Mechanism of dsRNA virus replication: Cloning, production and structural characterization of C-terminal domain of σNS

Bachelor´s Thesis

Student: Nelli Manoczki

Supervisor: RNDr. Zdenek Franta, Ph.D.
Co-Supervisor: Dr. Ivana Kuta-Smatanova
Institute of Chemistry and Biochemistry
Faculty of Science

University of South Bohemia in České Budějovice
Branišovská 1645/31a
370 05 České Budějovice

2019

Manoczki, N. V., 2019: Mechanism of dsRNA virus replication: Cloning, production and structural characterization of C-terminal domain of σNS. BSc. Thesis, in English, - 38 p. Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

A gene fragment encoding the C-terminal domain of the σNS protein of the avian reovirus was amplified via PCR. The amplicon was cloned using BsaI restriction enyzme into the pASK-IBA37+ expression vector and transformed into One Shot™ TOP 10 Chemically Competent *E. coli* cells. Colony PCR was performed and plasmids were sent for sequencing. Sequence verified plasmids were transformed into One Shot™ BL21 (DE3) or Rosetta-gami B (DE3) Chemically Competent *E. coli* cells. During Pilot expressions the conditions for production of soluble recombinant protein were optimized and according to them the recombinant protein was produced in large scale. The presence of the σNS-terminal domain was verified by Western Blot.

Affirmation

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 14.05.2019

.....

Manóczki Nelli Vanessza

Acknowledgement

I would like to express my greatest thankfulness to my supervisor, RNDr. Zdenek Franta, Ph.D. for his support, patience, teaching and explanatory skills during the practical realization of this project and later on during writing the thesis. Further I would like to thank the whole Institute of Chemistry and Biochemistry for the student-friendly atmosphere. Special thanks to James Tinnel for reading this thesis and checking it linguistically.

Abstract

Avian reovirus (AVR) is a doublestranded RNA retrovirus, infecting chicken embryos and thus causing considerable losses in the poultry industry. The non-structural σNS protein plays a key role in virus replication and packaging, but unfortunately our information about its structure is very limited. Previous attempts to acquire protein structure have failed, which could be caused by several disordered regions located within the protein. In this study, we tried to recombinantly produce the C-terminal domain of σNS (AA 152 - AA 367), which was identified previously by limited proteolysis. First, DNA sequence encoding for C-terminal domain of σNS was cloned into the pASK-IBA37+ vector carrying an N-terminal Histidin-tag, and propagated in the *E. coli* TOP10 cell line. Sequence verified plasmids were transformed either into *E. coli* BL21 (DE3) or *E. coli* Rosetta-gami B (DE3) expression cell lines. Protein production conditions were optimized performing PILOT expressions under various conditions. We found that the C-terminal domain of σNS is mostly present in inclusion bodies (insoluble phase) when using BL21 cells, while in the case of the Rosetta-gami cells, the domain can also be found in the soluble phase. The purification of recombinant protein from the soluble phase was attempted by using His-tag affinity Ni^{2+}-ion chromatography, but with no result. Western Blot analysis using anti-his antibody revealed that recombinant protein degrades in the soluble phase. Thus the desired C-terminal domain of σNS could not be produced for further structural characterization under these conditions and its recombinant production requires further optimization.

Table of Contents

List of tables……..……III
List of figures……..…...IV
Abbreviations……..……..VI
1.Introduction……..……1
1.1 About Viruses in General……..……1
1.2 Different Classification Systems……..……..1
1.2.1 Morphology……..……1
1.2.2 Chemical Composition and Mode of Replication…….…….3
1.2.3 The ICTV Taxonomy……..……...3
1.3 *Reoviridae*…….…….……4
1.4 The Avian Reovirus ARV…….…….……4
1.4.1 Their importance…….…….……..4
1.4.2 Pathology……..…….5
1.4.3 Life Cycle and Replication…….…...5
1.4.4 Viroplasms (Viral Factories)……..……6
1.4.5 Structure……..……...7
1.4.5.a Structural Proteins……..……7
1.4.5.b Non-structural Protein p19, p17, µNS…..……..9
1.4.5.c The Non-structural Protein σNS……..…….10
2.Aims……..……...13
3. Materials and Methods…….……14
3.1 Part A: Cloning of C-terminal σNS domain into the pASK-IBA37+ expression vector……..…….14
3.1.1 PCR amplification of C-terminal domain of ARV σNS protein…..…..14
3.1.2 Cloning of PCR amplicon into pASK-IBA37+ expression vector…...16
3.1.3 Transformation of *Escherichia coli* TOP10 cells…….……17
3.1.4 Verification of cloning success by performing colony PCR…….....18
3.2 Part B: Production and purification of recombinant protein……..……..18
3.2.1 Transformation of Escherichia coli BL21 (DE3) cells……..……18
3.2.2 Pilot expression……..……..18
3.2.3 Large scale protein production…….……....19

3.2.4 Transformation of Rosetta-gami B DE3 cell, pilot expression and large scale protein production……..……19
3.2.5 SDS-PAGE analysis……..……...20
3.2.5.a Pouring of polyacrylamide gels……..……..20
3.2.5.b Isolation of soluble and insoluble fraction from pilot expression……...……..21
3.2.6 Protein purification……..……22
3.2.7 Western Blot……..……...22
4. Results and Discussion……..……24
4.1 Part A: Cloning of C-terminal σNS domain into the pASK-IBA37+ expression vector……..…….24
4.2 Part B: Production and purification of recombinant protein……..……..27
4.2.1 Using *E. coli* BL21 DE3 cells……..……..……. 27
4.2.2 Using *E. coli* Rosetta-gami B DE3 cells……..……..……. 28
4.2.3 Troubleshooting……..…….30
5. Discussion……..…….33
6. Conclusion……..……35
7. Literature……..…….36

List of tables

Table 1. Sequences of the forward (sigmaNS_Cterm_162_pASK_F) and reverse (sigmaNS_wh_pASK_Nhis_R) primers for the ARV sigmaNS C-terminal domain (162-end AA).

Table 2. Gradient PCR for determining optimal annealing temperature.

Table 3. 1% Agarose gel, volumes needed for 150 mL.

Table 4. PCR of the σNS C-terminal domain, according to Q5® High-Fidelity DNA Polymerase Protocol by New England Biolabs.

Table 5. Restriction enzyme digestion in accordance with the protocol made by New England Biolabs.

Table 6. Detailed ligation protocol.

Table 7. The comparison of the different pilot expression reaction conditions using two different cell lines

Table 8. Protocol for preparing 12.5% polyacrylamide gel (resolving gel).

Table 9. Protocol for preparing 4% polyacrylamide gel (stacking gel).

Table 10. 4X Laemmli Sample Buffer.

Table 11. Used molecular ratios for the ligation.

Table 12. Concentration of isolated plasmids, which were sent for sequencing.

List of figures

Figure 1. An overview about the different shapes and structures of the virus types. Retrieved from: Cann, A. J. (2012). *Principles of molecular virology*. London: Elsevier.

Figure 2. Table of the virus taxinomy established by the ICTV committee. Modified and retreived from: The Organization of the ICTV. Retrieved on December 24, 2018, from https://talk.ictvonline.org/information/w/ictv-information/380/the-organization-of-the-ictv

Figure 3. A summary of the family members of *Reoviridae*, and its target organisms.

Figure 4. Simplified drawing of the dsRNA virus replication. Source: Benavente, J., & Martínez-Costas, J. (2007). Avian reovirus: Structure and biology. *Virus Research, 123*(2), 105-119. doi:10.1016/j.virusres.2006.09.005

Figure 5. The schematic structure of the avian reovirus particle Source: Benavente, J., & Martínez-Costas, J. (2007). Avian reovirus: Structure and biology. *Virus Research, 123*(2), 105-119.

Figure 6. A hypothesis of the working mechanism of the non-structural protein σNS. Source: Retrieved December 25, 2018, from Zamora, Paula F., et al. "Reovirus Nonstructural Protein σNS Acts as an RNA Stability Factor Promoting Viral Genome Replication." *Journal of Virology*, American Society for Microbiology Journals, 1 Aug. 2018, jvi.asm.org/content/92/15/e00563-18.

Figure 7. The σNS sequence after digestion with elastase. Source of the modified picture: "SigmaNS [Avian Orthoreovirus] - Protein - NCBI." *Current Neurology and Neuroscience Reports.*, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/protein/AGO32042.1?report=graph.

Figure 8. Vector pASK-IBA37+. Source: https://www.iba-lifesciences.com/isotope/2/2-1437-000-DS-2-1437-000-pASK-IBA37plus.pdf

Figure. 9. PCR amplification of the σNS C-terminal domain, using Q5 polymerase.

Figure. 10. Verification of cleaved vectors.

Figure. 11. Colony PCR.

Figure. 12. Pilot expression: *E. coli* BL21 DE3 cells, LB medium, 37°C, 225 rpm.

Figure. 13. Pilot expression: *E. coli* BL21 DE3 cells, LB medium, 30°C, 225 rpm.

Figure. 14. Pilot expression: *E. coli* BL21 DE3 cells, LB medium, 18°C, 225 rpm.

Figure. 15. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 37°C, 225 rpm.

Figure. 16. Pilot expression: Rosetta-gami B DE3 cells, LB medium, 37°C, 225 rpm.

Figure. 17. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 18°C, 225 rpm.

Figure. 18. Large scale protein production (Rosetta-gami B DE3 cells, TB medium, 37°C, 225 rpm).

Figure. 19. Western Blot of the fractions of the large scale protein production.

Figure. 20. Pilot expression: freshly transformed Rosetta-gami B DE3 cells, TB medium, 37°C, 225 rpm.

Figure. 21. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 37°C, 225 rpm.

Figure. 22. Western Blot of the fractions of the pilot expression (Rosetta-gami B DE3 cells, TB medium, 37°C), insoluble phase.

Abbreviations

AA – amino acid
APS - Ammonium persulfate
ARV – Avian reovirus
CNS – Central nervous system
CV – Column volume
DNA – Desoxyribonucleic acid
dNTPs - Deoxyribonucleotide triphosphates
dsRNA – Double-stranded RNA
EDTA – Ethylenediaminetetraacetic acid
GDP – Guanosine diphosphate
GMP – Guanosine monophosphate
gor - glutathione reductase
GTP – Guanosine triphosphate
HRP - Horseradish peroxidase
ICTV - International Comittee on Taxonomy of Viruses
IPTG - Isopropyl β-D-1-thiogalactopyranoside
LB – Lysogeny broth
OD – Optical density
OmpT - Outer membrane protease
ORF - Open reading frame
PCR – Polymerase chain reaction
PVDF - Polyvinylidene fluoride
RISC - RNA-induced Silencing Complexes
RNA – Ribonucleic acid
SDS - Sodium dodecyl sulfate
SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC - Super Optimal broth with Catabolite repression
ssRNA – Single-stranded RNA
TAE – Tris-acetate-EDTA
TB – Terrific broth
TBS – Tris buffered saline
TBS-T – Tris buffered saline and tween 20

TEMED – Tetramethylethylenediamine

Tris - *Tris*(hydroxymethyl)aminomethane

tRNAs – transfer RNAs

trxB - thioredoxin reductase

1. Introduction

1.1 About Viruses in General

Viruses are the smallest and the most abundant entities in our world. Their size ranges span between 30-90 nanometers, although there are a few exceptions, like eg. ebola being approximately 1000 nm long. They are characterised by the fact, that they can not reproduce themselves on their own, but are completely dependent on the host cell. It is because their genome is too small to code for ribosomes and sometimes even the own polymerase necessary for copying the genetic material is missing. Viruses contain only one type of nucleic acid, which is either RNA or DNA. In the case of RNA viruses, their genome is about 1.7-27 kbp long, while the genome of DNA viruses is bigger with approximately 3.2-200 kbp. The virus particles are surrounded by a protecting shell, capsid, which protects the virus from the destroying environment containing nucleases. Moreover, human and animal viruses have an additional lipid bilayer, an envelope (mostly coming from the host cell membrane), to prevent the genetic material from degradation. The cell attachment proteins can be found outside of the viral particle cell, which enable the adhesion to receptors or glycans on cell surfaces. The complete structure of the virus, comprising the genetic material, capsid and envelope if it is present, is called the virion. Since viruses have also no organelles and can not reproduce on their own, they are not considered to be living organisms. Their origin is unclear, but there are various hypotheses about their appearance on the Earth, including the controversial Panspermia Hypothesis, the Regressive hypothesis, the Cellular Origin or the Coevolution Theory. However proof for their presence can be found even among the earliest records of human activties; for example on ancient Egyptian depictions circa 3000 years B.C.[1]

1.2 Virus classification systems

The virus classification system is not uniform, but various aspects such as morphology, chemical composition and type of replication are used to classify the viruses.

1.2.1 Morphology

Two types of morphological features are distinguished: the helical symmetry and the icosahedral symmetry. In helical symmetry the capsid proteins (called protomers) are arranged in a helical way around the helical filaments, protecting the genetic material. These capsids

are usually rigid and have a rod-like or a filamentous shape, and can be further described by their length, width, pitch of helix and the number of protomers per helical turn. An example for this type is the tobacco mosaic virus. In the case of icosahedral symmetry, the virus capsid is shaped like a polyhedron, consisting of triangular faces faced to each other and vertices, and different rotational symmetries are possible. These viruses can have either polyhedral or spherical shape.[2] This thesis deals with the avian reovirus, which has an icosahedral morphology.[3] The protomers form oligomeric clusters, which are called capsomeres and this allows further classification according to capsomere numbers and patterns. Additionally, by determining the outer capsid and the envelope proteins of viruses, it is easier to find out the host range and antigenic composition of the virion.[2]

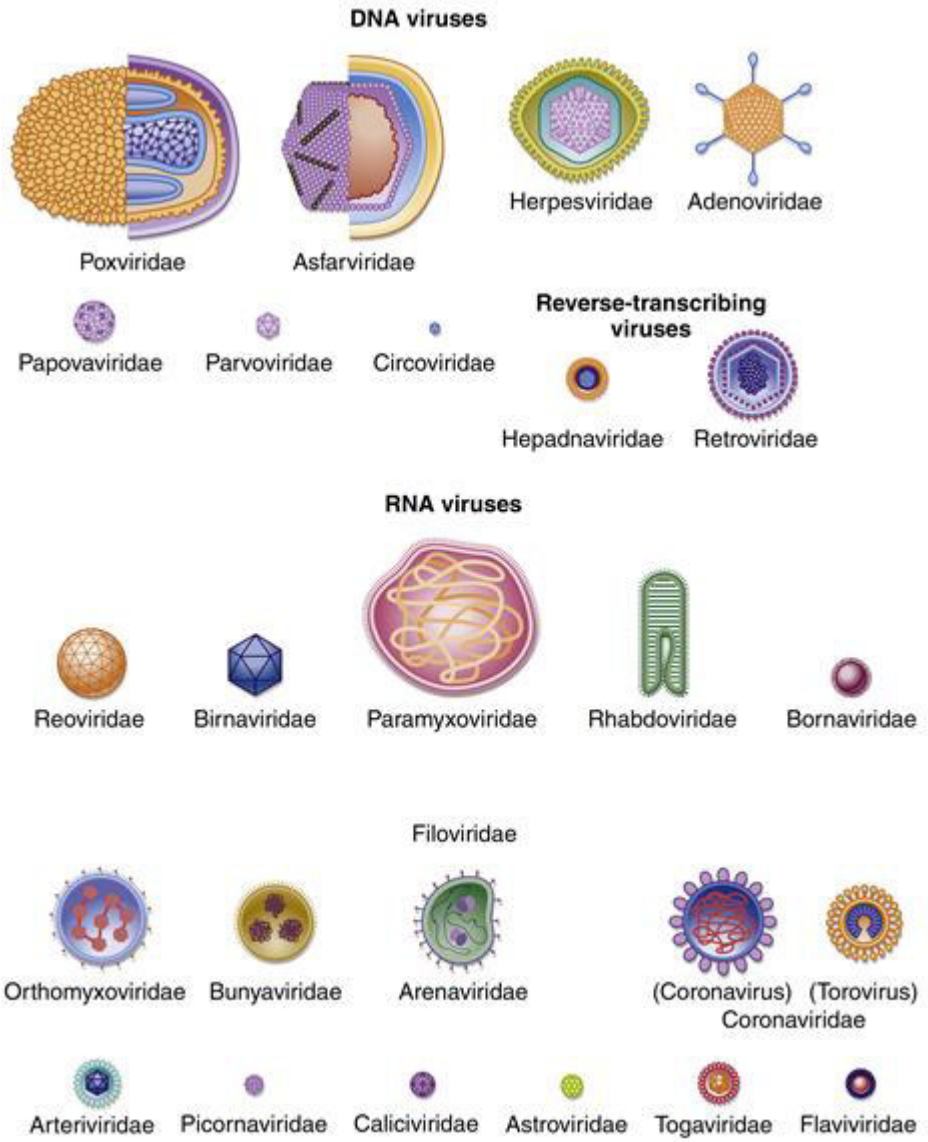

Figure 1. An overview about the different shapes and structures of the virus types. Retrieved from: Cann, A. J. (2012). *Principles of molecular virology*. London: Elsevier.

1.2.2 Chemical Composition and Mode of Replication

The genome of the viruses can be either RNA or DNA-based. In RNA viruses, the RNA can be single (ssRNA) or double-stranded (dsRNA) and these ones can form segments. During replication, only the sense strands carrying the genetic information can function as viral mRNAs. The replication of dsRNA viruses, like the avian reoviruses, is complex – they have the sense and antisense strands bound together by hydrogen bonds, forming a linear double-stranded molecule. Viruses, which contain DNA as genetic material, have mostly a linear, double-stranded genome, the only exceptions are the papovaviruses (including also the papillomaviruses), which have circular genomes.[2]

1.2.3 The ICTV Taxonomy

Futhermore it is possible to organize the virus species according to their evolutionary relationship. The International Comittee on Taxonomy of Viruses (ICTV) defined the terms like: order, family, sub-family, genus and species. The viruses were assigned into individual groups using modern nucleotide sequencing techniques. The current stand (after the meeting in Singapore, July 2017) consists of 6 groups, and is depicted in the picture below. The avian reoviruses belong to the *Reoviridae*, found in the group „Animal dsRNA and ssRNA- Viruses.[4]

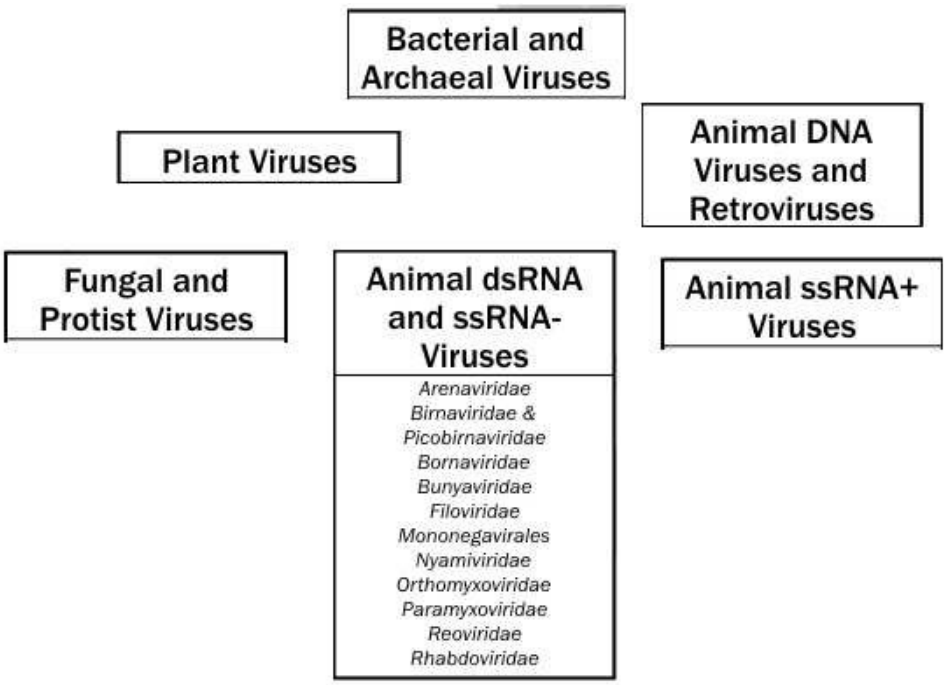

Figure 2. Table of the virus taxinomy established by the ICTV committee. Modified and retreived from: The Organization of the ICTV. Retrieved on December 24, 2018, from https://talk.ictvonline.org/information/w/ictv-information/380/the-organization-of-the-ictv

1.3 *Reoviridae*

Reovirus is a short name for Respiratory Enteric Orphan viruses. The family is unique in the absence of outer lipid envelope. The virions are naked, having a spheroidal shape surrounded by two icosahedral capsids. The core consists of double-stranded, segmented RNA.[5] There are around 87 virus species, infecting wide host range spanning from vertebrates to bacteria, which can be devided into 8 genera: *Orthoreovirus, Orbivirus, Coltivirus, Rotavirus, Aquareovirus, Cypovirus, Phytoreovirus,* and *Fijivirus*. Four of the eight genera infect humans, namely the *Orthoreoviruses, Orbiviruses, Coltiviruses,* and *Rotaviruses*. Although they attack the same hosts, they differ substantially in the type of diseases, their epidemological history and even in their antigens.[2,5] For instance, *Rotaviruses* cause severe diarrhea in children, while *Orbiviruses* are responsible for the blue-tongue disease in sheeps.[5] In this project we dealt with the avian reovirus, which belongs to the *Orthoreoviruses*.[2]

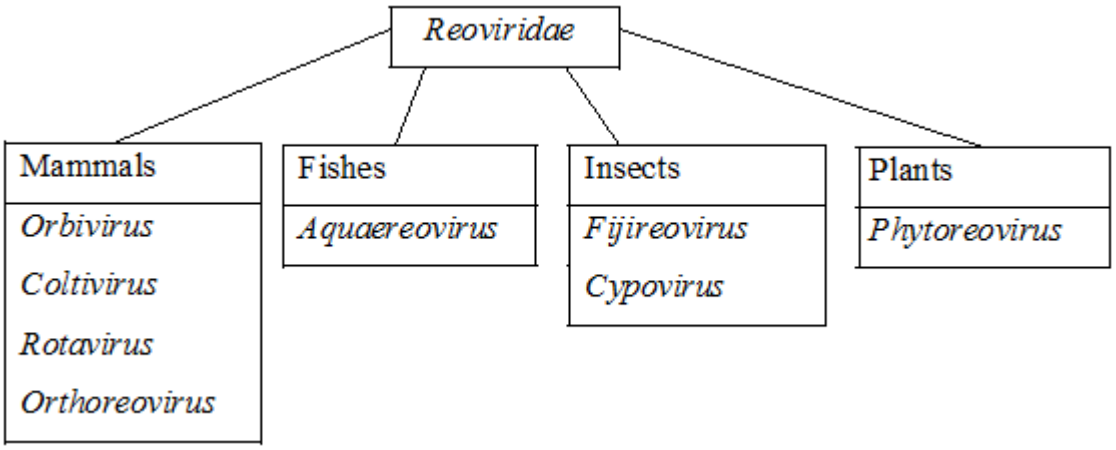

Figure 3. A summary of the family members of *Reoviridae*, and its target organisms.

1.4 The Avian Reovirus ARV

1.4.1 Their importance

ARV belongs to the *Orthoreoviruses*, which are primarly animal pathogenes.[2] This virus attacks birds but more importantly affects the poultry flocks in the food industry, causing significant economical losses.[7] As there is an increasing trend in the consumption of poultry meat, for instance in Europe the yearly consumption per capita increased from 11.2 kg (2000) to 20.3 kg (2007), the need for developing novel medications or vaccines against ARV becomes more urgent.[8] Moreover, the detailed understanding of the reoviruses biology in general could help us not only to protect the poultry flocks, but also to fight cancer by cancer cell oncolysis (reovirus induced cell death) using modified reoviruses. With this approach a

more specific anti-cancer therapy for the patients could be provided, but since there are still unclear parts of the virus cycle, the treatment has to be developed further till clinical practice becomes possible.[10]

1.4.2 Pathology

ARV mostly cause the asymptomatic infections, but it can induce tenosynovitis (an other name for this is the viral arthrytis syndrome). The disease is characterized by swollen joints and lesions in the gastrocnemius tendon. Besides it, the presence of the ARV was also observed in chickens suffering from enteric and respiratory diseases, myocarditis, hepatitis as well as from the stunting/malabsorption syndrome.[7] Moreover, new strain of ARV was isolated from a broiler chicken in Hungary in 2012 and associated with the CNS-disorder.[11] The chance to get infected with ARV is age-dependent, younger chickens get infected by ARV easier than older ones, indicating, that the older birds can acquire resistance against the virus. The fecal-oral route is the primary way of infection, but also infection via the respiratory tract or even transovarial infections have been recorded.[7]

1.4.3 Life cycle and replication

Infection happens, as the virus particle attaches to the cell surface via the cell attachment protein σC. This induces a conformational change, which triggers the clathrin mediated endocytosis of the virion into the host cell.[7] Then the virus particle gets uncoated by losing the outer capsid and the actively transcribing core is borned.[6] Plus-sense RNAs are transcribed from 10 dsRNA segments and get capped by the enzyme λC.[12] Then the positive strand viral RNAs leave the core through the turrets, enter the cell cytoplasm and get translated to produce the viral proteins. Plus-sense RNAs function as templates to produce the minus-sense RNAs, in order to reestablish the double-stranded genome structure.[6,12] Next the non-structural protein μNS forms viral factories and activates σNS and other viral factors.[6] Viral factories are inclusion bodies like structures, where the virus replication and assembly happen.[12] The protein σNS binds ssRNAs and help their interactions with each other and plays a role in forming an assortment complex, where structural proteins are gathered to build up the core again. As last, the outer layer is formed and the progeny virus is released.[6] More progeny viruses can produce viral proteins in big amounts and can continue to infect other cells.[12] Illustrated in Fig. 4 below.

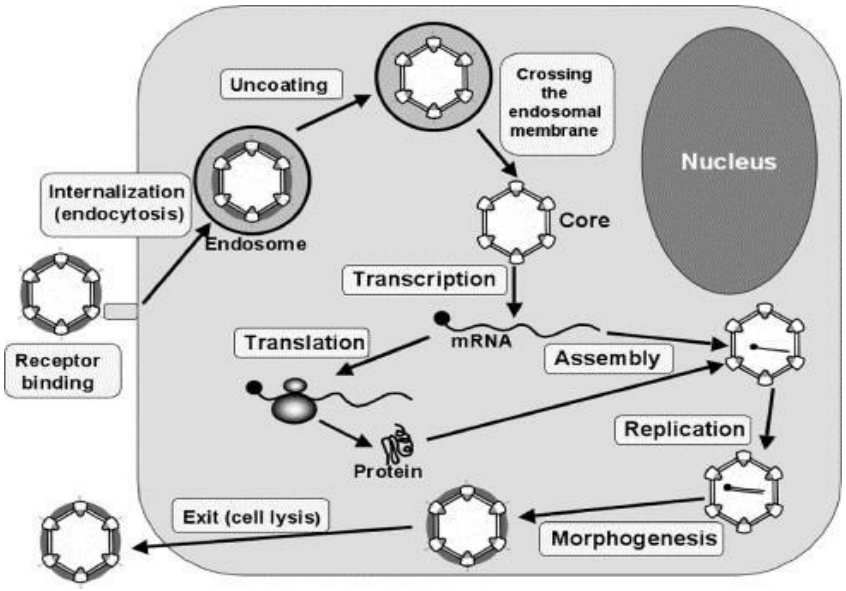

Figure 4. Simplified drawing of the dsRNA virus replication. Upon endocytosis of the virus the particle becomes uncoated. Positive strand RNAs are translated and the non-structural protein μNS mediates the establishing of the assortment complex (viroplasm) and recruits σNS, which acts as a chaperone to help the inter-segment RNA-RNA interactions. After replication by the RNA polymerase, the progeny virus is ready to leave the host cell. Source: Benavente, J., & Martínez-Costas, J. (2007). Avian reovirus: Structure and biology. *Virus Research, 123*(2), 105-119. doi:10.1016/j.virusres.2006.09.005

1.4.4 Viroplasms (viral factories)

Viroplasms are dense globular cytoplasmic structures, which contain factors for viral replication and morphogenesis.[13] These inclusions are not aggresomic structures, meaning that they show up no similarity with inclusions, where misfolded proteins are stored till degradation.[14] Their presence is important to prevent the positive strand RNAs from the cellular RNA-induced Silencing Complexes (RISC), which would degrade them. But before the ssRNAs are exported, the positive strands of the genome segments have to be selected – it can happen either by random incorporation or by selective packaging (which is the case in the avian reovirus).[6]

Their assembly due to the non-random selection of the genomic segments is possible through complex RNA-RNA and RNA-protein interactions. This process (the RNA assortment process) happens in the viral factories.[15] The attraction and concentration of the viral proteins into the viroplasms is made by the non-structural protein μNS, which controls it specifically and temporally.[16] This is possible due to its two coiled-coil domains, whose oligomerization induces the formation of viroplasms, besides μNS prevents the re-coating of the virus particle,

which would result in the termination of the replication.[6,12] However, not all viral proteins are attracted by μNS, because some are just subsequently activized – although unknown factors are not discovered yet. During the infection the inclusions get bigger, however smaller in number and their location is perinuclear.[17]

1.4.5 Structure

Avian reovirus has an icosahedral capsid with no envelope. ARV genome has approximately 23,5 kbp and consists of 10 linear segments of dsRNAs. The size of segments span from 0.9 to 3.8 kbp, thus, allowing for their classification according to the length: S (for small), M (for medium) and L (for large).[3] The segments are further divided to S1, S2, S3, S4 and M1, M2, M3 and L1, L2, L3, and with the exception of L2, the nucleotide sequence of all of them is known. Furthermore S1 is the only genomic segment, which codes for more than one primary translation product. All 10 segments of ARV code for 8 structural and 4 non-structural proteins. The proteins encoded by the S segments are called sigma (σ), by the M segments mu (μ) and by the L segment lambda (λ). They are further named with arabic alphabets refering to their size, determined by gel-electrophoresis.[7]

1.4.5.a Structural proteins

Eight structural proteins encoded by the ARV genome are λA, λB, λC, μA, μB, σB and σC. Their task is to build up the non-enveloped icosahedral structure with a diameter of 85 nm and with a buoyant density of 1.37 g/mL. The genetic material remains protected by two surrounding concentric protein shells, where μB, μBC, μBN, σB and σC form the outer capsid and the rest, λA, λB, μA, σA, forms the core (inner) shell.[7]

λA builds the inner core shell, protecting the virus genome and the RNA polymerase. It has been also detected in association with viroplasms in infected cells. λA has a hydrophilic amino-terminal region, which possesses an arm-like conformation with unknown function. Protein λB is only a minor component of the tiny viral core. It is the only protein, which protein sequence could not been determined so far, however the protein´s size and virion copy number is very similar to other RNA polymerases of the *Reoviridae* family, thus evoking its RNA polymerase function. Another protein encoded by the L segment is the λC (guanylyltransferase) protein, it spans from the inner core to the outer capsid, and its pentamers

form turrets. λC functions as the viral capping enzyme. It is the only structural protein binding GMP via a phosphoamide linkage, yielding cap structure after the GMP molecule is transferred to GDP or GTP acceptors. The turrets are built up by pentamers of λC and in the cavities the viral mRNAs at the 5´end get the caps while they exit the viral particle.[7]

Protein μ A is encoded by M segment and forms a small part in the inner shell. Although the aminoacid sequence of this protein has been recently identified, the protein function is still unknown.[6] μB is a major product of M2 segment, which is N-myristoylated. Protein μB is cleaved between ASN42-PRO43 by yet unknown protease yielding N-myristoylated cleavage products μBN and C-terminal protein μBC. Both μB and its cleavage products μBN and μBC are structural components of the outer capsid. Protein μBC plays except its structural function also an important role in internalization of ARV virion into the cell.[7]

The core shell protein σA encoded by the segment S, has various functions in ARV. Besides its ability to bind dsRNA in a sequence-independent manner σA also plays a key role in ARV morphogenesis. It is done by stabilising the core shell and helping the core-coating by outer shell proteins and has a role in the resistance to antiviral activity of interferon by blocking the activation of protein kinase, which is responsible for destroying foreign dsRNA in the body. While σA is found in the inner core, σB is present in the outer capsid. The function of σB protein is still unclear, however its rapid association with μB and μBC forming hetero-oligomeric complexes have been observed. The viral cell attachment protein σC is also encoded by the S segment. It is located on the outer capsid and can attach to avian cell monolayers, inducing specific antibody production.[7] The attachment to the host cell´s surface causes a conformational change in the virus particle, thus helping its endocytosis into the cell.[6] It has been shown that σC can even cause apoptosis when expressed in transfected cells, however this phenomena is not cleared yet, since viral gene expression is not required for ARV induced apoptosis.[7] The summary of the viral proteins located in the virion is shown in Fig. 5.

Besides 10 dsRNA segments, the viral particles contain also small, single-stranded oligonucleotides, rich in adenine, however their function is yet to be discovered.[7]

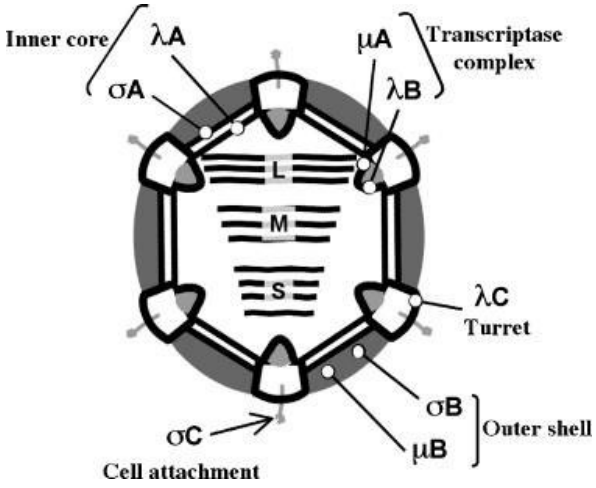

Figure 5. The schematic structure of the avian reovirus particle. The 10 dsRNA genome segments are protected by the inner core (σA, λA) and the outer core (σB, μB). The protein λC forms the turret, while λB and μA are small part of it. The cell attachment protein is the σC, providing contact surface on the future host cell. Source: Benavente, J., & Martínez-Costas, J. (2007). Avian reovirus: Structure and biology. *Virus Research, 123*(2), 105-119.

1.4.5.b Non-structural protein p10, p17, μNS

The non-structural proteins do not play any role in building up the viral capsid, but they are detected in the cytoplasm of infected cells, fulfilling functions, which are required for the replication. The M3 segments encodes for μNS and the segment S4 encodes for σNS. Two other non-structural proteins are encoded by S1 segment and are called p10 and p17.[7]

The μNS is 635 aminoacids long and has a size of approximately 70 kDa. The cleavage of this protein was observed in infected cells, resulting in the production of μNSN (~ 15 kDa) and μNSC (~55 kDa). This protein has a key role in the early virus morphogenesis by being the minimal factor needed for the formation of viral factory.[7,18] The two C-terminal coiled-coil domains take part in the oligomerization process thus helping the formation of viroplasms.[6] However it was shown, that only the bigger μNS isoform interacts with the protein σNS to form the viroplasm inclusions by mediating the association of σNS into these.[18] Furthermore μNS recruits even the structural protein λA to form the core.[17]

The protein p10 is encoded by the first open reading frame ORF of the segment S1 and has a size of 10.3 kDa. It is a transmembrane protein, causing cell-cell fusion in the host and its presence increases the permeability of the plasma membrane of the host cell, where its

hydrophobic N-terminal domain is supposed to play a key role. The other protein encoded by S1 segment is p17. Interestingly, it has no known homology with any other viral and even cellular proteins. P17 has the ability to bind DNA and possess a nuclear localization signal at its C-terminus. The role of thus protein remains unknown.[7]

1.4.5.c The non-structural protein σNS

The protein σNS is encoded by the gene segment S4. The protein consisting of 367 aminoacids has a size of about 40 kDa.[6] σNS binds the ssRNA of minimal size between 10-20 nucluotides, in sequence independent manner.[7] Protein was found to be present in the cytoplasm of infected cells in the form of large ribonucleoprotein complexes, associated with the non-structural protein μNS. The RNA-free form of σNS exists as a homodimer and homotrimer – which then form hexamers.[6] It was shown, that the binding of σNS to ssRNAs depends on the secondary structure of the RNAs and is mediated by 5 conservative basic residues, which are spread throughout the whole σNS sequence.[19] Upon binding of σNS, the ssRNA becomes unfolded, because σNS removes their secondary structure by melting the disulfide bonds between them.[3,19] Interestingly the RNA-binding does not depend exclusively on the size of the sequence, but also on the type of the sequence. An *in vitro* study showed, that σNS prefers poly(A), poly(U) and ssDNA, and binds them sequence-independently, while shows no binding for viral mRNA sequences.[20] In contrast to the non-structural proteins of other dsRNA viruses, which act as helicases, σNS does not require any ATP for its functions. This protein behaves like a chaperone while mediating the RNA-RNA interactions, which are required for the viroplasm assembly.[6] The chaperon function of σNS was confirmed by recent papers, whose authors suggest, that the driving force is the binding affinity between σNS and the RNAs.[21] The role of σNS is similar to the single-stranded DNA-binding proteins, which help the replication of genomic DNA, however ARV σNS binds RNA instead of DNA.[22]

A hypothesis for the working mechanism of the protein σNS is depicted in Fig. 6.

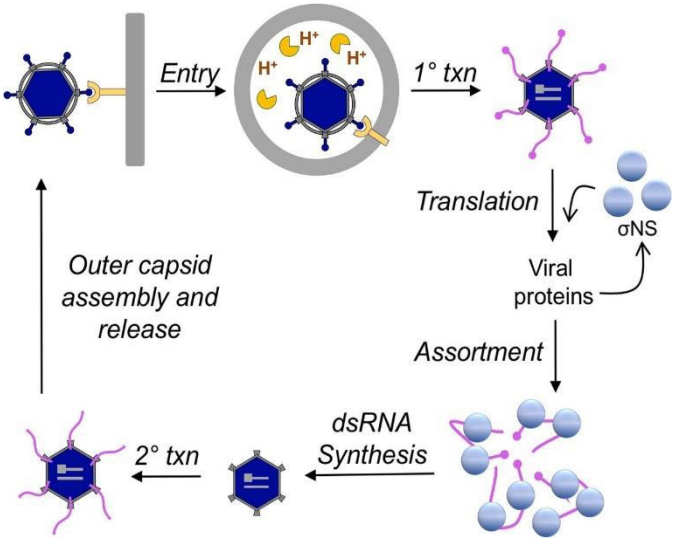

Figure 6. A hypothesis of the working mechanism of the non-structural protein σNS. During the endocytosis the 10 genomic segments do not interact with each other. After transcription and translation (primary transcriptase particle), the σNS is activated by μNS, forming a complex with the ssRNAs. It acts as a chaperone by melting the disulfide bonds (and thus changing the secondary structure of the RNAs) in order to facilitate the inter-segment RNA-RNA interactions between them, creating an assortment complex. After selecting the template RNAs, σNS leaves the complex and the RNA polymerase copies the choosen sequences. The cycle repeats as the progeny virus can infect further cells and produce the viral proteins. Source: Retrieved December 25, 2018, from Zamora, Paula F., et al. "Reovirus Nonstructural Protein σNS Acts as an RNA Stability Factor Promoting Viral Genome Replication." *Journal of Virology*, American Society for Microbiology Journals, 1 Aug. 2018, jvi.asm.org/content/92/15/e00563-18.

A recent study shows, that upon binding to ssRNA, the σNS hexamer transforms into an octamer, which destabilizes the helix activity, thus creating a new RNA-binding surface.[10] Moreover, it was found out, that once RNA is bound to the σNS, protein becomes protected from proteolytic digestion (J. Bravo manuscript in preparation). This could be explained by the fact, that either the σNS hexamer gets stabilised or that dynamics of the flexible regions becomes reduced. Due to the fact, that σNS has no sequence homology to any other known viral protein, and there is no crystal structure present, the exact function of the protein as well as the binding stoichiometry for the RNA - σNS complex has to be determined. [6]

Although there were several attempts to produce diffracting crystals from the whole σNS protein, they failed most likely due to the disoredered regions present in the protein sequence. These regions prevent the X-ray to diffract on the crystals, because they are too dynamic (flexible). Limited proteolysis approach was used to identify σNS dynamic regions yielding two regions: the N-terminal fragment and the C-terminal one (see Fig. 7.).[6] In this work we present the production of C-terminal domain of σNS in *E. coli* expression system.

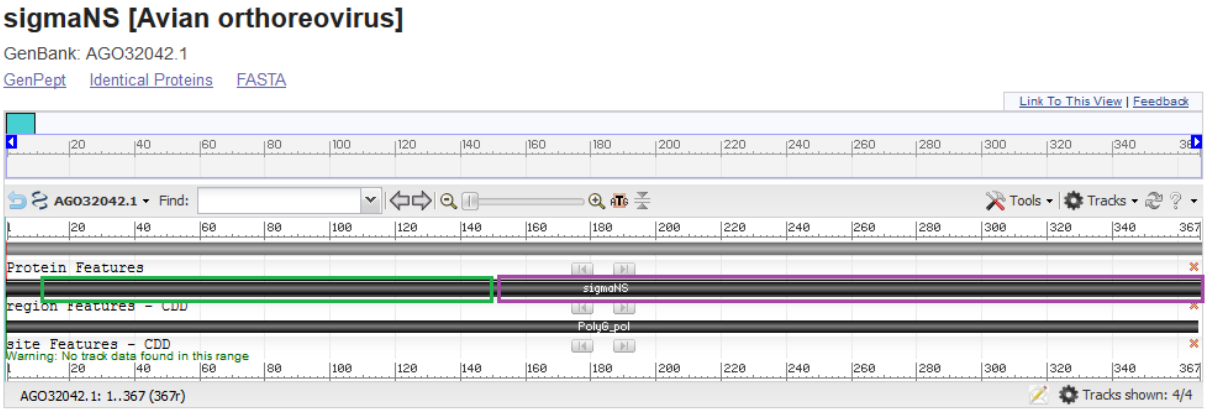

Figure 7. The σNS sequence after digestion with elastase. The disordered regions at bp 145-152 were cut out, resulting in two ends: N-terminal (12 bp to 145 bp, marked with green) and the C-terminal residue (152 bp to 367 bp, marked with lila). Source of the modified picture: "SigmaNS [Avian Orthoreovirus] - Protein - NCBI." *Current Neurology and Neuroscience Reports.*, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/protein/AGO32042.1?report=graph.

In the thesis by Higginbotham (2017) the C-terminal domain was cloned into the vector pET22 (Novagen), which has a T7 promotor and can be induced using IPTG. The recombinant protein was expressed in the BL21 DE3 cells at 21°C, however the protein was found in the insoluble phase. For further analysis, different solubilisation conditions were tested using various detergents and metal ions, and changing the concentration of NaCl, but the recombinant protein of the C-terminal domain remained in the insoluble phase. Therefore the C-terminally truncated end could not be crystallized either, because it was not possible to obtain this protein in the soluble phase.[23]

2. Aims

- To clone the C-terminal domain of σNS into expression vector.
- To optimize the protein production conditions.
- To purify the C-terminal domain of the non-structural protein σNS, which could be used to determine its structure by X-ray crystallization.

3. Materials and methods

3.1. Part A: Cloning of C-terminal σNS domain into the pASK-IBA37+ expression vector

3.1.1 PCR amplification of C-terminal domain of ARV σNS protein

The plasmid DNA containing whole σNS sequence (GenBank: AY303992.1) was given to our lab as a general gift by Dr. Roman Tuma, University of Leeds. Genespecific primers for the PCR amplification of the C-terminal domain were designed and verified using OligoAnalyzer 3.1 (https://eu.idtdna.com/calc/analyzer).

Table 1. Sequences of the forward (sigmaNS_Cterm_162_pASK_F) and reverse (sigmaNS_wh_pASK_Nhis_R) primers for the ARV sigmaNS C-terminal domain (162-end AA). The sequence of BsaI enzyme is underlined. Gene specific sequence in bold.

Forward primer:	
Sequence:	5'- ATGGTA GGTCTC A G CGC **CCTATTGACTACCTCGCTGCT**– 3´
Length:	21
GC-content:	52.4%
Melting temperature:	56.3°C
Reverse primer:	
Sequence:	5´ - ATGGTA GGTCTC A TATC TTA **CGCCATCCTAGCTGGAGA** – 3´
Length:	18
GC-content:	61.1%
Melting temperature:	56.5°C

A gradient PCR was performed to determine the optimal primer annealing temperature using Takara Ex Taq DNA Polymerase (RR001A, TAKARA BIO INC), and the following temperatures: 65°C, 56.2°C, 53.8°C, 48.6°C, 45°C, see Table 2. Results were analyzed on 1% agarose gel, see the preparation steps in Table 3. The samples and the DNA ladder (100 bp DNA #N3132S New England Biolabs) were loaded in the gel and run at constant voltage 100 V for 40 min. The picture was taken by the Syngene G:BOX Chemi XX9 and edited using the software GeneSys (Copyright © 2009 - 2018 Syngene, A Division of Synoptics Ltd, https://www.syngene.com/).

Table 2. Gradient PCR for determining optimal annealing temperature.

Gradient PCR	
Components (20 µL in total)	In µL
10X Takara Ex Buffer (Mg^{2+} plus)	2
20 mM dNTPs mixture	0.2
10 uM Forward primer	2
10 uM Reverse primer	2
Takara Ex Taq DNA Polymerase	0.1
Plasmid DNA (243 ng/µL)	0.5
Nuclease-free water	13.2

Table 3. 1% Agarose gel, volumes needed for 150 mL.

1% Agarose gel (150 mL in total)	
Components	
Agarose	1.5 g
50X TAE Buffer stock solution (1X TAE Buffer: 40mM Tris, 20mM acetic acid, 1mM EDTA, pH 7.6)	3 mL
dH_2O	147 mL
SYBR Gold Nucleic Acid Gel Stain (S11494 Invitrogen)	3 µL

Then the C-terminal domain was amplified by PCR using Q5 High-Fidelity polymerase (M0491S, New England Biolabs) following the protocol provided by the manufacturer (see Table 4). The correct size of the amplicon was verified by 1% agarose gel electrophoresis as described above in the gradient PCR section, with the exception of using 7.5 µL of the dye SERVA DNA Stain Clear G (HS: 38220000, SERVA) instead of SYBR Gold.

Table 4. PCR of the σNS C-terminal domain, according to Q5® High-Fidelity DNA Polymerase Protocol by New England Biolabs.

PCR	
Components (50 µL in total)	In µL
5X Q5 Reaction Buffer	10
10 mM dNTP mixture	1
10 uM Forward primer	2.5
10 uM Reverse primer	2.5
Q5 High Fidelity Polymerase	0.5
Plasmid DNA	1
Nuclease-free water	32.5

The amplicon was purified using Nucleospin: Gel and PCR clean up kit - Macherey-Nagel according to the protocol provided by the producer. The amount of purified DNA was measured by the spectrophotometer B-80-3004-31 (IMPLEN GmbH), yielding the concentration of 183 ng/µL.

3.1.2 Cloning of PCR amplicon into pASK-IBA37+ expression vector

Expression vector pASK-IBA37+ (2-1437-000, IBA Life Sciences), was chosen for the recombinant production of C-terminal domain of σNS (see Fig. 8 for the constitution of the vector). Both PCR amplicon as well as expression vector were digested using BsaI-HF restriction enzyme (R3535S, New England Biolabs) according to the protocol provided by the manufacturer using 1 µg of either plasmid or amplicon for the reaction. The detailed set up of the digestion reaction is shown in Table 5.

	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

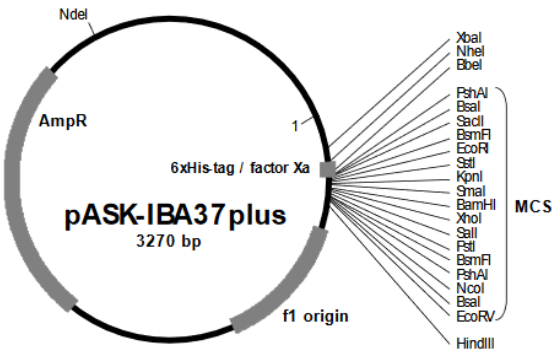

Figure 8. Vector pASK-IBA37+. Source: https://www.iba-lifesciences.com/isotope/2/2-1437-000-DS-2-1437-000-pASK-IBA37plus.pdf

Table 5. Restriction enzyme digestion in accordance with the protocol made by New England Biolabs.

Restriction enzyme digestion		
Components (in total 50µL)	In µL	In µL
Amplicon DNA (1µg)	5.5	---
Vector DNA (1µg)	---	4.1
10X CutSmart Buffer B7204S by New England Biolabs	5	5
BsaI-HF R3535S by New England Biolabs	1	1
Nuclease-free water	38.5	39.9

The restriction reaction was incubated at 37°C for 15 min. The enzyme was deactivated at 65°C for 20 min and the reaction was placed on ice until the next step. 10 µL of the digested vector was loaded in the 1% agarose gel to verify the digestion and the rest of the digested vector was dephosphorylated using (37°C for 30 min.) 2.5 µL of rSAP Shrimp Alkaline Phosphatase (M0371, New England Biolabs) and deactivated at 65°C for 5 min. Then both the digested PCR amplicon as well as digested and dephosphorylated vector were joined together by ligation using Instant Sticky-end Ligase Master Mix (M0370, New England Biolabs) according to the manufacturer´s instructions using the molar ratio of 1:3 vector and amplicon. The NEB Online Ligation Calculator (NEBioCalculator 1.8.1. https://nebiocalculator.neb.com/#!/main) was used to calculate the exact amounts. The detailed ligation protocol is shown in Table 6.

Table 6. Detailed ligation protocol.

Ligation	
Components (in total of 10 µL)	In µL
Vector pASK-IBA37+	3
Amplicon DNA	1.701
Water	0.299
Instant Sticky-end Ligase Master Mix M0370	5

3.1.3 Transformation of *Escherichia coli* TOP10 cells

The 50 µL of One Shot™ TOP 10 Chemically Competent *E. coli* cells (C404010, Invitrogen) were transformed by adding 5 µL of the ligation mixture into the cells. The mixture was incubated on ice for 30 min, followed by heat shock at 42°C for 30 s in the water bath. Cells

were placed on ice for 2 min, and then 250 µL of SOC medium (15544-034, Invitrogen) was added. The vial was shaked horizontally for 1 h at 37°C at 225 rpm (Eppendorf Innova S44i). Next 50 µL and 200 µL of cells were plated on two agar plates containing 100 µg/mL ampicillin (13399.3, SERVA). The petri dishes were left overnight in the incubator at 37°C.

3.1.4 Verification of cloning success by performing colony PCR

In the clean bench 20 cell colonies and one negative control (agar medium free of any colonies) were transferred in 1.5 mL eppendorf tubes containing 20 µL of dist. water using sterile tips. Then colony PCR was performed as described in Table 2, with the difference, that the initial denaturation step was set to 6 min. to lyse the bacterial cell walls. The cloning success was verified by running 1% agarose gel electrophoresis, as stated in Table 3. Positive clones (potentially carrying the σNS C-terminal domain) were propagated in 10 mL LB medium (48501.01, SERVA Liquid Broth Medium Powder) supplemented with 100 µg/mL ampicillin overnight (37°C, 225 rpm) in Eppendorf Innova S44i. Next day the plasmid DNA was isolated using NucleoSpin Plasmid: DNA, RNA and protein purification kit by Macherey-Nagel, according to the instructions of the manufacturer. Plasmid amount and purity was measured using the spectrophotometer (B-80-3004-31, IMPLEN GmbH) and sent to the company SEQme (https://www.seqme.eu/en/) to verify cloned sequences. The samples were prepared according to the protocol provided by the company. The obtained results were analyzed by the software Geneious (Copyright © 2005-2018 Biomatters Limited. https://www.geneious.com/).

3.2 Part B: Production and purification of recombinant protein

3.2.1 Transformation of *Escherichia coli* BL21 (De3) cells

One Shot™ BL21 (DE3) Chemically Competent *E. coli* cells (C600003, Invitrogen) were were transformed by plasmid containing verified sequence of σNS C-terminal domain as described in section 3.1.3.

3.2.2 Pilot expression

One cell colony was picked from the agar plate and placed in 10 mL of LB medium with 100 µg/mL of ampicillin. Cells were grown overnight in the horizontal shaker at 37°C and 220

rpm. Two falcon tubes (labelled as induced, uninduced) containing 10 mL LB medium with 100 µg/mL ampicillin were inoculated by 0.5 mL of the overnight culture. Cells were grown in the horizontal shaker at 37°C, till the cell density reached the OD range of 0.6-0.8. Cells were cooled down and the protein production was induced by adding 1 µL of anhydrotetracyclin (Alfa Asear, 2000 µg/mL in ethanol). Both induced and uninduced cells were incubated in the horizontal shaker at 37°C, 30°C or 18°C and 220 rpm. Protein production was monitored by harvesting (0.5 mL) right before induction and in 1,2,3,4,5,6 and 24 h intervals post induction. Uninduced cells were collected at the same time points. Collected cells were centrifuged at 11,000 g for 1 min and the pellets were frozen prior further analysis.

3.2.3 Large scale proten production

The large scale production of recombinant σNS C-terminal domain was carried out under condition identifed by pilot expression using 30°C. The overnight culture was inoculated into 800 mL of LB medium with ampicillin (100 µg/mL). Once the cell density reached the OD 0.6, protein production was induced by adding 80 µL of anhydrotetracyclin (2000 µg/mL). Cells were harvested after 6 h and spinned down at 4250 rpm at 4°C for 30 min (Hermle Z383K). Cell pellets were stored in 50 mL falcon tubes at -80°C until further analysis.

3.2.4 Transformation of Rosetta-gami B DE3 cell, pilot expression and large scale production

Rosetta-gami B DE3 Competent Cells (71136, Novagen) were transformed by *pASK-IBA37+* plasmid containing the correct σNS C-terminal DNA sequence as described in section 2.1, except having 20 min incubation time instead of 30 min. The pilot expression was carried out like written in section 3.2.2, with slightly different conditions: 37°C in LB, 37°C in TB (SIGMA Terrific Broth EZMIX^{TM} powder microbial growth medium), 18°C in TB. All media were supplemented with chloramphenicol (35 µg/ml) and ampicillin 100 µg/mL. Cells were obtained either from glycerol stock, fresh transformation or from the plate carrying a few days/a week old transformed cells. Since the temparature of 37°C and TB media were proved to be the most optimal ones, the cells were grown under these condition during the large scale production. The procedure remained the same like in the case of the BL21 cells. The cells were harvested after 2h, centrifuged and stored as described in section 3.2.2. See Table 7 for the

comparison of the two different pilot expressions using different cell lines and reaction conditions.

Table 7. The comparison of the different pilot expression reaction conditions using two different cell lines.

Comparison of pilot expressions		
	BL21 DE3 cell line	Rosetta-gami B DE3 cell line
Medium	LB	LB, TB
Temperature	18°C, 30°C, 37°C	18°C, 37°C
Time of induction	6h, 24h	6h, 24h
Ampicillin	100 µg/mL	100 µg/mL
Chloramphenicol	-	35µg/mL
Anydrotetracyclin	0.2 µg/mL	0.2 µg/mL
rpm	220	220

3.2.5 SDS-PAGE analysis

3.2.5.a Pouring of polyacrylamide gels

Resolving gels were prepared using Mini-Protean Tetra Handcast Systems (Bio-Rad) as summerized in Table 8.

Table 8. Protocol for preparing 12.5% polyacrylamide gel (resolving gel).

12.5% Resolving gel (for 1 gel)	
30% Acrylamide	2.08 mL
dH_2O	1.57 mL
1.5M Tris (pH=8.8)	1.25 mL
10% SDS	50 µL
10% APS	50 µL
TEMED	5 µL

Gels were overlayered with isopropanol and let to polymerize for about 45 min. Meanwhile the stacking gel was prepared as summerized in Table 9.

Table 9. Protocol for preparing 4% polyacrylamide gel (stacking gel).

4% Stacking gel (for 1 gel)	
30% Acrylamide/Bis solution 29:1 (SERVA)	340 µL
dH_2O	1.36 mL
0.5M Tris (pH=6.8)	250 µL
10% SDS	20 µL
10% APS	20 µL
TEMED	2 µL

Prior adding the stacking gel, the isopropanol was poured out and the edges of the resolving gels were washed with distilled water. Stacking gel was poured onto the resolving one and the comb was carefully inserted to avoid bubbles. The gel was let to polymerize for about 30 min. Prepared gels were wrapped into wet paper towels, put into plastic bags and stored in the fridge.

3.2.5.b Isolation of soluble and insoluble fraction from pilot expression

Cell pellets from the pilot expression were thawn and resuspended in 500 µL of lysis buffer (50mM potassium phosphate, 400mM NaCl, 100mM KCl, 10% glycerol, 0.5% Triton X-100, 10mM imidazole, pH 7.8). Cells were lysed by three steps of freezing in $N_{2(liq)}$ and thawing at 42°C. The soluble and insoluble cell fraction were separated by centrifugation at 22,240 rpm, 4°C for 5 min (SIGMA 3-30K). 4X reducing SDS-PAGE sample buffer (for 1 mL: 100 µL beta-mercaptoethanol, 900 µL Laemmli buffer 4x) were added and samples were boiled at 95°C for 5 min. See Table 10 for the preparation of 4X Laemmli Sample Buffer.

Table 10. 4X Laemmli Sample Buffer.

4X Laemmli Sample Buffer (in total 45 mL)	
SDS	4 g
Bromophenol blue	0.1 g
0.5M Tris (pH=6.8)	20 mL
Glycerol	20 mL
dH20	Filled up to 45 mL

20 µL of the samples and 10 µL of the protein marker (26616, PageRuler Prestained Protein Ladder, Thermo Fisher) were loaded in the 12.5% polyacrylamide gel and gel electrophoresis was performed at 150 mA, 100 V for 90 min. Then the gels were stained in the dye (Coomassie Brilliant Blue R250) on the rocking platform for 1 hour and after destaining in destaining solution (60% dist. water, 30% methanol, 10% acetic acid) a picture was taken using the software GeneSys and the camera Syngene G:BOX Chemi XX9.

3.2.6 Protein purification

Cell pellets from large scale production were thawn and resuspended in lysis buffer (50mM Tris HCl, 500mM NaCl, 1mM EDTA, pH = 7.6) supplemented with 1X protease inhibitor (SLBS9493, SIGMAFAST^{TM} 100X Protease Inhibitor Coctail Tablets EDTA-free). Cells were lysed using French press (EW FB50423, Fisherbrand), and ultracentrifugated for 1 h at 4°C at 25,000 g (himax CP90WX, HITACHI). The obtained soluble phase was purified via ÄKTA pure chromatography system (GE Healthcare Life Sciences), using HisTrap^{TM} HP-5 mL affinity column (GE Healthcare Life Sciences). The machine was calibrated by 3 column volumes (CV) of the equilibration buffer (20mM Tris, 50mM NaCl, 0.02% NaN_3, pH=7.8, vacuum filtrated and degased) and 3 CV of elution buffer (20mM Tris, 50mM NaCl, 1M imidazol, 0.02% NaN_3, pH=7.8, vacuum filtrated, degased) and again 3CV of equilibration buffer at flow rate 5 mL/min. Then the sample was loaded (1 mL/min) and flow-through was collected. The elution of the bounded protein from the column was done by increasing the concentration of the elution buffer (from 0-100% within 20 min).

3.2.7 Western Blot

Protein samples were separated on SDS-PAGE and the gel was placed in the prepared blotting sandwich. The blotting was performed for 30 min using Trans-Blot Turbo Transfer Pack (Bio-Rad). The PVDF membrane was washed 3x for 10 min in TBS-T (for 1L TBS-T: 100 mL 10X TBS, 1 mL 20% Tween, rest water), and blocked for 1 h in 20 mL of blocking solution (for 30 mL: 0.15 g blocking reagent, 3 mL 10X blocking buffer both from „ID:34460 Penta-His HRP Conjugate Kit“ by Qiagen) on rocking platform. The membrane was washed again with 3x for 10 min TBS-T and placed in a sealable plastic bag with 10 mL of blocking solution supplemented with 5 µL of anti-his specific antibodies (157028240, Penta-His HRP Conjugate, Qiagen) and incubated for 1h on rocking platform. The membrane was washed 2x

for 10 min with TBS-T and 1x for 10 min in TBS. The signal was detected using Bio-Rad „Opti-4CN Substrate Kit 170-8235 “ according to the manufacturer protocol. A digital image was taken using the software GeneSys and the camera Syngene G:BOX Chemi XX9.

.

4. Results

4.1. Part A: Cloning of C-terminal σNS domain into the pASK-IBA37+ expression vector

The optimal annealing temperature determined by gradient PCR was 56°C (data not shown). This temperature was used for further amplification of σNS C-terminal domain using Q5 High Fidelity Polymerase, yielding a 618 bp band. After the purification of the amplicon, its total DNA content was measured to be 183 ng/µL.

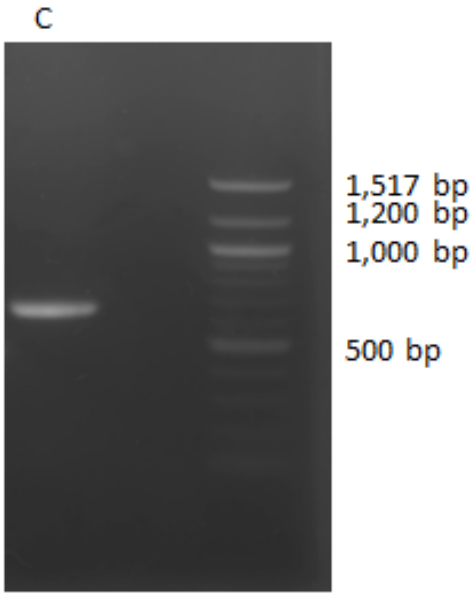

Figure 9. PCR amplification of the σNS C-terminal domain, using Q5 Polymerase. C: σNS C-terminal domain. Expected size: 618 bp. DNA ladder: 100 bp DNA #N3132S New England Biolabs, 1% agarose gel.

Then both the purified DNA sequence of interest and the expression vector pASK-IBA37+ were cut by the restriction enzyme BsaI-HF. Cleaved vector was dephosphorylated by adding Shrimp Alkaline Phosphatase and the cleavage was verified on 1% agarose gel (see Fig.10.) Both cleaved amplicon and the cleaved vector were ligated together using Instant Sticky-end Ligase Master Mix, at molar ratios shown in Table 11, using the NEBio calculator (New England Biolabs).

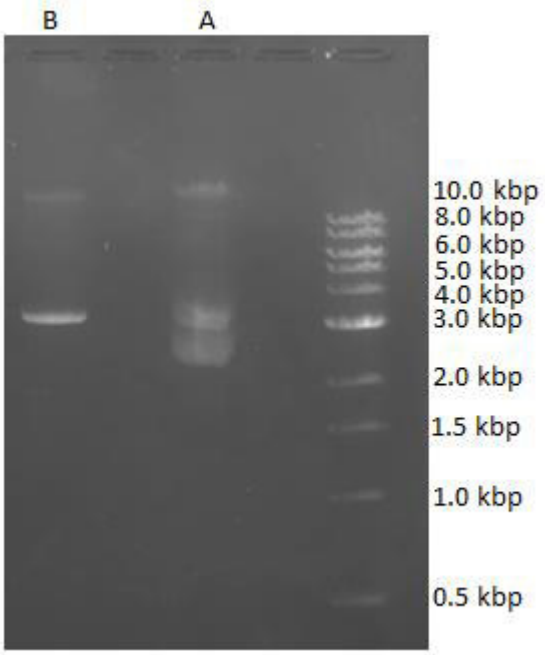

Figure 10. Verification of cleaved vectors. A: original, uncleaved vector, B: cleaved vector with expected size at about 3270 bp. DNA ladder: 1 kbp DNA ladder N3232S (New England Biolabs), 1% agarose gel.

Table 11. Used molecular ratios for the ligation.

Molecular ratios			
Vector : sample 1 : 3	vector	60 ng	3 uL
	sample	34.02 ng	1.701 uL
	water		0.299 uL
	In total		5 uL

One Shot TOP 10 Chemically Competent *E. coli* cells were transformed by the ligation mixture plated and grown overnight. A colony PCR was performed using gene specific primers and analyzed on agarose gel, see Fig. 11. 6 positive clones (number 1,3,4,5,6 and 9) were grown overnight and the plasmids were isolated, measured (see Table 12) and sent for sequencing. According to the sequencing results, all 6 clones carried the correct DNA sequence with no frame shift. The sample number 1 was chosen for the further work.

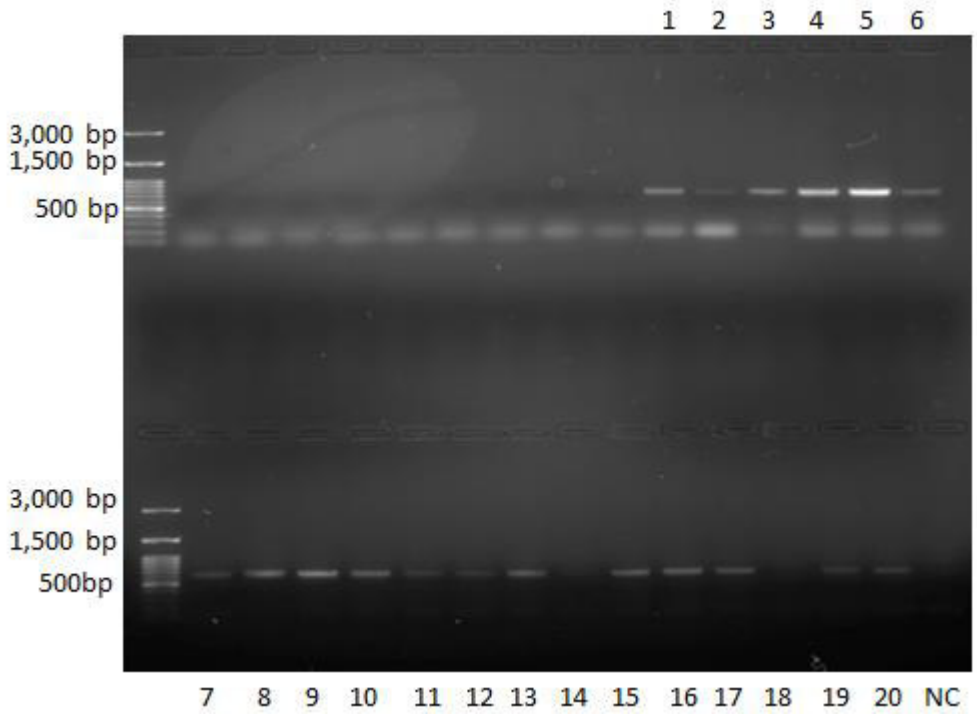

Figure 11. Colony PCR. Samples 1-20: positive controls, sample NC: negative control. Expected size of samples: 618 bp. DNA ladder: 100 bp DNA ladder H3 RTV Nippon Genetics, 1% agarose gel.

Table 12. Concentration of isolated plasmids, which were sent for sequencing.

Concentration of isolated plasmids	
Sample 1	70 ng/µL
Sample 3	35.4 ng/µL
Sample 4	67.9 ng/µL
Sample 5	44.4 ng/µL
Sample 6	76.3 ng/µL
Sample 9	43.4 ng/µL

4.2 Part B: Production and purification of recombinant protein

4.2.1 Using *E. coli* BL21 DE3 cells

First we tried to produce the recombinant protein in *E. coli* cells (BL21 DE3). To optimize the production conditions, pilot expression was carried out. We applied three different temperature (37°C, 30°C, 18°C) and collected samples at defined time points within 6h or 24h post induction with 0.2 µg/mL anhydrotetracycline. The resulting fractions were analyzed by SDS-PAGE. Our results clearly show, that the recombinant σNS C-terminal domain was found in the insoluble phase, just as described in the thesis by Higginbotham (2017). The majority of the target protein was identified at 30°C and 37°C 6 hours post induction. See Fig. 12-14.

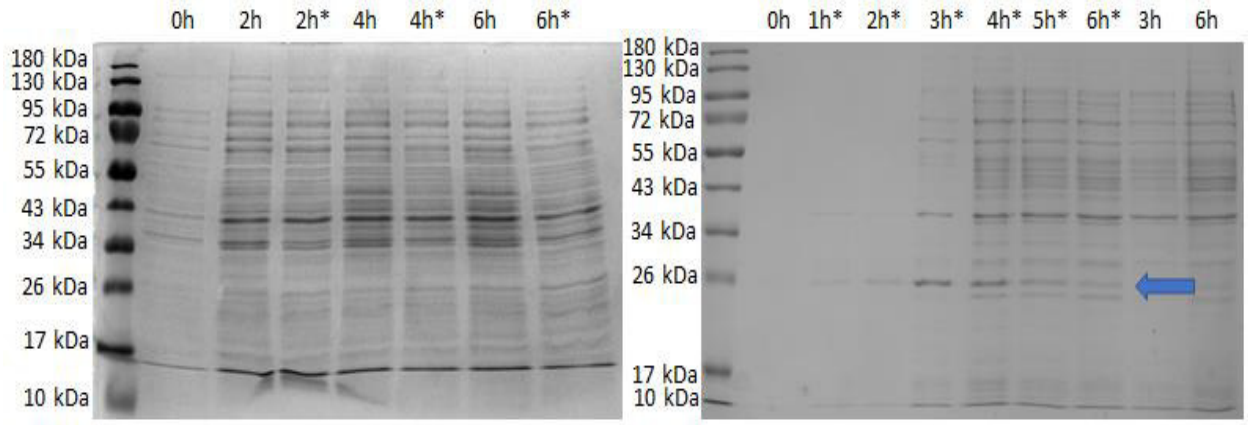

Figure 12. Pilot expression*: E. coli* BL21 DE3 cells, LB medium, 37°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

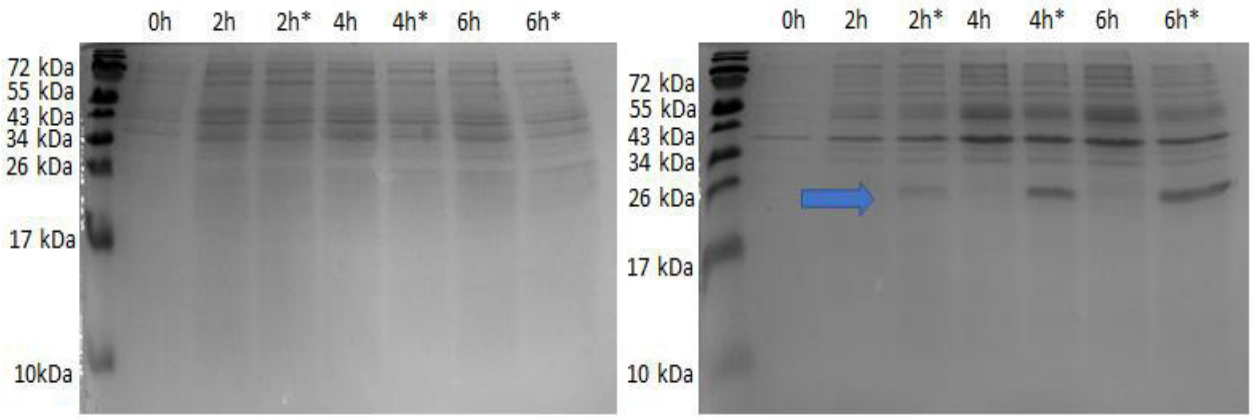

Figure 13. Pilot expression: *E. coli* BL21 DE3 cells, LB medium, 30°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

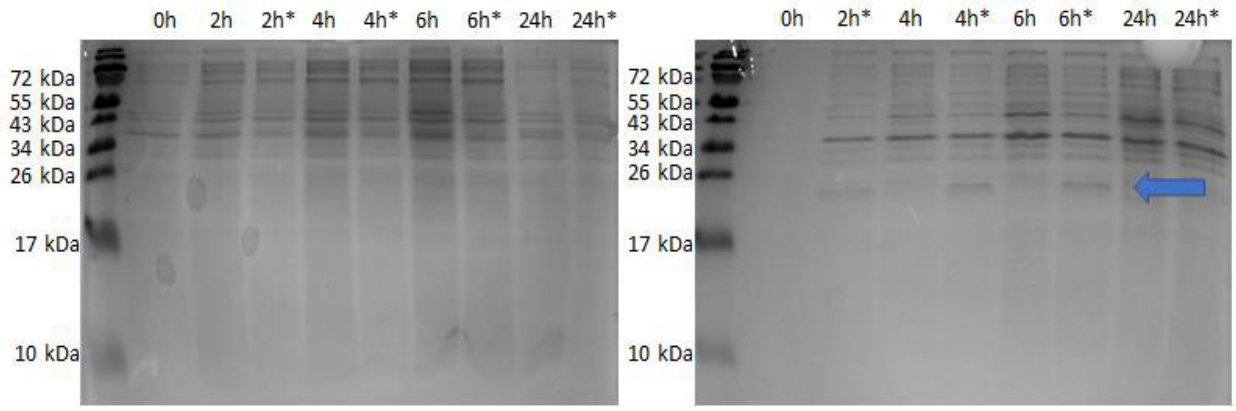

Figure 14. Pilot expression: *E. coli* BL21 DE3 cells, LB medium, 18°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

4.2.2 Using *E. coli* Rosetta-gami B DE3 cells

Because we did not get any soluble protein from the *E. coli* BL21 DE3 cell line, we decided to use a different *E. coli* expression strain, the Rosetta-gami B DE3 cells. The pilot expression was carried out using Terrific Broth medium to accelerate the growth of this rather slow cell line. The pilot expression analysis (see Fig.15-17) revealed the presence of C-terminal σNS in the soluble phase when growing the cells in TB medium at 37°C for 2 hours and induced by 1 µL anhydrotetracyclin. Therefore the large scale protein production was performed under these conditions.

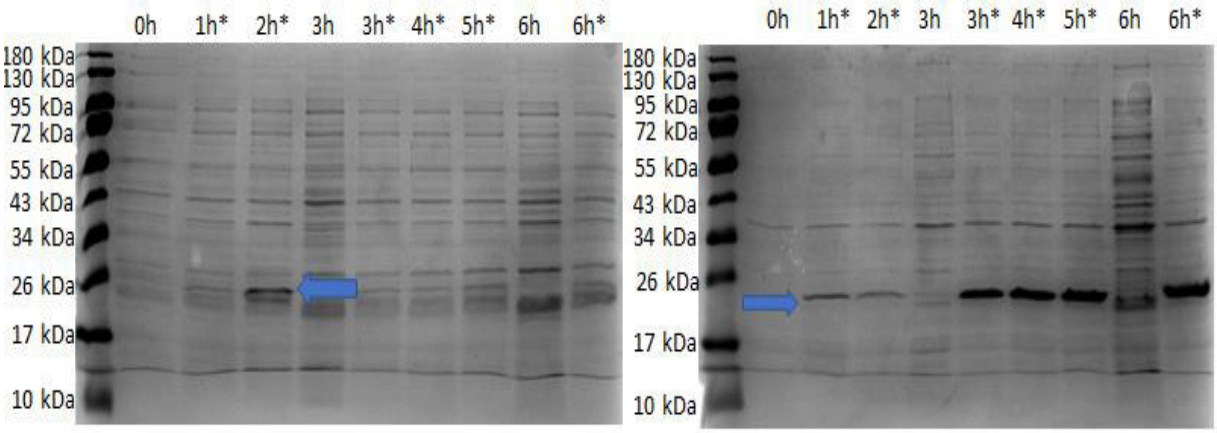

Figure 15. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 37°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

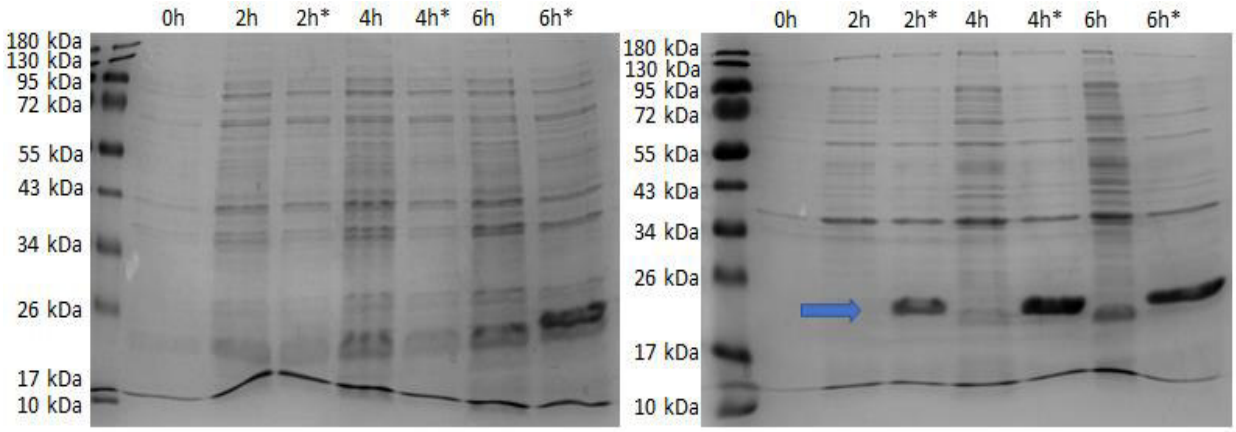

Figure 16. Pilot expression: Rosetta-gami B DE3 cells, LB medium, 37°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

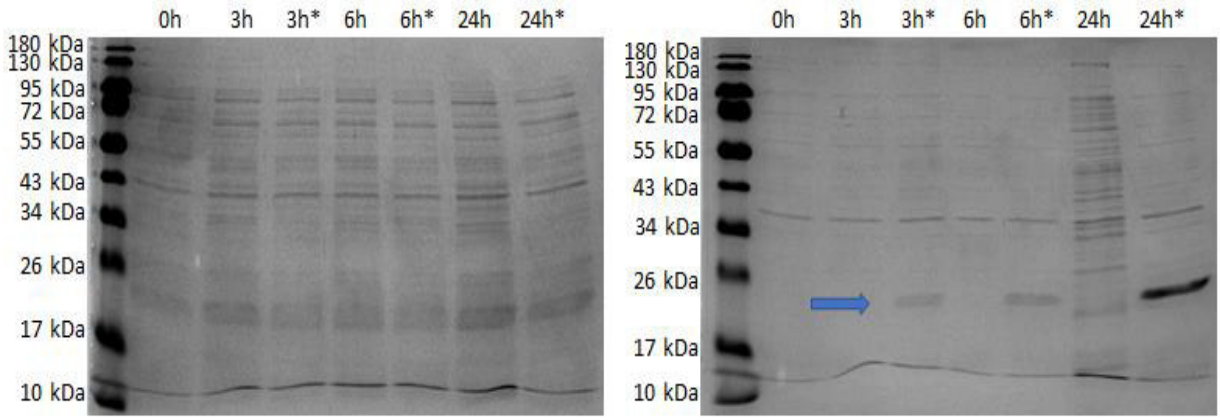

Figure 17. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 18°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin. Expected size of the C-terminal end domain: 25,012 kDa (expected range shown by arrow), 12.5% polyacrylamide gel.

During the large scale production, cells were growing very slowly. Instead of 2 hours incubation time, the cells reached the OD of 0.6 after 9 hours and only after then the induction by anhydrotetracyclin was possible. The purification of the produced recombinant protein found in the soluble phase was performed using the ÄKTA pure chromatography system. However, it was not successful (data not shown). To investigate the reason for the unsuccesful purification, aliquots were taken and analyzed by SDS-PAGE and were compared to the samples of the successful pilot expression. This analysis revealed, that the majority of recombinant protein was transferred from the soluble phase to the insoluble one. Moreover, the Western Blot analysis was done using using antibodies against the His-tags. The results clearly showed several protein bands on the membrane, indicating that the recombinant protein was degraded from the C-terminal part during the production. See Fig. 19.

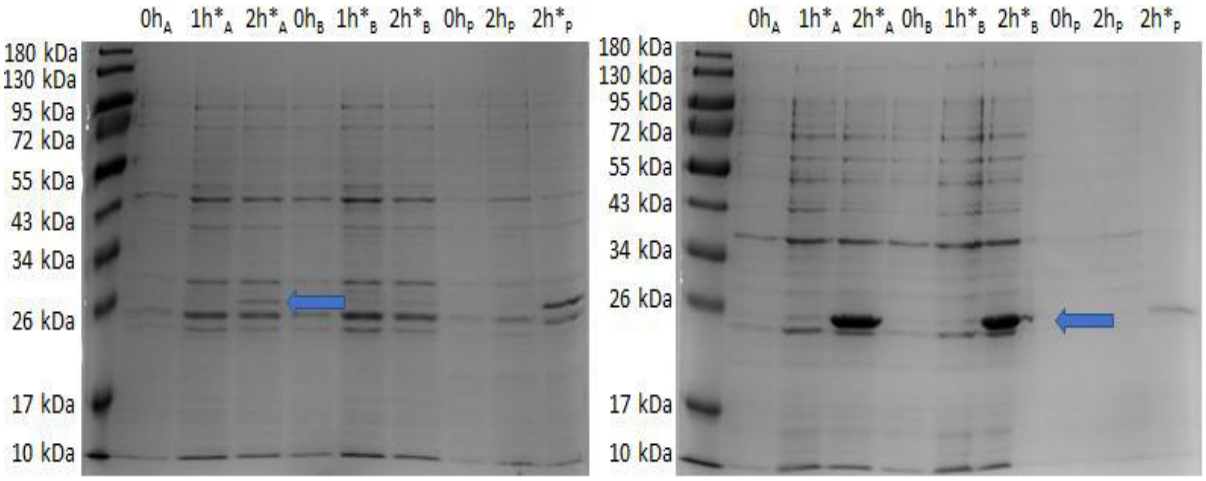

Figure 18. Large scale protein production (Rosetta-gami B DE3 cells, TB medium, 37°C, 220 rpm). Left: soluble phase, right: insoluble phase. *: induction with anhydrotetracyclin, A: ultra-yield flask number one, B: ultra-yield flask number two, P: pilot expression using the same conditions. The blue arrows show the position of the recombinant protein. 12.5% polyacrylamide gel.

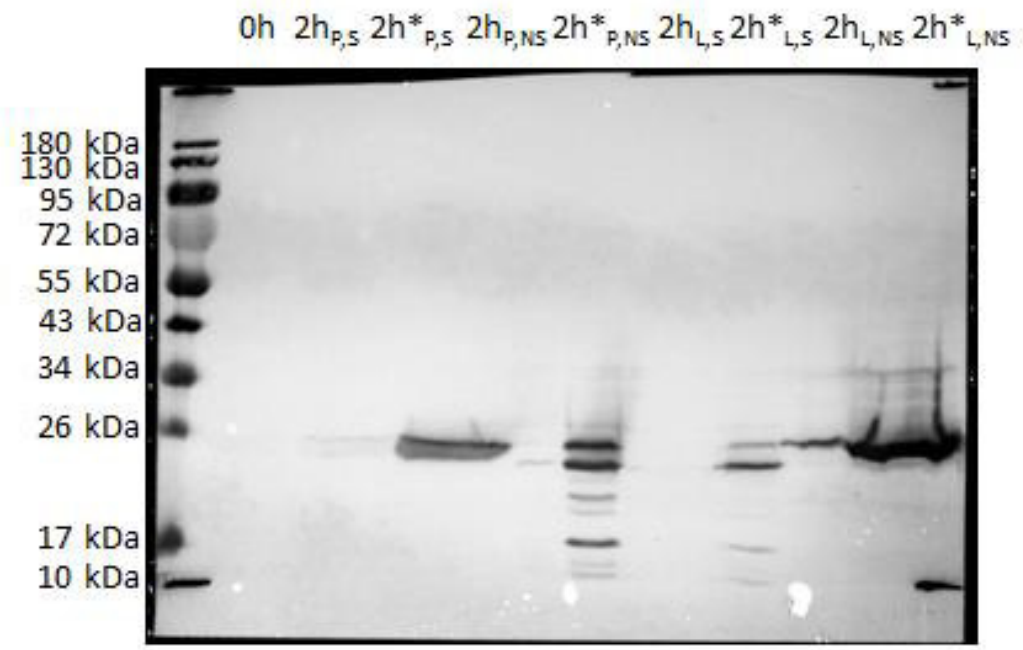

Figure 19. Western Blot of the fractions of the large scale protein production. *: induction with anhydrotetracyclin, A: outcome of large scale production (Rosetta-gami B DE3 cells, TB medium, 37°C), P: outcome of the pilot expression (Rosetta-gami B DE3 cells, TB medium, 37°C) using the same conditions. S: soluble fraction, NS: non-soluble fraction.

4.2.3 Troubleshooting

To find out, how to handle cells for further purification, new pilot expressions were performed using the above identified conditions (37°C, TB medium) with cells from various sources: a: glycerol stock, b: re-plated Rosetta-gami B DE3 cells. Besides the question, whether the C-

terminal domain is always degraded, another question had to be cleared, whether the age of the cells could have played a role in the problems during the large scale production. By measuring the OD of the overnight cultures before inoculating into the new medium, it was revealed, that old Rosetta-gami B DE3 cells are not viable when they are either plated (OD=0.182) or stored in glycerol stock (OD=0.010) for longer time. After growing all night they should have had an OD around 10. Rosetta-gami B DE3 cells were freshly transformed and the recombinant protein was produced under the same conditions as written in section 3.4.3. Unfortunately, the soluble form of recombinant protein was not found in any of tested conditions as analyzed via SDS-PAGE (Fig. 20-21).

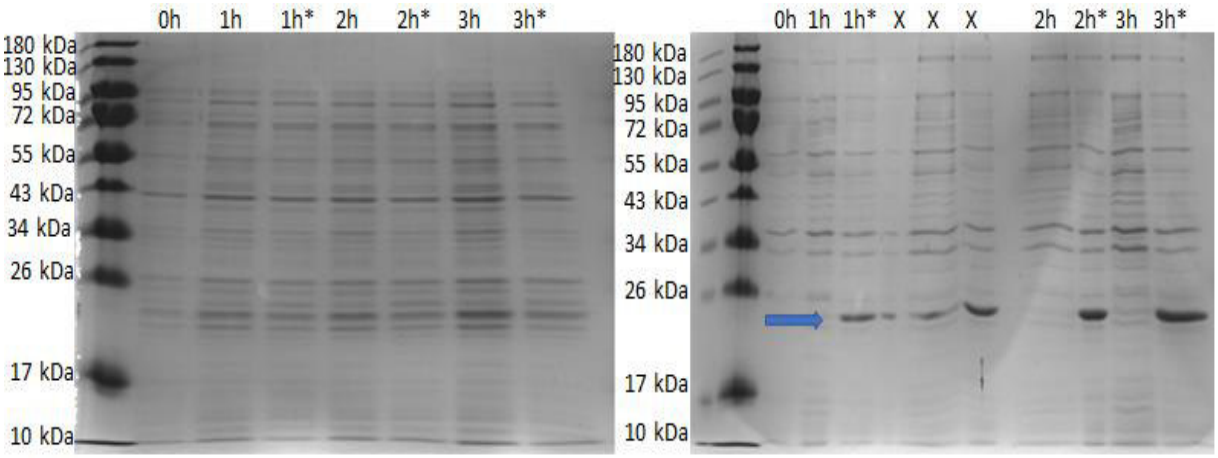

Figure 20. Pilot expression: freshly transformed Rosetta-gami B DE3 cells, TB medium, 37°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin, arrow: band position of the recombinant protein. Expected size of the C-terminal end domain: 25,012 kDa, 12.5% polyacrylamide gel. X: loading error, fraction diffused into neighbouring wells.

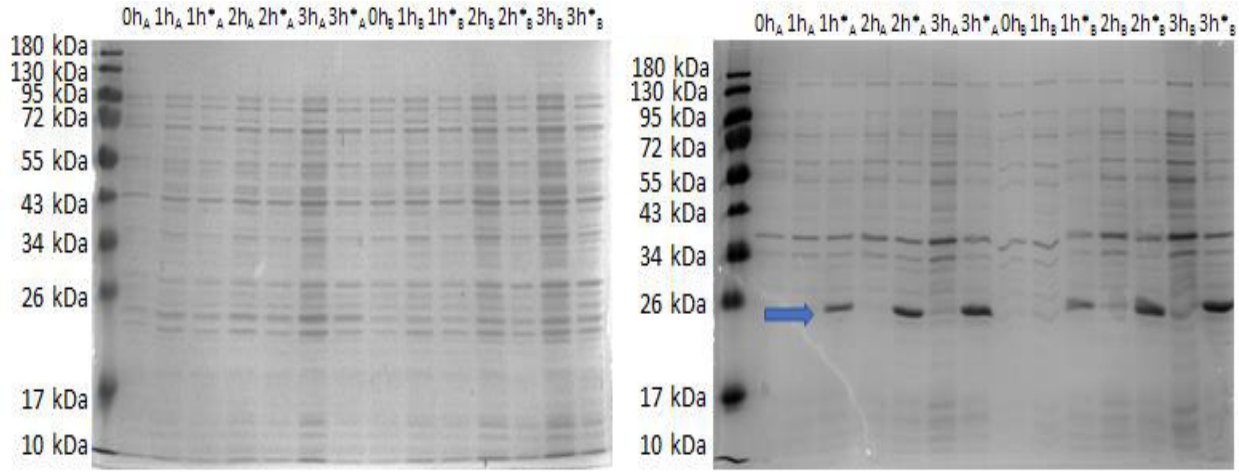

Figure 21. Pilot expression: Rosetta-gami B DE3 cells, TB medium, 37°C, 220 rpm. Left: soluble phase, right: insoluble phase. *: induced with anhydrotetracyclin, A: **glycerol stock cells**, B: **re-plated cells**, arrow: band position of the recombinant protein. Expected size of the C-terminal end domain: 25,012 kDa, 12.5% polyacrylamide gel.

Western Blot was performed using the fractions of the glycerol stock and frehsly plated cells of the pilot expression. It showed the same results like in the case of the first Western Blot, namely degradation of the recombinant protein. It proved, that the recombinant protein is degraded not even in the soluble phase, but also in the insoluble phase, indicating, that the protein is strongly unstable. See Fig. 22.

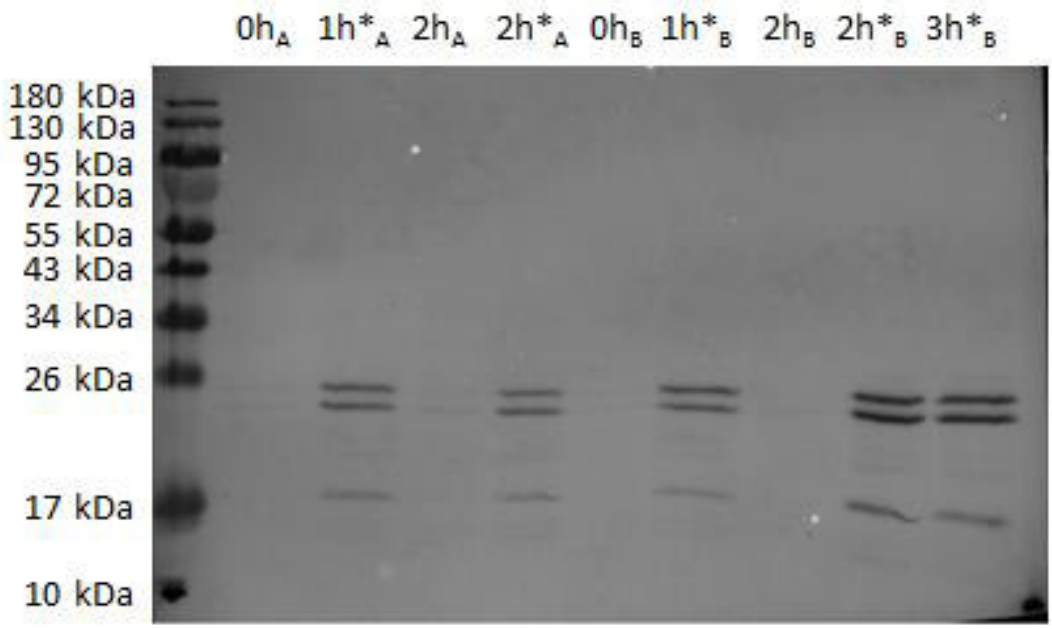

Figure 22. Western Blot of the fractions of the pilot expression (Rosetta-gami B DE3 cells, TB medium, 37°C), insoluble phase. A: glycerol stock cells, B: re-plated cells, *: induction with anhydrotetracyclin.

5. Discussion

In contrast to the thesis by Higginbotham (2017), where the DNA sequence encoding the C-terminal domain of σNS was cloned into pET22 (Novagen), containing of an IPTG inducible T7 promoter, we decided to use the vector pASK-IBA37+ (IBA Life Sciences). This vector allows us to produce a recombinant protein with N-terminal 6X Histidine-tag, which can be removed from the recombinant protein by Factor Xa after purification. Moreover, this vector has an anhydrotetracyclin promotor, enabling induction of the cells by adding anhydrotetracyclin and no leaking expression.

Here we decided to use *E. coli* expression system, because it is cheap, fast, easy to handle and yields the large amount of protein. Sequence verified recombinant plasmid was transformed into *E. coli* (BL21 DE3) cells. These bacterial host cells lack the Lon protease and the outer membrane protease OmpT, which enables the production of foreign protein at a higher expression rate without being degraded.

Our results using *E. coli* (BL21 DE3) cells clearly show, that the σNS C-terminal domain was produced in the insoluble phase, just as described previously by Higginbotham (2017). Interestingly, it has ben shown recently (Touris-Otero et al., 2015), that the C-terminally truncated protein of σNS (produced in baculovirus expression system) was also not detected in the soluble phase. The authors claimed, that the C-terminal domain of σNS might play an important role in solubility of the whole σNS protein. Our results present in this thesis do not support this theory, since we did not detect the C-terminal domain of the σNS protein in the soluble phase. However, it needs to be elucidated, whether this is due to the usage of different expression system, or really the protein characteristic.

In contrast to the thesis by Higginbotham (2017), where different cell lysis conditions were tested to acquire the recombinant protein in the soluble phase, we decided to modify the protein production conditions as well as to use a different *E. coli* expression strain, the Rosetta-gami B DE3 cells. These cells possess 2 important features over *E. coli* BL21 DE3 cells. Firstly, they possess mutations in both the glutathione reductase (*gor*) and thioredoxin reductase (*trxB*) genes. Both mutation greatly enhance the disulfide bond formation in the cytoplasm, which can stabilize produced proteins. Secondly, the cells carry an extra plasmid supplying the tRNAs for *E. coli* rare codons (AGG, AGA, AUA, CUA, CCC, GGA). Thus

these cells can enhance the production of recombinant protein in the soluble phase. To accelerate the growth of this rather slow cell line, Terrific Broth (TB) medium was used. This medium contains additional nutrients like various aminoacids, thus accelerating the growth of the bacterial cells.

As can be seen from our results (Fig.15), we were able to obtain the soluble form of the σNS C-terminal domain using Rosetta-gami DE3 cells, 37°C, TB medium, and 2 hours induction with anhydrotetracycline. However, during the large scale production, the cells grew extraordinarly slow and consequent SDS-PAGE analysis revealed that recombinant protein was only present in the insoluble phase. A possible reason for such slow growth remains to be elucidated, but we speculate that this could be caused by the fact that we used cells, which grew on LB plates for several days and were not freshly transformed. In addition, such a slow growth might have also influenced the bacterial cells to pack the produced protein in the insoluble phase.

Next we tried to produce soluble recombinant protein using the same conditions as stated above, but with different „cell age“. Unfortunately, non of these experimets led to a positive result (Fig.20-21). Therefore it is more likely, that the C-terminal domain of σNS is quite unstable and tends to degrade.

The Western Blot analysis using anti-His antibody confirmed our assumption. As shown in Fig.22, we did not observe a single band as expected, but many various size bands indicating the degradation of σNS C-terminal domain from its C-terminus. Since the recombinant protein is degraded at every conditions (soluble and insoluble phase), it is more likely, that the C-terminal domain is not stable, rather than it can be caused by proteolytic enzymes.

Recently, new crosslinking experiments were performed and showed that the C-terminal and N-terminal domain of σNS strongly interact with each other (R. Tuma, manuscript in preparation). This experiment shows, that the RNA binding sites might be formed across multiple monomers, because the RNA binding peptides on the C-terminal domain can be found in the proximity of the N-terminal domain. This finding supports our results and indicates that for successful recombinant production both domains have to be present.

6. Conclusion

Avian reovirus is an important pathogen, since it affects worldwide poultry, resulting in huge economical loss for the food industry. With the ongrowing need for poultry meat, parts because of its beneficial effects over red meat, parts because of the trend to consume more meat all over the world, it would be necessary to find an appropiate medication or vaccine to stop this disease. The non-structural protein σNS plays a key role in the avian reovirus´ replication cycle. By determining its structure with X-ray crystallization it would be possible to detect its exact role and thus attempts to stop the virus replication could be done. Various attempts to produce this protein were made, but there are flexible regions, which unable the protein´s structural characterization.

Here we worked with the C-terminal domain of σNS. The gene fragment encoding the C-terminal domain (AA 152 - AA 367) was cloned into pASK-IBA37+ and two various *E. coli* cell lines BL21 DE3 cells and Rosetta-gami B DE3 were used to produce the recombinant protein. Although we were able to obtain the soluble protein in one of the tested conditions during pilot expression, we did no succeed to repeat it in any of the large scale experiments. Moreover, our data shows, that the recombinant protein is degraded under all tested conditions.

Taken together, our results imply, that the *E. coli* expression systems are not the most suitable for the production of σNS C-terminal domain.

7. Literature

1. Shors, T. (2009). *Understanding viruses*. Burlington (Mass.): Jones & Bartlett Learning. p. 11-33

2. Baron, S. (1996). *Medical microbiology*. Galveston, TX: University of Texas Medical Branch at Galveston. Hans R. Gelderbom - Chapter 41: Structure and Classification of Viruses, Chapter 63: Rotaviruses, Reoviruses, Coltiviruses, and Orbiviruses

3. Borodavka, O. (2013). *Understanding the dynamics of viral RNA genomes using single-molecule fluorescence.* Ph.D. Thesis. The University of Leeds. p.1-14;137-167

4. International Commitee on Virus Taxonomy. (2016, December 15.). ICTV Taxonomy. Retrieved September 15., 2018, from https://talk.ictvonline.org/taxonomy/w/ictv-taxonomy This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

5. Britannica, T. E. (2017, July 14). Reovirus. Retrieved December 20, 2018, from https://www.britannica.com/science/reovirus

6. Bravo J. (2016). *Mechanism of genome assortment in Avian Reovirus* (Project Reference: 1827288). University of Leeds. p.1-19;49-51;66

7. Benavente, J., & Martínez-Costas, J. (2007). Avian reovirus: Structure and biology. *Virus Research, 123*(2), 105-119. doi:10.1016/j.virusres.2006.09.005

8. GLOBAL POULTRY TRENDS – European Chicken Meat Consumption. (n.d.). Retrieved December 23, 2018, from http://www.thepoultrysite.com/articles/1793/global-poultry-trends-european-chicken-meat-consumption/

9. Marangoni, F., Corsello, G., Cricelli, C., Ferrara, N., Ghiselli, A., Lucchin, L., & Poli, A. (2015). Role of poultry meat in a balanced diet aimed at maintaining health and wellbeing: an Italian consensus document. *Food & nutrition research*, *59*, 27606. doi:10.3402/fnr.v59.27606

10. Sahin, E., Egger, M. E., Mcmasters, K. M., & Zhou, H. S. (2013). Development of Oncolytic Reovirus for Cancer Therapy. *Journal of Cancer Therapy, 04*(06), 1100-1115. doi:10.4236/jct.2013.46127

11. Dandár, E., Bálint, Á, Kecskeméti, S., Szentpáli-Gavallér, K., Kisfali, P., Melegh, B., . . . Bányai, K. (2013). Detection and characterization of a divergent avian reovirus strain from a broiler chicken with central nervous system disease. *Archives of Virology, 158*(12), 2583-2588. doi:10.1007/s00705-013-1739-y

12. Miller, C. L., Broering, T. J., Parker, J. S., Arnold, M. M., & Nibert, M. L. (2003). Reovirus sigmaNS Protein Localizes to Inclusions through an Association Requiring the sigmaNS Amino Terminus. *Journal of Virology, 77*(8), 4566-4576. doi:10.1128/jvi.77.8.4566-4576.2003

13. Brandariz-Nuñez, A., Otero-Romero, I., Benavente, J., & Martinez-Costas, J. M. (2011). IC-tagged proteins are able to interact with each other and perform complex reactions when integrated into muNS-derived inclusions. *Journal of Biotechnology, 155*(3), 284-286. doi:10.1016/j.jbiotec.2011.07.010

14. Brandariz-Nunez, A., Menaya-Vargas, R., Benavente, J., & Martinez-Costas, J. (2010). Avian Reovirus muNS Protein Forms Homo-Oligomeric Inclusions in a Microtubule-Independent Fashion, Which Involves Specific Regions of Its C-Terminal Domain. *Journal of Virology, 84*(9), 4289-4301. doi:10.1128/jvi.02534-09

15. Borodavka, A., Dykeman, E. C., Schrimpf, W., & Lamb, D. C. (2017). Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *ELife, 6*. doi:10.7554/elife.27453

16. Brandariz-Nuñez, A., Menaya-Vargas, R., Benavente, J., & Martinez-Costas, J. (2010). IC-Tagging and Protein Relocation to ARV muNS Inclusions: A Method to Study Protein-Protein Interactions in the Cytoplasm or Nucleus of Living Cells. *PLoS ONE, 5*(11). doi:10.1371/journal.pone.0013785

17. Tourís-Otero, F., Martín, M. C., 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*(2), 361-374. doi:10.1016/j.jmb.2004.06.026

18. Touris-Otero, F., Martínez-Costas, J., Vakharia, V. N., & Benavente, J. (2004). Avian reovirus nonstructural protein μNS forms viroplasm-like inclusions and recruits protein σNS to these structures. *Virology, 319*(1), 94-106. doi:10.1016/j.virol.2003.10.034

19. Bravo, J. P., Borodavka, A., Barth, A., Calabrese, A. N., Mojzes, P., Cockburn, J. J., . . . Tuma, R. (2018). Stability of local secondary structure determines selectivity of viral RNA chaperones. doi:10.1101/293191, p.2-4;21-23

20. Touris-Otero, F., Martinez-Costas, J., Vakharia, V., & Benavente, J. (2005). Characterization of the nucleic acid-binding activity of the avian reovirus non-structural protein sigmaNS. *Journal of General Virology, 86*(4), 1159-1169. doi:10.1099/vir.0.80491-0

21. 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 Acids Research, 43*(14), 7044-7057. doi:10.1093/nar/gkv639

22. Gillian, A. L., Schmechel, S. C., Livny, J., Schiff, L. A., & Nibert, M. L. (2000). Reovirus Protein sigma NS Binds in Multiple Copies to Single-Stranded RNA and Shares Properties with Single-Stranded DNA Binding Proteins. *Journal of Virology, 74*(13), 5939-5948. doi:10.1128/jvi.74.13.5939-5948.2000

23. Higginbotham, V. (2017). *The preparation of Avian Reovirus C-Terminal Domain σNS protein for structural studies* (Unpublished bachelor's thesis). University of Leeds. p.2-21