University of South Bohemia in Česke Budějovice
Faculty of Science

Expression and purification of Tick-borne encephalitis virus Capsid protein

Bachelor Thesis

Jakob Schmelzer

Supervisor: RNDr. Zdenek Franta, PhD

Co-Supervisor: Paulina Duhita Anindita, DVM, PhD

Česke Budejovice

2022

Bibliographical reference:

Schmelzer, J., 2022: Expression and purification of Tick-borne encephalitis virus Capsid protein. BSc. Thesis, in English – 44p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Keywords:

Flavivirus, Tick-borne encephalitis virus, Capsid protein, Cloning, Expression, IPTG, BL21-CodonPlus (DE3)-RIPL, Immobilized Metal Affinity Chromatography

Annotation:

Tick-borne encephalitis virus (TBEV) is an important pathogen of the family *Flaviviridae* with an emerging health concern. Virus causes severe neurological disorders to human population in Europa and Asia. Despite effective vaccine, there is no specific antiviral treatment till now. The capsid protein encapsules the viral positive-sense single-stranded RNA and is essential in packaging and assembling of new virions within the host. A deeper understanding of these interactions and their implications should lead to new antiviral strategies and drug design. This thesis deals with the expression and purification of the capsid protein without membrane anchor using *Escherichia coli* expression system.

Declaration:

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

České Budějovice, date

....

Jakob Schmelzer

Acknowledgments:

I would like to express my gratitude to my supervisor, RNDr. Zdeněk Franta, Ph.D. and Paulina Duhita Anindita, DVM, PhD. They carefully guided me through all the practical work and thesis writing, had always an open ear ready for my questions and helped me to improve my scientific thinking.

My thanks also belong to doc. Roman Tuma, Ph.D. for his many advices and giving me sometimes the right push to think different.

Additionally I want to thank the rest of the community from the lab for being always friendly and supportive, which made the whole work experience unforgettable.

Table of contents

List of abbreviations

°C – degree Celsius
A – Ampere
BBB – blood-brain barrier
C - Capsid protein
C-core – Capsid core protein
C-anch – Capsid protein with a membrane bound anchor
CNS – central nervous system
DENV – Dengue Virus
E/M – membrane-associated proteins; Envelope and Membrane
HCV – Hepatitis C Virus
M – Molar
mM – milli molar
MR – mortality rate
NC – Nucleocapsid
NCRs – non-coding regions
NS – non-structural proteins
ORF – open reading frame
prM – precursor Membrane
PCR - Polymerase Chain Reaction
RdRp - RNA-dependent RNA polymerase
Rpm – rounds per minute (centrifuge)
RE - Restriction Endonucleases/Enzymes
RED – Restriction Enzyme Digestion
SP – structural proteins
+ssRNA – positive-sense single-stranded RNA
TBEV – Tick-borne Encephalitis Virus
V – Volts
YFV – Yellow Fever Virus
WNV – West Nile Virus
VLP – Virus-Like particles
ZKV – Zika Virus

1 Introduction

1.1 *Flaviviridae* family

This family of viruses consists of small enveloped viruses with positive sense, single-stranded RNA genomes (9.0 – 13 kb) which mostly infect mammals and birds and many are host-specific and pathogenic. The family received its name from the Yellow Fever virus (YFV, from Latin Flavus = "yellow"). There are 89 species divided into 4 genera within this family, such as *Hepacivirus* (Hepatitis C virus), *Flavivirus* (YFV, dengue virus (DENV), Zika virus (ZIKV) and tick-borne encephalitis virus (TBEV)), *Pegivirus* and *Pestivirus* (swine fever virus). These viruses are antigenically unrelated, but serological cross-reactive within each genus. The virion has a size of 40-60 nm with a single core protein (except for *Pegivirus*) and 2 or 3 envelope glycoproteins. *Pestivirus* (from Latin pestis = "plague") and *Hepacivirus* (from Greek hepatos = "liver") have similar replication strategies to flaviviruses, but are antigenically distinct and not arthropod-borne. These lineages could represent early diverging steps in the evolution of *Flaviviridae* (Cammisa-Parks *et al*., 1992).

The genomic RNA of all members of the family contains a single long open reading frame (ORF) which is flanked by a 5'- and 3'-terminal non-coding regions (NCRs) required for genome replication and translation (Lin *et al*., 2004). Upon viral entry to the cell, viral protein translation happens directly from genomic RNA containing a type I cap (*Flavivirus*) or an internal ribosome entry site (other genera) (Simmonds *et al*.,2017).

1.2 Flavivirus

Flaviviruses are causing severe human diseases like West Nile fever, Yellow Fever, Zika Fever which usually cause fever, headache or nausea up to different stages of inflammation of the brain and for many of them there are no vaccines available. Most flaviviruses are arthropod-borne viruses, either mosquito-borne (50%), tick-borne (28%), or

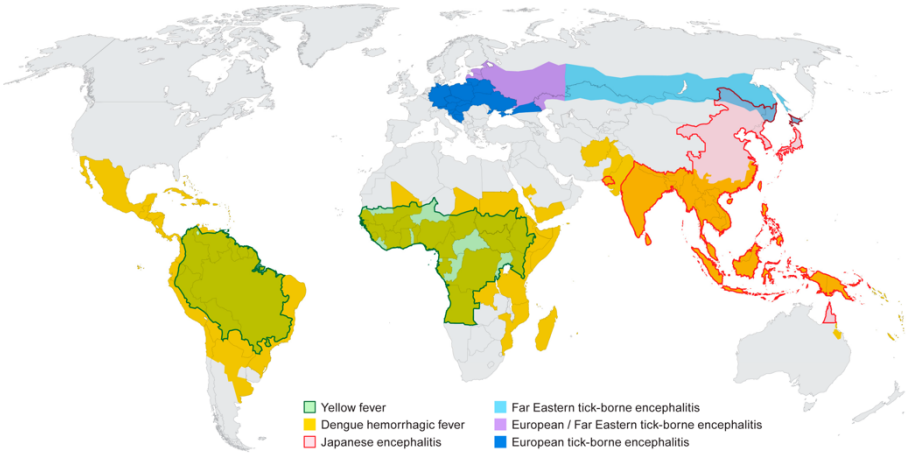


Figure 1 - Distribution of major Flaviviruses (Holbrook, 2017)

the rest transmits between rodents or bats without known arthropod vector (Simmonds *et al*., 2017). The common route of infection is via arthropod bite after its feeding on viraemic vertebrates (Blitvich and Firth, 2015).

As Humans are primarily infected via arthropod bite, they get alternatively infected by blood products, organ transplantation, non-pasteurized milk or aerosols. Individual species of flaviviruses are restricted to specific epidemic areas, e.g., dependent on specific arthropod species YFV presents in tropical or subtropical regions of Africa and South America while DENV spreads in tropical areas of Asia, Oceania, Africa, or Australia and TBEV is distributed in Europe and Northern Asia (Figure 1).

1.3 Tick-borne encephalitis virus

TBEV is a major tick-borne viral pathogen with a growing health concern. TBEV infection in human is mostly asymptomatic, however, in some cases it may cause severe neurological disorders such as meningitis, encephalitis or meningoencephalitis. Morbidity and mortality of TBEV infection vary on the viral subtype: European, Siberian and Far-Eastern subtype (Bogovic, 2015). The disease caused by the European and Siberian subtype is less severe, with a mortality rate (MR) of 0.5-2% and 2-3%, respectively. The Far-Eastern subtype causes the highest MR (up to 40%) (Bogovic, 2015).

Viral infection starts in subcutaneous tissues and is followed by Langerhans cells which transport the virus to the draining lymph nodes where it then enters the bloodstream and induces viremia. Later on, TBEV infects various peripheral organs and tissues, and crosses continuously the blood-brain barrier (BBB) without disrupting it, initiating brain infection (Ruzek *et al*., 2019). The phase of infection undergoing in the patient dictates the appearance of clinical symptoms. During the first phase when the virus has not yet crossed the BBB, the symptoms range from fever, headache, fatigue, myalgia, anorexia, nausea and/or vomiting. This phase can be abortive, which means that progression of disease may terminate and no neurological symptoms are developed (Bogovic et al., 2010). When the virus crosses the BBB, the patient develops full second phase neurological symptoms ranging from meningitis (inflammation of membranes around the brain) and encephalitis (inflammation of the brain) to the focal forms: meningoencephalitis, meningoencephalomyelitis (additional inflammation of the spinal marrow) and encephaloradiculitis (developing signs of damaged roots and peripheral nerves) (Ruzek *et al*., 2019).

The virion of TBEV is spherical in shape, has ~ 50 nm in diameter (see Figure 2) and consists of 3 structural proteins (SP): the capsid (C), envelope (E) and membrane (M) protein (see Figure 2). The virion can be distinguished into two main forms. The immature virion with its membrane-associated precursor (prM) proteins which are proteolytically cleaved during maturation into the membrane (M) protein giving rise to the mature virion (Stadler *et al.*, 1997), which is infectious for the cell. Multiple copies of the capsid (C) protein enclose a single copy of the genome and form the nucleocapsid (NC). The NC is surrounded by a lipid membrane, where E and M proteins are embedded to coat the lipid bilayer and form heterodimers (Füzik *et al.*, 2018). Since all flaviviruses are serologically-related, E protein is the major target for neutralizing antibodies. prM and NS1 proteins can also induce protective immunity to ensure a non-lethal infection in animals. The weight of the virion consists of around 17 % lipids derived from host cell membranes and 9 % account to carbohydrates (Simmonds *et al., 2017).*

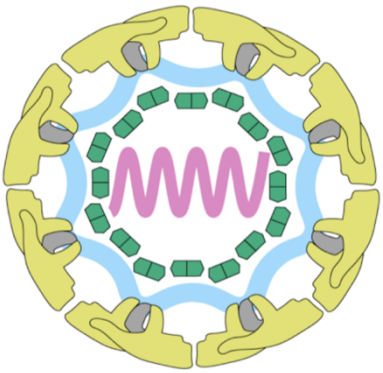


Figure 2 - Schematic representation of the TBEV virion. Viral genome (lilac), C protein (green), lipid membrane (light blue), E and M protein (yellow and grey, respectively) (Pulkkinen et al., 2018)

The genome of TBEV has a ~11 kilobase-long positive-sense single-strand RNA (+ssRNA) that encodes for the 3 already mentioned SP and additionally for seven non-structural proteins (NS): NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5 (Firth and Atkins, 2009); which are translated into a long polyprotein (see Figure 3) in infected cells. This polyprotein is then co- and post-translationally processed (see Figure 3) into the viral structural and non-structural proteins mainly by viral serine protease complex NS2B-NS3, which has both protease and helicase activities, but

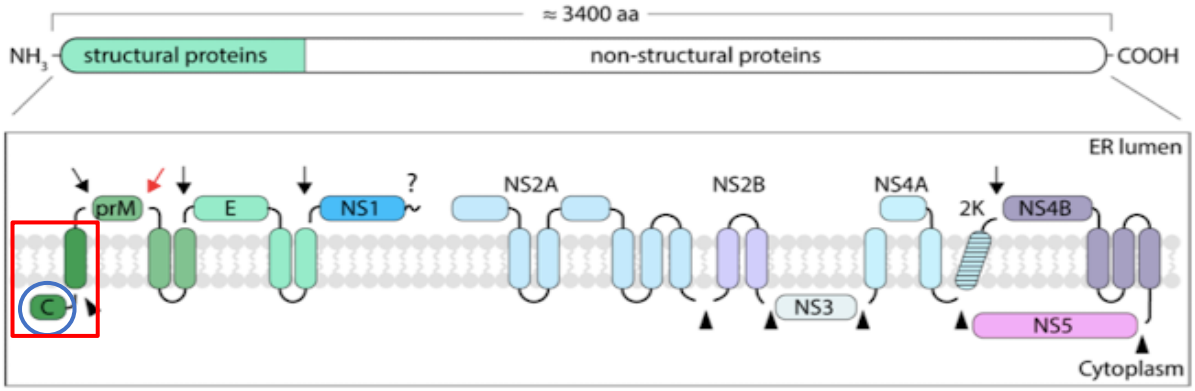


Figure 3 - Schematic representation of Flavivirus polyprotein and its cleavage products. Arrows and triangles indicate cleavage sites of viral serine protease (black arrows), furin (red arrow) and host signal peptidase (triangle), whereas the question mark indicates an unknown host protease (Pulkkinen *et al.*, 2018). The blue circle shows the main target protein (C-core) of the thesis while the red rectangle shows the additional membrane anchor.

also other proteases of host origin. Viral NS proteins play an important role in virus replication by creating specialized compartment in the membrane of host endoplasmatic reticulum (ER) called replication complex (RC). In RC the new viral RNA is synthetized and capped by NS3 and NS5 (Figure 4). Our information about the structure and assembly process of RC is rather limited but many studies are focusing on the roles of individual NS proteins. The budding process driven by the viral membrane proteins may be further stabilized by NS1 protein. It reduces the membrane attack complexes and prevents the destruction of infected cells. The structure of the integral transmembrane protein NS2A has yet to be determined, though studies suggest that NS2A is a key factor for encapsidation of the viral RNA. NS2B acts as a co-factor for the NS3 protease and is essential for its proper folding. The ATPase activity of NS3 is regulated by NS4A and has a nonenzymatic role in the assembly process. The C-terminal domain of NS3 contains RNA triphosphatase and helicase activities playing a role in new genome replication. A signal sequence, called the 2k peptide, separates NS4A from NS4B and directs the latter to the ER membrane, where it is cleaved off by the host signal peptidase. Integrated into the ER membrane NS4B performs multiple functions from replication complex formation to immunomodulation (Barnard *et al.*, 2021). The largest and most highly conserved

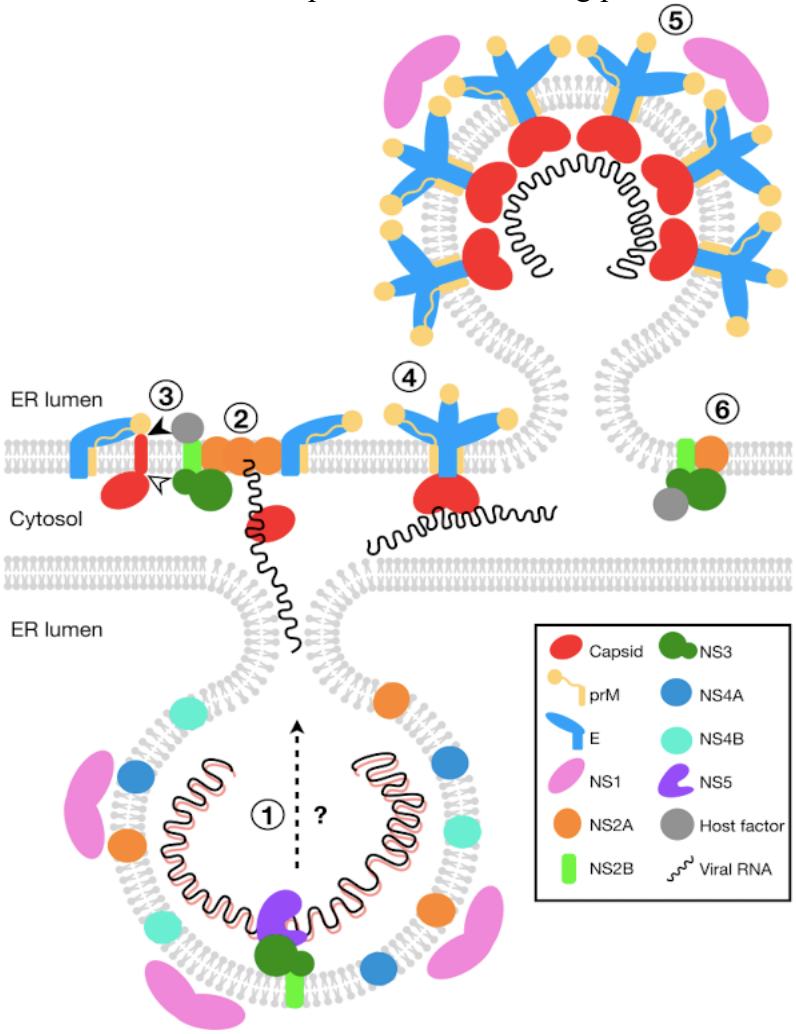


Figure 4 - Proposed Model for Flavivirus Assembly (Barnard *et al.*, 2021). (1) Newly replicated RNA may be transferred by RNA-binding proteins (potentially NS2A, NS3 or DEAD-box RNA helicas 56 (DDX56)) to the replication complex. (2) NS2A, prM-E complexes and NS2B-NS3 associate with the viral RNA. (3) C protein dimerizes through signalase and NS2B-NS3 protease (arrowheads) and associates with the viral RNA. (4) Assembly unit of the flavivirus is formed. (5) Viral membrane units drive the budding process (may be stabilized by NS1). (6) Releasing of the assembled immature virion into the ER lumen possibly by interactions with NS3 through the host endosomal complexes required for transport (ESCRT) machinery.

protein in flaviviruses is the NS5 which acts as the viral RNA-dependent RNA polymerase (RdRp) and its N-terminal Mtase domain is involved in the viral genome capping (Simmonds *et al.*, 2017).

1.4 TBEV life cycle

TBEV possess a complex life cycle (see Figure 5), which involves multiple steps. The virion enters the cell mainly via receptor-mediated endocytosis (Acosta *et al.*, 2009). Once in the cells the low pH in the late endosome triggers fusion of endosomal and virion membranes (Pulkkinen *et al.*, 2018). Membrane fusion is dependent on the correct lipid composition and cholesterol strongly enhances it (Stiasny & Heinz, 2004). The NC then disintegrates and releases the viral RNA, which serves as a template for polyprotein production. (Byk *et al.*, 2016)

After virion uncoating, the ribosomes of the rough endoplasmic reticulum (ER) start to synthesize the viral polyprotein. Subsequently, the NS proteins form a vesicle-like invagination of ER membrane (called replication complex) where viral genome replication occurs. Newly synthesized viral genome is captured and encompassed by multiple copies of the C protein. The NC complex assembles with E and M proteins and a lipid envelope to finally form a viral particle. The spiky immature particles maturate in the acidic trans-Golgi environment (Pulkkinen *et al.*, 2018). Since the maturation cannot be always completed, both immature and partially mature particles also egress from the infected cell along with the mature particle. Partially mature particles are still capable of infecting new cells while fully immature particles are non-infectious due to incapability of fusion (Guirakhoo *et al.*, 1991).

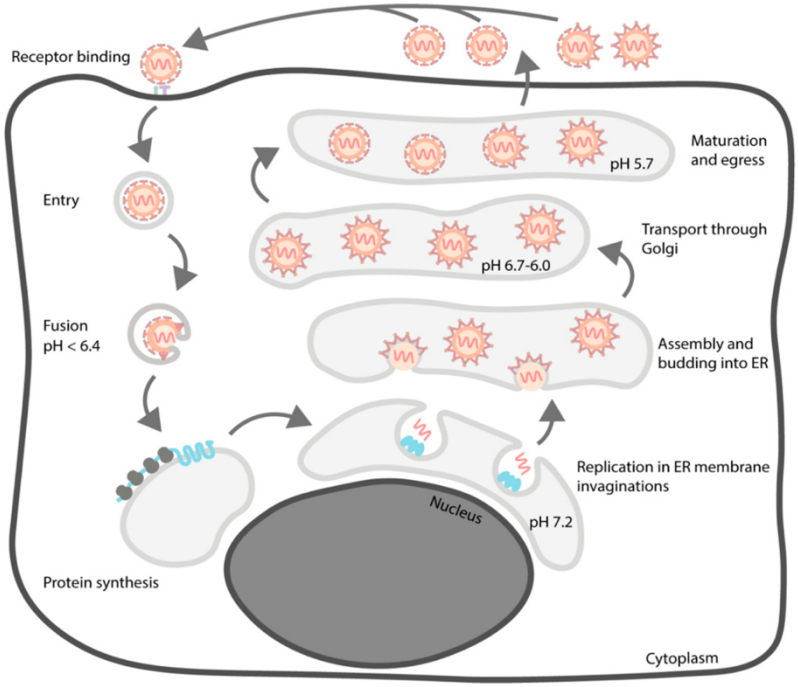


Figure 5 - Overview of the TBEV life cycle (Pulkkinen et al., 2018)

1.5 Capsid protein of tick-borne encephalitis virus

The understanding of the TBEV replication at the molecular level is crucial for development of specific anti-viral therapies. One of the critical events during the TBEV life cycle is the packaging of newly synthetized genomic RNA by C protein and formation of NC (Kaufman *et al.*,

2020). C protein is the smallest SP of flaviviruses containing approximately 100 amino acid residues with a high content of basic amino acids resulting in an extremely high positive charge (Ma *et al*., 2004). This qualifies C protein as a supercharged protein (very high net charge compared to molecular weight ratio) which might play a role in the ability of C protein to penetrate and deliver fully functional nucleic acids into cells (Freire *et al*., 2013).

C protein from other flaviviruses shares a low amino acid homology and similar 3D structure. The C protein from DENV, Zika virus (ZIKV) and West Nile virus (WNV) form antiparallel homodimers with an asymmetric charge distribution. Each monomer consists of four α-helices which process the dimerization and might mediate the interactions of C with biological membranes and lipid droplets. The crystallographic structure of ZIKV and WNV C protein suggests that the oligomeric arrangement of C incorporates the +ssRNA not only by electrostatic interactions but also through "tunnel" formed by the oligomers (Kaufman *et al*., 2020). There are no structural data for TBEV C protein available yet, but C protein of three other flaviviruses, DENV, ZIKV and Kunjin virus (KUNV; a variant of WNV), has been solved and they share the same fold despite low sequence identity. Therefore, a TBEV C protein model was generated using homology modeling in 1-TASSER server and ZIKV C protein, which predicted a similar overall fold (Pulkkinen *et al*., 2018).

C protein of TBEV consists of 96 amino acid residues and is organized most likely into four α-helices (α_1- α_4), where dimerization occurs between the helices of the two subunits α_2 and α_4 to form antiparallel dimers (see Figure 6). The helices α_1- α_3 of each monomer form a bundle with a hydrophobic surface which is believed to interact with host membranes. The two α_4 helices of the dimer from DENV, WNV and ZIKV form a surface that is rich in basic amino acids, which is most likely the RNA-binding domain of the dimer and it is believed that RNA-C binding occurs via non-specific electrostatic interactions (Pulkkinen *et al*., 2018). Flaviviral C protein shows a remarkable functional flexibility, since it is permissive to a large number of deletions and yet retains its RNA packaging activity. Therefore, it is important to gain a deeper understanding of TBEV C protein which may serve as an antiviral drug target (Oliveira *et al*., 2017).

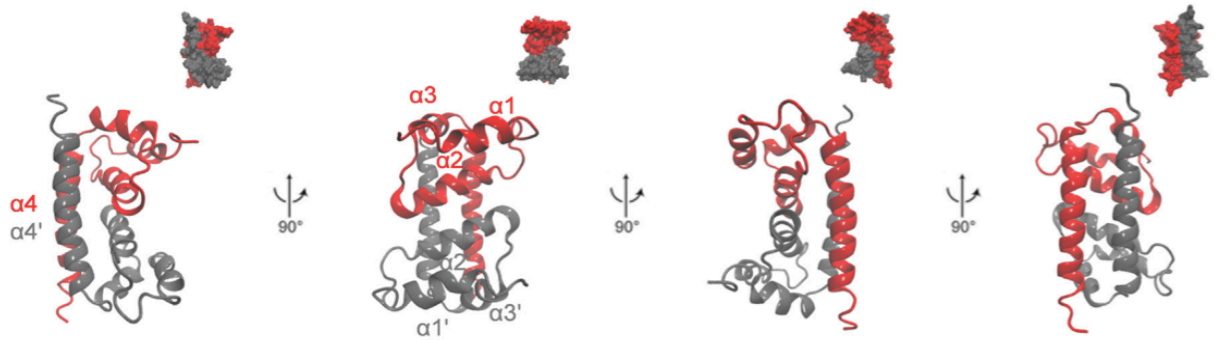


Figure 6 - Structural representation of capsid dimer showing α_1- α_4 in each monomer

2 Material and Methods

2.1 Generation of Capsid protein expression plasmid

2.1.1 Material

DNA fragments encoding either membrane-anchored Capsid (C-anch) or core Capsid (C-core) from TBEV strain HYPR were amplified from pUC57-HYPR-FRAGII plasmid (kindly provided by Dr. Daniel Ruzek). These DNA fragments were individually inserted into pET-19b and pASK-IBA37plus (previously prepared, IBA lifesciences) vectors for protein production in *E. coli* expression system.

2.1.2 Design of PCR primers

Amplification of C-anch and C-core was done using gene specific primers designed for the gene insertion into the corresponding vector. All primers were synthetized either by Generi Biotech or Eurofins Genomics. The list of primers is shown in Table *2-1*.

Table 2-1: Primers used for amplification

Primer name	**Sequence (5' to 3')**	**Cloning vector**	**Restriction enzyme**
CapC_anch_p19b_F	atggta **catatg** GTC AAG AAG GCC ATC CTG	pET-19b	NdeI
CapC_core_pASK37_R2	atg gta **ctcgag** TTA CCT CCT TTT TCC GCG TTT	pET-19b	XhoI

Note: Restriction sites are marked in bold and underlined

2.1.3 Gradient PCR amplification

To determine the most suitable annealing temperature of the primers for amplification of DNA insert, a gradient PCR was done using PPP Master Mix PCR reagent (Top-Bio). Composition of the PCR reaction is shown in Table *2-2*. For exact times and temperature used, see Table 2-3. The steps of denaturation, primer annealing and elongation were repeated 30 times.

Table 2-2: Composition of Gradient PCR reactions

Reagents	**Volume**
PPP Master Mix (75mM Tris-HCl, pH 8.8, 20 mM $(NH_4)_2SO_4$, 0.01 % Tween 20, 200 µM dATP, 200 µM dCTP, 200 µM dGTP, 200 µM dTTP, 2.5 U Taq Purple DNA polymerase, stabilizers and additives)	12.5 µL
Primer Forward (10 µM)	2.5 µL
Primer Reverse (10 µM)	2.5 µL
Template DNA	1 µL
dH_2O	6.5 µL
Total amount	25 µL for each reaction

Table 2-3: used quantities for Gradient PCR Amplification

Step	**Temperature in ° Celsius**	**Time in seconds**
Initial denaturation	94	60
30 consecutive cycles of		
Denaturation	94	15
Primer Annealing	53-61	15
Elongation	72	120
Final Elongation	72	350
Storing	4	∞

After gradient PCR amplification the reactions were analysed using gel electrophoresis with a 1.0 % agarose gel (SERVA electrophoresis) in 1X Tris-acetate-EDTA buffer (TAE). As a result, the annealing temperature of 55°C was determined as the most suitable for all primer pairs. All PCR reactions were performed on *Biometra TOne Thermal Cycler* (Jena Analytik)

2.1.4 Q5 High-Fidelity PCR

Next step is to amplify the DNA inserts with determined annealing temperature of 55° Celsius. The amplification was done by a Q5 High-Fidelity DNA Polymerase kit (New England Biolabs). The reaction composition is shown in Table *2-4*. Time and temperatures used for procedure are shown in Table *2-5*.

Table 2-4: Composition of reaction mixtures with Q5 High-Fidelity DNA Polymerase

Reagents	Volume
Q5 buffer	10 µL
10 mM dNTPs	1 µL
Primer Forward (10 µM)	2.5 µL
Primer Reverse (10 µM)	2.5 µL
Template DNA (CC/CM)	1 µL
Q5 polymerase	0.5 µL
dH_2O	32.5 µL
Total amount	50 µL

Table 2-5: Protocol for PCR amplification

Step	Temperature in °C	Time in seconds
Initial denaturation	94	30
30 consecutive cycles of		
Denaturation	94	10
Primer annealing	55	15
Elongation	72	30
Final extension	72	120

The reactions were analysed using gel electrophoresis (1.0 % agarose in 1X TAE buffer). The inserted fragments were cut out of the gel with a scalpel and isolated using the NucleoSpin Plasmid purification kit (Macherey-Nagel). The concentration of the DNA was measured using Nanophotometer Pearl (Implen).

2.1.5 Restriction Enzyme Digestion

Vectors and DNA inserts were digested with respective restrictions enzymes (RE). The reaction was prepared with the composition shown in Table *2-6*. The reactions were first incubated at 37°C for 15 minutes and then the reaction was stopped by heat inactivation at 65°C for 20 minutes. A 2.5 µL of Shrimp Alkaline Phosphatase were added to digested vectors followed by another incubation at 37°C for 30 minutes of incubation for further dephosphorylation. For deactivation the mixture was incubated for 5 minutes at 65°C. Reactions were purified using the NucleoSpin gel and PCR extraction kit (Macherey-Nagel) and the concentration was measured with a Nanophotometer Pearl (Implen). The samples were directly proceeded to continue with DNA ligation.

Table 2-6: Composition of RED reaction

Component	50 µL Reaction
DNA (C-anch or C-core)	1 µg
10X CutSmart Buffer	5 µL
NdeI	1.0 µL (20 units)
XhoI	1.0 µL (20 units)
Nuclease-free water	to 50 µL

2.1.6 DNA Ligation

After enzyme digestion, the linearized vector and digested DNA inserts were ligated using T4 DNA Ligase Kit with the molar ratio of insert to vector 3:1. The reaction (Table *2-7*) was incubated at 16°C overnight, heat inactivated at 65°C for 10 minutes and used to transform the *E. coli* DH5α cells.

Table 2-7: Setup of ligation reaction

10X T4 DNA Ligase Buffer	2 µL
Linearized vector	x µL
Digested insert	x µL
T4 DNA Ligase	1 µL
ddH_2O	Up to 20 µL

2.1.7 Transformation to *E. coli* DH5α competent cells

Super Optimal broth with Catabolites repression (S.O.C.) medium (New England Biolabs) and Luria-Bertani (LB) plates containing (100 µg/ml) Ampicillin were pre-warmed in an incubator at 37°C. The competent cells were thawed on ice out of the -80°C freezer for roughly 20 minutes. To start the transformation 2 µL of the ligated DNA product was added to the competent cells and gently mixed by finger flicking. Stressing the cells has to be avoided. The mixture was incubated on ice for 30 minutes, heat-shocked in a water bath at 42°C for 35 sec and again cooled down on ice for 2 minutes. Next 250 µL of S.O.C. medium was added to each vial and the competent cells were shaken at 37°C for 60 minutes at 220 rpm in a horizontal shaker. At last, each transformation reaction was separately spread (200 µL and 50 µL, respectively) on pre-warmed plates and incubated overnight at 37°C.

2.1.8 Colony PCR

To confirm the presence of plasmids containing C-anch or C-core in competent *E. coli* DH5α cells colony PCR was done using sets of primers (see Table 2-1). Single colony was picked carefully with a sterile pipette tip one by one from the plates and mixed into 20 µL of nuclease-free water. The remnants of the culture on the tip were placed on a new plate marked in a grid with numbers for each colony. A PCR reaction was performed using the Top-Bio PPP Master Mix with 1 µL of the template (Table *2-2*). Steps for the thermocycler were as followed:

Table 2-8: Times and Temperatures of colony PCR

Step	**Temperature in °C**	**Time in seconds**
Initial denaturation	94	360
30 consecutive circles of		
Denaturation	94	15
Primer annealing	55	15
Elongation	72	30
Final extension	72	420

The presence of cloned DNA fragment inside the cells was verified with gel electrophoresis on 1 % agarose gel together with a negative (without template) and positive control (template used from the beginning for gradient PCR amplification).

2.1.9 Isolation of plasmid and sequence verification

The positive clones were placed into 4 mL liquid LB media with Ampicillin (50µg/ml) and incubated at 37°C overnight. On the following day, the plasmids were isolated using NucleoSpin Plasmid kit (Macherey-Nagel) and eluted into 30 µL of distilled water. Concentrations of the purified plasmids were measured using the Nanophotometer Pearl (Implen). The sequence of the plasmid constructs were verified by sequencing using the Mix2Seq Kit NightXpress (Eurofins Genomics).

2.2 Protein expression

2.2.1 Transformation into Competent *E. coli* cells for protein expression

After sequence confirmation, the plasmids were transformed into various competent *E. coli* expression cells (see Table 2-9). The competent cells transformation was carried out according to the manufacturer's instructions.

2.2.2 Pilot Expression

For optimization of recombinant protein expression, single colony was picked from each transformed competent cells and incubated overnight in 10 mL of LB media (with appropriate antibiotic, see Table 3-2) at 37°C in a horizontal shaker overnight. Next day, 0.5 mL of culture was transferred into 10 mL of fresh LB media (containing appropriate antibiotic) and incubated at 37°C in a horizontal shaker until OD_{600} of 0.5 - 0.8 was reached. Then all cultures except the negative controls were induced with respective inducer, see Table 3-2. After induction, all cultures were incubated at either 18 - 20°C or 30 - 37°C in a horizontal shaker at 220 rpm. At specific time point (see Table 3-2), 1 mL of induced and uninduced culture was taken. The samples were centrifuged at 14,000 rpm for 1 minute, supernatant was discarded and the pelleted cells were stored at -20°C prior further analysis

Table 2-9: Properties of used cell lines for pilot expression

Cell line	**Properties**
OverExpress™ C43(DE3) (Lucigen)	Effective in expressing toxic proteins
Lemo21(DE3) (New England Biolabs)	Tunable T7 expression strain for difficult target proteins (membrane, toxic, insoluble)
BL21-CodonPlus (DE3)-RIPL (Agilent Technologies)	Containing genes encode for rare tRNAs
BL21(DE3)plysS (Invitrogen)	Reduce basal expression of heterologous genes, which are toxic

2.3 Analyses

2.3.1 SDS-PAGE Gel Electrophoresis

To examine the recombinant protein production from the pilot expression experiment, the pelleted cells were thawed, lysed and analysed through SDS-PAGE Gel Electrophoresis (Composition of gels shown in Table 2-10). The cells were resuspended in 150 µL of Lysis buffer (Table *2-11*) and lysed by repeated freezing in liquid Nitrogen and thawing at 42°C in a heating block for 3 times. Samples were then centrifuged at 14,000 rpm at 4°C for 15 min to separate between the insoluble protein and soluble protein fractions. Following the centrifugation, 150 µL of soluble fractions (supernatants) were transferred into a fresh tube and mixed with 50 µL 4X Laemmli sample buffer (prepared by addition of 100 µL beta-mercaptoethanol to 900 µL of 4X Laemmli sample buffer). The insoluble fractions were mixed with 1X Laemmli-PBS (prepared from 4X Laemmli sample buffer stock, see Table 2-12). All samples were boiled at 95°C for 5 min before loading into 12.5 % SDS-PAGE gel.

Table 2-10: Composition of SDS-PAGE gels for 8 gels

Running gel		**Stacking gel**	
Chemicals	**12.5 %**	**Chemicals**	**4 %**
30 % Acrylamide	16.64 mL	**30 % Acrylamide**	2.72 mL
H_2O	12.56 mL	**H_2O**	10.88 mL
1.5 M Tris (pH 8.8)	10 mL	**1 M Tris (pH 6.8)**	2 mL
10 % SDS	400 µL	**10 % SDS**	160 µL
10 % APS	400 µL	**10 % APS**	160 µL
TEMED	40 µL	**TEMED**	16 µL

Table 2-11: Composition of Lysis buffer

KH_2PO_4, pH = 7.8	50 mM
NaCl	400 mM
KCl	100 mM
Glycerol	10 %
TritonX-100	0.5 %
Imidazole	10 mM

Table 2-12: Composition of 4X Laemmli SB

0.5 M Tris (pH = 6.8)	20 mL
SDS	4 g
Glycerol	20 mL
Bromophenol blue	0.1 g
H_2O	Up to 45 mL

For each gel, 10 µL of PageRuler Protein Prestained Ladder (Thermo Fischer Scientific) was added as a marker. After electrophoresis, the upper stacking gel was disposed and the resolving gel was stained in Coomassie R-250 Brilliant Blue (see Table 2-13) for about 1 hour. Thereafter, the gels were de-stained in a solution containing: 60 % H_2O, 30 % methanol and 10 % acetic acid.

Table 2-13: Composition of staining solution

Coomassie 250 Brilliant Blue	0.625 g
Methanol	112.5 mL
ddH_2O	112.5 mL
Acetic acid	25 mL

2.3.2 Western Blot

The protein expression was also analysed using Western blot. Briefly, the protein samples described above were separated using SDS-PAGE gel electrophoresis and then transferred into a PVDF membrane using a Trans-Blot Turbo Transfer System (Bio-Rad). Before the transfer, PVDF membrane must be activated by soaking in methanol for 10-15 minutes and washed in a 1X transfer buffer (see Table *2-14*). Following electrophoresis, SDS-PAGE gels were washed in the transfer buffer for 3 min. The blotting sandwich was assembled and initiated on 25 V, 2.5 A for 5 min (low molecular weight protocol) using Trans-Blot Turbo Transfer System (Bio-Rad). Subsequently, the membrane was washed three times in 1X TBS-T buffer (see Table 2-15) on a shaking platform for 10 min each, then blocked in a blocking solution (5 g skim milk powder in 100 mL TBS-T buffer) for 1 hour on a shaking platform. The membrane was washed again 3 times in TBS-T for 10 min each and incubated at room temperature for 1 hour with antibody against histidine tag (Penta His HRP Conjugate, Qiagen) diluted in the blocking solution (1:2,500). Subsequently, the membrane was washed two times in TBS-T buffer for 10 min and one time in TBS (without Tween20) for 10 min. Finally, the signal was developed by incubating the membrane in a solution containing 3 mL Peroxide substrate and 3 mL of Luminol/Enhancer substrate (Bio-Rad) for 5 min and the blot was visualised using the Syngene G:BOX Chemi XT4 gel documentation system.

Table 2-14: Composition of SDS-PAGE transfer buffer

10X SDS-PAGE transfer buffer stock solution	
Tris	58.15 g
Glycine	29.3 g
SDS	3.75 g
Stock solution was diluted to 1X and 10 % ethanol was added	

Table 2-15: Composition of TBS-T

10X TBS stock buffer solution	
Tris	60.5 g
NaCl (pH = 7.6)	87.66 g
Stock solution was diluted to 1X and 1 % Tween20 was added	

2.3.3 Large-scale protein production with BL21-CodonPlus (DE3)-RIPL at 30°C for 2 hours with IPTG (1 mM)

After verification of the most suitable cell lines for protein production by SDS-PAGE and western blot, four 45 mL overnight cultures were prepared by shaking the transformed competent cells at 37°C overnight. These cultures were transferred into 800 mL of LB media on the next day to start the protein production. Appropriate antibiotics (see Table *3-2*) were added and samples were incubated at 30°C in a horizontal shaker till OD_{600} value between 0.5 and 0.8 was reached. Cultures were then induced and incubated at 30°C and collected after 2 hours for highest protein production yield followed by centrifugation at 4°C with 4000 rpm for 15 min. The pelleted cells were reconstituted with 50 mL LB medium and centrifuged again. Finally, the supernatant was discarded and cells were stored at -80°C until used for protein purification.

2.4 Protein Purification

2.4.1 Preparation of cell lysate

Pelleted cells were thawed on ice and 7.5 mL of Buffer A (Table 2-16; supplemented with one SIGMAFAST Protease Inhibitor Cocktail tablets, EDTA-free (Sigma-Aldrich) per 100 mL of Buffer) was added to each tube for resuspension of the pellets. Cells were lysed once with 10,000 psi and two times with 15,000 psi through the Cell disruptor Stansted Press (Stansted Fluid Power). The lysate was centrifuged in an ultracentrifuge with 28,000 rpm for 1 hour at 4°C.

2.4.2 Immobilized metal affinity chromatography

At first a HisTrap HP 5 mL column (Cytiva) was adjusted to ÄKTA Pure system (GE Healthcare) and washed with double-distilled and de-gassed water with a flowrate of 5 mL/min. After 10 min of washing, the system was equilibrated with buffer A (see Table 2-16) for 20 min with a flowrate of 1 mL/min, then with buffer B at the same flowrate for another 20 min and again with buffer A (10 mL/min) for 2 min. After equilibration, the supernatant obtained from the ultracentrifugation was loaded with the sample pump to the column at a flow rate of 1 mL/min. When UV value started to increase the flow-through fraction was collected and buffer A was used for equilibration until the UV value nearly reached zero. For elution of weakly-binding proteins from the column 5% buffer B, which corresponds to final concentration of 50 mM imidazole in the buffer flow, was used until UV value drops to the baseline. Fractions were collected using fraction collector with linear increase of imidazole concentration until reaching 100 % buffer B (within 20 min) and collection volume of 1 mL. After fraction collection was done, the column was washed with 100% buffer B for 10 min with a flowrate of 2.5 mL/min. Subsequently the flow was changed to buffer A with flowrate of 5 mL for 5 min, followed by double-distilled water and finally 20 % ethanol both for 5 min with a flowrate of 5 mL/min. Collected fractions were stored at 4°C and were analysed on a SDS-PAGE gel.

Table 2-16: Composition of Buffers for IMAC

Chemicals	Concentration
Buffer A (Equilibration buffer)	
Na-HEPES pH 7.5	20 mM
NaCl	300 mM
Buffer B (Elution buffer)	
Na-HEPES pH 7.5	20 mM
NaCl	300 mM
Imidazole	1 M

3 Results

Two capsid constructs were used to determine whether they are expressed via bacterial expression systems and if the target protein production remained in the soluble fraction. The C-anch with a size of 336 bp was cloned previously (not shown in this thesis). The production of whole capsid protein has been tested in various cell lines (see Table *3-2*) but without yielding any protein (data not shown). C-core without the membrane-bound anchor has an approximate size of 288 bp and yielded in some cases (see below) a soluble protein production.

Cloning of C-core into pET-19b

For identification of the most suitable annealing temperature of cloning primers a gradient PCR with a temperature range from 53-61°C was performed. Optimal annealing temperature based on the computed analysis was set at 55°C and verified on a 1.0 % agarose gel electrophoresis (see Figure 7).

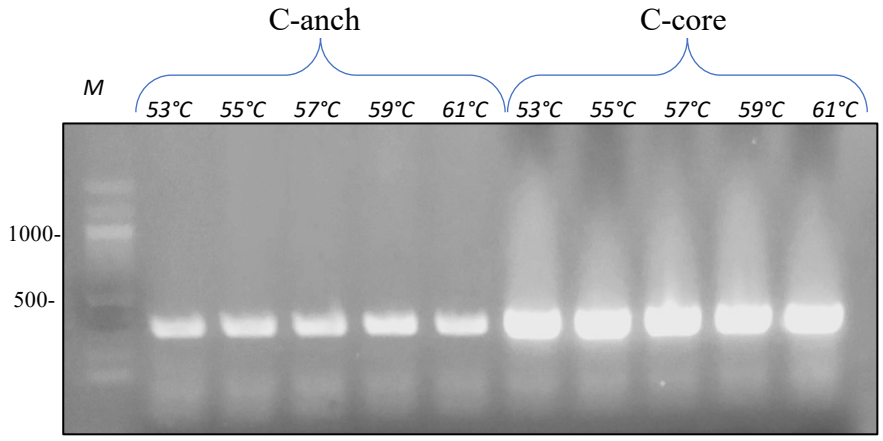


Figure 7 – Gel electrophoresis of Gradient PCR with C-anch (Anchor, 336 bp) and C-core (Core, 288bp) at different temperatures. M: 100 bp DNA Ladder (New England Biolabs)

Next, both genes were amplified using the Q5 High Fidelity DNA Polymerase. The amplification of C-core yielded a single band while the one of C-anch yielded an additional non-specific band (~ 1000 bp). This could be due to a nonspecific amplification when the primers bind to another site in DNA template. Because we did have the C-anch already cloned to pASK-IBA37plus expression vector, we decided to work on with C-core for subsequent cloning experiments. The C-core amplicon was gel-purified and treated with restriction enzymes, NdeI and XhoI, prior a ligation with NdeI-XhoI-treated pET-19b plasmid vector.

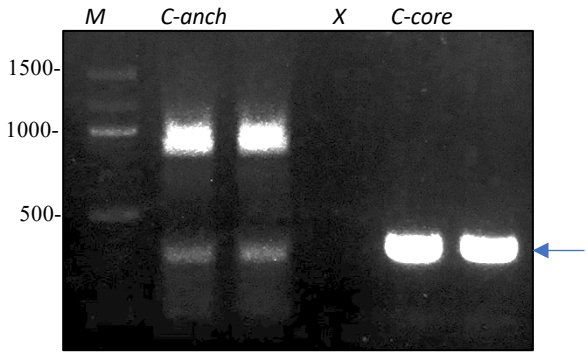


Figure 8 - Gel electrophoresis of insert amplification of C-anch (A) from Q5 insert amplification and C-core (C) from Top Bio amplification. M: 100 bp DNA Ladder (New England Biolabs). X – negative control containing no DNA template.

After ligation reaction using T4 DNA Ligase Kit (described in 3.1.6) the resulting pET-19b-C-core plasmid was transformed into *E. coli* DH5α cells. A colony PCR was done to verify the present of C-core gene in the plasmid using 28 colonies which were chosen from the LB agar plate. As a result, twelve colonies yielded clear visible band of size about 300 bp which correspond to the size of C-core. Samples 9, 18 and 22 were then inoculated into fresh liquid LB media overnight and the plasmids were extracted (see concentrations in the following Table). These plasmids were sent for sequencing to confirm the correct insertion of C-core into pET-19b plasmid. As all plasmids bear correct sequence of C-core with no frame shift thus ensuring the expression of the protein, plasmid number 18 was chosen and transformed into *E. coli* competent cells for protein production.

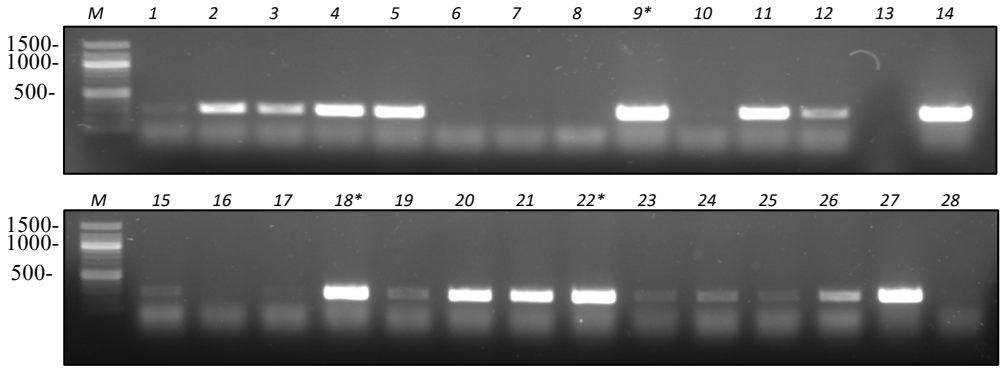


Figure 9 - Gel electrophoresis of colony PCR samples containing the pET-19b plasmid with C-core insert. M: 100 bp DNA Ladder (New England Biolabs). Samples marked with an asterisk were sent to sequencing.

Table 3-1: Concentration of extracted plasmids

Sample Nr.	Concentration in ng/µL
9	94
18	115
22	108

To find the most suitable condition for large-scale C-core production, a series of pilot expressions experiments were performed with BL21-CodonPlus (DE3)-RIPL and BL21(DE3)pLysS competent cells. Cells were grown with antibiotics Ampicillin (50µg/µL) and Chloramphenicol (35 µg/µL), 0-2% ethanol at 37°C based on established protocols by Chhetri, *et al*. (2015). When the cultures reached an OD_{600} of 0.6, they were cooled for 15 min in the fridge before getting induced with IPTG to a final concentration of 1 mM and incubated at 30°C or 18°C. The cells grew at 18°C, an aliquot of 1 mL was taken at 2, 4 and 6 h post induction and centrifuged at 4°C with 12000 rpm for 1 minute. Supernatant was removed and cell samples were stored in -20°C prior further analysis. Pelleted cells were all separated into soluble and insoluble fractions and analysed through SDS-PAGE gel electrophoresis. The protein profile was visualized by staining the gels with Coomassie R-250 Brilliant Blue. The estimated molecular weight of the C-core protein is about 13 kDa.

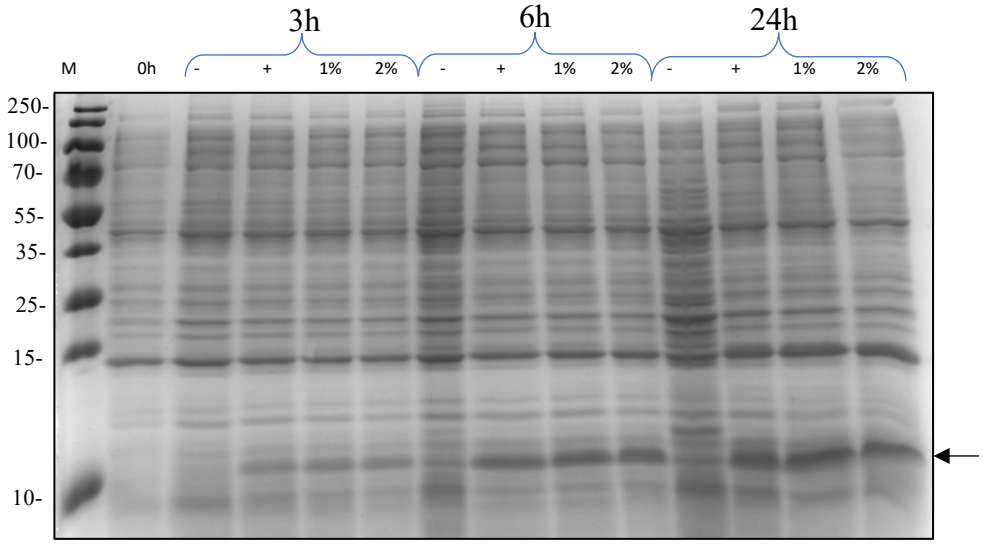


Figure 10 - SDS PAGE analysis of soluble protein expression in BL21-CodonPlus (DE3)-RIPL cells at 18° C. M: PageRuler Prestained Protein Ladder (Thermo Fischer) with corresponding size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

The protein band corresponding to the predicted size of C-core was observed from pilot expression experiment using BL21(DE3) CodonPlus competent cells at 18°C. The samples from each time point were separated into soluble and insoluble fractions. Following SDS-PAGE gel analysis of the fractions, specific signals at 13 kDa were detected mainly in the soluble fraction (compare Figure 10 and Figure 12). There was no significant difference in protein production whether ethanol was present or not. To confirm that the visible bands are the desired C-core protein, samples of the soluble fractions were analyzed using Western blot with anti-his tag monoclonal antibody (see Figure 11).

Observations from Western blot analysis showed that C-core protein was indeed produced in soluble fractions with no significant differences between cultures grown in usual LB media and LB media with addition of low concentration ethanol (see Figure 11), even though cultures grown with low concentration of ethanol should have shown an enhanced level of recombinant protein expression (Chhetri *et al*, 2015). Although the uninduced culture also showed a small protein production (shown as a thin band at 13 kD) at 24 h of incubation, this could be due to a leaky expression after long incubation time. Most significant difference is observed between 3 and 6 hours of expression but the contrast of produced protein between 6 hours and 24 hours is negligible.

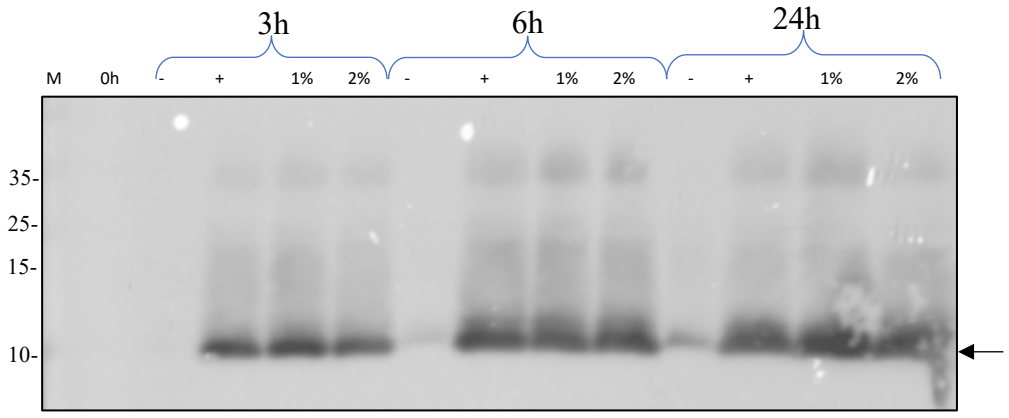


Figure 11 - Western blotting of soluble protein expression in BL21-CodonPlus (DE3)-RIPL cells at 18° C. M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

There was no recombinant protein expression observed in corresponding insoluble fractions at any specific timepoint analyzed. There are significant bands in the region of 25-27 kDa, which first gave the impression of a possible dimerization of C-core protein. However, subsequent Western blot analyses did not show any protein signal (data not shown).

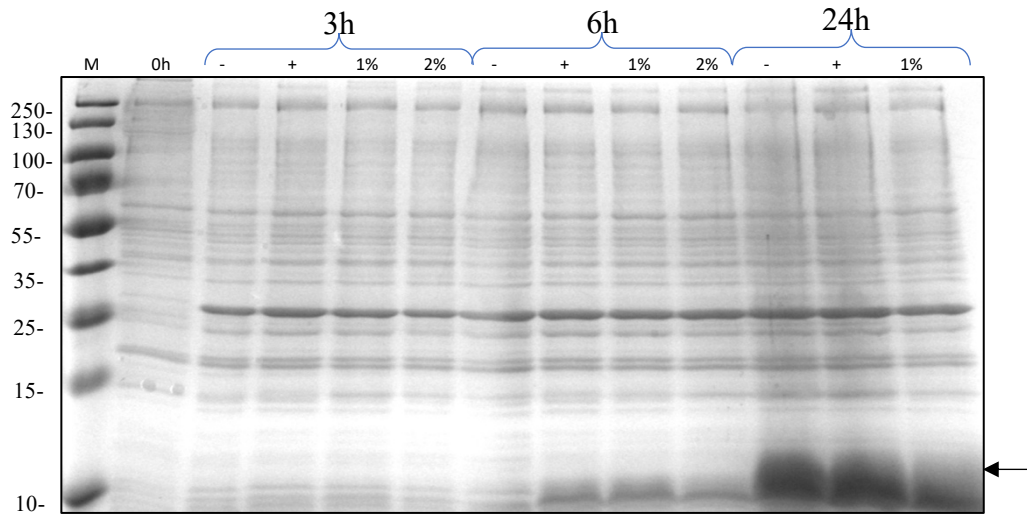


Figure 12 - SDS PAGE analysis of insoluble protein expression in BL21-CodonPlus (DE3)-RIPL cells at 18° C. M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

The expression of C-core protein in the soluble fraction of BL21-CodonPlus (DE3)-RIPL competent *E. coli* cells at 30°C was also confirmed using SDS-PAGE and Western Blot analysis. The results showed visible bands in the expected area of SDS-PAGE gel. There was a clear difference between the uninduced and induced cultures at all times and the thickness of the band does not decrease significantly at cultures grown in a 2 % ethanol solution. The insoluble fraction shows no protein expression (data not shown).

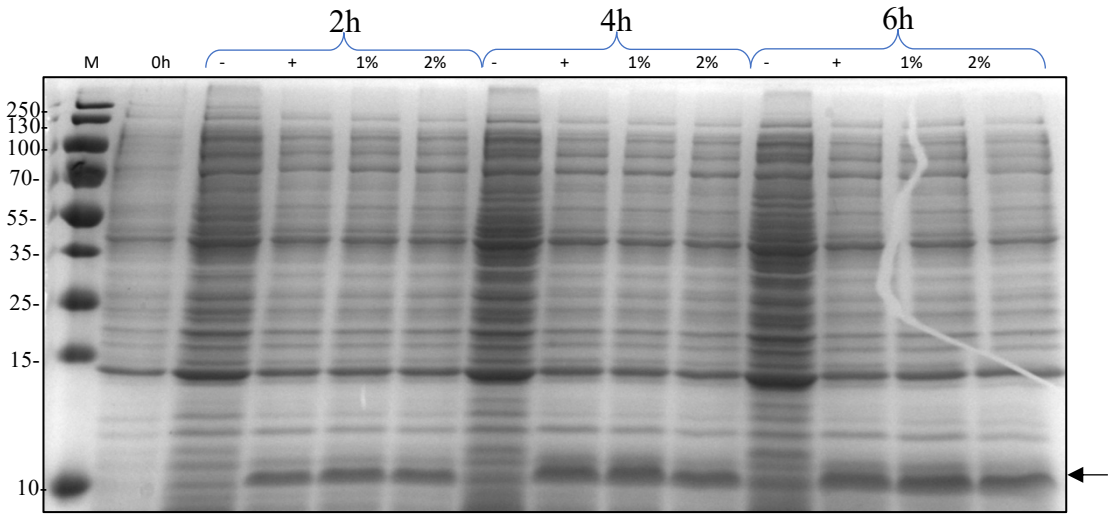


Figure 13 - SDS PAGE analysis of soluble protein expression in BL21-CodonPlus (DE3)-RIPL cells at 30° C. M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

The Western blot analysis of the soluble protein expression at 30°C (Figure 14) additionally confirms the presence of the C-core protein. It was noted that the intensity of expression decreases with longer incubation time which might be due to faster protein degradation in warm temperature. The uninduced cultures show leaky expression at all times except for time zero.

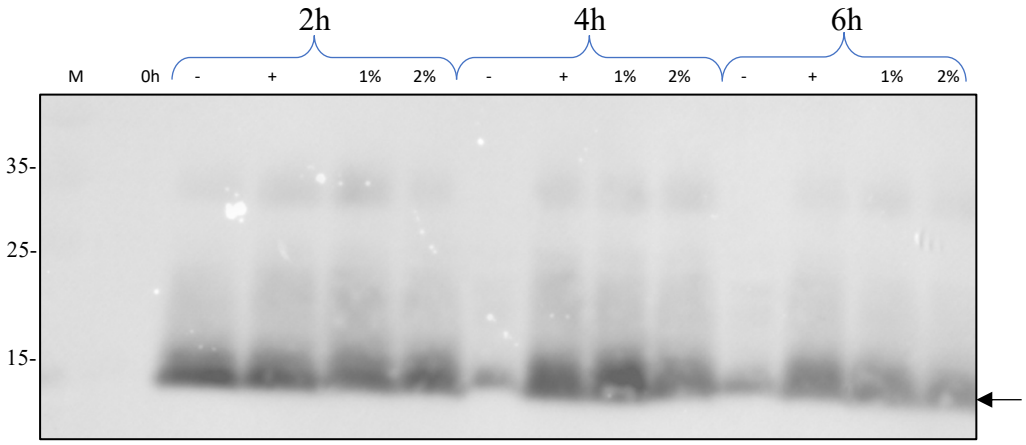
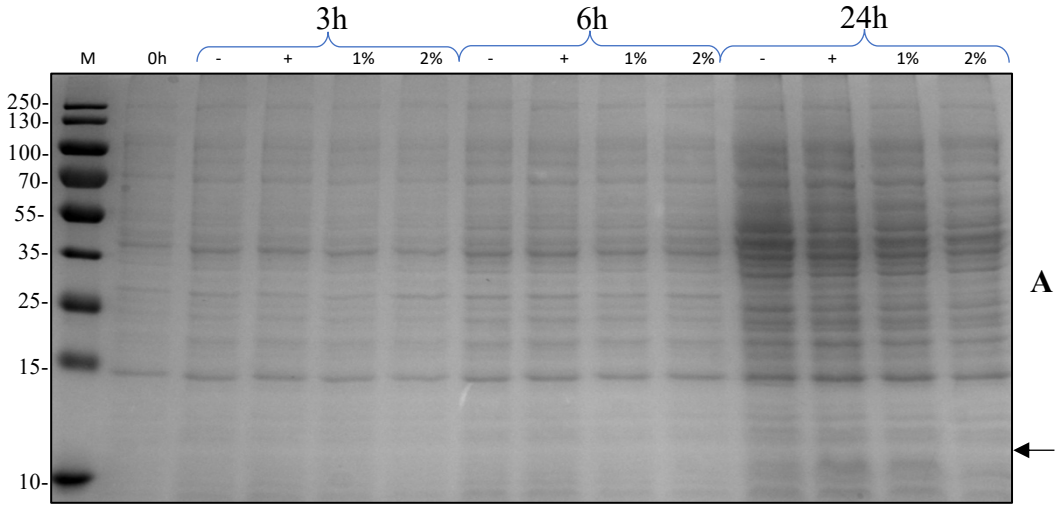


Figure 14 - Western blotting of soluble protein expression in BL21-CodonPlus (DE3)-RIPL cells at 30° C. M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

The pilot expressions using BL21(DE3)pLysS competent cells were unfortunately not successful at any tested condition. As shown in the following gels (Figure 15), no distinguishable band was observed in the expected area of SDS-PAGE gels. Western blot performed of the soluble fractions did not show any His-tag specific binding of antibodies either (data not shown).



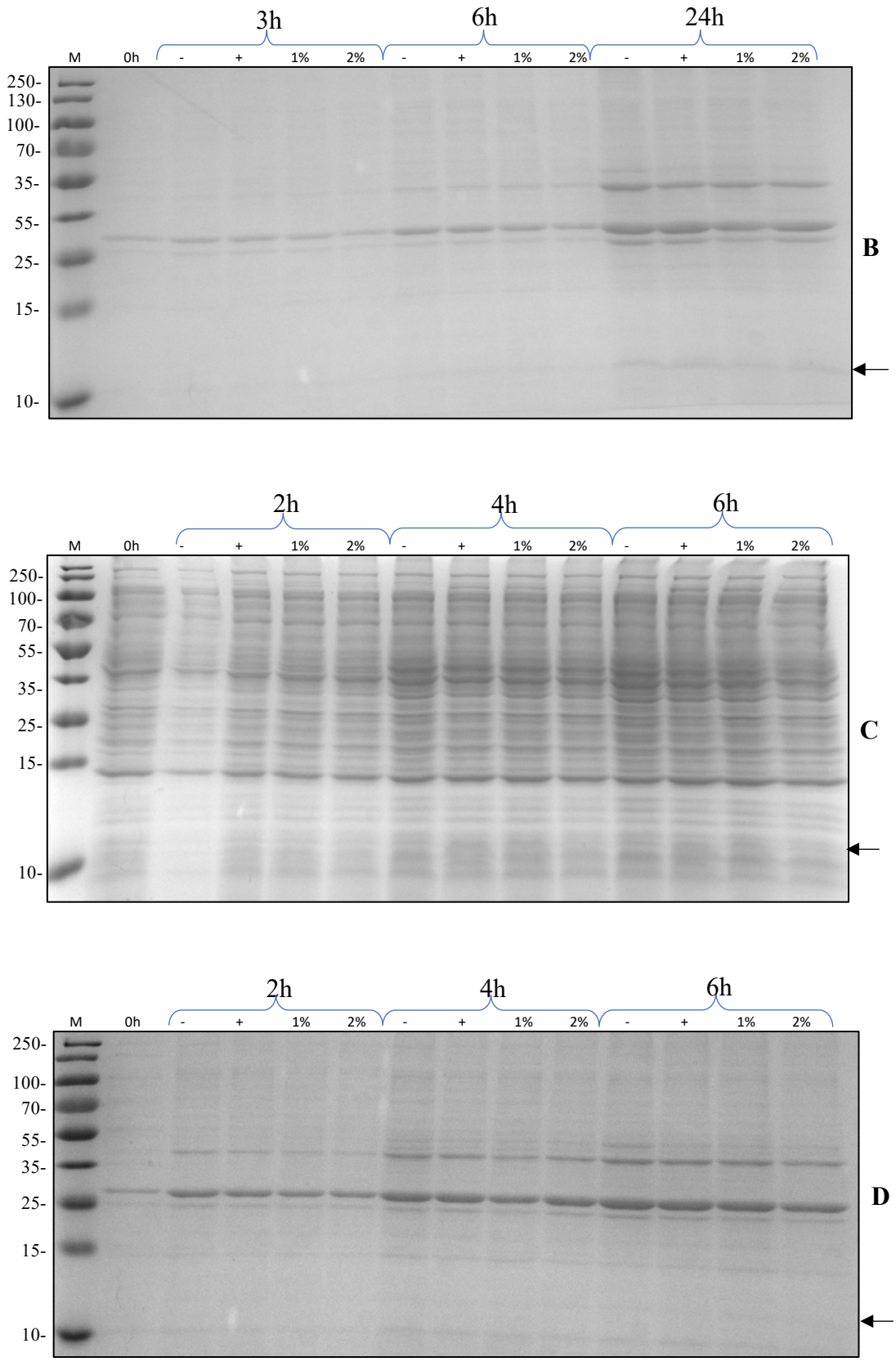


Figure 15 - SDS PAGE analysis of soluble and insoluble protein expression in BL21(DE3)plysS cells at 18°C (gel A and B) and 30° C (gel C and D). M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. – sign indicates uninduced cultures, + sign indicates induced cultures, % refers to cultures grown in 1-2 % ethanol solution. Arrow corresponds to expected bands of C-core protein.

The summary of all pilot expressions performed throughout the thesis are shown in Table 3-2with their respective properties used and if C-core protein production was successful or not (SDS-PAGE gels of successful and some unsuccessful expressions are shown above, further are not shown).

Table 3-2: Overview of all pilot expression

Competent cells	**Inducer**	**Antibiotic resistance**	**Temperature in °C**	**Time (hours)**	**Protein production**	
Lemo21(DE3) + C-anch	Anhydrotetracycline	AMP, CAM	20	2, 4, 24	soluble	-
					insoluble	-
Lemo21(DE3) + C-anch	Anhydrotetracycline	AMP, CAM	30	2, 4, 24	Soluble	-
					insoluble	-
OverExpress™ C43(DE3) + C-anch	Anhydrotetracycline	AMP, CAM	20	4, 13, 24	Soluble	-
					Insoluble	-
OverExpress™ C43(DE3) + C-anch	Anhydrotetracycline	AMP, CAM	30	2, 4, 6	Soluble	-
					Insoluble	+
BL21-CodonPlus (DE3)-RIPL + C-anch	Anhydrotetracycline	AMP, CAM	20	2, 24	Soluble	-
					Insoluble	-
BL21-CodonPlus (DE3)-RIPL + C-anch	Anhydrotetracycline	AMP, CAM	37	2, 24	Soluble	-
					Insoluble	-
BL21(DE3)plysS + C-anch	Anhydrotetracycline	AMP, CAM	20	2, 24	Soluble	-
					Insoluble	-
BL21(DE3)plysS + C-anch	Anhydrotetracycline	AMP, CAM	30	2, 24	Soluble	-
					Insoluble	-
BL21-CodonPlus (DE3)-RIPL + C-core	IPTG	AMP, CAM	18	3, 6, 24	Soluble	+
					Insoluble	-
BL21-CodonPlus (DE3)-RIPL + C-core	IPTG	AMP, CAM	30	2, 4, 6	Soluble	+
					Insoluble	-

BL21(DE3)plysS + C-core	IPTG	AMP, CAM	18	3, 6, 24	Soluble	-
					Insoluble	-
BL21(DE3)plysS + C-core	IPTG	AMP, CAM	30	2, 4, 6	Soluble	-
					Insoluble	-

Concentrations of inducer:	
Isopropyl-β-D-thiogalactopyranosid (IPTG)	1 mM
Anhydrotetracycline	200 µg per 1 liter *E. coli* culture
Concentration of antibiotics:	
Ampicillin (AMP)	50µg/µL
Chloramphenicol (CAM)	35µg/µL

Note: + indicates production of target protein

- indicates low or no production of target protein

underlined constructs refer to protein expression with addition of low ethanol concentrations (Chhetri, *et al*. 2015)

Based on the results from pilot expressions, BL21-CodonPlus(DE3)-RIPL competent cells were used to express C-core protein for 2 h at 30°C, since there were no significant differences visible from the individual batches the fastest option was used for expression. A large protein expression was performed using this condition (as described in section 2.3.3) and samples of 1 mL were analysed from the 800 mL cultures to again confirm the presence of the C-core protein by SDS-PAGE analysis (Figure 16). There is a clear difference between soluble and insoluble fractions, though no clear difference between samples taken right after inducing (0 hours) and samples taken after 2 hours, suggesting a leaky expression as previously observed during pilot expression experiment. The whole cultures were centrifuged and cell pellets were collected and stored at -80°C for further analysis.

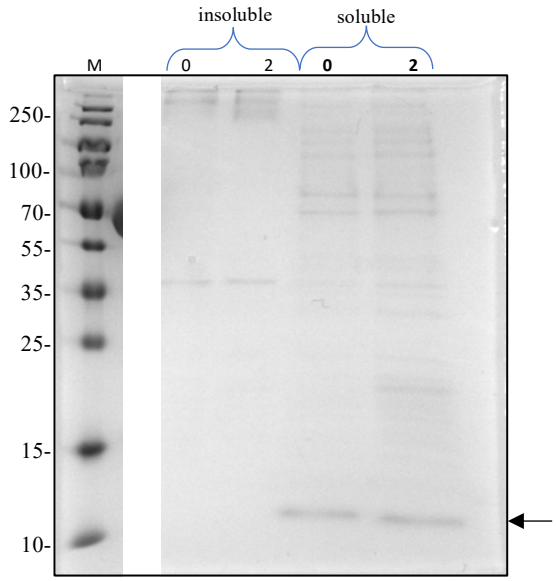
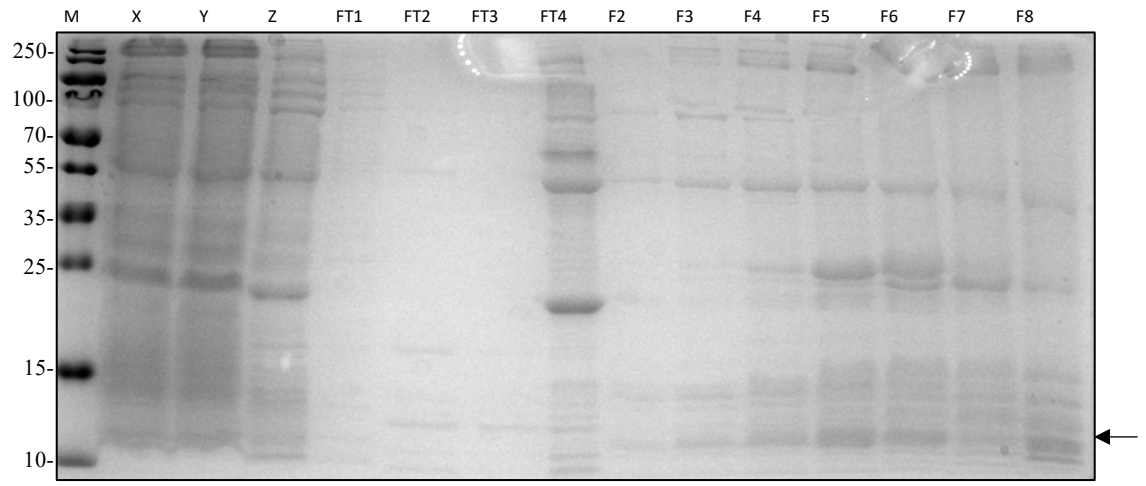
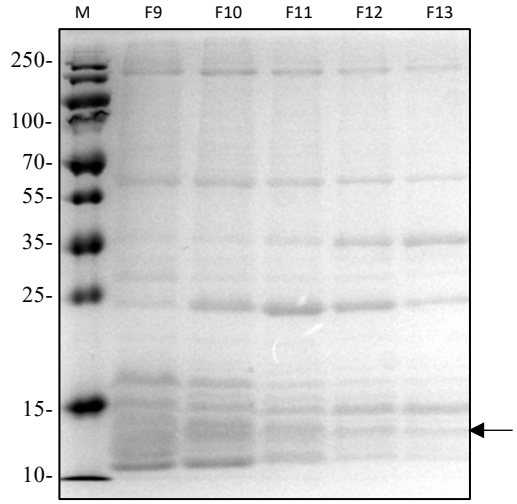


Figure 16 - **SDS PAGE analysis of upon-large-scale C protein production in BL21-CodonPlus (DE3)-RIPL cells at 30° Cd.** M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Numbers on top stand for the time point when the samples were collected. Arrow corresponds to expected area of C-core protein. Gap correlates to parts of the gel which were not part of Capsid expression experiment.

As described in 2.4.1. resuspension buffer containing protease inhibitor was added to the pelleted cells prior to cell lysis using French Press. Thereafter, the lysate was spun down using ultracentrifuge. The supernatant containing the soluble C-core protein was loaded onto HisTrap HP 5 mL column (Cytiva) in ÄKTA Pure System (GE Healthcare). After sample loading, the column was washed with 50 mM Imidazole to remove the weakly-bound *E. coli* proteins. Any proteins which came out from the column were collected and marked as soft-wash. The C-core protein was then eluted using a gradient of Imidazole (0.05 – 1 M). Collected fractions were analysed using SDS-PAGE and Western blot (Figure 17). For comparison after purification, cultures right after the expression (X), cultures after lysis (Y) and cultures after ultracentrifugation (Z) were added to the SDS-PAGE analysis.





Legend for SDS-PAGE analysis	
FT1	Flow-through during sample loading
FT2	Eluted proteins with low His-Tag affinity
FT3	Eluted proteins with low His-Tag affinity
FT4	Soft wash
F1-F13	Eluted fractions with increasing Imidazole (0.05 – 1 M)

Figure 17 – **SDS PAGE analysis from purification of the capsid core protein expression with BL21-CodonPlus (DE3)-RIPL at 30°C after 2 hours expression time.** M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. Arrow corresponds to expected bands of C-core protein.

Legend of Chromatogram	
Color	**Description**
Yellow	Flow-through
Red	Soft wash
Blue	Eluted fractions

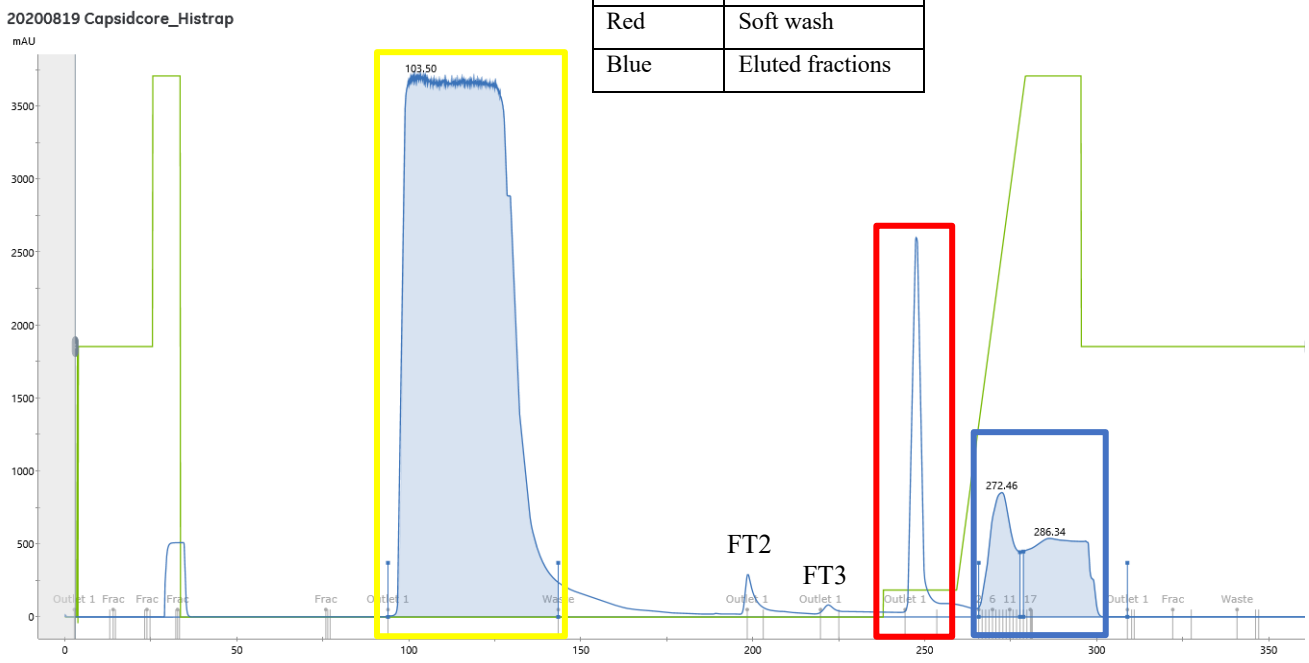


Figure 18 – Chromatogram of BL21-CodonPlus (DE3)-RIPL purification with IMAC through ÄKTA Pure System (GE Healthcare). X-axis refers to time spent, y-axis refers to UV absorbance, green line represents buffer B (see Table 2-16 for composition).

A Western blot analysis was done for fractions obtained after IMAC purification (Figure 19) and confirmed the presence of recombinant C-core protein (13 kDa). For comparison the fourth fraction of soft wash (FT4) from purification and the samples right after expression (2h and 0h) were added to the Blot. The FT4 shows no band at all because it contained proteins without 6x Histidine tag, which were eluted from the column during column washing with 50 mM Imidazole.

The peak fractions were found in fraction 7 and 8. The protein showed a sign of specific cleavage (seen as a double band in Figure 17), which could be due to degradation, proteolytic cleavage or problems with protein translation from the C-terminal. However, further investigation needs to be done in optimizing the expression, since the protein yield was continuously low.

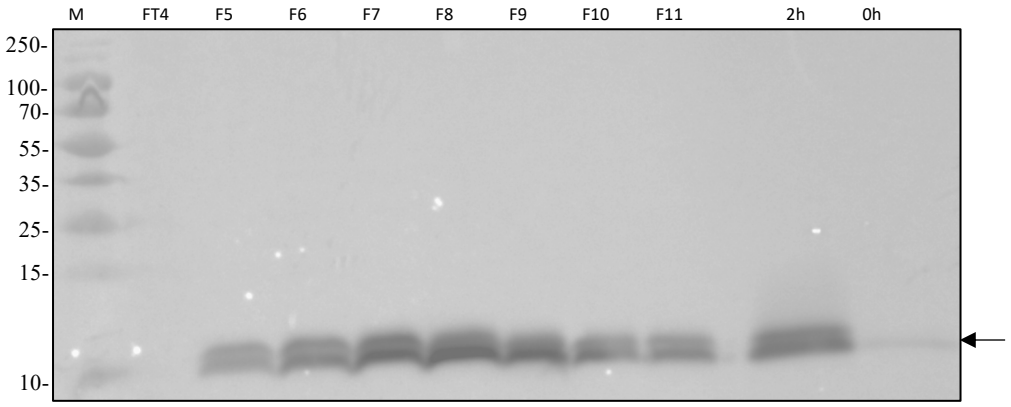


Figure 19 - **Western blotting of IMAC purification with pET-19B-BL-21 Codon plus at 30° C.** M: PageRuler Prestained Protein Ladder (Thermo Fischer) with associated size in kDa on the left. FT stands for flowthrough and F for Fractions collected after specific time points. On the right are shown soluble fractions directly after expression for comparison. Arrow corresponds to expected bands of C-core protein.

4 Discussion

Several strategies were made in current study to achieve a successful production of the recombinant C-core protein from TBEV using the bacterial (*E. coli*) expression system which is simple, cost-saving and harbour high protein yield.

The first cell lines used for unsuccessful expression of C-anch construct were Lemo21(DE3) and OverExpress™ C43(DE3) competent *E. coli* cells. Lemo21(DE3) allow the host features of BL21(DE3) (see below) while also providing a feature for tunable expression with varying level of lysozyme, which is modulated by adding L-rhamnose to the expression cultures. This should support the expression of toxic membrane proteins and proteins prone to insoluble expression. OverExpress™ C43(DE3) strain is derived from C41(DE3) and contain genetic mutations to express toxic and membrane proteins with the help of its chloramphenicol-resistant plasmid that produces small amounts of T7 lysozyme.

The two main cell lines used for expression of C-core were BL21(DE3)pLysS and BL21-CodonPlus (DE3)-RIPL, where only the latter lead to successful expression (see Results). Competent cells with the designation DE3 refer to the λDE3 lysogen, which implicates the T7

bacteriophage gene I, encoding the T7 RNA polymerase under the control of the lac UV5 promoter.

BL21(DE3)pLysS competent cells usually allow high-efficiency protein expression of any gene that is under the control of a T7 promoter and has a ribosome binding site, even with toxic proteins. This cell line is constructed with a resistance against Chloramphenicol and markedly reduce background expression by binding to, and inhibiting the low levels of T7 polymerase present before induction due to co-expression of the T7 lysozyme gene (Pan & Malcolm, 2000). However, as visible in Figure 15, there was no successful expression observable with these competent cells.

BL21-CodonPlus competent cells are derived from Agilent's BL21-Gold competent cells and enable efficient high-level expression of heterologous proteins in *E. coli* by overcoming codon bias. They are suitable for expressing proteins from all species, including mammals and the respective strain BL21-CodonPlus (DE3)-RIPL contain extra copies of rare *E. coli* tRNA genes (argU, ileY, leuW, proL) which corrects for codon bias and dramatically improves expression of sequences from other organisms (230280 | Agilent, 2022).

The approach to produce TBEV C protein in current study is similar to one performed by Jones *et al.* (2003) for DENV and YFC C proteins. The cDNAs of the mature C protein from those viruses were cloned into a pET30a expression plasmid (Novagen) which also carries a N-terminal His-Tag sequence as the expression plasmid used in this thesis (pET-19b). Protein expression was performed using BL21-CodonPlus (DE3)-RIL cells, which share similar properties as the BL21-CodonPlus (DE3)-RIPL (used in current study), except that it is engineered especially for AT-rich genomes. The protein expression was induced using final IPTG concentration of 1 mM and the optimal temperature for bacterial expression were 30°C (for DENV) and 25°C (for YFV). Another study also expressed C protein from Japanese Encephalitis Virus using pET30a vector system and *E. coli* BL21(DE3) competent cells (Poonsiri *et al.*, 2019).

In order to investigate other host cell nuclear proteins that bind the C from DENV in a study from Colpitts *et al.* (2011), the mature DENV serotype 2 C was cloned into the NTAP tagged expression vector and expressed in Huh-7 human liver cells. This method could be used to understand how C-anch proteins interact with human cells and binds to host specific histones, since there was no successful expression with bacterial strains most likely due to the embedment of the C-anch into the bacterial membrane.

Baculovirus or the Venezuelan equine encephalitis virus (VEE) expression system in insect cells represent a robust method for production of recombinant membrane proteins. It is highly effective, however, it can be time consuming and a relatively expensive expression system (Xia *et al.,* 2006). Another possibility is to use genetically engineered plants for expression, where transgenic tobacco and potato plants were created to express also the C protein from NV in a form of VLPs, which appeared identical to insect-cell derived VLPs (Mason *et al*., 1996). A common limitation of this system is the potential particle contamination from host components.

Further, eukaryotic expression system may serve as a promising approach to express C protein in its natural fold. The C protein from NV was expressed using yeast expression system (*Pichia pastoris*) and spontaneously formed virus-like particles (VLPs) which are morphologically and antigenically similar to insect cell-derived VLPs (Xia *et* al., 2006). This suggested that yeast expression system is a simple and effective alternative and the produced proteins are likely to be processed with natural folding and modification. Another advantage of yeast systems is the possibility of introducing post-translational modifications, such as glycosylation or phosphorylation to the structure of VLPs.

There are other expression systems which could be taken into account to assist drug development with each having unique features and advantages, especially an expression system which can lead to the assembly of VLPs. VLPs can be defined as an accumulation of macromolecular proteins which are replication-incompetent. They are multisubunit self-assembly-competent protein structures which should mimic the corresponding viruses in their overall structure. Even in the absence of frequently used adjuvants the nanometer-sized VLPs induce strong humoral and cellular immune response (T-cell activation) since antigen-presenting cells identify and degrade them. Newly constructed recombinant VLPs have significant advantages over native viruses, such as their availability in similar or even better functional properties. They can be expressed through bacterial, yeast, insect, plant, mammalian or cell-free systems, where the first tested expression system is usually through bacteria, since it is a generally accepted technology, despite their disadvantages (protein solubility problems, inability to generate proper disulfide bonds,…), which is cost and time saving (Zeltins, 2012).

5 Conclusion

Flavivirus Capsid protein the primary structural protein to interact with the viral genomic RNA within virus particles and is therefore necessary for efficient viral packaging, though it appears to have additional important roles aside from genome encapsidation.

Throughout this thesis it was shown that we only successfully expressed and isolated the C-core protein, while capsid protein with membrane anchor failed to express through bacterial expression systems. Most suitable conditions to express the desired protein were achieved with a pET-19b plasmid containing the C-core fragment and expression occurred using BL21-CodonPlus (DE3)-RIPL competent cells.

Every viral protein is theoretically an antiviral target, but the capsid protein has no enzymatic core and there are three important interactions of interest with the host cells which would allow antiviral strategies to focus on disrupting. The hydrophobic face of C protein interacts with ER membrane and lipid droplets to aid the assembly of virions and store it prior to packaging. An Antiviral treatment could therefore prevent capsid storage on lipid droplets, an entry into nucleus/nucleolus or prevent particle assembly. If the host proteins of the lipid metabolism are targeted for antiviral treatment, issues could be caused with particle formation and possibly lead to formation of VLPs (Sotcheff & Routh, 2020).

To hinder capsid-protein interactions, such as dimerization and nucleocapsid formation, it would result in viral attenuation, while reversely the stabilization of the capsid dimers may prevent genome release upon cell entry. Other capsid interactions such as capsid-RNA, or capsid-phospholipid membrane interactions with the associated host proteins also need to be studied more thoroughly to target these interactions for further antiviral strategies and vaccine design.

6 References

Acosta, E. G., Castilla, V., & Damonte, E. B. (2009). Alternative infectious entry pathways for dengue virus serotypes into mammalian cells. *Cellular Microbiology*, *11*(10), 1533–1549. https://doi.org/10.1111/j.1462-5822.2009.01345.x

230280 | Agilent. (2022). Agilent. Retrieved April 20, 2022, from https://www.agilent.com/store/productDetail.jsp?catalogId=230280&catId=SubCat3ECS_228019

Barnard, T. R., Abram, Q. H., Lin, Q. F., Wang, A. B., & Sagan, S. M. (2021, May). Molecular Determinants of Flavivirus Virion Assembly. *Trends in Biochemical Sciences*, *46*(5), 378–390. https://doi.org/10.1016/j.tibs.2020.12.007

BioLabs, N. E. (n.d.). NEBio Calculator. Retrieved 04 10, 2021, from http://nebiocalculator.neb.com/#!/ligation

Blitvich, B., & Firth, A. (2015). Insect-Specific Flaviviruses: A Systematic Review of Their Discovery, Host Range, Mode of Transmission, Superinfection Exclusion Potential and Genomic Organization. *Viruses*, *7*(4), 1927–1959. https://doi.org/10.3390/v7041927

Bogovic, P., Lotric-Furlan, S., & Strle, F. (2010). What tick-borne encephalitis may look like: Clinical signs and symptoms. *Travel Medicine and Infectious Disease*, *8*(4), 246–250. https://doi.org/10.1016/j.tmaid.2010.05.011

Bogovic, P. (2015). Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management. *World Journal of Clinical Cases*, *3*(5), 430. https://doi.org/10.12998/wjcc.v3.i5.430

Byk, L. A., Iglesias, N. G., de Maio, F. A., Gebhard, L. G., Rossi, M., & Gamarnik, A. V. (2016). Dengue Virus Genome Uncoating Requires Ubiquitination. *mBio*, *7*(3). https://doi.org/10.1128/mbio.00804-16

Cammisa-Parks, H., Cisar, L. A., Kane, A., & Stollar, V. (1992). The complete nucleotide sequence of cell fusing agent (CFA): Homology between the nonstructural proteins encoded by CFA and the nonstructural proteins encoded by arthropod-borne flaviviruses. Virology, 189(2), 511–524. https://doi.org/10.1016/0042-6822(92)90575-a

Chhetri, G., Kalita, P., & Tripathi, T. (2015). An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli. Molecular and Strutural Biophysics

Laboratory, Department of Biochemistry, North-Eastern Hill University, Shillong 793 022, India: MethodsX 2.

Colpitts, T. M., Barthel, S., Wang, P., & Fikrig, E. (2011). Dengue Virus Capsid Protein Binds Core Histones and Inhibits Nucleosome Formation in Human Liver Cells. *PLoS ONE*, *6*(9), e24365. https://doi.org/10.1371/journal.pone.0024365

Dobler, G. (2010). Zoonotic tick-borne flaviviruses. *Veterinary Microbiology*, *140*(3–4), 221–228. https://doi.org/10.1016/j.vetmic.2009.08.024

Firth, A. E., & Atkins, J. F. (2009). A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1' may derive from ribosomal frameshifting. Virology Journal, 6(1), 14. https://doi.org/10.1186/1743-422x-6-14

Freire, J. M., Veiga, A. S., Conceição, T. M., Kowalczyk, W., Mohana-Borges, R., Andreu, D., Santos, N. C., da Poian, A. T., & Castanho, M. A. R. B. (2013). Intracellular Nucleic Acid Delivery by the Supercharged Dengue Virus Capsid Protein. *PLoS ONE*, *8*(12), e81450. https://doi.org/10.1371/journal.pone.0081450

Füzik, T., Formanová, P., Růžek, D., Yoshii, K., Niedrig, M., & Plevka, P. (2018). Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-02882-0

Guirakhoo, F., Heinz, F. X., Mandl, C. W., Holzmann, H., & Kunz, C. (1991). Fusion activity of flaviviruses: comparison of mature and immature (prM-containing) tick-borne encephalitis virions. *Journal of General Virology*, *72*(6), 1323–1329. https://doi.org/10.1099/0022-1317-72-6-1323

Holbrook, M. (2017). Historical Perspectives on Flavivirus Research. *Viruses*, *9*(5), 97. https://doi.org/10.3390/v9050097

iba Solutions for Life Sciences. (2019). Retrieved May 11, 2020, from https://www.iba-lifesciences.com/details/product/2-1322-000.html

Jones, C. T., Ma, L., Burgner, J. W., Groesch, T. D., Post, C. B., & Kuhn, R. J. (2003). Flavivirus Capsid Is a Dimeric Alpha-Helical Protein. *Journal of Virology*, *77*(12), 7143–7149. https://doi.org/10.1128/jvi.77.12.7143-7149.2003

Kaufman, F., Dostálková, A., Pekárek, L., Thanh, T. D., Kapisheva, M., Hadravová, R., Bednárová, L., Novotný, R., Křížová, I., ČErný, J., Grubhoffer, L., Ruml, T., Hrabal, R., & Rumlová, M. (2020). Characterization and in vitro assembly of tick-borne encephalitis virus C protein. *FEBS Letters*, *594*(12), 1989–2004. https://doi.org/10.1002/1873-3468.13857

Khasnatinov, M. A., Tuplin, A., Gritsun, D. J., Slovak, M., Kazimirova, M., Lickova, M., Havlikova, S., Klempa, B., Labuda, M., Gould, E. A., & Gritsun, T. S. (2016). Tick-Borne Encephalitis Virus Structural Proteins Are the Primary Viral Determinants of Non-Viraemic Transmission between Ticks whereas Non-Structural Proteins Affect Cytotoxicity. PLOS ONE, 11(6), e0158105. https://doi.org/10.1371/journal.pone.0158105

Lin, K. C., Chang, H. L., & Chang, R. Y. (2004). Accumulation of a 3′-Terminal Genome Fragment in Japanese Encephalitis Virus-Infected Mammalian and Mosquito Cells. Journal of Virology, 78(10), 5133–5138. https://doi.org/10.1128/jvi.78.10.5133-5138.2004

Ma, L., Jones, C. T., Groesch, T. D., Kuhn, R. J., & Post, C. B. (2004). Solution structure of dengue virus capsid protein reveals another fold. *Proceedings of the National Academy of Sciences*, *101*(10), 3414–3419. https://doi.org/10.1073/pnas.0305892101

MACHEREY-NAGEL. (n.d.). MACHEREY-NAGEL. Retrieved 04 4, 2021, from https://www.mn-net.com/media/pdf/45/51/02/Instruction-NucleoSpin-Plasmid.pdf

Mansfield, K. L., Johnson, N., Phipps, L. P., Stephenson, J. R., Fooks, A. R., & Solomon, T. (2009). Tick-borne encephalitis virus – a review of an emerging zoonosis. *Journal of General Virology*, *90*(8), 1781–1794. https://doi.org/10.1099/vir.0.011437-0

Mason, H. S., Ball, J. M., Shi, J. J., Jiang, X., Estes, M. K., & Arntzen, C. J. (1996). Expression of Norwalk virus capsid protein in transgenic tobacco and potato and its oral immunogenicity in mice. *Proceedings of the National Academy of Sciences*, *93*(11), 5335–5340. https://doi.org/10.1073/pnas.93.11.5335

Oliveira, E. R., Mohana-Borges, R., de Alencastro, R. B., & Horta, B. A. (2017). The flavivirus capsid protein: Structure, function and perspectives towards drug design. *Virus Research*, *227*, 115–123. https://doi.org/10.1016/j.virusres.2016.10.005

Pan, S. H., & Malcolm, B. A. (2000). Reduced Background Expression and Improved Plasmid Stability with pET Vectors in BL21 (DE3). *BioTechniques*, *29*(6), 1234–1238. https://doi.org/10.2144/00296st03

PPP Mastermix. (n.d.). Retrieved May 11, 2020, from Top-Bio: http://www.top-bio.com/pcr-master-mixes-46.html?ppp-master-mix-1398#breadCrumbs

Poonsiri, T., Wright, G. S. A., Solomon, T., & Antonyuk, S. V. (2019). Crystal Structure of the Japanese Encephalitis Virus Capsid Protein. *Viruses*, *11*(7), 623. https://doi.org/10.3390/v11070623

Pulkkinen, L., Butcher, S., & Anastasina, M. (2018). Tick-Borne Encephalitis Virus: A Structural View. *Viruses*, *10*(7), 350. https://doi.org/10.3390/v10070350

Q5 High-Fidelity DNA Polymerase Kit. (n.d.). Retrieved May 11, 2020, from New England BioLabs Inc: https://www.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491

RED. (n.d.). New England BioLabs Inc. Retrieved May 11, 2020, from http://nebcloner.neb.com/#!/protocol/re/double/NdeI,XhoI

Ruzek, D., Avšič ŽUpanc, T., Borde, J., Chrdle, A., Eyer, L., Karganova, G., Kholodilov, I., Knap, N., Kozlovskaya, L., Matveev, A., Miller, A. D., Osolodkin, D. I., Överby, A. K., Tikunova, N., Tkachev, S., & Zajkowska, J. (2019). Tick-borne encephalitis in Europe and Russia: Review of pathogenesis, clinical features, therapy, and vaccines. *Antiviral Research*, *164*, 23–51. https://doi.org/10.1016/j.antiviral.2019.01.014

Simmonds, P., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Muerhoff, S., Pletnev, A., Rico-Hesse, R., Smith, D. B., & Stapleton, J. T. (2017). ICTV Virus Taxonomy Profile: Flaviviridae. Journal of General Virology, 98(1), 2–3. https://doi.org/10.1099/jgv.0.000672

Sotcheff, S., & Routh, A. (2020). Understanding Flavivirus Capsid Protein Functions: The Tip of the Iceberg. *Pathogens*, *9*(1), 42. https://doi.org/10.3390/pathogens9010042

Stadler, K., Allison, S. L., Schalich, J., & Heinz, F. X. (1997). Proteolytic activation of tick-borne encephalitis virus by furin. Journal of Virology, 71(11), 8475–8481. https://doi.org/10.1128/jvi.71.11.8475-8481.1997

Stiasny, K., & Heinz, F. X. (2004). Effect of Membrane Curvature-Modifying Lipids on Membrane Fusion by Tick-Borne Encephalitis Virus. *Journal of Virology*, *78*(16), 8536–8542. https://doi.org/10.1128/jvi.78.16.8536-8542.2004

Süss, J. (2008). Tick-borne encephalitis in Europe and beyond – the epidemiological situation as of 2007. *Eurosurveillance*, *13*(26). https://doi.org/10.2807/ese.13.26.18916-en

T4 DNA Ligase. (n.d.). Retrieved May 26, 2020, from New England BioLabs Inc.: https://www.neb.com/products/m0202-t4-dna-ligase#Product%20Information

Top-Bio. (n.d.). Retrieved Mai 11, 2020, from http://www.top-bio.com/files/1398_pl.pdf

Valenzuela, P., Medina, A., Rutter, W. J., Ammerer, G., & Hall, B. D. (1982). Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature*, *298*(5872), 347–350. https://doi.org/10.1038/298347a0

Vlachakis, D., Koumandou, V. L., & Kossida, S. (2013, May 7). A holistic evolutionary and structural study of flaviviridae provides insights into the function and inhibition of HCV helicase. PeerJ. https://peerj.com/articles/74/

Xia, M., Farkas, T., & Jiang, X. (2006). Norovirus capsid protein expressed in yeast forms virus-like particles and stimulates systemic and mucosal immunity in mice following an oral administration of raw yeast extracts. *Journal of Medical Virology*, *79*(1), 74–83. https://doi.org/10.1002/jmv.20762

Zeltins, A. (2012). Construction and Characterization of Virus-Like Particles: A Review. *Molecular Biotechnology*, *53*(1), 92–107. https://doi.org/10.1007/s12033-012-9598-4

7 Apendix

7.1 Information on Capsid Protein

DNA sequence

> Tick-borne encephalitis virus strain HYPR Core capsid, complete cds

ATGGTCAAGAAGGCCATCCTGAAAGGTAAGGGGGGCGGTCCCCCTCGACGAGTGTC
GAAAGAGACCGCAACGAAGACGCGTCAACCCAGAGTCCAAATGCCAAATGGGCTC
GTGTTGATGCGCATGATGGGGATCTTGTGGCATGCCGTAGCTGGCACCGCGAGAAA
CCCCGTATTGAAGGCGTTCTGGAACTCAGTCCCTCTGAAACAGGCCACAGCAGCACT
GCGGAAGATCAAAAGGACAGTGAGTGCCCTAATGGTTGGCTTGCAAAAACGCGGAA
AAAGGAGGTCAGCGACGGACTGGATGAGCTGGTTGCTAGTCATTACTCTGTTGGGG

Protein Sequence

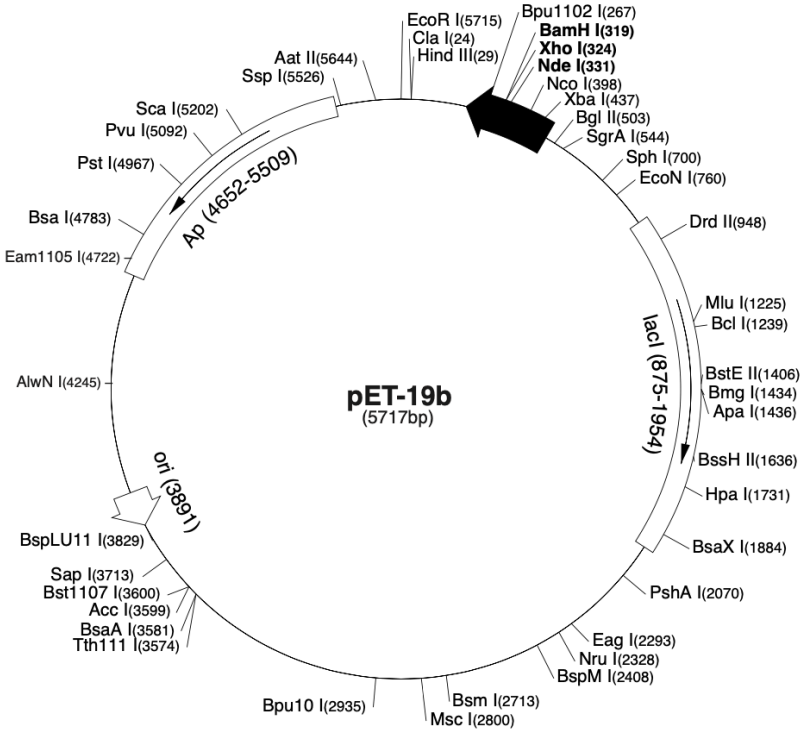
MVKKAILKGKGGGPPRRVSKETATKTRQPRVQMPNGLVLMRMMGILWHAVAGTARN
PVLKAFWNSVPLKQATAALRKIKRTVSALMVGLQKRGKRRSATDWMSWLLVITLLG

Highlighted area refers to capsid core sequence, primers are underlined, area not highlighted refers to the membrane-bound anchor

7.2 pET-19b properties

Description:	The pET-19b vector is developed for cloning and expression of recombinant proteins under the control of bacteriophage T7 transcription, which when fully induced, almost all cell's resources are converted to target gene expression
Affinity Tag	N-terminal His•Tag® followed by an enterokinase cleavage site and three cloning sites
Bacterial Expression	Expression is induced with a final concentration of 1 mM IPTG in shaking cultures with an OD_{600} = 0.8
Expression strain	Any *E. coli* strain with an T7 / T7lac promoter. Preferably pLysS
Resistance	Ampicillin

(Data retrieved from <https://www.merckmillipore.com>)



pET-19b sequence landmarks	
T7 promoter	472-488
T7 transcription start	471
His•Tag coding sequence	366-395
Multiple cloning sites (*Nde* I -*Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	875-1954
pBR322 origin	3891
bla coding sequence	4652-5509