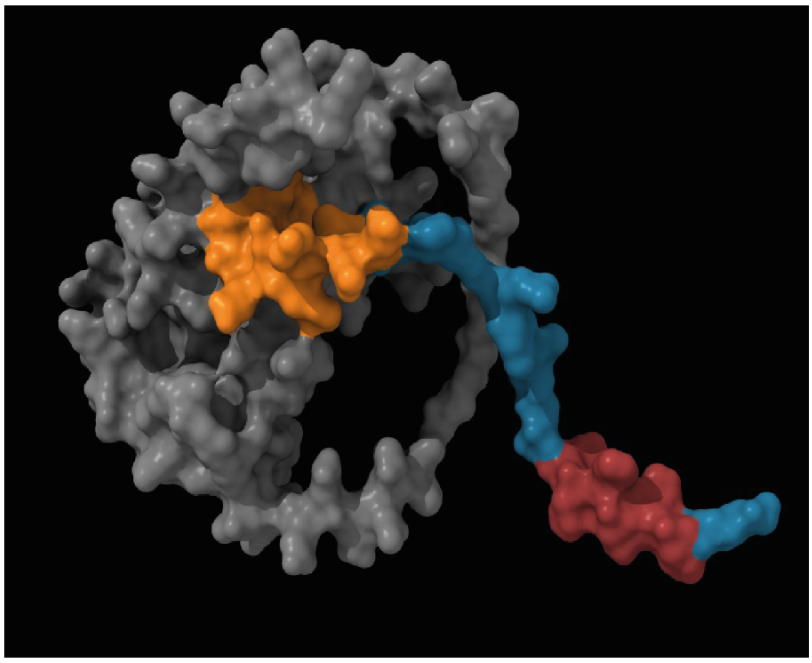

Production of recombinant NS2B-NS3 in *E. coli* and inhibition studies

Master thesis project
in Biotechnology
2024

AUTHOR
Isak Hallberg

SUPERVISOR
Javier Linares-Pastén

EXAMINER
Carl Grey

LUND UNIVERSITY

LTH
FACULTY OF ENGINEERING

Declaration of originality

I hereby declare that this master thesis solely consists of my own independent work, any sources and or indirect or direct aids have been appropriately acknowledged. I have read the Lund University policy of plagiarism and am fully aware of the consequences of academic dishonesty, and thereby certify that this master thesis adheres to the highest of standards in terms of originality and integrity.

2024-06-14	
Place and date of submission	Isak Hallberg

Abstract

The recent 2019 COVID pandemic elucidated the tremendous effects a viral outbreak can have on the societies of the world with almost 7 million cumulative deaths worldwide. It is therefore of high importance to prepare preventative measures to aid the handling of emerging viral threats before a new pandemic is realized. One of the more prevalent and rapidly spreading viruses today are the dengue virus (DENV). When looking closer at the viral cascade involved in infecting the host, one enzyme plays an especially important role - the NS2B-NS3 protease. Previous endeavors in the field of bioinformatics proved that the addition of a histidine-tag, that is used for purification, may be problematic in the case of the NS2B-NS3, as the tag seem to exhibit inhibitory effects on the activity of the enzyme. Based on these results, the aim of this study was to produce recombinant NS2B-NS3 with a TEV protease cleavage site added between the his-tag and the protease, allowing the removal of the his-tag through digestion with TEV protease. An attempt was also made to purposefully inhibit the activity of the NS2B-NS3, as inhibition of these viral enzymes may prove useful in the development of a possible anti-viral drug. Based on that premise, production was followed by in-silico and in-vitro docking experiments of 3 different ligands, a peptide derived from the *Chenopodium quinoa* plant, carvacrol and chlorogenic acid. The synthetic quinoa peptide was unfortunately not soluble and where therefore only studied in-silico. The results showed that the production of recombinant NS2B-NS3 can be optimized to reach a higher degree of activity by removing the his-tag by TEV protease digestion. The NS2B-NS3 with his-tag exhibited a V_{max} of 1.512 ± 0.103 µM/s, a K_m of 0.735 ± 0.015 mM and a k_{cat} of 0.304 ± 0.021 µM/s while the NS2B-NS3 without his-tag exhibited a V_{max} of 2.127 ± 0.104µM/s, a K_m of 0.940 ± 0.025 mM and a k_{cat} of 0.375 ± 0.018 s^{-1}, using DL-BAPNA as substrate. Carvacrol did not seem to exhibit any inhibitory effects on the NS2B-NS3, whereas chlorogenic acid seemed to be a potential inhibitor, decreasing the enzyme activity by ~24% at a concentration of 10 µM, resulting in a $V_{max,app}$ of 1.628 µM/s, a $K_{m,app}$ of 0.672 mM. and a $k_{cat,app}$ of 0.286 s^{-1}.

Keywords: DENV NS2B-NS3, TEV protease, viral protease inhibitor, bioinformatics, recombinant protein production, docking simulations, AutoDock VINA, ChimeraX, AlphaFold, Quinoa, Carvacrol, Chlorogenic acid

Acknowledgements

I am very grateful for having the chance to carry out my master thesis under the supervision of the calm, composed and highly devoted Javier Linares-Pastén. I want to thank Carl Grey for accepting the role as examiner of my project and for providing invaluable constructive criticism of the theories and conclusions presented in my thesis. I also want to thank Martin Hedström for listening in to my personal interests, aspirations, and goals and for directing me to the correct division here at LTH.

I am eternally grateful for the support and understanding given to me by my family, and my beloved Sarah.

The path to knowledge and discovery is the path less taken, hence it is often also the path with most resistance. I am very appreciative for the opportunity I have had to hone my traits and skills through this academical journey towards my engineering degree. Not only have I had the chance to develop my mindset and expand my knowledge, but I have built character and discipline, and found something reminiscent of a definition of myself. Through hardship, through difficulties, and through obstacles. It has been a long road, yet, I have enjoyed the ebb and flow of these trials. In the same way a ferocious sea yields an experienced sailor, the obstacles I have faced along the way is what has allowed this metamorphosis.

And for that, I am grateful.

"'Man cannot remake himself without suffering, for he is both the marble and the sculptor."
- Alexis Carrel

"We are what we repeatedly do. Excellence, then, is not an act, but a habit."
- Aristotle

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
AVG	Average
DENV	Dengue Virus
DHF	Dengue Hemorrhagic Fever
DNA	Deoxyribonucleic acid
DSS	Dengue Shock-Syndrome
ER	Endoplasmic Reticulum
PDB	Protein Data Bank
PI	Protease Inhibitors
RNA	Ribonucleic acid
TGN	Trans-Golgi Network

1 Introduction

The recent 2019 COVID pandemic elucidated the tremendous effects a viral outbreak can have on the societies of the world with almost 7 million cumulative deaths [1]. It is therefore of high importance to prepare preventative measures to aid the handling of emerging viral threats before a new pandemic is realized. One of the more prevalent and infectious viruses today are the Dengue virus, infecting up to 400 million people each year [2]. The Dengue virus is spread to humans through the bite of an infected mosquito belonging to the *Aedes* genus. What makes the Dengue virus especially susceptible for studying is that the mosquito responsible for its spread can be found in areas of the world where 4 billion people live, thus making it pose a major threat in the case of mutation [2]. Finding inhibitors for the proteases responsible for the virus's infectiousness makes for a great foundation in the pursuit of developing a potential anti-viral drug. In this study, an endeavor was made to produce recombinant NS2B-NS3 protease, a complexed non-structural transmembrane protein located within the endoplasmic reticulum (ER) that is involved in replication and assembly of the dengue flavivirus [3]. Previous studies in production of the NS2B-NS3 protease from the dengue virus have pointed to the fact that the added histidine-tag could act as an inhibitor for the enzyme. Based on these findings, a countermeasure has been implemented consisting of a linker sequence with an added TEV protease cleavage site that have been introduced between the protease and the histidine-tag. This allows the fused tag to be cleaved by TEV protease digestion after the initial purification, and then separated from the NS2B-NS3 solution using a *HisTrap* column in an Immobilized Metal Ion Affinity Chromatograph (IMAC), allowing the untagged enzyme to leave the column in the permeate. The isolated protease can then be used for inhibition studies with the three potential inhibitors that have been carefully selected for this project. Inhibition of these types of viral enzymes suppresses the infectious cascade that is set off when the host is infected by the virus, impeding replication of the virus molecules in the host cell. There are already some protease inhibitors today that are marketed and available as medication for specific viral infections such as HIV/AIDS, Hepatitis-C or the more recent COVID-19 [4], [5]. Discovery of novel inhibitors to viral proteases is one of the more auspicious approaches for developing drugs aimed at treating, or hopefully, preventing these types of diseases – with the end goal of increasing the survivability of patients.

2 Aim

The aim of this study is to perform a thorough bioinformatic analysis of the recombinant NS2B-NS3 protease derived from the dengue virus (DENV) including a structural prediction using AlphaFold. This will be followed by producing the recombinant NS2B-NS3 in *Escherichia coli*, cleaving the histidine-tag by TEV protease digestion, and determining if the removal of the tag affected the activity of the viral protease through a colorimetric assay. Finally, the ligands will be studied in-silico through docking simulations, and then the inhibitory effects will be assessed through a colorimetric inhibition assay, followed by determination of the kinetic constants of the reactions.

3 Scientific Background

To facilitate a deeper understanding of the contents and results of this study, a thorough compilation of background information has been assembled covering all knowledge required for the reader.

3.1 Dengue Virus (DENV)

3.1.1 Taxonomy, structure, and evolution

The dengue virus is a single-stranded positive-sense RNA virus, consisting of ~10,700 bases with a mature diameter of 500 Å that causes the viral infection called dengue fever [6]. The viral genome of the DENV consists of a 5' and 3' untranslated region, and an ORF, coding the three structural proteins – C protein (Capsid), PrM (Pre-Membrane) and E protein (Envelope) – as well as the 7 non-structural N-proteins - NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5, involved in replication through a diverse set of enzymatic activities [7].

DENV belongs to the family *Flaviviridae* of the genus *Flavivirus* and are one of the five viruses belonging to that genus that can infect humans, together with Zika, West Nile virus, tick-borne encephalitis, and yellow fever [8]. All these viruses are arthropod borne, meaning that they are spread by a vector organism, and in the case of DENV, the vector is the mosquito belonging to the family *Aedis aegypti* or *Aedis albopictus.* The DENV virus exist in five different serotypes; *type 1, 2, 3 and 4*, and a recently discovered *type 5* [9]. The fact that the *Aedis* mosquito dwells in some of the most populated areas of the world is what makes DENV pose a major threat as seen from an epidemiological perspective – infecting up to 400 million people annually, with 4 billion at risk of infection [2]. Regions expressing subtropical or tropical climate are the ones affected the worst by the disease, as this allows for an optimal environment for procreation of the *Aedis* mosquito. And with rising temperatures, the regions exhibiting these types of climates will only expand.

Based on previous research it has been deduced that all these flaviviruses - both those spread by mosquitos and those spread by ticks - share a common ancestor. This conclusion was based on the fact that mosquito borne viruses have been found in ticks, and vice-versa [10]. This may very well be a factor that explains the ubiquitous nature of the NS2B-NS3 protease as an important part of the infectiousness of all these different viruses.

3.1.2 Replication and viral life cycle

The DENV is dependent on the intrusive nature of the vector organism – in this case the *Aedis* mosquito – allowing virions to enter the target host's native cellular replication system. The virus is released from the salivary glands of the mosquito, in which replication has already started [11]. This allows virions to enter the host instantaneously, infecting a wide variety of cell populations, ranging from epithelial cells, fibroblasts, monocytes, macrophages, dendritic cells, B-cells, T-cells, hepatocytes, and endothelial cells [12]. The virus molecules then bind to multiple different surface receptors of the host's cells, allowing the virus to enter the cell through receptor-

mediated endocytosis. After internalization of the virus molecule, viral disassembly allows the encapsulated genome to be released into the cytoplasm of the cell. The released RNA is then used as a template for translating and replicating the virus in the endoplasmic reticulum (ER). The NS-proteins, of which NS2B-NS3 is one of the more prominent and multi-facetted, is then responsible for several enzymatic reactions that allow translation and replication of the viral genome. The synthesized positive-sense RNA then interacts with NS-proteins for assembly. The NS2 protease is responsible for recruiting C-prM-E polyproteins and NS2B-NS3 to the site of assembly, the latter being a serine protease that cleaves the viral polyprotein, releasing important building blocks used for constructing the mature capsid protein, ultimately forming the nucleocapsid. The virus particle then matures in the trans-Golgi network (TGN) after passing through the Golgi apparatus, undergoing glycosylation and post-translational modifications eventually leading to exocytosis from the host cell for continuation of the course of infection [12], [7], [13]. The mature DENV virion ultimately consists of a lipid bilayer with inserted transmembrane viral proteins, inside there are multiple encapsulated C proteins covering the positive-sense single-stranded RNA [14].

3.1.3 Diagnosis and symptoms

About one in four infected humans will be symptomatic when contracting the DENV virus. The most common symptoms include rashes, general aches and pains, nausea, vomiting, and fever. An infection usually lasts between 2 and 7 days, with the majority of patients being fully recovered about a week after the initial symptoms arose. About 1 in 20 people who contract the virus will develop what is called *severe dengue* which can lead to a more severe condition called *dengue hemorrhagic fever* (DHF), and in some cases, *dengue shock syndrome* (DSS). DHF is a life-threatening condition leading to thrombocytopenia and hemorrhaging, which can induce a decrease of the intravascular volume in the infected patient, ultimately leading to shock (DSS). These symptoms are also why the dengue virus is compared to the Ebola virus, as both diseases cause inflammation of the host's cardiovascular system [15]. Repeated infections of the DENV virus increases the likelihood of developing severe dengue. In the present day there is no medicine that can be used to treat an ongoing DENV infection, elucidating the importance of novel discoveries in this field [2].

3.2 Proteases – a subclass of enzymes

3.2.1 Enzymes and their function

Enzymes are a type of protein or protein complex that acts as a natural catalyst by minimizing the required energy for a reaction to take place and are used to convert substrates into products. In the world of biological nomenclature one can identify an enzyme by the addition of the suffix "*-ase*" after the name of the substrate on which the enzyme acts [16]. The most commonly used parable for explaining the function of an enzyme is the *lock-and-key* figure, as depicted in Figure 1 below.

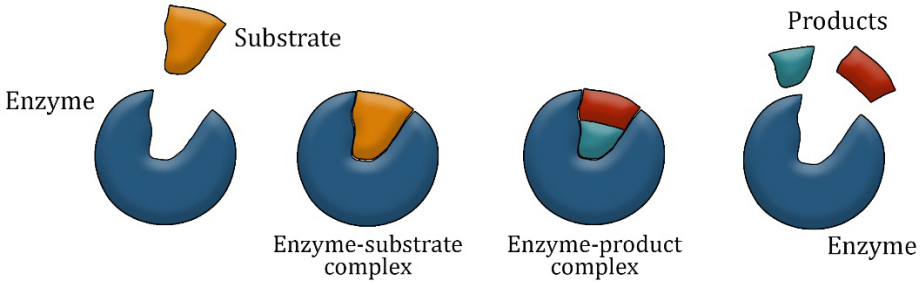

Figure 1: The classic lock-and-key representation of how an enzyme complexes and reacts with a substrate molecule to form products.

The lock-and-key explanation describes the fundamental function of an enzyme, i.e., the conversion of a substrate into products. The study of enzymes is most often focused on the kinetics of reactions. The most fundamental relationship that are applied in the field is the Michaelis-Menten equation (1):

$$V_0 = \frac{V_{max}[S]}{K_m + [S]} \qquad (1)$$

The equation contains the parameters V_{max} and K_m which aids in explaining and calculating how changes in substrate concentration affect the activity of the enzyme, and also allows for comparison between different enzymes based on their effectivity. In the equation V_{max} is the maximum velocity achieved and K_m the substrate concentration which yields half the maximum velocity. Other parameters used to evaluate enzyme productivity is the variable k_{cat} which is the maximum amount of substrate molecules that are converted to product each unit of time at saturation, k_{cat} is commonly referred to as the *turnover number* of the enzyme [17]. Determination of these constants, and comparison between different enzymes and inhibitors are often done using the linear inverse plot of the Michaelis-Menten equation called the Lineweaver-Burk plot (2).

$$\frac{1}{V_0} = \frac{K_m}{V_{max}[S]} + \frac{1}{V_{max}} \qquad (2)$$

Making the slope of the line K_m/V_{max}, the y-intercept $1/V_{max}$, and the x-intercept $1/K_m$.

3.2.2 Proteases - Function, mechanism, and applications

There are not a lot of things abundant enough to be considered a universal entity in our universe, however, the class of enzymes called *proteases* come very close. Proteases are a class of enzymes responsible for degrading proteins into single amino acids or smaller peptide chains by cleaving peptide bonds through hydrolysis. The degradation of products can also lead to indirect synthetic functions by cleaving and activating other enzymes and building blocks. Proteases play an important role in living organisms but are also used in a plethora of industrial applications. The utilization of proteases is ubiquitous in species ranging from microorganisms and plants to larger animals, clearly demonstrating the importance of the functions linked to this class of enzyme. The expression of proteases can have an influence on DNA replication, transcription, cell proliferation, differentiation, digestion of proteins, and a wide range of other cellular

functions, making it play a key role in all living organisms no matter the phylum or domain to which the species belongs [18], [19].

Proteases are classified in different categories depending on their mechanism of action and their regional specificity. The different types of proteases are *endo-* or *exopeptidases* based on their action away from, or in proximity of the termini, they are also divided into *cysteine*, *serine*, *aspartic*, *glutamic*, *threonine* and *metalloid proteases* depending on what residue is responsible for the catalytic activity of the enzyme. The reactivity of three of these classes – *aspartic*, *glutamic* and *metalloproteases* – utilizes a water molecule as a nucleophile, whereas in the other three classes the nucleophile is an amino acid residue within the molecular structure of the enzyme [19], [20].

The wide functionality, stability and susceptibility for genetic engineering makes proteases one of the more abundant enzymes used in the chemical and biotechnical industry today, with applications ranging from waste management, synthesis of compounds, production of foodstuffs and cleaning products, to disease control and medicine [18].

3.2.3 DENV NS2B-NS3

3.2.3.1 Function and mechanism

The NS2B-NS3 protease is a nonstructural protein with a molecular weight of ~26 kD that is involved in the replication of the flavivirus. The protease consists of two separate domains, one being the NS3 protease domain (*~11 kD*) which is a subpart of the full-length NS3 protein (~69 kD), and the other being the NS2B (*~15 kD*) acting as a cofactor to the NS3 by aiding in substrate recognition. The NS3 protease domain is a serine protease consisting of the N-terminal of the full-length NS3 which is the region responsible for the proteolytic activity of the NS2B-NS3, while the C-terminal of the full-length NS3 is responsible for RNA-helicase activity. The NS2B-NS3 complex undergoes a conformational change in the NS2B C-terminal domain when binding to the substrate, which leads to increased stability of a β-hairpin close to the active site of the enzyme. The two states in which the complex can exist is referred to as the *open* and *closed* conformation; where the former is required for proteolytic activity and the latter is required for recognizing the substrate binding site [21].

The active NS2B-NS3 protease is responsible for cleaving other proteases and tethering replicase complexes to intracellular membranes, i.e., it plays a role in a wide variety of enzymatic reactions involved in the replication process of the virus [21]. Apart from this, the protease also seem to play a role in regulating the host immune response by inhibiting production of *type-1-interferons* (IFN), cleaving the *stimulator of interferon* (STING) and *mediator of IRF3 activation* (MITA) genes [21], [22], [23]. It also cleaves the nonstructural region of the flavivirus polyprotein making it indirectly involved in RNA replication and virus assembly. As the NS2B-NS3 is initiating the mechanistic cascade responsible for replication, it makes for a promising drug target in the search of a potential cure to the dengue virus [24].

3.2.2.2 Structural analysis

The model used for comparison and as a reference during the structural analysis was obtained from *rcsb.org* with the accession number *6MO0*. However, this model was not structurally complete and did not contain two regions of the atomic structure of the recombinant protease; regions which is yet to be determined by x-ray crystallography. Instead, a model with a structure that was predicted by *AlphaFold* was used as the main model for analysis and docking in this study. The following sequence was used for predicting the structure of the NS2B-NS3:

GSSHHHHHHSSGLVPRGSHMENLYFQENLYFQGEYQGLMADLELERAADVRWEEQAEISGSSPILSITISEDGSMSI
KNEEEEQTLGGGGSGGGGAGVLWDVPSPPPVGKAELEDGAYRIKQKGILGYSQIGAGVYKEGTFHTMWHVTRGAVL
MHKGKRIEPSWADVKKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGLFKTNTGTIGAVSLDFSPG
TSGSPIVDKKGKVVGLYGNGVVTRSGAYVSAIANTEKSIEDNPEIEDDIFRK

The sequence included the recombinant NS2B-NS3 (teal), the added glycine linker which bind the two moieties (grey), the TEV protease cleaving site for removal of the 6x histidine-tag (tangerine) and the added histidine-tag (red).

The comparison of the predicted model and the original crystallographic model can be seen in detail in *Appendix A (*Figure A1*).* The structural prediction carried out in AlphaFold can be seen in Figure 2 below, showing the original unknown residues in tangerine, and the predicted structure in petrol blue.

Figure 2: The structural prediction made using AlphaFold, the tangerine residues are the original unknown region, and the petrol blue residues are the predicted region. Figure 2-A shows the predicted structure of the monomeric glycine linker, while Figure 2-B shows the predicted structure of an unknown region.

As depicted, the unknown regions seem to consist mainly of loops with a small number of beta-strands (Figure 2-A), and a minor antiparallel beta-sheet (Figure 2-B*)*. The missing structures are likely very mobile and therefore difficult to capture using x-ray crystallography [25].

The general structure of the NS2B-NS3 consist mainly of loops, with 19 beta-strands and 1 helical structure (Figure 3 & Figure 4). The loops span a rather large area compared to the clustered region containing the beta-strands and helix, allowing a higher mobility compared to the more stationary cluster. The 19 beta-strands make up 6 different anti-parallel beta-sheets of varying sizes that is distributed to one side of the protein structure. The residue forming the helical structure in the protease is one of the three

residues involved in the catalytic triad, i.e., the site responsible for the enzymatic activity of the protease [24].

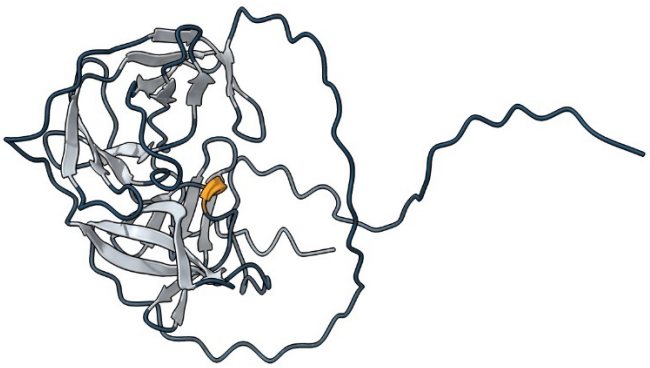

Figure 3: The secondary structure of the DENV NS2B-NS3, as predicted by AlphaFold. Beta-strands are depicted in white, loops are depicted in petrol blue and alpha helices are are depicted in tangerine.

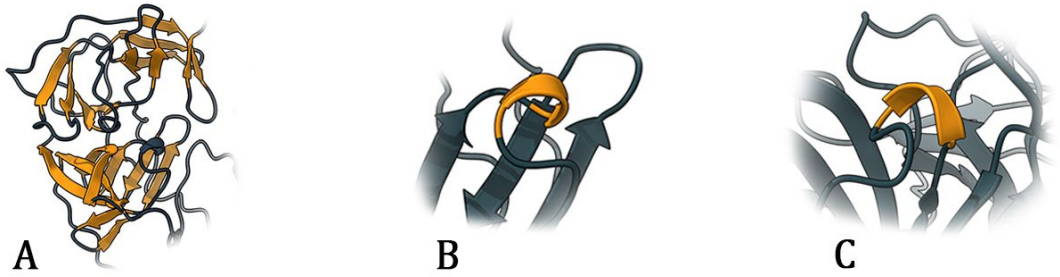

Figure 4: The composition of the secondary structure of the NS2B-NS3. The secondary structure in focus is highlighted in tangerine. Figure 4-A shows the beta-strands forming the sheets of the NS2B-NS3, Figure 4-B shows the alpha helix seen from the rotational axis, and Figure 4-C shows the alpha helix perpendicular to the rotational axis.

The long tail consisting of loops that stretches out towards the right in Figure 3 is the added linker histidine tag, the linker and its cleaving site. The different regions of the tag can be seen more precisely in Figure 5 below, showing the cleaving site, the linker sequence and the 6x-histidine tag.

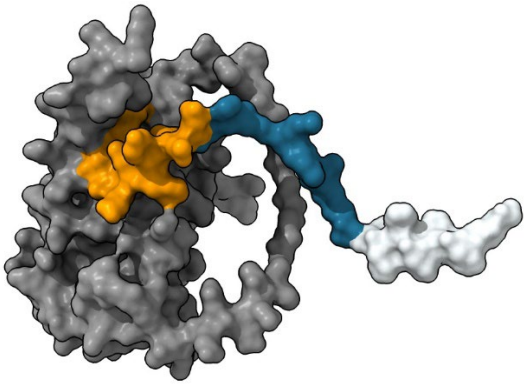

Figure 5: The molecular structure of the NS2B-NS3 with the added histidine tag, showing the surface of the protease. The TEV protease cleaving site is depicted in tangerine, the linker sequence in petrol blue and the histidine-tag in white.

The added cleaving sequence depicted in tangerine has the sequence ENLYFQG and is the cleaving site of the *Tobacco Etch Virus* (TEV) protease, which is the enzyme used for removal of the histidine tag [26]. To further minimize the risk of destructive interactions with the histidine tag and to allow for convenient separation in the subsequent purification steps, a non-active linker region (petrol blue, Figure 4*)* was added to the protease, separating the histidine residues (white*)* from the main structure (grey*).*

3.2.2.3 The catalytic triad and the allosteric site

The site responsible for the catalytic activity of the NS2B-NS3 protease is referred to as the catalytic triad and consists of the three residues *histidine 146* (H145), *aspartic acid 170* (D170) and *serine 230* (S230) (Figure 6). This triad is archetypical for the class of enzymes titled *serine proteases*, to which the NS2B-NS3 belongs.

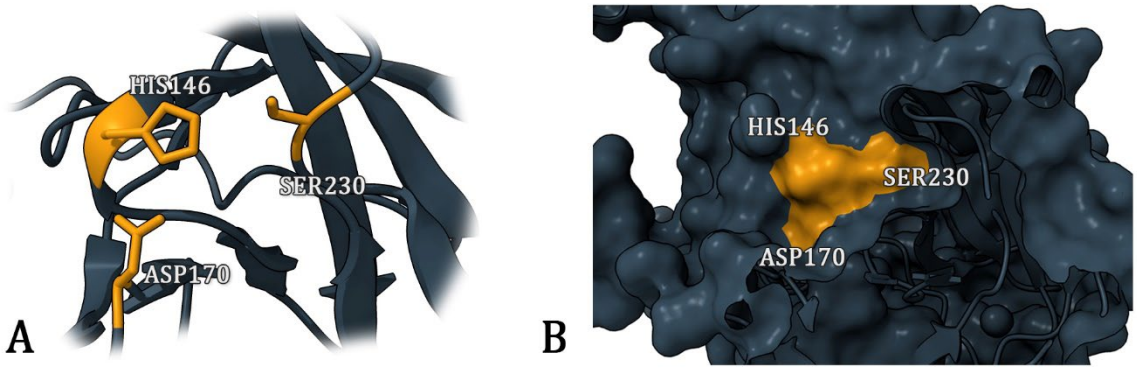

Figure 6: The catalytic triad of the NS2B-NS3 consisting of the residues HIS146, SER230 and ASP170 as depicted in tangerine. Figure 6-A shows the molecular structure of the residues while Figure 6-B shows the surface of the active site.

The NS2B-NS3 can exist in two main states, open (active) and closed (inactive), where the open state further exposes the enzymatic triad allowing for better access to the nucleophilic serine. Previous research has pointed to that the flexibility required for the NS2B-NS3 to achieve its open configuration can be hindered by obstructing a specific site on the enzyme, thus acting as an allosteric site inhibiting the activity of the protease [27]. Depicted below in Figure 7 is the location of the allosteric site obstructing the conformational change required to achieve full activity of the enzyme.

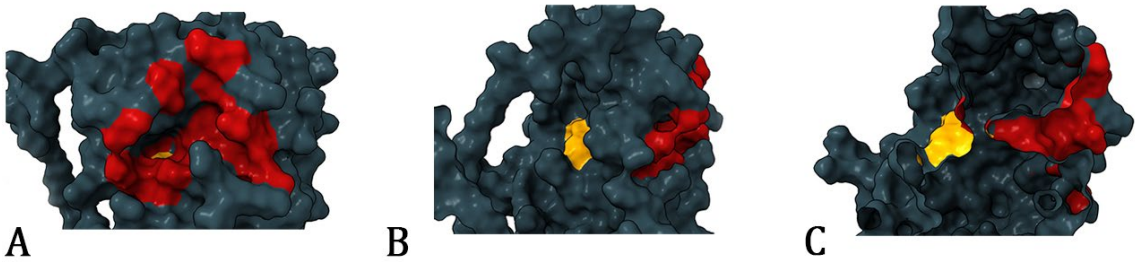

Figure 7: The location of the catalytic triad (tangerine) in relation to the allosteric site (red). Figure 7-*A shows the full allosteric site,* Figure 7-*B shows the catalytic triad and part of the allosteric site and* Figure 7-*C shows a cross-section of the internal proximity of the residues involved in both these sites.*

This site can pose as a great target for potential inhibitors as obstructing the mobility of the enzyme would suppress the activity of the NS2B-NS3.

3.4 Protease Inhibitors

3.4.1 Function, types of inhibition and mechanism

Protease inhibitors are a class of compounds having an inhibitory effect on important proteases involved in the replication of viruses. Viral proteases are responsible for cleaving the viral polyprotein and forming the structural capsid, which is involved in various attachment-, replication- and uncoating processes. Inhibition of these enzymes hinders the viral cascade responsible for the infectiousness of the virus, hence treating or ameliorating the complications of the infection.

There are two main categories of inhibition, *reversible* and *irreversible*. The formation of an irreversible inhibitor-enzyme complex is caused by the formation covalent bonds or by *slow binding*. The reversible inhibitors can be divided into four main cases – *competitive-, uncompetitive-*, *noncompetitive-,* and *mixed inhibition* – depending on how the inhibitor interacts with the receptor. It should be noted that the noncompetitive scenario can be described as a specific mixed inhibition, where the x-intercept remains constant in both the inhibited and uninhibited. Allosteric inhibition most often takes place as noncompetitive or uncompetitive inhibition depending on the properties of the ligand, but as the field of kinetics is highly complex there may be both intersections and exceptions [28], [29]. Figure 8 shows the four main cases of inhibition and how they affect the appearance of the Lineweaver-Burk plot.

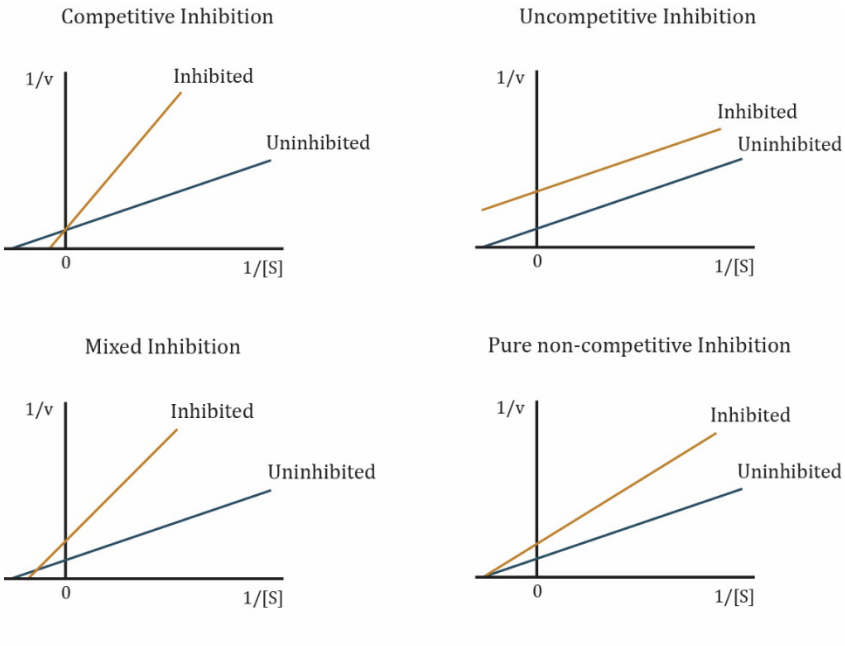

Figure 8: The Lineweaver-Burk plots showing the effect of the four main types of inhibition on the appearance of the slope and intercept of the plot. The y-axis shows the inverse of the initial rate of the enzyme while the x-axis shows the inverse of the substrate concentration.

3.4.2 Applications in biomedicine

The role of enzyme inhibitors in biomedicine has been under scrutinous research the past decades, but it is only during the past 30 years where the field have made noteworthy progress. Protease inhibitors have been successfully developed into drugs treating HIV (*Human Papillomavirus*), herpes virus, specific parasitic infections, aiding in cancer treatment and ameliorating cardiovascular diseases – making it a highly versatile tool for future research and development of novel biopharmaceuticals.

3.4.3 The ligands used in this study

Following the contemporary trends in the cross-border societal shift towards more sustainable, renewable, and naturally derived resources, the bioactive peptide used in this study have been derived from the ubiquitous and highly nutritious *Chenopodium quinoa,* otherwise known as the quinoa plant. Inhibitory studies will also be conducted using two plant derived chemical compounds, namely, *chlorogenic acid* found in coffee beans and *carvacrol* found in the oregano herb. The quinoa peptide was selected due to its potential competitive inhibition of the NS2B-NS3, while the chlorogenic acid and the carvacrol was believed to exhibit allosteric inhibition of the enzyme.

3.4.3.1 Quinoa derived peptide

The first ligand that were studied was a peptide from a hydrolysate derived from the common quinoa plant (*Chenopodium quinoa*) widely known for its nutritional value and role as an alternative carbohydrate source [29]. Based on a recent publication investigating novel bioactive peptides from quinoa hydrolysates, the peptide with sequence HVASGAGPW seemed to interact well with residues commonly found in the NS2B-NS3 protease. The chemical properties of the residues in the peptide also aligned well with the binding site at the catalytic triad and the allosteric site of the protease, further supporting this hypothesis. The production of this recombinant peptide would be in the form of a polypeptide, which later would be cleaved and purified using the same principles as in the production of the NS2B-NS3, i.e., using TEV protease cleavage sites.

Figure 9 shows the polypeptide construct that will be produced, its tertiary structure (A) and lipophilic profile (B & C).

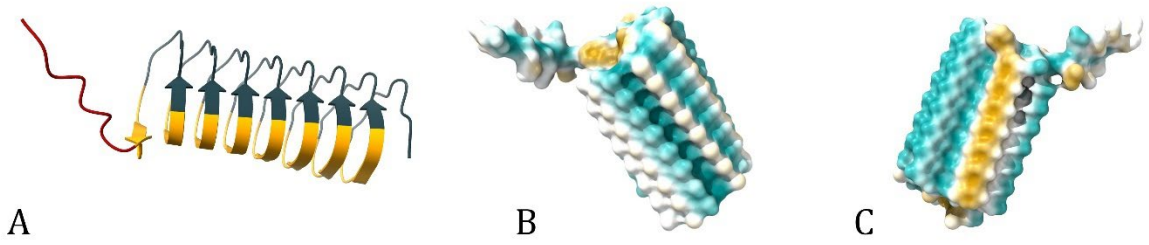

Figure 9: The tertiary structure (A) and the lipophilic profile of the produced polypeptide derived from the quinoa plant. Figure 9-A shows the peptide moieties (petrol blue), the cleavage sites (tangerine) as well as the his-tag in red. Figure 9-B and C shows the lipophilicity of the polypeptide with dark cyan being the most hydrophilic, white being neutral and orange being the most lipophilic.

The polypeptide construct would then undergo digestion by the TEV protease at the added cleavage sites containing the recognition sequence ENLYFQG depicted in tangerine, releasing the free peptide, and removing the his-tag (Figure 9-A*)*.

Depicted in Figure 10 below is the molecular structure (A), the lipophilic profile (B) and the electronegativity (C) of the peptide with the sequence HVASGAGPW.

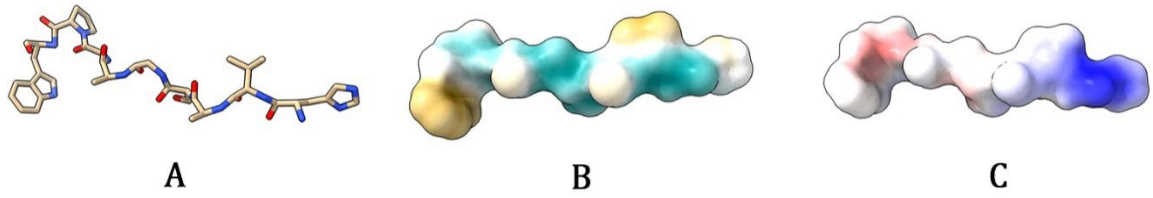

Figure 10: The molecular structure of the potential inhibitory peptide derived from quinoa (Chenopodium quinoa). Figure 10-A shows the atomic structure of the peptide. Figure 10-B shows the lipophilicity of the peptide structure with dark cyan being the most hydrophilic, white being neutral and orange being the most lipophilic. Figure 10-C shows the electronegativity of the peptide structure with red representing negative charge, white representing neutral and blue representing positive charge.

The general structure of the peptide (Figure 10-A) was an elongated chain, not very bulky and with a low level of inter-chain steric hindrance allowing the molecule to be more flexible when binding to the catalytic site. The lipophilic profile of the peptide (Figure 10-B) shows that the majority of the molecule is hydrophilic and polar, with two residues being predominantly lipophilic and non-polar, indicating that the molecule may have some surfactant-like properties. Figure 10-C shows the electronegativity of the molecule and that it is mostly neutral with a positively charged end.

3.4.3.2 Chlorogenic acid

Chlorogenic acid or *caffeoylquinic acid* is a compound that is predominantly found in berry-carrying plants, including the acclaimed *Coffea arabica,* the plant responsible for the worldwide production and consumption of coffee. Previous studies have uncovered several anti-viral and bioactive properties held by this compound, mainly based on medical practices originating from traditional Chinese herbal medicine. As the anti-viral mechanism of this compound to this day is unknown, chlorogenic acid was deemed a highly interesting candidate for inhibitory analysis on the viral protease in this study. As chlorogenic acid also is present in brewed coffee, establishing the bioactivity of this molecule may support previous findings in the potential health benefits of coffee consumption [30], [31]. Depicted below in Figure 11 is the molecular structure of chlorogenic acid.

Figure 11: The molecular structure of chlorogenic acid found in berry-carrying plants.

The chlorogenic acid molecule contains several functional groups ranging from simpler hydroxyl groups to esters and carboxylic acids, it also contains two rings, of which one is aromatic, making it susceptible to a wide array of reactions.

3.4.3.3 Carvacrol

Carvacrol, a phenolic monoterpenoid, is a bioactive compound that is predominantly found in aromatic herbs, especially oregano (*Origanum vulgare*) and thyme (*Thymus vulgaris*) and in the essential oils that are derived from these plants. Carvacrol have long been used in the food and cosmetics industry acting as an antioxidant and a fragrant, however a recent meta-study has shown that this molecule also exhibits potentially antimicrobial and anti-carcinogenic properties [32]. Depicted below in Figure 12 is the molecular structure of carvacrol.

Figure 12: The molecular structure of carvacrol found in aromatic herbs.

The carvacrol molecule is fairly small consisting of a phenol group with two carbon groups, of which one is branched in *para* configuration. The freely accessible hydroxyl group and the presence of a phenol moiety is what makes this compound exhibit antimicrobial properties. The mechanistic explanation being that it damages the cell membranes of the bacteria and inhibits some of the intracellular enzymatic functions [34]. Preclinical trials have also found that carvacrol seem to induce proapoptotic behavior of cancer cells from breast, liver, and lung carcinomas [33].

3.5 Methods

3.5.1 Recombinant protein production

Recombinant protein production is a biotechnological method where a gene of interest is cloned and produced in an expression host, allowing efficient and reliable production of a specific recombinant protein. The choice of vector and expression host is dependent on application, as different types of vectors and hosts offer different sets of advantages.

The first step of recombinant protein production is transforming host cells, most commonly *Escherichia coli,* with a plasmid vector or in some cases a virus molecule in which the gene of interest have been inserted. The ubiquitous usage of *E. coli* is due to its efficiency in terms of speed of growth, manipulability, and cost [34]. The vector also contains suitable promoters and regulatory sequences which allows the user to control the expression of the region of interest. The production is followed by protein purification, typically divided into (1) extracting through cell lysis, (2) precipitation or isolation based on differential solubilization, (3) centrifugation and (4) chromatographic methods [35]. There are a wide variety of methods that can be used to extract proteins, however, more gentle methods such as using detergents or sonication are most often preferred due to the sensitive nature of the protein's conformation. To facilitate more convenient separation, the recombinant proteins are often produced as fusion proteins with an added affinity tag which improves the solubility of the proteins and also allows for easier purification using affinity chromatography. A common tag that is used for recombinant production of proteins is a histidine tag, which is the separation method which will be used in this study, affinity chromatography with a column binding to the histidine-tag then allows convenient separation of the produced protein [36].

In certain applications a linker containing a cleavage site may be introduced between the affinity tag and the protein to allow for removal of the tag in the subsequent purification steps. In this study a linker has been introduced both based on this premise, but also since the tag itself may act as an inhibitor of the NS2B-NS3 protease and removal of this tag may yield an enzyme with higher productivity.

3.5.2 IMAC

Immobilized metal-ion affinity chromatography (IMAC) is a chromatographical method based on the tendency for interaction of certain peptide residues with specific metal-ions. The IMAC unit consists of a pump and a narrow column containing a solid support matrix, most often a specific type of magnetic nanoparticles or agarose gel. The binding portion of the column is filled with highly labile metal ions, most commonly Cu^{II}, Ni^{II} or Zn^{II} that interacts and retains proteins containing residues that have a tendency for metal-ion interactions [37]. Poly-histidine tags have a very strong tendency for metal-ion interactions thus enabling the user to separate recombinant proteins with an added histidine-tag from a mixture containing other residual proteins using specific columns [38]. The user can also utilize the system to achieve the reversed effect, namely, to separate out proteins not binding to the column.

3.5.3 SDS-PAGE

Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is one of the most widely applied techniques for separating proteins, confirming sample homogeneity, or assessing protein purity after purification with high resolution. The technique has been the golden standard since it was first introduced in the 1970s, with minimal changes made in comparison to how the technique is formulated today [39]. The SDS buffer denatures the proteins and provides a negative charge proportional to the size of the polypeptides in the mixture. An external electric field is then applied which forces the proteins to migrate through the gel towards the anode, the migrated distance is then

proportional to the size of the protein. The polyacrylamide gel used in SDS-PAGE is a transparent gel matrix with several wells where the sample mixtures are injected. A molecular weight ladder is added to one of the wells providing valuable information about the relationship between molecular weight and migration distance, allowing identification of the different species present in the mixture based on their molecular weight [39].

3.5.4 Mass Spectrometry (MS)

Mass spectrometry, or more commonly referred to as MS, is a technique used in analysis to identify compounds that are present in a sample based on their mass-to-charge ratio (m/z). Due to its efficacy and specificity this technique is widely used in research fields ranging from industrial production, hazard management, food, and pharmaceuticals along with others. A mass spectrometer identifies a compound by determination of molecular weight and analysis of the abundance of different naturally occurring isotopes; to allow for these measurements to take place the sample is first ionized into gaseous ions before subsequent separation and detection steps [40].

3.5.5 Nano differential scanning fluorimetry (DSF)

Nano differential scanning fluorimetry, or more commonly referred to as DSF, is a fast and convenient method for evaluating stability and conformation of proteins in a wide range of thermal environments. The technique can be expanded by using samples containing a matrix of screening conditions, e.g., varying pH range, buffer components, salt concentration, and can in this way establish what conditions yield the widest spectra of thermal stability. Theoretically, the method is based on measuring the fluorescence signal from hydrophobic residues that originally is protected when the protein is in its native folded state. As the temperature rises, the protein approaches its specific melting point and starts to unfold, exposing more of these hydrophobic residues that express a fluorescence signal [41].

3.5.5 AlphaFold

AlphaFold is a revolutionary technique based on artificial intelligence that was developed by DeepMind in 2018. Since its release it has helped scientists worldwide to predict structures of over 200 million proteins [42]. The predictions are based on previously determined structures, allowing an AI model to be trained on a vast dataset, identifying reoccurring physical and structural phenomena which offers the predictions a high level of validity [43]. The user simply enters an amino acid sequence and AlphaFold predicts the most probable configuration and generates a PDB model that can be used for further analysis.

3.5.6 UCSF Chimera

The structural study of the dengue derived protease was done using bioinformatic tools that allowed examination of the conformation of the residues constituting the protein. One such software is UCSF Chimera, which allows thorough analysis of the protein's conformation, including generating trajectories, density maps and performing sequence alignments [44]. This is done by loading a PDB file containing the atomic coordinates for the molecule in question (*available at RCSB PDB).* These models are based on

crystallographic structural determination of the proteins, which yield results of very high quality [45]. Molecular models can also be obtained by structural prediction using AlphaFold. The structure is then visualized in 3D-space where spatial measurements, bonding, mutations, and a plethora of other analytical tools can be used to evaluate the protein and its properties [44]. The software also makes it possible to generate high quality renders of proteins and molecules, making it suitable for producing illustrations for use in scientific publications.

3.5.7 YASARA and AutoDock VINA

Yet Another Scientific Artificial Reality Application or YASARA, is a molecular docking software that allows simulation of ligand-protein binding as well as visualization and modelling of molecules based on a plethora of different parameters. This allows the user to screen and evaluate binding dynamics before performing in-vitro studies, facilitating better efficiency and allocation of resources. YASARA uses the programming language Portable Vector Language (PVL), allowing visualization of large proteins. Since its release it has been cited in thousands of scientific articles, a testimony of its wide applicability in the bioinformatical field [46]. The previously mentioned UCSF Chimera is also a viable alternative to YASARA for performing molecular docking, however, screening of ligands is more efficient in YASARA as it allows multiple dockings as compared to UCSF Chimera [46]. The application used for docking simulations in UCSF Chimera is called AutoDock VINA and is also widely used due to its efficacy and rapidity.

3.5.8 Activity measurements

To assess the activity of the produced recombinant enzyme and to detect any inhibitory effects of the ligands included in this study, a chromogenic assay was developed using a chromogenic substrate. The substrate used in this study was *Nα-Benzoyl-D,L-arginine 4-nitroanilide hydrochloride* (DL-BAPNA), which is a chromogenic substrate used for assessing amidase, trypsin or balterobin activity of proteolytic enzymes. If in presence of an enzyme expressing any of these types of activity, the bond between the arginine and the p-nitroaniline moieties in DL-BAPNA is hydrolyzed, releasing the detectable chromophore p-nitroaniline which turns bright yellow and are detectable by spectrophotometry with an absorption maximum at 405 nm [47], [48].

4 Materials and methods

In this chapter, all materials and methods that have been used will be presented in depth. For additional information, please refer to the *Appendix*.

4.1 Materials

4.1.1 Strains of bacteria & plasmids

Two different bacterial strained were used during the work performed in the laboratory, these strains are listed in Table 1.

Table 1: List of the strains of bacteria used in the project.

Strain Name	Genotype	Purpose
DH5-α	Escherichia coli	Vector/plasmid
BL21(DE3)	Escherichia coli	Protein production

The plasmids used for production of polypeptides and the enzyme during the work performed in the laboratory is listed in Table 2.

Table 2: List of plasmids used for production of the recombinant NS2B-NS3 and the polypeptide.

Strain Name	Restriction Sites	Purpose
pET21b	*Nde1, Xho1* *Ampicillin resistant*	Production of polypeptide ligand
pET-28a	*Nde1, Xho1* *Kanamycin resistant*	Production of recombinant enzyme

4.1.2 Media

The media used for the different techniques as well as production of plasmids and proteins, as well as their composition and purpose are listed in Table 3.

Table 3: List of media used throughout the experimental part of the project.

Media Name	Composition	Purpose
LB-media	20 g of LB-powder per liter water.	To sustain a viable cell culture and provide the necessary nutrients for cell growth.
$CaCl_2$	1.665 g of $CaCl_2$ (150mM) in 100 mL millie-Q water.	Used in the washing of cells after growth. Preparing the cells for the coming transformation.
$CaCl_2$+Glycerol	1.665 g of $CaCl_2$ (150mM) in 75 mL millie-Q water and 25 mL glycerol (20%)	Used for storing the cells after washing.

4.1.3 Main disposable goods

- Eppendorf tubes
- Falcon tubes
- Amicon Concentration tubes
- Centrifugation tubes (300 mL, 100 mL, 5 mL)
- Syringes
- Needles
- Peleus pipettes

4.1.4 Main instruments

- ÄKTA Ready IMAC
- Bio-Rad SDS-PAGE
- Nano Drop
- NanoDSF Prometheus NT.48
- Ultra Centrifuge
- Shaking Incubator
- MultiSkan GO
- Bio-Rad GelDoc Go Imaging System

4.1.5 Software and online resources

The software and digital tools used for bioinformatic analysis are listed in Table 4.

Table 4: Lists of software used throughout the project.

Name	Purpose	Reference
AlphaFold	Prediction of the structure of peptides based on their sequence using artificial intelligence.	[42]
UCSF Chimera	Studying of the atomic models of the proteins (retrieved from PDB) and ligands, as well as docking simulations.	[44]
AutoDock VINA	Performing ligand-protein docking simulations to assess bioactivity of selected ligands.	[46]
PepSite2	Online tool for evaluating any potential bioactivity of a peptide sequence on a specific PDB structure.	[49]
AVPIden	Online tool for functional characterization and evaluation of antiviral peptides based on their sequence.	[50]

4.2 Bioinformatics

4.2.1 Structural prediction using AlphaFold

AlphaFold was run using the protein sequences for the viral protease and the peptide ligands using default settings to produce reliable PDB-files, for use in the subsequent bioinformatic studies. The following protocol was followed for structural prediction using AlphaFold.

1. Prepare the amino acid sequence for the protein that should be structurally predicted.
2. Input the amino acid sequence in the online workspace for AlphaFold.
3. Install necessary dependencies.
4. Run structural prediction.

5. Assess the preview of the 3D-structure for any obvious errors and evaluate the plots for establishing which model is the most statistically reliable.
6. Download the PDB-file.

4.2.2 Molecular modeling using UCSF Chimera

UCSF Chimera was used to open, model, and analyze the different PDB files of the NS2B-NS3 protease and the ligands used in this study. The numerous tools and options offered by the software allowed easy depiction of potential interactions, bonds, and structural conformations of the different molecules. The workflow in the software was split up into the following points.

1. Analyze the structure of the molecule.
2. Highlighting the different secondary structures.
3. Analyzing important parts of the molecules (*e.g. catalytic triad, allosteric site, polar/non-polar regions etc.*).
4. Export high-definition figures representing the main findings.

4.2.3 Docking simulations using AutoDock VINA in UCSF Chimera and YASARA

AutoDock Vina was used to analyze molecular dynamics and interactions between the selected ligands and the protein allowing deduction of potential binding-sites and providing a score corresponding to the AutoDock Vina is an addon to UCSF Chimera and is also the system implemented in YASARA for running docking simulations. Docking was performed on all ligands included in this study. The setup used generated 20 possible positions or sites of interactions for each ligand. The position exhibiting the highest score and the lowest RMSD (root-mean-square deviation) was deemed the most probable site of interaction. The workflow during docking was split up into the following points.

1. Load the PDB-structure for the protein and the ligand.
2. Set up the bounding-box for what site of the protein where docking should be simulated.
3. Set up any additional parameters.
4. Run the docking simulation.
5. Evaluate and export results.

4.2.4 Miscellaneous Online tools

To aid in selection of the potential inhibitory peptides for this study, two online tools were used. *PepSite2* and *AVPIden* was used in combination with comparison of published literature to assess and screen the viability of potential ligands for use in this study. The workflow using these tools were split into the following points.

1. Perform literature research of potential bioactive peptides.
2. Use PepSite2 to assess if the selected ligand would express any bioactivity towards the selected protein or class of enzyme.
3. Utilize AVPIden to assess if the peptide would interact with *Flaviviruses*.

4.3 Production of recombinant proteins

4.3.1 Preparative steps

The first steps consisted of preparing a working solution derived from the plasmid stock solution that were to be used in preparing the competent cells.

A set of 20 agar plates were also prepared for coming cultivations according to standard protocols, with media containing two types of antibiotics – kanamycin and ampicillin.

Before conducting any laboratory work all materials that would be used in proximity of the culture in the sterile hoods were autoclaved – this included medium, solutions and other materials that would come in contact with *E. coli* cells before, during or after growth.

To allow for production and inhibition studies of the polypeptide ligand, a plasmid containing the following sequence were ordered:

Quinoa derived construct

```
CATGTGGCGAGCGGCGCGGGCCCGTGGGAAAACCTGTATTTTCAGGGCCATGTGGCGAGC
GGCGCGGGCCCGTGGGAAAACCTGTATTTTCAGGGCCATGTGGCGAGCGGCGCGGGCCCG
TGGGAAAACCTGTATTTTCAGGGCCATGTGGCGAGCGGCGCGGGCCCGTGGGAAAACCTG
TATTTTCAGGGCCATGTGGCGAGCGGCGCGGGCCCGTGGGAAAACCTGTATTTTCAGGGC
CATGTGGCGAGCGGCGCGGGCCCGTGGGAAAACCTGTATTTTCAGGGCCATGTGGCGAGC
GGCGCGGGCCCGTGGGAAAACCTGTATTTTCAGGGCCATGTGGCGAGCGGCGCGGGCCCG
TGGGAAAACCTGTATTTTCAGGGC
```

The other chemical compounds that were intended for study, i.e. carvacrol and chlorogenic acid, were ordered from Sigma Aldrich.

4.3.2 Preparation of competent cells

Preparation of competent cells was performed in a batch of twenty Eppendorf tubes per strain, for two different strains. The general protocol followed for preparing the cells were as follows.

Note: *It is of high importance to keep a sterile environment when working with live cell cultures. All solutions need to be autoclaved in advance and all steps take place in a sterile bench. From step 3 it is crucial to keep the temperature of the solutions between 0 °C and 4 °C, use an ice bath at all times.*

1. From a stock solution containing the strain of interest, transfer 150 µL to a falcon tube with 3 mL of sterile LB media and leave overnight at 37 °C.
2. Transfer the content of the falcon tube to a shake flask containing 30 mL of sterile LB and keep at 37 °C shaking. Perform regular OD measurements at a λ_{600} until it reaches 0.3-0.4.
3. When adequate OD has been reached, transfer the content into a falcon tube and incubate on ice for 10 min. Then centrifuge for 10 min at 4500 rpm at 4°C and discard the supernatant.
   Note: *If only one strain is to be prepared, divide the contents into two separate falcon tubes to facilitate centrifugation.*
4. **First wash:** Add 7.5 mL of 150 mM $CaCl_2$ (sterile, pre-cooled in ice) and resuspend the pellets by pipetting slowly.
5. Centrifuge the solution for 10 min at 4500 rpm at 4 °C. Discard the supernatant carefully.
6. **Second wash:** Resuspend the pellet in a total volume of 3 mL using a solution containing 150 mM $CaCl_2$ and 20% glycerol (sterile, pre-cooled in ice).
7. From the final 3 mL volume, transfer aliquots of 150 µL to Eppendorf tubes, label, and store at -80 °C until use.

4.3.3 Transformation of competent cells

Transformation was carried out to insert the plasmid containing the gene into the host cell. The general protocol followed for transformation was as follows.

Note: *All steps except for the heating and cooling involved in heat shocking are done in sterile conditions, preferably in a sterile bench.*

1. Thaw two Eppendorf tubes containing a stock solution with the DH5-α and the BL21(DE3), stored in -80 °C.
2. Add 2 µL of the working solution containing the plasmid to each of the E. coli strains, incubate on ice for 15 min.
3. **Heatshock:** Put the Eppendorf tubes in a heat block at 42 °C for 2 min. Then incubate on ice for 10 min.
4. Add 300 µL sterile LB media to each tube, slowly pipet up and down 1-2 times to facilitate dispersion. Then incubate at 37 °C for 45 (shaking) or 60 (static) min.
5. After incubation, add 100 µL of each strain to two separate agar plates containing solid LB + added ampicillin for selection purposes.
6. Spread the droplet laterally first by tilting the plate, and then gently across the entire plate using a cell spreader.
7. Label the plates and incubate at 37 °C overnight.
8. Store the final grown culture in a cold room at 4 °C.

4.3.3.1 Troubleshooting unsuccessful transformations of the competent cells

As there were some difficulties with achieving a successful transformation using the NS2B-NS3 plasmids protein gels were run to confirm the plasmid-integrity before continuing with subsequent analysis. The protein gels were prepared as follows.

1. Prepare an agarose protein gel using the provided gel cast.
2. Let the gel set for about 45-60 min.
3. Remove the gel from the cast and load it in the rack and place it in the electrophoresis container.
4. Fill the container with loading buffer, covering the gel entirely.
5. Prepare samples containing stained loading buffer and sample in a 1 to 6 ratio.
6. Run the gel at 200 V for about 45-60 min.
7. Stain the gel for about 1 h.
8. Analyze the gel using an UV light source.

4.3.4 Production of recombinant proteins

4.3.4.1 NS2B-NS3

The protein production was carried out to allow for the transformed cells to produce the recombinant protein of interest. The general protocol followed for protein production was as follows.

Note: *Steps 1-5 are done in a sterile environment.*

1. Prepare an inoculum by adding 3 mL sterile LB media together with 3 µl of antibiotic (Kanamycin for NS2B-NS3) to a falcon tube.
2. Using a 5 µL inoculation loop, scoop up a selected colony and mix it with the media in the falcon tube.
3. Incubate overnight at 37 °C shaking at 150-200 rpm.
4. Transfer the inoculum to an Erlenmeyer flask containing 300 mL sterile LB media and 300 µL antibiotics.
5. Incubate at 37 °C shaking at 150-200 rpm and measure the OD_{600} until it reaches a value of 0.6-1.0.
6. Add 300 µL IPTG and leave the flasks at room temperature (or 20 °C) on a shaker overnight.
7. Transfer the content from the Erlenmeyer flask to a tube used for centrifugation. Run the centrifugation at 8000 rpm for 10 min.
   Note: *Remember to balance the tubes according to current instructions.*
8. Resuspend the pellet in 15 mL binding buffer and transfer the sample to a falcon tube.

9. Sonicate the content on ice for 15 min and transfer the content to an ultracentrifugation tube.
10. Centrifugate for 20 min at 12 000 rpm.
    Note: *Remember to balance the tubes according to current instructions.*
11. Collect the supernatant for IMAC purification.

4.3.4.2 Polypeptide ligand

The peptide ligand will be produced as a polypeptide which later will be cleaved using the TEV protease. The production of the polypeptide ligand followed the same protocol as when producing the NS2B-NS3, but with the added initial step of digesting the polypeptide with TEV protease to disconnect the peptide moieties from the polypeptide construct.

4.3.5 Protein purification – IMAC

Purification of the produced protein was performed using the ÄKTA IMAC. Before each run, the correct column was installed, in this case a HisTrap and the hose connecting the pump to the columns were prepared. The system was washed with millie-Q water before beginning the purification process. The sample containing the his-tag would then be trapped in the column, before being eluted into the sample collection tubes. The fractions of sample that did not get trapped in the column would be disposed in the specified waste container. Each purification was set to fraction the sample in 7 different collection tubes, of which 1 or 2 contained the final purified protein. The following protocol was used for running the purification using the ÄKTA IMAC:

1. Wash the system and allow the UV-signal to stabilize.
2. Prepare tubes for collection of sample fractions.
3. Connect the sample container to the inlet.
4. Run the purification.

The same general method and protocol was used when purifying the proteins after removal of the his-tag, but the tube that previously discarded the waste was now instead led into a sample tube allowing collection of the molecules that did not have an added his-tag.

4.3.6 Protein analysis – SDS-PAGE

For verification and analysis of the purification an SDS-PAGE was run. Pre-stained gels were used in combination with an unstained *ProteinLadder* (REF) to assess the size of the samples that were analyzed. The general protocol for SDS-PAGE was as follows:

1. Load the gel into the holder of the electrophoresis rack.
2. Fill the container with buffer, covering the gel and the wells entirely.
3. Prepare samples containing stained loading buffer and sample in a 1 to 6 ratio.
4. Load the samples and the weight ladder to the wells.
   NOTE*: It is convenient to prepare a schematic drawing of what samples should be*

loaded in the different wells. This also facilitates easier interpretation of the gels later.

5. Run the gel electrophoresis at 180 V for 15 min, followed by 200 V for ~45 min.
6. Analyze the gel using an UV light source.

4.3.7 Digestion using the TEV protease

An enzymatic digestion using TEV protease was carried out to remove the histidine tag from the recombinant enzyme and to cleave the produced polypeptide. The protocol followed were the same protocol provided with the commercial TEV protease kit from ThermoFisher [51]. Both the commercial TEV protease and the recombinant TEV protease were used in the same concentration throughout the experiment.

***NOTE**: The protocol used for TEV protease digestion was scaled up or down depending on the amount of fusion protein that were to be digested.*

1. Add the following to a microcentrifuge tube:

Fusion Protein	20 µg
20X TEV Buffer	7.5 µL
0.1 M DTT	1.5 µL
AcTEV Protease	1.0 µL
Water	to 150 µL

2. Incubate at 4 °C overnight.
3. Purify the sample using IMAC to isolate the protein moiety from the cleaved histidine tag.

4.3.8 Protein concentration

AMICON Ultra 10kDa centrifugal filter tubes were used to concentrate the final protein solution to obtain a more optimal working stock for the subsequent experiments and analysis. The general protocol for protein concentration were as follows.

1. Measure the sample concentration using NanoDrop.
2. Transfer sample content to the AMICON centrifugation tube.
3. Centrifugate the sample at 4500 rpm for 10 min.
4. Collect the retentate, i.e., the top layer.
5. Measure the concentration using NanoDrop to determine the new concentration of the concentrated sample. Label and store accordingly.

4.3.9 Analysis by mass spectroscopy (MS)

Samples were analyzed by mass spectroscopy to get more in-depth structural data of the produced peptides, and to confirm if cleavage of the his-tag fragment had been successful. The MS analysis was carried out by an associate working at the university, the samples sent for analysis were prepared as follows:

1. Collect sample in aliquots of 100 µL and store in the fridge at 4 °C.
2. Send samples for analysis.

The results were delivered as a report consisting of spectra of different ranges and thorough analysis of the most abundant fragments.

4.4 Inhibition studies

The studying of the inhibitory effects of the selected ligands were assessed and evaluated using a colorimetric microplate assay with a selected chromophore, in this case *Nα-Benzoyl-D, L-arginine 4-nitroanilide hydrochloride* or *DL-BAPNA*.

4.4.1 Initial Activity measurements

Initial activity measurements were carried out using a microplate assay that were analyzed in the MultiSkan Go. The MultiSkan Go were set to kinetic loop measuring the absorbance at 405 nm every 60 s for the set amount of time.

The general protocol for preparing the assay solutions in the microplates were as follows:

1. Add the components to the microplate wells in the following order:
    a. Water
    b. Buffer
    c. Enzyme
    d. Substrate

4.4.2 Assay development for activity measurements

The assay conditions had to be optimized and tailored for this specific application to provide an increased activity and more accurate and repeatable results. The parameters that were adjusted to allow for optimization were the buffer type, additives, buffer concentration, the reaction temperature, the enzyme concentration, the substrate concentration, and the run time of the experiment.

The following assay conditions were evaluated.

- TRIS-HCL 50 mM & 100 mM, pH 7 at 30 °C and 37 °C.
- TRIS-HCL 50 mM & 100 mM, pH 8.4 at 30 °C and 37 °C.
- TRIS-HCL 50 mM & 100 mM, pH 10 at 30 °C and 37 °C.
- Phosphate buffer 50 mM & 100 mM, pH 7 at 30 °C and 37 °C.
- Phosphate buffer 50 mM & 100 mM, pH 8.4 at 30 °C and 37 °C.

During the optimization of the protocol the concentration of substrate was kept at 1 mM for all runs, while the enzyme concentration was varied from 0.01 mg/mL to 0.1 mg/mL in increments of 0.02 mg/mL.

The final assay for determining the kinetics of the enzymatic reaction were run after the initial optimization of the reaction conditions. When running the kinetic assay the substrate concentration was varied from 10 µM to 2 mM.

***NOTE**: All measurements were made in duplicates to increase the reliability of the results.*

The MultiSkan Go was set to kinetic loop, measuring the absorption at 405 nm at intervals of 60 s for 240 min.

The general protocol for preparing the assay solutions in the microplates were as follows:

2. Add the components to the microplate wells in the following order:
    a. Water
    b. Buffer
    c. Enzyme
    d. Substrate

To further optimize the activity of the enzyme, the following compounds were also used to screen for the optimal assay conditions:

a. Chaps 1mM
b. NaCl 10mM
c. Glycerol 20% v/v

These additives were included in the assay used for the final activity and inhibition assays.

4.4.3 Inhibition assay

The inhibitions assays were carried out using the optimized conditions determined during the development of the assay. All measurements were made in duplicates to increase the reliability of the results.

The concentration of ligand was varied from 0 to 10 µM while keeping the concentration of enzyme and substrate constant. Four evenly distributed substrate concentrations were run - 0.25 mM, 0.5 mM, 1 mM, and 1.5 mM to allow for determination of the kinetic variables of the inhibited reaction. The MultiSkan Go was set to kinetic loop, measuring the absorption at 405 nm at intervals of 60 s for the set amount of time.

The final optimized protocol for preparing the assay solutions in the microplates were as follows:

2. Add the components to the microplate well in the following order
    a. Tris-HCl Buffer at pH 8.4 50 mM
    b. Chaps 1 mM
    c. NaCl 10 mM
    d. Glycerol 20% v/v
    e. Ligand/Inhibitor
    f. Enzyme 5.7 µM
    g. Water to 200 µL
    h. Substrate

The solutions were incubated at 37 °C for 30 min before introducing the substrate. The assay was run at 37 °C for around 4 h.

4.4.4 Data analysis

All basic plotting, linear regression, and datasheet formatting were carried out using the appropriate tools in Microsoft Excel. Any advanced statistical tools or more demanding analysis methods were carried out in MATLAB.

5 Results

5.1 Prediction of the structural model of the recombinant NS2B-NS3

The NS2B-NS3 used in this study consisted of the two enzyme moieties – NS2B and NS3 - bound with a glycine linker, hence, the structure of this construct was unknown and contained two regions that have yet to be determined by x-ray crystallography. AlphaFold was therefore used to predict the structure of the enzyme to allow for a better understanding of the conformation and function of the enzyme. Figure 13 shows the predicted structure overlaid with the experimentally determined structure.

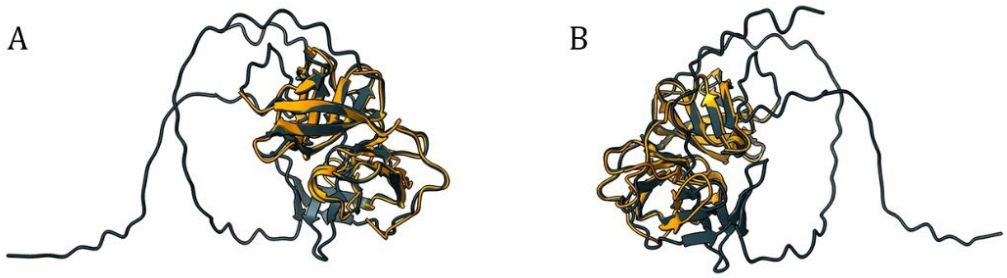

Figure 13: The anterior (A) and posterior (B) view of the predicted and crystallographic tertiary structure of the NS2B-NS3. The predicted structure is depicted in petrol blue and the experimentally determined structure in tangerine.

From Figure 13 it is evident that the AlphaFold prediction overlapped with the experimentally determined structure to a very high extent. A comparison between the prediction and the missing structure of the two regions that had not been determined by crystallography can be seen in Figure 14 below.

Figure 14: The missing (tangerine) and predicted structure (petrol blue) of the experimentally undetermined regions of the NS2B-NS3. Figure 14-A shows a predicted anti-parallel beta sheet found in the NS2B moiety, Figure 14-B shows the predicted structure of the glycine linker which binds the NS2B to the NS3.

Figure 14-A elucidates that one of the regions that were previously unknown consisted of a small coil with two antiparallel beta strands. The large strand in Figure 14-B containing the glycine linker consists of a longer coil with two separate beta strands in one of the ends. The final predicted model used for docking simulations and structural analysis can be seen in Figure 15 below.

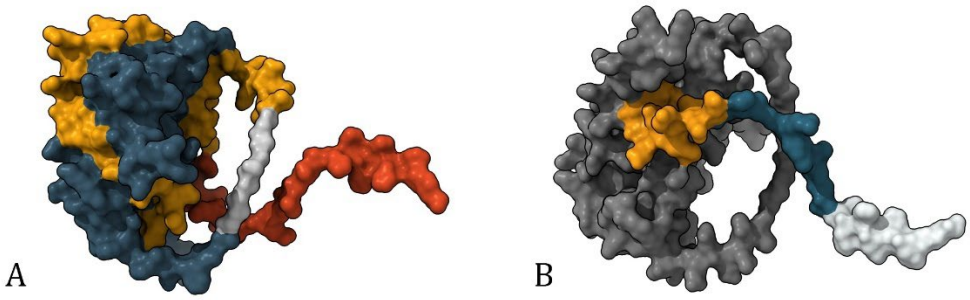

Figure 15: The surface model of the final predicted structure that was used for bioinformatic analysis and docking simulations. Figure 15-A shows the NS2B moiety depicted in petrol blue, the NS3 moiety in tangerine, the glycine linker in white and the added his-tag in orange. Figure 15-B shows the structure of the his-tag that was removed through enzymatic digestion, the cleavage site (ENLYFQG/S) is depicted in tangerine, the linker sequence in petrol blue and the 6x-his-tag in white.

Figure 15 clearly depicts the essential parts of the enzyme, as well as the cleaving site which were used to experimentally remove the histidine tag through TEV protease digestion before performing the subsequent activity measurements. The theoretical molecular weights of the different constructs (Figure 15-A) that were used in calculations were: NS2B-NS3 at ~26.4 kD and the his-tag and linker at ~3.7 kD.

5.2 Recombinant protein production

5.2.1 NS2B-NS3

5.2.1 Competent cell preparation

To assure that the culture had entered the exponential growth phase in the production of the competent cells', absorbance measurements were performed, as seen in Table 5. A value of about 0.3-0.4 was deemed optimal before continuing on with the subsequent steps of the protocol.

Table 5: OD measurements versus time taken when preparing competent cells of the two strains BL21(DE3) and DH5-α. The growth was terminated when the OD reached 0.3-0.4.

	T_0 (08:30)	**T_1 (09:10)**	**T_2 (09:40)**
BL21(DE3)	0.249	0.261	0.431
DH5-α	0.242	0.288	0.335

All measurements were performed in duplicates to assure a replicable and reliable value. The high initial value and the irregular trend seen between T_0 and T_1 is explained by an error in the protocol leading to a higher initial optical density, the solutions were diluted by a factor of 2 between T_0 and T_1 to assure adequate growth and optimal OD before proceeding.

Unfortunately, it became evident that the prepared competent cells were not viable, the subsequent transformation and production were therefore carried out with a commercial strain of competent BL21(DE3) provided by a division member.

5.2.2 Plating of the transformed cells

After introducing the plasmids into the cells, the strains were cultured on agar plates containing kanamycin. Figure 16 shows the colonies of the successfully transformed cells.

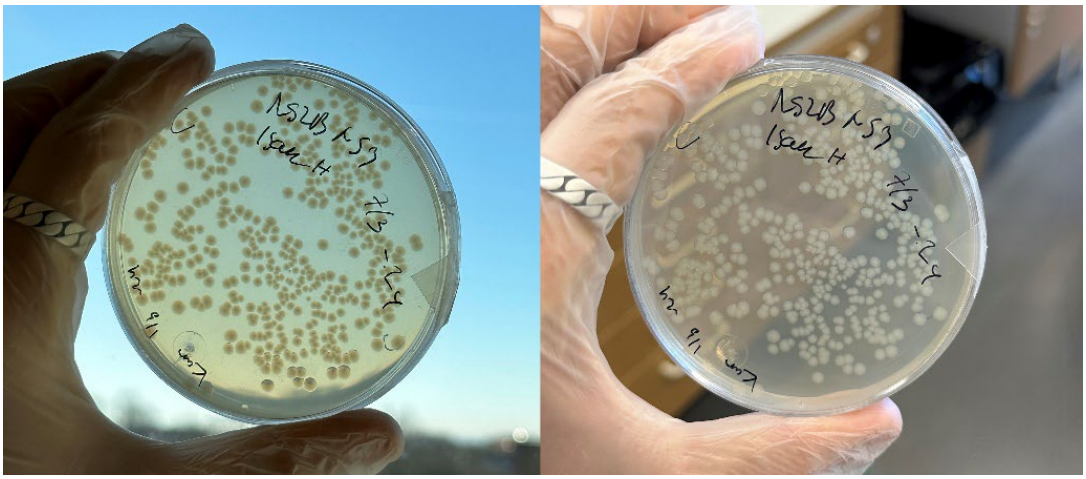

Figure 16: The successful plating of the transformed BL21(DE3) culture. These were the colonies used for production throughout the entirety of the project.

5.2.3 Protein production

The production of the recombinant NS2B-NS3 yielded a high amount of protein ranging from 3 to 4.5 mg/mL throughout the project. The concentration was determined using the *NanoDrop* spectrophotometer.

5.2.4 Protein purification and concentration measurements

SDS-PAGE was used to assure that the intended protein had been produced and to confirm the integrity of the protein. For better efficiency, a series of TEV protease digestions were also prepared to assess what conditions was optimal for later digestions. Both recombinant TEV protease and commercial TEV protease was used to compare and assess whether recombinant TEV protease could be used in later digestions, the results can be seen in Figure 17. The recombinant TEV protease was provided by a fellow master student – Fisnik Nerjovaj.

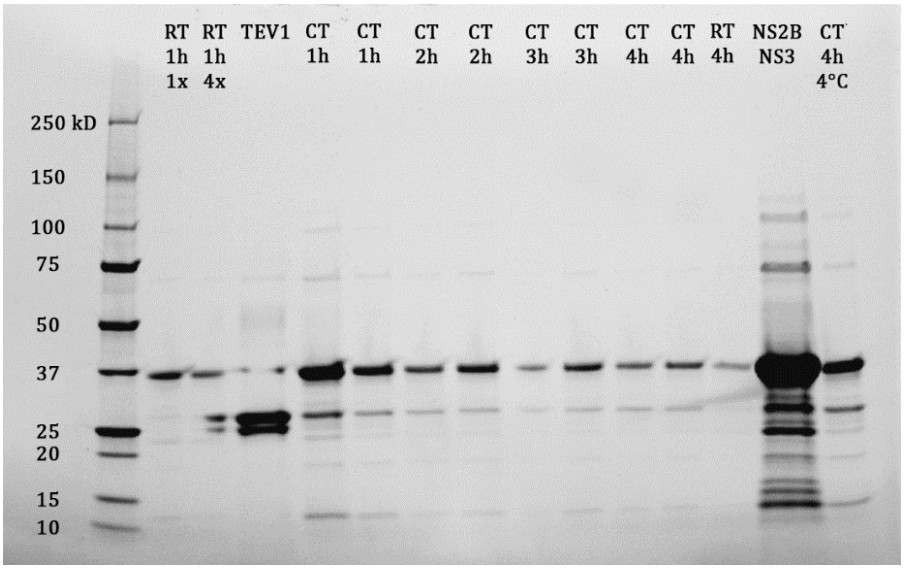

Figure 17: The SDS-PAGE gel showing the optimization of the digestion using the commercial TEV protease (CT 1-4h, CT 4h 4°C), trial digestions with the recombinant TEV protease at two different concentrations (RT 1h 1x, RT 1h 4x), purified recombinant TEV protease (TEV1) as well as a reference run of the crude NS2B-NS3 extract (NS2B-NS3).

As seen in Figure 17, the gel showed that the NS2B-NS3 had been produced (~35 kD). Due to limitations in the amount of wells a purified and diluted NS2B-NS3 sample was not run until later. This gel also showed that the optimal conditions for the enzymatic digestion using the commercial TEV protease was at 4°C for four hours, yielding a stronger band at ~35kDa, i.e., less degradation of NS2B-NS3. To further optimize the TEV reaction, a second batch of digestions were made but at 4°C overnight using both the commercial and the recombinant TEV protease (Figure 18). Before preparing the overnight digestions, the concentrations of the commercial and recombinant TEV protease were measured to make the reaction conditions identical. The commercial TEV protease used in this project had a concentration of 7.2 mg/mL.

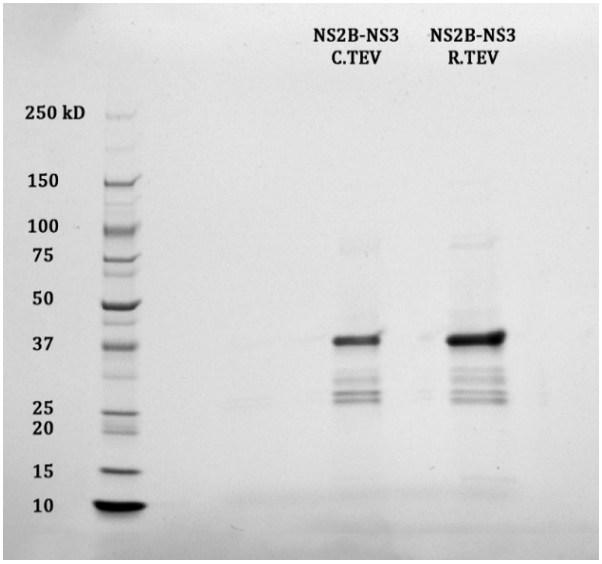

Figure 18: The SDS-PAGE gel comparing the enzymatic digestion of the NS2B-NS3 using the commercially available TEV protease (C.TEV) and the recombinantly produced TEV protease (R.TEV).

This showed that the recombinant TEV protease seemed to better preserve the integrity of the NS2B-NS3 compared to the commercial TEV protease. The cleaved fragment is not visible in the SDS-PAGE gel as the molecular weight of the tag and linker was ~3.5 kD. To confirm this conclusion the experiment was later repeated which yielded the same result. Using the now optimized protocol, a final gel was run using a negative control, a sample after digestion and a sample after digestion and purification (Figure 19).

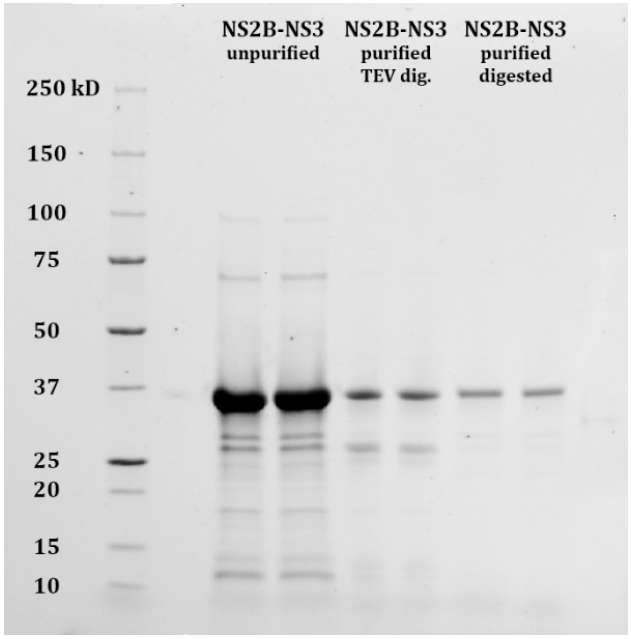

Figure 19: The SDS-PAGE gel showing the unpurified crude enzyme solution (unpurified), the reaction mixture for the enzymatic digestion containing both NS2B-NS3 and recombinant TEV protease (purified TEV dig.) and the final purified NS2B-NS3 without the his-tag (purified digested).

This gel clearly depicted that the optimized TEV protease digestion protocol yielded a sample containing isolated and intact NS2B-NS3 with minimal degradation.

The final production, purification and digestion scheme can be seen in Table 6 below.

Table 6: The purification and digestion scheme showing the amount of protein produced as well as the final yield after purification (IMAC) and digestion with TEV protease.

	Total Volume (mL)	**Concentration (mg/mL)**	**Total Protein (mg)**	**Yield (%)**
Undigested Enzyme	30	4.42	132.60	100
Final Digested Enzyme	4.50	1.52	6.84	5.20

5.2.4 Analysis by mass spectrometry (MS)

To better understand if the TEV protease digestion successfully removed the his-tag from the NS2B-NS3 domain, the samples used in Figure 18 was sent for analysis by mass spectroscopy.

From the spectra, there were some peaks that aligned well with the potential fragments of interest, however, as the sample were digested with trypsin, it was difficult to assess whether the tag remained in the sample or not. The spectra did, however, confirm that the NS2B-NS3 was present in both samples.

A BLAST and Mascot search was carried out to summarize and conclude the analysis of the MS data. The results showed that fragments were found in all parts of the protein except the part containing the 6xHis-tag (Figure 20*)*.

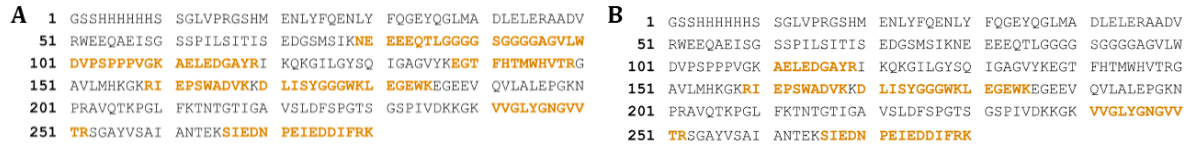

Figure 20: The identified fragments from the BLAST and Mascot search that was made based on the MS analysis of the digested and purified NS2B-NS3. Figure 20-A shows a 37% coverage in the sample digested with the recombinant TEV protease while Figure 20-B shows a 22% coverage that was found in the sample digested with the commercial TEV protease. Neither digestion resulted in identification of fragments matching the sequence of the his-tag.

A 37% coverage was found in the sample containing the digestion with the recombinant TEV protease (Figure 20-A), and a 22% coverage was found in the sample containing the digestion with the commercial TEV protease (Figure 20-B*)*.

5.2.5 Analysis by nano-DSF

Samples containing the NS2B-NS3 without his-tag were run in a *nano-differential-scanning-fluorometer* or nano DSF to assess the melting point of the enzyme in different buffer conditions. The enzyme was exposed to a temperature gradient ranging from 20 to 95 °C in Tris-HCl titrated to pH 5 through 10, the melting points were determined through fluorescence measurements of hydrophobic and hydrophilic residues at 350 nm and 330 nm. Figure 21 shows the melting points of the NS2B-NS3 plotted versus the pH of the buffer.

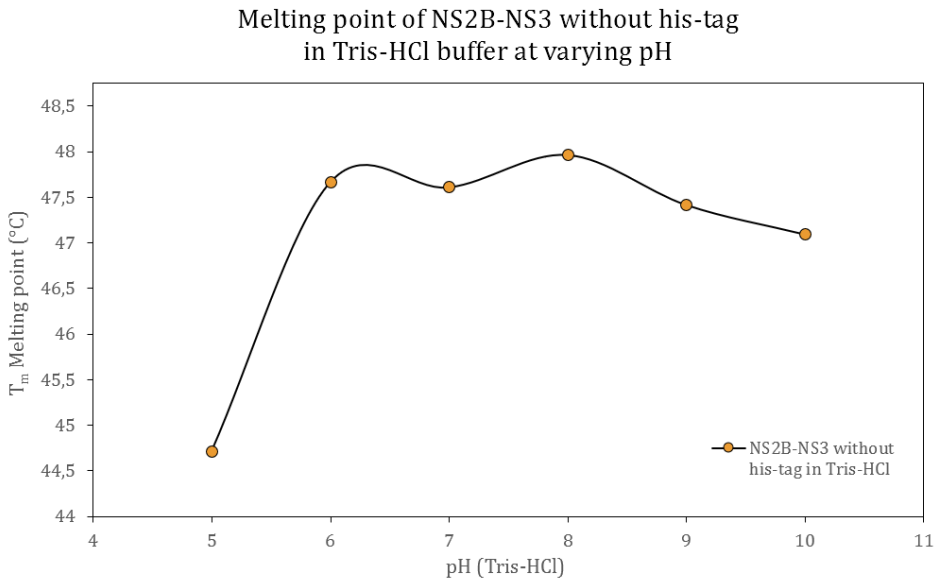

Figure 21: Plot of the different melting points (°C) of the NS2B-NS3 without his-tag at six different pH values (Tris-HCl buffer). The melting point of the enzyme reaches a maximum at around pH 8 (T_m = 47.7 °C) and a minimum at pH 5 (T_m = 44.6 °C). All capillaries were prepared and measured in triplicates.

The NS2B-NS3 without his-tag was stable at pH 6 through 10, with a maximum at around pH 8 (T_m = 47.7 °C) and a minimum at pH 5 (T_m = 44.6 °C) (Figure 21). A comparison between the ratio of unfolded and folded enzyme at low pH (pH 5) and high pH (pH 10) can be seen in Figure 22.

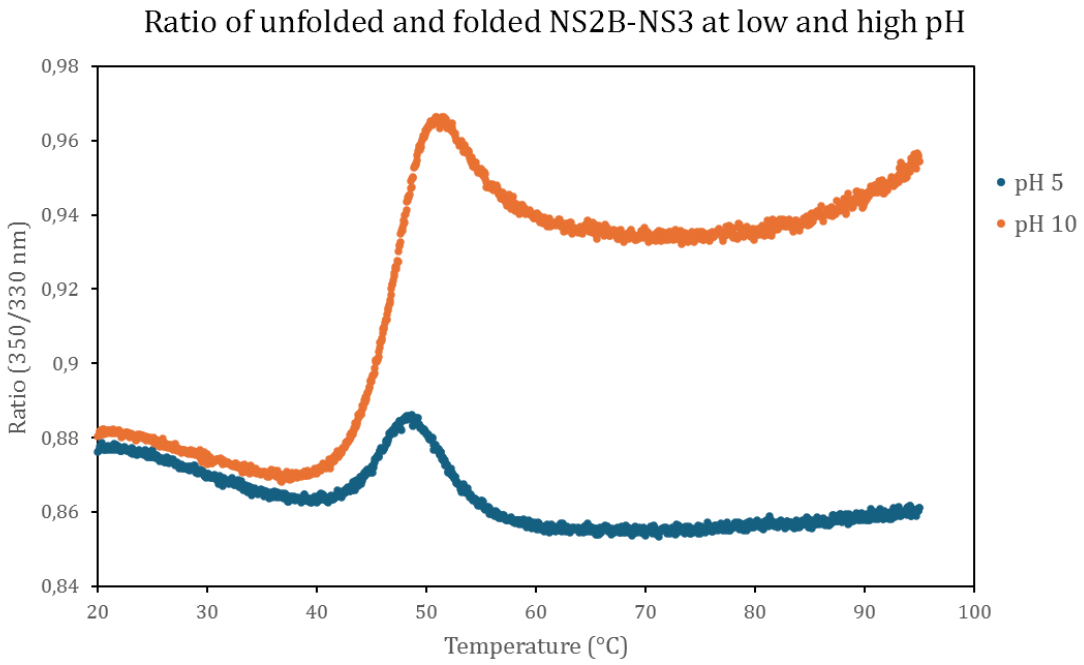

Figure 22: Plot of the ratio of native (330 nm) and unfolded (350 nm) NS2B-NS3 at pH 5 and pH 10 at different temperatures. The enzyme remains folded at a wider range of temperatures at lower pH compared to higher pH values.

This comparison revealed that the NS2B-NS3 without his-tag exhibited a refolding behavior at low pH when exceeding the melting point at 44.6 °C. The same behavior was observed at high pH but no to the same extent.

5.2.2 The quinoa derived peptide

The inhibitor derived from the quinoa plant was unfortunately not stable as a soluble polypeptide and were therefore not successfully produced. The peptide formed aggregates making it insoluble and inseparable from the cell pellet during the final centrifugation steps. Attempts of protein refolding were carried out using urea, but these were unsuccessful. Figure 23 shows the SDS-PAGE gel of the presence of the insoluble polypeptide in the cell debris pellet.

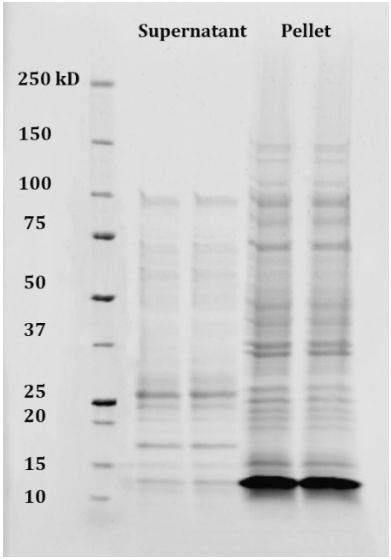

Figure 23: The SDS-page gel showing the supernatant and the cell pellet from the production of the polypeptide derived from quinoa. The sharp band at ~13 kD indicates the presence of the produced insoluble polypeptide.

Subsequent inhibition studies were therefore carried out solely using carvacrol and chlorogenic acid due to time constraints.

5.3 Activity measurements

Figure 24 shows the activity measurements comparing the native NS2B-NS3 with his-tag and the digested NS2B-NS3 without his-tag in the presence of the DL-BAPNA substrate.

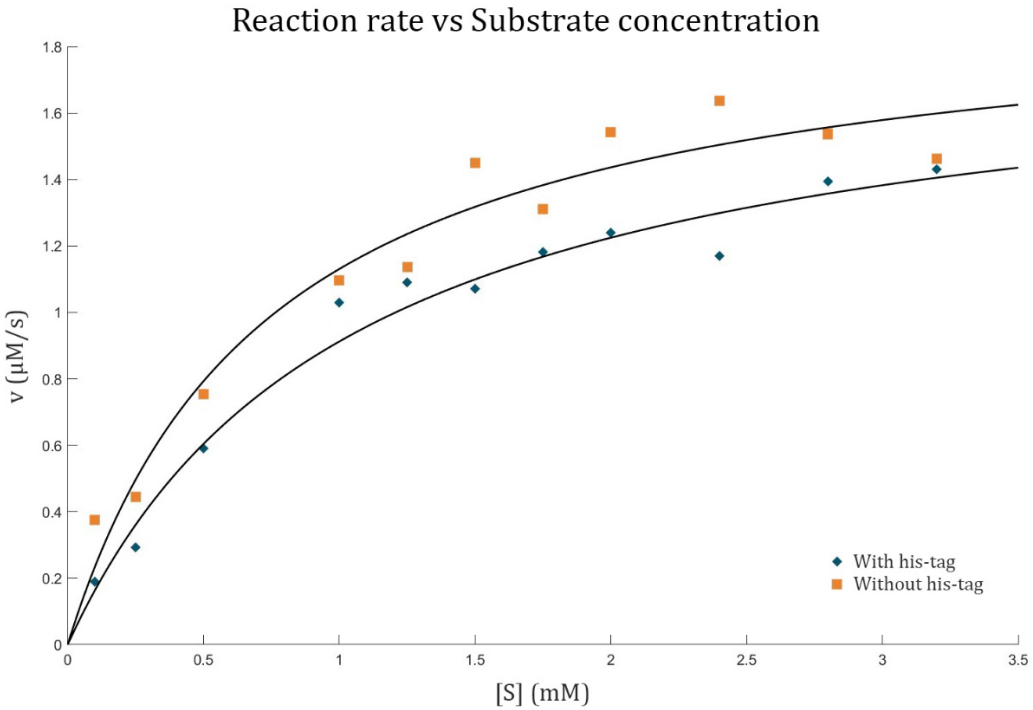

Figure 24: The Michaelis-Menten plot of the data collected from the colorimetric assay using DL-BAPNA comparing the activity of the NS2B-NS3 with his-tag (petrol blue) and the NS2B-NS3 without his-tag (tangerine). The black lines shows the fitted M-M equation that was generated using non-linear regression in MATLAB. The assay was carried out as described in the methods section. The concentration of NS2B-NS3 with and without his-tag was 5.7 µM. The concentration of substrate was varied between 0 and 3.2 mM.

The fitted Michaelis-Menten equation clearly shows that the NS2B-NS3 without his-tag exhibited greater activity compared to the NS2B-NS3 with his-tag at all assayed substrate concentrations. To confirm the validity of these results a paired double-sided t-test was carried out using the data from the triplicate measurements at all substrate concentrations, the results can be seen in Figure 25.

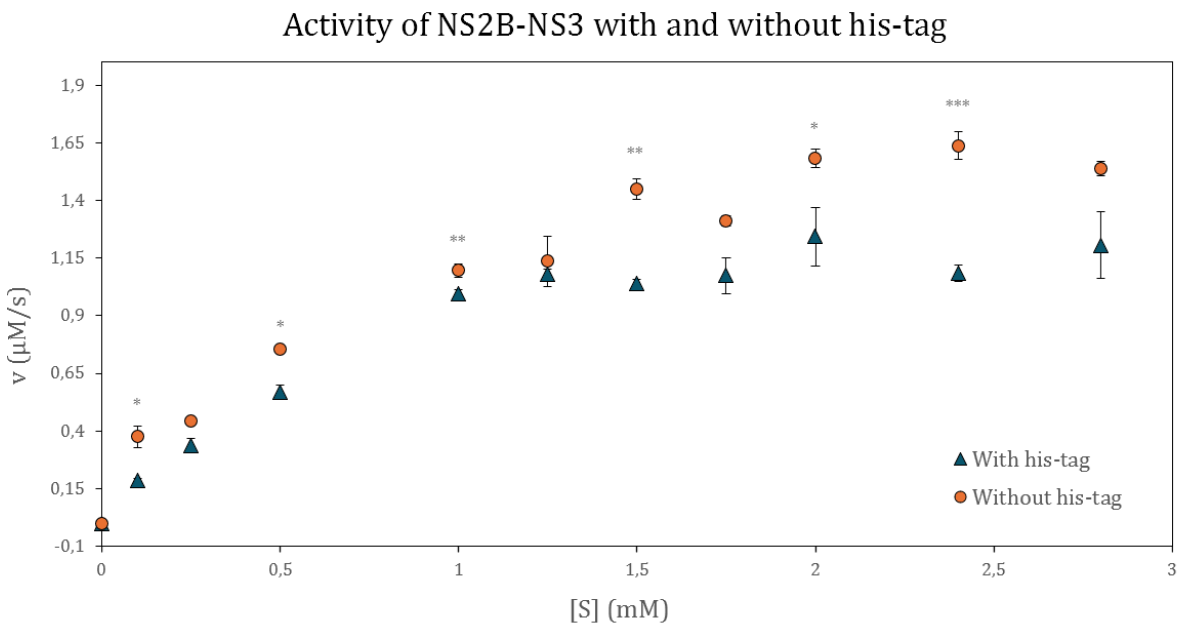

*Figure 25: The data-points with added significance test results for the colorimetric activity measurements of the NS2B-NS3 with his-tag (petrol blue) and the NS2B-NS3 without his-tag (tangerine). A paired double-sided t-test was carried out on the replications of the measurements to test for significance. The null-hypothesis was rejected at the following levels: * = p<0.1, ** = p<0.05, *** = p<0.01.*

The majority of the datapoints were significantly different but at varying confidence. A t-test was later carried out on the full dataset comparing the average activity of the enzyme over all substrate concentrations, yielding a p-value of 0.003, indicating that the null hypothesis could be confidently rejected. To determine the kinetic constants of the enzyme a Lineweaver-Burk plot was fitted to the inverse of the rate and substrate concentrations (Figure 26).

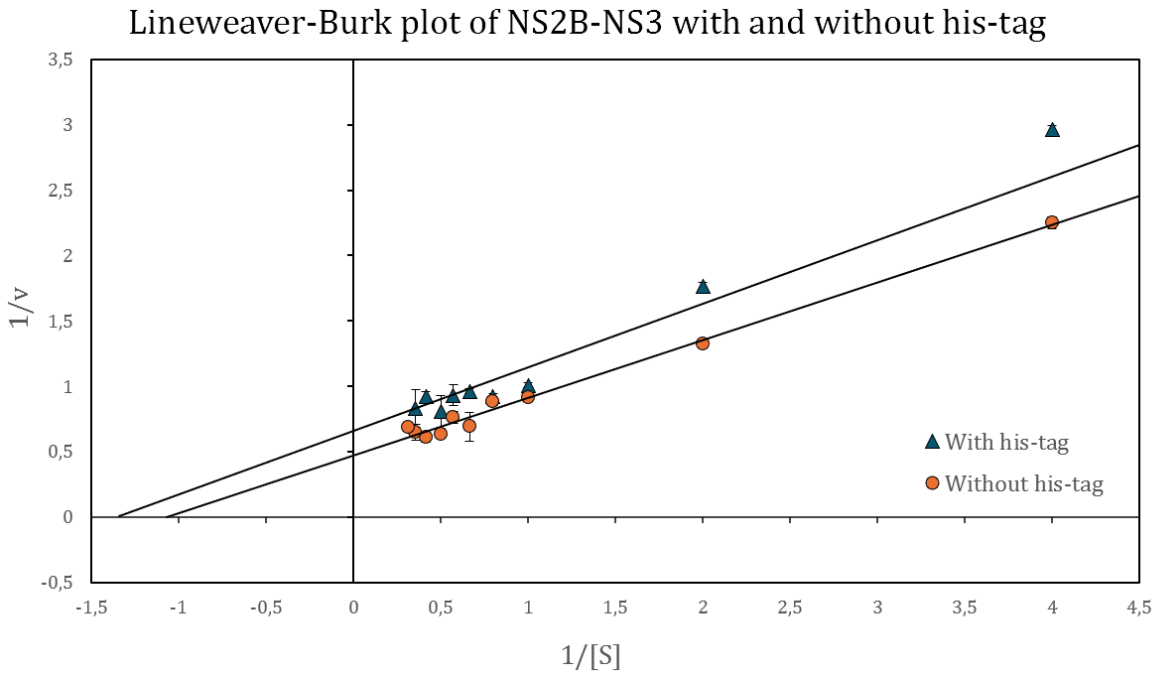

Figure 26: Lineweaver-Burk plot of the colorimetric activity measurements of the NS2B-NS3 with his-tag (petrol blue) and the NS2B-NS3 without his-tag (tangerine). The equations for the fitted lines was as follows: y =0.4863x + 0.6613, R^2=0.99 (With his-tag) and y = 0.4419x + 0.4702, R^2=0.99 (Without his-tag). The calculated V_{max}, K_m and K_{cat} values can be seen in Table 7.

The Lineweaver-Burk plot showed that the NS2B-NS3 without the his-tag had a higher V_{max}, K_m and k_{cat} value than the native NS2B-NS3 with the added his-tag. The calculated kinetic constants for the assay using DL-BAPNA as substrate can be seen in Table 7.

Table 7: The kinetic constants V_{max}, K_m and k_{cat} calculated from the Lineweaver-Burk plot of the activity measurements of the NS2B-NS3 with his-tag and the NS2B-NS3 without his-tag. The k_{cat} values were calculated assuming a 100% activity of the enzyme and using the theoretical molecular weights determined using ProtParam (NS2B-NS3 with his-tag – 30 175 Da and NS2B-NS3 without his-tag – 26 422 Da).

	V_{max} (µM/s)	**K_m (mM)**	**k_{cat} (s^{-1})**	**k_{cat}/K_m (mM^{-1} s^{-1})**
With his-tag	1.512 ± 0.103	0.735 ± 0.015	0.304 ± 0.021	0.414 ± 0.019
Without his-tag	2.127 ± 0.104	0.940 ± 0.025	0.375 ± 0.018	0.399 ± 0.008
Relative change	~141%	~128%	~123%	~96%

There was a larger difference in the calculated V_{max} values of the tagged and untagged enzyme compared to the K_m and k_{cat} values. However, there was no major difference in the *catalytic efficiency* (k_{cat}/K_m) of the tagged and untagged NS2B-NS3, meaning that the two enzymes perform quite similarly at low substrate concentrations.

5.4 Docking simulations

The interactions between the NS2B-NS3 and the ligands were also studied through in-silico docking simulations to further elucidate how and where the ligands bind to the NS2B-NS3 protease.

5.4.1 Quinoa Peptide

Figure 27 shows the three best docking results of the quinoa derived peptide to the catalytically active site of the NS2B-NS3.

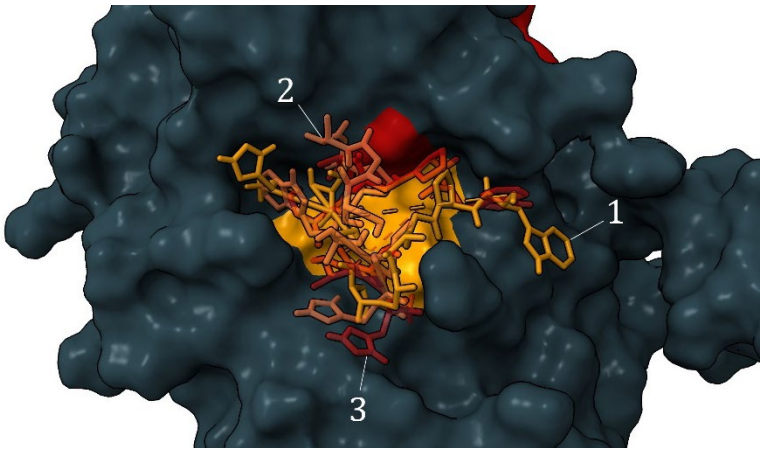

Figure 27: Docking simulation of the quinoa peptide (ligand) and the NS2B-NS3 without his-tag (receptor). The surface that is colored tangerine is the catalytic triad (HIS, ASP and SER) and the surface that is colored scarlet red is the allosteric site. The ligands are depicted in orange to red and in decreasing opacity in order of falling docking score. The docking scores can be seen in Table 8.

The interaction between the ligand and the catalytic site of the enzyme indicates that the quinoa peptide would be a potential competitive inhibitor of the NS2B-NS3. The docking results generated by AutoDock VINA can be seen in Table 8.

Table 8: The docking scores generated by AutoDock VINA when docking the quinoa peptide with NS2B-NS3. Position represents the indexing of the different orientations of the ligand, Score represents the change in Gibbs free energy of the ligand-receptor interaction, RMSD is the Root-Mean Square Deviation from the position with the best fit, and the Ligand H-bonds is the number of hydrogen bonds involved in the interaction between the ligand and the receptor.

Position	Score (kcal/mol)	RMSD	Ligand H-bonds
1	-7.8	0	6
2	-7.8	10.916	2
3	-7.7	28.475	1

The quinoa peptide exhibited a score of -7.8 kcal/mol in the two best positions (1 & 2) and -7.7 kcal/mol in the third best position (3). Docking position 1 also resulted in six ligand H-bonds.

5.4.2 Carvacrol

Figure 28 shows the docking results of the carvacrol molecule to the allosteric site of the NS2B-NS3.

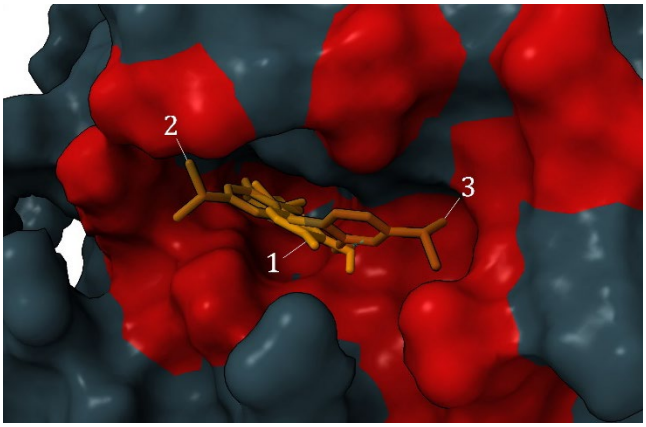

Figure 28: Docking simulation of carvacrol (ligand) and the NS2B-NS3 without his-tag (receptor). The surface that is colored scarlet red is the allosteric site. The ligands are depicted in orange to red and in decreasing opacity in order of falling docking score. The docking scores can be seen in Table 9.

The interaction between the carvacrol molecule and the allosteric site of the NS2B-NS3 showed that this ligand may exhibit potential non-competitive or uncompetitive inhibition of the enzyme. The docking results generated by AutoDock VINA can be seen in Table 9.

Table 9: The docking scores generated by AutoDock VINA when docking the carvacrol with NS2B-NS3. Position represents the indexing of the different orientations of the ligand, Score represents the change in Gibbs free energy of the ligand-receptor interaction, RMSD is the Root-Mean Square Deviation from the position with the best fit, and the Ligand H-bonds is the number of hydrogen bonds involved in the interaction between the ligand and the receptor.

Position	Score	RMSD	Ligand H-bonds
1	-5.6	0	0
2	-5.4	1.654	0
3	-5.4	4.093	1

The carvacrol molecule exhibited a score of -5.6 kcal/mol in the best position (1) and -5.4 kcal/mol in the second and third best position (2 & 3). Docking position 3 also resulted in one hydrogen bond. The carvacrol also interacted with the junction between the NS2B and NS3 moieties in close proximity of the allosteric site (Figure 29).

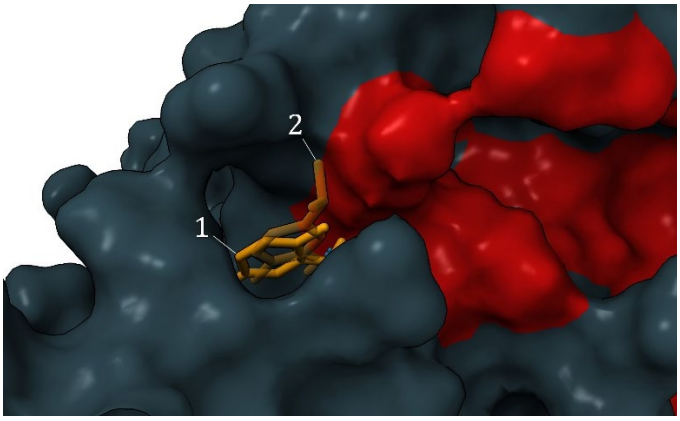

Figure 29: Docking simulation of carvacrol (ligand) and the NS2B-NS3 without his-tag (receptor). The surface that is colored scarlet red is the allosteric site. The ligands are depicted in orange to red and in decreasing opacity in order of falling docking score. The docking scores can be seen in Table 10.

The docking results of the carvacrol molecule at the junction of the NS2B and NS3 moieties can be seen in Table 10.

Table 10: The docking scores generated by AutoDock VINA when docking the carvacrol with NS2B-NS3. Position represents the indexing of the different orientations of the ligand, Score represents the change in Gibbs free energy of the ligand-receptor interaction, RMSD is the Root-Mean Square Deviation from the position with the best fit, and the Ligand H-bonds is the number of hydrogen bonds involved in the interaction between the ligand and the receptor.

Position	Score	RMSD	Ligand H-bonds
1	-6.0	0	0
2	-5.6	4.806	1

The carvacrol molecule exhibited a score of -6 kcal/mol in the best position (1) and -5.6 kcal/mol in the second-best position (2). Docking position 2 also resulted in one hydrogen bond.

5.4.3 Chlorogenic Acid

Figure 30 shows the docking results of the chlorogenic acid molecule to the allosteric site of the NS2B-NS3.

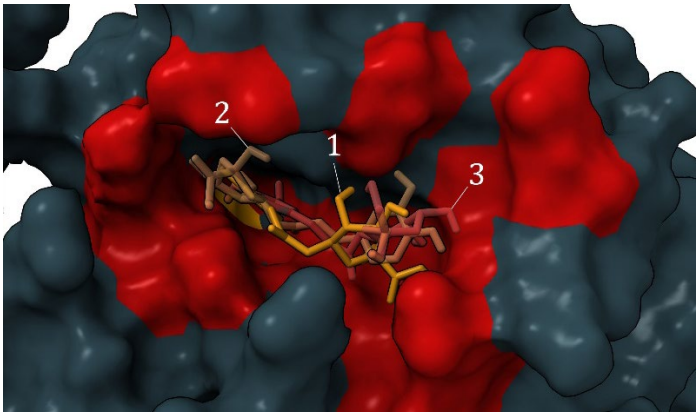

Figure 30: Docking simulation of chlorogenic acid (ligand) and the NS2B-NS3 without his-tag (receptor). The surface that is colored scarlet red is the allosteric site. The ligands are depicted in orange to red and in decreasing opacity in order of falling docking score. The docking scores can be seen in Table 11.

The interaction between the chlorogenic acid molecule and the allosteric site of the NS2B-NS3 showed that this ligand exhibits potential non-competitive or uncompetitive inhibition of the enzyme. The docking results generated by AutoDock VINA can be seen in Table 11.

Table 11: The docking scores generated by AutoDock VINA when docking the chlorogenic acid with NS2B-NS3. Position represents the indexing of the different orientations of the ligand, Score represents the Gibbs free energy of the ligand-receptor interaction, RMSD is the Root-Mean Square Deviation from the position with the best fit, and the Ligand H-bonds is the number of hydrogen bonds involved in the interaction between the ligand and the receptor.

Position	Score	RMSD	Ligand H-bonds
1	-7.9	0	1
2	-6.9	8.272	1
3	-6.7	3.933	1

The carvacrol molecule exhibited a score of -7.9 kcal/mol in the best position (1), a -6.9 kcal/mol in the second position (2) and a -6.7 kcal/mol in the third best position (3). All docking positions involved a hydrogen bond in the interaction with the receptor. The chlorogenic acid also interacted with the catalytically active site of the NS2B-NS3 (Figure 31).

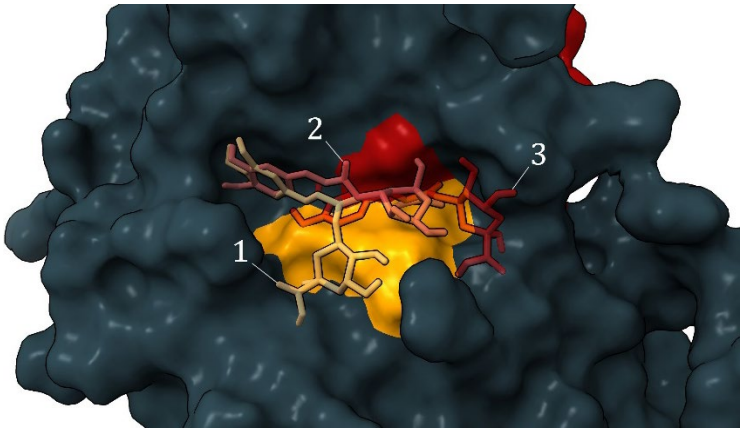

Figure 31: Docking simulation of the chlorogenic acid (ligand) and the NS2B-NS3 without his-tag (receptor). The surface that is colored tangerine is the catalytic triad (HIS, ASP and SER) and the surface that is colored scarlet red is the allosteric site. The ligands are depicted in orange to red and in decreasing opacity in order of falling docking score. The docking scores can be seen in Table 12.

The docking results of the chlorogenic acid molecule at the catalytic site of the NS2B-NS3 can be seen in Table 12.

Table 12: The docking scores generated by AutoDock VINA when docking the chlorogenic acid with NS2B-NS3. Position represents the indexing of the different orientations of the ligand, Score represents the Gibbs free energy of the ligand-receptor interaction, RMSD is the Root-Mean Square Deviation from the position with the best fit, and the Ligand H-bonds is the number of hydrogen bonds involved in the interaction between the ligand and the receptor.

Position	Score	RMSD	Ligand H-bonds
1	-6.5	0	1
2	-6.5	2.747	2
3	-6.4	2.508	3

The chlorogenic acid molecule exhibited a score of -6.5 kcal/mol in the best and second-best position (1 & 2) and a -6.4 kcal/mol in the third best position (3). All docking positions involved hydrogen bonds in the interaction with the receptor – resulting in 1, 2 and 3 bonds respectively.

5.5 Inhibition studies

Inhibition assays were performed using carvacrol and chlorogenic acid to experimentally determine if the ligands exhibited any inhibitory effects on the NS2B-NS3 without his-tag. The Lineweaver-Burk plot for inhibition of NS2B-NS3 with carvacrol can be seen in Figure 32 below.

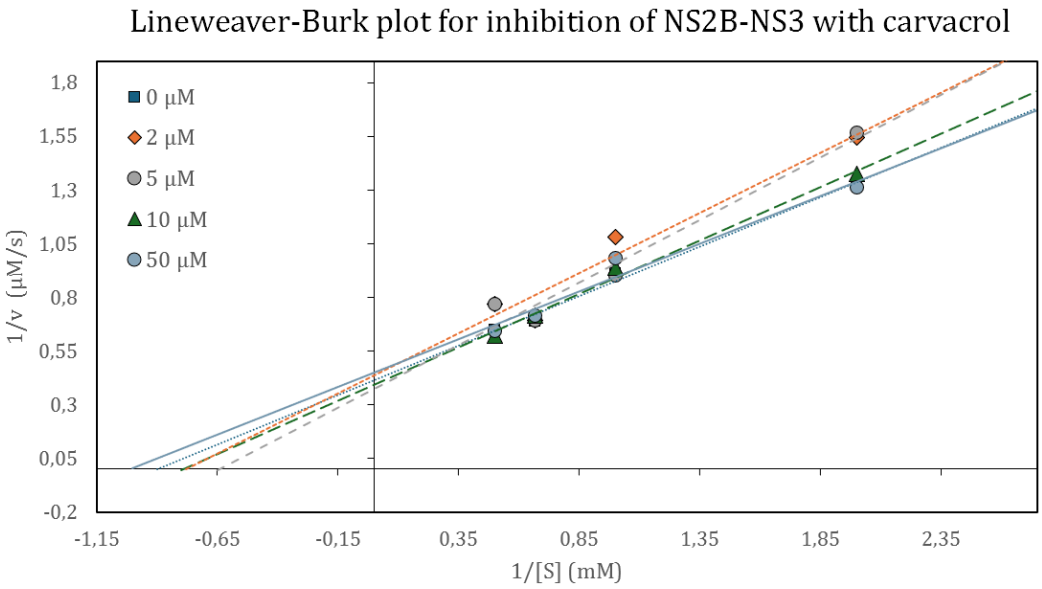

Figure 32: Lineweaver-Burk plot of the inhibition of NS2B-NS3 without his-tag (5.7 µM) with carvacrol. Measurements were made at five different inhibitor concentrations (0 µM, 2 µM, 5 µM, 10 µM and 50 µM) at four different substrate concentrations (0.25 mM, 0.5 mM, 1 mM, and 1.5 mM). All measurements were carried out in duplicates.

There was no clear inhibitory effect of the carvacrol on the activity of the NS2B-NS3 enzyme. There also seemed to be an irregular relationship between increasing inhibitor concentration and the activity of the NS2B-NS3. Hence, pointing to the fact that carvacrol may not be a potent inhibitor for this enzyme. To confirm this, a sample containing 50 µM carvacrol was also run, yielding no difference in results.

Continuing on with the other chemical ligand, the Lineweaver-Burk plot for the inhibition assay of chlorogenic acid can be seen in Figure 33 below.

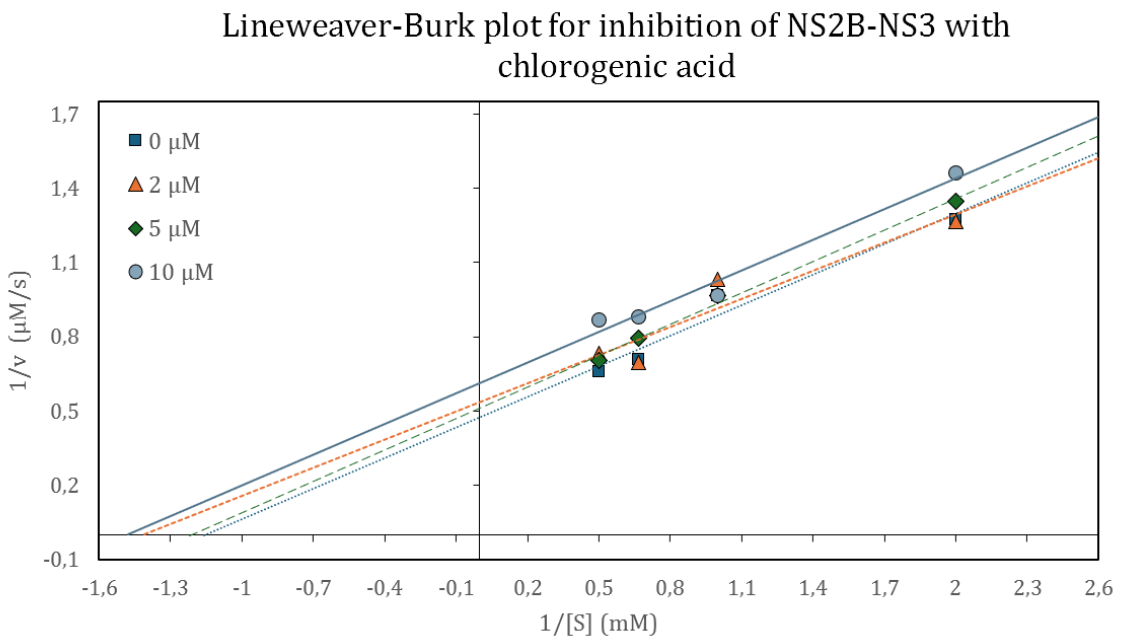

Figure 33: Lineweaver-Burk plot of the inhibition of NS2B-NS3 without his-tag (5.7 µM) with chlorogenic acid. Measurements were made at four different inhibitor concentrations (0 µM, 2 µM, 5 µM, 10 µM) at four different substrate concentrations (0.25 mM, 0.5 mM, 1 mM, and 1.5 mM). All measurements were carried out in duplicates.

The chlorogenic acid showed a gradually increasing inhibition of the NS2B-NS3 with increasing concentration. A 10 µM concentration of chlorogenic acid decreased the initial rate of the NS2B-NS3 with 23.5%. The $V_{max,app}$, $K_{m,app}$ and $k_{cat,app}$ was calculated from the Lineweaver-Burk plot of the carvacrol and chlorogenic acid inhibition assays and can be seen in Table 13 below.

Table 13: The apparent kinetic constants for inhibition of NS2B-NS3 without his-tag with carvacrol and chlorogenic acid (10 µM). The relative change compared to the pure uninhibited enzyme are presented within paranthesis.

	$V_{max,\,app}$ (µM/s)	**$K_{m,\,app}$ (mM)**	**$k_{cat,\,app}$ (s^{-1})**
Carvacrol	-	-	-
Chlorogenic acid	1.628 (-23.5%)	0.672 (-28.5%)	0.286 (-23.8%)

As there were no measurable inhibition from exposure to carvacrol, no kinetic variables could be determined for this compound. However, the exposure to 10 µM of chlorogenic acid resulted in a $V_{max,app}$ of 1.628 µM/s, a $K_{m,app}$ of 0.672 mM and a $k_{cat,app}$ of 0.286 s^{-1}.

6 Discussion

Despite recent advancements in the field of medicine and the major threat posed by a rapidly spreading virus, there still are no available medication for ameliorating or curing the dengue virus. Trying to find naturally derived sources of bioactive molecules is a sustainable and important route to go down trying to understand and treat these diseases. Another way to continue our development in this field is to improve the biological modeling conditions that are currently used for testing inhibitors, drugs, and bioactive compounds. Previous endeavors of inhibiting the NS2B-NS3 have used the his-tagged enzyme in their studies, which already seem to be inhibited by the presence of the tag. Removing the his-tag before conducting these types of assays could prove crucial to allow for a higher and more reliable catalytic activity, analogous to that of the wild-type enzyme involved in the infectiousness of the dengue virus.

6.1 Prediction of the structural model of NS2B-NS3

The development of artificial intelligence has facilitated a plethora of new approaches in the field of biotechnology, one of the more prominent being the ability to predict molecular conformations at a very high degree of certainty. This is confirmed when looking at the prediction of the produced NS2B-NS3 model compared to the experimentally determined structure from PDB (*6MO0*). The overlaid structures proved that the results from AlphaFold was highly reliable, as there were minute differences between the two models. The structures that were predicted and was missing from the experimental model retrieved from PDB were neither very complex. The glycine linker that linked the NS2B cofactor moiety to the NS3 protease moiety had high flexibility but did cross the region in front of the enzymatic triad, it could be interesting to see if a longer or shorter linker would have any effect on the enzymatic activity of the protease. However, due to the small nature of the glycine residue, the steric hinderance was most likely negligible.

6.2 Recombinant protein production

A large portion of the experimental part of the project was spent on analyzing unsuccessful production of the NS2B-NS3. A combination of protocol errors and compromised batches of competent cells led to the belief that the plasmid itself was degraded, however fortunately enough that was not the case. The yield when producing the recombinant NS2B-NS3 was very high, resulting in crude extracts having a concentration of 3-4 mg/mL, pointing to the fact that the NS2B-NS3 plasmid had a very high copy number. Throughout the entirety of the project no batch were produced unsuccessfully.

The step that lowered the yield of readily available NS2B-NS3 was the TEV protease digestion and the repeated purifications using the IMAC. Several trials were made to optimize the digestion however the improvements seemed to plateau after a certain point – one also had to factor in the amount of readily available TEV protease for finding an optima between efficiency and efficacy in these optimization trials. Further optimization could be made of the purification protocol used in the ÄKTA IMAC to avoid unnecessary dilution and loss of the protein due to the elution gradient.

Comparing the commercially available TEV protease with the recombinant TEV protease showed that the recombinant enzyme expressed a greater activity. However, that may very well be explained by the differing storage conditions of the enzymes, as the commercial kit was stored at -20 °C, and was thawed and refrozen after each use, whereas the recombinant TEV protease was used within 1-3 days after it had been produced. After five days of storage at 4 °C there seemed to be rapid denaturation of the produced TEV protease, with visible denaturation in the otherwise clear extract.

The SDS-PAGE gel showing the different incubation times of the TEV protease digestion as well as the unpurified NS2B-NS3 (Figure 17) was a great probe for determining the conditions for future digestions. However, as there were differences in the composition of the solutions, e.g. the commercial TEV protease was stored in glycerol, no direct conclusions can be drawn regarding the stability of the recombinant and commercial TEV protease from this data. The digestions were therefore always made at 4°C overnight to ensure stability and allow sufficient time for digestion.

The SDS-PAGE analysis unraveled a discrepancy in the molecular weight of the NS2B-NS3 as the gel showed a molecular weight of ~35 kD when the theoretical weight is closer to ~26 kD. This was probably caused by interactions with the detergent used in SDS-PAGE giving the molecule a higher intrinsic charge. After confirming the successful digestion and production of the NS2B-NS3 through SDS-PAGE, samples were sent for MS analysis to confirm if the his-tag had been entirely removed or not, and also allowing a better comparison between digestion using the commercial and the recombinant TEV protease. The BLAST and Mascot search did confirm that the peaks observed matched the NS2B-NS3 protease, but not to a large extent. However, as no fragment matched the his-tag and linker the digestion was assumed to have taken place successfully, this conclusion was also supported by the apparent difference in the measured activity between the undigested (with his-tag) and digested (without his-tag) NS2B-NS3. It would be beneficial to redo the MS analysis to confirm these results, but due to time constraints this was not possible. It also became evident that the samples containing the NS2B-NS3 had been digested using trypsin, which may explain the relatively low match from the BLAST and Mascot searches, however this does not explain the difference in the number of matches comparing the digestion with the recombinant and commercial TEV protease. A suggestion for future studies is to run the sample without trypsin digestion to see if a better identification is achieved and utilize the MALDI-TOF technique. The complete MS spectrum can be seen in *Appendix B.*

6.3 Characterization and activity measurements

Before conducting inhibition studies, characterization and activity measurements were carried out to develop a better understanding of the properties possessed by the NS2B-NS3 without his-tag. Determination of the stability of the enzyme was made through melting point determination at different pH levels. From previous literature it became evident that Tris-HCl was the most commonly used buffer for NS2B-NS3, and as this buffer also offers a wide pH range this was deemed optimal for analysis in the nano-DSF. The enzyme showed high stability at all pH except at pH 5, where the melting point were about 3°C lower, however there were no substantial difference (Figure 21). To better understand this difference, the ratio of unfolded and folded enzyme was plotted at a

high and low pH (5 and 10) (Figure 22). Both curves exhibited abnormal behavior where higher temperature seemed to make the enzyme refold, with the trend being especially pronounced at low pH levels. This phenomenon is probably explained by the formation of aggregates, decreasing the number of exposed hydrophobic residues which otherwise would fluoresce at the measured wavelength.

The initial activity measurements were made using one fixed concentration of DL-BAPNA to determine what concentration of NS2B-NS3 would be suitable for subsequent analysis. The optimal concentration of enzyme was determined to be 0.15 mg/mL or 5.7 µM. The raw activity measurements can be seen in *Appendix B*. Overall, the activity of the NS2B-NS3 with DL-BAPNA was quite low, especially when compared to fluorescence assays which seem to exhibit a higher activity at lower substrate concentrations [52].

When comparing the activity of the tagged and untagged NS2B-NS3 with DL-BAPNA, most of the data-points were statistically significant based on a paired double-sided t-test. Repeating the assay at parametrically spaced substrate concentrations may be beneficial for achieving a better fit of the plots. The NS2B-NS3 without his-tag had a larger V_{max} and k_{cat} than the NS2B-NS3 with his-tag (V_{max} = 2.127 ± 0.104 µM/s and k_{cat} = 0.375 ± 0.018 µM/s compared to V_{max} = 1.512 ± 0.103 µM/s and k_{cat} = 0.304 ± 0.021 µM/s). Note that the kinetic variables are calculated assuming a 100 % viable enzyme using the theoretical molecular weight acquired from *Expasy's ProtParam*. These results could prove useful in the process of designing assays that exhibit the same level of activity as in the host organism, which could contribute to more accurate and reliable development of novel bioactive substances. Hence, presenting a better way to perform inhibition studies in the future. To further expand these results, it would be important to assay different types of substrates to see if the difference in activity is persistent regardless of the substrate used. It would also be valuable to test the viability of the enzyme using a standard or control, to determine the ratio of active to unactive enzyme. This would also allow for comparison of the amount of active enzyme in the tagged and untagged NS2B-NS3 solutions which would be very advantageous.

6.4 Inhibition studies and molecular docking simulations

The theory on which the selection of inhibitors was based upon was that the main ligand to use for the inhibition assays was the quinoa derived peptide, as it exhibited a strong competitive inhibition of the NS2B-NS3 enzyme in the docking simulations. The other compounds were selected as these mainly interacted with the allosteric site of the enzyme and the fact that they continued the theme of being derived from natural and sustainable sources. Chlorogenic acid and carvacrol possessed similar properties but where of drastically different sizes, thus enabling a very versatile combination of inhibitory functions.

The docking results for the quinoa peptide resulted in a high free energy of binding as well as a large span in RMSD values. This indicates that the ligand would achieve strong binding to the site regardless of orientation, which increases the chance for interaction at the catalytic site of the NS2B-NS3. The docking results for carvacrol and chlorogenic acid was promising and showed rather strong binding scores and adequate RMSD values. Based on the docking results there were a possibility for multiple carvacrol

molecules to bind to the allosteric site of the NS2B-NS3 simultaneously, however this may be dependent on the concentration of inhibitor present. The intention was to test a combination of these inhibitors, allowing for competitive and non-competitive inhibition to occur simultaneously and also unravel what differences a larger or smaller allosteric inhibitor may have on the NS2B-NS3 protease.

As the quinoa peptide was impossible to produce as a soluble peptide, individual assays where run for carvacrol and chlorogenic acid instead. Carvacrol did not exhibit any inhibitory effects at the concentrations assayed, however as the trend in activity was quite irregular this may have been caused by the low number of experiments at different substrate concentrations. One hypothesis for the ineffectiveness of carvacrol as a potential inhibitor may be that it did not hinder the mobility of the NS2B-NS3 due to its small size and low steric hindrance. The chlorogenic acid did however exhibit distinct un-competitive inhibition of the NS2B-NS3 enzyme based on the slope of the fitted line. As the idea of testing these inhibitors was to potentially aid in the development of a novel drug, a lower range of inhibitor concentrations was initially tested. It should be noted that an un-competitive inhibitor is not the ideal candidate for drug-development as these types of inhibitors usually bind to the enzyme-substrate complex instead of the free enzyme. It would be interesting to further study this un-competitive behavior through docking simulations using the NS2B-NS3-substrate complex and determine how these results differ from using the free NS2B-NS3. Unfortunately, due time constraints and the availability of freshly produced NS2B-NS3, higher concentrations of inhibitor and greater number of substrate concentrations where not tested in the scope of this project. It would be necessary to redo the assay but at a greater number of substrate concentrations and replications to get more reliable results and a more accurate fit of the Lineweaver-Burk plots. Running the assays at higher inhibitor concentrations would also allow for determination of IC_{50} for chlorogenic acid which would be beneficial to further characterize the role of chlorogenic acid as a viral protease inhibitor.

7 Conclusion

The production of recombinant NS2B-NS3 can be optimized to reach a higher degree of activity by removing the histidine tag by TEV protease digestion. The NS2B-NS3 with his-tag exhibited a V_{max} of 1.512 ± 0.103 µM/s, a K_m of 0.735 ± 0.015 mM and a k_{cat} of 0.304 ± 0.021 µM/s while the NS2B-NS3 without his-tag exhibited a V_{max} of 2.127 ± 0.104 µM/s, a K_m of 0.940 ± 0.025 mM and a k_{cat} of 0.375 ± 0.018 µM/s. Carvacrol do not seem to exhibit any inhibitory effects on the NS2B-NS3, whereas chlorogenic acid seem to be a potential inhibitor, decreasing the enzyme activity by ~23.5% at a concentration of 10 µM, exhibiting a $V_{max,app}$ of 1.628 µM/s, a $K_{m,app}$ of 0.672 mM, and a $k_{cat,app}$ of 0.286 s^{-1}. The quinoa derived peptide was unfortunately not soluble and could therefore not be included in the assay.

Hopefully these results can aid in the pursuit of finding a potential cure for the dengue virus by presenting a novel way of performing kinetic studies of the NS2B-NS3 and by providing a stepping-stone for future research topics in the field of virology.

8 References

[1] WHO, ”Who coronavirus (COVID-19) dashboard,” 31 01 2024. [Online]. Available: https://covid19.who.int/.

[2] CDC, ”Data and maps. Centers for Disease Control and Prevention; 2023,” [Online]. Available: https://www.cdc.gov/dengue/statistics-maps/data-and-maps.html. [Använd 31 januari 2024].

[3] K. C. Li Q, ”Structures and dynamics of dengue virus nonstructural membrane proteins,” Multidisciplinary Digital Publishing Institute, 2022.

[4] K. E. P. A. K. Patick, ”Protease Inhibitors as Antiviral Agents,” *National Library of Medicine,* pp. 614-627, 1998.

[5] C. C. E. Team, ”Cleveland Clinic,” [Online]. Available: https://my.clevelandclinic.org/health/treatments/24937-protease-inhibitors. [Använd 01 02 2024].

[6] W. Z. et al. Kuhn, ”Structure of Dengue Virus: Implications for Flavivirus Organization, Maturation, and Fusion,” *National Library of Medicine,* pp. 717-725, 8 Mars 2002.

[7] A. m. e. a. Harapan Harapan, ”Dengue: A Minireview,” *Viruses,* vol. 12, nr 7th Pan-American Dengue Research Network Meeting: Lima 2020, p. 829, 2020.

[8] M. M. Amudhan Murugesan, ”Dengue Virus,” *National Library of Medicine,* pp. 281-359, 20 september 2019.

[9] C. V. R. b. C. S. J. c. L. C. V. G. d. Lt Col M.S. Mustafa a, ”Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control,” *ScienceDirect,* pp. 67-70, 2015.

[10] F. B. P. B. J. F. C. R. d. C. P. d. M. X. d. L. H Attoui 1, ”Sequence determination and analysis of the full-length genome of colorado tick fever virus, the type species of genus Coltivirus (Family Reoviridae),” *National Library of Medicine,* pp. 1121-1125, 2000.

[11] O. O. e. al., ”Potential role of vector-mediated natural selection in dengue virus genotype/lineage replacements in two epidemiologically contrasted settings,” pp. 1346-1357, 2021.

[12] C. P. ∗. S. T. ∗. A.-L. H. ∗. †. F. J. D. ∗. Silvio Urcuqui-Inchima ∗, ”Chapter 1 - Recent Developments in Understanding Dengue Virus Replication,” *Advances in Virus Research,* vol. 77, pp. 1-39, 2010.

[13] A. K., R. B. Eliana G. Acosta 1, ”Chapter One - Revisiting Dengue Virus–Host Cell Interaction: New Insights into Molecular and Cellular Virology,” *Advances in Virus Research,* vol. 88, pp. 1-109, 2014.

[14] A. B. e. a. Nikita Nanaware, ”Dengue Virus Infection: A Tale of Viral Exploitations and Host Responses,” *Viruses 2021,* 2021.

[15] S. Rajapakse, ”Dengue shock,” *Journal of Emergencies, Trauma and shock,* pp. 120-127, 4 januari 2011.

[16] M. A. İlknur Uçak, ”Chapter 13 - Enzymes,” *Nutraceutical and Functional Food Components,* pp. 537-571, 2021.

[17] R. Roskoski, ”Michaelis-Menten Kinetics,” *xPharm: The Comprehensive Pharmacology Reference,* 2007.

[18] e. a. Abdul Razzaq, ”Microbial Proteases Applications,” vol. 7, 12 juni 2019.

[19] J. S. B. Carlos López-Otín, ”Proteases: Multifunctional Enzymes in Life and Disease,” *JBC - Journal of biological chemistry,* pp. 433-437, 7 november 2008.

[20] A. M. T. M. S. G. e. a. Mala B. Rao, ”Molecular and Biotechnological Aspects of Microbial Proteases,” *Microbiology and molecular biology reviews,* pp. 597-635, september 1998.

[21] S. G. V. b. J. L. Dahai Luo a, ”The flavivirus NS2B–NS3 protease–helicase as a target for antiviral drug development,” *Antiviral Research,* vol. 118, pp. 148-158, 2015.

[22] Y. Y. et al. Bo Zhong 1, ”The Adaptor Protein MITA Links Virus-Sensing Receptors to IRF3 Transcription Factor Activation,” *Immunity,* vol. 29, pp. 538-550, 2008.

[23] A. M. M. P. R. P. S. G. M. B.-R. S. S. S. R. R.-M. C. F. M. N. B. F.-S. Sebastian Aguirre, ”DENV Inhibits Type I IFN Production in Infected Cells by Cleaving Human STING,” *PLOS Pathogens,* vol. October, 2012.

[24] B. E. M. M. et al. . Abdul Wahaab, ”Potential Role of Flavivirus NS2B-NS3 Proteases in Viral Pathogenesis and Anti-flavivirus Drug Discovery Employing Animal Cells and Models: A Review,” *MDPI Viruses,* 28 December 2022.

[25] N. S. Y. Z. M. D. C. L. a. Z. Y. Joshua Holcomb, ”Protein crystallization: Eluding the bottleneck of X-ray crystallography,” *AIMS Biophys.,* 2017.

[26] B. J. H. D. C. S. S. a. M. C. Heejin Nam, ”Tobacco etch virus (TEV) protease with multiple mutations to improve solubility and reduce self-cleavage exhibits enhanced enzymatic activity,” *FEBS Open Bio,* pp. 619-626, 18 March 2020.

[27] A. S. Aditi Gangopadhyay, ”Exploring allosteric hits of the NS2B-NS3 protease of DENV2 by structure-guided screening,” *Computational Biology and Chemistry,* vol. 104, Juni 2023.

[28] W. D. P. W. I. H. B. B. J. A. R. J. M. a. J. W. John Strelow, Assay Guidance Manual, Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004.

[29] I. M. Kovach, "Competitive Irreversible Inhibition of Enzymes in the Presence of A Substrate: Scope and Limitations," *Journal of Enzyme Inhibition,* vol. 3, pp. 201-212, 1991.

[30] E. F. V. B. P. K. N. G. C. K. Pathan S, "Nutritional composition of the green leaves of quinoa (chenopodium quinoa wild)," CCSENET, 2019.

[31] e. a. Robin Poole, "Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes," *The BMJ,* p. 359, 21 november 2017.

[32] Z. C. C. D. W. W. X. Yue Ding, "Antiviral activity of chlorogenic acid against influenza A (H1N1/H3N2) virus and its inhibition of neuraminidase," *Scientific Reports | Natureresearch,* p. 7, 10 april 2017.

[33] J. C. M. A. Zacharias E Suntres, "The bioactivity and toxicological actions of carvacrol," *Critical Reviews in Food Science and Nutrition* , pp. 304-318, 30 september 2014.

[34] D. J. W. e. al., "Antimicrobial Activity of Naturally Occurring Phenols and Derivatives Against Biofilm and Planktonic Bacteria," *Medicinal and Pharmaceutical Chemistry,* vol. 7, 2019.

[35] E. M. V. M. I. M. M. W. N. S. M. d. M. C. e. a. Mehdi Sharifi-Rad, "Carvacrol and human health: A comprehensive review," *Phtoterapy Research,* vol. 32, nr 9, pp. 1675-1687, 9 maj 2018.

[36] ,. R. M. M. K. S. B. M. A. C. M. N. H. Md. Fakruddin, "Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli," *ISRN Biotechnology,* 2013.

[37] C. O. J. Baolei Jia, "High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives," *Open Biology,* 01 August 2016.

[38] F. Baneyx, "Recombinant protein expression in Escherichia coli," *Current Opinion in Biotechnology,* pp. 411-421, October 1999.

[39] R. Shepherd, "Chromatographic Methods," *Comprehensive Coordination Chemistry II,* pp. 567-577, 2003.

[40] E. Sulkowski, "Purification of proteins by IMAC," *Trends in Biotechnology,* vol. 3, nr 1, pp. 1-7, 1985.

[41] W. J. W. D. H. P. Andrew B Nowakowski, ”Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions,” Metallomics, 2014.

[42] P. L. Urban, ”Quantitative mass spectrometry: an overview,” *Philos Trans A Math Phys Eng Sci.,* 28 October 2016.

[43] K. O. R. &. G. Gao, ”Theory and applications of differential scanning fluorimetry in early-stage drug discovery,” *Biophysical Reviews,* vol. 12, pp. 85-104, 31 Januari 2020.

[44] ”AlphaFold,” 2024. [Online]. Available: https://alphafold.ebi.ac.uk/. [Använd 02 Februari 2024].

[45] R. E. A. P. T. G. M. F. O. R. K. T. R. B. A. Ž. A. P. A. B. C. M. S. A. A. K. A. J. B. A. C. B. John Jumper, ”Highly accurate protein structure prediction with AlphaFold,” *Nature,* pp. 583-589, 2021.

[46] RBVI, ”RBVI UCSF Chimera Documentation,” [Online]. Available: https://www.rbvi.ucsf.edu/chimera/docindex.html. [Använd 02 Februari 2024].

[47] N. A. R. Group, ”Protein Data Bank: the single global archive for 3D macromolecular structure data,” *National Library of Medicine,* pp. 520-528, 8 Januari 2019.

[48] Y. E. Team, ”YASARA,” 2021. [Online]. Available: https://www.yasara.org/papers.htm.

[49] Sigma Aldrich, ”Sigma Aldrich,” 2024. [Online]. Available: https://www.sigmaaldrich.com/SE/en/product/sigma/b4875. [Använd 04 04 2024].

[50] Thermo Fisher Scientific, Pierce Protein Research, Rockford, Illinois, USA, ”Chapter 33 Performing and Optimizing Western Blots with an Emphasis on Chemiluminescent Detection,” *Methods In Enzymology,* 3 november 2009.

[51] A. S. G.-U. B. R. Evangelia Petsalaki, ”Accurate Prediction of Peptide Binding Sites on Protein Surfaces,” *PLOS Computational Biology,* 27 march 2009.

[52] L. Y. J.-H. J. Z. W. T.-Y. L. Yuxuan Pang, ”AVPIden: a new scheme for identification and functional prediction of antiviral peptides based on machine learning approaches,” *PubMed,* p. 22, 5 november 2021.

[53] Thermo Fisher, *Thermo Fisher Scientific AcTEV Protease.*

[54] a. D. D. e. a. Kok-Chuan Tiew, ”Inhibition of Dengue Virus and West Nile Virus Proteases by Click Chemistry-Derived Benz[d]isothiazol-3(2H)-one Derivatives,” *Bioorganic & Medicinal Chemistry,* vol. 20, nr 3, pp. 1213-1221, 1 february 2012.

[55] D. G. Gibson, Gibson Assembly Cloning Guide 2nd edition - RESTRICTION DIGESTFREE, SEAMLESS CLONING, SGIDNA, 2017, pp. 3-30.

[56] S. R. Daniel G. Gibson, ”Gibson Assembly® - Building a Synthetic Biology Toolset,” *NEB,* 2024.

[57] R. M. B. G. N. S. S. M. L. M. S. C. Christiano Argano, ”Protective Effect of Vitamin D Supplementation on COVID-19-Related Intensive Care Hospitalization and Mortality: Definitive Evidence from Meta-Analysis and Trial Sequential Analysis,” *Pharmaceuticals MDPI,* 16 januari 2023.

[58] C. N. N. P. S. R. D. T. Karunaithas Rasaratnam, ”A novel peptide isolated from garlic shows anticancer effect against leukemic cell lines via interaction with Bcl-2 family proteins,” pp. 1017-1028, 171 Februari 2021.

[59] U. L. Team, ”United Nations,” 2020. [Online]. Available: https://www.un.org/en/observances/tuna-day. [Använd 12 februari 2024].

[60] D. D. Bikle, ”Vitamin D and the skin: Physiology and pathophysiology,” *Rev Endocr Metab Disord,* vol. 1, pp. 3-19, 13 mars 2012.

[61] C. Aranow, ”Vitamin D and the Immune System,” *J Investig Med.,* pp. 881-886, 1 augusti 2011.

[62] ”Chapter 7 - Enzymes,” *Protocols in Biochemistry and Clinical Biochemistry,* pp. 61-67, 2021.

[63] G. A. A. M. e. a. Markossian S, Assay Guidance Manual, B. (. E. L. &. C. a. t. N. C. f. A. T. Sciences, Red., 2004.

9 Appendices

Appendices A – Bioinformatics & Recombinant protein production

Figure A1 shows the comparison of the predicted structure generated by AlphaFold and the experimentally determined structure.

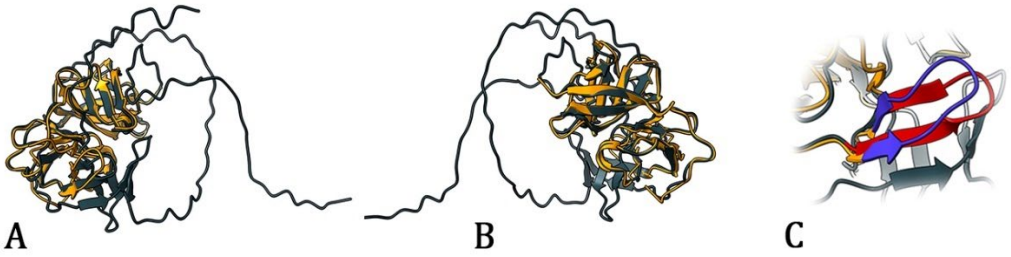

Figure A1: Comparison of the predicted structure and the experimentally determined structure. The predicted structure is depicted in petrol blue and the experimentally determined structure in tangerine. Figure A1(A) & (B) shows the comparison from two opposite angles, Figure A1(C) shows the largest difference between the two models, with the blue being the experimentally determined model and the red being the predicted model.

The optical density measurements (OD) during the first preparation of competent cells following a faulty protocol can be seen in Table A1.

Table A1: The optical density measurements (OD) taken during growth of cells in prepration of competent cell cultures. This was the first try, following a protocol that contained an error.

Strain / OD	**T_0 (08:30)**	**T_1 (09:10)**	**T_2 (09:40)**
BL21(DE3)	0.249	0.261	0.431
DH5-α	0.242	0.288	0.335

As the prepared competent DH5-α yielded colonies even though they were prepared in a faulty manner, this stock was kept (Figure A1*)*. However, the BL21(DE3) did not yield any colonies and were therefore remade (Table A2).

Table A14: The optical density measurements (OD) taken during growth of cells in prepration of competent cell cultures. This was the second try where the error in the protocol had been corrected. NOTE: Only the Bl21 strain was remade as this was the most important culture for recombinant protein production.

Strain / OD	**T_0 (08:00)**	**T_1 (08:30)**	**T_2 (09:15)**	**T_3 (09:45)**	**T_3 (10:05)**
BL21(DE3)	0.047	0.062	0.115	0.282	0.350

Figure A2 shows the plot of *Time vs OD* for the preparation of competent cells for strain BL21(DE3).

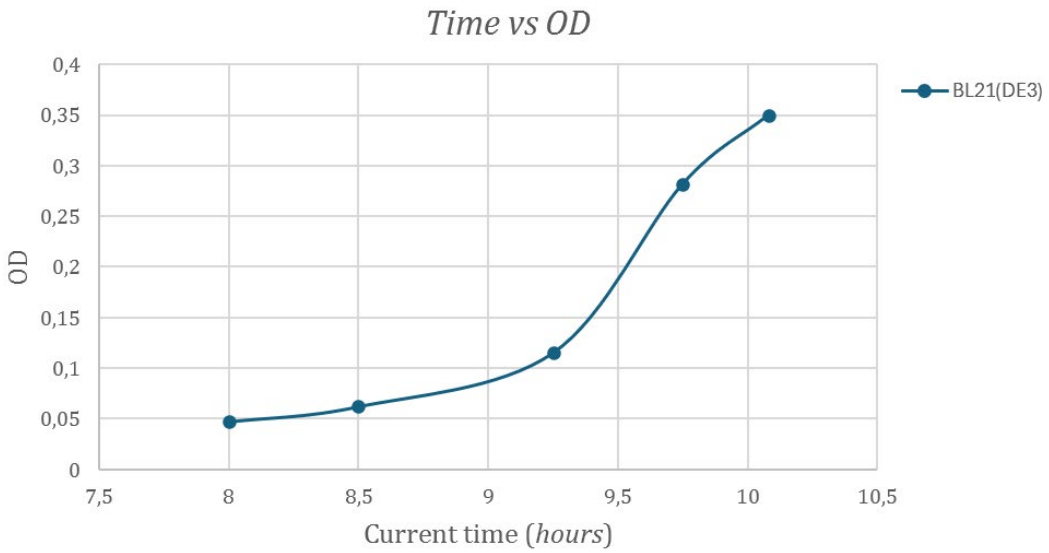

Figure A2: Current time vs OD from data acquired during competent cell preparation (BL21(DE3)) following the correct protocol. The cells were harvested at t=10:15 and an OD of 0,35.

Figure A3 shows the plot of time vs OD and the time of induction for producing the NS2B-NS3 protease using the BL21(DE3) strain.

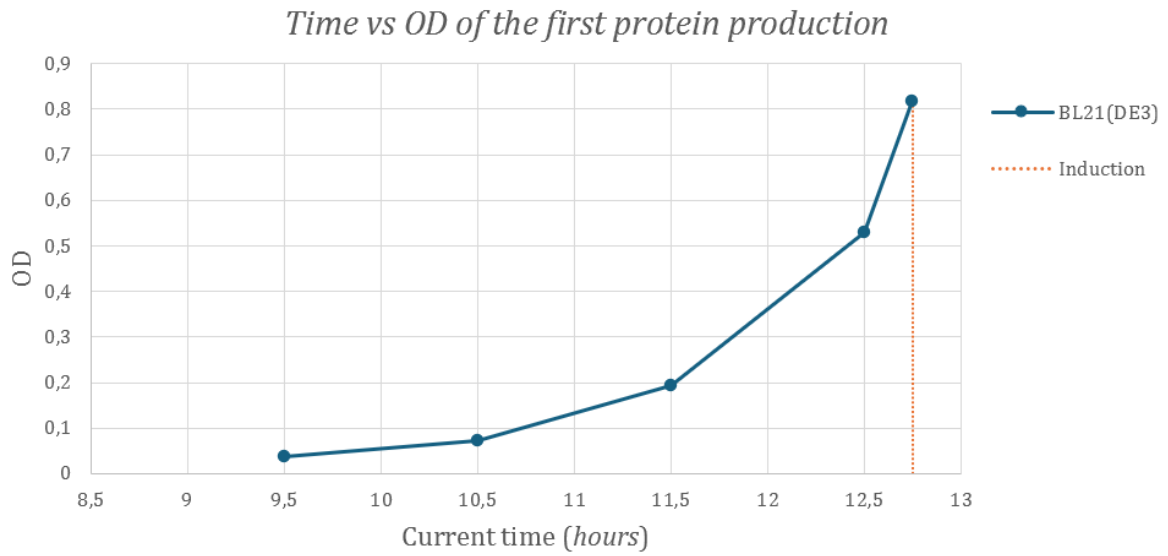

Figure A3: Current time vs OD during the first production of the NS2B-NS3 protease using BL21(DE3) strain. The induction took place at 12:43 when the OD had reached a value of 0.818 (see orange line).

Appendices B – Characterization and activity measurements

Figure B1 shows the MS spectra for digestion of NS2B-NS3 with recombinant and commercial TEV protease.

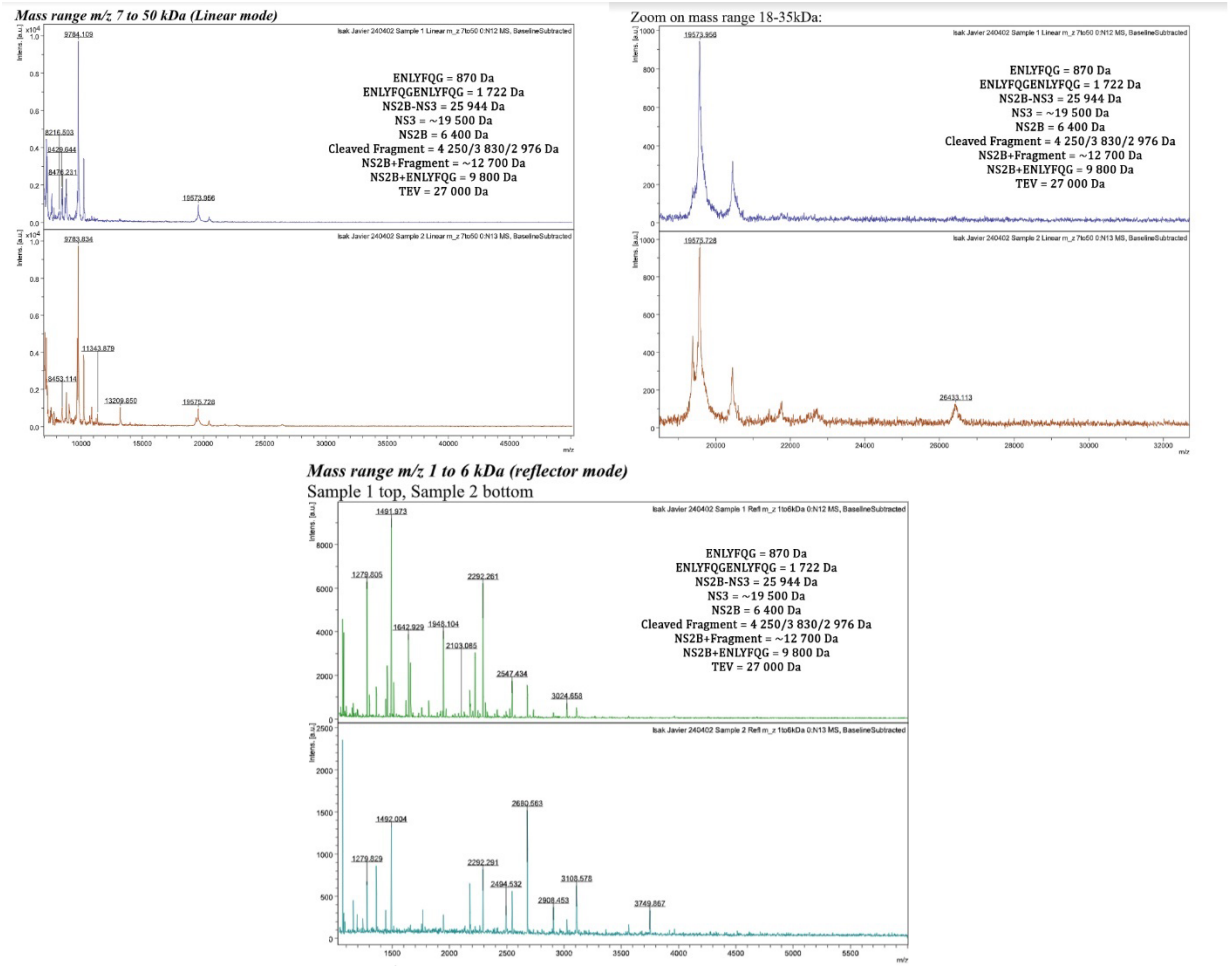

Figure B1: The generated MS spectra at different ranges and modes, with the added weights for all possible fragments that may have been present in the sample if it were not digested with trypsin. Sample 1 (top) is the NS2B-NS3 digested with recombinant TEV protease while Sample 2 (bottom) is the NS2B-NS3 digested with the commercial TEV protease.

Figure B2 shows the raw activity measurements for assaying which concentration of NS2B-NS3 to use for subsequent analysis.

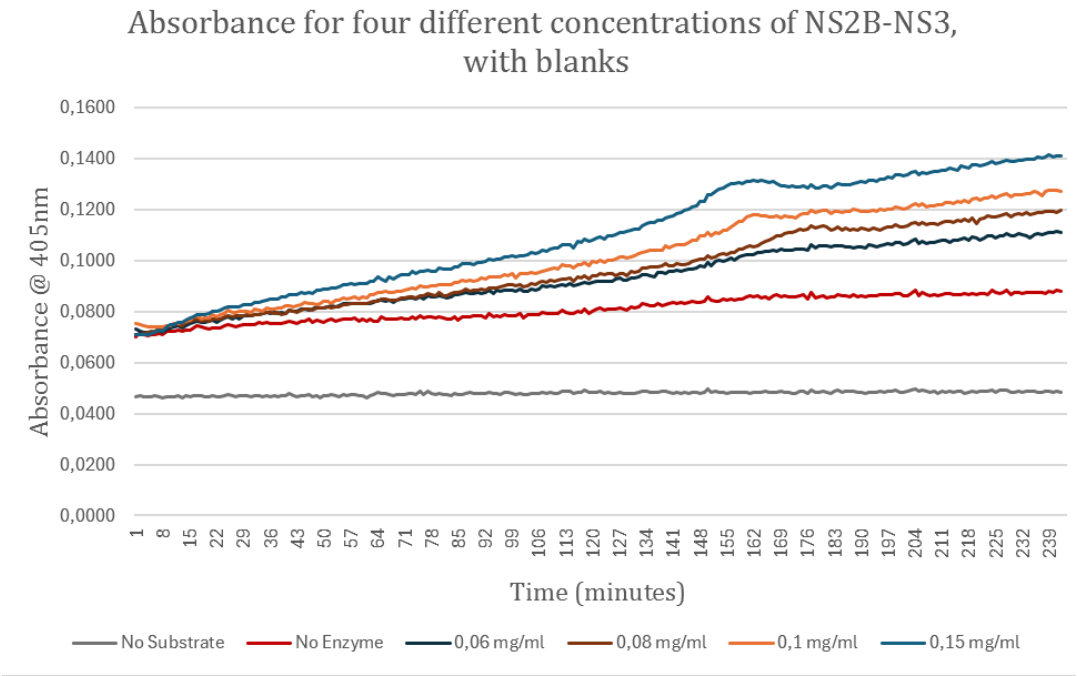

Figure B2: The raw absorbance measurements using four different concentrations of NS2B-NS3 with 1 mM DL-BAPNA with no CHAPS, NaCl, Glycerol or other additives.