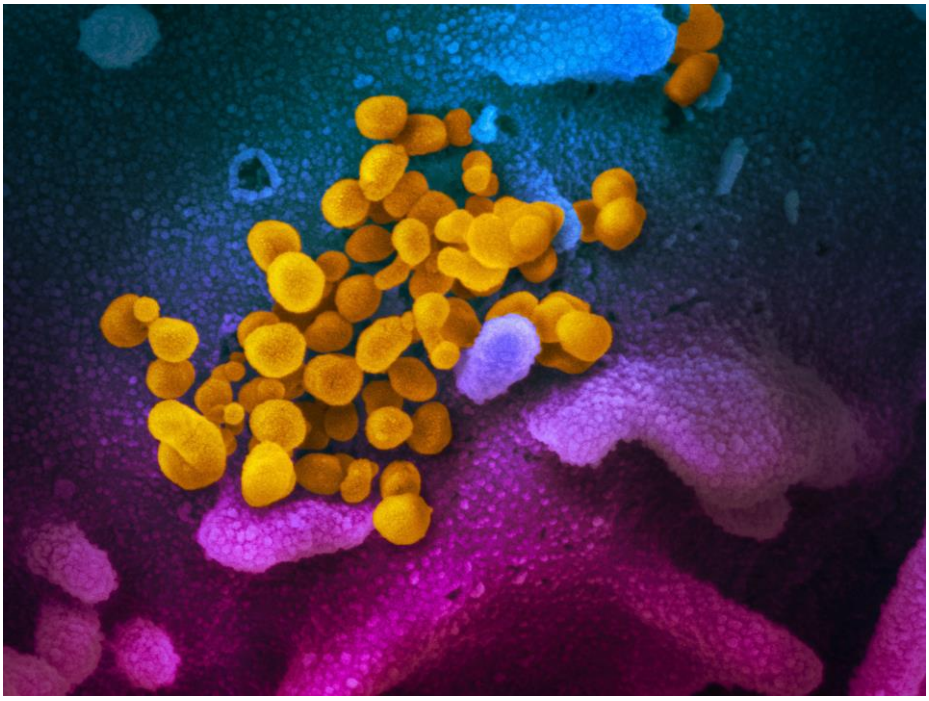

THE PRODUCTION OF RECOMBINANT 6LZE_A AND NS2B-NS3 IN E. COLI

Degree project in Biotechnology
2023

AUTHOR
Raluca Gabriela Svensson

SUPERVISOR
Javier Linares-Pastén

EXAMINER
Carl Grey

LUND UNIVERSITY
FACULTY OF ENGINNERING
DIVISION OF BIOTECHNOLOGY

Abstract

Even though the first coronavirus pandemic was registered in 1965, SARS-CoV-2 was better acknowledged when it caused the worldwide pandemic in 2019 in which millions of individuals lost their lives. Due to its facile way of spreading from one individual to another, the number of infected cases has gone up to more than 750 million to this day.

DENV is a very dangerous virus due to its asymptomatic profile in most of the infected cases. If left untreated it can lead to dengue hemorrhagic fever and for individuals who get infected with the virus for the second time, the consequences can be lethal.

Since there is currently no treatment for dengue infection, and due to the fast spreading of SARS-CoV-2 that still leads to deaths every day, it is important to understand how proteases are involved in viral replication and thus to find a way to stop this process in order to avoid spreading these viruses further.

In this report, 2 proteins (6LZE_A and NS2B-NS3) from these viruses were produced in *Escherichia coli* and their biological activity has been tested on DL-BAPNA substrate. Both of these proteins are structures of the main proteases that cleave the polyproteins, this step being very important in the replication process. Furthermore, computational modeling was used to understand the differences between their structures and how their catalytic mechanisms work. The results showed that working with these proteins has been proven difficult at times due to several failures and repetitions either because very few colonies grew on the Agar plates, cells died during the cultivation process, low concentrations, or failed purification. Hence, more time and research would be necessary to fully understand the reasons behind these issues. However, this report managed to highlight the different production times between the proteins and how SARS-CoV-2 seems to have better environmental stability than DENV due to its structure and docking characteristics.

Acknowledgments

This degree project was conducted in 2023 at Lund University, Faculty of Engineering, at the Division of Biotechnology.

First and foremost, I would like to express my gratitude to my supervisor, Javier Linares-Pastén, for accepting me as his student and for his helpful, calm, and promptitude character that he has. This project would not have been made possible without his aid and the support from other colleagues from the department which made me feel as if I was in a big family for a few months.

Second, I am very thankful to Carl Grey, for giving me a positive answer to my proposal of being my examiner, and for always being a nice and easy person to talk to and collaborate with. This journey would not have been possible without his implication.

Lastly, I will forever appreciate my family for their unconditional support and motivation during this entire experience, and my husband who made all of this possible by taking care of our baby so I can obtain my degree. Happily, my "mom brain" eventually faded and let me focus and use my cognitive functions.

Table of Contents

1 Background and aim

The SARS-CoV-2 outbreak started in Wuhan, China, in December 2019. From there, it spread worldwide from human to human resulting in the covid pandemic declared in March 2020 by the World Health Organization. The origin of the infection and the transmission pathways could not be definitively established but it is believed to be caused by exposure to infected animals due to its zoonotic profile [60].

The first coronavirus pandemic was registered in 1965, followed by another one in 2003 and 2012. Since this disease is considered highly infectious due to its rapid spreading and increased number of infected cases, it is classified as a Public Health Emergency of International Concern (PHEIC) [61].

To this day, there are over 270 million infected cases of SARS-CoV-2 registered in Europe and over 2 million deaths. Worldwide, there are over 750 million registered cases and almost 7 million deaths [62]. Research on how to inhibit coronavirus replication is still ongoing [17].

Dengue fever is caused by the DENV virus through mosquito bites. It is assumed that it was transmitted from nonhuman primates to humans approximately 1000 years ago in Africa or Asia [71] but the first major outbreak occurred in 1780 in Africa, Asia, and America [72].

Between 100 and 400 million infected cases of dengue fever are registered yearly in tropical or subtropical regions worldwide but the actual number can be even higher than that due to the asymptomatic tendency of the infection [73].

Since proteases are essential for viral reproduction because they help obtain functional proteins by cleaving the viral polyproteins pp1a and pp1ab [18], this master's thesis aim is to produce 2 recombinant proteases from SARS-CoV-2 (6LZE_A) and DENV (NS2B-NS3) in *E. coli* and test their biological activity on DL-BAPNA substrate which is a chromogenic substrate for proteolytic enzymes. When the bond between the arginine and the p-nitroaniline is hydrolyzed, the chromophore (p-nitroaniline) will be released and detected through colorimetric analysis. If the proteins are active they will hydrolyze the

substrate which will turn it from transparent to yellow. This substrate was previously tested on DENV but not on SARS-CoV-2, so it is expected that at least the NS2B-NS3 protease show activity on it. Furthermore, molecular modeling was performed as well to understand their structure and catalytic mechanism by observing the secondary structures of the proteins, the active site, the catalytic triad (for NS2B-NS3) or dyad (for 6LZE_A), and the docking with a Poly-histidine ligand and with the DL-BAPNA substrate.

The practical experiments were conducted at Lund University, Kemicentrum, Biotechnology Division supervised by Javier Linares-Pastén.

2 Introduction

2.1 SARS-CoV-2 – taxonomy and replication

SARS-CoV-2 is the virus that caused the worldwide pandemic in 2019 in which millions of individuals lost their lives. Anatomically, it is a single-stranded enveloped RNA virus, with a positive sense that belongs to the family *Coronoviridae,* genus *Betacoronavirus* and order *Nidovirales*, and has a diameter of approximately 65-125 nm [1]. Its name stands for severe acute respiratory syndrome-related coronavirus [2]. There are currently 39 species of coronaviruses, but only 3 of them stand out by having very high fatality rates. MERS (Middle East Respiratory Syndrome) has a fatality rate of 34.4% and appeared in 2012, SARS (Severe Acute Respiratory Distress Syndrome) has a mortality rate of 9.5% and emerged in 2003, but the coronavirus that caused the covid pandemic in 2019 has the lowest one of only 2.13% [3]. However, SARS-CoV-2 has an easier way of spreading from one individual to another compared to the other 2 types, which explains the increased number of infected cases of more than 750 million to this day [4].

Being a zoonotic virus, it can be transmitted from humans to animals and vice versa, and its origin is a horseshoe bat (*Rhinolophus*) from China. The virus has been further passed on from bats to civet cats,

raccoons, and badgers, and from there to humans [5]. House pets, zoo animals, and wild animals can also get infected [6].

The majority of the infected individuals present symptoms such as fever and cough, but approximately 15.6% are asymptomatic [7].

The virus spreads among individuals through mouth and nose droplets (from infected cases) that come in contact with the mucous membranes (eyes, mouth, nose) of healthy people. Touching a contaminated surface also disseminates the virus [8]. From here, the virus spreads to the rest of the body, especially in the lungs. Former studies showed that SARS-CoV-2 was found not only in the lungs of infected individuals but also in other organs, including the brain which was infected via olfactory nerves [9].

The replication process starts with the virus' S spike protein binding to a specific receptor (ACE2) in the host cell's membrane. Then transmembrane serine protease 2 (TMPRSS2) will cleave and activate the S protein followed by the release of the virus' large RNA into the cell which will recognize it as its own. The cell will then replicate the genetic material and further release it into new viral particles after self-destruct [10]. The process can be better observed in Figure 1.

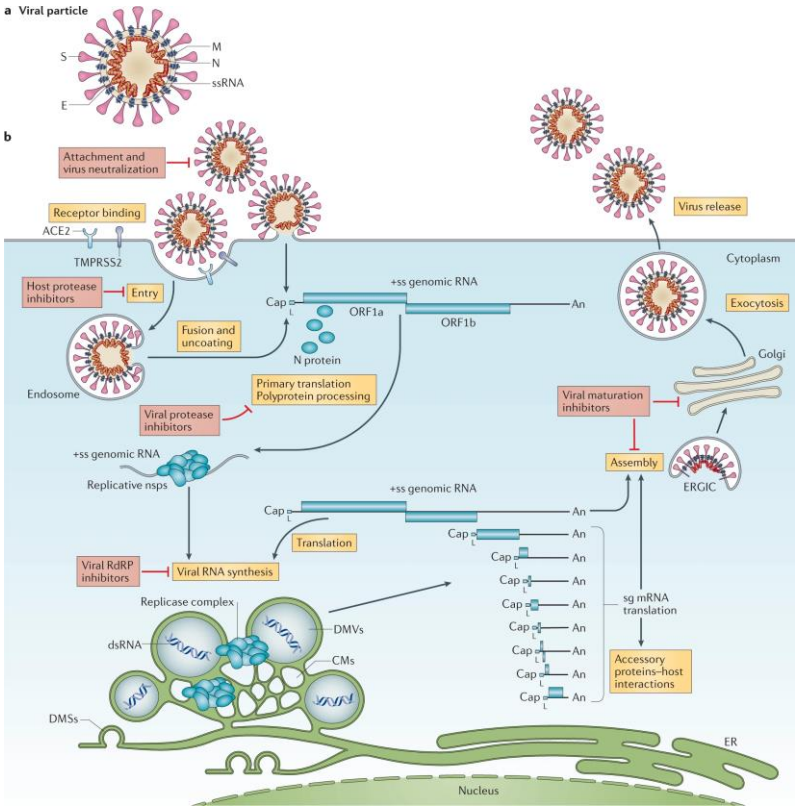

Figure 1. The replication process of SARS-CoV-2 (Source: [10])

2.2 Dengue virus – taxonomy and replication

Around 36,000 deaths are caused by dengue virus every year [73]. It is a single-stranded RNA virus with a positive sense that belongs to the genus *Flavivirus* in the family *Flaviviridae* and has a diameter of 50 nm. It was first isolated in 1943 by Ren Kimura and Susumu Hotta [71,73].

There are 4 serotypes of dengue virus (DEN-1, DEN-2, DEN-3, DEN-4), and they share the same genome in proportion of 65%. In 1970 DEN-1 and DEN-2 were predominantly found in America and Africa, while all 4 serotypes were found in Asia but nowadays they are all spread worldwide [71].

Even though the virus is transmitted to humans through female mosquito bites mainly from the *Aedes aegypti* species, pregnant women can also transmit the virus to the fetus resulting in low-weight newborns or pre-term births. However, only a few cases of maternal transmissions have been registered so far. Other rare and unusual transmission ways include blood transfusion and organ donation. Furthermore, humans can also infect mosquitos even if they are asymptomatic or pre-symptomatic [73].

Usually, the infection is asymptomatic but the most common symptoms include a 40°C fever, headache, nausea, rash, and eye and muscle pain [73]. If left untreated it leads to dengue hemorrhagic fever characterized by nose or gums bleedings and inability to breathe properly [74]. In severe cases, especially individuals who get infected with the virus for the second time there is a high risk of dying. For individuals who were already infected in the past, there is an approved vaccine available called Dengvaxia, but other possible vaccines are under evaluation. There is no current treatment for dengue fever but medicines such as Paracetamol are prescribed to alleviate the pain. [73].

The replication process starts with the virus binding to a skin cell. Then, the skin cell membrane traps the virus inside by creating a pouch around it called an endosome. Normally, cells use endocytosis in order to transport external molecules inside the cell for nourishment but DENV is using this process to infect the host cell. When the virus

gets deeper into the cell it is released into the cytoplasm resulting in the liberation of the viral genome that will then translate into a single polypeptide and replicate. After assembly on the endoplasmic reticulum, the particles mature through the trans-Golgy network where they become infectious cells that can further infect other cells [71]. The replication process can be seen in Figure 2.

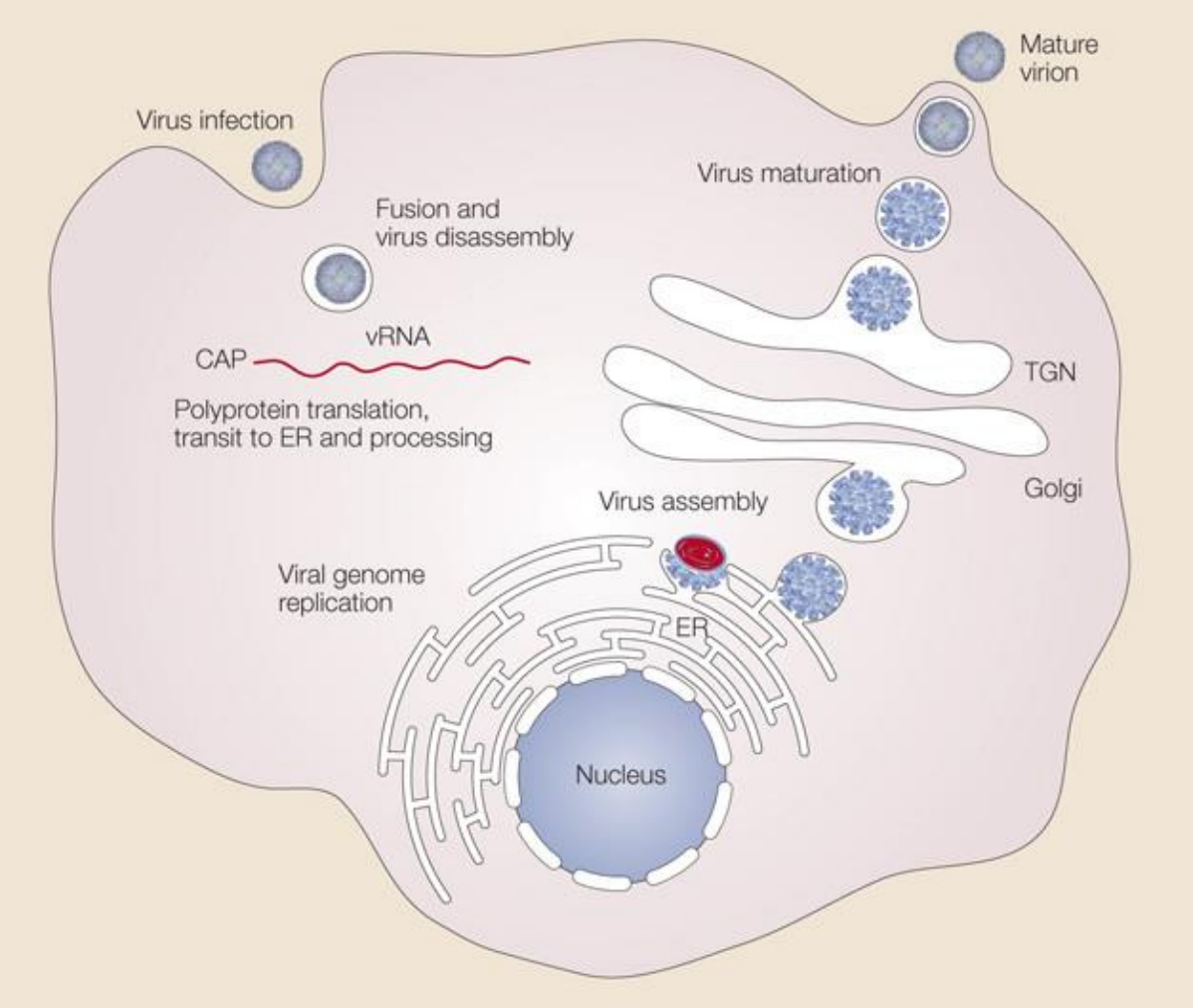

Figure 2. The replication process of the dengue virus (Source: Nature.com)

2.3 Proteases – sources, function, and importance

In 1825 the first protease was discovered, which was pepsin. Through the years, other enzymes were found, characterized, and crystallized, such as serine proteases (trypsin, chymotrypsin), cysteine proteases (papain), metalloendoproteinases (thermolysin), and carboxypeptidases A and B. A protease database called *MEROPS* offers information about more than a thousand peptidases and their inhibitors [11].

Different names for proteases are peptidases or proteinases, and they are responsible for the proteolysis reaction in which proteins break down resulting in smaller particles. They can be found in humans, animals, plants, and microbes [12].

They are classified into 6 classes: aspartic, serine, cysteine, threonine, glutamic, and metalloproteases but I will present in detail only the ones that can be found in viruses [13].

- Aspartic proteases are found in animals, plants, viruses, and fungi and they play a role in several physiological mechanisms such as digestion, pathogen protection, secondary malignant growths in breast cancer, regulation of blood pressure, and transformation of HIV regulatory proteins. They use a water molecule to cleave the protein bounds. Approximately 40% of the worldwide proteases come from microbes due to their diversity and proneness to genetic manipulation, and sources of aspartic proteases can be yeasts and molds. Some of their applications are in cheese production (coagulates milk), the baking industry (gluten helps the dough rise), the beer industry (haze removal), and the wine industry (maintaining its clarity) [14].
- Serine proteases are found in humans, animals, plants, fungi, bacteria, arthropods, helminths, protozoa, and viruses and they use a nucleophilic serine to break down peptide bounds. They are involved in a vast number of physiological and pathological mechanisms such as digestion, fertility, blood coagulation by snake venom, helping cells get infected by viruses, promoting tumor metastasis, and cartilage destruction in arthritis [15,16]. Their name comes from the fact that they use a serine residue to cleave

peptide bonds and they represent more than 30% of all proteases. Industrially, alkaline serine proteases are preferred, which are obtained through fermentation. They are mostly used as ingredients for detergents to help remove difficult stains (e.g. blood), but they also have applications in the leather industry (removing dirt and fat in the soaking stage), chemical industry (peptide synthesis), the medical industry (ointments, diagnosis, anticancer drugs), waste management (keratin degradation), the food industry (cheese, infant formulations, soft drinks, food for allergies and digestive disorders), and the silver industry (recovering silver from X-ray films). The production of alkaline serine proteases is mainly obtained from *Bacillus* sp. [21].

- Cysteine proteases or thiol proteases use nucleophilic cysteine thiol to break down proteins and the first cysteine protease discovered was papain from papaya [22]. They are found in all living organisms, but mainly in fruits (papaya, kiwi, pineapple) especially when they are unripe [23], and they represent the class of proteases most abundant in plants [24]. They are of high importance in different organisms: they play a role in plant growth, protein storage in seeds, and signaling pathways; in animals, they are involved in apoptosis, immune responses, and collagen degradation in atherosclerosis; some viruses (hepatitis C, polio) use them to cleave their polyprotein [25,26]. Industrially, they can be obtained in vivo or in vitro from plants (higher costs but standard parameters) or through traditional cultivations (cheaper but depends on environmental factors). Cysteine proteases have applications in the livestock feed industry (increases protein digestibility), and in the medical industry (edemas, infections, digestive disorders, inflammations, scar treatment, antibacterial and antifungal) [24].
- Glutamic proteases use a glutamic acid residue to break down proteins and they can be found mostly in filamentous fungi (*Ascomycota phylum*), but also in bacteria, viruses, and archaea [31,32,33]. They were first discovered in 2004 after previously thought to be aspartic proteases [34].
- Metalloproteases use a water molecule as a nucleophile and need a metal in order to catalyze hydrolysis of proteins [37]. They are found in animals, plants, viruses, archaebacteria, bacteria, and

nematodes [38]. They are important for cell proliferation and migration, tissue remodeling in the wound healing process, brain development, and are also involved in the growth of fetal organs [38,39]. Their applications include biomarkers for cancer and joint diseases [40,41]. Industrially, can be obtained by cloning human cancer tissue [42].

2.4 Protease inhibitors

The main role of proteases is to catalyze hydrolysis of protein bonds and they are responsible for numerous metabolic and regulatory functions in biological activities, thus disruptions in their proteolytic activity can lead to both diseases and their treatement [43]. Inhibitors hinder this process by binding to their active site making it possible to study their structure and mechanism. This can be useful when trying to find for example therapeutics in the medical industry for life-threatening diseases such as AIDS, arthritis, cancer, malaria, muscular dystrophy, SARS, and neurodegenerative diseases. Since viral proteases are responsible for obtaining the structural and non-structural proteins essential for viral replication, protease inhibitors that bind to the active site of the viral proteases and thus stopping the replication cycle have been used in medicine. They are also useful in agriculture as pest control agents which gives the benefit of reducing yield losses without using chemical agents. Commercially, they prolong the shelf life of food products and prevent proteolysis [44].

There are different kinds of protease inhibitors suitable for each class of proteases and their sources are microbial, animals, but mostly plants from their tubers and seeds (e.g. potamin-1 protein isolated from potatoes). Factors such as pH, structural stability, specificity, rate of binding, and cost are very important when choosing an inhibitor. When it comes to economical efficiency microbial sources are low-cost and have the advantage of growing fast in different environments and being receptive to genetic engineering. *Escherichia coli* can be used as a source of inhibitors against serine proteases, *Streptomyces* against serine, glutamic, and metalloproteases, *Aspergillus* or HIV for aspartic proteases, and *Saccharomyces* for

threonine proteases. The majority of inhibitors belong to the serine class [44].

According to previous research, Ebselen is a compound capable of inhibiting proteolysis of the main cysteine protease (Mpro) from SARS-CoV-2 by binding to the active site residues His41 and Cys145. Furthermore, compounds such as E04 and E24 which are pyrazole derivatives have also proved to inhibit viral replication [45]. Sitagliptin and Daclastavir HCI are suitable inhibitors for papain-like protease (PLpro), although the latter has decreased binding capabilities. Atazanavir, Nelfinavir, Lycorine HCI, MG-101, and Lomibuvir have antiviral properties towards Mpro. Among these, the most potent one is Lycorine which is found in the *Amaryllidaceae* plant [46].

When it comes to DENV there are a few compounds capable of inhibiting the viral protease such as diaryl (thio)ethers (especially for type 2) [30], and MB21 and Aprotinin that can be used for all 4 types by showing strong binding capabilities close to the active site [35]. Agathisflavone has proved to be the best protease inhibitor from a natural source. Furthermore, Hesperitin from *Ganoderma lucidium Antler* variant can inhibit the protease from type 2. When trying to find suitable protease inhibitors it is worth mentioning that the NS2B-NS3 protease has an affinity for substrates with dibasic residues (2 replaceable hydrogen atoms per molecule) at P1 and P2 sites such as lysine or arginine. P1 and P2 are substrate peptides sequences located between the cleavage site and the N- terminal [97]. Even though there is a significant number of protease inhibitors identified through the years, they have not entered clinical trials [36].

2.5 Recombinant protein production

The process of producing recombinant proteins involves modifying a protein by manipulating its DNA using an expression vector [47]. The desired gene is cloned in a vector, and after alteration inside the host, the target protein is obtained through heat shock where due to the high temperature small pores are caused in the competent cell's membrane allowing the plasmid to enter, followed by cell growth on

Agar plates, small-scale cultivation and finally purification. A representation of the process can be observed in Figure 3 [92].

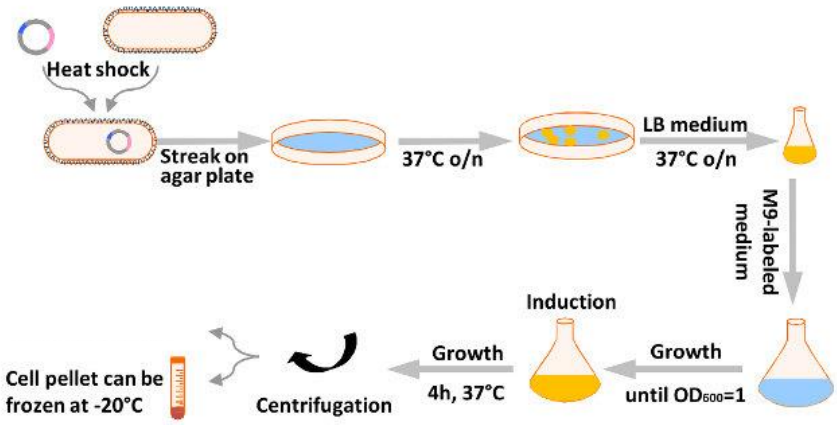

Figure 3. The main steps of recombinant protein production (Source: [101])

The expression vector contains the gene of interest, promoters, affinity tags, a coding sequence for tag removal, selection marker, and cloning sites. Plasmids contain replicons and it is preferred to choose high copy numbers to obtain increased production since replicons are DNA regions that are replicated from a sole replication origin. Promoters are also DNA regions that are responsible for initiating gene transcription through the binding of transcription factors and RNA polymerase. There are 2 types of promoters, constitutives (always active) and inducible (can be turned on and off in certain conditions). The most used promoter due to its efficiency is T7 which is induced with IPTG for expressing proteins in *E. coli*. Affinity tags such as His-tag are small peptides that increase the solubility of recombinant proteins, help the purification step, and can be easily detected during protein expression. They can be specific to either N- or C-terminus. They need to be removed from the desired protein to make sure it will not hinder its activity using tag removals which can be enzymes (His-tag TEV) or chemicals (CNBr). To confer specificity, antibiotics are used as selection markers with resistance for the plasmid used [49].

In 1978 was the first time when *Escherichia coli* was genetically modified in order to produce insulin [48]. Since then, *Escherichia coli* has become the preferred host due to its low cost, rapid growth, and manipulation ease. Before this revolutionary discovery, proteins were obtained and purified using animal and plant tissues [49,50]. Besides

bacteria, mammalian and insect cells, yeast, and plants, can also be used as hosts in industrial recombinant protein production [51].

Since success depends on a variety of factors, there are some challenges when it comes to the entire process, and these are a few of the possible errors:

- Low or no production of the target protein (due to toxic proteins)
- Accumulation of inclusion bodies and wrong disulfide bonds (due to incorrect folding)
- Obtaining a low-quality protein production (due to low solubility)
- Contamination from the environment, handling, or of the media
- Wrong cultivating conditions (pH, temperature, CO_2 supply)
- Post-translational modifications [49]

2.6 Applications in biomedicine

Proteases are very important in viral replication by breaking down proteins resulting in functional proteins so research on how to block or inhibit them is of great interest in stopping viral reproduction [93]. This is a complex process. For SARS-CoV-2 this starts with the histidine molecule from the catalytic dyad removing a proton from the cysteine molecule. Then, the main viral protease (Mpro) is autoactivated and cleaves the pp1a and pp1b polyproteins at 11 sites which produces mature viral proteins which are proteins present in the mature assembled virus particle [99]. For DENV, the NS3 protease together with its cofactor NS2B converts the polyproteins into functional proteins that will produce 10 viral mature proteins [100].

Recombinant proteins are of great importance in biotechnology. In research, they are involved in vivo and in vitro studies including cell biology, molecular biology, biochemistry, and inhibitors studies in pharmacology [52]. They are used to observe how proteins interact with each other, and they can also reveal the functions of certain genes. Techniques such as ELISA, Western Blot, and immunological assays use recombinant proteins or antibodies [53].

In biomedicine, they have applications in diagnostics and therapeutics using monoclonal antibodies that can replace insulin (by interacting with a MHC class II molecule called IA^{g7}) or as helping agents for

treating diseases such as hepatitis [52, 98]. They are also used to detect viruses and assess vaccines' efficiency. This is based on antibody detection against specific viruses using ELISA and specific and rapid kits can be purchased commercially for this purpose [54]. Using recombinant vaccines instead of other types of vaccines can solve some challenges such as difficulties in integrating the infectious agent in attenuated or inactivated vaccines, low immune response (inactivated vaccines), and health problems for weakened immune system patients (live-attenuated vaccines). For the production process *Escherichia coli*, insect or mammalian cells, or yeasts can be used as hosts. Some of the recombinant vaccines approved today include antiviral agents against Hepatitis B, Influenza, Malaria, SARS-CoV-2, Varicella, and Papilloma [55,56].

There are currently over 170 recombinant proteins worldwide that are used in biomedicine including growth factors, interferons, hormones, blood clotting agents, and enzymes (diabetes, multiple sclerosis, heart diseases, cancer, rheumatism, anemia, hepatitis, psoriasis, cystic fibrosis) [57]. These recombinant proteins need to be of high quality otherwise using unfitting cell lines can lead to severe side effects and unwanted immune responses [52].

In Sweden, *Protein Production Sweden* (PPS) is a national organization established in 2022 that produces and purifies different types of recombinant proteins for all researchers around the country, and it has 7 laboratories located at several Swedish universities [58].

There are optimistic prospects involving recombinant proteins that can cure melanoma, and lung cancer, or fight obesity (FGFBP3) and researchers are working on enhancing their activity [52,59].

3 Materials and methods

Two practical experiments were carried out during this report, one before the summer vacation and one after. Because of this, the ”Materials and methods” and ”Results” chapters are split into 2 parts mainly because they have slightly different parameters and also for a better understanding of the specific proteins that I am referring to in the report.

3.1 Production of the recombinant protein in E. coli – part 1

I will present a summary first and then I will go into details on specific chapters.

The experiment was carried out using synthetic genes from GenScript that were already inserted into the plasmid. The viral proteases used were 6LZE_A from coronavirus, and NS2B-NS3 from dengue virus since these are the main proteases essential for viral replication.

To insert the pET21b plasmid into the cells non-competent cells have to be transformed into competent cells. This is done through thermal shock which is a physicochemical method in which the freeze-dried tubes containing the synthetic genes are heated for a few minutes and then cooled down again. This process helps the molecules to resuspend [65].

The plasmid containing the synthetic gene and ampicillin-resistant gene (for selecting the transformed cells) was inserted in 2 types of competent cells: *E. coli* DH5 alpha is used for storing and propagating the plasmid, and *E. coli* BL21(DE3) for protein production. To introduce the plasmid into the cell they underwent thermal shock. After adding LB broth the cells were incubated on Agar plates, 1 for each *E. coli* strain used. One colony from each plate was incubated overnight and after inoculating in the cultivation media together with ampicillin the OD (optical density) was measured every hour using a spectrophotometer to evaluate the cell density. When it reached between 0.6 and 1 IPTG (*Isopropyl β-D-1-thiogalactopyranoside*) was added to start the induction of protein expression by binding to the lac repressor [64].

Spectrophotometry is a quantitative analysis that measures how much light a substance absorbs at a certain wavelength. This is done by passing a beam of light through a prism and then through the sample. A spectrophotometer is mainly composed of a light source, monochromator, sample holder, detector, and interpreter [63].

The protein purification of the samples was done using an ÄKTA Pure Purification System through immobilized metal chelate affinity chromatography (IMAC) which is using a chemical process called

chelation for immobilizing metals on a chromatographic medium. It can purify proteins, peptides, and nucleic acids. Its components mainly include inlet and outlet valves, 2 pumps (for continuous and accurate flow rates), a mixer (for homogenous buffer composition), a UV monitor (measures the absorbance), a conductivity monitor (measures buffer conductivity), and a fraction collector. It is connected to the UNICORN software where the results are shown. The purification chromatogram is supposed to show 2 peaks, the first one containing the elute unspecific protein, and the second one the elute target protein from which these specific tubes should be collected and analyzed further through SDS-PAGE to verify their purity [66].

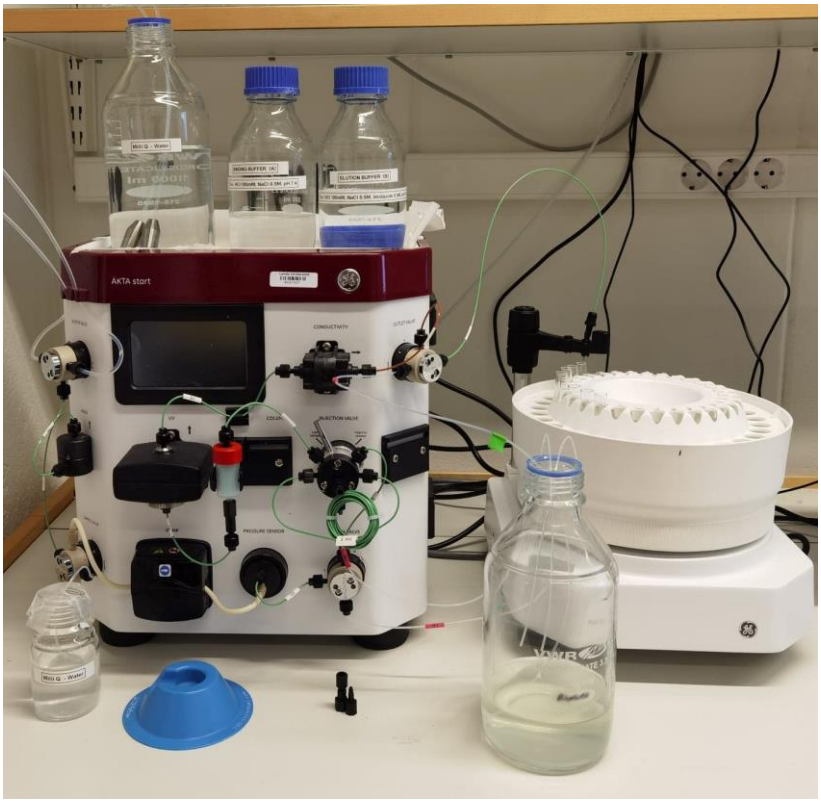

Figure 4. ÄKTA Pure Purification System

The protein concentration was assessed at A_{280} (for proteins) with Nanodrop which is a UV-Vis spectrophotometer that can measure sample volumes as small as 2 µL with high accuracy [69].

Figure 5. Nanodrop

SDS-PAGE (polyacrylamide gel electrophoresis) is an analytical method in which the proteins are separated based on their molecular weight. SDS (sodium dodecyl sulfate) is a detergent that denatures proteins and by being negatively charged it will cover the proteins in a negative charge as well after binding to their positive charge. Then, the proteins will migrate to the positive electrode through the gel matrix. The migration rate of proteins through the gel matrix depends on their charge, structure, and mass. Smaller proteins will migrate faster than larger ones. If the sample is pure only 1 band is going to be obtained according to its specific molecular mass [67].

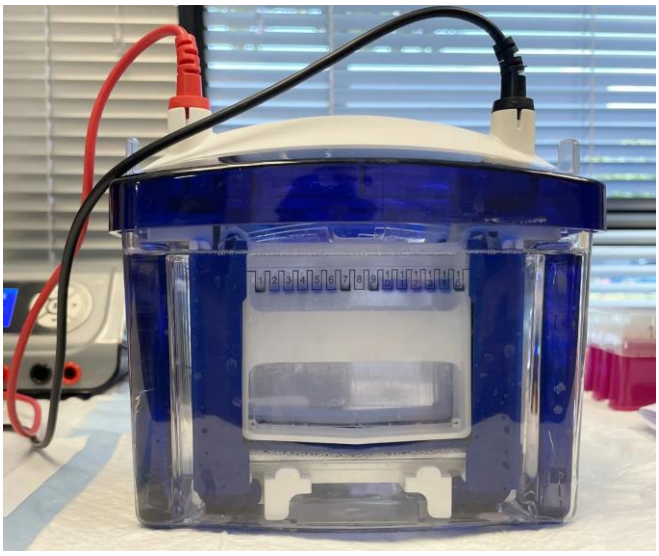

Figure 6. SDS-PAGE (Source: azurebiosystems.com)

To concentrate the samples a Pierce protein concentrator from Thermo Fisher was used.

Figure 7. Pierce protein concentrator (Source: www.fishersci.no)

The assessment of the biological activity of the compounds was done through Multiskan Go which is a microplate spectrophotometer that can test multiple samples in a time-efficient manner by measuring the absorbance in photometric assays [70].

Figure 8. Multiskan Go

3.1.1 Preparation of the competent cells

The competent cells were already prepared by other colleagues as follows: the tube with cells in the cultivation media was centrifuged and the LB was discarded. After adding 1.5 ml of $CaCl_2$ and glycerol (cryoprotectant) at a final concentration of 20% 150 µL were transferred to different tubes and stored at -80°C. The entire process was done on ice. These cells are now ready to receive plasmids for the recombinant protein transformation.

3.1.2 Thermal shock transformation

The freeze-dried tubes containing the synthetic genes were centrifuged at 5000 rpm for 1 minute to avoid losing all the molecules that could have ended up underneath the lid. After adding 50 µL of sterile Milli-Q water the tube was heated in a thermoblock at 50°C for 15 minutes and then transferred to ice. 2 µL of the sample was used to check the stock concentration in a Nanodrop at 280 nm. The result for 6LZE_A was 198 ng/µL and for NS2B-NS3 131 ng/µL. The samples were then diluted as follows: for 6LZE_A 20 µL of the plasmid was mixed with 40 µL of autoclaved Milli-Q water, and for NS2B-NS3 10 µL of the plasmid was mixed with 10 µL of Milli-Q water. The results were then 8.7 ng/µL and 14.2 ng/µL respectively.

The tubes containing 25 µL of *E. coli* BL21(DE3) competent cells were defrosted on ice from -80°C to 0-4°C and then 2 µL of the plasmid solution was added. The plasmid and the cells were carefully mixed using a micropipette. The tube was incubated on ice for 15 minutes, then transferred to a thermoblock at 42°C for 2 minutes, and then cooled again on ice for 20 minutes.

300 µL of LB broth (Luria Bertani) was then added and incubated at 37°C for approximately 50 minutes. 100 µL of this sample was transferred to 3 agar plates (with ampicillin) for each gene: 1 for *E. coli* DH5 alpha, 1 for *E. coli* BL21(DE3), and 1 for *E. coli* BL21(DE3) commercial variant, and then incubated overnight at 37°C.

3.1.3 Protein production

The next day only 4 plates out of 6 grew colonies, the BL21(DE3) commercial variant did not grow any colonies from either of the genes. Because of this, only 4 tubes were prepared, 1 per growing plate in which it was added 6 ml of LB broth and 6 µL of ampicillin and they were inoculated with 1 colony from each plate and left to incubate overnight at 37°C 200 rpm.

On the third day at 9:10, 2 Erlenmeyer flasks were used and in each of them were added 300 ml LB, 300 µL of ampicillin and 3 ml of *E. coli* BL21(DE3) from each viral strain. Then they were put in a shaker at 37°C and 180 rpm and the OD measurements were taken after 1 hour using a UV-Vis spectrophotometer at 600 nm single wavelength. When the OD reached 0.6 IPTG was added to start the induction. Since the lag phases were different between the proteins, the IPTG was added at different times.

At 12:25 300 µL of 1 mM IPTG was added to the NS2B-NS3 flask and moved to a shaker at room temperature (20°C) overnight.

At 15:30 IPTG was added into the 6LZE_A flask and moved to room temperature overnight.

The next day around 8:30, the samples were moved into centrifugation bottles with a capacity of 250 ml each and centrifuged for 10 minutes at 6000 rpm. The cell pellets were then collected and resuspended in 15 ml of binding buffer and transferred to falcon tubes which were further sonicated on ice for 15 minutes at 60 amplitude and 0.5 cycle. The samples were centrifuged again for 20 minutes at 12000 rpm and the supernatant was collected and kept at -4°C in falcon tubes until the next day.

3.1.4 Protein purification

To purify the samples, ÄKTA Pure Purification System was used based on affinity chromatography. The system was washed beforehand and in-between samples with Milli-Q water, and then the samples were purified. After measuring the concentration in Nanodrop, the samples were further stored in the fridge at 4°C.

3.1.5 Protein concentration

Since the protein concentration of the purified 6LZE_A sample was so low it was further concentrated using a "Pierce" protein concentrator (10 000 MWCO). It was then centrifuged for 10 min at 4500 rpm, and the concentration was assessed after another 15 minutes. The concentration in the top part of the tube was 0.71 mg/ml and in the bottom 0.15 mg/ml so this part was discarded and only the top was stored further at 4°C.

3.1.6 SDS-PAGE electrophoresis

In order to assess the purity of the compounds the SDS-PAGE method was used. This procedure was done together with other PhD students as follows:

16 µL of the purified samples were mixed with 4 µL of the loading buffer in Eppendorf tubes that were further heated for 10 minutes at 100°C. Then, 10 µL of the samples were loaded into the gel at 180V for 15 minutes and then another 30 minutes at 200V. The gel was carefully removed and washed with water and Coomassie brilliant blue was added as a staining agent. After leaving the gel with the staining agent for an hour destaining solution was used and left for a few hours before viewing the gel.

3.2 Production of the recombinant protein in E. coli – part 2

As stated previously, the experiment was redone after 2 months due to summer vacation.

3.2.1 Preparation of the competent cells

I prepared new competent cells together with other colleagues using the same method as described in the first part. However, this was only for learning purposes, and they were not used for the experiment.

3.2.2 Thermal shock transformation

Since there were no more colonies for 6LZE_A the entire process was redone but with slightly changed parameters. First, LB broth and agar plates were prepared and autoclaved. Then, the protein concentration of the 6LZE_A stock was evaluated with Nanodrop at 280 nm and the result was 19.9 ng/µL. A tube with *E. coli* BL21(DE3) competent cells was acquired from a Ph.D. student and after defrosting on ice 2 µL of the plasmid solution was added. After the plasmid and the cells were carefully mixed the tube was incubated on ice for 30 minutes, then transferred to a thermoblock at 42°C for 90 seconds, and then cooled again on ice for 10 minutes. 300 µL of LB broth (Luria Bertani) was then added and incubated at 37°C for approximately 50 minutes. The solution was further transferred to 2 agar plates containing ampicillin, approx 150 µL for each, and incubated overnight at 37°C.

3.2.3 Protein production

For 6LZE_A 2 tubes were prepared for each plate in which it was added 6 ml of LB broth and 6 µL of ampicillin and they were inoculated with 1 colony from each plate and left to incubate overnight at 37°C 200 rpm.

The next day 2 Erlenmeyer flasks (1 L) were used and in each of them were added 300 ml LB cultivation media, 300 µL of ampicillin in a concentration of 100 mg/ml and 3 ml of *E. coli* BL21(DE3) from each tube. They were put in a shaker at 37°C and 200 rpm and the OD measurements were taken after 30 minutes or 1 hour using a UV-Vis spectrophotometer at 600 nm single wavelength.

When the OD reached 0.6, 300 µL of 1 mM IPTG was added to the flasks and they were moved to a 200 rpm shaker at room temperature (20°C) overnight.

Glycerol stocks were also made by combining the 2 tubes with 6LZE_A resulting in 8 cryotubes that were further stored at -80°C in the 8th tower, box 3. This was obtained by adding 2 ml of Glycerol 80% into the remaining 6 ml of *E. coli* BL21(DE3) 6LZE_A by carefully mixing

them by hand for a few minutes before being transferred to the cryotubes.

The calculation formula for finding out how much Glycerol 80% to add was:

Ci: Glycerol 80%

Cf: Glycerol 20%

V: amount of sample in ml (6 ml)

Vi=(V x Cf)/(Ci-Cf)

For NS2B-NS3 the 2 plates with *E. coli* BL21(DE3) and *E. coli* DH5 alpha obtained in the first part were used to prepare 2 tubes with inoculum. These contained 6 ml of LB broth and 6 µL of ampicillin each and they were inoculated with 1 colony from each plate and left to incubate overnight at 37°C 170 rpm.

On the next day, 1 LB cultivation media flask was used containing 300 µL of ampicillin and 3 ml of *E. coli* BL21(DE3). Then they were put in a shaker at 37°C and 180 rpm and the OD measurements were taken after 1 hour using a UV-Vis spectrophotometer at 600 nm single wavelength.

The culture was eventually discarded due to the cells dying after 1 hour, and the process was redone. However, the samples left in the tubes were used for making glycerol stocks as follows. From the *E. coli* BL21(DE3) tube, 3 ml was left. Then 1 ml of Glycerol 80% was added and mixed carefully by hand for a few minutes and further transferred to 4 cryotubes each containing 1 ml and stored at -80C. From the *E. coli* DH5 alpha tube, 6 ml was left so 2 ml of Glycerol 80% was added and formed 8 cryotubes. The cryotubes were labelled accordingly with the strain name and date.

The next day, a glycerol stock tube with *E. coli* BL21(DE3) NS2B-NS3 was defrosted on ice and afterward, I added in a falcon tube 6 ml sterile LB broth, 6 µL ampicillin, and the content of the glycerol tube (approx 1 ml) while working in the sterile bench. This tube was further kept overnight at 37°C and 200 rpm.

In an LB cultivation media flask were added 300 µL of ampicillin and 3 ml of *E. coli* BL21(DE3) from the falcon tube. Then, it was put in a shaker at 37°C and 180 rpm and the OD measurements were taken every 30 minutes.

The rpm was changed from 180 to 200 after 1 hour and a half due to noticing that the culture was decreasing. IPTG was added to start the induction process and the flask was kept overnight at 20°C and 200 rpm.

The centrifugation and sonication process for 6LZE_A and NS2B-NS3 were done following the same method as previously described and the supernatant was collected and kept at -4°C in falcon tubes.

3.2.4 Protein purification

Only the NS2B-NS3 sample was purified because there was no Sodium Chloride (NaCl) at that moment so the Binding Buffer for the ÄKTA purification system could not be prepared.

3.2.5 Protein concentration

Since the 6LZE_A sample could not be purified at that moment and the NS2B-NS3 sample showed a negative value on the Nanodrop, the NS2B-NS3 produced in the first part was desalted and concentrated.

ÄKTA was also used for desalting and Phosphate Buffer and HCl 1 mM were prepared for this. The method for desalting was selected from the program and a desalting column was used instead. The sample was injected in the column using a 1.5 ml syringe and the process was repeated until the entire sample was desalted.

The Nanodrop concentration read at 280 nm was 0.35 mg/ml after desalting.

The sample was then concentrated as previously mentioned and the final concentration was 0.50 mg/ml.

3.2.6 SDS-PAGE electrophoresis

Since the purification step could not be performed on the 6LZE_A sample the SDS-PAGE assessment was also skipped.

3.3 Evaluation of the biological activity of the compounds

3.3.1 Enzyme activity assessment on DL-BAPNA substrate

DL-BAPNA (Nα-Benzoyl-DL-arginine 4-nitroanilide hydrochloride) substrate solution was prepared at a concentration of 10 mM. 10 mg of the powdered substrate was weighed, added to 2.3 ml of Milli-Q water, and vortexed. To ease the dissolving process 2.3 ml of DMSO solvent was added as well so the concentration changed from 10 mM to 5 mM.

If the formation of trypsin/substrate complex takes place then the bond between arginine and p-nitroaniline is hydrolyzed, resulting in the release of the chromophore (p-nitroaniline) which will turn the sample yellow and if there is no formation due to the presence of a competitive trypsin inhibitor the sample remains transparent (Figure 9).

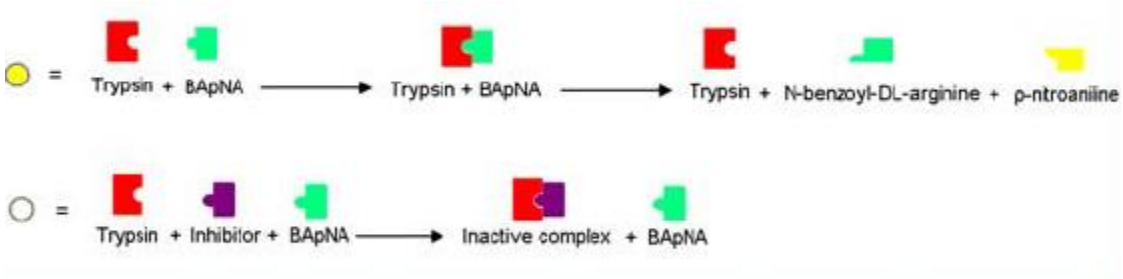

Figure 9. DL-BAPNA reaction (Source: [79])

To test the enzyme activity of the samples obtained in the first part 3 of the 96 wells of a microplate were used. All 3 of them contained 100 µL Phosphate buffer 100 mM pH 7.1, 50 µL of the substrate, 50 µL of Milli-Q water, and 10 µL of the sample. The first slot (1A) contained the 6LZE_A purified and concentrated sample, the second slot (2A) the NS2B-NS3 purified sample, and the third one was for control. The plate was then analyzed in a Multiskan go at 37°C, with 10 readings at 10-second intervals (400 nm wavelength). The results showed that the enzymes had no activity on this substrate.

The NS2B-NS3 that was desalted and concentrated in the second part was also evaluated using the same protocol and the result showed there was no activity. The substrate volume was then changed from 50 µL to 100 µL and the microplate was covered with aluminium foil and kept in an incubator at 37°C for an hour. The readings after that still showed no activity.

3.4 Molecular modeling

The protein sequences were provided by GenScript together with the synthetic genes and they are as follows:

> NS2B-NS3
GSHMLEADLELERAADVRWEEQAEISGSSPILSITISEDGSMSIKN
EEEEQTLGGGGSGGGGAGVLWDVPSPPPVGKAELEDGAYRIKQK
GILGYSQIGAGVYKEGTFHTMWHVTRGAVLMHKGKRIEPSWADV
KKDLISYGGGWKLEGEWKEGEEVQVLALEPGKNPRAVQTKPGL
FKTNTGTIGCVSLDFSPGTSGSPIVDKKGKVVGLYGNGVVTRSGA
YVSAIANTEKSIEDNPEIEDDIFRK

> 6LZE_A
SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVIC
TSEDMLNPNYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLK
LKVDTANPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRP
NFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTD
LEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFL
NRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDM
CASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGV

These sequences were copied in the NCBI BLAST PROTEIN database in order to obtain the PDB (Protein Data Bank) Accession codes for the proteins with 100% similarity or close to. NCBI BLAST is a tool used for comparing sequences between them through a database by calculating the matching percentage [75]. RCSB PDB is a global database for 3D structures of large biological molecules [76]. The Accession codes were then used in the Chimera software to visualize the 3D structures of the proteins, to identify the secondary structures, the active site together with the triad (for NS2B-NS3) or dyad (for 6LZE_A), and to observe the docking with the ligand. Since the structure for NS2B-NS3 was incomplete the prediction of the

incomplete region was done through AlphaFold linked to ChimeraX which is a tool that can build structures and analyze them [77]. Further, the ligand was built in Yasara with 6 histidines and then the docking was visualized in Chimera. The protein structures were docked with the DL-BAPNA substrate as well which was downloaded from NCBI. Yasara is a program used for visualizing and simulating molecules [19].

4 Results

4.1 Production of 6LZE_A and NS2B-NS3 – part 1

4.1.1 Protein production

The results for the cell growth on Agar plates for 6LZE_A can be seen in Figure 10 for 6LZE_A and for NS2B-NS3 in Figure 11. The plate for *E. coli* BL21(DE3) for 6LZE_A had only 2 colonies and for *E. coli* DH5 alpha had 15. The cells had a better growth rate for NS2B-NS3 than for 6LZE_A as seen in the pictures. However, the commercial variant did not grow any colonies on either of the proteins, so these ones were discarded.

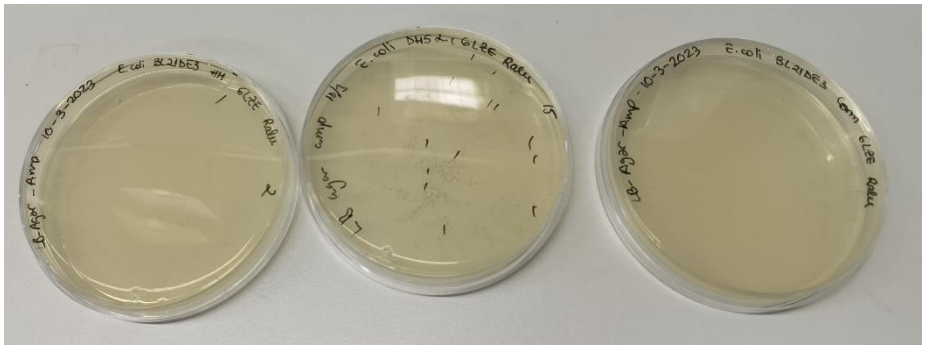

Figure 10. Cell growth for 6LZE_A

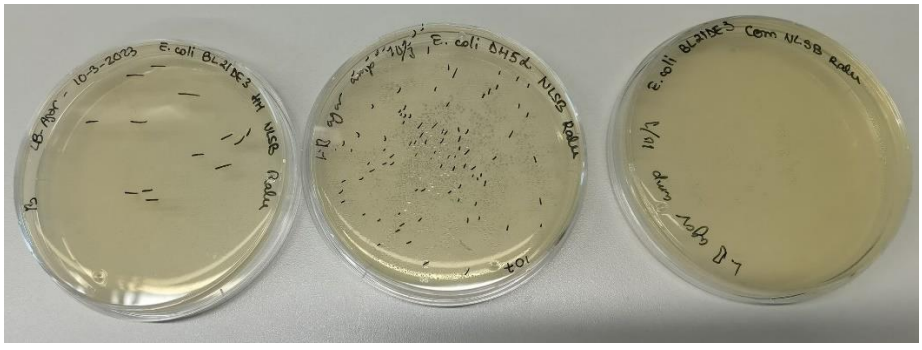

Figure 11. Cell growth for NS2B-NS3

The lag phase was very different between the cultivations. For 6LZE_A the lag phase took approximately 3.5 hours which is unusually long compared to 1 hour for NS2B-NS3 as seen in Figure 12.

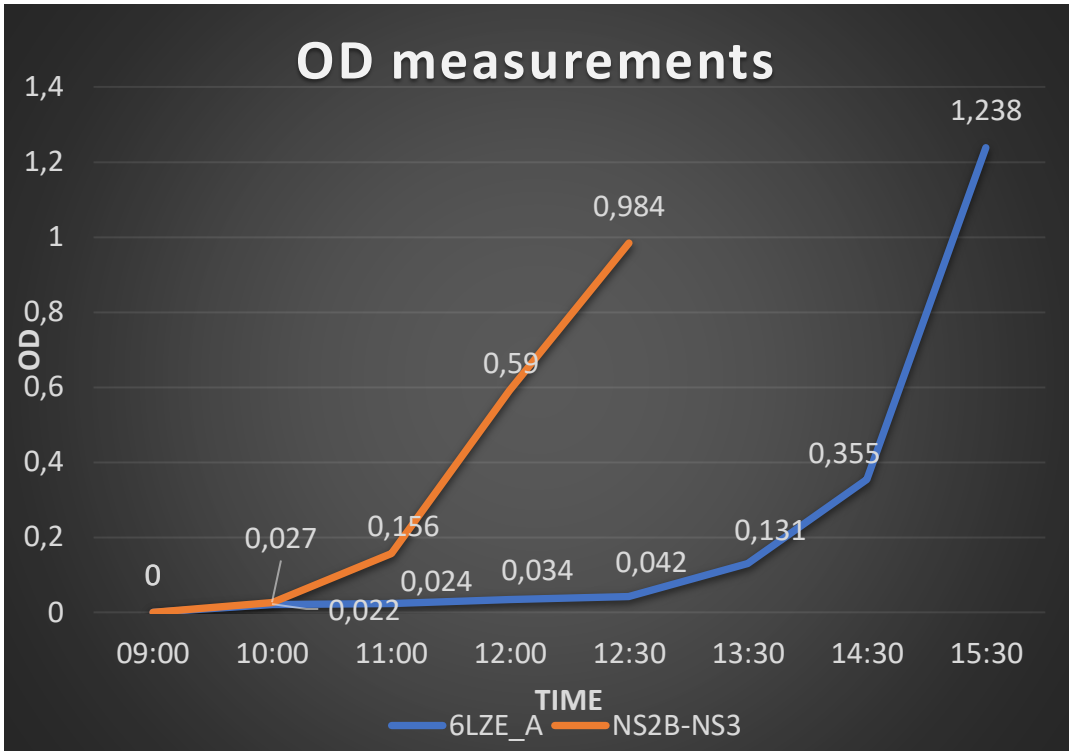

Figure 12. OD measurements for 6LZE_A and NS2B-NS3

4.1.2 Protein purification

The results for protein purification differed between the 2 proteins.

The 2 peaks shown in the purification chromatograms represented the unspecific proteins (the 1st peak), and the target protein (the 2nd peak).

For 6LZE_A the peak corresponding to unspecific proteins was higher than the one for the target protein as observed in Figure 13. The 8th and the 9th tubes were collected for further analysis and the rest were discarded. This sample was also in a very low concentration as seen in Figure 15 which is why it was then concentrated before being stored in a cold environment.

For NS2B-NS3 the target protein was in a higher concentration than the unspecific proteins as seen in Figures 14 and 15. Tubes 5-10 were collected further.

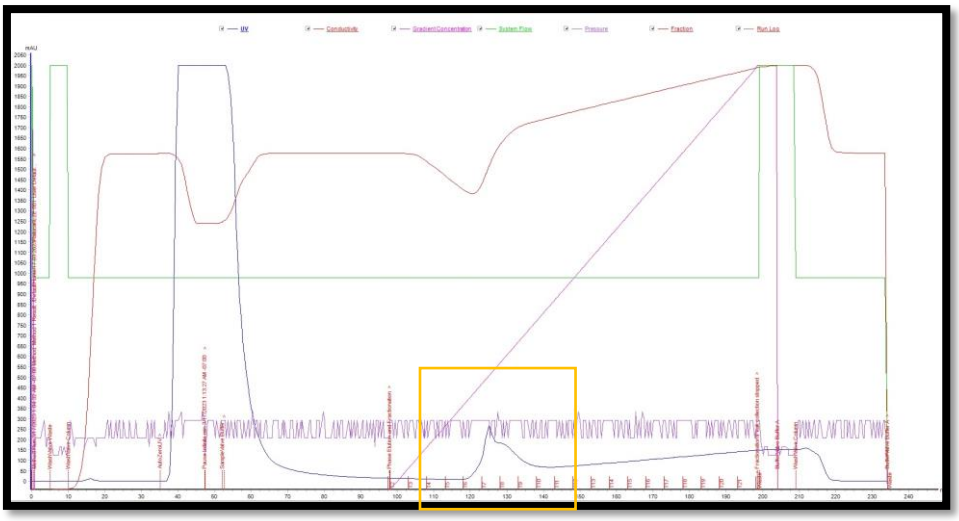

Figure 13. Purification chromatogram for 6LZE_A

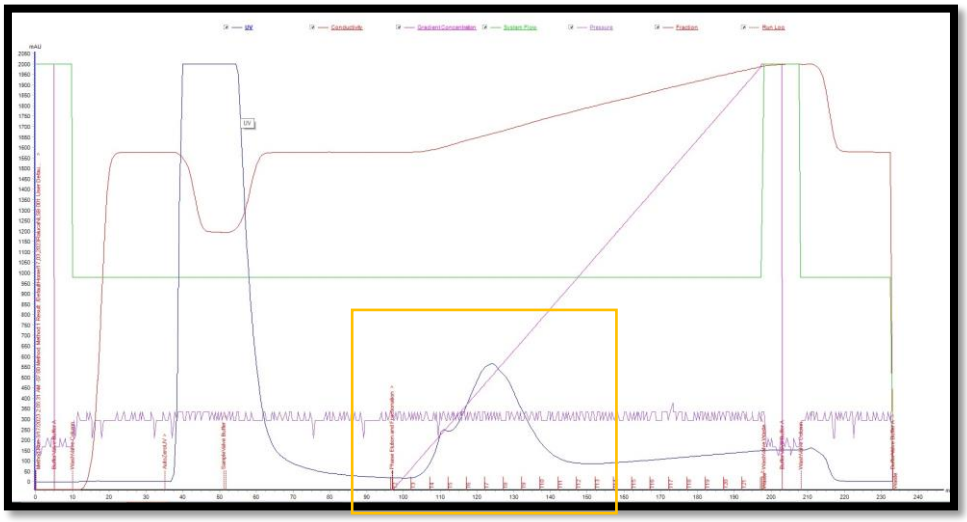

Figure 14. Purification chromatogram for NS2B-NS3

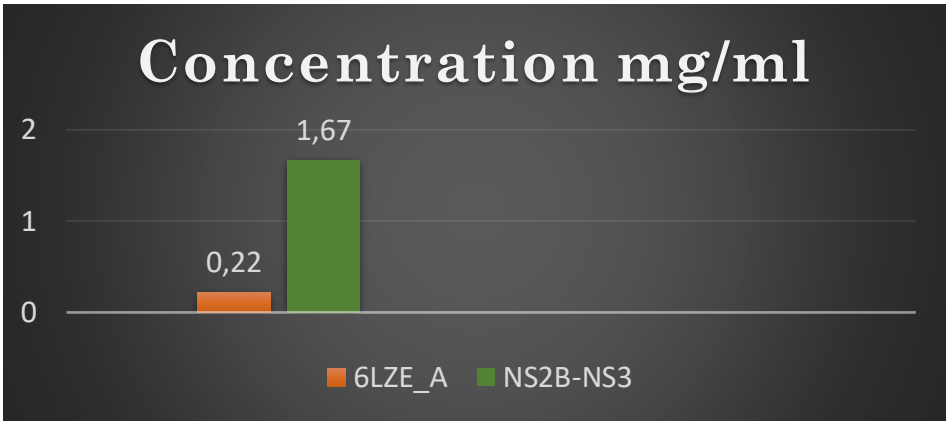

Figure 15. Protein concentration after purification

4.1.3 SDS-PAGE electrophoresis

The SDS-PAGE gel result can be observed in Figure 16. The molecular weights of both proteins seem to be similar around 30 kDa. The 6LZE_A band is fainter than the one for NS2B-NS3 but it also has less contamination than the other one, which seems to have several contaminants of different molecular weights.

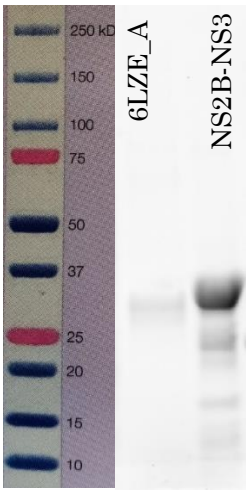

Figure 16. SDS-PAGE for 6LZE_A and NS2B-NS3

4.2 Production of 6LZE_A and NS2B-NS3 – part 2

4.2.1 Protein production

The cell growth for 6LZE_A on Agar plates can be observed in Figures 17 and 18. Compared to the first part the colonies were in a much higher number on both plates.

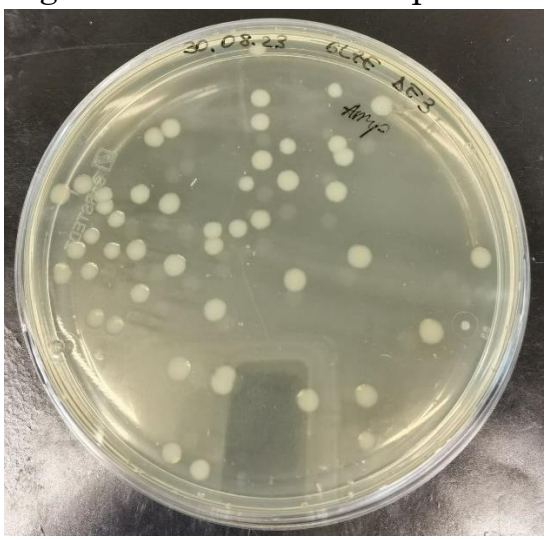

Figure 17. Cell growth 6LZE_A nr. 1

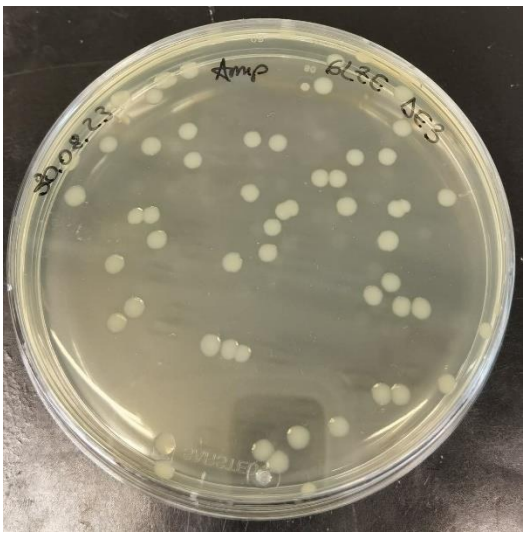

Figure 18. Cell growth 6LZE_A nr.2

The production times for both proteins followed the same curve as in the first part as seen in Figures 19 and 20 which was characterized by lag phases lasting 3.5 hours for 6LZE_A and 1 hour for NS2B-NS3.

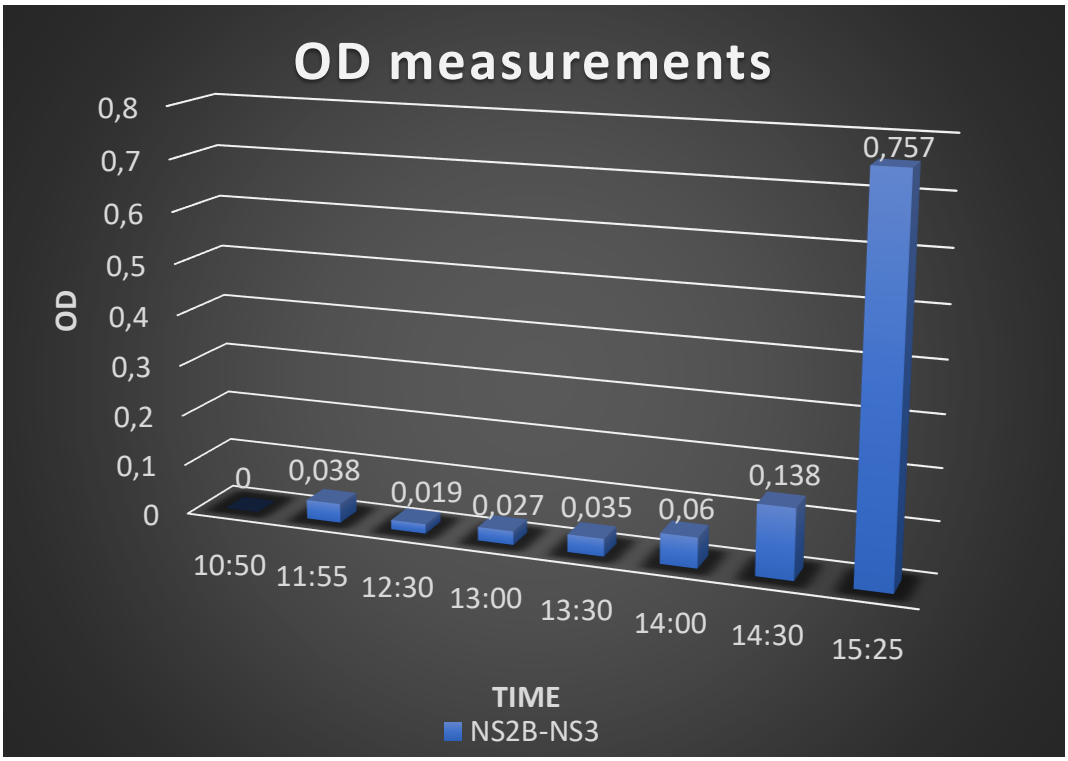

Figure 19. OD measurements for NS2B-NS3

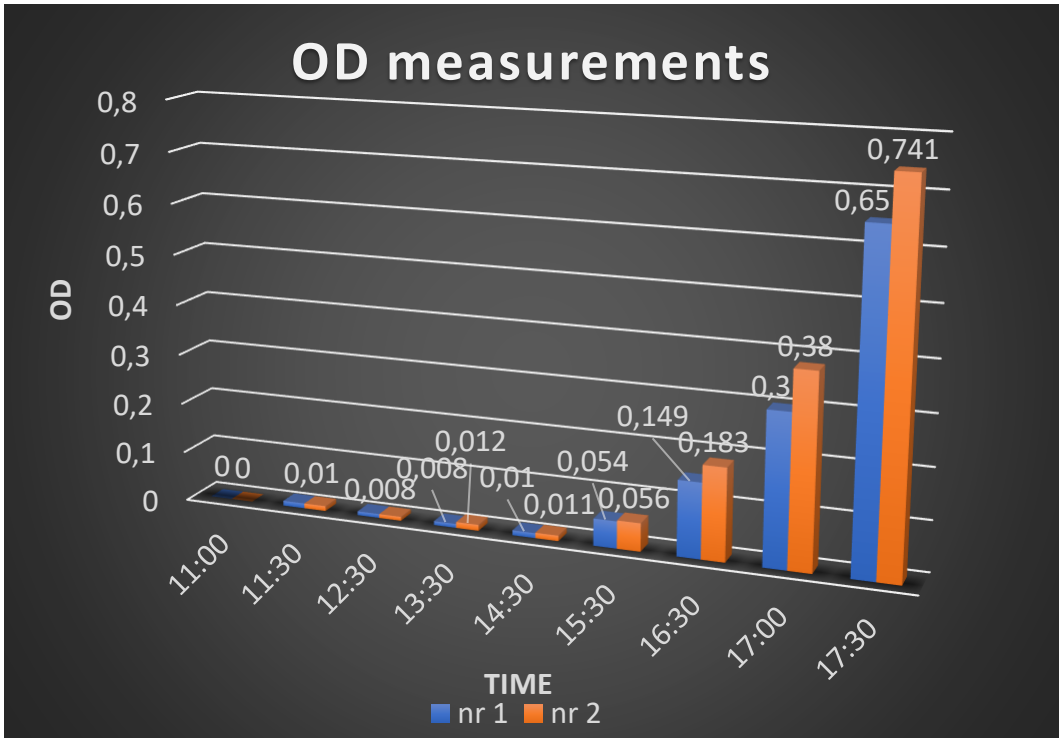

Figure 20. OD measurements for 6LZE_A

4.2.2 Protein purification

The purification process for NS2B-NS3 resulted in a very small single peak. The tubes 4 and 5 from Figure 21 were further collected and the concentration measured in the Nanodrop showed a value of -0.6 mg/ml so this sample was discarded.

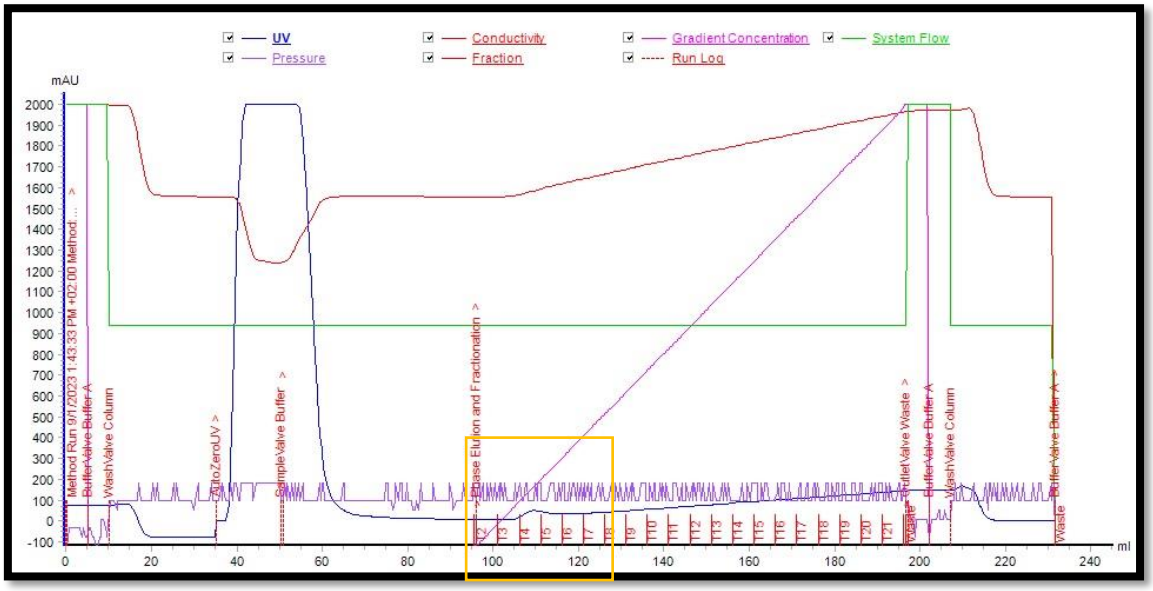

Figure 21. Purification chromatogram for NS2B-NS3

4.3 Evaluation of the biological activity of the compounds

4.3.1 Enzyme activity assessment on DL-BAPNA substrate

The activity of the proteins from the first part was assessed in Multiskan Go and the result is shown in Figure 23. The first slot corresponds to 6LZE_A, the second to NS2B-NS3, and the third is for control.

The microplate was also observed to see if any color change appeared. In Figure 22 it can be noticed that the samples were all transparent.

Figure 22. Microplate activity assessment for 6LZE_A, NS2B-NS3 and control (part 1)

	avg time [r	A01	A02	A03
1	0,0000	0,0881	0,0902	0,0866
2	0,1667	0,0877	0,0898	0,0865
3	0,3333	0,0871	0,0895	0,0863
4	0,5000	0,0878	0,0897	0,0864
5	0,6667	0,0878	0,0899	0,0869
6	0,8333	0,0882	0,0904	0,0870
7	1,0000	0,0882	0,0903	0,0867
8	1,1667	0,0884	0,0899	0,0868
9	1,3333	0,0886	0,0904	0,0865
10	1,5000	0,0883	0,0904	0,0849

Figure 23. Multiskan Go results for 6LZE_A, NS2B-NS3 and control (part 1)

For the second part, the activity assessments for NS2B-NS3 were done before and after incubation when the substrate concentration also increased. The results presented in Figures 24 and 25 show similar values for the samples (first slots) and controls (second slots).

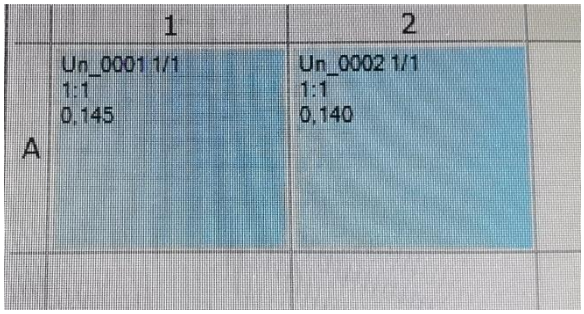

Figure 24. Multiskan Go readings for NS2B-NS3 and control (part 2)

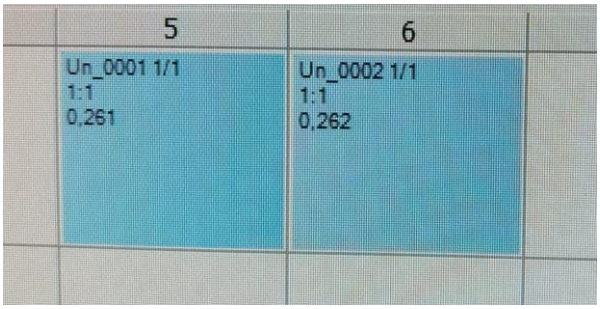

Figure 25. Multiskan Go readings for NS2B-NS3 and control after incubation (part 2)

4.4 Molecular modeling of NS2B-NS3

The sequence was copied in NCBI BLAST and 3 of the Accession numbers with the highest Percent Identity were further searched in PDB (Protein Data Bank).

Table 1. NCBI BLAST results for NS2B-NS3

Accession nr	Percent identity	Mutations
4M9F	100%	Yes
4M9K	99.60%	Yes
6MO0	99.19%	No

Since the first 2 proteins had mutations only 6MO0 was used further for structure analysis.

4.4.1 Structure analysis

The structure of the protein was analyzed in order to observe the secondary structures, the active site, and the catalytic triad. To analyze the structure of the protein the Accession number was loaded in the Chimera program and so the 3D structure was able to be visualized.

Figure 26. Secondary structures of 6MO0

In Figure 26 the secondary structures can be observed, thus there are 2 α-helix strands (red), and 15 ß-pleated sheets (purple) which are antiparallel. It can also be noticed that the structure is incomplete by having 2 regions with unknown atomic coordinates (dotted lines). In order to predict the missing region the sequence of the protein was copied in ChimeraX. The results can be seen in Figure 27.

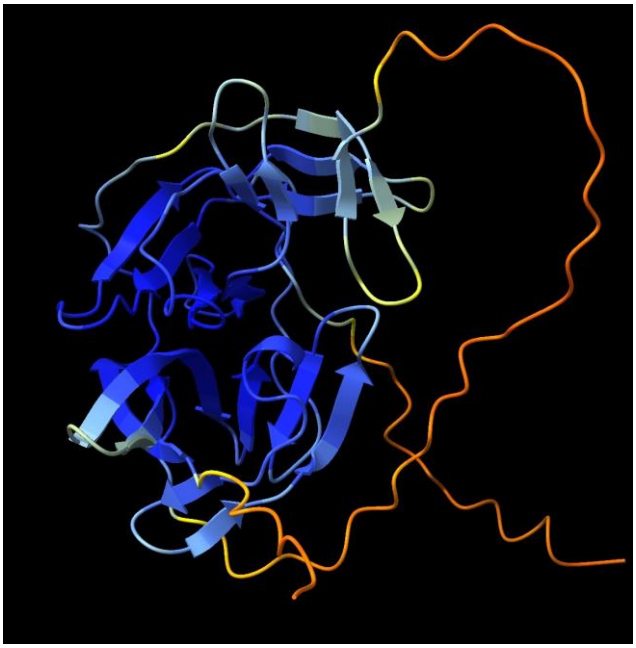

Figure 27. Structural prediction in ChimeraX

A Poly-Histidine ligand was built in Yasara for both proteins. The reason for this was that it was suspected that the Poly-histidine tag might have an inhibitory effect on the proteases. However, due to time limitations, the experimental part to test this hypothesis could not be concluded anymore. This would have been done by ordering the same genes but with the His-tag removed and following the same process as for the ones containing the His-tag except for the purification step that would require a His-tag to bind to the column so attaching one before the purification and removing it after would have been necessary. The Poly-Histidine ligand built can be seen in Figure 28.

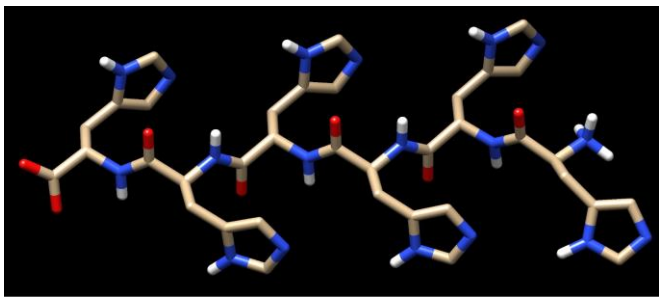

Figure 28. The ligand built in Yasara

The Chimera tool was also used to visualize the active site together with the Poly-histidine ligand and the result can be seen in Figure 29. There are more hydrophobic acids (orange) than hydrophilic amino acids (blue) in the active site.

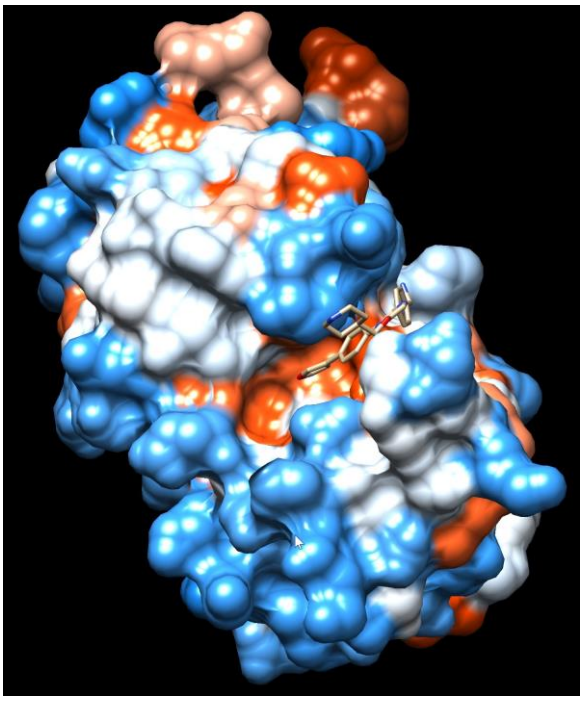

Figure 29. The active site of 6MO0

According to the literature the active site is comprised of a triad (H51, D75, and S135) and 4 subpockets: 1(D129, S135, Y150, Y161), 2(D75, D82, G83, N84, N152), 3(F85, Q86, L87), and 4(V154) [68]. Knowing this I searched for the ASP, HIS, and SER residues in Chimera and I identified the triad as HIS1051, ASP1075, and SER1135 (Figure 30).

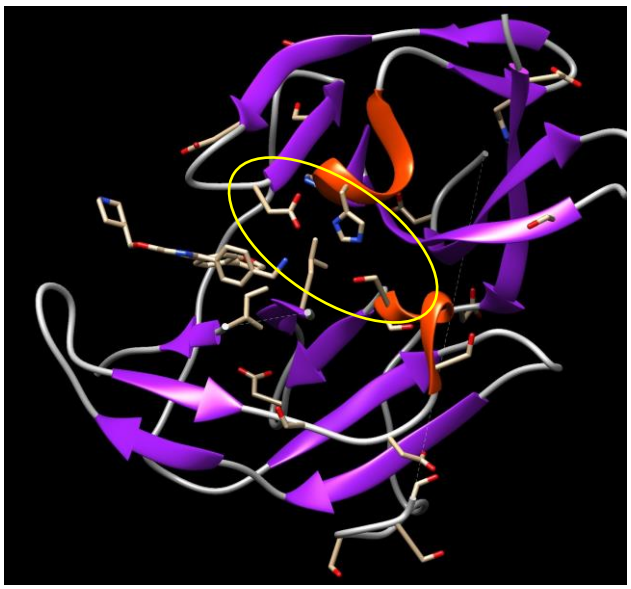

Figure 30. The triad of 6MO0

4.4.2 Docking with the Poly-Histidine ligand

The docking of the Poly-Histidine ligand to the active site can be seen in Figure 31. The docking was picked according to the highest score shown in Yasara. The distances between the triad and the ligand were measured in Chimera and can be seen in Figure 32. These distances can show if the ligand is bound in the right place or not and how strong this binding is. These distances are shown in Å and they are as follows:

Table 2. Distances in Å between the ligand and the triad

	ASP1075	SER1135	HIS1051
HIS1	6.594	14.640	11.491
HIS2	15.740	23.401	19.734
HIS3	6.001	12.386	9.747
HIS4	13.993	16.797	14.973
HIS5	17.252	21.571	20.161
HIS6	12.004	18.646	15.930

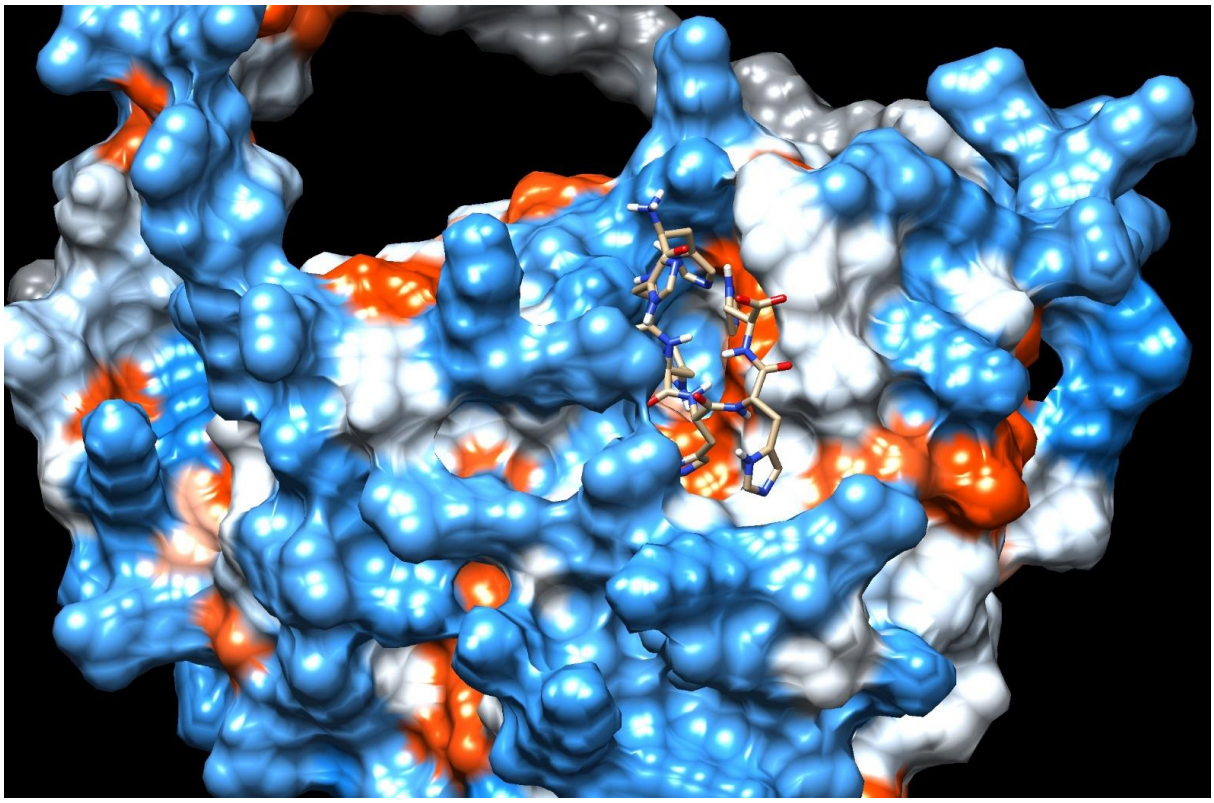

Figure 31. The docking of the ligand to the active site of NS2B-NS3

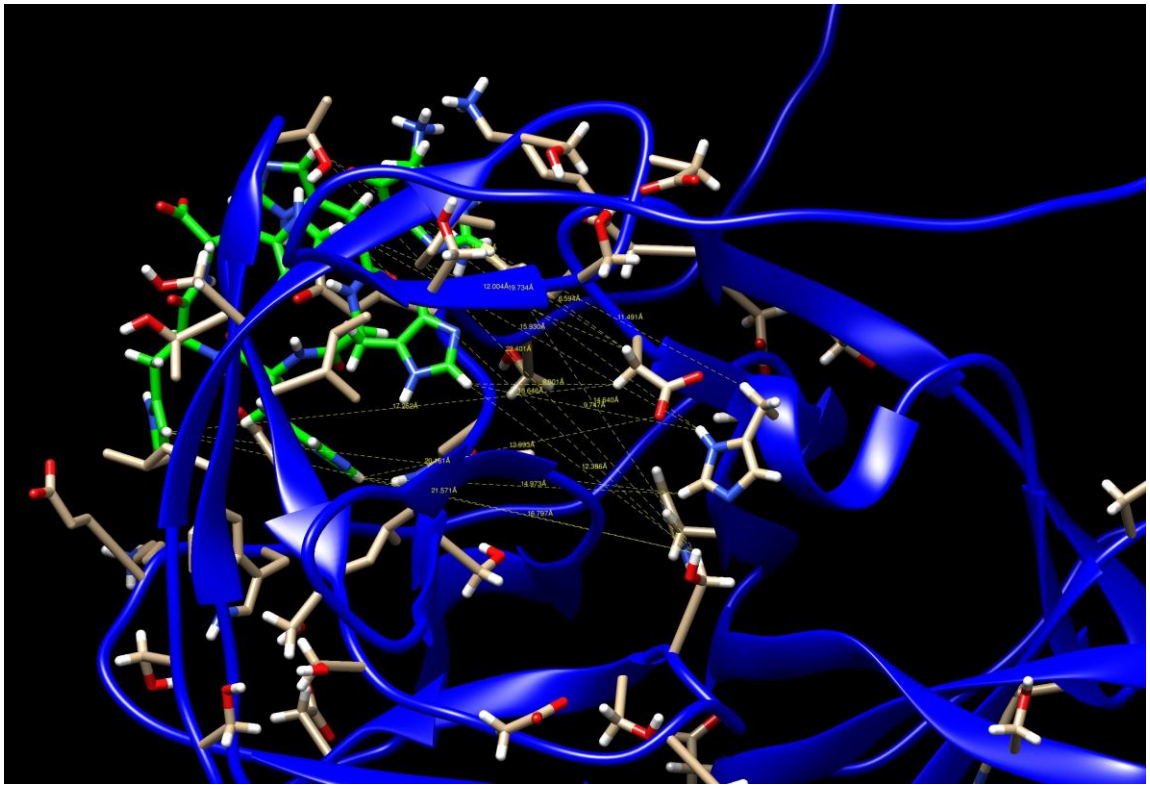

Figure 32. Distances between the triad and the ligand for NS2B-NS3

4.4.3 Docking with DL-BAPNA

The DL-BAPNA structure was downloaded from NCBI and used for docking to the active site of NS2B-NS3 which can be seen in Figure 34. The docking was picked according to the highest score shown in Yasara. The distances between the triad and DL-BAPNA were measured in Chimera and can be seen in Figure 35. These distances are shown in Å and they are as follows:

Table 3. Distances in Å between DL-BAPNA and the triad

	ASP1075	SER1135	HIS1051
DL-BAPNA	7.187	12.214	8.814

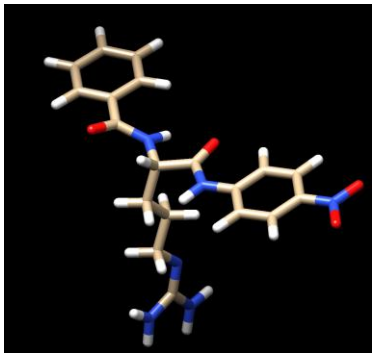

Figure 33. The DL-BAPNA structure

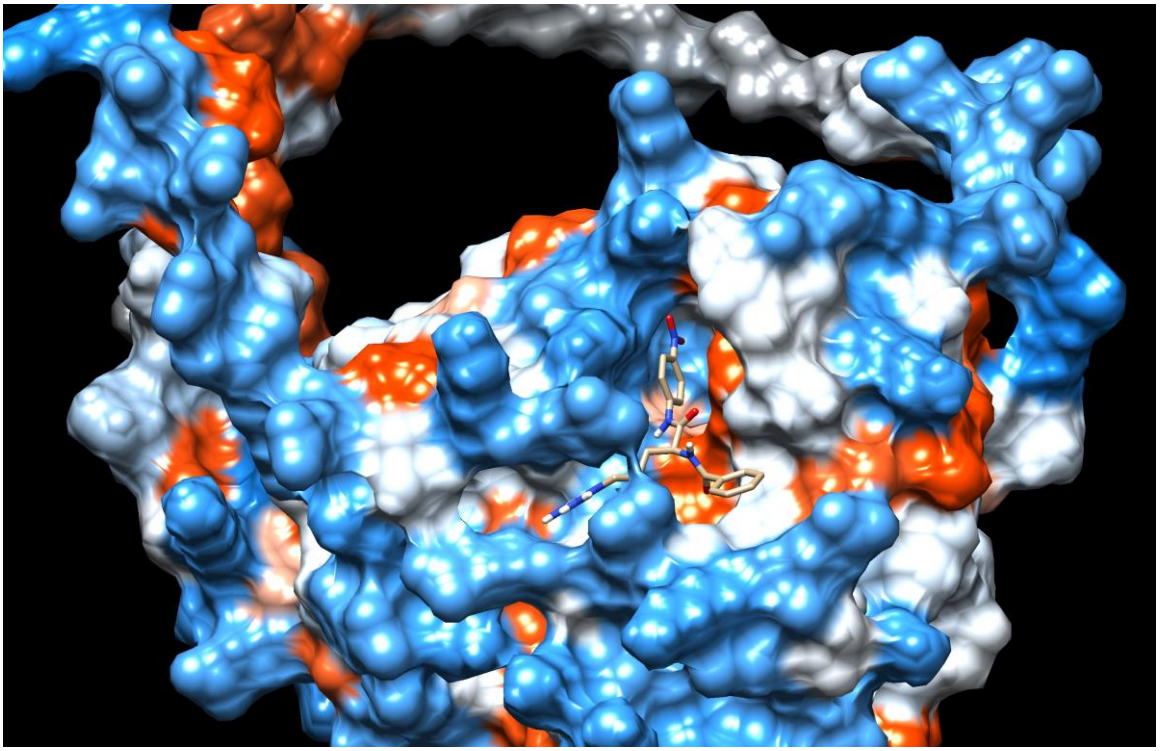

Figure 34. The docking of DL-BAPNA to the active site of NS2B-NS3

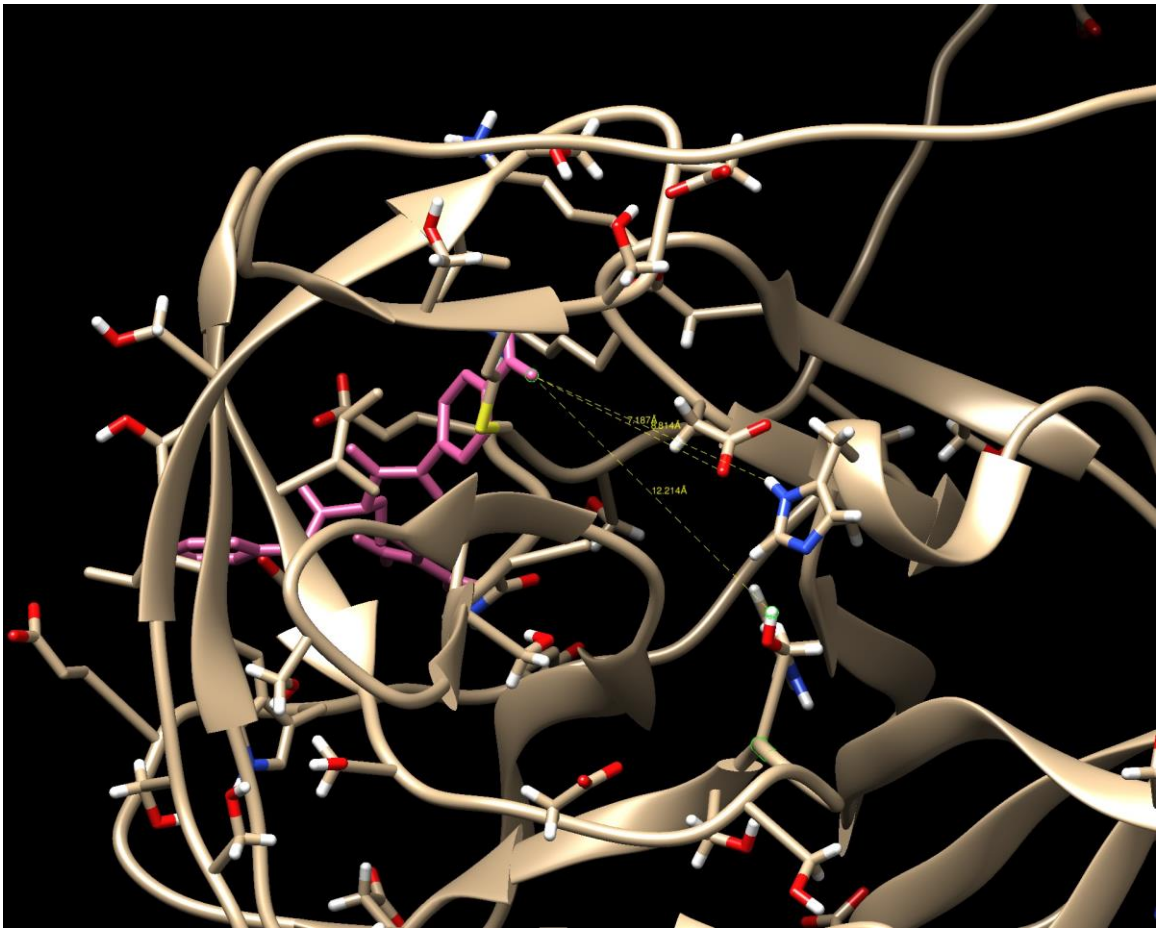

Figure 35. Distances between the triad and DL-BAPNA for NS2B-NS3

4.5 Molecular modeling of 6LZE_A

The sequence was copied in NCBI BLAST but the Accession numbers could not be found in PDB (Protein Data Bank) based on this search so instead a search was made for 6LZE and found the protein with the identical sequence as the one provided by GenScript.

4.5.1 Structure analysis

The structure of the protein was analyzed in order to observe the secondary structures, the active site, and the catalytic dyad. By using the Accession number 6LZE the 3D structure was observed and analyzed in the Chimera program. The secondary structures can be visualized in Figure 36. There are 9 α-helix strands (red), and 13 ß-pleated sheets (purple) which are antiparallel. 6LZE_A has 3 domains, 2 containing ß-pleated sheets, and 1 containing α-helix strands [94].

Figure 36. The secondary structures of 6LZE_A

The active site can be observed in Figure 37. It can be noticed that there are both hydrophilic and hydrophobic amino acids at the active site.

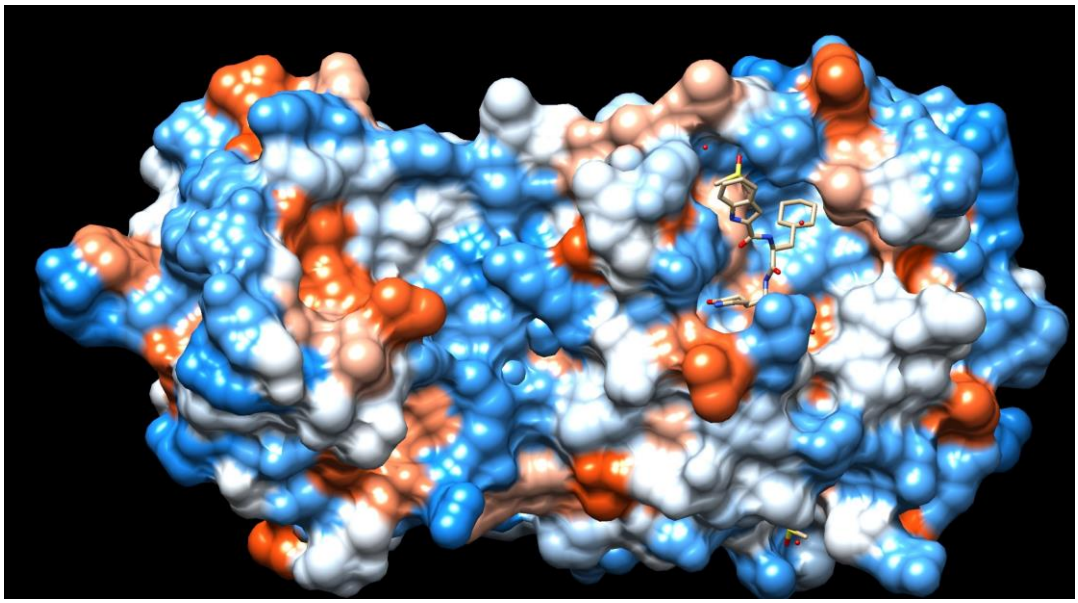

Figure 37. The active site of 6LZE_A

The active site of 6LZE_A consists of a Cys-His dyad according to the literature [82]. These have been identified as CYS145 and HIS41 and can be seen in Figure 38.

Figure 38. The dyad of 6LZE_A

4.5.2 Docking with the Poly-Histidine ligand

The same ligand was used for docking for both proteins. The docking to the active site of 6LZE_A can be seen in Figure 39. The docking was picked according to the highest score shown in Yasara. The distances between the dyad and the ligand were measured in Chimera and can be seen in Figure 40. These distances are shown in Å and they are as follows:

Table 4. Distances in Å between the ligand and the dyad

	HIS41	CYS145
HIS1	23.826	23.755
HIS2	28.456	27.312
HIS3	20.403	19.800
HIS4	22.535	20.754
HIS5	19.584	17.768
HIS6	16.798	16.260

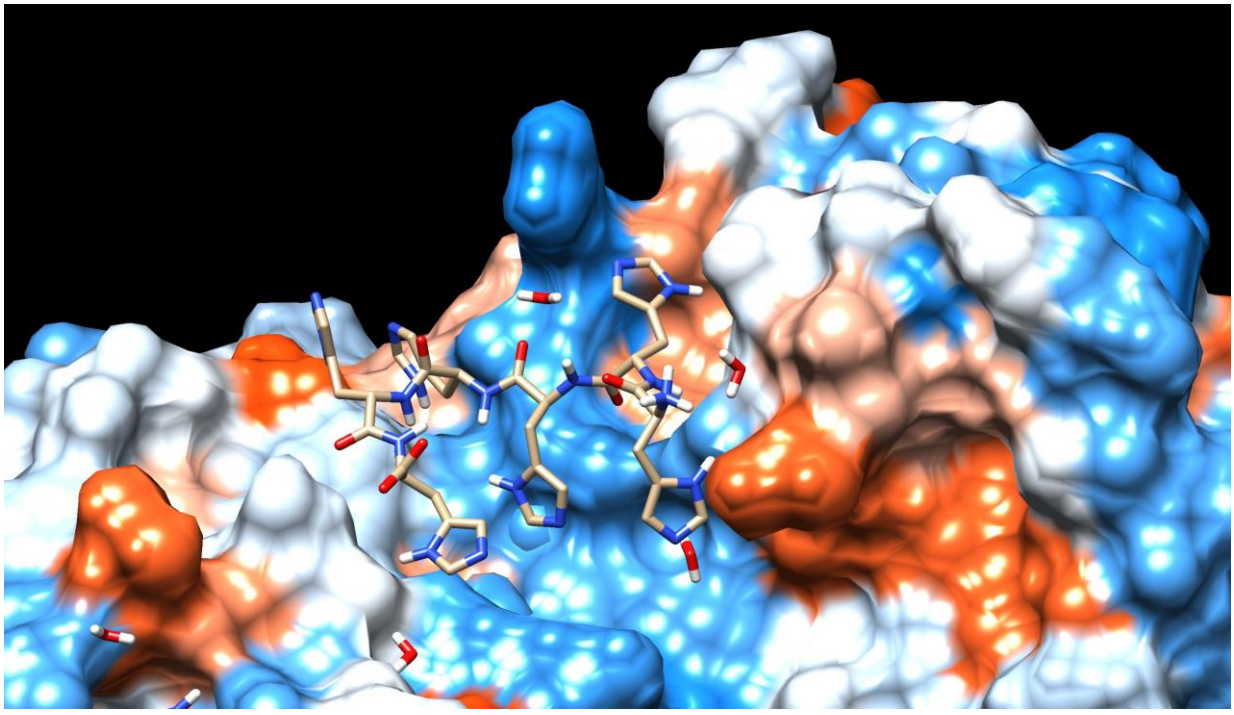

Figure 39. The docking of the ligand to the active site of 6LZE_A

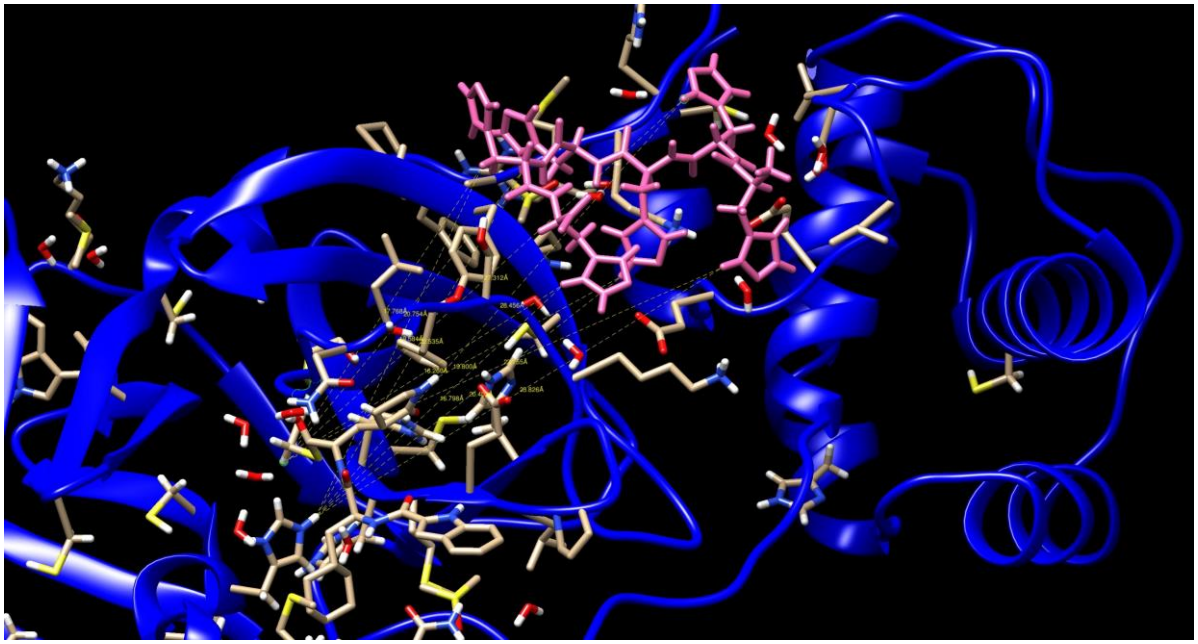

Figure 40. Distances between the dyad and the ligand for 6LZE_A

It was also noticed that there were many water molecules surrounding the ligand and at a very small distance from each other. The distances between the ligand's residues and water molecules were measured, and they can be seen in Table 5 in Å.

Table 5. Distances between the ligand and the closest water molecule

	HOH
HIS1	2.250
HIS1	3.718
HIS2	1.946
HIS2	2.615
HIS6	3.250

Since the distances between the ligand and the dyad were very high the docking was redone but first, the water molecules were removed in Chimera. The redocking can be observed in Figure 41, and the distances between the ligand and the dyad in Figure 42 and Table 6.

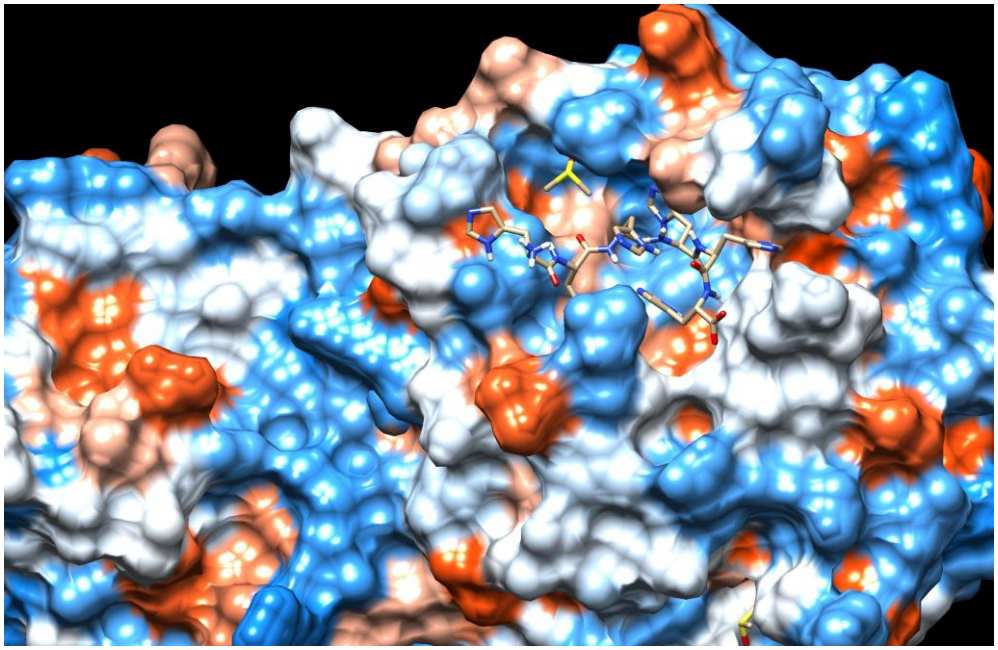

Figure 41. The redocking of the ligand to the active site of 6LZE_A

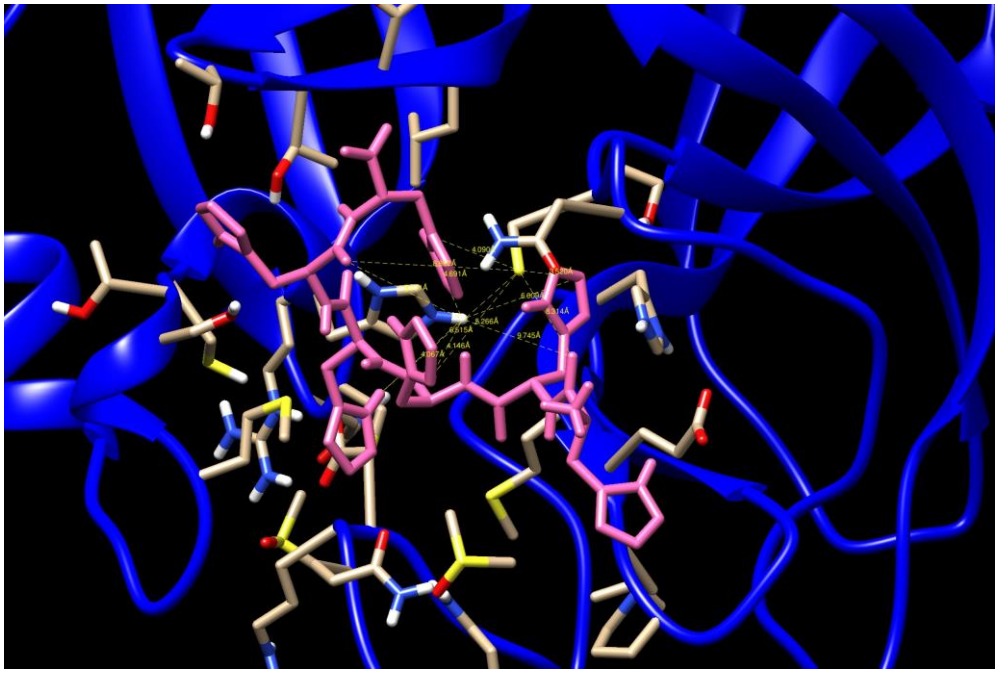

Figure 42. Distances between the dyad and the ligand for 6LZE_A after redocking

Table 6. Distances in Å between the ligand and the dyad after redocking

	HIS41	CYS145
HIS1	9.745	8.314
HIS2	6.009	3.520
HIS3	4.146	5.266
HIS4	4.067	6.515
HIS5	9.071	8.892
HIS6	4.691	4.090

4.5.3 Docking with DL-BAPNA

The 6LZE_A structure without the water molecules was used for docking between the active site of 6LZE_A and DL-BAPNA and can be seen in Figure 43. The docking was picked according to the highest score shown in Yasara. The distances between the dyad and the ligand were measured in Chimera and can be seen in Figure 44. These distances are shown in Å and they are as follows:

Table 7. Distances in Å between DL-BAPNA and the dyad

	HIS41	CYS145
DL-BAPNA	24.554	23.736

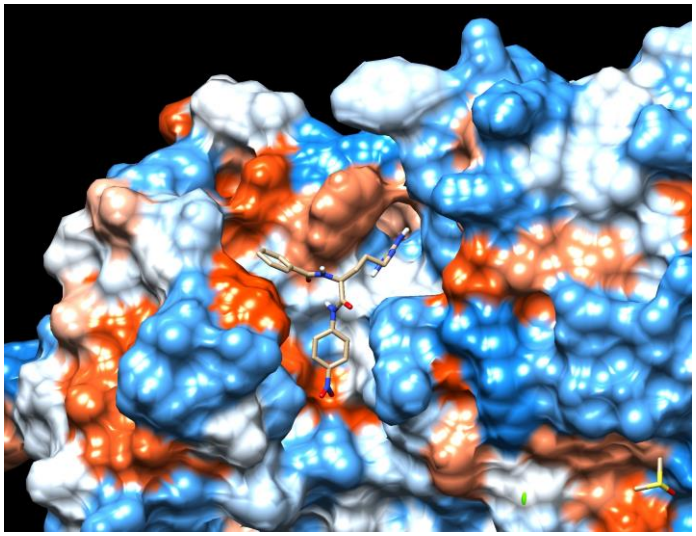

Figure 43. The docking of DL-BAPNA to the active site of 6LZE_A

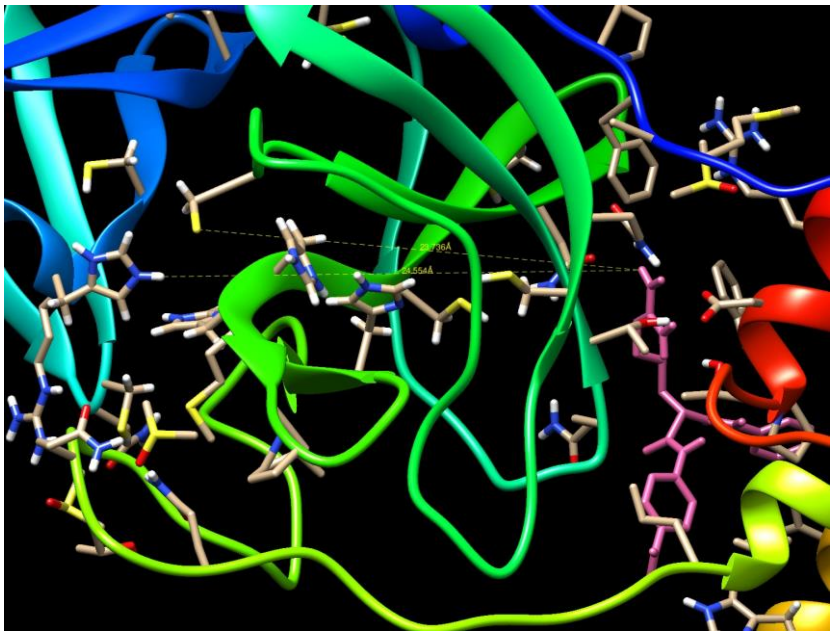

Figure 44. Distances between the dyad and DL-BAPNA for 6LZE_A

5 Discussion

5.1 Production of 6LZE_A and NS2B-NS3

5.1.1 Protein production

The production of both proteins seemed challenging at times with one purification failure (NS2B-NS3), 2 repetitions for 6LZE_A, and 3 for NS2B-NS3. Regarding the cell growth on Agar plates, it can be observed that when the parameters for the thermal shock transformation changed for 6LZE_A in the second part it resulted in a higher number of colonies than in the first part, so it is better if the tube containing the plasmid and the cells is incubated on ice for 30 minutes instead of 15, then heated at 42°C for 90 seconds rather than 120 and cooled on ice for 10 minutes.

The production times were similar in both parts for each protein. For 6LZE_A the process took approximately 6.5 hours and for NS2B-NS3 3.5 hours. The prolonged lag phase for 6LZE_A could be due to a survival mechanism that allows cells to cope with environmental stress [81]. It can also be noticed that at both times when the production for NS2B-NS3 in the second part died or diminished the rpm was set at 180. As soon as it increased to 200 the growth rate increased again and the culture was saved (Figure 19). Thus, it can be concluded that the oxygen level was too low for cells to multiply.

5.1.2 Protein purification

According to the ÄKTA purification chromatogram, the unspecific protein was in a much higher concentration than the target protein for 6LZE_A, while for NS2B-NS3 was the other way around in the first part.

The purification failure for NS2B-NS3 in the second part could be caused by the usage of Agar plates that might have been too old. Considering that they were a few months old it can be assumed that maybe the cells had low or no activity anymore hence the small single peak obtained in Figure 21. The reason for not producing new Agar

plates with NS2B-NS3 was a lack of competent cells and a time limit that did not allow me to invest more time in the matter.

5.1.3 SDS-PAGE electrophoresis

According to PDB, the molecular weight of 6LZE is 34.21 kDa [78]. In Figure 16 it can be noticed that the molecular weight of the 6LZE_A sample is indeed around 34 kDa and there is very little contamination. The faded line could be attributed to the low concentration of the purified protein.

The result for NS2B-NS3 showed multiple contaminations even though it was in a higher concentration than 6LZE_A since the line was stronger. According to the protein sequence, NS2B-NS3 should indeed have a molecular weight of approximately 34 kDa.

5.2 Evaluation of the biological activity of the compounds

Both assessments for testing the activity of the proteins showed that they had no activity. The readings had constant values instead of increasing values and the color of the sample did not change from transparent to yellow as expected when viewing the microplate. However, the SDS-PAGE results (Figure 16) showed that specific proteins for both proteases were present in the samples.

For 6LZE_A the negative result could be due to either usage of the wrong type of substrate or a very low concentration of the protein after purification. According to the literature, DL-BAPNA is a substrate that was previously used successfully for DENV but was not tested for SARS-CoV-2 [80].

It is also possible that the Poly-Histidine tag might have affected the activity of both proteins. In a study by Mira Behman, a His-tag attached to the C-terminal of the main protease (Mpro) of SARS-CoV-2 restrained the inhibitory activity of atazanavir towards the protease. The fact that both the N- and C-terminal are close to the active site of Mpro could be a reason for this inhibitory effect by the His-tag through binding competition with the inhibitor. It was also shown in this study that assay conditions such as pH and salinity play a major role in the

activity of the proteins and inhibitors [95]. As a response to Behman, Zhe Li further proved that the binding affinity when using the His-tag is 20 times lower than when using a tag-free protease due to the binding competition with the substrate or inhibitor [96].

5.3 Molecular modeling

5.3.1 Structure analysis

When comparing the structures of the two proteins it can be noticed that NS2B-NS3 has only 2 α-helix strands while for 6LZE_A the number of α-helix strands is almost as much as the number of ß-pleated sheets. Knowing that α-helices stabilization depends on local interactions while ß-pleated sheets stabilization depends on distance interactions it can be concluded that 6LZE_A has a higher environmental stability than NS2B-NS3 [84].

Some of the natural ligands for NS2B-NS3 studied previously are protegrin-1, conotoxins, and retrocyclin-1 [86].

For 6LZE_A it can be noticed in Figure 37 that the active site is mostly hydrophilic. Usually, the active site of proteins is mostly hydrophobic as for the NS2B-NS3 case. The catalytic dyad is buried right in the active site of 6LZE_A while the catalytic triad of NS2B-NS3 is further away from the active site.

According to research done by Sayed, the natural ligands for 6LZE_A with the highest potential are citriquinochroman, holyrine B, proximicin C, pityriacitrin B, anthrabenzoxocinone, and penimethavone A. From these, citriquinochroman found in *Penicillium citrium* showed perfect docking to the active site via Hydrogen bonds [82].

5.3.2 Docking

According to Table 2, the smallest distance between the Poly-Histidine ligand and the catalytic triad of NS2B-NS3 is between ASP1075 and HIS3 (6.001Å), and the longest is between SER1135 and HIS2 (23.401Å). Since the distance between ASP1075 and HIS3 is less than

8Å this means that these amino acids come in contact with each other [83].

The amino acids in the vicinity of the ligand have been identified as GLU (150, 153, 231), ASN229, LEU177, ILE36, LYS (135, 136, 152), and THR (182, 184). This means that glutamic acid, lysine, and threonine are predominant in the vicinity of the ligand.

Glutamic acid, lysine, and threonine are all important for the stability between hydrogen bonds [87,88,89].

The shortest distance between the DL-BAPNA substrate and the catalytic triad of NS2B-NS3 is 7.187Å and since it is less than 8Å it means that ASP1075 comes in contact with DL-BAPNA.

When it comes to 6LZE_A the shortest distance between the Poly-Histidine ligand and the catalytic dyad is between CYS145 and HIS6 (16.260Å), and the longest is between HIS41 and HIS2 (28.456Å). Compared to NS2B-NS3 the distances are bigger between the dyad and the ligand and they do not seem to come in contact with each other since the measurements are bigger than 8Å.

The amino acids in the vicinity of the ligand are TYR126, SER (139,284), HOH (540, 569, 611, 648, 658, 661, 676), LYS (5, 137), GLU (288, 290), ARG4, GLN127, and MET6. This means that the ligand is surrounded by many water molecules, but also serine, lysine, and glutamic acid.

Water molecules are very important in the binding process between the protein and the ligand by mediating the hydrogen bonds and they also play a role in increasing ligand affinity [85]. Since most of the distances shown in Table 5 are less than 3.60Å it means that most of the water molecules are bound to the ligand which correlated to the hydrophilic active site [91].

The role of serine is mainly for phosphorylation and catalytic activity [90].

When the docking between the Poly-Histidine ligand and the catalytic dyad of 6LZE_A was redone after removing the water molecules it can be noticed that the distances are substantially smaller. The shortest distance between the Poly-Histidine ligand and the catalytic dyad is

between CYS145 and HIS2 (3.520Å), and the longest one is between HIS41 and HIS1 (9.745Å). Since most of the distances are less than 8Å it seems that the molecules are coming in contact with each other. The differences between the two dockings, with and without water molecules show that the water molecules might interfere with the docking process.

The distances between the catalytic dyad of 6LZE_A and DL-BAPNA are very large and they show that the molecules do not come in contact with each other.

6 Conclusion

As shown in the report, there are some differences between the SARS-CoV-2 virus and DENV in regards to structure and catalytic mechanism.

The prolonged lag phase of 6LZE_A compared to NS2B-NS3 could be due to a survival mechanism that allows cells to cope with environmental stress. NS2B-NS3 has only 2 α-helix strands while for 6LZE_A the number of α-helix strands is almost as much as the number of ß-pleated sheets. Knowing that α-helices stabilization depends on local interactions while ß-pleated sheets stabilization depends on distance interactions it can be concluded that 6LZE_A has a higher environmental stability than NS2B-NS3. The presence of serine residues and water molecules could be a reason for the better stability of 6LZE_A.

The assessments for testing the activity of the proteins showed that they had no activity. One of the reasons could have been due to the usage of the wrong substrate or too low a concentration of the proteins. It is also possible that the Poly-Histidine might have interfered with the activity of the proteins as shown in other studies.

When checking the docking of the Poly-histidine ligand with 6LZE_A the one with the highest score shown in Yasara was picked. However, according to the distances between the ligand and the catalytic residues, it can be concluded that the docking was not in the right position for 6LZE_A when done the first time. Once the docking was redone after removing the water molecules the distances became substantially smaller meaning that the water molecules might have interfered with the docking on the first try.

Both dockings with Poly-Histidine and DL-BAPNA with the catalytic triad of NS2B-NS3 showed small distances that suggest that the molecules come in contact with each other. When comparing the two proteins, it seems that the dockings for NS2B-NS3 were a success while the results for 6LZE_A would need further research to understand what could be the problem for these very large distances.

Due to the difficulties that arose in producing these proteins because of several failures and repetitions either because very few colonies grew on the Agar plates, cells died during the cultivation process, low concentrations, or failed purification more time and research would be necessary for better understanding their mechanisms and finding an effective way to inhibit the viral replication process thus prohibiting the virus from spreading.

7 References

1. Malone, B., Urakova, N., Snijder, E.J. *et al.* (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol* **23**, 21–39. https://doi.org/10.1038/s41580-021-00432-z
2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536–544. https://doi.org/10.1038/s41564-020-0695-z
3. Pustake, M., Tambolkar, I., Giri, P., & Gandhi, C. (2022). SARS, MERS and CoVID-19: An overview and comparison of clinical, laboratory and radiological features. *Journal of family medicine and primary care*, *11*(1), 10–17. https://doi.org/10.4103/jfmpc.jfmpc_839_21
4. Worldometers. (2023). [online]. Available at: https://www.worldometers.info/coronavirus/
5. Spyros, L., Wei, X., Joseph, H., Xiaowei, J., David, L. (2021). The animal origin of SARS-CoV-2. *Science*, 373 (6558), pp. 968-970. DOI: 10.1126/science.abh0117
6. Centers for Disease, Control, and Prevention. (2022). [online]. Available at: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/animals.html
7. Pustake, M., Tambolkar, I., Giri, P., & Gandhi, C. (2022). SARS, MERS and CoVID-19: An overview and comparison of clinical, laboratory and radiological features. *Journal of family medicine and primary care*, *11*(1), 10–17. https://doi.org/10.4103/jfmpc.jfmpc_839_21
8. World Health Organization. (2021). [online]. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-how-is-it-transmitted
9. Kumar, A., Narayan, R. K., Prasoon, P., Kumari, C., Kaur, G., Kumar, S., Kulandhasamy, M., Sesham, K., Pareek, V., Faiq, M. A., Pandey, S. N., Singh, H. N., Kant, K., Shekhawat, P. S., Raza, K., & Kumar, S. (2021). COVID-19 Mechanisms in the Human Body-What We Know So Far. *Frontiers in immunology*, *12*, 693938. https://doi.org/10.3389/fimmu.2021.693938
10. V'kovski, P., Kratzel, A., Steiner, S. *et al.* (2021). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat Rev Microbiol* **19**, 155–170. https://doi.org/10.1038/s41579-020-00468-6
11. Bond J. S. (2019). Proteases: History, discovery, and roles in health and disease. *The Journal of biological chemistry*, *294*(5), 1643–1651. https://doi.org/10.1074/jbc.TM118.004156
12. Sino Biological. [online]. Available at: https://www.sinobiological.com/research/enzymes/what-are-proteases

13. López-Otín, C., & Bond, J. S. (2008). Proteases: multifunctional enzymes in life and disease. *The Journal of biological chemistry*, *283*(45), 30433–30437. https://doi.org/10.1074/jbc.R800035200
14. Jermen, M., Fassil, A.. (2018). The role of microbial aspartic protease enzyme in food and beverage industry. *Journal of Food quality*. https://doi.org/10.1155/2018/7957269
15. Sino Biological. [online]. Available at: https://www.sinobiological.com/research/enzymes/serine-protease-regulator
16. Neitzel, J. J. (2010) Enzyme Catalysis: The Serine Proteases . *Nature Education* 3(9):21.
17. Who Health Organization. (2023). [Online]. Weekly epidemiological update on COVID-19. Available at: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---25-january-2023
18. Zephyr, J., Kurt Yilmaz, N., & Schiffer, C. A. (2021). Viral proteases: Structure, mechanism and inhibition. *The Enzymes*, *50*, 301–333. https://doi.org/10.1016/bs.enz.2021.09.004
19. Yasara. [online]. Available at: https://bio.tools/yasara
20. Noha, A. Saleh, Wael, M. Elshemey. (2017). Structure-based drug design of novel peptidomimetic cellulose derivatives as HCV-NS3 protease inhibitors. *Life Sciences*, Volume 187, Pages 58-63, ISSN 0024-3205. https://doi.org/10.1016/j.lfs.2017.08.021.
21. Sundus H, Mukhtar H, Nawaz A. Industrial applications and production sources of serine alkaline proteases: a review. J Bacteriol Mycol Open Acces. 2016;3(1):191-194. DOI: 10.15406/jbmoa.2016.03.00051
22. Rawat, A., Roy, M., Jyoti, A., Kaushik, S., Verma, K., & Srivastava, V. K. (2021). Cysteine proteases: Battling pathogenic parasitic protozoans with omnipresent enzymes. *Microbiological Research*, *249*, 126784. https://doi.org/10.1016/j.micres.2021.126784
23. Domsalla, A., Melzig, M.F. (2008). Occurrence and properties of proteases in plant latices. *Planta Medica*. **74** (7): 699–711. doi:10.1055/s-2008-1074530
24. David Troncoso, F., Alberto Sánchez, D., & Luján Ferreira, M. (2022). Production of Plant Proteases and New Biotechnological Applications: An Updated Review. *ChemistryOpen*, *11*(3), e202200017. https://doi.org/10.1002/open.202200017
25. Grudkowska, M., Zagdańska, B. (2004). Multifunctional role of plant cysteine proteinases. *Acta Biochimica Polonica*. **51** (3): 609–24. doi:10.18388/abp.2004_3547
26. Chapman, H.A., Riese, R.J., Shi, G.P. (1997). Emerging roles for cysteine proteases in human biology. *Annual Review of Physiology*. **59**: 63–88. doi:10.1146/annurev.physiol.59.1.63
27. Sharma, A., & Gupta, S. P. (2017). Fundamentals of Viruses and Their Proteases. *Viral Proteases and Their Inhibitors*, 1–24. https://doi.org/10.1016/B978-0-12-809712-0.00001-0
28. Nguyen, T. T., Myrold, D. D., & Mueller, R. S. (2019). Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny

and Habitat. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.00413
29. Bampidis, V., Azimonti, G., Bastos, M. de L., Christensen, H., et al. (2019). Safety and efficacy of l-threonine produced by fermentation with *Corynebacterium glutamicum* for all animal species. *EFSA Journal, 17(3)*. https://doi.org/10.2903/j.efsa.2019.5603
30. Wu, H., Bock, S., Snitko, M., Berger, T., Weidner, T., Holloway, S., Kanitz, M., Diederich, W. E., Steuber, H., Walter, C., Hofmann, D., Weißbrich, B., Spannaus, R., Acosta, E. G., Bartenschlager, R., Engels, B., Schirmeister, T., & Bodem, J. (2015). Novel dengue virus NS2B/NS3 protease inhibitors. *Antimicrobial agents and chemotherapy*, *59*(2), 1100–1109. https://doi.org/10.1128/AAC.03543-14
31. Andrew, H., Nigel, S., Geoffrey, D., Stephen, G. (2004). Glutamic protease distribution is limited to filamentous fungi. *FEMS Microbiology Letters*, Volume 239, Pages 95–101. https://doi.org/10.1016/j.femsle.2004.08.023
32. Jensen, K., Østergaard, P. R., Wilting, R., & Lassen, S. F. (2010). Identification and characterization of a bacterial glutamic peptidase. *BMC biochemistry*, *11*, 47. https://doi.org/10.1186/1471-2091-11-47
33. Mann, K. S., Chisholm, J., & Sanfaçon, H. (2019). *Strawberry Mottle Virus* (Family *Secoviridae*, Order *Picornavirales*) Encodes a Novel Glutamic Protease To Process the RNA2 Polyprotein at Two Cleavage Sites. *Journal of virology*, *93*(5), e01679-18. https://doi.org/10.1128/JVI.01679-18
34. Fujinaga, M., Cherney, M. M., Oyama, H., Oda, K., & James, M. N. (2004). The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3364–3369. https://doi.org/10.1073/pnas.0400246101
35. Raut, R., Beesetti, H., Tyagi, P. *et al.* (2015). A small molecule inhibitor of dengue virus type 2 protease inhibits the replication of all four dengue virus serotypes in cell culture. *Virol J* **12**, 16. https://doi.org/10.1186/s12985-015-0248-x
36. Murtuja, S., Shilkar, D., Sarkar, B., Sinha, B. N., & Jayaprakash, V. (2021). A short survey of dengue protease inhibitor development in the past 6 years (2015–2020) with an emphasis on similarities between DENV and SARS-CoV-2 proteases. *Bioorganic & Medicinal Chemistry*, *49*, 116415. https://doi.org/10.1016/j.bmc.2021.116415
37. Rawlings, N. D., & Barrett, A. J. (1995). Evolutionary families of metallopeptidases. *Methods in enzymology*, *248*, 183–228. https://doi.org/10.1016/0076-6879(95)48015-3
38. Laronha, H., & Caldeira, J. (2020). Structure and Function of Human Matrix Metalloproteinases. *Cells*, *9*(5), 1076. https://doi.org/10.3390/cells9051076
39. Cabral-Pacheco, G. A., Garza-Veloz, I., Castruita-De la Rosa, C., Ramirez-Acuña, J. M., Perez-Romero, B. A., Guerrero-Rodriguez, J. F., Martinez-Avila, N., & Martinez-Fierro, M. L. (2020). The Roles of Matrix Metalloproteinases and Their Inhibitors in Human Diseases. *International*

journal of molecular sciences, *21*(24), 9739. https://doi.org/10.3390/ijms21249739
40. Roy, R., Yang, J., & Moses, M. A. (2009). Matrix metalloproteinases as novel biomarkers and potential therapeutic targets in human cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, *27*(31), 5287–5297. https://doi.org/10.1200/JCO.2009.23.5556
41. Bramono, D., Richmond, J., Weitzel, P., Kaplan, D., Altman, G. (2004). Matrix Metalloproteinases and Their Clinical Applications in Orthopaedics. *Clinical Orthopaedics and Related Research 428():p 272-285.* http://doi.org/10.1097/01.blo.0000144166.66737.3a
42. Yamamoto, K., Okano, H., Miyagawa, W., Visse, R., Shitomi, Y., Santamaria, S., Dudhia, J., Troeberg, L., Strickland, D. K., Hirohata, S., & Nagase, H. (2016). MMP-13 is constitutively produced in human chondrocytes and co-endocytosed with ADAMTS-5 and TIMP-3 by the endocytic receptor LRP1. *Matrix biology : journal of the International Society for Matrix Biology*, *56*, 57–73. https://doi.org/10.1016/j.matbio.2016.03.007
43. Flores-Gallegos, A. C., Delgado-García, M., Ascacio-Valdés, J. A., Villareal-Morales, S., Michel-Michel, M. R., Aguilar-González, C. N., & Rodríguez-Herrera, R. (2019). *Hydrolases of Halophilic Origin With Importance for the Food Industry. Enzymes in Food Biotechnology, 197–219.* https://doi.org/10.1016/B978-0-12-813280-7.00013-X
44. Marathe, K. R., Patil, R. H., Vishwakarma, K. S., Chaudhari, A. B., & Maheshwari, V. L. (2019). *Protease Inhibitors and Their Applications: An Overview. Studies in Natural Products Chemistry, 211–242.* doi:10.1016/b978-0-444-64185-4.00006-x
45. Sarah, H., Indrasena, R., Shashi, K., Matthew, B., Alex, E., Shaobo, W., William, B., Davey, S., Aaron, F., Mark, E., and Tariq, M. (2022). *Journal of Medicinal Chemistry*, *65* (4), 2866-2879. DOI: 10.1021/acs.jmedchem.1c00566
46. Narayanan, A., Narwal, M., Majowicz, S.A. et al. (2022). Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay. *Commun Biol* **5**, 169. https://doi.org/10.1038/s42003-022-03090-9
47. Enzolifesciences. [online]. Available at: https://www.enzolifesciences.com/science-center/technotes/2020/january/why-do-we-need-recombinant-proteins?/
48. Diabetes. [online]. Available at: https://diabetes.org/blog/history-wonderful-thing-we-call-insulin
49. Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00172
50. Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein science : a publication of the Protein Society*, *28*(8), 1412–1422. https://doi.org/10.1002/pro.3668
51. Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process

Development. *Frontiers in Bioengineering and Biotechnology*, *7*. https://doi.org/10.3389/fbioe.2019.00420
52. Evitria. [online]. Available at: https://www.evitria.com/journal/recombinant-antibodies/application-recombinant-proteins/
53. Genextgenomics. [online]. Available at: https://genextgenomics.com/introduction-to-recombinant-proteins-its-applications/
54. Spencer, A., Osorio, F. A., & Hiscox, J. A. (2007). Recombinant viral proteins for use in diagnostic ELISAs to detect virus infection. *Vaccine*, *25*(30), 5653-5659. https://doi.org/10.1016/j.vaccine.2007.02.053
55. de Pinho Favaro, M. T., Atienza-Garriga, J., Martínez-Torró, C., Parladé, E., Vázquez, E., Corchero, J. L., Ferrer-Miralles, N., & Villaverde, A. (2022). Recombinant vaccines in 2022: a perspective from the cell factory. *Microbial cell factories*, *21*(1), 203. https://doi.org/10.1186/s12934-022-01929-8
56. HHS. [online]. Available at: https://www.hhs.gov/immunization/basics/types/index.html
57. Pham, P. V. (2018). *Medical Biotechnology. Omics Technologies and Bio-Engineering, 449–469.* doi:10.1016/b978-0-12-804659-3.00019-1
58. Gu. [online]. Available at: https://www.gu.se/pps
59. Gupta, V., Sengupta, M., Prakash, J., & Tripathy, B. C. (2016). Production of Recombinant Pharmaceutical Proteins. *Basic and Applied Aspects of Biotechnology*, 77-101. https://doi.org/10.1007/978-981-10-0875-7_4
60. Webmd. [online]. Available at: https://www.webmd.com/covid/coronavirus-history
61. Chatterjee, P., Nagi, N., Agarwal, A., Das, B., Banerjee, S., Sarkar, S., Gupta, N., & Gangakhedkar, R. R. (2020). The 2019 novel coronavirus disease (COVID-19) pandemic: A review of the current evidence. *The Indian Journal of Medical Research*, *151*(2-3), 147-159. https://doi.org/10.4103/ijmr.IJMR_519_20
62. Who. [online]. Available at: https://www.who.int/europe/emergencies/situations/covid-19
63. Nilapwar, S. M., Nardelli, M., Westerhoff, H. V., & Verma, M. (2011). Absorption Spectroscopy. *Methods in Enzymology*, *500*, 59-75. https://doi.org/10.1016/B978-0-12-385118-5.00004-9
64. Gomes, L., Monteiro, G., & Mergulhão, F. (2020). The Impact of IPTG Induction on Plasmid Stability and Heterologous Protein Expression by *Escherichia coli* Biofilms. *International journal of molecular sciences*, *21*(2), 576. https://doi.org/10.3390/ijms21020576
65. Rahimzadeh, M., Sadeghizadeh, M., Najafi, F., Arab, S., & Mobasheri, H. (2016). Impact of heat shock step on bacterial transformation efficiency. *Molecular Biology Research Communications*, *5*(4), 257-261. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5326489/
66. Cytivalifesciences. [online]. Available at: https://www.cytivalifesciences.com/en/us/shop/chromatography/chromatography-systems/akta-pure-p-05844

67. Phosphosolutions. [online]. Available at: https://www.phosphosolutions.com/pages/SDS-PAGE-demystified
68. Abduraman, M. A., Hariono, M., Yusof, R., Rahman, N. A., Wahab, H. A., & Tan, M. L. (2018). Development of a NS2B/NS3 protease inhibition assay using AlphaScreen® beads for screening of anti-dengue activities. *Heliyon*, *4*(12). https://doi.org/10.1016/j.heliyon.2018.e01023
69. Bumc. [online]. Available at: https://www.bumc.bu.edu/camed/research/cores/analytical-instrumentation-core/nanodrop-uv-visible-spectrophotometers/
70. Fishersci. [online]. Available at: https://www.fishersci.pt/shop/products/multiskan-go-microplate-spectrophotometer/p-4530546
71. Nature. [online]. Available at: https://www.nature.com/scitable/topicpage/dengue-viruses-22400925/
72. Brathwaite Dick, O., San Martín, J. L., Montoya, R. H., del Diego, J., Zambrano, B., & Dayan, G. H. (2012). The history of dengue outbreaks in the Americas. *The American journal of tropical medicine and hygiene*, *87*(4), 584–593. https://doi.org/10.4269/ajtmh.2012.11-0770
73. Who. [online]. Available at: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue
74. Nih. [online]. Available at: https://www.nih.gov/news-events/nih-research-matters/how-dengue-virus-infects-cells
75. NCBI. [online]. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi
76. RCSB. [online]. Available at: https://www.rcsb.org/pages/about-us/index
77. Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for Structure Building and Analysis. *Protein science : a publication of the Protein Society*, e4792. Advance online publication. https://doi.org/10.1002/pro.4792
78. RCSB. [online]. Available at: https://www.rcsb.org/structure/6LZE
79. Coelho, L. (2013). Lectins and trypsin inhibitors from plants: biochemical characteristics and adverse effects on insect larvae. Nova Science publishers. ISBN: 978-1-62417-578-7
80. Brackney, D. E., Foy, B. D., & Olson, K. E. (2008). The effects of midgut serine proteases on dengue virus type 2 infectivity of Aedes aegypti. *The American journal of tropical medicine and hygiene*, *79*(2), 267–274.
81. Bertrand, R. L. (2019). Lag Phase Is a Dynamic, Organized, Adaptive, and Evolvable Period That Prepares Bacteria for Cell Division. *Journal of Bacteriology*, *201*(7). https://doi.org/10.1128/JB.00697-18
82. Razali, R., Asis, H., & Budiman, C. (2021). Structure-Function Characteristics of SARS-CoV-2 Proteases and Their Potential Inhibitors from Microbial Sources. *Microorganisms*, *9*(12), 2481. https://doi.org/10.3390/microorganisms9122481
83. Maljković, M. M. (2019). DistAA: Database of amino acid distances in proteins and web application for statistical review of distances. *Computational Biology and Chemistry*, *83*, 107130. https://doi.org/10.1016/j.compbiolchem.2019.107130
84. Kopeć, K., Pędziwiatr, M., Gront, D., Sztatelman, O., Sławski, J., Łazicka, M., Worch, R., Zawada, K., Makarova, K., Nyk, M., & Grzyb, J. (2019).

Comparison of α-Helix and ß-Sheet Structure Adaptation to a Quantum Dot Geometry: Toward the Identification of an Optimal Motif for a Protein Nanoparticle Cover. *ACS omega*, *4*(8), 13086–13099. https://doi.org/10.1021/acsomega.9b00505

85. Schiebel, J., Gaspari, R., Wulsdorf, T., Ngo, K., Sohn, C., Schrader, T. E., Cavalli, A., Ostermann, A., Heine, A., & Klebe, G. (2018). Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nature Communications*, *9*. https://doi.org/10.1038/s41467-018-05769-2
86. Saqallah, F. G., Abbas, M. A., & Wahab, H. A. (2022). Recent advances in natural products as potential inhibitors of dengue virus with a special emphasis on NS2b/NS3 protease. *Phytochemistry*, *202*, 113362. https://doi.org/10.1016/j.phytochem.2022.113362
87. Kagra, D., Prabhakar, P. S., Sharma, K. D., & Sharma, P. (2020). Structural Patterns and Stabilities of Hydrogen-Bonded Pairs Involving Ribonucleotide Bases and Arginine, Glutamic Acid, or Glutamine Residues of Proteins from Quantum Mechanical Calculations. *ACS omega*, *5*(7), 3612–3623. https://doi.org/10.1021/acsomega.9b04083
88. Betts, M.J. and Russell, R.B. (2003) Amino Acid Properties and Consequences of Subsitutions. In: Gray, I.C., Ed., Bioinformatics for Geneticists Barnes MR, Wiley. http://dx.doi.org/10.1002/0470867302.ch14
89. Huang, P., Nagy, P. I., Williams, F. E., & Peseckis, S. M. (1999). Roles of threonine 192 and asparagine 382 in agonist and antagonist interactions with M1 muscarinic receptors. *British Journal of Pharmacology*, *126*(3), 735-745. https://doi.org/10.1038/sj.bjp.0702301
90. Yan, W., Zheng, Y., Zeng, X., He, B., & Cheng, W. (2022). Structural biology of SARS-CoV-2: Open the door for novel therapies. *Signal Transduction and Targeted Therapy*, *7*. https://doi.org/10.1038/s41392-022-00884-5
91. Lu, Y., Wang, R., Yang, C. Y., & Wang, S. (2007). Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *Journal of chemical information and modeling*, *47*(2), 668–675. https://doi.org/10.1021/ci6003527
92. JoVE Science Education Database. (2023). Basic Methods in Cellular and Molecular Biology. Bacterial Transformation: The Heat Shock Method.
93. Pfizer. [online]. Available at: https://www.pfizer.com/news/articles/how_a_coronavirus_protease_inhibitor_works_to_fight_covid_19
94. Lokhande, K. B., Doiphode, S., Vyas, R., & Swamy, K. V. (2021). Molecular docking and simulation studies on SARS-CoV-2 M^{pro} reveals Mitoxantrone, Leucovorin, Birinapant, and Dynasore as potent drugs against COVID-19. *Journal of biomolecular structure & dynamics*, *39*(18), 7294–7305. https://doi.org/10.1080/07391102.2020.1805019
95. Behnam, M. A. M., & Klein, C. D. (2021). Inhibitor potency and assay conditions: A case study on SARS-CoV-2 main protease. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(36), e2106095118. https://doi.org/10.1073/pnas.2106095118
96. Li, Z., Liu, R., Zhan, C. G., Wang, X., & Luo, H. B. (2021). Reply to Behnam and Klein: Potential role of the His-tag in C-terminal His-tagged SARS-

CoV-2 main protease. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(36), e2108209118. https://doi.org/10.1073/pnas.2108209118

97. Timiri, A. K., Sinha, B. N., & Jayaprakash, V. (2016). Progress and prospects on DENV protease inhibitors. *European journal of medicinal chemistry*, *117*, 125–143. https://doi.org/10.1016/j.ejmech.2016.04.008
98. Zhang, L., Crawford, F., Yu, L., Michels, A., Nakayama, M., Davidson, H. W., Kappler, J. W., & Eisenbarth, G. S. (2014). Monoclonal antibody blocking the recognition of an insulin peptide-MHC complex modulates type 1 diabetes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(7), 2656–2661. https://doi.org/10.1073/pnas.1323436111
99. Zephyr, J., Kurt Yilmaz, N., & Schiffer, C. A. (2021). Viral proteases: Structure, mechanism and inhibition. *The Enzymes*, *50*, 301–333. https://doi.org/10.1016/bs.enz.2021.09.004
100. Norshidah, H., Leow, C. H., Ezleen, K. E., Wahab, H. A., Vignesh, R., Rasul, A., & Lai, N. S. (2023). Assessing the potential of NS2B/NS3 protease inhibitors biomarker in curbing dengue virus infections: *In silico* vs. *In vitro* approach. *Frontiers in cellular and infection microbiology*, *13*, 1061937. https://doi.org/10.3389/fcimb.2023.1061937
101. Danis, C., Despres, C., Bessa, L., Malki, I., Merzougui, H., Huvent, I., Qi, H., Lippens, G., Cantrelle, F., Schneider, R., Hanoulle, X., Smet-Nocca, C., Landrieu, I. (2016). Nuclear Magnetic Resonance Spectroscopy for the Identification of Multiple Phosphorylations of Intrinsically Disordered Proteins. Journal of Visualized Experiments. 10.3791/55001.

8 Solutions composition

8.1 LB broth

20 g of LB powder was dissolved in 1 L of distilled water and then autoclaved.

8.2 Agar plates

16 g of Agar powder and 20 g of LB powder were mixed in 1 L of distilled water and then autoclaved.

8.3 Binding buffer

0.62 g NaH_2PO_4 (20mM), 4.5 g Na_2HPO_4 (20mM), and 29.2 g NaCl (500mM) were mixed in 800 mL of distilled water. Then the pH was adjusted to 7.4 (1M NaOH or HCl) and more water up to 1 L was added.

8.4 Elution buffer

0.62 g NaH_2PO_4 (20mM), 4.5 g Na_2HPO_4 (20mM), 29.2 g NaCl (500mM), and 34.02 g Imidazole (500mM) were mixed in 800 mL of distilled water. Then the pH was adjusted to 7.4 (1M NaOH or HCl) and more water up to 1 L was added.

8.5 Phosphate buffer

2.9 g NaH_2PO_4, 7,7 g Na_2HPO_4, 8.77 g NaCl were mixed in 800 mL of distilled water. Then the pH was adjusted to 7.4 (1M NaOH or HCl) and more water up to 1 L was added.

8.6 HCl 1 M

5 mL of HCl 32% was mixed with 45 mL of Milli-Q Water.

9 List of Figures

10 List of Tables