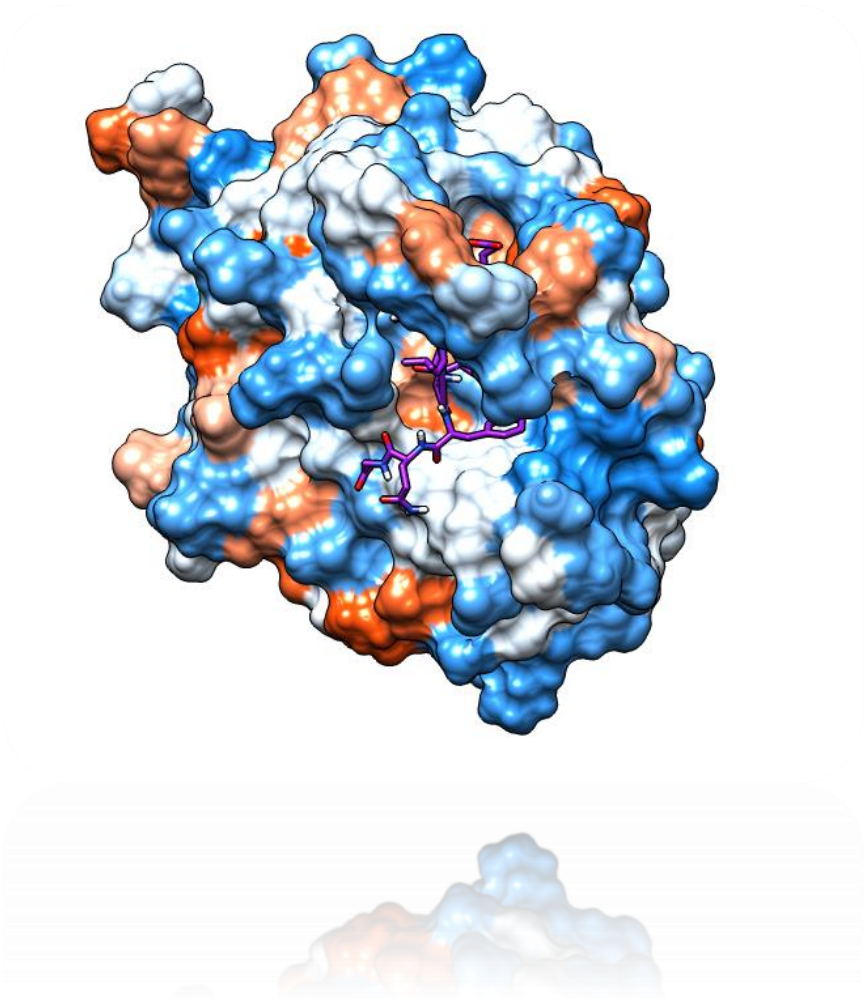
Production of recombinant TEV protease in *Escherichia coli* and analysis of enzymatic reactions



Master Thesis in Biotechnology by
Fisnik Nerjovaj

Supervisor: Javier Linares-Pastén

Examiner: Carl Grey

Lund University 2024

Faculty of Engineering

Division of Biotechnology

Popular science summary of:

"Production of recombinant TEV protease in *Escherichia coli* and analysis of enzymatic reactions"

Recent experiments, investigated in this thesis, may have shown that the TEV protease sequence recognition specificity is not as stringent as previously thought. This due to the possible observed hydrolysis of the substrate β-casein when digested with recombinant TEV protease. This new discovery may lead to new opportunities when using TEV protease as a tool in the lab as well as unexpected challenges.

The stringent sequence specificity of TEV protease has been studied by many and it is well-known by now that this recognition sequence consists of the amino acids ENLYFQ*G/S (where the asterisk marks the point of hydrolysation). The G, at the P1' position, can be substituted by several other amino acids, including S (Serine), without compromising the recognition pattern for the enzyme. The TEV protease has an auto-lysis tendency which is important to counter and can be done by fusing the fusion protein MBP (maltose-binding protein) to the protease, which prevents this detrimental function. The typical activity for the enzyme is to hydrolyse peptide bonds between Q (Glutamine) and G (Glycine), making the protease usable as a tool for precise removal of affinity tags. Recent experiments however have possibly shown that this specificity is not as strict as previously considered. The investigation in this thesis was conducted on the protein β-casein from bovine milk (UniProt accession no. P02666), which has an amino acid sequence similar to the recognition site for TEV protease between amino acids 131-142 (ENLHLPLPLLQS for β-casein compared to ENLYFQG for classic TEV protease recognition). The investigation started by docking the 12 amino acid long ligand to TEV protease which acts as the receptor in the UCSF Chimera program, which is a software allowing in-depth visualisation of molecular structures. The results showed a possible interaction between the thiol part of the cysteine (Cys151) in the active site of TEV protease to the carbonyl atom between amino acid residues Q and S on the β-casein ligand. This discovery led to the need for more experimental assays in the lab to further strengthen the results.

After successfully producing the recombinant his-tagged TEV protease in the host organism *E. coli*, the first assay to confirm a digestion of β-casein was the colorimetric assay using Folin-Ciocalteu´s phenol reagent. The reagent consists of the two acids phosphomolybdate and phosphotungstate, which are yellow in their oxidized form but turn blue after reacting with reducing phenolic compounds. The phenolic compounds are usually released upon digestion of substrates by enzymes. The blue colour is directly proportional to the concentration of the released reducing phenolic compounds and can be measured using a spectrophotometer to measure absorbance. The quantification can be done by comparing to a phenolic compound

used as a standard with known concentrations. In this case, tyrosine was used as a standard to measure the proteolytic activity. The assay showed that a phenolic compound was released upon interaction between β-casein and TEV protease, however further investigation was needed to confirm a digestion by the enzyme. Another assay used was SDS-PAGE technology, which showed slightly more bands on the β-casein control (undigested protein) compared to digested β-casein. To see if these bands corresponded to a peptide obtained after cleavage with TEV protease, a more detailed evaluation had to be done using MALDI-TOF mass spectrometry. At first glance, the results from these showed very little difference between the β-casein control (undigested) and the digested β-casein, however a peak at approximately 17.5 kDa could be observed on the digested sample of β-casein and not (or at least much less) on the unhydrolyzed control. Another small peak was observed at 7649.959 Da in the digested sample. Both polypeptide fragments obtained after digestion correspond well to the weights (when checked on Expasy´s ProtParam tool on the internet) of these peptides after a possible digestion at the hypothesized location of the β-casein. The first weight includes the N-terminal of the polypeptide to the Q (Glutamine) while the second polypeptide fragment includes the S (Serine) after digestion and downstream to the C-terminal. If TEV protease did, in fact, digest the β-casein at this position, there could be problems when using this protease in the future as a reliable tool for precise protein processing. However, to confirm these obtained results more extensive analysis would be needed.

The activity of the recombinant TEV protease was later investigated using a synthetic polypeptide constructed specifically as a target for the protease. This synthetic polypeptide had an 8-time repeated amino acid sequence where each repetition ended in the recognition site for TEV protease and a histidine-tag at the very end. When running the SDS-PAGE and mass spectrometry assay with both an undigested polypeptide and a digested sample, a difference could be observed. A peak at roughly 2403 kDa could be observed in the digested polypeptide but not in the control. This peak corresponded well to the weight of one of the 8-time repeated peptides, indicating a successful cleavage of this part of the polypeptide. The peptide products were later analysed for bioactive properties and showed a dual-inhibiting function (antihypertension and antidiabetic). This makes for a possible production of a multi-functional therapeutic agent in the future.

Abstract

Although different types of affinity tags have become a tool in the production and purification of recombinant proteins in various applications, the use of these have the risk of interfering with the biological activity and crystallization of proteins. There are both chemical and enzymatic ways to remove these tags, the enzymes include enteropeptidase, thrombokinase and protease to mention a few. The use of affinity tag-removing proteases has specifically proven to be useful. One of these enzymes include Tobacco Etch Virus (TEV) protease which recognizes the amino acids sequence ENLYFQ*G/S and cleaves between the Q and G/S, making it ideal for precise protein processing. Some limitations to these enzymes, like most other, are the fact that it has auto-cleaving properties which reduces the enzyme stability. In addition, the enzyme has solubility problems when it is produced by *Escherichia coli*. This can be circumvented by creating a fusion form of the protease with the chaperone MBP (maltose-binding-protein) which also prevents the auto-cleaving function. This thesis describes the production of a recombinant form of the Tobacco Etch Virus protease, analysis of the enzymatic reactions this enzyme has with different substrates and looks at various applications the obtained peptides can have. The substrates include a synthetic polypeptide created with the characteristic recognition site of TEV protease. The other substrate was β-casein which has a similar looking amino acid sequence compared to the recognition site for TEV protease. The methods leading this investigation included computational methods such as the visualisation and molecular modelling tool UCSF Chimera, where a docking could be achieved showing the interaction between the TEV protease and the substrates. Another computational tool included the use of A Plasmid Editor (ApE) program to construct the recombinant genes used for successful cultivation of the host organism *E. coli* and the expression genes for the produced polypeptides. The practical methods included a colorimetric assay with Folin-Ciocalteu's reagent showing activity based on a reduction reaction led by reducing phenolic compounds in β-casein, SDS-PAGE and mass spectrometry analysis of the hydrolysed polypeptides. All results showed that TEV protease had activity and was not as specific as previously thought since a hydrolysis of the β-casein could be observed. The hydrolysis of the synthetic polypeptide could play a beneficial role in the future when manufacturing therapeutic agents since the peptides derived from the synthetic polypeptide after digestion have dual-inhibiting bioactivity against hypertension and type 2 diabetes.

Key words: Tobacco Etch Virus protease, synthetic polypeptide, β-casein, UCSF Chimera, ApE (A Plasmid Editor), *Escherichia coli*, MBP (maltose-binding protein), SDS-PAGE, mass spectrometry, hypertension, type-2 diabetes.

Acknowledgments

Although I wrote this thesis alone, I have a lot of people to thank for believing in me and for pushing me in the right direction. I would like to thank my supervisor Javier Linares-Pastén for taking his time to educate me on everything from the theory of the subject at hand to the training of the various programs that were used to make this thesis possible. I also want to thank Carl Grey and Martin Hedström for taking their time to get to know what my biggest interest was regarding enzymes and then giving me Javier Linares-Pastén as the right supervisor. I also want to thank everyone at the division for their kindness and help with everything. Especially Savvina Leontakianakou who helped by providing not just competent cells but also information. And of course, my family and wife for believing in me and inspiring me to do my best.

Lund, June 2024

Fisnik Nerjovaj

Table of Contents

List of Tables & Equations

List of Figures

1. Introduction & Aim

The TEV (Tobacco Etch Virus) protease is an enzyme produced naturally from a virus on the tobacco plant. The protease has the capabilities to cleave certain sequences very specifically on polypeptides. This has many other uses, including removing affinity tags. The first aim of this thesis was to produce both the recombinant TEV protease and the polypeptide it is meant to cleave for further investigation as well as evaluating a possible cleavage of β-casein that has a similar recognition site (ENLHLPLPLLQS compared to ENLYFQG, which is the classic TEV protease recognition site).

To realise the aim the plan was as follows: the recombinant protease and synthetic polypeptide plasmids were synthesized in advance. From these genes the recombinant fusion proteins (6x his-tagged TEV protease coupled with the MBP (maltose-binding protein) chaperone and 6x histidine-tagged synthetic polypeptide) should be produced using *Escherichia coli* as a host organism. This is followed by protein purification using IMAC technology with a HisTrap coloum and SDS-PAGE for protein analysis. Finally, the enzyme reactions should be studied together with the products using mass spectrometry to see what fragments are obtained after hydrolysis with the recombinant protease. Also, in silico analysis using bioinformatics for analysing the sequence and 3D structure of these protein products and how the ligands (substrates) bind to the TEV protease active site should be conducted. The resources for this thesis will be provided by Kemicentrum at Lund University. These include a lab with all the required necessities, from cultivation of *Escherichia coli* with the plasmid containing genes for protein production to analysing equipment for the purified and cleaved proteins. Moreover, an office with a computer for both computational and writing requirements is provided.

Another aim of this project is to further improve on the understanding of the TEV proteases´ sequence specificity and to evaluate if there is a possibility that this protease can recognize and cleave sites with similar amino acid sequence in another protein. A more general aim is to understand more about the capabilities of enzymes as a tool in the laboratory environment and to further enhance the knowledge of enzymes in general.

2. Scientific background

To understand more about the project at hand, an extensive scientific literature study about enzymes, their reaction mechanism and the protein production, purification and analysis methods had to be conducted before putting that knowledge into practice.

2.1 Enzymes and their reaction mechanism

A chemical reaction occurs spontaneously when the overall free energy is negative ($\Delta G<0$). The overall free energy can be calculated from the following equation (see **Equation 1**) below:

EQUATION 1 GIBBS FREE ENERGY

$$\Delta G = \Delta G^0 + R * T \ln K \, , K = \left(\frac{[C][D]}{[A][B]}\right)$$

where ΔG^0 is the free energy for the reaction during standard conditions for both reactants and products at 1.0 M. The fraction between concentrations in the equation above is often termed the equilibrium constant K. At equilibrium, the ΔG value is equal to zero giving the following equation:

$0 = \Delta G^0 + RT \ln\left(\frac{[C][D]}{[A][B]}\right)$ which also can be written as $\Delta G^0 = -R * T \ln\left(\frac{[C][D]}{[A][B]}\right)$

A negative ΔG is achieved when the free energy for the product state is lower than for the reactants in a chemical reaction, leading to a spontaneous reaction from reactants to products (see **Figure 1**). During a chemical reaction, typically covalent bonds are broken and new bonds are formed. However, for a brief period, a chemical state in which both old and new bonds exist is first reached. This *transition state* has the highest free energy of the reaction and is therefore very unstable and exists only for a short time before it converts into the end products with new covalent bonds. The reaction, however, is considered reversible, which means that the reaction also can proceed back to the original reactants from the transition state

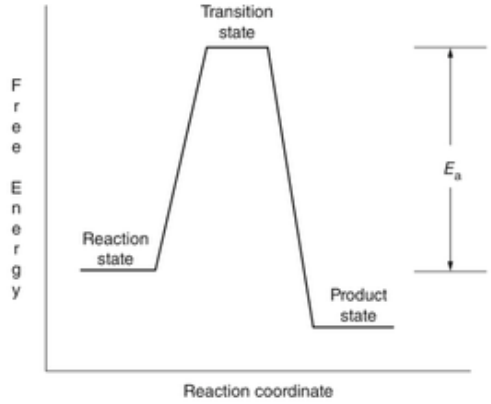


FIGURE 1 A DIAGRAM SHOWING THE FREE ENERGY FOR THE DIFFERENT STATES.

The figure also shows E_a, the molar activation energy required to reach the transition state. A higher activation energy corresponds to an exponentially lower reaction rate according to the Arrhenius equation see (**Equation 2**):

EQUATION 2 REACTION RATE CONSTANT

$$k = Ae^{\frac{-E_a}{R*T}}$$

Where k is the reaction rate constant at which a reaction occurs, and A is the frequency factor which describes the number of collisions that have energy enough to surpass the activation energy. The reaction rate decreases therefore exponentially with E_a needed to achieve transition state (see **Figure 2**) [1].

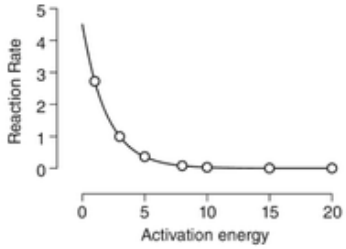


FIGURE 2 EXPONENTIAL DECREASE IN REACTION RATE AS A FUNCTION OF THE ACTIVATION ENERGY REQUIRED FOR THE TRANSITION STATE

2.2 Proteases

A protease is a type of enzyme that cleaves polypeptides by hydrolysing the peptide bonds generating shorter peptides or amino acids. This catabolic reaction can lead to protein products with altered functions than the originally cleaved polypeptide. The proteases can be divided into six groups based on the reaction mechanism. The groups are serine proteases, cysteine proteases, aspartic proteases, threonine proteases, metalloproteases and glutamic acid proteases [2].

2.3 Substrates

This section covers the substrates used in this thesis where the location of hydrolysis on these proteins by TEV protease is determined. The substrates include both the β-casein, which was evaluated first for potential peptide products, and the synthetic polypeptide (denoted SPP in the following sections to come)

2.3.1 β-casein

As mentioned earlier the β-casein, which has a molecular weight at roughly 25 kDa and a size of 224 amino acids, does not have the characteristic recognition site that TEV protease typically recognizes. Instead, a similar 12-amino acid long site (ENLHLPLPLLQS) is present between amino acids 131 to 142, where a potential cleavage between amino acids Q (Glutamine) and S (Serine) can occur. The entire β-casein sequence is as follows:

MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPF AQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVE PFTESQSLTLTDVENLHLPLPLLQ*SWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYP QRDMPIQAFLLYQEPVLGPVRGPFPIIV

A cleavage would potentially give at least two peptides with the following sequences:

MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIHPF AQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAPKHKEMPFPKYPVE PFTESQSLTLTDVENLHLPLPLLQ M_w = 17473 Da

SWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPF PIIV M_w = 7652.11 Da

These peptides are therefore expected to show up in the assays conducted in these experiments and weigh roughly 17473 and 7652 Da respectively [3]. The second peptide has

potential antidiabetic, antihypertension, antioxidative and immunomodulating effects when looking it up on the BIOPEP-UWM website which describes the bioactivity of this peptide [4].

2.3.2 Synthetic polypeptide (SPP)

The other polypeptide is synthesized alongside TEV protease and weighs around 20 kDa (193 amino acids). Much is unknown about this polypeptide since it is synthesized specifically for TEV protease digestion. The cleaved amino acid sequences are an 8-time repeated sequence ending with the recognition site for TEV protease after each repeated peptide. These peptides also have potential antidiabetic and antihypertension effects [4] when checking for bioactivity. The entire amino acid sequence is as follows:

MQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*GQHPHGLGALCAAPPSTENLYFQ*G LEHHHHHH*

After hydrolysis the obtained peptides should include:

MQHPHGLGALCAAPPSTENLYFQ M_w = **2480.14 Da**

GQHPHGLGALCAAPPSTENLYFQ M_w = **2406.13 Da**

GQHPHGLGALCAAPPSTENLYFQ

GQHPHGLGALCAAPPSTENLYFQ

GQHPHGLGALCAAPPSTENLYFQ

GQHPHGLGALCAAPPSTENLYFQ

GQHPHGLGALCAAPPSTENLYFQ

GQHPHGLGALCAAPPSTENLYFQ

GLEHHHHHH* M_w = **1139.5 Da**

With the corresponding weights written next to them.

2.4 TEV protease and its enzymatic mechanism

The Tobacco Etch Virus protease is a cysteine protease that has a weight of approximately 27 kDa (69.5 kDa when fused with the MBP chaperone as in this case) with an optimum temperature at 30-34 °C. The melting point for TEV protease is 44°C (which was observed during a Differential Scanning Fluorimetry experiment [5]. See **Appendix 3**, **Figure 25**) and pH optimum at 6-9, however the stability of the protease increased at 4 °C and the activity was deemed sufficient, making TEV protease suitable for reactions between temperatures 4-34 °C. This was shown during the experiments conducted in this thesis. Another observation and tested property of TEV protease is the denaturation of the enzyme which occurs after two weeks when stored at 4 °C. Since the TEV protease has a stringent sequence specificity, as described earlier, it can be used as a tool to hydrolyse desired sequence sites. One example of this enzymatic reaction is the removal of undesired histidine-tags in the final protein product. However, a drawback to the TEV protease is its ability for autolysis which generates a more

inactive truncated form of the protease. There are several ways to prevent this, one being by adding a fusion protein (e.g. MBP) to the protease which prevents the autolysis effect [6].

The TEV protease recognition site consists of the following amino acid sequence Glu-Asn-Leu-Tyr-Phe-Gln*Gly, where Gln and Gly at the end is required for cleaving. But the Gly at the P1' position is interchangeable with other amino acids without compromising the functionality of the protease. The auto-cleaving activity this enzyme has can be reduced by the MBP (maltose-binding protein) chaperone which is fused with the protease. These are later purified using a two-step affinity chromatography where the MBP is removed in the first step after the his-tagged protease is attached to the HisTrap and binding buffer washes all undesired proteins, including the fusion protein. This is followed by elution and fractionation of the his-tagged TEV protease in IMAC.

The catalytic reaction of TEV protease (see **Appendix 1** for complete overview) can be, when simplified, seen as two halves of a reaction which yields two different transition states. In the first part of the reaction, which is called the acylation step, the His_{46} is the general base that abstracts a proton from the catalytic cysteine (Cys_{151}) to produce the thiolate (RS^-) which acts as a nucleophile against the carbonyl atom on the target peptide. This leads to the formation of the first transition state (TS1) which has a tetrahedral formation. The newly produced and negative oxygen ion takes up electrons from the carbonyl double bond and is stabilised by the oxyanion hole which is constituted by the hydrogen bond interactions that occur with the backbone atoms of residues 149 and 151 to further stabilizing the transition state. The collapse of the first transition state leads to the release of the C-terminal product fragment and the formation of a covalent thioester enzyme-substrate complex. The next part of the reaction is the diacylation step and involves the release of the N-terminal product as well as the regeneration of the TEV protease. This is achieved when a hydroxide ion (obtained after His_{46} subtracts a proton from surrounding water molecules) acts as a nucleophile and attacks the thioester carbonyl atom and forms the second transition state (TS2) which also has the same tetrahedral shape and stabilized by the oxyanion hole. The catalytic cysteine residue is regenerated after the breakdown of TS2, and cleavage of the target peptide bond is complete [7].

2.5 Utilized methods

To be able to conduct the experiments needed for the production, purification and analysis of all proteins, it is of great importance to know the equipment and methods that are involved in these steps.

2.5.1 Nanodrop

The use of nanodrop provided the concentrations of all proteins involved in the experiments. However, the equipment was also used to determine the concentration of the plasmids during transformation of the BL21 (DE3) *E. coli* cells [8].

2.5.2 Agarose gel electrophoresis

By using gel electrophoresis, the recombinant TEV plasmid could be analysed to see if it was intact or have undergone some type of degradation. This method was used both on the working stock solution that was made (1:10 dilution with nuclease-free water) and the original commercial stock [9].

2.5.3 Recombinant gene technology

The first step in creating the ligand for the TEV protease is to determine the amino acid sequence of the polypeptide, which in this case is an 8-time repeated sequence of a smaller peptide, which has potential anti-hypertension and antidiabetic properties. This is followed by the recognition site for the recombinant TEV protease and the 6x his-tag. The start and stop codon and enzyme restriction sites (NdeI and XhoI) were also important to locate to ensure that the ligand is placed at the correct position in the plasmid pET21b for translation. This was achieved using the ApE (A plasmid Editor) program which allows the user to design and present DNA sequences and plasmids. The software was used to insert the target ligand between the desired restriction sites mentioned earlier. The recombinant pET21b plasmid can be seen in the figure below (see **Figure 3** which was obtained by using the ApE program) which has another important property, the ampicillin resistance genes that are going to be useful in the transformation process. [10]. The promoters, where the transcription for both of these expression genes start at, can also be seen in the figure below.

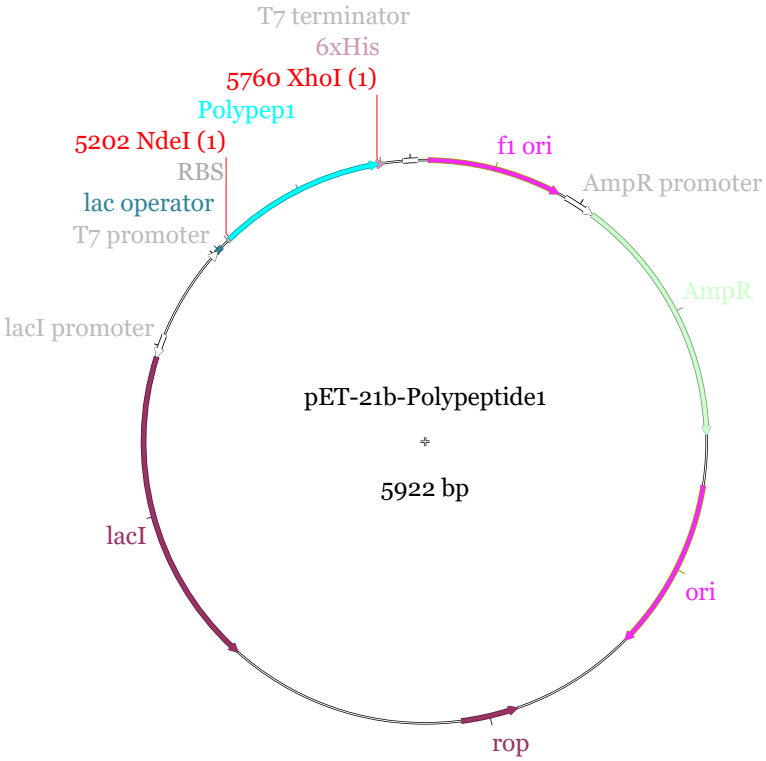


FIGURE 3 DIFFERENT SEGMENTS OF THE PET21B PLASMID WITH THE TARGET POLYPEPTIDE BETWEEN RESTRICTION SITES NDEI AND XHOI.

2.5.4 Competent cell preparation

Bacterial transformation is a process used to produce a quantitative amount of recombinant DNA copies. This is done through uptake of the DNA molecules through the cell wall from the external surroundings. The DNA is then incorporated into the genome of the target microorganism. Before transformation can occur, the recombinant plasmids are created using the abovementioned techniques. After that the transformation is divided into three steps:

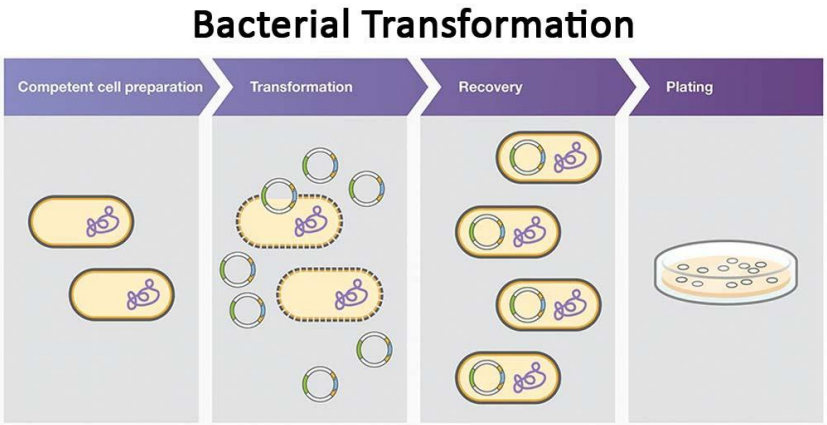


FIGURE 4 BACTERIAL TRANSFORMATION WORKFLOW

1) Competent cell preparation and transformation

Competent cell preparation is a method used to increase the competency (ability to undergo transformation) of the bacteria used for the cloning workflow. The most common bacteria used is *E. coli*, which naturally has a very low competency. That is the reason why these cells must undergo this step, to make them more efficient for the transformation to occur. There are a couple of ways to do this. The method used was a heat shock transformation, where a starter culture was prepared by inoculating a single fresh colony from an Eppendorf tube into a sterile liquid medium (LB broth) with the appropriate antibiotic as a negative control. It is important that the growing cultures are harvested at an optical cell density (OD_{600}) of roughly 0.3, which corresponds to the mid-log phase ensuring optimal growth for the entire population. These are then centrifuged and the supernatants were discarded. Afterwards, two washes with $CaCl_2$ were done with centrifugations in between. The supernatants were discarded also in these steps and the cell pellets were resuspended in a solution containing $CaCl_2$ and glycerol (20%).

The used transformation technique in this project was heat shock transformation. This means that the competent cells, in a mixture with the plasmids, were subjected to a sudden increase in temperature from 4 °C in an ice bath to roughly 42 °C for 1-2 min, which makes the cell membrane more permeable and ready for the transformation of the plasmids.

2) Cell recovery

After the heat shock, the now transformed cells were incubated in antibiotic-free liquid medium for one hour at 37 °C to allow the expression of antibiotic resistant genes from the pET21b-plasmid. This step is necessary for the bacteria to recover after the heat shock.

3) Cell plating

After the cultivation on LB medium, the plating on solid LB media follows. This is done by transferring the cells onto agar plates containing antibiotics used as a marker for the transformation. The pET21b plasmids have ampicillin resistance genes. If transformed properly into competent cells, they prevent the growth of undesired microorganisms. [11]

2.5.5 Recombinant protein production

The first step in protein production, after cultivation of the transformed *E. coli*, was to prepare an inoculum in LB media together with the antibiotic (Ampicillin for TEV protease production), as these cells are resistant to it and used as a selection marker. This inoculum was left in a shaking incubator at 37 °C overnight. The following day, the inoculum was mixed with a larger amount of LB media together with an equally larger amount of antibiotic into a shake flask. The mixture was left on a shaking incubator and the optical density (measuring the cell growth) was monitored until an OD of 0.6-1.0 was achieved after which IPTG (isopropyl-β-d-thiogalactoside) was added, which induces the protein expression [12].

2.5.6 Protein purification using IMAC technology

The protein purification step was done by using Immobilized Metal-ion Affinity Chromatography (IMAC) which is a liquid chromatography method where the column is coated with a certain immobilized metal-ion support. In this case it was coated with nickel ions, Ni^{2+}, which will allow for the his-tagged recombinant protein to attach to the column while other proteins get washed out. The column was first washed with an appropriate binding buffer (containing 100 mM Tris-HCl, 0.5 M NaCl at pH 7.4) which enables protein to attach more readily to the column, before adding the produced recombinant proteins into the column. The column was then washed again with the same binding buffer, to ensure a proper wash of unwanted proteins and the separating of the MBP from the TEV protease, before adding elution buffer (containing 100 mM Tris-HCl, 0.5 M NaCl, 0.5 M Imidazole at pH 7.4). The imidazole competes with the his-tagged proteins in the column and therefore elutes them to be fractionated separately. The fractions were collected and measured using a concentration measuring device like NanoDrop [13].

2.5.7 Protein analysis

To determine the activity of the TEV protease, various analytical methods were conducted to analyse both protease activity and determine concentration of the products after digestion. These included concentration measurement with NanoDrop, SDS-PAGE to determine the possible cleaved peptides, colorimetric assay to determine activity and mass spectrometry to determine the fragments after hydrolysis with the recombinant protease.

2.5.7.1 Colorimetric assay with Folin-Ciocalteu's reagent

The used colorimetric assay in these experiments included the Folin-Ciocalteu´s phenol reagent. The β-casein (0.5% (w/v)) acts as a substrate and was mixed with 50 mM phosphate buffer (pH 7.4). The procedure is based on the reduction of phosphomolybdate and phosphotungstate (colored yellow in their oxidized form and blue when reduced) by a released phenolic compound during proteolytic activity. The blue color could be observed in a

sample containing Folin-Ciocalteu´s reagent with a mixture of TEV protease and β-casein, indicating a possible digestion

Using a standard tyrosine sample (which is a phenolic compound) in different concentrations (5-50 μg/mL), the proteolytic activity of the recombinant protease could be determined by measuring the released concentration of tyrosine for the enzymatic reaction between TEV protease and β-casein. This can be done by making a calibration plot of the absorbance as a function of the concentration tyrosine and then comparing it to the sample solution [14].

2.5.7.2 SDS-PAGE

SDS-PAGE is a method used to separate proteins based on their molecular size and charge. These two factors determine how fast these travel through the gel. The SDS (sodium dodecyl sulfate) is a type of detergent used to denature the proteins being analyzed while the PAGE (polyacrylamide gel electrophoresis) describes the method of which these denatured proteins travel the gel, by using an electric current. When the analyzed proteins are boiled with SDS they gain a negative charge that corresponds to their molecular size which determines the speed of travel through the positively charged gel. Therefore, the smallest proteins are also the fastest traveling.

In the case of these experiments, the gel electrophoresis yielded information that corresponded well to the theory concerning the molecular masses of the proteins investigated. These results can be seen in **Figures 17** and **22** of the results section. [15]

2.5.7.3 Mass spectrometry

Mass spectrometry is a method used to analyze complex protein samples. These measurements are done in gas phase with ionized analytes. Ions are made by the first part of the mass spectrometer, the ion source. The second part, the mass analyzer, is responsible for separating the ionized analytes from the ion source by mass-to-charge ratio (m/z). The two most used techniques to ionize proteins in mass-spectrometry based proteomics are ESI (Electrospray ionization) and MALDI (Matrix-assisted laser desorption/ionization).

In this thesis, the used method to analyze the digestion fragments of β-casein and the synthetic polypeptide by TEV protease was MALDI. The results of these fragments with their corresponding intensities can be seen in the results section. [16]

2.5.7.4 Bioinformatics

To know more about the proteins examined in this thesis, bioinformatics was conducted to determine the primary, secondary, and tertiary structure of the recombinant proteins. This was done by searching the Protein Data Bank (PDB) on the web for an active Tobacco Etch Virus protease with the pdb code 1LVM. Using the UCSF Chimera program, a docking for the target ligands (ENLYFQG and ENLHLPLPLLQS) could be made with AutoDock Vina. This is illustrated for ENLYFQG in the figures below, which also shows the hydrophobicity for the TEV protease. As can be observed the active site is mostly hydrophilic (red) with slight hydrophobic parts (blue) [17].

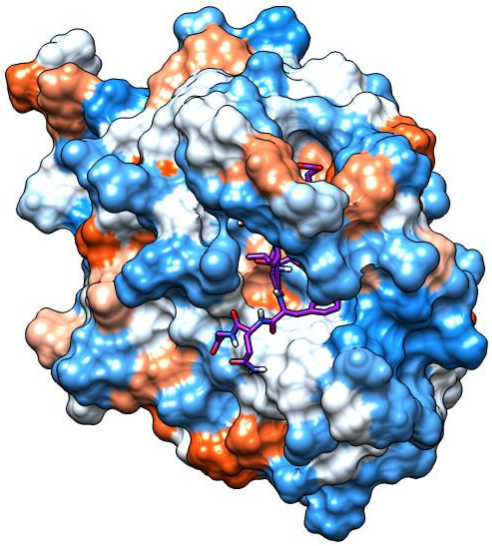


FIGURE 5 AN IMAGE SHOWING THE HYDROPHOBICITY SURFACE STRUCTURE OF TEV PROTEASE WITH THE LIGAND BOUND TO THE ACTIVE SITE

The important amino acid sequences in this analysis lies on the TEV protease recognition site (ENLYFQG) on the target protein and the catalytic triad this ligand binds to which in TEV protease consists of the catalytic triad mentioned earlier (His_{46}, Asp_{81} and Cys_{151}). This makes TEV protease a cysteine protease that has an enzymatic mechanism similar to the serine proteases.

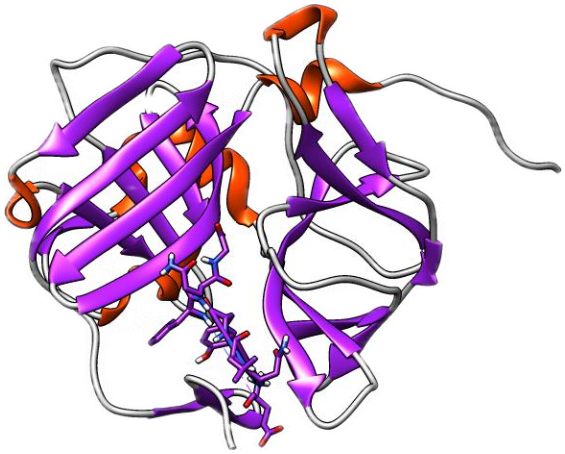


FIGURE 6 AN IMAGE SHOWING THE SECONDARY STRUCTURE OF TEV PROTEASE WHICH CONSISTS OF TWO ANTI-PARALLEL B-BARRELS (IN PURPLE).

2.6 Enzyme hydrolysis reactions

Before hydrolyzing the synthetic polypeptide (SPP), another substrate (β-casein) was digested using the recombinant TEV protease. This reaction may have yielded unexpected results since the TEV protease is extremely specific in its recognition pattern. The results for this digestion can be observed in the results section below which are obtained by using UCSF Chimera program, by measuring the activity of the reaction with FC (Folin-Ciocalteu´s) reagent as described earlier, by conducting an SDS-PAGE with potential hydrolyzed sample and its unhydrolyzed negative control and finally by evaluating the fragments obtained after these samples are run in the

mass spectrometer. Regarding the theoretical approach of this investigation, a docking using the recombinant TEV protease as a receptor and the target ligand ENLHLPLPLLQS was made [18].

3. Material & Methods

This section describes an in-depth look at all the materials used as well as the methods for different parts of the process from production and purification to protein analyses. NOTE! The protocols in this section were obtained from a Microsoft Teams group or provided by the supervisor.

3.1 Resuspension of pET21b synthetic plasmid

NOTE! Besides heating/cooling and centrifugation steps, be sure to work under sterile conditions to avoid decontamination.

1. Centrifuge tube containing plasmids (1-2 min) (be sure to balance with another tube by measuring weight beforehand).
2. Add 50 µL of nuclease-free water.
3. Put Eppendorf in heater (50 °C, 10 min).
4. Put on ice for 10 minutes.
5. From the stock solution, make working solution using nuclease-free water with a dilution factor of 1:10. Mix properly.
6. Put in freezer (-20 °C).

3.2 Competent cell preparation

NOTE! All solutions ($CaCl_2$, $CaCl_2$ + 20% glycerol and LB medium) and material (shake flasks, beakers and Eppendorf tubes) need to be autoclaved in advance and all steps (besides centrifugation) need to be done under sterile conditions to avoid decontamination. From step 3, it is important to keep cells on ice.

1. From an old batch take 10 µL and transfer it to a falcon tube with 3 mL of sterile LB media for *E. coli* (BL21 (DE3)) and leave overnight at 37 °C shaking.
2. Transfer 150 µL from each falcon tube to each shake flask containing 30 mL of sterile LB and keep at 37 °C shaking. Measure the OD_{600} until it goes up to 0.3 for both *E. coli* strains in the two tubes.
3. Divide the content into two 50 mL falcon tubes and incubate in ice for 10 min before centrifugation (4500 rpm and 4 °C) for 10 min. Discard the supernatant.
4. **First wash:** To each falcon tube add 15 mL of 150 mM calcium chloride ($CaCl_2$) and resuspend the pellets by slow pipetting 10 times.
5. Centrifuge for 10 minutes again and discard the supernatant.
6. **Second wash:** resuspend both pellets in a total volume of 3 mL each using 150 mM $CaCl_2$ + 20% glycerol.
7. From the final 3 mL of each tube, take aliquots of 150 µL and pour them on sterile Eppendorf tubes (labelled) and store at -80 °C.

3.3 Transformation

NOTE! Besides the heating/cooling steps, all steps are done under sterile conditions to avoid decontamination.

1. Using the working solution for the recombinant protein, add 2 µL of the pET21b plasmid (be sure to mix by pipetting before taking up and when added) to each aliquot of the *E. coli* (BL21 (DE3), stored in -80 °C) stored in ice and wait 15 minutes.
2. Put the Eppendorf tubes in a heat block (42 °C) for 2 minutes.
3. Put the tubes back in ice for 10 minutes.
4. Add 300 µL sterile LB to each tube, slowly pipet up and down 1-2 times and incubate at 37 °C for 45-60 minutes.
5. After incubation, add 100 µL of each strain to two agar plates containing solid LB with antibiotics used as a marker for the transformation.
6. Spread out the droplet sideways first by tilting these and then gently across the entire agar plates using a spreader. Wait 15 minutes and turn upside-down to prevent condensation to ruin the cultivation.
7. Label the agar plates accordingly and incubate at 37 °C overnight.

3.4 Protein Production & Purification

NOTE! Steps 1-5 are done in sterile environment.

1. Prepare an inoculum by adding 3 mL sterile LB media together with 3 µL of antibiotic (Ampicillin for TEV production) to a falcon tube. Using an inoculation loop, scoop up a colony and mix it in the falcon tube.
2. Incubate overnight at 37 °C and shaking at 150-200 rpm.
3. Transfer the inoculum to an Erlenmeyer flask containing 300 mL sterile LB media and 300 µL antibiotic.
4. Incubate at 37 °C and shaking at 150-200 rpm and measure the OD_{600} until it reaches 0.6-1.0.
5. Add 300 µL IPTG and leave at room temperature (or 20 °C) on a shaking device overnight.
6. Transfer the content from the Erlenmeyer flask to a plastic tube used for centrifugation at 8000 rpm for 10 min (mind the weight distribution). Discard supernatant!
7. Resuspend the pellet in 15 mL binding buffer and transfer to a falcon tube.
8. Sonicate the content on ice for 15 min and transfer the content to an ultra-centrifugation tube
9. Centrifugate for 20 min at 12000 rpm (mind the weight balance, compensate with binding buffer if necessary).
10. Collect the supernatant for IMAC purification.

For IMAC purification using ÄKTA equipment:

1. Use a HisTrap [19] column to bind the his-tagged protease.

2. Prepare two bottles containing both binding buffer (Tris-HCl 100 mM. NaCl 0.5 M, pH 7.4) and elution buffer (Tris-HCl 100 mM, NaCl 0.5 M, Imidazole 0.5 M, pH 7.4) according to lab protocol.
3. Purge all tubes on the IMAC with a manual run with MilliQ water before running the sample with a pre-determined method run.
4. Run the sample carefully (almost all sample) to prevent air to flow into the tube.
5. Collect the fractions from the elution for concentration determination with NANODROP.

Concentration analysis with NANODROP:

1. Clean both top and bottom sensors using 2 μL of MilliQ water.
2. Register a blank with MilliQ water.
3. Read your sample.

3.5 Protein analysis

3.5.1 SDS-PAGE

1. Mix 16 μL of sample with 4 μL of loading buffer in an Eppendorf tube.
2. Spin the tube to mix properly.
3. Heat the sample mixture with loading buffer for 10 minutes at 100 °C to (negatively charge the sample).
4. Spin the tube again.
5. Load 10 μL of the sample to an unstained gel (prepared beforehand).
6. Run the gel in an electrophoresis system for 45-60 minutes. First 15 minutes at 180 V and then 200 V for the rest of the time until the protein reaches the end.
7. Read the results at a gel imaging system.

3.5.2 Colorimetric Assay with Folin-Ciocalteu's reagent

1. Add TEV to 0.5 % (w/v) β-Casein in 50 mM phosphate buffer so that the final concentration for TEV is 0.1 mg/mL and final volume of 1 mL.
2. Let digestion occur for 1 hour at 20 °C.
3. End reaction by adding 1 mL of TCA (Trichloroacetic acid) [20] (10 % v/w) and incubate for 10 minutes at 20 °C.
4. Centrifuge the sample (10000 rpm, 10 min).
5. Take 150 μL of the supernatant and mix with 750 μL of Na_2CO_3 (0.5 M) and 225 μL of Folin-Ciocalteu's phenol in an Eppendorf tube.
6. Incubate for 30 minutes at 20 °C with an aluminium foil over to keep in dark.
7. Measure the absorbance (abs_{650}) in microplates together with tyrosine as a standard for concentration 5-50 μg/mL.

Tyrosine standard preparation and blank:

By using a final volume of 150 μL and a tyrosine initial concentration of 100 μg/mL, the volume of the tyrosine added can be calculated depending on the desired final concentration

of tyrosine. For 50 μg/mL for instance the added tyrosine volume is calculated to 75 μL (to get a final volume of 150 μL the rest can be filled with MilliQ water).

When using the microplate, it is favourable to make triplets of every sample and standard concentration. Keep in mind to make a blank for each measurement (even tyrosine standard) using 150 μL of MilliQ (or tyrosine standard) and mixing with same amount of Na_2CO_3 and Folin-Ciocalteu´s reagent as for the sample.

4. Results

The results obtained involves successful cultivation of the *E. coli* (BL21 (DE3)) host cells as well as successful production and purification of the proteins involved. The section concludes with various protein analysis assays.

4.1 Competent cell preparation

In the first attempt of preparing competent cells, the method was conducted using a faulty protocol where the inoculation volume before transformation was too large (3 mL instead of 150 μL). This led to the *E. coli* being produced at a much faster rate and the window at which these were suitable for harvest were much smaller compared to the window on the second try. The countermeasures provided by the supervisor was to dilute in 30 mL sterile LB medium, however it did not provide satisfying results. The results for the first try of producing competent cells can be seen in the appendix section (**Appendix 3**). As we did not get any colonies, a second run using a smaller inoculum was tested. This run had a longer time span for reaching the mid-log phase of the growth, due to the smaller inoculum, and was easier to measure. However, the results did not yield any competent cells either. To circumvent this issue, a commercial stock with competent BL21 (DE3) was provided and this gave growth on the agar plates (with ampicillin as a selection marker). The results for the bacterial growth can be observed in **Table 1** and **Figure 7A, 7B and 8** below:

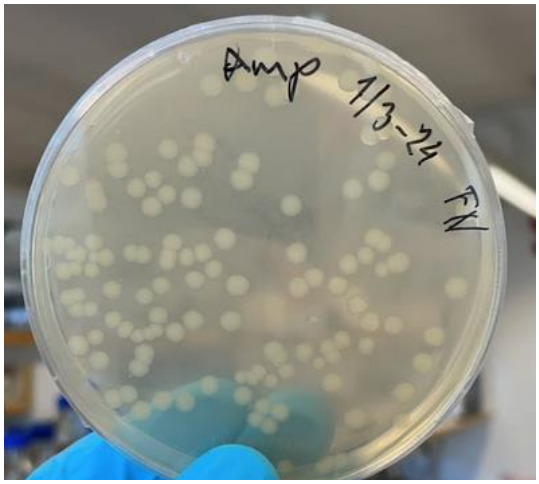


FIGURE 7A COLONIES OF BL21 (DE3) WITH PET21B PLASMID CONTAINING TEV PROTEASE RECOMBINANT EXPRESSION GENE

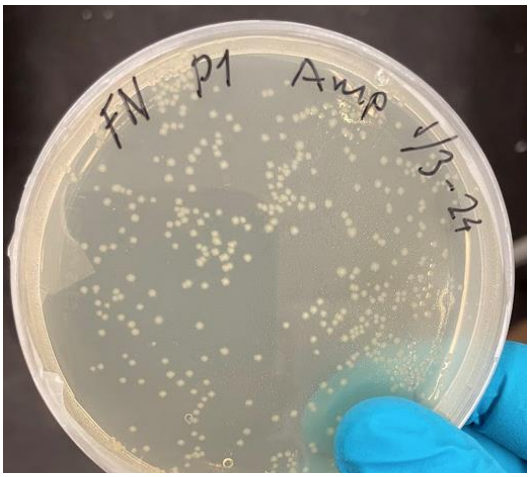


FIGURE 7B COLONIES OF BL21 (DE3) CONTAINING THE PET21B PLASMID WITH GENES FOR THE SYNTHETIC POLYPEPTIDE

TABLE 1 OD MEASUREMENTS FOR COMPETENT CELL GROWTH

E. coli strain / Time	BL21(DE3)
$T_{0, (08:09)}$	0.043
$T_{1, (08:52)}$	0.060
$T_{2, (09:40)}$	0.122
$T_{3(10:22)}$	0.297
$T_{final\ (10:39)}$	0.385

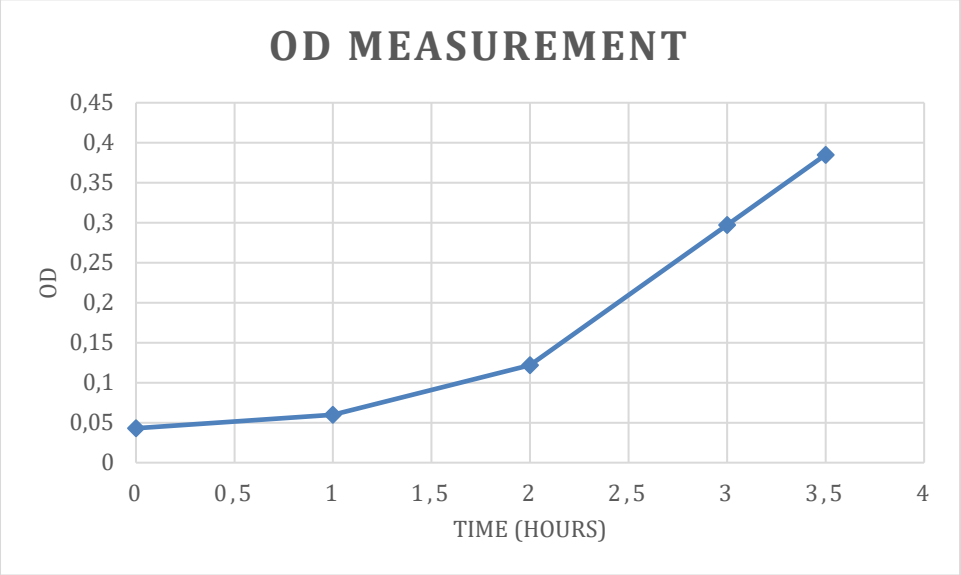


FIGURE 8 INCREASE IN OPTICAL DENSITY FOR BACTERIAL GROWTH.

4.2 Protein production and purification

The production process for both the recombinant proteins can be seen below. The recombinant TEV protease production is presented first followed by the production of the synthetic polypeptide (SPP). The production of β-casein was not necessary since this was purchased in advance.

4.2.1 Production and purification of recombinant TEV protease

The production of the recombinant TEV protease can be observed in this section with the optical density observed in **Table 2** and **Figure 9** below. After purification in the IMAC (see **Figure 10** below) the peak observed after 120 mL of elution yielded a protein concentration of 1.34 mg/mL when measured in the NanoDrop.

TABLE 2 OD MEASUREMENTS FOR COMPETENT CELL GROWTH FOR TEV PRODUCTION

Time	BL21 (DE3)
T_0 (09:21)	0.046
T_1 (10:37)	0.084
T_2 (11:39)	0.268
T_{final} (12:40)	0.776

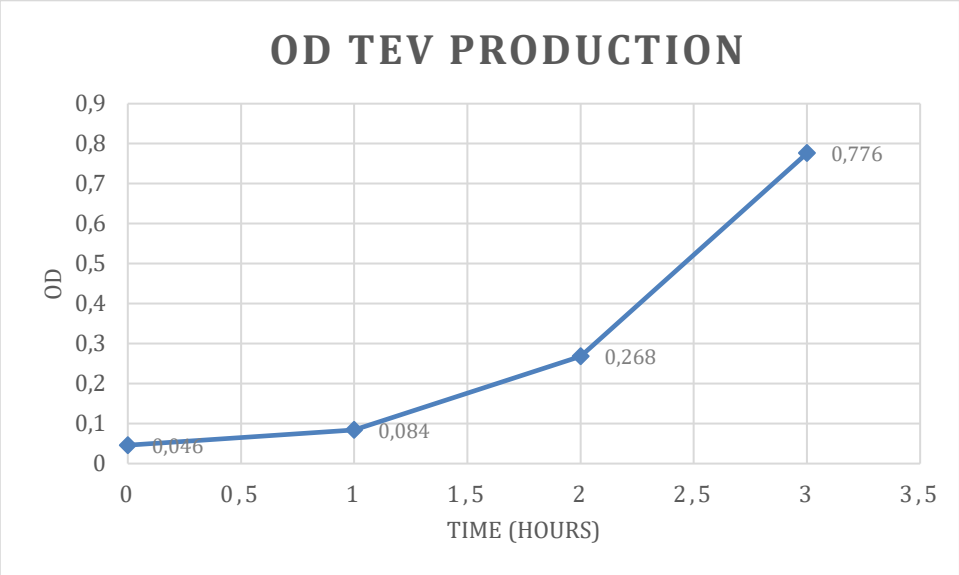


FIGURE 9 INCREASE IN OPTICAL DENSITY FOR BACTERIAL GROWTH DURING PROTEIN PRODUCTION FOR TEV PROTEASE

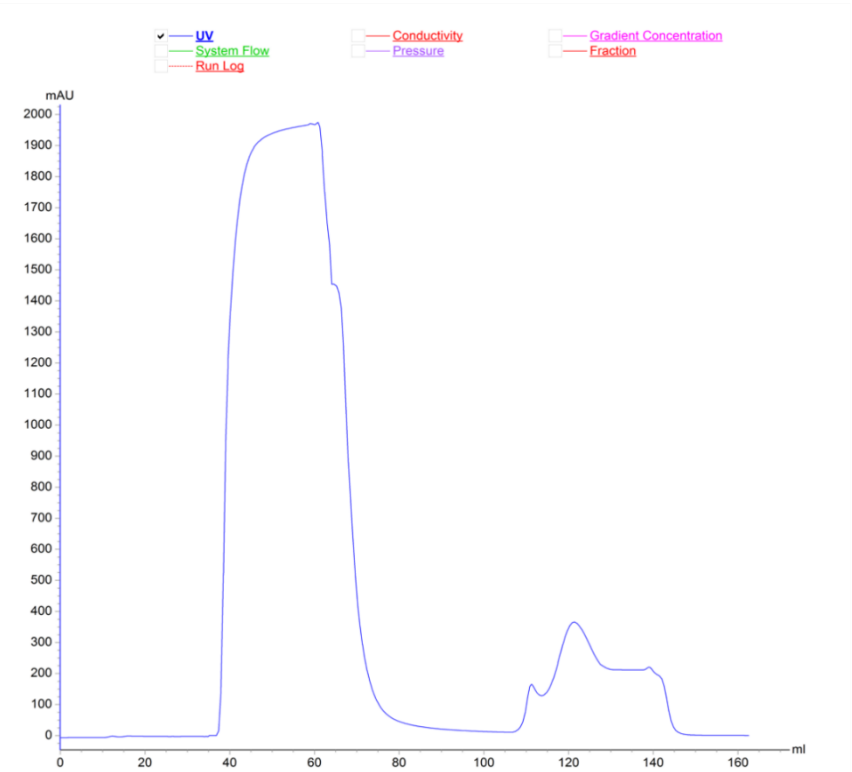


FIGURE 10 ABSORBANCE FROM SPECTROPHOTOMETER ELUTED FROM IMAC. A PEAK CAN BE OBSERVED AFTER 120 ML WHICH CORRESPONDS TO THE TEV PROTEASE ABSORBANCE PEAK.

4.2.2 Production and purification of Synthetic Polypeptide (SPP)

This section shows the production of the synthetic polypeptide which was produced alongside the recombinant TEV protease (see **Table 3** and **Figure 11** for OD measurements). After purification in the IMAC (see **Figure 12**) a very low yield of 0.18 mg/mL was obtained which led to the use of a protein concentrator tube to increase the concentration of the polypeptide to a final 1.07 mg/mL [21].

TABLE 3 OD MEASUREMENTS FOR COMPETENT CELL GROWTH FOR SPP PRODUCTION

Time	BL21 (DE3)
T_0 (08:56)	0.058
T_1 (10:08)	0.082
T_2 (11:23)	0.288
T_{final} (12:45)	0.614

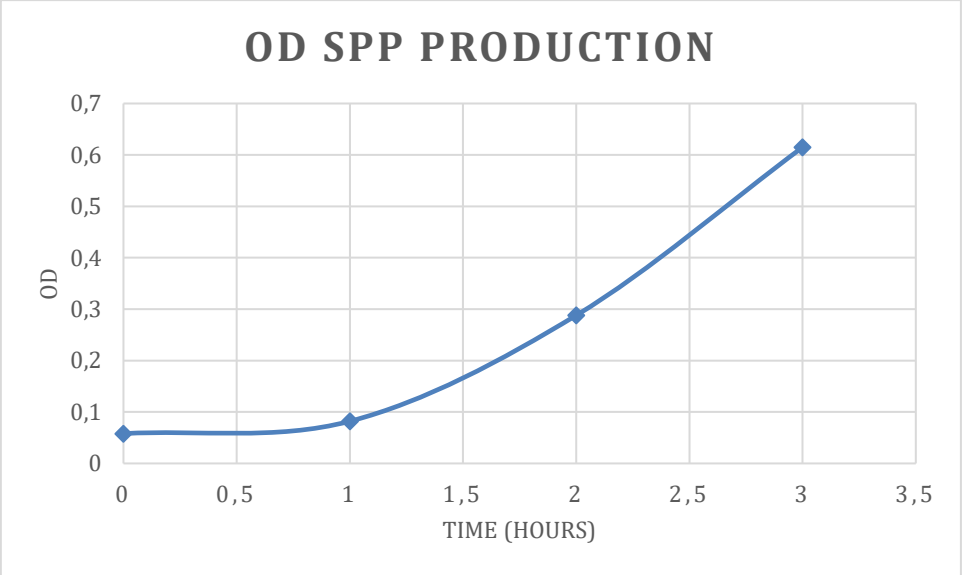


FIGURE 11 THE INCREASE IN OPTICAL DENSITY FOR BACTERIAL GROWTH DURING PROTEIN PRODUCTION FOR THE SYNTHETIC POLYPEPTIDE

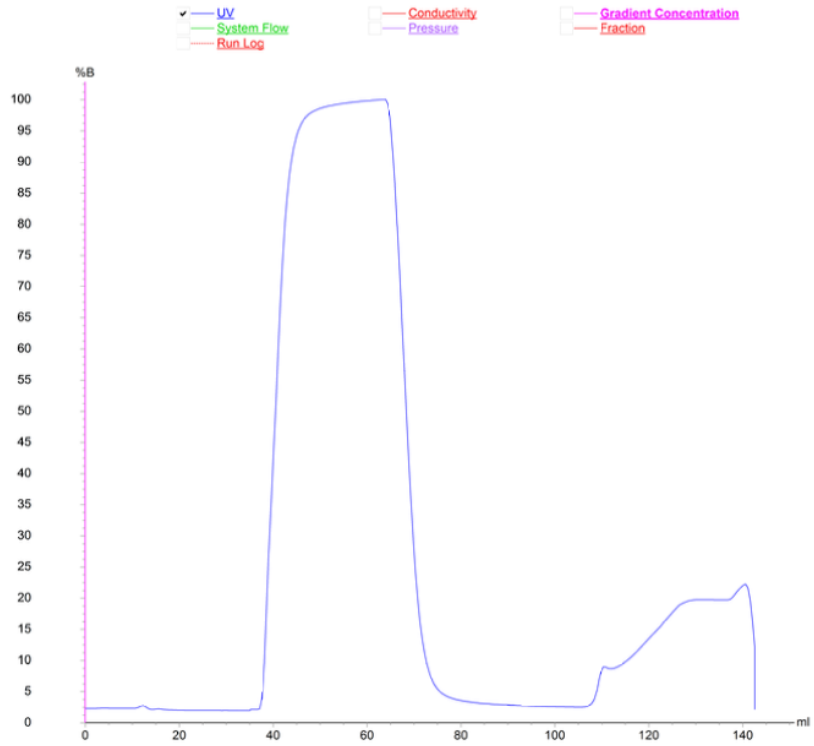


FIGURE 12 ABSORBANCE FROM SPECTROPHOTOMETER IN IMAC FOR THE PURIFICATION OF THE SYNTHETIC POLYPEPTIDE

4.3 Enzyme hydrolysis reactions

The first part of this section consists of the investigation of possible hydrolysis of β-casein with the recombinant TEV protease using various assays and computational methods. The second part is dedicated to analyzing the hydrolysis of the synthetic polypeptide using the same assays (except for the colorimetric assay) with the same recombinant protease.

4.3.1 Hydrolysis of β-casein

4.3.1.1 Docking of ENLHLPLPLLQS to TEV protease

The figure below (see **Figure 13**) shows a theoretical approach to the hydrolysis reaction between TEV protease and the target ligand ENLHLPLPLLQS obtained by docking this amino acid sequence as a ligand to the TEV protease, acting as a receptor. The docking was made using UCSF Chimera program. This interaction gives reason to further investigate if it means that TEV protease hydrolyses the β-casein at this location potentially giving the peptide products hypothesized earlier.

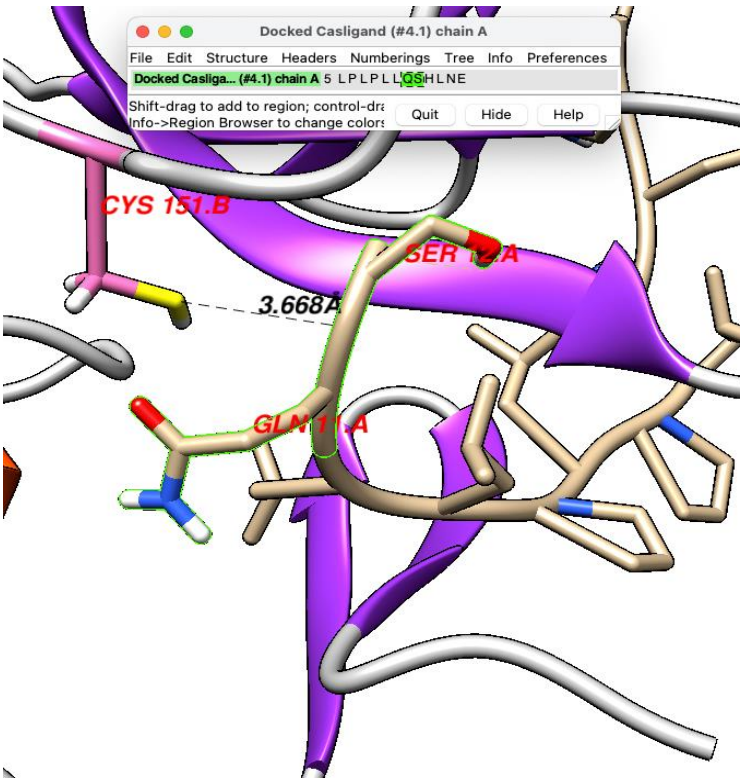


FIGURE 13 THE POSSIBLE NUCLEOPHILIC INTERACTION BETWEEN THE THIOL PART OF THE CYSTEINE IN TEV PROTEASE WITH A SIMILAR RECOGNITION SITE ON THE TARGET POLYPEPTIDE.

4.3.1.2 Colorimetric assay with Folin-Ciocalteu's reagent

The results for the colorimetric assay with Folin-Ciocalteu´s as a reagent was the assay used after successful docking. This assay measured the protease activity and can be observed in figures below (see **Figure 14**, **15** and **16**) where both color and absorbance can be seen. Every sample, including blank and the tyrosine standards with different concentrations, were made in triplicates and as can be seen below (in the **Figure 15** and **16**) the reaction mixture containing TEV protease and β-casein showed higher absorbance values than the first tyrosine concentration standard (5 mg/mL) also indicated by the slight blue color in the sample (row G and H in **Figure 14**).

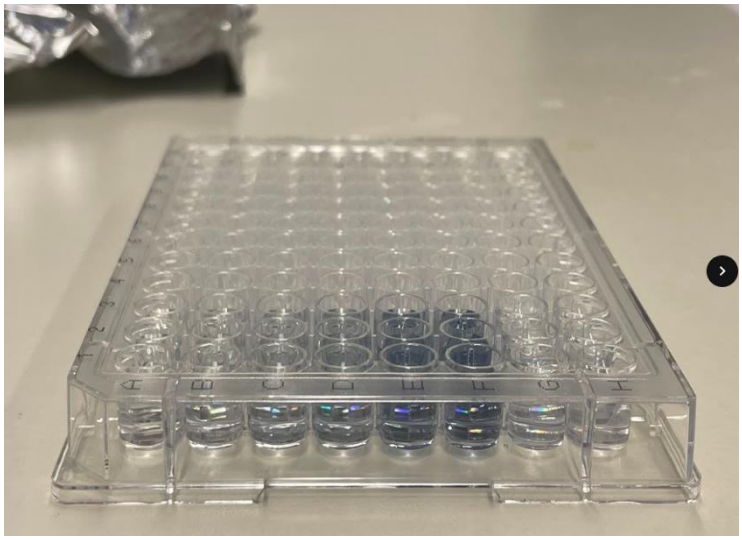


FIGURE 14 BLUE COLOUR OBTAINED IN THE REDUCTION REACTION WITH FC REAGENT. ROW A IS THE BLANK WHILE B-F IS TYROSINE STANDARD IN INCREASING CONCENTRATIONS. A SLIGHT BLUE COLOUR, INDICATING THE RELEASE OF TYROSINE, CAN BE OBSERVED ON ROW G AND H (TEV PROTEASE MIXED WITH B-CASEIN)

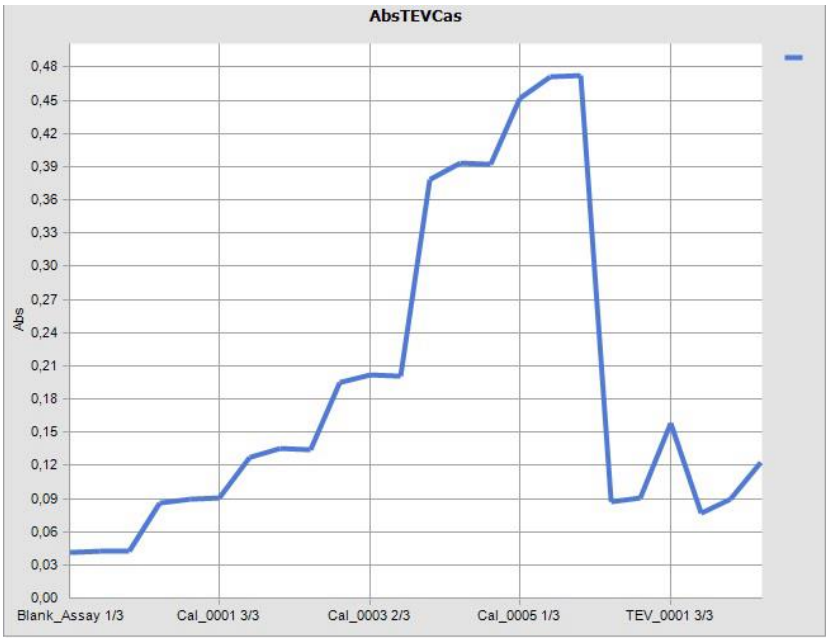


FIGURE 15 THE ABSORBANCE OF THE BLANK, TYROSINE STANDARDS AND THE HYDROLYSIS REACTION

	1	2	3
A	Blank_Assay blank 0,041	Blank_Assay blank 0,043	Blank_Assay blank 0,042
B	Cal_0001 1/3 5 calibrator 0,086	Cal_0001 2/3 5 calibrator 0,089	Cal_0001 3/3 5 calibrator 0,091
C	Cal_0002 1/3 10 calibrator 0,127	Cal_0002 2/3 10 calibrator 0,135	Cal_0002 3/3 10 calibrator 0,135
D	Cal_0003 1/3 20 calibrator 0,195	Cal_0003 2/3 20 calibrator 0,201	Cal_0003 3/3 20 calibrator 0,201
E	Cal_0004 1/3 40 calibrator 0,378	Cal_0004 2/3 40 calibrator 0,394	Cal_0004 3/3 40 calibrator 0,392
F	Cal_0005 1/3 50 calibrator 0,451	Cal_0005 2/3 50 calibrator 0,472	Cal_0005 3/3 50 calibrator 0,472
G	TEV_0001 1/3 1:1 unknown 0,087	TEV_0001 2/3 1:1 unknown 0,091	TEV_0001 3/3 1:1 unknown 0,159
H	TEV_0002 1/3 1:1 unknown 0,077	TEV_0002 2/3 1:1 unknown 0,089	TEV_0002 3/3 1:1 unknown 0,122

FIGURE 16 ABSORBANCE FOR BLANK, STANDARDS AND DIGESTION OF B-CASEIN.

4.3.1.3 SDS-PAGE

After the colorimetric assay an SDS-PAGE was conducted where the β-casein was analysed. The different samples were β-casein digested with TEV protease at 30°C for 1 hour, at 4 °C overnight and a negative control with no TEV protease digestion (see **Figure 17**) to be able to spot a difference between the samples.

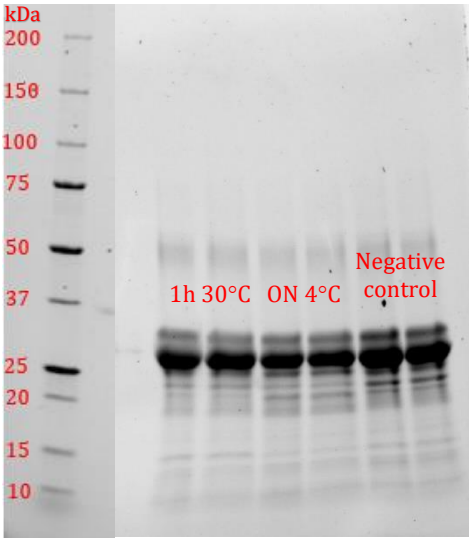


FIGURE 17 SDS-PAGE SHOWING THE DIFFERENT BANDS FOR DIGESTION OF B-CASEIN (AT 30 °C FOR 1H AND 4 °C OVERNIGHT) AND A NEGATIVE CONTROL WITH NO DIGESTION. ALL SAMPLES WERE MADE IN DUPLICATES

4.3.1.4 Mass Spectrometry

To get a more detailed view of the obtained SDS-PAGE results a mass spectrometer was used to determine the size of the fragments obtained after hydrolysis with the TEV protease and a negative control with no hydrolysation (see **Figures 18** and **19** below). The obtained fragments were then further analysed using a Mascot Search (MSMS) to identify the most prominent fragments obtained based on the amino acid sequences [22]. After possible digestion with TEV protease, the two expected polypeptides obtained should have a weight of approximately 17.5 kDa and 7.65 kDa respectively

MALDI spectra:

Casein 10 to 30kDa

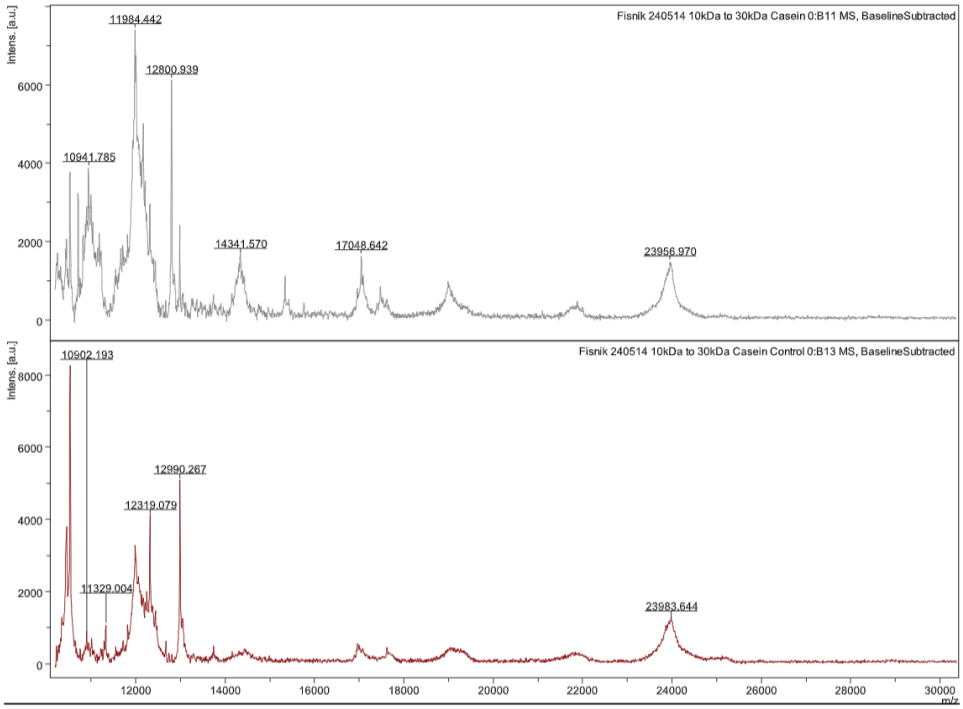


FIGURE 18 MASS SPECTRUM SHOWING PEAKS OBTAINED FOR BOTH DIGESTED (ABOVE) AND UNDIGESTED (BELOW) B-CASEIN AT 10-30 KDA. A MORE PROMINENT PEAK CAN BE OBSERVED AT 17048.642 DA IN DIGESTED SAMPLE AND THE SAME PEAK IS NOT AS PROMINENT ON THE CONTROL

Zoom around 7-8kDa:

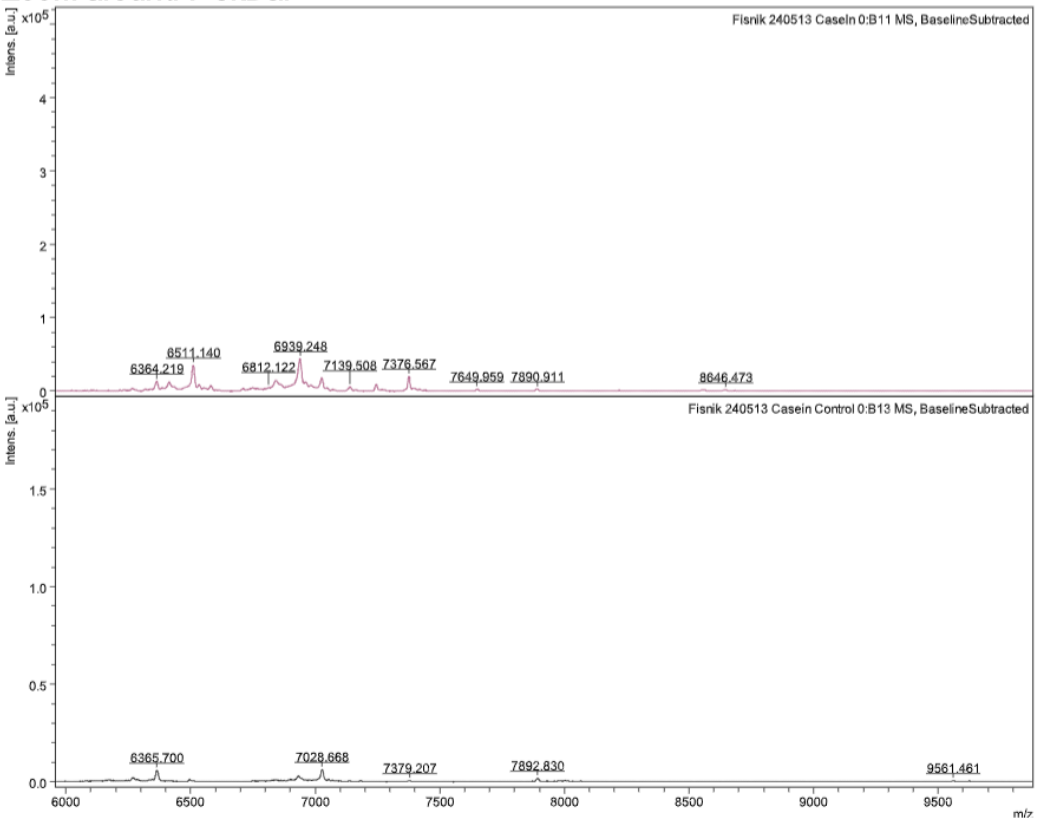


FIGURE 19 A ZOOMED-IN FIGURE AT 7-8 KDA FOR B-CASEIN (ABOVE) AND ITS NEGATIVE UNDIGESTED CONTROL (BELOW). THE PEAK AT 7649.959 DA IS NEW AND CAN ONLY BE SEEN IN DIGESTED SAMPLE

Mascot Search results

As seen before, one of the expected amino acid sequences for β-casein after hydrolysis with TEV protease was:

SWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV

The most prominent fragments obtained during a mascot search (MSMS) correspond to the following peptides (observed in **Figure 20**):

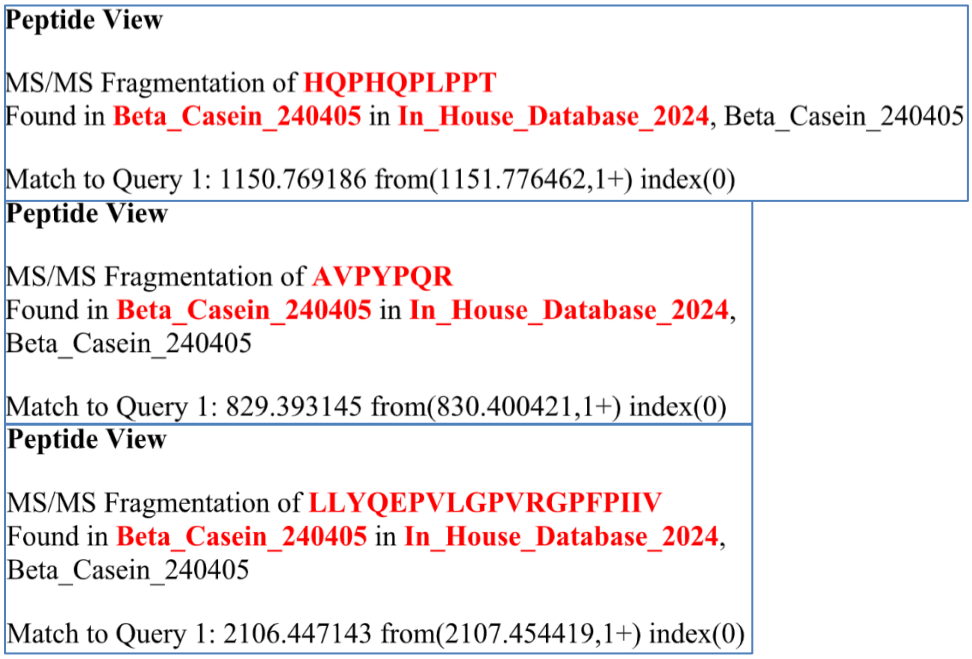


FIGURE 20 MASCOT SEARCH (MSMS) RESULTS OF THE MOST COMMON PEAKS IN THE MASS SPECTROMETRY RESULTS FOR B-CASEIN.

As can be seen in the mascot search results, these include the expected peptides from the bioinformatics studies. However, these were noticeable in both digested and undigested sample and do not prove a potential digestion at the hypothesized β-casein ligand. When investigating the properties of these peptides in the BIOPEP-UWM database the HQPHQPLPPT has a potential zinc binding activity. AVPYPQR has a potential ACE inhibiting activity [23] (antihypertension) while LLYQEPVLGPVRGPFPIIV has a potentially immunomodulating property.

4.3.2 Hydrolysis of the synthetic polypeptide

4.3.2.1 Docking of ENLYFQG to TEV protease

The image below in **Figure 21** shows a theoretical approach to the hydrolysis reaction between TEV protease and the target ligand ENLYFQG obtained by docking this amino acid sequence as a ligand to the TEV protease, acting as a receptor. The docking was made using UCSF Chimera program. This interaction gives reason to further analysis of the interaction between the recombinant TEV protease and the synthetic polypeptide to see if a hydrolysis reaction occurs.

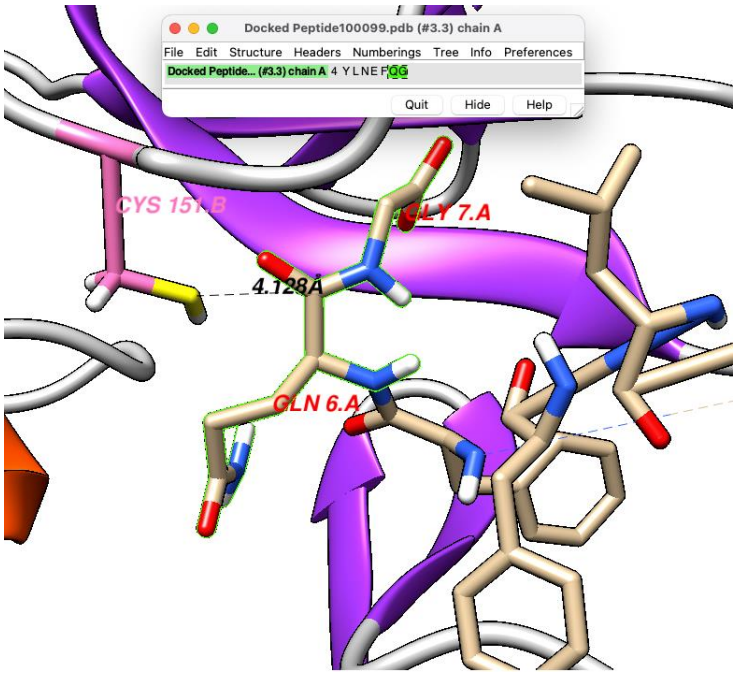


FIGURE 21 NUCLEOPHILIC INTERACTION BETWEEN THE THIOL PART OF THE CYSTEINE WITH THE CHARACTERISTIC RECOGNITION SITE FOR TEV ON THE LIGAND

4.3.2.2 SDS-PAGE

When performing the SDS-PAGE for the synthetic polypeptide a total SDS-PAGE (seen in **Figure 22**) including the TEV protease and the pellet for the synthetic polypeptide were included. The reason for including the pellet was due to the uncertainty if the target polypeptide would be fractionated in the IMAC or not.

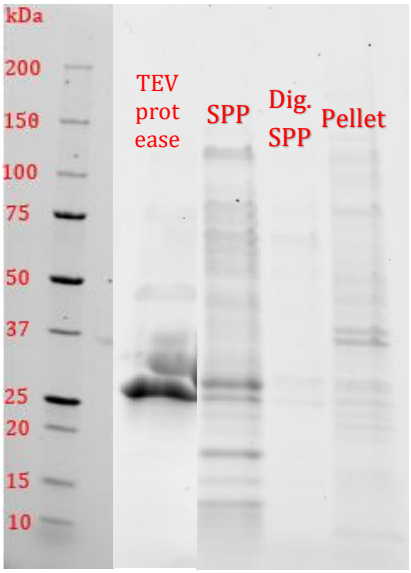


FIGURE 22 SDS-PAGE FOR RECOMBINANT TEV PROTEASE AND THE SYNTHETIC POLYPEPTIDE (UNDIGESTED, DIGESTED AND PELLET OBTAINED FROM THE PRODUCTION PHASE)

4.3.2.3 Mass Spectrometry

After the SDS-PAGE results, a MALDI-TOF mass spectrometer was used to determine the fragments obtained of the hydrolysed synthetic polypeptide. The results can be observed in the figure below (see **Figure 23**)

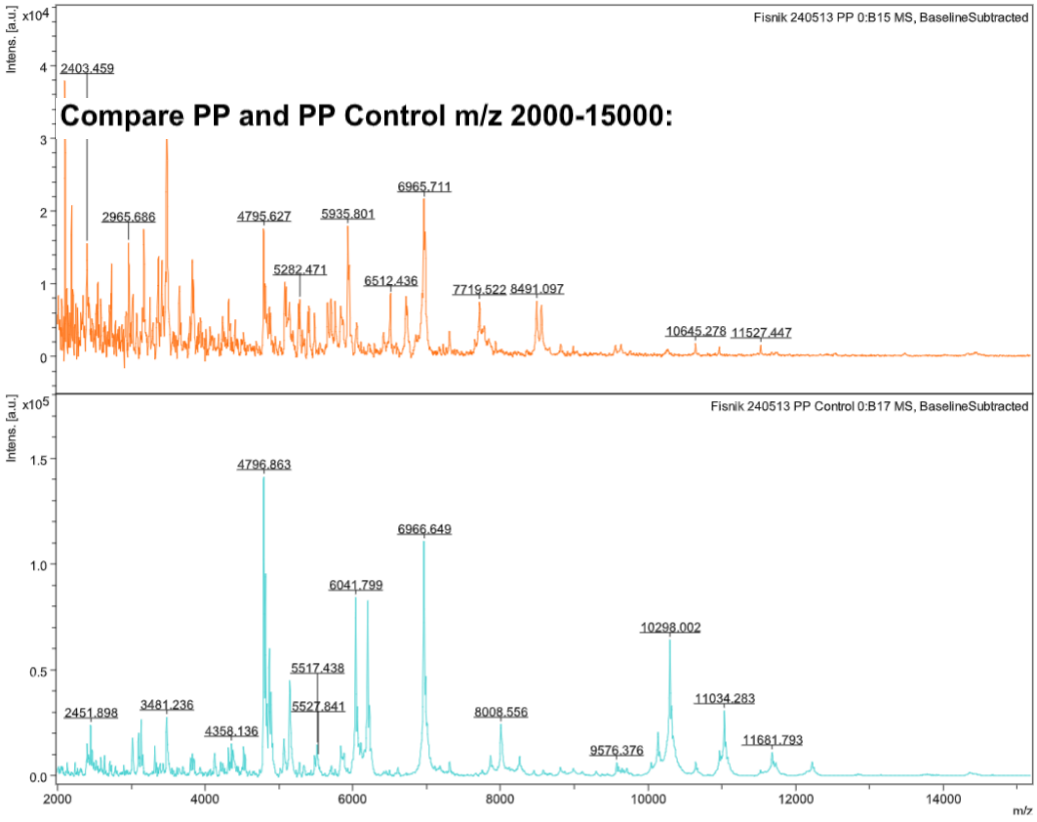


FIGURE 23 MASS SPECTRUM FOR THE HYDROLYSED SYNTHETIC POLYPEPTIDE AND ITS NEGATIVE UNHYDROLYZED FORM

When searching for potential bioactive properties for the 8-time repeated sequence GQHPHGLGALCAAPPSTENLYFQ in the BIOPEP-UWM web site, a match showing ACE inhibiting (antihypertension) and DPP-4 inhibiting (antidiabetic) activity was obtained for majority of the 2-4 amino acid long peptides within the cleaved sequence above [24].

5. Discussion

The Tobacco Etch Virus protease is an enzyme that has the capabilities to remove certain affinity tags on proteins that are necessary for purification processes but can have protein activity inhibiting effects. This thesis was focused on demonstrating the cleaving mechanisms of a recombinant form of TEV protease and evaluating possible cleaving of another protein than the target synthetic polypeptide, β-casein in this case. In order to do so, the necessary proteins had to first be produced successfully by using *E. coli* (BL21 (DE3)) as a host organism for the recombinant plasmids (pET21b) containing the necessary genes. Looking at the results (figures, graphs and tables under section **4.1** and **4.2**) for the cultivation, production and purification of these bacteria and proteins, one can see that a cultivation and production of the recombinant plasmids was successfully achieved. When analysing the hydrolysis reactions for both β-casein and the synthetic polypeptide several assays were performed to give a more reassuring result and conclusion of the mechanism of the protease in this case. Since the β-casein analysed in these experiments has a similar amino acid sequence as the classic TEV protease recognition sequence (ENLHLPLPLLQS compared to ENLYFQG), a possible hydrolysis of this substrate was investigated (results from section **4.3.1.1**) which gave a docking of the target ligand. This discovery led to performing a colorimetric assay to confirm an enzymatic reaction (see results from section **4.3.1.2**) and the results from this assay were also confirming a reaction between TEV protease and the β-casein, which gave reason for further analysis. The docking for the synthetic polypeptide was also a success, which was more expected (see section **4.3.2.1**).

The two SDS-PAGE experiments showed satisfying results. The first one (see figure **17**) showed no difference in bands between the different temperatures and periods of time at which the hydrolysis of the β-casein was done. This result gave an indication about TEV protease not having a specific temperature for optimal reaction. However, a very small difference could be observed on the undigested negative control where a band at roughly 20 kDa can be observed suggesting an unhydrolyzed polypeptide. When observing the results for the SDS-PAGE containing TEV protease and the synthetic polypeptide (section **4.3.2.2**) the TEV protease can be seen as a single band at roughly 27 kDa, which is to be expected indicating that the fusion protein MBP had been detached during the IMAC purification process. The synthetic polypeptide shows three bands; one digested, one undigested and a weight band for the pellet since there was a suspicion that the target polypeptide would appear there. However, this was not the case since there are not as many clear bands shown in that sample. The unhydrolyzed SPP however shows slightly darker bands, but these are spread out almost evenly across the gel due to a possible polymerization of the repeated sequences. The digested SPP is shown next to the undigested one with slightly faded bands indicating a possible digestion has occurred.

When analysing the mass spectrometer results for β-casein (see section **4.3.1.4**), new peaks at 17048.642 Da and 7649.959 Da were observed in digested sample compared to undigested control. These peaks could correspond to the masses of the obtained polypeptide after a TEV protease digestion (17048.642 Da corresponding to the polypeptide from the N-terminal of the β-casein to the Q (Glutamine in ENLHLPLPLLQS) whereas 7649.959 Da corresponds to the peptide from S (Serine in ENLHLPLPLLQS) and downstream to the C-terminal). As can be seen in the MSMS results for the expected amino acid sequences, that led the purpose for this investigation, are shown in the peaks that are most prominent on the MS results. However, there is not a significant fragmentation difference here and these appear in both samples and therefore does not give any substantial information about difference between both the digested and undigested β-casein. Since the samples were only subjected to the TEV protease enzyme, one can assume that these cleaved products are due to the enzyme if not other coincidental factors like impurities in the sample or adduct formation with potassium since potassium phosphate was used as a buffer in the sample [25]. If not coincidental however, this enzyme activity may contradict the previously stated stringent sequence specificity of TEV protease. Even if a very small intensity change is observed in the digested sample, can this new discovery debunk the precise protein processing functionality that TEV protease is known to have. This could lead to unwanted hydrolysis of polypeptides containing similar amino acid sequences as the recognition site for this protease. However, a more substantial analysis would need to be conducted to make this claim.

When analysing the hydrolysis of the synthetic polypeptide there is a clear difference between the hydrolysed sample and its negative control (see section **4.3.2.3).** Even smaller peaks are consistently different between the two samples. The unhydrolyzed polypeptide control has peaks at higher m/z values indicating heavier polypeptides while the digested sample has more low weight peptides. Furthermore, a clear peak at 2403 (see **Figure 23**) can be observed which is similar to the expected m/z ratio for one of the 8-time repeated amino acid sequences for the synthetic polypeptide (mW 2406). Another evidence of a successful digestion is the equally consistent distribution of peaks after the first peak which is to be expected since the synthetic polypeptide is a protein with repeating sequences. The reason to why the undigested sample has fragments could be due to impurities, but my guess is that, since the synthetic polypeptide has repeating sequences, these have formed oligomers/polymers after being ionized in the MALDI-TOF mass spectrometer, which may preserve non-covalent interactions.

6. Conclusion

In conclusion, the results obtained from this project were successful overall. The theoretical approach using docking of substrates to the receptor enzyme indicated a reaction between TEV protease and these substrates. This led to the several assays conducted in this project where the colorimetric assay further proved the possible hydrolysis of β-casein by TEV protease. The SDS-PAGE and mass spectrometry was later used to establish a more reassuring result showing obtained bands and fragments corresponding well to the hypothesized peptides after digestion with TEV protease. After each analysis step, whether it be theoretical or experimental, the

hydrolysis of β-casein was confirmed at the specific sequence that was similar to the TEV protease recognition site. This would imply that this enzyme can be more unspecific than previously thought. This discovery could potentially mean a risk of hydrolysing other proteins as well in the process when using TEV protease as a tool for removing affinity tags. Worth noting is that β-casein has a similar amino acid sequence (as the TEV protease recognition site) once in the entire polypeptide chain. Further unspecific digestion could therefore occur when purifying other proteins that have more similarities as this enzyme recognition site.

However, to make such a claim more extensive research would need to be conducted to reassure the hydrolysis of the polypeptide substrates investigated in this project. Performing even more assays would give a more reassuring answer to the hydrolyzation reactions. Another approach would have been to study the enzyme-inhibiting effects these peptide products formed after digestion potentially have. Both β-casein and the synthetic polypeptide showed properties of multi-inhibiting effects which can lead to formulation of multifunctional therapeutic agents in the future.

7. References

Bibliography

[1] R. A. Copeland, "The Transition State of Chemical Reactions," in *Enzymes: A practical introduction to structure, mechanism, and data analysis*, Wiley, 2023, pp. 24-27.

[2] G. T. T. A. I. Lye Yee Chew, "Application of Proteases for the Production of Bioactive Peptides," in *Enzymes in Food Biotechnology: Production, Applications, and Future Prospects*, Selangor Darul Ehsan, Academic Press, 2019, pp. 247-261 https://doi.org/10.1016/C2016-0-04555-2.

[3] "UniProt," 2024. [Online]. Available: https://www.uniprot.org/uniprotkb/P02666/entry#sequences. [Accessed 31 05 2024].

[4] I. A. D. M. Minkiewicz P., "BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities," *International Journal of Molecular Sciences,* pp. doi:10,3390/ijms20235978, 24 05 2019.

[5] F. B. H. &. V. M. Niesen, "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability," *nature protocols,* p. https://doi.org/10.1038/nprot.2007.321, 2007.

[6] D. S. Waugh, "An overview of enzymatic reagents for the removal of affinity tags," *Protein expression and purification,* pp. 283-293, 10.1016/j.pep.2011.08.005, 2011.

[7] A. B. M. H. W. D. D. C. Ashley Steere, "Proteopedia," 25 04 2022. [Online]. Available: https://proteopedia.org/wiki/index.php/Tobacco_Etch_Virus_%28TEV%29_Protease#Enzymatic_Mechanism. [Accessed 31 05 2024].

[8] C. D. Desjardins P, "NanoDrop Microvolume Quantitation of Nucleic Acids," *Journal of visualized experiments,* 2010.

[9] C. J. H. C. K. Y. Lee PY, "Agarose Gel Electrophoresis for the Separation of DNA Fragments," *Journal of visualized experiments,* 2012.

[10] E. M. J. M. Wayne Davis, "ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program," *Front Bioinform,* p. 10.3389/fbinf.2022.818619, 2022.

[11] "Thermo Fisher scientific," Thermo Fisher, 2006-2024. [Online]. Available: https://www.thermofisher.com/se/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/molecular-cloning/transformation/bacterial-transformation-workflow.html#preparation. [Accessed 08 03 2024].

[12] S. G. C. e. al., "Protein production and purification," *Nature methods,* vol. 5.2, 2008.

[13] J. A. a. J. J. F. Bornhorst, "Purification of proteins using polyhistidine affinity tags," *Methods in enzymology,* vol. 326, pp. doi:10.1016/s0076-6879(00)26058-8, 2000.

[14] M. e. a. Pérez, "The Chemistry Behind the Folin-Ciocalteu Method for the Estimation of (Poly)phenol Content in Food: Total Phenolic Intake in a Mediterranean Dietary Pattern," *Journal of agricultural and food chemistry,* Vols. 71,46, p. doi:10.1021/acs.jafc.3c04022, 2023.

[15] A. A. Al-Tubuly, "SDS-PAGE and Western Blotting," in *Diagnostic and Therapeutic Antibodies*, Humana Totowa, NJ, 2000, pp. 391-405, https://doi.org/10.1385/1592590764.

[16] M. M. Ruedi Aebersold, "Mass spectrometry-based proteomics," *nature,* no. 422, pp. 198-207, 2003.

[17] Y. B. M. S. M. R. Sania Safdar Butt, Molecular Docking Using Chimera and Autodock Vina Software for Nonbioinformaticians, vol. 1, JMIR Bioinform Biotech, 2020.

[18] D. H. e. al., "Integrated Interface for Protein Secondary Structure Detection and Evaluation in UCSF Chimera," in *Bioinformatics Research and Applications*, Beijing, Springer, p. 234.

[19] cytiva, "cytivalifesciences," 2024. [Online]. Available: https://www.cytivalifesciences.com/en/us/support/products/histrap-ff-17531901.

[20] M. M. A. S. S. Katyare, "An Improved Micromethod for Tyrosine Estimation," *Zeitschrift für Naturforschung C,* pp. 897-900, 2004.

[21] T. Scientific, "Protein Concentrators," 2024. [Online].

[22] T. e. a. Köcher, "Altered Mascot search results by changing the m/z range of MS/MS spectra: analysis and potential applications.," *Analytical and bioanalytical chemistry,* Vols. 400,8, pp. doi:10.1007/s00216-010-4572-0, 2011.

[23] P. S. A. I. e. a. Herman LL, "Angiotensin-Converting Enzyme Inhibitors (ACEI)," in *StatPearls*, StatPearls Publishing, 2023.

[24] B. K. Kasina SVSK, "Dipeptidyl Peptidase IV (DPP IV) Inhibitors," in *StatPearls*, StatPearls Publishing, 2023.

[25] A. K. K. Kruve, "Adduct Formation in ESI/MS by Mobile Phase Additives," *Journal of The American Society for Mass Spectrometry,* vol. 28, pp. 887-894, 2017.

Figures

TEV protease enzymatic mechanism:

https://proteopedia.org/wiki/index.php/Image:TEVMechanism.jpg

(Bacterial Transformation image)

https://microbenotes.com/bacterial-transformation/

Appendix
Appendix 1

TEV protease enzymatic mechanism:

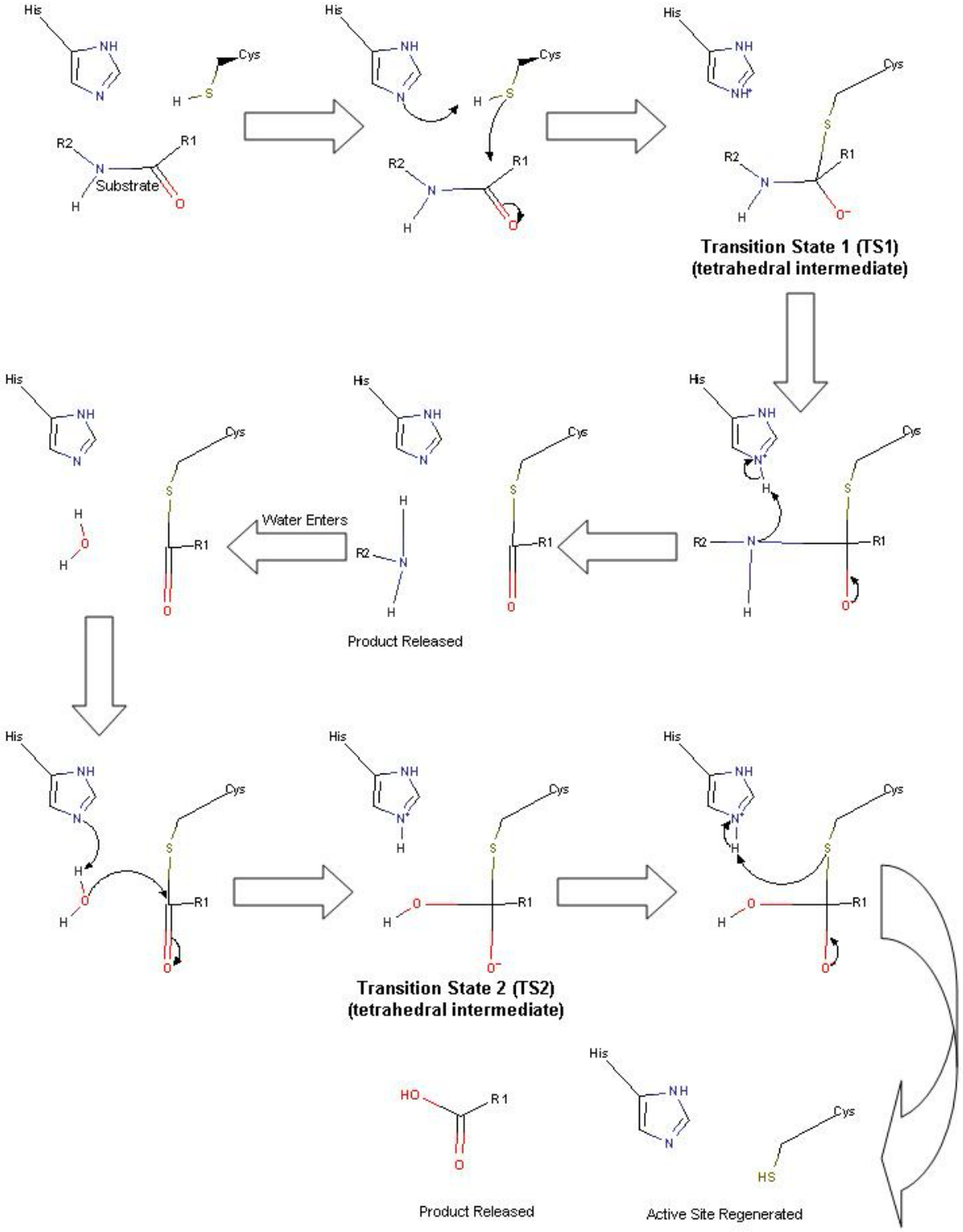


FIGURE 24 TEV PROTEASE ENZYMATIC MECHANISM

Appendix 2

TABLE 4 OD MEASUREMENTS (USING 3 ML INOCULATION + 30 ML LB)

E. coli strain / Time	BL21(DE3)	DH5α
$T_{0, (08:32)}$	0.233	0.241
$T_{1, (09:11)}$	0.265	0.211
$T_{2, (09:43)}$	0.334	0.301

Appendix 3

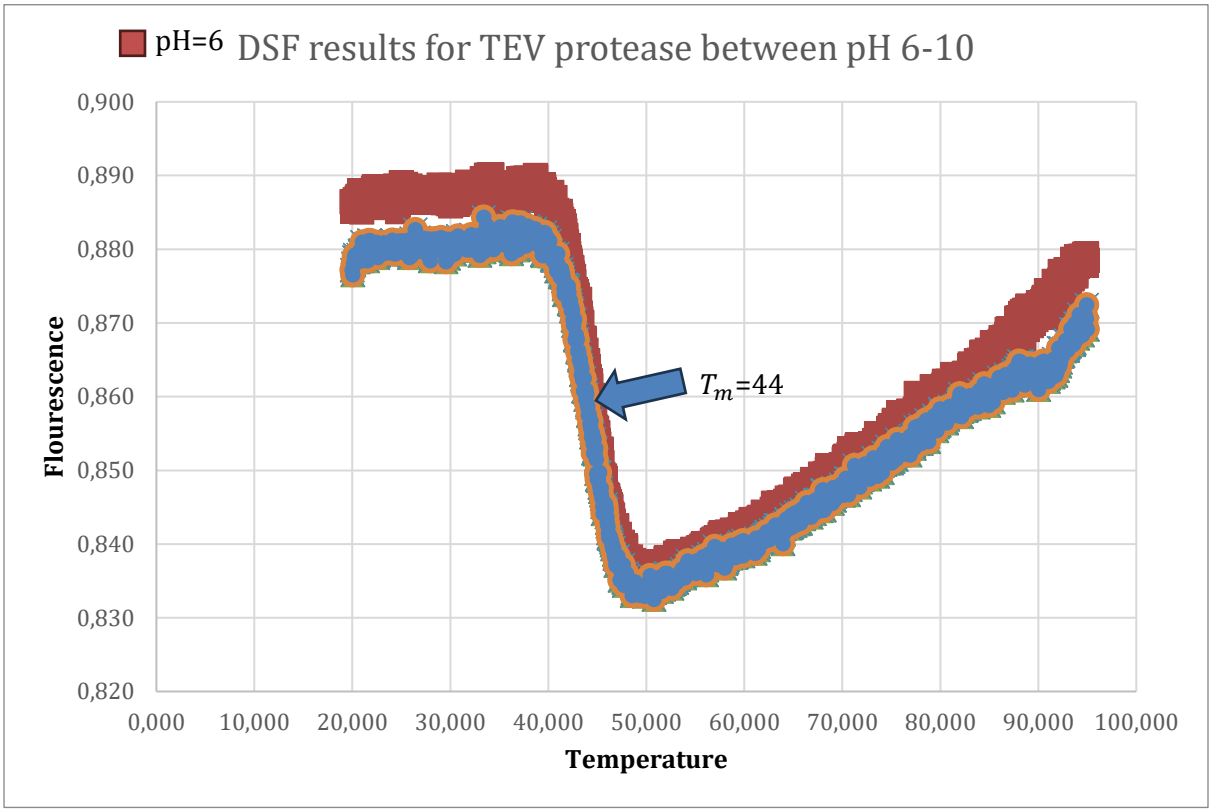


FIGURE 25 DSF RESULTS FOR TEV PROTEASE BETWEEN PH 6-10