Methods for production of recombinant enzymes for collagen degradation

Master thesis

Tea Eriksson & Elin Su

LUND UNIVERSITY

Faculty of Engineering, LTH

Department of Chemistry

Division of Pure & Applied Biochemistry

Spring 2020

Course:	KBKM05 Degree Project in Applied Biochemistry for Engineers
Extent:	30 Credits
Examiner:	Lei Ye
Supervisor:	Johan Svensson Bonde
Assistant supervisor:	Mohamad Takwa
Project period:	2020.01.20 - 2020.06.08

Populärvetenskaplig sammanfattning

Metoder för produktion av enzymer för nedbrytning av kollagen

Kollagenrika restprodukter kan användas i värdefulla livsmedels- och läkemedelsprodukter. Enzymatisk behandling tillåter kontrollerad och effektiv nedbrytning. I det här projektet har metoder utvecklats och testats för produktion av enzymer i bakterier.

Kollagen är det vanligaste proteinet i däggdjur. Proteinet finns i ben, vävnader och senor och ger mekanisk stabilitet och styrka. Den kompakta strukturen av kollagen gör molekylen svår att bryta ner. Det finns stora mängder av många olika restprodukter med höga kollagenhalter som har lågt värde och det finns därför intresse att omvandla kollagenet i de här materialen till värdefulla produkter. Detta skulle minska avfallshanteringen och därmed minska klimatpåverkan. Nedbrutet kollagen kan användas till en mängd olika tillämpningar som i kosttillskott, antioxidanter, hudvårdsprodukter och mervärdesmat. Marknaden drivs främst av stillahavsområdet i Asien där efterfrågan på näringsrik mat och dryck är hög.

Åtta enzymer har valts ut och studerats. Fyra av dessa är vanligt förekommande enzymer som bryter bindningar i proteiner. De resterande fyra enzymer är kollagenaser som kan specifikt bryta ner kollagen. Generna för dessa enzymer har optimerats för insättning i cirkulära DNA-molekyler (plasmider) och tillverkning i olika celler. I det här projektet har två strategier utvecklats för produktion av enzymer för nedbrytning av kollagen eller kollagenderiverade produkter. Strategierna utvecklades för produktion av enzymer i bakterien, *Escherichia coli* och jästen, *Pichia pastoris*. *E. coli* växer otroligt fort och kan fördubblas på tjugo minuter men saknar förmågan att modifiera proteiner vilket jäst kan åstadkomma. I jästen är systemet utvecklat för enzymproduktion utanför organismen. Bakteriesystemen är utvecklade för tillverkning av enzymerna inuti cellen, mellan inre och yttre cellmembranen samt utanför cellen. Fördelen med att producera enzymer inuti celler är att produktionen kan bli mycket hög. Nackdelarna är att det kan vara svårt att ta ut enzymerna för att det finns många andra proteiner i celler och det krävs också att man tar sönder cellerna.

Enzymer kan kombineras på många olika sätt för olika grad av nedbrytning av kollagen. Metoder för att analysera enzymerna och dess produkter har beskrivits. En försöksplanering designades för att med så få försök som möjligt upptäcka den mest effektiva kombinationen. För att få reda på vilket eller vilka enzymer som har bäst nedbrytningskapacitet behöver försöksplaneringen testas experimentellt.

Två enzymer valdes ut för experimentell framställning i *E. coli* mellan cellmembranen. Först måste de optimerade generna erhållas. Därefter behöver generna klippas ut och sättas in i plasmider som innehåller delarna som behövs för uttryck av protein. Plasmider med insatta gener kan tas upp av celler för att producera enzymerna. Selektering med antibiotika används för att välja ut de celler som bär på plasmiden med genen för enzymet. Selekteringen visade att experimenten inte hade lyckats. Experimenten fick avbrytas vid det här steget men med mer tid hade framställningen haft stor potential att lyckas.

Abstract

In this thesis eight enzymes have been identified with documented ability to degrade collagen or its hydrolyzed products. Collagen can be found in skin, bones, and connective tissue. Hydrolyzed collagen peptides are utilized in food, disease treatment, skincare and training supplements. The studied enzymes are: *Clostridium histolyticum* collagenases ColG and ColH, *Grimontia hollisae* collagenase, *Bacillus cereus* collagenase ColA, subtilisin, trypsin, pepsin and papain. Two organisms were used to develop cloning strategies, *Escherichia coli* and *Pichia pastoris*. For *E. coli* the enzymes have been optimized for cytoplasmic expression with pET22b+ expression vector, for periplasmic expression with pET22b+ in frame with pelB signal peptide and for extracellular expression with pAES40 vector with YebF protein. Extracellular expression of the enzymes with *P. pastoris* was optimized for pPICZα expression vector with α-factor signal sequence.

Two cloning strategies have been developed for the four expression systems which included subcloning, protein production and purification. Furthermore, possible methods for evaluating the produced enzymes were investigated. Additionally, a strategy to determine the most effective enzyme combination for breakdown of collagen was developed which resulted in a maximum of 16 tests. The optimized genes were compared between the species and a difference in codon bias was found as well as between different algorithms which showed that different tools gave diverse sequences. The cloning strategy developed for *E. coli* was experimentally tested with G. *hollisae* collagenase and subtilisin Carlsberg. The cloning was not successful as no growth of transformed cells were observed. It was discovered that the size of vector backbone and genes of interest caused difficulties for purification of genes, but it was postulated that with additional restriction enzymes or PCR and different competent cells the enzymes could have been successfully expressed.

Acknowledgement

First, we would like to thank our supervisor Johan Svensson Bonde at Lund University, department of pure and applied biochemistry. Johan has aided the work with helpful suggestions and guidance. Furthermore, he helped us to come up with a plan to continue this thesis after the practical work was stopped.

Secondly, we want to thank Ecozyme AB, Lund and Mohamad Takwa for giving us the opportunity to write and work on the thesis project giving us valuable experiences in cloning and project development. Additionally, we are thankful to be able to use the experimental results in our thesis even though the project was terminated early. We would also like to thank Rami Eskeif for all the help and support he provided during the practical work of the cloning experiments.

Lastly, we would like to thank each other for always working as a team and helping each other to solve problems and provide helpful tips and comment about each other's contributions.

Table of contents

1. Introduction

1.1. Aim

The aim of this thesis was to design experiments for production of recombinant enzymes capable of degrading native or degraded collagen and test a strategy with two enzymes experimentally. The idea was that our work could be used to convert substrate (collagen) into profitable products with many applications. Furthermore, the report aimed to investigate the actions of the enzymes and develop assays to compare their ability to degrade the substrate most efficiently and explore the advantages of using different expression systems.

1.2. Disposition of report

The thesis starts with an introduction including a section about collagen. Methods describe the strategies used for conducting literature studies, choosing and optimizing genes and experimental procedures. The experimental methods and materials are placed in method section, 2.3.1 and the experimental results and discussion under the result section, 3.6.1. As the enzymes, expression systems, cloning strategies and analysis of enzymes analyzed in this work were results of extensive literature studies these are placed in results. Then follows the discussion which discusses the results except for the practical work that has been discussed in the section above. In Appendix A, all DNA data for the theoretical cloning plan are found. Appendix D has the DNA used for the practical work. Appendix B and C contain detailed cloning strategies.

1.3. Division of work

The general strategy of this work was to divide it equally, but each student wrote or did the parts that felt most needed at that phase of the work. The amount of text written by each person does not reflect the amount of time spent on this thesis. The two enzymes used in the experimental work were decided together and Tea oversaw the optimization of the collagenase gene and Elin for the subtilisin gene. Both authors carried out laboratory work, but Elin conducted most. Tea

developed and performed the optimization of the genes, conducted the literature study about the four collagenases, wrote about the pET22b+ expression vector, yeast strain, cloning plan, the four first sections in “analysis of enzymes” and the results of experimental work. Elin conducted the literature study on the four remaining enzymes, developed the cloning assay for yeast, wrote about the pAES40 vector, the last two sections in “analysis of enzymes” and was additionally in charge of the referencing and managing throughout the work. The discussion was written together as were the conclusion, abstract and popular scientific summary.

1.4. Collagen

One of the most abundant proteins in mammals is collagen (Fratzl, 2008). There are 28 members of the collagen superfamily which all share the specific triple helical structure (Ricard-Blum, 2011). It is located in skin, tendon, bones and connective tissue and provides mechanical stability, toughness and strength. The protein has shown to be involved either indirectly or directly in cell differentiation and adhesion. Collagen has many applications such as in pharmaceuticals, artificial skin, glue, food, cosmetic surgery and cosmetics (Sagi & Afratis, 2019). It is common to use enzymes like pepsin, papain and trypsase to extract collagen protein (Yang & Shu, 2014). They act on the non-helix peptide chains of collagen and have usually no effect on helix peptide chains. The most common types of collagen in the body are type I, II and III which represents around 80-90% of all collagens (Lodish et al., 2000). Different types of collagen are presented in table 1.

Table 1. Structural features and representative tissues of different types of collagen (Lodish et al., 2000).

Type	Molecule Composition	Structural Features	Representative Tissues
Fibrillar Collagens			
I	$[\alpha1(I)]_2[\alpha2(I)]$	300-nm-long fibrils	Skin, tendon, bone, ligaments, dentin, interstitial tissues
II	$[\alpha1(II)]_3$	300-nm-long fibrils	Cartilage, vitreous humor
III	$[\alpha1(III)]_3$	300-nm-long fibrils; often with type I	Skin, muscle, blood vessels
V	$[\alpha1(V)]_3$	390-nm-long fibrils with globular N-terminal domain; often with type I	Similar to type I; also cell cultures, fetal tissues
Fibril-Associated Collagens			
VI	[α1(VI)][α2(VI)]	Lateral association with type I; periodic globular domains	Most interstitial tissues
IX	[α1(IX)][α2(IX)][α3(IX)]	Lateral association with type II; N-terminal globular domain; bound glycosaminoglycan	Cartilage, vitreous humor;
Sheet-Forming Collagens			
IV	$[\alpha1(IV)]_2[\alpha2(IV)]$	Two-dimensional network	All basal laminaes

The biosynthesis of the protein starts with synthesis of the procollagen α-chains on the ribosome which are then transported to the endoplasmic reticulum (Hulmes, 2008). There it undergoes post-translational modifications and assembly of procollagen molecules which includes for example hydroxylation of proline, O- and N-linked glycosylation, disulphide bonding, trimerization and folding of triple helix. The molecules are then imported to the Golgi apparatus where they are packaged into secretory vesicles before transport to the extracellular matrix. Procollagen processing ensues during or shortly after secretion followed by fibril assembly.

In 1994 the first high-resolution crystal structure of triple-helical collagen-related peptides was presented (Bella, Eaton, Brodsky & Berman, 1994). Collagen consists of three left-handed α-chains that are supercoiled along a common axis to form a supercoiled right-handed triple helix, see figure 1 (Engel & Bächinger, 2005). The monomeric building block for collagen fibres was theorized back in 1956 and called tropocollagen (Gross, 1956).

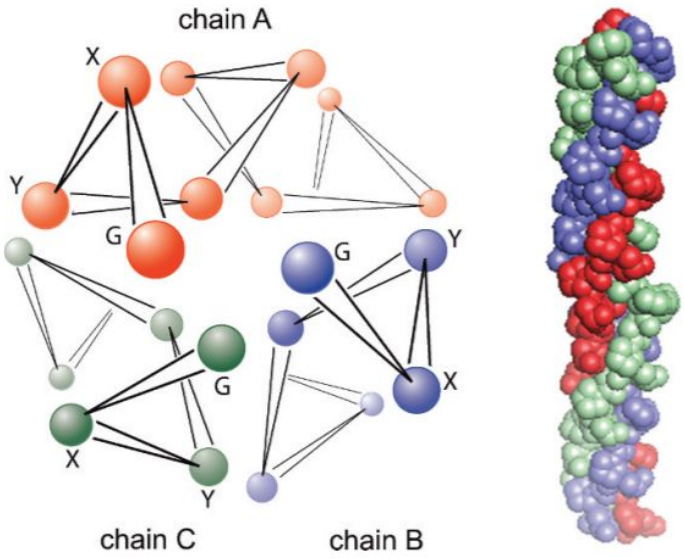

Figure 1. Triple-helical structure of collagen viewed along the molecular axis and from the side (Fratzl, 2008).

The α-chains can either be identical or heterotypic depending on the type of collagen (Hulmes, 2008). The main reappearing motif of the protein is the primary structure consisting of amino acids Glycine-X-Y, where X and Y are most commonly proline and hydroxyproline, respectively (Teramura et al., 2011). This motif is the driving force between the right-handed triple structure as hydrogen bonds are formed between the NH-group of glycine and backbone CO-group in the X-position of an adjacent helix (Engel & Bächinger, 2005). The stability of the triple helix is related to the percentage of proline and hydroxyproline as the cyclic sidechain imposes steric restrictions which affects the flexibility of the peptide bond. Additionally, propyl hydroxylation further stabilizes the structure by increasing the number of hydrogen bonds (Hulmes, 2008).

When the collagen molecules bind to each other they are displaced around 67 nm from each other and form up to a few hundred nanometers thick fibrils (Engel & Bächinger, 2005). Short parts of the collagenase molecules, located at the ends, do not form the triple helical structure. Between two hydroxylysine or lysine residues at the C-terminus, covalent aldol cross-links are formed with two analogous residues at the N-terminus of a neighbouring molecule which stabilizes the packaging of collagen and institutes stable fibrils.

Acidic or alkaline hydrolysis of collagen or alternatively heat treatment generates gelatin (Gómez-Guillén, Giménez, López-Caballero & Montero, 2011; Ikada & Tabata, 1998). The protein is polyampholyte having both anionic and cationic groups along with hydrophobic residues (Harhout & Metwally, 2019). Gelatin has many applications such as gelling, stabilising and thickening agent in the food industry and capsules (Poppe, 1992).

1.4.1. Collagen derived hydrolysates

Peptides derived from collagen and gelatin are most often obtained by enzyme hydrolysis by protein such as collagenases, alcalase, trypsin, pepsin and papain (Gómez-Guillén et al., 2011). With analytical methods, characterization and new applications for collagen and gelatin hydrolysates have been identified. Hydrolysed peptides have shown to be an effective dietary calcium source by accelerating the absorption of the metal (Jung et al., 2006). Moreover, gelatin hydrolysates have shown capability to act as an antihypertensive agent and have antioxidative properties (Byun & Kim, 2001; Mendis, Rajapakse & Kim, 2005). Additionally, collagen hydrolysate has shown potential productiveness in osteoarthritis and osteoporosis treatment as well as antimicrobial activity (Gómez-Guillén et al., 2010; Moskowitz, 2000).

Collagen hydrolysates from salmon and trout skins have shown to affect lipid absorption and metabolism is rats which may help suppress increase of plasma triglycerides which have documented effect on cardiovascular disease (Hokanson & Austin, 1996; Saito, Kiyose, Higuchi, Uchida & Suzuki, 2009). Fortified collagen hydrolysates are competent additives for advancement in skin moisture and elasticity as well as reducing wrinkles (Inoue, Sugihara & Wang, 2016). Furthermore, collagen hydrolysates ingestion with high levels of Hyp-Gly and Pro-Hyp have proven to aid healing of pressure ulcers as a supplement to traditional therapy (Sugihara, Inoue, & Venkateswarathirukumara, 2018). Moreover, seafood derived collagen peptides have shown promising applications as natural functional food ingredients and for production of functional beverages (Pal & Suresh, 2016; Volokitina, Ionova & Krasnova, 2020). Collagen peptides have been used as training supplements and have shown that in combination

with resistance exercise training (RET) increased body mass, fat-free mass and muscle strength more than RET alone (Oertzen-Hagemann, 2019).

The collagen hydrolysates market was valued over 700 million USD in 2017, is estimated to have a compound annual growth rate of 11.6% and reach over 1,300 million USD by 2023 (MarketsandMarkets, 2018). The market growth is highly attributed by rising demand for nutritional food and drinks especially in the Asia pacific region yet faces challenges in turns of strict food additive laws and an increasing opposition for using products obtained from animals. Collagen peptides are not only economical functional food items, the effective reuse of seafood by-products also minimizes waste treatment which can otherwise be an expensive and environment-impacting process (Pal & Suresh, 2016).

2. Methods

2.1. Choosing enzymes and expression systems

The proteins analyzed in this work were chosen by an extensive literature study. The study was conducted both at the start of the project and in the middle. The two main search drives for the literature work were *Lubsearch* (search engine provided by Lund University) and *Google scholar* and the search were based on published books and sources. References in relevant articles were often used to extend the study and to obtain the original source. The enzymes described in this work were chosen for either possessing the ability to degrade collagen, gelatin or collagenous material. The specific enzymes were chosen not only from literature but also for their availability in UniProtKB and NCBI Nucleotide databases.

The different expression systems used in this thesis were also chosen from literature studies. The host organisms were chosen based on the amount of research available, their effectiveness to produce recombinant proteins and commercially available expression vectors. The vectors were chosen depending on the possible final location of the protein production with either signal peptides or fusion proteins and affinity tags to aid purification.

2.2. Strategy and tools for choosing and analyzing genes for enzymes

After an enzyme of interest had been decided the gene needed to be obtained. To ensure that the searches were of high quality, the annotated database UniProtKB was used. The UniprotKB record contains information about the functionality of the enzyme, subcellular location, structure (if available), family and domains, the amino acid sequence and sequence databases. After finding an appropriate entry, the amino acid sequence was obtained and the six-character Uniprot ID noted. The next step was to obtain the DNA sequence for the protein. This was done by scrolling down to sequence databases and using one of the references. In this work, GeneBank has been used and research shows that it is a reliable source (Leray, Knowlton, Ho, Nguyen & Machida, 2019). From the gene, the coding sequence, CDS was obtained. If the open reading

frame for the protein was located on the complementary DNA strand, Reverse Complement tool was used to complement and reverse the DNA sequence.

For efficient production of recombinant proteins, DNA sequences of the enzymes were optimized for the chosen expression system. The choice of expression vectors determined the restriction sites used but in general, the strategy was to introduce as few modifications to the DNA sequences as possible when utilizing specific sites. Two different tools were used to optimize the sequences in this work. For the experimental work, protein DNA sequences were optimized when ordered with GeneArt Instant Designer, Thermo Fisher Scientific by GeneOptimizer algorithm. The deterministic algorithm functions by working through the sequence and randomizing sequences to assess against quality functions (Raab, Graf, Notka, Schödl & Wagner, 2010). The algorithm optimizes gene expression taking into account different aspects such as splicing, transcription, translation, GC content, mRNA degradation and mRNA secondary structure.

In the remainder of this thesis, GenScript's OptimumGene Gene Design system was used. The patented algorithm is claimed to be the most cited gene optimization algorithm and increases protein expression up to 100-fold (Genescript, n.d). OptimumGene Gene Design system considers a variety of factors such as codon adaptability, GC content, mRNA secondary structure, repeat sequences, ribosomal binding sites, inhibition sites and stable free energy of mRNA. When using the optimization tool, the host organism and excluded restriction sites were specified.

After the sequences were optimized, they were verified to translate the same amino acid as the original sequence with EMBOSS Transeq tool under EMBL-EBI. In order to compare the sequences, EMBOSS Needle pairwise sequence alignment tool under EMBL-EBI was used. The GC content of the sequences was evaluated using ENDMEMO DNA/RNA GC Content Calculator.

To determine the location of the affinity tag fused to the gene either articles or the 3D structure was evaluated using Swiss-PdbViewer 4.0 with RCSB PDB pdb files. The three-dimensional structure of a protein often has a hydrophobic core and a hydrophilic surface (Brändén & Tooze, 1999). Therefore, the tag should be attached at the terminus end that is most hydrophilic and/or that is most accessible. Additionally, some information regarding the molecular weight, theoretical pI and domains were found using Expasy's ProtParam and Compute MW pI tool. Structural alignment was performed using RCSB PDB Protein Comparison Tool with jFATCAT-rigid setting. Sequence Manipulation Suite: Codon Usage tool was used to compare codon usage between gene optimized sequences. All tools and databases used in this work is summarized in table 2.

Table 2. The tools and databases used throughout this report.

Description	**Tool or Database**
Protein sequence database	UniprotKB
Nucleotide (GeneBank) sequence database	NCBI Nucleotide
Nucleic acid sequence translation	EMBOSS Transeq
Pairwise sequence alignment	EMBOSS Needle
Customized DNA constructs	Invitrogen GeneArt Gene Synthesis
Codon optimization	GenSmart Codon Optimization
Protein structure	RCSB PDB
Analyzing protein structure	Swiss-PdbViewer 4.0
Chemical and physical protein parameters	ExPASy ProtParam
Compute pI	ExPASy Compute pI/MW tool
Reverse and/or complement DNA sequence	ENDMEMO DNA/RNA GC Content Calculator
Codon usage	Sequence Manipulation Suite: Codon Usage

2.3. Experimental procedures

Most of the experimental procedures and material were chosen according to their availability. The host cell *E. coli* BL21(DE3), pET22b+ vector, ligase and restriction enzymes were already provided by Ecozyme AB and therefore used. Most experiments were performed according to the user manuals of the material. The amount of substance of DNA fragments were calculated using NEBioCalculator - dsDNA: Mass to/from Moles Convertor using equation 1 and 2.

$$Amount\ of\ substance\ of\ dsDNA\ (mol) = mass\ of\ dsDNA\ (g)/molar\ mass\ of\ dsDNA\ (g/mol) \quad (1)$$

$$Molar\ mass\ of\ dsDNA\ (g/mol) = (number\ of\ base\ pairs\ of\ dsDNA \times 617.96\ g/mol) + 36.04\ g/mol \quad (2)$$

617.96 g/mol is the average molecular weight of a base pair excluding water molecules removed during polymerization and the 36.04 g/mol accounts for the two hydroxyl and two hydrogen atoms added back to the ends.

2.3.1 Methods and materials

pMK vectors with genes of interest

5 µg of *G. hollisae* collagenase and subtilisin Carlsberg genes were ordered from Thermo Fisher Scientific using Instant Designer Gene Synthesis Portal service. The gene sequences were optimized with their GeneOptimizer algorithm, dealing with sequence-related parameters involved in gene expression, such as transcription, splicing, translation and mRNA degradation. The genes were assembled from synthetic oligonucleotides and/or PCR products, inserted into pMK-T vectors, purified from transformed bacteria and verified by next generation sequencing.

pET22b+ expression vector

Two different plasmids preparations were used in the practical work. Firstly, pET22b+ vectors prepared in 2019 were used. Secondly, the plasmid was obtained from *E. coli* DH5α using E.Z.N.A.® Plasmid Mini Kit II Protocol with the Spin Method. The procedure was carried out according to the manufacturer's protocol. Glycerol stock with DH5α transformed with pET22b+ was used to streak an agar plate with ampicillin and was then incubated overnight. A single

colony was isolated from the agar plate and was used to inoculate 15 mL LB medium containing ampicillin (50 μg/mL). The culture was then incubated for 16 hours, at 37 °C with shaking at 225 rpm. The culture was centrifuged at 5,000 x g for 10 minutes at room temperature, the media was discarded and the additional 28 steps were implemented. The concentration of the vector was determined with NanoPhotometer (Implen) and stored at -20 °C.

Media

LB media for agar plates were prepared using 10 g tryptone, 5 g yeast extract, 5 g NaCl, 15 g agar, 1000 mL of Milli-Q water and adjusted to pH 7 with 2M NaOH. The media was autoclaved at 121 °C for 20 minutes. Filtered ampicillin (if added) was added after the media reached 55 °C to reach a final concentration of 100 μg/ml. The media was poured to petri dishes to approximately ⅓ of the height, left to solidify in the sterile hood, sealed with parafilm and stored at 4 °C upside down.

LB media for cell growth was prepared using the same recipe as the media for agar plates except for the exclusion of agar. It was stored at 4 °C (with ampicillin) or at room temperature.

The SOC media was prepared by mixing 0.2 g of tryptone, 0.05 g of yeast extract, 0.005 g of NaCl and 10 ml of Milli-Q water. The mixture was then autoclaved as above. Filtered glucose and $MgCl_2$ solutions were added to get final concentrations of 20 mM and 10mM, respectively. The media was stored at -20 °C.

Enzymatic digestion of DNA

The restriction enzymes used were Fastdigest NcoI (FD0573) and XhoI (K1991) from Thermo Scientific. The digestion was carried out according to the manufacturer's instruction with 1 μl of each enzyme, up to 1 μg of plasmid DNA, 2 μl 10X FastDigest™ Green Buffer™ and nuclease free water up to a total volume of 20 μl. The reaction was then incubated at 37 °C for 15 minutes.

Agarose gel electrophoresis

1% agarose gels were prepared by mixing 1 g agarose with 100 mL 1X TAE or TBE running buffer and microwaved until the agarose completely dissolved. After the mixture cooled down to 50-60 °C, 10 µL 10'000x GelRed nucleic acid stain was added and poured into a mini gel tray with a well comb and waited until solidified. The gel was placed into gel box and covered with 1X TBE or TAE running buffer. Digested DNA using Green FastDigest buffer could be directly loaded onto gel, otherwise samples were mixed in 5:1 ratio with 6x TriTrack DNA loading dye. The gel was run at 100 V until the dye line was approximately 80% down the gel. When finished, the gel was placed into an UV box on a glass plate and DNA band were extracted using a scalpel.

E.Z.N.A gel extraction kit

The kit was used according to the manufacturer's instructions and had an estimated DNA recovery yield of 85%. In the laboratory work, the spin protocol was used. The extracted DNA was stored at -20 °C.

Ligation of DNA

The ligation enzyme used in the ligation reaction was 1 µL T4 DNA ligase (Invitrogen) with the buffer, 4 µL 5x T4 DNA ligase buffer (Invitrogen) or 5 µL Anza T4 DNA ligase master mix (Invitrogen). Ligation procedure was carried out according to manufacturer's protocol with a molar ratio of insert to vector of 3:1 calculated using NEBioCalculator. The final volume was adjusted to 20 µL with autoclaved distilled water. The mixture was briefly centrifuged and incubated at room temperature for 5 minutes.

Competent cells

Competent cells used for transformation were: BL21(DE3) chemically competent cells prepared in 2019 and BL21(DE3) chemically competent cells prepared with $CaCl_2$ treatment (Das and Dash, 2015). The latter cells were made by growing an inoculum with a colony from an agar plate at 37 °C, 180 rpm for 24 hours in 5 mL LB media. 0.5 mL culture was transferred to 50 mL

LB and incubated at 37 °C and 225 rpm until an OD600 between 0.35 and 0.4 was reached and then chilled on ice. The pellets were collected by centrifugation at 6000 rpm for 5 minutes at 4 °C and resuspended in 20 mL ice-cold 50-mM $CaCl_2$ solution and incubated on ice for 20 minutes. The solution was centrifuged with the same conditions as before and pellets were resuspended in 2.5 ml ice-cold 50-mM $CaCl_2$ containing 10% glycerol and stored as 100 µL aliquots at −80 °C.

Transformation

Competent cells were thawed on ice and then transferred as 50 µL aliquots to chilled 1.5-mL microcentrifuge tubes. 1 µL of sample DNA was added to the cells, gently mixed and incubated for 30 minutes on ice. The cells were heat-shocked for 30 seconds in a 42 °C water bath, incubated on ice for 2 minutes, followed by addition of 250 µL room-temperature S.O.C. Medium. The tubes were placed on their side in a shaking incubator at 37 °C for 1 hour at 225 rpm and two different volumes (100 µL and 200 µL) were spread on agar plates with ampicillin (100 µg/ml) and incubated overnight at 37 °C.

Methods for colony PCR, protein expression and extraction can be found in appendix B.

3. Results

This section contains the results from the literature study and optimization (3.1-3.5) and the results and discussion from the experimental work (3.6).

3.1. Enzymes

The eight chosen enzymes are described in this section. Among other things, the structure if available are presented, the properties of the enzymes and the sequences. In order to prepare enzymes for cloning experiments their sequences had to be available. This meant that some commercially available enzymes with documented ability to degrade collagen, gelatin or collagen peptides were not applicable for this work as their sequences were not available. Some of the enzymes include neutrase, flavourzyme and protamex (Byun & Kim, 2001; Hong, Min & Jo, 2019).

3.1.1. Collagenase

Microbial collagenases are a part of the MEROPS peptidase M9 family which includes *Vibrio* and *Clostridium* collagenases (Duarte, Correia & Esteves, 2016). MEROPS is a manually curated database containing information about peptidases, their substrates and inhibitors (Rawlings, Barrett & Bateman, 2009). Most of the current knowledge about bacterial collagenases comes from studies conducted on the enzymes produced by *Clostridium histolyticum* (Teramura et al., 2011). Clostridial collagenases are multi-domain containing signal peptide, pro-peptide domain, collagenase unit, up to two linking domains e.g. PKD (polycystic kidney disease)-like domains and three CBD (collagen binding domains) (Duarte et al., 2016). Two collagenase classes are classified; class I has high activity against collagen but less activity towards synthetic peptides and class II has limited activity against collagen while higher towards synthetic peptides. Common bacterial collagenases are *C. histolyticum* collagenase G and *Clostridium perfringens* collagenase A which belongs to class I. A class II collagenase is *C. histolyticum* collagenase H. Microbial collagenases are in general less specific than those of zoological origin and can degrade both water-soluble and native molecules. Almost all types of

collagen are therefore prone to be attacked at various sites in the chains by bacterial collagenases. The enzymes make numerous cleavages within the triple helical regions in the Y-Glycine bond in the X-Y-Gly sequence (Teramura et al., 2011).

3.1.1.1 Clostridium histolyticum collagenases

Bacillus histolyticus was discovered in 1916 and was described as an obligate anaerobe rich in flagella and very motile (Oakley & Warrack, 1950). It was renamed to *C. histolyticum* in 1923. *C. histolyticum* is a Gram-positive rod-shaped microbe that often appears in pairs or as short chains (Hatheway, 1990). The bacterium produces powerful extracellular collagenases, isomers called ColG, ColH and ColT (Eckhard, Schoenauer & Brandstetter, 2013). These bacterial collagenases hydrolyze native collagen at multiple positions and are capable of complete degradation. ColG has a mass of 126 kDa and ColH 116 kDa. Both collagenases have a HEXXH zinc binding motif with an additional glutamate positioned 28-30 amino acids downstream. Additionally, they also require calcium for both collagenolytic and peptidolytic activity. In ColG there are 6.8 Ca^{2+} binding sites and 5.1 in ColH which were determined with atomic absorbance activity (Bond & Van Wart, 1984).

The domains of the enzymes are shown in figure 2. ColG consists of the domains S1, S2, S3a and S3b meanwhile ColH consists of S1, S2a, S2b and S3.

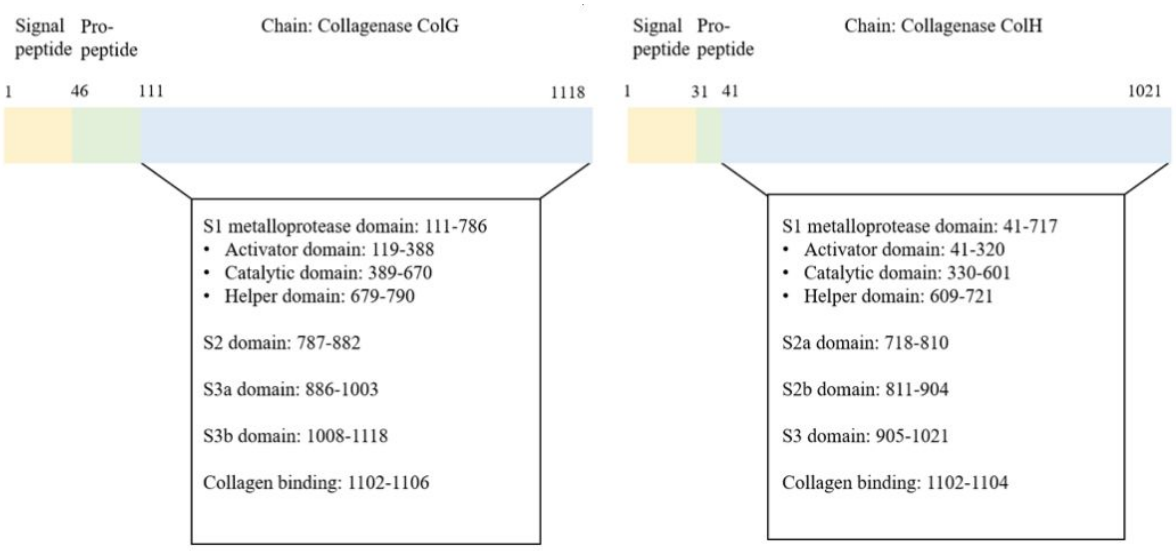

Figure 2. Comparison of the domains of ColG (Uniprot id: Q9X721) and ColH (Uniprot id: Q46085).

The S1 catalytic domain belongs to the M9B family of metallopeptidases (Bauer et al., 2012). The peptidase domains of ColG and ColH have 56 % sequence identity and 71 % sequence similarity (Eckhard et al., 2013). The area around the active site has 68 % sequence identity. The primary sequence of S2, S2a and S2b resembles PKD domains which have been shown to increase collagen binding (Bauer et al., 2012).

The S3, S3a and S3b domains are homologous and classified as bacterial PPC- (prepeptidase C-terminus) domains. A structural alignment using RCSB PDB Protein Comparison Tool (setting: jFATCAT-rigid) with S3b domain of ColG (RCSB id: 1NQD) and S3 domain of ColH (RCSB id: 3JQV) is shown in figure 3, these sequences also have 29.20 % sequence identity and 52.21 % sequence similarity. In ColH and ColG the domains bind either to insoluble collagen or to collagen fibrils in different tissues' extracellular matrix (Bauer et al., 2012).

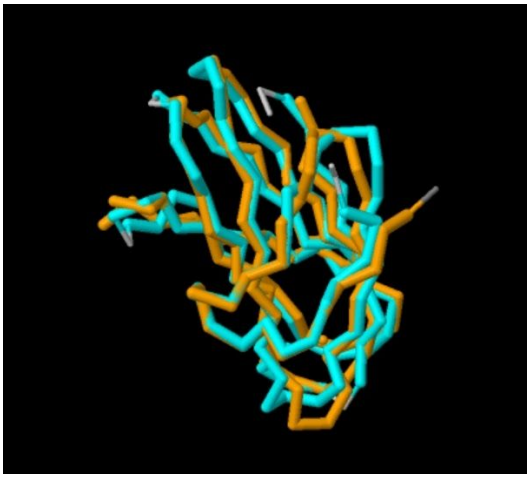

Figure 3. Structural alignment of S3b and S3 domains. S3b is shown in orange and S3 in turquoise.

Collagen zymogram experiments have shown that the enzymes are present not only in the 126 kDa and 116 kDa forms but also as 67, 78 and 82 kDa for ColG and 98 and 105 kDa for ColH (Matsushita et al., 1999). These gelatinases with identical N-terminus were produced by post-translational proteolytic removal of C-terminus fragments however, only the full-length collagenases were shown to digest insoluble collagen.

In 1987 researchers observed that collagenase type I and II from *C. histolyticum* degrades collagen type I with different mechanisms (French, Mookhtiar & Van Wart, 1987). Type I initially hydrolyzed collagen close to the C-terminus followed by a second cleavage at the N-terminus and further degradation which resulted in fragmented collagen. Type II on the other hand initially cleaved collagen within the interior of the molecule across all three chains followed by further degradation. The actions of the enzymes are complementary and fluorescent microplate collagen-degrading activity assays have shown synergy between type I and II collagenase with increasing percentages of the amount of type II enzyme (Breite, McCarthy & Dwulet, 2011). Furthermore, it has been observed that both type I and II collagenase was required for in vitro digestion of soluble calfskin collagen fibrils (McCarthy et al., 2008). The specific binding of collagen to ColG was described in 2011 to involve both the activator and catalytic domain (figure 4) (Eckhard, Schönauer, Nüss, & Brandstetter, 2011). Firstly, the collagen docks to the catalytic domain. After, HEAT repeats on the activator domain interacts with the triple helices which changes the structure of collagenase to semi-open conformation which allows degradation of collagen. HEAT repeats are repetitive clusters of short amphipathic alpha-helices (Yoshimura & Hirani, 2016). It was proposed in 2013 that this binding mechanism destabilizes the collagen molecule which presents vulnerable sites to the peptidase domain of *C. histolyticum* collagenases (Eckhard et al., 2013).

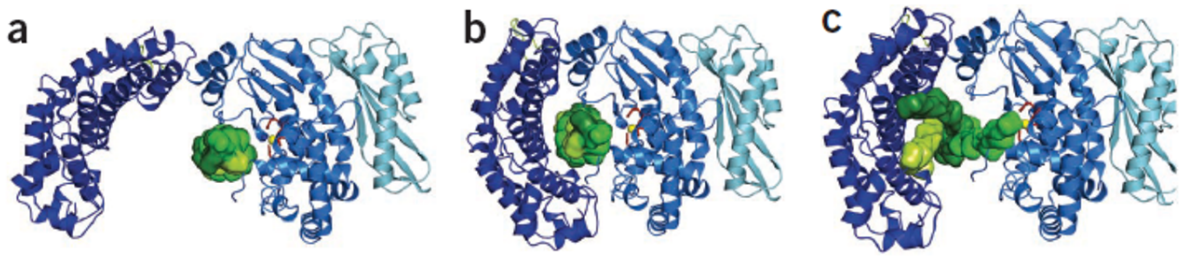

Figure 4. Binding of collagen to ColG. Picture a show the collagen docked to the catalytic domain, no hydrolysis occurs. Secondly, picture b depicts the activator HEAT repeats interacting with the triple helix which is a requirement for collagen hydrolysis. The semi-open confirmation in picture c allows for degradation across the α-chains. Once collagen is degraded, the conformation of collagenase relaxes back into an open-ground state conformation (Eckhard et al., 2013).

3.1.1.2 Grimontia (Vibrio) hollisae collagenase

Vibrio hollisae was discovered in 1982 and is a Gram-negative halophilic rod, 0.5 μm wide by 1.5-2 μm long often containing a single polar flagellum (Hickman et al., 1982). *G. hollisae* strain 1706B has proven to produce a more efficient collagen degrading collagenase than those of *C. histolyticum* (Teramura et al., 2011). The mature 74 kDa enzyme is 680 amino acids long (with pre-prodomain 84 kDa and 767 amino acids long), stable between pH 4.5-11 and has an optimum efficiency at pH 7-8. The collagenase is a zinc dependent metalloproteinase containing a HEXXH zinc-binding motif sequence important for electron transfer during enzyme catalysis (Menach, Hashida, Yasukawa & Inouye, 2013). The catalytic zinc ion is chelated by two histidine residues. The collagenase consists of a pre-pro region (amino acid 1-87), a catalytic domain (amino acid 88-615) with the active site HEYVH (Takita el al., 2018) and bacterial PPC (pre-peptidase C-terminal) domain (amino acid 688-749) (Tanaka et al., 2018). *G. hollisae* collagenase shows 12% and 11% sequence identity with ColG and ColH from *C. histolyticum* (Teramura et al., 2011).

Purified enzyme expressed in *Brevibacillus choshinensis* showed that the C-terminal of the mature enzyme gets autodegraded and the 74 kDa protein converts to a 60 kDa and 40 kDa enzyme respectively with the 60 kDa being most stable. The C-terminus of the 60 kDa recombinant collagenase is obtained by cleaving between serine and glycine in the sequence, GDSGAG in the linker region (U.S. Patent No. US 10,047,353 B2, 2018).

The PPC domain binds to type I collagen by recognizing its triple-helical conformation (Tanaka et al., 2018). The C-terminus domain coupled to Sepharose beads was used to test its proteolytic activity towards different types of collagen. The results showed that the collagenase binds to collagenase type II, III, IV and V besides type I (Tanaka et al., 2020). Additionally, the CBD of *G. hollisae* collagenase and *C. histolyticum* have different primary structures with sequence identities between 11% and 23% (Tanaka et al., 2018). In silico structural prediction shows that the different CBDs could be homologs.

Collagenolytic activity was compared for the *G. hollisae* and *C. histolyticum* collagenase using FITC-labelled type I collagen which showed the former having over four and six times higher specific activity and specificity constant, respectively (Teramura et al., 2011).

Recently another 62 kDa (1677 base pair) C-terminal truncated collagenase from the same 74 kDa protein was amplified using PCR and expressed in *Brevibacillus* expression system (Tanaka et al., 2020). The 62 kDa protein and above described 60 kDa protein had similar collagenolytic activity but the former obtained more stable transformants. The difference in domains of the different enzymes are presented in figure 5. The 74 kDA enzyme contains PPC domain and linker region, 62 kDA enzyme lacks PPC domain and has a shorter linker and the 60 kDa collagenase has an even shorter linker than the previous enzyme. Optimal activity for the protein was obtained between 7.5-9 pH at 30-40 °C and retained collagenolytic activity for 24 hours at 37 °C. The collagenase was proven to have three times higher specific activity against FITC-labeled collagen and synthetic FALGPA peptide compared to *C. histolyticum* collagenase and was able to degrade collagen type I to V.

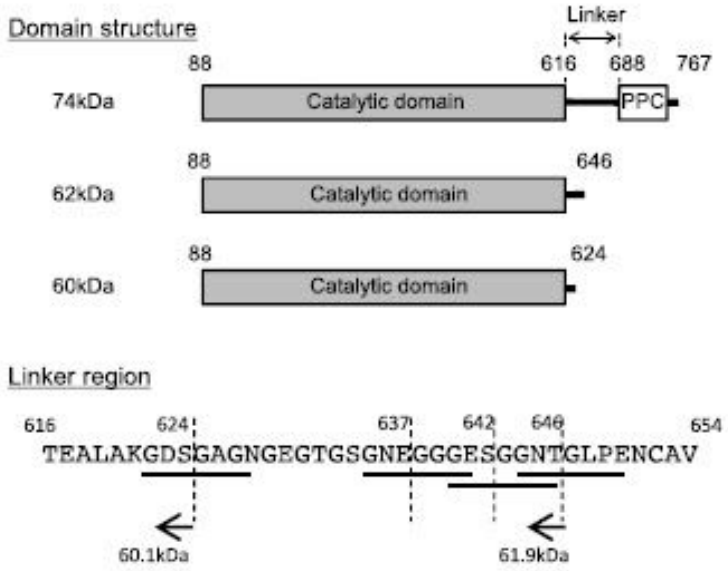

Figure 5. The domain structure of 74, 62 and 60 kDa G. hollisae collagenase (Tanaka et al., 2020).

3.1.1.3 Bacillus cereus collagenase

B. cereus is a Gram-positive, motile and spore-forming bacterium (Granum & Lund, 1997). The facultative anaerobic microorganism commonly found in soil causes two types of food poisoning, emetic and diarrhoeal type. *B. cereus* has been found to produce collagenolytic enzyme that can be regarded as a true collagenase, degrading soluble and insoluble collagens, gelatin and azocoll (azo dye-impregnated insoluble collagen) with similarities to the *C. histolyticum* collagenases (Makinen & Makinen, 1987). Lysates of the *B. cereus (*ATCC 14579) showed stark proteolytic activities (with molecular weights ranging from 70 to 125 kDa) when run on zymogram gels with gelatin during logarithmic growth phase (Abfalter, Schmidt & Wessler, 2015).

The domains of the 110 kDa collagenase ColA were predicted in 2011, using SMART and SignalP tools, see figure 6 (Abfalter et al., 2016). Pairwise sequence alignment using EMBOSS Needle shows 43.2% and 45.5% sequence identity between *B. cereus* ColA to ColG and ColH from *C. histolyticum* respectively.

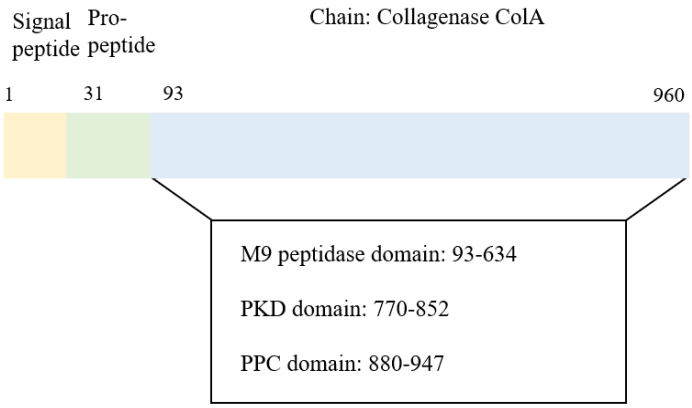

Figure 6. Domains of B. cereus collagenase ColA (Uniprot id: Q81BJ6)

Enzyme degradation of tropocollagen type I over time showed that ColA (without signal and propeptide) was more effective than ColG (Abfalter et al., 2016). The electrostatic potential surface of ColA and ColG was generated using PyMOL which showed that the surface properties

of the activator domains differed whereas for ColA, the surface was basic and for ColG acidic (shown as orange arrows in figure 7). It was postulated that this difference could explain the increased activity of ColA.

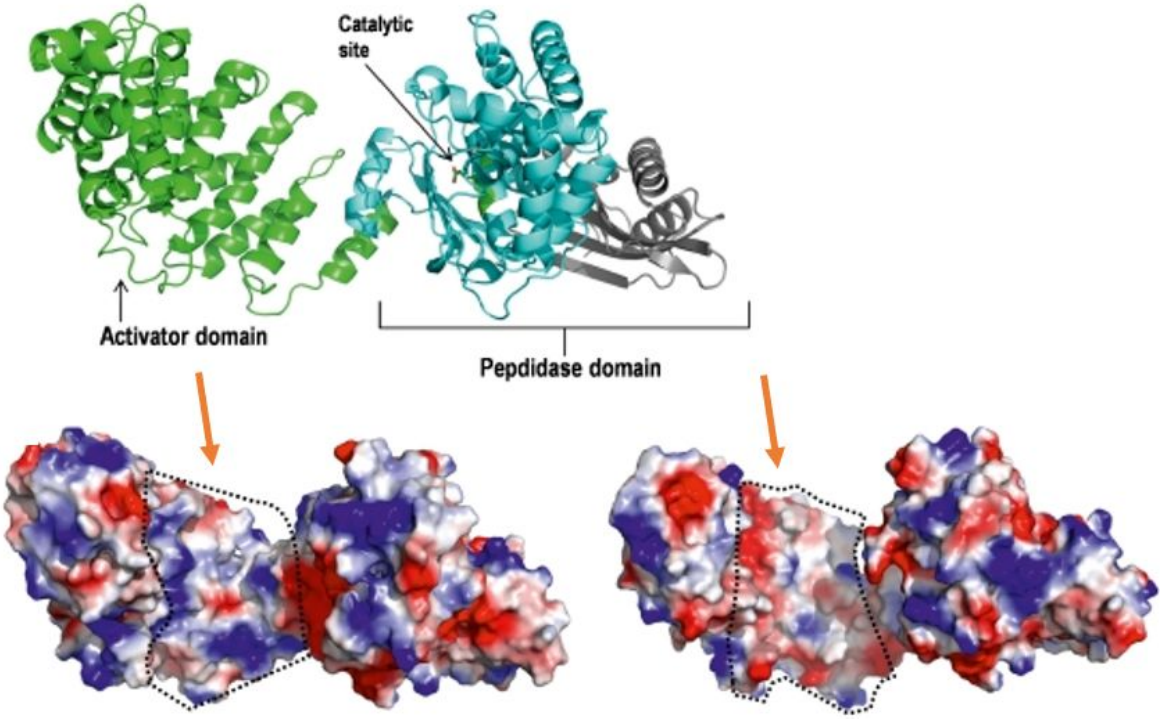

Figure 7. The top figure shows the predicted structure of ColA. Electrostatic potential surface of ColA and ColG generated using PyMOL is displayed at the bottom. The orange arrows point at the surface of the activator domains.

3.1.2. Subtilisin

Subtilisins are extracellular serine endopeptidases with origins from strains of *Bacillus subtilis* or related bacteria (Markland & Smith, 1971; Ottesen & Svendsen, 1970). There are several different subtilisins from different *Bacillus* strains, most commonly subtilisin Carlsberg (also known as subtilisin A, subtilopeptidase A and alcalase Novo) from *Bacillus licheniformis*, subtilisin BPN' (also known as subtilisin B, Nagarse, subtilopeptidase B, subtilopeptidase C and bacterial proteinase Novo) from *Bacillus amyloliquefaciens*, subtilisin 147 (or esperase) and subtilisin 309 (or savinase) from *Bacillus lentus*, Maxacal from *Bacillus alcalophilus* and subtilisin 168 (or subtilisin E) from *B. subtilis* (Graycar, Bott, Power & Estell, 2013). Subtilisins

are enzymes with broad specificity and application, for example they are found in detergents for removing protein-based stains (Renneberg, Berkling & Loroch, 2016).

Subtilisins are monomers consisting of approximately 275 amino acids and have a molar weight around 27 kDa (Markland & Smith, 1971). The enzyme is first expressed as a preproenzyme, where the propeptide is about 77 amino acids long and is required for correct folding of the enzyme (Ikemura, Takagi & Inouye, 1987; Strausberg, Alexander, Wang, Schwartz & Bryan, 1993). The amino acids involved in the catalytic activity of subtilisins are Asp32, His64 and Ser221 (numbering from subtilisin BPN') (Carter & Wells, 1988). The His64 needs to be deprotonated for the enzyme to be active, which makes the enzyme reliant of a basic pH with a pKa of approximately 7 (Philipp, Tsai, & Bender, 1979). Subtilisin Carlsberg has a central ß-sheet system surrounded by α-helices (figure 8). Subtilisins also have two calcium-binding sites, one with low affinity and one with high affinity (Graycal et al., 2013).

Commercial subtilisins are produced in different species of *Bacillus* (Chaplin, 2004). Subtilisin has also been produced in *E. coli*, however inclusion bodies are formed and require refolding (Ikemura et al., 1987; Zhang et al., 2005). Alcalase (subtilisin Carlsberg) gave the highest degree of hydrolysis of porcine skin when compared with other commercially available proteases (flavourzyme, bromelain, protamex, papain and neutrase) under several optimal conditions (Hong et al., 2019). When three different enzymes: Alcalase, Flavourzyme and trypsin were used to hydrolyze turkey head collagen, alcalase and flavourzyme gave the highest degree of hydrolysis (Khiari, Ndagijimana & Betti, 2014). However, the combination of the three gave the highest degree of hydrolysis.

Figure 8. 3D structure of Subtilisin Carlsberg. The figure was made in Swiss-PdbViewer 4.1.0 using a PDB file from RCSB Protein Data Bank with accession number 1BFU. The N-terminus, C-terminus and the amino acids in the catalytic triad (Asp32, His64 and Ser221) are marked.

3.1.3. Trypsin

Trypsin was discovered in 1866 by Wilhelm Kühne (Kühne, 1867). The serine protease is produced in the pancreas of a large number of vertebrates (Heissel, Frederiksen, Bunkenborg & Højrup, 2019). It is also the most used protease in proteomics and is relatively cheap. The molecular weight of trypsin from bovine pancreas is 23.7 kDa, the optimum pH for proteolytic activity is between 7 and 8, however the enzyme is most stable at pH around 2.3 (Wirnt, 1974). Trypsinogen is the zymogen of trypsin. The trypsinogen needs to be activated by enterokinase or already active trypsin. Trypsin hydrolyzes on the carboxy side of lysine and arginine residues but has reduced activity when followed by proline or surrounded by acidic residues (Heissel et al., 2019; Wirnt, 1974). Studies have shown that trypsin is capable of cleaving only unfolded regions

or non-helical regions of type I and III collagen (Mirigian et al, 2013).

The structure of trypsin has become the prototype for the S1 family of serine endopeptidases. The three-dimensional structure was first determined 1974 separately by two different groups (Huber et al., 1974; Stroud & Kay, 1974). Trypsin consists of a single polypeptide chain and as with other serine endopeptidase in its family, the catalytic residues are those in the catalytic triad consisting of aspartate, histidine and serine (Baird & Craik, 2013). The catalytic triad also forms a bridge between two ß-barrels domains. The three-dimensional fold, the catalytic triad and the residues surrounding it are always preserved between trypsins from different species. Other structural characteristics include a high affinity calcium-binding site formed by Glu70 to Glu80 which is required for its stability (Bode & Schwager, 1975). The structure of trypsin is shown in figure 9.

Several studies have shown successful results using trypsin to digest collagen (Jia et al., 2010; Khiari et al., 2014; León-López, Fuentes-Jiménez, Hernández-Fuentes, Campos-Montiel & Aguirre-Álvarez, 2019). Trypsin has been expressed in *E. coli* (Evnin, Vasquez & Craik, 1990; Higaki, Evnin & Craik, 1989; Vasquez, Evnin, Higaki & Craik, 1989). It has also been expressed in many other systems like *Saccharomyces cerevisiae* (Hedstrom, Szilagyi & Rutter, 1992) and *Pichia pastoris* (Halfon & Craik, 1996).

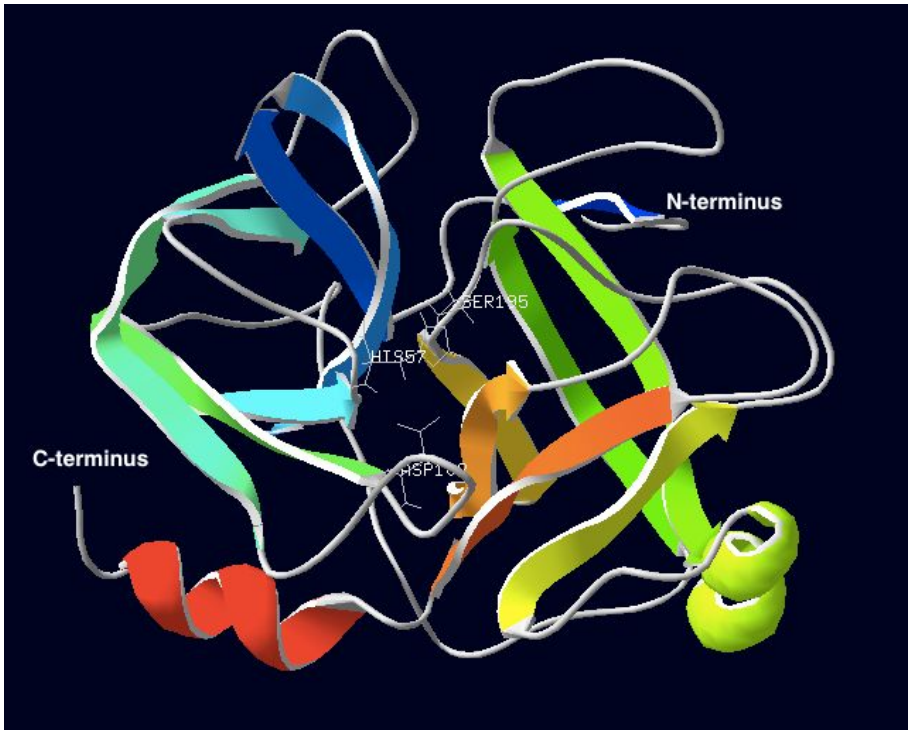

Figure 9. 3D structure of bovine trypsin. The figure was made in Swiss-PdbViewer 4.1.0 using a PDB file from RCSB Protein Data Bank with accession number 1UTN. The amino acids in the catalytic triad (His57, Asp102 and Ser195) are marked, as well as the N- and C-terminus.

3.1.4. Pepsin

Pepsin is an endopeptidase with broad specificity and is often considered to be the first enzyme to be discovered (Tang, 2013). It was discovered in the eighteenth century and named by T. Schwann in 1825. The major pig and human gastric proteolytic enzyme was named pepsin A to differentiate it from other similar enzymes. Pepsin B and C were isolated in the late 50s (Ryle & Porter, 1959). Pig pepsin B is another gene product from pepsin A and is a minor gastric proteolytic enzyme (Nielsen & Foltmann, 1995). Pepsin C is a gatricsin also isolated from human gastric juice (Richmond, Tang, Wolf, Trucco & Caputto, 1958).

Pig pepsin consists of a single chain with 326 residues and has a molecular weight of 34.6 kDa (Tang et al., 1973). The enzyme is an aspartic protease and the catalytic residues are Asp32 and Asp215, whereof Asp32 must be deprotonated and Asp215 the opposite to be able to bind to substrate (Antonov et al., 1978). The structure of pepsin is shown in figure 10. The propeptide

was postulated to be critical for pepsin folding (Tang, 2013). Pepsinogen is the zymogen of pepsin. Pig pepsinogen spontaneously converts to pepsin at pH below 5 (Herriott, 1938). Pepsin is much less specific than other proteases, like trypsin. Pepsin preferably cleaves after bulky hydrophobic amino acids residues, in general after phenylalanine and leucine (Fruton, 1970; Sachdev & Fruton, 1970). The structure contains two lobes, mostly ß strands and only a few short helical segments (Sielecki, Fedorov, Boodhoo, Andreeva & James, 1990). The enzyme hydrolyzes between pH less than 1 and up to almost pH 6 (Lin et al, 1992).

Commercial pepsin comes from acidified pig stomach homogenate, which makes it possible to digest impurity proteins before recovering the pepsin. However, commercial pepsin may contain nicks from autolysis and extra residues in the N-terminus (Tang et al., 1973). Recombinant pepsinogen has been produced in *E. coli* (Lin, Wong & Tang, 1989; Norihiro et al., 1988). The purification steps involve washing with a Triton buffer to get a pure protein and then the pepsinogen inclusion bodies are refolded with a urea solution (Tang, 2013).

Pepsinogen B has never been found in species other than pigs and dogs and the function is also unknown (Szecsi, 2013). However, pepsin B hydrolyzes gelatin efficiently but hydrolyzes hemoglobin poorly and the opposite can be said about pepsin A (Kageyama, 2006). Pepsin has been used to hydrolyze Alaska pollock skin (Jia et al., 2010), tuna backbone protein (Je, Qian, Byun & Kim, 2007), chicken bone extract (Nakade et al., 2010) and gelatin from jumbo squid skin (Mendis, Rajapakse, Byun & Kim, 2005).

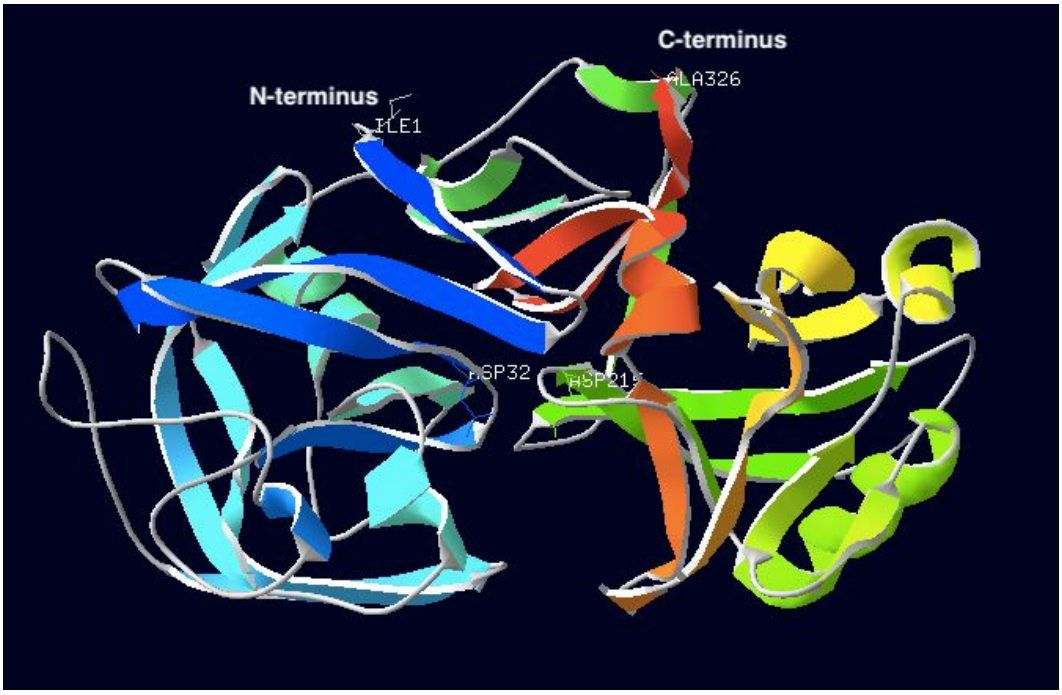

Figure 10. 3D structure of porcine pepsin. The figure was made in Swiss-PdbViewer 4.1.0 using a PDB file from RCSB Protein Data Bank with accession number 4PEP. The catalytic residues (Asp32 and Asp215), N-terminus and C-terminus are marked.

3.1.5. Papain

Papain is a cysteine peptidase with broad activity and got its name after being discovered in the latex of the papaya fruit (Storer & Ménard, 2013). The enzyme has endopeptidase, amidase and esterase activities (Glazer & Smith, 1971). Papain consists of a single chain with 212 amino acids and has a molecular weight of 23.4 kDa (Storer & Ménard, 2013). The enzyme has a pI of 8.75 and the catalytic residues are Cys25 and His159. The peptide chain forms two domains, which forms a groove and in this groove, the active site is found (Kamphuis, Kalk, Swarte & Drenth, 1984). The active site consists of seven subsites (S1-S4 and S1′-S3′) and are located on both sites of the catalytic site (Schechter & Berger, 1967). Each subsite can accommodate one amino acid residue of a substrate (P1-P4 and P1′-P3′). The preference for the S1 subsite is wide, it has been suggested that it has specificity for small and non-polar residues, larger residues and charged residues (Hilaire, Willert, Juliano, Juliano & Meldal, 1999). It has also been proposed that the S3 subsite has preference towards proline and valine, the S2 subsite has specificity

towards hydrophobic residues and specificity for small residues were dominant for the other subsites. The structure of papain is shown in figure 11.

Papain is stable and active from pH 4 to pH 10 and to temperatures up to 80 °C (Glazer & Smith, 1971). Papain can be extracted from papaya latex. It has also been produced as a precursor in baculovirus/insect cell system (Vernet et al., 1990), *E. coli* (Taylor et et al., 1992; Choudhury, Roy, Chakrabarti, Biswas & Dattagupta, 2009) and *S. cerevisiae* (Ramjee, Petithory, McElver, Weber & Kirsch, 1996). In *E. coli* the recombinant enzyme was produced in an insoluble form, requiring refolding for an active enzyme. However, it is possible to obtain approximately 400 mg of refolded propapain per liter culture of *E. coli* and the propapain could be autocatalytically activated at low pH (Choudhury et al., 2009).

Papain has been used to digest fish skin from Alaska pollock (Jia et al., 2010) and gelatin from cobia skin (Yang, Ho, Chu & Chow, 2008).

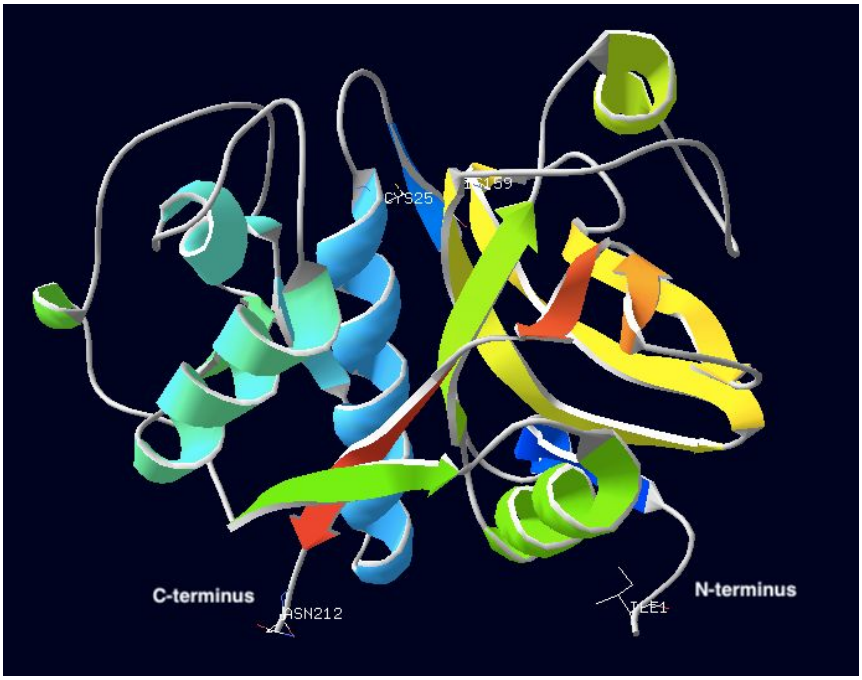

Figure 11. 3D structure of papain. The figure was made in Swiss-PdbViewer 4.1.0 using a PDB file from RCSB Protein Data Bank with accession number 1PPN. The catalytic residues (Cys25 and His159), the N-terminus and the C-terminus are marked.

3.2. Expression systems

There are an abundance of techniques and possible expression systems for producing recombinant proteins. Hosts are commonly bacteria (such as *E. coli*), yeast, insect and mammalian cells. Some comparable features of these expression systems are shown in table 3.

Table 3. Comparison of different expression systems. The table is adapted from Arya, Bhattacharya & Saini, 2008 and Toikkanen, n.d.

Factor	*E. coli*	Yeast	Insect cells	Mammalian cells
Recombinant protein yield	High	High	High	Low
Cell growth	Rapid	Rapid	Slow	Slow
Growth media	Simple and inexpensive	Simple and inexpensive	Complex and expensive	Complex and expensive
Protein folding	Refolding usually required	Refolding may be required	Proper folding	Proper folding
Post-translational modification	No	Yes, but N-or O-linked glycosylation different from higher eukaryotes	Yes, similar to higher eukaryotes	Yes, similar to native protein
Extracellular expression	No*	Yes	Yes	Yes

** E. coli does not efficiently secrete protein.*

In 1885, *E. coli* was first described and isolated (Escherich, 1885). *E. coli* is a Gram-negative bacteria and is one of the most common host cells used for protein production (Khow & Suntrarachun, 2012). This bacteria has unmatched rapid growth kinetics, a doubling time of twenty minutes, high cell densities, inexpensive simple growth media and fast transformation with exogenous DNA (Rosano & Ceccarelli, 2014). There are many well-documented expression systems using *E. coli* however, all genes can not be successfully expressed in the bacteria due to its inability to perform post-translational modifications such as glycosylation, disulfide linkages for proper protein folding and phosphorylation (Arya et al., 2008). Moreover, the bacteria lacks an efficient protein secretion mechanism to the medium. Additionally, high-level expression of

proteins in the bacteria can lead to formation of inclusion bodies, which complicates purification and often leads to lower enzyme production (Singh, Upadhyay, Upadhyay, Singh & Panda, 2015). Inclusion bodies are not limited to *E. coli* and can also form in yeast, insect and mammalian cells (Palmer & Wingfield, 2012).

Things to consider when expressing recombinant proteins in *E. coli* and other organisms are the host strain, type of vector, cultivation conditions and purification of the recombinant protein (Mergulhao, Monteiro, Cabral & Taipa, 2013). If plasmids are chosen the plasmid copy number, promoter and terminator are things to reflect on. When expressing enzymes in *E. coli* the final location will affect the production of the recombinant proteins. Expressing enzymes in different cellular locations affects properties such as production level, stability, host contamination and downstream processing. A comparison is presented in table 4.

Table 4. Comparison between recombinant protein production in different final locations (cytoplasm, periplasm and medium) (Mergulhao et al., 2013).

Factor	**Cytoplasm**	**Periplasm**	**Medium**
Production level	High	Low	Very low
Stability of the protein	High*	Moderate	High
Host protein contamination	High	Low	Very low
Downstream processing	Complex	Simple	Very simple

**The stability of the protein is high in the cytoplasm if inclusion bodies are formed, otherwise it is low due to degradation by the high level of proteases present in the cytoplasm.*

Yeast cells are widely used for recombinant protein expression and many advantages. It has been shown that yeast can produce higher quantities of isolated protein than native sources (Bonader & Bill, 2012). The fungi cells are easy to genetically manipulate and strains can be optimized for production of specific enzymes (Nielsen, 2014). Compared to bacterial systems, yeast cells are capable of post-translational modifications such as disulphide bonds formation and glycosylation (Liu et al., 2013). In comparison to mammalian cells, yeast cells grow more rapidly and large-scale cultures can be performed with fermentation (Nielsen, 2014). The fungi grow on

carbon sources, reaches high cell densities fast and are halo- and thermo-tolerant (Liu et al., 2013).

Things to consider when designing an expression system with yeast are the same as described for *E. coli* and other organisms above. Examples of yeasts that can be used are *S. cerevisiae*, *P. pastoris*, *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Arxula adeninivorans* and *Candida boidinii* (Liu et al., 2013). There are three different kinds of vectors for yeast: integration plasmids (YIp), episomal plasmids (YEp), and centromeric plasmids (YCp) (Vieira Gomes, Souza Carmo, Silva Carvalho, Mendonça Bahia & Parachin, 2018). YEp is a high-copy number plasmid inside the cell (5–30 copies) that allows strong gene expression but can be a considerable burden on cells leading to an increased plasmid instability. YCps are low-copy number plasmids that are more stable but the lower gene expression levels restrict their use. YIps integrate genes of interest into the native yeast chromosome which is beneficial because it decreases the selective pressure occurring after a recombinant strain is designed. It is also possible to use multiple integrations of an expression cassette to increase the total amount of the recombinant protein. Another thing to consider is that yeast can express the recombinant protein intracellular or extracellular.

3.2.1. Escherichia coli BL21(DE3)

The *E. coli* strain chosen for expression is BL21(DE3) due to the high-level production of recombinant proteins (Jeong, Kim & Lee, 2015). Additionally, the strain with a 4,557,508 base pair long genome lacks proteases such as Lon and OmpT that degrade target proteins (Hausjell, Weissensteiner, Molitor, Halbwirth & Spadiut, 2018; Jeong et al., 2009).

3.2.1.1. pET22b+ vector

The expression vector pET22b+ (figure 12) is one of the most developed vectors for cloning and expression of recombinant proteins in *E. coli* (Novagen, 1991). The gene expression is controlled by the T7 RNA polymerase from the T7 bacteriophage which transcribes genes downstream from the T7 promoter (base pair 360-377) (Tabor, 2001). This polymerase has such high activity and selectivity that most of the cell's resources are used to target gene expression and

consequently, more than 50% of the total cell protein can be composed of the desired product hours after induction. Protein expression can be induced by IPTG (Novagen, 1991).

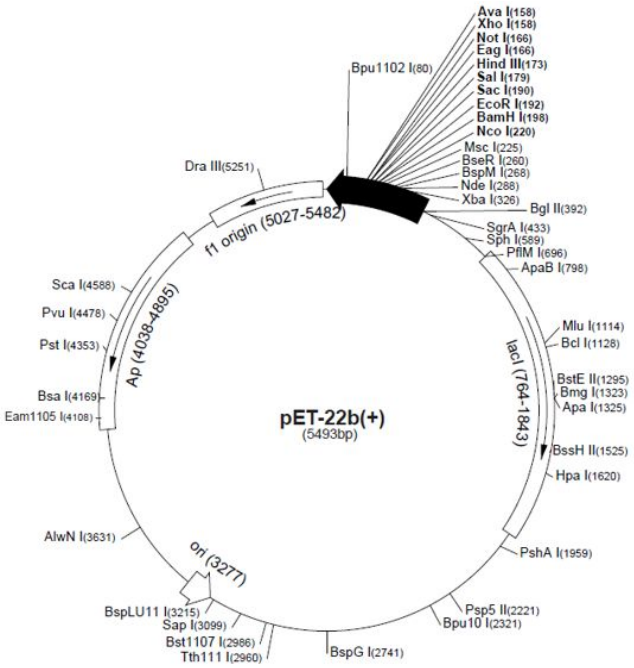

Figure 12. pET22b+ expression vector (Novagen, n.d.)

The cloning region of the pET22b+ vector contains the T7 promoter, lac operator, pelB leader, restriction sites, C-terminus histidine tag and T7 terminator (figure 13). The 63 base pair long pelB leader sequence directs synthesized polypeptides to the periplasm of *E. coli* where the sequence is removed by membrane bound signal peptidases (Singh et al., 2013).

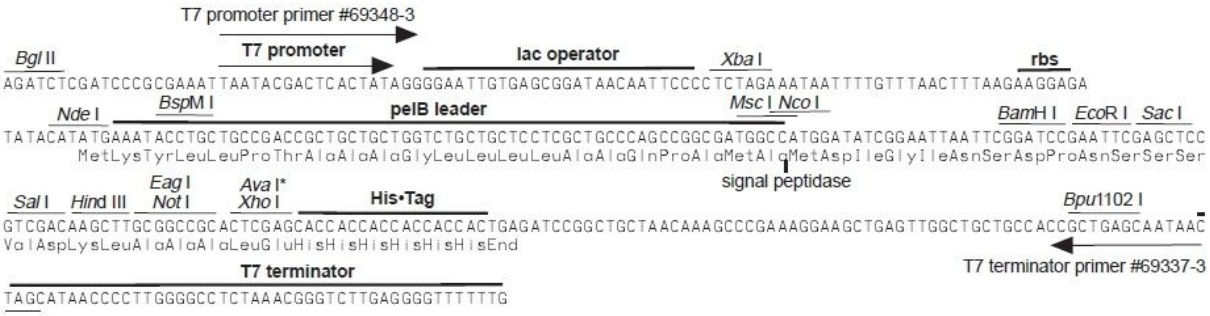

Figure 13. Expression and cloning region of pET22b+ (Novagen, n.d.)

3.2.1.2. pAES40 vector

YebF is a small protein (10.8 kDa) with unknown function that can be secreted to the medium by laboratory strains of *E. coli* (Zhang, Brokx & Weiner, 2006). It is also possible to attach a protein to the carboxyl terminal of YebF. The attached protein is efficiently secreted to the medium in an active form. YebF has a signal peptide and is first transported to the periplasm. The porin OmpF is required for the secretion (Prehna et al., 2012). OmpC is also involved in the secretion and OmpX is a potential binding partner.

The expression vector pAES40 has a tac promoter, lac repressor, ampicillin resistance, the protein yebF, multiple cloning sites, an enterokinase cleavage site between the cloning sites and yebF and a 6x his tag (figure 14) (ACES, n. d.). IPTG and lactose can be used for induction. Enterokinase is an enzyme that cleaves after lysine in the following sequence: Asp-Asp-Asp-Asp-Lys.

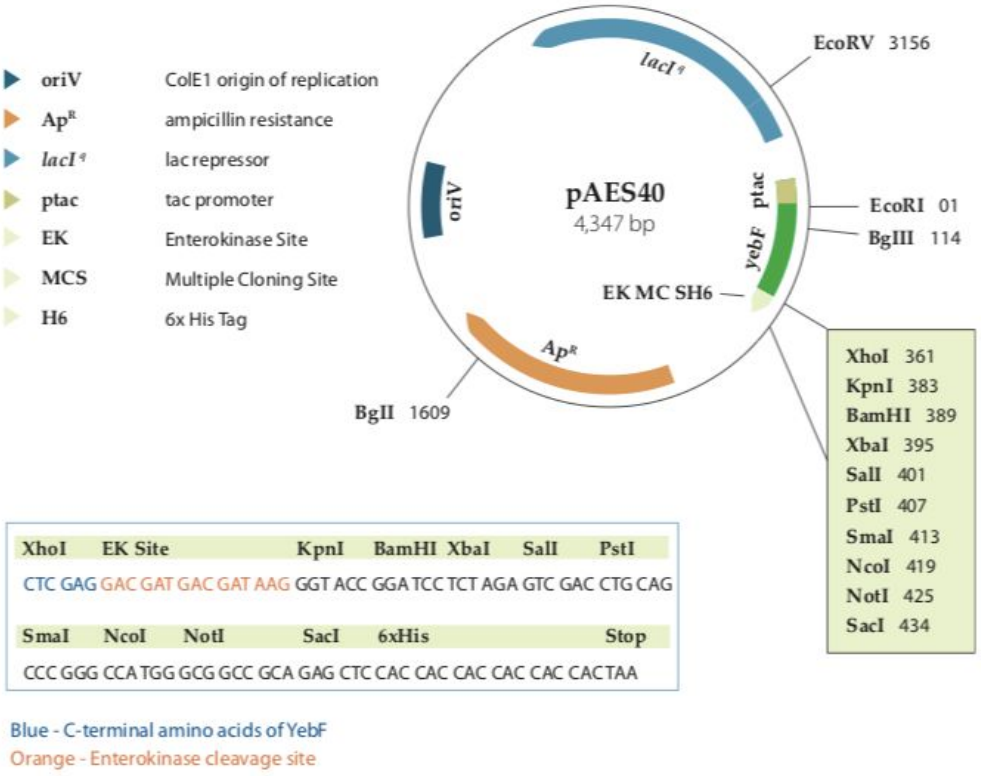

Figure 14. Map of the pAES40 plasmid and its cloning sites (ACES, n. d.).

3.2.2. Pichia pastoris

The haploid homothallic ascomycetous yeast *P. pastoris* is an exceptional expression system due to is ability to grow on minimum media at high cell densities with excellent extracellular proteins secretion yields (Cereghino & Cregg, 2000; Rivera, Gutiérrez-Ortega & Sandoval, 2018; Wegner, 1990). *Pastoris* does not effectively ferment sugars and hence why only small quantities of toxic fermentation products such as ethanol are generated (Cregg, 2007). This in turn results in carbon sources being processed into cell mass, yielding the exceptionally high cell yields. Compared to *S. cerevisiae*, *P. pastoris* produces less endogenous protein secretion and often, the recombinant protein is virtually the only protein in the cell culture which makes purification easy (Darby, Cartwright, Dilworth & Bill, 2012). Larger production of recombinant protein expression in *P. pastoris* requires high agitation and aeration often using continuously stirred tank reactors with dissolved oxygen levels from 20 to 30 % (Darby et al., 2012; Liu et al., 2013).

3.2.2.1. pPICZα vector

The pPICZα vector displayed in figure 15 contains the alcohol oxidase (*AOX1)* promoter, α-factor secretion signal, the highly effective affinity tags *c-myc* epitope and His_6-tag, Zeocin resistance marker and 5′ *AOX1* gene (Hilpert et al., 2001; Invitrogen, n. d.). The tightly regulated methanol induced *AOX1* promoter promotes particularly high amounts of proteins by uncoupling the growth phase from the production phase which excuses the cell from stress and can even allow production of *P. pastoris* toxic proteins (Ahmad, Hirz, Pichler & Schwab, 2014). The α-factor secretion signal consisting of signal and propeptide is successful in its secretion abilities (Lin-Cereghino et al., 2013; Liu et al., 2013,).

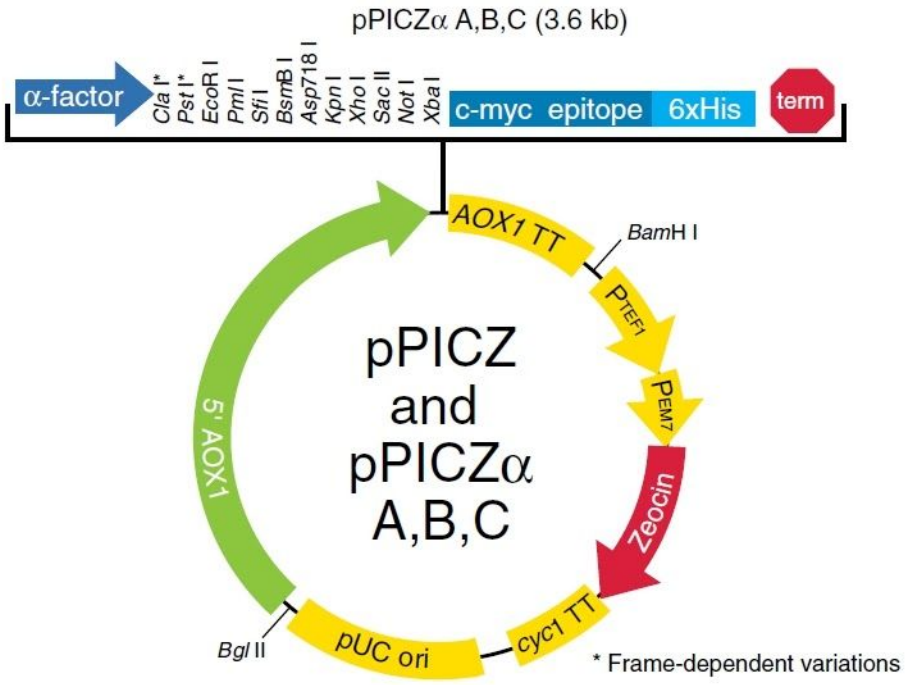

Figure 15. Vector map of pPICZ vectors (Invitrogen, n.d.).

In order to express protein with their native N-terminus, the XhoI restriction site at the end of the α-factor must be used to clone the gene flushed with the Kex2 cleavage site (figure 16) (Invitrogen, 2010a). Kex2 protease cleaves after the amino acids Glu-Lys-Arg and Glu-Ala repeats following the sequence may improve the efficiency of the enzyme. A proline directly after the arginine in Glu-Lys-Arg sequence will inhibit the Kex2 protease. The Zeocin resistance marker allows direct selection and can be used in *E. coli* thus eliminating the need for a second antibiotic maker which reduces vector size (Invitrogen, n. d). The 5′ *AOX1* integrates the target gene into the *Pichia* host genome by homologous recombination.

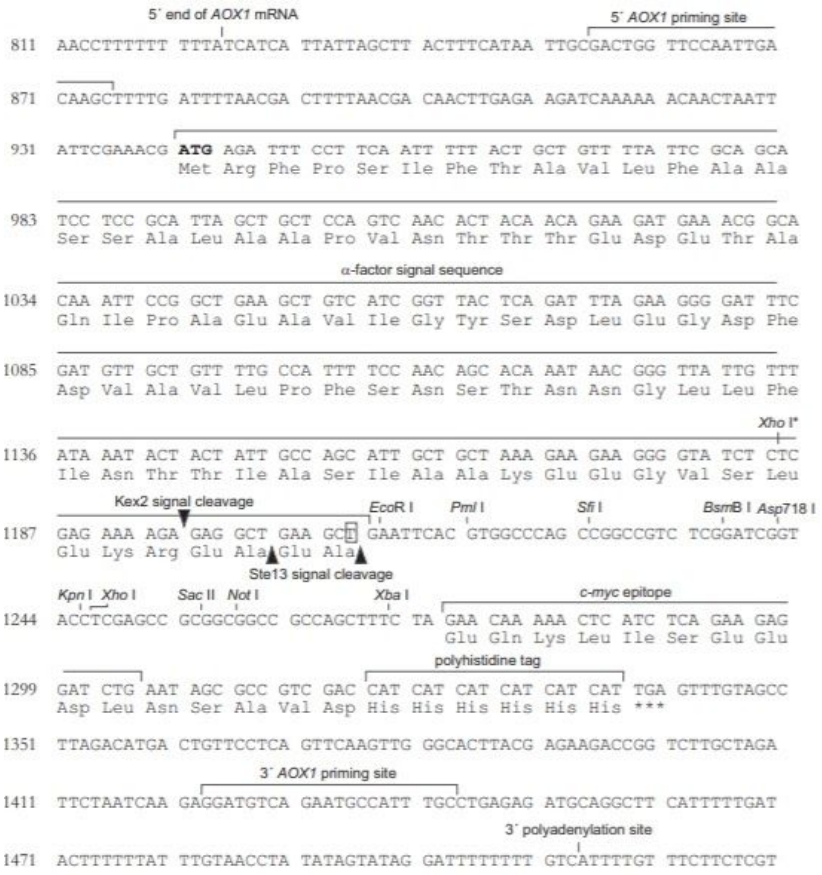

Figure 16. Multiple cloning sites for pPICZα A (Invitrogen, 2010a).

3.3. Gene optimizations for expression systems

The coding region for the eight chosen enzymes were optimized for the different expression systems used. For *E. coli* these were: cytoplasmic expression with pET22b+, periplasm expression with pET22b+ using pelB, extracellular expression with pAES40 utilizing YebF and for *P. pastoris*, extracellular expression with pPICZα using α-factor secretion signal. The results of the optimizations can be viewed in table 5. The optimization for intracellular expression in *E. coli* is not included in table 5 as it is identical to that of periplasmic expression. The restriction sites chosen for the insertion of the genes did not alter the amino acid sequences for the eight proteins however some extra residues were added to the terminal ends. The mean value of the base pairs changed with optimization was 24.8%. Full DNA sequences are found in Appendix A.

Table 5. Results of gene optimization from GenScript's OptimumGene Gene Design system. ColHol denotes the collagenase gene from G. hollisae.

Organism, vector and signal peptide	Gene	Restriction sites	Bp changes (%)
E. coli pAES40 YebF	ColG	KpnI, SacI	26.4
	ColH	KpnI, SacI	25.8
	ColHol	KpnI, SacI	22.1
	ColA	KpnI, SacI	24.5
	Sub	KpnI, SacI	26.7
	Pep	KpnI, SacI	23.4
	Trp	KpnI, SacI	25.0
	Pap	KpnI, SacI	26.0
E. coli pET22b+ pelB	ColG	NcoI, XhoI	26.9
	ColH	NcoI, XhoI	26.5
	ColHol	NcoI, XhoI	21.3
	ColA	NcoI, XhoI	24.8
	Sub	NcoI, XhoI	24.9
	Pep	NcoI, XhoI	25.1
	Trp	NcoI, XhoI	24.7
	Pap	NcoI, XhoI	25.4
P. pastoris pPICZα A α-factor	ColG	XhoI, NotI	23.4
	ColH	XhoI, XbaI	24.2
	ColHol	XhoI, NotI	23.9
	ColA	XhoI, NotI	22.8
	Sub	XhoI, NotI	24.9
	Pep	XhoI, XbaI	25.4
	Trp	XhoI, XbaI	28.2
	Pap	XhoI, XbaI	23.4

To get a better view of the GC content before and after codon optimization the results were plotted (figure 17). The GC content for *E. coli* is the mean values for using pET22b+ and pAES40 as these were almost identical. It was observed that the GC content in the optimized yeast genes were lower than for optimization in bacteria. In general, the GC content after optimization was around 50%.

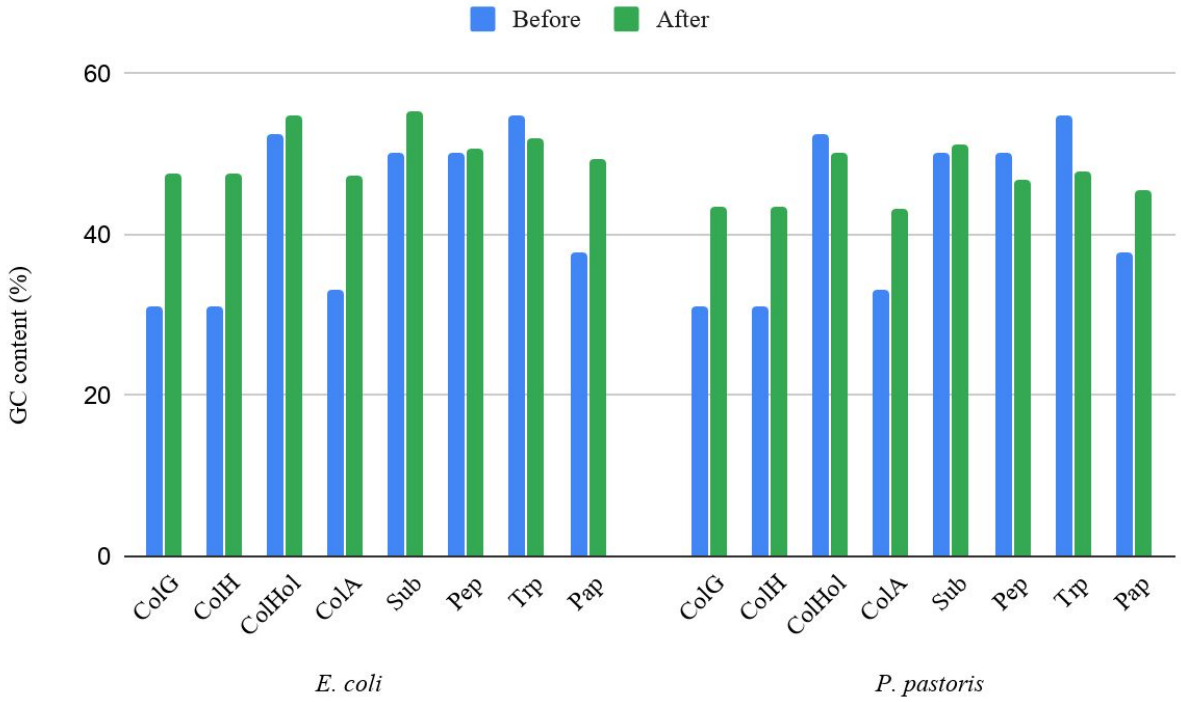

Figure 17. GC content before and after gene optimization for E. coli and P. pastoris.

To evaluate the differences between different tools for codon optimization, the *G. hollisae* collagenase and subtilisin Carlsberg gene were optimized with GenScript's OptimumGene Gene Design and Thermo Fisher Scientific's GeneOptimizer algorithm. The same DNA sequence was used for optimization in *E. coli* with restriction sites NcoI and XhoI. The optimization for collagenase showed 81.3% identity and subtilisin 75.1% identity. In order to verify the difference in codon utilization between the two algorithms, the two optimized subtilisin sequences were analyzed with Sequence Manipulation Suite: Codon Usage by bioinformatics.org (figure 18).

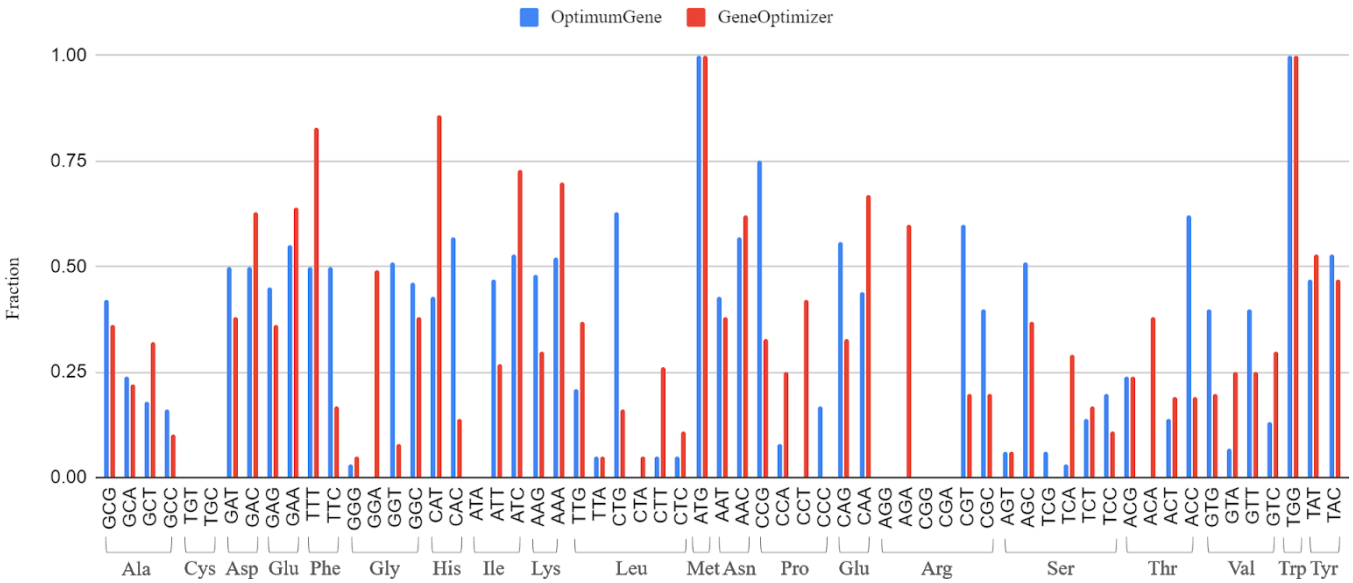

Figure 18. Comparison of codon usage for optimized subtilisin gene from OptimumGene and GeneOptimizer algorithm.

3.4. Cloning strategies

Below, strategies for cloning and expressing recombinant enzymes are presented. For detailed protocols, see "Cloning experiments", appendices B and C.

3.4.1. Protocol for cloning in *E. coli* with expression vectors

The cloning strategy for expressing enzymes in *E. coli* was adapted to be used either with pET22b+ or pAES40 vectors. Following manuals have been used to develop the cloning strategy: Addgene, 2018; BIO-RAD, 2011 Biotium, 2018; Invitrogen, 2002; Invitrogen, 2012a; Invitrogen, 2012b; Invitrogen, 2016; Jagannadha & Panda, 1996; Laemmli, 1970; Novagen, 1991; Omega BIO-TEK, 2019a; Omega BIO-TEK, 2019b; Thermo Scientific, 2015; Thermo Scientific, 2019. The final locations for the produced proteins are the cytoplasm, the periplasm and the culture media. The first step is subcloning in which the genes of interest and the expression vector are cleaved with restriction enzymes, followed by separation on agarose gel and extraction with gel extraction kit. The extracted DNA is then cloned into an expression vector in a 3:1 DNA insert to vector ratio. Transformation can be performed using chemical competent cells with heat shock. Cells with genes of interest as well as cells transformed with

self-ligated pET22b+ vectors are then grown on agar medium containing ampicillin. Depending on the number of clones of the plate with self-ligated vectors, a number of clones from each agar plate should be transferred to agar masterplates with sterile pipette tips and what's left on the pipettes to PCR tubes containing PCR mastermix with vector specific primers for colony PCR. Linearized vectors (cut with appropriate restriction enzyme) amplified with PCR should be run on agarose gel to validate positive clones.

To verify that the DNA sequence of the positive clones are correct plasmids should be extracted from cells on the masterplate and sent for sequencing. Positive and correctly expressed clones are then inoculated in a smaller volume till an OD600 of 2 is reached and transferred to a larger volume to an OD600 of 0.4 is achieved. Protein expression is induced with IPTG. Depending on the vector used, the enzymes should be expressed in different subcellular locations and to evaluate the effectiveness of the cloning, all subcellular locations should be evaluated for the presence of protein. Medium fraction that is concentrated or not is used for analyzing extracellular protein production. To evaluate periplasm protein production either osmotic chock or chloroform treatment can be utilized. In order to break the cytoplasm cell membrane sonication can be used. As all the expression vectors contain polyhistidine adhesion tags, proteins can be purified using metallic beads or HisTrap columns.

3.4.2. Protocol for cloning *P. pastoris* with pPICZα vector

The cloning strategy for expressing enzymes in *P. pastoris* was adapted from Invitrogen user manuals utilizing the yeast (2010a; 2010b). The first step in the process is propagation of the expression vector by transforming into recombination and endonuclease A deficient *E. coli* strain such as DH5α or JM109 and plating on LB agar plates with Zeocin. Bacterial glycerol stock should be prepared for storage of the vectors. The second step is subcloning of the genes of interest into the pPICZα vector using restriction enzymes, agarose gel, DNA extraction and ligation. The ligation mixture is then transformed into *E. coli* in this same way as described in propagation of the expression vector and plated. The transformants need to be evaluated for the presence of insert DNA with colony PCR and sequencing. Commercial 3′ *AOX1 Pichia* primer

and the 5′ *AOX1 Pichia* primer could be used and glycerol stocks should be made of the positive clones. Plasmids are extracted and linearized using a restriction enzyme, followed by heat inactivation of the restriction enzymes. The linearized plasmids are purified by phenol/chloroform extraction and ethanol precipitation.

The plasmids are integrated using the *AOX1* locus with *Pichia* strains such as Mut^+ (containing the *AOX1* gene and can metabolize methanol) X-33, Gs115, SMD1168H or Mut^S KM71H with either electroporation or chemical methods for transformation. Different media could be used for expression of recombinant protein for example buffered complex glycerol or methanol medium, which allows a wider pH range to optimize the protein production, decreases or prevents proteolysis of secreted protein and allows better growth and biomass accumulation. It also requires high aeration and a temperature of 30 °C. Mut^+ strains should be grown to OD600 of 1 and Mut^S strains to OD600 of 2-6. Methanol is used for protein induction and samples taken for Mut^+ strain at 0, 6, 12 and every 12 hours to 96 hours after induction and for Mut^S strain at every 24 hours (for 7 days) to analyze expression levels and determine the optimal time to harvest. Protein expression could be analyzed with SDS-PAGE, western blot with anti-his or anti-myc antibodies or functional assay and purified using the polyhistidine tag.

3.5. Analysis of enzymes

The following section describes how the results of the cloning experiments could be evaluated and how enzymes expressed should be tested. A total of eight enzymes in four systems, 32 enzymes can be expressed. The different steps included in these analyses are determination of enzyme concentration, molecular weight of enzymes, collagenolytic activity, protease activity, degree of hydrolysis and molecular weight of peptides. The plan of the analysis procedures is to first measure the concentration of protein before determining the molecular weight. Only the samples which have shown accurate molecular weight should be tested further on. For the collagenases and subtilisin, their collagenolytic activity should be assessed and only those that show breakdown of collagen should be assessed further. After this the protease activity and proteolytic products should be evaluated.

As collagenases digests collagen at Gly-X-Y bonds, they will not completely degrade collagen. Depending on the size of the hydrolysate that is desired, collagenases can be combined with the other proteases: trypsin, pepsin, papain and subtilisin, to achieve further degradation. Each collagenase can be tested with all combinations of trypsin, pepsin, papain or subtilisin according to the following equation: $x * \sum_{i=1}^{n} \frac{n!}{i!(n-i)!} = y$, where x is the number of collagenases, n is the number of other enzymes and y is the number of combinations. If there are four collagenases and four other enzymes there would be sixty combinations. However, it is also possible to test other combinations for example collagenase with collagenase and subtilisin with an enzyme that is not collagenase. That would lead to even more combinations. It is however difficult to do a detailed design of the experiment without having any knowledge of the results from the cloning experiments.

3.5.1. Concentration of enzymes

Before the enzymes would be analyzed it would be useful to determine the amount of protein that has been purified. A common convenient technique for this is Bradford protein assay (Ernst & Zor, 2010). Protein binds to coomassie brilliant blue dye under acidic conditions which induces a color shift from brown to blue, which is measured with absorbance at 595 nm. It is also possible to measure the absorbance at 280 nm to determine the concentration of the protein if the absorptivity value is known or standard curve is obtained.

3.5.2. Molecular weight of enzymes

The second step in analyzing the cloning results should be a size determination of the expressed proteins. One technique for doing this is SDS-PAGE which is simple, rapid, inexpensive and usually estimates molecular weight accurately within ± 10% (Matsumoto, Haniu & Komori, 2018). An alternative to SDS-PAGE is gel-filtration chromatography where a calibration curve of the protein is made and the elution volume of the protein is linearly correlated to the logarithm of its molecular weight (Fitzgerald, Leonard, Darcy, Sharma, & O'Kennedy, 2017).

3.5.3. Collagenolytic activity

A first evaluation of the enzyme's ability to degrade collagen could be tested with collagen zymogram which is one of the most widely used techniques for detecting proteolytic activity that allows for detection of active form (Snoek-van Beurden & Von den Hoff, 2005; Vandooren, Geurts, Martens, Van den Steen & Opdenakker, 2013). For this analysis, only the collagenases and subtilisin with documented collagen degrading ability should be tested. Activities of enzymes can be detected using collagen zymogram gels copolymerized with type I collagen sensitive enough to detect amounts down to 1 ng of protein (Inanc, Keles, & Oktay, 2017). Another reliable method for quantifying collagen degradation and enzyme activity is fluorescent detection of collagen cleavage using O-phthaldialdehyde (OPA) bound to the primary amine where fluorescence is measured using a microplate reader (Go, Horikawa, Garcia & Villarreal, 2008). The benefits of the method are high solubility of OPA in water, fast reactivity with a total experiment time of around 1 hour and low background fluorescence of OPA which yields good signal-to-noise ratio. Furthermore, the kinetic parameters of collagenases can be determined using synthetic peptide FALGPA with similar structure of collagen or other substrates (Teramura et al., 2011).

3.5.4. Protease activity

To measure and compare the activity of all proteases a non-specific protease activity assay developed by Merck using casein as a substrate can be used (Cupp-Enyard, 2008). When proteases digest casein tyrosine is released, Folin's phenol reagent reacts with free tyrosine and produces a blue colored chromophore related to the activity of enzyme which absorbance can be measured and quantified with spectrophotometry. Absorbance values are compared to a standard curve and activities can be determined in Units (amount micromoles of tyrosine released from casein per minute). Another homogeneous method for proteolytic activity determination is protease fluorescence resonance energy transfer (pro-FRET) assay that is based on cleavage of FRET peptide substrate which increases donor fluorescence (Zauner, Berger-Hoffmann, Müller, Hoffmann & Zuchner, 2011). Other alternatives are heterogeneous assays which require immobilisation of protease substrates on a solid surface which can interact with proteolytic

enzymes and produce different signals (Ong & Yang, 2017). Such signals can be detected with electrochemical sensors, surface plasmon resonance (SPR) sensors, enzyme-linked peptide protease assays and liquid crystal-(LC)-based assays. It is also possible to use zymograms with gelatin to check the protease activity (Tanaka et al., 2018).

3.5.5. Degree of hydrolysis

There are many different methods for determination of the degree of hydrolysis, which is defined as the percentage of cleaved peptide bonds in a protein hydrolysate. It can be determined by measuring the amount of nitrogen released during hydrolysis after precipitation with trichloroacetic acid, and the Kjeldahl method could be used for determining the total nitrogen content (Montoya et al., 2008; Spellman, McEvoy, O'cuinn & FitzGerald, 2003). Formol titration could also be used to measure the free amino groups (Nilsang, Lertsiri, Suphantharika & Assavanig, 2005; Pintado, Pintado, & Malcata, 1999). Osmometry is another option and the pH-stat method is possibly the most commonly used method for determination of the degree of hydrolysis (Rutherfurd, 2010). It is also possible to do spectrophotometric assays with different compounds such as, trinitrobenzenesulfonic acid (TNBS) (Adler-Nissen, 1979), OPA (Church, Porter, Catignani & Swaisgood, 1985; Nielsen, Petersen & Dambmann, 2001) and ninhydrin (Pearce, Karahalios & Friedman, 1988) which reacts with the primary amino groups.

In the TNBS method by Adler-Nissed (1979), the protein hydrolysate is dissolved in hot sodium dodecyl sulfate. A sample is then mixed with sodium phosphate buffer and TNBS, followed by incubation in the dark for 60 min at 50 °C. Absorbance is read at 340 nm after the reaction is quenched by addition of HCl. L-leucine solution is used as the standard. The OPA method by Nielsen et al. (2001) is more accurate, easier and faster to perform and more environmentally friendly than the TNBS method. In the OPA method, the OPA reagent (borax, SDS, OPA, DTT, ethanol and deionized water) is mixed with a diluted sample and incubated for two minutes. The absorbance is read at 340 nm and a serine solution is used as the standard. However, cysteine and OPA react weakly and form an unstable product (Chen, Scott & Trepman, 1979).

3.5.6. Molecular weight of peptides

It is possible to use size exclusion chromatography to determine the molecular weight distribution of the peptides (Khiari et al., 2014; Silvestre, Hamon & Yvon, 1994). Mass spectrometry could also be used to determine the molecular weight of the peptides.

3.6. Cloning experiments

The design of the experiment can be viewed in figure 19. First, the plasmids were cleaved with restriction enzymes. After that, the cut DNA was run on agarose gel for separation and extracted with an extraction kit. Then the DNA was ligated and transformed into *E. coli* Bl21(DE3). The transformed cells were plated and grown overnight. The experiments were not continued after this step.

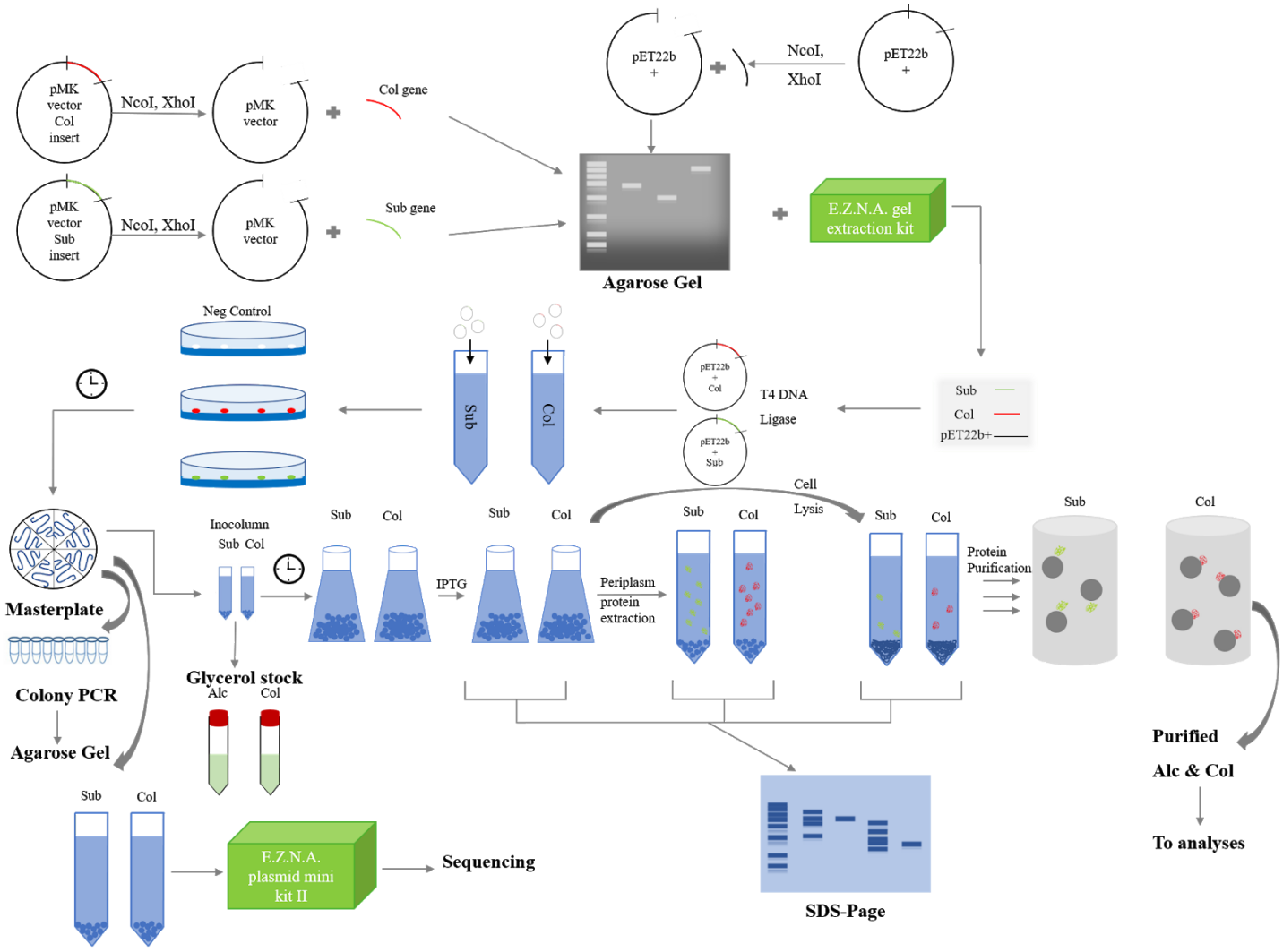

Figure 19. Process design for production of enzymes using E. coli bacteria. Firstly, plasmids are digested with restriction enzymes, NcoI and XhoI. After, the digested DNA is run on agarose gel for separation and extracted with E.Z.N.A gel extraction kit. The DNA is ligated using T4 DNA ligase and transformed into E. coli Bl21(DE3) and incubated while shaking to grow. The transformed cells are then plated and grown overnight. Depending on the number of colonies growing on the control plate, a number of colonies from each plate are streaked on a masterplate and what is left on the pipette tip is resuspended in PCR tubes for colony PCR. PCR samples are run on agarose gel and positive clones are sent for sequencing. Positive clones are inoculated in LB media. Parts of the cell culture are grown in a bigger volume with addition of IPTG to induce protein expression and the rest is made into glycerol stock. Cells are treated with osmotic shock to break outer cell membranes for periplasm protein recovery and with sonication for cell lysis. Extracellular, periplasmic and total cell fraction supernatants are run on SDS-PAGE to confirm presence of protein. The protein is purified using metal beads.

3.6.1. Results and Discussion

The results of the codon optimization using GeneOptimizer algorithm via Instant Designer Gene Synthesis Portal service is shown in figure 20. The plots displayed the quality of the used codon at the specific codon position and it was observed that there was only one codon that had codon quality under fifty (for optimized subtilisin). This indicates that the optimized sequence had substantially higher codon quality. Pairwise sequence alignment showed that 19.7% and 25.5% of the nucleotides were exchanged for collagenase respectively subtilisin.

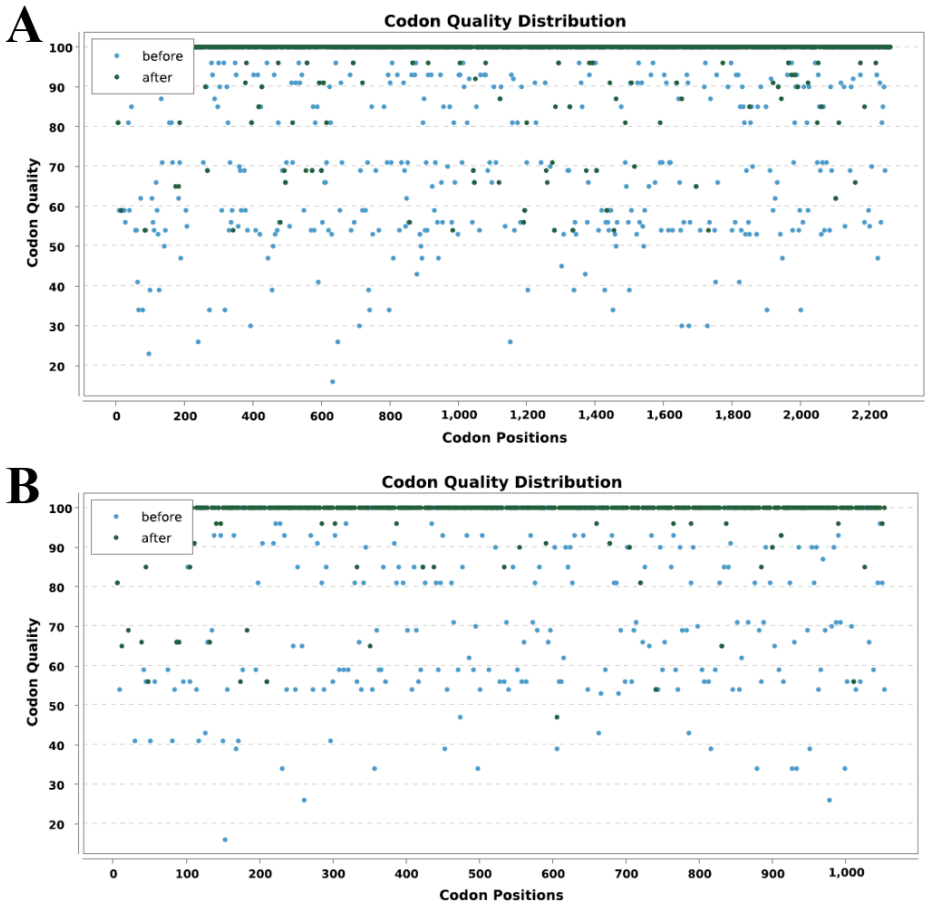

Figure 20. Codon optimization by GeneOptimizer. Graph A shows optimization for collagenase and B for subtilisin.

1 µg of collagenase-pMK, subtilisin-pMK and pET22b+ digested with restriction enzymes NcoI and XhoI and run on agarose gel are shown in figure 23, Gel A. In well number 2, with digested collagenase-pMK vector, 3 bands were present. One band was located at around 2500 bp

followed by two bands closely to each other between 1100 and 1400 bp. In the third well with the digested subtilisin vector only 2 bands were present, also between 1100 and 1400. As the two bands were consistent it was postulated that the pMK vector must have contained another restriction site. This was confirmed after reviewing the received DNA data for the plasmids. The sequence for subtilisin-pMK plasmid is presented in figure 21 with NcoI restriction site highlighted in red and XhoI in green. This gives rise to three fragments, Ncoi-XhoI: 1061 bp which is the subtilisin gene, XhoI-NcoI: 1234 bp and NcoI-NcoI: 1077 bp. Additionally, it was also observed that due to the size of the 1061 and 1077 bp gene fragments these are overlapping and impossible to separate on the gel.

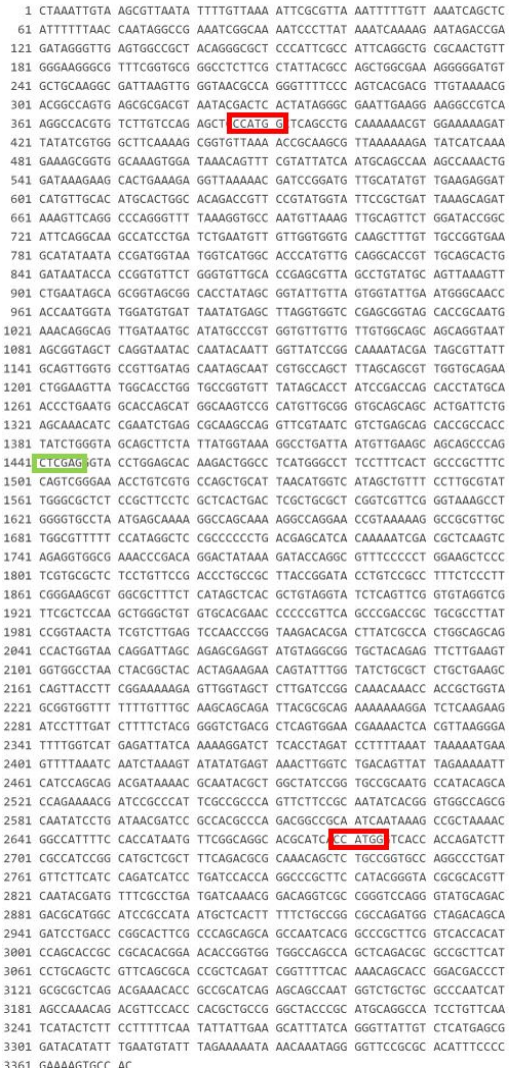

Figure 21. DNA of pMK vector with subtilisin gene. Red boxes mark the NcoI restriction sites and the green box marks the XhoI restriction site.

The collagenase, subtilisin and pET22b+ fragments were extracted from the gel, see gel B in figure 23. The concentration of eluted DNA using gel extraction kit was estimated with calculations and measured with nanodrop. The mass of the gene in the gel in table 6 was the theoretical maximum amount of DNA fragment calculated as the length of insert divided by plasmid length multiplied by the amount of DNA loaded.

As can be seen in table 6, the measured concentration of DNA was much higher than the estimated one and the mass of DNA larger than was originally used for the enzymatic digestion reaction. These values could be due to contamination. A_{260}/A_{280} measurements for collagenase, subtilisin and pET22b+ were 1.9, 1.7 and 2.3 whereas a pure DNA sample should have a value above 1.8. This indicated that at least the collagenase sample was rather pure. One might question the instrument's validity however such a grave error seems rather unlikely. Due to these results being discovered too late, the experiment was carried out as planned. The ligation protocol was used with the nanodrop measured concentration.

Table 6. The mass and concentration of DNA from digestion of plasmids to elution of DNA from extraction kit.

DNA in reaction	**Collagenase**	**Subtilisin**	**pET22b+**
DNA in digestion reaction (µg)	1.0	1.0	1.0
DNA loaded in gel (µg)	0.50	0.50	0.50
Mass of gene in gel (µg)	0.25	0.32	0.49
Theoretical DNA mass recovery from gel (85%) (µg)	0.21	0.27	0.42
Estimated concentration of eluted DNA (ng/µL)	7.0	8.9	14
Measured concentration of DNA (ng/µL)	38	41	40

After transformation and incubation overnight no colonies except for uncut plasmids were observed on plates. This could either imply that the ligation, transformation, agar plates or the competent cells were not working with the ligation mixture. The plates were checked by plating

E. coli DH5α transformed with pET22b+, which showed growth thus proving that the plates were not the issue.

As the values, both A_{260}/A_{280} and measured concentration for the pET22b+ vector seemed strange, the digested pET22b+ eluted from the gel extraction kit was run on agarose gel to verify that it was intact (see figure 23, gel C). Although the gel was faint it was observed that no DNA band around 5400 bp was present, suggesting that the digested plasmid could have been completely fragmented or not present at all due to unknown reasons.

To verify if the extraction kit worked the experiment was redone with concentrations of collagenase, subtilisin and pET22b+ measuring at 6.0, 5.5 and 6.5 ng/µL. These values were much closer to theoretical ones of 7.0, 8.5 and 14 ng/µL. However, the concentration of subtilisin should be higher than collagenase as it contains fragments from the pMK vector. Additionally, the A_{260}/A_{230} values which should be between 2.3-2.4 for pure DNA, were all low, 0.26, 0.11 and 0.18 (Koetsier & Cantor, 2019). This again suggested some form of contamination, possibly from the elution buffer containing guanidine thiocyanate which is a protein denaturant and has absorbance at 230 nm (von Ahlfen & Schlumpberger, 2010). It may be that the guanidine thiocyanate denatured the ligase in the ligation step. However, the purities ratio at lower concentrations (<20 ng/µL) are not reliable, especially the A260/A230 ratio (Koetsier & Cantor, 2019). The DNA fragments were ligated, transformed and plated and yet again, no colonies were observed after 2 days of incubation at 37°C.

The results questioned the advantages of purifying the DNA with gel extraction kit. A graphical representation of the outcomes of ligation with or without using the kit are presented in figure 22. The representation shows that, as it is not possible to separate the subtilisin-pMK fragments, running the sample on agarose gel and extracting the DNA does not have any sizable benefits.

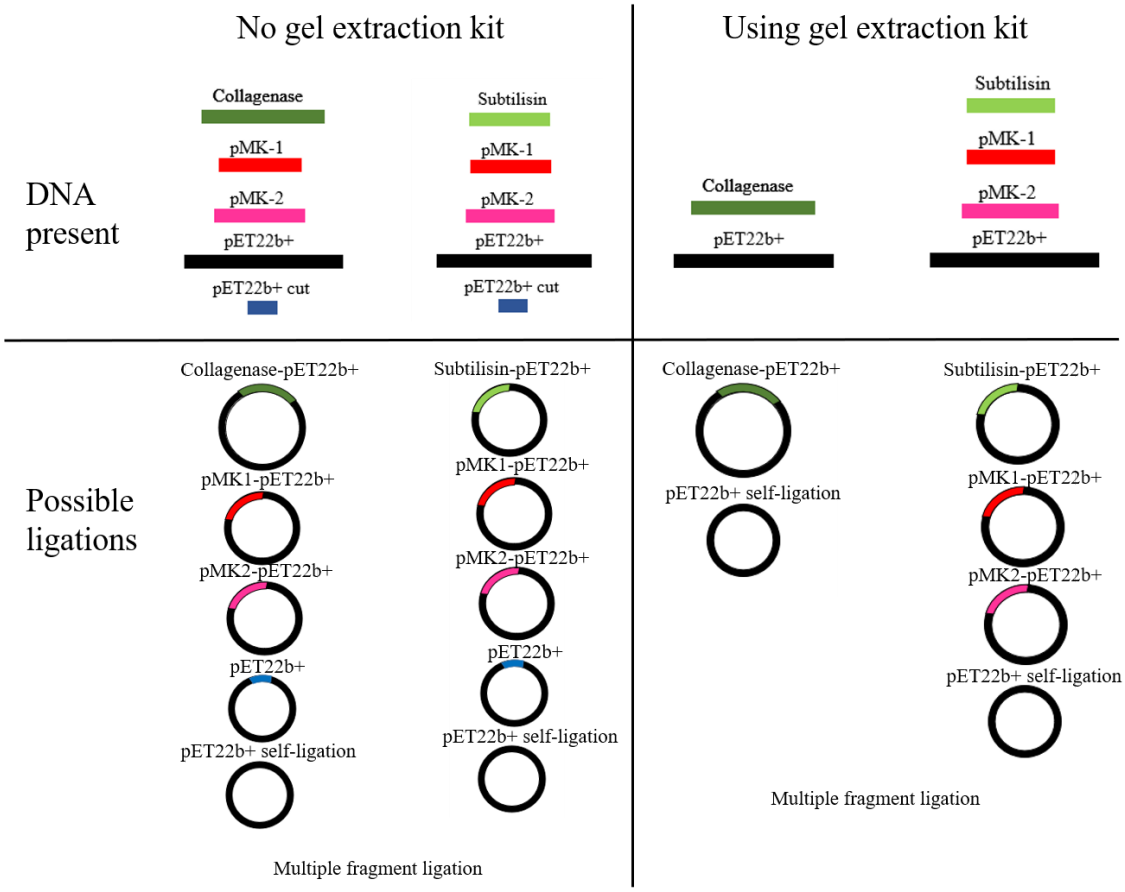

Figure 22. Possible ligated DNA of DNA fragments with or without running samples on agarose gel. The coiled vectors are multi-ligated plasmid DNA.

As the gel extraction kit had shown questionable results a new kit was used and 0.5 µg collagenase-pMK and pET22b+ were run on gel and extracted. Concentrations measured with nanodrop were 2.4 ng/µL and 5.4 ng/µL for collagenase and pET22b+ compared to calculated values of 3.5 and 7.0 ng/µL, respectively. This resulted in a sample recovery of 58% and 76% for collagenase and pET22b+ compared to the manufacturer's stated 85%. However, the accuracy of the 2.4 ng/µL measurement is questionable as it is outside the limit of quantification. Figure 23, gel D shows the digested collagenase and pET22b+ along with ligated collagenase-pET22b+ and subtilisin-pET22b+ from the previous experiment in well 4 and 5. No bands were observed suggesting either that too little DNA was present to be seen on the gel or that no ligation occurred.

As the ligation had not been successful, a ligase mastermix was exchanged for the ligase and ligase buffer used previously as it was suspected that the ligase buffer might be too old or the ligase inactive. To validate if this mastermix was active and if DNA not run on agarose gel and extracted with gel kit would ligate, a new enzymatic restriction reaction followed immediately by ligation was performed. The pET22b+ vector used in this experiment was isolated from *E. coli* DH5α due to prepared plasmid solution being out of stock. To investigate if a higher molar ratio would result in more successful ligation, the molar ratio of insert to vector of 7:1 was also tested. The plasmids would have been digested with XhoI before running on gel to get linearized DNA but due to this enzyme being out of stock the undigested plasmids were instead run on agarose gel. The results are seen in figure 23, gel E. In wells 2-5 are: col-pET22b+ 7:1 (mole ratio), col-pET22b+ 3:1, sub-pET22b+ 7:1 and sub-pET22b+ 3:1. The bands were very smeared and could either indicate nicked DNA or multiple fragment ligation. Well 6 and 7 showed self-ligated respectively undigested pET22b+ vectors. The undigested vector travelled further than previous linearized vectors had shown, which is to be expected as supercoiled vectors take up less space and experience less resistance from the gel hence why appears to travel faster. The nicked vector is also present above the supercoiled plasmid and appears to travel slower as its relaxed or open-state confirmation increases the size of the DNA molecule. These results indicated that the ligase mastermix was active and could be further used for ligation experiments.

The practical work could not be continued after these experiments due to the project being terminated. If laboratory tests would have been allowed to continue some measures could have been considered to successfully produce the enzymes. Firstly, the competent cells would have been evaluated since their storage time was unknown, which could affect the competence. This could be done with a control plasmid. New competent cells were produced but were not able to be tested. If the cells were not successfully transformed with ligated vectors, this could be an additional reason why no colonies were observed on agar plates. Secondly, if possible, additional restriction enzymes that would cut the pMK plasmid but not within the gene could have been used to avoid the issue of overlapping DNA bands on agarose gel. Thirdly, insert-specific primers could have been used to amplify the subtilisin gene in order to again avoid overlapping

DNA fragments. Lastly, a new subtilisin gene could have been ordered, either as geneblocks or inserted in different plasmids not containing more copies of the chosen restriction sites.

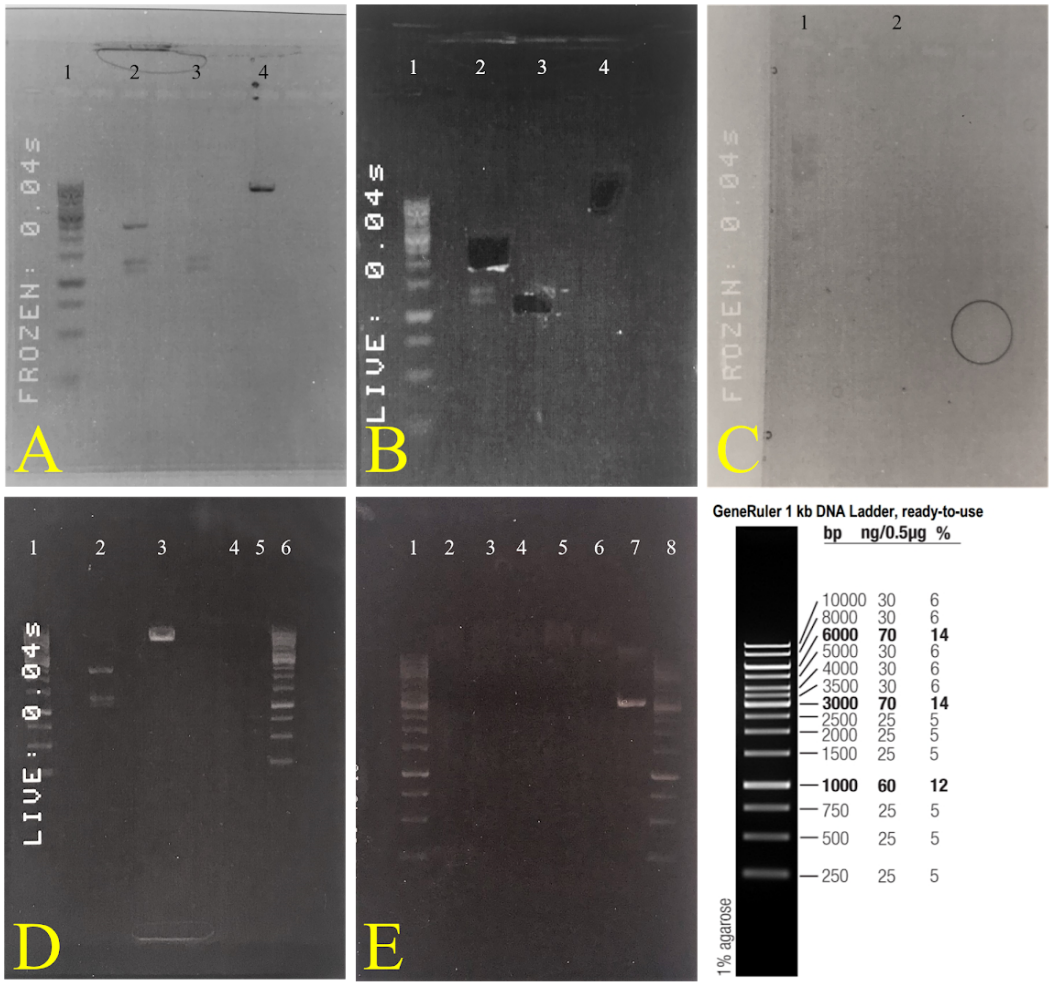

Figure 23. Agarose gels with digested or ligated DNA. Gel A shows: well 1: generuler, 2: digested collagenase-pMK vector, 3: digested subtilisin-pMK vector and 4: digested pET22b+. Gel B shows DNA bands cut out from Gel A. In well 2 the collagenase insert has been cut out, in well 3 all DNA bands have been extracted and in well 4 the pET22b+ vector. Gel C shows: well 1: generuler and 2: digested pET22b+. Gel D shows: well 1: generuler, 2: digested collagenase-pMK vector, 3: digested pET22b+, 4: ligated collagenase-pET22b+ vector, 5: ligated subtilisin-pET22b+ vector and 6: generuler. Gel E shows: well 1: generuler, 2: ligated digested collagenase-pMK and pET22b+ fmol ratio 7:1, 3: ligated digested collagenase-pMK and pET22b+ fmol ratio 3:1, 4: ligated digested subtilisin-pMK and vector fmol ratio 7:1, 5: ligated digested subtilisin-pMK and pET22b+ molecular ratio 3:1, 6: pET22b+ self ligation, 7: pET22b+ vector and 8: generuler. Generuler 1 kb DNA ladder used in all agarose gels is shown to the bottom right corner.

4. Discussion

The benefits with collagenase ColG and ColH are that they are well studied. For example, the structures are known. Collagenase ColA and *G. hollisae* collagenase are less studied but have been shown to be promising when it comes to their ability to degrade collagen. Subtilisin Carlsberg has broad specificity and has also been shown to degrade collagen well. However, the gene used in this work is the wildtype and the commercial Alcalase has been genetically improved. Something to consider is if it is worth producing and possibly mutating subtilisin and how much it differs from the commercial subtilisin. The remaining three enzymes; trypsin, pepsin and papain are commonly used enzymes. They might not be able to degrade untreated collagen, but collagen treated with an enzyme like collagenase or other treatments like acid can be further degraded by the enzymes. In order to increase the collagen degradation efficiency of the enzymes one option could be to use gene technology to optimize them. One alternative could be to do DNA shuffling with the collagenases to generate favorable mutants. To increase the stability of the enzymes, residues can be mutated with cysteine to induce formation of disulphide bridges however, these might not be expressed successfully in the reducing environment in the cytoplasm.

When choosing an expression system there are many things to consider. All chosen collagenases have a bacterial origin and all the other enzymes have been successfully expressed in *E. coli*. Therefore, it would be a good idea to begin with *E. coli*. It is also the simplest, cheapest and fastest system. These are the reasons why *E. coli* was chosen as the first organism and used in the experiments. Yeast was the second organism chosen because it is capable of post-translational modification, relatively simple to use and capable of secretion of the recombinant protein. The vector used for *pastoris* is a YIp, the strategy written is for a single integration but a way to further increase protein yield could be to integrate multiple expression cassettes. It could be wise to investigate other strains of bacteria as host cells to take advantage of fast and inexpensive cell growth but possess the ability to secrete protein for simpler purification such as *B. subtilis.*

The subcellular location for expression of the recombinant protein depends on different factors like the amount of enzymes needed and the purification of the protein. For the simplest purification of the recombinant protein it should be expressed to be secreted to the medium. In large scale, secretion to the medium would simplify the production. Downstream processing is also less complicated for expression in the periplasm compared to expression in the cytoplasm. However, there is of course a risk that it is not possible to express the protein in periplasm and medium. A benefit with cytoplasmic expression is that it gives high yield however a high yield often results in inclusion bodies. However, secretion of the enzyme results in lower yield and probably no formation of inclusion bodies. It has also been noted when using commercial vectors, it is not seldom that the protein gene gets extra amino acids on the terminals when using signal peptides or purification tags.

When optimizing the genes for cell hosts it has become clear that vectors with a limited number of restriction sites in the cloning site pose difficulties when cloning all enzymes using the same vector. Additionally, in order to utilize signal peptides and purification tags it is often required to alter the sequences. The strategy when designing the vectors was to minimize the distance between the gene of interest and the tags. This sometimes leads to either the first or last codon having to be exchanged or for another amino acid to be inserted. In order to make perfect subcloning different vectors could have been used or modifications after the subcloning to remove the extra amino acids. However, it was discussed that altering the genes slightly at the ends probably would not alter the activity or action of the enzymes substantially. It was also realized that histidine tags are not always optimum to use for purification. If the C- and/or N-terminus are changed during post-translational modifications, there is sometimes no optimum position for the affinity tag. Therefore, exploring other ways to aid purification could be an option.

When comparing genes optimized in *E. coli* and *P. pastoris* it was observed that for all enzymes, GC content was lower for the optimized yeast genes compared to *E. coli* (figure 17). *G. hollisae* collagenase and trypsin originally has a high GC content but after optimization it was decreased

in yeast and increased in bacteria. This indicated that there is a difference in overall GC content after optimization, which shows the difference in codon bias between the species. Lastly, codon optimization for *G. hollisae* collagenase and subtilisin Carlsberg with the two different algorithms used was compared and showed that even if the same sequence, same restriction sites and same organism is used, the sequence identity and codon usage are different. The algorithms optimized the DNA sequences according to many factors other than codon usage, which is why it was not surprising that the results differ. It is not easy to say which of the two used algorithms gives the best optimization. In order to verify if there was a difference both sequences would have to be cloned, expressed and the yield compared.

There are many ways in which organisms can be used to produce recombinant enzymes. The strategy developed for *E. coli* was a combination of many different protocols and henceforth not optimized or tested before. The results from the cloning experiments showed that enzymatic digestion and ligation of DNA is not always straightforward. Therefore, much consideration should go into designing the genes and what format they should be obtained in, either inserted in plasmids or as gene fragments but the latter might only be available for shorter sequences. It might also be a good idea to verify that ligation has been successful before transforming the cells. This could be done on agarose gel with preferable linearized DNA to verify size. However, the amount of ligated DNA needs to be enough to be visualized on gel. Transformation is a sensitive procedure, as it must be timed and with using the correct temperature as well as the cells being competent. One way to check the competence of the cells is by using a control plasmid. There are many ways to obtain the medium, periplasmic and cytoplasmic fraction. The ones that were investigated in this report all have documented efficiency.

To analyze which enzyme breaks down collagen best is rather straightforward but to determine what combination of enzymes would degrade collagen furthest would be very time consuming. It might be possible to do it theoretically by evaluating the amino acid sequence of collagen and predicting where the different enzymes could cut and analyzing the possible products. This however might not reflect the actions of expressed enzymes. Moreover, it is not possible to

foretell which enzymes would be successfully expressed. Furthermore, in order to test every possible combination of enzymes hundreds of tests would need to be conducted. To test all collagenases on their own with all combinations of other enzymes would result in sixty experiments. This is obviously not feasible. Instead the design should be planned so that as much information can be gathered from as little experiments as possible. Firstly, the collagenases and subtilisin should be evaluated individually. Secondly, an option could be to combine each collagenase (presuming they were successfully expressed) with subtilisin which has documented collagenolytic activity. The most successful combination could then be tested with the remaining three enzymes to evaluate which options would result in the most degraded collagen. This would greatly reduce the number of tests to a maximum of 16 presuming all collagenases are successfully expressed (figure 24). However, this design assumes that subtilisin is more effective than the three remaining proteases.

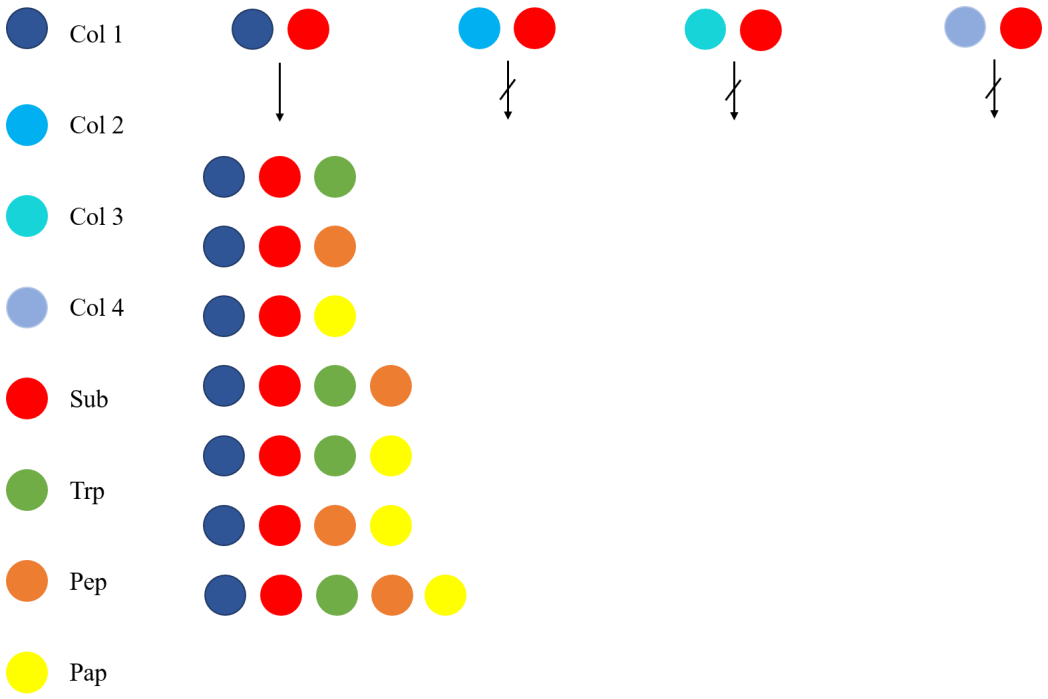

Figure 24. Possible design of experiment for optimization of collagen degradation. To the left, all enzymes and respective colours are shown. The design shows one collagenase-subtilisin combination being further tested with all combinations of trypsin, pepsin and papain.

As for cloning strategy there are many techniques available for analyzing the enzymes produced. A well-founded place to start is to determine the concentration of purified enzymes, for example to be able to calculate what amount to load on SDS-PAGE gels. Bradford assay is a very standard and straightforward technique and seems like a great choice to determine the concentration. To measure the molecular weight of recombinant both SDS-PAGE and gel filtration are rather simple and rapid techniques with a benefit for gel filtration in that samples do not need to be heated and denatured before analysis but more expensive equipment is required. Collagen zymogram which is commercially available is a simple assay to check the collagengenylotic activity. To get a quantitative estimation, assays utilizing FITC-labelled type I collagen or FALGPA can be used. There are abundant amounts of techniques to measure proteolytic activity, both who utilizes differences in fluorescence and electrochemistry. The methods that would be chosen would depend on what equipment is available, what would be most economical, the results desired and their accuracy. The same things can be said for determination of degree of hydrolysis.

Collagen can be found in animal skin and bones that have low value. With the enzymes examined in this report it could be possible to convert the collagen into products with a high value: hydrolysate, amino acids or peptides which have an evolving market value.

5. Conclusion

The aims of this thesis were to develop cloning strategies for production of collagenolytic enzymes, explore their actions and to design experiments to compare their efficiency. Two different cloning strategies for expression of eight enzymes for degradation of collagen have been developed. One strategy with two enzymes was evaluated experimentally. The expression of the two enzymes was not successful due to an unexpected shortage of time. It has also been noted that cloning is not an easy task and that there are many sensitive procedures and steps.

Due to the rigid structure of collagen it makes it less susceptible to enzymatic degradation and only a few enzymes have been shown to effectively break down the structural protein. With a growing market, it has been shown that collagen hydrolysates have multiple applications and effective methods for its breakdown desirable. Expression systems with *E. coli* are simple and frequently used but cannot perform post-translational modifications possible in organisms such as *P. pastoris* which also have a high yield of recombinant protein.

Cloning genes into expression vectors might alter the sequences at the terminal ends if signal peptides or tags are used since the gene needs to be in the same reading frame. Gene optimization with different tools gave different results but the genes need to be expressed to see if there are any differences. With eight different enzymes there are abundant ways to combine them and evaluate the collagen degrading ability but in order to make the study feasible, the number of tests should be optimized to obtain as much information as possible.

6. References

Abfalter, C., Schmidt, T., & Wessler, S. (2015). Proteolytic Activities Expressed by Gastrointestinal Pathogens Bacillus cereus, Listeria monocytogenes and Enterococcus faecium in Different Growth Phases. *British Microbiology Research Journal*, *7*(2), 62–70. doi:10.9734/bmrj/2015/16402

Abfalter, C. M., Schönauer, E., Ponnuraj, K., Huemer, M., Gadermaier, G., Regl, C., … Wessler, S. (2016). Cloning, Purification and Characterization of the Collagenase ColA Expressed by Bacillus cereus ATCC 14579. *Plos One*, *11*(9). doi:10.1371/journal.pone.0162433

ACES (n. d.). *YebF protein export kit* [User manual]. Retrieved from https://fnkprddata.blob.core.windows.net/domestic/data/datasheet/AES/0148-3.pdf

Addgene (2018). *Agarose Gel Electrophoresis* [User manual]. Retrieved from https://www.addgene.org/protocols/gel-electrophoresis/?gclid=Cj0KCQjwy6T1BRDXARIsAIqCTXoKI6uAFkl50GZjd5jdhqK9s5--nPuZvkyg9m3c5RLS24uH61EIFmQaAr-uEALw_wcB

Adler-Nissen, J. (1979). Determination of the degree of hydrolysis of food protein hydrolysates by trinitrobenzenesulfonic acid. *Journal of agricultural and food chemistry*, *27*(6), 1256-1262. doi:10.1021/jf60226a042

Ahmad, M., Hirz, M., Pichler, H., & Schwab, H. (2014). Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. *Applied microbiology and biotechnology*, *98*(12), 5301-5317. doi:10.1007/s00253-014-5732-5

Antonov, V., Ginodman, L., Kapitannikov, Y., Barshevskaya, T., Gurova, A., & Rumsh, L. (1978). Mechanism of pepsin catalysis: General base catalysis by the active-site carboxylate ion. *FEBS Letters, 88*(1), 87-90. doi:10.1016/0014-5793(78)80613-9

Arya, R., Bhattacharya, A., & Saini, K. S. (2008). Dictyostelium discoideum—a promising expression system for the production of eukaryotic proteins. *The FASEB Journal*, *22*(12), 4055-4066. doi:10.1096/fj.08-110544

Baird, T. T. and Craik, C. S. (2013). Chapter 575 - Trypsin. In Rawlings, N. D., & Salvesen, G. (Eds.), *Handbook of Proteolytic Enzymes*, 2594-2600. Elsevier. doi:10.1016/B978-0-12-382219-2.00575-5

Bauer, R., Wilson, J. J., Philominathan, S. T., Davis, D., Matsushita, O., & Sakon, J. (2012). Structural Comparison of ColH and ColG Collagen-Binding Domains from Clostridium histolyticum. *Journal of Bacteriology, 195*(2), 318-327. doi:10.1128/jb.00010-12

Bella, J., Eaton, M., Brodsky, B., & Berman, H. M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 A resolution. *Science*, *266*(5182), 75-81. doi:10.1126/science.7695699

BIO-RAD (2011). *Mini-PROTEAN Precast Protein Gels Instruction Manual and Application Guide* [User manual]. Retrieved from https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_1658100.pdf

Biotium (2018). *PI-GelRed-GelGreen-Quick-Start-Protocol* [User manual]. Retrieved from https://biotium.com/wp-content/uploads/2013/11/PI-GelRed-GelGreen-Quick-Start-Protocol.pdf

Bode, W., & Schwager, P. (1975). The refined crystal structure of bovine β-trypsin at 1·8 Å resolution. *Journal of Molecular Biology, 98*(4), 693-717. doi:10.1016/s0022-2836(75)80005-2

Bonander, N. & Bill, R. M. (2012). Optimising Yeast as a Host for Recombinant Protein Production (Review). In Bill, R. M. (Ed.). *Recombinant Protein Production in Yeast : Methods and Protocols*, 1-9. Humana Press Inc.

Bond, M. D., & Wart, H. E. (1984). Characterization of the individual collagenases from Clostridium histolyticum. *Biochemistry, 23*(13), 3085-3091. doi:10.1021/bi00308a036

Brändén, C & Tooze, J. (1999). *Introduction to Protein Structure.* New York: Garland Science.

Breite, A., Mccarthy, R., & Dwulet, F. (2011). Characterization and Functional Assessment of Clostridium

Histolyticum Class I (C1) Collagenases and the Synergistic Degradation of Native Collagen in Enzyme Mixtures Containing Class II (C2) Collagenase. *Transplantation Proceedings, 43*(9), 3171-3175. doi:10.1016/j.transproceed.2011.09.059

Byun, H. G., & Kim, S. K. (2001). Purification and characterization of angiotensin I converting enzyme (ACE) inhibitory peptides from Alaska pollack (Theragra chalcogramma) skin. *Process Biochemistry*, *36*(12), 1155-1162. doi:10.1016/S0032-9592(00)00297-1

Carter, P., & Wells, J. A. (1988). Dissecting the catalytic triad of a serine protease. *Nature, 332*(6164), 564-568. doi:10.1038/332564a0

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiology Reviews*, *24*(1), 45–66. doi:10.1016/S0168-6445(99)00029-7

Chaplin, M. (2004). *The use of enzymes in the detergents.* Retrived 2020-04-20 from http://www1.lsbu.ac.uk/water/enztech/detergent.html

Chen, R. F., Scott, C., & Trepman, E. (1979). Fluorescence properties of o-phthaldialdehyde derivatives of amino acids. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *576*(2), 440-455. doi:10.1016/0005-2795(79)90419-7

Choudhury, D., Roy, S., Chakrabarti, C., Biswas, S., & Dattagupta, J. K. (2009). Production and recovery of recombinant propapain with high yield. *Phytochemistry*, *70*(4), 465-472. doi:10.1016/j.phytochem.2009.02.001

Church, F. C., Porter, D. H., Catignani, G. L., & Swaisgood, H. E. (1985). An o-phthalaldehyde spectrophotometric assay for proteinases. *Analytical biochemistry*, *146*(2), 343-348. doi:10.1016/0003-2697(85)90549-4

Cregg, J. M. (2007). Introduction - Distinctions Between Pichia pastoris and Other Expression Systems. In Cregg, J. M. (Ed.). *Pichia protocols*, 1-10. New Jersey: Humana Press.

Cupp-Enyard, C. (2008). Sigma's non-specific protease activity assay-casein as a substrate. *JoVE (Journal of Visualized Experiments)*, (19), e899. doi:10.3791/899

Darby, R. A., Cartwright, S. P., Dilworth, M. V., & Bill, R. M. (2012). Which Yeast Species Shall I Choose? Saccharomyces cerevisiae versus Pichia pastoris (Review). In Bill, R. M. (Ed.). *Recombinant Protein Production in Yeast : Methods and Protocols*, 11-23. Humana Press Inc.

Das, S., & Dash, H. R. (2014). *Microbial biotechnology-a laboratory manual for bacterial systems*. Springer.

Duarte, A. S., Correia, A., & Esteves, A. C. (2014). Bacterial collagenases – A review. *Critical Reviews in Microbiology, 42*(1), 106-126. doi:10.3109/1040841x.2014.90427

Eckhard, U., Schönauer, E., & Brandstetter, H. (2013). Structural Basis for Activity Regulation and Substrate Preference of Clostridial Collagenases G, H, and T. *Journal of Biological Chemistry, 288*(28), 20184-20194. doi:10.1074/jbc.m112.448548

Eckhard, U., Schönauer, E., Nüss, D., & Brandstetter, H. (2011). Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis. *Nature Structural & Molecular Biology,18*(10), 1109-1114. doi:10.1038/nsmb.2127

Engel, J. and Bächinger H. P. (2005). Structure, stability and folding of the collagen triple helix. In Brinckmann, J., Notbohm, H., & Müller, P. K. (Eds.), *Collagen: Primer in structure, processing and assembly*, 7-33. Springer.

Ernst, O., & Zor, T. (2010). Linearization of the Bradford protein assay. *JoVE (Journal of Visualized Experiments)*, (38), e1918. doi:10.3791/1918.

Escherich, T. (1885). Die darmbakterien des neugeborenen und Säuglings Fortschritte des medizin. *München*, *3*, 512-522.

Evnin, L. B., Vasquez, J. R., & Craik, C. S. (1990). Substrate specificity of trypsin investigated by using a genetic selection. *Proceedings of the National Academy of Sciences, 87*(17), 6659-6663. doi:10.1073/pnas.87.17.6659

Fitzgerald, J., Leonard, P., Darcy, E., Sharma, S & O'Kennedy, R. (2017). Chapter 3 - Immunoaffinity Chromatography: Concepts and Applications. In Walls, D. & Loughran, S. T. (Eds.). *Protein*

Chromatography: Methods and Protocols. New York: Humana Press. doi:10.1007/978-1-4939-6412-3

Fratzl, P. (2008). Collagen: structure and mechanics, an introduction. In Fratzl, P. (Ed.), 1-13. *Collagen: Structure and Mechanics*. Springer. doi:10.1007/978-0-387-73906-9

French, M. F., Mookhtiar, K. A., & Wart, H. E. (1987). Limited proteolysis of type I collagen at hyperreactive sites by class I and II Clostridium histolyticum collagenases: Complementary digestion patterns. *Biochemistry,26*(3), 681-687. doi:10.1021/bi00377a004

Fruton, J. S. (1970). Specificity and Mechanism of Pepsin Action. In Desnuelle, P., Neurath, H., & Ottesen, M. (Eds.), *Structure–Function Relationships of Proteolytic Enzymes*, 222–236. New York: Academic Press. doi:10.1016/b978-0-12-211850-0.50022-7

Genscript (n. d.). *OptimumGene™ Codon Optimization - Most Cited Genetic Codon Designer for Protein Expression* [Pamphlet]. Retrieved from https://www.genscript.com/gsfiles/catalog/OptimumGene%20Codon%20Optimization%20Brochure.pdf?467077661

Glazer, A. N. & Smith E. L. (1971). Papain and other plant sulfhydryl proteolytic enzymes. In P.D. Boyer (Ed.), *The Enzymes, vol. 3*, 501-546. New York: Academic Press. doi:10.1016/s1874-6047(08)60405-9

Go, K., Horikawa, Y., Garcia, R., & Villarreal, F. J. (2008). Fluorescent method for detection of cleaved collagens using O-phthaldialdehyde (OPA). *Journal of Biochemical and Biophysical Methods*, *70*(6), 878–882. doi:10.1016/j.jbbm.2007.05.004

Gómez-Guillén, M., Giménez, B., López-Caballero, M., & Montero, M. (2011). Functional and bioactive properties of collagen and gelatin from alternative sources: A review. *Food Hydrocolloids, 25*(8), 1813-1827. doi:10.1016/j.foodhyd.2011.02.007

Gómez-Guillén, M. C., López-Caballero, M. E., López de Lacey, A., Alemán, A., Giménez, B., & Montero, P. (2010). Antioxidant and antimicrobial peptide fractions from squid and tuna skin gelatin. In Le Bihan, E. & Koueta N. (Eds.), *Sea by-products as a real material: New ways of application*, 89-115. Trivandrum: Transworld research network.

Granum, P. E., & Lund, T. (2006). Bacillus cereus and its food poisoning toxins. *FEMS Microbiology Letters*, *157*(2), 223–228. doi:10.1111/j.1574-6968.1997.tb12776.x

Graycar, T. P., Bott, R. R., Power, S. D., & Estell, D. A. (2013). Chapter 693 - Subtilisins. In Rawlings, N. D., & Salvesen, G. (Eds.), *Handbook of Proteolytic Enzymes*, 3148–3155. Elsevier. doi:10.1016/B978-0-12-382219-2.00693-1

Gross, J. (1956). The Behavior Of Collagen Units As A Model In Morphogenesis. *The Journal of Biophysical and Biochemical Cytology, 2*(4), 261-274. doi:10.1083/jcb.2.4.261

Halfon, S., & Craik, C. S. (1996). Regulation of Proteolytic Activity by Engineered Tridentate Metal Binding Loops. *Journal of the American Chemical Society, 118*(5), 1227-1228. doi:10.1021/ja9533813

Hatheway, C. L. (1990). Toxigenic clostridia. *Clinical Microbiology Reviews, 3*(1), 66-98. doi:10.1128/cmr.3.1.66

Hathout, R. M., & Metwally, A. A. (2019). Gelatin Nanoparticles. In Weissig, V. & Elbayoumi, T. (Eds.), *Pharmaceutical Nanotechnology,* 71-78. New York: Humana.

Hausjell, J., Weissensteiner, J., Molitor, C., Halbwirth, H & Spadiut, O. (2018). E. coli HMS174(DE3) is a sustainable alternative to BL21(DE3). *Microbial Cell Factories*, *17*(1), 1–9. doi:10.1186/s12934-018-1016-

Hedstrom, L., Szilagyi, L., & Rutter, W. (1992). Converting trypsin to chymotrypsin: The role of surface loops. *Science, 255*(5049), 1249-1253. doi:10.1126/science.1546324

Heissel, S., Frederiksen, S. J., Bunkenborg, J., & Højrup, P. (2019). Enhanced trypsin on a budget: Stabilization, purification and high-temperature application of inexpensive commercial trypsin for proteomics applications. *Plos One,14*(6). doi:10.1371/journal.pone.0218374

Herriott, R. M. (1938). Isolation, Crystallization, And Properties Of Swine Pepsinogen. *The Journal of General Physiology, 21*(4), 501-540. doi:10.1085/jgp.21.4.501

Hickman, F. W., Farmer, J. J., Hollis, D. G., Fanning, G. R., Steigerwalt, A. G., Weaver, R. E., & Brenner, D. J. (1982). Identification of Vibrio hollisae sp. nov. from patients with diarrhea. *Journal of Clinical*

Microbiology, *15*(3), 395–401. doi:10.1128/jcm.15.3.395-401.1982

Higaki, J. N., Evnin, L. B., & Craik, C. S. (1989). Introduction of a cysteine protease active site into trypsin. *Biochemistry, 28*(24), 9256-9263. doi:10.1021/bi00450a004

Hilaire, P. M., Willert, M., Juliano, M. A., Juliano, L., & Meldal, M. (1999). Fluorescence-Quenched Solid Phase Combinatorial Libraries in the Characterization of Cysteine Protease Substrate Specificity†. *Journal of Combinatorial Chemistry,1*(6), 509-523. doi:10.1021/cc990031u

Hilpert, K., Hansen, G., Wessner, H., Küttner, G., Welfle, K., Seifert, M., & Höhne, W. (2001). Anti-c-myc antibody 9E10: epitope key positions and variability characterized using peptide spot synthesis on cellulose. *Protein Engineering, 14*(10), 803–806. doi:10.1093/protein/14.10.803

Hokanson, J. E., & Austin, M. A. (1996). Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. *Journal of Cardiovascular Risk*, *3*(2), 213–219. doi:10.1177/174182679600300214

Hong, G., Min, S., & Jo, Y. (2019). Anti-Oxidative and Anti-Aging Activities of Porcine By-Product Collagen Hydrolysates Produced by Commercial Proteases: Effect of Hydrolysis and Ultrafiltration. *Molecules, 24*(6), 1104. doi:10.3390/molecules24061104

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., & Steigemann, W. (1974). Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. *Journal of Molecular Biology, 89*(1), 73-101. doi:10.1016/0022-2836(74)90163-6

Hulmes, D. J. S. (2008). Collagen diversity, synthesis and assembly. In Fratzl, P. (Ed.), 15-47. *Collagen: Structure and Mechanics*. Springer. doi:10.1007/978-0-387-73906-9

Ikemura, H., Takagi, H., & Inouye, M. (1987). Requirement of pro-sequence for the production of active subtilisin E in Escherichia coli. *Journal of Biological Chemistry*, *262*(16), 7859-7864. https://www.jbc.org/content/262/16/7859.full.pdf

Inanc, S., Keles, D., & Oktay, G. (2017). An improved collagen zymography approach for evaluating the collagenases MMP-1, MMP-8, and MMP-13. *BioTechniques*, *63*(4), 174–180. doi:10.2144/000114597

Inoue, N., Sugihara, F., & Wang, X. (2016). Ingestion of bioactive collagen hydrolysates enhance facial skin moisture and elasticity and reduce facial ageing signs in a randomised double-blind placebo-controlled clinical study. *Journal of the Science of Food and Agriculture*, *96*(12), 4077-4081. doi:10.1002/jsfa.7606

Invitrogen (n. d.). *Pichia pastoris expression system* [PDF file]. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fbrochures%2FB-067202_Pichia_Flyer.pdf&title=UGljaGlhIHBhc3RvcmlzIEV4cHJlc3Npb24gU3lzdGVt

Invitrogen (2002). *T4 DNA Ligase 1U/ul* [User manual]. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fbrochures%2F711-011113%2520TB-T4DNALigaseAp.pdf&title=UmFwaWQgbGlnYXRpb24gcHJvdG9jb2w=

Invitrogen (2010a). *pPICZα A, B, and C - Pichia expression vectors for selection on Zeocin™ and purification of secreted, recombinant proteins* [User manual]. Retrieved from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/ppicz_man.pdf

Invitrogen (2010b). *EasySelect™ Pichia Expression Kit - For Expression of Recombinant Proteins Using pPICZ and pPICZα in Pichia pastoris* [User manual]. Retrieved from http://tools.thermofisher.com/content/sfs/manuals/easyselect_man.pdf

Invitrogen (2012a). *User Guide: Double and Multiple Digestion of DNA* [User manual]. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0011984_Double_Multiple_Digestion_DNA_UG.pdf&title=VXNlciBHdWlkZTogRG91YmxlIGFuZCBNdWx0aXBsZSBEaWdlc3Rpb24gb2YgRE5B

Invitrogen (2012b). *User Guide: Fast Digestion of DNA*. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.ther

mofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0012413_Fast_Digestion_DNA_UG.pdf&title=VXNlciBHdWlkZTogRmFzdCBEaWdlc3Rpb24gb2YgRE5B

Invitrogen (2016). *Product Insert: Anza T4 DNA Ligase Master Mix* [User manual]. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0014310_anza_T4ligase_PI.pdf&title=UHJvZHVjdCBJbnNlcnQ6IEFuemEgVDQgRE5BIExpZ2FzZSBNYXN0ZXIgTWl4

Jagannadha Rao, K., & Panda, T. (1996). Release of periplasmic penicillin amidase from Escherichia coli by chloroform shock. *Bioprocess Engineering*, *14*(2), 101. doi:10.1007/BF00387964

Je, J. Y., Qian, Z. J., Byun, H. G., & Kim, S. K. (2007). Purification and characterization of an antioxidant peptide obtained from tuna backbone protein by enzymatic hydrolysis. *Process Biochemistry*, *42*(5), 840-846. doi:10.1016/j.procbio.2007.02.006

Jeong, H., Barbe, V., Lee, C. H., Vallenet, D., Yu, D. S., Choi, S. H., ... & Hur, C. G. (2009). Genome sequences of Escherichia coli B strains REL606 and BL21 (DE3). *Journal of molecular biology*, *394*(4), 644-652. doi:10.1016/j.jmb.2009.09.052

Jeong, H., Kim, H. J., & Lee, S. J. (2015). Complete genome sequence of Escherichia coli strain BL21. *Genome Announc.*, *3*(2), e00134-15. doi:10.1128/genomeA.00134-15

Jia, J., Zhou, Y., Lu, J., Chen, A., Li, Y., & Zheng, G. (2010). Enzymatic hydrolysis of Alaska pollack (Theragra chalcogramma) skin and antioxidant activity of the resulting hydrolysate. *Journal of the Science of Food and Agriculture*. doi:10.1002/jsfa.3861

Jung, W. K., Karawita, R., Heo, S. J., Lee, B. J., Kim, S. K., & Jeon, Y. J. (2006). Recovery of a novel Ca-binding peptide from Alaska Pollack (Theragra chalcogramma) backbone by pepsinolytic hydrolysis. *Process Biochemistry*, *41*(9), 2097-2100. doi:10.1016/j.procbio.2006.05.008

Kageyama, T. (2006). Roles of Tyr13and Phe219in the Unique Substrate Specificity of Pepsin B. *Biochemistry, 45*(48), 14415-14426. doi:10.1021/bi061467a

Kamphuis, I., Kalk, K., Swarte, M., & Drenth, J. (1984). Structure of papain refined at 1.65 Å resolution. *Journal of Molecular Biology,179*(2), 233-256. doi:10.1016/0022-2836(84)90467-4

Khow, O., & Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine*, *2*(2), 159–162. doi:10.1016/S2221-1691(11)60213-X

Khiari, Z., Ndagijimana, M., & Betti, M. (2014). Low molecular weight bioactive peptides derived from the enzymatic hydrolysis of collagen after isoelectric solubilization/precipitation process of turkey by-products. *Poultry Science, 93*(9), 2347-2362. doi:10.3382/ps.2014-03953

Koetsier, G. & Cantor, E. (2019). *A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers*. Retrieved from https://www.neb.com/-/media/catalog/application-notes/mvs_analysis_of_na_concentration_and_purity.pdf?la=en&rev=be7c8e19f4d34e558527496ea51623dc

Kühne, W. (1867). Ueber die Verdauung der Eiweisstoffe durch den Pankreassaft. *Archiv Für Pathologische Anatomie Und Physiologie Und Für Klinische Medicin, 39*(1), 130-174. doi:10.1007/bf01878988

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, *227*(5259), 680-685. doi:10.1038/227680a0

León-López, A., Fuentes-Jiménez, L., Hernández-Fuentes, A. D., Campos-Montiel, R. G., & Aguirre-Álvarez, G. (2019). Hydrolysed Collagen from Sheepskins as a Source of Functional Peptides with Antioxidant Activity. *International Journal of Molecular Sciences,20*(16), 3931. doi:10.3390/ijms20163931

Leray, M., Knowlton, N., Ho, S. L., Nguyen, B. N., & Machida, R. J. (2019). GenBank is a reliable resource for 21st century biodiversity research. *Proceedings of the National Academy of Sciences*, *116*(45), 22651-22656. doi:10.1073/pnas.1911714116

Lin-Cereghino, G. P., Stark, C. M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., ... & Huang, A. D. (2013). The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris.

Gene, *519*(2), 311-317. doi:10.1016/j.gene.2013.01.062

Lin, Y., Fusek, M., Lin, X., Hartsuck, J. A., Kezdy, F. J., & Tang, J. (1992). pH dependence of kinetic parameters of pepsin, rhizopuspepsin, and their active-site hydrogen bond mutants. *J. Biol. Chem., 267*, 18413-18418. https://www.jbc.org/content/267/26/18413.full.pdf

Lin, X., Wong, R. N. S., & Tang, J. (1989). Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen *J. Biol. Chem., 264*, 4482-4489. https://www.jbc.org/content/264/8/4482.full.pdf

Liu, L., Yang, H., Shin, H., Chen, R.R., Li, J., Du, G., & Chen, J. (2013). How to achieve high-level expression of microbial enzymes: strategies and perspectives. *Bioengineered*, *4*(4), 212–223. doi:10.4161/bioe.24761

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). *Molecular cell biology, 4th edition*. New York: W. H. Freeman and Company.

Maertens, B., Spriestersbach, A., von Groll, U., Roth, U., Kubicek, J., Gerrits, M., ... & Schäfer, F. (2010). Gene optimization mechanisms: a multi‐gene study reveals a high success rate of full‐length human proteins expressed in Escherichia coli. *Protein Science*, *19*(7), 1312-1326. doi:10.1002/pro.408

Makinen, K. K., & Makinen, P. L. (1987). Purification and properties of an extracellular collagenolytic protease produced by the human oral bacterium Bacillus cereus (strain Soc 67). *The Journal of Biological Chemistry*, *262*(26), 12488–12495.

MarketsandMarkets, (2018). *Collagen Peptides Market by Source (Cattle Hide & Bones, Pigskin, and Poultry & Fish), Application (Nutritional Products, Snacks & Cereals, Dairy Products, Beverages, and Meat & Poultry Products), and Region - Global Forecast to 2023*. Retrieved 2020-05-11 from https://www.marketsandmarkets.com/Market-Reports/collagen-peptide-market-248590980.html

Markland, F. S., & Smith, E. L. (1971). 16 Subtilisins: Primary Structure, Chemical and Physical Properties. In Boyer P. D. (Ed.), *Hydrolysis: Peptide Bonds The Enzymes*, 561–608. New York and London: Academic Press. doi:10.1016/s1874-6047(08)60407-2

Matsumoto, H., Haniu, H & Komori, N. (2018). Chapter 10 - Determination of Protein Molecular Weights on SDS-PAGE. In Kurien, B. T. & Hal Scofield, R. (Eds.). *Electrophoretic Separation of Proteins : Methods and Protocols*. New York: Humana Press. doi:10.1007/978-1-4939-8793-1_10

Matsushita, O., Jung, C., Katayama, S., Minami, J., Takahashi, Y., & Okabe, A. (1999). Gene Duplication and Multiplicity of Collagenases in Clostridium histolyticum. *Journal of Bacteriology, 181*(3), 923-933. doi:10.1128/jb.181.3.923-933.1999

Mccarthy, R., Spurlin, B., Wright, M., Breite, A., Sturdevant, L., Dwulet, C., & Dwulet, F. (2008). Development and Characterization of a Collagen Degradation Assay to Assess Purified Collagenase Used in Islet Isolation. *Transplantation Proceedings, 40*(2), 339-342. doi:10.1016/j.transproceed.2008.01

Mergulhao, F. J., Monteiro, G. A., Cabral, J. M., & Taipa, M. A. (2004). Design of bacterial vector systems for the production of recombinant proteins in Escherichia coli. *Journal of microbiology and biotechnology*, *14*(1), 1-14. https://www.researchgate.net/profile/Filipe_Mergulhao/publication/273132208_Design_of_bacterial_vector_systems_for_the_production_of_recombinant_proteins_in_Escherichia_coli/links/54f897990cf2ccffe9df573d.pdf

Menach, E., Hashida, Y., Yasukawa, K., & Inouye, K. (2013). Effects of Conversion of the Zinc-Binding Motif Sequence of Thermolysin, HEXXH, to That of Dipeptidyl Peptidase III, HEXXXH, on the Activity and Stability of Thermolysin. *Bioscience, Biotechnology, and Biochemistry, 77*(9), 1901-1906. doi:10.1271/bbb.130360

Mendis, E., Rajapakse, N., Byun, H., & Kim, S. (2005). Investigation of jumbo squid (Dosidicus gigas) skin gelatin peptides for their in vitro antioxidant effects. *Life Sciences, 77*(17), 2166-2178. doi:10.1016/j.lfs.2005.03.016

Mendis, E., Rajapakse, N., & Kim, S. K. (2005). Antioxidant properties of a radical-scavenging peptide purified from enzymatically prepared fish skin gelatin hydrolysate. *Journal of agricultural and food chemistry*, *53*(3), 581-587. doi:10.1021/jf048877v

Mirigian, L. S., Makareeva, E., Koistinen, H., Itkonen, O., Sorsa, T., Stenman, U., . . . Leikin, S. (2013). Collagen

degradation by tumor-associated trypsins. *Archives of Biochemistry and Biophysics,535*(2), 111-114. doi:10.1016/j.abb.2013.03.008

Montoya, C. A., Gomez, A. S., Lallès, J. P., Souffrant, W. B., Beebe, S., & Leterme, P. (2008). In vitro and in vivo protein hydrolysis of beans (Phaseolus vulgaris) genetically modified to express different phaseolin types. *Food chemistry*, *106*(3), 1225-1233. doi:10.1016/j.foodchem.2007.07.01

Moskowitz, R. W. (2000). Role of collagen hydrolysate in bone and joint disease. *Seminars in arthritis and rheumatism*, 30(2), 87-99. doi:10.1053/sarh.2000.9622

Nakade, K., Kamishima, R., Inoue, Y., Ahhmed, A., Kawahara, S., Nakayama, T., . . . Muguruma, M. (2008). Identification of an antihypertensive peptide derived from chicken bone extract. *Animal Science Journal, 79*(6), 710-715. doi:10.1111/j.1740-0929.2008.00584.x

Nielsen, P., & Foltmann, B. (1995). Purification and Characterization of Porcine Pepsinogen B and Pepsin B. *Archives of Biochemistry and Biophysics*, *322*(2), 417–422. doi:10.1006/abbi.1995.1483

Nielsen, K. H. (2014). Chapter Twelve - Protein Expression-Yeast. In Lorsch, J. (Ed.). *Methods in Enzymology*, Vol. 536, 133-147. Academic Press. doi:10.1016/B978-0-12-420070-8.00012-X

Nielsen, P. M., Petersen, D., & Dambmann, C. (2001). Improved method for determining food protein degree of hydrolysis. *Journal of food science*, *66*(5), 642-646. doi:10.1111/j.1365-2621.2001.tb04614.x

Nilsang, S., Lertsiri, S., Suphantharika, M., & Assavanig, A. (2005). Optimization of enzymatic hydrolysis of fish soluble concentrate by commercial proteases. *Journal of food Engineering*, *70*(4), 571-578. doi:10.1016/j.jfoodeng.2004.10.011

Norihiro, T., Yoshihiro, A., Yoshimi, T., Rie, U., Tomoaki, T., Takuji, S., . . . Kenji, T. (1988). Nucleotide sequence and expression in Escherichia coli of cDNA of swine pepsinogen: Involvement of the amino-terminal portion of the activation peptide segment in restoration of the functional protein. *Gene, 65*(2), 285-292. doi:10.1016/0378-1119(88)90465-9

Novagen (n. d.). *pET22b(+)* [PDF file]. Retrieved from https://www.helmholtz-muenchen.de/fileadmin/PEPF/pET_vectors/pET-22b_map.pdf

Novagen (1991). *pET System Manual* [User manual]. Retrieved from http://research.fhcrc.org/content/dam/stripe/hahn/methods/biochem/pet.pdf

Oakley, C. L., & Warrack, G. H. (1950). The Alpha, Beta and Gamma Antigens of Clostridium histolyticum (Weinberg & Seguin, 1916). *Journal of General Microbiology, 4*(3), 365-373. doi:10.1099/00221287-4-3-365

Oertzen-Hagemann, V., Kirmse, M., Eggers, B., Pfeiffer, K., Marcus, K,. de Marées, M & Platen, P. (2019). Effects of 12 Weeks of Hypertrophy Resistance Exercise Training Combined with Collagen Peptide Supplementation on the Skeletal Muscle Proteome in Recreationally Active Men. *Nutrients*, *11*(5), 1072. doi:10.3390/nu11051072

Omega BIO-TEK (2013). *D6942 D6943 D6945 Plasmid DNA Mini Kits I and II Protocol* [User manual]. Retrieved from https://www.omegabiotek.com/wp-content/uploads/2018/10/D6942.D6943.D6945-PROTOCOL-Plasmid-DNA-Mini-Kits-I-and-II-Combo.pdf

Omega BIO-TEK (2019). D2500 Gel Extraction Kit Protocol [User manual]. Retrieved from https://www.omegabiotek.com/product/e-z-n-a-gel-extraction-kit/

Ong, I. L. H., & Yang, K. L. (2017). Recent developments in protease activity assays and sensors. *Analyst*, *142*(11), 1867-1881. doi:10.1039/c6an02647h

Ottesen, M. & Svendsen I. (1970). [11] The Subtilisins. In Perlmann,G. E. & Lorand L. (Eds.), *Methods in Enzymology Proteolytic Enzymes*, 199–215. Elsevier. doi:10.1016/0076-6879(70)19014-8.

Pal, G. K., & Suresh, P. V. (2016). Sustainable valorisation of seafood by-products: Recovery of collagen and development of collagen-based novel functional food ingredients. *Innovative Food Science and Emerging Technologies, 37*(Part B), 201–215. doi:10.1016/j.ifset.2016.03.015

Palmer, I., & Wingfield, P. T. (2012). Preparation and extraction of insoluble (inclusion‐body) proteins from

Escherichia coli. *Current protocols in protein science*, *70*(1), 6-3. doi:10.1002/0471140864.ps0603s70
Philipp, M., Tsai, I., & Bender, M. L. (1979). Comparison of the kinetic specificity of subtilisin and thiolsubtilisin toward n-alkyl p-nitrophenyl esters. *Biochemistry,18*(17), 3769-3773. doi:10.1021/bi00584a020
Pearce, K. E. N., Karahalios, D. A., & Friedman, M. E. (1988). Ninhydrin assay for proteolysis in ripening cheese. *Journal of Food Science*, *53*(2), 432-435. doi:10.1111/j.1365-2621.1988.tb07723.x
Pintado, M. E., Pintado, A. E., & Malcata, F. X. (1999). Controlled whey protein hydrolysis using two alternative proteases. *Journal of Food Engineering*, *42*(1), 1-13. doi:10.1016/S0260-8774(99)00094-1
Poppe, J. (1992). Gelatin. In Imeson, A. (Ed.) *Thickening and Gelling Agents for Food*, 98-123. Boston: Springer.
Prehna, G., Zhang, G., Gong, X., Duszyk, M., Okon, M., McIntosh, L. P., ... & Strynadka, N. C. (2012). A protein export pathway involving Escherichia coli porins. *Structure*, *20*(7), 1154-1166. doi:10.1016/j.str.2012.04.01
Raab, D., Graf, M., Notka, F., Schödl, T., & Wagner, R. (2010). The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Systems & Synthetic Biology*, 4(3), 215-225. doi:10.1007/s11693-010-9062-3
Ramjee, M. K., Petithory, J. R., McElver, J., Weber, S. C., & Kirsch, J. F. (1996). A novel yeast expression/secretion system for the recombinant plant thiol endoprotease propapain. *Protein Engineering, Design and Selection*, *9*(11), 1055-1061. doi:10.1093/protein/9.11.1055
Rawlings, N. D., Barrett, A. J., & Bateman, A. (2009). MEROPS: The peptidase database. *Nucleic Acids Research, 38*(Suppl_1). doi:10.1093/nar/gkp971
Renneberg, R., Berkling, V., & Loroch, V. (2016). *Biotechnology for beginners*. Academic Press.
Ricard-Blum, S. (2010). The Collagen Family. *Cold Spring Harbor Perspectives in Biology, 3*(1). doi:10.1101/cshperspect.a004978
Richmond, V., Tang, J., Wolf, S., Trucco, R. E., & Caputto, R. (1958). Chromatographic isolation of gastricsin, the proteolytic enzyme from gastric juice with pH optimum 3.2. *Biochimica Et Biophysica Acta*, *29*(2), 453–454. doi:10.1016/0006-3002(58)90220-8
Rivera, I., Gutiérrez-Ortega, A & Sandoval, G. (2018). Functional Expression of Plant Lipases: The Case of CpLip1 from Carica papaya. In Sandoval, G. (Ed.). *Lipases and Phospholipases : Methods and Protocols*, 169-178. New York: Humana Press. doi:10.1007/978-1-4939-8672-9_11
Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology*, *5*, 172. doi:10.3389/fmicb.2014.00172.
Rutherfurd, S. M. (2010). Methodology for determining degree of hydrolysis of proteins in hydrolysates: a review. *Journal of AOAC International*, *93*(5), 1515-1522. doi:10.1093/jaoac/93.5.1515
Ryle, A. P., & Porter, R. R. (1959). Parapepsins: two proteolytic enzymes associated with porcine pepsin. *Biochemical Journal*, *73*(1), 75–86. doi:10.1042/bj0730075
Sachdev, G. P., & Fruton, J. S. (1970). Secondary enzyme-substrate interactions and the specificity of pepsin. *Biochemistry*, *9*(23), 4465–4470. doi:10.1021/bi00825a001
Sagi, I., & Afratis, N. A. (2019). *Collagen Methods and Protocols*. New York: Springer.
Saito, M., Kiyose, C., Higuchi, T., Uchida, N., & Suzuki, H. (2009). Effect of collagen hydrolysates from salmon and trout skins on the lipid profile in rats. *Journal of Agricultural and Food Chemistry, 57*(21), 10477–10482. doi:10.1021/jf902355m
Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., & James, M. N. (1990). Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8A˚resolution. *Journal of Molecular Biology, 214*(1), 143-170. doi:10.1016/0022-2836(90)90153-d
Silvestre, M. P., Hamon, M., & Yvon, M. (1994). Analysis of protein hydrolyzates. 1. Use of poly (2-hydroxyethylaspartamide)-silica column in size exclusion chromatography for the fractionation of casein hydrolyzates. *Journal of Agricultural and Food Chemistry*, *42*(12), 2778-2782. doi:10.1021/jf00048a024
Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M., & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*, *14*(1), 1–10. doi:10.1186/s12934-015-0222-8

Singh, P., Sharma, L., Kulothungan, S. R., Adkar, B. V., Prajapati, R. S., Ali, P. S. S., ... & Varadarajan, R. (2013). Effect of signal peptide on stability and folding of Escherichia coli thioredoxin. *PloS one*, *8*(5). doi:10.1371/journal.pone.0063442

Snoek-van Beurden, P. A. M., & Von den Hoff, J. W. (2005). Zymographic techniques for the analysis of matrix metalloproteinases and their inhibitors. *BioTechniques*, *38*(1), 73–83. doi:10.2144/05381RV01.

Spellman, D., McEvoy, E., O'cuinn, G., & FitzGerald, R. J. (2003). Proteinase and exopeptidase hydrolysis of whey protein: Comparison of the TNBS, OPA and pH stat methods for quantification of degree of hydrolysis. *International dairy journal, 13*(6), 447-453. doi:10.1016/S0958-6946(03)00053-0

Storer, A. C. & Ménard, R. (2013). Chapter 419- Papain. In Rawlings, N. D., & Salvesen, G. (Eds.), *Handbook of Proteolytic Enzymes*, 1858-1861. Elsevier. doi:10.1016/B978-0-12-382219-2.00418-X

Strausberg, S., Alexander, P., Wang, L., Schwarz, F., & Bryan, P. (1993). Catalysis of a protein folding reaction: Thermodynamic and kinetic analysis of subtilisin BPN' interactions with its propeptide fragment. *Biochemistry,32*(32), 8112-8119. doi:10.1021/bi00083a009

Stroud, R. M., Kay, L. M., & Dickerson, R. E. (1974). The structure of bovine trypsin : Electron density maps of the inhibited enzyme at 5 Å and at 2·7 Å resolution. *Journal of Molecular Biology, 83*(2), 185-208. doi:10.1016/0022-2836(74)90387-8

Sugihara, F., Inoue, N, & Venkateswarathirukumara, S. (2018). Ingestion of bioactive collagen hydrolysates enhanced pressure ulcer healing in a randomized double-blind placebo-controlled clinical study. *Scientific Reports*, *8*(1), 1–7. doi:10.1038/s41598-018-29831-7

Szecsi, P. B. (2013). Pepsin B. In Rawlings, N. D., & Salvesen, G. (Eds.) *Handbook of Proteolytic Enzymes*, 36–37. Elsevier. doi:10.1016/b978-0-12-382219-2.00004-1

Tabata, Y., & Ikada, Y. (1998). Protein release from gelatin matrices. Advanced drug delivery reviews, 31(3), 287-301. doi:10.1016/s0169-409x(97)00125-7

Tabor, S. (1990). Expression using the T7 RNA polymerase/promoter system. *Current protocols in molecular biology*, *11*(1), 16-2. doi:10.1002/0471142727.mb1602s11

Takita, T., Qian, J., Geng, H., He, Z., Nemoto, S., Mori, M and Yasukawa, K. (2018). Comparative studies on the activities of collagenases from Grimontia hollisae and Clostridium hystoliticum in the hydrolysis of synthetic substrates. *The Journal of Biochemistry, 163*(5), 425-431. doi:10.1093/jb/mvy009

Tanaka, K., Okitsu, T., Teramura, N., Iijima, K., Hayashida, O., Teramae, H., & Hattori, S. (2020). Recombinant collagenase from Grimontia hollisae as a tissue dissociation enzyme for isolating primary cells. Scientific Reports, 10(1). doi:10.1038/s41598-020-60802-z

Tanaka, K., Teramura, N., Hayashida, O., Iijima, K., Okitsu, T., & Hattori, S. (2018). The C-terminal segment of collagenase in Grimontia hollisae binds collagen to enhance collagenolysis. *FEBS Open Bio, 8*(10), 1691-1702. doi:10.1002/2211-5463.12510

Tang, J. (2013). Pepsin A. In Rawlings, N. D., & Salvesen, G. (Eds.) *Handbook of Proteolytic Enzymes*, 27–35. Elsevier. doi:10.1016/b978-0-12-382219-2.00003-x

Tang, J., Sepulveda, P., Marciniszyn, J., Chen, K. C. S., Huang, W.-Y., Tao, N., … Lanier, J. P. (1973). Amino-Acid Sequence of Porcine Pepsin. *Proceedings of the National Academy of Sciences*, *70*(12), 3437–3439. doi:10.1073/pnas.70.12.3437

Taylor, M. A., Pratt, K. A., Revell, D. F., Baker, K. C., Sumner, I. G., & Goodenough, P. W. (1992). Active papain renatured and processed from insoluble recombinant propapain expressed in Escherichia coli. *Protein Engineering, Design and Selection*, *5*(5), 455-459. doi:10.1093/protein/5.5.455

Teramura, N., Iijima, K., Hayashida, O., Tanaka, K., Hatori, S., Okitsu, T. and Takeuchi S. (2018). GRIMONITIA-HOLLISAE-DERIVED RECOMBINANT COLLAGENASE AND ENZYME AGENT FOR CELL AND TISSUE DISSOCIATION. U.S. Patent No. *US 10,047,353 B2*. Washington, DC: U.S. Patent and Trademark Office

Teramura, N., Tanaka, K., Iijima, K., Hayashida, O., Suzuki, K., Hattori, S., & Irie, S. (2011). Cloning of a Novel Collagenase Gene from the Gram-Negative Bacterium Grimontia (Vibrio) hollisae 1706B and Its Efficient

Expression in Brevibacillus choshinensis. *Journal of Bacteriology, 193*(12), 3049-3056. doi:10.1128/jb.01528-10

Thermo Scientific (2015). *User Guide: 6X TriTrack DNA Loading Dye* [User manual]. Retrieved from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0013168_6X_TriTrack_DNA_Loading_Dye_UG.pdf

Thermo Scientific (2019). *BL21(DE3) Competent CellsUser Guide (Pub.No. MAN0018595 A.0)* [User manual]. Retrieved from https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0018595_BL21_DE3competent_cells_UG.pdf&title=VXNlciBHdWlkZTogQkwyMShERTMpIENvbXBldGVudCBDZWxscw==

Toikkanen, J. (n. d.). *Expression of proteins in yeast systems* [PDF file]. Retrieved from http://www.biocenter.helsinki.fi/biotechgs/media/toikkanen.pdf

Vasquez, J. R., Evnin, L. B., Higaki, J. N., & Craik, C. S. (1989). An expression system for trypsin. *Journal of Cellular Biochemistry, 39*(3), 265-276. doi:10.1002/jcb.240390306

Vandooren, J., Geurts, N., Martens, E., Van den Steen, P. E., & Opdenakker, G. (2013). Zymography methods for visualizing hydrolytic enzymes. *Nature Methods*, *10*(3), 211–220. doi:10.1038/nmeth.2371

Vernet, T., Tessier, D. C., Richardson, C., Laliberte, F., Khouri, H. E., Bell, A. W., ... & Thomas, D. Y. (1990). Secretion of functional papain precursor from insect cells. Requirement for N-glycosylation of the pro-region. *Journal of Biological Chemistry*, *265*(27), 16661-16666. https://www.jbc.org/content/265/27/16661.long

Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F., & Parachin, N. S. (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms*, *6*(2), 38. doi:10.3390/microorganisms6020038

Volokitina, Z. V., Ionova, I. I. & Krasnova, I. S. (2020). Whey beverage with collagen hydrolysate from salmon skin and feijoa. *IOP Conference Series: Earth & Environmental Science*, *421*(3), 032017. doi:10.1088/1755-1315/421/3/032017

von Ahlfen, S. and Schlumpberger, M., (2010, 03). Effects of low A260/A230 ratios in RNA preparations on downstream applications. *Qiagen Gene Expression Newsletter,* 15, 7–8.

Wegner, G. H. (1990). Emerging applications of the methylotrophic yeasts. *FEMS microbiology reviews*, *7*(3-4), 279-283. doi:10.1111/j.1574-6968.1990.tb04925.x

Wirnt, R. (1974). Trypsin. In Bergmeyer, J. & Ulrich, H. (Eds.), *Methods of Enzymatic Analysis, volume 2,* 1013-1024. New York and London: Academic Press. doi:10.1016/B978-0-12-091302-2.50099-2

Yang, J. I., Ho, H. Y., Chu, Y. J., & Chow, C. J. (2008). Characteristic and antioxidant activity of retorted gelatin hydrolysates from cobia (Rachycentron canadum) skin. *Food chemistry*, *110*(1), 128-136. doi:10.1016/j.foodchem.2008.01.072

Yang, H., & Shu, Z. (2014). The extraction of collagen protein from pigskin. *Journal of chemical and pharmaceutical research*, *6*(2), 683-687. http://www.jocpr.com/articles/the-extraction-of-collagen-protein-from-pigskin.pdf

Yoshimura, S. H., & Hirano, T. (2016). HEAT repeats–versatile arrays of amphiphilic helices working in crowded environments?. *Journal of cell science*, *129*(21), 3963-3970. doi:10.1242/jcs.185710

Zauner, T., Berger-Hoffmann, R., Müller, K., Hoffmann, R., & Zuchner, T. (2011). Highly adaptable and sensitive protease assay based on fluorescence resonance energy transfer. *Analytical Chemistry*, *83*(19), 7356–7363. doi:10.1021/ac201274f

Zhang, G., Brokx, S., & Weiner, J. H. (2006). Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in Escherichia coli. *Nature biotechnology*, *24*(1), 100-104. doi:10.1038/nbt1174

Zhang, R. H., Xiao, L., Peng, Y., Wang, H. Y., Bai, F., & Zhang, Y. Z. (2005). Gene expression and characteristics of a novel fibrinolytic enzyme (subtilisin DFE) in Escherichia coli. *Letters in applied microbiology*, *41*(2), 190-195. doi:10.1111/j.1472-765X.2005.01715.x

Appendix A - DNA Data

Table of contents

Genbank and UniProt accession number of the enzymes

All accession numbers for the enzymes in Genbank and UniProt are collected below.

***Clostridium histolyticum* - Collagenase ColG**
UniProtKB - Q9X721
GenBank: D87215.1

***Clostridium histolyticum* - Collagenase ColH**
UniProtKB - Q46085
GenBank: AB014075.1

***Grimontia hollisae* - Collagenase**
Uniprot: F7IZI6
GeneBank: AB600550.1

***Bacillus cereus* - Collagenase ColA**
Uniprot: Q81BJ6
GeneBank: AE016877.1 (-1 reading frame)

***Bacillus licheniformi*s - Subtilisin Carlsberg**
Uniprot: P00780
GeneBank: X03341.1

***Canis lupus familiaris* - Pepsinogen B**
UniProtKB - Q8SQ41
GeneBank: AB082936.1

***Bos taurus* - Cationic Trypsin**
Uniprot: P00760
GenBank: BC134797.1

***Carica papaya* - Papain**
Uniprot: P00784
GenBank: M15203.1

Enzymes expressed in *E. coli* with pAES40 expression vector with YepF

The DNA sequences presented below contain N- and C-terminal restriction sites (in bold). The native signal peptides have been removed as Yebf protein will be utilized as well as the native Stop codon as the C-terminal His_6-tag contains Stop codon.

C. histolyticum - Collagenase ColG

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCAAACCCATAGAAAACACAAATGATACTTCAATTAAAAACGTCGAGAAGTTGCGTAATGCCCCGAACGAAGAAAATTCGAAGAAGGTCG
AAGATTCTAAGAACGATAAGGTTGAACATGTTAAGAACATTGAGGAGGCAAAGGTCGAACAGGTTGCTCCGGAGGTTAAAAGCAAATCTACTCT
GCGTTCCGCGAGCATCGCGAACACCAATAGCGAAAAATACGATTTCGAGTATCTGAACGGTCTGAGCTATACTGAACTGACCAATTTAATTAAG
AACATCAAGTGGAATCAGATTAACGGTTTGTTTAACTACTCTACGGGCAGCCAAAAGTTCTTCGGCGACAAGAACCGCGTACAAGCAATCATCA
ACGCTCTGCAAGAGAGCGGTCGTACATACACTGCGAACGATATGAAAGGCATTGAAACCTTCACTGAGGTGCTGCGCGCGGGTTTCTATCTTGG
CTACTACAACGATGGCCTGAGTTACTTAAACGACCGTAACTTCCAGGACAAGTGCATTCCGGCTATGATTGCAATTCAGAAAAATCCGAACTTC
AAACTGGGCACGGCTGTGCAAGATGAGGTGATCACCTCCCTGGGGAAACTGATCGGTAACGCCTCTGCGAATGCGGAGGTGGTTAATAATTGTG
TTCCGGTTTTGAAGCAGTTCCGTGAAAACCTGAACCAGTATGCTCCGGATTATGTCAAAGGTACGGCTGTTAACGAGCTGATCAAAGGGATCGA
GTTCGACTTCTCCGGTGCTGCCTATGAAAAGGACGTGAAGACAATGCCGTGGTATGGCAAAATCGATCCGTTTATTAATGAGTTGAAGGCGCTG
GGCCTCTACGGTAACATTACCAGCGCGACCGAGTGGGCCAGCGACGTAGGGATCTACTACCTATCCAAGTTTGGCCTGTACAGCACCAACCGTA
ATGATATCGTTCAGAGTCTGGAAAAGGCGGTCGATATGTATAAGTACGGCAAAATTGCATTCGTGGCCATGGAACGTATTACCTGGGACTACGA
CGGCATTGGTAGTAACGGCAAAAAAGTGGATCACGATAAATTTCTGGATGACGCGGAAAAGCACTACCTGCCAAAAACCTATACGTTCGACAAC
GGCACCTTTATCATCCGTGCAGGCGACAAGGTGAGCGAGGAGAAGATCAAACGTCTGTACTGGGCAAGCAGAGAAGTGAAATCTCAGTTTCATC
GTGTTGTTGGCAACGATAAAGCGCTGGAGGTCGGCAACGCGGACGACGTCCTGACCATGAAGATCTTCAACTCGCCGGAAGAGTACAAGTTCAA
TACGAACATCAATGGGGTGTCCACTGATAACGGTGGTCTCTACATCGAGCCGCGTGGTACTTTCTATACCTATGAGCGCACCCCGCAGCAGTCG
ATTTTTTCATTAGAAGAACTGTTCCGCCATGAATATACCCATTACCTCCAAGCGCGTTACTTGGTTGATGGTCTGTGGGGTCAAGGCCCGTTTT
ATGAAAAGAACCGCTTGACATGGTTTGACGAGGGCACGGCTGAATTCTTTGCAGGCTCGACCCGTACCAGCGGTGTTCTGCCGCGTAAATCTAT
CCTAGGTTATTTGGCGAAAGACAAAGTAGACCACCGCTATAGTCTTAAGAAAACCTTGAATAGTGGTTACGACGACTCCGATTGGATGTTTTAT
AACTATGGTTTTGCCGTGGCACATTATCTGTATGAGAAGGACATGCCGACCTTCATCAAGATGAATAAGGCCATCTTGAATACCGATGTTAAAA
GCTACGATGAAATTATCAAGAAATTGAGCGACGACGCCAACAAGAACACTGAGTACCAAAACCATATTCAAGAACTGGCGGATAAATACCAGGG
TGCCGGTATCCCACTGGTTAGCGACGACTACCTTAAGGACCACGGTTATAAGAAAGCTAGCGAAGTTTATAGCGAAATCAGCAAAGCAGCGAGC
CTCACTAATACTAGCGTGACTGCGGAAAAATCCCAATACTTCAACACGTTCACCCTGCGTGGCACCTACACCGGTGAAACCAGCAAAGGTGAAT
TTAAAGATTGGGATGAGATGAGCAAGAAGTTGGACGGCACGCTGGAGAGCCTGGCGAAGAATTCTTGGTCAGGCTACAAAACCTTGACCGCGTA
TTTCACGAACTACCGCGTTACCTCCGACAATAAAGTGCAGTATGATGTAGTTTTCCACGGCGTTCTCACCGATAATGCGGACATCAGCAATAAC
AAGGCTCCGATTGCGAAAGTCACCGGTCCGTCAACCGGCGCGGTCGGCCGTAACATCGAATTCTCCGGCAAAGACTCTAAAGACGAGGACGGCA
AGATCGTGTCCTACGATTGGGACTTCGGAGACGGTGCAACCTCACGCGGCAAGAATTCCGTGCACGCATATAAGAAGGCCGGCACCTACAACGT
GACCCTGAAGGTGACTGATGACAAAGGCGCAACCGCGACCGAGAGCTTTACAATTGAGATCAAAAACGAGGACACCACCACGCCGATAACCAAG
GAAATGGAACCGAATGATGACATCAAGGAGGCGAACGGTCCAATCGTGGAAGGGGTGACCGTGAAGGGCGACCTGAATGGCAGCGACGACGCGG
ATACCTTTTACTTCGATGTGAAAGAAGATGGTGACGTGACGATTGAGCTGCCGTATTCTGGTTCTAGCAACTTCACCTGGCTGGTGTACAAAGA
GGGCGACGACCAGAACCACATCGCGAGCGGCATTGATAAAAACAACTCAAAAGTTGGTACGTTTAAATCCACCAAAGGTCGCCACTACGTGTTT
ATTTACAAACACGACAGCGCGAGCAACATTAGCTACTCCCTGAATATCAAGGGTCTTGGTAATGAAAAACTGAAAGAGAAAGAGAACAACGATT
CGAGCGATAAGGCGACCGTTATTCCGAACTTTAATACCACCATGCAGGGTAGCCTTCTGGGTGACGACAGCCGTGATTATTACTCCTTTGAAGT
TAAAGAAGAAGGTGAAGTTAATATTGAGCTGGACAAAAAGGACGAGTTTGGTGTCACCTGGACCTTACATCCGGAGTCCAACATCAACGATCGT
ATCACCTATGGTCAAGTTGATGGCAACAAGGTCAGCAATAAAGTGAAGCTGAGACCTGGTAAATACTATTTGCTGGTTTATAAGTATTCTGGCT
CTGGTAATTACGAGCTGCGCGTGAACAAG**GAGCTC**

C. histolyticum - Collagenase ColH

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGCTGTAGATAAAAATAACGCTACAGCGGCAGTTCAAAACGAGAGCAAGCGCTACACTGTCAGCTATTTGAAGACCCTGAATTACTACG
ACCTGGTCGACTTGCTGGTTAAGACTGAGATCGAAAACCTGCCGGATTTGTTTCAGTATAGCAGCGACGCAAAAGAGTTCTATGGCAACAAAAC
CCGTATGAGCTTTATAATGGACGAGATCGGCCGCCGTGCGCCACAGTATACCGAGATCGACCACAAAGGGATCCCGACCCTGGTGGAAGTGGTG
CGTGCGGGCTTTTATTTAGGTTTCCACAACAAAGAGCTGAACGAAATCAACAAGAGAAGCTTCAAAGAACGTGTAATCCCGTCTATTTTGGCAA
TTCAGAAAAACCCGAACTTTAAGTTGGGCACGGAAGTCCAGGATAAGATCGTGTCTGCAACCGGTTTGCTGGCGGGTAACGAAACCGCACCTCC
GGAGGTGGTGAACAACTTCACCCCGATTCTGCAAGATTGTATTAAGAACATCGATCGTTATGCGCTCGACGATTTGAAGAGCAAAGCTTTGTTT
AATGTTCTGGCGGCACCGACATATGATATTACCGAATATTTGAGAGCCACCAAAGAAAAGCCGGAAAACACCCCGTGGTATGGCAAAATCGACG
GCTTCATCAACGAGCTGAAAAAATTGGCGCTGTACGGTAAAATTAATGATAATAATTCCTGGATAATCGACAATGGCATTTACCATATTGCCCC
TCTGGGTAAGCTGCATAGCAACAACAAAATCGGTATTGAAACCCTGACCGAAGTCATGAAAGTCTACCCGTATCTGAGCATGCAGCATCTTCAA
AGCGCTGACCAGATTAAACGCCACTATGACAGCAAAGATGCGGAGGGCAACAAGATCCCGCTTGACAAGTTCAAGAAAGAGGGCAAAGAGAAAT
ACTGCCCGAAAACCTACACGTTTGATGACGGTAAGGTCATCATTAAGGCAGGTGCGCGTGTTGAAGAGGAGAAAGTTAAGCGCCTGTATTGGGC
ATCCAAGGAAGTGAACTCGCAATTTTTCCGCGTTTATGGTATTGACAAGCCACTGGAAGAAGGTAATCCGGATGACATTCTGACCATGGTTATT
TACAATTCTCCGGAAGAGTACAAACTGAATAGCGTTCTGTACGGGTACGACACCAATAACGGTGGTATGTATATCGAACCAGAGGGCACTTTCT
TTACCTACGAACGTGAAGCGCAGGAGTCCACCTATACCTTGGAGGAACTGTTCCGTCACGAATACACGCATTACCTGCAAGGTCGTTACGCTGT
GCCGGGTCAATGGGGTCGTACTAAACTCTACGACAACGACCGTCTGACGTGGTATGAGGAGGGAGGCGCTGAACTGTTCGCGGGTAGCACCCGT
ACGTCGGGCATTCTGCCGCGTAAATCCATCGTGAGCAATATCCACAATACCACCCGTAATAACCGCTACAAACTATCTGACACCGTGCACAGCA
AGTACGGCGCGAGCTTTGAATTTTACAATTACGCCTGCATGTTTATGGATTATATGTATAACAAGGATATGGGTATTCTGAACAAGCTCAACGA
TTTGGCGAAAAACAACGATGTTGATGGCTATGACAACTACATTCGTGATCTCTCTTCTAACTACGCTTTGAACGACAAATATCAGGATCACATG
CAAGAGCGCATCGACAACTACGAAAACCTGACGGTTCCGTTTGTTGCAGACGACTACCTGGTTCGTCACGCATATAAAAATCCGAATGAAATCT
ACAGCGAGATCAGCGAAGTAGCCAAACTGAAGGACGCCAAGAGCGAAGTGAAAAAGTCGCAATATTTCAGCACGTTTACCTTACGCGGTTCGTA
CACAGGTGGTGCTAGCAAGGGTAAACTGGAAGATCAGAAAGCTATGAATAAGTTCATCGACGACTCCCTGAAAAAATTGGATACTTACAGCTGG
TCCGGCTATAAGACGCTGACCGCGTATTTCACCAATTATAAGGTTGACTCTTCTAATCGTGTCACCTATGACGTGGTGTTCCATGGTTACTTAC
CGAACGAGGGTGATAGCAAAAACTCGCTGCCGTATGGCAAGATCAACGGCACCTACAAGGGCACCGAGAAGGAGAAGATCAAATTCTCCAGCGA
GGGCTCATTTGATCCGGATGGAAAGATCGTGTCCTACGAGTGGGATTTCGGCGATGGGAATAAGTCCAACGAGGAGAACCCGGAACATAGCTAC
GACAAGGTTGGTACATACACGGTGAAACTTAAGGTGACCGACGACAAGGGAGAAAGCTCTGTCTCAACCACGACCGCTGAGATCAAGGACCTGT
CGGAGAACAAACTTCCGGTGATCTATATGCACGTGCCGAAGTCCGGTGCGCTGAATCAGAAAGTGGTGTTTTACGGTAAAGGTACTTACGACCC
GGATGGCTCTATAGCGGGCTATCAATGGGATTTTGGTGACGGTAGCGACTTCAGCTCTGAGCAGAACCCAAGCCATGTTTACACCAAAAAAGGT
GAATATACGGTTACCCTGCGCGTTATGGATAGCTCCGGCCAGATGAGCGAGAAGACCATGAAAATCAAGATCACCGACCCGGTGTATCCGATTG
GTACTGAGAAGGAACCGAATAACAGTAAGGAGACTGCGTCCGGTCCGATTGTTCCGGGTATCCCCGTTAGCGGTACGATTGAAAATACCAGCGA
CCAGGATTACTTCTACTTCGATGTTATTACCCCCGGCGAAGTTAAAATTGACATTAACAAACTCGGCTACGGTGGTGCGACCTGGGTTGTGTAC
GATGAAAACAACAACGCCGTCTCTTATGCGACCGATGACGGCCAAAACTTGTCCGGCAAGTTCAAAGCGGATAAACCGGGTCGTTACTACATCC
ACCTGTATATGTTCAACGGCAGCTACATGCCGTATCGTATTAACATTGAGGGCAGCGTTGGTCGC**GAGCTC**

G. hollisae - Collagenase

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCTCAGAAGTATTTGCTCACCCCGGAATGCCAGTTCAGCAGAAACTGCAACCGAATCTGCTGCAGCAGTCGACCCGTTTACAGCCGGAAC
AGCACATTCATGGCCTGGAACGTACGGACCGTCAGTACAGACCCACTGACGCCACCCAGCAACCGGAACCACCGACCTTGCTTAAGCGCCAGGT
GTCTGTTCAGCAGGACGCCGTGGAACAGTGCGATCTGAGTCAATTTCAGACCACCAGCAGCAATCAGCTGATGGCCGCTATCCGCCAACAAGGT
GCGTCTTGTGTTAACGCCCTGTTCAGCGCGGACACTGGTGTTCAAGAAGCTGCGTTCAGCTCCAATCATATGTACAACGTGGCCCAATACACCC
GCACCCTGGCTCAACAATACGCGGGTGGCGGCTCTGACGAGCTGGAGGCGCTGTACTTATACCTCCGTGCAGGTTACTATGCGGAATTCTATAA
CTCTAATATCACCTTTCTGAGCTGGGTTACCCCGGCGGTGAAAGGTGCGGTGGACGCGTTCGTGCAAAACGCGCATTTTTACGACAACGGTGAT
GCGCACGGCAAGGTTCTGAACGAAGTGATTATCACGATGGATAGCGCGGGACTGCAGCACGCATACTTAGATGTCGTCACCCAGTGGCTGACGC
GCTGGAACGCTCAATACGCTGAGCACTGGTATATGCGTAATGCGGTAAACGGCGTTTTTACCCTGTTGTTTGGTGGACAGTGGAATAATCAGTA
CACCTCTCTGATTGGTGAGCAGACGGCATTGGTGACCGCACTGCAAGCTTTCGCTCTGGACCGCACCAAAGTAAACAGCCCGACAGAGTTCATG
GCAGCTAATGCTGCACGTGAACTGGGTCGTCTGGCGCGTTATACGGACGCTACCATTGCGCCAAAGGTCACTGAGGGCTTAACCGCCATCTTTG
GCCAGTACCCGAGCTACGGCGATGGCGACGCGATCTGGCTCGGCGCGGCGGACACCGCTTCTTATTACGCAGACTGTAGCCAATTCAACATCTG

CGGTTTTGAAGATGCGCTCCGGGATGCGGCGTTGAATCAAACCTTTATCTGCAGCGACACGATCAAGATCCGATCCCAGGACATGAGCCAGGCG
CAACACCTGGCGGCCTGCGACAAAATGGCTTACGAGGAAAGCTTCTTCCACACCACTCTGGAGACAGGTAACCAGCCGGTTGCAGATGATCATA
ACACCCAACTGCAGGTGAACATTTTCAACAGCGACACTGACTACGGTAAGTATGCTGGTCCGATTTTCGGTATTGATACGAACAACGGTGGCAT
GTACCTCGAGGGCAATCCGGCAAACGTTGGTAATATTCCGAATTTTATCGCCTATGAGGCGTCCTACGCCAACCCGGACCACTTCGTGTGGAAC
CTGGAACACGAATACGTTCATTATCTGGACGGCCGTTTTAACATGTATGGCGACTTCGGTACGCCGACCGAGCTGGTGGTTTGGTGGTCTGAAG
GCGTGGCTGAATACGTGAGCCGTGTTAACGATAATCCGCAGGCGATTGCGACCATTCAGGATGGTTCCACCTATACTTTGGCGCAAGTTTTTGA
TACCACCTATGATGGTTTCGATGTGGATCGCATTTATCGTTGGGGTTACCTGGCGGTGCGTTTCATGTTCGAGCGCCATCCGGATGAGGTTCAG
CGTATGCTGTCGGCCACGCGTCAGGGTCGTTGGGCGGAGTATAAAGCTATCATCAGCGGTTGGGCAAATCAATACCAGAGCGAGTTTGCACAGT
GGACCGAGGCGTTGGCGAAAGGTGACAGCGGCGCAGGCAACGGTGAAGGCACCGGCTCCGGCAACGAAGGCGGCGGTGAGTCCGGTGGTAACAC
CGGTTTGCCGGAGAATTGTGCAGTCCTGCCGAAGATCTCGGATGGTCGCCTGGCGCTCGATGAAGCTGCGTGCCTGGCGGATACGGCGTCCGCG
AGCGATGTTCTTTGGTTTAGCATTCCGGCAGTCTCCGAATACCAGACCATCGCCATCACCGCAGGCAACGGCACCGGTGACCTGACCTTGGAGT
ACAGCAACCTGAACTGGCCTGACGGCACGAACGTGCAAGCCTCCTCAGCCAACATGGGTAACAGCGAGTGCATTATCCTGGAGCACCAGGCTAA
TTATTGGGGCTATTTGAAGGTGTCCGGCTCTTTTGAGAACGCGGCATTGTTGGTTGAAGCTGGTTCAAACCAATGCCGTCAA**GAGCTC**

B. cereus - Collagenase ColA

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGAAGAGACAGCTCCCTATAATATACTACAAATGAAACCGATGGGCACCGAGACCTCCAAGGACGAAATCGTTCATGCCACCAAGGCGG
ATGAGACCTTGAACTTCGAGGAACGATTGAAAATCGGTGACTTCAGCCAGAGACCTACCCTGGTAATGAAACGTGATGAAATTCAACTGAAGCA
AAGCTACACCCTGGCTGAACTGAATAAAATGCCGAACAGCGAACTGATCGATACCCTCTCGAAAATTTCCTGGAACCAAATTACTGATCTGTTT
CAATTCAATCAGGACACCAAAGCGTTTTACCAGAACAAAGAGCGGATGAATGTGATTATCAACGAGCTGGGACAGCGTGGTCGCACCTTTACTA
AAGAAAATAGCAAGGGCATTGAAACGTTCGTCGAAGTCTTGCGCAGCGCGTTCTACGTGGGCTACTATAACAACGAGCTGTCGTACCTTAAGGA
ACGCAGCTTCCACGATAAATGTCTGCCGGCGCTGAAAGCGATTGCCAAGAACCGTAATTTTACCCTCGGCACTGCGGAACAAGATCGTGTGGTT
ACCGCTTATGGCAAGCTGATTGGCAACGCTAGCTCCGACACCGAGACCGTTCAGTATGCCGTTAATGTGCTGAAGCACTACAACGACAACCTTA
GCACCTACGTTAGCGACTACGCGAAAGGTCAGGCGGTGTATGAGATCGTGAAGGGTATCGATTACGACATCCAAAGCTATCTGCAAGATACCAA
CAAGCAGCCGAACGAAACCATGTGGTATGGCAAAATTGACAACTTTATCAACGAGGTGAACCGTATTGCATTGGTCGGTAACATCACCAACGAG
AACAGCTGGCTGATCAATAACGGTATCTATTACGCGGGTCGTCTGGGCAAATTCCATAGCAATCCGAATAAGGGTCTGGAAGTGATCACACAGG
CTATGAACTTGTACCCGCGTCTGTCGGGCGCATACTTCGTGGCGGTAGAGCAGATGAAAACCAACTACGGCGGTAAAGACTATTCTGGCAACGC
TGTGGACCTGCAAAAGATCCGTGAGGAGGGCAAGCAGCAGTACCTGCCGAAAACCTATACCTTTGACGATGGTTCCATTGTTTTCAAAACCGGT
GACAAAGTGACTGAGGAAAAGATCAAGCGCCTGTATTGGGCAGCGAAGGAGGTGAAGGCGCAGTATCACCGTGTTATCGGTAACGACAAAGCAC
TTGAGCCGGGTAACGCTGATGACGTTTTGACCATTGTGATCTACAATAATCCGGACGAATACCAATTGAACCGCCAGCTCTACGGGTACGAAAC
GAATAACGGCGGCATATATATTGAAGAGAAGGGCACGTTCTTTACGTACGAGCGCACCCCGAAGCAGTCTATTTACTCTCTGGAAGAGCTGTTT
CGTCACGAATTCACCCATTATTTGCAAGGTCGTTACGAGGTCCCGGGATTATTTGGTTCTGGCGAGATGTATCAGAATGAGCGCTTGACCTGGT
TTCAAGAAGGCAACGCGGAATTCTTCGCGGGTAGCACGCGTACCAACAACGTTGTTCCGCGTAAAAGCATGATTAGCGGCCTGAGCAGCGACCC
GGCGAGCCGTTATACCGTCAAGCAGACCCTGTTTAGCAAGTACGGTAGCTGGGACTTTTACAAGTACTCGTTCGCTCTGCAATCATACTTATAC
AATCACCAGTTTGAGACCTTCGATAAACTGCAGGACCTGATTCGCGCGAACGACGTCAAGAACTATGATAGCTACCGCGAATCTTTGTCCAATA
ATACCCAGCTGAATGCCGAATATCAAGCGTATATGCAACAGCTGATTGATAATCAGGATAAGTACAACGTACCAAAAGTTACCAATGATTATCT
TATCCAACACGCACCGAAGCCGCTGGCGGAGGTGAAAAATGAAATCGTTGACGTGGCGAATATCAAAGATGCGAAAATCACAAAATATGAGAGC
CAGTTCTTTAACACGTTCACGGTGGAAGGTAAGTACACTGGTGGTACGTCCAAAGGCGAATCCGAAGATTGGAAAGCAATGAGCAAGCAGGTTA
ACCAGACGTTGGAACAGCTTTCTCAAAAAGGTTGGAGCGGTTATAAGACCGTTACCGCATATTTTGTGAACTACCGTGTCAATGCCGCAAATCA
ATTCGAGTACGATATCGTTTTCCACGGTGTTGCTACCGAGGAGAAAGAGAAAACTACGACCATCGTGAACATGAACGGTCCGTACTCTGGCATC
GTTAATGAAGAAATTCAGTTTCACAGCGATGGTACAAAGTCCGAGAACGGCAAGGTGATTAGCTATTTATGGAACTTCGGCGACGGCACGACCA
GTACGGAAGCTAATCCGACCCATGTTTATGGTGAAAAAGGCACGTACACTGTTGAGCTAACGGTAAAGGACAGTCGTGGTAAGGAGTCTAAAGA
GCAGACCAAGGTGACGGTAAAACAAGATCCACAGACCAGCGAATCCTATGAAGAGGAGAAGGTTCTGCCATTTAACACCTTGGTTAAAGGTAAC
CTGATCACCCCGGATCAAACCGACGTTTATACCTTTAATGTGACTGACCCTAAAGAAGTTGATATCAGCGTTGTCAACGAGCAAAACATTGGCA
TGACCTGGGTTCTGTACCATGAATCTGATATGCAAAACTACGTGGCCTGCGGTGAGGACGAGGGCAACGTGATCAAGGGAAAATTCGCCGCCAA
GCCGGGTAAATACTACCTGAATGTGTATAAATTCGACGATAAGAACGGCGAATACTCCCTGCTGGTTAAA**GAGCTC**

B. licheniformis - Subtilisin Carlsberg

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGCTCAACCCGCAAAGAATGTAGAAAAAGATTACATCGTTGGCTTCAAAAGCGGCGTTAAGACAGCGAGCGTTAAGAAAGATATCATTA
AAGAGAGCGGCGGCAAAGTTGACAAACAATTTCGTATTATCAACGCAGCGAAAGCAAAGCTGGACAAAGAGGCTCTGAAAGAGGTGAAGAACGA
CCCGGACGTTGCGTACGTTGAGGAAGATCACGTTGCGCATGCACTGGCGCAAACCGTCCCGTATGGTATCCCGCTGATTAAGGCCGATAAAGTA
CAGGCACAGGGTTTCAAGGGTGCGAATGTGAAGGTGGCTGTTTTAGACACCGGTATCCAGGCGAGCCACCCGGATTTGAATGTTGTGGGTGGTG
CCTCTTTTGTTGCTGGCGAAGCGTACAATACCGATGGTAACGGTCATGGTACGCATGTGGCGGGCACGGTTGCTGCTTTGGACAATACCACCGG
CGTGTTGGGCGTGGCGCCAAGCGTGTCTCTTTATGCCGTTAAGGTGCTCAACAGCTCTGGTTCGGGCACTTACAGCGGTATTGTCTCAGGCATC
GAATGGGCAACCACGAACGGTATGGATGTGATCAACATGAGCCTGGGCGGTCCGTCCGGCAGCACCGCAATGAAACAAGCGGTGGACAACGCCT
ACGCACGTGGTGTCGTGGTTGTGGCCGCGGCGGGTAACAGCGGATCTTCCGGGAATACCAACACCATTGGTTATCCGGCCAAGTATGATTCCGT
GATTGCCGTAGGTGCTGTGGACTCCAACAGCAACCGCGCAAGCTTCAGCTCCGTGGGGGCGGAATTGGAAGTGATGGCCCCTGGTGCGGGTGTT
TATTCTACCTACCCGACCAGCACCTACGCAACCCTGAACGGCACGTCCATGGCATCGCCGCACGTGGCGGGTGCGGCGGCTCTGATCCTGTCGA
AACACCCGAATCTGAGCGCTAGCCAGGTTCGCAACCGTCTGTCGAGTACCGCGACTTACCTGGGCAGCTCTTTTTATTACGGCAAGGGCCTGAT
TAATGTCGAGGCCGCGGCGCAG**GAGCTC**

C. lupus familiaris - Pepsinogen B

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGTAGAAAGGATAATTCTAAAGAAAGGAAAATCTATCCGCCAGGTCATGGAAGAAAGAGGTGTTTTGGAGACATTCTTACGTAATCATC
CGAAGGTGGACCCGGCAGCGAAATATTTGTTCAACAACGATGCGGTTGCTTACGAGCCGTTCACCAATTATCTTGACAGCTATTACTTCGGTGA
GATCTCCATCGGCACCCCGCCTCAAAACTTCCTGATTTTGTTCGACACTGGCTCCTCTAACCTGTGGGTTCCGTCTACGTACTGCCAGTCTCAG
GCATGTAGCAACCACAACCGTTTTAACCCGAGCCGTAGCTCGACGTACCAGTCGAGCGAGCAAACGTACACCTTAGCGTACGGCTTTGGCAGCC
TGACGGTACTGCTGGGTTATGATACCGTGACCGTTCAGAACATTGTGATCCACAACCAGCTGTTTGGTATGAGCGAGAATGAACCGAACTATCC
GTTCTACTACTCCTATTTCGACGGCATCCTGGGTATGGCGTATTCCAACCTGGCTGTAGATAATGGTCCAACTGTCTTGCAAAACATGATGCAG
CAAGGTCAGCTGACCCAGCCCATCTTTTCATTCTACTTCAGCCCGCAACCGACGTACGAATATGGTGGCGAGCTGATTCTGGGTGGCGTTGACA
CCCAATTTTATAGCGGTGAGATCGTGTGGGCACCGGTTACCCGTGAAATGTACTGGCAGGTCGCCATTGATGAATTTCTGATCGGTAATCAAGC
GACCGGCCTGTGCAGCCAGGGGTGCCAGGGTATCGTTGACACCGGAACCTTCCCGTTGACCGTGCCGCAACAATATCTCGACAGCTTTGTTAAA
GCAACCGGTGCGCAGCAAGATCAGAGTGGTAATTTTGTGGTTAATTGTAATAGCATCCAGTCCATGCCGACGATTACCTTCGTGATTAGCGGCA
GCCCGTTGCCACTTCCGCCAAGCACCTACGTGTTGAACAACAACGGCTATTGCACCTTAGGCATTGAAGTCACATACCTGCCGTCGCCGAACGG
CCAACCTCTGTGGATTCTGGGCGATGTTTTTCTGCGTGAGTACTACACCGTGTTCGACATGGCCGCTAATCGCGTGGGTTTTGCGCTGAGCAGC
GAGCTC

B. taurus - Cationic Trypsin

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGTAGATGATGACGATAAAATAGTTGGAGGGTACACGTGCGGCGCGAATACCGTTCCGTATCAGGTTAGCTTGAACTCCGGCTACCATT
TTTGCGGTGGTAGCTTGATCAACTCTCAGTGGGTTGTTTCCGCGGCGCACTGCTATAAAAGCGGGATCCAAGTGCGCCTGGGTGAAGATAATAT
TAACGTGGTCGAGGGCAACGAGCAGTTCATCTCTGCTTCGAAATCGATCGTCCACCCAAGCTACAATTCCAACACCCTGAACAACGACATCATG
CTGATTAAGCTGAAATCTGCCGCGTCCTTAAACAGCCGTGTTGCAAGCATCAGCCTTCCGACCTCATGCGCTTCTGCGGGCACCCAGTGTCTGA
TCAGCGGATGGGGTAATACCAAGAGCAGCGGCACGAGCTACCCGGACGTGCTGAAGTGCTTGAAGGCCCCTATTCTGTCCGATTCCAGCTGTAA
ATCTGCGTATCCGGGTCAAATTACCTCCAACATGTTCTGCGCAGGTTACCTGGAAGGCGGTAAGGACAGCTGCCAAGGTGATAGCGGCGGTCCG
GTGGTATGTAGCGGCAAACTGCAAGGTATCGTGTCTTGGGGCAGCGGTTGTGCGCAGAAAAACAAACCGGGTGTTTACACTAAGGTGTGCAATT
ATGTTAGTTGGATTAAGCAGACCATTGCAAGCAAT**GAGCTC**

C. papaya - **Papain**

N-terminal restriction site, KpmI: GGTACC
C-terminal restriction site, SacI: GAGCTC

GGTACCGTATATATGGGACTATCATTTGGGGATTTCTCCATCGTGGGCTACTCGCAAAACGACCTGACCAGCACCGAACGTCTGATTCAGTTGT
TCGAGAGCTGGATGCTGAAGCACAACAAGATCTACAAAAACATCGACGAGAAGATTTATCGTTTTGAGATCTTCAAGGACAACCTGAAATATAT
CGATGAAACCAATAAGAAAAACAACTCCTACTGGCTGGGTCTGAATGTTTTTGCGGATATGAGCAATGACGAATTTAAAGAAAAATACACCGGT
TCCATTGCGGGTAACTATACTACGACTGAGTTAAGCTACGAGGAAGTTCTGAACGACGGCGATGTGAATATTCCGGAGTACGTGGACTGGCGTC
AAAAAGGTGCGGTCACCCCGGTGAAAAACCAAGGTAGTTGCGGTTCTTGCTGGGCATTTAGCGCGGTGGTGACCATTGAAGGCATCATCAAAAT
CCGCACCGGCAACCTGAACGAGTATAGCGAACAGGAGTTGCTTGATTGTGATCGTAGAAGCTACGGGTGCAATGGTGGTTACCCGTGGTCAGCG
CTCCAGTTGGTTGCGCAGTATGGCATCCACTATCGCAATACCTATCCGTATGAGGGTGTCCAACGTTATTGCCGTAGCCGTGAAAAGGGCCCCT
ACGCGGCGAAAACCGATGGCGTTCGTCAGGTTCAACCGTATAACGAGGGCGCACTGCTGTACTCGATCGCCAATCAACCGGTCAGCGTTGTTCT
GGAAGCTGCTGGTAAAGACTTCCAGTTGTACCGCGGTGGTATTTTCGTGGGTCCGTGCGGCAACAAGGTGGATCATGCAGTCGCTGCCGTTGGT
TACGGCCCAAATTATATTCTGATTAAGAACAGCTGGGGTACAGGCTGGGGTGAAAATGGCTACATCCGTATTAAGCGCGGTACGGGCAACTCCT
ACGGCGTATGTGGTCTGTACACGTCTAGCTTCTATCCGGTTAAGAAC**GAGCTC**

Enzymes expressed in *E. coli* with pET22b+ expression vector with pelB signal sequences

The DNA sequences presented below contain N- and C-terminal restriction sites (in bold). The native signal peptides have been removed as pelB signal sequence will be utilized for periplasmic protein expression as well as the native Stop codon as the C-terminal His_6-tag contains Stop codon.

C. histolyticum - Collagenase ColG

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, a glycine (highlighted in yellow) was inserted. The first amino acid (Lys) of the ColG propeptide could not be fit to pelB without changing to another amino acid hence why the neutral glycine was added.

CCATGGGTAAACCCATAGAAAACACAAATGATACTTCAATAAAGAACGTGGAAAAGTTACGTAATGCGCCAAACGAAGAGAACTCTAAGAAGGT
CGAGGATTCAAAAAACGACAAAGTGGAGCACGTGAAAAATATCGAGGAGGCGAAAGTCGAGCAGGTTGCGCCAGAAGTTAAGTCCAAGTCGACG
CTGCGTAGCGCGTCGATCGCGAACACCAATAGCGAGAAATACGATTTCGAGTACCTGAACGGCCTGTCTTACACCGAATTGACCAATCTGATTA
AGAACATCAAGTGGAATCAAATTAACGGCTTGTTTAACTATAGCACCGGCAGTCAGAAATTCTTTGGTGACAAGAACCGTGTTCAAGCAATCAT
TAACGCTCTGCAAGAGAGCGGCCGTACCTATACCGCCAATGACATGAAAGGTATCGAGACATTCACCGAAGTTCTGAGAGCGGGTTTCTATCTG
GGTTATTATAACGATGGTCTGAGCTACCTTAACGATCGTAATTTCCAGGATAAATGTATTCCGGCAATGATTGCGATTCAGAAAAACCCGAACT
TCAAATTGGGCACGGCTGTGCAAGATGAAGTCATCACCAGCCTGGGTAAGCTGATCGGTAACGCGTCTGCAAATGCGGAAGTGGTAAATAACTG
CGTTCCGGTTTTGAAGCAGTTCAGGGAAAATCTGAACCAGTATGCACCGGATTACGTAAAGGGTACCGCGGTTAACGAACTGATTAAGGGCATC
GAGTTCGATTTCAGCGGTGCCGCATATGAGAAGGACGTGAAAACGATGCCTTGGTATGGCAAAATTGACCCGTTTATCAACGAGCTTAAGGCGC
TGGGCCTCTACGGCAATATCACCAGTGCTACTGAGTGGGCATCGGATGTGGGTATTTATTACTTGAGCAAGTTTGGTTTGTACTCTACGAACCG
CAATGACATTGTTCAAAGCTTGGAAAAGGCGGTGGACATGTATAAATACGGCAAAATCGCGTTTGTTGCGATGGAGCGTATCACCTGGGATTAT
GACGGCATTGGTAGCAATGGCAAAAAGGTGGATCACGATAAATTCCTGGATGACGCCGAGAAGCACTACCTCCCGAAAACCTACACCTTCGATA
ACGGCACGTTCATCATCCGTGCAGGCGACAAAGTGAGCGAAGAGAAGATCAAACGCCTGTACTGGGCAAGCCGTGAGGTGAAATCTCAGTTTCA
TCGTGTGGTCGGCAACGACAAAGCGCTGGAAGTGGGCAACGCCGATGATGTTCTAACGATGAAGATCTTCAACTCCCCGGAGGAATACAAGTTC
AACACCAATATCAACGGCGTTAGCACGGATAACGGCGGTCTTTACATTGAGCCGCGTGGCACGTTCTACACGTACGAGCGCACCCCGCAGCAGA
GCATTTTTTCTCTTGAGGAGTTGTTTCGTCATGAATACACCCATTATCTGCAAGCCCGTTACTTAGTCGACGGCCTGTGGGGTCAGGGCCCATT
TTACGAAAAGAACCGTCTGACTTGGTTTGACGAAGGCACTGCGGAATTCTTTGCTGGAAGCACTCGCACCTCTGGCGTCCTGCCGCGTAAAAGC
ATTCTGGGTTATCTGGCGAAGGACAAGGTTGACCACCGCTATAGCCTTAAGAAGACCCTGAATAGCGGCTACGACGACTCCGACTGGATGTTCT
ACAATTACGGTTTCGCGGTGGCGCATTATCTGTATGAGAAAGACATGCCGACCTTCATTAAGATGAATAAGGCCATTCTGAACACCGATGTTAA
ATCCTACGACGAGATTATCAAAAAATTGTCGGACGACGCGAACAAAAACACCGAATATCAGAATCACATCCAAGAACTCGCCGATAAGTACCAG
GGCGCGGGTATTCCGCTGGTGAGCGACGACTACTTAAAAGACCACGGTTATAAAAAGGCTTCGGAGGTGTATTCCGAGATCAGCAAGGCGGCAA
GCTTAACCAACACCTCCGTTACCGCTGAAAAAAGCCAATATTTCAATACCTTTACGTTGCGCGGTACCTACACCGGTGAAACCAGCAAGGGCGA
ATTCAAGGACTGGGATGAAATGAGCAAGAAACTGGATGGCACACTGGAAAGCCTGGCTAAAAATAGCTGGAGCGGTTACAAGACCCTCACTGCT
TACTTCACCAATTATCGTGTTACGTCCGACAACAAGGTTCAATATGATGTGGTTTTCCACGGCGTTCTGACTGACAATGCTGATATCTCTAACA
ATAAAGCGCCGATTGCGAAGGTAACGGGTCCGAGCACCGGTGCGGTTGGCCGCAACATCGAGTTTAGCGGTAAGGACTCGAAGGACGAGGACGG
TAAGATCGTTAGCTATGACTGGGATTTTGGCGATGGTGCTACCAGCCGTGGTAAAAACTCCGTGCATGCATATAAAAAGGCGGGTACGTATAAT
GTTACTCTGAAAGTCACCGATGATAAAGGTGCGACCGCTACTGAAAGCTTTACCATTGAAATTAAAAACGAAGACACCACAACCCCGATTACCA
AAGAAATGGAACCGAATGACGACATCAAGGAGGCGAATGGTCCGATCGTCGAGGGCGTGACCGTTAAGGGCGACCTGAACGGGTCGGACGACGC
GGATACCTTCTACTTCGACGTGAAGGAGGATGGCGACGTTACCATTGAGCTGCCGTATAGTGGTTCCAGCAACTTCACCTGGCTGGTGTACAAA
GAAGGTGATGACCAGAACCACATTGCTTCTGGCATTGATAAAAACAACAGCAAAGTCGGAACATTTAAAAGCACCAAAGGTCGCCACTACGTTT
TCATCTATAAGCACGATAGTGCAAGCAACATCAGCTATTCCTTGAACATCAAAGGTCTGGGTAACGAGAAGCTCAAGGAGAAGGAGAACAACGA
TTCCTCCGACAAAGCTACCGTGATCCCGAATTTTAATACGACCATGCAGGGTAGCTTGCTGGGTGACGATTCTAGAGATTACTATAGCTTTGAA
GTGAAGGAGGAGGGTGAAGTTAATATTGAGCTGGATAAGAAAGATGAGTTTGGAGTTACTTGGACCCTGCATCCGGAAAGCAATATCAACGATC
GTATTACCTACGGTCAAGTGGATGGTAACAAGGTGTCCAACAAAGTAAAATTGAGACCTGGTAAATATTACCTGCTGGTTTACAAATACTCTGG
CAGCGGGAACTACGAACTCCGTGTGAACAAA**CTCGAG**

C. histolyticum - Collagenase ColH

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of alanine (GCT) has been removed. This does not alter the gene sequence.

CCATGGCTGTAGATAAAAATAACGCTACAGCGGCAGTGCAGAATGAAAGCAAGAGATACACTGTGTCCTACTTGAAGACCCTGAACTACTATGA
CCTGGTTGATCTGCTGGTTAAGACGGAGATCGAAAACTTGCCGGACCTCTTCCAGTATAGCTCCGATGCGAAGGAGTTCTACGGTAATAAAACA
CGCATGAGCTTCATAATGGATGAAATCGGTCGTCGTGCGCCACAGTACACCGAAATCGACCACAAAGGGATTCCAACCCTGGTCGAGGTTGTTC
GTGCGGGTTTTTACCTGGGTTTCCACAATAAAGAGTTGAACGAAATTAACAAGAGAAGCTTCAAAGAACGTGTCATCCCGAGCATTCTGGCGAT
CCAGAAAAATCCAAATTTTAAGTTGGGAACGGAGGTTCAGGATAAGATTGTGAGTGCGACCGGTCTGCTGGCGGGTAACGAGACGGCGCCGCCT
GAGGTTGTGAATAACTTCACCCCGATTCTGCAAGACTGTATTAAAAACATCGACCGTTACGCCCTGGATGACTTGAAGTCCAAGGCACTGTTCA
ACGTGCTGGCCGCACCGACCTACGACATCACCGAGTATCTTCGCGCTACCAAGGAAAAGCCGGAGAACACCCCTTGGTATGGTAAAATCGACGG
CTTTATCAACGAATTGAAAAAACTCGCTCTGTACGGTAAGATCAATGACAATAATTCTTGGATTATCGATAATGGCATCTATCATATTGCTCCG
CTGGGTAAGCTGCATAGCAACAACAAGATTGGTATTGAAACCCTGACCGAAGTGATGAAAGTGTACCCGTATCTGTCTATGCAGCACCTGCAAA
GCGCTGATCAAATTAAGCGTCACTATGATTCCAAGGATGCGGAAGGAAATAAAATTCCGCTGGATAAATTCAAGAAAGAGGGTAAAGAGAAGTA
CTGCCCGAAAACCTACACCTTCGACGATGGCAAAGTGATTATCAAGGCTGGTGCGCGTGTCGAAGAAGAGAAGGTTAAACGCTTATACTGGGCT
AGCAAAGAGGTCAATTCGCAATTTTTTCGTGTTTATGGTATCGACAAACCGCTTGAGGAAGGCAACCCGGATGATATCCTGACGATGGTGATCT
ACAACTCACCGGAGGAGTATAAACTGAACAGCGTGCTGTACGGCTACGACACGAACAACGGTGGCATGTATATCGAACCGGAGGGTACCTTTTT
CACTTATGAGCGCGAAGCACAGGAGAGCACCTATACGCTGGAGGAGCTGTTCCGCCACGAATATACCCATTATCTGCAAGGTCGTTACGCGGTT
CCAGGCCAATGGGGTCGTACCAAACTCTACGACAACGACCGTCTGACCTGGTACGAGGAAGGTGGTGCGGAATTGTTCGCCGGCTCGACCCGAA
CCAGCGGCATCCTGCCGCGTAAATCTATCGTTTCCAATATTCATAACACCACCCGTAATAATCGTTACAAGCTGAGCGATACCGTGCACAGCAA
ATACGGCGCGAGCTTTGAATTTTACAATTATGCGTGCATGTTTATGGATTACATGTACAACAAGGACATGGGTATTCTTAACAAACTGAATGAC
TTGGCCAAAAACAACGATGTTGATGGTTACGACAACTATATTCGTGATCTGAGCTCCAACTACGCCCTTAACGATAAGTATCAGGACCACATGC
AGGAGCGCATCGACAACTATGAAAATCTGACGGTTCCGTTTGTGGCAGACGACTACCTGGTACGTCACGCGTATAAGAACCCGAATGAGATCTA
TAGCGAGATTAGCGAGGTTGCAAAATTAAAGGATGCAAAATCTGAGGTCAAAAAGTCCCAGTACTTCTCTACATTTACTCTGCGTGGTAGCTAT
ACTGGTGGTGCCAGCAAGGGCAAGTTGGAGGATCAAAAAGCCATGAATAAGTTCATCGACGACAGCTTGAAGAAGCTGGACACGTACTCGTGGT
CAGGCTACAAAACCCTGACCGCGTATTTTACGAATTACAAAGTGGATAGCAGCAACCGCGTGACGTACGACGTTGTTTTTCATGGTTATCTGCC
GAATGAGGGCGATTCGAAAAACAGCCTTCCGTATGGCAAGATCAACGGGACGTATAAAGGCACGGAGAAAGAAAAGATAAAGTTCAGCTCGGAG
GGTAGTTTCGATCCAGACGGCAAAATTGTCTCATATGAATGGGATTTTGGTGATGGCAACAAGAGCAATGAAGAGAACCCCGAGCACAGCTACG
ATAAAGTGGGCACTTACACTGTCAAGCTGAAGGTCACCGATGATAAGGGTGAATCTAGCGTGTCGACCACCACCGCGGAGATCAAAGACTTGTC
CGAGAACAAACTGCCGGTTATCTACATGCATGTTCCGAAATCTGGTGCTTTGAACCAGAAAGTTGTGTTCTACGGCAAGGGCACCTACGACCCG
GATGGCTCAATAGCGGGTTATCAATGGGATTTTGGCGACGGCTCCGACTTCAGCAGCGAACAGAACCCGAGCCATGTGTATACCAAGAAAGGTG
AATACACCGTAACCCTGCGCGTGATGGATAGCTCCGGCCAAATGAGCGAGAAAACCATGAAGATCAAAATTACCGACCCGGTTTATCCGATTGG
TACCGAAAAAGAGCCGAACAACAGCAAAGAAACTGCGAGCGGTCCGATCGTCCCGGGTATTCCGGTGTCAGGCACAATTGAGAACACCTCCGAT
CAAGACTATTTTTATTTCGACGTGATTACCCCGGGTGAAGTGAAGATCGACATTAATAAGTTAGGTTACGGCGGTGCTACCTGGGTTGTCTATG
ACGAAAACAACAACGCGGTTTCCTATGCGACGGACGACGGCCAGAATCTCAGCGGTAAGTTCAAAGCGGACAAACCGGGTCGCTACTACATCCA
CCTATACATGTTTAACGGAAGTTACATGCCGTATCGTATTAACATCGAAGGCTCTGTGGGCCGT**CTCGAG**

G. hollisae - Collagenase

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, a glycine (highlighted in yellow) was inserted. The first amino acid (Ser) of the ColG propeptide could not be fit to pelB without changing to another amino acid hence why the neutral glycine was added.

CCATGGGTTCAGAAGTATTTGCTCACCCCGGAATGCCAGTCCAGCAGAAATTACAGCCTAATTTGCTGCAGCAGAGCACCCGTCTCCAACCGGA
GCAACACATTCATGGTCTGGAGCGTACCGACCGGCAATACCGTCCGACCGACGCGACCCAGCAACCAGAACCGCCAACCCTGTTGAAGCGCCAG
GTTAGCGTTCAGCAGGATGCGGTCGAACAGTGTGATCTGAGCCAGTTTCAGACCACCAGCAGCAATCAGCTGATGGCAGCTATTCGTCAACAAG
GTGCCAGCTGTGTTAATGCTTTGTTCAGCGCTGACACGGGTGTGCAAGAAGCTGCGTTCTCCTCTAATCATATGTACAACGTGGCCCAATACAC
CAGAACGCTTGCACAACAATACGCGGGCGGCGGCAGCGACGAACTGGAAGCCCTTTACCTGTACCTGCGTGCAGGCTACTATGCGGAATTCTAT
AACTCCAACATCACCTTTCTGTCTTGGGTTACCCCTGCGGTTAAAGGCGCCGTTGATGCATTTGTGCAGAATGCGCACTTTTATGACAACGGCG

ATGCACACGGCAAGGTGCTGAATGAGGTTATTATTACGATGGACTCCGCCGGCCTGCAACACGCATACTTGGACGTTGTCACGCAGTGGCTGAC
CCGCTGGAATGCGCAATACGCCGAGCATTGGTATATGCGTAATGCGGTTAACGGCGTTTTCACCCTGCTGTTTGGTGGACAATGGAACAACCAA
TATACGAGCCTGATTGGTGAGCAGACCGCGCTGGTGACCGCCCTCCAAGCGTTCGCCCTTGACCGCACCAAAGTTAACTCGCCGACAGAGTTCA
TGGCAGCCAACGCTGCCCGTGAGCTGGGTCGCCTGGCACGCTATACCGACGCTACCATTGCACCGAAAGTGACCGAAGGTTTAACGGCGATCTT
CGGCCAGTACCCGAGCTACGGCGATGGCGACGCCATCTGGCTGGGTGCGGCGGACACCGCGTCCTACTACGCGGATTGCTCCCAATTCAACATT
TGTGGTTTCGAGGATGCGCTGCGTGATGCGGCTCTAAACCAGACCTTCATCTGCAGCGACACGATCAAGATCCGTTCGCAAGACATGTCCCAGG
CTCAACACCTGGCTGCCTGCGACAAAATGGCGTACGAAGAAAGCTTTTTTCATACCACGTTGGAAACAGGTAACCAGCCGGTTGCTGATGATCA
CAACACCCAGCTGCAAGTTAATATCTTTAACTCCGACACTGATTATGGCAAGTATGCAGGTCCGATTTTCGGTATTGATACGAACAATGGTGGC
ATGTATCTGGAGGGCAACCCGGCAAACGTGGGTAATATCCCGAATTTTATCGCTTATGAAGCGTCATACGCGAACCCGGACCACTTCGTGTGGA
ATCTGGAACACGAATACGTGCATTATCTGGATGGACGCTTCAACATGTATGGCGATTTTGGCACCCCGACTGAGCTGGTAGTTTGGTGGTCAGA
GGGCGTGGCAGAGTACGTGAGCCGAGTCAATGATAATCCGCAGGCAATCGCGACCATTCAGGACGGTTCTACGTACACTTTGGCGCAAGTGTTC
GATACGACCTATGATGGTTTCGACGTTGATCGTATTTATCGTTGGGGCTATCTGGCGGTTCGTTTTATGTTTGAACGTCATCCGGACGAGGTGC
AACGTATGCTGAGCGCGACCCGCCAGGGTCGTTGGGCGGAATACAAAGCAATCATCAGCGGTTGGGCAAACCAATACCAGAGCGAGTTCGCTCA
GTGGACCGAGGCATTGGCCAAAGGTGACAGCGGTGCGGGCAACGGCGAGGGCACCGGTAGCGGCAACGAAGGTGGTGGCGAAAGCGGCGGGAAC
ACCGGTTTACCGGAAAATTGCGCAGTCCTGCCGAAGATCTCTGACGGCCGTTTGGCGTTGGATGAGGCGGCCTGCCTGGCGGACACTGCGTCCG
CTAGTGATGTGCTGTGGTTTTCTATTCCGGCGGTGTCGGAGTATCAAACCATTGCGATCACCGCGGGTAATGGTACTGGTGACCTGACCTTGGA
GTACAGCAACCTGAACTGGCCGGACGGCACCAACGTTCAGGCTAGCTCTGCGAACATGGGTAACAGCGAGTGCATCATCCTAGAGCACCAGGCG
AACTATTGGGGTTACCTGAAGGTGTCTGGTTCCTTTGAAAACGCCGCTTTGCTTGTCGAGGCAGGTTCTAACCAGTGCCGTCAG**CTCGAG**

B. cereus - Collagenase ColA

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of Glu (GAA) has been removed. This does not alter the gene sequence.

CCATGGGAAGAGACAGCTCCCTATAATATACTACAAATGAAGCCGATGGGTACTGAGACCAGCAAAGACGAAATTGTTCATGCAACCAAAGCGG
ATGAGACACTGAATTTCGAAGAGCGCTTGAAGATCGGCGATTTTAGCCAGCGTCCGACTCTGGTAATGAAAAGAGATGAAATTCAGCTGAAGCA
GTCCTACACGCTGGCAGAACTGAACAAAATGCCGAATAGCGAGCTGATCGATACCCTATCCAAGATCTCCTGGAACCAGATTACCGACCTGTTC
CAATTCAACCAAGACACCAAAGCGTTTTACCAGAATAAAGAGCGCATGAATGTTATAATCAACGAACTGGGTCAGCGTGGTCGTACGTTTACCA
AGGAGAACAGCAAGGGCATCGAAACCTTCGTGGAAGTCTTACGTTCTGCATTTTACGTTGGTTACTACAATAATGAACTGAGTTACCTGAAAGA
GCGCAGCTTTCACGATAAGTGCCTGCCGGCACTGAAGGCAATTGCGAAGAACCGCAACTTCACGTTGGGTACTGCGGAACAGGACCGTGTTGTT
ACGGCGTATGGCAAACTCATTGGTAACGCGTCGTCAGACACCGAAACTGTACAGTACGCGGTTAACGTTTTAAAACATTATAACGATAACCTGA
GCACCTATGTTAGCGACTACGCGAAGGGTCAGGCGGTGTATGAAATTGTCAAAGGCATCGACTACGACATTCAAAGCTACTTGCAAGATACGAA
CAAGCAACCGAATGAAACAATGTGGTATGGTAAAATCGACAATTTTATCAACGAAGTGAACCGTATTGCCTTGGTTGGTAATATCACCAACGAA
AACTCTTGGCTGATTAACAACGGCATTTACTATGCCGGTCGTCTGGGTAAATTCCACAGCAACCCAAATAAAGGTCTGGAAGTCATTACCCAGG
CTATGAACCTGTACCCGCGCCTATCCGGCGCATACTTCGTGGCGGTGGAACAGATGAAGACGAATTACGGCGGTAAGGACTACAGCGGTAATGC
GGTTGATCTGCAGAAAATCCGCGAGGAAGGTAAACAGCAATATCTGCCGAAGACGTACACTTTTGACGACGGTTCTATTGTTTTTAAAACTGGT
GACAAGGTGACCGAGGAGAAAATTAAACGTTTGTACTGGGCTGCGAAAGAGGTGAAGGCTCAGTACCACCGTGTTATCGGCAACGATAAGGCGT
TGGAACCGGGTAACGCGGATGACGTTCTCACCATCGTGATCTACAACAATCCCGATGAGTATCAGTTGAATCGTCAGCTTTATGGTTATGAAAC
CAACAATGGTGGAATCTATATCGAAGAAAAGGGCACGTTCTTCACCTACGAACGCACCCCGAAACAAAGCATTTATTCCTTGGAGGAACTGTTT
CGTCATGAATTCACCCACTACCTCCAAGGTCGTTATGAGGTTCCGGGTCTGTTTGGCAGCGGTGAGATGTATCAGAACGAGAGACTGACCTGGT
TTCAAGAAGGCAACGCGGAGTTTTTCGCCGGTTCAACGCGTACCAACAACGTTGTTCCACGTAAAAGCATGATTTCCGGCCTGAGCAGCGACCC
GGCTAGCCGTTATACGGTGAAGCAAACGCTCTTCAGCAAGTACGGTTCCTGGGATTTCTATAAATACTCGTTCGCGCTGCAAAGCTATTTATAC
AACCACCAGTTCGAAACCTTTGATAAATTGCAAGACCTGATTCGCGCAAACGACGTCAAGAACTACGACAGCTATCGTGAATCTCTGTCTAACA
ACACCCAGCTGAATGCGGAGTACCAAGCTTATATGCAACAGCTCATCGACAACCAGGATAAATATAACGTGCCGAAAGTGACCAATGATTACTT
AATTCAACATGCTCCGAAACCGCTGGCAGAGGTGAAAAACGAGATCGTTGATGTTGCCAATATTAAAGACGCTAAAATTACCAAATACGAGTCC
CAATTTTTTAACACCTTCACCGTCGAGGGCAAGTACACCGGCGGCACCAGCAAGGGTGAATCCGAGGATTGGAAAGCCATGAGCAAGCAGGTTA
ACCAGACCTTGGAGCAACTGTCGCAAAAAGGATGGTCAGGCTACAAAACGGTGACGGCCTACTTCGTGAATTACCGCGTTAACGCTGCGAACCA
GTTCGAGTATGACATCGTGTTTCACGGCGTCGCGACCGAAGAGAAAGAGAAAACCACCACCATCGTCAATATGAACGGCCCTTATTCCGGCATC
GTGAATGAGGAGATCCAATTCCACAGCGATGGCACCAAGTCCGAGAATGGTAAGGTGATCAGCTATCTGTGGAATTTCGGTGATGGTACGACCA
GCACGGAAGCGAATCCGACGCACGTGTATGGCGAGAAGGGTACGTATACCGTTGAACTGACTGTAAAGGACAGCCGTGGTAAGGAGTCCAAGGA
GCAAACCAAAGTTACCGTTAAACAAGACCCGCAGACCTCTGAAAGCTATGAGGAAGAAAAGGTGCTGCCATTTAATACCCTTGTGAAAGGCAAC
CTGATTACTCCGGATCAGACAGATGTGTACACCTTCAACGTCACTGATCCGAAAGAGGTGGACATCAGCGTGGTTAACGAGCAGAATATTGGTA

TGACCTGGGTTCTGTACCATGAGTCGGATATGCAGAACTACGTGGCGTGTGGTGAAGATGAGGGCAACGTCATCAAAGGCAAGTTCGCCGCGAA
GCCGGGTAAGTACTACCTTAACGTTTATAAATTTGACGACAAGAACGGCGAATACTCTTTGCTGGTGAAG**CTCGAG**

B. licheniformis - Subtilisin Carlsberg

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of Ala (GCT) has been removed. This does not alter the gene sequence.

CCATGGCTCAACCCGCAAAGAATGTAGAAAAAGATTACATTGTTGGTTTTAAAAGCGGCGTGAAGACTGCGTCCGTGAAGAAGGACATCATCAA
AGAGAGCGGCGGTAAAGTGGACAAACAGTTTCGTATTATTAACGCAGCGAAGGCGAAACTGGACAAGGAGGCGCTCAAAGAGGTTAAGAACGAT
CCGGACGTTGCCTATGTTGAGGAAGATCACGTGGCGCATGCACTGGCGCAAACCGTGCCGTATGGTATCCCGCTGATTAAAGCCGATAAGGTGC
AGGCCCAAGGTTTTAAAGGTGCAAATGTTAAGGTGGCGGTCTTGGACACCGGTATCCAAGCGAGCCACCCAGATCTGAATGTTGTTGGTGGCGC
GTCGTTCGTGGCCGGCGAAGCGTACAATACCGATGGTAACGGCCACGGCACCCATGTTGCGGGCACTGTAGCGGCTTTGGACAACACCACGGGT
GTCTTAGGCGTGGCTCCGAGTGTTTCACTGTATGCCGTCAAGGTGTTGAATTCCTCTGGCAGCGGTACTTACAGCGGCATCGTGAGCGGGATCG
AATGGGCAACCACCAACGGCATGGATGTTATCAACATGAGCCTGGGTGGTCCGAGTGGTTCCACCGCAATGAAACAGGCGGTTGACAATGCGTA
CGCGCGTGGTGTGGTGGTGGTTGCTGCCGCGGGTAACAGCGGCAGCTCCGGCAACACCAATACCATTGGTTATCCGGCAAAGTATGATAGCGTT
ATTGCAGTCGGCGCGGTGGACAGCAACAGCAACCGCGCATCTTTCAGCTCTGTTGGCGCCGAACTTGAGGTTATGGCACCGGGTGCGGGCGTCT
ATAGCACGTACCCGACCAGCACGTACGCGACCCTGAATGGTACGAGCATGGCGTCTCCGCATGTTGCTGGCGCCGCTGCTTTGATTCTGAGCAA
ACACCCCAACCTGTCTGCAAGCCAGGTACGTAATCGCCTGTCCTCGACGGCTACCTACCTGGGTTCCTCCTTCTATTACGGCAAAGGTCTGATC
AACGTGGAAGCAGCGGCTCAG**CTCGAG**

C. lupus familiaris - Pepsinogen B

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of Val (GTA) has been removed. This does not alter the gene sequence.

CCATGGTAGAAAGGATAATTCTAAAGAAAGGAAAATCCATCCGTCAGGTTATGGAAGAACGTGGTGTTCTGGAAACCTTTCTGCGCAACCACCC
GAAGGTGGATCCGGCGGCGAAATATCTCTTCAACAACGATGCGGTGGCCTACGAGCCGTTTACCAATTATCTCGACTCCTATTACTTCGGTGAG
ATCAGCATCGGTACTCCGCCTCAAAACTTCCTGATCTTGTTCGACACCGGTTCTTCTAACCTTTGGGTTCCGTCCACCTACTGCCAATCGCAAG
CGTGTAGCAATCATAACCGCTTCAACCCGAGCCGTAGCTCGACCTACCAGTCTTCCGAACAGACCTACACCCTGGCGTACGGCTTTGGTTCACT
GACCGTTCTTCTGGGTTATGATACCGTCACGGTCCAGAATATTGTGATTCACAACCAGCTGTTCGGCATGAGCGAGAACGAACCGAATTATCCG
TTTTACTACAGCTATTTCGATGGTATCTTGGGCATGGCGTATAGCAACCTGGCTGTGGACAACGGTCCGACCGTCCTGCAAAACATGATGCAGC
AAGGTCAGCTGACGCAGCCGATTTTTAGCTTCTACTTCAGCCCACAGCCAACGTACGAGTATGGCGGTGAGCTGATCTTGGGAGGCGTTGACAC
CCAATTTTACAGCGGCGAGATCGTGTGGGCACCGGTTACTCGCGAGATGTATTGGCAGGTTGCGATTGATGAATTTCTGATTGGTAACCAGGCG
ACCGGTCTGTGCAGCCAAGGCTGCCAGGGCATCGTGGACACCGGTACATTCCCGCTGACGGTGCCGCAACAATATTTGGACAGCTTTGTTAAAG
CTACCGGCGCACAGCAGGATCAGAGTGGTAATTTCGTGGTGAATTGTAATAGCATTCAATCCATGCCGACAATTACCTTTGTGATCTCCGGCAG
CCCGCTGCCGTTACCACCGTCTACGTACGTTTTGAACAACAACGGCTATTGCACCTTGGGCATCGAAGTTACGTACCTGCCGTCGCCGAATGGC
CAGCCGTTATGGATTCTGGGTGATGTTTTCCTGCGTGAGTATTACACCGTGTTTGACATGGCCGCTAATCGTGTAGGTTTTGCACTGAGCTCG**C**
TCGAG

B. taurus - Cationic Trypsin

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of Val (GTA) has been removed. This does not alter the gene sequence.

CCATGGTAGATGACGATGATAAAATAGTTGGAGGGTACACGTGCGGCGCAAATACCGTTCCGTATCAGGTATCCCTGAATTCTGGCTACCACTT
CTGCGGTGGCAGCTTGATCAACTCGCAATGGGTTGTCAGCGCTGCGCACTGCTACAAAAGCGGGATCCAAGTGCGCCTGGGTGAAGATAATATT
AACGTGGTGGAGGGCAACGAGCAGTTCATCTCTGCGTCGAAATCTATTGTTCATCCGAGCTACAATAGCAACACTCTGAACAACGACATCATGC
TGATCAAGCTGAAGTCGGCGGCGTCCTTGAACTCTCGTGTTGCCTCCATTAGCCTTCCGACCAGTTGCGCCAGCGCTGGCACCCAGTGTCTGAT
CTCCGGTTGGGGTAATACCAAGTCCTCCGGCACCTCCTACCCGGATGTTCTGAAGTGCTTGAAGGCACCGATTCTGAGCGACAGCTCATGCAAA
AGCGCGTATCCAGGTCAGATCACCTCCAACATGTTTTGTGCGGGTTACCTGGAAGGCGGTAAGGACAGCTGCCAGGGTGATAGCGGCGGTCCGG
TCGTGTGCAGCGGCAAACTCCAAGGTATCGTGAGCTGGGGTAGCGGTTGTGCGCAAAAAAACAAACCGGGCGTTTATACGAAAGTTTGTAATTA
TGTGAGCTGGATTAAGCAGACCATTGCATCTAAC**CTCGAG**

C. papaya - Papain

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG
To keep in frame with the pelB signal sequence, The first G of Val (GTA) has been removed. This does not alter the gene sequence.

CCATGGTATATATGGGACTATCATTTGGGGATTTCTCGATTGTTGGTTACTCCCAGAATGACCTGACCTCAACCGAACGTCTGATTCAGCTGTT
TGAGAGCTGGATGCTGAAGCACAATAAGATCTACAAAAACATCGACGAGAAGATTTATCGCTTCGAGATCTTCAAGGACAACCTTAAATACATT
GATGAAACCAATAAGAAAAACAACAGCTACTGGCTGGGTCTGAATGTTTTTGCGGATATGAGCAACGACGAGTTCAAAGAAAAGTATACTGGCT
CTATAGCGGGTAACTATACCACGACTGAGTTAAGCTACGAGGAAGTTTTGAACGACGGCGACGTTAATATTCCGGAATATGTGGACTGGCGTCA
AAAAGGCGCGGTGACCCCGGTCAAAAACCAGGGTTCGTGCGGTTCTTGCTGGGCGTTTAGCGCCGTTGTGACCATTGAGGGTATCATCAAAATC
AGAACAGGCAACCTGAACGAGTACAGCGAACAGGAGCTGCTGGATTGTGATCGCCGTTCCTACGGCTGCAATGGTGGATATCCGTGGAGCGCGC
TGCAGCTGGTGGCGCAATATGGCATCCACTATCGCAACACCTATCCTTATGAGGGCGTACAACGTTATTGCCGTTCTCGTGAAAAGGGTCCGTA
CGCAGCGAAAACCGATGGCGTTCGTCAAGTGCAGCCGTACAATGAAGGCGCTCTGCTGTACAGCATCGCTAATCAACCGGTTTCCGTCGTGTTG
GAAGCAGCCGGTAAAGACTTCCAGTTGTACCGCGGTGGTATTTTCGTGGGTCCGTGTGGTAACAAAGTGGATCATGCAGTCGCTGCCGTGGGTT
ACGGCCCAAATTACATCCTCATCAAGAACTCCTGGGGCACCGGCTGGGGTGAAAACGGTTACATCCGTATTAAACGTGGTACGGGCAACAGCTA
TGGTGTTTGCGGTTTGTACACGAGCAGCTTTTATCCGGTCAAGAAC**CTCGAG**

Enzymes expressed in *E. coli* with pET22b+ expression vector for cytoplasmic expression

The DNA sequences presented below contain N- and C-terminal restriction sites (in bold). The native signal peptides have been removed as the proteins are to be expressed in the cytoplasm. A methionine (highlighted in yellow) has been added after every N-terminal restriction site to induce protein expression. The DNA sequence is not in frame with the pelB signal peptide but in frame with the C-terminal His_6-tag contains Stop codon.

C. histolyticum - Collagenase ColG

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGAAACCCATAGAAAACACAAATGATACTTCAATAAAGAACGTGGAAAAGTTACGTAATGCGCCAAACGAAGAGAACTCTAAGAAGG
TCGAGGATTCAAAAAACGACAAAGTGGAGCACGTGAAAAATATCGAGGAGGCGAAAGTCGAGCAGGTTGCGCCAGAAGTTAAGTCCAAGTCGAC
GCTGCGTAGCGCGTCGATCGCGAACACCAATAGCGAGAAATACGATTTCGAGTACCTGAACGGCCTGTCTTACACCGAATTGACCAATCTGATT
AAGAACATCAAGTGGAATCAAATTAACGGCTTGTTTAACTATAGCACCGGCAGTCAGAAATTCTTTGGTGACAAGAACCGTGTTCAAGCAATCA
TTAACGCTCTGCAAGAGAGCGGCCGTACCTATACCGCCAATGACATGAAAGGTATCGAGACATTCACCGAAGTTCTGAGAGCGGGTTTCTATCT
GGGTTATTATAACGATGGTCTGAGCTACCTTAACGATCGTAATTTCCAGGATAAATGTATTCCGGCAATGATTGCGATTCAGAAAAACCCGAAC
TTCAAATTGGGCACGGCTGTGCAAGATGAAGTCATCACCAGCCTGGGTAAGCTGATCGGTAACGCGTCTGCAAATGCGGAAGTGGTAAATAACT
GCGTTCCGGTTTTGAAGCAGTTCAGGGAAAATCTGAACCAGTATGCACCGGATTACGTAAAGGGTACCGCGGTTAACGAACTGATTAAGGGCAT
CGAGTTCGATTTCAGCGGTGCCGCATATGAGAAGGACGTGAAAACGATGCCTTGGTATGGCAAAATTGACCCGTTTATCAACGAGCTTAAGGCG
CTGGGCCTCTACGGCAATATCACCAGTGCTACTGAGTGGGCATCGGATGTGGGTATTTATTACTTGAGCAAGTTTGGTTTGTACTCTACGAACC
GCAATGACATTGTTCAAAGCTTGGAAAAGGCGGTGGACATGTATAAATACGGCAAAATCGCGTTTGTTGCGATGGAGCGTATCACCTGGGATTA
TGACGGCATTGGTAGCAATGGCAAAAAGGTGGATCACGATAAATTCCTGGATGACGCCGAGAAGCACTACCTCCCGAAAACCTACACCTTCGAT
AACGGCACGTTCATCATCCGTGCAGGCGACAAAGTGAGCGAAGAGAAGATCAAACGCCTGTACTGGGCAAGCCGTGAGGTGAAATCTCAGTTTC
ATCGTGTGGTCGGCAACGACAAAGCGCTGGAAGTGGGCAACGCCGATGATGTTCTAACGATGAAGATCTTCAACTCCCCGGAGGAATACAAGTT
CAACACCAATATCAACGGCGTTAGCACGGATAACGGCGGTCTTTACATTGAGCCGCGTGGCACGTTCTACACGTACGAGCGCACCCCGCAGCAG
AGCATTTTTTCTCTTGAGGAGTTGTTTCGTCATGAATACACCCATTATCTGCAAGCCCGTTACTTAGTCGACGGCCTGTGGGGTCAGGGCCCAT
TTTACGAAAAGAACCGTCTGACTTGGTTTGACGAAGGCACTGCGGAATTCTTTGCTGGAAGCACTCGCACCTCTGGCGTCCTGCCGCGTAAAAG
CATTCTGGGTTATCTGGCGAAGGACAAGGTTGACCACCGCTATAGCCTTAAGAAGACCCTGAATAGCGGCTACGACGACTCCGACTGGATGTTC
TACAATTACGGTTTCGCGGTGGCGCATTATCTGTATGAGAAAGACATGCCGACCTTCATTAAGATGAATAAGGCCATTCTGAACACCGATGTTA
AATCCTACGACGAGATTATCAAAAAATTGTCGGACGACGCGAACAAAAACACCGAATATCAGAATCACATCCAAGAACTCGCCGATAAGTACCA
GGGCGCGGGTATTCCGCTGGTGAGCGACGACTACTTAAAAGACCACGGTTATAAAAAGGCTTCGGAGGTGTATTCCGAGATCAGCAAGGCGGCA
AGCTTAACCAACACCTCCGTTACCGCTGAAAAAAGCCAATATTTCAATACCTTTACGTTGCGCGGTACCTACACCGGTGAAACCAGCAAGGGCG
AATTCAAGGACTGGGATGAAATGAGCAAGAAACTGGATGGCACACTGGAAAGCCTGGCTAAAAATAGCTGGAGCGGTTACAAGACCCTCACTGC
TTACTTCACCAATTATCGTGTTACGTCCGACAACAAGGTTCAATATGATGTGGTTTTCCACGGCGTTCTGACTGACAATGCTGATATCTCTAAC
AATAAAGCGCCGATTGCGAAGGTAACGGGTCCGAGCACCGGTGCGGTTGGCCGCAACATCGAGTTTAGCGGTAAGGACTCGAAGGACGAGGACG
GTAAGATCGTTAGCTATGACTGGGATTTTGGCGATGGTGCTACCAGCCGTGGTAAAAACTCCGTGCATGCATATAAAAAGGCGGGTACGTATAA
TGTTACTCTGAAAGTCACCGATGATAAAGGTGCGACCGCTACTGAAAGCTTTACCATTGAAATTAAAAACGAAGACACCACAACCCCGATTACC
AAAGAAATGGAACCGAATGACGACATCAAGGAGGCGAATGGTCCGATCGTCGAGGGCGTGACCGTTAAGGGCGACCTGAACGGGTCGGACGACG
CGGATACCTTCTACTTCGACGTGAAGGAGGATGGCGACGTTACCATTGAGCTGCCGTATAGTGGTTCCAGCAACTTCACCTGGCTGGTGTACAA
AGAAGGTGATGACCAGAACCACATTGCTTCTGGCATTGATAAAAACAACAGCAAAGTCGGAACATTTAAAAGCACCAAAGGTCGCCACTACGTT
TTCATCTATAAGCACGATAGTGCAAGCAACATCAGCTATTCCTTGAACATCAAAGGTCTGGGTAACGAGAAGCTCAAGGAGAAGGAGAACAACG
ATTCCTCCGACAAAGCTACCGTGATCCCGAATTTTAATACGACCATGCAGGGTAGCTTGCTGGGTGACGATTCTAGAGATTACTATAGCTTTGA
AGTGAAGGAGGAGGGTGAAGTTAATATTGAGCTGGATAAGAAAGATGAGTTTGGAGTTACTTGGACCCTGCATCCGGAAAGCAATATCAACGAT
CGTATTACCTACGGTCAAGTGGATGGTAACAAGGTGTCCAACAAAGTAAAATTGAGACCTGGTAAATATTACCTGCTGGTTTACAAATACTCTG
GCAGCGGGAACTACGAACTCCGTGTGAACAAA**CTCGAG**

C. histolyticum - Collagenase ColH

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGCTGTAGATAAAAATAACGCTACAGCGGCAGTGCAGAATGAAAGCAAGAGATACACTGTGTCCTACTTGAAGACCCTGAACTACT
ATGACCTGGTTGATCTGCTGGTTAAGACGGAGATCGAAAACTTGCCGGACCTCTTCCAGTATAGCTCCGATGCGAAGGAGTTCTACGGTAATAA
AACACGCATGAGCTTCATAATGGATGAAATCGGTCGTCGTGCGCCACAGTACACCGAAATCGACCACAAAGGGATTCCAACCCTGGTCGAGGTT
GTTCGTGCGGGTTTTTACCTGGGTTTCCACAATAAAGAGTTGAACGAAATTAACAAGAGAAGCTTCAAAGAACGTGTCATCCCGAGCATTCTGG
CGATCCAGAAAAATCCAAATTTTAAGTTGGGAACGGAGGTTCAGGATAAGATTGTGAGTGCGACCGGTCTGCTGGCGGGTAACGAGACGGCGCC
GCCTGAGGTTGTGAATAACTTCACCCCGATTCTGCAAGACTGTATTAAAAACATCGACCGTTACGCCCTGGATGACTTGAAGTCCAAGGCACTG
TTCAACGTGCTGGCCGCACCGACCTACGACATCACCGAGTATCTTCGCGCTACCAAGGAAAAGCCGGAGAACACCCCTTGGTATGGTAAAATCG
ACGGCTTTATCAACGAATTGAAAAAACTCGCTCTGTACGGTAAGATCAATGACAATAATTCTTGGATTATCGATAATGGCATCTATCATATTGC
TCCGCTGGGTAAGCTGCATAGCAACAACAAGATTGGTATTGAAACCCTGACCGAAGTGATGAAAGTGTACCCGTATCTGTCTATGCAGCACCTG
CAAAGCGCTGATCAAATTAAGCGTCACTATGATTCCAAGGATGCGGAAGGAAATAAAATTCCGCTGGATAAATTCAAGAAAGAGGGTAAAGAGA
AGTACTGCCCGAAAACCTACACCTTCGACGATGGCAAAGTGATTATCAAGGCTGGTGCGCGTGTCGAAGAAGAGAAGGTTAAACGCTTATACTG
GGCTAGCAAAGAGGTCAATTCGCAATTTTTTCGTGTTTATGGTATCGACAAACCGCTTGAGGAAGGCAACCCGGATGATATCCTGACGATGGTG
ATCTACAACTCACCGGAGGAGTATAAACTGAACAGCGTGCTGTACGGCTACGACACGAACAACGGTGGCATGTATATCGAACCGGAGGGTACCT
TTTTCACTTATGAGCGCGAAGCACAGGAGAGCACCTATACGCTGGAGGAGCTGTTCCGCCACGAATATACCCATTATCTGCAAGGTCGTTACGC
GGTTCCAGGCCAATGGGGTCGTACCAAACTCTACGACAACGACCGTCTGACCTGGTACGAGGAAGGTGGTGCGGAATTGTTCGCCGGCTCGACC
CGAACCAGCGGCATCCTGCCGCGTAAATCTATCGTTTCCAATATTCATAACACCACCCGTAATAATCGTTACAAGCTGAGCGATACCGTGCACA
GCAAATACGGCGCGAGCTTTGAATTTTACAATTATGCGTGCATGTTTATGGATTACATGTACAACAAGGACATGGGTATTCTTAACAAACTGAA
TGACTTGGCCAAAAACAACGATGTTGATGGTTACGACAACTATATTCGTGATCTGAGCTCCAACTACGCCCTTAACGATAAGTATCAGGACCAC
ATGCAGGAGCGCATCGACAACTATGAAAATCTGACGGTTCCGTTTGTGGCAGACGACTACCTGGTACGTCACGCGTATAAGAACCCGAATGAGA
TCTATAGCGAGATTAGCGAGGTTGCAAAATTAAAGGATGCAAAATCTGAGGTCAAAAAGTCCCAGTACTTCTCTACATTTACTCTGCGTGGTAG
CTATACTGGTGGTGCCAGCAAGGGCAAGTTGGAGGATCAAAAAGCCATGAATAAGTTCATCGACGACAGCTTGAAGAAGCTGGACACGTACTCG
TGGTCAGGCTACAAAACCCTGACCGCGTATTTTACGAATTACAAAGTGGATAGCAGCAACCGCGTGACGTACGACGTTGTTTTTCATGGTTATC
TGCCGAATGAGGGCGATTCGAAAAACAGCCTTCCGTATGGCAAGATCAACGGGACGTATAAAGGCACGGAGAAAGAAAAGATAAAGTTCAGCTC
GGAGGGTAGTTTCGATCCAGACGGCAAAATTGTCTCATATGAATGGGATTTTGGTGATGGCAACAAGAGCAATGAAGAGAACCCCGAGCACAGC
TACGATAAAGTGGGCACTTACACTGTCAAGCTGAAGGTCACCGATGATAAGGGTGAATCTAGCGTGTCGACCACCACCGCGGAGATCAAAGACT
TGTCCGAGAACAAACTGCCGGTTATCTACATGCATGTTCCGAAATCTGGTGCTTTGAACCAGAAAGTTGTGTTCTACGGCAAGGGCACCTACGA
CCCGGATGGCTCAATAGCGGGTTATCAATGGGATTTTGGCGACGGCTCCGACTTCAGCAGCGAACAGAACCCGAGCCATGTGTATACCAAGAAA
GGTGAATACACCGTAACCCTGCGCGTGATGGATAGCTCCGGCCAAATGAGCGAGAAAACCATGAAGATCAAAATTACCGACCCGGTTTATCCGA
TTGGTACCGAAAAAGAGCCGAACAACAGCAAAGAAACTGCGAGCGGTCCGATCGTCCCGGGTATTCCGGTGTCAGGCACAATTGAGAACACCTC
CGATCAAGACTATTTTTATTTCGACGTGATTACCCCGGGTGAAGTGAAGATCGACATTAATAAGTTAGGTTACGGCGGTGCTACCTGGGTTGTC
TATGACGAAAACAACAACGCGGTTTCCTATGCGACGGACGACGGCCAGAATCTCAGCGGTAAGTTCAAAGCGGACAAACCGGGTCGCTACTACA
TCCACCTATACATGTTTAACGGAAGTTACATGCCGTATCGTATTAACATCGAAGGCTCTGTGGGCCGT**CTCGAG**

G. hollisae - Collagenase

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGTCAGAAGTATTTGCTCACCCCGGAATGCCAGTCCAGCAGAAATTACAGCCTAATTTGCTGCAGCAGAGCACCCGTCTCCAACCGG
AGCAACACATTCATGGTCTGGAGCGTACCGACCGGCAATACCGTCCGACCGACGCGACCCAGCAACCAGAACCGCCAACCCTGTTGAAGCGCCA
GGTTAGCGTTCAGCAGGATGCGGTCGAACAGTGTGATCTGAGCCAGTTTCAGACCACCAGCAGCAATCAGCTGATGGCAGCTATTCGTCAACAA
GGTGCCAGCTGTGTTAATGCTTTGTTCAGCGCTGACACGGGTGTGCAAGAAGCTGCGTTCTCCTCTAATCATATGTACAACGTGGCCCAATACA
CCAGAACGCTTGCACAACAATACGCGGGCGGCGGCAGCGACGAACTGGAAGCCCTTTACCTGTACCTGCGTGCAGGCTACTATGCGGAATTCTA
TAACTCCAACATCACCTTTCTGTCTTGGGTTACCCCTGCGGTTAAAGGCGCCGTTGATGCATTTGTGCAGAATGCGCACTTTTATGACAACGGC
GATGCACACGGCAAGGTGCTGAATGAGGTTATTATTACGATGGACTCCGCCGGCCTGCAACACGCATACTTGGACGTTGTCACGCAGTGGCTGA
CCCGCTGGAATGCGCAATACGCCGAGCATTGGTATATGCGTAATGCGGTTAACGGCGTTTTCACCCTGCTGTTTGGTGGACAATGGAACAACCA
ATATACGAGCCTGATTGGTGAGCAGACCGCGCTGGTGACCGCCCTCCAAGCGTTCGCCCTTGACCGCACCAAAGTTAACTCGCCGACAGAGTTC
ATGGCAGCCAACGCTGCCCGTGAGCTGGGTCGCCTGGCACGCTATACCGACGCTACCATTGCACCGAAAGTGACCGAAGGTTTAACGGCGATCT
TCGGCCAGTACCCGAGCTACGGCGATGGCGACGCCATCTGGCTGGGTGCGGCGGACACCGCGTCCTACTACGCGGATTGCTCCCAATTCAACAT

TTGTGGTTTCGAGGATGCGCTGCGTGATGCGGCTCTAAACCAGACCTTCATCTGCAGCGACACGATCAAGATCCGTTCGCAAGACATGTCCCAG
GCTCAACACCTGGCTGCCTGCGACAAAATGGCGTACGAAGAAAGCTTTTTTCATACCACGTTGGAAACAGGTAACCAGCCGGTTGCTGATGATC
ACAACACCCAGCTGCAAGTTAATATCTTTAACTCCGACACTGATTATGGCAAGTATGCAGGTCCGATTTTCGGTATTGATACGAACAATGGTGG
CATGTATCTGGAGGGCAACCCGGCAAACGTGGGTAATATCCCGAATTTTATCGCTTATGAAGCGTCATACGCGAACCCGGACCACTTCGTGTGG
AATCTGGAACACGAATACGTGCATTATCTGGATGGACGCTTCAACATGTATGGCGATTTTGGCACCCCGACTGAGCTGGTAGTTTGGTGGTCAG
AGGGCGTGGCAGAGTACGTGAGCCGAGTCAATGATAATCCGCAGGCAATCGCGACCATTCAGGACGGTTCTACGTACACTTTGGCGCAAGTGTT
CGATACGACCTATGATGGTTTCGACGTTGATCGTATTTATCGTTGGGGCTATCTGGCGGTTCGTTTTATGTTTGAACGTCATCCGGACGAGGTG
CAACGTATGCTGAGCGCGACCCGCCAGGGTCGTTGGGCGGAATACAAAGCAATCATCAGCGGTTGGGCAAACCAATACCAGAGCGAGTTCGCTC
AGTGGACCGAGGCATTGGCCAAAGGTGACAGCGGTGCGGGCAACGGCGAGGGCACCGGTAGCGGCAACGAAGGTGGTGGCGAAAGCGGCGGGAA
CACCGGTTTACCGGAAAATTGCGCAGTCCTGCCGAAGATCTCTGACGGCCGTTTGGCGTTGGATGAGGCGGCCTGCCTGGCGGACACTGCGTCC
GCTAGTGATGTGCTGTGGTTTTCTATTCCGGCGGTGTCGGAGTATCAAACCATTGCGATCACCGCGGGTAATGGTACTGGTGACCTGACCTTGG
AGTACAGCAACCTGAACTGGCCGGACGGCACCAACGTTCAGGCTAGCTCTGCGAACATGGGTAACAGCGAGTGCATCATCCTAGAGCACCAGGC
GAACTATTGGGGTTACCTGAAGGTGTCTGGTTCCTTTGAAAACGCCGCTTTGCTTGTCGAGGCAGGTTCTAACCAGTGCCGTCAG**CTCGAG**

B. cereus - Collagenase ColA

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGAAGAGACAGCTCCCTATAATATACTACAAATGAAGCCGATGGGTACTGAGACCAGCAAAGACGAAATTGTTCATGCAACCAAAG
CGGATGAGACACTGAATTTCGAAGAGCGCTTGAAGATCGGCGATTTTAGCCAGCGTCCGACTCTGGTAATGAAAAGAGATGAAATTCAGCTGAA
GCAGTCCTACACGCTGGCAGAACTGAACAAAATGCCGAATAGCGAGCTGATCGATACCCTATCCAAGATCTCCTGGAACCAGATTACCGACCTG
TTCCAATTCAACCAAGACACCAAAGCGTTTTACCAGAATAAAGAGCGCATGAATGTTATAATCAACGAACTGGGTCAGCGTGGTCGTACGTTTA
CCAAGGAGAACAGCAAGGGCATCGAAACCTTCGTGGAAGTCTTACGTTCTGCATTTTACGTTGGTTACTACAATAATGAACTGAGTTACCTGAA
AGAGCGCAGCTTTCACGATAAGTGCCTGCCGGCACTGAAGGCAATTGCGAAGAACCGCAACTTCACGTTGGGTACTGCGGAACAGGACCGTGTT
GTTACGGCGTATGGCAAACTCATTGGTAACGCGTCGTCAGACACCGAAACTGTACAGTACGCGGTTAACGTTTTAAAACATTATAACGATAACC
TGAGCACCTATGTTAGCGACTACGCGAAGGGTCAGGCGGTGTATGAAATTGTCAAAGGCATCGACTACGACATTCAAAGCTACTTGCAAGATAC
GAACAAGCAACCGAATGAAACAATGTGGTATGGTAAAATCGACAATTTTATCAACGAAGTGAACCGTATTGCCTTGGTTGGTAATATCACCAAC
GAAAACTCTTGGCTGATTAACAACGGCATTTACTATGCCGGTCGTCTGGGTAAATTCCACAGCAACCCAAATAAAGGTCTGGAAGTCATTACCC
AGGCTATGAACCTGTACCCGCGCCTATCCGGCGCATACTTCGTGGCGGTGGAACAGATGAAGACGAATTACGGCGGTAAGGACTACAGCGGTAA
TGCGGTTGATCTGCAGAAAATCCGCGAGGAAGGTAAACAGCAATATCTGCCGAAGACGTACACTTTTGACGACGGTTCTATTGTTTTTAAAACT
GGTGACAAGGTGACCGAGGAGAAAATTAAACGTTTGTACTGGGCTGCGAAAGAGGTGAAGGCTCAGTACCACCGTGTTATCGGCAACGATAAGG
CGTTGGAACCGGGTAACGCGGATGACGTTCTCACCATCGTGATCTACAACAATCCCGATGAGTATCAGTTGAATCGTCAGCTTTATGGTTATGA
AACCAACAATGGTGGAATCTATATCGAAGAAAAGGGCACGTTCTTCACCTACGAACGCACCCCGAAACAAAGCATTTATTCCTTGGAGGAACTG
TTTCGTCATGAATTCACCCACTACCTCCAAGGTCGTTATGAGGTTCCGGGTCTGTTTGGCAGCGGTGAGATGTATCAGAACGAGAGACTGACCT
GGTTTCAAGAAGGCAACGCGGAGTTTTTCGCCGGTTCAACGCGTACCAACAACGTTGTTCCACGTAAAAGCATGATTTCCGGCCTGAGCAGCGA
CCCGGCTAGCCGTTATACGGTGAAGCAAACGCTCTTCAGCAAGTACGGTTCCTGGGATTTCTATAAATACTCGTTCGCGCTGCAAAGCTATTTA
TACAACCACCAGTTCGAAACCTTTGATAAATTGCAAGACCTGATTCGCGCAAACGACGTCAAGAACTACGACAGCTATCGTGAATCTCTGTCTA
ACAACACCCAGCTGAATGCGGAGTACCAAGCTTATATGCAACAGCTCATCGACAACCAGGATAAATATAACGTGCCGAAAGTGACCAATGATTA
CTTAATTCAACATGCTCCGAAACCGCTGGCAGAGGTGAAAAACGAGATCGTTGATGTTGCCAATATTAAAGACGCTAAAATTACCAAATACGAG
TCCCAATTTTTTAACACCTTCACCGTCGAGGGCAAGTACACCGGCGGCACCAGCAAGGGTGAATCCGAGGATTGGAAAGCCATGAGCAAGCAGG
TTAACCAGACCTTGGAGCAACTGTCGCAAAAAGGATGGTCAGGCTACAAAACGGTGACGGCCTACTTCGTGAATTACCGCGTTAACGCTGCGAA
CCAGTTCGAGTATGACATCGTGTTTCACGGCGTCGCGACCGAAGAGAAAGAGAAAACCACCACCATCGTCAATATGAACGGCCCTTATTCCGGC
ATCGTGAATGAGGAGATCCAATTCCACAGCGATGGCACCAAGTCCGAGAATGGTAAGGTGATCAGCTATCTGTGGAATTTCGGTGATGGTACGA
CCAGCACGGAAGCGAATCCGACGCACGTGTATGGCGAGAAGGGTACGTATACCGTTGAACTGACTGTAAAGGACAGCCGTGGTAAGGAGTCCAA
GGAGCAAACCAAAGTTACCGTTAAACAAGACCCGCAGACCTCTGAAAGCTATGAGGAAGAAAAGGTGCTGCCATTTAATACCCTTGTGAAAGGC
AACCTGATTACTCCGGATCAGACAGATGTGTACACCTTCAACGTCACTGATCCGAAAGAGGTGGACATCAGCGTGGTTAACGAGCAGAATATTG
GTATGACCTGGGTTCTGTACCATGAGTCGGATATGCAGAACTACGTGGCGTGTGGTGAAGATGAGGGCAACGTCATCAAAGGCAAGTTCGCCGC
GAAGCCGGGTAAGTACTACCTTAACGTTTATAAATTTGACGACAAGAACGGCGAATACTCTTTGCTGGTGAAG**CTCGAG**

B. licheniformis - Subtilisin Carlsberg

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGCTCAACCCGCAAAGAATGTAGAAAAAGATTACATTGTTGGTTTTAAAAGCGGCGTGAAGACTGCGTCCGTGAAGAAGGACATCA
TCAAAGAGAGCGGCGGTAAAGTGGACAAACAGTTTCGTATTATTAACGCAGCGAAGGCGAAACTGGACAAGGAGGCGCTCAAAGAGGTTAAGAA
CGATCCGGACGTTGCCTATGTTGAGGAAGATCACGTGGCGCATGCACTGGCGCAAACCGTGCCGTATGGTATCCCGCTGATTAAAGCCGATAAG
GTGCAGGCCCAAGGTTTTAAAGGTGCAAATGTTAAGGTGGCGGTCTTGGACACCGGTATCCAAGCGAGCCACCCAGATCTGAATGTTGTTGGTG
GCGCGTCGTTCGTGGCCGGCGAAGCGTACAATACCGATGGTAACGGCCACGGCACCCATGTTGCGGGCACTGTAGCGGCTTTGGACAACACCAC
GGGTGTCTTAGGCGTGGCTCCGAGTGTTTCACTGTATGCCGTCAAGGTGTTGAATTCCTCTGGCAGCGGTACTTACAGCGGCATCGTGAGCGGG
ATCGAATGGGCAACCACCAACGGCATGGATGTTATCAACATGAGCCTGGGTGGTCCGAGTGGTTCCACCGCAATGAAACAGGCGGTTGACAATG
CGTACGCGCGTGGTGTGGTGGTGGTTGCTGCCGCGGGTAACAGCGGCAGCTCCGGCAACACCAATACCATTGGTTATCCGGCAAAGTATGATAG
CGTTATTGCAGTCGGCGCGGTGGACAGCAACAGCAACCGCGCATCTTTCAGCTCTGTTGGCGCCGAACTTGAGGTTATGGCACCGGGTGCGGGC
GTCTATAGCACGTACCCGACCAGCACGTACGCGACCCTGAATGGTACGAGCATGGCGTCTCCGCATGTTGCTGGCGCCGCTGCTTTGATTCTGA
GCAAACACCCCAACCTGTCTGCAAGCCAGGTACGTAATCGCCTGTCCTCGACGGCTACCTACCTGGGTTCCTCCTTCTATTACGGCAAAGGTCT
GATCAACGTGGAAGCAGCGGCTCAG**CTCGAG**

C. lupus familiaris - Pepsinogen B

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGTAGAAAGGATAATTCTAAAGAAAGGAAAATCCATCCGTCAGGTTATGGAAGAACGTGGTGTTCTGGAAACCTTTCTGCGCAACC
ACCCGAAGGTGGATCCGGCGGCGAAATATCTCTTCAACAACGATGCGGTGGCCTACGAGCCGTTTACCAATTATCTCGACTCCTATTACTTCGG
TGAGATCAGCATCGGTACTCCGCCTCAAAACTTCCTGATCTTGTTCGACACCGGTTCTTCTAACCTTTGGGTTCCGTCCACCTACTGCCAATCG
CAAGCGTGTAGCAATCATAACCGCTTCAACCCGAGCCGTAGCTCGACCTACCAGTCTTCCGAACAGACCTACACCCTGGCGTACGGCTTTGGTT
CACTGACCGTTCTTCTGGGTTATGATACCGTCACGGTCCAGAATATTGTGATTCACAACCAGCTGTTCGGCATGAGCGAGAACGAACCGAATTA
TCCGTTTTACTACAGCTATTTCGATGGTATCTTGGGCATGGCGTATAGCAACCTGGCTGTGGACAACGGTCCGACCGTCCTGCAAAACATGATG
CAGCAAGGTCAGCTGACGCAGCCGATTTTTAGCTTCTACTTCAGCCCACAGCCAACGTACGAGTATGGCGGTGAGCTGATCTTGGGAGGCGTTG
ACACCCAATTTTACAGCGGCGAGATCGTGTGGGCACCGGTTACTCGCGAGATGTATTGGCAGGTTGCGATTGATGAATTTCTGATTGGTAACCA
GGCGACCGGTCTGTGCAGCCAAGGCTGCCAGGGCATCGTGGACACCGGTACATTCCCGCTGACGGTGCCGCAACAATATTTGGACAGCTTTGTT
AAAGCTACCGGCGCACAGCAGGATCAGAGTGGTAATTTCGTGGTGAATTGTAATAGCATTCAATCCATGCCGACAATTACCTTTGTGATCTCCG
GCAGCCCGCTGCCGTTACCACCGTCTACGTACGTTTTGAACAACAACGGCTATTGCACCTTGGGCATCGAAGTTACGTACCTGCCGTCGCCGAA
TGGCCAGCCGTTATGGATTCTGGGTGATGTTTTCCTGCGTGAGTATTACACCGTGTTTGACATGGCCGCTAATCGTGTAGGTTTTGCACTGAGC
TCG**CTCGAG**

B. taurus - Cationic Trypsin

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGTAGATGACGATGATAAAATAGTTGGAGGGTACACGTGCGGCGCAAATACCGTTCCGTATCAGGTATCCCTGAATTCTGGCTACC
ACTTCTGCGGTGGCAGCTTGATCAACTCGCAATGGGTTGTCAGCGCTGCGCACTGCTACAAAAGCGGGATCCAAGTGCGCCTGGGTGAAGATAA
TATTAACGTGGTGGAGGGCAACGAGCAGTTCATCTCTGCGTCGAAATCTATTGTTCATCCGAGCTACAATAGCAACACTCTGAACAACGACATC
ATGCTGATCAAGCTGAAGTCGGCGGCGTCCTTGAACTCTCGTGTTGCCTCCATTAGCCTTCCGACCAGTTGCGCCAGCGCTGGCACCCAGTGTC
TGATCTCCGGTTGGGGTAATACCAAGTCCTCCGGCACCTCCTACCCGGATGTTCTGAAGTGCTTGAAGGCACCGATTCTGAGCGACAGCTCATG
CAAAAGCGCGTATCCAGGTCAGATCACCTCCAACATGTTTTGTGCGGGTTACCTGGAAGGCGGTAAGGACAGCTGCCAGGGTGATAGCGGCGGT
CCGGTCGTGTGCAGCGGCAAACTCCAAGGTATCGTGAGCTGGGGTAGCGGTTGTGCGCAAAAAAACAAACCGGGCGTTTATACGAAAGTTTGTA
ATTATGTGAGCTGGATTAAGCAGACCATTGCATCTAAC**CTCGAG**

C. papaya - Papain

N-terminal restriction site, NcoI: CCATGG
C-terminal restriction site, XhoI: CTCGAG

CCATGGATGGTATATATGGGACTATCATTTGGGGATTTCTCGATTGTTGGTTACTCCCAGAATGACCTGACCTCAACCGAACGTCTGATTCAGC
TGTTTGAGAGCTGGATGCTGAAGCACAATAAGATCTACAAAAACATCGACGAGAAGATTTATCGCTTCGAGATCTTCAAGGACAACCTTAAATA
CATTGATGAAACCAATAAGAAAAACAACAGCTACTGGCTGGGTCTGAATGTTTTTGCGGATATGAGCAACGACGAGTTCAAAGAAAAGTATACT
GGCTCTATAGCGGGTAACTATACCACGACTGAGTTAAGCTACGAGGAAGTTTTGAACGACGGCGACGTTAATATTCCGGAATATGTGGACTGGC
GTCAAAAAGGCGCGGTGACCCCGGTCAAAAACCAGGGTTCGTGCGGTTCTTGCTGGGCGTTTAGCGCCGTTGTGACCATTGAGGGTATCATCAA
AATCAGAACAGGCAACCTGAACGAGTACAGCGAACAGGAGCTGCTGGATTGTGATCGCCGTTCCTACGGCTGCAATGGTGGATATCCGTGGAGC
GCGCTGCAGCTGGTGGCGCAATATGGCATCCACTATCGCAACACCTATCCTTATGAGGGCGTACAACGTTATTGCCGTTCTCGTGAAAAGGGTC
CGTACGCAGCGAAAACCGATGGCGTTCGTCAAGTGCAGCCGTACAATGAAGGCGCTCTGCTGTACAGCATCGCTAATCAACCGGTTTCCGTCGT
GTTGGAAGCAGCCGGTAAAGACTTCCAGTTGTACCGCGGTGGTATTTTCGTGGGTCCGTGTGGTAACAAAGTGGATCATGCAGTCGCTGCCGTG
GGTTACGGCCCAAATTACATCCTCATCAAGAACTCCTGGGGCACCGGCTGGGGTGAAAACGGTTACATCCGTATTAAACGTGGTACGGGCAACA
GCTATGGTGTTTGCGGTTTGTACACGAGCAGCTTTTATCCGGTCAAGAAC**CTCGAG**

Enzymes expressed in *P. pastoris* with pPICZα expression vector and α-factor for secretion

The DNA sequences presented below contain N- and C-terminal restriction sites (in bold). The native signal peptides have been removed as α-factor signal sequence will be utilized. XhoI followed by Lys and Arg residues are inserted before propeptide as Kex2 cleavage site is located after Lys and Arg. This way, the protein is flushed against the α-factor. The native Stop codon is also removed as the C-terminal His_6-tag contains Stop codon. Lys and Arg residues are written in red.

C. histolyticum - Collagenase ColG

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, NotI: GCGGCCGC
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site NotI is used as using XbaI site would call for the last amino acid of the enzyme having to be changed.

CTCGAGAAAAGAAAGCCCATAGAGAACACGAACGATACCTCCATAAAAAATGTAGAAAAGCTCAGAAACGCTCCCAACGAGGAAAATAGCAAGA
AGGTTGAGGATTCAAAGAACGACAAGGTCGAGCACGTCAAAAACATTGAAGAGGCTAAGGTAGAACAGGTGGCTCCCGAAGTTAAAAGTAAGTC
GACACTTCGCAGCGCCTCTATCGCTAATACGAACTCCGAAAAATACGACTTTGAGTATTTGAATGGTCTATCCTACACGGAGCTGACTAATCTT
ATAAAGAATATTAAGTGGAACCAGATTAACGGCCTATTTAATTACTCGACAGGATCACAAAAATTTTTCGGTGACAAAAATCGAGTTCAGGCTA
TAATAAATGCGCTGCAAGAGTCGGGTAGAACATATACTGCTAACGATATGAAGGGGATCGAAACCTTCACGGAGGTTCTGCGTGCGGGGTTTTA
TCTAGGGTACTATAACGACGGACTGTCATACCTGAACGACCGAAACTTCCAGGACAAATGCATTCCAGCCATGATAGCCATTCAGAAAAATCCG
AACTTCAAACTTGGCACGGCAGTCCAAGACGAAGTAATCACTAGTCTGGGGAAACTCATCGGAAATGCAAGTGCTAATGCTGAAGTAGTCAACA
ATTGTGTCCCCGTATTGAAACAATTCAGGGAGAACCTTAATCAATACGCGCCGGATTATGTTAAGGGTACAGCTGTCAATGAGTTGATTAAGGG
TATAGAGTTCGACTTCTCTGGGGCCGCATACGAAAAAGATGTTAAGACGATGCCGTGGTATGGAAAAATTGATCCATTTATTAATGAACTAAAG
GCATTGGGACTATACGGCAATATCACATCGGCAACTGAGTGGGCATCTGATGTGGGGATCTACTATCTGAGTAAATTTGGACTCTACTCCACAA
ATAGGAATGACATTGTGCAGTCCTTAGAAAAGGCCGTTGACATGTACAAATATGGAAAGATAGCCTTCGTTGCTATGGAGAGGATAACATGGGA
TTATGACGGGATTGGCTCGAACGGTAAAAAGGTTGACCATGACAAGTTTCTAGATGATGCCGAGAAACATTATCTACCAAAAACGTATACCTTT
GACAACGGTACCTTTATAATTCGGGCAGGGGACAAAGTTAGCGAGGAGAAGATCAAACGGCTTTATTGGGCTTCTCGCGAAGTAAAAAGCCAAT
TCCACAGAGTCGTGGGGAATGATAAGGCGTTAGAAGTTGGCAACGCGGATGATGTACTCACTATGAAGATCTTCAATTCTCCAGAGGAGTATAA
ATTCAATACGAATATAAACGGAGTTTCCACAGACAACGGCGGGTTGTATATAGAACCACGCGGAACATTCTATACCTACGAGCGGACCCCTCAG
CAGAGCATTTTTTCTTTGGAAGAATTGTTCCGGCATGAGTACACCCATTACCTCCAAGCACGATACCTGGTCGATGGCTTATGGGGTCAAGGCC
CTTTTTATGAAAAAAATCGCCTTACGTGGTTTGACGAAGGTACTGCGGAATTCTTTGCGGGTAGCACGCGGACATCTGGTGTCCTACCCAGAAA
GTCTATCTTGGGTTATCTCGCGAAGGATAAGGTGGATCACCGTTACAGTTTAAAGAAGACGCTCAACTCCGGCTATGATGATTCCGATTGGATG
TTCTACAATTACGGTTTTGCCGTGGCACATTATCTTTACGAAAAGGATATGCCTACATTTATTAAAATGAATAAGGCCATCCTAAATACAGACG
TAAAAAGTTACGACGAGATCATAAAAAAACTGTCAGATGACGCAAACAAAAATACAGAATACCAGAACCATATCCAAGAGCTCGCCGACAAATA
CCAAGGGGCCGGAATACCGTTAGTATCTGATGACTACTTGAAGGATCACGGTTATAAGAAGGCAAGCGAGGTCTACTCTGAGATTTCAAAGGCA
GCTTCACTGACAAATACGTCGGTTACTGCCGAAAAATCTCAGTATTTTAATACATTCACCTTACGAGGAACCTACACCGGAGAGACTTCGAAGG
GGGAATTTAAGGACTGGGACGAAATGAGTAAGAAGCTCGACGGTACTTTGGAGAGTCTTGCGAAAAATAGCTGGAGTGGCTACAAGACGCTTAC
TGCCTACTTTACGAACTATCGTGTAACTTCGGATAATAAGGTGCAGTATGATGTGGTGTTTCATGGCGTACTAACTGATAACGCGGATATCTCT
AACAACAAAGCGCCGATAGCTAAGGTTACGGGGCCTTCAACCGGGGCAGTTGGGAGAAATATAGAGTTTAGCGGGAAGGATAGTAAAGATGAGG
ATGGAAAGATCGTTTCGTACGATTGGGACTTCGGCGACGGCGCTACTAGCAGGGGAAAGAACTCAGTCCACGCATATAAAAGGCGGGAACCTA
TAACGTGACATTGAAAGTCACCGATGACAAAGGAGCTACTGCGACTGAAAGCTTTACAATAGAAATTAAAAACGAAGATACCACAACCCCGATA
ACGAAGGAAATGGAACCAAACGACGATATTAAGGAAGCCAACGGCCCAATAGTAGAAGGTGTGACGGTGAAAGGCGACCTAAACGGTTCCGACG
ATGCTGATACATTTTATTTCGACGTTAAGGAAGACGGAGATGTAACTATCGAGTTACCTTATTCTGGATCGTCTAATTTCACCTGGTTAGTCTA
TAAAGAGGGCGACGACCAAAACCACATTGCGTCGGGCATTGACAAAAACAACTCCAAAGTCGGTACCTTTAAGTCAACAAAAGGCAGGCACTAT
GTTTTCATCTACAAACATGACAGTGCATCCAATATCAGCTACTCATTAAATATAAAGGGACTTGGGAACGAAAAACTTAAAGAGAAAGAAAACA
ACGATAGTAGCGACAAGGCCACTGTAATTCCTAATTTCAATACTACAATGCAGGGTTCGCTGTTAGGGGACGACTCCCGAGATTATTACAGTTT
CGAAGTGAAAGAGGAGGGGGAAGTGAACATCGAGCTCGACAAAAAGGACGAATTCGGGGTGACCTGGACGTTACACCCTGAATCCAATATCAAC
GACCGCATTACCTATGGACAAGTGGATGGGAATAAAGTCTCGAATAAGGTCAAACTCCGTCCCGGTAAATATTATTTGCTGGTCTACAAATACT
CAGGCTCAGGTAATTATGAGCTTCGTGTAAACAAG**GCGGCCGC**

C. histolyticum - Collagenase ColH

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, XbaI: TCTAGA
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site XbaI is used. Using this restriction site requires the last codon of the enzyme sequence to be removed before the first nucleotide (T) of XbaI. In order to achieve this the last amino acid, Arg is mutated from CGA to CGT (written in green) and the T removed. This does not alter the sequence.

CTCGAGAAAAGAGCGGTTGACAAAAACAACGCTACAGCAGCTGTGCAGAACGAGAGTAAACGCTACACCGTCTCTTATCTAAAGACGCTCAACT
ATTATGACTTGGTAGATCTCTTAGTTAAAACCGAGATAGAGAACCTGCCTGACCTTTTTCAATATTCCAGCGACGCTAAGGAATTCTACGGAAA
CAAGACGCGAATGTCGTTCATTATGGATGAGATAGGCAGAAGGGCCCCACAATATACTGAGATTGACCATAAGGGTATACCGACGCTGGTCGAA
GTGGTGCGTGCGGGATTTTATCTGGGTTTCCATAATAAAGAATTAAATGAGATCAATAAGCGGTCGTTTAAGGAACGAGTCATACCGTCCATTC
TCGCAATTCAAAAAAACCCAAACTTCAAGCTGGGGACGGAGGTGCAAGACAAAATAGTCTCGGCTACAGGCTTGTTAGCGGGAAACGAAACGGC
GCCACCTGAGGTCGTGAATAATTTTACACCAATATTACAGGATTGCATCAAGAATATCGATCGTTACGCTCTGGACGATCTAAAAAGTAAAGCA
CTCTTCAATGTCCTTGCCGCACCAACTTATGACATCACTGAATACTTACGTGCCACGAAGGAAAAACCGGAAAATACGCCGTGGTACGGCAAAA
TAGACGGTTTCATTAATGAACTAAAAAAGCTAGCACTTTATGGTAAAATCAACGACAATAACAGTTGGATCATAGATAATGGAATTTACCACAT
TGCCCCACTGGGTAAATTACACTCGAATAACAAGATAGGAATAGAGACGTTGACGGAGGTCATGAAGGTATATCCATACTTAAGTATGCAGCAC
CTTCAGTCCGCAGACCAGATCAAGCGGCATTACGATAGCAAGGATGCAGAGGGGAATAAAATTCCCCTAGACAAGTTTAAAAAGGAAGGCAAAG
AAAAGTATTGTCCTAAGACGTACACCTTTGATGATGGCAAGGTAATAATAAAGGCCGGCGCCCGCGTAGAAGAGGAGAAAGTAAAGCGGCTATA
CTGGGCGTCGAAAGAAGTCAACAGTCAGTTTTTCCGTGTTTATGGAATCGATAAGCCCTTAGAAGAAGGTAACCCAGATGATATTTTGACAATG
GTAATCTATAATTCCCCCGAGGAGTACAAACTGAACTCAGTATTATATGGTTACGACACAAACAACGGAGGGATGTACATCGAACCAGAGGGAA
CATTTTTCACGTATGAACGGGAAGCCCAGGAATCCACGTATACCTTAGAAGAACTTTTTAGACATGAGTATACACACTACCTTCAAGGTAGGTA
CGCGGTGCCCGGACAATGGGGAAGGACTAAACTTTATGATAATGATCGCCTGACTTGGTATGAAGAAGGTGGTGCTGAACTCTTCGCAGGATCC
ACTAGGACATCTGGGATACTGCCTAGGAAGTCCATCGTCTCAAATATTCATAATACAACCCGGAATAACCGGTATAAACTCTCTGACACTGTAC
ATAGCAAGTATGGTGCCAGTTTCGAGTTCTACAACTACGCGTGCATGTTCATGGACTATATGTACAATAAAGATATGGGGATACTAAACAAATT
GAACGATCTGGCTAAGAACAATGATGTTGATGGTTACGACAACTATATTAGAGATTTATCTAGCAACTATGCACTAAACGACAAGTACCAAGAC
CACATGCAAGAAAGAATAGACAATTATGAGAATTTGACAGTTCCGTTTGTTGCAGATGACTATCTTGTACGCCATGCCTATAAGAACCCTAACG
AGATCTACTCAGAAATATCTGAGGTTGCCAAGTTAAAAGATGCAAAATCGGAGGTGAAGAAGTCACAATACTTTTCCACTTTTACCCTCCGCGG
TTCCTACACCGGCGGAGCATCTAAGGGTAAGTTAGAAGATCAGAAAGCCATGAATAAATTTATTGATGATAGCCTAAAGAAACTTGATACCTAT
TCGTGGTCAGGCTACAAAACCTTAACTGCTTACTTTACAAACTATAAAGTTGATAGCTCTAATCGAGTAACATACGACGTCGTCTTCCACGGCT
ACTTGCCTAATGAAGGAGACAGTAAGAACTCCTTGCCCTACGGCAAGATTAACGGGACATATAAGGGCACGGAAAAAGAAAAAATCAAGTTTAG
TTCTGAGGGCAGCTTTGACCCCGATGGCAAAATTGTGTCCTATGAGTGGGATTTCGGGGACGGAAACAAATCAAATGAAGAGAATCCTGAGCAC
TCGTACGATAAAGTAGGCACATACACTGTAAAGCTCAAAGTGACAGACGATAAGGGGGAAAGCAGTGTGTCGACTACTACGGCTGAAATCAAGG
ACCTCTCAGAGAATAAACTTCCGGTCATATATATGCATGTACCGAAGTCGGGGGCGTTGAATCAGAAAGTGGTGTTCTATGGCAAAGGCACTTA
CGACCCGGACGGATCAATAGCTGGCTATCAATGGGATTTTGGCGACGGTTCCGACTTCTCTTCAGAGCAGAACCCCTCCCACGTCTATACTAAA
AAGGGAGAGTACACTGTGACCTTGCGTGTTATGGACAGTAGTGGACAGATGAGCGAGAAAACTATGAAGATCAAGATTACGGACCCGGTCTACC
CCATCGGGACGGAGAAAGAACCAAATAATTCTAAAGAAACCGCGTCAGGGCCGATAGTACCTGGTATTCCTGTGAGCGGGACCATTGAAAATAC
CTCTGATCAAGATTACTTTTACTTCGACGTCATCACACCCGGGGAAGTTAAAATCGACATCAATAAACTTGGTTATGGTGGCGCGACCTGGGTT
GTATACGACGAAAACAACAATGCCGTTTCTTATGCGACCGATGACGGGCAGAACCTATCGGGGAAATTTAAGGCTGATAAACCGGGGCGATACT
ATATACACCTGTACATGTTCAACGGAAGCTATATGCCCTACAGAATCAATATTGAAGGGTCAGTTGGGCGT**CTAGA**

G. hollisae - Collagenase

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site NotI: GCGGCCGC
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site NotI is used.

CTCGAGAAAAGATCGGAGGTTTTTGCGCACCCGGGGATGCCGGTCCAACAGAAACTCCAGCCTAATTTACTTCAACAGAGTACTAGGTTGCAGC
CCGAACAACACATACATGGATTAGAAAGGACTGATCGACAATACAGGCCAACCGATGCTACCCAACAGCCCGAACCCCCCACGCTCTTAAAAAG
GCAAGTTTCAGTTCAGCAAGATGCCGTAGAACAATGCGACCTATCTCAATTTCAGACAACTTCCTCTAATCAACTGATGGCCGCTATTCGGCAA
CAGGGTGCTTCTTGCGTCAATGCCCTATTCTCGGCGGATACCGGAGTCCAGGAGGCAGCATTTTCTAGCAATCACATGTACAACGTTGCGCAGT
ACACCAGAACGTTAGCACAGCAGTATGCAGGAGGTGGATCCGACGAGCTCGAAGCGTTGTATTTGTATCTGCGCGCTGGCTACTACGCAGAGTT
TTATAACTCGAATATAACGTTTCTTTCCTGGGTCACCCCTGCGGTGAAAGGAGCTGTAGATGCCTTCGTTCAGAACGCTCATTTCTACGACAAT
GGTGACGCCCATGGGAAGGTGCTGAATGAAGTCATAATAACAATGGATAGCGCGGGCTTGCAGCACGCCTATCTTGACGTCGTGACACAGTGGC
TTACTCGATGGAATGCGCAGTATGCCGAGCATTGGTATATGAGGAATGCAGTGAACGGTGTGTTCACATTGTTGTTTGGAGGCCAGTGGAACAA
CCAATATACATCTCTGATCGGCGAGCAAACCGCGCTCGTGACGGCCCTACAGGCTTTTGCATTAGACAGAACGAAGGTAAACAGCCCGACCGAA
TTCATGGCAGCTAATGCCGCAAGAGAACTAGGGCGGCTAGCTCGCTACACCGATGCTACAATTGCACCTAAGGTGACGGAGGGTCTAACGGCCA
TCTTCGGCCAATACCCCTCCTATGGGGACGGGGACGCCATTTGGTTAGGGGCGGCCGACACGGCAAGTTACTACGCGGACTGCTCACAATTTAA
CATATGTGGTTTTGAGGATGCCCTACGAGACGCTGCTTTAAATCAGACATTTATATGCTCCGATACTATCAAAATTCGTTCGCAAGACATGAGC
CAAGCGCAGCACCTAGCTGCCTGCGACAAGATGGCATACGAGGAATCGTTCTTCCACACTACATTAGAAACTGGAAACCAACCAGTTGCCGACG
ATCATAACACCCAGCTTCAGGTTAACATCTTTAATTCAGACACTGATTACGGGAAGTACGCGGGGCCAATCTTCGGCATCGACACCAATAATGG
GGGCATGTATCTCGAAGGCAATCCTGCCAATGTGGGGAACATACCTAATTTCATTGCTTATGAGGCTAGCTACGCAAACCCAGATCATTTCGTA
TGGAACCTAGAGCACGAATACGTACATTACTTGGACGGTCGCTTTAACATGTACGGCGATTTCGGAACACCGACCGAACTGGTAGTATGGTGGA
GTGAAGGTGTCGCGGAATATGTATCGCGGGTTAACGATAACCCACAGGCAATAGCAACCATTCAGGATGGCTCGACGTACACTCTTGCACAGGT
TTTCGATACGACATATGACGGGTTCGATGTCGATCGCATCTATCGATGGGGATACCTCGCTGTCCGTTTTATGTTTGAACGTCACCCTGATGAG
GTGCAACGGATGCTGTCAGCCACGAGACAAGGTAGATGGGCAGAGTATAAAGCAATAATTAGTGGCTGGGCTAATCAGTACCAATCCGAGTTTG
CACAATGGACGGAGGCGCTCGCGAAGGGGGATTCGGGTGCGGGGAACGGTGAGGGGACTGGTAGTGGAAACGAAGGAGGAGGCGAGTCAGGTGG
CAATACCGGTCTCCCCGAAAATTGTGCCGTACTACCAAAAATCTCCGATGGTCGTCTTGCGTTGGATGAAGCCGCTTGTTTAGCGGACACTGCG
TCTGCAAGTGACGTTTTGTGGTTTAGTATCCCGGCCGTGTCTGAGTATCAAACGATTGCAATTACAGCGGGCAACGGCACAGGGGACCTGACCC
TGGAGTATTCCAACTTAAACTGGCCGGACGGTACTAACGTACAAGCGTCTTCAGCTAATATGGGAAACAGTGAGTGTATCATTCTTGAACATCA
GGCCAATTATTGGGGATATTTAAAAGTCAGCGGTTCATTCGAAAATGCGGCTCTGCTTGTCGAAGCTGGGAGCAATCAATGTCGGCAA**GCGGCC
GC**

B. cereus - Collagenase ColA

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, NotI: GCGGCCGC
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site NotI is used.

CTCGAGAAAAGAGAGGAAACAGCGCCCTACAACATCCTCCAGATGAAGCCGATGGGCACGGAAACTTCGAAAGATGAGATAGTCCATGCTACGA
AAGCGGATGAAACATTAAATTTTGAAGAGCGGCTCAAGATCGGCGACTTCAGCCAGCGCCCTACATTAGTAATGAAACGTGATGAGATTCAACT
AAAGCAATCATATACTCTGGCTGAACTCAACAAAATGCCAAACAGCGAGCTGATCGACACGCTAAGTAAAATCAGTTGGAACCAAATCACGGAT
TTGTTTCAATTTAACCAAGATACAAAAGCGTTCTATCAGAACAAGGAAAGGATGAACGTGATAATTAATGAGCTGGGACAGAGAGGCAGGACAT
TCACAAAAGAAAACTCTAAGGGGATCGAAACCTTCGTGGAGGTGCTTCGCTCCGCATTTTATGTGGGATATTACAACAACGAACTAAGCTATCT
TAAGGAGCGTTCGTTCCACGATAAGTGCCTGCCGGCACTCAAAGCCATTGCAAAGAATCGGAACTTTACTTTGGGTACAGCAGAACAGGACCGC
GTGGTTACCGCGTACGGAAAGCTAATAGGGAACGCTTCTTCAGACACCGAGACCGTTCAGTACGCGGTGAACGTTCTAAAGCACTACAATGACA
ATCTCTCCACGTATGTCTCTGACTACGCGAAAGGTCAGGCCGTATACGAGATTGTGAAAGGCATTGATTACGACATCCAGTCATATCTTCAAGA
TACCAATAAACAACCTAACGAAACGATGTGGTACGGTAAGATAGATAATTTCATTAATGAGGTTAACAGAATTGCATTAGTGGGCAATATCACA
AATGAAAATTCGTGGCTGATAAATAACGGAATATACTACGCGGGGAGATTAGGTAAATTCCATTCTAACCCTAACAAGGGCTTGGAGGTAATTA
CGCAAGCAATGAACCTCTATCCTCGACTATCTGGTGCATACTTTGTGGCGGTAGAACAGATGAAAACGAATTATGGGGGCAAAGATTACAGCGG
TAACGCAGTCGACCTTCAGAAAATACGTGAAGAAGGGAAGCAGCAATATTTGCCAAAGACTTATACATTTGACGATGGCAGTATCGTCTTCAAA
ACTGGAGATAAAGTCACAGAAGAGAAGATAAAAAGATTATACTGGGCCGCCAAGGAAGTAAAAGCCCAATACCACCGGGTTATCGGGAATGACA

AGGCTTTAGAACCTGGAAACGCCGATGACGTGTTAACCATCGTTATATATAATAATCCAGATGAATATCAACTGAACCGACAATTATATGGATA
CGAGACAAATAATGGAGGGATTTACATTGAGGAGAAAGGAACCTTCTTTACCTATGAACGCACGCCTAAACAGAGCATATATTCCCTGGAGGAG
CTTTTTCGTCACGAATTCACGCATTATCTTCAAGGCCGCTACGAAGTACCGGGACTGTTCGGGAGTGGTGAAATGTATCAAAATGAACGTCTGA
CTTGGTTTCAAGAGGGTAACGCTGAATTTTTCGCAGGATCCACAAGAACGAATAACGTTGTACCGCGGAAAAGCATGATCTCCGGTCTCTCATC
GGACCCCGCGAGTCGGTACACTGTTAAACAGACGTTGTTCTCGAAATACGGGTCTTGGGACTTTTATAAGTACTCGTTCGCACTTCAATCATAC
CTATATAATCACCAGTTTGAGACATTTGACAAGTTGCAGGATCTGATTCGAGCTAACGACGTCAAGAATTACGATTCCTACCGAGAAAGTCTGT
CCAACAACACTCAGCTAAATGCCGAGTATCAAGCCTACATGCAGCAACTTATTGACAATCAAGATAAATATAATGTACCGAAAGTCACCAATGA
TTACCTCATCCAACACGCTCCAAAGCCACTAGCCGAGGTTAAGAATGAAATCGTAGATGTGGCTAATATAAAGGATGCTAAAATTACCAAGTAC
GAGAGTCAGTTCTTTAACACTTTCACGGTTGAGGGCAAGTACACTGGAGGCACCTCCAAGGGAGAGTCTGAGGACTGGAAGGCCATGAGTAAAC
AAGTCAATCAGACCTTAGAACAATTGAGCCAGAAGGGATGGTCTGGGTATAAGACGGTCACCGCCTACTTTGTAAACTACAGGGTTAATGCGGC
GAATCAATTCGAGTATGATATTGTTTTTCATGGCGTTGCTACTGAGGAGAAAGAGAAAACTACTACCATAGTCAATATGAACGGGCCATACAGT
GGTATAGTCAACGAGGAAATACAATTCCATTCGGACGGCACAAAGTCAGAAAATGGGAAGGTGATTAGCTACCTATGGAACTTCGGTGATGGCA
CTACGTCTACTGAAGCCAACCCCACTCATGTTTATGGGGAAAAAGGAACGTATACTGTCGAATTAACCGTAAAAGATAGCAGGGGTAAGGAGTC
GAAGGAACAGACCAAGGTTACGGTGAAACAGGACCCCCAGACATCAGAGTCCTATGAAGAGGAAAAAGTACTCCCATTCAACACCTTGGTCAAG
GGTAACCTCATTACCCCCGACCAAACTGACGTGTATACCTTTAACGTTACAGATCCCAAAGAGGTAGACATCAGTGTGGTTAATGAACAGAATA
TAGGCATGACGTGGGTATTGTATCATGAATCGGACATGCAGAACTATGTCGCATGTGGGGAAGACGAAGGAAACGTGATCAAAGGTAAGTTTGC
AGCTAAGCCGGGTAAATATTATTTAAATGTATATAAGTTTGATGACAAGAACGGCGAGTACTCATTGCTTGTCAAA**GCGGCCGC**

B. licheniformis - Subtilisin Carlsberg

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, NotI: GCGGCCGC
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site NotI is used.

CTCGAGAAAAGAGCACAGCCCGCGAAAAACGTTGAAAAAGATTACATCGTCGGATTTAAATCAGGTGTTAAGACAGCGAGTGTCAAAAAGGACA
TAATCAAAGAGTCCGGCGGAAAAGTGGACAAGCAATTCCGTATTATAAATGCAGCTAAGGCGAAGCTCGACAAGGAAGCGCTGAAAGAGGTAAA
AAACGATCCCGACGTGGCGTATGTGGAAGAGGACCACGTAGCCCATGCTCTTGCTCAAACTGTGCCATACGGTATTCCGCTAATTAAGGCCGAC
AAAGTCCAGGCGCAGGGGTTCAAGGGAGCCAACGTCAAGGTCGCGGTTCTAGACACAGGCATTCAGGCTTCGCATCCTGACCTTAATGTCGTAG
GCGGGGCCAGCTTTGTTGCGGGTGAGGCGTACAACACTGATGGGAACGGACACGGGACACACGTAGCCGGAACCGTGGCTGCTCTGGATAATAC
AACTGGTGTACTTGGCGTTGCACCCAGTGTATCATTGTACGCTGTGAAAGTATTGAACTCTTCGGGATCTGGCACCTACTCCGGCATAGTCTCA
GGGATTGAATGGGCTACGACCAATGGCATGGATGTCATAAACATGTCTCTGGGTGGACCAAGTGGCAGCACGGCCATGAAGCAAGCAGTCGATA
ATGCGTATGCTCGTGGTGTTGTGGTTGTAGCCGCCGCAGGCAATTCCGGGAGTTCAGGAAATACGAATACAATCGGTTATCCTGCAAAGTACGA
TTCTGTTATCGCAGTGGGTGCCGTTGATAGCAATAGCAACAGGGCCTCCTTTTCGAGCGTCGGTGCTGAACTGGAGGTGATGGCACCGGGGGCC
GGCGTATATTCGACCTATCCGACTTCGACGTATGCTACGTTAAACGGTACTAGTATGGCTAGCCCACATGTTGCGGGGGCTGCAGCGTTAATCT
TGAGTAAGCACCCTAACCTCTCAGCCTCTCAAGTGAGAAATCGCTTATCCTCAACAGCAACCTATCTAGGATCCTCTTTCTACTACGGGAAAGG
ACTCATAAATGTAGAAGCAGCAGCGCAG**GCGGCCGC**

C. lupus familiaris - Pepsinogen B

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, XbaI: TCTAGA
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site XbaI is used. Using this restriction site requires the last codon of the enzyme sequence to be removed before the first nucleotide (T) of XbaI. In order to achieve this the last amino acid, Ser is mutated from TCC to TCT (written in green) and the T removed. This does not alter the sequence.

CTCGAGAAAAGAGTTGAACGAATCATATTGAAGAAGGGAAAATCCATACGTCAAGTTATGGAGGAGCGGGGAGTGCTTGAAACGTTTCTCCGGA
ACCATCCTAAGGTTGATCCTGCCGCTAAATACTTATTCAATAACGATGCCGTAGCGTATGAGCCATTTACCAATTACCTGGACTCTTATTATTT
CGGTGAGATTTCAATTGGCACCCCCCCACAAAACTTTCTCATACTATTCGATACAGGAAGCTCGAACTTATGGGTACCCTCAACTTACTGTCAA
AGCCAAGCTTGTAGTAATCATAATCGCTTTAACCCGTCCAGGAGCTCAACCTACCAGTCTAGTGAACAGACATACACTCTCGCGTATGGTTTCG
GCTCACTGACTGTCCTGTTGGGGTATGATACAGTGACTGTGCAAAACATAGTTATCCACAACCAGCTATTTGGAATGAGTGAAAATGAACCTAA
CTATCCATTTTATTACTCATATTTCGACGGTATCTTAGGCATGGCGTATTCGAATCTAGCGGTGGATAACGGTCCTACGGTTCTTCAGAATATG

ATGCAGCAGGGGCAACTAACCCAACCAATCTTTTCTTTCTACTTTAGTCCACAACCTACCTACGAGTATGGGGGCGAACTGATACTCGGAGGAG
TGGATACCCAGTTCTACTCGGGGGAGATAGTTTGGGCACCGGTAACTAGAGAAATGTATTGGCAAGTCGCAATCGACGAATTTTTGATTGGGAA
TCAGGCCACTGGATTGTGCTCTCAAGGGTGCCAGGGGATAGTAGACACAGGCACGTTCCCGCTAACGGTCCCCCAGCAATACTTAGACAGCTTT
GTGAAAGCCACGGGTGCACAACAGGATCAGTCCGGCAACTTCGTTGTCAATTGCAATTCTATCCAGTCGATGCCAACTATTACGTTTGTAATTT
CCGGCAGCCCGCTTCCTCTGCCCCCCTCAACGTACGTCTTGAATAACAACGGTTACTGTACACTTGGAATCGAGGTAACATATCTCCCCTCGCC
GAACGGTCAACCGCTTTGGATTCTCGGGGACGTCTTCTTACGCGAGTATTACACCGTATTTGACATGGCTGCAAATCGAGTGGGTTTCGCTCTG

AGTTC**TCTAGA**

B. taurus - Cationic Trypsin

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, XbaI: TCTAGA
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site XbaI is used. Using this restriction site requires the last codon of the enzyme sequence to be removed before the first nucleotide (T) of XbaI. In order to achieve this the last amino acid, Asn is mutated from AAC to AAT (written in green) and the T removed. This does not alter the sequence.

CTCGAGAAAAGAGTGGATGACGACGATAAGATTGTAGGCGGGTACACGTGCGGTGCAAATACTGTTCCGTATCAAGTTTCCTTGAACTCTGGCT
ATCATTTCTGTGGAGGTAGCCTAATAAATAGTCAGTGGGTTGTTAGCGCTGCCCACTGTTACAAGTCTGGCATCCAGGTCAGACTTGGGGAAGA
CAATATAAACGTCGTCGAAGGGAACGAGCAGTTCATATCTGCCAGTAAGTCCATTGTACACCCCAGTTACAACAGCAATACGCTCAACAATGAC
ATCATGTTGATCAAACTGAAAAGCGCGGCCTCACTGAACTCGCGGGTGGCTTCAATTTCGCTTCCTACGTCATGCGCCTCGGCTGGAACACAAT
GCCTAATTAGTGGCTGGGGTAACACCAAATCCAGTGGCACATCCTACCCAGACGTCTTAAAGTGTTTAAAAGCGCCCATACTCTCCGATTCTAG
TTGTAAATCTGCGTATCCTGGGCAAATTACCTCGAATATGTTTTGCGCAGGATACCTTGAGGGGGGTAAAGATTCGTGTCAGGGAGATAGCGGT
GGACCGGTGGTTTGCTCCGGCAAATTGCAAGGGATCGTATCGTGGGGCTCAGGTTGTGCGCAGAAGAATAAACCAGGAGTATATACTAAGGTAT
GCAATTATGTGTCATGGATAAAGCAAACCATCGCAAGCAA**TCTAGA**

C. papaya - Papain

N-terminal restriction site XhoI + Lys and Arg residue: CTCGAGAAAAGA
C-terminal restriction site, XbaI: TCTAGA
In order to be in frame with α-factor signal sequence, C-myc epitope and polyhistidine-tag, the c-terminal restriction site XbaI is used. Using this restriction site requires the last codon of the enzyme sequence to be removed before the first nucleotide (T) of XbaI. In order to achieve this the last amino acid, Asn is mutated from AAC to AAT (written in green) and the T removed. This does not alter the sequence.

CTCGAGAAAAGAGTATACATGGGCCTCTCCTTTGGGGACTTCTCTATTGTAGGTTACTCACAAAACGACCTCACGAGTACAGAGAGACTGATTC
AGCTCTTCGAGAGTTGGATGCTTAAACACAACAAGATATACAAGAACATCGATGAGAAGATATATCGATTTGAAATATTTAAAGACAATCTAAA
GTATATCGATGAAACTAATAAAAAAAATAATTCGTATTGGCTGGGATTAAACGTCTTCGCAGATATGTCCAACGACGAGTTTAAGGAAAAGTAC
ACCGGGTCAATAGCAGGAAATTATACTACGACTGAATTAAGTTATGAAGAAGTACTCAATGACGGCGATGTGAACATCCCGGAATATGTGGACT
GGCGTCAGAAAGGGGCTGTTACTCCCGTTAAGAATCAGGGAAGCTGTGGTTCATGCTGGGCCTTCAGCGCCGTTGTCACCATAGAGGGCATCAT
CAAAATTAGAACAGGCAATTTGAATGAGTACTCTGAACAGGAATTATTAGATTGTGACCGGCGAAGTTACGGTTGCAACGGAGGCTATCCGTGG
TCTGCTCTTCAGCTAGTGGCTCAGTACGGGATTCATTATAGGAATACATACCCATACGAGGGGGTCCAACGCTACTGTCGGTCCCGGGAGAAGG
GTCCTTACGCCGCAAAAACCGATGGAGTGCGTCAAGTTCAACCATACAATGAGGGGGCGCTGTTGTATTCGATTGCTAATCAACCCGTTTCTGT
TGTCTTAGAAGCCGCGGGCAAAGACTTCCAATTGTATAGAGGTGGTATATTTGTGGGCCCTTGCGGGAACAAAGTCGATCACGCGGTGGCGGCA
GTAGGATATGGACCTAACTACATTTTAATAAAGAATTCGTGGGGTACCGGATGGGGAGAGAACGGGTATATCCGCATCAAAAGGGGGACAGGTA
ACTCATATGGCGTCTGTGGTCTGTATACGAGCAGCTTTTACCCCGTAAAGAA**TCTAGA**

Appendix B - Cloning strategy for *E. coli* continuation

The continuation of the cloning strategy for expression of recombinant enzymes in *E. coli* is presented below.

Colony PCR

Prepare masterplates by taking colonies from agar plates with grown colonies. Streak single colonies on the agar masterplate and transfer what is left on the pipette tip to PCR tubes containing PCR mastermix. Prepare 20 reactions of PCR mastermix by mixing 387 µL deionized water, 50 µL 10x DreamTaq buffer, 5 µL forward primer, 5 µL reverse primer, 50 µL dNTP and 3 µL DreamTaq DNA polymerase (5 U/µL). These primers are available from previous work and are shown below:

Forward primer (5'-3'):

TAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGA
GATATACATATG

Reverse primer (5'-3'):

GCTAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATC

The PCR program is repeated for 30 cycles (denaturation to extension step) with tube setting according to: initial denaturation (95 °C, 3 minutes), denaturation (95 °C, 30 seconds), annealing (57 °C, 30 seconds), extension (72 °C, 2 minutes 17 seconds), final extension (72 °C, 10 minutes). Run linearized PCR samples (by digestion with suitable restriction enzyme) on 1% agarose gels according to the protocol in “Cloning experiments”. Mix the samples in a 5:1 ratio, DNA to 6x TriTrack DNA loading dye. Extract the plasmids from positive clones with Plasmid Mini Kit II and send for Sanger sequencing.

Cell culture and protein expression

Grow positive clones in 5 mL LB media with ampicillin (100 µg/ml) at 37 °C overnight till OD600 of 2 or more is reached. After that, add a volume of the inocolumn to a shake flask to reach an OD600 of 0.1 and a final volume of 100 mL. Grow the cultures till an OD600 of 0.4. Add IPTG to a final concentration of 0.5 mM to induce protein expression and incubate for an additional 2–3 hours. Use the remaining cell culture inoculum to make glycerol stocks by adding the same volume of 50% glycerol and inoculum in a 2 mL cryogenic tube. Store the bacterial stocks at -80 °C.

Medium fraction

The proteins in the medium could be examined with or without concentration. Without concentration, centrifuge 1.5 ml sample after induction for 2 min at 13000 rpm. Mix the supernatant with a sample buffer for SDS-PAGE analysis.

With concentration, harvest 40 ml of the culture by centrifugation at 10,000 × g for 10 min at 4°C. Transfer 1 ml of the supernatant to a microcentrifuge tube. The cell pellet could be used for the periplasmic fraction. The sample could be concentrated by TCA precipitation or spin filter concentration. With a spin filter, use a low MW cut-off filter (10 kDa or lower) and follow the manufacturer's instructions to concentrate 500 µl of medium approximately 10 times. Record the concentration factor and move the concentrated sample to a clean tube. Rinse the spin filter membrane with hot 2X SDS Sample Buffer using the same volume as the concentrated sample. Pool the 2X SDS Sample Buffer membrane rinse with the concentrated sample. Immediately heat for 3 min at 85 °C to denature the proteins and then store at –20 °C until SDS-PAGE analysis.

With TCA precipitation, add 100 µl of 100% TCA (w/v) to 1 ml of medium and vortex for 15 sec. Put on ice for at least 15 min. After that, centrifuge at 14,000 × g for 10 min, remove and discard the supernatant. Then, wash the pellet twice with 100 µl of acetone, by adding acetone, mixing and spinning for 5 min (14,000 × g). Remove and discard the acetone from the loose

pellet and allow the pellet to air dry. Afterwards, add 100 µl of 1X PBS and 100 µl of 4X SDS Sample buffer and resuspend by mixing. Immediately heat for 3 min at 85 °C to denature the proteins and then store at –20 °C until SDS-PAGE analysis.

Periplasm protein expression

One method for periplasmic protein extraction is osmotic shock. Resuspend cells in ice-cold 20% sucrose 2.5 mM EDTA, 20 mM Tris-HCl, pH 8.0, to a concentration of 5 OD550 and incubate on ice for 10 min. After that, centrifuge the cell solution at 15,000 x g for 30 seconds and resuspend the cell pellet in the same volume ice-cold 2.5 mM EDTA, 20 mM Tris-HCl, pH 8.0 and incubate on ice for 10 min. Lastly, centrifuge the solution at 15,000 × g for 10 min and collect the supernatant for analyses.

Chloroform extraction is another method for retrieving periplasmic protein. The following protocol utilizes 2 mL bacterial culture. Centrifuge cell culture for 10 minutes at 1,100 x g. Decant the supernatant and resuspend the cells in residual medium by vortexing. Add 20 µL of $CHCl_3$ followed by brief vortexing and 15 minute room temperature incubation. Add 0.2 ml of 0.01 Tris hydrochloride, pH 8 and separate the cells by centrifugation at 6,000 x g for 20 minutes. Remove supernatant using a pasteur pipette.

Total cell protein fraction

Cell lysis is carried out by centrifuging with 1 ml cell culture at 10,000 x g for 1 minute. Supernatant is removed and discarded. Furthermore, pellets are resuspended in 100 µl of 1X PBS to yield a concentration factor of 10X. 100 µl of 4X SDS Sample Buffer is added and sonicated with a microtip with the following settings: power level between 2–3, at 20–30% duty for 8–10 bursts (Branson Sonifier 450; conditions may vary with different equipment). Sample is immediately heated to denature proteins at 85 °C for 3 minutes and thereafter stored at -20 °C until usage.

SDS-PAGE

In order to confirm that the enzymes have been expressed, media, periplasmic and lysed cells samples are run on precast Mini Protean TGX gels 4-20% to evaluate the size of the proteins. Running buffer 10 v/v 10x Tris/Glycine buffer is used to fill the inner buffer chamber with 200 ml and 550 ml in the outer buffer chamber. Laemmli sample buffer (0.9% 4X Laemmli sample buffer, 0.1% 2-mercaptoethanol) is mixed with the sample in 0.25:0.75 ratio and 1 µl PageRuler prestained protein ladder is used to validate size. Mixed samples are heated at 90-100 °C for 5 minutes and run at 100 V for an expected 85-95 minutes. The gel is rinsed with distilled water and placed in a staining solution overnight (2.5 g Coomassie brilliant blue - R250, 100 mL Acetic Acid glacial 100%, 500 mL methanol, 400 mL Milli-Q water). The gel is further rinsed with ddH_2O and placed in a destaining solution (500 mL methanol, 100 mL Acetic Acid glacial 100%, 100 mL Milli-Q water) for 2 hours on a shaking table or until the background is clear.

Purification of enzyme

Protein purification can utilize the His_6-tags attached to the enzymes. A specific method has not been determined but using metallic beads, for example EziG from EnginZyme or HisTrap from GE Healthcare column could be two alternatives.

Appendix C - Yeast cloning strategy

This is a protocol with examples of steps for cloning and expression of the gene of interest in *P. pastoris* from Invitrogen's user manuals (2010a; 2010b). To clone and express the gene of interest the following steps could be executed: propagation of the expression vector, subcloning, transformation in *E. coli* and confirmation of successful clones, purification and linearization of the recombinant plasmid, transformation in *P. pastoris* strain, plating and selection, expression and purification of the recombinant protein.

Propagation of the expression vector

The pPICZα expression vector should be transformed into an *E. coli* strain that is recombination deficient (*rec*A) and endonuclease A deficient (*end*A) such as DH5α that was used in this thesis (see cloning experiment for transformation protocol). The pPICZα vector has Zeocin resistance and should be plated on LB agar plates (pH 7.5) with 25 μg/ml Zeocin. Bacterial glycerol stocks should also be prepared for storage of the vector (see cloning experiment for LB media and glycerol stock recipe). The transformed cells can be used to propagate the vector as described in the cloning experiment.

Subcloning

The gene of interest needs to be cloned into the pPICZα vector using specific restriction enzymes according to the gene design. To do that the gene must be obtained. The subcloning can then be performed according to the cloning experiment.

Transformation in *E. coli* and confirmation of successful clones

Once again, the ligation mixture should be transformed into an *E. coli* strain that is both recombination deficient (*rec*A) and endonuclease A deficient (*end*A) such as DH5α or JM109 and plated on LB agar plates (pH 7.5 containing 25 μg/ml Zeocin). The transformants must be verified to contain the insert DNA and the orientation of it. This could be done as described

before by colony PCR and sequencing (but the LB media should have the right antibiotic (Zeocin) and pH of 7.5). There are commercial primers for PCR for example 3′ *AOX1 Pichia* primer and the 5′ *AOX1 Pichia* primer. Glycerol stocks of the positive clones should be made.

Purification and linearization of the recombinant plasmid

The next step is to generate enough recombinant plasmids (5–10 μg) for the transformation and to linearize them. This could be done as described in the "cloning experiment" (but with LB media with the appropriate antibiotic and pH 7.5). The plasmids could then be linearized by digestion with a restriction enzyme. When the plasmids have been linearized the restriction enzymes could be heat inactivated and the linearized plasmids purified by for instance phenol/chloroform extraction and ethanol precipitation.

Transformation in *P. pastoris*, plating and selection

The pPICZα vector does not have the *HIS4* gene required for histidine synthesis which means that integration can only occur at the *AOX1* locus. *Pichia* strains that could be used are for example X-33, GS115, SMD1168H or KM71H. The first three strains have the *AOX1* gene and can therefore metabolize methanol (Mut^+). The latter strain has a disrupted *AOX1* gene and cannot metabolize methanol (Mut^S). It is possible to use electroporation or chemical methods for transformation. Spheroplasting is not recommended since it is not possible to directly plate spheroplasts on selective medium containing Zeocin.

To prepare the chosen *P. pastoris* strain for electroporation a preculture should be prepared in YPD medium (1% yeast, 2% peptone and 2% dextrose) and cultivated overnight at 30 °C. 500 ml of YPD should be inoculated with 0.1–0.5 ml of the overnight culture and grown overnight to an OD600 of 1.3–1.5. The cells should be centrifuged at 1500 × g for 5 minutes at 4 °C followed by pellet resuspension in 500 ml ice cold sterile water. Centrifugation as previous, then resuspension of the pellet with 250 ml of ice-cold sterile water. Centrifugation again as previous and then resuspension of the pellet in 20 ml of ice-cold 1 M sorbitol. Finally, a final

centrifugation and resuspension in 1 ml of ice-cold 1 M sorbitol. Keep the cells on ice and use the same day.

To transform with electroporation, 5–10 μg of linearized pPICZα is mixed in 5–10 μl water with 80 μl of the cells from above and transfer to an ice-cold 0.2 cm electroporation cuvette and incubate on ice for 5 minutes. Pulse the cells according to the specific device's instruction. Immediately add 1 ml of ice-cold 1 M sorbitol to the cuvette. Transfer the contents to a 15 ml tube and incubate at 30 °C without shaking for 1-2 hours. Spread different volumes (50-200 μl) on YPDS plates (YPD with 1 M sorbitol and 2% agar) containing appropriate concentration of Zeocin. Incubate plates for 2–3 days at 30 °C until colonies form. Pick 10–20 colonies and streak the single colonies on fresh YPD or YPDS plates with the appropriate concentration of Zeocin. PCR could be used to confirm the insertion of the gene or directly do small-scale expression. To perform PCR for confirmation, a protocol where the cells are lysed by the combination of enzyme, freezing and heating treatment could be used, for details see Invitrogen's user manual (2010b).

It is possible to determine the Mut phenotype which are transformants from the strains containing the *AOX1* gene such as X-33, GS115 or SMD1168H should be Mut^+. But there is however a chance that the recombination will occur at the 3′*AOX1* region also hence disrupting the AOX1 gene and creating a Mut^S. The phenotype could be determined with plates with or without methanol (Invitrogen, 2010b). If the strain KM71H was used there is no need to screen for phenotype because they will all be Mut^S. Knowing the phenotype makes it easier to determine the optimal expression conditions.

Expression of recombinant protein

Different media could be used for expression of the recombinant protein for example BMGY/BMMY (buffered complex glycerol or methanol medium) (1 liter water, 1% yeast extract, 2% peptone, 100 mM potassium phosphate (pH 6.0), 1.34% YNB, $4*10^{-5}$% biotin, 1% glycerol or 0.5% methanol). Buffered media allows a wide pH range for optimization of protein

production. The yeast extract and peptone stabilizes the secreted proteins by decreasing or preventing proteolysis and allows better biomass accumulation and growth. Aeration is important for expression and the temperature should always be 30 °C.

The chosen strain transformed with the vector without insert could be used as a control for background intracellular expression. For a Mut^+ strain, inoculate 25 ml of the media with a single colony in a baffled flask. Use 100 ml for a Mut^S strain. Grow at 30 °C in a shaking incubator (250–300 rpm) until culture reaches an OD600 of 2–6. Then harvest the cells by centrifuging at 1,500–3,000 × g for 5 minutes at room temperature. For a Mut^+ strain, decant supernatant and resuspend cell pellet to an OD600 of 1.0 in the chosen medium to induce expression in about 100–200 ml in a new baffled flask. Resuspend the cell pellet in 10-20 ml for the Mut^S strain. Put the flask back in the incubator to continue growth. Add 100% methanol to a final concentration of 0.5% methanol every 24 hours to maintain induction. For a Mut^+ strain, transfer 1 ml of the culture to a microcentrifuge tube at 0, 6, 12, 24, 36, 48, 60, 72, 84, and 96 hours after induction. Use the following time points for the Mut^S strain: 0, 24, 48, 72, 96, 120 and 144 hours. These samples should be used to analyze expression levels and to determine the optimal time to harvest after induction. Centrifuge at maximum speed in a microcentrifuge for 2–3 minutes at room temperature. To check the secreted protein expression, transfer the supernatant to other tubes. Store the supernatant and the cell pellets at –80 °C until use.

Protein expression could be analyzed with SDS-PAGE, western blot with anti-his or anti-myc antibodies or functional assay. The cells need to be broken with chemicals or physical forces to analyze the protein expression inside the cells, see user manual for details (Invitrogen, 2010b). The purification could be performed using the polyhistidine tag.

Appendix D - DNA for cloning experiment

The following sequences are optimized by Thermo Fisher Scientific's GeneOptimizer algorithm and used in the practical cloning experiment.

G. hollisae collagenase gene with NcoI and XhoI restriction sites highlighted in green.

CCATGGCTCATCATCACCATCATCATGAAGTTTTTGCACATCCGGGTATGCCGGTTCAGCAGAAACTGCAGCCGAATCTGCTGCAACAGAGCAC
CCGTCTGCAGCCGGAACAGCATATTCATGGTCTGGAACGTACCGATCGTCAGTATCGTCCGACCGATGCAACCCAGCAGCCTGAACCGCCTACA
CTGCTGAAACGTCAGGTTAGCGTTCAGCAGGATGCAGTTGAACAGTGTGATCTGAGCCAGTTTCAGACCACCAGTAGCAACCAGCTGATGGCAG
CAATTCGTCAGCAGGGTGCAAGCTGTGTTAATGCACTGTTTAGCGCAGATACCGGTGTTCAAGAAGCAGCATTTAGCAGCAATCATATGTATAA
TGTGGCCCAGTATACCCGTACACTGGCACAGCAGTATGCAGGCGGTGGTAGTGATGAACTGGAAGCACTGTATCTGTATCTGCGTGCAGGTTAT
TATGCCGAATTCTATAATAGCAACATCACCTTTCTGAGCTGGGTTACACCGGCAGTTAAAGGTGCAGTTGATGCATTTGTTCAGAACGCCCATT
TTTATGATAACGGTGATGCACATGGTAAAGTGCTGAACGAAGTGATTATTACAATGGATAGCGCAGGTCTGCAGCATGCATATCTGGATGTTGT
TACCCAGTGGCTGACCCGTTGGAATGCACAGTATGCCGAACATTGGTATATGCGTAATGCAGTGAATGGTGTTTTTACCCTGCTGTTTGGTGGT
CAGTGGAATAATCAGTATACCAGCCTGATTGGTGAACAGACCGCACTGGTTACCGCACTGCAGGCATTTGCACTGGATCGTACCAAAGTTAATA
GCCCGACCGAATTCATGGCAGCCAATGCAGCACGTGAACTGGGTCGTCTGGCACGTTATACCGATGCCACCATTGCACCGAAAGTTACCGAAGG
TCTGACCGCAATTTTTGGTCAGTATCCGAGCTATGGTGATGGTGACGCAATTTGGCTGGGTGCAGCCGATACCGCAAGCTATTATGCAGATTGT
AGCCAGTTTAACATCTGCGGTTTTGAAGATGCACTGCGTGATGCAGCCCTGAATCAGACCTTTATTTGTAGCGATACCATTAAAATCCGCAGCC
AGGATATGAGCCAGGCACAGCATCTGGCAGCATGTGATAAAATGGCATATGAAGAAAGCTTTTTCCACACCACACTGGAAACCGGTAATCAGCC
GGTTGCAGATGATCATAATACCCAGCTGCAGGTTAACATCTTTAATAGCGATACGGATTACGGCAAATATGCCGGTCCGATTTTTGGTATTGAT
ACCAATAATGGCGGTATGTACCTGGAAGGTAATCCGGCAAATGTTGGTAATATTCCGAACTTTATTGCCTATGAAGCCAGCTATGCAAACCCGG
ATCATTTTGTTTGGAATCTGGAACACGAATATGTGCATTATCTGGACGGTCGCTTTAATATGTATGGTGATTTTGGTACACCGACCGAACTGGT
TGTGTGGTGGTCCGAAGGTGTTGCAGAATATGTTAGCCGTGTTAATGATAATCCGCAGGCAATTGCAACCATTCAGGATGGTAGCACATATACC
CTGGCACAGGTTTTTGATACCACCTATGATGGTTTTGATGTGGATCGTATTTATCGCTGGGGTTATCTGGCAGTTCGTTTTATGTTTGAACGTC
ATCCTGATGAAGTTCAGCGTATGCTGAGCGCAACCCGTCAAGGTCGTTGGGCAGAATATAAAGCAATTATTAGCGGTTGGGCCAATCAGTATCA
GAGCGAATTTGCACAGTGGACCGAAGCACTGGCAAAAGGTGATAGCGGTGCAGGTAATGGTGAAGGCACCGGTAGCGGTAATGAAGGTGGTGGT
GAAAGCGGTGGTAATACCGGTCTGCCGGAAAATTGTGCAGTGCTGCCGAAAATTAGTGATGGTCGCCTGGCACTGGATGAAGCAGCCTGTCTGG
CGGATACCGCGAGCGCAAGTGATGTTCTGTGGTTTAGCATTCCGGCAGTGAGCGAATATCAGACCATTGCAATTACAGCCGGTAATGGCACCGG
TGATCTGACCCTGGAATATAGCAATCTGAATTGGCCAGATGGTACAAATGTTCAGGCAAGCAGCGCAAATATGGGTAATAGCGAATGTATTATC
CTGGAACATCAGGCCAATTATTGGGGCTATCTGAAAGTTAGCGGTAGCTTTGAAAATGCCGCACTGCTGGTTGAAGCAGGTAGCAATCAGTGTC
GTCAGTAACTCGAG

B. licheniformis subtilisin Carlsberg gene with NcoI and XhoI restriction sites highlighted in green.

CCATGGCTCAACCGGCGAAAAATGTTGAAAAGGATTATATTGTCGGATTTAAGTCAGGAGTGAAAACCGCATCTGTCAAAAAGGACATCATCA
AAGAGAGCGGCGGAAAAGTGGACAAGCAGTTTAGAATCATCAACGCGGCAAAAGCGAAGCTAGACAAAGAAGCGCTTAAGGAAGTCAAAAATG
ATCCGGATGTCGCTTATGTGGAAGAGGATCATGTGGCCCATGCCTTGGCGCAAACCGTTCCTTACGGCATTCCTCTCATTAAAGCGGACAAAG
TGCAGGCTCAAGGCTTTAAGGGAGCGAATGTAAAAGTAGCCGTCCTGGATACAGGAATCCAAGCTTCTCATCCGGACTTGAACGTAGTCGGCG
GAGCAAGCTTTGTGGCTGGCGAAGCTTATAACACCGACGGCAACGGACACGGCACACATGTTGCCGGTACAGTAGCTGCGCTTGACAATACAA
CGGGTGTATTAGGCGTTGCGCCAAGCGTATCCTTGTACGCGGTTAAAGTACTGAATTCAAGCGGAAGCGGAACTTACAGCGGCATTGTAAGCG
GAATCGAGTGGGCGACGACAAACGGCATGGATGTTATCAACATGAGTCTTGGAGGACCATCAGGCTCAACAGCGATGAAACAGGCGGTTGACA
ATGCATATGCAAGAGGGGTTGTCGTTGTGGCGGCTGCTGGGAACAGCGGATCTTCAGGAAACACGAATACAATCGGCTATCCTGCGAAATACG
ACTCTGTCATCGCAGTTGGCGCGGTAGACTCTAACAGCAACAGAGCTTCATTTTCCAGCGTCGGAGCAGAGCTTGAAGTCATGGCTCCTGGCG
CAGGCGTGTACAGCACTTACCCAACCAGCACTTATGCAACATTGAACGGAACGTCAATGGCTTCTCCTCATGTAGCGGGAGCAGCAGCTTTGA
TCTTGTCAAAACATCCGAACCTTTCAGCTTCACAAGTCCGCAACCGTCTCTCCAGTACGGCGACTTATTTGGGAAGCTCCTTCTACTATGGAA
AAGGTCTGATCAATGTCGAAGCTGCCGCTCAACTCGAG