Expression feasibility of recombinant enterokinase in *Nicotiana benthamiana*

Obakeng Goitseone Maema

04385861

Dissertation submitted in partial fulfilment of the degree:

Magister Scientia in Biotechnology

University of Pretoria

Faculty of Natural & Agricultural Sciences

School of Biological Sciences

Department of Plant Sciences

Supervisors: Dr Rachel Chikwamba

Dr Ereck Chakauya

June 2014

UNIVERSITY OF PRETORIA

DECLARATION OF ORIGINALITY

This document must be signed and submitted with every essay, report, project, assignment, dissertation and/or thesis.

Full names of student: OBAKENG GOITSEONE MAEMA
Student number: 04385861

Declaration

1. I understand what plagiarism is and am aware of the University's policy in this regard.

2. I declare that this thesis is my own original work. Where other people's work has been used (either from a printed source, Internet or any other source), this has been properly acknowledged and referenced in accordance with departmental requirements.

3. I have not used work previously produced by another student or any other person to hand in as my own.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

SIGNATURE OF STUDENT:.....……..…….

SIGNATURE OF SUPERVISOR:……..

Abstract

Enterokinase (EK) is the main enzyme used to cleave proteins that are produced in fusion format in the expression host. Recombinant bovine and human EK are widely used in cleavage of fusion proteins, because natural enzymes isolated from bovine intestines is prone to gut protease contamination and thus has ethical functional implications. The limiting factor in fusion protein technology is the isolation of the target protein from fusion protein. Enterokinase is predominantly used because of its unique cleavage pattern that cleaves without leaving N-terminal residues on the target protein. Owing to its attractive cleavage pattern and its versatility, EK is frequently used in laboratories. The demand for EK as an industrial reagent is high. Recombinant EK has been expressed and purified from various production systems but the yields are not high enough to meet the growing demand, resulting in high prices of recombinant EK, increasing costs of protein processing. Enterokinase is currently produced in *Escherichia coli*, Chinese hamster ovary (CHO) and yeast cells, but at very low yields resulting in limited availability. Enterokinase as a cleavage reagent remains active at various temperatures, wide pH range and in the presence of detergents and denaturants. These properties make it widely used in process which liberate proteins of interest from fusion constructs; which is the limiting step in recombinant protein yield. It is therefore worthwhile to explore other technologies to produce this high value protein and thus make it more affordable EK to researchers.

The current study explored the feasibility of producing recombinant EK in *Nicotiana benthamiana* as an alternative to the current production systems. Two research questions that were investigated were the role of sub-cellular targeting in the observed level of expression and accumulation of the target protein, as well as the functionality or otherwise of the protein produced in different organelles. Bovine EK light chain gene was cloned into tobacco mosaic virus (TMV)-deconstructed vector and expressed via *Agrobacterium* mediated transfer. Leaves of *N. benthamiana* were transfected by vacuum infiltration to deliver the transformed *Agrobacterium* into the leaves for transient expression targeted to the apoplast and the cytosol. Plant-made EK was isolated from the plant matrix and biochemically analysed by SDS-

PAGE gel, N-terminal peptide sequencing and western blot. EK was successfully visible as a protein band on an SDS-PAGE gel. The protein band was isolated and the sequence was confirmed by N terminal sequencing as well as western blot analysis. Using an anti-EK monoclonal antibody plant made EK was quantified by ELISA, and the expression amounts were established at 40 µg/g fresh weight when targeted to the apoplast. Cytosolic and apoplastic plant-made EK were partially purified and tested for kinetic activity. The activity assay measured cleavage of a thioester, Z-Lys-SBZL by plant-made EK. Plant-made EK was shown to be active at 23 U/ml plant extract. Apoplast targeted EK had higher expression and was more functionally active compared to cytosol targeted EK. In conclusion, the current study demonstrated that the active bovine EK light chain could be expressed in leaves of *N. benthamiana* at detectable amounts making plants a genuine contender for making recombinant EK.

Acknowledgements

The Almighty has been most merciful to me in my life, all the opportunities I have been granted – thanks be to God, and may your name be exalted, honoured and glorified. In all moments in my life, 2 Corinthians 12:9 reigns true.

My gratitude goes to all the CSIR employees, past and present who I have come into contact with and touched my life not just technically but professionally and socially.
A big thank you goes to all my lab mates who have shared the woes and beauty of lab work problems and successes; I could not have had a more knowledgeable group of scientists.

To my supervisors, Dr E Chakauya and Dr R Chikwamba; your invaluable input in this study has helped shaped this dissertation into what it is. Your zest and zeal throughout this study has made this journey with you rewarding. Outside of the formalities, all your wisdom has greatly helped my personal development.

A special appreciation goes to my family; for always being there for me, being encouraging and supportive throughout. I am particularly grateful to my mother, whose unwavering love and support has been eminent – to whom I dedicate this dissertation.

Thanks also goes to the CSIR for the financial support in this study.

List of symbols and abbreviations

%	Percentage
µM	micromolar
BLAST	Basic local alignment search tool
bp	basepair
BSA	Bovine serum albumin
cDNA	Complementary DNA
cds	Coding sequence
cm	centimeter
kDa	Kilo dalton
DNA	Deoxyribose nucleic acid
DsbA	Disulphide oxidoreductase
dsDNA	double-stranded deoxyribose nucleic acid
DTNB	5,5′-Dithiobis(2-nitrobenzoic acid)
EC	Enzyme commission
EK	Enterokinase
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
et al.	*et alii*, and others
GD4K-na	Gly-Asp-Asp-Asp-Asp-Lys-b-naphthylamide
GFP	Green fluorescent protein
GST	Glutathione S-transferase
GYT	Glycerol, yeast extract and tryptone medium
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HRP	horse radish peroxidase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IUB	International Union of Biochemistry
kb	kilo base
kV	Kilovolts
LA	Luria Agar
LB	Luria Bertani
LC MALDI	Liquid chromatography Matrix-assisted laser desorption/ ionization

LEW	Lysis-Equilibrium and Wash buffer
mAb	monoclonal antibody
MES	Morpholinoethanesulfonic acid
mg	milligrams
ml	millilitre
mm	millimeters
MS	Murashige & Skoog
NCBI	National Centre of Biotechnology Information
NEK	Non-codon optimised enterokinase
ng	nanograms
nm	nanometer
nmol	nanomole
^{o}C	Degrees Celsius
OD	Optical density
PCR	Polymerase chain reaction
PDB	Protein database
PEK	Plant codon-optimised enterokinase
pg	picograms
PVDF	Polyvinylidene fluoride
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	Super Optimal broth with Catabolite repression
T-DNA	Transfer DNA
TEMED	Tetramethylethylenediamine
T_i plasmid	Tumour inducing plasmid
TMB	3,3,5,5-tetramethylbenzidine
TNF	Tumour necrosis factor
TRAIL	Tumour necrosis factor-related apoptosis-inducing ligand
Trx	Thioredoxin
Tryp_SPc	Trypsin-like serine protease
UV	Ultraviolet
v/v	volume per volume
Vir	Virulence
w/v	weight per volume

YEP	yeast extract and bacto peptone
Z-Lys-SBzl	N-α-Cbz-L-lysine thiobenzyl ester
µF	microfarad
µl	microlitre
Ω	ohm

Table of Contents

List of Figures

List of tables

Research aims and objectives

The aim of the study is to transiently express functional EK in *Nicotiana benthamiana*

To achieve the main aim the following objectives were followed:

- To identify the EK genes for expression
- To clone EK genes into plant expression vectors
- To transiently express EK in *Nicotiana benthamina* leaves
- To isolate and biochemically characterise plant-made EK by SDS-PAGE, western blotting and protein sequencing
- To quantify plant-produced recombinant EK by ELISA
- To purify recombinant enterokinase from plant matrix
- To determine the activity of plant-made EK

Chapter 1: Introduction and literature review

1. Background

The biopharmaceutical industry publicly disclosed transaction values of 90 billion USD in 2011, derived from the sale of blood coagulants, proteins, vaccines, antimicrobial and anti-viral substances, antibodies and enzymes (Bieri, 2011). Enzymes are among the most versatile of all the biopharmaceutical applications, playing different roles across a wide range of industrial and research categories. For example a leading pharmaceutical company, Novozymes recorded recorded sales figures of 9 million USD in 2010, showing the magnitude of the need and utilization of enzymes from the biopharmaceutical industry (Basaran & Rodríguez-Cerezo, 2008).

Heterologous proteins are diverse in terms of the hosts in which they are expressed, these include viariability in stability, functionality and solubility. There are many factors to consider when scaling up heterologous protein expreession to pilot, demonstration or even commercial scale, thereby making them difficult targets. (Chelur *et al.*, 2008). Strategies for improving host strains and expression systems explore host-specific codon optimization, co-expression with chaperones, mRNA stability, post-translational modification, decreasing the amount of proteolytic degradation as well as the use of secretory pathways. When considering the strategies mentioned earlier for the successful expression of heterologous proteins the addition of fusion tags to an expression construct has been the most effective, ultimately improving the expression of even recombinant proteins that are difficult to express (Chelur *et al.*, 2008). Successes in these developments have led to the increasing use of recombinant proteins in the biopharmaceutical industry.

1.1 Fusion protein technology

Fusion protein technology is currently the most effective technology in enhancing yield of recombinant proteins (Lee *et al*., 2008). There are broadly two types of fusion tags; affinity tags and solubility-enhancing tags. Affinity tags aid in immobilising recombinant proteins but do not enhance solubility whereas solubility-enhancing tags explicitly enhance the solubility of the recombinant protein (Lee *et al*., 2008). The difficulties that are encountered when expressing recombinant proteins is that there will be over-expression and the protein aggregates into insoluble bodies or there is poor yield. The use of fusion protein technology overcomes these difficulties, simplifies target protein purification and improves solubility of recombinant proteins to be able to isolate them. Inclusion of a fusion tag to an expression construct allows for the purification of a protein whose biochemical properties are unknown (Arnau *et al.*, 2006). The tags can improve solubility and yield of many recombinant proteins. Whilst the inclusion of a fusion tag has many advantages, its addition can change protein conformation. These changes include partial loss or loss of biological activity and changes in toxicity. It is therefore necessary to remove the fusion tag before the recombinant protein is utilised. This is the limiting step in fusion protein technology, the separation of target protein from the fusion construct (Lee *et al*., 2008). Chemical cleavage of the fusion partner requires harsh reaction conditions which affects the overall yield of the target protein negatively. Enzymatic cleavage therefore, offers a better alternative for liberaton of the target protein (Kubitzki *et al.*, 2008). Enzymes typically used for chemical cleavage are factor Xa (fXa), thrombin (factor IIa, fIIa) and enterokinase (EK). Thrombin and factor Xa are derived from the blood coagulation cascade and EK is isolated from mammalian intestinal tract, involved in the activation of trypsin (Chelur *et al.*, 2008). Other non-commonly used enzymes include small ubiquitin-like modifier (SUMO), SUMO proteases (Ulp1, Senp2, and SUMOstar); mutated proteases from *Bacillus subtilis* (Chelur *et al.*, 2008). To these available methods of fusion tag removal, EK is an ideal enzyme due to its stringency, unique cleavage pattern and accomodative reaction conditions (Ogiwara & Takahashi, 2007).

1.1.1 What are fusion proteins?

Fusion partners are protein peptides added to N- or C-terminus of an expression construct, with the aim of facilitating the target protein in better expression, translocation, solubility or stability. These are peptides that could also facilitate protein secretion, protein labelling, cell surface expression, cytosol to membrane translocation and cytosol to nucleus translocation (Chelur *et al*., 2008). Fusion protein technology is the most effective technology in enhancing the production and solubility of recombinant proteins (Gasparian *et al*., 2007). Fusion protein technology enhances the production and purification of recombinant proteins such as antibodies, coagulation factors, vaccines and growth hormones. The addition of a fusion tag has been reported to facilitate the purification of a recombinant protein whose biochemical characteristics hasn't been determied (Arnau *et al.*, 2006; Gasparian *et al.*, 2011). Fusing the coding sequence of a gene of interest to that of a well-expressed fusion partner has several advantages. Fusion protein strategies typically place the gene of interest at the C-terminal of the highly expressed fusion partner, therefore allowing transcription to occur on the partner to allow high expression of the gene of interest (Gasparian *et al.*, 2011). The other advantages of having a fusion partner include aiding in specific cellular localization, formation or hydrophobic bodies and proteolytic or conformational stability (Waugh, 2005; LaVallie *et al.*, 1993). Fusion tags also aid in the purification of recombinant proteins either by affinity purification followed by liberation of the protein of interest (Waugh, 2005). Cleavage of the tag by incorporating a cleavage site and then using one of the widely available endoproteases such as EK, thrombin, urokinase or factor Xa (Gasparian *et al.*, 2011).

Affinity tags directly aid purification of recombinant proteins and solubility-enhancing tags indirectly aid in purification by making the recombinant protein more soluble and consequently more accessible for isolation (Arnau *et al*., 2006).

1.1.1.1 Types of fusion tags

Fusion tags are an amino acid sequence which have high affinity for a chemical/metal or specific biological ligand. The greater majority of affinity tags are

affinity tags that bind a small peptide on the expression construct to an immobilised (e.g., HIS_6-tags bind to immobilised metals). The other group of tags bind to an immobilised protein such as an antibody such as protein A affinity chromatography. Histidine tags (HIS_6-tags) are the most widely used affinity tags because they have a high metal affinity. Chelated metal ions are used as the affinity ligands that the HIS-tag on the recombinant protein binds (Chelur *et al*., 2008 & Arnau *et al.,* 2006). The metal ions are usually complexed with an immobilised chelating agent i.e. immobilised metal-ion affinity chromatography (IMAC). The basis of IMAC is that the immobilised metal chelate interact with the amino acid residues i.e histidine residues in the case of HIS_6-tags. Addition of affinity tags on either the N-terminus or C-terminus of the recombinant proteins allows access to the metal chelate, where the affinity tag can interact with the metal chelate thereby capturing the tagged protein out of a crude extract. The most commonly used IMACs are either nitrilotriacetic acid (NTA) or nickel in affinity chromatography (Ni–NTA) for immobilizing metals different chelating Sepharose matrices. (Arnau *et al.,* 2006).

Solubility-enhancing peptides added to an expression construct which increase the solubility of fusion proteins, thereby yielding a higher yield of recombinant protein. Fusion tags like glutathione S-transferase (GST) and maltose-binding protein (MBP) are generally used for protein purification because they are affinity tags. Fusion tags like thioredoxin (TRX), NusA, ubiquitin (Ub) and small ubiquitin-like modifier (SUMO) require additional affinity tags thereby increasing the input costs for purification (Chelur *et al*., 2008 & Lee *et al*., 2008). Some fusion tags are more effecient than others in increasing the solubility of many proteins, due to the variation of biochemical properties of the expression system and recombinant biochemical properties. Smaller affinity tags like GST are generally less effective than the large proteins like MBP and NusA in solubilising proteins (Chelur *et al*., 2008). It is therefore worth noting that variable combinations of fusion tags need to be explored for each target protein as they do not have the same biochemical properties. SUMO fusion tags are currently utilised as an alternative for increasing both the solubility and expression of low yielding and recombinant proteins. The SUMO fusion system is not dependent on the specific sequence with SUMO for cleavage the SUMO protease recognises the conformation of SUMO protein and cleanly excises the

target protein (Arnau *et al.*, 2006 & Chelur *et al*., 2008). The ideal enzymatic cleavage method requires that the cleavage should not alter the recombinant product, should not affect the internal target protein sequence and the method should be tolerant of a wide range of conditions. The cleavage conditions dictated by the fusion partner should not affect the efficacy of cleavage and the recombinant protein (Hunt, 2005). More importantly, the cleavage reagent should not be derived from mammalian source for concerns of contamination of infectious agents (Gasparian *et al.*, 2011, Esposito & Chatterjee, 2006)

1.1.2 Fusion protein cleavage reagents

Biochemical studies and therapeutic use of target proteins require that the fusion tags be removed and the maintenance of a stable soluble target protein of interest. Typically a linker is designed between the fusion tag and target protein, the linker can aid in refolding or be a protease cleavage site permitting *in vitro* excising of the fusion tag (Lee *et al.*, 2008). Enzymatic cleavage offers higher specificity than chemical cleavage, where chemical cleavage involves harsher reaction conditions for cleavage. Chemical cleavage involves the use of a unique methionine residue between the fusion tag and the target protein and then cleavage is performed by using cyanogen bromide or hydroxyxylamine. Chemical cleavage methods are unspecific, protein denaturation is eminent and there is side chain modification of amino acid residues in the target protein (Arnau *et al.,* 2006).

Efficacious site-specific hydrolysis of fusion proteins is essential to liberate the target proteins from their fusion partners. Several proteases have been identified and are being used in cleavage of fusion partners, due to their site-specific cleavage (Table 1-1). Thrombin is used to excise fusion tags in a fusion protein construct, with the inclusion of the thrombin cleavage site (Leu-Val-Pro-Arg*-Gly-Ser) on the N-terminal of the protein of interest. Cleavage by thrombin occurs between arginine and glycine, leaving gly-ser amino acid residues on the N-terminus of the recombinant product (Lu *et al.*, 1999 & Arnau *et al*., 2006). Factor Xa is a serine endopeptidase with cleavage site Ile-Glu/Asp-Gly-Arg*, cleaving after the arginine residue. Factor Xa and thrombin both have restrictive reaction conditions. Enterokinase is a serine

endopeptidase that has the cleavage site Asp-Asp-Asp-Asp-Lys*, cleavage occurs after the lysine residue (Lu *et al.*, 1999) (Figure 1-1). Fusion tag cleavage using other endoproteases (Table 1-1) results in amino acid residues remaining on the N-terminus of the target protein from the cleavage, which may change biochemical characteristics.

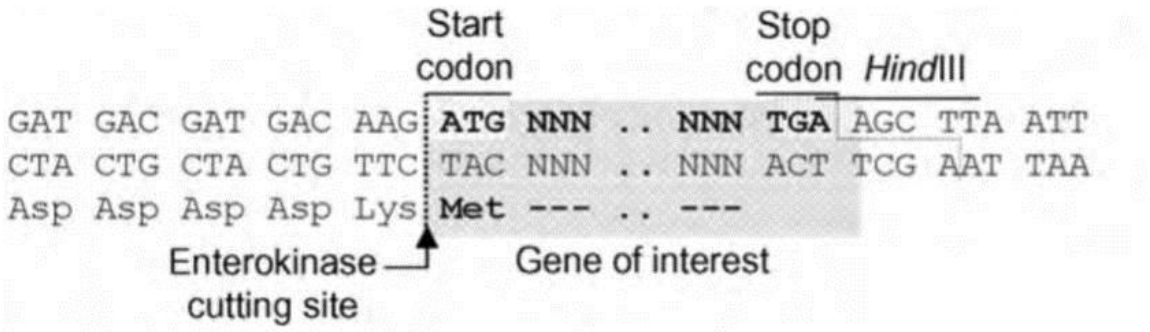


Figure 1-1: Schematic representation of the incorporation of EK cleavage site in a fusion protein partner, indicative of the cleavage site (Santana *et al.*, 2010)

Enterokinase is an ideal enzyme to use for fusion protein technology because it cleaves after its recognition sequence with high specificity, and has a wider temperature, buffer and pH range. High specificity of EK to D_4K site has led to its use in *in vitro* cleavage of fusion proteins (Gasparian *et al*., 2011, Kubitzki *et al.*, 2008). Enterokinase cleaves C-terminus without leaving undesired amino acids on the N-terminal of the target protein, which is why it is an ideal enzyme (Chun *et al*., 2011). The wide range of reaction conditions where EK retains its activity makes it a flexible enzyme wherein various buffers and physiological conditions the target protein will remain soluble. This is helpful where there is not much knowledge available on the target protein in a particular expression system, the target protein can then be removed using the fusion tag's biochemical properties (Liew *et al.*, 2007).

In some instances the target protein does not stay soluble after cleavage of the fusion tag because it was held in solution by the solubility enhancing tag, thus fusion tag removal makes the target protein revert back to its aggregate form. Enterokinase can aid in this predicament, since it is active across a number of detergents or

denaturants that will keep the target protein from re-folding while cleaving away the fusion partner (Esposito, 2006).

1.2 Enterokinase

Enterokinase is an important reagent in the biotechnological and biopharmaceutical industries for the cleavage of fusion proteins. Enterokinase has a high degree of specificity in recognition of its cleavage sequence and is tolerant to a wide range of reaction conditions, which meets the requirements of an ideal fusion tag cleavage agent (LaVallie *et al*., 1993). Several commercial EK are purified as holoenzymes from porcine or bovine intestines which is obtained in small quantities thereby making the production cost very expensive. Even highly purified native EK is prone to be contaminated by traces of other gut proteases (Fang *et al.*, 2004). To circumvent this challenge, recombinant EK is produced recombinantly in production systems and thus used for fusion partner cleavage (Yuan & Hua, 2002). The release of the fusion partner in a fusion construct is the most limiting step and it is where EK is most useful. The advantages of EK are that it remains active across a wide range of detergents and buffers for ease of concevience of isolation of target protein. The importance of EK is evident in research laboratories that require target proteins in high yields where use of fusion tags aids in expression or isolation of the target protein.

Table 1-1: Endoprotease fusion tag removal

Enzyme	Cleavage Site	Comments
Enterokinase	DDDDK*	Secondary sites at other basic amino acids
Factor Xa	IDGR*	Secondary sites at GR
Thrombin	LVPR***GS**	Secondary sites. Biotin labelled for removal of the protease
PreScission	LEVLFQ***GP**	GST tag for removal of the protease
TEV protease	EQLYFQ***G**	His-tag for removal of the protease
3C protease	ETLFQ***GP**	GST tag for removal of the protease
Sortase A	LPET***G**	Ca^{2+}-induction of cleavage, requires an additional affinity tag (e.g., his-tag) for on column tag removal
Granzyme	B D***X**, N***X**, M***N,** S***X**	Serine protease. Risk for unspecific cleavage
Intein	Self-cleavable	Artificial amino acids left after cleavage in some applications, on column (chitin-beads) cleavage
SUMO	Conformation	No affinity purification, requires HIS-tag
DAPase (TAGZyme)	Exo(di)peptidase	Cleaves at the N-terminus. His-tag (C-terminal) for purification and removal
Aeromonas aminopeptidase	Exopeptidase	Cleaves at the N-terminus, effective on M, L. Requires Zn
Aminopeptidase M	Exopeptidase	Cleaves at the N-terminus, does not cleave X-P
Carboxypeptidase A	Exopeptidase	Cleaves at the C-terminus. No cleavage at X-R, P
Carboxypeptidase B	Exopeptidase	Cleaves at the C-terminus R, K

The position of endoprotease cleavage is indicated with an asterisk (*). Residues in bold remain in the protein after endoprotease cleavage. Adapted from Arnau *et al*., 2006

Enterokinase (Enzyme Commission, EC 3.4.21.9) was discovered in I. P. Pavlov's laboratory by N. P. Schepowalnikow in 1899 (Kunitz, 1939). The name enterokinase is a poor description because as an enzyme it is a proteolytic enzyme and not a kinase. The International Union of Biochemistry (IUB) designation, enteropeptidase, is indicative of its proteolytic activity (Light & Janska, 1989).

Enterokinase is a serine protease bound to the mucosa membrane that catalyses the cleavage of peptide bonds. It is a heterodimer glycoprotein made up of two covalently linked subunits, heavy chain and light chain joined by a disulphide bond (Lu *et al*., 1999). The heavy chain contains an amino-terminal membrane spanning domain with very little effect on the recognition of small peptides, although the heavy chain influences inhibitor specificity and macromolecular substrate recognition (Gasparian *et al.*, 2003). The light chain known as the catalytic subunit, has a chymotrypsin-like protease domain exposed to the luminal contents (Figure 1-2) where it can cleave-activate trypsinogen (Chun *et al.*, 2011 & Huang *et al*., 2007). Native EK activates typsinogen to trypsin in the duodenal mucosa; cleavage of trypsinogen is at the EK cleavage site present in the activation peptide of trypsinogen, pictured in Figure 1-3. Specific cleavage of trypsinogen is at the C-terminal of the scissile bond of the lysine residue (Kim *et al*., 2008; Peng *et al*., 2004). Trypsinogen in mammals has a highly conserved amino terminus activation tetra aspartyl-lysl peptide, Asp-Asp-Asp-Asp-Lys (ASP_4-Lys, DDDDK, or D_4K) in the activation peptide (Kubitzki *et al*., 2008). The heterodimer form of EK has been isolated and purified from the following organisms; porcine, bovine, human, rat, mouse and ostrich intestines (Kitamoto *et al*, 1995). It is only human EK that has three subunits as shown in Table 1-2(Simeonov *et al*., 2011; Zhang *et al*., 2009).

The heterodimer form of EK has been extensively studied. Enterokinase has three histidine-serine catalytic residues, characteristic of serine proteases at H57, S195 and S214 being responsible for cleavage (Liew *et al.*, 2007). It has two polypeptide chains linking the heavy and light chains, for the bovine and porcine variants and three polypeptide chains in the human variant as tabulated in Table 1-2 (Song *et al.*, 2002). In most preparations, EK is a disulphide-linked heterodimer derived from a

single chain precursor. Both chains contains 30 – 50 % carbohydrates and 40 % of the molecular mass is provided by glycosylation (Kitamoto *et al*., 1995). The light chain of EK has four disulphide bonds C27 – C43, C121 – C186, C153 – C168 and C176 – C205 (Lu *et al.*, 1999).

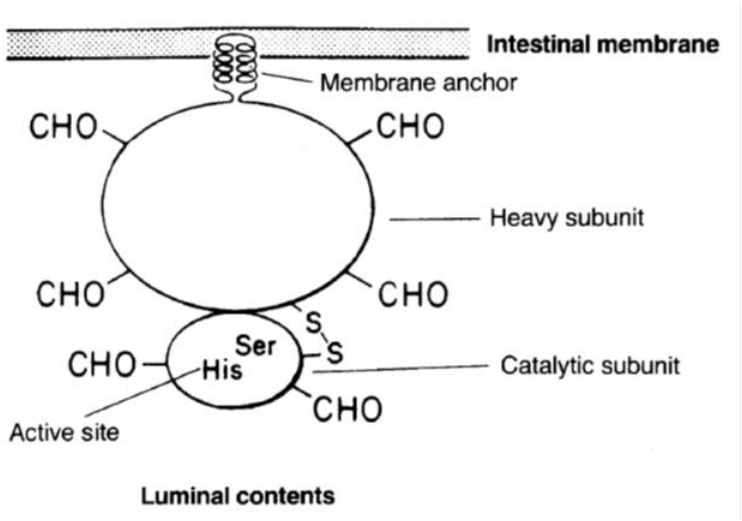


Figure 1-2 A schematic representation of EK in the duodenal mucosa. The heavy chain is the membrane anchor to the intestinal membrane and the active site residues on the light chain is directed towards the luminal contents (Light & Janska, 1989).

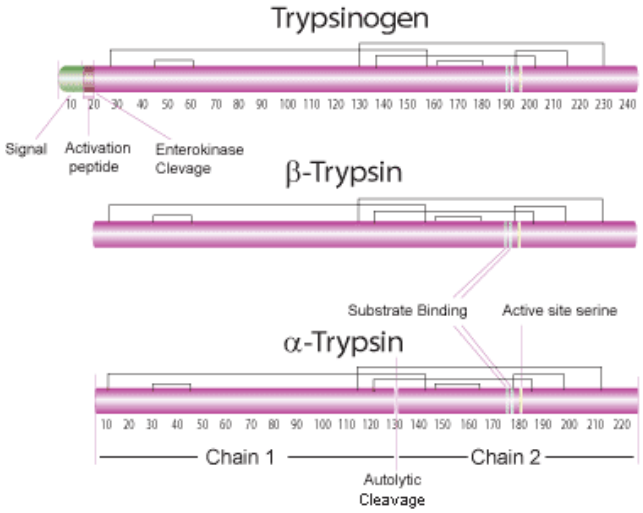


Figure 1-3: *In vivo* processing of trypsinogen by EK

(http://www.sigmaaldrich.com/catalog/product/sigma/e4906?lang=en&region=ZA)

trypsinogen + EK → trypsin + pro-region *(Val-Asp-Asp-Asp-Asp-Lys)*

Table 1-2: Molecular weights of EK subunits from different mammalian sources

		Molecular weight (kDa)	
Species	**Intact (kDa)**	**Heavy chain**	**Light chain**
Human	300	140 102*	54
Cow	150	115	35
Pig	200	134	62

***= human EK has three chain constituents**
(http://www.brenda-enzymes.org/php/result_flat.php4?ecno=3.4.21.9)

1.2.1 Discovery and isolation of EK

In 1910, Pavlov investigated the physiological controls of gastric and pancreatic secretions in dogs and discovered that food that is mechanically eaten by dogs elicits different gastric secretions than food directly introduced into the stomach of the dog. He noted the salivary, gastric and pancreatic differences as decreased digestive activity in the digestion of food (Maroux *et al.*, 1971). Gastric secretions were isolated from the dogs via a chronic gastric fistula which was a mixture of partially digested food along with salivary and gastric juices. It was this mixture that was used in clinical treatment of poor appetite (Maroux *et al.*, 1971). Schepowalnickow in Pavlov's laboratory discovered that a proteolytic activity was generated in pancreatic juice upon addition of duodenal juice. The agent responsible for this effect was enterokinase, which was responsible for the conversion of trypsinogen into trypsin (Maroux *et al.*, 1971; Smith, 1995). The kinetic activation assays performed on fresh beef extracts by Kunitz (1939) showed that EK exclusively converts crystalline trypsinogen to trypsin, which is a stoichiometric combination between trypsinogen and EK to form an active trypsin-kinase enzyme (Kunitz, 1939).

1.2.2 Importance of EK in mammals

Enterokinase is involved in mammalian digestion in a cascade of pancreatic digestive enzymes. Upon entry of food into the luminal area, EK is secreted from the crypts of Lieberkühn. The secretion of EK cleave-activates the zymogen trypsinogen to trypsin. Pancreatic trypsin is an important digestive enzyme because it is the sole

activator for the chymotrypsinogens, the procarboxypeptidases, proelastases and prophospholipase in the duodenum illustrated in Figure 1-4 (Simeonov *et al*., 2011). In humans, a nonsense or frameshift mutation in the ENTK/PRSS7 gene, on chromosome 21q21 leads to a rare recessively inherited disorder called congenital EK deficiency. Infants with congenital EK deficiency are deficient in the production of EK. Enterokinase is necessary to convert trypsinogen into trypsin which sequentially, activates the other pancreatic zymogens necessary for digestion. Duodenal juice of infants with congenital EK deficiency cannot activate trypsin *in vivo* and thus the digestive cascade is halted. Congenital EK deficiency is characterised by diarrhea, edema, hypoproteinemia and failure to thrive. It results in death in infants born with this deficiency (Holzinger *et al*., 2002).

Duodenal juice of infants with congenital EK deficiency can be supplemented with purified EK which results in activation of proteolytic enzymes and thus proper digestion occurs and the infants are able to thrive. Supplementation of purified EK works when there are inactive zymogens which are then activated. In the past, this condition was treated with administration of pancreatic juices. The pancreatic juices have trypsin present which has a feedback action to activate trypsinogen in the absence of EK, as illustrated in Figure 1-4 (LaVallie, 1997). Recombinant EK may be used as a therapy for congenital EK deficiency, where when formulated properly for oral administration, the enzyme enters the duodenum and activated trypsinogen and subsequently the inactive pancreatic zymogens. The presence of the heavy chain is crucial for correct substrate-enzyme complex formation with trypsinogen substrate (LaVallie, 1997; Light & Janska, 1989). The human varianat of EK gene may be used in gene therapy to treat congenital EK deficiency (LaVallie, 1997).

Enterokinase has, since its discovery been studied by researchers to characterise and elucidate its function, work that was pioneered by the separation of the heavy chain from the light chain by Light and Janska (1989). They investigated the function of the two subunits individually of native EK obtained from bovine intestines, once separated. Separation was achieved by limited reduction of the disulphide bonds of EK using 50 mM dithioerythritol pH 9 at 4 ^{o}C which caused the release of the light chain from the heterodimer. The isolated catalytic subunit displayed the same

functionality as the native heterodimer enzyme upon cleavage of small substrates such as trypsin. The catalytic subunit however, displayed lower and varied (from 7 to 60 %) activity towards larger substrate such as trypsinogen. This suggested that the histidine and serine residues of the catalytic subunit functioned as expected catalytic residues of other serine proteases within that family. The absence of the heavy chain no longer produced the full productive enzyme substrate complex, when larger substrates such as trypsinogen were used. Therefore, Light *et al.* (1989) was the first to separate EK into its two subunits and showed that the light chain of EK has catalytic activity while the heavy chain is required for larger substrate recognition and binding (Light & Janska, 1989).

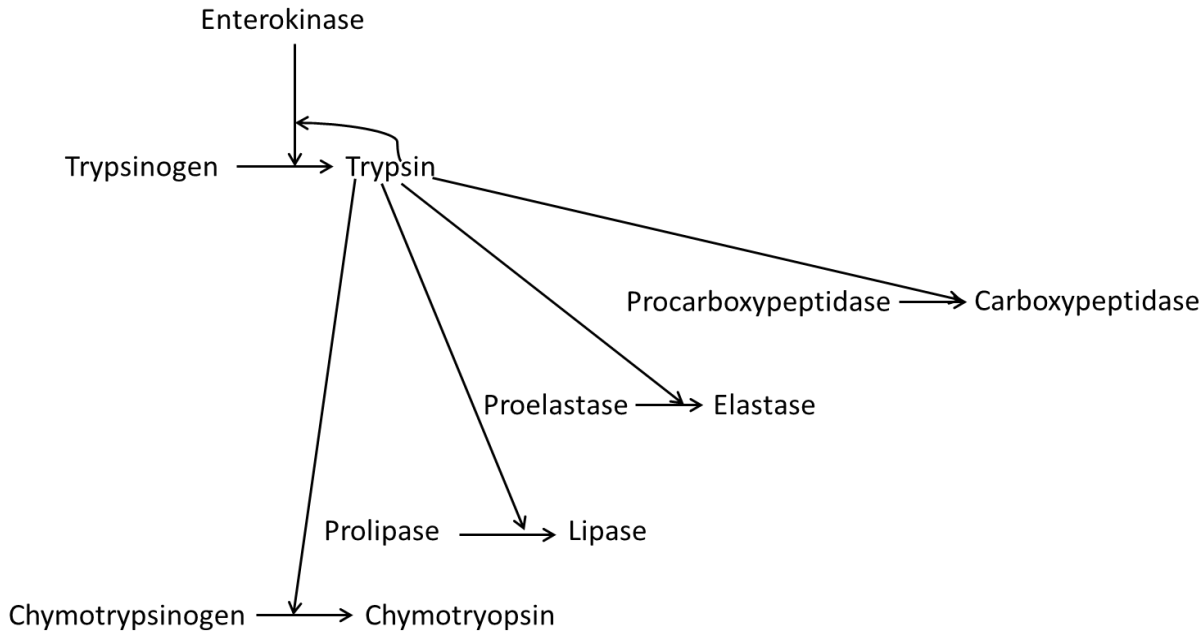


Figure 1-4: Activation cascade of trypsin by EK in the duodenal mucosa. (http://www.ncbi.nlm.nih.gov/bookshelf/picrender.fcgi?book=stryer&part=A1378&blobname=ch10f34.jpg**)**

1.2.3 Structural elucidation of EK

The heterodimer form of EK has been purified from bovine, porcine, human and ostrich intestine for structural elucidation and cDNA cloning. The catalytic subunit of EK has the desired activity and therefore most research was focused on the light chain. The serine protease domain contains a basic tetra-peptide consisting of R96-

R-R-K99 for porcine, mouse and human EK and K96-R-R-K99 for bovine and rat (Kitamoto *et al.*, 1995). The tetra-peptide region is not preserved among the serine proteases within the chymotrypsin-like serine proteases family, where it binds the acidic P2 - P5 residues of trypsinogen activation peptide in enzyme-substrate interaction. The catalytic mechanism which recognises the P1 lysine residue is conserved within the family, but the aspartyl side chains at P2 - P4 on the substrate are accommodated mainly by ionic interactions (Lu *et al*., 1999; Maroux *et al.*, 1971). Based on the sequence homology to other serine proteases, EK has a similar folding pattern to the representative serine proteases such as chymotrypsin and thrombin (Figure 1-5). The tertiary structure of EK has two six stranded β-barrels, that make up the bulk of the enzyme (Lu *et al*., 1999; Maroux *et al.*, 1971).

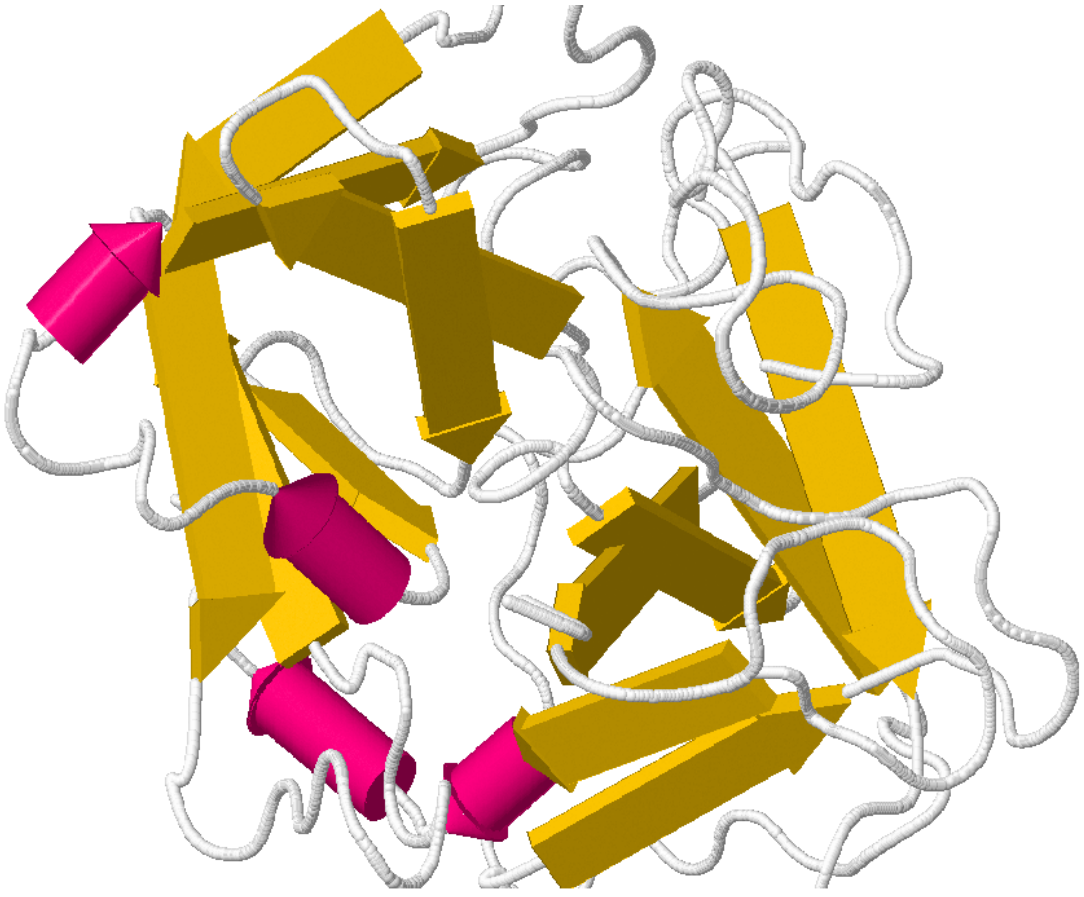


Figure 1-5: Rockets rendering model of EK light chain showing beta sheets and alpha helices. Flat arrow = β-sheet; round arrow = α-helix. http://helixweb.nih.gov/cgi-bin/moldraw?1351759222279

The catalytic centre contains serine proteases signature structural elements; the catalytic triad consisting of H57, S195 and S214 and the oxyanion hole. The oxyanion cavity is formed by the main N atoms 193 and 195 and the specificity pocket 1 which interacts with the P1 side-chain on the substrate (Lu *et al*., 1999). Lu *et al.* (1999) investigated the interactions of a VD_4K-chloromethane inhibitor on EK, which is structurally similar to trypsinogen activation peptide. Enterokinase was observed to bind with the catalytic residues H57 and S195 via the C-terminal Lys-P1 to the VD4K-cm inhibitor. The methylene carbon atom of the VD4K-cm was bound to the imidazole ring of H57. The carbonyl carbon atom of Lys-P1 formed a tetrahedral hemiketal with S195; it is the aliphatic part of the Lys-P1 side-chain that packs against the main chain atoms of P215, S214 and T213. This investigation showed the interactions of the trypsinogen activation peptide mimic with EK triad catalytic residues (Lu *et al*., 1999; Maroux *et al.*, 1971).

Lu *et al* (1999) determined that P1, P2 and P3 residues are critical in binding but a prerequisite is that P4 and P5 must be acidic for increased affinity of EK. The residue at P1 can either be lysine or arginine; residues at P2 and P3 must be acidic; being aspartic acid, glutamic acid or carboxy methylcysteine (Lu *et al*., 1999; Maroux *et al.*, 1971).

1.2.4 Conserved domains of EK

Enterokinase belongs to the trypsin-like serine protease (Tryp_SPc) family. As a member of the Tryp_SPc family, EK shares properties with trypsin-like serine proteases, characterised by the catalytic residues histidine and serine (H57, S195 and S214). Proteases in this family are synthesised as inactive precursor zymogens that are cleaved during limited proteolysis into their active forms. Active site prediction of EK light chain was derived from catalytic triad for serine proteases that exist in trypsin-like serine proteases (Huang *et al.*, 2007) (Figure 1-6). Bovine EK was aligned with trypsin-like serine protease to show their similarities in non/conserved domains and active site prediction (Marchler-Bauer *et al*. 2011). The alignment in Figure 1-6 shows the relationship of bovine EK light chain to the Tryp_SPc family, as being a highly conserved relationship because there is high

sequence homology between the two aligned sequences. This reiterates that EK belongs to the Tryp_SPc family and shares its characteristics, including the histidine and serine catalytic residues. The presence of the trypsin-like serine protease domain is sufficient for normal recognition of EK substrate and cleavage (Huang *et al.*, 2007).

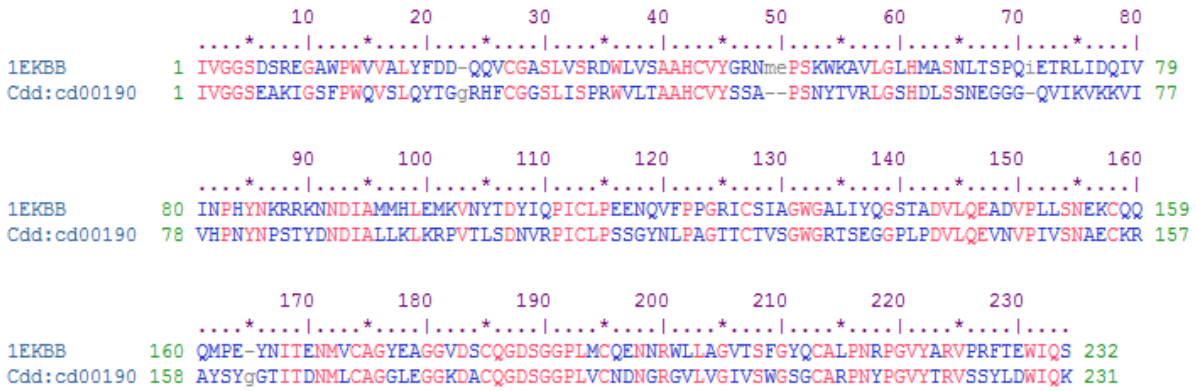


Figure 1-6: Conserved domains on bovine EK light chain. Alignment of bovine EK light chain protein sequence against conserved domains of trypsin-like protease family. Red = matches in alignment; blue = mismatch in alignment; grey = additional amino acid(s)

1.2.5 Biotechnological applications of EK

Enterokinase is documented as an enzyme that activates trypsinogen, biotechnological applications of this enzyme use the catalytic subunit that has been shown to retain its activity and specificity without the heavy chain (Light & Janska, 1989). The light chain retains its activity across a range of conditions; wide temperature (4 – 45 ^{o}C) and pH (4.5 – 9.5) range also in the presence of numerous detergents and denaturants. These properties of EK make it an enzyme that can be utilised in many reaction conditions without loss of activity (Tan *et al*., 2007).

Enterokinase was successfully used as a protease for cleavage in Trx/TRAIL fusion cleavage. The extracellular domain of human tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) is a candidate for clinical trials in cancer therapy. The extracellular portion was expressed with an added initiator Met codon (Gasparian *et al*., 2007). Human TRAIL is very stringent in its composition; the initial

peptide is valine but valine and threonine at P10 are poor substrates of methionine amino-peptidases type-I and type-II. The inclusion of the EK cleavage site enabled TRAIL to initiate with a methionine, which elevated the poor substrate interaction with amino-peptidases. Incorrectly processed N-terminus of TRAIL would result in high antigenecity (Gasparian *et al*., 2007). Enterokinase was the ideal protease to separate human TRAIL from thioredoixin fusion partner, expression yield of at 400 mg/L *E. coli* cell culture bioactive TRAIL was measured from 1.5 g/L trx/TRAIL fusion (Gasparian *et al*., 2007). This shows an excellent example of the applicability of EK as cleavage reagent in recombinant proteins intended for human therapeutics taking advantage of EK's unique cleavage.

1.2.5.1 Production of recombinant EK

Due to the ability of EK to cleave on the C-terminus of its recognition sequence made it an ideal candidate for use in fusion protein technology, without leaving amino acid residues on the N-terminus of the target protein (Lu *et al.,*1999). Enterokinase utility has been commercialised to a limited extent because the isolation from bovine and porcine has ethical implications and prone to protease contamination. The most commonly expressed mammalian EK is the bovine and human variant, rarely the porcine variant (Gasparian *et al*., 2007). The bovine variant has been widely studied and characterised. The annotated EK structure in Figure 1-7 shows the characterised features related to structure, composition and predicted features.

As a recombinant protein the catalytic subunit of EK has been shown to retain activity without the presence of the heavy chain subunit, therefore the expression of the catalytic subunit will be the focus of the study. The cDNA encoding EK light chain has been isolated from several sources; bovine, porcine, human, rat, mouse and medaka hosts (Gasparian *et al*., 2007). Recombinant EK has been expressed in various expression systems including CV-1 origin [SV40] (COS) cell lines, *Escherichia coli*, *Aspergillus niger, Pichia pastoris* and *Saccharomyces cerevisiae*. Where undesired or undetectable levels of EK were obtained different cellular targets

and use of fusion tags was explored. The results of previously expressed EK in different expression systems is summarised in Table 1-3.

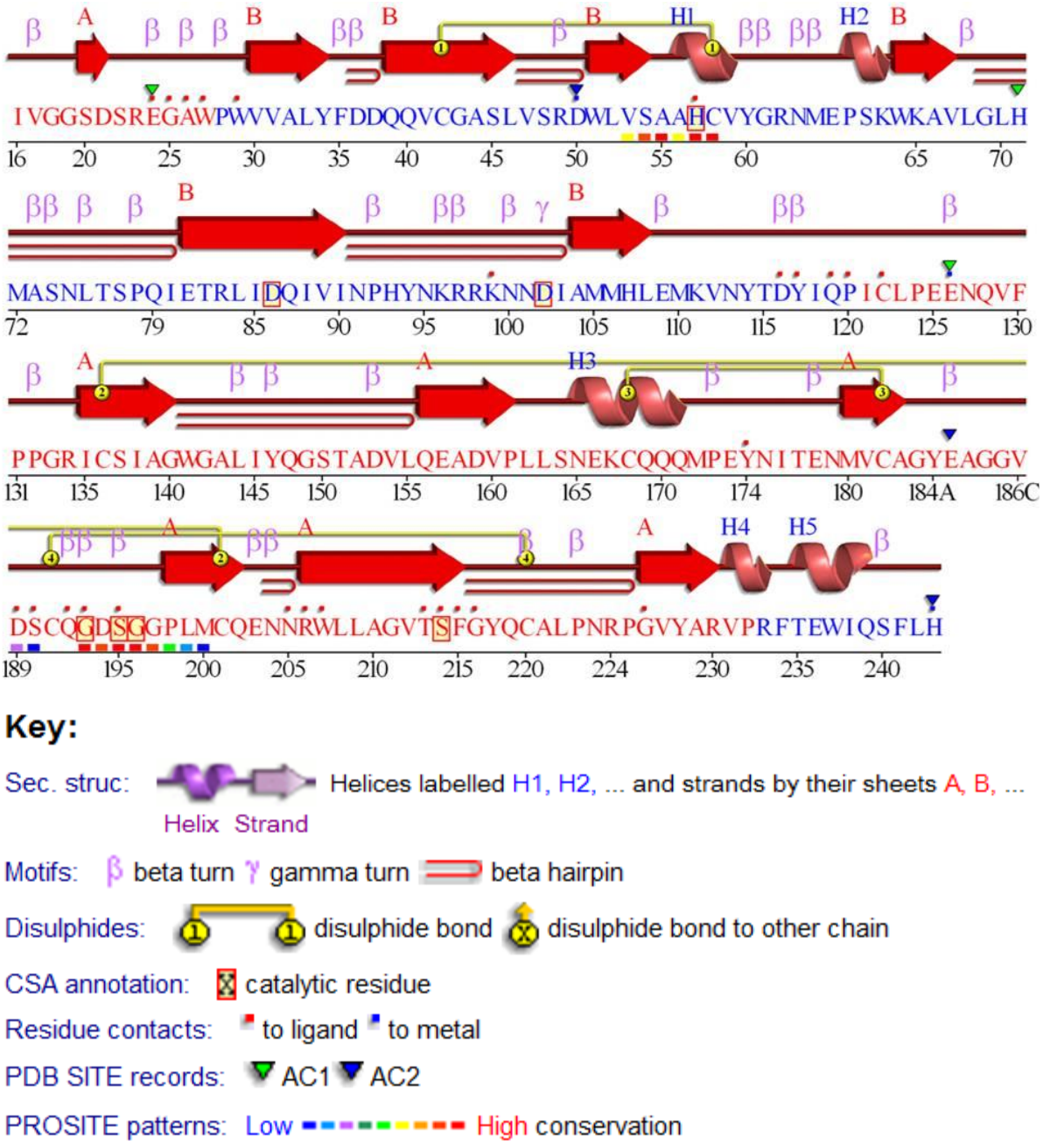


Figure 1-7: Annotated structure of EK indicative of the structural features and components (http://www.rcsb.org/pdb/explore/remediatedChain.do?chainId=B**)**

1.2.5.2 Expression yields of recombinant EK

The first report of recombinant expression of EK was in *E. coli*, which was at very low yields of 8 µg EK per 1 g cell paste, even with the use of novel secretory fusion partner disulfide oxidoreductase (DsbA) (Chun *et al.*, 2011). When bovine EK was successfully expressed and purified from *E. coli*, it was found to be partially active. It could only partially cleave D_4K containing fusion proteins (Chun *et al.*, 2011). Recombinant EK was expressed in *Saccharomyces cerevisiae* at 3.8 mg/L under optimised conditions using ethanol induction (Kim *et al.,* 2005). Tan *et al.* (2007) reported the expression of auto-cleaving GST-tagged EK in *E. coli* at 53 mg/500 ml cell culture, using IPTG and glucose induction. The purified EK shows specific activity of 110 ± 10 U/mg compared to EK purchased from SIGMA-ALDRICH at 20 U/mg (Tan *et al.*, 2007). Huang *et al.* (2007) also reported on the isolation of bioactive recombinant EK from *E. coli* at 6.8 mg/L (Huang *et al*., 2007). Expression of recombinant EK has been chronologically tabulated in Table 1-3.

Table 1-3: Summary of expression yields of recombinant EK

Host cell	Purified EK (mg/L)	Specific activity (U/mg purified protein)	Reference
Mammalian CHO cells	low amounts of secreted protein	low levels of activity	(LaVallie *et al*., 1993)
P. pastoris	6.3	Not determined	(Vozza *et al.,* 1996)
A. niger	1.9	19.88	(Svetina *et al.*, 2000)
S. cerevisiae	1	Not determined	(Choi *et al.*, 2001)
E. coli	43	720	(Yuan & Hua, 2002)
E. coli	20	no auto-cleavage activity	(Gasparian *et al.*, 2003)
P. pastoris	150	9000	(Peng *et al*., 2004)
P. pastoris	5.4	Cleavage activity towards fusion protein containing cleavage site	(Fang *et al.*, 2004)
P. pastoris	10.9	2.88×10^7	(Fang *et al*., 2004)
S. cerevisiae	3.8	active protein	(Kim *et al.*, 2005)
E. coli	106	110 ± 10	(Tan *et al.*, 2007)
E. coli	6.8	20	(Huang *et al*., 2007)
P. pastoris	479.99	13619.14 U/L	(Zhang *et al*., 2009)

Adapted from Chun *et al*., 2011

1.2.5.2.1 Kinetic parameters of recombinant EKn

The light chains of recombinant human and bovine EK are the commonly used for fusion tag removal processes. Gasparian *et al.* (2011) compared their kinetic parameters toward different substrates in Table 1-4. The catalytic constant (k_{cat}) of human EK to GD_4K-na substrate is 10-fold higher than bovine EK (Gasparian *et al.,* 2011). The table also indicates the relative activity units for both variants used in an assay of the human and bovine enzymes were the same for the substrate N-α-Cbz-L-lysine thiobenzyl ester (Z-Lys-SBzl) (Gasparian *et al.*, 2011). It is however worth noting that in using different mutant substrates, they discovered that bovine EK has higher specificity compared to human EK, in native and mutated cleavage substrates (Gasparian *et al.*, 2011). This clearly demonstrates that although human EK has 10-fold more activity than bovine EK, bovine EK is more desirable for cleavage due to its higher substrate specificity.

Table 1-4: Kinetic parameters of bovine and human EK to different substrates

	Bovine EK			Human EK		
Substrate	**kcat (s^{-1})**	**Km (μM)**	**kcat/Km ($mM^{-1}s^{-1}$)**	**kcat (s^{-1})**	**Km (μM)**	**kcat/Km ($mM^{-1}s^{-1}$)**
GD_4K-na*	49.3 ± 2.1	141.3 ± 3	347 ± 13	118 ± 5.1	34 ± 2	3500 ± 53
Z-Lys-SBzl	129 ± 4.0	120 ± 6	969 ± 30	133 ± 4.0	140 ± 7	950 ± 30

Adapted from Gasparian *et al.*, 2011

1.2.5.2.2 Non-specific cleavage by EK

The use of EK as a biotechnological reagent is limited to its sporadic cleavage at other sites in fusion protein cleavage, which consists of acidic and basic residues (Chun *et al*., 2011 ; Kim *et al*., 2008). Shahravan *et al.* (2008) showed that the non-specific cleavage by EK was significantly lowered when 1 – 4M urea was added to the fusion protein cleavage reaction. The work showed that EK as a reactive enzyme displays non-specific cleavage when it does not have access to the cleavage site due to aggregation, dimerization or multimerization of the fusion construct. Fusion protein denaturation by ≤ 4M urea prior to EK cleavage has been showed to deter

sporadic cleavage without modification of fusion construct or cleavage site (Shahravan *et al.,* 2008). Treatment by ≤ 4M urea allows the fusion construct cleavage site to be accessible to EK for cleavage and thus decreasing the sporadic cleavage at non-canonical sites, which had been reported prior to Shahravan's *et al* work.

Shahravan's *et al* work has lead to an inexpensive way of circumventing the secondary activity of EK, which is undesired in fusion construct cleavage (Shahravan *et al.,* 2008). Enterokinase has been expressed in many expression systems, with numerous optimisations ultimately expressed as an active enzyme in gradually increasing expression yields. As far as the production of EK goes, there is another expression host that EK has not to our knowledge been expressed in. This research project explores using a plant production systems to express functionally active EK .

1.3 Recombinant expression systems

Fermentation systems for microbial and animal cell cultures require expensive equipment and media for recombinant protein production; whereas plants are low cost – requiring only sunlight, water, $CO_{2,}$ fertilizer and can be grown in existing agricultural resources. Plants are simple to grow and easily manipulated for heterologous protein production. The expertise required for maintaining and running bioreactors requires specialised training meanwhile plant production systems expertise already exists as large scale growing, harvesting and processing of plant materials in the agricultural industry (Karg & Kallio, 2009). Protein production in plants is flexible as it can cheaply meet demand by scaling up, harvesting or planting the required number of plants (Teli & Timko, 2004). Plants have been for the past 10 years used to produce recombinant therapeutic proteins. The requirements for production of recombinant proteins from plants are less than fermantation systems due to already existing agricultural infrastructure for growing plants. Protein production is flexible and has been established to meet scale by demand (Karg & Kallio, 2009). Recent improvements include the advancement of expression vectors based on plant viruses that are able to expresss high yields of recombinant proteins of therapeutic value (Streatfield, 2007). Plant viruses are not zoonotic, they cannot infect mammals or humans that will be the intended hosts for the recombinant

proteins and plants can't be infected by human or mammalian viruses. Where deconstructed plant viruses are used to express a recombinant protein, the plant virus would not transfer to the intended host of use (Miele, 1997). Therefore, plant production systems do not harbour pathogens which are infective of mammalian or bacterial cell cultures – which are risks that exist for bacterial and mammalian cell cultures. Co-purification of potentially oncogenic sequences and blood-borne pathogens are additional risks of using mammalian and bacterial cell culture systems (Fischer & Emans, 2000).

Plants are preferred over prokaryotic expression systems in production of pharmaceutical proteins because they are capable of some post-translational modifications (Basaran & Rodríguez-Cerezo, 2008). Plants are able to carry out protein maturation events, such as signal peptide cleavage, disulphide bond formation, glycosylation and folding of complex recombinant antibodies, similarly to mammalian expression systems (Maliga and Graham, 2004). Protein maturation events are significantly important when produced recombinant proteins are intended to be bio-similars in human therapeutic use. Bacterial production systems do not possess post-translational machinery to process recombinant proteins, therefor they are unable to fold properly and lose their native function. Eukaryotic expression systems are preferred for recombinant proteins that require post-translational modification (Twyman *et al.*, 2003).

The demand for recombinant proteins in the biopharmaceutical industry grows the utilisation of fusion protein technology, for effective expression of proteins that would be difficult to express or purify. The limiting step with fusion protein technology is circumvented by the use of site-specific EK with its accommodative reaction conditions for better yields of high-value recombinant proteins (Basaran & Rodríguez-Cerezo, 2008). The production of EK, is met with several challenges that have escalated its price, which is the focal point of this research project (Gasparian *et al.,* 2011). This research project explores the use of plant production system to produce recombinant EK at more desirable amounts while retaining activity.

1.3.1 Molecular farming

Molecular farming, molecular pharming or biopharming is the expression of recombinant proteins in plants such as antibodies, blood coagulation proteins, vaccines, structural substances, antimicrobial and antiviral substances, growth hormones, enzymes and antibodies for industrial or biopharmaceutical applications (Bieri, 2011). Mammalian cell culture expression is an elusive and expensive process where the capital input for bioreactors is the most significant limiting step, whereas with transgenic plants or virus-infected plants there is existing infrastructure and plants require minerals, sunlight and water. The input requirements for plant production systems are far less than bioreactor requirements (Houdebine, 2009). Plants, like animals are complex and multicellular and therefor have more similar protein synthesis to animals than to those of yeast or bacteria, which may not be capable of producing those complex proteins with the required glycosylation patterns. Plant viruses have not been found to be infective to humans or mammals, thus plant production systems will not be able to transmit human or mammalian viruses (Frutos *et al.*, 2008). Recombinant protein purification of a target protein from plants is often easier than the labour- and cost intensive bacterial system. Plants can be engineered for multiple foreign genes to develop multicomponent vaccines. Plant expressed vaccines are more safe, versatile and stable than animal-derived vaccines. Potential transfer of animal viruses as by product of vaccine production in animal cells is eliminated by recombinant production in plants. There is greater control in expression of recombinant proteins in plants, higher accumulation of recombinant proteins can be obtained in targeted tissues. Plants, as eukaryotes are able to glycosylate recombinant proteins to elicit the immune responses where bacteria cannot, and yeast has been observed to over-glycosylate recombinant proteins (Frutos *et al.*, 2008*)*.

Horsch (2004) reported that there 3.5 fold cost difference to acquire 1 ton/year of a purified recombinant protein between a cGMP CHO site and a cGMP maize site. The CHO site would cost between USD 205 – 450 million at 1 g/L protein yield with 20 runs/year at 50 % recovery in 7 X 15000 L tanks, whereas in the maize site it would cost between USD 80 – 120 million at 1 kg/acre protein yield with 50 % recovery in a 200 – 2000 acre transgenic maize field. His report emphasised the significant cost

implications associated with mammalian systems and plants offer cost effective system at commercial scales (Horsch, 2004).

1.3.2 Examples of molecular farming

In the beginning plants produced 10-fold less recombinant proteins compared to bacterial and mammalian expression systems, where stably transformed plants took 18 months to produce. Developments in plant production systems have led to shorter recombinant protein production and higher yields i.e. up to 80 % total soluble protein (TSP) of GFP accumulation in *N. benthamiana* leaves by Icon Genetics (Thomas *et al.*, 2011). These developments were achieved by improving factors that affect overall recombinant protein yield in plants. The factors that affect protein yield are described in Section 1.3.3 (Thomas *et al.*, 2011).

1.3.3 Factors affecting production of plant-derived proteins

Synthesis of the gene of interest plays one of the factors in determining expression levels of recombinant proteins, where at nucleotide level optimisation of the codon usage is beneficial for genes derived from other organisms that will have a different codon bias to the intended expression host (Doran, 2006; Streatfield, 2007).

1.3.3.1 Codon optimization

Each organism has different codon usage frequencies and there is a general codon bias which is not the same between organisms. Plants and humans have different codon preferences, plants have a higher GC rich codon usage and mammals have a higher AT codon usage and there exists variations in codon usage among mammals or plants themselves (Gupta, 2003). Due to this difference in codon composition, optimisation of codon usage is required when expressing transgenes derived from other organisms (Gupta, 2003). Recombinant protein production more often than not, involves expression of a transgene in another production system and due to the codon bias that exists codon optimisation of the transgene is recommended. Codon optimisation is the adjustment or modification of the transgene codon preference toward the codon preference of the production system. In this process, nucleotides in

the codon are changed but the amino acid remains the same, where the low frequency codons are changed to higher frequency codons according to species frequency distribution table (Doran, 2006). Codon optimisation affects the production of recombinant proteins at nucleotide levels, to increase expression of the transgene (Gupta, 2003). In the process cis-acting elements such as tandem repeats, AU rich destabilising, splicing, stem-loops and putative signals are optimised out of the native sequence. The half-life of transgene mRNA is enhances when optimised codon are used (Doran, 2006 ; Streatfield 2007). The measure of codon bias is called codon adaptation index (CAI), which is a measure of the expression levels of a transgene in another organism. The removal of the negative elements in a transgene sequence varies the CAI value of the foreign gene (Laguía-Becher *et al*., 2010).

1.3.3.2 Choice of host

The success of expression of heterologous proteins in plants is attributed to many factors such as; codon optimisation of gene to be transcribed, expression system and target of recombinant protein. The physiology of plants are different, therefore the choice of host is a significant factor. Due to the differences in physiology, some plants will naturally express heterologous proteins better than others (Menkhaus *et al.*, 2004 ; Doran 2006). Soybean has a 40 % weight in protein compared to potato tubes, which has a 2 % protein weight; therefore if a recombinant protein would be expressed in the same total soluble protein in both crops there would be 20 X more protein accumulation in the soybean plant (Menkhaus *et al.*, 2004). The biomass of some crops may be attractive but other vital considerations include protein stability and ease of processing is important. *Arabidopsis thaliana* has been widely used as a model plant in recombinant protein production, as a crop of choice for proof-of-concept and scalability studies. However attractive *A. thaliana* is a model, its agronomic properties such as biomass or seed production make it impractical as a commercially relevant production host (Rybicki, 2010).

1.3.3.2.1 Non-food crop plants

Tobacco has been used as a model plant in recombinant protein production alongside *Arabidopsis thaliana*, for proof-of-concept and scalability studies (Rybicki, 2010). Cultivated tobacco known as *Nicotiana tabacum* is a herbaceous plant and it is the most commonly grown in the *Nicotiana* genus, because of the amount of biomass it produces. It grows up to heights between 1 to 1.5 meters. It is for this reason and the ease of genetic manipulation that *N. tabacum* has been used to produce large scale amounts as transgenic plants. Prior to performing large scale production of recombinant proteins; proof-of-concept expression is performed using *Arabidopsis thaliana* or *Nicotiana benthamiana*. *N. benthamiana* is preferred as it has higher biomass (Menkhaus *et al.,* 2007). The leaves of *N. benthamiana* are frail and can be injured in experiments in order to introduce transgenes for transformation via the vehicle organism *Agrobacterium tumefaciens*. *N. benthamiana* takes 6 - 8 weeks to maturity from seedling germination, thus allowing for continued transfection and recombinant protein production. It was reported that plant-produced recombinant toxin protein (LT-B toxin subunit) retained native functionality and orally vaccinated mice were able to neutralise the native toxin. Plant-made anti-LT-B protected orally-immunised animals against bacterial challenge (Rybicki, 2010).

1.3.3.2.2 Leafy crops

Edible leafy crops have been previously used for recombinant protein production; edibles such as lettuce, spinach, lupins and alfalfa. The same risks are shared as edible fruits and tubers with edible leafy crops because of the variation in expression of the therapeutic protein. The therapeutics expressed in leaves however, have to be processed immediately after maturity to prevent protein degradation which is not ideal because of therapeutics delivery amounts and schedule (Rybicki, 2010). Edible leafy hosts are not recommended for recombinant protein expression due to concerns of contamination into other food crops. The choice of host plant to use for recombinant expression is determined the physiology of the plant as well as the tissue/organelle expression is directed.

1.3.3.3 Protein targeting

Target organelle for expression of the heterologous proteins has been mentioned as one of the critical factors in achieving higher yields of protein expression (Schouten, 1996). Protein targeting is important as it determines the expression amounts as well as the stability of the expressed protein. There are various targets that proteins can be targeted to; cytosol, apoplast, chloroplast, endoplasmic reticulum (ER) or mitochondria. Schouten *et al*, 1996 observed a 200-fold increase in production of single chain antibody, and Menassa, 2001 also observed a 70-fold increase in expression of interleukin-10; both by the addition of an ER retention signal KDEL to the recombinant proteins (Conley *et al.*, 2010; Schouten *et al*., 1996). The use of animal amino-terminal signal peptides has been reported to increase the amount of recombinant protein accumulation than those same proteins targeted to the cytosol (Fischer *et al.*, 2004). The targeting of transgene expression to cellular compartments is beneficial to recombinant protein stability and therefore increased expression yields.

1.3.3.3.1 Chloroplast/plastid expression

Early work on plant production systems focused on nuclear transformation for expression of transgenes, meanwhile chloroplast expression or trans-plastomic expression has been able to achieve desirable protein yields. The high expression yields of chloroplast expression have been shown to have little effect on the phenotype of the plant, not as much as the changes brought about by nuclear transformation (Doran, 2006). Bacterial-derived recombinant proteins express easily in chloroplasts because they do not require post-translational processing such as glycosylation and the mechanisms for folding proteins. The highest level of expression in chloroplasts achieved was 3 g/kg of human papillomavirus (HPV)-16 L1 proteins in tobacco, which had the intended assembly and elicitation of neutralisation antibodies (Streatfield, 2007). This also highlighted that chloroplast expression could be used for expression of toxin agents, such as the *Bacillus anthracis* (anthrax) protective agent (PA) that may be toxic to other cell systems. Investigations of pathways in the chloroplasts have been identified which could make

recombinant proteins reach maturation (post-translational) that occurs in the ER and Golgi apparatus (Streatfield, 2007 ; Rybicki, 2010).

1.3.4 Plants as an expression system

Stable transformation and transient expression are two commonly used methods for recombinant protein production in plants (Rybicki, 2010). Stable transformation generates transgenic plants that have integration of the transgene to be expressed in the plant genome and is carried onto the further generations, whereas in transient expression the transgene is not integrated into the plant genome and not carried onto other generations (Streatfield, 2007).

1.3.4.1 Whole plant stable transformation

Transgenes can be stably integrated into the plant chromosome by micro-projectile bombardment, bombarding plant tissue with DNA-coated micro-projectiles that will integrate into the plant chromosome (Klein *et al.,* 1987). The most common method for stable transformation involves using the natural gene-transfer methods of *Agrobacterium tumefaciens* (Bally *et al.*, 2011). Regeneration of transgenic lines is followed by culturing the transformed calli and generation of transformed plants. Stable transformation of plants is an established technology applicable to a wide range of crops and expression vectors, most if the plant-derived recombinant proteins were produced using this technology (Horn *et al.*, 2004). There are however significant regulatory concerns in containment of transgenic plants, of which changes are inherent to new generations (Rybicki, 2010). Nuclear transformation occurs by random integration of transgenes into unpredictable location into the genome resulting in variable expression levels in transformation events and often gene silencing (Basaran and Rodriguez-Cerezo, 2008).

1.3.4.1.1 Cell culture system

The use of cell culture or tissue culture systems has received less commercialization efforts because the costs amount to almost the same as traditional bacterial and eukaryotic cell cultures systems, the significantly lowered yields show that there is a

need of technical expertise. The use of plant cell cultures was investigated for use due to high level of contamination in mammalian cell culture systems, even under good manufacturing practice (GMP) (Rybicki, 2010). In 2009 Genzyme reported contamination by a calicivirus in their CHO culture system that was producing commercial Gaugher disease therapeutic, Cerezyme®. That report lead to approval of rival Gaugher disease therapeutics to be made available due to the disease burden, even though the rival therapeutics were still in clinical trials. One of the competitor Gaugher diseases therapeutics made available was suspension-cultured carrot cells human glucocerebrosidase (rGCD) made by Israel-based Protalix (Rybicki, 2010). This showed the obvious advantage of plant cell cultures in that, they are not susceptible to mammalian cell culture systems contaminants.

1.3.4.2 Transient expression

Plant production systems started out with the use of transgenic plants used to produce proteins of interest, which was a lengthy process due to the length of generation of transgenic lines and the regulatory policies. The trend is slanted on using transient expression systems due to the convenience and turn-around time for production (Gleba *et al.*, 2007). Transient expression is limited to *Agrobacterium*-mediated somatic cell expression or virus-based expression. Viral based expression offers a more efficient expression of recombinant proteins. Once the virus vector has been introduced to the plant the virus is engineered to transcribe the transgene in high quantities. This allows for quick production of heterologous proteins without the lag time of transgenic lines regeneration but the engineered *Agrobacterium*, must be contained (Gleba *et al.,* 2007). The commonly used transient expression vectors used are described below.

1.3.4.2.1 Simple virus-based vectors

Two strategies exist for viral based expression, deconstructed virus and full virus approaches. Full virus approach uses gene insertion vectors, where the vector uses a fully functionally virus that retains all of its stability, infectivity and virulence to the host plant (Gleba *et al.,* 2007). The result is that the gene of interest will be expressed along with the wild-type viral genes. Deconstructed virus approach,

makes use of and engineered viral vector where a viral sequence has been exchanged with the gene of interest (Gleba *et al.,* 2007). Typically the coat protein is replaced, which in turn deactivates the cell-to-cell infectivity of the virus. Replacement of the coat protein therefore limits the scope of the engineered viral vector (Marillonnet, 2005). There are size limitations to the gene of interest's size in gene replacement vectors, there is a negative relation to the insert's size and vector stability. Larger insert sizes are lost from viral vectors when delivered by *in vitro*-generated ssRNA or cDNA inoculation (Gleba *et al.,* 2007). Isometric virions cannot accommodate larger transgene sizes, but rod-like plant viruses which can accommodate larger genomes are thus better for expression. There exists many viral vectors isolated from various plants such as; tobacco mosaic virus (TMV), cowpea mosaic comovirus (CPMV), cucumber mosaic viruses (CMV), potexviruses potato virus X (PVX), bean yellow dwarf mastrevirus (BeYDV) and the papaya mosaic virus (PapMV). One of the more famous TMV based vectors is the Geneware™ technology (Avesani *et al*., 2007; Rybicki, 2010). Typically simple virus based vectors achieve low expression yields due to re-engineering infectivity; which can be circumvented by delivering viral genomes using *Agrobacterium*.

1.3.4.2.2 Agroinfiltration vectors

Agrobacterium tumefaciens-mediated transient expression uses the infectivity *Agrobacterium tumefaciens* native infectivity to transfer transgenes into the plant (Rybicki, 2010). The Agrobacterium suspension is introduced into the plant via syringe or syringe, where the cultures infiltrate into the intercellular spaces in plant leaves i.e. Agroinfiltration, which results in the mobilisation of T-DNA and expression of the transgene occurs (Rybicki, 2010). Agroinfiltration is used to determine whether a transgene can be expressed or not and in which cellular compartment there is optimal expression. Agroinfiltration was used to express haemagglutinin (HA) from H5N1) and H1N1 (A/New Caledonia /20/99) in *N. benthamiana* leaves. The recombinant *HA* and H1N1 proteins had assembled correctly at yields of 50 mg/kg FW and the rHA was immunogenic (Rybicki, 2010).

1.3.4.2.3 *Agrobacterium*-mediated transfection

Agrobacterium tumefaciens is a soil-habiting bacterium used to transform the plant cell. The bacterium responds to the phenolic compounds released by a wounded plant; *Agrobacterium* moves towards the plant, attaches itself and infects the plant through the wound. It is through this wound that *Agrobacterium* induces transcription of its native virulence genes. The virulence genes are located on the tumor inducing (Ti) plasmid, which also contains a transfer DNA (T-DNA), it is the virulence genes that are responsible for host range, transcriptional activation, processsing and export of genes into the host for expression (Sala *et al.*, 2003). The T-DNA transferred by the *Agrobacterium* includes genes that stimulate growth of the gall tumour at the site of infectivity, enzymes responsible for opines which are used as a source of nitrogen which only *Agrobacterium* can metabolise and thrives in. The site of virulence gene integration into the host plant genome is where the bacterium transgenes are integrated into the plant genome (Sala *et al.*, 2003). *Agrobacterium*-mediated transfection uses the native infectivity of *Agrobacterium*, but the virulence genes have been engineered to introduce transgenes of interest into the plant genome i.e. disarmed *Agrobacterium*. Native transfection via *Agrobacterium* is described in Figure 1-8. Transfection initiates with the recognition and attachment of the bacterium to wounded cells (1), the plant signals activates the VirA/VirG virulence genes via signal transduction system (2). Transfer DNA is then mobilised by the VirD1/D2 protein complex (3 & 4) into the host cell cytoplasm along with other Vir proteins (5). A mature T-complex forms when VirE2-DNA associates with the T-strand, which then travels through the host cytoplasm (6) and imports into the cell nucleus (7). Once inside the nucleus the T-DNA is stripped of the chaperon Vir proteins and integrates into the host genome (8, 9 & 10) (Tzfira & Citovsky, 2006).

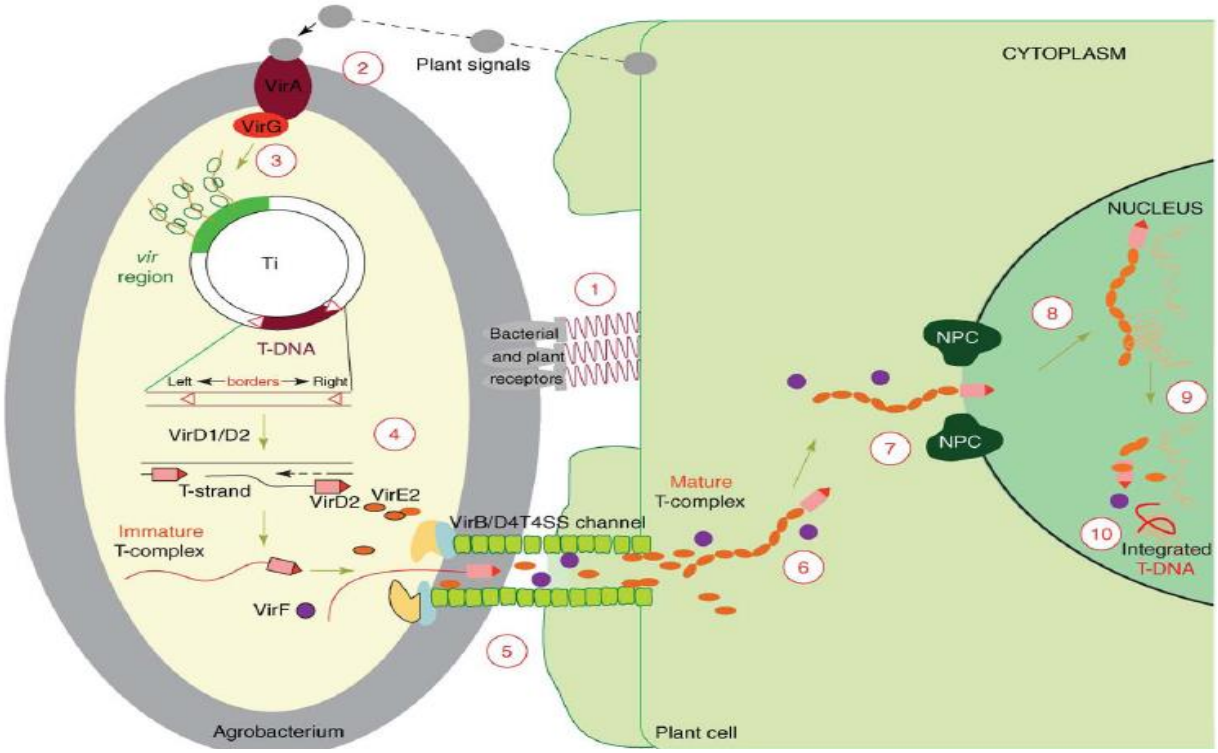


Figure 1-8: A model for *Agrobacterium*-mediated transfection. http://ars.els-cdn.com/content/image/1-s2.0-S0962892401022292-gr2.jpg. (Tzfira & Citovsky, 2006)

1.3.4.2.4 Agroinfiltrated viral vectors

Magnifection is a significantly important development in recombinant protein production as it combined agroinfiltration and the deconstructed TMV-based vector approach (Gleba *et al.,* 2005). In modular viral systems, viral components are split and inserted into *Agrobacterium* for co-inoculation; this reduces replicon size and increases vector stability. This modular system allows for other viral components to be retained and assemble *in trans* (Gleba *et al.*, 2007). Magnifection was developed by Icon Genetics as MagnICON® viral vectors that are delivered into tobacco as pro-vectors via *Agrobacterium* and assemble *in trans* for transgene expression. These vectors are limited to tobacco as a host, since they are based on a TMV virus, the host range can be circumvented by using the other plant viruses such as geminiviruses, nanoviruses and PVX vectors (Gleba *et al.,* 2005). Other improvements include the use of a silencing suppressor protein, p19 and an effective inducible In-plant Activation technology (INPACT) where transgenic expression

occurs via amplified replicon infectivity (Gleba *et al.,* 2005). This inducible system is particularly useful for expression of toxic molecules. MagnICON® viral vectors have been used to produce personalised single patient vaccines for non-Hodgkin lymphoma patients from 5 kg of *N. benthamiana*, rescuing the novelty of personalised drug-care and the use of plant production systems in addressing the demand for functional recombinant proteins in most industries (Rybicki, 2010).

In this report, MagnICON® viral vectors will be used for expression of recombinant enterokinase in the leaves of *Nicotiana benthamiana* via agroinfiltration using vacuum infiltration. Table 1-5 shows the documented expression yields of other recombinant proteins to show the feasibility of using the plant production systems coupled with the associated expression technology for optimal expression.

Table 1-5: Expression yield of recombinant molecules obtained by use of plant virus-mediated transient expression systems

Virus vector	Type of inoculation	Pharmaceutical protein	Highest yield
TMV	TMV-30B-based vector directed RNA transcript inoculation of spinach leaves	HIV-1 Tat protein (86 ama)	0.3 - 0.5 mg/g FW
TMV	TMV-30B-based vector directed RNA transcript inoculation of *Nicotiana benthamiana* leaves	HPV 16 L1 protein (531 ama)	0.03 x 10^{-3} mg/g FW
TMV	ToMV-TocJ-based vector directed RNA transcript inoculation of *N. benthamiana* leaves	The dengue virus envelope protein (102 ama fragment)	0.1 mg/g FW
TMV	*N. benthamiana* leaf agroinfiltration	*Mycobacterium tuberculosis* Ag85B and ESAT6 proteins (342 ama)	0.8 - 1.0 mg/g FW
TMV	TMV TTO1A vector directed RNA transcript inoculation of *N. benthamiana* leaves	Human scFv proteins (~ 30 kDa) derived from human tumor immunoglobulin genes of non-Hodgkin's lymphoma patients	100 - 800 µg/ml of leaf interstitial fluid
TMV	TMV launch vector pBID4 agroinfiltration of *N. benthamiana* leaves	hGH (~ 30 kDa)	0.7 mg/g FW
TMV and PVX	*N. benthamiana* leaf agroinfiltration with two populations of recombinant *Agrobacterium* encoding the mAb light chain and heavy chains	Anti-cancer full size mAb (~ 160 kDa)	100 - 300 µg/g FW affinity purified

TMV and PVX	MagnICON: *N. benthamiana* leaf agroinfiltration with two populations of recombinant A*grobacterium* encoding the mAb light chain and heavy chain	Anti-cancer full size mAb (~ 160 kDa)	0.3 - 0.5 mg/g FW affinity purified
TMV and PVX	MagnICON: *N. benthamiana* leaf agroinfiltration with two populations of recombinant *Agrobacterium* encoding the mAb light chain and heavy chain	Anti-West Nile virus mAb (~ 160 kDa)	0.8 mg/g FW leaves, affinity purified
TMV	MagnICON: *N. benthamiana* leaf agroinfiltration	*Yersinia pestis* F1 (262 ama) and V (150 ama) antigens	0.3 - 0.5 mg/g FW; 1.2 mg/g FW (purified product)
TMV	MagnICON: *N. benthamiana* leaf agroinfiltration	VV B5 (275 ama) antigenic domain	0.1 mg/g FW
TMV	MagnICON: *N. benthamiana* leaf agroinjection	*Plasmodium* antigen PyMSP4/5 (33 kDa)	1 - 2 mg/g FW
TMV	MagnICON: *N. benthamiana* leaf agroinfiltration	NV CP (~ 55 kDa)	0.8 mg/g FW
PVX	inoculation of *N. benthamiana* leaves with pPVX201-based cDNA construct	HBV nucleoplastid protein (21 kDa)	0.005 - 0.01 mg/g FW
PVX	inoculation of *N. benthamiana* leaves with pPVX201-based cDNA construct	*M. tuberculosis*ESAT6 antigen fusion with PVX CP (31 kDa)	0.5 - 1 % of the TSP
PVX and TMV	*N. benthamiana* leaf agroinfiltration	GFP fused with CAV VP1, VP2 and VP3 (40 - 51 kDa)	1.2 - 5.4 % of the TSP
BeYDV	*N. benthamiana* leaf agroinfiltration	HBc (21 kDa) and NV CP (58 kDa)	HBc: 0.8 mg/g FW NV CP: 0.34 mg/g FW

TMV =Tobacco mosaic virus , RNA = Ribonucleic acid, HIV = Human immunodeficiency virus, HPV = Human papilloma virus, FW = Fresh weight, scFv = Single-chain variable fragment, hGH = Human growth hormone, mAb = Monoclonal antibody, NV CP = Norwalk virus coat protein, HBV = , cDNA = Complementary DNA, PVX CP = Potato virus X coat protein, GFP = Green fluorescent protein, CAV VP1 VP2 VP3 = Chicken anemia virus virus particle 1, 2 & 3. Adapted from Komarova *et al*., 2010.

Chapter 2: Materials and methods

2.1 Source of general materials and methods

2.1.1 Chemical and non-chemical reagents

All reagents used in this study were of analytical grade and were obtained from SIGMA-ALDRICH (St. Louis, USA) unless otherwise stated. Restriction endonucleases, DNA Taq polymerase, T4 DNA ligase, DNA molecular markers and protein molecular markers were obtained from Fermentas (Ontario, Canada). Polyvinylidene fluoride (PVDF) membrane was obtained from Amersham Pharmacia Biotechnology (United Kingdom). Unless stated otherwise, all general DNA manipulations and basic molecular biology were as described by Sambrook and Fritsch, (1989). *Escherichia coli* DH10B was used as the general host for cloning and plasmid propagation. *Agrobacterium tumefaciens* GV3101::pMP90RK was used as the host to introduce transgenes into *Nicotiana benthamiana*. Enterokinase genes were synthesised at GenScript USA Inc. (Piscataway, NJ, USA) in two forms; plant codon optimised EK (PEK) and non-codon optimised EK (NEK). Both EK genes were synthesised to have a 3' HIS_6 tag for ease of purification.

2.1.2 Evolutionary relationship of EK

Nucleotide and protein sequences of EK were obtained from National Centre of Biotechnology Information (NCBI) and tabulated in Table 2-1. Alignments and comparisons of protein sequences were conducted with basic alignment search tool (BLAST) (Altschul *et al*., 1990) and CLUSTAL 2.1 using only EK catalytic subunits.

2.1.3 Plasmids

The plasmids used in this study are described and tabulated in Table 2-2. Schematics of the pro-vector T-DNA regions used in this study are provided in Figure 2-1.

Table 2-1: Sources of EK nucleotide and protein sequences showing description and references

Source	Description	Nucleotide accession number	Protein accession number
Human	Human enterokinase mRNA, complete cds	>gi|746412|	>gi|223942069|
Bovine	BOVEKCS *Bos taurus* enterokinase catalytic subunit mRNA, 3' end	>gi|416131|	>gi|27806097|
Porcine	*Sus scrofa* transmembrane protease, serine 15 (TMPRSS15), mRNA	>gi|47575833|	>gi|505123|
Rat	*Rattus norvegicus* transmembrane protease, serine 15 (Tmprss15), mRNA	>gi|157787049|	>gi|157787050|
Mouse	*Mus musculus* transmembrane protease, serine 15 (Tmprss15), transcript variant 1, mRNA	>gi|108936963|	>gi|1698878|

Table 2-2: Description and references of plasmids and constructs used in this study

Name	Description	Reference
pUC57	pUC19 derivative cloning vector, ampicillin resistant	Markausakas, A.; Dreguniene, G, 1997
pICH11599	MagnICON cloning pro-vector, ampicillin resistant	Marillonnet, 2005
pICH17388	MagnICON pro-vector that directs expression to the cytosol, ampicillin resistant	Marillonnet, 2005
pICH17620	MagnICON pro-vector that directs expression to the apoplast, ampicillin resistant	Marillonnet, 2005
pICH7410	MagnICON pro-vector with GFP incorporated, ampicillin resistant	Marillonnet, 2005
pICH14011	MagnICON pro-vector with the PhiC31 integrase recombination site, ampicillin resistant	Marillonnet, 2005
pGEM-T easy®	Cloning vector that possess thymidines at the 3´-ends to aid cloning of PCR products, ampicillin resistant	Promega, Madison, Wisconsin, 1992
pGEM-PEK	pGEM-T easy vector with plant codon optimised EK, ampicillin resistant	This study
pGEM-NEK	pGEM-T easy vector with non-codon optimised EK, ampicillin	This study

	resistant	
pICH-PEK	pICH11599 containing plant codon optimised EK, ampicillin resistant	This study
pICH-NEK	pICH11599 containing plant codon optimised EK, ampicillin resistant	This study

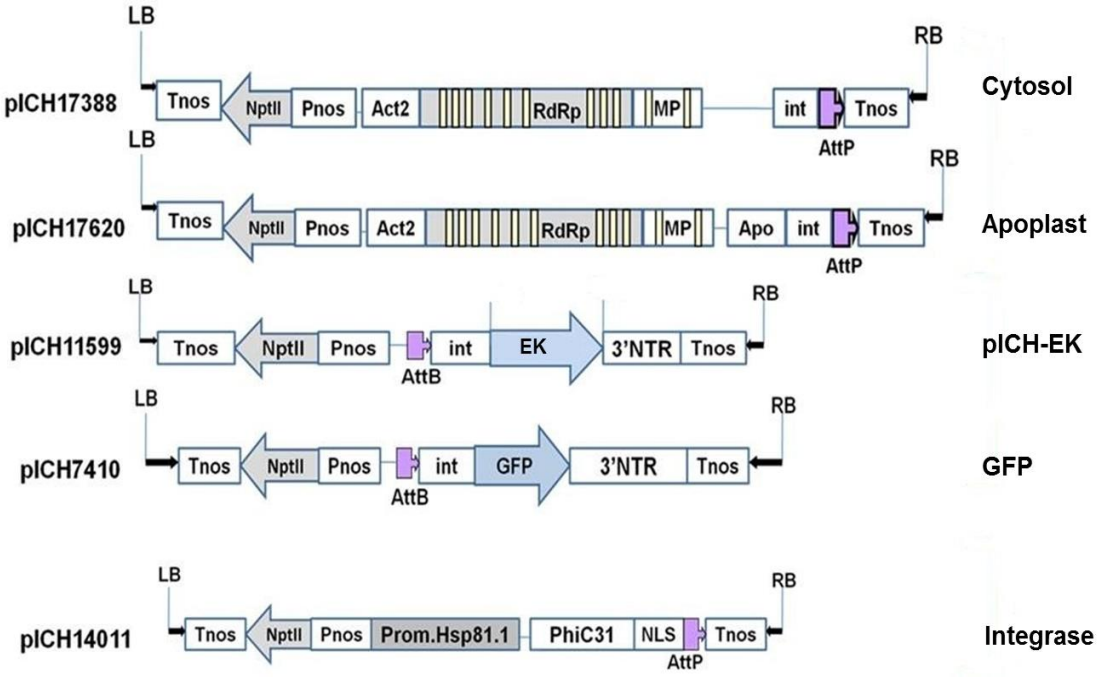


Figure 2-1: Schematic representation of the T-DNA regions of the magnICON pro-vector system used and generated in this study. A, Tobacco mosaic virus (TMV)-based provectors (5' modules -pICH17388 and pICH17620); B, 3' modules - cloning vector pICH11599 and green fluorescent protein (GFP) pICH7410. C, *Streptomyces* phage PhiC31 integrase module pICH14011 which mediates recombinational joining of the pro-vectors. LB and RB, binary left and right borders of the T-DNA region, respectively; RNA-dependent RNA Polymerase (RdRP), which catalyses the replication of RNA from a RNA template; Prom.Hsp81.1, *Arabidopsis* heat shock protein 81.1 promoter; Tnos and Pnos, nopaline synthase terminator and promoter, respectively; MP, movement protein; Act2, *Arabidopsis* actin 2 promoter; AttP and AttB, PhiC31 integrase recombination sites; int, intron; Apo, Calreticulin apoplast targeting signal; NLS, nuclear localization signal; nptII, nptII gene; 3'-NTR, 3' non-translated region; sLT-B is a plant-optimised synthetic gene encoding the *E. coli* heat labile enterotoxin B subunit (LT-B) (Giritch *et al*., 2006 & Mason *et al.,* 1998;)

2.1.4 Computer analysis

Primers were designed (tabulated in Table 2-3) and checked for hairpins and dimerization using oligo analyser 1.0.3 (Teemu Kuulasmaa). FinchTV (Geospiza)

was used to analyse the nucleotide sequencing data of PEK and NEK. Vector NTI® program was used for sequence analysis and alignments (Hall, 1999); and BLAST (Altschul *et al*., 1990) was used for database searches. Statistical analysis was carried out using Microsoft® Office Excel. Protein quantification and ELISA data was acquired using Bio-Tek Data Analysis Software v2.4 (Vermont, USA).

2.2 Microbiological techniques

2.2.1 Bacterial strains

The *Escherichia coli* and *Agrobacterium tumefaciens* strains used in this study are provided in Table 2-4, tabulated and described below.

Table 2-3: Primers used to amplify enterokinase (PEK and NEK) genes, synthesised by Integrated DNA technologies (IDT)

Primer name	Primer sequence (5' to 3 ')	Melting temperature (T_m)
PEK-F	ctcccatgggcattgtgggcggttcag	67.1 ^{o}C
PEK-R	gacggatccctactaggcggccgcgtggtgatgatg	70.6 ^{o}C
NEK-F	ctcccatgggaattgtcggaggaagtg	62.5 ^{o}C
NEK-R	gacggatccctactaggcggccgcgtggtgatggtgatgatg	71.5 ^{o}C

Table 2-4: Table of bacterial strains used in this study

Name	Full genotype and reference
E. coli DH10B	F^{-} endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ (Grant *et al.*, 1990)
E. coli BL21-Gold	F- *ompT hsdS(*rB- mB-) *dcm*+ Tetr *gal endA* The (Weiner *et al*. 1994)
Agrobacterium tumefaciens GV3101	MP90RK (pTiC58ΔT-DNA) (Koncz and Schell, 1986)

2.2.1.1 Preparation of electro-competent *E. coli* cells

One millilitre of overnight culture of *E. coli* DH10B was cultured in 100 ml of Luria broth (LB) medium and incubated at 37 ^{o}C in a shake incubator with vigorous shaking to an OD_{600} of 0.5. The culture was then chilled for 30 min on ice, then cells were harvested by centrifuging for 15 min at 4 000 g, at 4 °C. The pellet was washed

with 50 ml ice cold 10 % glycerol and then centrifuged and that pellet washed with 25 ml 10 % glycerol. That pellet was then re-suspended in ice cold GYT medium to a final volume of 200 µl. Aliquots of 50 µl of the competent cells was made and stored at -80 °C.

2.2.1.2 Preparation of *Agrobacterium tumefaciens* electro-competent cells

One millilitre of overnight culture of *Agrobacterium tumefaciens* GV3101::pMP90RK was cultured in 500 ml of Luria broth (LB) medium (rifampicin 25 mg/L and kanamycin 50 mg/L) and incubated at 28 ^{o}C in a shake incubator with vigorous shaking to an OD_{600} of 0.5. The culture was then chilled for 30 min on ice, then cells were harvested by centrifuging for 15 min at 4 000 g at 4 °C. The cells were kept on ice for subsequent steps. The pellet was washed with 50 ml ice cold 1 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, catalogue # H3375, SIGMA) pH 7.4 and centrifuged as above. The pellet was re-suspended in 250 ml 1 mM HEPES and centrifuged. The pellet was re-suspended in 10 ml 1 mM HEPES and centrifuged. The pellet was then re-suspended in 2 ml ice cold 10 % glycerol, aliquots of 50 µl of the competent cells was made and stored at -80 °C.

2.2.1.3 *Escherichia coli* transformation

The constructs pGEM-EK and pICH-EK (50 ng each) were used for transformation of *E. coli* DH10B. Constructs were added to ice-thawed 50 µl electro-competent *E. coli* DH10B and incubated on ice for 20 minutes. The *E. coli* was then transferred to pre-chilled 1 mm electroporation cuvettes (catalogue # 165-2093, BioRad), thereafter applying a voltage of 1.6 kV using the BioRad Genepulser apparatus (25 µF, 200 Ω; catalogue # 165-2660). The cuvettes were immediately transferred to ice without shaking. After 2 minutes SOC medium at room temperature was added to the cells to help recovery from electroporation. The cells in SOC medium were transferred to a McCartney bottle and shake incubated at 37 ^{o}C for 1 hour. Thereafter the transformed cells were plated onto YEP plates containing 50 mg/L carbenicillin at 37 ^{o}C overnight. Single colonies from overnight incubation were picked and inoculated into YEP (1 colony to 5 ml of YEP medium + 50 mg/L carbenicillin) and shake incubated at 37 ^{o}C overnight. Plasmid extraction followed as per protocol of Zyppy

mini-prep kit (Zyppy Plasmid Miniprep Kit catalogue # D4019 – Zymo Research Corp).

2.2.1.4 *Agrobacterium tumefaciens* transformation

The EK constructs, pICH-PEK and pICH-NEK (50 ng each) were used to transform *Agrobacterium tumefaciens*. Constructs were added to ice-thawed 50 µl electro-competent *Agrobacterium* and incubated, on ice for 20 minutes. The *Agrobacterium* was then transferred to pre-chilled 1 mm electroporation cuvettes, thereafter a pulse of 1.44 kV was applied using the BioRad Genepulser apparatus (25 µF, 200 Ω). The cuvettes were immediately transferred back to ice without shaking. After 2 minutes 950 µl LB medium at room temperature was added to the cells to help recovery from electroporation. The cells in LB medium were transferred to a McCartney bottle and shake incubated at 28 ^{o}C for 3 hours. Thereafter the transformed cells were plated onto LA plates (25 mg/L rifampicin, 50 mg/L kanamycin and 50 mg/L carbenicillin) at 28 ^{o}C for 36 hours. The transformed colonies that grew were inoculated into LA media (25 mg/L rifampicin, 50 mg/L kanamycin and 50 mg/L carbenicillin) at 28 ^{o}C overnight in shake incubation.

2.2.2 DNA preparation and analysis

2.2.2.1 Nucleic Acid manipulation

Unless otherwise stated, all nucleic acid manipulations and cloning were as described by Sambrook *et al.* (1989). These include small-scale DNA plasmid preparations, restriction enzyme digestion, and ligations. Restriction enzyme digestion was carried out to isolate EK genes from pUC57-EK, digest EK PCR products and verify pICH-EK constructs. Five hundred nanograms of DNA was digested with 1 U *Nco*I and *BamH*I both in a 30 µl reaction, using 1 X buffer Tango. The digestion was performed at 37 ^{o}C for 1 hour and visualised in agarose gel. Ligation was performed with 50 ng linear vector DNA (pGEM-T easy/pICH11599) and 150 ng insert DNA (EK) with 1U T4 DNA ligase in rapid ligation buffer. Ligation was performed overnight at 4 ^{o}C for production of maximum number of transformants.

2.2.2.2 Preparation of plasmid DNA

Zyppy mini-prep kit (Zyppy Plasmid Miniprep Kit catalogue # D4019 – Zymo Research Corp, CA USA) was used for plasmid DNA preparations according to manufacturer's protocol.

2.2.2.3 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was carried out using Fermentas Long PCR Enzyme Mix (catalogue # K0181). The reactions were carried out in 50 µl volumes containing; 50 mM $MgCl_2$, 1 µM forward/reverse primer, 0.2 µM dNTPs, 5 ng of DNA template and 2.5 units Taq polymerase. Reaction conditions were carried out as follows; 1 cycle at 95 ºC for 30 s, 94 ^{o}C for 30 s, 58 ºC for 30 s, 72 ^{o}C for 45 s; then repeated for 30 cycles. Final extension was at 72 ^{o}C for 4 minutes.

2.2.2.4 Agarose gel electrophoresis and isolation of DNA fragments

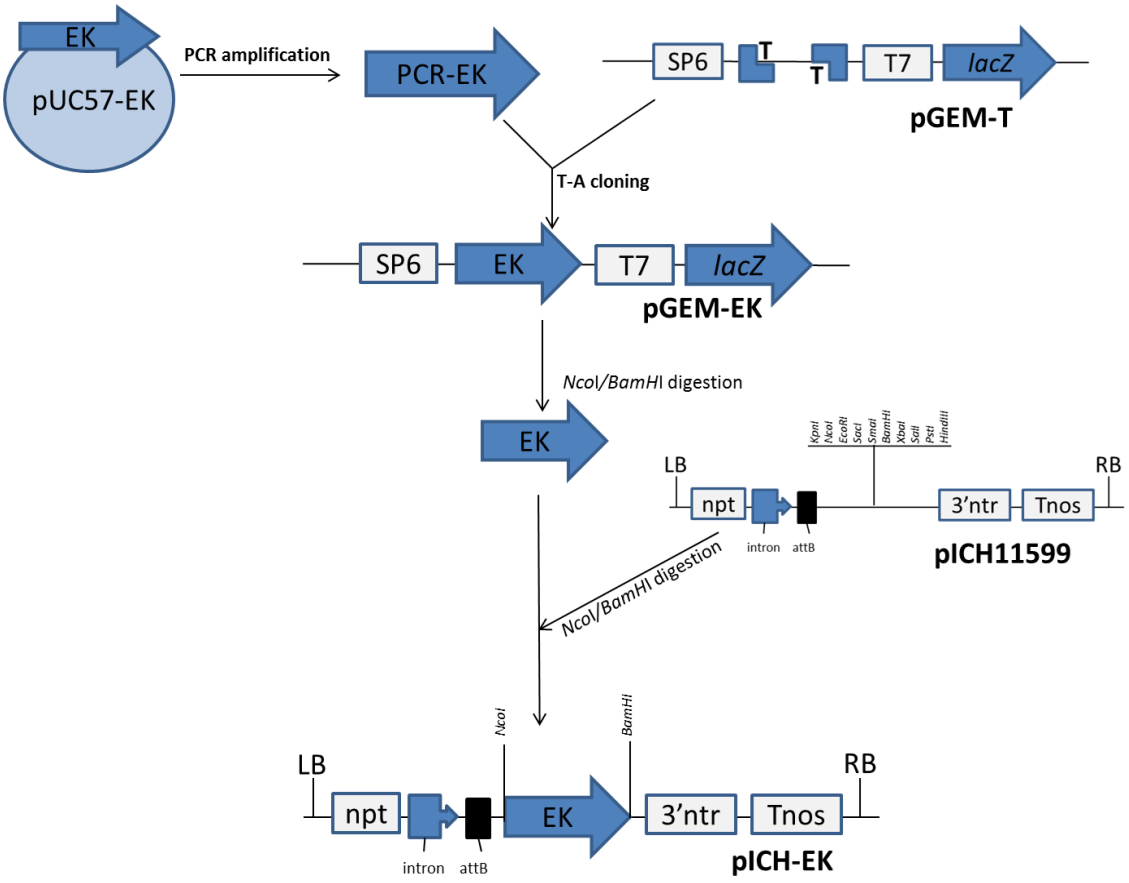
Polymerase chain reaction product or restriction enzyme digestion products were mixed with DNA loading buffer (0.25 % w/v bromophenol blue, 0.25 % w/v xylene cyanol FF, 30 % v/v glycerol) and then subjected to electrophoresis using 1 % agarose gel containing 1 µg/ml ethidium bromide. Gels were run in 1 x Tris-acetic acid-EDTA (TAE) buffer and photographed on a UV light transilluminator. Band sizes were estimated with 1 kb DNA ladder (catalogue # SM0331, Fermentas). DNA for further cloning was examined under longwave UV (254 nm), minimising the exposure to UV to prevent DNA nicking. Agarose gels with the desired fragment were excised and extracted from the agarose using QiaQuick gel extraction kit, QIAGEN catalogue number # 28704 (Venlo, Netherlands).

2.2.2.5 Sequencing of EK PCR products

Enterokinase amplicons were cloned into pGEM-T easy vector via T-A cloning. The EK amplicons were ligated into pGEM-T easy® vector to enable nucleotide sequencing. The pGEM-EK constructs were sent to Inqaba biotec (Pretoria, South

Africa). Nucleic acid sequences of PEK and NEK was determined by Sanger sequencing (Sanger, Coulson 1975).

A



B

NcoI

ccatggga attgtcggaggaagtgactccagagaaggagcctggccttgggtcgttgctctgtatttcgacgatcaac
aggtctgcggagcttctctggtgagcagggattggctggtgtcggccgcccactgcgtgtacgggagaaatatggag
ccgtctaagtggaaagcagtgctaggcctgcatatggcatcaaatctgacttctcctcagatagaaactaggttgattg
accaaattgtcataaacccacactacaataaacggagaaagaacaatgacattgccatgatgcatcttgaaatgaa
agtgaactacacagattatatacagcctatttgtttaccagaagaaaatcaagtttttcccccaggaagaatttgttctatt
gctggctgggggcacttatatatcaaggttctactgcagacgtactgcaagaagctgacgttccccttctatcaaatga
gaaatgtcaacaacagatgccagaatataacattacggaaaatatggtgtgtgcaggctatgaagcaggaggggta
gattcttgtcagggggattcaggcggaccactcatgtgccaagaaaacaacagatggctcctggctggcgtgacatc
atttggatatcaatgtgcactgcctaatcgcccaggggtgtatgcccgggtgccaaggttcacagagtggatacaaag
ttttctacat catcatcaccatcaccac gcggccgcc tagtag ggatcc

HIS_6 ***NotI*** ***BamHI***

Stop codon

Figure 2-2: Cloning strategy and annotated nucleotide sequence of EK. A = PCR-EK = EK PCR amplicons; SP6 = SP6 RNA polymerase; T7 = refactoring bacteriophage T7; lacZ = β-galactosidase gene; npt = neomycin phosphotransferase; attB = PhiC31 integrase recombination site; 3' ntr = 3' non-translated region ; Tnos = nopaline synthase terminator ; LB and RB = binary left and right borders respectively; EK = enterokinase. B= Indicated are the restriction

enzymes, HIS_6, stop codon, *Not*I restriction enzyme site and 3' *BamH*I restriction enzyme site.

2.2.3 Preparation of *Nicotiana benthamiana*

2.2.3.1 General plant husbandry

Nicotiana benthamiana (Kingdom: Plantae; order: *Solanales*; Family: *Solanaceae*; subfamily: *Nicotianoideae*; Genus: *Nicotiana*, species; *N. benthamiana*) seeds were obtained from University of Pretoria (Forestry and Agricultural Biotechnology Institute, South Africa). The seeds were sterilised using 70 % ethanol, then the ethanol was discarded and the seeds were soaked for 3 seconds in 100% bleach. The bleach was then discarded and the seeds were rinsed three times with distilled water. Sterile seeds were transferred into sterile bottles containing ½ MS macronutrients, 5 ml/l MS micronutrients, 30 g/L sucrose, 8 g/L agar, pH 5.8) medium. The seeds were grown in the tissue culture laboratory under a 16-h/8-hour photoperiod and a temperature regimen of 25 °C day/20 °C night. Three weeks after seeding, individual plantlets were picked out, transplanted in pots containing soil and grown in the tissue culture laboratory for 3-5 additional weeks under the same environmental conditions.

2.2.3.2 Transfection of *Nicotiana benthamiana*

Transfection of *Nicotiana benthamiana* with EK constructs was performed with the MagnICON vector modules; consisting of 3’ GFP and integrase pro-vectors as well as 5’ apoplast and 5’ cytosol targeting pro-vectors. The 3’ GFP pro-vector (pICH7410) module is used as an experimental control, to verify that the tri-partite pro-vector constructs assembled *in planta*. *Agrobacterium tumefaciens* transformed with pICH11599 was used as a negative control. The cytosol and apoplast targeting pro-vectors were used to target the gene of interest for expression to the organelle of interest. The 5’ module (pICH17388) contains the 5’ part of the viral vector including the RNA-dependent RNA polymerase, movement protein genes, the coat protein sub-genomic promoter, and a loxP site. The 3’ module (pICH11599) contains a loxP site, cloning sites for cloning of the gene of interest and the 3’ end of the viral vector. Both modules are assembled inside a plant cell with the help of a site-specific

recombinase (pICH14011) to form a fully functional RNA replicon. *Agrobacterium* was used to deliver various modules into plant cells. Glycerol stocks of MagnICON pro-vectors were inoculated into LB medium (mg/L rifampicin 25, 50 mg/L kanamycin and 50 mg/L carbenicillin) and shake incubated for 48 hours at 28 ^{o}C. The inoculum was inoculated into 500 ml LB medium and shaken at 150 rpm overnight at 28 ^{o}C. The overnight *A. tumefaciens* cultures with the MagnICON pro-vectors were sedimented at 2 500 rpm for 5 minutes at 4 ^{o}C and each culture re-suspended in infiltration media (10 mM $MgSO_4$, 10mM 2-[N-morpholinoethanesulfonic acid (MES, catalogue # M2933) hydrate pH 5.5). The cultures were mixed in equal tripartite proportions.

The infiltration constructs were diluted to D_{600} = 0.01 with infiltration media as depicted in Figure 2-3. One infiltration construct in a beaker was placed in a vacuum chamber where the aerial parts of the *Nicotiana benthamiana* plant were dipped. A vacuum was applied for 1 minute at a pressure of 24 – 25 mbar, thereafter the pressure was quickly released. The release of the vacuum facilitated the introduction of the constructs into the intracellular spaces of the leaves. The infiltrated plants were returned to the growth rooms under standard conditions (16 hour-light/8 hour-dark cycle, 55 – 65 % humidity). The same process was repeated for each tripartite solution for pICH11599, GFP expression, cytosol targeted expression of PEK and NEK and apoplast targeted expression of PEK and NEK (Marillonnet *et al.*, 2005).

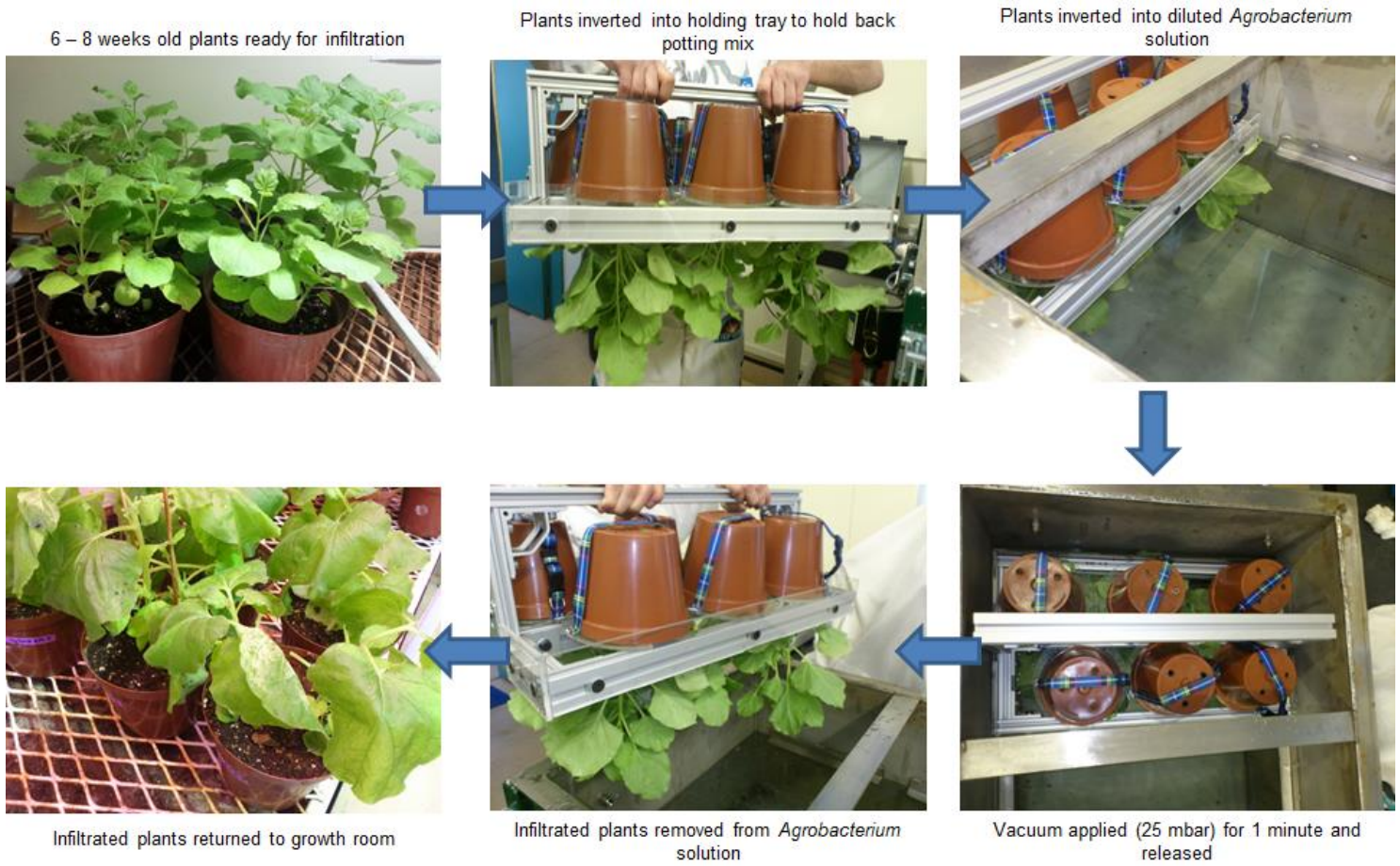


Figure 2-3: Agroinfiltration of *N. benthamiana* via vacuum infiltration. *Nicotiana benthamiana* plants were grown to 6-8 weeks in pots. The plants are then inverted into a holding tray with lids designed to hold back the potting mix, the holding tray loaded with plants was placed into transformed *Agrobacterium* infiltrated media. The chamber was closed and a vacuum of 25 mbar for 1 minute and the vacuum was then released. The infiltrated plants were removed from the chamber with the holding tray and placed back in the growth room. CSIR Biosciences.

2.2.3.3 Harvesting of plant material

Leaf material of the infiltrated plants was harvested at various time points, harvesting was done at 3-day intervals; 3, 6, 9, 12 days post infiltration (d.p.i.). Plants infiltrated with the green fluorescent protein (GFP) construct were illuminated under UV light to visualise fluorescence at each 3 day interval and leaf material collected. Leaf material from other infiltration constructs was collected, weighed and stored at -80 ^{o}C until protein extraction. Protein extraction of the harvested leaves was performed using phosphate buffered saline with tween20 (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4; PBST 0.02 % tween 20). Liquid nitrogen was used to grind the leaf material using a mortar and pestle. Extraction buffer was added in a 3:1 dilution ratio (ml:mg) of extraction buffer to leaf material weight. The leaf debris was sedimented using Eppendorf centrifuge 5417R (catalogue # 022621821, Eppendorf AG (Hamburg, Germany) at 10 000 rpm for 10 minutes at 4 ^{o}C and the supernatant was stored to be used for protein analysis.

2.3 Protein analysis

2.3.1 Protein quantification

Total soluble protein that was extracted from infiltrated *Nicotiana benthamiana* leaves was measured using Bradford analysis (Bradford, 1976). Bovine serum albumin (BSA, catalogue # 500-0206, BIORAD) was used at known concentrations to construct standard curve. Bradford reagent (catalogue # 500-0205, BIORAD) was used to determine total soluble protein concentration, using a 96 well microtitre plate (catalogue # 442404 NUNC, Roskilde, Denmark) in accordance with manufacturer's instructions. Protein concentration was determined by measuring absorbance at 595 nm using KC4 system (Analytical & Diagnostic Products, South Africa).

2.3.2 SDS-PAGE analysis

Protein samples (10 µg) were mixed with Laemmli loading dye (0.125 M Tris, 4 % w/v SDS, 10 % w/v 2-mercaptoehtanol, 20 % v/v glycerol, 0.004 % w/v bromophenol blue, pH 6.8) and heated in a water bath at 95 ^{o}C for 5 minutes. The samples were then separated by 12 % sodium dodecyl sulphate-polyacrylamide

electrophoresis (SDS-PAGE) using 1.5 mm acrylamide gels (Laemmli, 1970). Samples were separated via electrophoresis in SDS buffer (25 mM Tris, 192 mM glycine, 0.1 % w/v SDS, pH 8.3), using BIORAD Mini-PROTEAN Tetra cell unit as per manufacturer instructions (BIORAD, Hertz, UK). Electrophoresis was carried out at 120 V until the loading dye had run out of the cast gel sandwich. The gels were removed from the casting gel sandwich and stained as follows; staining solution I (10 % (v/v) acetic acid, 10 % (v/v) isopropanol & 0.125 % (w/v) coomasie brilliant blue G250) and staining solution II (10 % (v/v) acetic acid, 10% (v/v) isopropanol & 0.003 % (w/v) coomasie brilliant blue G250). The gels were de-stained with destaining solution (10 % (v/v) acetic acid, 5% (v/v) isopropanol) until the blue background had discoloured and protein bands were visible.

2.3.3 Protein band sequencing

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel was used to separate protein samples by molecular weight. The SDS-PAGE gel was stained and de-stained; bands of interest were excised and sent for sequencing to determine their amino acid sequence. Amino acids were sequenced by Liquid chromatography Matrix-assisted laser desorption/ ionization (LC MALDI) and sequences identified and compared at CSIR Biosciences, Molecular and Biomedical Technologies (MBT) group (Pretoria, South Africa).

2.3.4 Western blot analysis

Western blotting was used to detect plant-made EK using cross-reactive anti-EK antibody. Extracted proteins were separated by SDS-PAGE and probed on a membrane. The anti-EK western blot was performed as per Genscript®'s instructions (http://www.genscript.com/document/tech_guide/TM0590.pdf). After separation of 10 µg proteins via SDS-PAGE the gel was fixed in transfer buffer (25 mM Tris, 191 mM glycine, pH 8.3) while the polyvinylidene fluoride (PVDF) membrane was equilibrated in methanol and soaked in transfer buffer until transfer. Transfer of proteins onto PVDF membrane was carried out using trans-blot semi-dry electrophoretic transfer cell (catalogue # 170-3940, BIORAD) at 12 V, 200 mA for 2 hours. For immunodetection, proteins were transferred onto a PVDF membrane,

blocked with blocking buffer (PBST, 5 % fat-free milk, 0.05 % tween 20) overnight at 4 ^{o}C with agitation and then probed with primary antibody (anti-EK mAb, catalogue # A00889-200, Genscript) diluted in dilution buffer (PBST, 2.5 % fat-free milk, 0.05 % tween 20). This was followed by a secondary antibody incubation for 30 minutes with agitation (anti-mouse Ab - HRP, catalogue # A9917, SIGMA) diluted 1:10 000 in dilution buffer and subsequently detected with 3,3,5,5-tetramethylbenzidine (TMB) liquid substrate (catalogue # T 0565, SIGMA).

2.3.5 Protein quantification by ELISA

An indirect enzyme-linked immunosorbent assay (ELISA) assay was used to detect and quantify recombinant EK as described by Genscript®. (http://www.genscript.com/document/tech_guide/TM0589.pdf). Standard curve was constructed using commercial EK light chain (catalogue # P8070L, NEB), diluted in PBS as well as samples were incubated in 96 well microtitre plate (NUNC, A 62219) at 100 µl each for 2 hours at 37 ^{o}C (Themostar shake incubator, BMG LABTECH). The plate was washed 4 times with 200 µl of washing solution (PBST, 0.75 % tween 20) and blocked with blocking buffer (PBST, 0.5 % tween 20, 2.5 % fat-free milk powder) and incubated at 4 ^{o}C overnight with agitation. Primary antibody (anti-EK mAb, catalogue # A00889-200, Genscript) at 500 ng/ml was added to each well of the plate and incubated for 1 hour at 37 ^{o}C. The secondary antibody (anti-mouse IgG HRP Ab, catalogue # A9917, SIGMA) was diluted (1: 10 000) with blocking buffer, added to each well of the plate and incubated for 30 minutes at 37 ^{o}C. The plate was the washed 4 times with 200 µl of washing solution. Detection was made with distribution of 100 µl of 3,3,5,5-tetramethylbenzidine (TMB) liquid substrate (catalogue # T0440, SIGMA) into each well and plate was left for sufficient colour development. The reaction was stopped by addition of 1 M H_2SO_4. The optical density was measured at the range of 405 nm using Bio-Tek KC4 system (Analytical & Diagnostic Products, South Africa). Statistical analysis of the data was conducted with a student's t-test.

2.3.6 Protein purification by Protino Ni-TED

Plant-made EK purification was performed as per manufacturer's instructions. The Protino Ni-TED (tris-carboxymethyl ethylene diamine) packed columns (Macherey-nagel, catalogue # 745 150.50) were equilibrated with 320 µl 1 X Lysis-Equilibration-Wash (LEW) buffer (50 mM NaH_2PO_4, 300 mM NaCl, pH 8.0). The columns were allowed to drain by gravity. The prepared crude samples (extracted using 8M UREA, 3:1 ratio of extraction buffer to leaf material weight) were added to the pre-equilibrated columns and the columns were allowed to drain by gravity. The columns were washed with 320 µl 1 X LEW buffer. The washing step was repeated. The polyhistidine-tagged proteins were eluted in a collecting tube by adding 240 µl 1x Elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The collected samples, flow through, washes and elutes were collected and separated on 12 % SDS-PAGE.

2.3.6.1 Protein concentration by centrifugation

Protein concentration was conducted using the Macrosep centrifugal devices (PALL Life Sciences). According to the manufacturer's instructions, about 5 – 15 ml of sample was poured into the non-membrane side of the sample reservoir of Macrosep device (lot. 10550261) with a 30 KDa molecular weight cutoff. The reservoir was closed with a cap and centrifuged at 5000 rpm for 90 minutes or more to achieve desired concentrate volume using Eppendorf centrifuge 5810R. The retentate was collected and used for further analysis as it contained the EK protein.

2.3.6.2 Protein concentration by trichloroacetic acid (TCA) method

This is a modified method adapted from (Wang, Wu & Chen, 2006). One volume of trichloroacetic acid (TCA, 100 % w/v) stock solution was added to four volumes of protein samples, which was incubated for 10 minutes at 4 ^{o}C. The solution was subsequently centrifuged at 15000 rpm at 4 ^{o}C for 10 minutes using Eppendorf centrifuge 5417R. The supernatant was discarded and the precipitate was washed with 150 µl of 10% TCA once and then 200 µl acetone once and dried on a heating block at 95 ^{o}C. The dried pellet was re-suspended in SDS-PAGE loading buffer.

2.3.7 MagReSyn NTA protein purification

The EK purification was performed as per manufacturer's instructions using the MagReSyn microsphere (MagReSyn™ NTA, ReSyn Biosciences, CSIR Biosciences, Pretoria, catalogue # MR-NTA002). The MagReSyn microspheres were equilibrated, vortexed and washed three times in an eppendorf tube with 200 µl binding buffer (80 mM sodium phosphate pH 8.0; 40 mM imidazole and 0.5 M NaCl). Removal of the binding buffer was facilitated with the use of a magnet to attract the microspheres to the lining of the eppendorf and then the binding buffer was removed using a pipette. Binding buffer was discarded after the last wash, where protein extract was added to the washed microsphere. Binding buffer was added to a final volume of 300 µl to allow for binding of the HIS_6-tagged protein (extracted using PBST, 0.02 % tween 20, 3:1 ratio of extraction buffer to leaf material weight) to the microsphere for 5 minutes at room temperature. The binding buffer was removed and the microspheres washed with binding buffer 3 times. Binding buffer was discarded after the last wash, 20 µl of elution buffer (80 mM sodium phosphate pH 8.0, 500 mM NaCl and 500 mM imidazole) was added to elute the HIS_6-tagged proteins for 2 minutes at room temperature. Elution was repeated 3 times with elution buffer. The collected samples, flow through, washes and elutes were collected and separated on 12 % SDS-PAGE.

2.3.8 Activity assay of recombinant EK

Kinetic activity of EK was determined by monitoring the increase of 3-carboxy-4-nitrophenoxide yield by cleavage of Z-Lys-SBZL (N-α-Cbz-L-lysine thiobenzyl ester hydrochloride, catalogue # C3647, SIGMA) by EK at absorbance of 402 nm, for 5 minutes in a 96 well clear plate as described by (Lu *et al.*, 1997).

A kinetic assay was performed using Z-Lys-SBZL as a substrate, 5,5′-Dithiobis(2-nitrobenzoic acid, DTNB, catalogue # D8130, SIGMA) as a chromogen and commercial bovine EK light chain as a positive control (catalogue # P8070L, NEB). Enterokinase was diluted up to 40 ng/ml in assay buffer (50 mM Tris-HCl, pH 7.5). The serial dilutions of the substrate were constructed in the range of 0 – 500 µM in 50 µM increments in assay buffer to determine the range of activity detectable by the

assay, nitrobenzoic acid was added to each substrate concentration to 200 µM. Fifty microliters of EK was loaded in duplicate into appropriate wells in the microtitre plate, along with the relevant assay controls. Fifty microliters of substrate/DTNB solution was added to appropriate wells in the microtitre plate. Kinetic readings were performed by a standard colorimetric kinetic mode measuring the release of 3-carboxy-4-nitrophenoxide at 502 nm), using a Beckman DU850 UV/visible spectrophotometer at room temperature for 5 minutes. The molar extinction coefficient of EK under these conditions was 13.26 μM^{-1} cm^{-1}. The data was exported and analysed. The dilution curve of EK was measured as described above, using 200 µM substrate/DTNB. Enterokinase used in the dilution curve was used at 40, 4 and 0.4 ng/ml in duplicates. Substrate/DTNB solution was added to the enzyme in the microtitre plate and readings recorded. The Units/ml kinetic activity was calculated using the following equation

$$U/ml = \frac{\Delta abs^{*} \times V_S}{\varepsilon \times L \times V_T}$$

ε = extinction coefficient, 13.26 $\mu M^{-1}cm^{-1}$, L = path length; 0.32 cm,

V_S = volume of enzyme, V_T = total volume of reaction

* adjusted for substrate blank

Chapter 3: Results

3.1 Choice of EK protein

In order to check the feasibility of expressing enterokinase in *N. benthamiana* leaves, the first task was to choose the target protein sequence to use from the five sequences reported in the literature (see Section 1.1.2). This was done first by reviewing the literature on history of use, reaction specificities and then applying a bioinformatics approach to analyse the evolutionary relationship.

From the five sources of EK identified in the literature, it was established that the heavy chain is not necessary for fusion protein cleavage. Therefore, all further analysis put this into consideration. Thereafter, the enzyme activity and specificities was used to evaluate the different enterokinase sources as shown in Table 3-1. Although the mouse and rat EK have been isolated, no activity and specificity data has been reported and thus are not described in this analysis. Literature shows that the order in activity of the mammalian-derived enterokinase sources is human > bovine > porcine as shown in Table 3-1. In fact, Gasparian (2006) reported that human EK is 10-fold more active than both the bovine and porcine forms. However, when specificity is considered the desirable highly active human EK is defeated by non-specific activity. Bovine EK has been reported to be more specific towards substrate GD_4Kna than the more active human EK (Gasparian *et al*., 2006). With that in mind, bovine EK light chain (accession number L19663.1) was selected as the candidate gene of choice. Furthermore, the catalytic subunit was also chosen as it has been shown to have similar activity to the holoenzyme, whereas for larger substrates such as chymotrypsin, the presence of the heavy chain increases specificity by a factor of 5 (Simeonov *et al.*, 2011). Subsequently, a bioinformatics approach was used to understand the differences and similarities between the different EK sources and evaluate if there are any obvious sequence-derived differences that can explain the biochemical properties of the enzymes.

3.2 Evolutionary relationship of mammalian enterokinase

In order to establish the relationship between the enterokinase sequences from different sources the protein sequences were aligned using CLUSTAL 2.1. Sequence alignment showed a high level sequence homology in the mammalian EK

sources at 70.8 % sequence homology, 11.2 % conserved substitutions and 4.1 % semi-conserved substitutions across the mammalian sequences (Figure 3-1). There was a high level of conservation across the enterokinase enzymes with two possible groups visible: the bovine-human-porcine and rat-mouse EK groups (Figure 3-1). Two major differences were observed between the groups, where the rodent (mouse and rat) had a quad-peptide sequence (24SGDR) and a one amino acid insertion (137Y), both of which were absent in the bovine-human-porcine EK group.

To further visualise the grouping and similarities among the EK protein sequences, the evolutionary distances were plotted as a rooted tree diagram (Figure 3-2). The tree has two distinct branches, being the bovine-porcine-human and rat-mouse branches identified in the sequence alignment. The two branches are divergent from the root as: rat-mouse at 0.216 and bovine-porcine-human at 0.227. Although the rat and mouse EK seems similar, they are divergent by 0.06 between themselves. The bovine-porcine-human branch has a human sub-cluster which is divergent by 0.05 from the bovine-porcine sub-cluster. In the bovine-porcine sub-cluster, the bovine and porcine are divergent by 0.02. Overall, the phylogenetic tree showed that the rat-mouse cluster is more divergent compared to the bovine-porcine cluster being less divergent. The rat and mouse EK variants are more divergent whereas the bovine and porcine EK proteins have less divergence at 0.02 thus displaying homology. When one combines the information from the bioinformatics approach and the evolutionary analysis of the proteins, it can be concluded that the bovine light chain enterokinase gene could be the most suitable candidate for ectopic expression in tobacco leaves.

Table 3-1: Enzyme activity and specificity of mammalian enterokinases (Gasparian *et al.,* 2006)

Source	Activity (EKU/mg)	Specificity
Bovine EK	138	+++
Porcine EK	138	+
Human EK	1462	++

Adapted from Gasparian *et al*., 2006

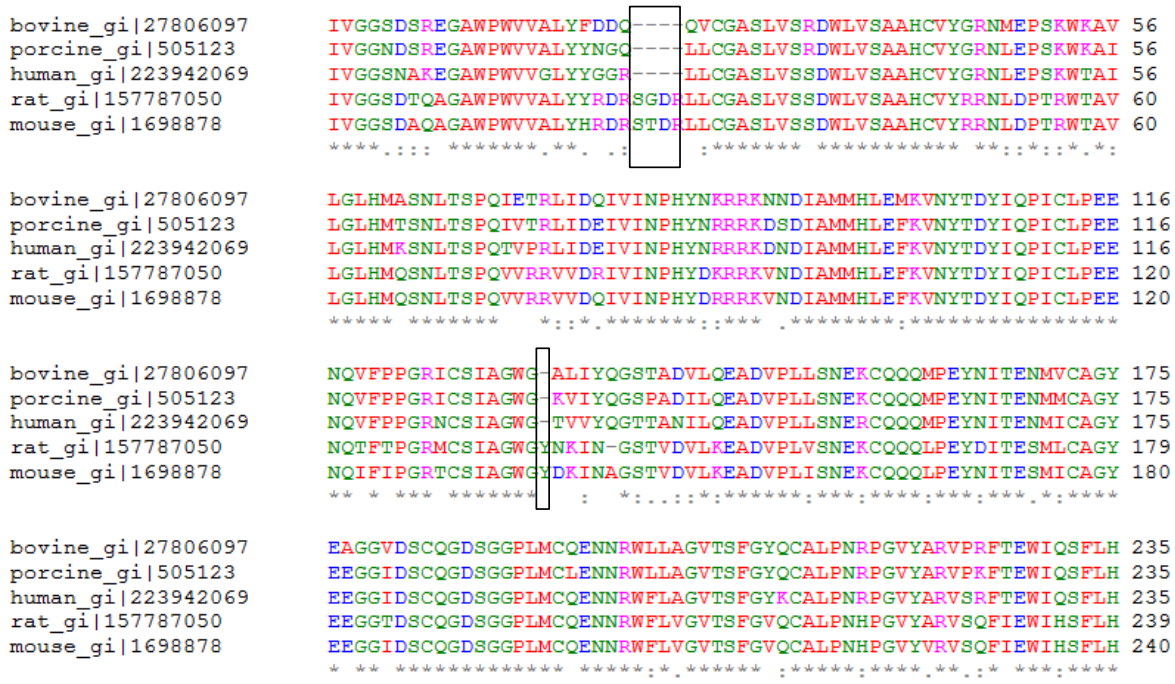


Figure 3-1: Enterokinase multiple sequence alignment of different mammalian sources. " * " – the amino in that column are identical in all sequences; " : " – conserved substitutions have been observed; " . " – semi-conserved substitutions are observed. Red = small hydrophobic amino acids, blue = acidic amino acids, magenta = basic amino acids, green = hydroxyl, amino and basic amino acids. http://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalw2-I20130130-082617-0478-52074348-pg&tool=clustalw2&showColors=true

3.3 Expression of bovine EK gene in plant leaves

To express the bovine EK light chain protein in *N. benthamiana*, two gene forms were synthesised as the plant-codon optimised (PEK) and non-codon optimised. native (NEK). As a result different codon preferences have to be evaluated (See Section 5.2.1 for the codon optimisation report and plasmids from Genscript). The genes were synthesised such that they incorporated a HIS_6 tag sequence at the C-terminus of the protein. The HIs_6 tag was incorporated for Ni^{2+} affinity purification (Tan *et al*., 2007). For easy cloning the genes also had *Nco*I and *BamH*I restriction enzyme sites at the 5′ and 3′-ends of the sequences.

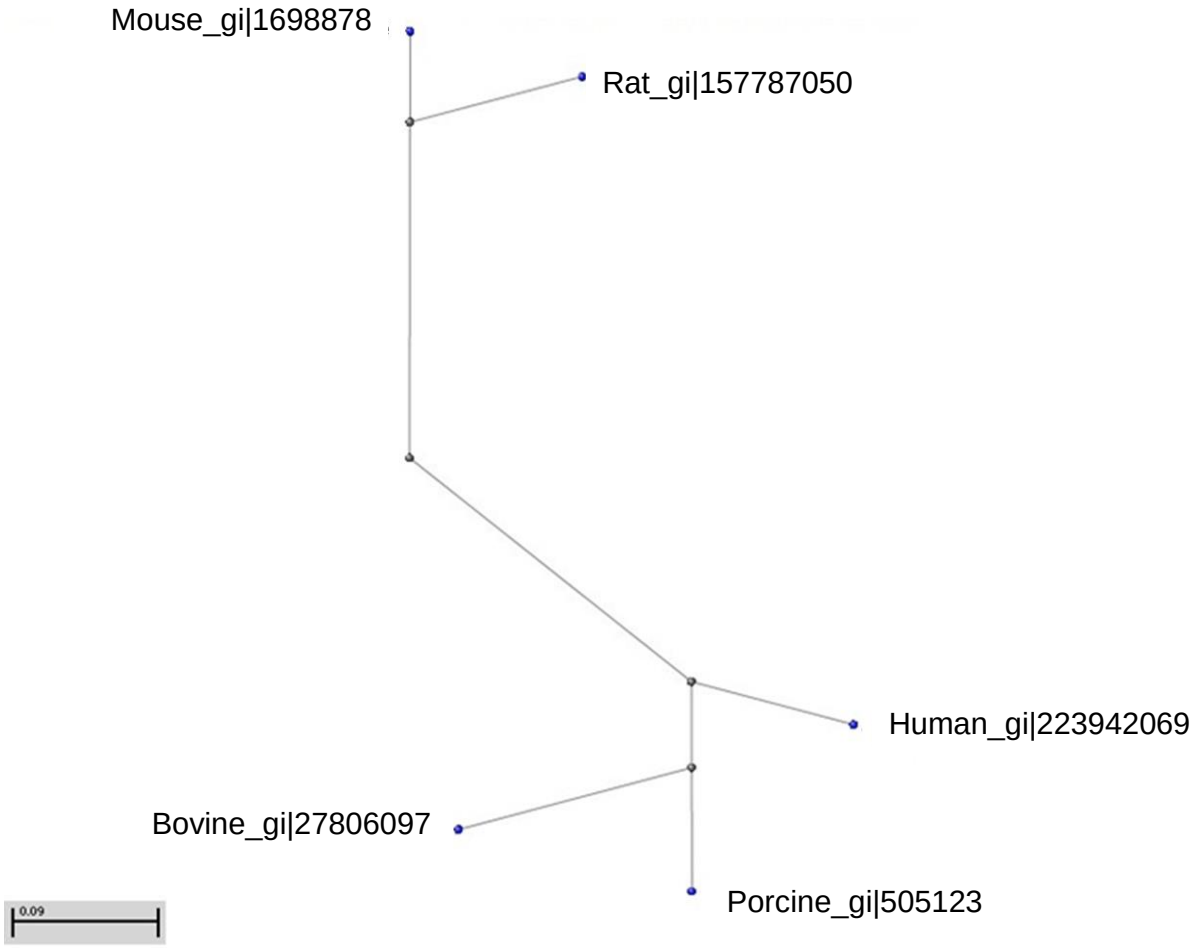


Figure 3-2: Rooted tree diagram showing evolutionary distances of mammalian EK protein sequences. Scale provided on the bottom left corner. http://www.ebi.ac.uk/Tools/services/web_clustalw2_phylogeny/toolresult.ebi?tool=clustalw2_phylogeny&jobId=clustalw2_phylogeny-I20121025-181604-0920-81483926-pg

3.3.1 Enterokinase gene optimisation

A total of 141 codons were altered in the NEK sequence to match the codon usage preferences of Nicotiana benthamiana, yielding an optimised-plant version of the EK sequence named PEK. The optimization strategy removed 58 % of all the infrequently used codons in *Nicotiana benthamiana*. The native gene sequence possesses tandem rare codons that reduce the efficiency of translation and disengage translational machinery. In order to avoid potential negative factors influencing protein expression, the following cis-acting elements were removed; 2 x polyA tails 255AATAAA and 377AATGGA, undesirable peaks and stem-loop

structures were optimised out and the restriction enzyme sites were retained. The final G + C content of the EK gene was lowered to 46.93 % and the codon adaptation index value (CAI) was increased to 0.76 for matching with tobacco genes as compared to the 0.76 CAI value of the NEK sequence. Adjustment of the G + C content was primarily to pro-long mRNA half-life (Doran, 2006). The nucleotide sequence of the PEK gene was 42 % homologous to the NEK gene. However, the amino acid sequence encoded by the PEK was identical to that of the NEK gene to ensure retention of the original EK sequence (Figure 5-2).

After verification of the synthetic genes by nucleotide sequencing and restriction digestion (Figure 3-3 A), the genes were amplified by PCR from the pUC57-EK template to generate a 758 bp fragment (Figure 3-3 B) that was then cloned into pGEMT-easy vector using the T-A cloning system. The resulting pGEM-PEK and pGEM-NEK plasmids were sent for nucleotide sequencing to ensure that no mutations, insertions or deletions had been incorporated during the PCR amplification. Sequencing of the EK amplicons was performed primers specific to the T7 promoter found upstream gene in the pGEMT-easy vector. The sequencing report was analysed and compared to the EK gene sequence in the provided pUC57-EK synthesis report. Alignments of the both the variants of EK nucleotide sequences was conducted and protein sequence alignment was performed to verify EK protein integrity. Nucleotide sequence alignments of both PCR-PEK and PCR-NEK compared to pUC57-PEK and pUC57-NEK with high sequence homology, as there was high sequence matches shown in Figure 3-4. The nucleotide sequence verification showed that the EK PCR products are highly similar to the PCR templates. Protein sequences of PCR-PEK and PCR-NEK were compared to the translated EK gene submitted for synthesis (Figure 3-5). The alignment showed that there is significant identical sequence homology, therefore showing that the PCR amplification of EK genes yielded correct EK protein products. The addition of 3' HIS_6 tag and restriction enzyme sites to PCR-PEK and PCR-NEK, are show as mis-matches amino acids (237 HHHHHHAAASRD), compared to the reference EK (accession # 1EKB_B). The EK PCR amplicons were then cloned into expression pro-vector as they were verified to have correct nucleotide sequence.

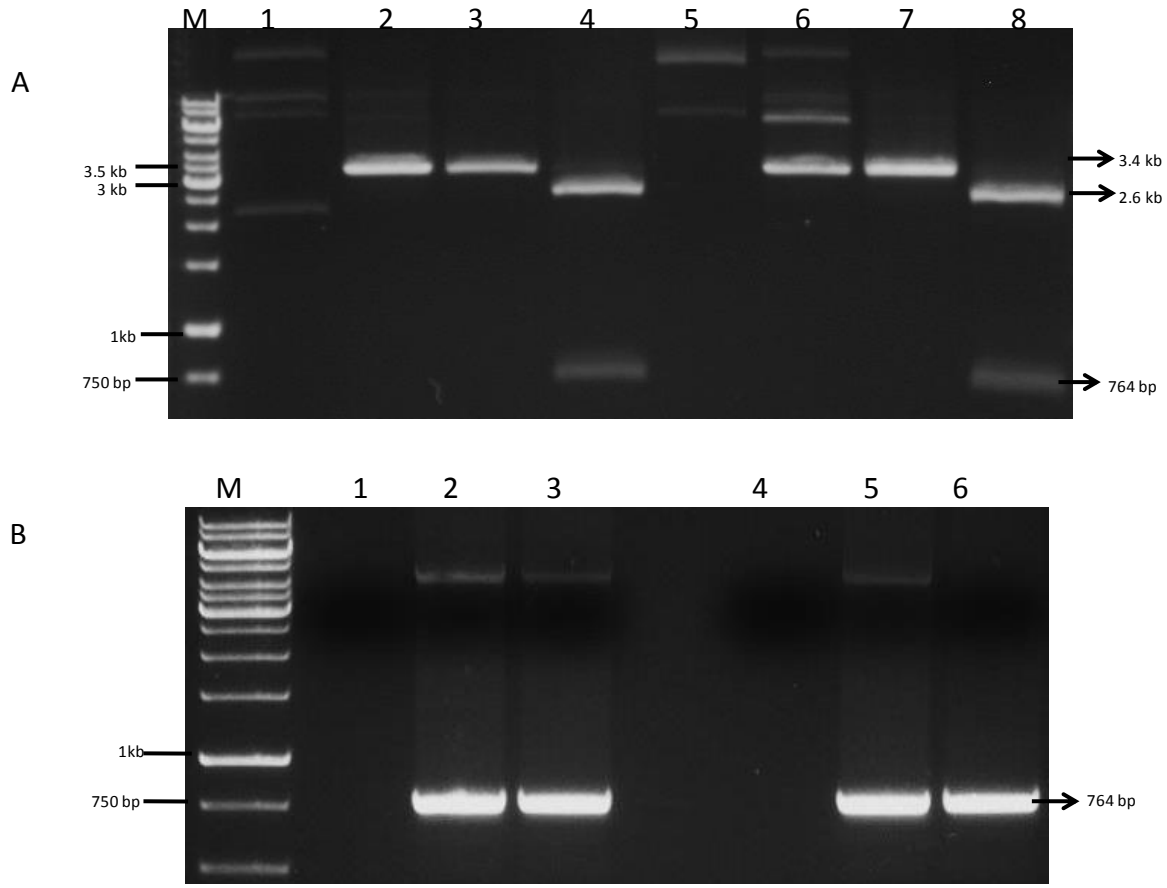


Figure 3-3: Restriction enzyme digestion characterisation of and PCR amplification of EK from pUC57-PEK and pUC57-NEK plasmids. A= Restriction enzyme digest separated on 1% agarose gel. ThermoScientific Generuler 1 kb DNA ladder (SM0311), 1= undigested pUC57-PEK, 2= NcoI digested pUC57-PEK, 3 = BamHI digested pUC57-PEK, 4 = NcoI/BamHI double digested pUC57-PEK, 5= undigested pUC57-NEK, 6= NcoI digested pUC57-NEK, 7 = BamHI digested pUC57-NEK, 8 = NcoI/BamHI double digested pUC57-NEK. B=. Amplification was performed with fermentas long PCR enzyme mix and separated on 1% agarose gel, Thermo Scientific Generuler 1 kb DNA ladder (SM0311); 1 PEK amplification without template DNA (negative control); 2 – 3 PEK amplification from pUC57-PEK; 4 NEK amplification without template DNA (negative control); 5 - 6 NEK amplification from pUC57-NEK

Having established the integrity of the EK gene products in pGEM-NEK and pGEM-PEK, the genes were then sub-cloned into pICH11599 expression pro-vector by digesting with *Nco*I and *BamH*I restriction enzymes and ligating into the same sites of pICH11599 expression vector. The resulting constructs were named pICH-PEK and pICH-NEK for the plant-codon optimised and native genes respectively. The

constructs were also confirmed by restriction digestion with *Nco*I and *BamH*I as shown in Figure 3-6 A.

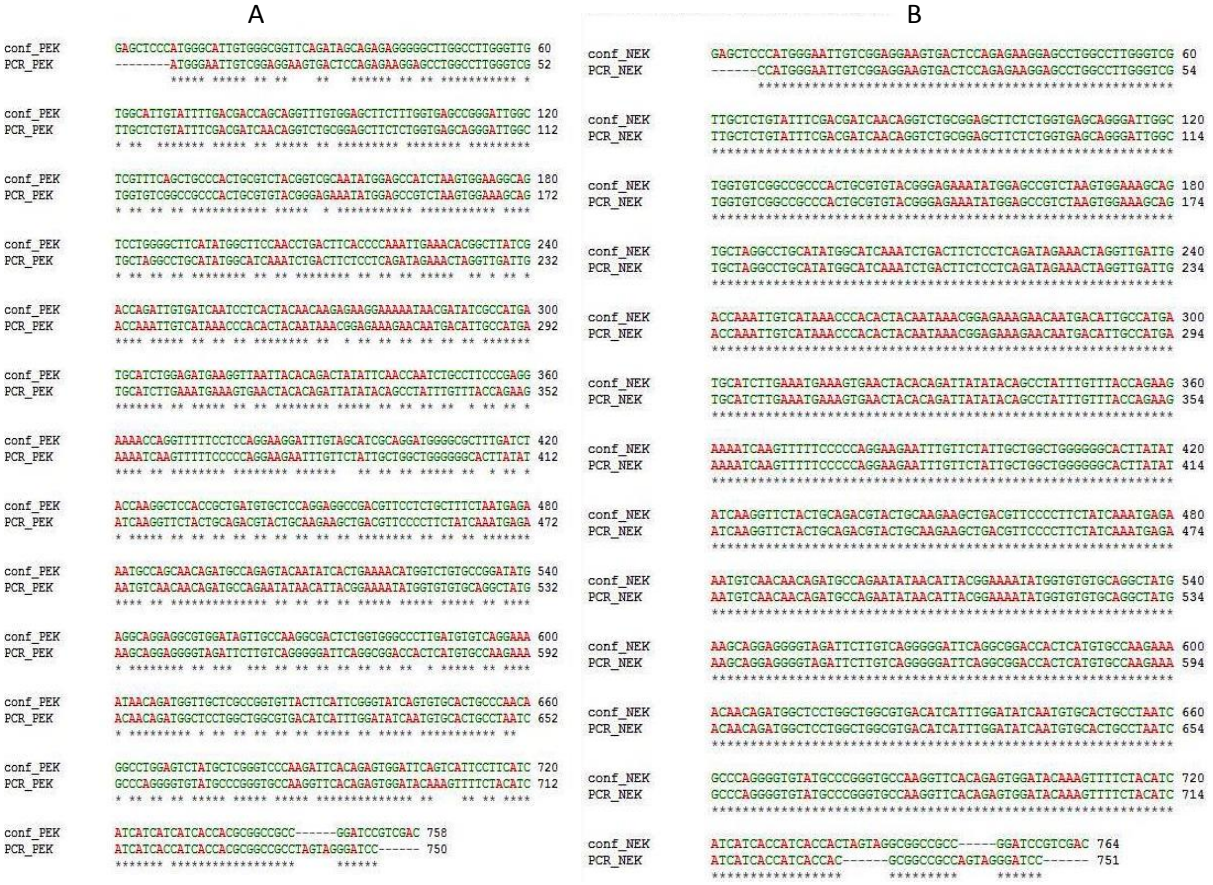


Figure 3-4: Nucleotide sequence multiple alignment of reference EK gene sequence and amplified EK genes. A = Conf_PEK = plant codon optimised EK sequence from pUC57-PEK synthesis report, PCR-PEK = PCR amplified plant codon optimised EK gene sequence, B = Conf_NEK = non-codon optimised EK sequence from pUC57-PEK synthesis report, PCR-NEK = non-codon optimised EK gene sequence, * = consensus between query and subject sequence; - = mis-match between query and subject sequence.

Figure 3-5: Protein sequence multiple alignment of reference EK, PEK and NEK. PEK = plant codon optimised bovine enterokinase light chain protein sequence as translated from the PCR amplification of PEK. NEK = non-codon optimised bovine enterokinase light chain protein sequence as translated from the PCR amplification of NEK. EK = bovine enterokinase light chain protein sequence as found on NCBI. * = consensus between query and subject sequence; - = mis-match between query and subject sequence.

The constructs pICH-PEK and pICH-NEK were transformed into *Agrobacterium tumefaciens* GV3101, and then the constructs were isolated and verified using colony PCR and restriction enzyme analysis (Figure 3-6 A & B). The constructs isolated from *A. tumefaciens* were digested with *Nco*I/*BamH*I, pICH-PEK and pICH-NEK dropped out the correct sized EK genes as indicated by the gel bands in lanes 1 - 3 (764 bp) and 4 - 6 (758 bp) as well as the correct sized linear plasmid of 6.7 kb (lanes 1 - 6) for the cloning pro-vector pICH11599. The restriction enzyme digestion verified the correct plasmids from the transformed *A. tumefaciens.* The generated constructs, pICH-PEK and pICH-NEK were transfected into *N. benthamiana* via *Agrobacterium* as described in Section 2.2.3.2 of the materials and methods.

The MagnICON® vector system is a deconstructed viral vector system based on cr-TMV (crucifer-infecting tobacco mosaic virus) which is an expression system that relies on *in planta* assembly of functional viral vectors from separated pro-vector modules. The *Agrobacterium* containing one of the construct, that is, pICH-PEK, pICH-NEK, pICH7410 (GFP experimental control) and pICH11599 (negative control) mixed with the appropriate integrase and 5′-vector modules were transfected into plant leaves using vacuum infection (see Section 2.2.3.2 for the method, Figure 2-2).

3.4 Plant phenotype and viability

Phenotype analysis of the infiltrated leaves was performed to observe the effect of the enterokinase protein accumulation should it be expressed. The leaves inoculated with pICH-PEK and pICH-NEK constructs were compared to a control samples infiltrated with an empty vector pICH11599. To check if the transfection method and pro-vector system was working well, the GFP construct was used an experimental positive control (Figure 3-7). When the GFP control leaves were visualised with UV light, a gradual increase in fluorescence intensity and spatial distribution was observed on the infiltrated leaves 3 to 12 days post infiltration as shown in Figure 3-7. When the phenotype of the pICH-PEK and pICH-NEK plant samples is considered, no necrosis or leaf curling was observed 3 days post infiltration compared to the control leaves, whereas the leaf samples 6 days post infiltration showed outer leaf curling. The same curling was observed in the leaves infiltrated with EK at 9 and 12 days post infiltration. The leaves infiltrated with EK at 12 days post infiltration developed tiny brown spots of necrosis. The symptoms were consistent in both the pICH-NEK and pICH-PEK regardless of subcellular target (i.e. cytosolic or apoplastic). The changes observed in the EK infiltrated leaves could be a first indication of expression by vector system in the leaves. This suggests that the *A. tumefaciens* was able to deliver the T-DNA with the constructs into the host plant and that all the vector components and buffers were working. The pro-vectors were able to assemble *in planta* and there was possible expression of EK.

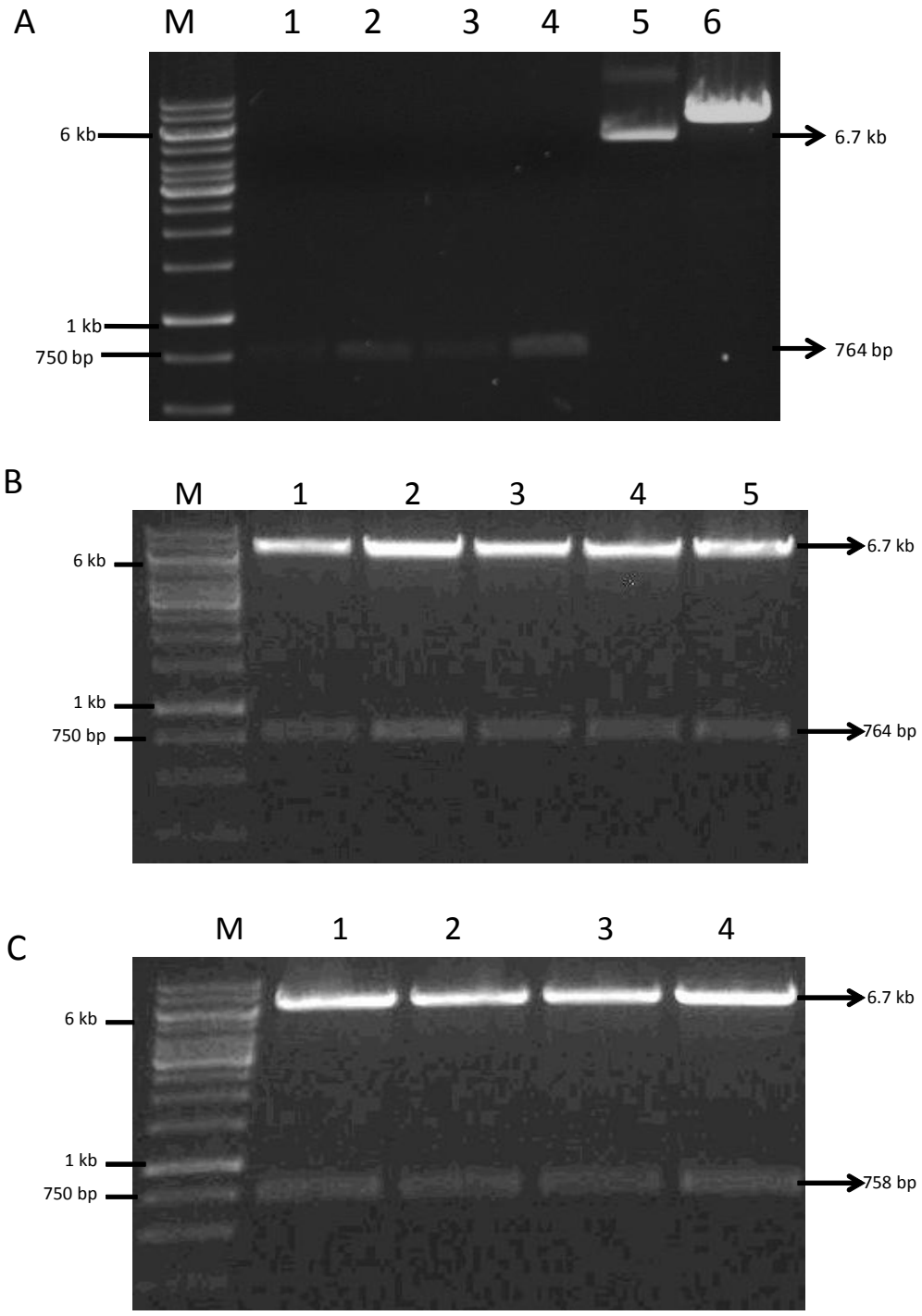


Figure 3-6: Digestion of PCR amplicons PEK, NEK and pICH11599 with *NcoI/BamH*I and pICH-PEK and pICH-NEK verification. (A) M = Thermo Scientific Generuler 1 kb DNA ladder (SM0311), 1 = undigested PCR-PEK, 2 = *NcoI/BamH*I digested PCR-PEK, 3 = undigested PCR-PEK, 4 = *NcoI/BamH*I digested PCR-PEK, 5= undigested pICH11599, 6 = *NcoI*/*BamH*I digested pICH11599. Restriction enzyme digestion of *Agrobacterium* derived multiple colonies of pICH-PEK (B) and pICH-NEK (C) 1 – 5 *NcoI/BamHI* digested pICH-EK constructs.

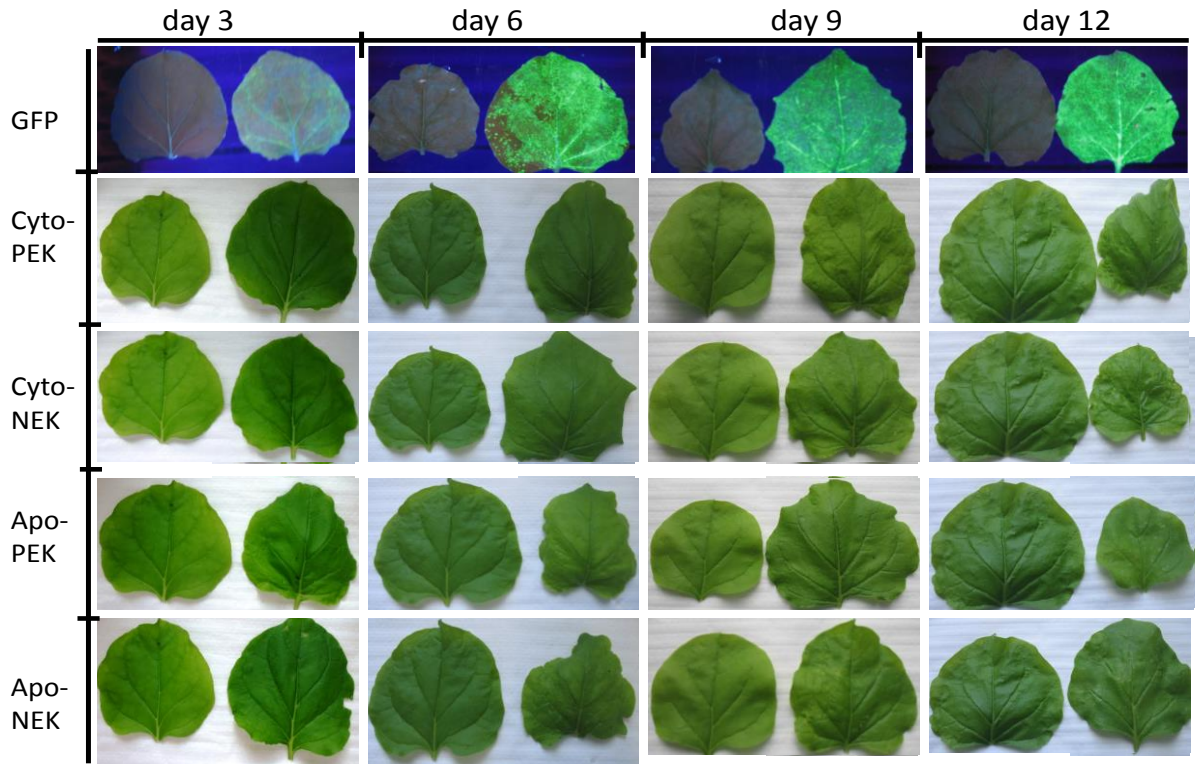


Figure 3-7: Phenotype analysis of infiltrated leaves. GFP: Ultraviolet (UV) light illumination of transfected *Nicotiana benthamiana* leaves to visualise GFP expression by magnICON® vector system. Control = *N. benthamiana* leaf transfected with pICH11599 (left); GFP = *N. benthamiana* leaf transfected with MagnICON vector system expressing GFP protein. Cyto-PEK = plant codon optimised EK targeted to the cytosol; cyto-NEK = non-codon optimised EK targeted to the cytosol; apo-PEK = plant codon optimised EK targeted to the apoplast; apo-NEK = non-codon optimised EK targeted to the apoplast

3.6 ELISA

The next investigation was to test if the *N. benthamiana* leaves were expressing EK and quantify the expression yields. This was done by conducting an indirect ELISA assay using commercial bovine EK light chain as the positive control following the manufacturer's protocol (see Section 2.3.5). Since the assay was not developed specifically for plant extracts, the first steps in the assay were to optimise for the amount of the standard and also the optimal concentration of the primary antibody.

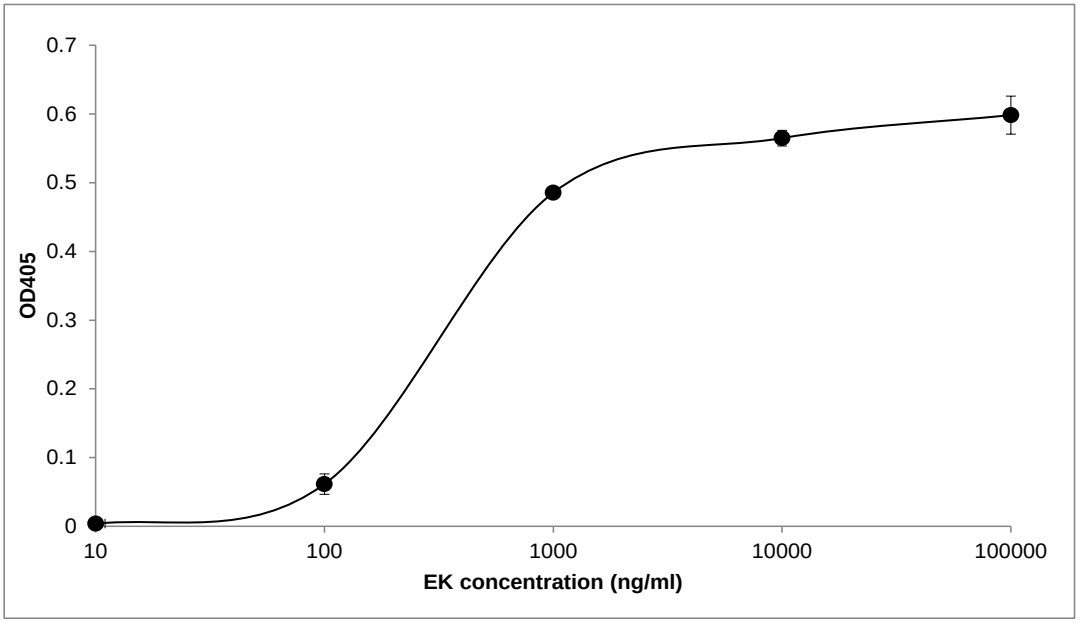
The range of the EK indirect ELISA was determined by using variable amounts of commercial EK and measuring the absorbance at OD_{405} a range of 0.0001 ng/ml – 100 µg/ml in a series of 10-fold dilution series. The linear range at which EK can be

detected was between 100 ng/ml – 1 µg/ml, an OD_{405} reading between 0.1 and 0.5 (Figure 3-8 A). The lower limit of the assay was determined to be at 10 ng/ml EK that can be detected at OD_{405} of 0.004, thereafter further dilutions of EK cannot be significantly detected for quantification. The higher limit of EK detection of 10 µg/ml; OD_{405} of 0.59, can be significantly detected, and serves as a starting point for determining the detection range not the detection limit, although it approaches the limit of the assay as the graph begins to plateau - inferring saturation of EK in the assay. The optimum concentration of EK detectable by the assay was found in the linear range between 100 – 1 000 ng/ml EK, therefore the upper limit of ≤ 1 µg/ml EK was used for standard curve construction, at less than 10 fold dilution – i.e. 5-fold dilution.

To optimise the primary antibody concentration variable concentrations of 0.5 µg/ml – 0.005 pg/ml were used in a 10-fold serial dilution. The range that the primary antibody detected EK was between 0.5 µg/ml and 0.5 pg/ml, shown in Figure 3-8 B. The upper limit of detection was determined by the starting concentration in the experiment of 0.5 µg/ml; OD_{405} of 0.01 and is not the absolute upper limit of the antibody. Ten-fold serial dilutions from 0.5 µg/ml antibody had a lower limit of detection at 0.05 ng/ml; OD_{405} of 0.009 of antibody being able to detect EK in the assay, with further dilutions there was insignificant detection of EK. The optimum concentration of antibody to use in the assay was determined by using the concentrations in the linear range, between 500 – 0.5 ng/ml. The concentration of primary antibody for EK detection was 500 ng/ml as it is in the middle of the inflection point in the detection range. Having established the optimal antibody and detection range of the standard curve, an analysis of EK expression was conducted with the plant extracts. Initially, a question was asked of the accumulation of the enterokinase over a period of 3 - 12 days post infiltration as measured by the ELISA assay.

Recombinant EK expression was measured at a 3 day time-course interval, for the two expression constructs PEK and NEK as well as subcellular targeting, which is summarised in Figure 3-9.

A



B

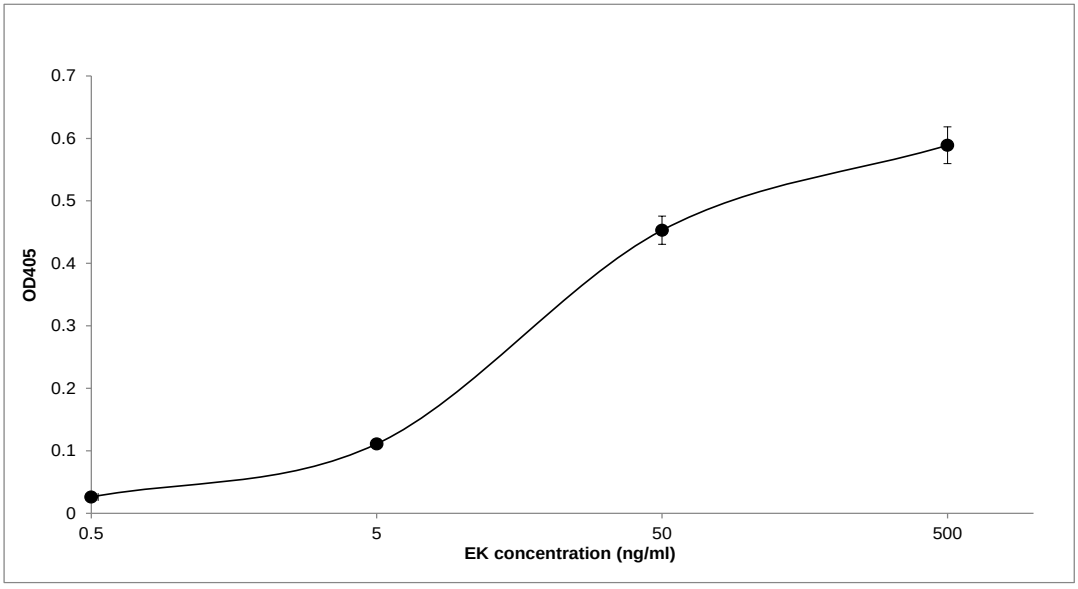


Figure 3-8: Determination of ELISA parameters. The detection range of the indirect ELISA was quantified by graph A, using OD405 = optical density measured at 405 nm to quantify measurable EK in the assay. The detection range of the primary antibody was quantified by graph B, using OD405 to quantify detection range of primary antibody in the assay. Enterokinase concentration measured in ng/ml in assay. OD reading variation indicated by error bars where N = 4. EK = enterokinase, measured in ng/ml.

In comparison to the apoplast EK construct the cytoplast construct expression was very low in both the plant codon optimised and non-codon optimised constructs. In the cytosol constructs, EK expression peaked in day 6 d.p.i leaves for both constructs, 41.7 µg/g FW NEK and 19.2 µg/g FW PEK and then declined between 9 – 12 days post infiltration (Figure 3-9). The decline in expression in both constructs post day 6 post infiltration, declined to look like day 3 expression which suggests a strong possibility of degradation of recombinant EK. When EK expression amounts, for both constructs was compared at day 6, the amount of the plant codon optimised EK gene was about half for apoplast (19.2 µg/g FW) that of the native EK gene (41.7 µg/g FW) suggesting that at optimal day of expression the native gene expressed much better. These observation culminated in the conclusion that native EK expression amounts were highest compared to plant codon optimised EK.

When one compares cytosol targeted EK, a different picture is evident. When cytosol targeted EK is compared, it is different to the expression pattern of apoplast targeted EK. Enterokinase infiltrated leaves at day 3 post infiltration, both plant codon optimised and non-codon optimised EK gene have similar expression amount of 3.6 and 3.6 µg/g FW respectively. Enterokinase expression amount then peaked at day 6 – 9 post infiltration to 15.5 µg/g FW for plant codon optimised EK gene. The non-codon optimised EK gene expression peaked at day 6 post infiltration (28.9 µg/g FW) and then declined to 13.8 µg/g FW along the time course. There was no observed difference in peak expression amounts for cytosol targeted plant codon optimised and native EK constructs, considering the high variation in expression in the day 6 cytosol NEK construct, standard deviation of 15.5 ± 4.5 and 28.9 ± 9.5 respectively (Figure 3-9). When put together, the data clearly shows that sub-cellular location has a significant effect on EK expression. Surprisingly, the non-codon optimised EK gene expresses better than the plant codon optimised EK gene, as well as for both subcellular targets. In addition, the data also shows that regardless of the subcellular targeting or EK optimisation (plant codon optimised or native) the optimum time for harvesting plant-made EK is day 6 post infiltration.

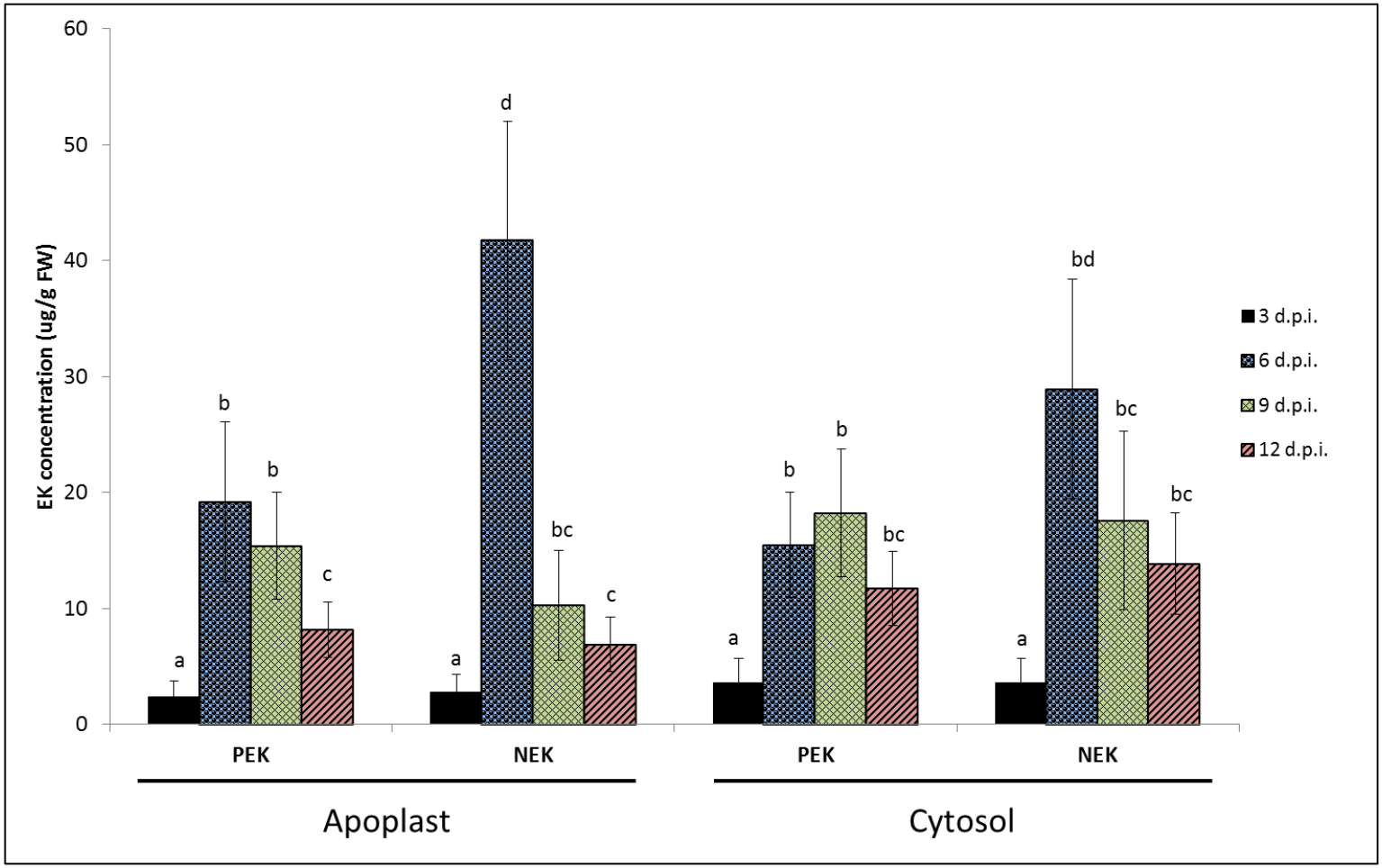


Figure 3-9: Time course accumulation of EK per leaf weight. Accumulation of EK per construct along a time course per leaf weight were quantified by indirect ELISA. Bars with the same letters are not siginificantly different (where $p > 0.05$) and error bars represent the standard error. PEK = plant codon optimised enterokinase, NEK = non-codon optmised enterokinase, d.p.i. = days post infiltration.

3.7 Statistical analysis of enterokinase expression

From the expression data this is evident; apoplast PEK: 6 = 9 > 12 > 3 d.p.i, apoplast NEK: 6 > 9 > 12 > 3 d.p.i, cytosol PEK: 6 = 9 > 12 > 3 d.p.i. and cytosol NEK: 6 > 9 = 12 > 3 d.p.i. Having ascertained the expression profile a student's t-test was conducted to analyse the significance of differences among the expression profiles. The expression samples that are similar were grouped by alphabet to make comparisons between those groups and the groupings were significantly different (Table 3-2). The analysis showed that the expression profiles of *b*, *bc* and *bd* sample groups are not significantly different and the sample groups are significantly different to each other. The statistical data can be shown as follows: *a* < *b*, *bc*, *c* < *bd* < *d*, when factoring in the incremental expression data. Grouping the similar data enabled us to quantify the expression profile variance and how different they are to each other. A full table of the student's t-test analysis is included in the Annexure: Table 5-1.

Table 3-2: Statistical significance of the EK expression group profiles

Description	*a*	*b*	*bc*	*c*	*bd*	*d*
Mean	3.10	16.39	11. 36	7.54	28.46	40.87
Standard deviation	0.60	1.60	1.03	0.90	0.65	1.23
Standard error	0.30	0.92	0.73	0.63	0.46	0.87
p-value		0.00236	0.00218	0.0485	0.00066	0.00669

3.8 Biochemical characterisation of recombinant EK

Having quantified the expression of enterokinase in leaf material, we set to characterise the plant-made EK by SDS-PAGE and western blot. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used the separate the expressed proteins and identification thereof was based on their molecular weight. Infiltrated *N. benthamiana* leaves were ground using 8 M urea and separated by electrophoresis on a polyacrylamide gel as described in Section 2.3.2 of the materials and methods.

The abundant protein RUBISCO is seen migrating at 52 kDa (A) in Figure 3-10, in all of the samples separated by SDS-PAGE. RUBISCO is the most abundant protein in plants due to its function in photosynthesis; it is seen migrating at 52 kDa. The MagnICON vector technology is based on the biology and transfection of tobacco with tobacco mosaic virus (TMV). One of the symptoms of viral infection in plants is chlorosis and stunted growth as a result of the degradation of RUBISCO, and thus chlorophyll in infected cells. This happens because of the rapid synthesis of the viral coat protein during viral replication. In the current study, the reduction in RUBISCO amount could be a quick indicator of the expression of the EK gene. Figure 3-10 shows the SDS-PAGE of the cytosol targeted EK. The RUBISCO protein band is observed at 52 kDa in the negative control (lane neg), the band intensity of the RUBISCO seem to decline from 3 to 9 dpi for both the pICH-PEK and pICH-NEK constructs suggesting possible accumulation of EK to compensate for the RUBISCO. When the EK plant samples were considered, three bands were observed that correlated with the decline in the RUBISCO and absent in the negative control. These were at the 34 kDa that correspond to the theoretical band for bovine EK light chain. Unfortunately, the SDS-PAGE (Figure 3-10) did not have the positive control because the limited amount available was spared for the western blot analysis (see Section 3.8.2 below). Nonetheless, the band was visible in 3 - 12 dpi for PEK samples and 3 -12 dpi. The other two bands were below 30 kDa (C & D), these proteins marked as band C were more pronounced in cytosol-PEK at 6 dpi and cytosol-NEK at day 6. The intensity of the 34 kDa bands was more prominent in the day 6 samples, which correlate with peak expression of EK, shown by ELISA. Looking at the apoplast targeted EK samples, the same banding pattern was observed (Figure 3-10) suggesting that EK was being expressed as these bands did not appear in the negative control. It was then sought to confirm the identity of the putative enterokinase band 34 kDa bands by N-terminal sequencing.

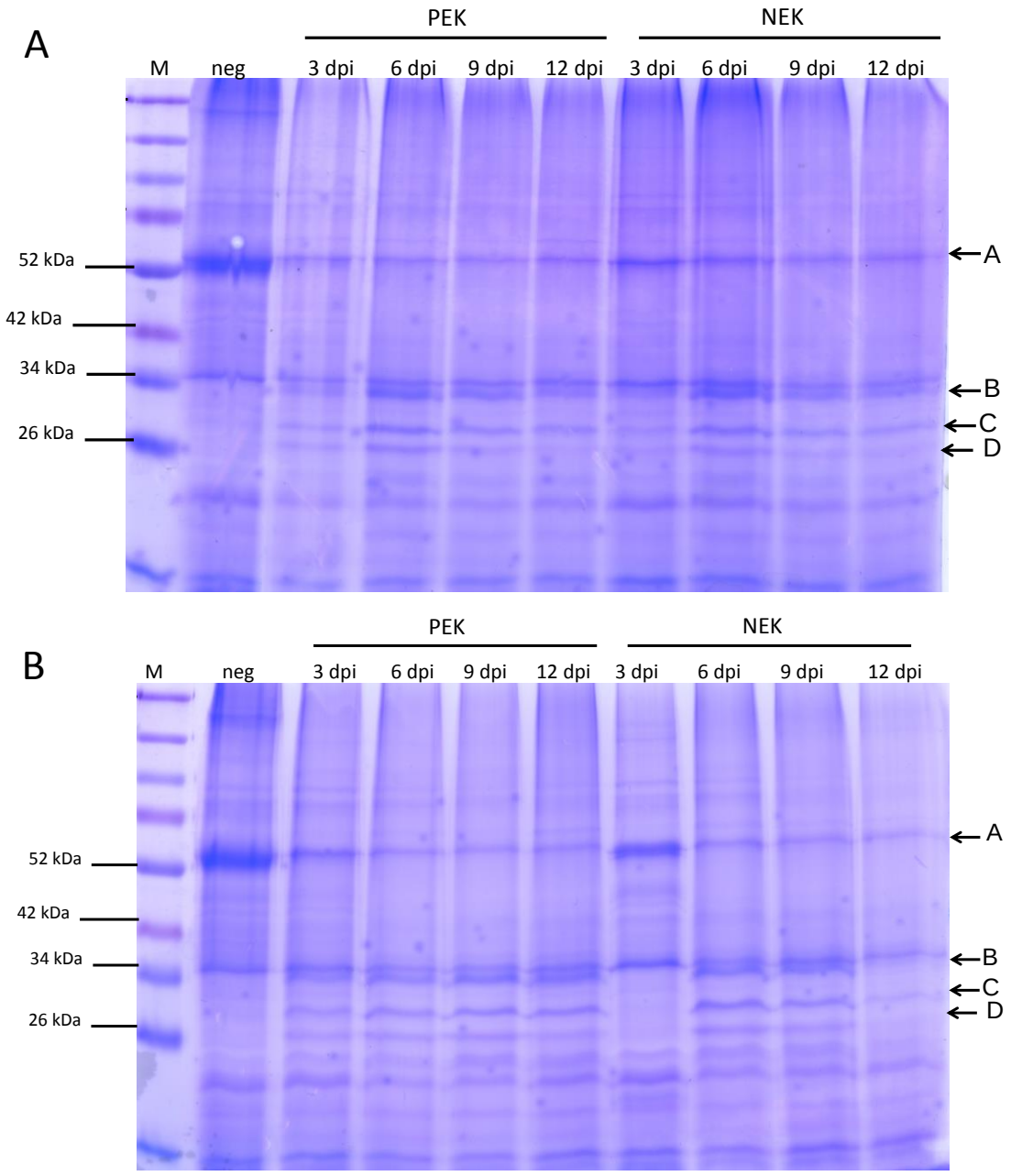


Figure 3-10: SDS-PAGE of plant made EK. M = Fermentas spectra multi-colour broad range protein ladder (SM1841), neg = pICH11599 infiltrated *N. benthamiana*, PEK = plant codon optimsed enterokinase, dpi = days post infiltration. A= cytosol targeted EK SDS-PAGE, B = apoplast targeted EK SDS-PAGE.

3.8.1 Amino acid sequencing

To identify the hypothesised protein, the 34 kDa band (marked B) was excised from the SDS-PAGE and sequenced by LC MALDI (Liquid chromatography Matrix-assisted laser desorption/ ionization). As expected, the peptide 84 LIDQIVINPHYNK was identified in EK bands targeted both the cytosol and apoplast. An additional 43 DWLVSAAHCVYGR peptide was identified in the apoplast-NEK at 6 dpi band. The two peptides identified are significant identifiers for bovine EK light chain protein (PDB reference #: 1EKB_B). BLAST analysis using the two peptides show positive identification of bovine EK light chain (Figure 3-11). The sequencing result therefore confirmed the identity of the 34 kDa as the bovine enterokinase. This provided further proof that EK was expressed in tobacco plants at amounts detectable by SDS-PAGE.

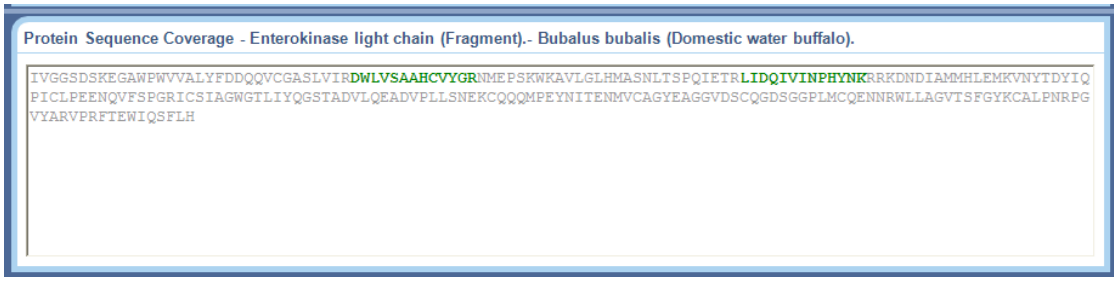


Figure 3-11: Sequencing visualisation of band of interest. Text in green shows the identified peptides that were used to ascertain protein identity of recombinant EK.

3.8.2 Western blot analysis

To further characterise the plant expressed EK, western blot analysis conducted using EK specific monoclonal antibodies as described in Section 2.3.4. After separating the proteins by gel electrophoresis and transfer onto a PVDF membrane, detection was done by ECL. The samples from the cytosol targeted EK and apoplast targeted EK were probed separately.

Monoclonal anti-EK was able to cross react with commercially available EK and produced a band on the membrane (lane +ve) migrating at 42 kDa (1), demonstrating capability of detection of EK. There were no bands detected in the

negative control (lane neg), showing that the mAb detects EK depicted in Figure 3-12 A. Of the cytosol targeted plant-made EK along the time course, the day 9 sample was detected migrating at 32 kDa. The other plant codon optimised cytosol targeted samples were not detected, possibly due to below detection amounts of EK. The cytosol-NEK samples were not detected by immunoblotting also due to insufficient amounts or highly diluted plant-made EK in the plant matrix.

Detection of apoplast targeted recombinant EK was successful as the positive control was detected (lane +ve) migrating at 42 kDa (1) and the negative control did not cross react with anti-EK to produce a band (lane neg) depicted in Figure 3-12 B. Detection of apoplast targeted recombinant EK was in the following samples; plant-codon optimised EK at 6, 9, 12 dpi and non-codon optimised EK at 6, 9, 12 dpi all the bands showing molecular weight of 30 kDa. The western blot analysis shows that plant-made EK was able to be detected, migrating at 32 kDa for the cytosol targeted EK and 30 kDa for the apoplast targeted EK. The detection of plant-made EK using anti-EK antibody illustrates that the antibody can detect plant-made EK and can be used to quantify yields of recombinant EK. Quantification of plant-made EK was quantified as described in Section 2.3.5, data was summarised and represented in Figure 3-9.

3.8.3 Protein purification

Once plant-made EK had been identified by western blot and quantified at detectable amounts by ELISA, protein purification of plant-made EK was conducted to further characterise EK once isolated from the plant matrix. The EK samples that expressed the highest amounts were purified; being the apoplastic and cytoplastic targeted EK at 6 days post infiltration for both variants of the gene optimisation, as described in Section 2.3.6. The non-bound proteins and non-histidine tagged proteins in the plant matrix are evident in the flow-through and wash lanes, as the plant matrix proteins. The target protein, plant-made EK was eluted and analysed in lanes 1,2 and 3 to confirm identity by molecular weight.

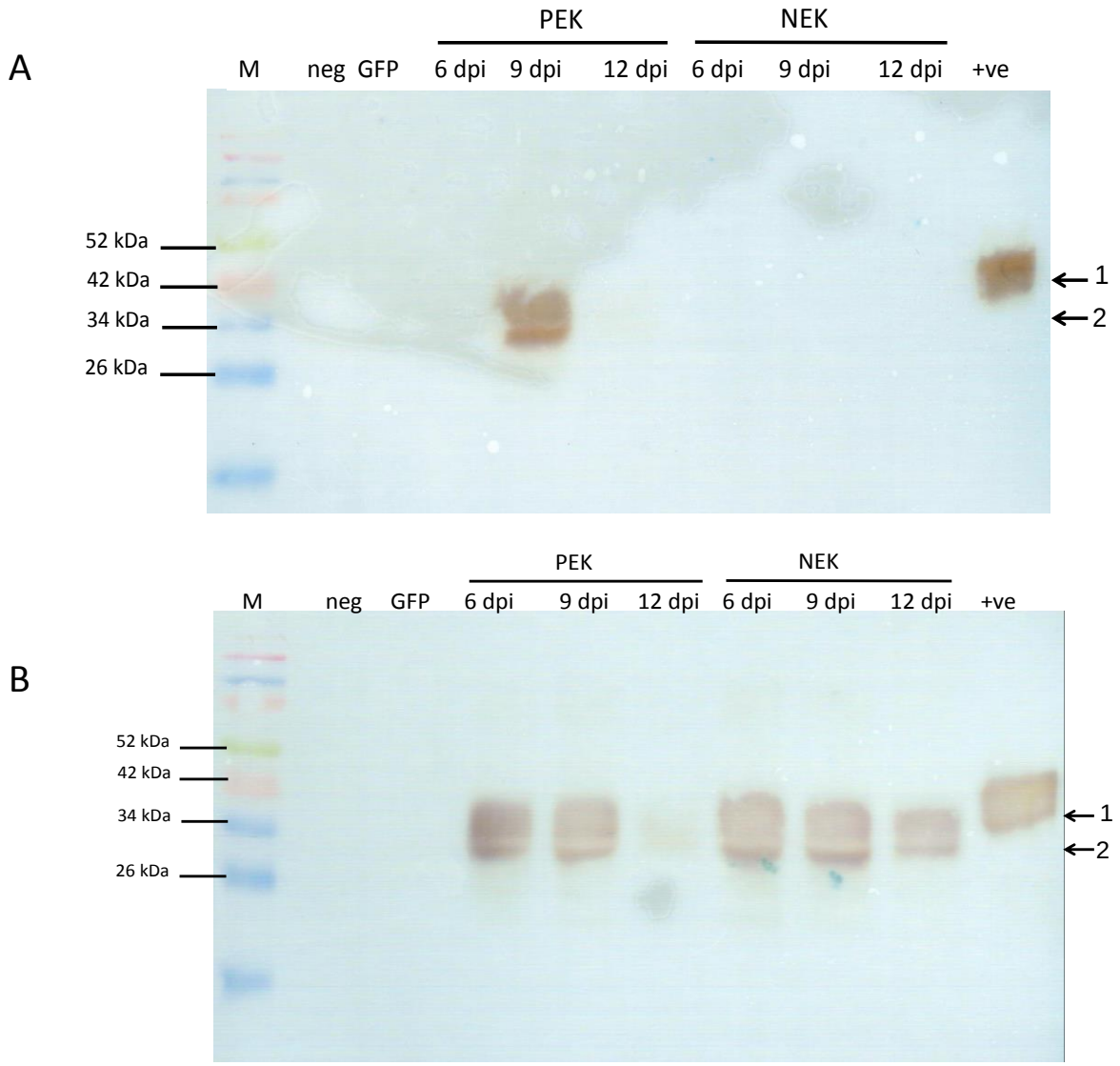


Figure 3-12: Western blot of plant made EK. M = Fermentas spectra multi-colour broad range protein ladder (SM1841), neg = pICH11599 infiltrated *N. benthamiana*, GFP = GFP epresssing *N. benthamiana* leaves, PEK = plant codon optimsed enterokinase, dpi = days post infiltration, 1 = positive control, bovine EK light chain (P8070L, NEB). A= cytosol targeted EK western blot, B = apoplast targeted EK western blot

The eluted protein in the predicted size was not identified for the native EK sequence for both sub-cellular targets, cytoplastic and apoplastic targeting. Purification of plant-made EK was unsuccessful, even when EK samples were concentrated prior to purification, under reducing and non-reducing conditions and using different purification columns and resins (not shown). Further characterization of plant-made EK was made with partially purified EK.

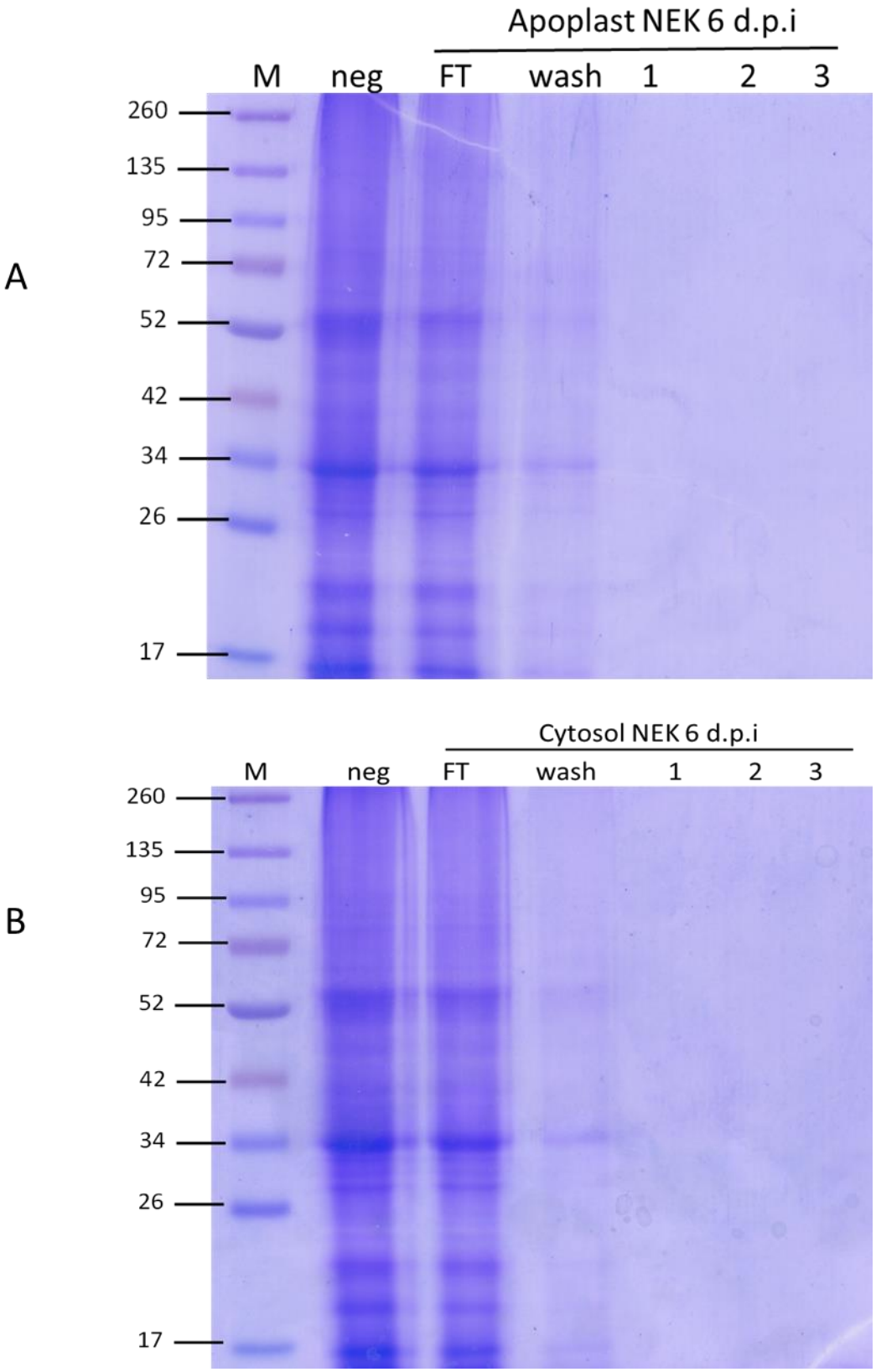


Figure 3-13: Purification of plant-made EK. M = Fermentas spectra multi-colour broad range protein ladder (SM1841), neg = pICH11599 infiltrated *N. benthamiana*, FT = flow through, 1, 2 and 3 = elutes 1, 2 and 3 respectively.

3.8.4 Kinetics assay parameters determination

The range of the kinetics assay was determined by using variable amounts of substrate Z Lys-SBzl, ranging from 0 to 500 µM in 50 µM increments and measuring the change in absorbance at OD_{402} a range for 5 minutes at room temperature at pH 7.5. Along with the variable substrate concentration, 40 ng/ml of commercially available EK and 200 µM of the chromogen DTNB were used. The linear range at which 40 ng/ml EK can catalyse Z Lys-SBzl in the dilution series, with an OD_{402} between 0.01 and 0.07 (Figure 3-14 A). The plot shows a linear relationship between increase in substrate and increase in change in OD_{402}, where the higher amounts of the substrate yield a higher change in OD_{402}. The increase in OD is attributed to the liberation of 3-carboxyl-4-nitrophenoxide as an indicator of kinetic activity of EK. The kinetic data point where there was 0 µM Z Lys-SBzl, there was 0.019 change in OD which inferred background kinetic activity captured by the assay.

The lower limit of the assay was determined to be at 50 µM Z Lys-SBzl that can be detected; OD_{402} of 0.035, < 50 µM dilutions of Z Lys-SBzl was not used for parameter determination as 50 µM was an initial concentration. There was a linear range observed for determination of activity from 50 µM Z Lys-SBzl as seen in Figure 3-14 A. The concentration of Z Lys-SBzl detectable by the assay was found in the linear range between 50 – 500 µM Z Lys-SBzl, without reaching saturation. The starting higher limit of EK activity at 200 µM Z Lys-SBzl; OD_{402} of 0.044, can be significantly detected, and serves as a starting point for determining the detection range not the assay limit, as the assay had not begun to show saturation in the form of a plateau. The concentration of 200 µM Z Lys-SBzl was used as the optimal concentration of substrate to use as it falls in the midpoint of the linear range of the kinetic assay.

To optimise the EK positive control to use in the assay, variable concentrations of EK were used as follows; 40, 4, 0.4 ng/ml in assay buffer, substrate/DTNB solution was used at 200 µM in assay buffer in the same reaction conditions as above. There is a dose response of the activity of EK at different dilutions. Each dilution of EK was plotted individually to visualise the kinetic reaction in change in OD_{402} over 5 minutes, the gradient of each dilution was used as the rate of the reaction in

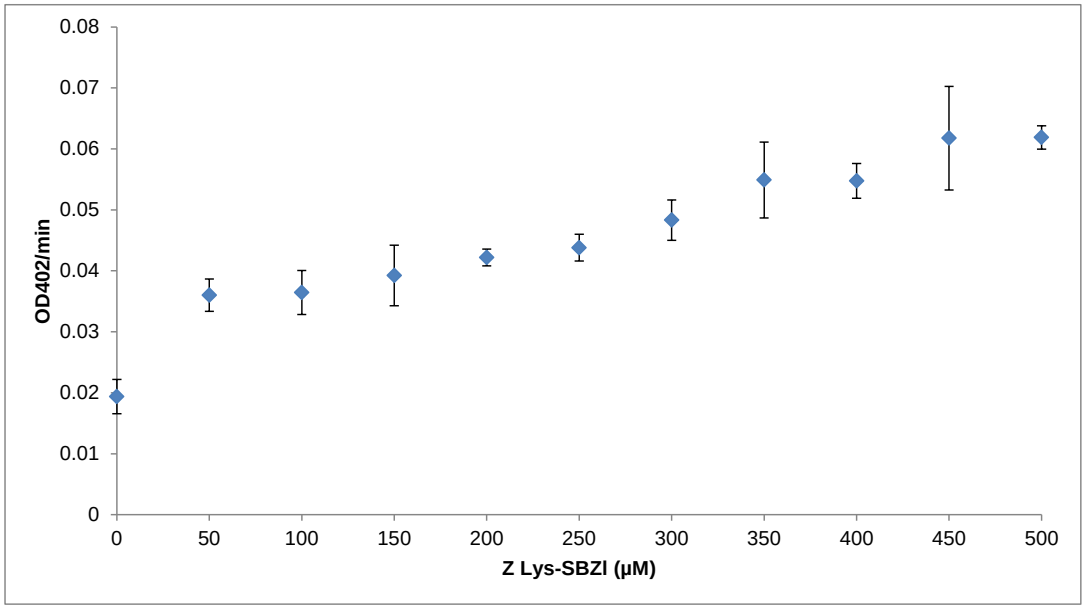
determining specific activity as described in Section 2.3.8. The starting concentration of EK at 40 ng/ml reached reaction completion after 40 seconds and the plots reached a plateau (not shown) at a reaction rate of 262 and starting off at OD_{402} of 0.4 and ending at 0.6 at the plateau (Figure 3-14 B). Ten-fold diluted EK, at 4 ng/ml displayed a linear dose response in the reaction at a reaction rate of 106.1 which is a 2 fold difference in reaction rate to 40 ng/ml EK. Hundred-fold diluted EK, at 0.4 ng/ml had a linear dose response in the reaction at a reaction rate of 69.6, which is also a reaction reduction of 1.5 fold compared to 4 ng/ml EK. The optimum EK concentration to use as a positive control in determining kinetic activity was 4 ng/ml, a linear response at sufficiently low enough amount of EK to use due to its price. Having determined the substrate concentration and EK concentration as a positive control to use in the assay, determination of kinetic activity of crude plant-made EK was carried out.

Using the optimised parameters described above, plant-made EK kinetic activity was determined along with all the relevant controls in duplicate. All the samples, including the negative controls were adjusted for the substrate blank. Fourty nanograms of EK was used as positive control and displayed 117.7 enzyme units per ml of plant extract (U/ml) which is a 5 fold difference in kinetic activity, using the equation provided in Section 2.3.2. Kinetic data of negative control (pICH11599) was measured at 1.2 U/ml, which is indicative the background activity in the plant extracts, in the form of other serine protease that are able to catalyse Z-Lys-SBZL (Figure 3-15). The GFP construct also exhibited similar activity.

There was no observed cumulative effect in enzyme kinetic activity per sample through the time course, rather the enzyme kinetic activity does not show a clear relationship in the time course. The apoplast targeted EK had a gradual increase in kinetic activity that declined after 6 dpi, for both plant codon optimised and non-codon optimised EK. The decline in kinetic activity can be attributed to decrease in expression of apoplast targeted EK, as quantified by ELISA (Figure 3-9). Kinetic activity of 23 U/ml EK in plant extract was obtained, in apoplast targeted EK at 6 dpi, where expression of EK was at its peak. This can be a representative kinetic activity data point of plant-made EK as it shows higher kinetic activity at the highest measured EK expression. The cytosol targeted EK, for both plant codon optimised

and the non-codon optimised constructs had overall kinetic activity of 1.1 U/ml, almost equivalent to the negative controls therefore the cytosol targeted EK kinetics are non-conclusive. We speculate that the kinetic activity is non-conclusive because the expression amount of cytosol targeted EK was significantly low to attain quantifiable kinetic activity data, mimicking background kinetic activity in the samples.

A



B

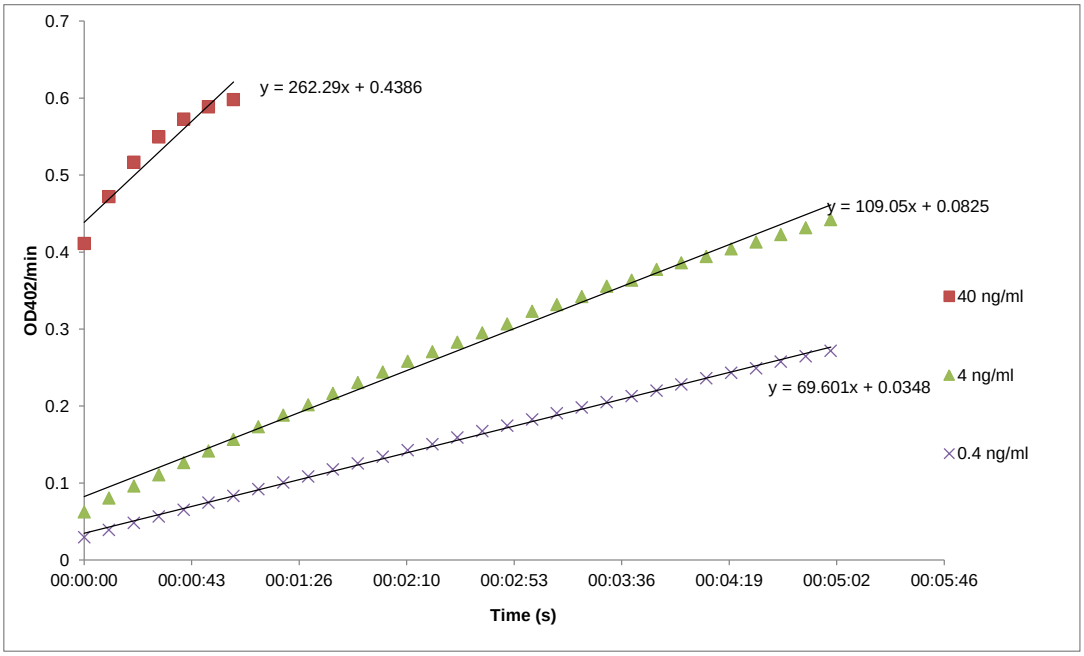


Figure 3-14: Kinetic assay standard curve. Kinetic change in substrate N-α-Cbz-L-lysine thiobenzyl ester hydrochloride was measured at OD402/min = optical density at 402 nm change over time when incubated in variable EK concentrations. OD reading variation indicated by error bars where N = 4. B. Kinetic activity optimal EK

substrate determination at 0.4, 4 and 40 ng/ml to obtain linear dose response reaction to be used as a determined parameter, 40 ng of EK was used.

Although attempts to purify plant-made EK were made, they were unsuccessful and kinetic activity characterisation was conducted with partially purified plant-made EK. Since purified plant-made EK was not obtained the presence of other serine proteases in the plant matrix cannot be discounted. Kinetics data of recombinant EK in plant matrix is not true representation of the activity of purified plant-made EK, as background protease activity is also captured. To accurately measure the activity of plant-made EK, purification has to be achieved and also specific activity measured to be compared to commercially available EK. Kinetic activity data of purified plant-made EK will give a better comparison of enzyme kinetics of various sources of recombinant EK. Specific activity of purified plant-made EK will also compare the specificity activity differences of various sources of recombinant EK, with which to measure the feasibility of plant-made EK. Purification of plant-made EK at the current expression amount makes purification challenging due to the high quantities of plant material required to isolate a few milligrams of purified plant-made EK, to perform definitive kinetic and specific activity assays. The expression amounts of plant-made EK would have to be optimised to achieve better yields to enable purification and specific activity assays. This was the first report of the cloning and expression of active bovine EK light in a plant production system.

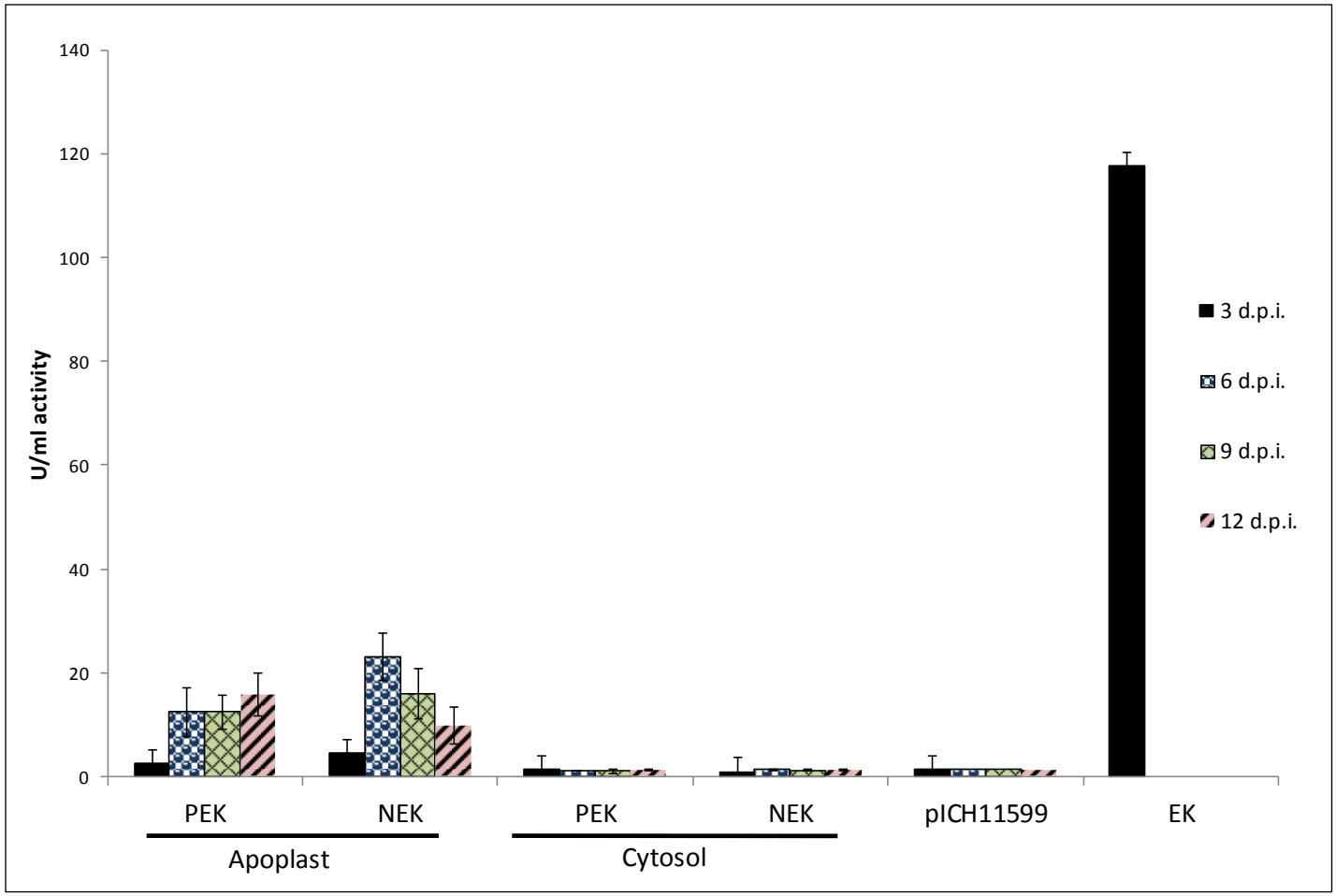


Figure 3-15: Kinetic activity of plant-made EK. Kinetic activity quantification of partially purified EK constructs comapred to commercially available purified EK. Kinetic activity measurements in enzyme units (U) per ml per measurement, kinetic activity standard deviation represented by error bars where N = 4. PEK = plant codon optimised enterokinase, NEK = non-codon optimised enterokinase, EK = commercial positive control, d.p.i. = days post infiltration.

Chapter 4: Discussion and conclusion

There is a growing demand to find alternative methods to produce enterokinase (EK) to complement the current yeast and *E. coli* systems (Mahammad, 2014). Due to its high kinetic and specific activity, EK has been the preferred reagent for protein expression strategies involving the use of fusion proteins of interest to functional peptides which enhance expression, secretion, solubility and purification (Basaran & Rodríguez-Cerezo, 2008). The expression of recombinant proteins is an industry of more than 90 billion USD (Bieri, 2011). Currently, recombinant EK is produced in *E. coli* and yeast however, low yields and erratic activity necessitates finding alternative production methods to meet the commercial demand of EK (Mahammad, 2014). In order to improve production technoeconomics in *P. pastoris* and *E. coli*, previous studies have focused on molecular level manipulation using fusion tags such as the maltose binding protein (MBP), optimizing fermentation conditions and multi-copy integration resulting in the best yield of EK at 479 mg/L (Zhang *et al.*, 2009). The current thesis explored overexpression of the bovine EK light chain in *Nicotiana benthamiana* as an alternative and novel production method. The bovine EK light chain gene was expressed in *N. benthamiana* as a native gene (NEK) and plant-codon-optimised (PEK) and analysed for enterokinase accumulation amounts in the leaves and enzyme activity of the plant-produced protein.

In the study expression of the EK protein was demonstrated initially SDS-PAGE (Figure 3-10), western blot analysis Figure 3-12) showing the expected 34 kDA band (Mahammad, 2014). This result was validated by amino acid sequencing of the protein band by LC MALDI (Figure 3-11), which provided conclusive evidence of the identity of the expressed recombinant protein by biochemical characterisation. Having identified the protein of interest, focus was put on quantifying the expression level by ELISA (Figure 3-9).

4.1 Expression of EK *in planta*

The expression data showed that at peak expression level (6 days post infiltration (dpi)), the trend in protein level expressed on a fresh weight basis was apoplast NEK > cytosol NEK > Apoplast PEK >cytosol PEK (40, 28, 19, 15 µg/g fresh weight (FW); Figure 3-9). The results suggest that the native EK gene produced relatively more

(about double) the amount of protein compared to the plant-codon optimised gene. Although caution is needed when drawing conclusions with this preliminary expression data, the results seem to suggest two discussion points: 1) that codon optimisation is not an exact science and does not always result in better protein expression, and 2) protein subcellular targeting is also empirical science where several constructs needs to be screened in order to get the best protein yield. The above assertions are consistent with some literature which show that codon optimisation to heterologous host preference (in this case *N. benthamiana*) does not a guarantee higher yields and that optimisation of constructs for higher expression is an empirical process. The codon adaptation index (CAI) value is considered a measure of the expression levels of a given gene in different organisms. Theoretically, increasing the CAI of a foreign gene results in a higher expression level of their proteins compared to the native gene. The CAI index of EK was increased from 0.75 to 0.77 with a decrease in A + T content from 51.7 % to 47.9 %. Codon optimisation of EK toward plant codon usage did not achieve higher levels of expression *in planta*, along the time course and the sub-cellular targets. This proved that codon optimisation towards the expression host codon usage does not guarantee higher expression amounts but is empirical. However, studies by Laguía-Becher *et al.*, (2010) have also shown that codon optimisation can increase the A + T content of the heterologous gene, resulting in a decrease in mRNA stability and reduced protein expression in *N. benthamiana*. In their study with human papillomavirus type 16 (HPV-16) L1 gene Laguía-Becher *et al.,* found that the human codon biased gene expressed better than the native gene, which in turn expressed better than the plant optimised gene. Furthermore, the highest expression of human L1 protein had the lowest A + T content, with possible better transcription and RNA processing. This illustrates that codon bias towards the expression host as well as a lowered A + T content doesn't guarantee better expression amounts, which is evident in this study (Maclean *et al*., 2007).

4.2 Sub-cellular targeting and time course

Targeting the EK protein to the apoplast produced 1.5 – 2 times more protein yield than in the cytosol irrespective of codon optimisation (Figure 3-9); these are results

that have been reported previously in recombinant proteins without addition of fusion peptides (Robert *et al*., 2013). This study reports on the expression of EK both in the cytosol and apoplast, albeit that the cytoplastic EK amounts were significantly lower than apoplastic EK. These results indicate that the stability of cytoplastic EK is shorter in comparison to apoplastic EK. When compared to the apoplast, the cytosol has less specific machinery involved in protein folding and processing (Schilliberg *et al*., 1999).The difference in yields could be explained by the difference in the rate of protein degradation between the two compartments, even where *ex vivo* proteolysis occurs during protein extraction and *in vivo* proteolysis occurring during in both compartments that house the 17 peptides associated with plant proteases (Hehle 2011). The literature reports that cytoplast targeted expression is the first choice for recombinant protein expression especially secreted proteins because it occupies much of the cellular volume (Paris, 1996, Santi *et al.,* 2006) but each recombinant protein has differences in biochemical properties and is empirical as shown by apoplastic EK > cytoplastic EK expression amounts. In their study with *Yersinia pestis* antigens F1 and V, Santi *et al.,* (2006) found that the cytosolic targeting was the most efficient for all antigens despite the F1 protein being normally a secreted protein. Interestingly, they also found that the F1 protein also showed high expression levels when targeted to the apoplast, but the cleavage of the targeting pre-sequences appeared incomplete, thus resulting in two bands corresponding to processed and unprocessed protein forms (Santi *et al.,* 2006). Results from this study support our results that the apoplast-targeted EK expresses the both EK gene sequences better than the cytosol-targeted EK.

We speculate that apoplastic plant-made EK accumulated has higher kinetic activity in the apoplast than the cytosolic EK due to the oxidative differences of these cellular targets. It is due to the oxidative nature of the apoplast that formation of disulphide bonds occurs more readily (Chakauya & Chikwamba, 2006), which explains the higher amount of detection and activity of apoplastic EK compared to the cytoplastic targeted EK gene. Protein disulphide isomerase (PDI) catalyses disulphide-coupled folding of proteins and its mechanism is well-studied, which can be used as fusion partner to help catalyse disulphide bond formation (Shin and Scheraga 2000; Shin *et al.* 2002). Chun *et al.,* (2011), used PDI as a fusion partner to produce an active EK

and it was observed that PDI helps EK acquire native disulphide linkages during oxidation period. With PDI fusion, EK was expressed in its soluble form, indicating that EK acquired the native conformation by forming correct disulphide bonds when exposed to oxidizing conditions (Chun *et al.*, 2011). One of the avenues to explore would be the use of a PDI fusion partner to plant-made EK to assist with correct folding, or targeting EK to more oxidised organelles such as the endoplasmic reticulum (ER). The use of inducible expression systems for EK *in planta* could possibly result in higher amounts of EK being harvested upon induction of transgene expression, which would limit the predominant *in vivo* proteolysis and recombinant EK expression would be induced prior to harvesting (Dugdale *et al*., 2013). Putting together all the above thoughts, one can agree with Santi and others (2006) that results from this study and others in the literature show that the inherent difficulty of predicting the expression of any particular antigen and that empirical verification is the only way to proceed (Santi *et al.,* 2006). They also concluded that the advantage of having a suite of vectors such as with the MagnICON system make it possible to conduct such empirical studies where a single antigen construct can be combined pairwise with various targeting elements in separate constructs, with several permutations.

Overall, the maximum protein yield of EK recorded in the current study (40 µg/g FW) is relatively several fold lower than other recombinant proteins of the same size and expressed with the same vector system (Table 1-5). For example, *Yersinia pestis* antigens F1 and V, proteins of 15 and 37 kDa were expressed at 1 - 2 mg/g FW. In a different study Komarova *et al.*, (2010) reported similar yields with *Plasmodium* antigen 322 of similar molecular weight to EK, using the same deconstructed viral vector system. Our yields are about 50 times less than those reported with similar size proteins and this cannot be attributed to a function of the molecular weight but possibly a number of interacting factors which explain the low productivity of current EK production systems. Zhang (2009) was able to show that EK yields of up to 479.99 mg/L EK in *P. pastoris*, while Vozza *et al., (*1996) recorded 6.3 mg/L. Some of the possible factors affecting the EK yields as recorded in the literature include degradation by proteases *in planta* (van der Hoorn, 2008; Pillay *et al.,* 2012) or *ex vivo* proteolytic degradation during downstream processing (Dugdale, 2013). In

other production systems such as *S. cerevesae* and *P. pastoris*, the different protease species that can degrade EK have been studied, but fewer have been studied in plants (Hehle, 2011), especially apoplastic proteins due to their low abundance (Witzel *et al.,* 2011). The second possible explanation causing low yields could be the different levels of glycosylation between the different hosts. It is well established that bovine EK is 35% carbohydrates and that different hosts glycosylate at different efficiencies (Liepnieks *et al*., 1979). For example, *P. pastoris* has been reported to sometimes excessively glycosylate recombinant proteins compared to the native mammalian protein (Zhang *et al.,* 2009). One could therefore speculate that since the plant-made EK glycosylation pattern has not been studied and that plants are known to add xylose and fructose, it is possible that this may affect the protein yields and efficacy. Having said that, the 40 µg/g FW yield recorded serves as a starting point in optimising the expression of EK in leaves.

4.3 Protein purification of plant-made EK

Plant-made EK was confirmed by molecular weight, sequence determination and cross reactivity with specific anti-EK antibody to verify biochemical characterization. Protein quantification illustrated that plant-made EK can be detected within the plant matrix and therefore isolation would enable further protein analysis to determine specific activity and kinetic activity. Protein purification with the aid of the HIS_6tag in the expression construct was attempted by Ni^{2+} metal immobilization of the recombinant EK (Figure 3-13) but without success. Purification of plant-made EK was unable to elute detectable EK even after concentrating the EK extract; therefore it was not possible to get purified EK protein for enzyme kinetic and specific activity. Success of the purification of plant-made EK would have displayed the full value of expression of EK in a plant expression system, to be able to adequately compare commercial EK to purified plant-made EK for further characterization. There are several reasons why it was difficult to purify the protein including 1) non-specific cleavage of the HIS_6-tag resulting in the target protein not immobilised to $Ni^{2+,}$ where Shahravan *et al.* (2008) extensively reported on the sporadic cleavage of recombinant EK which cannot be discounted from cleavage of the HIS_6tag and thus impeding purification via metal chromatography (Shahravan *et al.,* 2008; Liew *et al*.,

2007 and Choi *et al.* 2001). Secondly, it is possible that the plant EK associates with membranes or other components in the plant matrix. This is possible as it has been shown that protein yields below 1% total soluble protein (TSP) in plants make it difficult to purify the target protein from the plant matrix, where EK in this study is 0.04 % TSP (Marilonnet, 2004). Detergents are required to reduce the hydrophobic interaction of between proteins and membranes, but extraction with detergents increases the concentration of membrane-bound IgG, which makes recombinant proteins isolation more complex (Wilken and Nikolo, 2012). With more time and resources, one could have attempted to extract the apoplast targeted protein by apo-leaching. Apoplast leaching is a technique adapted from Joosten (2012) where an apoplast targeted construct can be lysed/ruptured into solution via vacuum infiltration by rupturing the cell wall and membrane then the recombinant protein would be liberated into solution (Joosten, 2012). Recombinant EK from other expression systems has been able to be purified and have sufficient kinetic and specific activity to be commercially available (Table 1-3; Chun *et al.*, 2011).

4.4 Plant-made EK kinetics

Despite the difficulties of getting the pure protein the crude plant protein extract was used to determine kinetic enzyme activity to confirm the functionality of the protein (Figure 3-15). The kinetic activity of plant-made EK is a critical consideration, since there have been reports of inactive recombinant EK from other production systems (Fang *et al.*, 2004). Plant-made EK was tested for kinetic activity in its partially purified form; where there were significant differences in sub-cellular targeting and effect of codon optimisation. There was no significantly measurable activity in cytosolic targeted EK as it had similar activity to the negative controls (1.1 U/ml), in both the plant codon optimised and native EK constructs. The apoplastic EK kinetic activity was 5-fold lower than the positive control (23 U/ml). Nonetheless, the result showed that plant-made EK has kinetic activity towards the EK substrate in the assay, while NEK was more active than PEK. Having said that, the above results should be used with care as the different activities could be just a reflection of the different titres of the protein of interest as EK is very sensitive and small fold differences in the amounts can result in several magnitude differences in activity. In

summary, one can speculate that if purification had been successful the plant-made EK would yield kinetic and specific activity data that is comparable to the other production systems, which will affirm the use of plants to produce functional EK.

4.5 Conclusion and Future perspectives

This work does show, for the first time that EK can be expressed in a plant production system at measurable amounts at the correct molecular weight and amino acid sequence which require further optimisation to meet production amounts for use in research laboratories. The same expression challenges in other EK production systems were observed in this report (Section 3.6), Expression of the EK in sufficient amounts is difficult, mainly due to very low refolding yields (Gasparian *et al*., 2003, 2006). EK contains four disulphide bonds and accumulates as inclusion bodies in the *E. coli* cytoplasm when expressed alone (Shin *et al.* 2002). The refolding process to obtain the native disulphide bonds is known to be very difficult, especially for proteins with an increased number of disulphide bonds (Shin and Scheraga 2000).

Nevertheless, we have successfully produced the bovine EK light chain in a soluble, active conformation in *N. benthamiana* leaves, providing an alternative expression system, where EK can be produced either with or without a fusion partner for auto-cleavage by recombinant EK (Tan *et al.*, 2007). The research objectives of this study were reached, in that bovine EK light chain was expressed, using the MagnICON vector system in *N. benthamiana* leaves. We observed that apoplastic EK was more active than cytosolic and native EK expressed highest at day 6 after infiltration. Expression feasibility of plant-made EK has been established, where native EK gene sequence was expressed in *N. benthamiana* leaves and optimal day for harvest is at day 6 for apoplastic EK that has the highest kinetic activity. Further work and optimisation are required on apoplastic EK, to establish the kinetic and specific activity once purification is successful. Hence, upon purification the yield may be drastically different compared to the partially purified EK expression achieved *in planta* with a strong impact on the process economics. Expression of plant-made EK has led to more production systems becoming available to produce or use in

research laboratories. The alternative methods of EK production therefore contribute to the advancement of production of other pharmaceutically important proteins that lead rise of the biopharmaceutical industry. For the first time a plant production system has been used to produce functional EK, to address the availability of EK as an industrial reagent for fusion protein technology.

Chapter 5: References and appendix

5.1 References

1. Altschul SF, Warren WM Gish, EW Myers, D. L. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, (215), 403–410.
2. Arnau, J., Lauritzen, C., Petersen, G. E., & Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification*, 48(1), 1–13. doi:10.1016/j.pep.2005.12.002
3. Avesani, L., Marconi, G., Morandini, F., Albertini, E., Bruschetta, M., Bortesi, L., Pezzotti, M. (2007*).* Stability of Potato virus X expression vectors is related to insert size: implications for replication models and risk assessment*. Transgenic Research*, 16(5), 587–97. doi:10.1007/s11248-006-9051-1
4. Bally, J., Job, C., Belghazi, M., & Job, D. (2011). Metabolic adaptation in transplastomic plants massively accumulating recombinant proteins. *PloS One*, 6(9), e25289. doi:10.1371/journal.pone.0025289
5. Basaran, P., & Rodríguez-Cerezo, E. (2008). Plant molecular farming: opportunities and challenges. *Critical Reviews in Biotechnology.* 28(3), 153–72. doi:10.1080/07388550802046624
6. Bieri, C. (2011). *IMAP Industry Report*. Retrieved from http://www.imap.com/imap/media/resources/Pharma_Report_2012_FINAL_2F6C8ADA76680.pdf. Accessed 22 March 2014.
7. Bradford, M.M. 1976. Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry. 72: 248–254, doi:10.1016/0003-2697(76)90527-3, PMID 942051.
8. Chakauya E., Chikwamba, R., Rybicki, E. P. 2006. Riding the tide of biopharming in Africa: considerations for risk assessment: science policy. South African Journal of Science. Volume 102(7) pages 284 – 288.
9. Chelur, D., Unal O., Scholtyssek, M., Strickler, J. (2008). Fusion Tags for Protein Expression and Purification. Retrieved from http://www.biopharminternational.com/biopharm/Upstream+Processing/Fusion-Tags-for-Protein-Expression-and-Purificatio/ArticleStandard/Article/detail/522166. Accessed 22 March 2014.

10. Chikwamba, R., Cunnick, J., Hathaway, D., McMurray, J., Mason, H., & Wang, K. (2002). A functional antigen in a practical crop: LT-B producing maize protects mice against *Escherichia coli* heat labile enterotoxin (LT) and cholera toxin (CT). *Transgenic Research*, 11(5), 479–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12437079
11. Choi, S. I. I., Song, H. W., Moon, J. W., & Seong, B. L. (2001). Recombinant Enterokinase Light Chain with Affinity Tag : Expression from *Saccharomyces cerevisiae* and Its Utilities in Fusion Protein Technology. doi:10.1002/bit.10082
12. Chun, H., Joo, K., Lee, J., & Shin, H.-C. (2011). Design and efficient production of bovine enterokinase light chain with higher specificity in E. coli. *Biotechnology Letters*, 33(6), 1227–32. doi:10.1007/s10529-011-0562-3
13. Conley, A. J., Zhu, H., Le, L. C., Jevnikar, A. M., Lee, B. H., Brandle, J. E., & Menassa, R. (2010). Recombinant protein production in a variety of *Nicotiana* hosts : a comparative analysis, 1–11. doi:10.1111/j.1467-7652.2010.00563.x
14. Doran, P. M. (2006). Foreign protein degradation and instability in plants and plant tissue cultures. *Trends in biotechnology*, 24(9), 426–32. doi:10.1016/j.tibtech.2006.06.012
15. Dugdale B., Mortimer C.L., Kato M., James T. A., Harding R. M., and. Dale J. L. 2013. In Plant Activation: An Inducible, Hyperexpression Platform for Recombinant Protein Production in Plants. *The Plant Cell.* Vol. 25: 2429–2443.
16. Esposito, D., & Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*, 17(4), 353–8. doi:10.1016/j.copbio.2006.06.003
17. Fang, L., Sun, Q.-M., & Hua, Z.-C. (2004). Expression of recombinant chinese bovine enterokinase catalytic subunit in P. pastoris and its purification and characterization. *Acta biochimica et biophysica Sinica*, 36(7), 513–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15248027.
18. Fischer, R., Emans N. 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research*. Volume 9(4) pages 279 – 299.
19. Fischer, R., Stoger, E., Schillberg, S., Christou, P., & Twyman, R. M. (2004). Plant-based production of biopharmaceuticals. *Current Opinion in Plant Biology*, 7(2), 152–8. doi:10.1016/j.pbi.2004.01.007

20. Frutos, R., Denise, H., Vivares, C., Neuhaus, J.-M., Vitale, S., Pedrazzini, E., Ma, J. (2008). Pharmaceutical proteins in plants. A strategic genetic engineering approach for the production of tuberculosis antigens. *Annuals of the New York Academy of Sciences*, *1149*, 275–80. doi:10.1196/annals.1428.089
21. Gasparian, M E, Ostapchenko, V. G., Dolgikh, D. A., & Kirpichnikov, M. P. (2006). Biochemical Characterization of Human Enteropeptidase Light Chain, 71(2). doi:10.1134/S0006297906020015
22. Gasparian, M E, Bychkov, M. L., Dolgikh, D. a, & Kirpichnikov, M. P. (2011). Strategy for improvement of enteropeptidase efficiency in tag removal processes. *Protein Expression and Purification*, 79(2), 191–6. doi:10.1016/j.pep.2011.04.005
23. Gasparian, M E, Ostapchenko, V. G., Yagolovich, A. V, Tsygannik, I. N., Chernyak, B. V, Dolgikh, D. a, & Kirpichnikov, M. P. (2007). Overexpression and refolding of thioredoxin/TRAIL fusion from inclusion bodies and further purification of TRAIL after cleavage by enteropeptidase. *Biotechnology Letters*, 29(10), 1567–73. doi:10.1007/s10529-007-9446-y
24. Gasparian, M E., Ostapchenko, V. G., Schulga, A. a., Dolgikh, D. a., & Kirpichnikov, M. P. (2003). Expression, purification, and characterization of human enteropeptidase catalytic subunit in Escherichia coli. *Protein Expression and Purification*, 31(1), 133–139. doi:10.1016/S1046-5928(03)00159-1
25. Giritch, A., Marillonnet, S., Engler, C., Van Eldik, G., Botterman, J., Klimyuk, V., & Gleba, Y. (2006). Rapid high-yield expression of full-size IgG antibodies in plants coinfected with noncompeting viral vectors. *Proceedings of the National Academy of Sciences of the United States of America*, 103(40), 14701–6. doi:10.1073/pnas.0606631103
26. Gleba, Y, Klimyuk, V., & Marillonnet, S. (2005). Magnifection--a new platform for expressing recombinant vaccines in plants. *Vaccine*, 23(17-18), 2042–8. doi:10.1016/j.vaccine.2005.01.006
27. Gleba, Y., Klimyuk, V., & Marillonnet, S. (2007). Viral vectors for the expression of proteins in plants. *Current Opinion in Biotechnology*, 18(2), 134–41. doi:10.1016/j.copbio.2007.03.002

28. Gleba, Y., Marillonnet, S., & Klimyuk, V. (2004). Engineering viral expression vectors for plants: the “full virus” and the “deconstructed virus” strategies. *Current Opinion in Plant Biology,* 7(2), 182–8. doi:10.1016/j.pbi.2004.01.003.
29. Grant S.G., Jessee J., Bloom F.R., Hanahan D. 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *PNAS*. Volume 87 pages 4645–4649.
30. Gupta, S. (2003). Codon Optimization Shubhra Gupta. Project Report, Arizona State University (pp. 1–12). Arizona.
31. Hehle V. K., Paul M. J., Drake P.M., Ma J. K. C and van Dolleweerd C. J. 2011. Antibody degradation in tobacco plants: a predominantly apoplastic process. *BioMed Central Biotechnology*. 11:128.
32. Holzinger, A., Maier, E. M., Bück, C., Mayerhofer, P. U., Kappler, M., Haworth, J. C., Moroz, S. P. (2002). Mutations in the proenteropeptidase gene are the molecular cause of congenital enteropeptidase deficiency. *American Journal of Human Genetics*, 70(1), 20–5. doi:10.1086/338456
33. Horsch, R. B. (2004). Plant-Made Pharmaceuticals (PMPs). In J. Boutwell (Ed.), *Plant-Made Pharmaceuticals (PMPs)”.* Cuba.
34. Horn, M. E., Woodard, S. L., & Howard, J. a. (2004). Plant molecular farming: systems and products. *Plant Cell Reports*, 22(10), 711–20. doi:10.1007/s00299-004-0767-1
35. Houdebine, L.-M. (2009). Production of pharmaceutical proteins by transgenic animals. *Comparative Immunology, Microbiology and Infectious Diseases*, 32(2), 107–21. doi:10.1016/j.cimid.2007.11.005
36. Huang, L., Ruan, H., Gu, W., Xu, Z., Cen, P., & Fan, L. (2007). Functional expression and purification of bovine enterokinase light chain in recombinant Escherichia coli. *Preparative Biochemistry & Biotechnology*, 37(3), 205–17. doi:10.1080/10826060701386695
37. Huang, Z., Santi, L., LePore, K., Kilbourne, J., Arntzen, C. J., & Mason, H. S. (2006). Rapid, high-level production of hepatitis B core antigen in plant leaf and its immunogenicity in mice. *Vaccine*, 24(14), 2506–13. doi:10.1016/j.vaccine.2005.12.024

38. Jenny, R. J., Mann, K. G., & Lundblad, R. L. (2003). A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression and Purification,* 31(1), 1–11. doi:10.1016/S1046-5928(03)00168-
39. Joosten M. H. A. J. 2012. Isolation of Apoplastic Fluid from Leaf Tissue by the Vacuum Infiltration-Centrifugation Technique. *Methods in Molecular Biology* Volume 835, 2012, pp 603-610.
40. Hunt, I. (2005). From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification*, 40(1), 1–22. doi:10.1016/j.pep.2004.10.018
41. Karg, S. R., & Kallio, P. T. (2009). The production of biopharmaceuticals in plant systems. *Biotechnology Advances*, 27(6), 879–94. doi:10.1016/j.biotechadv.2009.07.002
42. Kim, H. J., Kim, Y. H., Roh, Y. H., Seong, B. L., & Shin, C. S. (2005). Optimization of enterokinase fermentation using a recombinant Saccharomyces cerevisiae, 40, 717–722. doi:10.1016/j.procbio.2004.01.055
43. Kim, Y.-T., Nishii, W., Matsushima, M., Inoue, H., Ito, H., Park, S. J., & Takahashi, K. (2008). Substrate Specificities of Porcine and Bovine Enteropeptidases toward the Peptide Val-(Asp)4-Lys-Ile-Val-Gly and Its Analogs. *Bioscience, Biotechnology, and Biochemistry*, 72(3), 905–908. doi:10.1271/bbb.70732
44. Kitamoto, Y., Veile, R. a, Donis-Keller, H., & Sadler, J. E. (1995). cDNA sequence and chromosomal localization of human enterokinase, the proteolytic activator of trypsinogen. *Biochemistry*, 34(14), 4562–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7718557
45. Klein, T. M., E. D. Wolf, (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327: 70-73.
46. Komarova T. V., Baschieri S. Donini M., Marusic C., Benvenuto E., Dorokhov Y. 2010. Transient expression systems for plant-derived biopharmaceuticals. *Expert Reviews*. Volume 9(8), pages 859–876.
47. Koncz C, Schell J. 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Molecular Genetics*. Volume 204 pages 383–396

48. Kubitzki, T., Noll, T., & Lütz, S. (2008). Immobilisation of bovine enterokinase and application of the immobilised enzyme in fusion protein cleavage. *Bioprocess and Biosystems Engineering*, 31(3), 173–82. doi:10.1007/s00449-007-0191-3
49. Kunitz, M. (1939). *Formation of trypsin from crystalline trypsinogen by means of enterokinase.* The Journal of General Physiology. Retrieved from http://jgp.rupress.org/content/22/4/429.abstract.
50. 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. PMID 5432063.
51. Laguía-Becher, M., Martín, V., Kraemer, M., Corigliano, M., Yacono, M. L., Goldman, A., & Clemente, M. (2010). Effect of codon optimization and subcellular targeting on Toxoplasma gondii antigen SAG1 expression in tobacco leaves to use in subcutaneous and oral immunization in mice. *BMC biotechnology*, 10, 52. doi:10.1186/1472-6750-10-52
52. LaVallie, E. (1997). *Patent No. US6746859B1*. United States of America, International.
53. LaVallie, E. R., Rehemtulla, a, Racie, L. a, DiBlasio, E. a, Ferenz, C., Grant, K. L., Light, A,. (1993). Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. *The Journal of Biological Chemistry*, 268(31), 23311–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8226855
54. Lee, C., Sun, H., Hu, S., Chiu, C., Homhuan, A., Liang, S., Leng, C. (2008). An improved SUMO fusion protein system for effective production of native proteins, 1241–1248. doi:10.1110/ps.035188.108.6
55. Liew, O. W., Jenny Chong, P. C., Lim, Y. Z., Ang, C. X., Amy Lau, Y. C., Yandle, T. G., & Brennan, S. O. (2007). An SRLLR motif downstream of the scissile bond enhances enterokinase cleavage efficiency. *Biochimie*, 89(1), 21–9. doi:10.1016/j.biochi.2006.10.005
56. Light, a, & Janska, H. (1989). Enterokinase (enteropeptidase): comparative aspects. *Trends in Biochemical Sciences*, 14(3), 110–2. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2658218
57. Lu, D., Fütterer, K., Korolev, S., Zheng, X., Tan, K., Waksman, G., & Sadler, J. E. (1999). Crystal structure of enteropeptidase light chain complexed with an

analog of the trypsinogen activation peptide. *Journal of Molecular Biology*, 292(2), 361–73. doi:10.1006/jmbi.1999.3089

58. Lu, D., Yuan, X., Zheng, X., & Sadler, J. E. (1997). Bovine proenteropeptidase is activated by trypsin, and the specificity of enteropeptidase depends on the heavy chain. *The Journal of Biological Chemistry,* 272(50), 31293–300. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9395456
59. Maclean J., Koekemoer M., Olivier A. J., Stewart D., Hitzeroth I. I., Rademacher T., Fischer R., Williamson A.-L. and Rybicki E. P. 2007. Optimization of human papillomavirus type 16 (HPV-16) L1 expression in plants: comparison of the suitability of different HPV-16 L1 gene variants and different cell-compartment localization. *Journal of General Virology.* 88, 1460–1469.
60. Mahammad A., 2014. Production and purification of recombinant enteropeptidase expressed in an insect-baculovirus cell system, *Preparative Biochemistry and Biotechnology*, DOI: 10.1080/10826068.2014.907185.
61. Maliga, P. and I. Graham (2004). Molecular farming and metabolic engineering promise a new generation of high-tech crops. *Current Opinion in Plant Biology* 7: 149-151
62. Marchler-Bauer A. (2011), CDD: a Conserved Domain Database for the functional annotation of proteins, *Nucleic Acids Research*. 39D)225-9
63. Marillonnet, S., Giritch, A., Gils, M., Kandzia, R., Klimyuk, V., & Gleba, Y. (2004). In planta engineering of viral RNA replicons: efficient assembly by recombination of DNA modules delivered by Agrobacterium. *Proceedings of the National Academy of Sciences of the United States of America,* 101(18), 6852–7. doi:10.1073/pnas.0400149101
64. Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V., & Gleba, Y. (2005). Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants. *Nature Biotechnology*, 23(6), 718–23. doi:10.1038/nbt1094
65. Markausakas A., Dreguniene, G. 1997. Expression of Escherichia coli TehA gives resistance to antiseptics and disinfectants similar to that conferred by multidrug resistance efflux pumps. *Antimicrobial Agents Chemotherapy*. Volume 41(2) pages 440 – 444.

66. Maroux, S., Baratti, J., & Desnuelle, P. (1971). Purification and specificity of porcine enterokinase. *The Journal of Biological Chemistry*, 246(16), 5031–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5570436
67. Menassa R. Nguyen V., Jevnikar A., Brandle J. 2001. A self-contained system for the field production of plant recombinant interleukin-10. *Molecular Breeding.* Volume 8(2) pages 177 – 185.
68. Mason H., Haq T. A., Clements J.D., Arntzen C.J. 1998. Edible vaccine protects mice against *Escherichia coli* heat-labile enterotoxin (LT): potatoes expressing a synthetic LT-B gene. *Vaccine*. Volume 16 (13), pages 1336-1343.
69. Miele, L. (1997). Plants as bioreactors for biopharmaceuticals: regulatory considerations. *Trends in Biotechnology*, 15(2), 45–50. doi:10.1016/S0167-7799(97)84202-3
70. Menkhaus, T. J., & Glatz, C. E. (2004). Compatibility of column inlet and adsorbent designs for processing of corn endosperm extract by expanded bed adsorption. *Biotechnology and Bioengineering*, 87(3), 324–36. doi:10.1002/bit.20117
71. Ogiwara, K., & Takahashi, T. (2007). Specificity of the medaka enteropeptidase serine protease and its usefulness as a biotechnological tool for fusion-protein cleavage. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7021–6. doi:10.1073/pnas.0610447104.
72. Paris N. Stanley C. M. Jones R. L. Rogers J. C. 1996. Plant Cells Contain Two Functionally Distinct Vacuolar Compartments. *Cell Press.* DOI: 10.1016/S0092-8674(00)81256-8.
73. Pillay P., Kibido T., du Plessis M., van der Vyver C., Beyene G., Vorster B. J., Kunert K. J., Schlüter U. 2012. Use of Transgenic Oryzacystatin-I-Expressing Plants Enhances Recombinant Protein Production. *Applied Biochemistry and biotechnology*. 168(6), 1608 – 1620.
74. Peng, L., Zhong, X., Ou, J., Zheng, S., Liao, J., Wang, L., & Xu, A. (2004). High-level secretory production of recombinant bovine enterokinase light chain by Pichia pastoris. *Journal of Biotechnology*, 108(2), 185–192. doi:10.1016/j.jbiotec.2003.11.004.
75. Robert S., Khalf M., Marie-Claire Goulet, M-C., D'Aoust M-A, Sainsbury F. and Michaud D. 2013. Protection of Recombinant Mammalian Antibodies from

Development-Dependent Proteolysis in Leaves of *Nicotiana benthamiana*. *Public Library Of Science*. DOI: 10.1371/journal.pone.0070203.

76. Rybicki, E. P. (2010). Plant-made vaccines for humans and animals. *Plant Biotechnology Journal*, 8(5), 620–37. doi:10.1111/j.1467-7652.2010.00507.x
77. Sala, F., Manuela Rigano, M., Barbante, A., Basso, B., Walmsley, A. M., & Castiglione, S. (2003). Vaccine antigen production in transgenic plants: strategies, gene constructs and perspectives. *Vaccine*, 21(7-8), 803–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12531364
78. Sambrook, J., E. F. Fritsch. (1989). Molecular Cloning: A Laboratory Manual. New York, Cold Spring Harbor Laboratory Press.
79. Sanger F, Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*. Volume 25;94(3) pages 441–448.
80. Santana, M., Portillo, M. C., & Gonzalez, J. M. (2010). A cloning strategy to obtain recombinant proteins with identical primary structure to the native forms. *Journal of Biotechnology*, 149(1-2), 21–3. doi:10.1016/j.jbiotec.2010.06.010.
81. Santi L. Giritch A., Roy C. J., Marillonnet S., Klimyuk V., Gleba Y., Webb R., Arntzen C. J.,Mason H. S. 2006. Protection conferred by recombinant Yersinia pestis antigens produced by a rapid and highly scalable plant expression system. *PNAS*. doi: 10.1073/pnas.0510014103.
82. Schilliberg S. Fischer R. 1999. Transient expression of a tumor-specific single-chain fragment and a chimeric antibody in tobacco leaves. *Proceedings of the National Academy of Sciences*. USA., 96(20): 11128-11133.
83. Schouten, a, Roosien, J., Van Engelen, F. a, De Jong, G. a, Borst-Vrenssen, a W., Zilverentant, J. F., Bosch, D. (1996). The C-terminal KDEL sequence increases the expression level of a single-chain antibody designed to be targeted to both the cytosol and the secretory pathway in transgenic tobacco. *Plant Molecular Biology,* 30(4), 781–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8624409.
84. Shin H. C., Scheraga H. A. 2000. Catalysis of the oxidative folding of bovine pancreatic ribonuclease A by protein disulfide isomerase. *Journal of molecular biology.* DOI: 10.1006/jmbi.2000.3928.

85. Shin, E.S., Yang M. J., Jung K. H., Kwon E. J., Jung J. S., Park S. K., Kim J., Yun H. D. and Kim H. 2002. Influence of the Transposition of the Thermostabilizing Domain of Clostridium thermocellum Xylanase (XynX) on Xylan Binding and Thermostabilization. *Applied Environmental Microbiology*. doi: 10.1128/AEM.68.7.3496-3501.2002.
86. Shahravan, S. H., Qu, X., Chan, I.-S., & Shin, J. a. (2008). Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag. *Protein Expression and Purification*, 59(2), 314–9. doi:10.1016/j.pep.2008.02.015
87. Simeonov, P., Berger-Hoffmann, R., Hoffmann, R., Sträter, N., & Zuchner, T. (2011). Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield. *Protein Engineering, Design & Selection* : *PEDS*, 24(3), 261–8. doi:10.1093/protein/gzq104
88. Smith, G. P. (1995). Pavlov and Appetite, 30(2), 169–174.
89. Song, H., Choi, S., & Seong, B. L. (2002). Engineered Recombinant Enteropeptidase Catalytic Subunit : Effect of N-Terminal Modification, 400(1), 1–6. doi:10.1006/abbi.2001.2737
90. Streatfield, S. J. (2007). Approaches to achieve high-level heterologous protein production in plants. *Plant Biotechnology Journal*, 5(1), 2–15. doi:10.1111/j.1467-7652.2006.00216.x
91. Svetina, M., Krasevec, N., Gaberc-Porekar, V., & Komel, R. (2000). Expression of catalytic subunit of bovine enterokinase in the filamentous fungus Aspergillus niger. *Journal of Biotechnology*, 76(2-3), 245–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10656339
92. Tan, H., Wang, J., & Zhao, Z. K. (2007). Purification and refolding optimization of recombinant bovine enterokinase light chain overexpressed in *Escherichia coli*. *Protein Expression and Purification*, 56(1), 40–7. doi:10.1016/j.pep.2007.07.006
93. Teli, N. P., & Timko, M. P. (2004). Recent developments in the use of transgenic plants for the production of human therapeutics and biopharmaceuticals. *Plant Cell, Tissue and Organ Culture*, 79(2), 125–145. doi:10.1007/s11240-004-0653-0

94. Thomas, D. R., Penney, C. a, Majumder, A., & Walmsley, A. M. (2011). Evolution of plant-made pharmaceuticals. *International Journal Of Molecular Sciences*, 12(5), 3220–36. doi:10.3390/ijms12053220

95. Twyman, R. M., Stoger, E., Schillberg, S., Christou, P., & Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. *Trends In Biotechnology*, 21(12), 570–8. doi:10.1016/j.tibtech.2003.10.002

96. Tzfira, T., & Citovsky, V. (2006*).* Agrobacterium-mediated genetic transformation of plants: biology and biotechnology*. Current Opinion In Biotechnology*, 17(2), 147–54. doi:10.1016/j.copbio.2006.01.009.

97. van der Hoorn 2008. RAL: Plant Proteases: From Phenotypes to Molecular Mechanisms. *Annual Review of Plant Biology*. 59(1):191-223.

98. Vozza, L. A. *Production of a recombinant bovine enterokinase catalytic subunit in the methylotrophic yeast Pichia pastoris. Biotechnology* (Nature Publishing Company) 14.1 (1996): 77.

99. Waugh, D. S. (2005). Making the most of affinity tags. *Trends In Biotechnology*, 23(6), 316–20. doi:10.1016/j.tibtech.2005.03.012.

100. Weiner, G.L Costa, W Schoettlin, J Cline, E Mathur, J.C Bauer. 1994. Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction. *Gene*. Volume 151, pp. 119–123

101. Wilken L. R., Nikolo Z. L. 2012. Recovery and purification of plant-made recombinant proteins. *Biotechnology Advances*. Volume 30, Issue 2, March–April 2012, Pages 419–433.

102. Witzel K., Shahzad M., Matros A., Mock H-S and Mühling K. H. 2011. Comparative evaluation of extraction methods for apoplastic proteins from maize leaves. *Plant Methods*, 7:48 doi:10.1186/1746-4811-7-48.

103. Yuan, L.-D., & Hua, Z.-C. (2002). Expression, purification, and characterization of a biologically active bovine enterokinase catalytic subunit in Escherichia coli*. Protein Expression And Purification,* 25(2), 300–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12135563.

104. Zhang, J., Wang, X., Mao, X., & Wei, D. (2009). High-level Production of Bovine Enterokinase Light Chain Using Fed-batches by Recombinant Pichia pastoris*,* 23(2), 219–224.

5.2 Appendix

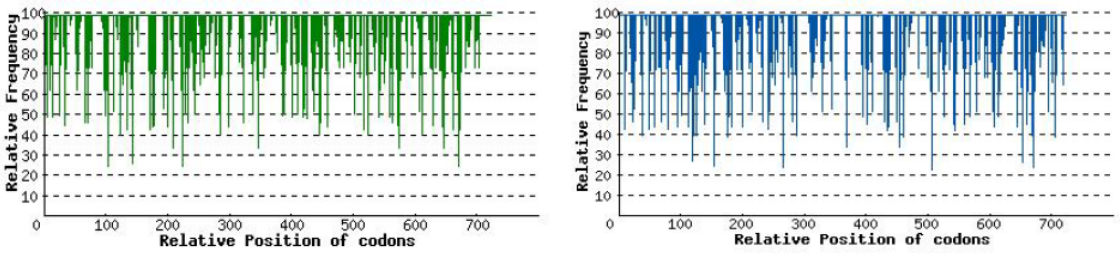
5.2.1 GenScript codon optimization report on EK genes

Results ***Nicotiana tabacum(Tobacco)***

1. Codon usage bias adjustment

Codon Adaptation Index (CAI)

After OptimumGene™ Optimization



0.76 CAI: 0.76

Figure 1a. The distribution of codon usage frequency along the length of the gene sequence. A CAI of 1.0 is considered to be perfect in the desired expression organism, and a CAI of > 0.8 is regarded as good, in terms of high gene expression level.

Frequency of Optimal Codons (FOP)

After OptimumGene™ Optimization

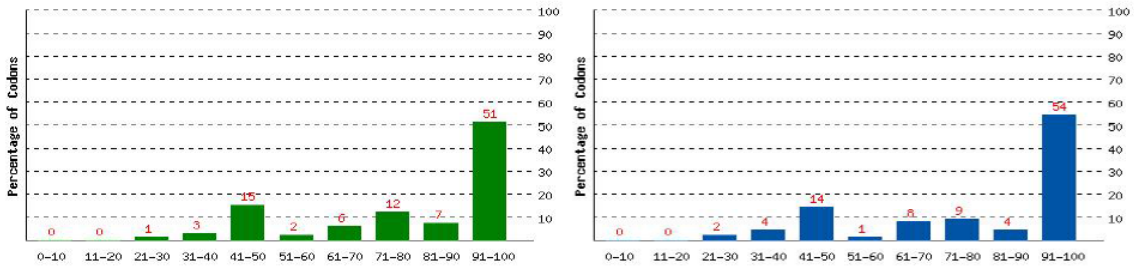
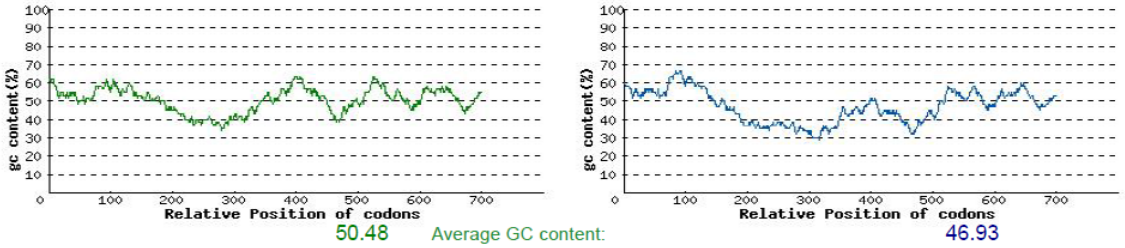


Figure 1b. The percentage distribution of codons in computed codon quality groups. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism.

2. GC Content Adjustment

GC Content Adjustment

After OptimumGene™ Optimization



50.48 Average GC content: 46.93

Figure 2. The ideal percentage range of GC content is between 30-70 %. Peaks of %GC content in a 60 bp window have been removed.

3. Restriction Enzymes and CIS-Acting Elements

Restriction Enzymes	Optimized	Original
* Green: filtered sites; Blue: checked sites (not filtered); Red: kept sites.		
NcoI(CCATGG)	1(7)	1(7)
BamHI(GGATCC)	1(747)	1(747)
NotI(GCGGCCGC)	1(738)	1(738)
SacI(GAGCTC)	1(1)	1(1)
SalI(GTCGAC)	1(753)	1(753)

CIS-Acting Elements	Optimized	Original
Splice(GGTAAG)	0	0
Splice(GGTGAT)	0	0
PolyA(AATAAA)	0	1
PolyA(ATTAAA)	0	0
Destabilizing(ATTTA)	0	0
PolyT(TTTTTT)	0	0
PolyA(AAAAAAA)	0	0
Splice(GTAAAA)	0	0
Splice(GTAAGT)	0	0
Splice(GTACGT)	0	0
PolyA(AATGAA)	0	1
PolyA(AATGGA)	0	0
PolyA(TATAAA)	0	0
PolyA(AATAAT)	0	0

4. Remove Repeat Sequences

After Optimization

Max Direct Repeat: Size:15 Distance:3 Frequency:2
Max Inverted Repeat: None
Max Dyad Repeat: None

Before Optimization

Max Direct Repeat: Size:11 Distance:113 Frequency:2
Max Inverted Repeat: None
Max Dyad Repeat: None

6. DNA Alignment (Optimized Region)

```
Optimized     9   ATGGGCATTGTGGGCGGTTCAGATAGCAGAGAGGGGGCTTGGCCTTGGGTTGTGGCATTG
Original      9   ATGGGAATTGTCGGAGGAAGTGACTCCAGAGAAGGAGCCTGGCCTTGGGTCGTTGCTCTG

Optimized    69   TATTTTGACGACCAGCAGGTTTGTGGAGCTTCTTTGGTGAGCCGGGATTGGCTCGTTTCA
Original     69   TATTTCGACGATCAACAGGTCTGCGGAGCTTCTCTGGTGAGCAGGGATTGGCTGGTGTCG

Optimized   129   GCTGCCCACTGCGTCTACGGTCGCAATATGGAGCCATCTAAGTGGAAGGCAGTCCTGGGG
Original    129   GCCGCCCACTGCGTGTACGGGAGAAATATGGAGCCGTCTAAGTGGAAAGCAGTGCTAGGC

Optimized   189   CTTCATATGGCTTCCAACCTGACTTCACCCCAAATTGAAACACGGCTTATCGACCAGATT
Original    189   CTGCATATGGCATCAAATCTGACTTCTCCTCAGATAGAAACTAGGTTGATTGACCAAATT

Optimized   249   GTGATCAATCCTCACTACAACAAGAGAAGGAAAAATAACGATATCGCCATGATGCATCTG
Original    249   GTCATAAACCCACACTACAATAAACGGAGAAAGAACAATGACATTGCCATGATGCATCTT

Optimized   309   GAGATGAAGGTTAATTACACAGACTATATTCAACCAATCTGCCTTCCCGAGGAAAACCAG
Original    309   GAAATGAAAGTGAACTACACAGATTATATACAGCCTATTTGTTTACCAGAAGAAAATCAA

Optimized   369   GTTTTTCCTCCAGGAAGGATTTGTAGCATCGCAGGATGGGGCGCTTTGATCTACCAAGGC
Original    369   GTTTTTCCCCCAGGAAGAATTTGTTCTATTGCTGGCTGGGGGGCACTTATATATCAAGGT

Optimized   429   TCCACCGCTGATGTGCTCCAGGAGGCCGACGTTCCTCTGCTTTCTAATGAGAAATGCCAG
Original    429   TCTACTGCAGACGTACTGCAAGAAGCTGACGTTCCCCTTCTATCAAATGAGAAATGTCAA

Optimized   489   CAACAGATGCCAGAGTACAATATCACTGAAAACATGGTCTGTGCCGGATATGAGGCAGGA
Original    489   CAACAGATGCCAGAATATAACATTACGGAAAATATGGTGTGTGCAGGCTATGAAGCAGGA

Optimized   549   GGCGTGGATAGTTGCCAAGGCGACTCTGGTGGGCCCTTGATGTGTCAGGAAAATAACAGA
Original    549   GGGGTAGATTCTTGTCAGGGGGATTCAGGCGGACCACTCATGTGCCAAGAAAACAACAGA

Optimized   609   TGGTTGCTCGCCGGTGTTACTTCATTCGGGTATCAGTGTGCACTGCCCAACAGGCCTGGA
Original    609   TGGCTCCTGGCTGGCGTGACATCATTTGGATATCAATGTGCACTGCCTAATCGCCCAGGG

Optimized   669   GTCTATGCTCGGGTCCCAAGATTCACAGAGTGGATTCAGTCATTCCTTCATCATCATCAT
Original    669   GTGTATGCCCGGGTGCCAAGGTTCACAGAGTGGATACAAAGTTTTCTACATCATCATCAC

Optimized   729   CAT
Original    729   CAT
```

Conclusion

A wide variety of factors regulate and influence gene expression levels, and our OptimumGene™ algorithm takes into consideration as many of them as possible, producing the single gene that can reach the highest possible level of expression.

In this case, the native gene employs tandem rare codons that can reduce the efficiency of translation or even disengage the translational machinery. We changed the codon usage bias in Bos taurus by upgrading the CAI from 0.67 to 0.73 , and in Nicotiana tabacum(Tobacco) by optimizing the CAI from 0.76 to 0.76 . GC content and unfavorable peaks have been optimized to prolong the half-life of the mRNA. The Stem-Loop structures, which impact ribosomal binding and stability of mRNA, were broken. In addition, our optimization process has screened and successfully modified those negative cis-acting sites as listed in the introduction.

```
CLUSTAL 2.1 multiple sequence alignment


PCR_PEK         GAGCTCCCATGGGCATTGTGGGCGGTTCAGATAGCAGAGAGGGGGCTTGGCCTTGGGTTG 60
PCR_NEK         GAGCTCCCATGGGAATTGTCGGAGGAAGTGACTCCAGAGAAGGAGCCTGGCCTTGGGTCG 60
                ************ ***** ** **    **   ****** ** ** ********** *

PCR_PEK         TGGCATTGTATTTTGACGACCAGCAGGTTTGTGGAGCTTCTTTGGTGAGCCGGGATTGGC 120
PCR_NEK         TTGCTCTGTATTTCGACGATCAACAGGTCTGCGGAGCTTCTCTGGTGAGCAGGGATTGGC 120
                * **  ******* ***** ** ***** ** ********* ******** *********

PCR_PEK         TCGTTTCAGCTGCCCACTGCGTCTACGGTCGCAATATGGAGCCATCTAAGTGGAAGGCAG 180
PCR_NEK         TGGTGTCGGCCGCCCACTGCGTGTACGGGAGAAATATGGAGCCGTCTAAGTGGAAAGCAG 180
                * ** ** ** ********** *****  * ********** ********** ****

PCR_PEK         TCCTGGGGCTTCATATGGCTTCCAACCTGACTTCACCCCAAATTGAAACACGGCTTATCG 240
PCR_NEK         TGCTAGGCCTGCATATGGCATCAAATCTGACTTCTCCTCAGATAGAAACTAGGTTGATTG 240
                * ** ** ** ******** ** ** ******** ** ** ** *****  ** * ** *

PCR_PEK         ACCAGATTGTGATCAATCCTCACTACAACAAGAGAAGGAAAAATAACGATATCGCCATGA 300
PCR_NEK         ACCAAATTGTCATAAACCCACACTACAATAAACGGAGAAAGAACAATGACATTGCCATGA 300
                **** ***** ** ** ** ******** **  * ** ** ** ** ** ** *******

PCR_PEK         TGCATCTGGAGATGAAGGTTAATTACACAGACTATATTCAACCAATCTGCCTTCCCGAGG 360
PCR_NEK         TGCATCTTGAAATGAAAGTGAACTACACAGATTATATACAGCCTATTTGTTTACCAGAAG 360
                ******* ** ***** ** ** ******** ***** ** ** ** ** * ** ** *

PCR_PEK         AAAACCAGGTTTTTCCTCCAGGAAGGATTTGTAGCATCGCAGGATGGGGCGCTTTGATCT 420
PCR_NEK         AAAATCAAGTTTTTCCCCCAGGAAGAATTTGTTCTATTGCTGGCTGGGGGGCACTTATAT 420
                **** ** ******** ******** ******   ** ** ** ***** **  * ** *

PCR_PEK         ACCAAGGCTCCACCGCTGATGTGCTCCAGGAGGCCGACGTTCCTCTGCTTTCTAATGAGA 480
PCR_NEK         ATCAAGGTTCTACTGCAGACGTACTGCAAGAAGCTGACGTTCCCCTTCTATCAAATGAGA 480
                * ***** ** ** ** ** ** ** ** ** ** ******** ** ** ** *******

PCR_PEK         AATGCCAGCAACAGATGCCAGAGTACAATATCACTGAAAACATGGTCTGTGCCGGATATG 540
PCR_NEK         AATGTCAACAACAGATGCCAGAATATAACATTACGGAAAATATGGTGTGTGCAGGCTATG 540
                **** ** ************** ** ** ** ** ***** ***** ***** ** ****

PCR_PEK         AGGCAGGAGGCGTGGATAGTTGCCAAGGCGACTCTGGTGGGCCCTTGATGTGTCAGGAAA 600
PCR_NEK         AAGCAGGAGGGGTAGATTCTTGTCAGGGGGATTCAGGCGGACCACTCATGTGCCAAGAAA 600
                * ******** ** ***  *** ** ** ** ** ** ** **  * ***** ** ****

PCR_PEK         ATAACAGATGGTTGCTCGCCGGTGTTACTTCATTCGGGTATCAGTGTGCACTGCCCAACA 660
PCR_NEK         ACAACAGATGGCTCCTGGCTGGCGTGACATCATTTGGATATCAATGTGCACTGCCTAATC 660
                * ********* * ** ** ** ** ** ***** ** ***** ********** **

PCR_PEK         GGCCTGGAGTCTATGCTCGGGTCCCAAGATTCACAGAGTGGATTCAGTCATTCCTTCATC 720
PCR_NEK         GCCCAGGGGTGTATGCCCGGGTGCCAAGGTTCACAGAGTGGATACAAAGTTTTCTACATC 720
                * ** ** ** ***** ***** ***** ************** **    ** ** ****

PCR_PEK         ATCATCATCATCACCAC------GCGGCCGCCGGATCCGTCGAC 758
PCR_NEK         ATCATCACCATCACCACTAGTAGGCGGCCGCCGGATCCGTCGAC 764
                ******* *********      **********************
```

Figure 5-1: PCR amplicons of EK alignment. Alignment shows that the nucleotides sequences of PCR_PEK and PCR_NEK are not the same and the mismatches are due to the plant codon optimisation.

```
CLUSTAL 2.1 multiple sequence alignment


PCR_NEK_1       MGIVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLG 60
1EKB            --IVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLG 58
PCR_PEK_1       MGIVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLG 60
                  ************************************************************


PCR_NEK_1       LHMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPICLPEENQ 120
1EKB            LHMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPICLPEENQ 118
PCR_PEK_1       LHMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPICLPEENQ 120
                ************************************************************


PCR_NEK_1       VFPPGRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAG 180
1EKB            VFPPGRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAG 178
PCR_PEK_1       VFPPGRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAG 180
                ************************************************************


PCR_NEK_1       GVDSCQGDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLHHHH 240
1EKB            GVDSCQGDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLH--- 235
PCR_PEK_1       GVDSCQGDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLHHHH 240
                ************************************************************


PCR_NEK_1       HHHAAASRDP 250
1EKB            ---------- 
PCR_PEK_1       HHHAAAGS-- 248
```

Figure 5-2: Translation of the PCR EK amplicons. Both the PCR variants were translated into their protein sequences and aligned to the reference EK sequence 1EKB

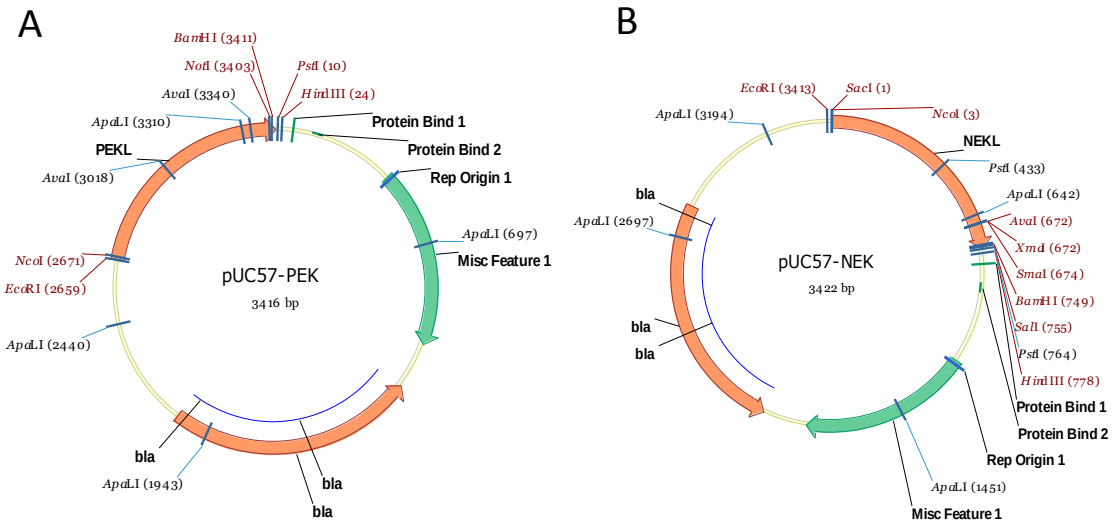


Figure 5-3: Synthesised plasmids containing EK, pUC57-EK. A = pUC57-PEK, B = pUC57-NEK. Protein bind 1 = lac repressor, *lacI*; protein bind 2 = CAP moiety binding site; Rep origin 1 = replication of origin of pMB1; Misc feature 1 = replication of pMB1; bla = gene that confers resistance to kanamycin, encodes beta-galactosidase; PEKL = plant codon optimised enterokinase gene,NEKL = native enterokinase gene.

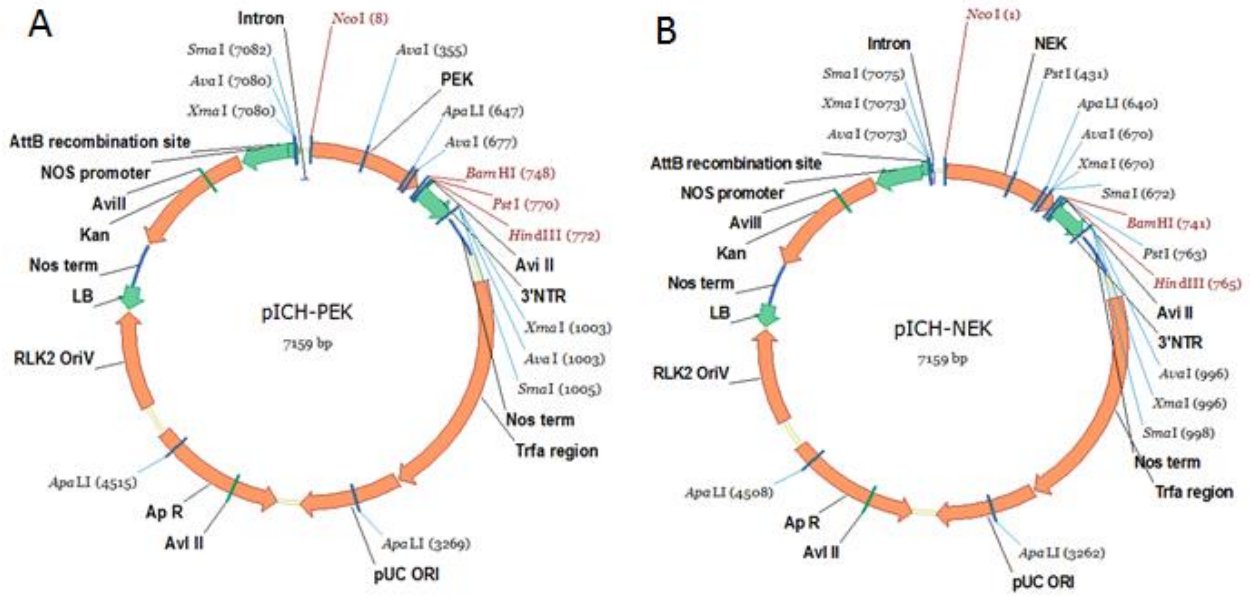


Figure 5-4: EK expression constructs; pICH-PEK (A) and pICH-NEK (B). PEK = plant codon optimised enterokinase gene, NEK = non-codon optimised enterokinase gene, 3' NTR = 3 non-translated region, nos term = nopaline synthase terminator, nos promoter = nopaline synthase promoter, pUC ori = origin of replication derived from pUC plasmid, ap R = amplicillin resuistance gene, LB & RB = binary left and right border of Ti plasmid, kan R = kanamycin resistance gene, Trfa region = plasmid RK2 replication initiator protein that attaches to RLK2 oriV, RLK2 oriV = origin of replication derived from RLK2 plasmid AttB recombination site = PhiC31 integrase recombination site.

Table 5-1: Student's test analysis of expression of recombinant EK from the constructs used in this study.

	a	*b*	*c*	*d*	*bc*	*bd*
Mean	3.10154	16.39661	7.54716	40.87376	11.00036	28.46589
Standard deviation	0.60181	1.60323	0.90051	1.23568	1.03277	0.65886
Standard error	0.30091	0.92563	0.63676	0.87376	0.73028	0.46589
Observations	4	3	2	2	5	2
df	3	2	1	1	4	1
p-value		0.00236	0.0485	0.00669	0.00218	0.00066

	b	*c*	*d*	*bc*	*bd*
Mean	7.54716	40.87376	11.00036	28.46589	18.4134
Standard deviation	0.90051	1.23568	1.03277	0.65886	1.16113
Standard error	0.63676	0.87376	0.73028	0.46589	0.82104
Observations	3	2	2	5	2
df	2	1	1	4	1
p-value		0.00427	0.0005	0.37691	0.00191

	c	*d*	*bc*	*bd*
Mean	40.87376	11.00036	28.46589	18.4134
Standard deviation	1.23568	1.03277	0.65886	1.16113
Standard error	0.87376	0.73028	0.46589	0.82104
Observations	2	2	5	2
df	1	1	4	1
p-value		0.0017	0.01288	0.00222

	d	*bc*	*bd*
Mean	11.00036	28.46589	18.4134
Standard deviation	1.03277	0.65886	1.16113
Standard error	0.73028	0.46589	0.82104
Observations	2	5	2
df	1	4	1
p-value		3.64E-05	0.01603

	bc	*bd*
Mean	28.46589	18.4134
Standard deviation	0.65886	1.16113
Standard error	0.46589	0.82104
Observations	5	2
df	4	1
p-value		0.00083