ISOLATION, PURIFICATION AND KINETIC CHARACTERIZATION OF PROLYL ENDOPEPTIDASE FROM *TRITICUM AESTIVUM* (WHEAT)

A Dissertation

Submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE (MSc) BIOCHEMISTRY

DEPARTMENT OF BIOCHEMISTRY AND MICROBIOLOGY

UNIVERSITY OF FORT HARE

By

ADRIAN MARK ABRAHAMS

SUPERVISOR: PROF GRAEME BRADLEY

JANUARY 2013

DECLARATION

I, hereby, declare that this dissertation submitted to the University of Fort Hare for the degree of Masters of Science in Biochemistry in the faculty of Science and Agriculture and the work contained herein is my original work with exemption to the citations and that this work has not been submitted at any other University in partial or entirety for the award of any degree.

DEDICATION

This dissertation is dedicated to the loving memory of my father, the late Andrew (Billy) Abrahams as well as to my grandmother, the late Lydia (mama Toeksie) Strydom. I suppose growing up wasn't all that easy but thank you for nurturing me and giving me the love I needed during your time here on earth throughout the years. You both knew I would get this far in life and continue to strive for even better in life. It is all through God's time as it was through God's time for you my father and you my grandmother to pass on to your heavenly home.

ACKNOWLEDGEMENTS

I would like to extend my most profound gratitude to God Almighty, the merciful and the passionate, for providing me the opportunity to step into the excellent world of science. Thank you for granting me the courage and strength throughout this research.

I shall start off by acknowledging GMRDC of the University of Fort Hare and NRF for funding this project.

It is my privilege and honour to express my most humble, deepest appreciation and deep sense of gratitude to my supervisor, Prof. Graeme Bradley, who has the attitude and the substance of a genius by continually and convincingly conveying a spirit of adventure with regards to research. Thank you for your tremendous knowledge, invaluable advice and help throughout my studies. This work would not have been possible without the technical assistance and friendship of Dr Ezekiel Green who contributed imperative and unique perspectives, throughout the research development, analysis and reporting periods. Your advice and feedback has been more helpful than you may realize. I'm also grateful to Lee-Anne Smit; I refer to her as my 'department mother'. You have throughout the years since I've known you provided immeasurable sapient and discerning council and have been a true mentor to me in the true sense of the words.

To all my colleagues, babes, nosphiwo, unathi badela, siziwe, Mr Mazomba, ntsaluba, biochemistry and microbiology honours 2012 group, members of the aemreg group and Mr Leonard Mabinya but most importantly I'm extremely grateful to my best friend and colleague, Ntombekhaya Nqumla. Thank you for your valuable suggestions and for the countless lively discussions we had. You were always willing to interact and never got bored of my frequent visits to you. There was never a dull moment with you around. The sleepless nights we made through in the lab. I appreciate you so much in my life!

I would also like to thank my friends, Mafika Nkosi, Sylvester Jeffries, Jonathan Stuurman, Mike Kurai, Abongile Pekana, Zimasa Dubeni, Nolonwabo Nontongana, Noluvo Maha, Eden Padayachee, Curswan, Chene, Delray, Delroy, Amanda, Nikki, Joy, Garth, Rudy, Pasi, Cedric and my loving netball girls (teddies) for always being there and supporting me with their friendly advice. A special thanks to my cousin Chivonne Strydom and dear friend/Afrikaans teacher, Mellissa Duiker who has always provided me with their immeasurable sapient and discerning council. I miss our lively informal meeting and random

conversations just about everything. Also to my other cousins Chesne, Shenda, Elandre, Mercia, Lucretia, Tella, Tammy Bowles and Weston for always checking up on me and spreading their love in every way possible. I miss our lively informal meetings too and random conversations (gossips. Lol). I would also like to express my special appreciation to Ayanda Bam for his friendship and continually motivating me over the years. Thank you for all those words of encouragement and for your friendship that's growing daily.

Above all, special thanks to my foundation which is my mom and dad (Jackie and Billy) who instilled in me from an early age the desire and skills to charter untamed waters. Mom, you have taught me everything I know by example. You are a phenomenal woman and I love you so much. My siblings Andre and Belisha who I've seen as a father and mother to me and who has taught me so much about life. We never had it easy, but you two have taught me how to make the best out of worst situations. Mom and Dad could not always provide for me but you two took the initiative and your hands were never closed for me. Thank you for teaching me about drive and work ethic.

I don't think there are enough words in this world to describe the other phenomenal woman in my life. Agnes (matie) Du Plessis you are the best gift in the universe to me. From a very young age you made sure there was always food on the table for us. Your sleeping hours that you had to sacrifice up for us, waking up at 4am to prepare our toasted lunch packs, making sure we wore clean clothes everyday are just but a few things. It all goes beyond that. Thank you for all your sacrifices (seen and unseen), support and encouragement in every moment through difficult times and wonderful years. You have been a continual source of support, strength and motivation and for that I am forever grateful. Also to those family members who have accompanied me and supported me so far on this life journey (Tokkie, uncle mike, aunty Marilyn, bowles, Abrahams and Prins families, juffrou Sharmel Williams and daughter Rene, Muller family as well as high school mates especially Sharni Stevens and Gerald brown). I give my sincere gratitude to you all. I couldn't have done it without any of you all.

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

CHAPTER 3: IN-SILICO ANALYSIS OF PROLYL ENDOPEPTIDASES IN *TRITICUM AESTIVUM* (WHEAT)

CHAPTER 4: DETECTION AND OPTIMIZATION OF PEP EXPRESSION DURING GERMINATION OF WHEAT SEEDS

CHAPTER 5: ISOLATION AND PURIFICATION OF PEP

LIST OF TABLES

CHAPTER 5: ISOLATION AND PURIFICATION OF PEP

CHAPTER 6: KINETIC CHARACTERIZATION OF PROLYL ENDOPEPTIDASE

LIST OF ABBREVIATIONS

ABA	Abscisic acid
APS	Ammonium persulphate
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
kDa	Kilo Daltons
mAU	milli absorbance unit
NCBI	National Centre for Biotechnology Information
PEP	Prolyl endopeptidase

pI	Isoelectric point
PMSF	phenylmethanesulfonylfluoride or phenylmethylsulfonyl fluoride
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
TFBs	Transcription Factor Binding Sites
TRIS	Tris Hydroxy Methyl Amino Methane
TRIS-HCl	Tris Hydroxy Methyl Amino Methane Hydrochloric Acid
Z-Gly-Pro-pNa	Carbobenzoxy-Glycine-Proline-paranitroaniline

BRIEF CHAPTER SYNOPSIS

CHAPTER 1:

This chapter discusses a concise but comprehensive literature review of PEP from various sources. The structural and expressional analysis, mechanism of action, substrate specificities are discussed here. The overall rationale and motivation for the present study are also discussed in this chapter followed by the hypothesis and specific objectives which this study seeks to actualize.

CHAPTER 2:

Chapter 2 provides a summary of the current study. This includes methods used, results and interpretation of what our findings mean.

CHAPTER 3:

Discussed here in this chapter is an inclusive comparative analysis of *cis*-acting regulatory elements present within the 5'-untranslated region of PEP genes in rice and Arabidopsis, using available bioinformatics tools. The cis-acting regulatory elements were predicted by scanning 2kbp upstream of the translational start site, using plant CARE and PLACE professional databases. It is assumed that these cis-acting regulatory may serve as target binding sites for functional promoters for the expression and regulation of PEP genes.

CHAPTER 4:

In this chapter we look at the detection and optimization of PEP expression during germination of wheat seeds. Various buffers were used to see which presents the best results.

CHAPTER 5:

Chapter 5 looks at the isolation and purification of PEP through the use of different chromatographic steps as well as different buffers in an attempt to purify PEP to homogeneity**.**

CHAPTER 6:

This chapter is dedicated to kinetically characterizing prolyl endopeptidase. It looks at the physicochemical properties of PEP as well as comparing them to several confirmed PEPs.

CHAPTER 7:

This chapter focusses on the inhibitory potential of class specific inhibitors on pep activity. It looks at the effect of leupeptin, E64, Pepstatin and PMSF on PEP activity. It also looks at the type of inhibition involved and to what class of enzymes PEP belongs.

CHAPTER 8:

Chapter 8 gives a general discussion and briefly outlines future research that should be carried out to provide further insight into the role of PEPs.

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

Prolyl endopeptidase, also known as prolyl oligopeptidase and post-proline cleaving enzyme belongs to the family of proteases and has been found to cleave peptides at internal proline residues (Polgar, 2002). According to all structural information in higher eukaryotes, PEP is considered to be a soluble cytoplasmic protein. No reports though have been documented concerning protein trafficking, protein modification or processing which would lead to its association with any other structures (O'Leary et al., 1996). Isolation, structure, function and how PEPs are currently being evaluated as potential drug targets for the treatment of celiac sprue will be reviewed here.

1.1.1Source

The first reported activity of PEP was done in 1971 as an oxytocin cleaving enzyme in the human uterus (Walter et al., 1971). Since this discovery, the enzyme has been isolated from a variety of sources ranging from archae (Harris et al., 2001) to eukaryotes (Venalainen et al., 2004). In resolving its structure, Fulop and colleagues (1998) showed a two domain structure with a unique β-propeller domain.

PEP activity has been described in several alternative locations and has mostly been linked to a variety of neurological disorders such as schizophrenia, amnaesia, depression as well as other disease states such as anorexia nervosa, bulimia nervosa and blood pressure regulation to mention a few (Polgar, 2002).

The enzyme has been found to be highly conserved in mammals and is broadly distributed in the body with high concentrations found in the brain (Goosens et al., 1996). Studies have also detected PEP activity in Alzheimer's disease as well as in Parkinson's disease. However, in Alzheimer's disease cases, conflicting findings were observed in PEP activity from that of the controlled state. Aoyagi and collaborators (1990) have reported enhanced activity of PEP compared with the controlled group. On the contrary, other findings suggest that PEP activity is decreased in specific regions of the brain of Alzheimer's disease patients compared to controls (Gibson et al., 1991; Ichai et al., 1994; Terwel et al., 1998 and Laitenen et al., 2001). Decreasing amounts of PEP was observed in the lung, liver, spleen, heart and kidney from that found in the brain (Yoshimoto et al., 1979; Taylor and Dixon., 1980).

Several studies have also detected PEP in bacteria (Juhasz et al., 2006; Oyama et al., 1997; Szwajcer-Dey et al., 1992) and archae (Harris et al., 2001) with very little work being done in plants (Yoshimoto et al., 1987).

1.1.2 Isolation of PEP

Literature reveals that purification procedures of PEP were developed from that reported by Yoshimoto et al.,(1988) with a few modifications amongst each of the organisms. These include a series of column chromatographies to purify the enzyme to homogeneity.

PEP has been isolated from several eukaryotic and prokaryotic sources. For instance, in isolating the enzyme from bacterium species (e.g pseudomona species), soil samples were collected and bacterial colonies were isolated and purified by single colony isolation procedures on YM agar medium. The isolated microorganisms were screened for their ability to produce prolyl endopeptidase by an enzyme assay and it was based on the increase in colour intensity due to the enzymatically liberated p-nitroaniline of a PEP substrate. The selected bacterium was then allowed to grow under certain conditions and later cultivated. After cultivation, the cells were harvested by centrifugation, cells resuspended in extraction buffer and then disrupted continuously with glass beads (Hiroshi et al., 1997). In other bacterium species, slight modifications were observed in isolating the enzyme especially with the extraction procedure. In species such as *Pyroccocus furiosis* for example, the gene coding for Pfu PEP in a pET11d vector (Harris et al., 2001) was subcloned into a pET3d expression vector using restriction sites flanking the coding region of the gene.

1.1.3 Physical characterization of PEP

Results obtained with respect to gene expression and distribution of the PEP gene from mice shows that the cytoplasmic PEP gene, Prep, maps to chromosome 10B2-B3 which spans a region of about 92kb in size and contains 15 exons. The catalytic triad of the enzyme is coded by exons 1-3 and 10-15 and that of the propeller domain by exons 3-10 (Kimura et al., 1999). Yoshimoto and colleagues (1988) have found PEP cDNA to be sequenced from a large variety of organisms ranging from bacteria, protozoa and animals. However, mammalian PEP has been widely studied from its various sources but more interest in the bacterial version has been pursued as a potential therapeutic agent. Despite the close sequence similarity between the mammalian and bacterial PEP, they however, do not share similar physiological roles.

The mammalian enzymes for which the sequence is available are 710 amino acids long (approximately 80kDa), globular and cytoplasmic (Venalainen et al., 2004) but some mature bacterial PEP are membrane bound and active extracellular (Xie and Sun., 2004). Harwood and colleagues (1997) have found that bacterial enzymes are able to digest peptides that are a bit longer as well as proteins and some are capable of auto proteolysis in contrast to PEP from mammals for which no peptide substrate longer than 30-mer has been found.

The first PEP crystal structure which had been resolved was the porcine PEP (Fig 1.1) (Fulop et al., 1998) and thereafter advancements have been made in resolving two homologous bacterial PEP from *Myxococcus Xanthus* and *Sphingomonas capsulate* (Shan et al., 2005). The approximately 75kDa protein in size has been purified from several eukaryotic as well as prokaryotic sources. It has also been crystallized and its structure resolved to high resolution (Fulop et al., 1998). PEP has been found to contain a cylindrical shape with a height of 60 Angstroms and a 50 Angstrom diameter as well as it consists of two domains: The peptidase domain (fig 1.1) which is formed by the N- and C-termini containing the catalytic triad (Ser, Asp and His). The other domain is the seven bladed β-propeller domain (fig 1.1) which is radially arranged around the central tunnel embedded within the cylinder where the narrow active site is located (Fulop et al., 1998). With the active site located in this large cavity at the interface of the two domains, it has been proposed that there is a narrow hole at the centre of the β-propeller through which the substrate enters. However, the rigid crystal structure of the β-propeller does not give sufficient evidence as to how the active site is approached by the substrate (Fulop et al., 1998). However, alternative routes for the substrate were proposed on resolving the 3D structure and kinetic analysis of mutated variants of PEP (Szeltner et al., 2004), along with molecular dynamic studies (Fuxreiter et al., 2005). Their studies showed that the substrate induces an opening at the interface of the peptidase and the β-propeller domains while entering into the active site and that rigorous movement of the propeller and peptidase domains are required for enzyme action. This mechanism was verified when the structure of *Sphingomonas capsulata* PEP in the open configuration was resolved (Shan et al., 2005)

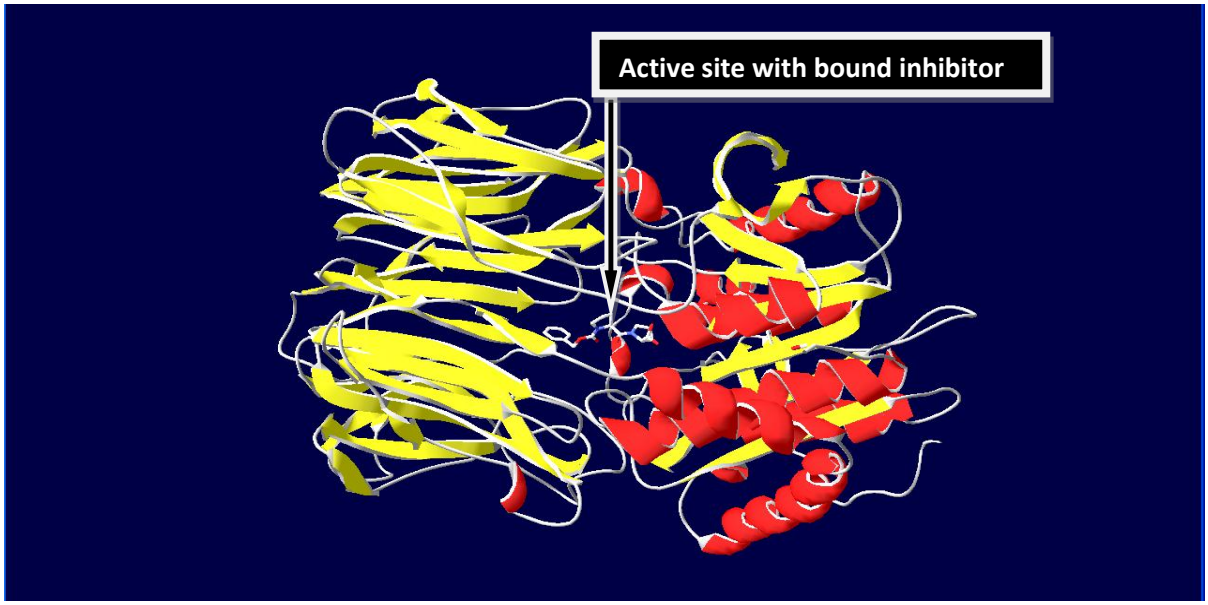


Fig 1.1: The structure of prolyl oligopeptidase with the inhibitor Z-Pro-Prolinal bound at the active site (Adapted from the Protein Data Bank code 1qfs). The seven bladed β-propeller domain (in yellow) is radially arranged around the narrow active site. The peptidase domain which is formed by the N- and C-termini (where the inhibitor is bound) containing the catalytic triad (Ser, Asp and His). The peptidase or catalytic domain is built up of residues in the region 1–72 and 428–710, and the residues between these two portions constitute the propeller domain.

1.1.4 Kinetic characterization of PEP

The catalytic mechanism of PEP can be seen as that of the general protease reaction mechanism via base-assisted catalysis by histidine which occurs through a tetrahedral intermediate, thus generating a covalent acyl enzyme complex that is subsequently hydrolysed. The hydrolysis also involves a second tetrahedral intermediate which breaks down via acid-assisted catalysis by histidine (Leung et al., 2000). PEP cleaves peptides at the C-terminal of an internal proline and will not cleave N-blocked peptides where the proline is the second amino acid (Rosenblum and Kozarich, 2003). The enzyme interacts maximally with amino acid residues of the substrate peptide: those in position P2, P3 and P4 from the N-terminal and those in positions P1’ and P2’ from the C-terminal of the proline which occupies the P1 position (Fulop et al., 2001). Cunningham and O’Connor (1997) as well as Rosenblum and Kozarich (2003) found that the highest reaction rates are obtainable when the P1’ site is a hydrophobic residue.

PEP has also been found to hydrolyse Ala-X bonds but at much more reduced reaction rates (Polgar, 1992). It has recently been reported that a bacterial PEP is able to cleave Val-X bonds at an even more reduced rate than the Ala-X bonds, but the physiological relevance of this fact is claimed to be negligible (Leprince et al., 2006). With regards to PEP catalytic efficiency (k_{cat}/K_m), it was reported to be in the same order of magnitude as for most standard substrates but the K_m is 2–3 orders of magnitude higher than the Pro-containing counterparts.

Inhibition studies using peptidase class specific inhibitors have revealed that PEPs belong to the serine peptidase family (Gass and Khosla., 2007). This is supported by the global protein alignment of PEP protein sequences from bacteria, mammals, plants and insects (fig 1.2) which show a conserved serine in the active site together with conserved Asp and Hist residues typical of the serine peptidase family. Vast majority of reported PEP inhibitors are substrate-like compounds. Most inhibitors that have been designed contain a proline or proline analogue residue due to the strong preference of PEP to cleave peptides specifically at the C-terminal of a proline residue. The most studied PEP inhibitors are Z-Pro-Prolinal, ONO-1603, SUAM-1221, JPT-4819 and S 17092 (Wilk et al., 1983; Fulop et al., 1998). Inhibition by compounds with electronegative substituents has shown that the addition of aldehyde (Wallen et al., 2002), hydroxyacetyl and cyano group (Bakker et al., 1990) at the P1 site results in slow-binding type of inhibition.

```
|OsPEPChr4|      ------------------------------------------------------------
|S.bPEP|          ------------------------------------------------------------
|ZmPEP1|          ------------------------MLLLHKSVPLRLLLPRRPLKTFLLPPPSRPRCLSLL 36
|ZmPEP2|          ------------------------------------------------------------
|OsPEPChr1|      ------------------------------------------------------------
|AtPEP1|          MLTAFASHARSHVFAFVTVPTITRRLRINILRQSPLSSSLLLNETFSNRPNSVSRRCYCS 60
|AtPEP2|          ---------------------------------------SLLLNETFSNRPNSVSRRCYCS 22
|AlPEP2|          MLTAFASHARSHVFAFVTVPTITRRLRINTLRQSSLSSFLLLNETFSNRPNSVSRCFCCS 60
|AlPEP1|          ------------------------------------------------------------
|HsPEP|           ------------------------------------------------------------
|DmPEP|           ------------------------------------------------------------
|DsPEP|           -----------------------------MHTIFRGPLLYRRICQIIVPKALPRNYIHR 30
T.cPEP|           ------------------------------------------------------------NOT 3
|T.oPEP|          ------------------------------------------------------------
|P.fPEP|          ------------------------------------------------------------
|N.mPEP|          ------------------------------------------------------------
```

```
|OsPEPChr4|      ----MGSLPAADKPLPALRYPPARRDDDIVDDYHGVTVPDPYRWMEELESEEVKGFVDAQ 56
|S.bPEP|         ----MGSLATADEPLPPLKYPQPRRDDDIVDDYHGVLVPDPYRWMEELDSKEVKEFVDAQ 56
|ZmPEP1|         PRAAMGSVAGDT--AARLAYPPARRDDSVVDDYHGVRIPDPYRWLEDPDSEETKEFVARQ 94
|ZmPEP2|         ----MGSVAGDT--AARLAYPPARRDDSVVDDYHGVRIPDPYRWLEDPDSEETKEFVARQ 54
|OsPEPChr1|      ----MGSVAGD---AARLSYPPTRRDNSVVDMYHGVPVADPYRWLEDPESEDTKEFVASQ 53
|AtPEP1|         SSAIMGSSSVF---GEQLQYPATRRDDSVVDDYHGVKIGDPYRWLEDPDAEEVKEFVQSQ 117
|AtPEP2|         SSAIMGSSSVF---GEQLQYPATRRDDSVVDDYHGVKIGDPYRWLEDPDAEEVKEFVQSQ 79
|AlPEP2|         SSAIMGSSSVL---GERLHYPATRRGDSVVDNYHGVKIGDPYRWLEDPDAEEVKEFVQNQ 117
|AlPEP1|         ----MGSLSAF---EERLQYPTAKRDESVVEDYHGVNVSDPYRWLEDPDAEEVKEFVEKQ 53
|HsPEP|          --------------MLSLQYPDVYRDETAVQDYHGHKICDPYAWLEDPDSEQTKAFVEAQ 46
|DmPEP|          -MATNVQKSS-KPIVEKFVYPVVRKDLSIVDDFHGTKIIDAYRWLEDPDSPETQAYVDSQ 58
|DsPEP|          ELPSNLTRAMPSTLSAKLEYPVARKDESVAEDFHGTQIKDVYRWLEDPDSTETEEFVNAQ 90
T.cPEP|          SECRETEDPROTEINMSFKYPDARRDETVKDNYFGTEITDPYRWLEDPDSEETKKYVDGQ 63
|T.oPEP|         -------------------------------------MDDPYMWMEDLQDERVLKLVEEE 23
|P.fPEP|         -------------------------------------MEDPYIWMENLEDERVLKIIEEE 23
|N.mPEP|         ----------------------------------MKSYPDPYRHFENLDSAETQNFAAEA 26
                                                     * *  :*: :   .
|OsPEPChr4|      AAVAEAVLSTCDDHRVRLRGQLTALFDHPRYRAPFKRAGSYFYLHNPGLQPHSALYVQHG 116
|S.bPEP|         SAVASAVLSTCD-HRGRLRGQLTALFDHPRFRAPFKRRGSYFYFHNPGLRPHSALYVQRG 115
|ZmPEP1|         AELAETVLAGCP-DRENLRREVTRLFDHPRHAAPFRRGNKYFYFHNSGLQAQSVLYMLDD 153
|ZmPEP2|         AELAETVLAGCP-DRENLRREVTRLFDHPRHAAPFRRGNKYFYFHNSGLQAQSVLYMLDD 113
|OsPEPChr1|      VELAESVLAGCF-DRENLRREVTRLFDHPRHGAPFRRGNKYFYFHNSGLQAQSVLYVQDS 112
|AtPEP1|         VKLTDSVLEKCE-TKEKLRQNITKLIDHPRYDSPFRQGDKYFYFHNTGLQAQSVLYMQDN 176
|AtPEP2|         VKLTDSVLEKCE-TKEKLRQNITKLIDHPRYDSPFRQGDKYFYFHNTGLQAQSVLYMQDN 138
|AlPEP2|         VKLTDSVLEKCE-TKEKLRQNITKLIDHPRYDSPFRQGDKYFYFHNTGLQAQSVLYMQDD 176
|AlPEP1|         VKLSESVLKTCE-TKEKLHDKFTEFIDYPRYDTPFKRGNSYFYFHNSGLQAQSVLHVQDD 112
|HsPEP|          NKITVPFLEQCP-IRGLYKERMTELYDYPKYSCHFKKGKRYFYFYNTGLQNQRVLYVQDS 105
|DmPEP|          NNISQPFLESCP-EWKKINNKLTKLWNYPKYGCPMKYGDYYYYFMNTGLQNQSVMYQQAS 117
|DsPEP|          NSISRPFLENGE-EWKKLNTKLTKLWNYPKYGCPMRYGNYYYYFMNTGLQNQSVMYQQKS 149
T.cPEP|          NAVTRPYLDGCS-FKESIKKKITQLWNYPKFSTPYRHGTKYYQYRNTGLQNQSVIYVQKD 122
|T.oPEP|         NKRFREFIGELS---DELFPEVWEYYSMPTLYGAKLTEKGIIVMFKERDRQ-----IIKW 75
|P.fPEP|         NKRFREFIGELS---DKLFPEVWEQFSQPTIGMARITKKGIIASYSEKDRV-----VIKW 75
|N.mPEP|         NAETRARFLNNDKARALSDGILAQLQDTRQIPFCQEHRARMYHFHQDAEYPKGVYRVCTA 86
                       :             .    .                .
|OsPEPChr4|      LGGG---EEPDVLLDPNTFSDDA--TVSLAMF-GVSHDGEHLAYGTSASGSDWVTIRVMR 170
|S.bPEP|         LGG----AEPDVLLDPNTFSDDG--TVSLGMV-GVSDDGEHLAYGTSASGSDWVTIRVMR 168
|ZmPEP1|         LDG-----KAEVLLDPNTLSKDG--TVALSTY-SISEDGNYIAYGLSESGSDWVSIHVMS 205
|ZmPEP2|         LDG-----KAEVLLDPNTLSKDG--TVALSTY-SISEDGNYIAYGLSESGSDWVSIHVMS 165
|OsPEPChr1|      LDG-----EAEVLLDPNALSKDG--TVALSTY-SVSKDGKYIAYGLSESGSDWVTIRVMN 164
|AtPEP1|         LDA-----EPEVLLDPNTLSDDG--TVALNTF-SVSEDAKYLAYGLSSSGSDWVTIKLMK 228
|AtPEP2|         LDA-----EPEVLLDPNTLSDDG--TVALNTF-SVSEDAKYLAYGLSSSGSDWVTIKLMK 190
|AlPEP2|         LDA-----EPEVLLDPNTLSDDG--TVALNTF-SVSEDAKYLAYGLSSSGSDWVTIKLMN 228
|AlPEP1|         LES-----EAEVLLDPNTLSDDG--TVSLNTL-SISEDAKYLAYGLSSSGSDWVTIKVMK 164
|HsPEP|          LEG-----EARVFLDPNILSDDG--TVALRGY-AFSEDGEYFAYGLSASGSDWVTIKFMK 157
|DmPEP|          LN-----GESKVFLDPNSLSEDG--TIALSQK-SFSDDGKFMAYGLSESGSDWIKIRIRN 169
|DsPEP|          LGDE---SESKVFLDPNTLSEDG--TIALTQK-AFSEDGKYMAYGLSESGSDWIKILIRD 203
T.cPEP|          LA-----SKAEIFLDPNTFSEDG--TVALSGT-AFSEDGQTFAYGLSSSGSDWLEIKFKD 174
|T.oPEP|         LGG-------KILVDSKRLEEELGDEVLLQGF-TADDAGKRLAYSFSIGGADEGITRIID 127
|P.fPEP|         FNG-------DVIVDSKELEREVGDEVLLQGF-TTDEEGEKLAYSFSIGGADEGITRIID 127
|N.mPEP|         ATYRSGYPEWKILFSVADFDELLGDDVYLGGVSHLVEQPNRALLTLSKSGGDTAYTLEVD 146
                          ::..   :.      : *       .  :      * .*.*
```

```
|OsPEPChr4|      VRDRRHLHDEICWVK-FSAIAWTRDGKGFFYSRFPAPKND-GAPLGAGIKTSVNLNHEVY 228
|S.bPEP|         VRDKEHLPDTMSWVK-FSRIAWTRDGLGFFYSRFPAPRGDVGGEMDSGIKTDVNLNHEVY 227
|ZmPEP1|         ITNKQPMPDKLSWVK-FSSISWTHDGKGFFYGRYPAP--RGGE-VDAGTETNINLNHQIY 261
|ZmPEP2|         ITNKQPMPDKLSWVK-FSSISWTHDGKGFFYGRYPAP--RGGE-VDAGTETNINLNHQIY 221
|OsPEPChr1|      IADKQTLSDKLSWVK-FSSISWTHDGKGFFYGRYPAP--REVE-LDAGTETNINLNHEIY 220
|AtPEP1|         IEDKKVEPDTLSWVK-FTGITWTHDSKGFFYGRYPAP--KEGEDIDAGTETNSNLYHELY 285
|AtPEP2|         IEDKKVEPDTLSWVK-FTGITWTHDSKGFFYGRYPAP--KEGEDIDAGTETNSNLYHELY 247
|AlPEP2|         IEDKKVEPDVLSWVK-FTGITWTHDSKGFFYGRYPAP--KEGEDIDAGTETNSNLYHELY 285
|AlPEP1|         IEDKKVESDSLSWVK-FSGITWTHDGKGFFYSRYPAP--QEGEKIDAGTETNSNLYHELY 221
|HsPEP|          VDGAKELPDVLERVK-FSCMAWTHDGKGMFYNSYPQQ-----DGKSDGTETSTNLHQKLY 211
|DmPEP|          AETGKDFDEVLEKVK-FSEISWTKDNKGFFYGRYPDQ-----DGKTDGSETQQNQNQKLY 223
|DsPEP|          AETGKDLSEVLEKVK-FSEISWTKDNKGFFYGRYPDQ-----DGKTDGSETKQNENQKLY 257
T.cPEP|          VETGKDYKEILKKVK-FSPMTWMHDNKGFFYGAYLDQ-----TGKADGSETKTNENQKLY 228
|T.oPEP|         LETG----ELIDEMK-PSVWNVAFLENGYYFGRFYRH------------DETPDGVKAPA 170
|P.fPEP|         LKTG----EVIEEIK-PSIWNITFLKDGYYFTRFYRK------------EKTPDGVNPPA 170
|N.mPEP|         LEAGELVEGGFHFPAGKNHVSWRDENSVWVCPAWDER-----------QLTESGYPREVW 195
                          :      .              :                 .  .
|OsPEPChr4|      YHFLGTDQSEDLLCWEDPDHPKYIYTPEVSEDGKYVILSVAETSEPVNKLYYCDLSALPD 288
|S.bPEP|         YHFLGTEQSQDVLCWRDPDHPKYIYIPEVTEDGKYVILSVSETSEPVNKLYYCDLSALAH 287
|ZmPEP1|         YHVLGSDQSEDILCWKDPEHPKYSFGASVTEDGKYIILGIYEGCDPVNKLYYCEISSLPQ 321
|ZmPEP2|         YHVLGSDQSEDILCWKDPEHPKYSFGASVTEDGKYIILGIYEGCDPVNKLYYCEISSLPQ 281
|OsPEPChr1|      YHVVGSDQSEDILCWKDPEHPKYSFGASVTEDGKYIILGTYEGCDPVNKLYYCEICTLPQ 280
|AtPEP1|         YHFIGTDQSQDILCWRDNENPKYMFGAEVTDDGKYLIMSIGESCDPVNKLYYCDMTSLSG 345
|AtPEP2|         YHFIGTDQSQDILCWRDNENPKYMFGAEVTDDGKYLIMSIGESCDPVNKLYYCDMTSLSG 307
|AlPEP2|         YHFLGTDQSQDILCWRDHENPKYMFGAEVTDDGKYLFMSIGESCDPVNKLYYCDMTSFSG 345
|AlPEP1|         YHFLGMDQSEDVLCWRDQDNPKHMFGSKVTDDGKYLIMSIEEGCDPVNKVYHCDLSSLPK 281
|HsPEP|          YHVLGTDQSEDILCAEFPDEPKWMGGAELSDDGRYVLLSIREGCDPVNRLWYCDLQQESS 271
|DmPEP|          YHRIGESQAKDVLVVQFNNEPSWRFQSVVSDCGKYLVLAIVKDCR-DNIIYYADLKPGEE 282
|DsPEP|          YHRVGESQDKDTLVVEFPEEPSWRIQSTVSDCGKYLILAIVKDCR-DNIVFYADLTPGAE 316
T.cPEP|          YHELGTDQSQDVVVVEFD-DPHLRIGAHVSHCGKYLVITGTKGCK-NNLLYFAQLDSG-K 285
|T.oPEP|         VRLFWKDEGGEKMVFGEGLSSGYFIGLRESTDGKWAMVVVTFGWN-KAEIYLGPIEEPEK 229
|P.fPEP|         ARMFWKDREGERMVFGEGLTSGYFMSIRKSSDGKFAIVTLTYGWN-QGEVYIGPIDNPQE 229
|N.mPEP|         LVERGKSFEESLPVYQIAEDGMMVNAWRYLDPQGSPIDLIEASDGFYTKTYLQVSAEGEA 255
                       .   .                          .            :
|OsPEPChr4|      GLEGMKGN--HGNAMLPFVKLVDEFEAYYALIANDDTQFTFLTNKNAPKYKLSRIDVN-- 344
|S.bPEP|         GLGGITKQSTHGNGMLPFVKLVDKFEAYYGLIANDGTEFTFLTNKDAPKYKLARVDVDDD 347
|ZmPEP1|         GIEGFRGT----QDLLPFVKLIDNFDAQYQVVANDGDEFTFLTNKSALKNKLVRVNIK-- 375
|ZmPEP2|         GIEGFRGT----QDLLPFVKLIDNFDAQYQVVANDGDEFTFLTNKSAPKNKLVRVNIK-- 335
|OsPEPChr1|      GIEGFKET----KGMLPFVKLIDNFDAQYHVVANDGDEFTFLTNRNAPKNKLVRVDIK-- 334
|AtPEP1|         GLESFRGS----SSFLPFIKLVDTFDAQYSVISNDETLFTFLTNKDAPKYKLVRVDLK-- 399
|AtPEP2|         GLESFRGS----SSFLPFIKLVDTFDAQYSVISNDETLFTFLTNKDAPKYKLVRVDLK-- 361
|AlPEP2|         GLESFRGS----SSFLPFIKLVDTFDAQYSVISNDETLFTFLTNKDAPKYKLVLVDLK-- 399
|AlPEP1|         GLEGFRGS----NALLPFVKLIDTFDAQYIAIANDATLFTFLTNKDAPKYKLVRVDLK-- 335
|HsPEP|          GIA----------GILKWVKLIDNFEGEYDYVTNEGTVFTFKTNRQSPNYRVINIDFRDP 321
|DmPEP|          IKSELDVK-----------IIVDKFESDYDYVTNVGSKVYFRTNKNAPNYRVIIIDFENP 331
|DsPEP|          ITSKLNVK-----------KIVEKFEADYDYITNEGSKIFFRTNKNAPNYQVIAIDFNNP 365
T.cPEP|          ITGKLKLT-----------EVVTEFVADFEYITNDKNLFYFHTNKNASNYRIVIIDFDNP 334
|T.oPEP|         WEKVYSAE-----------------VPAQPIEVVDGKVYILTKEGKGLGKVIAIEDG-- 269
|P.fPEP|         WKKVYSAS-----------------VPVEAIDVVNGKLYILTKEGKGLGKIIAIKNG-- 269
|N.mPEP|         KPLNLPND-------------CDVVGYLAGHLLLTLRKDWHRANQSYPSGALVAVKLNRG 302
                                         :          ::.      :  :.
```

```
|OsPEPChr4|      EPHSWMDILPEDEKAVLESACAVHGDKLLVNYLSDVKYVLQMRSLVTGELLHDIP--IDI 402
|S.bPEP|         APGSWTDVVPEDDKAVLESACAVHGHKLLVNYLSDVKYVLQMRSLVTGEFLRDIP--IDI 405
|ZmPEP1|         NPELWTDVLSEHEKDVLESADAVNNNQLLVNYMSDVKHILQIRDLRTGNFIHQLP--LEI 433
|ZmPEP2|         NPELWTDVLSEHEKDVLESADAVNNNQLLVNYMSDVKHILQIRDLRTGNFIHQLP--LEI 393
|OsPEPChr1|      KPELWTDILPEHERDVLESADAVNGNQLLVCYMSDVKHILQIRDLVTGNLLHKLP--LEI 392
|AtPEP1|         EPNSWTDVVEEHEKDVLASACAVNGNHLVACYMSDVKHILQIRDLKSGSLLHQLP--LDI 457
|AtPEP2|         EPNSWTDVVEEHEKDVLASACAVNGNHLVACYMSDVKHILQIRDLKSGSLLHQLP--LDI 419
|AlPEP2|         EPNSWTDVVEEHEKDVLASACAVNGNHLVACYMSDVKHILQIRDLKSGSLLHQLP--LDI 457
|AlPEP1|         EPSSWTDVIAEHEKDVLSTASAVNGDQLVVSYMSDVKHILQIRDLKSGSLLHRLP--VDI 393
|HsPEP|          EESKWKVLVPEHEKDVLEWIACVRSNFLVLCYLHDVKNILQLHDLTTGALLKTFP--LDV 379
|DmPEP|          AEENWETLIPEHEKDVLDWVKCVNEDKILLCYIRDVKTILQANSLRTGELIRQFD--LDI 389
|DsPEP|          AEDKWETLIAEHKSDVLDWVKCVDADKLLVCYIRDVKSVLQVNSLKDGTLLREFD--LDI 423
T.cPEP|          KESEWKDLISEHPKDVLDWAHAINENMLVVCYLQDVKNIMQLYDIKSSNKLHDFK--LDV 392
|T.oPEP|         ---EIEEVIPEDE-FPLEWAVIVG-NRILAGRLVHASHRLEVYSLD-GKKLDEIT--FDL 321
|P.fPEP|         ---KIDEVIPEGE-FPLEWAVIVR-DKILAGRLVHASYKLEVYTLN-GEKIKEIT--FDV 321
|N.mPEP|         ELGAAQLLFAPNETQALESVETTK-RFVVASLLENVQGRLKAWRFTDGKWQETELPRLPS 361
                       :.       *          ::   : ...  ::   :  .         .

|OsPEPChr4|      GSVNGISGRRDNSEVFIEFASFLTPGIIYRCDVSKETPEMNIYREISVG--GFDRTDFEA 460
|S.bPEP|         GTVNGISGRRVDSEVFIEFASFLTPGIIYRCDVSAEIPEMDVYREISVP--GFDRNEFEA 463
|ZmPEP1|         GSVSEISCRREDKEVFIGFTSFLSPGIIFRCNLASTIPEMKMFREISVP--GFDRTSFQV 491
|ZmPEP2|         GSVSEISCRREDKEVFIGFTSFLSPGIIFRCNLASTIPEMKMFREISVP--GFDRTSFQV 451
|OsPEPChr1|      GSVSEISCRREDMDVFIGFTSFLSPGIIYRCNLTSAIPEMKIFREISVP--GFDRTNFEV 450
|AtPEP1|         GSVSDVSARRKDNTFFFSFTSFLTPGVIYKCDLANESPEVKVFREVTVP--GFDREAFQA 515
|AtPEP2|         GSVSDVSARRKDNTFFFSFTSFLTPGVIYKCDLANESPEVKVFREVTVP--GFDREAFQA 477
|AlPEP2|         GSVSDVSARRKDNTFFFSFTSFLTPGVIYKCDLANESPEVKVFREVTVP--GFDREAFQA 515
|AlPEP1|         GSVCGVFARRKDTIFFFRFTSFLTPGVIYKCDLSHEAPEVTVFREIGVP--GFDRTAFQV 451
|HsPEP|          GSIVGYSGQKKDTEIFYQFTSFLSPGIIYHCDLTKEELEPRVFREVTVK--GIDASDYQT 437
|DmPEP|          GTIVGTSGKTKYSEIFYNFSSFLNPGTIYHYDFKYPDQPPKVFRDINLNLDGFNRDNYIV 449
|DsPEP|       GTIVGTSGEKKYSEIFYNFSSFLNPGSIYRYDFKTPEKSPSVFREIKLNLEGFRREDYAV 483
T.cPEP|          GTISAISGKKYHKEMFFSFCSFLTPNIIYKVDFDQGSIKETLFHETKVG--DFESSKYET 450
|T.oPEP|         PGSLYPLD-ADGKKVILRYESFTIPYRLYEFDG-----ELKLLDERKIEG------DFQV 369
|P.fPEP|         PGSLYPLD-KDEERVLLRYTSFTIPYRLYEFKD-----DLRLIEERKVEG------EFRV 369
|N.mPEP|         GALEMTDQPWGGDVVYLAASDFTTPLTLFALDLN--VMELTVMRRQPQQ---FDSDGINV 416
                            .     .*  *  ::  .         :                .

|OsPEPChr4|      KQVFYPSKDGTKIPMFIVSKKSIVLDGSHPTLLYGYGGFGMNMTPHFSVTRIVLMRNLGF 520
|S.bPEP|         KQVFYPSKDGTKIPMFIISKKGMDLDGSHPALLFGYGGFGVSITPQFSVTRVVLMRNLGF 523
|ZmPEP1|         KQVFVPSKDGTKIPMFIMSKKDIDLNGSHPTLLYGYGGFNISITPSFSVGRVVLCKNMGF 551
|ZmPEP2|         KQVFVPSKDGTKIPMFIMSKKDIDLNGSHPTLLYGYGGFNISITPSFSVGRVVLCKNMGF 511
|OsPEPChr1|      KQIFVNSKDGTKIPMFIMSKRDIELDGSHPTLLYGYGGFNISLTPSFSVSRVVLCKNMGF 510
|AtPEP1|         IQVFYPSKDGTKIPMFIVAKKDIKLDGSHPCLLYAYGGFNISITPSFSASRIVLSKHLGV 575
|AtPEP2|         IQVFYPSKDGTKIPMFIVAKKDIKLDGSHPCLLYAYGGFNISITPSFSASRIVLSKHLGV 537
|AlPEP2|         IQVFYPSKDGTKIPMFIVAKKDIKLDGSHPCLLYAYGGFNISITPSFSASRIVLSKHLGV 575
|AlPEP1|         TQVFYPSKDGTEIPMFIVARKDIKLDGSHPCLLYAYGGFSVSMTPFFSATRIVLSRHLGT 511
|HsPEP|          VQIFYPSKDGTKIPMFIVHKKGIKLDGSHPAFLYGYGGFNISITPNYSVSRLIFVRHMGG 497
|DmPEP|          EQIFYSSKDGTKVPMFIIRKK-ADTVQPRPCLLYGYGGFNISMLPTFGLTGLMFVDTFDG 508
|DsPEP|          EQIFYKSKDGTKVPMFIIRKK-RDNVEPRPCLLYGYGGFNISMLPSFGLSGLMFIDTFDG 542
T.cPEP|          KQVFYKSKDGTEIPMFIINKKGLVNDGSKPCLLYGYGGFNVNLTPSFGVSRLVFIENFDG 510
|T.oPEP|         SEDFAISKDGTRIHYFIVKGE----KDEKKAWVFGYGGFNIALTPRFFPQVIPFIR-RGG 424
|P.fPEP|         EEDFATSKDGTKVHYFIVKGE----RDEKRAWVFGYGGFNIALTPMFFPQVIPFLK-RGG 424
|N.mPEP|         QQFWTTSADGERIPYFHVGKN---AAPDMPTLVYAYGGFGIPELPHYLGSIGKYWLEEGN 473
                 : :  * ** .:  * :  .           ::.****.:   * :           .
```

```
|OsPEPChr4|      VSCIANIRGGGEYGEDWHKAGSLANKQNCFDDFIAAGEFLVSAGYTNPSRLCVEGASNGG 580
|S.bPEP|         VTCVANIRGGGEYGEDWHRAGSLANKQNCFDDFIAAGEFLVSAGYTSPARLCIEGGSNGG 583
|ZmPEP1|         VVCVANIRGGGEYGEEWHKAGALAMKQNCFDDFAACAEFLISNGYTSSRRLCIEGGSNGG 611
|ZmPEP2|         VVCLANIRGGGEYGEEWHKAGALAMKQNCFDDFAACAEFLISNGYTSSRRLCIEGGSNGG 571
|OsPEPChr1|      VVCVANIRGGGEYGEEWHKAGARAMKQNCFDDFIACAELLISAGYTSYRQLCIEGGSNGG 570
|AtPEP1|         VFCFANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 635
|AtPEP2|         VFCFANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 597
|AlPEP2|         VFCVANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 635
|AlPEP1|         VFCFANIRGGGEYGEEWHKSGALANKQNCFDDFISGAEYLVSAGYTQPRKLCIEGGSNGG 571
|HsPEP|          ILAVANIRGGGEYGETWHKGGILANKQNCFDDFQCAAEYLIKEGYTSPKRLTINGGSNGG 557
|DmPEP|          VLAYPNIRGGGEYGEKWHNGGRLLNKQNVFDDFQCAAEYLIKNNYTTKDRLAIQGGSNGG 568
|DsPEP|          VLAYPNLRGGGEYGEKWHNGGRLLNKQNVFDDFQTAAEYLIENKYTTKDRLAIQGGSNGG 602
T.cPEP|          VYALANIRGGGEYGDNWHNGGRFGKKQNCFDDFQYAAKYLVENKYTKVDKLIIQGGSNGG 570
|T.oPEP|         TFVMANLRGGSEYGEEWHRAGMRENKQNVFDDFIAVLGKLKAEGYK----VAAWGRSNGG 480
|P.fPEP|         TFIMANLRGGSEYGEEWHRAGMRENKQNVFDDFIAVLEKRKKEGYK----VAAWGRSNGG 480
|N.mPEP|         AFVLANIRGGGEFGPRWHQAAQGISKHKSVDDLLAVVSDLSERGISSPEHIGLQGGSNGG 533
                     .*:***.*:*  **...    *:: .**:            .    :   * ****
|OsPEPChr4|      LLVAACINQRPDLFGCALAHVGVMDMLRFHKFTIGRAWTCDFGCSEKEEEFH-WLIKYSP 639
|S.bPEP|         LLVAACMNQRPDLFGCALAHVGVMDMLRFHKFTIGRAWTCDFGCSDNEEEFH-WLIKYSP 642
|ZmPEP1|         LLVAASINQRPDLFGCALAHVGVMDMLRFHKFTIGHAWTTDYGCSDKEEEFQ-WLIKYSP 670
|ZmPEP2|         LLVAASINQRPDLFGCALAHVGVMDMLRFHKFTIGHAWTTDYGCSDKEEEFQ-WLIKYSP 630
|OsPEPChr1|      LLIAACVNQRPDLFGCALAHVGVMDMLRFHKFTIGHAWTTDYGCSDNEEEFH-WLIKYSP 629
|AtPEP1|         LLVGACINQRPDLYGCALAHVGVMDMLRFHKFTIGHAWTSDYGCSENEEEFH-WLIKYSP 694
|AtPEP2|         LLVGACINQRPDLYGCALAHVGVMDMLRFHKFTIGHAWTSDYGCSENEEEFH-WLIKYSP 656
|AlPEP2|         LLIGACINQ-ANLFGCALAHVGVMDMLRFHKFTIGHAWTSDYGCSENEEEFH-WLIKYSP 693
|AlPEP1|         ILVGACINQRPDLFGCALAHVGVMDMLRFHKFTIGHAWTSEFGCSDKEEEFH-WLIKYSP 630
|HsPEP|          LLVAACANQRPDLFGCVIAQVGVMDMLKFHKYTIGHAWTTDYGCSDSKQHFE-WLVKYSP 616
|DmPEP|          LLVGACINQRPDLFGAAVAQVGVMDMLRFHKFTIGHAWCSDYGNPSEKEHFD-NLYKLSP 627
|DsPEP|          LLVGSCINQRPDLFGAAVAQVGVMDMLRFHKFTIGHAWCSDYGNPSEKEHFD-NLYKFSP 661
T.cPEP|          LLVAACINQAPELFGAAICQVGVLDMLRFHKFTIGYAWKSDYGCSEEQEQFE-YLYKYSP 629
|T.oPEP|         LLVSATLVQRPDVMDAALIGYPVIDMFRFHKLYIGSVWVPEYGNPDDPKDRE-FLLKYSP 539
|P.fPEP|         LLVSATLTQRPDVMDSALIGYPVIDMLRFHKLYIGSVWIPEYGNPEDPKDRE-FLLKYSP 539
|N.mPEP|         LITAAAFVREPQSIGALVCEVPLTDMIRYPLLSAGSSWTDEYGNPQKYEVCKRRLGELSP 593
                 :: .:   : .:  .. :    : **:::     *  *  ::* ... :  .  * : **
|OsPEPChr4|      LHNVRRPWEKG-----HRRQQYPSTMLLTADHDDRVVPSHTLKFLATMQHVLCTSVKE-S 693
|S.bPEP|         LHNVRRPWEKEKWAAATGGGQYPATMLLTADHDDRVVPSHTLKFLATMQHVLRAGAEGGS 702
|ZmPEP1|         LHNVRRPWEQS----SGNNCQYPATMLLTADHDDRVVPLHSLKLLATLQHVLCTSTED-S 725
|ZmPEP2|         LHNVRRPWEQS----SGNNCQYPATMLLTADHDDRVVPLHSLKLLATLQHVLCTSTED-S 685
|OsPEPChr1|      LHNVRRPWEQS----FVNCCQYPAIMLLTADHDDRVVPLHSLKLLATLQYVLCTSIED-T 684
|AtPEP1|         LHNVKRPWEQQ----TDHLVQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLDN-S 749
|AtPEP2|         LHNVKRPWEQQ----TDHLVQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLDN-S 711
|AlPEP2|         LHNVKRPWEQQ----TDQFIQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLEN-S 748
|AlPEP1|         LHNVKRPWEQK----TDRFVQYPSTMLLTADHDDRVVPLHTYKLLATMQYELGLSLEN-S 685
|HsPEP|          LHNVKLPEADD--------IQYPSMLLLTADHDDRVVPLHSLKFIATLQYIVGRSRKQ-- 666
|DmPEP|          LHNVHTPNDASS--------EYPSTLILTADHDDRVSPLHSLKFAAALQEAVRDSPFQ-- 677
|DsPEP|          LHNVHTPKGAET--------EYPSTLILTADHDDRVSPLHSLKFIAALQEAVRDSKFQ-- 711
T.cPEP|          LHNIRVPQNGG---------QYPATLLLTADHDDRVVPLHSLKFIAELQNKIGRLPTQ-- 678
|T.oPEP|         YHNVK-------------EQKYPPTLLYTGLYDDRVHPAHALKFFMKLRAVS-------- 578
|P.fPEP|         YHNVDP------------KKKYPPTLIYTGLHDDRVHPAHALKFFMKLKEIG-------- 579
|N.mPEP|         YHNLSDG------------IDYPPALITTSLSDDRVHPAHALKFYAKLRETS-------- 633
                  **:               .**. :: *.  **** * *: *:   ::
```

```
|OsPEPChr4|      PQTNPIVARIDRKSGHGCGRSTQKIIDEAADRYAFAAKTMGISWID- 739
|S.bPEP|         PQTNPIIARIERNSGHCCGRSTQKIIDEAADRYAFAAKMMGVSWID- 748
|ZmPEP1|         PQTNPIIGRIDRKSGHGAGRPTQKMIDEAADRYSFMAKMLGASWTE- 771
|ZmPEP2|         PQTNPIIGRIDRKSGHGAGRPTQKMIDEAADRYSFMAKMLGASWTE- 731
|OsPEPChr1|      PQVNPIIGRIDVKSGHGAGRPTKKMIDEVADRYSFMANMLDASWTE- 730
|AtPEP1|         PQMNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNASWTE- 795
|AtPEP2|         PQMNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNASWTE- 757
|AlPEP2|         PQTNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNAAWTE- 794
|AlPEP1|         PQTNPIIARIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVDASWID- 731
|HsPEP|          --SNPLLIHVDTKAGHGAGKPTAKVIEEVSDMFAFIARCLNVDWIP- 710
|DmPEP|          --KNPLLLRVYKKAGHGAGKPTSKRIEEATDILTFMYRSLNIDVINL 722
|DsPEP|          --NNPVLLRVYQKAGHGAGKPTSKRIEEATDILTFLSKSLNVDTINV 756
T.cPEP|          --KNPLMIRIETRAGHGAGKPTSKIIEEVTDTFCFISRALNLTFSS- 722
|T.oPEP|         ---APVYLRVETKSGH-MGASPETRARELTDLLAFVVETLEA----- 616
|P.fPEP|         ---APVYLRVETKSGH-MGASPETRARELTDLLAFVLKTLS------ 616
|N.mPEP|         ---PQSWLYSPDGGGHTGNGTQREAADELACVLLFLKEFLG------ 671
                            .**  . .      * :    *  . :
```

Fig1.2: Global alignment of amino acid sequences of Zea Mays: ZmPEP1 (NM 001136920.1) and ZmPEP2 (ACN31271); Rice PEP: OsPEPChr4 (BAF15466) and OsPEPChr1 (BAF03701); Arabidopsis Lyrata: AlPEP (XP_002887627); Arabidopsis thaliana PEP: AtPEP1 (NP_173463) and AtPEP2 (NP_001117606); Drosophila PEP: DsPEP (XP_002036510) and DmPEP (XP_002002881); Sorghum bicolor PEP: SbPEP (XP_002446934); Homo sapiens PEP: HsPEP (NP_002717); Tribolium castaneum PEP: TcPEP (XP_972959.1); Pyrococcus furiosis PEP: *Pf*PEP(AAA73423), Thermococcus Onureus PEP: *To* PEP (ABA26946) and Neisseria meningitidis PEP:*Nm*PEP (YP_002342055). The alignment was performed using clustalW2 (Larkin et al., 2007) with default settings for protein alignments. Dashes indicate gaps and the numbers on the right represent the position of the last residue in the original sequence. Identical residues among the enzymes are marked as an asterisk and residues with conserved substitutions and semi conserved substitutions are marked as a colon and a dot, respectively.

1.2 PEPs involvement in Diseases

As mentioned in section 1.1.1, PEP activity has been found throughout the body, with highest concentrations detected within the brain (Irazusta, 2002). PEP activity has also been linked to a variety of neurological disorders, such as depression, schizophrenia, Alzheimer's disease as well as other diseases such as blood pressure regulation, anorexia and bulimia nervosa (Polgar, 2002).

PEPs have also been found to play a key role on learning and memory in animal models which have been tested for amnesia (Kamei et al., 1992; Portevin et al., 1996; Tsuru et al., 1988; Yoshimoto et al., 1988) as well as in the human uterus as an oxytocin cleaving enzyme since it has the ability of cleaving peptides at the C-terminal of the proline residue (Walter et al., 1971).

1.2.1 PEP activity in Alzheimer's disease

Reports have shown that PEP activity is enhanced in patients suffering from Alzheimer's disease in the intra-cerebral region of the brain (Aoyagi et al., 1990). On the contrary, conflicting results have been found by other groups that state that the activity of PEP is decreased in specific regions of the brain of Alzheimer's disease patients (Gibson et al, 1991). However, in their opinion, the reduction in PEP activity in Alzheimer's disease tissue is not specific to this disorder but a common factor to other neurodegenerative disorders such as Parkinson's disease and Huntington's disease (Mantle et al., 1996).

Changes in serum PEP have also been found as a result of several psychiatric disorders, such as depression, mania, schizophrenia, eating disorders and stress induced anxiety but, no specific mechanisms have been discovered to explain the involvement of PEP in these pathologies. Maes et al.,(1994, 1998) also reported that total serum PEP-like activity was found to be significantly lower in the plasma of patients with depression compared to normal individuals.

Maes et al.,(1995) also found PEP to be related to psychotic conditions such as schizophrenia and manic depression as they have found the enzymes' activity to have increased in mania and schizophrenia patients compared to the normal volunteers. PEP activity was then restored to normal levels in mania and schizophrenic patients that were treated with fluoxetine and valproic anti-depressants.

1.2.2 PEP and Cancer

Research reports with regards to PEPs role in cancer are rather sparse but Sedo et al.,(1991) reported that an increase in PEP activity was found in a number of primary lung tumours as compared with matched lung parenchyma from non-cancerous lung individuals. Zima et al.,(2001) reported a significant reduction though in PEP activity in rats treated with the anthracycline antibiotic doxorubicin, which is effective for the treatment of certain cancers but causes the likelihood of cardiomyopathy to develop.

1.2.3 PEP and inflammation

PEP has also been found to be related to the immunological or inflammatory disturbances in joints in rheumatoid arthritis as a significant increase in PEP activity was reported in the synovial fluids of patients with rheumatoid arthritis (Kamori et al., 1991 and Hashimoto et al., 2001). It has also been reported that PEP activity was reasonably high in the washing fluids of inflammatory foci in the *Mycobacterium tuberculosis* induced delayed-type allergic inflammation (Kakegawa et al., 2004)

Having comprehensively looked at the activity of PEP in some disease states and where they are also involved, the next section of this review will now focus on PEP as a potential treatment for celiac disease. It will also look at evolutionary trends as well as possible isolation and characterization methods of the enzyme.

1.3 PEP as a potential treatment for celiac sprue

Celiac disease is a disease characterized by enteropathy of the small intestine. When people with celiac disease consume gluten containing food products, an immune-mediated toxic reaction occurs that damages the lining of the small intestine called villi. Once these villi are destroyed, the person becomes malnourished (Wolters and Wijmenga., 2008).

Gluten which has been found to be the primary environmental factor, provoke the disease as the high proline and glutamine content is relatively resistant to proteolysis by gastric, pancreatic and intestinal brush border enzymes (Molberg et al., 2005). The undigested gluten peptides are deaminated by tissue transglutaminase, resulting in a better capacity pocket binding of HLA-DQ2 or HLA-DQ8 molecules on antigen presenting cells.

The absence of PEP from mammalian pancreatic proteases and the intestinal brush-border enzymes highlights the lack of a role of PEP activity in the assimilation of dietary proteins in mammals. A study done on 25 Dutch celiac patients revealed no major mutations in the PEP gene. Initially, bacterial PEP was identified as a potential family of enzymes for the oral treatment of celiac sprue. The role has now shifted over to plant-based sources that may possibly also serve as potential agents for the treatment of celiac disease. Plant PEP sources have been considered as gluten-detoxifying agents but uncertainty still lies in the enzyme's activity. It has however been hypothesized that plant PEP activity may be similar to that of other sources.

PEP's from *Flavobacterium meningosepticum*, *Sphingomonas capsulate*, *Thermococcus Aureus* and *Myxococcus Xanthus* show similar activities but these and those from plants exhibit substantial differences with respect to chain length and specificity in the key gluten-derived peptides with important physiological relevance to celiac sprue pathogenesis (Shan et al., 2004).

Research shows that glutamine and proline residues constitute approximately 30% and 20% respectively of gluten and further studies have also shown that even a small amount of gluten in food products causes health problems as well as affecting those already suffering from celiac disease. Therefore, the role of prolyl endopeptidase from plant based sources in treating individuals suffering from celiac disease may represent a significant step in drug development.

Furthermore, prolyl endopeptidase research should be viewed as a promising field in which medical advances are likely to be realized.

1.4 THE PROPOSED INVESTIGATION

Initially, the only accepted nutritional therapy for celiac disease involved the lifelong elimination of wheat, rye and barley from the diet (Lerner, 1996). However, clinical studies have shown that oats are tolerated by most patients suffering from celiac disease and may improve the nutritional content of the diet as well as the overall quality of a patient's life. Furthermore, clinical tests have shown that the elimination of gluten usually induces clinical improvements within days or weeks, though the patient may take months or even years to recover histologically especially adults in whom mucosal recovery may be incomplete (Hutchinson et al, 2008). Gluten-free products vary by country and are particularly expensive and hard to find in developing countries whereas in other countries such as the Netherlands, Sweden, Finland, Italy, United Kingdom and New Zealand, the government subsidizes these products. However, there has been considerable interest and advances in developing non-dietary therapies that may either replace or supplement the gluten-free diet such as the use of recombinant enzymes that digest the toxic gliadin fractions in the stomach or upper small intestine.

Prolyl endopetidases have gained importance as a potential treatment of celiac disease as well as other neuropsychiatric diseases and cognitive disturbances.

These enzymes belong to the serine protease family and have been found to hydrolyze small proline-containing peptides at the C-side of the proline residue (Gass and Khosla., 2007). Recently, it has been proposed that the enzyme can be used as a potential therapeutic agent in treating celiac sprue due to its unique ability to accelerate the cleaving of proline-containing gluten peptides in the gut lumen (Shan et al., 2002).

1.5 Problem statement

Extensive research has been done on the isolation of PEP from microorganisms but nothing according to my knowledge has been documented yet on cereal crops. Studies also show that even a small amount of gluten in food products cause health problems and affect those already suffering from celiac disease.

1.6 Hypothesis

We therefore hypothesize that cereal plants express glutamine and proline specific peptidases which are able to hydrolyze toxic gliadin (gluten proteins) fractions which can be used to alleviate the symptoms of celiac disease.

1.7 Aim

The present study is aimed at purifying and kinetically characterizing gliadin hydrolysing peptidases from *Triticum aestivum* (wheat) that may possibly lead to a novel and specific inhibitor to control disease.

1.8 Specific objectives

- Identification of plant Prolyl endopetidases in plants
- Isolation and purification of plant Prolyl endopetidases from a cereal crop (wheat)
- Kinetic analysis of purified Prolyl endopetidases

REFERENCES

1) Aoyagi T, Wada T, Nagai M, Kojima F, Harada S, Takeuchi T, Takahashi H, Hirokawa K and Tsumita T. (1990). Increased gamma-aminobutyrate aminotransferase activity in brain of patients with Alzheimer's disease. Chem. Pharm. Bull. (Tokyo) **38**: 1748–1749
2) Cunningham DF and O'Connor B. (1997). Identification and initial characterization of an N-benzyloxycarbonyl-prolyl-prolinal (Z-Proprolinal)-insensitive 7-(N-benzyloxycarbonyl-glycyl-prolyl-amido)-4-methylcoumarin (Z-Gly-Pro-NH-Mec)-hydrolysing peptidase in bovine serum. Eur. J. Biochem. **244**: 900–903
3) Fulop V, Szeltner Z, Renner V and Polgar L. (2001). Structures of prolyl oligopeptidase substrate/inhibitor complexes. Use of inhibitor binding for titration of the catalytic histidine residue. J. Biol. Chem. **276**: 1262–1266
4) Fulop V, Bocskei Z and Polgar L. (1998). Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. Cell. **94**: 161–170
5) Fuxreiter M, Magyar C, Juhasz T, Szeltner Z, Polgar L and Simon I. (2005). Flexibility of prolyl oligopeptidase: molecular dynamics and molecular framework analysis of the potential substrate pathways. Proteins. **60**: 504–512.
6) Gass J and Khosla C. (2007). Prolyl endopeptidases. Cell Mol Life Sci. **64**:345–55.
7) Gibson A, Edwardson J, McDermott J. (1991). Post mortem levels of some brain peptidases in Alzheimer's disease: reduction in proline endopeptidase activity in cerebral cortex. Neurosci. Res. Commun. **9**: 73–81.
8) Goossens F, De Meester I, Vanhoof G and Scharpe S. (1996). Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids. Eur. J. Clin. Chem. Clin. Biochem. **34**: 17-22.
9) Harris MN, Madura JD, Ming LJ and Harwood VJ.(2001). Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile Pyrococcus furiosus. J. Biol. Chem. **276**: 19310–19317.
10) Harwood VJ, Denson JD, Robinson-Bidle KA and Schreier HJ. (1997). Overexpression and characterization of a prolyl endopeptidase from the hyperthermophilic archaeon Pyrococcus furiosus. J. Bacteriol. **179**: 3613–3618.

11) Hashimoto Y, Kakegawa H, Narita Y., et al. (2001). Significance of cathepsin B accumulation in synovial fluid of rheumatoid arthritis. Biochem Biophys Res Commun **283**: 334-339.

12) Hutchinson MJ, Robins G and Howdle PD. (2008). Advances in Coeliac disease. Curr.Opin.Gastroenterol. **24**(2):129-134.

13) Ichai C, Chevallier N, Delaere P, Dournaud P, Epelbaum J, Hauw JJ, Vincent JP and Checler F. (1994). Influence of region specific alterations of neuropeptidase content on the catabolic fates of neuropeptides in Alzheimer's disease. J. Neurochem. **62**: 645–655.

14) Irazusta J, Larrinaga G, Gonzalez-Maeso J, Gil J, Meana JJ and Casis L. (2002). Distribution of prolyl endopeptidase activities in rat and human brain. Neurochem. Int. **40**: 337–345.

15) Juhasz T, Szeltner Z and Polgar L. (2006). Properties of the prolyl oligopeptidase homologue from Pyrococcus furiosus. FEBS Letters **580**: 3493–3497.

16) Kakegawa H, Matano Y, Inubushi T and Katunuma N. (2004). Significant accumulations of cathepsin B and prolyl endopeptidase in inflammatory focus of delayed-type hypersensitivity induced by Mycobacterium tuberculosis in mice. Biochem Biophys Res Commun.**316**:78–84.

17) Kamei H, Ueki T, Obi Y, Fukagawa Y and Oki T. (1992). Protective effect of eurystatins A and B, new prolyl endopeptidase inhibitors, on scopolamine-induced amnesia in rats. Jpn J Pharmacol. **60**: 377–380.

18) Kamori M, Hagihara M, Nagatsu T, Iwata H and Miura T. (1991). Activities of dipeptidyl peptidase II, dipeptidyl peptidase IV, prolyl endopeptidase and collagenase-like peptidase in synovial membrane from patients with rheumatoid arthritis and osteoarthritis. Biochem Med Metab Biol. **45**: 154-160.

19) Kimura A, Yoshida I, Takagi N and Takahashi T. (1999). Structure and localization of the mouse prolyl oligopeptidase gene. J. Biol.Chem. **274**: 24047–24053.

20) Laitinen KS, van Groen T, Tanila H, Venalainen JI, Mannisto PT and Alafuzoff I. (2001). Brain prolyl oligopeptidase activity is associated with neuronal damage rather than beta-amyloid accumulation. Neuroreport.**12**: 3309–3312.

21) Larkin MA, Blackshields G, Brown NP et al. (2007). ClustalW2 and ClustalX version 2. Bioinformatics. **23** (21): 2947-2948.

22) Leprince J, Cosquer D, Bellemere G, Chatenet D, Tollemer H, Jegou S, Tonon MC and Vaudry H. (2006). Catabolism of theoctadecaneuropeptide ODN by prolyl endopeptidase: identification of an unusual cleavage site. Peptides **27**: 1561–1569.

23) Lerner A, Blank M and Shoenfeld Y. (1996). Celiac disease and autoimmunity. Isr J Med Sci **32**:33–36.

24) Leung D, Abbenante G and Fairlie DP. (2000). Protease inhibitors: current status and future prospects. J. Med. Chem. **43**: 305–341.

25) Maes M, Goossens F, Scharpe S, Meltzer HY, D'Hondt P and Cosyns P. (1994). Lower serum prolyl endopeptidase enzyme activity in major depression: further evidence that peptidases play a role in the pathophysiology of depression. Biol. Psych. **35**: 545–552.

26) Maes M, Goossens F, Scharpe S, Calabrese J, Desnyder R and Meltzer HY. (1995). Alterations in plasma prolyl endopeptidase activity in depression, mania, and schizophrenia: effects of antidepressants, mood stabilizers, and antipsychotic drugs. Psychiatry Res. **58**: 217–25.

27) Maes M, Libbrecht I, Van Hunsel F, Lin AH, Bonaccorso S, Goossens F, De Meester I, De Clerck L, Biondi M, Scharpe S and Janca A. (1998). Lower serum activity of prolyl endopeptidase in fibromyalgia is related to severity of depressive symptoms and pressure hyperalgesia. Psychol. Med. **28**: 957–965.

28) Mantle D, Falkous G, Ishiura S, Blanchard PJ and Perry EK. (1996). Comparison of proline endopeptidase activity in brain tissue from normal cases and cases with Alzheimer's disease, Lewy body dementia, Parkinson's disease and Huntington's disease. Clin. Chim. Acta**. 249**: 129–139.

29) Molberg O, Uhlen AK, Jensen T, Flaete NS, Fleckenstein B, Arentz-Hansen H, et al. (2005). Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology. **128**:393–401.

30) O'Leary RM, Gallagher SP and O'Connor B. (1996). Purification and characterization of a novel membrane-bound form of prolyl endopeptidase from bovine brain. Int. J. Biochem. Cell. Biol. **28**: 441–449.

31) Oyama H, Aoki H, Amano M, Mizuki E, Yoshimoto T, Tsuru D and Murao S. (1997). Purification and Characterization of a Prolyl Endopeptidase from *Pseudomonas* sp. KU-22. Journal of fermentation. 84(6) 538-542.

32) Polgar L. (1992). Prolyl endopeptidase catalysis. A physical rather than a chemical step is rate-limiting. Biochem. J. **283** (3): 647–648.

33) Polgar L. (2002). The prolyl oligopeptidase family. Cell. Mol. Life Sci. **59**: 349–362.

34) Portevin B, Benoist A, Remond G., et al. (1996). New prolyl endopeptidase inhibitors: in vitro and in vivo activities of azabicyclooctane, azabicycloheptane, and perhydroindole derivatives. J Med Chem. **39**: 2379–2391.

35) Rosenblum JS and Kozarich JW. (2003). Prolyl peptidases: a serine protease subfamily with high potential for drug discovery. Curr. Opin. Chem. Biol. **7**: 496–504.

36) Sedo A, Krepela E and Kusafirek E. (1991). Dipeptidyl peptidase IV, prolyl endopeptidase and cathepsin B activities in primary human lung tumours and lung parenchyma. J Cancer Res Clin Oncol. **117**: 249-253.

37) Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM and Hosla C. (2002). Structural basis for gluten intolerance in celiac sprue. Science. **297**: 2275–2279.

38) Shan L, Marti T, Sollid LM, Gray GM and Khosla C. (2004). Comparative biochemical analysis of three bacterial prolyl endopetidases: implications for coeliac sprue. Biochem. J. **383**: 311–318.

39) Shan L, Mathews II and Khosla C. (2005). Structural and mechanistic analysis of two prolyl endopeptidases: role of interdomain dynamics in catalysis and specificity. Proc. Natl. Acad. Sci. USA **102**: 3599–3604.

40) Szeltner Z, Rea D, Juhasz T, Renner V, Fulop V and Polgar. (2004). Concerted structural changes in the peptidase and the propeller domains of prolyl oligopeptidase are required for substrate binding. J. Mol. Biol. **340**: 627–637.

41) Szwajcer-Dey E, Rasmussen J, Meldal, M and Breddam K. (1992). Proline-specific endopeptidases from microbial sources: isolation of an enzyme from a Xanthomonas species. Journal of bacteriology. **174** (8): 2454-2459.

42) Taylor WL and Dixon JE. (1980). Catabolism of neuropeptides by a brain proline endopeptidase. Biochem. Biophys. Res. Commun**. 94**: 9–15.

43) Terwel D, Bothmer J, Wolf E, Meng F and Jolles J. (1998). Affected enzyme activities in Alzheimer's disease are sensitive to antemortem hypoxia. J. Neurol. Sci. **161**: 47–56.

44) Tsuru D, Yoshimoto T, Koriyama N and Furukawa S. (1988). Thiazolidine derivatives as potent inhibitors specific for prolyl endopeptidase. J Biochem.**104**:580–586.

45) Venalainen JI, Juvonen RO and Mannisto PT. (2004). Evolutionary relationships of the prolyl oligopeptidase family enzymes. Eur. J. Biochem. **271**: 2705–2715.

46) Venalainen JI, Juvonen RO and Mannisto PT. (2004). Evolutionary relationships of the prolyl oligopeptidase family enzymes. Eur. J. Biochem. **271**: 2705–2715.

47) Wallen EA, Christiaans JA, Saario SM, Forsberg MM, Venalainen JI, Paso HM, Mannisto PT and Gynther J. (2002). 4-Phenylbutanoyl-2(S)-acylpyrrolidines and 4-phenylbutanoyl-Lprolyl-2(S)-acylpyrrolidines as prolyl oligopeptidase inhibitors. Bioorg. Med. Chem. **10**: 2199–2206.

48) Walter R, Shlank H, Glass JD, Schwartz IL, Kerenyi TD. (1971). Leucylglycinamide released from oxytocin by human uterine enzyme. Science **173**: 827–829.

49) Wilk S. (1983). Prolyl endopeptidase. Life Sci. **33**: 2149–2157.

50) Wolters MVand Wijmenga C. (2008). Genetic background of celiac disease and its clinical implications. Am J Gastroenterol. **103**(1):190-195.

51) Xie HX, Nie P and Sun BJ. (2004). Characterization of two membrane-associated protease genes obtained from screening out-membrane protein genes of Flavobacterium columnare G4. J. Fish. Dis. **27** (12): 719-729.

52) Yoshimoto T, Ogita K, Walter R, Koida M and Tsuru D. (1979). Post-proline cleaving enzyme. Synthesis of a new fluorogenic substrate and distribution of the endopeptidase in rat tissues and body fluids of man. Biochim. Biophys. Acta. **569**: 184–192.

53) Yoshimoto T, Sattar A, Hirose W and Tsuru D. (1987). Studies of prolyl endopeptidase from carrots (Daucus carota): purification and enzymatic properties. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology. **916** (1): 29-37.

54) Yoshimoto T, Sattar AK, Hirose W and Tsuru D. (1988). Studies on prolyl endopeptidase from shakashimeji (Lyophyllum cinerascens): purification and enzymatic properties. J. Biochem. **104**:622–627.

55) Zima T, Tesar V, Mantle D et al. (2001). Acute doxorubicin (Adriamycin) dosage does not reduce cardiac protein synthesis in vivo, but decreases diaminopeptidase I and proline endopeptidase activities. Exp Mol Pathol. **70**: 154-161.

CHAPTER 2

ABSTRACT

PEP activity has been described in several locations and has mostly been linked to a variety of neurological disorders such as schizophrenia, amnaesia, depression as well as other disease states such as anorexia nervosa, bulimia nervosa and blood pressure regulation. The enzyme has also been previously isolated from a variety of archae, microorganisms and several eukaryotic species but no prolyl endopeptidases have been isolated from plants. Plants have high levels of proline and glutamine rich peptides in seeds. We therefore hypothesize plants must express PEPs during germination.

Bioinformatics tools were used to identify known PEPs and putative plant PEPs. A global sequence alignment of putative plant PEPs and other known PEPs indicated that the active site amino acids Ser, His and Asp are conserved in putative plant PEP sequences. Furthermore, putative plant PEPs showed similar secondary structures to known PEPs and when a rice PEP was modelled onto porcine brain PEP structure, a high degree of similarity was found.

Germination studies of wheat seed showed an increase of PEP activity over time with maximum PEP activity reached after 4 days of germination, which remained at this level until 9 days of germination, implying a function for PEP in plant seed germination.

Wheat PEP was purified using ion exchange and gel filtration chromatography with a final yield of less than 1% and a relative purity (only 2 bands detected by SDS-PAGE).

The purified wheat PEP had a molecular weight of approximately 55kDa, substrate specificity for chymotrypsin-like substrates (N-Suc-Ala-Ala-Pro-Phe-pNa, K_m value of 0.58 mM, K_{cat} of 29.37 $s^{-1;}$ K_{cat}/K_m 50813.14s^{-1} M^{-1}); a pH optimum of 7.9; temperature optima of 37^oC and a high sensitivity to temperature as indicated by loss of activity at temperatures above 40^oC.

Inhibition studies using E64, Leupeptin and PMSF confirmed that the wheat PEP is from the Serine protease family and is most likely a trypsin-like protease.

CHAPTER 3

IN-SILICO ANALYSIS OF PROLYL ENDOPEPTIDASES IN *TRITICUM AESTIVUM* (WHEAT)

3.1 Abstract

Transcription plays a key role in the flow of information from static DNA to proteins that are active and is the most widely studied process in molecular and cell biology. It is shaped by the interactions between transcription factors that bind to *cis*-regulatory elements in the DNA, as well as chromatin structure and additional co-factors. However, each gene contains a unique set of *cis*-acting elements in its 5'regulatory region that ultimately determines its spatial and temporal expression. The objectives of this study were to identify genes expressing putative Prolyl endopetidases in plants and to isolate and characterize the purified PEPs. Prolyl endopeptidase encoding genes from *Arabidopsis thaliana* genome (chromosome 1), *Oryza sativa* genome (chromosome 1 and 4), *Zea mays* and *Sorghum bicolor* were identified and the 2kbp 5' regulatory region of each PEP genes in rice (*Oryza sativa*) and *Arabidopsis thaliana* was scanned for the presence of putative *cis*-acting regulatory elements identical with or similar to the motifs registered in PLACE and PLANTCARE. No sequences for *Triticum aestivum* (wheat) were obtained due to lack of genome sequences. Analysis of the 2kbp upstream promoter region revealed *cis*-regulatory elements in varying frequencies involved in hormone regulation, abiotic stress response as well as light response, thereby implicating plant PEP in development and stress response mechanisms. Sequence alignments, evolutionary trends, primary, secondary and tertiary structure prediction were also performed to confirm the identification of putative plant PEPs.

3.2 Introduction

Prolyl endopeptidase (PEP) is a cytosolic serine protease and has been found to be post-proline cleaving enzymes (Polgar, 2002). Crystallography studies on PEP from porcine muscle revealed that the enzyme is made up of a two domain structure: a catalytic domain which resembles a serine protease such as trypsin or chymotrypsin and a β-propeller domain which appears to form a barrel shape lid over the active site (Fulop et al., 2001). PEPs have been purified and characterized from various bacterial and mammalian sources. It is broadly distributed in all human tissues with the majority of its activity detected in the brain (Goossens et al., 1996). Despite its thorough enzymological and structural characterization, its exact function still remains obscure. Recently, it has been proposed that the enzyme may serve as a potential treatment for celiac disease due to its unique post-proline cleaving ability (personal communication).

Celiac disease is a disease characterized by enteropathy of the small intestine leading to villus atrophy (Wolters et al., 2008). It manifests itself in genetically predisposed individuals who are intolerant to gluten containing food products. Gluten provokes the disease due to its high proline and glutamine content. The absence of PEP from mammalian pancreatic proteases and the intestinal brush-border enzymes highlights the lack of a role of PEP activity in the assimilation of dietary proteins in mammals (Shan et al., 2004). As a result, the only accepted nutritional therapy involved the life-long elimination of gluten containing food products from a person's diet. However, advances have been made in developing non-dietary therapies in treating celiac sprue including the use of PEPs. Bacterial PEP was initially identified as a potential source of PEP enzymes, however, recently focus has shifted over to plant based sources that may serve as potential agents for the treatment of celiac disease. Plant PEP sources have been considered as gluten-detoxifying agents but there are still uncertainty in whether plants express PEP enzyme activity. We have therefore undertaken a study to identify PEP encoding genes in plants; to use computational approaches to identify the presence of putative *cis*-acting regulatory elements and its biophysical properties (functional and structural) as well as a wet bench approach to investigate whether PEP proteins play a role in seed germination.

3.3 Method

3.3.1 Search for the PEP gene

The retrieval of all known previously identified plant PEPs and plant Prolyl-like endopetidases genes were achieved through the use of Entrez accessed via the NCBI website (http://www.ncbi.nlm.nih.gov/Entrez). For all bioinformatics software accessed for this study, the default settings were used.

The keywords ("plant Prolyl endopetidases") were used for retrieving the sequences from Entrez. The sequences were saved in FASTA format in a text file. PEP genes identified were from *Arabidopsis thaliana*, rice, maize and *Sorghum bicolour*. No sequences for wheat were obtained.

3.3.2 Sequence Alignments

A global sequence alignment was performed on all PEP protein sequences encoded for by this gene identified in section 3.3.1using CLUSTALW2 (http://www.ebi.ac.uk/tools/clustalw2/index.html) (Larkin et al., 2007).

3.3.3 Promoter and Regulatory Elements Prediction

Recognition of promoter elements and regulatory elements in PEP was achieved by scanning the genomic DNA 2kb upstream of the transcriptional start site (TSS) with PLACE (Plant Cis-acting Regulatory DNA elements; http://www.dna.affrc.go.jp/PLACE/signalscan.html) according to the method of Higo et al., (1999) and plantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/). For each rice sequence, the 2000bp upstream region was scanned after the correct chromosome was identified where the gene that codes for PEP resides. This was done on the TIGR database (http://blast.jcvi.org/euk-blast/index.cgi?project=osa1). The same was done for *Arabidopsis thaliana* on the TAIR database (http://www.arabidopsis.org/). After obtaining the 2000bp upstream region, the sequences were then blasted to see whether they code for any proteins.

3.3.4 Subcellular Location and Signal Peptide Prediction

Subcellular location and signal peptide prediction were done for the PEP sequences ((NP_173463, BAF15466) using TargetP (http://www.cbs.dtu.dk/services/TargetP/) as described by Emanuelson et al.,(2007).

3.3.5 Secondary Structure Prediction

Protein secondary structures of PEP from Arabidopsis (NP_173463) and rice (BAF15466) were predicted using PSIpred (http://bioinfo.cs.ucl.ac.uk/psipred) as described by Jones (1999).

3.3.6 Tertiary Structure Prediction

The tertiary structure of PEP from rice (BAF15466) was predicted using SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold et al., 2006).

3.4 Results and Discussion

Similarity searches to the known bacterial PEP sequences (AAA73423 and ABA26946) were initially performed but no hits in plants were found. The search query was then confined to the keywords ("prolyl endopeptidase") and proteins from rice, *Arabidopsis thaliana*, *Zea mays*, *Sorghum bicolor*, *Homo sapiens*, *Drosophila* as well as a few bacterial and archae sequences were identified with some redundancy amongst the sequences. A global alignment was performed between the plant sequences and known PEP sequences from bacteria and *tribolium castaneum* to identify whether the PEPs in plants have similar primary sequences to known PEP and whether the active site, peptidase domain and β-propeller domain is conserved in the plant sequences. Once we identified whether these domains were conserved in the plant sequences, a BLAST was then performed using BLASTP (http://blast.ncbi.nlm.nih.gov/) in which the rice sequence (BAF15466) was then used as a query to identify other plant PEPs in the non-redundant protein database. Several other plant protein sequences were identified from *Sorghum bicolor* (XP_002446934), *Zea Mays* (NM 001136920.1, ACN31271) and *Vitus vinifera* (XP_002285910). No protein sequences were available in the non-redundant database for *Triticum aestivum* (wheat).

3.4.1 Sequence Alignment

Sequence alignment plays a key role in predicting a proteins possible function through sequence comparison with proteins of known function. Protein's that significantly match each other are expected to share a common structure and function especially if amino acids in the active site are conserved.

Multiple sequence alignments, in which three or more sequences are aligned, are useful in finding conserved regulatory patterns in nucleotide sequences and for identifying functional as well as structural domains in protein families (Bergeron, 2003). A global sequence alignments of PEPs from rice, Arabidopsis, sorghum, maize, drosophila, Human, insect and microorganisms were performed (figure 3.1) and significant similarities were observed.

```
|OsPEPChr4|      VSCIANIRGGGEYGEDWHKAGSLANKQNCFDDFIAAGEFLVSAGYTNPSRLCVEGASNGG 580
|S.bPEP|         VTCVANIRGGGEYGEDWHRAGSLANKQNCFDDFIAAGEFLVSAGYTSPARLCIEGGSNGG 583
|ZmPEP1|         VVCVANIRGGGEYGEEWHKAGALAMKQNCFDDFAACAEFLISNGYTSSRRLCIEGGSNGG 611
|ZmPEP2|         VVCLANIRGGGEYGEEWHKAGALAMKQNCFDDFAACAEFLISNGYTSSRRLCIEGGSNGG 571
|OsPEPChr1|      VVCVANIRGGGEYGEEWHKAGARAMKQNCFDDFIACAELLISAGYTSYRQLCIEGGSNGG 570
|AtPEP1|         VFCFANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 635
|AtPEP2|         VFCFANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 597
|AlPEP2|         VFCVANIRGGGEYGEEWHKAGSLAKKQNCFDDFISGAEYLVSAGYTQPSKLCIEGGSNGG 635
|AlPEP1|         VFCFANIRGGGEYGEEWHKSGALANKQNCFDDFISGAEYLVSAGYTQPRKLCIEGGSNGG 571
|HsPEP|          ILAVANIRGGGEYGETWHKGGILANKQNCFDDFQCAAEYLIKEGYTSPKRLTINGGSNGG 557
|DmPEP|          VLAYPNIRGGGEYGEKWHNGGRLLNKQNVFDDFQCAAEYLIKNNYTTKDRLAIQGGSNGG 568
|DsPEP|          VLAYPNLRGGGEYGEKWHNGGRLLNKQNVFDDFQTAAEYLIENKYTTKDRLAIQGGSNGG 602
T.cPEP|          VYALANIRGGGEYGDNWHNGGRFGKKQNCFDDFQYAAKYLVENKYTKVDKLIIQGGSNGG 570
|T.oPEP|         TFVMANLRGGSEYGEEWHRAGMRENKQNVFDDFIAVLGKLKAEGYK----VAAWGRSNGG 480
|P.fPEP|         TFIMANLRGGSEYGEEWHRAGMRENKQNVFDDFIAVLEKRKKEGYK----VAAWGRSNGG 480
|N.mPEP|         AFVLANIRGGGEFGPRWHQAAQGISKHKSVDDLLAVVSDLSERGISSPEHIGLQGGSNGG 533
                     .*:***.*:*  **...    *:: .**:           .    :   * ****

|OsPEPChr4|      LHNVRRPWEKG-----HRRQQYPSTMLLTADHDDRVVPSHTLKFLATMQHVLCTSVKE-S 693
|S.bPEP|         LHNVRRPWEKEKWAAATGGGQYPATMLLTADHDDRVVPSHTLKFLATMQHVLRAGAEGGS 702
|ZmPEP1|         LHNVRRPWEQS----SGNNCQYPATMLLTADHDDRVVPLHSLKLLATLQHVLCTSTED-S 725
|ZmPEP2|         LHNVRRPWEQS----SGNNCQYPATMLLTADHDDRVVPLHSLKLLATLQHVLCTSTED-S 685
|OsPEPChr1|      LHNVRRPWEQS----FVNCCQYPAIMLLTADHDDRVVPLHSLKLLATLQYVLCTSIED-T 684
|AtPEP1|         LHNVKRPWEQQ----TDHLVQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLDN-S 749
|AtPEP2|         LHNVKRPWEQQ----TDHLVQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLDN-S 711
|AlPEP2|         LHNVKRPWEQQ----TDQFIQYPSTMLLTADHDDRVVPLHSLKLLATLQHVLCTSLEN-S 748
|AlPEP1|         LHNVKRPWEQK----TDRFVQYPSTMLLTADHDDRVVPLHTYKLLATMQYELGLSLEN-S 685
|HsPEP|          LHNVKLPEADD--------IQYPSMLLLTADHDDRVVPLHSLKFIATLQYIVGRSRKQ-- 666
|DmPEP|          LHNVHTPNDASS--------EYPSTLILTADHDDRVSPLHSLKFAAALQEAVRDSPFQ-- 677
|DsPEP|          LHNVHTPKGAET--------EYPSTLILTADHDDRVSPLHSLKFIAALQEAVRDSKFQ-- 711
T.cPEP|          LHNIRVPQNGG---------QYPATLLLTADHDDRVVPLHSLKFIAELQNKIGRLPTQ-- 678
|T.oPEP|         YHNVK-------------EQKYPPTLLYTGLYDDRVHPAHALKFFMKLRAVS-------- 578
|P.fPEP|         YHNVDP------------KKKYPPTLIYTGLHDDRVHPAHALKFFMKLKEIG-------- 579
|N.mPEP|         YHNLSDG------------IDYPPALITTSLSDDRVHPAHALKFYAKLRETS-------- 633
                 **:               .**. :: *.  **** * *: *:   ::
```

```
|OsPEPChr4|      PQTNPIVARIDRKSGHGCGRSTQKIIDEAADRYAFAAKTMGISWID- 739
|S.bPEP|         PQTNPIIARIERNSGHCCGRSTQKIIDEAADRYAFAAKMMGVSWID- 748
|ZmPEP1|         PQTNPIIGRIDRKSGHGAGRPTQKMIDEAADRYSFMAKMLGASWTE- 771
|ZmPEP2|         PQTNPIIGRIDRKSGHGAGRPTQKMIDEAADRYSFMAKMLGASWTE- 731
|OsPEPChr1|      PQVNPIIGRIDVKSGHGAGRPTKKMIDEVADRYSFMANMLDASWTE- 730
|AtPEP1|         PQMNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNASWTE- 795
|AtPEP2|         PQMNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNASWTE- 757
|AlPEP2|         PQTNPIIGRIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVNAAWTE- 794
|AlPEP1|         PQTNPIIARIEVKAGHGAGRPTQKMIDEAADRYSFMAKMVDASWID- 731
|HsPEP|          --SNPLLIHVDTKAGHGAGKPTAKVIEEVSDMFAFIARCLNVDWIP- 710
|DmPEP|          --KNPLLLRVYKKAGHGAGKPTSKRIEEATDILTFMYRSLNIDVINL 722
|DsPEP|          --NNPVLLRVYQKAGHGAGKPTSKRIEEATDILTFLSKSLNVDTINV 756
T.cPEP|          --KNPLMIRIETRAGHGAGKPTSKIIEEVTDTFCFISRALNLTFSS- 722
|T.oPEP|         ---APVYLRVETKSGH-MGASPETRARELTDLLAFVVETLEA----- 616
|P.fPEP|         ---APVYLRVETKSGH-MGASPETRARELTDLLAFVLKTLS------ 616
|N.mPEP|         ---PQSWLYSPDGGGHTGNGTQREAADELACVLLFLKEFLG------ 671
                            .**  . .      * :    *  . :
```

Figure 3.1: A global alignment of confirmed PEP and putative plant PEP protein sequences of *Zea Mays*: ZmPEP1 (NM 001136920.1) and ZmPEP2 (ACN31271); Rice PEP: OsPEPChr4 (BAF15466) and OsPEPChr1 (BAF03701); *Arabidopsis Lyrata*: AlPEP (XP_002887627); *Arabidopsis thaliana* PEP: AtPEP1 (NP_173463) and AtPEP2 (NP_001117606); Drosophila PEP: DsPEP (XP_002036510) and DmPEP (XP_002002881); *Sorghum bicolor* PEP: SbPEP (XP_002446934); *Homo sapiens* PEP: HsPEP (NP_002717); *Tribolium castaneum* PEP: TcPEP (XP_972959.1); *Pyrococcus furiosis* PEP: *Pf*PEP(AAA73423), *Thermococcus Onureus* PEP: *To* PEP (ABA26946) and *Neisseria meningitidis* PEP: *Nm*PEP (YP_002342055). Only the sequences surrounding the known catalytic triad, serine (S), aspartate (D) and histidine (H) are shown. For a full sequence alignment showing the various domains see fig 1.3.

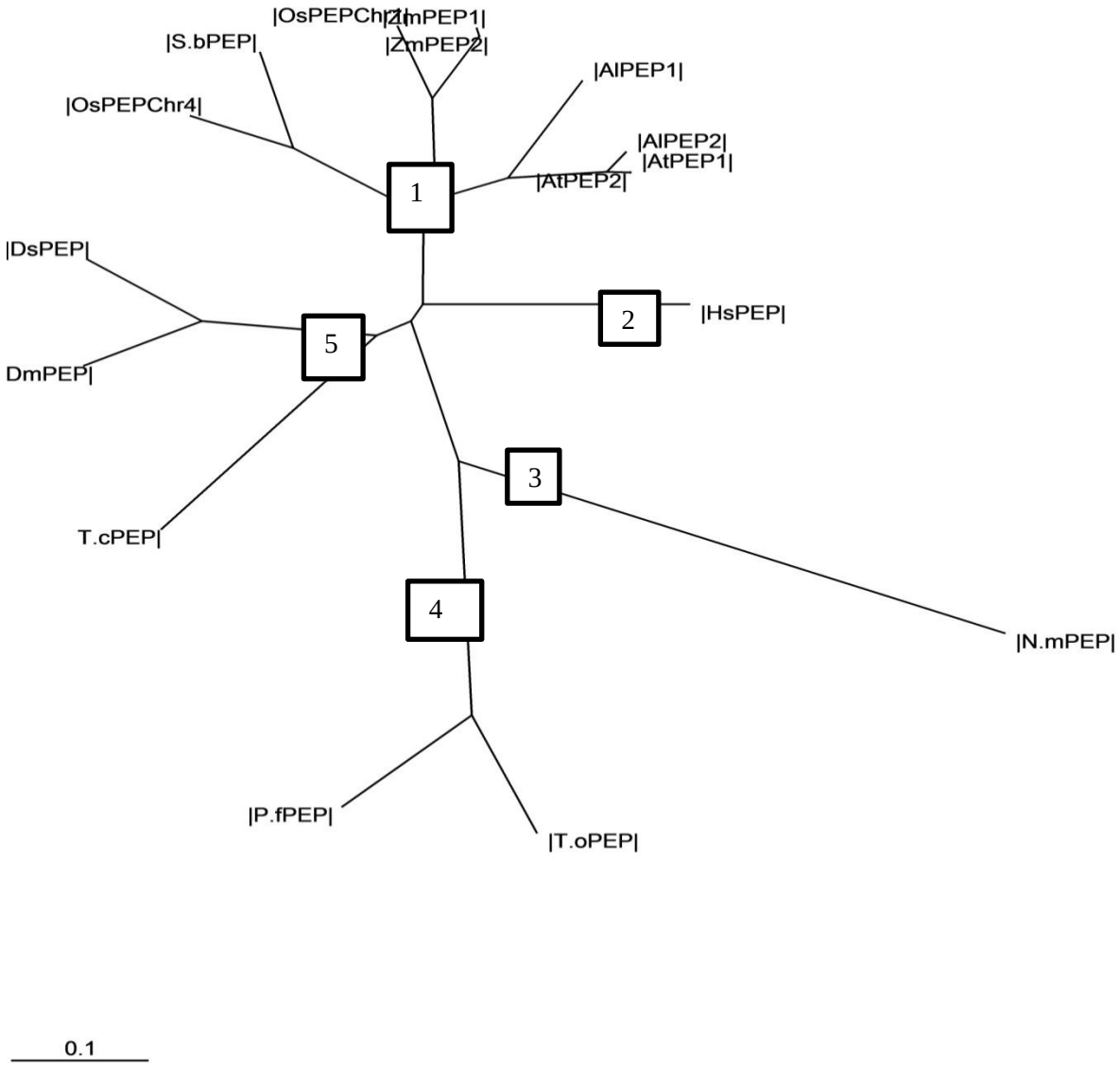


Figure 3.2: Phylogenetic tree of putative plant Prolyl endopetidases and PEPs from other organisms including bacterial, insect and human. *Zea Mays* PEP: ZmPEP1 (NM 001136920.1) and ZmPEP2 (ACN31271); Rice PEP: OsPEPChr4 (BAF15466) and OsPEPChr1 (BAF03701); *Arabidopsis Lyrata* PEP: AlPEP (XP_002887627); *Arabidopsis thaliana* PEP: AtPEP1 (NP_173463) and AtPEP2 (NP_001117606); Drosophila PEP: DsPEP (XP_002036510) and DmPEP (XP_002002881); *Sorghum bicolor* PEP: SbPEP (XP_002446934); *Homo sapiens* PEP: HsPEP (NP_002717); *Tribolium castaneum* PEP: TcPEP (XP_972959.1); *Pyrococcus furiosis* PEP: *Pf*PEP(AAA73423), *Thermococcus Onureus* PEP: *To* PEP (ABA26946) and *Neisseria meningitidis* PEP: *Nm*PEP (YP_002342055).

From the phylogenetic analysis, it would appear that there are 5 clusters of PEPs: 1) Plant PEP; 2) Mammalian; 3) Bacterial; 4) Archae and 5) Insect clusters. It is interesting to note that from the clusters observed, Drosophila and *Tribolium castaneum* differ more from the human PEP than do the plants *Oryza sativa, Arabidopsis thaliana*, *Sorghum bicolor* and *Zea mays*. The most probable reason for this observation is that these two insects may have diverged considerably faster than vertebrates. The presence of PEPs in these various clusters as seen from the phylogenetic tree also show that the enzyme clearly developed before the archaea, prokaryota and eucaryota diverged along their own evolutionary lines. This may thus suggest that all PEPs are possibly of ancient origin. Venalainen and colleagues (2004) who work on the evolutionary relationships of PEP family enzymes showed that the PEP family of enzymes were present in the last universal common ancestor (LUCA) of all life forms. Their work further indicates that the high conservation of PEP protein sequences from different species and their presence in the LUCA strongly suggest that these enzymes have important roles in physiological processes but the exact roles of these enzymes are still obscure. Evidently there was a need for peptidases that cleave only small peptides specifically after proline, lysine or arginine even during the early days of life.

3.4.2 Transcriptional signals and regulatory elements

A summary of the predicted *cis*-regulatory elements is represented in the form of a 3-D cylinder chart in figure 3.3. Most of the TFBS were shown to be related to hormonal regulation, light response, as well as abiotic stress. A detailed list of the *cis*-acting elements that have been found in the 2000bp 5' upstream regulatory region of the rice and Arabidopsis PEP gene sequences is given in appendix A (page 91).

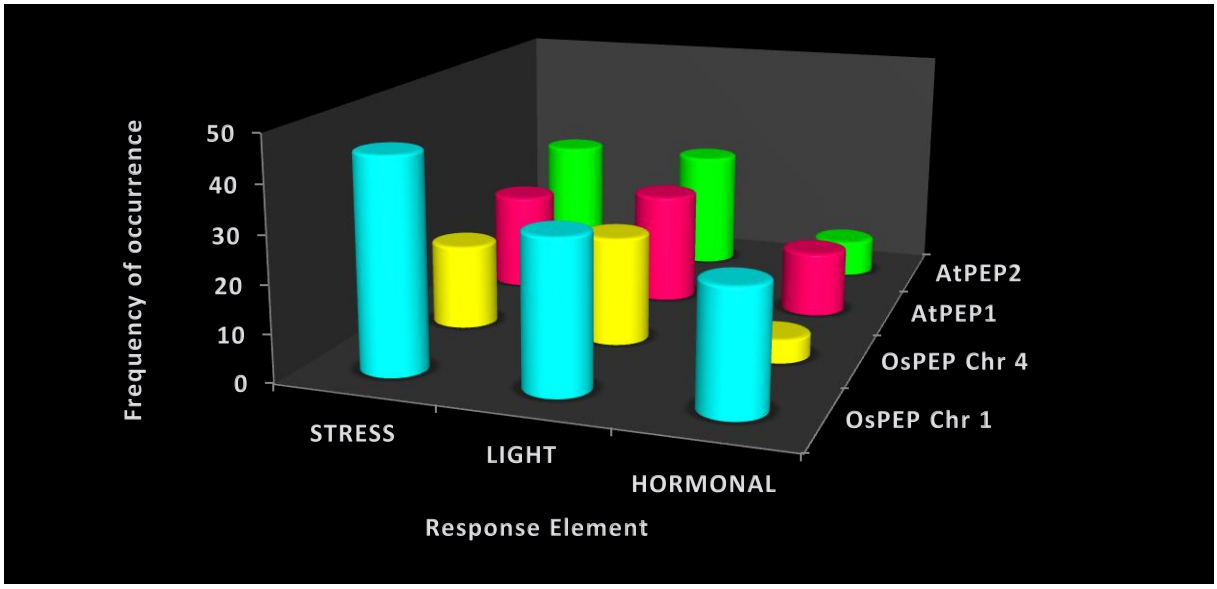


Figure3.3: A comparative in-silico analysis of the frequencies of the associated *cis*-acting elements identified in the promoter sequences of rice (*Oryza sativa* Japonica cultiva-group) and *Arabidopsis thaliana* prolyl endopetidase gene families. OsPEPChr1 (BAF03701); OsPEPChr4 (BAF15466); AtPEP1 (NP_173463) and AtPEP2 (NP_001117606). From the putative *cis*-acting regulatory elements that were identified in the 2kbps upstream of the 5'regulatory region starting from the translational start site of the rice and Arabidopsis PEP genes, variations were observed both in the frequency of occurrence and nature of elements that were identified but also to some extent similar elements in the two plants sequences were seen in the nature of the elements which would cause one to assume that the rice and Arabidopsis will have similar expression patterns. Vilo et al.,(2000) further attest that genes which have common *cis*-acting regulatory elements in the 5'regulatory region may assume similar expression patterns. Wang et al.,(2004) also draws attention to the fact that common *cis*-acting regulatory elements are the main factors of regulated genes and are often present in the regions where complex protein interaction occur. With reference to figure 3.3, PEP genes may well be regulated by or involved in developmental activities due to the high number of light response, abiotic stress as well hormonal regulated associated *cis*-acting regulatory elements present at the 5'regulatory region.

Light is used by plants as a signal for many physiological and developmental processes. One of the most extensively studied family of photoreceptors that plants use to perceive the presence and quality of light in their environment is phytochrome. Although phytochrome-regulated promoters are complex and can also respond to others signals specific DNA elements that are Ibox (conserved sequence upstream of light regulated genes) and GATA motif (required for high level, light regulated and tissue specific expression). The presence of

these TFBS helps confirm that PEP may possibly function in plant response to varying light conditions.

Plants encounter various environmental insults during a typical life cycle and as a result, have evolved mechanisms able to increase their tolerance of these insults through both physical adaptations and interactive molecular and cellular changes that begin after the onset of stress. The first step in switching on such molecular responses is to perceive the stress as it occurs and to relay information about it through a signal transduction pathway. These pathways eventually lead to physiological changes such as guard cell closure, or to the expression of genes and resultant modification of molecular and cellular processes. It is worth pointing out that the majority of stress responses associated *cis*-acting elements identified were specifically related to abiotic stress. Abiotic stress conditions cause extensive losses to agriculture production worldwide (Boyer., 1982; Bray et al., 2000). Individually, stress conditions such as drought, salinity or heat have been the subject of intense research (Bray et al., 2000; Cushman and Bohnert, 2000). However, in the field, crops as well as other plants are continually subjected to a combination of different abiotic stresses (Maffat, 2002, Heyne and Brunson, 1940; Craufurd and Peacock, 1993; Jiang and Huang, 2001). Recent studies have shown that plants have developed unique metabolic and molecular responses to a combination of drought and heat and cannot be directly extrapolated from the response of plants to each of these different stresses applied individually. (Pnueli et al., 2002; Rizhsky et al., 2002, 2004; Suzuki et al., 2005). Majority of the abiotic stress associated *cis* elements identified were various dehydration responsive elements such as ACGTAERD1, CBFHV, and CRTDREHVCBF2. Several MYB and MYC *cis*-elements were identified in the 5'regulatory regions in both the rice and Arabidopsis PEP genes. It has been reported that these *cis*-regulatory elements are necessary for the binding of transcriptional factors that are required for drought inducible gene expression. Drought, salinity and abscisic acid- induced members of NAC-domain transcription factors have been shown to bind at these sequences and promote expression of the downstream genes in arabidopsis and rice (Tran et al.,2004; Gao et al.,2010). Both drought and salinity causes extensive structural and physiological damage to plant cells/ tissues (Yancey et al, 1982; Mclure and Hanson, 1990). In addition, drought causes excessive transpiration water loss through the stomata pore (Lawlor and Cornic, 2002). Various low temperature responsive elements (LTR, LTRE1HVBLT49 and LTRECOREATCOR15), motifs similar to DRE which is responsible for the induction of cold-induced or regulated genes (Brown et al., 2001; Dunn et al., 1928) were identified in

both rice and Arabidopsis that cold acclimatization involved rapid cold-induced binding of the DRE or CRT binding T.Fs to the DRE leading to the expression of cold binding factor targeted gene, increasing Arabidopsis freezing tolerance.

During high temperature stress, heat shock protein helps to protect and refold cellular proteins as well as maintains a balance cellular homeostasis for plant survival (Vierling, 1991; Hartle, 1996; Larkindale and Vierling, 2008). Genes involved in these mechanisms contain HSE motifs that bind heat shock factors and are thus up-regulated during heat stress. The presence of a HSE motif in the 5'regulatory regions of rice and Arabidopsis PEP genes mat imply the ability of the genes to survive high temperature stress conditions.

With regards to the hormonal regulation associated *cis*-elements that were identified in the 2kbps upstream of the 5'regulatory region starting from the translational start site of the rice and Arabidopsis PEP genes, abscisic acid an abiotic stress hormone plays an important role in many aspects of plant development, in the regulation of stomatal aperture and in the initiation of adaptive responses to various environmental conditions (Shinozoki et al., 2000). Most regulatory regions of ABA inducible genes contain an ACGTG within 300 bp upstream of the transcription start site (Busk and Pages, 1998). It is worth noting that each of the rice and Arabidopsis PEP genes contains abscisic acid related *cis*-element and may hence be regulated by ABA. Studies have shown that ABA have antagonistic effects on various plant processes such as germination, development of seed as well as seedling growth (Finkelstein and Gibson, 2002). It may therefore be postulated that ABA regulates the expression of PEPs.

Gibberellin (GA) is an important phytohormone that participates in the regulation of many developmental processes in plants including those such as stem elongation (Richards et al., 2001), flower development and seed set (Derkx et al., 1994) that are of major importance to agriculture. Gibberellins responsive element (GARE) sequences plays a fundamental role in GA action as changes with the sequence may cause a substantial reduction in GA-driven gene expression (Gubler and Jacobsen, 1992). The presence of GARE motif in the 5' regulatory regions of some of the PEP genes may possibly be evidence for the significance of PEP in meeting energy and carbon needs of the cell during cellular development induced by gibberellins.

Auxins play a pivotal role in a variety of developmental as well as cellular processes in plants and these processes are assumed to involve Auxins-induced changes in gene expression (Hagen and Guifoyle, 2002). Some of these processes include cell division, elongation and

differentiation (Nishi et al., 1973), tropisms, root formation and apical dominance (Hagen et al., 1984; Theologis et al., 1985). ARE(auxin responsive element) was found in very low frequencies in both the arabidopsis genes and not in rice genes, thus its effects on the expression and/or regulation of these genes could either be to enhance or inhibit the genes transcription, relative to the type of plant tissue in which they are expressed.

Salicylic acid has been found to induce the expression of many defence related genes. Some of the genes' products have been identified and have direct effects on fungi and bacteria, but its role role in virus resistance still remains obscure (Carr and Klessig, 1989). It was assumed that salicylic acid assists plants in overcoming stress conditions by inducing or enhancing biochemical and physiological processes, most especially the induction of antioxidant activity which protects plants from oxidative damage (Borsani et al., 2001; Khodary, 2004). One may therefore assume that there may possibly exist a cooperative interaction between PEP's and stress response proteins as salicylic acid motifs have been identified in both rice and Arabidopsis genes with few frequency of occurrence in Arabidopsis.

Another important molecule found in plants, related to stress-signalling is methyl jasmonate. Methyl jasmonates have also been found to induce the expression of proteins/enzymes known to be in the chemical defence mechanisms of plants, proteins/enzymes such as proteinase inhibitor PI-2 (Farmer et al., 1992) and proline –rich proteins (Rushton and Somssich, 1998). Therefore, the presence of methyl jasmonate *cis*-elements in the 5' regulatory regions of some of the PEP genes may infer a possible role for PEP in wound or pathogen stress responses.

It is interesting to note that PEPs are implicated in abiotic/biotic stress responses although it is not sure what role they may play in these response mechanisms. This opens another field that one could possibly explore in the future.

3.4.3 Secondary structure, subcellular location and signal peptide prediction

A secondary structure prediction using PSIpred (Jones, 1999) was used to predict the presence of α-helices, β-sheets and coils from the putative PEP protein sequences of *Arabidopsis thaliana* and *Oryza sativa*. The comparison of AtPEP with OsPEP indicates that they are near congruent in predicted secondary structure (results not shown) which may imply common functionality. Moreover, this is expected as there was a >90% sequence

similarity between the sequences. SignalP prediction shows that PEP has no signal peptide as well as the prediction indicates that it is a non-secretory protein.

3.4.4 Tertiary structure prediction

The fold model (fig 3.4) of OsPEP was generated using the OsPEP sequence (1-730) (BAF03701) and it was based on the crystal structure of porcine brain PEP (1H2W). Building was completed using the SWISS-MODEL software (http://swissmodel.expasy.org/) (Arnold et al., 2006) and default settings. The template was selected from the SWISS-MODEL template library (SMTL), which was extracted from the protein Data Bank and the submitted OsPEP target sequence was found to have a structure identity of more than 50% with the template. One may therefore conclude that based on the structure, the putative rice PEP can be confirmed to have the same predicted structure as that of porcine brain PEP and therefore most likely to act as a functional PEP in the plant.

Figure 3.4: Modelled fold of OsPEP (red) and that of porcine brain (yellow). The OsPEP model was generated using SWISS-MODEL (Arnold et al., 2006) using the porcine brain PEP structure as a template.

3.5 Conclusion

This study, to our knowledge is the first to report work regarding *cis*-acting regulatory elements that are involved in the expression of PEP genes with regards to plant based sources. It was observed that PEP shows a relatively high sequence similarity to *Tribolium castaneum* PEP which infers a common function as the catalytic triad is conserved. With reference to the secondary and tertiary structure, this may also suggest a possibility of post-proline cleaving activity as there was a >50% structure similarity between PEP from rice and PEP from porcine brain. From the bioinformatics analysis, it has been shown that plants have functional PEP proteins that may be expressed during development or during biotic stress conditions.

3.6 References

(1) Arnold K, Bordoli L, Kopp J and Schwede T. (2006). The swiss-model workspace: A web-based environment for protein structure homology modelling. Bioinformatics. **22**:195-201.

(2) Bergeron B. (2003).The complete practical guide to bioinformatics for life scientists: Bioinformatics Computing. Pearson Education. Inc.

(3) Borsani O, Valpuesta V and Botella MA. (2001). Evidence for a role of salicylic acid in the oxidative damage generated by NaCl and osmotic stress in Arabidopsis seedlings. Plant Physiol. **126**:1024-1034.

(4) Boyer JS. (1982). Plant productivity and environment. Science **218**:443-448.

(5) Bray EA. et al. (2000). Responses to abiotic stresses. In Biochemistry and Molecular Biology Plants, pg 1158-1200.

(6) Brown APC, Dunn MA, Goddard NJ and Hughes MA. (2001). Identification of a low novel low-temperature response element in the promoter of barley (*Hordeum vulgare* L) gene blt101.1. *Planta* **213**:770-780.

(7) Busk PK and Pages M. (1998). Regulation of abscisic acid induced transcription. *Plant Molecular Biology* **37**:425-435.

(8) Carr JP and Klessig DF. (1989). The pathogenesis-related proteins of plants in Genetic Engineering, Principles and Methods 11(Setlow, J.K., ed.), pp 65-109, Plenum Press, New York.

(9) Craufurd PQ and Peacock JM. (1993). Effect of heat and drought stress on sorghum. Exp. Agric **29**:77-86.

(10) Cushman JC and Bohert HJ. (2000). Genomic approaches to plant stress tolerance. Curr Opin. Plant Biol **3**:117-124.

(11) Derkx MPM, Vermeer E and Karssen CM. (1994). Gibberellins in seeds of Arabidopsis thaliana: Biological activites, identification and effects of light and chilling on endogenous levels. Plant Growth Regulation **15**:223-234.

(12) Dunn MA, White AJ, Vural S and Hughes MA. (1998). Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (Hordeum vulgare L). Plant Molecular Biology **38**(4): 551-564.

(13) Emmanuelsson O, Brunak S, Von Heijne G and Nielson H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature Protocols. **2**:953-97.

(14) Farmer EE, Johnson RR and Ryan CA. (1992). Regulation of expression of proteinase inhibitor genes by methyl jasmonate and jasmonic acid. Plant Physiology **98**:995-1002.

(15) Finkelstein RR and Gibson SI. (2002). ABA and sugar interactions regulation development: cross-talk or voices in a crowd. Current Opinion Plant Biology **5**:26-32.

(16) Fulop V, Szeltner Z, Renner V and Polgar L. (2001). Structures of prolyl oligopeptidase substrate/inhibitor complexes. Use of inhibitor binding for titration of the catalytic histidine residue. J Biol Chem **276**: 1262–1266.

(17) Gao F., Xiong A., Peng R, Jin X, Xu J, Zhu B, Chen J and Yao Q. (2010). OsNAC52, a rice NAC transcription factor, potentially responds to ABA and confers drought tolerance in transgenic plants. Plant Cell Tissue Organ Culture **100**:255-262.

(18) Goossens F, De Meester I, Vanhoof G and Scharpe S. (1996). Distribution of prolyl oligopeptidase in human peripheral tissues and body fluids. Eur J Clin Chem Clin Biochem. **34**: 17-22.

(19) Gubler F and Jacobsen JV. (1992). Gibberellin-Responsive Elements in the promoter of a barley High-pI α-Amylase Gene. The Plant Cell **4**:1435-1441.

(20) Hagen G. and Guilfoyle T. (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Molecular Biology **49**:373-385.

(21) Hartle FU. (1996). Molecular chaperones in cellular protein folding. Nature **381**:571-579.

(22) Heyne EG and Brunson AM. (1940). Genetic studies of heat and drought tolerance in maize. J. Am. Soc. Agro **32**: 803-814.

(23) Higo K, Ugawa Y, Iwamoto M and Korenaga T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acid Research. **27**(1): 297-300.

(24) Jiang Y and Huang B. (2001). Drought and heat stress injury to two cool season turf grasses in relation to antioxidant metabolism and lipid peroxidation. Crop. Sci **41**:436-442.

(25) Jones DT. (1999). Protein secondary structure prediction based on position specific scoring matrices. Journal of Molecular Biology. **292**:195-205.

(26) Khodary SEA. (2004). Effect of salicylic acid on the growth, photosynthesis and carbohydrate metabolism in salt-stressed maize plants. Int. J. Agric. Biol **6**:5-8.

(27) Larkin MA, Blackshields G, Brown NP et al. (2007). ClustalW2 and ClustalX version 2. Bioinformatics. **23** (21): 2947-2948.

(28) Larkindale J. and Vierling E. (2008). Core genome responses involved in acclimation to high temperature. Plant Physiology **146**:748-761.

(29) Lawlor DW. and Cornic G. (2002). Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. Plant Cell and Environment **25**:275-294.

(30) McCure AF and Hanson AD. (1990). Drought and salt tolerance: Towards understanding and application. Trends in Biotechnology **8**:358:362.

(31) Moffat AS. (2002). Finding new ways to protect drought-stricken plants. Science **296**:1226-1229.

(32) Nishi T, Yamanda Y and Takahashi E. (1973). The role of auxins in differentiation of rice tissues culture in vitro. Journal of Plant Research **86**(3):183-188.

(33) Pnueli L et al. (2002). Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume Retama raetam. Plant. J **31**:319-330.

(34) Polgar L. (2002). The prolyl oligopeptidase family. Cell. Mol. Life Sci. **59**: 349–362.

(35) Richards DE, King KE, Ait-ali T and Harberd NP. (2001). How gibberellins regulates plant growth and development: A molecular genetic analysis of gibberellins signalling. Annual Review of Plant Physiology and Plant Molecular Biology **52**: 67-88.

(36) Rizhsky L et al. (2002). The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol **130**:1143-1151.

(37) Rizhsky L et al. (2004). When defense pathways collide: the response of Arabidopsis to a combination of drought and heat stress. Plant Physiol **134**: 1683-1696.

(38) Rushton PJ and Somssich IE. (1998). Transcription control of plant genes responsive to pathogens. Current Opinion in Plant Biology **1**:311-315.

(39) Shan L, Marti T, Sollid LM, Gray GM and Khosla C. (2004). Comparative biochemical analysis of three bacterial prolyl endopeptidases: implications for coeliac sprue. Biochem J **383**: 311–318.

(40) Shinozaki K and Yamaguchi-Shinozaki K. (2000). Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signalling pathways. Current Opinion in Plant Biology **3**:217-223.

(41) Suzuki N et al. (2005). Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional co-activator Multi-protein Bridging Factor 1c. plant Physiol **139**:1313-1322.

(42) Tran L-SP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita, M, Seki M, Shinozaki K and Yamaguchi-Shinozaki K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-

responsive *cis*-element in the early responsive to dehydration stress 1 promoter. Plant Cell **16**:2481-2498.

(43) Venalainen JI, Juvonen RO and Mannisto PT. (2004). Evolutionary relationships of the prolyl oligopeptidase family enzymes. Eur. J. Biochem. **271**: 2705–2715.

(44) Vierling E. (1991). The roles of heat shock proteins in plants. Annual Review of Plant Physiology and Plant Molecular Biology **42**:579-620.

(45) Vilo J, Brazma A, Jonassen I, Robinson A and Ukkonen E. (2000). Mining for putative regulatory elements in the yeast genome using gene expression data. Proceedings of International Conference on Intelligent Systems for Molecular Biology **8**:384-394.

(46) Wang Z, Dalkilic M and Kim S. (2004). Guiding motif discovery by iterative pattern refinement. ACM Symposium on Applied Computing, Nicosia, Cyprus. pg 162-166.

(47) Wolters MV and Wijmenga C. (2008). Genetic background of celiac disease and its clinical implications. Am J Gastroenterol. **103**(1):190-195.

(48) Yancey P, Lark MHS, Bowlus R. and Somero G. (1982). Living with water stress: Evolution of osmolyte systems. Science **217**:214-1222.

List of websites used:

(1) http://bioinfo.cs.ucl.ac.uk/psipred

(2) http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

(3) http://blast.ncbi.nlm.nih.gov/Blast.cgi

(4) http://swissmodel.expasy.org/)

(5) http://www.cbs.dtu.dk/services/TargetP/

(6) http://www.dna.affrc.go.jp/PLACE/signalscan.html

(7) http://www.ebi.ac.uk/tools/clustalw2/index.html

(8) http://www.ncbi.nlm.nih.gov/Entrez

CHAPTER 4

DETECTION AND OPTIMIZATION OF PEP EXPRESSION DURING GERMINATION OF WHEAT SEEDS

4.1 Introduction

The cereal prolamins have been found to be unique proteins in many respects. The wheat prolamins (glutenins and gliadins) are the major storage proteins of wheat grain and make up about 70-80 % of proteins found in wheat (Osborne 1907). The term prolamin refers to the proline (Pro) and glutamine (Gln) rich alcohol-soluble proteins that are typically found in cereals. Generally, the presence of a proline residue in an amino acid chain causes its conformation to be inflexibile, and thus renders itself resistant to proteolysis. Since the major cereal storage proteins, the prolamins, contain large amounts of proline, it is obvious that the cereal plant itself must possess tools for degrading proteins containing proline. Mikola (1986) has shown that wheat grain contains proline specific carboxypeptidases; the carboxypeptidases IV and V and are present in the endosperms of wheat grain. Tsuji et al., (2004) has also identified a serine oligopeptidase from wheat embryos that hydrolysed peptide bonds on the carboxyl side of basic amino acids. This chapter examines protein behaviour with specific reference to Prolyl endopeptidase that occurs during wheat germination. The chapter also looks at the differences between a buffered and non-buffered extraction as well as looking at the behaviour of PEPs through the use of various buffer systems.

4.2 Materials and Methods

4.2.1 Materials

Wheat grains (*Triticum aestivum*) were obtained from Kei Seeds and Feeds (PTY) Ltd. (East London, South Africa). All chemicals were purchased from Merck chemicals (PTY) Ltd. and were of analytical grade.

4.2.2 Method

Wheat seeds were germinated in a controlled environment (conviron) for a several days. The conditions in the conviron were 16hrs of light at 26^0C and 60% humidity. 1-10 g of seeds were frozen under liquid nitrogen, ground with a mortar and pestle to powder and 100 ml of extraction buffer (50 mM KCl, 10 mM Tris-HCl, pH 7.6) added. Extraction was allowed to proceed for 1hr under continuous stirring at 4^0C. The extract was filtered through glass wool and the filtrate centrifuged (using a JA10, Beckman Coulter rotor, 10000 rpm, 30 min) at 4^0C. The supernatant was analysed (SDS-PAGE, total protein and total activity). Aliquots were stored at -70^0C.

4.3 PEP activity

After each day's extraction, each fraction (145 µl) was mixed with 150 µl of universal buffer, 100 mM acetate–phosphate–borate (UB) of pH 7.9 in a 96-well microtiter plate using a multichannel pipette. The reaction was initiated by adding 5 µl of 6 mM Z-Gly-Pro-pNa dissolved in dimethylsulfoxide. The reaction mixture was incubated at 37^0C and the progress of activity was measured by monitoring the increase in absorbance of p-nitroanilide release at 405 nm in a micro plate reader (SynergyMx, Bio tek Instruments, USA) for 24 hours.

4.4 Total Protein Determination

Determination of protein content was done according to the bicinchoninic acid (BCA) method as well as the Qubit® Protein Assay Kits for use with the Qubit® 2.0 Fluorometer

4.5 Results and Discussion

4.5.1 Development of total protein extraction

Pilot run 1: Various total protein extraction conditions were evaluated to optimise the yield of PEP extracts. Approximately 20 volumes of germinated wheat seeds in 100 volume of extraction buffer (50 mM KCl, 10 mM Tris-HCl, pH 6.8) were used.

The expression profile of PEP shows very low activity levels and a lot of variability (fig 4.2). However, the total protein extraction appeared to be intact with no protein degradation as distinct bands were observed on analysing the SDS-PAGE (fig 4.1). We also started off with an extraction at pH 6.8 using Tris as the extraction buffer. Total protein yield for this extraction was 49 mg.

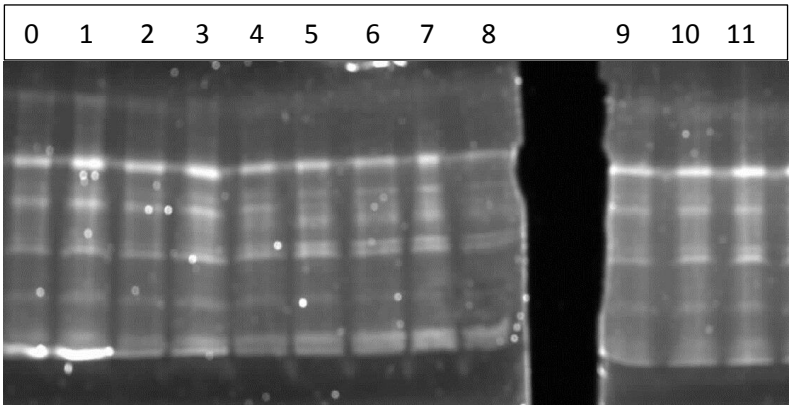


Figure 4.1: SDS-PAGE profile of proteins extracted from day 0 up to day 11(as indicated on the gel) germinating wheat seeds using total protein extraction of 20 volume of seed in 10 volume 50 mM KCl, 10 mMTris/HCl, pH 6.8.

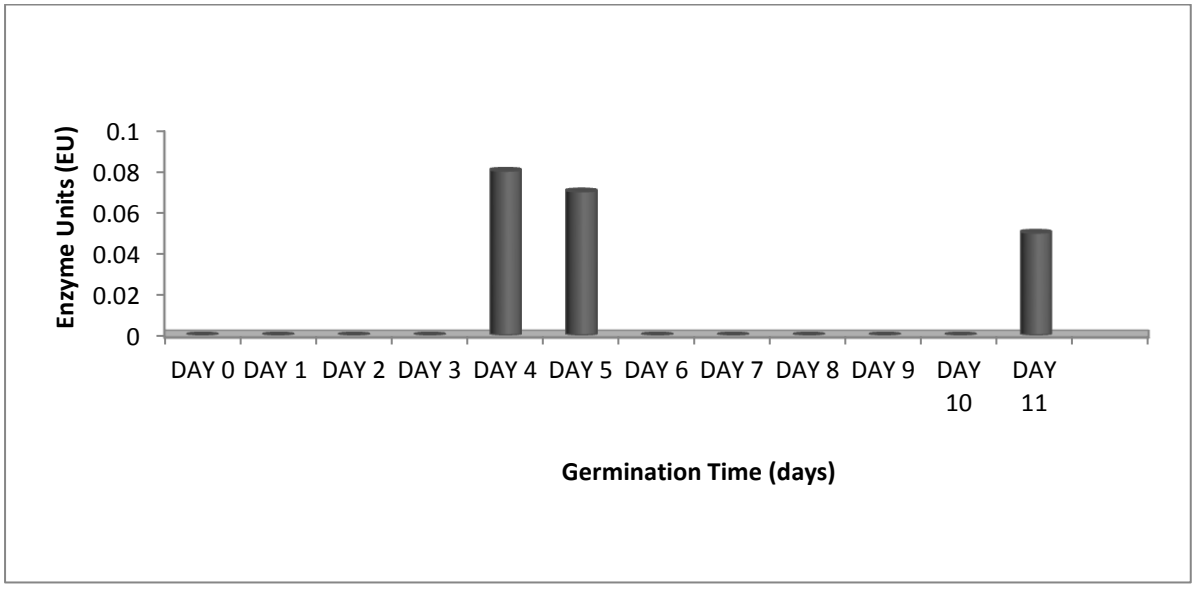


Figure 4.2: Changes observed in the expression of PEP activity during wheat seed germination over 11 days. One unit (U) of enzyme activity was defined as the amount of enzyme that results in the change of absorbance @ 405 nm of 0.001 when hydrolyzing 0.1mM (final concentration in the reaction mixture) Z-Gly-Pro-pNa substrate at 37^0C in 50mM UB, pH 7.9.

Pilot run 2: Comparison between using 20 volumes of germinated wheat seeds in 100 volume extraction buffer (50 mM KCl, 10 mM Tris-HCl, pH 6.8) and using 20 volumes of germinated wheat seeds in 100 volumes water.

On comparing the activity profiles of the buffered system with those of the unbuffered system, it would appear that the unbuffered extraction gave a better result of more PEP activity as well as activity over more days. Although the buffered system yielded variable results, there was no protein degradation observed when the buffered extraction was used (fig 4.3 a) whereas the unbuffered extraction showed extensive protein degradation (fig 4.3 b).

(a) (b)

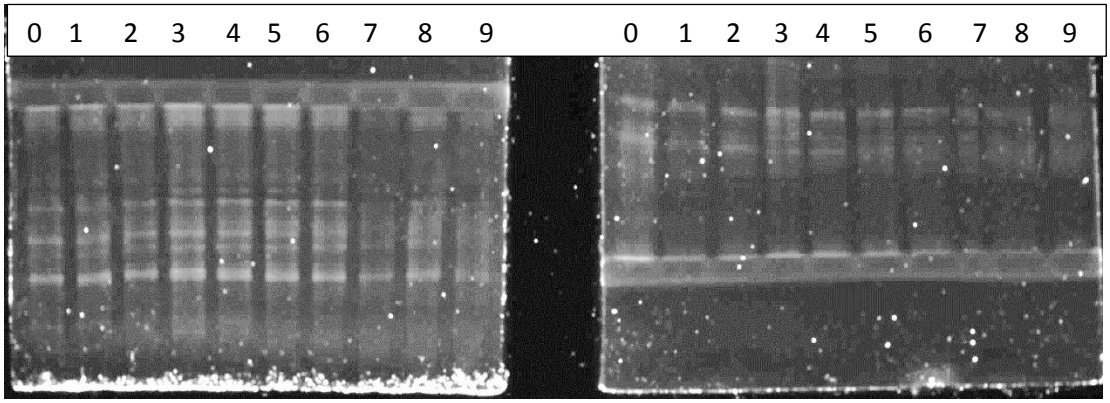


Figure 4.3: SDS-PAGE profile of proteins extracted from day 0 up to day 9(as indicated on the gel) germinating wheat seeds using total protein extraction of 20 volume of seed in 10 volume 50 mM KCl, 10 mM Tris/HCl, pH 6.8. The buffered system is on the right (a) and the non-buffered system is on the right (b).

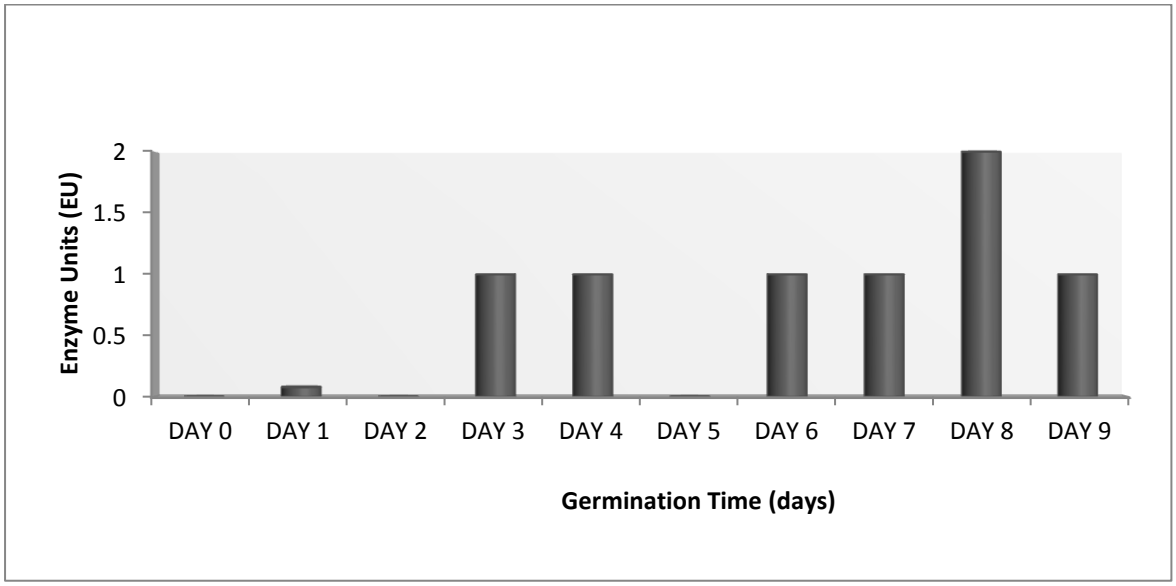


Figure 4.4: Changes observed in the expression of PEP activity during wheat seed germination over 9 days for an unbuffered system. One unit (U) of enzyme activity was defined as the amount of enzyme that results in the change of absorbance @ 405 nm of 0.001

when hydrolyzing 0.1 mM (in the reaction mixture) Z-Gly-Pro-pNa substrate at 37^0C in 50 mM UB, pH 7.9.

Pilot run 3: 200 germinating wheat seeds in 100 ml of 50 mM KCl, 10 mM Tris-HCl, pH 7.2. Higher levels of PEP activity were detected when using an extraction buffer of pH 7.2 instead of 6.8 (fig 4.5). The total protein yield also improved to 80.6 mg. The SDS-PAGE of the total protein extracts also showed that the protein remained intact (Data not shown). An increasing pH showed an increased yield and stability of the PEP activity across all 11 days of germination. It was decided to increase the pH to 7.6 as well as to change the buffer system from Tris/HCl to phosphate due to the higher pH.

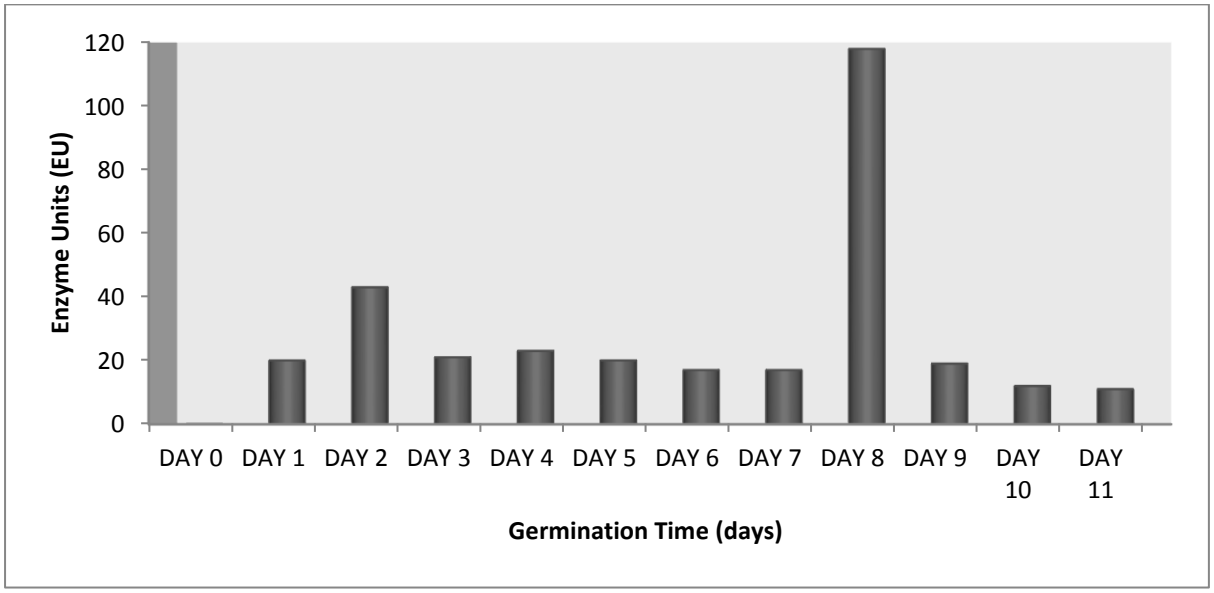


Figure 4.5: Changes observed in the expression of PEP activity during wheat seed germination over 11 days. One unit (U) of enzyme activity was defined as the amount of enzyme that results in the change of absorbance @ 405nm of 0.001 when hydrolyzing 0.1mM (in the reaction mixture) Z-Gly-Pro-pNa substrate at 37^0C in 50 mM UB, pH 7.9.

Pilot run 4: As a final pilot run in isolating the enzyme PEP, 200 germinated wheat seeds in 100 ml of extraction buffer were used. However, a slight modification was made with the initial buffer system used. The buffer system used this time was 50 mM KCl, 1mM $Na_2HPO_4.$ $2H_2O$, pH 7.6.

50 mM KCl, 1 mM $Na_2HPO_4.$ $2H_2O$, pH 7.6 as an extraction buffer resulted in the highest level of PEP activity with PEP activity showing increased levels from day 0 to day 4 after which it remained at a consistent level. The total protein yield also increased to 154 mg as

well as the SDS-PAGE of the total protein extracts also showed that the protein remained intact in that several bands were visible(Data not shown). It was therefor decided to use this buffer and 4/5 days germination for extraction of total protein /PEP activity for all isolations

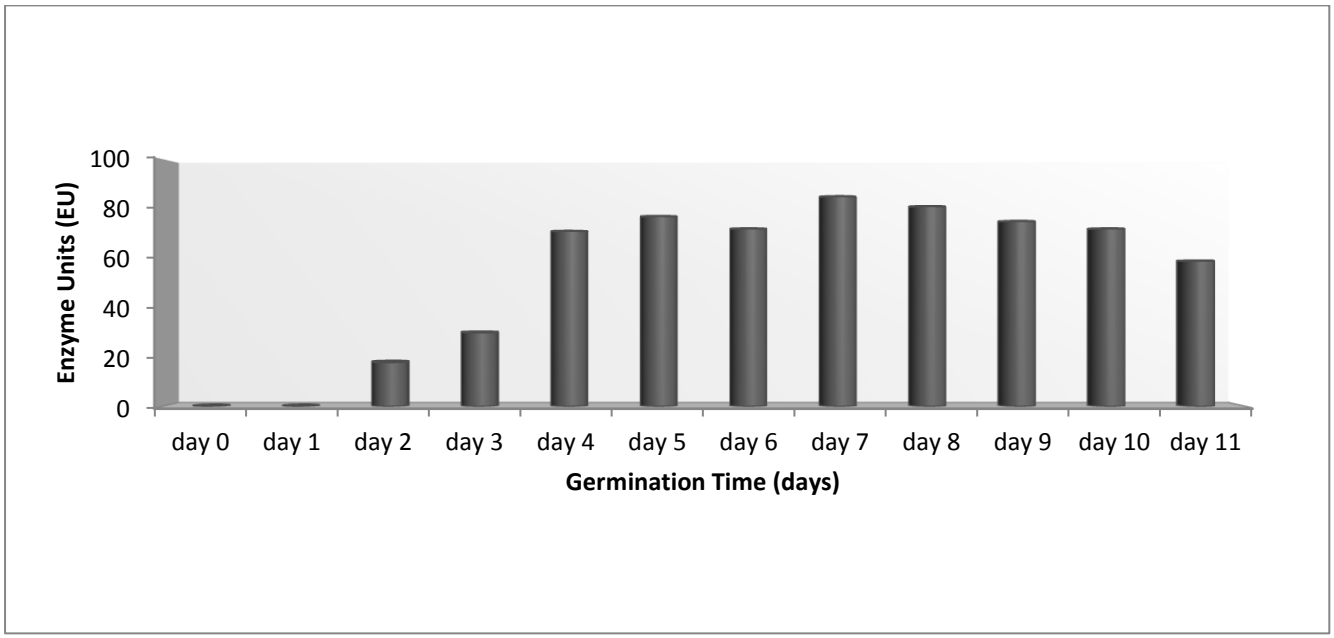


Figure 4.6: Changes observed in the expression of PEP activity during wheat seed germination over 11 days. One unit (U) of enzyme activity was defined as the amount of enzyme that results in the change of absorbance @ 405 nm of 0.001 when hydrolyzing 0.1 mM (in the reaction mixture) Z-Gly-Pro-pNa substrate at 37^0C in 50 mM UB, pH 7.9.

4.6 References

1) Osborne T B. (1907). The Proteins of the Wheat Kernel; publication 84; Carnegie Institution: Washington, DC.

2) Mikola L. (1986). Acid carboxypeptidases in grains and leaves of wheat, *Triticum aestivum* L. Plant Physiol.**81**: 823-829.

3) Tsuji A, Yuasa K and Matsuda Y. (2004). Identification of oligopeptidase B in higher plants. Purification and characterization of oligopeptidase B from quiescent wheat embryo, *Triticum aestivum*. J. Biochem. **136**:673-681.

CHAPTER 5

ISOLATION AND PURIFICATION OF PEP

5.1 Purification methods

5.1.1 Chromatographic instrument

An AKTA fast protein liquid chromatography system (pump P-920, monitor UPC-900 and fraction collector Frac-920) with ultra-violet (UV), pH and conductivity sensors was used. The system was controlled by Unicorn software version 5.1. Sample injections were done using 10ml sample loops (Pharmacia). Protein detection was done by reading the absorbance of the eluate at 280nm. Peak fractions (1ml) were collected automatically.

5.1.2 Column Chromatography

All purification steps performed were carried out at 4^0C and purity of the fractions was evaluated by SDS-PAGE and PEP activity of every fraction collected.

5.1.2.1 Anion exchange chromatography on a HiPrep DEAE FF16/10 column

Total protein extracts of varying volumes (50-100 ml) were applied separately onto a HiPrep DEAE FF16/10 column (GE HEALTH CARE). The column was pre-washed with 100 ml (5 column volumes (CV)) start buffer (low ionic strength) and equilibrated until A_{280} of elute reached baseline. The flow rate was set at 5ml/min and samples were eluted using a salt gradient (0-1M NaCl in equilibration buffer over 20 CV). Eluted peaks were collected in 1ml fractions and screened for PEP activity.

5.1.2.2 Cation exchange chromatography on HiPrep CM FF16/10 column

Total protein extracts of varying volumes (50-100 ml) were applied separately onto a HiPrep CM FF16/10 (GE HEALTH CARE). The column was pre-washed with 100 ml (5 CV) equilibration buffer and equilibrated until A_{280} of elute reached baseline. The flow rate was set at 5 ml/min and samples were eluted with a salt gradient 0-1M NaCl in equilibration buffer over 20 CV (400 ml). Eluted peaks were collected in 1ml fractions and screened for PEP activity

5.1.2.3 Size- exclusion chromatography on HiPrep 16/60 Sephacryl S-100 High Resolution column

Concentrated samples (5 ml) were applied separately onto a HiPrep 16/60 Sephacryl S-100 High Resolution column (GE HEALTH CARE). The column was pre-washed with equilibration buffer (with no salt gradient) and equilibrated until A_{280} of elute reached

baseline. The flow rate was set at 0.5 ml/min and samples were eluted with the same buffer. Fractions (1 ml) were collected and screened for PEP activity.

5.1.3 Ultra-filtration

Amicon® Ultra- 15 centrifugal filter devices were used to reduce the volume of the pooled fractions from ion exchange chromatography before preparations of the gel exclusion columns. The device provides fast ultra-filtration, with the capability for high concentration factors.

5.1.4 Ammonium sulphate precipitation

Concentration of fractions is necessary prior to gel exclusion chromatography as the Sephacryl S100 columns can only resolve proteins in a total volume of less than 5ml. Pooled fractions were precipitated by slow addition of finely ground ammonium sulphate to 80% saturation with constant stirring on a magnetic stirrer at 4^0C over a period of 4 hours. The precipitates were collected by centrifugation (JA10, Beckman Coulter rotor, 10000 rpm, 4^0C for 30 min). The precipitated proteins were re-suspended in 5ml buffer (20 mM Tris-HCl buffer, pH 5.0).

5.1.5 Dialysis

The resuspended A.S fraction was sealed in dialysis tubing (cut off less than 10 000) and dialyzed overnight at 4^0C in 400 times sample volume of 20 mM Tris-HCl buffer (pH 5.0) in a 2L measuring cylinder to ensure complete removal of ammonium sulphate and contaminating salts.

5.1.6 PEP activity

After each chromatography run, each fraction (145 µl) was mixed with 150 µl of universal buffer (100 mM acetate–phosphate–borate buffer, pH 7.9) in a 96-well microtiter plate using a multichannel pipette. The reaction was initiated by adding 5 µl of 20 mM Z-Gly-Pro-pNa dissolved in dimethylsulfoxide. The reaction mixture was incubated at 37^0C and the progress of activity was measured by monitoring the increase in absorbance of p-nitroanilide release at 405 nm in a micro plate reader (SynergyMx, Bio tek Instruments, USA) for 24 hours.

5.1.7 Protein determination

Determination of protein content was done using the bicinchoninic acid (BCA) method (Smith et al, 1985) as well as the Qubit® Protein Assay Kits (1999) for use with the Qubit® 2.0 Fluorometer

5.1.8 Determination of purity

In order to monitor the progress of purification of the enzyme, sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on fractions exhibiting PEP activity. Samples (20 µl) from each purification step dissolved in 10 µl sample buffer and standard molecular weight (5 µl, Lonza Group Ltd) were applied on a 12% resolving and 4% stacking SDS-PAGE and electrophoresed at 200 V for 45 min.

The gels were stained with various stains including Oriole fluorescent gel stain, SYPRO ruby protein gel stain and silver staining solution all from Bio-rad laboratories, Inc

5.2 Results and Discussion

All protein extracts used in the following pilot runs were prepared by using 200 seeds in 100 ml extraction buffer (50 mM KCl, 1 mM $Na_2HPO_4.2H_2O$, pH 7.6).

Pilot run 1: Ion exchange was evaluated as the first chromatography step in the purification of PEP. 50 ml of protein extract was clarified by centrifugation (using a JA10, Beckman Coulter rotor, 10 000rpm, 10 min, 4^{o}C). A HiPrep DEAE FF16/10 column was equilibrated with 5 column volumes (CV) of buffer A: (20 mM Tris-HCl, pH 6.8) before application of the clarified protein extract. The column was developed with 5 CV of buffer A until A_{280} reached baseline. Bound protein was eluted with a salt gradient (0-1M NaCl in buffer A) over 20 CV (figure 5.1).

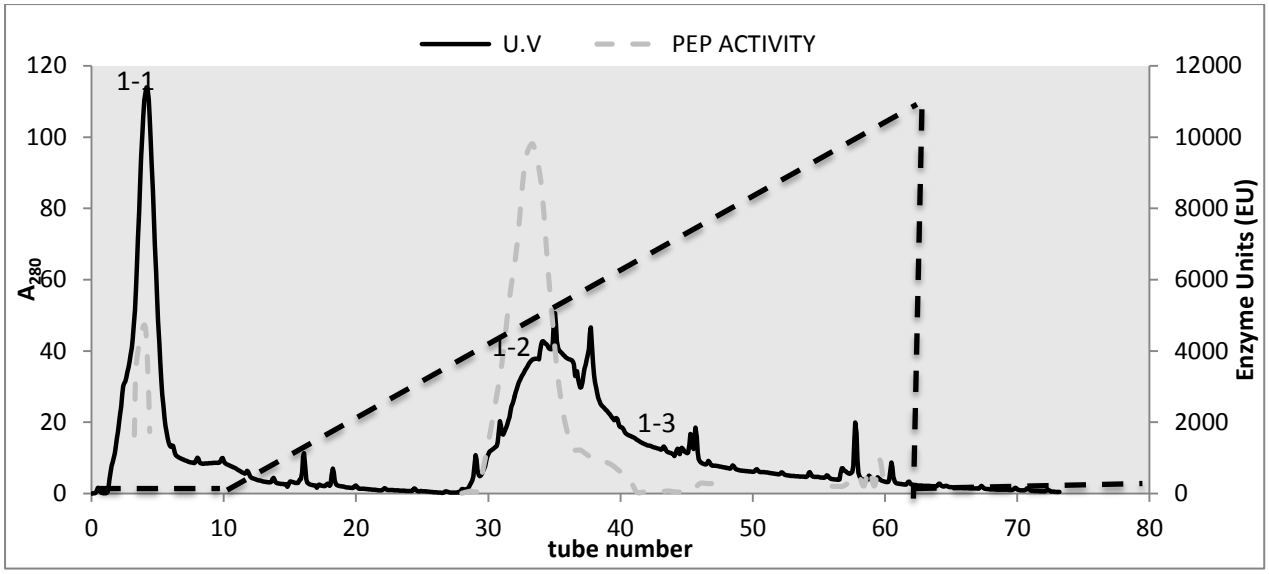


Figure 5.1: Purification of PEP from wheat seeds on a HiPrep DEAE FF16/10 column (FPLC). Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Black dashed line: gradient 0-1M NaCl in equilibration buffer. The numbers depict pooled fraction. Fractions containing PEP activity were pooled (fractions 1-1, 1-2, and 1-3) and submitted to SDS-PAGE to evaluate the purity of the fractions (fig 5.2)

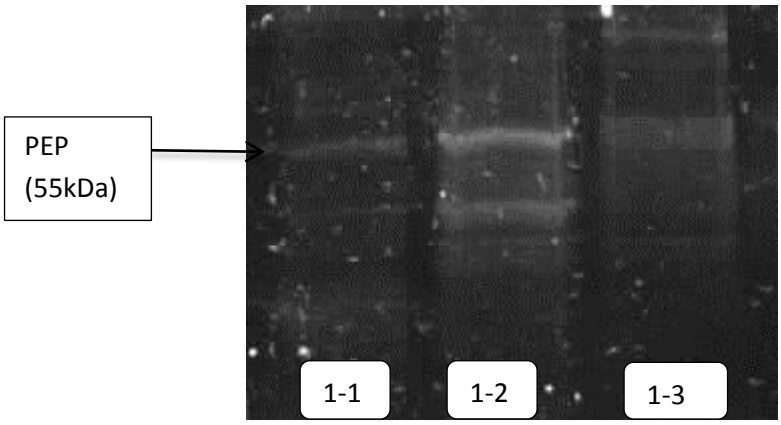


Figure 5.2: SDS-PAGE profile of proteins after chromatography on DEAE FF16/10 column. Lanes 1, 2 and 3 represent pooled fractions as indicated on the A_{280} profile (designated as 1-1, 1-2 and 1-3) in figure 5.1.

Three peaks containing PEP activity were obtained from the DEAE chromatography with one peak eluting before applying the salt gradient. This could reflect the presence of three different PEP proteins in the extract or the first peak could reflect overloading of the column.

The latter is unlikely though since Hi Prep DEAE FF16/10 has a binding capacity of 2.2 g protein and less than 100 mg total protein was loaded, well under the binding capacity of the column. From the SDS-PAGE, two prominent bands were observed around 50-70 KDa and these appear to be in line with the expected Molecular weights of PEPs. There were also several bands of lower molecular weight present in fraction 1-2. To remove these lower molecular weight proteins, fraction 1-2 was subjected to size exclusion chromatography of a Hi Prep Sephacryl S100 column.

Fractions 1-2 from the DEAE column containing PEP hydrolytic activity was concentrated with Amicon® Ultra- 15 centrifugal filter devices to 5 ml total volume and subjected to gel filtration on a HiPrep 16/60 Sephacryl S-100 High Resolution column equilibrated with 20 mM Tris-HCl, pH 6.8 buffer. Fractions were pooled where PEP activity was observed and pooled fractions were subjected to SDS-PAGE before storage at -70^{0}C.

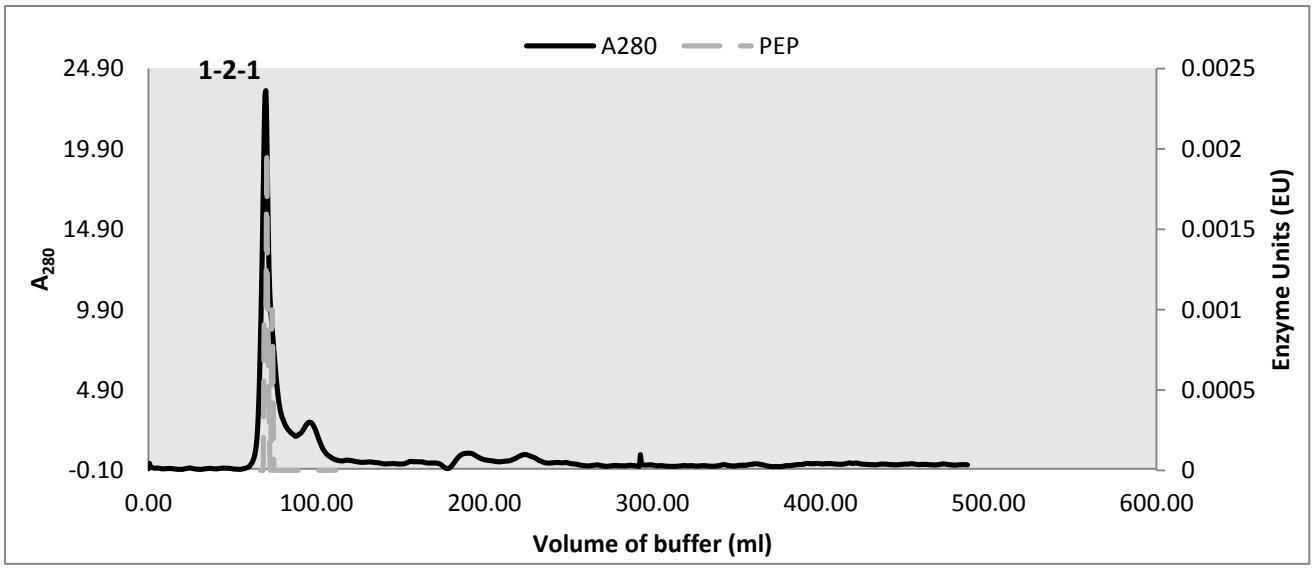


Figure 5.3: Purification of PEP from wheat seeds on a HiPrep 16/60 Sephacryl S-100 High Resolution column (FPLC). Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Fractions containing PEP activity were pooled as indicated by numbers (1-2-1) and submitted to SDS-PAGE to evaluate the purity of the fractions (fig 5.4)

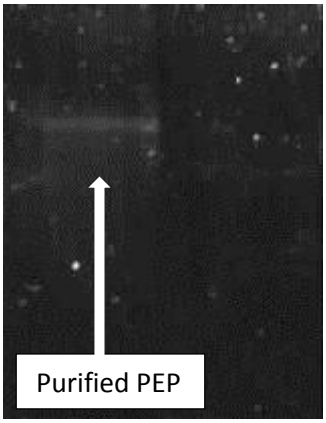


Figure 5.4: SDS-PAGE profile of purified PEPs after chromatography on a HiPrep 16/60 Sephacryl S-100 High Resolution column. The Lane designated as purified PEP represent pooled fractions (1-2-1) as indicated on the A_{280} profile in figure 5.3.

One major peak eluted at around 4-5 CV with several smaller peaks eluting later. The first peak containing PEP activity was pooled (fraction 1-2-1). SDS-PAGE of this fraction revealed one major band around 55 KDa.

Pilot run 2: 100 ml of clarified extract was loaded onto a HiPrep CM FF16/10 column equilibrated with 20 mM Na_2HPO_4. $2H_2O$ pH 7.6. The column was washed with 5 CV of equilibration buffer A and fractions eluted with a linear salt gradient (0-1M NaCl in buffer A) over 20 CV.

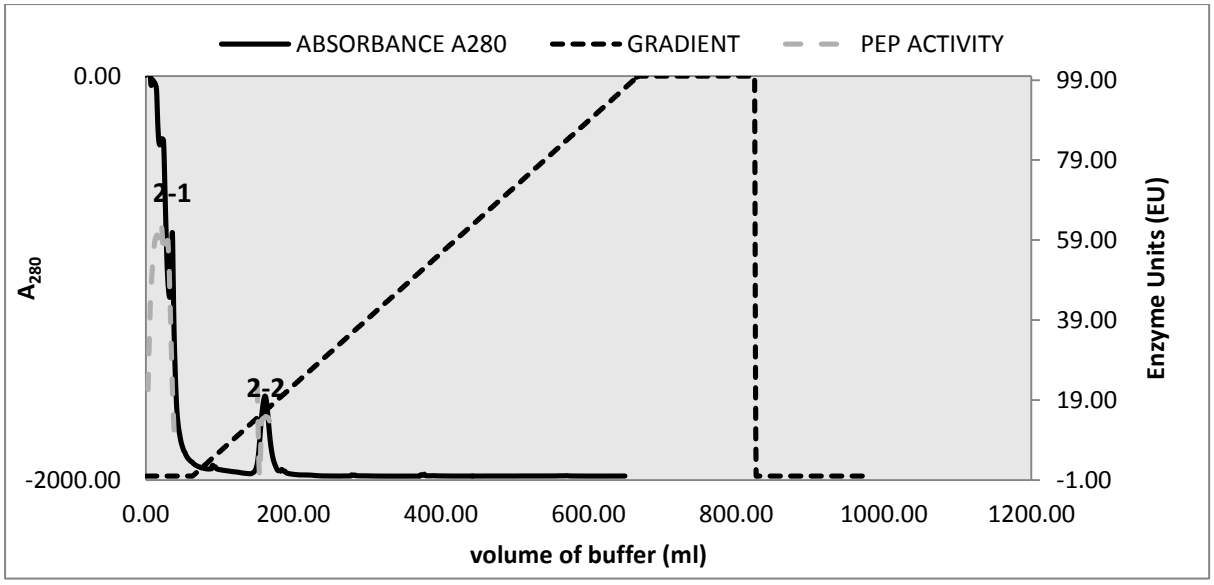
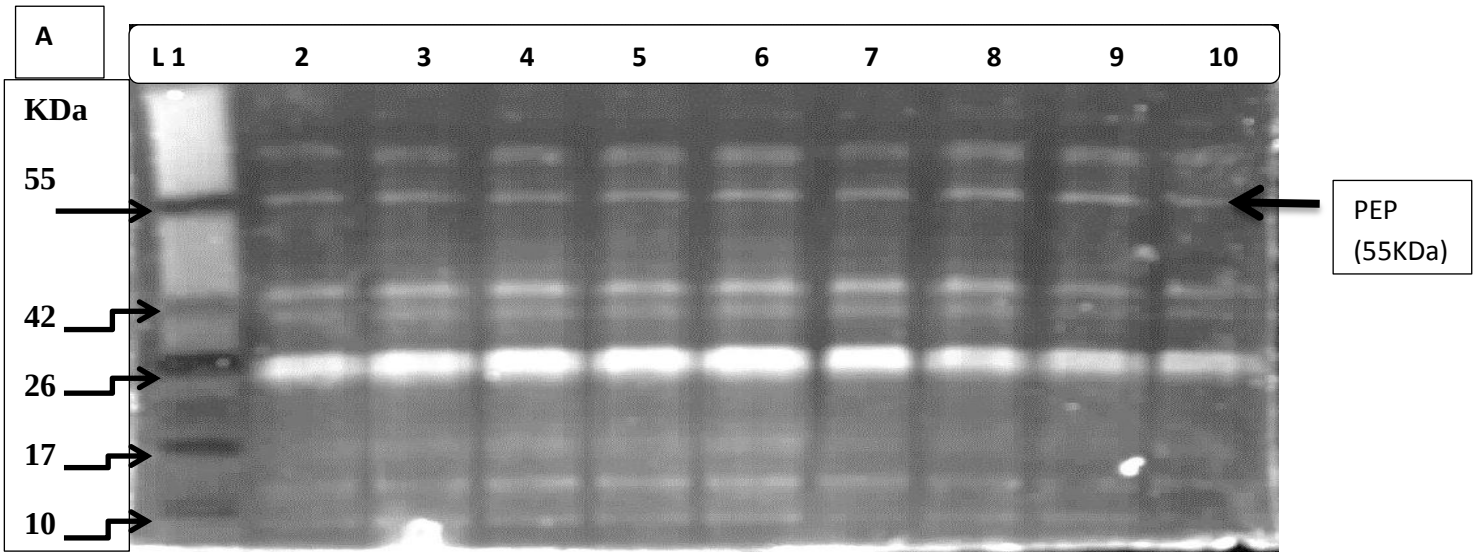


Figure 5.5: Purification of PEP from wheat seeds on a HiPrep CM FF16/10 column equilibrated with buffer A (20 mM Na_2HPO_4. $2H_2O$ pH 7.6) and developed with linear salt gradient (0-1M NaCl) over 20 CV. Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Dashed black line: salt gradient. Numbers (2-1 and 2-2) reflect pooled fractions.



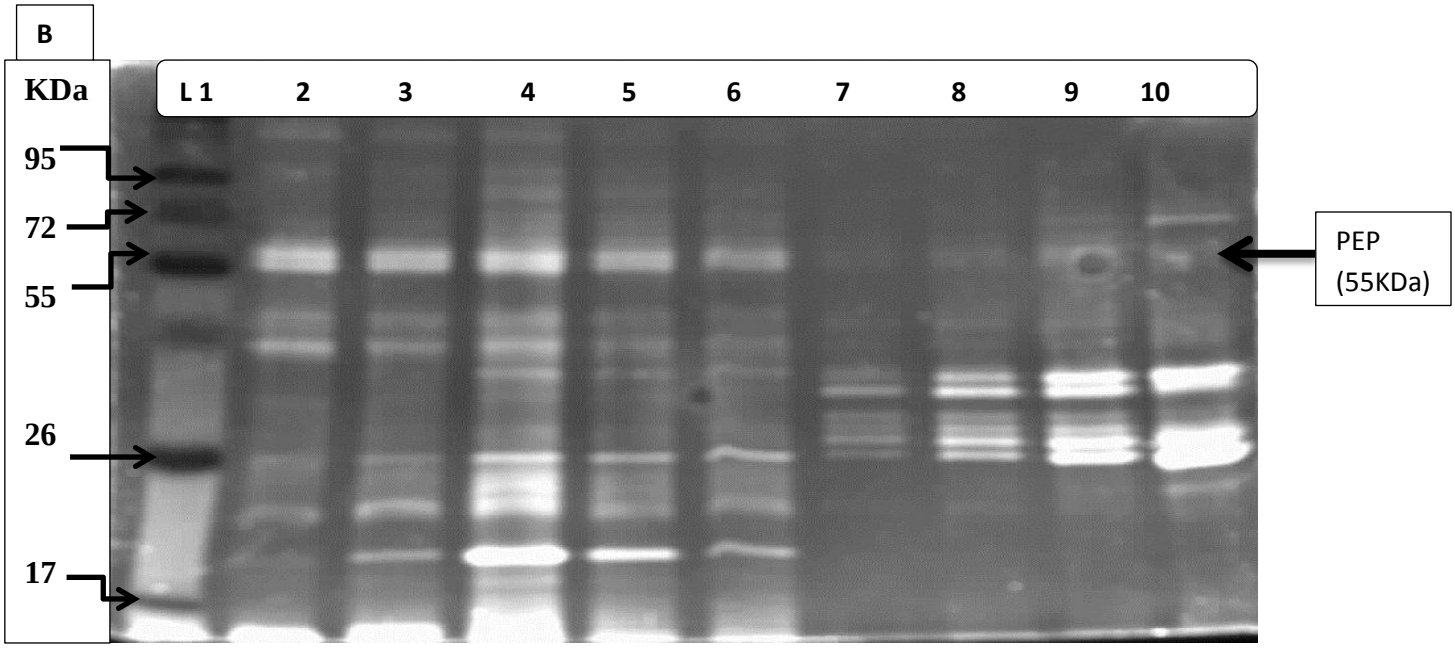


Figure 5.6a and b: SDS-PAGE profile of proteins after chromatography on HiPrep CM FF16/10 column. Lane 1 represents the molecular weight marker, lanes 2 to 10 represents fractions in the region designated as 2-1 indicated on the A_{280} profile in figure 5.5. Every third fraction was screened for PEP activity in triplicate. Arrow indicates the expected molecular weight of PEP.

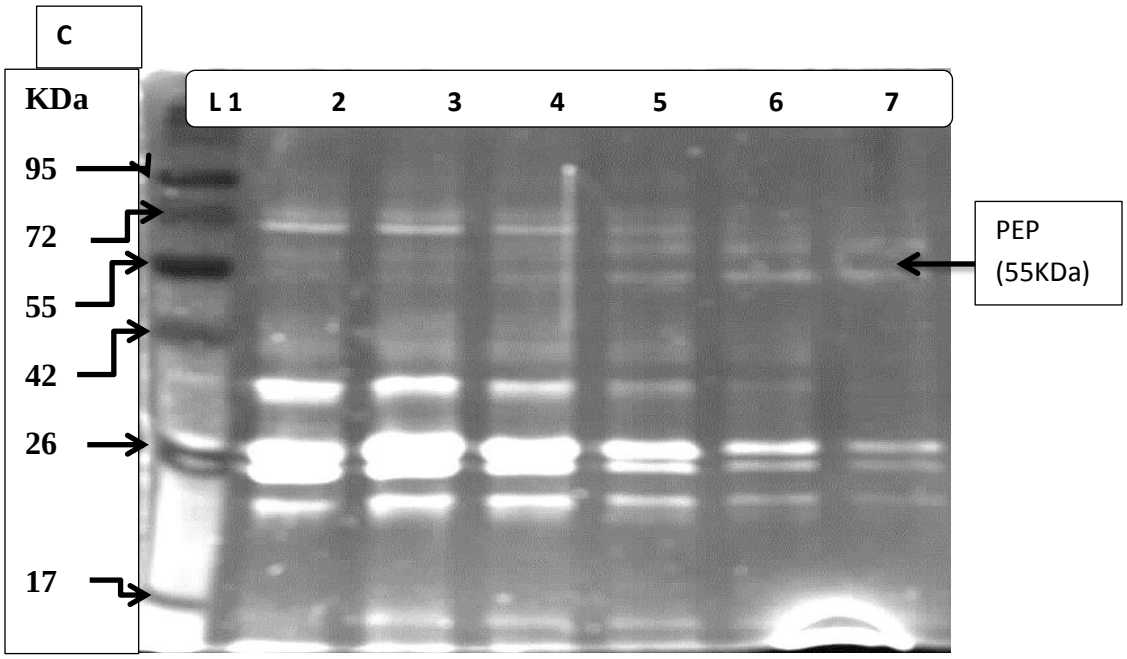


Figure 5.6c: SDS-PAGE profile of proteins after chromatography on HiPrep CM FF16/10 column. Lane 1 represents the molecular weight marker, lanes 2 to 10 represents fractions in the region designated as 2-2 indicated on the A_{280} profile in figure 5.5. Every third fraction was screened for PEP activity in triplicate.

Several peaks were obtained from the DEAE chromatography but these were split into two major pools containing PEP activity with one of these major pools eluting before applying the salt gradient. This could possibly reflect the presence of different PEP proteins in the extract or the first portion could reflect overloading of the column. The latter is unlikely though since Hi Prep CM FF16/10 has the same binding capacity as the Hi Prep DEAE FF16/10 and only 120mg total protein was loaded, which is well under the binding capacity of the column. From the SDS-PAGE (fig 5.6 a-c), several bands were observed indicating that quite a lot of other possible proteins were present and still intact. The major band around 55kDa for PEP was also still present. There was also a prominent band around 26 KDa which is exactly half of the expected mass of PEP (55 KDa). This may possible represent an additional subunit but this would have to be further investigated. There were also several bands of lower molecular weight present in and to remove these lower molecular weight proteins, fraction 2-1 was subjected to size exclusion chromatography of a Hi Prep Sephacryl S100 column.

Pooled fractions (2-1) was concentrated to Amicon® Ultra- 15 centrifugal filter devices 5ml and subjected to gel filtration on a HiPrep 16/60 Sephacryl S-100 High Resolution column in 20mM Na_2HPO_4. $2H_2O$ pH 7.6 at 0.5 ml/min. Fractions where activity was observed were subjected to SDS-PAGE before being pooled and stored at -70^{0}C.

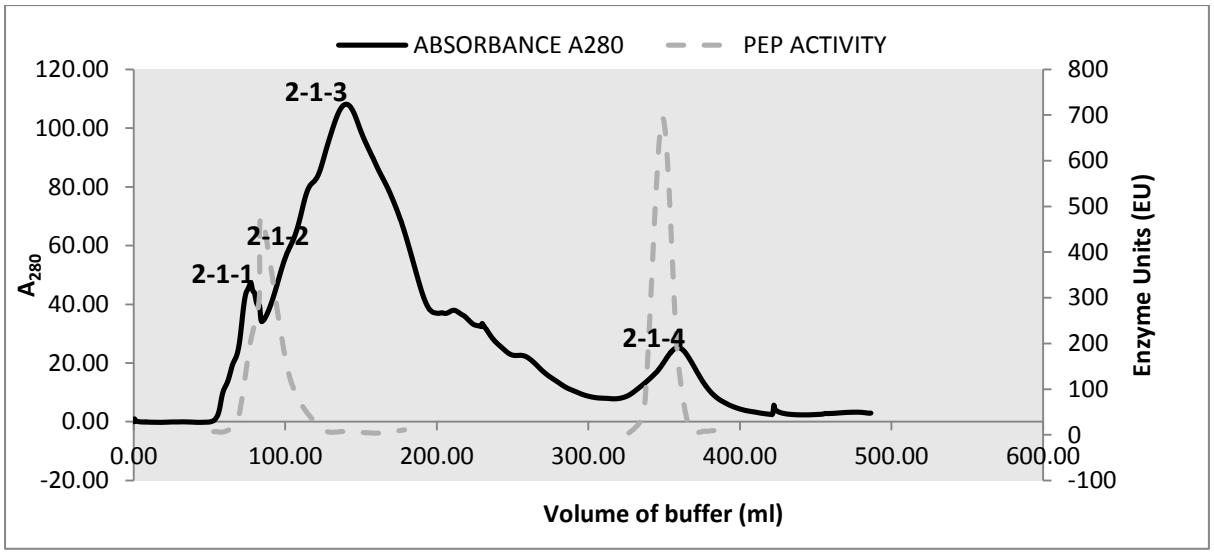


Figure 5.7: Elution profile of pooled fraction (2-1) subjected to size exclusion chromatography on a HiPrep 16/60 Sephacryl S-100 equilibrated with 20 mM Na_2HPO_4. $2H_2O$ pH 7.6. Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined

spectrophotometrically at 280 nm. Fractions were pooled as indicated by numbers 2-1-1 to 2-1-4.

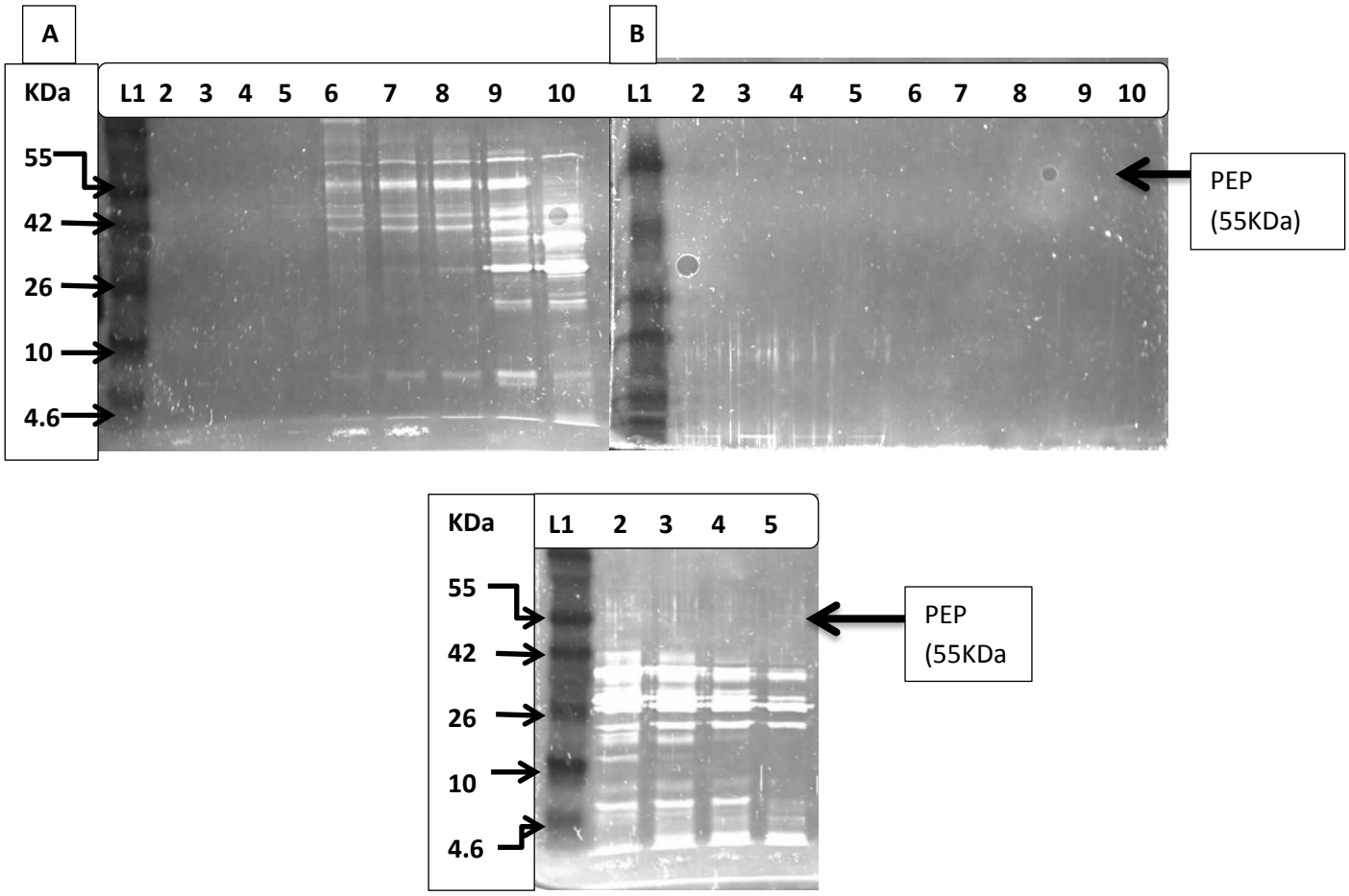


Figure 5.8a, b and c: SDS-PAGE profile of proteins peaks eluted after chromatography on a HiPrep 16/60 Sephacryl S-100 High Resolution column (FPLC). Lane 1 represents the molecular weight marker, lanes 2 to 5 (fig 5.8a) represents fractions in the region designated as 2-1-1; lanes 6 to 10 (fig 8a) represents fractions 13 to 40 in the region designated as 2-1-2 as indicated on the A_{280} profile in figure 5.7. Lanes 2 to 10 (fig 8b) represents fractions 48 to 68 in the region designated as 2-1-3. Lanes 2 to 5 (fig 8c) represents fractions 158 to 162 in the region designated as 1-2-4 as indicated on the A_{280} profile in figure 5.7. Every third fraction was screened for PEP activity in triplicate

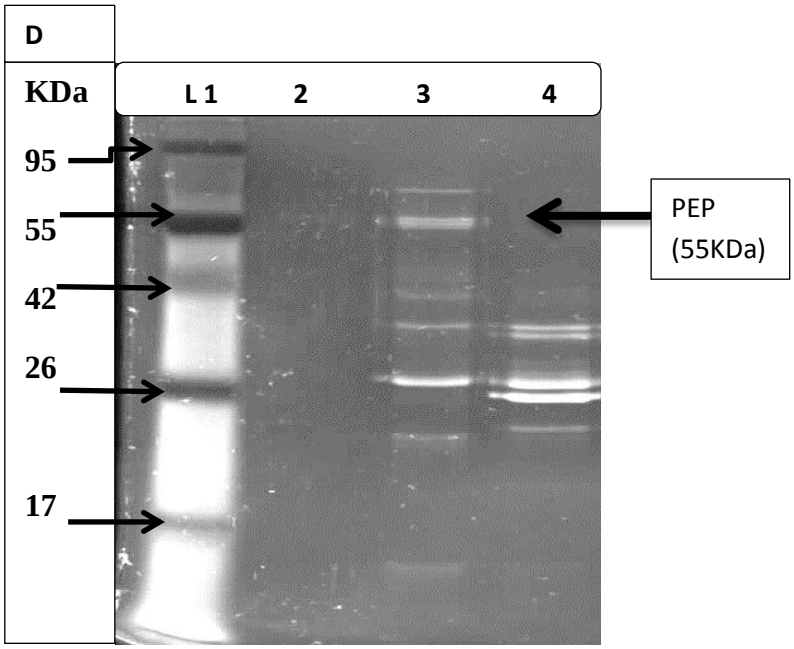


Fig 5.8d: Pooled fractions from peaks 2-1-1, 2-1-2 and 2-1-4 as indicated on the A_{280} profile in figure 5.7. Lane 1 represents the molecular weight marker, lane 2 represents pooled fractions 2-1-1, lane 3 represents pooled fractions 2-1-2 and lane 4 represents pooled fractions 2-1-4 in the region designated as 4 as indicated on the A_{280} profile in figure 5.7.

Four major peaks eluted at around 4, 5, 8 and 18 CV with several smaller peaks eluting in between. Fraction 1-2-1, 1-2-2 and 1-2-4 where hydrolytic activity was verified were pooled as shown on the SDS-PAGE analysis in fig 5.8 d. SDS-PAGE of these fractions revealed pooled fractions 2-1-2 contained the major band of interest around 55kDa as well as a band around 26 KDa which could possible infer that wheat PEP may contain more than one subunit. There were also other bands that appear to have coeluted with PEP. The purification table (table 1) also shows that the purification of PEP decreased possibly due to degradation of PEP as well as the yield was quite low.

Table 5.1: Purification of PEP from germinating wheat seeds. Enzyme activity was determined with 6 mM Z-Gly-Pro-pNa. Protein was determined spectrophotometrically at 280 nm

fraction	volume (ml)	[protein] mg/ml	total protein [mg]	activity [Change in A_{405}/min/ 0.145ml]	total activity [Change in A_{405}/min]	specific activity	purification factor	Yield [%]
Crude	120	1	120	3.533	423.96	3.533	1	100
Cation on CM FF16/10	18	0.567	10.21	0.157	2.826	0.276	0.078	0.6
Gel exclusion on Sephacryl S-100	33	0.184	6.1	0.395	13.035	2.14	0.61	3.08

The final set of pilot runs was performed with various columns at different pH but with the same buffer system (20 mM Tris-HCl + 1M NaCl). These were as follows: DEAE at pH 5 and 6 and CM at pH 8 and 9. No activity profile was done for these runs. These pilot runs were performed so as to see on which column abundant protein binds sufficiently at the various pH values specified. It was apparent from the A_{280} profiles (not shown) that sufficient protein binding was observed on the DEAE columns at both pH 5 and pH 6. It was then decided to use a DEAE at pH 5 as a final step.

100 ml of clarified extract was loaded onto a HiPrep DEAE FF16/10 column equilibrated with 20 mM Tris-HCl pH 5. The column was washed with 5 CV of equilibration buffer A and fractions eluted with a linear salt gradient (0-1M NaCl in buffer A) over 20 CV.

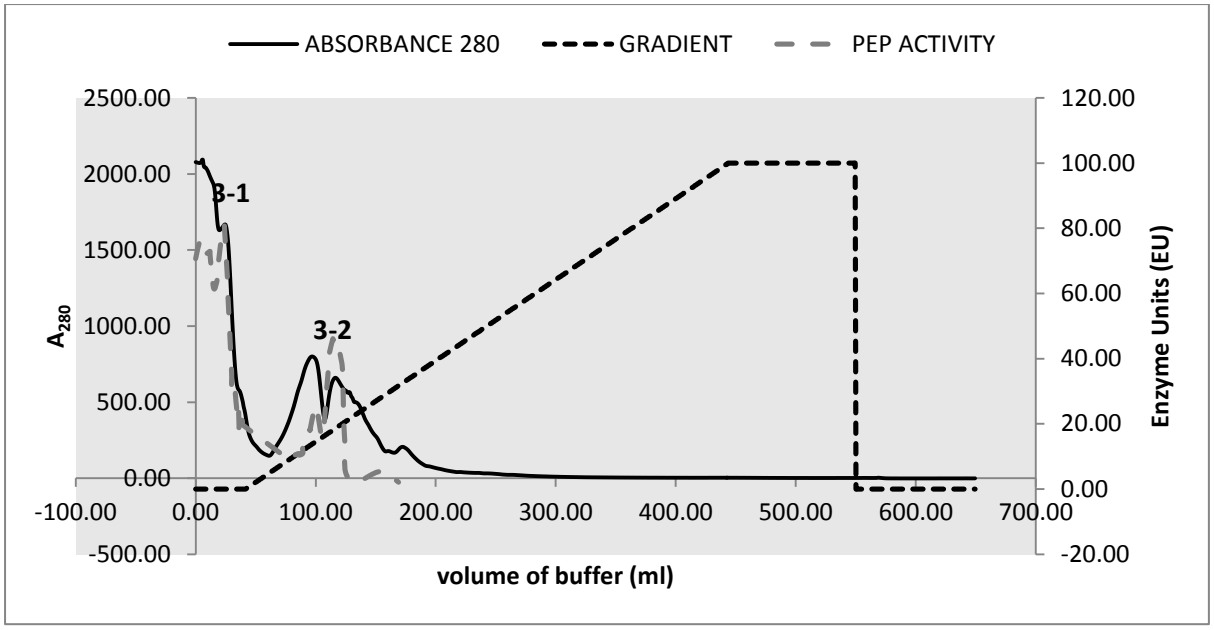
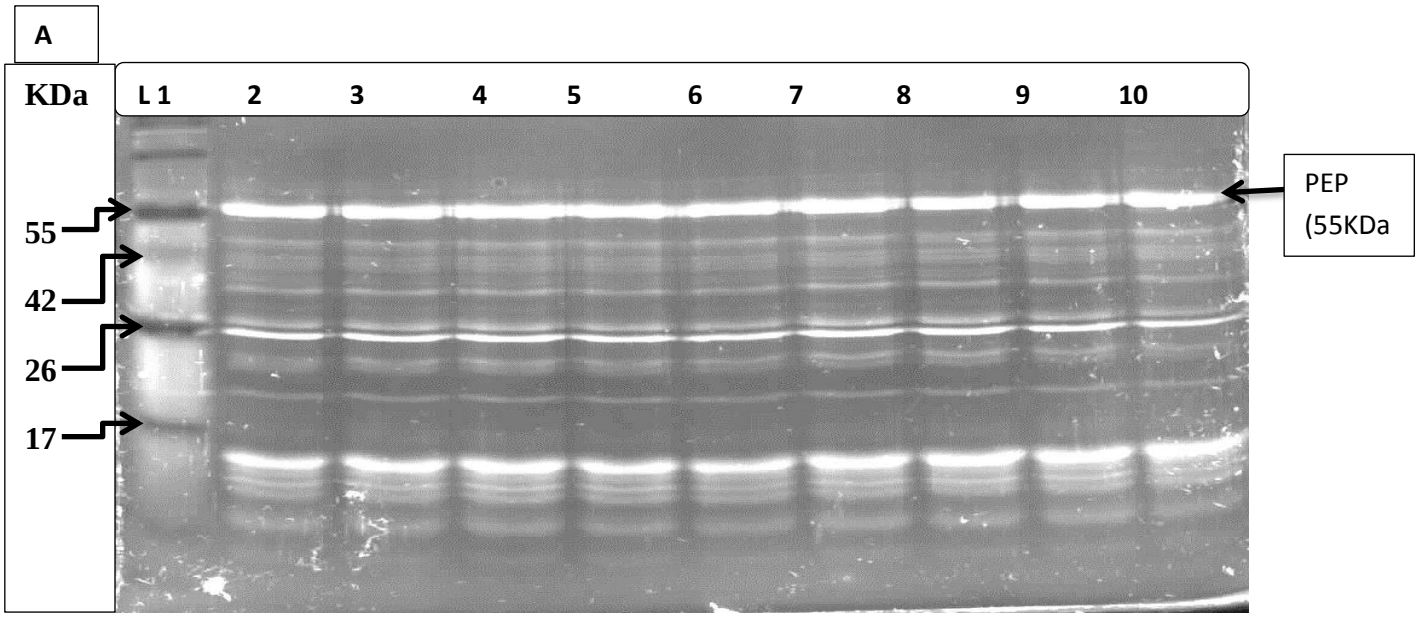
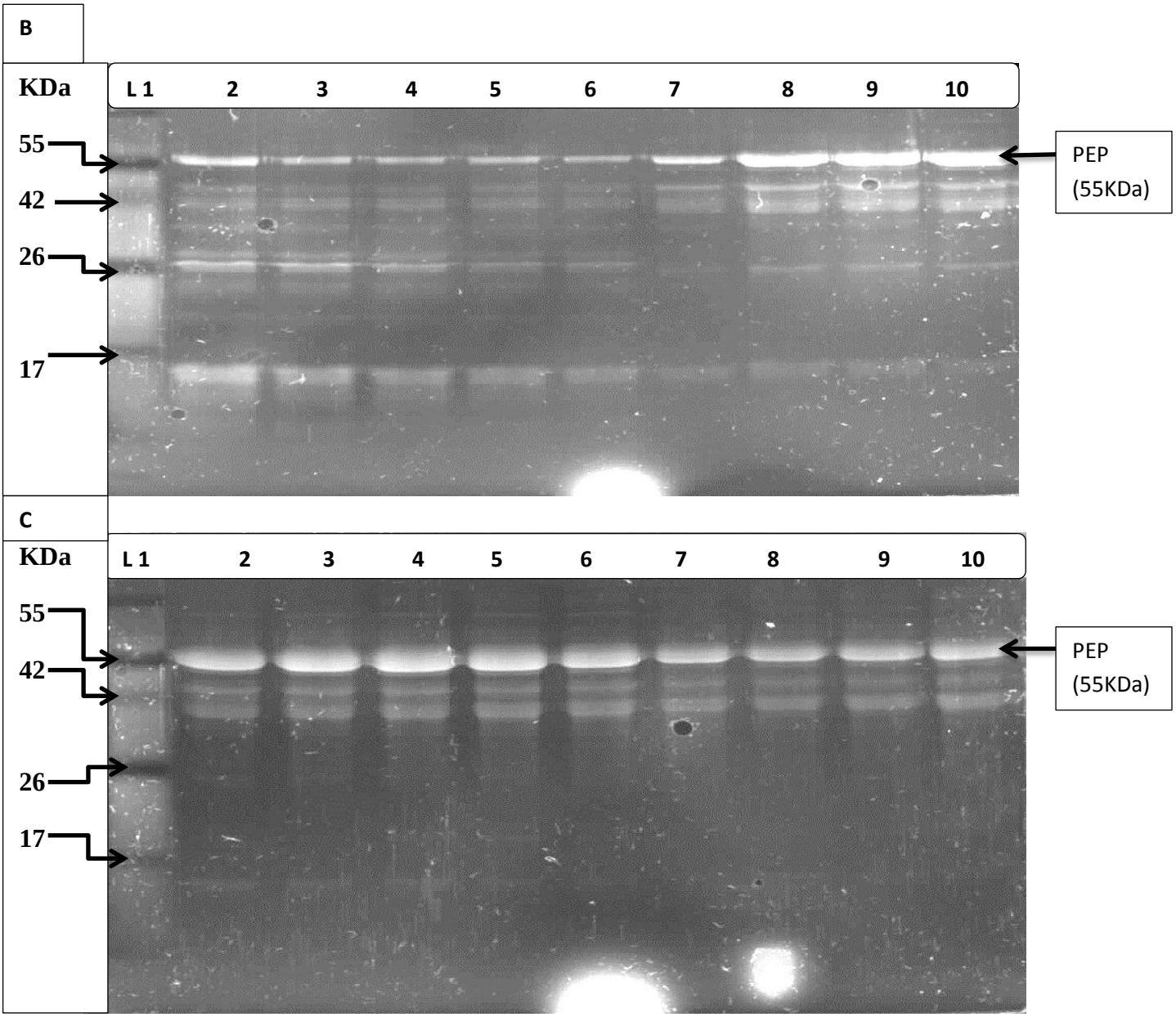


Figure 5.9: Purification of PEP from wheat seeds on a HiPrep DEAE FF16/10 column equilibrated with buffer A (20 mM Tris-HCl pH 5) and developed with linear salt gradient (0-1M NaCl) over 20 CV. Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Dashed black line: salt gradient. Numbers (3-1 and 3-2) reflect pooled fractions.




B
KDa
L 1
2
3
4
5
6
7
8
9
10
55
42
26
17
PEP
(55KDa)
C
KDa
L 1
2
3
4
5
6
7
8
9
10
55
42
26
17
PEP
(55KDa)

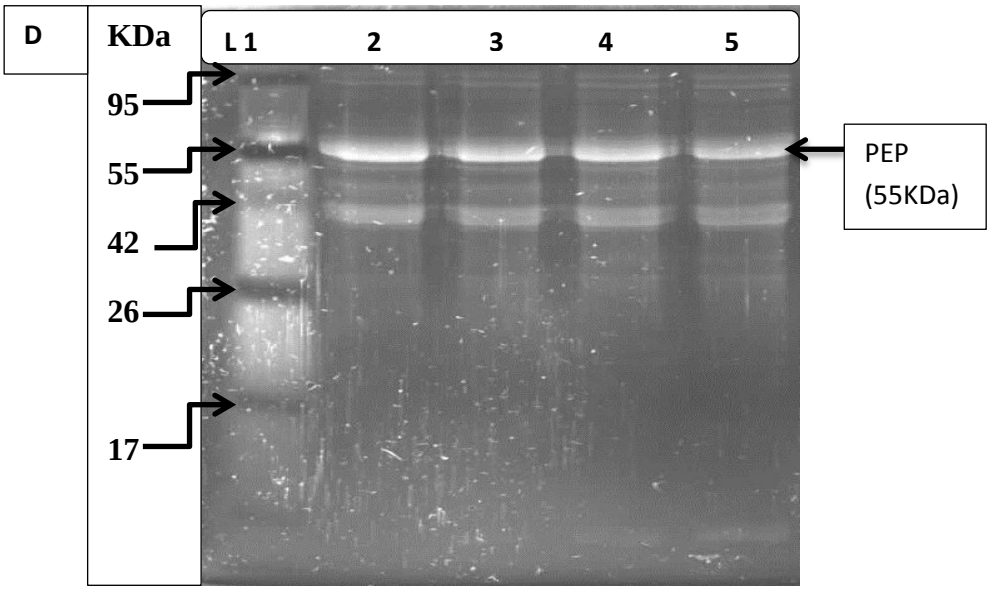

D
KDa
L 1
2
3
4
5
95
55
42
26
17
PEP
(55KDa)

Figure 5.10 a, b, c and d: SDS-PAGE profile of proteins peaks eluted after chromatography on a HiPrep DEAE FF16/10 column. Lane 1 represents the molecular weight marker, lanes 2 to 10 (fig 5.10 a) and 2-7 (figure 5.10 b) represents fractions in the region designated as 3-1; lanes 8-10 (fig 5.10 b), 2-10 (fig 5.10 c) and 2-5 (fig 5.10 d) represents fractions in the region designated as 3-2 as indicated on the A_{280} profile in figure 5.9. Every third fraction was screened for PEP activity in triplicate.

Three major peaks were obtained from the DEAE chromatography but these were split into two pools containing PEP activity with one of these pools eluting before applying the salt gradient. This may indicate that different PEP proteins in the extract or overloading of the column. Overloading of the column is unlikely though since Hi Prep DEAE FF16/10 has a binding capacity of 2.2 g total protein as the Hi Prep DEAE FF16/10 and only 375 mg total protein was loaded, which is beneath the binding capacity of the column. From the SDS-PAGE (fig 5.10 a-d), several bands were once again observed indicating that quite a lot of other possible proteins were present and still intact. The major band around 55kDa for PEP was also still present. There was also a prominent band around 26 KDa which is exactly half of the expected mass of PEP (55 KDa). This may possible represent an additional subunit but this would have to be further investigated. Several bands of lower molecular weight present and to remove these lower molecular weight proteins, both fractions (3-1 and 3-2) containing hydrolytic activity with the substrate Z-Gly-Pro-pNa were concentrated through ammonium sulphate precipitation and dialysed against excess buffer (20mM Tris-HCl pH 5) before being subjected to gel filtration on a HiPrep 16/60 Sephacryl S-100 High Resolution column. Fractions were pooled where activity was observed after gel filtration and aliquots of the purified enzyme were stored at -70^0C.

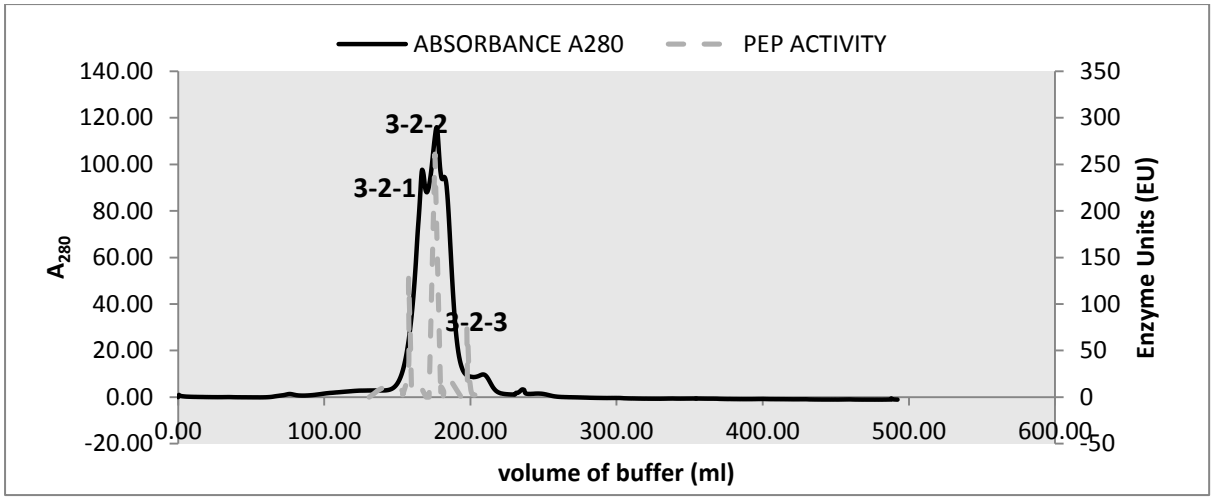


Figure 5.11: Elution profile of pooled fraction (3-1-2) from the DEAE column subjected to size exclusion chromatography on a HiPrep 16/60 Sephacryl S-100 equilibrated 20mM Tris-HCl pH 5. Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Fractions were pooled as indicated by numbers 3-2-1 to 3-2-3.

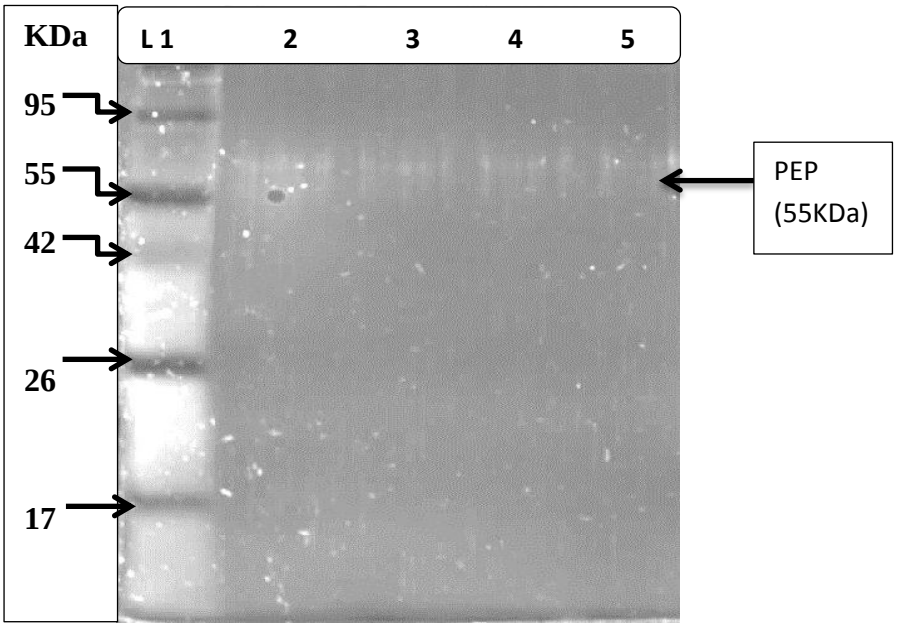


Figure 5.12: SDS-PAGE profile of proteins after chromatography on HiPrep 16/60 Sephacryl S-100 High Resolution column (FPLC). Lane 1 represents the molecular weight marker, lanes 2 to 5 represents fractions in the region designated as 3-2-2 as indicated on the A_{280} profile in figure 5.11. Every third fraction was screened for PEP activity in triplicate. This was the only SDS-PAGE gel that appeared to show remnants of protein bands from the three

regions as designated on the A_{280} profile in figure 5.11. No visible bands were seen on the SDS-PAGE gels (not shown) for fractions in the regions designated as 3-2-1 and 3-2-3 in figure 5.11.

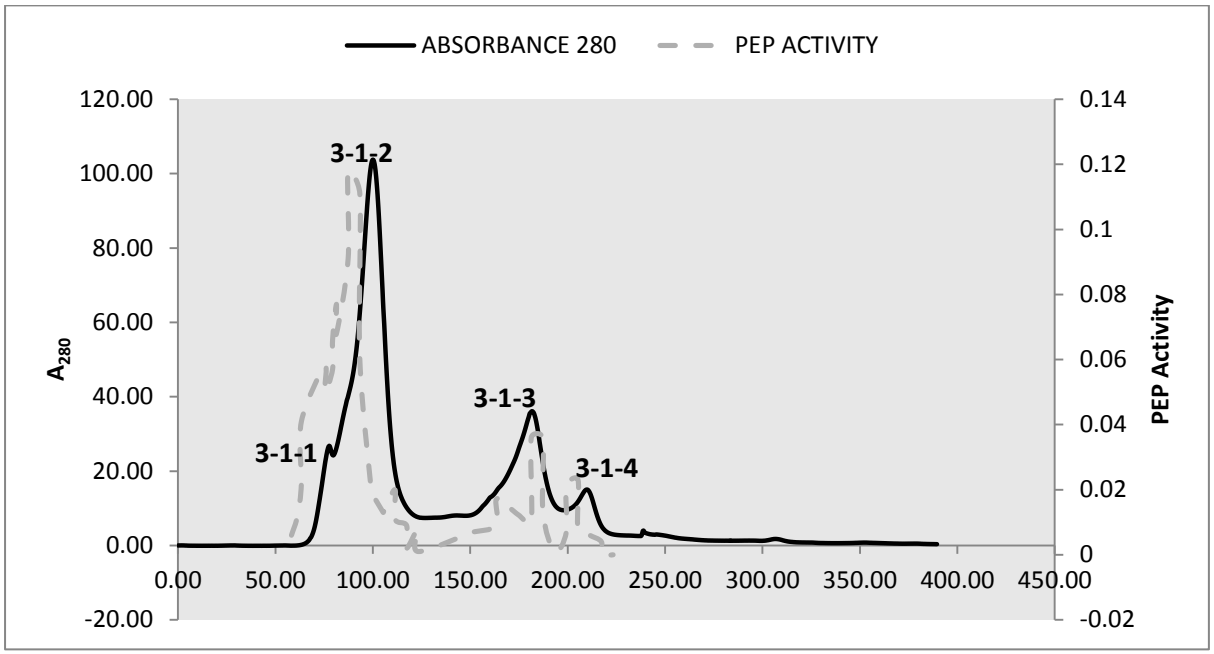


Figure 5.13: Elution profile of pooled fraction (3-1) from the DEAE column subjected to size exclusion chromatography on a HiPrep 16/60 Sephacryl S-100 equilibrated 20mM Tris-HCl pH 5. Grey dashed line: profile of proteolytic activity assayed at 405 nm with 0.1 mM final concentration Z-Gly-Pro-pNA; solid black line: profile of protein determined spectrophotometrically at 280 nm. Fractions were pooled as indicated by numbers 3-1-1 to 3-1-4.

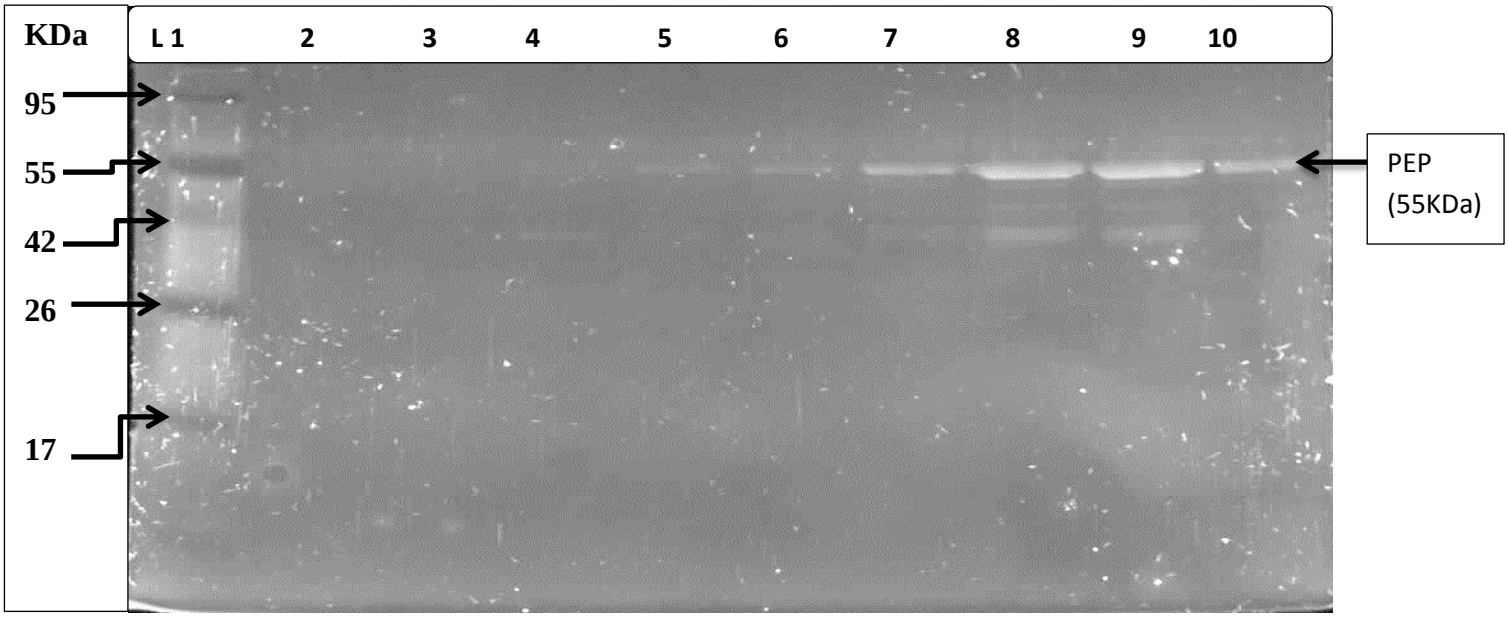


Figure 5.14: SDS-PAGE profile of proteins after chromatography on HiPrep 16/60 Sephacryl S-100 High Resolution column. Lane 1 represents the molecular weight marker, lanes 2 to 10 represents fractions in the region designated as 3-1-2 as indicated on the A_{280} profile in figure 5.13. Every third fraction was screened for PEP activity in triplicate. No bands were observed on the SDS-PAGE gels (not shown) in the regions designated 3-1-1, 3-1-3 and 3-1-4.

With reference to figures 5.11 and 5.13, three and four peaks eluted respectively. For fig 5.11, the peaks eluted around 8-10 CV with the peak showing abundant hydrolytic activity eluting around 8CV. For fig 5.13, the peaks eluted around 3-5 CV and 8-10 CV with the peak displaying major hydrolytic activity eluting around 4CV. SDS-PAGE profile reveals that only remnants of PEP was detected for the peak that displayed major hydrolytic activity in fig 5.11 designated as 3-2-2 and a near homogenous band for the peak that displayed major hydrolytic activity designated as 3-1-2 as indicated on the A_{280} profile in figure 5.13. This could possibly indicate several species of PEP or additional subunits. The purification results in table 2 also show that the purification factor of the enzyme increased almost four times for the portion that was designated as 3-1-2 but the yield was extremely low.

Table 5.2: Purification of PEP from germinating wheat seeds. Enzyme activity was determined with 6 mM Z-Gly-Pro-pNa. Protein was determined spectrophotometrically at 280 nm

fraction	volume (ml)	[protein] mg/ml	total protein [mg]	activity [Change in A_{405}/min/ 0.145ml]	total activity [Change in A_{405}/min]	specific activity	purification factor	Yield [%]
Crude	150	2.5	375	1.754	263.1	0.7016	1	100
Anion on DEAE								
3-1	33	0.736	24.3	0.627	20.691	0.851	1.21	6.48
3-2	40	0.532	21.3	0.722	28.88	1.356	1.93	5.68
Dialysis								
3-1	5	0.743	3.715	0.457	2.285	0.615	0.876	0.99
3-2	5	0.648	3.24	0.097	0.485	0.149	0.212	0.864
Gel exclusion								
3-2-1	10	0.019	0.19	0.15	0.02	0.105	0.149	0.05
3-2-2	5	0.0193	0.097	0.25	0.025	0.258	0.368	0.026
3-2-3	8	0.0186	0.149	0.05	0.072	0.483	0.688	0.0397
3-1-1	8	0.022	0.176	0.019	0.152	0.864	1.23	0.0469
3-1-2	18	0.0342	0.62	0.122	1.702	2.74	3.9	0.165
3-1-3	22	0.0174	0.383	0.041	1.342	3.48	4.96	0.102
3-1-4	11	0.0192	0.211	0.123	0.11	0.52	0.74	0.0562

Fractional precipitation of PEP crude extract with ammonium sulphate increased the specific activity from 0.7µmol to 3µmol of Z-Gly-Pro-pNa (table 2) as compared to the decline that is observed in table 1 where ammonium sulphate was not used as the precipitating agent. Ammonium sulphate draws water molecules away from the non-polar units of the proteins. This decrease in available water molecules increases the surface tension and enhances hydrophobic interactions, thus allowing a protein to precipitate from a solution. The fold purification also increased from 1 to approximately 4.

Ammonium sulphate maintains the protein in a folded state and also its low heat of solubilisation avoids the risk of protein denaturation that can occur when the sample temperature increases. The dialysis step removed salts and ions (that may limit the activity of the enzyme) through osmosis from a high concentrated solution (enzyme extract) to the dilute solution (buffer).

From both purification tables, one can see a noticeable drop in the enzyme recovered as a result of the size exclusion step. This drop may also possibly be due to autolysis or other contaminants that may have caused denaturation on the protein of interest.

5.3 Concluding remarks

Separation of PEP from germinating wheat seeds was achieved by gel filtration on a Sephacryl S100 column and appeared to be more successful using a slightly more acidic Tris buffer without salt. The elution profiles shows that PEP co-eluted with other proteins and was not homogenous according to SDS-PAGE analysis after the two-step purification. Our findings also estimated the molecular mass of PEP from wheat to be ~55kDa, according to SDS-PAGE analysis.

5.4 References

1) (1999). Invitrogen Quant-iT ™ Protein Assay Kit.J Biomol Screen **4**: 67-73.
2) Loponen J, Sontag-Strohm T, VenaLainen J and Salovaara H. (2007). Prolamin hydrolysis in wheat sourdoughs with differing proteolytic activities. J. Agric. Food Chem. **55**: 978-984.
3) Smith PK et al. (1985). BCA Protein Assay. Anal. Biochem. **150**: 76-85.

CHAPTER 6

KINETIC CHARACTERIZATION OF PROLYL ENDOPEPTIDASE

6.1 Introduction

A variety of approaches have been made to solve the mystery between gluten proteins and celiac sprue. Brandt et al.,(2007) mentions that although people eat bread and other grain products, they lack digestive prolyl endopeptidase. Sollid and Khosla (2005) have listed various approaches as to how celiac sprue may be treated. Of these, the use of proteolytic enzymes that cleave at a proline or glutamine and therefore can degrade harmful gluten peptides is one special area of interest. A Prolyl endopeptidase from wheat was therefore investigated in this study. There are two approaches for the use of these enzymes in treating celiac disease: 1) to hydrolyse them after the gluten has been ingested, in the gastrointestinal (GI) tract (the medical approach) or 2) to hydrolyse them prior to the gluten having been ingested, during food processing (the food technological approach). The detoxification of gluten by proteolysis is not a novel idea and neither is the use of more than one protease in an effective detoxification procedure. PEPs have thus gained special interest as a potential for celiac disease treatment individually or jointly with possibly other post-proline cleaving enzymes.

As reported in chapter one, although PEPs have been found in all major groups of living organisms, however their function still remains obscure. Many of the reported investigations have looked at PEP activity, inhibition studies as well as other kinetic characteristics of PEPs but in this study we present the first data and provide a more complete understanding of PEPs in germinating wheat seeds. To our knowledge, only one study has been conducted using germinated wheat as a raw material in sourdough fermentations. This study was aimed at evaluating the extent of prolamin hydrolysis during germinated wheat sourdough fermentation. One area of interest in their study looked at the activities of two proline-specific enzymes, prolyl oligopeptidase (POP) and dipeptidyl-peptidase IV (DPP IV) using specific 7-amino-4-methylcoumarin (AMC) substrates: succinyl (SUC)-Gly-Pro-AMC for POP and H-Gly-Pro-AMC for DPP IV (Loponen et al., 2007).

In this study we examined various kinetic properties of a purified wheat PEP including substrate specificity using specific p-nitroanilide substrates (Z-Gly-Pro-pNa, N-Suc-Ala-Ala-Pro-Phe-pNa and Glp-FA-pNA) as well as other kinetic characteristics such as pH optima, temperature optima, stability.

6.2 Method

6.2.1Characterization of PEP

6.2.1.1 pH optimum

To establish the optimum pH for the purified PEP, activity was determined at different pH values. PEP activity was measured in universal 100 mM acetate–phosphate–borate buffer (UB) at pH values ranging from 2 – 11 using the assay procedure described in section 4.3. Enzyme activity assays were done in triplicate at each pH.

6.2.1.2 Temperature optimum

The temperature profile for the purified PEP was determined over a temperature range of 4 - 100^0C at the optimum pH. The activity was determined as described in section 4.3. Enzyme activity assays were done in triplicate at each temperature.

6.2.1.3 Thermal stability

To determine the optimal thermal stability of PEP, purified PEP was incubated at 4, 10, 25, 30, 37, 40, 50, 60, 70 and 80 °C for 10min at the optimum pH. Samples were prepared in triplicate and residual enzyme activity was measured as described in section 4.3.

6.2.2 Kinetic studies

Kinetic studies were performed at 37 °C with substrate concentrations ranging from 0.17µM to 0.1mM final concentration in universal buffer (100 mM acetate–phosphate–borate buffer (UB) of pH 7.9). The kinetic parameters K_m and V_{max} were determined from a Lineweaver–Burk plot.

6.3 Results and Discussion

6.3.1 Characterization studies

6.3.1.1 pH-Optimum

The activity of most enzymes strongly depends on the pH of the medium therefore it was important to optimize the pH of PEP. Since PEP has been found in diverse species and in diverse environmental conditions the pH profile was explored. The optimum pH for activity with 0.1 mM final concentration Z-Gly-Pro-pNa was assayed using 100 mM universal buffer,

with pH values ranging from 2.0 to 11.0. 150 µl of 100 mM universal buffer of various pH values was added to 145 µl of purified PEP values and preincubated for 10 min at room temperature before adding 5 µl of 6 mM Z-Gly-Pro-pNa. A pH of 7.9 was determined to be the optimum pH for wheat PEP (figure 6.1).

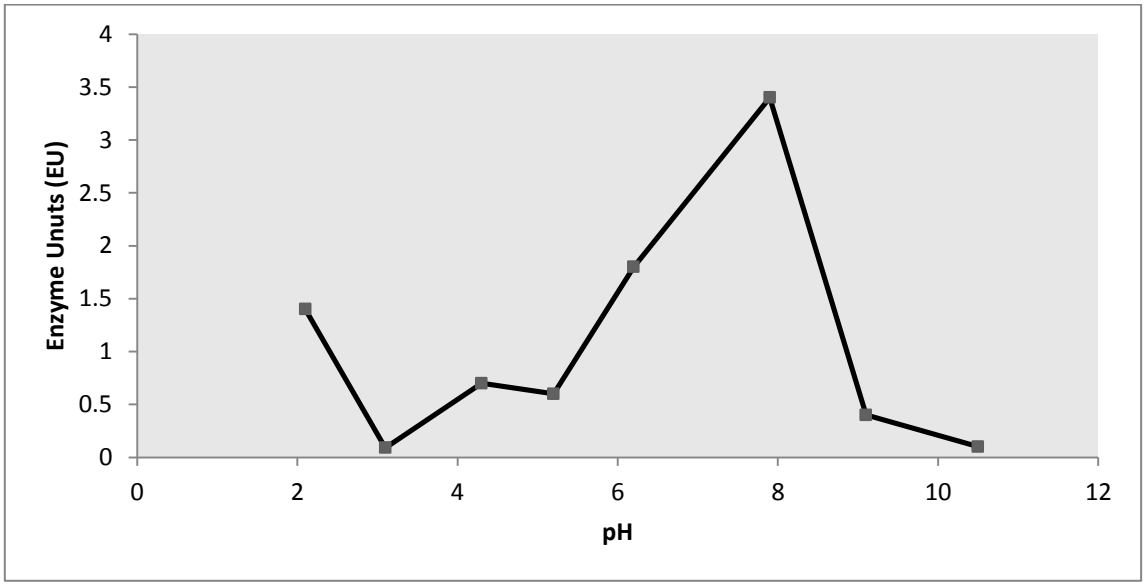


Figure 6.1: Effect of pH on the activity of PEP using Z-Gly-Pro-pNa as substrate. All assays were performed in triplicate and the mean activity plot against pH.

With reference to fig 6.1, the optimum pH for wheat appears to be in accordance with the findings of others but more closely related to the archae and insect group.

6.3.1.2 Temperature profile

To determine the optimum temperature for wheat PEP activity with 0.1 mM final concentration Z-Gly-Pro-pNa as substrate, assays were conducted at 4, 10, 25, 30, 37, 40, 50, 60, 70 and 80 °C in triplicate. Our findings indicate that the optimum temperature for PEP was 37^0C (figure 6.2). This finding is similar to previous observations of temperature optima of other PEPs (table 6.1). The sudden decrease in activity at temperature greater than 37^0C was thought to be due to enzyme instability or denaturation. A temperature stability study was therefore conducted.

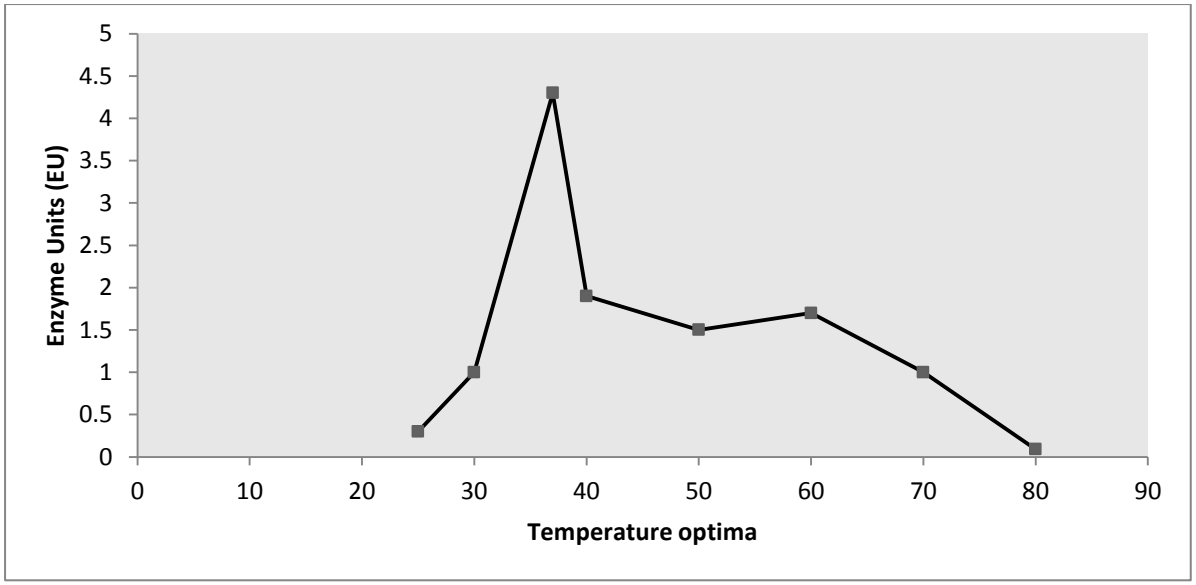


Figure 6.2: Effect of temperature on the activity of PEP with 0.1 mM final concentration Z-Gly-Pro-pNa as substrate and universal buffer at 7.9. Activity points represent the mean of triplicate assays

For temperature stability testing, the purified enzyme was incubated at 4, 10, 25, 30, 37, 40, 50, 60, 70, 80, 90 and 100 °C for 10 min. The solution of Z-Gly-Pro-pNa was added to the final concentration of 6 mM and the residual activity was measured as described earlier.

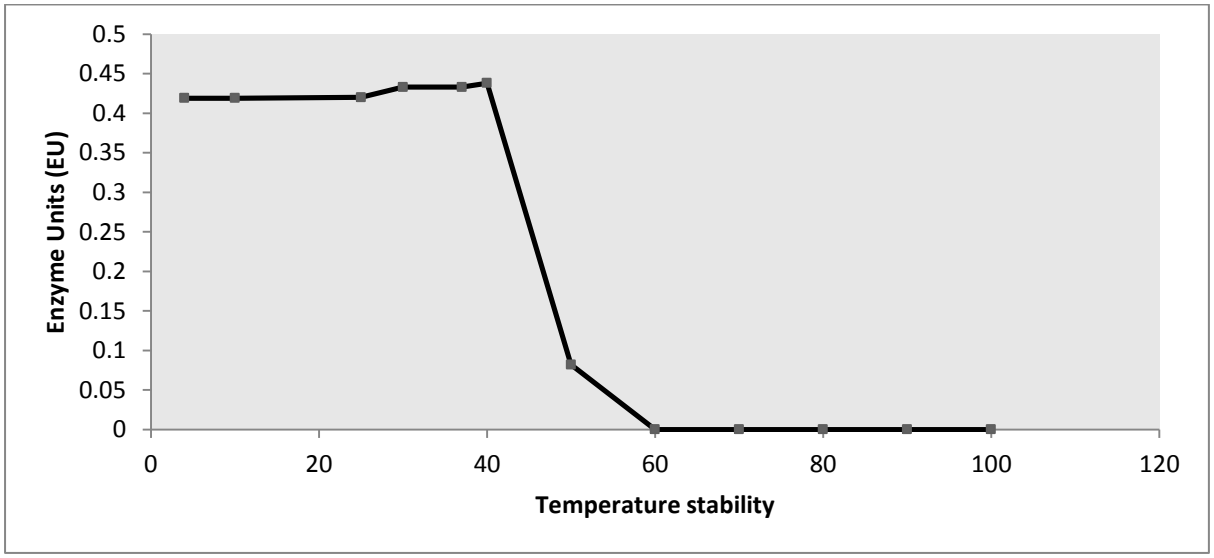


Figure 6.3: Temperature stability of PEP activity. Residual activity was measured using 0.1 mM final concentration Z-Gly-Pro-pNa as substrate in universal buffer. Activity points represent the mean of triplicate assays.

The thermal stability studies results show that wheat PEP is stable up to 40°C and a drastic decline occurs thereafter in the stability of the enzyme which could possible indicate denaturation. With reference to table 6.1, the stability of PEP from other sources appears to be in line with the findings of others with the exception of *pseudomonas* species. This is indicative that the enzyme cannot thrive at high to extreme temperatures.

Table 6.1: comparison of some physiochemical properties of several confirmed prolyl endopeptidase

Origin	*Pseudomonas sp. KU-22*	*F. meningosepticum*	*Xanthomonas sp*	*A. hydrophila*	Human brain	*Triticum aestivum*
Optimum pH	8.0	7.0	7.5	8	6.8	7.9
pH Stability	8.0-11.0	5.0-9.0	6-8.5	7-8.5	5.5-9.5	5.0-9.0
Optimum temperature	45°C	40°C	Not determined	30°C	37°C	37°C
Thermal stability	45°C	42°C	Not determined	35°C	37°C	40°C
Isoelectric point	4.9	9.6	7.2	5.5	4.75	Not determined
Molecular weight	76,000 Da	76,000 Da	74,000 Da	76,000 Da	79,600 Da	55,000 Da
Reference	Oyama et al., 1997	Yoshimoto et al., 1980	Szwajcer-dey et al., 1992	Knnatani et al., 1993	Kalwant and Porter., 1991	Present data

6.3.2 Estimation of kinetic parameters

The kinetic parameters K_m and V_{max} were determined for various p-nitroanilide substrates (Z-Gly-Pro-pNa, N-Suc-Ala-Ala-Pro-Phe-pNa and Glp-FA-pNA) using the Lineweaver-Burk plot (figures 6.4-6.6). The lowest K_m value of 0.58 mM was observed for N-Suc-Ala-Ala-Pro-Phe-pNa (table 6.2). The highest K_{cat} of 29.37 s^{-1} was for N-Suc-Ala-Ala-Pro-Phe-pNa .The K_{cat}/K_m ratio was the highest also for N-Suc-Ala-Ala-Pro-Phe-pNa (50813.14s^{-1} M^{-1}), and was the lowest for Z-Gly-Pro-pNa (858.77 s^{-1} M^{-1}) (table 6.2). According to the findings of Elpidina and Goptar on Post-proline cleaving peptidases in *Tenebrio Molitor*, this group of enzymes more effectively hydrolyze peptide bonds after Pro residues in peptide substrates with Ala in the P2 position more effectively (Personal communication).

Their findings also suggest that activity with substrates with Gly residues in the P2 position was very low and this could also be a determining factor as seen in our findings with the substrate Z-Gly-Pro-pNa.

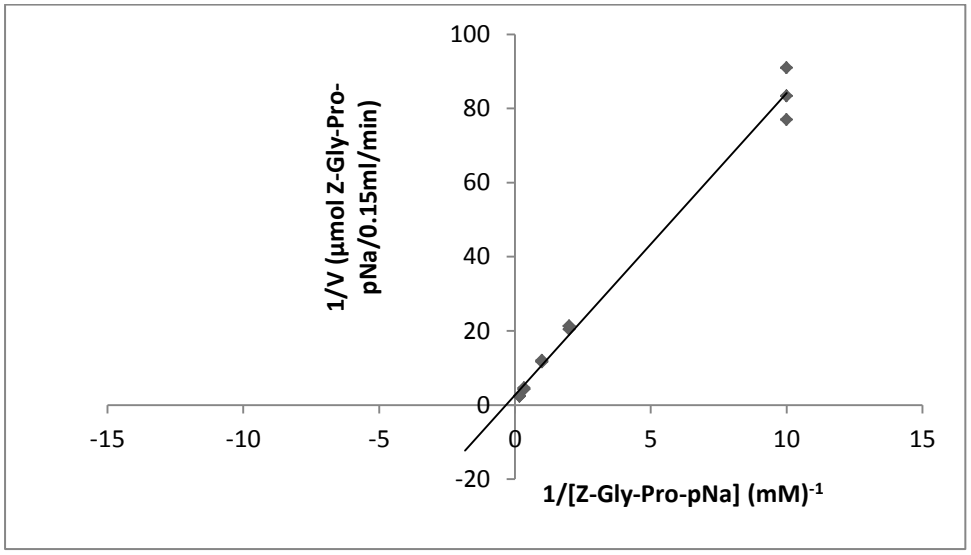


Figure 6.4: Lineweaver-Burk plot of PEP at various substrate concentrations. Assays were performed in triplicate at 37°C in universal buffer at pH 7.9.

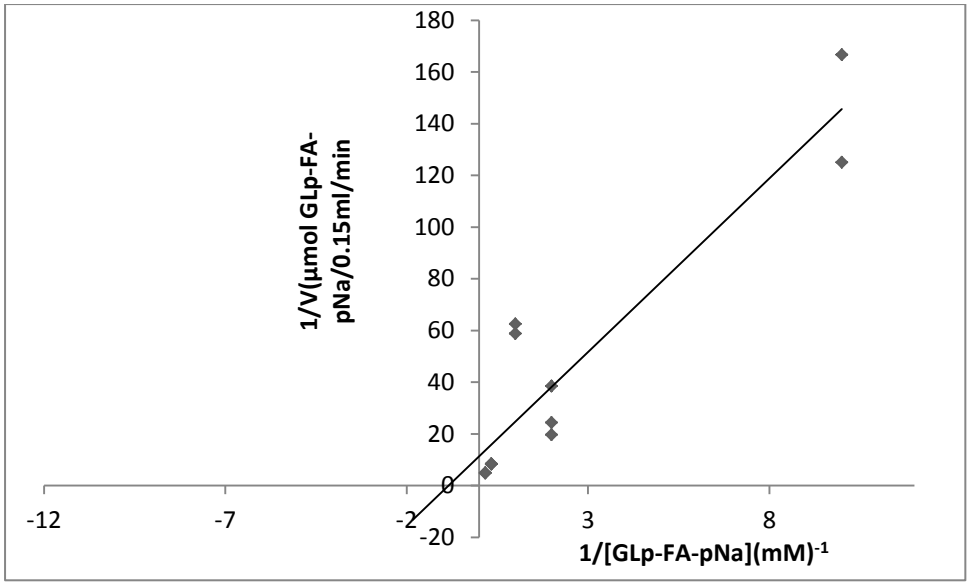


Figure 6.5: Lineweaver-Burk plot of PEP at various substrate concentrations. Assays were performed in triplicate at 37°C in universal buffer at pH 7.9.

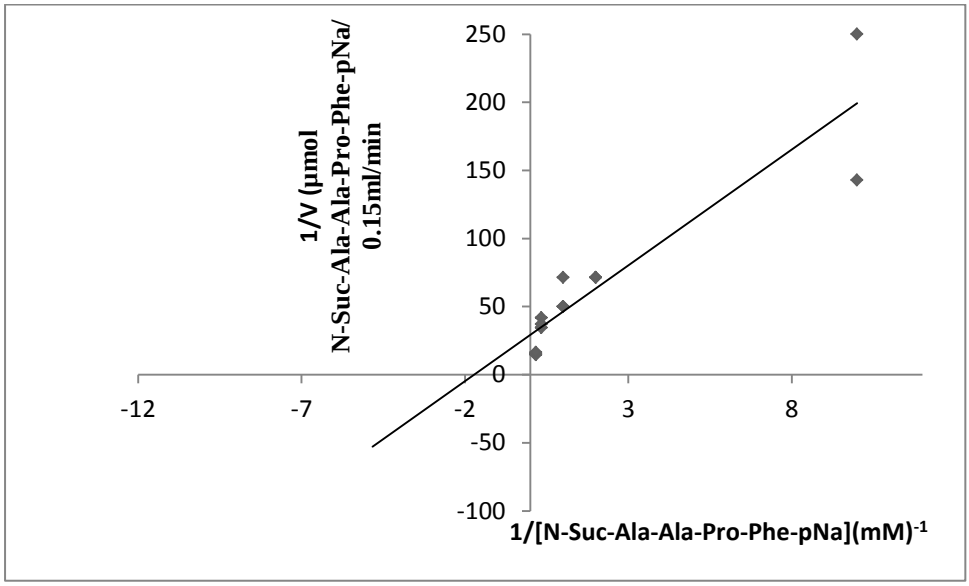


Figure 6.6: Lineweaver-Burk plot of PEP at various substrate concentrations. Assays were performed in triplicate at 37°C in universal buffer at pH 7.9.

The Michaelis- Menten plots clearly show that at low substrate concentrations, the rate of reaction increases more or less linearly with substrate concentration while at higher substrate concentrations the reaction becomes constant due to the active site being saturated with substrate. The Michaelis constant, K_m, can be obtained from the Michaelis-Mentin plot since it is the substrate concentration at which the reaction is half of its maximum value ($V_{max}/2$),

which is the reason why K_m is also called the half-saturation value. However, it is difficult to determine V_{max} for this hyperbolic plot and hence in determining K_m.

For this reason, Lineweaver-Burk plots were used because they allow more accurate measurements of the values of both K_m and V_{max} even at higher substrate concentrations.

Table 6.2: Kinetic parameters for the hydrolysis of p-nitroanilide substrates by PEP at 37°C and pH 8.

Substrate	K_m(M 10^{-3})	K_{cat} (s^{-1})	K_{cat}/K_m ($s^{-1}M^{-1}$)
Z-Gly-Pro-pNa	3.08×10^{-3}	2.645	858.77
N-Suc-Ala-Ala-Pro-Phe-pNa	5.78×10^{-4}	29.37	50813.14
Glp-FA-pNa	1.18×10^{-3}	11.35	9618.6

With regards to the hydrolysis of Z-Gly-Pro-pNa, it is evident that the microorganism family appears to have a stronger affinity and thus hydrolyses the proline specific substrate more effectively than our findings in a plant based source (table 3). Another determining factor could have been that our protein of interest was not purified to homogenous. Again, this however does not mean that the PEPs isolated from microorganisms appear to be a more effective enzyme than the plant based source one. Each one of these enzymes from its various sources is a unique enzyme with specific characteristics.

Table 6.3: Comparative analysis of K_m values for the hydrolysis of p-nitroanilide substrates Gly-Pro-pNa, N-Suc-Ala-Ala-Pro-Phe-pNa and Glp-FA-pNa by PEP.

Substrate	Organism	Km	Reference
Z-Gly-Pro-pNa	*Pyroccocus furiosis*	0.16	Wallen et al., 2003
	Elizabethkingia meningoseptica	0.25	Deutch et al., 2010
	Aeromonas caviae	0.81	Li et al., 2010
	Triticum aestivum	3.08	Present data
N-Suc-Ala-Ala-Pro-Phe-pNa	*Triticum aestivum*	0.578	Present data
Glp-FA-pNa	*Triticum aestivum*	1.18	Present data

6.4 Concluding remarks

The substrate specificity results indicate that the purified plant PEP is a chymotrypsin-like serine proteinase. The enzyme hydrolyzed a longer peptide substrate more efficiently (N-Suc-Ala-Ala-Pro-Phe-pNa), containing four amino acid residues with Phe in the P1 position, than the shorter proline specific substrate with proline occupying the P1 position (Z-Gly-Pro-pNa). PEP also appeared to have the highest affinity for N-Suc-Ala-Ala-Pro-Phe-pNa while the highest reaction rate (*k*cat) and efficiency (*k*cat/*K*m) were also observed for this substrate.

pH, temperature optima and temperature stability results appear to be in accordance with various other sources of known PEP from literature.

CHAPTER 7

INHIBITORY POTENTIAL OF CLASS SPECIFIC INHIBITORS ON PEP ACTIVITY

7.1 Introduction

7.1.1 Enzyme inhibition kinetics

Several substances can bring about a decline in the rate of an enzyme catalysed reaction. Some substances that may bring about this reduction can be non-specific protein denaturants whilst others may act in a specific manner as in the case of molecules/substances known as inhibitors. The loss of enzyme activity can be broadly categorized into reversible, where the activity may be restored by the removal of the inhibitor, or irreversible, in which the loss of the enzyme activity is time dependent and cannot be recovered (Copeland, 2000). Thus, enzyme inhibitors fall into two broad classes: reversible inhibitors (inhibitors causing reversible inactivation of enzymes) and irreversible inhibitors (inhibitors whose inhibitory effects cannot be reversed).

7.1.1.1 Reversible inhibitors

A typical property of all reversible inhibitors is that when the inhibitor concentration drops, the activity of the enzyme is regenerated. Reversible inhibitors bind to enzymes with non-covalent interactions such as hydrogen bonds, hydrophobic interactions and ionic bonds. Reversible inhibitors are further classified into four sub-classes: competitive inhibitors, non-competitive inhibitors, uncompetitive inhibitors and mixed inhibitors (Meleti, 1986).

7.1.1.2 Competitive inhibitors

The inhibitor binds to the same site on the enzyme as the substrate. The inhibitor only binds to the free enzyme and in most instances appears to be structurally very similar to the substrate. Once it has bound, it allows no space for the enzyme to react with its substrate. This type of inhibition can be overcome by sufficiently high concentrations of substrate (V_{max} remains constant) out-competing the inhibitor. However, the apparent K_m will increase as it takes a higher concentration of the substrate to reach the K_m point, or half the V_{max} (Meleti, 1986).

7.1.1.3 Non-competitive inhibitors

This type of inhibition can be seen as a form of mixed inhibition where the binding of the inhibitor to the enzyme reduces the enzymes activity but does not affect the binding of substrate. As a result, the extent of inhibition depends only on the concentration of the inhibitor. V_{max} will decrease due to the inability for the reaction to proceed as efficiently, but

Km will remain the same as the actual binding of the substrate will still function properly (Meleti, 1986).

7.1.1.4 Uncompetitive inhibitors

With Uncompetitive inhibition, the inhibitor binds only to the substrate-enzyme complex. This type of inhibition causes V_{max} and K_m to decrease (Meleti, 1986).

7.1.1.5 Mixed inhibitors

In mixed inhibition, the inhibitor can bind to the enzyme at the same time as the enzyme's substrate. However, the binding of the inhibitor affects the binding of the substrate, and vice versa. This type of inhibition can be reduced, but cannot be overcome by increasing concentrations of substrate (Meleti, 1986).

7.1.1.6 Irreversible inhibitors

Irreversible inhibitors usually covalently modify an enzyme and do not dissociate from the enzyme (Meleti, 1986).

7.2 Materials and methods

7.2.1 Materials

Lyophilized substrate (Z-Gly-Pro-pNa) was purchased from Sigma-Aldrich Co. (South Africa (Pty) Ltd). E64 (serine proteinases), pepstatin (aspartic proteinases) and leupeptin (cysteine and some serine proteinases) inhibitors were purchased from Sigma-Aldrich Co. (South Africa (Pty) Ltd). Monitoring the increase in absorbance at 405nm as p-nitroanilide is released was done using a micro plate reader (SynergyMx, Bio tek Instruments, USA).

7.2.2 Method

For inhibition studies, aliquots of the purified PEP enzyme preparation were preincubated with different concentrations of inhibitors for 1hr at 37°C in 100 mM UB, pH 7.9 and residual activity with appropriate p-nitroanilide substrate was assayed as described in section 4.3. Diagnostic inhibitors and their specificity included: E64 (serine proteinases) and pepstatin (aspartic proteinases) at 10.4 µM, 7.9 µM, 5.4 µM, 2.7 µM and 0 µM; leupeptin (cysteine and some serine proteinases), at104 µM, 79 µM, 54 µM, 27 µM and 0 µM final concentrations and PMSF (serine proteases) at 1.04 mM, 0.8mM, 0.5 mM, 0.27 mM and 0

mM. See appendix B for the various inhibitors stock concentrations that were prepared as well as the different solvents used in dissolving the inhibitors (pg 108).

7.2.2.1 Inhibition of PEP activity

Purified PEP (145 µl) was preincubated at 37°C for 1 hour prior to the addition of UB pH 7.9 and substrate. Residual activity was measured by adding 5 µl of 6mM substrate and 150 µl UB, pH 7.9 and the rate of increase in absorbance at 405 nm was measured with a microplate reader.

Comparison of reaction rates in the presence of an inhibitor with those for control reactions were used to calculate the degree of enzyme inhibition. All reactions were performed in triplicate

7.2.2.2 Determination of IC_{50}

The main objective of this part of the study was to explore the inhibitory potential of class specific inhibitors on the activity of PEP. The inhibitory potential of an inhibitor is described either as the concentration of an inhibitor at which the enzyme activity falls to 50% of its original value or as the molar concentration of the compound required to inhibit 50% of PEP activity (IC_{50}). In order to determine the IC_{50} values (50% inhibition), various class specific inhibitors were diluted to concentrations as specified in the appendix B.

7.2.2.3 Determination of the type of enzyme inhibition

Determination of the type of inhibition type is critical for the identification of mechanism of inhibition and the sites of inhibitor binding. Two methods were applied to determine the effect of inhibitors on both K_m and V_{max} values. This was the Dixon plots as well as the [S]/v against I plot.

7.3 Results and Discussion

7.3.1 Effect of inhibitors

The enzyme extract was examined in a series of class specific inhibition assays as shown in the method section previously. All assays were run in triplicate at pH 7.9 and 37ºC. The effects of varying the concentration of these class specific inhibitors on the activity of PEP were explored. (Figures 7.1a-d).

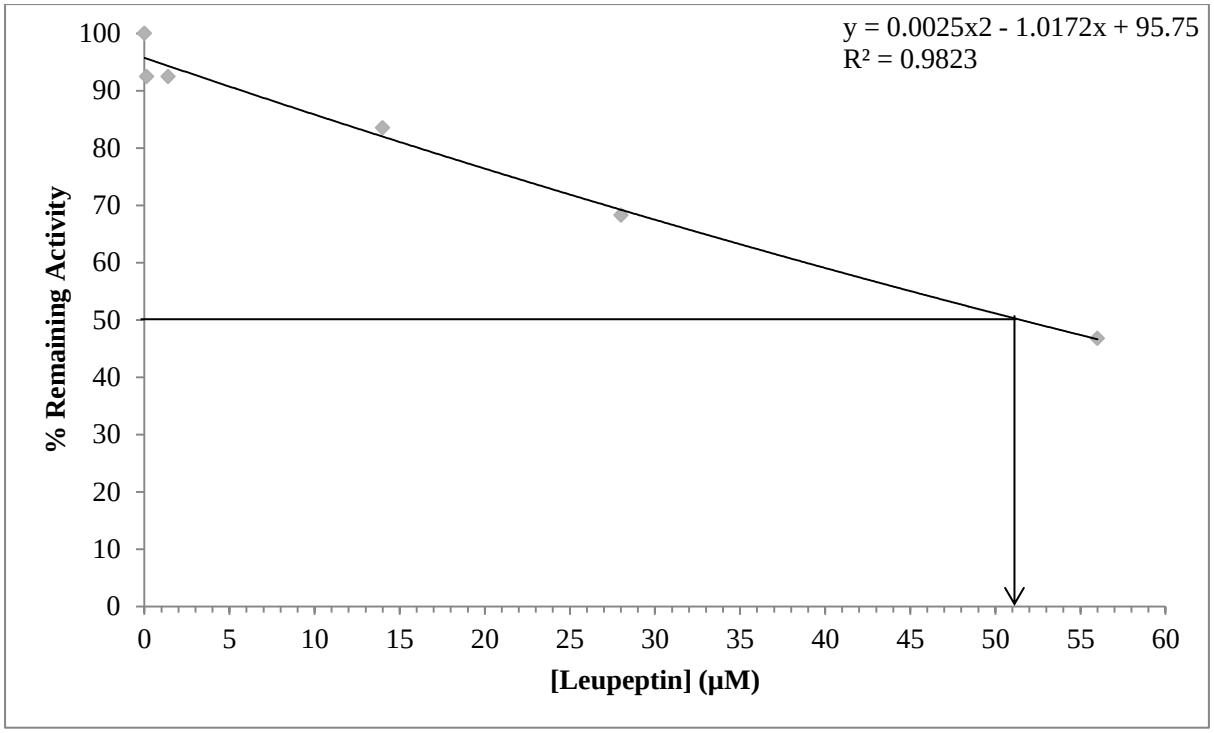


Figure 7.1a: Effect of leupeptin on PEP activity. PEP extracts in the absence of the inhibitor were used as positive controls. The data are the average of three experiments.

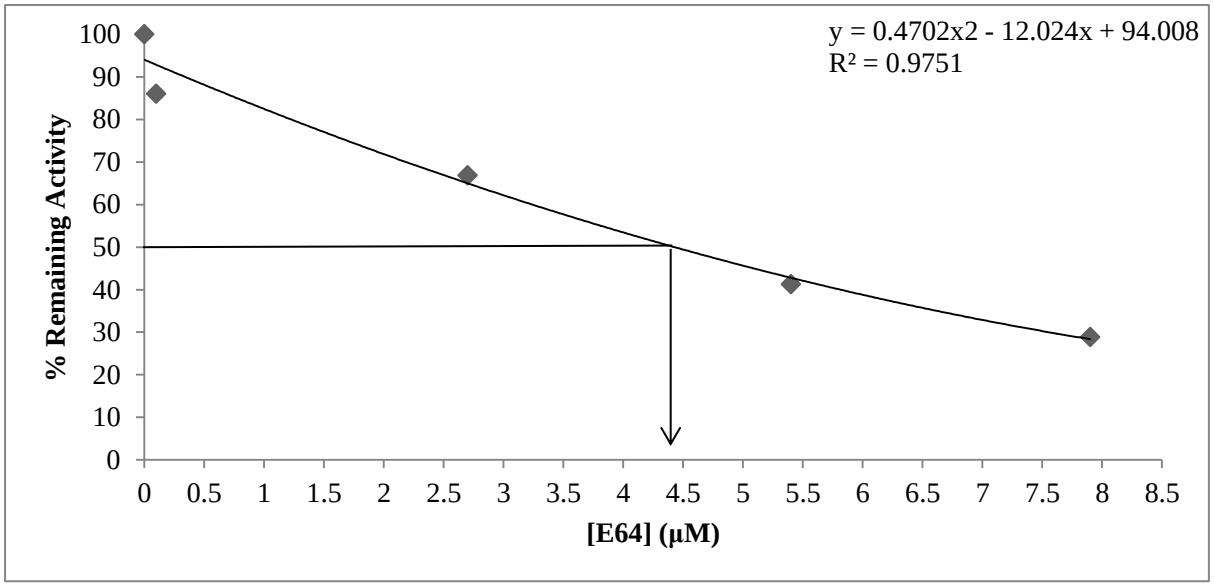


Figure 7.1b: Effect of E64 on PEP activity. PEP extracts in the absence of the inhibitor were used as positive controls. The data are the average of three experiments.

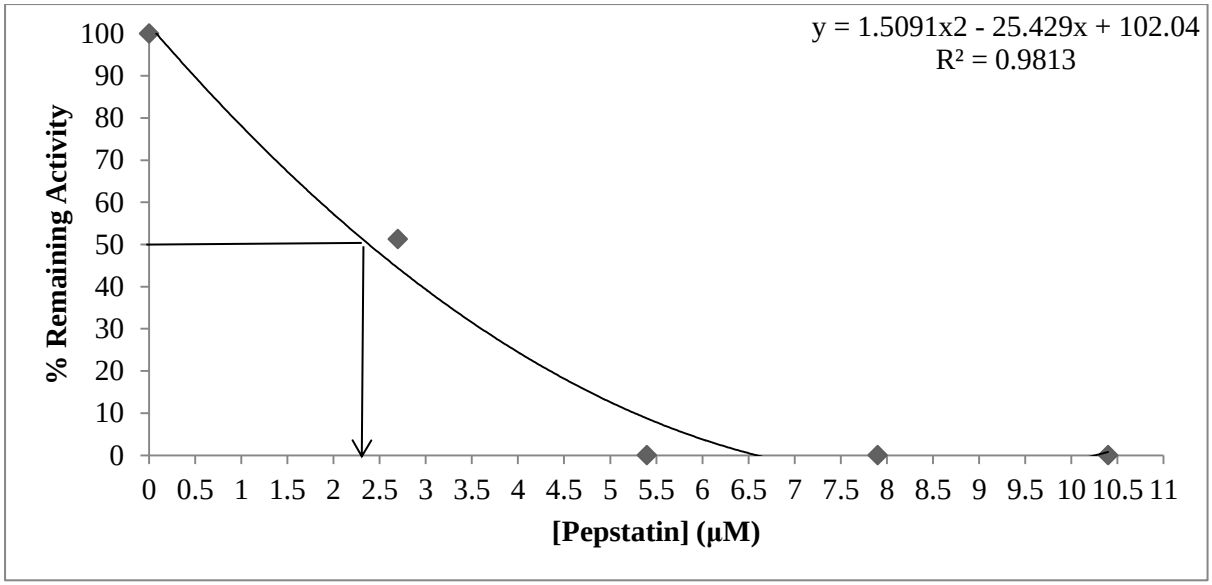


Figure 7.1c: Effect of Pepstatin on PEP activity. PEP extracts in the absence of the inhibitor were used as positive controls. The data are the average of three experiments.

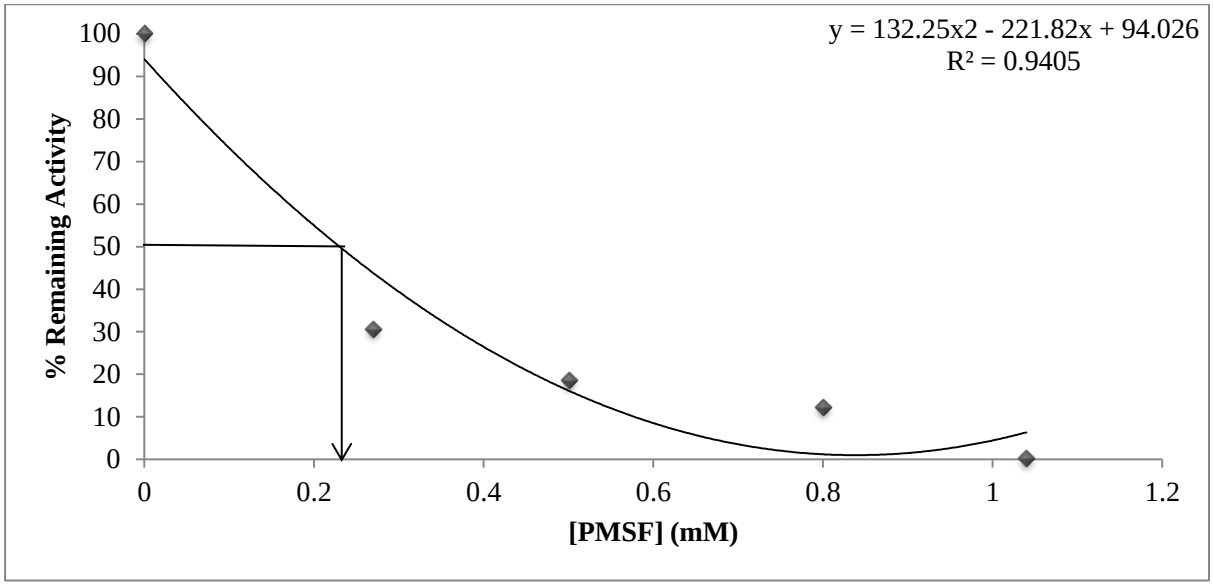


Figure 7.1d: Effect of PMSF on PEP activity. PEP extracts in the absence of the inhibitor were used as positive controls. The data are the average of three experiments.

The study of the effect of inhibitors of the active site groups demonstrated that the activity of PEP was completely inhibited by an inhibitor of serine proteases, PMSF (IC_{50} =0.22mM) as well as an inhibitor of aspartyl proteases, Pepstatin (IC_{50} =2.23µM). Inhibitors of cysteine proteinases, E-64 (IC_{50} =4.4µM) and Leupeptin (IC_{50}=51µM) (also an inhibitor of serine and threonine peptidases) demonstrated substantial inhibition on the activity of wheat PEP.

The PMSF inhibition data appears to be in accordance with the data of Elpidina and colleagues (2005) in which they have demonstrated complete inhibition by PMSF with a chymotrypsin-like Prolyl endopeptidase proteinase from the midgut of *Tenebrio molitor* larvae however, contradicting findings were shown by Bastos and colleagues (2010) in which their PMSF data showed very little inhibition (18%) of a PEP isolated from *Trypanosoma brucei*.

7.3.2 Determination of IC_{50}

As can be seen from figures 1a-d, 51μM, 4.4 μM, 2.23 μM and 0.22 mM concentration of Leupeptin, E64, Pepstatin and PMSF respectively was required to inhibit 50% of PEP activity. In order to establish the type of inhibition involved, Lineweaver-Burk plots were constructed using the concentrations that gave up to 50% residual activity. In this study, Leupeptin was used to determine the type of enzyme inhibition involved as E64 and PMSF are irreversible inhibitor.

7.3.3 Determination of the type of enzyme inhibition

The nature of inhibition of PEP activity by Leupeptin was determined by the method of Dixon (1953), in which the reciprocal velocity, 1/v, was plotted against the inhibitor concentration, [I]. The extrapolated lines at different substrates [s] values intersect at a single point, for which [I] =-$K_{i,}$ the inhibition constant. However, the limitation to this kind of plot is that it does not always distinguish competitive inhibition from mixed inhibition. For this reason, Cornish-Bowden (1974) recommended using an additional plot in conjunction with the method of Dixon to provide a more accurate and unambiguous indication of the type of inhibition.

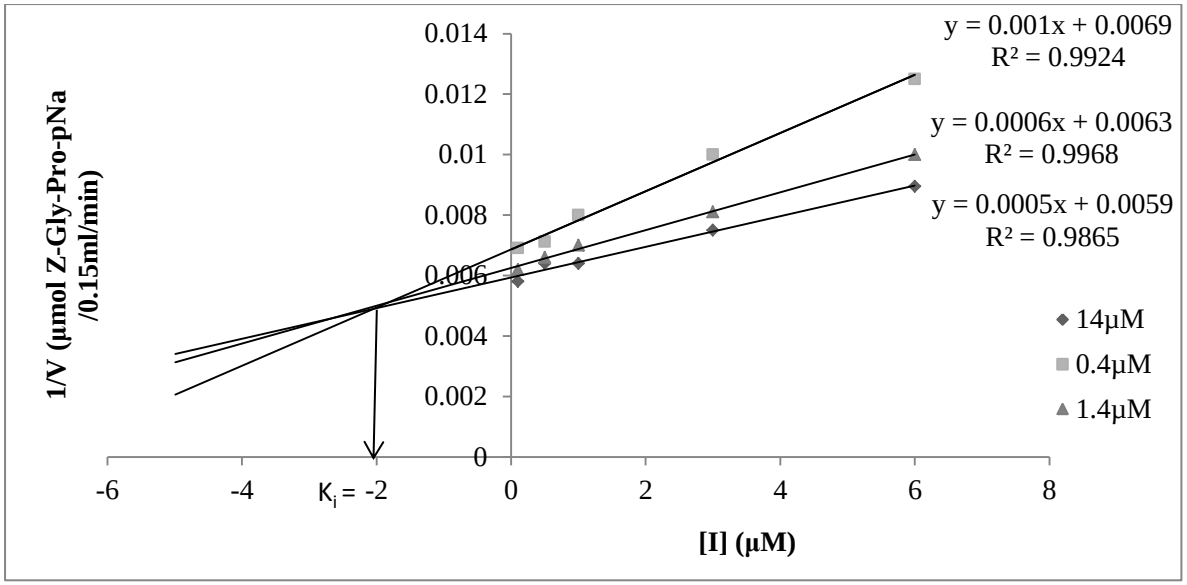


Figure 7.2: Dixon plot for PEP at different inhibitor and substrate concentrations. The data represents the mean triplicate assays.

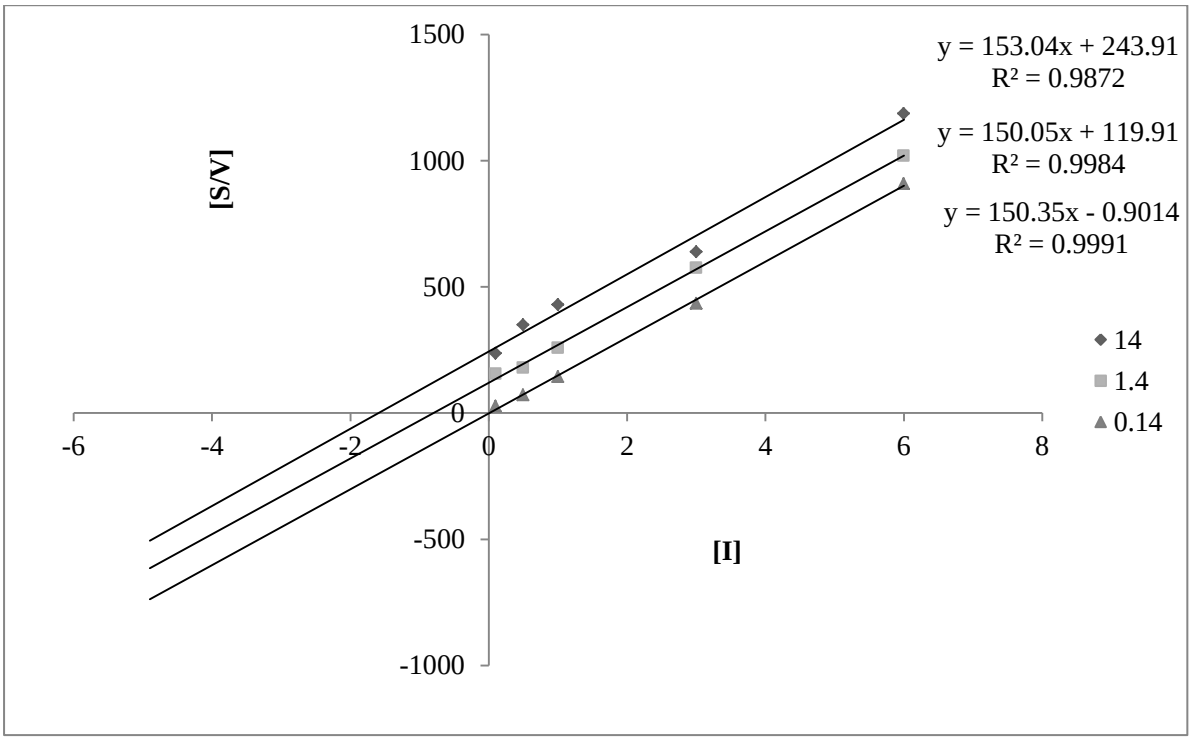


Figure 7.3: Plots of [s]/v against [I]. The data represents the mean of triplicate assays.

The nature of inhibition as predicted from the plots in figure 2 and figure 3 indicates competitive like type of inhibition against PEP as the comparative inhibitor displayed an increase in the K_m (affinity of the enzyme). The inhibitor constant, K_i, which is the concentration of inhibitor (in this study, Leupeptin) that is required to decrease the maximal

rate of the reaction to half of the uninhibited value, in the presence of a low substrate concentration, as extrapolated from the Dixon plot was approximately 2µM. It should also be noted that the lower the K_i, the lower the concentration of inhibitor needed to lower the rate and therefore the more effective the inhibitor.

7.3.4 Comparison of inhibitory activity towards PEP

On comparing the inhibitory activities of leupeptin, E64, pepstatin and PMSF on purified plant PEP, some interesting observations were found. All four class specific inhibitors appeared to show inhibitory potential towards plant PEP in a concentration manner. As expected, leupeptin and PMSF showed inhibition of PEP hydrolysis of Z-Gly-Pro-pNa as these are known serine protease inhibitors. However, unexpectedly inhibition towards PEP by E64, which is accepted as a specific inhibitor for cysteine peptidases, was also observed. Studies by Novillo et al., (1997) also reported inhibition of trypsin-like gut proteases from larvae of the lepidopteran species by E64 in a reversible competitive mechanism. They reported that this inhibition is only for trypsin-like proteases. This would therefore imply that the PEP from plants is a trypsin-like enzyme.

Pepstatin inhibition of plant PEP is an anomaly which is difficult to explain. If there were more than one species of PEP, one being a serine protease with the other an aspartic protease, there would have been some residual activity remaining after treatment with Pepstatin, due to the remaining activity of the serine-like PEP protease. However, Pepstatin totally inhibited the PEP activity at a concentration of 6.5µM, thereby implying that the purified plant PEP belonged to both the serine and the aspartic protease family. This would need to be investigated in future studies.

The inhibition studies would therefore indicate that the purified plant PEP is a serine-like protease, most likely of the trypsin family. This is in contradiction to the substrate specificity study which indicates the plant PEP to be a chymotrypsin-like protease. To confirm whether PEP is trypsin of chymotrypsin like would require further studies using a larger variety of substrates with either trypsin or chymotrypsin specificity built into the scissile bond. These substrates will have to be specifically synthesized as no such substrates are currently commercially available for PEPs.

7.4 References

1) Bastos IM, Motta FN, Charneau S, Santana JM, Dubost L, Augustyns K and Grellier P.(2010). Prolyl oligopeptidase of Trypanosoma brucei hydrolyzes native collagen, peptide hormones and is active in the plasma of infected mice. Microbes Infect.12(6):457-66
2) Copeland, RA. (2000). Enzymes. A practical introduction to structure, mechanism and data analysis. Wiley and Sons, New York.
3) Cornish-Bowden, A. (1974). A Simple Graphical Method for Determining the Inhibition Constants of Mixed,Uncompetitive and Non-Competitive Inhibitors. Biochem. J. 137:143-144
4) Dixon, M. (1953). Biochem. J. 55, 170-171
5) Meleti, T. (1986). In basic enzyme kinetics, Ed. Akademia Khado, Budapest

CHAPTER 8

GENERAL DISCUSSION, CONCLUSION AND FUTURE STUDIES

8.1 Discussion

Compared with the classic trypsin and subtilisin families, the prolyl endopeptidase family represents a relatively new class of serine peptidases. In this study, we have demonstrated the first isolation, purification and kinetic characterizing of a plant PEP, in particular from germinating wheat seeds. Oligopeptidase B, which was previously called protease II, an homologous enzyme to prolyl endopeptidase was identified in quiescent wheat embryo (Tsuji et al, 2004) but no purification studies of this enzyme were conducted. The enzyme displayed trypsin-like specificity (cleaving peptides at lysine and arginine residues), similar to what we are reporting here despite the fact that the enzyme we are reporting here also has a potential preference for a chymotrypsin-like substrate.

Separation of purified wheat PEP was achieved by anion exchange with elution using a salt gradient as well as gel filtration chromatography using a slightly acidic buffer without salt. The ion exchange elution profiles together with their corresponding SDS-PAGE gels clearly shows that PEP eluted with other peptides of various sizes very early in the linear salt gradient (0-1M NaCl over 20CV).

After the two step purification, PEP appeared to be relatively pure with only 2 additional bands of lower molecular weight observed using SDS-PAGE. The lower molecular weight bands could either be additional subunits or it could possibly infer the presence of multiple PEPs or other protein contaminants. We estimated PEP from germinating wheat seeds to be ~55kDa which is in line with the expected molecular weight range of 45-174kDa reported for known PEPs in literature.

The data on the substrate specificity indicates that PEP may be a chymotrypsin-like enzyme as it had a stronger preference for the chymotrypsin-like substrate with a possible extended binding site. The enzyme hydrolyzed a longer peptide substrate more efficiently (N-Suc-Ala-Ala-Pro-Phe-pNa) containing four amino acid residues with Phe in the P1 position than the shorter, proline specific substrate with proline occupying the P1 position (Z-Gly-Pro-pNa). PEP also appeared to have the highest affinity for N-Suc-Ala-Ala-Pro-Phe-pNa while the highest reaction rate (*k*cat) and efficiency (*k*cat/*K*m) were also observed for this substrate.

The effect of the active site inhibitors indicated that the purified enzyme is a potential serine proteinase as it was inhibited by PMSF and leupeptin. In addition it was also inhibited by E64, which in addition to inhibiting cysteine proteinases, has been shown to also inhibit trypsin-like enzymes, thereby implying that the purified plant PEP is a trypsin-like protease. This contradicts with the results of substrate specificity and only once a whole range of

substrates are used, would one be able to decisively conclude to which family of enzymes the purified plant PEP belongs.

8.2 Conclusion

This is the first report of a plant prolyl endopeptidase, purified from germinating wheat seeds. The results reported here in this thesis indicate that the enzyme is a serine protease belonging to either the trypsin or chymotrypsin-like family. The physical and kinetic properties of the purified plant PEP as similar to those reported for other known PEPs from bacterial, mammalian or insect origins.

8.3 Future work

As part of our future studies, we would like to explore the following possible avenues:

1. To further optimise germinating conditions of wheat seed for higher yield of PEP. This would include setting up different convirons and exploring different growth conditions to maximize germination of seeds. Also to try and use a variety of wheat seeds to see which of these gives a better yield in isolating PEP. Furthermore, to germinate more in bulk so as to have sufficient stock of protein exctracts.
2. We also plan to investigate the possible presence of multiple PEPs in plants as this is quite a novel enzyme in plants. Our results reported here shows that they may be a possibility of potential subunits for this enzyme as multiple bands and inhibition studies has shown.
3. Once we confirm the possibility of multiple plant PEPs, we will then further investigate whether PEPs belong to the chymotrypsin or trypsin family of proteases using a larger range of substrates. In this study we have reported only on two substrates, that being a chymotrypsin-like and prolyl endopeptidase specific substrates.
4. We will also like to establish the enzymatic mechanism for the inhibition of PEP through the use of a broader range of inhibitors that will confirm to which group of enzymes wheat PEPs belong to.
5. As a final objective, we would also like to elucidate a possible sequence for the purified wheat PEP through sequencing and to determine its molecular weight as our hypothesis we based on largely sequence alignments of other related organisms.

8.4 References

1) Tsuji A, Yuasa K and Matsuda Y. (2004). Identification of oligopeptidase B in higher plants. purification and characterization of oligopeptidase B from quiescent wheat embryo, Triticum aestivum. J.Biochem. **136**: 673-681

APPENDIX A

CIS-ACTING ELEMENT	FREQUENCY OF OCCURRENCE IN RICE CHR 1	FREQUENCY OF OCCURRENCE IN RICE CHR 4	FREQUENCY OF OCCURRENCE IN AtPEP1	FREQUENCY OF OCCURRENCE IN AtPEP2	TOTAL NUMBER OF OCCURRENCE	FUNCTION
3-AF1 binding site				1	1	light responsive element
5UTR Py-rich stretch				1	1	cis-acting element conferring high transcription levels
-10PEHVPSBD			2	2	4	"-10 promoter element" found in the barley (H.v.) chloroplast psbD gene promoter; Involved in the expression of the plastid gene psbD which encodes a photosystem II reaction center chlorophyll-binding protein that is activated by blue, white or UV-A light;
-300ELEMENT	1		1		2	Present upstream of the promoter from the B-hordein gene of barley and the alpha-gliadin, gamma-gliadin, and low molecular weight glutenin genes of wheat
AACACOREOSGLUB1			2		2	Core of AACA motifs found in rice (O.s.) glutelin genes, involved in controlling the endosperm-specific expression; AACA is also closely associated with the GCN4 motif in all rice glutelin genes and together have been shown to confer endosperm-specific enhancement to the truncated -90 CaMV 35S promoter
AAGAAMOTIF		2		1	3	
ABRE		1			1	Abscisic acid responsive element
ABREATCONSENSUS	4				4	ABA-responsive elements (ABREs) found in the promoter of ABA and/or stress-regulated genes; ABFs, a family of ABRE binding factors; ABF3 and ABF4 function in ABA signaling; Y=C/T;
ABRELATERD1	5	1	2	1	9	ABRE-like sequence required for etiolation-induced expression of erd1(early responsive to dehydration) in Arabidopsis
ABREOSRAB21	3				3	ABA responsive element(ABRE) of wheat Em and rice (O.s) rab21 genes
ABRERATCAL	5		2	1	8	"ABRE-related sequence" or "Repeated sequence motifs" identified in the upstream regions of 162 Ca(2+)-responsive upregulated genes
ABREZMRAB28	4				4	ABRE; ABA and water-stress responses; Found in maize (Z.m.) rab28; maize rab28 is ABA-inducible in embryos and vegetative tissues; Found in the Arabidopsis (A.t.) alcohol dehydrogenase (Adh) gene promoter; "ABRE2"; Found in the maize (Z.m.) Cat1 gene promoter; Responsible for the induction by ABA; Binding site of CBF2; Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance; The CBF genes do not appear to be autoregulated through the CRT/DRE sequence
ACE			1		14	Involved in light responsiveness
ACGTABREMOTIFA2OSEM	4				4	Experimentally determined sequence requirement of ACGT-core of motif A in ABRE of the rice gene, OSEM; See S000281; DRE and ABRE are interdependent in the ABA-responsive expression of the rd29A in Arabidopsis
ACGTATERD1	10	2	6	3	21	ACGT sequence required for etiolation-induced expression of erd1(early responsive element to dehydration) in Arabidopsis
ACGTCBOX			1		1	"C-box" according to the nomenclature of ACGT elements by Foster et al. (FASEB J

						8:192-200 (1994)); One of ACGT elements; Factors groups 1, 2 and 3 have affinity for C-box (Izawa et al. J Mol Biol. 230:1131-1144 (1993)); RITA-1 binding site (Izawa et al. 1994);
ACGTROOT1	4				4	"ACGT motif" related to root expression; Gene: synthetic; perfect palindromic sequence (PA) containing G-box-related sequence; transacting factor: TAF-1 ?; Binding of SGBF-1 (a Soybean G-box binding bZIP transcription factor) to ABRE is enhanced by SCOF-1 (a zinc finger protein); Transcription of SCOF-1 is induced by low temperature and ABA;
ACGTTBOX			1		1	"T-box" according to the nomenclature of ACGT elements by Foster et al. (FASEB J 8:192-200 (1994)); One of ACGT elements; See also ACGTABOX (S000130), ACGTCBOX (S000131), and CACGTGMOTIF (S000042);
AGCBOXNPGLB				1	1	Binding sequence of Arabidopsis AtERFs; AtERF1,2 and 5 functioned as activators of GCC box-dependent transcription; AtERF3 and 4 acted as repressors; AtERF proteins are stress signal-response factors; EREBP2 binding site; Conserved in most PR-protein genes; Rice MAPK (BWMK1) phosphorylates OS EREBP1, which enhance DNA-binding activity of the factor to the GCC box;
AMYBOX1			2		2	"amylase box"; Conserved sequence found in 5'-upstream region of alpha-amylase gene of rice, wheat, barley
AMYBOX2			1		1	"amylase box"; "amylase element"; Conserved sequence found in 5'upstream region of alpha-amylase gene of rice, wheat, barley
ANAERO1CONSENSUS			5		5	One of 16 motifs found in silico in promoters of 13 anaerobic genes involved in the fermentative pathway
ANAERO2CONSENSUS		3			3	One of 16 motifs found in silico in promoters of 13 anaerobic genes involved in the fermentative pathway
ANAERO3CONSENSUS		1	1	1	3	One of 16 motifs found in silico in promoters of 13 anaerobic genes involved in the fermentative pathway
ASF1MOTIFCAMV			1	1	2	"ASF-1 binding site" in CaMV 35S promoter; ASF-1 binds to two TGACG motifs; See S000023 (AS1); Found in HBP-1 binding site of wheat histone H3 gene; TGACG motifs are found in many promoters and are involved in transcriptional activation of several genes by auxin and/or salicylic acid; May be relevant to light regulation; Binding site of tobacco TGA1a; TGA1a and b show homology to CREB; TGA6 is a new member of the TGA family; Abiotic and biotic stress differentially stimulate "as-1 element" activity
ARE	2	1	1	2	6	cis-acting regulatory element essential for the anaerobic induction
ARFAT	1				1	ARF (auxin response factor) binding site found in the promoters of primary/early auxin response genes of Arabidopsis thaliana (A.t.); AuxRE; See S000337; Binding site of Arabidopsis ARF1 (Auxin response factor1); Sequence found in NDE element in Soybean (G.m.) SAUR (Small Auxin-Up RNA) 15A gene promoter; See S000359, S000360; Found in D1 or D4 element in Soybean (G.m.) GH3 promoter;

						This element was enriched in the 5'-flanking region of genes up-regulated by both IAA and BL
ARR1AT	10	18	11	12	51	"ARR1-binding element" found in Arabidopsis; ARR1 is a response regulator; N=G/A/C/T; AGATT is found in the promoter of rice non-symbiotic haemoglobin-2 (NSHB) gene
ATCT-MOTIF	1		1	1	3	part of a conserved DNA module involved in light responsiveness
BIHD1OS	2	3	2	1	8	Binding site of OsBIHD1, a rice BELL homeodomain transcription factor
BOXI			1	1	2	Light responsive element
BOXII		1			1	Light responsive element
BOXIII				1	1	protein binding site
BOXIINTPATPB			2	2	4	"Box I" found in the tobacco (N.t.) plastid atpB gene promoter; Conserved in several NCII (nonconsensus type II) promoters of plastid genes; Important for the activity of this NCII promoter
BOXIIPCCHS	4				4	Core of "Box II/G box" found in the parsley (P.c.) chs genes; Essential for light regulation
BOXLCOREDCPAL	1				1	Consensus of the putative "core" sequences of box-L-like sequences in carrot (D.c.) PAL1 promoter region; DCMYB1 bound to these sequences in vitro
BOXW1			2		2	Fungal elicitor responsive element
BP5OSWX		1			1	OsBP-5 (a MYC protein) binding site in Wx promoter
BS1EGCCR		1			1	"BS1 (binding site 1)" found in E. gunnii Cinnamoyl-CoA reductase (CCR) gene promoter; nuclear protein binding site; Required for vascular expression;
CAATBOX1	9	10	7	3	29	Responsible for tissue specific promoter activity of a pea legumin gene in tobacco
CAATBOX	5	11	6	6	28	common cis-acting element in promoter and enhancer regions
CACGTGMOTIF	5			1	6	"CACGTG motif"; "G-box"; Binding site of Arabidopsis GBF4; C. roseus G-box binding factor 1 (CrGBF1) and 1 (CrGBF2) can act as transcriptional repressors of the Str promoter via direct interaction with the G-box; See S000345; Essential for expression of beta-phaseolin gene during embryogenesis in bean, tobacco, Arabidopsis; Tomato Pti4 (ERF) regulates defense-related gene expression via GCC box and non-GCC box cis-element (Myb1 (GTTAGTT) and G-box (CACGTG));
CACTFTPPCA1	19	7	15	16	57	Tetranucleotide (CACT) is a key component of Mem1 (mesophyll expression module 1) found in the cis-regulatory element in the distal region of the phosphoenolpyruvate carboxylase (ppcA1) of the C4 dicot F. trinervia
CAREOSREP1	1		1	1	3	"CAREs (CAACTC regulatory elements)" found in the promoter region of a cystein proteinase (REP-1) gene in rice
CARGCW8GAT	1	2	1	2	6	A variant of CArG motif (see S000404), with a longer A/T-rich core; Binding site for AGL15 (AGAMOUS-like 15);
CATATGGMSAUR			1	1	2	Sequence found in NDE element in soybean (G.m.) SAUR (Small Auxin-Up RNA) 15A gene promoter; Involved in auxin responsiveness
CBFHV	1			1	2	Binding site of barley (H.v.) CBF1, and also of barley CBF2; CBF = C-repeat (CRT) binding factors; CBFs are also known as dehydration-responsive element (DRE) binding proteins (DREBs

CCAATBOX1	2	1	3	1	7	Found in the promoter of heat shock protein genes
CCAATBOX	1			1	2	MYBHv1 binding site
CEREGLUBOX2PSLEGA				1	1	"cereal glutenin box" in pea legumin gene (legA); sequence homologous to the cereal glutenin gene control element ("-300 element");
CGACGOSAMY3		1			1	"CGACG element" found in the GC-rich regions of the rice (O.s.) Amy3D and Amy3E amylase genes, but not in Amy3E gene; May function as a coupling element for the G box element
CGCGBOXAT	1	3	4	2	10	"CGCG box" recognized by AtSR1-6 (Arabidopsis thaliana signal-responsive genes); Multiple CGCG elements are found in promoters of many genes; Ca++/calmodulin binds to all AtSRs
CGTCAMOTIF	1	3			4	Involved in MeJA responsiveness
CIACADIANLELHC		2	1		3	Region necessary for circadian expression of tomato (L.e.) Lhc gene;
CIRCADIAN	1				1	Involved in circadian control
CPBCSPOR				1	1	The sequence critical for Cytokinin-enhanced Protein Binding in vitro, found in -490 to -340 of the promoter of the cucumber (CS) POR (NADPH-protochlorophyllide reductase) gene
CPRFPCCHS	3				3	"BoxII"; Binding site of CPRF-1, -2, -3 and -4(Common Plant Regulatory Factor) in the parsley (P.c.) light responsive chalcone synthase (CHS) gene promoter; CPRF proteins are bZIP class transcription factors; CPRF proteins participates in the light-mediated activation of the CHS gene in parsley; "ACE"; The proline-rich domains of CPRF1 and 4 activate transcription; CPRF1-containing bZIP heterodimer interacts with ACE in vivo;
CRTDREHVCBF2				1	1	Preferred sequence for AP2 transcriptional activator HvCBF2 of barley; "Core CRT/DRE motif"; HvCBF2 bound to a (G/a)(T/c)CGAC core motif (Xue, 2003); DNA binding is regulated by temperature
CURECORECR	11	3	4	5	23	GTAC is the core of a CuRE (copper-response element) found in Cyc6 and Cpx1 genes in Chlamydomonas; Also involved in oxygen-response of these genes
DOFCOREZM	10	13	21	11	55	Core site required for binding of Dof proteins in maize (Z.m.); Dof proteins are DNA binding proteins, with presumably only one zinc finger, and are unique to plants; Four cDNAs encoding Dof proteins, Dof1, Dof2, Dof3 and PBF, have been isolated from maize; PBF is an endosperm specific Dof protein that binds to prolamin box; Maize Dof1 enhances transcription from the promoters of both cytosolic orthophosphate kinase (CyPPDK) and a non-photosynthetic PEPC gene; Maize Dof2 supressed the C4PEPC promoter
DPBFCOREDCDC3	1				1	A novel class of bZIP transcription factors, DPBF-1 and 2 (Dc3 promoter-binding factor-1 and 2) binding core sequence; Found in the carrot (D.c.) Dc3 gene promoter; Dc3 expression is normally embryo-specific, and also can be induced by ABA; The Arabidopsis abscisic acid response gene ABI5 encodes a bZIP transcription factor; abi5 mutant have a pleiotropic defects in ABA response; ABI5 regulates a subset of late embryogenesis-abundant genes; GIA1 (growth-insensitivity to ABA) is identical to

						ABI5
E2FCONSENSUS		1			1	"E2F consensus sequence" of all different E2F-DP-binding motifs that were experimentally verified in plants
EBOXBNNAPA	12	7	5	7	31	E-box of napA storage-protein gene of Brassica napus
EECCRCAH1	6	2	3	1	12	"EEC"; Consensus motif of the two enhancer elements, EE-1 and EE-2, both found in the promoter region of the Chlamydomonas Cah1 (encoding a periplasmic carbonic anhydrase); Binding site of Myb transcription factor LCR1
ELRECOREPCRP 1						ElRE (Elicitor Responsive Element) core of parsley (P.c.) PR1 genes; consensus sequence of elements W1 and W2 of parsley PR1-1 and PR1-2 promoters; Box W1 and W2 are the binding site of WRKY1 and WRKY2, respectively; ERE; "WA box"; One of the W boxes found in the Parsley (P.c.) WRKY1 gene promoter; Required for elicitor responsiveness
EMBP1TAEM	4				4	Binding site of trans-acting factor EMBP-1; wheat (T.a.) Em gene; See S000015; Binding site of ABFs; ABFs (ABRE binding factors) were isolated from Arabidopsis by a yeast one-hybrid screening system; Expression ABFs is induced by ABA and various stress treatment; ABFs belongs to a distinct subfamily of bZIP proteins; Involved in ABA-mediated stress-signaling pathway
GAGMOTIF	1	1			1	Part of light responsive element
GAREMOTIF	1	1	1		3	Gibberellins-responsive element
GARE2OSREP1			1		1	"Gibberellin-responsive element (GARE)" found in the promoter region of a cystein proteinase (REP-1) gene in rice
GAREAT			2		2	GARE (GA-responsive element); Occurrence of GARE in GA-inducible, GA-responsible, and GA-nonresponsive genes found in Arabidopsis seed germination was 20, 18, and 12%, respectively
GATABOX	15	11	14	17	57	"GATA box"; GATA motif in CaMV 35S promoter; Binding with ASF-2; Three GATA box repeats were found in the promoter of Petunia (P.h.) chlorophyll a/b binding protein, Cab22 gene; Required for high level, light regulated, and tissue specific expression; Conserved in the promoter of all LHCII type I Cab genes
GBOXLERBCS	5				5	"G box"; Conserved sequence upstream of light-regulated genes; Sequence found in the promoter region of rbcS of tomato (L.e.) and Arabidopsis; Binding with GBF
GCMOTIF		2			2	Enhancer-like element involved in anoxic specific inducibility
GCCCORE	2			1	3	Core of GCC-box found in many pathogen-responsive genes such as PDF1.2, Thi2.1, and PR4; Has been shown to function as ethylene-responsive element; Appears to play important roles in regulating jasmonate-responsive gene expression; Tomato Pti4 (ERF) regulates defence-related gene expression via GCC box and non-GCC box cis elements (Myb1 (GTTAGTT) and G-box(CACGTG));
GCN4_motif				1	1	cis-regulatory element involved in endosperm expression
GCN4OSGLUB1		1		1	2	"GCN4 motif" found in GluB-1 gene in rice (O.s.); Required for endosperm-specific expression; See S000276; AACA and ACGT

						motifs was found sufficient to confer a detectable level of endosperm expression; See S000353, S000354; This motif is the recognition site for a basic leucine zipper transcription factor that belongs to the group of maize Opaque-2 (O2)-like proteins; Although all the RISBZ proteins are able to interact with the GCN4 motif, only RISBZ1 is capable of activating the gene expression;
GT1CONSENSUS	8	4	16	10	38	Consensus GT-1 binding site in many light-regulated genes, e.g., RBCS from many species, PHYA from oat and rice, spinach RCA and PETA, and bean CHS15; R=A/G; W=A/T; For a compilation of related GT elements and factors, see Villain et al. (1996); GT-1 can stabilize the TFIIA-TBP-DNA (TATA box) complex; The activation mechanism of GT-1 may be achieved through direct interaction between TFIIA and GT-1; Binding of GT-1-like factors to the PR-1a promoter influences the level of SA-inducible gene expression
GT1GMSCAM4	3	1	5	1	10	"GT-1 motif" found in the promoter of soybean (Glycine max) CaM isoform, SCaM-4; Plays a role in pathogen- and salt-induced SCaM-4 gene expression
GTGANTG10	7	10	11	6	34	"GTGA motif" found in the promoter of the tobacco (N.t.) late pollen gene g10 which shows homology to pectate lyase and is the putative homologue of the tomato gene lat56; Located between -96 and -93;
HDMOTIFPCPR2	1				1	HD (homeodomain) protein target site in parsley (P.c.) pathogenesis-related protein 2 (PR2); A potential in vivo target site
HEXAT			1		1	"Hex motif"; Binding site of Arabidopsis (A.t.) bZIP protein TGA1 and G box binding factor GBF1; TGA1 and members of the GBF family differ in their DNA binding properties; G-Box-like element
HEXMOTIFTAH3H4			1		1	"hexamer motif" found in promoter of wheat (T.a.) histone genes H3 and H4; CaMV35S; NOS; Binding with HBP-1A and HBP-1B; Binding site of wheat (T.a.) nuclear protein HBP-1 (histone DNA binding protein-1); HBP-1 has a leucine zipper motif; "hexamer motif" in type 1 element may play important roles in regulation of replication-dependent but not of replication-independent expression of the wheat histone H3 gene; See S000076, S000267; Rice OBF1-homodimer-binding site
HSE			1	1	2	cis-acting element involved in heat stress responsiveness
IBOX				2	2	"I box"; "I-box"; Conserved sequence upstream of light-regulated genes; Sequence found in the promoter region of rbcS of tomato and Arabidopsis; I box (Giuliano et al. 1988); Binding site of LeMYB1, that is a member of a novel class of myb-like proteins; LeMYBI act as a transcriptional activator
IBOX		1		2	3	"I box"; "I-box"; Conserved sequence upstream of light-regulated genes; Sequence found in the promoter region of rbcS of tomato and Arabidopsis; I box (Giuliano et al. 1988); Binding site of LeMYB1, that is a member of a novel class of myb-like proteins; LeMYBI act as a transcriptional activator
IBOXCORE	3	2	6	7	18	"I box"; "I-box"; Conserved sequence

						upstream of light-regulated genes; Conserved sequence upstream of light-regulated genes of both monocots and dicots
IBOXCORENT				2	2	"I-box core motif" in the CAMs (conserved DNA modular arrays) associated with light-responsive promoter regions
INRNTPSADB	2	1	3	4	10	Inr(initiator) elements found in the tobacco psaDb gene promoter without TATA boxes; light responsive transcription of psaDb depends on Inr, but not TATA box
INTRONLOWER	2				2	"3' intron-exon splice junctions"; Plant intron lower sequence; Consensus sequence for plant introns
IRO2OS	5				5	OsIRO2-binding core sequence; "G-box plus G"; Transcription factor OsIRO2 is induced exclusively by Fe deficiency;
L1BOXATPDF1	1		2	1	4	"L1 box" found in promoter of Arabidopsis thaliana (A.t.) PROTODERMAL FACTOR1 (PDF1) gene; Located between -134 and -127; Involved in L1 layer-specific expression; L1-specific homeodomain protein ATML can bind to the "L1 box"; Y=C/T; A cotton fiber gene, RD22-like 1 (RDL1), contains a homeodomain binding L1 box and a MYB binding motif
LECPLEACS2		1			1	Core element in LeCp (tomato Cys protease) binding cis-element (from -715 to -675) in LeAcs2 gene
LTR				1	1	cis-acting element involved in low-temperature responsiveness
LTRE1HVBLT49				1	1	"LTRE-1" (low-temperature-responsive element) in barley (H.v.) blt4.9 gene promoter; A new LTRE; A previously known LTRE is CCGAC
LTRECOREATCOR15	1		1	1	3	Core of low temperature responsive element of cor15a gene in Arabidopsis
MARTBOX			1	3	4	"T-Box"; Motif found in SAR (scaffold attachment region; or matrix attachment region, MAR);
MBS		2		1	3	MYB binding site involved in drought-inducibility
MBSI			1	1	2	MYB binding site involved in light responsiveness
MRE			1		1	MYB binding site involved in light responsiveness
MYB1AT			1	7	8	MYB recognition site found in the promoters of the dehydration-responsive gene rd22 and many other genes in Arabidopsis
MYB2AT				1	1	Binding site ATMYB2, an Arabidopsis MYB homolog; ATMYB2 is involved in regulation of genes that are responsive to water stress in Arabidopsis
MYB1LEPR		1			1	Tomato Pti4(ERF) regulates defence-related gene expression via GCC box and non-GCC box cis elements (Myb1(GTTAGTT), G box (CACGTG));
MYB2CONSENSUSAT	1	1		3	5	MYB recognition site found in the promoters of the dehydration- responsive gene rd22 and many other genes in Arabidopsis
MYBATRD22				1	1	Binding site for MYB (ATMYB2) in dehydration-responsive gene, rd22; MYB binding site in rd22 gene of Arabidopsis thaliana; ABA-induction; Located at ca. -141 of rd22 gene;
MYBCORE	2	3	2	3	10	Binding site for all animal MYB and at least two plant MYB proteins ATMYB1 and ATMYB2. ATMYB2 is involved in regulation of genes that are responsive to water stress in Arabidopsis.
MYBCOREATCY	1	1		3	5	"Myb core" in the 18 bp sequence which is

CB1						able to activate reporter gene without leading to M-phase-specific expression, found in the promoter of Arabidopsis thaliana cyclin B1:1 gene; the 18 bp sequence share homology with a sequence found in the N. sylvestris cyclin B1 promoter
MYBGAHV			2		2	Central element of gibberellin (GA) response complex (GARC) in high-pI alpha-amylase gene in barley (H.v.); Similar to c-myb and v-myb consensus binding site; GAmyb binds specifically to the TAACAAA box in vitro; GAmyb is the sole GA-regulated transcriptional factor required for transcriptional activation of the high-pI alpha-amylase; GARC consist of the pyrimidine, TAACAAA and TATCCAC boxes; GARE in RAmy1A gene; GARE and pyrimidine box in RAmy1A are partially involved in sugar repression
MYBPLANT	1				1	Regulate phenylpropanoid and lignin biosynthesis
MYBPZM	1				1	Core of consensus maize P (myb homolog) binding site; W=A/T; 6 bp core; Maize P gene specifies red pigmentation of kernel pericarp, cob, and other floral organs; P binds to A1 gene, but not Bz1 gene; Maize C1 (myb homolog) activates both A1 and Bz1 genes
MYBST1	1		4	3	8	Core motif of MybSt1 (a potato MYB homolog) binding site; MybSt1 cDNA clone was isolated by using CaMV 35S promoter domain A as a probe (Baranowskij et al. 1994); The Myb motif of the MybSt1 protein is distinct from the plant Myb DNA binding domain described so far;
MYCATERD1			2		2	MYC recognition sequence necessary for expression of erd1(early responsive to dehydration) in dehydrated Arabidopsis
MYCATRD22	1	3			4	Binding site for MYC (rd22BP1) in Arabidopsis dehydration-responsive gene, rd22
MYCCONSENSUSAT	12	7	5	7	31	MYC recognition site found in the promoters of dehydration-responsive gene rd22 and many other genes in Arabidopsis
NAPINMOTIFBN				1	1	Sequence found in 5' upstream region (-6, -95, -188) of napin (2S albumin) gene in Brassica napus (B.n.); Interact with a protein present in crude nuclear extracts from developing B. napus seeds;
NODCON1GM		1			1	One of two putative nodulin consensus sequences
NODCON2GM	1	2	2	1	6	One of two putative nodulin consensus sequences
NTBBF1ARROLB			1	2	3	NtBBF1(Dof protein from tobacco) binding site in Agrobacterium rhizogenes (A.r.) rolB gene; Found in regulatory domain B (-341 to -306); Required for tissue-specific expression and auxin induction;
O2-SITE		1	1		2	Involved in zein metabolism regulation
OSE1ROOTNODULE			2		2	One of the consensus sequence motifs of organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules
OSE2ROOTNODULE	1	1	2	1	5	One of the consensus sequence motifs of organ-specific elements (OSE) characteristic of the promoters activated in infected cells of root nodules
P1BS	1			1	2	PHR1-binding sequence found in the upstream regions of phosphate starvation responsive genes from several plant species; phr1 (phosphate starvation

						response 1) gene codes for PHR1 protein related to PSR1 gene in C. reinhardtii;
PALBOXAPC		1			1	Box A; Consensus; One of three putative cis-acting elements (boxes P, A, and L) of phenylalanine ammonia-lyase (PAL; EC 4.3.1.5) genes in parsley (P.c.); "None of these elements (boxes P, A, and L) alone, or the promoter region containing all of them together, conferred elicitor or light responsiveness. These elements appear to be necessary but not sufficient for elicitor- or light-mediated PAL gene activation
PBOX			1		1	Gibberellins responsive element
PE2FNTRNR1A		1			1	"pE2F (proximal E2F elemen)" at -143bp of tobacco (N.t.) RNR1a promoter; E2F factors involved in gene induction at the G1/S transition of the cell cycle; Important for regulating specific RNR1a (ribonucleotide reductase large subunit) gene expression in response to UV-C irradiation;
PIATGAPB	1				1	"PI" found in the Arabidopsis thaliana(A.T.) GAPB gene promoter; Located between -157 and -150; Mutations in the "PI" resulted in reductions of light-activated gene transcription; GAPB encodes the B subunit of chloroplast glyceraldehyde-3-phosphate dehydrogenase(GADPH) of A.T.;
POLASIG1		3	5	2	10	"PolyA signal"; poly A signal found in legA gene of pea, rice alpha-amylase; -10 to -30 in the case of animal genes. Near upstream elements (NUE) in Arabidopsis
POLASIG2			1	1	2	"PolyA signal"; poly A signal found in rice alpha-amylase; -10 to -30 in the case of animal genes. AATAAA; AATAAT; AATTAAA; AATAAG;
POLASIG3	2	1	5	3	11	Plant polyA signal; consensus sequence for plant polyadenylation signal.
POLLEN1LELAT52	6	3	13	7	29	One of two co-dependent regulatory elements responsible for pollen specific activation of tomato (L.e.) lat52 gene; Found at -72 to -68 region; See S000246 (POLLEN2LELAT52); AGAAA and TCCACCATA (S000246) are required for pollen specific expression; Also found in the promoter of tomato endo-beta-mannanase gene (LeMAN5) gene
PREATPRODH	1		1	1	3	"PRE (Pro- or hypoosmolarity-responsive element) found in the promoter region of proline dehydrogenase (ProDH) gene in Arabidopsis; Core of 9-bp sequence ACTCATCCT which is necessary for the efficient expression of ProDH in response to L-Pro and hypoosmolarity (Satoh et al., 2002); ATB2-binding site; Similar to GCN4 motif (ATGA(C/G)TCAT); ATB2 subgroup of bZIP transcription factors function as transcriptional activator for hypoosmolarity-inducible ProDH
PRECONSCRHSP70A		2	1	1	4	Consensus sequence of PRE (plastid response element) in the promoters of HSP70A in Chlamydomonas; Involved in induction of HSP70A gene by both MgProto and light
PYRIMIDINEBOXHVEPB1		1			1	Pyrimidine box found in barley (H.v) EPB1 (cysteine proteinase) gene promoter
PYRIMIDINEBOXOSRAMY1A			3	1	4	Pyrimidine box found in rice (O.s.) alpha-amylase (RAmy1A) gene; Gibberellin-respons cis-element of GARE and pyrimidine box are partially involved in sugar repression; Found in the promoter of barley alpha-amylase (Amy2/32b) gene which is

						induced in the aleurone layers in response to GA; BPBF protein binds specifically to this site
QELEMENTZMZM13	1				1	"Q(quantitative)-element" in maize (Z.m.) ZM13 gene promoter; Found at -107 to -102; Involved in expression enhancing activity; ZM13 is a maize homolog of tomato LAT52 gene; ZM13 is a pollen-specific maize gene;
RAV1AAT		4	6	4	14	Binding consensus sequence of Arabidopsis (A.t.) transcription factor, RAV1; RAV1 specifically binds to DNA with bipartite sequence motifs of RAV1-A (CAACA) and RAV1-B (CACCTG); RAV1 protein contain AP2-like and B3-like domains; The AP2-like and B3-like domains recognize the CAACA and CACCTG motifs, respectively; The expression level of RAV1 were relatively high in rosette leaves and roots;
RAV1BAT	1				1	Binding consensus sequence of an Arabidopsis (A.t.) transcription factor, RAV1; RAV1 specifically binds to DNA with bipartite sequence motifs of RAV1-A (CAACA) and RAV1-B (CACCTG); RAV1 protein contain AP2-like and B3-like domains; The AP2-like and B3-like domains recognize the CAACA and CACCTG motifs, respectively; The expression level of RAV1 were relatively high in rosette leaves and roots;
RBCSCONSENSUS			1		1	"rbcS general consensus sequence"; AATCCAA or AATCCAAC;
REALPHALGLHCB21		1		4	5	"REalpha" found in Lemna gibba Lhcb21 gene promoter; Located at -134 to -129; Binding site of proteins of whole-cell extracts; The DNA binding activity is high in etiolated plants but much lower in green plants; Required for phytochrome regulation
REBETALGLHCB21	1			1	2	"REbeta" found in Lemna gibba Lhcb21 gene promoter; Located at -114 to -109; A GATA sequence created at a position six nucleotides upstream could replace the function of REbeta; Required for phytochrome regulation
RHERPATEXPA7	3		1	1	5	"Right part of RHEs (Root Hair-specific cis-Elements)" conserved among the Arabidopsis thaliana A7 (AtEXPA7) orthologous (and paralogous) genes from diverse angiosperm species with different hair distribution patterns
ROOTMOTIFTAPOX1		4	8	6	18	Motif found both in promoters of rolD;
RYREPEATBNNAPA	1			1	2	"RY repeat" found in RY/G box (the complex containing the two RY repeats and the G-box) of napA gene in Brassica napus (B.n.); Found between -78 and -50; Required for seed specific expression; See S000262, S000263; dist B ABRE mediated transactivation by ABI3 adn ABI3-dependent response to ABA; a tetramer of the composite RY/G complex mediated only ABA-independent transactivation by ABI3; B2 domain of ABI3 is necessary for ABA-independent and ABA-dependent activation through the dist B ABRE;
RYREPEATLEGUMINBOX	1			1	2	"RY repeat (CATGCAY)" or legumin box found in seed-storage protein genes in legume such as soybean (G.m.)
S1FBOXSORPS1L21			2		2	"S1F box" conserved both in spinach (S.o.) RPS1 and RPL21 genes encoding the plastid ribosomal protein S1 and L21, respectively; Negative element; Might play a role in

						downregulating RPS1 and RPL21 promoter activity
S2FSORPL21	1				1	"S2F binding site" ("S2 site") in spinach RPL21 gene encoding the plastid ribosomal protein L21; S2 site (CATACAWW) is conserved in promoter region of many nuclear genes encoding plastid proteins; Leaf-specific, light-independent regulatory element; S2 site is related to but different from the light-responsive GT-1 binding site (Lagrange et al., 1997); For a compilation of related GT elements and factors,
SEF1MOTIF			2		2	"SEF1 (soybean embryo factor 1)" binding motif; sequence found in 5'-upstream region (-640; -765) of soybean beta-conglicinin (7S globulin) gene
SEF3MOTIFGM			1		1	"SEF3 binding site"; Soybean (G.m.) consensus sequence found in the 5' upstream region of beta-conglycinin (7S globulin) gene; AACCCA(-27bp-)AACCCA; SEF=soybean embryo factor; SEF2; SEF3; SEF4;
SEF4MOTIFGM7S	1	2	2	5	10	"SEF4 binding site"; Soybean (G.m.) consensus sequence found in 5'upstream region (-199) of beta-conglycinin (7S globulin) gene (Gmg17.1); "Binding with SEF4 (soybean embryo factor 4)";
SITEIIATCYTC	1	2	2		5	"Site II element" found in the promoter regions of cytochrome genes (Cytc-1, Cytc-2) in Arabidopsis; Located between -147 and -156 from the translational starts sites (Welchen et al., 2005); Y=C/T; See also S000308; Overrepresented in the promoters of nuclear genes encoding components of the oxidative phosphorylation (OxPhos) machinery from both Arabidopsis and rice (Welchen and Gonzalez, 2006);)
Skn-1_MOTIF		1	2	1	4	cis-acting regulatory element required for endosperm expression
SORLIP1AT	13	7			20	one of "Sequences Over-Represented in Light-Induced Promoters (SORLIPs) in Arabidopsis; Computationally identified phyA-induced motifs; SORLIP 1 is most over-represented, and most statistically singnificant; See also S000483, S000484, S000485, S000486 (all SORLIPs), and also S000487, S000488, S000489, S000490 (all SORLREPs); Over-represented in light-induced cotyledon and root common genes and root-specific genes
SORLIP5AT				1	1	one of "Sequences Over-Represented in Light-Induced Promoters (SORLIPs) in Arabidopsis; Computationally identified phyA-induced motifs; See also S000482, S000483, S000484, S000485 (all SORLIPs), and also S000487, S000488, S000489, S000490 (all SORLREPs); Over-represented in both light-induced cotyledon-specific and root-specific genes
SORLREP3AT	1				1	one of "Sequences Over-Represented in Light-Repressed Promoters (SORLREPs) in Arabidopsis; Computationally identified phyA-repressed motifs
Sp1		2			2	Light responsive element
SREATMSD		1			1	"sugar-repressive element (SRE)" found in 272 of the 1592 down-regulated genes after main stem decapitation in Arabidopsis;
SURECOREATSULTR11	4	1	2		7	Core of sulfur-responsive element (SURE) found in the promoter of SULTR1;1 high-affinity sulfate transporter gene in Arabidopsis; SURE contains auxin response

						factor (ARF) binding sequence (GAGACA)(see S000270 ARF:TGTCTC; its complementary seq is GAGACA), and this core sequence is a part of it; this core seq is involved in -S response; Beware of other SURE (sucrose responsive element)
SV40COREENHAN		2			2	"SV40 core enhancer"; Similar sequences found in rbcS genes
T/GBOXATPIN2		1	1		2	
TAAAGSTKST1	1		5	3	9	TAAAG motif found in promoter of *Solanum tuberosum* KST1 gene; target site for trans-acting StDof1 protein controlling guard cell-specific gene expression; KST1 gene encodes a K+ influx channel of guard cells.
TATABOX	5	7	34	24	70	Core promoter element
TATABOX2			4	1	5	"TATA box"; TATA box found in the 5'upstream region of pea legA gene; sporamin A of sweet potato; TATA box found in beta-phaseolin promoter (Grace et al.); sequence and spacing of TATA box elements are critical for accurate initiation
TATABOX5	1	2	2	5	10	"TATA box"; TATA box found in the 5'upstream region of pea (Pisum sativum) glutamine synthetase gene; a functional TATA element by in vivo analysis
TATABOXOSPAL			1	1	2	Binding site for OsTBP2, found in the promoter of rice pal gene encoding phenylalanine ammonia-lyase; OsTFIIB stimulated the DNA binding and bending activities of OsTBP2 and synergistically enhanced OsTBP2-mediated transcription from the pal promoter
TATCCAOSAMY		1	1		2	"TATCCA" element found in alpha-amylase promoters of rice (O.s.) at positions ca.90 to 150bp upstream of the transcription start sites; Binding sites of OsMYBS1, OsMYBS2 and OsMYBS3 which mediate sugar and hormone regulation of alpha-amylase gene expression
TATCCAYMOTIFOSRAMY3D			1		1	"TATCCAY motif" found in rice (O.s.) RAmy3D alpha-amylase gene promoter; Y=T/C; a GATA motif as its antisense sequence; TATCCAY motif and G motif (see S000130) are responsible for sugar repression
TBOXATGAPB	1		2	2	5	"Tbox" found in the Arabidopsis thaliana (A.T.) GAPB gene promoter; Located between -94 and -89 (T1) and also between -84 and -79 (T2); Mutations in the "Tbox" resulted in reductions of light-activated gene transcription; GAPB encodes the B subunit of chloroplast glyceraldehyde-3-phosphate dehydrogenase(GADPH) of A.T
TC-RICH REPEATS	2		1		3	Involved in defense and stress responsiveness
TCA-ELEMENT	1			1	2	cis-acting element involved in salicylic acid responsiveness
TCCC-MOTIF	1				1	Part of light responsive element
TGACGMOTIF			1	1	2	cis-acting regulatory element involved in the MeJA-responsiveness
TGACGTVMAMY			1		1	"TGACGT motif" found in the Vigna mungo (V.m.) alpha-Amylase (Amy) gene promoter; Located between -128 and -123; Required for high level expression of alpha-Amylase in the cotyledons of the germinated seeds
TRANSINITDICOTS				1	1	Context sequence of translational initiation codon in dicots
TRANSINITMON	1			1	2	Context sequence of translational initiation

OCOTS						codon in monocots
UP1ATMSD	1				1	"Up1" motif found in 162 of the 1184 up-regulated genes after main stem decapitation in Arabidopsis;
UPRMOTIFIIAT	2	1			3	"Motif II" in the conserved UPR (unfolded protein response) cis-acting element in Arabidopsis genes coding for SAR1B, HSP-90, SBR-like, Ca-ATPase 4, CNX1, PDI, etc.;
WBBOXPCWRKY1			2		2	"WB box"; WRKY proteins bind specifically to the DNA sequence motif (T)(T)TGAC(C/T), which is known as the W box; Found in amylase gene in sweet potato, alpha-Amy2 genes in wheat, barley, and wild oat, PR1 gene in parsley, and a transcription factor gene in Arabidopsis
WBOXATNPR1	1		5	2	8	"W-box" found in promoter of Arabidopsis thaliana (A.t.) NPR1 gene; Located between +70 and +79 in tandem; They were recognized specifically by salicylic acid (SA)-induced WRKY DNA binding proteins; See S000142 (SQ=TTGACC); See S000310 (SQ=TTTGACY); A cluster of WRKY binding sites act as negative regulatory elements for the inducible expression of AtWRKY18
WBOXHVISO1	1	1	5	2	9	SUSIBA2 bind to W-box element in barley iso1 (encoding isoamylase1) promoter
WBOXNTERF3	3	2	7	2	14	"W box" found in the promoter region of a transcriptional repressor ERF3 gene in tobacco; May be involved in activation of ERF3 gene by wounding
WRKY71OS	7	3	11	5	26	"A core of TGAC-containing W-box" of, e.g., Amy32b promoter; Binding site of rice WRKY71, a transcriptional repressor of the gibberellin signaling pathway; Parsley WRKY proteins bind specifically to TGAC-containing W box elements within the Pathogenesis-Related Class10 (PR-10) genes
WUSATAg		1			1	Target sequence of WUS in the intron of AGAMOUS gene in Arabidopsis
XYLAT				1	1	cis-element identified among the promoters of the "core xylem gene set";

APPENDIX B

PEP PMSF Inhibition Assay (Inhibits Serine proteases and some cysteine proteases):

Preparation of stock solution of 100 mM PMSF in Ethanol:

Dissolve 174.19 mg/mmol X 100 mmole/L X 0.005L = 87.1 mg in 5 mL of Ethanol

Preparation of 20 mM PMSF working solution in Ethanol:

Dilute by adding 20 µl of 100 mM stock solution to 80 µl ethanol

					Ethanol Control	**Positive Control**
Enzyme	145 µl	145 µl	145 µl	145 µl	145 µl	145 µl
PMSF (20 mM working solution)	8 µl (gives a final concentration of 1.04 mM)	6 µl (gives a final concentration of 0.8 mM)	4 µl (gives a final concentration of 0.5 mM)	2 µl (gives a final concentration of 0.27 mM)	0 µl (gives a final concentration of 0 mM)	0 µl (gives a final concentration of 0 mM)
Ethanol	2 µl	4 µl	6 µl	8 µl	10 µl	0 µl
Water	0 µl	0 µl	0 µl	0 µl	0 µl	10 µl
Incubate for 1 hour at assay temperature						
Universal Buffer, pH 7.6	140 µl	140 µl	140 µl	140 µl	140 µl	140 µl
Substrate (0.01-6 mM)	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl
Total Volume	300 µl	300 µl	300 µl	300 µl	300 µl	300 µl
Incubate for 8-12 hours at assay temperature						

PEP E64 Inhibition Assay (Inhibits only Cysteine proteases):

Preparation of stock solution of 1 mM E64 in water:

Dissolve 357.41 mg/mmol X 1 mmole/L X 0.01L = 3.57 mg in 10 mL of Water

Preparation of 200 µM E64 working solution in Water**:**

Dilute by adding 20 µl of 1 mM stock solution to 80 µl water

					Positive Control
Enzyme	145 µl	145 µl	145 µl	145 µl	145 µl
E64 (200 µM working solution)	8 µl (gives a final concentration of 10.4 µM)	6 µl (gives a final concentration of 7.9 µM)	4 µl (gives a final concentration of 5.4 µM)	2 µl (gives a final concentration of 2.7 µM)	0 µl (gives a final concentration of 0 µM)
Water	2 µl	4 µl	6 µl	8 µl	10 µl
Incubate for 1 hour at assay temperature					
Universal Buffer, pH 7.6	140 µl	140 µl	140 µl	140 µl	140 µl
Substrate (0.01-6 mM)	5 µl	5 µl	5 µl	5 µl	5 µl
Total Volume	300 µl	300 µl	300 µl	300 µl	300 µl
Incubate for 8-12 hours at assay temperature					

PEP Pepstatin Inhibition Assay (Inhibits Aspartic proteases only):

Preparation of stock solution of 1 mM Pepstatin in Methanol:Acetic Acid (9:1):

Dissolve 685.89 mg/mmol X 1 mmole/L X 0.01L = 6.86 mg in 9 mL of Methanol + 1 mL Acetic acid

Preparation of 200 µM Pepstatin working solution in Methanol:Acetic Acid (9:1):

Dilute by adding 20 µl of 1 mM stock solution to 80 µl Methanol:Acetic Acid (9:1)

					Positive Control	**Methanol:Acetic Control**
Enzyme	145 µl	145 µl	145 µl	145 µl	145 µl	145 µl
Pepstatin (200 µM working solution)	8 µl (gives a final concentration of 10.4 µM)	6 µl (gives a final concentration of 7.9 µM)	4 µl (gives a final concentration of 5.4 µM)	2 µl (gives a final concentration of 2.7 µM)	0 µl (gives a final concentration of 0 µM)	0 µl (gives a final concentration of 0 µM)
Water					10 µl	
Methanol: Acetic Acid (9:1)	2 µl	4 µl	6 µl	8 µl		10 µl
Incubate for 1 hour at assay temperature						
Universal Buffer, pH 7.6	140 µl	140 µl	140 µl	140 µl	140 µl	140 µl
Substrate (0.01-6 mM)	5 µl	5 µl	5 µl	5 µl	5 µl	5 µl
Total Volume	300 µl	300 µl	300 µl	300 µl	300 µl	300 µl
Incubate for 8-12 hours at assay temperature						

PEP Leupeptin Inhibition Assay (Inhibits Serine and Cysteine proteases):

Preparation of stock solution of 10 mM in water:

Dissolve 475.59 mg/mmol X 10 mmole/L X 0.005L = 23.77 mg in 5 mL of Water

Preparation of 2 mM Leupeptin working solution in Water:

Dilute by adding 20 µl of 10 mM stock solution to 80 µl water

					Positive Control
Enzyme	145 µl	145 µl	145 µl	145 µl	145 µl
Leupeptin (2 mM working solution)	8 µl (gives a final concentration of 104 µM)	6 µl (gives a final concentration of 79 µM)	4 µl (gives a final concentration of 54 µM)	2 µl (gives a final concentration of 27 µM)	0 µl (gives a final concentration of 0 µM)
Water	2 µl	4 µl	6 µl	8 µl	10 µl
Incubate for 1 hour at assay temperature					
Universal Buffer, pH 7.6	140 µl	140 µl	140 µl	140 µl	140 µl
Substrate (0.01-6 mM)	5 µl	5 µl	5 µl	5 µl	5 µl
Total Volume	300 µl	300 µl	300 µl	300 µl	300 µl
Incubate for 8-12 hours at assay temperature					

BCA Standard curve for protein determination