SUBSTRATE SPECIFICITY AND STRUCTURAL INVESTIGATION INTO PEPO AND PEPW: TWO PEPTIDASES FROM *LACTOBACILLUS RHAMNOSUS*

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science
in
Biochemistry

at Massey University, Palmerston North
New Zealand.

Karen Maree Yates
2005

ABSTRACT

The proteolytic systems of lactic acid bacteria have important roles in the maturation and flavour development of cheese. Lactic acid bacteria pepetidases contribute to the taste of cheese through the production of low-molecular weight peptides and free amino acids. Although some lactic acid bacteria peptidases have been structurally and enzymatically characterised for their substrate specificity, there are some that are yet to be completely biochemically characterised. The aim of the present study was to investigate the substrate specificity and three-dimensional structure of two peptidases that could potentially be used as a tool to modify and control cheese bitterness and possibly other flavour attributes from *Lactobacillus rhamnosus*, PepO and PepW.

The *pepW* gene was successfully cloned from *L. rhamnosus* into an *E. coli* expression system. Recombinant PepW was purified to homogeneity and was shown to exist as a hexamer of 50 kDa subunits. Recombinant PepO was expressed from a previously established *L. lactis* expression system and purified to homogeneity. PepO was shown to exist as a 70 kDa monomer, and function as a metallopeptidase.

PepO and PepW were shown to selectively hydrolyse chymosin-derived bovine β- and κ-casein peptides, and casein peptides extracted from Cheddar cheese. One conclusive PepO cleavage site that had not been previously characterised was identified. This was the β-casein peptide bond between Leu_6-Asn_7. Several possible PepO and PepW cleavage sites in α_{s1}-, β- and κ-casein were identified, suggesting that PepO has a broad endopeptidase activity, whilst PepW has a specific exopeptidase activity.

PepO and PepW crystals were successfully grown for structure determination by X-ray crystallography. Native data sets were collected for both PepO and PepW, and derivative data were collected for PepO. Structure determination was attempted using Multiple Isomorphous Replacement and Molecular Replacement techniques.

Results from the substrate specificity and structural investigation of the *L. rhamnosus* peptidases, PepO and PepW, are presented in this thesis.

ACKNOWLEDGMENTS

I would like to acknowledge the following people for their assistance throughout the course of this project:

My supervisors Dr Gill Norris and Dr Julian Reid for their advise, guidance and encouragement.

Professor Geoffrey Jameson for his help collecting and processing x-ray diffraction data, and explanation of crystal twinning. Dr Brian Anderson for his help in processing x-ray diffraction data.

The Centre for Molecular Biodiscovery for setting up the sitting drop crystallisation screens. Dr Julian Reid and Mr Aveenash Bala for carrying out the liquid-chromatography tandem mass spectrometry experiments. Trevor Loo for his help with Edman N-terminal sequencing and Dr Rosemary Brown for supplying the pET32a-Trx-His-rTEV vector.

Members of the X-lab who made working on this project enjoyable (most of the time).

Finally, thank you to my family and friends who have always encouraged me to pursuit my interests and helped me overcome my obstacles.

This work was carried out in collaboration with the Fonterra Co-operative Group Limited, and supported by the Foundation for Research Science and Technology, New Zealand.

TABLE OF CONTENTS

LIST OF FIGURES

PROTEOLYTIC ENZYMES: AN INTRODUCTION

PART 1:
SUBSTRATE SPECIFICITY OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

PART 2:

THREE-DIMENSIONAL STRUCTURE DETERMINATION OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

PROTEOLYTIC ENZYMES: AN INTRODUCTION

PART 1:
SUBSTRATE SPECIFICITY OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

CHAPTER 3

CHAPTER 4

CHAPTER 5

PART 2:
THREE-DIMENSIONAL STRUCTURE DETERMINATION OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

CHAPTER 2

CHAPTER 3

CHAPTER 4

ABBREVIATIONS

GENERAL ABBREVIATIONS

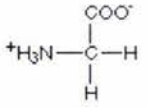
AA	Amino Acid
AC	Affinity Chromatography
ACE	Angiotensin Converting Enzyme
AIDS	Acquired Immune Disease Syndrome
AMoRe	Automated Molecular Replacement
AMP	Ampicillin
AMU	Atomic Mass Units
API	Atmospheric Pressure Ionisation
AS	Anomalous Scattering
ATP	Adenosine Tri-Phosphate
BH	Bleomycin Hydrolase
BLAST	Basic Local Alignment Search Tool
BTP	Bis-Tris Propane
CAD	Collect Assorted Data
CAPS	Cyclohexyl Amino Propane Sulfonic acid
CCP4	Collaborative Computational Project number 4
CCP4i	CCP4 graphical user Interface
CHES	2-Cyclo-Hexylamino Ethane Sulfonic acid
CID	Collision-Induced Dissociation
Da	Dalton
DNA	Deoxy-ribose Nucleic Acid
dNTP	Deoxyribo-Nucleotide Tri-Phosphate
Dpt	Di- and Tri-Peptide Transport System
DTT	Di-Thio-Threitol
EC	Enzyme Commission
EDTA	Ethylene-Di-amine-Tetra-Acetate
EPPS	4-(2-hydroxy-Ethyl)-1-Piperazine Propane Sulfonic acid
ESI	Electro-Spray Ionisation
FITC	Fluorescein Iso-Thio-Cyanate
FPLC	Fast Protein Liquid Chromatography
HCCA	α-Cyano-4-Hydroxy-trans-Cinnamic Acid
HEPES	2-(4-(2-Hydroxy-Ethyl)-1-Piperazinyl) Ethane Sulfonic acid
HIC	Hydrophobic Interaction Chromatography
HIV	Human Immunodeficiency Virus
HPLC	High Performance Liquid Chromatography
IEX	Ion Exchange Chromatography

IMAC	Immobilised Metal Affinity Chromatography
IPTG	Iso-Propyl-beta-*D*-Thio-Galactopyranoside
I/σ	Signal-to-noise ratio
IUBMB	International Union of Biochemistry and Molecular Biology
LAB	Lactic Acid Bacteria
LB	Luria Bertani Media
LC	Liquid Chromatography
MALDI	Matrix Assisted Laser Desorption Ionisation
MCP	Micro-Channel Plate
MCS	Multiple Cloning Site
MES	2-(*N*-Morpholino) Ethane Sulfonic acid
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
MOPS	3-(*N*-MOrpholino) Propane Sulfonic acid
MPD	2-Methyl-2,3-Pentane-Diol
MR	Molecular Replacement
mRNA	Messenger Ribo-Nucleic Acid
MRS	Man, Rogosa and Sharpe Media
MS	Mass Spectrometry
MSMS	Tandem Mass Spectrometry
Mw	Molecular Weight
M/z	Mass-to-charge ratio
NEP	Neutral Endo-Peptidase
NCBI	National Centre for Biotechnology Information
NIZO	Netherlands Institute for Dairy Research
OD	Optical Density
Opp	Oligopeptide Transport System
ORF	Open Reading Frame
PAGE	Poly-Acrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Poly-Ethylene Glycol
Pep	Peptidase
pI	Isoelectric Point
PIPES	PIPerazine-1,4-bis 2-Ethane Sulfonic acid
PITC	Phenyl-Iso-Thio-Cyanate
PMSF	Phenyl-Methane Sulfonyl Fluoride
PPG	Poly-Propylene Glycol
Prt	Proteinase
PSI	Pound-force per Square Inch

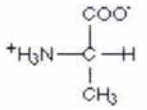
PTH	Phenyl-Thio-Hydantoin
PVDF	Poly Vinyli-Dene Fluoride
RE	Restriction Endonuclease
RP	Reverse Phase
RNA	Ribo-Nucleic Acid
RNase	Ribo-Nuclease
RSM	Reconstituted Skim Milk
SDS	Sodium Dodecyl-Sulphate
SEC	Size Exclusion Chromatography
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
TCA	Tri -Chloro-Acetic acid
TEV	Tobacco Etch Virus
TFA	Tri-Fluoro Acetic acid
TOF	Time Of Flight
TRIS	Tris-(hydroxymethyl)-aminomethane
Trx	Thioredoxin
UV	Ultra Violet

AMINO ACID ABBREVIATIONS

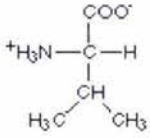
Amino acids with hydrophobic side groups:



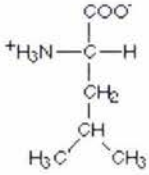
Glycine (Gly, G)



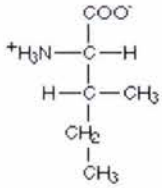
Alanine (Ala, A)



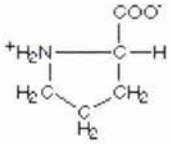
Valine (Val, V)



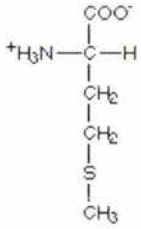
Leucine (Leu, L)



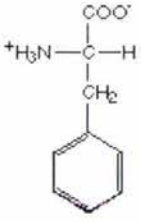
Isoleucine (Ile, I)



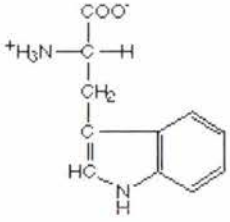
Proline (Pro, P)



Methionine (Met, M)

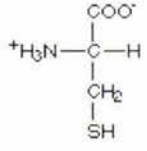


Phenylalanine

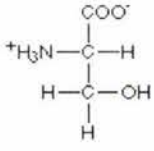


Tryptophan (Trp, W)

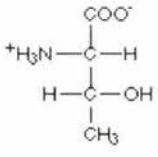
Amino acids with hydrophilic side groups:



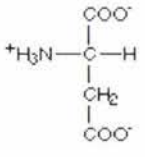
Cysteine (Cys, C)



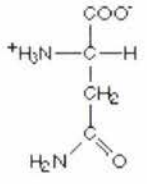
Serine (Ser, S)



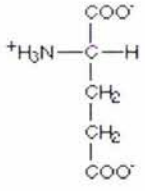
Threonine (Thr, T)



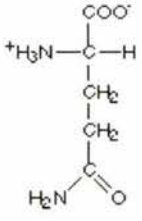
Asparatate (Asp, D)



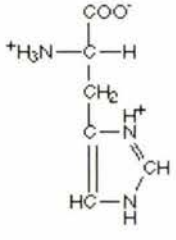
Asparagine (Asn, N)



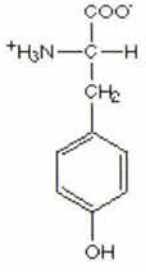
Glutamate (Glu, E)



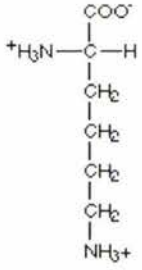
Glutamine (Gln, Q)



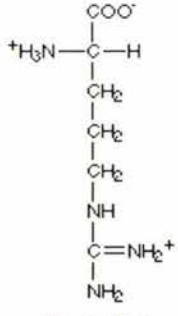
Histidine (His, H)



Tyrosine (Tyr, Y)



Lysine (Lys, K)



Arganine (Arg, R)

INTRODUCTION:

PROTEOLYTIC ENZYMES:
AN INTRODUCTION

PROTEOLYTIC ENZYMES: AN INTRODUCTION

1.1 PROTEOLYTIC ENZYMES

In recent years it has become clear that proteolytic enzymes mediate many important biological processes including: fertilization, cell division and differentiation, controlled cell death, blood coagulation and protein catabolism. Proteolytic enzymes are found in all organisms regardless of kingdom, and the importance of these enzymes is emphasized by the fact that approximately 2% of human genes encode proteolytic enzymes (Saklatvala 2003).

Proteolytic enzymes break peptide bonds linking the amino group of one amino acid with the carboxy group of an adjacent amino acid in a peptide chain (Figure 1.1).

Figure 1.1 Proteolytic Enzymes Catalyse the Hydrolysis of Peptide Bonds.

1.1.1 CLASSIFICATION

Proteolytic enzymes require water to be present in the catalytic site to break the peptide bond and are thus classified as hydrolases (EC 3). All enzymes that catalyse the cleavage of any chemical bonds using water in the mechanism are termed hydrolases. This family of enzymes encompass esterases, lipases, peptidases and proteases. While the systematic name of a hydrolase enzyme always includes the term 'hydrolase', the common name is formed by the name of the substrate with the suffix -ase. The term "peptidase" is synonymous with "peptide hydrolase" and has been recommended by the International Union of Biochemistry and Molecular Biology (IUBMB) as the general term for all enzymes that hydrolyse peptide bonds (NC-IUBMB 1992).

Peptidases (EC 3.4) are primarily classified on the basis of their action as either exopeptidases or endopeptidases. Exopeptidases (EC 3.4.11-19) cleave peptide bonds near the amino terminus or the carboxy terminus of a peptide substrate, while endopeptidases (EC 3.4.21-99) hydrolyze peptide bonds distant from the amino and carboxy termini of the peptide substrate (NC-IUBMB 1992; Barrett 2004) (Table 1.1).

Table 1.1 Peptidase Classification According to Enzyme Commission and Reaction Catalysed.

Peptidase	EC	Reaction Catalysed [a]
Exopeptidases	3.4.11-19	
Aminopeptidases	3.4.11	X↓X(n)
Dipeptidases	3.4.13	X↓X
Dipeptidyl-peptidases, tripeptidyl-peptidases	3.4.14	X-X(X)↓X(n)
Peptidyl-dipeptidases	3.4.15	X(n)↓X-X
Serine-type carboxypeptidases	3.4.16	X(n)↓X
Metallocarboxypeptidases	3.4.17	X(n)↓X
Cysteine-type carboxypeptidases	3.4.18	X(n)↓X
Omega peptidases	3.4.19	Y↓X(n)
Endopeptidases	3.4.21-99	
Serine endopeptidases	3.4.21	X(n)↓X(n')
Cysteine endopeptidases	3.4.22	X(n)↓X(n')
Aspartic endopeptidases	3.4.23	X(n)↓X(n')
Metalloendopeptidases	3.4.24	X(n)↓X(n')
Threonine endopeptidases	3.4.25	X(n)↓X(n')
Endopeptidases of unknown catalytic mechanism	3.4.99	X(n)↓X(n')

[a] Reaction Catalysed: ↓, cleavage site; n, number of residues; X, amino acid; Y, amino acid that is cyclized, or linked by an isopeptide bond.

1.1.2 EXOPEPTIDASES

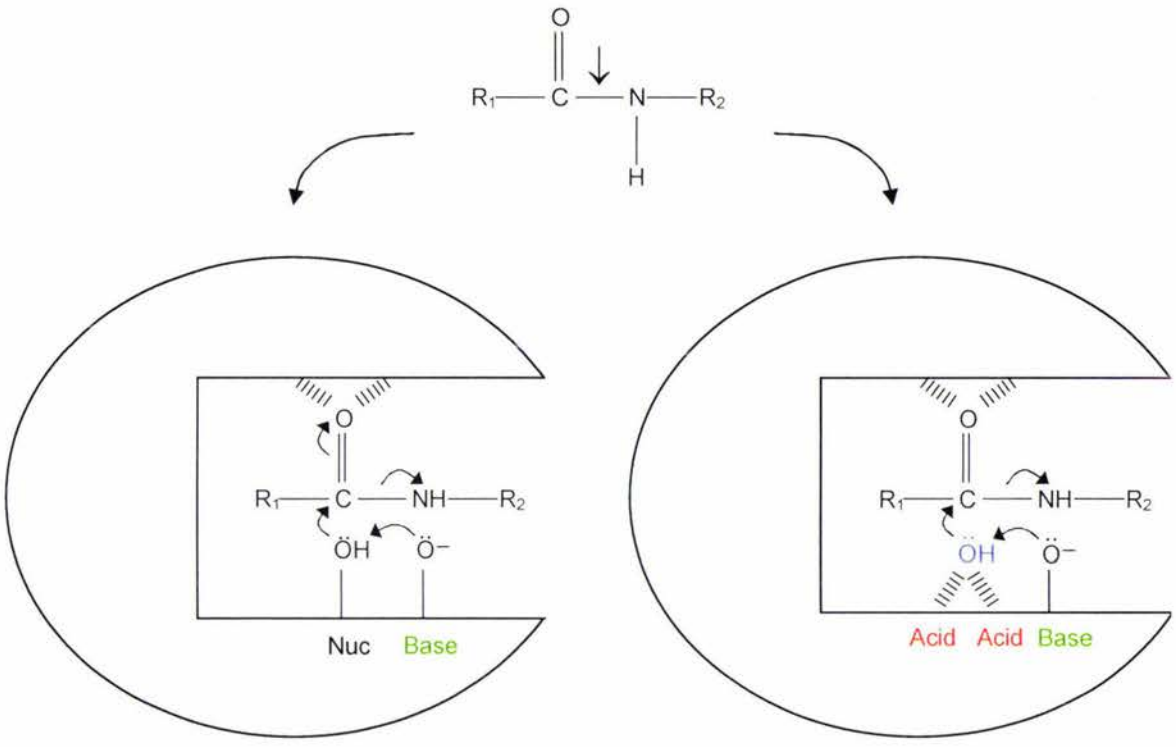
The exopeptidases can be further classified on the basis of the reaction catalysed. Exopeptidases that act at the amino-terminus of a peptide substrate can either release a single amino acid residue (aminopeptidases EC 3.4.11) or a dipeptide or a tripeptide (dipeptidyl-peptidases and tripeptidyl-peptidases EC 3.4.14). Those that act at the carboxy-terminus can also release either a single residue (carboxypeptidases EC 3.4.16-18) or a dipeptide (peptidyl-dipeptidases EC 3.4.15). Some exopeptidases are specific for certain dipeptides (dipeptidases EC 3.4.13) and some remove terminal residues that are cyclized, or linked by isopeptide bonds (peptide linkages other than those of α-amino to α-carboxyl groups) e.g. omega peptidases (EC 3.4.19) (Barrett 1992; NC-IUBMB 1992; Barrett 2004).

1.1.3 ENDOPEPTIDASES

Due to difficulties defining the substrate specificity of endopeptidases, an alternative classification system has been developed. In 1960 Hartley described four distinct types of catalytic mechanism used by peptidases; serine, cysteine, aspartic and metallo, and this observation developed into a practical system for peptidase classification (Hartley 1960; Barrett 1992). Peptidases are now classified under a hierarchical system built on the concepts of catalytic type, clan, family and peptidase (Barrett 2004; Rawlings 2004).

1.1.4 CATALYTIC TYPE

The concept of catalytic type of a peptidase depends upon the chemical nature of the groups responsible for catalysis. The currently recognized catalytic types are; Serine, Cysteine, Threonine, Aspartic, Metallo, Glutamic and some as yet unclassified. These catalytic types partition essentially into two catalytic mechanisms: covalent and non-covalent (general acid-base) catalysis (Salvesen 2003) (Figure 1.2).



1. Covalent Catalysis
Serine, Threonine, Cysteine

2. Non-covalent Catalysis
Aspartic, Glutamic, Metallo

Figure 1.2 The Covalent and Non-covalent Catalytic Mechanism of Peptidases. (Nuc, nucleophile; Acid, aspartic/glutamic acid; Base, usually histidine; OH, water molecule; ↓, cleavage site; ııı, hydrogen bond; ___ , covalent bond).

The covalent catalytic mechanism of the serine, cysteine and threonine peptidases involves the nucleophilic attack of the scissile peptide bond of the substrate, by a nucleophilic side chain of the enzyme generating a covalent enzyme-substrate intermediate. The nucleophile that attacks the carbonyl carbon of the scissile peptide bond may be the sulfhydryl group of a cysteine (cysteine peptidases EC 3.4.22), or the hydroxyl group of a serine (serine peptidases EC 3.4.21) or threonine residue (threonine peptidases EC 3.4.25), with the nitrogen of the imidazole of histidine often functioning as a base (Barrett 2004).

The catalytic mechanism of the aspartic, glutamic and metallopeptidases involves a non-covalent intermediate and differs from the serine, cysteine and threonine peptidases in that the nucleophile that attacks the scissile peptide bond is an activated water molecule rather than the nucleophilic side chain of an amino acid (Barrett 2004).

The aspartic peptidases (EC 3.4.23) and recently discovered glutamic peptidases (also EC 3.4.23) have either an aspartic acid or a glutamic acid residue at the catalytic site, that activates the water molecule, while the metallopeptidases (EC 3.4.24) contain a divalent cation, usually zinc, at the active site centre which in conjunction with a glutamic acid residue activates the water molecule. The imidazole ring of histidine often functions as a base for these peptidases as it does for the serine, cysteine and threonine peptidases (Barrett 2004; Rawlings 2004).

The catalytic mechanism of a peptidase can often be predicted by amino acid sequence analysis; for example metallopeptidases often contain a conserved HEXXH pentapeptide that has been shown in crystallographic studies to coordinate a metal ion at the active centre (Barrett 1992; Barrett 2004). However, the catalytic mechanism can only be reliably determined by establishing the susceptibility of the peptidase to specific protease inhibitors.

1.1.5 CLAN, FAMILY AND PEPTIDASE

Peptidases are further classified under the concepts of clan, family and peptidase. Clans and families are groups of homologous peptidases; a clan contains one or more families that appear to have come from a common origin. The clearest kind of evidence for homology at the clan level is three-dimensional structural similarity, but amino acid sequence similarity can also be revealing. The name of a clan is formed from the initial letter of the catalytic type (S, serine; C, cysteine; T, threonine; A, aspartic; M, metallo; G, glutamic; U, unclassified) followed by a serial second capital letter. Each peptidase family is named with a letter denoting the catalytic type followed by a sequentially assigned number (Barrett 2004; Rawlings 2004).

A peptidase family is a group of peptidases, all of which display a particular kind of peptidase activity, and are closely related in amino acid sequence (Barrett 2004). Each member of the group is assigned a unique identifier constructed of two parts, each part being separated by a decimal point: these are the family name and an arbitrary three-digit serial number (Barrett 2004; Rawlings 2004).

Due to the rapid expansion of research on proteolytic enzymes an updatable reference resource containing summary data on all peptidases has been established. The MEROPS peptidase database (www.merops.ac.uk) was first published on the World Wide Web in 1996 and currently provides information on 177 peptidase families, 49 clans and 2018 unique peptidases (Rawlings 2004) (Table 1.2).

Table 1.2 Statistics for the MEROPS Database.

Catalytic Type	Total	Family	Clan	Peptidase	PDB Entry
Serine	9484	39	10	773	145
Threonine	977	4	1	57	37
Cysteine	4771	58	7	459	75
Aspartic	1501	12	6	159	36
Metallo	7581	53	15	552	84
Glutamic	10	1	1	4	1
Unknown	575	10	9	14	1
Total	24899	177	49	2018	379

1.1.6 NOMENCLATURE FOR PEPTIDASE SUBSTRATE SPECIFICITY

The system used to describe peptidase substrate specificity was introduced in 1970 by Berger and Schechter, who proposed that the peptidase catalytic site is flanked by specificity subsites, each able to accommodate the side chain of a single amino acid residue. These subsites are called S (for subsites) and are numbered from the catalytic site, S1 ... Sn towards the amino-terminus of the substrate, and S1′ ... Sn′ towards the carboxy-terminus. The peptide substrate residues that they accommodate are called P (for peptide) and are numbered P1 ... Pn, and P1′ ... Pn′, respectively (Berger 1970; Barrett 1992; Barrett 2004) (Figure 1.3).

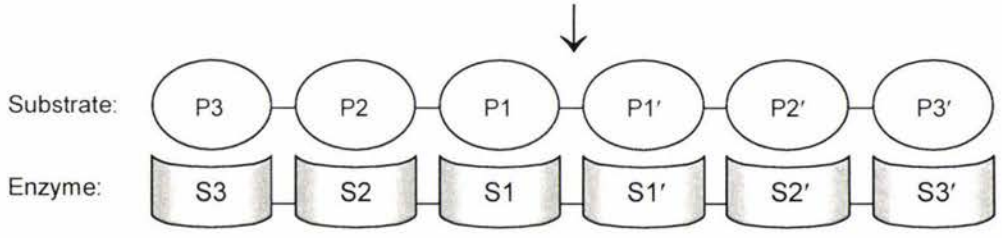


Figure 1.3 Nomenclature for Peptidase Substrate Specificity. (S, subsites; P, peptide residue; ↓, cleavage site)

1.2 PROTEOLYTIC ENZYMES IN BIOLOGY

Research into the proteolytic enzymes is rich and diverse. One of the earliest and best understood paradigms of proteolytic activity is zymogen activation. Many enzymes are synthesised as inactive precursors (zymogens) that are subsequently activated by cleavage of one or a few specific peptide bonds. Zymogen activation is responsible for a great variety of

physiological regulation, including protein digestion, and is particularly important when coupled to cascade activation as has been well documented for blood coagulation, apoptosis and the complement system (Neurath 1999).

Unlike the phosphorylation/dephosphorylation cascades, which are reversible, proteolytic cascades are irreversible, and therefore must be tightly regulated to prevent the serious deleterious effects of uncontrolled proteolytic activity (Amour 2004).

Deregulation of peptidase activity can have pathological effects and peptidases often have an important role in the development and manifestation of disease (Saklatvala 2003). The focus of current peptidase research is the search for inhibitors to a number of peptidases implicated in specific diseases. For example HIV peptidase inhibitors, and angiotensin converting enzyme (ACE) inhibitors have both been shown to be very effective drugs for the treatment of AIDS and cardiovascular disease, respectively (Pietzsch 2003).

By virtue of their ability to introduce restricted cleavages into a polypeptide substrate, peptidases are also valuable biochemical tools and have received a great deal of attention. In this context proteases and peptidases are also important in a number of industries particularly the dairy industry.

1.3 LACTIC ACID BACTERIA

Lactic acid bacteria (LAB) are gram-positive bacteria that are used in the dairy industry for the production of a variety of fermented milk products, including cheese and yoghurt. The principal role of LAB in the production of cheese and yoghurt is to ferment the lactose in milk to lactic acid, reducing the pH to approximately 4.5-5. This low pH prevents the growth of undesirable micro-organisms and contributes to casein precipitation, the initial step in cheese production (Reid 1994b).

It has been well established that many LAB, isolated from milk products, are multiple amino acid auxotrophs and therefore must obtain essential amino acids from the growth medium (Kunji 1996). The quantities of free amino acids and small peptides present in milk are not sufficient to support the growth of LAB to a high cell density. Subsequently, these bacteria have developed complex proteolytic systems to obtain essential amino acids from caseins, the primary proteins in milk.

The proteolytic systems of LAB have received an enormous amount of attention due to their importance in both the physiology of the bacteria themselves, and because of their involvement in the maturation and flavour development of cheese.

Proteolysis contributes to the taste of cheese through the production of low-molecular weight peptides and free amino acids. A number of free amino acids, mainly those with nonpolar or hydrophobic side chains, exhibit a bitter flavour (e.g. isoleucine, tryptophan, phenylalanine), as do many small peptides which exhibit flavours dependent on their hydrophobicity and amino acid composition (Lemieux 1992). Certain sequences in the caseins, including α_{s1}-casein(1-9) and β-casein(193-209), are particularly hydrophobic and, when excised by proteolytic enzymes result in bitter flavours in dairy products which are undesirable (Sousa 2001; Yvon 2001).

Although the direct contribution of peptides and amino acids to flavour is probably limited to a background flavour, they are precursors of many other flavour compounds. Conversion of amino acids through deamination, transamination, and decarboxylation to various amines, aldehydes, acids, and sulfur-containing compounds results in the development of different flavours (Sousa 2001). Flavour development in cheese is of major economic interest since the final flavour of cheeses partly determines consumer choice and because the development of flavour is a time consuming and an expensive process. An understanding of the proteolytic enzymes involved in LAB proteolysis is necessary to enable controlled synthesis of flavour compounds through judicious choice of LAB organisms. In the long term, the development of genetically modified LAB strains with improved fermentation properties, and substrate specificity will provide the ultimate tool for the control of flavour in various dairy products.

1.3.1 LACTIC ACID BACTERIA PROTEOLYTIC NETWORK

Research on the LAB proteolytic network has focussed to a large extent on the well characterised LAB, *Lactococcus lactis*. Most, if not all, of the components of the *L. lactis* proteolytic network have been identified and the majority of the enzymes have been purified and biochemically characterised. Furthermore the genetics of the corresponding genes have been studied (Stepaniak 2004).

The *L. lactis* proteolytic network is composed of several closely related but catalytically distinct cell-envelope proteinases (PrtP) which are involved in the initial cleavage of casein; cell membrane transport systems (Opp, DtpT, DtpP) that are involved in the uptake of small peptides and amino acids; and intracellular peptidases (Pep) that hydrolyse the casein-derived peptides (Kunji 1996; Christensen 1999) (Figure 1.4).

1.3.2 CELL ENVELOPE PROTEINASE

The *L. lactis* cell-envelope proteinases, PrtPs, are large monomeric serine endopeptidases with molecular weights of approximately 200 kDa. They are associated with the cell envelope through a membrane anchor and a cell wall-spanning region and are sometimes known as lactocepins (lactococcal cell envelope-associated proteinase) (Barrett 2004).

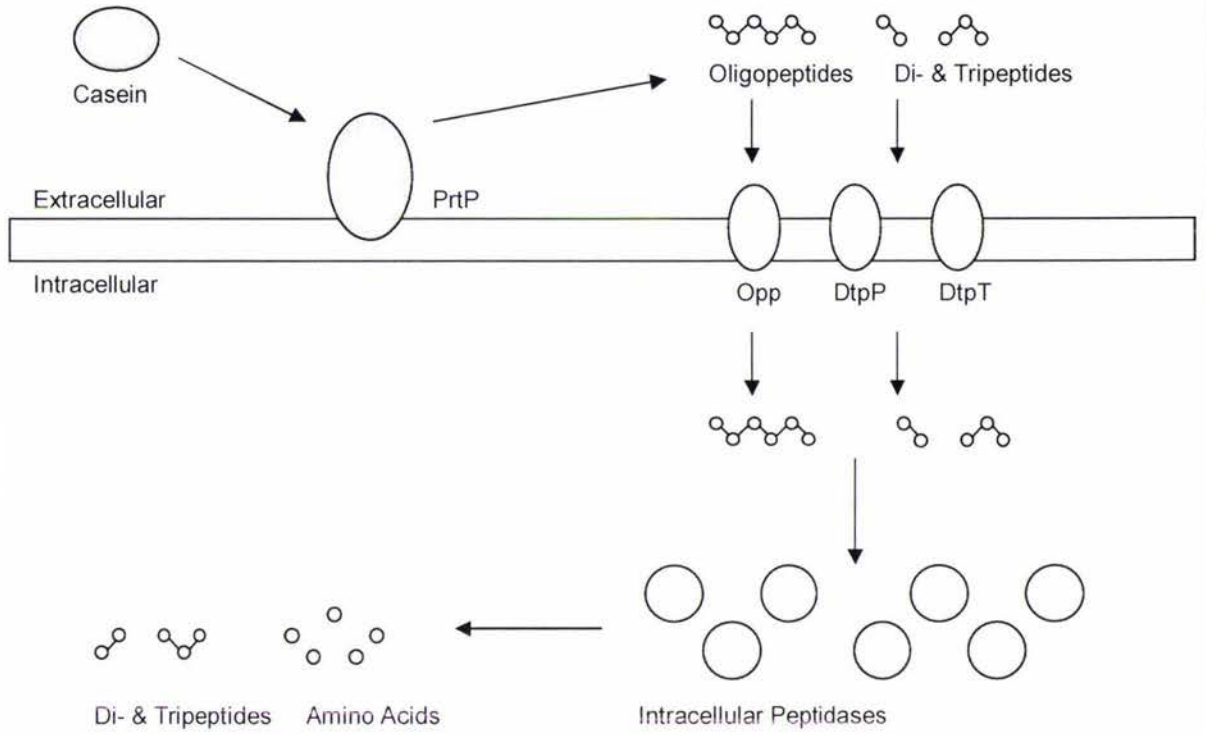


Figure 1.4 The Lactic Acid Bacteria Proteolytic Network. (PrtP, cell-envelope proteinase; Opp, DtpP, DtpT, cell membrane transport systems).

Individual strains of *L. lactis* produce only one form of cell-envelope proteinase, type I, type III or one of intermediate specificity. Both PI- and PIII-type peptidases hydrolyse β- and κ-caseins, although with different cleavage specificities, while α_{s1}-casesin is hydrolysed efficiently only by PIII-type peptidases (Barratt 2004). The enzyme is stabilised by Ca^{2+} (Exterkate 1999), and in the absence of Ca^{2+}, PrtP undergoes autoproteolysis and is released from the cell envelope. Whether associated with the cell envelope or not, PrtP is able to hydrolyse casein into small molecular weight peptides that can then be taken up by the cell via cell membrane transport systems (Barrett 2004).

1.3.3 CELL MEMBRANE TRANSPORT SYSTEMS

L. lactis has separate transport systems for amino acids, di- and tripeptides (DtpT and DtpP), and oligopeptides (Opp) in the cell membrane. The oligopeptide transport system (Opp) mediates the ATP-driven transport of peptides with four or more residues, and plays a central role in the proteolytic network. The di- and tripeptide transport systems DtpT and DtpP are also important components of this network, with DtpP preferentially transporting di- and tripeptides composed of hydrophobic, branched-chain amino acid residues, and DtpT transporting more-hydrophilic peptides (Foucaud 1995; Barrett 2004).

1.3.4 INTRACELLULAR PEPTIDASES

Following transport into the cell, the casein-derived peptides are further hydrolysed by intracellular peptidases. There are at least ten intracellular peptidases in the *L. lactis* proteolytic network, the majority of which are exopeptidases (Barrett 2004; Stepaniak 2004) (Table 1.3).

Table 1.3 *Lactococcus lactis* Intracellular Peptidases.

Peptidase	Molecular Mass [a] (kDa)	Structure [b]	Class [c]	Substrate [d]
Expopetidase				
Aminopeptidase PepA	38 - 45	Hexamer, trimer	M	Asp/Glu↓X(n)
Aminopeptidase PepC	50 - 54	Hexamer	C	X↓X(n)
Aminopeptidase PepN	89 - 95	Monomer	M	X↓X(n)
Dipeptidase PepV	46 - 52	Monomer	M	X↓X
Tripeptidase PepT	23 - 55	Monomer, dimer	M	X↓X-X
Prolidase PepQ	41 - 43	Monomer	M	X↓Pro
Prolinase PepP	43	Monomer	M	X↓Pro-X(n)
X-prolyl dipeptidyl aminopeptidase PepX	82 - 117	Monomer, dimer	S	X-Pro↓X(n)
Endopeptidase				
Endopeptidase PepO	70 - 71	Monomer	M	X(n)↓X(n')
Endopeptidase PepF	70	Monomer	M	X(n)↓X(n')

[a] Molecular Mass: calculated from known amino acid sequence or SDS-PAGE.

[b] Structure: predicted from comparison of gel filtration value with either amino acid sequence or SDS-PAGE values, or x-ray crystal analysis.

[c] Catalytic class: C, cysteine; M, metallo; S, serine peptidases.

[d] Substrate: ↓, cleavage site; X, amino acid.

1.3.4.1 EXOPEPTIDASES

L. lactis exopeptidases include the aminopeptidases (PepA, PepC, PepN, PepT, PepP), the dipeptidases (PepV, PepQ) and the dipeptidyl peptidase (PepX). To date, no carboxypeptidases have been identified in this organism.

A distinguishing feature of PepA is its thermal stability. Independent research groups have shown that PepA from *L. lactis* subsp. *lactis* and subsp. *cremoris* retain greater than 50% activity after incubation at 70°C for 1 hour (Niven 1991; Bacon 1994, Barrett 2004).

In contrast to PepA, the aminopeptidases PepC and PepN have very broad substrate specificities. PepN although it has a very broad substrate specificity, shows a marked preference for substrates containing Arg as the N-terminal residue (Tan 1991; Barrett 2004).

PepC, a cysteine aminopeptidase that is inhibited by SH-group reagents such as iodoacetamide, is active only in a reducing environment (Barrett 2004). While exhibiting a broad specificity, it cannot hydrolyse peptide bonds in which a proline residue is involved. The crystallographic structure of *L. lactis* PepC has recently been solved (Barrett 2004; Mistou unpublished results), revealing a hexameric ring structure which has a prominent central channel large enough to accommodate a peptide. The active site of each monomer has a papain-like fold and is situated within the central channel (Barrett 2004). Insertion of the C-terminus of each PepC monomer into its own active site leaves sufficient space for only one substrate residue, the P1 residue of the substrate, to enter the active site, and appears to be the underlying molecular basis for the strict aminopeptidase activity of *L. lactis* PepC (Mata 1999; Barrett 2004). Site directed mutagenesis, in which the last four C-terminal residues were deleted, converted the enzyme into an endopeptidase (Mata 1999) to prove this hypothesis.

Recently, two peptidases that have about 55% identity in their amino acid sequences to *L. lactis* PepC and that have a natural deletion of these C-terminal residues have been identified in *Lactobacillus delbrueckii* subsp. *lactis*. These have been named PepG and PepW (Courtin 2002). *L. delbrueckii* PepG has been shown to preferentially hydrolyse Leu rich di- and tripeptides from the amino terminal end of peptides, PepW in contrast, has been shown to be a strict dipeptidyl-peptidase. To date only one PepW cleavage site has been identified, Gly-Leu↓Leu-Gly (Courtin 2002).

PepP is proline specific and exclusively hydrolyses oligopeptides of up to 10 residues containing X↓Pro-Pro-(X)n or X↓Pro-(X)n sequences at their N-termini. No hydrolysis of di- and tripeptides with similar sequences occurs (Barrett 2004).

PepT is capable of removing the N-terminal residue from tripeptides, although it is unable to hydrolyse the peptide bonds of di- or oligopeptides. PepT is capable of hydrolysing tripeptides with Pro in either position P1 or P2', but is unable to cleave peptides having Pro in the P1' position (Barrett 2004).

The specificity of *L. lactis* PepV is strictly confined to dipeptides, and it prefers dipeptides with N-terminal hydrophobic residues (Christensen 1999).

PepQ activity is defined by a specificity for X-Pro peptides, where X can be any residue (Christensen 1999; Barrett 2004).

PepX, is a serine peptidase that is active as either a monomer or dimmer releasing X-Pro dipeptides from the N-terminus of oligopeptides. In addition to peptidase activity, PepX has esterase activity. The crystallographic structure of *L. lactis* subsp. *lactis* PepX has recently been solved showing the enzyme to be a dimmer related by 2-fold symmetry (Rigolet 2002).

The crystal structure of *L. lactis* PepX shows novel structural elements that could explain the selectivity of the enzyme for proline-containing peptides, and may help to explain its ability to function not only as a peptidase but also as an esterase (Rigolet 2002).

1.3.4.2 ENDOPEPTIDASES

To date only two distinct endopeptidases from *L. lactis*, PepF and PepO have been reported. Two copies of the *pepF* gene have been located in *L. lactis*; *pepF1* (located on a large plasmid of *L. lactis* subsp. *cremoris*) and *pepF2* (located on the chromosome of *L. lactis* subsp. *lactis*). They have 84% sequence identity and have a broad endopeptidase activity for peptides ranging in length from 7 to 23 amino acids (Barrett 2004).

PepO is able to cleave peptides ranging in length from 5 to at least 30 amino acids, although it is incapable of hydrolysing di- and tripeptides, or intact α-, β- or κ-casein (Tan 1991; Pritchard 1994; Baankreis 1995; Stepaniak 1995; Barrett 2004). While PepF has broad specificity, PepO has preference for peptide bonds with hydrophobic amino acids in the P1′ position (Barrett 2004).

Interestingly, the *pepO* gene is located immediately downstream of the genes encoding the oligopeptide transport system (Mierau 1993), which suggests that the two systems may be physiologically linked. However, culture of a *L. lactis pepO* mutant, generated by gene disruption, showed that PepO was not essential for the survival of the organism (Mierau 1993; Law 1997).

Single and multiple *L. lactis* peptidase knockouts have been constructed for the *pepA, pepC, pepN, pepT, pepX, pepV, pepO* and *pepF* genes. In general, no differences were reported for the growth rate of either single or multiple knockouts compared to the wild type strain when cells were grown in complex media (MRS or M17 broth) or complete amino acid defined media. When milk was the nutrient, however, there were noticeable reductions in growth rate for most of these mutants, which increased as the number of genes knocked out increased (Mierau 1996; Christensen 1999). This observation shows that individual peptidases are not essential for the survival of the organism, and also makes it possible to analyse the effect of each enzyme on cheese ripening and flavour development.

The effect of peptidases on cheese ripening and flavour development has also been evaluated by over-expressing individual peptidases. McGarry *et al.* (1994) and Christensen *et al.* (1995) studied the influence of *L. lactis* PepN on Cheddar cheese using *Lactococcus* strains engineered to overproduce PepN. No significant changes in cheese ripening or flavour development were observed (McGarry 1994; Christensen 1999). In contrast, Meyer and Spahni

(1998) who studied the influence of *L. lactis* PepX on Gruyere cheese found that PepX had an influence on cheese ripening and flavour development (Meyer 1998; Sousa 2001).

Flavour development in cheese is of major economic interest since the final flavour of cheeses partly determines consumer choice and because the development of flavour is a time consuming and an expensive process. For this reason, there is incentive to identify LAB peptidases with novel substrate specificity that could potentially be used to modify and control cheese bitterness and possibly other flavour attributes.

1.4 *LACTOBACILLUS RHAMNOSUS* PEPO

The *Lactobacillus rhamnosus pepO* gene is approximately 1.9-kbp long and encodes a monomeric 70.9 kDa enzyme. The amino acid sequence contains the HEXXH-pentapeptide characteristic of zinc metallopeptidases, and has 36% and 27% amino acid identity to *L. lactis* PepO and *H. sapiens* neutral endopeptidase (NEP), respectively (Christensson 2002) (Figure 1.6).

Although similar in sequence, the substrate specificity of *L. lactis* PepO and *L. rhamnosus* PepO is considerably different. *L. rhamnosus* PepO has been shown to hydrolyse peptide bonds located in the N-terminal part of α_{s1}-casein(1-23), whereas *L. lactis* PepO hydrolyses peptide bonds located within the C-terminal portion of this peptide (Christensson 2002) (Figure 1.5).

α_{s1}-casein(1-23)	R - P - K - H - P - I - K - H - Q - G - L - P - Q - E - V - L - N - E - N - L - L - R - F
L. rhamnosus PepO	↑ ↑ ↑ ↑
L. lactis PepO	↑ ↑ ↑ ↑

Figure 1.5 Cleavage Specificity of *L. rhamnosus* PepO and *L. lactis* PepO for α_{s1}-casein(1-23). (↑, cleavage site).

The unique substrate specificity of *L. rhamnosus* PepO has ramifications for flavour development in cheese. The peptides α_{s1}-casein(1-13) and α_{s1}-casein(1-17) are known to exhibit a very bitter flavour, and have been found to accumulate during cheese ripening (Christensson 2002). Because PepO is able to hydrolyse peptide bonds within these regions, it has the potential to reduce bitterness by preventing the accumulation of the peptide.

In order to understand the molecular basis for the unique substrate specificity of PepO, it is necessary to determine the three-dimensional structure of the enzyme. The crystal structure of *H. sapiens* neutral endopeptidase (NEP), which shares 27% amino acid identity with PepO, has

been solved (Oefner 2000). NEP (EC 3.4.24.11) also known as neprilysin, enkephalinase, CALLA and CD10, is a monomeric 80-92 kDa integral membrane glycosylated metallopeptidase that is involved in the physiological degradation of the peptides modulating blood pressure.

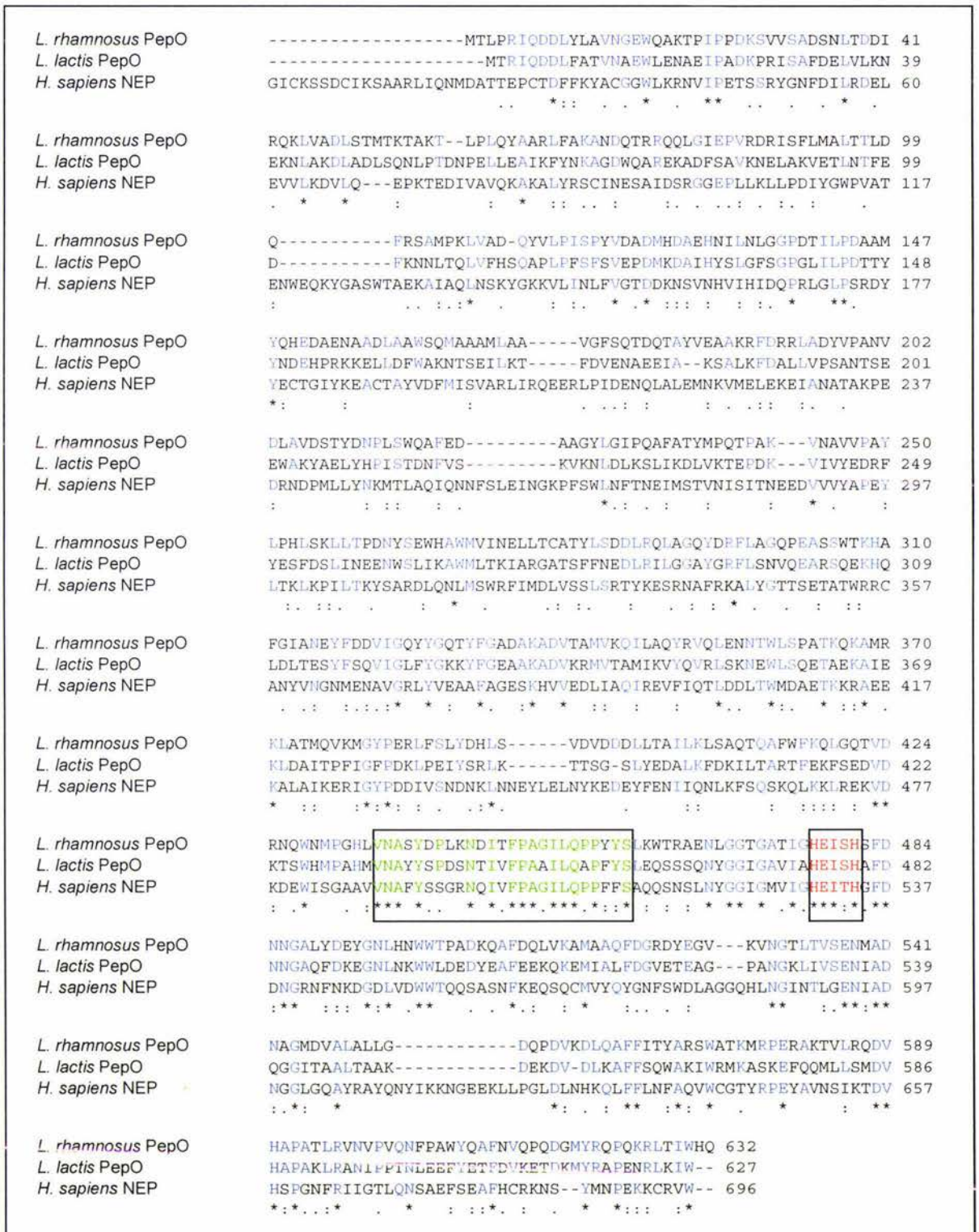


Figure 1.6 Sequence Alignment of PepO from *L. rhamnosus*, *L. lactis* and *H. sapiens* NEP. (Blue, identical residues; Green, highly conserved residues; Red, HEXXH petapeptide motif; Boxed, highly conserved domain; *, conserved residues; :, very similar residues;., similar residues. Generated using Clustal W, Chenna 2003).

The extracellular domain (residues 52-749), of NEP was crystallised in complex with the generic metallopeptidase inhibitor, phosphoramidon, and its structure solved to 2.1 Å resolution (Oefner 2000). The structure (Figure 1.7) reveals two largely α-helical domains which embrace a large central cavity containing the active site.

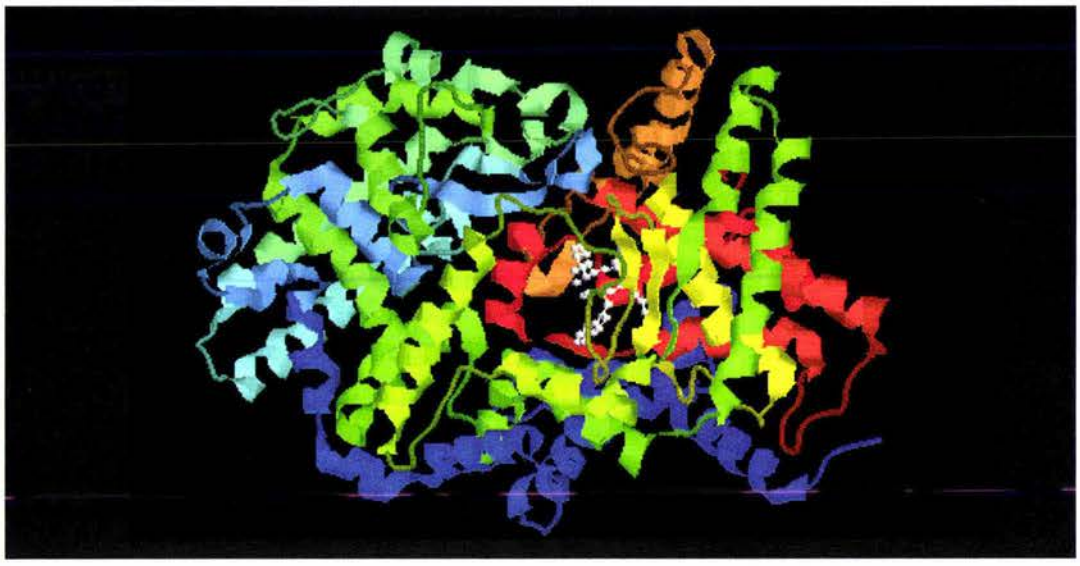


Figure 1.7 The Three-Dimensional Structure of *H. sapiens* NEP. (Cartoon representation coloured by residue from the N-terminus (blue) to the C-terminus (red). The phosphoramidon molecule is represented as a ball and stick model in white. Generated using Rasmol, Sayle 1995).

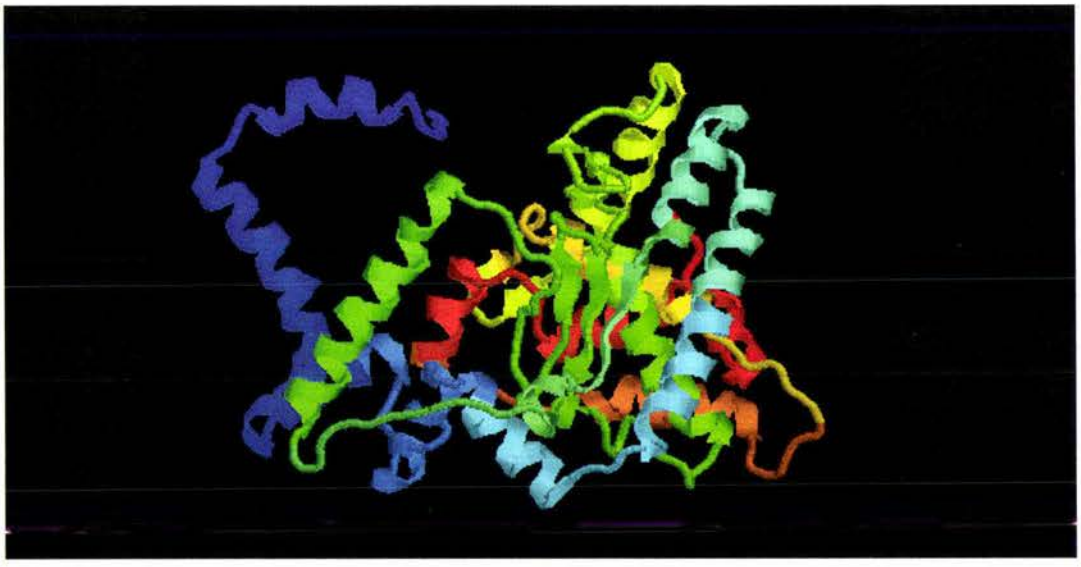


Figure 1.8 The Swiss-Model Predicted Three-Dimensional Structure of PepO. (Cartoon representation coloured by residue from the N-terminus (blue) to the C-terminus (red). Generated using Rasmol, Sayle 1995).

The inhibitor is bound to one side of this cavity and its binding mode provides a detailed understanding of the complex ligand-binding and substrate specificity determinants (Oefner 2000). The C-terminal domain of NEP shares 27% amino acid identity with PepO. Although structural investigations often require an atomic model with > 30% sequence identity, it is possible that this protein could aid in determining the three-dimensional structure of PepO. The Swiss-Model predicted three-dimensional structure of PepO using NEP as an atomic model is illustrated in Figure 1.8 (Schwede 2003).

1.5 *LACTOBACILLUS RHAMNOSUS* PEPW

Recently the *L. rhamnosus* HN001 genome has been sequenced, and analyses have revealed the presence of a 1.347 kb gene, *pepW*, encoding a putative protein of 448 amino acids, that has been named PepW (Lubbers unpublished results). PepW is 93%, 52%, 44% and 37% identical to aminopeptidase C (PepC) from *Lactobacillus casei*, *Lactococcus lactis*, and bleomycin hydrolase from *H. sapiens* (hBH) and *S. cerevisiae* (Gal6), respectively (Figure 1.9).

The name PepW was given due to incorrect sequence analysis of the *pepW* gene, which suggested that the putative peptidase, PepW, was missing the C-terminal residues needed to be classified as a PepC. Despite further analysis showing that the gene did indeed code these residues, the name of the gene and its protein product have been retained.

Although the substrate specificity of PepW is completely unknown, it is likely to have broad specificity, like the closely related LAB PepC peptidases (Barrett 2004). Substrate specificity, however, will only be unequivocally established experimentally using a number of different substrates. It may be informed by a three dimensional structure, which would allow comparisons with the structures of the yeast and human bleomycin hydrolases, which have been solved to 2.2 Å (Joshua-Tor 1995) and 2.6 Å resolution, respectively (O'Farrell 1999; Barrett 2004).

Bleomycin hydrolase (EC 3.4.22.40) is a highly conserved cysteine peptidase from bacteria to humans that hydrolyzes the anticancer drug bleomycin (Joshua-Tor 1995). The enzyme is an intracellular homohexameric cysteine peptidase of approximately 300 kDa (Barrett 2004). The structures of the *H. sapiens* and *S. cerevisiae* bleomycin hydrolases are very similar overall, adopting an unusual hexameric ring structure which has a prominent central channel of ~22 A diameter (Joshua-Tor 1995; O'Farrell 1999; Barrett 2004). The six papain-like active sites are situated within the central channel, in a manner resembling the organization of active sites in the proteasome (Joshua-Tor 1995). The three-dimensional structure of the yeast bleomycin hydrolase is shown in Figure 1.10.

Figure 1.9 Sequence Alignment of PepW from *L. rhamnosus*, *L. lactis* PepC, *H. sapiens* hBH, and *S. cerevisiae* Gal6. (Blue, identical residues; Green, highly conserved residues; Red, (Cys, His, Asn) catalytic triad; Boxed, highly conserved domains; *, conserved residues; :, very similar residues; ., similar residues. Generated using Clustal W, Chenna 2003)).

A striking feature of the molecule is the projection of the C-terminus of the protein into the active-site cleft. Intriguingly, the yeast bleomycin hydrolase has DNA-binding activity, and has been shown to bind to the upstream activating sequences (UAS) of the GAL system (hence the designation Gal6) (Joshua-Tor 1995). *L. lactis* PepC has also been shown to bind DNA (Xu 1994; Barrett 2004; Xu unpublished results), although human bleomycin hydrolase does not have this nucleic acid binding activity (Barrett 2004).

The Swiss-Model predicted three-dimensional structure of PepW obtained using a combination of native and mutant *H. sapiens* and *S. cerevisiae* bleomycin hydrolase structures as atomic models is illustrated in Figure 1.11.

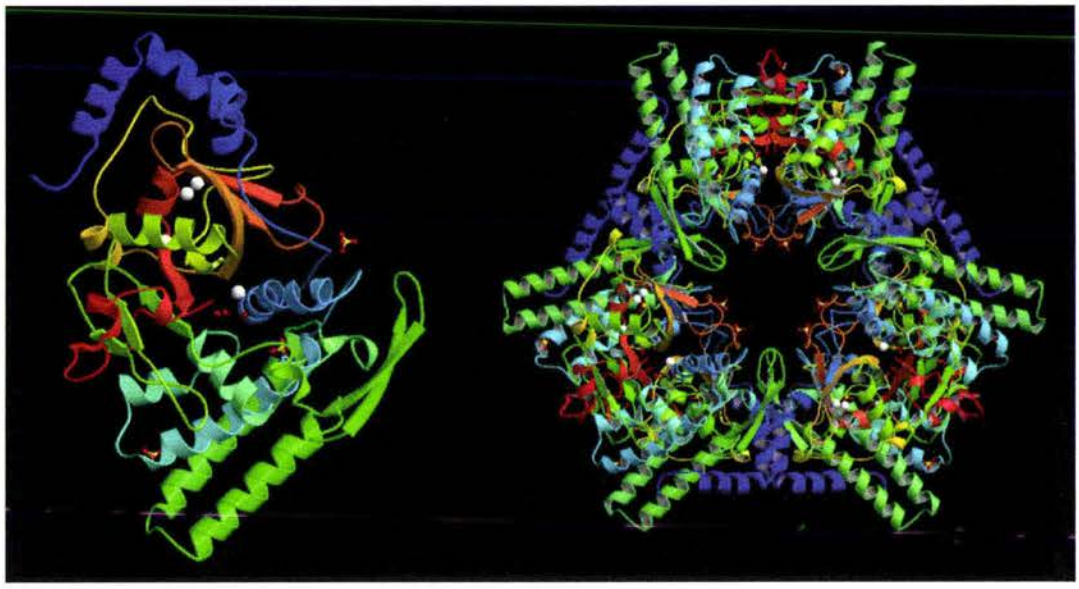


Figure 1.10 The Three-Dimensional Structure of the *S. cerevisiae* Bleomycin Hydrolase. (Cartoon representation coloured by residue from the N-terminus (blue) to the C-terminus (red). The mercury atoms (Hg) are represented as space-fill models in white. A monomer is shown on the left, with the hexamer being modelled based on crystallographic symmetry on the right, Berman 2002).

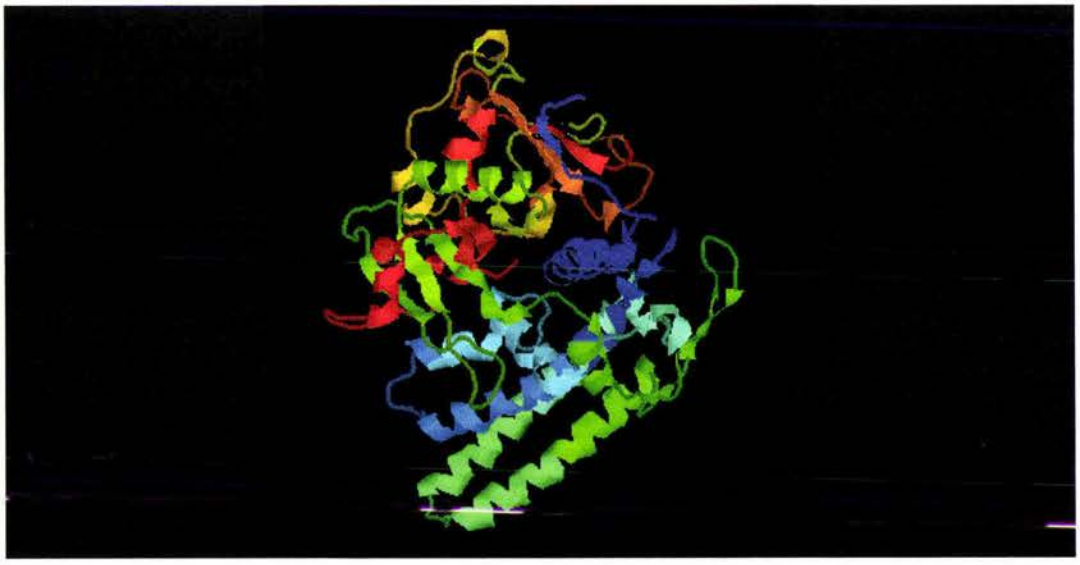


Figure 1.11 The Swiss-Model Predicted Three-Dimensional Structure of PepW. (Cartoon representation coloured by residue from the N-terminus (blue) to the C-terminus (red). Generated using Rasmol, Sayle 1995).

1.6 PROJECT AIMS

This research aims to investigate the substrate specificity and three-dimensional structure of the *L. rhamnosus* peptidases, PepO and PepW.

Aims:

1. Clone the *pepW* gene into a suitable expression system using standard molecular cloning techniques.

2. Isolate and purify recombinant PepO and PepW using traditional chromatographic methods for the former, and methods that exploit the purification tag of the latter.

3. Investigate the substrate specificity of PepO and PepW towards casein-derived substrates, using reverse-phase high pressure liquid chromatography, mass-spectrometry and protein sequencing techniques.

4. Grow suitable crystals of PepO and PepW for structure determination using x-ray diffraction.

5. Solve the three-dimensional structure of PepO and PepW using molecular replacement and/or multiple isomorphous replacement techniques.

Information obtained from this research will contribute to our fundamental understanding of the role of specific structural motifs for peptidases in general and in particular the *L. rhamnosus* peptidases, PepO and PepW. The structural and functional characterisation of *L. rhamnosus* PepO and PepW could in the future be used to inform the genetic engineering of proteolytic enzymes designed to develop new cheese flavours.

PART 1:

SUBSTRATE SPECIFICITY OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

INTRODUCTION

1.1 INTRODUCTION

Part 1 of this thesis focuses on the substrate specificity of the *L. rhamnosus* peptidases, PepO and PepW. While knowledge of the substrate specificity of PepO is limited, it has previously been shown to have a novel endopeptidase cleavage specificity towards α_{s1}-casein(1-23) that has ramifications for flavour development in cheese. The substrate specificity of PepW although completely unknown, is thought be a strict aminopeptidase activity. The aim of this section of work was to investigate the substrate specificity of the *L. rhamnosus* peptidases, PepO and PepW, towards α_{s1}-, β- and κ-casein-derived peptide substrates.

1.2 OBJECTIVES AND STRATEGIES

The objectives of this section of work were:

1) Clone the *L. rhamnosus pepW* gene into a suitable expression system using standard molecular cloning techniques.

2) Express and purify the recombinant *L. rhamnosus* PepO and PepW enzymes from *L. lactis* and *E. coli* respectively, using traditional chromatographic techniques, for the former and methods aimed at exploiting the purification tag on the latter.

3) Express and purify two *L. lactis* cell envelope proteinases PrtP using an extraction technique based on autoproteolysis.

4) Prepare bovine α_{s1}-, β- and κ-casein peptides by proteolytic digestion using chymosin and *L. lactis* cell-envelope proteinases.

5) Investigate the substrate specificity of PepO and PepW for the α_{s1}-, β- and κ-casein peptides using high-pressure liquid chromatography, mass-spectrometry and protein sequencing techniques.

MOLECULAR CLONING

2.1 INTRODUCTION

Molecular cloning is an important tool in the endeavour to understand the structure, function and regulation of individual genes and their products. This Chapter describes the cloning of the *L. rhamnosus pepW* gene into a suitable expression system to produce recombinant protein to allow the biochemical characterisation of the putative peptidase, PepW.

The first step in preparing a suitable expression system is to obtain the target gene. Knowledge of the DNA sequence allows selective amplification of the gene by the polymerase chain reaction. Following amplification, the gene is cloned into an appropriate expression system, the choice of which is critical for successful production of recombinant proteins.

In general, bacterial hosts are the easiest to use in terms of the preparation of the expression construct, the growth of the recombinant organism and the purification of the resulting protein. Although bacteria are often not suitable hosts for cloning eukaryote genes (due to problems arising from host-inability to carry out post-translational modification and protein folding), they are ideal host organisms for cloning bacterial genes. The difference in codon usage between gram positive and gram negative organisms must be taken into consideration when choosing a host organism to express the recombinant gene.

The well characterised bacterium, *E. coli*, is routinely used as a heterologous host, and as a consequence the greatest selection of cloning vectors exists for use with this host organism. The simplest *E. coli* cloning vectors, and the ones in most widespread use are those based on small bacterial plasmids. All plasmid vectors possess at least one DNA sequence that can act as an origin of replication, restriction endonuclease sites that allow the insertion of additional DNA, and a gene encoding antibiotic resistance as a selectable marker. More complex plasmid vectors, in particular expression vectors, possess additional components including a promoter, terminator and ribosome binding site.

The promoter is the most important component of an expression vector because it controls transcription initiation. The promoter is recognized by RNA polymerase and determines the rate at which mRNA is synthesized.

An efficient expression vector requires not only a strong, regulatable promoter, but also a ribosome binding site and a terminator. In most expression vectors these components form a cassette, into which the target gene is inserted. In some cassettes, the restriction endonuclease sites used to insert the target gene are not immediately adjacent to the ribosome

binding site, but are separated by a DNA sequence encoding either a gene, or gene fragment, called a tag. Insertion of the target gene so that it is in frame with this tag generates a fusion protein. Fusion tags are commonly used to facilitate detection and purification of the target protein, but can also influence protein solubility and/or stability. As fusion tags may alter the properties of the target protein, easy removal from the target protein is an advantage. This may be achieved by the inclusion of a sequence coding for a specific protease site before the target protein and after the tag, allowing subsequent enzyme assisted removal of the tag. It is important therefore that the protease used will not hydrolyse the target protein, necessitating the choice of proteases that are highly specific, and that recognise uncommon motifs.

2.2 METHODS

2.2.1 Agarose Gel Electrophoresis

DNA fragments were separated on the basis of size by agarose gel electrophoresis. Typically 10-20 μL DNA were mixed with 6x loading dye (0.2% bromophenol blue, 50% glycerol) and loaded onto a 1% agarose Tris-acetate-EDTA (40 mM Tris-HCl, 20 mM acetic acid, 2 mM EDTA, pH 8.0) gel. The gels were routinely run at 80 V until the dye front had migrated to the other end of the gel. The gels were stained with ethidium bromide and the DNA bands visualised by ultraviolet (UV) illumination. A 1 kb DNA ladder was routinely used, and quantitative DNA standards were used when required.

2.2.2 Genomic DNA Preparation

L. rhamnosus HN001 genomic DNA was isolated from a 20 mL culture grown anaerobically overnight at 37°C in MRS broth (Composition in g/L; peptone from casein, 10.0; meat extract, 80.0; yeast extract, 4.0; D(+)-glucose, 20.0; di-potassium hydrogen phosphate, 2.0; Tween 80, 1.0; di-ammonium hydrogen citrate, 2.0; sodium acetate, 5.0; magnesium sulphate, 0.2; manganese sulphate, 0.04). The cells were harvested by centrifugation (6,000 g for 10 min at 4°C), resuspended in 20 mL MRS and incubated at 37°C for 2 hr, before being harvested as above. They were then washed twice with 4 mL 50 mM NaCl, 5 mM EDTA, 30 mM Tris pH 8.0, then re-suspended in 0.5 mL 25% Sucrose, 1 mM EDTA, 20 mg/mL Lysozyme, 20 μg/mL Mutanolysin, 50 mM Tris pH 8.0, for lysis at 37°C for 1 hr. After this time 200 μL 20% (w/v) SDS was added and incubation continued at 65°C for 1 hr, followed by the addition of 10 μL Proteinase K (20 mg/mL) and incubation at 65°C for 15 min. An equal volume of a phenol/chloroform/-isoamyl alcohol (25:24:1), was then added to the solution which was incubated at room temperature for 5 min after which time the supernatant was separated from the genomic DNA by centrifugation (4,000 g for 10 min at 4°C). RNase was added to a final concentration of 100 μg/mL, and the solution incubated at 37°C for 30 min before re-extracting the DNA as above. An equal volume of chloroform/isoamyl alcohol (24:1) was added to the solution which was incubated at room temperature for 5 min before being centrifuged (13,000 g for 5 min at 4°C) to separate the aqueous phase. Genomic DNA (aqueous phase) was isolated by ethanol precipitation (overnight at 4°C), and the pellet resuspended in 100 μL 10 mM Tris pH 8.0.

2.2.3 Polymerase Chain Reaction

The polymerase chain reaction (PCR) was used to amplify the 1.3 kb *L. rhamnosus pepW* gene. Conditions for a typical PCR reaction were as follows: 100 ng of *L. rhamnosus* HN001 template

was amplified with 50 pM of both forward and reverse primers:

Forward primer: CAGCGGAATTCATGTCTGCAGAAATTACTTCAGGCG
Reverse primer: CTGTAGCGGCCGCTTATCTTGATGCCAAAGCACCCA

The mixture contained 2.5 units of Pwo DNA polymerase, 200 μM deoxyribo-nucleotide triphosphate (dNTP), in 1x standard Pwo polymerase buffer containing $MgSO_4$ in a total reaction volume of 20 μL. The thermo-cycling program used was: initial denaturation at 95°C for 5 min followed by 30 cycles of primer annealing at 55°C for 30 sec with extension at 72°C for 90 sec and denaturation at 95°C for 30 sec. A positive PCR control using a proven combination of primers and template was included, as was a negative control in which no template DNA was added to the reaction mixture. The PCR products were purified with a Roche High Pure PCR Product Purification Kit and PCR amplification confirmed by agarose gel electrophoresis.

2.2.4 Restriction Endonuclease Digests

200 ng of PCR product and 200 ng of pET32a-Trx-His-rTEV vector were digested with 5 units of *EcoRI* and *NotI*, in 1x REACT3™ buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, pH 8.0) in a reaction volume of 50 μL, at 37°C for 2-3 hr. Restriction endonuclease (RE) digestion was stopped by incubation at 65°C for 15 min and the digested products purified with a Roche High Pure PCR Product Purification Kit.

2.2.5 Ligation of DNA Fragments

50 ng RE-digested PCR product and pET32a-Trx-His-rTEV were ligated with 2 units of bacteriophage T4 DNA ligase, 1x ligase buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, pH 7.5) in a final volume of 10 μL, at 12-16°C overnight. Controls included ligation reaction mixtures containing RE-digested pET32a-Trx-His-rTEV DNA only, and a mixture containing PCR product and pET32a-Trx-His-rTEV without T4 DNA ligase.

2.2.6 Transformation

Transformation of *E. coli* Top10 and BL21(DE3) competent cells was carried out using ligated DNA fragments and plasmid DNA (section 2.2.7) respectively. Top10 competent cells were used in this research because they are resistant to phage contamination, which had been a problem in the laboratory. Typically, 100 μL Top10 or BL21(DE3) competent cells were incubated with 5 μL ligation reaction and 50 ng plasmid DNA respectively, at 4°C for 10 min. The reaction mixture was then heat shocked at 42°C for 60 sec, incubated at 4°C for a further 5 min, before being plated onto LB (Luria-Bertani) agarose plates (Composition in g/L: tryptone,

10.0; yeast extract, 5.0; NaCl, 5.0; agarose, 15.0) with ampicillin selection (100 μg/mL final concentration) and incubated at 37°C overnight. Controls included transformation of 100 μL competent cells with 50 ng uncut pET32a-Trx-His-rTEV vector, and 100 μL competent cells with no additional DNA. Single colonies were grown aerobically at 37°C in 4 mL LB liquid media (1% Tryptone, 0.5% yeast extract, 1% NaCl) with ampicillin selection (100 μg/mL ampicillin) overnight and stored as glycerol stocks at -80°C.

2.2.7 Plasmid DNA Preparation

Plasmid DNA from the Top10 transformants was isolated by inoculating 4 mL LB-AMP liquid media with a single colony and the culture grown aerobically at 37°C overnight. The cells were harvested by centrifugation (9,000 g for 1 min at 4°C) and the plasmid DNA isolated with a Roche High Pure Plasmid Isolation Kit. Plasmid DNA isolation was confirmed by agarose gel electrophoresis.

2.2.8 DNA Sequence Analysis

The insert DNA was subjected to DNA sequence analysis before proceeding further. Typically, 700 ng of plasmid DNA from the transformed Top10 cells was incubated with 10 pM T7 promoter or 10 pM T7 terminator primers and sequenced using the dye-termination method described by Rosenthal (1992) on an ABI Prism 377 automated DNA sequencer (Rosenthal 1992).

2.2.9 Sequence Analysis

DNA and amino acid sequence analyses of the *pepW* gene and its protein product were carried out using the Basic Local Alignment Search Tool (BLAST) available via the National Centre for Biotechnology Information (NCBI) web server (Atschul 1990).

2.3 RESULTS AND DISCUSSION

2.3.1 *L. RHAMNOSUS PEPW* CLONING

Analyses of the *L. rhamnosus* HN001 genome revealed the presence of a gene encoding a putative peptidase, PepW, that appears to have evolved as a result of gene duplication (Lubbers unpublished results) (Figure 2.1).

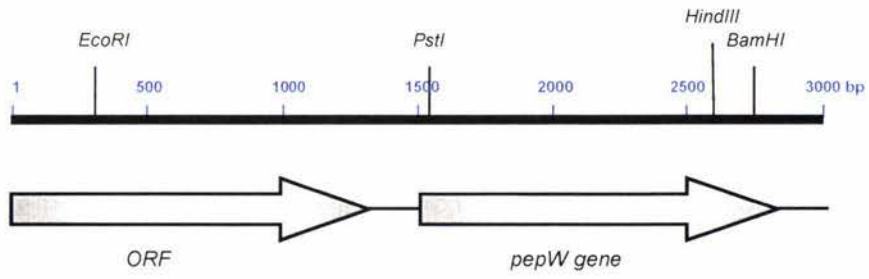


Figure 2.1 Operon Containing the *pepW* Gene.

The *L. rhamnosus pepW* gene is 1.347 kb in length and encodes a protein of 448 amino acids, that is 52% identical (according to BLAST) to the protein encoded by the open reading frame (ORF) within the operon, and 93%, 67% and 62% identical to aminopeptidase C (PepC) from *Lactobacillus casei*, *Lactobacillus plantarum*, and *Lactobacillus gasseri*, respectively (Atschul 1990). The putative PepW primary sequence contains domains that are conserved in PepC strongly suggesting that it is a cysteine peptidase (Atschul 1990) (Figure 2.2).

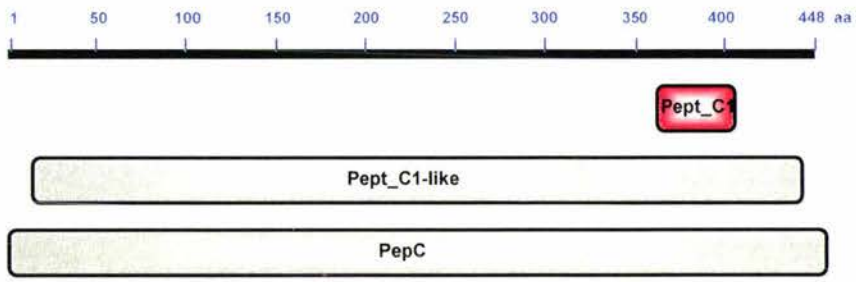


Figure 2.2 Conserved Domains of the Putative Peptidase, PepW. (Generated using BLAST, Atschul 1990).

To investigate the protein coded by this gene it was necessary to clone it into a suitable expression system, produce and purify the recombinant protein. This was achieved using a modified pET32a expression vector (Figure 2.3).

The pET32a vector contains strong T7 bacteriophage transcription and start and termination signals that provide high-level protein expression, induced by T7 RNA polymerase. The pET32a expression vector contains however, an unnecessarily large fusion construct (Trx-His-Thrombin-S-tag-

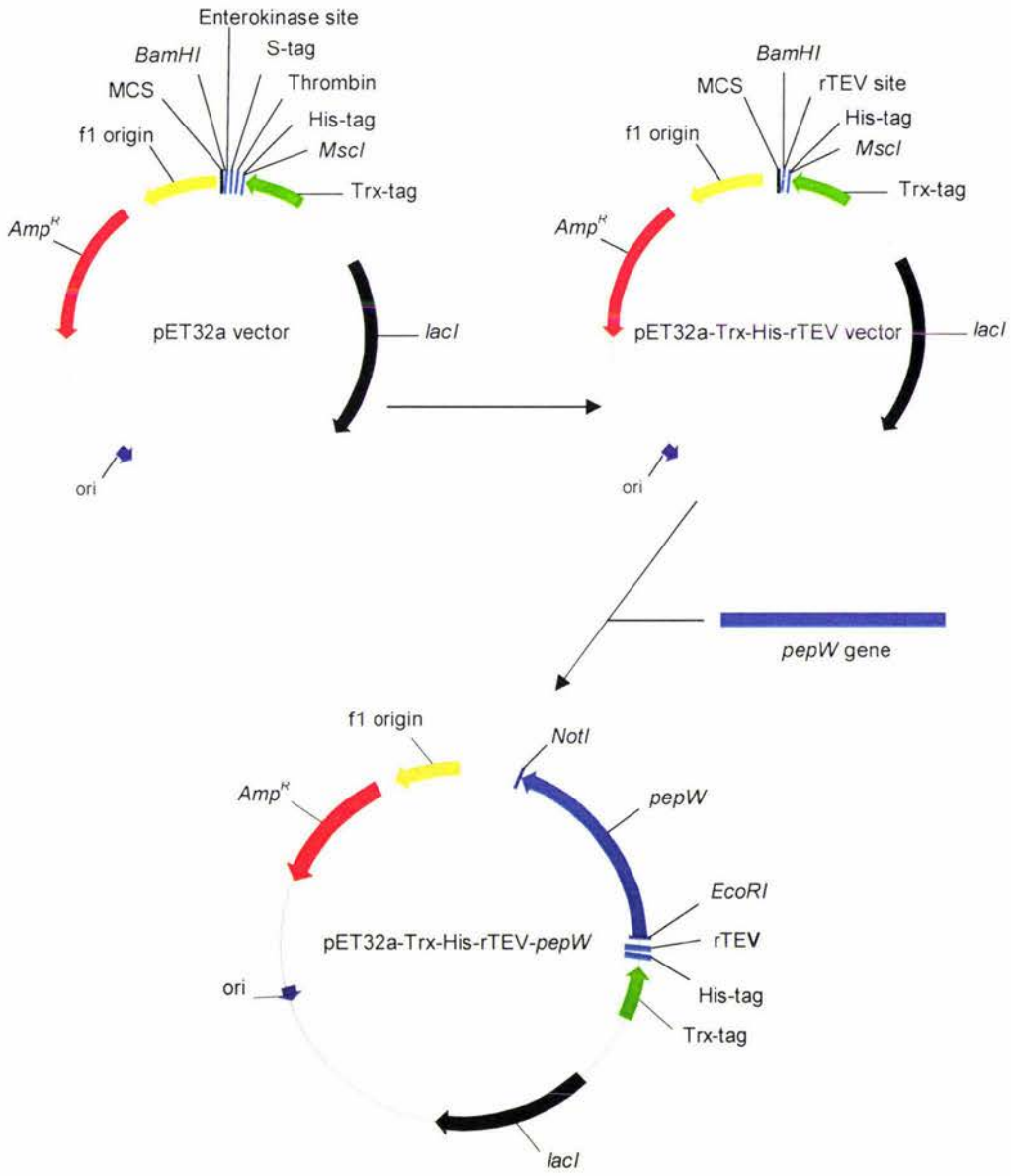


Figure 2.3 Cloning Strategy for the *pepW* Gene. The *pepW* gene was cloned into a modified pET32a vector, pET32a-Trx-His-rTEV, using the *EcoRI* and *NotI* sites within the multiple cloning site (MCS).

Enterokinase) that requires removal by enterokinase. A modified pET32a vector (a gift from Dr Rosemary Brown, Massey University, Palmerston North, New Zealand) was used, in which the large construct had been excised using *MscI* and *BamHI* and replaced with a sequence encoding a His-tag and rTEV protease cleavage site (His-rTEV) generating the construct, pET32a-Trx-His-rTEV (Figure 2.3).

The pET32a-Trx-His-rTEV construct contains in addition to the His-rTEV tag a 109 amino acid N-terminal thioredoxin (Trx) tag which itself is highly soluble and enhances the solubility of proteins fused to it (LaVallie 1993). The vector contains a strong T7 late promoter upstream from the fusion construct that is specifically recognised by T7 RNA polymerase, and a multiple cloning site downstream from the fusion construct. The *EcoRI* and *NotI* sites of the multiple cloning site were chosen as suitable insertion sites for the *L. rhamnosus pepW* gene, as the *pepW* gene does not contain recognition sequences for these enzymes. The *EcoRI* site is close to the rTEV protease cleavage site, resulting in the reduction of the number of extra residues remaining at the N-terminus of the protein after removal of the fusion tag.

PCR amplification of the 1.3 kb *pepW* gene from *L. rhamnosus* HN001 was successfully carried out using forward and reverse primers with flanking *EcoRI* and *NotI* sites respectively (Figure 2.4).

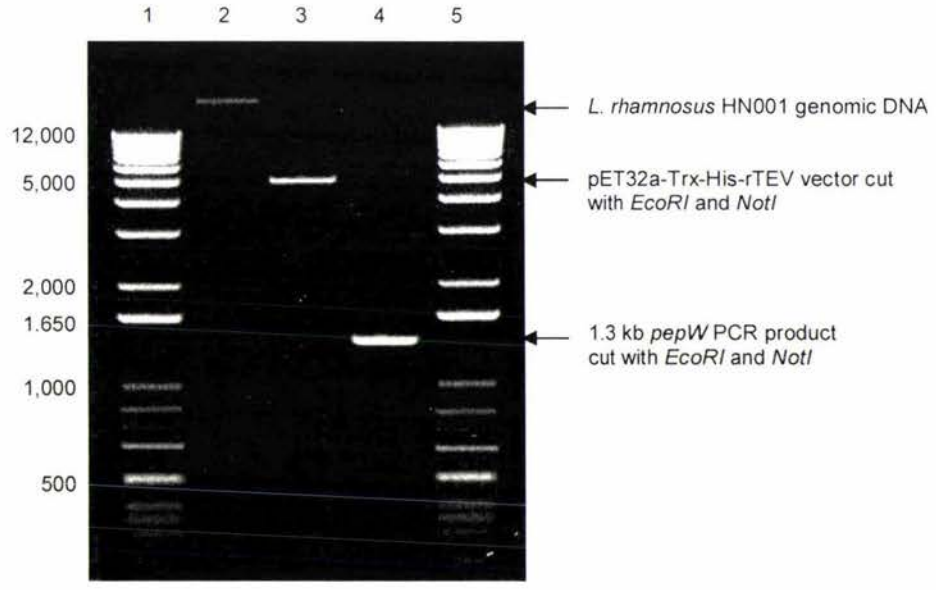


Figure 2.4 Agarose Gel Electrophoresis Analysis of *pepW* Cloning. (Lanes 1 and 5, 1 kb ladder (bp); 2, *L. rhamnosus* HN001 genomic DNA; 3, pET32a-Trx-His-rTEV vector; 4, PCR product).

Following RE digestion of the PCR product with *EcoRI* and *NotI*, it was successfully ligated into the pET32a-Trx-His-rTEV vector. *E. coli* Top10 competent cells were then transformed with the pET32a-Trx-His-rTEV-*pepW* construct and plated on LB-AMP plates. DNA sequence analysis confirmed that the *pepW* gene was successfully ligated in-frame and without errors into the *EcoRI* and *NotI* sites of the pET32a-Trx-His-rTEV vector (Appendix I).

The pET32a-Trx-His-rTEV-*pepW* construct was then used to transform BL21(DE3) cells for protein expression. BL21(DE3) cells contain a DE3 lysogen encoding the T7 RNA polymerase gene under control of the lacUV5 promoter. Induction with isopropyl-thiogalactoside (IPTG) activates transcription and expression of the T7 RNA polymerase, which in turn activates transcription and expression of the target gene downstream from the T7 promoter on the expression vector.

The expression and purification of *L. rhamnosus* PepW from this expression system will be discussed in Chapter 3.

PROTEIN EXPRESSION AND PURIFICATION

3.1 INTRODUCTION

The purification of a peptidase, like the purification of any other protein, should begin with consideration of the goals of the purification. These may range from investigations into the biochemical properties of isolated enzymes to detailed structural studies of purified proteins or protein subunits (Beynon 1989). Such considerations will dictate the degree of purification required to obtain protein in sufficient purity and quantity. A systematic approach exploiting protein properties can be used to develop a suitable protein purification strategy (Table 3.1).

Table 3.1 Protein Properties Exploited in Purification.

Protein property	Technique
Charge	Ion exchange (IEX), Chromato-focussing, Isoelectric Focussing
Hydrophobicity	Hydrophobic interaction Chromatography (HIC)
Size	Size exclusion Chromatography (SEC)
Biorecognition	Affinity Chromatography (AC)

The selection and combination of the purification techniques are dependent upon the biochemical properties of the target and host organism proteins. If little is known about the target protein, an effective general approach is to carry out ion exchange (IEX), followed by hydrophobic interaction chromatography (HIC) and finally size exclusion chromatography (SEC).

Ion exchange chromatography (IEX) separates proteins according to the net surface charge of proteins and is dependent upon pH. When the pH of the solution is above a protein's isoelectric point (pI) it will be negatively charged and will bind to an anion exchanger, conversely at a pH below its pI it will be positively charged and will bind to a cation exchanger. Typically, conditions are chosen under which the target protein, binds to the matrix while the majority of contaminating protein do not. An alternative strategy is one in which the target protein does not bind, in contrast to most of the impurities. The ionic interaction between protein and matrix is enhanced at low ionic strength, and reduced at high ionic strength. IEX chromatography is commonly used as the first step in protein purification because it has a high binding capacity.

Hydrophobic interaction chromatography (HIC) separates proteins according to differences in hydrophobicity. The separation is based on the reversible interaction between a hydrophobic stationary phase and the hydrophobic regions of a protein. This interaction is enhanced by a

high ionic strength buffer, which makes HIC a suitable purification step following IEX. HIC like IEX has a high binding capacity and is commonly used towards the beginning of the purification strategy.

Size exclusion chromatography (SEC) separates molecules according to differences in size. Unlike IEX and HIC, proteins do not bind to the stationary phase. SEC is often used as a final purification step as only a small volume of concentrated protein sample can be applied to the column.

If the target protein has been fused to a purification tag, affinity chromatography may be used. Affinity chromatography (AC) separates proteins on the basis of a reversible interaction between the target protein and a specific stationary phase ligand. The advantage of AC is the high selectivity for the target protein. Often AC can achieve the purification required without the need for additional purification steps, although its removal may require additional steps.

In all protein purification strategies the number of steps necessary to reach the required purification should be minimised, without compromising the quality of the end product. Protein purification should be monitored at each purification step. A protein purification table recording the protein yield, volume, activity and specific activity where appropriate, is a useful tool for monitoring the purification procedure.

In addition to the normal challenges associated with protein purification, there are specific problems associated with the purification of proteolytic enzymes which complicates the interpretation of a purification table. Proteolytic degradation is a major concern, and can occur as a result of host proteolytic enzymes and autoproteolysis. While purification tables are both useful and helpful, they should be viewed cautiously and always with the above-mentioned considerations in mind (Beynon 1989).

3.2 METHODS

3.2.1 Growth of Genetically Modified Bacteria

Bacterial expression systems using *E. coli* and *L. lactis* host organisms were used in this research. *E. coli* BL21 transformed with pET32-Trx-His-rTEV-*pepW* was used to express *L. rhamnosus* PepW, and *L. lactis* subsp. *lactis* NZ9000 transformed with pLB004 (a gift from Fonterra Research, Palmerston North, New Zealand) was used to express *L. rhamnosus* PepO. *E. coli* and *L. lactis* cultures were grown aerobically at 25°C and anaerobically at 30°C respectively.

3.2.2 Induction of Protein Expression

E. coli BL21 pET32a-Trx-His-rTEV-*pepW* was grown aerobically at 25°C in 200 mL LB liquid media containing a final ampicillin concentration of 100 μg/mL. When an OD_{600nm} of 0.5 was reached protein expression was induced with IPTG (0.25-0.75 mM IPTG final concentration). Following induction, the culture was grown for a further 3 hr and harvested. As a control, a culture was grown as above without the addition of IPTG. *L. lactis* subsp. *lactis* NZ9000 pLB004 was grown anaerobically at 30°C in 1 L M17 media (Composition in g/L: beef extract, 5.0; yeast extract, 2.5; peptone, 5.0; soytone, 5.0; D(+)-lactose, 5.0; ascorbic acid, 0.5; sodium-β-glycerophosphate, 19.0; magnesium sulphate, 0.25) containing 1 mL chloramphenicol (5 mg/mL) in ethanol. When an OD_{600nm} of 0.5 was reached protein expression was induced with 1.5 mL nisaplin (10mg/mL) in 0.1 M HCl. Following induction the culture was grown for a further 3 hr and then harvested. As a control, a culture was grown as above without the addition of nisaplin.

3.2.3 Harvesting and Lysis of Cells

Bacteria were harvested by centrifugation (5,000 g for 15 min at 4°C) and then lysed. The *E. coli* cells were lysed by sonication at 50 Hz using a VirTis VirSonic Digital 550 Ultrasonic Cell Disrupter. *L. lactis* cells were lysed by two passes through either a Wabash French pressure cell (Aminco Instruments Co.) at 15 kPSI, or a 'ZPlus' 1.1 kW Benchtop Cell Disrupter (Constant Systems, Model code Z5/40/AE/GA) at 15 kPSI. Cell debris was removed by centrifugation (13,000 g for 15 min at 4°C), and the supernatant and pellet collected for solubility analysis. Samples were taken prior to centrifugation for whole cell extract analysis. Protease inhibitors were not added to the cell lysate to prevent possible inhibition of target peptidase activity.

3.2.4 Immobilised Metal Affinity Chromatography

5.0 mL protein (approximately 5.0 mg/mL) samples were loaded onto a Chelating sepharose (Amersham) column (10 x 100 mm) loaded with nickel and equilibrated with 0.5 M NaCl, 50 mM

Tris pH 8.0. Proteins bound to the column were eluted with an increasing imidazole gradient (0-100%) with 1 M Imidazole, 0.5 M NaCl, 50 mM Tris pH 8.0 at a flow rate of 1.0 mL/min over 50 minutes. Purification was carried out using a low pressure BIO-RAD Econo System at 4°C, and monitored at 280 nm using a SEKONIC SS-250F recorder. All further chromatography was carried out on a BioLogic DuoFlow System (BIO-RAD) using Version 3.0 software, Model 2128 Fraction collector collecting 1.0 mL fractions, and BioLogic QuadTec UV-Vis Detector measuring absorbance at 214 nm and 280 nm.

3.2.5 Anion Exchange Chromatography

20.0 mL protein samples (approximately 20 mg/mL) were dialysed against 50 mM Tris pH 8.0 before loading onto a Source-^{15}Q (Amersham) column (20 x 100 mm) equilibrated with 50 mM Tris pH 8.0. Proteins bound to the column were eluted by a linear salt gradient of 0-1.0 M NaCl (in 50 mM Tris pH 8.0), at a flow rate of 1.0 mL/min over 50 minutes. Purification was carried at 4°C.

3.2.6 Hydrophobic Interaction Chromatography

5 mL protein samples (approximately 5 mg/mL) were loaded onto a Source-phenyl (Amersham) column (15 x 100 mm) equilibrated with 1.25 M Ammonium sulphate, 50 mM Tris pH 8.0. Proteins bound to the column were eluted with a descending salt gradient of 1.25-0.0 M Ammonium sulphate (in 50 mM Tris pH 8.0), at a flow rate of 0.4 mL/min over 75 minutes. Purification was carried out at room temperature.

3.2.7 Gel Filtration Chromatography

250 μL PepO (approximately 10mg/mL) and PepW (approximately 5mg/mL) protein samples were loaded onto a Superdex^{TM}200 column (10 x 300 mm) (Amersham) equilibrated with either 0.15 M NaCl, 50 mM Tris pH 8.0 (PepO), or 0.1 M NaCl, 2 mM DTT, 50 mM Tris pH 7.5 (PepW), and the proteins were eluted isocratically at a flow rate of 0.25 mL/min over 120 minutes. Purification was carried out at 4°C. Sigma Gel Filtration Molecular Mass Markers (12,400 to 200,000 Da) were used to calibrate the column for molecular weight determination of the eluting proteins.

3.2.8 Proteolytic Removal of Purification Tag

The removal of the purification Trx-His-tag was carried out with recombinant Tobacco Etch Virus (rTEV) protease cleavage. Typically 10 mL protein (approximately 0.75 mg/mL) was digested with 200 μL 1 mg/mL rTEV at room temperature for 15-20 hr and cleavage was monitored by SDS-PAGE.

3.2.9 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to the protocol of Davis (1964). Typically 10-20 μL protein samples were mixed with (2-4 μL) 5x loading dye (0.25 M Tris-HCl pH 6.8, 6% SDS, 40% glycerol, 0.04% bromophenol blue) boiled at 100°C for 1 min and loaded onto either a 7.5% or 12% acrylamide SDS-PAGE separating gel with a 4% stacking gel. The gels were routinely run at 200 V until the dye front had migrated to the other end of the gel, after which the gels were stained with Coomassie Blue (R-250). Broad and low range markers were routinely used in the 7.5% and 12% gels respectively.

3.2.10 Edman N-terminal Sequencing

Purified protein was run on a 12% SDS-PAGE gel under standard conditions and electro-blotted onto a polyvinylidene fluoride (PVDF) membrane at 60 V for 90 min. The blot was briefly stained in Coomassie Blue to visualise the protein band and dried. The band was excised and the amino acid sequence derived using the chemical degradation method developed by Edman (1950) on an Applied Biosystems 476A automated protein sequencer.

3.2.11 High-Performance Liquid Chromatography (HPLC)-based Peptidase Activity Assay

The high-performance liquid chromatography (HPLC)-based peptidase assays were carried out using a Phenomoenex Jupiter C_{18} column (5μ, 4.6 x 250 mm), on a Dionex Summit HPLC system connected to a UVD-340S Photo Array Detector, using Chromeleon Version 6.4 Software. Peptidase activity was detected using a discontinuous assay based on reverse phase (RP) HPLC separation of the reaction products from substrate. Typically, 100 μL of 0.65 mg/mL α_{s1}-casein(1-23) was incubated with 40 μL protein sample at room temperature for 0-2 hr. 20 μL aliquots were taken at time intervals (0-2 hr) and an equal volume of 2% trifluoroacetic acid (TFA) added to stop proteolysis. 60 μL 1% TFA was added to give a total volume of 100 μL, the samples centrifuged (10,000 g for 1 min at 4°C) to remove precipitated protein and 80 μL of the supernatant injected onto the column which had been pre-equilibrated with 0.1% TFA in deionised water. The peptides bound by the column were eluted by an increasing gradient of 0.08% TFA in Acetonitrile and monitored at 214 nm and 280 nm. The disappearance of the substrate peak and appearance of product peaks indicated peptidase activity.

3.2.12 Concentration and Storage of Enzymes

Protein concentrations were determined using the Bradford Assay as described by Bradford (1976) or by their absorbance at 280 nm. Protein samples were concentrated to a minimum of

2.5 mg/mL by ultrafiltration using Vivaspin concentrators with a molecular weight cut off of 10 kDa, filtered (0.22 μm filters) to prevent growth of microorganisms and stored at either 4°C for short periods or -20°C or -80°C for extended periods.

3.2.13 Amino Acid Sequence Analyses

Amino acid sequence analyses of PepO and PepW was carried out using ProtParam a fully automated program that allows the computation of various physical and chemical parameters including: molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity, available from the Expasy web server (Gasteiger 2003).

3.3 RESULTS AND DISCUSSION

3.3.1 *L. RHAMNOSUS* PEPO EXPRESSION AND PURIFICATION

Recombinant *L. rhamnosus* PepO was expressed in a *Lactococcus lactis* nisin-inducible expression system developed by the Netherlands Institute for Dairy Research (NIZO, Ede, The Netherlands). The nisin-inducible expression system is based on a two-component signal transduction pathway involving a cell surface histidine-kinase, NisK, and a response-regulator protein, NisR (Kuipers 1997). Nisin signals transcriptional activation by activating NisK, which in turn activates NisR, enabling it to bind to and activate transcription from the *nisA* promoter upstream of the *L. rhamnosus pepO* gene (Figure 3.1).

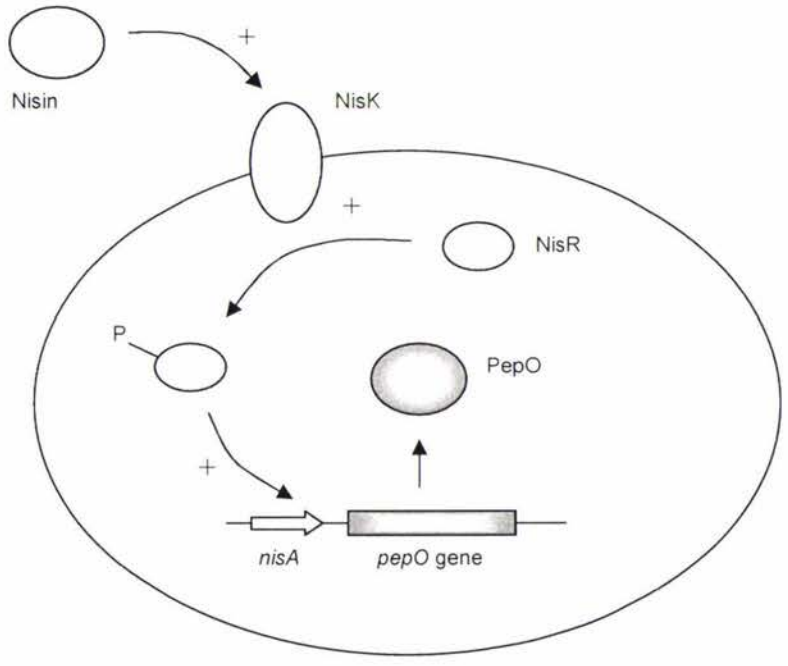


Figure 3.1 *Lactococcus lactis* Nisin-Inducible Expression System for *L. rhamnosus* PepO. (NisK, cell surface histidine-kinase; NisR, response-regulator protein; P, phosphate group; +, activation)

The *L. lactis* nisin-inducible expression system provides a tightly controlled, food-grade system for the overproduction of LAB peptidases. *L .lactis* is a well characterised LAB that is routinely used in the dairy industry, and nisin is a 3.5 kDa antimicrobial peptide that is widely used as a food preservative. Nisaplin, an impure solution of nisin, effectively induced the overexpression of recombinant PepO in the *L. lactis* expression system. Approximately 90% of the overexpressed protein was soluble, as estimated by SDS-PAGE (Figure 3.2).

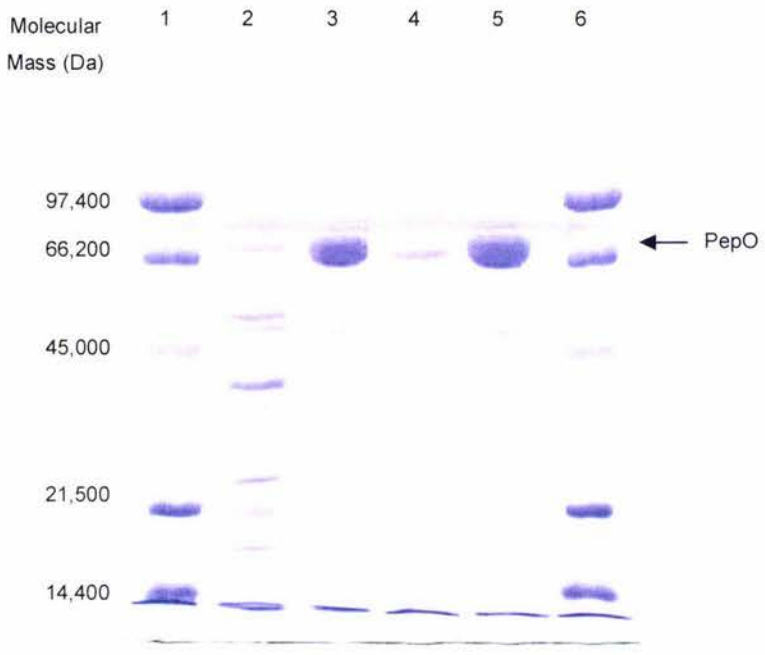


Figure 3.2 SDS-PAGE Analysis of Nisin Induced Expression and Solubility of PepO. (Lanes 1 and 6, low-range marker; 2, uninduced and; 3, nisin-induced cell extract; 4, insoluble fraction and; 5, soluble fraction of the nisin-induced cell extract).

Recombinant PepO had a mass of about 70 kDa as estimated by SDS-PAGE, close to the theoretical mass of the protein (70.9 kDa). Overloaded protein bands often appear to run further than when loaded at 'normal' levels (i.e. 1-2 μg protein/band) (Figure 3.2).

Peptidase activity was assayed using a previously established HPLC-based activity assay using α_{s1}-casein(1-23) (Christensson 2002), however, no significant difference was observed between the cell free extracts of uninduced and induced samples. The peptidase activity detected was therefore not due to PepO, and it was not possible to verify that the overexpressed protein was *L. rhamnosus* PepO, or that the protein was active.

As the interpretation of peptidase activity assays on impure protein samples can often be erroneous, due to the presence of endogenous peptidases, substrates and/or inhibitors. Purification of the overexpressed protein was carried out, and peptidase activity assays and Edman N-terminal sequencing of the purified enzyme used to confirm that the protein was indeed *L. rhamnosus* PepO.

The purification strategy chosen for PepO was based on one previously established purification (Christensson 2002), and involved ion exchange, hydrophobic interaction and size exclusion

chromatography. Purification was monitored at each step by SDS-PAGE, and peptidase activity assays.

Ion exchange chromatography (IEX) which separates proteins according to the net surface charge was chosen as the first step in the purification procedure. Because of its low calculated pI (4.86), anion exchange chromatography was used. However, even at pH 8.0 PepO bound only weakly to the resin being eluted at approximately 0.05 M NaCl. Fortunately, the majority of *E. coli* proteins remained bound to the column at this salt concentration providing a very efficient purification step (Figure 3.3).

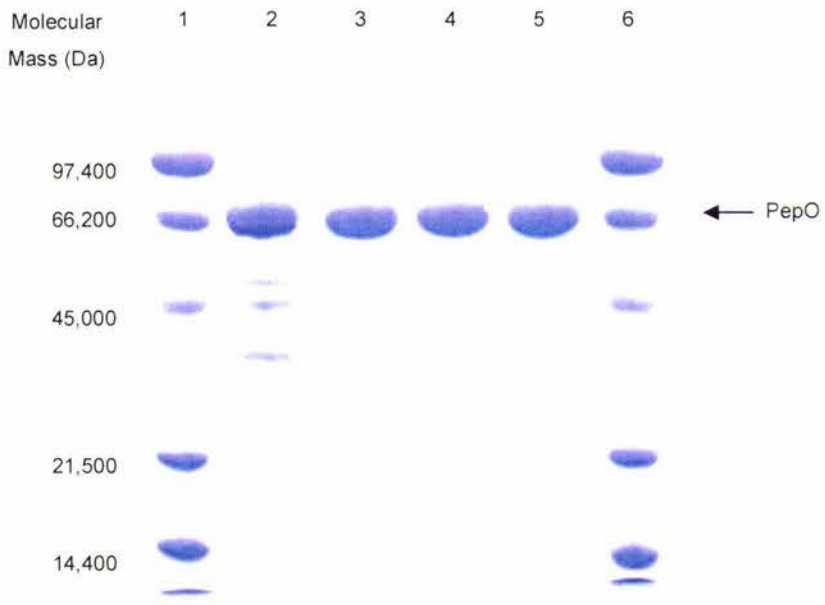


Figure 3.3 SDS-PAGE Analysis of PepO Purification. (Lanes 1 and 6, low-range marker; 2, nisin-induced cell extract; 3 IEX purified; 4, HIC purified; 5 SEC purified PepO).

IEX purified PepO samples displayed more peptidase activity than the crude cell extracts (Figure 3.4).

HIC was selected as the second step in the purification procedure because the separation principle is complementary to ion exchange. HIC was carried out using a phenyl substituted resin (Source-Phe, Amersham) with a lyotrophic salt [$(NH_4)_2SO_4$]. The protein mixture was introduced to the column in 2 M $(NH_4)_2SO_4$. PepO bound strongly to the Source-Phe resin eluting at approximately 0.2 M $(NH_4)_2SO_4$, while most contaminating proteins eluted earlier at

> 1 M $(NH_4)_2SO_4$. The HIC purification step removed all minor contaminants, giving > 99% pure protein, as estimated by SDS-PAGE (Figure 3.3). HIC-purified PepO samples had a higher level of peptidase activity than both the crude cell extracts and IEX purified samples (Figure 3.4). A small percentage of the HIC purified PepO was used for crystallisation trials (Part 2, Chapter 2).

Size exclusion chromatography using a Superdex^{TM}200 resin (Amersham) which gives excellent resolution between 10-200 kDa was carried out as a final purification step to transfer the enzyme into a low ionic strength buffer and to remove minor contaminants (Figure 3.3). PepO eluted from the Superdex^{TM}200 column as a single peak within the range of 68-72 kDa, demonstrating that the enzyme exists as a monomer.

Peptidase activity assays revealed that the SEC purified PepO had a significantly higher level of activity towards as1-casein(1-23) than the cell extract, IEX and HIC purified samples (Figure 3.4). A small percentage of the SEC purified PepO was taken for crystallisation trials (Part 2, Chapter 2).

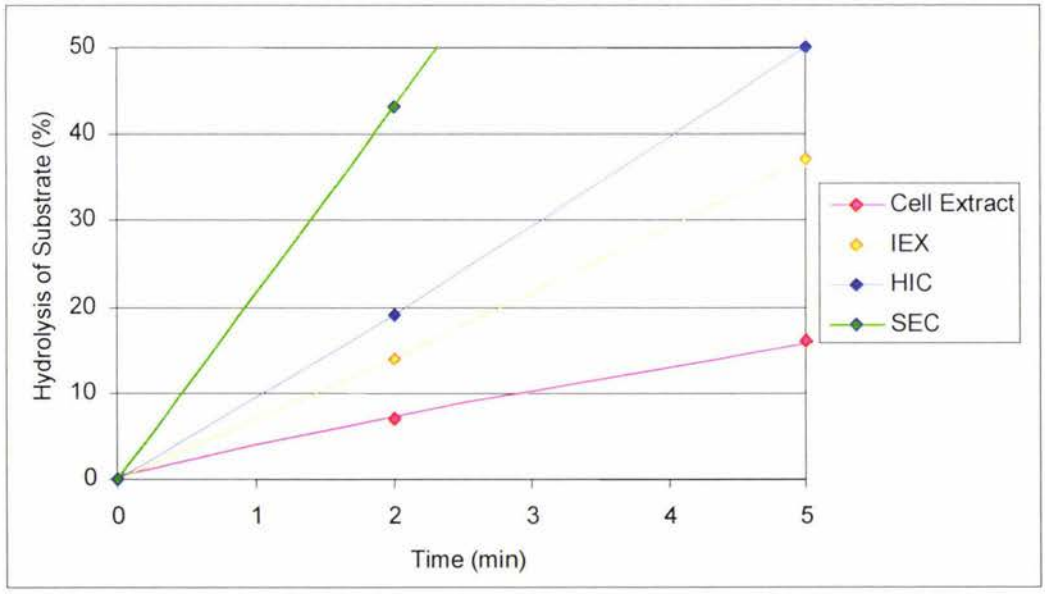


Figure 3.4 PepO Peptidase Activity During Protein Purification. Peptidase activity was monitored by an HPLC-based activity assay that measured the initial rate of hydrolysis of α_{s1}-casein(1-23). (Refer to Table 3.2)

A purification table summarising the purification of *L. rhamnosus* PepO to sufficient purity and quantity for structural and functional investigations is shown below (Table 3.2).

Table 3.2 Purification Table of *L. rhamnosus* PepO.

Purification Step	Total Protein (mg)	Volume (mL)	Total Activity [a] (units)	Specific Activity [b] (units/mg)	Yield [c] (%)	Purification [d] (Fold)
Cell Extract	1880	47	32.8	0.02	100	1
IEX	286	26	46.4	0.16	141	8
HIC	180	24	59.0	0.33	180	2
SEC	38	8	355.4	9.35	1084	6

[a] Total Activity, 1 unit of PepO activity is defined as 1 pM of substrate hydrolysed per minute at room temperature and pH 6.5.

[b] Specific Activity was calculated by dividing the total activity (units) by the total protein (mg).

[c] Yield, expressed as a percentage, was calculated by dividing the total activity of each purification step by the total activity of the first purification step.

[d] Purification, expressed in fold, was calculated by dividing the specific activity of each purification step by the specific activity of the previous purification step.

Interpretation of the PepO purification table (Table 3.2) is difficult because the total activity and the percentage yield both increased as the purification proceeded. For most enzyme purifications, both of these parameters decrease as the purification proceeds, due to the inevitable loss of enzyme that occurs at each step. The total activity of the cell extract, IEX and HIC protein samples, was relatively low compared to that of the SEC protein sample. This was surprising because *L. lactis* has a number of enodogenous peptidases (including *L. lactis* PepO) that could potentially hydrolyse the α_{s1}-casein(1-23) substrate. The low total activity in the initial purification steps could possibly result from the presence of an inhibitor in the medium, substrate competition or the absence or competition for metal ions. PepO is a metallopeptidase (Chapter 5) that requires a metal ion for activity. The absence of this metal ion during the initial purification steps could result in the unexpectedly low total activity, however, metal ions were not added at any stage during the purification. It is unlikely that PepO activity was suppressed by the ionic strength of the buffer as the buffers contained a similar ionic strength following IEX (0.05 M NaCl), HIC (0.2 M $(NH_4)_2SO_4$) and SEC (0.15 M NaCl). It was beyond the scope of this project to investigate this further.

3.3.1 *L. RHAMNOSUS* PEPW EXPRESSION AND PURIFICATION

Recombinant *L. rhamnosus* PepW was expressed in an *E. coli* IPTG-inducible expression system as described in Chapter 2. IPTG induces expression by displacing the *lac* repressor from the *lacUV5* promoter upstream of the gene encoding the T7 RNA polymerase. The T7 RNA polymerase binds to the T7 promoter upstream of the *pepW* gene and initiates transcription (Figure 3.5).

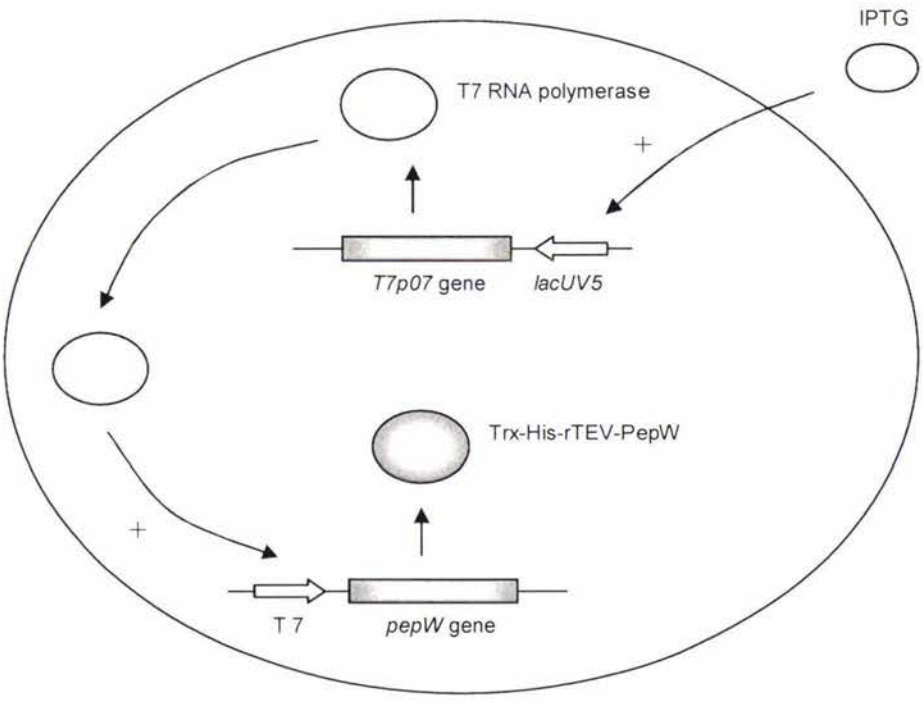


Figure 3.5 *E. coli* IPTG-Inducible Expression System for *L. rhamnosus* PepW. (lacUV5, lacUV5 promoter; T7, T7 promoter; T7p07, T7 RNA polymerase gene; +, activation).

The *pepW* gene encodes a protein of 448 residues with a theoretical molecular mass of 50.8 kDa. The *trx:his:rtev:pepW* fusion gene however, encodes a protein of 588 residues with a theoretical mass of 66.1 kDa. The *E. coli* IPTG-inducible expression system effectively overproduced the Trx-His-rTEV-PepW fusion protein (Figure 3.6), with approximately 80% of the protein being soluble, as estimated by SDS-PAGE. Above 0.5 mM IPTG, cleavage of the fusion tag was observed in the soluble fractionated samples, although this was not observed in the whole cell extracts (Figure 3.7).

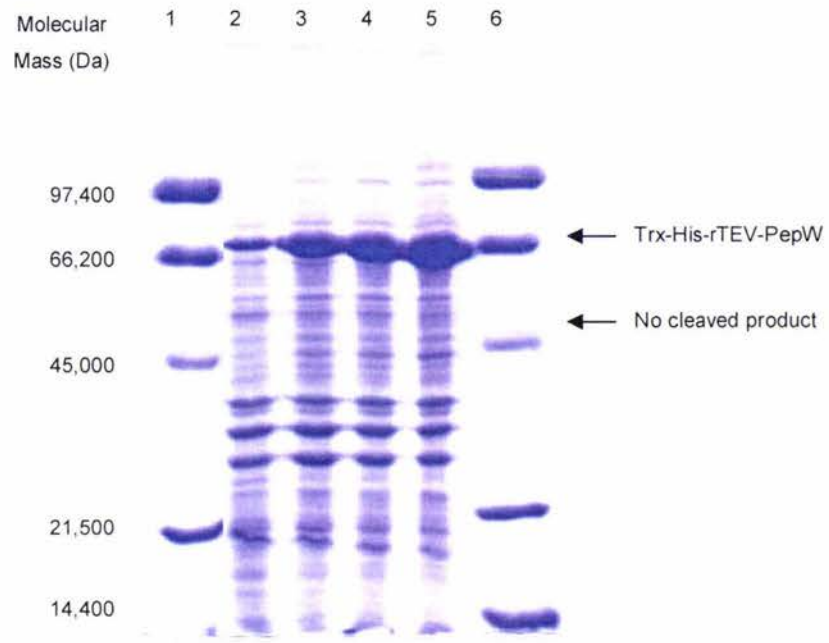


Figure 3.6 SDS-PAGE Analysis of IPTG Induced PepW Expression. (Lanes 1 and 5, low-range marker; 2, uninduced cell extract; 3, 0.25 mM IPTG; 4, 0.5 mM IPTG and; 5, 0.75 mM IPTG induced cell extract).

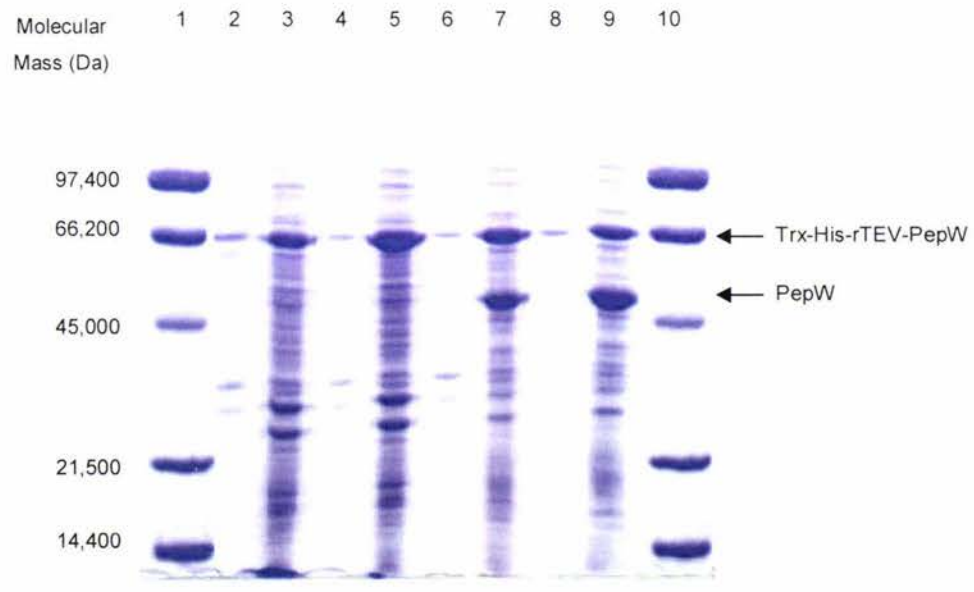


Figure 3.7 SDS-PAGE Analysis of PepW Solubility. (Lanes 1 and 10, low-range marker; 2, pellet (insoluble) and 3, soluble fraction of uninduced sample; 4, insoluble and 5, soluble fraction of 0.25 mM IPTG sample; 6, insoluble and 7, soluble fraction of 0.50 mM IPTG sample; 8, insoluble and; 9, soluble fraction of 0.75 mM sample).

It is difficult to conceive of an explanation as to why autocleavage occurred in the soluble fractionated samples and not the whole cell extracts. Samples for the whole cell extract and solubility analyses were taken from the same sonicated culture. The only difference between the samples was that the soluble fractionated samples were centrifuged (13,000 g for 15 min at 4°C) prior to SDS-PAGE analysis. It is therefore possible that cleavage was brought about by some *E. coli* proteases or that the protein product of the more highly induced samples is not folded properly and is unstable. This was not investigated further.

As cleavage of the fusion tag prior to purification prevents its use in affinity chromatography, 0.25 mM IPTG was used to induce expression of PepW. The first step in the purification of PepW used immobilised metal affinity chromatography (IMAC) exploiting the His-tag. Because there is no established assay for PepW, purification was monitored purely by SDS-PAGE. IMAC purification selectively and efficiently purified the enzyme as shown in Figure 3.8.

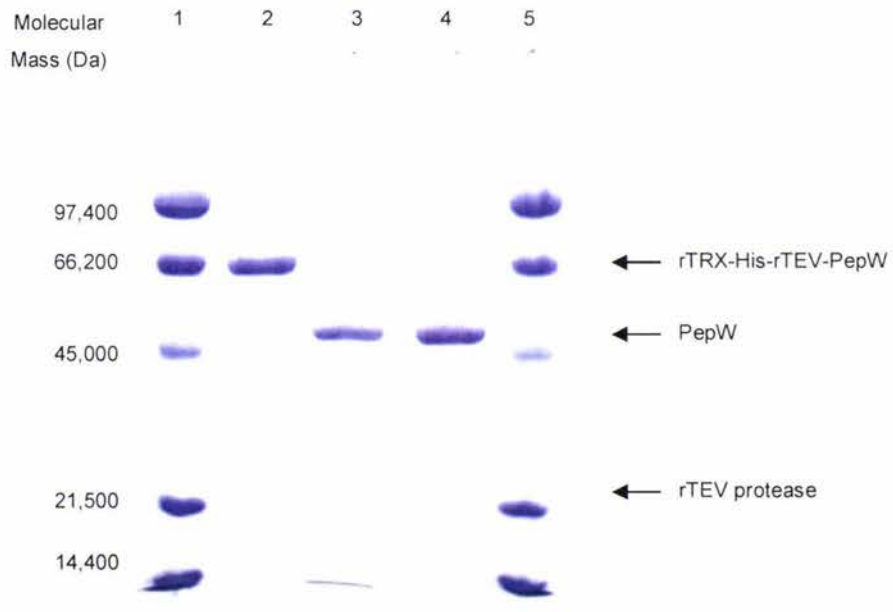


Figure 3.8 SDS-PAGE Analysis of PepW Purification. (Lanes 1 and 5, low-range marker; 2, IMAC purified; 3, rTEV treated; 4 SEC purified PepW).

Removal of the Trx-His-rTEV tag was carried out using rTEV protease followed by size exclusion chromatography (Superdex^{TM}200) (Amersham) to remove the rTEV protease, Trx-His-rTEV tag and other minor contaminants. It also served to transfer the enzyme into a low ionic strength buffer. Approximately 25.0% of PepW eluted from the Superdex^{TM}200 column as a broad peak corresponding to a range of molecular weights from 45.0-55.0 kDa. The remaining

75% eluted in a volume that corresponds to a theoretical molecular weight of 290-310.0 kDa (if the standard curve was extended), demonstrating that the enzyme exists predominantly as a hexamer, a small percentage of PepW exists as a monomer. The *S. cerevisiae* bleomycin hydrolase has also been found to exist as a hexamer and monomer, both of which are active (Zhendong 1996).

Substrate specificity investigations of the *L. rhamnosus* peptidases are discussed in Chapter 5.

PREPARATION OF SUBSTRATES

4.1 INTRODUCTION

An important property of any peptidase is its substrate specificity. Despite all efforts, it is at present not possible to computationally predict substrate specificity for the majority of peptidases. To determine the substrate specificity it is necessary to isolate and purify the peptidase before biochemically characterising individual peptide substrates and products. This Chapter describes the preparation of potential substrates that were used to characterise the substrate specificity of the *L. rhamnosus* peptidases, PepO and PepW, in Chapter 5.

Knowledge of the substrate specificity of PepO is limited and PepW is completely unknown. To investigate the substrate specificity of these enzymes potential substrates that would be available to the enzyme in a starter culture were generated by partially recreating the LAB proteolytic network in vitro (Figure 4.1).

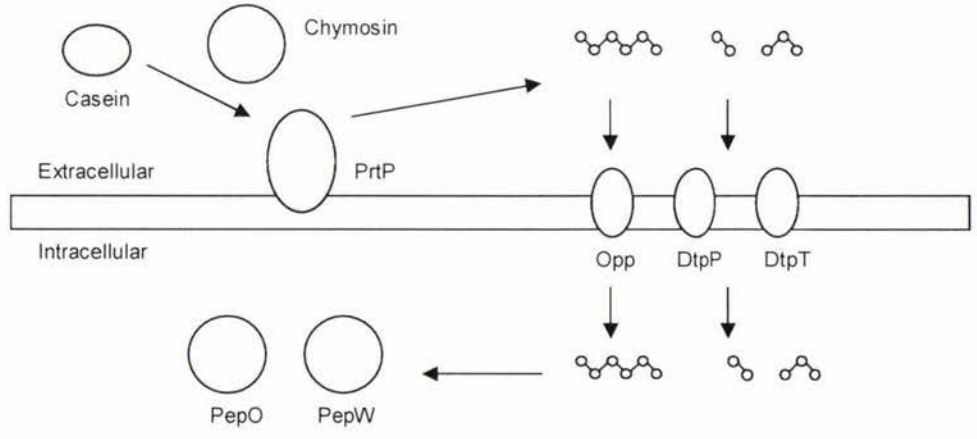


Figure 4.1 The LAB Proteolytic Network. (PrtP, *L. lactis* cell-envelope proteinase; Opp, oligopeptide and; DtpP, DtpT, di- and tripeptide transport systems).

The substrates that both PepO and PepW are most likely to encounter are derived from proteolytic digestion of the bovine casein by chymosin and *L. lactis* cell-envelope proteinases (PrtP). There are four main caseins in bovine milk; α_{s1}-, α_{s2}-, β-, and κ-casein, and together these make up approximately 80% of total bovine milk protein (Fox 2000) (Table 4.1).

Table 4.1 Properties of Bovine Casein (Fox 2000).

	Molecular Weight (kDa)	Proline Residues (per M)	Phosphate Groups (per M)	Total Milk Protein (%)
α_{s1}-casein	23.5	17	8-9	31.0
α_{s2}-casein	25.1	10	10-13	8.5
β-casein	24.0	35	4-5	28.0
κ-casein	19.0	20	1-3	12.5
Total casein	N/A	N/A	N/A	80.0

Bovine caseins contain a high percentage of prolines that prevent the formation of ordered secondary structures. This lack of structural order results in considerable exposure of hydrophobic residues resulting in the formation of casein micelles. The formation of casein micelles is not due solely to the proline content, but also to the high degree of phosphorylation of the proteins. Phosphate groups esterified to serine and threonine residues bind Ca^{2+} and other polyvalent cations, causing charge neutralisation which promotes micelle formation (Fox 2000).

The casein micelle can be destabilised by the proteolytic activity of chymosin a traditional enzyme used in cheesemaking. Chymosin (EC 3.4.23.4) is a 35.6 kDa aspartic peptidase, that primary cleaves the Phe_{105}-Met_{106} bond in κ-casein. Cleavage of this peptide bond has been well characterised and is essential to the coagulation step of cheesemaking. Chymosin cleaves additional peptide bonds (although at a lower rate) within α_{s1}-, β-, and κ-casein. Considerable cleavage variation attributed to temperature, pH and salt concentration has been observed, and no definite rules concerning the cleavage specificity of chymosin have been established. The chymosin cleavage sites and soluble peptides of α_{s1}-, β-, and κ-casein that have been identified are summarised in Table 4.2 (Appendix II) (Visser 1977; McSweeney 1993; Fox 2000).

The cell-envelope proteinases from *L. lactis* subsp. *cremoris* strains H2 and SK11 are PI- and PIII-type proteinases, respectively, and both hydrolyse β- and κ-caseins, although with different cleavage specificities. In contrast, α_{s1}-casesin is hydrolysed efficiently only by PIII-type proteinases (Barrett 2004). Definite rules concerning the cleavage specificity of PI- and PIII-type peptidases have not been established, although some preferences have been observed (Barrett 2004). Both the PI- and PIII-type proteinases have preference for large hydrophobic residues in the P1 and P4 positions, and preference for a Pro in the P2 position. The H2 and SK11 PrtP cleavage sites and peptides of α_{s1}-, β-, and κ-casein that have been identified are summarised in Appendix II (Reid 1991a; Reid 1991b; Reid 1994b).

Table 4.2 Chymosin-derived Soluble Peptides of α_{s1}-, β-, and κ-casein.

	Peptide	Sequence
α_{s1}-casein	1-11	R-P-K-H-P-I-K-H-Q-G-L
	1-23	R-P-K-H-P-I-K-H-Q-G-L-P-Q-E-V-L-N-E-N-L-L-R-F
	24-28	F-V-A-P-F
	24-32	F-V-A-P-F-P-E-V-F
	24-40	F-V-A-P-F-P-E-V-F-G-K-E-K-V-N-E-L
	101-142	K-K-Y-K-V-P-Q-L-E-I-V-P-N-S-A-E-E-R-L-H-S-M-K-E-G-I-H-A-Q-Q-K-E-P-M-I-G-V-N-Q-E-L
	143-149	A-Y-F-Y-P-E-L
	150-153	F-R-Q-F
	150-156	F-R-Q-F-Y-Q-L
	154-159	Y-Q-L-D-A-Y
	154-164	Y-Q-L-D-A-Y-P-S-G-A-W
	157-164	D-A-Y-P-S-G-A-W
	165-179	Y-Y-V-P-L-G-T-Q-Y-T-D-A-P-S-F
	165-199	Y-Y-V-P-L-G-T-Q-Y-T-D-A-P-S-F-S-D-I-P-N-P-I-G-S-E-N-S-E-K-T-T-M-P-L-W
β-casein	192-209	Y-Q-E-P-V-L-G-P-V-R-G-P-F-P-I-I-V

In addition, potential casein substrates were extracted from Cheddar cheese. Proteolysis of Cheddar cheese is well characterised and the peptides isolated and identified are summarised in Appendix II (Alli 1998; Fox 2000). The major peptides are derived from the proteolytic digestion of α_{s1}-casein by chymosin, and β-casein by plasmin and *L. lactis* PrtP enzymes. Plasmin (EC 3.4.21.7) is a 91 kDa indigenous milk serine proteinase that primarily cleaves the Lys_{28}-Lys_{29}, Lys_{105}-His_{106}, and Lys_{107}-Glu_{108} peptide bonds of β-casein (Barrett 2004; Fox 2000).

4.2 METHODS

4.2.1 *L. lactis* Cell-Envelope Peptidase Extraction

L. lactis subsp. *cremoris* strains H2 and SK11 were grown anaerobically at 30°C in reconstituted skim milk (RSM) that had been autoclaved at 10 PSI for 10 min. Immediately prior to inoculation (1% innoculum), sterile 2.5 M Sodium β-glycerophosphate (30 mL per 1 L RSM) was added to buffer against the pH decrease which results from lactic acid production during fermentation. The culture was incubated at 30°C for 12-14 hr at which point the pH decreased to about pH 5.0. The culture was then neutralised by the addition 10 M NaOH and cooled to 5-10°C before the addition of 25% Sodium citrate (60 mL per 1 L culture) to chelate Ca^{2+}. The culture was then incubated at room temperature for 20 min and harvested by centrifugation (12,000 g for 10 min at 4°C). The cells were washed twice with ice-cold 50 mM Sodium phosphate pH 6.4 (250 mL per 1 L culture), and then resuspended in 50 mM sodium phosphate pH 6.4 (25 mL per 1 L culture). The cells were incubated at 30°C for 30 min, centrifuged (12,000 g for 10 min at 4°C) and the supernatant containing proteinase activity pooled, concentrated to approximately 5 mg/mL and stored at -80°C.

4.2.2 *L. lactis* Cell-Envelope Proteinase Activity Assay

L. lactis cell-envelope proteinase activity was measured using a previously established spectrofluorometer based activity assay with fluorescein isothiocyanate (FITC) labelled β-casein as the substrate (Reid 1994b). Typically, 100 μL of FITC labelled β-casein (1% protein) was incubated with 100 μL enzyme sample with 100 μL of 0.5 M Sodium phosphate pH 6.4 at 25°C for 0-60 min. 60 μL aliquots were taken at 0 and 60 min and added to 150 μL of 5% (w/v) trichloroacetic acid (TCA) to stop proteolysis. The samples were centrifuged (10,000 g for 1 min at 4°C) to remove precipitated protein, 100 μL of the supernatant added to 3 mL of 0.5 M Tris-HCl pH 8.5 and the fluorescence measured using a Perkin Elmer Luminescence Spectrometer LS50B spectrofluoromter (excitation = 490 nm, emission = 525 nm). An increase in fluorescence indicates proteolytic activity because small FITC-labelled peptides are soluble in TCA while FITC-labelled β-casein is not. Controls included a 100 ng/mL FITC standard in 50 mM Tris-HCl pH 7.2, FITC-β-casein without proteolytic enzyme, and FITC-β-casein digested with 1 mg/mL Trypsin.

4.2.3 Preparation of FITC-β-casein

FITC which specifically labels primary amines (lysine side chains and N-terminal amino groups) was conjugated to β-casein by mixing 5 mg of FITC (Sigma) and 200 mg β-casein (Sigma) in 20 mL 150 mM NaCl, 50 mM Sodium carbonate buffer pH 9.5 in the dark overnight. Unreacted

FITC was removed by dialysis in the dark overnight against the following: 2 L 0.5% Charcoal (twice), followed by 2 L 50 mM Tris-HCl pH 8.5 and 2 L 50 mM Tris-HCl pH 7.2. FITC-labelled β-casein was stored in the dark at -20°C.

4.2.4 Preparation of α_{s1}-casein(1-23)

α_{s1}-casein(1-23) was prepared by dissolving 2.5 mg α_{s1}-casein in 100 mL deionised water. The pH was adjusted to pH 5.8 to optimize proteolytic digestion, 300 µL of 20 mg/mL Chymosin was added and the solution was incubated at room temperature for 30 min. The pH was adjusted to pH 9.0 with 10 M NaOH to irreversibly inactivate chymosin and incubated at room temperature for a further 30 min. The pH was then adjusted to pH 7.5 with 1 M HCl and incubated at 50°C for 20 min before 5 mL 10% Calcium chloride was added to precipitate α_{s1}-casein. The sample was centrifuged (13,000 g for 20 min at 18°C) to remove the non-hydrolysed α_{s1}-casein, and the supernatant containing the α_{s1}-casein(1-23) peptide was purified by RP-HPLC using a Jupiter C_{18} prep column (Phenomenex) (10 µ, 10 x 250 mm) equilibrated with 0.1% TFA in deionised water. Elution of the bound α_{s1}-casein(1-23) peptide, which elutes with a retention time of approximately 16.0 minutes, was achieved by applying a linear gradient of 0-100% Acetonitrile (0.08% TFA), and monitored at 214 nm and 280 nm. The α_{s1}-casein(1-23) peptide was manually collected, lyophilised using a Dura-Dry™ µP Freeze-dryer and mass analysis carried out with MALDI-TOF mass spectrometry (Section 5.2.4).

4.2.5 Preparation of Bovine Casein Peptides

Bovine casein peptides were generated by incubating 200 µL 10 mg/mL α_{s1}-, β- or κ-casein in 50 mM Tris-HCl pH 6.5 with 20 µL (20 mg/mL) Chymosin and/ or 20 µL (5 mg/mL) *L. lactis* strain H2 or SK11 PrtP, at room temperature for 20 hr. The reaction was terminated by incubation at 100°C for 15 min. Proteolysis was monitored by RP-HPLC. Aliquots (20 µL) of the inactivated digest were added to 20 µL 2% TFA, and 60 µL 1% TFA added to give a total volume of 100 µL. The samples were centrifuged (10,000 g for 1 min at 4°C) to remove precipitated protein and 80 µL of the supernatant injected onto a Phenomoenex Jupiter C_{18} column (5 µ, 4.6 x 250 mm) equilibrated with 0.1% TFA in deionised water. Elution of the bound peptides was achieved by applying a linear gradient of 0-100% Acetonitrile (0.08% TFA), and monitored at 214 nm and 280 nm.

4.2.6 Preparation of Cheese Extracted Peptides

Casein peptides were extracted from an young Cheddar cheese produced with standard industry LAB starter organisms. 1 g of grated cheese in 1 mL deionised water was incubated at 60°C for 5 min and then homogenised using a high speed blender. The sample was centrifuged

(10,000 g for 15 min at 4°C) and the aqueous solution containing peptides collected. Salt and minor contaminants were removed using Maxi-clean C_{18} cartridges (Alltech), and the solutions were then lyophilised using a Dura-Dry™ μP Freeze-dryer and stored at -20°C. Cheese peptides were resuspended in 50 mM Tris-HCl pH 6.5 to give a final concentration of 10 mg/mL and analysed by RP-HPLC as described in section 4.2.5.

4.3 RESULTS AND DISCUSSION

4.3.1 *L. LACTIS* CELL ENVELOPE PROTEINASE EXTRACTION

The *L. lactis* subsp. *cremoris* strain H2 and SK11 cell-envelope proteinases (PrtP) were purified using a strategy that exploits autoproteolysis of the proteinase from the cell-envelope. LAB cell-envelope proteinases consist of an N-terminal subtilisin-like proteinases domain and a C-terminal domain that extends into the cell-envelope (Barrett 2004). It is thought that the N-terminal domain binds Ca^{2+} ions inducing a conformation that prevents detachment of the enzyme from the cell-envelope. In the absence of Ca^{2+}, a structural rearrangement occurs exposing a sequence which is highly susceptible to autoproteolytic attack (Barrett 2004). Autoproteolysis of the LAB cell-envelope proteinase occurs at multiple sites and generates active proteinase fragments with molecular mass ranging from 80-180 kDa, depending partly on the *L. lactis* strain used (Barrett 2004).

Autoproteolytic release of the *L. lactis* subsp. *cremoris* strain H2 and SK11 cell-envelope proteinases predominantly generated 140 kDa and 180 kDa fragments of the H2 cell-envelope proteinase and a 180 kDa fragment of the SK11 cell-envelope proteinase (Figure 4.2).

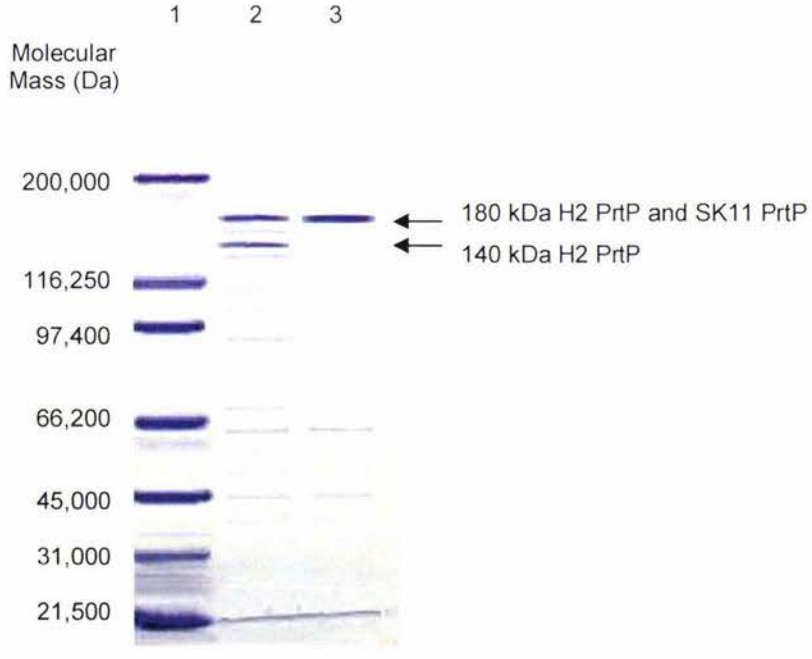


Figure 4.2 SDS-PAGE Analysis of *L. lactis* subsp, *cremoris* Strains H2 and SK11 PrtP Following Autoproteolytic Release from the Cell-Envelope. (1, broad-range marker; 2, H2 PrtP; 3, SK11 PrtP).

Although the 140 kDa and 180 kDa PrtP enzymes were the predominant proteins present in the PrtP solutions, a number of low molecular weight proteins, thought to be intracellular proteins were also present. Further purification was not carried out to remove these proteins as pure PrtP enzymes were not required. It is likely that a number of the intracellular proteins are peptidases and these may contribute to the preparation of casein peptide substrates. Proteolytic activity of the PrtP enzymes was confirmed with a spectrofluorometric assay using FITC labelled β-casein (Figure 4.3).

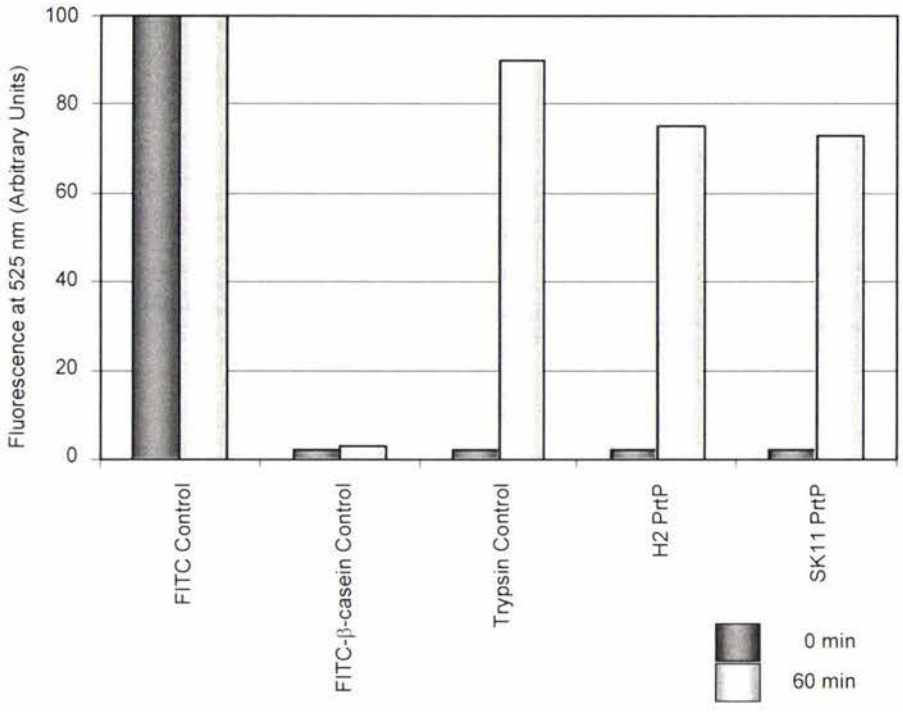


Figure 4.3 Proteolytic Activity of the *L. lactis* subsp. *cremoris* Strains H2 and SK11 PrtP Enzymes.

Figure 4.3 shows the progress of the hydrolysis of FITC labelled β-casein peptides at 0 and 60 minutes. FITC, and FITC-β-casein controls were used to ensure that fluorescence of FITC was and was not detected at 595 nm, respectively, and a Trypsin control used to confirm that the FITC labelled β-casein could be hydrolysed.

Similar proteolytic activities for both H2 and SK11 PrtP enzymes were observed, although the different cleavage specificities of the PrtP enzymes for β-casein made it difficult to compare activities.

4.3.2 BOVINE CASEIN PEPTIDE SUBSTRATES

Potential bovine α_{s1}-, β-, and κ-casein peptide substrates were generated by proteolytic digestion with chymosin, and/or the H2 and SK11 PrtP enzymes. The α_{s1}-casein(1-23) substrate used to monitor PepO purification was generated by chymosin digestion and successfully purified using RP-HPLC chromatography (Appendix III). The α_{s1}-casein(1-23) peptide mass (2764.2 Da) was confirmed by MALDI-TOF mass spectrometry (data not shown).

Potential α_{s1}-, β-, and κ-casein peptide substrates were prepared by proteolytic digestion with either chymosin (Appendix III) or the H2 or the SK11 PrtP enzymes (Appendix III), or the combined action of chymosin and the H2 or SK11 PrtP enzymes (Appendix III). Successful heat inactivation of chymosin and the H2 and SK11 PrtP enzymes was confirmed by repeating substrate preparation digests using heat inactivated enzymes (results not shown).

Proteolytic digestion of α_{s1}-, β-, and κ-casein by chymosin and/or H2 and SK11 PrtP enzymes generated complex mixtures of casein peptides. The peptides were separated by RP-HPLC using 0-75% Acetonitrile/0.08% TFA gradient over a period of 25 min. The chromatogram was very crowded and several attempts were made to improve the resolution of the peaks. This proved to be impossible without further fractionation of the peptide mixtures. Although the vast number of peptides in the mixtures made it difficult to separate individual peptides, the separation achieved allowed identification of at least some of the peptides using Edman sequencing, MS, MSMS and LC-MSMS techniques (Chapter 5)

4.3.3 CHEESE EXTRACTED PEPTIDE SUBSTRATES

Casein peptides were extracted from Cheddar cheese using a simple centrifugation procedure and analysed by RP-HPLC chromatography (Appendix III).

Casein peptides, either extracted from Cheddar cheese, or generated by the proteolytic digestion of chymosin and the *L. lactis* strain H2 and SK11 PrtP enzymes, were then used to investigate the substrate specificity of the *L. rhamnosus* peptidases, PepO and PepW (Chapter 5).

SUBSTRATE SPECIFICITY

5.1 INTRODUCTION

In this Chapter the substrate specificity of the *L. rhamnosus* peptidases, PepO and PepW, will be discussed.

PepO is a putative metallo-endopeptidase that is already known to have a unique cleavage specificity for the peptide α_{s1}-casein(1-23) (Figure 5.1). Nothing is known, however, about its ability to cleave other casein peptides.

	5 6 7 8 9 10
α_{s1}-casein(1-23)	R - P - K - H - P - I - K - H - Q - G - L - P - Q - E - V - L - N - E - N - L - L - R - F
	↑ ↑ ↑ ↑

Figure 5.1 *L. rhamnosus* PepO has Unique Cleavage Specificity for α_{s1}-casein(1-23). (↑, cleavage sites of *L. rhamnosus* PepO).

PepW is a putative cysteine aminopeptidase that is completely uncharacterised. To establish the peptide bond specificity of PepO and PepW it was necessary to investigate the substrates hydrolysed and the products produced by these enzymes.

To investigate the substrate specificity of a peptidase, it is essential to isolate the peptidase from all contaminating peptidolytic enzymes, since even traces of these enzymes can interfere with substrate specificity analysis. Peptide substrates often contain multiple peptide bonds that could be potentially cleaved by the peptidase. As hydrolysis of a peptide substrate proceeds, product peptides are generated with similar general properties to the substrate from which they were derived. These product peptides become new substrates for the peptidase, often making it difficult to establish which products correspond to which initial substrate.

The most commonly used technique to investigate peptidase substrate specificity involves an HPLC-based separation of the substrate and product peptides followed by identification of their amino acid sequences. Amino acid sequence analysis is commonly carried out using a combination of techniques including: Edman N-terminal Sequencing; Mass Spectrometry; and Tandem Mass Spectrometry.

5.1.1 EDMAN N-TERMINAL SEQUENCING

Edman N-terminal sequencing uses the Edman reaction to systematically cleave and identify amino acids from the N-terminus of a protein or peptide. In the Edman reaction, phenylisothiocyanate (PITC) reacts with the N-terminus of the peptide to form a cyclic intermediate that facilitates selective hydrolysis of the adjacent amide bond (Edman 1950; Kinter 2000). The cleaved amino acid is converted to a phenylthiohydantoin (PTH) derivative for analysis, while the peptide (shortened by one amino acid) is subjected to another cycle for the Edman reaction (Kinter 2000). At each cycle of the Edman reaction, the cleaved PTH-amino acid derivative is separated by RP-HPLC and identified by comparison with PTH-amino acid standards.

Edman N-terminal sequencing provides an accurate analytical method for sequencing peptides, however, there are a number of limitations that make this method unsuitable for high throughput work. The most significant limitation is the time required to complete each cycle. Each cycle takes approximately 45 min (instrument dependent), and because it is often necessary to identify at least five amino acids for each peptide (in combination with the peptide's mass) to accurately identify the peptide sequence, no more than two to three samples can be analysed per day. Furthermore, very pure samples are required because the N-terminal amino acid of all peptides present in a sample will give rise to a PTH-amino acid, which can create serious problems when assigning which PTH-amino acid corresponds to which peptide. Also the chemistry of the Edman reaction prevents the identification of unmodified cysteine and many modified residues and requires that the N-termini of the peptides are not blocked.

5.1.2 MASS-SPECTROMETRY

Mass spectrometry (MS) is a technique that separates ions on the basis of mass, and can provide both mass and sequence information for a molecule. MS typically involves 3 steps: (i) ionisation of the molecule; (ii) separation of the ions on the basis of their mass-to-charge ratio; and (iii) detection of the separated ions.

The most commonly used ionisation methods are Atmospheric Pressure Ionisation (API) and Matrix Assisted Laser Desorption/Ionisation (MALDI). API allows the continuous production of multiply charged ions and the analysis of high molecular weight compounds using mass analysers of limited mass range. API produces small positively charged droplets by pneumatic nebulization (i.e. the forcing of the solution through a needle to which a high positive voltage is applied while also applying a continuous axial flow of nebulising gas). The positively charged droplets move towards the negatively charged (relative to the needle) instrument, during which time the solvent of the droplet evaporates, Coulombic repulsion occurs and the droplet splits

into a population of smaller, charged droplets. This process continues until de-solvated gas phase ions are produced.

MALDI allows the pulsed production of predominantly singly-charged ions by protonation. The MALDI method requires that the molecule is dissolved in a solution of a UV-absorbing matrix compound on a stainless steel probe. The matrix solution protonates the molecule and upon drying crystallises incorporating the target molecule in the crystal structure. Pulses of UV laser light are used to vaporize the crystals and analyte ions into the gas phase.

Following ionisation, the ions are separated by a mass-analyser according to their mass-to-charge ratio. The most commonly used mass-analysers include Time-Of-Flight (TOF), and Quadrupole mass analysers. Mass-analysers can be used in tandem and this is referred to as Tandem Mass Spectrometry (MSMS) (Section 5.1.3).

In a TOF mass analyser the ions are introduced as a 'packet' of ions resulting in all of the ions receiving the same initial kinetic energy. The ions then enter a field-free region where they travel at velocities that are inversely proportional to their mass-to-charge (m/z) ratio. Because of this inverse relationship, ions with low mass-to-charge ratios travel more rapidly than ions with high mass-to-charge ratios (Kinter 2000). This enables the detector to record all ions as they arrive at the detector, and allows the calculation of their mass-to-charge ratios based on the time of their arrival.

In a Quadrupole mass analyser, electric fields are used to separate ions according to their m/z value. A quadrupole consists of four parallel rods or poles through which the ions being separated are passed. By varying the strength and frequencies of the electric fields applied to the poles, it is possible to induce a stable trajectory for an ion with any given m/z value, thus causing it to pass through the path of the mass analyser to the detector while all other ion are not detected (Kinter 2000). By incrementally changing the field it is possible to detect all ions in a sample.

The data output of mass spectrometry is a plot of the relative intensity against the mass-to-charge ratio of individual ions. The mass (M) of the ion is given by:

$$M = [nH \times m/z] - nH$$

Where *m/z* is the mass-to-charge ratio of an ion, *n* is the number of protons attached to the ion, and H is the mass of a proton (1.00794 Da). The calculated mass can then be used to identify the molecule.

5.1.3 TANDEM MASS SPECTROMETRY

Tandem Mass Spectrometry (MSMS) commonly involves the use of two or more mass analysers in a series separated by a collision cell. The first mass analyser is used to select an ion (parent ion) with a certain mass-to-charge ratio, which is subjected to fragmentation by collision with an inert gas in the collision cell. The second mass analyser is then used to separate the fragment ions according to their mass-to-charge ratios. The fragment ions (daughter ions) can then used for structural or sequence analysis.

Although peptide ions can be fragmented at any bond in the peptide, fractionation most commonly occurs at the peptide bond. When a singly charged peptide is fragmented at a single peptide bond, two fragments are formed: if the charge is retained on the N-terminal fragment of the peptide ion it is called a b-ion; if the charge is retained on the C-terminal fragment it is called a y-ion. In this case, only the fragment retaining the charge is detected as the other fragment is neutral and thus impossible to detect (Kinter 2000) (Figure 5.2).

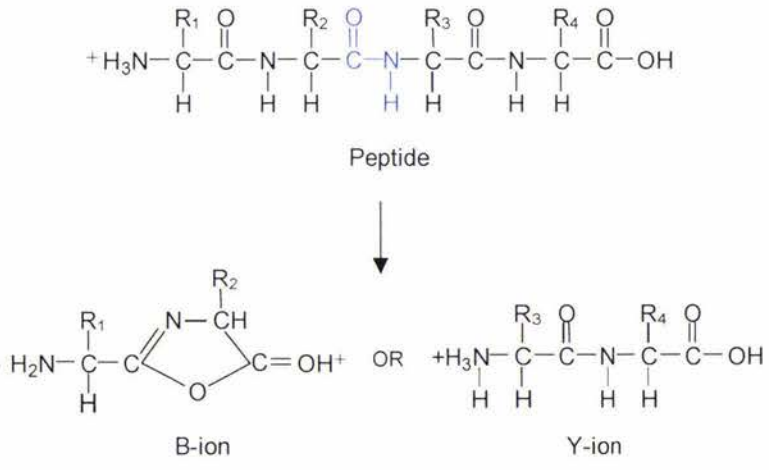


Figure 5.2 Formation of the B-ion and Y-ion Series from Fragmentation of a Singly-Charged Precursor (Kinter 2000).

The mass difference between adjacent members of either the b-ion or y-ion series can be calculated to determine the amino acid sequence of the peptide. Upon peptide fragmentation each amino acid residue has produces a characteristic immonium ion, with the exception of leucine and isoleucine, and lysine and glutamine, which produce immonium ions with the same mass-to-charge ratio. The immonium ions are useful for detecting and confirming the presence of many of the amino acid residues in a peptide, although no information regarding the position of these amino acid residues in the peptide sequence can be ascertained from the immonium ion (Kinter 2000).

5.2 METHODS

5.2.1 Peptidase Substrate Specificity Assay

Peptidase activity was detected using a discontinuous High-Pressure Liquid Chromatography (HPLC)-based Peptidase Activity Assay (Section 3.2.11). Typically, 100 μL of either 0.65 mg/mL α_{s1}-casein(1-23) or 10 mg/mL bovine casein peptides in 50 mM Tris-HCl pH 6.5 (Section 4.2.5) were incubated with 40 μL enzyme sample at room temperature for 0 and 2 hr. To investigate the peptidase catalytic mechanism, enzyme samples were preincubated for 1 hr with various inhibitors: 100 μM Aprotinin, 100 μM Leupeptin, 10 μM Pepstatin, 10 mM EDTA, 100 μM Phosphoramidon. Peptidase activity was observed as the disappearance of substrate peaks and appearance of product peaks. Individual substrate and product peaks were manually collected, concentrated by lyophilisation using a Dura-Dry™ μP Freeze-dryer and analysed by Edman N-terminal Sequencing and Mass Spectrometry.

5.2.2 Edman N-terminal Sequencing

Individual substrate and product peptides from the peptidase activity assays were subjected to Edman N-terminal sequencing on an Applied Biosystems 476A (Perkin Elmer) automated protein sequencer using Model 610A data analysis software.

5.2.3 Matrix-Assisted Laser Desorption/Ionisation Time Of Flight

Individual substrate and product peaks from the peptidase activity assays were subjected to Matrix-Assisted Laser Desorption/Ionisation Time Of Flight (MALDI-TOF) analyses on a Micromass M@LDI™ instrument using MassLynx version 3.5 control and processing software with MaxSpec™ automated laser system tuning and real time data selection. Concentrated protein samples (1 μL) were mixed with 1 μL 40 mg/mL α-cyano-4-hydroxy-trans-cinnamic acid (HCCA), 10 mg/mL nitrocellulose matrix in 50% acetone and 50% Isopropanol, on a stainless steel probe according to the rapid evaporation method (Jimenez 1998). The MALDI-TOF instrument was run in positive-ion, reflectron mode, using a nitrogen UV laser (λ = 337 nm) and a micro-channel plate (MCP) detector. MALDI-TOF was carried out using an operating voltage of 20 kV and laser pulse energy of 60 V. An external mass spectrum calibration was performed using a 1 pM Angiotensin I (Monoisotopic mass, 1296.69 Da) standard.

5.2.4 Atmospheric Pressure Ionisation Tandem Mass Spectroscopy

Atmospheric Pressure Ionisation Tandem Mass Spectroscopy (API-MSMS) was used to obtain peptide mass and amino acid sequence information for individual substrate and product

peptides. The lyophilised peptides were reconstituted in a 50% acetonitrile, 1% formic acid solution in ultra-pure bottled water, before analysis by API-MSMS. The peptide sample was introduced into the mass spectrometer through a fused silica capillary tube (50 μm inner diameter), using a syringe pump (Harvard Instruments) at a rate of 5.0 μL/min. API-MSMS was carried out in positive ionization mode on an Applied BioSystems API 300 triple quadrupole mass spectrometer using an Electro-Spray Ionisation (ESI) source and Analyst 4.0 instrument control software. The instrument was calibrated using a polypropylene glycol (PPG) standard (Applied Biosystems) and the conditions for ionization optimized. Parameter setting were as follows: ion spray voltage, 4500.0 kV; sampling orifice voltage, 30.0 V; focusing ring voltage, 150.0 V. Ion fragmentation was achieved by using collision-induced dissociation (CID) voltages that were m/z dependent. The MSMS data were analysed using Mascott software (Perkins 1999).

5.2.5 Liquid Chromatography Tandem Mass Spectroscopy

Liquid Chromatography Tandem Mass Spectroscopy (LC-MSMS) was carried out by Dr Julian Reid and Mr Aveenash Bala (Fonterra Research Centre, Palmerston North, New Zealand). This was used to obtain peptide mass and amino acid sequence information for substrate and product peptide mixtures (Section 5.2.1). The substrate and product peptide mixtures (0.2-1.0 μL) were loaded onto a PepMap C_{18} column (3 μm, 100 A pore size, 75 μm x 150 mm dimensions) (LC Packings) on an Ultimate nano HPLC system (LC Packings) and eluted using a gradient of 0-60% acetonitrile/10.2% formic acid. The eluted peptides were sprayed into the mass spectrometer using PicoTip™ Emitter 20 μm diameter (10 μm emitter) glass tips and MSMS carried out over a total time of 65 minute per sample on a Q-Star XL hybrid mass spectrometer (QqTOF configuration) (Applied Biosystems). A 3-part Information Dependent Acquisition (IDA) method was used: (1) TOF-MS survey scan between 50-2000 AMU with a cumulative time of 1 second; (2) MSMS product ion scan of the first most intense ion between 50-2000 AMU with a cumulative time of 3 seconds and; (3) MSMS product ion scan of the second most intense ion between 50-2000 AMU with a cumulative time of 3 seconds. The switch criteria between part (1), (2) and (3), was a mass-to-charge ratio of 380.0, which exceeds 20 counts/second. Parameter settings were as follows: ion source gas 1, 5.0; curtain gas, 12.0; ion spray voltage, 1900.0 V, declustering potential, 60.0V; declustering 2 potential, 15.0 V; focussing potential, 150.0 V; collision energy, rolling; collision gas, 6.0; ion release delay, 6.0 V; ion release width, 5.0 V. The IDA advanced parameter was set to ignore peaks within a 3 AMU window of selected ions.

5.3 RESULTS AND DISCUSSION

5.3.1 *L. RHAMNOSUS* PEPO SUBSTRATE SPECIFICITY

5.3.1.1 RP-HPLC ANALYSIS

PepO activity was not inhibited by the aspartic, cysteine and serine inhibitors pepstatin, leupeptin, and aprotinin, respectively (Figure 5.3). However, it was strongly inhibited by EDTA a metal chelator, and weakly inhibited by phosphoramidon, a metabolite produced by *Streptomyces tanashiensis*, which has been found to specifically inhibit the closely related metallopeptidase NEP from *H. sapiens*. PepO is therefore classified as a metallopeptidase.

This finding is consistent with amino acid sequence analysis of the enzyme. The metallopeptidases share a conserved HEXXH pentapeptide motif that has been shown in crystallographic studies to coordinate the metal ion at the active centre. The HEXXH pentapeptide motif is relatively common, but can be more stringently defined for metallopeptidases as the Jongeneel consensus sequence abXHEbbHbc, where 'a' is most often valine or threonine, 'b' is an uncharged residue, and 'c' a hydrophobic residue. (Jongeneel 1989; Rawlings 1995). The metallopeptidases contain a divalent cation, usually zinc, but sometimes cobalt, manganese, nickel or copper, at the active site that activates the water molecule, making it a better nucleophile. Establishing the nature of the metal ion at PepO's active site was beyond the scope of this project.

Although it has been established that PepO is an endopeptidase, knowledge of the enzyme's substrate specificity is limited. To investigate this, various bovine casein peptides were prepared (Chapter 4) and the cleavage of these peptides by PepO analysed using RP-HPLC. PepO was found to be unable to cleave peptide bonds of intact α_{s1}-, β- or κ-casein (data not shown), but was found to cleave specific peptide bonds in the chymosin-derived α_{s1}-casein(1-23) peptide, β- and κ-casein peptides (Appendix IV). Minor hydrolysis by PepO was observed for β- and κ-casein peptides derived from proteolytic digestion by either *L. lactis* subsp. *cremoris* strain H2 or SK11 cell-envelope proteinases (PrtP), and by the combined proteolytic activity of chymosin and the *L. lactis* cell-envelope proteinases (Appendix IV). PepO was also found to cleave specific peptide bonds in casein peptides extracted from Cheddar Cheese (Appendix IV). The focus of this research was directed at the major β-casein, κ-casein and cheese substrate peptides only.

From the RP-HPLC analysis alone it is not possible to establish whether PepO is acting as an endopeptidase or an exopeptidase. To investigate this further, the substrate and product peptides isolated by RP-HPLC were analysed by a combination of mass spectrometry and

sequencing techniques including: Edman N-terminal sequencing, Matrix Assisted Laser Desorption Ionisation Mass Spectrometry (MALDI-TOF MS), Electro Spray Ionization Tandem Mass Spectrometry (ESI-MSMS) and Liquid Chromatography (LC) Tandem Mass Spectrometry (MSMS) using either infusion or LC for sample introduction.

5.3.1.2 EDMAN N-TERMINAL SEQUENCING ANALYSIS

Edman N-terminal sequencing was carried out for the major PepO substrate and product peptides from β-casein, κ-casein and Cheddar cheese isolated by RP-HPLC (Figures 5.3, 5.4 and 5.5). The α_{s1}-casein substrate and products identified by RP-HPLC were not further investigated as the substrate specificity for the substrate peptide, α_{s1}-casein(1-23), has previously been characterised (Christensson 2002). The peptide sequences obtained for the major PepO substrate and product peptides are presented in Tables 5.1, 5.2 and 5.3.

Several of the peptides could not be identified by Edman sequencing and this is likely to be due to low peptide concentration. It is difficult to determine the peptide concentration from the absorbance measured at either 214 nm or 280 nm as the tryptophan and tyrosine content of each residue is unknown. Although the concentration of lyophilized peptides may be determined by weight this is problematic when working with small volumes that may contain anywhere from 10% to as much as 70% bound water by weight. Blocked N-termini are relatively rare, and as these peptides were generated by hydrolysis, extremely unlikely.

The sequences of the peptides that were identified by Edman sequencing were unambiguous and provided accurate analysis of the peptide. The drawback of Edman sequencing is however, that it is often not possible to identify the C-terminal residue of the peptide. To further characterise the substrate and product peptides accurate masses of the peptides were required.

5.3.1.3 MALDI-TOF MS ANALYSIS

Matrix Assisted Laser Desorption Ionisation Time-Of-Flight (MALDI-TOF) MS analysis alone is often not sufficient to conclusively identify peptides. However when combined with sequence data the mass determined by MALDI-TOF analysis provides convincing evidence of a peptides identity. MALDI-TOF was used to ascertain the masses of peptides identified by Edman sequencing and based on the known sequence of the parent protein, it was then possible to deduce complete peptide sequence.

MALDI-TOF analysis proved to be relatively problematic and provided only a few reliable peptide masses. In some cases no obvious masses were present in the spectra. While in other cases clear mass spectra were obtained, however the masses did not correspond to any

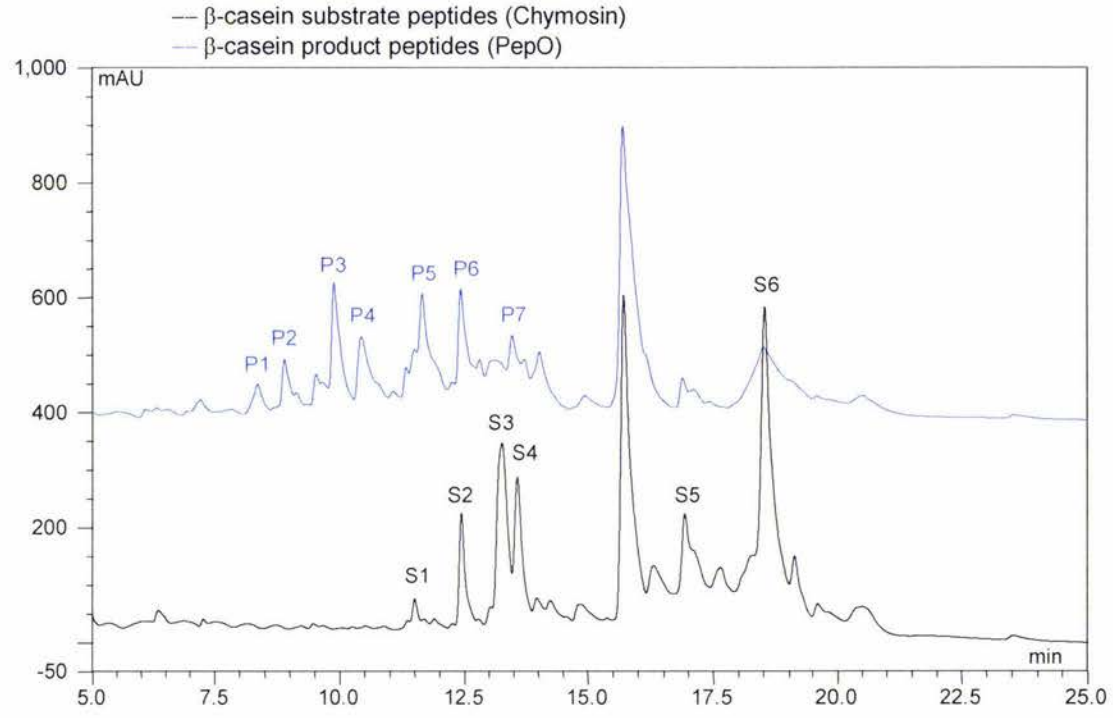


Figure 5.3 RP-HPLC Analysis of PepO Substrate (Chymosin-Generated β-casein Peptides) and Product Peptides. Peaks labelled S1-S6 (Substrate peptides) and P1-P7 (Product peptides) are referred to in Table 5.1.

Table 5.1 Mass and Sequence Analysis of PepO Substrate and Product Peptides from β-casein.

RP-HPLC Peak	MALDI-TOF [a] (m/z)	API-MS [a] (m/z)	API-MSMS [b]	Edman Sequence	Peptide [c]
S1	N/D	N/D	x	S-Q-S-K-V-L-P-V-	β-casein(166-?)
S2	N/D	N/D	x	N/D	N/D
S3	N/D	N/D	x	R-E-L-E-E-L-N-	β-casein(1-?)
S4	N/D	N/D	x	M-H-Q-P-H-Q-P-L-	β-casein(144-?)
S5	N/D	N/D	x	R-E-L-E-E-L-N-V-	β-casein(1-?)
S6	N/D	N/D	x	N/D	N/D
P1	N/D	N/D	x	N/D	N/D
P2	N/D	N/D	x	Y-Q-E-	β-casein(193-?)
P3	N/D	N/D	x	K-A-V-P-Y-P-Q-	β-casein(176-?)
P4	N/D	N/D	x	M-H-Q-P-H-Q-P-	β-casein(144-?)
P5	N/D	N/D	x	R-D-M-P-I-Q-A-	β-casein(183-?)
P6	N/D	N/D	x	V-M-F-P-P-Q-	β-casein(155-?)
P7	N/D	N/D	x	N/D	N/D

[a] Mass-to-charge (m/z) ratio, measured in Thompson (Th).

[b] Tandem Mass Spectrometry data used to identify peptide; ✓, yes; x, no.

[c] ?, C-terminus was not determined.

N/D, Not Determined.

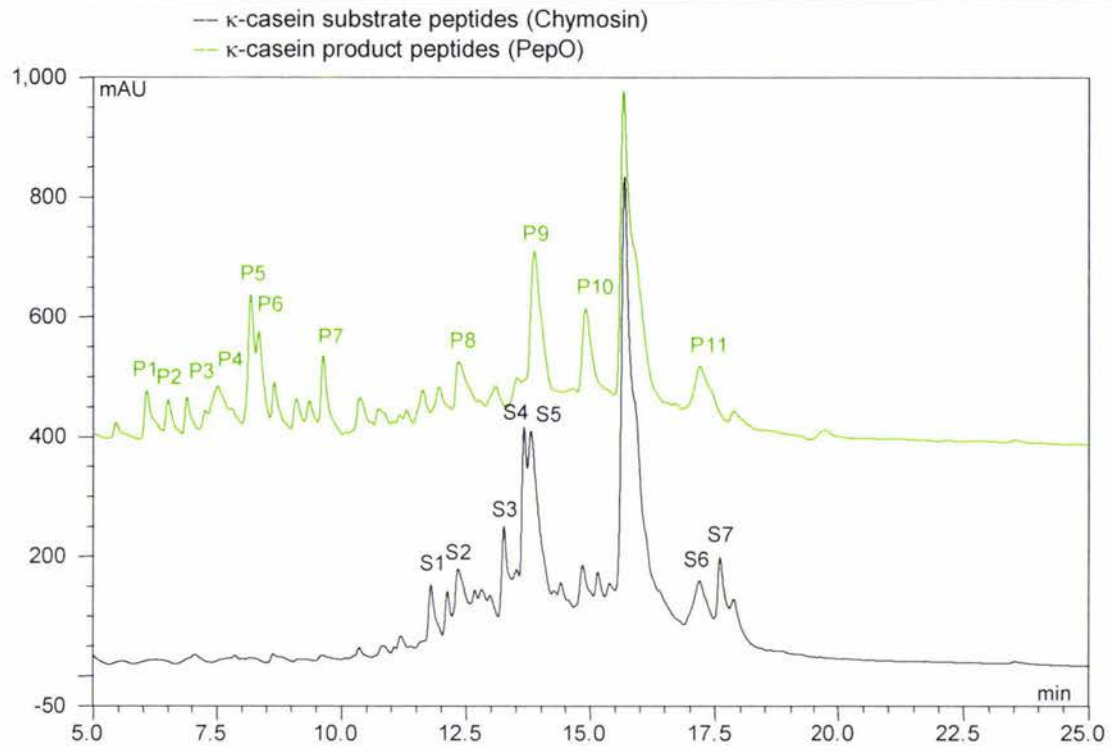


Figure 5.4 RP-HPLC Analysis of PepO Substrate (Chymosin-Generated κ-casein Peptides) and Product Peptides. Peaks labelled S1-S7 (Substrate peptides) and P1-P11 (Product peptides) are referred to in Table 5.2.

Table 5.2 Mass and Sequence Analysis of PepO Substrate and Product Peptides from κ-casein.

RP-HPLC Peak	MALDI-TOF [a] (m/z)	API-MS [a] (m/z)	API-MSMS [b]	Edman Sequence	Peptide [c]
S1	N/D	1056.8	✓	S-R-Y-P-S-Y-G-L-	κ-casein(33-41)
S2	N/D	N/D	x	N/D	N/D
S3	N/D	N/D	x	N/D	N/D
S4	N/D	400.5	✓	A-R-H-P-H-P-	κ-casein(96-105)
S5	N/D	N/D	x	N/D	N/D
S6	N/D	N/D	x	N/D	N/D
S7	N/D	N/D	x	N/D	N/D
P1	N/D	N/D	x	N/D	N/D
P2	N/D	457.4	✓	N/D	κ-casein (62-66)
P3	511.4	N/D	x	N/D	N/D
P4	657.3	657.3	✓	N/D	κ-casein(106-111)
P5	N/D	N/D	x	N/D	N/D
P6	N/D	714.7	✓	N/D	κ-casein(96-101)
P7	N/D	N/D	x	N/D	N/D
P8	N/D	N/D	x	N/D	N/D
P9	N/D	796.7	✓	N/D	κ-casein(25-30)
P10	N/D	636.5	✓	N/D	κ-casein(55-59)
P11	N/D	N/D	x	N/D	N/D

[a] Mass-to-charge (m/z) ratio, measured in Thompson (Th).

[b] Tandem Mass Spectrometry data used to identify peptide; ✓, yes; x, no.

[c] ?, C-terminus was not determined.

N/D, Not Determined.

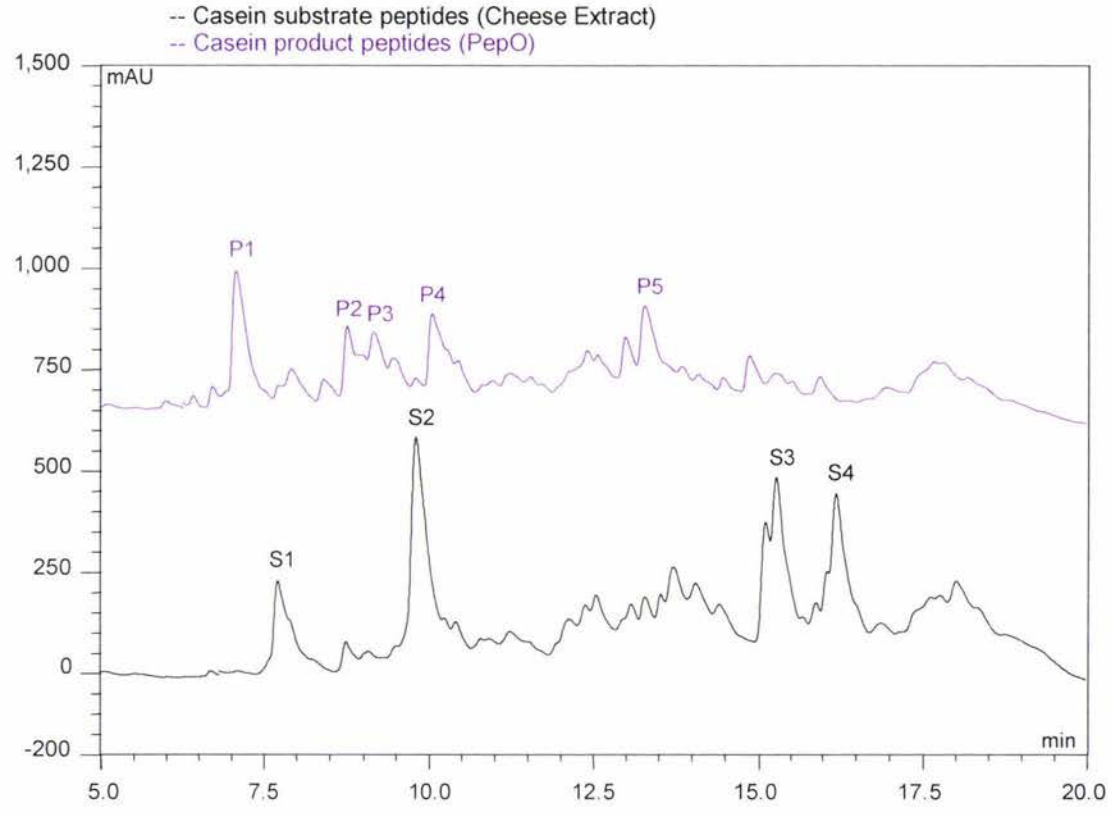


Figure 5.5 RP-HPLC Analysis of PepO Cheddar Cheese Extracted Casein Substrate and Product Peptides. Peaks labelled S1-S4 (Substrate peptides) and P1-P5 (Product peptides) are referred to in Table 5.3.

Table 5.3 Mass and Sequence Analysis of PepO Cheddar Cheese Casein Substrate and Product Peptides.

RP-HPLC Peak	MALDI-TOF [a] (m/z)	API-MS [a] (m/z)	API-MSMS [b]	Edman Sequence	Peptide [c]
S1	1141.1	N/D	x	R-P-K-H-P-I-	α_{s1}-casein(1-9)
S2	N/D	N/D	x	N/D	N/D
S3	N/D	791.7	✓	N/D	α_{s1}-casein(18-23)
	N/D	905.6	✓	N/D	α_{s1}-casein(17-23)
S4	N/D	N/D	x	N/D	N/D
P1	N/D	634.3	✓	N/D	α_{s1}-casein(1-5)
P2	N/D	414.4	✓	N/D	α_{s1}-casein(10-13)
	N/D	474.3	✓	N/D	α_{s1}-casein(15-18)
	N/D	486.5	✓	N/D	α_{s1}-casein(11-14)
	N/D	717.5	✓	N/D	α_{s1}-casein(14-19)
P3	N/D	542.7	✓	N/D	α_{s1}-casein(9-13)
	N/D	679.2	✓	N/D	α_{s1}-casein(8-13)
P4	N/D	594.5	✓	N/D	β-casein(63-68)
P5	N/D	788.5	✓	N/D	β-casein(1-6)
	N/D	838.5	✓	N/D	β-casein(61-68)

[a] Mass-to-charge (m/z) ratio, measured in Thompson (Th).

[b] Tandem Mass Spectrometry data used to identify peptide; ✓, yes; x, no.

[c] ?, C-terminus was not determined.

N/D, Not Determined.

possible casein peptides. Careful analysis of these masses was carried out, taking into consideration possible adducts that could result in the observed mass.

An α-cyano-4-hydroxy-trans-cinnamic acid (α-HCCA) matrix was used and it is known that sodium and potassium ions are able to induce the formation of α-HCCA adducts. The most abundant α-HCCA matrix adducts are typically in the range of m/z 800–1100 Da, which is part of the mass range used for peptide identification (Zhu 2003). Although these α-HCCA adducts are ubiquitous in MALDI-TOF mass spectra, they are particularly evident at low sample concentrations and can interfere with peptide ionization and mass spectrum interpretation (Zhu 2003). Contaminants including detergents, salts, and substances which ionize more readily than the sample itself also complicate mass spectrum interpretation as these will suppress sample ionization and in many cases completely obliterate mass spectra.

In attempts to improve the mass spectra, lower concentrations of the α-HCCA matrix were used, and desalting of peptide samples with zip-tips was implemented. Unfortunately this did not significantly improve the mass spectra for the large majority of the peptide samples. Sometimes ions do not fly due to the absence of an appropriate charge, ion suppression, or other reasons. This cannot be easily determined for individual ions in an experiment like this. Peptide Masses obtained from the reproducible MALDI-TOF analyses are presented in Tables 5.1, 5.2 and 5.3.

5.3.1.4 ESI-MSMS ANALYSIS

Tandem mass spectrometry is considered an important alternative to Edman N-terminal sequencing for obtaining peptide sequence information. Electro-Spray Ionization Tandem Mass Spectrometry (ESI-MSMS) was carried out to identify and confirm major substrate and product peptides identified by MALDI-TOF and Edman N-terminal Sequencing (Figures 5.3, 5.4 and 5.5; Tables 5.1, 5.2 and 5.3)

In general, the mass spectra obtained provided relatively unambiguous assignment of the peptide. One of the disadvantages of MSMS is the complex and often ambiguous process for the interpretation of the mass spectra. An ideal mass spectrum would consist of either complete a- and x-ions series, or b- and y-ion series, or c- and z-ions series (Figure 5.6) (Chapman 1996; Kinter 2000). However, this seldom occurs, as internal cleavages, secondary fragmentation, side-chain cleavages, and other peptide backbone series usually complicate the spectrum (Chapman 1996).

Another aspect that can complicate mass spectra interpretation is the charge state of the ion. Often MS spectra consist of a distribution of multiply-charged ions and it is necessary to determine the charge state of the ions to determine the molecular mass.

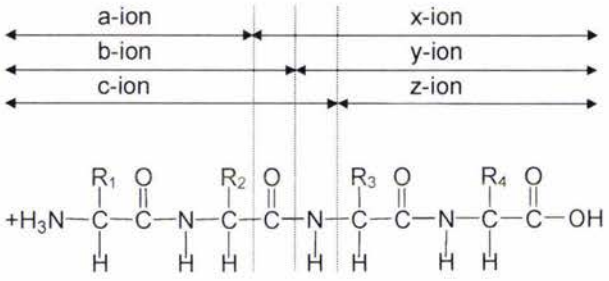


Figure 5.6 Bonds that must be Fragmented to Produce the a-, b-, c- and x-, y-, z-ions.

Both the charge state and fragment ion-series were analysed for each peptide, before the MSMS spectra was submitted to a Mascot MS/MS Ion Search. The Mascot searches were carried out against a milk protein data bank with a 50 ppm mass tolerance to ensure accurate peptide identification.

Although the combined sequencing and mass spectrometry approach identified a large number of the major substrate and product peptides observed by RP-HPLC, very few product peptides could be matched to the substrate peptides. From the cleavage sites of the product peptides which matched to a substrate peptide, it was possible to speculate two possible PepO cleavage sites in κ-casein; Pro_{101}-His_{102} and Lys_{111}-Lys_{112}. Only one conclusive PepO cleavage site, β-casein Leu_6-Asn_7, was identified. Although this peptide did not correspond to a substrate peptide, it retains the N-terminus of the protein.

5.3.1.5 LC-ESI-MSMS ANALYSIS

To further investigate PepO cleavage sites in β-casein, non-fractionated β-casein substrate (Chymosin-generated) peptide mixtures, and non-fractionated β-casein product peptide mixtures were analysed by Liquid Chromatography Tandem Mass Spectrometry (LC-ESI-MSMS) (Table 5.4).

Although LC-ESI-MSMS identified a large number of peptides, it was still not possible to match many of these to the substrate peptides (Table 5.4 and Appendix V). From the cleavage sites of the product peptides which matched to a substrate peptide, it was possible to speculate a number of possible PepO cleavage sites in β-casein; Asp_{129}-Val_{130}, Glu_{131}-Asn_{132}, Asn_{132}-Leu_{133}, Leu_{133}-His_{134}, Leu_{135}-Pro_{136}, Thr_{154}-Val_{155}, Ser_{168}-Lys_{169}, Gln_{175}-Lys_{176}, Lys_{169}-Val_{170}, Val_{173}-Pro_{174}, Arg_{183}-Asp_{184}, Pro_{196}-Val_{197}, Pro_{206}-Ile_{207}.

Table 5.4 LC-ESI-MSMS Analysis of PepO Substrate (Chymosin-Generated β-casein Peptides) and Product Peptides from β-casein.

RP-HPLC Peak [a]	LC-MS [b] (m/z)	Peptide [c]	Sequence
S1'	1473.04	128-140	T-D-V-E-N-L-H-L-P-L-P-L-L
S2'	1688.15	128-142	T-D-V-E-N-L-H-L-P-L-P-L-L-Q-S
S3'	1698.06	143-156	W-M-H-Q-P-H-Q-P-L-P-P-T-V-M
S4'	2466.58	143-163	W-M-H-Q-P-H-Q-P-L-P-P-T-V-M-F-P-P-Q-S-V-L
S5'	2280.47	144-163	M-H-Q-P-H-Q-P-L-P-P-T-V-M-F-P-P-Q-S-V-L
S6'	2480.61	144-165	M-H-Q-P-H-Q-P-L-P-P-T-V-M-F-P-P-Q-S-V-L-S-L
S7'	986.69	157-165	F-P-P-Q-S-V-L-S-L
S8'	2876.93	164-189	S-L-S-Q-S-K-V-L-P-V-P-Q-K-A-V-P-Y-P-Q-R-D-M-P-I-Q-A
S9'	2676.78	166-189	S-Q-S-K-V-L-P-V-P-Q-K-A-V-P-Y-P-Q-R-D-M-P-I-Q-A
S10'	2461.67	168-189	S-K-V-L-P-V-P-Q-K-A-V-P-Y-P-Q-R-D-M-P-I-Q-A
S11'	1780.99	193-208	Y-Q-W-P-V-L-G-P-V-R-G-P-F-P-I-I
S12'	1880.33	193-209	Y-Q-W-P-V-L-G-P-V-R-G-P-F-P-I-I-V
P1'	968.49	6-14	L-N-V-P-G-E-I-V-E
P2'	791.41	58-64	L-V-Y-P-F-P-G
P3'	1452.74	58-70	L-V-Y-P-F-P-G-P-I-H-N-S-L
P4'	678.32	59-64	V-Y-P-F-P-G
P5'	1025.50	59-67	V-Y-P-F-P-G-P-I-H
P6'	1139.55	59-68	V-Y-P-F-P-G-P-I-H-N
P7'	1299.66	59-70	V-Y-P-F-P-G-P-I-P-N-S-L
P8'	753.43	81-87	P-V-V-V-P-P-F
P9'	866.50	81-88	P-V-V-V-P-P-F-L
P10'	994.55	81-89	P-V-V-V-P-P-F-L-Q
P11'	1319.75	81-92	P-V-V-V-P-P-F-L-Q-P-E-V
P12'	1450.76	81-93	P-V-V-V-P-P-F-L-Q-P-E-V-M
P13'	2401.18	94-114	G-V-S-K-V-K-E-A-M-A-P-K-H-K-E-M-P-F-P-K-Y
P14'	2158.07	97-114	K-V-K-E-A-M-A-P-K-H-K-E-M-P-F-P-K-Y
P15'	1673.79	101-114	A-M-A-P-K-H-K-E-M-P-F-P-K-Y
P16'	1998.93	101-117	A-M-A-P-K-H-K-E-M-P-F-P-K-Y-P-V-E
P17'	910.40	108-114	E-M-P-F-P-K-Y
P18'	688.32	115-120	P-V-E-P-F-T
P19'	817.35	115-121	P-V-E-P-F-T-E
P20'	904.38	115-122	P-V-E-P-F-T-E-S
P21'	1119.48	115-124	P-V-E-P-F-T-E-S-Q-S
P22'	1256.71	130-140	V-E-N-L-H-L-P-L-P-L-L
P23'	1028.61	132-140	N-L-H-L-P-L-P-L-L
P24'	914.57	133-140	L-H-L-P-L-P-L-L
P25'	801.49	134-140	H-L-P-L-P-L-L
P26'	627.30	136-140	P-L-P-L-L
P27'	1281.59	144-154	M-H-Q-P-H-Q-P-L-P-P-T
P28'	651.37	170-175	V-L-P-V-P-Q
P29'	1183.00	174-183	P-Q-K-A-V-P-Y-P-Q-R
P30'	1037.58	197-206	V-L-G-P-V-R-G-P-F-P

[a] RP-HPLC Peak; Substrate (S) and Product (P) peptides are sorted based on position in b-casein and not elution from RP-HPLC (data not shown). The peptides do not correspond to peaks in Table 5.1.

[b] Mass-to-charge (m/z) ratio; measured in Thompson (Th). Tandem Mass Spectrometry data was used to identify the peptide in all cases.

[c] Peptide; identified by residue number.

These results do not show any type of consensus regarding the residues flanking the scissile peptide bond and suggest that PepO has broad substrate specificity. PepO substrate specificity is therefore likely to be determined not by sequence but by peptide structure.

Casein molecules have an extraordinarily high composition of proline residues which perturb the peptide backbone. To investigate possible substrate recognition it would be necessary to have structural information regarding both the substrate and enzyme.

5.3.2 *L. RHAMNOSUS* PEPW SUBSTRATE SPECIFICITY

5.3.2.1 SUBSTRATE SPECIFICITY ANALYSIS

PepW is a putative cysteine aminopeptidase that is completely uncharacterised. To establish the reaction catalysed by PepW its substrate specificity towards bovine casein peptides was investigated. Due to time constraints the catalytic mechanism of PepW was not investigated and the substrate specificity investigation was limited to β- and κ-casein peptides derived from Chymosin digestion (Chapter 4).

PepW was expressed as a fusion protein with Thioredoxin (Trx) attached to its N-terminus. To investigate whether or not the Trx-PepW fusion protein influenced PepW peptidase activity, substrate specificity investigations were carried out for both PepW and Trx-PepW (data not shown). PepW and Trx-PepW were found to hydrolyse the same substrate peptides generating the same product peptides. The only difference observed was the rate of hydrolysis, and this is likely to be due to the lower concentration of PepW used in the reaction with the fusion protein. PepW concentrations were determined using Bradford assays and absorbance at 280 nm measurements. Although the PepW and Trx-PepW Molar concentrations were the same, the amount of PepW in the Trx-PepW solution was considerably lower, thereby explaining the reduced rate of hydrolysis.

PepW was found to cleave several β- and κ-casein substrate peptides (Figures 5.7, 5.8). Identification of the β- and κ-casein substrate and product peptides was carried out with Edman N-terminal Sequencing, MALDI-TOF and API-MSMS (Tables 5.5 and 5.6).

Unfortunately, very few product peptides were identified preventing an accurate analysis of PepW substrate specificity. However, the κ-casein(35-?) product peptide derived from hydrolysis of the κ-casein(33-41) substrate peptide, demonstrated that PepW acts as an exopeptidase rather than an endopeptidase.

Although the results obtained are not sufficient to speculate the reaction catalysed, it is suspected that PepW is an aminopeptidase as apposed to a dipeptidyl-peptidase. The dipeptidyl-peptidase activity suggested by the removal of a dipeptide from the κ-casein(33-41) substrate peptide to generate the κ-casein(35-?) product peptide is unlikely to be correct. It is more likely that κ-casein(35-?) did not result from direct hydrolysis of κ-casein(33-41) but by hydrolysis of an unidentified κ-casein(34-?) product peptide.

To establish the catalytic mechanism and substrate specificity of PepW further investigations are required.

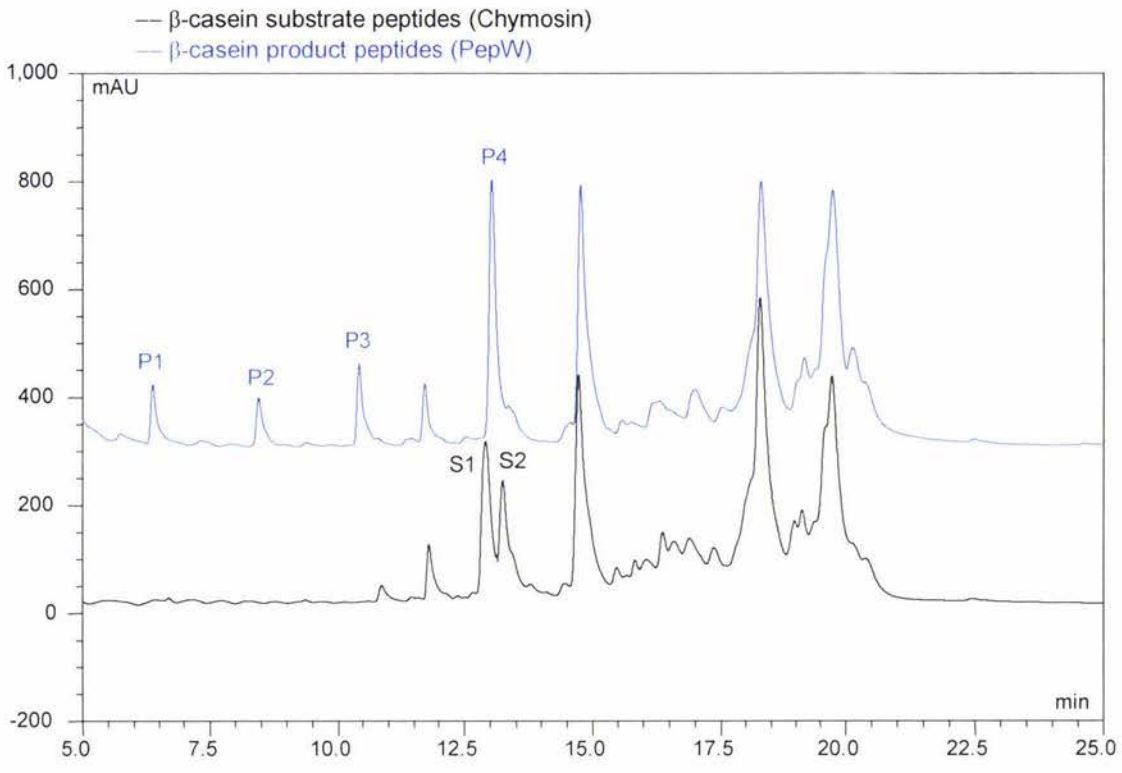


Figure 5.7 RP-HPLC Analysis of PepW Substrate (Chymosin-Generated β-casein Peptides) and Product Peptides. Peaks labelled S1-S2 (Substrate peptides) and P1-P4 (Product peptides) are referred to in Table 5.5.

Table 5.5 Mass and Sequence Analysis of PepW Substrate and Product Peptides from β-casein.

RP-HPLC Peak	MALDI-TOF [a] (m/z)	API-MS [a] (m/z)	API-MSMS [b]	Edman Sequence	Peptide [c]
S1	N/D	N/D	x	R-E-L-E-E-L-N-	β-casein(1-?)
S2	N/D	N/D	x	M-H-Q-P-H-Q-P-L-	β-casein(144-?)
P1	N/D	N/D	x	N/D	N/D
P2	N/D	N/D	x	N/D	N/D
P3	N/D	N/D	x	N/D	N/D
P4	N/D	N/D	x	N/D	N/D

[a] Mass-to-charge (m/z) ratio, measured in Thompson (Th).

[b] Tandem Mass Spectrometry data used to identify peptide; ✓, yes; x, no.

[c] ?, C-terminus was not determined.

N/D, Not Determined.

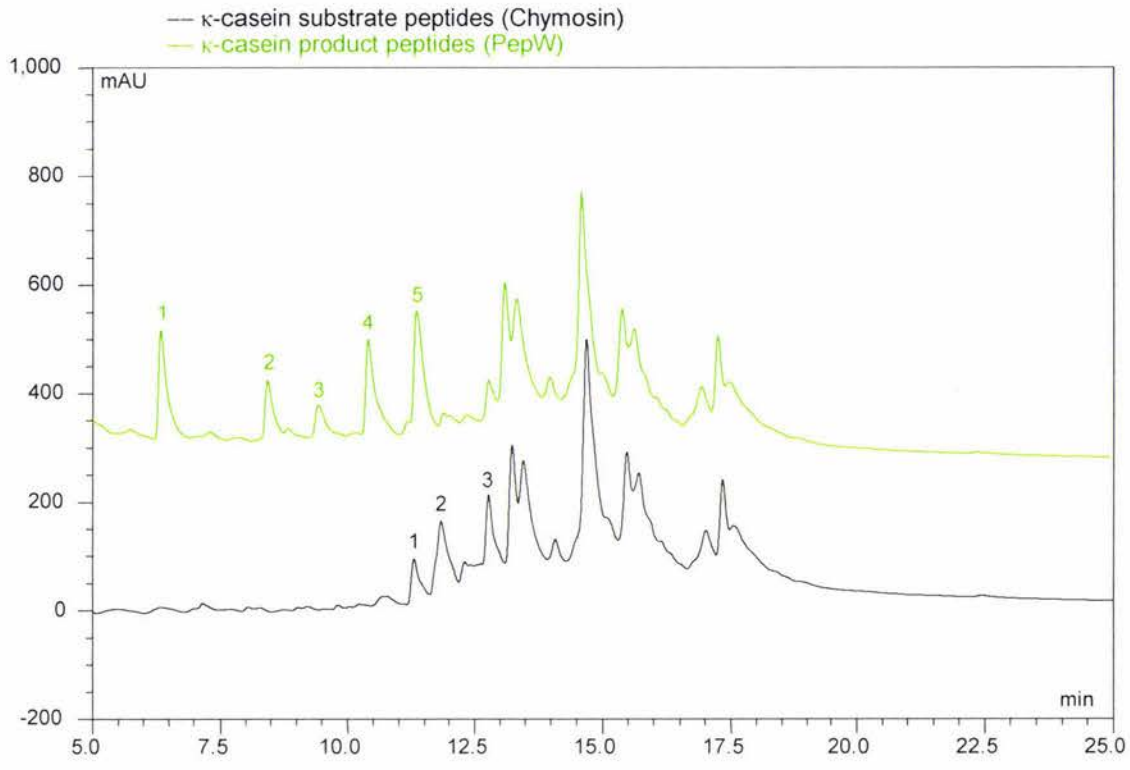


Figure 5.16 RP-HPLC Analysis of PepW Substrate (Chymosin-Generated κ-casein Peptides) and Product Peptides. Peaks labelled S1-S3 (Substrate peptides) and P1-P5 (Product peptides) are referred to in Table 5.6.

Table 5.6 Mass and Sequence Analysis of PepW κ-casein Substrate and Product Peptides.

RP-HPLC Peak	MALDI-TOF [a] (m/z)	API-MS [a] (m/z)	API-MSMS [b]	Edman Sequence	Peptide [c]
Substrate	N/D	1056.8	✓	S-R-Y-P-S-Y-G-L-	κ-casein(33-41)
Substrate	N/D	946.5	x	N/D	N/D
Substrate	N/D	600.0	✓	N/D	κ-casein(96-105)
Product 1	N/D	N/D	x	N/D	N/D
Product 2	N/D	N/D	x	N/D	N/D
Product 3	N/D	N/D	x	N/D	N/D
Product 4	N/D	N/D	x	Y-P-S-Y-G-L-N-	κ-casein(35-?)
Product 5	N/D	N/D	x	N/D	N/D

[a] Mass-to-charge (m/z) ratio, measured in Thompson (Th).

[b] Tandem Mass Spectrometry data used to identify peptide; ✓, yes; x, no.

[c] ?, C-terminus was not determined.

N/D, Not Determined.

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 CONCLUSIONS

Part 1 of this thesis investigated the substrate specificity of the two peptidases, PepO and PepW both produced by *L. rhamnosus*. The 1.347-kb *pepW* gene was successfully cloned from *L. rhamnosus* into an *E. coli* expression system utilising a modified pET32a vector with T7-promoter, IPTG induction and an N-terminal fusion Trx-His-rTEV-tag. Recombinant PepW was purified to homogeneity using immobilised metal affinity chromatography which exploited the His-tag and size exclusion chromatography was used to separate the tag from PepW, following removal of the fusion tag by recombinant TEV protease.. PepW was shown by size exclusion chromatography to exist predominantly as a hexamer of 50 kDa PepW subunits.

Recombinant PepO was expressed from an *L. lactis* expression system that utilised a *nisA*-promoter and Nisin induction. PepO was purified to homogeneity, without the use of a fusion tag, by ion exchange chromatography, hydrophobic interaction chromatography and size exclusion chromatography. It was shown by size exclusion chromatography to exist as a 70 kDa monomer, and by RP-HPLC analysis to be inhibited by EDTA and phosphoramidon, classifying it as a metallopeptidase.

Substrate specificity of PepO and PepW towards bovine α_{s1}-, β- and κ-casein peptides as well as casein peptides extracted from Cheddar cheese was then investigated. PepO and PepW selectively hydrolysed chymosin-derived bovine β- and κ-casein peptides, and casein peptides that had been formed in Cheddar cheese. Albeit to a lesser extent, PepO was also found to hydrolyse β- and κ-casein peptides derived from both chymosin and *L. lactis* H2 PrtP, and from chymosin and *L. lactis* SK11 PrtP hydrolysis.

In order to establish the exact cleavage points both substrate and product peptides were isolated and analysed by a combination of sequencing and mass spectrometry techniques that included: Edman N-terminal Sequencing, MALDI-TOF MS, ESI-MSMS and LC-ESI-MSMS.

Although a considerable amount of information was obtained from these experiments, it was not sufficient to completely characterise the substrate specificities of either PepO or PepW. Only one conclusive PepO cleavage site that had not been previously characterised was identified. This cleavage site was the β-casein peptide bond between Leu_6-Asn_7. Several possible PepO and PepW cleavage sites in α_{s1}-, β- and κ-casein were identified, suggesting that PepO has a broad endopeptidase activity, whilst PepW has a specific exopeptidase activity. Further work is required to conclusively determine both PepO and PepW substrate specificity.

6.3 FUTURE DIRECTIONS

To determine with certainty the substrate specificity of PepO and PepW it will be necessary to match product peptides produced by these enzymes to the corresponding substrate peptides. This is often difficult because as hydrolysis of a peptide substrate proceeds, product peptides are generated which become new substrates (secondary substrates) for the peptidase.

An approach that would prove useful would be to fractionate peptide substrates and individually analyse the hydrolysis of these by PepO and PepW. This would limit the number of secondary substrate peptides, and allow a more controlled investigation into the substrate specificity of these enzymes.

The incentive to investigate PepO and PepW substrate specificity was to establish if these enzymes could potentially be used to modify and control cheese bitterness and possibly other flavour attributes. It is therefore necessary to investigate the substrate specificity of PepO and PepW under physiological/application conditions. This could be achieved by the overexpression of PepO and PepW in cheese, or in vitro by adjusting the reaction conditions including temperature, ionic strength, and pH.

Flavour development in cheese is of major economic interest since the final flavour of cheeses partly determines consumer choice and because the development of flavour is a time consuming and an expensive process. An understanding of the *L. rhamnosus* PepO and PepW substrate specificity may enable controlled synthesis of flavour compounds through judicious choice of LAB organisms. In the long term, the development of genetically modified LAB strains with improved fermentation properties, and substrate specificity will provide the ultimate tool for the control of flavour in various dairy products.

PART 2:

THREE-DIMENSIONAL STRUCTURE DETERMINATION OF THE *L. RHAMNOSUS* PEPTIDASES, PEPO AND PEPW.

INTRODUCTION

1.1 INTRODUCTION

Part 2 of this thesis focuses on the investigation into the three-dimensional structure of the *L. rhamnosus* peptidases, PepO and PepW. PepO shares 27% amino acid identity (according to BLAST) to the *H. sapiens* neutral endopeptidase, NEP (PDB accession 1DMT), and PepW shares 44% and 37% amino acid identity (according to BLAST) to the *H. sapiens* bleomycin hydrolase, hBH (PDB accession 1CB5) and the *S. cerevisiae* bleomycin hydrolase, Gal6 (PDB accession 1GCB) respectively. Typically, proteins with greater than 30% amino acid identity have conserved structural motifs. The aim of this section of work was to investigate the three-dimensional structures of the *L. rhamnosus* peptidases, PepO and PepW, using x-ray crystallography.

1.2 OBJECTIVES AND STRATEGIES

The objectives of this section of work were:

1) Grow suitable *L. rhamnosus* PepO and PepW crystals for x-ray crystallography.

2) Collect native data sets for the *L. rhamnosus* PepO and PepW crystals.

3) Solve the three-dimensional structures of *L. rhamnosus* PepO and PepW by Molecular Replacement using *H. sapiens* NEP, and *H. sapiens* hBH and *S. cerevisiae* Gal6, as search models respectively.

   If Molecular Replacement proved to be unsuccessful:

4) Derivatize *L. rhamnosus* PepO and PepW crystals with heavy atom(s).

5) Collect derivative data sets for the *L. rhamnosus* PepO and PepW derivative crystals.

6) Solve the three-dimensional structures of *L. rhamnosus* PepO and PepW using Multiple Isomorphous Replacement (MIR), or Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS).

CRYSTALLISATION

2.1 INTRODUCTION

The first prerequisite for solving a three-dimensional structure of a protein by x-ray crystallography is a well ordered crystal that strongly diffracts x-rays (Martin 1996). To grow crystals, the protein molecules must be brought into a thermodynamically unstable state known as supersaturation. This is usually achieved by the gradual removal of solvent and proceeds via three stages: nucleation, crystal growth and cessation of growth (Martin 1996). These three stages can be described with reference to the crystallisation phase diagram (Ducruix 1992) (Figure 2.1).

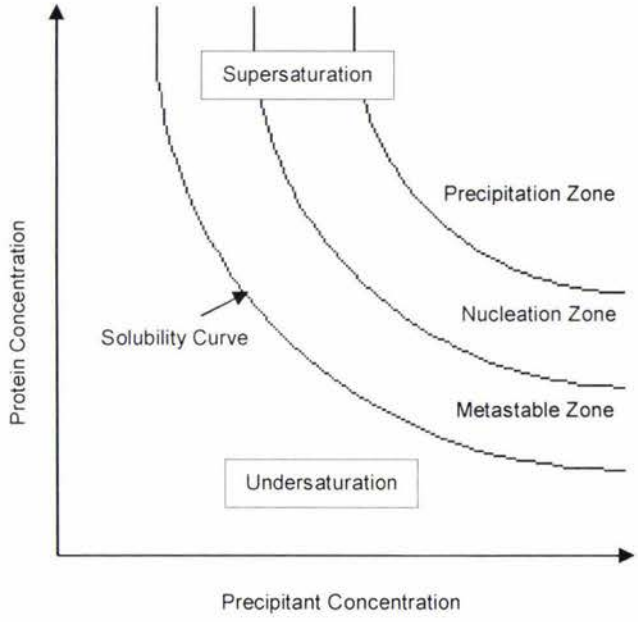


Figure 2.1. Crystallisation Phase Diagram. The crystallisation phase diagram partitions essentially into two zones; undersaturated and supersaturated, separated by the solubility curve. The supersaturated zone is subdivided into three zones: the metastable zone; nucleation zone; and the precipitation zone, defined by increasing levels of supersaturation (Modified from Ducruix 1992).

A large number of extrinsic and intrinsic factors influence the crystallisation process of protein molecules including: protein purity and concentration, the presence of ligands or cofactors, precipitating agent, pH, and temperature (Table 2.1).

Table 2.1 Factors Affecting Crystallisation

Extrinsic Factors	Intrinsic Factors
Temperature	Purity
pH	Proteolytic digestion
Ionic strength	Single point mutations
Protein concentration	Aggregation state
Precipitant type and concentration	Glycosylation state
Ligands, cofactors	
Crystallisation method	

Successful protein crystallisation typically requires that the protein is 95-100% pure, and exists as a single conformational species. Temperature, pH and ionic strength all influence the crystallisation process by affecting both protein solubility and stability.

Finding the best combination among many crystallisation parameters is a multivariate task and is often limited by the amount of pure protein available. Often the best approach to crystallisation is to use techniques and reagents that have proven successful for other proteins (Gilliland 1996). In this research crystallisation trials were performed using the vapour diffusion technique using commercially available sparse matrix crystallisation conditions to effectively cover a number of precipitant, pH conditions and additives (Jancarik 1991). The vapour diffusion technique utilises evaporation and diffusion of water between solutions of different concentrations as a means of approaching and achieving supersaturation of protein molecules (McPherson 1990).

The most commonly used vapour diffusion technique is the hanging drop method, although the same effect is achieved with the sitting drop method. In these methods a drop of protein solution is mixed with an equal volume of precipitant at subsaturation, and is then placed in a sealed vessel above a reservoir containing the same precipitant. The difference in precipitant concentration between the drop and the reservoir causes water to evaporate from the drop until the concentration of the precipitant in the drop equals that of the reservoir. Under the correct conditions, nucleation will result followed by crystal growth.

Crystallisation trials should be examined microscopically once a day for the first week, then once a week there after. Results should be recorded to indicate whether the drop is clear, contains precipitate and/or crystals. If crystals appear in the drop it is necessary to determine whether or not they are protein or salt crystals. Protein crystals differ from salt crystals in a number of aspects: (i) protein crystals are very fragile whilst salt crystals are usually hard; (ii) protein crystals show signs of birefringence under the microscope; (iii) protein crystals will be present in only the drop and not the reservoir. If multiple crystals with the same morphology are

present in the drop, it is possible to check the crystal more thoroughly by Coomassie Blue staining or SDS-PAGE analysis, or alternatively performing activity assays.

Once it is established that the crystals are protein crystals the next step is to analyse the crystal quality by x-ray diffraction. Preliminary indications of high crystal quality come from optical clarity and sharply faceted crystal surfaces. However, the only reliable way to tell if the crystals are suitable for x-ray crystallography is by subjecting them to x-ray diffraction.

2.1 METHODS

2.2.1 Handling and Storage of Protein Samples

Protein purification was carried out as described in Part 1 Chapter 3 in the shortest time possible to ensure fresh and pure protein for crystallisation. Protein samples were concentrated to concentrations of approximately 11.5 mg/mL for PepO (in 0.2 M ammonium sulphate, 50 mM Tris pH 8.0) and 5.0 mg/mL for PepW (in 0.1 M NaCl, 2 mM DTT, 50 mM Tris pH 7.5). Where possible crystallisation trials were set up immediately following protein concentration, in all other cases protein solutions were aliquoted and stored at -80°C. Prior to crystallisation the protein solutions were equilibrated to the temperature at which the crystallization experiments were to be set up and filtered through 0.22 μm filters (Ultrafree-Millipore)to remove any particulate matter by centrifugation.

2.2.2 Hanging Drop Method

Crystallisation experiments were carried out using the vapour-diffusion hanging drop method (McPherson 1990). A drop containing 1 μL concentrated protein and 1μL precipitant was mixed on a siliconized cover slide (Hampton). This was inverted and sealed with vaseline over a well containing 400 μL precipitant. Crystallisation trials were set up at 4°C and room temperature, using Linbro trays.

2.2.3 Sitting Drop Method

Crystallisation experiments were also carried out using the vapour-diffusion sitting drop method (McPherson 1990). The sitting drop crystallisation screens were set up using the Centre for Molecular Biodiscovery Crystallisation facility (University of Auckland, New Zealand). A modified Perkin-Elmer Multiprobe robot was used to transfer the crystallisation reagents from pre-made stock into Intelliplates, and a Cartesian Honeybee robot was used to set up the nanolitre crystallisation experiments. A drop containing 100 nL concentrated protein and 100 nL crystallisation reagent was mixed on a platform of a sealed 96-well Intelliplate (Hampton Research) above a well containing 75 μL crystallisation reagent. Crystallisation trials were set up at room temperature.

2.2.4 Crystallisation Reagents

The crystallisation reagents used included: Hampton Research crystal screens 1 and 2; PEG screen; PEG/Ion screen; Footprint I screen; Top 67 screen; and an Ammonium sulphate screen. The list of reagent formulations for these screens is provided in Appendix VI.

2.2.5 Optimisation of Crystallisation Conditions

Once preliminary crystallization conditions were found, optimization was carried out using the vapour-diffusion hanging drop method by screening around the preliminary conditions. In the case of PepO the following optimisation conditions were used: (i) the protein concentration was decreased from 11.5 mg/mL to 9.0 mg/mL with 0.5 mg/mL increments; (ii) the precipitant concentration was examined over a range of 14.0-22.0% PEG 8000 with 1% increments; (iii) the calcium acetate concentration was varied over a range of 0.0-0.25 M in 0.05 M increments; (iv) the pH was screened over a range of pH 4.0-7.0 using increments of 0.5 pH units; (v) the temperature was reduced to 4°C; and (vi) co-crystallisation of PepO with the inhibitor phosphoramidon (0.1-0.6 mM) was trialled. For PepW, reproduction of the successful sitting drop crystallisation experiments using the hanging drop method.

2.2.6 Crystal Characterisation

Crystallisation progress was monitored at the temperature at which the crystallisation trials were carried out using a Nikon Petrographic Model SMZ-2B Stereo-microscope, (1) immediately after the trials were set up, (2) at 2 day intervals for the first week, (3) once a week for 2 months and then (4) once a month. Crystallisation was recorded accordingly as: (1) clear drop, (2) phase separation, (3) regular granular precipitant, (4) birefringent precipitate or microcrystals, (5) rosettes or spherulites, (6) needles (1D growth), (7) plates (2D growth), (8) single crystals (3D growth < 0.2 mm), (9) single crystals (3D growth > 0.2 mm). Images of any crystals were recorded at room temperature at x10 magnification using a Leica MZ12 Stereo-microscope, Optronics Digital Microscope Camera and MagnaFire Version 2.1B software.

2.2.7 Heavy Atom Derivatives

Derivatization of PepO crystals with heavy atoms was carried out by soaking the crystals for varying lengths of time at 4°C in 1 μL of 100 mg/mL of Ethyl mercury chloride (CH_3CH_2HgCl), Thiomersal ($C_9H_9HgO_2SNa$), Uranyl acetate ($UO_2CH_3CO_2$) and Potassium tetra-chloroplatinate (K_2PtCl_4) in mother liquor. Derivatization of PepO with heavy atoms was also carried out by crystallisation trials in which 100 mg/mL Praeseodynium (III) nitrate ($PrNO_3$) was present in the mother liquor (Table 2.2). In retrospect, the heavy atom concentration (100mg/mL) was far too high. However, crystal cracking and/or disintegration was rarely observed.

Table 2.2 Heavy Atom Derivative Crystallisation Conditions.

Heavy Atom	Concentration	Soak [a]
$C_9H_9HgO_2SNa$	100 mg/mL	1 day
$C_9H_9HgO_2SNa$	100 mg/mL	1 week
CH_3CH_2HgCl	100 mg/mL	6 hours
CH_3CH_2HgCl	100 mg/mL	1 day
CH_3CH_2HgCl	100 mg/mL	1 week
$UO_2CH_3CO_2$	100 mg/mL	1 day
$UO_2CH_3CO_2$	100 mg/mL	1 week
K_2PtCl_4	100 mg/mL	8 days
$PrNO_3$	100 mg/mL	N/A

[a] The mother liquor was: 15% PEG 8000, 0.1 M $Na(CH_3)_2AsO_2$ pH 6.5, 0.2 M $CaCH_2CO_2$.

2.3 RESULTS AND DISCUSSION

2.3.1 *L. RHAMNOSUS* PEPO CRYSTALLISATION

2.3.1.1 INITIAL CRYSTALLISATION SCREENS

Initial crystallisation screens were performed at room temperature using the hanging drop method in conjunction with the commercially available Hampton Research crystallisation screens (Appendix VI). Out of the 98 crystallisation conditions tested, crystals formed within 1-2 days in the following conditions:

(a) 30% MPD, 0.1 M sodium acetate pH 4.6, 0.02 M calcium chloride
(b) 28% PEG 400, 0.1 M sodium hepes, pH 7.5, 0.2 M calcium chloride
(c) 20% Isopropanol, 0.1 M sodium acetate pH 4.6, 0.2 M calcium chloride
(d) 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate

The crystals had visibly different morphologys suggestive of different crystal systems (Figure 2.2).

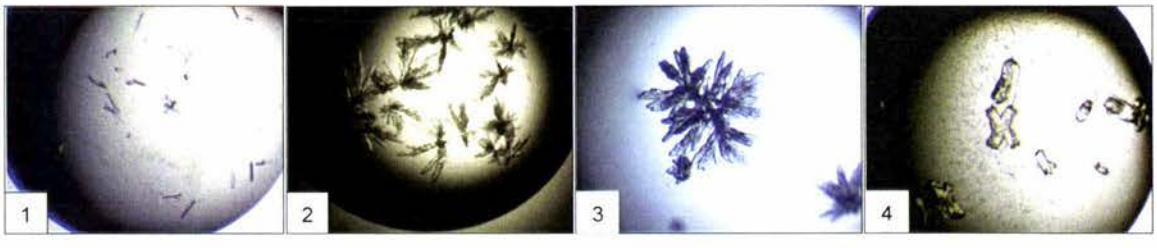


Figure 2.2 PepO Crystals Obtained from Initial Crystallisation Screens at Room Temperature. (1, crystals from condition (a); 2, crystals from condition (b); 3, crystals from condition (c); 4, crystals from condition (d)).

The presence of a calcium containing salt appears to be a critical determinant in the growth of these crystals, and is likely to play a structural role either co-ordinating intramolecular or intermolecular interactions within these crystals.

Only very few of the crystals had sharp crystal surfaces suggestive of high quality crystal structures. X-ray diffraction analysis of these crystals showed that only those grown in condition (d) diffracted x-rays and were of suitable quality for x-ray crystallography (Chapter 3).

Crystal nucleation and growth in condition (d) was evident within 30 minutes of setting up the crystal trials, and the cessation of crystal growth occurred within 1-2 days. After a period of

approximately 2 weeks, the protein in solution began to precipitate and within 1 month deterioration of the crystal surfaces was apparent (Figure 2.3).

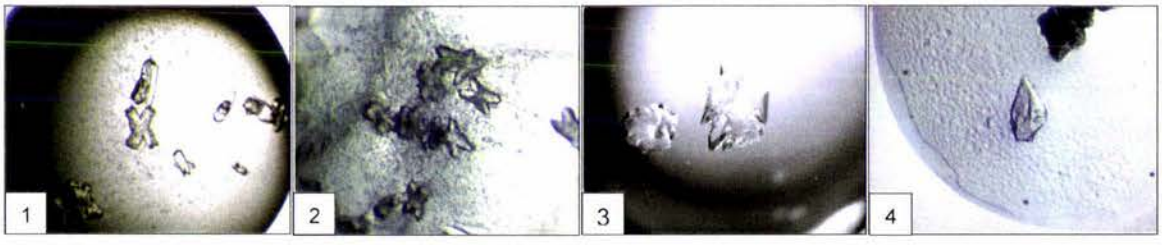


Figure 2.3 PepO Crystals Grown in Conditions (d) and (e). (1, 1 day and; 2, 1 month old crystals obtained from condition (d) at room temperature; 3, crystals obtained from condition (d) at 4°C; 4, crystals obtained from condition (e) at 4°C. Note: photographs were taken at room temperature, which resulted in crystal deterioration of 3 and 4).

2.3.1.2 OPTIMISATION SCREENS

In an attempt to improve crystal quality, and prevent deterioration, crystal trials were conducted at 4°C. Decreasing the temperature effectively reduces the rate of equilibration and hence reduces crystal nucleation and growth rates. Crystal nucleation and growth at 4°C was evident within 1-2 days and crystal growth cessation occurred within 1-2 weeks. The majority of crystals obtained had a different morphology than those obtained at room temperature (Figure 2.3).

In order to obtain single large crystals suitable for structure determination by x-ray diffraction, several refinement screens were set up around condition (d). The crystallisation parameters optimised included: protein, precipitant and salt concentrations; pH; temperature and additives. All of the following optimisation trials were carried out at 4°C.

In theory, decreasing the protein and/or precipitant concentration should reduce the rate at which supersaturation is reached and thus reduce the rate of nucleation and crystal growth. Reducing the number of nuclei from which crystals can grow and slowing crystal growth should produce single large, high-quality crystals. When the protein concentration was decreased from 11.5mg/mL to 9.0mg/mL, no significant difference was observed, however, when the precipitant concentration was decreased from 18.0% to 14.0% PEG 8000, a reduced rate of nucleation and crystal growth was observed. In contrast, when the precipitant concentration was increased from 18.0% to 22.0% PEG 8000, the rate of nucleation and crystal growth increased, and above 20.0% PEG 8000 a skin formed over the drop which appeared to inhibit crystal growth.

The salt concentration was adjusted to find the optimum concentration for crystallisation. Salts normally influence the crystallisation process through competition for water molecules, although,

in this case the calcium ions are likely to influence crystallisation through specific protein-ion interactions. Crystallisation was optimal between 0.05-0.20 M calcium acetate. At concentrations below 0.05 M the solution did not appear to reach supersaturation, and at concentrations exceeding 0.20 M, the protein precipitated.

The optimum pH was found to be dependent upon the precipitant concentration. At 18.0% PEG 8000 the optimum pH was pH 5.75; above 18.0% PEG 8000 the optimum pH shifted to pH 5.5; below 18% PEG 8000 the optimum pH shifted to pH 6.0.

Additives including 1.0-25.0% glycerol, and 1.0-10.0% ethanol were added to the mother liquor in an attempt to improve crystal quality, but had no apparent effect. The optimum solutions for producing reproducible large single PepO crystals were:

(e) 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate

Crystals obtained in condition (e) were single, had sharp faces (Figure 2.3), and x-ray diffraction analysis demonstrated that they diffracted to 2.00 A resolution (Chapter 3).

2.3.1.3 HEAVY ATOM DERIVATIVE SCREENS

Multiple isomorphous replacement (MIR) is the predominant method for determining the initial phasing of completely new macromolecular structures (Chapter 4). As it is not the focus of this section to discuss the MIR technique, it will only be introduced here to rationalize the crystallisation of heavy atom derivatives. The MIR approach requires at least two and in many cases numerous heavy atom derivative crystals of the macromolecule. The introduction of a heavy atom (defined as an atom with atomic weight greater than that of rubidium) into a crystal significantly changes the intensity distribution of the diffraction pattern (Rossmann 2001). The difference in intensity between native and heavy atom data can be used to calculate phase information by determining the location of the heavy atom(s) in the cell.

The heavy atom derivative and native crystals need to be isomorphous, i.e. the difference in cell dimensions of the native and heavy atom crystal unit cells should be less than 5.0% over all axes (Crick 1956; Rossmann 2001). Heavy atom derivatization is typically carried out by adding a small volume of a heavy atom solution to the mother liquor, or by transferring the native crystal(s) to a drop containing the heavy atom dissolved in the mother liquor. Both approaches were taken to produce heavy atom derivative crystals of PepO.

The crystallisation of PepO seemed to be dependent on the presence of Ca^{2+} ions, it was decided to try to replace Ca^{2+} with Pr^{3+} in the mother liquor. The lanthanoid, Prasoedymium

(PrIII), has a similar ionic radii to calcium (6 and 8 co-ordinate ionic radii of 113 pm and 126.6 pm respectively), and is a good anomalous scatterer.

Derivative crystals grew when the Ca^{2+} was replaced with between 0.025-0.100 M prasoedymium nitrate. Better crystals, however, grew in the following solution:

(f) 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.08 M prasoedymium nitrate

These crystals had a considerably different morphology to those obtained in previous crystallisation trials (Figure 2.4).

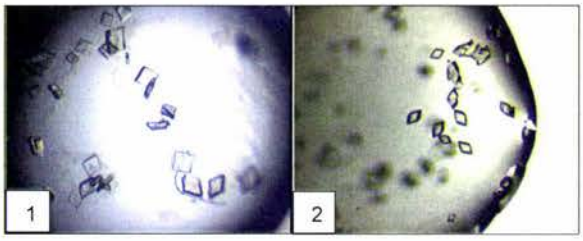


Figure 2.4 PepO Crystals Grown with Prasoedymium and Phosphoramidon. (1, PepO heavy atom derivative crystals with praseodymium; 2, PepO crystals complexed with phosphoramidon.

Although x-ray diffraction analysis revealed that the crystals were protein crystals and that they diffracted, they were not isomorphous to the native PepO crystals (Chapter 3).

Heavy atom derivatization was also carried out by soaking PepO crystals in mother liquor containing various heavy atoms. The choice of the heavy-atoms used in derivatization was made after analysis of the amino acid composition of PepO. The amino acid sequence of PepO (632 amino acids) is unusual in that it contains only one cysteine residue (Appendix I). A heavy-atom that specifically targets cysteine residues should therefore in theory, provide a heavy atom derivative containing a single heavy atom.

Cysteine is highly reactive as the thiolate anion, the predominant form above pH 8.0 (Petsko 1985), as it is an excellent nucleophile and reacts strongly with both mercury (Hg) and platinum (Pt) (Petsko 1985). Platinum also binds histidine and methionine residues, and similarly binding decreases below pH 7.0. The PepO sequence contains 13 histidine residues and 20 methionine residues.

composition of the protein but also upon the pH, buffer, and additive salts present in the mother liquor.

The crystals grew at pH 6.5, which is a less than ideal pH for targeting cysteine for heavy atom derivatization as only a small portion of the sidechains will be deprotonated. On the other hand, about 50% of the histidine residues will be deprotonated allowing interaction with positively charged metal ions.

The crystals grew from solutions containing PEG, sodium cacodylate and calcium acetate. PEG does not react with most heavy atom compounds and is therefore a favourable media for heavy atom derivatization (Petsko 1985). Cacodylate (methyl arsonate) is a phosphate mimic, and a commonly used buffer in crystallisation studies. Unlike phosphate, cacodylate does not interfere with heavy atom binding. Although acetate salts can complex uranyl and lanthanide heavy atoms and decrease their reactivity with the protein, it is more suitable than phosphate, citrate or hydroxide salts (Blundell 1976; Petsko 1985). The protein solution used in the crystallisation experiments contained 0.2 M ammonium sulphate. In solutions above pH 6.0, ammonium sulphate produces the strong nucleophile NH_3, which will compete with the proteins for the heavy metals, preventing binding.

Temperature also has a strong influence on heavy atom derivatization. Most heavy-atom binding will be significantly slower at 4°C than at room temperature. Soaking time and heavy-atom concentrations are the other remaining considerations. Reported soaking times range from several hours to months, and concentrations of heavy atoms as low as 0.05 mM and as high as 100 mM have been used (Blundell 1976; Petsko 1985).

Preliminary heavy atom derivative screens were carried out with lower quality crystals. Heavy atom derivatization was carried out by soaking PepO crystals in 100 mg/mL Thiomersal ($C_9H_9HgO_2SNa$), Ethyl mercury chloride (CH_3CH_2HgCl), Uranyl acetate ($UO_2CH_3CO_2$) and Potassium tetra-chloroplatinate (K_2PtCl_4) in mother liquor at 4°C for various lengths of time.

The PepO derivative crystals retained the visible quality observed for the PepO native crystals over a period of approximately 1 week, after which time crystal deterioration was apparent. Crystal cracking was occasionally observed for some of the crystal derivatives. Crystal cracking is often indicative of derivatization at a lattice contact, denaturation of the crystallized protein, or induction of a conformational change that cannot be accommodated within the crystal lattice.

X-ray diffraction analysis is necessary to assess whether the derivative crystals are isomorphous, whether the diffraction differences resulting from the heavy atoms is measurable.

X-ray diffraction analysis of the PepO derivative crystals showed that although most of the heavy atoms did not disrupt the crystal lattice, the heavy atoms did not appear to bind to the protein (Chapters 3 and 4).

2.3.1.4 CO-CRYSTALLISATION SCREENS

In Part 1, Chapter 5 it was established that phosphoramidon, a metabolite produced by *Streptomyces tanashiensis*, inhibits the peptidase activity of PepO. Phosphoramidon specifically inhibits metallopeptidases, and has previously been crystallised in complex with *H. sapiens* neutral endopeptidase (NEP) (Oefner 2000). Co-crystallisation of an enzyme with an inhibitor can be beneficial to crystallisation because it can induce a conformational change in the protein making it more amenable to crystallisation. Co-crystallisation experiments have the additional advantage in that the way the inhibitor binds to the enzyme can show how a substrate may bind, and therefore the structural basis of the catalytic mechanism. Crystals obtained from co-crystallisation experiments with phosphoramidon, grew in the following solution:

(g) 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate, 0.5 mM phosphoramidon

These crystals had a similar morphology to the PepO derivative crystals grown in the presence of praseodymium (Figure 2.4), and although they showed signs of birefringence under the microscope, and appeared only in the crystallisation drop and not in the reservoir solution, x-ray diffraction analysis showed them to be salt crystals (Chapter 3).

2.3.1.5 SECONDARY CRYSTALLISATION SCREENS

In order to obtain a different crystal form suitable for x-ray crystallography, crystal trials were carried out at room temperature using the sitting drop vapour diffusion method in conjunction with a number of sparse matrix screens including: PEG screen, PEG/ION screen, Ammonium Sulphate screen, Footprint I screen and Top 67 screen. Of the 192 crystallisation conditions tested, crystals formed within 1 month in the following 15 conditions:

(a') 20% PEG 3350, 0.2 M ammonium chloride pH 4.0
(b') 20% PEG 3350, 0.2 M calcium chloride
(c') 28% PEG 6000, 0.2 M MES pH 6.1
(d') 28% PEG 6000, 0.2 M BTP pH 6.7
(e') 28% PEG 6000, 0.2 M HEPES pH 7.3
(f') 20% PEG 8000, 0.2 M CHES pH 9.5
(g') 21% Methoxy-PEG 5000, 0.2 M acetic acid pH 4.6
(h') 21% Methoxy-PEG 5000, 0.2 M MES pH 6.1
(I') 21% Methoxy-PEG 5000, 0.2 M BTP pH 6.7

(j') 21% Methoxy-PEG 5000, 0.2 M HEPES pH 7.3
(k') 28% Methoxy-PEG 5000, 0.2 M PIPES pH 6.7
(l') 28% Methoxy-PEG 5000, 0.2 M MOPS pH 7.3
(m') 28% Methoxy-PEG 5000, 0.2 M EPPS pH 7.9
(n') 30% MPD, 0.1 M sodium acetate pH 4.6, 0.2 M calcium chloride
(o') 40% PEG 300, 0.1 M sodium cacodylate pH 7.0, 0.2 M calcium acetate

The crystals obtained in these conditions were very small due to the small scale of the crystallisation trials carried out. A selection of the different crystal morphologies obtained is shown in Figure 2.5.

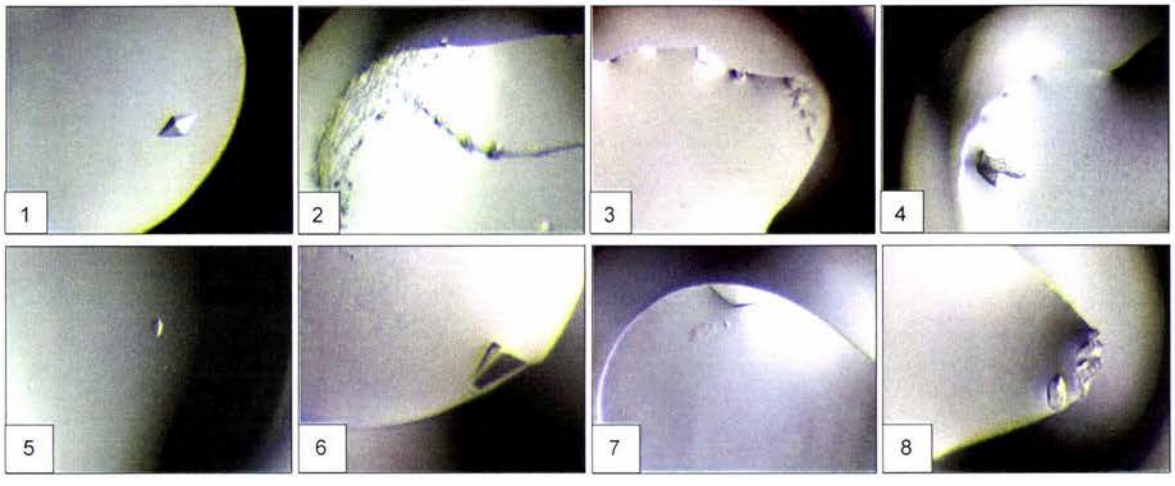


Figure 2.5 PepO Crystals Obtained from Secondary Crystallisation Screens. (1, crystals from condition (a'); 2, crystals from condition (c'); 3, crystals from condition (d'); 4, crystals from condition (f'); 5, crystals from condition (g'); 6, crystals from condition (k'); 7, crystals from condition (l'); crystals from condition (m')).

Due to time constraints x-ray diffraction analysis of these crystals were not carried out to establish the quality and suitability of these crystals for x-ray crystallography.

2.3.1.6 DUMB LUCK CRYSTALS

"Crystals fully recognize that patience is a virtue and will reward those who practice it" (Boyle 1996). Over a period of 10-12 months single large crystals grew in the initial crystallisation screens, at room temperature in the following conditions:

(a'') 1.5 M Lithium sulphate monohydrate, 0.1 M sodium hepes pH 7.5
(b'') 1.0 M Lithium sulphate monohydrate, 0.1 M tri-sodium citrate dehydrate pH 5.6, 0.5 M ammonium sulphate

The common component in these crystallisation conditions was lithium sulphate. Crystals with similar morphology were obtained in these two conditions, however, images were taken only of condition (b") for fear that the heating of the crystal during image capture may damage the crystals (Figure 2.6).

Figure 2.6 Dumb Luck PepO Crystals Obtained from Condition (b").

As discussed in the following section the crystal obtained in condition (a") diffracted to 1.80 A resolution and were suitable for structure determination by x-ray crystallography.

2.3.2 *L. RHAMNOSUS* PEPW CRYSTALLISATION:

2.3.2.1 INITIAL CRYSTALLISATION SCREENS

Initial crystallisation screens were performed at room temperature using the sitting drop vapour diffusion method in conjunction with a number of sparse matrix screens including: Hampton Research crystal screens 1 and 2, PEG screen, PEG/ION screen, MPD screen, Clear Strategy screen, Precipitant Synergy Reagent Formulation screen, Ammonium Sulphate screen, Footprint I screen and Top 67 screen. Out of the 480 crystallisation conditions tested, crystals formed within 1 month in the following 24 conditions:

(a) 20% PEG 3350, 0.2 M di-ammonium tartrate
(b) 20% PEG 3350, sodium di-hydrogen phosphate
(c) 20% PEG 3350, 0.2 M magnesium formate pH 5.9
(d) 20% PEG 3350, 0.2 M tri-potassium citrate monohydrate pH 8.3
(e) 20% PEG 3350, 3 M sodium chloride, 0.1 M magnesium chloride pH 6.5
(f) 20% PEG 3350, 0.2 M sodium chloride
(g) 20% PEG 3350, 0.2 M tri-sodium citrate
(h 25% PEG 3350, 15% MPD, 0.2 M lithium sulphate pH 6.5
(i) 30% PEG 4000, 0.1 M tri-sodium citrate di-hydrate pH 5.6, 0.2 M ammonium acetate
(j) 7% PEG 6000, 0.2 M MOPS pH 7.3
(k) 14% PEG 6000, 0.2 M BTP pH 6.7
(l) 28% PEG 6000, 0.2 M citric acid pH 4.9
(m) 10% PEG 8000, 10% PEG 1000, 0.2 M lithium sulphate
(n) 10% PEG 8000, 10% PEG 1000, 0.2 M KSCN
(o) 10% PEG 8000, 10% PEG 1000, 0.2 M potassium bromide.
(p) 14% PEG 8000, 0.2 M acetic acid pH 4.9
(q) 15% PEG 8000, 0.5 M lithium sulphate monohydrate
(r) 20% PEG 8000, 0.05 M mono-potassium di-hydrogen phosphate
(s) 7% Methoxy-PEG 5000, 0.2 M acetic acid pH 4.9
(t) 14% Methoxy-PEG 5000, 0.2 M citric acid pH 4.9
(u) 21% Methoxy-PEG 5000, 0.2 M acetic acid pH 4.9
(v) 28% Methoxy-PEG 5000, 0.2 M cacodylic acid pH 6.1
(w) 8% PEG 20000, 8% PEG 550 MME, 0.2 M magnesium chloride
(x) 40% MPD, 0.2M tri-lithium citrate tetra-hydrate

The PepW crystals grew in a number of different precipitant, salt and buffer types over a pH range of 4.9-8.3. The theoretical pI value of PepW is 5.23, and is within the pH range for successful crystallisation. The majority of crystals obtained had sharply faceted crystal surfaces suggestive of high quality crystal structures (Figure 2.7).

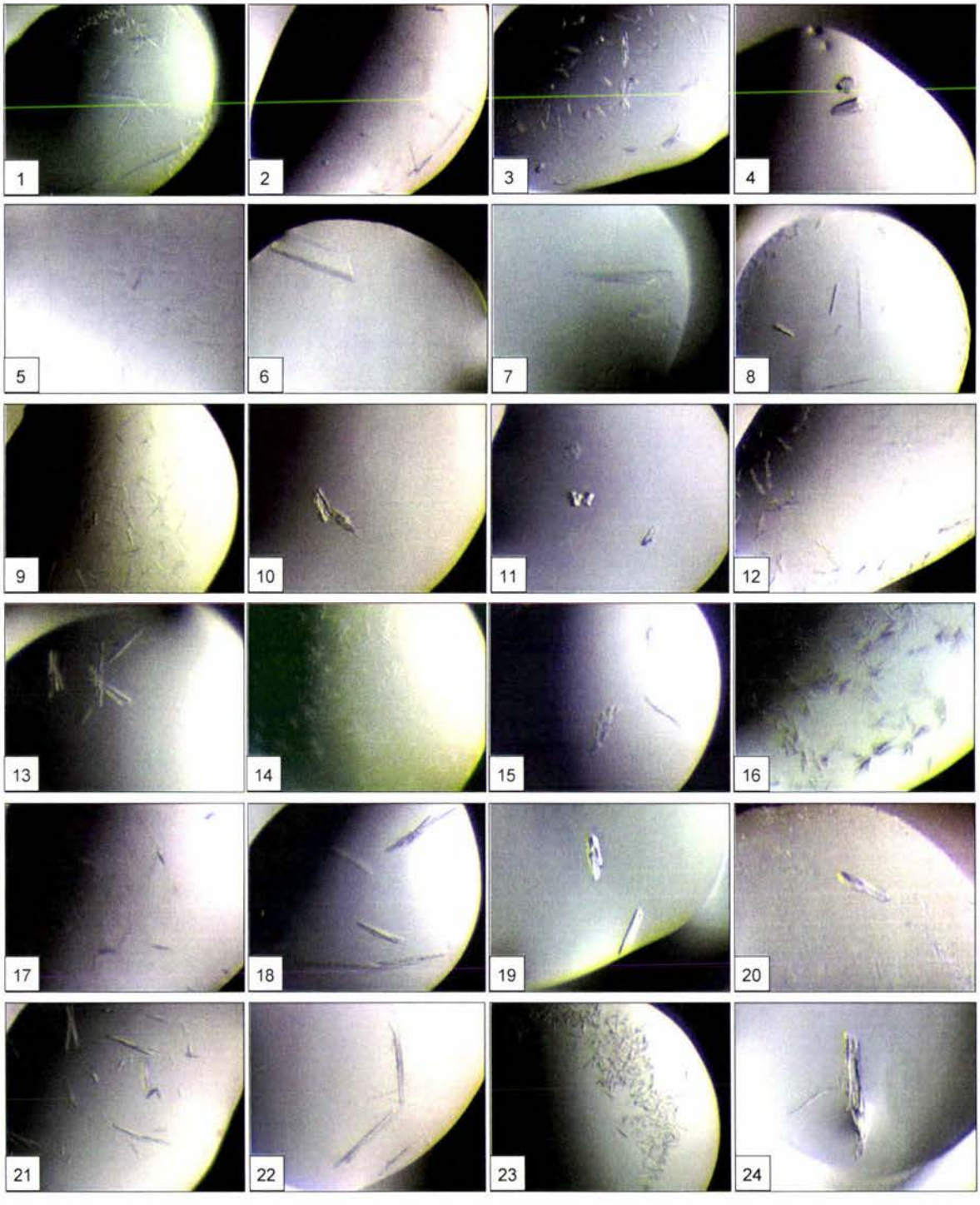


Figure 2.7 PepW Crystals Obtained from Initial Crystallisation Screens at Room Temperature. (1, condition (a); 2, condition (b); 3, condition (c); 4, condition (d); 5, condition (e); 6, condition (f); 7, condition (g); 8, condition (h); 9, condition (i); 10, condition (j); 11, condition (k); 12, condition (l); 13, condition (m); 14, condition (n); 15, condition (o); 16, condition (p); 17, condition (q); 18, condition (r); 19, condition (s); 20, condition (t); 21, condition (u); 22, condition (v); 23, condition (w); 24, condition (x)).

Physical analysis of the crystals revealed that the majority of these crystals were hollow and were extremely fragile. X-ray diffraction analysis of the more stable crystals revealed that none of the crystals diffracted and were therefore not suitable for x-ray crystallography (Chapter 3).

Within 2 months large single crystals had grown in the following 3 conditions:

(a') 20% PEG 3350, 0.2 M ammonium di-hydrogen phosphate
(b') 28% PEG 6000, 0.2 M MES pH 6.1
(c') 40% MPD, 0.1 M MES pH 6.0

The crystals grown in conditions (a') and (c') had similar morphologies, the crystal grown in condition (b') was however considerably different. Images were taken only of condition (a') for fear that heating of the crystal during image capture may damage the crystals (Figure 2.8).

Figure 2.8 PepW Crystal Grown in Condition (a').

The crystals grown in conditions (a') and (c') were very fragile and x-ray analysis revealed that these crystals did not diffract (Chapter 2). The crystal grown in condition (b') was much harder and a concern was that it was a salt crystal. X-ray diffraction analysis however revealed that the crystal grown in condition (b') diffracted to 3.0 A resolution and was suitable for structure determination by x-ray crystallography.

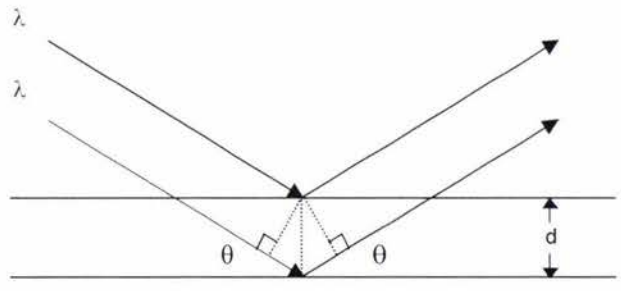
X-RAY DIFFRACTION OF CRYSTALS

3.1 INTRODUCTION

Once crystals have been obtained the next step is to examine their quality and suitability for x-ray crystallography by x-ray diffraction.

Crystals are ordered three-dimensional structures that can be described as a repetition or lattice of identical unit cells. The unit cell is defined by three edge lengths *a*, *b*, *c*, and three interaxial angles α, β, γ; the locations of the atoms within the unit cell are defined by three Cartesian co-ordinates *x*, *y*, *z*. The unit cell is made up of the smallest possible component that, when repeated, is representative of the entire crystal. Although the x-ray diffraction from one unit cell is not significant the repetition of unit cells within a crystal amplifies the diffraction enough to give a detectable signal (Charter 1997).

The diffraction of x-rays by crystals is based on the interference phenomenon described by Braggs Law. Braggs Law states that a radiation will constructively interfere when partially reflected from regularly spaced surfaces (planes) produces a path difference equal to an integral number of wavelengths (Figure 3.1) (Bragg 1913).



Braggs Law: $n\lambda = 2d \sin \theta$

Figure 3.1. Braggs Law of Diffraction, where *n* is an integer (the order of scattering), λ is the wavelength of the radiation, *d* is the spacing between Bragg planes and θ is the angle of scattering (Bragg 1913).

In order for the diffraction to be in phase, the Bragg planes must pass through the same points in all the unit cells in the crystal. There are many possible planes of atoms within the crystal and these are defined by three Miller indices (*h*, *k*, *l*), which are the reciprocals of the intercepts that the set of planes makes with the axes (*a*, *b*, *c*) of the unit cell. For example, a plane parallel to axes a and b will have the Miller indices 001 (or multiple thereof) and a plane with indices 232 means that intercepts are 1/2, 1/3, 1/2 on *a*, *b*, and *c* axes respectively.

Each family of planes is described by one point (*h*, *k*, *l*) in a reciprocal lattice. The reciprocal lattice is an important concept in crystallography and arises from the inverse relationship of Bragg's law between *d* and θ.

All geometrical considerations of diffraction can be rationalized using the concept of the Ewald sphere, which illustrates Bragg's law of diffraction. The Ewald construction uses the concept of the reciprocal lattice to aid in interpreting which crystal planes are in the correct orientation to give constructive interference (Figure 3.2) (Ewald 1921).

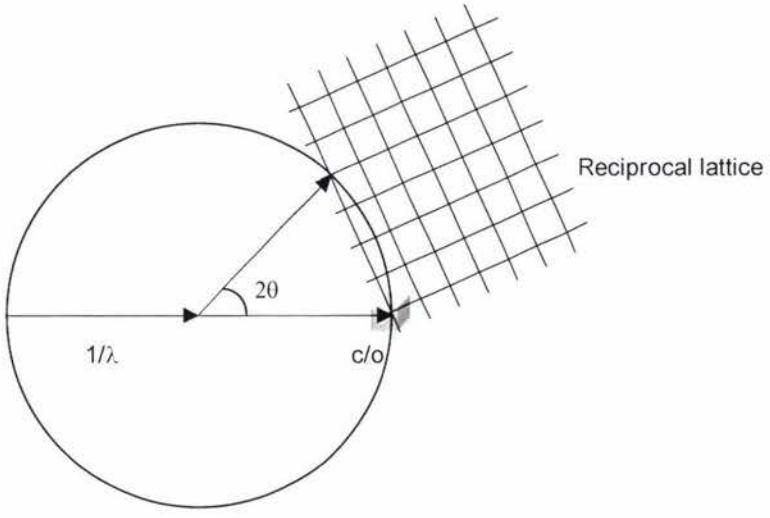


Figure 3.2 The Ewald Construction. The Ewald sphere (with radius 1/λ) and the reciprocal lattice are drawn to the same scale, with the reciprocal lattice origin (o), and the crystal origin (c) located 1/λ from the centre of the circle on the opposite side of the incident beam (Ewald 1921; Rossmann 2001).

When a reciprocal-lattice point intersects the Ewald sphere, it obeys Bragg's law for constructive interference and will appear as a discrete spot in a diffraction pattern, referred to as an *hkl reflection*. To obtain sufficient hkl reflections that will be representative of the three-dimensional positioning of all atoms within the unit cell it is necessary to rotate the crystal (equivalent to rotating the reciprocal lattice) so that all unique reciprocal-lattice points pass through the Ewald sphere. This is done by mounting the crystal on a goniometer head that can be rotated 360° about the phi (ϕ) axis (Figure 3.3).

Initial diffraction images of a crystal are collected with a 0.5° oscillation at phi 0° and 90°. These initial images are necessary to determine firstly whether or not the crystal diffracts, the resolution limit of the diffraction, unit cell dimensions and possible crystal system and symmetry operations.

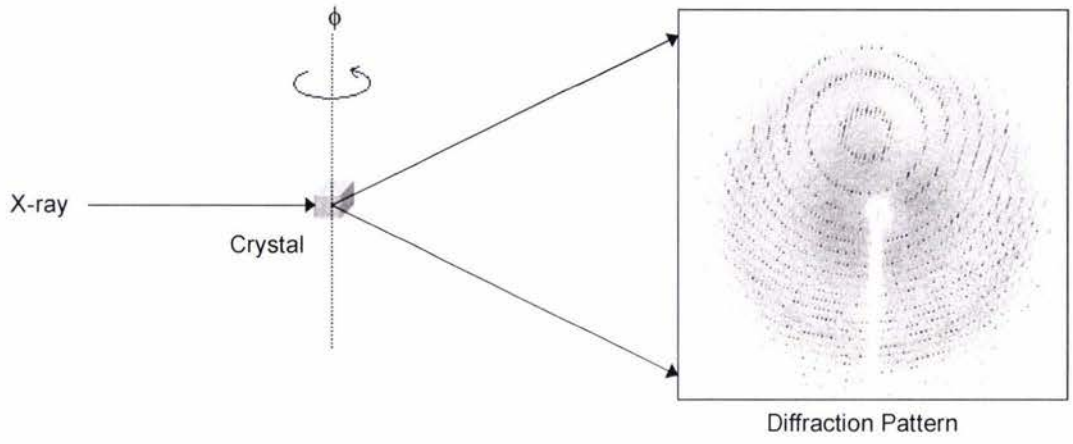


Figure 3.3 Schematic Representation of X-ray Diffraction and Crystal Rotation.

The only reliable way to tell if a crystal will diffract is by x-ray diffraction analysis. It is not uncommon to find a crystal that will not diffract x-rays. This indicates a lack of order within the crystal, and most often a new crystal form must be identified to obtain crystals that diffract. The resolution limit of diffracting crystals is dependent upon the number of unit cells in the x-ray beam and the order of the crystal. In general, large crystals will amplify the diffraction signal more strongly than small crystals, however, this is dependent upon the number and order of unit cells within the crystal. The unit cell dimensions are determined by the spacing between diffraction spots, closely spaced diffraction spots indicate large unit cell axes, while distantly spaced spots indicate small unit cell axes. The crystal system and symmetry operations of a crystal are determined by examination of the symmetry in the diffraction pattern (reciprocal space).

There are seven possible crystal systems which arise from the basic space-filling shapes that unit cells can adopt (Table 3.1) (Rossmann 2001). The seven crystal systems can be further defined by point and space symmetry operations. Symmetry operations including rotation axes, mirror planes, inversion centres and rotation inversion axes, about a point gives rise to the 32 point groups (Table 3.1). The 11 centrosymmetric point groups are called the Laue groups. Space symmetry involves translational symmetry including lattice centring, screw axes and glide planes, and gives rise to the 230 space groups. Of the 230 space groups only 65 are compatible with enantiomorphic biological structures, some of which are more common than others.

Table 3.1 The Crystal Systems and Point Groups

Crystal System	Point Group [a]	Unit Cell Dimensions
Triclinic	1, $\bar{1}$	$a \neq b \neq c$, $\alpha \neq \beta \neq \gamma$
Monoclinic	2, *m*, 2/*m*	$a \neq b \neq c$, $\alpha = \gamma = 90^\circ$, $\beta > 90^\circ$
Orthorhombic	222, 2*mm*, *mmm*	$a \neq b \neq c$, $\alpha = \beta = \gamma = 90^\circ$
Trigonal	3, 3*m*, 32, $\bar{6}2m$, $\bar{3}$	$a = b = c$, $\alpha = \beta = \gamma \neq 90^\circ$
Tetragonal	4, 4*mm*, 4/*m*, 422, $\bar{4}2m$, 4*mmm*, $\bar{4}$	$a = b \neq c$, $\alpha = \beta = \gamma = 90^\circ$
Hexagonal	6, 6*mm*, 6/*m*, $\bar{6}$, 3*m*, 622, 6/*mmm*	$a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$
Cubic	23, *m*3, $\bar{4}3m$, 43, *m*3*m*	$a = b = c$, $\alpha = \beta = \gamma = 90^\circ$

[a] Point group nomenclature: 1-6, rotation axis; $\bar{1}$-$\bar{6}$, rotation inversion axis; *m*, mirror plane; /, axis perpendicular to a mirror plane.

Knowledge of the crystal system and symmetry operations of a crystal can be used to compute a suitable x-ray diffraction data collection strategy. To obtain a complete data set it is necessary to rotate the crystal so that all unique reciprocal-lattice points pass through the Ewald sphere. The minimum rotation required is dictated by the crystal system, space group symmetry and crystal orientation. A highly symmetric crystal system such as the cubic system will produce a highly symmetric diffraction pattern. Therefore a limited portion of the Ewald sphere of reflections will contain all information representative of the unit cell, because all other reflections can be obtained by applying symmetry relationships (Wiencek 1999).

The objective of a data collection strategy is to obtain a complete data set at the highest possible resolution that the crystal will support, with the highest data accuracy. There are many considerations to be made in deciding upon a data collection strategy including: rotation and oscillation range, crystal-to-detector distance, and exposure time for each image.

As already mentioned the rotation range required for a crystal is dependent upon the crystal symmetry. In general a large rotation range will result in multiple measurements of equivalent reflections, leading to more accurate data. The size of the oscillation range is also important, and depends on the size of the unit cell, and crystal-to-detector distance. In general, fine oscillations will give a more accurate profile of the reflections and increase the signal-to-noise (I/σ) ratio, thereby increasing the data accuracy.

The crystal-to-detector distance required for data collection is dependent upon the unit cell dimensions. If a crystal is diffracting well, it would be beneficial to move the detector close to the crystal to capture the high resolution reflections. However, if the unit cell of the crystal is large there will be significant overlap of reflections; in this case it is better to sacrifice the resolution for completeness of the data and set the crystal-to-detector distance so that reflection profiles separate (Dauter 1999). In general, the further away the detector is the less noise and

the better the I/σ ratio of the intensities. However, this is dependent upon air scatter and beam divergence.

Another important consideration is the exposure time for each image. In general a longer exposure time will generate a greater signal-to-noise ratio of the intensities which will increase data quality.

Once a suitable data collection strategy has been established and the data collected, the reflection intensities of each diffraction image are indexed and the mosaicity (degree of crystalline uniformity) of the crystal determined. The lower the mosaicity the more ordered the crystal. The intensities on adjacent images are then integrated, a process by which equivalent reflections are combined. Systematic absences in the data are then determined and a space group assigned. Scaling and averaging is then carried out to correct for systematic errors and a statistical assessment of data quality is finally calculated.

Data quality can be measured in terms of completeness, redundancy, signal-to-noise (I/σ) ratio statistics and the calculated *R:factor*. A high quality data set will be 95-100% complete, will have a high redundancy (number of times that equivalent reflections are measured) and will have a high I/σ ratio. The R-factor (difference in intensities of symmetry-equivalent reflections) provides the most reliable estimation as to the quality of the data, where the lower the R-factor the higher the data quality.

Finally, the corrected intensities are converted to structure factor amplitudes. The intensity of a reflection is proportional to the square of the amplitude of the reflected beam and can be used to calculate what is known as the structure factor. The structure factor *F*(*hkl*) of a reflection is expressed by both the amplitude I*F*(*hkl*)I and the phase φ (*hkl*) of that reflection:

$$F(hkl) = |F(hkl)| \exp [2\pi i \varphi (hkl)]$$

Although the structure factor amplitudes may be calculated from the diffraction pattern, all information concerning the structure factor phases is lost. This is the phase problem of crystallography, and will be discussed in Chapter 4. In order to determine the three-dimensional structure of a molecule, information concerning both the structure factor amplitudes and phases must be obtained.

3.2 METHODS

3.2.1 Selection of a Suitable Cryoprotectant

A suitable cryoptotectant was selected using the Hampton Research Crystal Cryo™ screen when crystals were grown in Hampton Research Crystal screens 1 and 2. Alternatively, cryoprotectants were selected by a process of trial and error which involved incrementally increasing the glycerol concentration in the crystallisation condition until a clear drop was achieved at the desired cryogenic temperature of 110 K.

3.2.2 X-ray Diffraction Apparatus

X-ray diffraction analysis was carried out using a Rigaku MicroMax-007 rotating copper anode x-ray generator (λ = 1.5418 nm, 800 W, 40 kV), with Osmic "Blue" confocal optics, R-Axis IV++ image plate detector, and an Oxford Series 700 cryostream, in conjunction with CrystalClear 1.3.6 and d*TREK software.

3.2.3 X-ray Diffraction Data Collection

X-ray diffraction data were collected at 110 K (section 3.2.2) at a crystal-to-detector distance of either 100 mm or 180 mm, an exposure time of 5.0-14.0 min, with a rotation range of up to 200.0 degrees and an oscillation range of 0.25-0.50 degrees, depending upon the individual crystal.

3.2.4 X-ray Diffraction Data Processing

X-ray diffraction data was primarily processed using CrystalClear 1.3.6 and d*TREK software.

3.3 RESULTS AND DISCUSSION

3.3.1 *L. RHAMNOSUS* PEPO X-RAY DIFFRACTION

3.3.1.1 PRELIMINARY X-RAY DIFFRACTION

Preliminary x-ray diffraction analysis of the PepO crystals grown at room temperature in conditions (a)-(d) of the initial crystallisation screens (Section 2.3.1.1), showed that only those grown in condition (d) diffracted and were of suitable quality for structure determination by x-ray crystallography (Table 3.2).

Table 3.2 X-ray Diffraction Analysis of PepO Crystals Grown in Conditions (d) and (e).

Crystal	Diffraction	Resolution	Unit Cell Dimensions					
			a	b	c	α	β	γ
Condition (d) [a]	✓	2.60 Å	117.72	117.72	206.53	90.0	90.0	90.0
Condition (e) [b]	✓	2.00 Å	117.60	117.60	206.09	90.0	90.0	90.0

[a] Condition (d); 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate at room temperature.
[b] Condition (e); 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate at 4°C.

Crystals grown at room temperature in condition (d) have large unit cells with cell edge lengths of 117.72 Å, 117.72 Å and 206.53 Å, interaxial angles of 90.0°, and diffracted x-rays to 2.60 Å resolution. The larger crystals obtained in condition (e) at 4°C, had the same unit cell dimensions as those grown in condition (d), and strongly diffracted to 2.00 Å resolution (Table 3.2). These crystals were used to collect a native data set (Section 3.3.1.2).

3.3.1.2 PEPO NATIVE DATA SET A

The data collected from crystals grown in condition (e) revealed that the crystals were apparently tetragonal, belonging to point group 422, with cell dimensions of a = b = 117.60 Å and c = 206.09 Å. The long c-axis required the crystal-to-detector distance to be set to 180.0 mm to prevent overlap of the reflections. This crystal-to-detector distance did not significantly compromise the resolution which was limited to 2.0 Å by crystal quality.

Data were collected for several crystals at 110 K with a minimum rotation range of 100.0° with an oscillation range of 0.5°. An exposure time of 10.0 minutes per image was used to enhance the signal-to-noise (I/σ) ratio of the high resolution data and improve data quality.

The data collected indexed and integrated well, and once scaled and averaged produced good statistics. The data collection statistics for two of the best native data sets processed in tetragonal crystal systems are presented in Table 3.3.

Table 3.3 Summary of Data Collection Statistics for PepO Native Data Sets 1 and 2 in Tetragonal Crystal Systems.

	Data Set 1 [a]		Data Set 2 [a]	
Space group	$P42_12$		P422	
Resolution range (Å)	38.90 – 2. 00	(2.07 – 2.00)	39.20 – 2.00	(2.07 – 2.00)
R-merge	0.085	(0.327)	0.061	(0.262)
Completeness (%)	95.2	(70.4)	89.1	(54.6)
I/σ	16.6	(4.3)	13.7	(3.3)
Average redundancy	7.97	(3.89)	2.91	(1.73)
Rotation range (°)	1.0-113.0		1.0 - 194.0	
Oscillation range (°)	0.5		0.5	
Exposure time (min)	10.0		10.0	
Crystal-to-detector distance (mm)	180.0		180.0	
Total number of reflections	743955		254123	
Number of unique reflections	93392		87299	

[a] Values in () are for the last resolution shell.

The native data sets processed in space groups P422 and $P42_12$ (numbers 89 and 90 respectively). The P422 and $P42_12$ space groups each have one 4-fold axis of rotation about the c-axis and two perpendicular 2-fold axes of rotation about the a- and b-axes. Both crystals are expected to have the same space group.

The native data sets gave good overall R-merge (difference in intensities of symmetry-equivalent reflections) values between 6.0-9.0%. The I/σ ratio, average redundancy and completeness were all very good. The completeness of Data Set 1 was very good and the completeness of Data Set 2 was also good considering that this crystal was orientated such that one of the axes was approximately perpendicular to the x-ray beam. Efforts were made to merge data sets together to improve the completeness and average redundancy; however, it proved too impossible to do so.

Data Sets 1 and 2 were further processed and were found to be twinned (Chapter 4). Twinning is colloquially a crystal growth anomaly in which multiple crystalline domains are oriented to form a single crystal. Although software exists to unravel twinned data, the data proved too impossible to process successfully despite many efforts to do so.

Extensive attempts that were made to solve the PepO structure using Data Sets 1 and 2 without correcting for twinning, by Molecular Replacement (MR) and Multiple Isomorphous Replacement (MIR) techniques (Chapter 4), produced no conclusive results. The inability to solve the structure was initially thought to be due to incorrect space group assignment as a result of twinning. The Data Sets were consequently reprocessed in a lower symmetry space group. The data collection statistics for Data Sets 1 and 2 reprocessed in monoclinic crystal systems are presented in Table 3.4.

Table 3.4 Summary of Data Collection Statistics for PepO Native Data Sets 1 and 2 in Monoclinic Crystal Systems.

	Data Set 1 [a]		Data Set 2 [a]	
Space group	P2		$P2_1$	
Resolution range (Å)	39.22 - 2.00	(2.07 - 2.00)	39.21 - 2.00	(2.07 - 2.00)
R-merge	0.068	(0.278)	0.055	(0.257)
Completeness (%)	88.0	(47.3)	86.3	(41.8)
I/σ	10.2	(2.7)	11.1	(2.9)
Average redundancy	2.23	(1.49)	1.95	(1.42)
Rotation range (°)	1.0-113.0		1.0 - 194.0	
Oscillation range (°)	0.5		0.5	
Exposure time (min)	10.0		10.0	
Crystal-to-detector distance (mm)	180.0		180.0	
Total number of reflections	743188		636650	
Number of unique reflections	332852		326295	

[a] Values in () are for the last resolution shell.

The Data Sets processed in space groups P2 and $P2_1$ (numbers 17 and 18 respectively). Because the data was collected as tetragonal and converted to a lower symmetry the number of unique reflections recorded and the average redundancy decreased significantly. The % completeness decreased considerably for Data Set 1, but decreased only slightly for Data Set 2, due to the orientation of the crystal relative to the x-ray beam. The data-quality as measured by the R-merge value appeared to improve. Although a decrease in R-merge was expected due to lowering the number of symmetry-equivalent reflections it is difficult to establish whether the difference in R-merge is significant.

Further unsuccessful attempts were made to solve the PepO structure with the monoclinic native data sets (Chapter 4).

3.3.1.3 PEPO DERIVATIVE DATA

Prior to establishing that the crystals were twinned, heavy atom derivative crystals were prepared (Section 2.3.1.3) and analysed by x-ray diffraction (Table 3.5) in order to identify suitable derivatives for Multiple Isomorphous Replacement (MIR) (Chapter 4).

Table 3.5 X-ray Diffraction Analysis of PepO Derivative Crystals.

Derivative	Diffraction	Resolution	Isomorphous *	Unit Cell Dimensions					
				a	b	c	α	β	γ
Native	✓	2.00 Å	N/A	117.72	117.72	206.53	90.0	90.0	90.0
$C_9H_9HgO_2SNa$ [1]	✓	2.80 Å	✓ (i)	118.06	118.06	205.54	90.0	90.0	90.0
$C_9H_9HgO_2SNa$ [2]	✓	2.80 Å	✓ (ii)	114.38	114.89	204.09	90.0	90.0	90.0
CH_3CH_2HgCl [3]	✓	3.20 Å	✓ (iii)	118.19	118.19	206.16	90.0	90.0	90.0
CH_3CH_2HgCl [4]	✓	3.50 Å	x	83.84	83.84	205.95	90.0	90.0	90.0
CH_3CH_2HgCl [5]	✓	3.00 Å	x	53.72	53.72	1835.98	90.0	90.0	120.0
$UO_2CH_3CO_2$ [6]	✓	3.00 Å	✓ (iv)	117.44	117.44	206.33	90.0	90.0	90.0
$UO_2CH_3CO_2$ [7]	✓	3.00 Å	x	80.55	80.55	203.53	90.0	90.0	90.0
$UO_2CH_3CO_2$ [8]	✓	3.50 Å	✓	120.34	120.34	207.0	90.0	90.0	90.0
K_2PtCl_4 [9]	✓	4.00 Å	✓	118.05	118.08	202.25	90.0	90.0	90.0
$Pr(NO_3)_3$ [10]	✓	3.10 Å	x	98.83	115.65	136.49	90.0	90.0	90.0

[1] $C_9H_9HgO_2SNa$: 100 mg/mL $C_9H_9HgO_2SNa$, in condition (e), 1 day at 4°C.

[2] $C_9H_9HgO_2SNa$: 100 mg/mL $C_9H_9HgO_2SNa$, in condition (e), 1 week at 4°C.

[3] CH_3CH_2HgCl: 100 mg/mL CH_3CH_2HgCl, in condition (e), 6 hours at 4°C.

[4] CH_3CH_2HgCl: 100 mg/mL CH_3CH_2HgCl, in condition (e), 1 day at 4°C.

[5] CH_3CH_2HgCl: 100 mg/mL CH_3CH_2HgCl, in condition (e), 1 week at 4°C.

[6] $UO_2CH_3CO_2$: 100 mg/mL $UO_2CH_3CO_2$, in condition (e), 1 day at 4°C.

[7] $UO_2CH_3CO_2$: 100 mg/mL $UO_2CH_3CO_2$, in condition (e), 1 week at 4°C.

[8] $UO_2CH_3CO_2$: 100 mg/mL $UO_2CH_3CO_2$, in condition (e), 1 week at 4°C.

[9] K_2PtCl_4: 100 mg/mL K_2PtCl_4, in condition (e), 8 days at 4°C.

[10] $Pr(NO_3)_3$: 100 mg/mL $Pr(III)NO_3$, in condition (e), at 4°C.

* (i), (ii), (iii) and (iv), Derivatives for which data sets were collected; N/A, Not Applicable.

An early study by Crick and Magodoff (1956) proposed a rule of thumb that a change in any of the cell dimensions by more than 5.0% would result in a lack of isomorphism (Crick 1956; Rossmann 2001). According to this 6 of the 10 derivatives (Table 3.5) are potentially isomorphous with the native data. In addition to deviations in the cell dimensions the resolution limit should not substantially decrease. Typically the resolution limit of the derivative crystal should not decrease by more than 25% of the native crystal resolution limit (Drenth 1999). Unfortunately the resolution limit for all of the derivative crystals exceeded a 25% decrease in resolution.

To investigate whether or not the potentially isomorphous derivatives obtained would be suitable for MIR complete data sets were collected for 4 of the 6 derivatives. The data collection statistics for the 4 derivative data sets are presented in Table 3.6.

The derivative data sets although having unit cell dimensions that deviated by less than 5.0% of the native data sets, processed in a range of space groups; C2, P2, $P2_1$ and $P4_1$. The inaccuracy of the space group assignment is reflected by the poor data collection statistics. Although the data sets had high % completeness (with the exception of $UO_2CH_3CO_2$ [6]), they had very high R-merge values in the range of 13.0-21.1%, low resolution, low average redundancy and I/σ values, representative of low quality data sets.

The derivative data sets were further processed to determine if they could be used to obtain initial phases using Multiple Isomorphous Replacement (MIR). Unfortunately, the derivatives were not suitable for structure determination by MIR (Chapter 4).

One approach that was not taken, but may prove successful, is to collect native and derivative data sets from the same crystal. This technique has been successful in the structure determination of cyclohydrolase (Nar 1995), proteosome (Lowe 1995), and a number of other proteins (Rossmann 2001).

Table 3.6 Summary of Data Collection Statistics for PepO Derivative Data Set Crystals.

	(i) $C_9H_9HgO_2SNa$ [a]		(ii) $C_9H_9HgO_2SNa$ [a]		(iii) CH_3CH_2HgCl [a]		(iv) $UO_2CH_3CO_2$ [a]	
Space group	P2		C2		$P4_1$		$P2_1$	
Unit Cell	a = b = 118.06, c = 205.54		a =114.38, b = 114.89, c = 204.09		a = b = 118.19, c = 206.16		a = b = 117.44, c = 206.33	
	$\alpha = \beta = \gamma = 90.0$		$\alpha = \beta = \gamma = 90.0$		$\alpha = \beta = \gamma = 90.0$		$\alpha = \beta = \gamma = 90.0$	
Resolution range (Å)	39.35 - 2.80	(2.90 - 2.80)	50.06 - 2.80	(2.90 - 2.80)	39.04 - 3.20	(3.31 - 3.20)	39.15 - 3.00	(3.11 - 3.00)
R-merge	0.174	(0.409)	0.132	(0.308)	0.200	(0.384)	0.211	(0.319)
Completeness	97.7	(98.2)	97.7	(96.6)	99.2	(99.8)	58.1	(60.6)
I/σ	4.3	(1.8)	5.4	(2.9)	5.4	(3.1)	3.4	(2.3)
Average redundancy	1.99	(1.94)	2.43	(2.35)	3.83	(3.62)	1.65	(1.57)
Rotation range (°)	1.0-94.0		1.0-117.0		1.0-94.0		1.0-95.0	
Oscillation range (°)	0.5		0.5		0.5		0.5	
Exposure time (min)	10.0		10.0		10.0		10.0	
Crystal-to-detector distance (mm)	180.0		180.0		180.0		180.0	
Total number of reflections	270017		154263		176749		107683	
Number of unique reflections	135965		63591		46103		65291	

(i) $C_9H_9HgO_2SNa$: 15% PEG 8000, 0.1 M $Na(CH_3)_2AsO_2$ pH 6.5, 0.2 M $CaCH_3CO_2$, 100 mg/mL $C_9H_9HgO_2SNa$, 1 day at 4°C.

(ii) $C_9H_9HgO_2SNa$: 15% PEG 8000, 0.1 M $Na(CH_3)_2AsO_2$ pH 6.5, 0.2 M $CaCH_3CO_2$, 100 mg/mL $C_9H_9HgO_2SNa$, 1 week at 4°C.

(iii) CH_3CH_2HgCl: 15% PEG 8000, 0.1 M $Na(CH_3)_2AsO_2$ pH 6.5, 0.2 M $CaCH_3CO_2$, 100 mg/mL CH_3CH_2HgCl, 6 hours at 4°C.

(iv) $UO_2CH_3CO_2$: 15% PEG 8000, 0.1 M $Na(CH_3)_2AsO_2$ pH 6.5, 0.2 M $CaCH_3CO_2$, 100 mg/mL $UO_2CH_3CO_2$, 1 day at 4°C.

[a] Values in () are for the last resolution shell.

3.3.1.4 PRELIMINARY X-RAY DIFFRACTION

Preliminary x-ray diffraction analysis of the crystals grown in condition (g) (Section 2.3.1.4), identified these not as protein but salt crystals. Salt crystals can easily be recognised by diffraction patterns in which strong diffraction spots are observed at high but not low resolution. This is because salt crystals have much smaller unit cells than protein crystals.

X-ray diffraction analysis was not carried out for the crystals grown in conditions (a')-(o') (Section 2.3.1.5) because there was not sufficient time to do so, and because a suitable crystal had by that time been identified.

Preliminary x-ray diffraction analysis of the single crystal grown in condition (a'') (Section 2.3.1.6) showed that this crystal strongly diffracted x-rays and was of a different crystal form and unlikely to be twinned (Table 3.7).

Table 3.7 X-ray Diffraction Analysis of PepO Crystals Grown in Condition (a'') and (g).

Crystal	Diffraction	Resolution	Unit Cell Dimensions					
			a	b	c	α	β	γ
Condition (a'') [a]	✓	1.80 Å	56.32	83.75	143.82	90.0	90.0	90.0
Condition (g) [b]	x	N/A	N/A	N/A	N/A	N/A	N/A	N/A

[a] Condition (a''); 1.5 M lithium sulfate monohydrate, 0.1 M sodium hepes pH 7.5 at room temperature.

[b] Condition (g); 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate, 0.5 mM phosphoramidon at 4°C.

N/A, Not Applicable.

The crystal grown in condition (a'') had a much smaller unit cell than the crystals used to obtain Native Data Sets 1 and 2, with cell edge lengths of 56.32 Å, 83.75 Å and 143.8 Å, interaxial angles of 90.0°, and diffracted x-rays to 1.80 Å resolution. This crystal was used to collect a native data set (Section 3.3.1.5).

3.3.1.5 PEPO NATIVE DATA SET B

The data collected from the crystal grown in condition (a'') revealed that the crystal was orthorhombic, belonging to point group 222, with cell dimensions of a = 56.32 Å, b = 83.75 Å and c = 143.82 Å. To prevent overlap of the reflections the crystal-to-detector distance was set to 180.0 mm; this, however, restricted the collection of the high resolution data. To collect as much of the high resolution data as possible the data collection was carried out over a rotation

of 100° with an oscillation range of 0.3° at 110 K. An exposure time of 5.0 minutes per image gave a good I/σ ratio for the high resolution data. The data collected indexed and integrated well to a resolution limit of 1.85 Å; however, once scaled and averaged the data produced poor merging statistics. To improve the statistics, the data were integrated with a resolution limit of 2.20 Å, and scaled and averaged. The data collection statistics for PepO Native Data Set 3 are presented in Table 3.8.

Table 3.8 Summary of Data Collection Statistics for PepO Native Data Set 3 in an Orthorhombic Crystal System.

	Data Set 3 [a]	
Space group	P222	
Resolution range (Å)	39.21 – 2.20	(2.28 – 2.20)
R-merge	0.063	(0.085)
Completeness (%)	96.4	(95.7)
I/σ	15.2	(10.1)
Average redundancy	3.78	(3.61)
Rotation range (°)	1.0-77.0	
Oscillation range (°)	0.3	
Exposure time (min)	5.0	
Crystal-to-detector distance (mm)	180.0	
Total number of reflections	129359	
Number of unique reflections	34188	

[a] Values in () are for the last resolution shell.

Native Data Set 3 processed in space group P222 (number 16), was 96.4% complete with an R-merge value of 6.3% to 2.20 Å resolution. The I/σ ratio, average redundancy and completeness were all indicators of a high quality data set.

Native Data Set 3 was used to successfully obtain some initial phase information by Molecular Replacement (MR) (Chapter 4).

3.3.2 *L. RHAMNOSUS* PEPW X-RAY DIFFRACTION

Preliminary x-ray diffraction analysis of the PepW crystals was very difficult, as the crystals were very small and extremely fragile. This made mounting them problematic, and only 10 of the PepW crystals (Section 2.3.2.1) were analysed by x-ray diffraction. Of these crystals only 1 diffracted and this was the crystal grown in condition (b') (Section 2.3.2.1) (Table 3.9).

Table 3.9 X-ray Diffraction Analysis of PepW Crystals Grown in Condition (b').

Crystal	Diffraction	Resolution	Unit Cell Dimensions					
			a	b	c	α	β	γ
Condition (b') [a]	✓	3.00 Å	142.22	142.22	98.27	90.0	90.0	120.0

[a] Condition (b'); 28% PEG 6000, 0.2 M MES pH 6.1 at room temperature.

The crystal grown in condition (b') had unit cell dimensions of 142.22 Å, 142.22 Å and 98.27 Å, interaxial angles of 90.0°, 90.0° and 120.0° and diffracted x-rays to 3.00 Å resolution. This crystal was used to collect a native data set (Section 3.3.2.1).

3.3.2.1 PEPW NATIVE DATA SET A

The data collected from the crystal grown in condition (b') revealed that the crystal was trigonal/hexagonal, belonging to point group 3 and space group P3 (number 143), with unit cell dimensions of a = b = 142.22 Å, and c = 98.27 Å.

The unit cell dimensions and resolution limits allowed data collection to be carried out at a crystal-to-detector distance of 100.0 mm. Data was collected at 110 K over a rotation range of 82.0° with an oscillation range of 0.25°. An exposure time of 14.0 minutes per image was used to enhance the I/σ ratio of the high resolution data and improve data quality.

Unfortunately the crystal deteriorated during the data collection due to malfunctions associated with the cryostream. The data collected indexed and integrated poorly, and it was not possible to continue with data processing.

SOLVING THE PHASE PROBLEM

4.1 INTRODUCTION

Once diffraction data have been collected the next step is to determine the three-dimensional crystal structure. If both the structure factor amplitudes and the phases were known, then the electron density of the protein can be directly calculated by the Fourier transformation:

$$\rho(xyz) = 1/v \sum \sum \sum |F(hkl)| \exp[2\pi i \varphi(hkl)] \exp[-2\pi i (hx + ky + lz)]$$

where $\rho(xyz)$ is the electron density in 'real space' at the fractional Cartesian coordinates (x, y, z) along the unit cell axes, $|F(hkl)|$ and $\varphi(hkl)$ are the structure factor amplitude and phase of the reflection respectively, with the specified Miller indices (h, k, l) in reciprocal space, and v the volume of the unit cell.

In order to perform the Fourier transformation the complex structure factors are required. Although the structure factor amplitudes may be calculated from the intensities of the diffraction pattern, all information concerning the structure factor phases is lost. This is the underlying phase problem of crystallography.

In 1934 Patterson presented a method for locating the atomic positions in small molecules without knowledge of the phase angles (Patterson 1934). This method uses the structure factor amplitudes in a Fourier transform called the Patterson Function:

$$P(uvw) = 1/v \sum \sum \sum |F(hkl)|^2 \cos[2\pi i (hu + kv + lw)]$$

Where $P(uvw)$ is the Patterson function in 'Patterson space' at coordinates (u, v, w) along the unit cell axes, $|F(hkl)|^2$ is the square of the structure factor amplitude, which is equivalent to the measured intensities and the expression $\cos[2\pi i (hu + kv + lw)]$ is the frequency factor involving the products of the coordinates (h, k, l and u, v, w) in both 'reciprocal' and 'Patterson space'.

The Patterson function is essentially a convolution of the structure factor, and generates a Patterson map of vectors between atoms (Figure 4.1). Patterson maps consist of intramolecular (from one atom in the molecule to another atom in the same molecule) and intermolecular vectors (from an atom in one molecule to an atom in a different molecule) originating from a single point, the origin of the Patterson function. Because the peaks represent both positive and negative vectors, Patterson maps are always centrosymmetric (Figure 4.1).

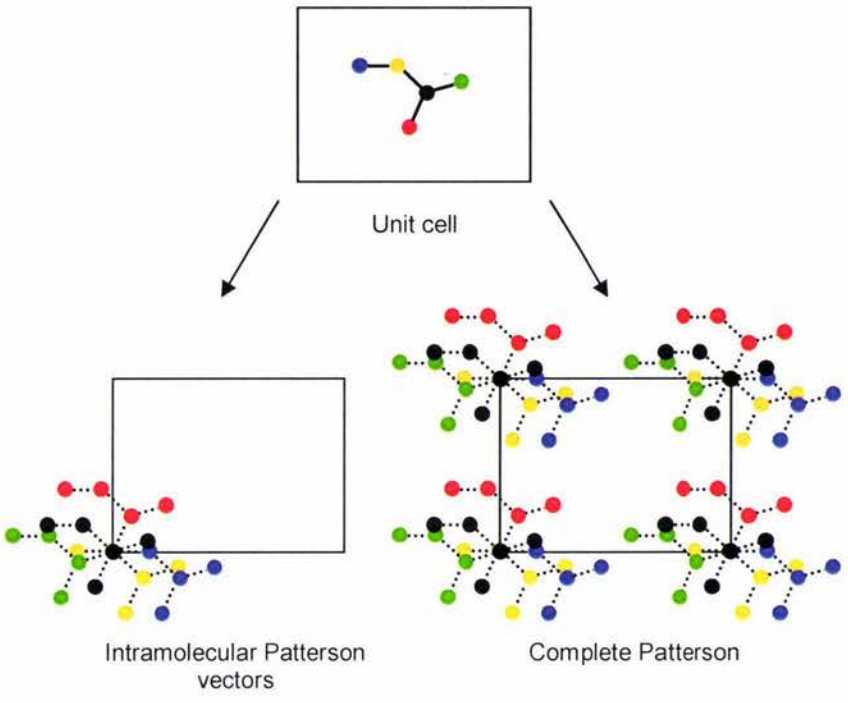


Figure 4.1 The Patterson Map. The construction of a Patterson map can be regarded as consecutively placing each atom of the molecule at the Patterson origin and drawing vectors from that atom to all other atoms in the unit cell. (Colours are representative of the vectors resulting from when that coloured atom in the unit cell is placed at the Patterson origin).

The symmetry operations of a crystal are often represented in a Patterson map by clusters of vector peaks on specific lines or planes of the map, called Harker lines and sections (Giacovazzo 2002). The Patterson function must be deconvoluted to determine the positions of the atoms within the unit cell and this can be done by direct methods, molecular replacement, multiple isomorphous replacement and multiple wavelength anomalous dispersion methods.

4.1.1 MOLECULAR REPLACEMENT

Molecular Replacement (MR), first described by Rossmann and Blow (1962), uses a search model from which estimates of the structure factor phases can be computed. A suitable search model is chosen based on the probability of three-dimensional structural similarity with the protein of unknown structure. Structural similarity correlates well with sequence identity above a threshold of approximately 30% amino acid sequence identity. It becomes progressively more difficult to obtain phase information by MR methods as the sequence identity decreases. The use of MR is expected to increase as more structures are solved and deposited in the Protein

Data Bank (PDB) (Berman 2000) because it will be more likely that an accurate search model will be available for a given MR problem.

The objective of MR is to orient and position the search model such that it coincides with the position in the unit cell of the unknown protein. In the simplistic case where only one molecule is present in the asymmetric unit, six variables, three rotational and three translational, will approximately describe the transformation required to orient and position the search model. There have been many attempts to improve and/or alter the way molecular replacement is done including real and reciprocal space searches. However, traditional MR methods use a Patterson-based approach to do this.

The Patterson function can be calculated for both the search model and for the unknown protein structure. When the model is oriented and positioned correctly in the unit cell of the unknown protein structure the two Patterson maps should coincide. As mentioned above the Patterson map consists of intramolecular and intermolecular vectors, and these can be used separately to compute the correct orientation and position respectively. The intramolecular vectors depend only upon the orientation of the molecule and not the position in the unit cell, therefore these can be exploited in what is known as a rotation function. The intermolecular vectors depend both on the orientation of the molecule and its position within the unit cell. Once the orientation is known the intermolecular vectors can be exploited in what is known as a translation function (Figure 4.2).

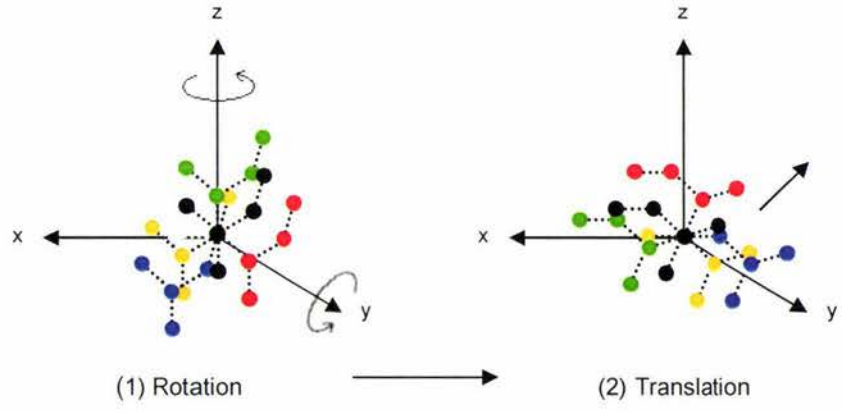


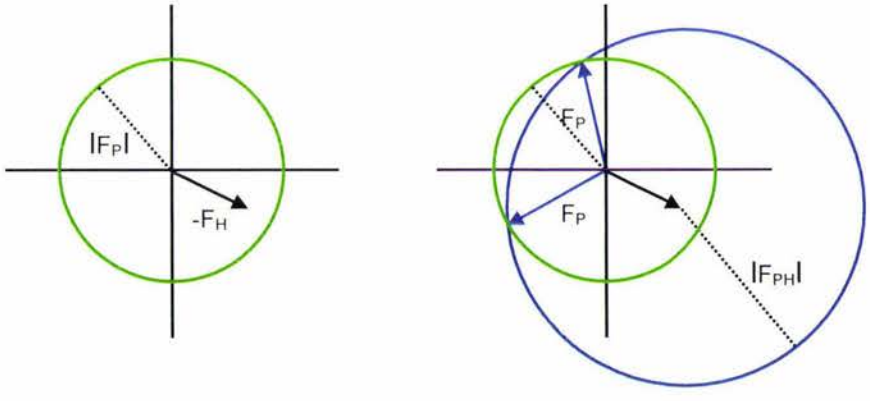
Figure 4.2 The Molecular Replacement Technique. (1) Rotation function; the intramolecular vectors of the crystal structure (coloured) are rotated such that they overlay with the intramolecular vectors of a search model (grey). (2) Translation function; the intermolecular vectors of the crystal structure are then used to determine the location of the molecule within the unit cell.

An initial set of phases can be determined by placing the known structure into the unit cell of the unknown structure using the rotation and translation operations obtained. These phases combined with the observed amplitudes can be used to generate an electron density map using the Fourier transformation.

4.1.3 MULTIPLE ISOMORPHOUS REPLACEMENT

Multiple Isomorphous Replacement (MIR) is a method that obtains phase information by comparison of a native data set with multiple isomorphous derivative data sets. Isomorphous derivatives should be identical to the native protein structure except for the presence of one or more heavy atoms bound to the protein molecules. The lattice, space group, cell dimensions, position and conformation of the protein molecule within the unit cell should be preserved. The coordinates of the heavy atom(s) can be derived via either direct methods or Patterson methods.

The Patterson-based method involves the construction of a difference Patterson where the only remaining vectors are those of the heavy atom(s). From the difference Patterson the position of the heavy atom in the unit cell can be determined. The vector of the heavy atom F_H combined with the amplitudes of the native data set I F_P I and derivative data set(s) I F_{PH} I can be used to construct a Harker diagram from which the phase can be solved (Figure 4.3).



(1) Native Data (2) Native and Derivative Data

Figure 4.3 The MIR Harker Diagram. (1) Native Data; the Harker diagram is constructed by drawing a circle (green) of radius $|F_P|$ to indicate all possible positions for the native vector, F_P. (2) Native and Derivative Data; from the native data circle centre the heavy atom vector, $-F_H$, is drawn, and from its end point a circle (blue) of radius $|F_{PH}|$ is drawn to indicate all possible positions for the vector F_{PH}. The intersection of the two circles corresponds to the possible vectors, F_P.

As illustrated in the Harker diagram a single heavy atom derivative gives two equally possible protein phase solutions. This ambiguity can be solved if a different heavy atom derivative is available because the correct solution will coincide in both cases. Conversely, the anomalous scattering (AS) data of the first derivative can be utilised to resolve the phase ambiguity.

4.1.4 ANOMALOUS SCATTERING

Anomalous Scattering (AS) of x-ray radiation by heavy atoms results in small but measurable differences in intensity between reflections hkl and -h-k-l, and leads to a breakdown of Friedel's law and can be used in combination with MIR to resolve the phase ambiguity. If two or more derivatives and anomalous data are used, then it is called Multiple Isomorphous Replaceent with Anomalous Scattering (MIRAS) (Howard 2002; Rhodes 1993; Rossmann 2001).

The MIRAS Patterson-based method involves the construction of a difference Patterson where from which the phase can be solved (Figure 4.4).

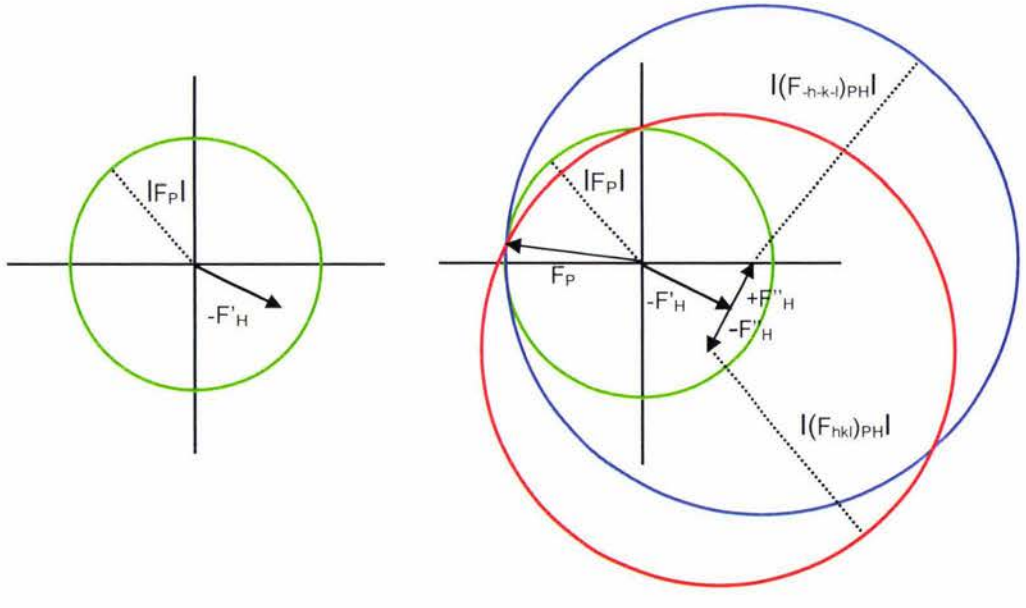


Figure 4.4 The MIRAS Harker Diagram. (1) Native Data; the Harker diagram is constructed as described in Figure 4.3. (2) Native and Derivative Data; from the native data circle centre the heavy atom vector, $-F'_H$, is drawn, and from its end point the anomalous scattering data, $-F''_H$ and $+F''_H$, are drawn, and from their end points a circle (blue) of radius $|(F_{-h-k-l})_{PH}|$, and a circle (red) of radius $|(F_{hkl})_{PH}|$ are drawn. The intersection of the three circles corresponds to the vector, F_P. (Modified from Rhodes 1993).

4.1.5 SOLVING THE PHASE PROBLEM

Solving the phase of the structure factor is the underlying problem of x-ray crystallography. Once the structure factor phases have been established a three-dimensional electron density map is generated, from which a protein structure model can be established. The model is then refined and validated to give an accurate three-dimensional structure.

4.2 METHODS

4.2.1 Computer Software

Data processing of the native and derivative data sets was carried out using the Collaborative Computational Project Number 4 (CCP4) Version 5.0.2 suite of programs including: AMoRe, CAD, Dtrek2mtz, Hklview, Matthews_coef, Scaleit and Truncate, in conjunction with the CCP4 graphical user interface (CCP4i) Version 1.3.20 on a Unix operating system (Bailey 1994; Potterton 2003). Additional non-CCP4 programs used included the Yeates Merohedral Crystal Twinning Server (Yeates 1997), Turbo-Frodo (Roussel 1991), Swiss-Model Version 3.5 (Sayle 1995; Schwede 2003).

4.2.2 Structure Prediction

Three-dimensional structure predictions of PepO and PepW were carried out using Swiss-Model Version 3.5, a fully automated protein structure homology-modelling server, accessible via the Expasy web server (Gasteiger 2003; Schwede 2003).

4.2.3 File Conversions

The Dtrek ascii output from the Crystal Clear 1.3.6 Software was converted to pseudo-structure factors (F), using the CCP4 programs, Dtrek2mtz and Truncate (French 1978; Bailey 1994). Dtrek2mtz converts the ascii output into a binary mtz format, which is compatible with other CCP4 programs. Truncate takes the intensities (I) and standard deviations (σI) from the Dtrek2mtz output, corrects for negative values by putting them on an absolute scale and converts them to the corresponding pseudo-structure factors (F) and standard deviations (σF).

4.2.4 Preliminary Data Analysis

The diffraction data in mtz format was processed using the CCP4 programs Truncate, Matthews_coef and HKLview (Matthews 1968; French 1978; Bailey 1994). Truncate was used to plot a Wilson distribution of the logarithm of the intensity values against resolution in order to calculate temperature correction factors, and to plot a Cumulative Intensity-distribution of the intensity data to identify possible twinning. Matthews_coef was used to calculate the number of molecules present in the asymmetric unit, and Hklview was used to interpret the intensity data of pseudo-precession images and establish correct space groups.

4.2.5 Automated Molecular Replacement (AMoRe)

Molecular Replacement was carried out using the CCP4 program Automated Molecular

Replacement (AMoRe) (Navaza 2001). AMoRe uses the (F) and (SigF) values in mtz format, with knowledge of the number of molecules per asymmetric unit, to perform tabling, rotation and translation functions against a search model pdb file. Tabling is a process that converts a pdb file into a set of structure factors and involves: translation of the pdb file such that the molecules' centre of mass is at the Patterson origin, construction of a P1 (triclinic crystal system) box of sufficient size around the Patterson vectors to make fractional co-ordinates (x,y,z) for each atom in the molecule, and calculation of the structure factors (Navaza 2001). AMoRe automatically performs rotation and translation functions against the search models and calculates the best solutions in terms of the correlation coefficient. The AMoRe rotation function was carried out using a 3.0 Å maximum resolution, and B-factor of 30.0, the translation function was carried out using default settings. The output of the rotation and translation functions are three Eulerian angles (α, β, γ), three translational co-ordinates (Tx, Ty, Tz) and measurements of how well these values match the Patterson maps from the diffraction data and search models. The *H. sapiens* neutral endopeptidase (NEP) (PDB accession 1DMT) and Swiss-Model generated PepO structure were used as search models.

4.2.6 Multiple Isomorphous Replacement

Multiple Isomorphous Replacement (MIR) was attempted using the CCP4 programs: Collect Assorted Data (CAD), and Scaleit (Howell 1992; Bailey 1994). CAD combines native and derivative data sets into one file, while Scaleit puts the data sets on the same scale using an overall scale factor and B-factor. The output of Scaleit provides useful statistics, including the weighted R-factor, that can be used to establish if the data sets are isomorphous and whether or not they are suitable for MIR.

4.2.7 Visual Analysis of Structures

Rasmol Version 2.7.1.1 (Sayle 1995) was used to analyse the Swiss-Model generated structures, and the CCP4 program Turbo-Frodo was used to visually analyse non-refined crystal structures (Jones 1978).

4.3 RESULTS AND DISCUSSION

4.3.1 *L. RHAMNOSUS* PEPO PHASE PROBLEM

4.3.1.1 PEPO NATIVE DATA SET A

PepO Native Data Set 1 and 2 (Section 3.3.1.2) intensities (I) were converted to pseudo-structure factors (F), using the CCP4 programs Dtrek2mtz and Truncate (French 1978; Bailey 1994). The pseudo-structure factors (F) of Native Data Sets 1 and 2 were shown to follow the Wilson distribution (Figure 4.5).

The Wilson distribution for a macromolecule should deviate from a straight line above 3.0 Å resolution with a characteristic dip at 5.5 Å and a peak at 4.0 Å resolution. The intensity peak at 4.0 Å resolution can be ascribed to the presence of many non-bonded distances of approximately 4.0 Å in the protein, and the decrease in intensity at 5.5 Å is due to the less ordered solvent content relative to the crystal content of the unit cell (Drenth 1999). The linear relationship between the logarithm of intensity against resolution beyond 3.0 Å is used to determine an absolute scale and temperature correction factor (French 1978; Drenth 1999; Rossmann 2001). The temperature correction factor, *B:factor*, is a measurement of the atomic vibration, orientational and positional disorder within the crystal, where the higher the B-factor the greater the atomic vibration and disorder. Both Data Sets 1 and 2 had characteristic Wilson distributions and relatively low B-factors of 24.0 and 25.3 respectively, (Figure 4.5).

In addition, Truncate was used to analyse the intensity statistics to identify possible twinning. Twinning is colloquially a crystal growth anomaly in which multiple crystalline domains are oriented to form a single crystal. There are two fundamentally different types of twinning, epitaxial and merohedral (Yeates 1997). Epitaxial twinning occurs when different crystalline domains are oriented such that the diffraction patterns are not superimposable. Whereas, merohedral twinning occurs when different crystalline domains are oriented such that they coincide exactly in three-dimensions generating superimposable diffraction patterns (Yeates 1997).

The possibility of twinning was initially overlooked as it was not suggested by the diffraction pattern, scale and averaged statistics nor the Wilson distribution. To identify twinning Cumulative Intensity-distributions were generated for both the theoretical and observed centric (with phase restrictions) and acentric (without phase restriction) intensity data (Figure 4.6) (Howells 1950). The theoretical acentric intensity data are evenly distributed and therefore follow a near linear distribution. Twinned acentric intensity data, however, are not evenly distributed and follow a sigmoidal distribution. Under twinning conditions each reflection (intensity) is contributed to by more than one crystalline domain. It is unlikely that both

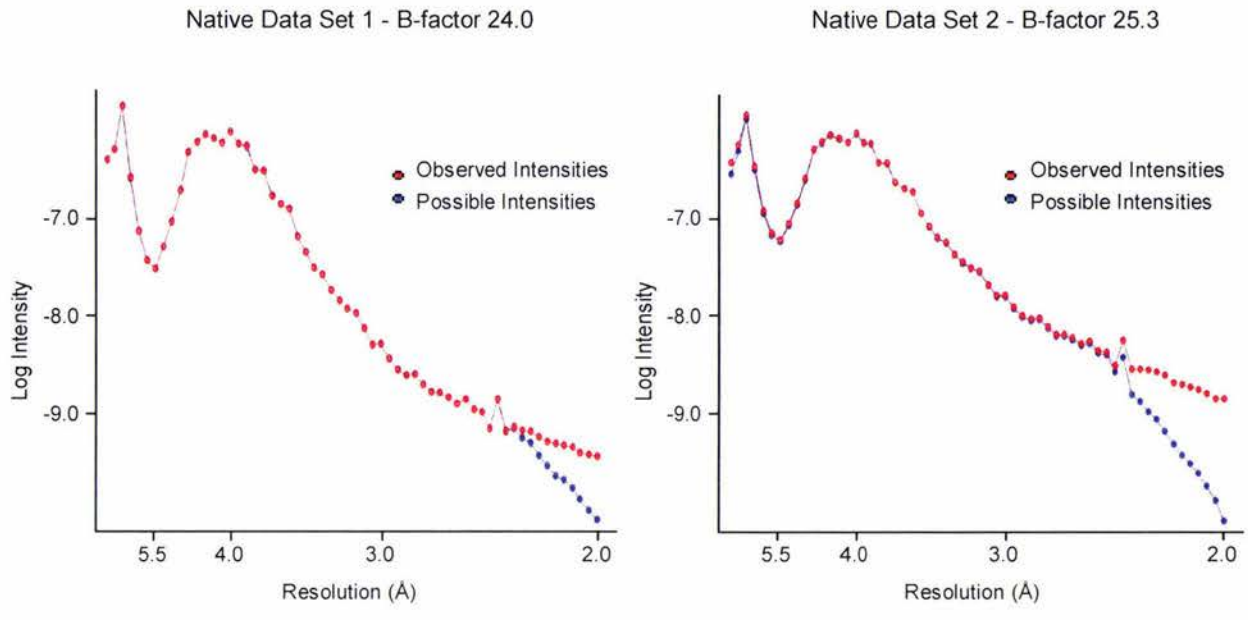


Figure 4.5 Wilson Plot for (left) PepO Native Data Set 1 and (right) PepO Native Data Set 2. (Generated with Truncate, French 1978).

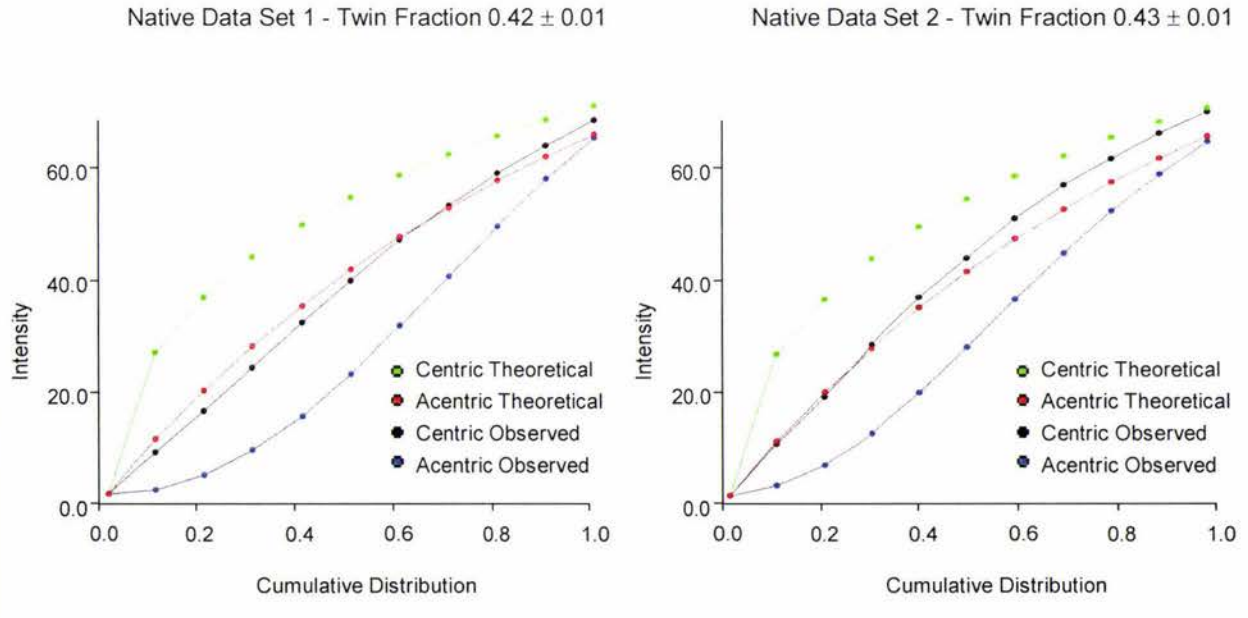


Figure 4.6 Cumulative Intensity-Distribution Curves for (left) PepO Native Data Set 1 and (right) PepO Native Data Set 2. (Generated with Truncate, Howells 1950).

contributions will have very high or that both will have very low intensities, therefore very few extreme values are observed (Terwissha van Scheltinga 2003).

The deviation from the theoretical distribution can be used to estimate the extent of twinning, measured by the twin fraction. By convention, the fraction of the minor crystalline component is called the twin fraction, α (Terwissha van Scheltinga 2003). A twin fraction of 0.0 corresponds to non-twinned data, between 0.0-0.5 partial twinning and 0.5 perfect twinning. The Yeates Crystal Twinning Server showed that Data Sets 1 and 2 were merohedrally twinned, with twin fractions of 0.42 and 0.43 respectively, (Yeates 1997) (Figure 4.6).

Twinning presents fundamental problems that seriously impair the determination of macro-molecular crystal structures. Although structure determination by Multiple Isomorphous Replacement (MIR) techniques is possible in theory (Yeates 1987) only a few cases have been reported, mostly from crystals with twin fractions less than 0.1 (Fisher 1980; Cheah 1994; Terwissha van Scheltinga 2001). The majority of protein structures solved from twinned crystals have been determined using Molecular Replacement (MR) techniques, and include the structure of the *H. sapiens* R121D lactoferrin mutant from a crystal with a twin fraction of 0.43 (Jameson 2002).

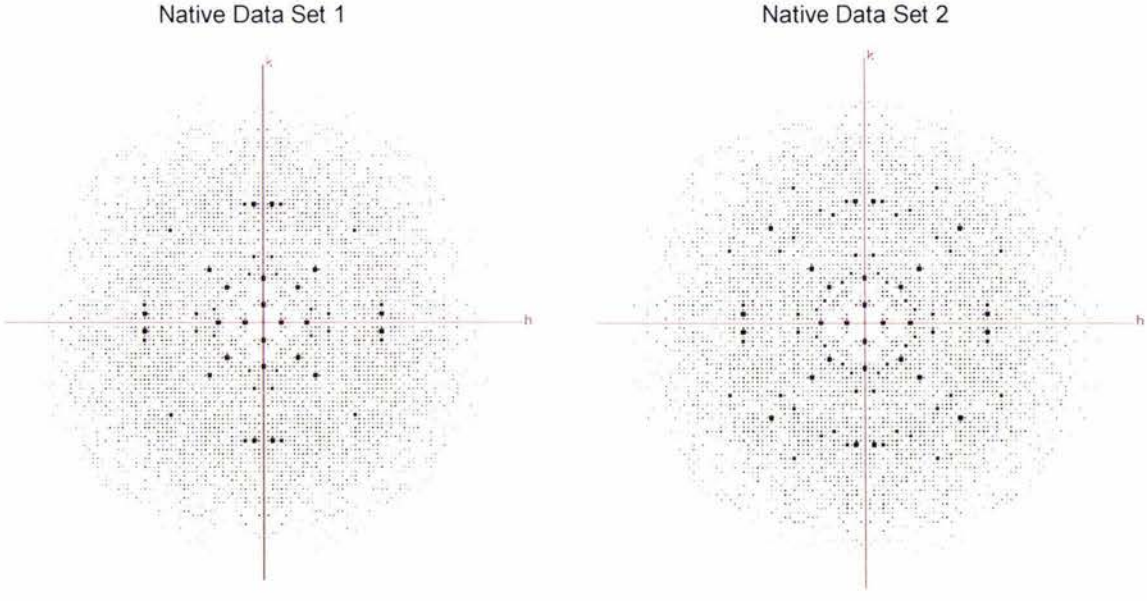


Figure 4.7 HkIview Analysis of (left) PepO Native Data Set 1 and (right) PepO Native Data Set 2. (Psuedo-precession images are of hk0).

It is often not necessary to carry out data correction for twinning prior to Molecular Replacement. Although twinning introduces additional rotational symmetry into the observed Patterson map it is possible to establish the molecular orientation by performing a suitable rotation search. To establish the translation orientation, however, the correct space group must be known and this is often complicated by twinning operations.

To determine the correct space group of the two data sets, analysis of the systematic absences and equivalent reflections in the pseudo-precession images was carried out using the CCP4 program Hklview (Bailey 1994). Although both data sets were clearly tetragonal (Figure 4.7), the correct space groups were somewhat ambiguous. The Data Set 1 and 2 systematic absences and equivalent reflections were characteristic of space groups $P42_12$ and P422 respectively. However, it is possible that the symmetry operations observed are a result of twinning operations.

To determine the number of molecules present in the asymmetric unit of the crystals the Matthew's coefficient was calculated using the CCP4 program Matthews_coeff (Matthews 1968). The Matthew's coefficient, V_M, can be calculated from knowledge of: the unit cell volume, *V unit cell*; the molecular weight of the molecule, *Mr*; the number of asymmetric units according to the crystal symmetry operations, *z*; and the number of molecules in the asymmetric unit, *n*.

$$V_M = \frac{V\ unit\ cell}{z \times n \times Mr} = \frac{117.60\ \text{Å} \times 117.60\ \text{Å} \times 206.09\ \text{Å}}{z \times n \times 70{,}900\ \text{Da}}$$

Native Data Set 1 and 2 crystals have unit cell volumes of $a \times b \times c$ where $a = b = 117.60$ Å and $c = 206.09$ Å, 8 asymmetric units in the tetragonal space groups ($P42_12$ and P422) and 2 asymmetric units in the monoclinic space groups (P2 and $P2_1$), and an unknown number of 70.9 kDa PepO molecules within each asymmetric unit. The number of molecules in the asymmetric unit has a strong correlation with the Matthew's coefficient, V_M, and the solvent content of the crystal (Matthews 1968). The most commonly observed values of V_M are in the vicinity of 2.15 $Å^3$/Da, which corresponds to approximately 50% solvent content (Matthews 1968). For the PepO crystals used to collect Native Data Sets 1 and 2 to have a 50% solvent content, 2 molecules per asymmetric unit will be present in the tetragonal crystal systems, and 8 molecules per asymmetric unit will be present in the monoclinic crystal systems. Regardless of the space group there will be 16 PepO molecules in the crystal unit cell.

4.3.1.2 MOLECULAR REPLACEMENT FOR PEPO NATIVE DATA SET A

With knowledge of the number of molecules per asymmetric unit in the different crystal systems, and the availability of a suitable search model, Molecular Replacement (MR) was used to try to

calculate approximate phases for the PepO structure. The *H. sapiens* neutral endopeptidase (NEP) structure (PDB accession 1DMT) and a Swiss-Model predicted PepO structure were used as search models. The three-dimensional structure of the C-terminal domain (residues 243-360) of PepO was predicted using Swiss-Model a fully automated protein structure homology-modelling server, accessible via the Expasy web server (Gasteiger 2003; Schwede 2003). According to Swiss-Model, PepO has 31.6% amino acid sequence identity to *H. sapiens* NEP over this region. Crystal structures of NEP with various specific inhibitors have been solved and are deposited in the Protein Data Bank (Berman 2000; Oefner 2000; Oefner 2004). The NEP pdb files (PDB accessions: 1DMT, 1R1I, 1R1H, 1R1J) were used to predict the model shown in Figure 4.8 representing residues 243-630 of PepO.

Figure 4.8 Swiss-Model Predicted Structure of PepO. (Cartoon representation of residues 243-630 coloured by residue number from the N-terminus (blue) to the C-terminus (red). Generated using Rasmol, Sayle 1995).

The core β-sheet structure and adjacent helices of the predicted structure may be representative of the true PepO structure as these regions contain highly conserved amino acid sequences between the two structures and include the HEXXH pentapeptide motif found in metallopeptidases. The structure for the N-terminus of PepO (residues 1-242) was not modelled as there is not sufficient sequence similarity between PepO and NEP, or indeed any other protein in the PDB for these residues.

Molecular Replacement (MR) was carried out using the CCP4 program *A*utomated *Mo*lecular *Re*placement program (*AMoRe*) (Navaza 2001). Data Sets 1 and 2 were not corrected for

twinning (detwinned) prior to MR, as the twinning operations were unknown. MR using twinned data often produces overlapping rotation function solutions that can be used to establish the twinning operations, allowing data correction. If the correct space group is known, the rotation function solutions may be used separately in the subsequent translation functions without data correction, to determine the molecular orientation and location within the unit cell.
MR attempts for Data Set 1 using *H. sapiens* NEP (PDB accession 1DMT) as the search model failed to find statistically valid solutions (Tables 4.1, and 4.2).

Table 4.1 AMoRe Solutions for Data Set 1 Processed in Tetragonal using *H. sapiens* NEP as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	53.33	70.61	283.29	0.2130	-0.3977	0.0134	24.4
2	73.67	24.77	195.83	0.0372	-0.3522	0.4720	24.1
3	38.48	43.46	171.53	0.3634	0.4612	0.4427	23.6
4	63.06	72.65	274.86	0.1187	0.3881	0.0791	23.0
5	19.08	45.59	181.09	-0.0075	-0.4901	0.0003	23.7

[a] α, β, γ, Eulerian angles measured in degrees.
[b] Tx, Ty, Tz, Translation coordinates.
[c] Corr, Correlation coefficient.

Table 4.2 AMoRe Solutions for Data Set 1 Processed in Monoclinic using *H. sapiens* NEP as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	235.19	68.15	82.46	0.4652	0.0000	0.5123	21.3
2	274.88	82.84	72.51	0.0389	0.0000	0.0397	21.6
3	347.10	28.05	186.75	0.5162	0.0000	0.1084	21.3
4	55.16	69.25	39.45	0.0248	0.0000	0.4358	21.6
5	335.71	23.24	199.21	0.4898	0.0000	0.4882	20.9
6	331.97	74.56	279.21	0.4701	0.0000	0.0006	22.3
7	267.52	86.25	69.17	0.0389	0.0000	0.0316	21.6
8	241.89	69.91	281.65	0.5137	0.0000	0.0931	21.5
9	312.75	44.18	172.91	0.4845	0.0000	0.0156	21.3
10	72.41	63.13	275.39	0.1020	0.0000	0.4268	22.0

[a] α, β, γ, Eulerian angles measured in degrees.
[b] Tx, Ty, Tz, Translation coordinates.
[c] Corr, Correlation coefficient.

MR attempts were repeated for Data Set 1 using the Swiss-Model predicted PepO structure as the search model and likewise failed to find statistically valid solutions (Tables 4.3, and 4.4).

Table 4.3 AMoRe Solutions for Data Set 1 Processed in Tetragonal using the Swiss-Model Predicted PepO Structure as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	40.00	69.41	99.10	0.0220	0.0462	0.4934	27.0
2	44.42	53.20	291.90	-0.0258	-0.4233	0.4456	27.3
3	36.08	67.01	277.86	-0.3542	0.3673	0.0156	26.7
4	26.48	78.13	97.27	0.3505	-0.3764	0.4936	26.9
5	53.45	42.61	212.45	0.0320	-0.1270	0.0041	27.3

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

Table 4.4 AMoRe Solutions for Data Set 1 Processed in Monoclinic using the Swiss-Model Predicted PepO Structure as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	332.99	72.34	275.28	0.4749	0.0000	0.0005	22.4
2	275.44	73.20	180.86	0.0621	0.0000	0.4681	21.6
3	62.00	72.48	274.34	0.0377	0.0000	0.4891	22.0
4	234.92	67.06	80.83	0.4566	0.0000	0.4855	21.3
5	335.69	23.08	198.83	0.4874	0.0000	0.4889	20.9
6	243.07	72.16	275.15	0.4692	0.0000	0.0821	21.5
7	311.60	44.13	168.59	0.4686	0.0000	0.0034	21.3
8	50.60	71.10	34.63	0.0250	0.0000	0.4109	21.6
9	266.58	85.95	70.50	0.0376	0.0000	0.0287	21.6
10	343.84	24.09	194.21	0.4124	0.0000	0.0964	21.3

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

A statistically valid solution is a solution that has a significantly higher correlation coefficient than all others. Given that there are 2 molecules per asymmetric unit in the tetragonal crystal system, 2 statistically valid solutions would be expected. Similarly, 8 statistically valid solutions would be expected in the monoclinic crystal system. The correlation coefficient values obtained, however, were all very similar and a definite solution was not apparent.

MR attempts for Data Set 2 using *H. sapiens* NEP (PDB accession 1DMT) as the search model (Tables 4.5 and 4.6) and the Swiss-Model predicted PepO structure (Tables 4.7 and 4.8) also failed to find statistically valid solutions. The twinning operations were therefore not identified and subsequently Native Data Sets 1 and 2 were not detwinned.

Table 4.5 AMoRe Solutions for Data Set 2 Processed in Tetragonal using *H. sapiens* NEP as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	28.16	46.79	173.33	0.4218	0.4724	0.4593	24.8
2	55.68	67.36	80.74	0.3768	-0.4207	0.0063	24.2
3	71.26	38.99	222.59	0.4257	0.4539	0.0280	23.6
4	68.31	24.18	187.72	0.2998	-0.4394	0.0128	24.1
5	63.73	65.79	269.05	-0.4653	-0.4916	0.4964	25.5

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

Table 4.6 AMoRe Solutions for Data Set 2 Processed in Monoclinic using *H. sapiens* NEP as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	321.01	90.03	329.11	0.2122	0.0000	0.4861	21.4
2	242.73	71.59	274.79	0.1317	0.0000	0.0749	21.3
3	332.46	72.36	275.07	0.1503	0.0000	0.4358	21.3
4	62.26	71.62	275.57	0.3684	0.0000	0.0749	21.3
5	197.77	82.07	70.42	0.1688	0.0000	0.0108	21.2
6	18.93	82.04	70.70	0.3313	0.0000	0.0108	21.2
7	308.65	44.45	171.34	0.1625	0.0000	0.0038	21.2
8	151.99	72.36	275.42	0.3442	0.0000	0.0213	21.2
9	249.52	51.09	114.70	0.3810	0.0000	0.0142	21.2
10	159.27	51.43	115.12	0.2934	0.0000	0.4782	21.2

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

The rotation and translation values obtained for the tetragonal and monoclinic data sets are considerably different, due to symmetry operations. The higher the symmetry the more compact the rotation function becomes, but the greater the translation function becomes. For the triclinic space group the translation search is irrelevant and similarly for other low symmetry space groups, e.g. monoclinic space groups only require a 2-D search along the x- and z-axes because the position along the y-axis is arbitrary (Crowther 1967).

The inability to identify possible MR solutions suggested that the space group assignment may be incorrect. Although the molecular orientation may in fact be correct, an incorrect space group would result in an inaccurate positioning of the molecular within the unit cell. MR was not repeated with the data reprocessed in alternative space groups.

Table 4.7 AMoRe Solutions for Data Set 2 Processed in Tetragonal using the Swiss-Model Predicted PepO Structure as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	23.50	50.59	301.13	0.0188	0.0674	0.4768	28.2
2	84.51	69.78	352.15	0.0121	0.0417	0.4433	28.7
3	23.68	26.41	34.20	-0.0143	0.4767	0.4860	30.0
4	73.89	49.72	246.10	0.3846	-0.4664	0.4834	28.1
5	86.90	69.56	351.58	0.0124	0.0419	0.4433	27.6

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

Table 4.8 AMoRe Solutions for Data Set 2 Processed in Monoclinic using the Swiss-Model Predicted PepO Structure as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	61.68	30.35	274.71	0.2625	0.0000	0.4680	24.4
2	114.59	79.31	96.99	0.0622	0.0000	0.0215	24.4
3	55.98	29.93	275.73	0.3804	0.0000	0.4892	24.3
4	194.62	40.46	329.94	0.1684	0.0000	0.4284	24.5
5	195.15	39.40	328.26	0.1684	0.0000	0.4284	24.5
6	44.53	54.13	293.12	0.3059	0.0000	0.0250	24.5
7	116.02	79.15	97.93	0.3060	0.0000	0.4604	24.4
8	202.69	26.24	34.51	0.2741	0.0000	0.0174	24.4
9	238.70	29.80	274.69	0.3625	0.0000	0.4429	24.3
10	176.08	72.71	89.69	0.2620	0.0000	0.4751	24.3

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Translation coordinates.

[c] Corr, Correlation coefficient.

4.3.1.3 MULTIPLE ISOMORPHOUS REPLACEMENT FOR PEPO DERIVATIVE DATA

Twinning presents fundamental problems that seriously impair the determination of macro-molecular crystal structures by Multiple Isomorphous Replacement (MIR) techniques. The general strategy for structure determination of twinned data is to correct for twinning (detwin) prior to MIR in order to extract the heavy atom signal.

Although MIR for crystals with low twin fractions has been successful, until recently MIR had not been used to solve the structure for crystals with high twin fractions. In 2001, Terwissha van Scheltinga *et al.* successfully solved by MIR the structure of Deacetoxycephalosporin C Synthase from merohedrally twinned crystals with a twin fraction of 0.5.

MIR was attempted for Data Set 1 and 2 without correcting for twinning, using the CCP4 programs Collect Assorted Data (CAD) and Scaleit (Howell 1992; Bailey 1994). CAD combines native and derivative data sets into one file, while Scaleit puts the data sets on the same scale using an overall scale factor and B-factor. Scaleit compares the data sets and calculates several statistics which are useful in establishing isomorphism and suitability for MIR. The most useful of these statistics is the weighted R-factor, which evaluates the fractional isomorphous difference and compares it with the expected agreement with the native data.

In general, the weighted R-factor values should be within the range of 0.1-0.3 if the derivative is isomorphous and thus suitable for phasing. Typically, values less than 0.1 suggest that the heavy atom is not bound, while values higher than about 0.3 suggest an unacceptable level of non-isomorphism (Rossmann 2001). Values between 0.1-0.3 suggest, but do not guarantee, that the derivative is worth pursuing (Rossmann 2001). The weighted R-factors for the derivative data sets were all found to be below 0.1 (Table 4.9).

Table 4.9 Scaleit Statistics for PepO Derivative Data Sets.

	Unit Cell Dimensions						Weighted R-factor [a]	Suitable for MIR [a]
	a	b	c	α	β	γ		
Native	117.62	117.62	206.09	90.0	90.0	90.0	N/A	N/A
(i) $C_9H_9HgO_2SNa$	118.06	118.06	205.53	90.0	90.0	90.0	0.068	x
(ii) $C_9H_9HgO_2SNa$	114.38	114.89	204.09	90.0	90.0	90.0	N/D	x
(iii) CH_3CH_2HgCl	118.19	118.19	206.16	90.0	90.0	90.0	N/D	x
(iv) $UO_2CH_3CO_2$	117.44	117.44	206.32	90.0	90.0	90.0	0.040	x

[a] N/A, not applicable; N/D, not determined; x, not suitable.

It is difficult to establish whether or not the low weighted R-factors observed for the derivative data are a consequence of non-binding of the heavy atoms to the crystal, or as a consequence of twinning. To determine this, it would be necessary to detwin the native and derivative data sets.

4.3.1.4 PEPO NATIVE DATA SET B

PepO Native Data Set 3 intensities (I) were converted to pseudo-structure factors (F), and were shown to follow the Wilson distribution with a calculated B-factor of 23.2, which is comparable with Native Data Sets 1 and 2 (Figure 4.9). The Cumulative Intensity-distribution of Native Data Set 3 did not suggest twinning (Figure 4.9).

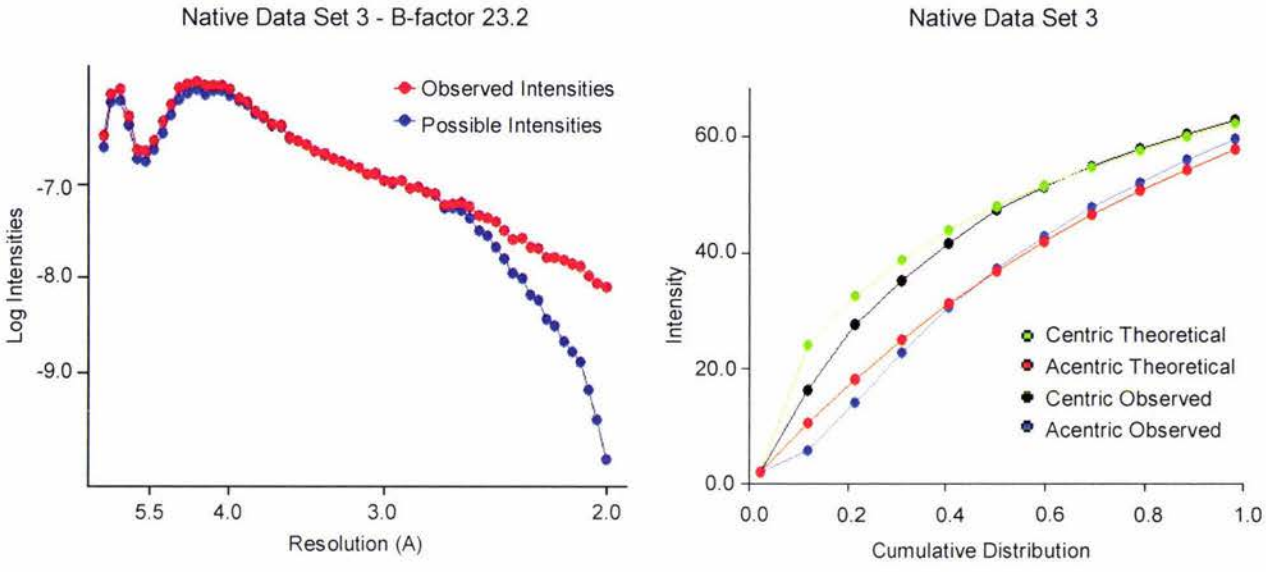


Figure 4.9 Wilson Distribution and Cumulative Intensity-Distribution for PepO Native Data Set 3. (Generated with Truncate, French 1978).

The crystals have a unit cell volume of a x b x c where a = 56.38 Å, b = 83.75 Å and c = 143.82 Å, 4 asymmetric units in the P222 space group, and an unknown number of 70.9 kDa PepO molecules within each asymmetric unit.

$$V_M = \frac{V \text{ unit cell}}{z \times n \times Mr} = \frac{56.38 \text{ Å} \times 83.75 \text{ Å} \times 143.82 \text{ Å}}{4 \times n \times 70{,}900 \text{ Da}}$$

If one monomer is present in the asymmetric unit, the Matthew's coefficient is calculated to be 2.5, correlating to a solvent content of 48.2%.

4.3.1.5 MOLECULAR REPLACEMENT FOR PEPO NATIVE DATA SET B

Molecular Replacement was carried out for Native Data Set 3 using the CCP4 *A*utomated *Mo*lecular *Re*placement program (*AMoRe*), and *H. sapiens* NEP (PDB accession 1DMT) and the Swiss-Model predicted PepO structure as search models.

MR attempts for Data Set 3 using *H. sapiens* NEP (PDB accession 1DMT) as the search model failed to find a statistically valid solution (Tables 4.10).

Table 4.10 *AMoRe* Solutions for PepO Native Data Set 3 using *H. sapiens* NEP as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	26.50	70.43	352.38	0.0195	0.2030	0.0621	9.9
2	98.720	88.23	214.15	0.3507	0.0942	0.0547	9.3
3	137.96	47.24	44.67	0.4523	0.3988	0.0942	9.3
4	82.25	48.61	312.69	0.1788	0.3451	0.2841	10.0
5	97.65	74.00	123.47	0.4074	0.4560	0.4019	8.6

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Tranlsation coordinates.

[c] Corr, Correlation coefficient.

MR attempts for Data Set 3 using the Swiss-Model predicted PepO structure as the search model however, found a statistically valid solution (Table 4.11).

Table 4.11 AMoRe Solutions for PepO Native Data Set 3 using the Swiss-Model Predicted PepO Structure as the Search Model.

	α [a]	β [a]	γ [a]	Tx [b]	Ty [b]	Tz [b]	Corr [c]
1	37.79	73.89	343.71	0.2802	0.0154	0.4011	30.5
2	39.72	74.65	342.33	0.2832	0.2153	0.3683	30.0
3	142.47	65.33	113.68	0.3525	0.4363	0.0062	28.2
4	93.87	87.02	16.56	0.2014	0.2501	0.3418	28.3
5	121.07	41.16	262.73	0.0963	0.1708	0.4270	28.1

[a] α, β, γ, Eulerian angles measured in degrees.

[b] Tx, Ty, Tz, Tranlsation coordinates.

[c] Corr, Correlation coefficient.

The correlation coefficients of solutions 1 and 2, 30.5 and 30.0 respectively, are higher than all other solutions suggesting a possible structure solution. The molecular orientation (measured in Eulerian angles) and position (measured in translation coordinates) of the two solutions are almost equivalent, suggesting that 1 molecule in the asymmetric unit had been found.

Solution 1 was analysed using the CCP4 program Turbo-Frodo (Jones 1978) to locate the molecules within the unit cell and identify possible molecular overlap. There are four molecules present in the unit cell, consistent with the Matthew's coefficient calculation, which appear to overlap slightly (Figures 4.10).

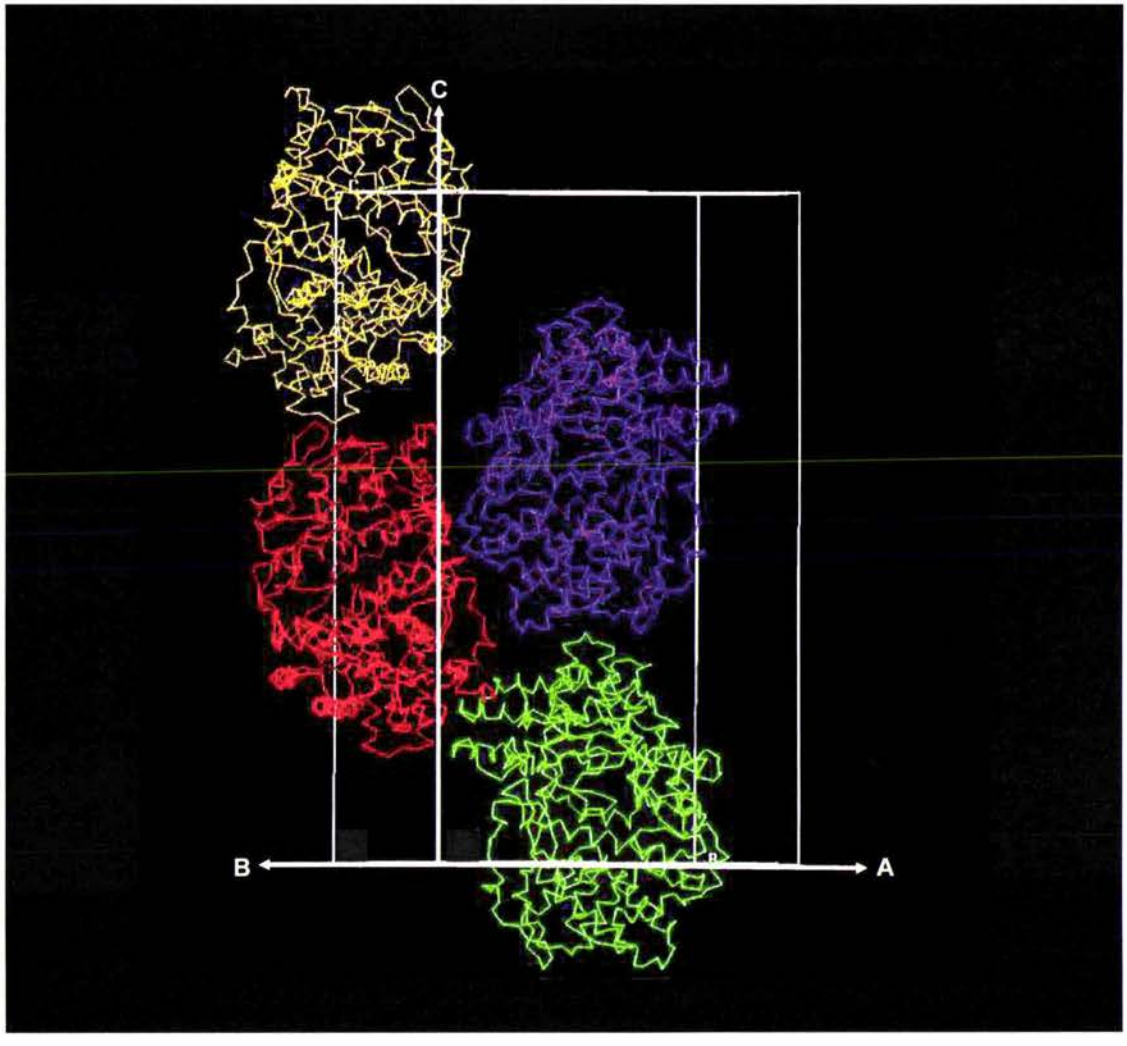


Figure 4.10 Visual Analysis of PepO Native Data Set 3 Unit Cell. (View along a-, b- and c-axis. Generated from AMoRe Solution 1 using Turbo-Frodo).

The full extent of molecular overlap is not illustrated in Figures 4.10 as only the Cα-backbone of the molecules is shown. The molecules within the unit cell are arranged in a $P2_12_12_1$ space group and not the P222 space group that was initially expected (Figure 4.11).

Unfortunately, although the resulting model was approximately placed in the unit cell, it could not be built into the electron density and the structure could not be solved.

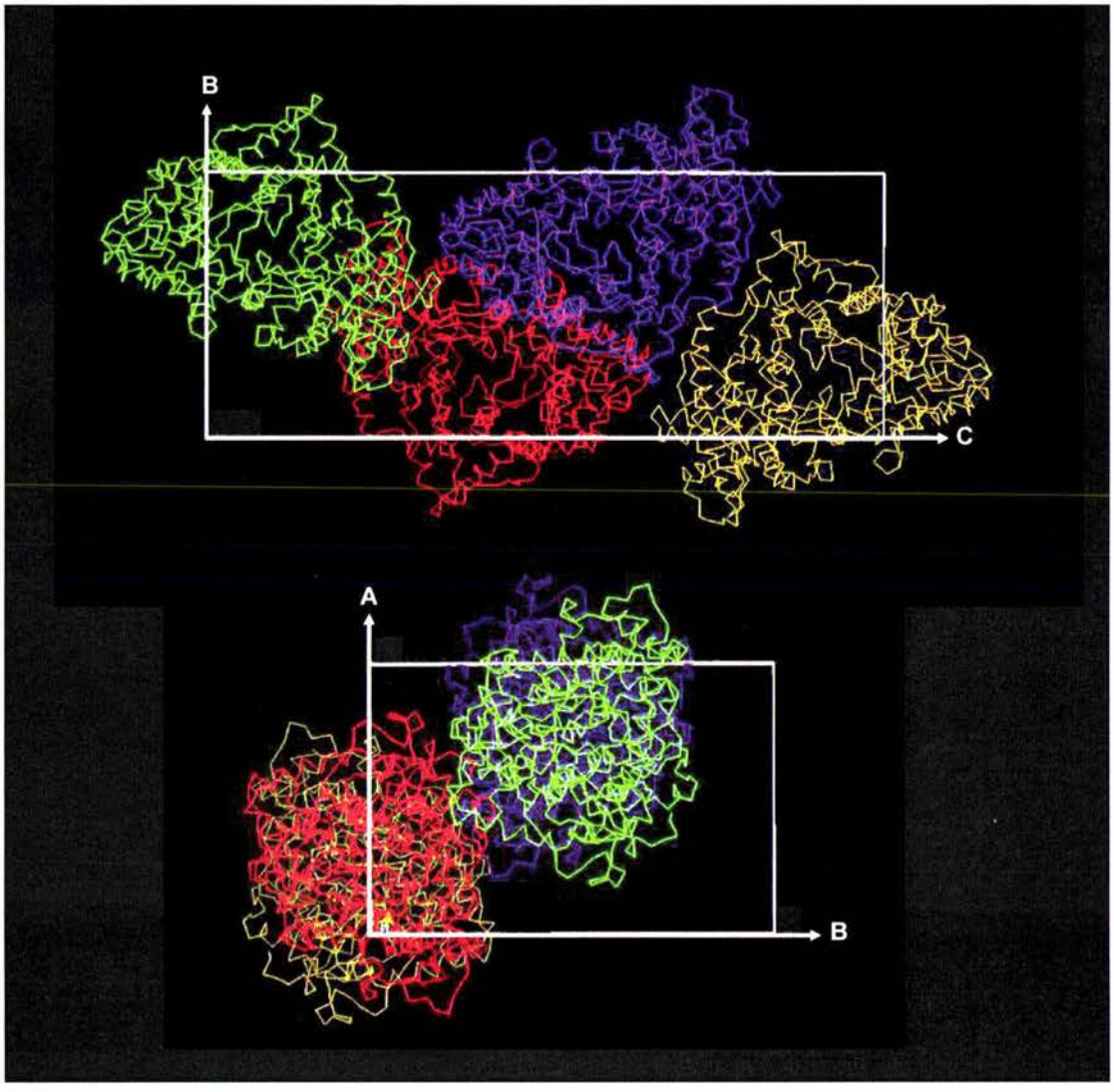


Figure 4.11 Visual Analysis of PepO Native Data Set 3 Unit Cell as Viewed Along the b- and c-axis (top) and a- and b-axis (below). (Generated from AMoRe Solution 1 using Turbo-Frodo).

4.3.2 *L. RHAMNOSUS* PEPW PHASE PROBLEM

4.3.2.1 THREE-DIMENSIONAL STRUCTURE PREDICTIONS

The three-dimensional structure of PepW was predicted using Swiss-Model a fully automated protein structure homology-modelling server, accessible via the Expasy web server (Gasteiger 2003; Schwede 2003). According to Swiss-Model, PepW has 52.70%, 52.45%, 39.93% and 39.82% amino acid identity to *H. sapiens* bleomycin hydrolase (hBH) (PDB accessions 1CB5A, 1CB5B, 1CB5C), *H. sapiens* hBH C73S/A455 deletion mutant (PDB accessions 2CB5A, 2CB5B), *S. cerevisiae* bleomycin hydrolase (Gal6) (PDB accessions 1GCB), *S. cerevisiae* Gal6 C73A/K454 deletion mutant (PDB accession 3GCB), and *S. cerevisiae* Gal6 C73A mutant (PDB accession 1A6R) respectively. The *H. sapiens* and *S. cerevisiae* bleomycin hydrolase pdb files (PDB accessions: 1CB5A, 1CB5B, 1CB5C, 2CB5A, 2CB5B, 1GCB, 3GCB, 1A6R) were used to predict the three-dimensional structure of PepW (Figure 4.12).

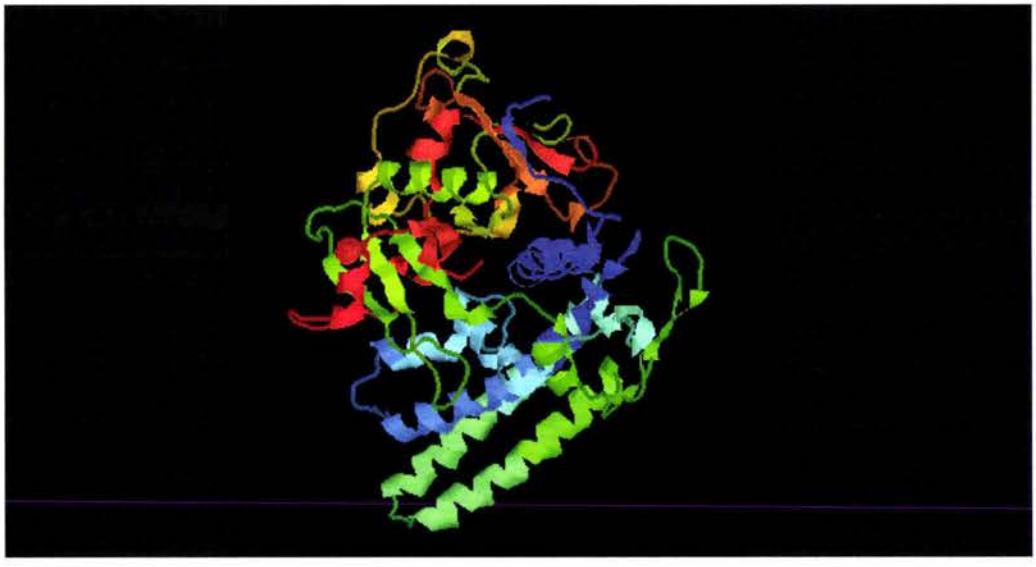


Figure 4.12 Swiss-Model Predicted Structure of PepW. (Cartoon representation coloured by group: blue, N-terminal; red, C-terminal).

The Swiss-Model generated PepW structure is not necessarily representative of the actual three-dimensional structure of PepW. The bleomycin hydrolases and related proteins are however, highly conserved and therefore may have very similar structures. To experimentally determine the crystal structure of PepO it is necessary to obtain accurate diffraction data from which the phase problem can be solved. As discussed in Chapter 3 the quality of the PepW native data set was poor and as such an attempt to solve the phase problem was not carried out.

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 CONCLUSIONS

Part 2 of this thesis investigated the three-dimensional structures of the *L. rhamnosus* peptidases, PepO and PepW. In order to understand the molecular basis for the unique substrate specificities of PepO and PepW, attempts were made to determine the three-dimensional structures of the enzymes by x-ray crystallography.

PepO and PepW crystals were successfully grown using sitting- and hanging-drop vapour-diffusion techniques. PepO crystals suitable for x-ray crystallography were grown in 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate at 4°C (1-2 weeks) and 1.5 M lithium sulphate monohydrate, 0.1 M sodium hepes pH 7.5 at room temperature (10-12 months). Derivatization of the PepO crystals grown in 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate at 4°C was attempted with mercury (Hg), platinum (Pt), uranly (U) and praseodymium (Pr) salts. PepW crystals suitable for x-ray crystallography were grown in 28% PEG 6000, 0.2 M MES pH 6.1 at room temperature (2 months).

Preliminary diffraction data for the PepO and PepW crystals is described in Table 5.1.

Table 5.1 X-ray Diffraction Analysis of PepO and PepW Crystals.

Crystal [a]	Diffraction	Space Group	Resolution	Unit Cell Dimensions					
				a	b	c	α	β	γ
PepO A	✓	P422, $P42_12$	2.00 Å	117.60	117.60	206.09	90.0	90.0	90.0
PepO B	✓	P222	1.80 Å	56.32	83.75	143.82	90.0	90.0	90.0
PepW	✓	P3	3.00 Å	142.22	142.22	98.27	90.0	90.0	90.0

[a] PepO Native A, 15% PEG 8000, 0.1 M sodium cacodylate pH 6.5, 0.2 M calcium acetate at 4°C (1-2 weeks)
PepO Native B, 1.5 M Lithium sulphate monohydrate, 0.1 M sodium hepes pH 7.5 at RT (10-12 months)
PepW Native A, 28% PEG 6000, 0.2 M MES pH 6.1 at RT (2 months)

Native data sets were collected for both PepO and PepW, and derivative data were collected for PepO. The PepO A native data processed in space groups P422 and $P42_12$, and were found to be merohedrally twinned with twin fractions of 0.42-0.43. Derivate data for PepO A processed in a number of different space groups ($P4_1$, $P2_1$, P2, C2), and hence were not suitable for structure determination using Multiple Isomorphous Replacement (MIR). Molecular Replacement (MR) was attempted to solve the PepO structure using *H. sapiens* NEP (PDB accession 1DMT) and a Swiss-Model predicted PepO structure, as search models, but both failed to find statistically valid solutions.

The PepO B native data processed in the space group P222, and showed no signs of twinning. Molecular replacement was used to try and solve the structure, but failed to give a statistically valid solution using the *H. sapiens* NEP as a model. A possible solution was found however, using the Swiss-Model predicted PepO structure. Unfortunately, although the resulting model was approximately placed in the unit cell, it could not be built into the electron density.

The PepW native data processed in space group P3, but because of instrument failure during data collection, and a lack of available crystals, insufficient data was collected to solve the structure.

5.2 FUTURE DIRECTIONS

To determine the three dimensional structures of *L. rhamnosus* PepO and PepW further work is required. Although suitable PepO crystallisation conditions have been found that produce suitable crystals for x-ray crystallography, a suitable search model for MR is not available, and suitable heavy atom derivatives for MIR have not been found.

As a new crystal form has been found which lacks a twinning problem, these crystals will no longer be used for structure determination and all efforts will be placed into obtaining heavy atom derivatives of type 'B' crystals. An alternative solution would be to introduce seleno-methionine residues and determine the PepO structure using Multiple Anomalous Dispersion (MAD) methods.

To successfully determine the PepW structure it will be necessary to re-crystallise PepW and collect a complete and accurate native data set. Because of the high degree of sequence similarity to *S. cerevisiae* bleomycin hydrolase, Gal6 (PDB accession 1GCB) and *H. sapiens* bleomycin hydrolase, hBH (PDB accession 1CB5A, 1CB5B, 1CB5C), molecular replacement will be the method of preference for structure solution.

The incentive to investigate the PepO and PepW structures was to aid the understanding of the substrate specificity of the enzymes. It will therefore be beneficial to attempt co-crystallisation of both enzymes with either a substrate or inhibitor.

Once the crystal structures of PepO and PepW are known it will be possible to conduct informed site-directed mutagenesis studies to either modify or enhance the substrate specificities. The structural and functional characterisation of *L. rhamnosus* PepO and PepW could in the future be used to inform the genetic engineering of proteolytic enzymes designed to develop new cheese flavours.

APPENDIX I

L. RHAMNOSUS PEPW SEQUENCE

```
 M  S  A  E  I  T  S  G  D  L  D  Q  F  K  Q  D  L  Q  A  T  P  A  A  N  A  L  Q  K  A  V
ATGTCTGCAGAAATTACTTCAGGCGATTTGGATCAGTTCAAACAGGATCTTCAAGCCACACCTGCCGCCAACGCCTTACAAAAAGCGGTC
         10        20        30        40        50        60        70        80        90

 M  N  N  G  I  N  A  T  A  E  N  T  D  S  K  V  A  M  T  P  T  F  S  I  E  L  D  T  G  A
ATGAACAATGGTATTAATGCAACTGCCGAGAATACCGATAGCAAAGTTGCCATGACACCAACGTTCTCAATTGAACTTGATACCGGTGCT
        100       110       120       130       140       150       160       170       180

 V  S  N  Q  K  Q  S  G  R  C  W  M  F  A  A  L  N  T  M  R  H  G  I  Q  A  Q  F  K  I  K
GTGTCAAACCAAAAACAAAGCGGCCGGTGCTGGATGTTCGCCGCCTTGAATACTATGCGTCATGGCATTCAGGCACAGTTTAAGATCAAG
        190       200       210       220       230       240       250       260       270

 D  F  E  L  S  Q  N  Y  T  F  F  W  D  K  F  E  K  S  N  Y  F  Y  E  N  V  L  K  T  A  D
GATTTTGAACTGTCCCAAAACTACACCTTCTTCTGGGACAAGTTTGAAAAGTCCAACTATTTTTATGAAAATGTCTTAAAAACCGCTGAT
        280       290       300       310       320       330       340       350       360

 Q  P  L  Y  S  R  K  V  A  F  L  L  A  T  P  Q  Q  D  G  G  Q  W  D  M  L  S  A  L  I  E
CAACCACTCTACAGCCGTAAAGTTGCCTTTCTTTTGGCTACCCCGCAACAAGATGGCGGCCAATGGGATATGTTGTCAGCCTTAATCGAA
        370       380       390       400       410       420       430       440       450

 K  Y  G  I  V  P  K  S  V  M  P  E  T  Y  S  S  S  K  S  N  E  L  N  G  L  L  N  L  K  L
AAGTATGGCATTGTGCCGAAGTCAGTAATGCCTGAAACCTACAGTTCTAGCAAGAGTAACGAATTAAATGGTTTGCTCAACTTGAAACTG
        460       470       480       490       500       510       520       530       540

 R  K  D  A  V  T  L  R  K  L  V  A  D  K  A  S  D  A  D  I  E  A  A  K  Q  K  M  L  A  E
CGTAAAGATGCGGTCACCCTACGCAAGTTAGTTGCTGATAAAGCTAGCGATGCCGACATTGAAGCAGCCAAGCAAAAAATGCTGGCTGAA
        550       560       570       580       590       600       610       620       630

 D  Y  R  I  W  A  Y  T  L  G  N  P  P  T  K  F  D  F  E  Y  R  D  D  D  K  N  Y  H  I  D
GACTATCGCATCTGGGCATACACGTTAGGCAACCCGCCAACCAAATTTGACTTTGAATACCGCGATGATGATAAGAATTATCACATTGAT
        640       650       660       670       680       690       700       710       720

 R  E  L  T  P  Q  T  F  F  K  K  Y  V  G  W  N  L  D  D  Y  Q  S  I  I  N  A  P  T  A  D
CGCGAATTGACGCCGCAAACCTTCTTCAAGAAGTATGTCGGCTGGAATCTCGATGACTATCAAAGTATCATCAACGCCCCAACCGCTGAC
        730       740       750       760       770       780       790       800       810

 K  P  Y  K  H  L  Y  T  V  E  M  L  G  N  V  V  G  G  R  E  V  R  H  L  N  L  D  I  D  T
AAGCCTTACAAGCACCTTTACACCGTTGAAATGCTGGGCAATGTCGTAGGTGGTCGTGAAGTTCGCCACTTGAATCTTGACATTGATACC
        820       830       840       850       860       870       880       890       900

 F  K  D  L  A  I  K  Q  L  K  A  G  E  S  V  W  F  G  S  D  V  G  Q  S  S  D  R  Q  L  G
TTCAAAGACTTGGCAATCAAGCAACTCAAGGCCGGCGAATCAGTTTGGTTCGGTTCCGATGTTGGTCAAAGCTCCGATCGTCAACTTGGC
        910       920       930       940       950       960       970       980       990

 I  L  D  T  N  I  Y  K  K  D  D  L  F  N  T  D  F  T  M  T  K  A  E  R  L  D  Y  G  E  S
ATTTTGGACACCAACATTTACAAGAAAGATGACCTGTTCAACACCGACTTCACCATGACTAAAGCTGAACGCCTTGACTATGGCGAAAGC
       1000      1010      1020      1030      1040      1050      1060      1070      1080
```

```
 L  M  T  H  A  M  V  L  T  G  V  D  L  V  D  G  K  P  T  K  W  K  V  E  N  S  W  G  E  K
TTGATGACTCACGCAATGGTCTTGACCGGCGTTGATCTTGTAGACGGCAAGCCAACCAAGTGGAAAGTTGAAAACTCTTGGGGTGAAAAA
       1090       1100       1110       1120       1130       1140       1150       1160       1170

 V  G  E  K  G  Y  F  V  A  S  D  A  W  F  D  Q  F  V  Y  Q  V  V  I  S  K  K  Y  L  P  A
GTCGGCGAAAAAGGCTACTTTGTCGCAAGTGATGCGTGGTTTGACCAATTCGTCTACCAAGTTGTTATTTCCAAGAAGTATCTCCCAGCT
       1180       1190       1200       1210       1220       1230       1240       1250       1260

 E  L  Q  D  V  I  K  N  E  Y  D  K  P  T  V  L  A  P  W  D  P  M  G  A  L  A  S  R  *
GAACTACAGGATGTCATCAAGAACGAATACGACAAGCCAACCGTTCTTGCACCTTGGGATCCAATGGGTGCTTTGGCATCAAGATAA
       1270       1280       1290       1300       1310       1320       1340       1350
```

L. RHAMNOSUS PEPO SEQUENCE

```
 M  T  L  P  R  I  Q  D  D  L  Y  L  A  V  N  G  E  W  Q  A  K  T  P  I  P  P  D  K  S  V
ATGACATTACCAAGAATTCAAGATGATTTGTACCTAGCCGTCAATGGCGAATGGCAAGCGAAGACGCCGATTCCACCTGACAAAAGTGTT
         10         20         30         40         50         60         70         80         90

 V  S  A  D  S  N  L  T  D  D  I  R  Q  K  L  V  A  D  L  S  T  M  T  K  T  A  K  T  L  P
GTGAGTGCGGATAGTAATCTGACCGATGATATTCGCCAAAAACTAGTGGCTGATCTAAGCACGATGACGAAAACAGCCAAAACTTTGCCG
        100        110        120        130        140        150        160        170        180

 L  Q  Y  A  A  R  L  F  A  K  A  N  D  Q  T  R  R  Q  Q  L  G  I  E  P  V  R  D  R  I  S
CTCCAGTATGCAGCGCGGTTGTTTGCCAAAGCCAATGACCAAACCCGCCGTCAGCAGCTAGGCATTGAGCCAGTTCGTGATCGGATAAGC
        190        200        210        220        230        240        250        260        270

 F  L  M  A  L  T  T  L  D  Q  F  R  S  A  M  P  K  L  V  A  D  Q  Y  V  L  P  I  S  P  Y
TTTTTGATGGCGCTCACGACGCTTGATCAATTTCGCAGCGCTATGCCCAAACTGGTTGCTGATCAATACGTCTTACCGATCAGTCCTTAC
        280        290        300        310        320        330        340        350        360

 V  D  A  D  M  H  D  A  E  H  N  I  L  N  L  G  G  P  D  T  I  L  P  D  A  A  M  Y  Q  H
GTTGATGCTGATATGCACGATGCCGAGCATAATATTCTGAATCTTGGCGGGCCAGACACAATTTTACCTGATGCGGCGATGTACCAACAT
        370        380        390        400        410        420        430        440        450

 E  D  A  E  N  A  A  D  L  A  A  W  S  Q  M  A  A  A  M  L  A  A  V  G  F  S  Q  T  D  Q
GAAGATGCCGAAAATGCGGCGGATCTGGCAGCGTGGTCGCAGATGGCAGCTGCCATGCTGGCTGCGGTAGGATTCAGTCAGACTGATCAA
        460        470        480        490        500        510        520        530        540

 T  A  Y  V  E  A  A  K  R  F  D  R  R  L  A  D  Y  V  P  A  N  V  D  L  A  V  D  S  T  Y
ACAGCATATGTTGAAGCGGCTAAACGATTTGATCGGCGTTTGGCTGATTATGTGCCAGCAAATGTTGACTTAGCGGTAGATAGCACGTAT
        550        560        570        580        590        600        610        620        630

 D  N  P  L  S  W  Q  A  F  E  D  A  A  G  Y  L  G  I  P  Q  A  F  A  T  Y  M  P  Q  T  P
GACAATCCATTGAGCTGGCAGGCGTTTGAAGATGCGGCCGGTTATTTGGGGATCCCACAAGCCTTTGCAACTTACATGCCGCAAACACCG
        640        650        660        670        680        690        700        710        720

 A  K  V  N  A  V  V  P  A  Y  L  P  H  L  S  K  L  L  T  P  D  N  Y  S  E  W  H  A  W  M
GCGAAAGTCAATGCGGTTGTACCGGCTTATCTTCCGCACTTAAGCAAACTACTGACGCCGGACAATTATTCAGAATGGCACGCATGGATG
        730        740        750        760        770        780        790        800        810
```

```
 V  I  N  E  L  L  T  C  A  T  Y  L  S  D  D  L  R  Q  L  A  G  Q  Y  D  R  F  L  A  G  Q
GTGATTAACGAATTGCTAACCTGCGCCACTTACCTCAGTGATGATTTACGTCAATTGGCCGGACAGTATGATCGGTTTTTGGCTGGTCAA
        820        830        840        850        860        870        880        890        900

 P  E  A  S  S  W  T  K  H  A  F  G  I  A  N  E  Y  F  D  D  V  I  G  Q  Y  Y  G  Q  T  Y
CCTGAGGCGTCATCGTGGACGAAACACGCTTTTGGGATTGCCAACGAGTATTTTGACGATGTGATTGGTCAGTATTATGGTCAAACCTAC
        910        920        930        940        950        960        970        980        990

 F  G  A  D  A  K  A  D  V  T  A  M  V  K  Q  I  L  A  Q  Y  R  V  Q  L  E  N  N  T  W  L
TTTGGTGCCGACGCTAAGGCAGATGTGACGGCCATGGTTAAGCAAATTCTTGCGCAATACCGCGTGCAGCTAGAAAACAACACTTGGCTG
       1000       1010       1020       1030       1040       1050       1060       1070       1080

 S  P  A  T  K  Q  K  A  M  R  K  L  A  T  M  Q  V  K  M  G  Y  P  E  R  L  F  S  L  Y  D
AGTCCGGCTACGAAGCAAAAGGCGATGCGCAAGTTAGCCACGATGCAAGTCAAAATGGGGTATCCGGAGCGACTCTTTTCCTTGTATGAT
       1090       1100       1110       1120       1130       1140       1150       1160       1170

 H  L  S  V  D  V  D  D  D  L  L  T  A  I  L  K  L  S  A  Q  T  Q  A  F  W  F  K  Q  L  G
CACTTGAGCGTGGATGTTGACGATGATTTGTTGACGGCAATTCTGAAACTTAGCGCACAGACGCAGGCCTTTTGGTTTAAACAGTTAGGC
       1180       1190       1200       1210       1220       1230       1240       1250       1260

 Q  T  V  D  R  N  Q  W  N  M  P  G  H  L  V  N  A  S  Y  D  P  L  K  N  D  I  T  F  P  A
CAGACGGTGGATCGGAATCAATGGAATATGCCGGGACACTTGGTGAATGCCAGTTATGATCCGCTGAAAAATGACATCACTTTTCCCGCT
       1270       1280       1290       1300       1310       1320       1330       1340       1350

 G  I  L  Q  P  P  Y  Y  S  L  K  W  T  R  A  E  N  L  G  G  T  G  A  T  I  G  H  E  I  S
GGTATCTTGCAGCCGCCGTATTACTCACTCAAATGGACCCGGGCGGAAAACCTCGGAGGGACAGGCGCAACGATCGGTCATGAAATCTCG
       1360       1370       1380       1390       1400       1410       1420       1430       1440

 H  S  F  D  N  N  G  A  L  Y  D  E  Y  G  N  L  H  N  W  W  T  P  A  D  K  Q  A  F  D  Q
CATTCGTTTGATAATAACGGGGCGCTGTATGATGAATATGGTAATTTGCATAACTGGTGGACACCAGCGGATAAGCAGGCATTTGATCAG
       1450       1460       1470       1480       1490       1500       1510       1520       1530

 L  V  K  A  M  A  A  Q  F  D  G  R  D  Y  E  G  V  K  V  N  G  T  L  T  V  S  E  N  M  A
CTGGTAAAAGCGATGGCGGCACAGTTTGATGGCCGTGACTATGAAGGAGTCAAGGTCAACGGTACACTGACCGTTAGTGAAAACATGGCG
       1540       1550       1560       1570       1580       1590       1600       1610       1620

 D  N  A  G  M  D  V  A  L  A  L  L  G  D  Q  P  D  V  K  D  L  Q  A  F  F  I  T  Y  A  R
GATAACGCCGGCATGGATGTGGCGTTGGCGTTACTAGGCGATCAGCCGGATGTTAAGGATCTGCAGGCATTCTTCATCACTTACGCTCGT
       1630       1640       1650       1660       1670       1680       1690       1700       1710

 S  W  A  T  K  M  R  P  E  R  A  K  T  V  L  R  Q  D  V  H  A  P  A  T  L  R  V  N  V  P
TCATGGGCCACCAAAATGCGACCGGAGCGGGCTAAAACTGTTTTGCGGCAAGATGTTCATGCGCCGGCTACCTTACGCGTGAATGTGCCG
       1720       1730       1740       1750       1760       1770       1780       1790       1800

 V  Q  N  F  P  A  W  Y  Q  A  F  N  V  Q  P  Q  D  G  M  Y  R  Q  P  Q  K  R  L  T  I  W
GTGCAAAACTTTCCTGCATGGTACCAGGCATTTAATGTTCAGCCACAAGATGGTATGTATCGGCAACCACAGAAGCGGCTGACGATTTGG
       1810       1820       1830       1840       1850       1860       1870       1880       1890

 H  Q  *
CATCAGTAA
       1899
```

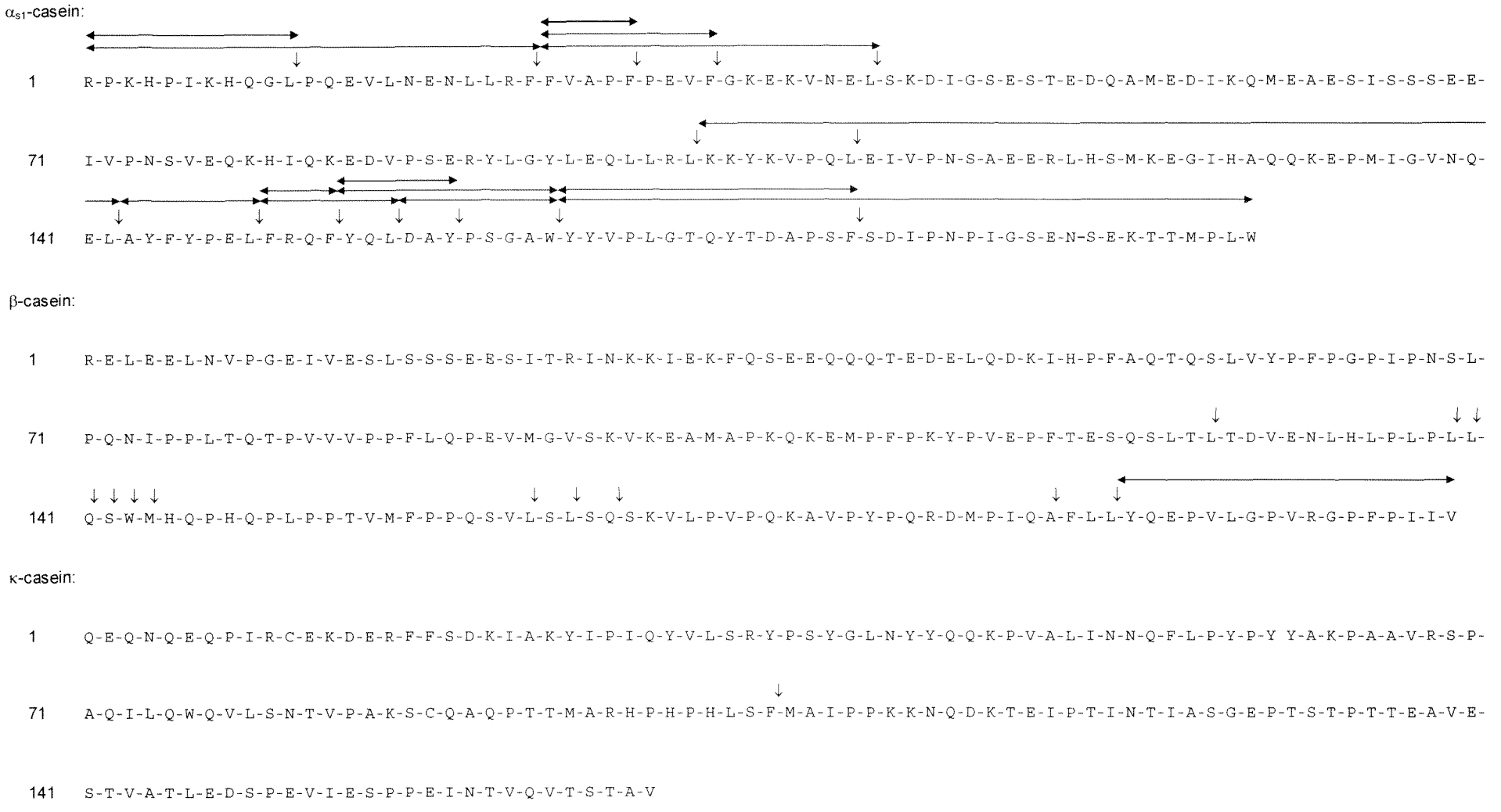


Figure 2.1 Chymosin Cleavage Sites and Soluble Peptide Products of α_{s1}-, β-, and κ-casein (Visser 1977; McSweeney 1993; Fox 2000).
(↓, Cleavage sites; ←→, soluble peptides produced by chymosin).

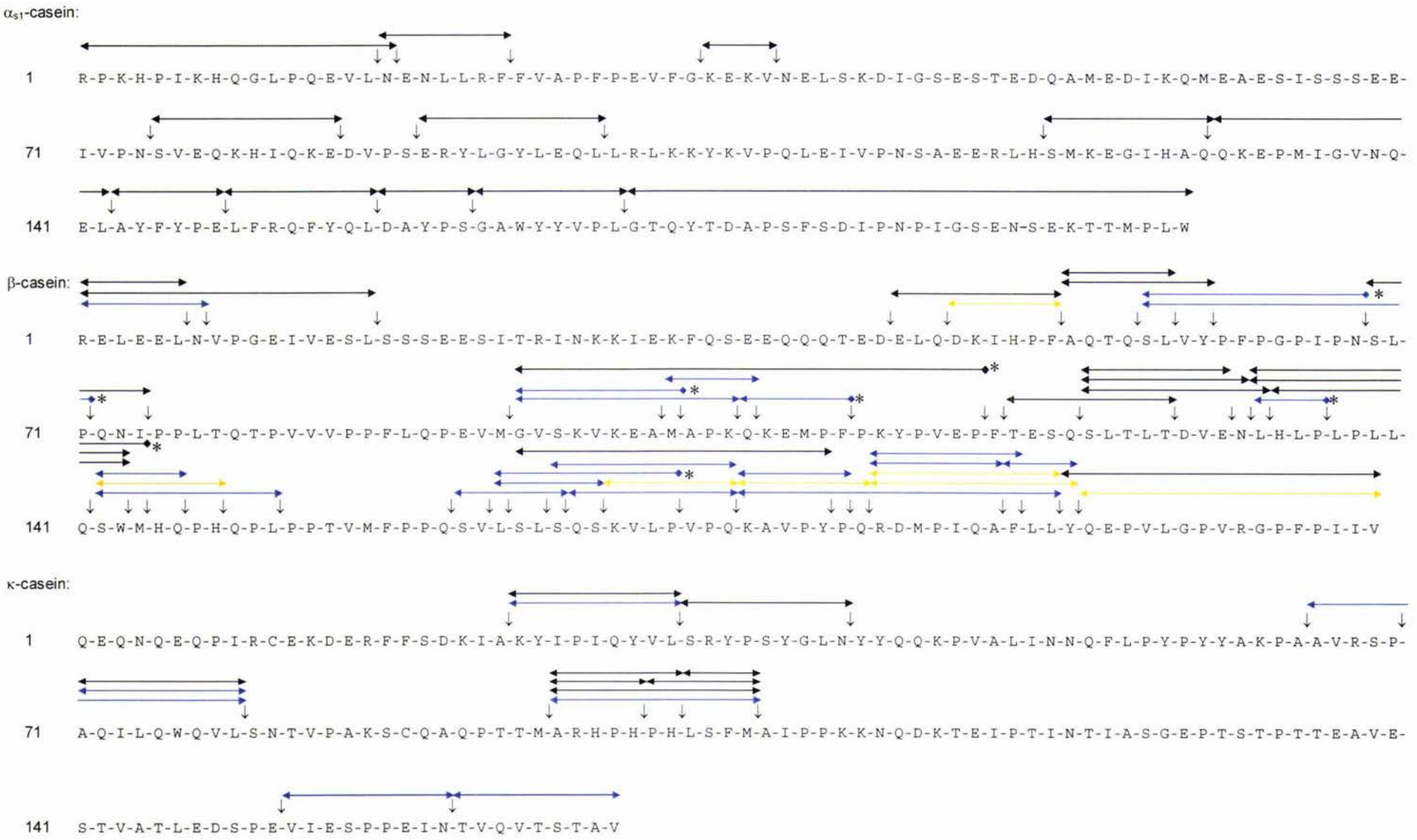


Figure 2.2 *L. lactis* subsp. *cremoris* stratins H2 and SK11 PrtP Cleavage Sites and Soluble Peptide Products of α_{s1}-, β-, and κ-casein (Reid 1991a; Reid 1991b; Reid 1994b). (↓, Cleavage sites; ←→, ←→, ←→, soluble peptides produced by SK11, H2, and both SK11 and H2 PrtP enzymes, respectively; ←→*, C-terminus unknown).

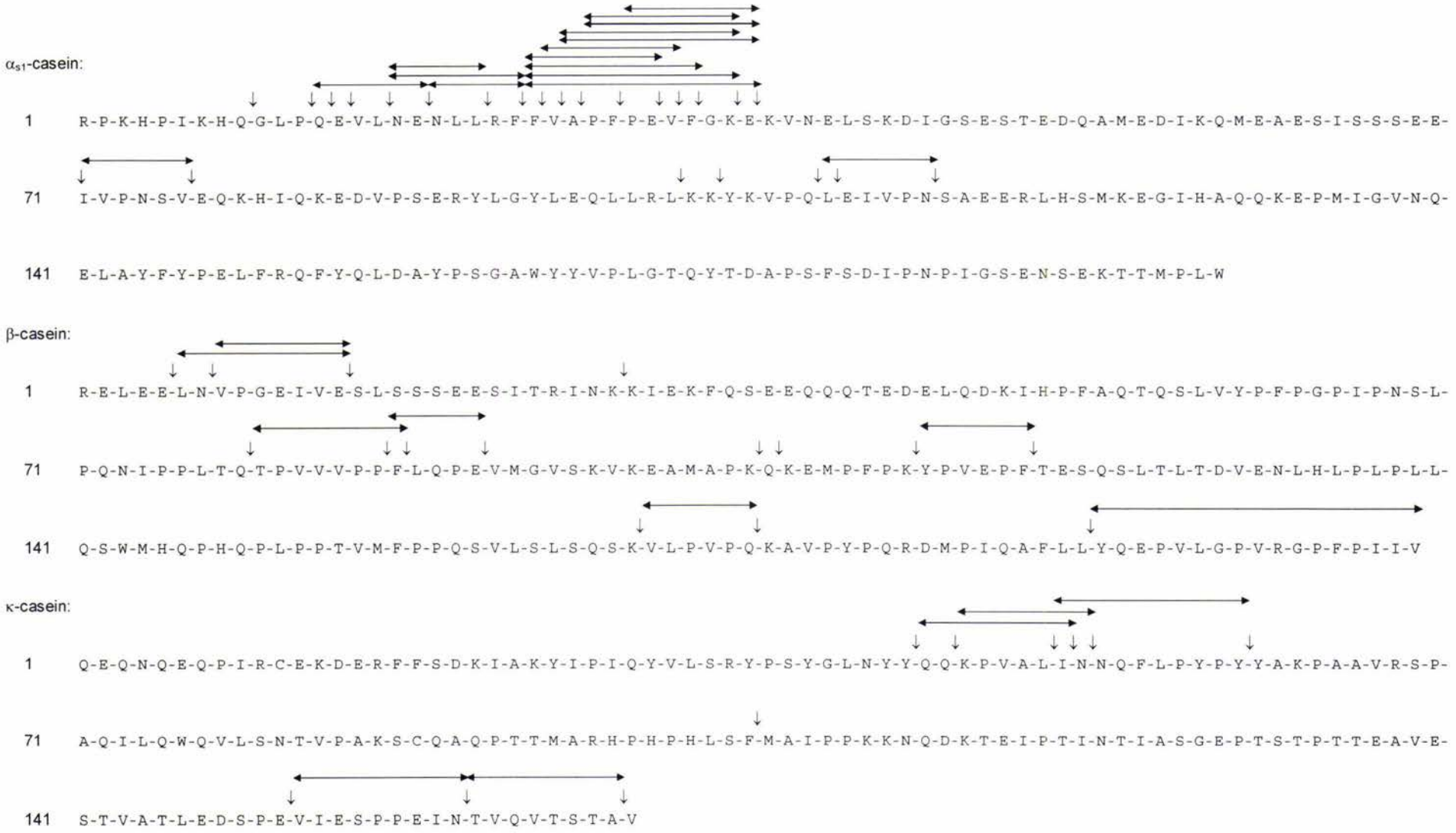


Figure 2.3 α_{s1}-, β-, and κ-casein Peptides Identified in Cheddar Cheese (Alli 1998; Fox 2000).
(↓, Cleavage sites; ←→, soluble peptides produced by proteolytic enzymes).

APPENDIX III

RP-HPLC ANALYSES OF α_{s1}-, β- and κ-CASEIN SUBSTRATE PEPTIDES

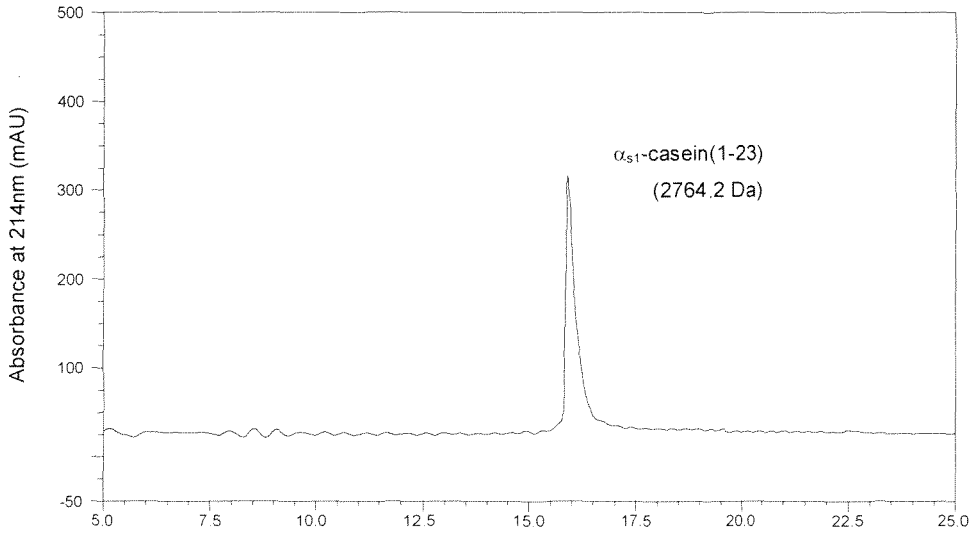


Figure 3.1 RP-HPLC Analysis of α_{s1}-casein(1-23).

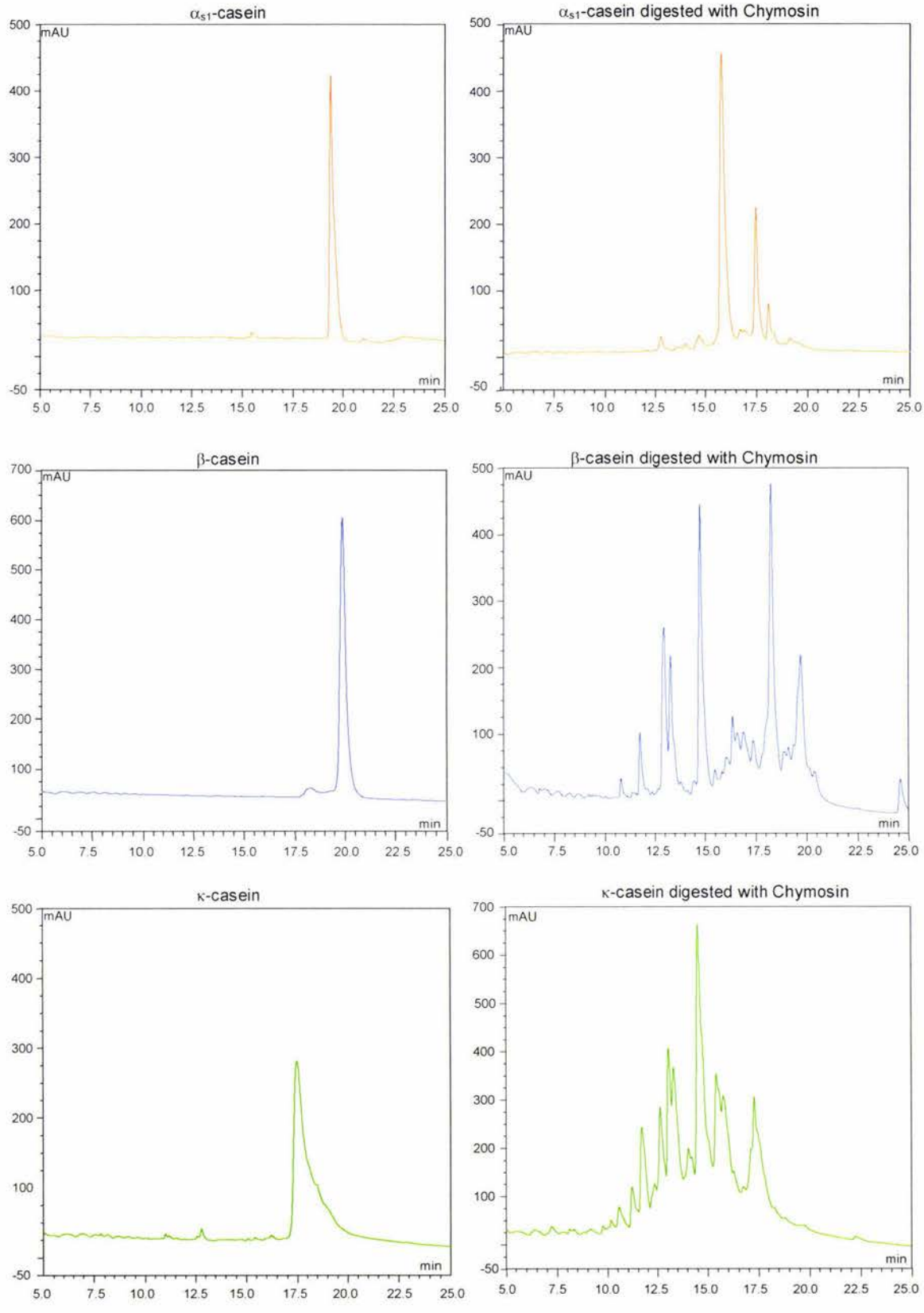


Figure 3.2 RP-HPLC Analysis of Undigested Caseins (left panel) and Casein Peptides Generated from Chymosin Digestion of α_{s1}-, β-, and κ-casein (right panel).

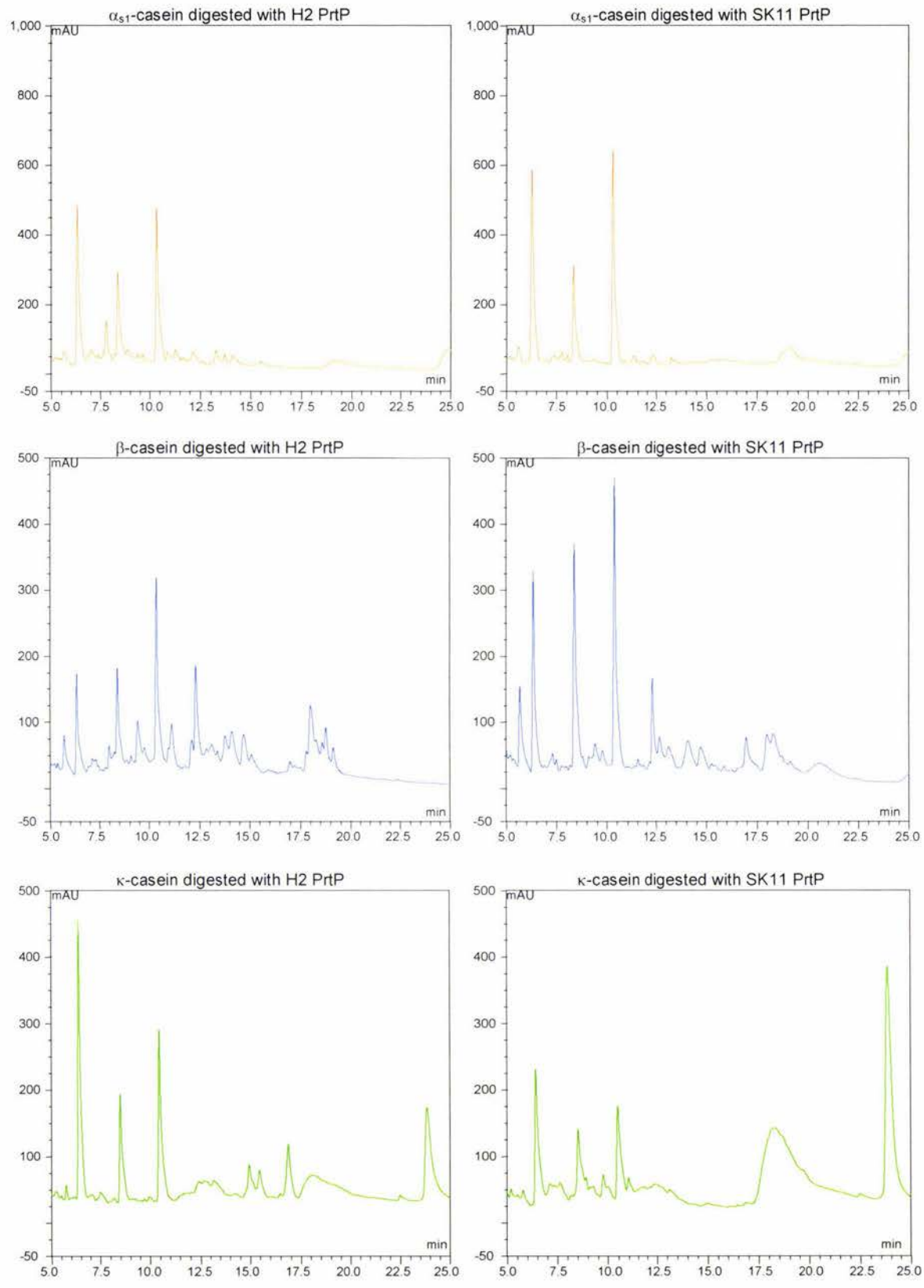


Figure 3.3 RP-HPLC Analysis of Casein Peptides Generated from H2 (left panel) and SK11 (right panel) PrtP Digestion of α_{s1}-, β-, and κ-casein.

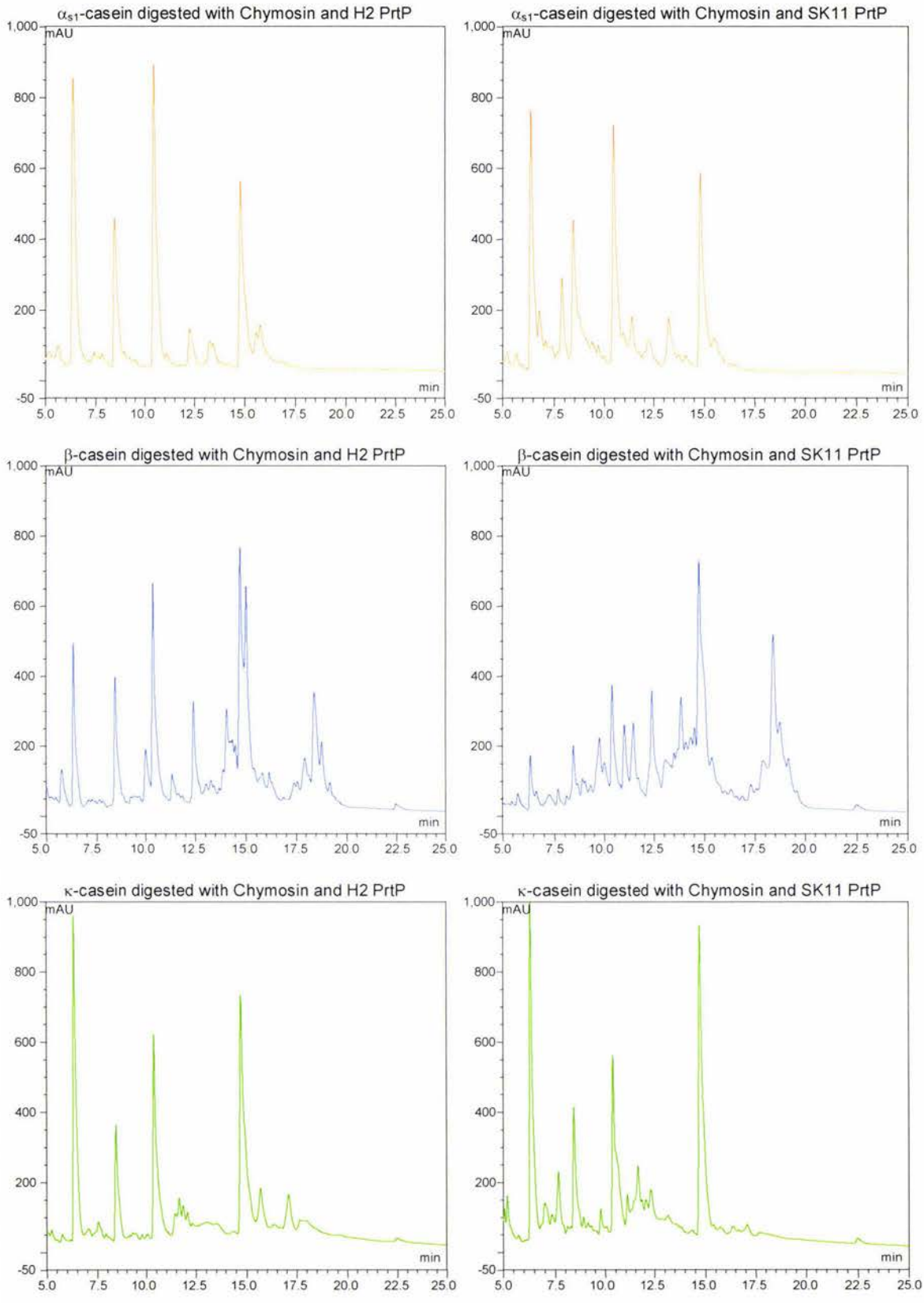


Figure 3.4 RP-HPLC Analysis of Casein Peptides Generated either by the action of Chymosin and either H2 PrtP (left panel) or by Chymosin and SK11 PrtP (right panel) on α_{s1}-, β-, and κ-casein.

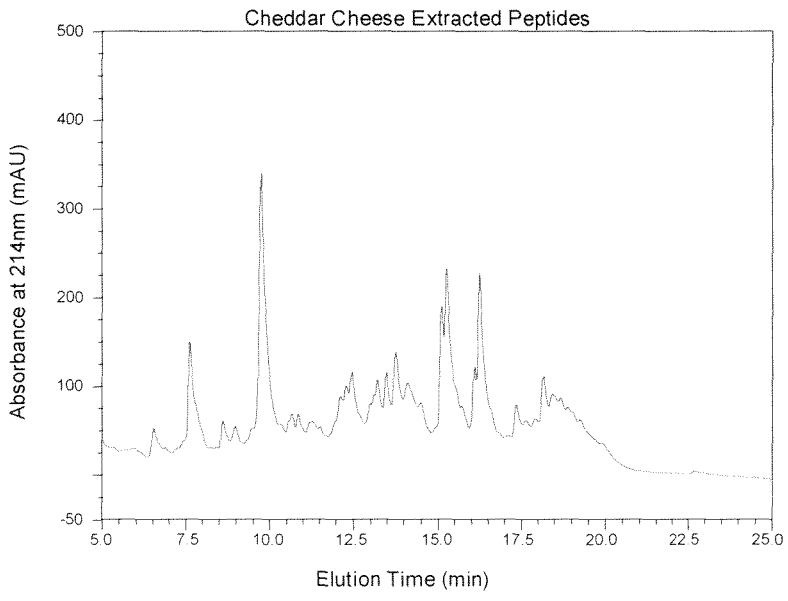


Figure 3.5 RP-HPLC Analysis of the Peptides Extracted from Cheddar cheese.

APPENDIX IV

RP-HPLC ANALYSES OF *L. RHAMNOSUS* PEPO SUBSTRATE SPECIFICITY

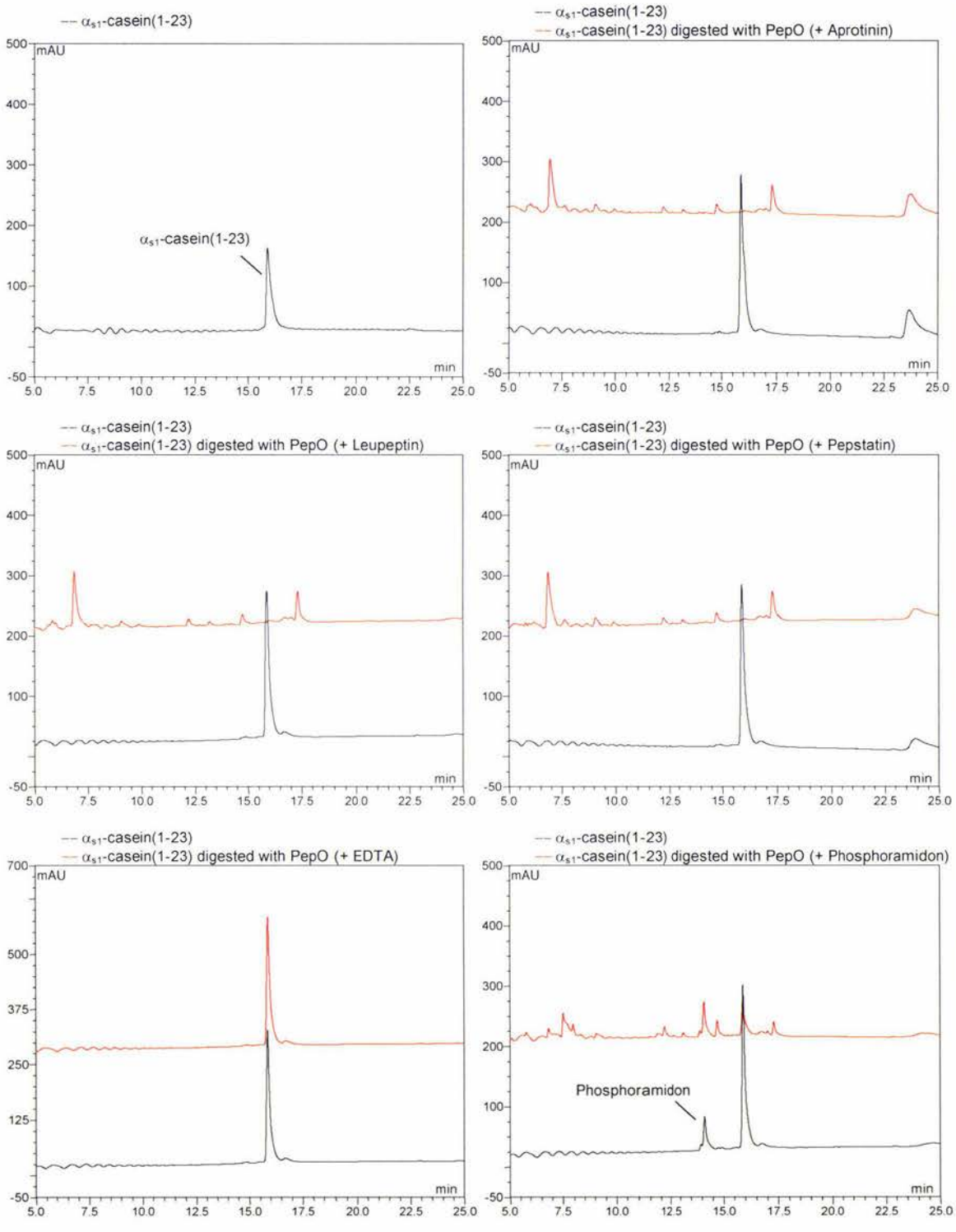


Figure 4.1 RP-HPLC Analysis of PepO Inhibition. (Serine peptidase inhibitors; Aprotinin, Leupeptin, Cysteine peptidase inhibitor, Leupeptin; Aspartic peptidase inhibitor, Pepstatin; Metallo peptidase inhibitors; EDTA and Phosphoramidon).

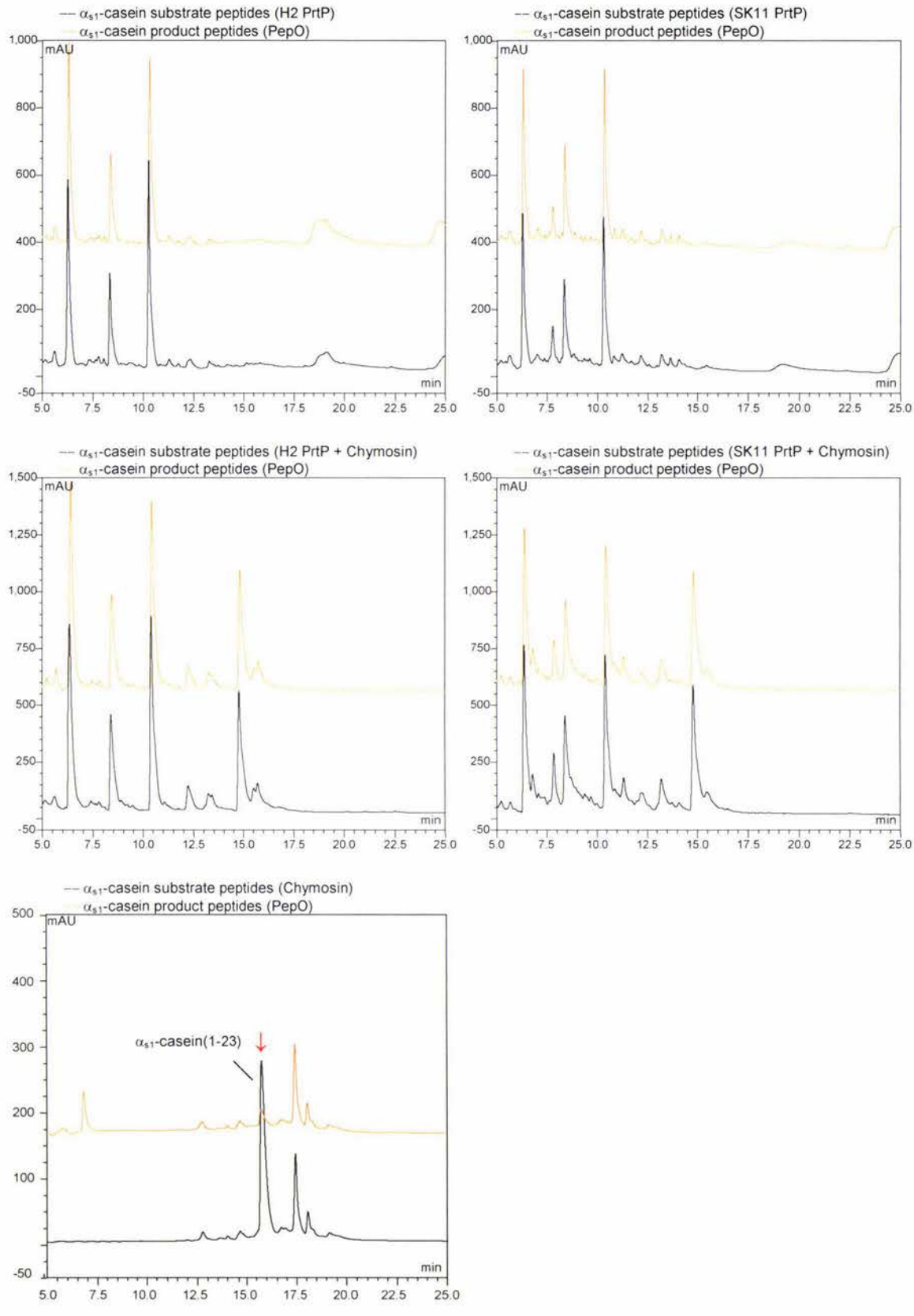


Figure 4.2 RP-HPLC Analysis of PepO Substrate Specificity for α_{s1}-casein Peptides Generated by *L. lactis* H2 and SK11 PrtP enzymes and Chymosin. (↓, Peptides hydrolysed by PepO)

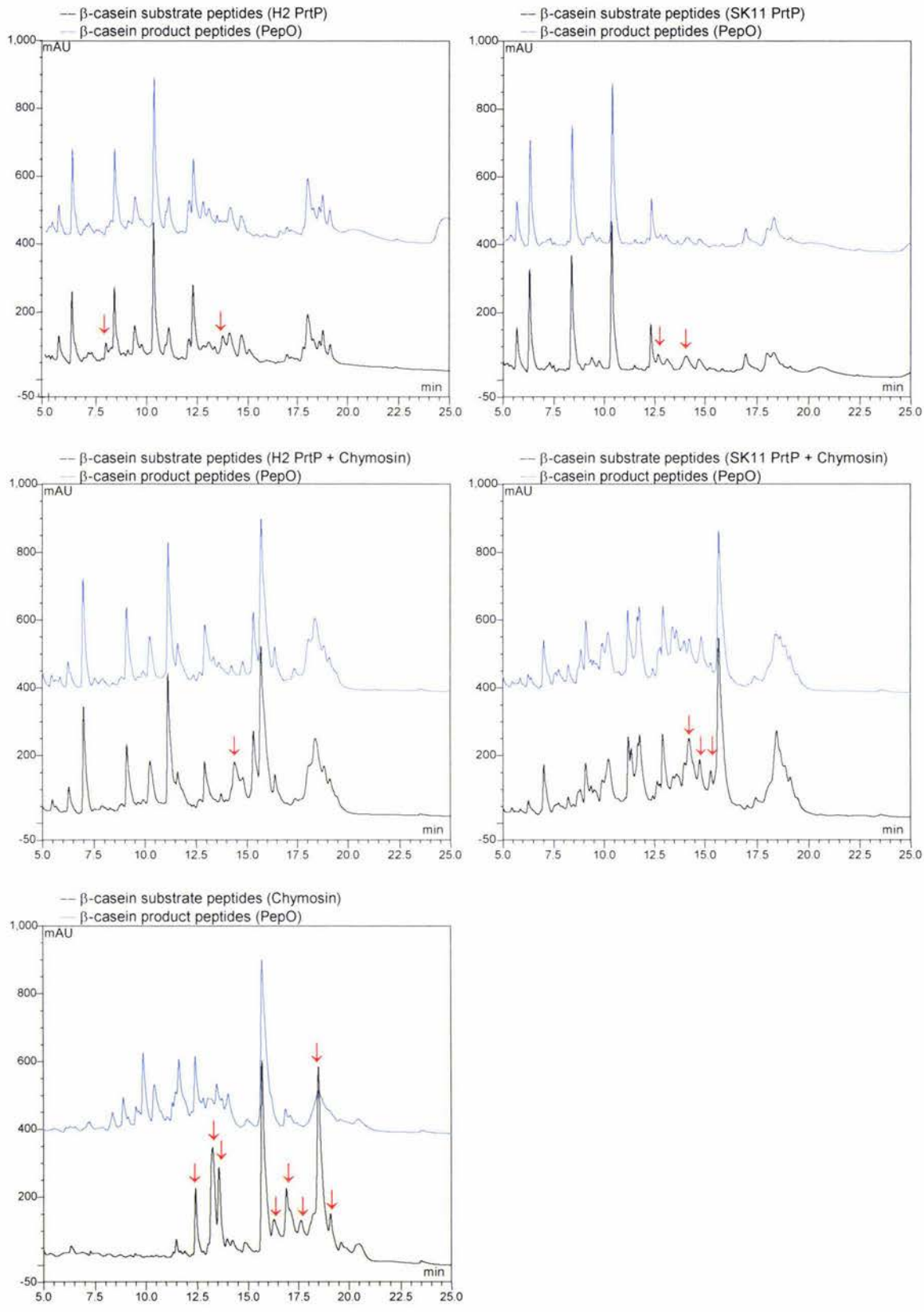


Figure 4.3 RP-HPLC Analysis of PepO Substrate Specificity for β-casein Peptides Generated by *L. lactis* H2 and SK11 PrtP enzymes and Chymosin. (↓, Peptides hydrolysed by PepO).

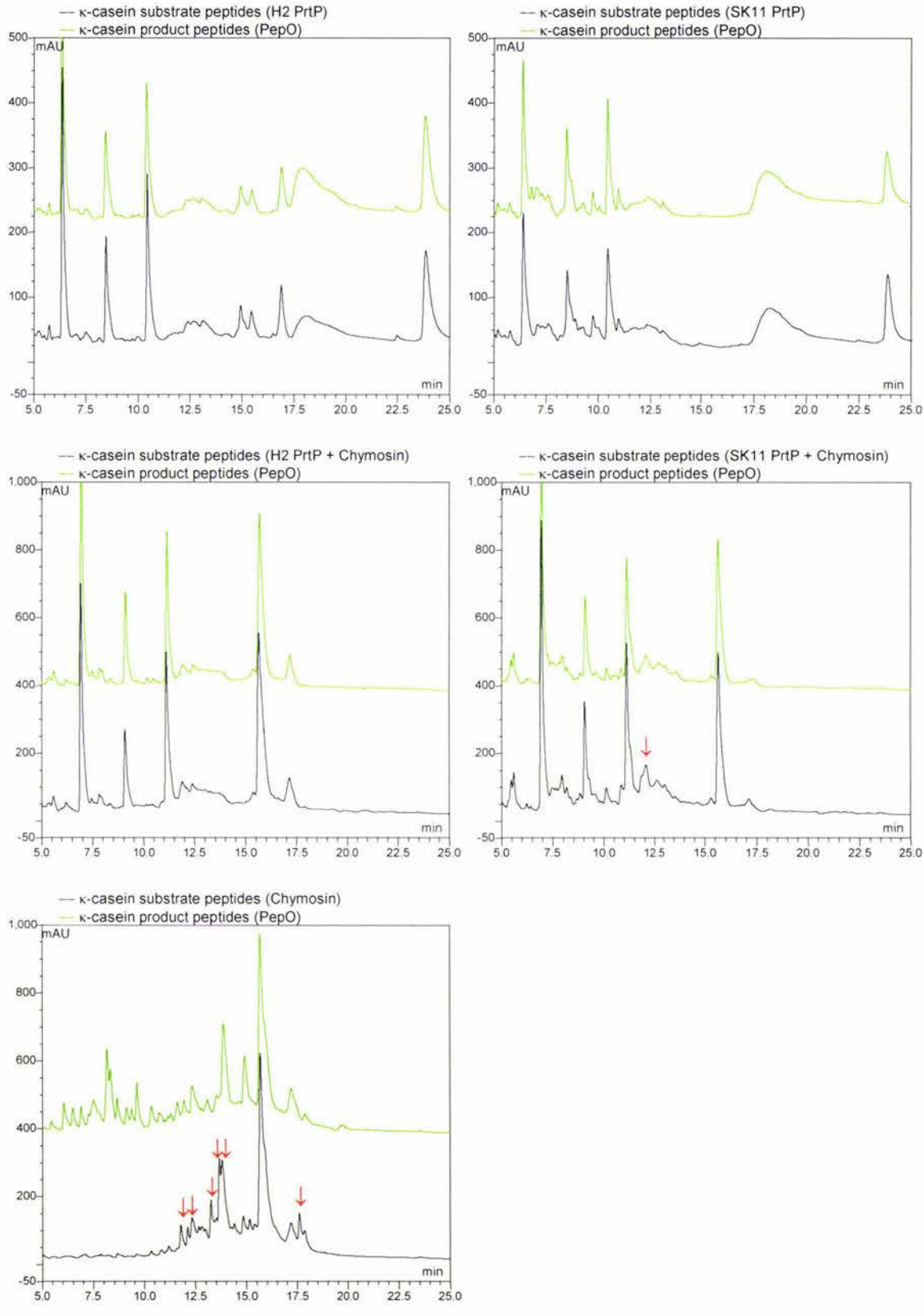


Figure 4.4 RP-HPLC Analysis of PepO Substrate Specificity for κ-casein Peptides Generated by *L. lactis* H2 and SK11 PrtP enzymes and Chymosin. (↓, Peptides hydrolysed by PepO).

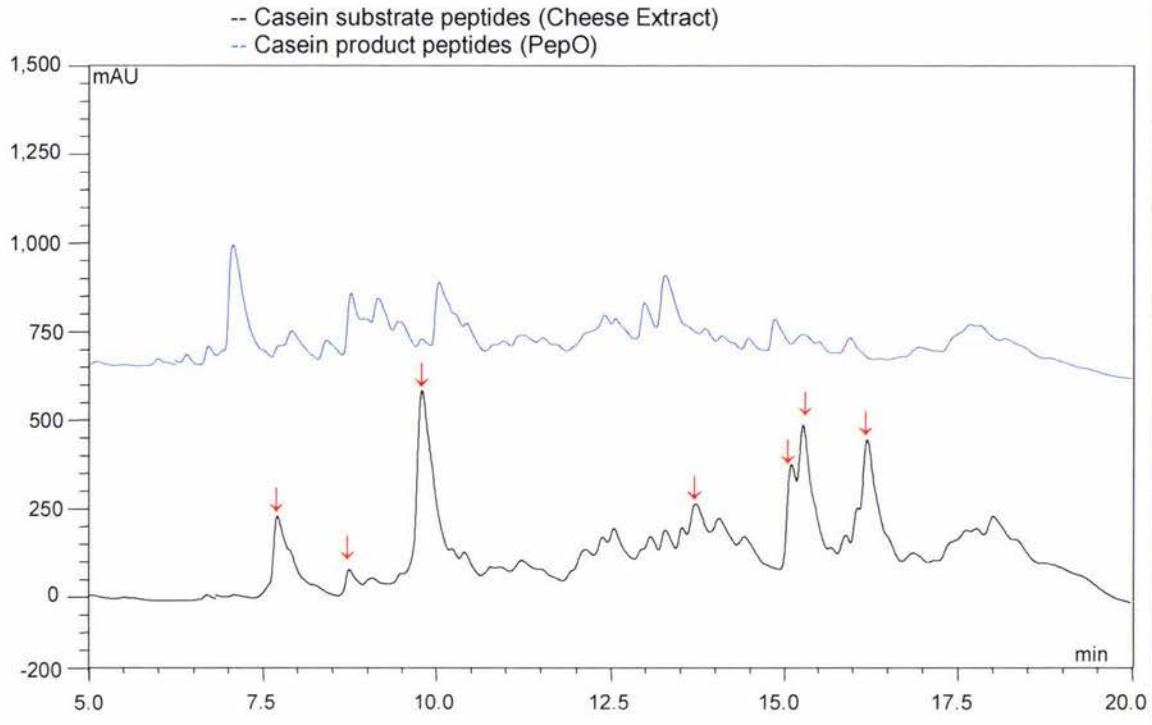


Figure 4.5 RP-HPLC Analysis of PepO Substrate Specificity for Cheddar Cheese Peptides. (↓, Peptides hydrolysed by PepO).

β-casein Peptide Substrates:

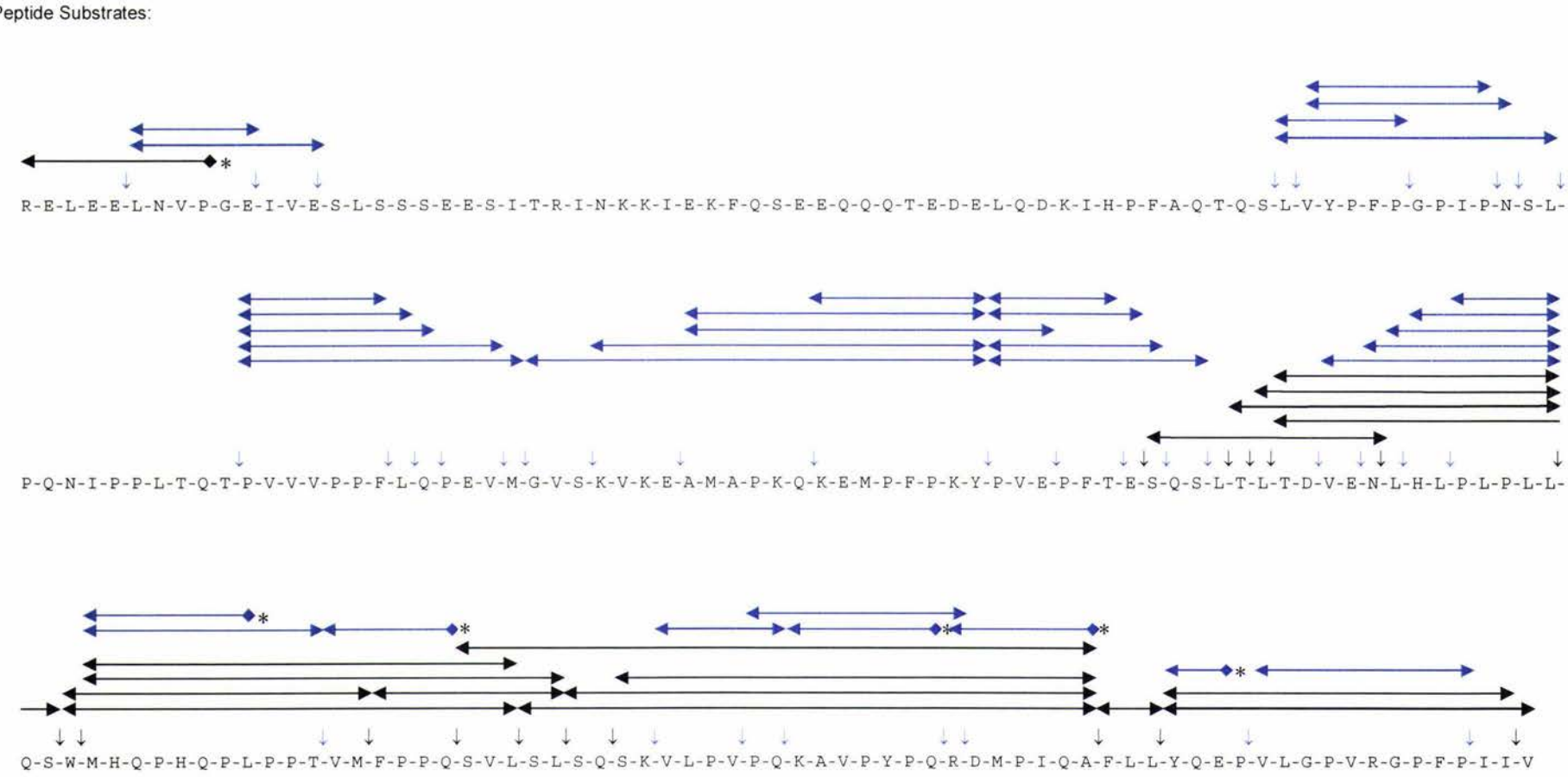


Figure 5.1 The Soluble PepO Substrate Peptides (◆⟷) from β-casein and the Product Peptides (⟷) Resulting from their Hydrolysis by PepO. Chymosin cleavage sites (↓), and possible PepO cleavage sites (↓) are shown. C-terminus unknown (◆⟷*).

κ-casein:

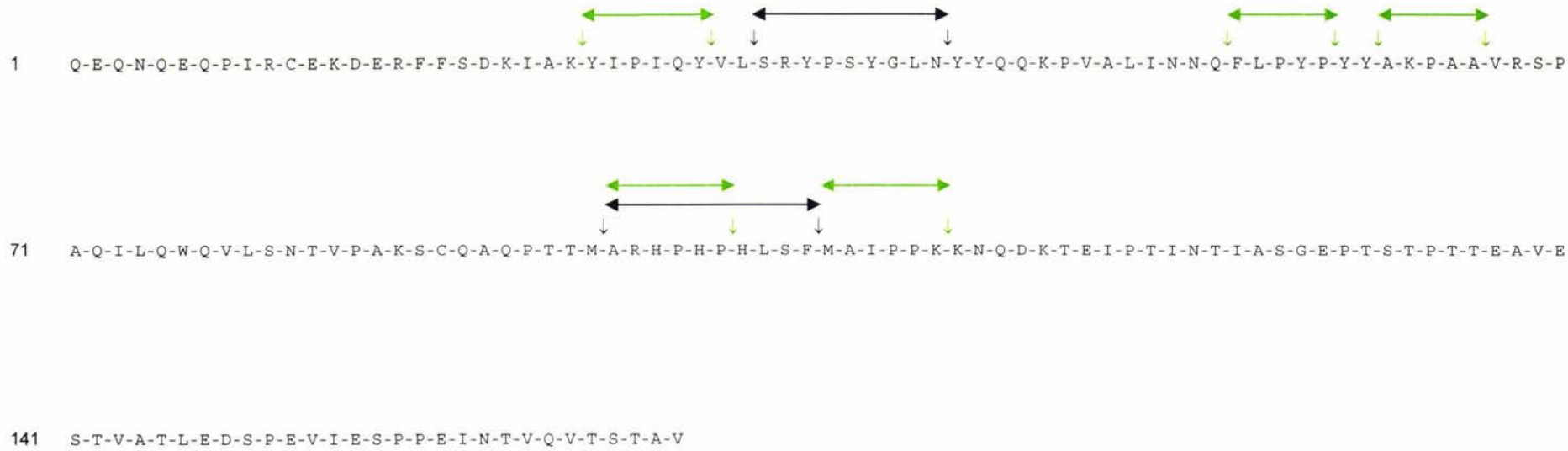
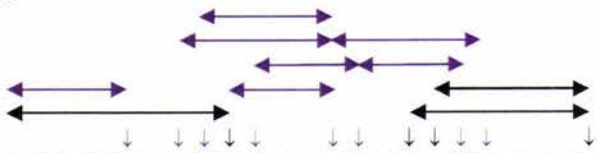


Figure 5.2 The Soluble PepO Substrate Peptides (⟷) from κ-casein and the Product Peptides (⟷) Resulting from their Hydrolysis by PepO. Chymosin cleavage sites (↓), and possible PepO cleavage sites (↓) are shown.

α_{s1}-casein:



```
1    R-P-K-H-P-I-K-H-Q-G-L-P-Q-E-V-L-N-E-N-L-L-R-F-F-V-A-P-F-P-E-V-F-G-K-E-K-V-N-E-L-S-K-D-I-G-S-E-S-T-E-D-Q-A-M-E-D-I-K-Q-M-E-A-E-S-I-S-S-S-E-E-

71   I-V-P-N-S-V-E-Q-K-H-I-Q-K-E-D-V-P-S-E-R-Y-L-G-Y-L-E-Q-L-L-R-L-K-K-Y-K-V-P-Q-L-E-I-V-P-N-S-A-E-E-R-L-H-S-M-K-E-G-I-H-A-Q-Q-K-E-P-M-I-G-V-N-Q-

141  E-L-A-Y-F-Y-P-E-L-F-R-Q-F-Y-Q-L-D-A-Y-P-S-G-A-W-Y-Y-V-P-L-G-T-Q-Y-T-D-A-P-S-F-S-D-I-P-N-P-I-G-S-E-N-S-E-K-T-T-M-P-L-W
```

β-casein:

```
1    R-E-L-E-E-L-N-V-P-G-E-I-V-E-S-L-S-S-S-E-E-S-I-T-R-I-N-K-K-I-E-K-F-Q-S-E-E-Q-Q-Q-T-E-D-E-L-Q-D-K-I-H-P-F-A-Q-T-Q-S-L-V-Y-P-F-P-G-P-I-P-N-S-L-

71   P-Q-N-I-P-P-L-T-Q-T-P-V-V-V-P-P-F-L-Q-P-E-V-M-G-V-S-K-V-K-E-A-M-A-P-K-Q-K-E-M-P-F-P-K-Y-P-V-E-P-F-T-E-S-Q-S-L-T-L-T-D-V-E-N-L-H-L-P-L-P-L-L-

141  Q-S-W-M-H-Q-P-H-Q-P-L-P-P-T-V-M-F-P-P-Q-S-V-L-S-L-S-Q-S-K-V-L-P-V-P-Q-K-A-V-P-Y-P-Q-R-D-M-P-I-Q-A-F-L-L-Y-Q-E-P-V-L-G-P-V-R-G-P-F-P-I-I-V
```

Figure 5.3 The Soluble PepO Substrate Peptides (⟷) from Cheddar Cheese and the Product Peptides (⟷) Resulting from their Hydrolysis by PepO. Cheese proteolytic enzyme cleavage sites (↓), and possible PepO cleavage sites (↓) are shown.

APPENDIX VI

CRYSTALLISATION REAGENTS

Hampton Screen 1:

	Precipitant	Buffer	Salt
1	30% MPD	0.1 M Sodium acetate trihydrate pH 4.6	0.02 M Calcium chloride
2	0.4 M Potassium Sodium tartrate	-	-
3	0.4 M Ammonium dihydrogen phosphate	-	-
4	2.0 M Ammonium sulfate	0.1 M Tris-HCl pH 8.5	-
5	30% MPD	0.1 M Sodium HEPES pH 7.5	0.2 M tri-Sodium citrate
6	30% PEG 4000	0.1 M Tris-HCl pH 8.5	0.2 M Magnesium chloride
7	1.4 M Sodium acetate	0.1 M Sodium cacodylate pH 6.5	-
8	30% Isopropanol	0.1 M Sodium cacodylate pH 6.5	0.2 M tri-Sodium citrate
9	30% PEG 4000	0.1 M tri-Sodium citrate dihydrate pH 5.6	0.2 M Ammonium acetate
10	30% PEG 4000	0.1 M Sodium acetate trihydrate pH 4.6	0.2 M Ammonium acetate
11	1.0 M Ammonium dihydrogen phosphate	0.1 M tri-Sodium citrate dihydrate pH 5.6	-
12	30% Isopropanol	0.1 M Sodium HEPES pH 7.5	0.2 M Magnesium chloride
13	30% PEG 400	0.1 M Tris-HCl pH 8.5	0.2 M tri-Sodium citrate
14	28% PEG 400	0.1 M Sodium HEPES pH 7.5	0.2 M Calcium chloride
15	30% PEG 8000	0.1 M Sodium cacodylate pH 6.5	0.2 M Ammonium sulfate
16	1.5 M Lithium sulfate	0.1 M Sodium HEPES pH 7.5	-
17	30% PEG 4000	0.1 M Tris-HCl pH 8.5	0.2 M Lithium sulfate
18	20% PEG 8000	0.1 M Sodium cacodylate pH 6.5	0.2 M Magnesium acetate
19	30% Isopropanol	0.1 M Tris-HCl pH 8.5	0.2 M Ammonium acetate
20	25% PEG 4000	0.1 M Sodium acetate trihydrate pH 4.6	0.2 M Ammonium sulfate
21	30% MPD	0.1 M Sodium cacodylate pH 6.5	0.2 M Magnesium acetate
22	30% PEG 4000	0.1 M Tris-HCl pH 8.5	0.2 M Sodium acetate
23	30% PEG 400	0.1 M Sodium HEPES pH 7.5	0.2 M Magnesium chloride
24	20% Isopropanol	0.1 M Sodium acetate trihydrate pH 4.6	0.2 M Calcium chloride
25	1.0 M Sodium acetate	0.1 M Imidazole pH 6.5	-
26	30 % MPD	0.1 M tri-Sodium citrate dihydrate pH 5.6	0.2 M Ammonium acetate
27	20% Isopropanol	0.1 M Sodium HEPES pH 7.5	0.2 M tri-Sodium citrate
28	30% PEG 8000	0.1 M Sodium cacodylate pH 6.5	0.2 M Sodium acetate
29	0.8 M Potassium sodium tartrate	0.1 M Sodium HEPES pH 7.5	-
30	30% PEG 8000	-	0.2 M Ammonium sulfate
31	30% PEG 4000	-	0.2 M Ammonium sulfate
32	2.0 M Ammonium sulfate	-	-
33	4.0 M Sodium formate	-	-
34	2.0 M Sodium formate	0.1 M Sodium acetate trihydrate pH 4.6	-
35	0.8 M Sodium dihydrogen phosphate 0.8 M Potassium dihydrogen phosphate	0.1 M Sodium HEPES pH 7.5	-
36	8% PEG 8000	0.1 M Tris-HCl pH 8.5	-
37	8% PEG 4000	0.1 M Sodium acetate trihydrate pH 4.6	-
38	1.4 M tri-Sodium citrate	0.1 M Sodium HEPES pH 7.5	-
39	2% PEG 400, 2.0 M Ammonium sulfate	0.1 M Sodium HEPES pH 7.5	-
40	20% Isopropanol, 20% PEG 4000	0.1 M tri-Sodium citrate dihydrate pH 5.6	-
41	10% Isopropanol, 20% PEG 4000	0.1 M Sodium HEPES pH 7.5	-
42	20% PEG 8000	-	0.05 M Potassium Phosphate
43	30% PEG 1500	-	-
44	0.2 M Magnesium formate	-	-
45	18% PEG 8000	0.1 M Sodium cacodylate pH 6.5	0.2 M Zinc acetate
46	18% PEG 8000	0.1 M Sodium cacodylate pH 6.5	0.2 M Calcium acetate
47	2.0 M Ammonium sulfate	0.1 M Sodium acetate trihydrate pH 4.6	-
48	2.0 M Ammonium dihydrogen phosphate	0.1 M Tris-HCl pH 8.5	-
49	2% PEG 8000	-	1.0 M Lithium sulfate
50	15% PEG 8000	-	0.5 M Lithium sulfate

Hampton Screen 2:

	Precipitant	Buffer	Salt
1	10% PEG 6000	-	2.0 M Sodium chloride
2	0.5 M Sodium chloride, 0.01 M Magnesium chloride	-	0.01 M Hexa-decyl-tri-methyl- Ammonium bromide
3	25% Ethylene Glycol	-	-
4	35% Dioxane	-	-
5	5% Isopropanol	-	2.0 M Ammonium sulfate
6	1.0 M Imidazole pH 7.0	-	-
7	10% PEG 1000, 10% PEG 8000	-	-
8	10% Ethanol	-	1.5 M Sodium chloride
9	2.0 M Sodium chloride	0.1 M Sodium acetate pH 4.6	-
10	30% MPD	0.1 M Sodium acetate pH 4.6	0.2 M Sodium chloride
11	1.0 M 1,6-Hexanediol	0.1 M Sodium acetate pH 4.6	0.01 M Cobaltous chloride
12	30% PEG 400	0.1 M Sodium acetate pH 4.6	0.1 M Cadmium chloride
13	30% PEG Monomethyl ether 2000	0.1 M Sodium acetate pH 4.6	0.2 M Ammonium sulfate
14	2.0 M Ammonium sulfate	0.1 M tri-Sodium citrate pH 5.6	0.2 M Potassium sodium tartrate tetrahydrate
15	1.0 M Lithium sulfate monohydrate	0.1 M tri-Sodium citrate pH 5.6	0.5 M Ammonium sulfate
16	2% Ethylene imine polymer	0.1 M tri-Sodium citrate pH 5.6	0.5 M Sodium chloride
17	35% tert-Butanol	0.1 M tri-Sodium citrate pH 5.6	-
18	10% Jeffamine M-600	0.1 M tri-Sodium citrate pH 5.6	0.01 M Ferric chloride
19	2.5 M 1,6-Hexanediol	0.1 M tri-Sodium citrate pH 5.6	-
20	1.6 M Magnesium sulfate	0.1 M MES pH 6.5	-
21	2.0 M Sodium chloride	0.1 M MES pH 6.5	0.1 M Sodium phosphate, 0.1 M Potassium phosphate
22	12% PEG 20,000	0.1 M MES pH 6.5	-
23	10% Dioxane	0.1 M MES pH 6.5	1.6 M Ammonium sulfate
24	30% Jeffamine M-600	0.1 M MES pH 6.5	0.05 M Cesium chloride
25	1.8 M Ammonium sulfate	0.1 M MES pH 6.5	0.01 M Cobaltous chloride
26	30% PEG Monomethyl ether 5000	0.1 M MES pH 6.5	0.2 M Ammonium sulfate
27	25% PEG Monomethyl ether 550	0.1 M MES pH 6.5	0.01 M Zinc sulfate
28	1.6 M tri-Sodium citrate dihydrate pH 6.5	-	-
29	30% MPD	0.1 M HEPES pH 7.5	0.5 M Ammonium sulfate
30	10% PEG, 5% MPD	0.1 M HEPES pH 7.5	-
31	20% Jeffamine M-600	0.1 M HEPES pH 7.5	-
32	1.6 M Ammonium sulfate	0.1 M HEPES pH 7.5	0.1 M Sodium chloride
33	2.0 M Ammonium formate	0.1 M HEPES pH 7.5	-
34	1.0 M Sodium acetate	0.1 M HEPES pH 7.5	0.05 M Cadmium sulfate
35	70% MPD	0.1 M HEPES pH 7.5	-
36	4.3 M Sodium chloride	0.1 M HEPES pH 7.5	-
37	10% PEG 8000, 8% Ethylene glycol	0.1 M HEPES pH 7.5	-
38	20% PEG 10,000	0.1 M HEPES pH 7.5	-
39	3.4 M 1,6-Hexanediol	0.1 M Tris pH 8.5	0.2 M Magnesium chloride
40	25% tert-Butanol	0.1 M Tris pH 8.5	-
41	1.0 M Lithium sulfate monohydrate	0.1 M Tris pH 8.5	0.01 M Nickel(II) chloride
42	12% Glycerol anhydrous	0.1 M Tris pH 8.5	1.5 M Ammonium sulfate
43	50% MPD	0.1 M Tris pH 8.5	0.2 M Ammonium phosphate
44	20% Ethanol	0.1 M Tris pH 8.5	-
45	20% PEG Monomethyl ether 2000	0.1 M Tris pH 8.5	0.01 M Nickel(II) chloride
46	20% PEG Monomethyl ether 550	0.1 M Bicine pH 9.0	0.1 M Sodium chloride
47	2.0 M Magnesium chloride hexahydrate	0.1 M Bicine pH 9.0	-
48	10% PEG 20,000	0.1 M Bicine pH 9.0	2% Dioxane

2-Methyl-2,4-Pentanediol (MPD) Screen:

	Precipitant	Buffer	Salt
1	40% (v/v) MPD	-	0.2 M Cadmium chloride

2	40% (v/v) MPD	-	0.2 M Potassium fluoride
3	40% (v/v) MPD	-	0.2 M Ammonium fluoride
4	40% (v/v) MPD	-	0.2 M Lithium chloride anhydrous
5	40% (v/v) MPD	-	0.2 M Magnesium chloride hexahydrate
6	40% (v/v) MPD	-	0.2 M Sodium chloride
7	40% (v/v) MPD	-	0.2 M Calcium chloride dihydrate
8	40% (v/v) MPD	-	0.2 M Potassium chloride
9	40% (v/v) MPD	-	0.2 M Ammonium chloride
10	40% (v/v) MPD	-	0.2 M Sodium iodide
11	40% (v/v) MPD	-	0.2 M Potassium iodide
12	40% (v/v) MPD	-	0.2 M Ammonium iodide
13	40% (v/v) MPD	-	0.2 M Sodium thiocyanate
14	40% (v/v) MPD	-	0.2 M Potassium thiocyanate
15	40% (v/v) MPD	-	0.2 M Lithium nitrate
16	40% (v/v) MPD	-	0.2 M Magnesium nitrate hexahydrate
17	40% (v/v) MPD	-	0.2 M Sodium nitrate
18	40% (v/v) MPD	-	0.2 M Potassium nitrate
19	40% (v/v) MPD	-	0.2 M Ammonium nitrate
20	40% (v/v) MPD	-	0.2 M Zinc sulphate heptahydrate
21	40% (v/v) MPD	-	0.2 M Sodium formate
22	40% (v/v) MPD	-	0.2 M Potassium formate
23	40% (v/v) MPD	-	0.2 M Ammonium formate
24	40% (v/v) MPD	-	0.2 M Lithium acetate dihydrate
25	40% (v/v) MPD	-	0.2 M Magnesium acetate tetrahydrate
26	40% (v/v) MPD	-	0.2 M Sodium malonate
27	40% (v/v) MPD	-	0.2 M Sodium acetate trihydrate
28	40% (v/v) MPD	-	0.2 M Calcium acetate hydrate
29	40% (v/v) MPD	-	0.2 M Potassium acetate
30	40% (v/v) MPD	-	0.2 M Ammonium acetate
31	40% (v/v) MPD	-	0.2 M Lithium sulphate monohydrate
32	40% (v/v) MPD	-	0.2 M Lithium sulphate heptahydrate
33	40% (v/v) MPD	-	0.2 M Cesium chloride
34	40% (v/v) MPD	-	0.2 M Ferric chloride
35	40% (v/v) MPD	-	0.2 M Ammonium sulphate
36	40% (v/v) MPD	-	0.2 M di-Sodium tartrate dehydrate
37	40% (v/v) MPD	-	0.2 M Potassium sodium tartrate tetrahydrate
38	40% (v/v) MPD	-	0.2 M di-Ammonium tartrate
39	40% (v/v) MPD	-	0.2 M Sodium dihydrogen phosphate hydrate
40	40% (v/v) MPD	-	0.2 M Potassium bromide
41	40% (v/v) MPD	-	0.2 M Sodium bromide
42	40% (v/v) MPD	-	0.2 M di-Potassium hydrogen phosphate
43	40% (v/v) MPD	-	0.2 M Ammonium dihydrogen phosphate
44	40% (v/v) MPD	-	0.2 M di-Ammonium hydrogen phosphate
45	40% (v/v) MPD	-	0.2 M di-Ammonium hydrogen phosphate
46	40% (v/v) MPD	-	0.2 M tri-Lithium citrate tetrahydrate
47	40% (v/v) MPD	-	0.2 M tri-Potassium citrate monohydrate
48	40% (v/v) MPD	-	0.2 M di-Ammonium hydrogen citrate
49	10% (v/v) MPD	0.1 M Citric acid pH 4.0	
50	10% (v/v) MPD	0.1 M Sodium acetate anhydrous pH 5.0	-
51	10% (v/v) MPD	0.1 M MES pH 6.0	-
52	10% (v/v) MPD	0.1 M HEPES pH 7.0	-
53	10% (v/v) MPD	0.1 M Tris pH 8.0	-
54	10% (v/v) MPD	0.1 M Bicine pH 9.0	-
55	20% (v/v) MPD	0.1 M Citric acid pH 4.0	-
56	20% (v/v) MPD	0.1 M Sodium acetate anhydrous pH 5.0	-
57	20% (v/v) MPD	0.1 M MES pH 6.0	-
58	20% (v/v) MPD	0.1 M HEPES pH 7.0	-
59	20% (v/v) MPD	0.1 M Tris pH 8.0	-
60	20% (v/v) MPD	0.1 M Bicine pH 9.0	-

61	40% (v/v) MPD	0.1 M Citric acid pH 4.0	-
62	40% (v/v) MPD	0.1 M Sodium acetate anhydrous pH 5.0	-
63	40% (v/v) MPD	0.1 M MES pH 6.0	-
64	40% (v/v) MPD	0.1 M HEPES pH 7.0	-
65	40% (v/v) MPD	0.1 M Tris pH 8.0	-
66	40% (v/v) MPD	0.1 M Bicine pH 9.0	-
67	65% (v/v) MPD	0.1 M Citric acid pH 4.0	-
68	65% (v/v) MPD	0.1 M Sodium acetate anhydrous pH 5.0	-
69	65% (v/v) MPD	0.1 M MES pH 6.0	-
70	65% (v/v) MPD	0.1 M HEPES pH 7.0	-
71	65% (v/v) MPD	0.1 M Tris pH 8.0	-
72	65% (v/v) MPD	0.1 M Bicine pH 9.0	-
73	10% (v/v) MPD	0.1 M HEPES sodium salt 0H 7.5	0.1 M tri-Sodium citrate dehydrate
74	12% (v/v) MPD	0.1 M Tris-HCL pH 8.5	0.05 M Magnesium chloride heptahydrate
75	15% (v/v) MPD	0.1 M Sodium acetate pH 4.6	0.02 M Calcium chloride dehydrate
76	15% (v/v) MPD, 5% (w/v) PEG 4000	0.1 M Imidazole-HCl pH 8.0	-
77	15% (v/v) MPD	0.1 M Sodium citrate pH 5.6	0.2 M Ammonium acetate
78	15% (v/v) MPD	0.1 M MES sodium salt pH 6.5	0.2 M Magnesium acetate
79	15% (v/v) MPD	0.1 M HEPES sodium salt pH 7.5	0.2 M tri-Sodium citrate dehydrate
80	20% (v/v) MPD	0.1 M HEPES sodium salt pH 7.5	0.1 M tri-Sodium citrate dehydrate
81	20% (v/v) MPD	0.1 M Imidazole-HCl pH 8.0	-
82	40% (v/v) MPD, 4% (w/v) Glycerol	-	0.2 M Sodium chloride
83	30% (v/v) MPD	0.1 M Sodium acetate pH 4.6	0.02 M Calcium chloride dehydrate
84	30% (v/v) MPD	0.1 M Sodium citrate pH 5.6	0.2 M Ammonium acetate
85	30% (v/v) MPD	0.1 M MES sodium salt pH 6.5	0.2 M Magnesium acetate tetrahydrate
86	30% (v/v) MPD	0.1 M HEPES sodium salt pH 7.5	0.5 M Ammonium sulphate
87	30% (v/v) MPD	0.1 M HEPES sodium salt pH 7.5	0.2 M tri-Sodium citrate dihydrate
88	30% (v/v) MPD, 5% (w/v) PEG 4000	0.1 M HEPES sodium salt pH 7.5	-
89	30% (v/v) MPD, 10% (w/v) PEG 4000	0.1 M Imidazole-HCl pH 8.0	-
90	30% (v/v) MPD, 20% (w/v) \ethanol	-	-
91	35% (v/v) MPD	-	-
92	35% (v/v) MPD	0.1 M Imidazole-HCl pH 8.0	-
93	40% (v/v) MPD	0.1 M Tris-HCL pH 8.5	-
94	47% (v/v) MPD	0.1 M HEPES sodium salt pH 7.5	-
95	47% (v/v) MPD, 2% (w/v) Tert-	-	-
96	50% (v/v) MPD	-	-

Poly Ethylene Glycol (PEG) Screen:

	Precipitant	Buffer	Salt
1	7% PEG 6000	0.2 M Citric acid-KOH pH 4.9	-
2	14% PEG 6000	0.2 M Acetic acid-KOH pH 4.9	-
3	21% PEG 6000	0.2 M Citric acid-KOH pH 4.9	-
4	28% PEG 6000	0.2 M Acetic acid-KOH pH 4.9	-
5	7% MPEG 5000	0.2 M Acetic acid-KOH pH 4.9	-
6	14% MPEG 5000	0.2 M Citric acid-KOH pH 4.9	-
7	21% MPEG 5000	0.2 M Acetic acid-KOH pH 4.9	-
8	28% MPEG 5000	0.2 M Citric acid-KOH pH 4.9	-
9	7% PEG 6000	0.2 M Succinic acid-KOH pH 5.5	-
10	14% PEG 6000	0.2 M Malic acid-KOH pH 5.5	-
11	21% PEG 6000	0.2 M Succinic acid-KOH pH 5.5	-
12	28% PEG 6000	0.2 M Malic acid-KOH pH 5.5	-
13	7% MPEG 5000	0.2 M Malic acid-KOH pH 5.5	-
14	14% MPEG 5000	0.2 M Succinic acid-KOH pH 5.5	-
15	21% MPEG 5000	0.2 M Malic acid-KOH pH 5.5	-
16	28% MPEG 5000	0.2 M Succinic acid-KOH pH 5.5	-
17	7% PEG 6000	0.2 M Cacodylic acid-KOH pH 6.1	-
18	14% PEG 6000	0.2 M MES-KOH pH 6.1	-
19	21% PEG 6000	0.2 M Cacodylic acid-KOH pH 6.1	-

20	28% PEG 6000	0.2 M MES-KOH pH 6.1	-
21	7% MPEG 5000	0.2 M MES-KOH pH 6.1	-
22	14% MPEG 5000	0.2 M Cacodylic acid-KOH pH 6.1	-
23	21% MPEG 5000	0.2 M MES-KOH pH 6.1	-
24	28% MPEG 5000	0.2 M Cacodylic acid-KOH pH 6.1	-
25	7% PEG 6000	0.2 M PIPES-KOH pH 6.7	-
26	14% PEG 6000	0.2 M BTP-HCl pH 6.7	-
27	21% PEG 6000	0.2 M PIPES-KOH pH 6.7	-
28	28% PEG 6000	0.2 M BTP-HCl pH 6.7	-
29	7% MPEG 5000	0.2 M BTP-HCl pH 6.7	-
30	14% MPEG 5000	0.2 M PIPES-KOH pH 6.7	-
31	21% MPEG 5000	0.2 M BTP-HCl pH 6.7	-
32	28% MPEG 5000	0.2 M PIPES-KOH pH 6.7	-
33	7% PEG 6000	0.2 M MOPS-KOH pH 7.3	-
34	14% PEG 6000	0.2 M HEPES-KOH pH 7.3	-
35	21% PEG 6000	0.2 M MOPS-KOH pH 7.3	-
36	28% PEG 6000	0.2 M HEPES-KOH pH 7.3	-
37	7% MPEG 5000	0.2 M HEPES-KOH pH 7.3	-
38	14% MPEG 5000	0.2 M MOPS-KOH pH 7.3	-
39	21% MPEG 5000	0.2 M HEPES-KOH pH 7.3	-
40	28% MPEG 5000	0.2 M MOPS-KOH pH 7.3	-
41	7% PEG 6000	0.2 M EPPS-KOH pH 7.9	-
42	14% PEG 6000	0.2 M Tris-Hcli pH 7.9	-
43	21% PEG 6000	0.2 M EPPS-KOH pH 7.9	-
44	28% PEG 6000	0.2 M Tris-Hcli pH 7.9	-
45	7% MPEG 5000	0.2 M Tris-Hcli pH 7.9	-
46	14% MPEG 5000	0.2 M EPPS-KOH pH 7.9	-
47	21% MPEG 5000	0.2 M Tris-Hcli pH 7.9	-
48	28% MPEG 5000	0.2 M EPPS-KOH pH 7.9	-
49	7% PEG 6000	0.2 M BTP-HCl pH 8.5	-
50	14% PEG 6000	0.2 M TAPS-KOH pH 8.5	-
51	21% PEG 6000	0.2 M BTP-HCl pH 8.5	-
52	28% PEG 6000	0.2 M TAPS-KOH pH 8.5	-
53	7% MPEG 5000	0.2 M TAPS-KOH pH 8.5	-
54	14% MPEG 5000	0.2 M BTP-HCl pH 8.5	-
55	21% MPEG 5000	0.2 M TAPS-KOH pH 8.5	-
56	28% MPEG 5000	0.2 M BTP-HCl pH 8.5	-
57	7%	0.2 M AMPSO-KOH pH 9.1	-
58	14%	0.2 M Boric acid-KOH pH 9.1	-
59	21%	0.2 M AMPSO-KOH pH 9.1	-
60	28%	0.2 M Boric acid-KOH pH 9.1	-
61	7% MPEG 5000	0.2 M Boric acid-KOH pH 9.1	-
62	14% MPEG 5000	0.2 M AMPSO-KOH pH 9.1	-
63	21% MPEG 5000	0.2 M Boric acid-KOH pH 9.1	-
64	28% MPEG 5000	0.2 M AMPSO-KOH pH 9.1	-

PEG/Ion Screen Formulation:

	Precipitant	Buffer	Salt
1	20% PEG 3350	-	0.2 M Sodium fluoride
2	20% PEG 3350	-	0.2 M Potassium fluoride
3	20% PEG 3350	-	0.2 M Ammonium fluoride
4	20% PEG 3350	-	0.2 M Lithium chloride
5	20% PEG 3350	-	0.2 M Magnesium chloride
6	20% PEG 3350	-	0.2 M Sodium chloride
7	20% PEG 3350	-	0.2 M Calcium chloride
8	20% PEG 3350	-	0.2 M Potassium chloride
9	20% PEG 3350	-	0.2 M Ammonium chloride

10	20% PEG 3350	-	0.2 M Sodium Iodide
11	20% PEG 3350	-	0.2 M Potassium iodide
12	20% PEG 3350	-	0.2 M Ammonium iodide
13	20% PEG 3350	-	0.2 M Sodium thiocyanate
14	20% PEG 3350	-	0.2 M Potassium thiocyanate
15	20% PEG 3350	-	0.2 M Lithium nitrate
16	20% PEG 3350	-	0.2 M Magnesium nitrate
17	20% PEG 3350	-	0.2 M Sodium nitrate
18	20% PEG 3350	-	0.2 M Potassium nitrate
19	20% PEG 3350	-	0.2 M Ammonium nitrate
20	20% PEG 3350	-	0.2 M Magnesium formate
21	20% PEG 3350	-	0.2 M Sodium formate
22	20% PEG 3350	-	0.2 M Potassium formate
23	20% PEG 3350	-	0.2 M Ammonium formate
24	20% PEG 3350	-	0.2 M Lithium acetate
25	20% PEG 3350	-	0.2 M Magnesium acetate
26	20% PEG 3350	-	0.2 M Zinc acetate
27	20% PEG 3350	-	0.2 M Sodium acetate
28	20% PEG 3350	-	0.2 M Calcium acetate
29	20% PEG 3350	-	0.2 M Potassium acetate
30	20% PEG 3350	-	0.2 M Ammonium acetate
31	20% PEG 3350	-	0.2 M Lithium sulfate
32	20% PEG 3350	-	0.2 M Magnesium sulfate
33	20% PEG 3350	-	0.2 M Sodium sulfate
34	20% PEG 3350	-	0.2 M Potassium sulfate
35	20% PEG 3350	-	0.2 M Ammonium sulfate
36	20% PEG 3350	-	0.2 M di-Sodium tartrate
37	20% PEG 3350	-	0.2 M Potassium sodium tartrate
38	20% PEG 3350	-	0.2 M di-Ammonium tartrate
39	20% PEG 3350	-	0.2 M Sodium dihydrogen phosphate
40	20% PEG 3350	-	0.2 M di-Sodium hydrogen phosphate
41	20% PEG 3350	-	0.2 M Potassium dihydrogen phosphate
42	20% PEG 3350	-	0.2 M di-Potassium hydrogen phosphate
43	20% PEG 3350	-	0.2 M Ammonium dihydrogen phosphate
44	20% PEG 3350	-	0.2 M di-Ammonium hydrogen phosphate
45	20% PEG 3350	-	0.2 M tri-Lithium citrate
46	20% PEG 3350	-	0.2 M tri-Sodium citrate
47	20% PEG 3350	-	0.2 M tri-Potassium citrate
48	20% PEG 3350	-	0.2 M di-Ammonium hydrogen citrate

Footprint Screen:

	Precipitant	Buffer	Salt
1	15% PEG 600	0.2 M Imidazole malate pH 5.5	-
2	10% PEG 4000	0.2 M Imidazole malate pH 7.0	-
3	7.5% PEG 10000	0.2 M Imidazole malate pH 8.5	-
4	0.75 M $(NH_4)_2SO_4$	0.15 M Sodium citrate pH 5.5	-
5	-	0.8 M Sodium dihydrogen phosphate pH 7.0	-
6	0.75 M Citrate	10 mM Sodium borate pH 8.5	-
7	24% PEG 600	0.2 M Imidazole malate pH 5.5	-
8	15% PEG 4000	0.2 M Imidazole malate pH 7.0	-
9	12.5% PEG 10000	0.2 M Imidazole malate pH 8.5	-
10	1.0 M $(NH_4)_2SO_4$	0.15 M Sodium citrate pH 5.5	-
11	-	1.32 M Sodium dihydrogen phosphate pH 7.0	-
12	1.0 M Citrate	10 mM Sodium borate pH 8.5	-
13	33% PEG 600	0.2 M Imidazole malate pH 5.5	-
14	20% PEG 4000	0.2 M Imidazole malate pH 7.0	-
15	17.5% PEG 10000	0.2 M Imidazole malate pH 8.5	-

16	1.5 M $(NH_4)_2SO_4$	0.15 M Sodium citrate pH 5.5	-
17	-	1.6 M Sodium dihydrogen phosphate pH 7.0	-
18	1.2 M Citrate	10 mM Sodium borate pH 8.5	-
19	42% PEG 600	0.2 M Imidazole malate pH 5.5	-
20	25% PEG 4000	0.2 M Imidazole malate pH 7.0	-
21	22.5% PEG 10000	0.2 M Imidazole malate pH 8.5	-
22	2.0 M $(NH_4)_2SO_4$	0.15 M Sodium citrate pH 5.5	-
23	-	2.0 M Sodium dihydrogen phosphate pH 7.0	-
24	1.5 M Citrate	10 mM Sodium borate pH 8.5	-

PEG Footprint Screen:

	Precipitant	Buffer	Salt
1	30% MPEG 550	0.1 M HEPES pH 8.2	-
2	18% PEG 600	0.1 M HEPES pH 7.5	-
3	18% MPEG 2000	0.1 M Sodium cacodylate pH 6.5	-
4	8% PEG 4000	0.2 M Imidazole malate pH 6.0	-
5	12% MPEG 5000	0.1 M Sodium acetate pH 5.5	-
6	9% PEG 10000	0.1 M Ammonium acetate pH 4.5	-
7	40% MPEG 550	0.1 M HEPES pH 8.2	-
8	27% PEG 600	0.1 M HEPES pH 7.5	-
9	27% MPEG 2000	0.1 M Sodium cacodylate pH 6.5	-
10	15% PEG 4000	0.2 M Imidazole malate pH 6.0	-
11	18% MPEG 5000	0.1 M Sodium acetate pH 5.5	-
12	15% PEG 10000	0.1 M Ammonium acetate pH 4.5	-
13	50% MPEG 550	0.1 M HEPES pH 8.2	-
14	36% PEG 600	0.1 M HEPES pH 7.5	-
15	36% MPEG 2000	0.1 M Sodium cacodylate pH 6.5	-
16	20% PEG 4000	0.2 M Imidazole malate pH 6.0	-
17	24% MPEG 5000	0.1 M Sodium acetate pH 5.5	-
18	22.5% PEG 10000	0.1 M Ammonium acetate pH 4.5	-
19	60% MPEG 550	0.1 M HEPES pH 8.2	-
20	45% PEG 600	0.1 M HEPES pH 7.5	-
21	45% MPEG 2000	0.1 M Sodium cacodylate pH 6.5	-
22	30% PEG 4000	0.2 M Imidazole malate pH 6.0	-
23	36% MPEG 5000	0.1 M Sodium acetate pH 5.5	-
24	27% PEG 10000	0.1 M Ammonium acetate pH 4.5	-

Ammonium Sulfate Screen:

		Buffer	Salt
1	3.2 M Ammonium sulfate	pH 4.0	-
2	3.2 M Ammonium sulfate	pH 5.0	-
3	3.2 M Ammonium sulfate	pH 6.0	-
4	3.2 M Ammonium sulfate	pH 7.0	-
5	3.2 M Ammonium sulfate	pH 8.0	-
6	3.2 M Ammonium sulfate	pH 9.0	-
7	2.4 M Ammonium sulfate	pH 4.0	-
8	2.4 M Ammonium sulfate	pH 5.0	-
9	2.4 M Ammonium sulfate	pH 6.0	-
10	2.4 M Ammonium sulfate	pH 7.0	-
11	2.4 M Ammonium sulfate	pH 8.0	-
12	2.4 M Ammonium sulfate	pH 9.0	-
13	1.6 M Ammonium sulfate	pH 4.0	-
14	1.6 M Ammonium sulfate	pH 5.0	-
15	1.6 M Ammonium sulfate	pH 6.0	-
16	1.6 M Ammonium sulfate	pH 7.0	-

17	1.6 M Ammonium sulfate	pH 8.0	-
18	1.6 M Ammonium sulfate	pH 9.0	-
19	0.8 M Ammonium sulfate	pH 4.0	-
20	0.8 M Ammonium sulfate	pH 5.0	-
21	0.8 M Ammonium sulfate	pH 6.0	-
22	0.8 M Ammonium sulfate	pH 7.0	-
23	0.8 M Ammonium sulfate	pH 8.0	-
24	0.8 M Ammonium sulfate	pH 9.0	-

Top 67 Screen:

	Precipitant	Buffer	Salt
1	50% PEG 400	0.1 M Acetate pH 5.1	0.2 M Lithium sulfate
2	20% PEG 3000	0.1 M Citrate pH 5.5	-
3	20% PEG 3350	0.2 M di-Ammonium hydrogen citrate pH 4.85	-
4	30% MPD	0.1 M Sodium acetate pH 4.6	0.02 M Calcium chloride
5	20% PEG 3350	0.2 M Magnesium formate pH 5.9	-
6	20% PEG 1000	0.2 M Phosphate-citrate pH 4.2	0.2 M Lithium sulfate
7	20% PEG 8000	0.1 M CHES pH 9.5	-
8	20% PEG 3350	0.2 M Ammonium formate pH 6.2	-
9	20% PEG 3350	0.2 M Ammonium chloride pH 4.0	-
10	20% PEG 3350	0.2 M Potassium formate pH 7.6	-
11	50% MPD	0.1 M Tris pH 8.5	0.2 M Ammonium dihyrogen phosphate
12	20% PEG 3350	0.2 M Potassium nitrate pH 6.3	-
13	-	0.1 M Citric acid pH 4.0	0.8 M Ammonium sulfate
14	20% PEG 3350	0.2 M Sodium thiocyanate pH 7.0	-
15	20% PEG 6000	0.1 M Bicine pH 9.0	-
16	10% PEG 8000, 8% ethylene glycol	0.1 M HEPES pH 7.5	-
17	40% MPD, 5% PEG 8000	0.1 M Cacodylate pH 7.0	-
18	40% Ethanol, 5% PEG 1000	0.1 M Phosphate-citrate pH 5.2	-
19	8% PEG 4000	0.1 M Sodium acetate pH 4.6	-
20	10% PEG 8000	0.1 M Tris pH 7.0	0.2 M Magnesium chloride
21	20% PEG 6000	0.1 M Citric acid pH 5.0	-
22	50% PEG 200	0.1 M Cacodylate pH 6.6	0.2 M Magnesium chloride
23	-	1.6 M Sodium citrate pH 6.5	-
24	20% PEG 3350	0.2 M tri-Potassium citrate hydrate pH 8.3	-
25	30% MPD	0.1 M Sodium acetate pH 4.6	0.02 Calcium chloride
26	20% PEG 8000	0.1 M Phosphate-citrate pH 4.2	0.2 M Sodium chloride
27	20% PEG 6000	0.1 M Citric acid pH 4.0	1.0 M Lithium chloride
28	20% PEG 3350	0.2 M Ammonium nitrate pH 4.3	-
29	10% PEG 6000	0.1 M HEPES pH 7.0	-
30	-	0.1 M HEPES pH 7.5	0.8 M Sodium dihydrogen phosphate
31	40% PEG 300	0.1 M Phosphate-citrate pH 5.2	-
32	10% PEG 3000	0.1 M Acetate pH 6.2	0.2 M Zinc acetate
33	20% Ethanol	0.1 M Tris pH 8.5	-
34	25% 1,2-propanediol, 10% Glycerol	0.1 M Sodium phosphate pH 6.8	-
35	10% PEG 20000, 2% Dioxane	0.1 M Bicine pH 9.0	-
36	-	0.1 M Acetate pH 4.6	-
37	10% PEG 1000, 10% PEG 8000	-	-
38	24% PEG 1500, 20% Glycerol	-	-
39	30% PEG 400	0.1 M HEPES pH 7.5	0.2 M Magnesium chloride
40	50% PEG 200	0.1 M Sodium phosphate pH 7.2	0.2 M Sodium chloride
41	30% PEG 8000	0.1 M Acetate pH 4.5	0.2 M Lithium sulfate
42	70% MPD	0.1 M HEPES pH 7.5	0.2 M Magnesium chloride
43	20% PEG 8000	0.1 M Tris pH 8.5	-
44	40% PEG 400	0.1 M Tris pH 8.4	0.2 M Lithium sulfate
45	40% MPD	0.1 M Tris pH 8.0	-
46	25.5% PEG 4000, 15% glycerol	-	0.17 M Ammonium sulfate

47	40% PEG 300	0.1 M Cacodylate pH 7.0	0.2 M Calcium acetate
48	14% 2-propanol, 30% Glycerol	0.07 M Acetate pH 4.6	0.14 M Calcium chloride
49	16% PEG 8000, 20% Glycerol	-	0.04 M Potassium dihydrogen
50	-	0.1 M Cacodylate pH 6.5	0.1 M Sodium citrate
51	2.0 M Ammonium sulfate	0.1 M Cacodylate pH 6.5	0.2 M Sodium chloride,
52	10% 2-propanol	0.1 M HEPES pH 7.5	0.2 M Sodium chloride
53	1.26 M Ammonium sulfate	0.1 M Tris pH 8.5	0.2 M Lithium sulfate
54	40% MPD	0.1 M CAPS pH 10.1	-
55	20% PEG 3000	0.1 M Imidazole pH 8.0	0.2 M Zinc acetate
56	10% 2-propanol	0.1 M Cacodylate pH 6.5	0.2 M Zinc acetate
57	-	0.1 M Acetate pH 4.5	1.0 M di-Ammonium hydrogen
58	-	0.1 M MES pH 6.5	1.6 M Magnesium sulfate
59	10% PEG 6000	0.1 M Bicine pH 9.0	-
60	14.4% PEG 8000, 20% Glycerol	0.08 M Cacodylate pH 6.5	0.16 M Calcium acetate
61	10% PEG 8000	0.1 M Imidazole pH 8.0	-
62	30% Jeffamine M-600	0.1 M MES pH 6.5	0.05 M Cesium chloride
63	-	0.1 M Citric acid pH 5.0	-
64	20% MPD	0.1 M Tris pH 8.0	-
65	20% Jeffamine	0.1 M HEPES pH 6.5	-
66	50% Ethylene glycol	0.1 M Tris pH 8.5	0.2 M Magnesium chloride
67	10% MPD	0.1 M Bicine pH 9.0	-

Clear Strategy Screen (CSS) I:

	Precipitant	Buffer	Salt
1	25% PEG 2000	-	0.3 M Sodium acetate
2	25% PEG 2000	-	0.2 M Lithium sulfate
3	25% PEG 2000	-	0.2 M Magnesium chloride
4	25% PEG 2000	-	0.2 M Potassium bromide
5	25% PEG 2000	-	0.2 M Potasium thiocyanate
6	25% PEG 2000	-	0.8 M Sodium formate
7	15% PEG 4000	-	0.3 M Sodium Acetate
8	15% PEG 4000	-	0.2 M Lithium sulfate
9	15% PEG 4000	-	0.2 M Magnesium chloride
10	15% PEG 4000	-	0.2 M Potassium bromide
11	15% PEG 4000	-	0.2 M Potasium thiocyanate
12	15% PEG 4000	-	0.8 M Sodium formate
13	10% PEG 8000, 10% PEG 1000	-	0.3 M Sodium Acetate
14	10% PEG 8000, 10% PEG 1000	-	0.2 M Lithium sulfate
15	10% PEG 8000, 10% PEG 1000	-	0.2 M Magnesium chloride
16	10% PEG 8000, 10% PEG 1000	-	0.2 M Potassium bromide
17	10% PEG 8000, 10% PEG 1000	-	0.2 M Potasium thiocyanate
18	10% PEG 8000, 10% PEG 1000	-	0.8 M Sodium formate
19	8% PEG 20000, 8% PEG 550	-	0.3 M Sodium Acetate
20	8% PEG 20000, 8% PEG 550	-	0.2 M Lithium sulfate
21	8% PEG 20000, 8% PEG 550	-	0.2 M Magnesium chloride
22	8% PEG 20000, 8% PEG 550	-	0.2 M Potassium bromide
23	8% PEG 20000, 8% PEG 550	-	0.2 M Potasium thiocyanate
24	8% PEG 20000, 8% PEG 550	-	0.8 M Sodium formate

Precipitant Synergy Reagent Formulation:

	Precipitant	pH	Salt	Additive
1	2% PEG 400	pH 5.5	2.0 M Ammonium sulfate	-
2	10% Glycerol	pH 6.5	2.0 M Ammonium sulfate	0.1 M Magnesium sulfate
3	1% MPD	pH 7.5	2.0 M Ammonium sulfate	-
4	5% PEG 400	pH 8.5	2.0 M Ammonium sulfate	0.1 M Magnesium sulfate

5	2% PEG 400	pH 5.5	4.0 M Sodium chloride	0.1 M Magnesium chloride
6	5% MPD	pH 6.5	3.0 M Sodium chloride	0.1 M Calcium chloride
7	5% Isopropanol	pH 7.5	4.0 M Sodium chloride	-
8	5% Isopropanol	pH 5.5	2.5 M Sodium phosphate	-
9	2% PEG 400	pH 6.5	2.0 M Sodium phosphate	-
10	20% Glycerol	pH 7.5	2.5 M Sodium phosphate	-
11	8% MPD	pH 8.5	1.0 M Sodium phosphate	-
12	1% MPD	pH 4.5	2.0 M Ammonium citrate	-
13	5% Isopropanol	pH 6.5	2.0 M Ammonium citrate	-
14	5% PEG 400	pH 7.5	2.0 M Ammonium citrate	-
15	5% Isopropanol	pH 4.5	2.0 M Lithium sulfate	0.1 M Magnesium sulfate
16	5% PEG 400	pH 5.5	2.0 M Lithium sulfate	0.1 M Magnesium sulfate
17	8% MPD	pH 6.5	2.0 M Lithium sulfate	-
18	2% PEG 400	pH 8.5	2.0 M Lithium sulfate	-
19	1% MPD	pH 4.5	1.0 M Lithium sulfate	0.1 M Magnesium sulfate
20	25% MPD	pH 5.5	0.75 M Ammonium citrate	-
21	12% Isopropanol	pH 6.5	1.5 M Ammonium sulphate	-
22	30% Isopropanol	pH 6.5	1.3 M Sodium chloride	0.1 M Calcium chloride
23	10% PEG 400	pH 7.5	4.0 M Sodium chloride	-
24	20% PEG 400	pH 7.5	0.8 M Sodium phosphate	-
25	15% Isopropanol	pH 8.5	1.0 M Ammonium citrate	-
26	15% Isopropanol, 2.5% PEG 3350	pH 8.5	2.0 M Sodium formate	-
27	30% MPD, 25% PEG 1500	pH 4.5	-	-
28	30% MPD, 15% PEG 8000	pH 5.5	-	0.1 M Calcium chloride
29	30% MPD, 10% PEG 3350	pH 6.5	-	0.2 M Ammonium sulfate
30	30% MPD, 4% PEG 1500	pH 7.5	-	-
31	30% MPD, 8% PEG 8000	pH 8.5	-	0.5 M Sodium chloride
32	30% Isopropanol, 4% PEG 3350	pH 4.5	-	0.1 M Calcium chloride
33	30% Isopropanol, 10% PEG 1500	pH 5.5	-	0.2 M Lithium sulfate
34	40% Isopropanol, 15% PEG 8000	pH 6.5	-	-
35	20% Isopropanol, 15% PEG 3350	pH 7.5	-	0.2 M Ammonium citrate
36	30% Isopropanol, 30% PEG 3350	pH 8.5	-	-
37	40% PEG 400, 20% PEG 8000	pH 4.5	-	-
38	40% PEG 400, 5% PEG 3350	pH 5.5	-	-
39	40% PEG 400, 15% PEG 1000	pH 6.5	-	0.15 M Sodium phosphate
40	40% PEG 400, 8% PEG 8000	pH 7.5	-	-
41	25% PEG 400, 20% PEG 3350	pH 8.5	-	0.1 M Magnesium chloride
42	3% MPD, 30% PEG 1500	pH 5.5	-	0.2 M Magnesium sulfate
43	10% Isopropanol, 30% PEG 1500	pH 6.5	-	0.1 M Calcium chloride
44	20% PEG 400, 30% PEG 1500	pH 7.5	-	-
45	8% MPD, 30% PEG 1500	pH 8.5	-	-
46	15% Isopropanol, 25% PEG 3350	pH 4.5	-	0.2 M Ammonium citrate
47	5% PEG 400, 25% PEG 3350	pH 5.5	-	-
48	15% MPD, 25% PEG 3350	pH 6.5	-	0.2 M Lithium sulphate
49	4% Isopropanol, 25% PEG 3350	pH 7.5	-	0.1 M Calcium chloride
50	10% PEG 400, 20% PEG 8000	pH 5.5	-	0.5 M Sodium chloride
51	3% MPD, 20% PEG 8000	pH 6.5	-	-
52	10% Isopropanol, 20% PEG 8000	pH 7.5	-	0.2 M Ammonium sulfate
53	20% PEG 400, 20% PEG 8000	pH 8.5	-	0.1 M Magnesium chloride
54	25% PEG 3350	pH 4.5	-	0.1 M Calcium chloride
55	5% Isopropanol, 7.5% PEG 3350	pH 4.5	-	-
56	1% PEG 4000	pH 5.5	-	-
57	1.5% MPD, 12% PEG 1500	pH 5.5	-	-
58	20% PEG 3350	pH 6.5	-	0.1 M Magnesium chloride
59	4% PEG 8000	pH 6.5	-	-
60	0.5% PEG 4000	pH 7.5	-	-
61	10% PEG 3350	pH 7.5	-	-
62	2% PEG 8000	pH 8.5	-	-
63	5% PEG 4000	pH 8.5	-	-
64	15%PEG 8000	pH 8.5	-	-

REFERENCES

Alli I., Okoniewska, M., Gibbs, B. F., Konishi, Y. (1998) Identification of Peptides in Cheddar Cheese by Electrospray Ionization Mass Spectrometry. *International Dairy Journal.* **8,** 643-649.

Amour A., Bird, M., Chaudry, L., Deadman, J., Hayes, D., Kay, C. (2004) General Considerations for Proteolytic Cascades. *Biochemical Society Transactions.* **32,** 15-16.

Atschul S. F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology.* **215,** 403-410.

Baankreis R., van Schalkwijk, S., Alting, A.C. Exterkate, F.A. (1995) The Occurrence of Two Intracellular Oligoendopeptidases in *Lactococcus lactis* and their Significance for Peptide Conversion in Cheese. *Applied Microbiology and Biotechnology.* **44,** 386–392.

Bacon C. L., Jennings, P.V., Ni Fhaolain, I., O'Cuinn, G. (1994) Purification and Characterisation of an Aminopeptidase A from the Cytoplasm of *Lactococcus lactis* subsp. *cremoris* AM2. *International Dairy Journal.* **4,** 503–519.

Bailey S. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallographica Section D.* **50,** 760-763.

Barrett A. J., Rawlings, N.D. (1992) Enzyme Nomenclature. *Methods in Enzymolgy* **244,** 1-15.

Barrett A. J., Rawlings, N.D., Woessner, J.F. (2004) *Handbook of Proteolytic Enzymes,* Second Edition. Elsevier Academic Press, London.

Berger A., Schechter, I. (1970) Mapping the Active Site of Papain with the aid of Peptide Substrates and Inhibitors. *Philosophical Transactions Of The Royal Society Of London B.* **257,** 249-264.

Berman H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalo, I.N., Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research.* **28,** 235-242.

Beynon R. J., Bond, J.S. (1989) *Proteolytic Enzymes: A Practical Approach.* Information Research, Oxford.

Blundell T. L., Johnson, L.N. (1976) *Protein Crystallography.* Academic Press, London.

Bradford M. (1976) A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry.* **72,** 248-254.

Bragg W. H., Bragg, W.L. (1913) The Reflection of X-rays by Crystals. *Proceedings of the Royal Society of London Series A:Containing Papers of a Mathematical and Physical Character.* **88,** 428-428.

Chapman J. R. (1996) *Protein and Peptide Analysis by Mass Spectrometry.* Humana Press, Totowa.

Charter C. W., Sweet, R.M. (1997) Response Surface Methods for Otimizing and Improving Reproducibility of Crystal Growth. *Methods in Enzymology.* **276,** 74-99.

Cheah E., Carr, P.D., Suffolk, P.M., Vasudevan, S.G., Dixon, N.E., Ollis, D.L. (1994) Structure of the *Escherichia coli* Signal Transducing Protein P-II. *Structure.* **2,** 981-990.

Chenna R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., Thompson, J. D. (2003) Multiple Sequence Alignment with the Clustal Series of Programs. *Nucleic Acids Research.* **31,** 3497-3500.

Christensen J. E., Dudley, E.G., Pederson, J.A., Steele, J.L. (1999) Peptidases and Amino Acid Catabolism in Lactic Acid Bacteria. *Antonie van Leeuwenhoek.* **76,** 217–246.

Christensson C., Bratt, H., Coolbear, T., Holland, R., Lubbers, M.W., O'Toole, P.W., Reid, J.R. (2002) Cloning and Expression of an Oligopeptidase, PepO, with Novel Specificity from *Lactobacillus rhamnosus* HN001 (DR20). *Applied and Environmental Microbiology.* **68,** 254-262.

Courtin P., Nardi, M., Wegmann, U., Joutsjoki, V., Ogier, J.C., Gripon, J.C, Palva, A., Henrich, B., Monnet, V. (2002) Accelerating Cheese Proteolysis by Enriching *Lactococcus lactis* Proteolytic System with *Lactobacilli* Peptidases. *International Dairy Journal.* **12,** 447-454.

Crick F. H. C., Magdoff, B.S. (1956) The Theory of the Method of Isomorphous Replacement for Protien Crystals. *Acta Crystallographica.* **9**, 901-908.

Crowther R. A., Blow, D.M. (1967) A Method of Positioning a Known Molecule in an Unknown Crystal Structure. *Acta Crystallographica.* **23**, 544-548.

Dauter Z. (1999) Data-Collection Strategies. *Acta Crystallographica Section D.* **D55,** 1703-1717.

Davis B. J. (1964) Disc Electrophoresis-II: Method and Application to Human Serum Proteins. *Annals of the New York Academy of Sciences.* **121,** 404-427.

Drenth J. (1999) *Principles of Protein X:ray Crystallography.* Second Edition. Springer-Verlag, New York.

Ducruix A., Giege, R. (1992) *Crystallization of Nucleic Acids and Proteins: A Practical Approach.* Oxford University Press, New York.

Edman P. (1950) Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica.* **4,** 277-282.

Ewald P. P. (1921) The Calculation of Optical and Electrostatic Grid Potential. *Zeitschrift Fur Kristallographie.* **56,** 129-156.

Exterkate F. A., Alting, A.C. (1999) Role of Calcium in Activity and Stability of the *Lactococcus lactis* Cell Envelope Proteinase. *Applied and Environmental Microbiology.* **65,** 1390-1396.

Fisher R. G., Woods, N.E., Fuchs, H.E., Sweet, R.M. (1980) 3-Dimensional Structures of C-Phycocyanin and B-Phycoerythrin at 5.0 A Resolution. *Journal of Biological Chemistry.* **255,** 5082-5089.

Foucaud C., Kunji, E.R.S., Hagting, A., Richard, J., Konings, W.N., Desmazeaud, M., Poolman, B. (1995) Specificity of Peptide Transport Systems in Lactococcus lactis: Evidence for a Third SYstem Which Transports Hydrophobic Di- and Tripeptides. *Journal of Bacteriology.* **177,** 4652-4657.

Fox P. F., Guinee, T.P., Cogan, T.M., McSweeney, P.L.H. (2000) *Fundamentals of Cheese Scinece.* Aspen Publishers, Gaithersburg.

French G. S., Wilson, K.S. (1978) On the Treatment of Negative Intensity Observations. *Acta Crystallographica Section A.* **34,** 517-525.

Gasteiger E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. (2003) ExPASy: the Proteomics Server for In-Depth Protein Knowledge and Analysis. *Nucleic Acids Research.* **31,** 3784-3788.

Giacovazzo C., Monaco, H. L., Artioli, G., Viterbo, D., Ferraris, G., Gilli, G., Anotti, G., Catti, M. (2002) *Fundamentals of Crystallography,* Second Edition. Oxford University Press, Oxford.

Gilliland G. L., Ladner, J.E. (1996) Crystallization of Biological Macromolecules for X-ray Diffraction Studies. *Current Opinion in Structural Biology.* **6,** 595-603.

Hartley B. S. (1960) Proteolytic Enzymes. *Annual Review of Biochemistry.* **29,** 45-72.

Howard G. C., Brown, H.W. (2002) *Modern Protein Chemistry: Practical Aspects.* CRC Press, Boca Raton.

Howell P. L., Smith, G.D. (1992) Identification of Heavy-Atom Derivatives by Normal Probability Methods. *Journal of Applied Crystallography.* **25,** 81-86.

Howells E. R., Phillips, D.C., Rogers, D. (1950) The Probability Distribution of X-ray Intensities: 2. Experimental Investigation and the X-ray Detection of Centres of Symmetry. *Acta Crystallographica.* **3,** 210-214.

Jameson G. B., Anderson, B. F., Breyer, W. A., Day, C. L., Tweedie, J. W., Baker, E. N. (2002) Structure of a Domain-opended mutant (R121D) of the Human Lactoferrin N-lobe Refined from a Merohedrally Twinned Crystal Form. *Acta Crystallographica Section D.* **58,** 955-962.

Jancarik J., Kim, S.H. (1991) Sparse Matrix Sampling: A Screening Method for Crystallization of Proteins. *Journal of Applied Crystallography.* **24,** 409-411.

Jimenez C. R., Huang, L., Qiu, Y., Burlingame, A.L. (1998) *Sample preparation for MALDI mass analysis of peptides and proteins., in Current Protocols in Protein Science.* Wiley-Interscience, New York.

Jongeneel C. V. (1989) A Unique Signature Identifies A Family Of Zinc-Dependent Metallopeptidases. *Federation of European Biochemical Societies Letters.* **242,** 211-214.

Joshua-Tor L., Xu, E.H., Johnston, S.A., Rees, D.C. (1995) Crystal Structure of a Conserved Protease that Binds DNA: The Bleomycin Hydrolase, Gal6. *Science.* **269,** 945-950.

Kinter M., Sherman, N.E. (2000) *Protein Sequencing and Identification using Tandem Mass Spectrometry.* Wiley-Interscience, New York.

Kuipers O. P., de Ruyter, P.G.G.A., Kleerebezem, M., de Vos, W.M. (1997) Controlled Overproduction of Proteins by Lactic Acid Bacteria. *Trends in Biotechnology.* **15,** 135-140.

Kunji E. R. S., Mierau, I., Haging, A., Poolman, B, Konings, W.N. (1996) The Proteolytic Systems of Lactic Acid Bacteria. *Antonie van Leeuwenhoek.* **70,** 187-221.

LaVallie E. R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F., McCoy, J.M. (1993) A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytolasm. *BioTechnology.* **11,** 187-193.

Law J., Haandrikman, A. (1997) Proteolytic Enzymes of Lactic Acid Bacteria. *International Dairy Journal.* **7,** 1-11.

Lemieux L., Simard, R.E. (1992) Bitter Flavour in Dairy Products. II. A Review of Bitter Peptides from Caseins: their Formation, Isolation, and Identification, Structure Masking and Inhibition. *Lait.* **72,** 335–382.

Lowe J., Stock, D., Jap, B., Zwickle, P., Baumeister, W., Huber, R. (1995) Crystal Structure of the 20S Proteosome from the Archaeon *T. acidophilum* at 3.4 A resolution. *Science.* **268,** 533-539.

Martin I. M. B. (1996) *Scientific Data Visualisation and Digital Image Processing for Structural Biology*. Purdue University, West Lafayette.

Mata L., Gripon, J., Mistou, M. (1999) Deletion of the Four C-terminal Residues of PepC Converts an Aminopeptidase into an Oligpeptidase. *Protein Engineering.* **12,** 681-686.

Matthews B. W. (1968) Solvent Content of Protein Crystals. *Journal of Molecular Biology.* **33,** 491-497.

McGarry A., Law, J., Coffey, A., Daly, C., Fox, P.F., Fitzgerald, G.F. (1994) Effect of Genetically Modifying the *Lactococcal* Proteolytic System on Ripening and Flavor Development in Cheddar Cheese. *Applied and Environmental Microbiology.* **60,** 4226-4233.

McPherson A. (1990) Current Approaches to Macromolecular Crystallization. *European Journal of Biochemistry.* **189,** 1-23.

McSweeney P. L. H., Olson, N.F., Fox, P.F., Healy, A., Hojrup, P. (1993) Proteolytic Specificity of Chymosin on Bovine α_{s1}-casein. *Journal of Dairy Research.* **60,** 401-412.

Meyer J., Spahni, A. (1998) Influence of X-prolyl-dipeptidylaminopeptidase of *Lactobacillus debrueckii* subsp. *lactis* on Proteolysis and Taste of Swiss Gruyere Cheese. *Milchwissenschaft.* **53,** 449–453.

Mierau I., Tan, P.S.T., Haandrikman, A.J., Kok, J., Leenhouts, K.J., Konings, W.N., Venema, G. (1993) Cloning and Sequencing of the Gene for a *Lactococcal* Endopeptidase, an Enzyme with Sequence Similarity to Mammalian Enkephalinase. *Journal of Bacteriology.* **175,** 2087-2096.

Mierau I., Kunji, E.R.S., Leenhouts, K.J., Hellendoorn, M.A., Haandrikman, A.J., Poolman, B., Konings, W.N., Venema, G., Kok, J. (1996) Multiple-Peptidase Mutants of *Lactococcus lactis* are Severely Impaired in their Ability to Grow in Milk. *Journal of Bacteriology.* **178,** 2794-2803.

Nar H., Huber, R., Meining, W., Schmid, C., Weinkauf, S., Bacher, A. (1995) Atomic Structure of GTP Cyclohydrolase I. *Structure.* **3,** 459-466.

Navaza J. (2001) Implementation of Molecular Replacement in *AMoRe*. *Acta Crystallographica Section D.* **57,** 1367-1372.

NC-IUBMB (1992) *Enzyme Nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes.* Academic Press, Orlando.

Neurath H. (1999) Proteolytic Enzymes, Past and Future. *Proceedings of the National Academy of Science.* **96,** 10962-10963.

Niven G. W. (1991) Purification and Characterization of Aminopeptidase A from *Lactococcus lactis* subsp. *lactis* NCDO-712. *Journal in Genetics and Microbiology.* **137,** 1207-1212.

Oefner C., D'Arcy, A., Hennig, M., Winkler, F.D., Dale, G.E. (2000) Structure of Human Neutral Endopeptidase (Neprilysin) Complexed with Phosphoramidon. *Journal of Molecular Biology.* **296,** 341-349.

Oefner C., Roques, B.P., Fournie-Zaluski, M., Dale, G.E. (2004) Structural Analysis of Neprilysin with Carious Specific and Potent Inhibitors. *Acta Crystallographica Section D.* **60,** 392-396.

O'Farrell P. A., Gonzalez, F., Zheng, W., Johnston, S.A., Joshua-Tor, L. (1999) Crystal Structure of Human Bleomycin Hydrolase, a Self-Compartmentalizing Cysteine Protease. *Structure.* **7,** 619-627.

Patterson A. L. (1934) A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. *Physical Review.* **46,** 372-376.

Perkins D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. (1999) Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis.* **20,** 3551-3567.

Petsko G. A. (1985) Preparation of Isomorphous Heavy-atom Derivatives. *Methods in Enzymology.* **114,** 147-156.

Pietzsch J. (2003) *Signalling Scissors: New Perspectives on Proteases in: The 3rd Horizon Symposia.* Nature Publishing Group, Brescia.

Potterton E., Briggs, P., Turkenburg, M., Dodson, E. (2003) A Graphical User Interface to the CCP4 Program Suite. *Acta Crystallographica Section D.* **59,** 1131-1137.

Pritchard G. G., Freebairn, A.D., Coolbear, T. (1994) Purification and Characterisation of an Endopeptidase from *Lactococcus lactis* susp. *cremoris* SK11. *Microbiology.* **140,** 923-930.

Rawlings N. D., Barrett, A.J. (1995) Evolutionary Families of Metallopeptidases. *Methods in Enzymology.* **248,** 183-228.

Rawlings N. D., Tolle, D.P., Barrett, A.J. (2004) MEROPS: the Peptidase Database. *Nucleic Acids Research.* **32,** 160-164.

Reid J. R., Huat Ng, K., Moore, C.H., Coolbear, T., Pritchard, G.G. (1991a) Comparison of bovine β-casein hydrolysis by PI and PIII-type Proteinases from *Lactobacillus lactis* supsp. *cremoris. Applied Microbiology and Biotechnology.* **36,** 344-351.

Reid J. R., Moore, C.H., Midwinter, G.G., Pritchard, G.G. (1991b) Action of a Cell Wall Proteinase from *Lactococcus lactis* subsp. *cremoris* SK11 on bovine α_{s1}-casein. *Applied Microbiology and Biotechnology.* **35,** 222-227.

Reid J. R., Coolbear, T., Pillidge, C.J., Pritchard, G.G. (1994a) Specificity of Hydrolysis of Bovine κ-Casein by Cell Envelope-Associated Proteinases from *Lactococcus lactis* Strains. *Applied and Environmental Microbiology.* **60,** 801-806.

Reid J. R., Huat Ng, K., Moore, C.H., Coolbear, T., Pritchard, G.G. (1994b) *A Study of Bovine Casein Hydrolysis by the Lactococcal Cell Envelope:Associated Proteinase.* Massey University, Palmerston North, New Zealand.

Rhodes G. (1993) *Crystallography Made Crystal Clear: A Guide For Users Of Macromolecular Models.* Academic Press, California.

Rigolet P., Mechin, I., Delage, M. M., Chich, J. F. (2002) The Structural Basis for Catalysis and Specificity of the X-Prolyl Dipepdidyl Aminopeptidase from *Lactococcus Lactis. Structure.* **10,** 1383-1394.

Rosenthal A., Charnock-Jones, D.S. (1992) New Protocols for DNA Sequencing with Dye Terminators. *DNA Sequence.* **3,** 61-64.

Rossmann M. G., Arnold, E. (2001) *International Table for Crystallography.*, Volume F. Kluwer Academic, London.

Roussel A., Cambillau, C. (1991) *Turbo-Frodo, Silicon Graphics Geometry Partners Directory.* Silicon Graphics, Mountain View, CA.

Saklatvala J., Nagase, H., Salvesen, G. (2003) *Proteases and the Regulation of Biological Processes.* Protland Press., London.

Salvesen G. S. (2003) *Summary: New Perspectives on Proteases, in The 3rd Horizon Symposia.* Nature Publishing Group, Brescia.

Sayle R., Milner-White, E.J. (1995) RasMol: Biomolecular Graphics For All. *Trends in Biochemical Sciences.* **20,** 374.
Schwede T., Kopp, J., Guex, N., Peitsch, M.C. (2003) SWISS-MODEL: An Automated Protein Homology-Modelling Server. *Nucleic Acids Research.* **31,** 3381-3385.

Sousa M. J., Ardo, Y., McSweeney, P.L.H. (2001) Advances in the Study of Proteolysis During Cheese Ripening. *International Dairy Journal.* **11,** 327-345.

Stepaniak L., Fox, P.F. (1995) Characterization of the Principal Intracellular Endopeptidase from *Lactococcus lactis* subsp. *lactis* MG1363. *International Dairy Journal.* **5,** 699–713.

Stepaniak L. (2004) Dairy Enzymology. *International Journal of Dairy Technology.* **57,** 153-171.

Tan P. S. T., Pos, K.M., Konings, W.N. (1991) Purification and Characterization of an Endopeptidase from *Lactococcus lactis* subsp. *cremoris* Wg2. *Applied and Environmental Microbiology.* **57,** 3593–3599.

Terwissha van Scheltinga A. C., Valegard, K., Ramaswamy, S., Hajdu, J., Andersson, I. (2001) Multiple Isomorphous Replacement on Merohedral Twins: Structure Determination of Deacetoxycephalosporin C Synthase. *Acta Crystallographica Section D.* **57,** 1776-1785.

Terwissha van Scheltinga A. C., Valegard, K., Hajdu, J., Andersson, I. (2003) MIR Phasing using Merohedrally Twinned Crystals. *Acta Crystallographica Section D.* **59,** 2017-2022.

Visser S., Slangen, K.J. (1977) On the Specificity of Chymosin (Rennin) in its Action on Bovine β-casein. *Netherlands Milk Dairy Journal.* **31,** 16-30.

Wiencek J. M. (1999) New Strategies for Protein Crystal Growth. *Annual Review Biomedical Engineering.* **1,** 505-534.

Xu H. E., Johnston, S.A. (1994) Yeast Bleomycin Hydrolase is a DNA-Binding Cysteine Protease. *Journal of Biological Chemistry.* **269,** 21177-21183.

Yeates T. O., Rees, D.C. (1987) An Isomorphous Replacement Method for Phasing Twinned Structures. *Acta Crystallographica Section A.* **43,** 30-36.

Yeates T. O. (1997) Detecting and Overcoming Crystal Twinning. *Methods in Enzymology.* **276,** 344-358.

Yvon M., Rijnen, L. (2001) Cheese Flavour Formation by Amino Acid Catabolism. *International Dairy Journal.* **11,** 185-201.

Zhendong P., Sebti, M. (1996) Cys^{102} and His^{398} Are Required for Bleomycin-Inactivating Activity but Not for Hexamer Formation of Yeast Bleomycin Hydrolase. *Biochemistry.* **35,** 10751-10756.

Zhu X., Papayannopoulos, I.A. (2003) Improvement in the Detection of Low Concentration Protein Digests on a MALDI TOF/TOF Workstation by Reducing a-Cyano-4-hydroxycinnamic Acid Adduct Ions. *Journal of Biomolecular Techniques.* **14,** 298-307.