CLONING AND EXPRESSION OF A MODIFIED ORYZACYSTATIN INHIBITOR GENE AND AN INVESTIGATION OF ITS INHIBITORY CAPABILITIES

CAROLINE JOANNE HAWORTH

Thesis presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Biochemistry
UNIVERSITY OF CAPE TOWN

MAY 1997

To my parents, Ron and the late Audrey Haworth

Acknowledgements

I am greatly indebted to my supervisor, Professor Jennifer Thomson and to my co-supervisor Dr. Dennis Maeder for their endless support, guidance and encouragement during this project.

A special thanks goes to Professor Eugene Dowdle for all his helpful discussions on enzyme kinetics and yeast expression and for proofreading this manuscript.

I would also like to say a special thanks to Professor Soichi Arai at the Department of Agricultural Chemistry, University of Tokyo, Japan, for supplying the antibodies to oryzacystatin. Also to Dr. Lisa Jouanin at the Department of Cellular Biology, National Institute for Agricultural Research, France, for supplying the OC I inhibitor in the pET plasmid.

Thanks also goes to Di James for her helpful advice on DNA sequencing, Associate Professor Wolf Brandt for his assistance with the protein N-terminal sequencing and Liesel Stevens for technical assistance with the yeast work. A special thanks goes to Nikki Campbell for her assistance with the design and compilation of the computer graphics for this thesis.

To my fiancé, Daryl, words are not enough to thank you for all your support, encouragement and patience. I could not have done it without you.

To my family, a very big thank you for all your steadfast support and encouragement.

Finally, thanks to my colleagues in the departments of Biochemistry and Microbiology for their support and invaluable assistance.

Financial support is acknowledged and greatly appreciated from AECI, the Foundation for Research and Development (FRD) and the Departments of Microbiology and Biochemistry.

ABSTRACT

Cysteine proteinase inhibitors have shown potential as biocontrol agents for the protection of plants against insect and pathogen attack. With the advent of protein and genetic engineering such inhibitors can now be modified in order to improve their effectiveness. Because cystatins have already been isolated from plants, they provide a good starting point for developing modifications which may improve their function as biocontrol agents. The purpose of this project, therefore, was to design a potentially improved analogue of the rice cysteine proteinase inhibitor, oryzacystatin I, through molecular modelling studies. The gene sequence for this modified protein was then synthesised and expressed for kinetic analysis and insect trial assays.

A prediction of the oryzacystatin I (OC I) tertiary structure was made using Biograf software on an Evans and Sutherland workstation. This structure was based on the known structures of stefin B and chicken cystatin whose co-ordinates are published in the Brookhaven data files. Chicken cystatin is one of the most potent inhibitors of papain in the cystatin superfamily. This is believed to be due, in part, to an increased binding of the cystatin to papain through its amino-terminal region with the residues Leu^{7} to Gly^{9} playing a particularly important role.

In order to determine whether the chicken cystatin N-terminal sequence could enhance the binding of OC I to papain, the amino-terminus of chicken cystatin was substituted for the OC I N-terminus. Such manipulations of the protein structure, it was hoped, would give some insight into the interactions of cystatins in the amino-terminus. The rest of the OC I sequence remained unmodified and three models suggesting possible structures for this 'hybrid' inhibitor were designed.

For synthesis of the gene, the amino acid sequence of this 'hybrid' inhibitor was back translated to a 282-bp gene sequence using *E. coli* codon usage. Since this gene was too long to be synthesised in a single step on an automatic DNA synthesiser, the gene sequence was scanned for unique restriction sites or these were generated by modifications of the codon usage. This enabled division of the gene into smaller sections. Cloning of this gene proved problematic as a result of errors in the sequence that arose as a result of inefficient DNA synthesis. A number of methods were therefore explored to purify the full length DNA sequences from the truncated sequences. These methods included HPLC purification and the use of biotin-labelled primers for purification using streptavidin Dynabeads. These techniques, as well as the use of larger controlled pore glass columns for synthesis and a division of the gene into four parts, resulted in the successful synthesis of this gene which was then cloned into pUC18.

Protein expression of the hybrid oryzacystatin (HO) gene was achieved using the pMAL vector system. In these vectors, the gene is cloned 3' to the sequence which codes for the *E. coli* maltose binding protein (MBP). This gives rise to a fusion composed of the MBP and the gene of interest. In this case the MBP and the HO

TABLE OF CONTENTS

ABBREVIATIONS USED IN TEXT AND FIGURES

GENERAL

A	Adenine
AMC	aminomethyl coumarin
amp	ampicillin
bp	base pair(s)
BCIP	5-bromo-4-chloro-3-indoylphosphate
°C	degrees centigrade
DIG	digoxygenin
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
E-64	L-trans-epoxysuccinyl-leucylamido(4-guanidino)butane
EDTA	ethylenediaminetetra-acetic acid (disodium salt)
g	Gram
G	guanine / guanosine
GCG	Genetics Computer Group
HO	Hybrid Oryzacystatin
HPLC	High performance liquid chromatography
IPTG	isopropylthio-β-D-thiogalactopyranoside
Kb	kilobase
kDa	kilodalton
K_i	equilibrium dissociation constant
K_m	Michaelis-Menton constant
LB	Luria-Bertani broth

M	molar
MBP	maltose binding protein
ml	millilitre
MD	minimal dextrose medium
MM	minimal methanol medium
M_r	relative molecular weight
NBT	p-nitro blue tetrazolium chloride
ND	not determined
nm	nanometer
N-terminal	amino terminal
OC I	Oryzacystatin I
OC II	Oryzacystatin II
OD	Optical density
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
[S]	substrate concentration
SBTI	soybean trypsin inhibitor
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	thymine / thymidine
TE	Tris/EDTA buffer
TEMED	N,N,N',N'-tetramethyl ethylene diamine
Tris	tris(hydroxymethyl)aminoethane
TAE	Tris/Acetate/EDTA buffer
TBE	Tris/Borate/EDTA buffer
U	uracil
V_i	velocity in the presene of inhibitor
V_o	initial velovity
V_{max}	maximum velocity

w/v	weight per volume (in g per 100 ml)
YNB	yeast nitrogen base
YPD	Yeast peptone dextrose
X-gal	5-Bromo-4-chloro-3-indolyl-B-galactoside
Z-Phe-Arg-AMC	benzylcarbonylphenylalanyl-arginyl-aminomethylcoumarin

ONE AND THREE LETTER CODES FOR AMINO ACIDS

One Letter Code	Three Letter Code	Full name	One Letter Code	Three Letter Code	Full name
A	Ala	Alanine	M	Met	Methionine
D	Asp	Aspartic acid	N	Asn	Asparagine
C	Cys	Cysteine	P	Pro	Proline
E	Glu	Glutamic acid	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arganine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

CHAPTER 1

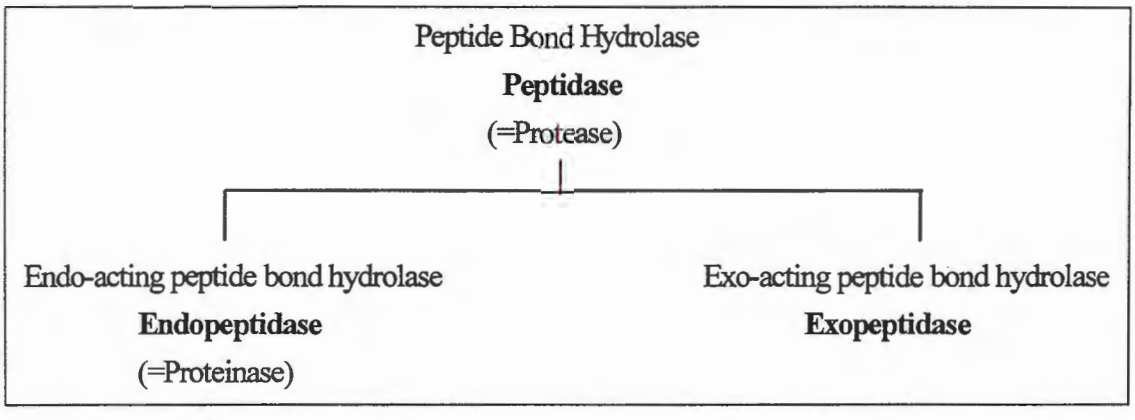
Literature Review: Cysteine Proteinases and their Inhibitors

1.1 Introduction

Proteases and their inhibitors are a widespread group of proteins which occur naturally in many different physiological systems. These include digestion, blood clotting, fertilisation and seed germination. These proteins have also been implicated in the processing and inactivation of peptide hormones, in tumour metastasis and in cellular invasion (Polgár *et al.*, 1986).

Proteases or proteinases act by cleaving peptide bonds. Because of the large number of enzymes that function in this way, there has often been some confusion in the terms used to describe them. Barrett & McDonald (1986) have provided the following clarifying definition:

> The term protease is used to describe a peptide bond hydrolase that may be either an endopeptidase or an exopeptidase. The term 'proteinase,' on the other hand, refers exclusively to the endopeptidases. The relationship between these terms is illustrated below:



(from Barrett & McDonald, 1986)

This review will focus on the endopeptidases since firstly, the large majority of proteases are endo-acting peptidases and secondly this project deals with a cysteine endopeptidase inhibitor from rice seeds, namely oryzacystatin I. Hence the term 'proteinase' will be used. This review will also discuss very briefly the various characteristics of the different classes of proteinases. Emphasis will be placed on the cysteine proteinases and their inhibitors since the oryzacystatin inhibitor used in this study belongs to this class. Furthermore, with the advent of protein engineering, it is possible to modify the structures of protein inhibitors to improve their inhibitory capabilities against target enzymes. This review will thus discuss the potential of cysteine proteinase inhibitors (cystatins) to act as agents against viruses and human parasites, and to serve as agricultural biocontrol agents against insect pests.

1.2 Classification of Proteinases into their Different Mechanistic Classes

Proteinases can be classified according to:

(1) their catalytic mechanisms

(2) the pH range over which they are active (i.e. acid, neutral or alkaline)

(3) their ability to hydrolyse specific proteins (keratinase, elastase, collagenase etc.)

(4) their sensitivity to specific inhibitors (North, 1982; Barrett, 1986).

Based on these criteria, several mechanistic classes of proteinases have been recognised. These include the serine, cysteine, aspartic and metallo-proteinases whose different characteristics and properties are shown in Table 1.1.

Table 1.1: The four classes of proteinases and their different properties (adapted from North, 1982 and Barrett *et al.*, 1986)

Proteinase	Example	Active Site Residue	pH Range	Inhibitors
Serine (EC 3.4.21)	Chymotrypsin Trypsin Elastase	Serine Histidine	7.0 - 10.0	Bowman-Birk Kunitz Lima bean
Cysteine (EC 3.4.22)	Papain Cathepsins L, B, H, and S Calpain	Cysteine Histidine	5.0 - 7.0	Cystatins E-64[a] Iodoacetate
Aspartic (EC 3.4.23)	Cathepsin D Penicillopepsin	Two Aspartic Acid Residues	Below 4.5	Pepstatin
Metallo (EC 3.4.24)	Thermolysin Collagenases	Metal, usually Zinc	7.0	EDTA[b]

[a]L-*trans*-epoxysuccinyl-leucylamido(4-guanidino)butane
[b]ethylenediaminetetra-acetic acid

This table shows that each class of proteinase has distinct characteristics and is inhibited by a number of class specific inhibitors. In general proteinase inhibitors act by binding to the active site of the proteinase to inhibit proteolytic activity. The inhibitors achieve this by preventing access of substrates to the proteinase catalytic site through steric hindrance. Only a small portion of the inhibitor's structure, the 'inhibitory' site, is involved in binding directly to the proteinase (Creighton & Darby, 1989).

Some proteinase inhibitors, however, are able to act on more than one type of proteinase, i.e. they have 'dual' functionality. For example, the microbial inhibitors, leupeptin and antipain, inhibit some serine proteinases as

well as many cysteine proteinases (Bode & Huber, 1992a). The *Streptomyces* subtilisin inhibitor has also been shown to inhibit not only subtilisin, but also a metallo endopeptidase (Kajiwara *et al.*, 1991). Similarly, an inhibitor from potato was shown to inhibit the aspartyl proteinase, cathepsin D, as well as the serine proteinase, trypsin (Bode & Huber, 1992a).

Thus it appears that proteinases (and their inhibitors) are an extremely diverse group, each with a number of different characteristics. Since the cysteine proteinases are of particular relevance to this study they will be discussed more fully in the following section.

1.3 The Cysteine Proteinases

The cysteine proteinases (as shown in Table 1.1) generally show maximal activity, on either synthetic substrates or proteins, at slightly acidic or neutral pH (Barrett, 1986) and all require activation by thiol reagents (North, 1982). The catalytic mechanisms which involve the active site residues Cys^{25} and His^{159} (papain numbering), will be discussed in more detail in Chapter 2. This section will deal mainly with the types of cysteine proteinases, their occurrence and their possible functions.

Cysteine proteinases are a diverse group of proteins and are found in most living organisms (Barrett, 1986). Most cysteine proteinases belong to the large papain superfamily. This includes papain, (isolated from the latex of *Carica papaya,* Drenth *et al.*, 1971), many cysteine proteinases, with properties similar to papain, from the protozoa and higher plants (North, 1982), the lysosomal cathepsins, B, H, L and S, and the calcium-dependent cysteine proteinases, the calpains. These latter enzymes are generally found in solution in the cytoplasmic fraction of the cells and require Ca^{2+} for activation (Barrett, 1986).

These above-mentioned proteinases are considered to fall into a single evolutionary superfamily (Barrett, 1986) while bacterial and viral cysteine proteinases have evolved independently. Two examples of bacterial cysteine proteinases are the proteinase from *Streptococcus* species (Tai *et al.*, 1976) and clostripain from *Clostridium histolyticum* (Siffert, 1976). Viruses known to contain cysteine proteinases include the polio virus (Argos *et al.*, 1984) and many plant viruses such as the plum pox potyvirus (Garcia *et al.*, 1993). These proteinases are responsible for cleaving viral polyprotein precursors into their functional units (Argos *et al.*, 1984; Garcia *et al.*, 1993).

The two largest groups of cysteine proteinases i.e. those found in plants and mammals, are considered in more detail in the following sub-sections.

1.3.1 Plant Cysteine Proteinases of the Papain Superfamily

All four mechanistic classes of proteinases have been found to occur in plants, particularly in the seeds and tubers (Ryan, 1973). However, since this review focuses on the cysteine proteinases, only the properties of this class of enzymes in plants will be discussed. Furthermore, as this project deals with a cysteine proteinase inhibitor isolated from rice plants (oryzacystatin I), it seems pertinent to discuss the possible roles that cysteine proteinases might play in plants.

Examples of some of the cysteine proteinases found in plants include papain, aleurain, ficin, actinidin, bromelains and oryzain (Koizumi *et al.*, 1993). It is known that during germination there is a marked specific increase in the activity of cysteine proteinases. These function to digest the reserve protein and to provide amino acids for the developing seedlings (Cervantes *et al.*, 1994; Watanabe *et al.*, 1991). Cysteine proteinases may also be involved in the degradation of proteins stored in vacuoles, thus releasing products for the synthesis of new proteins during stress or for osmotic adjustment (Guerrero *et al.*, 1990).

Plant cysteine proteinases are often induced in plants in response to hormones. The transcription of barley aleurain and rice oryzains, for example, is activated by gibberellic acid and repressed by abscisic acid (Rogers *et al.*, 1985; Watanabe *et al.*, 1991). Furthermore, the production of ethylene by plants has also been shown to regulate the expression of cysteine proteinase genes during the germination of chickpea (Cervantes *et al.*, 1994).

Environmental stresses such as cold or water deficiency have also been shown to affect the regulation of cysteine proteinase genes (Bond & Butler, 1987). The functions of proteinases induced by these adverse conditions have not, however, been fully elucidated. During cold or desiccation stress, conformational changes in the proteins may occur and the breakdown of cellular structures and leakage of solutes and electrolytes may lead to the denaturation of polypeptides (Guerrero *et al.*, 1990). Researchers have postulated, therefore, that cysteine proteinases may be involved in the degradation of proteins that have been denatured by dehydration or cold stress. Thus, activation of cysteine proteinases may influence biochemical pathways by increasing the rates of protein turnover.

The cysteine proteinases therefore appear to play a physiological role in plants. One might assume that this could have important implications for biocontrol strategies that involve cloning of cysteine proteinase inhibitors into plants. However, transgenic plants producing high levels of cysteine proteinase inhibitors have been shown to be normal in appearance, self-fertile, and able to produce viable seeds (Masoud *et al.*, 1993).

1.3.2 The Mammalian Cysteine Proteinases, the Cathepsins and Calpains

The cathepsins are found mainly in the lysosomes in the cells of higher animals. The name 'cathepsin' is derived from the Greek word meaning, 'to digest' (Bond & Butler, 1987), and these proteinases are believed to be responsible for much of the bulk turnover of proteins in the cell. They are essentially small proteins, which differ from one another in pH optima and substrate specificity. Some of them are true endopeptidases (such as cathepsins L and S; Bond & Butler, 1987), while others exhibit aminopeptidase (cathepsin H; Jerala *et al.*, 1994) or carboxypeptidase (cathepsin B) activity (Musil *et al.*, 1991). Besides the endopeptidases, lysosomes also contain the cysteine exopeptidases such as dipeptidyl peptidase I (cathepsin C; Barrett, 1987). These proteinases complete the degradation of proteins by their rapid action on oligopeptides. A brief summary of the lysosomal proteinases and their properties is given in Table 1.2.

The cathepsins have also become important in the development and progression of a variety of human diseases (Katunuma & Kominami, 1986). For example, the two most commonly studied lysosomal cysteine proteinases, cathepsins L and B, have been shown to degrade components of the extracellular matrix in diseased states such as muscular dystrophy (Moreau *et al.*, 1990) and osteoporosis (Abrahamson *et al.*, 1987). Furthermore, malignant tissues have also been shown to secrete a 'cathepsin B-like' cysteine proteinase that is thought to be involved in the destruction of the extracellular matrix as well as facilitating the detachment of metastatic cells from primary tumour masses (Katunuma & Kominami, 1986).

Table 1.2: The properties of lysosomal cysteine proteinases.

Examples	Activity	Activity Requirements	Location	pH- Optima	Inhibitors	References
Cathepsin H (EC 3.4.22.16)	Endo-Peptidase	SH-reducing agent	Lysosomes	5.0	Alkylating reagents, Cystatin inhibitors	Bond & Butler, 1987
Cathepsin B (EC 3.4.22.1)	Endo-Peptidase	SH-reducing agent	Lysosomes	5.0	Alkylating reagents, Cystatin inhibitors	Bond & Butler, 1987
Cathepsin L (EC 3.4.22.15)	Endo-Peptidase	SH-reducing agent	Lysosomes	5.0	Alkylating reagents, Cystatin inhibitors	Bond & Butler, 1987
Cathepsin S (EC 3.4.22.27)	Endo-Peptidase	SH-reducing agent	Lysosomes	3.5	Alkylating reagents, Cystatin inhibitors	Isemura *et al.*, 1991
Cathepsin C (EC 3.4.14.1)	Exo-Peptidase	SH-reducing agent	Lysosomes	7.0	Alkylating reagents, Cystatin inhibitors	Barrett, 1987
Calpains (EC 3.4.22.17)	Endo-peptidase	Ca^{2+}, SH-reducing agent	Cytosol	7.0 - 8.5	Calpastatin, Kininogens, E-64	Bond & Butler, 1987; Murachi, 1983

Other mammalian cysteine proteinases include the calpains. These are a group of cysteine endopeptidases which require calcium ions for activity (Murachi, 1983). They are heterodimers of 110 kDa and are composed of two subunits of 80 and 30 kDa (Bond & Butler, 1987). The amino acid sequence around the active-site

cysteine residue shows 33% homology with papain. These inhibitors have been identified in most tissues and all have similar properties. Some of the properties of the calpains are shown in Table 1.2.

1.4 The Cystatin Superfamily

Cysteine proteinase inhibitors or cystatins, are very tight-binding, non-covalent inhibitors of the cysteine proteinases (Turk & Bode, 1991). Crystal and solution structures of native and recombinant forms of cystatins have been elucidated (Bode *et al.*, 1988; Stubbs *et al.*, 1990; Diekmann *et al.*, 1993; Engh *et al.*, 1993; Martin *et al.*, 1995; Tate *et al.*, 1995). From these models and from primary structure sequence homology, it is apparent that the cystatins have a very similar basic structure. Interaction of the enzyme and the inhibitor occurs mainly through hydrophobic interactions between the inhibitor and the complementary active-site cleft of the enzyme. The inhibitor forms three distinct binding loops which form the contact regions with the enzyme. These are the amino terminus, a first hairpin loop and a second hairpin loop. Machleidt *et al.* (1995) have described a so-called 'elephant-trunk model' (Figure 1.1) which shows very clearly the three contact areas.

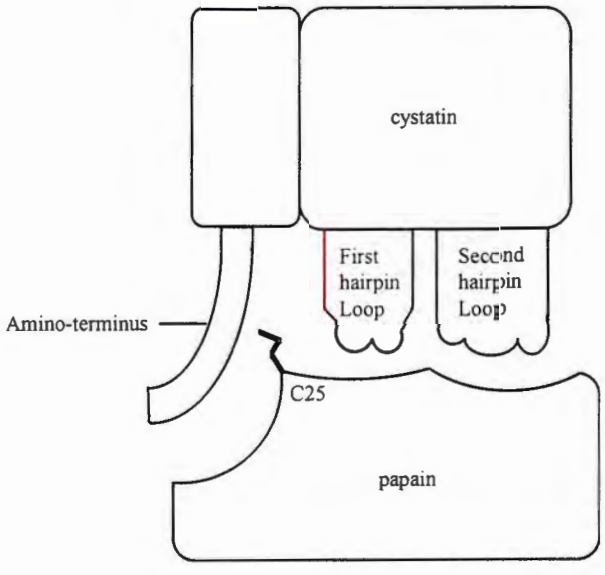


Figure 1.1: *The 'elephant-trunk model' which depicts the three binding areas of contact of the inhibitor with the cysteine proteinase (from Machleidt et al., 1995).*

The significance of these three binding areas, and especially that of the amino terminus of cystatins, has long been debated. Since this issue is considered in greater detail in Chapter 2, this section rather gives an overview of the different families of cystatins and their properties.

Members of the cystatin superfamily have been subdivided into a number of families based on:

1. sequence identity or primary structure characteristics
2. the number of disulphide bonds present
3. the molecular mass of the protein (Barrett, 1986).

By these criteria, four families were defined which, with the discovery of cystatins in plants, have been extended to include a fifth family, known as the "phytocystatins" (Kondo *et al.*, 1991). The five cystatin families are:

- Stefins (family I)
- Cystatins (family II)
- Kininogens (family III)
- The non-inhibitory proteins of human histidine-rich glycoproteins and α–2-HS-glycoproteins (family IV)
- Phytocystatins (plant cystatins; family V)

A number of hypotheses have been put forward as to how the different cystatins diverged during evolution. Rawlings and Barrett (1990) suggested that:

- the archetype cystatin, like current members of family I, had no disulphide bonds
- the family II cystatins came into existence with the introduction of disulphide bridges into the secondary structure
- a gene duplication produced the family IV cystatins
- a divergence, followed by a second gene duplication gave rise to the family III cystatins

Abe *et al.* (1987) suggested that the plant and the animal cystatins evolved from a cognate ancestral gene.

The main characteristics of each of the cystatin families are described below.

1.4.1 Family I (Stefins)

This family contains the proteins which lack disulphide bonds (Abrahamson *et al.*, 1986; Turk *et al.*, 1995). They are small, acidic proteins consisting of approximately 100 amino acid residues and are mainly involved in the regulation of protein degradation. They are generally found intracellularly, although there are some indications that they may function extracellularly. In terms of activity the stefins are potent inhibitors of papain and the cathepsins H and L (Turk *et al.*, 1986).

Stefins characterised at the molecular level include stefins A, B, C (Järvinen, 1978; Green *et al.*, 1984) and PLCPI (pig leukocyte cysteine proteinase inhibitor; Lenarčič *et al.*, 1993) which has now been designated stefin D1 (Lenarčič *et al.*, 1996). Porcine stefin D2 has also been identified (Lenarčič *et al.*, 1996) as have

two rodent counterparts of stefins A and B called rat cystatins α and β. Stefins A, C and D1 contain no cysteine residues. Stefin B contains one cysteine residue which can form intermolecular disulphide bridges and is thought to be responsible for the observed formation of dimeric complexes (Barrett, 1987).

The stefins have been isolated from a number of different sources and all have many sequence similarities. Stefin A, for example, has been found in high concentrations in various types of epithelial cells (Rinne *et al.*, 1985; Hopsu-Havu, *et al.*, 1985; Takeda *et al.*, 1989), in polymorphonuclear leukocytes (Brzin *et al.*, 1983) and in lymphoid follicular dendric cells (Alavaikko *et al.*, 1985). This association with epithelial barriers to microbial invasion and with cells involved in protection against infective agents has led to the suggestion that stefin A may be involved in the defense mechanism (Katunuma & Kominami, 1986). Furthermore, since stefin A has been detected in squamous cell carcinomas of the lung, skin, vulva, cervix and oesophagus, it may also play a role in tumour invasion (Rinne *et al.*, 1980; Lah *et al.*, 1990).

In contrast to stefin A, stefin B is uniformly distributed among various cell and tissue types (Barrett *et al.*, 1986). The most likely role of stefin B is in the regulation of proteolysis in the cytoplasm following rupture of lysosomes (Barrett *et al.*, 1986). It could also play a role in the control of intracellular and extracellular breakdown (Stubbs *et al.*, 1990).

1.4.2 Family II (Cystatins)

This group contains the proteins with disulphide bonds and is slightly larger than the stefins, consisting of approximately 115 amino acid residues (Barrett *et al.*, 1986). They occur in high concentrations in human seminal plasma, in saliva and chicken egg white. Examples include chicken cystatin, human cystatin C, the human salivary cystatins and beef colostrum cystatin. A further member of this family, cystatin M, has been isolated from human mammary cells and tissues (Sotiropoulou *et al.*, 1997).

Of the family II cystatins, chicken cystatin is the best known representative and its X-ray crystal structure has been determined (Bode *et al.*, 1988). It is a tight binding inhibitor of the cysteine proteinases ficin, papain, cathepsin B and dipeptidyl peptidase I. Chicken cystatin can be purified as two major isoforms (Turk *et al.*, 1983; Anastasi *et al.*, 1983): a phosphorylated (pI 5.6) and unphosphorylated form (pI 6.5; Laber *et al.*, 1989). This latter, more basic form lacks eight residues on its N-terminus (Curin *et al.*, 1986). Chicken cystatin has no free thiol group and contains four cysteine residues in the C-terminal segment. These form two successive disulphide bridges (Grubb *et al.*, 1984). Because it occurs in such high concentrations in chicken egg white, chicken cystatin is thought to play a role as a defensive agent against pathogens (Barrett *et al.*, 1986).

Three cysteine proteinase inhibitors of this family have also been isolated from saliva, namely cystatin S, cystatin SA and cystatin SN (Isemura *et al.*, 1991). In addition to these S-type cystatins, human saliva also contains cystatin C - another family II member that has 50% homology with the S-type cystatins (Abrahamson

et al., 1986). Cystatin C, like chicken cystatin, is one of the more potent inhibitors of papain and lysosomal cysteine proteinases. It appears to be expressed in all tissues (Abrahamson *et al.*, 1990; Dickinson *et al.*, 1993) with the highest concentrations being found in seminal plasma and cerebrol spinal fluid (Abrahamson *et al.*, 1986). Cystatin C is thought to play a role in physiological control.

1.4.3 Family III (Kininogens)

This family contains the plasma kininogens. These are well known for their involvement in inflammation (Sharma & Mohsin, 1990; Stewart, 1993), in the blood coagulation cascade (Kato *et al.*, 1981; Turk & Bode, 1991) and for the inhibition of papain-like cysteine proteinases (Müller-Esterl *et al.*, 1985). Members of this group have large molecular weights - usually above 65 kDa (Turk *et al.*, 1986) - and typically up to nine disulphide bonds. They are acidic glycoproteins and occur in multiple isoforms with pI values ranging from 4.0 to 5.2 (Auerswald *et al.*, 1993). They have been found in plasma and other mammalian secretions (Turk & Bode, 1991). In mammals, three types of kininogens have been found differing in size, structure and function. They are:

- H-kininogen (M_r 80 - 120 kDa)
- L-Kininogen (M_r 56 - 68 kDa)
- Rat T-kininogen (M_r 68 kDa) (Müller-Esterl *et al.*, 1986).

Human plasma contains H-kininogen (high molecular weight form) with M_r of 83.5 kDa (Turk *et al.*, 1996), and L-kininogen (low molecular weight form) with M_r of 50 - 68 kDa (Okamoto & Greenbaum, 1983). Human kininogens H and L are single-chain plasma glycoproteins each composed of an amino-terminal heavy chain, the kinin segment and a C-terminal light chain. The heavy chains can be divided into three segments, D1, D2 and D3 with the kinin segment making up D4. The light chains of both kininogens differ from one another, and can be divided into domains D5 (LK) and D5 and D6 (HK). All eighteen cysteine residues of LMW kininogen form disulphide bonds. The disulphide bonds are intra-domain with the exception of a link between domain 1 and the light chain (Müller-Esterl *et al.*, 1986; Salvesen *et al.*, 1986). The structure of the H and L-kininogens are shown in Figure 1.2.

Salvesen *et al.* (1986) showed that domains D2 and D3 of the kininogens are potent inhibitors of papain and cathepsin L but poor inhibitors of cathepsin B. Domain D2 differs from other cysteine proteinase inhibitors in that it can inhibit chicken calpain - no other cystatin is known to inhibit the calpains. Domain D1, on the other hand, has proven ineffective against any protease tested so far (Bradford *et al.*, 1993; Turk & Bode, 1991). Turk *et al.* (1995) showed that intact human L-kininogen could bind two molecules of papain-like cysteine proteinases with high affinity and more recently that H-kininogen can also bind two molecules of cathepsin S, cruzipain and papain independently (Turk *et al.*, 1996).

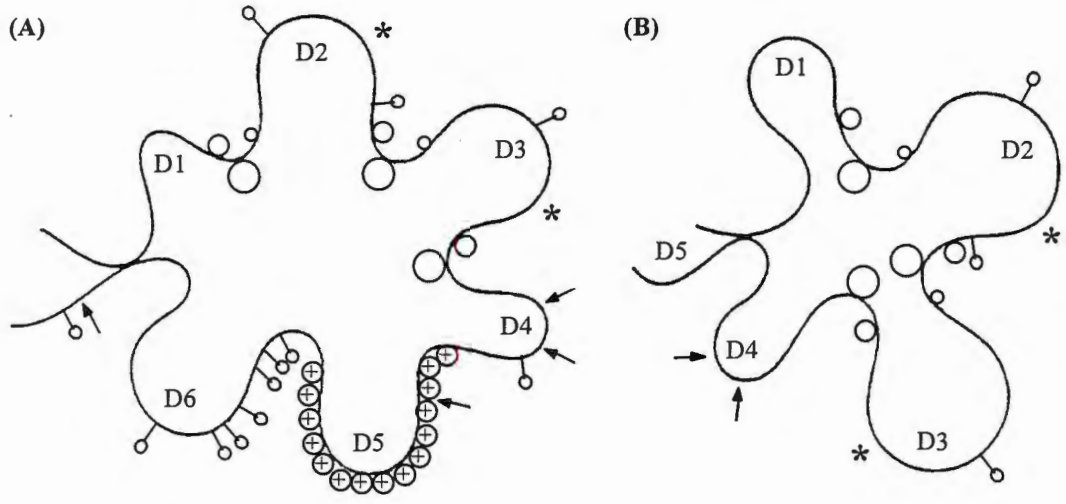


Figure 1.2: *The general structure of human kininogens is represented as comprising multiple domains. (A) HMW kininogen has six domains (D1 - D6) and (B) LMW kininogen has five domains (D1 - D5)*

symbols: * - *putative reactive site* ♀ - *potential disulphide loop*
♀ - *carbohydrate attachment site* ⊕ - *histidine-rich region*
↓ - *cleavage site for kallikrein (from Müller-Esterl, 1986)*

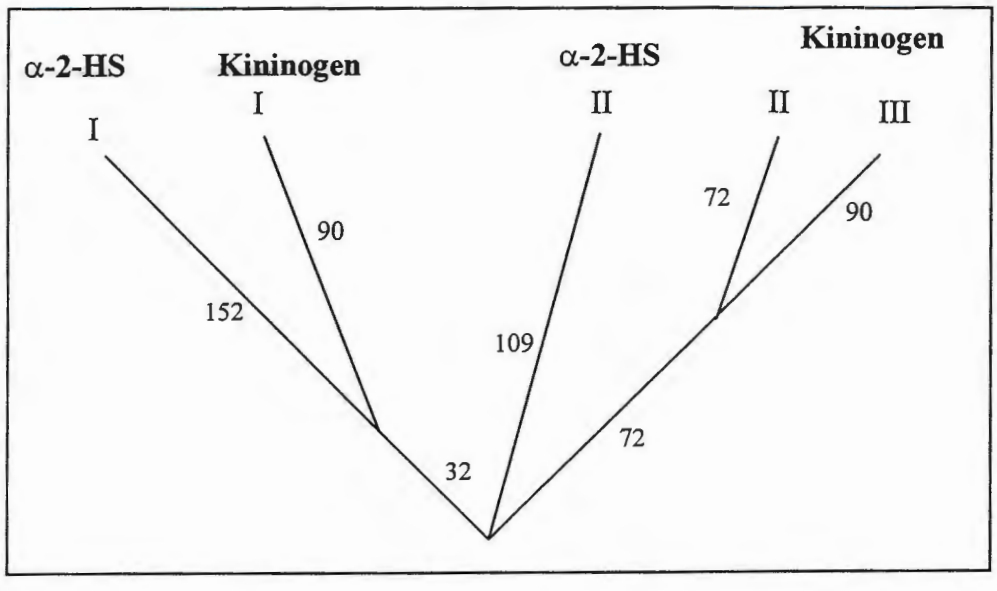


Figure 1.3: *The phylogeny of the cystatin domains of the kininogens and α–2-HS generated by the MATTOP tree program described in Hunt et al. (1985).*

1.4.4 Family IV (α-1-HS and α-2-HS-glycoproteins)

In discussing the evolution of the cystatins Calkins & Sloane (1995) suggested that, despite their similarities to the kininogens, the α-1-HS-glycoprotein (α-1-HS; Ohkubo *et al.*, 1984) and the α-2-HS-glycoprotein (α-2-HS; Elzanowski *et al.*, 1988) should be included as a separate family. These two proteins consist of two repeated patterns of a single cystatin-like molecule (Rawlings and Barrett, 1990; Elzanowski *et al.*, 1988). They occur in very high levels in the matrix of bone and dentine. This association with mineralised tissues and the affinity for calcium and barium ions suggests that these proteins play a role in mineral balance (Elzanowski *et al.*, 1988). Figure 1.3. shows the phylogeny of the cystatin domains of the kininogens and α–2-HS as suggested by Hunt *et al.* (1985).

1.4.5 Family V (Phytocystatins)

The discovery of cystatins in plants added a further group to the cystatin superfamily. This group, which contains all the plant cystatins, has become known as the Phytocystatins. Despite being placed in a family on their own, plant cystatins are similar to the family I cystatins in that they lack disulphide bonds and to the family II cystatins in their overall amino acid homology (Abe *et al.*, 1987b). Some authors suggest that many of these plant cystatins such as oryzacystatin, should be included in the stefin family as the phytostefins (Turk & Bode, 1991), since they contain no disulphide bonds and are more similar in amino acid composition to the stefins.

Table 1.3: Cysteine proteinase inhibitors isolated from plants

Plant	Reference
Potato	Rodis *et al.*, 1974
Pineapple stem	Perlstein & Kezdy, 1973; Hirano et al., 1995
Cowpeas	Rele *et al.*, 1980; Fernandes *et al.*, 1993, Valevski *et al.*, 1991
Mung beans	Baumgarner & Chrispeels, 1976
Seeds of the leguminous Bauhinia tree	Goldstein *et al.*, 1973
Pumpkin seeds	Zimacheva *et al.*, 1988
Maize	Abe *et al.*, 1980; Abe *et al.*, 1992; Abe *et al.*, 1994
Rice	Abe & Arai, 1985; Kondo *et al.*, 1990
Brazilian beans	Olivia *et al.*, 1988
Soybean	Brzin *et al.*, 1990 ; Hines *et al.*, 1991
Chinese cabbage	Lim *et al.*, 1986
Wisteria floribunda seeds	Hirashiki *et al.*, 1990
Wheat, Barley, Rye	Fossom, 1970
Papaya	Song *et al.*, 1986
Tomato	Akers & Hoff, 1980
Millet	Tashiro & Maki, 1986

Corn cystatin, isolated from maize, was one of the first cystatins to be discovered in plants (Abe & Arai, 1980). Subsequently, however, cystatins were found in rice (Oryzacystatins I and II, Abe & Arai, 1985; Kondo *et al.*, 1990) and cysteine proteinase inhibitors in a number of other plant species (Table 1.3).

1.5 The Phytocystatins

In the plant kingdom, cysteine proteinase inhibitors are usually found in seeds and tubers and are particularly prevalent in the families Graminacae, Leguminacae and Solanaceae (Ryan, 1973). However, their role in plants has not been fully elucidated.

The following sub-sections describe two cystatins that have been thoroughly investigated.

1.5.1 The Oryzacystatins

Oryzacystatin I (OC I), was first isolated from rice endosperm by Abe & Arai, (1985). This cystatin contains 102 amino acids, has no disulphide bonds and shares sequence homology with other members of the cystatin superfamily, particularly the family I cystatins (Kondo *et al.*, 1990; Abe *et al.*, 1987b).

Following the isolation of the OC I inhibitor from rice, a second cystatin in rice was discovered. This was named oryzacystatin II (OC II; Kondo *et al.*, 1990). This inhibitor is composed of 107 amino acid residues and shows 55% identity in its amino acid sequence to OC I (Kondo *et al.*, 1990). Figure 1.4 shows an amino acid sequence alignment of OC I and II.

```
OC I:      M-----SS--DGGPVLGGVEPVGNENDLHLVDLARFAVTEHNKKANSLL  42
OC II:     MAEEAQSHAREGGRHPRQ--PAGRENDLTTVELARFAVAEHNSKANAML  47

OCI:       EFEKLVSVKQQVVAGTLYYFTIEVKEGD-AKKLYEAKVWEKPW       84
OC II:     ELERVVKVRQQVVGGFMHYLTVEVKEPGGANKLYEAKVWERAW       90

OCI:       MDFKELQEFKPVDASANA     102
OC II:     ENFKQLQDFKPLDDATA      107
```

Figure 1.4: *Sequences of oryzacystatin I (OC I) and oryzacystatin II (OC II). Commonly conserved amino acid residues are shown in bold with underlining (from Kondo et al., 1990).*

Despite their similarities these two rice cystatins have slightly different expression patterns and perhaps different physiological roles. Evidence for this came from the studies of Watanabe *et al.* (1991), who showed

that the mRNA of OC II is expressed constantly throughout the maturation stages of rice and has even been detected in mature seeds. OC I mRNA, on the other hand, was found to reach a maximum level two weeks after flowering and then gradually decreased to an undetectable level at ten weeks. By this time the ripening of the seeds was complete. OC I mRNA was not detected in mature seeds.

The two rice cystatins have also shown differences in their specificity to different proteinases. For example, OC I has been shown to inhibit papain more efficiently than OC II, while OC II inhibits cathepsin H more efficiently than OC I. The inhibitory characteristics of OC I and II are shown in Table 1.4.

Table 1.4: Inhibitory characteristics of OC I and OC II (from Kondo *et al.*, 1990)

Inhibitor	Ki*	
	Papain	Cathepsin H
Oryzacystatin I	3.02×10^{-8} M	0.83×10^{-6} M
Oryzacystatin II	0.79×10^{-6} M	1.00×10^{-8} M

**The above kinetic studies were carried out using BANA (N-Benzoyl-DL-arginine-2-naphthylamide) as the substrate (Kondo et al., 1990).*

In rice, oryzains α and β are believed to be the two proteinases analogous to papain. These proteinases are therefore inhibited more efficiently by OC I. Another rice cysteine proteinase oryzain γ, on the other hand, is similar to cathepsin H and this proteinase is thus considered to be the natural target of OC II in rice seeds (Abe *et al.*, 1987b; Watanabe *et al.*, 1991). Investigations into the pattern of mRNA expression of oryzacystatin have also led to the deduction of possible functions of these inhibitors in plants. As mentioned previously, the mRNA of the cystatins accumulates in seeds before germination. They are not, however, expressed at germination, which is the time at which the oryzains are expressed (Watanabe *et al.*, 1991). Thus, the presence of oryzacystatins before germination may prevent the degradation of seed storage proteins by the exogenous proteinases of destructive insects or the premature breakdown of the protein by the oryzains. The disappearance of oryzacystatins at the time of germination would then release proteolytic activity for the digestion of storage proteins.

1.5.2 Plant 'Multicystatins'

Besides the single cystatins, other types of cysteine proteinase inhibitors have been found in plants. One of these, the 'multicystatin' was first observed in potato by Rodis (1974). Potato multicystatin (PMC) is quite unique amongst the plant cystatins in that it is able to form crystal-like units at alkali pH. It consists of eight tandem cystatin-like domains (of approximately 10 kDa), which are linked by proteolytically sensitive junctions. They therefore differ from the kininogens, which contain additional regions unrelated to cystatin.

Despite its unique features, PMC is 43 - 50% identical to the rice and corn cystatins, suggesting that all plant cystatin genes may be derived from a common ancestor (Waldron *et al.*, 1993).

PMC has been shown to be active against the cysteine proteinases papain, chymopapain and ficin (Rodis & Hoff, 1984; Krizaj *et al.*, 1993). These crystals occur in the subphellogen layer of potato tubers as well as throughout the tuber and are considered to be the first line of defence against insect attack. The crystals can be solubilised at low pH (Rodis & Hoff, 1984) and by proteinase treatment (Walsh & Strickland, 1993). Since the pH of the midgut of insects using cysteine proteinases for digestion is typically acidic (Murdock *et al.*, 1987), ingested crystals would thus be solubilised and the multicystatin broken up into its inhibitory cystatin units (Walsh & Strickland, 1993).

1.6 The Functions of Cystatins

As can be seen from the previous sections, cystatins have been isolated from numerous sources. Experimental evidence suggests that cystatins not only play important roles in the protection of cells from unfavourable proteolysis by cysteine proteinases, but also in biological defence systems against invaders (Bieth, 1986). An example of such protection is provided by the cysteine proteinase inhibitor stored in the large granules of horseshoe crab hematocytes. This cystatin serves as a defence against invading microbes and, together with other defence molecules, is released in response to external stimuli (Agarwala *et al.*, 1996). Similarly, many insects have also been shown to contain proteinase inhibitors that protect against proteases from invading fungi and microorganisms (Eguchi, 1993).

Thus, these proteinase inhibitors could show potential as a means of controlling the proteinases of human and plant parasites and viruses. Cystatins could also be used as biocontrol agents to protect plants from insect attack through inhibition of the insect digestive proteinases. These potential applications of the cystatins in the pharmaceutical and agricultural fields have been made more attractive by the development of techniques to improve the capabilities of inhibitors through modelling studies and site-directed mutagenesis, thus making them more specific for their target enzymes.

The following subsections discuss examples from the literature where cystatins have shown potential in the area of host protection.

1.6.1 Cystatins and Control of Human Parasites

Cysteine proteinases have been isolated from the lysates of many different parasitic protozoa and nematodes (North, 1992; Atkinson *et al.*, 1995). Some examples include; the pathogenic nematode, *Haemonchus*

contortus (Pratt *et al.*, 1990; Cox *et al.*, 1990); *Trypanosoma brucei* (Lonsdale-Eccles & Grab, 1986); *Leishmania mexicana mexicana* (North, 1992) and *Toxoplasma gondii* (Irvine *et al.*, 1992).

Many parasites depend on proteolysis for nutrition, for entry into tissues or cells or to counteract, directly or indirectly, the host's defence mechanisms (Irvine *et al.*, 1992). Furthermore, parasites often have complex life-cycles and proteinases are important in protein turnover during cell differentiation and in the transformation from one stage to another. For these reasons, it may be beneficial to identify functional differences between the proteinases of the host and those of the parasite that might be exploited by appropriately tailored inhibitors.

Proof of the potential for proteinase inhibitors in host protection can be seen from the work of Silva *et al.* (1995), who showed that on infection by *Trypanosoma cruzi,* cystatin S was induced in rats as a defence mechanism. They suggested that this increase in cystatin S on infection may be related to the presence of *T. cruzi* proteinases acting as antigens. Hence the synthesis of inhibitors by the host may be a biological response to tissue injury. This in turn, serves to protect the host from the harmful effects of enzymes and/or facilitate tissue repair (Silva *et al.*, 1995). Various cystatins have also been shown to inhibit the cysteine proteinase (trypanopain) from *Trypanosoma brucei in vitro* (Lonsdale-Eccles & Grab, 1986).

1.6.2 Cystatins and Control of Viruses

The use of proteinase inhibitors as protective mechanisms may also extend to plants in protecting them against various viruses and plant pathogens. For example, cysteine proteinase inhibitors have been identified in resistant plant species which are able to affect the protein processing of the cowpea mosaic virus (Ponz *et al.*, 1987). In fact a number of plant viruses, such as the comoviruses and the potyviruses, as well as a number of other types of viruses, require cysteine proteinases for the processing of their precursor polyproteins. These protein cleavages are often central to the replication cycle of the viral parasites and if these are prevented, the infection process is aborted (Korant *et al.*, 1986). In the case of the potyviruses, the cysteine proteinase responsible for this processing is located at the carboxyl-terminal third of the polyprotein. The role of this enzyme is to process itself out of the polyprotein and to then cleave the coat proteins from their precursor. This precedes their assembly into viral particles (Carrington *et al.*, 1990).

In view of the importance of these cysteine proteinases in the virus life cycle they have been considered as putative targets for anti-viral therapy using proteinase inhibitors. Molecular modelling approaches and genetic engineering techniques could also be used to increase the inhibitory capabilities of the inhibitor against specific viral proteinases.

Turk *et al.* (1983) investigated the effect of cystatins on the cysteine proteinases of the neurovirulent poliomyelitis virus type I. At 100 μg/ml concentrations of chicken cystatin these authors observed an

alteration in protein processing and a decrease in the total quantity of viral protein that was dependent on the concentration of chicken cystatin and on whether or not the cells were exposed to the cystatin prior to infection. At high concentrations of chicken cystatin there was no viral protein synthesis and the pattern of cellular protein synthesis was preserved. This showed that the ability of the poliovirus to abolish cellular protein synthesis was blocked by chicken cystatin. Cystatins thus appear to alter the pattern of viral polypeptide processing (Turk *et al.*, 1983).

Similar antiviral effects have been observed with HSV (Herpes simplex virus type I) by oryzacystatin (Aoki *et al.*, 1995) and cystatin C (Bjorck *et al.*, 1990; Kondo *et al.*, 1992). In fact, the molar concentrations of cystatin C that gave total inhibition of HSV replication were found to be lower than that of acyclovir, a drug most frequently used against HSV infections. These results suggest that cysteine proteinase inhibitors might play a role as inhibitors of viral replication and that peptide derivatives that mimic their proteinase-binding centres, might be useful as antiviral agents. Cystatin C does not inhibit polio virus while chicken cystatin and oryzacystatin I do (Bjorck *et al.*, 1990; Aoki *et al.*, 1995). The reason for these differences is still unclear.

1.7 Potential of Cystatins as Biocontrol Agents

This last section, aims to assess the advantages and disadvantages in the use of cysteine proteinase inhibitors against insect pests and whether transformation of proteinase inhibitor genes into plants will serve to enhance their resistance.

The potential use of proteinase inhibitors to protect crop plants from insect damage is well recognised. Early demonstrations showed that proteinase inhibitors are capable of repressing the activity of proteinases from the insect digestive tract (Gatehouse & Boulter, 1983; Wolfson & Murdock, 1987). This is an important consideration for biocontrol since economic losses per annum from insect infestation runs into billions of dollars worldwide.

A number of hypotheses have been suggested regarding the efficiency of proteinase inhibitors in this particular context. One of the ways they are thought to act is through tight binding to the target digestive proteinases, which leads to a general decrease in the proteolytic digestive functions and hence to delays in growth and development. A second proposed mode of action is thought to be via feedback mechanisms. Here the presence of the inhibitor causes an increase in the production of midgut proteinases. This in turn results in amino acid utilisation and wasted energy.

Much of the work carried out using proteinase inhibitors as biocontrol agents have been done using serine proteinase inhibitors. Serine proteinase inhibitors transformed into tobacco plants, for example, have successfully conferred resistance to *Manduca sexta* larvae (Johnson *et al.*, 1989). Despite the successes with

serine proteinase inhibitors, cysteine proteinase inhibitors have proven variable in their potential as biocontrol agents in plants. They have, nevertheless, been successful in a number of cases, thus encouraging their use as protective agents in plants, especially when used in conjunction with other forms of biocontrol. Furthermore, there are many advantages in using cystatins rather than other types of proteinase inhibitors for biocontrol. For example, unlike the serine proteinases, cysteine proteinases are not secreted as intestinal digestive enzymes in higher animals, but are found in the midguts of several families of Hemiptera and Coleoptera. Thus the presence of high concentrations of inhibitors in plants should have no detrimental effects on humans and animals. Furthermore, trial studies have shown that cystatins do inhibit midgut cysteine proteinase *in vivo*. For example, oryzacystatin has been shown to retard the growth of the red flour beetle (Chen *et al.*, 1992; Liang *et al.*, 1991), as has E-64 against the bean weevil (Hines *et al.*, 1990) and the Colorado potato beetle (Wolfson & Murdock, 1987). These functions have led to the proposal, that inhibitors such as oryzacystatin be expressed in transgenic cereals to protect seeds against infestation by stored grain Coleoptera (Liang *et al.*, 1991).

1.7.1 The Possibility of Transforming Cysteine Proteinase Inhibitors into Plants

As has been described in section 1.3.1, cysteine proteinases appear to play certain physiological roles in plants. Since OC I is known to inhibit proteinases of plant origin such as papain and endogenous rice proteinases, the concept of introducing these inhibitors into transgenic plants has been controversial. Masoud *et al.* (1993) placed the OC I gene under the control of the CaMV 35S promoter in tobacco. High levels of the inhibitor in this case, did not interfere with the plants' normal developmental processes. The tobacco plants containing the OC I gene were found to be normal in appearance, self-fertile, and produced viable seeds.

Similarly, the OC I gene has been introduced and expressed in poplar trees where it accumulated to high levels, especially in older leaves and successfully conferred resistance against the poplar leaf beetle, *Chrysomela tremulae,* with no detrimental effects on the poplar trees (Leplé *et al.*, 1995).

There thus appears to be potential for cystatins such as oryzacystatin to be used in transgenic plants despite the possible physiological roles that these proteinases play in the plant itself. Because of the limited expression in some plant organs it may be necessary to express the inhibitors in those specific organs which form part of the diet of the target insect pest.

1.7.2 Problems Associated with Cystatins as Biocontrol Agents.

1. The Wide Diversity of Insect Digestive Enzymes

All four classes of proteinases are represented within the insect order and different types of insects utilise different proteinases for their digestion. Members of the Hemiptera and Coleoptera, for example, use cysteine and/or aspartic proteinases (Houseman & Downe, 1980) while pepsin-like enzymes are found in some Diptera (Pendola & Greenberg, 1975). Trypsin-like enzymes, on the other hand, are common in Lepidoptera (Applebaum, 1985). There are no definite rules to this classification, however, since the pH profiles in the midguts of some insects have been shown to exhibit multiple peaks, suggesting that these insects utilise more than one type of proteinase (Wolfson & Murdock 1989).

The reason for such variation in insect digestive enzymes is not particularly clear. Houseman *et al.* (1985) postulated that insect adaptation from using serine proteinases to that of aspartic and cysteine proteinases may have occurred as two evolutionary events. The first was the loss of serine proteinases following a change to sap feeding by Homopteran-like insects (thought to be the ancestors of the entire Hemiptera order; Terra & Ferreira, 1994; Houseman & Downe, 1983). The second event was the advent of predatory habits. With higher protein sources or even seed feeding, catheptic lysosomal proteinases became the dominant enzymes for digestion (Houseman *et al.*, 1985). Alternatively, it has been suggested (Ryan *et al.*, 1990) that in both the Hemiptera and Coleoptera the presence of catheptic proteinases in the midguts may have been an adaptation to feeding on food sources rich in trypsin inhibitors. In this case, the use of alternative proteinases could have been developed to overcome the potential toxic effects of these ingested inhibitors.

The problems associated with this large diversity in insect digestive proteinases is illustrated by the work of McManus *et al.* (1994) who introduced a chymotrypsin inhibitor gene into tobacco and so conferred resistance against the green looper (*Chrysodeixis erisoma;* Lepidoptera: Noctuidae). It did not, however, inhibit other closely related noctuids. This suggests that a single proteinase inhibitor gene may not be universally effective against a wide range of insect pests.

Further evidence for such variability in digestive proteinases within an insect order has come from studies on the Coleoptera. Although these insects are thought to use cysteine proteinases for digestion (Wolfson & Murdock, 1989; Michaud *et al.*, 1995a), not all species of beetle use them. Proteolytic activity in the midgut homogenates of the milkweed beetle, for example, was not inhibited by cysteine proteinase inhibitors while the midgut of the mealworm, *Tenebrio molitor,* has been shown to utilise a trypsin-like serine proteinase for digestion (Applebaum, 1985). The carpet beetle *Attagenus megatoma* also relies, at least in part, on serine proteinases for protein digestion (Baker, 1976) as does *Costelytra zealandica* (the grass grub; Christeller & Shaw, 1989). It is thus clear that proteolytic activity varies within a species as well as from species to species. Thus, the potential for an inhibitor to be active against specific insects depends on the proteinase type(s)

present in the midgut. Table 1.5 shows a number of insects and the types of midgut digestive proteinases identified. From the table, it is clear that some insects utilise more than one type of proteinase for digestion.

Table 1.5: Proteinases identified in insect midguts.

Insect	Proteinases in midgut	Inhibitors	Reference
Mexican bean beetle *(Epilachna varivestis)*	Cysteine	E-64	Murdock *et al.*, 1987
Cowpea weevil (*Callosobruchus maculatus*)	Cysteine	E-64	Murdock *et al.*, 1988
Red flour beetle (*Tribolium castaneum*)	Cysteine	E-64 OC I	Murdock *et al.*, 1987; Chen *et al.*, 1992; Liang *et al.*, 1991
Callosobruchus chinensis	Cysteine	OC I; OC II	Kuroda *et al.*, 1996
Riptortus clavatus	Cysteine	OC I; OC II	Kuroda *et al.*, 1996
Bean weevil, (*Zabrotes subfascaitus*)	Cysteine	E-64	Silva & Xavier-Filho, 1991
Mealworm (*Tenebrio molitor*)	Cysteine, Serine	Bowman-Birk; E-64	Murdock *et al.*, 1987
Poplar leaf beetle (*Chrysomela tremulae*)	Cysteine	OC I	Leplé *et al.*, 1995
Carpet beetle (*Attagenus megaloma*)	Cysteine, Serine	Bowman-Birk; E-64	Baker, 1976
Colorado potato beetle (*Leptinotarso decemlineata*)	Cysteine, Aspartic	OC I; OC II; E-64; pepstatin	Wolfson & Murdock, 1987 Michaud *et al.*, 1993
Milkweed beetle	Cysteine, Serine	Bowman-Birk; E-64	Applebaum, 1985
Rice weevil, (*Sitophilus oryzae*)	Cysteine, Aspartic	OC I; Antipain; E-64	Liang *et al.*, 1991
Southern corn rootworm *(Diobrotica undecimpunctata)*	Cysteine	PMC	Orr *et al.*, 1994
Grass grub (*Costelytra zealandica*)	Serine	Bowman-Birk	Christeller & Shaw, 1989

In conclusion therefore, it appears that there may be disadvantages in transforming single proteinase inhibitor genes into plants because of the large diversity in digestive proteinase activity amongst insects. Furthermore, insects with multiple proteinases in their guts may be able to resist the effects of single proteinase inhibitors by increasing the secretion of proteinases insensitive to these inhibitors. A possible solution to this problem would be to clone more than one type of inhibitor gene into plants. In this way, plants would have resistance against a much wider range of insect pests.

2. Inhibitory Capabilities of Cystatins

Kinetic studies carried out with cystatins have shown that the inhibitory constants for different cysteine proteinases vary. This presents a further problem for biocontrol since different members of the Coleoptera and Hemiptera use different cathepsins for their midgut protein digestion. OC II, for example, is a better inhibitor of the midgut proteinases of the Colorado potato beetle, than is OC I (Michaud *et al.*, 1993). A possible reason for this is that the major gut proteinases found in this insect are cathepsins B and H (Michaud *et al.*, 1993) and although both OC I and OC II are fairly poor inhibitors of cathepsin B, OC II is a better inhibitor of cathepsin H than is OC I (Kondo *et al.*, 1990).

It is therefore important to know the specific proteolytic systems used by the insects to be controlled and either use the appropriate inhibitor or, by protein engineering and molecular modelling, develop inhibitors with both improved inhibitory capabilities and a wider inhibitory range.

3. High Concentrations of Cystatins are Required for Effective Inhibition

Results obtained from the incorporation of cystatins into the diets of insects have depended very much on the trial conditions, particularly on the concentrations of inhibitors used. E-64, for example, prolonged the developmental time and increased larval mortality of the cowpea weevil, *Callosobruchus maculatus*, in direct proportion to its concentration in artificial seeds (Murdock *et al.*, 1988). Similarly, OC I and OC II, when added to the diets of two bean insects, *Callosobruchus chinensis* (Coleoptera) and *Riptortus clavatus* (Hemiptera) at concentrations of 0.3 - 0.5% (w/w), retarded the growth of these insects (Kuroda *et al.*, 1996).

At concentrations of 1% inhibitor in the diets, the insects died. In contrast, studies carried out with *Tribolium castaneum* (Chen *et al.*, 1992) required the addition of 10% (w/w) OC I to the diet in order to achieve inhibition of gut proteinases. Thus the outcome, especially with assays in which the insect guts are extracted and analysed (i.e. *in vitro* assays), depends very much on the conditions used, including the type of diet and the concentrations of inhibitor. These will also differ with different insects depending on the types of proteinases present in the midguts and whether or not there is more than one class of proteinase involved.

For Coleoptera it is often necessary to have inhibitors in excess in order to avoid recovery of digestive proteolytic functions and/or to ensure the inhibition of newly synthesised proteinases. In general, proteinase inhibitors need to be present in plants in relatively high concentrations (greater than 5% of the soluble protein) in order to be toxic to insects. This presents potential problems since normally plant cystatins, such as oryzacystatin, occur in rice seeds at low levels (approximately 0.001 - 0.002%; Kondo *et al.*, 1989), which may be insufficient for effective protection against insect pests. In fact, the effects of proteinase inhibitors at low concentrations only amount to a developmental delay of one or two days (Hilder et al., 1987). There are thus

difficulties in delivering sufficient cysteine proteinase inhibitors to interfere with insect growth and development in a useful way.

A further problem is that the oryzacystatin gene is not expressed in the leaves, stems or roots of the rice plant. This cystatin would thus be ineffective against rice insect pests, such as *Lagynotomous elongatus* and *Saccarosydne procerus* that eat the leaves of growing plants. *Sitophilus zeamais,* on the other hand, invades the interior of the rice seeds where the inhibitors are expressed (Irie *et al.*, 1996). It may thus be necessary to produce transgenic plants that express higher concentrations of inhibitors and target these inhibitors to specific plant organs.

4. Insect Resistance/ Adaptation to the Presence of Inhibitors in their Diets

In many of the studies carried out in which insects were fed proteinase inhibitors either mixed with artificial diets or, present in the plant material itself, the insect mortality has generally been so low that protection is incomplete despite the high levels of inhibitors. It has been suggested that this may be due to the ability of insects to overcome the detrimental effects of proteinase inhibitors by over-producing proteinases insensitive to the inhibitor (Broadway & Duffy, 1986). Evidence for this apparent 'resistance' of insects over time was observed in studies with the Colorado potato beetle (Orr *et al.*, 1994). Here the addition of phenyl-methyl-sulphonyl-fluoride (PMSF) to the insect diets caused an increase in the production of cysteine proteinases and the insects were able to recover from exposure to some extent. Larvae of the Southern corn rootworm were also seen to recover, with time, from acute exposure to potato multicystatin (PMC). Orr *et al.* (1994) proposed that these larvae must have avoided a drastic reduction in their digestive capacity through a 2-3-fold increase in the secretion of proteinases 'insensitive' to the PMC and that such larvae possess feedback mechanisms for monitoring and maintaining an adequate level of proteinase activity. In the field this may show itself as a delay in the developmental time which might not translate to any significant reduction in crop damage. Some researchers, however, consider that this over-production itself may be detrimental to insects in terms of energy wastage and amino acid utilisation (Broadway & Duffy, 1986; Bolter & Jongsma, 1995).

In general the ability to counteract the effects of proteinase inhibitors appears to be limited. Often 70% of insect digestive enzymes are affected by the proteinase inhibitors, and the production of inhibitor insensitive proteinases is often insufficient to restore normal levels of activity. The simultaneous delivery of a few proteinase inhibitors with different specificities may help by minimising the effects of insect adaptation seen with the secretion of proteinase inhibitor 'insensitive' proteinases.

5. Variation in the Results obtained with *in vitro* versus *in vivo* studies

A further problem observed with proteinase inhibitor tests stems from the fact that the encouraging results seen with *in vitro* tests, are often not observed in *in vivo* assays. That is, proteinase inhibitors appear to drastically reduce the activity of cysteine proteinases from gut extracts, but this is not translated into increased mortality rates in the *in vitro* assays. Chen *et al.* (1992), for example, found that oryzacystatin was very effective in inhibiting the gut proteinases of *T. castaneum in vitro*, but that high levels (10 % w/w), were required to inhibit growth *in vivo*.

Furthermore, some *in vivo* studies have shown proteinase inhibitors to be susceptible to proteolysis by insect midgut proteinases "insensitive" to the inhibitor. For example, studies carried out with the black vine weevil (*Otiorynchus sulcatis*) using OC I and OC II showed that OC II was susceptible to proteolysis (Michaud *et al.*, 1995b). When the inhibitor was incubated with the black vine weevil midgut proteinases and then sequenced, it was found that the C-terminus of OC II had been truncated. This initial cleavage led to a conformational destabilisation of the inhibitor. This susceptibility of certain inhibitors, such as OC II, could have major implications when planning the use of plant cystatins for insect pest control. Here multicystatins such as PMC may be more useful as its individual cystatin domains are less susceptible to proteolysis.

One may conclude that cystatins do show potential as possible biocontrol agents in plants, although a number of problems are associated with their use in this regard. These and possible solutions are summarised in Table 1.6.

Table 1.6: Problems and solutions in the use of cystatins as potential biocontrol agents.

PROBLEMS	SOLUTIONS
Wide diversity in insect digestive enzymes	Transgenic plants containing multiple inhibitors
Specificity of inhibitors for certain cysteine proteinases	Protein engineering to improve the capabilities of the inhibitor against more than one target enzyme
Insect resistance to the presence of inhibitors in their diets	Transgenic plants containing multiple inhibitors

1.7.3 Further Approaches to Improve Cystatins for Biocontrol

In addition to the solutions summarised in Table 1.6, a number of other approaches can be used to overcome these problems namely:

1. Use of multicystatin inhibitors in transgenic plants rather than single cystatins e.g. potato multicystatin (PMC).
2. Use of inhibitors in conjunction with other forms of biocontrol e.g. the *Bacillus thuringiensis* (Bt) toxin.

These are discussed below:

1. Advantages of using 'Multicystatins' rather than Individual Cystatins

The use of multicystatins, such as PMC, has been shown to give better inhibitory activity than the single cystatins such as oryzacystatin and chicken egg white cystatin. Orr *et al.* (1994) investigated the use of PMC against the cysteine proteinase activity of the southern corn rootworm larvae *in vivo* and *in vitro* and its growth inhibitory effect was far more potent than either oryzacystatin I or chicken cystatin. Furthermore, its activity was comparable to that of the potent, irreversible cysteine proteinase inhibitor E-64 against bean weevil and cowpea weevil (Orr *et al.*, 1994).

2. Use of inhibitors with other forms of Biocontrol Agents

A possible means of obtaining efficient protection in plants would be to use different pest resistant factors that act through different mechanisms. Such dual approaches have been used successfully in transgenic plants. Boulter *et al.* (1990) expressed both the cowpea trypsin inhibitor gene and/or the lectin gene in tobacco. Progeny that inherited and expressed both genes showed an enhanced resistance to *H. virescens* when compared to plants that expressed only one of the two genes.

A further example, is provided by the use of proteinase inhibitors in conjunction with Bt-toxin. Using this approach, Hilder *et al.* (1987) used serine proteinase inhibitors in conjunction with the Bt toxin and found that the mortality of tobacco budworms was 2- to 20-fold higher than that obtained with the Bt toxin alone. This approach may help to prevent the development of insects that are resistant to this toxin but not to the proteinase inhibitor.

1.8 Aims of this Thesis

It appears that there are many advantages to improving the potential of proteinase inhibitors as biocontrol agents. One approach (as mentioned briefly in section 1.7.2) would be to optimise the interactions between proteinase inhibitors and their target proteinases. This could be carried out by screening many naturally occurring inhibitors with *in vitro* enzyme assays using gut extracts from insect pests. The gene encoding the inhibitor with the strongest interaction could then be expressed in transgenic plants. This strategy does however, have its limitations, since, as mentioned previously, positive *in vitro* results do not always predict effective *in vivo* inhibition.

however, have its limitations, since, as mentioned previously, positive *in vitro* results do not always predict effective *in vivo* inhibition.

A second option could be to develop molecular modelling approaches. Many of the interactions of proteinase inhibitors with their cognate enzymes have been studied in detail at the structural level and functional residues on the inhibitors have been well characterised. One could thus use protein engineering methods to try and increase the strength of these interactions. It is this latter point which was applied in this thesis.

The success of manipulating cystatins via protein engineering approaches to improve biocontrol has been demonstrated by Urwin *et al.* (1995a) who developed a variant of natural OC I that carried a mutation in the form of a deleted Asp^{86} residue. This inhibitor was found to retard the growth of the plant nematode *Globodera pallida,* and had a greater inhibitory effect than that of the natural OC I. In fact, it reduced the size of *G. pallida* females to the extent that fecundity was greatly affected (Urwin *et al.*, 1995a). Thus, modification of inhibitors has shown potential as a means of increasing their biocontrol capabilities.

Although successes such as this are encouraging, the rational engineering of protein inhibitors is still an uncertain and somewhat unpredictable science. The knowledge of structure-function relationships is as yet incomplete and the current lack of structural data on insect and nematode proteases means that the modelling approach relies more heavily on speculation rather than on the sound definitive basis that would support effective prediction.

The aim of this thesis was to investigate further the possible interactions of cystatins by means of structural modifications. Oryzacystatin I was used as the model inhibitor because plant cysteine proteinases have shown potential as biocontrol agents against certain insect pests (Liang *et al.*, 1991; Leplé *et al.*, 1995; Michaud *et al.*, 1995a; Irie *et al.*, 1996). Oryzacystatin I was used rather than oryzacystatin II as it has proven to be a better inhibitor in the majority of insect trials (Michaud *et al.*, 1995b). It has also been shown to be less susceptible to proteolysis in *in vivo* assays.

The inhibitory constants of a number of cystatin inhibitors were obtained from the literature and compared to oryzacystatin. It was found that OC I, which has a Ki of about 10^{-8} M for papain, binds about 4 orders of magnitude less efficiently than does, for example, chicken egg white cystatin (10^{-12} M). The fact that these two cystatins display such a marked difference in their affinity for the same proteinase suggested that the plant cystatin could be modified to improve its efficiency against both papain and possibly insect proteinases. Molecular modelling studies were thus carried out to determine where in OC I changes could be made without disrupting the entire structure. Following a study of the cystatin structures, it was decided to substitute the N-terminus of oryzacystatin with the chicken cystatin N-terminus, for the reasons discussed in Chapter 2. The inhibitory capabilities of this 'hybrid' cystatin were then compared with those of the natural inhibitors. The aim of this substitution was to determine whether the chicken cystatin N-terminal sequence could enhance the

binding of OC I to papain. Such manipulations of the inhibitor structure may also give some insight into the interactions of cystatins in the amino-terminus. Furthermore, should the inhibitory activity of this 'hybrid oryzacystatin' be enhanced, it could improve its potential as a possible biocontrol agent in plants.

CHAPTER 2

Design and Molecular Modelling of the Hybrid Oryzacystatin Gene

PART 1: Literature Review

2.1 Introduction

Several approaches have been used to study cysteine proteinase inhibitors. These include crystallographic structure determination of both free and proteinase-bound inhibitors (Bode *et al.*, 1988; Stubbs *et al.*, 1990), molecular modelling studies of enzyme-inhibitor interactions (Bode *et al.*, 1990), the use of recombinant cystatins modified by site-directed mutagenesis (Shibuya *et al.*, 1995; Machleidt *et al.*, 1995; Björk *et al.*, 1995) and kinetic analyses of enzyme:inhibitor interactions (Björk *et al.*, 1989).

Many of the models proposed for the interactions of cystatins with their cysteine proteinases suggest that this mechanism differs from that of serine proteinase inhibitors. Instead of forming covalent bonds with the active site residues, as is seen with serine proteinase inhibitor complexes (Laskowski & Kato, 1980; Bode & Huber, 1992b), cystatins interact with minimal conformational changes (Lindahl *et al.*, 1992; Pol *et al.*, 1995). Steric hindrance of substrate binding is achieved by close interactions of the binding regions of the cystatins with the active site cleft of the proteinase (Bode & Huber, 1992b; Auerswald *et al.*, 1992).

Models have been proposed to describe the interactions between cystatins and their cysteine proteinases (Machleidt et al., 1995) and the importance of the various binding areas of cystatins have long been debated. In designing the hybrid oryzacystatin gene that is the focus of this thesis, these and the structure-function relationships as outlined in the following sections were taken into account.

2.2 Cysteine Proteinase Structure and Enzymatic Mechanisms

Of the cysteine proteinases, only papain (from papaya fruit, *Carica papaya;* Drenth *et al.,* 1971), calotropin DI (from the Indian Madar plant, *Calotropis giganta*; Heinemann *et al.*, 1982), actinidin (from kiwifruit, *Actinidia chinensis;* Carne & Moore, 1978), human cathepsin B (Musil *et al.*, 1991) and carican (papaya proteinase Ω; Pickersgill *et al.*, 1991) have had their 3-D structures deduced by X-ray diffraction data.

Papain is one of the most studied enzymes of the cysteine proteinase group. Many of these studies have concentrated on the chemical, physical or the kinetic properties of papain. This enzyme consists of a single polypeptide chain which is folded to form a cleft between two regions, known as the L and R domains. Each domain has a central core of hydrophobic residues, with the interface between the two being predominantly polar in nature. The L-domain is mostly α-helical with three α helices, designated A (residues 24 - 42, a long α helix through the centre of the molecule at the interface between the two domains), B (residues 50 - 57) and C (residues 67 – 78) (Figure. 2.1). The R-domain, on the other hand, is a twisted anti-parallel β-sheet which is folded over in the form of a barrel. Its interior is filled with hydrophobic side-chains. Two α-helices; D (residues 117 - 127) and E (residues 138 - 143) on the molecular surface, seal opposite ends of this barrel.

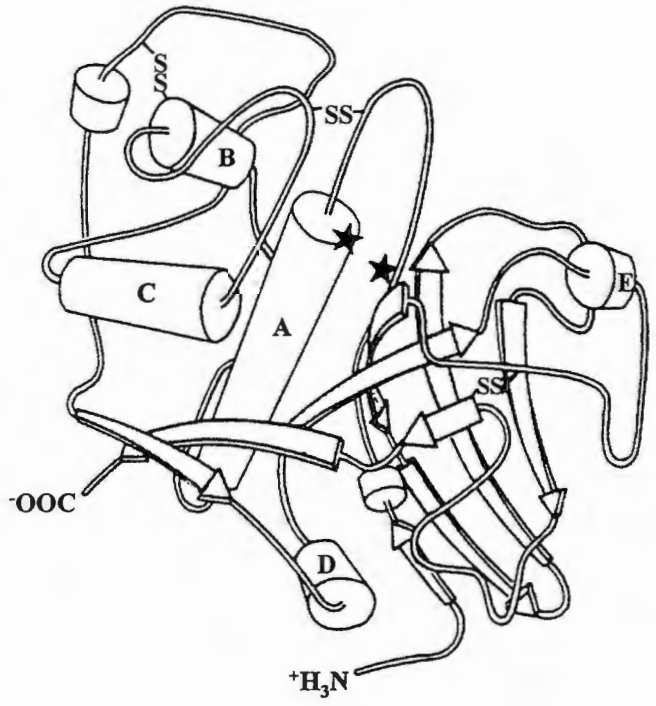


Figure 2.1: *Diagram of a model showing the chain folding patterns of papain. The stars represent the positions of the catalytic cysteine and histidine residues. The cylinders represent the α helices and the arrows the β-sheets.*

The cleft between the L and R domains contains the catalytic site residues. Cys^{25} forms part of the L-domain while the catalytic site histidine residue forms part of the R-domain (Drenth *et al.*, 1971; Baker & Drenth, 1986). Figure 2.2 shows the positions of the active site residues of papain and the coincident stefin B binding site. From this model it is clear that the stefins and cystatins must prevent substrate binding through steric hindrance.

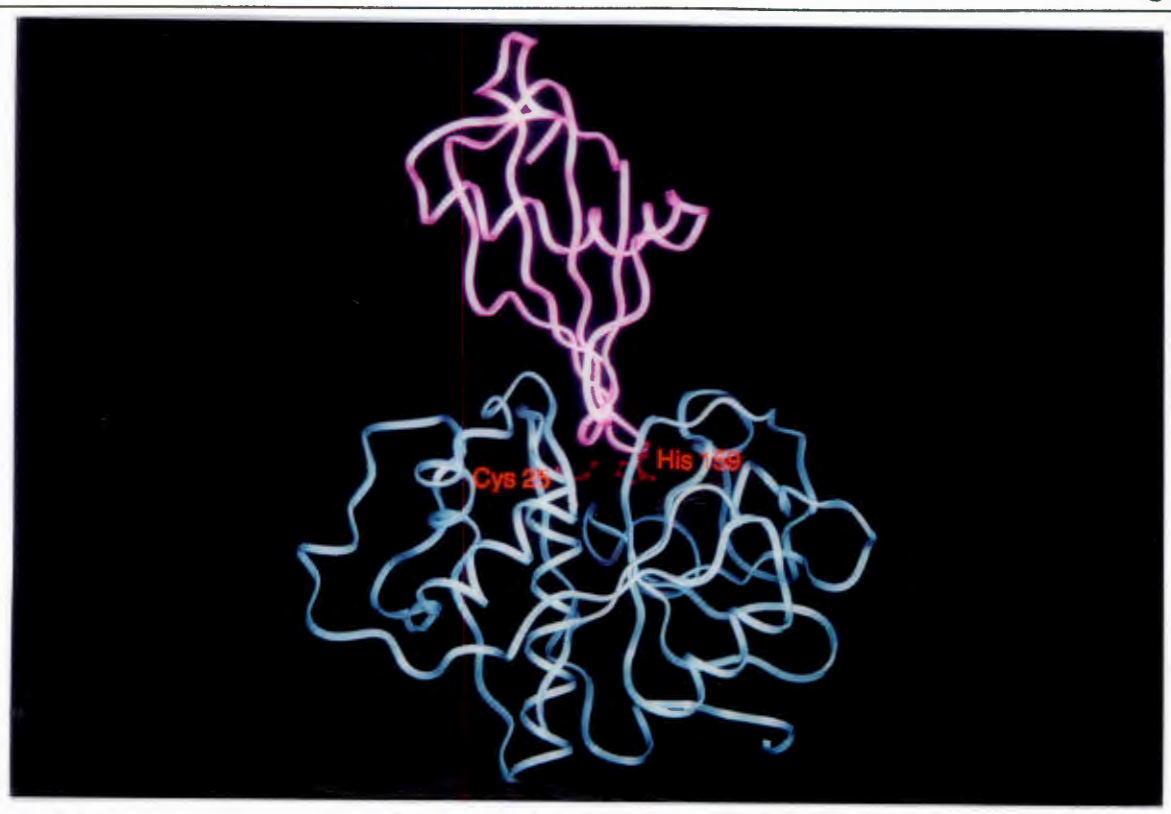


Figure 2.2: *Ribbon model of stefin B (pink) docked onto papain (white). The positions of the active site residues Cys^{25} and His^{159} are indicated.*

Cathepsin B is the only member of the cathepsins whose X-ray crystallographic structure has been elucidated (Musil *et al.*, 1991). Unlike the other cysteine proteinases, cathepsin B can act as both an endopeptidase and a peptidyldipeptidase (Musil *et al.*, 1991). This latter capacity to remove C-terminal dipeptides has been attributed to the presence of a 20-residue insertion termed the occluding loop, which is formed by residues 108 - 119. This loop blocks the right-hand side of the substrate binding cleft through steric hindrance. The residues His^{11} to Asn^{113} of the occluding loop have been shown to collide with the second hairpin loop of cystatins (Musil *et al.*, 1991), thus making the fit with cathepsin B less efficient than with papain. This may explain the reduction in affinity for the interaction of cathepsin B with chicken cystatin and stefin B (Thiele et al., 1990) and with cystatin C (Illy *et al.*, 1997). The loss of affinity is less with stefin A as it has a shorter second hairpin loop (Popovič *et al.*, 1988).

Many studies have also been carried out with cathepsin H which has aminopeptidase activity. These have highlighted the role of an additional disulphide linked peptide in cathepsin H which originates from a part of this enzyme known as the mini-chain. This peptide is probably located in the S2 region of the active site cleft and blocks the left-hand side of the substrate binding site, thus mediating the aminopeptidase activity (Rothe *et al.*, 1994). These structural differences between the various cysteine proteinases are thought to account for their different susceptibilities to inhibitors (Björk *et al.*, 1994).

2.2.1 Catalytic Mechanism

Crystallographic studies (Drenth *et al.*, 1971) have implicated the thiol residue Cys^{25} (papain numbering) and also the imidazole group of His^{159} (papain numbering) as the catalytic site residues of cysteine proteinases.

The basic steps involved in the hydrolysis of polypeptide substrates are shown schematically in Figure 2.3.

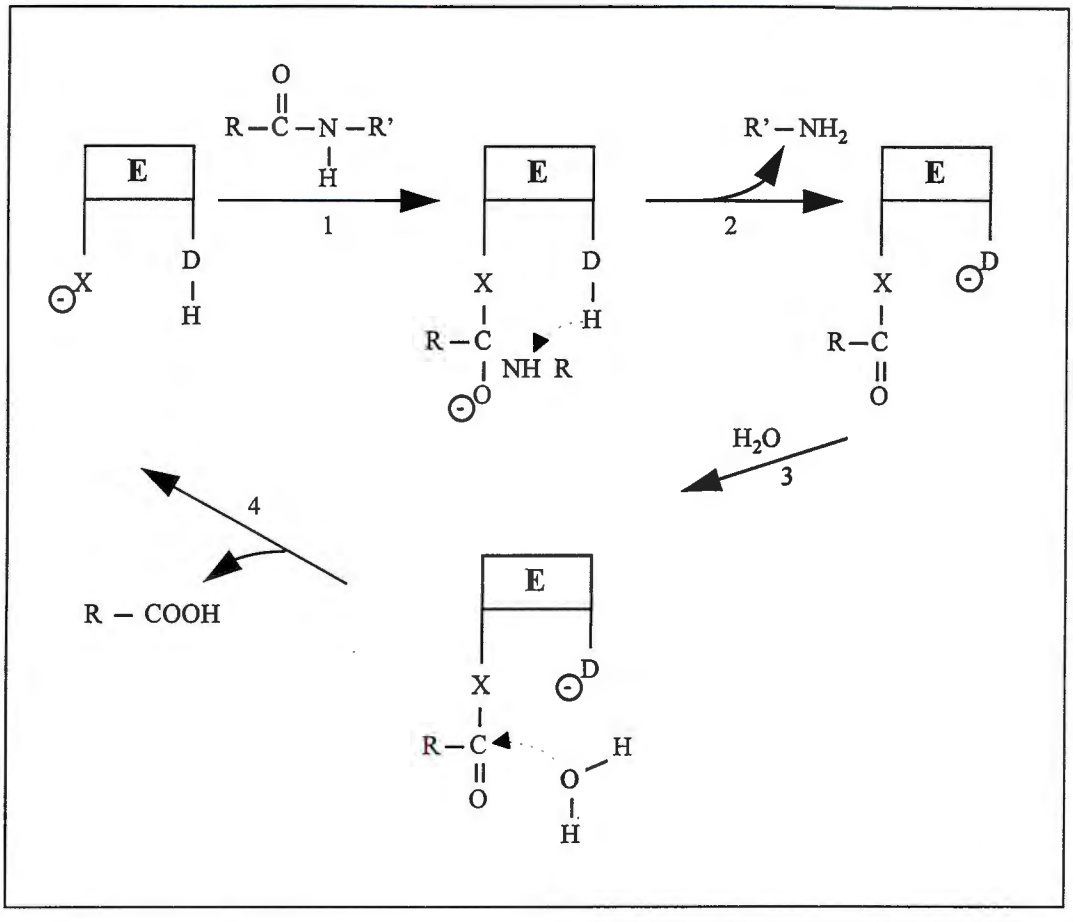


Figure 2.3: *A generalised scheme for proteolytic cleavage. X is the essential nucleophile, while DH is a proton donor that is able to protonate the leaving group (from Baker & Drenth, 1986).*

Formation of an enzyme-substrate complex occurs (the Michaelis complex; step 1). The nucleophilic group of the enzyme then attacks the carbon atom of the peptide bond to be cleaved, resulting in the formation of a tetrahedral intermediate that may be stabilised by interaction of the oxyanion ($-C-O^-$) with an appropriate group on the enzyme. Bond cleavage is then completed by proton donation (by the donor group D) to the leaving group. This leaving group can then diffuse away (step 2), leaving the acylated enzyme which in this case is a thioester. This can then be hydrolysed by nucleophilic attack of OH^- (from a water molecule; step 3). Release of the product results in the regeneration of the free enzyme (step 4).

2.3 Structure of Cystatins

The structures of a number of cystatins have been elucidated by X-ray crystallography. The first of these was that of chicken cystatin (Bode *et al.*, 1988). The crystal structure of stefin B in complex with papain (Stubbs *et al.*, 1990) and the NMR structure of free stefin A in solution (Martin *et al.*, 1995; Tate *et al.*, 1995) were subsequently identified. The solution structures of natural phosphorylated (Dieckmann *et al.*, 1993) and a recombinant variant Arg-Glu-Phe[Met]Ile29,leu 89 of chicken cystatin have also been solved by NMR analysis (Auerswald *et al.*, 1994). More recently, a preliminary X-ray crystallographic analysis of human salivary cystatin has been carried out (Ramasubbu *et al.*, 1996).

These and kinetic studies have shown that cystatins and stefins are tight binding, reversible inhibitors. Their basic structure consists of a long central α-helix, wrapped in a five-stranded anti-parallel β-pleated sheet, with a subsidiary helix in the case of cystatins (family II cystatins) and a strand in the case of the stefins (family I cystatins). There is a highly conserved "QXVXG" region (first hairpin loop) in all cystatins which is flanked by the projecting amino terminus and a second hairpin loop also containing conserved residues. A schematic representation of the chicken cystatin structure bound to papain is shown in Figure 2.4.

The wedge-shaped hydrophobic edge of the inhibitor is complementary in shape to the active-site cleft of papain. Docking experiments with papain (Bode *et al.*, 1988) have shown that both the cystatin loops interact with areas that lie adjacent to papain's catalytic residues. The amino terminus loops over the catalytic Cys^{25} residue of papain and interacts with the S1 and S2 subsites of papain. The first hairpin loop is believed to confer most of the stability to the complex, whereas interactions of the N-terminus with papain appear to strengthen complexes with the enzyme (Bode *et al.*, 1990; Turk & Bode, 1991; Bode & Huber, 1992). However, as will be reviewed later, the importance of the interactions of the various cystatin binding regions may be dependent on the particular cysteine proteinase to which the inhibitor is bound.

2.4 Primary Structure Sequence Similarities between the Cystatins

The cystatin superfamily shows strong homology amongst its members. Sequence alignments carried out, (Figure 2.5) show that one of the most highly conserved residues of fragments is the QVVAG region or variants thereof (QLVSG in chicken cystatin, Figures 2.4 and 2.5). The strong homology amongst the members in this region suggests that it plays an important part in the functioning of the inhibitors. Other common features include a conserved Gly^{9} residue (chicken cystatin numbering) and a conserved Pro^{103} - Trp^{104} region (chicken cystatin numbering). These Pro and Trp resides are also conserved in the third domain of the kininogens (Kellermann *et al.*, 1986) which may suggest an importance of this region for function.

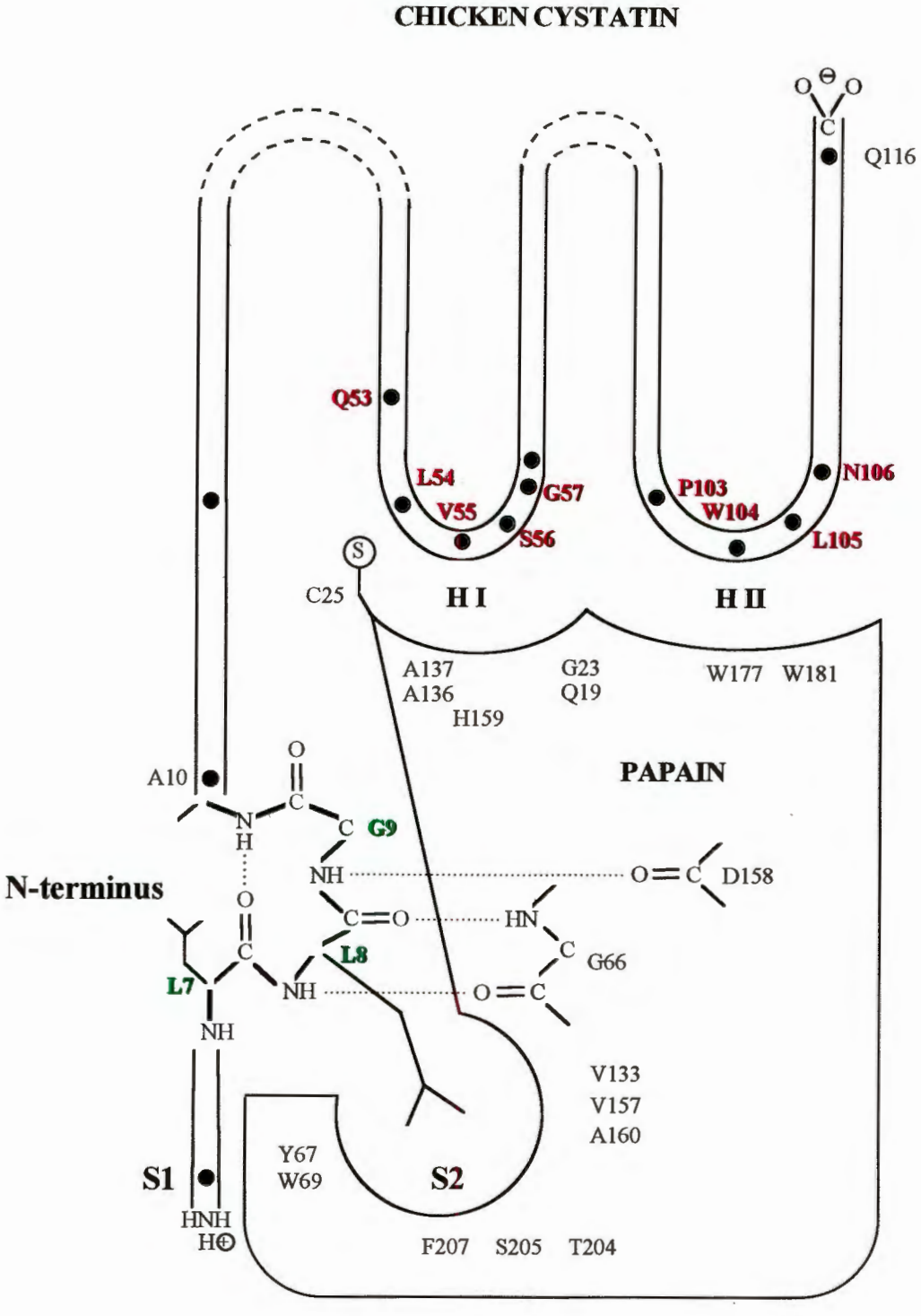


Figure 2.4: *A diagrammatic model showing the interaction of chicken cystatin with papain (adapted from Bode et al., 1988). The residues of chicken cystatin which are important for interaction with papain are highlighted in colour. H I, first hairpin loop; H II, second hairpin loop.*

Figure 2.5: *Primary amino acide sequence alignment of a number of cystatins.*

a: human stefin A (Machleidt etal., 1983); b: rat stefin a (Takio et al., 1984); c: human stefin B (Ritonja et al., 1995)

d: human cystatin C (Grubb andLofberg, 1982); e: human cystatin S (Isemura etal., 1984); f: chicken cystatin (Schwabe et al., 1984)

g: human kininogen domain 1; h: human kininogen domain 2; i: human kininogen domain 3 (Salvesen etal., 1986).

The experimentally determined structures of the cystatins (Bode *et al.*, 1988; Stubbs *et al.*, 1990) suggest that the conserved Gly^{9} residue may confer some flexibility to the amino terminus. Despite the conserved regions in the cystatins, there are also regions of high variability. These are often the site of modifications, for example, phosphorylation found on Ser^{80} in chicken cystatin (Dieckmann *et al.*, 1993) and glycosylation found in the kininogen domains (Kellermann *et al.*, 1986).

As mentioned previously (section 2.3), cystatins (family II) have a second helix (amino acid residues 69 - 91 in chicken cystatin), while stefins have a strand. Other differences between these two families are as follows:

- the stefins do not contain the two disulphide bonds found in the family II cystatins.
- the carboxy-termini of the stefins (family I cystatins) often extend beyond those of the cystatins in family II by 9 residues.
- compared to the family I cystatins, the family II cystatins have much longer N-terminal regions. The conserved glycine in these cystatins is often preceded by some 10 amino acid residues.

2.5 The Functions of the Cystatin Binding Loops

A number of substitution and deletion experiments have been carried out in order to establish structure/function relationships between the different regions of the cystatins and their inhibitory activity. These have identified the amino-terminus, the first hairpin loop and the second hairpin loop as the most important regions.

2.5.1 Importance of the Amino-Terminus

Although it has been estimated that 36 - 40% of the binding energy in the formation of the complex between chicken cystatin and papain is contributed by the N-terminal residues (Martin *et al.*, 1994; Lindahl *et al.*, 1992), the importance of the N-terminus in the binding of members of the cystatin superfamily to their cognate enzymes remains a controversial issue that has yet to be resolved.

A survey of the literature has shown that the family I and II cystatins interact differently with their cognate enzymes so that the relative importance of the different binding region varies. For this reason, the different families are affected to different degrees by modifications of their amino terminus and a separate discussion of the family I and II cystatins is indicated.

1. Importance of the N-terminus in Family I Cystatins

The best characterised representatives of this family are stefins A and B (Barrett *et al.*, 1986; Turk & Bode, 1991). Despite the fact that NMR analyses of the structure of stefin A (Martin *et al.*, 1995; Tate *et al.*, 1995) have shown a number of similarities in the overall pattern, amino acid substitution and deletion experiments in the N-terminus of stefins A and B have very different effects on the inhibitory activity of these proteins.

In experiments with stefin A, deletion of the first eight residues (Shibuya *et al.*, 1995) or the first six residues (Pol *et al.*, 1995) from the N-terminus led to complete inactivation of the inhibitor when tested with papain or cathepsin B. Truncation of the first two residues caused no change in activity against papain, whilst deletion of Pro^3 led to inactivation of the inhibitor (Shibuya *et al.*, 1995). Substitution experiments, on the other hand, in which Pro^3 was replaced with small amino acids such as leucine, had no effect on inhibitor activity. This suggests that, while an amino acid residue in this position is important for the functioning of the N-terminus, the proline residue is not required for turn formation or main chain rotation but is important for the contacts it provides with the proteinase (Shibuya *et al.*, 1995). Studies carried out on a truncated N-terminus of cystatin-α from rat epidermis showed similar results to those found with human stefin A (Takeda *et al.*, 1985). This is not surprising, since these two cystatins are highly related (Barrett *et al.*, 1986).

Shibuya *et al.* (1995) have used 2D ^{15}N-^{1}H HSQC NMR spectra to study further the effects of amino-terminal truncation on stefin A. Four-residue truncation of the amino-terminus was found to cause changes in the conformation of the V^{47} and V^{48} residues that make up part of the conserved QVVAG region of the first hairpin loop. The authors concluded therefore, that N-terminal truncation may cause a conformational change in the first loop which in turn, may also cause a conformational change in the second binding loop. The resulting lack of sufficient contact between the hairpin loops and the proteinase would severely affect inhibitory activity. The N-terminus of stefin A must therefore contribute to both the contact with the proteinase and maintain the conformational integrity of the tripartite wedge of the inhibitor.

Elongation of the N-terminal region of stefin A had only a minor effect on the affinity of the inhibitor for papain and cathepsin H (Pol *et al.*, 1995). These findings are contrary to results obtained with stefin B where elongation of the N-terminus resulted in weaker binding to cathepsin H (Jerala *et al.*, 1994). Pol *et al.* (1995) suggested that this was due to the fact that the N-terminal regions of the two stefins interact with cathepsin H in different ways.

Oryzacystatin differed from stefin A in that the amino-terminal truncation of the first 21 residues had little or no effect on the inhibition of papain (Abe *et al.*, 1988). Similarly, the removal of the N-terminal residues from stefin B resulted in little change in inhibitory activity (Jerala *et al.* 1990). However, Stubbs *et al.* (1990) have shown from X-ray crystallographic data of stefin B that Met^2 and Ser^3 (Ser^3 had been substituted for Cys^8 in

the crystal structure protein because of possible reactivity of the Cys^3 side chains), are responsible for most of the contacts with the proteinase in the N-terminal region.

Thiele *et al.* (1990) made several truncated forms of stefin B. Removal of the first two amino acids and the substitution of Cys^3 with a hydrophobic phenylalanine did not significantly change the inhibition constants. The introduction of a negatively charged glutamic acid in place of the Cys^3, on the other hand, increased the K_i for both papain and cathepsin B (Thiele *et al.*, 1990). The glutamic acid obviously interfered electrostatically with both the active-site cleft of the proteinase. These authors suggested, therefore, that it is electrostatic interactions which are important in the contacts of stefin B at the amino-terminus. It thus appears that only certain residues can be accommodated at the contact area with the proteinase.

Since the effect of N-terminal truncation of the different stefins are inconsistent, many believe that the main contact area of stefins with their cognate enzymes comes from the first (QVVAG) binding loop with the N-terminal 'trunk' helping to pilot the inhibitor into the active site cleft following binding to the enzyme (Stubbs *et al.*, 1990; Thiele *et al.*, 1990). This would imply that the N-terminus is flexible and Martin *et al.* (1995) have indeed shown, by NMR, that the N-terminus of free stefin A does not populate any single conformation to a significant degree but changes its conformation on binding. Jerala (1992) also suggests that the conformation of the N-terminal residues of free and bound stefins differ. It seems, therefore, that the N-terminal 'trunk' of stefin A bends and aligns itself to the groove of the enzyme upon binding to the proteinase so helping to 'anchor' the inhibitor. The contribution of the amino-terminus of stefin B to binding in this region is, however, negligible (Jerala, 1992). This may explain why truncation of the N-terminal part of stefin B does not weaken binding to the enzyme.

Tate *et al.* (1995) have used the differences between the inhibitors to propose a basis for the classification of the cystatins. They argue that cystatins can essentially be divided into two groups. Those in group I include cystatins such as oryzacystatin and stefin B that do not require the N- and/or C-terminal segments to maintain the active conformation of the first hairpin loop. Those in group II, on the other hand, such as stefin A, do require both the N- and C-termini. In the case of stefin A, for example, truncation of a few residues after Pro^3 from the N-terminus would cause the N-terminal peptide segment to pull apart from the C-terminal region thus causing a change in the conformation of the first hairpin loop (Tate *et al.*, 1995). In the case of stefin B (Jerala *et al.*, 1991) and oryzacystatin 1 (Abe *et al.*, 1988) where the N- and C-terminal interactions are not required to maintain their active conformation, these deletions are without effect. Further experiments are needed to determine the validity of this theory, especially since there is now conflicting evidence on the effect of N-terminal truncation of oryzacystatin (Abe *et al.*, 1988; Urwin *et al.*, 1995b).

2. Importance of the N-terminus in Family II Cystatins

The function of the N-terminus of family II cystatins has, as in the case of the family I cystatins, been elucidated by inhibition experiments with N-terminally truncated forms of the inhibitor. Detailed analyses of the N-terminal segments of chicken cystatin and human cystatin C indicate that, when family II cystatins inhibit their target proteinases, residues near the N-terminus bind in a substrate-like manner to one or more of the unprimed substrate-binding subsites in the enzyme. The experimental results indicate that the specificity of different cystatins for the different cysteine proteinases is partially determined by the residues present in this area of the enzyme (Björk *et al.*, 1994).

Chicken cystatin has been shown to form complexes with cysteine proteinases even after their catalytic sites have been inactivated by bulky active-site-directed reagents (Anastasi *et al.*, 1983). Keilová & Tomášek (1974) showed that a complex could be formed between chicken cystatin and cathepsin B even if the free SH-group of the catalytic site was blocked with mercury. These findings raise the possibility that, in the complex, the surface of the proteinase and inhibitor are not closely associated in the region of Cys^{25} (papain numbering).

Truncation of chicken cystatin from its full length (N-terminal serine) to the truncated form beginning with Leu^{7}, causes no change in the inhibitory activity (Machleidt *et al.*, 1989). Truncation beyond Leu^{8}, leads to an approximately 5000-fold lower affinity for papain (Machleidt *et al.*, 1989). Truncation of chicken cystatin beyond Gly^{9}, on the other hand, caused no further loss of inhibitory activity. Thus, Gly^{9} and residues 1 to 6 account for a substantially smaller portion of the binding energy than do the two leucine residues (Auerswald *et al.*, 1995).

Bode *et al.* (1988) claimed that the functional importance of Leu^{8} stemmed from its interaction with the S_2 subsite of papain. Lindahl *et al.* (1992) showed that this residue only contributes approximately one-third of the unitary free energy of binding of the N-terminal region of cystatin to papain. Leu^{7}, on the other hand, contributes a larger portion of the binding energy (Lindahl *et al.*, 1992). The general consensus (Machleidt *et al.*, 1989; Abrahamson *et al.*, 1991; Lindahl *et al.*, 1992; Björk *et al.*, 1994) is therefore, that Leu^{7} and Leu^{8} are important for interactions of the N-terminus, while residues N-terminal to Leu^{7} stabilise the interactions between the enzyme and the inhibitor. These results are in accord with the X-ray studies on chicken cystatin which show that the N-terminus is flexible. Removal of this region shows no alteration in the 3-D conformation of the rest of the protein (Bode *et al.*, 1990) and the loss of affinity following truncation is due to the loss of contacts in the N-terminal region. The N-terminus thus contributes to stability by keeping the inhibitor anchored to the proteinase and thereby reducing the rate of dissociation (Björk *et al.*, 1994).

The function of the N-terminal segment of chicken cystatin is not entirely straightforward in the sense that the effect of its removal depends upon the structure of the active-site region of the proteinase with which it is interacting. Thus the reduced affinity after N-terminal truncation of chicken cystatin is due to either a

decreased association rate constant or an increased dissociation rate constant, or both, depending on the target enzyme (Björk *et al.*, 1994). For example, the decreased affinity of N-terminally truncated chicken cystatin, is due to an increased dissociation rate constant in the case of papain, to both a decreased association constant and increased dissociation rate constant for actinidin, and finally, to a decreased association constant for cathepsin B. The differences between papain and actinidin may be due to the fact that these enzymes differ in their S_2 subsites (Drenth *et al.*, 1971) and may thus bind the N-terminal regions of cystatins differently. In the case of cathepsin B, the presence of the occluding loop causes collision around Trp^{104} in chicken cystatin which prevents close contact of this residue with the active site of cathepsin B (Björk *et al.*, 1994).

Truncations and substitutions in the N-terminus of cystatin C have also been extensively studied (Abrahamson *et al.*, 1991; Hall *et al.*, 1993; Lindahl *et al.*, 1994; Hall *et al.*, 1995). As with chicken cystatin, it appears that the N-terminus of this inhibitor makes an important contribution towards its inhibitory activity. For example, truncated forms of human cystatin C starting with Leu^9 and Val^{10} before the conserved Gly^{11} (Gly^9 in chicken cystatin) has the same affinity for papain as the full length sequence. However, a truncated form starting with Gly^{12} is a 1000-fold weaker inhibitor (Brzin, 1984).

The glycine residue (Gly^{11} in cystatin C and Gly^9 in chicken cystatin) is highly conserved in the family II cystatins and in the other cystatin families. Experiments that have replaced the evolutionary conserved Gly^{11} residue with different amino acids (Hall *et al.*, 1993), even those with small side-chains such as a serine or alanine have shown decreased affinities for papain, cathepsin B and ficin (Lindahl *et al.*, 1994). Substitutions with bulky side-chains had the same effect. This further confirms the hypothesis that the conserved glycine residue, by conferring flexibility on the N-terminal region of the inhibitor, allows maximum interaction between this segment and the substrate binding pockets of the target enzyme. The glycine residue that is similarly conserved in family I is also responsible for flexibility of these cystatins.

Despite the findings that truncation of the N-terminus affects the binding of chicken cystatin to the endopeptidases, Nicklin & Barrett (1984) showed that the N-terminus of chicken cystatin makes no appreciable contribution to the inhibition of the exopeptidase, dipeptidyl aminopeptidase I (cathepsin C). Thus, the N-terminal contact region of cystatins seems to be indispensable for effective inhibition of cysteine endoproteinases, but not for the inhibition of the exopeptidase cathepsin C.

In conclusion, it appears that in the family II cystatins, the conserved glycine residue, as well as one or two residues preceding this glycine, are responsible for the tighter binding of cystatins to papain. The amino terminus may be important in family II cystatins to compensate for less favourable contacts made in the hairpin loops such as the QLVSG sequence in chicken cystatin, for example. Family I cystatins (with the exception of stefin A) can be truncated to only one residue before the glycine residue without loss of activity (Machleidt *et al.*, 1989). Furthermore, in the case of both the cystatins and the stefins, the amino terminal residues have distinct but differing effects on the binding to papain, actinidin, ficin, and the cathepsins B, H and L.

Cathepsin L is perhaps the most sensitive proteinase to changes in the N-terminal region of cystatins (Auerswald *et al.*, 1995). It can thus be concluded that the importance of the N-terminus may depend on the cysteine proteinase and thus the amino acid residues involved in the contacts with the inhibitor.

Table 2.1 gives a summary of some of the mutation experiments reported in the literature, and the K_i obtained with different cysteine proteinases as a result of these mutations.

2.5.2 Variants of the First Hairpin Loop

X-ray analysis of chicken cystatin has revealed three contact regions between the inhibitor and papain. The most important of these is the first hairpin loop containing the QXVXG region that has been shown, by sequence alignments, to be highly conserved and therefore involved in the inhibitory activity of the protein (Barrett *et al.*, 1986; Abrahamson *et al.*, 1987; Turk & Bode, 1991). Experimental results using site-directed mutagenesis have for the most part, established the validity of this conclusion.

Studies on papain have shown that this enzyme does not have a binding pocket that is selective and specific for the QXVXG sequence of the inhibitor as there is no specific interface for this arrangement of amino acids (Yamamoto *et al.*, 1992). It is rather the hairpin loop structure that this sequence contributes to the tertiary structure of the inhibitor that is important. Most of the binding energy comes from the tight interaction of the first hairpin (QVVAG) loop with the S_1-pocket of the enzyme (Thiele *et al.*, 1990).

Of the amino acids in the first hairpin loop, the most highly conserved are Gln^{53} (which forms a H-bond with Gly^{9} of the N-terminus), Val^{55} and Gly^{57} (chicken cystatin numbering, see Figure 2.5). Mutations in any of these positions results in a reduced affinity for the cysteine proteinases by up to three orders of magnitude (Turk & Bode, 1991). The extent to which substitutions in any of these regions affects inhibition is very much dependent on the various cysteine proteinases.

Nikawa *et al.* (1989) probed exchanges in this conserved region by replacing the QVVAG of cystatin A with KVVAG or QVTAG. They found that these replacements did not significantly affect the inhibitory activity of cystatin A towards papain or the cathepsins B, H and L. In the case of Oryzacystatin I, on the other hand, substitutions of Pro and Leu for Q^{53} and G^{57} respectively or Asp for the second Valine (V^{55}) resulted in an increase in the K_i (Abe *et al.*, 1991). It has been suggested (Björk *et al.*, 1995) that the increased K_i from variations in the QVVAG region are mainly due to an increased dissociation of the complexes.

Table 2.1: Summary of mutation experiments reported in the literature and the effect on the K_i

Sequence	* K_i (nM)				
	Papain	Cathepsin H	Cathepsin L	Cathepsin B	Reference
Family I Cystatins					
Human Stefin A					
MIPGGLSEAKP...	1.2	ND	ND	17.7	Shibuya *et al.*, 1995
IPGGLSEAKP...	2.45	ND	ND	20.5	Shibuya *et al.*, 1995
GLSEAKP...	414	ND	ND	ND	Shibuya *et al.*, 1995
LGGLSEAKP...	6.35	ND	ND	293	Shibuya *et al.*, 1995
Human Stefin B					
MMCGAPSATQ...	0.12	ND	ND	73	Barrett *et al.*, 1986
LLCGAPSATQ...	0.004	ND	ND	113	Thiele *et al.*, 1990
LLFGAPSATQ...	0.005	ND	ND	273	Thiele *et al.*, 1990
LLEGAPSATQ...	0.120	ND	ND	881	Thiele *et al.*, 1990
SATQ...	0.008	ND	ND	ND	Thiele *et al.*, 1990
CGAPSATQ...	0.010	ND	ND	ND	Thiele *et al.*, 1990
Family II Cystatins					
Chicken Cystatin					
SEDRSRLLGAPVPVDEND...	0.006	ND	ND	4	Björk *et al.*, 1994
AEFMVPVDEND...	166	ND	1.6	≥ 20 000	Auerswald *et al.*, 1994
GAPVPDEND...	12	ND	ND	1800	Björk *et al.*, 1994
LLGAPVPDEND...	0.05	ND	ND	ND	Lindahl *et al.*, 1992
LGAPVPDEND...	0.12	ND	ND	ND	Lindahl *et al.*, 1992
GAPVPDEND...	33.9	ND	ND	ND	Machleidt *et al.*, 1989
APVPVDEND...	31.6	ND	ND	ND	Machleidt *et al.*, 1989
Cystatin C					
SSPGKPPRLVGGPMDASVEE.	0.0011	ND	ND	0.3	Björk *et al.*, 1994
SSPGKPPRLVGGPMDASVEE.	0.0011	ND	ND	0.28	Lindahl *et al.*, 1994
SSPGKPPRL**R**GGPMDASVEE.	0.13	ND	ND	4.2	Lindahl *et al.*, 1994
SSPGKPPRLV**A**GPMDASVEE.	ND	ND	ND	8.1	Björk *et al.*, 1995
SSPGKPPRLV**E**GPMDASVEE.	1.7	ND	ND	18000	Björk *et al.*, 1995
SSPGKPPRLV**W**GPMDASVEE.	0.8	ND	ND	9600	Björk *et al.*, 1995
GGPMDASVEE.	ND	ND	ND	1.3	Björk *et al.*, 1995
GGPMDASVEE.	ND	2.15	2.11	101	Abrahamson *et al.*, 1991
SSPGKPP**G**LVGGPMDASVEE.	ND	0.27	<0.01	1.9	Hall *et al.*, 1995
SSPGKPPR**G**VGGPMDASVEE.	ND	0.01	<0.01	46	Hall *et al.*, 1995
SSPGKPPRL**G**GGPMDASVEE.	ND	10	<0.01	500	Hall *et al.*, 1995
SSPGKPPRLV**A**GPMDASVEE.	0.024	ND	ND	4.4	Hall *et al.*, 1993
SSPGKPPRLV**S**GPMDASVEE.	0.077	ND	ND	14	Hall *et al.*, 1993
SSPGKPPRLV**W**GPMDASVEE.	2.4	ND	ND	>50	Hall *et al.*, 1993
SSPGKPPRLV**E**GPMDASVEE.	3.5	ND	ND	>50	Hall *et al.*, 1993
SSPGKPPRLV**R**GPMDASVEE.	3.8	ND	ND	>50	Hall *et al.*, 1993
GGPMDASVEE.	29	ND	ND	13	Hall *et al.*, 1993

** The above K_i determinations were carried out using Z-Phe-Arg-AMC as the substrate in the case of papain and the cathepsins B and L. The substrate Arg-AMC was used for cathepsin H.*

Amino acids shown in bold and underlined represent changes made by mutation experiments.

ND = not determined.

Peptide inhibitors that mimic the QVVAG loop of cystatin surfaces have been synthesised (Lalmanach *et al.*, 1993; Serveau *et al.*, 1994). These were shown to retain significant inhibitory activity towards papain. However, the inhibition by this QVVAG peptide was time dependent. Peptide backbone cleavage was found to occur at the A-G bond following long incubation with papain whereas the QVVAG sequence is not cleaved in the natural inhibitors. This may be due to the particular spatial conformation of this sequence within the natural cystatin structure (Lalmanach *et al.*, 1993; Serveau *et al.*, 1994).

It is clear that the first hairpin loop is extremely important in the binding of cystatins to their cysteine proteinases. This loop may contribute most of the free energy of binding.

2.5.3 The Second Binding Loop

In comparison with the N-terminus or the first binding loop, relatively little is known of the importance of the second hairpin loop for the stability of the molecule or its contribution to inhibitory activity. It is however, generally believed that all three regions must contribute to the "overall fit" of the inhibitor into the active site cleft of the enzyme.

The second binding loop, like the first is quite hydrophobic in nature. Pro^{103} and Trp^{104} (chicken cystatin numbering) are well conserved in the family II cystatins and in the third domain of the kininogens (Kellermann *et al.*, 1986) suggesting that they are functionally important. No counterpart to this conserved region has been found in stefin A. In fact, stefin A and rat cystatin α are devoid of any aromatic amino acids in their second binding loop, whereas most of the other cystatins have a his/trp residue on this loop (Bode *et al.*, 1988).

Support for the contribution of the second hairpin loop to the function of chicken cystatin was obtained by Nycander & Björk, (1990) who modified Trp^{104} with a 2-hydroxy-5-nitrobenzyl group and showed that the product had a 100,000-fold lower affinity for papain than the intact inhibitor.

Similar results were obtained for oryzacystatin I, where substitution of amino acids in the second binding region resulted in a significantly higher Ki value, indicating the importance of this region for the inhibition of proteinases (Urwin *et al.*, 1995). Complete loss of the second hairpin loop of oryzacystatin i.e. a large deletion at the carboxyl-terminus, also resulted in weaker inhibitory activity (Abe *et al.*, 1988).

Despite the importance of the second loop, mutations in this region have been shown to have a significantly lower effect on the stability of the molecule and on the affinity for proteinases than mutations in the N-terminus or the first hairpin loop (Auerswald *et al.*, 1992;. Auerswald *et al.*, 1995; Stubbs *et al.*, 1990). On the other hand, Trp^{104} of chicken cystatin and the N-terminus have been shown to be equally important for the inhibition of cathepsin S, whereas Trp^{106} is more important than the N-terminus in the case of cathepsin H and less important for cathepsin L (Hall *et al.*, 1995; Auerswald *et al.*, 1995).

These results further support the observation that modifications of the inhibitors have effects that differ depending on the enzyme that is used to test them. Binding to cathepsin B has been shown to be even more affected by insertion of a 'wrong' second hairpin loop than variations in the first loop (Auerswald *et al.*, 1995). This is due to steric collision with the occluding loop of cathepsin B.

2.5.4 Studies Involving 'Whole Cystatin Region' Substitutions

Relative to the substitution and deletion experiments, very little has been done with whole sequence replacement of cystatin binding areas. Those experiments which have been carried out have yielded interesting results. When, for example, stefin B was used as the master protein onto which different blocks similar either to cystatin C or kininogen were ligated (Jerala *et al.*, 1994), both amino termini were longer than that of the original stefin B and both substitutions introduced an ionic residue. The affinity and kinetic properties of these hybrid stefins was nevertheless found to be largely unaffected for papain and cathepsin L. These modifications did not improve inhibition of cathepsin B, which is only moderately inhibited by stefin B, and inhibition of cathepsin H was found to be weaker than that of the original stefin (K_i 0.14 nM and 0.9 nM compared to 0.02 mM for wild-type stefin B) and approached that of human cystatin C (0.28 nM) and the second domain of human kininogen (1.2 nM). Most differences that were observed could be ascribed to a decrease in association constants of both mutants, indicating that the rate of complex formation is altered by adding a charged group to the amino terminus (Jerala *et al.*, 1994).

Auerswald *et al.* (1996) substituted the N-terminus as well as the first and second hairpin loops of chicken cystatin for the corresponding regions of human kininogen domain 2 in a single and combined manner. The characteristic feature of this second kininogen domain is that it can inhibit calpain (Müller-Esterl *et al.*, 1986; Bradford *et al.*, 1993). Temporary inhibition of calpain was achieved with hybrids carrying the N-terminus of kininogen alone and with those carrying both the N-terminus and the first hairpin loop. Hybrids of the second loop were weaker inhibitors of calpain. This suggested, therefore, that the N-terminus and first hairpin loop play the most important role in the calpain inhibition by human kininogen domain 2.

In this project, the N-terminus of oryzacystatin was substituted for amino acid residues from the N-terminus of chicken cystatin. The effects of the substitutions on oryzacystatin inhibition of papain have been investigated.

2.5.5 General Conclusions

It can perhaps be concluded that both hairpin loops, as well as the N-terminal trunk of the inhibitor complex, contribute to complex formation. The first hairpin loop possibly maintains its conformation so as to fit properly into the active site cleft of the enzyme. The N-terminus and the second hairpin loop, which often have weaker contacts with the proteinase, adopt a conformation which stabilises the interactions between the first binding loop and the enzyme (Tate *et al.*, 1995). In fact, the importance of 'correct conformation' of the hairpin loops for binding has been shown (Machleidt *et al.*, 1995). In these experiments, papain activity was only transiently inhibited and reappeared with time. This was found to be due to a selective cleavage of the Gly^{9}- Ala^{10} bond in the N-terminal region of chicken cystatin (Machleidt *et al.*, 1995) and the Gly^{11}- Gly^{12} bond in cystatin C variants (Grubb *et al.*, 1990). This resulted in truncated inhibitors with lower affinities. These results suggest therefore, that distorted contacts of one of the hairpin loops affects the binding of the N-terminal contact area. Covalent interaction of the Gly-Ala or Gly-Gly bond with the active site cysteine residue of papain can thus occur, and the bond is cleaved in a substrate-like manner (Machleidt *et al.*, 1995). This is the so called 'trunk model' shown in Figure 2.6 (from Machleidt *et al.*, 1995).

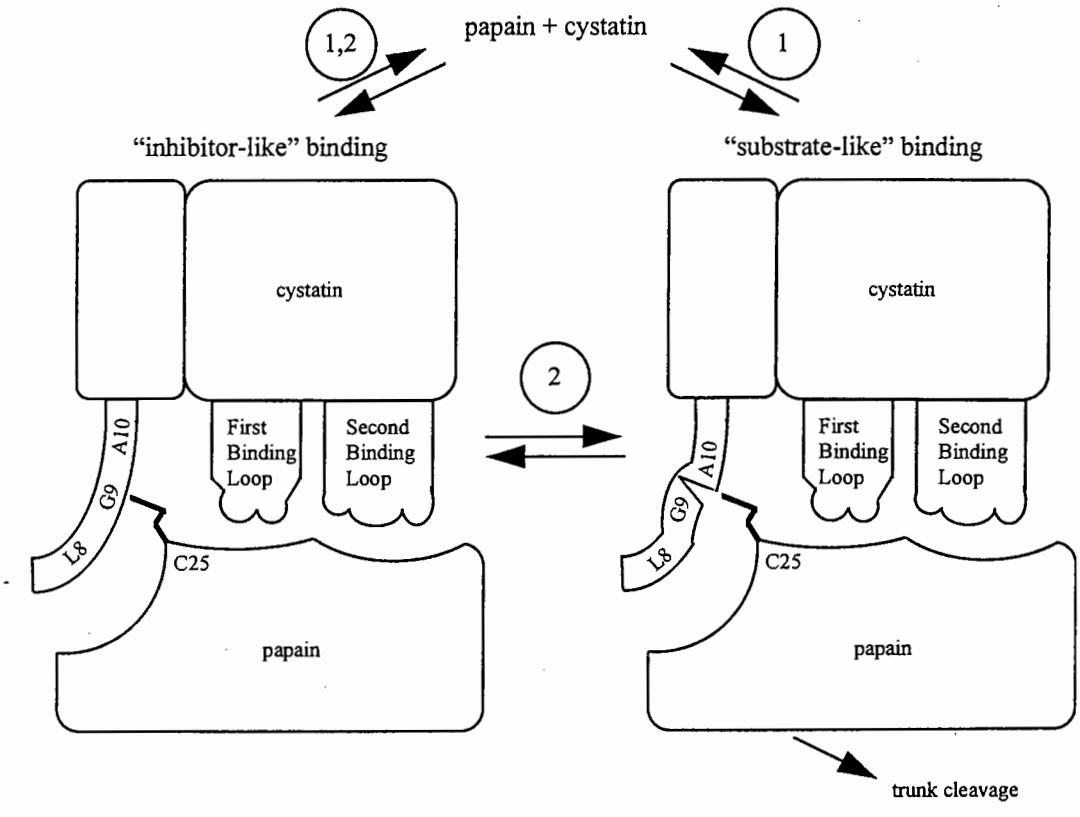


Figure 2.6: *The 'elephant-trunk' model showing cystatin binding to papain. Cleavage of the Gly^{9}-Ala^{10} bond is thought to occur via one of two possible mechanisms. (1) competition of inhibitor-like and substrate-like complexes for free papain; or (2) rearrangement of the inhibitor-like conformation by 'slippage' in the hydrophobic environment (from Machleidt et al., 1995)*

PART 2: Molecular Modelling

2.6 Aims of the Molecular Modelling

From the literature, it is clear that the regions interacting in the complex of cystatins with their cysteine proteinases have been identified, and site-directed mutagenesis can be employed to probe the importance of each residue or region for inhibition. The aim of the molecular modelling was two-fold:

- To determine a possible structure for oryzacystatin.
- By studying the structures of chicken cystatin and the generated OC I model, determine which sections of the amino terminus of OC I could be replaced by those in chicken cystatin.
- Having decided on the areas to substitute, determine how this new 'hybrid' inhibitor might interact with papain.

Any improvement in the inhibitory capabilities of this hybrid inhibitor would be beneficial, since a long term goal of this project is to utilise protein engineering methodology for the development of a proteinase inhibitor which has strong inhibitory activity against physiologically relevant cysteine proteinases such as those found in the insect order Coleoptera.

2.7. Reasons for using Papain in the Modelling Studies

Papain was used as the model proteinase in the modelling studies for the following reasons:

- It is well-characterised (Polgar, 1973; Angelides & Fink, 1979; Berti *et al.*, 1990; Matsumoto *et al.*, 1994) and is often used in the study of protein-inhibitor interactions with cystatins.
- It is available in pure form and in large quantities. This would be useful for the kinetic studies performed in this project.
- Its X-ray structure has been elucidated and its co-ordinates are readily available for use in computer modelling and docking studies with cystatins.

Since the long-term goal of this project is to look at insect biocontrol, it is perhaps more appropriate to use insect cysteine proteinases in the modelling studies, specifically those from the Coleoptera. However, very little work has been done on insect proteinases, and most have not been characterised. Thus, since the foremost aim of this project was to investigate the effects of substitutions in the amino terminus, papain would be sufficient as a model proteinase.

2.8 Oryzacystatin Secondary Structure Predictions

Although coordinates for the X-ray crystallographic structure of oryzacystatin are not available, the tertiary structures of a number of cystatins are known and from these and sequence alignments it appears that all cystatins have a similar basic structure. Given the high degree of homology seen in the primary sequence alignments (Figure 2.5), one may assume that oryzacystatin I (OC I) should form a very similar structure to that of the other cystatins and hence base prediction on this assumption.

```
                1             2         3         4         5         6
      A12345678901   A23456789012345678901234567890123456789012345678901AB2345
OC I: MSSDGGPVLGGV---EPVGNENDLHLVDLARFAVTEHNKKANSLL--EFEKLYSVKQQVVAGTLYYFTIEVK

          ccccc----ccctttaaaaaaaaaaaaaccccββββββββββββββββββttttββββββββ——ββββ
CC:    SEDRSRLLGAP----VPVDENDEGLQRALQFAMAEYNRASNDKYSSRVVRVISAKRQLVSGIKYI--LQVE
SB:         MMCGAP----SATQPATAETQNIADQVRSQLEEKTNKKF--PVFKAVSFKSQVVAGTNYF--IKVH
          ttttc----ccccccaaaaaaaaaaaaaaaaaaaaaattt--βββββββββββttttttββββ—ββββ

           7         8         9         10          11
      6789012345678901234567890123456789012A345A6789012345ABCDEFGHIJK
OC I: EGD----------------------AKKLYEAKVWEK-PWMDFKELQEFKPVDASANA

      βββββββttaaaaaaaaaaaaccccc-βββββββββ-βββββ-ttt-ttββββββββ-β
CC:   IGRTTCPKSSGDLQSCEFNDEPEMAKYTTCTF-VVYSI-PWL-NQIKLLESKC-Q
SB:   VGD----------------------EDFVHL-RVFQSLPHENKPLTLSNYQTNKAKHDELTYF--
      βtt----------------------tββββββ-βββββββttttββββββββββββttttββββ—
```

Figure 2.7: *Primary sequence alignments of oryzacystatin I (OC I) with chicken cystatin (CC) and stefin B (SB) (numbering is as for chicken cystatin, extra amino acids in the other cystatin structures are represented using letters from the alphabet; adapted from Turk & Bode, 1991). The predicted secondary structures for chicken cystatin (Bode et al., 1988) and stefin B (Stubbs et al., 1990) are shown above and below the sequences respectively.*

Key:

α = α-*helix* *t* = β–*turn*

β = β-*sheet* *c* = *random coil*

The Chou and Fasman algorithm (Chou & Fasman, 1978) was used to predict local structural elements, such as α-helices, β-sheets and β-turns and this revealed a possible secondary structure similar to that of other cystatins (Bode *et al.*, 1988; Stubbs *et al.*, 1990). No α-helices or β–sheets were predicted for the carboxy or

amino-termini. An almost equal number of α-helix former and helix breakers were shown to occur in the first 18 amino acids. This would make up the N-terminal region of OC I and suggests that this region has no definite α-helix or β-sheet structure. Residues Asp^{18} to Lys^{34} (numbering as in Figure 2.7, above) showed strong α-helix formation, which is similar to results shown for stefin B and chicken cystatin. The last 60 residues showed a high probability for β-sheet formation, with some α-helix.

An alternative method for modelling OC I was to consider the known structures of chicken cystatin and stefin B (Figure 2.7) and together with their primary sequence alignments (as reported by Turk & Bode, 1991) with each other and OC I, predict a secondary structure for OC I.

Thus, from the above sequence alignments, patterns of homologous residues are evident amongst the sequences which can be used to infer similar secondary structure. Figure 2.8 shows a predicted secondary structure for OC I as deduced from primary sequence alignments and Chou-Fasman secondary structure predictions.

```
                1             2         3         4         5         6
      A12345678901   A2345678901234567890123456789012345678901234567890AB2345
OC I: MSSDGGPVLGGV---EPVGNENDLHLVDLARFAVTEHNKKANSLL--EFEKLYSVKQQVVAGTLYYFTIEVK
      ccccccctttc---ccccαααααααααααααααttttttcc--βββββββββββtttttβββββββββββ

                                      10          11
      678                        234567A89012A345A6789012345ABCDEF
OC I: EGD------------------------AKKLYEAKVWEK-PWMDFKELQEFKPVDASANA
      βtt------------------------βββββββββββtt-ttttβββββββββββββttttt
```

Figure 2.8: *Secondary structure prediction for OC I (numbering as for chicken cystatin, extra amino acids in the OC I structure are represented using letters from the alphabet).*

Key:

α = α-*helix* *t* = β–*turn*

β = β-*sheet* *c* = *random coil*

A number of observations can be made from these primary and secondary structure predictions. For example:

- The QVVAG region is common to both stefin B and oryzacystatin, while chicken cystatin has a similar sequence in the form of the QLVSG region. These sequences have been shown by X-ray crystallographic data to be involved in the formation of a hairpin loop (Bode *et al.*, 1988; Stubbs *et al.*, 1990) and so this corresponding region in OC I would be expected to take on a similar conformation. Furthermore, the importance of this region in the inhibitory activity of OC1 has been demonstrated (Abe *et al.*, 1991). Thus, from the secondary structure predictions, it seems likely that oryzacystatin forms a similar wedge in the binding region as chicken cystatin and stefin B.

- From the alignment, it can be seen that the PWMD sequence of oryzacystatin is equivalent to the PHEN sequence in stefin B and the PWL sequence in chicken cystatin. The tryptophan residue is thus conserved and this loop in oryzacystatin was presumed to form a similar conformation to that of the other two cystatins.

- There are regions of homology in the amino-terminus with respect to the conserved glycine residue.

2.9 A Tertiary Structure Model of Oryzacystatin I

To construct a model of the natural OC I inhibitor, it was decided to use one of the known cystatin structures (available in the Brookhaven data bank) to serve as a 'backbone model'. The question of which cystatin should be used for this purpose then arose and the following points were considered in making this choice:

- Firstly, although OC I is a member of family V in the cystatin superfamily, it has a high degree of sequence homology with the family I cystatins and indeed, Turk & Bode (1991) have suggested that it could be considered a member of this family. The only member of family I whose structure was available at the time of modelling was stefin B (Stubbs *et al.*, 1990). This structure (available in the Brookhaven data bank) was therefore used rather than chicken cystatin (Bode *et al.*, 1988) which belongs to the family II cystatins.

- Secondly, the coordinates for stefin B represent the inhibitor in the bound form (i.e. complexed to papain). I thought this was more informative in terms of the contacts the amino terminus makes with papain. It has been suggested (Martin *et al.*, 1995), that owing to the flexibility of the amino terminus, changes in the position of this region occur on binding.

In constructing a model of OC I, greater emphasis was placed on the three binding regions of this cystatin, since these are the main areas which interact with the proteinase. Therefore, I assumed that any insertions or deletions needed to change the stefin B structure into the OC I structure could be accommodated in the loop

region joining helix I to strand B (Stubbs *et al.*, 1990), or in the open β-hairpin loop region from Val^{66} - Asp^{93} (stefin B numbering; Stubbs *et al.*, 1990). This is however, discussed more fully in section 2.11.

Alignment of the sequences of oryzacystatin I and stefin B (Figure 2.9) showed two obvious differences between the two structures:

- the OC I amino terminus extends beyond the stefin B amino terminus by 6 residues.
- the stefin B carboxy terminus extends beyond the OC I C-terminus by 5 residues.

In the case of OC I, the shorter C-terminus is compensated for by a longer N-terminus – a feature that should allow the two termini to come into contact with each other when the inhibitor is in the unbound form as was shown with stefin A (Tate *et al.*, 1995). In this conformation the N- and C-termini interact very closely with one another to form a more compact structure, making the N-terminus less susceptible to proteolytic attack. The authors do, however, caution that the relative orientation of the N-terminal segment in their structure was in sharp contrast to the random conformation of the N-terminal segments in the solution structure of cystatin A as deduced by Martin *et al.* (1995) at pH 5.5. With respect to OC I the question therefore arose as to how the extended N-terminus would lie when free in solution or when bound to the proteinase.

```
                1             2         3         4         5         6
      A12345678901   A2345678901234567890123456789012345678901234567890AB2345678
OC1:  MSSDGGPVLGGV---EPVGNENDLHLVDLARFAVTEHNKKANSLL--EFEKLYSVKQQVVAGTLYYFTIEVKEGD-
SB:         MMCGAP----SATQPATAETQNIADQVRSQLEEKTNKKF--PVFKAVSFKSQVVAGTNYF--IKVDVGE-


       9         10          11
      901234567A89012A345A6789012345ABCDEFGHIJK
OC1:  ---AKKLYEAKVWEK-PWMDFKELQEFKPVDASANA
SB:   ---EDFVHL-RVFQSLPHENKPLTLSNYQTNKAKHDELTYF
```

Figure 2.9: *Sequence alignment of oryzacystatin I (OC I) and stefin B (SB) (numbering as for chicken cystatin, extra amino acids in the OC I structure are represented using letters from the alphabet; adapted from Turk & Bode, 1991). The similarities between the sequences are highlighted in bold.*

As no model of a cystatin with as long an amino terminus was available to determine the interactions in this area, a possible model was inferred from a consideration of the interactions of the amino acid residues of the inhibitor and those of papain in this region. Biograf software (Biodesign, Pasedena, U. S. A) on an Evans and Sutherland computer was then used to complete the model as follows:

Primary sequence alignments were used as a point of departure to model the N-terminus of OC I onto that of stefin B. In the OC I model, the Gly^{9}-Gly^{10}-Val^{11} residues of OC I (numbering as in Figure 2.9) replaced the Gly^{4}-Ala^{5}-Pro^{6} sequence of stefin B (stefin B numbering; Stubbs *et al.*, 1990). The additional Glu^{11A} residue in the OC I sequence (Figure 2.9), was inserted after the Val^{12} (ex-Pro^{6}) residue in the stefin B structure, as there was no corresponding residue in stefin B.

The extra amino acids required for the construction of the OC I N-terminus were M^{A} to L^{8} (numbering as in Figure 2.9). These were then built onto the end of the N-terminus beyond Gly^{9} (ex stefin B Gly^{4}). This produced a long extended strand which projected off the papain surface while the other amino acids in this region remained fixed. A series of steps consisting of alternating energy minimisations and simple dynamics were then set up to minimise this structure. Both papain and the N-terminal 17 residues of OC I were included in this energy expression. This was done to minimise bad through-space contacts and to regularise the steric hindrances between the N-terminus and papain in this region.

Most of the movement during these energy minimisation steps was found to occur in the N-terminal 5 residues (A – 4). Residues Gly^{5} – Leu^{8} (numbering as in Figure 2.9) showed little movement, possibly due to steric hindrance from papain. This model suggests therefore, that on binding, the amino acids from Gly^{5} to Leu^{8} could lie along the surface of papain. Gly^{9} of OC I in this model kept its open type II turn following the energy minimisations and simple dynamics. This residue may, therefore, provide the anchorage for the contacts in the amino terminus as is seen with Gly^{9} of chicken cystatin (Bode *et al.*, 1988) and the Gly^{4} residues of stefin B (Stubbs *et al.*, 1990) and stefin A (Martin *et al.*, 1995).

To determine if it were plausible for the N-terminus of OC I to 'wrap' around papain rather than projecting out into solution, the carbonyl bond of the Gly^{4} residue was rotated. This moved the first four residues of OC I down close to papain (Figures 2.10 and 2.11). This new structure was then energy minimised with papain also being included in the energy expression.

The result of this rotation brought the first methionine residue, for example, into the vicinity of the Trp^{67}, Val^{57} and Phe^{207} residues of the enzyme. Whether or not these first few residues of OC I do make contact with papain could only be deduced from X-ray crystallography studies. It is clear, however, that any contacts made by these four amino acids with papain would not strengthen the binding in this area. This was in accord with Urwin *et al.* (1995b). These authors found that removal of the first 21 amino acids from the N-terminus of OC I did not affect its inhibitory activity towards papain. This is in contrast to chicken cystatin whose N-terminal Leu^{7} and Leu^{8} residues make contacts with similar papain residues in this region. However, if these last four residues projected out into solution, they could be susceptible to proteolytic attack. Unlike chicken cystatin, no natural N-terminally truncated forms of oryzacystatin have been isolated.

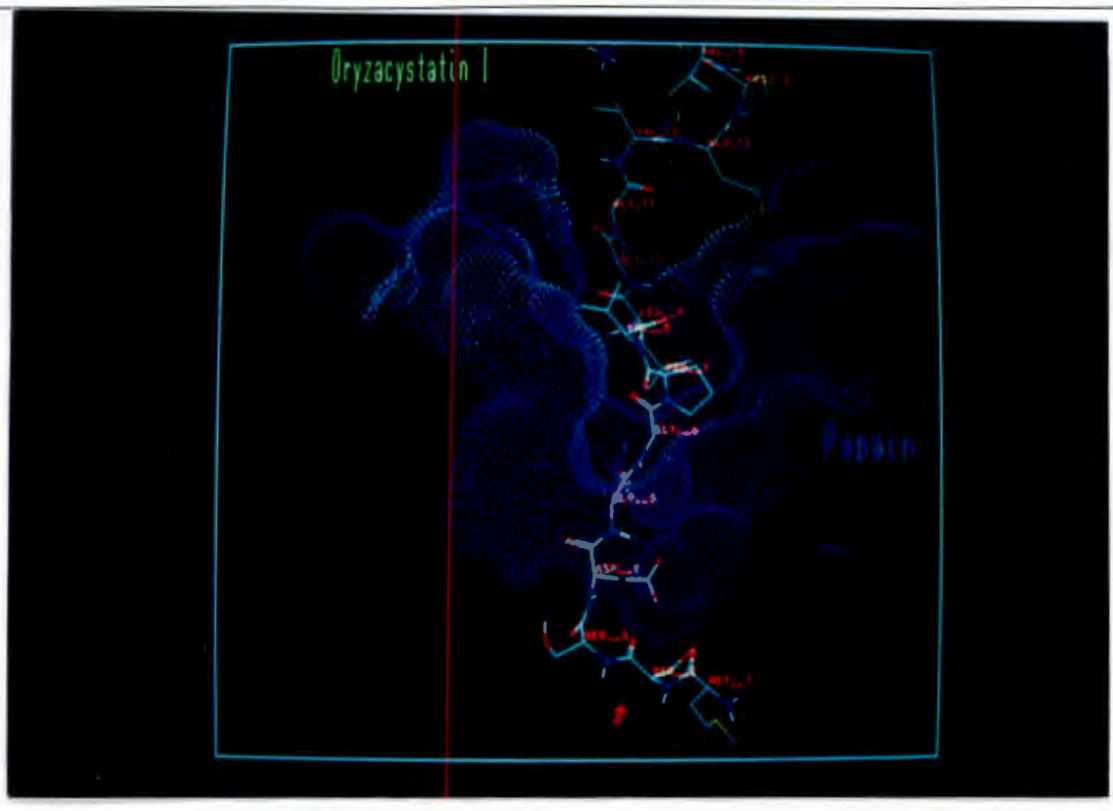


Figure 2.10: *The half-bonded backbone structure of the OC I amino-terminus along the groove of papain (blue). The final four amino terminal residues (indicated by the red arrow) are bending down.*

Figure 2.11 shows a ribbon model for a proposed structure of OC I with the N-terminus 'wrapping' around papain.

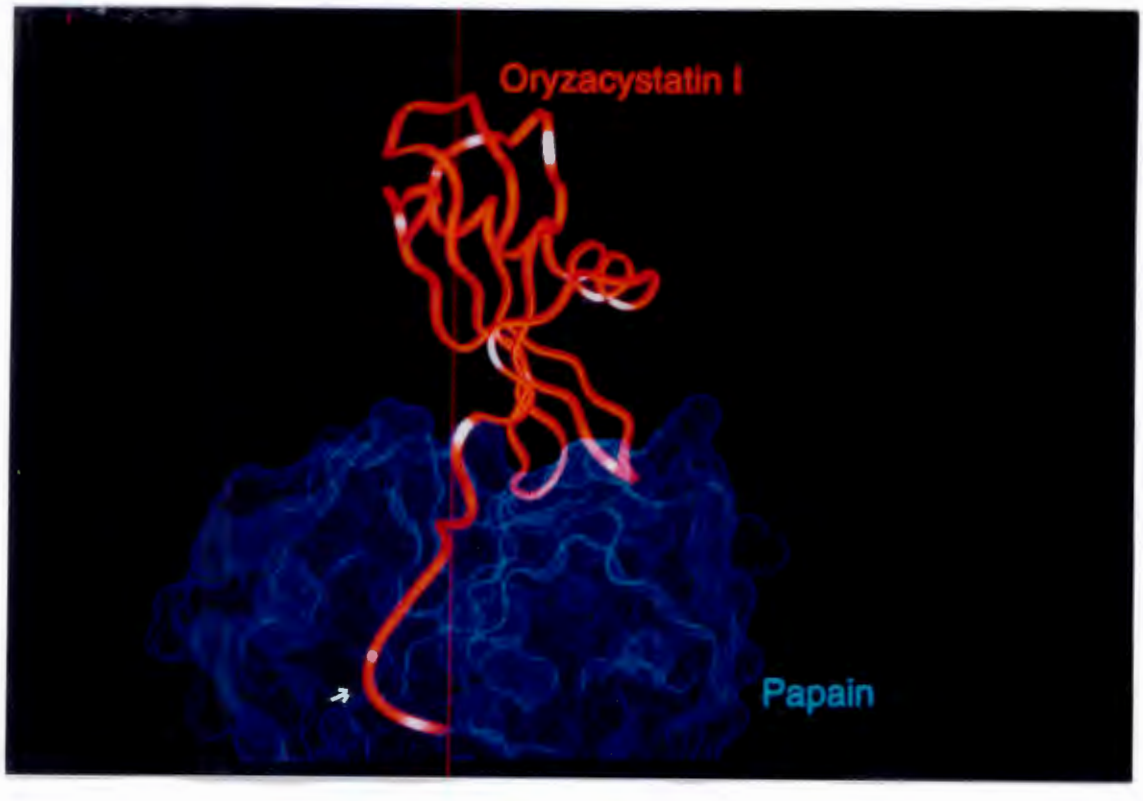


Figure 2.11: *A ribbon model of a proposed structure of OC I on binding to papain. The arrow indicates the amino-terminus.*

2.9.1 Observations from the Model

The literature offers conflicting views of the functions of the N-terminus. Urwin *et al.* (1995b) found that the K_i value of an amino-terminally truncated oryzacystatin mutant lacking Gly^4 (numbering as in Figure 2.9) was comparable to that of the full-length oryzacystatin, while the removal of Gly^9, on the other hand, drastically reduced the inhibitory capabilities of this inhibitor (Urwin *et al.*, 1995b). In contrast, Abe *et al.* (1988) showed that oryzacystatin hybrid mutants lacking the amino-terminal 21 residues but having an additional 15 amino acids from the cloning vector, efficiently inhibited papain activity.

From the sequence alignments of OC I with stefin B (Figure 2.9) it appeared that Gly^9 of oryzacystatin could correspond to Gly^9 in chicken cystatin. This is different to the alignment proposed by Abe *et al.* (1988), who suggested that the Gly^4 residue of oryzacystatin is analogous to the Gly^9 residue in chicken cystatin. According to the docking models of chicken cystatin (Bode *et al.*, 1988), however, the Gly^9 residue of this cystatin appears to be involved in a tight turn formation which brings Leu^8 in contact with the S2-subsite binding region of papain. From the alignment, oryzacystatin has a leucine residue (Leu^8) in the equivalent position to that in chicken cystatin. Since conserved residues often occur in equivalent positions in the cystatins, one would expect the Gly^9 residue to contribute to the turn formation as is seen in the case of chicken cystatin. Furthermore, if Gly^9 of chicken cystatin were aligned with Gly^4 of OC I, this would put the Pro^{12} residue (numbering as in Figure 2.9) in the α–helix region and proline is generally regarded as a helix breaker (Chou & Fasman, 1978). Although the presence of a proline residue in the α-helix of stefin A is described (Martin *et al.*, 1995; Tate *et al.*, 1995), it nevertheless seemed probable that chicken cystatin and oryzacystatin would interact in a similar manner and that the alignment of Gly^9 of OC I with Gly^9 of chicken cystatin and Gly^4 of stefin B (stefin B numbering – Stubbs *et al.*, 1990) seems more feasible.

OC I has the same QVVAG sequence as stefin B and would hence form a similar structure. A study of the model of OC I with papain shows that this QVVAG loop exerts its inhibitory effect by forming a rigid loop structure over the catalytic site residues as is suggested by Yamamoto *et al.* (1992). This conformation of the first binding loop is very important for the cystatin inhibitory activity since it is primarily responsible for the tight binding to proteinases (Thiele *et al.*, 1990) and has been demonstrated experimentally (Abe *et al.*, 1991) to be important for the inhibitory activity of OC I.

With regard to the second binding loop of cystatin, substitutions of amino acids in this region (i.e. a tryptophan residue replacing His^{104} of stefin B (numbered as in Figure 2.9), methionine replacing Glu^{105} and aspartic acid replacing Asn^{105A}), resulted in changes which could be well accommodated in the structure of the second binding loop, following energy minimisations and simple dynamics. This occurred as the amino acid side chains found different locations in space with respect to papain. It is clear from the model that the second hairpin loop (which contains the conserved tryptophan residue) does not make as strong a contact with papain as the first hairpin loop, but rather makes contact with the enzyme through its side-chains. As in the case of the first hairpin loop, the importance of this second loop in the binding of OC I to papain, has been

demonstrated (Abe *et al.*, 1988; Urwin *et al.*, 1995a). Removal of this loop resulted in weaker inhibitory activity. A comparison of the OC I and stefin B models is shown in Figure 2.12. This shows the small changes that have occurred in the second loop by insertion of the OC I sequence. It also shows the extended amino-terminus. The other changes seen (i.e. in the first loop after the α-helix), occurred through substitutions of the OC I sequence onto the stefin B structure. This section of the OC I model is similar to model 3 and so the changes made in this area will be discussed in terms of the models in section 2.11.

Carboxy-terminal contacts of oryzacystatin with papain appeared from the modelling studies to be similar to those of chicken cystatin i.e. there is no contact with papain. Stefin B, on the other hand, was shown by X-ray crystallographic studies to contact papain through the side-chains of residues Tyr^{124} and Phe^{125} (Stubbs *et al.*, 1990). The importance of these C-terminal contacts of stefin B is controversial. Despite the X-ray crystallographic findings, deletion of the C-terminus resulted in little change in the inhibitory activity of stefin B (Jerala *et al.*, 1991). In the case of oryzacystatin, removal of the C-terminal 11 residues also showed no change in the activity of the inhibitor (Abe *et al.*, 1988). This is in agreement with the modelled structure of OC I which suggests that there are no contacts with papain in the C-terminal regions.

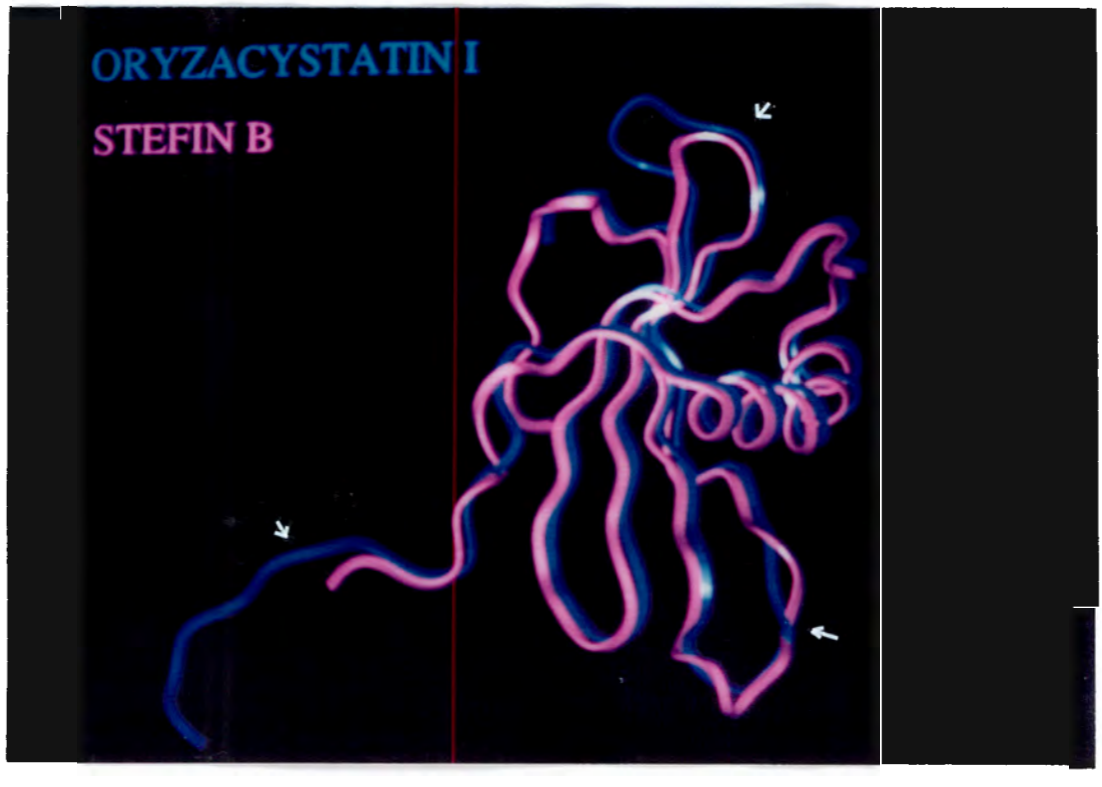


Figure 2.12: Ribbon models of OC I and stefin B showing the differences between them. The arrows indicate the changes made by amino acid substitutions to form the OC I structure.

2.10 Construction of the Hybrid Oryzacystatin

The OC I model together with the protein sequence alignments, secondary structure predictions and published Ki values for cystatins, were used to consider possible sites for introducing mutations that might improve the inhibitory capacity of OC I and, at the same time, shed light on the possible functions of the amino terminus. In this regard, the approach that appealed to me as the most interesting and potentially the most informative was that in which the N-terminus of oryzacystatin would be removed and the N-terminus of chicken cystatin inserted in its place. The reason for this choice was as follows:

- It is clear that, along with human cystatin C, chicken cystatin is one of the most potent inhibitors of papain, with a Ki in the region of 0.006 nM (Machleidt *et al.*, 1989).

- It has been proposed (Auerswald *et al.*, 1992) that chicken cystatin has strong interactions in its amino-terminus in order to compensate for less desirable contacts in the QLVSG binding loop. In fact, Machleidt *et al.* (1991) estimated that 36% of the binding energy is involved in the formation of a complex between the N-terminal residues of chicken cystatin and papain. The N-terminal residues of stefin B on the other hand contribute less to the binding energy in complex with papain (Thiele *et al.*, 1990). A similar effect is thought to occur with OC I (Abe *et al.*, 1988). Hence, the amino terminal contacts in the family II cystatins may be more important than in the family I cystatins (Tate *et al.*, 1995). Thus, by substituting the N-terminus of chicken cystatin for the oryzacystatin I N-terminus it was hoped that this would create a hybrid oryzacystatin with stronger interactions in its N-terminal region and thus a more potent inhibitor.

In deciding how much of the N-terminus of oryzacystatin to substitute or replace, the following points were considered:

- Firstly, according to Abe *et al.* (1988), the amino acids in the N-terminus of OC I do not make any contribution to the inhibitory capability of the inhibitor. Thus, alteration of any of these residues would not be expected to change the inhibitory activity of the inhibitor. This is, of course, debatable in view of the recent findings of Urwin *et al.* (1995b) who found that removal of Gly^{10} of OC I greatly decreased its inhibitory activity.

- Secondly, according to the sequence alignment of OC I with chicken cystatin (Figure 2.13), it is evident that there are regions of homology between the two sequences in the N-terminal region. For example, the sequence 'END' is common to both sequences. In chicken cystatin this is the start of the α-helix (Bode *et al.*, 1988). A similar prediction was made in this area for OC I using Chou-Fasman predictions of protein secondary structure. Thus, since the 'END' sequence could form the start of the α–helix in both sequences, the addition of the chicken cystatin sequence onto the N-terminus before these three residues should not affect the structure of the rest of the protein, particularly considering that the is flexible.

```
        1             2         3         4         5         6
        A12345678901   A2345678901234567890123456789012345678901AB2345
OC I:   MSSDGGPVLGGV---EPVGNENDLHLVDLARFAVTEHNKKANSLL--EFEKLYSVKQQVVAGTLYYFTIEVK
CC:      SEDRSRLLGAP----VPVDENDEGLQRALQFAMAEYNRASNDKYSSRVVRVISAKRQLVSGIKYI--LQVE

            7         8         9         10          11
        6789012345678901234567890123456789012A345A6789012345ABCDEFGHIJK
OC I:   EGD----------------------AKKLYEAKVWEK-PWMDFKELQEFKPVDASANA
CC:     IGRTTCPKSSGDLQSCEFNDEPEMAKYTTCTF-VVYSI-PWL-NQIKLLESKC-Q
```

Figure 2.13: *Sequence alignment of oryzacystatin I (OC I) and chicken cystatin (CC) (numbering as for chicken cystatin, extra amino acids in the OC I structure are represented using letters from the alphabet; adapted from Turk & Bode, 1991). The similarities between the sequences are highlighted in bold.*

- A final consideration was the extent to which the N-terminus of chicken cystatin was to be substituted. It is known that the full length and the truncated forms of chicken cystatin (starting at Leu^7) are equally active (Machleidt *et al.*, 1989). Since this hybrid inhibitor was to be synthesised, the shorter N-terminal sequence is preferable in terms of both cost and ease of synthesis. Thus, the first 18 residues of oryzacystatin were replaced with the truncated chicken cystatin sequence to generate the following sequence for the proposed hybrid oryzacystatin:

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

The chicken cystatin residues are shown in bold with underlining.

2.10.1 The Hybrid Oryzacystatin Models

Having decided upon the sequence substitutions and hence the sequence alignment, the structure was then modelled to determine whether the N-terminal substitutions were likely to have any structural effects in this region. Since stefin B was used as the basic structure for modelling OC I it was used again for the HO which should have essentially the same structure as OC I but with changes in the position of the amino-terminus.

Before modelling the HO structure, a series of alignments were carried out using the GCG computer programme (Devereux *et al.*, 1984). With computer alignments it is possible to vary the gap weight and so obtain a number of different alignments. Thus, when aligning, it is important to bear in mind the structural characteristics of the protein so that the sequence is aligned structurally as well as sequentially.

From a comparison of the hybrid and stefin B sequences, it was clear that, in order to model the HO onto the stefin B backbone, a number of substitutions and deletions of amino acids needed to be accommodated. Furthermore, since different alignments were possible, the positions of these deletions and insertions, with respect to the stefin B sequence, would vary. Three models were generated. Their features were as follows:

The HO - Model 1

The first alignment carried out on the GCG program used a gap weight of 1 and a length weight of 0.1. From this alignment (Figure 2.14), a table was drawn up to highlight the areas common to stefin B (shaded), as well as the areas where deletions (DEL) and insertions into the stefin B structure should occur in order to accommodate the oryzacystatin sequence (Table 2.2). This table was then used as the basis to produce the first model.

Model 1 represents the construction made according to the alignment shown in Figure 2.14. The stefin B molecule was used as the backbone structure. Using the table created (Table 2.2), amino acids in the stefin B structure that differed from the HO sequence were replaced with the corresponding hybrid amino acid residues.

The insertions and deletions were then studied. The first modification in this alignment was to delete the Ser^{8} and Ala^{13} - Gln^{15} (stefin B sequence, see Table 2.2.). The first deletion was catered for by deleting the Ser^{8} residue and forming a bond between Leu^{7} and Gly^{9} (HO sequence). The second deletion, following the 'GAPV' sequence (residues 9 - 12, stefin B numbering, Table 2.2), was catered for by moving the 'GAPV' up by three residues, i.e. the N-terminal 6 residues were docked onto the rest of the amino terminus.

```
Gap Weight: 1.000     Average Match: 0.540
      Length Weight: 0.100  Average Mismatch: -0.396
      Quality: 55.7         Length:  109
      Ratio: 0.605          Gaps:    8
 Percent Similarity: 58.824  Percent Identity: 35.294

 Hybrid Sequence (HO) x Stefin B Sequence (SB)

HO     1 ll.gapv...pvdendlhlvdlarfavteh.nkkansllefeklvsvkqq 45
         :: |||     |...:. |:.| .| .:.|  |||   :  | | ||.| |
SB     1 MMSGAPSATQPATAETQHIADQVRSQLEEKYNKK...FPVF.KAVSFKSQ 46

HO    46 vvagtlyyftievkegdakklyeakvwek.pwmdfk..elqefkpvd... 86
         ||||| |::.:.| ::|   : . :|::  |:   |:  |.:.:...
SB    47 VVAGTNYFIKVHVGDED...FVHLRVFQSLPHEN.KPLTLSNYQTNK... 93

HO    87 asana... 92
         | ...
SB    94 AKHDELTYF 102
```

Figure 2.14: *Sequence alignment of oryzacystatin I (OC I) and stefin B (SB) generated using the GCG computer programmes (Devereux et al., 1984). Alignment was carried out using a gap weight of 1.0 and a length weight of 0.1.*

This sequence was then processed by repeated steps of energy minimisation and simple dynamics to give a theoretical structure with the amino terminus moved away from the surface of papain. Rotating the carbonyl bond of the HO Gly^{9} (stefin B numbering) did not bring the amino terminal leucine residues into sufficient contact with the papain surface, so that one would not expect an improvement in the inhibitory activity from such a structure.

Insertions of amino acids were also required to convert the stefin B structure into that of the HO. From the alignment in Figure 2.14, the 'ANS' insertion (residues 39A - 39C; see Table 2.2) occur outside the binding loop regions of the inhibitor. There is also a deletion of the Tyr^{36} residue of stefin B in this area. Thus, the tyrosine residue was deleted, a bond formed between His^{35} and Asn^{37}, and the inserts at residues 39A - 39C and at 45A (Table 2.2) were accommodated by bulging the loop, following the α-helix, outwards. This region from Thr^{30} to Val^{48} was then energy minimised.

A similar procedure was carried out for the other insertion, AKK (residues 93A - 93C; Table 2.2), since these occurred in the vicinity of the second loop. Hence they could be accommodated by bulging this second loop outwards.

Table 2.2: Positions of substitutions, insertions and deletions for Model 1.

Stefin B Residue	Stefin B Number	Hybrid Residue
Met	6	Leu
Met	7	Leu
Ser	8	DEL
Gly	9	Gly
Ala	10	Ala
Pro	11	Pro
Ser	12	Val
Ala	13	DEL
Thr	14	DEL
Gln	15	DEL
Pro	16	Pro
Ala	17	Val
Thr	18	Asp
Ala	19	Glu
Glu	20	Asn
Thr	21	Asp
Gln	22	Leu
His	23	His
Ile	24	Leu
Ala	25	Val
Asp	26	Asp
Gln	27	Leu
Val	28	Ala
Arg	29	Arg
Ser	30	Phe
Gln	31	Ala
Leu	32	Val
Glu	33	Thr
Glu	34	Glu
Lys	35	His
Tyr	36	DEL
Asn	37	Asn
Lys	38	Lys
Lys	39	Lys
	39A	Ala
	39B	Asn
	39C	Ser
Phe	40	Leu
Pro	43	Leu
Val	44	Glu
Phe	45	Phe
	45A	Glu
Lys	46	Lys
Ala	47	Leu
Val	48	Val
Ser	49	Ser
Phe	50	Val
Lys	51	Lys
Ser	52	Gln
Gln	53	Gln
Val	54	Val
Val	55	Val
Ala	56	Ala

Stefin B Residue	Stefin B Number	Hybrid Residue
Gly	57	Gly
Thr	58	Thr
Asn	59	Leu
Tyr	60	Tyr
Phe	61	Tyr
Ile	62	Phe
Lys	63	Thr
Val	64	Ile
His	65	Glu
Val	66	Val
Gly	67	Lys
Asp	68	Glu
Glu	92	Gly
Asp	93	Asp
	93A	Ala
	93B	Lys
	93C	Lys
Phe	94	Leu
Val	95	Tyr
His	96	Glu
Leu	97	Ala
Arg	98	Lys
Val	99	Val
Phe	100	Trp
Gln	101	Glu
Ser	102	Lys
Leu	102A	DEL
Pro	103	Pro
His	104	Trp
Glu	105	Met
Asn	105A	Asp
	105B	Phe
Lys	106	Lys
Pro	107	DEL
Leu	108	DEL
Thr	109	Glu
Leu	110	Leu
Ser	111	Gln
Asn	112	Glu
Tyr	113	Phe
Gln	114	Lys
Thr	115	Pro
Asn	115A	Val
Lys	116	Asp
Ala	117	Ala
Lys	118	Ser
His	119	Ala
Asp	120	Asn
Glu	121	Ala
Leu	122	DEL
Thr	123	DEL
Tyr	124	DEL
Phe	125	DEL

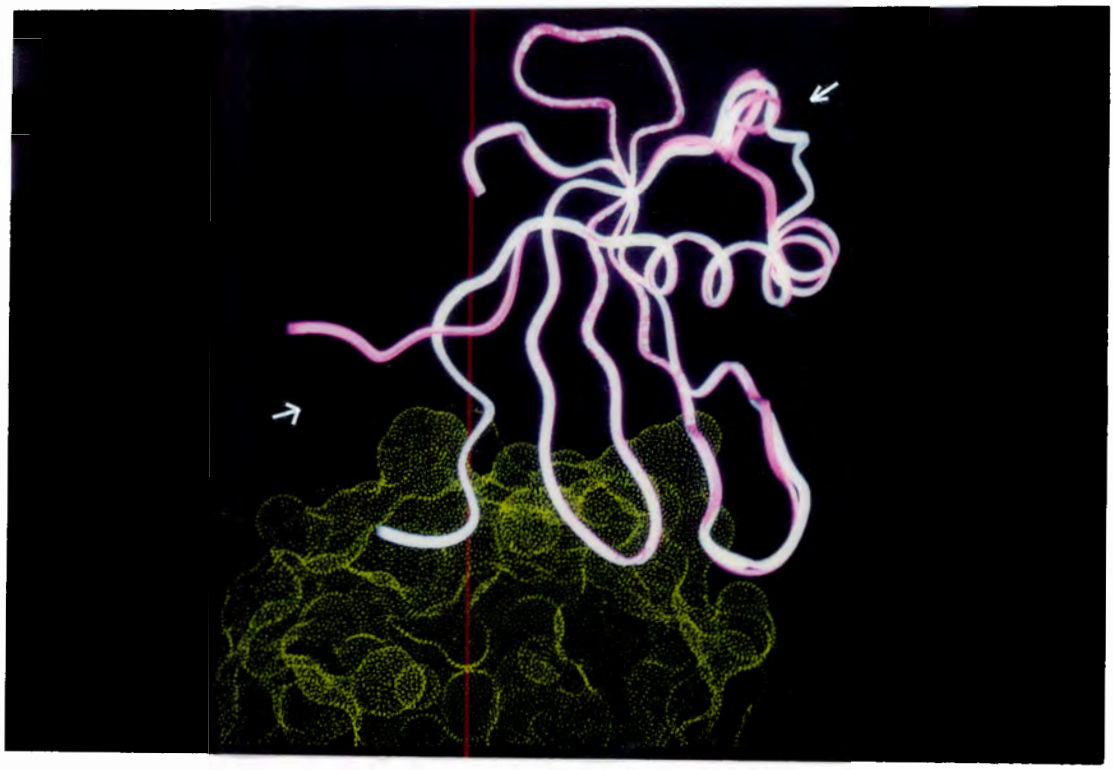


Figure 2.15: *Ribbon structures of stefin B (white) and model 1 (pink). The different positions of the amino-termini (arrow 1) and the loops generated by insertions (arrow 2) are shown. The papain surface is shown in yellow.*

The Hybrid Oryzacystatin - Model 2

A second alignment was generated as shown in Figure 2.16. In this alignment the early deletions in the amino terminus seen in model 1 have been moved to the end of the α-helix. Structurally, this results in removal of the last few turns of the α-helix. Thus, to determine the structural effects of this new alignment, a second model, was created.

Thus, to generate this model, the stefin B amino acids which differed from the HO sequence, were replaced with the HO residues as in model 1 (see Table 2.3). In this alignment (Figure 2.16), the first methionine residue of stefin B is removed, and two leucine residues replace the Met^{7} and Ser^{8} residues.

The second deletion occurs at residues Glu^{33}-Tyr^{36} in the stefin B sequence. These residues were thus deleted and a peptide bond was then formed between His^{32} and Asn^{37} (stefin B numbering) of the OC I sequence. The insertions, Ala^{39A}-Ser^{39C} required to form the HO model, were built in following Lys^{39} (Table 2.3). Energy minimisations and simple dynamics steps were then carried out over this region. However, despite these steps, there was still a large net energy in this region. An analysis of the surface of this modified area (see Figure 2.17), showed the formation of a 'hole'. It was therefore concluded that this model was not a compact structure and may in fact be implausible.

```
Gap Weight: 0.500     Average Match: 0.540
Length Weight: 0.500  Average Mismatch: -0.396
       Quality: 54.6          Length:  107
         Ratio: 0.593         Gaps:    10
Percent Similarity: 57.471  Percent Identity: 34.483

Hybrid Sequence (HO) x Stein B Sequence (SB)              . .

HO     1 .llgapvpvdendlhlvdlarfavteh....nkkansllefeklvsvkqq 45
         : ||| ...   ...:. .:.. ..      |||   :  | | ||.| |
SB     1 MMSGAPSATQPATAETQHIADQVRSQLEEKYNKK...FPVF.KAVSFKSQ 46

HO    46 vvagtlyyftievkegdakklyeakvwe.kpwmdfkelqefkpvdas. 91
         ||||| |:.:.:| ::|..    ::|   .|: :.| | ::.. .|.
SB    47 VVAGTNYFIKVHVGDED...FVHLRVFQSLPHEN.KPLSNYQTNKAK. 95

HO    92 ana.... 92

SB    96 HDELTYF 102
```

Figure 2.16: *Amino acid sequence alignment of stefin B (SB) and the hybrid oryzacystatin (HO) generated through GCG (Devereux et al., 1984). A gap weight of 0.5 was used and a length weight of 0.5.*

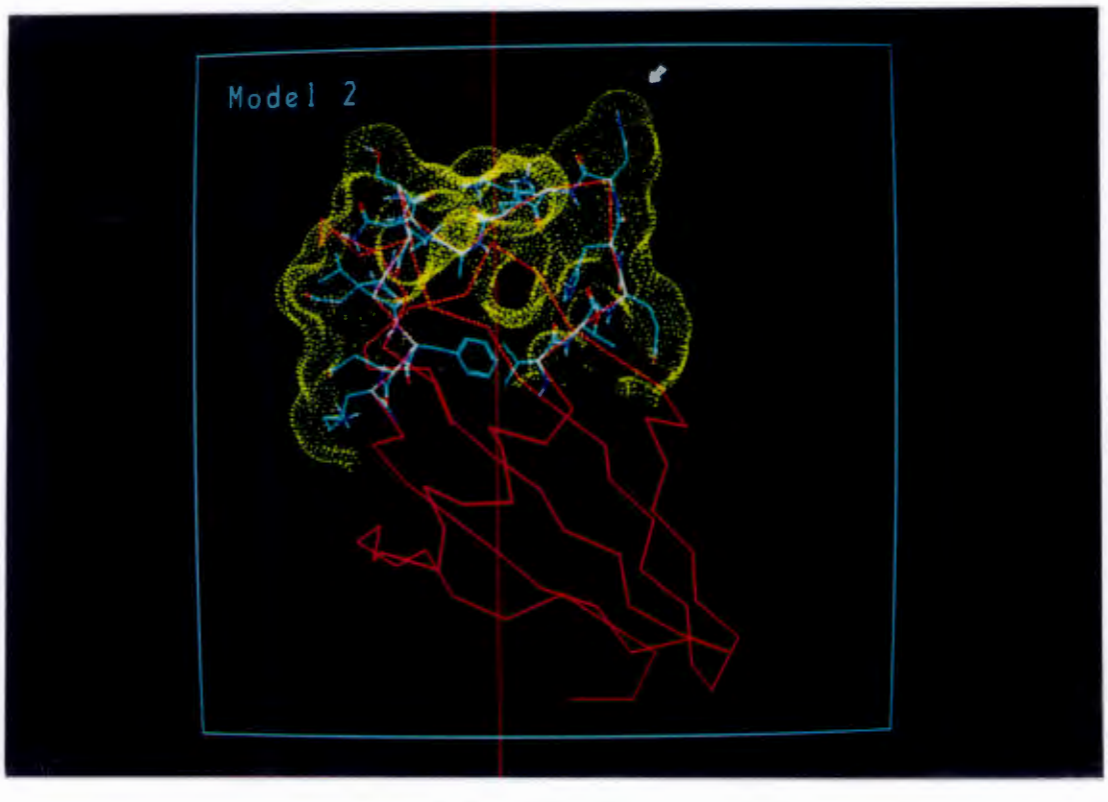


Figure 2.17: *Cα-trace of model 2. The amino acid side-chains (cyan) and the surface of the outer loops (yellow) are shown. The arrow indicates the problem areas, generated by the sequence alignment, which could not be accommodated.*

Table 2.3: Positions of substitutions, insertions and deletions for Model 2.

Stefin B Residue	Stefin B Number	Hybrid Residue
Met	6	DEL
Met	7	Leu
Ser	8	Leu
Gly	9	Gly
Ala	10	Ala
Pro	11	Pro
Ser	12	Val
Ala	13	Pro
Thr	14	Val
Gln	15	Asp
Pro	16	Glu
Ala	17	Asn
Thr	18	Asp
Ala	19	Leu
Glu	20	His
Thr	21	Leu
Gln	22	Val
His	23	Asp
Ile	24	Leu
Ala	25	Ala
Asp	26	Arg
Gln	27	Phe
Val	28	Ala
Arg	29	Val
Ser	30	Thr
Gln	31	Glu
Leu	32	His
Glu	33	DEL
Glu	34	DEL
Lys	35	DEL
Tyr	36	DEL
Asn	37	Asn
Lys	38	Lys
Lys	39	Lys
	39A	Ala
	39B	Asn
	39C	Ser
Phe	40	Leu
Pro	43	Leu
Val	44	Glu
Phe	45	Phe
	45A	Glu
Lys	46	Lys
Ala	47	Leu
Val	48	Val
Ser	49	Ser
Phe	50	Val
Lys	51	Lys
Ser	52	Gln
Gln	53	Gln
Val	54	Val
Val	55	Val
Ala	56	Ala

Stefin B Residue	Stefin B Number	Hybrid Residue
Gly	57	Gly
Thr	58	Thr
Asn	59	Leu
Tyr	60	Tyr
Phe	61	Tyr
Ile	62	Phe
Lys	63	Thr
Val	64	Ile
His	65	Glu
Val	66	Val
Gly	67	Lys
Asp	68	Glu
Glu	92	Gly
Asp	93	Asp
	93A	Ala
	93B	Lys
	93C	Lys
Phe	94	Leu
Val	95	Tyr
His	96	Glu
Leu	97	Ala
Arg	98	Lys
Val	99	Val
Phe	100	Trp
Gln	101	Glu
Ser	102	Lys
Leu	102A	DEL
Pro	103	Pro
His	104	Trp
Glu	105	Met
Asn	105A	Asp
	105B	Phe
Lys	106	Lys
Pro	107	DEL
Leu	108	DEL
Thr	109	Glu
Leu	110	Leu
Ser	111	Gln
Asn	112	Glu
Tyr	113	Phe
Gln	114	Lys
Thr	115	Pro
Asn	115A	Val
Lys	116	Asp
Ala	117	Ala
Lys	118	Ser
His	119	Ala
Asp	120	Asn
Glu	121	Ala
Leu	122	DEL
Thr	123	DEL
Tyr	124	DEL
Phe	125	DEL

To try to solve this problem, amino acid side-chains of hydrophobic residues (such as Leu[43], stefin B numbering), which had their side chains pointing outwards, were rotated inwards into the more hydrophobic environment. A number of constraints on torsion angles were also set. However, despite these changes, there was still no improvement in the structure of this model. It was thus concluded, that this alignment was not structurally feasible.

The Hybrid Oryzacystatin - Model 3

The alignment used to generate model 3 is shown in Figure 2.18. In this alignment, as with model 2, only a single deletion (the first methionine residue) occurs in the N-terminus.

In model 3 (Figure 2.19) insertions occurred in the top loop area outside of the α-helix (see Table 2.4). These areas are not important for binding of the inhibitor to the proteinase active site and changes in this region should have no effect on the inhibitory capabilities of this inhibitor. As mentioned previously, Auerswald *et al.* (1994) showed that deletion of the outer α-helical loop of chicken cystatin had no significant effect on the inhibitory activity towards papain, actinidin as well as the cathepsins B and L.

```
Gap Weight:  1.000     Average Match:  0.540
Length Weight:  1.000  Average Mismatch: -0.396
       Quality:  46.5          Length:  102
         Ratio:  0.505           Gaps:  2
 Percent Similarity: 45.652  Percent Identity: 23.913

Hybrid Sequence (HO) x Stefin B Sequence (SB)

HO     1 .llgapvpvdendlhlvdlarfavtehnkkansllefeklvsvkqqvvag 49
         : ||| :.:... .  .:|  . .: :.| |. :.. | ||.| |||||
SB     1 MMSGAPSATQPATAETQHIADQVRSQLEEKYNKKFPVFKAVSFKSQVVAG 50

HO    50 tlyyftievkegdakklyeakvwek.pwmdfk..elqefkpvdasana.. 92
         | |:.:.:| ::|     ::::|. . |.. :|. .| . ....|
SB    51 TNYFIKVHVGDED...FVHLRVFQSLPHEN.KPLTLSNYQTNKAKHDELT 100
```

Figure 2.18: *Sequence alignment of stefin B (SB) and the hybrid oryzacystatin (HO). The alignment was generated through GCG (Devereux et al., 1984). A gap weight of 1.0 was used and a length weight of 1.0.*

Table 2.4: Positions of substitutions, insertions and deletions for Model 3.

Stefin B Residue	Stefin B Number	Hybrid Residue
Met	6	DEL
Met	7	Leu
Ser	8	Leu
Gly	9	Gly
Ala	10	Ala
Pro	11	Pro
Ser	12	Val
Ala	13	Pro
Thr	14	Val
Gln	15	Asp
Pro	16	Glu
Ala	17	Asn
Thr	18	Asp
Ala	19	Leu
Glu	20	His
Thr	21	Leu
Gln	22	Val
His	23	Asp
Ile	24	Leu
Ala	25	Ala
Asp	26	Arg
Gln	27	Phe
Val	28	Ala
Arg	29	Val
Ser	30	Thr
Gln	31	Glu
Leu	32	His
Glu	33	Asn
Glu	34	Lys
Lys	35	Lys
Tyr	36	Ala
Asn	37	Asn
Lys	38	Ser
Lys	39	Leu
Phe	40	Leu
Pro	43	Glu
Val	44	Phe
Phe	45	Glu
Lys	46	Lys
Ala	47	Leu
Val	48	Val
Ser	49	Ser
Phe	50	Val
Lys	51	Lys
Ser	52	Gln
Gln	53	Gln
Val	54	Val
Val	55	Val
Ala	56	Ala
Gly	57	Gly
Gly	57	Gly
Thr	58	Thr
Asn	59	Leu

Stefin B Residue	Stefin B Number	Hybrid Residue
Tyr	60	Tyr
Phe	61	Tyr
Ile	62	Phe
Lys	63	Thr
Val	64	Ile
His	65	Glu
Val	66	Val
Gly	67	Lys
Asp	68	Glu
Glu	92	Gly
Asp	93	Asp
	93A	Ala
	93B	Lys
	93C	Lys
Phe	94	Leu
Val	95	Tyr
His	96	Glu
Leu	97	Ala
Arg	98	Lys
Val	99	Val
Phe	100	Trp
Gln	101	Glu
Ser	102	Lys
Leu	102A	DEL
Pro	103	Pro
His	104	Trp
Glu	105	Met
Asn	105A	Asp
	105B	Phe
Lys	106	Lys
Pro	107	DEL
Leu	108	DEL
Thr	109	Glu
Leu	110	Leu
Ser	111	Gln
Asn	112	Glu
Tyr	113	Phe
Gln	114	Lys
Thr	115	Pro
Asn	115A	Val
Lys	116	Asp
Ala	117	Ala
Lys	118	Ser
His	119	Ala
Asp	120	Asn
Glu	121	Ala
Leu	122	DEL
Thr	123	DEL
Tyr	124	DEL
Phe	125	DEL
Tyr	124	DEL
Phe	125	DEL

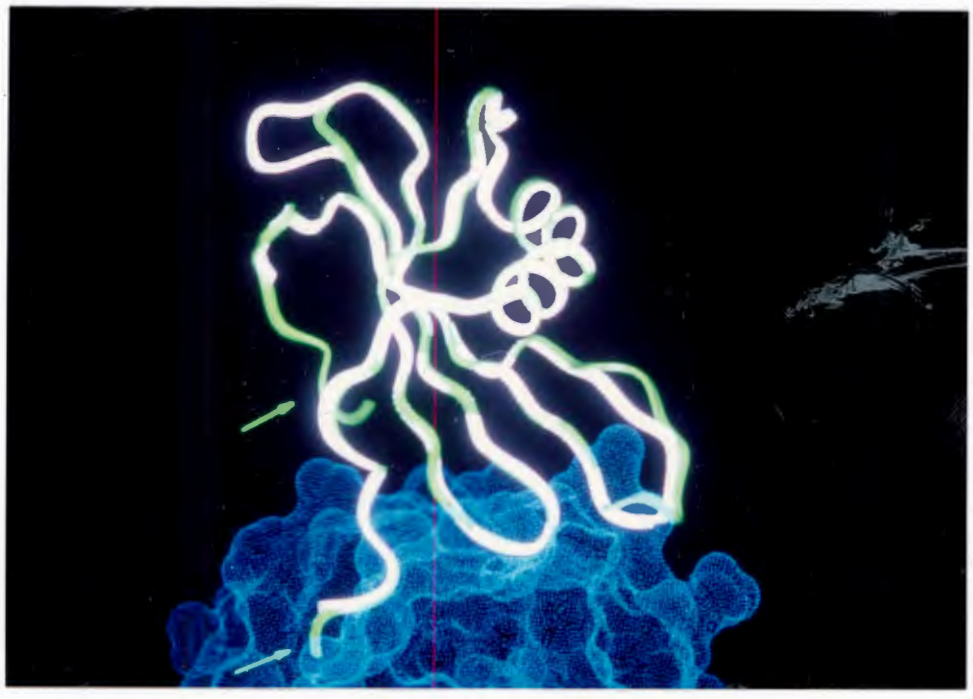


Figure 2.19: *Ribbon structure of model 3 (green) and stefin B (white). The papain surface is shown in blue and the different N- (bottom arrow) and C-termini(top arrow) are indicated.*

The Amino-terminal Interactions

The main area of interest in these models is the amino-terminus since the strength of the interactions in this area may have an impact on the inhibitory constants. In chicken cystatin the residues located N-terminal to Ala^{10}, that is, Leu^{7}, Leu^{8} and Gly^{9} are of major importance for the binding to papain subpockets S_2, S_3 and S_4 (Machleidt *et al.*, 1989). Thus it was predicted that Leu^{7} and Leu^{8} of HO should make similar contacts. The alignment of stefin B with the HO (Figure 2.18) shows the glycine of HO in equivalent positions to that of stefin B. It would hence be presumed to form a similar open type II turn conformation. This glycine residue in stefin B provides a flexible region allowing the N-terminal segment to adopt a conformation suitable for interaction with the substrate-biding pockets of the enzyme (Stubbs *et al.*, 1990).

Figure 2.20 shows the N-terminal 5 residues of model 3 and in Figure 2.21, the hydrophobic areas of papain are also shown. The leucine residues have their side-chains directed towards the hydrophobic residues of papain. These hydrophobic residues of the enzyme in this area include Tyr^{61}, Tyr^{67}, Pro^{68}, Trp^{69}, Val^{133}, Val^{157}, Ala^{160}, Phe^{207}. Leu^{7} of the HO (stefin B numbering) has its side-chains directed over a hydrophobic pocket (filled with water molecules) as is indicated by the white arrow (Figure 2.20). A serine residue is on the other side of this pocket where it is shielded from contacts with the leucine residues by the water molecules. Thus there appears to be strong hydrophobic interactions of the residues in this region. In the interactions of cystatins with papain it is the steric fit between the hydrophobic parts of the inhibitor and the enzyme which

dominates the interactions (Stubbs *et al.*, 1990). Whether or not these contacts would improve the hybrid could only be deduced from kinetic studies (see Chapter 6).

A further observation made from this model is that the glycine residue lies close to the active site residues of papain, as does the first hair-pin loop which lies close to the depression of the cleft between Cys^{25} and His^{159} of papain. Energy minimisations on the N-terminus brought the N-terminal four residues close to the active-site residues, although not too close to be cleaved at this site.

The single amino acid deletion at the amino terminus seen from the alignments of models 2 and 3 provides a better scenario than that shown in model 1. In models 2 and 3, the N-terminus lies in close contact with the enzyme, forming similar contacts to those seen with chicken cystatin (Stubbs *et al.*, 1990). Model 1, however, has no contact with the enzyme in this region. Of all the alignments, that of model 3 was considered to be structurally the most feasible.

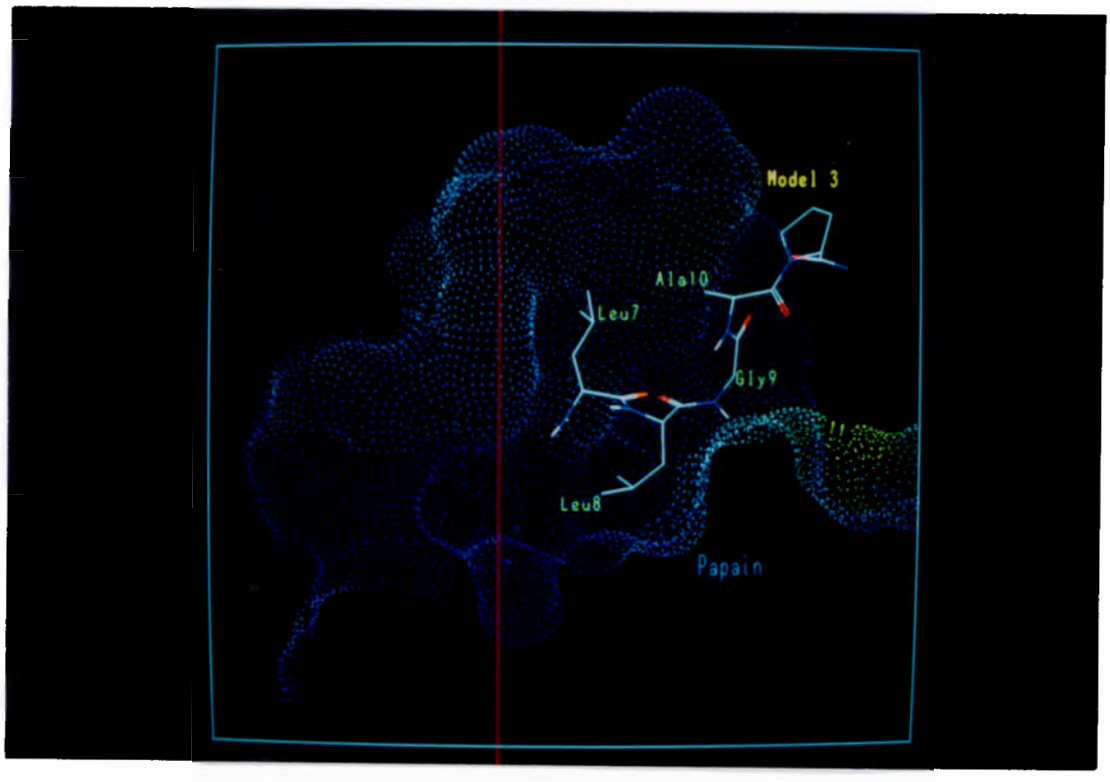


Figure 2.20: *Half-bonded structure of the N-terminal four residues. The papain electrostatic surface is shown.*

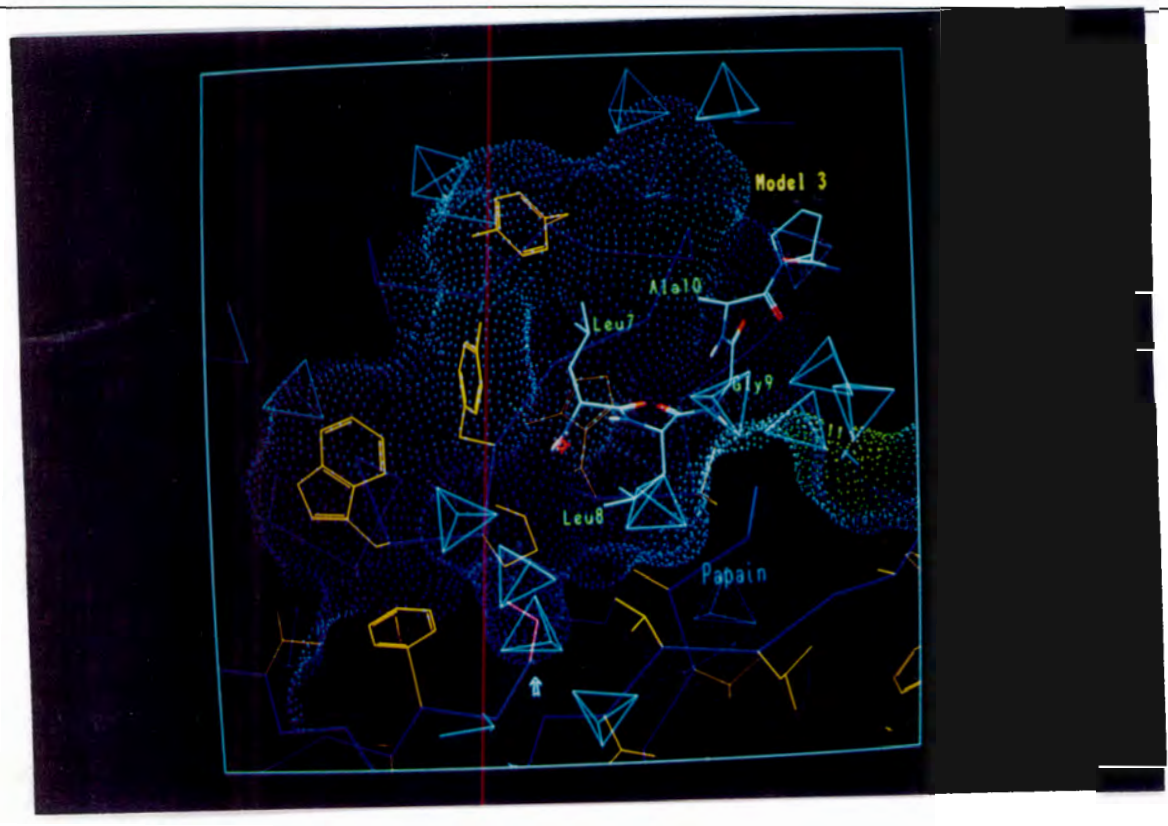


Figure 2.21: Half-bonded structure of the N-terminal four residues. The hydrophobic amino acid side-chains of papain amino acids are shown in orange. Water molecules are denoted by the tetrahedra.

2.12 Conclusions

It is clear from a review of the literature that the importance of the cystatin binding loops is still a matter of debate. The study that I have made of this has led me to suggest the following:

- The N-terminal contacts do play an important role since contacts in this area help anchor the inhibitor and so enable the other loops (particularly the first) to fit correctly into binding site cleft of the proteinase. The contacts may, however, be more important in some cystatin families (e.g. family II) than others (e.g. family I). Thus the loss of activity seen with truncation, particularly in the family II cystatins, could as suggested by the trunk model (Machleidt *et al.*, 1995), be due to loss of conformation in the rest of the loops. This suggests that the cystatin binding loops do not act independently of each other but rather in a concerted fashion so that each contributes to the stability of the interactions of the others.

- The importance of the various binding regions may be dependent on the stereochemistry of the proteinase, i.e. the fit of the inhibitor into the proteinase active site. Papain and actinidin, for example, are two similar enzymes with widely different susceptibilities to different cysteine proteinase inhibitors. These differences have been attributed to the differences in the catalytic-site reactivity, to small changes in the S_2-subsites (Baker & Drenth, 1987), or to differences in the electrostatic properties of the two enzymes (Pickersgill, *et*

al., 1988). My inclination, however, is to agree with Björk *et al.* (1994) who suggest that the extent of inhibition is determined by steric interactions between the inhibitor and the enzyme. This would explain the many different results seen in the literature for substitution/truncation and deletion experiments.

The three different alignments of HO and stefin B have given 3 alternative configurations for the HO structure. Of these, model 3 appears to be the most compatible with the structural characteristics of the cystatins, with the obvious reservation that the actual structure of OC I or the HO can only be deduced definitively by NMR or X-ray crystallography.

Despite possible variations in the structure of the hybrid protein, it should shed some light on the function of the amino-terminus. If the amino-terminus does interact, one would expect that the substitution of the chicken cystatin amino terminus would improve the inhibitory activity since chicken cystatin is such a potent inhibitor. Or, at the very least it should have an intermediary effect between that of natural chicken cystatin and natural oryzacystatin.

CHAPTER 3

Automated DNA Synthesis of the Hybrid Oryzacystatin Gene and Cloning into *Escherichia coli*

3.1 Introduction

The supply of proteins of clinical, industrial or research interest is often limited by their low natural availability and gene cloning and expression can provide a more abundant source. In this case, large quantities of the HO inhibitor were required for study and hence the aim was to clone the gene into a suitable expression system thereby producing the inhibitor in large quantities for inhibition assays.

This chapter will thus describe the synthesis and cloning of the hybrid oryzacystatin inhibitor gene into a suitable bacterial system. Cloning of the amplified gene into an expression vector for production of the protein is discussed in Chapter 4.

3.1.1 Considerations Prior to Gene Design

Two general considerations led to the selection of an *E. coli*-based expression system to produce the HO gene, for example:

1. The technology for DNA manipulation and cell culture compared with that of other expression systems is well established.
2. It is the most frequently used expression system, a large amount of information is available and genetic techniques have been well described. This information can be used to address any difficulties encountered in the expression of a particular gene.

Two options for producing the hybrid gene were considered. The first was to obtain the two inhibitor genes, oryzacystatin I and chicken cystatin. These two sequences could then be cut and joined together at the junction as determined by the hybrid model. The second option would be to synthesise the modified gene on an automatic DNA synthesiser.

After careful consideration, the second option was chosen for the following reasons:

1. Synthesis of the gene enables the selection of gene codons optimal for the expression system being used which can improve expression levels. De Boer *et al.* (1983) showed that the use of genes with low codon usage for the organism in which it is expressed could result in poor expression levels. For this gene, the codons used were as reported for *E. coli* by Wada *et al.* (1981).

2. Synthesis also makes it possible to introduce restriction sites and other features necessary for cloning into the required vector systems.

PART 1: Cloning of the Gene in Two Parts

3.2 Considerations in the Design of the Gene for Cloning and Expression

Prior to synthesis it is necessary to consider the methodology to be used for cloning since this may have a bearing on the gene design. The modified HO has 94 amino acids. This results in a gene sequence of 282 bases, which is too long to be synthesised in a single step on an automatic DNA synthesiser without loss of purity and fidelity (D.P. Botes, personal communication). Thus, to try and reduce the incidence of any errors, it was decided to divide the gene into smaller fragments which would then be cloned, in a step-wise manner, into pUC18 for gene amplification.

A number of methods have been used for cloning multiple sections of a gene, including the oligonucleotide-directed double-strand break repair method of Mandecki (1986) and PCR methods such as that of Michaels *et al.* (1992). It was decided to use this latter method as it has a number of advantages, for example:

- Low amounts of template DNA are required, reducing the cost of synthesis.
- It enables the inclusion of unique restriction sites which, by designing them on the end of a primer, can be incorporated into the template. These restriction sites can then aid in cloning of the gene into the required vector.

A further consideration in the design of synthetic genes is that of the stop and start codons, especially (as in this case) if these are not supplied by the expression vector. The TAA termination sequence was thus used as the stop codon at the end of the gene sequence (Figure 3.1). The start codon was not included in the gene design since it was to be supplied by the pMAL system of vectors (New England Biolabs) that were to be used to express the inhibitor following gene amplification. In the pMAL system (discussed in Chapter 4), foreign genes are expressed as a fusion linked to the maltose binding protein whose coding sequence is provided by the transcription start codon.

3.2.1 Dividing the Gene into Two Templates

To decide where division of the gene should occur, the sequence was scanned for restriction sites using the Genetics Computer Group (GCG) database of 181 restriction enzymes (version 7.1; Devereux *et al.*, 1984). As no restriction sites were found that were unique to the HO compared with the pUC18 vector, and which cut the sequence into two manageable pieces, the sequence was then scanned for partial restriction sites. The GCG program is able to determine where in a sequence changes in the wobble base of degenerate codons will introduce restriction sites and still code for the same amino acids. Thus, using this method, a unique *Nae*I site

was generated by changing the codon of alanine from its preferred codon GCG to GCC. This approach allowed for the introduction of new restriction sites and produced two shorter sequences of 147 (template A) and 143 bases (template B) for synthesis (Figure 3.1).

As cloning was to be carried out in a step-wise manner, template A followed by template B, it was necessary to include the Ala-Gly amino acids at the end and at the beginning of templates A and B respectively. Template A could then be cut with *Nae*I, thus opening up the vector for insertion of template B (also cut with *Nae*I; Figure 3.1). These extra amino acids were thus included when synthesising the templates.

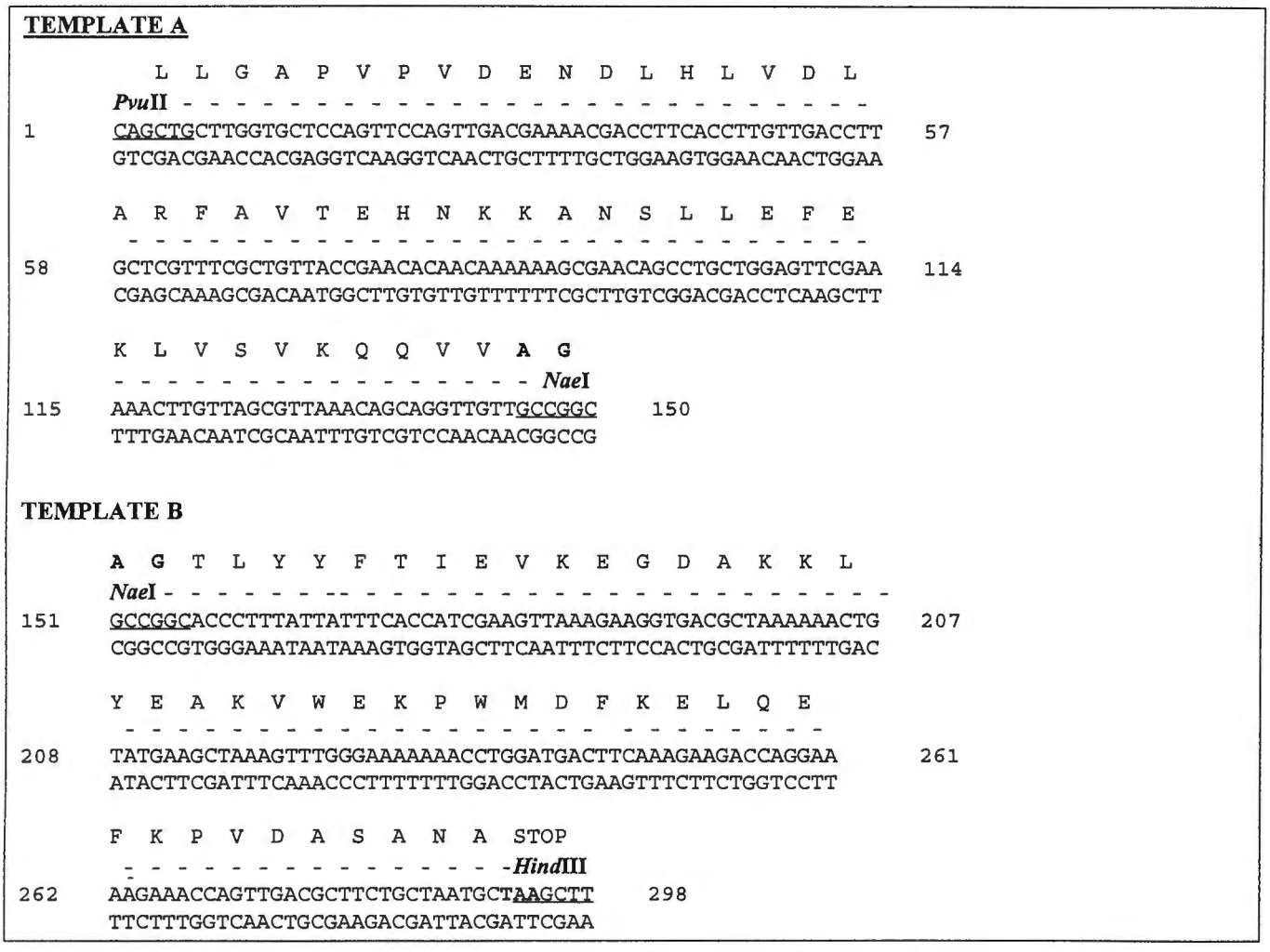


Figure 3.1: *The DNA and amino acid sequence of the synthetic HO gene showing the division of the gene into two fragments. Restriction enzymes were incorporated into the sequence by modification of the codon usage. The first 3 bases of the PvuII site are contributed by the forward primer to template A. Similarly, the last 4 residues making up the HindIII site are contributed by the reverse primer to template B. The repeated amino acid residues are shown in bold as is the 'TAA' stop codon.*

Primer sequences were designed to incorporate restriction sites for easy insertion and excision of the cloned gene from the vector. Since the pUC18 vector was only to be used as a vehicle for gene amplification with subsequent excision of the insert and recloning into the pMAL vector for expression, a restriction site for *Pvu*II

was included at the start of the gene sequence (Figure 3.1). This was possible because the codon for the amino acid at the start of the gene sequence, leucine, is CTG and the cutting site for *Pvu*II is CAG ↓ CTG. This blunt-ended restriction site was useful in that on cloning into the expression vector, no undesirable pUC18-derived amino acids residues would be added to the amino-terminus. The first three bases of the *Pvu*II recognition sequence were included in the forward primer sequence for template A (Table 3.1). A *Bam*HI site was also engineered onto both the forward and the reverse primers to enable cloning into the *Bam*HI site of the pUC18 vector (Figure 3.2).

The reverse primer for template A contained the *Nae*I restriction site. This restriction site requires 'buffering sequences' 3' to the *Nae*I site, as this enzyme does not cut well at the end of a gene sequence (Promega manual, 1992). These were provided by the *Bam*HI sequence. A sixteen base-pair overlap with the template was also included for both primers (Table 3.1).

For template B, a *Hind*III restriction site was incorporated, following the stop codon, into the reverse primer sequence. This was done to enable removal of the gene from pUC18 (using *Pvu*II and *Hind*III) and to allow 'blunt/sticky' directional cloning into the pMAL vectors (which contain *StuI* and *Hind*III restriction sites in the multiple cloning site; Appendix D).

A *Nae*I site was included in the forward and reverse primer sequences for template B (see Table 3.1), since the vector containing template A (named pUC18a), was to be cut with *Nae*I for cloning in of template B (Table 3.2).

Before synthesis, the primer sequences were optimised using the computer program 'Primer' (Version 5.0; Whitehead Institute for Biomedical Research, Cambridge, MA). Primers with restriction sites are shown in Table 3.1.

Table 3.1: Primer sequences and their restriction enzyme sites required for cloning

Primer	Sequence	Restriction Sites Present
Forward, Template A	5' CTGCATGTGCGCACAG***CTGCTTGGTGCTCCAG*** 3'	*Bam*HI, *Pvu*II
Reverse, Template A	5' CAGGTGCGCTT***GCCGGCAACAACC*** 3'	*Bam*HI, *Nae*I
Forward, Template B	5' GCGGGCGGC***GCCGGCACCCTTTATTA*** 3'	*Nae*I
Reverse, Template B	5' CATGCCGGCGTACAAGC***TTAGCATTAGCAG*** 3'	*Nae*I, *Hind*III

The underlined sequence show the restriction sites in the order as indicated in the last column.

The overlap regions with the template are shown in bold and italics.

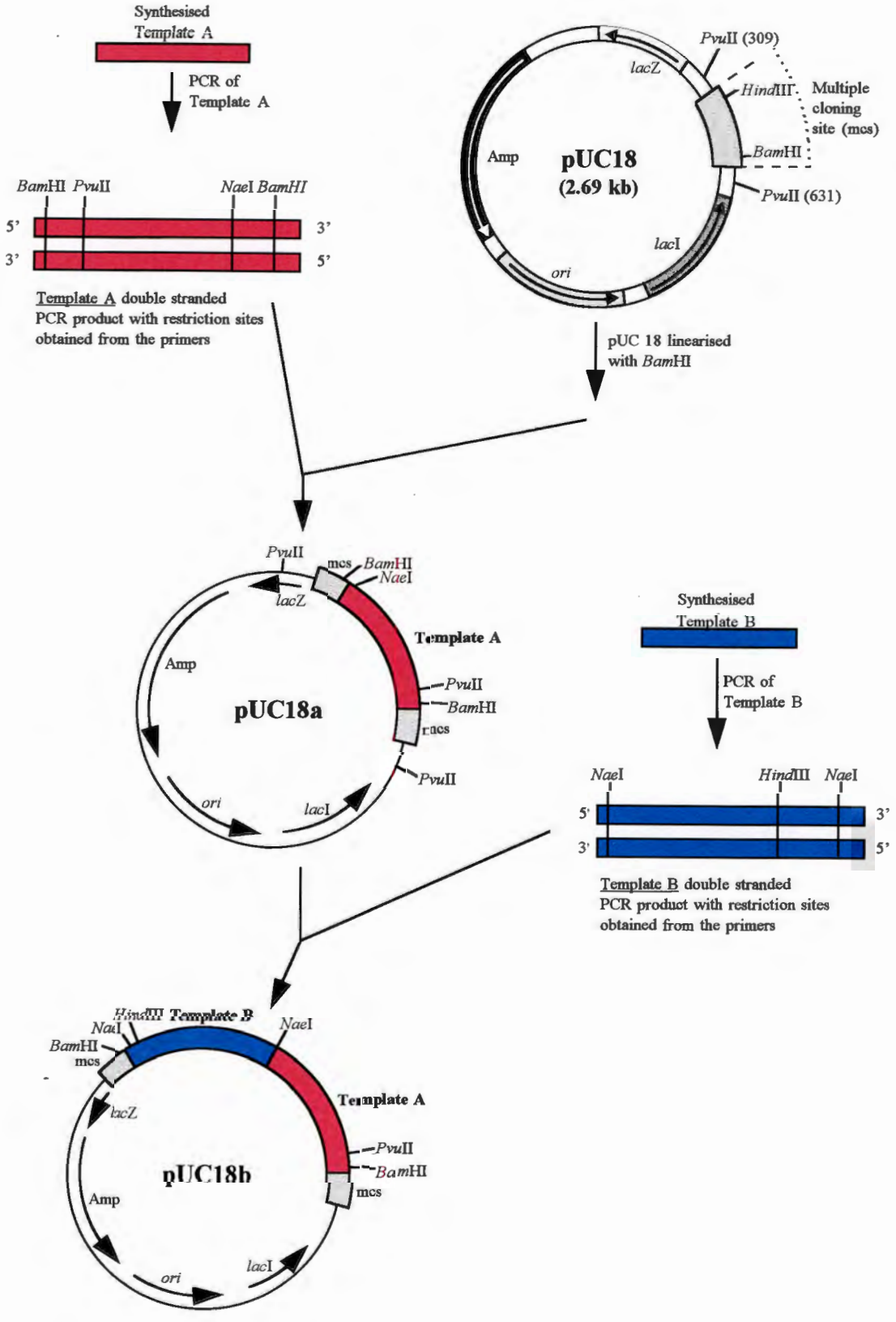


Figure 3.2: *Schematic diagram showing the planned cloning procedure for templates A and B. Template A was amplified by PCR, cut with BamHI and cloned into the BamHI-linearised pUC18 to generate pUC18a. Template B was to be inserted by cutting the PCR product with NaeI and inserted into the NaeI linearised pUC18a to generate pUC18b.*

3.3 Materials and Methods

3.3.1 Bacterial Strains, Plasmids and Growth Conditions

E. coli strain JM109 (*recA1, endA1, gyrA96*, thi, *hsdR17*, *supE44*, *relA*1, λ⁻, Δ(*lac-proAB*), [F', *traD*36, *proAB*, $lacI^q$ ΔZM15]) was used for these experiments. The cultures were grown in Luria broth (LB) at 37°C with vigorous shaking. Transformed strains carrying pUC18a and b were grown as above but with 100 μg/ml ampicillin (Sigma).

3.3.2 Gene Synthesis

Oligonucleotides were synthesised with nucleoside-O-(2-cyanoethyl)-N,N-diisopropyl phosphoramidites. These and other reagents for synthesis were obtained from Bioresearch. Synthesis was performed on a model 6500 Autogen™ DNA synthesiser (Milligen Bioresearch, Burlington, U.S.A). Primers were synthesised as for the template oligonucleotides. This was carried out in a glass column containing 1 μmole (for the templates) and 0.2 μmole (for the primers) of an appropriate 5'-O-dimethoxytrityl deoxynucleoside-3'-O-succinyl-derivatised 1000 Å control pore glass support at a loading of 18 μmol/g.

For the template A reverse primer containing biotin (for purification of the PCR product; section 3.4.2), a biotin phosphoramidite (Glen Research) was added during synthesis as described (Nelson *et al.*, 1992). The reaction kinetics of this phosphoramidite is virtually identical to a nucleoside phosphoramidite and has a 12 – 15 minute coupling optimum.

3.3.3 Purification of Oligonucleotides by Reverse-phase HPLC.

Following synthesis the last unblocking step was omitted in order to retain the terminal dimethoxytrityl (DMT) group. This 'DMT-containing' synthesis product was dissolved in Buffer A (50 mM LiCl, 10 mM NaOH, pH 12.0) and loaded onto a Pharmacia Mono Q HR 5/5 ion exchange column. Elution was via a linear gradient with Buffer B (3 M LiCl, 10 mM NaOH, pH 12.0). The tritylated oligonucleotides were recovered by precipitation with ethanol/acetone (1:3), and centrifugation at 15 000 rpm for 1 hour. The oligonucleotide pellet was then resuspended in concentrated ammonia and dried. The final DMT group was removed by adding 100 μl 80% acetic acid for 1 hour at room temperature. Recovery was by butanol precipitation and pellets were resuspended in 50 μl TE buffer (10 mM Tris; 1 mM EDTA, pH 7.6). The purity of the template oligonucleotides was confirmed by electrophoresis on 10% non-denaturing polyacrylamide gels (Appendix C).

3.3.4 Gene Amplification by PCR

PCR amplification of oligonucleotide templates was carried out in a 50 µl reaction volume. PCR cycles were optimised for template, primer and magnesium concentrations and annealing temperature. An average cycle consisted of a denaturation step at 93°C for 20 seconds, an annealing step of 55°C for 30 seconds and an extension step of 72°C for 1 minute. DNA amplification cycles were controlled by a custom-made thermocycler (JDI Model 8012). Promega buffers and *Taq* DNA polymerase were used for PCR. Following amplification, 10 µl aliquots were removed and electrophoresed on 10% non-denaturing polyacrylamide gels to check the PCR product was the predicted size.

3.3.5 Removal of DNA Bands from Polyacrylamide Gels

A modification of the method of Maniatis *et al.* (1982) was used to purify DNA from polyacrylamide gels. Full-length PCR products or templates were removed from 10% non-denaturing polyacrylamide gels as follows:

The required DNA bands were removed using a razor blade, chopped into small pieces, and placed into 1.5 ml Eppendorf tubes. The gel pieces were covered with elution buffer (0.5 M Ammonium Acetate; 10 mM Magnesium Acetate; 0.1% SDS; 1 mM EDTA, pH 8.0) and incubated overnight at 37°C in a shaking air incubator. The DNA fragments were recovered from the supernatant solution by precipitating in 2 volumes of ethanol and centrifuging at 14 000 rpm for 20 minutes. Pellets were washed in 70% ethanol, dried and resuspended in TE buffer.

3.3.6 Streptavidin Purification of Biotin Labelled PCR Products

PCR products in which biotin labelled primers had been used during PCR were purified using magnetic Dynabeads (Dynal, Norway) which had streptavidin attached. Forty microlitres of the beads was placed in a 1.5 ml Eppendorf tube and washed twice with 1 × PBS (phosphate-buffered saline; Sambrook *et al.*, 1989), to remove the sodium azide used to store the beads. The beads were then equilibrated with PCR buffer (Promega) without MgCl. This buffer was then removed by placing the Eppendorf in the magnetic holder supplied by the manufacturer (Dynal, Norway). Following equilibration of the beads, 50 µl of the PCR product was added to the beads and these were placed on a roller (to keep the beads from settling) for 15 minutes at room temperature. The non-bound DNA (supernatant), was then removed from the beads by again placing the Eppendorf in the magnetic holder. The beads were then resuspended in restriction enzyme buffer H (Boehringer Mannheim) and the bound DNA was then removed from the beads by digestion with uracil dehydrogenase (Gibco). Purified DNA products were visualised on 10% non-denaturing polyacrylamide gels stained with ethidium bromide.

3.3.7 Standard molecular methods

All standard molecular genetic techniques were carried out using the methods described in Sambrook *et al.* (1989). Restriction enzymes were obtained from Boehringer Mannheim. For the cloning of template A, pUC18 was 'maxi-prepped,' cut with *Bam*HI, and purified from a 1% agarose gel in TAE buffer using the GENECLEAN^R II kit (Bio 101, Inc.) according to the manufacturer's instructions. The purified vector was digested and the 5'-phosphates removed using calf intestinal phosphatase (Boehringer Mannheim). The vector was ligated to the *Bam*HI-cut template using T4 DNA ligase (Promega). The dimethyl sulphoxide (DMSO) transformation method of Chung and Miller (1988) was used to prepare competent *E. coli* cells.

3.3.8 Screening for Positive Clones using PCR

For PCR screening, a few colonies were selected from plates and grown in 3 ml of LB containing 100 μg/ml ampicillin. 10 μl aliquots were then used in the PCR reaction. 10 pmoles of the pUC18 forward and reverse primers were used for the PCR. An average cycle consisted of a denaturation step at 93°C for 20 seconds, an annealing step of 50°C for 30 seconds and an extension step of 72°C for 1 minute. After 30 cycles, a final extension step of 5 minutes at 72°C was used. After PCR the aqueous phase was removed and electrophoresed on a 1% agarose gel. Positive clones were identified by an increase in size of the PCR band from the control (vector only, or vector plus insert). The insert sequences were confirmed using the dideoxy chain termination sequencing method (Sanger *et al.*, 1977) with the Sequenase kit (version 2.0; United States Chemical Corp).

3.4 Results and Discussion

3.4.1 Synthesis and PCR Amplification

Template A and the forward and reverse primers were synthesised as described (section 3.3.2) and the quality of these oligonucleotides is shown in Figure 3.3 (lanes 2 - 4). From the template only band in lane 4, it is clear that template A did not consist of a single 147-bp species but was contaminated with a number of shorter 'failure' sequences. These arise in even the most efficient synthesis strategies as a result of incomplete coupling of nucleoside 3'-phosphoramidites with the controlled pore glass (CPG) supported 5'-hydroxyl DNA (Horn & Urdea, 1988). In addition, shorter fragments also arise when the oligonucleotide is removed from the CPG by treatment with concentrated ammonia. This cleaves the DNA at any apurinic sites (Horn & Urdea, 1988). The primers to template A, forward (lane 3) and reverse (lane 2) were purer, possibly because the sequences were much shorter.

PCR of the 'mixed' template A species was highly inefficient and in order to obtain full length PCR product conditions had to be optimised for template, primer and magnesium concentrations, as well as for optimal annealing temperatures (results not shown). Figure 3.3 (lane 5) shows the most optimal amplification product obtained.

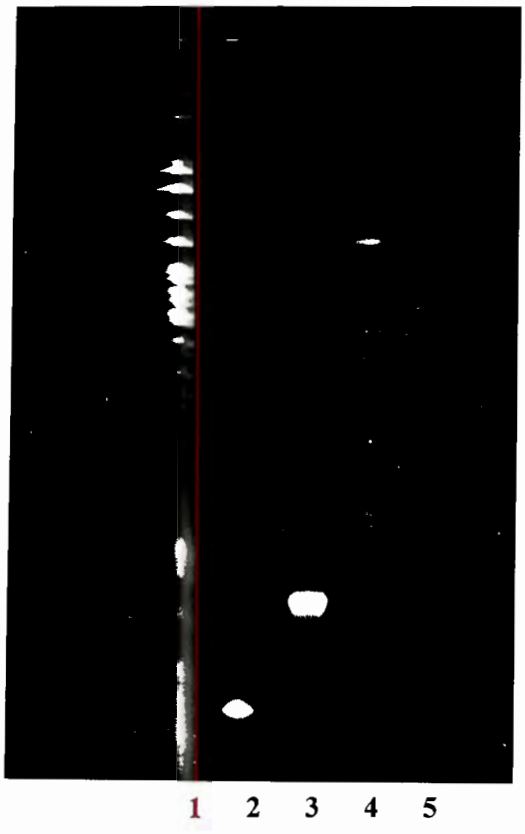


Figure 3.3: *10% Non-denaturing polyacrylamide gel showing the synthesised oligonucleotides and the results of PCR amplification for template A. Lanes 1, molecular weight markers pBR322 digested with HpaII; 2, reverse PCR primer; 3, forward PCR primer; 4, template A only; 5, PCR product for template A.*

The presumed full length PCR products (top band in lane 5) was then excised from the polyacrylamide gel and eluted, cut with *Bam*H1 and ligated to the *Bam*H1-linearised pUC18. After transformation, positive clones were detected by PCR screening and the sequence checked by DNA sequencing. This showed that the PCR products were not full length, despite the removal of the upper-most PCR band from the polyacrylamide gel prior to cloning. Twelve residue deletions were observed from the 3' end indicating that the smaller PCR products had been cloned. Purification of the top template band from the polyacrylamide gel before PCR did not solve this problem. These base deletions were clearly too small to be detected by PCR screening because the differences between those with deletions and the full length PCR products were not detectable on a 1% agarose gel. There were also other random errors in the sequence that may have been introduced by PCR or

resulted from errors made during synthesis. Furthermore, the efficiency of the synthesis process may have been too small to accommodate the long length of this template (D. P. Botes, personal communication).

3.4.2 Use of Biotinylated Primers for Purification of the Full Length PCR Product

To overcome the problems encountered in 3.4.1, two steps were taken:

- Firstly, it was decided to resynthesise the template using a larger CPG glass column. The 147-base template A had previously been synthesised on a 1000 Å column. In this case we decided to use a 2000 Å CPG column as this larger size would accommodate the larger DNA length and might result in a purer template with fewer truncated sequences.
- A second step was to use biotinylated oligonucleotides for affinity purification as has been documented (Arnold & Hodgson, 1991; Hultman *et al.*, 1989). Since biotin can be randomly incorporated into the oligonucleotide during synthesis (Hultman *et al.*, 1989), incorporation of biotin onto the reverse primer, followed by affinity purification using streptavidin Dynabeads, would ensure the recovery of only the full-length PCR product.

Dynabeads containing streptavidin covalently attached to the bead surface were obtained from Dynal (Norway). Streptavidin is a protein (M_r of approximately 66,000) which is made up of four identical subunits, each containing a high affinity binding site for biotin (KD = 10^{-15} M; Argarana *et al.*, 1986). During affinity purification the biotin binds to the streptavidin (in this case, only the full-length product will be biotinylated since only the reverse primer contains biotin; Figure 3.4) and the shorter PCR products are removed by washing steps. Cleavage of the desired product from the streptavidin is facilitated by the presence of a uracil residue which is added onto the 3' end of the reverse primer during synthesis. The enzyme uracil DNA glycosylase is then used to cleave the DNA at the uracil residue, thus releasing the full length PCR product.

The newly synthesised oligonucleotide (template A), made on the 2000 Å CPG column was found to be only slightly purer than that synthesised previously. It was therefore necessary to use the biotin purification described above to further purify the template.

The new template A was amplified by PCR using the reverse primer which had biotin attached. The full length PCR product was then purified using the streptavidin Dynabeads. The results of this purification method are shown in Figure 3.5.

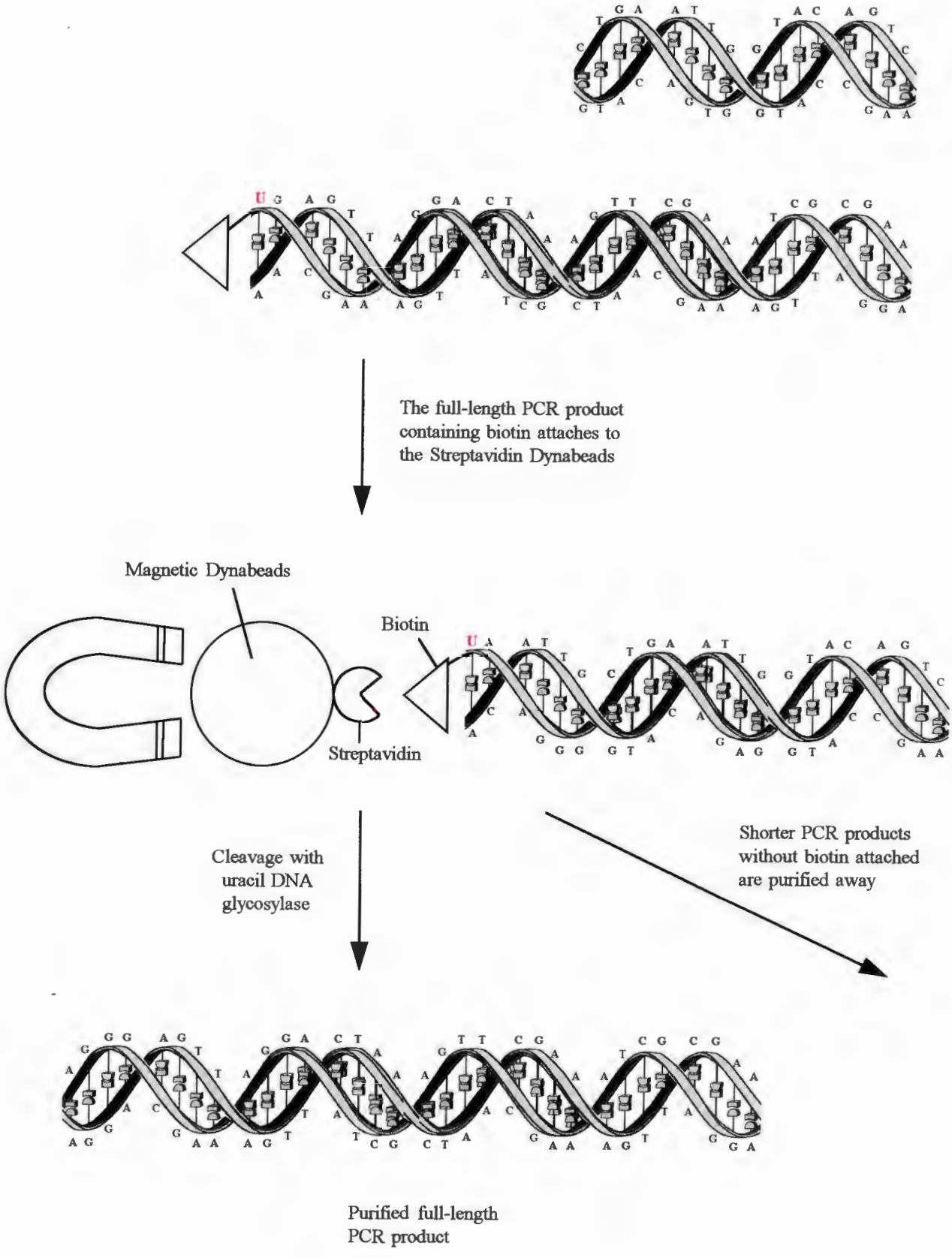


Figure 3.4: *A schematic diagram showing the purification of the full length biotinylated PCR product using streptavidin Dynabeads. The process is as described in the text.*

Uracil DNA glycosylase was used for the removal of the full length PCR product from the Dynabeads (lane 4). However, cleavage using this enzyme was found to be extremely slow and inefficient leading to poor yields of cleaved product. To overcome this problem, the PCR product was cleaved off the Dynabeads using the *Bam*H1 restriction site present at the end of the reverse primer for template A (lane 5), a process found to be far more efficient. The incubation time was also considerably less (1 hour as opposed to 4 hours with uracil DNA glycosylase).

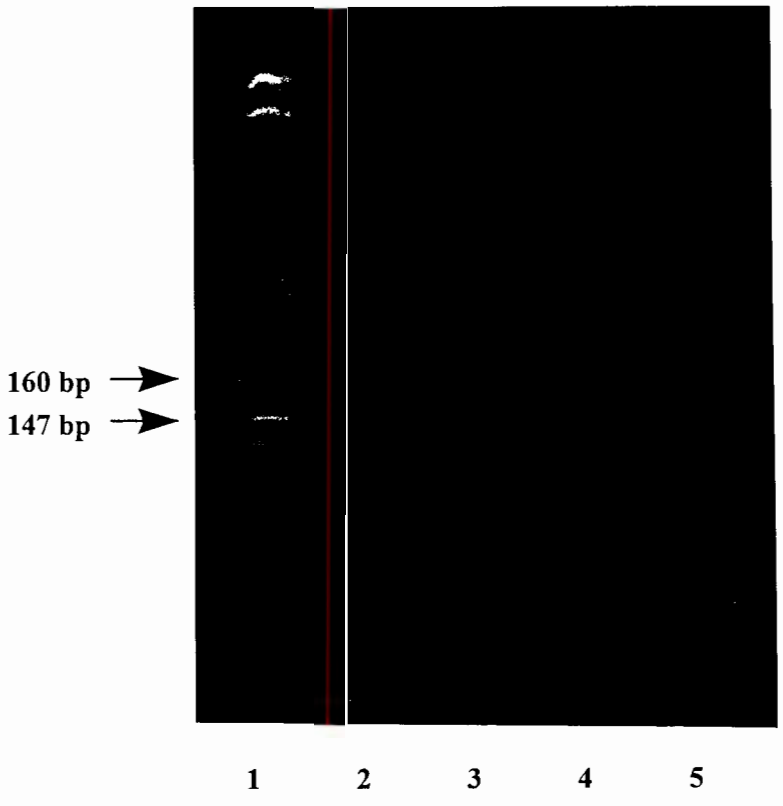


Figure 3.5: *10% non-denaturing polyacrylamide gel showing the results of affinity purification using streptavidin Dynabeads. Lanes 1, molecular weight markers, pBR322 digested with HpaII; 2, template A; 3, template A PCR product; 4, uracil DNA glycosylase cleaved PCR product; 5, BamH1 cleaved PCR product.*

This *Bam*HI-cut template A was then ligated to the dephosphorylated *Bam*HI-cut pUC18 vector and transformed into *E. coli*. Of the 37 colonies that appeared on the ampicillin-LA plates, 20 were selected for PCR screening. Of these, 12 colonies showed an increase in size by PCR screening (results not shown). Four of these were then selected and the plasmid purified by CsCl ultracentrifugation and sequenced.

DNA sequencing results showed that although the template used was now full length, it still contained errors. These were found to occur towards the 5' end of the DNA in the form of a deletion and two incorrect base substitutions. The error occurred in all four clones and extended to all subsequent clones tested. It was

concluded that these errors were more likely to have occurred during synthesis than during the PCR amplification since it is possible to amplify 3 - 6 kb of DNA with minimal errors (Ponce & Micol, 1991). Thus, although the higher pore glass size had increased the purity of the synthesised oligonucleotides, the templates were still too long for synthesis and as a result there was a loss of fidelity. This is especially true since the error occurred at the 5' end of the DNA which is where the final bases are added.

PART 2: Design of the Gene in Four Parts

3.5 Considerations in the Design of the Gene

Despite the small errors in the sequence being minor, they would nevertheless have major impacts. For example, there was a single G missing near the beginning of the sequence, which would have caused a frame shift and resulted in meaningless amino acid sequences. There are methods available to correct these errors such as site directed mutagenesis. However, since the resources were available for whole gene synthesis, it was decided to determine ways to improve this process. The following modifications to the cloning approach were made:

1. The gene was divided into even smaller sections which shortened the oligonucleotides to be synthesised. It has been shown that synthesis and cloning of a 516-bp synthetic *Erythrina* trypsin inhibitor was more successful if the DNA used was of shorter, more manageable lengths (Teixeira, 1992).

2. The synthesis process was improved by coupling the oligonucleotides to the larger pore size columns as before (i.e. 2000 Å).

3. A combination of anion exchange HPLC in denaturing conditions combined with non-specific hydrophobic dimethyl trityl (DMT) selection was used to purify the oligonucleotides. This step would ensure that the full-length template was separated from the failure sequences.

For HPLC purification, the DMT group (which is attached to the last base during the synthesis process) is not removed after synthesis but is kept attached. Hence only the full length sequences will be tritylated (Völker, 1993). During HPLC purification, this DMT group binds specifically (hydrophobic interaction) to the column matrix. The full-length sequences are thus retarded and elute last, well separated from any of the shorter 'trityl-free' failure sequences. With only the full length oligonucleotides present in the PCR reaction, more efficient PCR amplification could occur since it is the smaller failure sequences which are more likely to be amplified at the expense of the full length product as they occur in higher molar abundance.

To find restriction sites which would enable the gene to be divided into four parts, the gene was again scanned for partial restriction sites using the GCG program. The result was the introduction of three restriction sites

(*Xho*I, *Nae*I and *Sty*I) which divided the gene into four parts of 105, 48, 72 and 81 bases. These were named templates A, B, C and D respectively (Figure 3.6).

TEMPLATE A

```
          L  L  G  A  P  V  P  V  D  E  N  D  L  H  L  V  D  L
        PvuII - - - - - - - - - - - - - - - - - - - - - - - - - -
1       CAGCTGCTTGGTGCTCCAGTTCCAGTTGACGAAAACGACCTTCACCTTGTTGACCTT   57
        GTCGACGAACCACGAGGTCAAGGTCAACTGCTTTTGCTGGAAGTGGAACAACTGGAA

        A  R  F  A  V  T  E  H  N  K  K  A  N  S  L  L  E
58       - - - - - - - - - - - - - - - - - - - - - - -XhoI
        GCTCGTTTCGCTGTTACCGAACACAACAAAAAAGCGAACAGCCTGCTCGAG  108
        CGAGCAAAGCGACAATGGCTTGTGTTGTTTTTTCGCTTGTCGGACGAGCTC
```

TEMPLATE B

```
        L  E  F  E  K  L  V  S  V  K  Q  Q  V  V  A  G
        XhoI - - - - - - - - - - - - - - - - - - - - NaeI
109     CTCGAGTTCGAAAAACTTGTTAGCGTTAAACAGCAGGTTGTTGCCGGC      156
        GAGCTCAAGCTTTTTGAACAATCGCAATTTGTCGTCCAACAACGGCCG
```

TEMPLATE C

```
        A  G  T  L  Y  Y  F  T  I  E  V  K  E  G  D  A  K  K  L
        NaeI - - - - - - - - - - - - - - - - - - - - - - - - - -
157     GCCGGCACCCTTTATTATTTCACCATCGAAGTTAAAGAAGGTGACGCTAAAAAACTG   213
        CGGCCGTGGGAAATAATAAAGTGGTAGCTTCAATTTCTTCCACTGCGATTTTTTGAC

        Y  E  A  K  V
         - - - - StyI
214     TATGAAGCCAAGGTT     228
        ATACTTCGGTTCCAA
```

TEMPLATE D

```
        A  K  V  W  E  K  P  W  M  D  F  K  E  L  Q  E
        StyI - - - - - - - - - - - - - - - - - - - - - -
229     GCCAAGGTTTGGGAAAAACCATGGATGGACTTCAAAGAACTGCAGGAA      276
        CGGTTCCAAACCCTTTTTGGTACCTACCTGAAGTTTCTTGACGTCCTT

        F  P  V  D  A  S  A  N  A  STOP
         - - - - - - - - - - - - HindIII
277     TTCCCAGTTGACGCTTCTGCTAATGCTAAGCTTG      310
        AAGGGTCAACTGCGAAGACGATTACGATTCGAAC
```

Figure 3.6: *DNA and amino acid sequences of the HO templates showing the division of the gene into four fragments. Restriction enzymes were incorporated into the sequence by modification of the codon usage. The repeated amino acids are shown in bold as is the 'TAA' stop codon.*

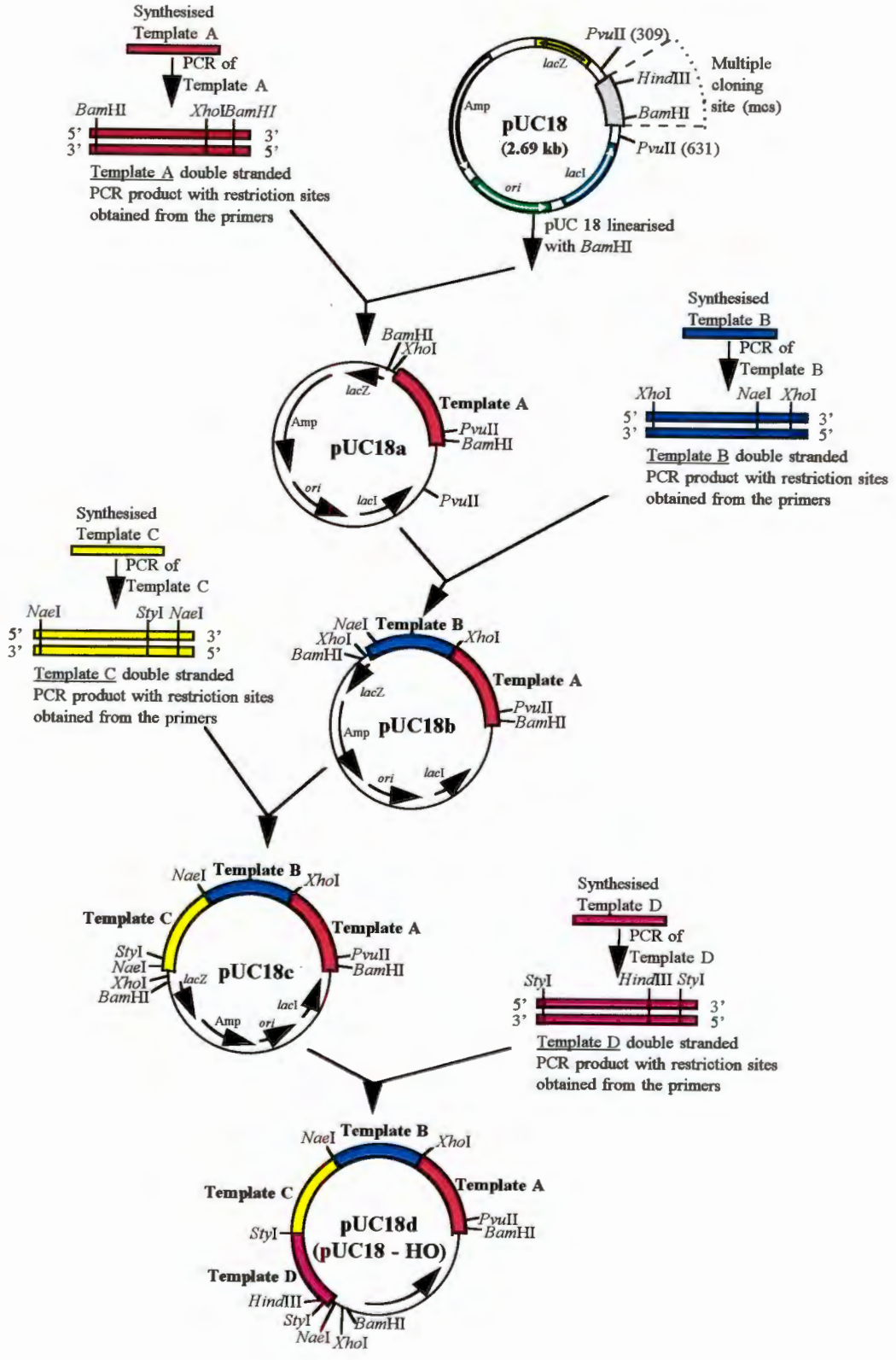


Figure 3.7: *Schematic diagram of the cloning procedure used to clone the gene in 4 parts into pUC18. Template A was amplified by PCR, cut with BamHI and cloned into the BamHI-linearised pUC18 to generate pUC18a. The template B PCR product was cut with XhoI and inserted into the XhoI linearised pUC18a to generate pUC18b. The template C PCR product was cut with NaeI and inserted into the NaeI linearised pUC18b to generate pUC18c. Finally the template D PCR product was cut with StyI and inserted into the StyI linearised pUC18c to generate pUC18d.*

Primers, which contained these restriction sites, were then designed as was done previously (section 3.2.1). For example, in the new template A, *Bam*HI was included in the forward and reverse primer sequences so that this template could be cloned into the *BamHI* site of the linearised plasmid. An *Xho*I restriction site was engineered into the reverse primer sequence, just before the *Bam*HI site enabling pUC18 carrying template A (pUC18a), to be linearised using *Xho*I for cloning in of template B. Template B had *Xho*I on each primer set so that it could be cloned in at this point. A *Nae*I restriction site was engineered into the reverse primer sequence of template B, just before the *Xho*I sequence. This enabled the new vector named pUC18b, to be linearised with *Nae*I for cloning in of template C, and so on (Figure 3.6). Primers were checked for any adverse secondary structure using the program 'Primer' (Version 0.5; Whitehead Institute for Biomedical Research, Cambridge, MA). Primer sequences are shown in Table 3.2.

Table 3.2: Primer Sequences and their restriction enzyme sites required for cloning

Primer	Sequence	Restriction Enzymes Present
Forward, Template A	5'CGCGTGCGCAGCA***GCTGCTTGGTGCTCCAGTTC*** 3'	*Bam*HI, *Pvu*II
Reverse, Template A	5' TCTGTGCGCAGGA***CTCGAGCAGGCTGTTC*** 3'	*Bam*HI, *Xho*I
Forward, Template B	5' CTAGGCGCT***CTCGAGTTCGAAAAAC*** 3'	*Xho*I
Reverse, Template B	5' CACATACTCGAGA***GCCGGCAACAACC*** 3'	*Xho*I, *Nae*I
Forward, Template C	5' CGTAT***GCCGGCACCCTTTATTA*** 3'	*Nae*I
Reverse, Template C	5' CTATGCCGGC***AACCTTGGCTTC*** 3'	*Nae*I, *Sty*I
Forward, Template D	5' CTAG***CCAAGGTTTGGAAAA*** 3'	*Sty*I
Reverse, Template D	5'ATACCAAGGTGGC***AAGCTTAGCATTAGCAGA*** 3'	*Sty*I, *Hind*III

The underlined sequence shows the restriction sites in the order as indicated in the last column.
The overlap region with the template is shown in bold and italics.

3.6 Results and Discussion

Template A was synthesised using a 2000 Å CPG column and was purified by HPLC (Figure 3.8). It is clear that despite the use of a larger pore size column, there was still a large proportion of failure sequences (lanes 3-10). The previous 147-bp template A is shown in lane 2. Lane 11, shows the purified 'DMT-containing' peak which eluted last off the ion exchange HPLC column. This DMT group was finally removed by the addition of 80% acetic acid.

The HPLC procedure was carried out for all four templates. Since they now consisted of only a single species, PCR was much more efficient. Figure 3.9 shows the full length PCR products for template B (lane 1), C (lane 3) and D (lane 5) in relation to the purified templates. Template A also showed a single full-length band, although it was not run on this same PAGE gel.

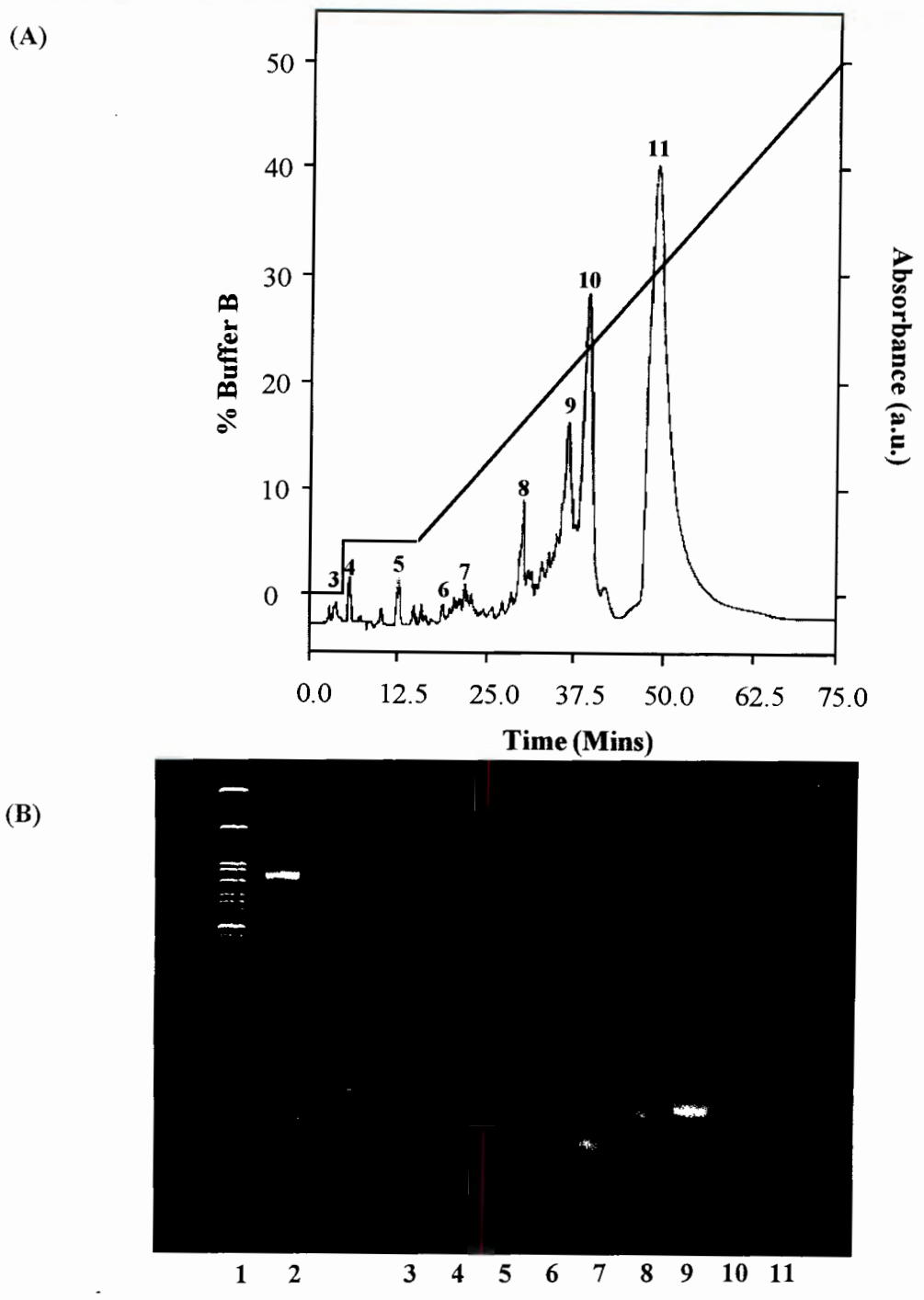


Figure 3.8: *Purification of oligonucleotides by HPLC using a Pharmacia Mono Q HR 5/5 ion exchange column.*

(A) HPLC profile of the purification. Elution was via a linear gradient of 3 M LiCl, 10 mM NaOH, pH 12.0. The peak numbers correspond to the lane numbers in (B).

(B) 10% non-denaturing polyacrylamide gel. Carrying 20 μl of samples eluted from the column, processed as described in Materials and Methods (section 3.3.3). Lanes 1, molecular weight markers, pBR322 digested with HpaII; 2; previous template A synthesised (section 3.4.1); Lanes 3 – 10, are the different peaks which eluted off the HPLC column. Lane 11, final 'DMT-containing' peak which is the full length template A.

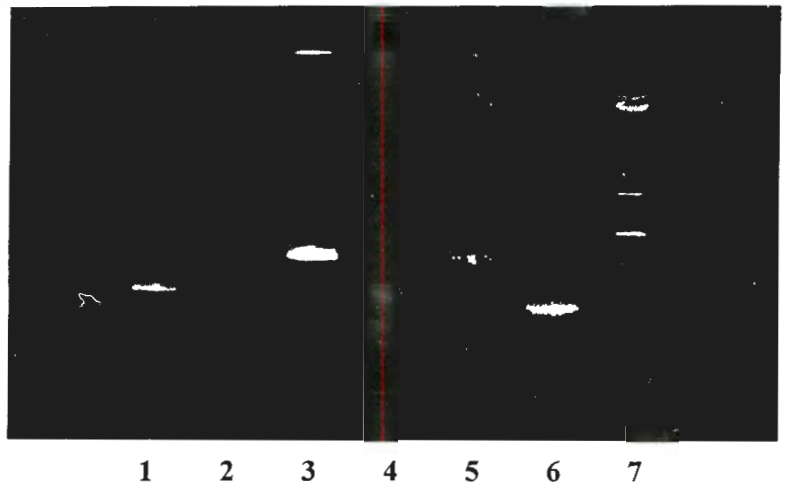


Figure 3.9: *10% non-denaturing polyacrylamide gel showing the full length PCR products in relation to the purified templates. Lanes 1, template B PCR product; 2, template B; 3, template C PCR product; 4, template C; 5, template D PCR product; 6, template D; 7, molecular weight markers, pBR322 digested with HpaII.*

After each successive HO portion was cloned in, the resultant plasmid was termed pUC18a to d, depending on the oligonucleotide most recently inserted (Figure 3.7). Because the templates contained the same restriction sites on each end, the orientation of the insert also had to be confirmed at each stage. This was done by restriction enzyme mapping of 'miniprepped' DNA or of the products of PCR screening. The insert DNA at each stage was then sequenced to confirm orientation and that the sequence was correct before continuing further.

Sequencing results showed that each cloned template was not only full length but was correct. Once all the templates had been cloned the gene was again sequenced from beginning to end using the forward and reverse pUC18 primers. This was to make a final check to ensure that there was no frame shift or error in the gene sequence.

In conclusion therefore, all 282 bases of the HO gene have been successfully cloned into the pUC18 vector. With regard to the synthesis of long lengths of DNA, the following conclusions can be drawn:

1. Sufficiently sized CPG column should be used to accommodate the required length of DNA to be synthesised.
2. The length of DNA used for synthesis should not be too long such that the fidelity of the sequence is compromised.
3. Long lengths of DNA (i.e. greater than 30 bases) should be purified, especially if used for PCR.

CHAPTER 4

Expression of the Hybrid Oryzacystatin Gene in *Escherichia coli*

PART 1: Cloning, Expression and Purification of the Hybrid Oryzacystatin in the pMAL System.

4.1 Introduction

High-level expression of eukaryotic genes in microorganisms has often proved difficult to achieve despite the use of strong promoters (Nagai & Thøgersen, 1984). The most common problems encountered with recombinant proteins include insolubility, aggregation or incorrect folding (Marston, 1986). To overcome these problems, many eukaryotic proteins have been efficiently produced as hybrids in which expressed genes are fused to the coding sequences of *E. coli* genes such as β-galactosidase.

A similar approach has been developed by New England Biolabs in the form of the pMAL vectors. In this system, target proteins are fused to a maltose binding protein (MBP; Maina *et al.*, 1988). This protein has a high affinity for amylose such that a cross-linked amylose resin can be used as an affinity matrix to purify high yields of the fusion protein in a single step. It is this pMAL system that was used for the expression and purification of the hybrid inhibitor. The characteristics of the pMAL vector system and the purification method are discussed more fully in the following section.

4.1.1 The pMAL Series of Vectors for High-level Expression of Foreign Genes in *E. coli*

In the pMAL vector system, the cloned gene is inserted downstream from the mal*E* gene, which encodes the MBP (New England Biolabs manual). This, as mentioned previously, results in the expression of a MBP fusion.

The *tac* promoter controls transcription of this fusion complex. Strong *rrnB* ribosomal RNA transcription terminators have been cloned downstream of the *lacZα* gene to prevent transcription from *ptac* from interfering with plasmid functions (Maina *et al.*, 1988; Figure 4.1). The promoter is kept in a switched off or repressed state by the *lac* repressor (the product of the *lacI* gene), which is also carried by the pMAL vectors.

In order to separate the target protein from the MBP, a sequence coding for the cleavage site of blood coagulation protease, factor Xa, has been inserted between the *malE* gene and the multiple cloning site (MCS;

Figure 4.1). This provides a mechanism for cleavage of the fusion protein, to release the target protein from the MBP (Maina *et al.*, 1988).

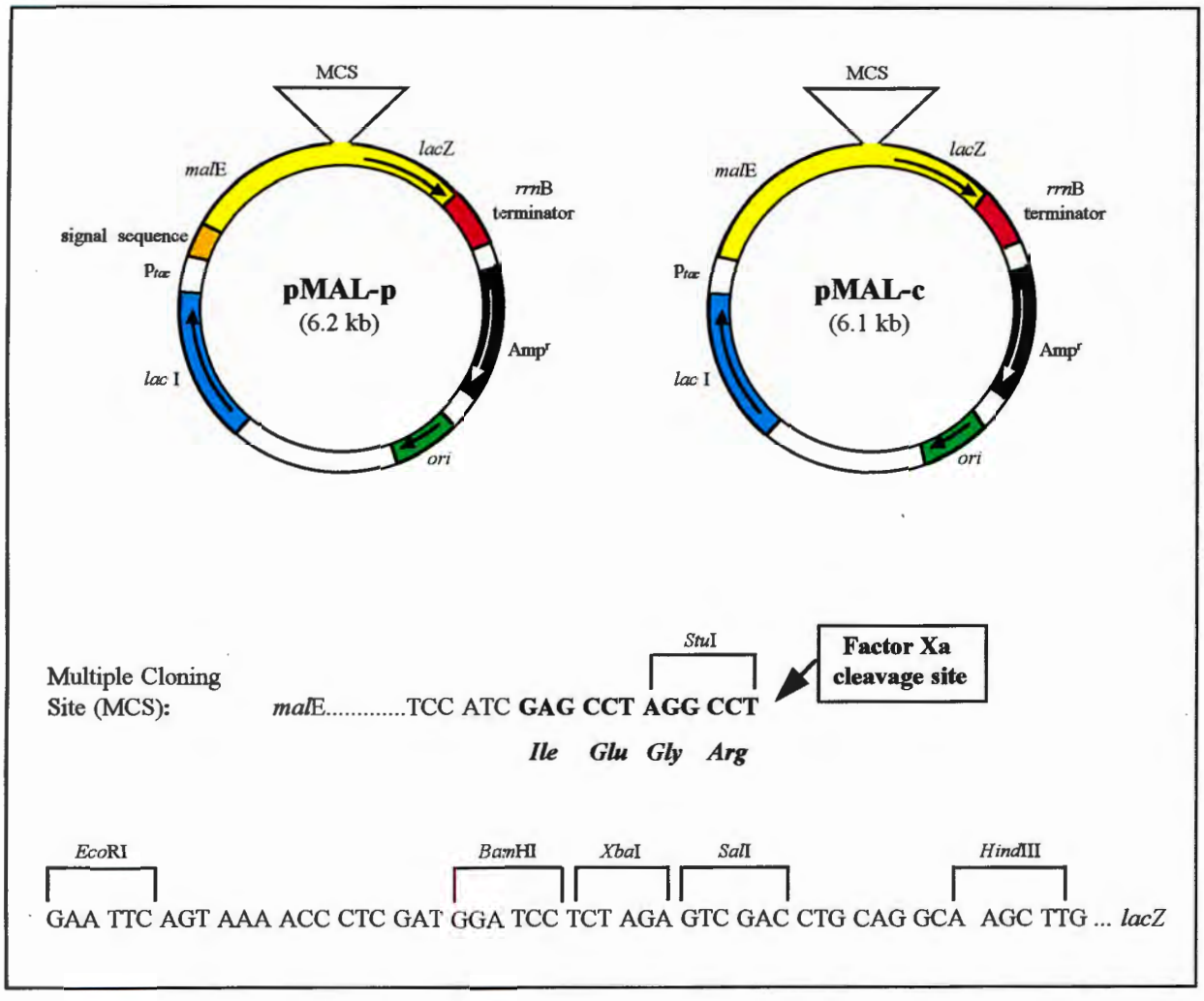


Figure 4.1: *The pMAL vectors, pMAL-p and pMAL-c. These vectors contain the inducible p_{tac} promoter positioned to transcribe a malE-lacZα gene fusion. The factor Xa cleavage site which is located near the StuI restriction site (of the multiple cloning site), is shown in bold. pMAL-p contains the signal sequence of the MalE gene for periplasmic expression.*

There are two types of vectors in this system, pMAL-p and pMAL-c (see Figure 4.1). For the purpose of this study, the pMAL-p vector was used. This vector contains the signal sequence of the m*alE* gene which directs the MBP fusion proteins to the periplasm. The pMAL-c vector lacks this signal sequence and the product thus remains localised in the cytoplasm. Such localisation may be undesirable if a protein contains disulphide bonds as the reducing environment of the cell prevents their formation. In this case, the HO has no disulphide bonds. However, since cystatins have been shown to be extremely susceptible to proteolytic degradation, periplasmic expression was considered preferable to expression in the cytoplasm.

The strategy for expression and purification of proteins using the pMAL system is shown schematically in Figure 4.2. The cleared lysate of induced cells is run through a cross-linked amylose column. After washing, the fusion protein is eluted with a dilute maltose solution. The purified fusion protein is then cleaved with the protease factor Xa and the target protein is purified away from the MBP domain by repassage through the cross-linked amylose column (Mania *et al.*, 1988).

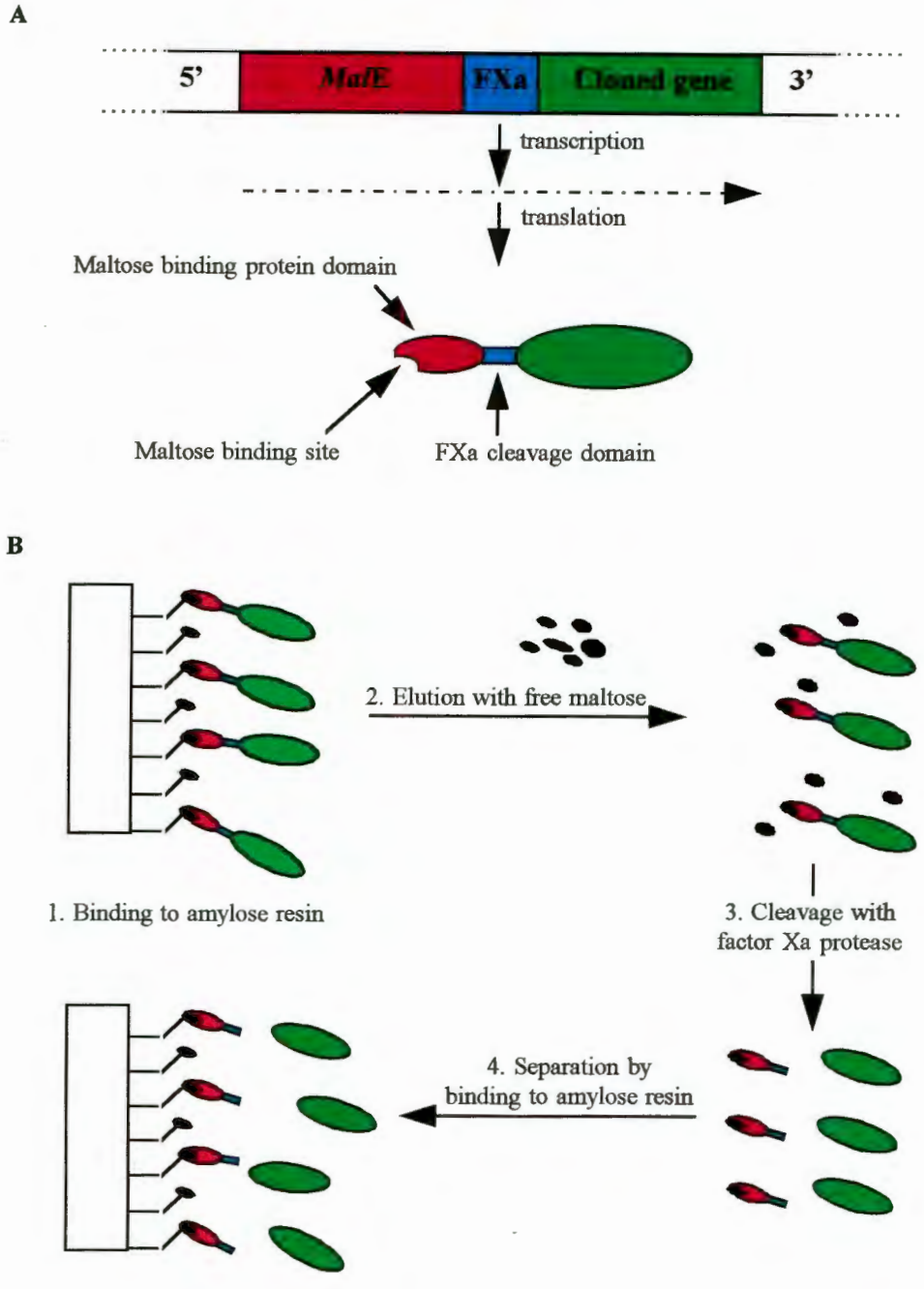


Figure 4.2: *Schematic diagram showing (A) a portion of a pMAL vector with the MalE gene, the coding region for factor Xa and a hypothetical gene cloned into the polylinker and the protein that results from expression of this gene. (B) the purification of the fusion protein from the cross-linked amylose resin. Fxa indicates the recognition site for factor Xa (adapted from Maina et al., 1988).*

4.2 Materials and Methods

4.2.1 *Escherichia coil* Strains and Culture Conditions

All plasmids constructed for protein expression were cloned into *E. coli* strain TBI. *E. coli* cells were grown at 37°C in Luria broth (LB) containing 0.2% glucose with shaking at 300 rpm. Strains carrying the plasmid were grown as above, with the addition of 100 μg/ml ampicillin (Sigma) to the medium.

4.2.2 Construction and Analysis of Recombinant Plasmids

Unless otherwise stated all molecular biology techniques were carried out according to Sambrook *et al.* (1989). The HO insert was removed from pUC18 by cutting with the restriction enzymes *Pvu*II and *Hind*III (Boehringer Mannheim). This 282-bp DNA fragment was gel purified from a 10% non-denaturing polyacrylamide gel as described in section 3.3.5.

For preparation of the pMAL vector, 10 μg of pMAL-p was cut with *Stu*I and *Hind*III and purified from a 1% agarose gel in TAE buffer (0.4 M Tris; 0.4 M Acetic Acid; 0.002M EDTA) using the GENECLEAN^R II kit (Bio 101 Inc.) according to the manufacturer's instructions. The 5'-phosphates were removed using calf intestinal phosphatase (Boehringer Mannheim) and ligated to the HO insert using T4 DNA ligase (Promega). Colonies carrying the correctly inserted plasmid, designated pMAL-HO, were detected by colony blots using a digoxygenin (DIG) labelled probe of the HO gene (see section 4.2.3). These were then confirmed by 'miniprepping' the DNA and digesting with various restriction enzymes. DNA sequencing was carried out using primers designed to anneal to the multiple cloning site of the pMAL vector. The designed sequences were as follows:

Forward primer: 5' GCAGCGACATTTTGCTGCCGTTCGGGCCCGCATTAGTAT 3'

Reverse primer: 5' TCGAGCTCGGTACCCGGCCG 3'

4.2.3 Colony Blot Hybridisation and DIG Detection

Colonies were replica plated onto two LB plates containing 100 μg/ml ampicillin. Following growth overnight at 37°C, Hybond N membranes (Amersham International) were blotted over the colonies onto the surface of the agar plates. The membranes were then placed onto filter-paper soaked in denaturing buffer (0.5 M NaOH; 1.5 M NaCl) for 3 minutes and onto filter-paper soaked in neutralising buffer (0.5 M NaCl; 1 mM EDTA; 0.5 M Tris-Cl, pH 7.0) for 7 minutes. The membranes were then washed in 5 × SSC (0.6 M NaCl, 0.06 M Sodium Citrate) for 5 minutes, allowed to dry and UV cross-linked at 595 nm for 5 minutes. A probe labelled with DIG (Boehringer Mannheim) was made by PCR amplifying the insert. The PCR method used is as described in Chapter 3, section 3.3.4, except that DIG-labelled dNTPs were used at concentrations as described by the

manufacturer (Boehringer Mannheim). The primers used for PCR were the forward primer to template A (the first template in the '4-part' cloning strategy) and the reverse primer to template D (the last template in the '4-part' cloning strategy). Detection of the DIG-labelled probe was as described by the manufacturer (Boehringer Mannheim).

4.2.4 Expression of the Hybrid Oryzacystatin Fusion Protein in *E. coli*

Overnight cultures of *E. coli* TBI transformed with parental (pMAL-p) or recombinant (pMAL-HO) plasmids were grown in one litre LB containing 2 g/l glucose. Cultures were grown at 37°C until the A_{600} reached 0.4 (approximately 2 hours). Isopropylthio-β-D-galactopyranoside (IPTG; Sigma) was then added to 0.3 mM. Preliminary studies of the time required for induction (see section 4.3.2) had shown maximum expression after 24 hours, so cultures were grown for a further 24 hours before isolation of the periplasmic fraction.

4.2.5 Periplasmic Fraction Isolation

Cells were grown at 37°C for 24 hours, pelleted by centrifugation at 5000 rpm for 20 minutes at 4°C and resuspended in 20% sucrose; 30 mM Tris-Cl, pH 8.0. This suspension was incubated for 20 minutes at room temperature with stirring. The cells were then pelleted, as before, resuspended in 5 mM $MgSO_4$ and incubated at 4°C for 20 minutes. The suspension was centrifuged and the supernatant kept as the 'cold osmotic shock' fluid.

4.2.6 Optimisation of Induction Time

The induction time with IPTG that gave optimum concentrations of fusion protein was determined as follows. Five millilitre samples were withdrawn from a one litre culture at various times after the addition of IPTG and these were used to prepare the cold osmotic shock fluid (periplasmic fraction) as described above. Twenty microlitre samples of this was used for SDS-PAGE gels. The concentration of fusion protein in the cold osmotic shock fluid was determined by incubation of 0.5 ml of the fluid with 0.5 ml of packed amylose resin in column buffer (0.2 M NaCl; 1 mM EDTA; 20 mM Tris-Cl, pH 7.4). This was then spun down at 14,000 rpm in a bench top centrifuge. Following washing with column buffer, the fusion was again spun down, and eluted from the amylose resin with 10 mM maltose. The protein concentrations in the eluate were determined using the Bio-Rad protein assay (Bradford, 1976) as described in Appendix C.

4.2.7 Column Purification of the Fusion Protein

The periplasmic extract was diluted 1:5 with column buffer (0.2 M NaCl; 1 mM EDTA; 20 mM Tris-Cl, pH 7.4) and loaded onto a pre-swollen amylose affinity column (2.5 × 10 cm) at a flow rate of 1 ml/min. The

column was then washed with 8 column volumes of column buffer to remove the last of any unbound proteins. The fusion protein was finally eluted with column buffer containing 10 mM maltose. Five millilitre fractions were collected using a fraction collector and the A_{280} of each 5 ml fraction was read on a Beckman DU-64 spectrophotometer. The inhibitory activity of the various fractions was determined as described in Appendix C.

4.2.8 Polyacrylamide Gel Electrophoresis (PAGE) and Western Blotting

Fusion proteins were separated on 13% SDS-PAGE gels with 6.4% stacking gels as described by Laemmli (1970; Appendix C). Electrophoresis was performed at 4°C at constant current (15 to 40 mA) until the bromophenol blue tracking dye reached the bottom of the gel. The proteins were identified by staining with Coomassie brilliant blue R-250 (Chen *et al.*, 1993).

The 10-kDa hybrid inhibitor was electrophoresed in triton gels using the loading buffer and procedure as described (Schägger & von Jagow, 1987). Electrophoresis was performed at 4°C at constant current (15 to 40 mA) until the tracking dye reached the bottom of the gel. Separated proteins were visualised using the silver staining method of Merril *et al.* (1983).

For Western blotting, proteins were electroblotted onto nitrocellulose and processed according to Towbin *et al.* (1979) with modifications (Appendix C). The antibodies used were:

- anti-OC I (a gift from Professor Soichi Arai, Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, Japan) used at a 1/1000 dilution.
- anti-MBP antibodies (New England Biolabs) used at a 1/10 000 dilution.

4.2.9 Digestion of the Fusion with Factor Xa

The fusion product, eluted from the amylose column, was diluted to 400 ml with 100 mM NaCl; 10 mM Tris-Cl, pH 8.0. This was then dialysed against 5 litres of 0.1 M NaCl; 10 mM Tris-Cl, pH 8.0, with a number of changes over 48 hours at 4°C. The dialysed protein was concentrated by ultrafiltration under pressure using a P10 membrane (Amicon Inc., U.S.A) and the concentration of the fusion protein was determined using the BioRad protein assay (Bradford, 1976; Appendix C). Factor Xa was then added at a ratio of 1:200 (factor Xa:fusion), and incubated for 12 hours at 25°C.

4.2.10 Purification by Ion Exchange Chromatography

Following digestion with factor Xa, the proteins were dialysed against 5 litres of column buffer (25 mM NaCl; 20 mM Tris-Cl, pH 8.0) at 4°C. A 1 × 10 cm column was packed with 6 ml of Q-Sepharose resin (Pharmacia)

and equilibrated in column buffer before placing the dialysed proteins over the column. The column was washed with 5 column volumes of column buffer and proteins were eluted using a salt gradient from 25 mM NaCl, 20 mM Tris-Cl, pH 8.0 to 500 mM NaCl, 20 mM Tris-Cl, pH 8.0. Fractions (2.5 ml) were collected and the A_{280} was read on a Beckman DU-64 spectrophotometer.

4.3 Results and Discussion

4.3.1 Construction of the Hybrid Oryzacystatin Fusion Protein

The synthesised gene was subcloned into the pMAL vector after removal from pUC18 by digesting with *Pvu*II and *Hind*III. This resulted in the removal of a 282-bp fragment (Figure 4.3) which was gel purified to separate the insert from the rest of the plasmid. This was particularly necessary since *Pvu*II cuts in the pUC18 plasmid (Appendix D) so the correct size band had to be isolated.

The recognition site for the factor Xa protease consists of four amino acids, Ile-Glu-Gly-Arg (Nagai & Thøgersen, 1988). The ideal cloning procedure would be to insert the gene of interest directly 3' to the coding sequence for factor Xa. To do this, the gene of interest must be cloned into the *Stu*I site in the vector multiple cloning site (New England Biolabs manual). By cutting the insert at the specifically engineered *Pvu*II site a blunt-ended fragment resulted, with the first amino acid (leucine) at the start of the gene. Cutting the vector and the insert with *Hind*III allowed directional ligation of the *Pvu*II-*Hind*III-cut insert into the desired position and in frame with the *malE* and factor Xa sequences.

The insert and vector were ligated and transformed into *E. coli* strain TBI by the method of Chung & Miller (1988). When the putative recombinants were transferred onto fresh LB plates containing 100 μg/ml ampicillin, 80 μg/ml 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal) and 0.1 mM IPTG, the majority of white colonies did not grow well. This could be attributed to the over expression of the foreign fusion protein on the plates. Some fusion proteins have been shown to be toxic to *E. coli* when IPTG and X-gal are included in LB plates, even at induction levels of 10 μM IPTG (New England Biolabs Manual). The identity of positive clones was thus further verified by colony hybridisation using a DIG-labelled probe of the HO gene. The DNA of a few of these positive clones was extracted using the 'miniprep' method (Sambrook *et al.*, 1989) and digested with restriction enzymes present in the insert (Figure 4.4). The *Stu*I and *Xba*I restriction sites (lanes 2 and 3) are lost from the multiple cloning site once the insert is cloned (see Figure 4.1) and thus would not be expected to cut. They do cut the pMAL-p plasmid, as is shown in lanes 6 and 7. The restriction enzyme, *Sty*I, was used to clone in template D of the insert in the 4-part cloning strategy (Chapter 3). It is not present in the pMAL-p vector (Appendix D) as can be seen from the uncut band (lane 5). The pMAL-HO plasmid does cut with *Sty*I (lane 1), however, which verifies the presence of the HO gene in the plasmid.

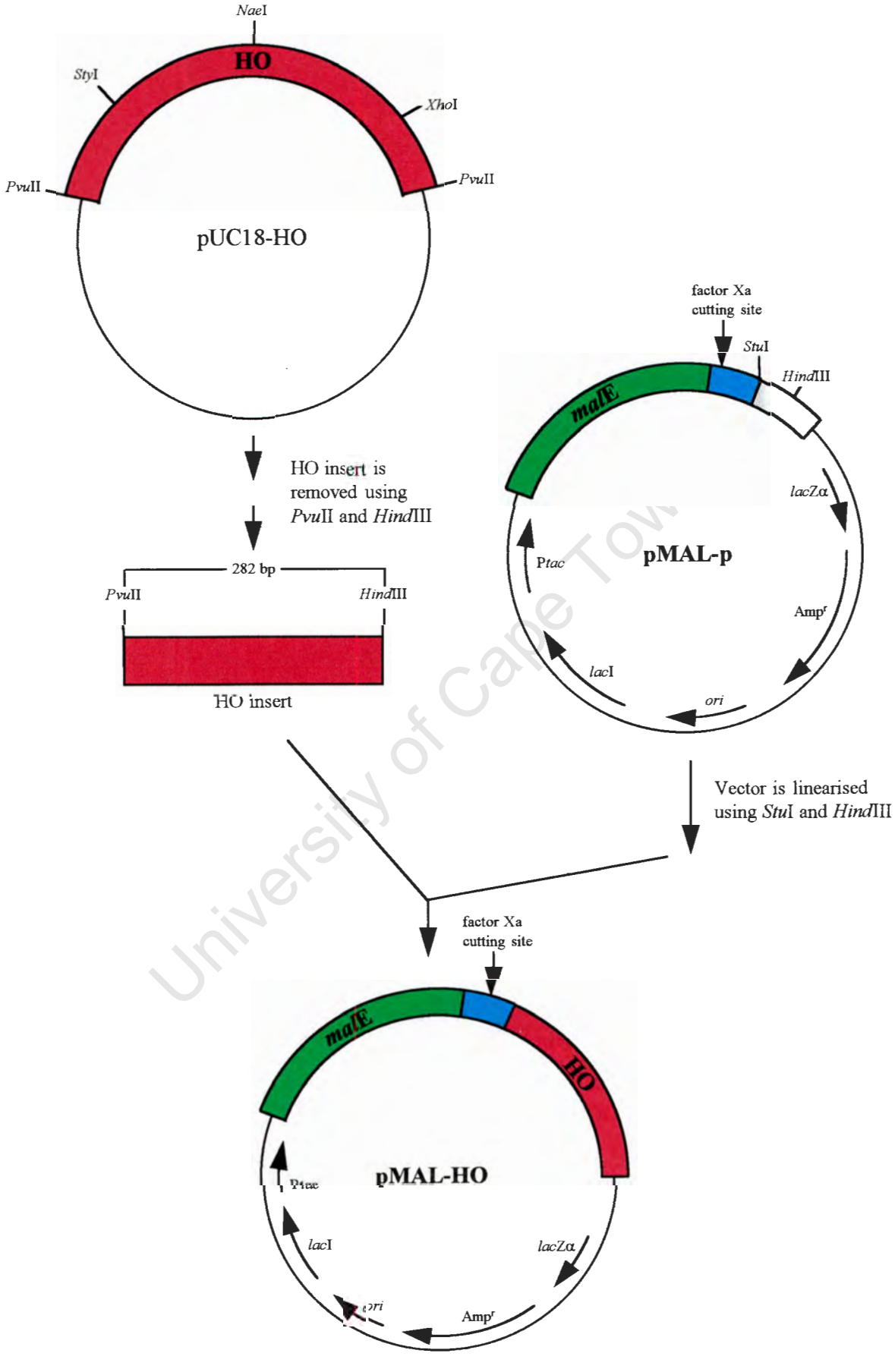


Figure 4.3: *Schematic diagram showing the cloning procedure. The insert was removed from pUC18-HO by cutting with PvuII and HindIII and ligated into the pMal-p plasmid (linearised with StuI and HindIII) to produce pMAL-HO.*

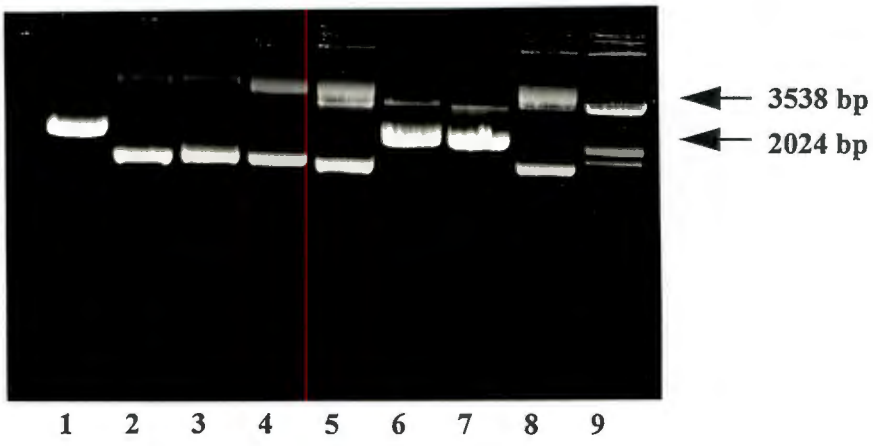


Figure 4.4: *1 % agarose gel showing plasmid DNA, pMAL-p & pMAL-HO, digested with various restriction enzymes. Lanes 1-4 are the recombinant plasmid pMAL-HO and lanes 5-8 are the pMAL-p plasmid only. Lanes 1 & 5, StyI cut plasmid; lanes 2 & 6, StuI cut plasmid; lanes 3 & 7, XbaI cut plasmid; lanes 4 & 8, uncut plasmid; 9, molecular weight markers λ digested with EcoRI and HindIII.*

The DNA was sequenced using the dideoxy chain termination method of Sanger *et al.* (1977) to ensure that the gene was correct and in frame with the factor Xa cutting sequence.

4.3.2 Analysis of Expression of the Fusion Protein

Expression was induced (section 4.2.4) by addition of IPTG to a mid-log phase culture and then incubated for a further 24 hours to allow for the accumulation of the fusion protein. Samples of the cold osmotic shock solution were taken before and after induction and analysed on a 13% SDS-PAGE gel to detect the presence of the fusion (Figure 4.5). No fusion product was observed for the controls, i.e. the uninduced and induced bacterial periplasmic fractions without plasmids (lanes 2 and 3) and fractions containing the pMAL-p (lanes 4 and 5). Bacterial periplasmic fractions (lanes 2 and 3) did show expression of *E. coli* periplasmic proteins as is seen in lanes 4 – 6, however, these were faint and did not photograph well.

On induction with IPTG, pMAL-p (lane 5) showed accumulation of a 42-kDa protein. This is close to the predited size of 42,700 Da (lane 8; Duplay *et al.*, 1984). In lane 7, there is an accumulation of a predominant product of approximately 52 kDa following induction of pMAL-HO with IPTG. This is in agreement with the expected size of 52 kDa for the fusion protein since the MBP is a 42-kDa protein and the HO protein is approximately 10 kDa. A MBP-HO fusion would not be expected to include the α–fragment of the *lacZ* gene as the stop codon at the end of the HO gene would prevent its inclusion during translation. Some induction can be seen in the uninduced lanes, the fusion protein (lane 6) and the MBP (lane 4), because the strong *tac* promoter often causes readthrough even in uninduced cultures. A similar result was observed when expressing a chitinase protein (*ChiA*) under the control of the *tac* promoter (Downing, 1997).

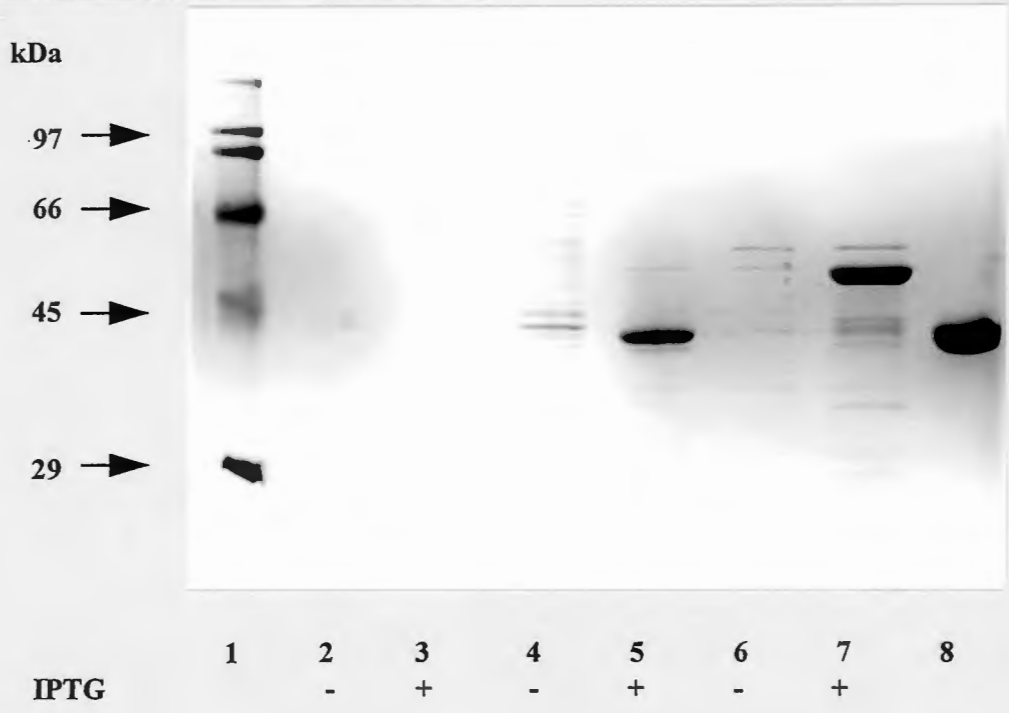


Figure 4.5: *13 % SDS-PAGE gel showing results of protein expression before and after induction with IPTG. Lane 1, molecular weight markers; lanes 2 & 3, the TBI periplasmic fractions; lanes 4 & 5, pMAL-p plasmid only; lanes 6 & 7, pMAL-HO periplasmic fractions; lane 8, MBP control. The addition of IPTG is indicated by a (+).*

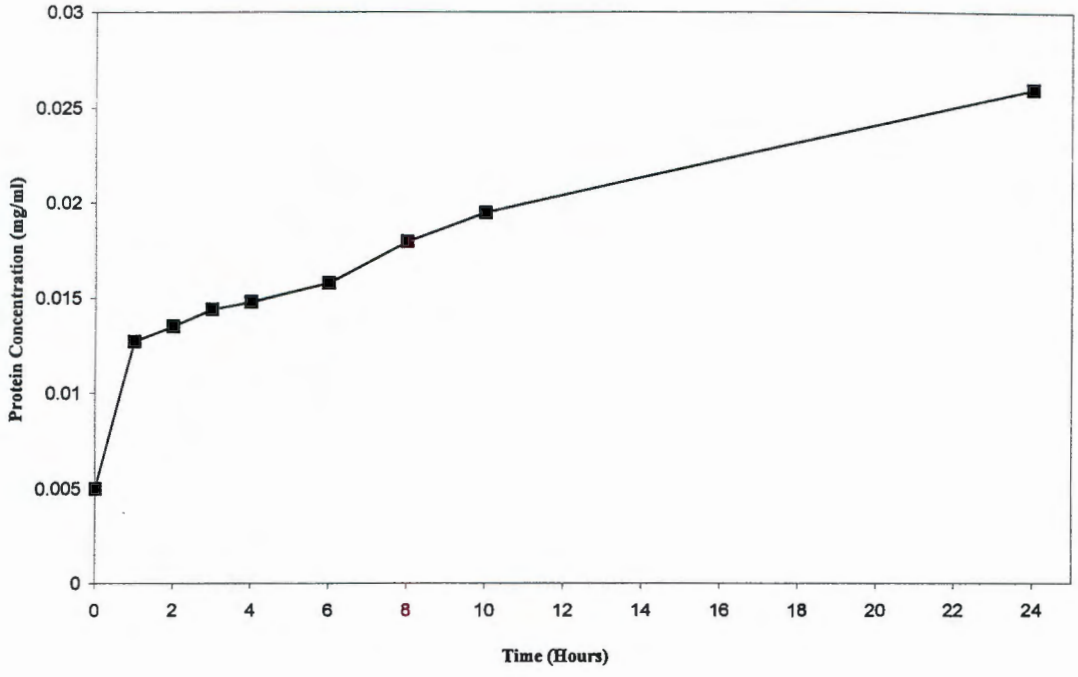
4.3.3 Optimisation of Induction Time

The experiment described in section 4.2.6 was conducted to determine the best time to harvest the cells following induction with IPTG. As can be seen from both the SDS-PAGE gel (Figure 4.6 (B); lanes 7 - 13) and the results of the Bio-Rad protein assays (Figure 4.6. (A)), the fusion protein showed optimal expression after 24 hours induction, well after stationary phase had been reached. A similar result was seen with the MBP (lanes 3 – 6). This pattern has also been shown with the expression of the Bt-toxin under the control of the *tac* promoter (Ge *et al.*, 1990). These authors suggested that the reason the longer incubation times further increased the amount of expressed protein, was that *E. coli* cells do not lyse in the stationary phase and therefore have more time to synthesise the expressed protein.

4.3.4 Western Blot Analysis of the Expressed Fusion Protein

SDS-PAGE gels were analysed by western blotting using antibodies to both the maltose binding protein (New England Biolabs; Figure 4.7) and to OC I (Figure 4.8).

(A)



(B)

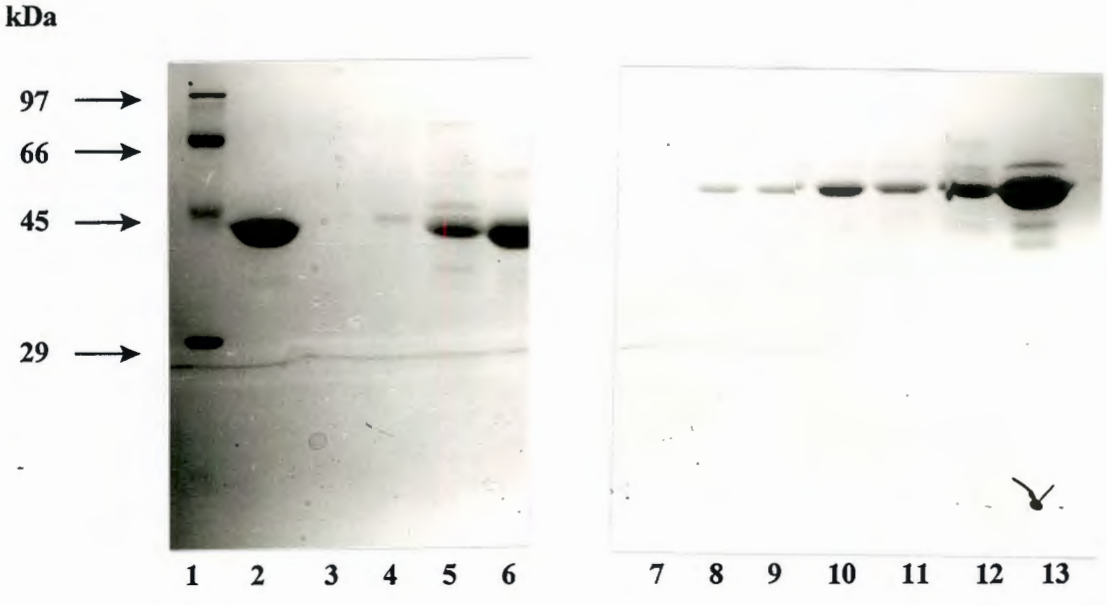


Figure 4.6: (A) Levels of expressed fusion protein over time. (B) 13% SDS gel showing fusion protein produced over time. 20 µl of the cold osmotic shock fluid was loaded onto the gel. Lanes 1, molecular weight markers; 2, MBP control; 3 - 6, pMAL-p expressed for 0, 1, 3 and 24 hours after induction; lanes 7 - 13, pMAL-HO at 0, 1, 2, 3, 5, 10 and 12 hours after induction with IPTG.

As expected, the antibody to the MBP bound to the MBP only (Figure 4.7; lane 1) and the pMAL-p expressing the MBP (lane 5). Since the fusion also contains the MBP, the antibody bound to the 52-kDa fusion protein in lanes 6 and 7. The uninduced lanes (lane 4, MBP and lane 6, fusion protein) show some expression, since as mentioned above, there is often read-through from the *tac* promoter. In lane 7, other bands lower than the 52-kDa fusion protein were visible, which were possibly degradation products. The middle of the three major bands were similar in size to the MBP and is either endogenous MBP or a cleavage product of the fusion protein. No MBP production was detected in the uninduced and induced bacterial periplasmic fractions containing no plasmid (lanes 2 and 3).

In the western blot carried out using antibodies to OC I (Figure 4.8), the OC I antibody bound to the fusion protein (lanes 6 and 7). More than one band is seen in lane 7, presumably due to premature termination of the fusion protein or its digestion with *E. coli* proteases (Maina *et al.*, 1988). As in the case with the MBP antibody, the uninduced periplasmic fraction showed a band (lane 6). This is as a result of read-through from the *tac* promoter. As expected, no bands were seen for the uninduced and induced bacterial fractions containing no plasmid (lanes 2 and 3), in the MBP only (lane 1), or the pMAL-p uninduced (lane 4) and induced (lane 5).

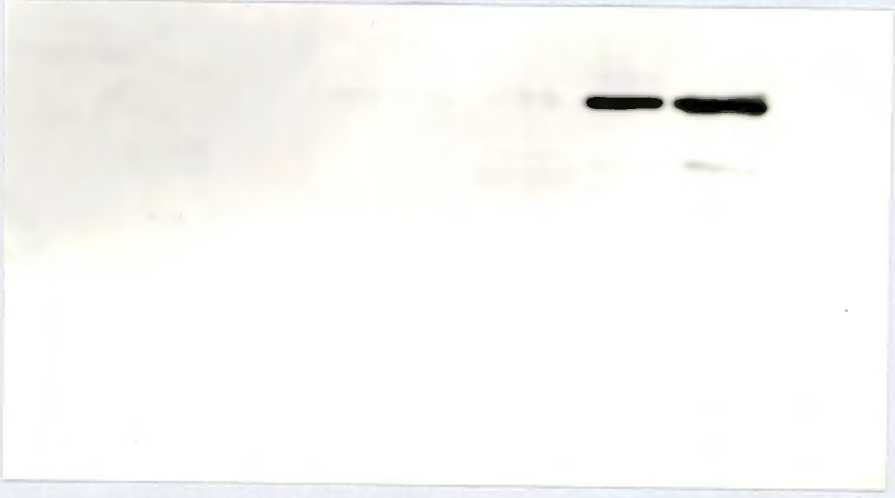
4.3.5 Purification of the Expressed Fusion Protein

A one litre culture of TB1 containing the pMAL-HO plasmid was induced, processed by isolating the cold osmotic shock fluid and purified on a column of cross-linked amylose. The elution profile is shown in Figure 4.9 (A). A large fraction of the protein eluted with the column buffer while the fusion protein remained attached to the column. On addition of 10 mM maltose, the fusion protein was eluted. The presence of the fusion protein in this final fraction was confirmed by inhibitory assays with papain (see Appendix C) and by SDS-PAGE gel analysis (Figure 4.9 (B)). Lane 2 shows the total protein before purification over the amylose column while lane 3 shows the purified protein fusion which eluted with the 10 mM maltose. This appears as a single band with a molecular weight of 52 kDa.

Crude extracts from one litre of cell suspension contained approximately 42 mg of periplasmic protein (Table 4.1). Following purification over the amylose column, 7 mg of pure fusion protein were isolated.

	1	2	3	4	5	6	7
IPTG		-	+	-	+	-	+

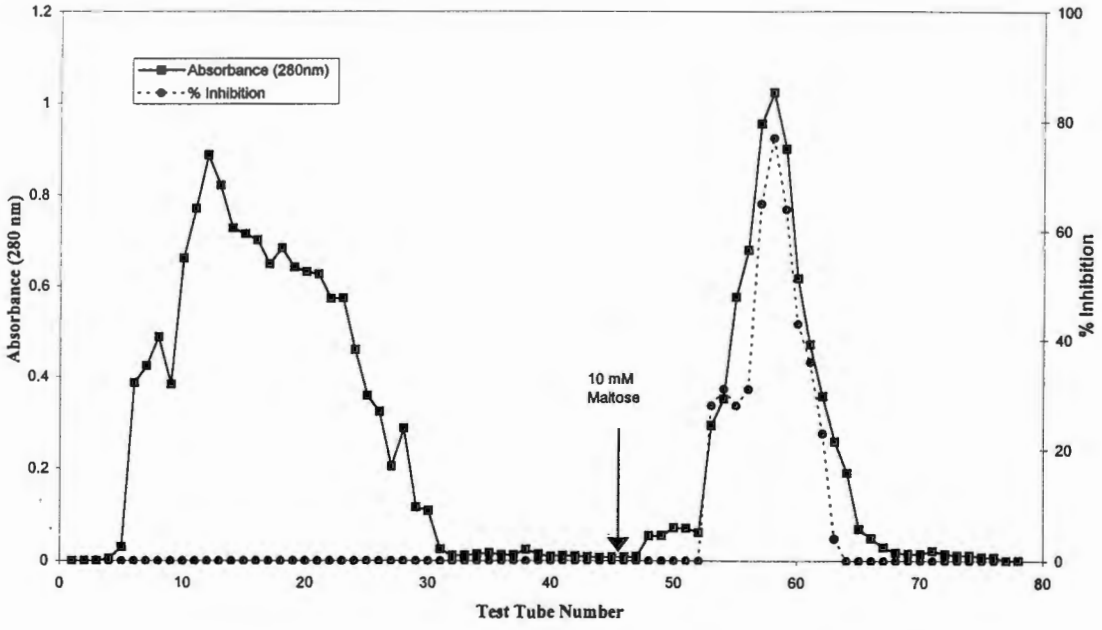
***Figure 4.7**: Western blot of a 13% SDS-PAGE gel. Proteins were detected using antibodies to the MBP. Lane 1, MBP control (New England Biolabs); TBI uninduced (lane 2) and induced (lane 3); pMAL-p uninduced (lane 4) and induced (lane 5); pMAL-HO uninduced (lane 6) and induced (lane 7). The addition of IPTG is indicated by a (+).*



	1	2	3	4	5	6	7
IPTG		-	+	-	+	-	+

***Figure 4.8**: Western blot of a 13% SDS-PAGE gel. Proteins were detected using antibodies to the natural OC I inhibitor. Lane 1, MBP control (New England Biolabs); TBI uninduced (lane 2) and induced (lane 3); pMAL-p uninduced (lane 4) and induced (lane 5); pMAL-HO uninduced (lane 6) and induced (lane 7). The addition of IPTG is indicated by a (+).*

(A)



(B)

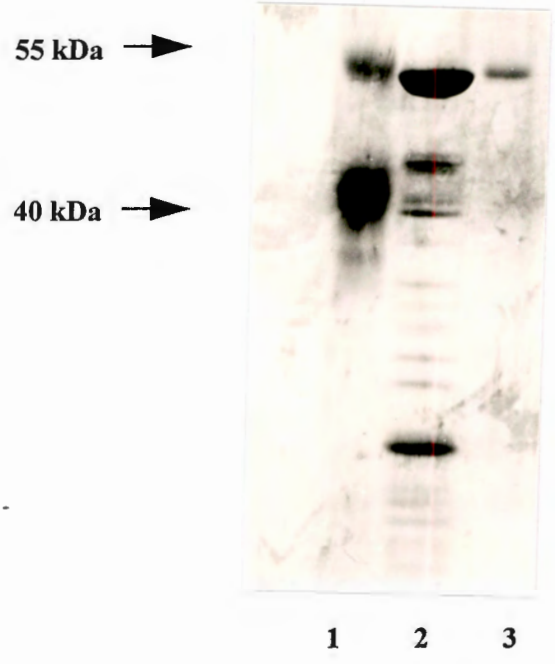


Figure 4.9: *(A) Cold osmotic shock proteins were loaded onto a 2.5 × 10 cm column equilibrated in 0.2 M NaCl; 1 mM EDTA; 20 mM Tris-Cl, pH 7.4. The fusion protein was eluted by the addition of 10 mM maltose. The % inhibition was determined as described in Appendix C.* ***(B)*** *Silver stained, 13% SDS-PAGE gel showing the results before and after purification on the affinity column. Lane 1, molecular weight markers; 2, total periplasmic proteins before placing over the column; 3, fusion protein eluted with 10 mM maltose.*

Table 4.1: Purification of the Fusion Protein.

Stage	Volume (ml)	Fusion Protein[b] (mg)	Activity (U)[c]	Specific Activity U/mg	Purification	Yield %
Cold osmotic shock[a]	400	42.5	85836	2019	1	100
Amylose column	30	6.9	19440	27811	13.7	23

[a] Data in the table refers to an initial starting volume of 400 ml of periplasmic extract after all other cellular proteins from the origional 1 litre culture have been removed.

[b] Measured as total protein by the Bio-Rad protein assay (Bradford, 1976; Appendix C).

[c] One unit is arbitrarily defined as the amount of inhibitor that can reduce the activity of 0.5 μM papain by 50% under the assay conditions.

4.3.6 Digestion of the Fusion Protein with Factor Xa Protease

The pooled, eluted fractions from the amylose column contained 10 mM maltose. This maltose had to be removed by dialysis before cleavage with factor Xa to enable further purification over the affinity column. Although experiments with fluorescent microscopy have shown that the rate of dissociation of the complex between maltose and the MBP is extremely high (Silhavy *et al.*, 1975), to enhance diffusion of maltose from the dialysis bag, the protein solution was diluted to 200 ml with 100 mM NaCl; 10 mM Tris-Cl, pH 8.0, giving a concentration of approximately 1.2 μg/ml. This was then dialysed against the above buffer (used for dilution) and concentrated by ultrafiltration under pressure with a 10-kDa molecular weight cut-off membrane (Amicon Inc., U.S.A).

To separate the HO from the MBP the fusion protein was incubated with factor Xa. In order to determine the optimal time for digestion, the fusion protein was added at a ratio of fusion:factor Xa of 200:1. Samples were incubated at various time intervals at 25°C and then analysed on a triton SDS-PAGE gel (Schägger & von Jagow, 1987; Figure 4.10). Lanes 4 - 7 show the digestion of the fusion protein with factor Xa. Over the time period of the digestion the 52-kDa protein band disappeared, and two new lower bands of 42-kDa and 10-kDa (lower arrow) appeared, representing the two domains of the MBP-HO fusion. It appears that the most efficient cleavage occurred after 12 hours incubation at 25°C. Trial digests at 37°C and 42°C showed no increase in the efficiency in the time required for or in the efficiency of cleavage (results not shown). Similar results were obtained even after incubation for 12 hours with a 5-fold higher concentration of enzyme (results not shown). The fact that the fusion protein was cleaved into its two domains under native conditions suggests

that the 4-amino acid recognition site was accessible to the protease. Furthermore, factor Xa did not appear to cleave the MBP or the hybrid oryzacystatin as no further cleavage products were obtained.

From the SDS-PAGE gel, it appears that the amount of cleaved product (the 10-kDa inhibitor) is extremely low. Guan and Dixon (1991) showed that cleavage of many proteins in their laboratory with factor Xa was slow and often extremely inefficient. Other researchers (Dr. John Hastings, personal communication) have made similar observations, which suggests that there may be some inherent problems with this cleavage system.

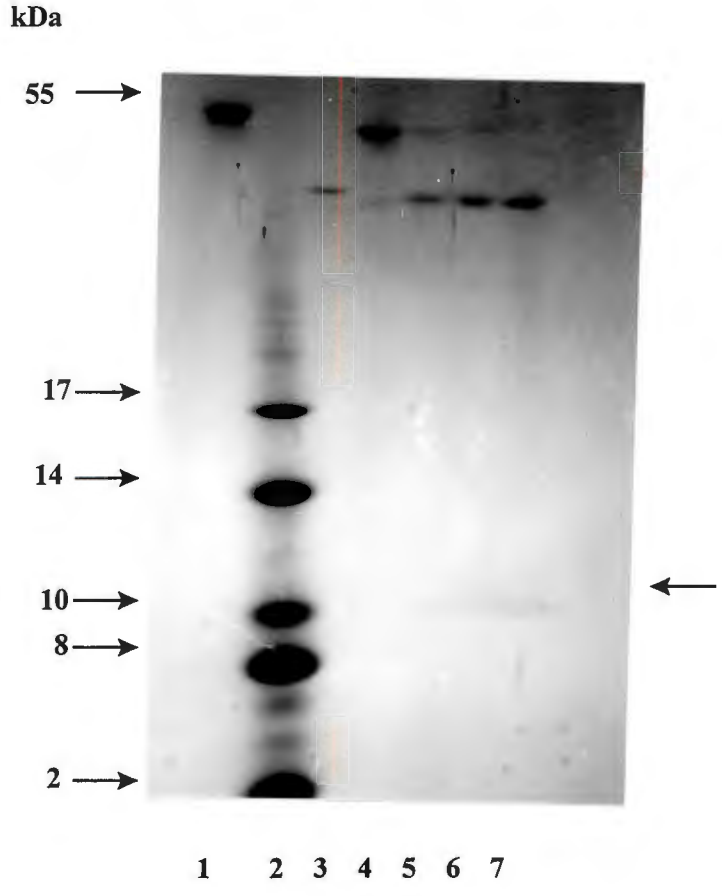


Figure 4.10: *Triton-SDS gel showing the results of factor Xa cleavage with time. Lane 1, 55 kDa molecular weight marker; 2, small molecular weight markers; 3, MBP; lanes 4 – 7, digests over 0, 2, 6 and 12 hours respectively. The arrow indicates the position of the HO inhibitor.*

4.3.7 Purification of the HO Inhibitor by Ion Exchange Chromatography

Rather than placing the factor Xa-cut fusion protein over the amylose column as described by Mania *et al.*, 1988, it was decided to purify the HO by ion exchange chromatography. The advantage of this method is that it allows for separation of not only the MBP, but also the factor Xa protease, from the HO. This is necessary for N-terminal sequencing which requires a pure protein.

Three protein peaks were obtained and the column fractions for each were pooled and dialysed against 5 mM phosphate buffer to remove the NaCl. The different peaks were then analysed by SDS-PAGE (Figure 4.11). Lane 2 (control) is a partial digest of the MBP-HO fusion showing the three possible band sizes. The uppermost band is the uncut fusion protein, the second and third are the two cleaved domains of the fusion protein the MBP and the HO inhibitor respectively. Lane 3 shows the purified HO which eluted between 0.025 M and 0.1 M NaCl. Lane 4 is the MBP. This eluted between 0.1 and 0.15 M NaCl while the factor Xa protease, the concentration of which was too low for the protein to be visualised on the gel, eluted at approximately 0.4 M NaCl. A total yield of 0.025 mg of pure HO inhibitor was obtained.

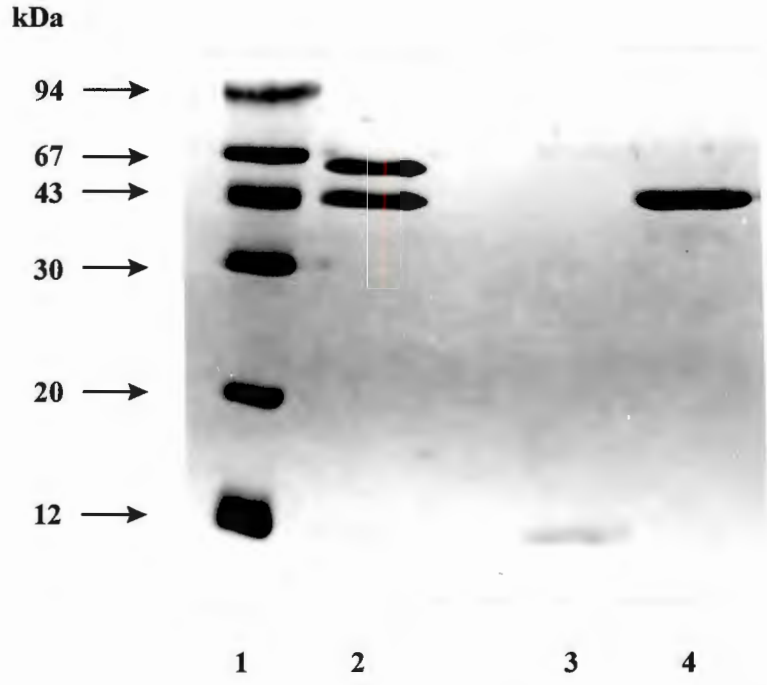


Figure 4.11: *13% SDS-PAGE gel showing the results of ion exchange protein purification. Lanes 1, molecular weight markers; 2, partial digest of the fusion protein with factor Xa; 3, purified HO inhibitor – eluting between 0.025 M and 0.1 M NaCl; 4, MBP – eluting between 0.1 M and 0.15 M NaCl.*

N-terminal sequencing of the purified HO inhibitor was then carried out to ensure that the factor Xa protease had produced the required N-terminus for the HO inhibitor. For this, the Edman degradation sequencing method was used. In this method, the α-amino group of the polypeptide chain is coupled under alkaline conditions to phenylisothiocyanate which allows the terminal amino acid to be cleaved from the remainder of the peptide under acidic conditions. The resulting amino acid derivative (an anilinothiazolinone) is converted into the more stable phenylthiohydantoin (PTH) form which can be analysed by HPLC while the remaining peptide is subjected to further degradation signals.

Sequenceing of the ion exchange sample showed high background from glycine and Tris contaminants and thus only the first two leucine residues and a proline residue in the fifth cycle could be determined with any accuracy. Purification by HPLC on a C_{18} reverse phase column would have been a better option as it would have removed the contaminants. However, since the sequence (obtained by subtracting the background from the previous cycle) appeared to be correct and since factor Xa, had cut at the correct cutting site in the pMAL system under the same experimental conditions as described in section 4.2.9 (Pei-yin Ma, personal communication; Teixura, 1992), and K_i data obtained with the cut inhibitor was in agreement with the results determined by Urwin *et al.*, 1995b using a similar inhibitor, the sequence was assumed to be correct.

PART 2: Assessment of the Inhibitory Capabilities of the Fusion Protein

4.4 Introduction

Although many proteins expressed as fusion hybrids have been shown to be biologically active (Smith & Johnson, 1988), it is questionable as to whether the 10-kDa HO protein would still be active and able to bind to the cysteine proteinase when attatched to such a large 42-kDa MBP which can offer steric hindrance.

Contrary to expectations, however, a survey of the literature showed that cystatin fusions are, in fact, biologically active. For example, Michaud *et al.* (1994) expressed oryzacystatins I and II as fusion proteins linked to glutathione-S-transferase (GST), a 26-kDa protein. These oryzacystatin fusions were found to be biologically active when assayed with papain. Similarly, Thiele *et al.* (1990) constructed inhibitor fusions of stefin B and mutant variants with MS-2 polymerase. These fusions were found to be tight-binding inhibitors of papain (Thiele *et al.*, 1990) as were stefin A fusions with adenylate kinase (Kaji *et al.*, 1990).

We thus proceeded to test the MBP-HO protein for its ability to inhibit papain in the hope that should the fusion protein show activity against cysteine proteinases, it could be used in biological control assays and so provide the convenience of a more abundant supply than was available as purified, cleaved inhibitor.

4.5 Materials and Methods

4.5.1 Fluorimetric Assays to Test for Activity

To test for activity of the fusion protein, 200 nM fusion inhibitor was incubated with 300 nM papain in papain assay buffer (Appendix B) in a total volume of 140 µl at 30°C for 30 minutes. 10 µl of this incubation mix was then removed and assayed with 5 µM substrate (Z-Phe-Arg-AMC; Sigma). Activity was detected using a Perkin-Elmer MDF-43A fluorescence spectrophotometer at an excitation wavelength of 380 nm and an emission wavelength of 460 nm.

Distilled water or a solution of a non-cystatin fusion protein (i.e. MBP-phosphinothricin acetyl transferase (PAT); a gift from Kenneth Palmer, Microbiology Department, University of Cape Town, Cape Town, South Africa), replaced the MBP-HO in negative control assays. The PAT protein is a 22-kDa protein from *Streptomyces hygroscopicus* (Thompson *et al.*, 1987) which was also expressed using the pMAL system and recovered as a fusion protein.

4.5.2 Assays to Determine whether the Inhibitor was Cleaved by Papain

A modification of the method of Machleidt *et al.* (1995) was used for these assays. Papain (0.4 µM) was incubated with the MBP-HO fusion (4 µM) or the MBP-PAT fusion (4 µM) in papain assay buffer (Appendix B) at 30°C. The incubation was carried out for various lengths of time as indicated in Figures 4.13 (A) and (B). The reaction was stopped at each stage by the addition of 1 mM iodoacetate. Results were analysed by SDS-PAGE gels (Appendix C) and fusion bands were scanned using a Hoefer GS300 scanning densitometer.

4.5.3 Gelatin Gel Assays to Test the Stability of the Inhibitor/Enzyme Complexes

Cystatin/papain complexes were subjected to mildly-denaturing gelatin-PAGE gels as described by Michaud *et al.* (1993) and Heussen & Dowdle (1980). For this assay, various cystatin inhibitors (50 µM) were incubated with 2 µM papain and 25 µl papain assay buffer (Appendix B) for 15 minutes at 37°C. Following incubation, this mix was added to 6 µl of gelatin-PAGE sample buffer (Heussen & Dowdle, 1980). Negative controls for inhibitory activity were prepared by incubating the proteinase with both 1 mM PMSF and 50 µM soybean trypsin inhibitor (SBTI). Positive controls for proteinase activity consisted of papain incubated with 50 mM Tris-Cl, pH 7.0 instead of the inhibitors.

Proteins were then electrophoresed on polyacrylamide gels containing 0.1% w/v gelatin. To avoid migration of the gelatin out of the resolving gel, the ratio of acrylamide to N,N'-methylenebisacrylamide was adjusted to

29:1.1. Gels were run overnight at constant current (18 mA) at 4°C and transferred to a 25% v/v aqueous solution of Triton-X-100 for 30 minutes at room temperature, to allow for renaturation of proteinases. The gels were then immersed in papain assay buffer (Appendix B) for one hour at 37°C. Gels were then stained in Coomassie Brilliant blue R250 stain as described by Chen *et al.* (1993).

4.5.4 Preparation of the Fusion Protein for Insect Trial Assays

Cold osmotic shock proteins of parental (pMAL-p) or recombinant pMAL-HO plasmids, were isolated as described previously (sections 4.2.4 and 4.2.5). These were then freeze-dried. Samples of the cold osmotic shock protein were analysed on 13% SDS-PAGE gels to estimate the expression levels. The amount of fusion or MBP present in the total periplasmic fraction was quantitated using a Hoefer GS300 scanning densitometer.

4.5.5 Insect Assays

Insect assays were carried out by D. Van der Linde at the South African Bureau of Standards (S.A.B.S.), in Pretoria according to the following protocol (D van der Linde, personal communication):

The test was conducted in plastic containers (top diameter, 120 mm; bottom diameter, 90 mm and depth 65 mm). Five replicates were used for the controls (MBP only and the no protein control), and the experiment (HO-MBP fusion protein). 200 g of a standard breeding medium mix was placed in each test container. This consisted of 2250 g white bread flour, 180 g fish meal and 270 g full cream milk powder. 0.1% of the fusion protein (equivalent to 400 mg of the MBP-HO fusion periplasmic fraction, and 400 mg of the MBP periplasmic fraction) was dissolved in distilled water and sprayed onto the surface of the rearing mixture using an Aerograph spray gun. For the untreated controls, distilled water was sprayed onto the surface of the breeding medium and mixed in. Protein-containing samples were seen to form a crust on top of the medium when dry. This crust was then ground to a fine powder using a coffee blender, and mixed in with the medium. Twenty adult test insects of the confused flour beetle (*Tribolium confusum*) were placed in each container and the containers were covered with glass panels to prevent the insects escaping. Observations for larval activity and F1 adults were made.

4.6 Results and Discussion

4.6.1 Fluorimetric Assays

The MBP-HO was tested for inhibitory activity as described (section 4.5.1) and the results are shown in Figure 4.12. From the figure, it is clear that the MBP-HO-fusion protein inhibited papain whereas 200 nM of the MBP-PAT fusion protein and 200 nM MBP failed to inhibit the enzyme.

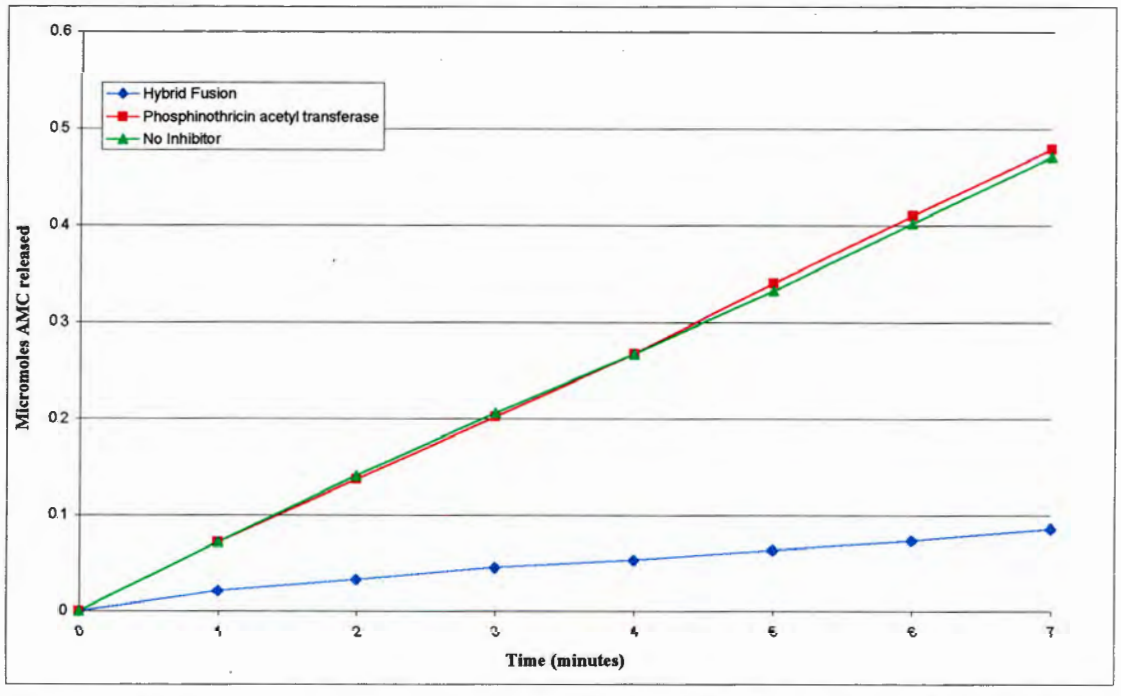
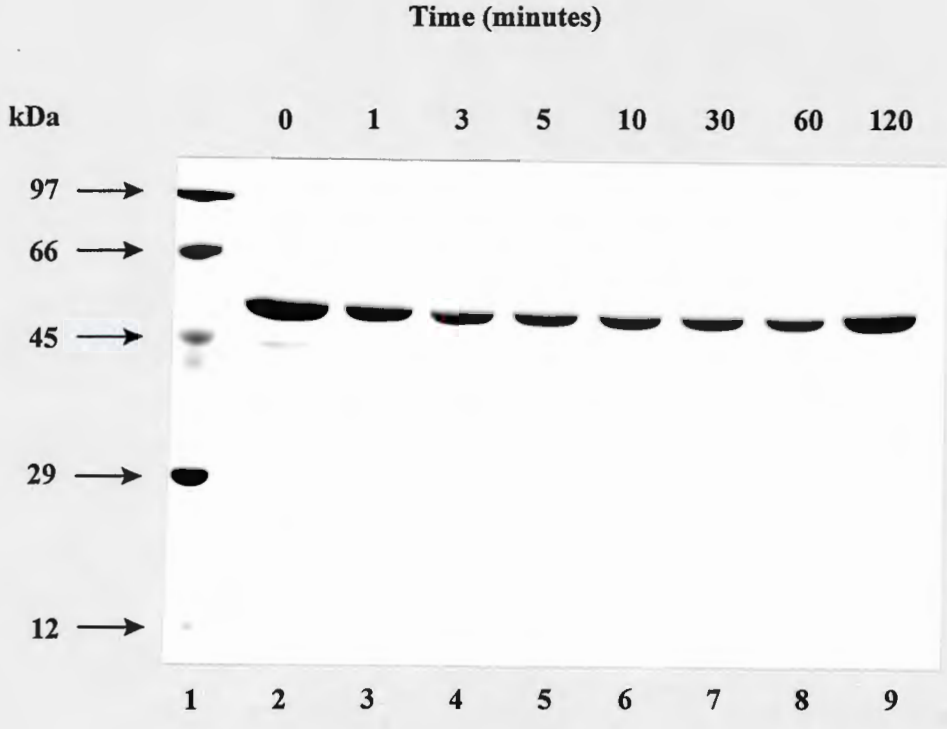


Figure 4.12: *Enzyme activity profiles of papain incubated with and without fusion proteins*

To see if the MBP-HO protein was acting as a substrate for papain, the MBP-HO fusion (4 μM) was incubated with papain (0.4 μM) for the different times shown in Figure 4.13 (A) and the reaction was stopped by the addition of 1 mM iodoacetate. As can be seen from the 13% SDS-PAGE gel, there appears to be no digestion of the MBP-HO, even after 60 minutes, a far longer period than used in the fluorimetric assay. Densitometric scanning of the PAGE gel showed no change in the relative concentrations of the bands with time.

To determine whether the MBP-PAT fusion product showed any difference in results, 4 μM of this fusion protein was incubated at the same time intervals as used for the MBP-HO fusion. Results showed complete digestion of this fusion protein even after 5 minutes incubation with papain (results not shown). Hence this experiment was repeated using a shorter time scale. The results are shown in Figure 4.13 (B). In this case digestion occurred even after 10 seconds incubation with papain and the fusion was completly digested after 2 minutes. On cleavage with papain there is the appearance of an approximately 42-kDa band which is the size of the MBP. The cleavage site is possibly at the start of the PAT sequence since no cleavage sites for papain (i.e. a Phe, Tyr, Val or Leu at the P_2 position; Berger & Schechter, 1970) were found in the MBP sequence. A

(A)



(B)

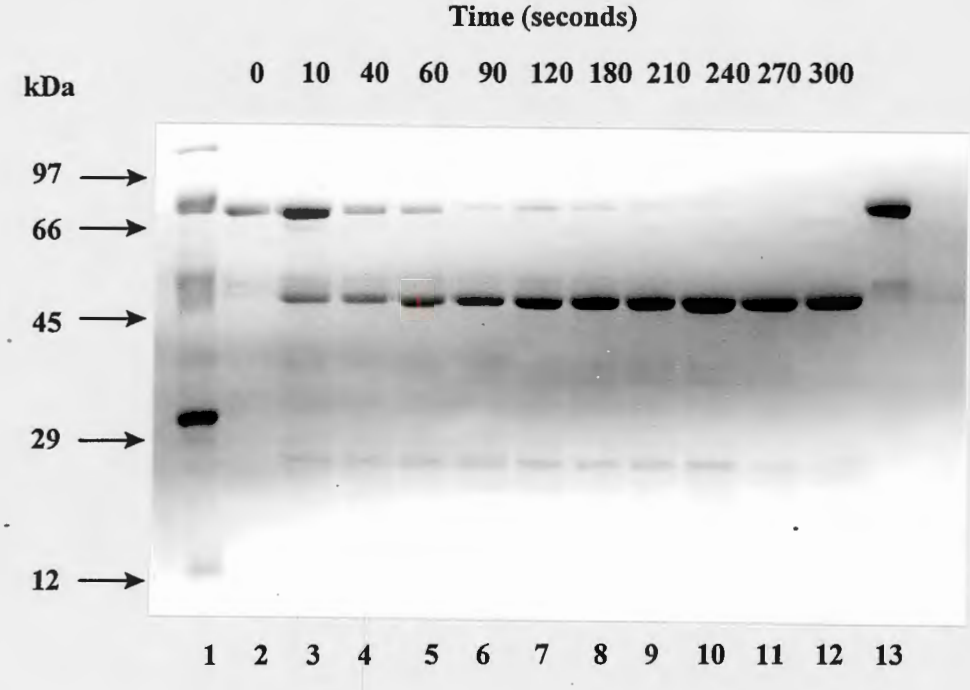


Figure 4.13: *(A) 13% SDS-PAGE gel showing results of incubation of the fusion protein with papain at 37℃. The incubation times are indicated (lanes 2 - 9) and the reaction was stopped by the addition of 1 mM iodoacetate. Lanes 1, molecular weight markers.* ***(B)*** *13% SDS-PAGE gel showing results of incubation of the MBP-PAT fusion protein with papain as described above. The incubation times are indicated (lanes 2 -12). Lanes 1, molecular weight markers; 13, MBP-PAT without papain.*

22-kDa band can be seen which is the size of the PAT protein (Thompson *et al.*, 1987) and is similar to papain (M_r 23 000 Da; Cunningham, 1965).

A further point that must be noted, is that no enzyme:inhibitor complexes of the HO-MBP fusion with papain were seen presumably because the samples were boiled in the presence of SDS and iodoacetate for 10 minutes before electrophoresis. Others (Nicklin & Barrett, 1984; Anastasi *et al.*, 1983; Barrett *et al.*, 1979), using a similar experimental approach have also failed to illustrate complex formation. However, should this be a theoretical requirement, complex formation could be shown using experiments such as non-denaturing electrophoresis at the appropriate pH (without boiling) or gel exclusion chromatography of labelled components (Heussen-Schemmer & Dowdle, 1993). Hence this study does not reflect the stability of inhibitory papain-fusion complexes during SDS-PAGE, but rather, the aim was to show that unlike the 'non-cystatin' fusion protein, the HO fusion is not degraded on interaction with papain, but acts as an inhibitor.

4.6.2 Gelatin Gel Electrophoresis

To study the behaviour of different papain-inhibitor complexes under mildly denaturing conditions of SDS-PAGE, mixtures of papain with either E-64 (an irreversible cysteine proteinase inhibitor that originates from *Aspergillus japonicus;* Hanada *et al.*, 1978), the HO, OC I, chicken cystatin, PMSF or SBTI were electrophoresed at 4°C in the presence of 1% SDS, in a polyacrylamide gel containing gelatin as co-polymerised substrate (Heussen & Dowdle, 1980; Michaud *et al.*, 1993, Michaud *et al.*, 1996).

After electrophoresis the SDS was removed by immersing the gel in Triton-X-100 and bands of lytic activity were developed by incubating the gel in assay buffer at 37°C. The bands of gelatin hydrolysis were revealed as clear areas against a dark background when the gel was stained with Coomassie brilliant blue R250.

The results (Figure 4.14) showed (predictably) that the serine proteinase inhibitors, PMSF and SBTI, had no discernable effect upon the migration or the activity of the enzyme (lanes 3 and 4). The E-64 (lane 2) inhibited papain completely, OC I (lane 5) and HO (lane 7) very nearly so and chicken cystatin (lane 6) to an appreciable but lesser extent.

Since residual enzyme activity with the M_r of papain represents free enzyme and since HO, chicken cystatin and OC I were all used in the same molar enzyme:inhibitor ratio, one may conclude that the chicken cystatin:papain complex is less stable in the presence of 1% SDS than are either HO:papain or OC I:papain complexes. This is of interest considering that the K_i of chicken cystatin is several orders of magnitude lower (in the 10^{-14} M range; Auerswald *et al.*, 1992) than that for OC I which has a K_i in the 10^{-9} M range (Arai *et al.*, 1991). One may also conclude from the results that the inhibitor:enzyme complexes are not catabolically active.

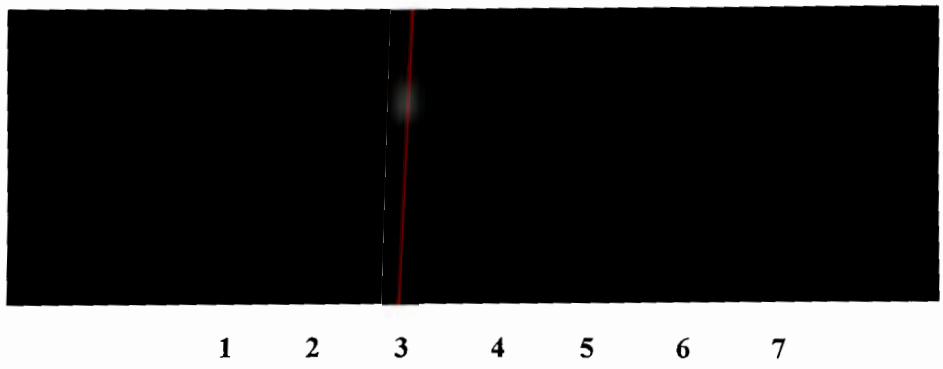


Figure 4.14: *13% SDS-gelatin-PAGE gel showing the stability of cystatin/papain complexes following electrophoresis. Lanes 1, papain only; 2, E-64/papain complex; 3, PMSF and papain; 4, SBTI and papain; 5, HO-fusion protein/papain complex; 6, Chicken cystatin (CC)/papain complex; 7, OC I/papain complex.*

4.6.3 Insect Trial Assays

Since the MBP-HO was an effective inhibitor of papain and is available in relatively large amounts, this compound was used for insect trial assays using the confused flour beetle, *Tribolium confusum*. This insect was chosen for two reasons:

1. It was readily available as a test system at the S.A.B.S. in Pretoria, South Africa.
2. Initial investigations of casein hydrolysis by the midgut proteinases of *T. confusum* have shown maximal activity between pH 6.5 and 6.9 (Birk *et al.*, 1962). This suggests that these insects use catheptic proteinases for digestion.

Crude periplasmic proteins containing 55% HO-MBP (determined by densitometric scaning of SDS-PAGE gels) were isolated as described in section 4.2.5. These and control MBP fractions were freeze-dried and sent to the S.A.B.S. in Pretoria.

The results of the insect assays are shown in Table 4.3. Problems were encountered in the administration of the fusion protein in the trial assays as the proteins were not easily crushed with the media. Repeated attempts to do this resulted in the proteins becoming sticky and difficult to mix in with the dry media. The proteins thus had to be dissolved in water and sprayed onto the nutrient medium (D. Van der Linde, personal communication).

Table 4.3: Insect trial assay results using the confused flour beetle, *Tribolium confusum.*

Sample	Number of Tribolium Confusum adults that have emerged				
			Replicates		
	1	2	3	4	5
Proteinase Inhibitor	>600	>700	>700	>600	>400
Control Protein (MBP)	>800	>700	>700	>500	>600
Untreated Control	>700	>600	>700	>400	>500

Results showed very little difference between the controls (the maltose binding protein and the untreated control) and the fusion. This could be attributed to a number of reasons:

1. There may have been other proteinases besides cysteine proteinases in the insect midgut which could have overcome any toxic effects of the inhibitor.
2. The concentrations of inhibitor used in this assay (0.1%) may not have been sufficient to deliver toxic effects. Should this be the case, the use of this inhibitor in transgenic plants would be questionable since it would be difficult to achieve higher concentrations in plant tissues.
3. In this study, an artificial diet was used. In many cases, the relative effectiveness of proteinase inhibitors has been shown to vary with dietary proteins of differing digestibility and/or quality (Kuroda *et al.*, 1996).
4. The procedure used in this experiment, for adding inhibitor to the dry medium may not have been effective for obtaining adequate mixing with the medium or dosage.

4.7 Conclusions

The HO fusion protein has been successfully cloned and expressed in the pMAL system. 42.5 mg of periplasmic proteins per litre were obtained from the periplasmic fraction following incubation of the induced pMAL-HO culture at 37°C for 24 hours. This was then purified on a cross-linked amylose column to give 7 mg of purified fusion protein.

Sufficient yields of this fusion protein were obtained and the inhibitor showed activity when assayed against papain in fluorescent assays and in the gelatin assay. It was also confirmed that this protein acts as an inhibitor rather than as a substrate of papain.

The system was, however, unsatisfactory in as much as low yields of inhibitor (0.025 mg) were obtained on cleavage with factor Xa. Attempts to optimise the conditions for factor Xa cleavage failed to improve the yield. Sufficient purified HO inhibitor was, however, obtained to enable N-terminal sequencing. This confirmed that factor Xa had cut at the correct site between the MBP and the HO proteins.

The usefulness of this inhibitor as an agent for biocontrol could not be fully assessed by the insect assay that was used. A better option would be to use the HO inhibitor protein (rather than the fusion) in biocontrol assays. Furthermore, because of the insect's ability to overcome the toxic effects of proteinase inhibitors (discussed in chapter 1, section 1.7), these assays should perhaps be:

1. expanded to include a range of concentrations of inhibitors
2. conducted in conjunction with other inhibitors such as the serine or aspartic proteinase inhibitors, with appropriate controls designed to provide more efficient ways of administering the HO inhibitor.

In conclusion therefore, the HO inhibitor had been successfully cloned and expresed in the pMAL system. However, the yields of 0.025 mg of fusion from one litre of culture were low. To try and improve this yield, another expression sysem was explored. This is described in Appendix A.

CHAPTER 5

Expression of Natural Oryzacystatin I in the pET System

5.1 Introduction

The previous chapters have described the synthesis, cloning and expression of the modified hybrid inhibitor. However, in order to study this hybrid inhibitor in relation to its two derivative proteins (i.e. the natural chicken and oryzacystatin inhibitors), it was necessary to obtain these two inhibitors. Chicken cystatin is commercially available (Sigma) but oryzacystatin on the other hand is not and I had either to express it in a genetic vector, or isolate it from rice seeds. I chose the former and was grateful to receive the natural OC I gene cloned into the pET-3a plasmid as a kind gift from Dr Lisa Jouanin (Department Of Cellular Biology, National Institute for Agricultural Research, Cedex, France).

The pET system is useful for the expression of foreign proteins. Target genes are cloned under the control of the strong bacteriophage T7 RNA polymerase promoter in the host cell. For protein production, a recombinant plasmid is transferred to an *E. coli* host strain, which contains a chromosomal copy of the gene coding for T7 RNA polymerase. These bacterial hosts are lysogens of bacteriophage DE3, a lambda derivative that has the immunity region of phage 21 and which carries a DNA fragment containing the *lacI* gene, the *lacUV5* promoter, and the gene for T7 RNA polymerase (Studier & Moffatt, 1986). Once a DE3 lysogen is formed, the only promoter that will direct transcription of the T7 RNA polymerase gene is the *lacUV5* promoter, which is inducible by IPTG. Addition of IPTG to the growing culture of the lysogen, induces the production of T7 RNA polymerase, which in turn transcribes the target DNA in the plasmid (Rosenburg *et al.*, 1987).

This section therefore, will describe the expression and purification of the OC I protein from the pET system. The expressed protein was then used in the kinetic studies and its activity compared with that of the hybrid oryzacystatin.

5.2 Materials and Methods

5.2.1 *Escherichia coli* Strains and Culture Conditions

pET-3a plasmid (Novagen) and pET-3aOC I (Leplé *et al.*, 1995) in *E. coli* strain BL21(DE3) (Novagen) were used for these experiments. The cultures were grown in M9ZB media (see Appendix B) at 37°C with vigorous shaking and plasmids were maintained by selection in the presence of 100 μg/ml ampicillin. For expression of the native inhibitor, cultures containing the parental (pET-3a) or the recombinant (pET-3aOC I) plasmid were

grown until the A_{600} reached 0.9 (approximately 5 hours). Expression was induced by the addition of 0.4 mM IPTG in the presence of 200 µg/ml rifampicin. The culture was then incubated for a further three hours at 37°C before pelleting the cells.

5.2.2 Isolation of Total, Soluble Cell Proteins

Isolation of the soluble, cytoplasmic cell proteins was carried out according to the method of Leplé *et al.* (1995). The cells were first pelleted at 5000 rpm for 20 minutes at 4°C and resuspended in 50 mM Tris, pH 7.5; 0.2 mM NaCl; 2 mM EDTA and 10% glucose. Cells were frozen overnight at -20°C. The following day, the cells were thawed and sonicated on ice (3 times for 40 seconds at 100 W) on a Soniprep 150 sonicator. This was repeated over 3 minutes and the suspension was centrifuged for 1 hour at 15 000 rpm. The soluble supernatant was heated at 65°C for 15 minutes and centrifuged as before to remove the heat-sensitive proteins. The supernatant was concentrated at 4°C by pressure ultrafiltration using a Diaflo YM3 membrane (Amicon Corp. Beverly, MA, U.S.A.).

5.2.3 Gel Exclusion Chromatography

For gel exclusion chromatography a Sephadex G25 gel filtration matrix was used and prepared according to the manufacturer's instructions (Pharmacia). Once the column was packed, it was then equilibrated in 10 mM Tris, pH 7.5 before placing 25 ml of the concentrated supernatant over the column. Five millilitre fractions were then collected using a fraction collector, and the A_{280} of each 5 ml fraction was read on a Beckman DU-64 spectrophotometer. Samples were analysed by SDS-PAGE and Western blotting (Appendix C). Fractions containing inhibitory activity (assayed as described in Appendix C) were pooled and freeze-dried.

5.2.4 Ion Exchange Chromatography

Ion exchange chromatography was carried out using DEAE Trisacryl M (Sigma) according to the method of Leplé *et al.* (1995). Forty millilitres of the column matrix was packed into a 20 cm × 2.5 cm column according to the manufacturer's instructions (Sigma). This was then equilibrated in 8 column volumes of 25 mM Tris-HCl, pH 7.5. The freeze-dried OC I extract was then dissolved in 30 ml distilled water and placed over the column. This was followed with 30 ml of distilled water and then a discontinuous salt gradient from 0.15 to 0.6 M NaCl was used to elute the proteins (Leplé *et al.*, 1995). Following ion exchange, the salt was removed by dialysis at 4°C against 5 mM potassium phosphate buffer, pH 6.5.

5.2.5 Affinity Purification

Cm-papain-Sepharose was prepared according to the method of (Anastasi *et al.*, 1983) such that a final concentration of 2.3 mg of protein per ml of gel was attained. This was then packed into a column, 2.5 cm × 10 cm. The column was equilibrated in 50 mM phosphate buffer, pH 6.5, containing 0.5 M NaCl and 0.1% Brij 35 (Anastasi *et al.*, 1983). After absorption of the proteins, the column was washed in this same buffer, but with 10% glycerol in place of the Brij 35. Non-specifically bound material was eluted with 50 mM K_3PO_4, pH 11.5 containing 0.5 M NaCl and 10% glycerol. The collected fractions were then adjusted to pH 7.4 by dialysis at 4°C against 5 mM potassium phosphate buffer, pH. 7.4.

5.2.6 Polyacrylamide Gel Electrophoresis and Western Blotting

Samples were prepared for electrophoresis on a 20% SDS-PAGE gel as follows:
1 ml volumes were removed from *E. coli* cultures and pelleted by centrifugation in a microcentrifuge at 14 000 rpm for 5 minutes. Cell pellets were then resuspended in 50 µl sample loading buffer. Immediately before loading onto SDS-PAGE gels, samples were boiled for 5 minutes to disrupt the cells. 20 µl of this suspension (approximately 30 µg of protein) was loaded. 20% SDS-PAGE gels were prepared as described in Appendix C. For Western blotting, proteins were electrophoresed onto nitrocellulose and analysed according to Towbin *et al.* (1979) with modifications (Appendix C). OC I antibodies were kindly supplied by Professor Soichi Arai (Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, Japan).

5.3 Results and Discussion

5.3.1 Analysis of the Expressed Oryzacystatin I Protein

Expression was induced by the addition of IPTG (and rifampicin; 200 µg/ml) to a late log phase culture. The antibiotic inhibits transcription by the host RNA polymerase and thus reduces the background synthesis of host RNA and proteins (Novagen manual).

One millilitre aliquots were removed from the cell cultures at various times following induction. The expression results of the pET-3aOC I are shown in Figure 5.1. The uninduced culture (lane 2), shows almost no expression of OC I, whilst lanes 3 - 5, show an accumulation of a 12-kDa protein. 12 kDa is slightly larger than the published size of 11,800 kDa for natural OC I as shown by Abe *et al.* (1987a). However, there is often a slight over-estimation of the molecular weight of cystatins when compared with commonly used molecular weight marker proteins (Abrahamson *et al.*, 1986). A band size of 12 kDa for OC I has been observed on SDS-PAGE gels by Abe *et al.* (1987a).

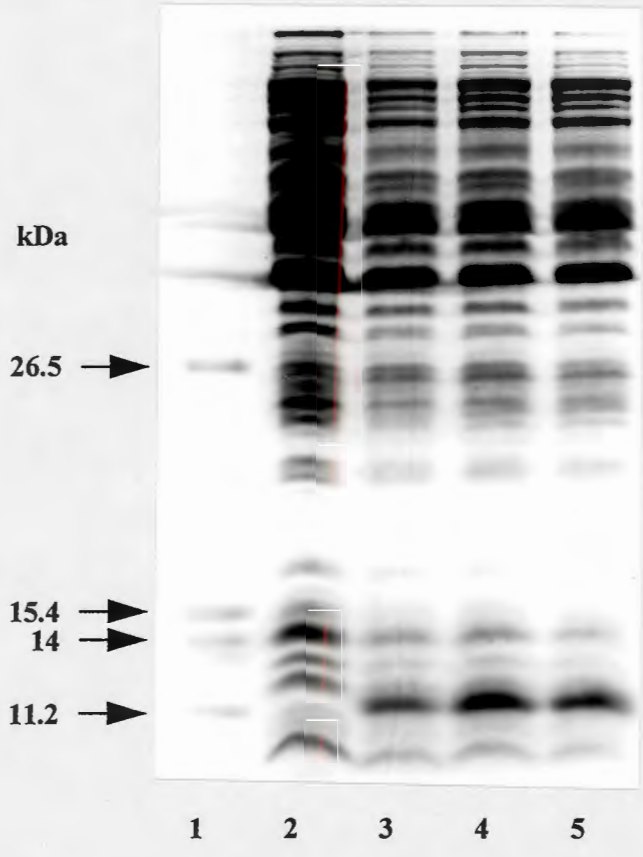


Figure 5.1: 20% SDS polyacrylamide gel showing the results of expression of pET-3aOC I. Lane 1, molecular weight markers (total chicken histones); lanes 2 - 5, pET-3aOC I at 0, 1, 2, and 3 hours following induction with IPTG.

Figure 5.2: Western blot of a 20% SDS' polyacrylamide gel. Proteins were detected using antibodies to natural OC I. Lane 1, pET-3a uninduced; lanes 2-4, pET-3a 1,2 and 3 hours post-induction; lane 5, pET-3aOC I uninduced; lanes 6–8, pET-3aOC I induced.

The identity of this 12-kDa protein was further confirmed by western blot analysis using an antibody to natural OC I (Figure 5.2). The antibodies bound to a protein of approximately 12 kDa at one, two and three hours post-induction (lanes 6 - 8). No bands were seen for the uninduced pET-3aOC I (lane 5), nor for the pET3a plasmid only in the uninduced (lane 1) or in the induced cultures (lanes 2 - 4).

5.3.3 Purification of the Expressed OC I Protein by Gel Exclusion Chromatography and Ion Exchange.

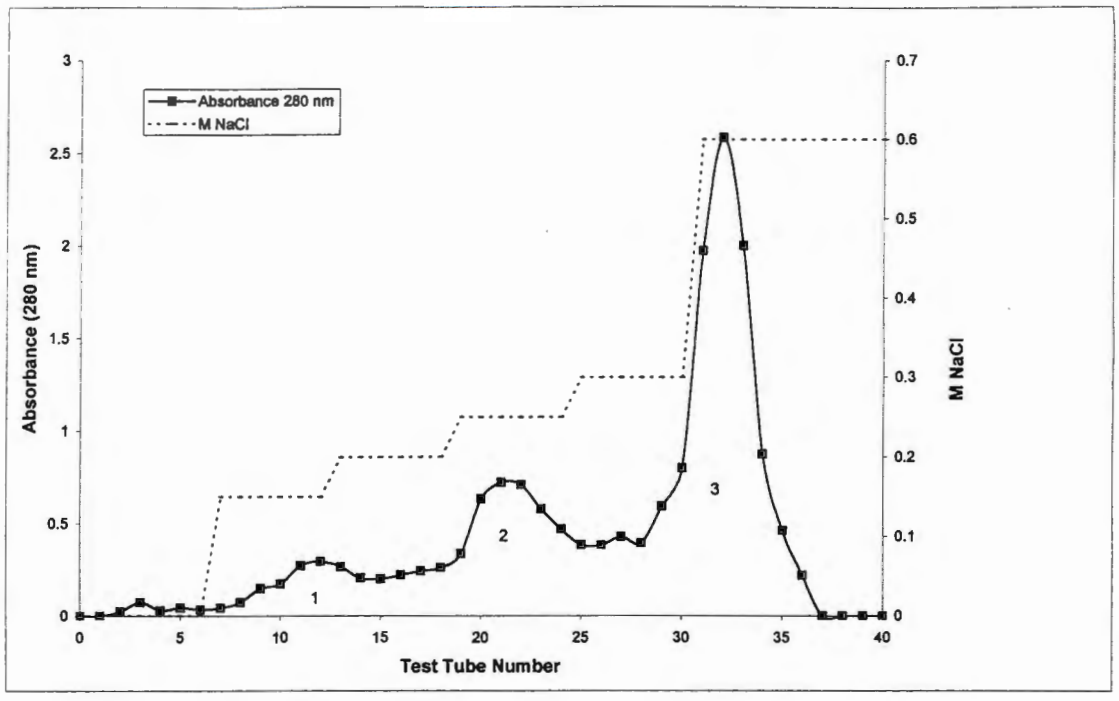
Once expression had been optimised, 1 litre cultures of pET-3aOC I were induced, the cells harvested and the soluble cytoplasmic proteins were then isolated. Cystatins have been shown to be heat stable (Kondo *et al.*, 1990). Thus, as a purification step, cytoplasmic proteins were heated for 15 minutes at 65°C. The heat labile proteins were then removed by centrifugation and the heat stable fraction isolated by gel exclusion chromatography on a Sephadex G25 matrix. This aimed to remove the high concentrations of salts, glucose and rifampicin present in the medium.

To determine which fractions contained the inhibitor, each 5 ml fraction collected was tested for inhibitory activity against papain using the fluorimetric substrate Z-Phe-Arg-AMC (Sigma). Inhibition was determined as the percentage decrease in enzyme activity following the addition of 0.1 ml aliquots from the 5 ml fractions. Calculations were carried out as described in Appendix C.

Anastasi *et al.* (1983) and Keilová and Tomášek (1974) have found that freeze drying of chicken cystatin leads to a loss of inhibitory activity. Since such findings were not reported by Leplé *et al.*, (1995), and since we observed no significant loss in inhibitory activity following freeze-drying, the column fractions were pooled and freeze-dried before placing over the ion exchange column as described by Leplé *et al.* (1995).

For ion-exchange, proteins were eluted using a discontinuous salt gradient according to the method of Leplé *et al.* (1995). Peaks were eluted in 5 ml fractions and the A_{280} of these fractions were read on a Beckman DU-64 spectrophotometer. The column profile is shown in Figure 5.3 (A). A number of peaks eluted with the different molar salt concentrations and aliquots were run on 20% SDS-PAGE gels to check for the presence of the inhibitor in the various peaks (Figure 5.3(B)). From the SDS-PAGE gel, the OC I protein was found to elute between 0.15 and 0.2 M NaCl. Lane 1 shows the presence of a 12-kDa protein which had previously been shown (by Western Blot analysis, section 5.3.1), to be the natural OC I inhibitor. From lanes 2 and 3, it is clear that a number of contaminating proteins have been removed. In lane 2, however, a number of other protein bands could be seen which eluted at the same time as the OC I protein. Further purification methods were therefore necessary to remove these contaminating proteins, although this protein had been used, following ion exchange, by Leplé *et al.* (1995) for their insect assays.

(A)



(B)

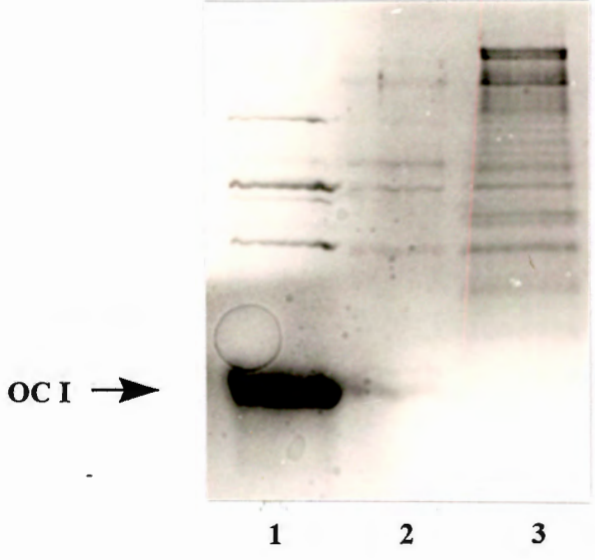


Figure 5.3: *(A) Column profile of proteins eluted from the ion exchange column (DEAE Trisacryl M) equilibrated in 25 mM Tris-Cl, pH 7.5. Proteins were eluted with a discontinuous salt gradient from 0.15 to 0.6 M NaCl (indicated by the dotted line).* ***(B)*** *20% SDS-PAGE gel of fractions eluted from the ion-exchange column. Lanes 1 – 3 are peaks 1, 2 and 3 respectively.*

To further purify the expressed protein use was made of an affinity column. Fractions containing the OC I protein from the ion-exchange column (as determined by SDS-PAGE analysis) were dialysed against 5 mM phosphate buffer (pH 6.5) to remove the NaCl. This was then passed through a column of Cm-papain linked to Sepharose 4B. The column was washed with phosphate buffer, pH 6.5 and the bound proteins were eluted by changing the buffer pH to 11.5 (Anastasi *et al.*, 1983). The elution profile of the affinity column is shown in Figure 5.4. Only a single protein eluted with the high pH buffer and this was shown to inhibit papain (see Chapter 6).

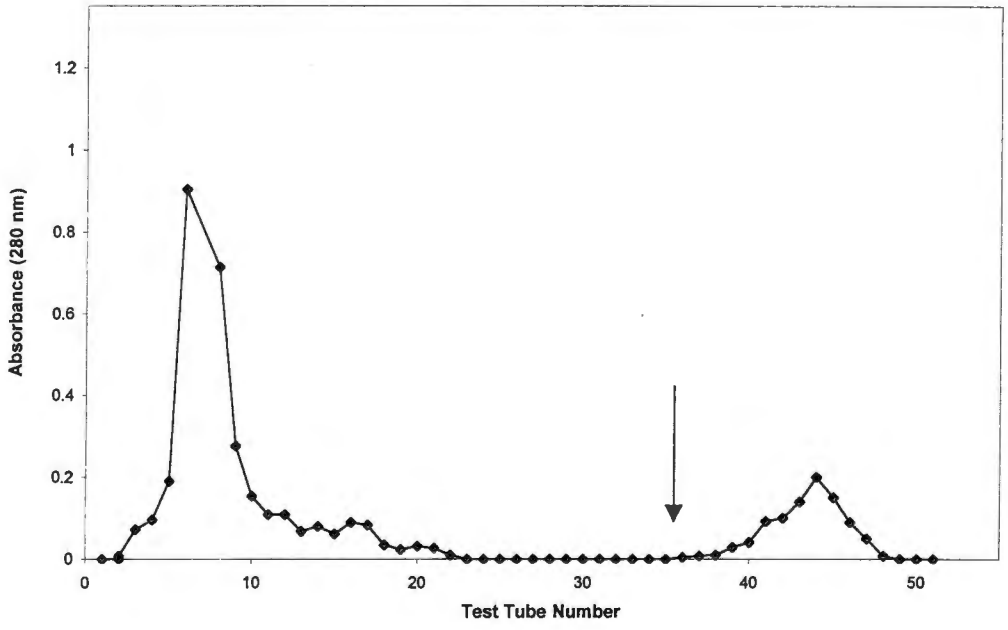


Figure 5.4: *Affinity chromatography profile of proteins eluted from the affinity column of Cm-papain-Sepharose. 10 ml of proteins were applied to the column equilibrated in buffer (50 mM phosphate buffer, pH 6.0; 0.5 M NaCl; 0.1% Brij 35). The column was washed with (50 mM phosphate buffer, pH 6.0; 0.5 M NaCl;10% glycerol) followed by elution of the bound protein with 50 mM K_3PO_4, pH 11.0; 0.5 M NaCl; 10% glycerol (indicated by the arrow).*

560 µg of pure OC I protein was obtained from the one litre pET-3aOCI culture following purification (Table 5.1).

Table 5.1: Purification of the OC I protein from *E. coli*

Stage	Volume (ml)	OC I Protein (mg)[a]	Activity ([b]Units)	Specific Activity (U/mg)	Purification	Yield %
Total Cytoplasmic Proteins	40	98.6	62695	636	1	100
Sephadex Column	55	22	47826.1	2174	3.4	76.3
Ion Exchange Chromatography	15	11.4	34168.8	2997.3	4.7	54.5
Affinity Chromatography	10	0.56	22541	40252	63.3	35.9

[a] Measured as total protein by the Biorad (USA) Protein Assay (Bradford, 1976; Appendix C)

[b] 1 Unit is arbitrarily defined as the amount of inhibitor required to reduce the activity of 0.5 μM papain by 50% under the assay conditions.

In conclusion, the pET-3aOC I system supplied by Leplé *et al.* (1995) was used successfully to express and purify sufficient natural OC I for the kinetic studies. This purified protein was then used in enzyme kinetic assays. This will be discussed in the following chapter.

CHAPTER 6

Enzyme Kinetic Studies

6.1 Introduction

In order to define the function of the hybrid inhibitor and to relate this to the activity of the parent inhibitors from which it was derived, the appropriate kinetic parameters were measured. The experimental approaches that were used for this purpose were taken largely from the literature and were based upon the generally accepted assumptions: (a) that cystatin activity is reliably measured by the inhibition of papain-catalysed hydrolysis of the synthetic, fluorogenic substrate benzylcarbonylphenylalanyl-arginyl-aminomethylcoumarin (Z-Phe-Arg-AMC; Barrett & Kirschke, 1981) (b) that the hybrid inhibitor was a slow, tight-binding inhibitor whose interactions with papain could be defined kinetically by the standard equations for this class of inhibitors (Henderson, 1972; Cha, 1978; Morrison, 1982); and, (c) that reliable estimates of the kinetic parameters can be obtained by regression analysis of progress curves that express the concentration of accumulated product (in this case AMC) as a function of time after the addition of pre-incubated mixtures of papain and inhibitor to substrate in the reaction cuvette.

6.2 Materials and Methods

6.2.1 Enzyme Assay

Purified papain (EC 3.4.22.2) was obtained from Boehringer Mannheim and used without further purification. Stock solutions of papain were made up in 0.1 M phosphate buffer and stored frozen at –70°C in 200 µl volumes. A tube was thawed, stored on ice for the duration of the experiment and then discarded.

Papain activity was measured fluorimetrically by recording the accumulation of AMC released by the hydrolysis of Z-Phe-Arg-AMC as a function of time. A Perkin Elmer MPF-43A fluorescence spectrophotometer, fitted with a temperature-controlled cell holder and a chart recorder was used for this purpose. The instrument was standardised with a solution of 0.2 µM 7-amino-4-methylcoumarin (AMC; Sigma) containing the substrate, Z-Phe-Arg-AMC (Sigma) at the concentrations described for the different assays. Concentrations were chosen so that the maximum amount of substrate hydrolysis did not exceed 2%. Fluorescence was measured at an excitation wavelength of 380 nm and an emission wavelength of 460 nm. Chart recordings of fluorescence as a function of time were read manually for linear data or by digitising for non-linear data.

Assays were carried out in phosphate buffer, pH 6.8. According to Mason *et al.* (1985), assays for inhibitory

activity should be carried out above pH 6.0, preferably at pH 6.8, to prevent the dissociation of inhibitory complexes into active enzyme and inhibitor. Assays for inhibitory activity thus involve a compromise between the optimal for proteolytic activity of the enzyme and stability of the cystatin:enzyme complex.

6.2.2 Active-Site Titration of Papain with E-64

Active-site titrations of papain were carried out using the method of Barrett & Kirschke (1981). Three milligrams of E-64 (Sigma) were dissolved in 0.1 ml DMSO and diluted to 1 mM with distilled water. Working inhibitor solutions of 1, 2, 3, 4, 5, 6, 7 and 8 μM E-64 were then made. A solution of papain (approximately 10 μM) in 0.1% Brij 35 was used to prepare mixtures comprising: enzyme solution - 25 μl; papain assay buffer - 50 μl and E-64 solutions - 25 μl. These were incubated at 30°C for 30 minutes and diluted to 5 ml by the addition of 4.9 ml of 0.1% Brij 35. Ten microlitres of this solution was then assayed for residual enzyme activity. The assay buffer was kept as a 4 × concentrated stock solution and diluted, for use, to the 1× buffer containing 100 mM KH_2/Na_2 phosphate pH 6.0; 1 mM EDTA; 2 mM DTT. The DTT was added freshly on the day of use.

The result (Figure 6.1) showed that the standardised papain solution had an active concentration of 8.66 μM.

6.2.3 Determination of the K_m for the Hydrolysis of Z-Phe-Arg-AMC Catalysed by Papain

Papain (50 pM) was incubated with the phosphate-assay buffer and 0.1% Brij 35 at 30°C for 2 minutes. The reaction was then started by the addition of various concentrations of Z-Phe-Arg-AMC ranging from 1.5 to 100 μM and the fluorescence released was measured continuously. Values for K_m and V_{max} were derived from non-linear fitting of the recorded curve to the Michaelis-Menten equation using the FigP program (Elsevier-Biosoft, Cambridge, UK).

6.2.4 Active-site Titration of the Inhibitors with Titrated Papain

Measurement of the inhibitory constant (K_i) for the papain:inhibitor complexes required estimates of the reactants. The molarity of the papain solution was determined, as described above, by titration with E-64. The standardised papain solution was then used to measure the concentration of active inhibitor sites in stock solutions as follows:

Papain (0.3 μM final active concentration; 42 pmoles) was incubated with 70 μl papain assay buffer and varying dilutions of inhibitor solutions (see Appendix E) for 30 minutes at 30°C in a total volume of 140 μl.

This was then diluted to 500 µl and 10 µl were used for the assay. Titrations were performed at a papain/K_i ratio of >100. Reactions were started by the addition of Z-Phe-Arg-AMC to a final concentration of 5 µM. The protein content of the inhibitor solution was then measured using the Bio-Rad protein assay and specific activity (µmoles inhibitory activity/mg protein) was calculated assuming 1:1 stochiometry.

6.2.6 K_i Determinations for the Various Cystatins

For the determination of equilibrium constants, continuous rate assays were carried out using 20 µM Z-Phe-Arg-AMC. Variable concentrations of inhibitors (80 nM – 240 nM) were used and reactions were started by the addition of 20 µl papain (activated in assay buffer at 30°C for 2 minutes) at 8 nM final active concentration. All experiments were performed under pseudo first order conditions i.e. at a molar ratio of inhibitor to enzyme of >10:1.

Apparent K_i values ($K_{i(app)}$) were calculated as the slope of the plot of $[I]/1-V_i/V_o$) versus V_o/V_i (Henderson, 1972). K_m values of 63 µM (determined as described in section 6.2.3) were used to calculate the substrate independent K_i using the relationship:

$$K_{i(app)} = K_i (1 + [S]/K_m)$$

All determinations of V_o and V_i were based on assays with less than 2% hydrolysis and a linear correlation coefficient, at steady state, of greater than 0.990.

For chicken cystatin, the K_i value was obtained from estimates of K_{on} and K_{off}. Papain (0.1 nM) was incubated in papain assay buffer at 30°C for 2 minutes in a total volume of 30 µl. The reaction was started by addition of the activated papain to the reaction cuvette which contained 10 µM Z-Phe-Arg-AMC and inhibitor (0.5 nM – 2 nM). Values for K_{obs}, V_o and V_i were obtained by curvilinear fitting to the equation:

$$[P] = \upsilon_i t + (\upsilon_0 - \upsilon_i)(1 - e^{-k_{obs}t}) / k_{obs} \quad \text{.....(Cha, 1976)}$$

where

[P] = the concentration of product formed by hydrolysis of the substrate

t = time

V_o = initial rate of substrate cleavage

V_i = final rate of substrate cleavage

k_{obs} = the observed pseudo first order rate constant for the reaction of the inhibitor with the enzyme

using FigP (Elsevier-Biosoft, Cambridge, UK).

6.3 Results and Discussion

6.3.1 Active-site Titrations of Papain and Inhibitors

Before commencing the fluorimetric assays, the instrument was calibrated to give a maximum reading with 0.2 μM AMC. This ensured that no more than 2% of the substrate present was hydrolysed over the assay period and that any decrease in the velocity measured was due to the addition of inhibitors and not to substrate depletion (Salvesen & Nagase, 1989).

Linear regression analysis of the data established a conentration of 8.66 μM for the standardised papain solution. Figure 6.1 shows the plot of residual enzyme activity against E-64 concentration (for data, see Appendix E).

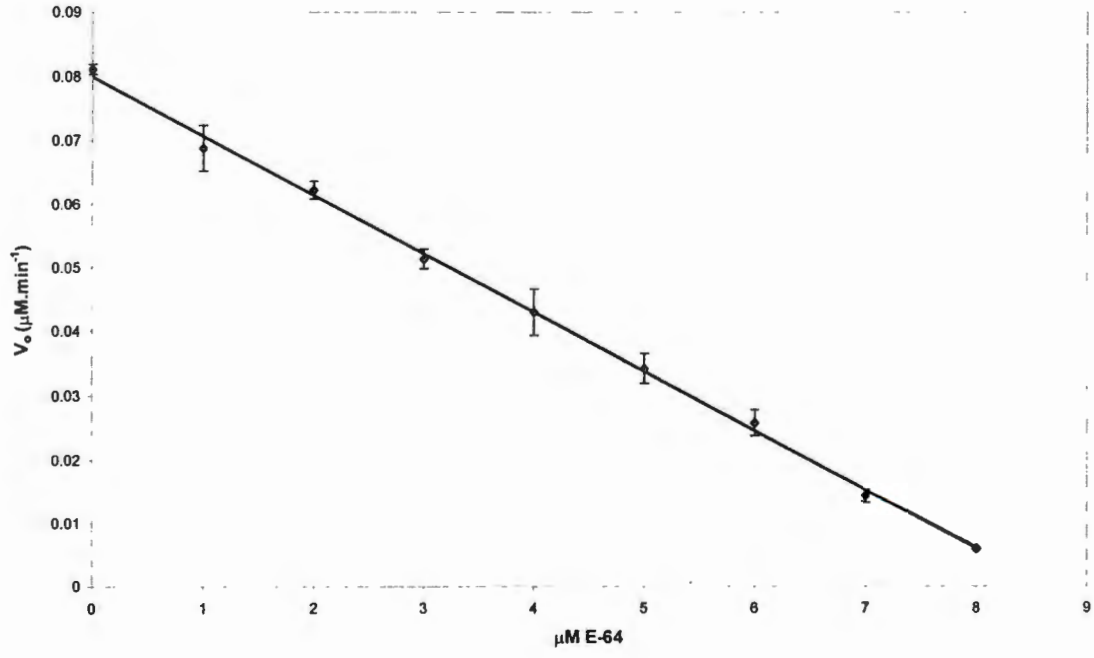


Figure 6.1: Active-site titration of papain with E-64. Papain was incubated with E-64 for 30 minutes at 30 ℃ and residual enzyme activity was detected by continuous fluorimetric assays using 5 μM Z-Phe-Arg-AMC. Error bars show the 95% confidence limits.

Inhibitor solutions were then titrated against the standardised papain to determine their active concentrations. For the active site titrations of the inhibitors, a papain:K_i ratio of >100 was used since, using an enzyme concentration of the same order of magnitude as the K_i, gives curvilinear titration curves that do not extrapolate to a reliable intersection on the abscissa. Table 6.1 shows the specific activity determined for the various inhibitors (for data, see Appendix E).

Table 6.1: Activity of papain and inhibitors as determined by active-site titration

Inhibitor	Specific Activity	
	[a]μmoles/mg protein	[b]moles/mole
Chicken Cystatin	0.0692	0.91
HO Fusion	0.0166	0.86
HO	0.0153	0.71
OC I	0.0402	0.45

[a]μmoles of active sites of inhibitor per mg of protein determined using the Bio-Rad protein assay.

[b]moles of active sites per mole which assumes purity of the protein added and exact mass determination.

6.3.2 K_m Determination

To determine the K_m for Z-Phe-Arg-AMC, I incubated papain (50 pM) with various concentrations of substrate (2.5 – 100 μM) and measured the reaction velocitiy (V_o) for each substrate concentration ([S]). Values for K_m and V_{max} were obtained by using the FigP program (Elsevier-Biosoft, Cambridge, UK) to fit the observed values for V_o to the Michaelis-Menten equation:

$$\upsilon_0 = \frac{V_{max}[S]}{K_m + [S]}$$

The results are shown in Figure 6.2 (for data, see Appendix E).

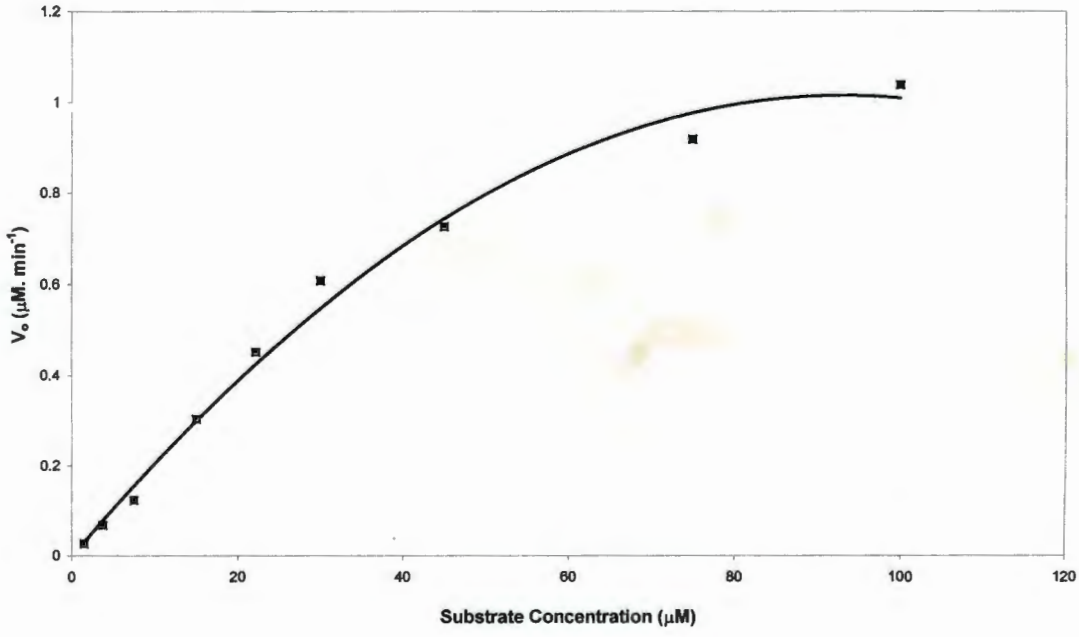


Figure 6.2: *Michaelis-Menten Plot for papain.*

A K_m value of 63 μM was obtained for the substrate under the assay conditions described in section 6.2.3 (Figure 6.2). This is similar to the K_m values of 60 μM obtained by Lindahl *et al.* (1992) and 65 μM by Zucker *et al.* (1985).

6.3.3 Determination of Equilibrium Constants (K_i)

A number of different methods have been used to determine the equilibrium dissociation constant (K_i) of cystatins and their cognate enzymes. These include the determination of the initial reaction velocities (V_i) under presteady state conditions in which activated enzyme is added to a mixture of inhibitor and substrate. The results are then fitted (by non-linear regression) to the Cha equation (Cha, 1975). Other methods that have been used are those described by Henderson (1972) in which the reaction is started by the addition of enzyme to the substrate. The relative steady state velocities before (V_o) and after (V_i) the addition of the cystatin inhibitor are then used to calculate the apparent K_i value ($K_{i(app)}$).

Protocols, such as those used by Bjork *et al.*, 1989; Turk *et al.*, (1996); Lindahl *et al.* (1994), that depend upon the analysis of pre- and post-steady state reaction velocities are best used when computer-controlled experimental systems are available for the automatic addition of reagents and for monitoring the subsequent accumulation of fluorescent product.

Since I did not have access to the necessary equipment and had, instead, to rely on manual timing of the experiment and a strip chart recorder for output, I chose to use the approaches described by Kondo *et al.* (1990), in which initial reaction velocities are measured in the absence of inhibitor (V_o) and in the presence of different concentrations of inhibitor (V_i) after co-incubation with enzyme for a long enough period for equilibrium to have been established.

Each assay took approximately 60 minutes to complete since this period of time was needed before linear recordings of V_i were obtained. Control incubations, without inhibitor, for 0, 20 and 60 minutes gave similar V_o values for the enzyme.

The data obtained for V_i and V_o for the different inhibitors are shown in Appendix E. The points in these final graphs represent the means of several experimental observations. To ensure that residual enzyme activity could be detected at the high concentrations of inhibitor used in these assays the substrate was used at concentrations of 25 μM. A lower amount of enzyme (0.8 nM) was used for the HO as there was less inhibitor available for assay. Apparent K_i values ($K_{i(app)}$) were calculated as the slope of the plot of $[I]/1-V_i/V_o$ versus V_o/V_i (Henderson, 1972) and the K_i was determined from the relationship:

$$K_{i(app)} = K_i (1 + [S]/K_m).$$

using a K_m value of 63 μM. The corrected K_i values obtained are shown in Table 6.2.

Table 6.2: K_i data for the different inhibitors

Inhibitor	K_i (M)
Chicken Cystatin	5.95×10^{-12}
Oryzacystatin I	1.58×10^{-9}
HO-MBP Fusion	2.89×10^{-9}
HO	3.6×10^{-9}

Analysis of the data (Appendix E) showed considerable variation between the data sets, (i.e. the standard deviation at a given inhibitor concentration was greater than the differences between the average points at different inhibitor concentrations) so that very large differences in inhibitor concentration were required to document significant changes in V_i.

Estimates of K_i obtained for the different oryzacystatin variants (Table 6.2) were slightly higher than described in the literature. Represented values for K_i of oryzacystatin range from 7.0 nM to 36.4 nM (Abe *et al.*, 1988; Kondo *et al.*, 1990; Arai *et al.*, 1991; Michaud *et al.*, 1994; Urwin *et al.*, 1995). K_i values obtained for HO were similar to those observed with the natural OC I. If anything, the natural OC I was a marginally better inhibitor. No significant difference was observed between the fused and the free hybrid inhibitors.

Because chicken cystatin is such a tight-binding inhibitor that does not readily dissociate from the complex, the K_i for this inhibitor was obtained from the k_{off} and k_{on} data determined in pre-steady state experiments where the enzyme was added to a mixture of the inhibitor and substrate (described in Turk *et al.*, 1996). The data were fitted by non-linear least squares regression to the equation:

$$[P] = \upsilon_i t + (\upsilon_0 - \upsilon_i)(1 - e^{-k_{obs}t}) / k_{obs}$$

(Cha, 1975), where [P] = the concentration of product formed by hydrolysis of the substrate

t = time

V_o = initial rate of substrate cleavage

V_i = final rate of substrate cleavage

k_{obs} = the observed pseudo first order rate constant for the reaction of the inhibitor with the enzyme

Detailed evaluation and graphic presentation of results (Appendix E) was performed using FigP (Biosoft Cambridge).

k_{obs} was found to be linearly dependent on inhibitor concentration (Appendix E). These observations indicate that, under the conditions used, the interactions followed the mechanism E + I = EI; i.e., the combination of enzyme and inhibitor occurred as a simple, reversible, bimolecular reaction (Morrison,1982). According to

this mechanism, the linear dependence of K_{obs} on inhibitor concentration is described by:

$$k_{obs} = k_{-1} + k_{+1}[I]$$

where k_{-1} and k_{+1} represent the 'off' and the 'on' rate constants, respectively.

Using the equation it should, theoretically be possible, by plotting k_{obs} as a function of [I], to obtain values for both the 'on' and 'off' rate constants – the former as the slope of the curve and the latter as the intercept on the coordinate axis. The value that I obtained (9.1 $M^{-1}s^{-2}$) is similar to that described by Bjork *et al.* (1990).

The 'off' rate , however, is so slow that I was not able to get a reliable estimate of the k_{off} and the derived rate for K_i (10^{-12} M) should be regarded as no better than the upper estimate. Bjork *et al.* (1990) using a superior experimentan system, recorded a K_i of 10^{-14} M.

6.4. Conclusions

The results that I obtained show that the substitution of the N-terminal 21 residues of oryzacystatin for the N-terminal 15 residues of chicken cystatin did not generate a significantly better inhibitor. In fact, a comparison of the inhibitory constants of the hybrid oryzacystatin with the natural component inhibitors showed the HO to have a K_i that is similar to the natural OC I and significantly higher than that of the truncated chicken cystatin inhibitor. Similar K_i results were seen for the HO and that fused to the MBP. This is in agreement with results obtained by Michaud *et al.* (1994), Thiele *et al.* (1990) and Kaji *et al.* (1990).

In terms of the molecular models, these kinetic results suggest that model 1 for the HO inhibitor (Chapter 2) is an unlikely structure. Urwin *et al.* (1995b) showed that Gly^{10} of OC I was important for maintaining its inhibitory activity and that the loss of this residue resulted in an increased K_i. They suggested that the importance of this residue lies in the fact that it is involved in a turn formation which brings the Leu^9 residue of OC I into contact with the proteinase for tighter binding in this region. In model I, however, there is no contact between the amino terminus of the HO and papain and hence, from the results described by Urwin *et al.* (1995b), one would expect such an inhibitor to have a greatly increased K_i. My results do not support this.

Since model 2 was found to be improbable, the most likely structure for the HO is model 3. In this model the amino-terminal contacts of the HO, containing the substituted chicken cystatin N-terminus, makes similar contacts with the N-terminal region of papain to the natural chicken cystatin. The importance of the N-terminal region of chicken cystatin has been demonstrated (Machleidt *et al.*, 1989; Abrahamson *et al.*, 1992; Lindahl *et al.*, 1992; Bjork *et al.*, 1994) and loss of the amino-terminal residues Leu^7 to Gly^9 have resulted in large increases in the inhibitory constants of chicken cystatin. The question that therefore arises is if the N-terminus does play such an important role, why was there such a small change in the inhibitory constant of the HO compared to the natural OC I?

From the results and in answer to this question, the following points can be made:

1. Urwin *et al.* (1995) showed that deletion of the N-terminal 21 residues from oryzacystatin resulted in the inhibitor becoming completely inactive. Abe *et al.*, (1988) who used a similar strategy, (although their inhibitor had an additional 15 amino acids at the N-terminus derived from the pUC18 vector) found no change in the inhibitory capabilities of the molecule. These results, as well as those shown here, suggest that a pre-requisite for inhibitory activity of the inhibitor is the presence, in this area, of residues that provide some contact with the proteinase. In this way, the N-terminus may help to pilot the molecule over the active site of the enzyme and thus, as long as there are no residues in the amino terminus that provide steric hindrance with papain, a degree of inhibition is obtained. Although not determined in this project, there may in fact be differences in the association and dissociation constants between the natural and hybrid rice inhibitors.

2. Shibuya *et al.* (1995) have shown that changes in the N-terminus can affect the positioning of the first hairpin loop which has been shown to be important in the cystatins for conferring inhibitory activity (Abrahamson *et al.*, 1987; Abe *et al.*, 1988; Turk & Bode, 1991). Furthermore, Bode *et al.* (1988) showed that interactions occur not only between the cystatin and the proteinase but also between the different cystatin loop regions (for example, Gln^{53} of chicken cystatin forms a H-bond with Gly^{9} of the N-terminus). Although I was not able to show (through molecular modelling) any changes in the interactions of the different loops of OC I by replacing the N-terminus with that of chicken cystatin, it must be remembered that the OC I part of this modelled molecule was based on the structure of stefin B and there may be slight differences to the actual OC I inhibitor. Thus the interaction of the substituted region and the two binding loops of the HO may have been different from that of the natural chicken cystatin and this could have been responsible for the increased K_i of the HO. This is particularly important since chicken cystatin has a different sequence (i.e. QLVSG as opposed to QVVAG). Shibuya *et al.* (1995) showed that the contacts between the cystatin binding loops are essential and any changes may affect the association constant and hence the K_i. I agree with Urwin *et al.* (1995), Machleidt *et al* (1995) and Shibuya *et al.* (1995) who suggest that the inhibitory regions of cystatins do not act independently but all contribute to the inhibitory capabilities of the inhibitor.

3. Many of the effects of mutations on the K_i's have been the result of significant changes in the properties of the substituted amino acid residues. For example, in experiments with stefin B, changes in the K_i were observed when ionic residues such as glutamine replaced Cys^{3}. In my case, however, substituted amino acid residues were all very similar in the form of their hydrophobics and charges, and thus I would not have expected a large change in K_i on this account.

4. It has been suggested that binding of the N-terminus in family II cystatins compensates for less favourable contacts made in the other hairpin loops such as the QLVSG region of chicken cystatin (Machleidt *et al.*, 1989). The results I obtained with the HO indicate that the N-terminus in the family I cystatins may not be as important as in the family II cystatins, since substitution of the whole N-terminal region of OC I for the chicken N-terminus did not significantly change the inhibitory constant from that of the natural inhibitor.

5. Finally, experiments in which the N-terminus of stefin B was replaced by the N-termini of cystatin C or kininogen (Jerala, 1994), yielded protein whose activity was largely unaffected as far as papain and cathepsin L were concerned. Inhibition of cathepsin H by these hybrids, however, was considerably weaker than that of the natural stefin. Since only the inhibitory effects against papain were investigated in this project, it would be interesting to determine the effect of such a substitution on other cysteine proteinases such as the cathepsins.

Thus in summary, the interactions between the three binding loops appear to be important for inhibitory activity of the cystatins, particularly in the family II cystatins. The flexible N-terminus being responsible for the positioning of the cystatin over the active site residues of the cysteine proteinase.

In terms of biocontrol, this inhibitor was shown to be slightly less effective an inhibitor of papain than the natural inhibitor. To improve its inhibitory capabilities for use in plants, therefore, one would need to study other substitutions such as those in which other loop regions of chicken cystatin were introduced as mentioned by Urwin *et al.* (1995b).

REFERENCES

Abe, K., Arai, S., Kato, H., and Fumaki, M. (1980) Thiol-protease inhibitors occurring in endosperm of corn. Agric. Biol. Chem., *44*, 685 - 686.

Abe, K., and Arai, S. (1985) Purification of a cysteine proteinase inhibitor from rice, *Oryza sativa L. japonica.* Agric. Biol. Chem. *49*, 3349 – 3350.

Abe, K., Kondo, H., and Arai, S. (1987) Purification and characterisation of a rice cysteine proteinase inhibitor. Agric. Biol. Chem., *51*, 2763 - 2768.

Abe, K., Emori, Y., Kondo, H., Arai, S., and Suzuki, K. (1988) The NH_2-terminal 21 amino acid residues are not essential for the papain-inhibitory activity of oryzacystatin, a member of the cystatin superfamily. J. Biol. Chem., *263*, 7655-7659.

Abe, K., Kondo, H., Watanabe, H., Emori, Y., and Arai, S. (1991) Oryzacystatins as the first well-defined cystatins of plant origin and their target proteinases in rice seeds. Biomed. Biochem. Acta, *50*, 637 - 641.

Abrahamson, M., Barrett, A. J., Salvesen, G., and Grubb, A. (1986) Isolation of six cysteine proteinase inhibitors from human urine. J. Biol. Chem., *261*, 11282 - 11289.

Abrahamson, M., Ritonja, A., Brown, M. A., Grubb, A., Machleidt, W., and Barrett, A. J. (1987) Identification of the probable inhibitory reactive sites of the cysteine proteinase inhibitors human cystatin C and chicken cystatin. J. Biol. Chem., *262*, 9688 - 9694.

Abrahamson, M., Mason, R. W., Hansson, H., Buttle, D. J., Grubb, A., and Ohlsson, K. (1991) Human cystatin C: role of the N-terminal segment in the inhibition of human cysteine proteinases and its inactivation by leucocyte elastase. Biochem. J., *273*, 621 - 626.

Agarwala, K., L., Kawabata, S., Hirata, M., Miyagi, M., Tsunasawa, S., and Iwanaga, S. (1996) A cysteine protease inhibitor stored in the large granules of horseshoe crab hematocytes: purification, characterisation, cDNA cloning and tissue localisation. J. Biochem., *119*, 85 - 94.

Alavaikko, M., Rinne, A., Järvinen, M., Jokinen, K., and Hopsu-Havu, H. K. (1985) Acid cysteine proteinase inhibitors, a new characteristic of reticulum cells of human lymphoid secondary follicles. Acta Histochem., *77*, 1 - 6.

Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J. H., Sayers, C. A., Sunter, D. C., and Barrett, A. J. (1983) Cystatin, a protein inhibitor of cysteine proteinases. Biochem. J., *211*, 129 - 138.

Angelides, K. J., and Fink, A. L. (1979) Mechanism of thiol protease catalysis: detection and stabilization of a tetrahedral intermediate in papain catalysis. Biochem., *18*, 2363 - 2369.

Aoki, H., Akaike, T., Abe, K., Kuroda, M., Arai, S., Okamura, R., Negi, A., and Maeda, H. (1995) Antiviral effect of oryzacystatin, a proteinase inhibitor of rice, against herpes simplex virus type I *in vitro* and *in vivo.* Antimicrobial Agents Chemo., *39*, 846 -849.

Applebaum, S. W. (1985) Biochemistry of digestion. In: Comprehensive Insect Physiology, Biochemistry and Pharmacology, (Kerkut, G. A., and Gilbert, L. I., eds), Pergamon Press, Vol. 4, New York, pp. 279 – 311.

Arai, S., Watanabe, H., Kondo, H., Emori, Y., and Abe, K. (1991) Papain-inhibitory activity of oryzacystatin, a rice seed cysteine proteinase inhibitor, depends on the central Gln-Val-Val-Ala-Gly region conserved among cystatin superfamily members. J. Biochem., *109*, 294-298.

Argarana, C. E., Kuntz, I. D., Birken, S., Axel, R., and Cantor, C. R. (1986) Molecular cloning and nucleotide sequence of the streptavidin gene. Nucl. Acids Res., *14*, 1871 - 1882.

Argos, P., Kamer, G., Nicklin, M. J. H., and Wimmer, E. (1984) Similarity in gene organisation and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucl. Acids Res., *12*, 7251 - 7267.

Arnold, C., and Hodgson, I. J. (1991) Vectorette PCR: A novel approach to genomic walking. PCR methods and applications, *1*, 39 - 42.

Atkinson, H. J., Urwin, P. E., Hansen, E., and McPherson, M. J. (1995) Designs for engineered resistance to root-parasitic nematodes. Tibtech., *131*, 369 - 374.

Auerswald, E. A., Genenger, G., Assfalg-Machleidt, I., Machleidt, W., Engh, R. A., and Fritz, H. (1992) Recombinant chicken egg white cystatin variants of the QLVSG region. Eur. J. Biochem., *209*, 837 - 845.

Auerswald, E. A., Rossler, D., Mentele, R., and Assfalg-Machleidt, I. (1993) Cloning, expression and characterisation of human kininogen domain 3.

Auerswald, E. A., Nägler, D. K., Schulze, A. J., Engh, R. A., Genenger, G., Machleidt, W., and Fritz, H. (1994) Production, inhibitory activity, folding and conformational analysis of an N-terminal and an internal deletion variant of chicken cystatin. Eur. J. Biochem., *224*, 407 - 415.

Auerswald, E. A., Nägler, D. K., Assfalg-Machleidt, I., Stubbs, M. T., Machleidt, W., and Fritz, H. (1995) Hairpin loop mutations of chicken cystatin have different effects on the inhibition of cathepsin B, cathepsin L and papain. FEBS Lett., *361*, 179 - 184.

Auerswald, E. A., Nägler, D. K., Gross, S., Assfalg-Machleidt, I., Stubbs, M. T., Eckerskorn, C., Machleidt, W., and Fritz, H. (1996) Hybrids of chicken cystatin with human kininogen domain 2 sequences exhibit novel inhibition of calpain, improved inhibition of actinidin and impaired inhibition of papain, cathepsin L and cathepsin B. Eur. J. Biochem., *356*, 1 - 9.

Baker, J. E. (1976) Properties of midgut proteinases in larvae of *Attagenus megatoma*. Insect Biochem., *6*, 143 – 148.

Baker, E. N., and Drenth, J. (1986) In: Biological Macromolecules and Assemblies, vol 3, (Jurnak, F. A., and McPherson, A., eds.), John Wiley and Sons Inc., pp. 313 - 368.

Barr, K. A., Hopkins, S. A., and Sreekrishna, K. (1992) Protocol for efficient secretion of HSA developed from *Pichia pastoris*. Pharm. Eng., *12*, 48 - 51.

Barrett, A. J., Brown, M. A., and Sayers, C. A. (1979) The electrophoretically "slow" and "fast" forms of the α–2-macroglobulin molecule. Biochem. J., *181*, 401 - 418.

Barrett, A. J. (1980) Flourimetric assays for cathepsin B and cathepsin H with methylcoumarylamide substrates. Biochem J., *187*, 909 - 912.

Barrett, A. J., and Kirschke, H. (1981) Cathepsin B, cathepsin H and cathepsin L. Meth. Enzymol., *80*, 535 - 561.

Barrett, A. J., and McDonald, J. K. (1986) Nomenclature: protease, proteinase and peptidase. Biochem. J., *237*, 935.

Barrett, A. J. (1986) An Introduction to the Proteinases. In: Proteinase Inhibitors. (Barrett, A. J., and Salvesen, G., eds), Elsevier Science Publishers, New York, pp. 3 - 21.

Barrett, A. J., Rawlings, N. D., Davies, E., Machleidt, W., Salvesen, G., and Turk, V. (1986) Cysteine proteinase inhibitors of the cystatin superfamily. In: Proteinase Inhibitors. (Barrett, A. J., and Salvesen, G., eds). Elsevier Science Publishers, New York, pp. 515 - 569.

Barrett, A. J. (1987) The cystatins: a new class of peptidase inhibitors. TIBS, *12*, 193 - 196.

Bathurst, I. C., Brennan, S. O., Carrell, R.W., Cousens, L. S., Brake, A. J., and Barr, P. J. (1987) Yeast *KEX2* protease has the properties of a human proalbumin converting enzyme. Science, *235*, 348 – 350.

Baumgartner, B., and Chrispeels, M. J. (1976) Partial characterisation of a protease inhibitor which inhibits the major endopeptidase present in cotyledons of mung beans. Plant Physiol., *58*, 1-6.

Bennetzen, J. L., and Hall, B. D. (1982) Codon selection in yeast. J. Biol. Chem., *257*, 3026-3031.

Berger, A., and Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. Phil. Trans, Roy, Soc, Lond, B257, 249 – 264.

Berti, P. J., Faerman, C. H., and Storer, A. C. (1990) Cooperativity of papain-substrate interaction energies in the S_2 to S_2' subsites. Biochem., *30*, 1394 - 1402.

Birk, Y., Harpaz, I., Ishaya, I., and Bondi, A. (1962) Studies on the proteolytic activity of the beetles *Tenebrio* and *Tribolium*. J. Insect Physiol., *8*, 414 – 429.

Bjorck, L., Grubb, A. and Kjellen, L. (1990) Cystatin C, a human proteinase inhibitor, blocks replication of herpes simplex virus. J. Virol., *64*, 941 - 943.

Björk, I., Ylinenjärvi, K. and Lindahl, P. (1990) Equilibrium and kinetic studies of the interaction of chicken cystatin with four cysteine proteinases. Biol. Chem. Hoppe-Seyler, *371, Suppl.*, 119 - 124.

Björk, I., Pol, E., Raub-Segall, E., Abrahamson, M., Rowan, A., and Mort, J. (1994) Differential changes in the association and dissociation rate constants for binding of cystatins to target proteinases occurring on N-terminal truncation of the inhibitors indicate that the interaction mechanism varies with different enzymes. Biochem J., *299*, 219 - 225.

Björk, I., Brieditis, I., and Abrahamson, M. (1995) Probing the functional role of the N-terminal region of cystatins by equilibrium and kinetic studies of the binding of Gly-11 variants of recombinant human cystatin C to target proteinases. Biochem. J., *306*, 513 - 518.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J, *7*, 2593-2599.

Bode, W., Engh, R., Musil, D., Laber, B., Stubbs, M., Huber, R., and Turk, V. (1990) Mechanism of interaction of cysteine proteinases and their protein inhibitors as compared to the serine proteinase-inhibitor interactions. Biol. Chem. Hoppe-Seyler, *371, Suppl.*, 111 - 118.

Bode, W., and Huber, R. (1992a) Structural basis of the proteinase-protein inhibitor interaction. In: Innovations in Proteinases and their Inhibitors. (Avilés, F. X., ed), Walter de Gruyter, New York, pp. 91 – 121.

Bode, W., and Huber, R. (1992b) N atural protein proteinase inhibitors and their interaction with proteinases. Eur. J. Biochem., *204*, 433 - 451.

Bond, J. S., and Butler, P. E. (1987) Intracellular Proteases. Ann. Rev. Biochem., *56*, 333 - 364.

Bolter, C. J., and Jongsma, M. A. (1995) Colorado potato beetles (*Leptinotarsa decemlineata*) adapt to proteinase inhibitors induced in potato leaves by methyl jasmonate. J. Insect. Physiol., *41*, 1071 - 1078.

Boulter, D., Edwards, G. A., Gatehouse, A. M. R., Gatehouse, J. A. and Hilder, V. A. (1990) Additive protective effects of different plant-derived insect resistance genes in transgenic tobacco plants. Crop Protection, *9*, 3551 - 3554.

Bradford, M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. Anal. Biochem., *72*, 248-254.

Bradford, H. N., Jameson, B. A., Adam, A. A., Wassell, R. P., and Colman, R. W. (1993) Contiguous binding and inhibitory sites on kininogens required for the inhibition of platelet calpain. J. Biol. Chem., *268*, 26546 - 26551.

Brake, A. J., Merryweather, J. P., Coit, D. G., Herberlein, U. A., Masiarz, G., R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P., and Barr, P. J. (1984) α-Factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci., *81*, 4642 - 4646.

Brake, A. J. (1990) α-Factor leader-directed secretion of heterologous proteins from yeast. Meth. Enzymol., *185*, 408 - 420.

Broadway, R. M., and Duffey, S. S. (1986) Plant proteinase inhibitors: mechanism of action and effect on the growth and digestive physiology of larval *Heliothis zea* and *Spodoptera exiqua*. J. Insect Physiol., *32*, 827 – 833.

Brzin, J., Kopitar, M., and Turk, V. (1983) Protein inhibitors of cysteine proteinases. 1) Isolation and characterisation of stefin, a cystosolic proteinase inhibitor of cysteine proteinases from human polymorphonuclear granulocytes. Hoppe-Seylers. Z. Physiol. Chem., *364*, 1475 - 1480.

Brzin, J., Popovič, T., and Turk, V. (1984) Human cystatin C, a new protein inhibitor of cysteine proteinases. Biochem. Biophys. Res. Commun., *118*, 103 - 109.

Brzin, J., Ritonja, A., Popovič, T., and Turk, V. (1990) Low molecular mass protein inhibitor of cysteine proteinases from soybean. Biol. Chem. Hoppe-Seyler, *371. Suppl.*, 167 - 170.

Buckholz, R. G. and Gleeson, M. A. G. (1991) Yeast systems for the commercial production of heterologous proteins, Bio/Tech., *9*, 1067 - 1071.

Calkins, C. C., and Sloane, B. F. (1995) Mammalian cysteine proteinase inhibitors: Biochemical properties and possible roles in tumour progression. Biol. Chem. Hoppe-Seyler, *376*, 71 - 80.

Carne, A., and Moore, C. H. (1978) The amino acid sequences of the tryptic peptides from actinidinn, a proteolytic enzyme from the fruit of *Actinidia chinensis*. Biochem J., *173*, 73 - 83.

Carrington, J. C., Freed, D. D., and Oh, C. (1990) Expression of polyviral polyproteins in transgenic plants reveals three proteolytic activities required for complete processing. EMBO J., *9*, 1347 - 1353.

Cervantes, E., Rodrigues, A., and Nicholas, G. (1994) Ethylene regulates the expression of a cysteine proteinase gene during germination of chickpea (*Cicer arietinum L.*). Plant Mol. Biol., *25*, 207 – 215.

Cha, S. (1975) Tight-binding inhibitors - I. Biochem. Pharmacol., *24*, 2177 – 2185.

Chappell, C. L., and Dresden, H. H. (1986) Characterization of a cysteinyl proteinase from the human parasite, *Schistosoma mansoni*. In: Cysteine proteinases and their inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 199 - 208.

Chen, M. S., Johnson, B., Muthukrishnen, S., Kramer, K. J., Morgan, T. D., and Reeck, G. R. (1992) Rice cystatin: bacterial expression, purification, cysteine proteinase inhibitory activity and insect growth suppressing activity of a truncated form of the protein. Prot. Expr. Purif., *3*, 41 - 49.

Chen, H., Cheng, H., and Bjerknes, M. (1993) One-step coomassie brilliant blue R-250. Staining of proteins in polyacrylamide gel. Anal. Biochem., *212*, 295 – 296.

Chou, P. Y., and Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequences. Adv. Enzymol., *47*, 45 – 48.

Christeller, J. T., and Shaw, B. D (1989) The interaction of a range of serine proteinases inhibitors with bovine trypsin and *Costelytra zealandica* trypsin. Insect Biochem., *19*, 233-241.

Chung, C. T., and Miller, R. H. (1988) A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucleic. Acids, Res., *16*, 3580.

Clare, J. J., Romanos, M. A., Rayment, F. B., Rowedder, J. E., Smith, M. A., Payne, M. M., Sreekrishna, K., and Henwood, C. A. (1991) Production of mouse epidermal growth factor in yeast: high-level secretion using *Pichia pastoris* strains containing multiple gene copies. Gene, *105*, 205-212.

Cox, G. N., Pratt, D., Hageman, R., and Boisevenue, R. J. (1990) Molecular cloning and primary sequence of a cysteine protease expressed by *Haemonchus contortus* adult worms. Mol. Biochem. Parasitol., *41*, 25 - 34.

Cregg, J. M., Barringer, K. J., Hessler, A. Y. and Madden, K. R. (1985) *Pichia pastoris* as a host system for transformations. Mol. Cell. Biol., *5*, 3376 – 3385.

Cregg, J. M., and Madden, K. R (1987). Development of yeast transformation systems and construction of methanol-utilisation-defective mutants of *Pichia pastoris* by gene disruption. In: Biological Research on Industrial yeasts. Stewart, G. G., Russell, I., Klein, R. D., and Hiebsch, R. R. (eds.). CRC Press, Boca Raton, Florida.

Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G., and Stillman, C. A. (1989) Functional characterisation of the two alcohol oxidase genes from the yeast *Pichia pastoris*. Mol. Cell. Biol., *9*, 1316 - 1325.

Cregg, J. M., Vedvick, T. S., and Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. Bio/Tech., *11*, 905-910.

Creighton, T. E., and Darby, N. J. (1989) Fuctional evolutionary divergence of proteolytic enzymes and their inhibitors. TIBS, *14*, 319 - 324.

Cunningham, L. (1965) In: Comprehensive Biochemistry (Florkin, M., and Storer, E. H., eds), Vol. 16, Elsevier Publishing Co., Amsterdam, pp. 85.

De Boer, H. A., Comstock,, L. J., and Vasser, M. (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. Proc. Natl. Acad. Sci. U.S.A., *80*, 21 – 25.

Devereux, J., Haeber, P., and Smithies, O. (1984) A comprehensive list of sequence analysis programs for the vax. Nucl. Acids. Res., *12*, 387 - 395.

Diekmann, T., Mitschang, L., Hofmann, M., Kos, V., Turk, V., Auerswald, A., Jaenicke, R., and Oschkinat, H. (1993) The structures of native phosphorylated chicken cystatin and of a recombinant unphosphorylated variant in solution. J. Mol. Biol., *234*, 1048 - 1059.

Dolenč, I., Turk, B., Kos, J., and Turk, V. (1996) Interaction of human cathepsin C with chicken cystatin. FEBS Lett., *392*, 277 – 280.

Downing, K. J. (1997) Biological control of a plant pathogen and pest by expression of a cloned *Serratia marcescens chiA* gene and *Bacillus thuringiensis cry1A(c)* gene in endophytic bacteria. PhD thesis, University of Cape Town, Cape Town, South Africa.

Drenth, J., Jansonius, J. N., Koekkoek, R., and Wolthers, B. (1971) The structure of papain. Adv. Protein Chem., *25*, 79 - 115.

Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W. and Hofnung, M (1984) Sequences of the *MalE* gene and its product, the maltose-binding protein of *Escherichia coli* K12. J. Biol. Chem., *259*, 10606 - 10613.

Eguchi, M. (1993) Protein protease inhibitors of insects and comparison with mammalian inhibitors. Comp. Biochem. Physiol., *105*, 449 - 456.

Ellis, S. B., Brust, P. F., Koutz, P. J., Waters, A. F., Harpold, M. M., and Gingeras, T. R. (1985) Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. Bio/Tech., *7*, 160 - 164.

Elzanowski, A., Baker, W. C., Hunt, L. T., and Seibel-Ross, E. (1988) Cystatin domain in α-2-HS-glycoprotein and fetuin. FEBS Lett., *227*, 167 - 170.

Engh, R. A., Diekmann, T., Bode, W., Auerswald, A., Turk, V., Huber, R., and Oschkinat, H. (1993) Conformational variability of chicken cystatin: comparison of structures determined by X-ray diffraction and NMR spectroscopy. J. Mol. Biol., *234*, 1060 - 1069.

Fernandes, K. V. S., Sabelli, P., Barrette, D. H. P., Richardson, M., Xavier-Filho, J., and Shewry, P. R. (1993) The resistance of cowpea seeds to bruchid beetles is not related to levels of cysteine proteinase inhibitors. Plant Mol. Biol., *23*, 215 - 219.

Fossum, K. (1970) Proteolytic enzymes and biological inhibitors. III. Naturally occurring inhibitors in some animal and plant materials and their effect upon enzymes of various origins. Acta Pathol. Microbial. Scand. Microbiol., *78*, 741 - 754.

Garcia, J. A., Cervera, M., Riechmann, J. L., and Lopez-Otin, C. (1993) Inhibitory effects of human cystatin C on plum pox potyvirus proteases. Plant Mol. Biol., *22*, 697 - 701.

Gatehouse, A. M. R., Gatehouse, J. A., Dobie, P., Kilminster, A. M., and Boulter, D. (1979) Biochemical basis of insect resistance in *Vigna unguiculata*. J. Sci. Food. Agric., *30*, 948 - 958.

Gatehouse, A. M. R, and Bouter, D. (1983) Assessment of the antimetabolic effects of trypsin inhibitors from cowpea (*Vigna unquiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus*. J. Sci. Agric., *34*, 345 – 350.

Gatehouse, A. M. R., Butler, K. J., Fenton, K. A., and Gatehouse, J. A. (1985) Presence and partial characterisation of a major proteolytic enzyme in the larval gut of *Callosobruchus maculatus*. Entomol. Exp. Appl., *39*, 279 - 286.

Ge, A. Z., Pfister, R. M., and Dean, D. H. (1990) Hyperexpression of a *Bacillus thuringiensis* delta-endotoxin-encoding gene in *Escherichia coli*: properties of the product. Gene, *93*, 49 - 54.

Goldstein, Z., Trop, M., and Birk, Y. (1973) Multifunctional proteinase inhibitor from *Bauhinia* seeds. Nature, New Biol., *246*, 29 - 31.

Green, G. D. J., Kembhaui, A. A., Davies, M. E., and Barrett, A. J. (1984) Cystatin-like cysteine proteinase inhibitors from human liver. Biochem. J., *218*, 939 - 946.

Grubb, A., Lofberg, H., and Barrett, A. J. (1984) The disulphide bridges of human cystatin C (γ–trace) and chicken cystatin. FEBS Lett., *170*, 370 - 374.

Grubb, A., Abrahamson, M., Olafsson, I., Trojnar, J., Kasprzykowska, R., Kasprzykowski, F., and Grzonka, Z. (1990) Synthesis of cysteine proteinase inhibitors structurally based on the proteinase interacting N-terminal region of human cystatin C. Biol. Chem. Hoppe-Seyler, *371, Suppl.*, 137 - 144.

Guan, C., Li, P., Riggs, P. D., and Inouye, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusions to maltose binding protein. Gene, *67*, 21 - 30.

Guan, K., and Dixon, J. E. (1991) Eukaryotic proteins Expressed in *Escherichia coli:* an improved thrombin cleavage and purification procedure of fusion with glutathione S-transferase. Anal. Biochem., *192*, 262 - 267.

Guerrero, F. D., Jones, J. T., and Mullet, J. E. (1990) Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes. Plant Mol. Biol., *15*, 11 - 26.

Hadfield, C., Raina, K. K., Shashi-Menon, K., and Mount, R. C. (1983) The expression and performance of cloned genes in yeast. Mycol. Res., *97*, 897 - 944.

Hall, A., Håkansson, K., Mason, R. W., Grubb, A., and Abrahamson, M. (1993) Importance of the evolutionary conserved glycine residue in the N-terminal region of human cystatin C (Gly-11) for cysteine endopeptidase inhibition. Biochem. J., *291*, 123 - 129.

Hall, A., Håkansson, K., Mason, R. W., Grubb, A., and Abrahamson, M. (1995) Structural basis for the biological specificity of cystatin C. J. Biol. Chem., *270*, 5115 – 5121.

Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J., and Tanaka, I. (1978) Isolation and characterization of E-64, a new thiol proteinase inhibitor. Agric. Biol. Chem., *42*, 523 – 528.

Hatano, K., Kojima, M., Tanokura, M., and Takahashi, K. (1995) Primary structure, sequence-specific H-NMR assignments and secondary structure in solution of bromelain inhibitor VI from pineapple stem. Eur. J. Biochem., *232*, 335 - 343.

Heinikoff, S., and Cohen, E. H. (1984) Sequence responsible for the transcription termination on a gene segment in *Saccharomyces cerevisiae*. Mol. Cell Biol., *4*, 1515 – 1520.

Heinemann, U., Pal, G. P., Hilgenfeld, R., and Saenger, W. (1982) Crystal and molecular structure of the sulfhydryl protease calotropin DI at 3.2 Å resolution. J. Mol. Biol., *161*, 591 - 606.

Henderson, P. J. F. (1972) A linear equation that describes the steady-state kinetics of enzymes and subcellular particles interacting with tightly bound inhibitors. Biochem. J., *127*, 321 – 333.

Heussen, C., and Dowdle, E. B. (1980) Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulphate and co-polymerised substrates. Anal. Biochem., *102*, 196 - 202.

Heussen-Schemmer, C., and Dowdle, E. B. (1993) Erythrina protease inhibitor: kinetics of interaction with tissue plasminogen activator. South African Journal of Science, *89*, 131 – 135.

Hilder, V. A., Gatehouse, A. M. R., Sheerman, S. E., Barker, R. F. and Boulter, D. (1987) A novel mechanism of insect resistance engineered into tobacco. Nature, *330*, 160 - 163.

Hines, M. E., Osuala, C. I., and Nielsen, S. S. (1991) Isolation and partial characterisation of a soybean cystatin cysteine proteinase inhibitor of coleopteran digestive proteolytic activity. J. Agric. Food Chem., *39*, 1515-1520.

Hirashiki, I., Ogata, F., Yoshida, N., Makisumi, S., and Ito, A. (1990) Purification and complex formation analysis of a cysteine proteinase inhibitor (cystatin) from seeds of *Wisteria floribunda*. J. Biochem., *108*, 604-608.

Hopsu-Havu, V. K., Joronen, I., Rinne, A., and Jarvinen, M. (1985) Production of acid and neutral cysteine proteinase inhibitors by a cultured human skin epithelial line. Arch. Dermatol. Res., *277*, 452 - 456.

Horn, T., and Urdea, M. S. (1988) Solid supported hydrolysis of apurinic sites in synthetic oligonucleotides for rapid and efficient purification on reverse-phase cartridges. Nucl. Acids Res., *16,* 11559-11571.

Houseman, J. G. and Downe, A. E. R. (1980) Endoproteinase activity in the posterior midgut of *Rhodnius prolixus Stal* (Hemiptera: Reduviidae). Insect Biochem., *10*, 363 - 366.

Houseman., J. G., and Downe, A. E. R. (1983) Cathepsin D-like activity in the posterior midgut of Hemipteran insects. Comp. Biochem. Physiol., *75B*, 509 – 512.

Houseman, J. G., Morrison, P. E., and Downe, A. E. (1985) Cathepsin B and aminopeptidase in the posterior midgut of *Phymata wolffii* (Hemiptera: Phymatidae). Can. J. Zool., *63*, 1288 – 1291.

Hultman, T., Stahl, S., Hornes, E., and Uhlen, M. (1989) Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucl. Acids, Res., *17*, 4937 - 4946.

Hunt, L. T., George, D. G., and Barker, W. C. (1985) The prokaryote-eukaryote interface. Biosystems, *18*, 223 – 240.

Illy, C., Qurashi, O., Wang, J., Purisima, E., Vernet, T. (1997) Role of the occluding loop in cathepsin B activity. J. Biol. Chem., *272*, 1197 – 1202.

Innis, M. A., Holland, M. J., and MacCabe, P. C. (1985) Expression, glycosylation and secretion of an *Aspergillus* glucoamylase by *Saccharomyces cerevisiae*. Science, *228*, 21-26.

Irie, K., Hosoyama, H., Takeuchi, T., Iwabuchi, K., Watanabe, H., Abe, M., Abe, K., and Arai, S. (1996) Transgenic rice established to express corn cystatin exhibits strong inhibitory activity against insect gut proteinases. Plant Mol. Biol., *30*, 149 - 157.

Irvine, J. W., Coombs, G. H., and North, M. J. (1992) Cystatin-like cysteine proteinase inhibitors of parasitic protozoa. FEBS Micro. Lett., *96*, 67 -72.

Isemura, S., Saitoh, K., and Minakata, K. (1991) Identification of full sized forms of salivary (S-type) cystatins (Cystatin SN, Cystatin SA, Cystatin S, and two phosphorylated forms of Cystatin S) in human whole saliva and determination of phosphorylation sites of Cystatin S. J. Biochem., *110,* 648-654.

Järvinen, M. (1978) Purification and some characteristics of the human epidermal SH-protease inhibitor. J. Invest. Dermatol., *71*, 114 - 118.

Jerala, R., Tritenjak-Prebanda, M., Kroon-Žitko, L., Lenarčič, B., and Turk, V. (1990) Mutations in the QVVAG region of the cysteine proteinase inhibitor stefin B. Biol. Chem. Hoppe-Seyler, *371*, *Suppl,* 157 - 160.

Jerala, R., Kroon-Žitko, L., Kopitar, N., Popovič, T., and Turk, V. (1991) Deletion of the carboxy terminal part of stefin B does not have a major effect for binding to papain. Biochem. Biomed. Acta., *50*, 627 – 629.

Jerala, R. (1992) Conformation and fluctuations of free stefin B: A molecular dynamics study. Biol. Chem. Hoppe-seyler, *373*, 447 – 452.

Jerala, R., Kroon-Žitko, L., and Turk, V. (1994) Improved expression and evaluation of polyethyleneimine precipitation in isolation of recombinant cysteine proteinase inhibitor Stefin B. Protein Expr. Purif., *5*, 65 - 69.

John, R. A. (1992) In: Enzyme Assays, A Practical Approach, (Eisenthal, R., and Danson, M. J., eds), IRL Press, Oxford, pp. 59 - 92.

Johnson, R., Narvaez, J., and Ryan, C. A. (1989) Expression of proteinase inhibitors I and II in transgenic tobacco plants; effects on natural defence against *Munduca sexta* larvae. Proc. Natl. Acad. Sci. U.S.A., *86*: 9871 - 9875.

Johnston, M. (1987) A model fungal gene regulatory mechanism: the *GAL* genes of *Saccharomyces cerevisiae*. Microbiol. Rev., *51*, 458 - 476.

Kaji, H., Samejima, T., Kumagai, I., Hibino, T., Miura, K., and Takeda, A. (1990) Efficient preparation of human recombinant cystatin A by *Escherichia coli.* Biol. Chem. Hoppe-Seyler, *371, Suppl,* 145 – 150.

Kajiwara, K., Fujita, A., Tsuyuki, H., Kumazaki, T., and Ischi, I. (1991) Interactons of *Streptomyces* serine-protease inhibitors with *Streptomyces griseus* metalloendopeptidase II. J. Biochem., *110*, 350 - 354.

Kato, H., Nagasawa, S., and Iwanaga, S. (1981) High molecular weight kininogens. Meth. Enzymol., *80*, 172 - 198.

Katunuma, N., and Kominami, E. (1986) Distributions and localizations of lysosomal cysteine proteinases and cystatins. In: Cysteine proteinases and their Inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 219 - 227.

Keilova, H., and Tomasek, V. (1974) Effect of papain inhibitor from chicken egg white on cathepsin B1. Biochem. Biophys. Acta, *334*, 170 -186.

Kellermann, J., Lottspeich, J., and Müller-Esterl, W. (1986) Completion of the primary structure of human high-molecular-mass kinin. The amino acid sequence of the entire heavy chain and evidence for its evolution by gene triplication. Eur. J. Biochem., *154*, 471 - 478.

Koizumi, M., Yamaguchi-Shinozaki, Tsuji, H., and Shinozaki, K. (1993) Structure and expression of two genes that encode distinct drought inducible cysteine proteinases in *Arabidopsis thaliana.* Gene, *129*, 175 - 182.

Kondo, H., Emori, Y., Abe, K., Susuki, K., and Arai, S. (1989) Cloning and sequence analysis of the genomic fragment encoding oryzacystatin. Gene, *81*, 259 - 265.

Kondo, H., Abe, K., Nishimura, I., Watanabe, H., Emori, Y., and Arai, S. (1990) Two distinct cystatin species in rice seeds with different specificities against cysteine proteinases. J. Biol. Chem., *265,* 15832-15837.

Kondo, H., Abe, K., Emori, Y., and Arai, S. (1991) Gene organisation of oryzacystatin II, a new cystatin superfamily member of plant origin, is closely related to that of oryzacystatin I but different from those of animal cystatins. FEBS Lett., *278*, 87 - 90.

Kondo, H., Ijiri, S., Abe, K., Maeda, H., and Arai, S. (1992) Inhibitory effect of oryzacystatins and a truncation mutant on the replication of poliovirus in infected vero cells. FEBS Lett., *299*, 48 – 50.

Korant, B. D., Brzin, J., and Turk, V. (1985) Cystatin, a protein inhibitor of cysteine proteases alters viral protein cleavages in infected human cells. Biochem. Biophys. Res. Commun., *127*, 1072 -1076.

Korant, R., Towatari, T., Ivanoff, L., Kettner, C., Cordova, A., and Petteway, S. (1986) Viruses as vectors for cysteine proteinases. In: Cysteine proteinases and their Inhibitors. (Turk, V., ed), Walter de Gruyter, Berlin, 295 – 305.

Koutz, P., Davies, G. R., Stillman, C., Barringer, K., Cregg, J., and Thill, G. (1989) Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast, *5*, 167 – 177.

Križaj, I., Drobnic-Kosorok, M., Brzin, J., Jerala, R., and Turk, V. (1993) The primary structure of an inhibitor of cysteine proteinases from potato. FEBS Lett., *333*, 15 - 20.

Kuroda, M., Ishimoto, M., Suzuki, K., Kondo, H., Abe, K., Kitamura, K., and Arai, S. (1996) Oryzacystatins exhibit growth-inhibitory and lethal effects on different species of bean insects, *Callosobruchus chinensis* (Coleoptera) and *Riptortus clavatus* (Hemiptera). Biosci. Biotech. Biochem., *60*, 209 – 212.

La Grange, D. C. (1995) Molecular cloning, characterisation and expression of the β-xylanase cDNA gene XYN2 of the fungus *Trichoderma reseei* in the yeast *Saccharomyces cerevisiae*. MSc Thesis, University of Stellenbosch, Stellenbosch, South Africa.

Laber, B., Krieglstein, K., Henschen, A., Kos, J., Turk, V., Huber, R., and Bode, W. (1989) The cysteine proteinase inhibitor chicken cystatin is a phosphoprotein. FEBS Lett., *248*, 162 - 168.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of the bacteriophage T. Nature, *277*, 680 - 685.

Lah, T. T., Kokaij-Kunovar, M., and Turk, V. (1990) Cysteine proteinase inhibitors in human cancerous tissues and fluids. Biol. Chem. Hoppe-Seyler, *371, Suppl*, 199 - 203.

Lalmanach, G., Hoebeke, J., Moreau, T., Brillard-Bourdet, M., Ferrer-Di Martino, M., Borras-Cuesta, F., Gauthier, F. (1993) Interaction between cystatin-derived peptides and papain. J. Protein Chem., *12*, 23 -31.

Lantz, M. S., and Ciborowski, P. (1994) Zymographic techniques for detection and characterisation of microbial proteases. Meth. Enzymol., *235*, 563 – 594.

Laroche, Y., Storme, V., De Meutter, J., Messens, J., and Lauweeys, M. (1994) High-level secretion and very efficient isotopic labelling of tick anticoagulant peptide (TAP) expressed in the methylotropic yeast *Pichia pastoris*. Bio/Tech, *12*, 1119-1124.

Laskowski, M., and Kato, I (1980) Proteinase inhibitors of proteinases. Ann. Rev. Biochem., *49*, 593 - 626.

Lenarčič, B., Ritonja, A., Dolenč, I., Stoka, V., Berbic, S., Pingercar, J., Strukelj, B., and Turk, V. (1993) Pig leukocyte cysteine proteinase inhibitor (PLCPI), a new member of the stefin family. FEBS Lett., *336*, 289 - 292.

Lenarčič, B., Križaj, I., Žunec, P., and Turk, V. (1996) Differences in specificity for the interactions of stefins A, B and D with cysteine proteinases. FEBS Lett., *395*, 113 – 118.

Leplé, J. C., Bonadé-Bottino, M., Augustin, S., Pilate, G., Lê Tãn, V. D., Delplanque, A., Cornu, D., Jouanin, L. (1995) Toxicity of *Chrysomela tremulae* (Coleoptera *L Chrysomelidae*) of transgenic poplars expressing a cysteine proteinase inhibitor. Molecular Breeding, *1*, 319 -328.

Liang, C., Brookhart, G., Feng, G. H., Reeck, G. R., and Kramer, K. J. (1991) Inhibition of digestive proteinases of stored grain coleoptera by oryzacystatin, a cysteine proteinase inhibitor from rice seed. FEBS Lett., *278*, 139-142.

Lim, C. O., Lee, S. I., Chung, W. S., Park, S. H., Hwang, I., and Cho, M. J. (1996) Characterisation of a cDNA encoding cysteine proteinase inhibitor from Chinese cabbage (*Brassica campestris L. ssp. pekinensis*) flower buds. Plant Mol. Biol., *30*, 373 -379.

Lindahl, P., Abrahamson, M., and Björk, I. (1992) Interaction of recombinant human cystatin C with the cysteine proteinases papain and actinidin. Biochem. J., *281*, 49 -55.

Lindahl, P., Nyacander, M., Ylinenjärvi, K., Pol, E., and Björk, I. (1992) Characterisation by rapid-kinetic and equilibrium methods of the interaction between N-terminally truncated forms of chicken cystatin and the cysteine proteinases papain and actinidin. Biochem. J., *286*, 165 - 171.

Lindahl, P., Ripoll, D., Abrahamson, M., Mort, J. S., and Storer, A. C. (1994) Evidence for the interaction of valine-10 in cystatin C with the S_2-subsite of cathepsin B. Biochemistry, *33*, 4384 - 4392.

Lonsdale-Eccles, J. D., and Grab, D. J. (1986) Proteases in African trypanosomes. In: Cysteine proteinases and their inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 189 - 197.

Mach, L., Mort, J. S., and Glossl, J. (1994) Non-covalent complexes between the lysosomal proteinase cathepsin B and its propeptide account for stable, extracellular, high molecular mass forms of the enzyme. J. Biol. Chem., *269*, 13036 - 13040.

Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, A. E., Wiegand, G., Kos, J., Turk, V., and Bode, W. (1989) Mechanism of inhibition of papain by chicken egg white cystatin: inhibition constants of N-terminally truncated forms and cyanogen bromide fragments of the inhibitor. FEBS Lett., *243*, 234-238.

Machleidt, W., Thiele, U., Assfalg-Machleidt, A. E., Forger, D., and Auerswald, E. A. (1991) Molecular mechanism of inhibition of cysteine proteinases by their protein inhibitors: kinetic studies with natural and recombinant variants of cystatins and stefins. Biomed. Biochim. Acta, *50*, 613 - 620.

Machleidt, W., Nägler, D. K., Assfalg-Machleidt, I., Stubbs, M. T., Fritz, H., and Auerswald, E. A. (1995) Temporary inhibition of papain by hairpin loop mutants of chicken cystatin. Distorted binding of the loops results in cleavage of the Gly^{9}-Ala^{10} bond. FEBS Lett., *361*, 185 - 190.

Maina, C. V., Riggs, P. D., Andres, G. G., Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, l. A., and Di, Guan, C. (1988) An *Escherichia coli* vector to express and purify foreign proteins by fusion to and separation from the maltose-binding protein. Gene, *74*, 365-373.

Mandecki, W. (1986) Oligonucleotide-directed double strand break repair in plasmids of *Escherichia coli*: A method for site-specific mutagenesis. Proc. Natl. Acad. Sci. U.S.A., *83*, 7177 – 7181.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Marston, F. O. A. (1986) The purification of eukaryotic polypeptides synthesised in *Escherichia coli*. Biochem J., *240*, 1-12

Martin, J. R., Jerala, R., Kroon-Žiko,L., Žerovnik, E., Turk, V. and Waltho, J. P. (1994) Structural characterisation of human stefin A in solution and implications for binding to cysteine proteinases. Eur. J. Biochem., *225*, 1181 - 1194.

Martin, J. R., Craven, C. J. Jerala, R., Kroon-Žitko, L., Žerovnik, E., Turk, V., and Waltho, J. P. (1995) The three-dimensional solution structure of human stefin A. J. Mol. Biol., *246*, 331 - 343.

Mason, R. W., Greem, G. D. J., and Barrett, A. J. (1985) Human liver cathepsin L. Biochem J., *226*, 233 - 241.

Masoud, S. A., Johnson, L. B., White, F. F., and Reeck, G. R. (1993) Expression of a cysteine proteinase inhibitor (oryzacystatin-I) in transgenic tobacco plants. Plant Mol. Biol., *21*, 655 - 663.

Masoud, S. A., Ding, X., Johnson, L. B., White, F. F., Reeck, G. R. (1996) Expression of a corn bifunctional inhibitor of serine proteinases and insect α-amylases in transgenic tobacco. Plant Sci., *115*, 59 - 69.

Matsumoto, K., Murata, M., Sumiya, S., Kitamura, K., and Ishida, T. (1994) Clarification of substrate specificity of papain by crystal analysis of complexes with covalent-type inhibitors. Biochem. Biophys. Acta, *1208*, 268 - 276.

Matsumoto, I., Watanabe, H., Abe, K., Arai, S., and Emori, Y. (1995) A putative digestive cysteine proteinase from *Drosophila melanogasteir* is predominantly expressed in the embryonic larval midgut. Eur. J. Biochem., *227*, 582 - 587.

McManus, M. T., White, D. W. R., McGregor, P. G. (1994) Accumulation of a chymotrypsin inhibitor in transgenic tobacco can affect the growth of insect pests. Transgenic. Res., *3*, 50 - 58.

Merril, C. R., Goldman, D., and Van Keuren, M. L. (1983) Silver staining methods for polyacrylamide gel electrophoresis. Meth. Enzymol., *96*, 230 - 239.

Michaels, M. L., Hsiao, H. M., and Miller, J. H. (1992) Using PCR to extend the limit of oligonucleotide synthesis. Bio/Tech., *12*, 45 - 47.

Michaud, D., Nguyen-Quoc, B., and Yelle, S. (1993) Selective inhibition of Colorado potato beetle cathepsin H by oryzacystatins I and II. FEBS Lett., *331*, 173 - 176.

Michaud, D., Nguyen-Quoc, B., and Yelle, S. (1994) Production of oryzacystatins I and II in *Escherichia coli* using the glutathione S-transferase gene fusion system. Biotechnol. Prog., *10*, 155 - 159.

Michaud, D., Bernier-Vadnis, N., Overney, S., and Yelle, S. (1995a) Constitutive Expression of digestive cysteine proteinase forms during development of the Colorado potato beetle, *Leptinotarsa decemlineata* Say (*Coleoptera: Chrysomelidae*). Insect Biochem. Molec. Biol., *25*, 1041 -1048.

Michaud, D., Cantin, L., and Vrain, T. C. (1995b) Carboxy-terminal truncation of oryzacystatin II by oryzacystatin-insensitive insect digestive proteinases. Arch. Biochem. Biophys., *322*, 469.

Michaud, D., Cantin, L., Bonadé-Bottino, Jouanin, L., and Vrain, T. C. (1996) Identification of stable plant cystatin/nematode proteinase complexes using mildly denaturing gelatin/polyacrylamide gel electrophoresis. Electrophoresis, *17*, 1373 - 1379.

Mizuno, K., Nakamura, T., Ohishima, T., Tanaka, S. and Matsuo, H. (1989) Characterisation of *KEX2*-encoded endopeptidase from yeast *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun., *159*, 305-311.

Moreau. T. H., Hoebeke, T., lalmanach, M., Haltab, M., and Gauthier, F. (1990) Simulation of the inhibitory cystatin surface by a synthetic peptide. Biochem. Biophys. Res. Commun., *167*, 117 – 122.

Morrison, J. F. (1982) The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. TIBS, 7, 102 – 105.

Müller-Esterl, W., Iwanaga, S., and Nakanishi, S. (1986) Kininogens revisited. TIBS, *11*, 336 - 339.

Murachi, T. (1983) Calpain and calpastatin. TIBS, *8*, 167 - 169.

Murdock, L. L., Brookhart, G., Dunn, P. E., Foard, D. E., Kelley, S., Kitch, L., Shade, R. E., Shukle, R. H., and Wolfson, J. L. (1987) Cysteine digestive proteinases in Coleoptera. Comp. Biochem. Physiol., *87*, 783 - 787.

Murdock, L. L., Shade, R. E., and Pomeroy, M. A. (1988) Effects of E-64, a cysteine proteinase inhibitor, of cowpea weevil growth, development and fecundity. Enviro. Entomol., *17*, 467-469.

Musil, D., Zucic, D., Turk, D., Engh, R. E., Mayr, I., Huber, R., Popovič, T., Turk, V., Towatari, T., Kaunuma, N., and Bode, W. (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. EMBO J., *10*, 2321 -2330.

Nagai, K., and Thøgersen, H., C. (1984) Generation of β-globulin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli.*, Nature, 309, 810-812.

Nagai, K., and Thøgersen, H., C. (1988) Synthesis and sequence-specific proteolysis of hybrid proteins produced in *Escherichia coli.* Meth. Enzymol., *153*, 461 - 481

Nelson, P. S., Kent, M., and Muthini, S. (1992) Oligonucleotide labeling methods 3. Direct labelling of oligonucleotides employing a novel, non-nucleosidic, 2-aminobutyl-1,3-propanediol backbone. Nucl. Acids, Res., *20*, 6253 – 6259.

Nicklin, M. J. H., and Barrett, A. J. (1984) Inhibition of cysteine proteinases and dipeptidyl peptidase I by egg-white cystatin. Biochem. J., *223*, 245 - 253.

Nikawa, T., Towatari, T., Ike, Y., and Katunuma, N. (1989) Studies on the reactive site of the cystatin superfamily using recombinant cystatin A mutants: Evidence that the QVVAG region is not essential for cysteine proteinase inhibitory activities. FEBS Lett., *255*, 309 - 314.

North, M. J. (1982) Comparative biochemistry of the proteinases of eukaryotic microorganisms. Microbiol. Reviews, *46*, 308-340.

North, M. J., Scott, K. I., and Lockwood, B. C. (1988) Multiple cysteine proteinase forms during the life cycle of *Dictostelium discoideum* revealed by electrophoretc analysis. Biochem. J., *254*, 261 – 268.

North, M. J. (1992) The characteristics of cysteine proteinases of parasitic protozoa. Biol. Chem. Hoppe Seyler, *373*, 401 - 406.

Nycander, M., and Björk, I. (1990) Evidence by chemical modification that tryptophan-104 of the cysteine-proteinase inhibitor chicken cystatin is located in or near the proteinase-binding site. Biochem. J., *271*, 281 – 284.

Ohi, H., Ohtani, W., Okazaki, N., Furuhata, N., and Ohmura, T. (1996) Cloning and characterization of the *Pichia pastoris PRC1* gene encoding carboxypeptidase Y. Yeast, *12*, 31 - 40.

Ohkubo, I., Kurachi, K., Takasawa, T., Shikawa, H., and Sasaki, M. (1984) Isolation of a human cDNA for the α2-thiol proteinase inhibitor and its identity with low molecular weight kininogen. Biochemistry, *23*, 5691 - 5697.

Okamoto, H., and Greenbaum, L. M. (1983) Isolation and structure of T-kinin. Biochem. Biophys. Res. Commun., *112*, 701 - 708.

Olivia, M. L. V., Sampaio, M. U., and Sampaio, C. A. M. (1988) Purification and partial characterisation of a thiol protease inhibitor from *Enterolobium contortisiliquum* beans. Biol. Chem. Hoppe-Seyler, *369*, 229 - 232.

Oshika, Y., Yamada, T., Nakagawa, S., Fujishima, A., Kawase, M., Ishibashi, Y., and Fukuda, T. (1994) Human parathyroid hormone: efficient synthesis in *Escherichia coli* using a synthetic gene, purification and characterisation. Int. J. Peptide Protein Res., *43*, 441 - 447.

Orr, G., Strickland, J. A., and Walsh, T. A. (1994) Inhibition of *Diabrotica* larval growth by a multicystatin from potato tubers. J. Insect Physiol., *40*, 893 - 900.

Pendola, S., and Greenberg, B. (1975) Substrate specific analysis of proteolytic enzymes in the larval midgut of *Calliphora vicina*. Ann. Ent. Soc. Am., *68*, 341 - 345.

Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, S. L., and Fischhoff, D. A. (1991) Modification of the coding sequence enhances plant expression of insect control proein genes. Proc. Natl. Acad. Sci. U.S.A., *88*, 3324-3328.

Perlstein, S. H., and Kezdy, F. J. (1973) Isolation and characterisation of a protease inhibitor from commercial stem bromelain acetone powder. J. Supramol. Struct., *1*, 249 - 254.

Pickersgill, R. W. (1988) The electrostatic fields in the active site clefts of actinidin and papain. Biochem. J., *254*, 235 – 238.

Pickersgill, R. W., Rizkallah, P., Harris, G. W., and Goodenough, P. W. (1991) Determination of the structure of papaya protease omega. Acta Crystallogr. Sect. B: Struct. Sci., *B47*, 766-771.

Pol, E., Olsson, S., Estrada, S., Prasthofer, T. W., and Björk, I. (1995) Characterisation by spectroscopic, kinetic and equilibrium methods of the interaction between recombinant human cystatin A (stefin A) and cysteine proteinases. Biochem. J., *311*, 275 - 282.

Polgár, L. (1973) On the mode of activation of the catalytically essential sulfhydryl group of papain. Eur. J. Biochem., *33*, 104 - 109.

Polgár, L., Asboth, B., and Korodi, I. (1986) Mechanism of action of cysteine proteinases: 1) Differences from serine enzymes 2) The second thiol group of chymopapain. In: Cysteine proteinases and their inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 327 - 338.

Ponce, M. R., and Micol, J. L. (1991) PCR amplification of long DNA fragments. Nucl. Acids Res., *20*, 623.

Ponz, F., Glascock, C. B., and Bruening, G. (1988) An inhibitor of polyprotein processing with the characteristics of a natural virus resistance factor. Mol. Plant-Microbe Interactions, *1*, 25 - 31.

Popovič, T., Brzin, J., Kos, L., Lenarčič, B., Machleidt, W., Ritonja, A., Hanada, K., and Turk, V. (1988) A new purification procedure of human kidney cathepsin H, its properties and kinetic data. Biol. Chem. Hoppe-seyler, *369, Suppl.*, 175 - 183.

Pratt, D., Cox, G. N., Milhausen, M. J., and Boisvenue, R. J. (1990) A developmentally regulated cysteine protease gene family in *Haemonchus contortus*. Mol. Biochem. Parasitol., *43*, 181 -192.

Ramasubbu, N., Weaver, T., Tseng, C. C., Bobek, L. A., and Levine, M. J. (1996) Preliminary X-ray crystallographic analysis of human salivary cystatin. Acta Cryst., D52, 869 – 870.

Rawlings, N. D., and Barrett, A. J. (1990) Evolution of proteins of the cystatin superfamily. J. Mol. Evol., *30*, 60 - 70.

Rele, M. V., Vartak, H. G., and Jagannanthan, V. (1980) Proteinase inhibitors from *Vigna unguiculata* subsp. *cylindrica*. Occurrence of thiol proteinase inhibitors in plants and purification from *Vigna unguiculata* subsp. *cylindrica*. Arch. Biochem. Biophys., *204*, 117-128.

Richardson, M. (1977) The proteinase inhibitors of plants and microorganisms. Phytochemistry, *16*, 159 - 169.

Rinne, A., Järvinen, M., Räsänen, O., and Dorn, A. (1980) Occurrence of the epidermal SH-protease inhibitor in normal human epithelia and in some human neoplasms. An immunological study. Acta. Histochem. Suppl., *74*, 75 - 79.

Rinne, A., Kirschke, H., Järvinen, M., Hopsu-Havu, V. K., Wiederanders, B., and Bohley, P. (1985) Localisation of cathepsin H and its inhibitors in the skin and other stratified epithelia. Arch. Dermatol. Res., *277*, 190 - 194.

Rodis, P. (1974) Nature and function of cuboidal protein crystals in *Solaneum tuberosum* L. PhD Thesis, Purdue University, West Lafayette, IN.

Rodis, P., and Hoff, J. E. (1984) Naturally occurring protein crystals in potato. Plant Physiol., *74*, 907-911.

Rogers, J. C., Dean, D., and Heck, G. R. (1985) Aleurain: a barley thiol protease related to mammalian cathepsin H. Proc. Natl. Acad. Sci. U.S.A., *82*, 6512 - 6516.

Rogers, B. L., Pollock, J., Klapper, D. G., and Griffith, I. J. (1993) Sequence of the proteinase-inhibitor cystatin homologue from the pollen of *Ambrosia artemisiifolia* (short ragweed). Gene, *133*, 219 - 221.

Romanos, M. A., Clare, J. J., Beesley, K. M., Rayment, F. B., Ballantine, S. P., Makoff, A. J., Dougan, G., Fairweather, N. F., and Charles, I. G. (1991) Recombinant *Bordetella pertussis* pertactin p69 from the Yeast *Pichia pastoris*. High Level Production and Immunological Properties. Vaccine, *9*, 901 - 906.

Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) Foreign gene expression in yeast: a review. Yeast, *8*, 423-488.

Romanos, M. A. (1995) Advances in the use of *Pichia pastoris* for high-level gene expression. Current Opinion Biotech., *6*, 527-533.

Rosenburg, A. H., Lade, B. N., Chui, D., Lin, S., Dunn, J. J., and Studier, F. W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene, *56*, 125 - 135.

Rothe, M., Zichner, A., Auerswald, E. A., and Dodt, J. (1994) Structure/function implications for the aminopeptidase specificity of aleurone. Eur. J. Biochem., *224*, 559 - 565.

Ryan, C. A. (1973) Proteolytic enzymes and their inhibitors in plants. Annu. Rev. Plant Physiol., *24*, 173 - 196.

Ryan, C. A. (1989) Proteinases inhibitor gene families: strategies for transformation to improve plant defence against herbivores. Bioassays, *10*, 20 - 24.

Ryan, C. A. (1990) Protease Inhibitors in plants: genes for improving defences against insects and pathogens. Ann. Rev. Phytopath., *28*, 425-449.

Salvesen, G., Parkes, C., Rawlings, N. D., Brown, M. A., Barrett, A. J., Abrahamson, M., and Grubb, A. (1986) Cystatin-like domains of LMW-kininogen, and speculations on the evolution of cystatins. In: Cysteine proteinases and their inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 413 - 428.

Salvesen, G., Parkes, C., Abrahamson, M., Grubb, A., and Barrett, A. J. (1986) Human low-M_r kininogen contains three copies of a cystatin sequence that are divergent in structure and in inhibitory activity for cysteine proteinases. Biochem. J., *234*, 429 - 434.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning. A laboratory manual. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, R., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A., *74*, 3650 - 3654.

Sarath, G., De La Motte, R. S., and Wagner, F. W. (1989) In: proteolytic Enzymes, A Practical approach, (Beynon, R. J., and Bond, J. S., eds), Oxford University Press, Oxford, pp. 25 - 55.

Schägger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulphate-polyacrylamide gel electrophoresiss for the separation of proteins in the rangee from 1 to 100 kDa. Anal. Biochem., *166*, 368 - 379.

Scorer, C. A., Buckholz, R. G., Clare, J. J., and Romanos, M. A. (1993) The intracellular production and secretion of HIV-1 envelope protein in the yeast *Pichia pastoris*. Gene, *136*, 111-119

Serveau, C., Juliano, L., Bernard, P., Moreau, T., Mayer, R., and Gauthier, F. (1994) New substrates of papain, based on the conserved sequence of natural inhibitors of the cystatin family. Biochimie, *76*, 153 - 158.

Sharma, J. N., and Mohsin, S. S. J. (1990) The role of chemical mediators in the pathogenesis of inflammation with emphasis on the kinin system. Exp. Pathol., *38*, 73 - 96.

Sharp, P. M., and Bulmer, M. (1988) Selected differences among translation termination codons. Gene, *63*, 141-145.

Sharp, P. M. and Cowe, E. (1991) Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast, *7*, 657-678.

Shibuya, K., Kaji, H., Ito, T., Tsujikami, A., Tate, S., Takeda, A., Kumagai, I., Hirao, I., Miura, K., Inagaki, F., and Samejima, T. (1995) Human cystatin A is inactivated by engineered truncation. The NH_2-terminal region of the cysteine proteinase inhibitor is essential for expression of its inhibitory activity. Biochem., *34*, 12185 – 12195.

Siffert, O., Emöd, I. and Keil, B (1976) Interaction of clostripain with natural trypsin inhibitor affinity labelling by Na-p-nitrobenzyloxycarbonyl arginine chloromethyl ketone. FEBS Lett., *66*, 114 – 119.

Silhavy, T. J., Benson, S. A., and Emr, S. D. (1983) Mechanisms of protein localisation. Microbiol. Rev., 47, 313 – 344.

Silva, C. P., and Xavier-Filho, J. (1991) Comparison between the levels of aspartic and cysteine proteinases of the larval midguts of *Callosobruchus maculatus* (F.) and *Zabrotes subfasciatus* (BOH.) (Coleoptera: Bruchidae). Comp. Biochem. Physiol., *99B*, 529 - 533.

Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene, *67*, 31 - 40.

Song, I., Taylor, M., Baker, K., and Bateman, J, R. C. (1995) Inhibition of cysteine proteinases by *Carica papaya* cystatin produced in *Escherichia coli*. Gene, *162*, 221 - 224.

Sotiropoulou, G., Anisowicz, A., and Sager, R. (1997) Identification, cloning, and characterisation of cystatin M, a novel cysteine proteinase inhibitor, down-regulated in breast cancer. J. Biol. Chem., *272*, 903 – 910.

Sreekrishna, K., Potenz, R. H. B., Cruze, J. A., McCombie, W. R., Parker, K. A., Nelles, L., Mazzaferro, P. K., Holden, K. A., Harrison, R. G., Wood, P. J., Phelps, D. A., Hubbard, C. E., and Fuke, M. (1988) High level expression of heterologous protein in methylotropic yeast *Pichia pastoris*. J. Basic Microbiol., *28*, 265 - 278.

Stewart, J. M. (1993) The kinin system in inflammation. Agents Actions Suppl., *42*, 145 –147.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarčič, B., and Turk, V. (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J., *9*, 1939-1947.

Tai, J. Y., Kort, A. A., Liu, T. Y. and Elliott, S. D. (1976) Primary structure of streptococcal proteinase III. Isolation of cyanogen bromide peptides: complete covalent structure of the polypeptide chain. J. Biol. Chem., *251*, 1955 - 1959.

Takeda, A., Kaji, H., Nakaya, K., Aoki, Y., Nakamura, Y., and Samejima, T (1985) Amino acid sequence of derivatives of newborn rat epidermal thiol proteinase inhibitors. Biochem. Int., 11, 557 - 564.

Takeda, A., Kaji, H., Nakaya, K., Nakamura, Y., and Samejima, T. (1989) Comparative studies on the primary structure of human cystatin As from epidermis, liver, spleen and leukocytes. J. Biochem., *105*, 986 – 991.

Tashiro, M., and Maki, Z. (1986) Isolation of protein inhibitors of papain, trypsin and α-amylase in the grain of foxtail millet. Agric. Biol. Chem., *50*, 2955 - 2957.

Tate, S., Ushioda, T., Utsunomiya, N., Shibuya, K., Ohyama, Y., Nakano, Y., Kaja, H., Inagaki, F., Samejima, T., and Kainosho, M. (1995) Solution structure of a human cystatin A variant, Cystatin A^{2-98} M65L, by NMR spectroscopy. A possible role of the interactions between the N- and C-termini to maintain the inhibitory active form of cystatin A. Biochem, *34*, 14637 - 14648.

Teixura, L. (1992) Synthesis and Expression of the *Erythrina* trypsin/tissue plasminogen activator (tpa) inhibitor encoding-gene. Genetic dissection to correlate the interaction of *Erythrina* and soybean trypsin inhibitors with tpa. PhD Thesis, University of Cape Town, Cape Town, South Africa.

Terra, W. R., Ferreira, C., and Garcia, E. S. (1988) Origin, distribution, properties and functions of the major *Rhodnius prolixus* midgut hydrolases. Insect Biochem. *18*, 423 -434.

Terra, W. R., and Ferreira, C. (1994) Insect digestive enzymes: properties, compartmentalisation and function. Comp. Biochem. Physiol., *109B*, 1 - 62.

Thie, N. M., and Houseman, J. G. (1989) Identification of cathepsin B, D and H in the larval midgut of Colorado potato beetle, *Leptinotarsa decemlineata Say* (Coleoptera: Chrysomelidae). Insect Biochem., *20*, 313 - 318.

Thiele, U., Assfalg-Machleidt, I., Machleidt, W., and Auerswald, E. A. (1990) N-Terminal variants of recombinant stefin B: effect on affinity for papain and cathepsin B. Biol. Chem. Hoppe-Seyler, *371, Suppl.*, 125 - 136.

Thomas, M. P., Verma, C., Boyd, S. M., and Brocklehurst, K. (1995) The structural origins of the unusual specificity's observed in the isolation of chymopapain and actinidin by covalent chromatography and the lack of inhibition by chymopapain M by cystatin. Biochem J., *306*, 39 - 46.

Thompson, C. J., Movva, N. R., Tizard, R., Crameri, R., Davies, J. E., Lauwereys, M., and Botterman, J. (1987) Characterization of the herbicide-resistance gene *bar* from *Streprrtomyces hygroscopicus*. EMBO J., *6*, 2519 – 2523.

Towbin, H., Staehelin, T., and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets. Proc. Natl. Acad. Sci. U.S.A., *76*, 4350 – 4354.

Tschopp, J. F., Sverlow, G., Kosson, R., Craig, W., and Grinna, L. (1987) High level secretion of glycosylated invertase in the methylotropic yeast *Pichia pastoris*. Bio/Tech., *5,* 1305 - 1308.

Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U., and Machleidt, W. (1983) Protein inhibitors of cysteine proteinases III. Amino acid sequence of cystatin from chicken egg white. Hoppe-Seyler's Z. Physiol., Chem., *364*, 1487 – 1496.

Turk, V., Brzin, J., Lenarčič, B., and Sali, A (1986) Human stefins and cystatins: their properties and structural relationships. In: Cysteine proteinases and their inhibitors (Turk, V. ed), Walter de Gruyter, Berlin, pp. 429 – 441.

Turk, V., and Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. FEBS Lett., *285*, 213 - 219.

Turk, V., and Bode, W. (1992) Lysosomal cysteine proteinases and their inhibitors cystatins. In: Innovations in proteases and their inhibitors. (Avilés, F. X., ed), Walter de Gruyter, New York, pp. 161 – 178.

Turk, B., Križaj, I., Kral, B., Dolenč, I., Popovič, T., Bieth, J. G., and Turk, V. (1993) Bovine stefin C, a new member of the stefin family. J. Biol. Chem., *268*, 7323 - 7329.

Turk, B., Ritonja, A., Björk, I., Stoka, V., Dolenč, I., and Turk, V. (1995) Identification of bovine stefin A, a novel protein inhibitor of cysteine proteinases. FEBS Lett., *360*, 101 - 105.

Turk, B., Stoka, V., Turk, V., Johansson, G., Cazzulo, J. J., and Björk, I. (1996) High-molecular-weight kininogen binds two molecules of cysteine proteinases with different rate constants. FEBS Lett., *391*, 109 – 112.

Urwin, P. E. , Atkinson, H. J., Waller, D. A., and McPherson, J. (1995a) Engineered oryzacystatin-I expressed in transgenic hairy roots confers resistance to *Globodera pallida.* Plant J., *8*, 121 - 131.

Urwin, P. E., Atkinson, H. J., and McPherson, J. (1995b) Involvement of the NH_2-terminal region of oryzacystatin-I in cysteine proteinase inhibition. Prot. Eng., *8*, 1303 - 1307.

Valevski, K., Fernandes, S., Campos, F. A. P., Do Val, R. R., and Xavier-Fiho, J. (1991) The expression of papain inhibitors during the development of cowpea seeds. Plant. Sci, *74*, 179 – 184.

Varughese, K. I., Ahmed, F. R., Carey, P. R., Hasnain, S., Huber, C. P., and Storer, A. C. (1989) Crystal structure of papain-E-64 complex. Biochem., *28*, 1330 - 1332.

Völker, J. (1993) The impact of global and local composition on the stability of triple helical DNA. PhD Thesis, University of Cape Town, Cape Town, South Africa.

Wada, K., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T., and Ikemura, T. (1991) Codon usage tables. Nucl. Acids Res., *19, Suppl.*, 1981 - 1986

Waldron, C., Wegrich, L. M., Owens Merlo, P. A., and Walsh, T. A. (1993) Characterisation of a genomic sequence coding for potato multicystatin, an eight domain cysteine proteinase inhibitor. Plant Mol. Biol., *23*, 801 - 812.

Walsh, T. A., and Strickland, J. A. (1993) Proteolysis of the 85-kilodalton crystalline cysteine proteinase inhibitor from potato release functional cystatin domains. Plant Physiol., *103*, 1227 - 1234.

Watanabe, H., Abe, K., Emori, Y., Hosoyama, H., and Arai, S. (1991) Molecular cloning and gibberellin-induced expression of multiple cysteine proteinases of rice seeds (Oryzains). J. Biol. Chem., *266,* 16897-16902.

Weiman, K. F., and Nielsen, S. S. (1988) Isolation and partial characterisation of a major gut proteinase from larval *Acanthoscelides obtectus say* (Coleoptera, Bruchidae). Comp. Biochem. Physiol., *89B*, 419 - 426.

White, C. E., Hunter, M. J., Meininger, D. P., White, L. R., and Komives, E. A. (1995) Large-scale expression, purification and characterisation of small fragments of thrombomodulin: the roles of the sixth domain and of methionine 388. Prot. Eng., *8*, 95 - 103.

Wolfson, J. L., and Murdock, L. L. (1987) Suppression of larval Colorado potato beetle growth and development by digestive proteinase inhibitors. Entomol. Exp. Appl., *44,* 235-240.

Wolfson, J. L., and Murdock, L. L. (1989) Diversity in digestive proteinase activity among insects. J. Chem. Ecol., *16,* 1089-1103.

Yamamoto, A., Tomoo, K., Doi, M., Ohishi, H., Inoue, M., Ishida, T., Yamamoto, D., Tsuboi, S., Okamoto, H., and Okada, Y. (1992) Crystal structure of papain-succinyl-Gln-Val-Val-Ala-Ala-p-nitroanilide complex at 1.7-Å resolution: non-covalent binding mode of a common sequence of endogenous thiol protease inhibitors. Biochem., *31*, 11305 - 11309.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning and host strains: Nucleotide sequences of M13mp18 and pUC19 vectors. Gene, *33*, 103 – 119.

Zaret, K. S., and Sherman, F. (1982) DNA sequence required for efficient transcription termination in yeast. Cell, *28*, 563 – 573.

Zimacheva, A. V., Ievleva, E. V., and Mosolov, V.V.A (1988) A cysteine proteinase inhibitor from pumpkin seeds. Biokhimiya, *109*, 50742.

Zucker, S., Buttle, D. J., Nicklin, J. H., and Barrett, A. J. (1985) The proteolytic activities of chymopapain, papain and papaya proteinase III. Biochim. Biophys. Acta, 828, 196 – 204.

Appendix A

Expression of the Hybrid Oryzacystatin in the Yeast *Pichia pastoris*

A.1 Introduction

The production of the HO gene as a fusion product in bacteria was successful in that a large amount of biologically active fusion protein was produced. Cleavage of the fusion with factor Xa, however, gave poor yields of liberated inhibitor, so it was decided to explore other expression systems. To this end, the hybrid inhibitor was expressed in a yeast system. The yeast *Saccharomyces cerevisiae* has often been used for the expression of foreign genes for the advantages that yeast offers over other expression systems i.e:

- The organisms can be grown rapidly and to high cell density (Bucknolz & Gleeson, 1991).
- The production of proteins in yeast often leads to higher levels of heterologous protein expression than in bacterial systems.
- Yeasts are eukaryotes and hence their intracellular environment is often more suitable for the correct folding of eukaryotic proteins.

The low yields of expressed proteins and difficulties in secreting some proteins in *Saccharomyces* have, however, led to the use of alternative yeast expression systems. One of these is *Pichia pastoris* developed by Invitrogen, San Diego. The *Pichia pastoris* expression system was used in this study. The following sections give a brief review of this yeast expression system and discuss the expression results.

A.1.1 *Pichia pastoris* as an Expression Host

Pichia pastoris is a methylotrophic yeast that is capable of using methanol as its sole carbon source. It has been used successfully by a number of groups to produce recombinant proteins (Cregg *et al.*, 1987; Clare *et al.*, 1991; Scorer *et al.*, 1993) and often produces higher yields of expressed protein than many other yeast expression systems.

The first step in the metabolism of methanol by this yeast is the oxidation of methanol to formaldehyde using molecular oxygen and the enzyme alcohol oxidase (*AOX1*). This reaction generates formaldehyde and hydrogen peroxide. To avoid toxicity of the latter, methanol metabolism takes place in the peroxisome which sequesters these toxic by-products from the rest of the cell (Ellis *et al.*, 1985). The *AOX1* enzyme has a poor affinity for oxygen and so the yeast compensates by generating large amounts of this enzyme. It is the *AOX1* promoter which has been used in the *Pichia pastoris* system to drive heterologous protein expression (Cregg *et al.*, 1989).

There are two genes in *P. pastoris* that code for alcohol oxidase - *AOX1* and *AOX2* (Cregg *et al.*, 1989). The *AOX1* gene is responsible for the vast majority of alcohol oxidase activity in the cell, and in methanol grown cells the *AOX1* gene message represents approximately 5% of the total poly A^+ RNA levels (Cregg *et al.*, 1993). The *AOX1* promoter is under the control of both a general carbon catabolic repression/ derepression mechanism and a C-specific induction mechanism (Johnston, 1987). In this system, growth on glucose represses transcription, even in the presence of the inducer, methanol. Hence, before induction, this yeast is grown on glycerol as the carbon source. The second functional alcohol oxidase gene, *AOX2*, encodes a protein that is 97% identical to and has almost the same specific activity as *AOX1* (Cregg *et al.*, 1993). Clones lacking the *AOX1* gene grow much slower on methanol and this allows for the isolation of Mut^S strains (Cregg *et al.*, 1989; Koutz *et al.*, 1989). The reason these transformants can grow on methanol is because of the second alcohol oxidase gene, *AOX2*, which provides a less efficient source of alcohol oxidase (Cregg *et al.*, 1987).

A.1.2 Heterologous Protein Secretion In Pichia

The goal in this case, and with many other expression systems in *Pichia pastoris,* was secretion of the foreign protein into the media. This has the advantage in that *Pichia* secretes very low levels of its own native proteins and so it provides a first step in the purification of the protein of interest.

The pPIC9 vector, which is used for secretion of foreign proteins in this yeast, was used in this study (Figure A.1). This expression vector contains the *Saccharomyces cerevisiae* pre-pro-α–mating factor leader sequence which is responsible for directing the secretion of several proteins from the cell (Clare *et al.*, 1991). It is a 13-residue peptide that is secreted by cells of the α-mating type and functions by acting on cells of the opposite mating type to promote efficient conjugation. This leads to the formation of diploid cells (Brake *et al.*, 1984).

For the processing of this pre-pro-α-factor signal peptide a number of different proteolytic activities are required (Clare *et al.*, 1991). For example, the glycosylated pro-α-factor is cleaved firstly by an endoproteinase that cleaves on the carboxyl side of the Lys-Arg sequence (Figure A.2). This endoproteinase is encoded by the *KEX2* gene and is used in conjunction with the α-factor leader sequence to direct the secretion of small peptides (Clare *et al.*, 1991). An investigation into the properties of this endoproteinase (Clare *et al.*, 1991) has shown it to be a membrane-bound, calcium-dependent serine protease which is homologous to subtilisin and other related proteases. Since the inhibitor being cloned was a cysteine proteinase inhibitor and not a serine proteinase inhibitor, cloning of this HO protein into a vector containing this protease was not expected to interfere with the protease action.

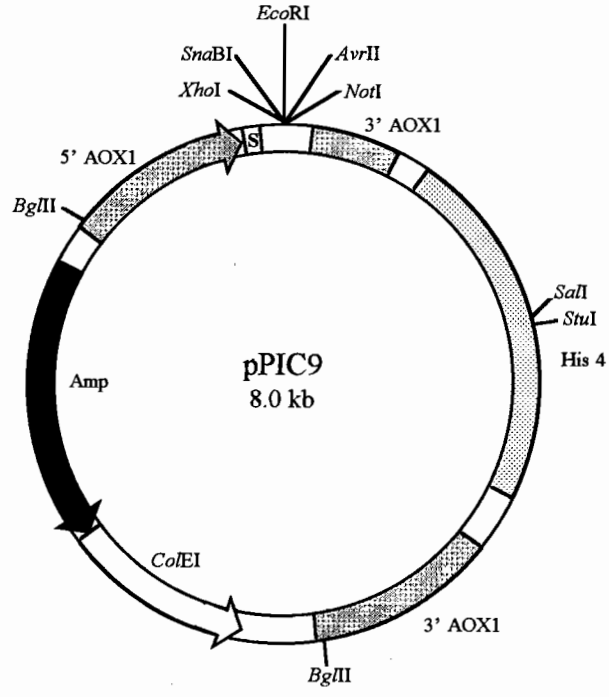


Figure A.1: *Map of the pPIC9 vector showing the muliple cloning site and positions of the AOX1 and His 4 genes. S, α–factor signal sequence.*

```
mRNA AOX1
5' end (773)                                                    αF start
773    ....+.....+.....917  AAAAACAACTAATTATTCGAAGGATCCAAACG ATG AGA TTT CCT TCA ATT TTT ACT
                                                               Met ARG Phe Pro Ser Ile Phe Thr

973    GCA GTT TTA TTC GCA GCA TCC TCC GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT
       Ala Val Leu Phe Ala Ala Ser Ser Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp

1033   GAA ACG GCA CAA ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG GAT TTC
       Glu Thr Ala Gln Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe

1093   GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG TTT ATA AAT ACT
       Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu Phe Ile Asn Thr

                                                                  XhoI          ↓
1153   ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA TCT CTC GAG AAA AGA GAG GCT GAA
       Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val Ser Leu Glu Lys Arg Glu Ala Glu

          *SnaBI     EcoRI   AvrII      NotI            stop                    mRNA AOX1
1213   GCT TAC GTA GAA TTC CCT AGG GCG GCC GCG AAT TAA                            3'end (1418)
       Ala Tyr Val Glu Phe Pro Arg Ala Ala Ala Asn ...TTCGCCTTAGACATGACTGTT.....TTGTCA
```

Figure A.2: *The pPIC9 promoter and multiple cloning region. The KEX2 cleavage site (↓) and the pre-pro α–factor signal sequence (αF) are shown. * represents the STE13 cleavage site.*

Preliminary cleavage of the signal sequence by the *KEX2* gene product occurs between arginine and glutamine in the sequence Glu-Lys-Arg* Glu-Ala-Glu-Ala where * is the site of cleavage. The STE13 gene product then further cleaves the Glu-Ala repeats. This *STE13* gene encodes a membrane-bound, heat-stable di-peptidylaminopeptidase which is responsible for removal of Glu-Ala sequences after the signal cleavage site (Romanos *et al.*, 1992). The α-factor signal sequence, the cleavage sites and the multiple cloning site of pPIC9 are shown in Figure A.2.

Secreted proteins are often glycosylated. However, an advantage of *Pichia* over *Saccharomyces,* is that *Pichia* does not hyperglycosylate. Thus, although both *S. cerevisiae* and *P. pastoris* have mainly N-linked glycosylation of the high-mannose type, the length of the oligosaccharide chains added post-translationally to proteins in *Pichia* is much shorter than those added in *S. cerevisiae* (i.e. 8-14 in *Pichia* as opposed to 50-150 in *S. cerevisiae;* Tschopp *et al.*, 1987). Furthermore, *S. cerevisiae* core oligosaccharides have terminal 1,3 glycan linkages whereas *P. pastoris* does not. It is believed that the 1,3 glycan linkages in glycosylated proteins produced from *S. cerevisiae* are primarily responsible for the hyper-antigenicity that makes them particularly unsuitable for therapeutic use (Cregg *et al.*, 1993). This is predicted to be less of a problem for glycoproteins generated in *P. pastoris* since these resemble the glycoprotein structure of higher eukaryotes (Cregg *et al.*, 1993).

The recognition site for glycosylation is Asn-X-Ser/Thr (Brake, 1990). This sequence is not present in the HO sequence and hence glycosylation was not considered to be a problem in secretion of this protein.

A.1.4 Recombination and Integration in *Pichia*

Pichia pastoris has no stable episomal vectors and so those used for transformation into the yeast are integration vectors. Integration into the genome has the advantage in that it confers stability through meiosis (Romanos *et al.*, 1995). There are three ways in which the vector expression cassette can be inserted into the yeast genome, depending on the restriction enzymes used to linearise the vector. This is illustrated in Figure A.3 and is described in more detail below.

(A) The vector is cut with *Bgl*II

Digestion with *Bgl*II produces a fragment with ends homologous to the 5' and 3' ends of *AOX1.* On transformation, these are then targeted to transplace into the *AOX1* locus (Figure A.3 (A)). The chromosomal *AOX1* is thus disrupted by this event and Mut^S transformants are produced. Multiple insertions can also occur.

(B) The vector is cut with *Sac*I

A second method of integration is to linearise the vector with *Sac*I. This targets integration by a single cross-over event into the *AOX1* locus (Figure A.3 (B)). The *AOX1* gene remains intact in this case, and the resultant clones have the Mut$^+$ phenotype. Multiple insertions by repeated single cross-overs can also occur.

(C) The vector is cut with *Stu*I

Linearising the vector with *Stu*I targets integration into the chromosomal *his4* locus by a single cross-over mechanism (Figure A.3 (C)). These transformants are again Mut$^+$ and multiple integrations can occur.

Thus in summary, the gene of interest is cloned into the multiple cloning site of the vector, this vector is then integrated, in one of the above ways, into the yeast genome.

A.2 Materials and Methods

A.2.1 Yeast Strains

P. pastoris, GS115 (*his4*; Cregg *et al.*, 1995), SMD1168 (*his4, pep4;* Invitrogen, San Diego, USA) were used as host strains for the expression of the HO in yeast. *E. coli* strain DH5α was used for the plasmid pPIC9 (Invitrogen, San Diego, USA) and pPIC9-HO constructions and propagations.

A.2.2 Cloning of the Hybrid Oryzacystatin Gene in *Pichia pastoris*

All molecular biology methods were carried out according to the methods of Sambrook *et al.* (1989). Restriction enzymes were from Boehringer Mannheim. The insert was removed from the pUC18 vector by cutting with the restriction enzymes *Pvu*II and *Hind*III and gel purified as discussed previously (section 3.3.5). For cloning into the *Sna*BI site of the pPIC9 vector, the *Hind*III site was blunt-ended using Klenow (Boehringer Mannheim). This 282-bp fragment was then cut with *Sna*BI and cloned behind the *AOX1* promoter in pPIC9 (Figures A1 and A.2). Transformation into the *E. coli* host DH5α was carried out using the method of Chung & Miller (1988).

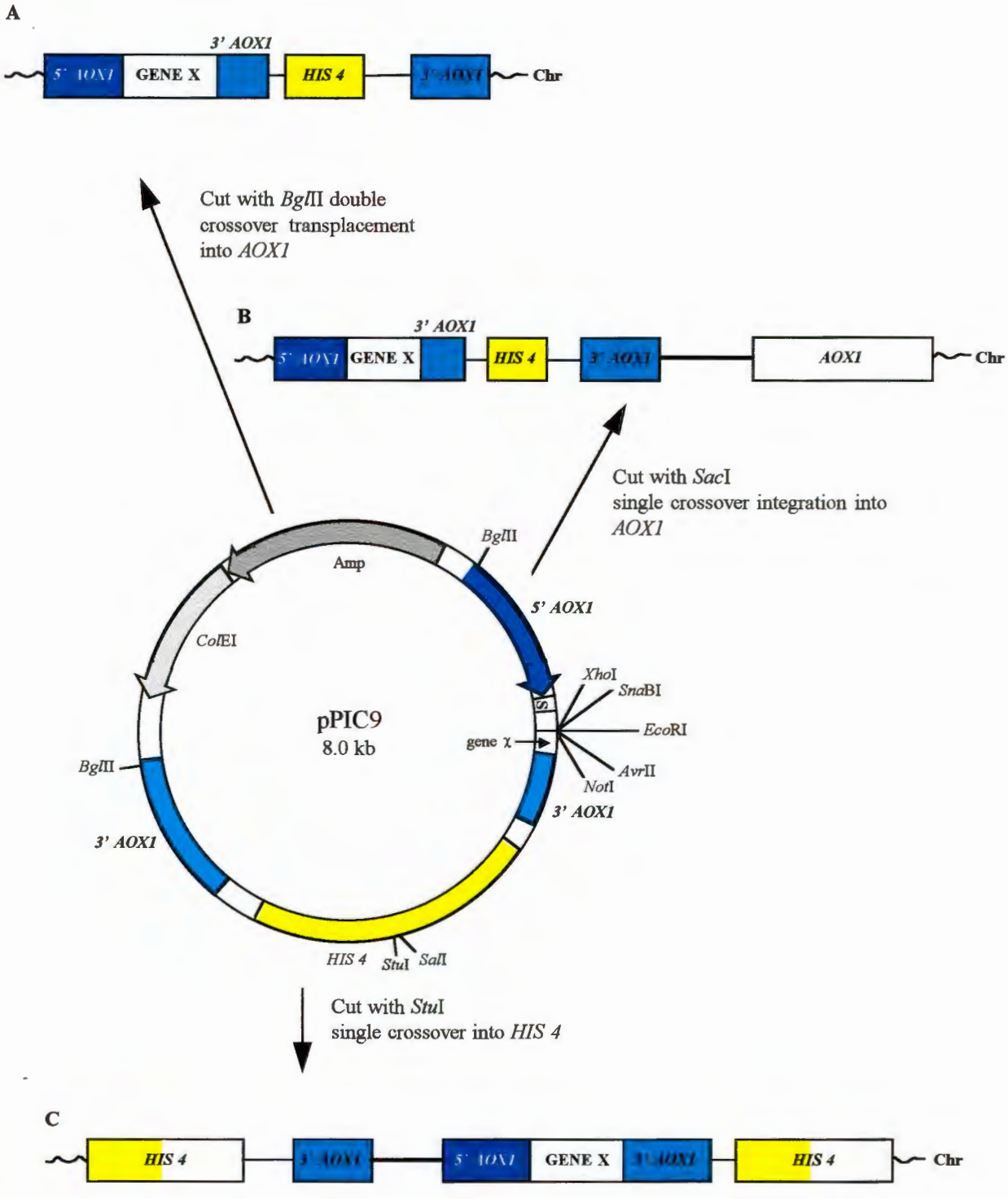


Figure A.3: *The 3 models of chromosomal integration of pPIC9 into the P. pastoris genome. (A) transplacement of the expression cassette into AOX1; (B) transplacement of the whole vector into AOX1; (C) integration of the whole vector into His4 (from Clare et al., 1991).*

A.2.3 Preparation of DNA for Transformation into the Yeast Strain GS115

The recombinant pPIC9 vector (pPIC9-HO) was linearised with *Bgl*II according to the manufacturer's instructions (Boehringer Mannheim), run on a 1% agarose gel and the uppermost-band excised and purified using the GENECLEAN kit according to the manufacturer's instructions (Bio 101). The concentration and purity of the DNA was determined by scanning the DNA at a wavelength of 220 nm to 310 nm on a Beckman DU-64 spectrophotometer.

A.2.4 LiCl Transformation Procedure and Electroporation

The transformation of the *Bgl*II cut fragment into *P. pastoris* was as outlined in the *Pichia* manual (Invitrogen, San Diego), with slight modifications. The yeast strains were grown until the A_{600} reached 1.2. Cells were centrifuged at 5000 rpm for 10 minutes, the cell pellet washed in 100 ml of distilled water, dissolved in 100 ml of LiCl transformation buffer (0.1 M LiCl; 1 mM EDTA; 0.01 M Tris-Cl, pH 7.4) and incubated at 30°C for 45 minutes. 1.25 ml of a 2 M DTT stock solution was then added and the cells were incubated at 30°C for a further 15 minutes before pelleting at 5000 rpm for 10 minutes. Cells were washed several times in 30 ml of ice cold 1 M sorbitol and the final pellet was resuspended in 0.5 ml of 1 M sorbitol.

For electroporation, 10 µg of DNA was added to 40 µl of yeast cells and electroporated using a Bio-Rad GenePulser at 1.5 kV, 25 µF and 400 ohm. Immediately after delivery of the high voltage pulses, 1 ml of ice cold 1 M sorbitol was added and 200 µl aliquots were then spread onto RDB plates (Appendix B). Plates were incubated at 30°C for three to five days, or until colonies appeared.

A.2.5 Screening for Recombinant Clones

Transformation of the GS115 (or SMD1168) strain using the *Bgl*II fragment favours recombination at the *AOX1* locus (Cregg *et al.*, 1993). The His^+ transformants from the RDB plates (used to identify His^+ transformants) were therefore patched onto minimal methanol (MM) and minimal dextrose (MD) plates (Appendix B). The Mut phenotype of the recombinant *Pichia* was then identified by visual screening for the absence of normal growth on the methanol-containing plates. This was then further verified by inoculating the clones into MD and MM broth (Appendix B).

A.2.6 Isolation of Chromosomal DNA

Yeast cells were grown to stationary phase in 10 ml YPD (1% yeast extract, 2% peptone, 2% glucose) medium at 30°C. Cells were then pelleted at 3000 rpm for 5 minutes and the pellet washed in distilled water. Cells were transferred into a 2 ml eppendorf tube and resuspended in 200 µl disrupt buffer (25% Trion-X-100; 1%

SDS; 100 mM NaCl; 1 mM EDTA; 10 mM Tris-Cl, pH 8.0) and 200 µl of a phenol/chloroform/-isoamylalcohol mix at a ratio of 25:24:1. To this 0.3 g of acid-washed glass beads (Sigma) were added. This was vortexed until 90% of the cells were broken (the extent of the cell breakage was determined using a Zeis Standard WL Microscope). 200 µl of TE (10 mM Tris; 1 mM EDTA, pH 8.0) was then added and the cells centrifuged in a micro-centrifuge at 14 000 rpm for 5 minutes. The aqueous layer was then transferred to a fresh 2 ml eppendorf tube and chromosomal DNA was precipitated with 100% ethanol. Following centrifugation, the pellet was resuspended in TE with the addition of 30 µl of a 1 mg/ml RNAse A solution (prepared according to Sambrook *et al.*, 1989). This was incubated at 37°C for 10 minutes. 10 µl of 4M ammonium acetate and 1 ml 100% ethanol were added and the DNA again precipitated by centrifugation. The DNA pellet was then washed in 70% ethanol, and resuspended in 100 µl TE. The concentration and purity of the DNA was tested by scanning the DNA as described in section A.2.3 above.

A.2.7 PCR Analysis of Pichia Integrants

Amplification of this chromosomal DNA (isolated as described in section A.2.6) was carried out using the 5' *AOX1* and 3' *AOX1* sequencing primers (Invitrogen, San Diego). A typical PCR reaction was as follows:

10 × PCR buffer (Promega)	5 µl
Chromosomal DNA (1 µg)	5 µl
100 mM dNTPs (25 mM each; Boehringer Mannheim)	1 µl
20 pmoles 5' *AOX1* primer	2 µl
20 pmoles 3' *AOX1* primer	2 µl
Sterile water	to 50 µl
Taq Polymerase (5U/µl) (Promega)	0.25 µl

An average cycle consisted of a "hot start" of 2 minutes at 94°C. This was followed by a denaturation step at 93°C for 20 seconds, an annealing step of 50°C for 30 seconds and an extension step of 72°C for 1 minute. After 30 cycles, a final extension step of 5 minutes at 72°C was used. Following PCR, the aqueous phase was removed and electrophoresed on a 0.8% agarose gel stained with ethidium bromide for visualisation of the PCR bands.

A.2.8 Large-Scale Protein Expression: Media Composition and Culture Conditions

Yeast cells containing the cloned gene, or the pPIC9 vector only, were pre-cultivated in 10 ml YPD (1% yeast extract, 2% peptone, 2% glucose) medium at 30°C for 24 hours. This pre-culture was then inoculated into a 5 litre Erlenmeyer flasks containing 500 ml minimal medium (13.4 g/l yeast nitrogen base without amino acids, 400 µg/l biotin, 1% (v/v) glycerol) at 30°C with vigorous shaking. After the cells (originally inoculated at

about 10^7 cells/litre) reached saturation (48-72 hours), they were harvested by centrifugation and resuspended in 500 ml of minimal medium containing 0.5% (v/v) methanol in place of the glycerol. The incubation was continued for a further 7 days, during which time additional methanol (5 ml/litre) was added every 24 hours. Cells were harvested by centrifugation, and the supernatant kept. The supernatant was then concentrated by ultracentrifugation on an amicon P6 membrane (Amicon Inc., USA). Protein concentrations in yeast media were determined using the Bio-Rad assay (Bradford, 1976; Appendix C).

A.2.9 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed according to Laemmli (1970) as described (Appendix C).

A.3 Results and Discussion

A.3.1 Subcloning into the Yeast Expression Vector pPIC9

The pPIC9 vector, described in section A.1.3, was used as the cloning vector for the HO insert. The synthesised HO insert was removed from the pUC18 vector rather than from pMAL because restriction sites present in pUC18 enabled the coding sequences for the HO to be removed without including vector sequences, and cutting with *Pvu*II generated a blunt-ended site at the beginning of the gene which could then be cloned into the blunt-ended *Sna*BI site of the yeast vector.

The insert was therefore removed from pUC18 using *Pvu*II and *Hind*III. The *Hind*III site on the insert was then blunt-ended by filling in with dideoxy-nucleotide-tri-phosphates using the enzyme Klenow (Boehringer Mannheim). The insert was cloned into the *Sna*BI-linearised pPIC9 and transformed into *E. coli* strain DH5α using the method of Chung and Miller (1988).

Positive clones were identified by PCR screening using the 5' and 3' *AOX1* primers (Invitrogen, San Diego). PCR products were then electrophoresed on 1% agarose gels and clones carrying the HO insert were identified as an increase in the banding size from 494-bp (control, pPIC9 only) to 779-bp (pPIC9 + HO).

Since both the vector and the insert were blunt-ended, it is possible for the gene to be inserted into the vector in either orientation. To determine which clones had the correct orientation, these 'positive' clones were 'mini-prepped' (Sambrook *et al,*. 1989) and the DNA mapped by digesting with *Xho*I.

Seven out of the ten positive clones selected contained the correct orientation. Two of these were 'maxi-prepped' and sequenced to confirm that:

- the reading frame (for secretion) was correct

- the ATG was in the proper context for eukaryotic translation initiation
- the HO sequence was correct.

However, because the distance between the 5' and 3' AOX1 primers was so large (741-bp), it was never possible to sequence the 5'-HO sequence. Hence, a second primer was designed using the 'Primer' computer programme and the following primer sequence was generated. This primed close to the multiple cloning site of pPIC9:

5' AAATACTACTATTGCCAGC 3'

Once the sequence of the HO and the correct reading frame for the secretion signal had been confirmed, the vector was linearised with *Bgl*II for transformation. This generated a fragment that was able to displace the chromosomal *AOX1* gene by double-crossover recombination (transplacement; see Figure A.3) and, as opposed to cutting at other restriction sites (*Stu*I and *Sac*I), generates the highest frequency of multi-copy integrants (Romanos, 1995).

A.3.2 Transformation of pPIC9-HO into GS115 and Screening of Recombinants

In order to target the *Bgl*II linearised fragment to the *AOX1* gene locus in the chromosome, the pPIC9-HO DNA was mixed with the LiCl yeast transformation mix. This was then electroporated and the transformants selected for histidine prototrophy (His^+) on RDB plates (Appendix B). This is possible because the GS115 yeast strain, a histidine rich autotroph, has been made defective in histidinol dehydrogenase. The pPIC9 vector, however, contains the *his4* gene and thus serves as a selectable marker for transformation.

As shown in Figure A.3, correct insertion of the fragment should result in the deletion of the entire *AOX1* gene. Such alcohol oxidase disruption mutants (Mut^S) grow slowly on methanol. The reason these transformants can grow on methanol at all is due to the presence of the second alcohol oxidase gene, *AOX2*, which provides a less efficient source of alcohol oxidase (Cregg *et al.*, 1987).

144 His^+ transformants were plated onto MM and MD for determination of their Mut phenotype. However this screening was not conclusive as the yeast tended to grow slowly. It was thus difficult to see the difference between slow growth and the even slower growth expected for the Mut^S transformants. The phenotype of clones suspected to be Mut^S were thus confirmed by inoculating into MD and MM broth (see Appendix B) in which the difference between Mut^S and Mut^+ clones was much more obvious (Figure A.4).

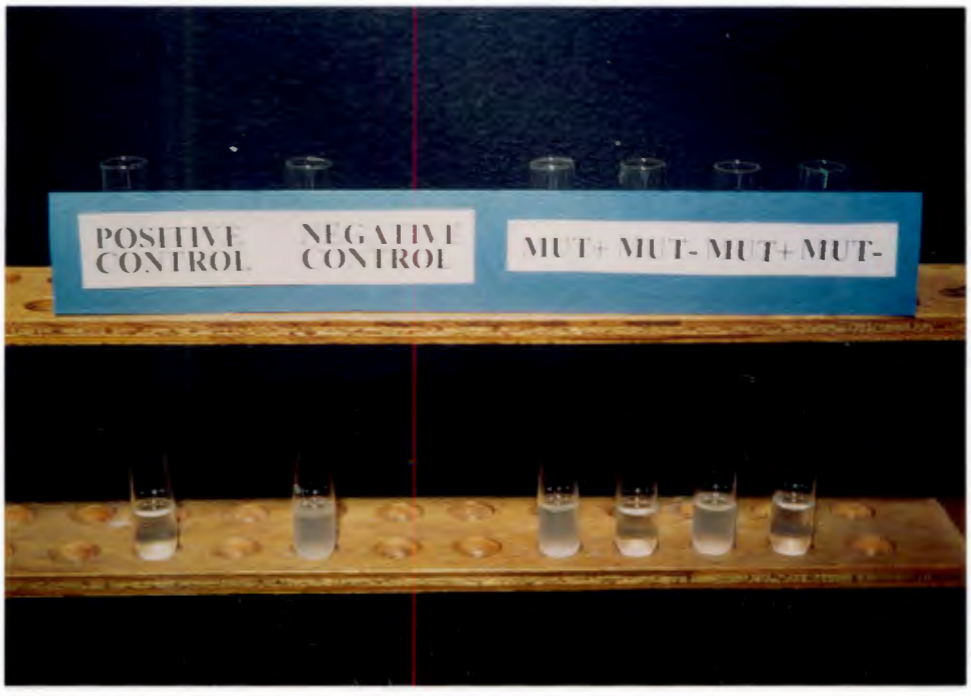


Figure A.4: *Screening for the Mut^S phenotype using MM and MD broth. The positive control (for the Mut^S phenotype) was a strain in which the AOX1 gene had been replaced and expressed the albuman gene while the negative control (for the Mut^S phenotype) was a strain which had the AOX1 gene and which expressed the β-galactosidase gene (both supplied by Invitrogen, San Diego).*

Of the His$^+$ transformants only 8% were Mut^S. This indicates that in the majority of cases the transforming DNA integrates without *AOX1* disruption. Similar results have been found by other researchers (Clare *et al.*, 1991; Sreekrishna *et al.*, 1989).

To confirm the phenotypes at the genetic level, PCR was then carried out using the 5' and 3' *AOX1* primers. To do this, genomic DNA was isolated from the *Pichia* transformants (both Mut$^+$ and Mut^S) as well as from appropriate controls. Ten microlitre samples from each PCR reaction were then run on 0.8% agarose gels (Figure A.5). The differences between Mut^S and Mut$^+$ phenotypes on the genetic level can easily be identified. Lane 6 (pPIC9 only) shows the presence of a single 494-bp band which is the size of the pre-pro α–factor and the multiple cloning region (see Figure A.2). In lane 5, the GS115 shows the presence of a 2.2 kb band which is the size of the *AOX1* gene. From the *Pichia* transformants in lanes 1-4, it can be seen that lanes 1,3 and 4 contain the insert as there is an 779-bp (size of pPIC9 region + HO) band present. Of these, only lane 1 is Mut^S. Lane 2, however, although a His$^+$ transformant, does not contain the insert.

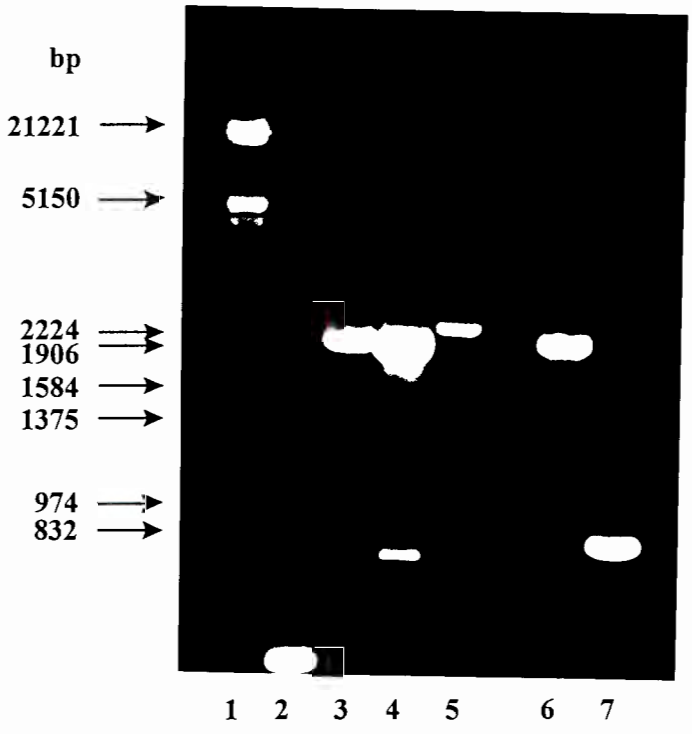


Figure A.5: *0.8% agarose gel showing the results of PCR amplification of chromosomal DNA from transformants. Lanes 4 - 7, Pichia transformants; 3, GS115 PCR product; 2, pPIC9 only; 1, molecular weight marker, lambda DNA digested with HindIII and EcoRI.*

A.3.3 Screening for High Levels of Expressed Protein.

Cregg *et al.* (1993) have shown that independently isolated *P. pastoris* strains transformed with the same expression vector routinely display a range of product levels. This clonal variation has been observed even with clones containing the same number of expression cassettes (Cregg *et al.*, 1993). Thus, in order to determine which clones give the best product expression levels, it is necessary to screen a significant number of transformants.

A small-scale expression study was therefore carried out. For this test, 10 ml cultures of a number of both Mut^S and Mut^+ clones, containing the insert, were grown to an OD of 17 (8.5×10^8 cells/ml). The optical density of the medium at 600 nm (OD_{600}) was used to estimate the cell density, assuming that 1 OD_{600} is equivalent to 5×10^7 cells/ml. Once this OD was reached, the cells were placed in 10 ml minimal media containing methanol in 100 ml volume Erlenmeyer flasks for a further 2 days. Extracts of the cell medium, before and after induction, were then analysed by SDS-PAGE for the presence of the HO.

In all the clones tested the levels of HO in the culture medium were found to be < 0.001 μg/ml in the YNB minimal medium. There was also no difference between the Mut^+ and the Mut^S strains (results not shown). Similar results were obtained with all other clones tested. The lack of expression could be due to a number of reasons, for example:

- The flasks used in this expression may have been too small for adequate aeration. This is a critical requirement for *Pichia* expression, as dissolved molecular oxygen is necessary for growth in methanol (J. M. Cregg, personal communication).
- Induction in methanol was only carried out over 48 hours. This may not have been sufficient for large scale production of the inhibitor.
- Richer medium containing yeast extract and peptone may be better for expression. Clare *et al.* (1991) showed that yeast induced in rich media produced higher levels of mouse epidermal growth factor than yeast grown in minimal media.

From the above results, it was clear that the culture conditions for growth of the yeast needed to be fully optimised.

A.3.4 Optimisation of the Culture Medium for Expression

The first step in the optimisation process involved a study of different media. Often different media may be optimal depending on the particular protein being expressed. For example, Clare *et al.* (1991) found yields of mouse epidermal factor to be higher in richer media (BMMY) than minimal media (Appendix B), however, they found that proteolytic degradation of their expressed protein was less of a problem in the minimal media, especially if supplemented with casamino acids. Hence, it was decided to test a range of transformants in BMGY media (Appendix B). This rich medium contains yeast extract and peptone. Transformants were grown as 100 ml volumes in 1 litre Erlenmeyer flasks to provide good aeration. Cultures were grown in this BMGY medium containing glycerol at 30°C until a large cell density of about 8.5×10^8 cells/ml (17 OD units) was obtained. Cultures were then resuspended in fresh volumes of the same medium, but containing 0.5% methanol instead of glycerol as the sole carbon source. Aliquots were withdrawn on each day to determine the extent of HO production by SDS-PAGE. However, an analysis of the results showed similar low levels of expression to that observed in minimal media.

Romanos (1995) observed that heterologous secretion is more demanding than intracellular expression and is not guaranteed to work. Sometimes the secretory pathway can be blocked due to malfolding of the protein (Romanos *et al.*, 1992). Thus, to determine whether such a breakdown in the secretary system had occurred in this case, the cellular fraction was analysed. Extracellular and intracellular protein preparations were extracted according to the method of La Grange (1995). Proteins retained in the periplasmic space were also isolated by sphaeroplasting (White et al., 1995) to determine whether a build up of any non-secreted protein within the cell

had occurred. However, SDS-PAGE analyses of the cellular and periplasmic fractions showed no HO protein (results not shown).

A further point to note, however, was that any HO protein expressed was observed on day 1 of expression in methanol. None or very little HO protein was detected in the medium beyond day 2. It appeared therefore, that the low yields could be a result of proteolysis. Romanos (1995) observed proteolysis to be a common problem in the expression of many proteins in yeast. Proteolysis rather than a lack of oxygen was considered the most likely problem since:

- no 'fermentation odours' were detected in the yeast culture which might indicate anaerobic metabolism
- no increase in protein yield was seen even if molecular oxygen was bubbled into the culture vessel.

There are a number of options available to minimise protein degradation. Clare *et al.* (1991) demonstrated that the yield of mouse epidermal growth factor could be improved by firstly, adding 1% casamino acids to the medium, and secondly, by buffering the YNB medium to pH 6.0 using a phosphate buffer (Clare *et al.*, 1991). Proteolysis of human immunodeficiency virus type I envelope glycoprotein, on the other hand, was minimised by maintaining a low pH (Scorer *et al.*, 1993).

In this study, proteolysis of the HO inhibitor was observed even though the cells were cultured in buffered medium (pH 6.0) containing the peptide components of yeast extract and peptone. The addition of 3% casamino acids to either medium did increase the amount of HO produced (results not shown), but a decrease in the protein concentration with time was still observed (Figure A.7).

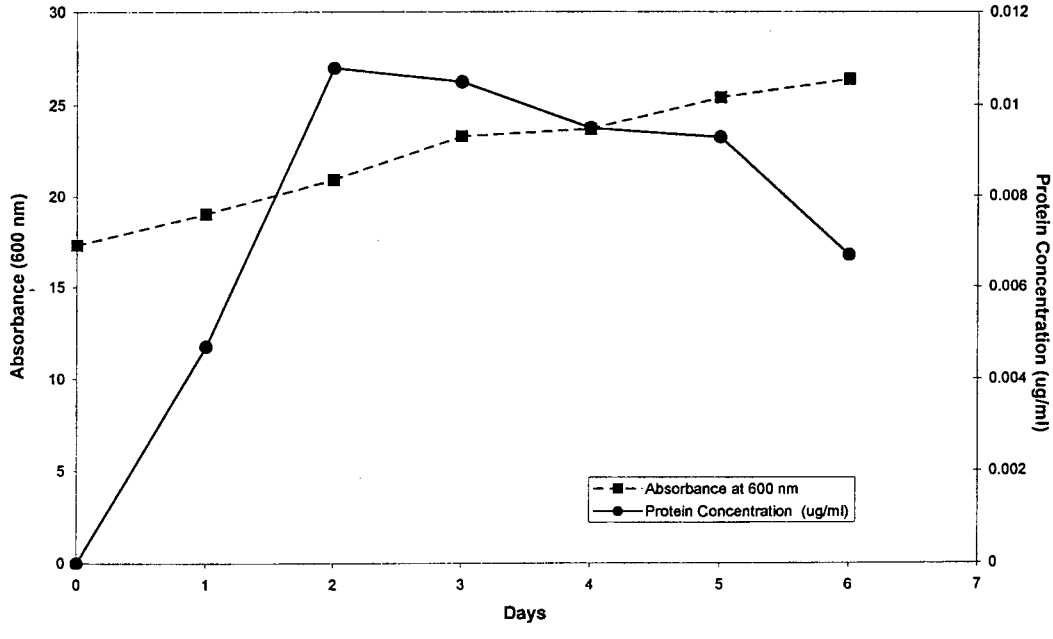


Figure A.7: *Yeast strain GS115 carrying the pPIC9-HO integrated plasmid was grown for 6 days. Aliquots of the media were removed on each day, the A_{600} was read and after pelleting the cells, the supernatant was tested for protein concentration using the Bio-Rad protein assay (Bradford, 1976).*

Whether this proteolysis was occurring during or after secretion into the yeast medium was not clear. Human albumen (Bathurst *et al.*, 1987) and *Aspergillus* glucoamylase (Innis *et al.*, 1985) which were secreted from *S. cerevisiae*, have been shown to be processed by the *KEX2* protease during secretion. *P. pastoris,* as mentioned previously (section A.1.2), also possesses an analogue to *KEX2* protease during secretion (Buckholz and Gleeson, 1991). This *KEX2* protease of *S. cerevisiae* exhibits a substrate specificity toward the carboxyl sides of Lys-Arg, Arg-Arg and Pro-Arg sequences (Mizuno *et al.*, 1989). These cleavage sites are not, however, present in the HO sequence. Thus, if the substrate specificity of the *KEX2* homologue of *P. pastoris* is identical to that of *KEX2* of *S. cerevisiae*, unknown protease(s) rather than the *KEX2* homologue must be responsible for the breakdown of the protein product.

A.3.5 Approaches to Reduce Proteolysis

Romanos *et al.* (1992) have reported that the proteolysis of some secreted products could be reduced by using the protease-deficient strain SMD1168 i.e. deficient in a gene (*PEP4*) that is homologous to protease A in *S. cerevisiae*. I therefore decided to try this option.

By using YNB medium that had been buffered to pH 6.0 and supplemented with 3% Casamino acids, secreted HO levels were increased to approximately 0.055 μg/ml. These levels are still disappointing, particularly since levels of expression of some proteins in this yeast system have been as high as 4 g/l as in the case of human serum albumin (Barr *et al.*, 1992). Discussions with others at a *Pichia* expression conference in San Diego, California, indicated that not all proteins could be expressed in *Pichia* and in such cases, other expression systems should be considered. Other possible reasons for the poor expression levels are outlined below:

1 The use of a fermenter might have been better for large scale production. Romanos *et al*, (1991) found that due to oxygen limitation, *P. pastoris* shake-flask inductions are often sub-optimal when compared to fermenter inductions.

2 A second problem could be RNA truncation. Some consensus sequences considered to be a problem in *S. cerevisiae* (which may be similar in *P. pastoris*), include (TAG)..(T rich)..TA(T)GT..(A + T rich)..TTT (Zaret and Sherman, 1984) and TTTTTATA (Henikoff and Cohen, 1984). Scorer *et al*, (1993) found termination to be a problem when attempting to express the HIV-1 envelope glycoprotein in yeast. This protein had AT-rich consensus sequences similar to the problematic TTTTTATA sequence. The HO sequence also contains many AT runs, for example: 151.....160

5' TTTATTATTT 3'

3 The HO gene codon usage had been optimised for expression in bacteria rather than yeast. According to Hadfield *et al.* (1983), a major factor influencing translational efficiency in the expression of foreign genes in yeast, is codon usage. To determine whether the codon usage for the HO gene was sufficient for yeast expression, the so called 'codon bias index' (CBI) described by Bennetzen & Hall (1982) was used.

The CBI is a measure of directional codon bias towards the use of 25 out of the 61 possible codons for the 20 amino acids. It measures the occurrence of any of the 25 optimal codons in a gene sequence. This reaches a maximum value of 1.0 in a gene with extreme codon usage. If all the possible codons for amino acids are used equally, the CBI value is 0.0. A negative value indicates the occurrence of very few optimal codons in a specific gene (Sharp & Cowe, 1991). The codon bias for the HO gene in yeast was 0.311. This is not optimal for yeast (W.H. van Zyl, personal communication), and hence the presence of the minor codons may have resulted in limited translation owing to the demand placed upon rare tRNAs, thus resulting in low concentrations of recombinant HO protein. Temporary but specific pauses during polypeptide elongation have been attributed to secondary structure formation on mRNA. These translational pauses can result in mRNA degradation and reduce the proportion of DNA synthesised in its active conformation.

Appendix B : Media, Buffers and Solutions

B.1 Bacterial Media

B.1.1 Luria-Bertani Medium (LB)

Media Component	/1000 ml
Bacto-tryptone	10 g
Yeast Extract	5 g
NaCl	5 g

Sterilise by autoclaving

For Luria agar media, add agar to 1.5% before autoclaving.

B.1.2 M9ZB Media

Media Component	/1000 ml
Bacto-tryptone	10 g
Glucose	40 g
$MgSO_4,7H_2O$	20 g
KH_2PO_4	30 g
NH_4Cl	10 g
Na_2HPO_4	60 g
NaCl	50 g

pH to 7.2

B.2 Yeast Media

B.2.1 RDB Plates

Dissolve 186 g sorbitol and 20 g agar in 700 ml of water and autoclave. When cool the medium is maintained at 45 °C and a prewarmed mix of the following solutions are added:

100 mls 20% Dextrose

100 mls 10 × YNB (13,4% yeast nitrogen base with ammonium sulphate and without amino acids)

2 ml 0.02% Biotin

10 mls 100 × amino acids*

88 ml sterile water

*100 × Amino Acids

Dissolve 500 mgs each of L-glutamic acid, L-methionine, L-lysine, L-leucine, and L-isoleucine in 100 mls of water. Filter sterilise and store at 4°C.

B.2.2 YPD

10 g yeast extract

10 g peptone

100 ml 20% Glucose stock solution

Distilled Water to 1000 ml

B.2.3 Minimal Methanol (MM) Media

1.34% YNB

4×10^{-5} % biotin

0.5% methanol

B.2.4 Minimal Dextrose (MD) Media

1.34% YNB

4×10^{-5} % biotin

1% dextrose

To make plates, 15 g/l agar is added to the above solutions

B.2.5 BMGY

1% yeast extract

2% peptone

100 mM potassium phosphate, pH 6.0.

1.34% YNB

4×10^{-5} % biotin

1% glycerol

B.2.6 BMMY

1% yeast extract

2% peptone

100 mM potassium phosphate, pH 6.0.

1.34% YNB

4×10^{-5} % biotin

0.5% methanol

B.2.7 Minimal Media for Large Scale Protein Expression

1.34% YNB

4×10^{-5} % biotin

100 mM potassium phosphate, pH 6.0.

1% glycerol initially and then 0.5% methanol is added to induce protein expression.

B.3 Enzyme Assay Solutions

B.3.1 4 × Papain Assay Buffer

350 mM KH_2PO_4	49 g
50 mM Na_2HPO_4	6.8 g
4 mM EDTA	1.5 g
pH to 6.8	

2 mM DTT (Dithiothreitol) is added just before use.

B.3.2 0.1% (m/v) Brij 35

0.5 g Brij 35
Distilled water to 500 ml

B.3.3 1 mM Substrate Solution

Dissolve 1 mg Z-Phe-Arg-AMC in 1.5 ml DMSO (Dimethyl sulphoxide). Stock solutions were aliquoted and stores at -20 C. Solutions were thawed for use and diluted to working concentrations

B.3.4 20 μM Substrate Stock Solution

Stock substrate solution (100 μl) was diluted to 5 ml with distilled water. Solutions were discarded after use.

B.3.5 10 mM E-64 Stock Solution

E-64 (3.8 mg) was disoolved in 100 μl DMSO and diluted to 1 ml with distilled water. This stock solution was diluted to a 10 μM working solution when required.

APPENDIX C : General Techniques

C.1 10% Non-denaturing Polyacrylamide gel electrophoresis

Non-denaturing polyacrylamide gels were run using a Hoefer SE600 vertical slab electrophoresis unit (Hoefer Scientific Instruments, San Fransisco, USA).

Component	Millilitres
Acrylamide[a]	33.3
10 × TBE[b]	20
10% Ammonium persulphate	0.7
TEMED	0.035
Distilled Water	to 100 mls

[a]30% Acrylamide stock solution

- 29 g acrylamide
- 1 g N, N' – methylenebisacrylamide

[b]10 × TBE (2 l) was made as follows:

- 216 g Tris-base
- 110 g boric acid
- 80 ml of 0.5 M EDTA (pH 8.0)

C.2 SDS Polyacrylamide gel electrophoresis (SDS-PAGE)

Discontinuous SDS-PAGE was done according to the method of Laemmli (1970), using a Hoefer SE600 vertical slab electrophoresis unit (Hoefer Scientific Instruments, San Fransisco, USA). Gel spacers were 1.5 mm thick.

Component	Millilitres
13% Resolving Gel	
30% Acrylamide (as above)	20
1.25 M Tris, pH 8.8	12.5
10% SDS	0.5
10% Ammonium persulfate	0.5
TEMED	0.02
Distilled Water	to 50 mls
6.7% Stacking Gel	
30% Acrylamide	1.7
0.375 M Tris, pH 6.8	1.25
10% SDS	0.1
10% Ammonium persulfate	0.1
TEMED	0.01
Distilled Water	to 10 mls

SDS-PAGE tank and loading buffers were as described by Laemmli (1970)

C.3 Western Blotting

For the Western blotting assay, proteins were transferred by the electroblotting procedure to a nitrocellulose filter (Amersham International) according to the method of Towbin *et al.*, (1979). Electrotransfer was carried out using a Hoeffer Transphor (TE 42) unit in transfer buffer (0.196 M glycine, 0.025 M Tris, and 20% methanol, pH 8.3) at constant voltage (20 V) for 6 hours at 4°C. Once blotted and dried, the nitrocellulose membrane was then blocked in phosphate buffered saline (1 × PBS; phosphate buffered saline) containing 4,5% milk powder for 1 hour at room temperature. The primary antibody was then added (at dilutions as indicated in the relevant sections). Incubation was for 1 hour at room temperature. The blot was then washed three times in wash solution (1 x PBS + 0.5% Tween 20). The primary antibody was detected with a secondary antibody conjugate (goat anti-rabbit IgG conjugated to alkaline phosphatase; Sigma), at a 1/5000 dilution. The membrane was washed again with three 10 minute washes in 1 × PBS containing 0.5% Tween 20, followed by 30 minutes in 1 × PBS. Colourimetric visualisation was performed using the chromogenic

substrates, BCIP and NBT (Sambrook *et al.*, 1989). Colour was allowed to develop, in the dark, at room temperature and stopped by washing the nitrocellulose membrane in deionised water.

C.4 Biorad Protein Assay

Protein concentrations were determined according to the instructions in the Bio-Rad Laboratories manual which is based on the method of Bradford (1976). Two standard curves were prepared in triplicate using BSA fraction V (Boehringer Mannheim) as the protein standard. Microgram dilutions of this BSA were made and the volume made up to 800 millilitres with distilled water or the buffer the protein to be determined was dissolved in. To this was added 200 µl BioRad dye reagent concentrate. The tubes were then mixed by vortexing and incubated at room temperature for 20 minutes. The OD_{595} was then read against the reagent blank (water + dye reagent). The OD_{595} versus the standard concentrations was then plotted and the regression curve (y=mx+c) was determined using the computer software package Microsoft Excel for Windows 95 (Microsoft, USA).

C.5 Inhibitory Assays of Column Fractions

Assays for inhibitory activity were carried out using the fluorimetric substrate Z-Phe-Arg-NHMec as described by Barrett and Kirschke (1981). 200 pM papain was incubated with papain assay buffer (87.5 mM KH_2PO_4, 12.5 mM Na_2HPO_4, 2 mM DTT, 4 mM EDTA, pH 6.8) and water (in the case of the control samples) or 20 µl of column samples, for 20 minutes at 30°C. For all assays percent inhibition was expressed as a percentage using the equation:

$$\% \ Inhibition = \frac{control \ fluorescence \ - sample \ fluorscence}{control \ fluorescence} \ x \ 100$$

APPENDIX D: Bacterial Plasmids

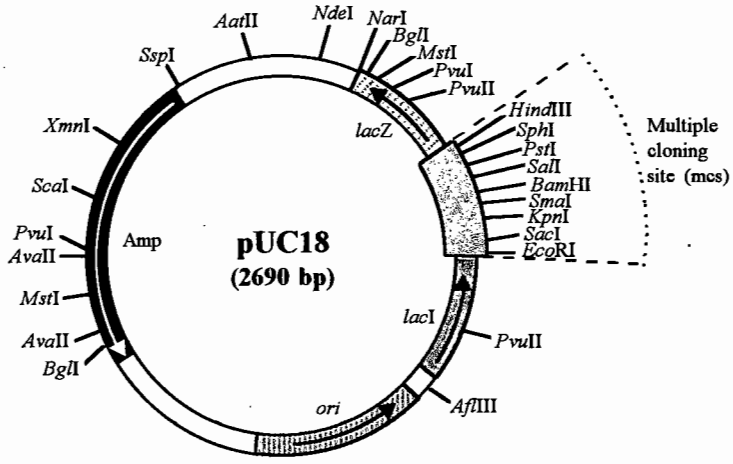


Figure D1: *Physical map of the plasmid pUC18 (Yanish-Perron et al., 1985), a high copy number cloning vector.*

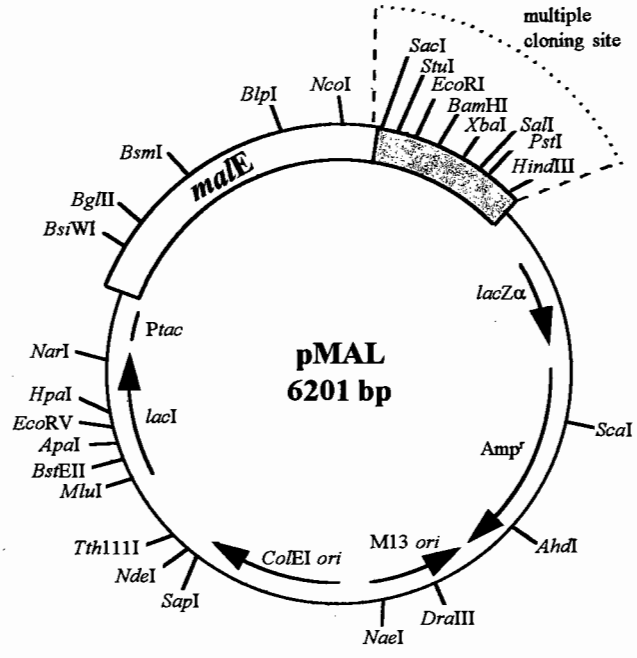


Figure D2: *Physical map of the expression vector pMAL (from New-England Biolabs Manual).*

APPENDIX E: Kinetic Data

E.1. Active-Site Titrations

E.1.1 Papain Active site Titration

Inhibitor Conc. (μM E-64)	Experiment 1		Experiment 2		Experiment 3		Average V_o	Std dev
	V_o	r^2	V_o	r^2	V_o	r^2		
0	8.02×10^{-8}	0.9999	8.13×10^{-8}	0.9992	8.18×10^{-8}	0.9999	8.11×10^{-8}	0.0008
1	6.84×10^{-8}	0.9974	7.26×10^{-8}	0.9999	6.55×10^{-8}	0.9999	6.88×10^{-8}	0.0036
2	6.24×10^{-8}	0.9998	6.34×10^{-8}	0.9992	6.08×10^{-8}	0.989	6.23×10^{-8}	0.0014
3	4.93×10^{-8}	0.9996	5.21×10^{-8}	0.9999	5.27×10^{-8}	0.9994	5.14×10^{-8}	0.002
4	4.22×10^{-8}	0.9998	4.19×10^{-8}	0.9949	4.47×10^{-8}	0.9997	4.29×10^{-8}	0.0036
5	3.15×10^{-8}	0.9986	3.54×10^{-8}	0.999	3.57×10^{-8}	0.9937	3.42×10^{-8}	0.0023
6	2.78×10^{-8}	0.999	2.58×10^{-8}	0.9974	2.35×10^{-8}	0.9989	2.57×10^{-8}	0.002
7	1.55×10^{-8}	0.9986	1.35×10^{-8}	0.9952	1.43×10^{-8}	0.9934	1.44×10^{-8}	0.001
8	6.1×10^{-9}	0.982	5.8×10^{-9}	0.9999	6.1×10^{-9}	0.9826	6×10^{-9}	0.0002

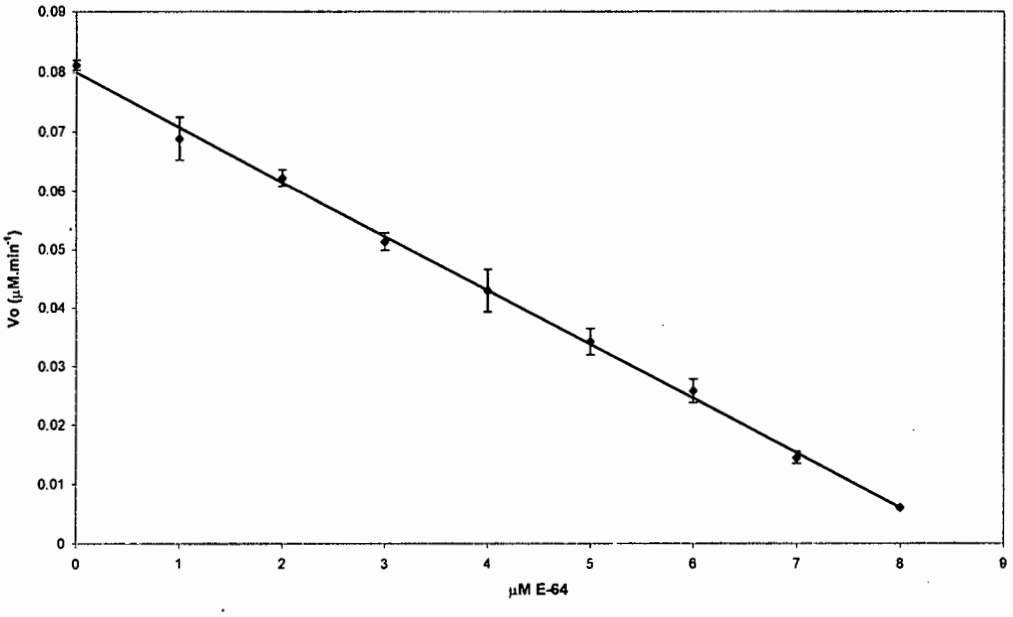


Figure E.1: *Standardisation of papain by titration with E-64. A solution of papain (10 μM) was prepared from the commercial product and titrated with E-64 as described in section 6.2.2. The extrapolated linear regression curve intercepted the abscissor at a point corresponding to a papain concentration of 8.66 μM. Error bars show 95% confidence limits.*

E.1.2 Chicken Cystatin Active-Site Titration

Inhibitor Conc. (ng Chicken)	Experiment 1		Experiment 2		Experiment 3		Average Vo	Std dev
	Vo	r^2	Vo		Vo			
0	1.14×10^{-7}	0.9997	1.11×10^{-7}	0.9999	1.13×10^{-7}	0.9999	1.13×10^{-7}	0.002
100	9.24×10^{-8}	0.9992	9.54×10^{-8}	0.9997	9.66×10^{-8}	0.9998	9.48×10^{-8}	0.002
200	7.4×10^{-8}	0.9997	7.06×10^{-8}	0.9995	7.38×10^{-8}	0.9999	7.28×10^{-8}	0.0019
300	5.64×10^{-8}	0.9999	5.89×10^{-8}	0.9997	5.66×10^{-8}	0.999	5.73×10^{-8}	0.00013
400	3.99×10^{-8}	0.9990	3.66×10^{-8}	0.9992	3.98×10^{-8}	0.9999	3.88×10^{-8}	0.0002
500	2.22×10^{-8}	0.9999	1.22×10^{-8}	0.9985	1.04×10^{-8}	0.9999	1.28×10^{-8}	0.0002
600	1.59×10^{-8}	0.992	1.21×10^{-8}	0.999	1.03×10^{-8}	0.9999	1.28×10^{-8}	0.003
700	4.5×10^{-9}	0.9999	4.7×10^{-9}	0.9999	4.3×10^{-9}	0.9999	4.5×10^{-9}	0.0002

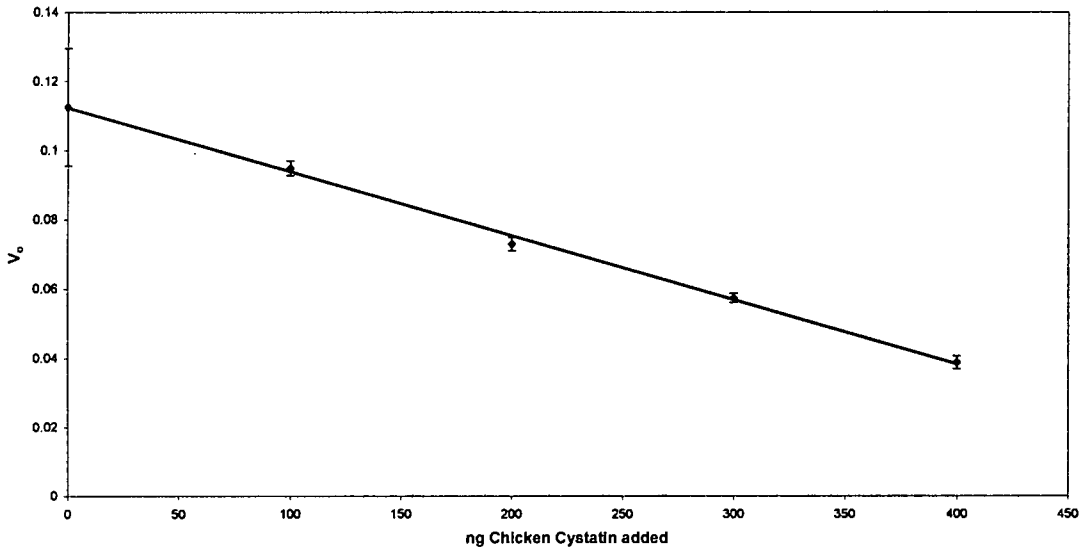


Figure E.2: *Active-site titration of chicken cystatin using E-64 standardised papain. Papain was incubated with chicken cystatin for 30 minutes at 30°C and residual enzyme activity was detected by continuous fluorimetric assays using 5 μM Z-Phe-Arg-AMC. Error bars show 95% confidence limits.*

E.1.3 Oryzacystatin I Active-Site Titration

Inhibitor Conc. (µg OC I)	Experiment 1		Experiment 2		Experiment 3		Average V_o	Std dev
	V_o	r^2	V_o	r^2	V_o	r^2		
0	7.44×10^{-8}	0.9997	7.08×10^{-8}	0.9969	7.01×10^{-8}	0.9981	7.17×10^{-8}	0.003
200	5.69×10^{-8}	0.9992	5.94×10^{-8}	0.9998	5.97×10^{-8}	0.9997	5.87×10^{-8}	0.0016
400	4.4×10^{-8}	0.996	4.36×10^{-8}	0.9997	4.38×10^{-8}	0.9995	4.38×10^{-8}	0.0002
600	3.11×10^{-8}	0.9992	3.12×10^{-8}	0.9993	3.4×10^{-8}	0.9992	3.11×10^{-8}	0.0017
800	1.42×10^{-8}	0.9994	1.4×10^{-8}	0.9992	1.53×10^{-8}	0.9995	1.45×10^{-8}	0.9994
1000	7.3×10^{-9}	0.9994	7.4×10^{-9}	0.9874	7.3×10^{-9}	0.9984	7.3×10^{-9}	0.9994

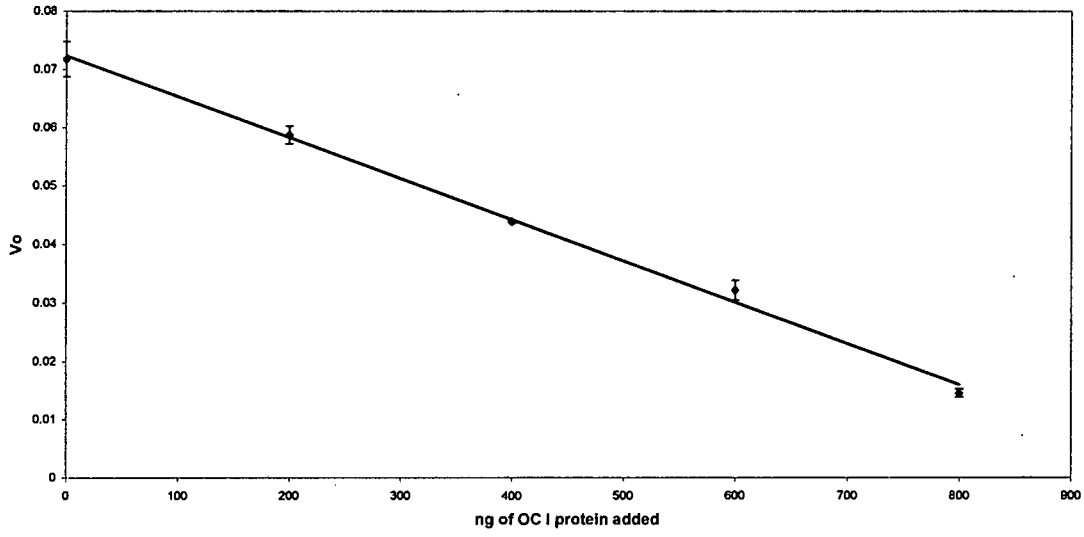


Figure E.3: *Active-site titration of OC I against E-64 standardised papain. Papain was incubated with OC I for 30 minutes at 30 ℃ and residual enzyme activity was detected by continuous fluorimetric assays using 5 µM Z-Phe-Arg-AMC. Error bars show 95% confidence limits.*

E.1.4 Fusion Active Site Titration

Inhibitor Conc. (μg HO-MBP)	Experiment 1		Experiment 2		Experiment 3		Average V_o	Std dev
	V_o	r^2	V_o	r^2	V_o	r^2		
0	8.37×10^{-8}	0.9981	8.49×10^{-8}	0.9998	9.31×10^{-8}	0.9998	8.72×10^{-8}	0.0051
500	6.85×10^{-8}	0.9973	6.53×10^{-8}	0.9995	7.02×10^{-8}	0.9984	6.8×10^{-8}	0.002
1000	5.34×10^{-8}	0.9988	5.73×10^{-8}	0.9954	6.32×10^{-8}	0.9977	5.79×10^{-8}	0.005
1500	3.64×10^{-8}	0.9998	3.42×10^{-8}	0.9996	3.48×10^{-8}	0.9995	3.64×10^{-8}	0.0011
2000	2.71×10^{-8}	0.9999	2.66×10^{-8}	0.9999	2.84×10^{-8}	0.9997	2.74×10^{-8}	0.0009
2500	1.96×10^{-8}	0.9999	1.81×10^{-8}	0.9994	1.93×10^{-8}	0.9999	1.9×10^{-8}	0.00008

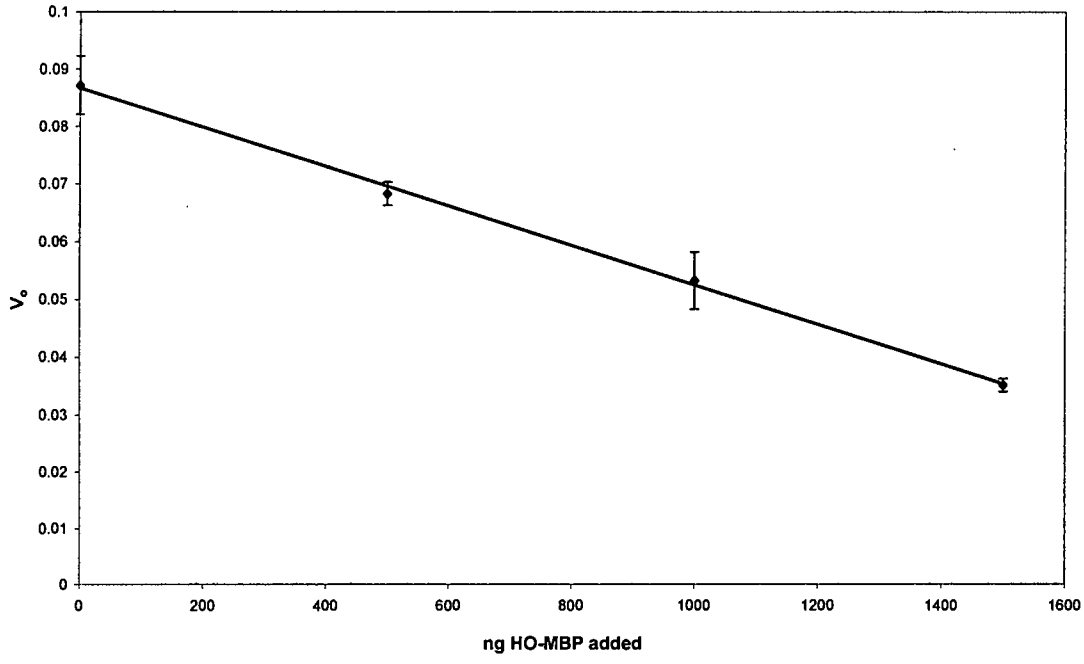


Figure E.4: *Active-site titration of HO-MBP fusion against E-64 standardised papain. Papain was incubated with the fusion for 30 minutes at 30℃ and residual enzyme activity was detected by continuous fluorimetric assays using 5 μM Z-Phe-Arg-AMC. Error bars show 95% confidence limits.*

E.1.5 Hybrid Oryzacystatin (Cut with Fxa) Active Site Titration

Inhibitor Conc. (nM HO)	Experiment 1		Experiment 2		Experiment 3		Average V_o	Std dev
	V_o	r^2	V_o	r^2	V_o	r^2		
0	9.48×10^{-8}	0.9997	9.41×10^{-8}	0.999	9.35×10^{-8}	0.9998	9.41×10^{-8}	0.0006
500	7.93×10^{-8}	0.9996	7.9×10^{-8}	0.9997	7.60×10^{-8}	0.9993	8.06×10^{-8}	0.0018
1000	6.21×10^{-8}	0.9999	5.57×10^{-8}	0.9991	6.17×10^{-8}	0.999	5.98×10^{-8}	0.0035
1500	4.09×10^{-8}	0.9999	4.26×10^{-8}	0.9997	3.6×10^{-8}	0.9989	3.98×10^{-8}	0.0034
2000	2.88×10^{-8}	0.9998	2.79×10^{-8}	0.9999	2.61×10^{-8}	0.9996	2.76×10^{-8}	0.0014
2500	6.9×10^{-9}	0.9979	1.1×10^{-9}	0.9968	1.04×10^{-8}	0.9968	9.49×10^{-9}	0.0022
3000	5.03×10^{-8}	0.9843	0.00503	0.9899	3.91×10^{-8}	09999	4.66×10^{-8}	0.0006

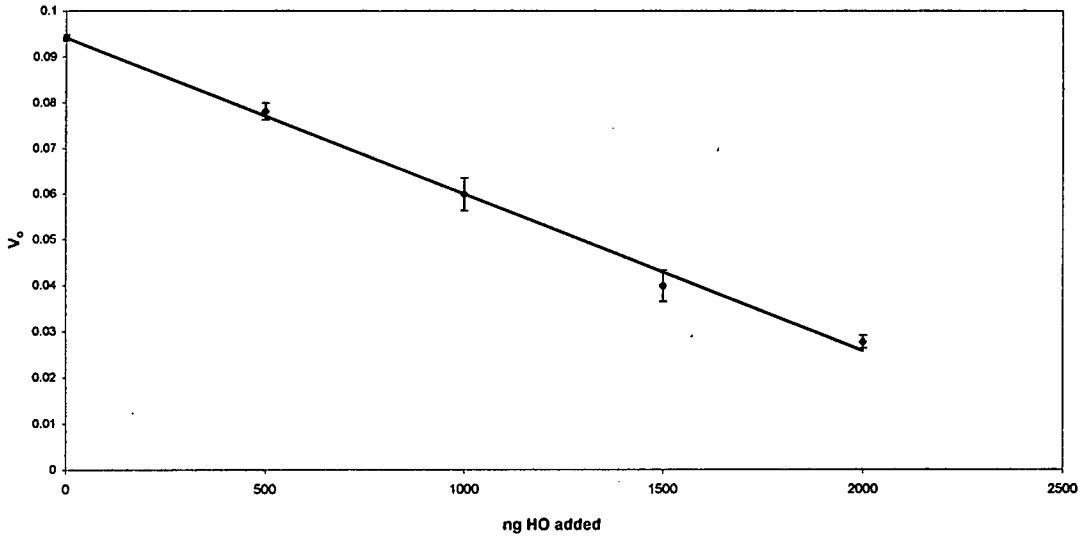


Figure E.5: *Active-site titration of HO against E-64 standardised papain. Papain was incubated with HO for 30 minutes at 30 ℃ and residual enzyme activity was detected by continuous fluorimetric assays using 5 μM Z-Phe-Arg-AMC. Error bars show 95% confidence limits.*

E.2 Determination of K_m and V_{max} Values for Papain

c	Experiment 1		Experiment 2		Experiment 3		Average V_o	Std dev
	V_o	r^2	V_o	r^2	V_o	r^2		
1.5	2.53×10^{-8}	0.998	2.73×10^{-8}	0.999	2.68×10^{-8}	0.994	2.64×10^{-7}	0.01
3.75	4.20×10^{-8}	0.999	3.83×10^{-8}	0.998	$3,71 \times 10^{-8}$	0.998	3.77×10^{-8}	0.0025
7.5	1.23×10^{-7}	0.999	1.24×10^{-7}	0.999	1.21×10^{-7}	0.999	1.23×10^{-7}	0.0016
15	3.12×10^{-7}	0.999	3.03×10^{-7}	0.999	2.94×10^{-7}	0.999	3.03×10^{-7}	0.0092
22.5	4.64×10^{-7}	0.999	4.32×10^{-7}	0.999	4.54×10^{-7}	0.999	4.50×10^{-7}	0.016
30	6.36×10^{-7}	0.999	6.07×10^{-7}	0.999	5.81×10^{-7}	0.999	6.08×10^{-7}	0.02
45	6.65×10^{-7}	0.999	7.25×10^{-7}	0.999	6.86×10^{-7}	0.999	6.92×10^{-7}	0.03
75	8.16×10^{-7}	0.999	8.01×10^{-7}	0.999	7.98×10^{-7}	0.993	8.05×10^{-7}	0.009
100	1.04×10^{-7}	0.997	1.03×10^{-7}	0.995	8.9×10^{-8}	0.998	9.88×10^{-8}	0.008

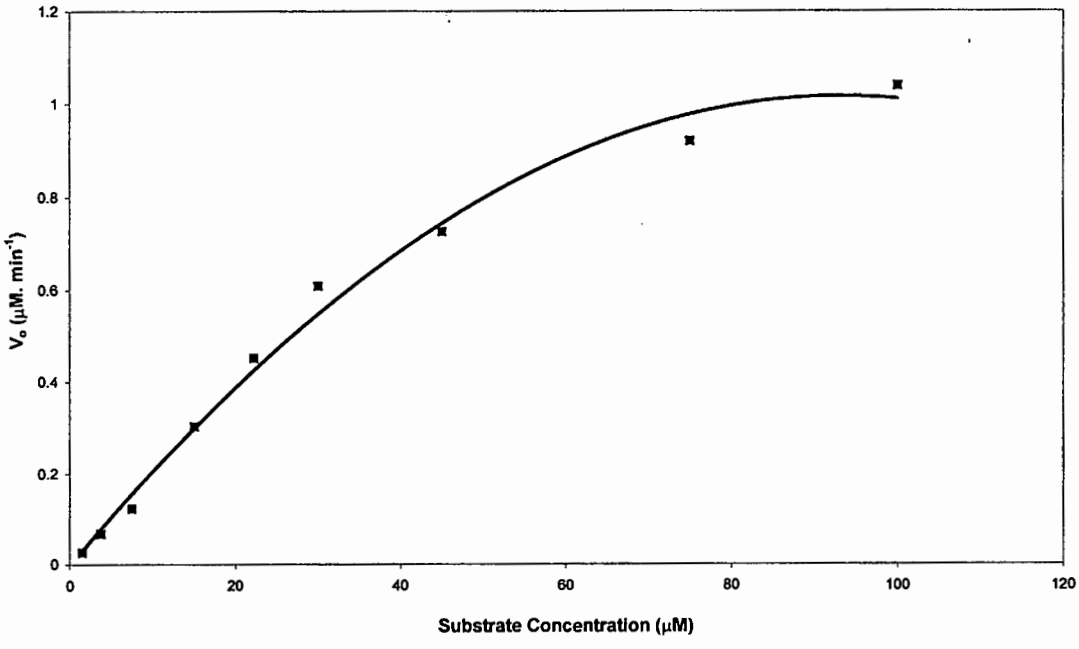


Figure E.6: Michaelis-Menten plot for papain.

E.3 Determination of Equilibrium Constants (K_i)

E.3.1 Data For K_i Determination of the Hybrid Protein

Inhibitor Conc. (nM HO)	Experiment 1		Experiment 2		Experiment 3		Average Vo	Std dev
	Vo	r^2	Vo	r^2	Vo	r^2		
0	2.03×10^{-9}	0.9999	1.42×10^{-9}	0.9999	1.95×10^{-9}	0.9999	1.49×10^{-9}	4.66×10^{-9}

Inhibitor Conc. (nM HO)	Experiment 1		Experiment 2		Experiment 3		Average Vi	Std dev
	Vi	r^2	Vi	r^2	Vi	r^2		
8	3.46×10^{-11}	0.9998	2.98×10^{-11}	0.9997	3.22×10^{-11}	0.9995	3.22×10^{-11}	2.40×10^{-12}
16	3.56×10^{-11}	0.9998	2.81×10^{-11}	0.9991	3.72×10^{-11}	0.9993	3.38×10^{-8}	4.83×10^{-12}
24	3.47×10^{-11}	0.9993	3.23×10^{-11}	0.9991	3.26×10^{-11}	0.9998	3.32×10^{-11}	1.310×10^{-12}
32	3.22×10^{-11}	0.999	3.09×10^{-11}	0.9992	3.26×10^{-11}	0.9991	3.19×10^{-11}	8.89×10^{-13}
40	2.89×10^{-11}	0.9997	2.96×10^{-11}	0.9992	3.21×10^{-11}	0.9996	3.02×10^{-11}	1.68×10^{-12}

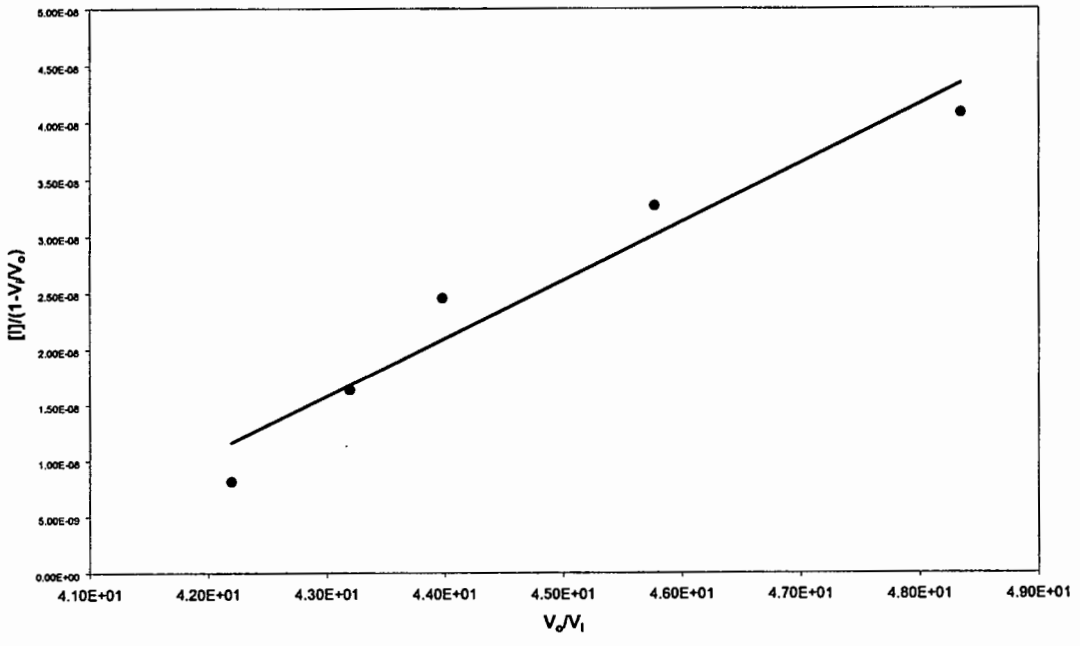


Figure E.7: *Henderson plot for the determination of the K_i of the HO inhibitor.*

E.3.2 Data For K_i Determination of the Fusion Protein

Inhibitor Conc. (nM Fusion)	Experiment 1		Experiment 2		Experiment 3		Average Vo	Std dev
	Vo	r^2	Vo	r^2	Vo	r^2		
0	8.14×10^{-11}	0.999	8.15×10^{-11}	0.998	8.01×10^{-11}	0.999	8.10×10^{-11}	7.81×10^{-11}

Inhibitor Conc. (nM Fusion)	Experiment 1		Experiment 2		Experiment 3		Average Vi	Std dev
	Vi	r^2	Vi	r^2	Vi	r^2		
20	6.96×10^{-11}	0.999	6.78×10^{-11}	0.998	7.08×10^{-11}	0.999	6.94×10^{-11}	1.39×10^{-12}
40	6.88×10^{-11}	0.999	6.73×10^{-11}	0.999	6.1×10^{-11}	0.996	6.57×10^{-11}	2.06×10^{-11}
60	6.46×10^{-11}	0.999	6.96×10^{-11}	0.999	6.26×10^{-11}	0.999	6.56×10^{-11}	3.61×10^{-12}
80	6.31×10^{-11}	0.999	5.97×10^{-11}	0.998	6.5×10^{-11}	0.991	6.26×10^{-11}	2.69×10^{-12}
100	5.89×10^{-11}	0.992	6.12×10^{-11}	0.994	5.87×10^{-11}	0.991	5.96×10^{-11}	1.39×10^{-11}

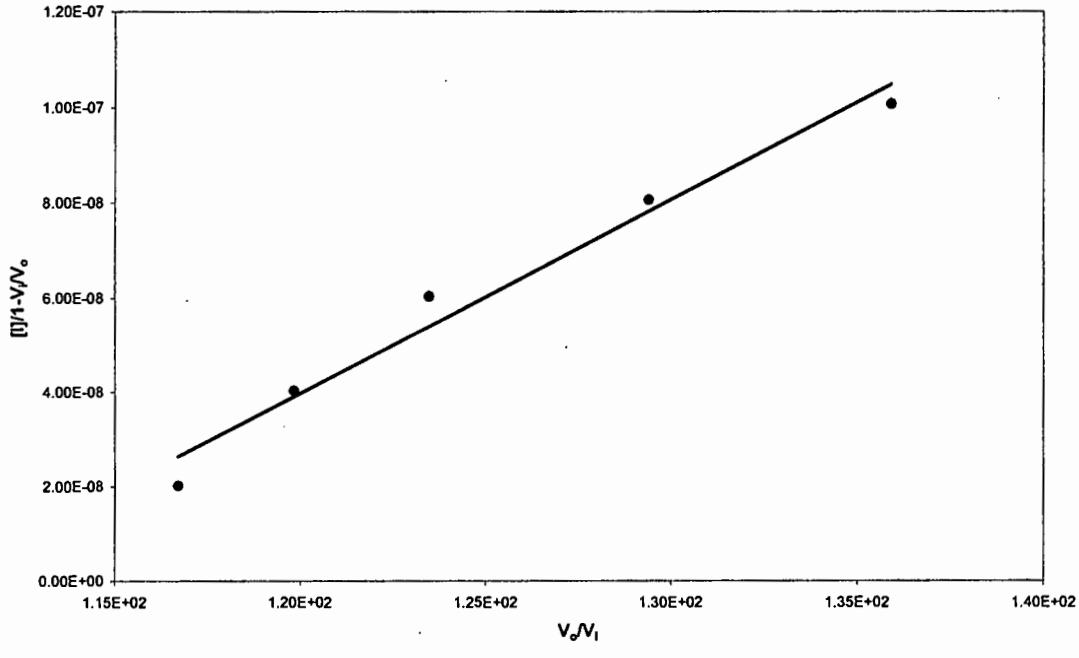


Figure E.8: *Henderson plot for the determination of the K_i of the Fusion protein.*

E.3.3 Data For K_i Determination of Oryzacystatin I

Inhibitor Conc. (nM OC I)	Experiment 1		Experiment 2		Experiment 3		Average Vo	Std dev
	Vo	r^2	Vo	r^2	Vo	r^2		
0	4.03×10^{-11}	0.999	4.25×10^{-11}	0.999	4.3×10^{-11}	0.999	4.2×10^{-11}	1.41×10^{-10}

Inhibitor Conc. (nM OC I)	Experiment 1		Experiment 2		Experiment 3		Average Vi	Std dev
	Vi	r^2	Vi	r^2	Vi	r^2		
20	4.03×10^{-11}	0.997	3.96×10^{-11}	0.999	3.71×10^{-11}	0.998	3.9×10^{-11}	1.67×10^{-12}
40	3.46×10^{-11}	0.999	3.89×10^{-11}	0.999	3.75×10^{-11}	0.999	3.7×10^{-11}	2.19×10^{-12}
60	3.71×10^{-11}	0.999	3.39×10^{-11}	0.998	3.16×10^{-11}	0.997	3.42×10^{-11}	1.63×10^{-12}
80	3.23×10^{-11}	0.999	2.97×10^{-11}	0.999	3.11×10^{-11}	0.999	3.09×10^{-11}	9.89×10^{-13}
100	2.76×10^{-11}	0.999	3.1×10^{-11}	0.999	3.14×10^{-11}	0.999	3×10^{-11}	2.83×10^{-13}

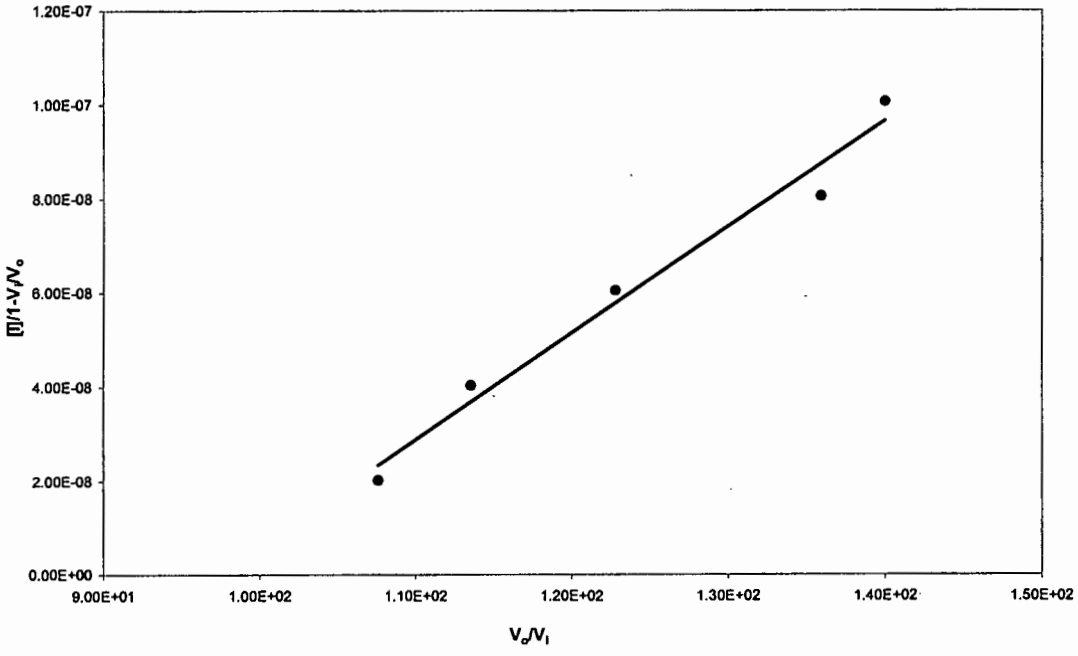


Figure E.9: Henderson plot for the determination of the K_i of OC I.

E.3.4 K_I Determination of Chicken Cystatin

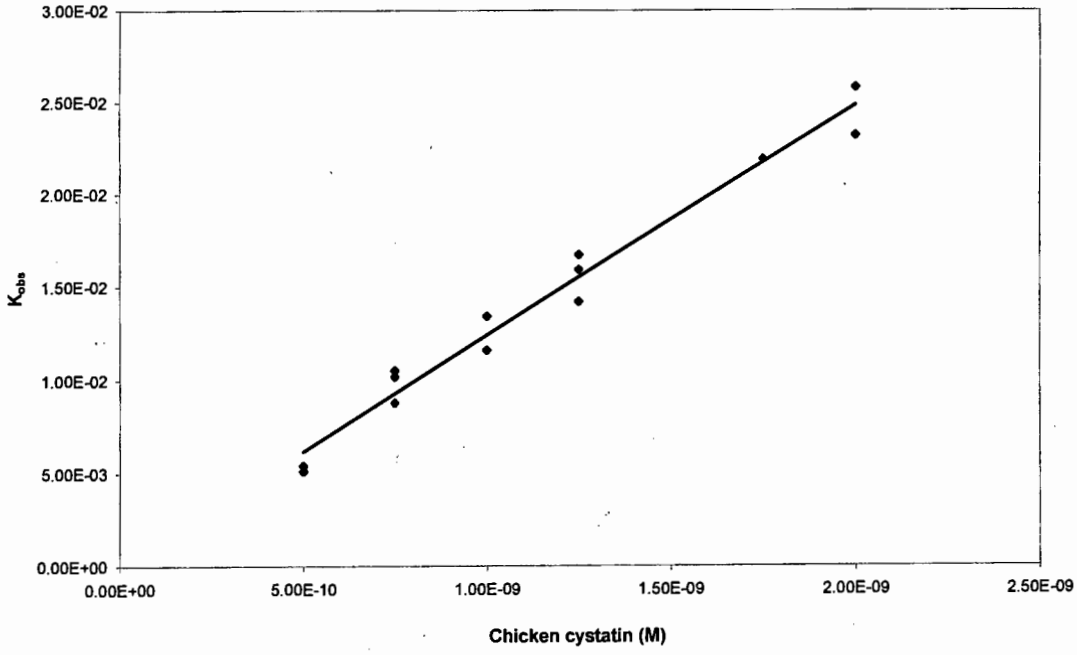


Figure E.10: *Plot of k_{obs} versus inhibitor concentration for chicken cystatin.*