

CLONING AND EXPRESSION OF A CHIMERIC PROTEASE
INHIBITOR ENCODING GENE IN *ESCHERICHIA COLI* AND
PICHIA PASTORIS

by

AARON MATSEBULA

Submitted in fulfilment of the requirements
for the degree of
Master of Science
in the Faculty of Science,
University of Cape Town

The University of Cape Town has been given
the right to reproduce this thesis in whole
or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

ACKNOWLEDGMENTS

I wish to thank the following people in the preparation of this work:

Professor D.P. Botes, my late supervisor, for proposing this project which exposed me to a wide range of recombinant DNA techniques.

I also thank A.M. Roseman who laid the ground work on this project by carrying out all the molecular modelling studies.

The assistance provided by Professor H. Klump, my supervisor, during the last phases of the project is gratefully acknowledged.

Dr. D.L. Maeder, my co-supervisor, for his valuable comments and advice pertaining to technical problems in protein purification.

Associate Professor W. Brandt for sequencing numerous peptide samples.

Dr. J.A. Rodrigues who performed amino acid analysis on many peptide samples.

Finally, general help by fellow students and staff of the Department of Biochemistry is gratefully appreciated.

SUMMARY

Squash family protease inhibitors are small peptides of 27-32 residues, hence they are ideal subjects for structure-function studies. Their small size is within the reach of peptide chemical synthesis, which enables one to produce enough peptide material for experimental purposes within a reasonable time frame.

A high degree of structural similarity has been noted between squash seeds (*Curcubita maxima*) trypsin inhibitor (CMTI-I) and carboxypeptidase inhibitor (CPI) from potatoes [1]. For CPI, inhibitory activity is located at the carboxy-terminus whereas in CMTI-I it is amino-terminus localized. Bode et al. [1], proposed that CMTI-I could be elongated on the C-terminus to make it a carboxypeptidase inhibitor.

In separate studies, Le-Nguyen et al. [2] synthesized a chimeric peptide inhibitor consisting of the sequence of EETI-II (a squash family member) plus the C-terminal tetrapeptide of CPI.

In this project, a multispecific inhibitor peptide based on the sequence of CMTI-I elongated with the four C-terminus residues of CPI was investigated using recombinant DNA techniques. The proposed hybrid gene sequence of the inhibitor protein was cloned and expressed in the prokaryote *Escherichia coli* and the unicellular eukaryote *Pichia pastoris* (a yeast). The amount of recombinant product produced by the two hosts was compared. In protein production, the number of extraction (purification) steps tend to determine the yield of final product. The cost of resins and equipment used during the purification is also a factor which must be considered.

It was found that protein secretion into the medium by *P. pastoris* facilitated purification of the inhibitor peptide, and the yields were reasonably good at laboratory scale level. On the contrary, expression of the inhibitor peptide as a fusion protein destined for translocation into the periplasm of *E. coli* required tedious extractions. The purification also necessitated the use of expensive and non regeneratable resins.

CONTENTS

Acknowledgments	i
Summary	ii

Part 1

Review On Protease Inhibitors

1.1	General	1
1.2	Protease inhibitors in plants	7
1.2.2	Background	7
1.2.2	Protease inhibitor families in plants	8
1.2.3	Effects of plant protease inhibitors on the digestive processes of insects and higher animals	9
1.2.4	The use of protease inhibitors as defensive proteins against plant pests and pathogens	11
1.2.4.1	Indirect evidence on the successful use of protease inhibitors as antimetabolites	14
1.2.4.2	Direct tests on roles of protease inhibitors in plant leaves to defend against pests	15
1.2.4.3	Protease inhibitors as likely defense agents against plant pathogens	17
1.2.4.4	Difficulties with the use of protease inhibitors	18

Part 2

Design of a chimeric protease inhibitor

2.1	Topological similarity of CMTI and CPI	20
2.2	Natural variants of CMTI and CPI	23
2.3	Enzyme inhibitor interactions	24
2.3.1	CPA/CPI	24
2.3.2	Trypsin/CMTI	26
2.4	Model structure of the hybrid peptide	26
2.4.1	The hybrid sequence	30

Part 3

Design and synthesis of a CMCPI2 encoding gene

3.1	Design of gene coding for CMCPI2	31
3.2	Synthesis and purification of the 135 mer oligonucleotide	34
3.3	Synthesis of a CMCPI2 encoding gene	37
3.3.1	Screening of transformants	40
3.4	Construction of pMAL-M2 (the pMAL-pk/CMCPI2 fusion)	42
3.5	Expression from pMAL vectors	44
3.5.1	Expression of MBP-CMCPI2 fusion protein	47
3.5.2	Purification of MBP-CMCPI2 fusion protein	49
3.5.3	Characterization of peptides generated by enterokinase cleavage	53
3.5.4	Discussion	56

Part 4

Cloning and expression of a CMCPI2 encoding gene in *Pichia pastoris*

4.1	Introduction	57
4.2	Design and synthesis of CMCPI2 inhibitor gene	61
4.3	Construction of pHIL-S1::CMCPI2*	65
4.4	Expression of the gene CMCPI2* in <i>Pichia pastoris</i>	68
4.4.1	Characterization of the purified inhibitor peptide	74
4.4.2	Discussion	75
4.4.3	Conclusion	75

Part 5

Materials And Methods

5.1	Suppliers of materials	76
5.2	Recipes	77
5.3	Methods	79

Literature Cited	95
-------------------------	-----------

LIST OF ABBREVIATIONS

dNTP	deoxynucleotide triphosphate
EtBr	ethidium bromide
rpm	revolutions per minute
EDTA	ethylenediaminetetraacetic acid
SDS	sodium dodecylsulphate
DMTr	dimethoxytrityl
bp	base pair(s)
kDa	kilo Daltons
tRNA	transfer ribonucleic acid
PCR	polymerase chain reaction
A ₂₆₀	absorbance reading at 600nm
HPLC	high performance liquid chromatography
UCT	University of Cape Town
EK	enterokinase recognition sequence
fXa	factor Xa cleavage site

One and three-letter codes for amino acids

A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartic acid	P	Pro	proline
E	Glu	glutamic acid	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
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

Part 1

Review On Protease Inhibitors

1.1 General

Protease inhibitor proteins are ubiquitous in the animal, plant and micro-organism kingdoms as they constitute an abundant group of proteins [3-5]. For example, the blood of higher animals contains cumulatively over 200 mg of various protease inhibitor proteins per 100 ml serum [3]. Storage organs of plants such as seeds and tubers, contain from 1 to 10% of their proteins as inhibitors of various types of proteolytic enzymes [6] and some fruits contain even up to 50% of their proteins as inhibitors of serine proteinases [7].

There are two groups of proteins/peptides which could be classified as inhibitors since they inhibit enzymatic reactions. The first group consist of the high molecular weight α_2 -macroglobulins which are found in plasma of all animals and are able to inhibit proteases from the four mechanistic classes; i.e., serine-, cysteine-, aspartyl-, and metallo-proteinases. Inhibition is achieved by means of a promiscuous "bait region". The second group is made up of class specific proteinase inhibitors. They have reactive sites which offer a particular peptide bond, designated P_1 - P_1' by the notation of

Schechter and Berger [8], as bait to the reactive site of the cognate protease. Usually, the nature of the P₁ residue determines the specificity of the inhibitor. Proteases and proteinase inhibitors differ profoundly in the variability of the active site. Changes in the reactive site of an inhibitor alter specificity but may not reduce inhibitory activity, whereas variations in the active site of an enzyme tends to destroy proteolytic activity [9,9a].

Class specific inhibitors mediate numerous physiological processes. Inhibitors of metallo-proteases are not very common. In plants a small carboxypeptidase inhibitor (CPI) from potatoes has been studied in detail [10]. When fed to newly hatched chicks the inhibitor was not found to be deleterious to their growth at levels in which serine proteinase inhibitors from potatoes were toxic [11]. Collagenases are the main human metallo-proteinases and are found in leukocytes, the plasma, fibroblasts and synovial fluids. The activation of these enzymes during an inflammatory response is controlled by inhibitors. If the level of inhibitor protein is low, rheumatoid arthritis, pulmonary emphysema and other cases of tissue degradation may develop [12]. Aspartate proteinase inhibitors are also not very common. However, in mammals pepsin inhibitors have been reported. In addition, renin inhibitors have been used in the control of hypertension [12]. The cystatin (cysteine proteinase inhibitors) super family is made up of at least three distinct but related families of cysteine proteinase inhibitors [13]. Major plasma proteins called kininogens are known to be involved in the inflammatory process, including activation of the intrinsic clotting cascade. More recently, it was shown that they may also regulate the

activity of other cysteine proteases such as cathepsin B, H, and L, as well as calpain I and II [14]. Cysteine proteinase inhibitors have been found to be active against the enzymes of the picornavirus responsible for polio [15].

The largest number of class specific proteinase inhibitors belong to the serpin (serine) superfamily. They have been isolated and characterised from vast sources such as plants and animals including micro-organisms. They are grouped into families based on sequence similarity, topological similarity and mechanism of binding. The following list was compiled by Ryan [16].

Plant

Squash Inhibitor Family
 Barley Trypsin Inhibitor Family
 Potato II Family
 Potato I Family
 Soybean Proteinase Inhibitor Family (Bowman-Birk)
 Soybean Trypsin Inhibitor Family (Kunitz)

Animal

Bovine Pancreatic Trypsin Inhibitor Family
 (Kunitz, Kunin)
 Pancreatic Secretory Trypsin Inhibitor Family
 (Kazal)
 Ascaris I Inhibitor Family
 Chelonianin Family

Serpin Family (mechanistically distinct)

Hirudin Family

Micro-organism

Streptomyces Subtilisin Inhibitor Family

The broad function of proteinase inhibitors in biological systems is to prevent uncontrolled proteolysis within cells, organelles, or fluids where limited proteolysis is important for biochemical and physiological processes. In so doing they aid in the maturation of many enzymes and the activity of numerous fully processed molecules. Proteinase inhibitors also protect proteins of cells, fluids or tissues from foreign proteolytic enzymes. In addition, they are implicated in signalling receptor activation or clearance. In general, the regulation of proteolysis in nature is still not well understood since the specific roles of most known proteinase inhibitors are not well defined.

Travis and Salvesen [17] implicated proteinase inhibitors in the control of the blood coagulation cascade, fibrinolysis and the complement response and in the control of fertilisation and digestion. Eglin c, a potent serine proteinase inhibitor isolated from the leech, *Hirudo medicinalis*, inhibits chymotrypsin, subtilisin, and the leukocyte proteinases elastase and cathepsin G [18]. Due to its multispecificity, it has attracted attention as a possible therapeutic agent against various pathogenic elastase-like tissue

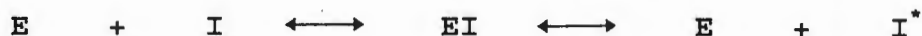
agents, blood clotting disorders and inflammatory processes.

Inhibitor deficiency is known to be linked to the development of pulmonary emphysema [19] and plaques which characterise Alzheimer's disease contain inhibitors of proteinases [20]. In plants, proteinase inhibitors are ubiquitous and are thought to play a role in the defense systems against insect pests and in protein storage [21,21a,21b]

Proteinase inhibitors prevent access of substrate to proteinase's catalytic sites by steric hindrance. There are two well documented models of protein inhibitor interactions. Small (composed of 27 to about 180 amino acids) serine proteinase inhibitors interact with substrate in a "substrate-like" (canonical) mechanism i.e., the standard mechanism. Other inhibitors, CPI for example, display product-like inhibition. The unravelling of new crystal structures uncovers novel modes of inhibition. For instance, the cystatin/stefin-papain complexes and the interaction of hirudin with thrombin [21c] implicate modes of interaction which are only partial substrate-like in character.

In substrate-like binding, intra- and inter-molecular interactions of inhibitor primary binding segments with inhibitor cores (through spacer elements) and with enzyme active sites mutually stabilize one another. Protein inhibitors interacting via this model have a compact shape, a hydrophobic core which is often disulfide interconnected, and a domain segment forms a scaffold for the exposed proteinase binding loop. The binding loop is flat in shape and fits into the active site of cognate proteinase. Residues from P_3 to P_3' (flanking

the scissile peptide bond) have a characteristic canonical conformation i.e., when analyzed in different crystal environments [22,22a] or complexation states or by NMR techniques in solution [23], the same inhibitors exhibit similar conformations. The canonical interaction may be represented as follows:



Notation; E- cognate enzyme, I - virgin inhibitor, EI- stable enzyme inhibitor complex and I*- modified inhibitor.

The EI complex is rapidly formed and usually dissociates very slowly.

Note that I* is cleaved at the scissile peptide bond P₁-P₁'.

Interactions of P₁ side chain with the specificity pocket of enzyme (S₁ subsite) are energetically the most important, hence are the primary determinants of specificity of a given protein inhibitor for a particular proteinase. For example, R or K at P₁ confer trypsin-like specificity in serine proteinases. A secondary binding segment of about 4 residues in length is also present but its interactions are not very specific. It appears that the relatively rigid inhibitor peptide loop does not lose as much conformational freedom on binding as a flexible substrate would, so that the enzyme-inhibitor complex forms with a favourable, relatively low entropy loss. Coupled with the electrostatic and van der Waal's interactions between the two highly complementary structures, this gives a large and negative free energy of association for the formation of the stable complex [24].

In product-like binding, the remaining interactions are strong enough to prevent fast dissociation. Take CPA/CPI interactions for an illustration. The projecting C-terminus residues (4) of CPI form the

primary contact region which inserts into the active site of CPA. Glycine 39 is cleaved off but remains buried by the rest of the inhibitor moiety. A few secondary contacts confer stability. The primary contact region is flexible, but upon binding it becomes rigid. The combination of primary and secondary contacts keep the truncated inhibitor in position i.e., fast dissociation or further intrusion into the active site of enzyme is prevented.

As more new crystal structures of protein inhibitors in complex with their proteinases become available new modes of interactions are being discovered. The interaction of hirudin with thrombin shows that the nucleophilic Ser195 of thrombin is not blocked, nor is its specificity pocket used by hirudin residues. It is filled instead with several structured water molecules. The globular domain of hirudin contacts characteristic thrombin surface patches adjacent to the thrombin active site.

1.2 Protease Inhibitors In Plants

1.2.1 Background

Inhibitors of proteolytic enzymes are found in large proportions in vulnerable plant tissues such as floral organs, seeds, tubers and some fruits. In particular, they have been isolated in members of the

Leguminosae, Gramineae, Liliaceae, Solanaceae and Curcubitaceae families. Plant protein inhibitors tend to be small in molecular mass; typically within the range 3Kda to 50Kda. In general, they act as regulatory agents that control endogenous enzymes and the degradation of storage proteins during seed maturation. Inhibitor peptides also play a role as protective agents against pest and pathogens attacking plants.

1.2.2 Protease Inhibitor Families In Plants

In plants, a total of about ten protease inhibitor families has been identified and most are specific for serine proteinases [25]. Ryan [26] tabulated the following list:

1. Soybean Trypsin Inhibitor (Kunitz) Family
2. Bowman-Birk Inhibitor Family
- 3 Barley Trypsin Inhibitor Family
4. Potato Inhibitor I Family
5. Potato Inhibitor II Family
6. Squash Inhibitor Family
7. Ragi 1-2/Maize Bifunctional Inhibitor Family
8. Carboxypeptidase A, B Inhibitor Family
9. Cysteine Proteinase Inhibitor Family (Cystatins)
10. Aspartyl Proteinase Inhibitor Family

The first seven are serine protease inhibitors and the last three

include families that are specific for either cysteine, aspartyl, or metallo-classes of proteases.

1.2.3 Effects Of Plant Protease Inhibitors On The Digestive Processes Of Insects And Higher Animals

Animals require proteolytic enzymes to degrade macromolecules such as proteins into simpler components which could be absorbed by their intestines. In mammals digestion is achieved extracellularly in organs such as the stomach, intestines, rumen or midgut. Numerous microbes use the extracellular digestion strategy by secreting hydrolytic enzymes into the surrounding environment. The resulting monomeric (oligomeric) amino acids are then taken up.

Since proteases play an important role in the digestive physiology of vertebrates and insects, considerable focus has been directed at the effects of protease inhibitors which are often present in food. For humans, protease inhibitors do not pose any problems because foods which contain a high concentration of inhibitors are cooked, thus inactivating the inhibitor proteins. In 1917 it had been noticed that soybeans would not support growth of rats unless cooked for several hours. Later (1940) Kunitz and Bowman discovered that beans contain trypsin inhibitors.

The adverse effects of inhibitors of proteolytic enzymes in foods are very complex. Inhibitors do not only reduce the proteolytic

activities of digestive hydrolytic enzymes; it was observed that when trypsin inhibitors are added to artificial diets of rats they stimulate pancreatic secretion to above normal levels [27]. This implicates the involvement of trypsin in the regulation of pancreatic secretions. The regulation involves the degradation of a 'monitor' peptide which is released into the gut where it regulates secretion of cholecystokinin (CCK), a circulating polypeptide hormone [28]. It appears that the interaction of inhibitors with trypsin and other proteases interferes with normal processing of the monitor peptide, which then abnormally stimulates a complex feedback mechanism that leads to chronic physiological responses in animals. Consequently, the presence of protein inhibitors in large concentrations can lead to hypersecretion by pancreatic tissues, loss of proteolytic activity in the gut, loss of appetite, starvation and eventual death.

There is no direct evidence to indicate the presence of CCK-like hormones in insects. However, insects might have feedback mechanisms which control protease production. For example, Baker et al. [29] showed that the secretion of hydrolytic enzymes in insect guts depends upon midgut protein content and not food volume. Protease secretion regulation may be attributed to two mechanisms which might not be mutually exclusive; involving either direct effects of food components on midgut epithelial cells, or a hormonal effect which is triggered by food ingestion [30]. Larvae of *Heliothis zea* and *Spodoptera* fed on artificial diets supplemented with soybean trypsin inhibitor (SBTI) plus potato Inhibitor II experienced elevated trypsin-like activities in their digestive tracts [31]. This is similar to the negative feedback of higher animals mentioned earlier.

1.2.4 The Use of Protease Inhibitors As Defensive Proteins Against Plant Pests And Pathogens

Man's need to produce more food on arable soils which are being eroded each year has necessitated the development of new crop breeding methods and protection strategies. Numerous plants have been modified by breeding and selection to the extent that they are nearly unrecognizable with respect to their native progenitors. Recombinant DNA techniques have speeded up genetic modification of agronomically important crops. The plants are grown in new, often hostile environments where they become prey to predators and pathogens against whom they have not evolved suitable counter measures. Moreover, insects, viruses and other pathogens have been transported far from their native habitats. This presents the pathogens and insects with new prey species.

The threat of pathogens and pests on crop yields led to the development of agrochemicals. The exclusive use of chemical pesticides has resulted in rapid build-up of resistance to such chemicals [32]. Insecticides tend to be non-selective therefore disturb the balance between pests and natural predators. Some pesticides being non-biodegradable, accumulate in the foodchain causing serious environmental consequences and concerns for human health [33].

One alternative to the use of agrochemicals would require the exploitation of resistant plant varieties which includes the use of

genetically engineered insect-resistant crops. Neurath [34] proposed that primitive organisms accomplished control of proteolysis by means of protease inhibitors. Plant tissues are very rich in protease inhibitors which could interact with pests and pathogens that attempt to consume them [35]. The serine inhibitor family has been isolated and characterised from plants more frequently than cysteine-, metallo-, and aspartyl-proteinase inhibitors. Numerous studies concerning the possible role of cysteine and serine proteinase inhibitors in natural plant defense systems have been done. It should be specified that the defensive role of proteinase inhibitors is only part of a complex interaction between the many defensive chemicals that are present or induced in plants and predators and pathogens that attack them [36,37]. Inducible chemicals which occur constitutively include antibiotics, alkaloids, terpenes and proteins in the form of enzymes, lectins and enzyme inhibitors. Upon pathogen, pest or mechanical wounding, a plant initiates production of the defensive chemicals mediated by potential signals. The signals could be oligosaccharides derived from the cell wall of pathogens, insects and the plant themselves [38]; lipids such as eicosapentanoic and arachidonic acid [39]; action potentials [40]; abscissic acid [41]; phytoalexins [42]; or hormones such as indole-3-acetic acid (IAA) [43]. Recently, the peptide systemin [44,45] as well as salicylic acid have been shown to be signal molecules in the systemic induction of defenses in response to wounding. Induced chemicals are mostly the product of complex biochemical pathways whose components are encoded by many genes. Some, are under control of single genes. Single genes which code for inducible proteins have practical advantages since

they could be isolated and transferred from one species to another. Many protease inhibitors are products of single genes [46].

In considering the capability of any proteinase inhibitor in a plant tissue to inhibit a foreign protease which is released into the digestive tract of a herbivore or secreted by a micro-organism, the mechanistic class and peptide bond specificity of the protease must be known. Structural aspects of the inhibitor that determine its ability to interact specifically with a particular enzyme should also be evaluated. In addition, the association constant of the interaction between protease and inhibitor should be of sufficient strength. Typically, association constants ranging from 10^6 to 10^{10} (M^{-1}) and sometimes higher have been reported [3]. These considerations require a better understanding of the digestive physiology of insects and the effects of protease inhibitors on these processes. A knowledge of proteases present in midguts of insects is also crucial. In the order Lepidoptera, which encompasses numerous crop pests, the pH in the gut is in the alkaline range of 9 to 11 where serine proteinases and metallo exopeptidases are most active [30]. As a result, these herbivores should be targeted with serine- and metallo-proteinase inhibitors. Serine proteinases have been shown to have antinutritional effects against several lepidopteran insects [47-49].

In the case of several families in the orders Hemiptera and Coleoptera, cysteine proteinases are the main enzymes which appear to play a role in the digestion of food proteins. The midguts of these insects are characterised by mildly acidic pHs which are near the pH

optimum of cysteine proteinases (pH-5). Most coleopteran insects are seed- and leaf-eaters hence they are major crop pests. Serine proteinases have not been identified in coleopteran pests. Thus, larvae of the coleopteran cowpea weevil can consume cowpea tissues, which are rich in serine proteinases, without any adverse effects on digestion [50]. Similarly, the larvae of colorado potato beetle (another coleopteran) can readily consume potato and tomato tissues [51]. It should be clear that cysteine proteinase inhibitors are the best defense candidates to obstruct digestion of plant proteins in Coleoptera and Hemiptera insect pests.

1.2.4.1 Indirect Evidence For The Successful Use Of Protease Inhibitors As Growth Inhibitors

In the 1950's Lipke et al. [52] showed that fractions from soybeans inhibited growth as well as proteolytic activity in vitro of the mealworm, *Tribolium confusum*. Likewise, the growth of *Tribolium castaneum* larvae was affected by purified fractions of soybean proteins which were added to artificial diets [53]. Moreover, purified bowman-birk trypsin inhibitor (BBTI) at 5% of diet inhibited growth of the same larvae [54].

Diets containing a combination of protease inhibitors unearthed more facts on the subject. Soybean trypsin inhibitor (SBTI) mixed with inhibitors of bovine trypsin when fed to larvae of european cornborer at 2-5% of diets showed that only the SBTI inhibited growth and delayed pupation [55]. The corn inhibitors on their own had no

effects on growth or pupation. Broadway and Duffer [33] investigated the effects of SBTI and potato Inhibitor II (anti trypsin and chymotrypsin) on the growth and digestive physiology of larvae of *H. zea* and *S. exigua*. At a level of about 10% in the diets, growth of the larvae was hindered. The addition of methionine to the diets reversed the adverse responses. It was concluded that the inhibition of protein digestion alone did not only cause the adverse effects but also resulted in hyperproduction of digestive enzymes which enhanced the loss of sulphur amino acids by the insects.

The blood sucking insect (*Haematobia irritans* L.) when fed blood containing 0.02% (w/v) leupeptin or 0.1% (w/v) SBTI showed a reduction in fecundity [56]. In addition, the stable flies (*Stomoxys calcitrans*) raised on SBTI encapsulated in erythrocytes failed to produce eggs and a 50% mortality was observed.

1.2.4.2 Direct Tests On Roles Of Protease Inhibitors In Plant Leaves To Defend Against Pests

The first results were reported by Hilder et al. [58]. Transgenic tobacco expressing a foreign cowpea trypsin inhibitor (CpTI) accumulated this inhibitor peptide at relatively high level (1%). He noticed that the transgenics were more resistant to feeding by larvae of *Heliothis virescens* than untransformed control plants or plants that did not express the gene. Cowpea trypsin inhibitor is an antinutrient against a wide range of insects [50], such as *Heliothis*,

Spodoptera, Diabrotica, and Tribolium. All are agronomically important pests.

Johnson et al [59] transformed tobacco plants with a gene coding for tomato and potato Inhibitor II (having trypsin and chymotrypsin inhibitory activities). Other tobacco plants were transformed with a gene coding for tomato Inhibitor I (having chymotrypsin inhibitory activity). In both groups of plants, the regulation of expression was under control of the constitutive cauliflower mosaic virus (CaMV) promoter. Leaves expressing Inhibitor II protein at 50 $\mu\text{g/g}$ tissue or above caused severe growth inhibition of *Manduca sexta* larvae growth. Larvae feeding on native control plants were unaffected. In addition, plants expressing Inhibitor II at 100 $\mu\text{g/g}$ leaf tissue caused larvae to grow even less and some died. On the contrary, larvae feeding on leaves containing Inhibitor I at levels greater than 100 $\mu\text{g/g}$ leaf tissue were not affected. The results indicate that trypsin-inhibitor activity of Inhibitor II was largely responsible for the adverse effects on insect growth.

Recently, it has been demonstrated that an array of inhibitors of proteases such as trypsin, chymotrypsin, elastase, carboxypeptidase and even toxins acting in concert may add to defensive potentials of plant tissues. McIntosh et al. (1990) [60] illustrated in feeding trials with *H. virescens* that a fusion protein comprising *Bacillus thuringiensis* var. *kurstaki* HD-1 protein fused to a squash trypsin inhibitor (CMTI) potentiated sublethal concentrations of the Bt protein alone. Clearly, numerous aspects of proteinase inhibitor functions are not fully understood. Boulter (1990) [61] used progeny

whose transgenic parents expressed either the CPTI gene or a pea lectin gene. Progeny that inherited and expressed both genes demonstrated an enhanced resistance to *H. virescens* when compared to plants that expressed either one of the two genes. This multifunctional approach closely resemble the situation in nature. Finally, a gene which could be useful in this approach is that of cathepsin D. The product of the gene is capable of inhibiting trypsin, chymotrypsin and cathepsin D.

1.2.4.3 Protease Inhibitors As Likely Defense Agents Against Plant Pathogens

During the nineteenth century the fungus, *Phytophthora infestans*, caused widespread famine and waves of migration from Ireland and other european countries. The discovery of fungicides brought the disease caused by the fungus in plants under control. Bacterial plant diseases are still a problem and we have to find an alternative to the copper sulfate treatment in order to reduce expenses and possible environmental costs.

Only indirect evidence for the involvement of protease inhibitors in the control of microbes has emerged. The availability of transgenic plants expressing foreign proteinase inhibitor genes should allow direct testing. Peng and Black [62] showed that levels of trypsin inhibitor increased in leaves in varieties of tomato that were resistant to *phytophthora infestans* over those found in susceptible varieties. An increase in proteinase inhibitory activity in melon

plants infected with *Colletotrichum lindemuthianum* or treated with an elicitor from this organism was noted by Roby et al. [63].

Suspension-cell cultures of tobacco treated with elicitor obtained from *phytophthora parasitica* showed an increased proteinase inhibitor synthesis [64]. Activities of proteinases from *Fusarium solani* [65] and *C. lindemuthianum* [66] were reduced by proteinase inhibitors isolated from healthy bean and tomato plants. It was also proven that a cocktail of proteinase inhibitors isolated from potato tubers [67], green tomato fruits [68], and pumpkin reduced growth of microbes present in spoiled tissues of these organs. Presumably, proteolysis of protein substrates by enzymes secreted by the micrograms was limited.

Inhibitors directed against microbial proteinases are present in plants, especially legumes. For example, inhibitors of the serine class of enzymes secreted by *Bacillus subtilis* are common in seeds or vegetative tissues of legumes, cereals and tuberous crops [69].

Inhibitors of subtilisins have been identified in at least three of the inhibitor families i.e., potato inhibitor I family, the kunitz family and the cereal/amylase inhibitor family.

1.2.4.4 Difficulties With The Use Of Protease Inhibitors

Numerous strategies for cloning genes into genomes of dicotyledons are available, but some monocotyledonous plant species have turned out to be difficult to transform. The successful strategies include: differential expression with respect to genotype, tissue-specificity

or physiological conditions; transposon tagging; biochemical characterisation of binding sites for race-specific elicitors; and shotgun cloning. Sucking pests such as aphids do not rely on proteolysis for digestion and therefore do not appear to have digestive proteases. They depend on free amino acids in the phloem as a nitrogen source. For a given insect species, some inhibitors are found to be effective antimetabolites whereas others are not. In addition, adaptations in the type of enzyme used for digestion by insects have been observed; seed weevils feeding on plant tissues rich in inhibitors of serine proteinases have replaced these digestive enzymes with a thiol protease [50,70]. The identification and isolation of genes that can improve crop yields is still in its infancy. A good deal remains to be learned of the fundamental biochemistry, physiology and molecular biology of plant processes and of plant interactions with their environments. It should also be noted that no proteinase inhibitor tested is as effective against any pest as are agrochemicals and toxins.

Part 2

Design of a chimeric Protease Inhibitor

Availability of crystal structures of proteinases in complex with their inhibitors has allowed the design of several inhibitors. CPI and CMTI are competitive protein inhibitors which bind tightly to the active site of cognate proteinases.

The structure of CMTI has been solved in complex by X-ray analysis [71] and in solution by NMR [72]. Similarly, the structure of CPI has been studied in solution using NMR [73] and by X-ray crystallography in complex with CPA [74]. The structure co-ordinates of both complexes have been deposited in the Brookhaven databank. Thus, molecular modelling was conducted by a knowledge based approach i.e., using the native structures as a starting point [75].

The aim was to produce a hybrid structure that retained the crucial structural aspects of both CMTI and CPI. Computer modelling was conducted using the BIOGRAF program (Biodesign, Pasadena, USA) on an Evans and Sutherland workstation.

2.1 Topological Similarity of CMTI and CPI

The structures of CMTI and CPI were aligned on the workstation on the basis of the least squares fit to the distances between the cysteine alpha carbons; a root mean square (rms) deviation reported by Bode et al. [71], of 1.82Å, was obtained.

Crucial features which were observed are as follows:

- a. The backbones of the two structures follow the same path in space, Figure 1 and Figure 2.
- b. The disulphide pairing is identical in the two molecules i.e., the three cysteine residues of the amino terminal half are connected with those of the carboxy terminal half in identical spatial relationship.

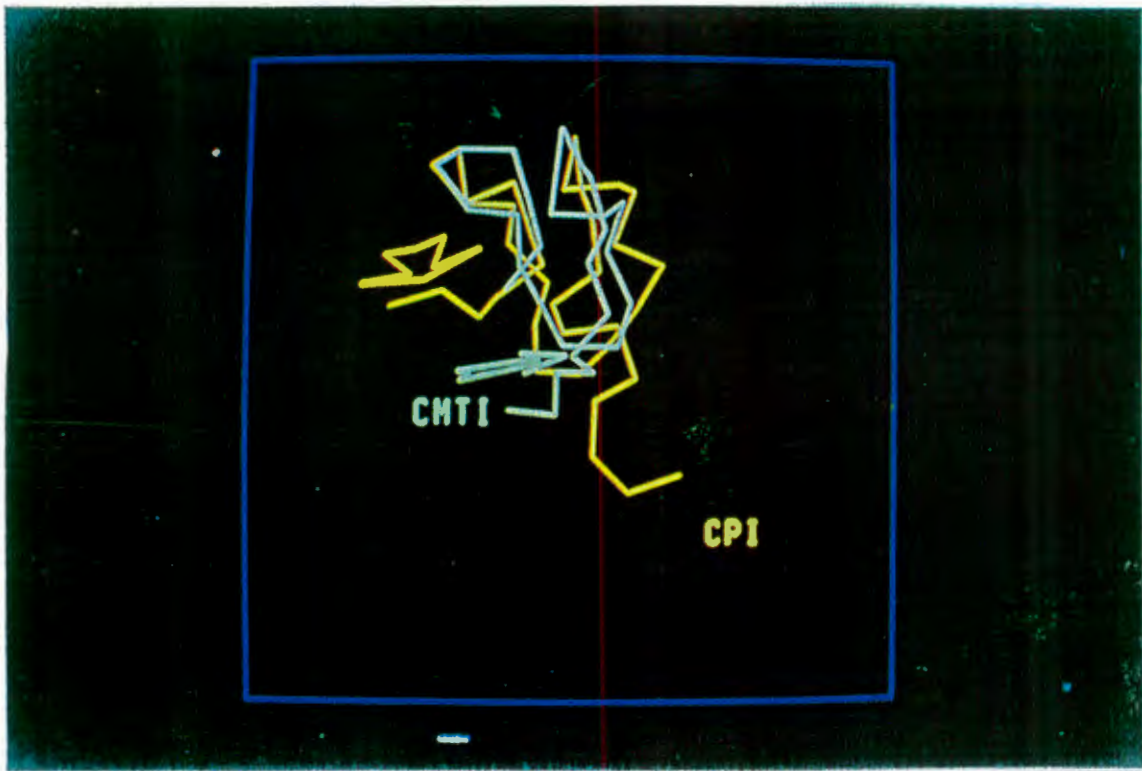


Fig. 1. C_{α} drawing of the superimposed backbones of CPI (yellow) and CMTI (white).

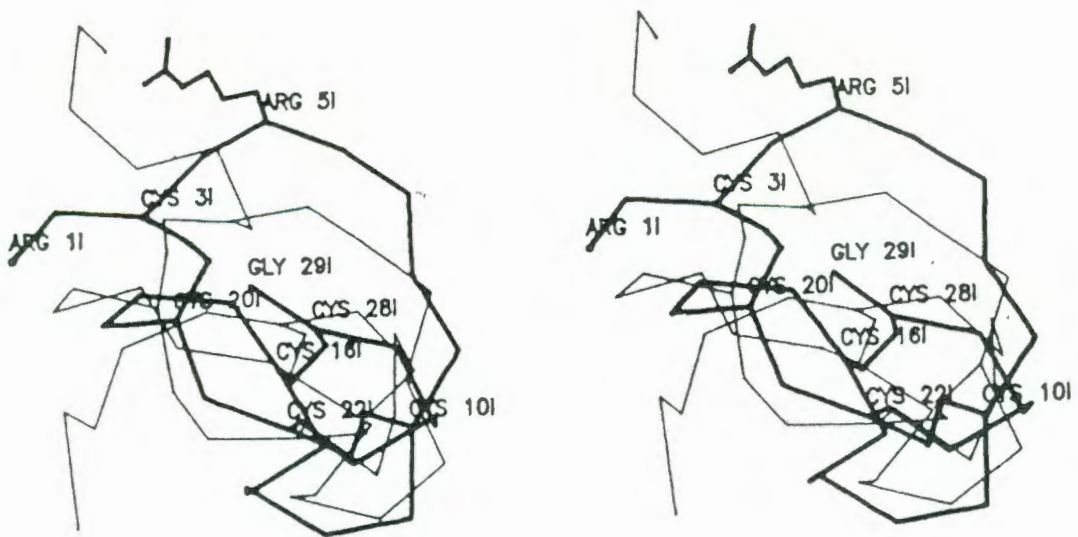


Figure 2. Stereo view on the alignment of the two backbones. CPI in bold lines and CMTI in thin lines. From reference 1.

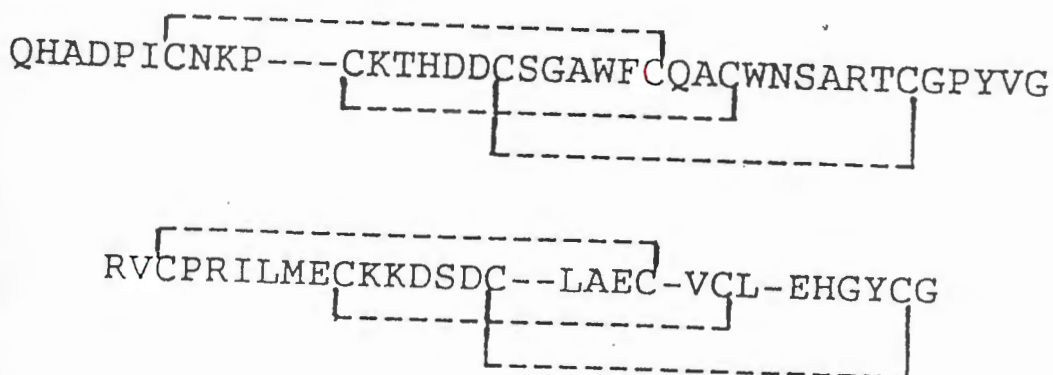


Fig. 3. Alignment of the amino acid sequences of CPI and CMTI indicating the identical disulfide pairing.

The disulphide bridges are shown in Figure 2. They contribute to a core around which the rest of the molecule is folded. The first two disulphide bridges are of right-handed whereas the third one is of left-handed conformation.

c. The active sites of the inhibitors are isolated; for CPI, the C-terminus is the exopeptidase binding site while the NH_2 -terminus of CMTI provides the endopeptidase binding site.

The alignment of the sequences of CMTI and CPI is depicted in Figure 3. The two peptides share only 4 identical residues besides the six cysteines.

Topological similarity suggests a common evolutionary origin for the two peptides. The globins are another series of proteins which show a high degree of structural homology. In such proteins, many mutations

to the primary sequence could be tolerated without any serious defect in function.

2.2 Natural Variants Of CMTI And CPI

A search through the protein sequence data bank yielded numerous analogues of CMTI-I which are displayed in Figure 4. It is evident that the residues located at the reactive site (P_1), the disulfide bonds and the COOH termini are completely conserved in all the squash family inhibitors. The exception is the bitter gourd inhibitor (MCEI-III) whose reactive site is directed against elastase.

CPI has an analogue in the tomato plant, carboxypeptidase inhibitor from tomato fruit (CTI). See Figure 4 for the alignment of the sequences of CPI and CTI.

The alignment of the natural sequences enables us to pinpoint mutations which could be tolerated because there is flexibility in the primary structures.

	5	10	15	20	25	30
TGTI-I (Towel gourd)	I C P R I L M P C S S D S D C L A E C I C L E N - G F C G					
TGTI-II (Towel gourd)	G I C P R I L M P C K T D D D C M L D C R C L S N - G Y C G					
TTI-I (Trichosanthes)	C P R I L M P C K V N D D C L R G C K C L S N - G Y C G					
TTI-II (Trichosanthes)	C P R I L M P C Q V N D D C L R G C K C L S N - G Y C G					
HMTI (Hami melon)	V G C P R I L M K C K T D D D C L L G C K C L S N - G Y C G					
MCTI-I (Bitter gourd)	E R R C P R I L K Q C K R D S D C P G E C I C M A H - G F C G					
MCTI-II (Bitter gourd)	R I C P R I W M E C K R D S D C M A Q C I C V D - - G H C G					
MCEI-I (Bitter gourd)	R I C P L I W M E C K R D S D C L A Q C I C V D - - G H C G					
CMTI-I (Squash)	R V C P R I L M E C K K D S D C L A E C V C L E H - G Y C G					
CMTI-II (Squash)	R V C P R I L M K C K K D S D C L A E C V C L E H - G Y C G					
CMTI-IV (Squash)	H E E R V C P R I L M K C K K D S D C L A E C V C L E H - G Y C G					
CPGTI-I (Zucchini)	R V C P K I L M E C K K D S D C L A E C I C L E H - G Y C G					
CPTI-I (Summer squash)	R V C P K I L M E C K K D S D C L A E C I C L E H - G Y C G					
CPTI-II (Summer squash)	H E E R V C P K I L M E C K K D S D C L A E C I C L E H - G Y C G					
CSTI-I (Cucumber)	M V C P K I L M K C K H D S D C L L D C V C L E D I G Y C G V S					
CSTI-IV (Cucumber)	M M C P R I D M K C K H D S D C L P G C V C L E H I E Y C G					
EETI-I (Ecballium elaterium)	G C P R I L M R C K Q D S D C L A G C V C G P N - G F C G S P					
CVTI-I (Citrullus vulgaris)	G R R C P R I Y M E C K R D A D C L A D C V C L Q H - G I C G					
BDTI-I (Bryonia dioica)	R G C P R I L M R C K R D S D C L A G C V C Q K N - G Y C G					

CPI GHANPIC-NK-P-CXTHNNCSGANFCQACWNSARTCGPYVG
 QT V H S Q GT TRF G CTI

Figure 4. Sequences of CMTI and CPI, and their homologues.

2.3 Enzyme Inhibitor Interactions

Most of the information was found in the literature and it was supplemented with observations from the model structures.

2.3.1 CPA/CPI

Figure 5a illustrates the CPK wireframe model of the inhibitor/enzyme complex and the highly complementary surfaces of the two molecules can be seen. In the C_{α} wireframe model, the extended C-terminus of CPI can be seen to be penetrating into the active site of CPA; Figure 5b.

A summary of primary and secondary contacts is listed bellow:

- * The four C-terminus residues (P36 to G39) of CPI form the primary binding segment which inserts into the active site of CPA (binding subsites (S_3 , S_2 , S_1 , S_1')). This segment is very flexible and becomes ordered upon binding.

- * A secondary contact region extends from residue W27 to S30.

Hydrophobic interaction through amino acid side chains are involved in the strengthening of primary contacts.

- * Hydrogen bonding;

Asp29c Ile247a, side chain N to carbonyl

Ser30c Ser246a, hydroxyl to carbonyl

(where c denotes CPI and a indicating CPA)

As indicated, all the primary and secondary contact residues are located in the region 22 to 39.



Fig. 5a. Illustration of the tight binding of CPI (green) and CPA (yellow).

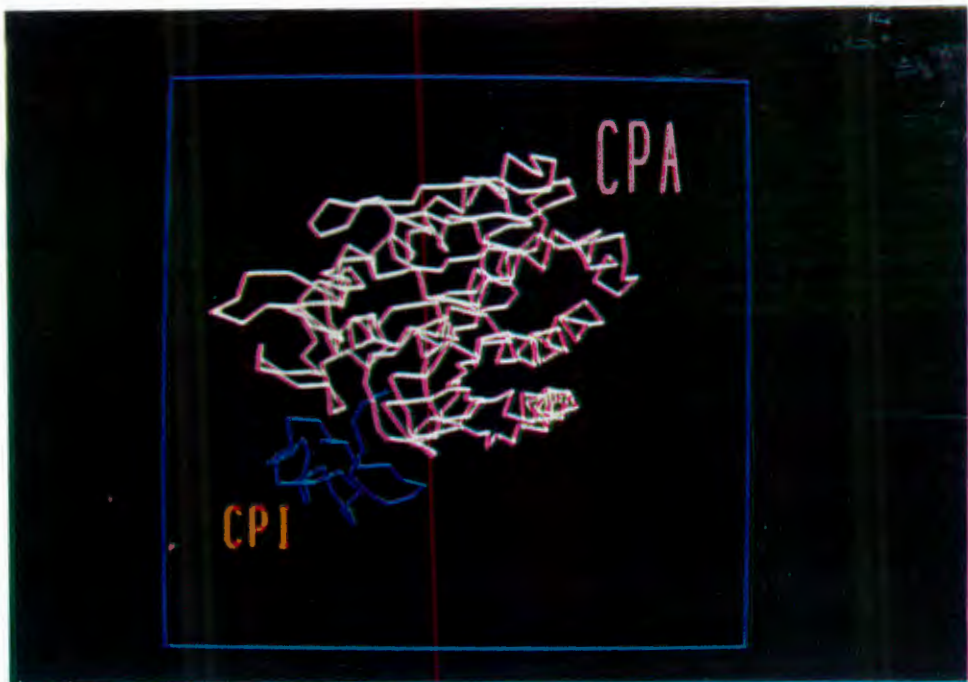


Figure 5b. C_{α} wireframe model of the same view as above.

2.3.2 Trypsin/CMTI

The tight binding between trypsin and CMTI can be seen in Figure 6a. Figure 6b shows the exposed binding loop of the inhibitor in the active site of trypsin. Trypsin cleaves on the carboxy-terminal side of Arg or Lys residues which are large, positive and hydrophilic and can be accommodated in the binding pocket and held in place by electrostatic interactions. The mechanism of action of trypsin is characterised by the concerted attack on the carbonyl carbon of the peptide bond by Ser195 to yield an oxyanion tetrahedral intermediate stabilized in an "oxyanion hole" in which the negative charge on the carbonyl carbon is stabilized by hydrogen bonds to the enzyme. The primary binding loop of CMTI is Val2 to Met8. The hydrophobic side chains of this loop (Pro4, Ile6, Leu7, and Met8) make up, together with Val2, Tyr27, Leu17 and Ala18, a curved hydrophobic surface.

2.4 Model Structure Of The Hybrid Peptide

The sequence was designed by first isolating, then joining together the appropriate N-terminus structure (residues R1 to L17) from the CMTI file with the C-terminal structure (residues G20 to G39) from the CPI file. The final hybrid structure was arrived at by docking the CPI substructure onto the CMTI backbone, then creating a peptide bond between L17 and G20. The hybrid peptide was then subjected to energy minimisation under the DREIDING force field.

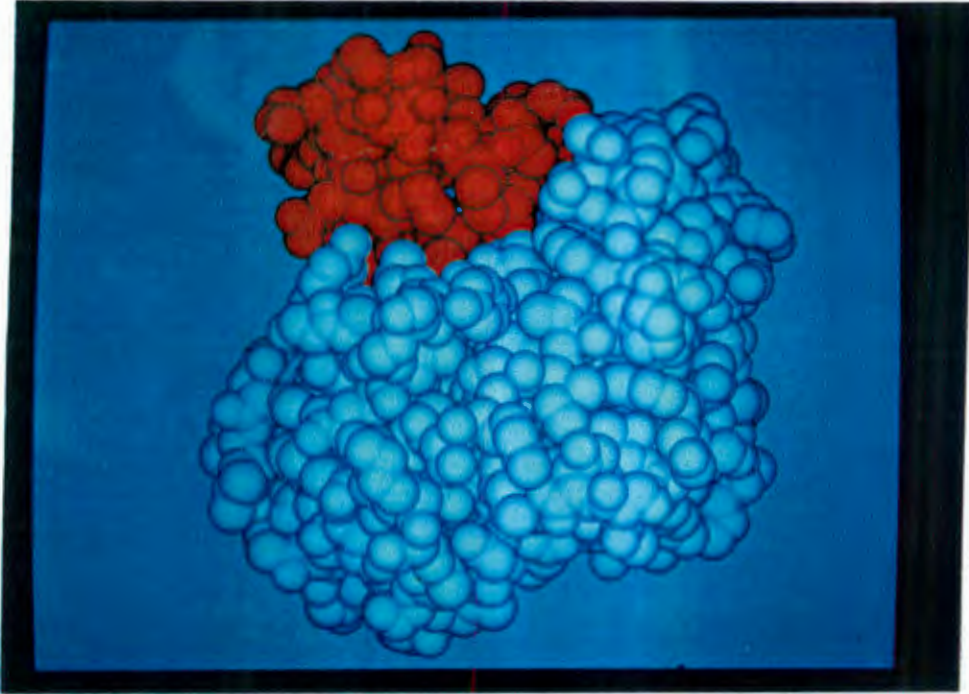


Fig. 6a. Interaction of CMTI-I and trypsin in the crystal complex.

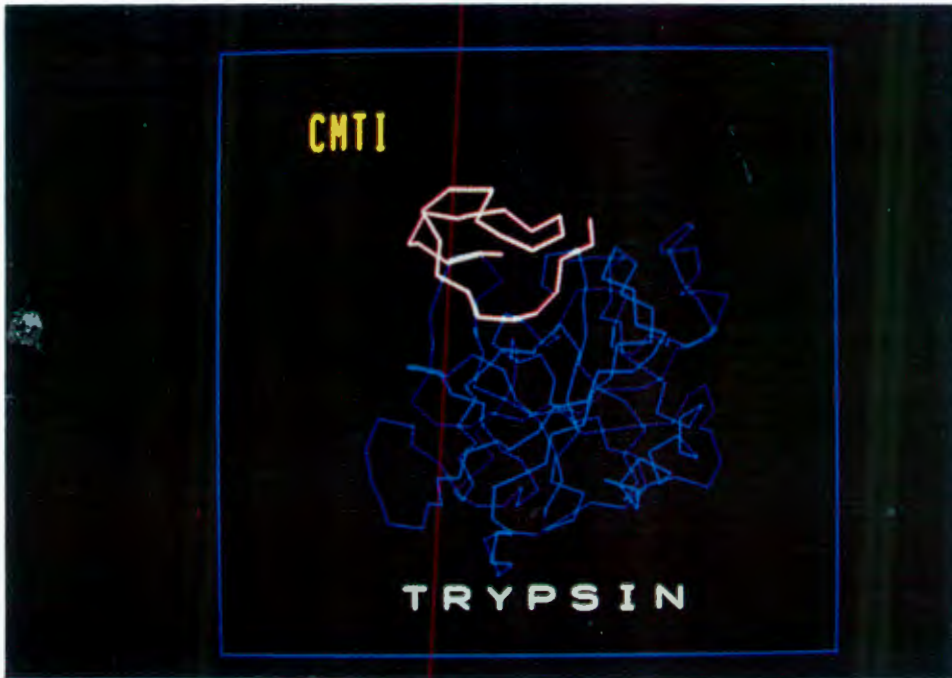


Figure 6b. Wireframe C_{α} model of the same view as above.

Figure 7 depicts the backbones of the hypothetical hybrid protein, CMTI and CPI. The hybrid closely resembles each of its "parents". In order to see if the hybrid peptide could accommodate both trypsin and CPA without any steric hindrance, it was displayed with its active segments bound to the proteases as shown in Figure 8 and 9.



Fig. 7. The alignment of the backbones of CPI, CMTI and the hybrid peptide.

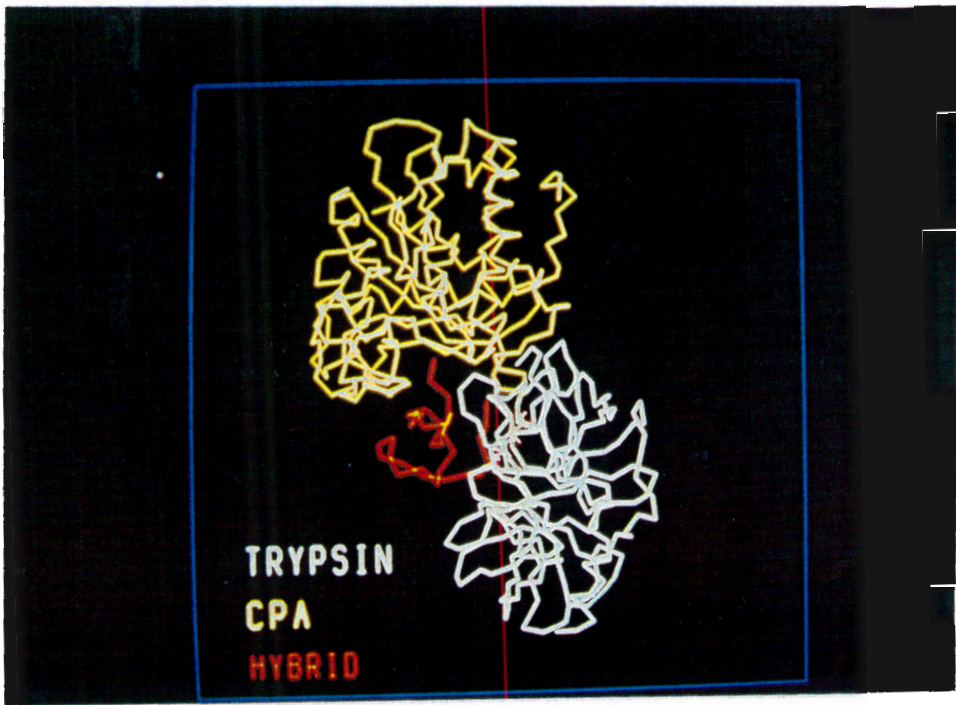


Figure 8. Accessibility of the active segments of the hybrid peptide to CPA and trypsin. C_{α} model.

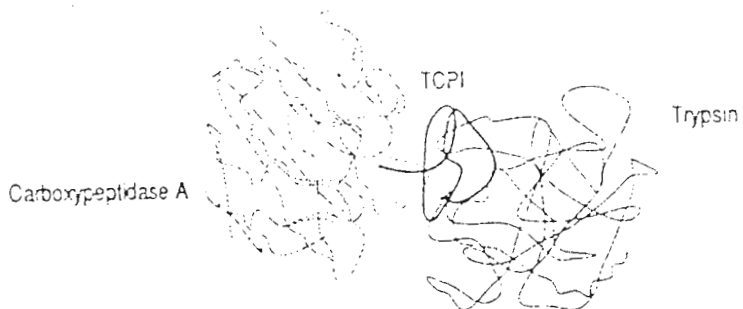


Fig. 9. Model of a similar trimolecular complex. Hybrid in bold, CPA in dotted lines and trypsin in medium thick lines. From reference 2.

2.4.1 The Hybrid Sequence

GVCPMILMECKKDSDCIAECVCGPNGYCGPYV

The sequence shown above was arrived at by modelling studies and by information from the literature. The rationale of the sequence is summarised below:

G

R1 was changed to glycine because it extends into the active site of CPI. The side chain of R1 in CMTI is involved in a hydrogen/ion pair with G29, fixing the amino and carboxy termini to the core of the inhibitor structure [71]. This carboxy terminus is no longer present. Homologous inhibitors show sequence variation at this position therefore the ionpair/hydrogen-bonding interaction is not crucial for activity or folding.

GPN

These residues were taken from the sequence of EETI-II, a squash family inhibitor. They have been implicated as a nucleation site for folding in some proteins.

PYV

These amino acids constitute the CPI's active site, and it was fully preserved with the exception of G39 which was omitted. G39 is cleaved off by CPA but the inhibitor remains active.

M

The R at position 5 (P_1) of CMTI sequence was changed to M in order to create a multispecific inhibitor. McWherter et al. [77] investigated sequence variations of CMTI-III and found that several substitutions at the P_1 site altered the specificity of the inhibitor to inhibit at least three other proteinases (human leukocyte elastase, human leukocyte cathepsin G and chymotrypsin).

Part 3**Design And Synthesis Of A CMCPI2 Encoding Gene****3.1 Design Of A Gene Coding For CMCPI2**

Gene design required the following considerations. The organism in which the gene would be expressed influenced the codon choice for each amino acid. The cloning vectors should be chosen and their characteristics studied. In this project, two vectors are used; one for gene synthesis and the other for gene expression. Vector choice governs the type of restriction sites that could be engineered in the gene for the purposes of excising the gene out of the synthesis host and subsequent subcloning it into an expression vector.

At the outset, gene design involved translating the known amino acid sequence of a peptide into a DNA sequence using preferred codons for the chosen expression host, the main reason being that you want to produce protein in abundance from the gene. Accuracy of gene

transcription is also assured when highly utilized codons are used. In single-cell organisms, Grosjean and Fiers [78] indicated a direct relationship between the level of expression of a gene and its codon bias. The genetic code comprises 60 codons which code for 20 amino acids, 3 stop codons and one start codon. This means that most of the amino acids are coded for by more than one codon (degenerate codons). In a given species, characteristic genes prefer certain codons. In this work, codons used were as reported by Ken-nosuke Wada et al. [79].

While it is important to use preferred codons, it does not mean that the inclusion of low readthrough codons would lower the level of expression of a gene. Good expression of genes with a low number of preferred codons is possible. For example, in *E. coli* the chloramphenicol acetyltransferase genes has codons similar to that of poorly expressed genes. When expressed under a strong promoter, Robinson et al. [80] observed high levels of expression. In fact, some natural genes possess rare codons which are thought to constitute pause sites that enable the nascent polypeptide to fold properly. Proteins with more than one disulphide bond tend to use pause sites as a means of allowing the correct order of -SH pairing to form [81]. Note that it is also wise to use different codons for repeated amino acids so as not to exhaust a particular species of tRNA synthetases during expression.

The coding strand of the gene is displayed in Figure 10 with the amino acids it codes for given below. A *Sma*I and a *Hind*III site were included to allow excision of the gene from the synthesis vector. As

included to allow excision of the gene from the synthesis vector. As indicated in Figure 10, the gene consists of 32 amino acid residues (99 bases) and terminates with a TAA stop codon. At the 5' and the 3' ends of the gene is a 15 base long sequence which would allow overlap between oligonucleotide and synthesis vector sequences. In total, an oligonucleotide sequence consisting of 135 bases was designed.

```

5' TCGAGCTCGGTACCC GGG GTG TGC CCG ATG ATA TTG ATG GAA TGT
   g  v  c  p  m  i  l  m  e  c

   AAA AAA GAT AGT GAT TGC TTA GCG GAA TGT
   k  k  d  s  d  c  l  a  e  c

   GTG TGT GGA CCG AAT GGT TAT TGT GGA CCT
   v  c  g  p  n  g  y  c  g  p

   TAT GTG TAA AAGCTT GGGGATCCTCTAGAG 3'
   y  v  *  HindIII

```

Fig. 10. Assignment of codons to the multispecific protease inhibitor gene. The sequence in bold (CCCGGG) is a Sma I restriction site, * - denotes a stop codon (TAA) and the underlined sequences are the two 15 bases long arms.

The sequence coding for the 32 amino acid peptide (CMCPI2) was then scanned for restriction sites using the GCG program database of 189 restriction sites. This allows one to check if unwanted restriction sites have been introduced during the assignment of codons to the amino acids. Ideally, the restriction enzymes (*Sma*I and *Hind*III) should not digest within the gene sequence. The folding of the transcribed RNA molecule was examined using MFOLD and PLOTFOLD from GCG; by changing codon usage for specific amino acids, the amount of energy required to denature any self complementary structure was minimised to a value of -7.6 Kcal/mol. This energy minimisation was done in conjunction with the optimisation of codon usage, while at the same time it was seen to use different codons for repeated amino acids.

3.2 Synthesis And Purification Of The 135 Mer Oligonucleotide

The oligonucleotide was synthesized in one run as stipulated in section 5.3.6. During synthesis, not all reactions go to completion especially for sequences longer than 80 bases. As a result there are incomplete oligonucleotides mixed together with the complete product. Incomplete sequences do not contain the 5'DMTr group, as the group is removed before addition of each consecutive phosphoramidite. Only new bases contain the 5'DMTr, hence an addition that does not go to completion lacks this group.

The presence of the DMTr group was used to separate completed

oligonucleotides from aborted sequences. The HPLC ion-exchange resin binds all oligonucleotides, as they are negatively charged. As the concentration of Cl^- (LiCl) increases, shorter sequences are eluted first since they have fewer negative charges and longer sequences are eluted last. The DMTr group is uncharged, but three phenyl groups interact non-specifically with the resin via hydrophobic interactions, causing the oligonucleotide which has the DMTr group to elute later than if it lacked the DMTr group. Consequently, there is always a gap between the elution of the longest incomplete oligonucleotide and the oligonucleotide of interest. Figure 11 shows the elution of the 135 bases long oligonucleotide and its incomplete sequences.

The purification step was conducted for the purposes of estimating the concentration of full length sequence using the HPLC chromatogram. In the following gene synthesis section, crude oligonucleotide preparations were used. Primers for PCR purposes and DNA sequencing were also used in a crude form because they were all less than 40 bases in length.

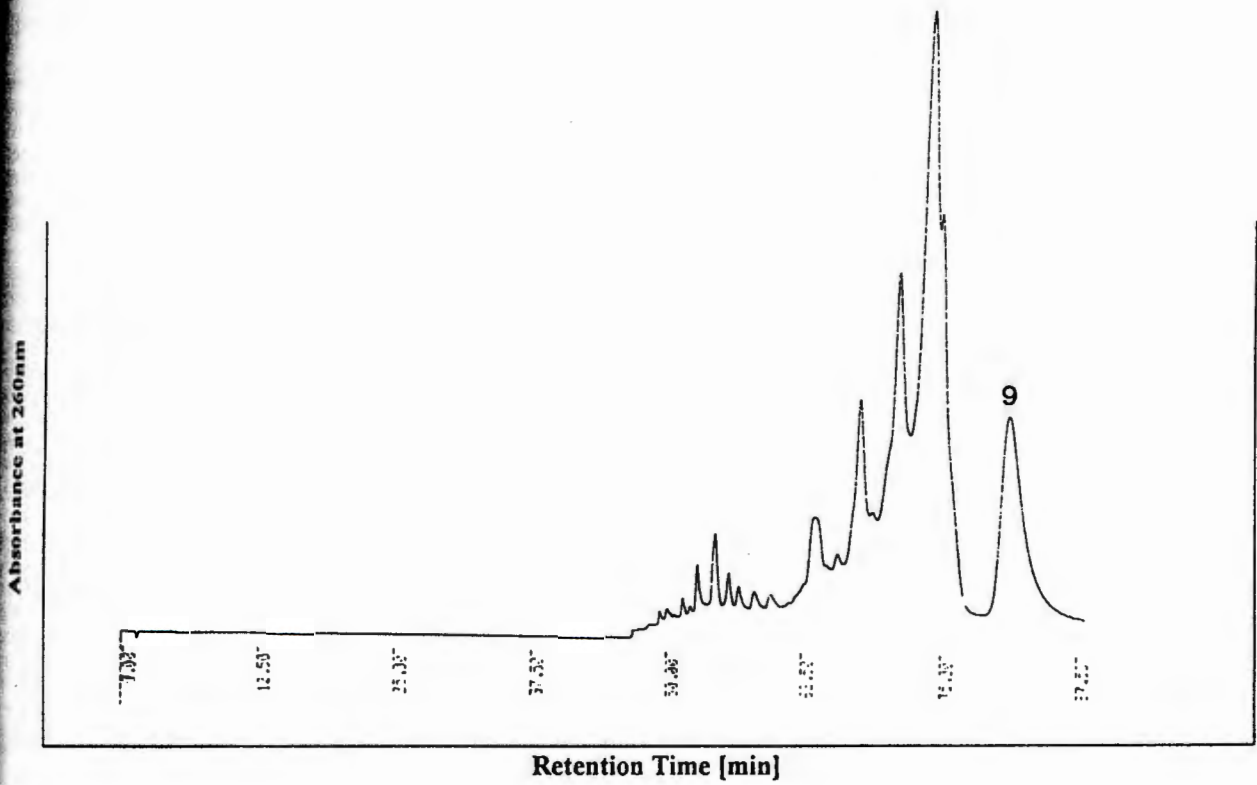


Figure 11. HPLC purification of the 135 bases long oligonucleotide. Peak 9 denotes oligonucleotides synthesized to completion, which constituted about 10% of the synthesis.

3.3 Synthesis Of A CMCPI2 Encoding Gene

The gene synthesis method described by Mandecki [82] was made use of. The method stems from the observation that an oligonucleotide with a sequence derived from DNA sequences on the two sides of a DNA double-strand break (spanning the break) can direct the repair of the break, Figure 12. The method is simple and affordable compared to PCR strategies of gene synthesis.

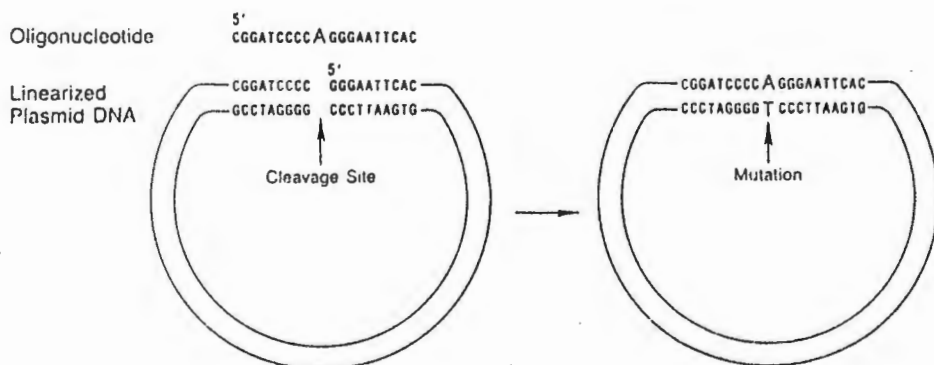


Figure 12. Principles of the mutagenesis method. The plasmid molecule is pWM300 cleaved at the Sma I site. The oligonucleotide directs repair of the gap after transformation into an *E. coli* host.

As mentioned earlier, the oligonucleotide had arms of 15 bases at the 5' and 3' ends of the CMCPI2 gene. The arms provide an overlap between the oligonucleotide and the DNA sequences of Sma I linearised

synthesis vector, pWM528. The cloning strategy is depicted in Figure 13. The parent of this vector (pWM521) is the first totally synthetic plasmid of *E. coli* which was made through the *FokI* method [83]. The plasmid contains synthetic modules for β -lactamase (*bla*) encoding gene, *lacZ* gene fragment and an origin of replication. Its multicloning site (MCS) is the same as that of pUC18. The *bla* gene allows selection for clones carrying the plasmid in medium containing a β -lactam antibiotic. And the *lacZ* gene fragment expression results (if in an α -complementing strain of *E. coli* such as JM83) in blue colonies on solid medium containing 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal). The cloning of a gene in the MCS disrupts expression of the *lacZ* gene fragment and results in white host colonies.

Two mechanisms have been proposed for the transfer of information from oligonucleotide to plasmid DNA in *E. coli*. First, it is envisioned that oligonucleotide spanning the break anneals to the strand of denatured plasmid which creates a substrate for the nick-translation/gap-repair process. Exonucleolytic repair of plasmid DNA and filling of gap by DNA polymerase may occur [84]. Secondly, a break in a plasmid DNA strand may be a signal for DNA polymerase which is in the process of replicating plasmid DNA to switch template from the plasmid DNA strand to the oligonucleotide. After passing the break, the DNA polymerase would switch template again to continue replicating the plasmid DNA [85].

3.3.1 Screening OF Transformants

JM83 cells transformed with the mixture of the oligonucleotide and denatured plasmid were plated on LB solid medium supplemented with ampicillin and X-gal. Twenty seven transformants were obtained; 17 blue and 10 white. The white clones were the ones suspected to be carrying the CMCPI2 gene.

The white clones were ascertained for the presence of insert through PCR screening as shown in Figure 14. The primers used are shown in Figure 15. A sample of JM83 cells transformed with supercoil pWM528 was also PCR screened. The product from this PCR serves as a negative control. If there is no insert in the plasmid, a 126bp product is expected. With the desired insert the product should be 231bp. The authenticity of the cloned CMCPI2 gene was further assessed by sequencing. Two clones that were sequenced using the pWM/pUC primers gave the expected gene sequence. From now on, pWM528 carrying the CMCPI2 gene is designated pWM528::CMCPI2.

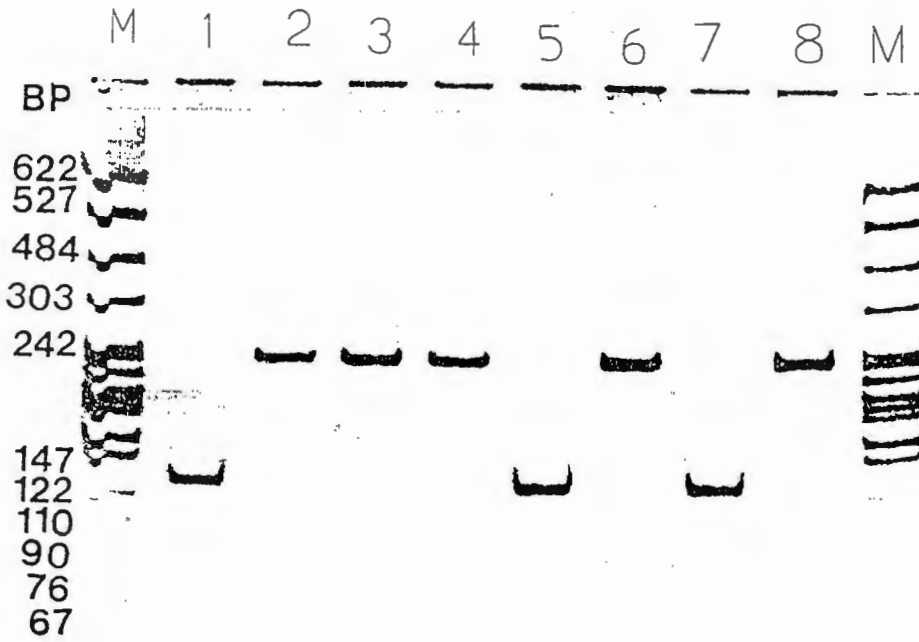


Fig. 14. Non-denaturing polyacrylamide gel electrophoresis of PCR screening results. The product of the supercoil pWM528 control (lane 1) runs lower than those samples which contain the insert (panels 2, 3, 4, 6,8) and PCR products of clones lacking insert (lanes 5, 7) have the same mobility as the control.

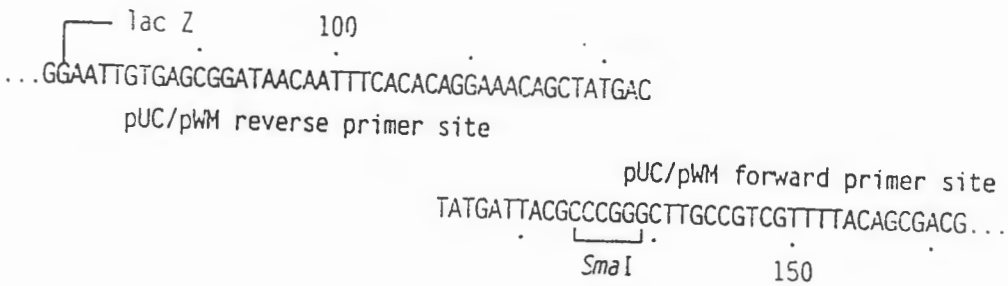


Figure 15. pWM521 (lacks MCS which is present in pWM528) sequence showing the positions of the forward and reverse primers used in PCR screening.

3.4 Construction Of pMAL-M2 (the pMAL-pk/CMCPI2 Fusion)

pMAL-pk is a derivative of pMAL-p, and contains an enterokinase cleavage sequence engineered in its MCS downstream of the factor Xa sequence. The pMAL vectors will be explained in the next section.

pWM528::CMCPI2 was SmaI / HindIII digested to excise the gene fragment which was then isolated by non denaturing polyacrylamide gel electrophoresis; Figure 16, the gene DNA (105 bp) in lane 2 is comigrating with marker DNA of 110 bp in Lane 1.



Fig. 16. Results of the isolation of CMCPI2 gene from a mixture of Hind III / Sma I digested pWM528::CMCPI2 by non-denaturing polyacrylamide gel electrophoresis.

The gene was then recovered from the gel by electroelution. The concentration of the CMCPI2 DNA was quantified and then ligated to the pMAL-pk vector which had been treated as indicated in Figure 17. In brief, pMAL-pk was digested with EcoRI followed by treatment with nuclease P1 which removes the overhangs created by EcoRI. Nuclease P1 has a single-strand-specific endonuclease activity (double stranded substrates are hydrolyzed very slowly) and a 3'nucleotidase activity

[86]. As a result, the blunt ending step was done for a limited time of 25 minutes which had been determined as appropriate. To terminate the activity of nuclease P1, the reaction mixture was extracted with phenol/chloroform. DNA obtained at this stage was digested with *Hind* III. The end result is pMAL-pk with a blunt and a sticky end into which the excised CMCPI2 gene was ligated.

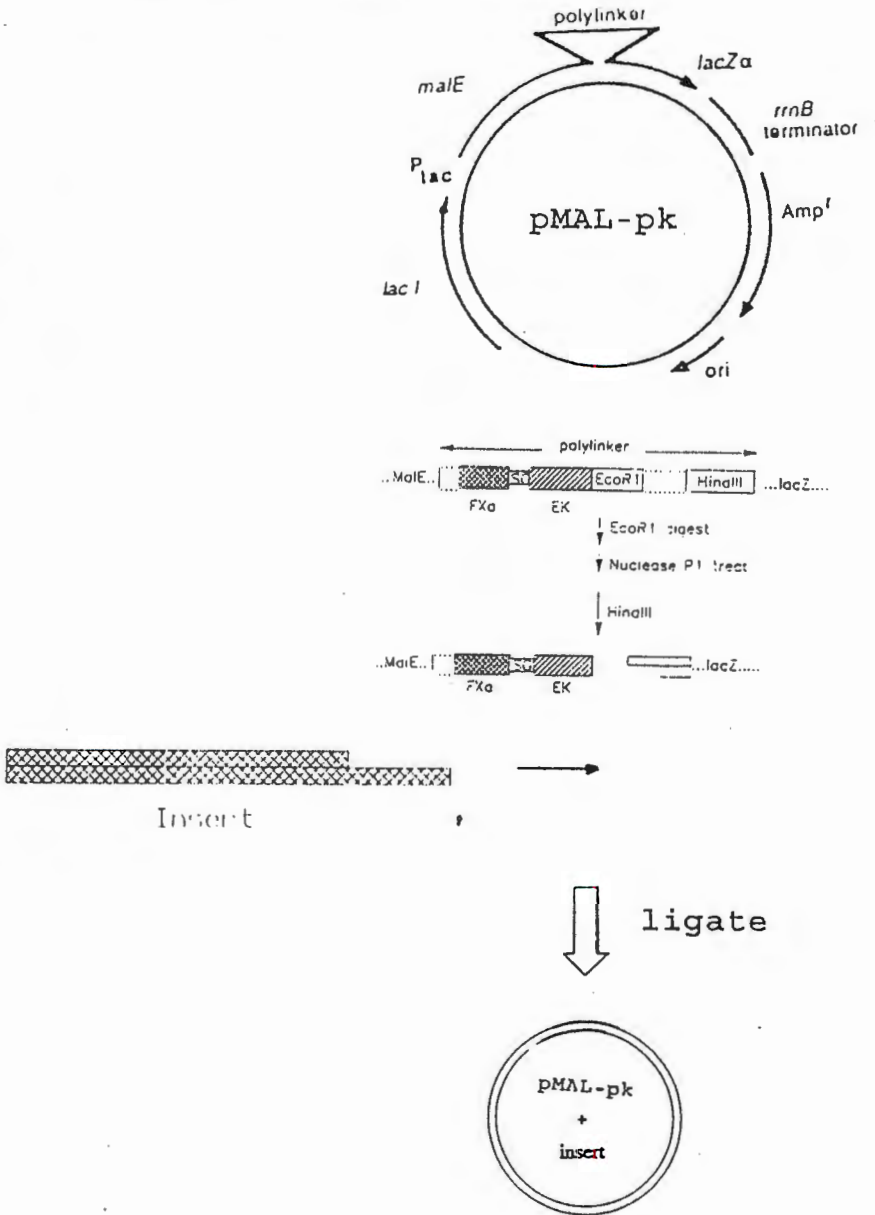


Figure 17. Construction of the plasmid pMAL-pk::CMCPI2.

3.5 Expression From pMAL Vectors

The pMAL expression vectors are supplied by New England Biolabs, and provide a simple system for expression of genes and purification of the resultant proteins; Figure 18a. Figure 18b shows a schematic drawing of pMAL-p; its polylinker region and that of pMAL-pk are indicated. These vectors supply the *malE* (which codes for 'maltose binding protein' or MBP) fused through a polylinker region to the *lacZ α* gene. Transcription of the fusion complex is controlled by the *tac* promoter, and strong *rrnB* ribosomal RNA transcription terminators have been cloned downstream of the *lacZ α* gene to prevent transcription from *ptac* from interfering with plasmid functions. The promoter is kept in a switched off or repressed state by *lac* repressor, the product of the *lacI* gene, which is also carried by the pMAL vectors. Cloning of a sequence (containing a stop codon) into the polylinker and derepression of the promoter by isopropyl- β -D-thiogalactoside (IPTG) interrupts the *malE-lacZ α* fusion. This changes colony colour from blue to white in an α -complementing *E. coli* host (TB1). Furthermore, cloning in this region gives rise to a protein fusion consisting of MBP and the target protein. To separate the two domains, a sequence coding for the cleavage site of enterokinase is provided in the vector pMAL-pk. The vectors also contain the entire signal sequence of the *malE* gene which directs expression of fusion proteins into the periplasm of *E. coli* thus facilitating extraction of the fusion protein. In addition, the periplasm offers a much more oxidative environment which is conducive to disulphide bond

formation. The formation of a fusion protein may aid in protection of the mRNA and prevent digestion of the protein by proteases.

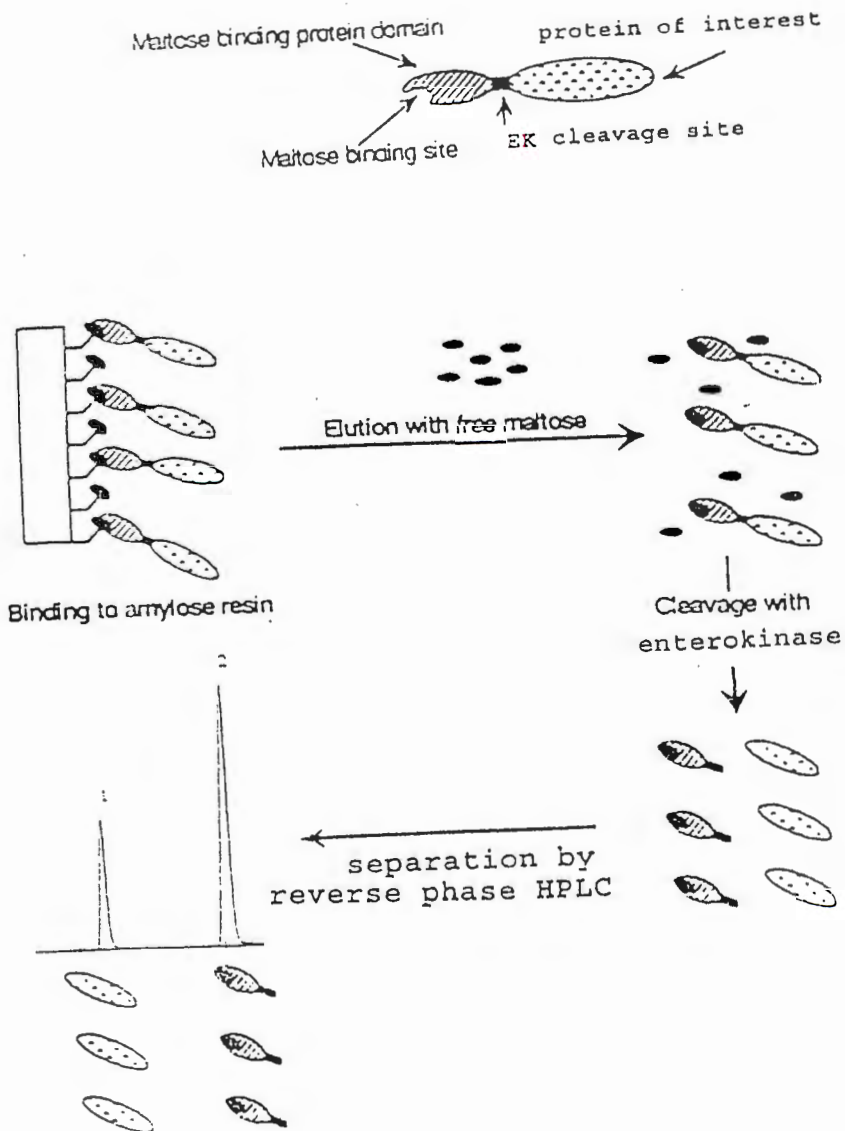
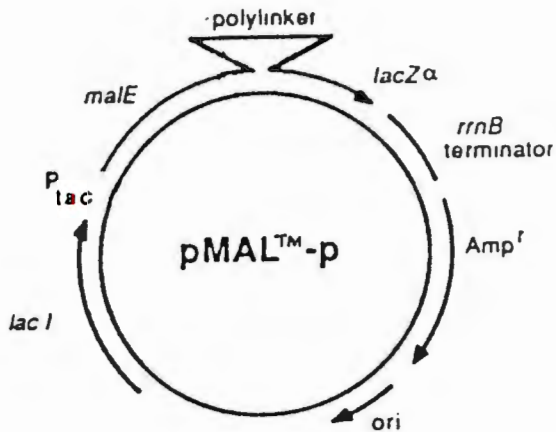


Fig. 18a. Purification scheme of MBP-CMCPI2 fusion protein and its enterokinase digests fragments.



Polylinker:

malE... TCG AGC TCG GTA CCC GGC CGG GGA TCC ATC GAG GGT AGG CCT Ile Glu Gly Arg

[Sac I] [Kpn I] [Eag I] [BamHI] [Stu I] Factor Xa cleavage site

[EcoRI] [BamHI] [Xba I] [Sal I] [Pst I] [Hind III]

GAA TTC AGT AAA ACC CTC GAT GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG...lacZα

pMAL-pk polylinker:

malE.....IleGluGlyArgSerGlyAspAspAspAspLys.....lacZα

fXa EK

Fig.18b. Plasmid map of pMAL-p and its polylinker region. The polylinker region of its derivative pMAL-pk is also shown.

3.5.1 Expression Of MBP-CMCPI2 Fusion Protein

The plasmid pMAL-M2 was introduced into *E.coli* strain TB1. PCR screening was again used to determine whether the gene, CMCPI2 was present in MCS of the vector. The correct frame of the gene was assessed by sequencing two clones which were positive by PCR screening. The final result was to choose a suitable transformant, TB1(pMAL-M2) to be used in fusion protein production.

Expression was induced by the addition of IPTG to mid log phase of the culture and incubation further to allow the fusion protein to accumulate. Samples taken before induction and one to two hours thereafter, were run with suitable controls on 14% SDS polyacrylamide gel - see Figure 19. Before induction, there appears to be no fusion protein accumulation (lane 2). This shows that there is no fusion protein synthesis from pMAL-M2 in TB1 before induction and that the repressor protein coded for by pMAL-M2 prevents detectable levels of expression from the *tac* promoter. However, one and two hrs post induction a thick band of slightly greater Mr than the 45 kDa marker protein band appears (lane 1). Pure MBP has a molecular mass of 42 kDa [87]. Since CMCPI2 peptide has a calculated molecular mass of 3498.14 Da as estimated by GCG; its fusion with MBP should give a protein of about 45 kDa. Hence it was concluded that the thick band represented expression of the cloned inhibitor gene fused to the MBP gene.

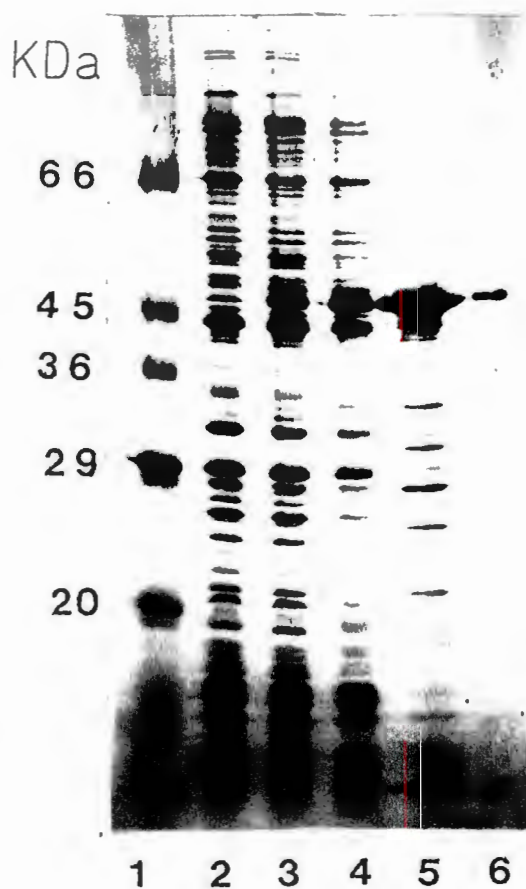


Fig. 19. Expression and purification of MBP-CMCPI2 fusion protein.

- Lane 1) Protein molecular weight markers
2) TB1(pMAL-pk) uninduced
3) TB1(pMAL-pk) 1 hr post induction
4) TB1(pMAL-pk) 2 hrs after induction
5) TB1(pMAL-pk) periplasmic extract
6) Amylose resin purified MBP-CMCPI2

3.5.2 Purification Of MBP-CMCPI2 Fusion Protein

A one litre log phase culture of TB1(pMAL-M2) was induced and periplasmic proteins were extracted by cold osmotic shock. The periplasmic preparation was then loaded on a column of cross-linked amylose. MBP has been shown to bind oligosaccharides with high affinity [88]. The interactions involve mostly hydrogen bonds between the oligosaccharide hydroxyls and charged side chains in the MBP binding groove. Van der Waal's interactions further strengthen the affinity of MBP for oligosaccharides. The result is that only the MBP-CMCPI2 fusion protein binds to the amylose resin. This is then eluted with 10 mM maltose. The maltose acts as a competitor for binding on the fusion protein. As can be seen in Figure 19, lane 6, only the MBP-CMCPI2 fusion protein band is visible. Lane 5, indicates that periplasmic expression facilitates protein purification because there are fewer protein in that compartment. Through the use of the affinity column, high yields of pure fusion protein were obtained in one purification attempt. The fusion protein at this stage is in 10 mM maltose which was removed by dialysis. Since the affinity of maltose for the MBP groove is high, fractions containing the fusion protein were diluted ten times in column buffer before dialysis against 10 litres of 10 mM Tris (pH 8.0), 100 mM NaCl with 4 changes over 24 hrs at 4°C. To remove salt, the fusion protein fraction was dialyzed exhaustively against distilled water, and later concentrated by freeze drying.

Separation of CMCPI2 from its fusion partner was achieved through the

use of enterokinase (enteropeptidase; E.C. 3.4.21.4) which is a serine protease that recognises the amino acid sequence $\text{NH}_2\text{-Asp-Asp-Asp-Asp-Lys-}\downarrow\text{-X}$. The hydrophilicity of this sequence is likely to aid in the adoption of a highly exposed conformation in the three dimensional folding of a protein. Hence in the fusion construct it would be located at the surface, facilitating processing by enterokinase. The natural substrate of enterokinase is trypsinogen which it cleaves at the C-terminus to release active trypsin [89]. The aspartic acid residues in the recognition sequence can be substituted by glutamic acid [90]. Cleavage of MBP-CMCPI2 fusion protein was carried out for 1, 3, 6, 12, and 24 hrs as depicted in Figure 20. After 6 hrs of digestion (lane 5), two digestion products can be seen. This indicates that the protein is being cleaved at two sites. Presumably, one of the site is that of enterokinase. Attempts to enhance cleavage site accessibility by denaturing the fusion protein with 6M guanidine hydrochloride prior to digestion were not successful. The enterokinase digest of the fusion protein was separated through the use of reverse phase HPLC. Proteins bind on the resin via hydrophobic interactions and are eluted in the order of increasing hydrophobicity using a gradient of acetonitrile. Figure 21 shows the results of the separation; three major peptide peaks were resolved, no. 1: CMCPI2 processed at a wrong site, no. 2: CMCPI2 cleaved at the enterokinase site and no. 3: MBP. The assignment of peaks (1 and 2) to peptides was done through N-terminus sequencing.

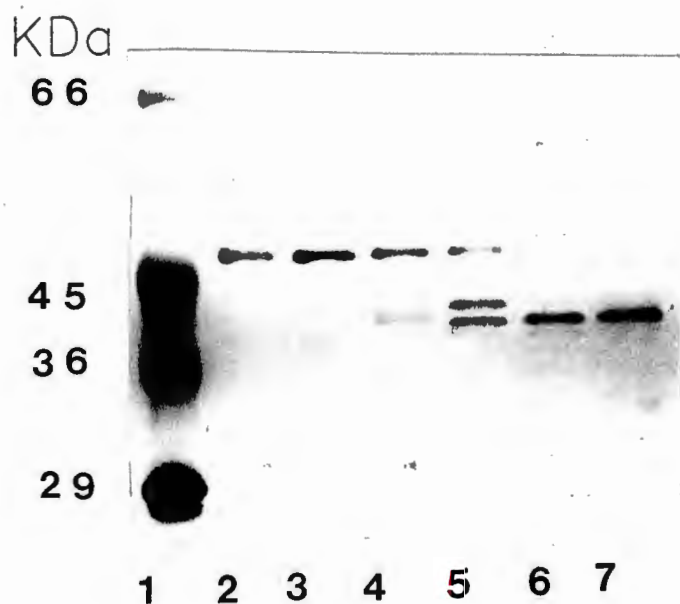


Figure 20. Cleavage of MBP-CMCPI2 fusion protein.

Lane 1) Calibration proteins

2) Uncut fusion protein

Lanes 3, 4, 5, 6, 7: incubation of the fusion protein with enzyme for 1, 3, 6, 12 and 24 hr(s) respectively

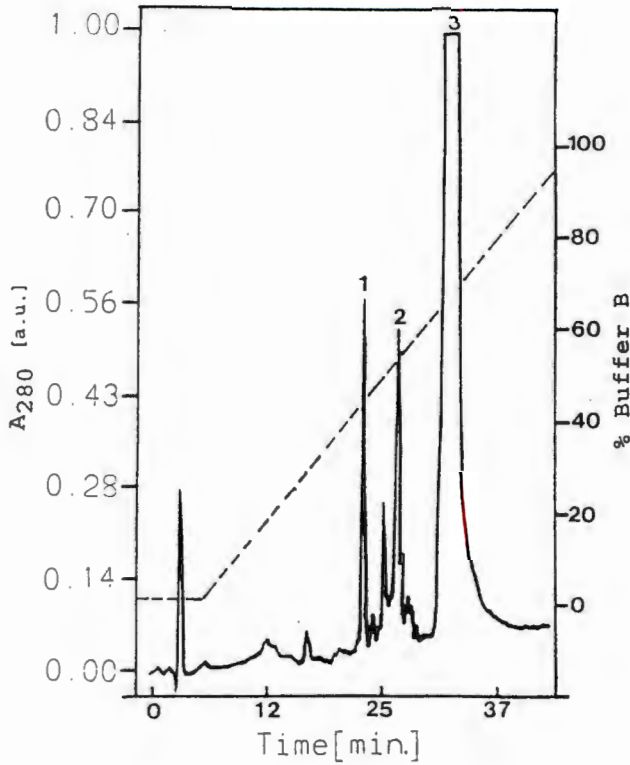


Fig. 21. Reverse phase separation of an enterokinase digest of MBP-CMCPI2 fusion protein. Column, TSK-GEL ODS-80T_m (15cm x 4.6mm ID, 5 μ m); elution solution A (0.1% TFA in water), buffer B (0.085% TFA in 80% acetonitrile); sample, 1mg digest in 1ml buffer A ; elution- 60 min linear gradient from 10% to 100% buffer B; flow rate , 0.7ml/min.

- Peak 1) Truncated CMCPI2**
2) Properly processed CMCPI2
3) MBP

3.5.3 Characterisation Of Peptides Generated By Enterokinase Cleavage

To ensure that Enterokinase cleavage had produced the required N-terminus of peak 1 and peak 2 peptides, a sample of each was partially sequenced to determine the first few residues of the N-terminus. The results are presented in Table 1a and 1b. The processing of each peptide by enterokinase is schematically shown at the bottom of table 1. It is evident that the truncated CMCPI2 was cleaved at the factor Xa site. Initially, it was suspected that the enterokinase enzyme was contaminated; an assessment of the preparation by reverse phase HPLC gave one peak, indicating high purity. The supplier of enterokinase was also changed and the pattern of digestion as assessed by SDS polyacrylamide gel electrophoresis and reverse phase HPLC were identical to the ones mentioned above. Light et al. [90] investigated limited digestion of protein substrates containing sequences of one or more acidic residues preceding a basic residue (R, K); he found that bovine serum albumin was in fact capable of digestion at these sites as a S-carboxymethyl derivative. In this study, the Gly preceding the Arg in factor Xa recognition sequence lacks side chains which could shield the Glu preceding it. As a result cleavage could take place.

A)

Cycle	PTH Amino Acid	Yield (pmoles)
1	S	166
2	G	1060
3	D	424
4	D	531
5	D	572
6	D	532
7	K	132
8	G	213
9	V	193
10	(C)	-
11	P	37
12	M	72
13	I	27

B)

Cycle	PTH amino acid	Yield (pmoles)
1	G	204
2	V	121
3	(C)	-
4	P	10
5	M	11

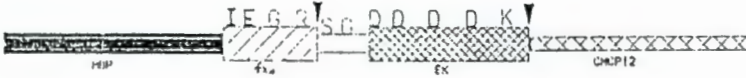


Table 1) Results of automatic sequencing of the products of

enterokinase digestion of MBPCMCPI2 fusion protein. The amino acid residues that produced largest proportions and their yields are reported.

A) N-terminal sequence of peak 1 peptide (truncated CMCP12).

B) N-terminal sequence of peak 2 peptide (properly processed CMCP12)

To understand the varying yields of the PTH-amino acids shown in table 1 a and 1 b , it is necessary to describe briefly the stepwise degradation of proteins by the method of Edman which was used in this

sequencing reaction. To prepare the N-terminal amino acid for removal from the protein, phenyl-iso-thio-cyanate (PITC) is coupled to its α -amino group to produce a phenyl-thio-carbamyl-peptide (PTC-peptide). The PTC peptide bond is cleaved by causing the N-terminal PTC-amino acid to cyclize under acid conditions producing an anilino-thiazolinone (ATZ) derivative of the amino acid. The ATZ-amino acids are unstable and are converted to PTH-amino acids usually by heating to 80°C. It is at this stage of the conversion of ATZ-amino acids to PTH-amino acids where much of the problem arises; some PTH-amino acids tend to decompose during the conversion and decrease the yield of those amino acids.

As indicated by the N-terminus sequencing results, peak no. 2 fraction contained properly processed CMCP12 peptide. Further attempts to ascertain authenticity of the peptide involved analyzing the composition of its amino acids. The results are displayed in Table 2. At this stage, it is evident that the inhibitor peptide has been expressed in *E. coli* although in minute quantities {10 to 40 μ g per fusion protein isolated from a litre of TB1(pMAL-M2) culture}.

Table 2) Amino acid composition of CMCPI2 inhibitor in mol/mol

Amino Acid	CMCPI2
G	3.7 (4)
V	2.8 (3)
C	- (6)
P	- (3)
M	1.9 (2)
E	2.2 (2)
K	1.5 (2)
D/N	3.2 (2)
S	1.1 (1)
L	1.9 (2)
I	0.9 (1)
Y	1.8 (2)
A	1.2 (1)
Total	(32)

Notation: -, not determined and (), theoretical values

3.5.4 Discussion

The pMAL expression vector chosen for this work has proven to be a very good producer of MBP-CMCPI2 fusion protein (yields as high as 21.31 mg per litre of *E. Coli* culture were obtained), which was translocated to the periplasm of *E. coli*. The idea here was that on export to the periplasm the protein would be able to fold to form correct disulphide bridges. The MBP fusion partner lacks any cysteine residues which could interfere with disulphide bond formation within the target protein. The fusion protein was purified by affinity chromatography to homogeneity as assessed by SDS-polyacrylamide gel electrophoresis. However, processing of the fusion peptide by enterokinase at the factor Xa cleavage site (NH₂-Ile-Glu-Gly-Arg-↓-) was unexpected and resulted in low yields of the peptide of interest. Only about half of the fusion protein was correctly processed in

analytical scale digests. Preparative scale digests with the serine proteinase, enterokinase, were not very successful and could be due to inhibition of the enzyme by the inhibitor, CMCPI2. These drawbacks coupled with high costs of the cross-linked amylose resin and enterokinase discouraged further attempts to produce CMCPI2 inhibitor using the pMAL expression system. Nonetheless, correctly processed CMCPI2 was obtained as indicated by the N-terminus sequencing result (NH₂-Gly-Val-Cys*-Pro-Met; *- not determined). It was then decided to express CMCPI2 in the yeast, *Pichia pastoris*. The cloning strategy, expression and purification of the inhibitor peptide is described in the next section.

Part 4

Cloning and Expression of A CMCPI2 Encoding Gene In *Pichia pastoris*

4.1 Introduction

Pichia pastoris is one of the methylotrophic yeasts which includes *Torulopsis*, *Candida* and *Hansenula*. These organisms are all capable of using methanol as a sole carbon source. The initial step in the metabolism of methanol involves its conversion to formaldehyde which makes use of the enzyme alcohol oxidase I (AOXI) [91]. The synthesis of AOXI is regulated at the level of transcription by a promoter which is induced by methanol. It was this discovery which led to the

investigation of *P. pastoris* as a vehicle for the synthesis of heterologous proteins. Early successes in the production of recombinant proteins were achieved using the bacterium *Escherichia coli* as a host. However, the *E. coli* host is not capable of expressing authentic complex recombinant proteins such as those of mammalian origin and hence it was necessary to investigate other host systems. The first yeast to be investigated was the well-studied baker's yeast *Saccharomyces cerevisiae*. As unicellular eukaryotes, the yeasts are capable of accurate post-translational processing and modification of many mammalian proteins. In general, the yields of foreign proteins produced in baker's yeast have been low because of poor plasmid stability during production runs [92]. This resulted in expression systems which may be difficult to scale-up to production volumes. The second factor contributing to low yields is that many *S. cerevisiae* secreted proteins are not found free in the culture medium, but are retained in the periplasmic space in a cell associated form [93]. This can lead to tedious purification schemes and lower yields of heterologous proteins. Obviously, there was a need to look at alternative yeast hosts for stable high-level production of appropriately modified recombinant proteins. *P. pastoris* is one of the hosts which was originally developed for industrial processes in the 1970s as a source of single-cell protein.

As a host for the production of recombinant proteins, *P. pastoris* has the subcellular machinery for performing post-translational protein modifications of eukaryotes. Secretory proteins are synthesized on ribosomes which are tightly bound to the rough endoplasmic reticulum (ER) and then translocated into the lumen of the rough ER, where the

signal sequence is cleaved. This is an important feature of any secretory system. Cleavage is carried out by a signal peptidase at a specific recognition site. A dibasic endopeptidase analogous to Kex2 of *S. cerevisiae* is implicated in cleaving secretory proteins at the recognition sequence -Lys-↓-Arg or Arg-↓-Arg [94]. The processing takes place in the Golgi complex where the endopeptidase is located. Disulfide bond formation is a post-translational modification which occurs within the rough ER. Native disulfide bond formation is an integral aspect of the protein folding pathway, and may play a part in protein assembly. In yeast, the enzyme protein disulfide isomerase (PDI) enhances isomerisation, reduction and formation of disulfide bonds [95]. This takes place in the lumen of the rough ER where the redox state is more oxidizing than that of the cytosol [96]. PDI is a homodimer and each of the monomers has two active sites characterised by the sequence -CGHC-, which is similar to that of prokaryotic / eukaryotic thioredoxin. It is proposed that the most N-terminus cysteine provides the essential reactive thiol group. The mechanism of action of PDI is not fully understood but it is thought that it facilitates protein folding so that the cysteines to be linked are brought in close proximity to each other.

The addition of a limited number of high-mannose oligosaccharides to specific Asp residues is also carried out in the rough ER. *P. pastoris* has the advantage of not being capable of hyper-glycosylation as compared to *S. cerevisiae* which can add α -1-3 glucan linked oligosaccharides that confer hyper-antigenicity to the foreign protein; such proteins are unsuitable for therapeutic use. Glycosylation is a very important process for some proteins as many

studies have shown that abnormal or inefficient glycosylation of proteins leads to their misfolding and concomitant retention in the ER [97].

The use of *P. pastoris* expression vectors especially those which secrete proteins into the media simplifies protein purification and makes use of inexpensive media components. The observation that few proteins are secreted by the organism constitute a first step in purification; secreted foreign proteins usually comprise a large fraction of proteins in the medium. The inducer, methanol of foreign protein expression is inexpensive compared to IPTG used in most *E. coli* expression vectors.

High cell-count fermentations are possible with *P. pastoris* secreting expression strains. This leads to higher yields of foreign proteins but also results in accumulation of protein degrading proteases. The degradation of secreted heterologous proteins by native proteases can be prevented in many ways. *P. pastoris* expressing strains are capable of growing across a relatively broad range of pH 3.0 to 7.0. This allows one to adjust the pH to a value where protein degradation by native proteases is limited. Amino-acid rich supplements can be added to the medium to reduce protein degradation by acting as excess substrates of the enzymes. The use of host strains lacking a gene (PEP4) which encodes a vacuolar protease that is responsible for proteolytic activation of other vacuolar proteases has proven useful [98].

Very good yields are possible with *P. pastoris* expression systems.

For example, the secretory product of the gene encoding tick anticoagulant peptide accumulated in the medium to about 1.7 g/L [99]. The same gene has been expressed in *S. cerevisiae* but product accumulated to levels which were about seven times less. Because of variations among experimental designs, it is difficult to make direct comparisons between the productivities of the two. Nonetheless, it serves as a good qualitative comparison. In the above example, only one copy of the gene was integrated into the genome of *P. pastoris*; other reports indicate a direct correlation between gene copy number and protein levels [100].

Compared to foreign protein production in *E. coli*, *P. pastoris* expression protocols are very long and require good planning. However, the genetic manipulations are similar and the medium used in both cases is inexpensive since it is a mixture of salts, trace elements and carbon source. *P. pastoris* expression systems are fairly easy to use compared to mammalian tissue culture and the baculovirus system.

4.2 Design And Synthesis Of CMCP12 Inhibitor Gene

Codons assigned to the amino acids of the protein inhibitor gene were the same as in the design for expression in *E. coli*. It would have been ideal to make use of preferred *P. pastoris* codons, but few genes have been isolated and fully sequenced from this organism. *S.*

cerevisiae is the only yeast whose genes have been well characterised in sequence. A summary of codon usage in *S. cerevisiae* and *E. coli* is tabulated by Ken-nosuke Wada et al. [79]; it is apparent that codon usage between the two organisms is similar. It has also been demonstrated that *P. pastoris* and *S. cerevisiae* show a lot of genetic cross functionality. For example, *P. pastoris* HIS4 gene functionally complements *S. cerevisiae* his4 mutants and the *S. cerevisiae* HIS4 gene functionally complements *P. pastoris* his4 mutants; other cross complementing genes that have been identified include LEU2, ARG4, and URA3. In addition, some *E. coli* genes have been produced in both yeasts in a biologically active form; Tschopp et al. [101] expressed *E. coli* β -galactosidase in both organisms. As a result, one is tempted to conclude that the inhibitor gene which was designed for expression in *E. coli* would also be expressed in *P. pastoris*. However, it is not claimed that the gene would be expressed optimally and with predicted authenticity.

The only change which was made in the CMCPI2 peptide was at position 1 where G was substituted with R for the reason of allowing processing at the K- \downarrow -R site during expression. The changes in the gene sequence which is carried by pWM528 were to be made through the use of PCR primers incorporating necessary mismatches as depicted in Figure 22. For the purposes of gene subcloning into the expression *E.coli-P. pastoris* shuttle vector, pHIL-S1, a *Sma* I site was introduced at the 5' end while a *Xho* I site was engineered at the 3' end. The proposed piece of DNA was then checked for the presence of unwanted restriction sites; *Bgl* II, *Sma* I and *Xho* I should not cut within the gene.

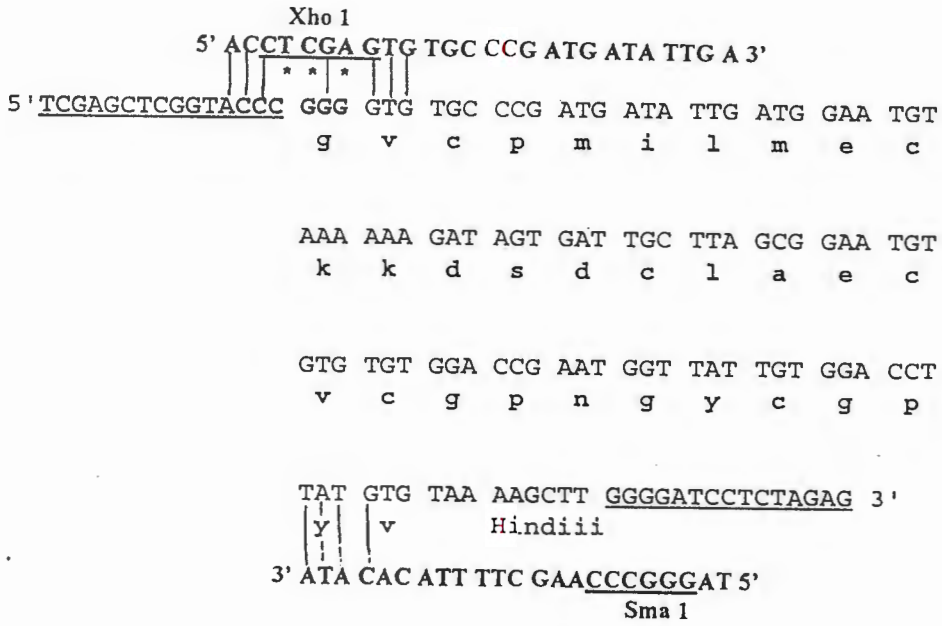


Figure 22. PCR primers indicating the mismatches (*) which would allow site specific mutation of the CMCPI2 gene into a variant which starts with R thus allowing creation of an Xho 1 restriction site.

PCR amplification was conducted in a mixture consisting of a Pst 1 linearised pWM528::CMCPI2 which facilitated annealing of primer to the linear template compared to a circular template. Thereafter, the PCR product was loaded on a 12% nondenaturing polyacrylamide gel in order to assess its size - see Figure 23. The observed bands in lanes 2 to 6 comigrated with the 120bp marker DNA in lanes 1 and 7. From the known nucleotide sequence, the PCR product is 117bp long. Since the obtained band has a molecular mass which is close to the expected one, it was concluded that the bands comprise the insert. To prepare the amplified DNA for the next step, it was treated with phenol/chloroform, digested with Sma 1 / Xho 1, and named CMCPI2^{*}.

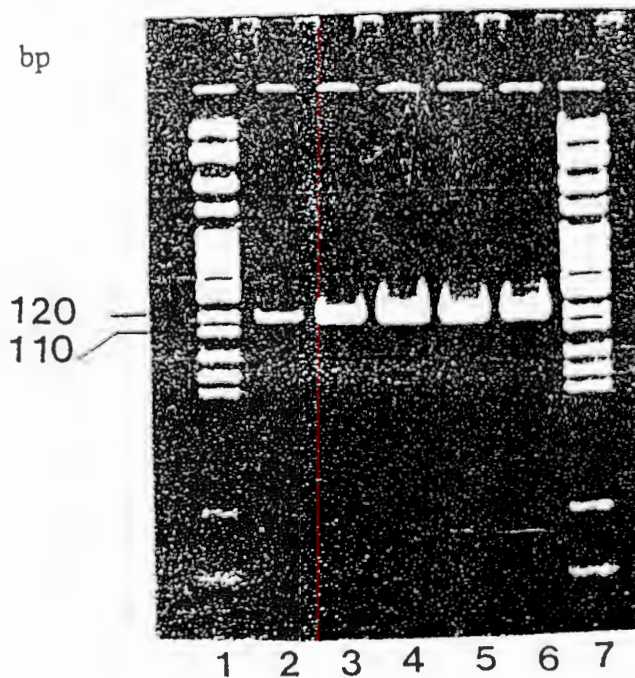


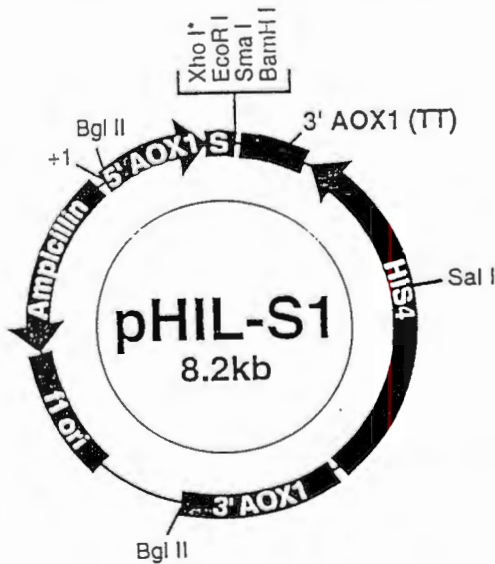
Fig 23. Results of PCR amplification of CMCPI2^{*} from the plasmid pWM528::CMCPI2.

Lanes 1,7) Calibration DNA

Lanes 2-6) PCR product at increasing concentration

4.3 Construction Of pHIL-S1::CMCPI2*

As mentioned, the plasmid pHIL-S1 is an *E. coli*-*P. pastoris* shuttle vector with sequences designed for selection in both host. Figure 24 shows a typical map of the plasmid. The left half of the plasmid is a portion of pBR322; there is an ampicillin resistance gene (AMP^r) and the f₁-bacteriophage origin of replication (f₁ ori). The EcoRI site present in this segment in the parent plasmid has been eliminated. Note that all these elements enable one to generate and propagate the vector in any *E. coli* strain which is recombination deficient (recA⁻) and carries a selectable F' episome such as TOP10F'.



- If the *Xho* I site is used for cloning, the signal cleavage site (GCT CGA) must be recreated in order for the cleavage of the fusion protein to occur. This can be achieved by regenerating the *Xho* I site.

Figure 24. A map of the plasmid pHIL-S1.

DNA elements comprising the rest of the plasmid are derived from the genome of *P. pastoris*, except for the short regions of pBR322, used to link the yeast elements. The 5' AOX1 and the 3' AOX1 have DNA elements from the gene encoding the enzyme alcohol oxidase and together aid in site directed integration into the genome of a host strain such as GS115. The *P. pastoris* HIS4 gene makes complementation possible in the defective his4 gene in the host GS115 which can then grow in medium lacking the amino acid histidine. The gene PHO1 codes for acid phosphatase and provides a secretion signal sequence which targets protein expression into the medium, and it helps initiate translation of genes cloned downstream by its ATG start codon.

Briefly, the construction of PHIL-S1::CMCPI2* involved ligation of Sma I / Xho I digested PHIL-S1 and CMCPI2* DNA. The end result is the recreation of the Xho I site at the fusion junction of the gene and the PHO1 signal sequence. The recreated Xho I site acts as Kex2 like protease cleavage sequence that would enable processing of the recombinant protein during expression. In order to propagate and screen for recombinant clones, the ligated mixture was transformed into CaCl₂ competency induced TOP10F' *E. coli* cells. Clones carrying the PHIL-S1::CMCPI2* were screened for the insert by PCR screening and the results are displayed in Figure 25. 21 clones were screened; 10 were identified positive (lanes 3, 7, 8, 9, 10, 11, 15, 19, 20 and 21), 9 had no inserts and 2 clones did not give rise to a PCR product (lane 22 and 12). The control PCR product (lane 2) has a calculated length of 123bp and clones whose PCR products comigrated with it were deemed negative. Any clones which had an insert produced PCR products of high mobility and theory tells us that it should be 233bp long.

The difference in mobility of PCR products of the control and positive clones accounts for the presence of the expected DNA fragment. Two clones which were positive as per PCR screening were sequenced and the orientation of the sequence was determined to be in the right order.

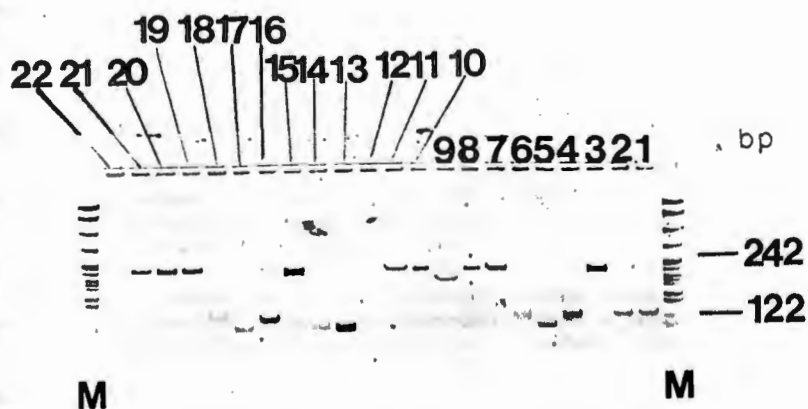


Figure 25. Non-denaturing polyacrylamide gel electrophoresis of PCR screening results.

4.4 Expression Of The Inhibitor Gene CMCP12* In *P.*

pastoris

The use of *P. pastoris* expression vectors stems from the discovery that methanol-regulated promoters, such as that controlling the alcohol oxidase I gene, are highly expressed and tightly regulated. In glucose or glycerol-grown cells, alcohol oxidase is absent but constitutes as much as 30% in of the total protein in methanol-grown cells [104]. Therefore it was postulated that a foreign gene regulated by the AOX1 promoter would be expressed at high level, even in cells that contained only a single copy of the AOX1 promoter-gene construct (expression cassette). The expression cassette is inserted into the host genome, thereby avoiding potential plasmid instability problems which are common in *S. cerevisiae* expression systems. The host harbouring the expression cassette is first grown on repressing carbon sources, such as glucose or glycerol, an unlimited mass of cells is generated without significant selection for mutants defective in foreign protein production. The cells are allowed to deplete the repressing carbon source and upon methanol addition, the expression cassette is switched on. A number of vectors for intracellular or secretion expression are supplied by Invitrogen Corporation.

The construction of GS115 strain carrying the desired expression-cassette involved digestion of PHIL-S1::CMCP12* with Bgl11. After digestion, two plasmid fragments are generated. One carries the expected expression cassette flanked by portions of the AOX1 gene. 10

μg of the digested DNA was transformed into GS115 cells which had been made competent by treatment with CsCl as laid down by the kit supplier. The strain GS115 is a histidine requiring auxotroph meaning that it is defective in histidinol dehydrogenase (*his4*); only the transformed expression cassette has the *HIS4* gene which then complements GS115. Once the transforming DNA is in the cell, it integrates into genomic DNA according to the dictates of the *AOX1* sequences. After the homologous recombination event has taken place, the entire *AOX1* in GS115 is displaced by the expression cassette. GS115 recombinants have a slow growth phenotype on medium containing methanol compared to the wild type strain. The reason being that they no longer produce alcohol oxidase I, the product of the *AOX1* gene and they depend on the poorly expressed *AOX2* gene for the utilisation of methanol. Screening for *AOX1*-disrupted transformants was done on MM plates, which are supplemented with methanol. A suitable control, GS115 (*HIS4*⁺, *Mut*⁺, albumin secreted) was also plated; *Mut*⁺ - capable of using methanol. 53 transformants were obtained; ten were found positive i.e., *HIS4*⁺ and *Mut*⁻.

To further prove the presence of *CMCPI2** gene in the genome of GS115, the recombinants were grown up for the isolation of their genomic DNA. Then the DNA was used in a PCR screening protocol utilizing primers directed against the 3' *AOX1* and the 5' *AOX1* elements. In this way we were able to amplify a DNA piece which was loaded on a 12% non-denaturing polyacrylamide gel. A control DNA PCR product obtained from wild type GS115 was also run in parallel. The results are shown in Figure 26. Lane 2 indicates a DNA band corresponding to between 309bp and 242bp in length. It is the PCR product of the

control. From the known sequence, the product should be 263bp long. Lane 4 shows a DNA band of just under 484bp; based on calculations, the experimental PCR product should be 353bp in size. The difference in size of the two bands is big enough to account for the presence of the expression cassette in GS115 genomic DNA.

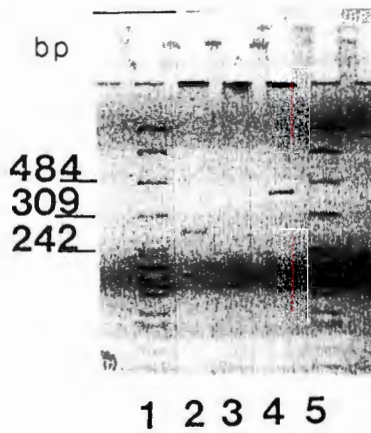


Fig. 26. PCR screening of GS115 genomic DNA for the presence of the gene fragment CMCP12'.

The next step was to establish the level of foreign protein production by the ten recombinant GS115 clones. 100 ml of BMGY (contains glycerol as a sole carbon source) medium was placed in 500 ml baffled flasks and inoculated with a small amount of cells from a different clone on MD plates. The cultures were incubated at 30°C with vigorous shaking for two days. During this step, the repressing carbon source glycerol is used up in the build-up of biomass. Thereafter, the cells were removed from the broth by centrifugation and resuspended in BMMY (supplemented with 0.5% methanol) medium and further incubated at 30°C. A time-course study after 1, 2 and 3 days post induction by removing only a portion of the supernatant for protein analysis by reverse phase HPLC was conducted. 20 µl of supernatant was made 50% in column equilibration buffer (0.1% TFA in water) and loaded on a C₁₈ analytical column. Elution followed with a gradient of buffer B (0.085% TFA in 80% acetonitrile). Based on the chromatogram in Figure 21, the approximate retention time of the inhibitor peptide was known. Outcomes of the time-course study are shown in Figure 27a. A peak after 28 minutes (retention time) appears to increase in area during the expression phase. On the other hand, uninduced clones do not seem to be producing a detectable amount of a peptide with a retention time of 28 minutes. All the different recombinants assayed did not show a marked difference in the level of protein production as estimated by the area under the 28 minutes peak. The results should not be interpreted as if only one protein is secreted into the medium by *P. pastoris* strain GS115(CMCPI2*). In fact, at very high sensitive detector setting other proteins peaks corresponding to *P. pastors* secreted proteins were observed; especially after more than two days expression induction. At some

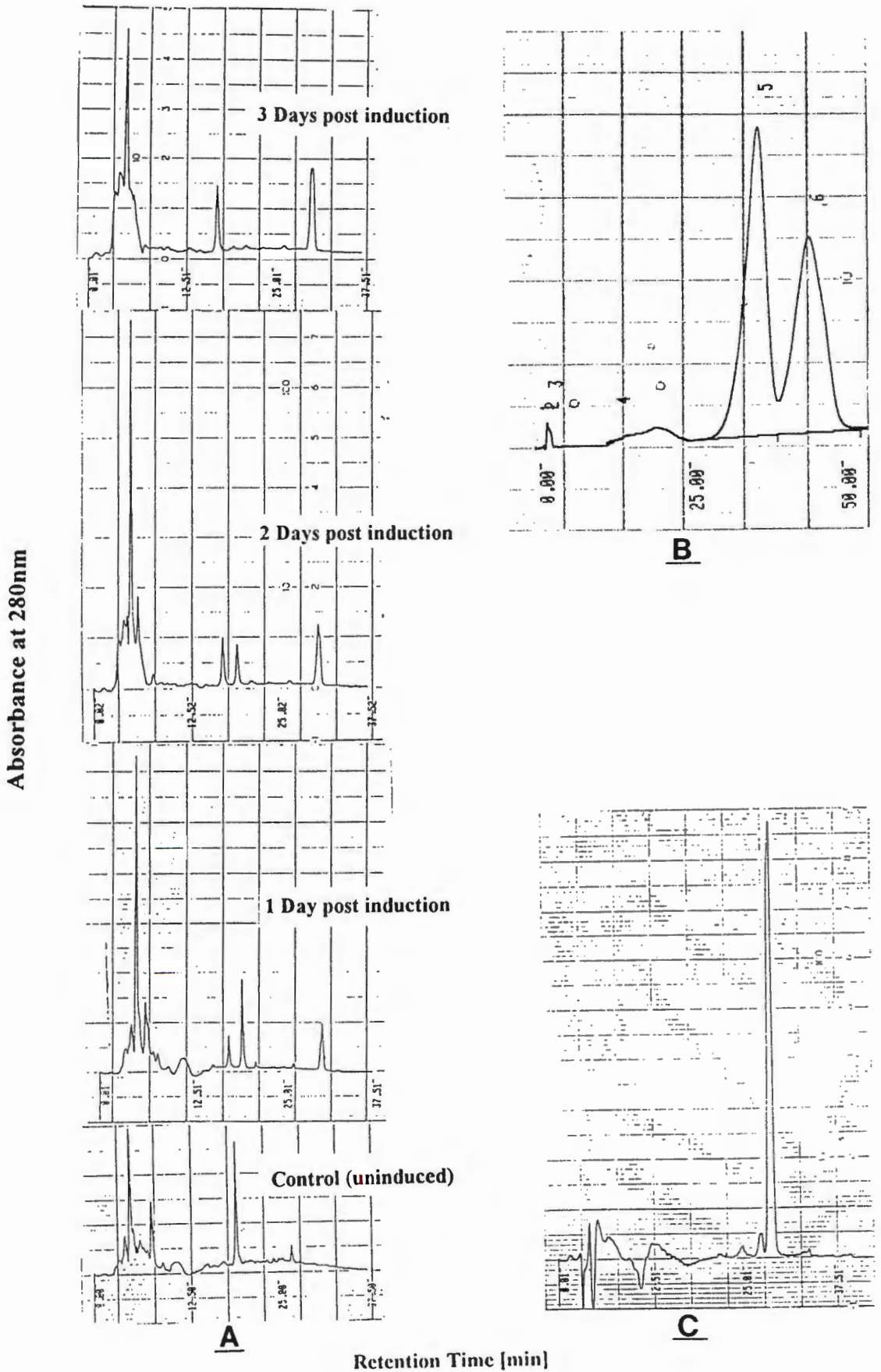


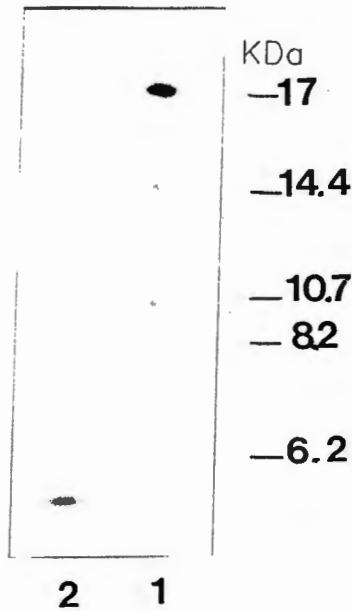
Fig 27. Analysis of supernatants for the expression of the inhibitor peptide by reverse phase HPLC (A); purification of the inhibitor protein by an-ion exchange chromatography (B); peak 5 and 6 fractions (B) were then analyzed by reverse phase HPLC, only peak 6 was rich in the inhibitor peptide (C).

instances, the 28 minutes peak started declining in area probably due to peptide degradation by acid protease's. For this reason, large scale protein expressions were carried out for not more than two days.

Supernatants derived from large scale protein productions were made to 0.06M Tris (pH 8.7), diluted by 50% in distilled water, filtered through a Whitman filter paper and then a 0.45 μ m acetate membrane before being pumped (5ml/min) on to a Super Q-650S resin packed in a stainless steel column (2.0 cm diameter; 17.5cm length). The strong anion-exchange resin had been equilibrated in buffer A (0.06M Tris/HCl, pH 8.7). At this pH, the inhibitor protein which has a calculated isoelectric point of 4.62 should be negatively charged and therefore bind on the resin. The column was then developed with a step gradient of 0.5M NaCl; Cl⁻ competes with proteins for binding sites on the resin resulting in elution. A number of peak fractions were collected but only the last one indicated to be rich in the 28 minutes peptide - see Figure 27b. To remove salt, the pooled fractions were dialyzed against 50mM Tris (pH 8.0) using dialysis tubing with a pore size cut-off of 2.00kDa. To concentrate the protein, the fractions were freeze dried. The material at this stage was not very pure because of the usage of a large column. Coloured impurities from the fermentation broth imparted a beige colour to the material. It was then essential to further purify the protein using the C₁₈ column mentioned earlier. A protein yield of 2.5 mg (estimated by Bio-Rad protein assay) per litre of BMGY was obtained.

4.4.1 Characterisation Of The Purified Inhibitor Peptide

First, it was necessary to prove homogeneity of the purified inhibitor peptide; tricine-SDS polyacrylamide gel electrophoresis followed by silver staining gave rise to one protein band migrating with a molecular mass of less than 6.2kDa (Figure 28). This is consistent with the value 3.498Da predicted from the amino acid sequence. 1 nmol of the protein was subjected to five cycles of N-terminal sequencing and the expected sequence of CMCPI2* (NH₂-R-V-{C}-P-M-; C was not determined) was confirmed.



4.4.2 Discussion

The Invitrogen *Pichia pastoris* Kit for the expression of foreign proteins has proven to be a good tool in the production of the recombinant chimeric protein inhibitor. However, shake-flask cultures have the disadvantage that cell growth conditions are not easily controlled. Aeration and medium buffering could be inadequate for high culture density. The accumulation of yeast by-products of metabolism could have influenced the yield of the final product.

The reverse phase HPLC chromatograms indicate that the secretion of the inhibitor peptide into the medium was an effective first purification step which separates product from most other cellular components.

The N-terminus of the inhibitor peptide was sequenced and the results interpreted as showing that the protein had been correctly processed by endogenous yeast Kex2 like proteases. But it still remains to be proven whether the rest of the peptide was authentically produced. Tricine-SDS polyacrylamide gel analysis of the inhibitor peptide indicate that it was purified to homogeneity.

4.4.3 Conclusion

From this study on the effectiveness of production of a foreign

inhibitor peptide in a bacterium and a yeast, it was clear that the yeast host was a better producer. However, the two expression systems can not be quantitatively compared due to differences in experimental design.

It remains to be established if the putative multispecific inhibitor peptide is in fact active. This will bring the design section to completion and will allow the hypotheses to be judged.

Part 5

Materials And Methods

5.1 Suppliers Of materials

<u>Supplier</u>	<u>Material</u>
Boehringer-Mannheim	Restriction enzymes
Pharmacia Fine Chemicals	SuRE CUT Enzyme Buffers
Promega	MonoQ HR10/10 Anion Exchange HPLC Column
Stratagene	Taq Track Sequencing Kit
DNAzyme	Taq 10x buffer
Sigma	Taq DNA polymerase
Biozyme Laboratories Ltd	10x Taq buffer
Boehringer-Mannheim	DNA polymerase
TosoHass	10x DNA polymerase buffer
BDH Laboratory Supplies	IPTG, CsCl, RNase A, X-gal, Trypsin,
Merck	Tricine, Ampicillin, Agarose, and Chymotrypsin
Macherey- Nagel, Düren	Enterokinase (EC 3.4.21.9)
Difco and Sigma	Nuclease P1 (EC3.1.30.1)
R. Chauhan (UCT)	RP HPLC Column (ODS-80TM)
Pei-Yin Ma (UCT)	Acrylamide (sequencing grade)
Jenni Poter (UCT)	All other chemicals (analytical grade)
	NUCLEOBOND Ax100 Kit
	Ingredients for yeast media
	pWM528 and JM83
	pMAL-pk, TB1
	pHIL-S1, GS115

5.2 Recipes

1) Luria Broth (LB: Luria-Bertani) [102] per litre

Tryptone	10g
Yeast extract	5g
NaCl	5g

adjust pH to 7.5 using 1M NaOH

2) Alkaline Lysis Buffers [103]

Solution I:

Glucose	50mM
EDTA	10mM
Tris/HCl pH 8	25mM

This solution is autoclaved and kept in the fridge

Solution II:

NaOH	0.2M
SDS	1%

Kept at room temperature

Solution III:

per 100ml

K Acetate	60ml 5M
Glacial acetic acid	1.5ml
water	38.5ml

3) DNase free RNase preparation

RNase A is dissolved at 20 mg/ml in 0.3M sodium acetate (pH 5.5) and aliquoted into eppendorf tubes which are incubated in boiling water for 10 min.

4) TE (Tris-EDTA) Buffer

Tris (pH 8.0)	10mM
Na ₂ EDTA	1mM

5) TER Buffer (to make 10 ml)

0.1ml {1M Tris, pH 8.0}	10mM Tris
0.02ml {0.5M EDTA, pH 8.0}	1mM EDTA
0.02ml 10mg/ml RNase A	20µg/ml RNase A

add sterile water to the 10 ml mark

6) Preparation of phenol [102]

To 500 ml of redistilled phenol add 0.5 ml of hydroxyquinoline and mix. Add 500 ml of 50 mM Tris base (pH 10.0), stir for 10 min and stand at room temperature to allow phases to separate. Remove the aqueous phase and replace it with the same amount of 50 mM Tris (pH 8.0). Repeat the process once again. The equilibrated phenol can be stored at room 4°C under 50 mM Tris pH 8.0

7) PCR product phenol-chloroform extraction

The PCR sample is made to 180 μ L with TE, in an eppendorf. 20 μ L of phenol is added and the sample is mixed well. 200 μ L of chloroform/isoamylalcohol (24:1) is added and the sample is mixed again. The eppendorf is centrifuged in a microfuge for 5min at 4°C. The top layer is carefully removed into a new eppendorf, avoiding the white interface layer. An equal volume of cold isopropanol and 1/10 volume 3M sodium acetate is added to the sample. It is mixed and incubated on ice for 5 min, after which it is centrifuged in a microfuge for 10min at 4°C. The supernatant is removed and the pellet is washed with 200 μ L 70% ethanol (centrifuge for 2 min at 4°C with tube re-orientation). The ethanol is removed, and the pellet is briefly dried in a Speedvac. The pellet is finally resuspended in TE.

8) TBE (Tris-Borate-EDTA) 10 times concentrated

Tris (pH 8.0)	0.89M
Boric acid	0.89M
EDTA	0.2M

9) Denaturation solution for bridge mutagenesis

(10 times concentrated)

Tris (pH 8.0)	50mM
KCl	100mM
MgSO ₄	50mM
DTT	5mM

10) DNA loading buffer

Sucrose is dissolved to 40% in TE and a spatula tip of Bromophenol is added.

11) Denaturation solution for bridge mutagenesis

(Two times concentrated)

KCl	20mM
Tris/HCl (pH 8.0)	10mM
MgSO ₄	10mM
DTT	1mM

11) pMAL expression system buffers**Column buffer:**

Tris/HCl (pH 7.4)	20mM
NaCl	200mM
EDTA	1mM
Sodium azide	1mM
DTT	1mM

12) Media for yeast growth, screening and induction of protein secretion**a. YPD (for growth)**

To prepare a litre; 10 g of bacto yeast extract and 20 g of peptone are dissolved in 900 ml distilled water and the mixture is autoclaved for 20 minutes. To this 100 ml of 10x dextrose (200 g of D-glucose

dissolved in 1000 ml water and autoclaved) is added. The solution is kept at room temperature.

b. MD (for screening)

To prepare MD plates; 200 ml of autoclaved water was combined with 3.75 g agar and the mixture was autoclaved. The mixture is allowed to cool to about 60°C. Thereafter, 25 ml of 10x YNB (33.5 g yeast nitrogen base lacking amino acids dissolved in 250 ml autoclaved water and then filter sterilized), 0.5 ml 500x biotin (20 mg of biotin in 100 ml water, filter sterilized) and 10x dextrose were added. The solution is then poured into sterile petri dishes.

c. MM (for screening)

MM. plates have the same composition as MD plates with the exception of the addition of methanol to 0.5%..

d. BMGY (for protein expression)

To make a litre: 10 g of yeast extract, 20 g peptone and 700 ml water were combined, then autoclaved. The solution was then allowed to cool to room temperature prior to the addition of the following ingredients; 100 ml 10x yeast nitrogen base, 2 ml 500x biotin, 100 ml 1M phosphate buffer(pH 6.0), and 100 ml 10x GY (10% glycerol, sterile)

e. BMMY (for induction of protein expression)

The composition of this medium is similar to BMGY, except that it lacks 10x GY and contains 0.5% methanol.

5.3 Methods

5.3.1 Extraction (purification) of *E. coli* plasmid DNA and *Pichia pastoris* genomic DNA

The method used for the isolation of plasmid DNA was adapted from the alkaline lysis method of Sambrook et al . [104]

A) Large scale lysis by alkali

- ◆ 10 ml LB was inoculated with a loop of the desired clone from a - 70°C stock in a McCartney bottle containing the appropriate antibiotic. Incubated overnight at 37°C with vigorous shaking.

- ◆ 500 ml of LB (containing 100 µg/ml antibiotic) in a 2L flask is then inoculated with 1 ml of the overnight culture and the mixture is incubated at 37°C overnight with shaking.
- ◆ The cells are harvested by centrifugation (20 min., 4°C) at 5000 rpm in a JA10 rotor of a BECKMAN J2-21 centrifuge.
- ◆ The supernatant is decanted and the pellet is resuspended in 10 ml solution I.
- ◆ Incubated on ice for 5 min.
- ◆ 10 ml of solution II is added, mix gently and let stand on ice for 5 min.
- ◆ To this was added 10 ml of solution III and then followed with rigorous mixing.
- ◆ Clear off precipitate by centrifugation (20 000 rpm in a JA20 rotor for 20 min., 4°C). The clear solution was aspirated off and placed in a clean tube.
- ◆ Plasmid DNA was precipitated by adding an equal volume of isopropanol and incubation at -20°C for 30 min.
- ◆ The DNA is washed with 70% ethanol, centrifuged (10 000 rpm, 4°C, 10 min.). The supernatant is decanted and the pellet is air dried briefly, and resuspended in 4 ml TE. 4.5 g CsCl was added and dissolved and 150µl EtBr (10mg/ml) added.
- ◆ The preparation was centrifuged at 55 000 rpm at 20°C for 16 hrs. The resultant plasmid band was extracted and the EtBr removed through phase separation using isoamylalcohol. CsCl was removed by dialysis against TE buffer for 3 hrs.

To the mixture, 3 volumes of ethanol are added followed by incubation at -20°C for 1 hr. The DNA pellet is collected by centrifugation (20

000 rpm, 10 min., JA-20 rotor), washed in 70% ethanol, and air dried. The pellet is finally dissolved in 1 ml TE.

B) One step 'miniprep' method for the isolation of plasmid DNA

The method used was as reported by Kamal Chowdhury [105]

- ◊ Take 0.5 ml of overnight *E. Coli* culture in an eppendorf.
- ◊ Add 0.5 ml of phenol:chloroform:isoamylalcohol (25:24:1)
- ◊ Mix by vortexing at maximum speed for 1 min.
- ◊ Centrifuge at 12 000 rpm for 5 min. in a microfuge. During the spin prepare eppendorfs with isopropanol. After the centrifugation, carefully remove 0.45 ml of the upper aqueous phase leaving the interface undisturbed and add it to the isopropanol. Mix well and spin immediately at 12 000 rpm for 5 min.
- ◊ Pour off the supernatant, add carefully 0.5 ml of 70 % ethanol to the side of the tube and pour off. Repeat the washing once more. Vacuum dry the pellet and suspend in 20 to 100 microliters of TER buffer. About 5-10 microliters of this DNA can be cleaved with appropriate restriction enzyme(s) for analysis.

C) Nucleobond AX PC-Kit 100 for the purification of plasmids

The procedure suggested by the supplier was followed. Buffers were also provided by the supplier.

- Volume of E. Coli overnight culture; 100ml.
- The provided cartridge is equilibrated with buffer N2 (2ml)
- Bacterial cells are collected by centrifugation and the resultant pellet is resuspended carefully in buffer S1 (4 ml).
- 4 ml of buffer S2 is added. The suspension is mixed gently and incubated at room temperature for 5 minutes.
- Buffer S3(4ml) is added, the suspension is mixed gently by inverting the tube 6-8 times until a homogenous suspension is formed. The suspension is incubated at on ice for 5-10 minutes and centrifuged at high speed (>10 000 rpm) at 4°C for 38 minutes.
- The supernatant is removed carefully from the white precipitate and loaded on the NUCLEOBOND AX cartridge, equilibrated with buffer N2.
- The cartridge is washed with 2 times 4 ml buffer N3.
- The plasmid DNA is eluted with buffer N5 (2ml).
- The purified plasmid DNA is precipitated with 0.8 volumes of isopropanol, preequilibrated to room temperature, and centrifuged at high speed (>10 000 rpm) at 4°C. The DNA is washed in 70% ethanol, briefly air dried and redissolved in appropriate amount of TE buffer.

D) Isolation of *P. Pastoris* genomic DNA

The method used in this work was first reported by Hoffman, C.S. and Winston, F. [104a]

- * Grow 10 ml yeast cultures in YPD medium to saturation.
- * Harvest cells for 5 min at 5 000 rpm and resuspend in 0.5 ml water.
- * Decant the supernatant and briefly vortex the tube to resuspend the pellet in the residual liquid.
- * Add 200 μ l of the following reaction mixture (25 Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA), and vortex briefly.
- * Add 0.2 ml phenol:Chloroform:isoamylalcohol (25:24:1). Then add 0.3 g acid-washed glass beads and vortex for three minutes.
- * Add 0.2 ml TE and spin for 5 min. at 13 000 rpm. Transfer aqueous supernatant into a new tube and phenol/chloroform this mixture two times. Precipitate the DNA by adding 1.0 ml of absolute ethanol and mix.
- * Centrifuge for 2 min. At 13 000 rpm, then resuspend pellet in 0.4 ml TE buffer.
- * Add 10 μ l 4M ammonium acetate (pH 5.5) plus 1.0 ml ethanol and invert tube to mix. Spin for 2 min., dry pellet and resuspend in 50 μ l TE.

5.3.2 Plasmid Digests

High, medium and low salts buffers supplied as a 10 × stock were used for most digests. Between 1 and 2 µg of plasmid DNA were digested in volume of 10 to 20 µl with 1 to 3 units of restriction enzyme per microgram of DNA. Digests were performed at the optimal temperature for the particular enzyme for 2 hrs or overnight. Reaction mixtures were incubated in a water bath set at the required temperature. Note that digests with Nuclease P1 were carried out for 25 min. in 0.02 M EDTA at 37°C.

In the case of double digest, a common buffer in which both enzymes are active was chosen and the digest was done overnight.

5.3.3 Gel electrophoresis

1) Agarose

DNA preparations were routinely analyzed by agarose gel electrophoresis, using a flatbed apparatus and 0.8, 1.0 and 3.0 % agarose in TBE buffer (pH 8.0) containing EtBr at 0.5 µg/ml. The running buffer (1 × TBE) in which the gel was submerged also contained EtBr at the same concentration. Gel loading buffer containing sucrose and bromophenol blue as a tracking dye was added to the samples before loading, and electrophoresis was carried out at 4-8 volts/cm for 2-5 hrs.

2) SDS Polyacrylamide gel electrophoresis (SDS PAGE)

Periplasmic extracts and purified fusion proteins were electrophoresed on 13%/14% SDS PAGE. 15 cm long gels were run consisting of a ± 2 cm stacking gel and 12 cm of separating gel. Samples were dissolved by boiling for 3 minutes in an appropriate volume of SDS sample buffer containing 20 μ l/ml of 2-mercaptoethanol. Gels were run at 10 mA until the bromophenol blue of the loading buffer had entered the stacking gel then current increased to 18 mA, and then ran until the bromophenol had just run off the bottom of the gel.

3) Purification Of Small DNA Fragments From Non-

Denaturing Polyacrylamide Gels

- Double digestions of plasmids carrying insert DNA for subcloning were loaded on 12% 15 cm long polyacrylamide gels. Usually, 5 μ g DNA was loaded per well and electrophoresis was done in vertical set ups using TBE at one times the strength as a running buffer. The electrophoresis was conducted for 2-3 hrs at 2 V/cm. Samples were made 10% in sucrose before loading and a separate lane was loaded with loading buffer containing bromophenol blue as a tracer.
- After the electrophoresis, the gel plates were detached from the tank and the plates were separated ensuring that the gel was left

on one plate. The attached gel was stained for 20 minutes in 0.5 $\mu\text{g/ml}$ EtBr

- Under a longwave UV light lamp, the DNA bands were cut out with a razor blade.
- The gel pieces were placed in 1.5 ml tubes which contained elution buffer (0.5M ammonium acetate, 1mM EDTA), sufficient to cover the gel pieces.
- To elute the DNA, the tubes are incubated at 37°C with overnight shaking.
- Supernatants were pipetted off and DNA fragments were precipitated with 2 volumes of ethanol, chilled at -20°C for 30 min. The fragments were recovered by centrifugation (12 000rpm, 4°C, 10 min.)
- Pellets were rinsed in 70% ethanol, dried and resuspended in TE buffer.

4) Tricine-SDS PAGE

Small proteins were dissolved in SDS loading buffer and loaded on 14% gels including a stacking gel as described by Herman Schägger and Gebhard von Jagow [106] and the protein bands were stained by silver staining.

5.3.4 Cloning

1) Ligation

100 µg of digested vector DNA was added to 2x equimolar, 3x equimolar and 4x equimolar insert DNA. Water was added to 8 µl and 1 µl of 10x bacteriophage T4 DNA ligase buffer was added. Then 5 units of T4 DNA ligase was added. The reaction was incubated at 4°C overnight.

2) Transformations

a. **Competence induction.** Fresh overnight culture in LB from a single colony (JM83 or TB1) off a fresh plate was grown. 1 ml of this was added to 100 ml LB and grown to $A_{600} \pm 0.5$. The cells were incubated on ice for 15 minutes. 40 ml of this was centrifuged at 4 000 rpm in a JA20 rotor of a Beckman J2-21 centrifuge at 4°C for 5 min. And immediately dissolved in 8 ml of 0.1 M CaCl_2 at 4°C. This was recentrifuged as above and resuspended in and the same solution and kept on ice for 45 minutes then harvested as before and redissolved in 1.6 ml of the same calcium chloride solution, then kept on ice.

b. **Transformation.** The ligation mixture (5µl) was made to 100 µl with water and incubated on ice in an eppendorf. To this was added 200 µl competent cells and stood on ice for 15 minutes. After which, it was directly transferred to a water bath at 42°C for 3 minutes, then back to ice for 5 ml. To this was added 0.8 ml LB and incubated at 37°C with shaking for 30 minutes. The cells were spun down and resuspended in 100 µl LB of which all was plated on LB plates containing 100

µg/ml ampicillin. The plates were kept at 37°C in an upside-down position overnight to allow growth of recombinants.

5.3.5 DNA sequencing

DNA was sequenced using the Taq Track^{RM} Sequencing System supplied by Promega according to the manufacturers instructions. In all cases double stranded DNA was used as a template. 10 µg of template DNA (CsCl purified or Nucleobond purified) was diluted to 18 µl with sterile double distilled water and 2 µl of 2 M NaOH added. This was incubated at 37°C for 6 min. The solution was neutralized by the addition of 4 µl sodium acetate (pH 5.5) and the DNA precipitated by the addition of 150 µl ice cold absolute ethanol. It was then centrifuged for 20 min in a microfuge at 4°C. The resultant pellet was washed in 70% ethanol and briefly air dried. The DNA was then ready for priming for which the manufacturer's protocol was used. The sequencing of a particular DNA template made use of forward and reverse primers specific for that particular vector.

Products were electrophoresed at 86W on a 6% polyacrylamide sequencing gel, that was kept at 42°C using a water jacket. The gel was dried and placed in an autoradiograph cassette with a sensitized X-ray film. The autoradiograph was exposed overnight at -70°C, and then developed.

5.3.6 Oligonucleotide Synthesis

a. Synthesis. Oligonucleotides were synthesized with nucleoside-o-(2-cyanoethyl)-N,N-diisopropyl phosphoramidites. These and other reagents for synthesis were from Biosearch. Synthesis was performed on a model 6500 AutogenTM DNA synthesizer in a glass column containing 1 μ mole (for the long synthesis oligos) and 0.2 μ mole (for short oligos such as primers) of an appropriate 5'-O-dimethoxytrityl deoxynucleoside-3'-O-succinyl-derivatized 1000 Angstrom control pore glass support at a loading of 18 μ mol/g.

b. Preparation of oligonucleotides for purification by anion exchange chromatography

The synthesis products are first removed from the support medium by cleavage in concentrated ammonia; add 1 ml 32% ammonia and allow to stand at room temperature for 2 hrs. Centrifuge in a microfuge and remove the supernatant into a fresh eppendorf tube. Seal the tube tightly and incubate at +70°C for 1 hr. Precipitate the oligonucleotide in butanol (1ml butanol for 100 μ l oligo).

c. Purification of oligonucleotides

The precipitated oligonucleotide is resuspended in buffer A (50mM LiCl, 10mM NaOH, pH 12.0). The sample was run over a MonoQ HR10/10 ion exchange HPLC column. A LiCl concentration gradient (50mM to 3M) was used to elute the oligonucleotides. Fractions containing the desired oligonucleotide were detritylated to remove the DMTr group.

The desired HPLC samples were precipitated overnight at -70°C with 6 volumes of acetone/ethanol (3:1). The sample was centrifuged at 20 000 rpm, at 4°C . The pellet was resuspended in 100 μl water and reprecipitated with 9 volumes of butanol. 500 μl cold, concentrated ammonia was added to the pellet and then evaporated in a Speedvac. The pellet was dissolved in 100 μl 80% acetic acid and left at room temperature for 1 hr. The DNA was precipitated with 9 volumes of butanol and resuspended in TE buffer.

5.3.7 Bridge Mutagenesis Synthesis Of A Gene

Three aliquots of 50 ng of plasmid linearized with *Sma*I restriction enzyme was mixed with complementary oligonucleotide at 10 fold, 100 fold and 1000 fold molar excesses. To this was added 15 μl of denaturation buffer (section 5.2) and the total volume was made up to 30 μl with sterile water. The mixture was boiled for 3 min. And allowed to cool at room temperature for 5 minutes. The cooled mixture was centrifuged to collect condensate and placed on ice. To each tube, 100 μl competent JM83 [*ara*, Δ (*lac-pro*), *strA*, *thi*, Φ 80d *lacZ* M15] cells were added. Transformed cells were plated on LB containing 100 $\mu\text{g}/\text{ml}$ Of ampicillin and 20 $\mu\text{g}/\text{ml}$ X-gal. Transformants were selected according to the appropriate change in colour from that of the parental on the same medium.

5.3.8 Gene Expression

Bacterial strains containing the gene to be expressed were grown at 37°C with shaking to A_{600} 0.3-0.6 in LB with ampicillin. Cells were

harvested by centrifugation (4 000 rpm, JA20 rotor) under sterile conditions and resuspended in fresh sterile LB at 37°C, containing ampicillin (100 µg/ml) and then induced by adding IPTG to a final concentration of 0.3 mM and shaking for a further 3 hrs at 37°C.

The cells were pelleted by centrifugation (4 000 rpm, 10min, 4°C), and resuspended in 30 mM TrisHCl, 20% sucrose, pH 8.0 (8 ml for each 0.1 g cells wet weight). Incubate for ten minutes at room temperature. Cells were then centrifuged (8 000 rpm, 4°C, 10 min.) and dissolved in ice cold 5 mM MgSO₄ (8 ml for each 0.1 g cells wet weight). The suspension was mixed gently for 10 min. in an iced water bath. After that, it was centrifuged as above and the fusion protein was isolated from the resultant supernatant.

5.3.9 The Maltose Binding Protein Purification System

Two 500 ml cultures of E. Coli TB1(CMCP12^{*}) were grown to mid log phase and induced with IPTG. After 3 hrs post induction, the cells were harvested and were subjected to cold osmotic shock as mentioned above. The supernatant was diluted by half in column buffer, then loaded to a cross-linked amylose column which had been equilibrated in column buffer. A washing step with 100 ml of column buffer was carried out, followed by elution of the fusion protein with 10 mM maltose. Approximately 12 fractions were collected, dialyzed and concentrated by freeze drying. The fusion protein was then electrophoresed on a 13% SDS polyacrylamide gel.

5.3.10 Fusion Protein Digestion With Enterokinase

The protein preparation was made 50mM Tris, pH 7.5 and 25 µg was incubated with 0.6 µg enzyme at 37°C for 24 hrs.

5.3.11 PCR Screening Of Transformants

a) **E. Coli colonies.** Colonies were picked from the plates into separate test tubes containing 10 ml LB broth and grown overnight at 37°C with shaking. A $1/20$ dilution was made into 100 µl of water and boiled for 5 min. 25 ml of the boiled sample was added to an equal volume of Stratagene 1× PCR buffer (10 mM Tris/HCL, pH 8.8, 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin) containing 0.2 µM of each dNTP, 20 pmol of each plasmid primer, and 0.005% Tween 20. 1 unit of Stratagene Taq polymerase was added to the combined sample, which was overlaid with 30 µl paraffin oil. 30 cycles of 10 sec at 92°C, 30 sec at 55°C and 60 sec at 72°C were performed, with an additional 5 min. Extension at 72°C after the last cycle. Products were visualized by agarose gel electrophoresis on a 3% gel containing TBE buffer and EtBr, or by PAGE on a 12% gel and visualized after staining in EtBr.

B) **Pichia pastoris clones.** Total genomic DNA of recombinants was used in PCR; 0.5 µg of template DNA in a volume of 25 µl was added to the PCR buffer mentioned above. The PCR reaction and visualization of products was carried out as mentioned.

5.3.12 PCR Amplification

1 ng of template DNA was added to 5 μ l 10x Dynazyme buffer containing 0.2 μ M of each dNTP, 10 pmol of each primer and the volume was made to 50 μ l with sterile water. 2.5 units of Dynazyme DNA polymerase was added to the sample which was overlaid with 30 μ l of paraffin oil. 30 cycles of 20 sec At 94°C, 30 sec at 50°C, and 50 sec at 72°C were performed, with an additional 180 sec extension at 72°C after the last cycle.

5.3.12 Amino Acid Sequence Analysis

N-terminal sequence analysis was performed on a gas-liquid solid phase sequencer constructed by Hewick et al. [107] and slightly modified as described by Brant et al. [108]. The reagents, solvents and degradation cycle used have been published in the above references. The converted PTH amino acids were identified by an isocratic HPLC system on a 3 x 250mm 3 μ Lichrospher C₁₈ (Bischoff) column as described by Lottspeich [107b].

5.3.13 Amino Acid Composition Analysis

Gas phase acid hydrolysis was performed on sample of 5-10 nmols in acid washed, pyrolysed test tubes. 5.7 M HCL and 0.1% phenol was utilized in an evacuated vessel. The hydrolysis was carried out at 110°C for 24 hrs.

A standard mixture of 10 nmols of free amino acids was hydrolyzed together with the samples and used for quantitation. Norleucine was added as an internal standard

For detection purposes, the hydrolysates were derivatised using the Dabsyl method [108, 108a], and separated by reverse phase HPLC.

Literature Cited

1. Bode, W., Greyling, H.J., Huber, R., Otlewski, J. And Wilusz, T. (1989) *FEBBS LETTERS* 242, 285-292
2. Le-Nguyen, D., Mattras, H., Coletti-Previero, M-A. And Castro, B. (1989) *Biochem. Biophys. Res. Comm.* 162: 1425-1430
3. Laskowski, M. And Kato, I. (1980) *Ann. Rev. Biochem.* 49: 593-626
4. Liener, I.E. (1979) *Rev. Int. Biochem.* 27: 97-122
5. Richardson, M. (1977) *Phytochemistry (Oxf.)* 16: 159-169
6. Ryan, C.A. (1981) *Proteinase Inhibitors. In The Biochemistry of Plants, A Comprehensive Treatise* 6, (ed. P.K Stumpf and E.E. Conn.), pp. 351- 371. Academic Press New York
7. Pearce, G., Liljegren, D. And Ryan, C.A. (1987) *Proteinase Inhibitors In Wild Tomato Species. In Tomato Biotechnology* (ed. D.J. Nevins and R.A Jones) pp. 139- 144. Alan R. Liss, New York.
8. Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* 27: 157-162
9. Hill, R.E., Shaw, P.H, Boyd, P.A., Baumann, H. And Hastie, N.D. (1984) *Nature* 311: 175-177
- 9a. Hill, R.E. and Hastie, N. D. (1987) *Nature*, 326: 96-99
10. Hass, G.M. and Ryan, C.A. (1987) *Methods In Enzymol* 80: 778-791
11. Pearce, G., McGinnis, J. And Ryan, C.A. (1983) *Proc. Soc. Exp. Biol. Med.* 173: 447- 453
12. Kay, J., Valler, M.J. and Dun, B.M. (1983) In "Proteinase Inhibitors: Medical and Biological Aspects," (ed. By Katunuma, N., Umezawa, H. And Holzer, H.). Japan Sci. Soc. Press, Tokyo/Springer-Verlag, Berlin
13. Barrett, A.J. (1987) *TIBS* 12: 193-196
14. Higashiyama, S. Ohkubo, I. Ishiguro, H., Kunimatsu, M., and Sawaki, M. (1986) *Biochemistry* 25: 1669-1675
15. Korant, B.D., Chow, N., Lively, M. And Powers, J. (1979) *Proc. Natl. Acad. Sci. USA* 76: 2992-2995

16. Ryan, C.A (1989) *BioEssays* 10: 20-24
17. Travis, J. And Salvesen G.S.(1983) *Ann. Rev. Biochem.* 52: 655
-709
18. Seemüller, U., Fritz, H. And Eulitz, M. (1981) *Methods In
Enzymol.*80: 804-816
19. Correll, R.W. and Travis, J. (1985) *TIBS* 10: 20-24
20. Abraham, C.R. and Potter, H. (1989) *Biochemistry* 7: 147-153
21. Bewley, J.D. and Black, M. (1986) "Seed: Physiology of
Development and Germination", Plenum Press, New YORK.
- 21a. Khan, A.A. (1977) "The Physiology and Biochemistry of Seed
Dormancy and Germination", North Holland Publishing
Company.
- 21b. Ryan, C.A. (1973) *Ann. Rev. Plant Physiol.* 24: 173-196
22. Eigenbrot, C., Randal, M., Kossiakoff, A.A. (1990) *Protein
Eng.* 3: 591-598
- 21c. Rydel, T.J., Tulinsky, A., Bode, W. & Hubber, R. (1992) *J. Mol.
Biol.* 221, 583-601
- 22a. Papamokos, E., Wober, E., Bode, W., Huber, R., Empie,
M.W., Kato, I. And Laskowski, M. (1982) *J. Mol. Biol.* 158:
515-537
23. Wagner, G., Braun, W., Havel, T.F., Schauman, T. Go, N. And
Wüthrich, K. (1987) *J.mol. Biol.* 196: 611-639
24. Greenblatt, H.M., Ryan, C.A., and James, M.N.G. (1989) *J. Mol.
Biol.* 205: 201-228
25. Xavier-Filho, J. And Campos, F.A.P. (1989) *Proteinase
Inhibitors. In Toxicants of Plant Origins, (ed. P.R.
Cheeke) Vol. III. Boca Raton, F. CRC Press In Press*
26. Ryan, C.A. (1990) *Ann. Rev. Phytopathol.* 28: 425-449
27. Scheeman, B.O., Chaing, I., Smith, L.B. and Lymman, R.L.
(1977) *J. Nutr.* 107: 281-288
28. Lu, L. Lonie, D.S., May. D. and Ow-yang, C. (1987)
Dig. Dis. Sci. 32: 1175
29. Baker, J.E., Woo, S.M., and Mullen, M.A. (1984)
Entomol. Exp. Appl. 36: 97-105
30. Appleham, S.W. (1985) *Biochemistry of digestion. In
comprehensive Insect Physiology, Biochemistry and
Pharmacology (ed. G.A. Kerkut and L.I. Gilbert) 4:279-311.
New York Pergamon*

31. Broadway, R.M. And Duffey, S.S. (1986) *J. Insect. Physiol.* 32: 673-680
32. Metcalf, R.L. (1986) The ecology of insecticides and the chemical control of insects. In ecological theory and intergrated pest management (ed. M. Kogan) pp. 251-297. New York: John Wiley and Sons.
33. Dirham, B. (1993) The pesticides Hazard. London: The Pesticide Trust.
34. Neurath, H. (1984) *Science* 224: 350-363
35. Bond, J. And Butler, P.E. (1987) *Ann. Rev. Biochem.* 56: 333-364
36. McKey, D. (1979) The distribution of secondary compounds within plants. Rosenthal and Janzen (1979), pp. 55-133
37. Bowles, D.J. (1990) *Ann. Rev. Biochem.* 59: 873-907
38. Ryan, C.A. (1988) *Biochemistry* 27: 8879-8883
39. Bostock, R.M. Kuc, J.A. and Laine, R.A. (1981) *Science* 212: 67-69
40. Davies, E. (1987) Plant Response to Wounding. In the biochemistry of plants, D.D. Davies, pp. 43-364. New York: Academic Press
41. Pena-Cortes, H., Sanchez-Serrano, J.J., Mertens, R. Wilmitzer, L. And Prat, S. (1989) *Proc. Natl. Acad. Sci. USA* 87: 9851-9855
42. Schmidt, W.E. and Ebel, J. (1987) *Proc. Natl. Acad. Sci. USA* 84: 4117-41121
43. Thornburg, R.W. An, G. Cleveland, T.E., Johnson, R. And Ryan, C.A. (1987) *Proc. Natl. Acad. Sci. USA* 84: 744-748
44. Pearce, G., Strydom, D., Johnson, S. And Ryan. C.A. (1991) *Science* 253: 895-898
45. McGurl, B., Pearce, G., Orozco-Cardenas, M. And Ryan, C.A. (1992) *Science* 255: 1570-1573
46. Ryan, C.A. (1989) *BioEssays* 10: 20-24
47. Broadway, Q.M. (1989) *J. Chem. Eco.* 15: 2101-2113
48. Shukle, R.H., Murdock, L.L. (1983) *Environ. Entomol.* 12: 787-791
49. Christeller, J.T., Laing, W.A., Markwick, N.P. and Burges, E.P.J. (1992) *Insect. Biochem. Molec. Biol.* 22: 735-746
50. Gatehouse, A.M.R. and Boulter, D. (1983) *J. Sci. Food Agric.* 34:

345-3450

51. Murdock, L.L, Brookhart, G., Dun, P.E., Foard, D.E., Kelley, S. et al. (1987) *Comp. Biochem. Physiol.* **B87**: 783-787
52. Lipke, H., Fraenkel, G.S. and Liener, I. (1954) *Food. Chem.* **2**: 410-411
53. Birk, Y. And Applebaum, Y. (1960) *Enzymologia* **22**: 318-326
54. Birk, Y. (1985) The Bowman-Birk Inhibitor. *Int. J. Pept. Protein Res.* **25**: 113-131
55. Steffens, R., Fox, F.R. and Kassel, B. (1978) *J. Agric. Food Chem.* **26**: 170-175
56. Spates, G.E. and Harris, R.L. (1984) *Southwest. Entomol.* **4**: 399-403
57. Deloach, J.R. and Spates, G.E (1980) *J. Eco. Entomol.* **73**: 590-594
58. Hilder, V.A., Gatehouse, A.M.R., Sheerman, S.E., Barker, R.F. and Boulter, D. (1987) *Nature* **330**: 160-163
59. Johnson, R., Narvaez, J., An, G. And Ryan, C.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**: 9871-9875
60. McIntosh, S.C., Kishore, G.M., Perlate, F.J., Marrone, P.G., Stone, T.B., Sims, S.R. and Fuchs, R.L. (1990) *J. Agric. Food Chem.* **36**: 1145-1152
61. Boulter, D., Edwards, G.A., Gatehouse, A.M.R., Gatehouse, J.A. and Hilder, V.A. (1990) *Crop Protect.* **9**: 351-354
62. Peng, J.H. and Black, L.L. (1976) *Phytopathology* **66**: 958-963
63. Roby, D., Toppan, A. And Esquere-Tugaye, M-T. (1987) *Physiol. Mol. Plant Physiol.* **30**: 453-460
64. Rickaner, M., Fournier, F. And Esquerre- Tugaye, M-T. (189) *Physiol.* **90**: 1065-1070
65. Mosolov, V.V., Longinova, M.D., Fedurkina, N.V. and Benken, I.I. (1976) *Plant Sci. Lett.* **7**: 77-80
66. Mosolov, V.V., Longinova, M.D., Malova, E.L. and Benken, I.I. (1979) *Planta* **144**: 265-269
67. Senser, F., Belitz, H.D., Kaiser, K.P. and Santarius, K. (1974) *Z. Lebensum, Anters. Forsch.* **155**: 100-101
68. Brown, A.E. and Adikaram, N.D.B. (1983) *Phytopathol. Z.* **106**: 239-251

69. Garcia-Olmedo, F., Salcedo, G., Sanchez-Monge, R., Gomez, L., Royo, J. And Cabonero, P. (1987) *Oxf. Surv. Plant. Mol. Cell Biol.* 4: 275-334
70. Liang, C., Brookhart, G., Feng, G.H, Reeck, G.R. and Kramer, K.J. (1991) *FEBBS Lett.* 278: 137-142.
71. Bode, W., Greyling, H.J., Huber, R., Meyer, E., Otlewski, J. And Wilusz, T. (1989) *FEBBS Lett.* 242: 285-292
72. Holak , T.A., Wilusz, T. And Otlewski, J. (1989) *J. Mol. Biol.* 210: 635-648
73. Clore, G.M., Gronenborn, A.M., Nilges, M. And Ryan, C.A. *Biochemistry* 26: 8012-8023
74. Rees, D.C and Lipscomb, W.N. (1982) *J. Mol. Biol.* 160: 475-498
75. Blundell, T.C. Sibanda, B.L., Sternberg, M. And Thornton, J.M. (1987) *Nature* 326: 374-352
76. Hass, G.M. and Hermodson, M.A. (1981) *Biochemistry* 20: 2256-2260
77. McWerter W., Horst W.F., Campbell, E.J. and Glover, G.I. (1989) *Biochemistry* 28: 5708-5714
78. Grosjean, H. And Fiers (1982) *Gene* 19: 199-209
79. Ken-nosuke W., Wada, Y., Doi, H., Ishibashi, F., Gojobori, T. And Ikemura, T. (1991) *Nucleic Acids Research* 19: Supplement, 1981-1986
80. Robinson, M., Lilley, R., Little, S., Emtage, J.S., Yarraton, G., Stephens, P., Millican, A., Eaton, M. And Humphreys, G. (1984) *Nucl. Acid Res.* 12: 6663-6671
81. Swanfield et al. (1987) *Prot. Eng.* 1: 228
82. Mandecki, W. (1986) *Proc. Natl. Acad. Sci. USA* 83: 7177-7181
83. Mandecki, W., Hayden, M.A., Shallcross, M.A. and Stotland, E. (1990) *Gene* 94: 103-107
84. Kornberg, A. (1987) *DNA replication* (Freeman San Franscisco), 138-139
85. Bonner, D.M. (1961) *Heredity* (Pretice Hall, Englewood Cliffs, N.J.), pp. 54-57
86. Furnichi, Y. And Miura, K. (1975) *Nature* 253: 374
87. Duplay, P., Fowler, A. Bedouelle, H., Zabin, I., Sourin, W. And Hofnung, M. (1984) *J. Biol. Chem.* 259: 10606-10613
88. Spurlino, J.C., Guang-Ying Lu, and Quicho, F.A. (1991) *J.*

Biological Chemistry 266: 5202-5219

89. Keil, B. (1971) In the enzymes, (Boyer, P.D., ed.) vol. 3: 249-275, Academic Press.
90. Light, A., Savitri, H.S. and Liepnieks, J.J (1980)
Anal. Biochemistry 106: 199-206
91. Veenhuis, M., van Dijken, J.P and Harder, W. (1983)
Adv. Microb. Physiol. 24: 1-82
92. Da Silva, N.A. and Bailey, J.E. (1991) *Biotechnology and Bioengineering* 37: 309-317
93. De Nobel, J.G. and Barnett, J.A (1991) *Yeast* 7: 313-323
94. Gluschkof, P. And Fuller, R.S. (1994) *EMBO J.* 13: 154
95. LaMantia, M. And Lennarz, W.J. (1993) *Cell* 74: 899-908
96. Hwang, C., Sinskey, A.J. and Lodish, H.F. (1992) *Science* 257: 1496-1502
97. Gallagher, P.J., Hennenberg, J.M., Sambrook, J.F. and Gething, M -J. H. (1992) *J. Virol.* 66: 7136-7145
98. Hirsch, H.H., Rendules, P.S. and Wolf, D.H. (1989) *Yeast Saccharomyces cerevisiae* proteinases: structure characterization and function , pp. 134-200. In : Molecular and cell biology of yeast. Walton, E.F. and Yarranton, G.T. (eds.). Blackie and Van Nostrand Reinhold, New York
99. Laroche, Y., Storme, V., DeMentter, J., Messens, J. And Lauwereys, M. (1994) *Biotechnology* 12: 1119-1124
100. Thill, G.P., Davis, G.R., Stillman, C., Brierley, R., Engel, M., Buckholz, R., Kenney, J., Proviw, S., Vedvick, T. And Siegel, R.S. (1990) Positive and negative effects of multicopy intergrated expression vectors on protein expression in *Pichia pastoris*. pp. 477-490. In: Proceedings of the sixth international symposium on genetics of microorganisms, vol. II. Heslot, H. et al. (eds.). Societe Francaise de Microbiologie Paris
101. Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, A. And Gingeras, T. (1987) *Nucleic Acids Research* 15: 3859-3876
102. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G. and Sruhl, K. (1987) *Current protocols in molecular biology*. John Wiley and Sons, New York
103. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning, A laboratory Manual* 1st Edition. CSH Laboratory Press
104. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular*

Cloning, A Laboratory Manual 2nd Edition. CSH Laboratory Press.

- 104a. Hoffman, C.S. and Winston, F. (1987) *Gene* 57: 267-272
105. Kamal Chowdhury (1991) *Nucleic Acids Res.* 19: 2792
106. Schägger, H. And von Jagow, G. (1987) *Anal. Biochemistry* 166: 368-379
107. Hewick, R.M, Hunkapillar, M.W., Hood, L.E. and Dreyer, W.J (1981) *J. Biol. Chem.* 256: 7990-7997
- 107a. Brant, W.F., Chauhan, M. and von Holt, C. (1984) *FEBBS Lett* 174: 228-232
- 107b. Lottspeich, F. (1985) *J. Chromatog.* 326: 321-327
108. Knecht, R and Chang, J. *Analytical Chem.* 58: 2375-2379
- 108a. Chang, J. and Knecht, R (1981) *Biochem. J.* 199: 547-555