Biosynthesis of *Cucurbita Maxima* Trypsin Inhibitor I in the Methylotrophic Yeast *Pichia pastoris*

Thesis submitted for the completion of the degree of Master of Science.
September 1996.

Jennifer Hüsler
Department of Biochemistry
University of Cape Town
Republic of South Africa

Acknowledgments

I would like to thank my supervisors Prof. H. H. Klump, Prof. W. F. Brandt and Dr. D. Maeder for their support, and the Foundation of Research and Development for funding this research. My thanks also to Prof. J. Otlewski of the University of Wraclaw for providing the native CMTI I inhibitor for this study. Last, but by no means least, my thanks to Dr Paul Hüsler for his input and patience.

Abstract

Squash inhibitors are the smallest natural serine protease inhibitors. Their compact, rigid nature has enabled detailed examination of their 3D structure by NMR and X-ray crystallography. Being of a convenient size to synthesise chemically, the effects on activity of selective substitutions and deletions within the sequence have also been investigated. Thus, this family of inhibitors is considered useful as a model system for the study of protein-protein interactions.

Cucurbita maxima trypsin inhibitor I (CMTI I) may be thought of as representative of the squash inhibitors, for which there is detailed structural and functional information available. It is a 29 amino acid protein, with the tri-disulphide bridging pattern common to all squash inhibitors.

There are only a few examples of squash inhibitors being produced by recombinant DNA technology. As this technique offers a relatively cheap way of producing large amounts of these proteins, further investigation is required. Problems have been experienced with the expression of disulphide-rich proteins in *E. coli*, as the cytosol of this microorganism is not conducive to their folding. Furthermore extraction of these proteins from the periplasmic space is often required, resulting in a reduction in yield. To overcome these shortcomings, the methylotrophic yeast *Pichia pastoris* was investigated as an alternative means of expression, although at the inception of this work, no disulphide-rich proteins of this size had been expressed in *P. pastoris*. It was a challenge to investigate the feasibility of producing squash inhibitors in this expression host and to compare the activity of the recombinant inhibitor to that of native CMTI I.

An oligonucleotide coding for the sense strand of the CMTI I sequence was chemically synthesised. PCR amplification was used to introduce a recognition site for the enzyme enterokinase at the 5' end of the CMTI gene sequence in addition to restriction sites for cloning into plasmid PIC9. Determination of the orientation and sequence of the gene were carried out by first cloning the construct into *E. coli*. The recombinant plasmid was then transformed into spheroplasts of *P. pastoris* using electroporation.

Transformants exhibiting a *methanol-utilisation-slow* phenotype were screened for inclusion of the gene at the genomic *AOX I* locus by PCR.

Clones containing the modified *CMTI I* gene were screened for the recombinant product, by direct assay of induced culture supernatants for anti-tryptic activity. A procedure was then developed to isolate the inhibitor. The isolate was subsequently identified using amino acid analysis and protein sequencing. The *N*-terminus of this protein was trimmed by digestion with enterokinase and the product identified as a recombinant form of CMTI I, equivalent to the native protein.

Inhibition constants were determined for both recombinant inhibitors as well as for native CMTI I. The inhibition constants are 3.3×10^{-7}M, 2.11×10^{-7}M and 2.17×10^{-7}M respectively for the fusion protein, recombinant CMTI I and native CMTI I. The results show that all three inhibitors have Ki of similar magnitude. As the structure and activity of squash inhibitors are intimately linked, it can be readily concluded that *P. pastoris* is capable of synthesising and folding active CMTI I. It is further shown that extension of the *N*-terminus of CMTI I by these eight amino acids had little effect on the inhibition constant.

Table of Contents

Illustrations

Figures

Tables

1.0 Serine Proteases and Their Inhibitors

1.1 Introduction

Proteases are enzymes which cleave peptide bonds. They were discovered around the turn of the century through studies of the properties of gastric juices. For instance, a component of acid chyme was shown to convert the enzyme prosecretin to secretin (Bayliss and Starling, 1902). As the number of proteases discovered grew, so did the need for their classification into various mechanistic sets. Methods used in the 1960's to this end, involved detailed studies of their kinetics for the investigation of substrate specificity and susceptibility of proteases to various inhibitors. Through chemical analysis common features were noticed, such as the requirement of an acid environment for activity of a group subsequently termed the acid proteases. At the forefront of techniques used to characterise proteases in the 1960s and '70s was X-ray crystallography. This enabled elucidation of the folding properties as well as situation and geometry of the active site. Examples include the X-ray structure of papain (Drenth, J. J. N., 1971) and carboxypeptidase A (Lipscomb, W. N.,1969).

Despite two proteases sharing the same mechanism of action, their tertiary structures may be quite different. This is exemplified by chymotrypsin and subtilisin which are both serine proteases, yet differ both in sequence and conformation. Despite the essential residues for catalysis being at quite different positions in their primary sequence, substrates bind to both enzymes in a very similar manner. Thus proteases which exhibit similar folding and geometry of the active site, yet may have low sequence homology, were classified into sub-groups.

Chymotrypsin, trypsin and elastase are a sub-group of serine proteases which are closely related in terms of tertiary structure and yet have only 40-50% amino acid homology. This has led to the suggestion of a common primordial gene for these enzymes which mutated in structurally important sequences, but not in the catalytic residues. The evolutionary significance of proteases is demonstrated by the fact that trypsin has been found in a wide variety of organisms including cattle, sheep, dogfish, lungfish and *Streptomyces*, (Reich, Rifken and Shaw, 1975).

The first proteases characterised were those found in abundance in various organs and bodily fluids. They may be considered to be of a catabolic nature, their purpose to degrade proteins for nutritional and other reasons. With the advent of more refined techniques of detection, less abundant proteases which are more difficult to isolate were discovered. Hence, molecular biology techniques were used for probing genomic DNA for coding sequences with homology to known protease genes. These proteases tend to be more versatile in both structure and function. They exhibit a limited processing of proteins, often for the adaptation of a protein to its physiological environment. As regulatory enzymes, their activity is required to be specific and short-lived in order to avoid all out devastation of essential proteins in their vacinity. It is here that inhibitors play an important role in regulating proteases.

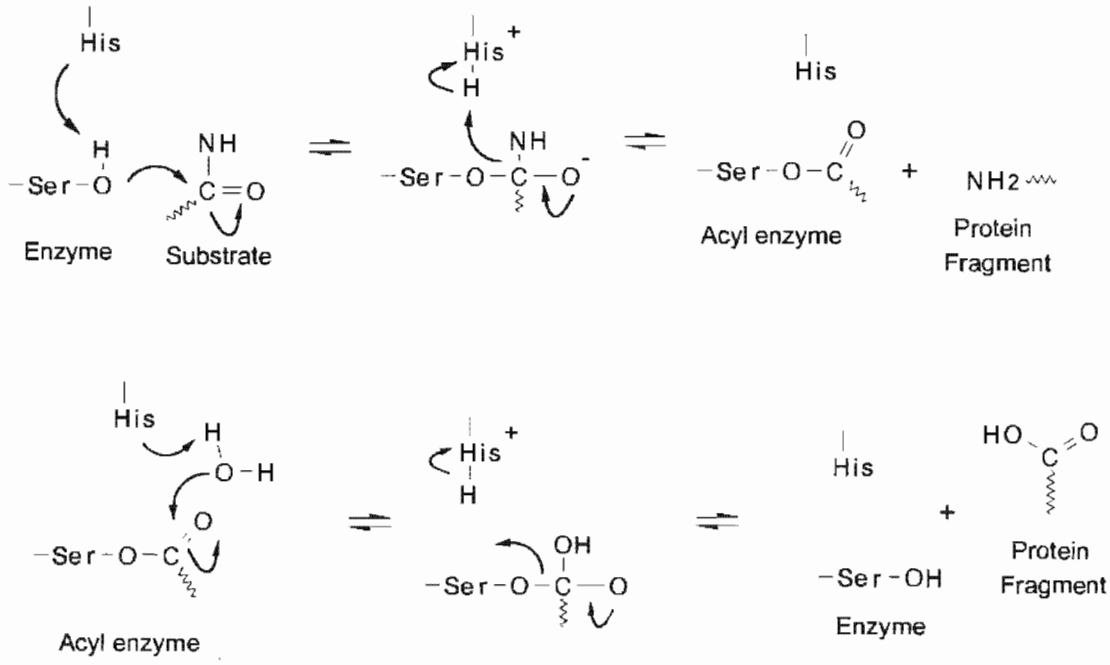
1.2 The Serine Proteases

The serine proteases are the largest family of proteases. Serine in the active cleft is pivotal to their mechanism of action, hence their name. The shape of this cleft determines the substrate for the enzyme as only certain residues will "fit". For instance, trypsin has a long, narrow cleft which accommodates basic amino acid side chains such as those of arginine and lysine; while in chymotrypsin this is shallow and wider, hence more bulky residues such as valine will bind.

Common to all serine proteases is a means of "charge transfer" which enables a bond to be established between Ser195 (trypsin) and the carbonyl carbon of the peptide bond to be cleaved on the substrate (see figure 1). Serine is usually an unreactive amino acid but, its position in the enzyme active centre next to a positively charged histidine ring (His95 in trypsin), allows serine to loose a proton to His95 and so become negatively charged. Protonation of His95 is in turn favoured by the presence of aspartic acid (Asp102 in trypsin) which has a negatively charged side chain. As this transfer of negative charge comes into play a simultaneous cleavage of the substrate peptide bond occurs. The *N*-terminal portion of the now cleaved substrate moves away leaving the *C*-terminal portion still bound to the enzyme. At this point water binds to the enzyme in place of the departed polypeptide, transferring a proton to His95. In effect the charge transfer mechanism is reversed with Ser195 once again becoming neutral, and release of the *C*-terminal portion of the cleaved substrate. Thus is the active cleft of a serine protease regenerated for proteolysis of another substrate.

Figure 1.

Mechanism of Action for the Serine Proteases.



1.3 Serine Protease Inhibitors

Serine protease inhibitors have been grouped into at least 16 families based on homologies between their primary sequences, tertiary structures and mechanism of binding. Many of these inhibitors crystallise and hence their structures have been elucidated from X-ray diffraction patterns (see below). Most serine protease inhibitors react in a substrate-like mechanism, as competitive inhibitors. That is, they are not irreversibly bound to the enzyme, but rather compete with a substrate for the active site. The notation most commonly adopted to describe where an amino acid lies in the primary sequence of an inhibitor, relative to the reactive site bond, is that of Schechter and Berger (1967). The residue at which peptide bond cleavage occurs is named the P1 residue, the residue bonded to its *C*-terminus being P1'. Residues to the *N* and *C* terminals of these two amino acids are termed as follows:

.....P3 - P2 - P1 - P1'- P2' - P3'.....

where the arrow indicates the site of cleavage.

Serine protease inhibitors share a common general structure whereby the active site is situated on a loop that projects from the core of the inhibitor, and is able to insert into the cleft of the enzyme. Apart from this feature they can be unrelated in structure. This binding loop is exposed to the solvent even though these residues tend to be hydrophobic in nature. It is held in place in the enzyme cleft by a supporting scaffold provided by the inhibitor core, as well as interactions between the core and residues which flank the reactive site (Bode and Huber, 1992). The binding loop usually fits quite rigidly in the enzyme cleft. In the case of trypsin inhibitors, hydrogen bonds are formed between the binding loop and residues Gly193N and Ser195N of trypsin. However, it is the cooperative and stabilising nature of the interaction between the inhibitor core and the binding loop which contributes most to this rigidity.

Both X-ray crystallography and NMR have proven extremely useful in unraveling the structures of both free and enzyme bound inhibitors. X-ray crystallography has provided

Both X-ray crystallography and NMR have proven extremely useful in unraveling the structures of both free and enzyme bound inhibitors. X-ray crystallography has provided detailed information on the tertiary structures of free bovine pancreatic trypsin inhibitor (BPTI) (Deisenhofer and Steigermann, 1975); BPTI bound to β bovine trypsin (Huber *et al*, 1974; Marquart *et al*, 1983); anhydrotrypsin (Huber *et al*, 1975; Marquart *et al*, 1983) and bovine trypsinogen (Bode *et al* 1978; Marquart *et al* 1983). NMR was used to illucidate the structure of the potato CI-2 inhibitor (Clore, G. M., *et al*, 1987). Grasberger *et al*, (1994) used NMR spectroscopy to show the *Ascaris* trypsin inhibitor is active at low pH because only then does it fold into the canonical shape of a serine protease inhibitor.

The specificity of inhibitors of the chymotrypsin-like proteases is determined by the side chain of the P1 residue, as this interaction is the most important in terms of energy required (Laskowski *et al*, 1987). Changing this residue can often alter specificity to that of another enzyme. By synthesising variants of *Cucurbita maxima* trypsin inhibitor III (CMTI III) with substitutions at the P1 position, McWerter *et al*, (1989) produced inhibitors against human leukocyte elastase and cathepsin G. Natural CMTI III inhibits trypsin as it contains arginine at the P1 position and this experiment demonstrates the simplicity with which inhibitor specificity can be changed. A few natural inhibitors have been shown to inhibit more that one protease: an interesting and unusual example being that of ecotin. This 142 amino acid protein is found in the periplasm of *E. coli* and exhibits activity against trypsin, chymotrypsin, elastase, factor Xa and kallikrein. Methionine was found to be the active site residue of this inhibitor in all cases. The side chain of methionine was found to "mimic" the "preferred" side chain for fitting in to the enzyme cleft. For instance, when bound to trypsin, it was found in an extended conformation similar to lysine (McGrath, M. E., *et al* 1995).

The reaction mechanism for serine protease inhibitors may be expressed by following the general model:

Figure 2.

$$E + I \underset{Koff}{\overset{Kon}{\Leftrightarrow}} EI \underset{Kon^*}{\overset{Koff^*}{\Leftrightarrow}} E + I^*$$

where, E is the enzyme; I the free inhibitor; EI the enzyme-inhibitor complex and I* the cleaved inhibitor (Finkenstadt, W. R., *et al* 1974; Quast, V., *et al* 1978).

rate of reaction for K_{on} is much faster than that for K_{off}*, implying the EI complex forms rapidly, with a slow dissociation to yield the modified inhibitor. Despite this, the complexes formed between unmodified and modified inhibitor and the enzyme were found to be equally stable in thermodynamic terms (Laskowski, M. Jr., Sealock, R. W., 1971). At neutral pH, the value for the hydrolysis of the peptide bond by the cognate enzyme, as expressed by the ratio K_{cat} / K_m, is high for an inhibitor, compared to that of a substrate (Laskowski and Kato, 1980). Small changes in the conformation of the binding loop on complex formation, as investigated by X-ray crystallography and NMR (see above), have provided some in-sight in to the differences between substrate and inhibitor binding.

Making use of the structural, thermodynamic and kinetic data available, recent work has utilised computer modeling to predict which residues are essential for activity. These ideas are subsequently validated by measuring activity following the appropriate amino acid substitutions and/or deletions. This may be done by chemically synthesising the varient or by cloning the gene with appropriate changes to the base sequence. Some groups have taken modeling experiments a step further by designing chimeric proteins able to inhibit more than one protease. These "multi-headed" inhibitors (having more than one enzyme binding site) do exist in nature. For instance, Rhodes *et al* (1960) showed that inhibitors known as ovomucoids from a number of avian species are single, double or triple headed. Le-Nguyent *et al*, (1989), chemically synthesised a 32 residue protein with the active site of *Ecballium elaterium* trypsin inhibitor, a squash inhibitor, and the *C*-terminal tetrapeptide of potato caboxypeptidase inhibitor. This synthetic, "double headed" inhibitor showed some loss of activity to trypsin in addition to an almost identical inhibition of carboxypeptidase A.

1.4 The squash family of inhibitors

Inhibitors found in the seeds of the *Cucurbitaceae* are commonly known as the squash family of inhibitors. Polanowski *et al* (1980), were one of the first groups to isolate two trypsin inhibitors, with molecular mass of 3300D each, from squash seeds. They were later named *cucurbita maxima* trypsin inhibitors I and III (CMTI I, CMTI III), according to the elution times of peaks with anti-tryptic activity after chromatography. CMTI I and CMTI III differ only at residue 9: in CMTI I this is Glu and in CMTI III it is Lys. Examples of this group of plants include summer squash, zucchini, and watermelon from which inhibitors have all been isolated (Otlewski, J.and Wilusz, T., 1985; Leluk, J., *et al* 1983; Polanswski, A., *et al* 1987.).

Squash inhibitors constitute the smallest natural serine protease inhibitors, ranging in length from 27 to 32 amino acids. Due to their stable, rigid conformation, in addition to being small enough to be chemically synthesised, there has been much interest in this group as a model system for the study of protein-protein interaction. As they are relatively insensitive to pH and temperature, harsh conditions can be employed in their extraction. For example, Matsuo *et al* (1992) isolated two squash inhibitors from bottle gourd, by first extracting with 10mM $(NH_4)2CO_3$; followed by acetone precipitation and four stages of reversed phase HPLC.

The reactive site always lies between residues five and six: the side chain of the P1 residue determining specificity (Otlewski, J., 1993). Because this is arginine, or in some cases lysine, squash inhibitors with few exceptions, display anti-tryptic activity. It can be argued that this resulted from investigators isolating new squash inhibitors by screening solely for antitryptic activity, for instance, using trypsin immobilised on an affinity column. Consequently, the enzyme/s they inhibit *in vivo* may well remain unknown, although several groups claim to be using a variety of proteases in their isolation procedure. Indeed, exceptions to the anti-tryptic activity of squash inhibitors have come to light with the discovery of several elastase inhibitors from the seeds of bitter gourd, in which the P1 residue is leucine (Hara, S., *et al,* 1989; Hamato, N., *et al,* 1995).

The inhibitor binding loop consists of residues 2 to 8 in the primary sequence, of which Val2, Pro4 and Ile6 are in contact with the enzyme. The positions of six half-cystines are conserved,

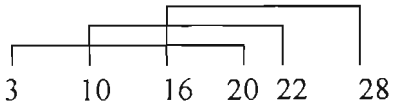
conserved, which are responsible for the structure-function relationship of these inhibitors. An aromatic residue is found at either position 7 or 27, but not at both, and a His-Glu-Glu tripeptide is situated at the *N*-terminus of two squash inhibitors. From the latter it has been postulated that this family of inhibitors arose as a product of limited proteolysis of one gene product, instead of as the products of individual gene expression (Otlewski, J.,1993). The majority of conserved residues are those involved in direct contact with the enzyme, apart from the six half-cystines. The sequences of ten squash inhibitors are tabulated as follows:

Table 1: Sequences of ten squash inhibitors

SQUASH INHIBITOR	SEQUENCE 1 5 6 10 15 20 25 29	REFERENCE
CMTI I	RVCP**R**I**L**MEC**K**KDSDC**L**AEC**V**C**L**EHGY**C**G	Wilutz et al (1983)
CMTI IV	HEERVCP**R**I**L**MK**C**KKDSDC**L**AEC**V**C**L**EHGY**C**G	Wieczorek et al 1985)
CPTI II	RVCP**K**I**L**MEC**K**KDSDC**L**AEC**I**C**L**EHGY**C**G	Wieczorek et al 1985)
CPTI III	HEERVCP**K**I**L**MEC**K**KDSDC**L**AEC**I**C**L**EHGY**C**G	Wieczorek et al (1985)
EETI II	GCP**R**I**L**MRC**K**QDSDC**L**AGC**V**C**G**PNGF**C**G	Favel et al (1989)
MCTI I	*ERRCP**R**I**L**KQC**K**RDSDC**P**GEC**I**C**M**AHGF**C**G	Hara et al (1989)
MCEI I	RICP**L**I**W**MEC**K**RDSDC**L**AQC**I**C**V**DGH C**G**	Hara et al (1989)
LCTI I	RICP**R**I**L**MEC**S**SDSDC**L**AEC**I**C**L**EQGF**C**G	Hakateyama et al 1991
BHTI III	RRCP**R**I**Y**MEC**K**HDSDC**L**ADC**V**C**L**PQGI**C**G	Matsuo et al (1992)
TTI II	CP**R**I**L**MPC**Q**VNDDC**L**RGC**K**C**L**SNGY**C**G	Huang et al (1990)

Modified from Otelwski, J. (1993). (* =pyroglutamic acid)
Highlighted letters represent concerved cysteins and the reactive site P1-P1' residues

Squash inhibitors exhibit little secondary structure: there are two reverse turns between residues 17 to 20 and 23 to 26; one turn of an 3_{10} helix from Asp13 to Cys 16 and a short antiparallel β-sheet from Cys20 to Gly29 (Otlewski, J., 1993). Thus tertiary structure is highly dependent on disulphide bridge formation. The topology of disulphide bridges in squash inhibitors is as follows:



Various methods were used to establish this pattern, as previous structural predictions, based on comparing the sequences of squash inhibitors to that of wheat germ agglutinin, proved incorrect. For example, 1H 2D NMR was performed on a synthetic form of *Ecballium elaterium* (squirting cucumber) trypsin inhibitor II (Heitz. A., *et al*, 1989) in order to elucidate S-S conectivities. Hara, S., *et al* (1989) elucidated the disulphide bridging pattern of natural *Momordica charantia* (bitter gourd) trypsin inhibitor II (MCTI II) by partially digesting the intact molecule with thermolysin, after attempts at cyanogen bromide digestion failed. Fragments generated were separated by HPLC and those containing cysteins determined by amino acid analysis. These peptides were subsequently reduced, the cysteins S-pyridylethylated and a further separation by HPLC performed. Sequencing of these final peptides enabled them to establish the disulphide bridging pattern. In all cases studied so far, disulphide topology has been found to conform to the above pattern.

Structural similarities have been noticed between small proteins relying on a tri-disulphide pattern for tertiary structure. They have been named the "knottins" and as such, squash inhibitors may be included in this class along side a number of unrelated proteins, such as the ω-toxin from the funnel web spider. (Liang Lin, S. and Nussinov, R., 1995).The structure of these proteins is roughly that of a letter T, with three loops held in place by two of the disulphide bridges. The third bridge is used to link the active site to the second loop: for squash inhibitors this being the active site loop at the *N*-terminus. Many other proteins show similarities to the knottins but tend to have more extensive secondary structures. What is important about the T-knot scaffold is that secondary structures are distinctly lacking in this group such that the disposition, number and pattern of disulphides must be subject to certain architectural rules to ensure full protein activity.

Investigations into structure-function relationships have centered around chemical synthesis of squash inhibitors and their analogues, with substitutions and deletions in the amino acid sequence. Kupryszewskki *et al* (1986) chemically synthesised CMTI III by the Merryfield

solid phase procedure. By comparing the immunigenicity of rabbit antibody raised against natural CMTI III to that of their synthesised peptide, they were able to confirm the inhibitor as the 29 amino acid, 3 disulphide-bridged structure proposed by Wilusz *et al* in 1983. This was further confirmed by UV and CD spectroscopy. Initially there had been some disagreement over the sequence and number of disulphide bridges, with a 28 amino acid, 2 disulphide-bridged structure being published for CMTI III two years earlier (Nowak *et al*, 1981).

Rolka *et al* (1991) chemically synthesised CMTI III substituting Phe for Arg5 in the P1 position in order to attain anti-chymotryptic activity. To enhance the potency of this analogue they introduced part of the sequence of the avian ovomucoid third domain in positions P4...P3', which is known to be a strong chymotrypsin inhibitor. The first analogue, with simply a substitution of Phe at P1, was shown to be a poor inhibitor of chymotrypsin while the latter construct was strongly anti-chymotryptic, with an inhibition constant (Ki) of 6×10^{-11} M. Prior to this they had shown replacement of the P1 residue with Val produced an elastase inhibitor (Rolka, K., 1989). They concluded that the P1 residue therefore determines specificity, with amino acids in the binding loop providing contacts with the enzyme specificity pocket and hence being important for enzyme-inhibitor complex formation.

Substitutions outside the binding loop have provided further evidence in support of the above. CMTI III truncated at the *N*-terminus by 2 residues, and by one residue at the *C*-terminus, was fully active, as were three other analogues containing substitutions in non-contact regions. (Rozycki, J., *et al* 1994). Last year, this group, in a more dramatic step, synthesised an analogue of CMTI III with the removal of four and simultaneous substitution of six amino acids (Rozycki, J., *et al* 1995). This inhibitor showed a Ki of the same order as CMTI III and is the shortest polypeptide inhibitor to date, being just 25 residues in length.

The potential of squash inhibitors and their analogues as therapeutic compounds has received some attention in recent years. They have been targeted against proteases of the blood coagulation system, in what has usually been a gun-shot type of approach. Hayashi *et al* (1994) isolated 8 squash inhibitors from various members of the *Cucurbitaceae* and investigated their activities toward enzymes of the intrinsic and extrinsic blood coagulation cascades. Only MCTI II caused a significant prolongation of the extrinsic blood coagulation

time due to its inhibition of Factor XIIa, with a K_i of 5.6×10^{-8} M. Earlier, Wynn and Laskowski (1990), reported inhibition of Factor XIIa by CMTI III.

1.5 *Cucurbita Maxima* Trypsin Inhibitor I

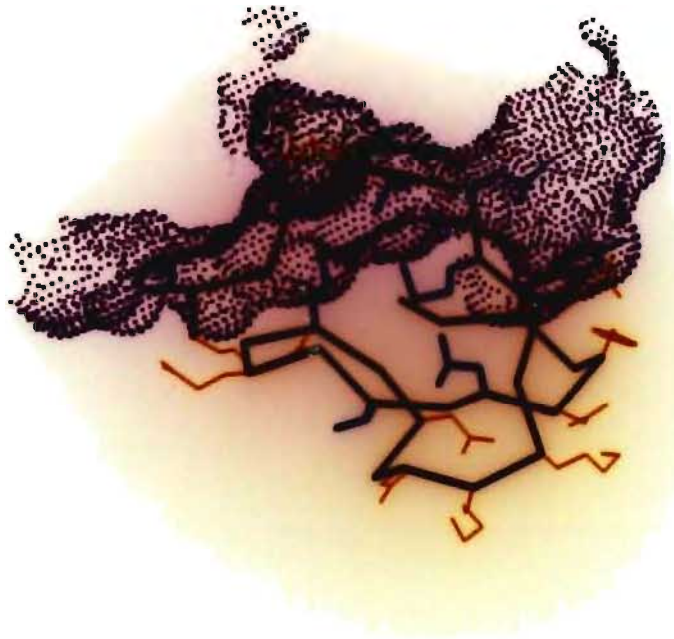
CMTI I has been mentioned a number of times in the preceding sections because it is the best characterised of the squash inhibitors. In 1989, collaboration between the Max-Planck institute in Germany and the University of Wroclaw, Poland, provided the solution structure of CMTI I by NMR ([1]Holak, Gondol, Otlewski, Wilusz, 1989). At the same time these groups published a comparison of the conformation of the binding loop in free solution (by NMR), and in a trypsin bound state (by X-ray crystallography). Their data was stringent enough to afford accurate calculations which showed there to be very little difference between the structure of the bound and unbound binding loop, ([2]Holak, Bode, Huber, Otlewski, Wilutz, 1989). Using a different approach the 3D structure was further refined by NMR in 1991. (Nilges, M., *et al*, 1991).

The amino acid sequence of CMTI I is given in table 1. It is a 29 amino acid polypeptide, the P_1-P_1' active site bond lying between Arg5 and Ile6. CMTI I may be considered a strong inhibitor of β bovine trypsin with inhibition constants in the order of 10^{-8}-10^{-12} M. A discrepancy exists between Ki reported from different groups, which may be explained by non uniformity in the assays used as well as difficulties in measuring high binding constants (Otlewski, J., 1993).

Otlewski and Zbyryt (1994) investigated the hydrolysis and resynthesis of the peptide bond in the reaction of CMTI I with β bovine trypsin. The reaction was performed in four different buffers ranging in pH from 3.2 to 8.3. At various time intervals, an aliquot was removed and subjected to analytical ion-exchange HPLC. The ratio of two peaks corresponding to the cleaved and intact forms of CMTI I were calculated at each time. A value of 21.6 was obtained for the ratio of Kon / K*on, showing that the association constant for the intact inhibitor is higher than for the cleaved inhibitor. The specificity index, K_{cat} / K_m for the CMTI I-β bovine trypsin interaction is high, but this value is dependent on an extremely low K_m for the inhibitor rather than a high value for K_{cat}. In this way CMTI I, like other squash inhibitors,

differs from a substrate. Added to this, thermodynamic data provided further evidence that slow hydrolysis of the active site bond was due to conformation of the inhibitor binding loop "locking the protease in a highly populated noncatalytic state" (Otlewski and Zbyryt 1994). They concluded that CMTI I conforms nicely to the standard mechanism.

Figure 3.



Computer simulated model of CMTI I in the active site cleft of trypsin (Arg5-Ile6 active site bond in yellow). Curtesy of Dr. D. Maeder, Dept. of Biochemistry, University of Cape Town.

2.0 The Production of Recombinant Proteins in Yeast

2.1 Introduction

The manipulation of microorganisms for the production of heterologous proteins has had a significant impact on both science and industry. The bacterium *Escherischia coli* continues to play a central role in cloning, for which there are many commercial strains and expression plasmids available. However, there are several short-comings to bacterial expression. Although high yields can be obtained, the product may not be a satisfactory representation of the native protein. Proteins are often sequestered in the periplasmic space in insoluble inclusion bodies, necessitating their extraction and purification. As this usually involves denaturing and renaturing the protein, the final product may be incorrectly folded, thus impairing activity. Microorganisms like *E. coli* cannot glycosylate proteins and the *N*-terminal methionine is often retained. Cloning mammalian genes in *E. coli* can also be problematic if the foreign mRNA requires extensive methylation. In addition, bacteria do not splice introns correctly.

The genetics and physiology of *Saccharomyces cerevisiae*, otherwise known as Bakers or Brewers yeast, have been extensively investigated. With this knowledge came the realisation that yeast could be used as an alternative to bacteria for the expression of recombinant proteins. Yeast are eukaryotes, and as such, their potential for producing eukaryotic proteins of superior quality to those produced by bacteria, has been of particular interest. Yeast can reproduce both sexually and asexually, depending on environmental conditions. Their genetic material is carried on 34 chromosomes. Vectors for *S. cerevisiae* are based on the naturally occuring, 2 micron plasmid which is found in high copy number and segregates through meiosis in a stable manner.

Yeast show transcriptional and translational features of higher organisms. For instance, yeast mRNA is 5' capped and is extended by poly(A) at the 3' end, while translation initiates from the first AUG. Cellular organelles similar to those found in higher eukaryotes are present,

such as mitochondria, rough endoplasmic reticulum and lysosomes. Proteins of homology to those in plants and animals have also been characterised including histones, peptide hormones and actin.

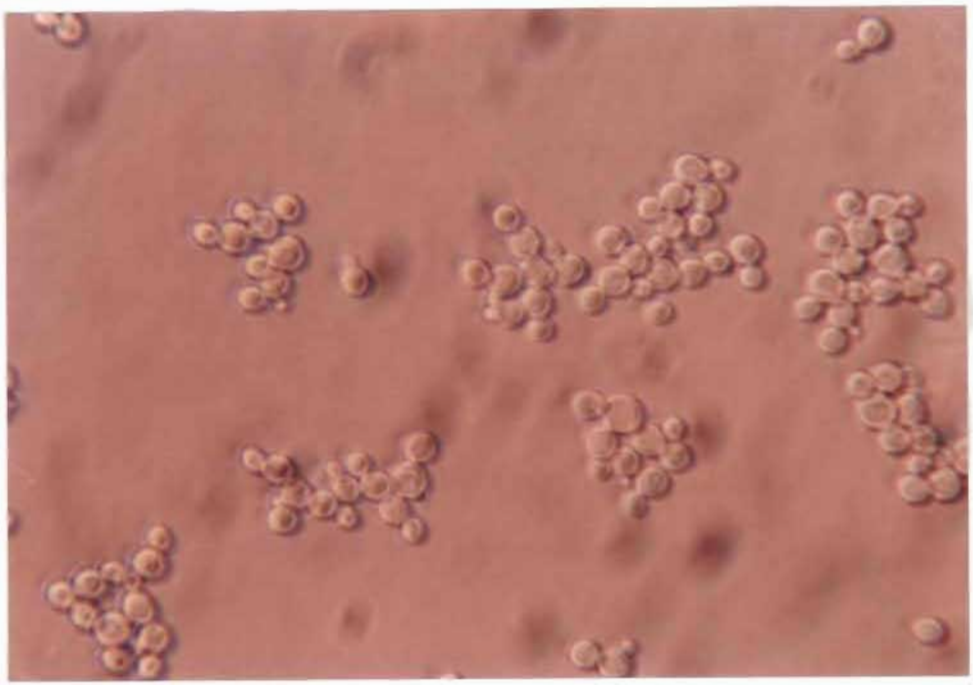
Several factors make *S. cerevisiae* attractive as an expression host: It can be manipulated in a prokaryotic manner, for which protocols are well established and numerous. Many auxotrophic markers for selection of transformants are available, in addition to commercial host strains. As *S. cerevisiae* contains a eukaryotic secretory pathway, recombinant proteins may be modified in a manner similar to that of higher organisms. For instance, unlike bacteria, *S. cerevisiae* is able to glycosylate proteins. Recombinant proteins can be secreted into the medium allowing for easier isolation. Finally, of commercial interest is the classification of *S. cerevisiae* as a "generally regarded as safe" (GRAS) organism. The FDA awarded its first license for a yeast-produced therapeutic compound, for the production of hepatitis B surface antigen (HBsAg), in 1987. HBsAg was produced as unglycosylated, but nevertheless, immunoreactive particles of the same molecular weight as those produced by *E. coli* (Valenzuela, P., 1982).

Even so, problems have arisen from the application of *S. cerevisiae* as an expression host. Yields tend to be poor, usually only up to 5% of total cell protein (TCP). This can be due to plasmid instability although various types of yeast vectors, containing stabilising sequences, have been developed to overcome this. For instance, the inclusion of both autonomously replicating and centromeric sequences has resulted in vectors that behave like circular minichromosomes and are thus stable throughout meiosis. Difficulties in scaling up have been reported as well as susceptibility to toxicity resulting from the intracellular accumulation of some recombinant products (Struhl, K., 1983). Similar to *E. coli*, some proteins are not secreted but sequestered in the periplasmic space, from which they must be extracted and purified. This often becomes necessary for large proteins which do not easily pass through the cell wall of *S. cerevisiae*, thus leading to a reduction in yield (Romanos, M. A., *et al* 1992). Finally, *S. cerevisiae* has a tendency to hyperglycosylate proteins which can lead to an impairment in activity. Consequently, alternative yeast genuses have been studied as tools for recombinant protein expression.

2.2 *Pichia pastoris* as an Expression Host

Pichia is a genus of yeast first isolated in 1922, France, from the exudate of the chestnut tree (*The Yeast*, 1970). Members of this genus are able to utilise methanol as their sole carbon source. Alcohol oxidase is the first enzyme in the methanol utilisation pathway, catalysing the oxidation of methanol to formaldehyde (Anthony, C., 1982). It is found in abundance in *Pichia pastoris* and under optimal conditions represents up to 30% TCP. Alcohol oxidase is the product of two near-homologous genes: *AOX1* and *AOX2*, of which *AOX1* produces the vast majority. Transcription of the *AOX1* gene is directed by a tightly regulated, methanol-induced promoter. The exploitation of this gene and its promoter have enabled the development of *Pichia pastoris* as a powerful tool for the expression of recombinant proteins.

Budding cells of *P. pastoris*, strain GS115, taken at 400x magnification



All foreign genes expressed in *P. pastoris* have been subject to the principle of separating the stages of biomass formation from product formation. This division is known to be advantageous in avoiding the depletion of proteins and factors necessary for cell growth, which can arise during the over-expression of a foreign protein. The *AOX1* promoter is both strongly repressed by glycerol and glucose, and induced by methanol. Thus, the basic strategy involves cloning the gene of interest downstream from the *AOX1* promoter, obtaining satisfactory cell density by growth on glycerol, and then replacing the glycerol with methanol for induction of the promoter and hence protein expression.

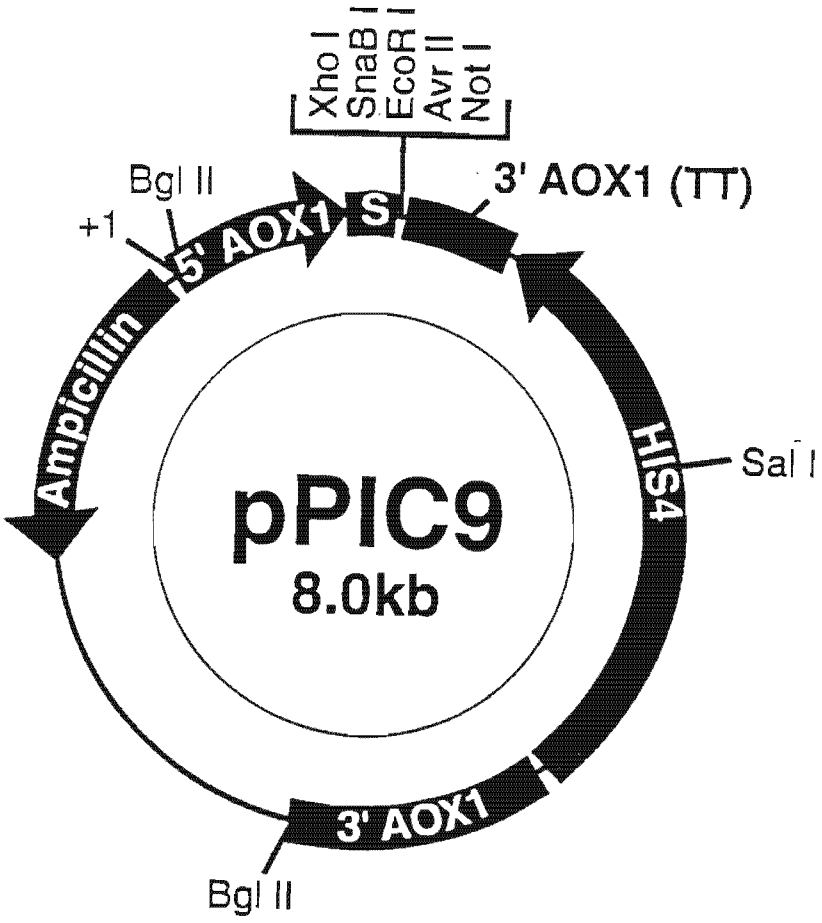
P. pastoris has no stable episomal vectors and as such, integrating vectors have been developed (Romanos, M., 1995). Integration into the genome is advantageous in conferring stability through meiosis. This allows a foreign gene to be passed on through generations without having to re-transform the host, as may be necessary for genes carried on episomal vectors. Vectors for both intracellular and extracellular protein production have been produced. Those enabling protein secretion into the medium contain the gene for a secretary signal upstream from the promoter. Although a heterologous signal sequence may be employed, it is usually better to use one from yeast. The most commonly used has been the α mating factor (α MF) signal sequence of *S. cerevisiae*. Other signal sequences include that from the acid phosphatase (*PHO5*) gene, and the invertase *SUC2* gene (Romanos, M., 1995).

Common features of vectors for transforming *P. pastoris* include:

- Bacterial sequences (usually pBR322) and an antibiotic marker for cloning in *E. coli*.
- *HIS4* for selection of transformants in *P. pastoris*.
- The *AOX1* promoter region.
- A multiple cloning site downstream from the promoter.
- *AOX1* genomic sequences for transcription termination.
- *AOX1* genomic sequences to afford integration of linearised plasmid at the *AOX1* locus.

Plasmid PIC9 is a typical secretory vector for *P. pastoris* and is shown below:

Figure 4.



Where TT is transcription termination and S the prepro region of the yeast αMF signal sequence.

(Reproduced from *Invitrogen* manual 3).

It is usual to generate several types of transformants, depending on the sites at which the plasmid is linearised and the genotype of the host cell. GS115 has been the host cell line most often used, containing a mutation in the histinadol dehydrogenase 4 (*his4*) gene to allow for selection of HIS4 transformants. For instance, cutting pPIC9 with *Bgl* II (see above) will generate two fragments. One contains bacterial sequences and has no further part to play in the cloning procedure. The other, larger fragment, containing the gene construct, is directed to integrate at the *AOX1* locus because it has ends homologous to this gene.

In up to 25% of GS115 transformants, a transplacement event will take place such that the *AOX1* gene is replaced by the *Bgl* II-cut fragment. Thus, having no structural *AOX1* gene, these transformants grow slowly on methanol because they now rely on the less efficient *AOX2* gene for alcohol oxidase production. These clones have been named methanol utilisation slow mutants (Mut^S). An equal number of *Bgl* II generated, GS115 transformants may have a methanol utilisation positive (Mut^-) phenotype, suggesting an integration event at either the *AOX1* or *HIS4* loci, with the gene remaining intact. In addition, a number of "transformants" will be generated containing no vector and are likely to be clones in which the *his4* gene has been converted to the wild type *HIS4* phenotype (Romanos, M., 1995).

Cutting pPIC9 with *Sal* I will result in a population of clones, the majority of which will exhibit a Mut^- phenotype, because the insert was directed to insert, as opposed to transplace, at the *HIS4* or *AOX1* loci. It has not yet been proven that either phenotype is preferable for the production of any one type of protein. However, it is useful to obtain both Mut^S and Mut^+ phenotypes to increase the chances of isolating a super-secretor. Some groups favour the faster growing Mut^- clones, on the basis of protein secretion being limited to secretary vesicles located in the budding tips of yeast (Romanos, M., 1995). However, Cregg *et al* (1987), in an attempt to explain the superior expression levels from their Mut^S transformants, suggested the longer generation time on methanol allowed more time for the complex folding of recombinant HBsAg particles.

A number of investigators have shown that gene multiplicity is an important factor in obtaining high yields. Both Mut^- and Mut^S clones have the remarkable ability to insert multiple copies of the gene, as tandem repeats, at the targeted genomic locus. The exact means by which this happens is not yet known, but it has been suggested that transforming fragments self-ligate before repeated head-to-tail insertion into the genome (Clare,[1] J. J., *et al*, 1991). This group showed a relationship between yield and the number of tandem repeats of a foreign gene, by isolating several clones of both Mut^+ and Mut^S phenotype, expressing the gene for tetanus toxin fragment C. A Mut^S clone, containing the maximum of 14 gene copies, produced a top yield of 10.5% TCP in shake flask culture and 27% by fermentation. Shortly afterwards, they used hybridisation probing of genomic DNA to isolate transformants containing multiple copies of the mouse epidermal growth factor (mEGF) gene. A clone

containing 19 integrated copies of the mEGF gene was found, which under fermentation conditions secreted 450µg/ml of mEGF (Clare,[2] J. J., *et al*, 1991).

Listed below is a selection of proteins expressed in *P. pastoris,* together with points of interest from each publication:

Table 2: Selection of proteins expressed in *P. pastoris*

RECOMBINANT PRODUCT	YEAR	NOTES	REFERENCE
Hepatitis B surface antigen	1987	High % of product assembled into 22nm particles, similar to those found in humans. Superior secretion in Mut^S clones.	Cregg. J. M., *et al*, (1987) *Bio/Tech 5*, 479-485.
Human tumor necrosis factor	1989	Antitumor protein originally expressed in *E. coli* in a partially soluble form. Synthesised as a fully soluble protein in *P. pastoris*. Attained 0.108 $g.l^{-1}.h^{-1}$ by continuous culture fermentation.	Sreekrishna. K., *et al*, (1989) *Biochem. 28*, 4117-4125.
Mouse epidermal growth factor	1991	Investigated effects of media composition on extracellular proteolysis and hence yield of this small, disulphide rich protein.	Clare[2] J. J., *et al*, (1991) *Gene 105*, 205-212.
Tetanus toxin fragment C	1991	Investigated effects on expression of: • site of genomic integration • *Mut* phenotype • gene dosage • comparison of yield from shake flask cultures and fermentation	Clare[1] J. J., *et al* (1991) *Bio/Tech 9*, 455-460.
HIV-ENV protein	1994	Used HIV-ENV gene product as a marker to show there is a relationship between copy number and resistance to the antibiotic G418. Showed expression in general is limited by gene copy number.	Scorer. C. A. *et al*, (1994) *Bio/Tech 12*, 181-184.

Porcine leukocyte 12-lipoxygenase	1994	Intracellularly produced oxygenase with characteristics more similar to that expressed by *E. coli* than the baculovirus-produced enzyme.	Reddy. R. G., *et al* (1994) *Biochem. Biophys. Res. Comm. 205*, 381-388.
Human cathepsin E.	1994	Time course assay showed a 90KDa product was converted to an 84KDa intermediate before a 82KDa protein in the media. Means of processing not yet known. Activity very similar to human cathepsin E.	Yamada. M., *et al* (1994) *BBA 1206*, 279-285.
Tick anticoagulant peptide	1994	Developed rapid screening procedure in 96-well culture plates, for His^+ transformants producing anti-Factor Xa activity in 200µl cultures.	Laroche. Y., *et al*, (1994) *Bio/Tech 12*, 1119-1124.
Mouse 5-HT5A seratonin receptor	1995	Fused *c-mcy* tag between gene and *AOX1* promoter to allow immono-detection of product. Found expression up to 8 times superior with proteolytic minus host strain.	Markus. H. W. *et al* , (1995) *FEBS letts.* 451-456.
Bovine enterokinase	1996	Cloned in pro region of α MF signal sequence 5' to gene, to allow for correct *N*-terminal processing by the KEX2 protease in *P. pastoris*. Fully active, glycosylated protein with correct *N*-terminus produced.	Vozza. L. A., *et al*, (1996) *Bio/Tech 14*, 77-81.

In several cases, *P. pastoris* has proven a superior expression host to *S. cerevisiae*. Comparisons of yield have been made although this is not easy due to differences in experimental procedures. Referring back to expression of HBsAg, the yield from *S. cerevisiae* was 1-2 monomers per 100mg of protein, equivalent to 0.02g/l (Valenzuela, P., 1982), in contrast to 2.3 monomers (0.04g/l) in *P. pastoris* (Cregg, J. M., *et al*, 1987). The immunogenic outer membrane protein, pertactin (P69) from *B. pertussis*, a microorganism responsible for whooping cough in infants, was co-expressed in *S. cerevisiae* and *P. pastoris* (Romanos, M. A., *et al*, 1991). With over 50 copies of the recombinant vector, *S. cerevisiae* produced 0.1% TCP as P69 with 100% solubility. *P. pastoris* produced P69 at a level of 0.5% TCP from a clone containing 30 copies of the gene, but with less than 10% of the product in a soluble form.

Usually there is a remarkable leap in yield from shake flask culture to controlled fermentation conditions. The ease with which protein expression in *P. pastoris* can be scaled up and it's ability to grow on inexpensive media has allowed a conversion to fermentation in most investigations. Protein expression in *S. cerevisiae* on the other hand can be difficult to scale up. In the above study, over 3g/l of P69 was obtained in high density fermentation of *P. pastoris* containing this gene (Romanos, M. A., *et al*, 1991). A Mut^+ clone containing the tick anticoagulant peptide gene, produced 21.4mg/l in shake flask culture which increased to 1.7g/l during fermentation (Laroche, Y., *et al*, 1994).

P. pastoris has been turned to as an alternative to *S. cerevisiae* for the expression of glycosylated proteins. One such protein is invertase, which is encoded by the *SUC2* gene of *S. cerevisiae*. Tschopp, J. F., *et al* (1987), studied invertase expressed in three systems: Recombinant invertase synthesised by *P. pastoris*; from wild-type *S. cerevisiae* and from the *sec18* mutant of *S. cerevisiae*, in which a mutation in the endoplasmic reticulum is present. The carbohydrate content of each invertase produced were characterised by observing their buoyant densities in a cesium chloride gradient. Invertase produced from wild type *S. cerevisiae* was higher in carbohydrate content than either of the other two forms, which were near-identical. This experiment demonstrated the difference in glycosylation patterns of *S. cerevisiae* and *P. pastoris*. A subsequent investigation of the size distribution and structure of *N*-linked glycan side chains from invertase produced by *P. pastoris*, showed a lack of terminal α1,3-linked mannose residues that are characteristic of *S. cerevisiae* core oligosaccharides (Ginna, L. S. and Tschopp, J. F., 1988). Hence, *P. pastoris* glycosylates in a manner more akin to higher organisms. As the length of glycan side chains can be important for protein folding and their passage through the secretory pathway, the advantages of using *P. pastoris* for the expression of glycosylated proteins is apparent.

It is now generally accepted that secretion of recombinant proteins from *P. pastoris* is largely protein-specific. Thus for any investigator contemplating *P. pastoris* as a means of protein production , it is necessary to gather as much information before hand about the folding characteristics and activity of that protein. For instance, strain GS115 is known to produce a number of extracellular proteases which place protease-susceptible products in danger of

degradation. To counteract unwanted proteolysis, protease-deficient host strains have been developed such as *Invitrogen's* SMD1168. Other strategies include dropping the pH of the media to a point at which protease activity is impaired and adding protein hydrolysates, to provide polypeptide chains that act as a "decoy" for protease action (Clare[2], J. J., *et al*, 1991).

Further to the point of expression levels being protein specific, maximising gene dosage has actually resulted in a reduction in yield in some cases. For instance, secretion of the HIV-1 gp120 envelope protein decreased with clones containing any copy number above one (Scorer, C.A., *et al*, 1993). It has been shown that a protein requiring extensive folding, or containing a retention signal, may cause a blockage in the secretary pathway. This is exemplified by the large surface protein of hepatitis B virus, which accumulates and subsequently causes enlargement of the rough endoplasmic reticulum in mammals (Biemans, R., *et al*, 1991). A bottle neck in the secretary pathway, which is more likely to occur with super-secretors, can lead to cellular toxicity (Romanos, M. A., 1992). As mentioned above, this may not be solely as a result of foreign protein accumulation but can also be due to a limitation of resources for the production of proteins and factors necessary for cell maintenance.

3.0 Synthesis and Cloning of the *CMTI I* gene in *E. coli*

3.1 Summary

A 93 mer oligonucleotide (oligo), containing the sequence for the *CMTI I* gene, was synthesised using solid phase phosphoramidite chemistry. Two primers containing restriction sites for *Eco* RI and *Avr* II, to enable cloning into plasmid PIC9, were also synthesised. Forward primer design included an enterokinase cleavage site immediately 5' to the first codon of the *CMTI I* gene. Preferred codon usage for *S. cerevisiae* was utilised for each sequence. The polymerase chain reaction (PCR) was used to amplify the modified *CMTI I* gene with the addition of linkers containing restriction sites for *Eco* R1 and *Avr* II. Restriction mapping of the PCR product resulted in the expected number and size of fragments. The PCR product was then ligated to pPIC9 cut at the multiple cloning site with *Eco* R1 and *Avr* II, and transformed into *E. coli*. Transformants were screened for the presence of an insert by PCR amplification of the plasmid borne gene. Three clones produced PCR products of the expected electrophoretic mobility, suggesting inclusion of the gene insert. Sequencing of plasmid DNA isolated from one such positive clone showed the modified *CMTI I* gene to be in frame and of the correct sequence.

3.2 Materials and Methods

For media recipes and bacterial strains, see appendix I.

3.2.1. Oligonucleotide synthesis

Oligonucleotides (oligos) were synthesised in a 3' to 5' direction on an AutoGen™ 6500 DNA synthesiser, using solid phase phosphoramidite chemistry. Oligos were cleaved from the resin by incubation in aqueous ammonia at room temperature for 1.5 hours and precipitated in butanol in a ratio of 1:10 v/v. The tube was shaken vigorously for one minute and the DNA pelleted by spinning at 12000rpm, 4°C for 5mins, in a microcentifuge. Butanol was poured off and the pellet resuspended in 200µl of buffer A (see below) for HPLC purification.

3.2.2. HPLC purification and deprotection of oligonucleotides

Oligos were purified by anion exchange HPLC using a Mono Q FPLC column. Elution buffers were as follows:

A: 50mM LiCl
10mM NaOH
pH 12.0

B: 3M LiCl
10mM NaOH
pH 11

A linear gradient from 100% A to 100% B was run over 90 min. The peak containing DNA to which the trityl group was still attached, was collected and the DNA allowed to precipitate in 6 volumes of 1:3 v/v ethanol to acetone, for 3 hours at -20°C. DNA was pelleted by spinning at 12000rpm, 4°C for 20 min. This step was repeated on the supernatant following centrifugation, to retrieve as much DNA as possible.

Pellets were air dried prior to removal of the tritly group, (located on the 5' phosphate of each oligo), as follow: The pellets were redissolved in 50µl of cold aqueous ammonia, which was subsequently dried off, under vacuum, overnight at room

temperature. 80µl of 80% acetic acid was added for half an hour to remove the trityl group. The detritylated oligos were reprecipitated in butanol and the pellets dissolved in 200µl of 1xTE buffer. The purity of each oligo was checked on a 20% denaturing polyacrylamide gel (c.f. section 3.2.9.5).

3.2.3. Concentration and prediction of annealing properties of Oligos

Oligos were diluted 1:100 in distilled water (dH_2O) and their UV spectra read between 220nm and 280nm on a Beckman DU 650 spectrophotometer. The concentration of each oligo was calculated taking the optical density of 1 unit being equivalent to approximately 37µg/ml for single stranded DNA. The computer program *Oligo*, version 3.4, was used to predict possible primer-primer interactions and secondary structures. Using this software, melting temperatures (Tm) for both forward and reverse primers bound to the 93mer oligo were predicted as 38.3°C and 47.3°C respectively, in a buffer of 1.5mM $MgCl_2$, 50mM KCl, 10mM Tris-HCl, pH 8.0. Tm in this case refers to the melting temperature of the double helical region between primer and template DNA. To determine whether the primers interacted with each other, approximately 1µg of both primers were placed in the PCR buffer (1.5mM $MgCl_2$, 50mM KCl, 10mM Tris-HCl, pH 8.0) and subjected to UV melting between 10 and 100°C on a Unicam 1700 UV spectrophotometer.

3.2.4. PCR amplification of the *CMTI I* gene

Primers were diluted in dH_2O to give a final concentration of 1pmol/µl. dNTPs were diluted in 1xTE to a concentration of 8mM. A reaction cocktail was set up containing all reagents except the enzyme and primers, in order to provide a uniform mix for each reaction. All PCR was carried out in reaction volumes made up to 50µl with dH_2O and covered with 50µl of paraffin to minimise evaporation. With the addition of primers and lastly enzyme, PCR reactions were set up as follows:

SAMPLE NO.	TEMPLATE ng	REVERSE PRIMER pmol	FORWARD PRIMER pmol	TAQ. POL. units	BUFFER µL OF 1.5mM $MgCl_2$	dNTPs 8.0mM µl
1	500	20	20	1	2	0.5
2	500	15	15	1	2	0.5
3	500	10	10	1	2	0.5
Control	-	20	20	1	2	0.5

A two step annealing procedure was used in the PCR reaction. The steps were as follows:

1. Denature: 94°C for 20s
2. Anneal reverse primer: 43°C for 30s
3. Anneal forward primer: 37°C for 30s
4. Extend: 72°C for 30s
5. Final extension: 72°C for 300s

Steps 1 to 4 were repeated for 30 cycles, with step 5 as a final extension. Following PCR, 10µl were removed from each tube and subjected to 1.5% agarose gel electrophoresis.

3.2.5. Restriction Analysis of the PCR Product

PCR was repeated and the products pooled. Four samples of 10µl each, were incubated separately for one hour at 37°C, with one unit each of the following restriction endonucleases:

1. *Hinf* I
2. *Eco* RI
3. *Avr* II
4. *Eco* RI plus *Avr* II

Each digest was heat inactivated at 90°C for 5 min and then run on a 12% non-denaturing polyacrylamide gel.

3.2.6. Preparation of pPIC9

E. coli strain TOP10F' (*Invitrogen*) was grown in LB broth to an optical density (600nm) of between 0.3-0.6. Competent cells were made from a 20ml culture and subsequently transformed with plasmid PIC9 (*Invitrogen*) by the method of Chung *et al*, (1989). One such transformant was picked in order to isolate plasmid DNA. Plasmid PIC9 was maxipreped from a culture volume of 500ml by the alkaline lysis method (*Current Protocols*, 1989, unit 1.7.). High speed centrifugation in a cesium chloride gradient was used to purify plasmid DNA (*Current Protocols*, 1989, unit 1.7.5.), which was then dialysed against 1xTE for 3 hours at 4°C. Precipitation of pPIC9 was carried out by the addition of 3M sodium acetate to a final concentration of 0.3M and 3 volumes of ethanol (*Current Protocols*, 1989, unit 2.1.1). DNA was left to precipitate at -20°C for 3 hours and the pellet recovered by spinning for 10 min, 4°C at 12000rpm. pPIC9 was dissolved in 200µl of 1xTE and the optical density (260nm) recorded. pPIC9 was stored at -20°C.

3.2.7. Restriction digest of pPIC9

40µl containing 27.2µg of pPIC9 was simultaneously digested with *Eco* R1 and *Avr* II. Sticky ends were prevented from religation by the removal of 5' phosphate with calf intestinal phosphatase (*Current Protocols*, 1989, 3.10.1.). The reaction was stopped by incubating the mixture at 75°C for 15 min. DNA was phenol/chloroform extracted and precipitated with ethanol and 10% v/v 3M sodium acetate (c.f. section 3.2.6). The pellet was redissolved in 50µl of 1xTE and divided in two aliquots. One sample was split into five aliquots and run on a 0.7% low melting point agarose gel. Bands of linearised pPIC9 were excised from the gel and the DNA extracted using a *Geneclean* kit (*US Biochemical corporation*). Purification by phenol chloroform extraction followed by precipitation with ethanol/ 10% v/v 3M sodium acetate (c.f. section 3.2.6) was once again carried out on both *Genecleaned* and non-*Genecleaned* pPIC9. The concentration of each DNA was calculated from optical density (260nm) readings; one OD unit being equivalent to 50µg/ml of double stranded DNA. DNA was dissolved in 50µl of 1xTE and stored at -20°C.

3.2.8. Preparation of the PCR product for ligation

Following PCR, taq polymerase was removed by incubating 1hour at 37°C, with 1µl of 20mg/ml proteinase K to every 20µl of PCR reaction mix. All protein was then removed by repeated phenol-chloroform extractions and the DNA subsequently precipitated with ethanol/ 10% v/v 3M sodium acetate (c.f. section 3.2.6). The OD (260nm) was measured in order to calculate DNA concentration (assuming 1OD≅ 50µg/ml double stranded DNA). A double digest with *Eco* R1 and *Avr* II was carried out simultaneously, by incubating the DNA with one unit of each enzyme at 37°C. The reaction mix was heat inactivated at 70°C for 5 min. Proteins were removed by phenol-chloroform extraction and DNA precipitated using ethanol/ 10% v/v 3M sodium acetate (c.f. section 3.2.6).

3.2.9. Ligation of the PCR Product to pPIC9 and transformation of *E. coli*

The following six ligation reactions were carried out:

1. 30ng "*Genecleaned*" linear pPIC9 with 200, 600 and 1000ng PCR product
2. 30ng "non-*Genecleaned*" linear pPIC9 with 200, 600 and 1000ng PCR product

All samples were ligated by overnight incubation at 15°C with 1 unit each of T4 ligase (*Current Protocols*, 1989, 3.14.1.).

Competent cells of *E. coli* strain TOP10F' were prepared as above. 100µl of cells were added to each ligation mix. The following controls were set up:

1. *Uncut pPIC9 plus cells*: to establish efficiency of transformation of super-coiled pPIC9 and the degree of cell competency.
2. *Cells only*: as a control for the selection of transformants
3. *Non-CIPed, linearised pPIC9 plus cells*: to establish whether the double digest of the plasmid was complete, as any contaminating supercoiled plasmid will cause a background of transformants lacking the gene insert.
4. *CIPed, linearised plasmid plus cells*: to establish efficacy of "CIPing" procedure.

Transformation was carried out by the method of Chung *et al*, (1989), and transformants selected for on LB agar plates containing ampicillin.

3.2.9.1 PCR screen of transformants

Two 21 mer oligonucleotide primers with sequence homology to either side of the multiple cloning site (MCS) of pPIC9 (sequences taken from the *Invitrogen* manual, ver 3), were synthesised and purified as above. Tm values were calculated as 51.7°C and 54.3°C (in 1.5mM $MgCl_2$, 50mM KCl, 10mM Tris-HCl, pH 8.0) for forward and reverse primers (*Oligo* ver3.4). Single colonies of transformants were placed in 1ml of LB and grown overnight at 37°C. 5µl of the resulting cultures were diluted with 95µl of dH_2O and 25µl of these diluted cultures were boiled for 3 min. In all PCR, template DNA was provided from the addition of the above cultures, with a non-transformant being used as a negative control. The polymerase used was *Dynazyme*. Reactions were as follows:

TEMPLATE µl of boiled culture	REVERSE PRIMER pmol	FORWARD PRIMER pmol	TAQ. POL. units	BUFFER µL OF 1.5mM $MgCl_2$	dNTPs 8.0mM µl	dH_2O µl
25	50	50	1.5	5	1.5	15

PCR was carried out as follows:

1. denature: 94°C for 20s
2. anneal: 49°C for 30s
3. extend: 72°C for 60s
4. final extension: 72°C for 300s

Steps 1 to 3 were repeated for 30 cycles with step 4 as a final extension.

Following PCR, 10µl was removed and subjected to 1.5% agarose gel electrophoresis.

3.2.9.2 Sequencing of the gene insert

Clone 1 (see results) was chosen for sequencing. Plasmid DNA was first maxipreped by the alkaline lysis method and purified in a cesium chloride gradient, (c.f. section 3.2.6). Sequencing was carried out by the dideoxynucleotide method (*Current Protocols*, 1991, unit 7.4), using a *Promega TaqTrack* kit. Primers for sequencing were those previously synthesised for PCR screening of the MCS. 1.2µl of 15pmol/µl of recombinant PIC9 template was used per reaction. Sequencing gels were run at

50V/cm, using a water jacket to provide a constant temperature of 45°C. Sequencing gels were dried on *3M Whatman* filter paper wrapped in polythene, before exposure to X-ray film for 24 hours. Autorads were developed by emmersing the film in X-ray developer (*AGFA* G127) for 5 min, prior to treatment with fixing solution (*AGFA* G333c rapid fix) for 5 min and washing with water.

3.2.9.3 Agarose gel electrophoresis

High grade agarose (*Sigma*) was dissolved in 1xTBE, to give the percentage of weight to volume required. Ethidium bromide was added to a final concentration of 0.5µg/ml for visualisation of DNA. Agarose was dissolved by boiling the solution in a microwave oven. Molten agarose was poured into a 8cm x 15cm horizontal gel chamber containing a comb at one end. After the agarose had set, the comb was removed, and the gel immersed in 1xTBE. Samples were dissolved in loading buffer containing 1xTBE and 1µl of bromophenol blue/ xylene cyanol tracking dye. Gels were run at 6V/cm for 1.5 hours and DNA visualised on a UV light box at 254nm.

3.2.9.4 Non-denaturing polyacrylamide gel electrophoresis

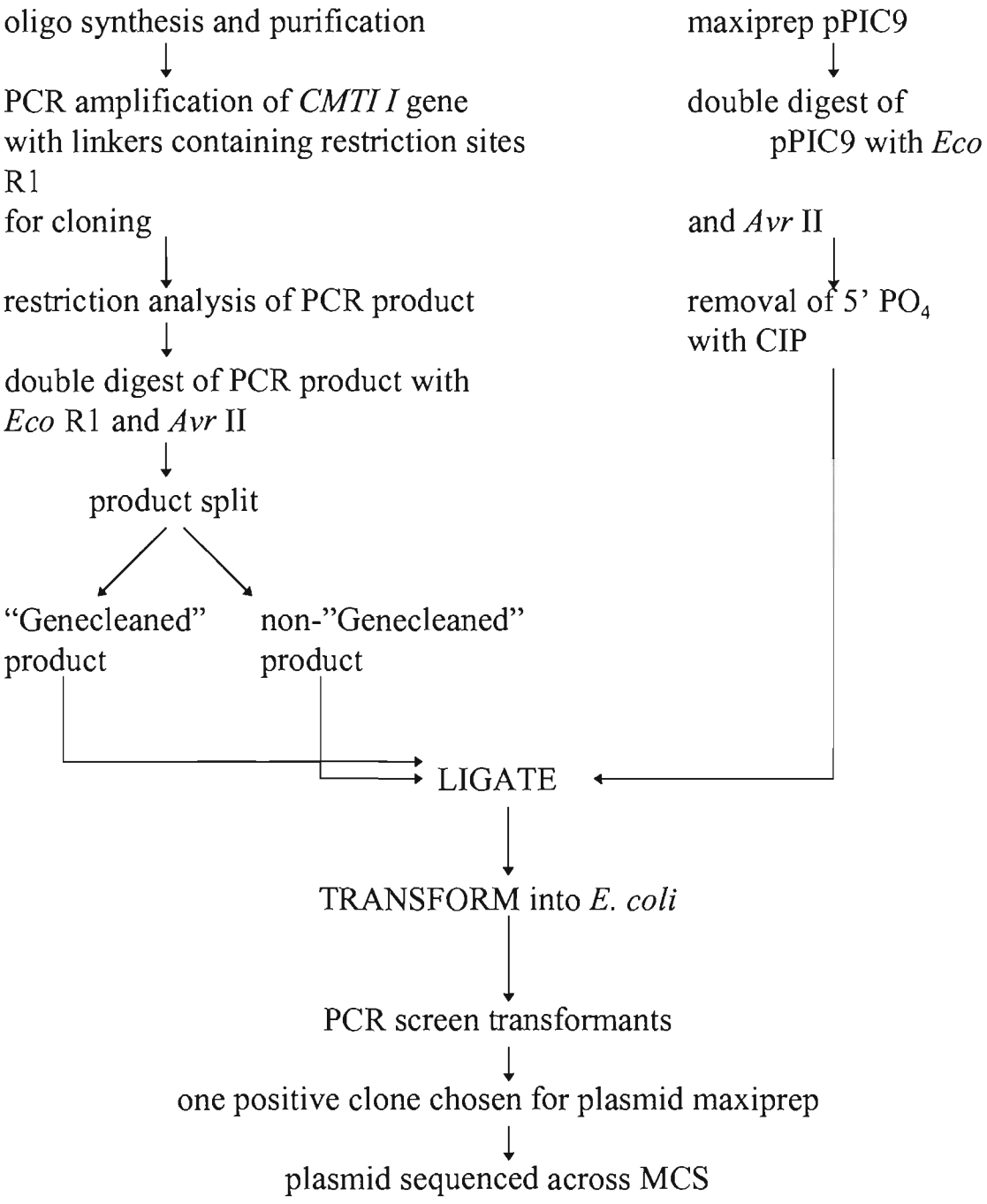
29% acrylamide/ 1% bisacrylamide stock solution was diluted by the addition of 10xTBE and dH_2O to give the percentage acrylamide in solution required for the separation of DNA fragments, in a total volume of 25ml. A 140mm x 9mm x 1mm gel was polymerised between vertical glass plates, by the addition of 130µl of ammonium peroxidisulphate stock solution and 20µl TEMED. Samples were run at 5V/cm in 1xTBE with tracking dye in one lane, until the Bromophenol Blue band just entered the bottom tank buffer. Gels were soakfor 10 min in a solution of 5µg/ml of ethidium bromide solution and visualised on a UV light box at 254nm.

3.2.9.5. Denaturing polyacrylamide gel electrophoresis

38% acrylamide/ 2% bisacrylamide stock solution was diluted by the addition of 10xTBE and dH_2O, plus urea to a final concentration of 7M. The gel was polymerised between vertical plates as above. Samples were prepared by diluting 1:2 with recrystallised formamide, and heating to 55°C for 20 min prior to loading. Gels were run hot at 20V/cm, using 1xTBE as the electrophoresis buffer.

Figure 5.

Summary of cloning strategy



3.4. Results and Discussion

When designing primers for PCR amplification of the 93 mer oligo, care was taken to minimise formation of secondary structure and self-annealing of the single stranded primer sequences. This would increase the chances of obtaining one, uncontaminated PCR product for cloning. To this end, several permutations of the base composition were studied (aided by *Primer/GCG* software) in order to obtain primers with the least secondary structure.

For the reverse primer, no sequence could be found that would have less than four consecutive base pairs. This would probably result in the primer looping back onto itself forming a hairpin structure, which could impair polymerase activity during PCR. To overcome this, an extra codon: AGC, was placed between the *Avr* II palindrome and the stop codon in the reverse primer (c.f. figure 7). Accordingly, the complimentary codon was placed at the end of the 93 mer oligo, which brought the total number of codons for binding of the reverse primer to six. The inclusion of AGC reduced the number of consecutive AT or CG base pairs within the reverse primer to three. According to the laws of thermodynamics this structure is too unstable to form a hairpin at this ionic strength.

Extra bases were placed on the 5' ends of each primer to enhance the efficiency of binding of taq polymerase. Because they lay "outside" of the restriction sites these codons would not be included in the cloned gene, such that manipulation of their sequence also provided a flexible means of minimising hairpin formation. Furthermore, it should be noted that inclusion of a start codon (ATG) at the 5' end of the gene sequence was unnecessary, as this was provided upstream in the *AOX I* promoter region of pPIC9.

To determine whether the primers interacted, UV melting (260nm) experiments of an equimolar mixture of both primers were carried out. No thermal denaturation was observed and primer design was therefore deemed satisfactory.

Figure 6.

UV melting spectrum (260nm) of 1μg each of forward and reverse primers

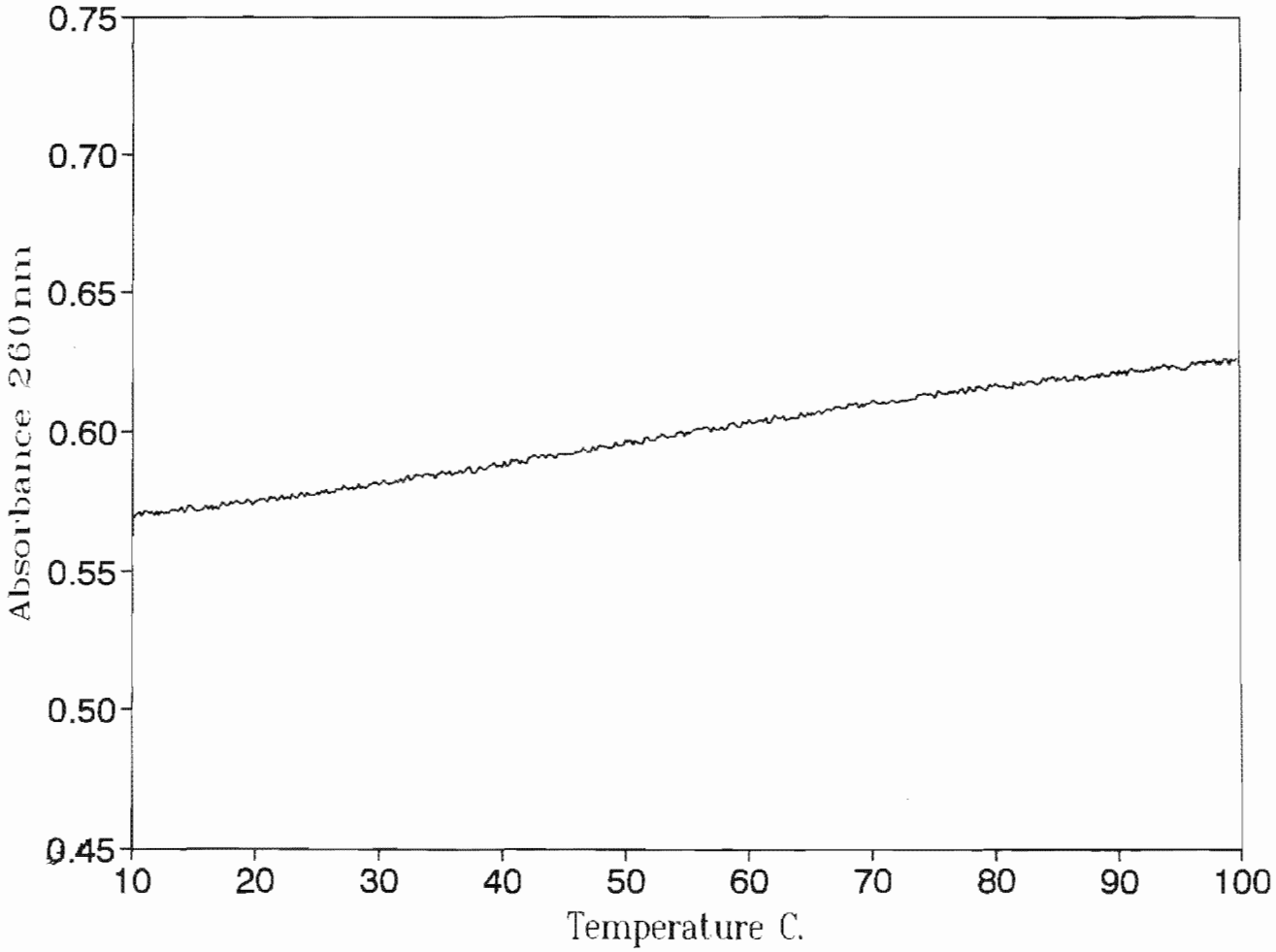


Figure 7.

PCR amplification of the *CMTI I* gene with the simultaneous addition of an enterokinase cleavage site at the 5' end and linkers containing restriction sites for cloning in pPIC9

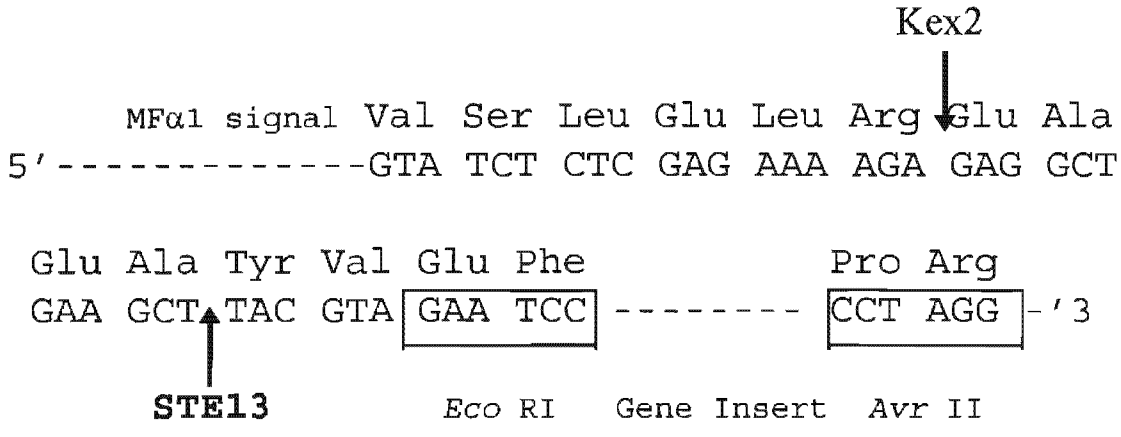
KEY

antisense strand
sense strand
forward primer
reverse primer
| :base annealment between 93mer oligo and primers
:direction of transcription
: site of cleavage for enterokinase

The multiple cloning site of pPIC9 lies immediately downstream from the prepro region of the *S. cerevisiae* α MF signal sequence. The Kex2 protein produced by *P. pastoris* performs the initial processing of the signal sequence, by cleavage between Arg and Glu as illustrated below (c.f. figure 8). Further cleavage of remaining Glu-Ala repeats are carried out by the *STE13* gene product. (Romanos, M. A. *et al*, 1992). As the diagram shows, codons for Tyr and Val remain between this last cleavage site and the *Eco* RI site. Including the two codons of the *Eco* RI site itself, this brings the number of extra amino acids that should remain on the *N*-terminus of the gene product, to four. To facilitate correct processing of the *N*-terminus at a later stage, the gene sequence for the recognition site for enterokinase was included in the forward primer design. Enterokinase is a serine protease which converts trypsinogen to trypsin and has the recognition sequence: Asp-Asp-Asp-Asp-Lys. Cleavage occurs at the peptide bond formed between lysine and the residue at its *C*-terminus, and so the gene sequence for this recognition site was placed immediately 5' to the first codon of the modified *CMTI I* gene.

Figure 8.

Processing of the signal sequence at the *N*-terminus of the fusion protein

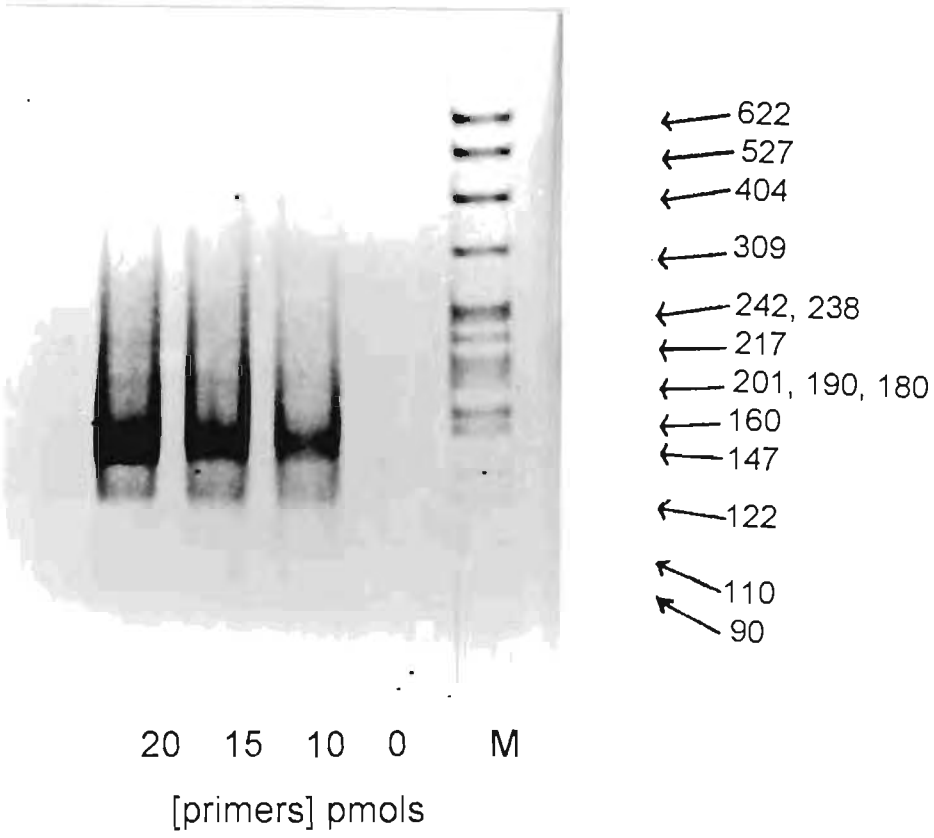


The sequence of each oligo was optimised for *Saccharomyces* codon usage. This does not mean, that had codons been optimised for *E. coli* usage, misincorporation of amino acids into the product would have necessarily occurred. Codon preferences between *E. coli* and *S. cerevisiae* differ only for proline, glutamine and cysteine (Ikemura, T., *et al*, 1982; Sharp, P. M., *et al*, 1988; Bennetzen, J. L., and Hall, B. D., 1982), which are amino acids all present in CMTI I. Documentation does exist of lower yields occurring for highly-expressed proteins in which codons rarely used are present (Romanos, M. A., 1992). Therefore *Saccharomyces* codon usage, (which is near-identical to that of *P.pastoris*, (*Invitrogen*)) was considered advantageous.

Due to a 9°C difference in the Tm of forward and reverse primers, PCR amplification of the gene was initiated by a two step annealing of primers. This strategy should in principle increase binding efficiency of the forward primer, which has a lower melting temperature. The effects of primer concentration on PCR product formation are shown below. The intensity of bands show an increase in product formation with increasing amounts of primers. From figure 5 it is evident that 20pmol of each primer can be used with no visible side products:

Figure 9.

10% non-denaturing PAGE of PCR products



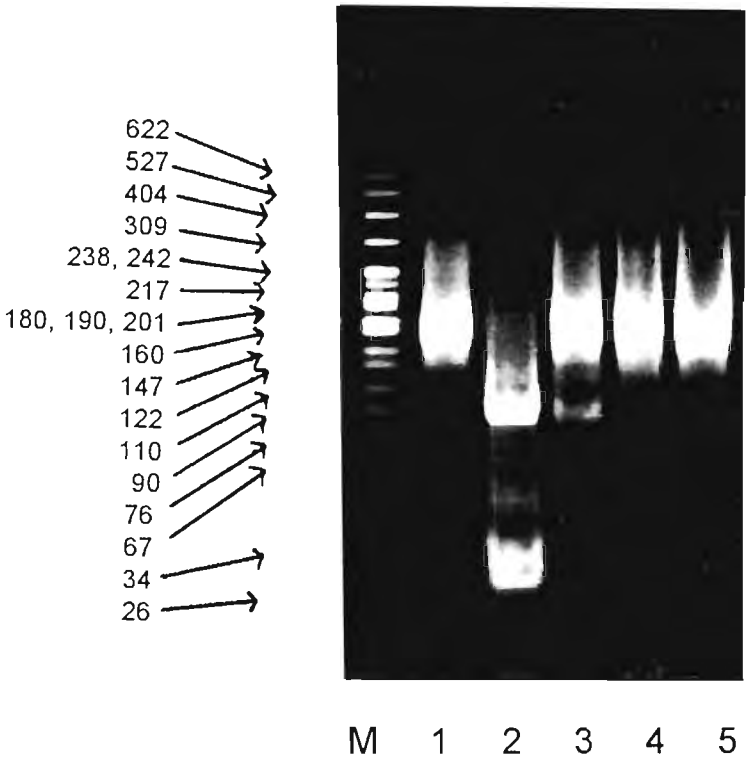
M = pBR322-*Hpa* II digest

(numbers with arrows represent base pairs)

To determine if the PCR product contained the correct sequence, a restriction digest was carried out with enzymes known to cut the insert. Restriction sites other than those known to occur (*Eco* RI and *Avr* II) were located using the program *GCG*. Three cleavage sites for the restriction endonuclease *Hinf* I were predicted, which should generate four fragments of 12, 23, 26 and 32 base pairs (bp) each.

Figure 10

12% non-denaturing PAGE of restriction analysis of the PCR product



M pBR322-*Hpa* II digest

1. uncut PCR product
2. *Hpa* II digest
3. *Eco* RI digest
4. *Avr* II digest
5. *Eco* RI - *Avr* II double digest

(numbers with arrows represent base pairs)

In figure 6, a slight downward shift in the mobility of fragments resulting from a double digest with *Eco* RI/*Avr* II is visible in lane 5. Single digests with each of these enzymes did not produce a visible shift. This is not surprising, as it is not possible to resolve such small changes in the molecular weight of the PCR product. In lane 3, a band appearing between 67- 76bp suggests a partial digest by *Hinf* I. Two much smaller fragments, representing the 26bp and 32bp fragments, are also visible in lane 3. The 23bp and 12bp fragments are not visible due to inability to resolve fragments of this size on a 12%

polyacrylamide gel. Even so, the electrophoretic mobilities of both uncut DNA and fragments generated by restriction digest, have the expected length.

The results for transformation were as follows:

Table 3.

	NUMBER OF TRANSFORMANTS	
ng of insert	"GENECLEANED" pPIC9 + INSERT	"NON-GENECLEANED" pPIC9 + INSERT
200	ND	8
600	1	4
1000	5	15
control 1	too numerous to count	
control 2	ND	
control 3	148	
control 4	10	

ND = non detected

Clearly the *Genecleaned* plasmid did not produce as many transformants as the plasmid that was not purified. *Genecleaning* was used to remove any contaminating supercoiled plasmid that remained from an incomplete digest with *Eco* RI and *Avr* II. This avoids the situation in which transformants are produced without the gene insert. It is possible that sodium iodide, used to dissolve agarose surrounding the DNA in this procedure, may have been present is trace amounts during ligation and thus have interfered with the action of T4 ligase. The controls implied the following:

control 1: competency of cells was satisfactory and the ability of pPIC9 to transform *E. coli* was good.

control 2: means of selection was satisfactory and no contaminating, ampicillin-resistant microorganisms were present.

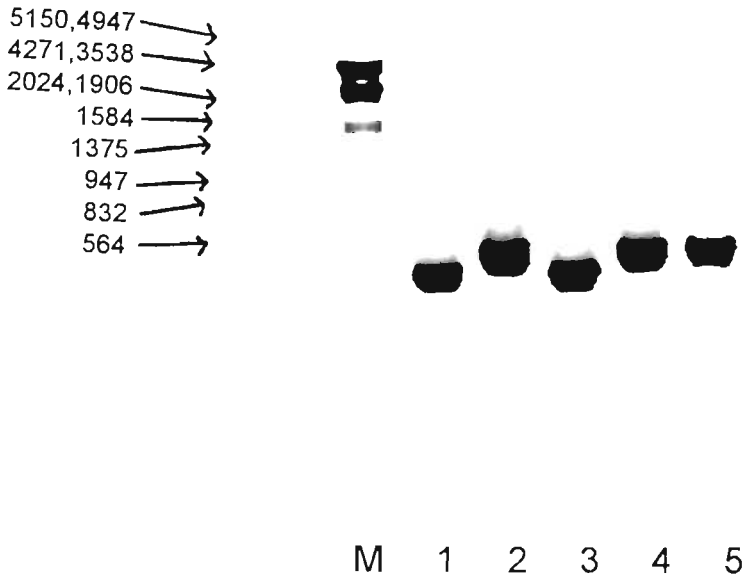
control 3: an amount of contaminating supercoiled plasmid was still present with linearised pPIC9, conferring ampicillin resistance on these (false-positive) clones. Thus the *Bgl* II digest was not complete.

control 4: showed the efficacy of using CIP: this process was 93% efficient, as calculated from the ratio of colonies produced by controls 3 and 4.

Four colonies were picked at random for PCR screening (denoted clones 1 to 4). Primers synthesised for this purpose were those recommended by *Invitrogen* (manual ver 3) for sequencing. These 21 mer oligos had sequence homology to areas within the *AOX1* promoter region, 3' and 5' to the MCS, such that the inclusion of an insert would cause a change in the electrophoretic mobility of the PCR product. Results of the PCR screen are shown below:

Figure11

1.5% agarose gel electrophoresis of PCR screen



M: bacteriophage lambda, *Eco* RI - *Hind* III double digest

1. pPIC9
2. clone 1
3. clone 2
4. clone 3
5. clone 4

(numbers with arrows represent base pairs)

Clones 1, 3 and 4 show the expected shift in electrophoretic mobility, suggesting these transformants contain the gene. (Note that the largest fragment resulting from the lambda digest (21221bp), did not appear due to the percentage agarose). Clone 1 was chosen for sequencing and was subsequently maxipreped for this purpose.

As the forward primer lay some 389bp upstream from the *Eco* RI site, difficulties were experienced with sequencing in this direction over the entire length. However, as the reverse primer lay just 100bp downstream from the *Avr* II site, a clear sequence could be read in this direction, running the gel for about 3 hours at 50V/cm. Sequencing from the reverse primer was repeated in order to identify GC artifacts and ensure the insert was in frame with the pPIC9 reading frame. The sequence could be read clearly upto 6 base pairs "outside" of the *Eco*RI and *Avr* II sites and was found to be correct and in frame (see below).

Figure 12.

Sequence read from the reverse primer in a 5' to 3' direction

5' end *Avr* II

TTA AAT CCG CGG CCG *CCT AGG* CGA TTA ACC ACA GTA
TTA ATT CGC GGC CGC *CCT AGG* CGA TTA ACC ACA GTA
TTA ATT CGC GGC CGC *CCT AGG* CGA TTA ACC ACA GTA

ACC GTG TTC CAA ACA AAC ACA TTC AGC CAA ACA GTC
ACC GTG TTC CAA ACA AAC ACA TTC AGC CAA ACA GTC
ACC GTG TTC CAA ACA AAC ACA TTC AGC CAA ACA GTC

AGA GTC CTT CTT ACA TTC CAT CAA GAT TCT TGG ACA
AGA GTC CTT CTT ACA TTC CAT CAA GAT TCT TGG ACA
AGA GTC CTT CTT ACA TTC CAT CAA GAT TCT TGG ACA

GAC TCT CTT GTC ATC GTC ATC *GAA TTC* TAC GTA AGG
GAC TCT CTT GTC ATC GTC ATC *GAA TTC* TAC GTA AGG
GAC TCT CTT GTC ATC GTC ATC *GAA TTC* TAC GTA AGC
Eco RI

CTT CAG CTA
CTT CAG CTA
CTT CAG CTA
3'end

Key

first experimental sequence
second experimental sequence
expected sequence
codons in italix: recognition sites for *Eco* RI and *Avr* II

4.0 Cloning of the modified *CMTI I* gene into *P. pastoris* and selection of Mut^S clones

4.1 Summary

Plasmid DNA, containing the *CMTI I* gene insert (referred to as cmti-PIC9 in the following sections), was transformed into spheroplasts of *P. pastoris*. A novel method of transformation was used, whereby spheroplasts were electroporated in the presence of linearised, cmti-PIC9. Forty three transformants were selected by their ability to grow on agar lacking histidine. To identify clones of Mut^S phenotype, their rate of growth was studied on minimal media containing methanol. Nine of these transformants grew slowly on methanol and were selected as likely candidates for containing the gene inserted into the genome. Total DNA was isolated from these putative Mut^S clones and PCR amplification of genomic DNA carried out, using primers with homology to the *AOX I* locus. Three clones produced a PCR product of identical electrophoretic mobility to cmti-PIC9.

4.2 Materials and Methods

For media recipes see appendix I.

4.2.1 *Bgl* II digestion of cmti-PIC9

Plasmid DNA was prepared by the alkaline lysis method (*Current Protocols*, 1989, unit 1.7.). RNA was removed by incubation for 1 hour at 37°C with 100µg/ml of RNAse A. A restriction digest of 20µg of DNA with 2 units of *Bgl* II was then carried out for 1 hour at 37°C. Linearised cmti-PIC9 was subsequently purified by phenol-chloroform extractions and precipitated with ethanol/ 10% v/v 3M sodium acetate (c.f. section 3.2.6).

4.2.2 Preparation of a growth curve for *P. pastoris*

P. pastoris, strain GS115, was provided as a stab culture (*Invitrogen*). GS115 was streaked for single colonies on YPD agar and grown for 2 days at 30°C. Single colonies were used to inoculate 3 McCartney bottles containing 5ml of YPD each. Starter cultures were grown overnight, on a shaker, at 30°C and the following day, used to inoculate 3 conical flasks containing 50, 200 and 500ml of YPD. A 50ml culture of *S. cerevisiae* was also included in the study. Cultures were grown for 24 hours and 1ml removed every hour, from 8 to 24 hours post inoculation (hpi) (collected by repeating the experiment such that readings were taken from 0 to 12 hours and 12 to 24 hours). The optical density (600nm) was recorded for each sample and graphs plotted of absorbance versus hpi. Samples with an optical density greater than 2 were further diluted and the OD then multiplied by the dilution factor, in order to remain within the working range of the spectrophotometer.

4.2.3 Preparation of spheroplasts

100ml of spheroplast suspension was prepared according to the method of Cregg, J. M. *et al* (1985). In addition to this protocol, another 200µl of 0.3 mg/ml lyticase (*Sigma*) and 200µl of 1M DTT were added, 2 hours after the initial addition of enzyme. The degree of spheroplasting was established at regular time intervals as follows. Every 20 min, 1µl of culture was removed and diluted 1:100 with 5% SDS. The absorbance (800nm) of this solution was recorded and the degree of spheroplasting calculated using the formula:

$$\% \text{ spheroplasting} = 100 - [(OD_{800} \text{ at time t}/OD_{800} \text{ at time 0}) \times 100]$$

When approximately 70% spheroplasting was attained, the solution was divided in two and labeled as sample 1 and 2. Sample 1 was pelleted by gentle centrifugation at 1500rpm, 4°C for 2 min. The pellet was washed in 1M sorbitol, 10mM Tris (pH 7.5), 10mM $CaCl_2$, repelleted and stored on ice. Sample 2 was prepared for electroporation, by washing once with 200ml of ice cold water and twice with 200ml of 1M sorbitol. Spheroplasts of sample 2 were pelleted as above and also left on ice.

4.2.4 Preparation of carrier DNA for transformation

The sodium salt of type III DNA from Salmon testes (*Sigma*) was dissolved in 1xTE to a final concentration of 10mg/ml. The solution was stirred overnight at 4°C to ensure the DNA was completely dissolved and frozen in 1ml aliquots at -20°C. One aliquot was divided into four equal portions and sonicated for 7, 15, 30 and 60 seconds respectively. Samples were cooled on ice immediately following sonication to prevent heat denaturation. DNA was then extracted with phenol-chloroform and precipitation with ethanol/10% v/v 3M sodium acetate, (c.f. section 3.2.6). The optical density (260nm) was recorded for each sample, and 1µg run on a 0.8% agarose gel to visualise the extent of sheering. In this way, sonication conditions which produced DNA fragments between the desired molecular weights of 2-15Kb, were used to prepare a fresh sample of carrier DNA for transformation.

4.2.5 Transformation

Pellets of samples 1 and 2 were resuspended in 40% PEG 3350, 1M sorbitol, 25mM EDTA, (pH 8.0) in an approximate 1:1 (w/v) ratio of biomass to solution. Sample 1 was incubated with 50µg of carrier DNA and 1µg of *Bgl* II-cut cmti-PIC9, for 10 min at room temperature. Sample 2 was treated in a similar manner, but with just 100ng of *Bgl* II-cut cmti-PIC9 and no carrier DNA. At this stage, controls of 50µl spheroplasts were set up for each sample, without plasmid DNA.

Cells of sample 1 were harvested by centrifugation as above and the PEG solution aspirated off. The pellet was resuspended in a 10-fold volume of 1M sorbitol, 10mM Tris

10mM $CaCl_2$ (pH 7.5), and left for a further 10 min at room temperature to encourage the recovery of spheroplasts. Finally, cells were suspended in 10ml of molten RD agarose at 45°C which was spread on to RDB agar. Plates were incubated at 30°C until colonies appeared.

Standard electroporation curves were prepared in 1M sorbitol. 50µl aliquots of 1M sorbitol were placed in cuvettes and pulsed with 0.9, 1.2 and 1.5KV. For each voltage, characteristic decay times were obtained depending on the preset resistance, using a constant capacitance of 25µF. Resistances which gave decay times around 10ms were used for electroporation. 50µl aliquots of sample 2 were electroporated under the following conditions:

ALIQUOT	CUVETTE GAP SIZE, cm	KV/cm	RESISTANCE ohms	CAPACITANCE µF
1	0.2	0.9	400	25
2	0.4	1.8	600	25
3	0.4	1.8	800	25
control	0.4	1.8	800	25

Immediately following electroporation, 1ml of ice-cold 1M sorbitol was added to the cuvette to provide an osmotically stable environment for the recovery of the yeast. Cells were suspended in 10ml of molten RD agarose at 45°C, which was spread on to RDB agar. Plates were incubated at 30°C until colonies appeared.

4.2.6 Preparation of glycerol stocks

Transformants were assigned a number for identification, streaked on to YPD agar and grown for 2 days at 30°C. Single colonies were placed in testubes containing 1ml of YPD and grown overnight, shaken, at 30°C. Glycerol stocks were prepared by mixing 0.5ml of overnight cultures with 0.5ml of sterile, 50% glycerol, and stored at -70°C.

4.2.7 Screening transformants for phenotype

One ml cultures of each transformant were grown from single colonies as above. A Mut^s control strain was also included in the screen, containing the gene for bovine serum albumin (BSA) cloned at the *AOX 1* locus (*Invitrogen*). Two 10µl aliquots of each culture were placed in testubes, one of which contained 1ml of minimal dextrose media (MMD) and the other, 1ml of minimal methanol media (MMM). Cultures were grown to mid log phase and the amount of biomass formation compared by eye to that of the Mut^s control strain. Clones exhibiting turbidity of the broth identical to that of the control were selected.

4.2.8 Preparation of total DNA

Total DNA was isolated from the selected clones as follows: 2ml cultures were grown overnight in YPD at 30°C and the cells harvested by centrifugation at 12000rpm, 4°C, 3min. Total DNA was extracted by vortexing the cell pellet in a breaking buffer containing glass beads (*Current Protocols*, (1989) unit 13.11.3.). DNA was separated from proteins by repeated phenol/chloroform extractions and precipitation in ethanol/ 10% v/v 3M sodium acetate (c.f. section 3.2.6). The optical densities at 260nm and 280nm were recorded for each isolate. The 260nm reading was used to calculate approximate DNA concentration (1OD ≡ 50µg/ml double stranded DNA) and the ratio of A_{260}:A_{280}, was used to give a rough idea of protein still present.

4.2.9 PCR screen of genomic DNA

PCR was performed using the above total DNA preps to provide template DNA. Primers were those previously used for sequencing, having homology to 3' and 5' *AOX I* sequences. PCR was carried out using the polymerase *Thermoprime Plus* (*Advanced Biotechnologies*). Conditions were first optimised using cmti-PIC9 as template DNA. Controls were set up as follows:

positive control: cmti-PIC9

negative control 1: total DNA isolated from the BSA-secreting control

negative control 2: pPIC9

Reagents were mixed, diluted to a final volume of 50µl with dH_2O and covered with 50µl of paraffin prior to PCR. The reaction mix for each clone was as follows:

TEMPLATE DNA ng	FORWARD PRIMER pmols	REVERSE PRIMER pmols	dNTPs mM	$MgCl_2$ mM	10 x BUFFER µl	taq polymerase units
200	50	50	1.0	6	5	0.5

PCR was carried out as follows:

1. denature: 94°C for 20s
2. anneal: 49°C for 30s
3. extend: 72°C for 60s
4. final extension: 72°C for 300s

Steps 1 to 3 were repeated over 30 cycles with step 4 as a final extension.

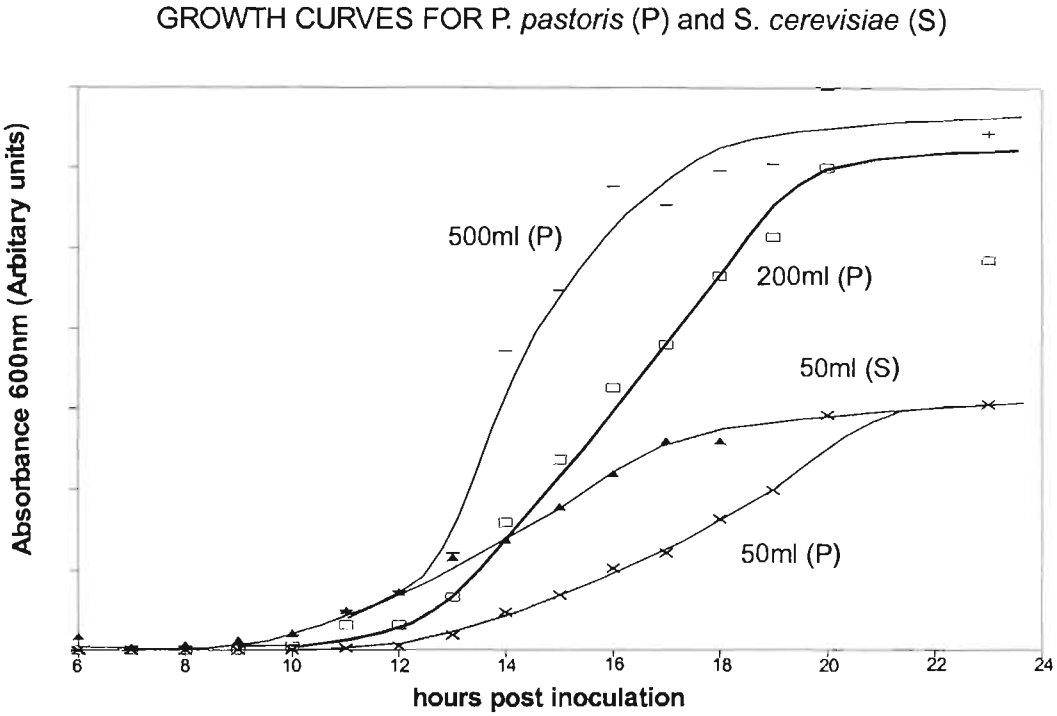
Following PCR, 10µl was removed from each tube and subjected to 1.5% agarose gel electrophoresis.

4.3 Results and Discussion

There are three techniques for transforming yeast cells: by rendering cells competent for the uptake of DNA using lithium acetate; by electroporating cells in the presence of DNA and by transforming spheroplasts. The choice of method will depend on the purpose for which transformants are intended, the strain used and availability of time and equipment. The majority of cloning experiments in *P. pastoris* have employed the spheroplasting technique. This was refined for GS115 by Cregg *et al* in 1985, based on the method developed for *S. cerevisiae* by Hinnen *et al* (1987). Examples of proteins expressed after spheroplast transformation of *P. pastoris*, include expression of the gene for bovine lysozyme (Digan, M. E., *et al*, 1989) and the tetanus toxin fragment C (Clare[1], J. J., *et al*, 1991). Electroporation has also been successful as a means of transforming *P. pastoris*, for example, in the expression of tick anticoagulant peptide (Laroche, Y., *et al*, 1994). The lithium acetate method has been less popular, especially in recent years. One group to use lithium acetate was Cregg, J. M., *et al*, (1987), in the expression of hepatitis B surface antigen.

For all three methods it is important to know the stage of cell growth and hence number of cells/ml, (indicated by the optical density of the culture at 600nm) before transformation is attempted. For *S. cerevisiae*, transformation using lithium acetate is optimal for a culture with an optical density (600nm) of 0.3-0.5. For electroporation the preferred OD lies between 1.3-1.5 and for spheroplasting 0.5-1.0 is recommended (*Current Protocols*, 1989, unit 13.7.1). There are few in-depth explanations for these differences, rather they refer to conditions investigators have found to work best for that particular method of transformation. Methods to transform *P. pastoris* are based on the techniques documented for *S. cerevisiae*, with little or no alteration. Yet it cannot be assumed that *P. pastoris* has exactly the same mean generation time or nutrient requirements. Thus it was considered necessary to study the growth patterns of *P. pastoris* prior to transformation. Figure 13 shows growth curves for *P. pastoris* (P) and *S. cerevisiae* (S):

Figure 13.



In order to attain the desired density of cells for transformation, different volumes of YPD were inoculated with the same volume of starter culture and the rate of biomass formation was studied. Both 50ml cultures did not appear to grow exponentially, suggesting the media volume to be inadequate to support logrithmic growth from an inoculum of this density. These graphs provided guide lines as to the culture volume required to attain the specified cells/ml from an overnight incubation. These results also indicate the mean generation time for *S. cerevisiae* is slightly shorter than for *P. pastoris* in this media, as shown by a shift to the left of the growth curve of *S. cerevisiae*. Studying the growth of *P. pastoris* enabled a general feel for the organism, and throughout this project the above graphs were frequently referred to.

Prior to spheroplasting, attempts were made to transform GS115 using the lithium acetate method. This was initially the method of choice because it is relatively short and simple, with transformation efficiency expected at 10^5-10^6 cells/μg transforming DNA for *S. cerevisiae* (*Current Protocols*, 1989, unit 13.7.1). However, every attempt resulted in a

smear of growth on selective plates. Contamination was unlikely as only yeast were detected under the microscope when a sample was examined. In addition, they had a characteristic yeast odor and yet distinct colonies, even after several dilutions, could not be obtained. After several attempts, spheroplasting was turned to as a second means of transformation. Interestingly, since that time, similar reports came to light at the *Invitrogen* conference in San Diego, CA, in March 1996, in which investigators claim to have experienced problems transforming *P. pastoris* via the lithium acetate method. Indeed, the updated manual (ver.E) from *Invitrogen* warns that the response of *P. pastoris* to lithium acetate has been found to be poor.

Spheroplasts are yeast cells from which the cell wall has been removed, leaving the cell membrane intact. These cells appear more rounded under the microscope compared to ordinary cells, and exhibit a "ghost" membrane. Spheroplasts have the ability to take up exogenous DNA. Preparing spheroplasts is a lengthy process requiring certain expertise, in addition to lyticase, an enzyme used to digest components of the cell wall, being an expensive enzyme. However, spheroplasting is the only technique for which transformants of *P. pastoris* containing multiple copies of the foreign gene have been reported. This makes it an attractive choice for transformation, as such clones have been shown to express recombinant proteins at higher levels than those containing just one copy (c.f. section 2.2).

Enzymes used to partially digest the cell wall of yeast, are aided by the presence of a reducing agent such as DTT or β-mercaptoethanol. A second addition of enzyme and 1M DTT was necessary because the rate of spheroplast formation was unsatisfactorily slow. After spheroplasting a number of times I found it was preferable to add a larger amount of enzyme for a shorter time, rather than incubating for longer with less enzyme, as cell lysis was apparent under the microscope after about 2.5 hours. However, transformation frequencies remained unsatisfactorily low and in an attempt to boost this, electroporation of spheroplasts was investigated.

The following figure shows the *Bgl* II cleavage of cmti-PIC9 was complete (lane 2). The larger fragment contained the gene insert. Uncut cmti-PIC9 (lane 1) exhibited two forms of plasmid DNA, which is quite usual.

Figure 14.

0.7% agarose gel electrophoresis of *Bgl* II digest of cmti-PIC9

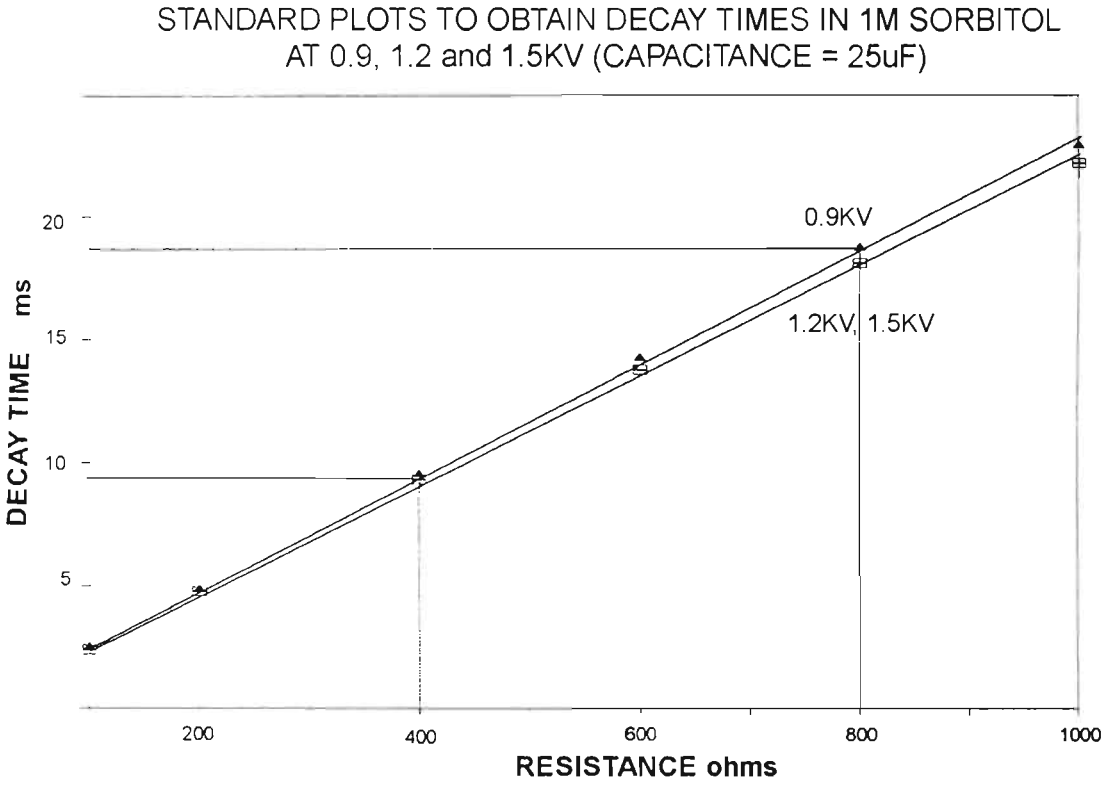
1 2

1. 1µg cmti-PIC9
2. 1µg *Bgl* II digested cmti-PIC9

Spheroplast stock solution was split in two in order that one sample be subjected to electroporation after the addition of transforming DNA. Uptake of DNA by spheroplasts is aided by the presence of carrier DNA (large fragments of double stranded DNA), and polyethylene glycol (PEG) of a specific molecular weight. Carrier DNA was not included in the transformation of sample 2 because it is known to inhibit uptake of plasmid DNA during electroporation. In addition, the amounts of transforming DNA differed: just 100ng of cmti-PIC9 was added to sample 2 as amounts in excess of this cause a drop in transformation frequency from electroporation (*Current Protocols*, 1989, unit 13.7.3). In order to keep all-things-equal, both samples were incubated with transforming DNA for 10 min, although it is unlikely that spheroplasts of sample 2 took up DNA at this point, due to a suboptimal amount of cmti-PIC9 and lack of carrier DNA.

Prior to transformation, standard curves for electroporation in 1M sorbitol were plotted. The following graphs show that changing the voltage had a negligible effect on decay time.

Figure 15.



Electroporation conditions were chosen to give decay times around 10ms, as recommended by *Invitrogen*. Thus a voltage of 0.9KV and resistance settings between 400 and 800 ohms were chosen (note that only resistance settings in multiples of 200Ω were possible with this apparatus). A control was included, in which spheroplasts without transforming DNA were electroporated at 800Ω.

Results of transformation of samples 1 and 2 were as follows:

Table 4.

SAMPLE	cmti-PIC9 ng	KV/cm	RESISTANCE Ω	DECAY TIMES ms	No COLONIES
1	1000	not electroporated	N/A	N/A	none
2(a)	100	0.9	400	6.9	32
2(b)	100	1.8	600	11.7	11
2(c)	100	1.8	800	13.3	none
control 1	none	not electroporated	N/A	N/A	none
control 2	none	1.8	800	13.3	none

Colonies were not recovered from sample 1 and the greatest number of transformants resulted from electroporation of spheroplasts from sample 2, at 0.9 KV/cm and 400Ω. The controls showed a) the selection procedure for transformants on agar lacking histidine was successful and b) no auxotrophic revertants or contaminants able to grow on the selective agar were present.

At the *Invitrogen* conference, March 1-4, 1996, opinion was still divided as to whether electroporation or spheroplasting gave the best results for transformation of *P. pastoris* (*personal communication*). Our results show that combining these two techniques is a viable option. However, it cannot be said whether transformants arose from the uptake of DNA into spheroplasts or merely ordinary cells remaining in the spheroplasting solution. Had sample 1 produced transformants, comparing the number of colonies from sample 1 and 2 would have shed some light on this. Indeed, the absence of colonies from sample 1 was surprising: certainly spheroplasts appeared viable under the microscope prior to transformation and it is possible that the amount of DNA used (1μg) was insufficient. To our knowledge, electroporation of spheroplasts has not been attempted before and may be

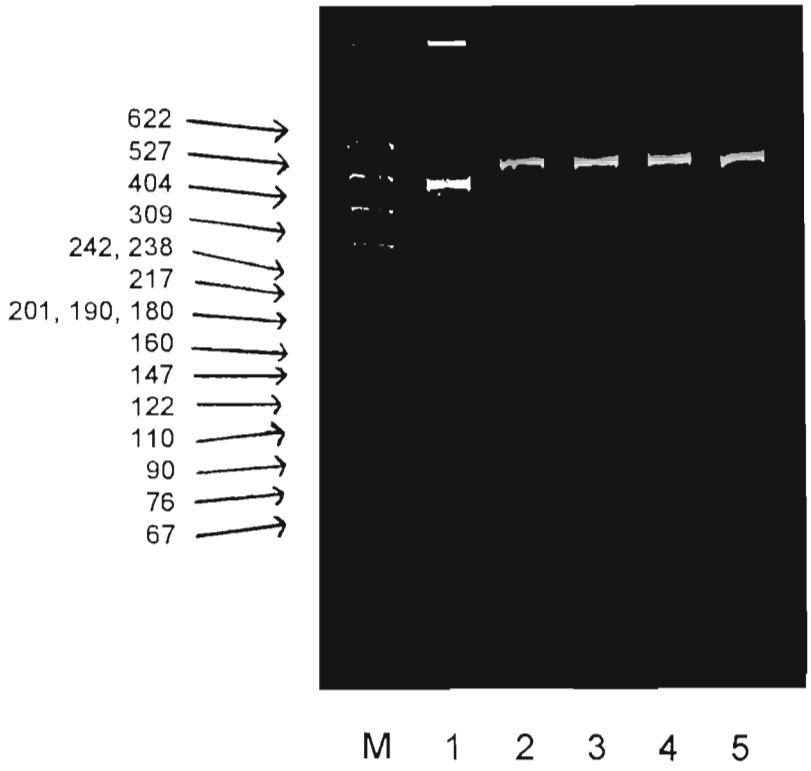
worthy of further investigation for strains which are difficult to transform by conventional means.

The initial screen for Mut^S transformants was carried out in broths, (Prof. Dowdle, *personal communication*) as we found the respective test using MM and MD agar (*Invitrogen* manual 3) produced a large number of false positive colonies. The inclusion of a control strain of *methanol-utilisation-slow* (Mut^S) phenotype was important in enabling the detection of clones of this type, as well as a means of establishing expression conditions in later experiments. The Mut^S clone provided by *Invitrogen* secreted bovine serum albumin (BSA) into the media. The *BSA* gene is inserted at the *AOX I* genomic locus via a transplacement event. To keep the experiment simple, we chose to initially screen for just Mut^S clones.

Of the 43 transformants, 9 displayed a similar rate of growth in MMM to the BSA-secreting control strain. As expected these tubes were almost clear, showing that the initial inoculum of 10µl had hardly grown, whereas cultures grown on dextrose as the food source (MMD) were opaic. These clones were considered as putative positives until the absence or presence of the *CMTI I* gene insert in the genome had been shown by PCR. Results of the PCR screen of genomic DNA are shown overleaf:

Figure 16.

10% non-denaturing PAGE of PCR screen of genomic DNA



M: pBR322 *Hpa* II digest

1. pPIC9 (no insert)
2. cmti-PIC9
3. clone 9
4. clone 21
5. clone 29

Clones 9, 21 and 29 (lanes 3, 4 and 5) produced a PCR product of identical mobility to the positive control in lane 2 (cmti-pPIC9). This showed that these three clones contained the gene inserted into the genome. DNA from the BSA-secreting control did not produce a product, which would have acted as a second negative control. As PCR conditions were optimised using cmti-PIC9 as template DNA, conditions may have been unsuitable for amplification of a much larger insert, this being some 30000bp for the *BSA* gene. The possibility of having obtained a product from amplification of episomal plasmid DNA was ruled out as *P. pastoris* does not have stable episomal vectors (c.f. section 2.2).

The PCR does not positively identify the site of integration, nor provide any information as to the possibility of multiple copies inserted into the genome. For this, hybridisation to a DNA probe by Southern blotting is required on both counts. However, the plasmid was directed to integrate at the *AOX I* locus, as cleavage with *Bgl* II would produce a fragment with 5' and 3' ends homologous to *AOX I* sequences. Further, as a Mut^S phenotype was observed, thereby implying disruption of the *AOX I* structural gene, it is highly likely that these clones did indeed arise from integration of the gene at the *AOX I* locus via a transplacement event.

5.0 Expression, isolation and identification of recombinant CMTI I

5.1 Summary

Expression of the cmti-PIC9 gene product from *P. pastoris*, was carried out in a defined medium, based on the method of Clare *et al*, (1991). Extracellular expression, directed by the yeast αMF signal sequence, enabled recovery of the recombinant product from the culture supernatant. The fusion protein, now referred to as *f*CMTI, was detected by spectroscopic assay of the culture supernatant for anti-tryptic activity, using N^{α} -Benzoyl-L-arginine-4-nitroanilide hydrochloride (BANA) as a substrate. A five step purification procedure was developed for the isolation of *f*CMTI. At each stage, the peak containing *f*CMTI was identified by assay for anti-tryptic activity. Amino acid analysis and protein sequencing were then used to confirm the identity of the isolate. A second batch of *f*CMTI was expressed and isolated in the same way, in order to process the *N*-terminus by cleavage at the enterokinase site. The modified product was isolated by HPLC and identified as recombinant CMTI I by protein sequencing and amino acid analysis.

5.2 Materials and Methods

For media recipes see appendix I

5.2.1 Expression conditions

Clones were streaked for single colonies on YPD agar and grown for 2 days at 30°C. A starter culture was grown for 36 hours at 30°C, by inoculating 5ml of YPD with three or four colonies. Cells were harvested the following morning by centrifugation at 3000rpm, 4°C for 10 min. The pellet was resuspended in 50ml of minimal glycerol broth (MGB) in a 250ml conical flask and shaken for 7 hours at 30°C. Cells were harvested as above and the pellet resuspended in 500ml of minimal methanol broth (MMB), in a two litre reagent bottle. The culture was aerated by pumping air through sterile perspex tubing, via a 0.22 micron filter to remove air-borne contamination. A second, sterilised tube allowed for the removal of carbon dioxide produced from aerobic respiration of yeast, and was also fitted at the end with a 0.22 micron filter. The culture was shaken for 4.5 days at 30°C, with the addition of 5ml methanol every 24 hours.

As a control, a second culture was set up on a smaller scale, in which the pellet of cells grown for 7 hours in MGB was resuspended in fresh MGB. Cells from both cultures were harvested by centrifugation at 12000rpm, 4°C for 10min. The culture supernatants were decanted and stored on ice prior to dialysis.

5.2.2 Time course assay using the BSA-secreting control (*Invitrogen*)

A small scale expression of BSA from the Mut^S control strain was carried out in accordance with the above protocol. 1ml aliquots of culture were removed at approximate 12 hour intervals, for 4.5 days post induction. Cells were harvested at 12000rpm for 1 min in a microcentrifuge. 100µl of each sample was reduced in volume by freeze drying, to an amount that could be loaded into the lanes of a gel (about 34µl), and mixed with 6µl of 6xSDS loading buffer (*Current Protocols*, 1989, unit 10.2.6). Samples were stored at -20°C prior to analysis by SDS-PAGE (c.f.

section 5.3.2). 10µl of "large molecular weight proteins" (*Sigma*) and 1µg of commercial BSA were used as markers. Gels were stained with Coomassie Blue prior to photography (section 5.3.2).

5.2.3 Dialysis and freeze drying

Cellulose ester dialysis tubing, with a cut off pore size of 1000D (*Spectra/Por*), was prepared by several washes in distilled water. 50ml aliquots of culture supernatant were dialysed against 5litres of dH_2O for 6 hours at 4°C, with changes of dH_2O every 2 hours. The contents of each bag were then poured into round bottomed freeze drying flasks, frozen by rotating the flasks in an ethanol bath at -20°C, and placed under vacuum for 24 hours.

5.2.4 Gel filtration

Sephadex (G50 fine) was swollen in an elution buffer of 20mM Tris-HCl (pH 8.0), according to the manufacturers instructions. The gel was degassed and poured into a glass column, 85cm long, with a diameter of 3cm. Sephadex was allowed to pack and a fiber glass filter then placed on top of the column. The column was sealed at either end producing a closed system and allowed to equilibrate in the elution buffer. The outlet tube was connected to a UV detector, which measured the optical density of the eluent at 229nm. The eluent was collected in testubes, in 3ml fractions, using an automatic fraction collector. The output from the detector was monitored on a PC using *HPLC* software (courtesy of Dr. D. Maeder, U.C.T.). Column parameters were obtained from the elution profiles for dextran blue and vitamin B12, which elute in the outer and inner volumes respectively.

Freeze dried solids were redissolved in a minimum amount of dH_2O and 6ml loaded onto the column at a time for chromatography. Fractions corresponding to the apex of each peak were freeze dried in testubes and the material obtained redissolved in 300µl dH_2O. 100µl aliquots were assayed for anti-tryptic activity (c.f. section 5.2.8) and an average rate of reaction determined.

5.2.5 Reverse phase HPLC using trifluroacetic acid (TFA)

The fraction from Sephadex column filtration which exhibited anti-tryptic activity was split into four equal portions and subjected to reverse phase HPLC on a *Vydac* C18 column, using the following buffers:

Buffer A: 0.1% TFA
HPLC-grade dH_2O

Buffer B: 0.1% TFA
90% acetonitrile

A gradient was run from 100% A to 100% B over 50 min and peaks collected in 2ml eppendorf tubes. Samples were dried down under vacuum and resuspended in 300µl of dH_2O. Each peak was then assayed for anit-tryptic activity (c.f. section 5.2.8).

5.2.6 Reverse phase HPLC using hydrofluorobutyric acid (HFBA)

The fraction exhibiting anit-tryptic activity from the first HPLC isolation, was subjected to reverse phase HPLC using the same column and the following buffers:

Buffer A: 0.1% HFBA
HPLC-grade dH_2O

Buffer B: 0.1% HFBA
100% acetonitrile

A gradient was run from 100% A to 100% B over 45 min and the major peak collected. The isolate was assayed for anit-tryptic activity (c.f. section 5.2.8), and characterised by amino acid analysis and sequencing.

5.2.7 Detection of anit-tryptic activity in culture supernatants

Bovine pancreatic trypsin (*Sigma*) was dissolved in dH_2O to a final concentration of 5mg/ml. A 10mM stock solution of N^{α} -Benzoyl-L-arginine-4-nitroanilide hydrochloride (BANA, *Merck*) was made, by first dissolving 43mg in 2ml of methanol, and adding dH_2O to a final volume of 10ml. The assay buffer was 100mM Tris-HCl, 20mM $CaCl_2$, (pH 8.0). 0.5ml of induced culture supernatant was placed in a 3ml quartz cuvette with 2.5µl of trypsin stock solution and 2.45ml of buffer. The solution was mixed and left at room temperature for 10 min, after which 50µl of BANA stock solution was added. A stopper was fitted to the cuvette and the contents

mixed by inverting 3 or 4 times. The cuvette was immediately placed in a Beckman DU 650 spectrophotometer and changes in optical density (405nm) measured every 0.5s over 200s. Assays were repeated in an identical manner with uninduced culture supernatants. The rate of reaction, measured as mols of nitroanaline released per second, was calculated from the change in absorbance/ min, using the formula of Lambert:

$$A_{450} = l.c.\varepsilon$$

where A_{450} is the absorbance of the solution at 405nm; l is the path length of the cuvette; c is the concentration of the substance being monitored and ε is the extinction coefficient in l. mol^{-1}. cm^{-1} .To obtain the amount of nitroanaline released in M/s, this equation was rearranged and expressed as follows:

$$\frac{dc}{dt} = \left(\frac{dA_{405}}{dt}\right) / 9620$$

Where the extinction coefficient of BANA, $\varepsilon_{405} = 9620$ (l. mol^{-1}. cm^{-1})
(*Merck*, E.C. 3.4.21.4)

5.2.8 Detection of anti-tryptic activity in fractions from gel filtration and HPLC

Fractions collected from column chromatography and HPLC, were assayed for the gene product by detection of anti-tryptic activity. In each case, fractions coinciding with each peak were freeze dried and redissolved in 300µl dH_2O. Three 100µl aliquots were then assayed for anti-tryptic activity in a volume of 1ml, using the procedure in section 5.2.7. The same final concentrations of the stock solutions were used. In order to minimise pipetting errors, an average rate was calculated from the three independent readings.

Just prior to the assay, the pH of the solution was checked with pH paper to confirm that it had not drifted away from 8.0. To check that no other substance interfered with

the activity of trypsin, control assays were carried out on fractions collected during a base line reading from chromatography, and the rate of reaction calculated as above.

5.2.9 Performic acid hydrolysis and amino acid analysis

The UV spectrum of the isolate was measured between 200 and 300nm. Protein concentration was then estimated from the optical density of 1 unit (230nm) ≅ 3.4mg/ml protein. A volume containing approximately 5nmols was placed in an acid-cleaned capillary tube and the sample freeze dried.

In order to measure the amounts of cystein, performic acid oxidation was carried out prior to amino acid analysis. Performic acid was made by mixing 5µl of H_2O_2 and 95µl of formic acid and allowing the reaction to proceed at room temperature for 2 hours. 2µl of performic acid was then added to the dried sample and left on ice for 3.5 hours. The oxidised sample was then diluted with 200µl of dH_2O and freeze dried.

The sample was dissolved in 100µl of dH_2O, and degraded to single amino acids by a 20 hour gas hydrolysis with boiling HCl. In order to prevent the degredation of amino acids, hydrolysis was performed in a sealed chamber flushed with nitrogen gas, to remove oxygen. The inclusion of 5% phenol was used to mop up any excess oxygen. One fifth of the hydrolysed sample was reacted with orthophthaldehyde, which fluoresces when bound to amino acids. Amino acids were separated by cation exchange HPLC and detected by monitoring excitation (338nm) and emission (425nm). Buffers used for HPLC were:

Buffer A: 0.2M sodium nitrate, pH 3.1

Buffer B: 0.02M sodium borate, pH 9.5

Elution times were standardised using 1nmol of each amino acid. 1nmol of Norleucine was included in each analysis as an internal standard. A blank run was also performed to enable base line subtraction of dirt peaks.

5.2.9.1 Protein sequencing

Sequencing of the *N*-terminal amino acids was carried out using Edman degredation (Edman, P. and Begg, G., 1967). 20pmol of purified inhibitor was dissolved in 15µl of dH_2O and the *N*-terminal amino group coupled to phenylthiohydantoin on a solid support. Sequential degredation of amino acids from the *N*-terminus was performed in acidic conditions. The *N*-terminal residue could be identified during each round of degredation by its elution time from HPLC.

5.2.9.2 Enterokinase cleavage of *f*CMTI

One tube containing 30µg of bovine enterokinase (*Bohringer Manheim*) was dissolved in 30µl dH_2O to give a concentration of 1mg/ml. 30µl of enterokinase solution was added to approximately 200µg of *f*CMTI dissolved in a buffer of 50mM Tris-HCl, 0.5% Tween 80, (pH 8.0). The solution was incubated at 37°C for 15 hours, after which the reaction was stopped by placing the tube in boiling water for 5 min. The cleaved *f*CMTI product was then purified by reverse phase HPLC using 0.1% HFBA as the ion pair (c.f. section 5.2.6). The fraction corresponding to the major peak was freeze dried, resuspended in dH_2O and characterised by amino acid analysis and sequencing.

5.2.9.3 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE)

SDS PAGE was carried out using a modified version of the method developed by Laemli (1970). (c.f. *Current Protocols*, (1989), unit 10.2). Gels were run at 5V/cm until the tracking dye entered the bottom tank buffer. To minimise over-heating gel plates were cooled with a fan. Bands were visualised after soaking for three hours in Coomassie Blue and destaining overnight in 25% ethanol, 10% acetic acid (*Current Protocols*, 1989, unit 10.2.6).

5.4 Results and Discussion

It is generally accepted that expression levels from *P. pastoris* are largely protein specific (*Invitrogen* conference, March 1996, *personal communication*). As such, expression conditions should be optimised for each protein and in doing so special attention must be paid to the composition of the media. Clare[2] J. J. *et al* (1991), investigated the effects of media composition on levels of extracellular expression of the mouse epidermal growth factor (mEGF), from both *S. cerevisiae* and *P. pastoris*. In both cases it was shown that yields of mEGF were greater from expression in rich media than in defined media. For *P. pastoris* they found that adding cas amino acids (hydrolysed casein) to the defined media, or lowering the pH, resulted in a reduction in proteolytic activity and consequently an improvement in yield. It is believed that small polypeptide chains in rich media, (provided by yeast extract and peptone), divert the activity of these extracellular proteases. Such "decoy" substrates were provided by the addition of cas amino acids to their defined media.

Attempts to express *f*CMTI were first carried out in the rich induction medium; buffered methanol-complex medium, (BMMY, *Invitrogen* manual 3). During the isolation procedure an oily brown residue was obtained, which was difficult to remove. This residue derived from the complex components of BMMY (i.e., yeast extract and peptone) and difficulties were envisaged in separating these substances from a small protein such as *f*CMTI. To overcome this problem, a medium of known composition was used instead. The major component of this medium was yeast nitrogen base (YNB), which contains all the essential amino acids, vitamins and salts to support the growth of yeast. The media and method adopted were based on those developed by Clare[1], J. J. *et al* (1991). Cas amino acids were however, excluded, as this component would cause problems similar to that of BMMY.

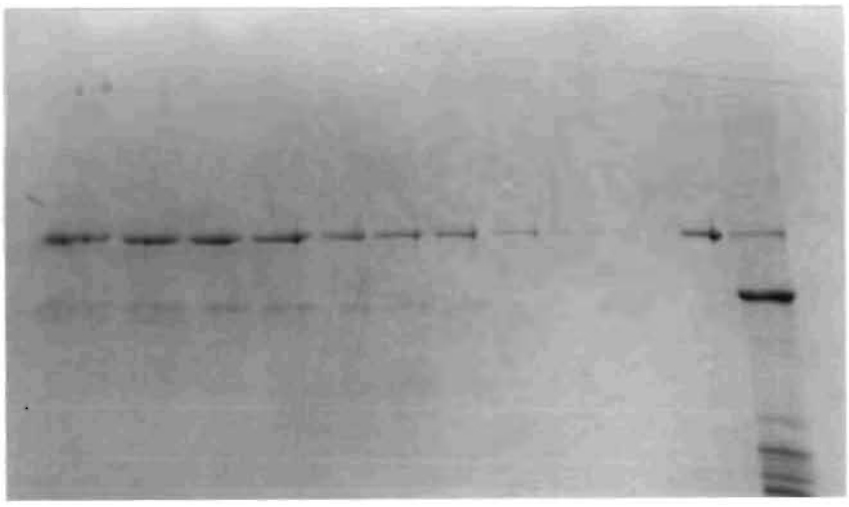
Cregg *et al* stated: "we have found that induction is highly sensitive to factors such as agitation rate and dissolved oxygen concentration." This finding was echoed at the *Invitrogen* conference in March of this year (*personal communication*), where the importance of using baffled flasks to increase surface area for oxygenation, was

agreed upon. Due to the unavailability of this apparatus, our system involved pumping sterilised air into a culture, which was agitated by placing the bottle on a shaker. In this way adequate aeration was provided for the needs of expressing cells.

In order to verify conditions as being suitable for expression, a time course assay of the BSA-secreting Mut^s control strain was carried out. Figure 12 shows a 10% SDS PAGE gel containing samples of induced culture supernatant taken at approximate 12 hour intervals, for up to 4.5 days after the initial induction with methanol (c.f. section 5.2.2):

Figure 17.

10% SDS PAGE of time course assay



108 96 87 72 63 48 40 24 18 0 BSA M
hours post inoculation

M: molecular weight markers for SDS PAGE (*Sigma*): bovine serum albumin (66KD); egg albumin (45KD); glyceraldehyde-3-phosphate dehydrogenase (36KD); carbonic anhydrase (29KD); trypsinogen (24KD).

BSA: 1µg of commercial bovine serum albumin

Recombinant BSA did not appear until 40 hpi, showing this to be the minimum time required for expression of detectable amounts of BSA, using Coomassie stain. The amount of recombinant BSA produced after 108 hpi was calculated by comparing the results of a densitometric scan of this band and the band produced by 1μg of commercial BSA ((lane "BSA") scans not shown). This was determined as 4μg of recombinant BSA, which extrapolates to an estimated yield of 40mg/l. The results of the time course assay showed that conditions were suitable to induce satisfactory levels of expression from the Mut^s control strain. However, it could not be inferred from this data that production of *f*CMTI would necessarily result in the same yields.

As a point of interest, an extra band can be seen in each lane which did not appear in the BSA standard (Figure 17). Another gel, in which the bands were visualised by the more sensitive silver staining method, revealed a ladder of bands directly under each band of recombinant BSA (data not shown). These were likely to be degredation products from the action of extracellular proteases. The time course assay therefore served a dual purpose, by drawing attention to the problem of extracellular proteolysis of recombinant proteins expressed by *P. pastoris*. This has been documented in reviews by Cregg, J. M., *et al*, (1993) and Romanos, M.A., (1995).

Fortunately, native CMTI I is known to be extremely resistant to proteolysis by enzymes other than trypsin (Otlewski, J., 1993). A few cases of cleavage at different sites within the inhibitor do exist, for instance porcine pepsin was shown to cleave the Leu7-Met8 bond (Otlewski, J., *et al*, 1994). Even so, the inherent, protease-resistant nature of CMTI I allowed us to consider the possibility of proteolysis as a minor concern.

For the expression of recombinant proteins in *P. pastoris*, it is desirable to separate the stages of biomass formation and protein synthesis (c.f. section 2.2). For Mut^s clones, which grow slowly on methanol, (doubling time of 18 hours, *Invitrogen*) it was necessary to bulk up in terms of total biomass prior to induction. Accordingly, biomass was increased 10 fold at every stage. This provided a large number of cells, which would have an additive effect on yield. A dense starter culture was obtained

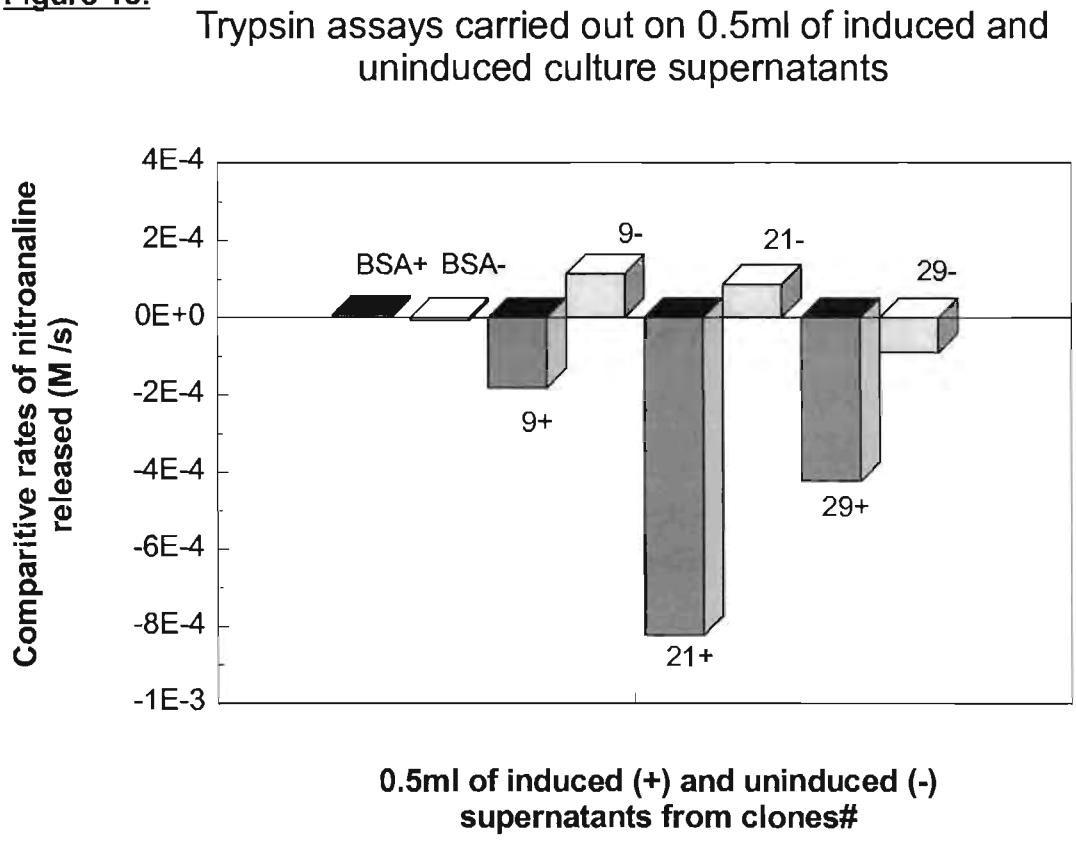
(about $1x10^{20}$ cells/ml) by growth of the initial inoculum for 36 hours in rich media. This was followed by a 7 hour incubation in defined media containing 2% glycerol as the carbon source. This would provide cells in the early log phase of growth (Figure 13, section 4.3), with the *AOX I* promoter tightly repressed. With the addition of methanol, cells would start expressing about 40 hours post induction, according to the results of the time course assay (c.f. Figure 17).

In order to detect *f*CMTI, attempts were first made to visualise the recombinant product using SDS PAGE methods developed for the resolution of small proteins. A time course assay was carried out under the conditions outlined in section 5.2.2 and SDS-PAGE performed, using a gradient of acrylamide between 12-26% of the polyacrylamide solution (DeWald D. B. *et al*, 1986). Samples were also run on a tricine-SDS gel, this discontinuous gel system being particularly useful for the resolution of proteins under 10KD. (Schagger H. and von Jagow G., 1987). Silver staining of either gel did not produce distinct bands under a molecular weight of 5KD. This was attributed to a large salt front obscuring bands in this region, caused by small components of the media present in each sample. To overcome this, 1ml aliquots of the culture supernatant were dialysed to remove media components, and reduced in volume by freeze drying prior to electrophoresis. Yet, a band approximating to the molecular weight of the cmti-PIC9 gene product could still not be visualised with certainty. Gel electrophoresis was therefore thought to be an unsuitable means of detecting this small protein, in as yet, unknown quantities.

Anti-tryptic activity was detected by screening the supernatants of expressing cultures, of the clones previously shown to contain a genomic copy of the cmti-PIC9 gene. Assays were carried out using N^{α} -Benzoyl-L-arginine-4-nitroanilide hydrochloride (BANA) as a substrate, and hence competitor, for trypsin (see next section for description of the trypsin-BANA interaction). For each clone, an uninduced culture was also set up to provide a control for any substance in the media which may have caused trypsin inhibition. An induced culture of the BSA-secreting Mut^{s} clone was also screened for anti-tryptic activity, as a control for possible trypsin inhibition arising from expressing cultures . The results are shown in the following

bar chart, in which an average rate from induced and uninduced culture supernatants of the Mut^s (*BSA*) control (which were near identical), was subtracted from the rates obtained from culture supernatants of clones 9, 21 and 29:

Figure 18.

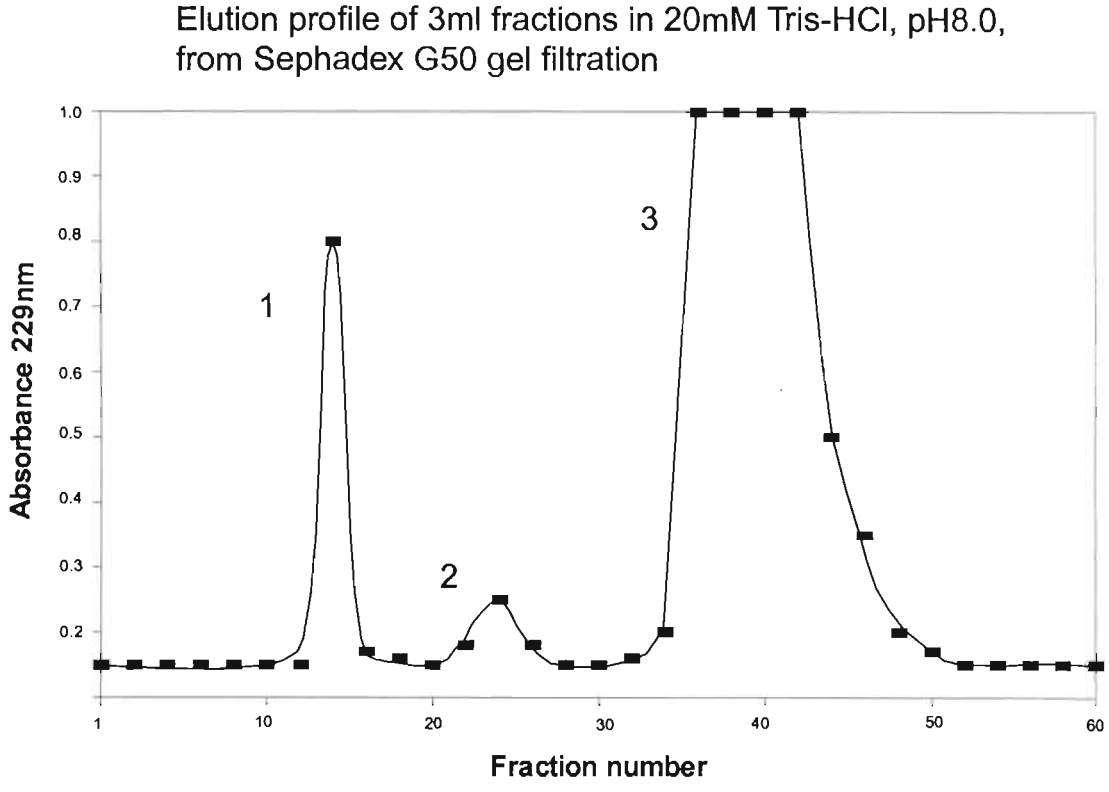


Induced culture supernatant from clone 21 (bar 21+) exhibited the greatest anti-tryptic activity. The controls (uninduced supernatants) showed some variability in rate of reaction, notably, uninduced clone 29 appeared to exhibit a low level of trypsin inhibition. However, this was considered to be within the experimental error of the assay, as it is important to realise that these assays were not carried out under analytical circumstances. Indeed, the assays in this section were not intended for any accurate measurement of inhibition but merely as a means of detecting the recombinant product. Nevertheless, the rates obtained were clear in showing clone 21 as the best candidate for expression of *f*CMTI.

An isolation procedure was developed, whereby (at each stage) the fraction exhibiting trypsin inhibition was purified until a homogenous protein was obtained. The first stage involved the removal of the majority of media components by dialysis. Gel filtration was then used to separate large proteins from the active fraction, as well as any remaining media. Finally, two sequential steps of reverse phase HPLC were used to purify the inhibitor to homogeneity.

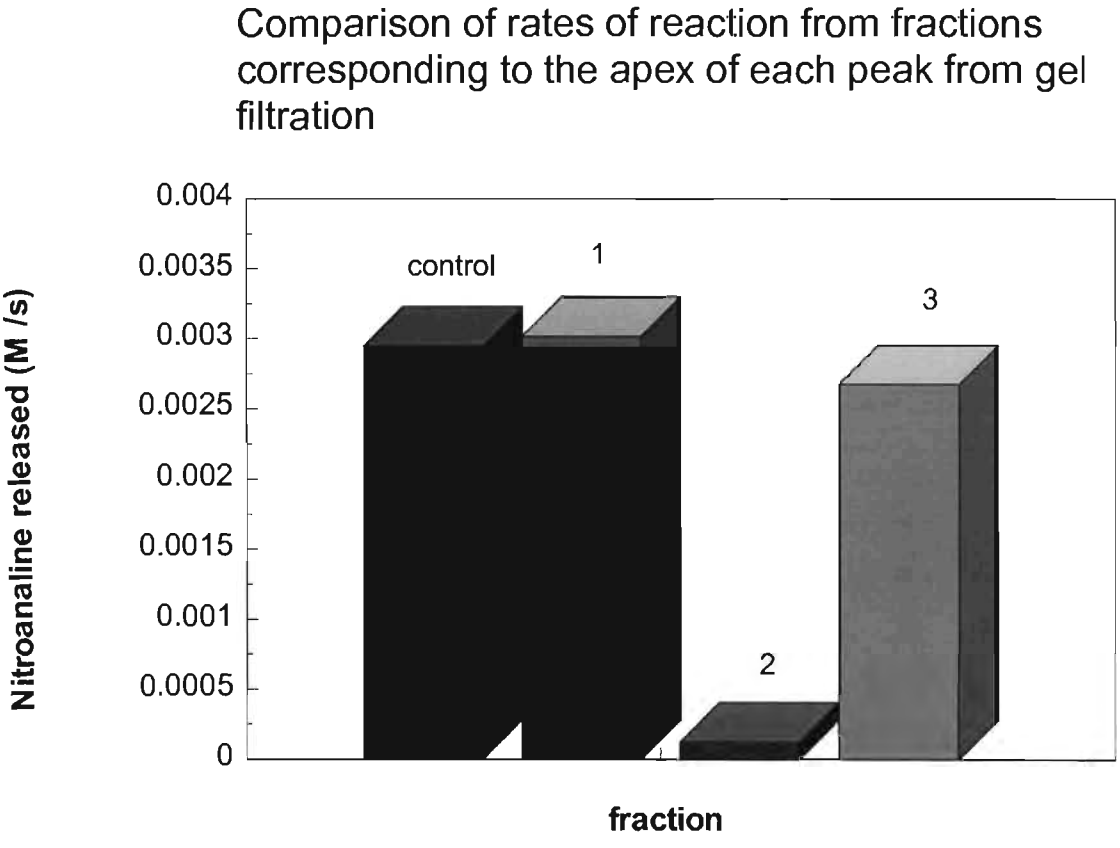
A defined medium was chosen to enable the separation, by dialysis, of *f*CMTI (M_r predicted as 3863D), from components in the media of 1000D and below. Dialysis was kept to a maximum of 6 hours, with frequent changes of water, to minimise losses of *f*CMTI. Following freeze drying of the dialysed solution, some media components were still apparent due to a distinctive yellow-green tinge. Subsequent passage through a Sephadex G50 column enabled separation of *f*CMTI from these remaining contaminants, as well as from the bulk of other proteins. The elution profile from gel filtration is shown below:

Figure 19.



Sephadex G50 fine was chosen in order to give the greatest resolution of proteins around the predicted molecular weight of *f*CMTI. Peak 1 eluted in the outer volume and contained large proteins of 30KD and above. Peak 3 eluted in the inner volume and probably contained salts and residual media components: as implied by the distinctive colour and odor of fractions 40 to 48. Peak 2 eluted between the inner and outer volumes, thus making it the prime candidate for recovery of *f*CMTI. Trypsin assays were performed on fractions corresponding to the apex of each peak. The rates of reaction obtained are compared in the following bar chart:

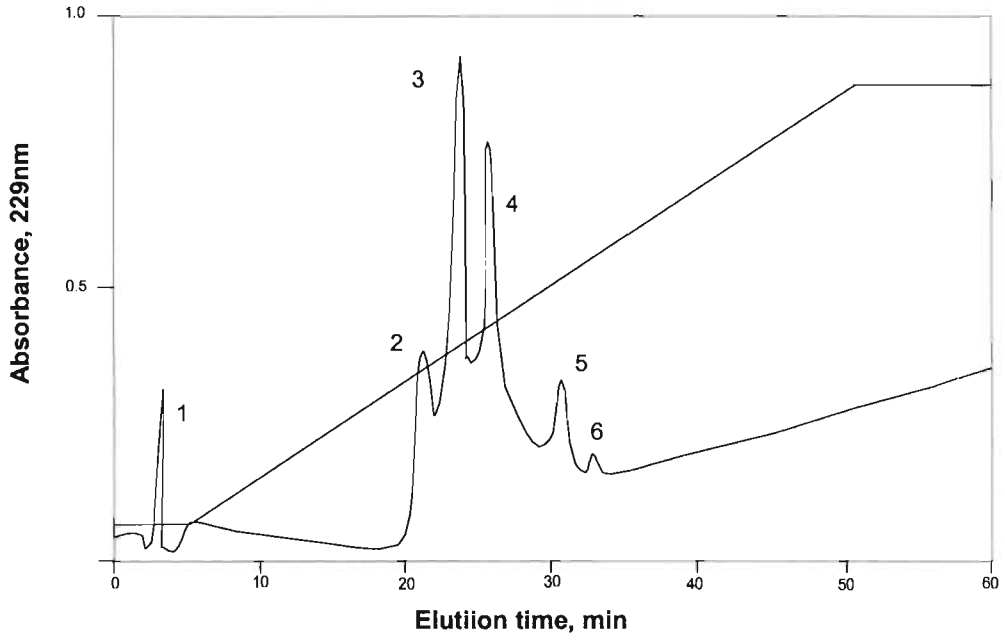
Figure 20.



It is clear that peak 2 contained a trypsin inhibitor(s). The control consisted of an assay carried out on fraction 1, which would have contained the elution buffer and any contaminants eluting at low levels from the column. Fractions 24 to 28 were then freeze dried and subjected to reverse phase HPLC, using 0.1% TFA as the ion pair. The following chromatogram was obtained:

Figure 21.

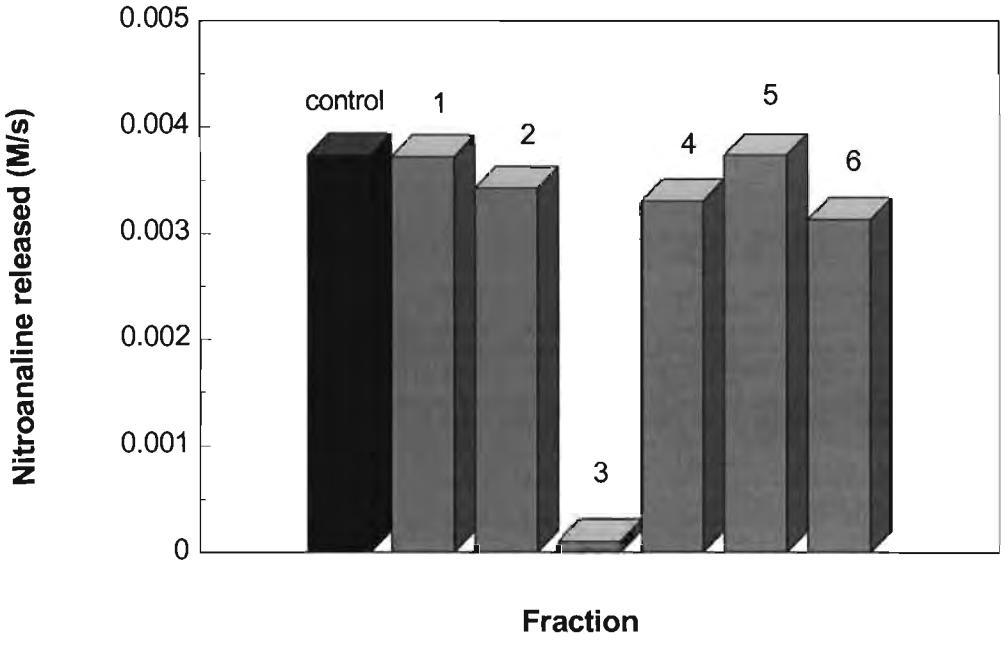
HPLC chromatogram of peek 2 from gel filtration



In order to identify the active fraction, peaks 1 to 6 were freeze dried and assayed for anti-tryptic activity. The rates of reaction obtained are compared in the following bar chart:

Figure 22.

Comparison of rates of reaction from fractions corresponding to peaks collected during HPLC

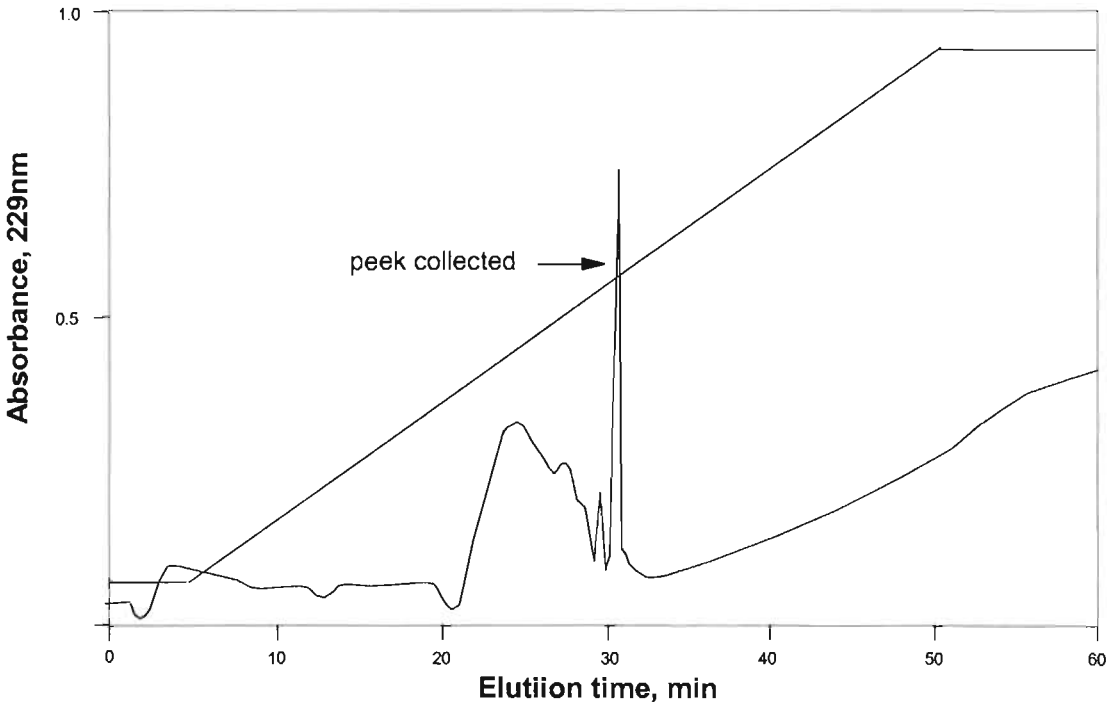


The peak exhibiting trypsin inhibition (peak 3) was easily identified. The control included an aliquot collected before commencement of the gradient (in buffer A), in which the absorbance had not moved from the base line. Peak 3 eluted 22.5 minutes in to the gradient, at 40.5% acetonitrile.

A further purification step was carried out on the active fraction, by submitting it to reverse phase HPLC using a different ion pair. The following chromatogram was obtained from HPLC using buffers containing 0.1% HFBA:

Figure 23.

Second HPLC purification of the active fraction



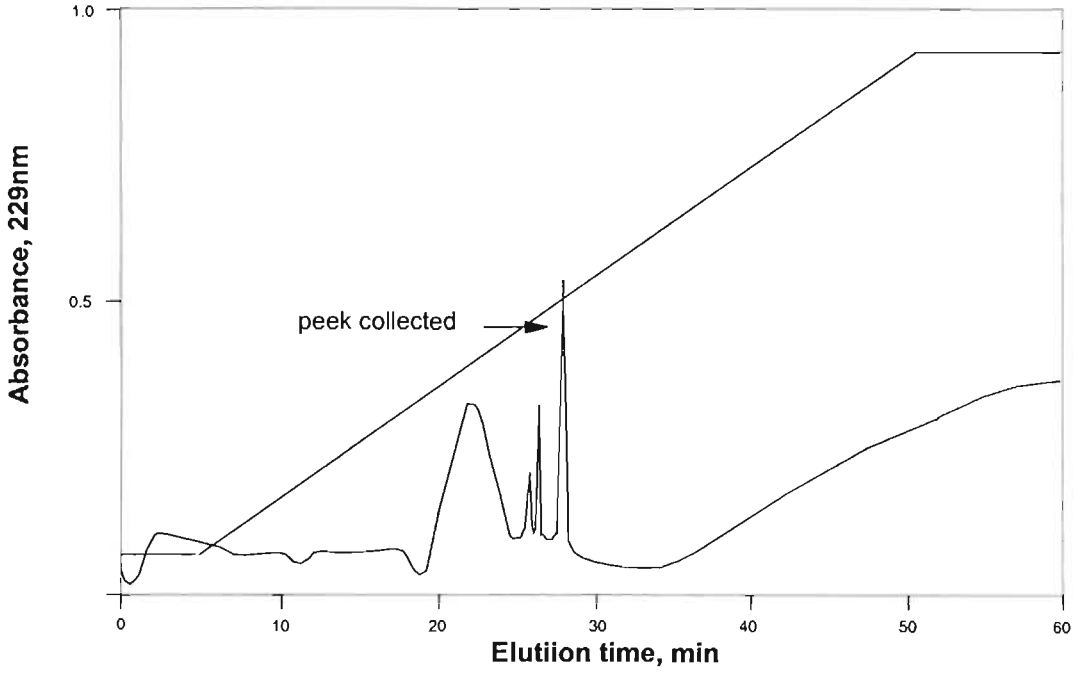
The major peak was collected, eluting 30 minutes in to the gradient at 66% acetonitrile.

Amino acid analysis was performed on the isolate as well as a sample of native CMTI I extracted from the seeds of squash (donated by Prof. J. Otlewski, University of Wroclaw, Poland). The analyses showed, within the experimental error (see section 6), that the protein isolated was indeed the recombinant fusion protein (c.f. figure 25).

The fusion protein was incubated with enterokinase to obtain recombinant CMTI I with the correct residues at the *N*-terminus. Enterokinase cleaves the peptide bond between the Lys of its recognition site and first residue (Arg) of CMTI I (c.f. figure 7, section 3.4). Another batch of *f*CMTI was expressed and isolated for this purpose. The products of enterokinase cleavage were separated by reverse phase HPLC, using 0.1% HFBA as the ion pair. Figure 19 shows the chromatogram obtained:

Figure 24.

HPLC chromatogram of the fusion protein after cleavage with enterokinase



The major peak eluted 28 minutes in to the gradient, at 62% acetonitrile. Cleavage was indicated by a reduction in the elution time of the major peak, suggesting this protein to be less hydrophobic. The extra peak (when compared to figure 23), eluting at 26 minutes was assumed to be enterokinase. The identity of the major peak was established by amino acid analysis and *N*-terminal sequencing of this fraction.

Figure 25 shows the results of amino acid analysis for *f*CMTI, enterokinase-cleaved *f*CMTI and native CMTI I. The results show the recombinant form of CMTI I had the same composition as the natural inhibitor, within the experimental error margin.

Prior to amino acid analysis, each protein was oxidised using performic acid in order to measure the number of cysteins present. Cystein must be converted to cysteic acid in order to survive the harsh conditions of hydrolysis. The amounts of cysteic acid measured in each analysis implied between 4-6 cysteins were present in each inhibitor, suggesting that some degredation of cysteic acid nevertheless occurred.

One notable discrepancy in the analyses was the absence of tyrosine in the native inhibitor. It can be assumed that this was due to degredation of tyrosine (an amino acid known to do so during this procedure), rather than to its actual absence. The harsh conditions of performic acid oxidation were also responsible for the absence of methionine and proline, which cannot be measured using this particular method of amino acid analysis. In addition, a dirt peak, which is known to vary in size, was responsible for the large area under the histidine peak in the native inhibitor.

Protein sequencing provided the first five amino acids at the *N*-terminus of each inhibitor. For *f*CMTI the sequence was:

Val - Glu - Phe - Asp - Asp

Cleaved *f*CMTI and native CMTI both had the following sequences:

Arg - Val - X - Pro - Arg

Where X represents an amino acid that could not be identified by sequencing. The native sequence of CMTI I was published in 1983 by Wilutz *et al,* in which the first five amino acids were found to be:

Arg - Val - Cys - Pro - Arg

As cystein is known to degrade under conditions of acid hydrolysis, it is likely that X was indeed cystein and that a recombinant form of CMTI I, with the same composition as the native protein, had indeed been synthesised.

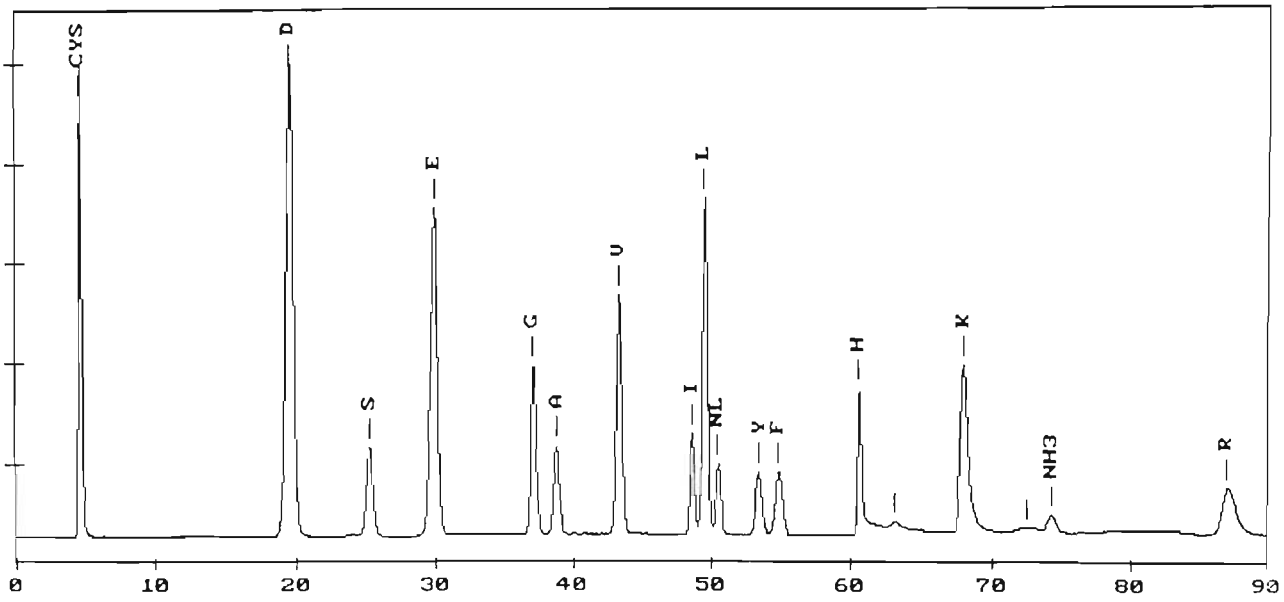
The occurrence of Val-Glu-Phe on the *N*-terminus of *f*CMTI was expected, as codons for these residues form part of the multiple cloning site of pPIC9, lying 5' to the first amino acid (Asp) of the *f*CMTI gene (c.f. figure 8, section 3.4). Interestingly, amino acid analysis of the batch of *f*CMTI produced for enterokinase cleavage revealed an abnormal amount of tyrosine, and on sequencing the *N*-terminus the following residues were obtained:

Tyr - Val - Glu - Phe - Asp

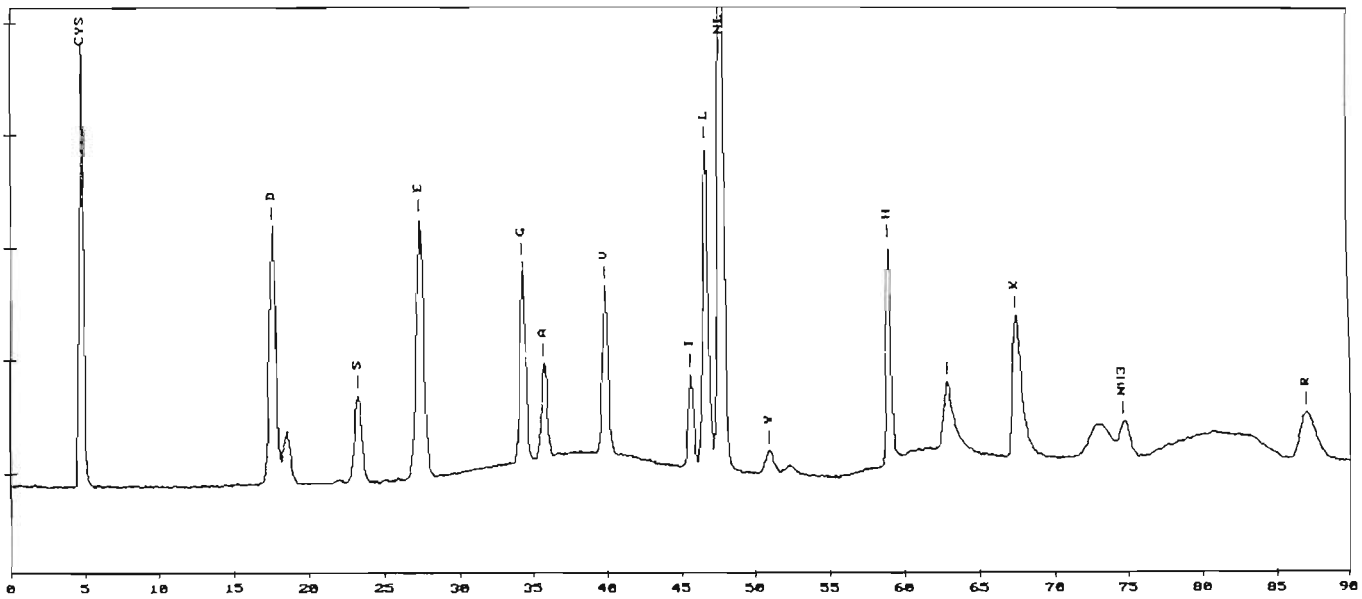
This sequence implied a variation in the site of cleavage of the *STE13* gene product, an enzyme produced by *P. pastoris* to trim extra amino acids remaining after Kex2 cleavage of the signal sequence (c.f. section 3.4). Cleavage was expected before the tyrosine in the latter sequence because it was preceded by Glu-Ala repeats: the recognition sequence for STE13. Why this variation in processing occurred can only be speculated at from the data obtained in this study. However, it can be said that expression of CMTI I as a fusion protein provided a degree of control in obtaining a protein with the correct *N*-terminal sequence at a later stage.

Figure 25.

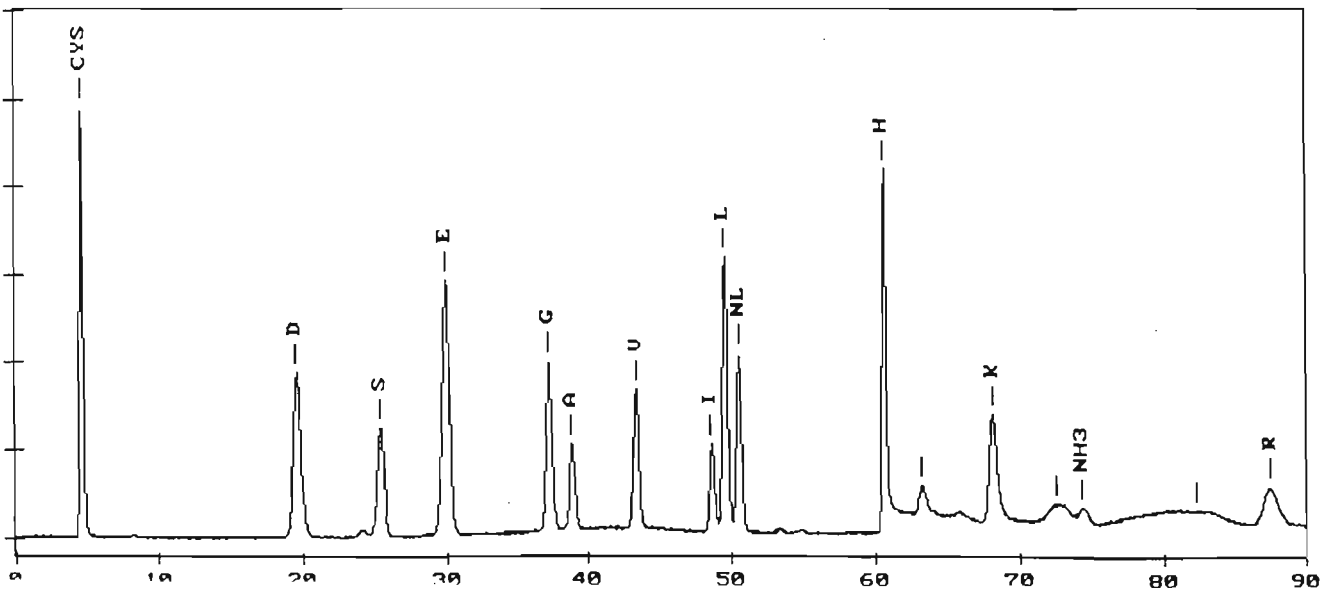
recombinant *f*CMTI



recombinant CMTI I



native CMTI I



6.0 Determination of Ki for *f*CMTI, recombinant CMTI I and native CMTI I.

6.1 Summary

Assays of trypsin activity were performed in the presence of *f*CMTI, recombinant CMTI I and native CMTI I, to obtain Ki for each inhibitor. Native CMTI I was used as a control, to which the activity of the recombinant inhibitors were compared. Spectroscopic assays were performed in an analytical manner using N-Benzoyl-L-arginine-4-nitroanilide hydrochloride as a substrate for trypsin. Lineweaver-Burke plots were constructed with increasing concentrations of inhibitor to determine the apparent Km values. Ki values for the three inhibitors were obtained from plotting apparent Km versus inhibitor concentration. Inhibition constants were: $3.3x10^{-7}$M, $2.11x10^{-7}$M and $2.17x10^{-7}$M for *f*CMTI, recombinant CMTI I and native CMTI I respectively. No reduction in activity was observed for recombinant CMTI I containing eight extra amino acids at the *N*-terminus. As Ki for the recombinant inhibitors are almost identical to that of native CMTI I, structural and folding determinants for their activity are presumed to be correct.

6.2 Materials and Methods

6.2.1 Stock solutions

10^{-5}M trypsin

24mg of pancreatic bovine trypsin * (*Sigma*) was weighed out to one decimal place and dissolved in 100ml of 100mM Tris/HCl, 20mM $CaCl_2$, (pH 8.0) in a volumetric flask. 2ml aliquots were stored at -20°C.

10^{-2}M BANA

43mg of N-Benzoyl-L-arginine-4-nitroanilide hydrochloride (*Merck*) was weighed out to one decimal place and dissolved, while stirring, in 2ml methanol. 100mM Tris/HCl, 20mM $CaCl_2$ (pH 8.0) was added to a final volume of 10ml in a volumetric flask.

Assay Buffer

100mM Tris/HCl, 20mM $CaCl_2$, pH 8.0.

Inhibitors

The concentration of each inhibitor was calculated from the results of amino acid analysis. A mean value in nmols was obtained from the amount of each of the amino acids, with the exclusion of those known to degrade during hydrolysis.

6.2.2 Instrumentation

All assays were performed on a Beckman DU 650 spectrophotometer, using the kinetics software therein to provide rates of reaction. This was given as change in absorbance (at the specified wavelength) per minute. Error margins for each rate were also provided based on regression analysis of the plot. Gilson pipettes were assessed for precision by weighing distilled water on a micro balance. The average error was calculated by regression analysis and factored back in to the pipette settings for all assays.

---

* **This product is given as being predominantly the α form of bovine pancreatic trypsin.**

6.2.3 Assays to obtain Km and Vmax for trypsin

Relative substrate concentrations between 1 and 10 were calculated to give values between zero and no greater than 1.5 Km for trypsin (Erlanger, B.F. *et al*, 1961). 100µl of trypsin stock solution was added to amounts of BANA corresponding to these concentrations in a quartz cuvette and the cuvette placed on a balance. Assay buffer was immediately added to the mixture until the weight of the solution reached 1g exactly. The cuvette was stoppered, inverted 3 or 4 times to mix the contents and placed in the sample holder of the spectrophotometer. Changes in absorbance (405nm) were read at 0.5 second intervals over 200 seconds, at room temperature (22°C). The change in absorbance per minute and standard deviations were recorded and saved to disk. The cuvette was rinsed with dH_2O between readings, and with 10% SDS to remove any residual trypsin which may have adhered to the glass. A reciprocal plot of rates of reaction against substrate concentration was used to obtain 1/Vmax and -1/Km (c.f. section 6.2.5).

6.2.4 Inhibition assays

A 1:10 dilution of inhibitor stock solution was made in distilled water. 100µl of trypsin stock solution was incubated at room temperature (22°C) for 10 min with a known amount of inhibitor. Five relative concentrations of BANA were calculated between 0 and 1.5 of the experimentally determined Km. A series of reactions were thus set up by adding increasing amounts of substrate to the pre-incubated trypsin-inhibitor mix. After addition of BANA to the cuvette, the volume was immediately made up to 1ml with assay buffer as above. Spectroscopic monitoring of the rate of reaction was carried out c.f. section 6.2.4. Apparent Km values were obtained from plotting 1/S versus 1/V.

6.2.5 Determination of Km and Ki

The Michaelis-Menten equation was used to determine the maximum velocity of reaction (Vmax) and the Michaelis constant (Km) for the interaction of trypsin with BANA:

$$\frac{V}{V_{max}} = \frac{[S]}{K_m + [S]} \qquad (1)$$

Where V (M/s) represents the rate of reaction, Vmax the maximum rate, Km (M^{-1}) the Michealis constant, and [S] substrate concentration. For convenience, this was rearranged into the Lineweaver-Burke equation to obtain linear plots:

$$\frac{1}{V} = \frac{K_m}{V_{max}}\frac{1}{[S]} + \frac{1}{V_{max}} \qquad (2)$$

Km can be obtained from the slope of equation 2 and the y-intercept as follows:

$$K_m = \frac{\text{slope}}{\text{y intercept}} \qquad (3)$$

The range of [S] was chosen to give values conveniently spaced between 0-1.5 Km for the reciprocal plot.

The following reaction scheme describes the mechanism of action for a competitive inhibitor:

$$E + S \Leftrightarrow ES \rightarrow E + P$$
$$+$$
$$I$$
$$\Updownarrow$$
$$EI$$

Where E is the enzyme; S the substrate; P the product; ES the enzyme-substrate complex; I the inhibitor and EI the enzyme-inhibitor complex. The inhibition constant Ki, is defined as:

$$K_i = \frac{[E][I]}{[EI]} \qquad (4)$$

From this model, the velocity equation for competitive inhibition under steady state conditions may be expressed as:

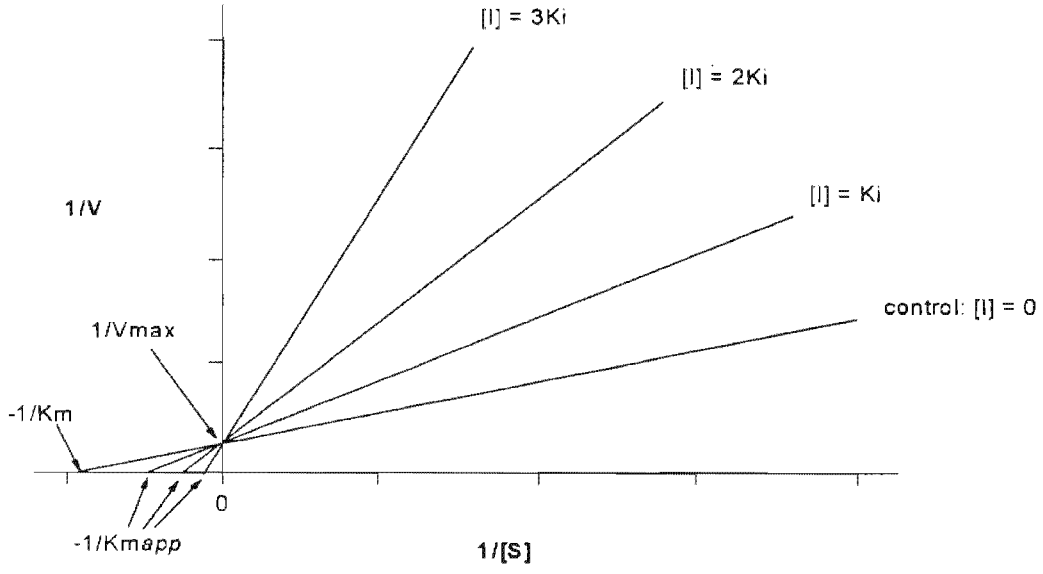
$$\frac{V}{V_{max}} = \frac{[S]}{Ks\left(1+\frac{[I]}{Ki}\right)+[S]} \qquad (5)$$

This equation may be rearranged into a reciprocal form as follows:

$$\frac{1}{V} = \frac{K_m}{V_{max}}\left(1+\frac{[I]}{Ki}\right)\frac{1}{[S]}+\frac{1}{V_{max}} \qquad (6)$$

This provided a number of apparent Km values from the intercept of each plot with the x axis (-1/Km apparent). The following diagram describes the effects of increasing concentrations of a competitive inhibitor on the velocity of reaction, in a reciprocal form:

Diagram 6.



Where each plot has a slope = Km app/Vmax, and:

$$Km_{app} = Km\left(1 + \frac{[I]}{Ki}\right) \quad (7)$$

or:

$$Km_{app} = Km + \frac{Km}{Ki}[I] \quad (8)$$

The standard derivations (sigma), were calculated by the method of Bevington, P. R. (1969), for Km and Ki. According to equation 8, a plot of Km app vs. [I] will result in a linear plot with a positive slope and y-intercept:

$$\text{slope} = \frac{Km}{Ki} \quad (9)$$

$$\text{y-intercept} = Km \quad (10)$$

It follows from equations 9 and 10 that the Ki values can be obtained from:

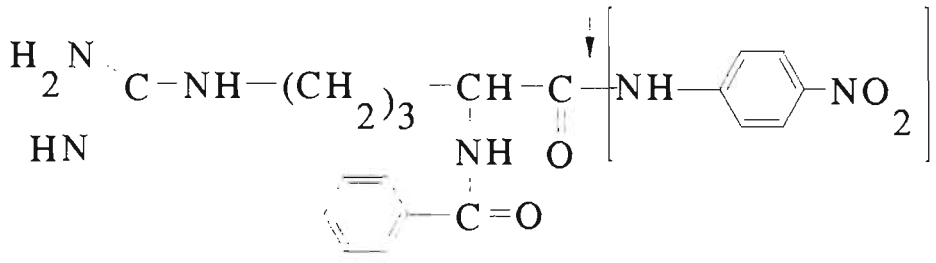
$$Ki = \frac{\text{y intercept}}{\text{slope}} \quad (11)$$

6.3 Results and Discussion

Kinetic experiments were carried out in order to compare the inhibitory properties of the two recombinant inhibitors, to that of native CMTI I. The binding characteristics of squash inhibitors rely primarily on conformation, which in turn is largely dependent on disulphide bridge formation and topology (c.f. section 1.3). This data would therefore show whether *P. pastoris* was a suitable expression host for the biosynthesis of these small, disulphide rich proteins in an active form. In addition, the effects on Ki of the extra amino acids at the *N*-terminus of the fusion protein, would be shown.

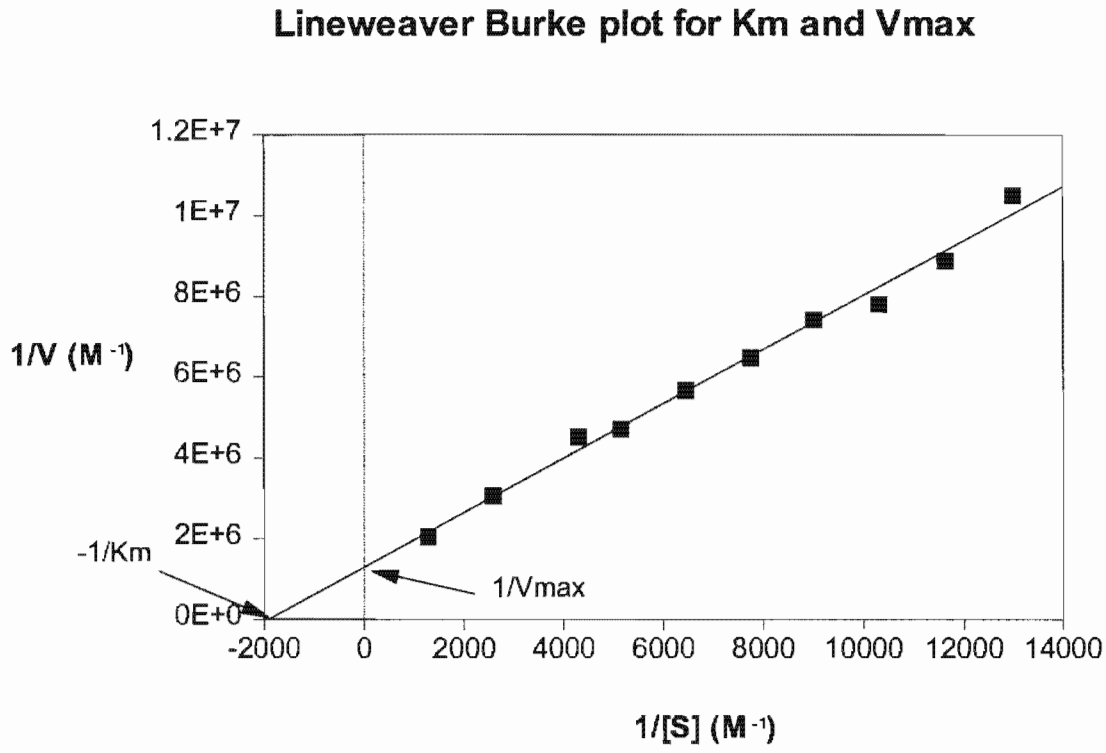
The first experiment involved a determination of the Michaelis constant (Km) for the reaction of bovine trypsin with N-Benzoyl-L-arginine-4-nitroanilide hydrochloride (BANA). The structure of BANA is shown below, the arrow indicating the site of trypsin cleavage. Cleavage occurs at the peptide bond of arginine, with the liberation of 4-nitroaniline (in brackets). This substance absorbs in the visible range of the spectrum at 405nm, such that the reaction mix became increasingly yellow.

Structure of BANA



Rates of reaction were measured with increasing amounts of substrate, the concentrations of which did not exceed 1.5 Km (Erlanger, B.F. et al, 1961). Amounts of substrate were chosen such that reciprocal values would produce evenly spaced points on the subsequent Lineweaver-Burke plots. Assays were performed in as uniform a manner as possible, with pipettes being calibrated prior to use. The Lineweaver-Burke plot for Km and Vmax is shown below:

Figure 27.



The rate of reaction was expressed as amount of nitroanaline liberated per second (M/s). Using equation 2, Km for trypsin was calculated as $52.2 \times 10^{-5}M \pm 6.23 \times 10^{-6}$.

Inhibitor concentrations were determined by amino acid analysis (c.f. section 5.2.9):

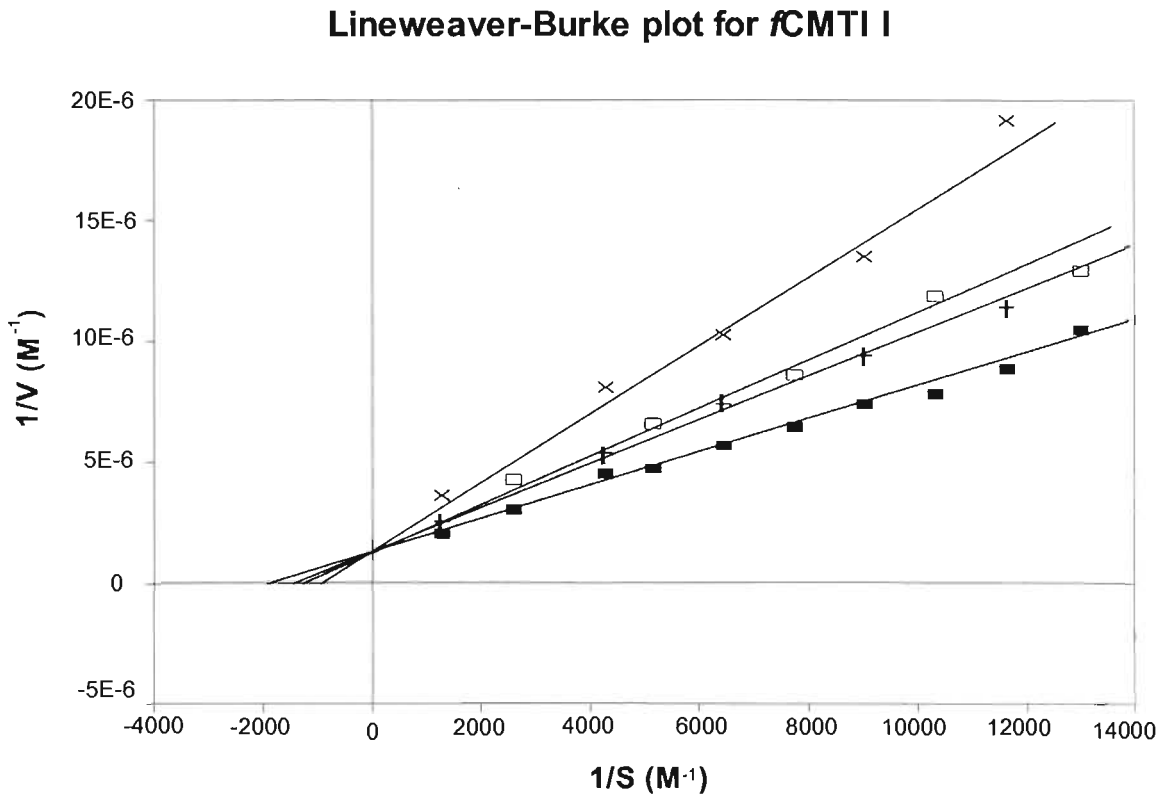
*f*CMTI: $7.0 \times 10^{-5}M \pm 1.72 \times 10^{-5}$

native CMTI I: $8.7 \times 10^{-5}M \pm 1.99 \times 10^{-5}$

recombinant CMTI I: $9.4 \times 10^{-5}M \pm 1.56 \times 10^{-5}$

To measure the activity of each inhibitor, rates of reaction were determined over a range of inhibitor concentrations and plots of 1/[S] vs. 1/V constructed. Lineweaver-Burke plots in the presence of *f*CMTI are shown in figure 28:

Figure 28.



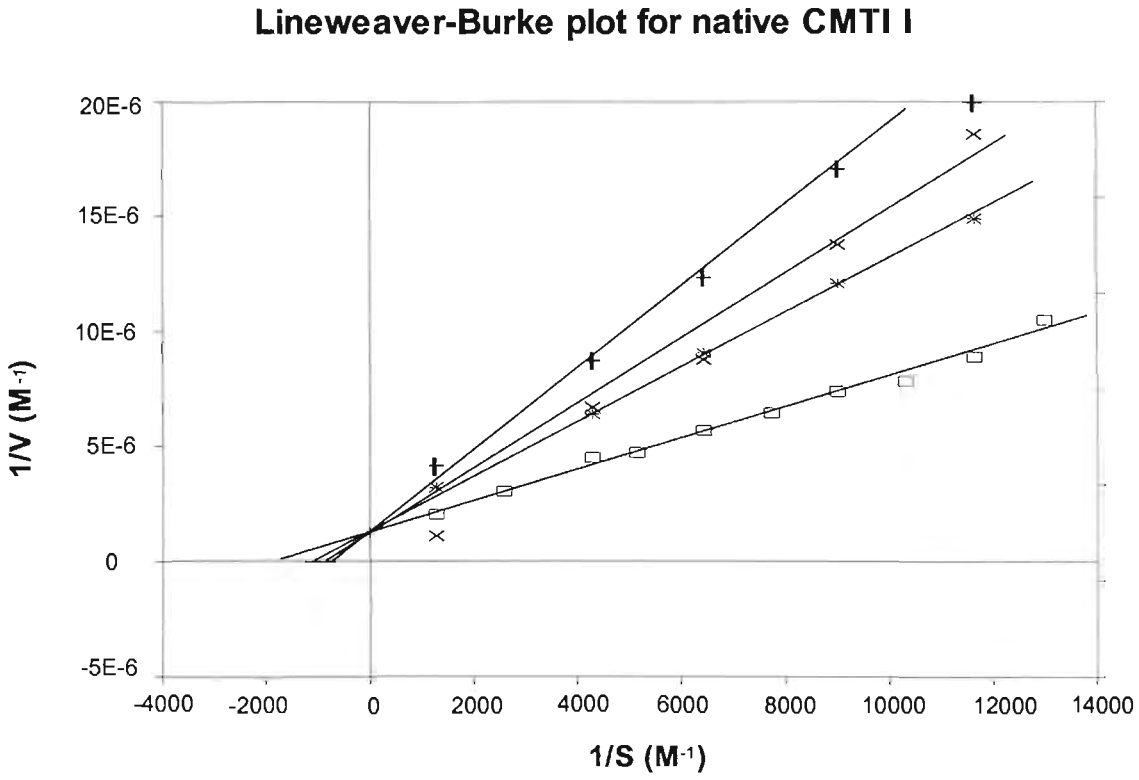
Apparent Km values with increasing concentrations of *f*CMTI are given below:

Table 5: Km apparent for *f*CMTI I.

µl of 1:10 inhibitor stock solution added	[inhibitor] M	Km apparent M	standard deviation sigma
0	0	5.22×10^{-4}	$\pm 6.23 \times 10^{-6}$
5	3.50×10^{-8}	5.78×10^{-4}	$\pm 1.45 \times 10^{-5}$
15	1.05×10^{-7}	6.95×10^{-4}	$\pm 1.30 \times 10^{-5}$
20	1.40×10^{-7}	7.41×10^{-4}	$\pm 1.19 \times 10^{-5}$

The experiments were repeated for recombinant and native CMTI I, after cleavage of the *N*-terminal peptide from *f*CMTI with enterokinase (figures 29 and 30):

Figure 29.



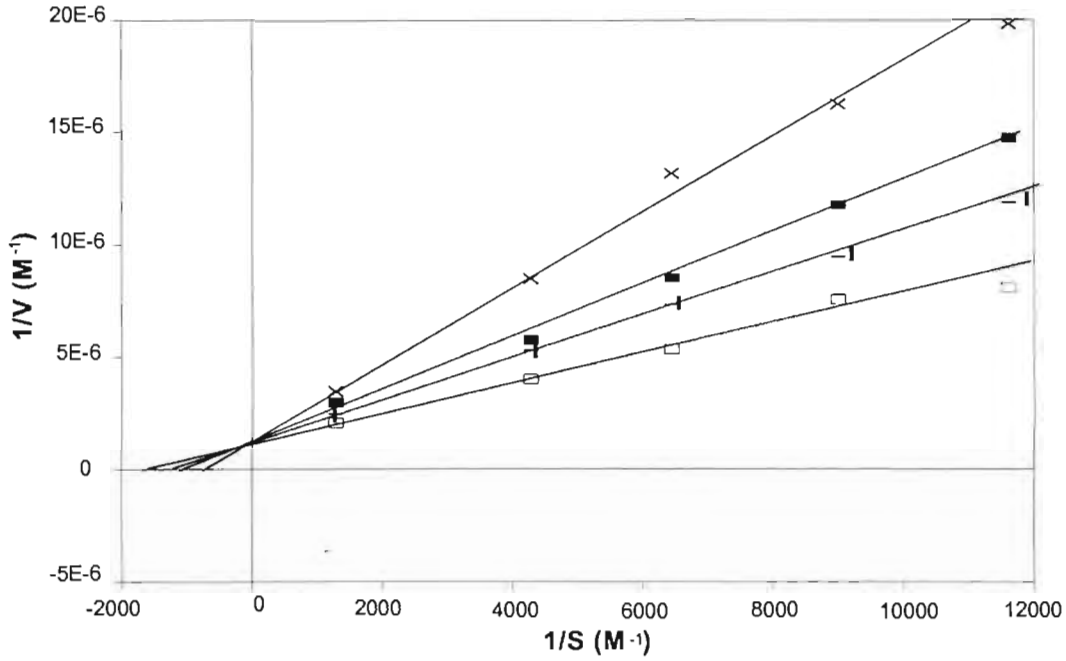
Km app values for the native inhibtitor are given below:

Table 6: Km apparent for native CMTI I.

µl of 1:10 inhibitor stock solution added	[inhibitor] M	Km apparent M	standard deviation sigma
0	0	5.22×10^{-4}	$\pm 6.23 \times 10^{-6}$
13.1	1.14×10^{-7}	6.70×10^{-4}	$\pm 2.0 \times 10^{-4}$
29.9	2.60×10^{-7}	9.19×10^{-4}	$\pm 6.3 \times 10^{-4}$
39.2	3.41x10-7	12.99×10^{-4}	$\pm 1.6 \times 10^{-4}$

Figure 30.

Lineweaver-Burke plot for recombinant CMTI I



Km app values for recombinant CMTI I are given below:

Table 7: Km apparent for recombinant CMTI I

µl of 1:10 inhibitor stock added	[inhibitor] M	Km apparent M	standard deviation sigma
0	0	4.1×10^{-4}	$\pm 1.1 \times 10^{-4}$
9.4	8.0×10^{-8}	6.49×10^{-4}	$\pm 3.0 \times 10^{-5}$
18.5	1.73×10^{-7}	8.15×10^{-4}	$\pm 1.7 \times 10^{-4}$
37.0	3.46×10^{-7}	11.83×10^{-4}	$\pm 5.4 \times 10^{-4}$

To obtain Ki values for the three inhibitors, apparent Km values, as predicted from the Lineweaver-Burke plots (c.f. tables 5,6,7), were plotted against inhibitor concentration. Figures 31, 32, and 33 depict the results for *f*CMTI, native CMTI I and recombinant CMTI I.

Figure 31.

Plot of Km vs. inhibitor concentration for the fusion protein (*f*CMTI)

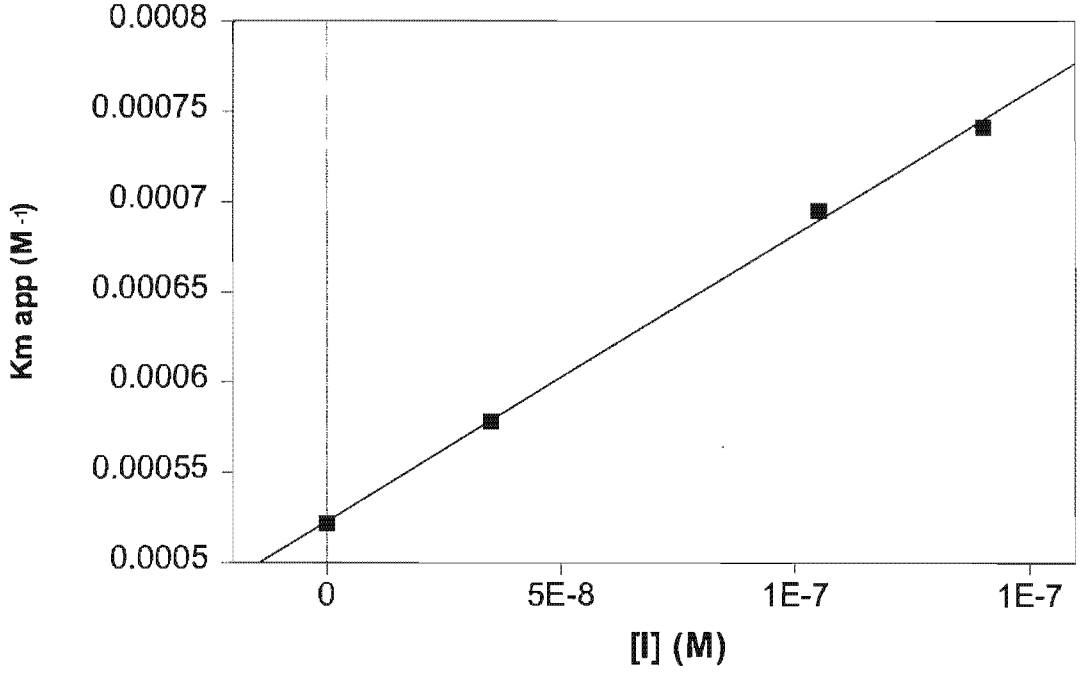


Figure 32.

Plot of Km vs. inhibitor concentration for native CMTI I

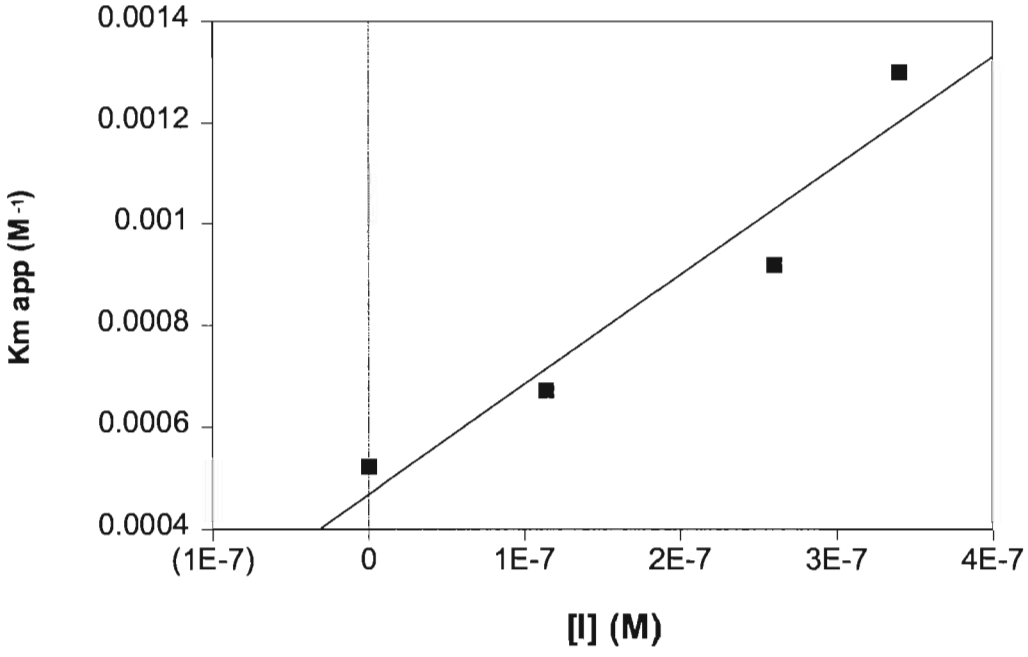
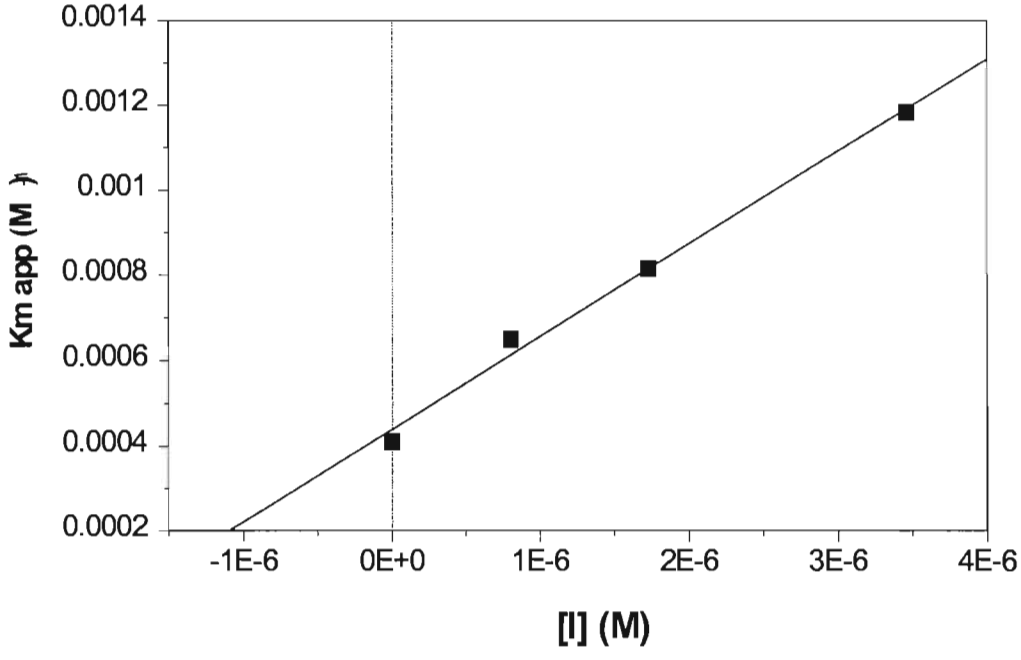


Figure 33.

Plot of Km vs. inhibitor concentration for recombinant CMTI I



Regression analysis of the above plots (figures 31, 32 and 33) enabled calculation of inhibition constants for each inhibitor according to eq.11. The results are given in table 8:

Table 8: Ki for *f*CMTI I, recombinant CMTI I and native CMTI I.

INHIBITOR	**Ki (M)**	**standard derivation (sigma)**
*f*CMTI	3.30×10^{-7}	$\pm 6.1 \times 10^{-9}$
recombinant CMTI I	2.11×10^{-7}	$\pm 4.2 \times 10^{-9}$
native CMTI I	2.17×10^{-7}	$\pm 9.1 \times 10^{-9}$

The standard derivations are very small and the correlation (R^2) values for each plot, which ranged between 0.924-0.999, were satisfactory.

The inhibition constants for all three inhibitors were very similar, especially Ki for the native and recombinant-cleaved forms. This infers that recombinant CMTI I was produced as a correctly folded protein and therefore the topology of the disulphide bridges matches the pattern of the native inhibitor (c.f. section 1.4). Had the recombinant protein contained mismatched disulphide bridges, or incomplete disulphide formation, Ki would have been much lower. This is supported by work done by Rolka, K., *et al* (1992), who synthesised three analogues of CMTI III* , each with a different cystein-cystein pair substituted with Gly and Ala. Elimination of any one disulphide bridge resulted in a decrease in the association constant by several orders of magnitude. Earlier, Kupryszewski *et al* (1985) had shown that two incorrect disulphide bridges of an analogue of CMTI III resulted in an inactive molecule.

*** CMTI III differs from CMTI I only at residue 9, which is Lys instead of Glu.**

ranging from 10^{-8} to 10^{-12} M. This may be due to the trypsin used in this study being an undefined mixture of both α and β forms, with a predominance of the former. The α form is known to have a lower affinity for BANA than the β form, (Schroeder and Shaw, 1968) which may explain this result. However, as the Ki values of the recombinant inhibitors were of the same order as that of native CMTI I, a comparison of their activity was still possible.

The extra residues on the *N*-terminus of the batch of *f*CMTI for which Ki was determined were:

Val-Glu-Phe-Asp-Asp-Asp-Asp-Lys

The results show that addition of this particular peptide to the *N*-terminus of CMTI I did not result in any marked reduction in the inhibition constant. It would be interesting to do structural studies on this elongated inhibitor to see how these residues interact with the core of the inhibitor and/or trypsin, if indeed they do. However, such research lies outside of the scope and intention of this MSc. thesis.

Concluding Remarks

Two active forms of CMTI I were synthesised by recombinant DNA technology in the methylotrophic yeast *P. pastoris*. Inhibition constants were determined as $3.3x10^{-7}$M, $2.11x10^{-7}$M and $2.17x10^{-7}$M for *f*CMTI, recombinant CMTI I and native CMTI I respectively. Comparison of these values strongly suggests that the recombinant inhibitors were folded in the correct manner. This answered the initial question posed regarding the ability of *P. pastoris* to sythesise and fold these small, disulphide rich proteins.

A few other squash inhibitors have been cloned. Xiao-Ming, C., *et al* (1992) cloned a synthetic gene for *Trichosanthes* trypsin inhibitor (TTI) mutated at position 6. This inhibitor was expressed both in *E. coli* and *S. cerevisiae*, the latter expression system producing superior yields of over 2mg/l. The inhibitor was first expressed as a fusion protein containing 3 extra amino acids on the *N*-terminus of the yeast-expressed protein, and showed inhibitive properties similar to natural TTI.

Wen, L., *et al* (1993) expressed CMTI V in *E. coli*, also initially as a fusion protein. CMTI V is actually 68 amino acids in length with just one disulphide bridge, and is classified as a pumpkin inhibitor (although still a member of the *Cucurbitaceae*). The fusion protein contained 7 extra amino acids on the *N*-terminus, and showed a reduction of 50% in inhibition towards its target enzyme (β-factor XIIa) as compared to native CMTI V. Bolewska, K., *et al* (1995) expressed a synthetic gene for a CMTI I analogue in *E. coli*, modified at the P'3 residue by a substitution of Met8 with Leu. The Ki for this analogue was the same as native CMTI I, and the yield obtained was 5mg/l.

It is difficult to give an accurate measurement of recombinant CMTI I produced here, as it was subjected to a four step purification, prior to obtaining the correct sequence by cleavage with enterokinase. From the absorbance (229nm) of peaks (peak 3) from the first stage of HPLC, the yield of the fusion protein was estimated as 21.4mg/l. This yield is typical of shake flask cultures of *P. pastoris* and advocates this system as an attractive alternative to producing squash inhibitors.

ABBREVIATIONS

BANA	N-Benzoyl-L-arginine-4-nitroanilide hydrochloride
cmti-PIC9	plasmid PIC9 containing the modified *CMTI I* gene
dH_2O	distilled water
DTT	dithiothreotol
hpi	hours post inoculation
Ki	inhibition constant (M^{-1})
Km	Michaelis constant (M^{-1})
MCS	multiple cloning site
MGY	minimal glycerol media (non-induction media)
MMD	minimal media containing dextrose
MMM	minimal media containing methanol
MMY	minimal methanol media (induction media)
oligos	oligonucleotides
RD	regeneration dextrose (agarose)
RDB	regeneration dextrose base (agar)
TCP	total cell protein
YNB	yeast nitrogen base without amino acids
YPD	yeast peptone dextrose media

Appendix 1

Bacterial and yeast strains

E. coli strain TOP10F' was provided by the *Invitrogen corporation*, San Diego, CA, USA., as a stab culture. TOP10F' genotype is as follows:

F' {*pro*AB, *lac*I^{q}, *lac*ZΔM15, Tn*10* (Tet^{R})} *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lac*ZΔM15, ΔlacX74, *deo*R, *rec*A1, *ara*D139, Δ(ara-leu)7697, *gal*U, *gal*K, *rps*L(Str^{R}), *end*A1, *nup*Gλ-

P. pastoris strain GS115 was provided by the *Invitrogen corporation*, San Diego, CA, USA., as a stab culture.

GS115 genotype is *his-4*.

plasmid PIC9

pPIC9 is a 8023bp fusion vector produced by the *Invitrogen corporation*. (For description see introduction, *Pichia pastoris as an expression host*).

Buffers

1xTE buffer

Diluted from a stock solution of 10xTE buffer

10xTE stock solution

100mM Tris-HCl, pH 7.6

10mM EDTA, pH 8.0

1xTBE

Diluted from a stock solution of 10xTBE

10xTBE stock solution

108g Tris base

55g Boric acid

40ml 0.5M EDTA, pH 8.0

dH_2O to 1litre

Media

LB broth

5g yeast extract powder

10g tryptone

10g sodium chloride

dH_2O to one litre

Broth was aloquoted into appropriate glassware, autoclaved for 15min at 14psi.

LB agar + ampicillin

LB broth made as above with the addition of 15g agar

Ampicillin from a 50mg/ml stock added as 1µl per ml of LB.

YPD (yeast peptone dextrose media)

10g yeast extract

20g bacteriological peptone

900ml dH_2O

Solution was stirred until all solids were dissolved, poured in to 1l reagent bottle and autoclaved at 14psi for 15min. For agar plates, 7g of agar was added prior to sterilisation. When cool, 100ml of 10% dextrose solution was added. YPD was stored at 4°C.

RD and RDB(regeneration dextrose top and base agar)

As taken from the *Invitrogen* manual verE.

For top agar, 186g of sorbitol was dissolved in 700ml of dH_2O and 10g agar added.

For base agar, 186g of sorbitol was dissolved in 700ml of dH_2O and 20g of agar added.

Both media were sterilised by autoclaving as above and placed in a water bath at 60°C.

The following reagents were then added and plates poured for RDB. Top agar was retained molten at 45°C for the addition of transformants following spheroplasting.

Media components added to each agar:

- 100ml of sterile 10% dextrose solution
- 100ml of YNB stock solution
- 2ml of biotin stock solution
- 10ml of amino acid stock solution
- 88ml of sterile water

Stock solutions

YNB without amino acids (*DIFCO*)

13.4g was dissolved in 100ml dH_2O by heating gently, and sterilised through a 0.2 micron filter under pressure. Stored at 4°C.

Biotin stock solution

20mg biotin was dissolved in 100ml dH_2O and filter sterilised as above. Stored at 4°C.

Amino acid stock solution

500mg of *L*-Glu, *L*-Met, *L*-Lys, *L*-Leu, and *L*-Ile were dissolved in 100ml dH_2O, filter sterilised and stored at 4°C.

MGB

For 200ml:

40ml (2%) of 10% glycerol

20ml of 10% YNB (yeast nitrogen bases with out amino acids (*DIFCO*))

400µl biotin

MMB

For 200ml:

20ml 10% YNB

20ml of 10% methanol

20ml of 1M phosphate buffer (*Invitrogen* manual ver 3)

References

Anthony. C., (1982) *Academic press*, New York.

Bayliss, W. M. and Starling, E. H., (1902) *J. Physiol 28*, 325.

Bennetzen. J. L., and Hall, B. D., (1982) *J. Bio. Chem. 257*, (6) 3026-3031.

Bevington, P. R. (1969) *Data reduction and error analysis for the physical sciences.* McGraw-Hill, N.Y.

Biemans, R., *et al*, (1991) *DNA. Cell. Biol. 10*, 191-200.

Bode, W. and Huber, R. (1992) *Eur. J. Biochem. 204*, 433-451.

Bode, W., *et al* (1978) *J. Mol. Biol. 118*, 99-112.

Bolewska, K., *et al*, (1995) *FEBS Lets 377*, 172-174.

Chung *et al*, (1989) *PNAS. USA. 89*, 2172-2175.

Clare[1] J. J., *et al* (1991) *Bio Tech 9*, 455-460.

Clare[1] J. J., *et al* (1991) *Bio/Tech 9*, 455-460.

Clare[2] J. J., *et al* (1991) *Gene 105*, 205-212.

Clore, G. M., *et al*, (1987) *Prot. Eng. 1*, 313-318.

Cregg, J. M. *et al*, (1993) *Bio/Tech 11*, 905-910.

Cregg, J. M., *et al*, (1993) Bio/Tech *11*, 905-910.

Cregg. J. M. *et al*, (1985) *Mol. Cell. Biol. 5*, 3376-3385.

Cregg. J. M., *et al*, (1987) *Bio/Tech 5*, 479-485.

Deisenhofer, J. and Steigermann, W., (1975) *Acta Crystallogr. B31*, 238-250.

DeWald, D. B. *et al*, (1986) *Anal Biochem 154*, 502-508.

Digan, M. E., *et al*, (1989) *Bio/Tech. 7*, 160-164.

Drenth, J. J. N., *et al* (1971) *The Enzymes* (ed. P. Boyer), 3rd ed., *3*, 484 *Academic Press N. Y.*

Edman, P. and Begg, G., (1967) *Eur. J. Biochem 1*, 80.

Erlanger, B. F. *et al*, (1966) *Arch. Biochem. Biophys. Acta 112*, 116.

Favel, A., *et al* (1989) *Int. J. Peptide. Prot. Res. 33*, 202-208.

Finkenstadt, W. R., *et al* (1974) Bayer Symp. 5, 389-411.

Ginna, L. S. and Tschopp. J. F., (1988) *Yeast 5*, 106-115.

Grasberger *et al*, (1994) *Structure. 2*, 669-678.

Hakateyama, T., *et al*, (1991) *Agric. Biol. Chem. 10*, 2641-2642.

Hamato, N., *et al*, (1995) *J. Biochem. 117*, 432-437.

Hara, S., *et al*, (1989) *J. Biochemistry 105*, 88-92.

Hayashi, K., *et al*, (1994) J. Biochem. *116*, 1013-1018.

Heitz, A., *et al* (1989) *Biochemistry 28*, 2392-2398.

Hinnen, A. P., *et al* (1987) *Mol. Cell. Biol. 14*: 43-50.

[1]Holak, T. A., Gondol, D., Otlewski, J. and Wilusz, T., (1989) *J. Mol. Biol. 210*, 635-648.

[2]Holak, T. A., Bode, W., Huber, R., Otlewski, J. and Wilusz, T., (1989) *J. Mol. Biol. 210*, 649-654.

Huang, Q., *et al* (1992) *FEBS Lett. 297*, 143-146.

Huang, Q., *et al*, (1990) *Science in China 33*, 1192-1200.

Huber, R., *et al* (1975) *Biophys. Struct. Mech. 1*, 189-201.

Huber, R., *et al*, (1974) *J. Mol. Biol. 89*, 73-101.

Ikemura, T., *et al*, (1982) *J. Mol. Bio. 158*, 573-597.

Kupryszewski, G., *et al*, (1985) *Experientia 41*, 1422-1423.

Kupryszewskki, G., *et al*, (1986) *Int. J. Peptide Protein Res. 27*, 245-250.

Laemmli, U. K. and Favre, M. (1973) *J. Mol. Biol. 80*, 575-599.

Laroche, Y., *et al*, (1994) *Bio/Tech 12*, 1119-1124.

Laskowski, M. Jr. and Kato, I. (1980) *Annu. Rev. Biochem. 49*, 593-626.

Laskowski, M. Jr., Sealock, R. W. (1971) *The Enzymes 3*, 375-473.

Laskowski, M., *et al*, (1987) *Biochemistry 26*, 202-221.

Le-Nguyent, D., *et al*, (1989) *Biochem. Biophys. Res. Com, 162*, 1425-1430.

Leluk, J., *et al*, (1983) *Acta. Biochim. Polon. 30*, 127-138.

Liang Lin, S. and Nussinov, R. (1995) *Nat. Str. Bio, 2*, 835-837.

Lipscomb, W. N., *et al*, (1969) *PNAS* 64, 28.

Markus, H. W. *et al*, (1995) *FEBS lett.* 451-456.

Marquart, M., *et al*, (1983) *Acta Crystallogr B39*, 480-490.

Matsuo, M., *et al*, (1992) *Biochim. Biophys. Acta. 1120*, 187-192.

McGrath, M. E., *et al*, (1995) *Protein Sci. 4*, 141-148.

McWerter *et al*, (1989) *Biochemistry 28*, 5708-5714.

Nilges, M., *et al*, (1991) *J. Mol. Biol. 219*, 499-510.

Nowak, K., *et al*, (1981) *Hoppe-Seylers Z. Physiol. Chem. 362*, 1017-1019.

Otlewski, J. (1990) *Biol.Chem. Hoppe-Seyler. 371 suppliment.*, 23-28.

Otlewski, J. (1993) *The Squash Inhibitors of Serine Proteinases* in *Innovations on Proteases and their Inhibitors* (Aviles, F.X., Ed.) pp 369-390, de Gruyter, Berlin

Otlewski, J. (1995) *Prot. eng. 8*, issue S, 55.

Otlewski, J. and Wilusz, T. (1985) *Acta. Biochim. Polon. 32*, 285-293.

Otlewski, J. and Zbyryt. T., (1994) *Biochemistry 33*. 200-207.

Otlewski, J., *et al*, (1994) *Biochemistry 33*, 208-213.

Polanowski, A., *et al*, (1980) *Acta. Biochim. Pol. 27*, 371-382.

Polanswski, A., *et al*, (1987) *Acta. Biochim. Polon. 34*, 395-406.

Quast, V., *et al*, (1978) *Biochemistry 17*, 1675-1682.

Reddy, R. G., *et al*, (1994) *Biochem. Biophys. Res. Comm.* 205, 381-388.

Reich, Rifken and Shaw, (1975) *Proteases and Biological Control*. Cold Spring Harbor Laboratory.

Rhodes *et al*, (1960) *J. Biol. Chem. 235*, 1686-1693.

Rolka, K., (1989) *Biol. Chem. Hoppe-Seyler, 370*, 499-502.

Rolka, K., *et al*, (1991) *Biol. Chem. Hoppe-Seyler, 372*, 63-68.

Rolka, K., *et al*, (1992) *Bio. Chem. Hoppe-Seyler 373*, 1055-1060.

Romanos, M. A., *et al*, (1991) *Vaccine 9*, 901-906.

Romanos, M. A., *et al*, (1992) *Yeast, 8*, 423-488.

Romanos, M., (1995) *Curr Opin Biotech 6*, 527-533.

Rozycki, J., *et al*, (1994) *Biol. Chem. Hoppe-Seyler, 375*, 21-23.

Schagger, H. and von Jagow, G., (1987) *Anal Biochem 166*, 368-379.

Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res Commun 27*, 157.

Schroder, D. D. and Shaw, E. (1968) *J. Bio. Chem. 243*, 2943-2949.

Scorer, C. A., *et al*, (1993) *Gene 136*, 111-119.

Scorer, C. A., *et al*, (1994) *Bio Tech 12*, 181-184.

Sharp, P. M., *et al*, (1988) *N. A. R. 16*, (17) 8207-8211.

Sreekrishna, K., *et al*, (1989) *Biochem 28*, 4117-4125.

Struhl, K., (1983) *Nature 305*, 391-397.

The Yeast, (1970) Ed. Lodder. J. North-Holland publishing co. Amsterdam/London.

Tschopp, J. F., *et al*, (1987) *Bio Tech 5*, 1305-1308.

Valenzuela, P., *et al.* (1982) *Nature 298*, 347-350.

Vozza, L. A., *et al*, (1996) *Bio'Tech 14*, 77-81.

Wen, L., *et al.* (1993) *Protein Expression and Purification 4*, 215-222.

Wieczorek, M., *et al.* (1985) *Biochem. Biophys. Res. Commun. 126*, 646-652.

Wilusz *et al*, (1983) *Hoppe-Seylers Z. Physiol. Chem. 364*, 93-95.

Wilutz, T., *et al*, (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 93-95

Wynn, R. and Laskowski, M. Jr., (1990) *Biochem. Biophys. Res. Comm. 166*, 1406-1410.

Xiao-Ming, C., *et al.* (1992) *J. Biochem. 112*, 45-51.

Yamada, M., *et al.* (1994) *BBA* 1206, 279-285.