Synthesis and expression of the *Erythrina* trypsin/tissue plasminogen activator (tPA) inhibitor encoding-gene. Genetic dissection to correlate the interaction of *Erythrina* and Soybean trypsin inhibitors with tPA.

Avelino V. Teixeira

Thesis presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry,

Faculty of Science

University of Cape Town

Cape Town, August 1992.

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

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

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

University of Cape Town

CONTENTS

		PAGE
ACKNOWLEDGEMENTS		i
ABSTRACT		ii
ABBREVIATIONS USED IN TEXT AND FIGURES		vi
CHAPTER 1	INTRODUCTION	1
CHAPTER 2	DESIGN AND SYNTHESIS OF AN ETI-ENCODING GENE	50
CHAPTER 3	EXPRESSION OF THE SYNTHETIC ETI GENE IN *E.COLI*	87
CHAPTER 4	DETERMINATION OF THE ACTIVITY OF val-ETI AND asp_7ETI AGAINST t-PA	169
CHAPTER 5	CONCLUSIONS	187
CHAPTER 6	MATERIALS AND METHODS	193
APPENDIX		217
REFERENCES		229

ACKNOWLEDGEMENTS

I am indebted to the following people:

My supervisor Professor Dawie Botes for his guidance and assistance during the course of this work.

Ms T. Pitout for assistance in synthesis and purification of all oligonucleotides used in this work.

Dr C. Heussen-Schemmer, Ms Sandra Coetzee and Professor E.B.D. Dowdle for their assistance with all tPA assays in this work.

Dr P. Smith and Mrs F. Davids of the Biochemistry Department, UCT for their assistance with all antibody preparations and Western blotting.

The Foundation for Research Development for their financial support.

My Mother and late Father for their constant encouragement.

My Wife Dianne for her continuous support and abundant patience during the course of this work and my two little girls Rachel and Angie for being themselves.

ABSTRACT

A trypsin inhibitor had previously been isolated from *Erythrina caffra*, a member of the Leguminosae family. The inhibitor, *Erythrina* trypsin inhibitor (ETI), is unique among plant-derived inhibitors, in that in addition to trypsin, it inhibits chymotrypsin and tissue plasminogen activator (tPA). ETI was previously sequenced and its crystal structure determined from which it could be seen that ETI has good homology to soybean trypsin inhibitor (STI). However, STI does not inhibit tPA.

From the three-dimensional structure of ETI it was known that the amino-terminus of the molecule forms a finger-like structure stabilized by hydrogen bonds and hydrophobic interactions. In addition, the N-terminal finger region is located in close proximity to the reactive site loop and the N-terminal residue (Val) is bound up in the finger region. In STI the N-terminal region is located in close proximity to the reactive site loop and is folded into a structure similar to that in ETI. While the crystal structure of STI was not detailed enough to determine secondary interactions such as hydrogen bonds it was hypothesized that the N-terminal region is stabilized as in ETI. It was further hypothesized that the N-terminal residue of STI (Asp), because of its hydrophilic nature, is not involved in the structured N-terminal finger region of this protein. This leaves this Asp residue of STI free to form an ion pair with Lys at position 60 in trypsin when STI and trypsin interact.

When amino acid sequences of trypsin and the B-chain of tPA are aligned for optimum homology it is seen that there are a number of insertion sequences in tPA that are thought to be accommodated in the form of protrusions. One of these can be seen to occur in the region that lies opposite the Lys60 region of trypsin. It is suggested in this work that the N-terminal Asp of STI and this protrusion of tPA sterically prevent the two proteins from approaching close enough to allow for binding and inhibition to occur.

Plant codon preferences differ sufficiently from codon preferences of *E.coli* to make high levels of expression in *E.coli*, of a gene cloned from a plant, unpredictable. For this reason, a gene for ETI was designed from the published amino acid sequence. In order to alter the gene and produce modified forms of ETI, it was designed to allow removal and replacement of sections by strategically placed unique restriction sites.

The design was divided into 8 sections for each of which the non-coding strand was synthesized with additional 15 base arms to anchor the oligonucleotide (oligo) across the gap created by digestion of the synthesis plasmid by *Bsm*I. The entire gene was assembled in consecutive steps in the 3′- 5′ direction, by bridging the gap created by digestion of the plasmid containing the prior oligo with *Bsm*I with the sequential oligo and allowing *E.coli* to fill in the missing strand. At each step and on completion, the gene was sequenced to confirm that no mutation had occurred during synthesis.

The ETI-encoding gene was originally designed for expression under the control of the *tac* promoter in a system that allowed for accumulation in the cytoplasm. This system proved unsuccessful. The gene was then cloned into the pMal gene and expressed as a fusion to maltose binding protein (MBP). The two domains of the protein in this system are separated by the cleavage site for factor X_a (FX_a). The MBP domain allowed for purification of fusion protein but FX_a was unable to separate the two domains and produced a fragment smaller than ETI. FX_a produced a similar sized fragment from pure natural ETI and it was concluded that ETI, even in the form of a fusion with MBP was able to inhibit FX_a which is a serine protease with reactive site similar to that of trypsin. As the facile enzymatic approach to producing ETI with a natural Val N-terminus had failed a Met residue was engineered just upstream of the required N-terminal residue. This construct was expressed in pET12a which supplies a signal sequence for export to the periplasm. Most of the expressed protein was located in the cytoplasm but because the periplasm is an environment conducive to the formation of disulphide bridges only periplasmic protein was isolated. The remaining two N-terminal amino acids after signal sequence cleavage in the periplasm were cleaved carboxy-terminal to the engineered Met residue with CNBr.

Two forms of ETI were produced, one with the sequence of natural ETI including a Val N-terminus (val-ETI), the other with an Asp residue N-terminal to Val (asp-ETI), to simulate the N-terminal region of STI. The active sites in each were titrated against trypsin and assayed against tPA. The results

showed val-ETI to have much the same activity towards tPA as natural ETI while asp-ETI had activity towards tPA similar to that of STI. This evidence indicates strongly that the N-terminal Asp of STI prevents its binding to and inhibiting tPA.

ABBREVIATIONS USED IN TEXT AND FIGURES

ETI	*Erythrina* trypsin inhibitor
FX	factor ten
FX_a	factor ten - activated form
kDa	kilo Daltons
PA	plasminogen activator
STI	soybean trypsin inhibitor
tPA	tissue plasminogen activator
tPAI	tissue plasminogen activator inhibitor
uPA	urokinase plasminogen activator
WTI	wingbean trypsin inhibitor

One and three-letter codes for amino acids

A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartic acid	P	Pro	proline
E	Glu	glutamic acid	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	Lysine	W	Trp	tryptophan
L	Leu	leucine	Y	Tyr	tyrosine

CHAPTER 1

INTRODUCTION

1.0 Contents

1.1 Preface 1

1.2 Proteases 2

1.2.1 A brief look at some of the serine proteases in blood clotting and fibrinolysis. 7

1.2.2 Plasminogen activation 9

1.3 Protease inhibitors 16

1.3.1 Plasminogen activator inhibitors 22

1.3.2 Erythrina trypsin inhibitor 25

1.4 Aims of this work 46

1.1 Preface

Biologically active proteins are often synthesized at a location in the organism that is distant from where they are required to function. In addition, they may need to be stored until their activity is required. Whatever the case, their activity is such that untimely performance of the specific biological function for which they are responsible can be detrimental to the organism. It is for this reason that living organisms have developed sophisticated mechanisms to prevent the activity of such proteins until it is required. Preventing the activity of physiologically active proteins can be achieved by synthesizing them as inactive precursors (zymogens). These are subsequently converted to active forms by cleavage of often only one peptide bond (limited proteolysis) by very selective

proteolytic enzymes (proteases). Another mechanism widely utilized by all classes of organisms for control of active proteins is to simultaneously synthesize proteins that bind specifically to a protein or group of proteins and inhibit its activity.

The subject of this study is an inhibitor protein that was isolated from an *Erythrina* species. It is an inhibitor of the powerful digestive proteases: trypsin, chymotrypsin and the plasma protein, tissue plasminogen activator (tPA). As a result, the inhibitor is known as *Erythrina* trypsin inhibitor (ETI). ETI will be described later in Chapter 1 after some of the properties of proteases and their inhibitors have been introduced in relation to the target proteases of ETI: trypsin and tPA.

1.2 Proteases

Hydrolysis of peptides in neutral solutions is slow because there are high energy barriers between starting materials and products. Enzymes allow the reactants to follow a different pathway and new intermediate states are established in which the enzyme stabilizes the states of highest energy thereby lowering internal energy barriers (Stroud, 1974).

Proteolytic enzymes from the gastrointestinal tract of humans were among the first recognized enzymes and have counterparts in blood plasma, sperm and other tissues as well as in simpler organisms such as bacteria where they serve in a control

function (Walsh, 1975). Proteases have been arranged into families which are mechanistically, but not necessarily structurally related viz i) aspartic ii) cysteine iii) metallo and v) serine proteinases (Bond and Butler, 1978).

The aspartic family of proteinases include inter alia the very important gastric enzymes such as chymosins, gastrins and pepsins as well as the renin group from the kidneys. These have been placed in this family because they share a common catalytic mechanism provided by the carboxyl side- chains of two aspartic acid residues which are positioned distant from each other on the polypeptide chain. For example, in pepsin, Asp32 and 215 are implicated in catalysis and they are brought into close proximity in a deep cleft between β-sheet regions by the folding of the protein. The cysteine proteases, of which the papains are an example use a nucleophilic cysteine residue for their reaction mechanism. Metalloproteases are characterized by the use of a chelated metal atom, such as Zn as an electron sink to polarize the peptide bond under attack. For example, carboxypeptidase A uses a hydrophobic pocket and Arg145 to bind the amino acid C terminal to the target peptide bond. This is then attacked in a concerted manner by a general acid (Tyr248) and a general base (Glu270).

The serine proteases are a very large family of related enzymes which are important in biological control. They are also of interest to the current work as the target enzymes of ETI: trypsin, chymotrypsin and t-PA all belong to this family of

enzymes. The detailed structure and mechanism of action of one member of a mechanistic set can provide a model for understanding the structure, mechanism, function and specificity of other members of the same set. As a result, serine proteases are further divided into trypsin-like (the subfamily to which the target enzymes of ETI belong) and subtilisin-like enzymes (Stroud, 1974).

The specificity and mechanism of action of an enzyme is determined by two key factors.

First is the presence of three groups of important amino acids. which are involved in contact with the substrate.

1) The amino acids involved in cleavage (the active site amino acids). In the three serine protease digestive enzymes trypsin, chymotrypsin and elastase, the most important amino acids in this group are Ser195, His57 and Asp102 (chymotrypsin numbering) and their immediate neighbours. These 3 units are brought together by the folding of the polypeptide chain and to form the active site.

2) Those amino acids that form the binding site (or pocket) which is the region of the enzyme that accommodates the side-chain of the amino acid on the carboxy side of the target peptide bond to orient it for cleavage. This is the region that determines specificity of the different members of a mechanistic family of enzymes. Trypsin is the most selective of the digestive enzymes discussed above. It has a large binding pocket at the back of which is an Asp residue

which is negatively charged. Trypsin cleaves on the carboxy-terminal side of Arg or Lys residues which are large, positive and hydrophilic and can be accommodated in the binding pocket and held in place by electrostatic attraction. Chymotrypsin is specific to peptide bonds next to amino acids with large hydrophobic side chains. The elastase binding pocket is blocked to all but the smallest side chains in the substrate and hence it cleaves bonds adjacent to Gly, Ala or Ser the smallest of the amino acids (Stroud, 1974).

3) Those apart from the catalytic and binding sites discussed above that are located on the enzyme and stabilize the enzyme-substrate complex (Walsh, 1975).

The second factor determining specificity and mechanism of action of an enzyme is the overall structure of the enzyme. To illustrate this the serine proteases trypsin, chymotrypsin and elastase are structurally similar and have analogous assemblies of components at their active centers but alignment of their amino acid sequences reveals only 40-50% identity. Evidently the important sequences that control conformation have been preserved. The three differ in specificity and this is explained by changes in the amino acids that shape their binding sites (Walsh, 1975).

The need for timely activity of physiologically active proteins has been discussed earlier. For example, the digestive enzymes trypsin and chymotrypsin are very active enzymes and would degrade the cells of the pancreas and its ducts if they

were not produced as the inactive proenzymes, trypsinogen and chymotrypsinogen respectively. Activation of trypsin is achieved by cleavage of the peptide bond between the 15th and 16th amino acid of trypsinogen by enterokinase, another proteolytic enzyme produced by the intestinal mucosa, releasing a 6 amino acid fragment. The new amino terminal at unit 16 interacts ionically with Asp194, altering the orientation of Lys145 and inducing a structural change that generates the specific binding site. Each molecule of trypsin released can activate another molecule of trypsinogen which then activates the other digestive proenzymes (Stroud, 1974).

The catalytic mechanism of trypsin, which acts as a model for the mechanism of action of the trypsin-like serine proteases, is characterized by the concerted attack on the carbonyl carbon of the peptide bond by Ser195 and His57 to yield an oxyanion tetrahedral intermediate stabilized in an "oxyanion hole" in which the negative charge on the carbonyl carbon is stabilized by hydrogen bonds to the enzyme. The charge relay system consists of a proton exchange between the side chain imidazole of His57 and the carboxylate of Asp102. This releases the amine portion of cleaved peptide and produces an acyl enzyme. A water molecule replaces the amine and the process is reversed to deacylate the enzyme (Walsh, 1975; Kraut, 1977; Carter and Wells, 1988). It is important to note that the unusual electronic state of His57 and Asp102 in the interior of the protein allows for the efficient shuttle of protons which is the key to the enzymatic activity (Stroud, 1974). Serine proteases employ the same three amino acids to hydrolyse

peptide bonds. Their specificity results from the way in which they accommodate their substrates.

1.2.1 A brief look at some of the serine proteases in blood clotting and fibrinolysis.

When tissues are broken the first requirement of the body's defence system is to stop the leakage of blood. This process is begun immediately since the injury itself releases activating agents, one of which is Hageman factor that initiates a cascade of enzyme-controlled reactions. Some of these are serine proteases, and are activated in sequence. The last in this cascade is thrombin, a serine protease very much like trypsin (but after between Arg and Gly) which removes 4 short highly charged peptides from fibrinogen converting it to insoluble fibrin. Fibrinogen is a major blood glycoprotein that plays an essential role in haemostasis and the maintenance of blood viscosity (Cook and Ubben, 1990). The fibrin matrix is initially held together by hydrogen bonds but fibrin stabilizing factor (FSF or factor XIII), which is activated by thrombin, activates fibrinoligase from its precursor. Fibrinoligase links the fibrin molecules covalently by cross-linking the side chains of Gln and Lys residues between fibrin molecules. The resulting mass of tangled threads is what is called a clot (Lorand, 1972).

Thrombin is produced as an inactive precursor called prothrombin (also known as coagulation factor or Factor II), which is a single polypeptide chain with intrachain disulphide

bonds and carbohydrate side chains. The conversion of prothrombin to thrombin is brought about by thromboplastin (Factor III) in the presence of Ca^{2+} (factor IV). Thromboplastin is derived from various tissues through the mediation of platelets, platelet factors and other factors viz proaccelerin (accelerator globulin or Factor V); proconvertin (serum prothrombin conversion accelerator, SPCA or Factor VII); antihemophilic globulin (AHG) (antihemophilic factor, AHF, also called Factor VIII); and the Christmas factor (antihemophilic factor B or Factor IX). On activation, prothrombin is cleaved to produce two polypeptide chains linked by a disulphide bridge. The A chain contains 36 residues and B chain 259 residues with carbohydrate attached. The sequence of the B chain is homologous to that of trypsin, chymotrypsin and elastase. Thrombin is involved only in the release of fibrinopeptides and cleaves only one Arg-Lys bond at the NH_2 terminus of the α and β chains of fibrinogen (Mann, 1987).

Conditions other than wounding can lead to clot formation in the blood vessels, for example a defect such as an atherosclerotic deposit on the vessel wall. The formation of a thrombus (clot) on this defect almost invariably results in the interruption of blood flow through that vessel (Haber *et al.*, 1989). Thus it is crucial that there be a mechanism for the breakdown of clots in the vascular system; this is known as the fibrinolytic system. The enzyme ultimately responsible for dissolution of clots is the serine protease plasmin, which is produced as the proenzyme plasminogen. The native form of human plasminogen, Glu-plasminogen, is a 90 kDa glycoprotein with ±

2% carbohydrate and is a single polypeptide of 791 amino acids. All mammalian plasminogens exist in two variant forms differing in their carbohydrate composition. In the human system the carbohydrate difference is due to the presence of an oligosaccharide on Asn228 of variant 1 but not on variant 2. Activation of Glu-plasminogen occurs through cleavage of a single peptide bond by a plasminogen activator (PA). The Glu-plasmin formed consists of two polypeptides (25 and 65 kDa) linked by two disulphide bonds. The smaller subunit contains the active site and resembles in sequence thrombin, trypsin, chymotrypsin, and elastase. The larger subunit contains several fibrin binding sites with an affinity for Lys and an 8 kDa activation peptide at the amino terminal end which is released by autocatalysis after the initial activation. When fibrin is formed, Glu-plasminogen is specifically bound to it and can be activated by a PA for breakdown of the clot (Saksela, 1985).

1.2.2 Plasminogen activation

The activation of plasminogen is regulated at several levels. Firstly, the proactivator needs activation. Secondly, the active activator is bound by specific inhibitors. Thirdly, other molecules such as fibrin affect the activation rate. Fourthly, the active plasmin creates a positive feedback effect by activating proactivator. In addition, plasmin itself is inhibited by a variety of proteinase inhibitors (Saksela, 1985). Plasminogen activators hydrolyze the Arg560-Val561 peptide bond of plasminogen and convert it to active plasmin (Lijnen and Collen, 1988). In addition, streptokinase, a protein from *Streptococci* can also activate plasminogen when

complexing with it. In doing so streptokinase induces a conformational change which exposes the active site of plasminogen allowing it to function as a plasminogen activator (Van Zonneveld *et al.,* 1986).

Two types of plasminogen activator have been shown to be responsible for fibrinolysis in the human vascular system: tissue plasminogen activators (tPA) and urokinase plasminogen activator (uPA) both of which are trypsin-like serine proteases but are immunologically distinct. UPA was originally isolated from urine where it is the major PA but subsequently both PAs have been isolated from many human cell types (Rijken *et al.*, 1981). Both plasminogen activators are produced as single chain proenzymes and require cleavage of a single bond for activation (Wallen *et al.*, 1983). While plasmin and trypsin can activate PAs, activating activity has also been attributed to other serine proteases such as plasma kalikrein, Factor XI_a and Factor XII_a (Skriver *et al.*, 1982; Laug *et al.*, 1983). However, tPA is active in both the single and two chain forms which is unusual for serine proteases as most of the members of this group are produced as zymogens and require proteolytic cleavage to a two chain form to release full enzymatic activity (Urano, 1989; Rijken, 1982).

Active uPA is a 54 kDa protein of two subunits (30 and 24 kDa) linked by a disulphide bond. The active site is located on the 30 kDa fragment (Saksela, 1985). Native tPA has molecular weight of 70 kDa and is glycosylated at 3/4 of the possible glycosylation positions (Saksela, 1985). Single-chain tPA, by proteolytic cleavage at residues 275-276, is converted into a

two-chain activator (30 and 40 kDa) linked by a disulphide bond (Pennica *et al.*, 1983;). The light chain contains the active site (the serine protease domain) and the heavy chain contains four identifiable domains: a fibronectin-like finger, an epidermal growth factor homologous region and two kringle structures (triple disulphide loops). The structure of uPA is homologous to that of tPA, except that it has only one kringle domain and it does not posses the fibronectin-like finger (Bennett *et al.*, 1991).

A central goal of tPA biochemistry is to relate the details of this structure to the diverse functionalities displayed by the enzyme. Physical analysis, for example by crystallography, has been difficult because of the large size of tPA and its heterogeneous glycosylation in mammalian cells. Enzyme kinetic studies are complicated by the size, complexity and heterogeneity of tPA's substrate (plasminogen) and its most important modulators, fibrin and tissue plasminogen activator inhibitor 1 (tPAI-1). Thus much of the hope for understanding the relationship between the structural elements of tPA and its various functional properties has become invested in the study of mutant proteins (Bennett, 1991).

The 43 residue amino terminal region of the tPA heavy chain (the fibronectin-like finger domain) is homologous with the finger domains responsible for the fibrin affinity of fibronectin (Banyai *et al.*, 1983). This region, the growth factor domain and to a lesser extent the kringle 2 domain of tPA has been shown to be involved in binding of the molecule to fibrin (van Zonneveld *et al.*, 1986; Markland *et al.*, 1989). It

is suggested that kringle 2 carries the Lys binding sites that are utilized for purification of tPA (Cleary *et al.*, 1989; Bennett *et al.*, 1991). Kalyan *et al.*, (1988) have gone further and suggest in addition, a role for the fibronectin-like finger domain (and the NH_2-terminus of the light chain) in clearance of tPA from plasma *in vivo* (tPA has a high turnover rate) mediated by receptors in the liver (Bakhit, 1987).

Evidence suggests that there are several different types of tPA variant (Pohl *et al.*, 1984a). About 50% of the single chain form lacks the N-terminal sequence Gly-Ala-Arg, which has been removed by plasmin. In addition, the heavy chain occurs in 37 and 40 kDa variants. The difference has been attributed to the carbohydrate structure (Pohl *et al.*, 1984b).

Several hormones and hormone-like substances alter the secretion rate of PAs. Secretion of uPA particularly is extremely sensitive to various stimuli affecting the cell cycle which suggests some still unknown role for this enzyme. Mapping of the genomic structure of cDNA produced from the uPA gene has revealed a potential hormone receptor binding site and a sequence in the genome that is possibly involved in activation of transcription by cAMP (Nagamine *et al.*, 1984).

The fibrin binding properties of tPA ensures that the activation of plasminogen is localized predominantly on the surface of a fibrin clot and in addition, enhances its ability to activate plasminogen. Plasminogen and tPA form a cyclic ternary complex bound to fibrin. This substantially increases the catalytic efficiency of tPA for free plasminogen over that

of unbound tPA or tPA in the absence of fibrin. UPA binds poorly to fibrin and lacks such fibrin-mediated plasminogen activation. (Ranby, 1982; Hoylearts *et al.*, 1982). The fibrin dependency and higher fibrin-binding properties of tPA make it a specific and efficient thrombolytic (clot digesting) agent. This suggested that tPA could be used clinically to specifically facilitate breakdown of clots without the side-effects of bleeding associated with uPA and streptokinase (van de Werf, 1984). However as tPA binds to fibrin during clot formation, clots prepared from tPA poor plasma are only slowly degraded after addition of tPA (Brommer, 1984). As tPA fibrinolytic therapy activates plasminogen indiscreetly and plasmin modifies platelet function and degrades circulating fibrinogen and clotting Factors V and VIII as well as thrombus bound fibrin, plasminogen activator therapy also carries the risk of hemorrhage (Haber, 1989). Thus, tPA fibrinolytic activity may possibly be best suited to preventative treatment, in appropriate and well controlled doses, in cases in which tPA can be shown to be lacking.

The role of plasmin in fibrinolysis has been discussed above but it has also been shown that several cell types other than cells of the vascular system produce tPA and that plasmin contributes directly and indirectly via conversion of latent collagenase into active collagenase to the degradation of and turnover of the extracellular matrix (Kruithof, 1988). The intracellular activator proenzyme has been localized to the Golgi-derived secretory vesicles and may be bound to the internal surface of the vesicle membrane, where it is able to

activate extracellular plasminogen after fusion of the vesicle with the plasma membrane (Laug *et al.*, 1983). The cascade reactions leading to plasminogen activation provide effective proteolytic activity which can thus be regulated at the level of single cells and used for strictly controlled proteolysis and degradation of extracellular matrixes (Saksela, 1985; Lawrence *et al.*, 1990). Thus plasmin, and by inference, plasminogen and its activators are important in many human physiological processes. Some of these are discussed below.

Plasmin plays an important role in reproduction, for example, a few hours before ovulation, the granulosa cells of the ovaries secrete high levels of tPA and uPA (controlled by follicle stimulating hormone, FSH) which activates plasmin and causes disruption of the follicle wall in preparation for release of the egg (Beers, 1975; Moscatelli and Rifkin, 1988). In culture, the sertoli cells of the testis also secrete PAs under regulation of FSH. The activated plasmin is required to degrade the matrix between primary spermatocytes and the basal membrane of the seminiferous tubules (Lacroix, 1982). After fertilization of the egg, trophoblasts, which are derivatives of the trophoectodermal cell layer of the embryo, attach it to the uterine wall during implantation, in doing so they have been shown in mice, to secrete PAs (Strickland *et al.*, 1976). Plasmin is also implicated in the conversion of the prohormone of placental lactogen into its more active form (Russell *et al.*, 1979).

Plasmin is also involved in general metabolism and endocrinology as there have been suggestions that it may be involved in insulin metabolism (Grieger and Binder, 1984). It is known that the cells of the pituitary gland of the rat secrete uPA and tPA. In pituitary cell cultures, extracellular ACTH has been found to be exclusively of the inactive proenzyme form when the culture medium is depleted of plasmin (Granelli-Piperno and Reich, 1983).

Plasmin has been shown to be important in proliferation of cancer cells which by their nature penetrate structures separating various tissues and cell types such as the endothelium and basement membrane of the vessel wall. The PAs secreted by cancer cells assist in this process by mediating degradation of the glycoproteins fibronectin and laminin found in the basement membrane and the extracellular matrix (Liotta *et al.*, 1981). The collagenous components present in these structures are resistant to plasmin digestion but latent collagenases secreted by the cells themselves or by adjacent tissues can be activated by plasmin (O'Grady *et al.*, 1981). Three types of enzyme seem to be needed for effective matrix breakdown: PA, plasminogen and collagenases. The conclusion is drawn that plasminogen activation is required during the very early stages of the metastatic process when, for example infiltrative growth is required. Proteinase inhibitors leupeptin and antipain (*Streptomyces* and *Actinomyces*) can inhibit carcinogen and radiation induced transformation of cells and both substances are known to inhibit PAs (Kinsella, 1980). Furthermore, exposure of various fibroblasts to ultraviolet light or other agents producing DNA damage induces

synthesis and secretion of PAs (Miskin, 1980). Clearly, solely the production of PAs does not define a cell as malignant but rather, possible defects in the control of the enzyme activity (or the activation of the secreted proenzyme) at the pericellular level may account for several of the growth properties of individual cells in the heterogeneous population of tumor cells (Dexter and Calabresi, 1982; Lawrence *et al.*, 1990).

1.3 Protease inhibitors

As discussed above, plasmin is a wide spectrum protease that is responsible for fibrinolysis and indirectly for turnover of extracellular matrix by activation of latent collagenases. It was mentioned previously that one method for regulation of plasminogen activation is by the use of specific inhibitors which bind to and prevent the activity of plasminogen activators. Protein inhibitors of proteinases are found in all living organisms where their function is the prevention of any unwanted proteolysis. They comprise ± 10% of the human plasma proteins and belong to one of two main groups. The first group, ∝-2-Macroglobulin (∝2M), is a 725 kDa glycoprotein with 8-11% carbohydrate. While it can inhibit all four mechanistic classes of proteases the main function of ∝2M is to bind and remove active proteases from the system. It is only when the local or systemic concentrations of an inhibitor become lowered significantly, for example in physiological disturbances such as disseminated intravascular coagulation, during which the majority of the anti-thrombin III (the specific inhibitor of thrombin) is exhausted, and during thrombolytic therapy, in

which α2 anti-plasmin (the specific inhibitor of plasmin) levels drop precipitously that α2M exerts any significant control of the coagulation and fibrinolytic systems (Travis and Salvesen, 1983). For this reason it is seen by some as a backup system for protease control. Functionally, α2M has 3 centers a) Bait region which is susceptible to proteolysis by proteinases. b) Internal thioester region, possibly involved in conformational changes of the molecule and binding of the protease. The inhibited enzyme is irreversibly bound because the protease cleaves a peptide bond of the bait region. This results in conformational change of the inhibitor which is thought to wrap around the protease and to inhibit it by steric hindrance. In this complex, the active site of the protease is blocked to large substrates but remains accessible to small substrates (Barrett and Starkey, 1973). c) The receptor recognition region which is the region of α2M revealed after conformational change of the molecule. Complex formation signals the start of degradation and the complex is absorbed into the reticuloendothelial system by receptor mediated endocytosis and destroyed (Van Leuven, 1982; Travis and Salvesan, 1983.

The second group of inhibitors contains those with class-specific reactive sites. These are the specific inhibitors of the aspartic, cysteine, metallo and serine proteases (Travis and Salvesen, 1983). The protein inhibitors directed against serine proteinases are not all homologous and can be grouped into at least 11 different families based on sequence similarity, topological similarity, and mechanism of binding (Bode and Huber, 1992) and are listed below.

1. Bovine pancreatic trypsin inhibitor (BPTI) (Kunitz).
2. Kazal.
3. Soybean trypsin inhibitor (STI-Kunitz).
4. Streptomyces subtilisin inhibitor (SSI).
5. Potato inhibitor 1 (PI-1) family.
6. Potato inhibitor 2 (PI-2).
7. Chelonianin.
8. Bowman-Birk.
9. Squash seed inhibitors.
10. Serpins.
11. Hirudin

Of these, the serpins and the STI-Kunitz family are of special significance to this work as the physiological plasminogen activator inhibitors (PAIs) are members of the serpin family while ETI and soybean trypsin inhibitor (STI, which will be discussed later) are members of STI-Kunitz family. Most of these serine-protease-directed inhibitors react competitively with their cognate enzymes according to a common, substrate-like standard mechanism (Laskowski and Kato, 1980). This group of 'canonical' inhibitors (to which ETI and STI belong) comprises relatively small proteins (or protein domains of multiheaded inhibitors) of between 29 and ± 190 amino acid residues. They all posses an exposed binding loop of characteristic conformation, but are otherwise unrelated in structure (Bode and Huber, 1992). In all cases the inhibitors (or their individual domains) have a compact shape and contain a hydrophobic core which in some cases consists mostly of the cross-connecting disulphide bridges. The core inhibitor forms a scaffold on which the exposed protease binding loop is held. Interactions between the loop and individual elements of the

core give additional stability to the loop which has a flat shape that fits into the active-site cleft of its cognate serine proteases. The reactive site loop is relatively flexible compared to the core of the inhibitor and acts as if hinged to this core. This was observed by superposing the structures of certain inhibitors in their free state and when complexed with a cognate enzyme. In doing so the major differences were found in the active site loop whereas the core of the molecule was seen to be structurally highly conserved (Onesti *et al.*, 1991). Regions of the domain are also involved in bringing the inhibitor and protease together to allow for inhibition and stabilizing this interaction. Inhibition, in this group of inhibitors is achieved by competitively blocking the enzyme's active site to its substrate and all enzymatic activity of the protease is prevented. Under appropriate conditions of time, pH and relative concentrations of inhibitor and enzyme, the inhibitor is cleaved at a single scissile bond that is situated at the reactive site and is positioned over the catalytic site of the enzyme in a stable enzyme:inhibitor complex. By convention, the residue on the amino-terminal side of the scissile bond is referred to as P1 while that on the carboxy terminal side as P'1. Trypsin inhibitors have Arg or Lys as the P1 residue; Ala, Leu, Ile and Ser have been found at the P'1 position (Tschesche et al., 1976). X-ray analyses reveal that the geometry of the reactive sites of certain inhibitors is such that they should act as good substrates. Indeed some inhibitors are hydrolyzed by their cognate proteases while blocking access of their active sites to substrates. At neutral pH and typical enzyme:inhibitor concentrations, because kcat and km for the process of inhibition are very much smaller than

for normal reactions, hydrolysis of the bond is extremely slow (Laskowski and Kato, 1980). In the case of those inhibitors obeying the standard mechanism, hydrolysis of the reactive site peptide bond does not proceed to completion and an equilibrium is established between the form of the inhibitor with peptide bond intact (virgin inhibitor) and the form with peptide bond cleaved(modified inhibitor). The following simplified kinetic scheme summarizes the reaction:

E + I <-> EI <-> C <-> EI* <-> E + I*

Where E = protease.

I and I* = virgin and modified inhibitor respectively.

C = stable complex.

The stable complex C is formed from either virgin or modified inhibitor, both of which are thermodynamically strong inhibitors however the apparent rate of association of modified inhibitor with enzyme is generally slower than the corresponding rate for virgin inhibitor. X-ray analysis reveals that the reactive-site peptide bond is intact within the complex, but that the carbonyl function is partially tetrahedral rather than trigonal, as it is in its uncomplexed state. The distortion of this bond is caused by attraction of the carbonyl oxygen to the oxyanion binding pocket and is envisaged as one of the attractions that help stabilize the enzyme inhibitor complex. Other interactions including hydrogen bonds, Van der Waals forces, and possibly salt bridges, in the close contact area between the reactive site and the corresponding substrate-binding region of the protease all contribute to the characteristic stability of the complex (Travis and Salvesen, 1983).

The serpins are a superfamily of homologous 50-100 kDa glycoproteins which, like the canonical inhibitors, have a reactive centre located on an exposed loop, referred to as the strained loop, situated near the carboxy terminus of the molecule (Loeberman et al., 1984; Huber and Carrell, 1989). As with the canonical inhibitors, they interact with their target proteases by providing a 'bait' residue (P1 residue) located on the strained loop that mimics the normal substrate of the target protease. In doing so they form 1:1 complexes with the target protease and as the intermediate acyl-enzyme complex is hydrolyzed only very slowly, the enzyme is in effect trapped in an inactive state (Cohen et al., 1977; Stephens et al., 1987; Bruch et al., 1988). Several factors contribute to the protease specificity of serpins: 1) the amino acid sequence, especially residue P1, around the reactive site which is highly variable between serpins and is postulated to have undergone much more rapid evolutionary variation than the remainder of the molecule. 2) specific domains in the serpins involved in specialized functions such as a site in α2-antiplasmin which binds to a lysine-binding site in kringle 1 and/or kringle 4 of plasmin. The initial docking is the rate limiting step and thus of major importance for the rapid reaction between the active site of plasmin and the reactive site of α2-antiplasmin. 3) The absence or presence of small highly charged non-protein molecules which interact with the inhibitor at a site remote from the active site and are referred to by Chmielewska et al. (1988) as second site interactions. These interactions have the effect of increasing the stability of the complex formed between the inhibitor and the protease, for example, heparin interacts with antithrombin III and increases its association

rate constant by about 1000 fold (Jordan et al., 1979). The serpin family contains most of the plasma protease inhibitors, some examples of which are: 1) Alpha-1-protease inhibitor, a 53 kDa protein which inhibits inter alia plasmin, uPA and tPA. It has the physiological role of controlling elastolytic activity secreted by neutrophils (Murano et al., 1980; Rijken et al., 1983). 2) Antithrombin III, a 58 kDa protein that inhibits inter alia plasmin, trypsin, chymotrypsin and uPA. 3) Alpha-2-antiplasmin (α2AP), is a 65-70 kDa glycoprotein that is an extremely efficient, fast-reacting inhibitor of plasmin and the primary inhibitor of plasmin in circulation. However several reports indicate that it can also inactivate trypsin, plasma kallikrein, chymotrypsin, thrombin, Factor Xa and Factor XIa. The inhibition of all of these enzymes, with the exception of trypsin is very slow and it is unlikely that α2AP plays any role in controlling these activities (Lucas et al., 1983). 4) Inter-alpha-trypsin-inhibitor, a 180 kDa glycoprotein that inhibits trypsin, chymotrypsin and plasmin. 5) Protease nexin (PN) is a glycoprotein of 47 kDa. Besides uPA and two chain tPA it inhibits thrombin rapidly and also trypsin, plasmin and factor Xa. The wide inhibitory spectrum of PN and the different functions attributed to PN suggest that the function of PN is probably not inhibition of PAs, but rather the general control of local serine protease activity (Scott and Baker, 1983; Scott et al., 1985). Some important specific physiological inhibitors of plasminogen activators are also members of the serpin family.

1.3.1 Plasminogen activator inhibitors

Several cell types grown in vitro secrete proteins which are capable of directly and specifically inhibiting the plasminogen activation reaction. Three proteins have thus far been identified that efficiently react with tPA and uPA: PA inhibitor 1 (PAI-1), PA inhibitor 2 (PAI-2) and PA inhibitor 3 (PAI-3) (Kruithof, 1988). The indication is that all three PAIs each react equally well with single chain and two chain forms of tPA and the same is true of uPA (Chmielewska et al., 1980).PAI-1 was first identified in conditioned media of human endothelial cells and rat hepatoma cells. It is found in a wide variety of other sources such as plasma, platelets, placenta and conditioned media of fibrosarcoma cells and hepatocytes. PAI-1 levels are regulated by glucocorticoids, endotoxin, interleukin-1, transforming growth factor β, tumor necrosis factor and thrombin. The inhibitor has been purified from endothelial cells, hepatoma cells, fibrosarcoma cells, smooth muscle cells and melanoma cells. It was found to be a 52 kDa glycoprotein that is responsible for 60% of PA inhibitory activity (Sprengers and Kluft, 1987). In accordance with the Arg specifity of tPA and uPA all PAIs have an Arg at the active site and tPA in particular has been shown to cleave PAI-1 at the scissile bond Arg346-Met347 (Lindall et al., 1990). In the PAI-1-tPA inhibition complex, residues 296-302 and 304 of tPA which are located on the lip of the active site cleft and are not involved in catalyltic functions are also involved in stabilization of the intermediate complex (Madison et al., 1989; Monge et al., 1991). The complex can be further

stabilized by second site interactions with PAI-1. In this capacity, fibrinogen or heparin can increase the inhibition constant fourfold and vitronectin increases it sixfold (Chmielewska et al., 1988; Edelberg et al., 1991). Furthermore, the work of Erlich et al. (1990) has shown that vitronectin can dramatically enhance the reactivity of PAI-1 towards thrombin without decreasing its capacity to inhibit PA. They propose, as a result, that PAI-1 is a versatile inhibitor which in the presence of vitronectin can modulate both coagulation and fibrinolysis.

PAI-2 is found in human placenta where it is localized in the trophoblastic epithelium, pregnancy plasma and is secreted by leukocytes, leukocyte-derived cell lines and fibrosarcoma cells. Its biosynthesis and secretion is regulated by endotoxin and by phorbol ester, the latter stimulates PAI-2 gene transcription over fiftyfold. PAI-2 was purified from placenta, leukocytes and phorbol ester stimulated U-937 cells. It exists in two different forms, a 47 kDa nonglycosylated form and a 60 kDa glycosylated form, with similar PA inhibition characteristics. The reaction rate of PAI-2 with uPA is at least one order of magnitude slower than that of PAI-1 (Sprengers and Kluft, 1987). PAI-1 and PAI-2 share only about 25% homology with each other at the active site whereas PAI-1 and PAI-2 both share some 33% and 38% homology with other serpins. Furthermore, a complete comparison of the primary structure of PAI-1 and PAI-2 failed to show other regions of homology significantly greater than those with other serpins. It thus appears that PAI-1 and PAI-2 are two evolutionary very distant serpins that somehow acquired the same enzyme

specificity (Lawrence et al., 1990). They constructed three chimeric mutants of PAI-1 where the strained loop was replaced with the similar region from PAI-2, antithrombin III or with an artificial serine protease inhibitor superfamily consensus strained loop. Their results suggested that structures outside of the strained loop are responsible for the major differences in specificity between PAI-1 and PAI-2.

PAI-3 is a 51 kDa protein purified from urine which has been shown to react slowly with the two chain forms of u-PA and t-PA and with thrombin (Sprengers and Kluft, 1987).

1.3.2 *Erythrina* trypsin inhibitor

The trees and shrubs of the genus *Erythrina* which belongs to the Leguminosae family, are distributed throughout tropical to warm regions of the world (Raven, 1974). As is the case with other legumes, it was established that seeds from the Southern African species of *Erythrina*, viz. *E.acanthrocarpa*, *E.caffra*, *E.decora*, *E.latissima*, *E.lysistemon* and *E.zeyheri* contain large concentrations of trypsin and chymotrypsin inhibitors (Joubert *et al.*, 1981). A number of these were purified by gel filtration on Sephadex G-50 followed by ion-exchange chromatography with DEAE-cellulose and DEAE-sepharose and numbered according to the fraction in which they were located (Joubert, 1982). The *Erythrina* protease inhibitors were found to resemble other inhibitors of the STI Kunitz family of serine protease inhibitors. They have molecular weights of approximately 20 kDa and contain about 170 amino acids that include four half cysteine residues. The inhibitors can be

allocated to three groups on the basis of their relative abilities to inhibit chymotrypsin, trypsin and tPA. Group a inhibitors were relatively specific for chymotrypsin (which they inhibited strongly) but were relatively poor inhibitors of trypsin and had no apparent effect on tPA. Proteins in this group had blocked amino-termini. Group b proteins inhibited trypsin strongly and chymotrypsin less effectively. They had no effect upon tPA. None of the proteins of this group had free amino-terminal amino acids. Group c comprised proteins that were potent inhibitors of trypsin, α-chymotrypsin and tPA. All of the proteins in this group have the same free amino-terminal amino acid (valine) and all inhibit tPA strongly (Heussen *et al.*, 1987). This genus is unusual in that other plant-derived Kunitz-type inhibitors that have been tested have no effect on plasminogen activators (Zimmerman *et al.*, 1978; Dano and Reich, 1979). One of the group c inhibitors, isolated in the DE-3 fraction of *E.latissima* was examined further and was found to have the selective ability to inhibit tPA without affecting the action of uPA. (Heussen *et al.*, 1984). When coupled to agarose it was used to purify both one- and two-chain forms of tPA.

Another group c inhibitor from the DE-3 fraction of *E.caffra* called *Erythrina* trypsin inhibitor (ETI) was found to have the same selective properties but a higher specific tPA inhibitory activity (Joubert *et al.*, 1987). This is remarkable as the primary structures of these two inhibitors are identical except for a difference of four amino acids at positions 65, 72, 123 and 160. Both are proteins of 172 amino acids (about 20 kDa) with four half-cysteine residues and a scissile bond at Arg63-Ser64 and both bear good homology to STI and winged bean

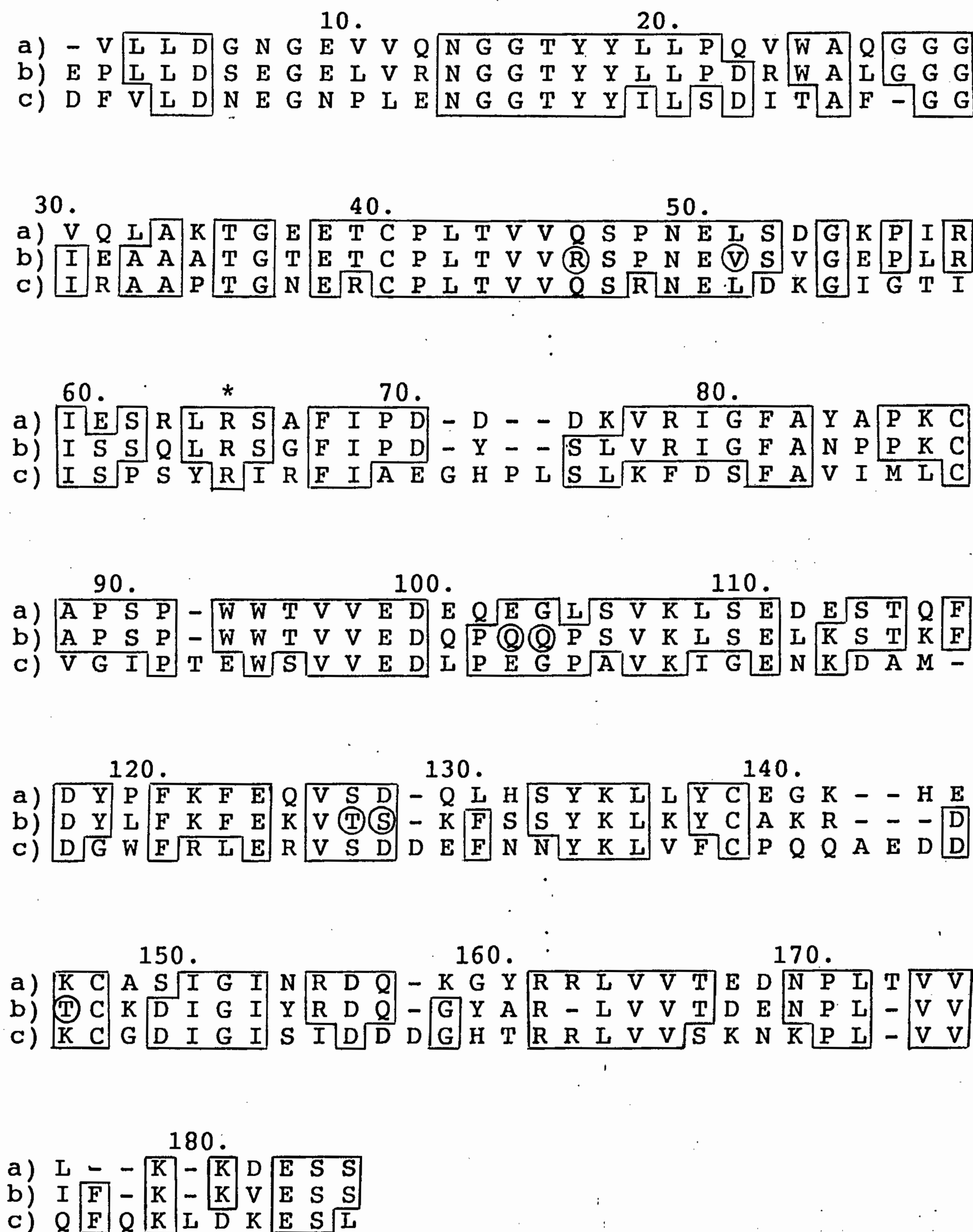


Figure 1.1 Comparison of the amino acid sequence of Kunitz-type trypsin inhibitors from three seeds :a) *E.caffra* (ETI), b) winged beans (WTI) and c) soybeans (STI). * = trypsin reactive site. Circles in boxed regions indicate invariant amino acids. From Joubert and Dowdle (1987).

trypsin inhibitor (WTI) as shown in Figure 1.1. However, STI and WTI are inactive against t-PA (Joubert and Dowdle, 1987). It is argued by Joubert and Dowdle (1987) that the change at position 72 from an acidic Glu residue in *E.latissima* trypsin/t-PA inhibitor to a basic Lys residue in ETI may be responsible for the difference in t-PA- affinity between the two molecules. WTI has a very hydrophobic Leu residue at the corresponding position and has no t-PA inhibitory activity, they suggest that hydrophilicity in this region may be required for inhibition.

The position of the scissile bond was determined by treatment of ETI with catalytic amounts of trypsin at pH 3.5. This resulted in cleavage of the inhibitor into two fragments of molecular weights 12.4 kDa and 7.5 kDa. The cleaved peptide bond was resynthesized by incubation of the purified trypsin-cleaved products with trypsin at pH 8. ETI was not cleaved by tPA under the same conditions that led to its being cleaved by trypsin. In addition, tPA was not able to resynthesize the cleaved peptide bond of trypsin-cleaved ETI products. This indicated that the interaction of ETI with trypsin is kinetically different from its mode of interaction with tPA (Schemmer-Heussen and Dowdle, in press). The cleavage of ETI resulted in an apparent decrease in hydrophobicity of the molecule and this was used to separate modified from unmodified ETI by chromatography on phenyl-sepharose after trypsin and trypsin-inhibitor complexes had been removed from the reaction mixture by ion exchange chromatography on DEAE-cellulose.

The specific inhibitory activity of ETI for both trypsin and tPA was however reduced when the inhibitor was cleaved at the reactive site. On average 5.8 times more cleaved than native inhibitor was needed to inhibit 50% of the trypsin activity in this assay. In the case of tPA 9.8 times more cleaved inhibitor was required. When the Arg residue at P1 had been removed by treatment of cleaved ETI with carboxypeptidase B, the specific inhibitory activity towards trypsin was reduced 146-fold and that towards tPA 243-fold. Thus the integrity of the ETI reactive site Arg63 is important for activity against both trypsin and tPA (Heussen-Schemmer *et al.*, 1991).

To confirm this observation, Heussen-Schemmer *et al.*(1991), converted the reactive site Arg of ETI to citrulline by peptidyl arginine deiminase. ETI and STI are very similar in their primary structure as can be seen in Figure 1.1. It was shown by Takahara *et al.* (1985) that treatment of STI with peptidyl arginine deiminase specifically modified the inhibitor at its reactive site Arg preventing its ability to inhibit trypsin. ETI modified in this manner had its ability to inhibit trypsin or tPA reduced to undetectable levels in the chromogenic assay.

X-ray crystallographic data on complexes between STI and trypsin (Sweet *et al.*, 1974) and between pancreatic trypsin inhibitor and trypsin (Ruhlmann *et al.*, 1973; Huber *et al.*, 1974) have shown that in addition to points of contact at the active site there is contact between enzyme and inhibitor at a

small number of residues in both molecules. This was also seen to be the case with interaction between PAI-1 and tPA as discussed previously. Secondary contact sites are probably also important in the interaction between tPA and ETI, since tPA which has been inactivated at its active site serine residue with diiso-propyl-fluorophosphate will still bind to immobilized ETI (Schemmer-Heussen and Dowdle, in press). ETI has been crystallized and its structure solved and will be discussed in more detail later. This has shown that ETI, like STI, has an external loop structure that contains the reactive site, Arg63-Ser64 through which ETI interacts with both trypsin and tPA. The external loop structure is typical of the canonical group of serine proteinase inhibitors which was shown (as discussed earlier) to be stabilized by the rest of the molecule (the core molecule).

To examine the ability of the active site loop region of ETI to function as an inhibitor of trypsin and tPA a synthetic hexadecapeptide Glu-Ser-Arg-Leu-Arg*-Ser-Thr-Phe-Ile-Pro-Asp-Asp-Asp-Lys-Val-Arg which includes the trypsin reactive site (Arg*-Ser) of ETI was produced and tested for its inhibitory activity. On a molar basis, approximately ten thousand times more peptide than native ETI was required to achieve the same degree of inhibition of trypsin. No inhibition of tPA by the linear peptide was observed (Heussen-Schemmer *et al.*, 1991). These results illustrate the need for contact regions in the inhibitor other than those surrounding the P1 residue and for the overall tertiary structure of ETI to confer a suitable conformation on the reactive site loop.

The salient features of ETI are:

1) It is an inhibitor of trypsin, chymotrypsin and tPA.

2) The active site residue is Arg63 and the scissile bond is between Arg63-Ser64.

3) The active site is on an external loop that relies on structure that is imparted and stabilized by the rest of the molecule, for its ability to inhibit the activity of trypsin and tPA.

The three dimensional structure of ETI was determined to 2.5 Å resolution by the isomorphous replacement method (Onesti *et al.*, 1991). It is composed largely of twelve β-strands joined by long loops. Six of these form an antiparallel, roughly cone shaped barrel. The wider end is capped by the remaining antiparallel β-strands arranged in pairs and joined to the barrel by six loops extending outside and perpendicular to the axis of the barrel. This gives the molecule a roughly spherical shape with diameter of ±40 Å (44 x 40 x 40 Å). The structure is stabilized by two disulphide bridges: Cys39-Cys83 and Cys132-Cys139, which are well removed from Arg63 and do not contribute to its stability. The active site loop, the N-terminal loop and a short segment bearing the C-terminus protrude from the bottom of the barrel. ETI shows approximate 3-fold symmetry about the axis of the barrel, the repeating unit consisting of four sequential β-strands and the connecting loops. The structure is shown in Figure 1.2.

The structure of the active-site loop (residue 61 to 67) is stabilized by hydrogen bonds and not by secondary structure

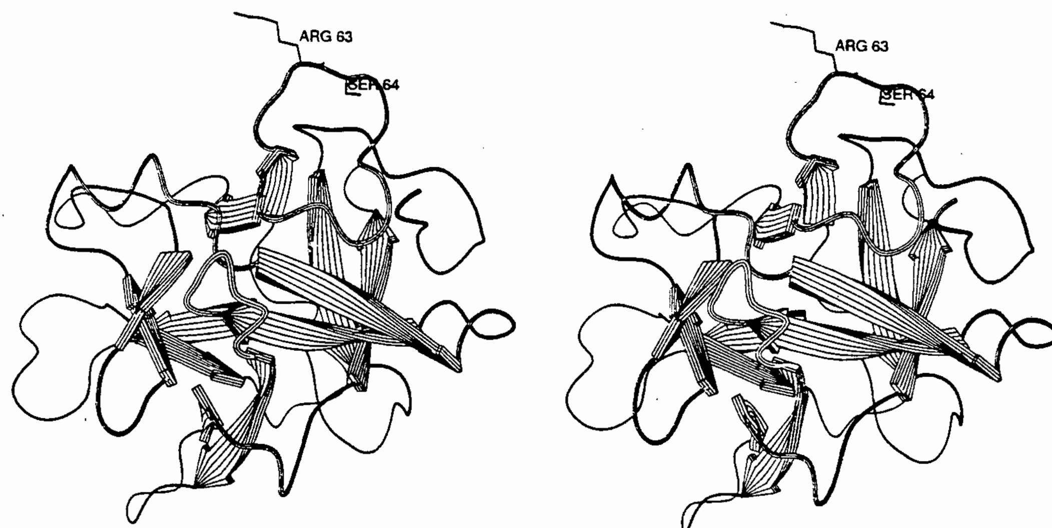


Figure 1.2 A stereo-ribbon drawing of ETI showing the twelve β-strands and their interconnecting loops. The top righthand six of these are arranged in antiparallel pairs into the form of a roughly cone shaped barrel. The axis of the barrel lies in a north-easterly direction in this figure from Bode and Huber (1992). At the bottom lefthand of the figure, the six β-strands are arranged antiparallel in pairs and close off the wider end of the barrel. At the uppermost righthand of the figure, the active site loop, the N-terminal finger and the C-terminus can be seen to project from the narrow end of the barrel and lie in close proximity to each other.

elements such as disulphide bridges that could limit its conformational freedom as is usual with other peptide protease inhibitors. Hydrogen bonds of particular importance are from two residues within the loop, Leu62 (oxygen atom) and Ser60 (hydroxyl group and nitrogen atom), to the side chain amide group of Asn12. In addition, the residues 65 to 67 of the active site loop are involved in hydrogen bonding with residues 1 to 4 and 8 to 12. The three segments form the sides of a triangle which is stabilized by a molecule of water at its centre and antiparallel β-type hydrogen bonds at the angles. In addition to stabilizing the active site loop these interactions give structure and stability to a finger-like region that contains the N-terminal Val of ETI and involves residues 1 through 12. The interactions stabilizing the N-terminal finger have been modelled by Onesti *et al.* (1991) and depicted in the form of a schematic diagram shown in Figure 1.3. This same region was modelled by computer (Maeder, 1989 unpublished) and shown in Figure 1.4 as the carbon backbone and molecular surface of the region in relation to the reactive site loop. There are a number of hydrophobic residues, viz. Val1, Leu2, Leu3, Val9, Val10, Ala65, Phe66, Pro67, Ile68, involved in the N-terminal region and it appears from the structures shown in Figures 1.3 and 1.4, that hydrophobic interactions add to the stability of this N-terminal structure. Through these interactions, the N-terminal finger of ETI is fixed to the reactive site loop and rendered immobile relative to the loop. As a result, in ETI-protease complexes, the N-terminal finger is positioned close to the protease.

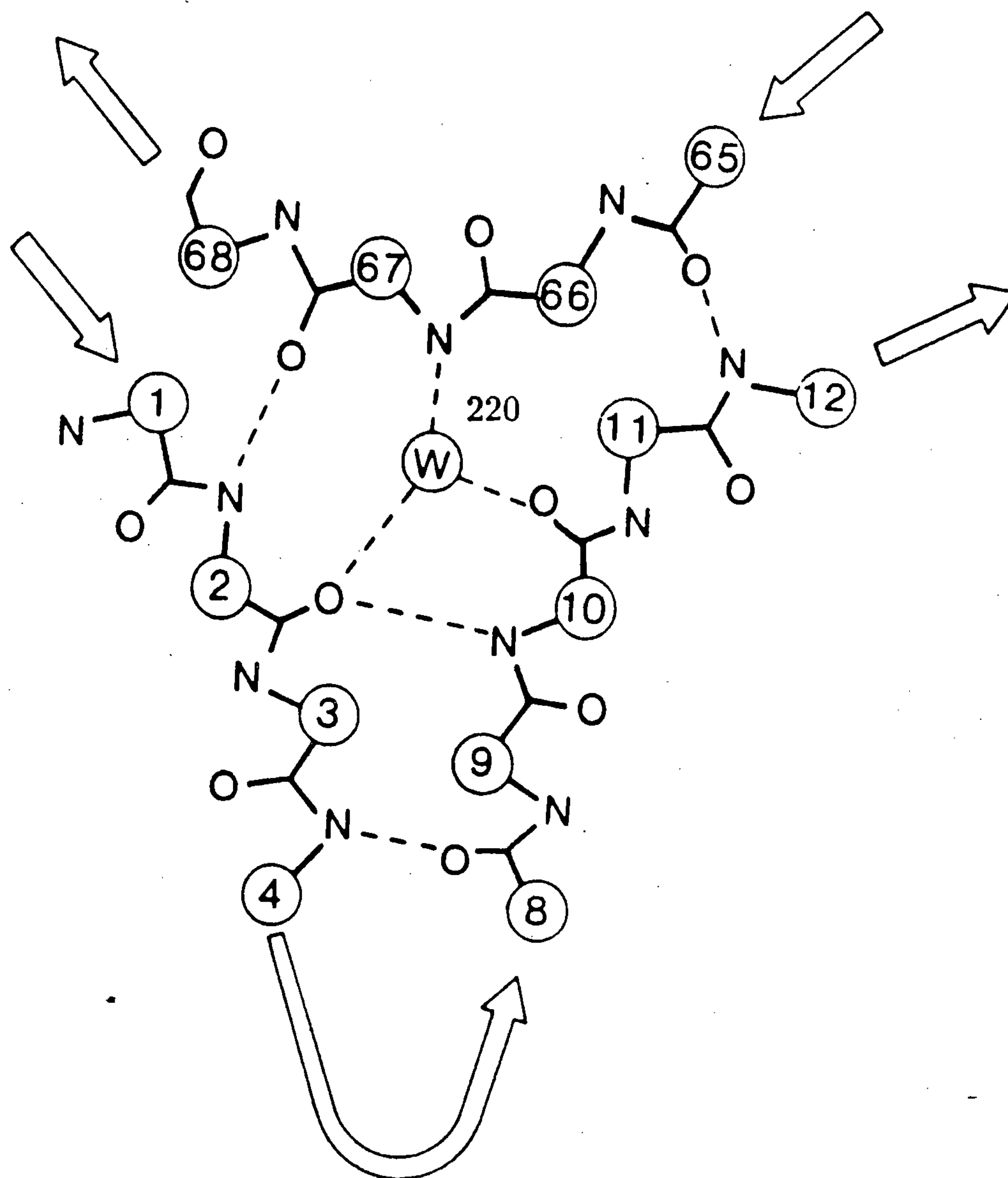


Figure 1.3 Schematic diagram of the amino-terminal finger region of ETI from Onesti *et al.* (1991). Interactions between residues 65 to 67 of the reactive loop interact with residues 1 to 4 and 8 to 12 in a triangular formation with a water molecule to produce and stabilize this amino terminal finger-like region of the molecule.

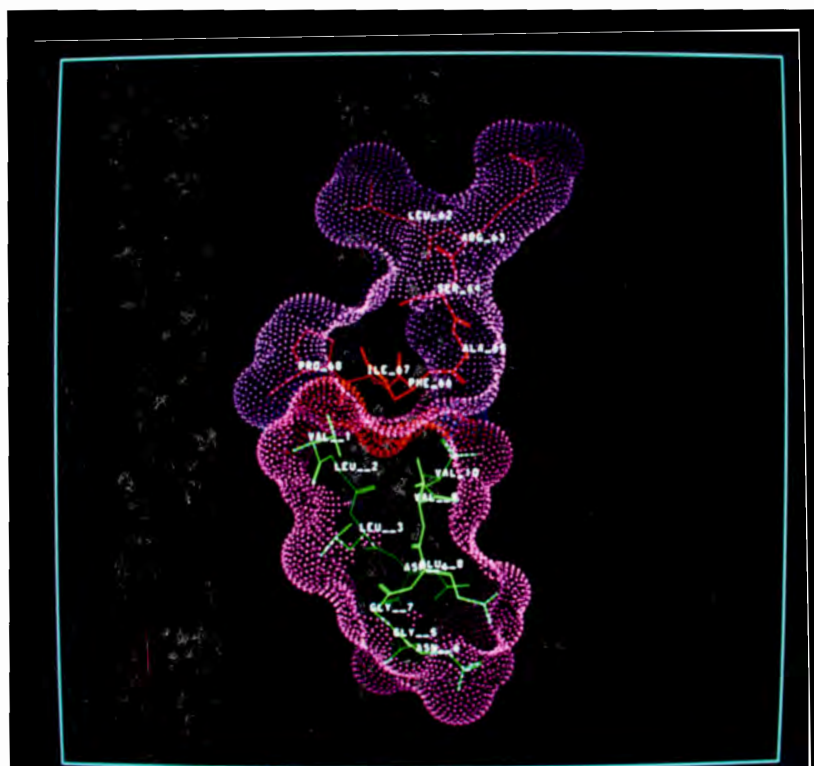


Figure 1.4 Computer model of the N-terminal region of ETI in relation to the reactive site loop (Maeder, 1989 unpublished). The backbone of ETI in this region and the molecular surface is shown.

It was shown in Figure 1.1 that there is good homology between ETI and STI and between ETI and WTI. The homology between ETI (a 172 amino acid protein) and WTI (171 amino acids) is greater at 113 homologous amino acids, than that between ETI and STI (181 amino acids) where there are 75 homologies. In addition, STI is similar to ETI also from the point of view that STI can be reversibly hydrolyzed at its scissile bond Arg63-Ile64. Nonetheless, while they inhibit trypsin, both STI and WTI do not inhibit tPA. Thus it was considered that comparison of the three-dimensional structures of these three molecules would help to understand the unique ability of ETI to inhibit tPA. However, the crystal structure for WTI has not as yet been determined.

The structure of the complex between STI and porcine trypsin has been determined at 2.6 Å resolution (Sweet *et al.*, 1974; Blow *et al.*, 1974). The resulting electron density maps of some parts of the inhibitor were not clear and a precise interpretation could be made of the residues 1-93 of STI. Another short stretch of density which appeared clearly on the map was tentatively assigned to residues 116 to 122. For the segments 94 to 115 and 123 to 176, only a Cα tracing was available. Despite these difficulties, a model for STI was derived with a certain amount of confidence. This structure shows an unusual symmetrical three-lobed fold consisting of a β-sheet barrel of six antiparallel strands interconnected by irregular loops (McLachlan, 1979). This gives an overall spherical structure to STI with diameter of $\pm$ 35 Å. Two disulphide bridges are found, both of which are on the surface

of the molecule. Cys39-Cys86 is found on the bottom half, the half that contains the carboxy terminal section of the molecule. Cys138-Cys145 is on the top half of STI along with its amino terminal section. The region that interacts with trypsin is found on a curved loop of residues 61 to 66 which protrudes from the core of the inhibitor and contains the active site Arg63 and scissile bond Arg63-Ile64. In addition, the carbonyl group of Asp1 of STI forms an ion pair with the ϵ-amino group of Lys60 of trypsin. This is the only ion pair interaction between STI and trypsin outside the primary binding site. The STI structure that was modelled by Sweet *et al.*, (1974) from, inter alia, the data discussed above is shown in Figure 1.5.

STI has a topology very similar to that of ETI conferred by the degree of amino acid homology between the two inhibitors. Eighty three Cα positions were superposed in the two structures with an r.m.s. deviation of 1.34 Å as shown in Figure 1.6a (Onesti *et al.*, 1991). The superposition of the two models shows that qualitatively the β-strands are similar while some of the loops are not. In particular, the loop between residues 61 and 67 (the reactive site loop), is shifted such that the reactive site of ETI is displaced by ± 4 Å with respect to the reactive site loop of STI. The differences in conformation do not disappear even if a more local superposition of atoms is performed, overlapping residues 60 to 65 gives an r.m.s. difference of 0.64 Å. This is shown in Figure 1.6b. As a consequence of the active site displacement and due to the

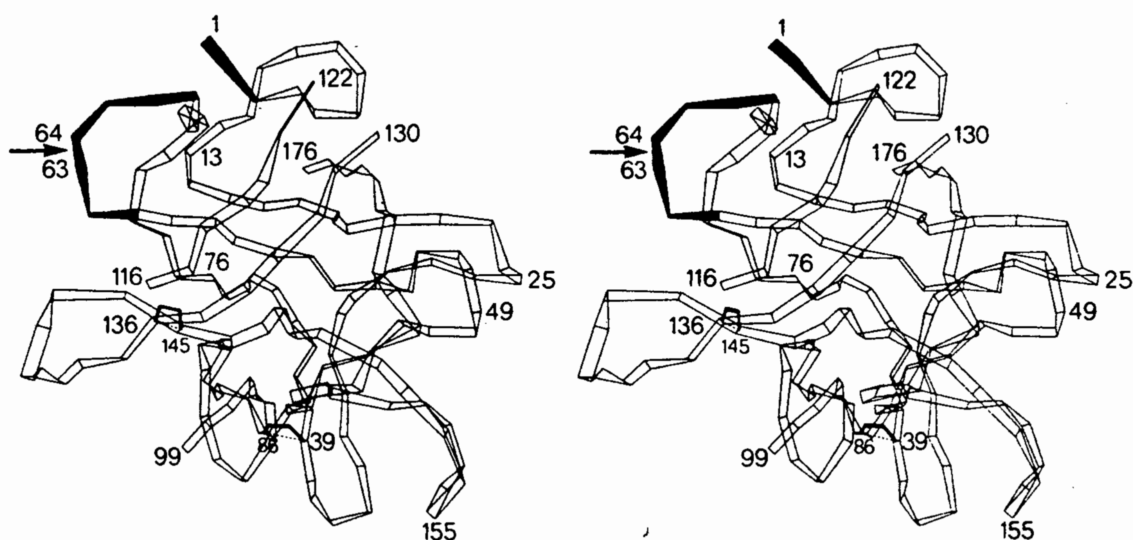


Figure 1.5 The three-dimensional structure of STI as determined from the crystal structure of the complex with trypsin and the amino acid sequence. The residues 107-115, 123-129 and 177-181 of STI could not be found, thus the assignment of residues 116-122 was tentative. The two disulphide bridges are shown, the residues that make contact with the trypsin are blackened and the arrow indicates the scissile bond (taken from Sweet *et al.*, 1974 and Blow *et al.*, 1974)

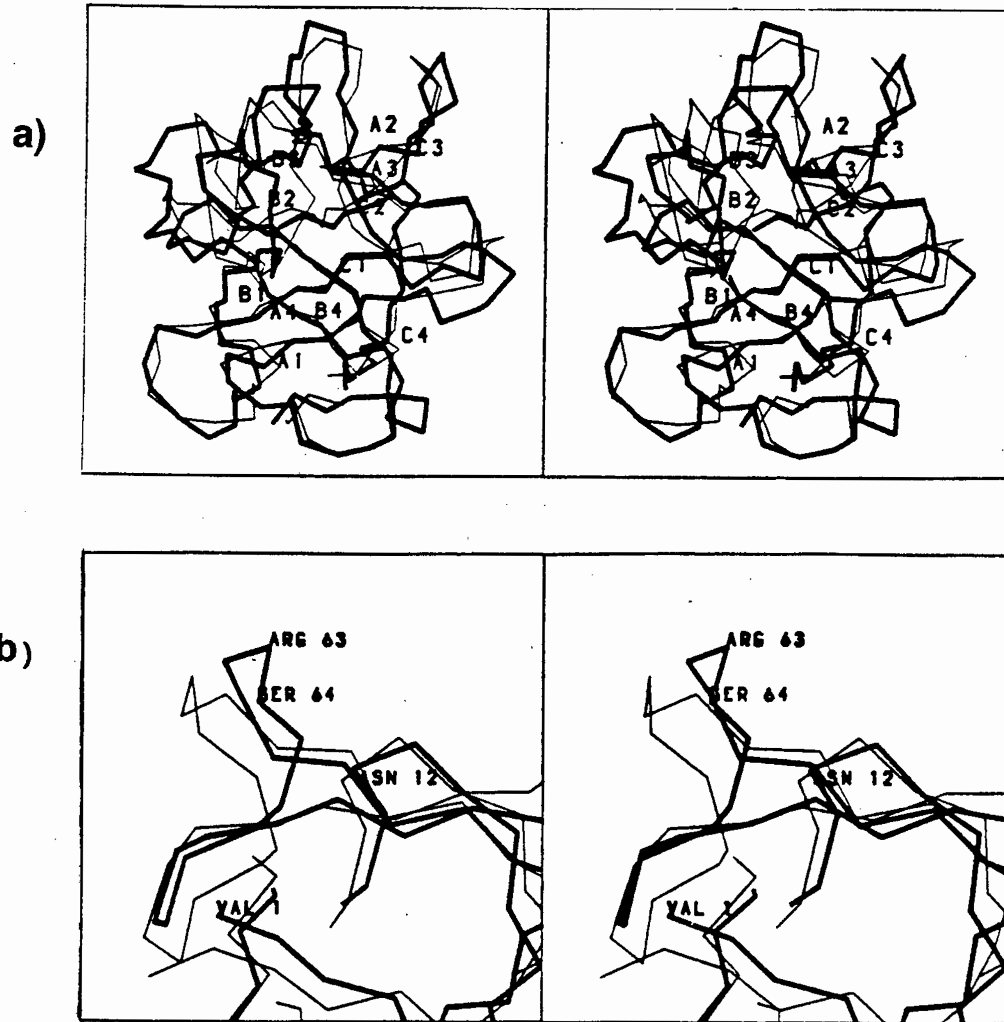


Figure 1.6 Stereoscopic superposition of the Cα backbone of ETI (thick line) and STI (thin line). From Onesti *et al.* (1991).

a) ETI and STI have similar topology and 83 Cα positions can be superposed in the two structures. The axis of the barrel runs in a north-easterly direction as in Figure 1.2 but here the reactive site loop (loop A4-B1 in this diagram) and N-terminal finger region (on loop A1) are at the bottom left corner of the diagram. From this diagram it can be seen that STI has an N-terminal region structured similarly to that in ETI.

b)Examination of superposition of the reactive site loop regions of ETI and STI with scissile bond, Asn12 and N-terminus indicated. From this it can be seen that the active site of ETI in the crystal lattice is displaced relative to that of STI. In addition, the position of the N-terminal residue of STI in relation to its reactive site loop is indicated.

large size and globular shape of ETI, when the loop is modelled such that the residues P4-P'3 of ETI overlap with the corresponding STI residues, the alignment of the two molecules has to be altered in relation to the STI:trypsin alignment. This gives rise to new potential interactions that are not present in the STI:trypsin complex. Onesti *et al.* (1991) argue that since ETI and STI show high sequence and structural homology and the strength of their interaction with trypsin is of the same order of magnitude, it seems unlikely that the region in each molecule in contact with trypsin is very different. Thus they suggest that the observed displacement may be as result of intermolecular contacts in the crystal lattice involving the reactive site loop and the loop 20 to 23 of a symmetry-related molecule. Within these regions the guanidinium group of Arg63 makes strong hydrogen bonds with the carbonyl of Pro20 on the adjacent molecule, while the hydrophobic segment of the Arg side chain is packed against the aromatic group of Trp23 (Onesti *et al.*, 1991). Taking this into account, allowed the building of an ETI:trypsin complex that is very similar to the model built by Sweet *et al.* (1974) of the STI:trypsin complex.

From Figure 1.6a it can be seen that the N-terminus of STI is arranged into a finger-like structure similar to that of ETI. The STI model is not of sufficient quality to allow an analysis of secondary structure elements in terms of hydrogen bonds. However, Figure 1.6b indicates that the two N-terminal regions follow each other closely and that the N-terminus of STI appears to be a residue longer than that of ETI as expected from the alignment of homologous sequences of the two proteins

(Figure 1.1). Furthermore, where the sequence of the ETI and STI N-terminal finger regions are not homologous, there has been a conservation of the hydrophobic and hydrophilic residues as can be seen in Figure 1.1. This suggests that the structure of this finger-like region of STI is stabilized by hydrogen bonds and hydrophobic interactions as in ETI.

It is reasoned that Asp1 of STI is not included in the structure of the N-terminal finger because of the need of its hydrophilic side chain to remain outside of the area of hydrophobic interactions between Phe2, Phe66, Ile67 and perhaps Ala68. This leaves the carboxyl group of the Asp1 side chain free to form an ion pair with Lys60 of trypsin. The proximity of the Asp1 side chain of STI to the side chain of the trypsin Lys60 side chain when the two molecules interact is shown in Figures 1.7 and 1.8. Both figures, modelled from the X-ray crystallographic coordinates of Sweet *et al.* (1974), show the the active site cleft of trypsin in relation to the reactive site loop and N-terminal region of STI when the two molecules interact.

Val1 of ETI has neither the proximity nor the side chain groups needed to form an ion pair interaction with Lys60 of ETI. The proximity of these groups is indicated in Figure 1.9 in which ETI was modelled on the Cα skeleton of STI. The fact that the strength of inhibition of trypsin by STI and ETI is of the same order of magnitude indicates that ETI does not require the ion pair interaction in this region to inhibit trypsin. But in the case of STI, it may well be that the ion pair formed between

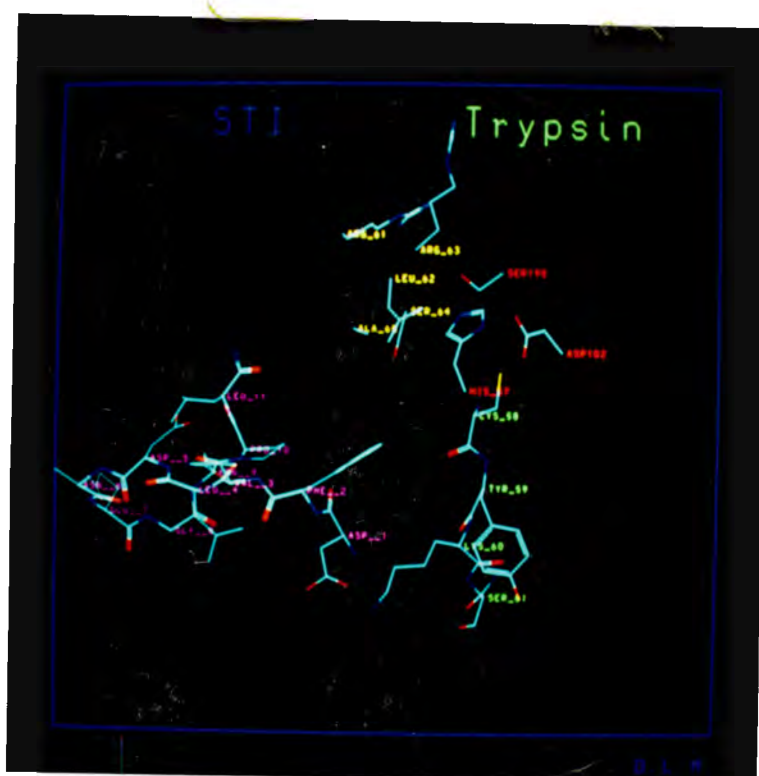


Figure 1.7 Model based on the crystallographic data of Sweet *et al.* (1974) of the area of interaction between STI and trypsin (Maeder, 1989 unpublished). Emphasizes the proximity of the side chains of Asp1 of STI and Lys60 of trypsin when the reactive site loop of STI is positioned over the corresponding residues of trypsin.

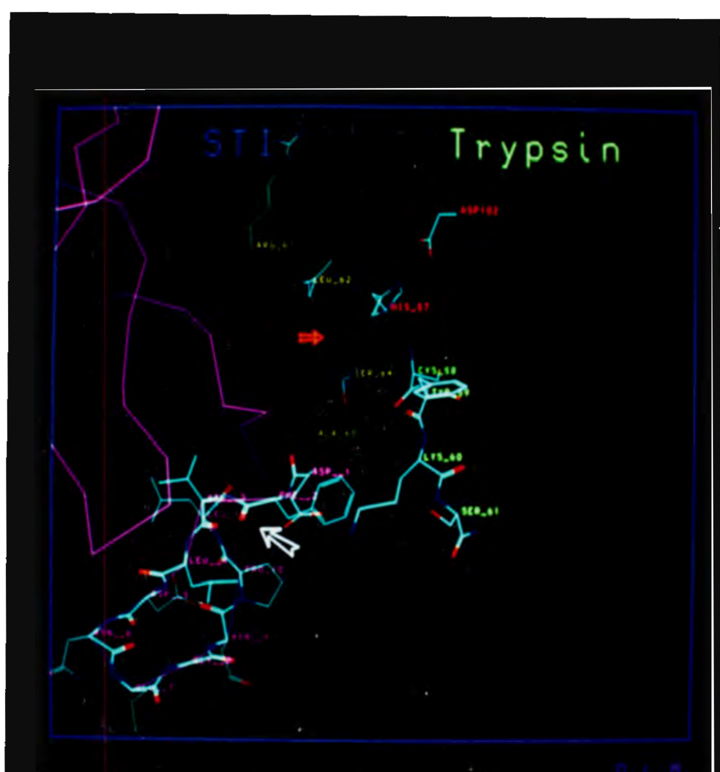


Figure 1.8 Model of the same region as shown in Figure 1.7 viewed from a different angle. From this view it appears that the aromatic ring of Phe2 of STI interacts with Lys60 of trypsin which is not the case as can be seen in Figure 1.7. Here the N-terminal finger of STI is indicated with a white arrow and the reactive site loop (indistinct in this photo.) by a red arrow.

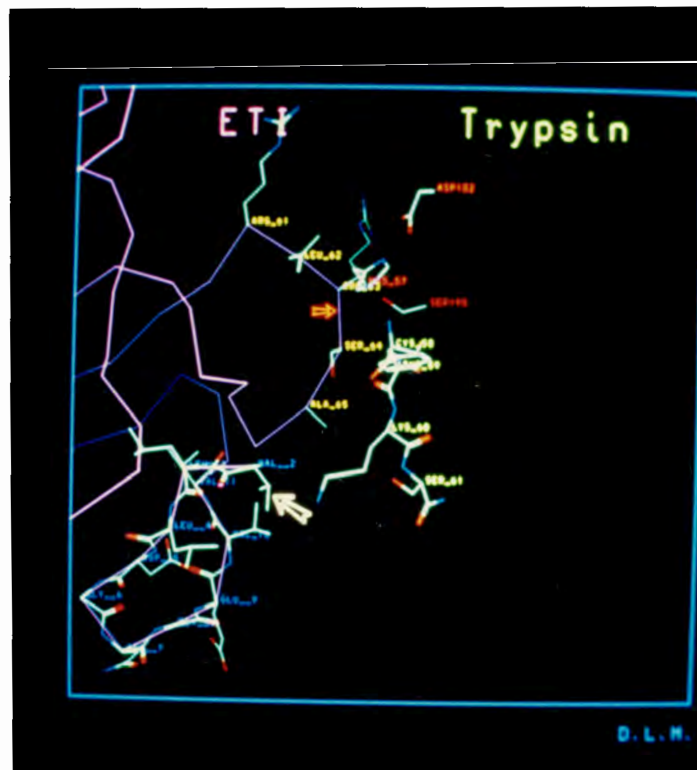


Figure 1.9 Model of the reactive site and N-terminal finger region of ETI modelled on the Cα backbone of STI is shown bound to the active site of trypsin. The N-terminal Val of ETI, numbered Val2 by STI numbering is shown with a yellow arrow and the reactive site of ETI by an orange arrow (Maeder, 1989 unpublished).

itself and trypsin is essential to the stability of the inhibition complex.

However, it has been hypothesized, on the basis of computer aided molecular modelling studies, that the N-terminal Asp of STI is the key to the inability of this protein to inhibit tPA (Maeder, unpublished). While there is currently no crystal structure available for tPA, its C-terminal domain (residues 260 to 527) is homologous with a number of serine proteinases of known structure (Onesti *et al.*, 1991). Alignments of the amino acid sequence of the C-terminal domain of tPA with the sequences of trypsin and chymotrypsin suggests that there are a number of insertions in tPA not seen in either trypsin or chymotrypsin. Based on the work of Bode *et al.* (1989) on thrombin, these insertions are likely to be accommodated in the form of protrusions. As discussed earlier, the B-chain of human α-thrombin is homologous to chymotrypsin and trypsin. Thrombin cleaves similarly to trypsin, carboxy terminal to basic amino acid residues, but thrombin is much more specific and cleaves fewer bonds in macromolecular substrates. Comparison of the crystallographic structures of trypsin, chymotrypsin and thrombin has shown that the specificity properties of thrombin towards macromolecular substrates are mainly determined by its insertions. Some of these result in a narrower and deeper active site cleft in thrombin in comparison with that of chymotrypsin and trypsin, while others prevent inhibition of it by protease inhibitors that inhibit trypsin and/or chymotrypsin (Bode *et al.*, 1989). Thus the insertion of five amino acids between positions 62 and 63 (trypsin numbering) is thought to be accomodated by a protrusion in this area. Part of the

aligned sequences of trypsin and tPA, including the region in question, is shown in Figure 1.10. It is hypothesized that the presence of this protrusion, not seen in trypsin, clashes with Asp1 of STI preventing it sterically from approaching close enough to tPA to form an STI:tPA inhibitory complex. As ETI has one less residue in its N-terminal finger region than STI, it is able to approach close enough to complex with tPA. Models of the interaction of STI and ETI with tPA (modelled on chymotrypsin structure) to illustrate the concepts described above are shown in Figures 1.11 and 1.12.

```
     .42              .50               .60            .63
a)...C G G S L I N S Q W V V S A A H C Y K S G - - - - - I Q
b)...C G G I L I S S C W I L S A A H C F Q E R F P P H H L T

            .70                  .80
a)   V R L G E D N I N V V E G N E Q...
b)   V I L G R T Y R V V P G E E E Q...
```

Figure 1.10 Partial alignment of the amino acid sequences of: a) trypsin and b) tPA showing the region of interest around Lys60 the residue that is involved in ion pairing with STI. Five residues have been inserted in tPA between residues 63 and 62 by trypsin numbering.

1.4 Aims of this work

The unique ability of ETI to inhibit tPA as well as trypsin and chymotrypsin as well, has endowed it with potential commercial value. In addition, the accumulated knowledge on ETI and its homology with other important protease inhibitors makes it a good candidate for further study. Questions on the nature of the interaction between inhibitor and protease can be addressed by making modifications to the protein. The simplest, and in many cases the only way to achieve this is to make

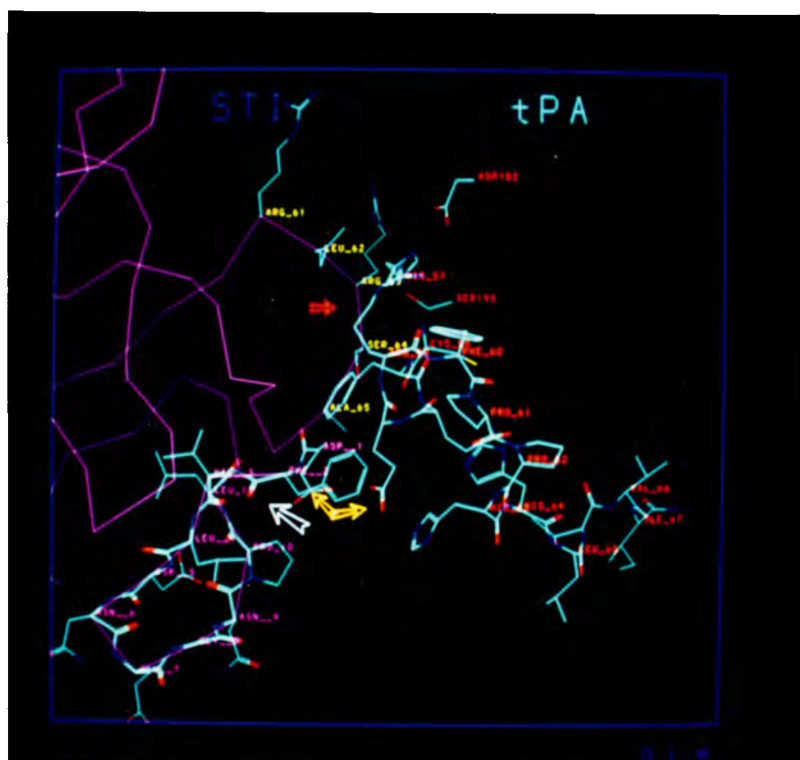


Figure 1.11 Model of the interaction of STI and tPA showing the reactive site loop (red arrow) and N-terminal finger region of STI (white arrow), the active site cleft of tPA and the approach of Asp1 of STI to the insertion sequence of tPA which prevents the formation of the STI:tPA complex. Here STI is modelled on the coordinates of Sweet *et al.* (1974) and tPA is modelled on chymotrypsin structure.

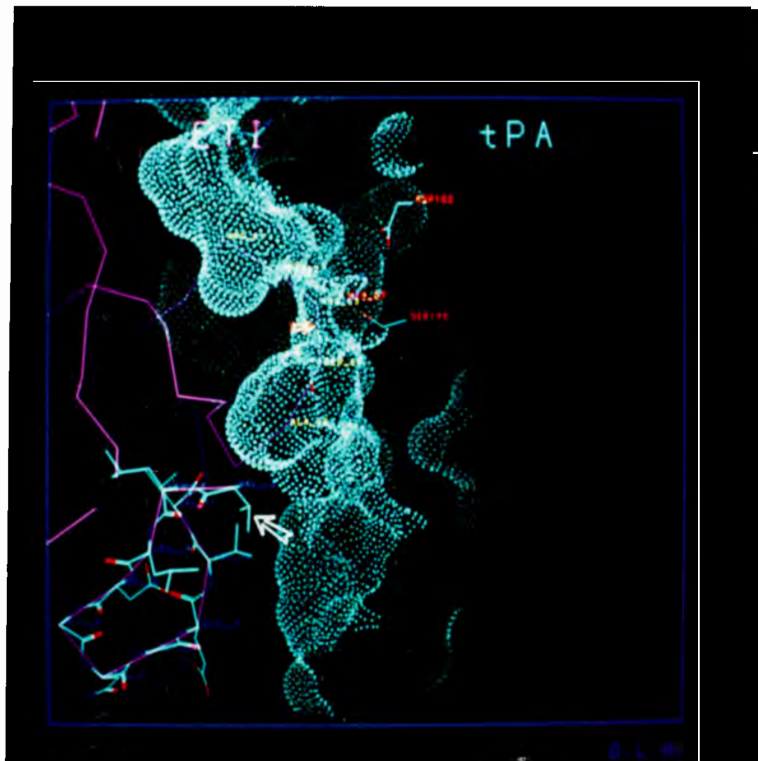


Figure 1.12 Model of ETI:tPA complex showing proximity of Val1 of ETI (white arrow) to the tPA surface when the reactive site loop of ETI is bound to the tPA active site (orange arrow). Again ETI was modelled on the STI backbone and tPA is modelled on chymotrypsin.

modifications to the gene that encodes the protein, and to measure the change in activity produced in the expressed protein. Gene expression technology has been well established in E.coli. However, plant gene codon preferences differ sufficiently from codon preferences of E.coli (Murray et al., 1989) to make high levels of expression in E.coli, of a gene cloned from a plant, unpredictable. Thus the aims of this work were to design a gene encoding ETI that would allow for easy manipulation to produce modified forms of the protein. The unaltered gene was to be expressed in an appropriate E.coli expression system to produce ETI with tPA specific inhibitory activity of the same order of magnitude as ETI isolated from *E.caffra*. Activity of modified forms of ETI could be compared with that determined for the unmodified protein.

The synthetic ETI-encoding gene could then be altered to produce any number of modified forms of ETI as an aid to understanding its inhibitory activities and protease inhibtion in general. The question to be addressed in this work is the hypothesis that the N-terminal Asp of STI prevents it from inhibiting tPA. To this end an altered form of ETI would be produced in which the N-terminus had been extended by an Asp residue. The prediction, based on the model described above is that this form of ETI would have much reduced activity against tPA as compared with the unaltered protein from this gene.

CHAPTER 2

DESIGN AND SYNTHESIS OF AN ETI ENCODING GENE

2.0 Contents

2.1	Considerations prior to design.	50
2.2	Design of an ETI-encoding gene.	60
2.3	Alteration of pUC9 to produce the synthesis vector pSyn.	68
2.4	Synthesis and purification of oligonucleotides for construction of the synthetic ETI gene.	74
2.5	Synthesis of the ETI encoding gene.	78
2.6	Discussion.	83

2.1 Considerations prior to design

The purpose of synthesizing an ETI-encoding gene was to produce the protein in abundance from a cloned gene and, most importantly, to have a useful gene to allow modifications of the protein as a method for studying interactions between ETI and its target enzymes. To this end it was necessary to have, at regular intervals in the gene, unique restriction sites for tailoring ETI to different studies.

The first step in designing a gene is to translate the known amino acid sequence of the protein for which that gene codes into a DNA sequence using preferred codons for the chosen host. Prior to designing, a decision has to be made regarding the organism(s) that will be used for eventual expression, as this will influence the codon choice for each amino acid. Therefore,

it is important at the outset, to have some insight into the effect that codon choice can have on translation rate and accuracy as it may be necessary to use less favoured codons, in order to introduce restriction sites. Secondly, an expression vector should be chosen if this is to be different from the synthesis vector. This will influence the restriction sites required for excision of the gene from the synthesis vector and for insertion into the expression vector.

The establishment of the genetic code has given rise to 61 sense codons that code for 20 amino acids and 3 stop codons. The result is that most of the amino acids are coded for by more than one codon. Where more than one codon is used to code for an amino acid, these are called degenerate codons. When the frequency of use of these degenerate codons is examined, it is seen that different taxonomic groups favour certain codons for particular amino acids. In addition, it is reported that within species, individual genes tend to prefer characteristic codon distributions (Gouy and Gautier, 1982; Bennetzen and Hall, 1982).

Indeed, in unicellular organisms there appears to be a direct relationship between the level of expression of a gene and its degree of codon bias (Grosjean and Fiers, 1982). In these organisms there is a group of genes coding for example, ribosomal proteins, the elongation factors and outer membrane proteins for which codon choice is strongly biased toward a subset of the available codons which are thus called the major codons (Shields and Sharp, 1987; Sharp *et al.*, 1986). In this highly biased group there may be as much as a 100-fold

variation in the usage frequency for preferred as opposed to avoided or rare codons. It is observed that the higher the level of protein production the greater the tendency to use only major codons in the gene (Bennetzen and Hall, 1982; Grosjean and Fiers, 1982).

Studies of codon usage patterns in various species have been undertaken and the results expressed in tables from which the preferred codon for an amino acid in highly biased (or highly expressed), as well as those for lowly biased genes, can be determined. A table of relative synonymous codon usage (RSCU) values was produced by Sharp *et al.* (1988) for inter alia, *E.coli* and *S. cerevisiae.* An RSCU value is obtained by dividing the number of codons observed by the number of codons expected if all codons for that amino acid were used equally. The observed codons for determining these RSCU values were taken from the highest 10% and lowest 10% of 165 *E.coli* genes and 154 *S.cerevisiae* genes. From these and similar tables, the preferred codon for an amino acid that would be utilized by a highly expressed gene can be determined. A summary of these codons for *E.coli* and *S.cerevisiae* as used in the design of the ETI-encoding gene is presented in Table 2.1.

Amino Acids	Codons				Compromise
	E.coli		S.cerevisiae		
	High	Low	High	Low	
Ala	GCU	GCG	GCU	GCA	---
Cys	UGC	UGC	UGU	UGU	UGC
Asp	GAC	GAU	GAC	GAU	---
Glu	GAA	GAA	GAA	GAA	---
Phe	UUC	UUU	UUC	UUU	---
Gly	GGU	GGC	GGU	GGC	---
His	CAC	CAU	CAC	CAU	---
Ile	AUC	AUU	AUC	AUU	---
Lys	AAA	AAA	AAG	AAA	AAA
Leu	CUG	CUG	UUG	UUA	CUG
Met	AUG	AUG	AUG	AUG	---
Asn	AAC	AAU	AAC	AAU	---
Pro	CCG	CCG	CCA	CCA	CCA
Gln	CAG	CAG	CAA	CAA	CAA
Arg	CGU	CGC	AGA	AGA	CGU
Ser	UCU	AGC	UCU	UCU	---
Thr	ACC	ACC	ACC	ACA	---
Val	GUU	GUG	GUU	GUA	---
Trp	UGG	UGG	UGG	UGG	---
Tyr	UAC	UAU	UAC	UAU	---

Table 2.1 A summary of the preferred codons for highly expressed and lowly expressed genes in E.coli and S.cerevisiae and the compromise codons used in design of the ETI gene when differences exist in major codon preferences between the two species.

It has been postulated that the preferred codons reflect the optimization of codon: anticodon interaction energies (Grosjean and Fiers, 1982). However, examination of codon preference tables such as the RSCU tables of Sharp *et al.* (1988) shows that different organisms use different subsets of major codons. Andersson and Kurland (1990) argue further that every codon is used as a major codon by some system and as a minor codon by another. This, they feel, indicates strongly that the intrinsic nature of the codons cannot account for their preferred usage.

There is evidence to suggest that translation rates can differ at different codons. For example, Sorensen *et al.* (1989) inserted a piece of a ribosomal gene into the *lacZ* gene either directly or flanked by a few frame shifting bases, leaving the reading frame of the *lacZ* gene unchanged. This gave minor codons in one reading frame and major codons in the other. The translation time for these modified *lacZ* mRNAs was found to be faster by six-fold for major codons than for minor codons. In addition, Andersson and Kurland (1990) quote results of Sorensen and Pedersen (Abstr. 13th Int. Transfer RNA Meet. 1989, abstr. no. mo-am-13). These workers measured the translation rates of strings of GAA and GAG, respectively the major and minor codons for Glu. Both are thought to be served by the same tRNA, $tRNA_2^{Glu}$, but the GAA string is translated about 3 times faster *in vivo* than the GAG string. Thus it can be concluded that the major codons are recognized and translated at higher rates than minor codons.

Available data indicates that the relative amounts of the different tRNA species are not constant but vary in systematic

ways. For example, it was found that the concentration of the major Leu isoacceptor CUG increases progressively at higher growth rates. Likewise, the level of a minor Leu species that can also translate the major Leu codon ($tRNA_3^{Leu}$, CUA, CUG) increases almost threefold while concentration of the remaining Leu species decreases as growth rate increases. In parallel to this, the heterogeneity of the protein population of a cell is also growth rate dependent. Under low growth conditions, the proteins produced by bacteria are fairly equally represented in their protein population. In contrast at highest growth rates, mRNA that codes for key growth components such as translational, transcriptional and membrane proteins (which are coded for by major codons), represents most of the mRNA species. From this it is clear that under good growth conditions the concentrations of the cognate tRNAs for the major codons increase. Conversely, their concentrations decrease as growth conditions deteriorate. From this evidence the conclusion can be drawn that translation rates for most major codons of an organism are limited by the availability of their cognate tRNA species. Furthermore, it is suggested that the selective pressure for major codon preferences in primitive organisms such as *E.coli* and *S.cerevisiae* is a strategy to maximize their growth rates in relatively rich media (Pedersen, 1984; Sorensen *et al.*, 1989; Curran and Yarus, 1989; Andersson and Kurland, 1990).

While it is important to utilize the major codons required by the expression organism, introduction of less preferred codons, where necessary, will not of necessity lower the expression level of the synthetic gene. In fact, high expression levels

are routinely achieved for genes with a relatively high content of less preferred codons. For example, the chloramphenicol acetyltransferase gene which has codon usage similar to that of poorly expressed genes in *E.coli* has been expressed to high levels when cloned under the control of a strong promoter (Robinson *et al.*, 1984). The same phenomenon was reported (Kniskern *et al.*,1986) for *S.cerevisiae* in which the hepatitis B virus core antigen was expressed to high levels even though it has a high level of less preferred codons for this organism.

It may seem that the above argument indicates that it is immaterial to ensure the incorporation into the sequence of as many major codons as possible in order to achieve high expression of a synthetic gene. However, it would seem prudent to utilize every strategy possible to ensure good expression of the planned gene before the investment of the resources required to synthesize a gene. In support of this, there are examples where judicious use of major codons in a synthetic gene has led to increased expression over that of cloned genes that have poor representation of highly utilized codons for that organism. For example, Williams *et al.* (1988) synthesized the gene for human interleukin-2 (IL-2). Here the percentage of preferred codons for *E.coli* was increased from 43% as found in the native cDNA to 85% in the synthetic sequence. Both genes were placed under the control of the *trc* promoter and expressed in JM101, and it was shown that the synthetic gene was able to direct up to 16 times more IL-2 in *E.coli* than the native cDNA sequence even though both genes produce mRNA with similar half lives.

The presence of rare codons, or even short strings of such codons, in a gene is thought by some to constitute pause sites that transiently slow down the movement of ribosomes over mRNA and allow the proper folding of a protein domain. This could be of particular importance when the incomplete polypeptide can take up different secondary structures. A pause site could allow one domain to form before the rest of the polypeptide is produced and interferes with folding of the first domain. In addition, where a protein has more than one disulphide bridge, a pause in synthesis at an appropriate position in the polypeptide chain may allow the correct sequence of -SH pairings to form (Purvis *et al.*, 1987; Swafield *et al.*, 1987).

While ETI was known, from its similarities to STI, not to be a multi-domained protein it does require the two disulphide bridges Cys39-Cys83 and Cys132-Cys139 (Sweet *et al.*, 1974; Onesti *et al.*,1991). In addition, the need for pause sites was not considered to be important in designing an ETI gene for expression in *E.coli* because heterologous proteins have rarely been recovered, in a native state from the cytoplasm of this organism, regardless of the inclusion of any required pause sites. The reason for this is that foreign proteins often are not able to attain their native conformation in the cytoplasm of *E.coli*, especially where the structure is stabilized by disulphide bridges (of which ETI has two) as these apparently cannot form in the reducing conditions found in *E.coli* (Goeddel, 1990). In addition, accumulation of heterologous proteins in *E.coli* often causes precipitation of these proteins

which are sequestered into inclusion bodies. Stringent chemical conditions are then required to release the proteins from the inclusion bodies. Such treatment of inclusion bodies often results in denaturation of all the resident proteins and conditions have to be established which allow for refolding of the required protein (Marston, 1987). Where the protein has been exported from the cell to conditions which may be more conducive to disulphide bridging, pause sites are also of no value since the folding takes place after complete translation of the protein.

It is often not possible to avoid rare codons in the design of a gene because of the need to introduce unique restriction sites. Nonetheless, Robinson *et al.* (1984) advise avoidance of strings of codons which are rarely used in the expression organism. They have found that such a string can lead to unnecessary pausing of mRNA translation and even uncoupling of ribosomes, which can lower the expression level or even prevent it altogether. Moreover, at strings of rare codons, translational accuracy could be compromised (Dix and Thompson, 1989). This guideline is very important in the design of a compromise sequence - one that can be utilized by more than one host, for example *E.coli* and *S.cerevisiae*, which is the strategy adopted by Adams *et al.* (1988) in their design of a gene for HIV transactivator protein.

It is very important to appreciate that the rate of translation of an mRNA species will normally not affect the expression level of the corresponding protein unless the mRNA species can

capture a major fraction of the ribosomes in the cell. This objective can be achieved by the mRNA becoming the dominant, or one of the dominant, mRNA species of the cell. Thus, in the design of a synthetic gene, it is important not only to use as many major codons of the intended expression organism as possible, but also to place the gene under control of strong regulatory elements ie. promoter and ribosome binding site. This will ensure high transcription level and allow it to become a dominant mRNA species of the cell. Nevertheless, all mRNA is inherently unstable and in bacteria the half lives for most messages are a small fraction of cell generation time. Blundell *et al.* (1972) have shown that each *E.coli* mRNA species decays at a unique rate with half-lives ranging from 30 seconds to more than 8 minutes at 37°C. This is thought to be a strategy for rapid adaptation to changes in growth conditions (Kennell, 1986). Thus, even when conditions of major codon usage and strong regulatory elements are present, mRNA may not accumulate to high levels in the cell.

As yet no unique recognition target for initiation of decay has been identified. The most likely primary target is thought to be any single stranded region of the mRNA or regions of weak secondary structure (Kenell, 1986).

2.2 Design of an ETI-encoding gene

As has been discussed, the trypsin and plasminogen activator inhibitor from *Erythrina caffra* contains 172 amino acids (Joubert and Dowdle, 1987). Because of the predicted time and expense of synthesizing a gene for ETI and the uncertainty of being able to express a eukaryotic gene in *E.coli*, it was decided to design the gene so that it could be expressed by *E.coli* and if necessary by *S.cerevisiae*.

These two organisms differ in major codon preference for only 6 amino acids. Using the RSCU values of Sharp *et al.* (1988), which are summarized in Table 2.1, codons for these 6 amino acids were chosen that would not be disfavoured by either organism.

For 3 of these amino acids viz Cys, Lys and Leu the major codon for *E.coli* was used as it could be utilized by *S.cerevisiae*. The rest were assigned codons that could be read by both organisms and the sequence was terminated with a TAA stop codon.

The sequence was then scanned for restriction sites using the GCG program database of 181 restriction enzymes. These were compared with restriction sites for pUC9 (the intended vector for synthesis) to find those sites unique to ETI ie sites that occur once in ETI and not in pUC9. The only such site found was *Kpn*I.

Consequently, the sequence was scanned for partial restriction sites using the GCG program which determines where in a sequence, changes in the wobble base of degenerate codons would introduce restriction sites and still code for the same amino acid. Where such sites would be useful and unique, the appropriate bases were changed so as to introduce the restriction site. This approach allowed the introduction of 15 new restriction sites viz *Xba*I, *Apa*I, *Ban*I, *Nhe*I, *Bsp*EI, *Bgl*II, *Sna*BI, *Nco*I, *Stu*I, *Hind*III, *Bcl*I, *Cla*I, *Bst*EII, *Hpa*I and *Sac*I.

Not all the restriction sites introduced are unique to pUC9/ETI. The restriction site *Hind*III occurs in the pUC9 multicloning site but was considered useful in the position into which it was introduced in ETI. In addition, *Bsp*EI which does not appear in pUC9 was introduced twice in ETI on either side of the trypsin/tPA reactive site loop (Joubert *et al.*, 1987). This allows for removal of the reactive site region with a single digest making this a useful construct for future work aimed at investigation of the reactive site of ETI. The unique restriction sites are shown in Figure 2.2 while the changes to the major codon sequence that were required for these sites is shown in Figure 2.1 and its related notes.

```
i.          1   2   3                                                               20
ii.     H2N-Val Leu Leu Asp Gly Asn Gly Glu Val Val Gln Asn Gly Gly Thr Tyr Tyr Leu Leu Pro
iii.        --- CTG CTG GAC GGT AAC GGT GAA GTT GTT CAG AAC GGT GGT ACC TAC TAC CTG CTG CCG
iv. GAA TTC ATG --T --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --A

i.                  24  25                      31
ii. Gln Val Trp Ala Gln Gly Gly Gly Val Gln Leu Ala Lys Thr Gly Glu Glu Thr Cys Pro Leu Thr
iii.CAG GTT TGG GCT CAG GGT GGT GGT GTT CAG CTG GCT AAA ACC GGT GAA GAA ACC TGC CCA CTG ACC
iv. --- --- --- --C --A --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- ---

i.          45      47                                  55      57                  63
ii. Val Val Gln Ser Pro Asn Glu Leu Ser Asp Gly Lys Pro Ile Arg Ile Glu Ser Arg Leu Arg Ser
iii.GTT GTT CAG TCT CCG AAC GAA CTG TCT GAC GGT AAA CCG ATC CGT ATC GAA TCT CGT CTG CGT TCT
iv. --- --- --A --- --A --- --- --- --- --- --- --- --A --- --G --- --- --- --- --- AGA ---

i.          67                      73                                  81          85
ii. Ala Phe Ile Pro Asp Asp Asp Lys Val Arg Ile Gly Phe Ala Tyr Ala Pro Lys Cys Ala Pro Ser
iii.GCT TTC ATC CCG GAC GAC GAC AAA GTT CGT ATC GGT TTC GCT TAC GCT CCG AAA TGC GCT CCG TCT
iv. --- --- --T --- --- --- --- --- --A --- --- --- --- --- --- --- --A --- --- --- --A ---

i.  87                                  96      98              102 103
ii. Pro Trp Trp Thr Val Val Glu Asp Glu Gln Glu Gly Leu Ser Val Lys Leu Ser Glu Asp Glu Ser
iii.CCG TGG TGG ACC GTT GTT GAA GAC GAA CAG GAA GGT CTG TCT GTT AAA CTG TCT GAA GAC GAA TCT
iv. --A --- --- --- --- --- --- --- --- --A --- --C --- --- --- --G --T --- --- --- --- ---

i.      110             114                     119         122 123                     130
ii. Thr Gln Phe Asp Tyr Pro Phe Lys Phe Glu Gln Val Ser Asp Gln Leu His Ser Tyr Lys Leu Leu
iii.ACC CAG TTC GAC TAC CCG TTC AAA TTC GAA CAG GTT TCT GAC CAG CTG CAC TCT TAC AAA CTG CTG
iv. --- --A --- --- --- --A --- --- --- --- --A --- --- --T --A --- --- --- --- --- --- --T

i.                          137 138     140 141
ii. Tyr Cys Glu Gly Lys His Glu Lys Cys Ala Ser Ile Gly Ile Asn Arg Asp Gln Lys Gly Tyr Arg
iii TAC TGC GAA GGT AAA CAC GAA AAA TGC GCT TCT ATC GGT ATC AAC CGT GAC CAG AAA GGT TAC CGT
iv. --- --- --- --- --- --- --G --G --- --A --G --- --- --- --- --- --- --- --- --- --- ---

i.                                   161 162                 167 168     170 171
ii.  Arg Leu Val Val Thr Glu Asp Tyr Pro Leu Thr Val Val Leu Lys Lys Asp Glu Ser Ser-OH
iii. CGT CTG GTT GTT ACC GAA GAC TAC CCG CTG ACC GTT GTT CTG AAA AAA GAC GAA TCT TCT TAA TGC AGG GAT CC
iv.  --- --- --- --- --- --- --- --- -- -T-A --- --- --- --- --G --G --- --G AGC --- --- --- --- --- --
```

Figure 2.1 Sequence of *Erythrina* trypsin inhibitor.

i. Position of amino acid whose codon has been changed from preferred to less preferred for *E.coli* - see notes below.
ii. Shows published amino acid sequence for *Erythrina* trypsin inhibitor (ETI).
iii. Translation of amino acid sequence into preferred codons for *E.coli*.
iv. Base changes made to the sequence in iii above to introduce unique restriction sites or to accommodate expression by *S.cerevisiae*

Notes to Figure 2.1

1	Val changed to Met - see text.
2	Leu from CTG to CTT and
3	Leu to CTA to introduce *Xba*I site.
20	Pro from CCG to compromise codon CCA for expression in *S.cerevisiae* as per text.
24	Ala from GCT to GCC introduces *Apa*I site.
25	Gln CAG to compromise codon CAA.
31	Leu to CTA gains *Nhe*I site.
45	Gln CAG to compromise CAA.
47	Pro CCG to CCA compromise.
55	As above.
57	Arg CGT to CGG to introduce *BspE*I site.
63	Arg to AGA to gain *Bgl*II site.
67	Ile from ATC to ATT gives a second *BspE*I site.
73	Val GTT to GTA introduces *Sna*BI.
81	Pro compromise codon.
85	As above.
87	As above and introduces unique *Nco*I site.
96	Gln compromise codon.
98	Gly GGT to GGC to gain *Stu*I site.
102	Lys AAA to AAG and
103	Leu to CTT produces a *Hind*III site in ETI.
110	Gln CAA compromise.
114	Pro CCA compromise.
119	As for 110.
122	Asp GAC to GAT to gain *Bcl*I site.
123	As for 110.
130	Leu CTG to compromise CTT.
137	Glu GAA to GAG to break string of A's.
138	Lys AAA to AAG breaks string of A's.
140	Ala GCT to GCA and
141	Ser TCT to TCG makes *Cla*I site.
161	Pro proceeds rare codon for Leu at 162 thus *E.coli* preferred codon CCG used.
162	Leu from CTG compromise to TTA gains unique *Hpa*I site.
167	Lys AAA to AAG breaks string of A's.
168	As above
170	Glu GAA to GAG and
171	Ser TCT to AGC introduces *Sac*I site.

```
      EcoRI      XbaI                                  KpnI
      GAATTCATGCTTCTAGACGGTAACGGTGAAGTTGTTCAGAACGGTGGTACCTACTACCTG
   1  ---------+---------+---------+---------+---------+---------+ 60

                   ApaI                  NheI
      CTGCCACAGGTTTGGGCCCAAGGTGGTGGTGTTCAGCTAGCTAAAACCGGTGAAGAAACC
                ^
  61  ---------+---------+---------+---------+---------+---------+ 120
                1

                                                             BspEI
      TGCCCACTGACCGTTGTTCAATCTCCAAACGAACTGTCTGACGGTAAACCAATCCGGATC
                      ^
 121  ---------+---------+---------+---------+---------+---------+ 180
                      2

                   BglII      BspEI           SnaBI
      GAATCTCGTCTGAGATCTGCTTTCATTCCGGACGACGACAAAGTACGTATCGGTTTCGCT
                             ^
 181  ---------+---------+---------+---------+---------+---------+ 240
                             3

                                NcoI                          StuI
      TACGCTCCAAAATGCGCTCCATCTCCATGGTGGACCGTTGTTGAAGACGAACAAGAAGGC
                                  ^
 241  ---------+---------+---------+---------+---------+---------+ 300
                                  4

              HindIII
      CTGTCTGTTAAGCTTTCTGAAGACGAATCTACCCAATTCGACTACCCATTCAAATTCGAA
                                        ^
 301  ---------+---------+---------+---------+---------+---------+ 360
                                        5

             BclI
      CAAGTTTCTGATCAACTGCACTCTTACAAACTGCTTTACTGCGAAGGTAAACACGAGAAG
                                               ^
 361  ---------+---------+---------+---------+---------+---------+ 420
                                               6

         ClaI                       BstEII
      TGCGCATCGATCGGTATCAACCGTGACCAGAAAGGTTACCGTCGTCTGGTTGTTACCGAA
                                                      ^
 421  ---------+---------+---------+---------+---------+---------+ 480
                                                      7

             HpaI                        SacI          BamHI
      GACTACCCGTTAACCGTTGTTCTGAAGAAGGACGAGAGCTCTTAATGCAGGGATCC
 481  ---------+---------+---------+---------+---------+------^536
                                                              8
```

Figure 2.2. DNA sequence for synthetic ETI gene showing the restriction sites unique within the context of the synthesis vector pSyn.1.

Numbered arrows indicate the division points for construction of oligos for the synthesis of the entire gene.

There were some deviations from the intended strategy discussed above. The natural Val NH_2 terminus of ETI was replaced with Met because *E.coli* requires an initiator codon to precede the gene coding sequence. In most *E.coli* genes sequenced (91%), this codon is AUG, the codon for Met. A further 8% utilize GUG, a minor codon for Val, while only 1% have been recorded that use UUG in this position (Stormo, 1986). An initiator AUG codon in *E.coli* elicits the insertion of a formyl methionine (f-Met) at this position.

After translation, f-Met peptides are deformylated and the N-terminal Met may be removed depending on the nature of the second amino acid. Ala, Ser, Gly, Pro, Thr or Val as second amino acid permit N-terminal Met to be removed, whereas Arg, Asn, Asp, Gln, Glu, Ile, Leu, Lys and Met do not (Tsunasawa *et al.*, 1985). However, Marston (1986) has found that methionyl-aminopeptidase, the N-terminal Met removal enzyme of *E.coli*, functions poorly with recombinant proteins regardless of the amino acid following the N-terminal Met.

Rather than rely on the N-terminal Met removal system of *E.coli*, it was decided to forfeit the Val and utilize a Met in this position. Met is a hydrophobic amino acid and should keep the N-terminal finger region of ETI intact and not interfere with the interaction of ETI with tPA and was considered an acceptable replacement for Val. Nonetheless, if necessary, the 5′ end of the gene can be altered to conform to other vectors or hosts by using the *Xba*I site that was placed near to the 5′ end of the gene.

The two restriction sites *Eco*RI and *Bam*HI were chosen to be the sites of excision of the synthetic gene as they are compatible for forced cloning into the expression vector of choice at this stage viz pBtacI from Boehringer Mannheim. Another set of changes made to the original sequence for ETI were made to break up a string of 5 and 6 adenosines respectively. This was done to avoid misreadings during bridge mutagenesis (the technique that was to be used for the synthesis of this gene). The termination codon TAA was placed directly after the coding sequence. Termination of protein synthesis can be achieved with two other termination codons viz TGA or TAG, recognized by *E.coli* release factors RF2 and RF1 respectively, both of which are rare in the cell relative to charged tRNA species. TAA was used in the ETI-encoding gene as it is recognized by both RF1 and RF2 and thus is more efficient a terminator than TGA or TAG (Maitra *et al.*, 1982; Andersson and Kurland, 1990). These positions of change are shown in Figure 2.1.

The completed sequence was examined by the Fold program of GCG which determines the most stable theoretical mRNA secondary structure. The constraints used were minimum stem of six base pairs and minimum loop size of three bases. The resultant structure is shown in Fig 2.3.

The structure presented in Fig 2.3 is one with extensive double stranded regions which produces a thermodynamically stable structure with a ΔG of about -143 KCal/mol. Such a structure should protect the mRNA from RNAse degradation.

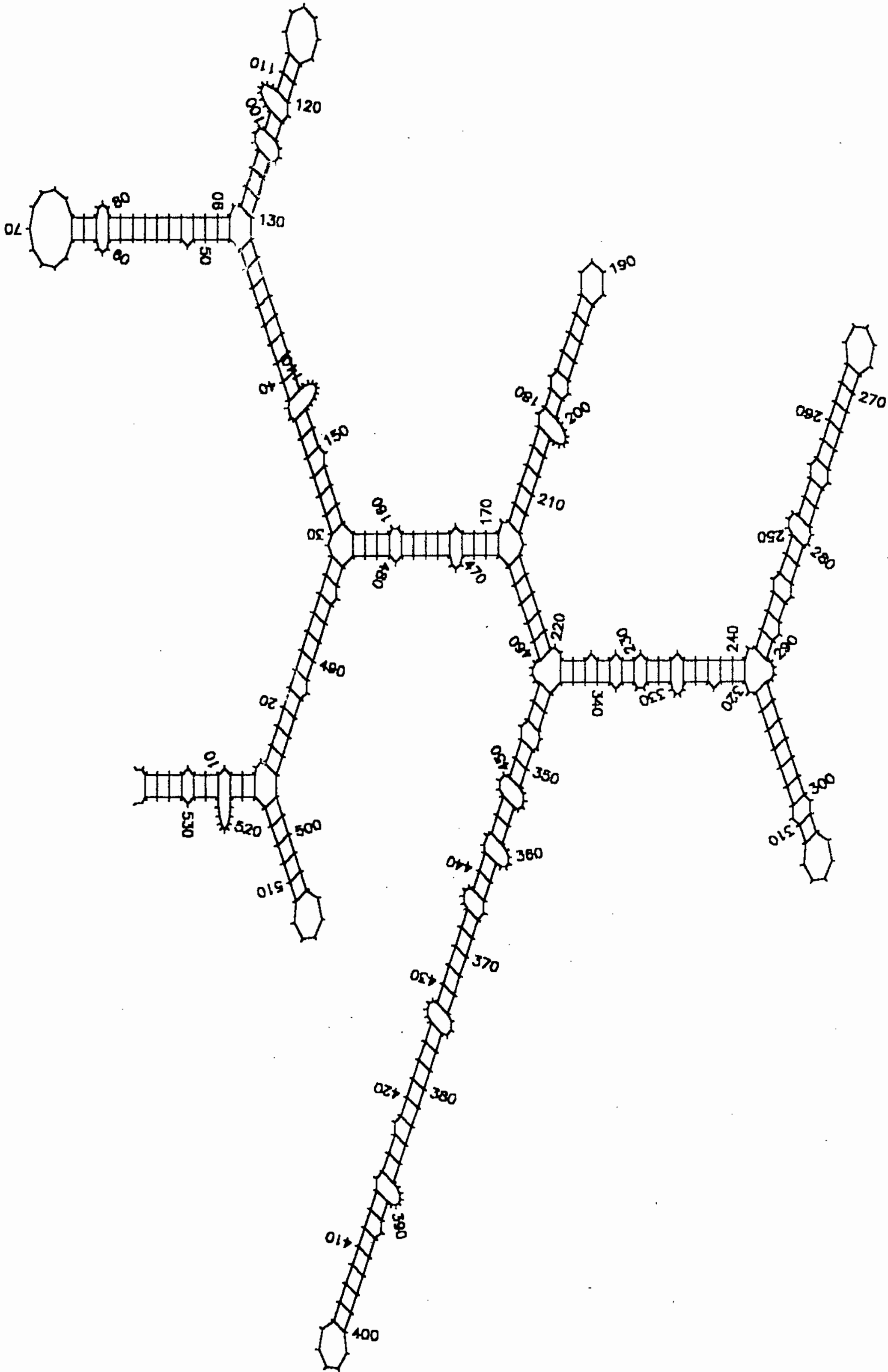


Figure 2.3 Most stable computer generated theoretical mRNA secondary structure for ETI as produced by the Fold program of the GCG program portfolio.

From the point of view of accessibility of a compact mRNA to the ribosomes, the work of Sorensen *et al.* (1989) on the *lacZ* gene has shown that theoretically stable mRNA structures do not necessarily cause translation delays. However, this may not be the case with the mRNA produced by the ETI encoding gene designed for this work. If this turned out to be the case, the option of placing a leader sequence upstream of the ETI encoding mRNA to enhance its translational efficiency would be used. This approach has been used with success by a number of workers eg Schoner *et al.*, (1984) and Sung *et al.*, (1986). Thus, the theoretical mRNA structure produced by the synthetic ETI gene appears to be satisfactory and no consideration was given to changing bases in order to alter it.

2.3 Alteration of pUC9 to produce the synthesis vector pSyn

The gene was to be synthesized by the serial cloning of oligonucleotides as described by Hayden and Mandecki (1988). The method requires the use of a synthesis vector containing a unique *Bsm*I site. The restriction enzyme *Bsm*I cuts one strand of DNA within its recognition site and the other strand just outside of the recognition site as shown below:

These workers inserted a *Bsm*I site into pUC9 by oligonucleotide-directed double-strand break-repair or bridge mutagenesis (Mandecki, 1986). This technique forms the basis of

the gene synthesis methodology of Hayden and Mandecki (1988) and makes use of alpha-complementation as an initial screen to select for cells containing plasmid into which DNA had been correctly inserted.

Alpha-complementation was described in 1967 by Ullman *et al.* They showed that deletion of the 3′ end of the *lac*Z gene gives rise to an inactive N-terminal fragment of β-galactosidase called the α-fragment. This deleted form of the *lac*Z gene is referred to as the *lac*Z′ or *lac*Zα gene and is incorporated in many plasmid vectors such as the pUC series of vectors. Conversely, deletions of the 5′ end of the *lac*Z gene expressed as Δ(*lac-proAB*), gives rise to an inactive C-terminal of β-galactosidase called the ω-fragment. This fragment is usually supplied by an F′ plasmid resident in the host *E.coli* cell (eg. JM105) and is controlled by the *lac*Z promoter and its operator. In cells such as JM105 the F′ plasmid also codes for Lac repressor which prevents expression of the *lac*Z ω-fragment. Cells of JM105, that contain a vector such as pUC9, grown in the presence of isopropyl-β-D-thiogalactosidase (IPTG) which derepresses the gene for the ω-fragment, are able to synthesize this fragment. When the two fragments, α- and ω- are present in a cell they complement, which can be detected by activity on the synthetic lactose analogue 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) which gives a blue reaction product (Ausubel *et al.*, 1987).

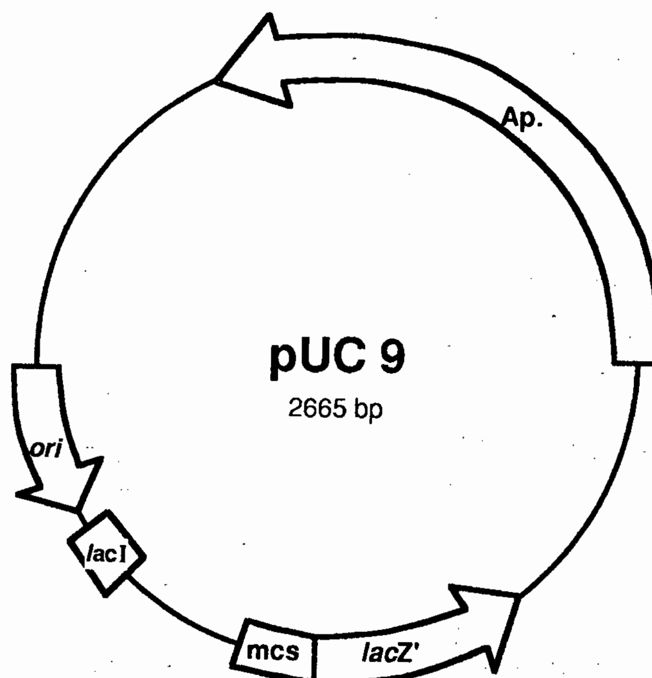
The *E.coli* strain JM83 [*ara*, *Δ(lac-proAB)*, *rpsL(=strA)*, φ80,*lac*ZΔM15], which was used for synthesizing the ETI-encoding gene, is also an ω-fragment producing strain. However,

it does not produce Lac repressor and as a result ω-fragment is produced continuously and pUC containing cells do not require derepression with IPTG to allow assembly of an active β-galactosidase complex (Viera and Messing, 1982).

Rather than apply to the authors Hayden and Mandecki (1988) for their synthesis plasmid, a similar plasmid was produced as part of this work. This exercise was used to gain experience in double-strand break-repair and to gauge the efficacy of the procedure.

To this end a 37 base (37-mer) oligonucleotide (oligo) was synthesized and purified as described in Chapter 6.7. Fifteen bases of the 5′ and 3′ ends of this oligo were designed to be complementary to the coding strand of pUC9 on either side of the *Sma*I cleavage site. This is shown in Figure 2.4. The methodology here requires only a few simple steps which are discussed in detail in Chapter 6. Briefly, the cleaved plasmid was incubated with the bridging oligo in a buffer that allows reannealing of the plasmid which was denatured by incubation in boiling water and then cooled to room temp. *E.coli* JM83 was then transformed with the bridged plasmid and plated on agar with the appropriate antibiotic.

a)



b)

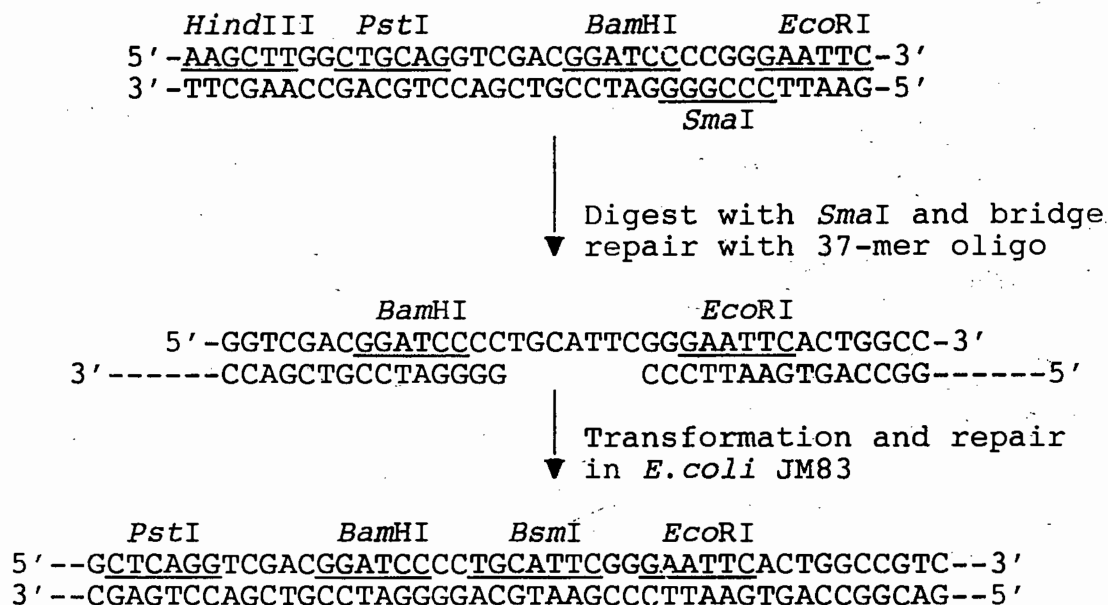


Figure 2.4 Strategy for insertion of *Bsm*I site into pUC9 to produce pSyn synthesis vector.

a) pUC9 and its multicloning site.

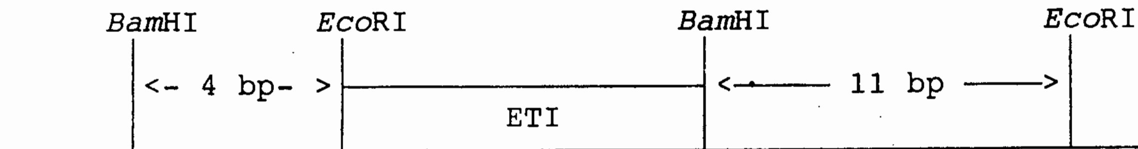
b) Production of synthesis vector. pUC9 is cut with *Sma*I oligo-directed break-repair performed with linear pUC9. This is transformed into JM83 and oligo-directed repair of the gap in pUC9 inserts the *Bsm*I site.

The resulting plates contained in excess of 300 colonies each, all of which were white, while the control JM83(pUC9) cells were all blue. The bridging of the gap created by *Sma*I digestion of pUC9 with the 37 mer oligo had given rise to an insertion of 7 bases (*Bsm*I site) into pUC9 ie an insertion of 2 X 3 bases (2 codon equivalents) + 1. Thus the *lac*Z' gene of pUC9 had been forced out of frame with its promoter by one base pair giving a nonsense β-galactosidase α-fragment unable to complement with the ω-fragment of β-galactosidase of the cloning host, *E.coli* JM83. This results in colonies with no β-galactosidase activity which remain white as they are unable to degrade the X-gal in the medium.

Ten such colonies taken from each of 3 separate platings were streaked to single colonies on Luria agar containing 100μg/ml of Ampicillin and 20μg/ml of X-gal ($LAA^{100}X$) plates. From 20 of these, minipreps of plasmid DNA were made and the resulting plasmid DNA tested for the presence of a *Bsm*I site. All were found to be linearized with *Bsm*I, while the control pUC9 was uncut by this enzyme. This indicated that these 20 plasmids each had at least one copy of the *Bsm*I site. To verify the results, the sequence across the multicloning site (mcs) of 2 of these plasmids was determined by Sanger dideoxy sequencing. Both were found to have a single copy of the *Bsm*I site in an otherwise intact pUC9 mcs. One of these was selected for further work and called pSyn.

The gene for ETI had been designed to be excised from the synthesis vector (pSyn) by *Eco*RI and *Bam*HI digestion. However, as can be seen from Fig 2.4 the mcs of pUC9 and hence pSyn

contains both an *Eco*RI and a *Bam*HI site and the final product of ETI in pSyn would have appeared as shown below:



It can be seen from the above sketch that on attempting to remove ETI from pSyn with *Eco*RI and *Bam*HI, the entire gene would be released, carrying pSyn sequences either 5′ or 3′ to itself depending on which enzyme was used first.

Forced cloning into pBtac1 (ie *Eco*RI-*Bam*HI fragment into an *Eco*RI and *Bam*HI digested vector) would require removal by digestion with *Eco*RI of 4 base pairs of DNA, from the 5′ end of gene when the initial digest had been with *Bam*HI. Similarly, the removal of 11 base pairs of pSyn DNA from the 3′ end of the gene with *Bam*HI would be required when *Eco*RI had been the initial cutter. However, the removal of such small pieces of DNA by digestion with either of these two enzymes is reportedly not reliable (New England Biolabs Catalogue 1990-1991).

To avoid this problem, the *Eco*RI site was removed from the mcs of pSyn. This was achieved by *Eco*RI digestion, removal of the resultant 4 base overhangs with Mung bean nuclease digestion followed by blunt end ligation to close the plasmid. The resultant plasmid, pSyn.1 had lost 4 bases ie 1 X 3 (or 1 codon equivalent) + 1. Thus, the inserted base which put the *lac*Z′ gene of pSyn out of frame was removed and the *lac*Z′ gene of pSyn.1 was put in frame with its promoter. Colonies of JM83

transformed with this plasmid were all blue when plated on LAA^{100}X plates. The method is outlined in Chapter 6 and shown in Figure 2.5 below.

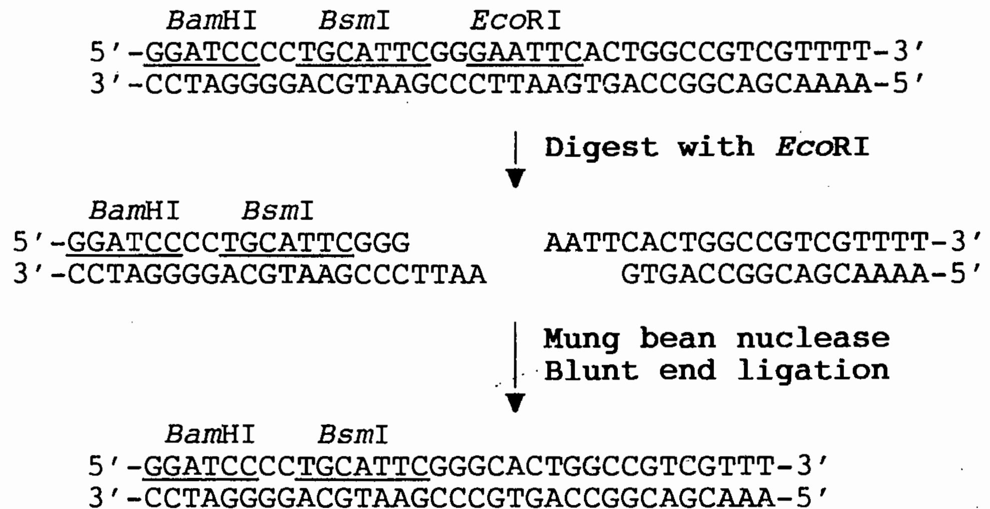


Figure 2.5 **Strategy for removal of the *Eco*RI site from pSyn the modified pUC9 synthesis vector**

2.4 Synthesis and purification of oligonucleotides for construction of the synthetic ETI gene

Synthesis of any large gene requires some form of sequential fragment assembly. This is certainly the case with gene synthesis by serial cloning (Hayden and Mandecki, 1988). With this methodology it is essential to supply 15-base hybridization arms at the 5′ and 3′ ends of the coding sequence to anchor each successive oligo fragment of the gene across the gap created by *Bsm*I digestion of the synthesis plasmid.

The sequence of ETI was divided into 8 portions (see Figure 2.2). The length of each portion was chosen to put the *lacZ'* gene of pSyn.1 alternately out of sequence then back into

sequence. Selection for the incorporation of each new insert would be based on change of colony colour from blue to white or vice versa. Eight oligos having the sequence of each of these portions were synthesized (see Figure 2.6). Each of these was designed with a 15-base arm at the 5′ and 3′ ends.

In oligo 1 the arms are homologous to pSyn.1 on either side of the *Bsm*I cleavage. All the oligos had the same 3′ arm for homology to pSyn.1 and for regeneration of the *Bsm*I site. Subsequent oligos had 5′ arms complementary to the coding strand produced by the 3′ end of the preceding insert (see Figure 2.7).

Full length oligos free of truncated sequences (also known as failure sequences) is a requirement for successful gene synthesis. Truncated sequences arise in even the most efficient of synthesis strategies as a result of incomplete coupling of nucleoside 3′-phosphoramidites with the controlled pore glass (CPG) supported 5′-hydroxyl DNA (Horn and Urdea, 1988). These authors recommend purifying oligos on phenyl derivatized silica syringe columns to separate the hydrophobic 5′-dimethoxytrityl (DMT) protected full length product from unprotected failure products.

In addition, shorter fragments can also arise during removal of the oligo from the CPG by treatment with concentrated NH_4OH which cleaves the DNA at any apurinic sites (Horn and Urdea, 1988). These sites arise during synthesis by random acid hydrolysis of purine bases without affecting the phosphodiester

OLIGO-1

5'-**GGT CGA CGG ATC CCC** GAA TTC ATG CTT CTA GAC GGT AAC GGT GAA GTT GTT
CAG AAC GGT GGT ACC TAC TAC CTG CTG CCA CAG G **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGO-2

5'-**ACC TGC TGC CAC AGG** TTT GGG CCC AAG GTG GTG GTG TTC AGC TAG CTA AAA
CCG GTG AAG AAA CCT GCC CAC TGA CCG TT **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGI-3

5'-**TGC CCA CTG ACC GTT** GTT CAA TCT CCA AAC GAA CTG TCT GAC GGT AAA CCA
ATC CGG ATC GAA TCT CGT CTG AGA TCT GCT T **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGO-4

5'-**GTC TGA GAT CTG CTT** TCA TTC CGG ACG ACG ACA AAG TAC GTA TCG GTT TCG
CTT ACG CTC CAA AAT GCG CTC CAT CTC CA **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGO-5

5'-**TGC GCT CCA TCT CCA** TGG TGG ACC GTT GTT GAA GAC GAA CAA GAA GGC CTG
TCT GTT AAG CTT TCT GAA GAC GAA TCT ACC C **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGO-6

5'-**AAG ACG AAT CTA CCC** AAT TCG ACT ACC CAT TCA AAT TCG AAC AAG TTT CTG
ATC AAC TGC ACT CTT ACA AAC TGC TTT AC **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGO-7

5'-**TAC AAA CTG CTT TAC** TGC GAA GGT AAA CAC GAG AAG TGC GCA TCG ATC GGT
ATC AAC CGT GAC CAG AAA GGT TAC CGT CGT C **TGC ATT CGG GCA CTG**-3'
*Bsm*I

OLIGI-8

5'-**AAG GTT ACC GTC GTC** TGG TTG TTA CCG AAG ACT ACC CGT TAA CCG TTG TTC
TGA AGA AGG ACG AGA GCT CTT AAT GCA GGG ATC C **TGC ATT CGG GCA CTG**-3'
*Bsm*I

Figure 2.6. Sequences of the oligonucleotides required for the synthesis of ETI.
The 15 base 5′ and 3′ anchoring arms are shown in bold.

*Bsm*I
5'-GATCCCCTGCATTCGGGCACT-3'
3'-CTAGGGGACGTAAGCCCGTGA-5'

BsmI digestion

5'-GATCCCCTG CATTCGGGCACT-3'
3'-CTAGGGG ACGTAAGCCCGTGA-5'
pSyn.1 coding strand

Double strand break repair with oligo 1 a 70 base oligo fragment that encodes the EcoRI site and the amino terminal region of ETI. Followed by transformation to allow oligo directed gap repair in JM83.

5'-GATCCC———TGCATTCGGGCACT-3'
3'-CTAGGG———ACGTAAGCCCGTGA-5'
Insert 1

BsmI digestion and double strand break repair with oligo 2 a 65 base oligo encoding the next portion of ETI

complementary arm to insert 1
5-'GATCCC——— ———TGCATTCGGGCACT-3'
3'-CTAGGG——— ACGTAAGCCCGTGA-5'

Transformation, oligo directed gap repair in JM83

5'-GATCCC———TGCATTCGGGCACT-3'
3'-CTAGGG———ACGTAAGCCCGTGA-5'
Insert 1 Insert 2

Figure 2.7 The strategy for serial cloning of oligonucleotides as described by Hayden and Mandecki (1988)

bonds of the backbone. The 5′ end of the cleaved product carries a 5′-DMT and is purified along with 5′-protected full length product.

However, apurinic sites are susceptible to mildly alkaline solutions of amines and were cleaved in the presence of 1M lysine pH9 (without removal of other oligos from the CPG) and the soluble 5′-DMT truncated fragments were washed away. The remaining 3′-bound material was removed from the support with NH_4OH. As only full length oligos are protected with 5′-DMT, these are readily purified away from the rest of the mix on passage through a rapid reverse-phase cartridge containing phenyl derivatized silica.

2.5 Synthesis of the ETI encoding gene

The gene synthesis protocol of Hayden and Mandecki (1988), as discussed in Chapter 2.3 and outlined in Figure 2.7, was followed. Prior to each serial addition of oligonucleotide to extend the ETI sequence, the plasmid was linearized by *Bsm*I digestion. After each round of bridge mutagenesis up to 20 colonies of the correct colour, were streaked to single colonies on $LAA^{100}X$ plates. This was done to ensure their stability, by monitoring their maintenance of correct colour on X-gal containing medium, and to separate them from satellite colonies which invariably form around antibiotic resistant colonies.

Of the apparently stable colonies, 12 were chosen for plasmid minipreps and digested with the appropriate restriction enzyme

that had been introduced on that particular insert. Four plasmids that were shown to contain the new site were extracted on large scale by plasmid maxipreps, purified twice by CsCl ultracentrifugation and sequenced by Sanger dideoxy sequencing. After each successful insertion of successive ETI portions the plasmid containing the correct sequence was termed pLT1 to 8 depending on the number of the oligo most recently inserted. The results of each successive round of insertion are summarized in Table 2.2.

From the insertion of oligo 3 onwards the development of blue colouration of colonies required about 48 hours to develop (16 hrs at 37°C, approximately 5 hrs at room temp. and the remainder at 4°C), whereas for oligo 1 and 2 this required only about 16 hrs at 37°C. Furthermore, the blue that did develop was very pale and difficult to discern. Nonetheless, a selection was made at this stage and those colonies streaked on LAA^{100}X as described earlier. At a second round of growth, easily discernible blue colour developed after 12-16 hours. It is possible that the initial slow development of colour was due to the repair process.

ETI plasmid (a)	Insertion length (bases) (b)	Colour selection (c)	Total colonies (d)	% colour/ enzyme site (e)/(f)	Correct sequence (g)
pLT1	70	white	100	38/12	4
pLT2	65	blue	117	23/10	4
pLT3	67	white	125	25/11	4
pLT4	65	blue	130	30/9	3
pLT5	67	white	152	32/8	2
pLT6	65	blue	103	20/6	4
pLT7	67	white	119	22/7	4
pLT8	70	blue	99	21/9	4

Table 2.2 Shows summary of results for each successive insertion of oligonucleotides in the synthesis of an ETI encoding gene.

(a) pSyn.1 with the number of the most recently added insert.

(b) Number of bases inserted across the gap created by *Bsm*I digestion.

(c) Selected colour for the number of bases added.

(d) Total number of colonies produced from 3 plasmid:oligo bridgings of 1:10, 1:100 and 1:1000 on 3 separate plates.

(e) Percentage of the total number of colonies that on initial selection were of the correct colour.

(f) Of the 12 colonies of correct colour that were chosen for minipreps, the number with the restriction site of the most recent insert.

(g) The number out of 4 with the correct restriction site that had the correct sequence.

It can be seen from Table 2.2, that from 20-38% of total colonies at each insert produced the correct colour. Of those selected for minipreps not all could be digested by the newly inserted restriction site. This was thought to be as a result of incorrect repair of the gap created by the bridging oligo. In the case of pLT4, one out of 4 sequenced plasmids and in pLT5, two out of 4 sequenced
plasmids, the unique restriction sites contributed by these oligos were left intact while error was introduced elsewhere in the new sequence.

Upon construction of the entire ETI encoding gene in pSyn.1, plasmid pLT8 was purified as before and both strands of the synthetic ETI gene were sequenced using 4 primers (see Figure 2.8). The entire sequence of this gene was found to be free of any alterations from the designed sequence.

```
TGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGGCTGCAGGTCGACGGATCCCC
---------+---------+---------+---------+---------+---------+---------+---------+
ACACTCGCCTATTGTTAAAGTGTGTCCTTTGTCGATACTGGTACTAATGCGGTTCGAACCGACGTCCAGCTGCCTAGGGG
                5'-TCACACAGGAAACAGC-3'  REV 01 --->

EcoRI
GAATTCATGCTTCTAGACGGTAACGGTGAAGTTGTTCAGAACGGTGGTACCTACTACCTGCTGCCACAGGTTTGGGCCCA
---------+---------+---------+---------+---------+---------+---------+---------+
CTTAAGTACGAAGATCTGCCATTGCCACTTCAACAAGTCTTGCCACCATGGATGATGGACGACGGTGTCCAAACCCGGGT

AGGTGGTGGTGTTCAGCTAGCTAAAACCGGTGAAGAAACCTGCCCACTGACCGTTGTTCAATCTCCAAACGAACTGTCTG
---------+---------+---------+---------+---------+---------+---------+---------+
TCCACCACCACAAGTCGATCGATTTTGGCCACTTCTTTGGACGGGTGACTGGCAACAAGTTAGAGGTTTGCTTGACAGAC

ACGGTAAACCAATCCGGATCGAATCTCGTCTGAGATCTGCTTTCATTCCGGACGACGACAAAGTACGTATCGGTTTCGCT
---------+---------+---------+---------+---------+---------+---------+---------+
TGCCATTTGGTTAGGCCTAGCTTAGAGCAGACTCTAGACGAAAGTAAGGCCTGCTGCTGTTTCATGCATAGCCAAAGCGA
                                                  ---> REV 02  5'-AAGTACGTATCGGTTTC-3'

                                              <--- FOW 02  3'-CGGACAGACAATTCGA-5'
TACGCTCCAAAATGCGCTCCATCTCCATGGTGGACCGTTGTTGAAGACGAACAAGAAGGCCTGTCTGTTAAGCTTTCTGA
---------+---------+---------+---------+---------+---------+---------+---------+
ATGCGAGGTTTTACGCGAGGTAGAGGTACCACCTGGCAACAACTTCTGCTTGTTCTTCCGGACAGACAATTCGAAAGACT

AGACGAATCTACCCAATTCGACTACCCATTCAAATTCGAACAAGTTTCTGATCAACTGCACTCTTACAAACTGCTTTACT
---------+---------+---------+---------+---------+---------+---------+---------+
TCTGCTTAGATGGGTTAAGCTGATGGGTAAGTTTAAGCTTGTTCAAAGACTAGTTGACGTGAGAATGTTTGACGAAATGA

GCGAAGGTAAACACGAGAAGTGCGCATCGATCGGTATCAACCGTGACCAGAAAGGTTACCGTCGTCTGGTTGTTACCGAA
---------+---------+---------+---------+---------+---------+---------+---------+
CGCTTCCATTTGTGCTCTTCACGCGTAGCTAGCCATAGTTGGCACTGGTCTTTCCAATGGCAGCAGACCAACAATGGCTT

                                                       BamHI
GACTACCCGTTAACCGTTGTTCTGAAGAAGGACGAGAGCTCTTAATGCAGGGATCCTGCATTCGGGCACTGGCCGTCGTT
---------+---------+---------+---------+---------+---------+---------+---------+
CTGATGGGCAATTGGCAACAAGACTTCTTCCTGCTCTCGAGAATTACGTCCCTAGGACGTAAGCCCGTGACCGGCAGCAA

<---FOW 01  3'-ACCCTTTTGGGACCGC-5'
TTACAACGTCGTGACTGGGAAAACCCTGGCGGTTACCCAA
---------+---------+---------+---------+
AATGTTGCAGCACTGACCCTTTTGGGACCGCCAATGGGTT
```

Figure 2.8 The sequence of pLT8 showing the 4 oligonucleotide primers used to sequence the synthetic ETI gene.

pSyn.1 sequence is shown in bold and arrows show direction of sequence.

2.6 Discussion

The strategy adopted for the design of a gene encoding ETI will be summarized as a set of general guidelines for future gene designs. The amino acid sequence was initially translated strictly into major codons for *E.coli*. To make the sequence acceptable for possible expression in *S.cerevisiae*, 2 codons (for Pro and Gln) of the sequence were changed such that they would still be acceptable to *E.coli*. The sequence was scanned for restriction sites and partial restriction sites using a large database of enzymes. Restriction sites unique to the ETI/pSyn.1 vector were introduced into the sequence by change of the wobble base of that codon but without compromising the ability of *E.coli* to utilize the sequence. Two other changes were made to the *E.coli* major codon sequence in order to break up two strings of A's but again without compromising the intended expression host. A very useful ETI-encoding sequence was designed that can be used for investigations beyond this work into the inhibitory mechanisms of ETI. The design has made allowance for alterations to the amino terminus and the ease by which this region is altered will be demonstrated in Chapter 3. Furthermore, the carboxy terminus can be altered in the same manner and the coding sequence for the amino acids at and surrounding the scissile bond can also be altered in the same manner with relative ease.

The decision was made to synthesize the gene by the bridge mutagenesis method for three important reasons viz:

a) The reduction in cost, as only 1.5 times the length of the gene was needed in synthetic oligo (taking into account the 5' and 3' 15 base arms required on each oligo), as

opposed to at least double the length of the gene as is required in a double strand synthesis approach. The 0.5 times gene length is contributed by the required arms for anchoring each oligo across the gap created by digestion with *Bsm*I. This amount can be reduced somewhat by increasing the distance between the arms ie., by increasing the size of each oligo.

b) The accuracy of the repair of the gap in *E.coli*, which as can be seen in Table 2.2 is reasonably high.

c) The relative ease of the synthesis approach which, after oligo synthesis, requires no further chemical manipulations for insertion of each new gene segment.

From Table 2.2 it can be seen that only between 20 and 38% of the transformants that were obtained after each bridging step were of the required colour. However, it was not established how much of this was contributed by either uncut or reclosed cut plasmid. Thus the accuracy of the procedure cannot be judged from the results obtained at this step. Of a selection from the correct colour of transformants, from 50 to 100% were found to carry the restriction site coded for on the newly inserted oligos. Thereafter, of a sample taken from those with the required restriction site, in only two cases, viz pLT4 and 5, 1 out of 4 and 2 out of 4 respectively, did not have the correct sequence coded for by the latest insert.

Technically, bridge mutagenesis is very simple. However, gene synthesis by serial cloning using bridge mutagenesis is time consuming because of the cumbersome screening procedure that was used for this work. For synthesis and final sequencing of

the completed gene required 8 months of normal working days by this worker. The time required to complete a similar project would be much reduced as the technique has now been established and the laboratory well equipped for the work.

To further expedite a future synthesis project it has been suggested by Mandecki (pers. comm.) that the method of gene synthesis by the *Fok*I method would be preferable to serial cloning. This method was described by Mandecki and Bolling (1988) and later used by Mandecki *et al.* (1990) to synthesize an entire plasmid of some 2050 base pairs as was discussed in the introduction to this chapter. However, prior to the start of this work some experiments were done to determine the suitability of the *Fok*I method for this current work. It was determined at this stage that the enzyme *Fok*I produced unreliable cleavage in the hands of this worker and the decision was made to synthesize by serial cloning. However as the *Fok*I method is highly recommended by Mandecki and bearing in mind their spectacular success in synthesis of an entire plasmid, the technique should be perfected and established for use in this laboratory.

In addition, it was also suggested that to expedite screening, post-bridging transformants should be screened with a sample of end-labeled oligo most recently used for bridging. This method was used in parallel with the successful screening approach described earlier for selecting transformants after bridging with oligos 4, 5 and 6. However, it offered no saving of time and no improvement in accuracy to the first screening step.

Thus, in summary, the method of gene synthesis by serial cloning using bridge mutagenesis as described by Hayden and Mandecki (1988) was used with great success in the synthesis of a gene encoding ETI.

CHAPTER 3

EXPRESSION OF THE SYNTHETIC ETI GENE IN *E.coli*

3.0 Contents

3.1 Introduction 88

3.1.1 E.coli as a vehicle for expression of cloned genes 89

3.1.2 Increasing the stability of foreign proteins in E.coli 90

3.1.2.1 lon Mutations 91

3.1.2.2 *htpR* Mutations 92

3.1.2.3 Other protease mutants 93

3.1.3 General characteristics of expression plasmids 94

3.1.3.1 Plasmid replication 94

3.1.3.2 Gene expression at the transcription level 95

3.1.3.2.1 Transcription termination 99

3.1.3.2.2 Transcription control 100

3.1.4 Efficiency of translation in E.coli 104

3.1.5 Protein translocation in *E.coli* 105

3.1.6 Cytoplasmic expression of proteins in *E.coli* 109

3.2 Expression in pbtac1TM vector 113

3.2.1 The pBtac1TM expression vector 113

3.2.2 Construction of pTETI expression vector 115

3.2.3 Expression of ETI from pTETI 116

3.3 Expression from pMALTM-c expression vector 119

3.3.1 The pMAL expression vector 119

3.3.2 Alteration of the N-terminus of synthetic ETI 124
3.3.3 Construction of pMetI the pMAL-c/ETI fusion 125
3.3.4 Expression of the fusion protein coded for by pMetI 126
3.3.5 Discussion 142
3.4 Expression of two forms of synthetic ETI in the pET expression system 145
3.4.1 The pET expression system 145
3.4.1.1 The pET expression vectors 147
3.4.1.2 Hosts for expression from pET vectors 149
3.4.1.3 Proposed reconstruction of the synthetic gene for ETI 151
3.4.2 Reconstruction of the 5′ end of the synthetic ETI gene 153
3.4.3 Construction of pET 12a/ETI fusions 154
3.4.4 Determination of the appropriate strain for expression 155
3.4.5 Expression, cleavage and purification of ETI from pPETI.1 and pPETI.2 158
3.4.6 Discussion 167

3.1 Introduction

It was shown in Chapter 2 that the gene designed for ETI was accurately synthesized. Furthermore, it appears to be stably maintained in the synthesis vector pSyn.1 and the host JM83 and

thus is amenable to be excised and placed in a suitable vector/host system for expression or, more desirably, for overexpression.

3.1.1 *E.coli* as a vehicle for expression of cloned genes

The decision had been taken earlier to use an *E.coli*-based expression system to produce ETI from this synthetic gene. The advantage of this organism is the well established technology for DNA manipulation and cell culture, as compared with that of other expression systems. In addition, there is a large body of knowledge on the genetics and physiology of *E.coli* which can be drawn on to address difficulties in the expression of a particular gene. In support of this, there are many genes both cloned and synthetic that have been successfully expressed in *E.coli* (Marston, 1986).

Nonetheless, *E.coli* is not suitable for expression of all genes. The most important drawbacks to the use of this organism for production of proteins are its limited capacity to secrete proteins and its inability to perform certain post-translational modifications of proteins such as acetylation and glycosylation (Balbas and Bolivar, 1990).

ETI requires neither glycosylation nor acetylation but does require two disulphide bridges to stabilize the correct structure for activity. In the reducing environment of the *E.coli* cytoplasm these disulphide bonds are unlikely to form (Goeddel, 1990). However, *E.coli* expression systems are versatile and the problem may be overcome by exporting the

protein to an oxidizing environment, either the periplasm or the extracellular medium (Kohno *et al.*, 1990). This strategy has been very successful for the expression of a number of proteins (Marston, 1986) and will be discussed later. Alternatively, Kohno *et al.* (1990) indicate that disulphide bonds can be formed accurately and in high yield *in vitro* by supplying the *in vivo* redox state of the natural environment of the native protein. In doing so, they indicate, the primary amino acid sequence will direct the normal folding pathway of the protein.

Despite the high degree of control that current expression technology allows over *E.coli* expression systems, in many instances, success is achieved empirically and the practice is described by Stader and Silhavy (1990) as an art. This must certainly reflect how much is still to be learned about gene expression in *E.coli*.

The choices of expression vector and host and combinations of these for *E.coli* are numerous. Thus, before attempting expression it is appropriate to review briefly, current views on the requirements for overexpression of recombinant proteins in *E.coli*.

3.1.2 Increasing the stability of foreign proteins in *E.coli*

Many proteins, both foreign and native are rapidly degraded by proteolytic enzymes when expressed in *E.coli*. Proteolysis in bacteria is a very selective and carefully regulated process aimed at recycling amino acids and preventing accumulation of

nonfunctional or denatured polypeptides that could be toxic to the cells (Goldberg and Goff, 1986).

The reason for the instability of particular recombinant proteins is not well understood. Nonetheless, it appears that a single degradation system is involved in the breakdown of proteins in *E.coli.* Thus one approach to stabilization of such proteins has been the use of strains carrying one or more mutations in the genes that control the degradation system to decrease intracellular proteolysis (Miller, 1987). However, no one protease mutant will stabilize all foreign proteins. Therefore, when it has been established that proteolysis is the cause of poor protein yield, the correct strain for expression must be established empirically.

Strains of *E.coli* with mutations in the *lon* and *htpR* genes are known to lower levels of proteolytic activity and have been used successfully for expression of proteins (Miller, 1987).

3.1.2.1 *lon* Mutations

Lon is a protease that seems to be active against unfolded or misfolded proteins. Strains with *lon* mutations degrade such proteins more slowly than their lon^+ parents (Gottesman, 1990).

Both deletion and insertion mutations in *lon* are available and can be introduced into other strain backgrounds by P1 transduction (Gottesman, 1990). However, *lon* mutations produce two major phenotypes viz mucoidy and UV sensitivity, which may make them unsuitable as expression hosts. These phenotypes can

be eliminated by appropriate secondary mutations (Gottesman, 1989). In addition, *lon* mutations can affect transformation efficiency and plasmid stability and may not be suitable for selection of transformants after ligation (Goldberg and Goff, 1986).

Mucoidy phenotype is seen when *lon* mutants are grown at temperatures of 34°C and below and on minimal media, but serious problems of mucoidy can be avoided when these strains are grown on rich media or at temperatures of 37°C or above. Genetic solutions to prevent capsule synthesis by the mucoidy phenotype include mutations in the *gal* operon, *cps* gene cluster or *rcs*B and *rcs*A genes (Gottesman, 1990).

In response to UV damage, *lon* mutants form long nonseptated filaments, which are frequently not viable. In addition, *lon* mutants tend to show some filamentation on rich growth media. Filamentation is lethal when cells are transferred from minimal to rich media. Filamentation is due to the stabilization of the SOS inducible cell division inhibitor sulA. Generally, this phenotype is not troublesome enough to require remedies. However, mutations inactivating *sulA* have no deleterious effects on cell growth and should eliminate the filamentation phenotype (Gottesman, 1990).

3.1.2.2 *htpR* Mutations

The *htpR* gene codes for σ^{32}, the subunit that confers specificity on RNA polymerase for gene promoters required for the heat shock response (Grossman et al.,1984). As lon itself

is a heat shock protein, strains deficient in this factor are generally defective in proteolysis. Strains with *lon* and *htpR* mutations appear to stabilize foreign proteins better than either one alone (Miller, 1987).

However, the most commonly used *htpR* allele is an amber mutation in the gene, used in strains that provide conditional suppression of the mutation, for example, *sup* temperature sensitive hosts. Such hosts allow for good growth at 32°C, poor growth at 37°C and no growth above this temperature (Gottesman, 1990).

3.1.2.3 Other protease mutants

Mutations in another heat shock gene *dnaJ* have been used for stabilizing foreign proteins in *E.coli* (Gottesman, 1990). In addition, *clp* mutations enhance stability of proteins in strains with defective lon protease (Goldberg and Goff, 1986).

Proteases located in the periplasm may degrade foreign proteins that have been exported to this region. In particular, mutations in *ompT* inactivate an outer membrane-localized protease that has been implicated in protein degradation in extracts from *E.coli* (Sedgwick, 1989; Grodberg and Dunn, 1988).

Mutations in *degP* inactivate a periplasmic protease which can cleave phoA fusions to release the phosphatase portion. This protease may degrade periplasmic proteins and thus inactivating

it may help to stabilize some exported proteins (Strauch *et al.*, 1989).

3.1.3 General characteristics of expression plasmids

The minimal elements that are required of an expression plasmid vector are as follows:

1) Well characterized origin of replication.
2) A gene encoding a selection marker for plasmid propagation and maintenance.
3) A multicloning region into which the gene of interest can be cloned under control of a strong regulatable promoter and in frame with the coding sequence for a strong ribosome binding site.

3.1.3.1 Plasmid replication

The primary function of naturally occurring plasmids is self replication. The region on a plasmid that controls this function as well as the plasmid copy number (number of plasmids per cell) is found at the origin of replication. Relaxed replicating plasmids, such as pBR322 and its derivatives have been preferentially used for expression systems (Balbas and Bolivar, 1990).

Of particular importance to the expression of ETI is that in relaxed plasmids, copy number is inversely proportional to the growth rate. Thus at high growth rates the number of plasmids per cell is low. This, coupled with defective partitioning of plasmid DNA between cells during cell division, invariably

leads to segregational instability (Klotsky and Schwartz, 1987). The plasmid can be gradually lost from the growing population by segregational instability unless a gene for selection, usually an antibiotic resistance gene, is carried by the plasmid. When the concentration of the selective antibiotic in the growth medium is reduced by the action of the resistance gene product, sensitive cells (those with few or no copies of the plasmid) are allowed to grow. Such plasmid-deficient cells may outcompete high-dose plasmid-containing cells as the latter suffer 'metabolic drag' induced by diversion of metabolic energy to their plasmids. Multiple additions of antibiotic to cultures appears to be the only solution currently available (Balbas and Bolivar, 1990).

A more insidious form of plasmid instability is that of structural instability resulting from some physical change (eg. deletion, insertion or rearrangement), usually to the DNA of interest, while the remainder of the plasmid including the selectable gene is unchanged. No method is yet available for accurate prediction of such instability. Nonetheless, the problem can be controlled by alternative strategies for gene cloning or by use of $RecA^-$ strains (Balbas and Bolivar, 1990).

3.1.3.2 Gene expression at the transcription level

As discussed in Chapter 2, for overexpression of a gene, the mRNA derived from it must become the major mRNA species in the cell. To achieve this, the gene must be placed under the control of a strong promoter.

Gene transcription or mRNA synthesis is catalyzed by RNA polymerase. The core enzyme of this polymerase, when combined with a dissociable subunit called sigma (σ), which confers specificity to the interaction between RNA polymerase and binding sites on the DNA called promoters, is known as a holoenzyme (Losick and Pero, 1981). *E.coli* has multiple forms of RNA polymerase which differ only in their σ subunits (Grossman *et al.*, 1984). Each holoenzyme form transcribes a family of genes controlled by its cognate promoter. Recombinant genes in *E.coli* are usually placed under the control of promoters recognized by holoenzyme $E\sigma^{70}$. In general the core promoter is that DNA region that is critically important for complex formation with holoenzyme in the absence of other factors (Gralla, 1990).

The core DNA promoter elements for $E\sigma^{70}$ lie in an approximately 29 base pair region between positions -7 and -36 with respect to the start point of transcription. The region is bounded by two hexanucleotide consensus sequences near position -10, viz 5′-TATAAT-3′ and position -35, viz 5′-TTGACA-3′. These two sequences are separated by a spacer region of 17 $\pm$ 1 base pairs. Polymerases containing other σ factors recognize other promoter sequences (McClure, 1985). These core DNA promoter elements were identified by both functional and statistical analyses. Most importantly, they have been extensively mutagenized and the damaged elements were shown to be associated with defective transcription (Gralla, 1990).

Transcription is a multistep process which can conveniently be divided into initiation, elongation and termination events.

Firstly, holoenzyme scans the DNA and encounters a promoter. The bases in the -35 and -10 regions form molecular contacts with the holoenzyme and may hold it long enough for a holoenzyme/DNA complex to form and DNA melting to occur. The strength of the promoter influences the stability of this complex and hence whether initiations will occur (Gralla, 1990). It is thought that the simultaneous binding of polymerase to the two conserved regions of core promoter induces conformational changes in the DNA. This leads to melting of the DNA in the promoter and formation of an open complex or transcription bubble.

In the final open complex the polymerase extends from position -45 to +20. Thus DNA sequences both upstream and downstream from the core promoter can have an important influence on expression, for example the upstream activator regions which are usually binding sites for regulatory proteins (Kuhnke et al., 1987). It has also been suggested that non core DNA sequences can influence promoter function by causing bending or allowing melting. *E.coli* σ^{70} promoters are in general quite rich in AT pairs and the sequences near -45 show some tendency toward being conserved (Hawley and McClure, 1983; Boroweic and Gralla, 1986).

Within the open complex RNA polymerase synthesizes a short primer strand which is paired with the template strand. If this template is allowed to grow to about 10 nucleotides, at which point it is strongly bound to the template, initiation is completed successfully. However, the enzyme frequently aborts initiation and the short primer cannot bind to the template.

The enzyme repeatedly produces and releases short initiation products while being held firmly at the promoter in a cycling pathway. This cycling slows down transcription initiation and can limit expression (Carpousis et al., 1982). The phenomenon is more pronounced with some promoters than with others, giving rise to the term slow start (eg. *lac*UV5 promoter) and rapid start (eg. T7 promoter) promoters. At slow start promoters RNA polymerase seems to have difficulty in escaping the forces that positioned it for initiation in the first place. In rapid start promoters the non transcribed regions of the promoter may assist in polymerase escape and promoter clearance (Gralla, 1990). In addition, there is some evidence to suggest that the initial transcribed region plays a role in setting the rate of polymerase escape and promoter clearance (Deuschle *et al.*, 1986; Carpousis *et al.*, 1982). Thus a cloned sequence itself may influence the level of expression from a particular promoter and hence the perceived strength of the promoter. As there are currently no rules for prediction of the influence a cloned gene will have on its promoter, the correct vector for expression must be determined empirically.

Once holoenzyme has synthesized a short, stable RNA initiation primer, σ factor is released and the core enzyme, which is no longer complexed to the promoter, elongates the RNA chain. The properties of elongating core polymerase can be altered by factors that associate with it. For example, NusA protein binds to core RNA polymerase, preventing σ from binding and may, to some extent, alter patterns of pausing and termination (Gralla, 1990).

3.1.3.2.1 Transcription termination

Elongation of mRNA continues until transcription is terminated, at which point, the mRNA/RNA polymerase/DNA complex dissociates. In bacteria, termination is achieved either by the action of a protein termination factor called rho or by factor independent methods (Balbas and Bolivar, 1990). Both methods utilize a site on the DNA which causes pausing of the elongation complex at the end of a coding sequence to achieve termination. Factor independent termination is usually used in expression systems and makes use of two GC rich segments that in the transcript have the potential to form a stem-loop structure. Following this is a downstream run of 4-8 A residues. It is believed that the GC rich segments cause pausing of polymerase. The nascent transcript forms a hairpin structure at this region displacing it from the template. Transcribing the A rich segment then completes the destabilization of the transcription complex and termination is achieved.

These termination regulatory elements are not indispensable in an expression plasmid vector but their presence at the end of a highly expressed gene is advantageous as transcription terminators act as barriers to elongation. Thus they minimize sequestering of RNA polymerase that may be engaged in unnecessary transcription. They also restrict the mRNA length to the minimum, thus limiting energy expense. Furthermore, as strong transcription may interfere with plasmid replication, transcription terminators may isolate the origin of replication and allow plasmid copy number to be maintained (Balbas and

Bolivar, 1990). In addition, it is believed that the presence of strong terminators at the end of a gene can greatly increase the message half life (Schmeissner et al.,1984).

However, the gene to be expressed may itself contain one or more sequences that lead to termination. A general solution to such a problem is offered by vectors that utilize the λ antitermination system to override transcription stop signals (Das, 1990; Gottesman, 1980). This is achieved by cocloning of the λ *Nut*L site which produces the λ N gene product. This binds *E.coli* NusA protein which is complexed with RNA polymerase core enzyme and alters the response of elongation complex to transcription terminators (Salstrom and Szybalski, 1978).

3.1.3.2.2 Transcription control

A common feature of bacteria is their ability to adapt readily to changes in their environment. For example, the genes for lactose utilization are activated once *E.coli* cells sense the presence of lactose or a similar compound and the lactose utilizing enzymes are induced. Many other genes are similarly kept in a turned-off state until their products are needed. Still other genes, concerned primarily with anabolism, are kept in a turned-on state unless the product of the anabolic sequence is present, and then the genes are repressed or turned off.

Both enzyme induction and repression are achieved through control of transcription, primarily at the initiation step. An example of gene regulation at the level of transcription, that

soluble form. In general 3 steps are followed to achieve this purpose:

Firstly, the forces holding the inclusion bodies together must be broken and the proteins solubilized. Solubilization agents include:

a) 5-8M guanidine chloride.

b) 6-8M urea.

c) Detergents.

d) Alkaline pH (>9).

e) Organic solvents.

Shoemaker *et al.* (1985) discovered a small amount of monomer with disulphide bridges on cytoplasmic expression of prochymosin in *E.coli*. However, it is known that the cytoplasm of *E.coli* is kept in a reduced state (Fahey, 1977). It was presumed that this disulphide bonding occurred on exposure to air, during cell lysis or in an oxidizing environment in the inclusion bodies. It is reasoned that any disulphide bridging that occurs under such conditions is likely to be incorrect and most protocols include thiol reagents such as mercaptoethanol to disrupt these.

Secondly, the protein can be purified by a number of strategies including ion exchange or gel chromatography, centrifugation or by extraction of proteins into iso-butanol or methylbutanol. The choices here depend on the protein in question and the denaturant used (Marston, 1987).

Thirdly, the disrupted proteins must be slowly introduced to the appropriate conditions to allow the correct intramolecular interactions including disulphide bonds to form. Generally, dialysis is employed to remove denaturant, dilute and introduce the denatured protein to an environment conducive to folding (Marston, 1987).

Clearly there are a large variety of expression systems for *E.coli* from which to choose. While the choice is to some extent empirical, at least a general understanding of expression technology is required before attempting expression.

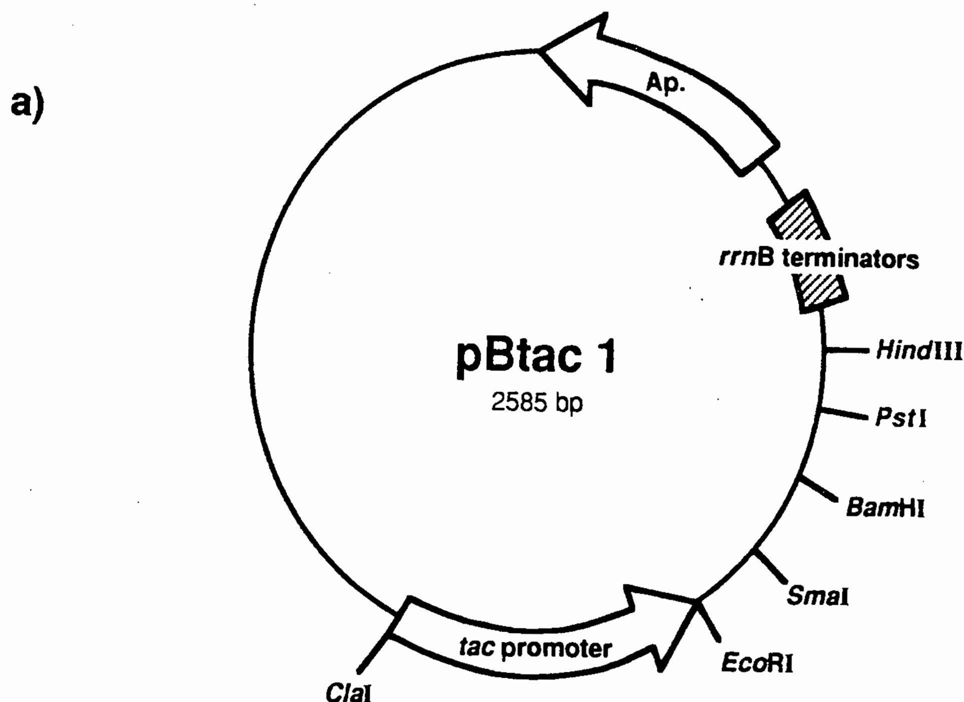
3.2 Expression in pbtac1™ vector

3.2.1 The pBtac1™ expression vector

As discussed in Chapter 2 the synthetic gene for ETI was provided with restriction sites for cloning into the expression vector pBtac1. The pBtac expression vectors were obtained from Boehringer Mannheim and supply the strong *tac* promoter which is regulated by the repressor of the *lacZ* gene, the *lacZ* ribosome binding site, strong *rrnB* ribosomal RNA transcription terminators, which have been inserted downstream of the multicloning site of pUC8 and a gene for ampicillin resistance (see Figure 3.1).

The *tac* promoter is a hybrid promoter consisting of the -35 region of the *trp* promoter and the -10 region of the *lac*UV5 promoter (De Boer *et al.*, 1983). The *lac*UV5 promoter was derived from a UV induced mutation of the *lacZ* promoter. This mutation changed the *lac* -10 region from 5'- TATGTT-3' to the consensus 5'-TATAAT-3' -10 sequence (Carposius et al., 1982). However, *lac*UV5 has a non-consensus 5'-TTTACA-3' -35 region. Conversely, the *trp* promoter has a consensus 5'-TTGACA-3' -35 region but a non-consensus 5'-TTAACT-3' -10 region. In addition, the *lac*UV5 promoter supplies a good S-D sequence, viz 5'-AGGA-3' in the ribosome binding site, while the *trp* promoter has a less than ideal 5'-GGTA-3' S-D sequence. The fusion of these two pieces of DNA, one containing the -35 region of the *trp* promoter and the other containing the -10 region of the *lac*UV5 promoter, its *lac* operator and its

a)



b)

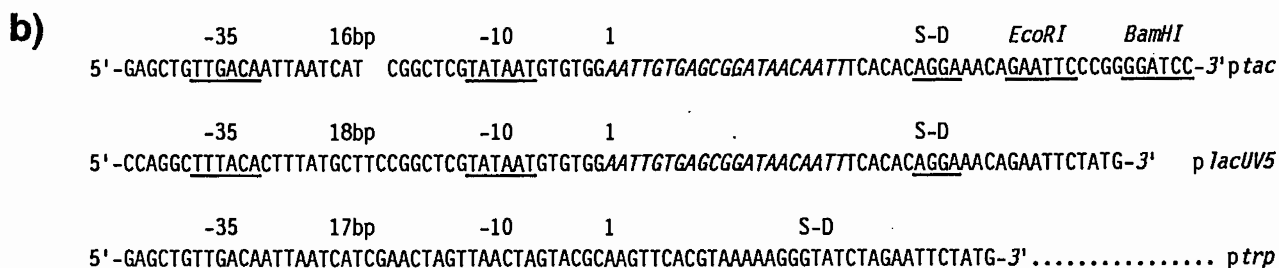


Figure 3.1 Plasmid pBtac1 and the origin of the tac promoter.

a) Map of pBtac1 (not to scale). Size is indicated in base pair (bp). The *tac* promoter and ampicillin resistance gene with its promoter are indicated by arrows.

b) Shows the *tac* promoter (p*tac*) and other control elements which are derived from the *trp* and *lac*UV5 promoters (p*trp* and p*lac*UV5). The -35, -10, Shine-Dalgarno (S-D) sequence and restriction sites are indicated by underlining. The transcription start sites are indicated by 1. The *lac* operator is indicated for *tac* and *lac*UV5 by italicized sequence.

ribosome binding site (see Figure 3.1) has given rise to a promoter that conforms with the consensus sequence for promoters and is far more efficient than either of the parental promoters (De Boer *et al.*, 1983).

The *tac* promoter is repressed in *lacI*^{q} strains of *E.coli*, eg JM105 which has resident the F′ episome carrying, inter alia, the *lacI*^{q} mutant form of the *lacI* gene, allowing for overexpression of the *lac* repressor (Yannish-Perron *et al.*, 1985). In such a host the *tac* promoter is induced by IPTG. This promoter has been used to give high yields of expressed product from certain recombinant genes. For example, De Boer *et al.* (1983) found the *tac* promoter to be 5-10 times more efficient than the *lac*UV5 promoter for the production of hGH. Interferon was produced to high levels by Amman *et al.* (1983) and in addition, they produced the *c*I repressor of bacteriophage λ to 30% of total cell protein using the *tac* promoter. More recently, Smith *et al.* (1990) recovered, correctly folded and active, the eukaryotic protein, horseradish peroxidase isoenzyme C (HRP C), at a yield of 2-3% of total cell protein. In this example, HRP C was produced from a synthetic gene driven by the *tac* promoter. Thus the *tac* promoter was considered to be ideal for cytoplasmic expression of synthetic ETI.

3.2.2 Construction of pTETI expression vector

The expression vector pTETI was constructed by directional cloning of the synthetic gene for ETI into pBtac1. The gene for

ETI was removed from pLT8 by digesting firstly with the unique *Eco*RI site in pLT8 (see Chapter 2.3) to linearize the plasmid. Linearized plasmid was purified, precipitated and digested further with *Bam*HI releasing the synthetic ETI gene. The gene-bearing DNA fragment was gel-purified from the rest of the plasmid.

To receive ETI, pBtacl was digested firstly with *Bam*HI, purified, and precipitated. Linearized plasmid was digested with an excess of *Eco*RI to allow some cleavage at the *Eco*RI site which was located 4 base pairs from the end of the linearized DNA. The large fragment was gel-purified as before. The two pieces of DNA were ligated and used to transform JM105 to ampicillin resistance. Transformants were selected at random, DNA was extracted by the miniprep method discussed in Chapter 6 and digested with the appropriate restriction enzymes to verify single insert and orientation. DNA was extracted from promising transformants by the maxiprep method and sequenced by Sanger dideoxy-sequencing using the Fow-02 and Rev-02 primers shown in Chapter 2.5.

3.2.3 Expression of ETI from pTETI

E.coli JM105(pTETI) was grown in LB (Luria broth) containing 100 μg/ml of ampicillin to late log phase (A600 0.6-0.8), the *tac* promoter controlling the synthetic ETI gene was then induced by addition of IPTG to 1 mM. The cells were allowed to grow for a further 3 hrs under inducing conditions. As a control JM105 (pBtacl) was subjected to the same conditions of

growth and induction as JM105(pTETI). Samples were taken prior to induction and after 3 hrs of post-induction growth.

Figure 3.2 shows an SDS-PAGE gel after staining and destaining of samples run as described above. From this it is clear that there is no difference in banding pattern between the three samples and no accumulation of any band after induction of pTETI in particular of ETI size range (approximately 20 kDa). Thus the synthetic gene for ETI was not expressed at detectable levels in this vector/host system.

It was established in Chapter 3.1 that current understanding of gene expression requires that the optimal expression system for any gene be chosen empirically. Thus no further work was performed to establish the reason for failure of the JM105(pTETI) system to express ETI and this can only be speculated upon at this point. Such reasons may include: the ribosome binding site being caught up in secondary structure, degradation of mRNA prior to translation or degradation of newly synthesized ETI in the cytoplasm before it can accumulate to form inclusion bodies.

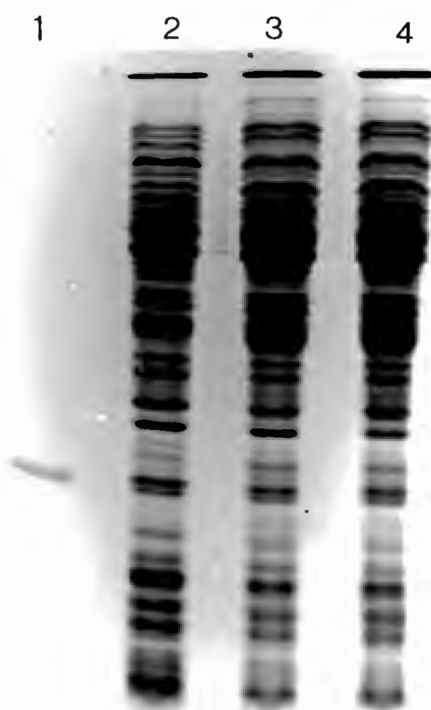


Figure 3.2 **PAGE gel of samples from the attempt to express the synthetic ETI gene of JM105 (pTETI).**

Lane 1) ETI purified from seeds of *Erythrina caffra* and kindly supplied by Serevac South Africa
2) JM105(pTETI) just prior to induction
3) JM105(pTETI) 3 hrs after induction

4) JM105(pBtac1) 3 hrs after induction

3.3 Expression from pMAL^{TM}-c expression vector

3.3.1 The pMAL expression vector

The pMAL expression vectors are supplied by New England Biolabs (NEB) and provide a single system for expression of genes and purification of the resultant proteins (see Figure 3.3). These expression vectors supply the *malE* gene (which codes for maltose binding protein or MBP) fused through a polylinker region to the *lacZ*α gene. Transcription of the fusion complex is controlled by the *tac* promoter and strong *rrnB* ribosomal RNA transcription terminators have been cloned downstream of the *lacZ*α gene to prevent transcription from p*tac* from interfering with plasmid functions. The promoter is kept in a switched off or repressed state by lac repressor, the product of the *lacI* gene, which is also carried by the pMAL vectors. Cloning of a coding sequence (which contains a stop codon such as TAA) into the polylinker and derepression of the promoter by IPTG interrupts the *malE-lacZ*α fusion. This changes colony colour from blue to white on X-gal in an α-complementing *E.coli* host such as TB1. Furthermore, cloning into this region gives rise to a fusion protein consisting of MBP and the target protein. To separate the two domains after purification, a sequence coding for the cleavage site of blood coagulation factor X_a protease has been inserted between the *malE* gene and the polycloning site. This provides a mechanism for cleavage of the fusion protein to release target protein from MBP (Maina *et al.*, 1988).

a)

Ap.
ori
lacI
rrnB terminators
pMAL-c
6145 bp
ptac
lacZ
MalE Δ2-26
polylinker

b)

```
              SacI                                                   StuI     EcoRI
malE...TCG AGC TCG GTA CCC GGC CGG GGA TCC ATC GAG GGT AGG CCT GAA TTC AGT AAA ACC CTC GAT GGA TCC TCT
                                                 Ile Glu Gly Arg
                                                              ^
                                                              |
                                                       Factor Xa cleavage site
                        HindIII
AGA GTC GAC CTG CAG GCA AGC TTG...lacZα
```

Figure 3.3 The pMAL-c vector

a) Map of pMAL-c. This contains the inducible p*tac* promoter positioned to transcribe a *malE-lacZ*α gene fusion. The *lacI* gene encodes the lac repressor, which turns off transcription from p*tac* until IPTG is added. The *rrnB* terminators prevent transcription from interfering with plasmid functions.

b) The polylinker provides several restriction endonuclease sites to insert the gene of interest, fusing it to the *malE* gene.

MBP is an *E.coli* K-12 periplasmic protein of 370 amino acids (42 kDa) that is involved in the transport of maltose and maltodextrins across the bacterial envelope, as well as in the chemotaxis towards these sugars (Kellerman and Ferenci, 1982; Duplay *et al.*, 1984). MBP is fused with target protein in pMAL vectors to facilitate purification and offers the following advantages: firstly, MBP binds with high affinity to cross-linked amylose matrix, and is released from the matrix by 10 mM maltose. Thus high yield and purity can be achieved in a single affinity chromatography step. Secondly, the purification takes place under physiological conditions and denaturation is unlikely to occur during these steps. Thirdly, the fused protein can be directed to the periplasm by incorporation of a *malE* export signal sequence, as is the case with the pMAL-p vector. The periplasm offers an environment conducive to disulphide bond formation and MBP does not contain any cysteine residues that could interfere with disulphide bond formation within the target protein (Guan *et al.*, 1988). In addition to the advantages discussed above, formation of a fusion protein can afford protection to the mRNA and to the protein itself. A fusion may overcome the inability of p*tac* of pBtac1 to express ETI.

Factor X_a (FX_a) is a serine protease involved in blood coagulation. It is derived from cleavage of a specific bond of factor X (FX) which is a plasma glycoprotein of 55 kDa composed of a heavy chain of 38 kDa and a light chain of 17 kDa. A number of non-physiological mechanisms for the activation of FX have been described, for example, a protein present in Russell's viper venom. Action of Russell's viper venom on FX

results in the release of a glycopeptide of 11 kDa from the amino terminal end of the heavy chain (Fujikawa *et al.*, 1972). This gives rise to FX_a, a protein of 44 kDa that recognizes the tetrapeptide Ile-Glu-Gly-Arg and cleaves at the carboxy-terminal side of Arg, between it and any amino acid that follows. As a result it can be used to produce a protein with any desired N-terminal amino acid (Maina *et al.*, 1988). In mammalian blood, FX_a converts prothrombin to thrombin during the coagulation process (Jackson, 1985).

Nagai and Thogersen (1984) and Nagai *et al.* (1985) used a construct containing a FX_a cleavage site to generate β-globin for which they required a native N-terminus. To achieve this, and to prevent secondary structure formation of the mRNA in the region of the S-D sequence, they inserted the sequence for FX_a between the 31 N-terminal amino acids of λ cII protein and Val 1 of human β-globin and produced this hybrid in high yield in *E.coli*. They then cleaved the hybrid specifically at the single Arg and liberated the authentic β-globin chain.

The strategy for expression and purification of proteins using the pMAL system is shown schematically in Figure 3.4. It involves cloning into the multicloning site of a pMAL vector, preferably into the *Stu*I site to avoid vector derived amino acids after cleavage with FX_a. Expression is started as with pBtac1 by induction with IPTG and allowing synthesis to continue for an optimal period that has been established empirically. Cells are then lysed, cell debris is centrifuged out and the remaining solution is passed over an amylose resin to allow the hybrid protein to bind to the amylose and the

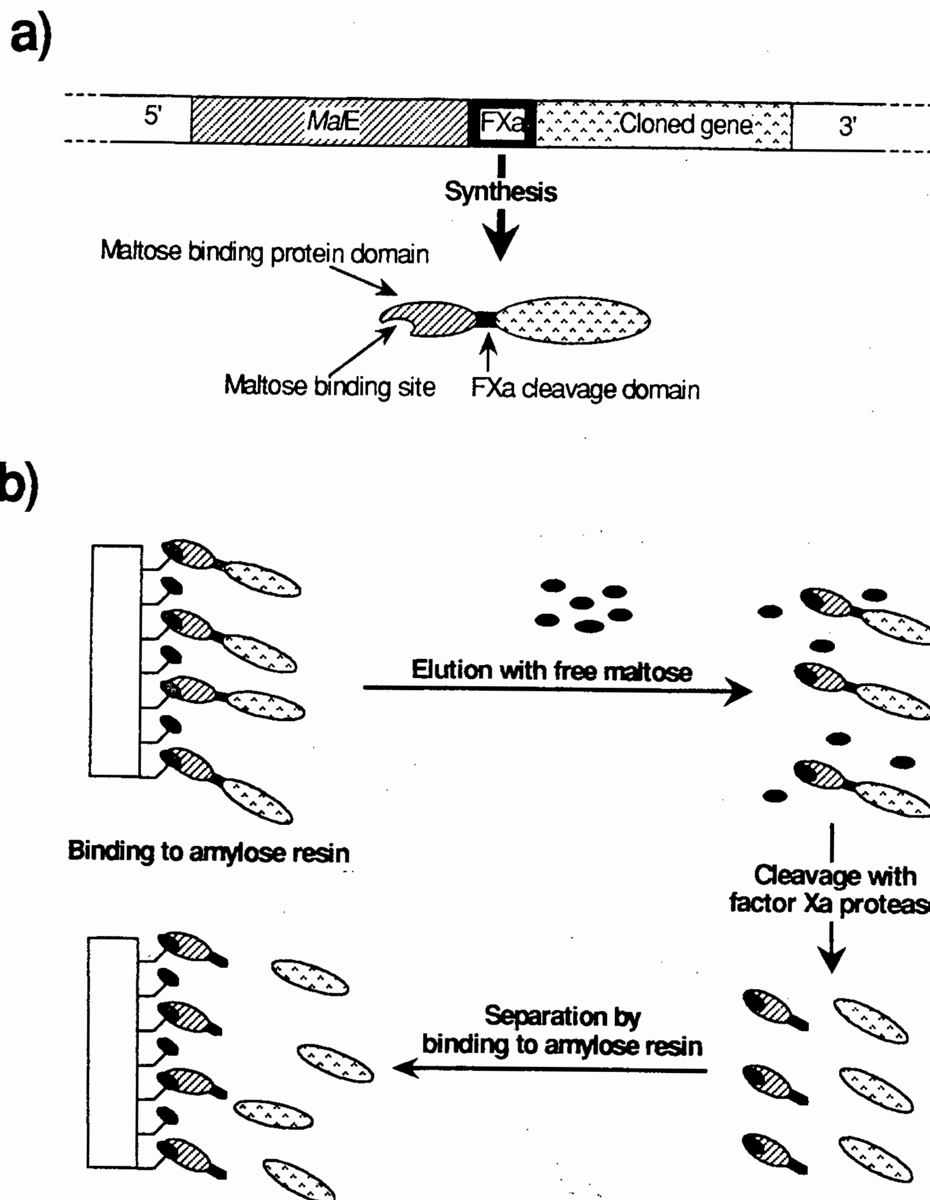


Figure 3.4 Expression and purification with the pMAL vectors.

a) Schematic representation of a portion of a pMAL vector showing the *malE* gene, the coding region for factor X_a and a hypothetical gene cloned into the polylinker and the protein that results from expression of this gene.

b) Schematic representation of the synthesis and purification of MBP fusion protein with a hypothetical protein. FX_a indicates the recognition site for factor X_a. The cleared lysate of induced cells is run through a crosslinked amylose column. After washing, the fusion protein is eluted with a dilute maltose solution. The purified fusion protein is cleaved with factor X_a and the target protein purified away from the MBP domain by repassage through the crosslinked amylose column (Maina *et al.*,1988).

remaining proteins to flow through the column. Purified hybrid is eluted from the column with 10 mM maltose and after removal of maltose, target protein is released by digestion with FX_a. MBP is then removed from solution by passing it again over the amylose column. This strategy was used successfully to express both a bacterial β-galactosidase and an invertebrate paramyosin gene and to purify the resultant proteins (Maina *et al.*, 1988).

3.3.2 Alteration of the N-terminus of synthetic ETI

As the pMAL expression system has the potential to produce an N-terminus of choice, it was decided to alter the synthetic ETI gene to produce a Val-Leu-Leu amino terminus as occurs in natural ETI rather than the Met-Leu-Leu which was designed for expression in pBtac1.

A unique XbaI (TCTGAG) site had been placed near the 5′ end of the gene. Cutting at this site separated the Met-Leu-Leu coding sequence and a C from the coding strand of the synthetic ETI gene (see below).

```
                   Met Leu Leu Asp Gly
            ...GAATTC ATG CTT CTA GAC GGT...
            ...CTTAAG TAC GAA GAT CTG CCA...

                            | XbaI digestion
                            ▼
...GAATTC ATG CTT                    CTA GAC GGT...
...CTTAAG TAC GAA GAT C                  TG CCA...
```

The gap created was bridged by an oligonucleotide carrying the sequence for an *Xho*I restriction endonuclease site, the amino acids Val, Leu, Leu and a G to replace the G of the GAC codon for Asp which would otherwise be lost in the bridging:

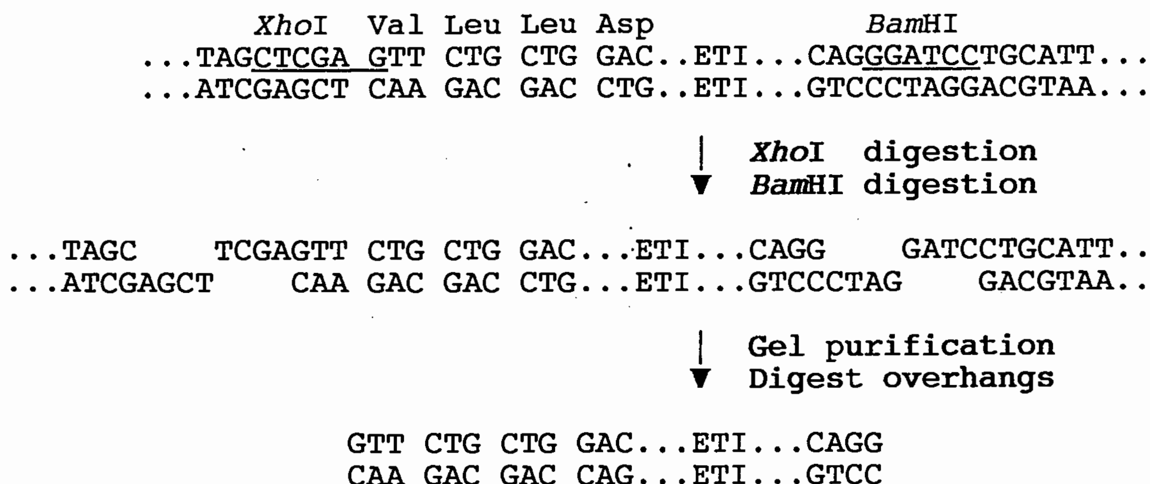
```
      15 base 5'-arm  XhoI  Val Leu Leu   15 base 3'-arm
   5'-AATTCATGCTTCTAG CTCGA GTT CTG CTG G ACGGTAACGGTGAAG-3'
..GGGCTTAAGTACGAAGATC                     TGCCATTGCCACTTCAACA
```

The altered plasmid, pLT8.1 was screened for the unique *Xho*I site introduced by bridge mutagenesis and the modified ETI gene sequenced as described earlier.

The *Xho*I site was introduced 5′ to the ETI gene in pLT8.1 as it was absent in pLT8. In addition, the *Xho*I site ends in a G which is the 1st base of the codon for Val and when the overhang after *Xho*I digestion is removed, this leaves a coding sequence that begins with the codon for Val. These properties were utilized to remove ETI from pLT8.1 and to prepare it for insertion into pMAL-c.

3.3.3 Construction of pMetI the pMAL-c/ETI fusion

Plasmid pLT8.1 was digested firstly with *Xho*I to linearize it and then with *Bam*HI as there are two of these sites in pLT8.1 (as discussed in Chapter 2.3). This releases an approximately 550 base pair (bp) fragment which was gel purified and the 5′ overhangs at each end of the gene-bearing fragment were removed by digestion with mung bean nuclease. This gave rise to a blunted DNA fragment bearing the gene for ETI.



The pMAL vectors have been designed with a *Stu*I site which gives blunt ends and allows for cloning directly 3′ to the sequence coding for FX_a such that on cleavage with factor X_a no vector derived sequence is carried with the target protein. This site was used for cloning the blunt fragment carrying the ETI gene. As the fragment can be cloned in either orientation or in more than one copy, to characterize plasmids were digested with *Bst*EII which has a single site in the vector and one in the ETI-coding sequence. The expected fragments of *Bst*EII digestion are shown in Figure 3.5

Clones selected by the initial *Bst*EII screen were sequenced, using the ETI primers shown in Chapter 2.5 to ensure that the gene had remained intact. The clone selected was called pMETI and used for subsequent expression work.

3.3.4 Expression of the fusion protein coded for by pMETI

The plasmid pMETI was introduced into *E.coli* strain TB1 by transformation and the stability of the plasmid in this host determined by restriction digestion and sequencing. In this

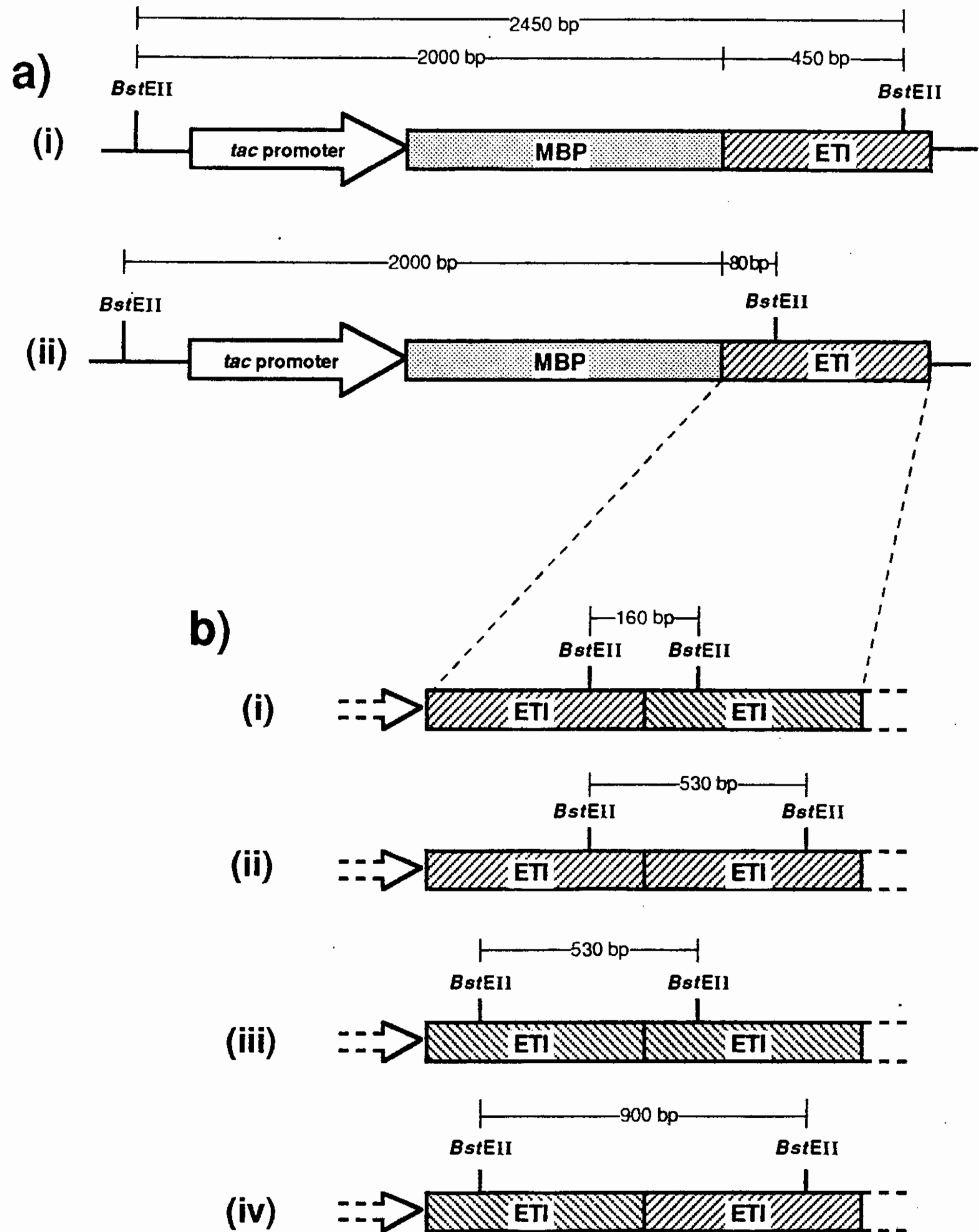


Figure 3.5 Diagram showing possible band sizes (sizes in base pairs) from BstEII digestion of putative clones of ETI in pMAL-c

a) Predicted sizes of bands that would be removed from pMAL-c bearing a single copy of ETI.

i) ETI in correct orientation for expression from the *tac* promoter removes a ± 2 500 base pair band leaving a ± 4 100 bp band.

ii) Incorrect orientation of ETI produces two bands of ± 2000 and 4 600 bp each.

b) In addition to the above bands, the presence of one or more of the four bands shown here was used as an indicator of multiple ETI bands.

manner, a suitable transformant, TB1(pMETI), was selected for production of ETI.

Expression was induced, as discussed in Chapter 6, by addition of IPTG to a mid log phase culture and incubating further to allow for accumulation of the fusion protein. Samples taken prior to induction and 3 hrs thereafter, were run with suitable controls on a 15% polyacrylamide gel and shown in Figure 3.6. The controls show no apparent accumulation of any protein in the host *E.coli* TB1 after induction (lanes 5, 6 and 7). In addition, there appears to be no accumulation of any protein in TB1 (pMETI) prior to induction (lane 2) as compared with the host lanes. This indicates that there is no apparent fusion protein synthesis from pMETI in TB1 prior to induction and that the repressor protein coded for by pMAL-c prevents detectable levels of expression from the *tac* promoter. However, 3 hrs after induction, TB1(pMAL-c), shown in lane 4, produced a heavy band which appeared to be slightly larger than the 45 kDa marker protein band of lane 1. Pure MBP, shown very faintly in lane 8, is a 42 kDa protein (Duplay *et al.*, 1984) thus it was concluded that the large band of lane 4 was a fusion of MBP and the α-fragment of β-galactosidase. Lane 3 shows TB1(pMETI) 3 hrs after induction, and here a dominant band has appeared which migrates slightly ahead of the 66 kDa marker band seen in lane 1. As MBP is a 42 kDa protein and ETI is approximately 20 kDa, a fusion of these two would be expected to yield an approximately 62 kDa band on gel electrophoresis. An MBP-ETI fusion would not be expected to include α-fragment as the stop codon of ETI would prevent its inclusion during translation. From Figure 3.6, it appears that the MBP-ETI fusion protein

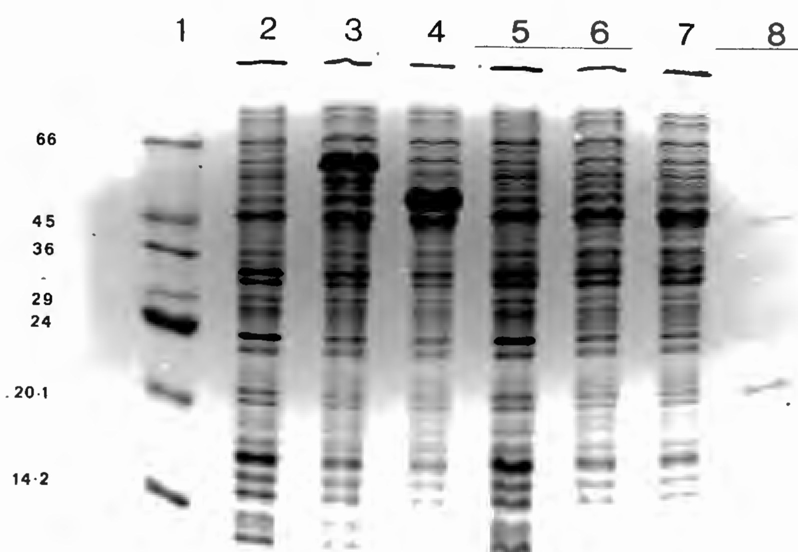


Figure 3.6 Expression of ETI-MBP fusion protein

Lane 1) Protein molecular weight markers (kDa)
2) TB1(pMETI) uninduced time 0hrs
3) TB1(pMETI) induced time 3hrs
4) TB1(pMAL-c) induced time 3hrs
5) TB1 uninduced time 0hrs
6) TB1 induced time 1hr
7) TB1 induced time 3hrs
8) ETI(natural) + MBP

represents about 25% of the total protein produced by this expression system.

As was discussed previously, overproduction of a protein in *E.coli* often results in its being sequestered in inclusion bodies. Thus, before purification could be attempted, it was necessary to determine the state of the expressed fusion protein in the cytoplasm, ie to determine if it was soluble in the cytoplasm or if it had been sequestered in inclusion bodies. To do this the pellet was resuspended in a lysis buffer (see Chapter 6) containing tween and β-mercaptoethanol and lysed by freeze-thawing and sonication. The cell debris was removed by centrifugation and a sample from the supernatant and the pellet was removed for analysis. The results are shown in Figure 3.7 from which it can be seen in lane 5 that under the conditions of cell lysis used for this work the fusion protein is soluble and located in the supernatant. The insoluble fraction contains an insignificant amount of the fusion protein and is shown in lanes 6 and 7, and could be as a result of supernatant remaining with the pellet after cell lysis.

A 1 litre culture of TB1(pMETI) was induced, lysed and the lysate purified on a column of cross-linked amylose and the eluate collected in 10 fractions. Aliquots from each fraction were pooled and a sample from each pool run on a gel. From Figure 3.8 it can be seen that pools 5 and 6 (lanes 6 and 7) appear to have the highest concentrations and purity of the fusion protein and it was decided to use the fractions from which these pools were derived. In addition to the fusion protein, a band of roughly 45 kDa is also purified. It is

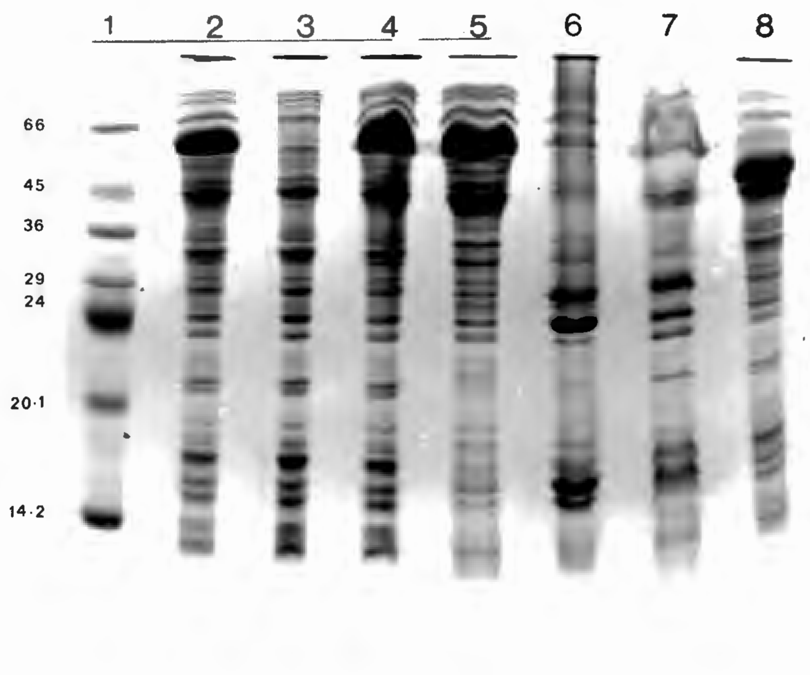


Figure 3.7 **Determination of location of the ETI-MBP fusionprotein after cell lysis**

Lane
1) Protein molecular weight markers (kDa)
2) TB1(pMETI) induced time 3hrs
3) TB1(pMETI) uninduced time 0hrs
4) TB1(pMETI) induced time 3hrs
5) TB1(pMETI) induced time 3hrs supernatant of lysed cells
6) TB1(pMETI) induced time 3hrs solubilized pellet of lysed cells
7) TB1(pMETI) induced time 3hrs pellet remaining from lane 6
8) TB1(pMAL-c) induced time 3hrs

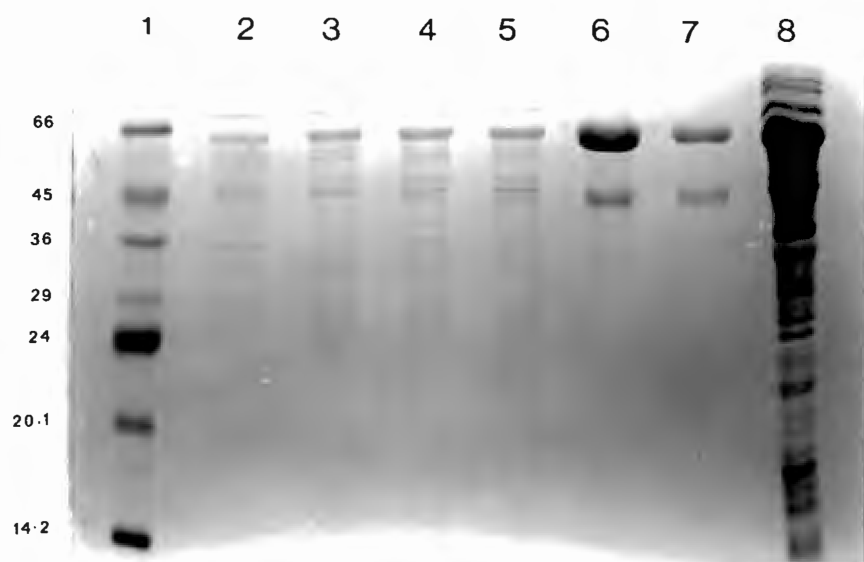


Figure 3.8 **Pooled eluates after affinity chromatography of partially purified TB1(pMETI) cell lysate**

Lane
1) Protein molecular weight markers (kDa)
2) Pool 1
3) Pool 2
4) Pool 3
5) Pool 4
6) Pool 5
7) Pool 6
8) TB1(pMETI) induced time 3hrs

believed that this band is a MBP-β-galactosidase-α-fragment fusion produced by a subpopulation of the TB1 (pMETI)population that had lost the ETI gene. This band is also seen in the results presented by Maina *et al.*(1988).

The selected fractions were 10 mM in maltose which had to be removed so that after cleavage with FX_a, free ETI could be separated from free MBP by loading again on the cross-linked maltose resin. It was decided to remove the maltose from the pooled fractions by dialysis. Experiments with fluorescent spectroscopy have shown that the rate of dissociation of complex between maltose and MBP is high (Silhavy *et al.*,1975). Nonetheless, to ensure diffusion of maltose from the dialysis bag, it is essential to make the solution very dilute in MBP, thus the pooled fractions were diluted to 400ml with 10 mM Tris (pH 8), 100 mM NaCl giving a concentration of about 80 μg/ml. This was dialyzed against 40L of 10 mM Tris (pH 8), 100 mM NaCl with 6 changes over 24hrs at 4°C.The dialysate was concentrated to 15 ml with a millipore immersible CX-30 ultrafilter. The concentrate was dialysed against 1 L of FX_a buffer for 24hrs with 3 changes of buffer at 4°C. A 50 μl sample was taken and incubated at a constant 22°C for 24 hrs with 2.5 μl of 200μg/ml FX_a (ie in this instance FX_a was used at a w/w ratio of 1% the amount of fusion protein). Samples were taken for electrophoresis and shown in Figure 3.9. However it was apparent from this gel that incubation with FX_a even for 24hrs (lanes 3-6) did not cleave the ETI-MBP fusion protein at the FX_a recognition site to release a band that comigrated with ETI(natural) in lane 2. Nonetheless, some cleavage had occurred judging by the appearance of a very faint band, which was seen

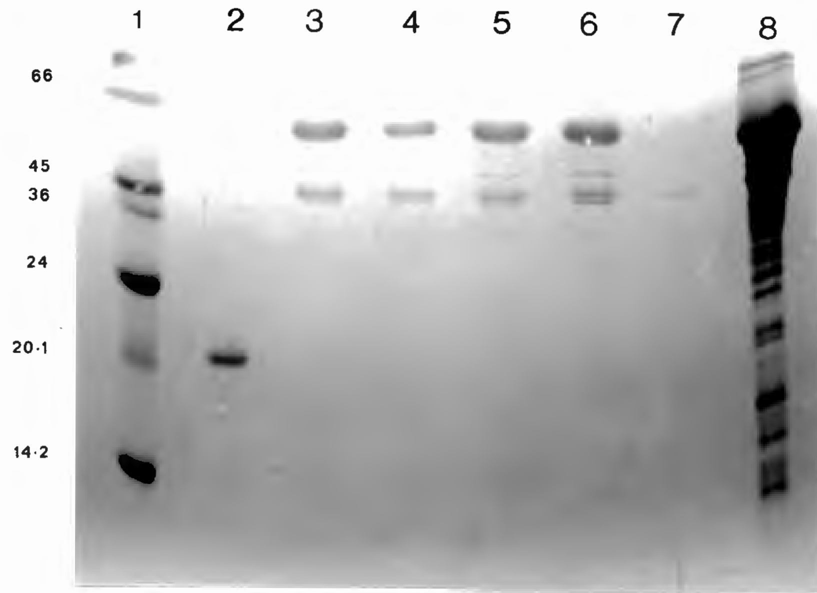


Figure 3.9 **Attempt to cleave ETI-MBP fusion protein with factor Xa**

Lane 1) Protein molecular weight marker (kDa)
2) ETI(natural)
3) ETI-MBP incubated with FX_a 2.5hrs
4) ETI-MBP incubated with FX_a 4 hrs
5) ETI-MBP incubated with FX_a 8 hrs
6) ETI-MBP incubated with FX_a 24 hrs
7) MBP
8) TB1(pMETI) induced time 3hrs

after 4 hrs of digestion and had migrated slightly ahead of the ± 45 kDa MBP-β-galactosidase-α-fragment band. After 8 hrs of digestion (lane 5) this band had increased in intensity and another faint band appeared at approximately 50 kDa, between the above mentioned ± 45 kDa band and the fusion protein and these increased further in intensity in the 24 hr sample (lane 6). In addition, the 24 hr sample gave rise to an extremely faint band that migrated at about 14 kDa.

It was postulated that the MBP-β-galactosidase-α-fragment fusion protein was cleaved slowly, freeing MBP (42 kDa) and gave rise to the band that migrated slightly ahead of 45 kDa. The smaller β-galactosidase-α-fragment was believed to either have run off this gel or to be too dilute to be discerned under the conditions in which this gel was run.

The appearance of the ± 50 kDa protein after 8 hrs of digestion and the appearance of the approximately 14 kDa band in the 24 hr sample suggest that some cleavage of the MBP-ETI fusion protein had occurred. MBP does not contain the sequence Ile-Glu-Gly-Arg required for FX_a cleavage (Maina *et al.*, 1988; Duplay *et al.*, 1984). However, Lauritzen *et al.* (in press, from the abstract) produced bovine pancreatic trypsin inhibitor (BPTI) extended at the N-terminus by 6 amino acids, N-Gly-Ser-Ile-Glu-Gly-Arg after expression from a pMAL vector and cleavage with FX_a. This indicated that FX_a had recognized an alternative site at the C-terminus of MBP, possibly under the influence of the cloned sequence. Such a cleavage was not suspected in this case however, as it would have given rise to cleavage products of which one would be larger but closer in

size to ETI than any band produced in this digest. Thus, it was concluded that cleavage was taking place in the ETI domain of the fusion protein.

That no ETI sized fragment was produced even after 24hrs of digestion indicated that, for this fusion protein, FX_a could not recognize or contact its recognition site between the MBP and ETI domains. There were two possible explanations for the inability of FX_a to separate the two domains:

1) The fusion had adopted a secondary structure that obscured the FX_a recognition site from this enzyme. Furthermore, the structure adopted left exposed an ETI sequence Ile-Glu-Ser-Arg that starts at residue 58 (see Figure 2.1) and could offer a cleavage site for FX_a. Further examination of the sequence of ETI revealed that cleavage at this putative FX_a site within the ETI domain of the MBP-ETI fusion would give a product of approximately 14 kDa leaving a fragment of about 50 kDa. At the time of choice of this expression vector it was believed that the true Fx_a cleavage site supplied by pMAL-c would be used in preference to the partial site within ETI.
2) As discussed earlier, FX_a is a serine protease and ETI is an inhibitor of serine proteases, known to inhibit trypsin, chymotrypsin and plasminogen activator. Furthermore, in many cases the inhibitor functions as a substrate for the enzyme and under appropriate conditions of time, pH and relative concentrations of inhibitor and enzyme, inhibitor can be cleaved at a single scissile bond situated at its reactive site. For ETI, the scissile bond lies between Arg63 and Ser64 (Figure 2.1) and cleavage at

this bond in the MBP-ETI fusion will also produce two bands of about 14 and 50 kDa each. However, in order to inhibit FX_a in this way, ETI must have folded into a structure approximating that of natural ETI. When choosing this expression vector it was reasoned that being part of a fusion protein of which it represents one third would prevent ETI from folding into its correct structure. In addition, the choice of the cytoplasm (rather than the periplasm) as the site of accumulation of the expressed protein was made to limit the ability of the protein to form disulphide bridges and further prevent it from inhibiting FX_a.

Examining the first possibility required that the fusion protein be denatured in an effort to expose the FX_a recognition site. To this end, samples of the purified fusion were dialysed, as suggested by NEB, against 100 volumes of 20mM Tris (pH 8), 6M guanidine hydrochloride or against 6M urea for 4 hrs at room temperature. To remove denaturant and prepare for digestion they were then dialysed against 2X 100 volumes of FX_a buffer for 4 hrs each. Thereafter, the samples were digested using FX_a at 1% and at 3% of the total protein concentration.

The results for guanidine hydrochloride as denaturant are shown in Figure 3.10, from which it can be seen that at all time intervals sampled, 3% FX_a has given increased digestion over that produced by the 1% FX_a recommended by NEB. In addition, the ± 14 kDa band appeared, albeit faintly after 8 hrs with 3%

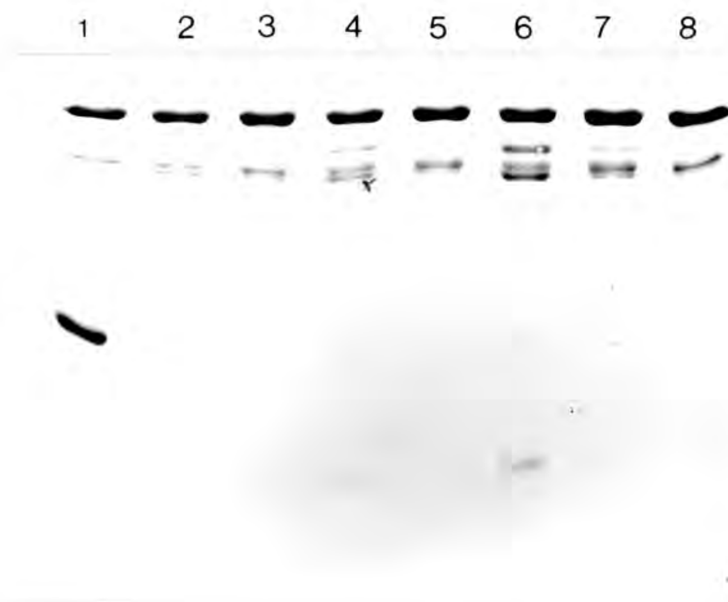


Figure 3.10 Further attempt to cleave ETI-MBP fusion protein after denaturation with guanidine hydrochloride

Lane 1) ETI-MBP fusion protein after affinity purification and concentration + ETI
2) ETI-MBP digested with FX_a (3%) for 4hrs
3) ETI-MBP digested with FX_a (1%) for 4hrs
4) ETI-MBP digested with FX_a (3%) for 8hrs
5) ETI-MBP digested with FX_a (1%) for 8hrs
6) ETI-MBP digested with FX_a (3%) for 24hrs
7) ETI-MBP digested with FX_a (1%) for 24hrs
8) ETI-MBP control with no FX_a for 24hrs

FX_a. This increased markedly after 24 hrs by which time an extremely faint band that comigrated with ETI appeared. From this it seemed, that upon dialysis against FX_a buffer, most of the protein had regained the structure that shielded FX_a from its recognition site and allowed for the formation of the ±14 kDa fragment.

The next step was to determine whether ETI was cleaved on interaction with FX_a and by implication that it inhibits also this protease. The approach used was to incubate purified natural ETI with FX_a at 3% of the total protein and under the same conditions that had been used in the attempt to cleave the fusion protease. A sample of this was used for electrophoresis and is shown in Figure 3.11. In lane 2 of this gel, digested ETI had produced a band of much the same size as seen on digestion of the fusion protein. The remainder of the reaction mix was then used to determine the sequence of the N-terminus of ETI and any cleavage products of ETI in the solution. The results of the limited sequencing is shown in Tables 3.1 a) and b).

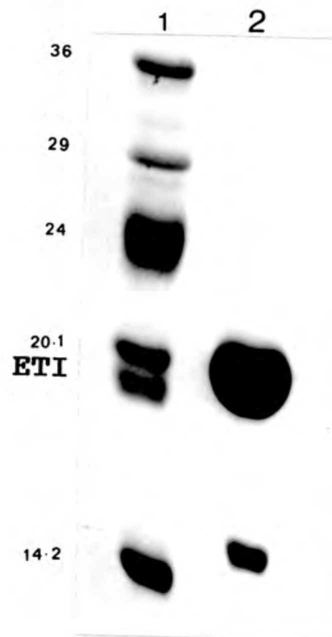


Figure 3.11 Determination of effect of FXa on ETI(natural)

Lane 1) Protein molecular weight markers + ETI (natural)
2) ETI(natural) digested with FX_a (3%) for 8 hrs

a)

cycle	PTH amino acid	YIELD (pmoles)
1	Val	104
2	Leu	171
3	Leu	184
4	Asp	48
5	Gly	90

b)

cycle	PTH amino acid	YIELD (pmoles)
1	Ser	47
2	Ala	117
3	Phe	94
4	Ile	31
5	Pro	95

Table 3.1 Results of automatic sequencing of the products of FX_a digestion of purified natural ETI. The amino acid residues that produce the two largest peaks and their yields are reported.

a) The sequence that could be aligned with the known sequence for ETI.

b) The second sequence which appeared concomitantly with that shown in a.

As ETI was seen in Figure 3.11 to be cleaved on reaction with FX_a two N-terminii were expected in the sequence. As these had not been purified from each other, each sequencing cycle had to

produce phenyl-thio-hydantoin (PTH) derivatives of sequential amino acids of each chain. Thus, for each cycle, the two highest yielding residues were recorded and one matched with the known N-terminal sequence of ETI. The second sequence was aligned with a sequence within ETI (see Figure 2.1). In this way it was found that the second sequence started at what was residue 64 in the sequence of ETI. This indicates that ETI has been cleaved at the scissile bond between Arg63 and Ser64 of ETI and hence that ETI is an inhibitor of FX_a and suggests that conditions of pH and inhibitor:enzyme ratio in the digest are conducive to cleavage of the inhibitor.

To understand the varying yields of the PTH-amino acids shown in Tables 3.1 a) and b), it is necessary to describe briefly the stepwise degradation of proteins by the method of Edman which was used in this sequencing reaction. To prepare the N-terminal amino acid for removal from the protein, phenyl-iso-thio-cyanate (PITC) is coupled to its α-amino group to produce a phenyl-thio-carbamyl-peptide (PTC-peptide). The PTC peptide bond is cleaved by causing the N-terminal PTC-amino acid to cyclize under acid conditions producing an anilino-thiazolinone (ATZ) derivative of the amino acid. The ATZ-amino acids are unstable and are converted to PTH-amino acids usually by heating to 80^oC. It is at the stage of conversion of ATZ-amino acids to PTH-amino acids where much of the problem arises; some PTH-amino acids tend to decompose during the conversion and decrease the yield of those amino acids (Han *et al.*, 1985).

3.3.5 Discussion

The need to produce forms of ETI with specified N-termini has been discussed but it is clear that this requires a specialized expression system. It was primarily for this reason that the pMAL expression system was chosen to express the ETI-encoding gene. The system provides a leader sequence in the form of MBP which it was reasoned would protect the expressed ETI from degradation.

A drawback to the use of this system is that FX_a is used to separate ETI from MBP to which it is fused. FX_a is a serine protease much like trypsin and tPA and could be inhibited by ETI. In addition, ETI has a sequence of residues 58 to 62 that is remarkably similar to the FX*a* recognition site. It was reasoned that if both sites were equally available to FX_a it would be more likely to utilize its precise recognition site in preference to the partial site in ETI. Furthermore, it was reasoned that as ETI contributed only about one third of the mass of the fusion protein it would not fold to the correct conformation to inhibit FX_a. If on the other hand it did fold correctly the MBP at the N-terminus of ETI was predicted to prevent it from binding to and inhibiting FX_a. Finally, Lauritzen *et al.* (1991) had been able to produce active bovine trypsin inhibitor (BTI), although with an altered N-terminus, using the pMAL expression and purification system. Considering the fact that BTI is also a serine protease inhibitor, it was considered that this would be an ideal system for the expression and facile purification of ETI.

While it was shown that good expression levels of the fusion protein were achieved, FX_a was unable to cleave at its

recognition site and free ETI from the fusion. Possible reasons for this were that FX_a could not reach its recognition site because of the structure adopted by the fusion protein or that ETI was in fact inhibiting FX_a. To examine the first possibility, the fusion protein was denatured and incubated with FX_a but no ETI sized cleavage product was produced. All attempts to cleave the fusion produced amongst others a $\pm$ 14 kDa fragment. A band of the same size was seen when purified natural ETI was incubated with FX_a which indicated that ETI does inhibit FX_a. This was confirmed on sequencing the amino-terminii of the products of incubation of pure natural ETI with FX_a.

Thus the evidence indicated that the ETI domain of the ETI-MBP fusion protein was able to inhibit FX_a. Taking into account that ETI has MBP attached to its N-terminus indicates that the kinetics of interaction in this case are different from that of ETI with trypsin or tPA.

That ETI could have attained the structure required for activity after being expressed in the cytoplasm where disulphide bridges are reputed to be unable to form can be explained by the methods used for purification. In addition to extensive exposure to air, the fusion was exposed to lysis buffer containing β-mercaptoethanol and tween and at a later stage after other purification steps it was diluted several-fold and dialysed against a buffer which was conducive to folding.

3.4 Expression of two forms of synthetic ETI in the pET expression system

3.4.1 The pET expression system

The pET expression system developed by Tabor and Richardson (1985) and Studier and Moffatt (1986) involves a number of vectors into one of which the target gene for expression is cloned under control of a promoter of bacteriophage T7. RNA polymerase of T7 in turn is supplied in a controlled manner by one of a number of *E.coli* hosts developed for this purpose.

There are 17 known promoters in the genome of bacteriophage T7 and these have a highly conserved sequence of 23 continuous base pairs starting from -17 to +6 relative to the start site of the RNA chain (Dunn and Studier, 1983). Individual promoters are identified by a ϕ followed by the number of the gene first transcribed from the promoter, except for *ϕOL* and *ϕOL*, which are thought to be parts of the replication origins located near the left and right ends of the DNA (Dunn and Studier, 1981). Assays of these promoters *in vivo* and *in vitro* show that all 17 are utilized by T7 polymerase to a greater or lesser degree (Golomb and Chamberlin, 1974). One of these promoters *ϕ10* (shown in Figure 3.12) is utilized in all of the pET expression vectors (Studier *et al.*, 1990).

conserved sequence	TAATACGACTCACTATAGGGAGA
ϕ10 promoter	ACTTCGAAAT TAATACGACTCACTATAGGGAGA CC
ϕ3.8 promoter	CGTGGATAAT TAATTGAACTCACTAAAGGGAGA CC

Figure 3.12 Promoters ϕ10, used in the pET expression system, ϕ3.8, an example of a classII promoter and the 23 base pair conserved sequence (from Dunn and Studier, 1983). Within the conserved sequence, boldtype indicates nucleotides that are unchanged in the conserved region of all 17 promoters. Within the *ϕ3.8* sequence, boldtype indicates nucleotides that are different from the conserved sequence. In all sequences, +1, the first nucleotide of the RNA chain, is underlined. In addition, spaces separate the conserved region from the rest of the sequence.

The promoters for T7 RNA polymerase in T7 DNA have been classified into three groups viz, class II, III and replication promoters based on their location, utilization and nucleotide sequence (Dunn and Studier, 1983). The five class III promoters to which *ϕ10* belongs, appear to be the strongest *in vitro* and *in vivo* and all have the same 23 base pair conserved sequence located at positions -17 to +6 (Golomb and Chamberlin , 1974). The class II promoters, all of which appear to be weaker than class III promoters, are different at a number of positions in the -17 to +6 region from class III promoters (Saito *et al.*, 1980).

The conserved sequence required to make a T7 promoter is of sufficient length to make unlikely its occurrence by chance in any DNA unrelated to T7 DNA. Furthermore, no such promoters are known to exist in the DNA of *E.coli*, the natural host of bacteriophage T7, which favours transcription of viral rather than host genes during infection (Studier and Moffatt, 1986).

The RNA polymerase of bacteriophage T7, the product of viral gene 1, is a single subunit protein of 10.7 kDa molecular weight and it exhibits *in vivo* and *in vitro*, almost absolute specificity for T7 promoters. In addition, T7 RNA polymerase initiates RNA chains efficiently and is a very active enzyme which elongates chains about five times faster than does *E.coli* RNA polymerase (Golomb and Chamberlin, 1974; Davanloo *et al.*, 1984). Furthermore, its progress along a DNA strand appears to be stopped, albeit inefficiently, by the only T7 terminator, *ϕT* and its progress appears to be unhindered by the factors that cause termination of transcription by *E.coli* RNA polymerase (Carter *et al.*, 1981). Therefore, T7 RNA polymerase should be able to produce complete transcripts from almost any DNA that is linked to a T7 promoter (Studier and Moffatt, 1986).

3.4.1.1 The pET expression vectors

These vectors contain the strong *ϕ10* promoter and the efficient translation initiation signals for gene 10 inserted into the *Bam*HI site of the multicopy plasmid pBR322 oriented such that the *bla* gene is expressed from the T7 promoter along with the target gene (Tabor and Richardson, 1985). There are a number of differences between the various pET expression vectors. The one chosen for this work, specifically for its ability to direct export of the expressed protein to the periplasm, was pET12a (Fig 3.13).

Each of the three pET12 vectors has a unique *Bam*HI in a different reading frame relative to the initiation codon; the

```
                  Met Arg Ala Lys Leu Leu Gly Ile Val Leu Thr Thr Pro Ile Ala Ile
AAGAAGGAGATATACATATG CCG GCG AAA CTC CTA GGA ATA GTC CTG ACA ACC CCT ATC GCG ATC
    SD

Ser Ser Phe Ala Ser Thr Gly Ser
AGC TCT TTT GCG TCG ACG GGA TCC
             SalI       Bam HI
```

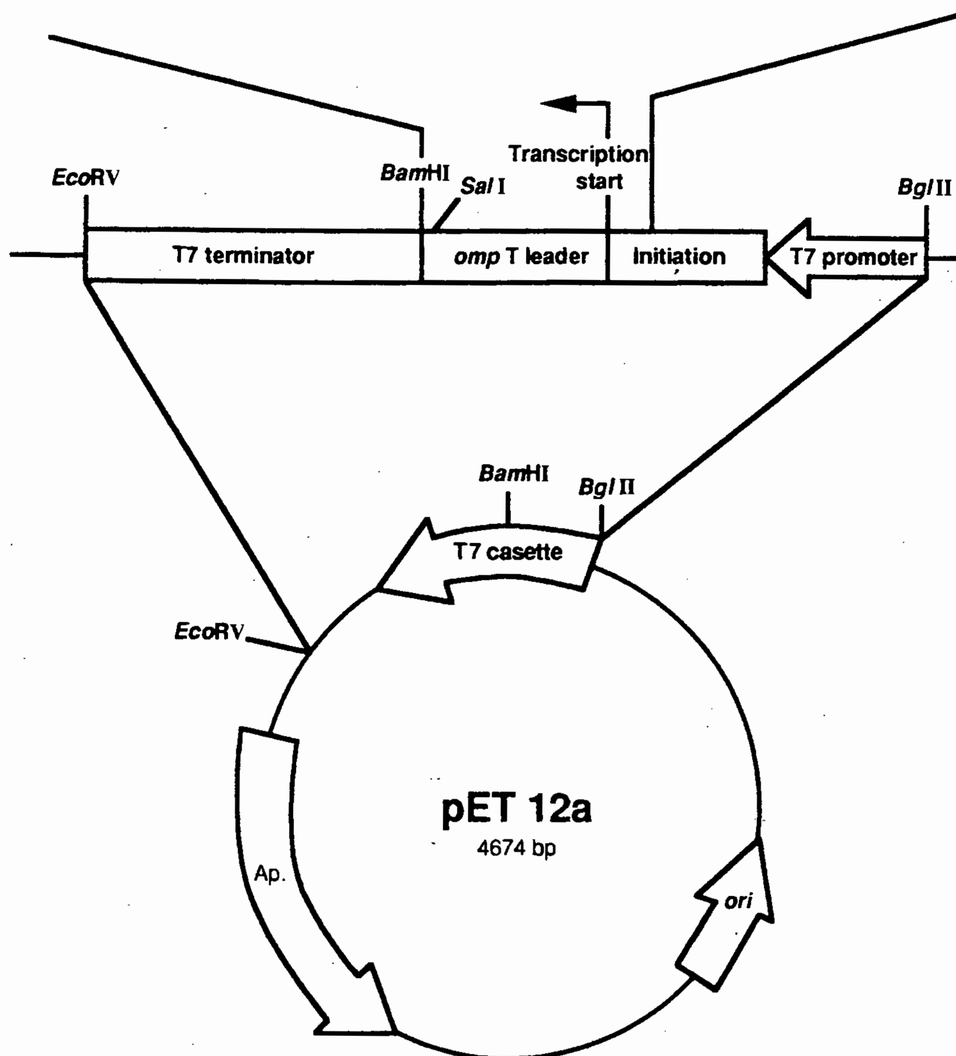


Figure 3.13 Map of pET 12a.

This contains the T7 ϕ10 promoter, translation initiation signals of T7 gene *10* inserted into pBR322 such that ampicillin resistance gene is expressed from the T7 promoter. Downstream of the T7 casette is cloned a T7 DNA fragment bearing the Tϕ terminator.

The *ompT* peptide starts with Met and ends with the Ala at the *Sal*I site. The leader peptide is cleaved on transport to the periplasm between this Ala and the Ser which follows.

suffix a, b, or c indicates whether the GGA, GAT or ATC triplet of the *Bam*HI site is in the open reading frame. These vectors produce a fusion protein consisting of 22 amino acids of the outer membrane protein specified by *ompT* of *E.coli* attached to the N-terminus of the target protein. The ompT leader peptide directs transport of the protein into the periplasm and is itself removed in the process. In addition, these vectors contain the Tϕ transcription terminator of T7 (Rosenberg *et al.*, 1987; Grodberg *et al.*, 1988).

3.4.1.2 Hosts for expression from pET vectors

The two hosts commonly used for pET expression are the *E.coli* K12 strain HMS174 (F^- *recA* $r^-_{k12}m^+_{k12}$ Rif^R) and the B strain BL21 (F^- *ompT* $r^-_B m^-_B$). BL21 is deficient in the lon and the ompT proteases that can degrade proteins during purification and for this reason was chosen as the host for production of ETI (Studier *et al.*, 1990).

To deliver T7 RNA polymerase to the cell, Studier and Moffatt (1986) established a DE3 lysogen in the host. Bacteriophage DE3 is a λ derivative that carries *inter alia* the *lacI* gene, the *lac*UV5 promoter, the beginning of the *lacZ* gene and the gene for T7 RNA polymerase. Once the lysogen is formed, T7 RNA polymerase is transcribed from the *lac*UV5 promoter which requires IPTG induction because of the presence of the *lacI* gene. In the resulting mRNA, T7 polymerase has its own transcription start site and hence is not produced as a fusion with the product of the beginning of the *lacZ* gene. Thus, addition of IPTG to 0.4 mM to a growing culture of BL21(DE3) or

HMS174(DE3) induces the T7 RNA polymerase, which in turn transcribes the target DNA in the plasmid.

However, the basal level of T7 RNA polymerase activity promotes some transcription of the target gene in the uninduced cell. As a result, some genes whose products are sufficiently toxic cannot be established in BL21(DE3) or HMS174(DE3). A method of reducing the basal activity of the polymerase is to supply the cell with T7 lysozyme a natural inhibitor of T7 RNA polymerase (Moffatt and Studier, 1987).

T7 lysozyme binds to T7 RNA polymerase and prevents transcription. However, it also degrades the peptidoglycan layer of the *E.coli* cell wall but appears unable to pass through the cell membrane to reach this layer. Lysozyme is provided to the cell from a clone of the T7 lysozyme gene in the plasmid pACYC184. A plasmid having this gene oriented to be expressed from the *tet* promoter of pACYC184 is referred to as pLysE; cells carrying this plasmid accumulate substantial levels of lysozyme. When the gene is in the opposite orientation it is known as pLysS and cells carrying this plasmid accumulate much lower levels of lysozyme. Neither lysozyme plasmid interferes with transformation of pET vectors in cells that carry it. Nonetheless, they do slow the growth rate of their host cells, pLysE to a much greater extent than pLysS but their presence is necessary to increase tolerance to toxic gene products and to stabilize unstable plasmids (Studier *et al.*, 1990).

The pET expression system has been successfully used by a number of workers to express a variety of genes. For example, Freedman *et al.* (1988) produced a 150 amino acid glucocorticoid receptor fragment in *E.coli* using the T7 system. Their yield, after purification, was 2 mg/L of starting material.

Wheat high molecular weight (WHM) glutenin was expressed in BL21(DE3)pLysS from pET3a at high yields (no data supplied), where other expression systems had yielded no WHM glutenin due to the toxicity of this protein to *E.coli* cells (Galili, 1989). A further example of expression of toxic gene products was described by Sano and Cantor (1990), where they expressed the extremely lethal streptavidin gene in BL21(DE3)pLysE with a yield of 39-65 mg/L of culture. In contrast, procathepsin D which is not toxic to *E.coli*, was produced from pET3a using the basal level activity of T7 RNA polymerase in BL21(DE3) (Conner and Udey, 1990). Thus it was decided to use the pET expression system to produce forms of ETI that could be further manipulated to achieve a desired N-terminus.

3.4.1.3 Proposed reconstruction of the synthetic gene for ETI

It has been established that ETI has a structured N-terminal region stabilized by hydrogen and hydrophobic interactions. The N-terminal region lies in close proximity to the protease when ETI is in complex with trypsin or tPA. Ideally it was required that the expressed ETI have the same sequence, including the N-terminal Val, as natural ETI in order to produce a protein with the same activity as natural ETI.

STI is very similar to ETI in its sequence and structure (as described in Chapter 1) and has an N-terminal region structured much like that of ETI. One difference between the two regions is that STI has an Asp N-terminal residue. It has been hypothesized that because of the sequence homology of ETI and STI and the hydrophilic nature of Asp it is not incorporated into the structure of the STI N-terminus. This leaves the residue free to form an ion pair with Lys60 of trypsin but interferes with the interaction of STI and tPA. One of the aims of this work was to add an Asp residue to the N-terminus of ETI to make it like STI in this region to investigate the abovementioned hypothesis. Thus there was a need to produce two forms of ETI differing in only their N-terminal residues.

Direct expression (ie. with no leader sequence) of the ETI-encoding gene was shown to be unsuccessful (Chapter 3.2). It was then attempted to express the gene as part of a fusion protein with MBP which produced good yields. However, FX_a appears to be inhibited by ETI and cannot cleave the MBP-ETI fusion (Chapter 3.3). Nonetheless, this exercise indicated the need for a leader sequence to assist in the expression of the ETI-encoding gene. It also showed that ETI may inhibit other serine proteases that would cleave it from leader sequences to which it could be fused. The pET12 series of vectors supply a leader sequence in the form of a signal sequence for export to the periplasm but the signal sequence cleavage system of *E.coli* leaves these proteins with an N-terminal Ser (Rosenberg *et al.*, 1987; Studier *et al.*, 1990). While this may have been suitable to make ETI that functions as the natural form of the protein it could not be used to produce ETI with an Asp N-terminal

residue. As ETI has no internal Met, it was decided to modify the 5′ end of the synthetic gene to generate ETI with an N-Ser-Met-Val... N-terminal region which can be cleaved in 70% formic acid, by CNBr, between Met and Val leaving ETI with the natural Val N-terminus (Gross, 1967; Boyot *et al.*, 1990).

To add an Asp residue to the N-terminus of ETI, the same approach was used and the 5′ end of the synthetic gene modified to produce Asp between Met and Val in the sequence above. The resultant protein could be cleaved with CNBr to produce a variant ETI possessing an Asp residue at its N-terminus.

3.4.2 Reconstruction of the 5′ end of the synthetic ETI gene

The plasmid pLT8 was cleaved with *Xba*I and the gap bridged (in the manner discussed in Chapter 2.3.2) with one of two oligonucleotides: oligo- Val which introduces an *Mlu*I restriction site and amino acid residues Ser, Met and Val in that order, or oligo-Asp which introduces an Asp residue between Met and Val to the above order. The resultant plasmids, called pLT8.val and pLT8.asp, respectively were each digested firstly with *Mlu*I to linearize the plasmid and then with *Bam*HI to remove the ETI gene.

The ETI gene-bearing fragment was gel purified and the overhangs resulting from the *Mlu*I and *Bam*HI digests were removed by digestion with mung bean nuclease. This gave rise to two blunt ended fragments carrying the gene for ETI both of which start at their 5′ ends with the codon for Ser. Success in altering the 5′ end of the gene with oligos-Val and -Asp was

determined by initially establishing the presence of the *Mlu*I site and then by sequencing.

3.4.3 Construction of pET 12a/ETI fusions

Plasmid pET12a was digested with *Sal*I which cuts between Ala and Ser codons at the end of the *ompT* coding sequence shown in Figure 3.13. The *Sal*I overhangs were removed from the linearized plasmid with mung bean nuclease to produce a linearized blunt ended plasmid which had lost the coding sequence for Ser required by signal peptidase to cleave off the ompT leader peptide on export to the periplasm. To this was ligated one of the blunt ended ETI-carrying fragments described above. In the correct orientation, the ETI fragment supplied the Ser codon lost on blunt ending linearized pET12a. Initially the ligation of a single insert and its orientation was determined by restriction digests and this was confirmed by sequencing the selected clones. Sequencing primers used to prime outside the ETI sequence were synthesized for the purpose and called pET12a.F01 of sequence: 5′-CCTAGGAATAGTCCTG-3′ which primes ± 30 base pairs 3′ of the *Sal*I site in pET12a and pET12a.R01 of sequence: 5′-GCTCAGCGGTGGCAGC-3′ which primes about 50 base pairs 5′ of the *Sal*I site in pET12a. Internal primers used for this sequencing are those used for sequencing ETI as discussed previously. The two clones of confirmed sequence were called pPETI.1 from oligo-Val (ie the sequence required to produce natural ETI sequence) and pPETI.2 from oligo-Asp (ie the sequence to produce ETI with Asp N-terminus).

3.4.4 Determination of the appropriate strain for expression

To test for toxicity of the expression product of pPETI.1 and 2, pET expression host strains were transformed with equivalent amounts of the two plasmids. The results are shown in Table 3.1. If one or both forms of ETI were innocuous to the hosts all seven strains would transform with equal frequency. At some level of toxicity, the DE3 lysogens (where the basal level of T7 lysogen is highest) would fail to be transformed while the lysogen containing pLysS would transform at the same frequency as the parental host. At higher levels of toxicity, even pLysS lysogens would not transform and pLysE lysogens may be needed.

STRAIN	# Transformants per PLASMID *	
	pPETI.1	pPETI.2
HMS174	200	155
HMS174(DE3)	0	5
HMS174(DE3)pLysS	195	150
HMS174(DE3)pLysE	198	154
BL21	201	156
BL21(DE3)	1	0
BL21(DE3)pLysS	200	160

Table 3.1 Numbers of transformants per pET expression strain for ETI plasmids pPETI.1 and pPETI.2 in order to determine stability of the plasmids in these strains. * Numbers reported above are the average for three transformations.

The results shown in Table 3.1 indicate that the product of both plasmids was somewhat toxic to the *E.coli* hosts as both plasmids failed to transform DE3 lysogens. It may be that it is not ETI itself but the large volume of mRNA or expressed product that is harmful to the cells. However, it was clear that T7 lysozyme would be required to dampen the basal activity of T7 polymerase and that a pLysS lysogen could be used as these gave equivalent numbers of transformants to pLysE lysogens. Nonetheless, there was no clear indication from these results as to which strain, *E.coli* HMS174 or BL21 would be best for this work. To answer this question, HMS174(DE3)pLysS, pLysE and BL21(DE3)pLysS were transformed with pPETI.1 and grown in 20 ml liquid cultures and induced at the appropriate time. Samples taken just prior to and 3 hrs after induction were lysed and subjected to electrophoresis and the results shown in Figure 3.14.

Examination of lanes 2, 3 and 4 of Figure 3.14 shows an accumulation of two bands in each of these in comparison with lane 5, the uninduced BL21(DE3)pLysS. The lower of the two bands comigrates with natural ETI and it was concluded that this represents full length ETI from which the signal sequence had been cleaved. The slower of the two bands was believed to be ETI plus the secretion signal sequence. In lanes 2 and 4 this upper band is the heavier of the two which indicates that the cell's export machinery is unable to cope with the rate of production of the protein. Similar sample volumes were collected for the three cultures shown in Figure 3.14 and each was treated in the same manner and similar volumes added to each lane. Thus it was concluded that the fact that lane 4

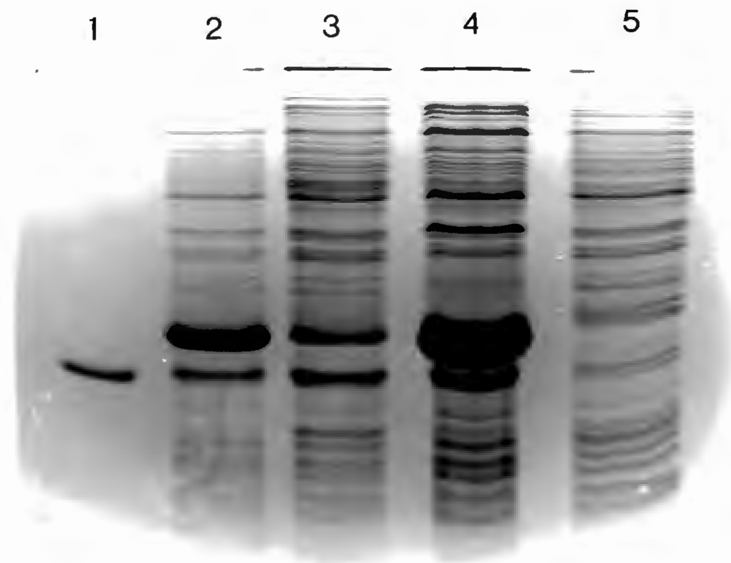


Figure 3.14 Determination of stability of pPETI.1 in different hosts.

Lane 1) Purified natural ETI
2) *E.coli* HMS174(DE3)pLysS(pPETI.1) induced
3) ------- do ------ pLysE(pPETI.1) induced
4) *E.coli* BL21(DE3)pLysS(pPETI.1) induced
5) ------------ do -------------- uninduced

contains a greater amount of both forms of ETI than either of the other two lanes indicated that BL21(DE3)pLysS is the most suitable host for expression of the ETI gene of pPETI.1.

To confirm the ability of this host to express the ETI-encoding gene of pPETI.2 a culture of BL21(DE3)pLysS(pPETI.2) was grown and induced along with the pPETI.1 containing BL21(DE3)pLysS strain described above. Samples were taken and processed as described before and electrophoresed, the results of this are shown in Figure 3.15. From lane 3 of this gel it is clear that asp-ETI, the product of the ETI gene of pPETI.2 is produced to much the same level as val-ETI, the product of pPETI.1 (lane 5). No asp-ETI or val-ETI is produced prior to induction as can be seen in lanes 2 and 4. In addition, both forms of asp-ETI (lane 3) migrate slightly ahead of the two val-ETI bands (lane 5) and was accredited to the increased negative charge of asp-ETI which is contributed by the Asp residue.

3.4.5 Expression, cleavage and purification of ETI from pPETI.1 and pPETI.2

Large scale cultures (1 litre) of pPETI.1, pPETI.2 and pET 12a containing strains were grown, the cells harvested and sphaeroplasted to release the periplasmic proteins which were collected in a final volume of 240 ml of sphaeroplasting solution. The pET12a containing strain was included as a control for background inhibitor activity. To concentrate the periplasmic proteins, the solution was 80% saturated with $(NH_4)_2SO_4$ and the precipitated proteins collected by

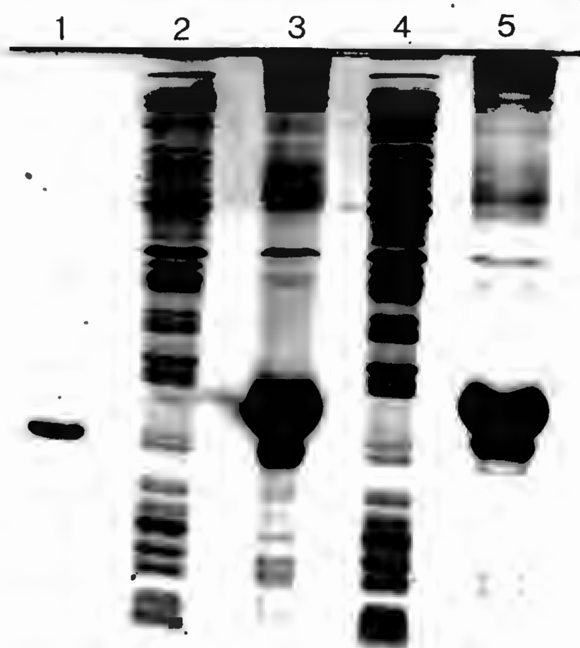


Figure 3.15 Expression of ETI from pPETI.1 and 2 in BL21(DE)pLysS

Lane	
1)	Purified natural ETI
2)	*E.coli* BL21(DE3)pLysS(pPETI.2) uninduced
3)	------------- do --------------- induced
4)	*E.coli* BL21(DE3)pLysS(pPETI.1) uninduced
5)	------------- do --------------- induced

centrifugation. The pellet was resuspended in sterile distilled water (sddw) and the $(NH_4)_2SO_4$ removed by dialysis. A light precipitate was formed after about 24 hrs of dialysis and was removed by centrifugation for later examination. The cleared dialysate was freeze-dried and resuspended in 3 ml of saline from which samples were taken for PAGE electrophoresis The results of this are shown in Figure 3.16.

Figure 3.16 shows that ETI has been recovered and is to be found in the cleared dialysate (lanes 3 and 5). There is very little of either form of ETI found in the precipitate that forms on extended dialysis. From the samples of this precipitate that were run (lanes 2, 4 and 6), it is clear that a number of contaminating proteins have been removed from the solution when comparing these with samples of the cleared dialysate (lanes 3, 5 and 7).

Cleared dialysate derived from pPETI.1, pPETI.2 and pET12a carrying strains as described above, was incubated with CNBr under the appropriate conditions to induce cleavage at the carboxy-side of Met residues. After the required period for cleavage, the solution was diluted, freeze-dried and redissolved in 4 ml of 100 mM Tris pH8. Each was loaded onto a Sephadex G50 column and eluted with 100mM Tris pH8. Fractions were collected and the presence of the respective ETI established by PAGE and these were pooled (equivalent fractions of the control were also pooled), freeze-dried and redissolved in 100mM Tris at approximately 4 mg/ml in a total of 250 μl for ETI solutions and 50 μl for the control. From more accurate concentration determinations made in biuret, it was established

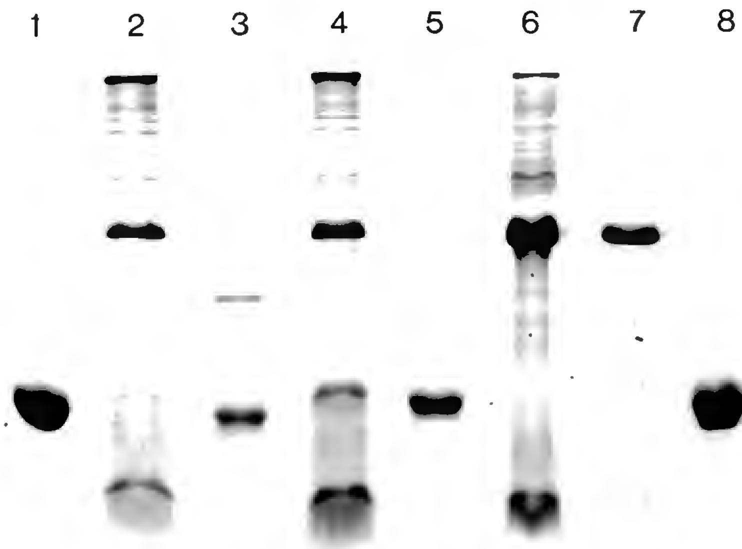


Figure 3.16 **Examination of the fractions collected during precipitation and dialysis of periplasmic proteins of BL21(DE3)pLysS(pPETI.1 and .2).**

Lane 1) Purified natural ETI
2) Pellet of precipitate formed on dialysis - pPETI.2 strain
3) Cleared dialysate - pPETI.2 carrying strain
4) Pellet of precipitate formed on dialysis - pPETI.1 strain
5) Cleared dialysate - pPETI.1 carrying strain
6) Pellet of precipitate formed on dialysis - BL21(DE3)pLysS(pET 12a) control
7) Cleared dialysate - BL21(DE3)pLysS(pET 12a) control
8) Purified natural ETI

that the concentration of protein in val-ETI, asp-ETI and control solutions was 3.64, 4.32 and 4.4 mg/ml respectively. A 1μl sample of asp-ETI, a 1.5 μl sample of val-ETI and 15 μl of control was run on the gel shown in Figure 3.17.

Lanes 5 and 7 of this gel suggest that the upper bands seen in the uncleaved supernatant material (lanes 4 and 6) have been removed. This may be as result of CNBr cleavage having reduced them to a small enough size to be separated from ETI on gel exclusion chromatography. Lane 3 was loaded with cleaved and treated control and appears to have lost most bands seen in lane 2, the untreated control material. This indicates that a great deal of contaminating host and vector proteins that have coprecipitated with ETI appear to be removed by CNBr cleavage followed by gel exclusion chromatography.

In addition, there is no apparent difference in migration between cleaved and uncleaved ETI of both forms to be seen in Figure 3.17, even though CNBr cleavage removes two amino acids from the protein. This was to be expected as 15% PAGE will not show such a small difference in size and this can be seen in the closeness of the bands produced by ETI with and without its signal sequence as seen in Figures 3.14 and 3.15.

In an effort to confirm that the bands seen in Figures 3.14 and 3.15 were correctly identified as ETI and ETI with signal sequence , western blots were performed. These compared natural ETI, total lysates of BL21(DE3)pLysS(pPETI.1 and .2) and cleaved product of cleared dialysates of pPETI.1, .2 and pET12a containing strains. The blots were probed with rabbit

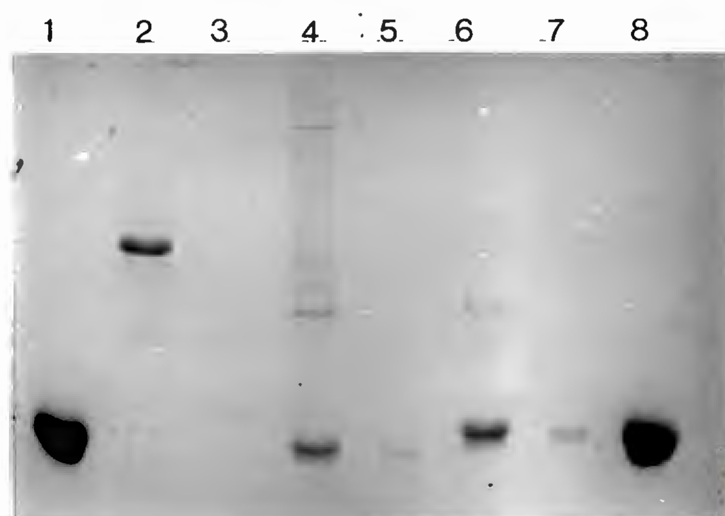


Figure 3.17 Examination of products of CNBr cleavage of proteins expressed from pPETI.1 and .2 carrying BL21(DE3)pLysS.

Lane 1) Purified natural ETI
2) Cleared dialysate - BL21(DE3)pLysS(pET12a) control
3) Cleaved product of BL21(DE3)pLysS(pET12a) cleared dialysate
4) Cleared dialysate - pPETI.2 carrying strain
5) Cleaved product of pPETI.2 cleared dialysate
6) Cleared dialysate - pPETI.1 carrying strain
7) Cleaved product of pPETI.1 cleared dialysate
8) Purified natural ETI

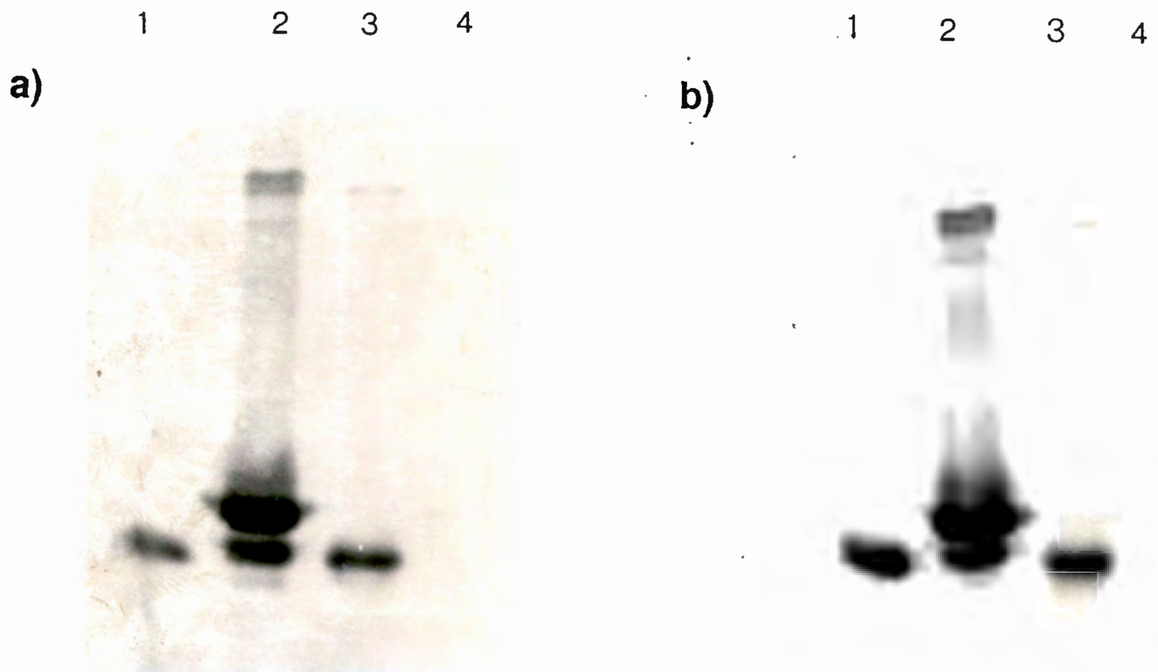


Figure 3.18 Western blots to establish composition of ETI containing preparations.

a)
Lane
1) Purified natural ETI
2) Total lysate of BL21(DE3)pLysS(pPETI.2)
3) Cleaved product of cleared dialysate of pPETI.2 strain
4) ---------------- do ------------------ control strain

b)
Lane
1) Purified natural ETI
2) Total lysate of BL21(DE3)pLysS(pPETI.1)
3) Cleaved product of cleared dialysate of pPETI.1 strain
4) -----------------do---------------------control strain

anti-ETI antibodies raised against purified natural ETI (Figures 3.18a and b) which cross react with purified natural ETI to produce a single band in lane 1. In addition, these antibodies reacted with total lysate of BL21(DE3)pLysS(pPETI.1 and .2) in lane 2 of both gels to produce two main bands. The upper band of this doublet in both gels appears in the same position relative to the lower as seen in PAGE (Figures 3.14 and 3.15) and its cross reactivity with these antibodies confirms it to be ETI with the 20 amino acid ompT signal sequence. Lanes 2 also carry two faint high molecular weight bands which are believed to be doublets of the two lower bands. Lane 3 of both gels a and b contains asp-ETI and val-ETI respectively and a faint single high molecular weight band that comigrates with the lower of the faint doublet of lanes 2 and is believed to be a doublet of the respective ETI form. Finally, lane 4 contains cleared dialysate of the control which can be seen to contain no proteins that cross react with the antibodies.

To ensure that CNBr cleavage had produced the required N-terminus of both ETI species, a sample of each (5 μl) was partially sequenced to determine the first five residues of the N-terminus. The results are presented in Tables 3.3a and b.

a)

cycle	PTH amino acid	YIELD (pmoles)
1	Val	173
2	Leu	148
3	Leu	145
4	Asp	73
5	Gly	103

b)

cycle	PTH amino acid	YIELD (pmoles)
1	Asp	149
2	Val	147
3	Leu	148
4	Leu	125
5	Asp	93

Table 3.3 Results of automatic sequencing of a) val-ETI and b) asp-ETI. Tables show amino acid residue that produces the largest peak in each cycle and the yield for that peak.

The partial amino acid sequence of cleaved product of cleared dialysate of BL21(DE3)pLysS(pPETI.1) and BL21(DE3)pLysS (pPETI.2) is shown in Table 3.3 a and b respectively. The amino acid residue with the highest yield in each cycle was accepted as the residue to be found at that position, with respect to the N-terminus, of the particular ETI under investigation. The two sequences shown were what was expected after CNBr cleavage

of the two forms of ETI isolated from the periplasm. In both sequences, yield of Ser in cycle 1 is low (<1 pmole for val-ETI and 18 pmole for asp-ETI) This indicates that most of the ETI in each preparation has been cleaved at its unique Met residue by the CNBr.

3.4.6 Discussion

The pET expression system chosen for this work has proven to be a prolific producer of ETI. It was seen that the cell's export machinery was unable to keep up with production and the majority of the expressed protein has remained in the cytoplasm of BL21(DE3)pLysS. As it was essential to show that active ETI with natural sequence could be produced from the synthetic ETI gene, the most cautious approach was adopted and only periplasmic material was used for further work. The logic here was that on export to the periplasm, the protein would be able to fold according to the dictates of its primary sequence and form correct disulphide bridges. CNBr cleavage, under the conditions used, it was reasoned, would cause only denaturation of the protein but not reduction of the disulphide bridges. Hence on removal of CNBr and formic acid, refolding would be quick and efficient. The same reasoning and approach was used to produce the altered form of ETI.

The N-terminus of the final product of both forms of ETI were sequenced and the results interpreted as showing that the proteins had been correctly cleaved and purified. The final test of the success of the approach to production of synthetic

ETI and its altered form is the assaying of these proteins. This will be discussed in Chapter 4.

Success of this approach may lead to the requirement for large amounts of either form of ETI or of some other form of the protein. In this eventuality, consideration should be given to recovery of the ETI plus signal peptide form of the protein from the cytoplasm. The signal peptide should be removed by CNBr with the same degree of success as were the Ser and Met in this work, even though the signal peptide has an N-terminal Met. Furthermore, cleaved ETI in the appropriate environment will fold into the correct form for activity which is the indication derived from the work in Chapter 3.3.

CHAPTER 4

DETERMINATION OF THE ACTIVITY OF val-ETI AND asp-ETI AGAINST t-PA

4.0 Index

4.1 Introduction 169

4.2 Results and Discussion. 179

4.2.1 Titration of inhibitor active sites with trypsin 179

4.2.2 Assays of ETI preparations for tissue plasminogen activator inhibiting activity. 182

4.1 Introduction

In Chapter 3, the expression of two forms of a synthetic gene for ETI viz, pPETI.1 and pPETI.2 were discussed. CNBr cleavage of the resultant proteins, and purification of the products gave rise to a natural form of ETI (val-ETI) and ETI with an Asp residue beyond Val at the N-terminus (asp-ETI).

It was necessary to assay first, the level of affinity of val-ETI for both trypsin and t-PA and compare these results with those obtained for natural ETI prepared from *E.caffra* seeds. Results comparable to that of natural ETI would indicate that conditions had been met for efficient and correct folding and disulphide bridge formation especially after the harsh

conditions required for CNBr cleavage. Furthermore, such results would allow comparisons to be made between the activities of val-ETI, natural ETI and asp-ETI and conclusions could be drawn on the effect of adding an Asp at the N-terminus of ETI.

To obtain this information an established assay for ETI against 2-chain t-PA was used as described by Heussen-Schemmer and Dowdle (in Press). In this report they have shown that the inhibition of t-PA by ETI at 25°C is slow and a steady state velocity in the presence of inhibitors is only attained after 15 minutes, whether the enzyme is preincubated with the inhibitor before the assay or not. Under both assay conditions, the steady state velocity of the reaction was the same, showing that ETI is a freely reversible inhibitor of 2-chain t-PA. Furthermore, in their examination of complex formation of ETI with 2-chain t-PA, Heussen-Schemmer and Dowdle (in press) found that with a t-PA:ETI molar ratio as high as 17:1, only 27% of the inhibitor was complexed with this enzyme, indicating that the complex between t-PA and ETI is relatively loose. Thus, ETI could be described by the classification of Cha (1980) as a semi-tight binding inhibitor. However, in the absence of inhibitor, these same authors have found that the steady state of the t-PA catalyzed reaction was reached so rapidly that the pre-steady state range was undetectable. Thus they reasoned that for the definition of the mechanism of the reaction and for the determination of Ki, the specific velocity equation of Baici (1981, which will be discussed) would be required.

However, ETI was found to behave differently against trypsin. Preincubation of the enzyme with inhibitor, gave a much lower rate of product release than when enzyme was added last. This is typical of an irreversible, tight-binding reaction (Heussen-Schemmer and Dowdle, in press).

Enzyme inhibitor assays essentially determine the decrease in activity of the enzyme against which the inhibitor is active. One of the prerequisites for a successful assay system is that the reaction being catalyzed should be capable of being accurately monitored. Thus Heussen-Schemmer and Dowdle (in press) have utilized the substrate methylsulphonyl-D-cyclohexyltyrosyl-glycyl-arginine paranitroanilide which is a direct chromogenic substrate for t-PA standardization described by Magnotti (1988). The action of t-PA on this substrate releases nitroanilide, the accumulation of which can be monitored at an excitation wavelength of 405mm and an emission absorbance wavelength of 680nm.

The specific velocity plot of Baici (1981) was derived using the basic model shown in Figure 4.1. In addition, to derive the equation, it was assumed that the inhibitor binds in a rapid equilibrium fashion and that the catalytic step is irreversible.

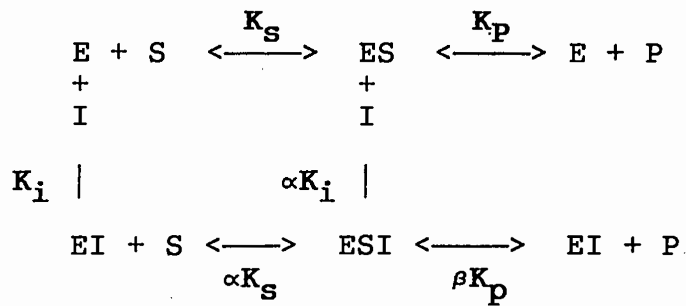


Figure 4.1 Basic Model from which specific velocity plot of Baici (1981) was derived.

E = enzyme, S = substrate, I = inhibitor,

P = product, K_p = catalytic constant

$K_i = [E][I] / [EI]$

$K_s = [E][S] / [ES]$

$\alpha K_i = K'_i = [ES][I] / [ESI]$

For these conditions, the following equation called the specific velocity equation was derived:

$$\frac{Vo}{Vi} = \frac{[I]\,(1/K'_i - 1/K_i)}{1 + \beta\,[I]/K'_i}\;\frac{\sigma}{(1+\sigma)} + \frac{1 + [I]/K_i}{1 + \beta[I]/K'_i} \qquad (1)$$

where $\sigma = [S]/K_m$ (Baici, 1981)

Measurement of the rate of steady state release of nitroanilide as a function of substrate concentration in the absence of inhibitor (Vo) allowed Heussen-Schemmer and Dowdle (in press) to calculate a K_m of 0.14 mM for 2-chain t-PA. Similar measurements of the reaction rates (Vi) of t-PA with different concentrations of ETI as a function of substrate concentration, was made. The combined data were used to construct a specific velocity plot which gave a family of curves that intersect at Vo/Vi = 1 and $\sigma/1+\sigma$ = 1 indicating that ETI functions as a

purely competitive inhibitor of t-PA (Heussen-Schemmer and Dowdle, in press).

In the case of pure competitive inhibition, $\alpha=\infty$ and $\beta=0$ and under these conditions the specific velocity equation (1) becomes:

$$Vo/Vi = - [I] \sigma / K_i (1+\sigma) + 1 + [I]/Ki \qquad (2)$$

Equation (2) simplifies to:

$$\frac{Vo}{Vi} = \frac{[I]}{Ki (1 + [S]/Km)} + 1 \qquad (3)$$

Using equation (3) and plotting Vo/Vi as a function of [I], Heussen-Schemmer and Dowdle (in press) obtained an average Ki of 2.88 X 10^{-8}M for 2-chain t-PA with pure ETI.

The reaction velocities Vo and Vi are determined by the slope, between 15 and 30 minutes, of the A405/A680 vs time plot and as the ratio of these velocities is used in equation (3), the reaction product concentration in molar terms are not required for the calculations. Furthermore, as can be seen from equation (3), knowledge of initial enzyme concentration is not required for the calculations. Nonetheless, an approximate initial enzyme concentration is required to provide conditions appropriate for analysis of the steady state kinetic data by the method of Baici (1981). These are reported by Heussen-Schemmer and Dowdle (in press) to be:

a) initial inhibitor concentration must be > tenfold larger than initial enzyme concentration;

b) substrate concentration must be large enough to ensure that less than 10% of it is hydrolysed during the reaction.

For the calculation of Ki from equation (3), accurate concentrations of substrate and inhibitor are required. Substrate is obtained in a high purity form and accurate concentrations can be made of this. However, to determine the Ki of ETI against t-PA (ie. the affinity of the inhibitor for t-PA) it is necessary to establish accurately, the concentration of active inhibitor. Concentration of active inhibitor will invariably be lower than that for total protein concentration (as determined for example by biuret reaction) which can be attributed to the presence in the sample of inactive inhibitor and other protein impurities.

In the case of ETI, it has been shown that for inhibitory activity, the overall tertiary structure is required to confer a suitable conformation on the active site loop (Heussen-Schemmer *et al.*, 1991). Any enzyme or inhibitor protein preparation will invariably be contaminated with both protein of correct sequence which is incorrectly folded and hence is inactive, and other proteins unrelated to the one of interest (Kezdy and Kaiser, 1970). This is a very important consideration when assaying val-ETI and asp-ETI both of which have been denatured by 70% formic acid and CNBr treatment and then purified. Prior to CNBr cleavage, ETI had been allowed to form disulphide bridges on export to the periplasm. It was reasoned that these bonds should remain intact in redistilled

formic acid and pure CNBr while the rest of the protein denatured. Under renaturing conditions the folding would be assisted by the disulphide bridges and provided these were of correct sequence, the protein would be likely to fold to give the tertiary structure that allows for the inhibiting activity of ETI. Nonetheless, it was expected that a percentage would not refold correctly and that in addition, unrelated contaminating protein could be expected.

Thus it was decided to determine the amount of active inhibitor in val-ETI, asp-ETI, a sample of purified natural ETI, a sample of pure STI and a control sample, by titrating active sites with trypsin. Active site titrations were suggested by Kezdy and Kaiser (1970) as a method of determining the molarity of active sites in a preparation of an active protein independently of the purity of the sample. The inhibitor active site concentration determined by titration against trypsin can be used reliably to determine the inhibitor's Ki for t-PA for the following reasons:

a) ETI reacts with trypsin in an irreversible, tight-binding manner and on a 1:1 molar ratio (Heussen-Schemmer and Dowdle, in press);

b) interaction of ETI with t-PA is dependent to a large extent on the integrity of the trypsin reactive site (Heussen-Schemmer *et al.*, 1991), ie. ETI molecules active against trypsin are likely to be active against t-PA.

As with inhibitor assays, titrations of inhibitor active sites require determination of decrease of enzyme activity. In choosing a substrate for trypsin active site titrations,

advantage was taken of the facts that serine proteases, of which trypsin is an example, can hydrolyze a wide variety of esters and in doing so form a covalent intermediate, the acyl enzyme (Chase and Shaw, 1970). Thus, p-nitrophenol-guanidino benzoate (pNPGB), an ester of p-nitrophenol and guanidino-benzoic acid, which on cleavage by trypsin releases the light absorbing or chromogenic alcohol viz p-nitrophenol, according to the scheme in Figure 4.2, is the substrate of choice for trypsin active site titrations (Chase and Shaw, 1970).

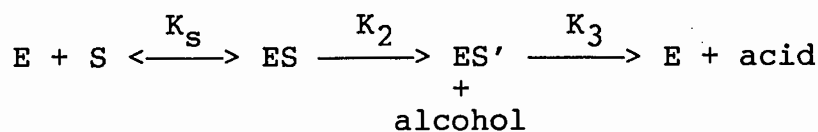


Figure 4.2 Scheme for reaction of trypsin with ester substrate such as pNPGB (from Chase and Shaw, 1970). E=enzyme, S=substrate (pNPGB), ES=enzyme substrate complex, ES'=acyl enzyme.

However, it can be shown that the active enzyme concentration is equal to the observed burst of p-nitrophenol only if $k_2 >> k_3$ and the titrant concentration is greater than the apparent Michaelis constants. Chase and Shaw (1970) have observed in their work that no effect of pNPGB on burst size is observed down to [pNPGB] = 2×10^{-6} M with trypsin.

Using pNPGB as a substrate for active site titrations of trypsin has a number of advantages. First, p-nitrophenol is an excellent leaving group and its esters, which are stable in aqueous solutions at neutral pH and room temperature, have high

values of K_2/K_3. In addition, in its ionized form (pKa = 7.04), p-nitrophenol has high absorbance at 410 nm and can be detected at 10^{-6} M with < 2% error while its ester (pNPGB in this case) has negligibly small ϵ410. Furthermore, pNPGB has sufficiently high affinity for trypsin to allow active site determinations to be independent of substrate concentration provided it is supplied in excess over [E] (Kezdy and Kaiser, 1970).

Thus it was decided to use the method described by Chase and Shaw (1970) for active site titrations of trypsin using pNPGB. The method requires that enzyme and inhibitor or enzyme alone, be incubated in a cuvette for 10 minutes at room temperature. Substrate is added to the cuvette, mixed rapidly and placed immediately in a spectrophotometer for absorbance at 410 nm (A410) to be recorded every 60 seconds for 5.5 minutes. After deduction of values obtained in a similar manner for a reference cuvette the A410 valves were plotted against Time. Extrapolation of this plot to Time=0 gives the maximum A410 of the pre-steady state or burst portion of the reaction. Subtraction of inhibited from uninhibited burst A410 and dividing this valve by ϵ410 for p-nitrophenol viz 16595, gives the molarity of the original ETI solution, from which can be deduced the specific activity of that solution.

The initial burst of product (p-nitrophenol) can be explained by the stoichiometric reaction of enzyme active sites with pNPGB to produce the acyl intermediate ES′ shown in Figure 4.2. This is followed by a slow further production of p-nitrophenol due to the following factors:

1) Hydrolysis of the acyl enzyme and reaction of the freed enzyme with a second molecule of pNPGB ie. turnover (this is small in trypsin because K_3 of Figure 4.2 is very small relative to K_2).

2) Nonspecific hydrolysis of pNGPB by other nucleophilic groups of the enzyme or buffer (ie. non-specific hydrolysis) which is small in the case of trypsin (Chase and Shaw, 1970). Nonspecific hydrolysis produced by the buffer is blanked out by addition of pNPGB to the reference cuvette and following the reaction as for the sample cuvette. However, any inequalities in addition of pNPGB will result in an erroneous rate of postburst nitrophenol production.

In addition, Chase and Shaw (1970) emphasize a number of important considerations regarding their titration technique which was utilized in this work (see Chapter 6). These are listed and discussed below.

1) The titration is carried out at pH of 8.3. This pH was selected as a compromise between the instability of both trypsin and pNPGB at higher pH and the rapid change of extinction coefficient of p-nitrophenol at lower pH. Furthermore, at pH 8.3 p-nitrophenol is approximately 92% ionized and sensitivity is near maximal. Under such conditions, only small errors are introduced by small changes of pH ($\pm$ 0.2 units) caused by the buffering effect of a buffered sample.

2) Veronal was selected as titration buffer because of the lower rate of non-enzymatic hydrolysis of pNPGB in it than in other buffers of useful capacity at this pH.

4.2 Results and Discussion.

4.2.1 Titration of inhibitor active sites with trypsin

The concentration of trypsin active sites was determined for known inhibitors: purified natural ETI and STI and for putative inhibitors: val-ETI, asp-ETI and expression control as described in Chapter 6.15.1. Release of p-nitrophenol was measured by recording the absorbance of the reaction solution at 410nm at 60 second intervals between 30 seconds and 5.5 minutes. After correction for non-active site background release of p-nitrophenol at equivalent time intervals from the start of the reaction, the A410 values were plotted as a function of Time and shown in Figure 4.3.

The plots were extrapolated to Time=0 to obtain the pre-steady state or burst A410 value. Difference between burst A410 for the pure trypsin solution and that for trypsin incubated with each inhibitor was used to calculate the concentration of active protein in μmol/L. The results are shown in Table 4.1.

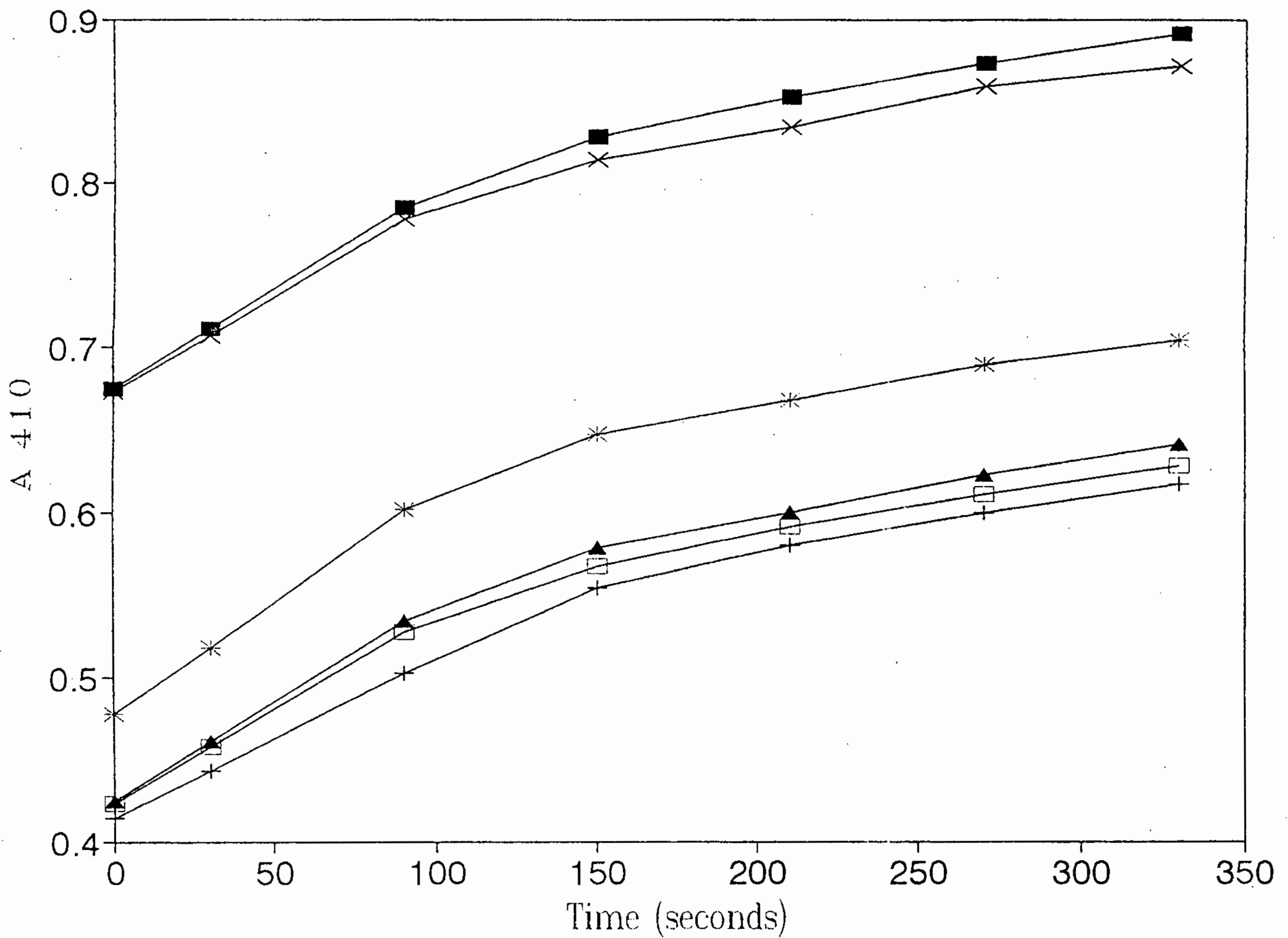


Figure 4.3 **Titration of trypsin alone and trypsin in the presence of inhibitors or putative inhibitors**

■ - Trypsin alone

× - expression control

□ - STI

✳ - val-ETI

▲ - asp-ETI

+ - Natural ETI

Sample	Burst A410	[Active site] µmol/L a	mg/ml a	[Total prot] b	Activ. c	Ac.si % d
Trypsin	0.675	406.69	9.76	11.06	36.77	88.25
Control	0.673	1.21	--	4.4	0.275	--
STI	0.423	151.83	3.05	3.56	42.65	85.67
Natural ETI	0.414	157.25	3.15	3.80	41.38	82.90
val-ETI	0.477	119.30	2.39	3.64	32.78	65.66
asp-ETI	0.425	150.63	3.01	4.32	34.87	69.68

Table 4.1 Active site concentrations determined as discussed in the text.

a. Titratable active site concentrations.

b. Total protein concentration determined in biuret reaction in mg/ml.

c. Activity in µmole/g

d. Active site concentration as a % of the total protein concentration.

The activity values shown in Table 4.1 of each test sample indicates the concentration of active protein in µmole of active protein/g of total protein (determined in biuret reaction). A sample of pure ETI should have an activity of approximately 50 µmole/g ie when 100% of the sample is active protein. However, as discussed previously, some preparations of active proteins contain contaminants and inactive protein and will have less than 100% of their total as active protein. This can be seen in Table 4.1 with the pure sample of natural ETI titrated in this work in which only ± 83% of the total is active. The expressed forms of ETI had been subjected to

denaturing conditions during CNBr cleavage followed by only partial purification. Thus the concentration of active protein in these preparations was expected to be lower than that for purified natural ETI. The specific activity of val-ETI was found to be about 33 μmole/g (about 66% active) while that for asp-ETI was 35 μmole/g (about 70% active).

From these results it can be concluded that while approximately 30% of the preparation, in the case of asp-ETI and 34% in the case of val-ETI, is inactive protein, these two forms of ETI expressed from synthetic genes do inhibit trypsin. Furthermore, considering the parallel expression, treatment and purification of val-ETI and asp-ETI, the results indicate that there is little difference between the two in their ability to bind to and inhibit trypsin. This was predicted by molecular modelling studies .

It can be seen from Table 4.1, that there is a small measure of background inhibition contributed by components of the preparation other than the expressed ETI. Nonetheless, its contribution was considered to be negligible for the purposes of this titration.

4.2.2 Assays of ETI preparations for tissue plasminogen activator inhibiting activity.

Concentration of active proteins in each preparation to be assayed against t-PA had been determined (Chapter 4.2.1). Steady state release of nitroanilide was followed by recording A405/A680 of the reaction. Reaction velocities were determined

from the slope of plots of A408/A680 as a function of time. These were determined in the absence of inhibitor (Vo) and in the presence of various concentrations of inhibitor (Vi).

The rate Vo/Vi was determined for each dilution of the various samples and plotted as a function of active inhibitor concentration, [I]. The data were fitted by linear regression to the equation (3):

$$\frac{Vo}{Vi} = \frac{[I]}{Ki\ (1+[S]/K_m)} + 1$$

The slope of the resulting straight lines were determined, from which the K_i and specific activity of the assayed samples were calculated. These are shown in Table 4.2. The specific velocity plots are shown in Figure 4.4 while the data for these plots is tabulated in Appendix A. The necessary calculations to derive Ki and specific activity are described in Chapter 6.15.2 and shown in Appendix A.

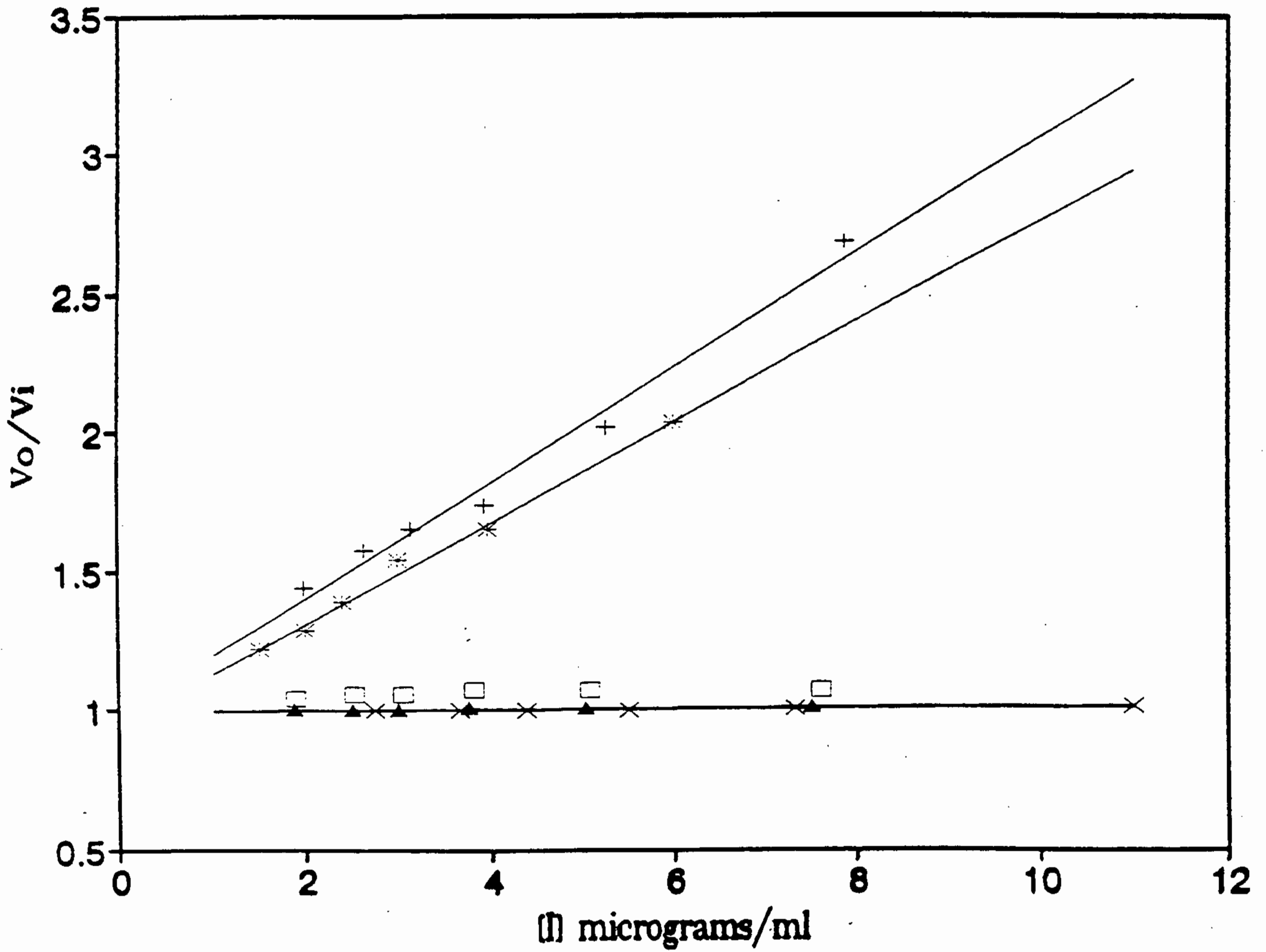


Figure 4.4 Determination of inhibition constants. Data are fitted to the simplified forms of the specific velocity equations (see text).

× - control

▲ - asp-ETI

□ - STI

✳ - val-ETI

\+ - Natural ETI.

Sample	Ki $\mu g/ml$	Ki moles/L	Specific activity units/mmole
Control	97.967	----	----
val-ETI	0.677	3.386×10^{-8}	1.701×10^{9}
asp-ETI	56.405	2.820×10^{-6}	2.040×10^{7}
Natural ETI	0.594	2.970×10^{-8}	1.939×10^{9}
SBTI	66.549	3.311×10^{-6}	1.740×10^{7}

Table 4.2 Ki's and specific inhibitory activity for expressed ETI's and controls against t-PA. One inhibitor unit is the amount of inhibitor that will inhibit one unit of enzyme activity by 50%. One unit of enzyme activity is equal to 10 pmol of substrate hydrolyzed per minute.

A sample of purified natural ETI, should have Ki of 2.88×10^{-8} M and specific activity of 2×10^{9} units/mmole (Heussen-Schemmer and Dowdle, in Press). This may reflect differences in titration and assay applications. Results obtained in this work indicate that 1.14 times more of the val-ETI preparation than natural ETI was required to inhibit the same amount of t-PA. In this case there may be substances in the preparation that interfere with the ability of t-PA to interact with ETI but do not affect t-PA's ability to cleave Spectozyme-tPA. Nonetheless these imply that val-ETI is very much like natural ETI in its ability to inhibit t-PA.

On the other hand, about 111.4 and 94.95 times more active STI and asp-ETI respectively are required than natural ETI to reduce, by 50%, the activity of t-PA in this assay. Thus, adding an Asp to the N-terminus of ETI as in asp-ETI, appears to greatly decrease its ability to inhibit t-PA. In fact, the added Asp residue has made asp-ETI very STI-like in their abilities to inhibit tPA.

In the case of the expression control preparation, Ki and specific activity are reported although it shows negligible inhibitory activity to trypsin. The purpose here is to show that this preparation also displays negligible inhibitory activity to t-PA and the constants for val-ETI and asp-ETI were thus determined with confidence, using the active site concentrations determined by titration.

CHAPTER 5

CONCLUSIONS

An ETI-encoding gene was designed using the published amino acid sequence. The gene was synthesized by serial additions of 8 oligonucleotides into a synthesis vector designed for the purpose. It was ascertained from the literature that the correct system for expression of a gene is, in most cases, largely empirically determined. In this case the synthetic ETI gene was originally designed to be placed under the control of a strong promoter in a simple expression vector, pBtac1. Met was designed to be the N-terminal amino acid in place of Val as *E.coli* requires a Met codon for initiation and it was reasoned that this amino acid could act in the same capacity as Val at this position. However, no detectable levels of expressed protein was detected from this expression system. It was suspected at this stage that a leader sequence was required to assist expression by opening the mRNA to contact with the ribosomes, or to protect against degradation of either the mRNA or the expressed protein. This was confirmed when the gene failed to express from the expression vector, pET 11a which also supplies no leader sequence. These results have not been presented in this work as they offered no more illumination than the results gained from expression of pBtac1.

In the MBP expression system, the target protein is expressed, fused at its N-terminus to MBP. This system was used very successfully to express and purify an ETI-MBP fusion protein. In this context MBP acted as a leader peptide and high levels of expression were achieved from the synthetic ETI gene.

Apparently the presence of a 5' leader sequence of *E.coli* origin has provided protection to ETI and allowed it to accumulate to high levels in the cell. However, even as part of a fusion protein, ETI was able to fold into a form enabling it to inhibit the serine protease which was to free it from MBP.

At this point it was decided to use the secretion vector pET12a to produce ETI with ompT, the periplasmic secretion signal peptide, at its N-terminus. The aim here was primarily to allow expression of the synthetic gene and secondly to promote export to an environment conducive to disulphide bridging. As the signal peptidase of *E.coli* would only cleave off the ompT peptide to leave ETI with a Ser N-terminus, a Met was engineered between Ser and the natural Val which had been inserted into its correct position. The Ser and Met were removed by CNBr cleavage to leave ETI with the same amino acid sequence as found in the natural protein isolated from *Erythrina caffra*. This was confirmed by sequencing the amino terminal region of this protein and by assaying its activity against trypsin and t-PA. However most of the expressed ETI with the signal sequence attached, was left in the cytoplasm as the cell's export machinery could not keep up with the level of expression. It was suggested that, because of the apparent ability of ETI isolated from the cytoplasm to fold correctly and form disulphide bridges, as was seen with the MBP system, the cytoplasmic ETI + signal sequence be isolated, cleaved and purified for an even greater yield of functional ETI.

The expressed ETI was called val-ETI and was incubated with trypsin, one of the proteases against which ETI is active and to which it binds irreversibly in a 1:1 molar ratio, to determine the concentration of active molecules in the sample. From this, the amount of active ETI as a percentage of the total protein was determined. An amount of 66% of val-ETI was found to be active protein compared to 83% in natural ETI. This was considered to be a comparable activity as val-ETI had undergone relatively harsh treatment and some loss of activity was expected. The concentration of active protein determined in this way was used in determining the inhibition constant of val-ETI against t-PA. At the same time, purified natural ETI (which had been donated by Seravac) was also assayed against t-PA. The Ki determined for val-ETI was found to be very similar to that determined for natural ETI. In essence this was the culmination of the synthesis of val-ETI, a project that had started at the translation of the published amino acid sequence to a DNA sequence. At the start there was some apprehension as to whether *E.coli* would be able to synthesize this eukaryotic protein, although it is a relatively simple one requiring only two disulphide bridges in a single chain protein. Nonetheless, as a form of insurance, the gene was designed to be utilized by *S.cerevisiae* as well. After engineering into the design a number of unique restriction sites, theoretical structures that the resultant mRNA could adopt were looked at. The most stable of these turned out to be a complex structure with extensive double stranded regions. This was acceptable from the point of view of protecting the mRNA from degradation. However, there was the danger of any ribosome binding site that was attached to this mRNA being caught up in the stable secondary structure

and being unavailable to the ribosomes for translation. This, of course would halt expression at this point and may have been the explanation for unsuccessful expression from pBtac1 and pET11a. Both of these vectors are for the direct expression of proteins in the cytoplasm of *E.coli*.

The remedy was cloning the synthetic gene at the 3' end of a leader sequence which allowed good levels of expression. In these cases it is believed that the leader sequence has made the sequence more available to the ribosomes.

It was suggested that a very useful ETI-encoding gene had been designed. In addition to being able to produce ETI almost indistinguishable from the natural protein, changes can easily be made to the sequence to produce altered forms of the protein as a way to investigate its function in relation to its structure. This was put to the test in altering the sequence to allow insertion in pET12a. Then an altered form of the protein was produced to address a question on the role of the Asp N-terminus of STI in its inability to inhibit t-PA.

STI is an inhibitor of trypsin but not of t-PA. ETI and STI bear a great deal of similarity in their sequence and structure. However, they differ firstly in the scissile bond which is the point of contact of the protein with its substrate, trypsin. This bond, between Arg63 and Ser64 in ETI and between Arg63 and Ile64 in STI is cleaved under the right circumstances. The other point of difference between the two proteins relates to their N-termini ETI has been shown to have a structured N-terminal region which does not interefere

with the interaction of the protein with trypsin or tPA. STI is believed to have a similar structure to its N-terminal region, by virtue of sequence homology with ETI and three-dimensional structure overlay. STI has an N-terminal Asp residue which, because of its hydrophilic nature does not participate in the structure of the N-terminal region. Furthermore, while this Asp residue participates in STI:trypsin complex formation by ion pairing with the side chain of Lys60 of trypsin, it interferes with the interaction of STI and tPA either sterically or possibly by charge repulsion.

For this reason, the N-terminus of val-ETI was extended by an Asp to produce asp-ETI. This new form of ETI was seen to have activity towards trypsin similar to that of val-ETI but to have very little activity against t-PA as compared with val-ETI. This result supports the hypothesis, based on computer modeling that the Asp residue of STI prevents docking of this protein with tPA, thereby thwarting inhibitor:protease complex formation.

Another inhibitor, WTI was discussed briefly and seen to be more like ETI in its primary structure than STI but no crystal structure has been determined for this protein. As with STI, WTI is an inhibitor of trypsin but not of tPA. Based on sequence homology with ETI and STI, WTI appears to have the same N-terminal structure as that of STI. The implication from this is that its N-terminal Glu is responsible for the inability of this protein to inhibit tPA.

However more work is required to understand the unique ability of ETI to inhibit t-PA. For example, the scissile bond of val-ETI could be altered to resemble that of STI and the effect on its activity assayed. The synthetic ETI gene was designed with such a use in mind and as an ETI very similar in activity to natural ETI was produced, such changes can be made and their effects on activity compared reliably with those of val-ETI.

CHAPTER 6

MATERIALS AND METHODS

6.0 Contents

6.1 Extraction and purification of plasmid DNA 195

6.1.1 Small scale preparations (Miniprep) 195

6.1.2 Large scale lysis by alkali 196

6.2 Plasmid digests 197

6.3 Gel electrophoresis 198

6.3.1 Agarose 198

6.3.2 SDS Polyacrylamide gel electrophoresis (SDS PAGE) 198

6.4 Purification of digested DNA fragments from agarose gels 198

6.5 Cloning 200

6.5.1 Ligations 200

a) Ligation of cohesive termini 200

b) Ligation of blunt-ended DNA 200

6.5.2 Phosphatase treatment of vector DNA. 200

6.5.3 Mung bean nuclease removal of overhangs from digested DNA. 201

6.5.4 Transformation 201

a) Competence induction. 201

b) Transformation. 202

6.6 DNA sequencing 202

6.7 Photographic techmiques 203

6.7 Oligonucleotide synthesis 203

a) Synthesis. 203

b) Purification 203

6.8 Gene synthesis 205

6.9 Gene expression 205

6.10 The Maltose binding protein purification system 206

6.11 Detection of expressed ETI with anti - ETI antibodies 206

a) Production of antisera. 206

b) Western blots. 207

6.12 Sphaeroplasting E.coli BL21(DE3)pLysS (pPETI.1 or.2) cells to release periplasmic proteins 208

6.13 Concentration of periplasmic proteins released by sphaeroplasting 209

6.14 Cyanogen bromide cleavage of proteins extracted from the periplasm 210

6.15 Inhibitor assays 210

6.15.1 Active site titratiion 210

6.15.2 Assays of inhibitory activity for tissue plasminogen activator 213

6.16 Amino acid sequence analysis 216

6.1 Extraction and purification of plasmid DNA

The method used in this work was adapted from the alkaline lysis method of Sambrook *et al.* (1989).

6.1.1 Small scale preparations (Miniprep)

a) A single colony of the plasmid containing culture was grown overnight in 8ml of Luria broth (LB) (Appendix B) in a sterile McCartney bottle with the appropriate antibiotic(s).

b) Five 1 ml aliquots of this culture was centrifuged in a 1 ml microfuge tube and resuspended in 100 μl of ice cold Solution 1 (Appendix B) by vortexing and pipetting vigorously.

c) To this was added 200 μl of room temperature Solution 2 (Appendix B) and mixed thoroughly by hand until cleared then stood on ice for 5 minutes.

d) 150 μl of ice cold Solution 3 (Appendix B) was added and vortexed vigorously whereupon a white flaky precipitate was produced.

e) This was centrifuged in a microfuge for 5 minutes at room temp. and the supernatant transfered to a fresh tube

f) To the supernatant was added 450 μl of equilibrated phenol (Appendix B), vortexed thoroughly and centrifuged for 2 minutes and the upper aqueous phase recovered.

g) This was repeated with a 1:1 mix of phenol and chloroform, chloroform and 900 μl of ether. The aqueous phase after ether extraction was found below the less dense ether.

h) Plasmid was precipitated by addition of 500 μl of absolute ethanol (-20°C), centrifuged for 2 min. and the pellet washed with 1 ml 70 % ethanol then left to air dry briefly and dissolved in 15 μl TE (Appendix B) with RNAse A (50 μg/ml) (Appendix B).

6.1.2 Large scale lysis by alkali

a) A single colony of the required plasmid containing strain was grown overnight in 500 ml of LB in a 2 L flask with the appropriate antibiotics and vigorous shaking.

b) The cells were harvested by centrifugation at 7000 rpm in a JA10 rotor of a Beckman J2-21 centrifuge for 10 minutes at 4°C and the pellet resuspended in 3.6 ml of Solution 1 as in 6.1.1.

c) 8 ml Solution 2 was added and mixed gently but thoroughly until clear then stood on ice for 5 min.

d) To this was added 4 ml of solution 3 and mixed vigorously. The solution was cleared of precipitate by centrifugation at 17 000 rpm in a JA20 rotor for 20 min.

e) The plasmid was precipitated, centrifuged, air dried briefly and dissolved in 4 ml TE. 4.5 g of CsCl was added and dissolved and 150 μl of EtBr 10 mg/ml added.

f) The preparation was centrifuged at 55 000 rpm at 20°C for 16 hrs. The resultant plasmid band was extracted, redissolved and centrifuged for a second time.

6.2 Plasmid digests

High, medium and low salt buffers supplied by Boehringer Mannheim as a 10 x stock were used for most digests except for *Bsm* I which was supplied, by New England Biolabs with a 10 x stock buffer. Between 1 and 2 μg of plasmid DNA were digested in volume of 20 - 30 μl with 1 -3 U of enzyme per μg of DNA. Digests were performed at the optimal temperature for the particular enzyme for 5 hrs in the case of *Bsm* I and for 2 hrs for all others. The reaction tube was incubated in a heating block set at the required temperature and incubated in oven, set at the same temp., to prevent condensation from forming on those parts of the reaction tube not in direct contact with the heating block.

In the case of double digests, a sample of the digest mix was electrophoresed to determine complete digestion, in which case the digest was treated with phenol, phenol/$CHCl_3$, $CHCl_3$ and ether and precipitated with LiCl as before. The resultant digested plasmid pellet was dissolved in a 1 x strength buffer required for the 2nd enzyme and digested as before.

6.3 Gel electrophoresis

6.3.1 Agarose

DNA preparations were routinely analysed by agarose gel electrophoresis, using a flatbed apparatus and 1% agarose in TBE buffer (Appendix B), pH 8 containing ethidium bromide (EtBr) at 0.5 μg/ml. The running buffer (TBE) in which the gel

was submerged also contained EtBr at the same concentration. Gel loading buffer containing sucrose and bromophenol blue as a tracking dye was added to samples before loading and electrophoresis was carried out at 4-8 volts/cm for 1-2 hrs.

6.3.2 SDS Polyacrylamide gel electrophoresis (SDS PAGE)

All cell extracts and purified proteins were electrophoresed on 15 % acrylamide SDS PAGE (for recipes see Appendix B). 14 cm long gels were run consisting of a ± 2 cm stacking gel and 12cm of separating gel. Samples were dissolved by boiling for 2 min. in the case of pure samples and 5 min. for cell lysates, in an appropriate volume of SDS sample buffer containing 20 μl/ml of 2-mercaptoethanol. For cell lysates, cell were first harvested then dissolved in the lysis buffer. Gels were run at 10 mA until the bromophenol blue of the loading buffer had entered the stacking gel then current increased to 15 mA.and run until the bromophenol had just run off the bottom of the gel.

Gels were submerged in Page Blue 83 stain (45% methanol, 45% H_2O, 10% acetic acid, 0.2% Page Blue) for 1 hr at 37°C with gentle shaking and destained in 20% acetic acid until the background had cleared.

6.4 Purification of digested DNA fragments from agarose gels

There are many good methods available for this purpose, the one found best for this work was the Freeze squeeze method described by Antoniw (1985).

a) The digest reaction mix was loaded into a preparative well of a 1% agarose gel with no EtBr in the gel or the running buffer and electrophoresed as before.

b) The gel was stained in 0.5 μg/ml EtBr in water for 10 min. in the dark. DNA bands were observed with long wavelength (302 nm) UV and cut out of the gel.

c) Gel slices were incubated in ± 500 μl of Freeze squeeze buffer in the dark for 30 min. with occasional inverting.

d) Small (0.5 ml) polypropylene microfuge tubes each had a hole pierced at the bottom with a wide gauge needle. The hole was plugged with a small amount of siliconized glass wool and the gel slice placed on this and the tube closed.

e) The tube with its gel slice was dropped into liquid N_2 and left for ± 5 min. then placed inside and hanging on the lip of a 1 ml microfuge tube from which the lid had been cut.

f) This was centrifuged in a microfuge for 5 min. at 4^oC then the gel refrozen and respun.

g) The eluate was vortexed with an equal volume of phenol, then phenol/$CHCl_3$, $CHCl_3$ alone and two volumes of ether.

h) DNA was precipitated with two volumes of absolute ethanol after addition of 1/10th final volume of 4 M LiCl, washed with 70% ethanol and dissolved in an appropriate volume of TE.

6.5 Cloning

6.5.1 Ligations

a) **Ligation of cohesive termini** (according to Sambrook *et al.*, 1989). 0.1 μg of digested and phosphatased vector DNA was added to equimolar, 2 x equimolar, 3 x equimolar and 4 x equimolar insert DNA. Water was added to 7 μl and the solution warmed to 45°C for 5 min to melt any cohesive termini that have reannealed, then incubated on ice until the start of the reaction. To this was added 1 μl each of 10 x bacteriophage T4 DNA ligase buffer, 5 mM ATP and bacteriophage T4 DNA ligase at 1 U/μl. The reaction was incubated at 16°C for 1 - 4 hrs or at 14°C for 16 hrs. 2 μl of this was used to transform competent *E.coli*.

b) **Ligation of blunt-ended DNA** (adapted from Sambrook *et al., 1989)*. Higher concentrations of both vector and insert DNA were required for successful ligation. 100 μg of phosphatase treated vector and an equimolar amount of insert were mixed on ice with water and 1.5 μl of 10 x ligase buffer, 2.5μl of 5mM ATP. To this was added 5 μl of 60% PEG 6000 at room temp. and 1 μl of T4 DNA ligase at 15 U/μl and incubated at 20°C for 16 hrs. This was used to transform competent *E.coli*.

6.5.2 **Phosphatase treatment of vector DNA** (Sambrook *et al.*, 1989).

Digested or mung bean nuclease treated plasmid was phenol treated and precipitated as described above and dissolved in calf intestinal phosphatase or CIP buffer (Boehringer

Mannheim). The DNA was treated with CIP at 2 units per pmole of termini. For 5′ overhangs, the reaction was incubated at 37°C for 30 min. then a further aliquot of CIP was added and incubated for a further 30 min. For blunted DNA, after the 30 min. incubation the reaction was incubated for a further 45 min. at 55°C

6.5.3 Mung bean nuclease removal of overhangs from digested DNA (New England Biolabs 1988-1989 catalogue).

The digested DNA was phenol treated and precipitated as before and dissolved in mung bean nuclease MBN buffer at a concentration of 1 μg of DNA per 10 μl of buffer. To this was added 0.5 units of MBN per μg of DNA and incubated at 30°C for 30 min after which it was made 0.2 M in NaCl.

6.5.4 Transformation

a) **Competence induction.** Fresh overnight culture in LB with the appropriate antibotics (25 μg/ml of chloramphenicol for BL21(DE3)pLysS) from single colony off a fresh plate was grown. 1 ml of this was added to 100 ml of LB + antibotic if necessary and grown to A550 of ± 0.6. Cells incubated on ice for 15 min. 40 ml of this was centrifuged at 5000 rpm in JA20 rotor of a Beckman J2-21 centrifuge at 4°C for 5 min and immediately dissolved in 8 ml of 0.1 M $MgCl_2$ at 4°C. This was recentrifuged as above and resuspended in 8 ml 0.1 M $CaCl_2$ at 4°C and kept on ice for 1 hr then harvested as before and redissolved in 1.6 ml $CaCl_2$ at 4°C then kept on ice.

b) **Transformation.** Ligation solution was made up to 100 μl with water and incubated on ice in polypropylene microfuge tube. To this was added 200 μl of competent cells and stood on ice for 15 min. after which transferred directly to a waterbath at 42°C for 3 min. then back to ice for 5 min. To this was added 0.8 ml of LB and incubated at 37°C with shaking for 1 hr. Cells were spun down and resuspended in 200 μl of LB of which all or dilutions thereof were plated on appropriate plates which had been dried in an incubator and after plating the plates were dried for 10 min. without lids in the incubator.

6.6 DNA sequencing

DNA was sequenced using a SEQUENASE ^{TM} sequencing kit supplied by United States Biochemical according to the manufactures instructions. In all cases double stranded DNA was used as template. 10 μg of template DNA (2 x CsCl gradient purified) was diluted to 18 μl with sterile double distilled water and 2 μl of 2 M NaOH added. This was incubated at 37°C for 5 min. The solution was neutralized and the DNA precipitated by addition of 150 μl of absolute ethanol and 4 μl of 3 M Na-acetate and centrifuged in a microfuge for 20 min. The resultant pellet was washed with 70% ethanol then air dried. The DNA was then ready for priming for which the manufacturers protocol was followed.

6.7 Photographic techmiques

SDS PAGE gels stained with Page blue - 83 were photographed on a light box with Ilford Pan - F film and a red filter. Film was developed using Agfa Rodinal according to the manufacturers instructions.

6.7 Oligonucleotide synthesis

a) **Synthesis.** Oligonucleotides were synthesised with nucleoside-*O*-(2-cyanoethyl)-N,N-diisopropyl phosphoramidites. These and other reagents for synthesis were from Biosearch. Synthesis was performed on a model 6500 Autogen^{TM} DNA synthesizer in a glass column containing 1 μmole (for the long synthesis oligos) and 0.2 μmole (for short oligos eg. primers) of an appropriate 5'-*O*-dimethoxytrityl deoxynucleoside-3'-*O*-succinyl-derivatized 1000 Angstrom control pore glass support at a loading of 18 μmol/g.

b) **Purification** (Horn and Urdea, 1988).

i) After synthesis, the resin was dried, transfered to a microfuge tube and washed with 1 ml CH_3CN then air dried.

ii) Resin was incubated for 1 hr at room temp. in 1 ml tri - butylamine/pyridine (1:9 v/v).

iii) Tri - butylamine/pyridine was removed and the resin washed 5 x with CH_3CN then air dried.

iv) 1 ml of lysine pH 9 was added to the resin and incubated at 60°C for 1½ hrs.

v) Lysine was removed and the resin washed 2 x with distilled water and 3 x with CH_3CN.

vi) Resin was transfered to a glass vial, air dried and 1 ml concentrated NH_4OH added, vial closed immediately and incubated for 1 hr at room temp. then at 60^oC for 24 hrs.

vii) Solution was cooled, ammonia allowed to evaporate and the solution added to 1 ml 0.1 M Tris pH 9.5 and 8 ml TNE (Appendix B) in glass scintillation vial.

viii) A disposable rapid reverse phase phenyl silica cartridge (Baker - 10 SPE ^{TM}, Baker Chemical Co., Phillipsburg,NJ) was equilibrated with 10 ml methanol, followed by 10 ml 50% v/v methanol/100mM TEAA followed by 10 ml of 20 mM TEAA.

ix) The oligo. containing solution was loaded onto the equilibrated column and washed with 5 ml of 20 mM TEAA followed by 10 ml of 30% v/v methanol/100mM TEAA.

x) The oligo was eluted in 4 ml 75% v/v methanol/100mM TEAA and evaporated to dryness.

xi) The residue was dissolved in 100 μl of 80% aqueous acetic acid in which it was incubated for 30 min. then evaporated to dryness, dissolved in 1 ml of 5 x TE pH 7.4 and centrifuged to remove any precipitate. The concentration was determined from the A260 value.

6.8 Gene synthesis

The gene for ETI was synthesised by the method of oligonucleotide-directed double-strand break repair established by Mandecki (1986). Three aliquots of 50 ng of plasmid linearized with *Bsm* I restriction enzyme was mixed with complementary oligonucleotide at 10 fold, 100 fold and 1000 fold molar excesses. To this was added 3 μl of 10 x concentrated denaturation buffer (Appendix B) and the solution made up to 30 μl with water. The mixture was boiled for 3 min. and allowed to cool for 5 min at room temp. on a polystyrene block. The cooled mixture was centrifuged to collect condensate and placed on ice. To this was added 200 μl of competent *E.coli* JM83 [*ara*, Δ(*lac-pro)*, *str*A, *thi*,Φ80d *lac*Z ΔM15]. Transformed cells were plated on LB containing 100 μg/ml of ampicillin and 20 μg/ml of 5-bromo-4-chloro-3-indolyl β-galactoside (X-gal). Transformants were selected according to the appropriate change in colour from that of the parental on the same medium.

6.9 Gene expression

All strains containing a gene to be expressed were grown at 37°C with shaking to A550 of 0.6 - 0.8 in LB with the appropriate antibiotic(s). Cells were harvested at 5000 rpm in a JA20 rotor under sterile conditions and resuspended in fresh, sterile LB at 37°C, containing antibiotic then induced by making the culture 0.4mM in isopropyl-β-D-thiogalactoside (IPTG) and shaking for a further 3 hrs at 37°C. In all cases ampicillin was required in the culture medium and futher aliquots of this were added, in addition to the initial 100

μg/ml, every hour after the start of growth and after induction. This practice and that of harvesting and resuspension of cells was adopted in an effort to prevent plasmidless cells from outgrowing plasmid containig cells in a medium depleted of ampicillin by secreted β-lactamase.

6.10 The Maltose binding protein purification system

Two 500ml cultures of *E.coli* TB1 (pMETI) were grown to mid log phase and induced with IPTG. After 3 hrs of growth, the cells were harvested and lysed as described earlier. The supernatant of the lysate was found to be about 20 mg/ml in total protein and was diluted to about 2.5 mg/ml. Cross-linked amylose resin (1.5 g) supplied by NEB, was swollen in 50 ml of column buffer and packed into a 50 ml syringe to act as a column. Diluted supernatant was loaded onto the amylose resin which was then washed with 50 ml of column buffer + 0.25% tween and then with 80 ml of column buffer with no tween. Bound MBP-ETI fusion protein was eluted in about 300 ml of column buffer containing 10 mM maltose and collected in approximately 10 ml fractions. Samples of the fractions (20 μl) were pooled 5 at a time and a 10 μl sample from each pool run on a polyacrylamide gel.

6.11 Detection of expressed ETI with anti - ETI antibodies

a) **Production of antisera.** Purified natural ETI (a donation from Serevac) was dissolved at 200 μg/ml in phosphate buffered saline (PBS - see Appendix B) and emulsified with an equal volume of Freunds incomplete adjuvant. Rabbits were primed with three intramuscular injections of 1 ml at

1 week intervals and boosted thereafter at 4 - 6 weeks over several months. Rabbits were bled from a marginal ear vein at weekly intervals starting 3 weeks after the first injection. The serum or primary antibody fractions (10 - 20 ml) were stored separately at - 20°C and titred to determined which had the highest anti-ETI antibody concentration. Dilutions (10 times from 10^{-1} - 10^{-8} in Tris buffered saline with 0.1% Tween 20) were made of each primary antibody fraction, and incubated with ETI (diluted in PBS) in an ELISA plate reaction. Reaction with the non antigenic molecules in the antigen solution was blocked with 3% BSA in TBS/Tween. Anti-ETI antiboby remaining bound to ETI in the wells was detected with goat-anti-rabbit alkaline phosphatase conjugate. The reaction was developed with 1 mg/ml of nitrophenylphosphate in 10% diethanolamine, pH 9.8.

b) Western blots. Gels run as described before and transfered electrophoretically to Hybond^{TM} as described in Ausubel (1987). The nitrocellulose was blocked in 5% milk powder buffer and was incubated, 1hr at room temperature, with antibody diluted 10^{-5} in TBS/Tween at 100 μl/cm^2 of nitrocellulose. This was washed 2 x 10 min each in TBS/Tween. The washed nitrocellulose was incubated further with diluted (1/1000) goat-anti-rabbit alkaline phosphatase conjugate (made in the Department of Biochemistry, UCT) at 100 μl conjugate/cm^2 at room temp. for 30 min. Nitrocellulose was washed in TBS/Tween and then in Developing buffer (Appendix B). The reaction was developed in BCIP, NBT developing solution until the

solution took on a purple tinge or until the background became dark in colour then washed in TBS/Tween.

6.12 Sphaeroplasting *E.coli* BL21(DE3)pLysS (pPETI.1 or.2) cells to release periplasmic proteins

The method of Marston (1987) was adapted and used to release periplasmic proteins separate from cytoplasmic proteins.

a) Cells were harvested from two 500 ml cultures 3 hours after induction by centrifugation at 5000 rpm in a JA10 rotor for 5 minutes in a Beckman J2-21 centrifuge.

b) The entire pellet from each 500 ml culture was resuspended in 20 ml Tris pH 8 then pool both of these.

c) Immediately added 40 ml 200mM Tris pH 8 , 1M sucrose and mixed very gently.

d) As soon as possible 40 ml of 100mM EDTA was added and mixed very gently.

e) To this 120 μl of 10 mg/ml lysozyme was added in 200mM Tris pH 8 and mixed very gently again.

f) Immediately after added 120 ml sterile double distilled water (sddw) and mixed gently.

g) 20 μl sample was taken, added to 980 μl sddw and the absorbance at 450 nm monitored, sphaeroplasting was seen to be complete as A_{450} ceased to decrease. This took about 22 minutes.

h) Mg^{2+} was added 20 mM ie 20 ml of 1 M $MgCl_2.6H_2O$, to stabilize sphaeroplasts.

i) Sphaeroplasts were harvested at 4000 rpm for 5 minutes in a JA20 rotor at room temp.

j) Supernatant at final volume of 240 ml, which contained periplasmic proteins, was poured off and stored on ice.

6.13 Concentration of periplasmic proteins released by sphaeroplasting

The expressed ETI's were dissolved in a volume of 240ml and hence too dilute to for further work. It was thus decieded to concentrate the periplasmic proteins by precipitation with $(NH_4)_2SO_4$. It was previously established that 80% saturation of samples of the harvested periplasmic solution precipitated a fraction of proteins that appeared to contain all of the expressed ETI's. The precipitation was performed as follows:

a) The solution was made up to 250ml in a 500ml beaker with sddw and cooled on ice.

b) 145g of $(NH_4)SO_4$ was added and stirred on ice until all salt had dissolved.

c) Precipitated protein was collected by centrifugation at 10000rpm in a JA10 rotor in a Beckman J2-21 centrifuge for 15 minutes at 0^oC.

d) The pellet was resuspended in 8ml of sddw and dialysed against running water for about 24hrs.

e) The light precipitate which formed at about 24hrs of dialysis was removed by centrifugation at 15 000rpm in a JA20 rotor at 4^oC for 15 minutes.

f) Freeze dried the supernatant and resuspended in 3ml of saline.

6.14 Cyanogen bromide cleavage of proteins extracted from the periplasm

a) To 2.6 ml of each solution of concentrated periplasmic proteins, in a 20 ml McCartney bottle, was added 0.4 ml of sterile distilled water (sddw) and 7 ml of redistilled formic acid.

b) To this was added 60 mg of recondensed CNBr and a small magnetic stirrer bar.

c) N_2 was blown across the top of the solution to displace the air as best possible, the bottle capped tightly, sealed with parafilm and wrapped in foil to exclude light.

d) This was stirred on a magnetic stirrer for 12 hrs at room temperature, after which a further 60 mg of CNBr was added and the above procedure in c above repeated. The solution was then stirred for a further 12 hrs.

e) After the required incubation period the solution was diluted 15 times with sddw, 50 ml aliquoted per freeze dry flask and freeze dried. The freeze dried material was weighed out and dissolved at 4 mg/ml in 100 mM Tris pH 8.

6.15 Inhibitor assays

6.15.1 Active site titratiion

Reagents:

a) Trypsin (type XIII from bovine pancreas, TPCK treated, supplied by Sigma) was dissolved in 1 mM HCl at approximately 10 mg/ml.

b) A 0.1 M, pH 8.3 solution of veronal (Na barbital) was prepared containing 0.02 M $CaCl_2$.

c) p-Nitrophenol guanidinobenzoate (NPGB) was dissolved in dimethyl formamide at 16.8 mg/ml to give a 0.05 M solution. This was diluted by addition of 4 volumes of acetonitrile to give a 0.01 M solution.

d) Inhibitor (purified natural ETI and soybean trypsin inhibitor (STI)) and putative inhibitor (expression control, syn-ETI and asp-ETI) were dissolved at approximately 4 mg/ml in Tris-HCl pH 8.

Procedure. Cuvettes were prepared as shown in Table 6.1:

Reagent	Volume (ml)/cuvette			
	A	B	A′	B′
Trypsin 10 mg/ml	0.2	0.2	-	-
HCl 1 mM	-	-	0.2	0.2
Veronal buffer	0.69	0.69	0.69	0.69
Tris-HCl 0.1M, pH 8	-	0.1	-	0.1
Inhibitor solution	0.1	-	0.1	-

Table 6.1 Plan showing the preparation of cuvettes for titration of trypsin active sites. Where in the table above, B and B′ represent the titration of trypsin active sites and its reference reaction respectively. A and A′ represent titration of trypsin active sites remaining after incubation with an inhibitor solution and the related reference reaction respectively.

To each reaction being followed, 10 μl of NPGB solution was added and rapidly mixed by capping the top of the cuvette and inverting twice. As fast as possible, the cuvette was placed in the spectrophotometer and the absorbance at 410 nm (A410) recorded at 30 seconds after addition of the titrant. The A410 values were recorded at 60 second intervals thereafter up to 5.5 minutes.

As a single-beam spectrophotometer was utilized for these titrations, each reaction was followed separately of its reference reaction. The recorded values for the reference reaction were deducted from the appropriate reaction values.

Corrected A410 values were plotted as a function of time. The resultant curve from each set of data was extrapolated to Time = 0 to determine ΔA (ie the burst A410 for trypsin) and ΔB (ie the burst A410 for the remaining trypsin active sites after incubation with the inhibitor in question).

The molarity of the original ETI solution was calculated from the following formula: $[\text{ETI in moles/L}] = \frac{[(\Delta B - \Delta A) * 10]}{16595}$

ie the difference between trypsin active site concentration and concentration of active sites remaining after incubation, corrected for 10 times dilution, divided by $\epsilon 410$ for p-nitrophenol (16 595 M^{-1} cm^{-1}).

6.15.2 Assays of inhibitory activity for tissue plasminogen activator

Reagents:

a) Tris-imidazole buffer stock solution (according to American Diagnostica) was made as described below.
Solution A: 3.03 g Tris, 1.7 g imidazole, 50 ml HCl (1M) was made up to 100 ml with distilled water.
Solution B: 4.04 g Tris, 2.27 g imidazole, 1.95 g NaCl was made up to 100 ml with distilled water.
Solutions A and B were mixed to give 200 ml of solution of pH 8.4 at 25°C
To the 200 ml of the solution above was added 23.4 g of NaCl to obtain a concentrated solution of ionic strength = 3.0. Aliquots were stored at -20°C.

b) Tris-Triton-Imidazole (TTI) buffer: 10 ml of Tris-imidazole stock was mixed with 0.2 ml of 10% Triton X-100 and made up to 100 ml with distilled water.

c) Substrate: Spectrozyme t-PA was supplied at 10 μmol/vial. The contents of one vial was dissolved in 0.5 ml of distilled water. To this was added 4 ml of TTI to obtain a 4 mM solution.

d) t-PA: t-PA was dissolved in 0.1M Tris pH 8.1 containing 1% Triton X-100. One chain t-PA was converted to 2-chain t-PA by treatment with plasmin linked to Sepharose CL-4b. This was assayed and diluted to ± 500 I.U./ml.

e) Inhibitors: Inhibitor and putative inhibitor samples taken from stock solutions of which active site concentrations had been determined.

Procedure: Inhibitors were diluted 200, 300, 400, 500, 600, and 800 times with TTI buffer. 100 μl aliquots of each dilution or TTI were added to the 1st and 2nd rows of a 96-well flat bottomed polystyrene plate (Greiner). To each of these was added 50 μl of the substrate Spectrozyme t-PA this was incubated at 25°C for 5 minutes. The reaction in each well was started simultaneously by the addition, with a multipipette, of 50 μl of the t-PA solution or TTI to the wells. This is shown in Table 6.2 below.

Reagent	Well of row 1 and 2 of Greiner plate							
	A	B	C	D	E	F	G	H
Inhibitor	1/500	TTI	1/200	1/300	1/400	1/500	1/600	1/800
Substrate	s	s	s	s	s	s	s	s
Incubate								
Enzyme	TTI	t-PA	t-PA	t-PA	t-PA	t-PA	t-PA	t-PA

Table 6.2 Plan showing position of each reaction in each well of the reaction plate. Where s = Substrate

A Titertek Multiskan MCC/340 microplate reader was used to measure absorbance of the released nitroanilide at 405/680 nm. Readings were taken after 15 minutes of incubation at 25°C just prior to which, at 14.5 minutes, the machine was zeroed on column A of the reaction plate. Subsequent readings were taken every 4 minutes up to 43 minutes.

Calculations:

Plots of A405/A680 as a function of time were made for each reaction well and the slope between 15 and 30 minutes determined for each plot and an average slope calculated from the duplicate of each reaction.

The reaction in well B is uninhibited t-PA incubated with substrate and the average slope = V_0 while the slope of the remaining wells (C-H) = V_i.

For each inhibitor, V_0/V_i was calculated and this was plotted as a function of [I] (the concentration of inhibitor active sites as determined by titration, in μg/ml.

The data was fitted by linear regression to the equation:

$$V_0/V_i = 1/K_{iapp} * [I] + 1$$

and the slope of this line determined. The apparent K_i was calculated as the reciprocal of the slope. This is the inhibitor concentration in μg/ml that inhibits t-PA by 50%. As $K_{iapp} = K_i \ (1 + [s]/K_m)$ where [s] is the concentration of substrate in each well which in this case is 1 mM and K_m for 2-chain t-PA has been determined by Heussen-Schemmer and Dowdle (in press) as 0.14 mM with Spectrozyme t-PA

The specific activity of pure ETI is defined by the above mentioned workers as 1 inhibitory unit/μg which has a K_i of 0.576 μg/ml (= $2.88*10^{-8}$ M). One inhibitory unit of inhibitor activity is that amount of inhibitor that will inhibit one chromogenic unit of enzyme activity by 50% and one unit of enzyme activity is equal to 10 pmol of substrate hydrolysed per minute.

Thus, the specific activity of an unknown sample is calculated by dividing 0.576 by the K_i. Furthermore, as pure ETI has specific activity of 100 000 units/mg in the flourometric assay, 1 unit determined by chromogenic assay is equal to approximately 100 fluorometric units. In order to compare specific activities determined by fluorometric assays which appear in the earlier literature with results derived in this work, all specific activities for inhibition of t-PA were reported in terms of the old units.

6.16 Amino acid sequence analysis

Sequence analysis was performed on a gas-liquid solid phase sequencer constructed as outlined by Hewick *et al.* (1981) and slightly modified as described by Brandt *et al.* (1984) The reagents, solvents and degradation cycle used have been published in the above references. The converted PTH amino acids were identified by isocratic HPLC system on 3 x 250 mm 3 μ Lichrospher C_{18} (Bishoff) column as described by Lottspeich (1985).

APPENDIX A

ASSAY VALUES, REGRESSION DATA AND CALCULATIONS

I Table of observed values for titration of inhibitor with trypsin

Sample	A410 values at time (T) in seconds						
	T=0	30	90	150	210	270	330
Trypsin	0.675	0.711	0.785	0.828	0.852	0.873	0.891
natural ETI	0.414	0.443	0.502	0.554	0.580	0.600	0.617
val-ETI	0.477	0.518	0.602	0.647	0.668	0.690	0.704
asp-ETI	0.425	0.461	0.534	0.578	0.600	0.623	0.641
STI	0.423	0.457	0.527	0.567	0.591	0.611	0.628
Control	0.673	0.707	0.778	0.814	0.834	0.859	0.871

Table A.1 Absorbance at 410 nm at the time intervals shown. Experimental conditions as described in Chapters 4.1, 4.2 and 6.15. The T=0 column shows the values where the extrapolated graph intersects the A410 axis as shown in Figure 4.3.

II Observed values and regression data for titration of inhibitors, putative inhibitors and controls against t-PA

Table A 2 See next four pages. Observed absorbance values from chromogenic assays of dilutions of inhibitors, putative inhibitors and controls against t-PA taken every 4 minutes from 15 to 43 minutes. Alongside the absorption vs time data from dilutions of each sample is shown the regression analysis (reported as regression output) relating to that set of data.

Uninhibited

A405/A680	Time
0.202	15
0.256	19
0.306	23
0.343	27
0.398	31
0.453	35
0.507	39
0.56	43

Regression Output:

Constant		0.010705
Std Err of Y Est		0.00562
R Squared		0.998246
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.01267	
Std Err of Coef.	0.000217	

1/200

A405/A680	Time
0.145	15
0.173	19
0.197	23
0.218	27
0.245	31
0.273	35
0.296	39
0.32	43

Regression Output:

Constant		0.052557
Std Err of Y Est		0.001859
R Squared		0.999207
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.006235	
Std Err of Coef.	7.17E-05	

1/300

A405/A680	Time
0.159	15
0.195	19
0.224	23
0.242	27
0.28	31
0.316	35
0.345	39
0.376	43

Regression Output:

Constant		0.044188
Std Err of Y Est		0.004886
R Squared		0.996406
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.007687	
Std Err of Coef.	0.000188	

1/400

A405/A680	Time
0.165	15
0.201	19
0.229	23
0.255	27
0.293	31
0.329	35
0.363	39
0.396	43

Regression Output:

Constant		0.040229
Std Err of Y Est		0.003864
R Squared		0.998035
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.008229	
Std Err of Coef.	0.000149	

1/500

A405/A680	Time
0.169	15
0.206	19
0.242	23
0.273	27
0.314	31
0.349	35
0.388	39
0.425	43

Regression Output:

Constant		0.031298
Std Err of Y Est		0.002235
R Squared		0.999464
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.009119	
Std Err of Coef.	8.62E-05	

1/600

A405/A680	Time
0.178	15
0.22	19
0.25	23
0.287	27
0.331	31
0.374	35
0.414	39
0.454	43

Regression Output:

Constant		0.027125
Std Err of Y Est		0.004349
R Squared		0.998271
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.009875	
Std Err of Coef.	0.000168	

1/800

A405/A680	Time
0.184	15
0.229	19
0.27	23
0.305	27
0.35	31
0.392	35
0.435	39
0.478	43

Regression Output:

Constant		0.028378
Std Err of Y Est		0.002545
R Squared		0.999467
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.010414	
Std Err of Coef.	9.82E-05	

asp-ETI

Uninhibited

A405/A680	Time
0.198	15
0.253	19
0.303	23
0.354	27
0.408	31
0.467	35
0.514	39
0.554	43

Regression Output:		
Constant		0.006533
Std Err of Y Est		0.005102
R Squared		0.998611
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012926	
Std Err of Coef.	0.000197	

1/200

A405/A680	Time
0.198	15
0.245	19
0.298	23
0.349	27
0.4	31
0.451	35
0.505	39
0.551	43

Regression Output:		
Constant		0.005134
Std Err of Y Est		0.001863
R Squared		0.999809
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012741	
Std Err of Coef.	7.19E-05	

1/300

A405/A680	Time
0.196	15
0.25	19
0.306	23
0.352	27
0.401	31
0.454	35
0.502	39
0.559	43

Regression Output:		
Constant		0.006887
Std Err of Y Est		0.003039
R Squared		0.999495
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.01278	
Std Err of Coef.	0.000117	

1/400

A405/A680	Time
0.2	15
0.247	19
0.302	23
0.35	27
0.406	31
0.459	35
0.505	39
0.556	43

Regression Output:		
Constant		0.006217
Std Err of Y Est		0.002651
R Squared		0.999619
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012824	
Std Err of Coef.	0.000102	

1/500

A405/A680	Time
0.199	15
0.249	19
0.303	23
0.354	27
0.406	31
0.457	35
0.506	39
0.559	43

Regression Output:		
Constant		0.006354
Std Err of Y Est		0.001196
R Squared		0.999923
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012854	
Std Err of Coef.	4.61E-05	

1/600

A405/A680	Time
0.203	15
0.255	19
0.305	23
0.358	27
0.409	31
0.461	35
0.512	39
0.563	43

Regression Output:		
Constant		0.010048
Std Err of Y Est		0.000574
R Squared		0.999982
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012869	
Std Err of Coef.	2.21E-05	

1/800

A405/A680	Time
0.213	15
0.262	19
0.306	23
0.347	27
0.407	31
0.461	35
0.525	39
0.569	43

Regression Output:		
Constant		0.012357
Std Err of Y Est		0.008142
R Squared		0.996452
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.012893	
Std Err of Coef.	0.000314	

NATURAL ETI

Uninhibited

A405/A680	Time
0.186	15
0.239	19
0.286	23
0.34	27
0.397	31
0.45	35
0.505	39
0.559	43

Regression Output:

Constant		-0.01728
Std Err of Y Est		0.002796
R Squared		0.999609
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013363	
Std Err of Coef.	0.000108	

1/200

A405/A680	Time
0.119	15
0.139	19
0.159	23
0.18	27
0.204	31
0.221	35
0.239	39
0.257	43

Regression Output:

Constant		0.045095
Std Err of Y Est		0.002214
R Squared		0.998244
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.004988	
Std Err of Coef.	8.54E-05	

1/300

A405/A680	Time
0.137	15
0.161	19
0.19	23
0.218	27
0.247	31
0.271	35
0.296	39
0.321	43

Regression Output:

Constant		0.037223
Std Err of Y Est		0.002137
R Squared		0.99908
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.006652	
Std Err of Coef.	8.24E-05	

1/400

A405/A680	Time
0.146	15
0.18	19
0.211	23
0.243	27
0.277	31
0.304	35
0.333	39
0.362	43

Regression Output:

Constant		0.033458
Std Err of Y Est		0.002672
R Squared		0.998928
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.007708	
Std Err of Coef.	0.000103	

1/500

A405/A680	Time
0.15	15
0.183	19
0.214	23
0.247	27
0.284	31
0.312	35
0.343	39
0.377	43

Regression Output:

Constant		0.028988
Std Err of Y Est		0.001889
R Squared		0.999514
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.008095	
Std Err of Coef.	7.29E-05	

1/600

A405/A680	Time
0.161	15
0.198	19
0.227	23
0.261	27
0.272	31
0.332	35
0.367	39
0.401	43

Regression Output:

Constant		0.031307
Std Err of Y Est		0.009964
R Squared		0.987837
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.008485	
Std Err of Coef.	0.000384	

1/800

A405/A680	Time
0.17	15
0.205	19
0.243	23
0.278	27
0.322	31
0.355	35
0.392	39
0.428	43

Regression Output:

Constant		0.029753
Std Err of Y Est		0.002194
R Squared		0.999502
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.009289	
Std Err of Coef.	8.46E-05	

STI

Uninhibited

A405/A680	Time
0.198	15
0.233	19
0.292	23
0.333	27
0.393	31
0.457	35
0.522	39
0.549	43

Regression Output:		
Constant		-0.01256
Std Err of Y Est		0.011326
R Squared		0.993532
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013265	
Std Err of Coef.	0.000437	

1/200

A405/A680	Time
0.185	15
0.238	19
0.289	23
0.336	27
0.391	31
0.446	35
0.498	39
0.553	43

Regression Output:		
Constant		-0.01293
Std Err of Y Est		0.002618
R Squared		0.999644
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013101	
Std Err of Coef.	0.000101	

1/300

A405/A680	Time
0.184	15
0.238	19
0.29	23
0.34	27
0.387	31
0.449	35
0.5	39
0.552	43

Regression Output:		
Constant		-0.01312
Std Err of Y Est		0.003202
R Squared		0.999469
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013125	
Std Err of Coef.	0.000124	

1/400

A405/A680	Time
0.181	15
0.24	19
0.291	23
0.336	27
0.389	31
0.447	35
0.499	39
0.552	43

Regression Output:		
Constant		-0.01401
Std Err of Y Est		0.003531
R Squared		0.999355
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013134	
Std Err of Coef.	0.000136	

1/500

A405/A680	Time
0.181	15
0.236	19
0.288	23
0.338	27
0.384	31
0.45	35
0.5	39
0.55	43

Regression Output:		
Constant		-0.01691
Std Err of Y Est		0.004206
R Squared		0.999094
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013199	
Std Err of Coef.	0.000162	

1/600

A405/A680	Time
0.194	15
0.239	19
0.299	23
0.345	27
0.399	31
0.456	35
0.509	39
0.56	43

Regression Output:		
Constant		-0.00783
Std Err of Y Est		0.00333
R Squared		0.999433
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013205	
Std Err of Coef.	0.000128	

1/800

A405/A680	Time
0.192	15
0.238	19
0.293	23
0.343	27
0.401	31
0.456	35
0.509	39
0.556	43

Regression Output:		
Constant		-0.01058
Std Err of Y Est		0.003415
R Squared		0.999407
No. of Observations		8
Degrees of Freedom		6
X Coefficient(s)	0.013244	
Std Err of Coef.	0.000132	

From each regression analysis is required the slope of the line that fits the data by linear regression. The slopes are reported as X Coefficient(s). These represent Vo for the uninhibited enzyme and Vi for the same amount of enzyme inhibited by increasing dilutions of inhibitor, putative inhibitor or control. A summary of the slopes derived for each sample is presented in Table A 3.

Amount or. dilution of inhibitor	Slopes of fitted line to each set of data				
	control	val-ETI	asp-ETI	nat-ETI	STI
0	0.013229	0.01267	0.012926	0.013363	0.013893
1/200	0.013092	0.006235	0.012741	0.004988	0.013021
1/300	0.013167	0.007687	0.012780	0.006652	0.013024
1/400	0.013205	0.008229	0.012824	0.007708	0.013125
1/500	0.013214	0.009119	0.012854	0.008095	0.013199
1/600	0.013217	0.009875	0.012869	0.008485	0.013205
1/800	0.013223	0.010414	0.012893	0.009289	0.013384

Table A 3 A summary of the slopes derived from the fitted data shown in Table A 2.

The ratio of the rate of uninhibited enzyme (V_0) to the rate of inhibited enzyme (Vi), at each concentration of inhibitor, was determined. This was plotted against the concentration of inhibitor ([I]), corresponding to each Vi, according to the equation: $Vo/Vi = [I]/Ki_{app} + 1$ (the specific velocity equation described in Chapter 4.1) in this equation $Ki_{app} = Ki\ (1 + [s]/K_m)$ and for this reaction [S] = 1mM and Km = 0.14mM (see Chapter 6.15) ie $Ki_{app} = 8.14.Ki$. A straight line was

fitted to each set of data by linear regression. The sets of data with their regression analyses are shown in Table A 4. In addition, Table A 4 has a two extra columns labeled X and Y in which the corresponding Y axis (Vo/Vi) values are displayed for X axis ([I]) values of 1 and 11 to facilitate the plotting of all the graphs on one set of axes (Figure 4.4).

The specific activity of each sample was calculated from the slope of the fitted graph (the X coefficient(s) values of the regression outputs) of Vo/Vi vs [I]. The calculation was performed as shown below using the data from natural ETI as an example: the slope = $1/Ki_{app}$ or $1/slope = Ki_{app}$ in the case of ETI slope = 0.206823 thus Ki_{app} = 4.835. As described above Ki = $Ki_{app}/8.14$ thus in the case of natural ETI Ki = 0.594 μg/ml. In Chapter 4.1 it was shown that pure ETI has a specific activity of 1 unit/μg and that has a Ki = 0.576 μg/ml. Therefore, the specific activity of the natural ETI sample used in this work was calculated as 0.576/0.594 ie 0.9696969 units/μg or 96969.69 old units/mg which is equal to 1.94×10^9 units/mmole. The Ki and specific activity of each sample is presented in Table 4.3.

Table A 4 <u>Displayed on following page</u>. Shows the Vo/Vi ratios with the corresponding inhibitor concentration ([I]) in μg/ml, the corresponding [I] values for Vo/Vi values of 1 and 11 (shown under the headings of X and Y), and the regression output for each set of data.

Control

Vo/Vi	[I]	X	Y	Regression Output:		
1.0105	11	1	0.997263	Constant		0.996009
1.0047	7.33	11	1.009801	Std Err of Y Est		0.00091
1.0018	5.5			R Squared		0.955528
1.0011	4.4			No. of Observations		6
1.0009	3.67			Degrees of Freedom		4
1.0005	2.75					
				X Coefficient(s)	0.001254	
				Std Err of Coef.	0.000135	

val-ETI

Vo/Vi	[I]	X	Y	Regression Output:		
2.032	5.98	1	1.13022	Constant		0.948809
1.648	3.98	11	2.944325	Std Err of Y Est		0.02991
1.539	2.99			R Squared		0.991951
1.389	2.39			No. of Observations		6
1.283	1.99			Degrees of Freedom		4
1.216	1.49					
				X Coefficient(s)	0.181411	
				Std Err of Coef.	0.008171	

asp-ETI

Vo/Vi	[I]	X	Y	Regression Output:		
1.015	7.53	1	1.001404	Constant		0.999226
1.011	5.02	11	1.023185	Std Err of Y Est		0.00072
1.008	3.76			R Squared		0.979861
1.006	3.01			No. of Observations		6
1.004	2.51			Degrees of Freedom		4
1.003	1.88					
				X Coefficient(s)	0.002178	
				Std Err of Coef.	0.000156	

Natural ETI

Vo/Vi	[I]	X	Y	Regression Output:		
2.679	7.88	1	1.199098	Constant		0.992275
2.009	5.25	11	3.267329	Std Err of Y Est		0.064168
1.734	3.94			R Squared		0.983695
1.651	3.15			No. of Observations		6
1.575	2.63			Degrees of Freedom		4
1.439	1.97					
				X Coefficient(s)	0.206823	
				Std Err of Coef.	0.013314	

STI

Vo/Vi	[I]	X	Y	Regression Output:		
1.0125	7.63	1	1.001838	Constant		0.999991
1.0107	5.08	11	1.020303	Std Err of Y Est		0.002074
1.01	3.81			R Squared		0.811915
1.005	3.05			No. of Observations		6
1.0045	2.54			Degrees of Freedom		4
1.0016	1.91					
				X Coefficient(s)	0.001846	
				Std Err of Coef.	0.000444	

APPENDIX B

RECIPES

B.I DNA techniques

1) Broth

Luria Broth (Ausubel *et al.*, 1987)

per litre		
	Tryptone	10g
	Yeast extract	5g
	NaCl	5g
	NaOH (1M)	1ml

2) Alkaline lysis buffers (Maniatis *et al.*, 1982)

Solution I:

Glucose	50mM
EDTA	10mM
Tris-HCL pH 8	25mM

This is autoclave and kept in the fridge.

Solution II:

NaOH	0.2M
SDS	1%

Kept at room temp.

Solution III To make 100 ml of solution add:

Potassium acetate	60 ml of 5M solution
Acetic acid	1.5 ml
H_2O	28.5

3 Phenol preparation (Ausubel *et al.*, 1987)

To 500 ml of redistilled phenol 0.5g of hydroxyquinoline was added and dissolved. 500ml of 50mM Tris base (pH 10) was added and stirred for 10 min.,then stood at room temp. to allow phases to separate. The aqueous phase was removed and replaced with the same amount of 50mM Tris (pH 8) and the process repeated twice. The equilibrated phenol was stored, at 4°C under 50mM Tris (pH 8).

4 DNAse free RNAse A preparation

RNAse A was dissolved at 20mg/ml in 0.3M Na acetate (pH 5.5). This was aliquoted into polypropylene eppendorf tubes and in these incubated in boiling water for 10min and stored up to 3 months at 4°C.

5 TE (Tris-EDTA)

Tris (pH 8)	10mM
Na_2EDTA	1mM

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

Tris (pH 8)	0.89M
Boric acid	0.89M
Na_2EDTA	0.2M

7 **DNA loading buffer**

Dissolve sucrose to 40% in TE (pH 8) and add a tiny amount of Bromophenol on the very tip of a spatula.

8 **Denaturation solution for bridge mutagenesis**
(10 times concentrated)

Tris (pH 8)	50mM
KCl	100mM
$MgSO_4$	50mM
DTT	5mM

9 **TNE buffer** (used in oligonucleotide purification)

Tris (pH 7)	100mM
EDTA	5mM
NaCl	0.5mM

10 **Freeze squeeze buffer** (extraction of DNA from agarose)

NaOAc (pH 7)	300mM
EDTA	1mM

11 **Mung bean nuclease buffer**

NaOAc (pH 5)	50mM
NaCl	30mM
$ZnSO_4$	1mM

12 **Ligation buffer for bacteriophage T4 ligase**
(10 times concentrated)

Tris (pH 7.6)	200mM
$MgCl_2$	50mM
DTT	50mM

B II Western blots and protein detection

1 **Phosphate buffered saline (PBS)**

Dissolve in 800ml of double distilled water :

NaCl	8g
KCl	0.2g
Na_2HPO_4	1.44g
KH_2PO_4	1.2 g

Make up to 1L with water pH is 7.2

2 **SDS running buffer** (10 times conc.)
For transfer of proteins from PAGE gels to nitrocellulose.

Tris (pH 8)	30.28g
Glycine	144.1 g
SDS	1.0 g

Dissolved in water and made up to 1L

3 Transfer buffer For transfer of proteins to nitrocellulose

SDS running buffer 10x	400ml
methanol	800ml
water	2.8L

4 TBS tween

NaCl	100mM
Tris (pH 7.5)	10mM
Tween 20	0.1%

5 Blocking buffer for nitrocellulose

Tris (pH 7.5)	50mM
NaCl	150mM
Tween 20	0.05%
Azide	0.02%
Milk powder	5%

6 Developing buffer

Diethanolamine	1M
NaCl	0.1M
$MgCl_2$	0.5mM

Adjust pH to 9.6

7 BCIP/NBT developing solution

Dissolve:
Bromo chloro indolyl phosphate (BCIP) @ 50 mg/ml in 100% DMF
Nitroblue tetrazolium (NBT) @ 75 mg/ml in 70% DMF

To develop blot after incubation with goat-anti-rabbit conjugate and washing in developing buffer. Incubate in developing buffer @ 50 $\mu l/cm^2$ and to this add 4.4 μl BCIP and 3.3 μl NBT per ml of developing buffer.

B III Protein gels

1 SDS sample buffer (2x concentrated)
(for dissolving and running proteins on SDS PAGE gels)

Tris	1.52g
Glycerol	20ml
SDS	2g

Add bromophenol for colour and dissolve in 40 ml of water, adjust pH to 6.8 and make up to 100 ml.
Just prior to use add 20 μl/ml of 2-mercaptoethanol.

2 Stacking gel buffer (4x concentrated)

Tris (pH 6.8)	0.5M
SDS	0.4%

Filter through 0.45 μm filter and store at 4^oC.

3 Running gel buffer (4x concentrated)

Tris (pH 8.8)	1.5M
SDS	0.4%

Filter through a 4.5μm filter and store at 4°C

4 Acrylamide stock solution

Acrylamide	30%
N',N'-methylenebisacrylamide	0.8%

Filter and store as before.

5 Stacking gel

30% acrylamide stock	1.3ml
4x stacking gel buffer	2.5ml
Water	6.1ml
10% Ammonium persulphate	50μl
TEMED	10μl

6 Separating gel

30% Acrylamide stock	10ml
4x stacking gel buffer	5ml
Water	5ml
10% ammonium persulphate	132μl
TEMED	26μl

7 SDS electrophoresis buffer (5x concentrated)
For running SDS PAGE gels

Tris (pH 8)	15.1g
Glycine	72.0g
SDS	5g

Make up to 1L with water. Do not alter pH as this is 8.3 after dilution to 1x.

REFERENCES

Adams, S.E., I.D. Johnson, M. Braddock, A.J. Kingsman, S.M. Kingsman and R.M. Edwards (1988). Nucl. Acid. Res. *16*:4287-4298.

Alessi, M.C., P.J. Declerk, M. De Mol, L. Nelles and D. Collen (1988). Eur. J. Biochem. *175*:531-540.

Amann, E., J. Brosius and M. Ptashne (1983). Gene. *25*:167-178

Andersson, S.G.E. and C.G. Kurland (1990). *54*:198-210.

Andreasen, P.A., A. Riccio, K.G. Welinder, R. Douglas, R. Sartorio, L.S. Nielsen, C. Oppenheimer, F. Blasi and K. Dano (1986). FEBS Lett. *209*:213-218.

Antoniw, J.F. (1985). Anal.Biochem. *132*:14-19.

Ausubel. F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl (1987). Current Protocols in Molecular Biology. John Wiley and Sons, New York.

Baici, A. (1981). Eur. J. Biochem. *119*:9-14.

Baici, A. and M. Gyger-Marazzi (1982). Eur. J. Biochem. *129*:33-41.

Bakhit, C., D. Lewis, R. Billings and B. Malfroy (1987). J. Biol. Chem. *262*:8716-8720.

Balbas, P. and F. Bolivar (1990). Meth. Enzymol. *185*:14-37.

Banyai, L., A. Varady and L. Patthy (1983). FEBS Lett.*163*:37-41.

Barrett, A.J. and P.M. Starkey (1973). Biochem. J. *133*:709-724.

Beers, W.H. (1975). Cell *6*:379-386.

Bennett, W.F., N.F. Paoni, B.A. Keyt, D. Botstein, A.J.S. Jones, L. Presta, F.M. Wurm and M.J. Zoller (1991). J. Biol. Chem. *266*:5191-5201.

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

Blow, D.M., J. Janin and R.M. Sweet (1974). Nature *249*:54-57.

Blundell, M., E. Craig and D. Kenell (1972). Nature. *238*:46-48.

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

Bode, W., I. Mayer, U. Baumann, R. Huber, S.R. Stone and J. Hofsteege (1988). EMBO J. *8*:3467-3475.

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

Borowiec, J.A. and J.D. Gralla (1986). Biochemistry. *25*:5051

Boyot, P., L. Pillet, F. Ducanel, J. Boulain, O. Tremeau and A. Menez (1990). FEBS. *266*:87-90.

Brandt, W.F., H. Alk, M. Chauhan and C. von Holt (1984). FEBS Lett., *174*:228-232.

Brommer, E.J.P. (1984). Thromb. Res. *34*:109-115.

Brosius, J. (1984). Gene. 27:161-172

Bruch, M., V. Weiss and J. Engel (1988). J. Biol. Chem. *263*:16626-16630.

Carpousis, A.J., J.E. Stefano and J.D. Gralla (1982). J. Mol. Biol. *157*:619-621.

Carrell, R. and J. Travis (1985). Trends Biochem. Sci. *10*:20-24.

Carrell, R.W. and J. Travis (1985). Trends Biochem. Sci. *10*:20-24.

Carter, A.D., C.E. Morris and W.T. McAllister (1981). J. Virol. *37*:636-642.

Chase, T. and E. Shaw (1970). Method. Enzymol. *19*:20-27.

Cheng, Y.S., D.Y. Kwoh, T.J. Kwoh, B.C. Saltveldt and D. Zipser (1981). Gene. *14*:121-

Chmielewska, J., M. Ranby and B. Wiman (1980). Biochem. J. *251*:327-323.

Chmielewska, J., M. Ranby and B. Wiman (1988). Biochem. J. *251*:327-332.

Cohen, A.B., L.D. Greunke, J.C. Craig, and D. Geczy (1977). Proc. Natl. Acad. Sci. USA. *74*:4311-4314.

Collen, D. (1976). Eur. J. Biochem. *69*:209-216.

Colucci, M. J.A. Paramo and D. Collen (1985). J. Clin. Invest. *75*:818-824.

Conner, G.E. and J.A. Udey (1990). Cell Biol. *9*:1-9.

Cook, N.S. and D. Ubben (1990). Trends Pharmacol. Sci. *11*:444-451.

Curran, J.F. and M. Yarus (1988). J. Mol. Biol. *203*:75-83.

Curran, J.F. and M. Yarus (1989). J. Mol. Biol. *209*:65-77.

Dano, K. and E. Reich (1979). Biochem. Biophys. Acta. *566*:138-151.

Das, A (1990). Method. Enzymol. *185*:93-112

Davanloo, P., A.H. Rosenberg, J.J. Dunn and F.W. Studier (1984). Proc. Natl. Acad. Sci. U.S.A. *81*:2035-2039.

De Boer, H.A., L.J. Comstock and M. Vasser (1983). Proc. Natl. Acad. Sci. USA. *80*:21-25.

Declerck, P.J., M. De Mol, M. Alessi, S. Baudner, E. Paques, K.T. Preissner, G. Muller-Berghaus and D. Collen (1988). J. Biol. Chem. *263*:15454-15461.

Deuschle, U., W. Kammerer, R. Gentz and H. Bujard (1986). EMBO. J. *5*:2987-

Dexter, D.L. and P. Calabresi (1982). Biochim. Biophys. Acta. *695*:97-112.

diGuan, C., P. Li, P.D. Riggs and H. Inouye (1988). Gene. *67*:21-30.

Dix, D.B. and R.C. Thompson (1989). Proc. Natl. Acad. Sci. USA. *86*:6888-6892.

Ducand, F., J-C. Boulain, O. Tremeau and A. Menez (1989) Prot. Engin. *3*:139-143

Dunn, J.J. and F.W. Studier (1981). J. Mol. Biol. *148*:303-330.

Dunn, J.J. and F.W. Studier (1983). J. Mol. Biol. *166*:477-535.

Duplay, P., H.Bedouelle, A.Fowler, I.Zabin, W.Saurin and M.Hofnung (1984). J.Biol.Chem. *259*:10606-10613.

Edelberg, J.M., C.F. Reilly and S.V. Pizzo (1991). J. Biol. Chem. *266*:7488-7493.

Eigenbrot, C., M. Randall and A.A. Kossiakoff (1990). Protein Eng. *3*:591-598.

Eismann, E., B. von Wilcken-Bergman and G. Muller-Hill (1987). J. Mol. Biol. *195*:949-

Erlich, H.J., R. Klein Gebbink, J. Keijer, M. Linders, K.T. Preissner and H. Pannekoek (1990). J. Biol. Chem. *265*:13029-13035.

Finkenstat, W.R. and M. Laskowski Jr. (1965). J. Biol. Chem. *240*:962-963.

Flashner, Y. and J. Gralla (1988). Proc. Natl. Acad. Sci. U.S.A. *85*:8968-8973.

Freedman, L.P., B.F. Luisi, Z.R. Korszun, R. Basavappa, P.B. Zigler and K.R. Yamamoto (1988). Nature. *334*:543-546.

Freundl, R., H. Schwarz, M. Klose, N.R. Movva and V. Henning (1985). EMBO J. *4*:3593-3598.

Fujikawa, K. M.E.Legaz and E.W.Davie (1972). Biochem. *11*:4892-4899.

Galili, G. (1989). Proc. Natl. Acad. Sci. U.S.A. *86*:7756-7760.

Gaun, C., P.Li, P.D.Riggs and H.Inouye (1988). Gene *67*:21-30.

Goeddel, D.V. (1990). Meth. Enzymol. *185*:3-7.

Gold, L. (1990). Meth. Enzymol. *185*:11-

Goldberg, A.L. and S.A. Goff (1986). In Reznikoff, W. and L. Gold (Eds). Maximizing gene expression. Butterworths, Boston.

Golder, J.P. and R.W. Stephens (1983). Eur. J. Biochem. *136*:517522.

Golomb, M. and M. Chamberlin (1974). J. Biol. Chem. 249:2858-2863.

Gottesman, M.E., S. Adhya and A. Das (1980). J. Mol. Biol. *140*:57-65.

Gouy, M. and C.Gautier (1982). Nucleic Acids Res. *10*:7055-7074.

Gralla, J.D. (1990). Meth. Enzymol. *185*:37-54

Granelli-Piperno, A. and E. Reich (1983). J. Cell Biol. 97:1029-1037.

Grieger, M. and B.R. Binder (1984). J. Biol. Chem. *259*:2976-2981.

Grodberg, J. and J.J. Dunn (1988). J. Bacteriol. *170*:245-

Grodberg, J., M.D. Lundrigan, D.L. Toledo, W.F. Mangel and J.J. Dunn (1988). Nucl. Acid Res. *16*:1209.

Grosjean, H. and W. Fiers (1982). Gene. *18*:199-209.

Gross, E. (1967). Method. Enzymol. *11*:238-255.

Grossman, A., J. Erickson and C. Gross (1984). Cell. *38*:383-390.

Grottesmann, S. (1989). Ann. Rev. Genet. *23*:163-187.

Grottesmann, S. (1990). Meth. Enzymol. *185*:119-129.

Haber, E., T. Quetermous, G.R. Matsueda, and M. S. Runge (1989). Science. *243*:51-56.

Harris, T.J.R. (1983). In Williamson (Ed). Genetic Engineering. vol4. Academic Press, London.

Hawley, D.K. and W.R. McClure (1983). Nucl. Acids Res. *11*:2232-2239.

Hayden, M.A. and W. Mandecki (1988). DNA. *8*:571-577.

Hedner, U. (1979). In: Collen, D., B. Wiman and M. Verstraete (Eds) Inhibitors of coagulation and fibrinolysis. Elsevier/North-Holland, Amsterdam.

Heussen, C., F. Joubert and E.B.D. Dowdle (1984). J. Biol. Chem. *259*:11635-11638.

Heussen-Schemmer, C., Merrifield, E.H. and Dowdle, E.B.D. (1991). Thromb. Haemostas. *66*:226-231.

Heussen-Schemmer, C. and E.B. Dowdle (1992). S. A. J. Sci. (in press).

Hewick, R.M., M.W. Hunkapillar, L.E. Hood and W.J. Dreyer (1981). J. Biol. Chem. *256*:7990-7997.

Horn, T. and Urdea, M.S. (1988). Nucl. Acid. Res. *16*:11559-11571.

Hoylearts, M. D.C. Rijken, H.R. Lijnen and D. Collen (1982). J. Biol. Chem.*257*:2912-2919.

Hsiung, H.M., A. Cantrell, J. Luirink, B. Oudega, A.J. Veros and G.W. Becker. Bio/technology. 7:267-272.

Huber, R. and R.W. Carrell (1989). Biochemistry *28*:8951-8966.

Huber, R., D. Kukla, W. Bode, P. Schwager, K. Bartels, S. Deisenhofer and W. Steigmann (1974). J. Mol. Biol. *89*:73-101.

Jackson, C.M. (1984). In: Spaet, T.H. (Ed), Progress in haemostasis and thrombosis, Vol.7. Grune and Stratton, New York.

Jallat, S., D. Carvallo, L.H. Tessier, D. Roecklin, C. Roitsch, F. Ogushi, R.G. Crystal, and M. Courtney (1986). Protein Eng. *1*:29-35.

Jordan, R., D. Beeler and R.D. Rosenberg (1979). J. Biol. Chem. *254*: 2902-2913.

Joubert, F.J. (1982). Int. J. Biochem. *14*:187-193.

Joubert, F.J. (1986). S. Afr. J. Chem. *39*:169-175.

Joubert, F.J. and E.B.D. Dowdle (1987). Thromb. Haemostas. *57*:356-360.

Joubert, F.J., E.H. Merrifield and E.B.D. Dowdle (1987). Int. J. Biochem. *19*:601-606.

Kalyan, N.K., S.G. Lee, J. Wilhelm, K.P. Fu, W. Hum, R. Rappapport, R.W. Hartzell, C. Urbano and P.P. Hung (1988). J. Biol. Chem. *263*:3971-3978.

Kato, C., T. Kobayashi, T. Kudo, T. Furuseto, Y. Murakami, T. Tanaka, H. Baba, T. Oishi, E. Odtsuka, M. Ikehara, T. Yanagida, H. Kato, S. Moriyama and K. Honkoshi (1987). Gene. *54*:197-202

Kellerman, O.K. and T.Ferenci (1982). Method.Enzymol. *90*:459-463.

Kennell, D.E. (1986). In: Resnikoff, W. and L. Gold (Ed.) Maximizing gene expression. Butterworths, Boston.

Kezdy, F.J. and E.T. Kaiser (1970). Method. Enzymol. *19*:3-20.

Kinsella, A.R. and M. Radman (1980). Proc. Natl. Acad. Sci. USA 77:3544-3547.

Klotsky, R.A. and I. Schwartz (1987). Gene. *55*:141.

Knauer, D.J., and D.D. Cunningham (1984). Trends Biochem. Sci.*9*:231-233.

Kniskern, P.J., A. Hagopian, D.L. Montgomery, P. Burke, N.R. Dunn, K.J. Hofman, W. J. Miller and R.W. Ellis (1986). Gene. *46*:135-141.

Kohno, T., D.F. Carmichael, A. Sommer and R.C. Thompson (1990). Meth. Enzymol. *185*:187-195.

Kruithof, E.K.O. (1988). Enzyme. *40*:113-121.

Kuhnke G., H.J. Fritz and R. Ehring (1987). EMBO J. *6*:507-

Lacroix, M., F.E. Smith and I.B. Fritz (1982). Mol. Cell. Endocrinol.*26*:259-267.

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

Laug, W.E., B. Dewald, J. Schnyder and M. Baggiolini (1983). Cancer Res. *43*:22-27.

Lauritzen, C., E. Tuchsen, P.E. Hansen and O. Skovgaard (1991). Biochemistry, in press, from abstract supplied by Dr. P. Riggs New England Biolabs.

Lawrence, D.A., L. Strandberg, J. Ericson and T. Ny (1990). J. Biol. Chem. *265*:20293-20301.

Lawrence, D.A., L. Strandberg, T. Grundstrom and T. Ny (1989). Eur. J. Biochem. *186*:523-533.

Lijnen, H.R. and D. Collen (1988). Enzyme *40*:90-96.

Lindahl, T.L., P. Ohlsson and B. Wiman (1990). Biochem. J. *265*:109-113.

Liotta, L.A., R.H. Goldfarb, R. Brundage, G.P. Siegal, V. Terranova and S. Garbisa (1981). Cancer Res. *41*:4629-4636.

Loeberman, H., R. Tokuoka, J. Deisenhofer, and R. Huber (1984). J. Mol. Biol. *177*:531-556.

Lorand, L. (1972). Ann. Rev. N.Y. Acad. Sci. *202*:6-26.

Losick, R., J. Pero (1981). Cell. *25*:585.

Lottspeich, F. (1985). J. Chromatog. *326*:321-327.

Lucas, M.A., L.J. Fretto and P.A. McKee (1983). J. Biol. Chem. *258*:4249-4256.

Madison, E.L., E.J. Goldsmith, R.D. Gerard, M. Gething and J.F. Sambrook (1989). Nature. *339*:721-724.

Magnotti, R.A. (1988). Anal. Biochem. *170*:228-237.

Maina, C.V., P.D.Riggs, A.G.Grandea, B.E.Slatko, L.S.Moran, J.A.Tagliamonte, L.A.McReynolds and C.Guan (1988). Gene. *74*:365-373.

Mandecki, W. (1986) Proc. Natl. Acad. Sci. USA 83, 7177-7181.

Mandecki, W., M.A. Hayden, M. A. Shallcross and E. Stotland (1990). Gene *94*:103-107.

Mann, K.G. (1987). TIBS. *12*:229-233.

Marotti, K.R., D. Belin and S. Strickland (1982). Devel. Biol. *90*:154-159.

Marston, F.A.O. (1987). In Glover, D.M. (Ed.) DNA cloning volume III. IRL Press, Oxford.

Marston, F.O.A. (1986). Biochem. J. *240*:1-12.

Matsuda, M., S. Iwanaga and S. Nakamura (1972) . Thromb. Res. *1*:619-630.

McClure, W.R. (1985). Ann. Rev. Biochem. *54*:171-

McLachlan, A.D. (1979). J. Mol. Biol. *133*:557-563.

Miller, C.G. (1987). In Neidhardt, F.C. (Ed.). *Eschrichichia coli* and *Salmonella typhimurium*. American Society for Microbiology, Washington, D.C.

Mimuro, J. and D.J. Loskutoff (1989). J. Biol. Chem. *264*:936-939.

Miskin, R. and E. Reich (1980). Cell *19*:217-224.

Moffatt, B.A. and F.W. Studier (1987). Cell. *49*:221.

Monge, J.C., C.L. Lucore, E.T.A. Fry, B.E. Sobel and J.J. Billadello (1989). J. Biol. Chem. *264*:10922-10925.

Moroi, M., and N. Aoki (1976). J. Biol. Chem. *251*:5956-5965.

Morrison, D.C. and R.J. Ulevitch (1978). Am. J. Pathol. *93*:526-617.

Moscatelli, D. and D.B. Rifkin (1988). Biochim. Biophys. Acta. *948*:67-85.

Moses, E. and H. Hinz (1983). J. Mol. Biol. *179*:765-776.

Mosher, D.F. and A. Vaheri (1980). Biochim. Biophys. Acta *647*:113-122.

Mossing, M. and T. Record (1986). Science. *233*:889-892.

Murano, G., D. Aronson, L. Williams and L. Brown (1980). Blood *55*:430-436.

Nagai, K. and C.Thogersen (1984). Nature. *309*:810-812.

Nagai, K. and H.C.Thogersen (1990). Method. Enzymol. *185*:461-481.

Nagai, K., M.F.Perutz and C.Poyart (1985). Proc.Natl.Acad.Sci.USA. *82*:7252-7255

Nagamine, Y., D. Pearson, M.S. Altus and E. Reich (1984). Nucleic Acids Res. *12*:9525-9541.

Neurath, H. (1964). Protein-digesting enzymes. Sci. Am. *211(6)*:68-79.

Nielsen, L.S., P.A. Andreasen, J. Grondahl-Hansen, L. Skriver and K. Dano (1986). FEBS Lett. *196*:269-273.

O'Grady, R.L., L.I. Upfold and R.W. Stephens (1981). Int. J. Cancer *28*:509-515.

Oka, T., S. Sakamoto, K.I. Miyoshi, T. Fuwa, K. Yoda, M. Yamasaki, G. Tamura and T. Miyake (1985). Proc. Natl. Acad. Sci. U.S.A. *82*:7212-7216.

Onesti, S. P., Brick and D.M. Blow (1991). J. Mol. Biol. *217*:153-176.

Owen, M.C., S.O. Brennan, J.H. Lewis, R.W. Carrell (1985). New Engl. J. Med. *309*:694-698.

Ozawa, K., and M. Laskowski Jr. (1966). J. Biol. Chem. *241*:3955-3961.

Pastan, I., M. Willingham, W. Anderson and M. Gallo (1977). Cell*12*:609-617.

Pedersen, S. (1984). In: Clark, B.F.C. and H.U. Petersen (Ed.). Gene expression: the translational step and its control. Munksgaard, Copenhagen.

Pennica, D., W.E. Holmes, W.J. Kohr, R.N. Harkins, G.A. Vehar, C.A. Ward, W.F. Bennett, E. Yelverton, P.H. Seeburg, H.L. Heynecker, D.V. Goeddel and D. Collen (1983). Nature. *301*:214-221.

Pohl, G., L. Kaplan, M. Einarsson, P. Wallen and H. Jornvall (1984a). FEBS Lett. *169*:29-32.

Pohl, G., M. Kallstrom, N. Bergsdorf, P. Wallen and H. Jornvall (1984b). Biochemistry *23*:3701-3707.

Preissner, K.T., S. Holzhuter, C. Justus and G. Muller-Berghaus (1989). Blood *74*:1989-1996.

Priestle, J.P., H. Schar and M.G. Grutter (1988). EMBO. 7:339-343.

Purvis, I.J., A.J.E. Bettany, T.C. Santiago, J.R. Coggins, K. Duncan, R. Easton and A.J.P. Brown (1987). J. Mol. Biol. *193*:413-417.

Ranby, M. (1982). Biochim. Biophys. Acta. *704*:461-469.

Raven, P.H. (1984). *Erythrina* (*Fabaceae*): achievements and opportunities. Lloydia. *37*:321-331.

Rijken, D.C. and D. Collen (1981). J. Biol. Chem. *256*:7035-7041.

Rijken, D.C., G. Wijngaards and J. Welbergen (1981). J. Lab. Clin. Med. *97*:477-486.

Rijken, D.C., I. Juhan-Vague and C. Collen (1983)... J. Lab. Clin. Med. *101*:285-294.

Rijken, D.C., M. Hoylaerts and D. Collen (1982). J. Biol. Chem. *257*:2920-2925.

Robbins, K.C., L. Summaria, B. Hsieh and R.J. Shah (1967) . J. Biol. Chem. *242*:2333-2342.

Robbins, K.C., P. Bernabe, L. Arzadon and L. Summaria (1972). J. Biol. Chem. *247*:6757-6762.

Robbins, K.C., P. Bernabe, L. Arzadon and L. Summaria (1973). J. Biol. Chem. *248*:7242-7246.

Robinson, M., R. Lilley, S.Little, J.S. Emtage, G. Yarranton, P. Stevens, A. Millican, M. Easton and G. Humphreys (1984). Nucl. Acid. Res. *12*:6663-6671.

Rosenberg, A.H., B.N. Lade, D.S. Chui, S.W. Lin, J.J. Dunn and F.W. Studier (1987). Gene. *56*:125-135.

Rosenberg, R.D. and P.S. Damus (1973). J. Biol. Chem. *248*:6490-6505.

Ruhlmann, A., D. Kukla, P. Schwager, K. Bartels and R. Huber (1973). J. Mol. Biol. 77:417-436.

Russell, J. A.B. Schnieder, J. Katzhendler, K. Kowalski and L.M. Sherwood (1979). J. Biol. Chem. *254*:2296-2301.

Saito, H., S. Tabor, F. Tamanoi and C.C. Richardson (1980). Proc. Natl. Acad. U. S. A. 77:3917-3921.

Saksela, O. (1985). Biochim. Biophys. Acta. *823*:35-65.

Saksela, O., A. Vaheri, W. Schleuning, P. Mignatti and A. Barlati (1984). Int. J. Cancer 33:609-616.

Salstrom, J. and W. Szybalski (1978). J. Mol. Biol. *124*:195-202.

Sambrook, J., E.F. Fritsch and T. Maniatis (1989). Molecular Cloning a laboratory manual 2nd Edition. CSH Laboratory press.

Sano, T. and C.R. Cantor (1990). Proc. Natl. Acad. Sci. U.S.A. *87*:142-146.

Schecter, I. and A. Berger (1967). Biochem. Biophys. Res. Commun. *27*:157-165.

Schmeissiner, U., K. McKenney, M. Rosenberg and D. Court (1984). J. Mol. Biol. *176*:39.

Schoner, B.E., H.M. Hsiung, R.M. Belagaje, N.G. Mayne and R.G. Schoner (1984). Proc. Natl. Acad. Sci. U.S.A. *81*:5403-5407.

Scott, R.W. and J.B. Baker (1983). J. Biol. Chem. *258*:10439-10444.

Scott, R.W., B.L. Bergman, A. Bajpai, R.T. Hersch, H. Rodriguez, B.N. Jones, C. Barreda, S. Watts and J.B. Baker (1985). J. Biol. Chem. *260*:7029-7034.

Sedgwick, B. (1989). J. Bacteriol. *171*:2249-2253.

Sharp, P.M., E.Cowe, D.G. Higgins, D.C. Shields, K.H. Wolfe and F.Wright (1988). Nucl. Acid. Res. *16*:8207-8211.

Sharp, P.M., T.M.F. Tuohy, and K.R. Mosurski (1986). Nucl. Acid. Res. *14*:5125-5143.

Shields, X.C. and P.M. Sharp (1987). Nucl. Acid. Res. *15*:8023-8040.

Shuman, H.A., T.J. Silhavy and J.R. Beckwith (1980). J. Biol. Chem. *255*:168-174.

Silhavy, T.J. S.Szmelcman, W.Boos and M.Schwartz (1975). Proc.Natl.Acad.Sci.USA. *72*:2120-2124.

Silhavy, T.J., S.A. Benson and S.D. Emr (1983). Microbiol.Rev. *47*:313-344.

Skriver, L. L.S. Nielsen, R. Stephens and K. Dano (1982). Eur. J. Biochem. *124*:409-414.

Smith, A.T., N. Santama, S. Dacey, M. Edwards, R.C. Bray, R.N.F. Thorneley and J.F. Burke (1990). J. Biol. Chem. *265*:13335-13343.

Sorensen, M.A., C.G. Kurland and S. Pedersen (1989). J. Mol. Biol. *207*:365-377.

Sottrup-Jensen, L., T.M. Stepanik, T. Kristensen, D.M. Wierzbicki, C.M. Jones, P.B. Lonblad, S. Magnusson and T.E. Petersen (1984). J. Biol. Chem. *259*:8318-8327.

Sprengers, E.D. and C. Kluft (1987). Blood. *69*:381-387.

Stader, J.A. and T.J. Silhavy (1990). Method. Enzymol. *185*:166-187.

Starkey, P.M. and A.J. Barrett (1977). In: Barrett (Ed.) Proteinases in Mammalian cells and tissues. Elsevier/North-Holland, Amsterdam.

Stephens, A.W., B.S. Thalley and C.H.W. Hirs (1987). J. Biol. Chem. *262*:1044-1048.

Stormo, G.D. (1986). In: Resnikoff, W. and L. Gold (Eds). Maximizing gene expression. Butterworths, Boston.

Straey, S. and D. Crothers (1987). Cell. *51*:699.

Strauch, K.L., J. Johnson and J. Beckwith (1989). J. Bacteriol. *171*:2689-2694.

Strickland, S., E. Reich and M.I. Sherman (1976). Cell *9*:231-240.

Stroud R.M. (1974). A family of protein-cutting proteins. Sci. Am. *231(1)*:74-89.

Studier, F.W.and B.A.Moffat (1986). J.Mol.Biol. *189*:113-130.

Studier, J.J., A.H. Rosenberg, J.J. Dunn and J.W. Dubendorf (1990). Meth. Enzymol. *185*:60-89.

Sugita, T., H. Shinagawa, K. Makino and A. Nakata (1985). J. Biol. Chem. *99*:1247-1250.

Summaria, L., B. Hsieh and K.C. Robbins (1967). J. Biol. Chem. *242*:4279-4283.

Sung, W.L., F-L. Yao, D.M. Zahab and S.A. Narang (1986). Proc. Natl. Acad. Sci. USA. *83*:561-565.

Swaffield, J.C., I.J. Purvis and A.J.P. Brown (1987). Prot. Eng. *1*:228.

Swamy, K.H.S. and A.L. Goldberg (1982). J.Bacteriol. *149*:1027-1033.

Sweet, R.M., H.T. Wright, J. Janin, C.H. Chothia and D.M. Blow (1974). Biochem. *13*:4212-4228.

Swenson, R.P. and J.B. Howards (1979). J. Biol. Chem. *254*:4452-4456.

Tabor, S. and C.C. Richardson (1985). Proc. Natl. Acad. Sci. U.S.A. *82*:1074-1078.

Takahara, H, H. Okamoto and K. Sugawara (1985). J. Biol. Chem. *260*:8378-8383.

Talmidge, K. and W. Gilbert (1982). Proc. Natl. Acad. Sci. U.S.A. *79*:1830-1833.

Travis, J. and G. Salvesan (1983). Annu. Rev. Biochem. *52*:655-709.

Tschesche, J. and S. Kupfer (1976). Hoppe-Seyler's Z Physiol. Chem. *357*:769-776.

Tsunasawa, S., J.W. Steward and F.S. Sherman (1985). J.Biol.Chem. *260*:5382-5391.

Ullman, A., F. Jacobs and J. Monod (1967). J. Mol. Biol. *24*:339-343.

Urano, S., A.R. Metzger and F.J. Castellano (1989). Proc. Natl. Acad. Sci. USA. *86*:2568-2571.

Van de Werf, F., P.A. Ludbrook, S.R. Bergman, A.J. Tiefenbrunn, K.A. Fox, H. De Geest, M. Verstraete, D. Collen and B.E. Sobel (1984). N. Engl. J. Med. *310*:609-613.

Van Leuven, F. (1982). Trends Biochem. Sci. 7:185-187.

Van Zonneveld, A. (1986). J. Cell. Biochem. *32*:169-178.

Vaughn, D.E., P.J. DeClerck, M. DeMol and D. Collen (1989). J. Clin. Invest. *84*:586-591.

Vetterlein, D., P.L. Young, T.E. Bell and R. Roblin (1979). J. Biol. Chem. *254*:575-578.

Wallen, P., G. Pohl, N. Bergsdorf, M. Ranby, T. Ny and H. Jornvall (1983). Eur. J. Biochem. *132*:681-686.

Wallen, P., M. Ranby, N. Bergsdorf and P. Kok (1981). Prog. Chem. Fibrinolysis Thrombolysis. *5*:16-23.

Walsh, K.A. (1975). Unifying concepts among proteases. In: Reich, E., D.B. Rifkin and E. Shaw (eds). Proteases and Biological Control. Cold Spring Harbour Laboratory.

Walther, P.J., H.M. Steinman, R.L. Hill and P.A. McKee (1974). J. Biol. Chem. *249*:1173-1181.

Werb, Z., M.C. Burleigh, A.J. Barrett and P.M. Starkey (1974). Biochem. J. *139*:359-368.

Williams, D.P., D.Regier, D. Akiyoshi, F. Genbauffe and J.R. Murphy (1988). Nucl. Acid. Res. *16*:10453-10467.

Wilson, E.L., M.L.B. Becker, E.G. Hoal and E.B.D. Dowdle (1980). Cancer Res. *40*:933-938.

Wiman, B, L. Boman and D. Collen (1978). Eur. J. Biochem. *87*:143-146.

Wiman, B., A. Almquist, O. Sigurdardottir and T. Lindahl (1988). FEBS Lett. *242*:125-128.

Wolff, S.M. and J.V. Bennett (1974). N. Engl. J. Med. *291*:733-734.

Wun, T. W. Schleuning and E. Reich (1982). J. Biol. Chem. *257*:3276-3283.

Wun, T., L. Ossowski and E. Reich (1982). J. Biol. Chem. *257*:7262-7268.

Yanish-Perron, C., J. Vieira and J. Messing (1985). Gene *33*:103-109.

Zimmerman, M., J.P. Quigley, B. Ashe, C. Dorn, R. Goldfarb and W. Troll (1978). Proc. Natl. Acad. Sci. USA. *75:750-753*.