Novel genes for insect resistance in transgenic plants

Gatehouse, Laurence Neil

NOVEL GENES FOR INSECT RESISTANCE IN TRANSGENIC PLANTS

A thesis submitted by Laurence Neil Gatehouse, M.Sc. (Dunelm)

in accordance with the requirements of the University of Durham

for the degree of Doctor of Philosophy.

" Fe ddechreuwn a disgled o dê "

A. Plater

Department of Biological Sciences

September 1995

Abstract

From a cDNA library of potato tuber, cDNAs for the potato carboxypeptidase inhibitors were isolated and characterised by DNA sequencing. One full length clone of each type was used to make plant expression constructs, and these constructs used to transform tobacco. An insect bioassay, conducted using the self crossed progeny from the highest expressing transgenic line, revealed that the expression of potato carboxypeptidase inhibitor in tobacco increased the susceptibility of the plants to attack by *Heliothus virescens* .

A cDNA library of whole *Manduca sexta* larvae was constructed in λZAP II. Oligos were designed to fit the strongly conserved region of insect haemolymph trypsin inhibitor protein sequences and to a region of published protein sequence from *Manduca* haemolymph trypsin inhibitor A (MHTI A) and these were used to PCR a fragment of a MHTI A cDNA. This fragment was used to screen the cDNA library and a number of clones for MHTI A were isolated, along with cDNAs for a previously unknown related protein. These cDNAs were characterised by DNA sequencing.

One of the MHTI A cDNAs was used to make plant and *E. coli* expression constructs and these were sent, for subsequent bioassays of the resultant transgenic plants and of bacterially expressed protein, to Horticulture Research International. While these assays were seriously flawed, there were strong indications from both the plant and the artificial diet bioassays that MHTI expression did enhance insect resistance.

A cDNA library of whole *Diabrotica undecimpunctata* larvae was constructed in λZAP II. Oligos were designed to fit each of the three strongly conserved regions of protein sequence of mammalian and nematode microsomal aminopeptidases. These were used to PCR fragments from both *Manduca* and *Diabrotica* cDNA templates. These PCR products were characterised by DNA sequencing and the *Diabrotica* PCR products used to screen the cDNA library. Two cDNAs were isolated, neither of which were full length, but which were of sufficient length for protein expressed from them to be likely to be functional as an aminopeptidase. *E. coli* expression constructs were made from each cDNA and bacterial expression was demonstrated.

Pilot work on the feasibility of using antibodies as anti-insect proteins was conducted and the antibodies shown to be reasonably resistant to *Diabrotica* gut proteases. It was also demonstrated that antibodies could be produced that were active at the extreme pHs (3.5 - 11) found in insect guts. While many questions have been left unanswered, this project has successfully demonstrated the viability of such novel approaches to the enhancement of insect resistance in plants by genetic engineering.

Declaration

I declare that no part of this work has been submitted by me for any degree in this or any other university. All of the work presented was conducted by me, except where otherwise stated in the text.

Laurence N. Gatehouse.

Contents

Figures

Acknowledgements

I would like to thank the many people who have helped in the production of this thesis and the research on which it is based.

I grateful acknowledge the financial support of AXIS Genetics Company Ltd., Cambridge and thank the Department of Biological Sciences, Durham University for the use of their facilities.

The following people were all especially helpful and I grateful acknowledge their aid and support: Dr. Marta Evans for many useful discussions; John Gilroy; Dr. Angharad Gatehouse for help with the insect bioassay; Professor Boulter for the support he has given; Mrs. Julia Bartley for DNA sequencing; Dr. Dave Bown; Rachel Down; Dr. Kevin Powell; Gillian Davison for insect rearing; all the people at AXIS Genetics especially Dr. Andy Merryweather, Dr. Marian Longstaff and Dr. Bill Hamilton; Dr. Bill Deakin for becoming captain and everyone else who played 5-a-side football; my colleagues in New Zealand, especially Libby Burgess, who helped in the motivation stakes with more than just a deadline; Sue Maskell; Apple for manufacturing computers which did not drive me completely insane; everyone in Lab. D2 or Laboratory 1 (as it became), and my colleagues throughout the department.

In particular, I would like to thank Dr. Nick Harris for being my internal supervisor and apologise to him for the trouble and for the fact that the SCAB project failed to proceed to the point at which it would have become interesting for him; Dr. John Gatehouse for the many suggestions given and discussions we had over the various projects; Dr Heather Edmonds for the support during writing this thesis; and Dr. Vaughan Hilder for help with the statistics and the invaluable discussions over the design of the experiments and the strategy behind them.

Finally, I would like to thank my father for the help, both financial and moral, which he has given; Heather for all the support and for marrying me; and Heather's parents, Mike and Jen Edmonds, for everything.

Abbreviations

The following abbreviations were used;

aa	amino acid
α-AI	alpha-amylase inhibitor (from *Phaseolus vulgaris*)
AB	antibody
ABTS	2,2'-azino-di-(3-ethylbenzthiazoline-6-sulphonate) (ABTS™ Boehringer Mannheim Inc.)
ATP	adenosine triphosphate
AU	absorbance units
bp	base pairs
BAP	6-benzylaminopuridine
BCTI	bovine colostrum trypsin inhibitor
BPTI	bovine pancreatic trypsin inhibitor
Bt	*Bacillus thuringiensis* (delta endotoxins)
BSA	bovine serum albumin
cDNA	copy DNA
conA	concanavillin-A, jackbean lectin
cpm	counts per minute (of radioactivity)
CpTI	cowpea trypsin inhibitor
C-terminal	carboxy-terminal (of proteins and peptides)
CTP	cytidine triphosphate
d	deoxy-
Da	Daltons (molecular weight)
dd	dideoxy-
dd water	double distilled and deionised water
DEPC	diethyl pyrocarbonate
DEPC water	diethyl pyrocarbonate treated dd water
DMF	dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
ds	double stranded
DTT	dithiothreitol

E. coli	*Escherichia coli*
EDTA	ethylene diamine tetra-acetic acid
ELISA	enzyme linked immunosorbent assay
Enhance PM	Enhance Perfect Match (Stratagene)
EtBr	ethidium bromide
FAPP	N-(2-furanacryloyl)-L-phenylalanine-L-phenlyalanine
g	gravities (of centrifugation)
GNA	*Galanthus nivalis* agglutinin, snowdrop lectin
GTP	guanosine triphosphate
HCl	hydrochloric acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulphonis acid
HTI	haemolymph trypsin inhibitor
IPTG	isopropyl-ß-D-thiogalactopyranoside
kb	kilobase pairs
lac	ß-galactosidase gene system
MAP	microsomal aminopeptidase
MHTI A	*Manduca* haemolymph trypsin inhibitor A
MHTI B	*Manduca* haemolymph trypsin inhibitor B
miniprep	Plasmid minipreparation method or the product of it.
MOPS	3-(N-morpholino)propanesulphonic acid
MPI	*Manduca* protease inhibitor
MPI C	*Manduca* protease inhibitor C
Mr	molecular weight
mRNA	messenger RNA
NAA	napthaleneacetic acid
NCR	Northern corn rootworm
N-terminal	amino-terminal (of proteins and peptides)
OD_{xxx}	optical density at xxx nm
O/N	overnight literally or an *E. coli* colony grown up in 10 ml of broth plus selection if appropriate overnight
oligo	DNA oligomer

ORF	open reading frame (of nucleic acid sequence)
PAGE	polyacrylamide gel electrophoresis
PCI	potato carboxypeptidase inhibitor
PCR	polymerase chain reaction
PEG (xxxx)	polyethylene glycol (of average molecular weight xxxx)
pfu	plaque forming units (of phage)
PHA-E	erythro-agglutinating Phytohemagglutinin (from *Phaseolus vulgaris*)
PMSF	phenylmethylsulphonyl fluoride (a potent inhibitor of proteolysis)
pNA	paranitroanilide (derivative of)
pol	polymerase
polyA	polyadenosine
polyA+	of RNA with a polyA tail, i.e. mRNA
RACE	rapid amplification of cDNA ends
RF	replicative form
RNA	ribonucleic acid
RT	reverse transcriptase
S	Svedbergs
SCAB	single chain antibody
SCR	Southern corn rootworm
SDS	sodium dodecyl sulphate
seq.	sequence
ss	single stranded
TCA	trichloroacetic acid
TI	trypsin inhibitor (protein)
tris	tris-(hydroxymethyl)-aminomethane
tRNA	transfer RNA
TTP	thymidine triphosphate
u	unit (of enzyme)
UV	ultra violet (light)
v/v	volume per volume

WCR	Western corn rootworm
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indoyl-ß-D-galactopyranoside

Introduction

This thesis was originally intended to concentrate on the production of a single chain antibody (SCAB) library and the subsequent isolation of antibody lines which were deleterious to insects when incorporated into artificial diets (see this chapter, microsomal aminopeptidase). The research was funded by AXIS Genetics Ltd. Cambridge and the single chain antibody project, in particular, was thought to be commercially sensitive, consequently this thesis has been submitted on a restricted access basis. The collapse of the AXIS Genetics insect resistance program halted all work on the SCAB project and caused the other parts of the work to be wound up as rapidly as possible.

General introduction

Cultivated plants are the primary source for food production, with the only significant exception being the deep sea fisheries. They are also a major source of raw materials for many industries. Hence, agriculture is the industry on which most economic activity is ultimately based and efficient agricultural systems are the foundation on which civilisation is built. Historically many civilisations have declined or even vanished due to minor climatic changes or new plant pests or pathogens which made their agricultural practises unviable. There is a widespread popular belief that such events could not occur in modern times in developed countries. However, it has been predicted that global warming may cause widespread collapse of, at the least, the traditional agricultural systems. The effects of the mass migration of populations, in response to what would be a near global famine, may be imagined.

The rise in world population provides a strong incentive to optimise agricultural practises. One major constraint on agricultural production is due to the losses caused by pests and diseases. Figure 1 shows the world production and losses due to various pests of a number of important crop plants (Oerke 1994). This shows that over 40% of all primary agricultural production is lost due to pests, diseases and weeds. Losses of crop plants due to insect predation is approximately 15.1% overall (but ranges from 0 to 100%), that due to weeds is approximately 13.6% and that due to diseases approximately 13.5%. Insects, as well as acting directly as pests, are a major vector for diseases of plants, thus they are the major agent of crop plant losses world wide.

Worldwide losses in major cereal and non-cereal crops

Crop	Yield	Potential yield	Losses (%)		
	(1988-1990 million tonnes)		Insects	Weeds	Disease
Cereals					
Rice	548	1109	21.5	17.2	15.9
Wheat	545	848	9.3	13.1	13.3
Maize	428	766	14.5	13.1	10.8
Barley	169	272	8.8	11.0	9.9
Non-cereals					
Potato	268	458	16.1	8.9	16.3
Soyabean	99	150	11.2	12.9	9.2
Cotton	51	87	17.4	13.2	10.5
Coffee	6	10	14.7	10.7	16.0
Weighted mean			15.1	13.6	13.5

Figure 1. Estimated worldwide losses in a number of major cereal and non-cereal crops by cause of loss. Data from Oerke (1994).

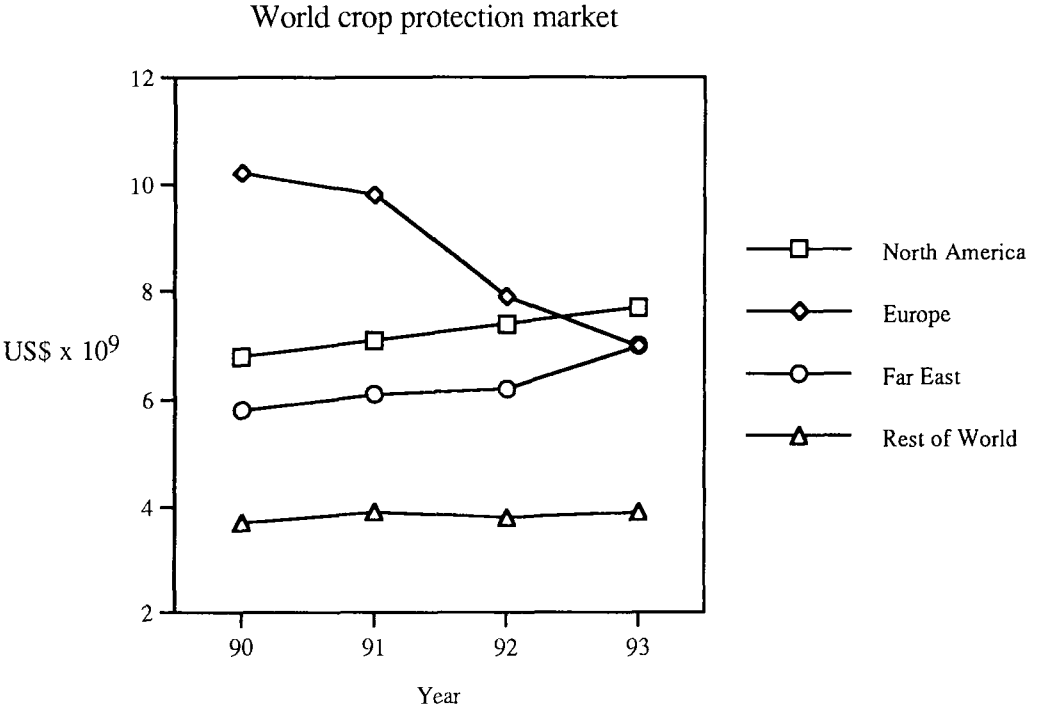


Figure 2. World crop protection market by region. Data from ECPA (1994).

The importance of these losses is perhaps best judged by the amount spent on crop protection agrochemicals; this was $25.6 billion in 1993 (ECPA 1994). This sum has been declining in recent years (figure 2) due to the decrease in usage in Europe because of the reforms in the common agricultural policy, but is rising elsewhere. This market is split, as shown in figure 3, with about 30% spent on insecticides, i.e. the world insecticide market was worth approximately $7.7 billion in 1993 (BAA 1994).

Agrochemical usage on this scale, apart from the expense, raises a range of concerns typified by the message of the green movement and the increasing popularity of "organic produce", i.e. that large scale usage of artificial chemicals is environmentally damaging. Whether these concerns are fully justified or not, the public perception of them dictates that alternatives to conventional protective agrochemicals must be developed.

A second constraint on the use of conventional agrochemicals for controlling pests is the development of resistance in target pests. It is not uncommon for examples of resistance in a major pest to be noted in the first year of use of a new pesticide (Metcalf 1986) and examples of resistance developing after long term usage are numerous. Furthermore, in some cases, the indiscriminate use of pesticides has, in some cases, worsened the problem of insect damage by eliminating the wide range of predatory species which normally play an important role in limiting the population of pests. An example of this is the rice brown plant hopper (*Nilaparvata lugens*) which was converted from a secondary pest of little importance, to a major pest with a greater impact on rice yield than the original pests (Heinrichs and Mochida 1983).

The elimination of the use of agrochemical pest control agents is, in the foreseeable future at least, not a realistic possibility, but a significant reduction in their use is both necessary and desirable. The route to this goal is to modify agricultural practices to produce integrated pest control programmes. These would combine judicious use of pesticides with crop rotation, field sanitation, the use of pest free seeds and most importantly the use of crop plant varieties which are inherently resistant to the pest species which occur in the location where they are to be grown (Meiners and Elden 1978).

Plants are a rich and diverse source of secondary compounds, many of which have been shown to have a protective role against various competitors, pests and pathogens. In recent years much effort has been put into the use of these for crop protection by both conventional breeding and by plant genetic engineering. However, historically, crop plants have been selected for high and dependable yield, palatability (to humans or as

	Percentage of total
Herbicides	45.9
Insecticides	30.0
Fungicides	18.7
others	6.4

Figure 3. Pest control agrochemical usage by product type for 1993. Data from BAA (1994).

fodder), low mammalian toxicity, nutritional value and adaption to certain environmental conditions. As a result few cultivated species are as inherently resistant to pests and pathogens as their wild relatives (Feeny 1976).

The breeding of crops for resistance has many advantages over the use of external control methods. The major ones being: The defensive factors are always present in the plant, hence the protection is provided when needed not when the infestation is detected by the farmer or on the schedule of when that field is to be sprayed; furthermore the pests are treated at their most sensitive stage. The protection is season long, independent of the weather, involves no application costs and protects those parts of the plant which are difficult to protect using conventional agrochemicals, for example the pod boring pest of oil seed rape *Ceutorhyncus assimilis* which attacks the developing ovules. The defensive factors are confined to the plant and do not leach into the environment and hence only those pests which attack the crop are exposed. This means that predatory species present on and around the crop plants will not be targeted and will, by predating pest species, add an extra layer of protection to the plants inherent ones. The active factors, if carefully chosen, will not be toxic to man and animals.

Public disquiet over the use of pesticides will be assuaged by the use of inherently resistant crop plants as pesticide usage on them will be lower than on susceptible varieties. However, public acceptance of plants genetically engineered for resistance is more problematical, with much resistance existing to the whole concept of genetic engineering. Whether this is resistance to the new or a more deep seated fear, only time will tell, but it is worth noting that for the only genetically engineered plant product to go on general sale to the public so far, Flavour Saver Tomatoes (Calgene) where the uronidase gene has been turned off by antisensing thus improving storage properties, demand has outstripped supply.

Finally there is the prospect of considerable financial savings to the farmer who would face a reduced pesticide bill which should more than offset the premium price seed companies would charge for pest resistant crop varieties. This reduction in overheads should eventually benefit the consumer and should, in conjunction with a "green" image of reduced pesticide usage, aid in the acceptance of the new genetic engineering technology.

Genetic engineering techniques will be required as part of a breeding programme to produce these new resistant crop plants as they allow desired genes encoding resistance factors to be transferred to plants quickly and without transferring other undesirable

characteristics. This allows the rapid development of new resistant varieties with otherwise the same properties as tried and tested ones. It also makes it feasible to produce resistant types of localised traditional varieties which, in more traditional agricultural systems, tend to be out performed by premium varieties with pesticide and fertiliser inputs, but are better adapted to local conditions and practises. Conventional breeding programmes make the production of varieties of this type uneconomic as the distribution would be too small scale and these traditional varieties are seen as little more than potentially useful sources of novel characteristics for breeding.

The other great advantage of genetic engineering over conventional plant breeding is that it allows traits which do not occur in the gene pool of that crop plant to be transferred in to them without the problems of species incompatibility. The genes for these traits can, in principle, come from any available outside sources. Although problems of incompatibility do sometimes arise (eg. poor expression of Bt in plants, poor processing of monocot introns in dicots) they can usually be overcome by judicious engineering of the gene sequences. The limitations of this approach are mainly technical: some plants cannot as yet be transformed and regenerated; the transfer of pathways, for example for secondary metabolite synthesis, requires many genes which are regulated in concert and is beyond the present state of knowledge; and much work is yet to be done on the identification and isolation of 'useful' genes. There are also regulatory and public acceptance barriers.

The present technology allows the transfer of genes which are coupled usually to a strong promoter and are not regulated, other than in a crude fashion for example by tissue type, with respect to other genes in the plant. Thus, the 'useful' genes which can be utilised at present are those whose action is direct via their primary protein product. It would also be possible to use antisense constructs which act to switch off endogenous gene(s) to, for example, disrupt a biochemical pathway to some secondary product which is an attractant or feeding stimulus to a pest of that species by antisensing one of the critical enzymes in the pathway. However, schemes of this type require detailed knowledge of these pathways which is only available in a limited number of cases. A further problem with this approach is that it is not just switching off the production of a single compound, but of every compound downstream of the antisensed enzyme step, and thus, the results would be extremely difficult to predict. This strategy is further complicated by the fact that many secondary metabolites which act as protective factors against most species of

phytophageous insects also act as specific attractants for others.

In order for insecticidal proteins to act upon an insect, they must be, in some way, taken up by the insect and hence must act in either the mouth, via taste receptors, or in the gut. This is in marked contrast with agrochemical insecticides most of which act, in the main, by binding to and disrupting the action of receptors for neurotransmitters. In fact 95% of all chemical insecticides act on one of three targets in the insect nervous system. It is possible to isolate genes which would express proteins which bind to these or similar targets in the insect nervous or endocrine system, for example the neuropeptide insect diuretic hormone. This would be expected, if released in quantity into the insect haemolymph, to disrupt the insect's water balance and seriously reduce its fitness. However there are no routes for getting a protein expressed in a plant into the haemolymph of insects where these targets are accessible. An alternative strategy here would be to use an engineered insect virus, such as baculovirus, carrying a gene for the anti-insect protein, as the delivery system. However, the use of viruses in this way not only removes the protection from the plant to an outside applied source, but is also subject to huge regulatory, safety and public acceptance problems.

Anti-insect genes

A wide range of single gene products have been investigated as potential insect resistance factors and these can be classified as follows:

Enzyme inhibitors

Enzyme inhibitors can be subdivided by their function into protease inhibitors, and inhibitors of other gut enzymes such as α-amylases, but their mode of action is essentially similar. They act, primarily at least, by binding to and inhibiting the digestive enzymes in the insect gut, this blocks the breakdown of complex molecules to simple absorbable ones and hence causes deleterious effects to the insects by reducing the absorption of nutrients. However, the actual mechanism of action is more complex; nutrient loss due to over production of digestive enzymes also plays a major role in the effects of this class of proteins and there are other less well understood effects.

The production of trypsin-like protease, the major proteolytic enzyme in

Spodoptera exigua and *Heliothus zea* , when these insects were fed 0.18% wet weight soybean trypsin inhibitor and potato inhibitor II in artificial diets (an amount thought comparable to that found in tomato, Gustafson and Ryan 1976) was observed to be greatly increased, and the insects showed reduced growth and development (Broadway and Duffy 1986). This is analogous to the effects of some protease inhibitors when fed to mammals (Leiner 1980) where hypertrophy of the pancreas and over production of digestive enzymes is observed. Broadway and Duffey (ibid) proposed that the deleterious effect on the insects was due to loss of sulphur-containing amino acids which were depleted by the increased trypsin synthesis. The nutrient loss hypothesis is supported by the observations of Gatehouse and Boulter (1983) where the antinutritional effects of cowpea trypsin inhibitor (CpTI) on *Callosobruchus maculatus* could be ameliorated by supplementation of their diet with methionine.

Broadway *et al.* (1986) suggest that enzyme inhibitors should be viewed as a single component of the complex interaction between the insect (and the physiology of its digestive processes) and the plant (and its nutritive value and other defences) which may explain why there are varieties of cowpea with high levels of CpTI which never the less are susceptible to infestation by bruchids (Xavier-Filho *et al.* 1989).

Burgess *et al.* (1994), in a study of the effects of a wide range of protease inhibitors on black field cricket (*Teleogryllus commodus*), report that low levels of inhibitors (in the 0.1% w/v region) had significant deleterious effects in diets with low to moderate levels of protein supplied as caesin, but at high protein levels greater amounts of protease inhibitors were required to show an effect and these effects were smaller. This parallels the observations of Broadway and Duffy (1988) in that the effectiveness of proteinase inhibitors as anti-insect proteins increases with diminishing protein quality.

However, Burgess *et al.* (1994) observed modest decreases in the activity of some proteolytic enzymes in the gut of black field cricket when fed the protease inhibitors at levels which were showing a significant deleterious effect, which contradicts the overproduction of enzymes hypothesis. This was in contrast to their own earlier results (Burgess *et al.* 1991) on the same insect which had shown increases in various protease activities with some of the inhibitors used in the later study. The differences in response may be explained by the duration of the trials or by the age of the insects (Grant, Dorward and Pusztai (1993) showed that young and old rats are very much more susceptible to the effects of diets containing protease inhibitors than mature animals). This emphasises that

the insects physiological response to a diet containing enzyme inhibitors is complex and involves feedback mechanisms which determine and adjust the enzyme levels in the gut.

There are many examples of protease inhibitors that have been shown to be active as inhibitors of the digestive enzymes of species of insects in *in vitro* experiments on gut extracts, and/or metabolically deleterious to insect species in artificial diet bioassay trials *in vivo* (Gatehouse *et al.* 1979; Hilder *et al.* 1987; Broadway and Duffy 1986; Burgess *et al.* 1991 and 1994; Johnson *et al.* 1989 and 1995; Edmonds 1994). Most studies have concentrated on inhibitors of serine proteases, which are the dominant activities in the guts of Lepidoptera (Applebaum 1985), with rather less studies of inhibitors of cysteine proteases, which are usually the dominant activities in the guts of Coleoptera (Colepicolo-Neto *et al.* 1987). This is due more to the greater abundance and relative ease of isolation of the former over the latter, rather than a greater interest in resistance against lepidopteran over coleopteran pests. Inhibitors of aspartic acid proteases (one has been reported in potato by Ritonja *et al.* 1990) and metallo-proteases (inhibitors of carboyxpeptidase-A have been isolated from potato and tomato, Rancour and Ryan 1968, Hass and Ryan 1980) have been little studied in their effects on insects and their gut enzymes. An inhibitor of the metalloprotease carboxypeptidase-A isolated from potato is one of the subjects of this thesis and is discussed in detail below.

There have also been a number of studies of inhibitors of enzymes other than proteases, mainly inhibitors of α-amlyase. Kidney bean (*Phaseolus vulgaris*) α-amlyase inhibitor, which is a protein closely related to the seed lectins, is an effective growth inhibitor of adzuki bean weevil (Ishimoto and Kitamura 1988). One of the wheat α-amlyase iso-inhibitors, designated 0.28, is a strong inhibitor of a number of insect α-amlyase enzymes without being active against the mammalian enzyme, and was significantly detrimental in artificial diet to *Callosobruchus maculatus* (Gatehouse *et al.* 1986).

Reviews on enzyme inhibitors of plant origin include those by Richardson (1977) and (1991), Bode and Huber (1992), Broadway *et al.* (1986), Fossum (1970), Green and Ryan (1972), Ryan (1990) and Xavier-Filho (1991).

One example of the production of insect resistant plants using an enzyme inhibitor gene is provided by the trypsin inhibitor from cowpea (CpTI). CpTI is a serine protease inhibitor of the Bowman -Birk family (Birk *et al.* 1963) with two inhibitory sites per molecule and is homologous to a wide range of proteins found in the Leguminosae and

Gramineae. It is a small protein of about 80 amino acids and is the product of a small gene family. Most isoforms of this protein are trypsin/chymotrypsin inhibitors but the one actually transfered to transgenic plants was a trypsin/trypsin inhibitor (Hilder *et al.* 1987).

It had been shown that the active constituent in a line of cowpea that was resistant to the cowpea weevil (*Callosobruchus maculatus*) was CpTI, which was present in the seed in this line at higher levels than in susceptible lines (Gatehouse *et al.* 1979). This protein was isolated and characterised (Gatehouse and Boulter 1980). Feeding trials in artificial diets incorporating purified CpTI at physiological levels showed this protein to be an effective anti-metabolic agent against a wide range of insects, including genera such as *Heliothus* , *Spodoptera* , *Diabrotica* and *Tribolium* , all of which cause losses of major economic importance (Gatehouse and Boulter 1983). Furthermore although normally cooked, cowpeas can be eaten raw (Peterson 1984), and in feeding trials rats fed raw cowpea meal showed no abnormalities in growth (Hilder *et al.* 1987).

A cDNA for CpTI was isolated (Hilder *et al.* 1989), used to make a plant expression construct, with the CpTI gene being driven by the strong constitutive CAMV35S promoter, and introduced into into tobacco (Hilder *et al.* 1987). The best expressing plant obtained, which expressed at about 1% of total soluble leaf protein (an amount roughly equal to the concentration of CpTI in the seeds of the resistant line of cowpea) was significantly more resistant towards *Heliothus virescens* larvae in whole plant bioassays. Later bioassays also showed enhanced levels of resistance to a wide range of insect pests including *Manduca sexta* and *Spodoptera littoralis* .

Lectins

Lectins (also called agglutinins) are carbohydrate-binding proteins of non-immune origin that agglutinate cells or precipitate polysaccharides or glycoconjugates (Goldstein *et al.* 1980). They occur in many plant species and in many organs and tissues, although the highest concentrations are found in seeds and other storage tissues. Most of the lectins characterised, are secretory proteins which accumulate in vacuoles, intercellular spaces and the cell wall.

Structurally they are a diverse group of proteins although many lectins belong to an homologous family of proteins based on an amino acid chain of approximately 220 residues as typified by soybean lectin (Lotan *et al.* 1975). Chrispeels and Raikhel (1991)

suggest that many lectins have evolved by gene duplication and divergence, as is indicated by the lectin and lectin-like genes in kidney bean (*Phaseolus vulgaris*) which include the two lectins isoforms phytohaemeagglutinin L and E, the α-amylase inhibitor and the arcelin group of storage proteins. Alternatively, they suggest, some lectins may have evolved by incorporation of lectin domains into other genes, as is suggested by the presence of the characteristic lectin domain which is found in hevein, wheat germ agglutinin, rice lectin, tomato lectin and nettle lectin.

Lectins are usually classified by their sugar binding specificities, usually by their binding to a single residue, for example snowdrop lectin (*Galanthus nivalis* agglutinin GNA) is mannose specific. However, it is well known that binding to oligomers and polysaccharides is often much stronger than to the monomer, and sugar specificity has not proved to be a guide to the insecticidal activity of a given lectin.

Toxicity of members of this group of proteins in mammals (Evans *et al.* 1973, Leiner 1980) and birds (Jayne-Williams and Burgess 1974) has been well documented and detailed studies of mechanism of action have been undertaken (Pusztai, Clarke and King 1979. King, Pusztai and Clarke 1980a and 1980b, de Olivera 1986).

A role for lectins in the defence of plants against insects was first proposed by Janzen, Juster and Leiner (1976) who demonstrated the toxicity of *Phaseolus vulgaris* haemeagglutinin in artificial diet to the bruchid beetle *Callosobruchus maculatus*. This was confirmed by Gatehouse *et al.* (1984) although they found that a highly purified preparation of the lectins was less effective in retarding insect development than a more crude preparation. The contaminating protein was shown to be the α-amylase inhibitor (Huessing *et al.* 1991), which is structurally very similar and so would be expected to co-purify with the lectins. However, Murdock *et al.* (1990) showed that rigorously purified *Phaseolus vulgaris* haemeagglutinin was not toxic to *Callosobruchus maculatus* larvae.

Lectins found to be active against *Callosobruchus maculatus* include wheat germ agglutinin, nettle lectin and rice lectin (Huessing *et al.* 1991a and 1991b) which all caused marked retardation of larval development, while thornapple and tomato lectins had a lesser effect (Murdock *et al.* 1990). *Ostrinia nubilalis* (European corn borer) and *Diabrotica undecimpunctata* (Southern corn rootworm) were found to be highly sensitive to wheat germ agglutinin, castor bean lectin and *Bauhinia purpurea* lectin, but insensitive to lectins from kidney bean, garden pea, soybean and mistletoe (Czapla and Lang 1990). They also observed anti-insect activity against southern corn rootworm for lectins from

pokeweed, *Codium fragile* , *Mandura pomifera* , *Baneiraea simplicifolia* (lectin II), Jackfruit, hairy vetch and elderberry.

Powell *et al.* (1993) observed, in a test with eight lectins, activity against *Nilaparvata lugens* (rice brown plant hopper) and *Nephotettix cinciteps* (rice green leaf hopper) for the lectin from snowdrop (*Galanthus nivalis* agglutinin GNA), and against rice brown plant hopper only for wheat germ agglutinin. Gatehouse *et al.* (1991) showed that winged bean lectins are involved in resistance of the seeds of this plant to non-pest species. Shukle and Murdock (1983) showed that soybean lectin was toxic to *Manduca sexta* larvae. Edwards (1988) showed that expression of pea lectin in transgenic tobacco resulted in enhanced resistance to *Heliothus virescens* larvae.

The mode of action of lectins in insects is thought to similar to that in animals. The lectin binds to suitably glycosylated targets in the insect gut or mouth parts. Binding to receptors in the mouthparts may explain the antifeedant effect of some lectins such as GNA against rice brown plant hopper (Powell *et al.* 1995). Ultrastructural studies (Gatehouse *et al.* 1984, Sauvion 1995, Powell 1993) have shown that lectins active against an insect can be detected bound to the midgut epithelial cells and a range of structural changes analogous to those seen in mammals can be observed, such as changes in the length of the microvilli in the brush border membrane and disruption of the epithelial cell layer. This disruption leads to breakdown in the transport of nutrients and potentially allows the absorption of harmful substances as well as leakage of nutrients back into the gut.

Toxins

The crystal delta-endotoxin proteins of the bacterium *Bacillus thuringiensis* and the ribosome inactivating proteins (RIPs) of plant origin, are both toxins of single gene origin which have been investigated in detail as anti-insect proteins.

Bacillus thuringiensis delta endotoxins

The gram positive bacterium *Bacillus thuringiensis* produces proteins which are toxic to a variety of insect species. Spore preparations of the bacterium have been used for about the last thirty years as a biological insecticide (Dulmage 1980), and *B. thuringiensis* is the current market leader bio-insecticide. The insecticidal activity resides in crystalline

inclusion bodies, produced during sporulation of the bacteria, which are are composed of proteins (termed delta endotoxins) specifically toxic to a variety of insects. Different strains of *B. thuringiensis* differ in their spectra of insecticidal activity. Most are active against Lepidoptera, but some strains are specific to Diptera (Yamamoto and McLaughlin 1981; Goldberg and Margalit 1977) and Coleoptera (Herrnstadt *et al.* 1986; Krieg *et al.* 1983) are known. Some strains have also been reported as active against nematodes (Bone *et al.* 1988; Freitelson *et al.* 1992). In insects the crystals dissolve in the alkaline conditions of the mid gut and release proteins of relative molecular weight 65 to 160 kDa (ibid and Calabrese, Nickerson and Lane 1980); these are proteolyticly processed by midgut proteases to yield smaller toxic fragments.

Bt insect toxins are highly specific, that is they are not toxic to other organisms including mammals, and hence are potentially suitable for use as plant protection agents in food plants. This potential was first realised by Vaeck *et al.* (1987) who used a cloned Bt toxin gene to make expression constructs which were transformed into tobacco. The constructs which contained truncated versions of the gene expressed better than the whole gene, with the best expressers (0.004% of total leaf protein) able to kill all the *Manduca sexta* larvae put on them in within three days. Resistance was shown to be correlated with level of expression of Bt toxin protein and the Bt expressers were shown to be significantly more resistant to *Manduca sexta* larvae than control plants in all parameters measured.

In order to achieve the protein expression levels needed to give protection against pests less sensitive than *Manduca sexta* , much work was undertaken to improve the promoters and add enhancer elements to the expression constructs. However, it was found necessary to reconstruct the truncated toxin gene to alter the codon usage without altering the protein sequence to achieve the desired levels of expression (Perlak *et al.* 1992).

The target of Bt toxins in susceptible insects is the cells of the mid gut epithelium where the toxin, after proteolytic processing to the active fragments, inserts into the plasma membrane but not into the cytoplasm (Lüthy and Ebersold 1981, Gill *et al.* 1992). The sequence of events is: 1-5 minutes, increased glucose uptake by the gut cells and first histopathological signs in the columnar cells; 5-10 minutes, midgut paralysis, cessation of feeding increased blood pH and decreased lumen pH, apical membrane becomes permeable to dyes, swelling of the columnar cells, blebbing of the microvilli and first histopathological signs in the goblet cells; 10-30 minutes, increased potassium turnover in

the gut cells, increased blood potassium concentration, a decrease in transport of glucose and leucine to the blood and general metabolic breakdown of the gut cells; 30-60 minutes, gut cells lyse and slough from the basement membrane and the insect becomes paralysed; 1-3 days, death by starvation and/or septicaemia if the insect had eaten *B. thuringiensis* spores (Knowles 1994).

The Bt toxins are known to bind with high affinity to specific receptors found on the brush border of the columnar cells lining the insect mid gut, with the pattern of binding being highly complex. A given toxin may have more than one binding site in a single insect and more than one toxin may compete for binding to a single site (Hoffman *et al.* 1988, Van Rie *et al.* 1989 and 1990). After binding, the toxins insert rapidly and irreversibly into the plasma membrane of the gut cell (Van Rie 1989), resulting in the formation of a pore or lesion in the membrane. How this pore leads to cell lysis is disputed, osmotic lysis by uptake of water through the pore, or the pore acting as an ion channel which leaks potassium causing the cell to depolarise and the pH to rise hence killing the cells, are the main hypotheses (Knowles 1994).

The identification of two receptors for Bt toxins in *Manduca sexta* gut, an aminopeptidase-N (Knight, Crickmore and Ellar 1994) and a cadherin (a protein involved with cell-cell recognition and cell-cell binding) (Vadlamudi *et al.* 1995), supports the complex pattern of binding of the toxins observations (above).

Resistance to Bt toxins has been reported for a number of species in the laboratory, including many of economic importance such as the colorado beetle (*Leptiontarsa decemlineata*)and the diamondback moth (*Plutella xylostella*). However, only the latter has been reported to have developed high levels of resistance in the field as a consequence of repeated use of Bt (Shelton and Wyman 1992), although the indian meal moth (*Plodia interpunctella*) is thought to have evolved low levels of resistance in stored grain treated with Bt (McGaughey and Whalon 1992).

The use of single toxin genes in transgenic plants, as opposed to the mixture of toxins produced by the bacterium, will make the evolution of resistance easier and although Bt toxin genes are the most effective transgenes for insect resistance known, a strategy relying on them alone would be ill advised.

Ribosome inactivating proteins

RIPs are proteins which specifically remove a single adenosine from the conserved stem-loop structure present in all 23S, 26S and 28S ribosomal RNAs. This results in the permanent inactivation of the ribosome and kills the cell as protein synthesis is blocked. RIPs are widely distributed throughout the plant kingdom and fall into two classes; type I RIPs which consist of a single polypeptide chain which carries the RIP active site, and type II RIPs which consist of two polypeptide chains linked by disulphide bonds, one being the RIP the other being, in all cases so far characterised, a galactose binding lectin.

In mammals the type II RIPs are the more potent of the two as the lectin subunit binds to targets on the cell surface aiding the entry of the RIP subunit into the cell where it is active. Ricin a type II RIP from castor bean (*Ricinus communis*) is extremely poisonous it is estimated that one molecule is sufficient to kill a cell. The type I RIPs typified by saporin (from *Saponaria officinalis*) are also highly toxic though less so than the type II RIPs.

Gatehouse *et al.* (1990) have reported the extreme toxicity of both saporin and ricin to two species of Coleoptera, *Callosobruchus maculatus* and *Anthomonas grandis* , with LD_{50} values of less than 0.01% dry weight and curiously no apparent difference between the toxicity of the two types of RIP. In the same report, no toxicity against Lepidoptera was observed, suggesting that the species tested are able to hydrolyse the RIPs in their alkaline gut before entry to cells is achieved.

The outstanding problem with the use of RIPs as plant protection proteins is their extreme mammalian toxicity and unless classes of RIPs can be found, or engineered, that are insect specific they are generally considered to be too dangerous to be released in plants intended for mammalian consumption.

Enzymes

A number of enzymes have been investigated as potential anti-insect proteins. Examples include chitinase, lipoxygenase and cholesterol oxygenase. In each case the enzyme was selected for its potential ability to attack some component of the insect gut since, it was assumed, this would be deleterious to the insect. A second consideration for the use of such enzymes as defence proteins in transgenic plants is that they either have no substrate in the plant, for example chitinase, or that they can be produced as a proenzyme or compartmentalised, so that they do not harm or act in the plant.

Powell *et al.* (1993 and 1995) found lipoxygenase, in artificial diet bioassays, to have a significant deleterious effect against rice brown plant hopper. However, the use of lipoxygenase as a transgene for insect resistance would affect fatty-acid oxidative metabolism, as was observed by Deng *et al.* (1992). An introduced soybean lipoxygenase gene driven by a strong promoter in tobacco greatly increased the production of C_6 aldehydes, although overall lipoxygenase activity in leaves was unaffected as the introduced protein was rapidly turned over. These enzymes are also associated with the development of objectionable beany, grassy and green flavours in soybeans (Kitamura, Kikuchi and Harada 1987). Thus any strategy involving this enzyme would require cellular compartmentalisation and probably tissue specific expression of the introduced enzyme.

Chitinase is one of the proteins whose synthesis is induced by wounding or by ethylene, and is associated with defence against fungal and bacterial infection in plants (Boller *et al.* 1983, Chrispeels and Raikhel 1991). Since chitin does not occur in plants and many plants can accumulate chitinase proteins with no apparent penalty, a suitably chosen plant chitinase gene would not be expected to adversely affect a plant in which it was expressed. Chitin is a major component of insects including the peritrophic membrane in their guts. Thus it could be speculated that it might have a deleterious effect on insects to which it has been fed.

We have recently demonstrated that different lines of potato plants expressing bean chitinase have a significant deleterious effect on aphids (*Myzus persicae*) compared to controls (R. Down personal communication; the chitinase gene was previously isolated by myself, unpublished work).

The insect gut contains a mixture of proteases, only some of which can be bound by a single inhibitor, and other potentially useful anti-insect proteins may need protection against the insect gut proteases to be effective. So for practical use anti-insect genes will be used in combinations which are complementary in action. For example Boulter *et al.* (1990) found that the effects of pea lectin and cowpea trypsin inhibitor expressed together in transgenic tobacco were additive, and Burgess *et al.* (1994) found that a combination of potato inhibitor I and wheat germ inhibitor I, which are inhibitors of SAAPLpNA hydrolysing enzymes (elastase-like) and trypsin respectively, had the same effect as potato inhibitor II, which is an inhibitor of both these proteases, at the same level of

incorporation in artificial diets.

This principle is well established in nature, where plants rely on multiple strategies and use a wide range of compounds and proteins for defence. Many defence-related proteins are multifunctional, for example, the type II RIPs, which have much higher mammalian toxicity than the type I RIPs consist of a RIP and a lectin (see above), and the kidney bean chitinase is a multi-domain protein consisting of a chitin-binding lectin, a proline-rich hinge region with similarities to the extensins (hydroxyproline rich cell wall structural proteins), and the chitinase domain (see above).

Potato carboxypeptidase inhibitor

Potato carboxypeptidase inhibitor (PCI) is a small protein of molecular weight 4300 Daltons which is found mainly in potato tubers and, to a lesser extent, throughout the plant (Hass, Derr and Makus 1979). Its function in tubers is thought to be both as a storage protein and as a defence mechanism, the latter by arresting the digestive proteolysis of invading pests (Hass and Ryan 1981). Supporting evidence for the defence role of PCI is given by its localisation in the cortical tissue (surface layers) of tubers (Graham and Ryan 1981), which is the initial point of attack by any invading pest.

Mechanical wounding of potato plant leaves induces an increase in expression of PCI both in the wounded leaves and in unwounded ones (Graham and Ryan 1981). This is similar to the wound induction of the serine protease inhibitors in potato, and this response is effected by the same wound hormone, called the proteinase inhibitor inducing factor. This observation also supports the defence function argument.

PCI is localised at the cellular level in the central vacuole of leaf cells (Holländer-Czytko, Andersen and Ryan 1985), which is again similar to the potato serine proteases. The expression in tobacco at quite high levels (>1%) of a number of serine proteases, and other proteins, which are localised to the vacuole, for example cowpea trypsin inhibitor and soybean Kunitz trypsin inhibitor, has been observed to have little or no effect on the fitness of the plants (V. A. Hilder personal communication). Thus, it would not be expected that expression of PCI in tobacco and its subsequent accumulation in the vacuoles, would be deleterious to the plants.

From the protein sequence data of Hass *et al.* (1979) PCI is known that at least two genes code for PCI. There are five observed isoforms but these, they pointed out, can be

derived from only two genes by different post-translational processing.

PCI has a tightly folded, cysteine cysteine bonded, three dimensional structure of great stability (Leary *et al.* 1979). It is resistant to temperatures up to 90°C and extremes of pH (pH 2-11.5) for extended periods. Even in the tuber it is resistant to denaturation by cooking and intact, functional PCI protein can be extracted from commercial instant mashed potato flakes (Pearce and Ryan 1983). Such stability is ideal for a protein intended as an insect defensive agent as it would be expected to survive the conditions of the insect gut. Its ability to survive cooking may have raised questions over its safety when fed to humans and other mammals except that potato has been a staple food crop for centuries and no ill effects from eating cooked tubers has been reported (apart from those which have been exposed to light and "gone green". The green portions of potatoes produce alkaloids.).

PCI inhibits carboxypeptidase A by a one-to-one binding mechanism (Ako, Hass, Grahn and Neurath 1976; Hass and Ryan 1981), and accurate X-ray crystal structures are known for both the inhibitor and for its complex with carboxypeptidase A (Rees and Lipscomb 1982). The mechanism of action is that the C-terminus of PCI is bound into the active site of the carboxypeptidase enzyme in a manner which mimics the enzyme substrate interaction. Indeed the PCI C-terminal glycine (if present) is slowly cleaved off and remains trapped in the enzyme active site pocket. Unlike a substrate however, the inhibitor remains bound to the enzyme and the strength of binding is not significantly affected by the loss of the C-terminal glycine. Iso-inhibitors both with, and lacking the C-terminal glycine, are effective inhibitors of carboxypeptidases.

PCI has been found to be an inhibitor of carboxypeptidase enzymes from a wide range of animals (Hass and Ryan 1981). These include bovine carboxypeptidases A and B, porcine carboxypeptidases A and B, shrimp carboxypeptidases A and B, a carboxypeptidase from lungfish, a carboxypeptidase from limpet, and a carboxypeptidase from a cabbage looper caterpillar (exact species unspecified). This last observation suggests that as insect carboxypeptidases are susceptible to inhibition by PCI, expression of this protein *in planta* will affect the digestive processes of predating insects. No plant-derived carboxypeptidase activities have been found that are inhibited by PCI, thus PCI is unlikely to interfere with the plants normal metabolism if expressed *in planta* .

A protein homologous to potato carboxypeptidase inhibitor has been isolated from the fruit of the related species tomato (Hass and Ryan 1979). It was found to be coded for

by a single gene (Martineau, McBride and Houck 1992). Wounding was found to give strong (greater than 100 fold) induction in leaves at the RNA level, but no changes in protein level were detected, suggesting a strong post-translational or post-transcriptional control of this gene. The protein in the fruit was apparently all synthesised early in development as RNA levels are very high in anthesis stage ovaries but decrease rapidly during fruit development. The accumulation of inhibitor protein in this organ matches the RNA levels. This pattern of expression does not seem defence related, particularly the failure of wound induced up-regulation of the mRNA level to be reflected in increased inhibitor protein synthesis. It suggests that in tomato, if not potato, carboxypeptidase inhibitor protein is involved in some developmental process. However, to date, no carboxypeptidases identified in solanaceous plants have been found to be inhibited by these proteins (Walker-Symmons and Ryan 1980; Ryan 1981; Martineau, McBride and Houck 1992).

Manduca haemolymph trypsin inhibitor

The partial protein sequences of two trypsin inhibitor proteins from the haemolymph of *Manduca sexta* have been published (Ramesh, Sugumaran and Mole 1988) and are homologous to the group mammalian trypsin inhibitors typified by bovine pancreatic trypsin inhibitor BPTI (Kunitz) (Lorand 1976, Laslowski and Kato 1980). They have, however, closer homology to another typical protein of this family, bovine coloestrum trypsin inhibitor (BCTI) (Cechova and Muszynska 1970, Cechova, Jonakova and Sorm 1971, Cechova and Ber 1974).

By searching the national protein database at Daresbury using the gcg package of programs (gcg 1994), a large number of other related protein sequences were found mainly from mammals but including some insect proteins; three chymotrypsin inhibitors from silkworm (*Bombyx mori*) (Sasaki 1978, Eguchi, Haneda and Iwamoto 1982, Sasaki and Kobayashi 1984), a hypothetical acrosin inhibitor from *Drosophila* (Schmidt *et al.* 1989) and a protein inhibitor from *Sarcophaga bullata* (flesh fly) (Papayannopoulos and Biemann 1992). The more homologous proteins from other orders tended to be multidomain proteins where one or more of the domains were of the BPTI/Kunitz type. They included proteins from Red Sea turtle eggwhite; two domains the first of which is of the BPTI/Kunitz type (Kato and Tominaga 1979), chicken α3 chain of

collagen VI; complex multiple domains with a mosaic structure at the carboxy-terminus region consisting of glycoprotein Ib-like, fibronectin type III and BPTI/Kunitz modules (Bonaldo and Colombatti 1989), a number of mammalian tissue factor proteins involved in the inhibition of blood clotting; three homologous repeats of BPTI/Kunitz domains (Enjyoji *et al.* 1992, Wun *et al.* 1988, Van der Logt, Reitsma and Bertina 1991, Wesselschmidt, Girard and Broze 1990, Warn-Cramer, Broze and Komives 1992) and bovine coloestrum trypsin inhibitor (ibid) which, like the insect proteins, consists of only a single domain.

The sequences have a number of conserved regions, particularly around the cystein residues which, by disulphide bonding, form the loops which give the molecule its stability against disruption by cleavage of the reactive bond (Laslowski and Kato 1980).

The reactive bond is between residues 18 and 19 for BCTI, and is in the equivalent position in other homologous proteins (Laslowski and Kato 1980), just downstream of the second cysteine residue of the mature protein. The amino acid residue on the 5' side of the reactive bond, residue 18 for BCTI, generally determines the specificity of the serine protease the inhibitor is active against with argenine (R) or lysine (K) in those inhibitors active against trypsin or trypsin-like enzymes (Laskowski and Sealock 1971) as is seen in both MHTI A and B. Those inhibitors with tyrosine (Y), phenyalanine (F), tryptophan (W), leucine (L) and methionine (M) in this position inhibit chymotrypsin and chymotrypsin-like enzymes and often subtilisin also, and those with alanine (A) or serine (S) inhibit elastase-like enzymes. These rules are very rough and exceptions are known (Laskowski and Kato 1980).

Ramesh, Sugumaran and Mole (1988) speculated that the physiological function of the MHTIs is the suppression of the activation of the haemolymph phenoloxidase, a key enzyme responsible for the melanisation and killing of invading foreign organisms. Activation of this enzyme occurs by the proteolytic cleavage of a proenzyme by a serine protease present in both haemolymph and cuticle (Ashida and Dohke 1980). High levels of haemolymph phenoloxidase would lead to unwanted self melanisation. Thus phenoloxidase activity is controlled via the MHTI control of the activating enzyme. Other functions of MHTI may be postulated by analogy to bovine pancreatic trypsin inhibitor where the function is to inhibit self activation of trypsinogen to trypsin by self proteolysis in the pancreas (Lorand 1976, Laslowski and Kato 1980). Insects lack a pancreas but might use the haemolymph trypsin inhibitors as a mechanism to prevent digestive

proenzymes which leak into the haemolymph from becoming activated, or to suppress active digestive enzymes which have leaked from the gut. A consequence of this hypothesis would be that the range of haemolymph inhibitors in a given species should at least cover the range of gut serine protease enzymes found in that insect.

Microsomal aminopeptidases and the single chain antibody project

The original single chain antibody project which evolved into the microsomal aminopeptidase (MAP) project was to be a novel method of generating anti-insect genes for expression in plants. The objective was to raise antibodies against total insect gut protein. Mice would be immunised with insect gut preparations and then would be used to make spleen RNA. This RNA would be used to make cDNA and the heavy and light antibody variable regions would be amplified from it by PCR. Eventually an M13 type phage library expressing the antibodies as randomised single chains, essentially the heavy and light variable regions joined together in a random combination by a linker, translationally fused to the end one of the phage coat proteins, would be produced.

This system means that each phage carries an antibody as part of its coat protein, in which form it is active, and the gene for the antibody is carried in the phage genome (Pharmacia P-L Biochemicals Recombinant Phage Antibody System Data sheet). Useful antibodies could be selected from the library and the selected antibody genes minus the coat protein could be engineered into plants by the usual methods. The single chain antibody (SCAB) genes were to have been selected by the ability of the phage line carrying them to kill or seriously slow the development of a target insect when fed to them in artificial diet.

The strength of this scheme was that no assumptions about the targets of the antibodies are made, the antibodies were to be selected purely for their deleterious effect on insects when fed to them. This means that completely novel gut proteins could be selected as targets. The draw back of this scheme is the screening problem.

The immune system of an individual is capable of expressing more than 10^7 different combinations of light and heavy chain (Harlow and Lane 1988, Kimball 1983) and the randomising step in the recombinant SCAB library preparation means that a library approaching this size would be needed as relatively few lines could be expected to be effective as anti-insect factors. The library would be biased towards gut protein binding SCABs by the original inoculation but the vast majority of antibodies raised would bind to

structural proteins or proteins not located on the gut inner surface. The randomising of the combination of light and heavy chains would produce novel combinations which did not occur in the individual. This increases the chances of producing a SCAB with the desired characteristics but at the price of producing huge numbers of SCAB lines with characteristics of no interest.

It is deemed likely that any antibody eventually identified in the screening would be very insecticidal, at least as effective as Bt. The prime reason for the lack of commercial interest in plant-derived insect resistance genes was seen to be their low (but statistically significant) level of effect compared to Bt. Thus any positive results from this proposed screening programme should be exceptionally valuable.

The difficulties of such a screening programme are, however, so great that it was considered that some initial demonstration of the potential of the SCAB technology was needed to justify it. A specific target antigen in the insect gut was, therefore, identified. Microsomal aminopeptidase (MAP) was selected as a handle to use to screen this SCAB library.

MAPs are a family of integral membrane proteins present in the gut wall whose function is the digestion of polypeptides to amino acids; they are also present in a number of other tissues where their function is less clear (Vallee and Auld 1990, Olsen *et al.* 1988, Taylor 1993a, Taylor 1993b). The protein has a short N-terminal, membrane-spanning, region and a long C-terminal region which contains the enzyme active site and which is accessible to the gut contents. It is located in the brush borders of the gut epithelial membrane (Maroux and Feracci 1983). MAPs represent an important type of digestive enzyme, and it has been shown that MAP enzyme activity could be blocked by monoclonal antibody lines (Ashmun, Shapiro and Look 1992). Hence some of the SCAB lines against MAP would block aminopeptidase activity. This could be expected to be deleterious to a target insect as the inhibition of other clases of digestive enzyme has been shown to be deleterious to a range of insects (see the introduction Enzyme inhibitors).

I also had, through AXIS Genetics Ltd., confidential information about a patent application (Munn 1992). This shows that antibodies raised in sheep against the microsomal aminopeptidases of a nematode *Haemonchus contortus* protected the sheep against infection with these blood-living parasites. The degree of protection was correlated with the titre of the antibodies, and protective antibodies were also raised by inoculation with a recombinant protein expressed by the bacculo virus system in insect cells.

The above was considered to be sufficient grounds to decide on insect gut aminopeptidase as a specific target antigen for the SCAB approach. This selection appeared to be vindicated by the subsequent identification of aminopeptidase-N, which is a microsomal aminopeptidase, as the target protein for binding of the *Bacillus thuringiensis* Cry1A(c) delta-endotoxin in *Manduca sexta* gut brush border membrane (Knight, Crickmore and Ellar 1994; Knight, Knowles and Ellar 1995). However, it has also been shown that a cadherin like protein specifically binds the Bt Cry1A(b) toxin in *Manduca sexta* gut (Vadlamudi, Ji and Bulla 1993; Vadlamudi *et al.* 1995). That the different Bt toxins bind to different receptors is hardly unexpected, however the observation that there are a range of potential binding targets in the insect gut supports the random SCAB library approach to generating novel anti-insect genes.

Insect MAP would be produced, either from insect guts or by over expression of a cloned insect MAP gene in a heterologous system. This protein would be used as a specific antigen to immunise mice for the production of anti-MAP antibodies and for the production of anti-MAP SCABs and to provide antigen for use in the panning of such a specific and the proposed random gut protein SCAB library. A proportion of the selected MAP binding SCAB lines would block MAP enzymatic activity and hence should be deleterious to the target insect. This 'belt and braces' approach justified the considerable investment in resources that the SCAB library would represent.

Microsomal aminopeptidases have a wide range of names, reflecting their original isolation and preferred amino acid for enzymatic activity, although often the same, or very homologous, proteins from the same organism were isolated independently and given different names, frequently without the aminopeptidase activity being detected. The names frequently used for mammalian proteins are gp150 and 160 (glycoprotein 150 and 160), CD13 (cell determinant 13) and a number of other cell-differentiation, antigenic determinants, aminopeptidase-N or leucine aminopeptidase (N for neutral and leucine in reference to their favoured substrates which are hydrophobic amino acid residues.) and aminopeptidase-A (acid aminopeptidase as they have greatest activity against aspartic acid residues.).

Microsomal aminopeptidases are characterised by a number of shared features; They are membrane bound by means of a short transmembrane domain near the N-terminus with the N-terminus inside the cell and the major part of the molecule, including the active site and the C-terminus, on the outside of the cell. There are a number of homologous

soluble proteins which lack this N-terminal anchor but none have been reported from higher organisms (Vallee and Auld 1990; Taylor 1993). They bind, at their active sites, a single zinc ion which is required for enzymatic activity. This zinc ion is bound via the histidines (H) in the conserved sequence AHELAHQWFG and the glutamic acid (E) in the conserved sequence LWLNEGFA located 20 amino acids down stream of the second histidine (see figures 16 and 18 in the MAP results chapter). A number of other classes of enzymes also use a HExxH motif in their active sites (Jongeneel, Bouvier and Bairoch 1989) which is characteristic of zinc binding. They are highly and variably, depending on where they are expressed, glycosylated (Munn 1992; Wu *et al.* 1990); the deglycosylated protein usually runs at ~110 kDa on SDS PAGE whereas the glycosylated forms run 30 to 50 kDa larger. Finally they are homologous to each other with regions of conserved amino acids occurring throughout their sequences (see figure 78). Their function in tissues other than gut is usually ascribed to processing signal peptides and as surface determinants of cell types involved in recognition and control of growth and differentiation (ibid).

Diabrotica species and the southern corn rootworm (*Diabrotica undecimpunctata*)

The genus *Diabrotica* is largely neotropical and includes some of North America's most destructive and economically important insect pests. The larval stage is most pernicious in the case of maize, causing reduction in yield both by actual damage to the root system and through fallen plants which cause harvesting problems. Adults also cause damage, feeding on the aerial parts of the plants, often causing poor kernel set, and can act as vectors for some plant diseases, such as maize chlorotic mottle virus and cowpea severe mosaic virus (Gergerich, Scott and Fulton 1986).

Rootworm control is undertaken mainly by the use of soil insecticides to attack the soil-living larval stage but problems with resistance are reducing the effectiveness of this approach. Resistance to heptachlor and cyclodiene are common and resistance to many other insecticides, including DDT, fonofos, malathion, aldrin and diazinon, have been observed. The resistant strains also tend to have a wider distribution than the susceptible ones, presumably matching the spread of maize cultivation in the last 50 years. Furthermore, the adaption of soil bacteria to use the applied insecticides as a carbon source, hence making the actual concentration of insecticide in the soil much less that the

amount applied, is becoming an increasing problem in fields that have been routinely treated (Felsot *et al.* 1982). In 1986 the losses due to larval feeding on the roots of maize in the USA was estimated at around $1 billion with about $40-50 million for aerial spraying of extensive out breaks and $50-100 million of damage caused by attacks on other crops (Edmonds 1994).

The western corn rootworm (WCR), *Diabrotica virgifera virgifera* , and northern corn rootworm (NCR), *D. barberi* , are widely distributed throughout the central and mid-western United States where their primary and most economically important host is field maize (Chiang 1973). The NCR and WCR have a similar life-cycle, the eggs overwinter in the soil and hatch in late spring (Branson 1987). The larvae feed on only a limited range of plants in the Gramineae family, the most important host plant being maize. In mid-summer adult emergence begins, the adults feeding on the pollen, silks and leaves of maize, along with other plant species. A few weeks later, egg laying starts and continues until the beetles are killed by frost.

The southern corn rootworm (SCR), *D. undecimpunctata howardii* , has a broader host range and is an economic pest of peanuts, sweet potato and maize in the south-eastern United States. Unlike the northern and western corn rootworms, the SCR overwinter as hibernating adults, which, unable to survive the sub-zero conditions of the central and northern regions of North America, migrate to the southern United States for winter. During the following warm season these adults migrate again, extending their range into most of North America east of the Rocky Mountains (Krysan 1986).

The eggs, laid in spring, soon hatch and the polyphagous larvae feed on the roots of many species including members of the Cucurbitaceae, Leguminosae, Solanaceae, Compositae, Cyperaceae and Convolvulaceae, as well as plants of the Gramineae family (Krysan, 1986). Adult emergence begins in mid-summer.

In all species, the newly hatched larvae feed primarily on the root hairs and outer cortical tissue. As they become older and the food requirement increases, they burrow deeper into the cortical parenchyma, causing channelling to occur in the stele, thus causing greater damage to the plant. Feeding is not limited to the roots, in the insectary larvae will readily feed on the kernel and stem of newly germinated maize, but in the field, where maize plants are more established, the only susceptible point for attack may be the new roots and root hairs.

Diabrotica was also suitable as a target species because it is amenable to bioassay

in artificial diets (Edmonds 1994) which could be adapted to allow screening of SCABS in microtitre plates, and is readily reared in large numbers in the laboratory.

Aims and objectives

The hypothesis of this thesis is that by the expression of a protein from a gene introduced into a plant by genetic engineering, the susceptibility of that plant to a range of insect pests can be significantly decreased. As is discussed above, this has been demonstrated with a range of genes and the aim of the projects reported here is to further demonstrate insect resistance in transgenic plants by the expression of novel genes.

One method of achieving this would be to target an enzyme activity, which occurs in the insect gut, with an inhibitor which would be expected to be active against that enzyme. The potato carboxypeptidase inhibitor project targeted carboxypeptidase activity and no reports of the effects of inhibition of this class of digestive enzyme in insects have been published. The *Manduca* haemolymph trypsin inhibitor project targeted the serine protease digestive enzymes and many reports exist of inhibition of this class of enzymes being deleterious to insects (see above). However, in this case the inhibitor is of insect origin, rather than plant origin, and it was expected that it would be more specific for, and resistant to, insect proteases than inhibitors from other sources. The microsomal aminopeptidase inhibitor project targeted the aminopeptidase activity. No reports of the effects of inhibiting gut aminopeptidase have been published.

The other approach was by the expression of single chain antibodies in plants which interacted with target sites in the insect gut and interfere with normal metabolism. This would be highly deleterious to the insect. Work was initiated but unfortunately, due to the collapse of the AXIS genetics insect resistance project, resources became unavailable and the project was halted in its early stages.

Materials and methods

Materials

Plant material

All plant material was grown in a growth room set to 8 hours dark/16 hours light and 20/25°C on either John Innes number 2 or Fisons Levington compost unless stated otherwise.

Insect material

Manduca sexta were grown from eggs supplied by S. Reynolds (University of Bath) on tobacco plants.

Diabrotica undecimpunctata were reared by the method of Branson et al (1975) as modified by Edmonds (1994);

Eggs, obtained from French Agricultural Research Inc. (Lamberton, MN), were reared to adult and a culture maintained, using techniques chosen for their closeness to the natural environment of the corn rootworm and the available facilities. Adults were maintained at 25°C (+/-2°C), 10-15% RH and with a 16 hour light/8 hour dark lighting regime. The beetles were held in perspex cages and provided with artificial diet (a 1:1:1 mixture of yeast, honey and wheat germ) and cabbage as food. The diet was replaced fortnightly and the cabbage replaced at least twice weekly. The cabbage was also used to provide cover for the oviposition site, a tray of damp, dark-coloured pebbles. Eggs were collected from the oviposition site twice weekly. Adults were transferred to clean cages fortnightly to reduce risk of disease.

The eggs were washed from the pebbles onto a mesh sheet, retaining the stones in a metal sieve and were collected into one area of the sheet, this folded so the eggs were uppermost and placed in a closed container until hatching began. Eggs were maintained in total darkness at 25°C and kept damp. As larval emergence began, the egg sheets were transferred to mats of seedling maize (see below), and left for 1 week, allowing larvae to hatch and move onto the rearing medium away from the egg mass. Trays of seedling maize thus infested with first instar larvae were kept moist and at 25°C. After 1 week the original

mat was split in half and fresh half mats used to provide more food for the developing larvae. Larvae migrated towards the fresh seedlings, allowing the old (dead) half to be replaced with another new half when necessary (usually after 1 week). When larvae reached the prepupal stage, the trays were caged using a wire framework and a mesh. Emerging adults were removed regularly and transferred to the adult cage. Once a prepupal stage was reached no further fresh maize was added.

Maize (*Zea mays* variety Kodak F1, Dawn F1 or Earle F1) mats were grown in trays 25 x 35 cm, in a 1:1 mixture of Levington M3 potting compost and vermiculite (medium grade). Kernels were washed in ethanol to remove the fungicide treatment, rinsed thoroughly, several times and soaked in water for 24 hours. After further washing the maize was sown. Maize mats used for larval rearing were usually planted 7 days prior to use.

Bacterial strains

The *E. coli* strain used for all routine work was DH5α (Gibco BRL) although occasionally JM109 was used for PCR cloning by the pGEMT method (Promega). The *Agrobacterium tumefaciens* strain used was LBA4404. All of these are standard strains. The strain used for *E. coli* expression was BL21(DE3) which was supplied with the pET vectors.

Chemical and biological reagents

Most chemical reagents were supplied by B.D.H. Chemicals Ltd. Poole, Dorset, U.K. and were of the purest grade available.

The Sigma Chemical Co. Poole, Dorset, U.K. supplied most specialist chemicals, in particular, ethidium bromide, adenosine triphosphate (ATP), spermidine, bovine serum albumin (BSA), dithiothreitol (DTT), HEPES (N-2-hydroxyethylpiperazine-N-2-ethane sulphonic acid), Tris (tris-hydroxymethylaminomethane), ampicillin, chloramphenicol, tetracycline, kanamycin sulphate, streptomycin, rifampicin, RNAse A, yeast tRNA, polyadenylic acid (polyA), herring sperm DNA, all deoxynucleotide triphosphates, all dideoxynuceotide

triphosphates, lysozyme.

Koch-Light Ltd. Haverhill, Suffolk, U.K. supplied sodium chloride, caesium chloride, acetic anhydride, iso-amyl alcohol, chloroform.

NBL Gene Sciences Ltd. Cramlington, Northumberland, U.K. supplied carbenicillin and some DNA modifying enzymes.

Pharmacia Fine Chemicals, Uppsala, Sweden, supplied Sephadex G-50, Sepharose 6B-CL, Ficoll 400.

Schliecher and Schull, Anderman and Co. Ltd. Kingston-upon-Thames, Surrey, U.K. supplied nitrocellulose filters (BA85, 0.45 μm).

Whatmann Ltd. Maidstone, Kent. U.K. supplied 3MM paper, DEAE-cellulose (DE81) paper, glass fibre (GF and GC) filters.

Difco Laboratories, Detroit, Michigan, U.S.A. supplied Bactotryptone, Bacto-Agar, Bacto-Yeast Extract.

Becton Dickinson and Co. Cockeysville, M.D. U.S.A. supplied BBL Tripticase peptone.

Collaborative Research Inc. Waltham, M.A. U.S.A. supplied Oligo dT_{12-18}.

Restriction endonucleases were mainly supplied by Bethesda Research Laboratories (UK) Ltd. Cambridge, U.K.. or The Boehringer Corporation (London) Ltd. Lewes, East Sussex, U.K., or New England Biolabs, C.P. Laboratories Ltd. Bishops Stortford, Herts, U.K.

The Boehringer Corporation (London) Ltd. also supplied Calf intestine alkaline phosphatase, endonuclease-free *E. coli* DNA polymerase I, T4 polynucleotide kinase, T4 DNA ligase, S1 nuclease, T4 DNA polymerase, *E. coli* DNA polymerase I large

fragment (Klenow polymerase).

Bethesda Research Laboratories (UK) Ltd. also supplied *E. coli* DNA polymerase I large fragment (Klenow polymerase), EcoR I synthetic linkers, agarose (electrophoresis grade) and *E. coli* strain DH5α.

Amersham International plc. Amersham, Bucks, U.K. supplied radiochemicals and TetZ polymerase.

Spectrum Medical Industries, Inc. 1100 Rankin Rd. Houston, Texas 77073-4716, U.S.A. supplied 2000 M_r cutoff dialysis tubing.

Novagen Inc. 597 Science Dr. Madison, WI, 53711, U.S.A. supplied pET *E. coli* expression vectors and kit.

Kirkegaard and Perry Inc. 2 Cessna Court, Gaithersburg, MD 20879, U.S.A. supplied ELISA reagents and antibodies.

Stratagene Ltd. 140 Cambridge Innovation Centre, Cambridge Science Park, Milton Rd. Cambridge, CB4 4GF, U.K. supplied Enhance PM, Taq Extender, pSK plasmid vector and the λZAP II kit.

Clontech Laboratories Ltd. 4030 Fabian Way, Palo Alto, CA 94303-4607, U.S.A. supplied the 5'RACE kit.

National Diagnostics, 1013-1017 Kennedy Blvd. Manville, New Jersey, 08835, U.S.A. supplied Ecoscint A scintillation fluid.

FMC BioProducts, 191 Thomaston Street, Rockland, ME 04841, U.S.A. supplied MetaPhor agarose.

Promega Corporation, 2800 Woods Hollow Rd. Madison, WI 53711-5399, U.S.A. supplied pGEMT PCR cloning vector and other plasmid vectors and the

PolyATract mRNA isolation kit.

Frequently used media, buffers and other solutions

Acrylamides	30 g acrylamide, 0.8 g bisacrylamide per 100 ml
dd water	double distilled and deionised water
denaturing buffer	1.5 M sodium chloride, 0.5 M sodium hydroxide
Denhardts solution	50 x Denhardts is 1% Ficoll 400, 1% polyvinylpyrrolidine, 1% BSA
Destain	40% Methanol, 7% Glacial Acetic Acid
gel loading beads	10 mM Tris-HCl pH 8.0, 10 mM EDTA, 30% w/v glycerol, 0.1% agarose and either 0.1% w/v xylene cyanol and 0.1% bromophenol blue, or 0.1% fast orange for when fragments <200 bp are expected. The mixture was prepared, autoclaved, allowed to set, then extruded through a fine needle with a syringe twice.
L agar	L broth plus 15 gm/l Bacto agar or 6 gm/l for top or soft agar
L broth	Bacto-tryptone 10 g, yeast extract 5 g, sodium chloride 5 gm to 1 litre with water.
MSO medium	MS salts (Murashige and Skoog medium salts) 4.7 gm/l, sucrose 30 gm/l, pH to 5.8, plus 8 gm/l agar for solid medium.
neutralising buffer	1.5 M sodium chloride, 0.5 M Tris-HCl pH 7.4
NZY agar	NZY broth plus 15 gm/l Bacto agar or 6 gm/l for top or soft agar
NZY top agarose	NZY broth plus 0.7% agarose
NZY broth	per litre; sodium chloride 5 gm, magnesium sulphate heptahydrate 2 gm, yeast extract 5 gm, NZ amine 10 gm, pH to 7.5 with sodium hydroxide
PBS	phosphate buffered saline; 8 gm of sodium chloride, 0.2 gm potassium chloride, 1.44 gm disodium hydrogen phosphate and 0.24 gm potassium dihydrogen phosphate in 1 litre of water to pH 7.4 with HCl.
PBS + T	PBS plus 0.02% Tween 20
PNK buffer	50 mM Tris-HCl pH 7.5, 10 mM magnesium chloride, 10 mM

	DTT
prewash solution	50 mM Tris/HCl pH 8.0, 1 M sodium chloride, 1 mM EDTA, 0.1% SDS
Stain	40% Methanol, 7% Glacial Acetic Acid, 0.05% Kenacid Blue R or Coomassie Brilliant Blue
Resolving Buffer	3.0 M Tris HCl pH 8.8
2 x SDS Sample Buffer	0.2 M Tris pH 6.8, 20% glycerol, 2% SDS, 0.002% Bromophenol blue
SM buffer	per litre; sodium chloride 5.8 gm, magnesium sulphate heptahydrate 2 gm, 50 mM Tris/HCl pH 7.5, .01% gelatin
Stacker Buffer	0.5 M Tris HCl pH 6.8
Stain	40% Methanol, 7% Glacial Acetic Acid, 0.05% Kenacid Blue R or Coomassie Brilliant Blue
SSC	1 x SSC is 0.15 M sodium chloride,. 15 mM trisodium citrate, pH 7.0 with sodium hydroxide.
SSPE	1 x SSPE is 0.18 M sodium chloride, 10 mM sodium dihydrogen phosphate pH 7.7, 1 mM EDTA.
Stain	40% Methanol, 7% Glacial Acetic Acid, 0.05% Kenacid Blue R or Coomassie Brilliant Blue
TBE buffer	1 x TBE is 10.8 gm tris, 5.5 gm boric acid, 0.93 gm EDTA (the disodium salt) to 1 litre pH ~8.3.
TE buffer	10 mM tris/HCl pH 7.5, 1 mM EDTA
YEB agar	YEB broth plus 15 gm/l Bacto agar
YEB broth	per litre; Bacto peptone 5 gm, yeast extract 1 gm, beef extract (brain/heart infusion) 5 gm, magnesium sulphate heptahydrate 2 gm, sucrose 5 gm
YT agar	YT broth plus 15 gm/l Bacto agar or 6 gm/l for top or soft agar
YT broth	Bactotryptone 8 g, yeast extract 5 g, sodium chloride 5 g, to 1 litre with water

Methods

Standard biochemical and molecular biological techniques

All standard techniques are standard practise in the department of Biological Sciences, Durham University or based on those in Current Protocols in Molecular Biology (Ausubel *et al.* 1992) or Molecular Cloning: A laboratory manual (Sambrook, Fritsch and Maniatis 1989) unless a reference is specifically given.

Plastic ware, glassware, general apparatus and reagents

All apparatus coming directly or indirectly in contact with nucleic acids, bacteria, enzymes, and or sensitive or sterile reagents etcetera were sterilised by autoclaving. Those items which cannot be autoclaved were sterilised by prolonged immersion in 80% ethanol and, if appropriate, flaming. All glassware for nucleic acid work was siliconized before autoclaving using "Repelcote" (Hopkins and Williams, Romford, U.K.).

All solutions for DNA manipulations, except gels, were autoclaved where possible or made up using sterile water and sterile stock solutions in sterilised containers, minimising the use of non-sterile reagents or put through a sterilising filter. All water was distilled or deionised, or, where noted as dd grade double distilled and deionised.

Alcohol precipitation of DNA

DNA solutions were made to 0.3 M sodium acetate, usually by the addition of 1/10th volume of 3 M sodium acetate pH 5.2, then 2 to 3 volumes of ethanol were added, mixed, and the solution kept at -20°C or -80°C for 15-20 minutes or overnight.

The precipitated DNA was pelleted by centrifugation; for small (up to 2 ml) samples at ~12,000 g, 4°C for 10 to 30 minutes (M.S.E. Microcentaur microcentrifuge or equivalent), or for larger samples at ~25,000 g, 4°C for 20-30 minutes (Sorvall RC-5B centrifuge). The pellet was washed in 70% ethanol one to three times and gently vacuum dried before being resuspended usually in dd water or T.E. buffer . When the concentration of DNA was low or the size small (<200 bp), transfer RNA (tRNA) or glycogen was added prior to the addition of the ethanol as a carrier. When the sodium concentration in the DNA solution was high, potassium acetate pH 5.2 to 0.5 M or ammonium acetate and magnesium acetate pH 5.2 to 0.5 M and 0.1 M respectively were used instead of sodium acetate. When it was necessary to minimise the total volume 0.6 to

1.0 volumes of iso propanol were used in place of ethanol and the mixture kept at -20°C for at least 20 minutes before centrifugation.

Phenol extraction of DNA solutions

Proteins were removed from DNA solutions by extracting twice with ~1 volume of phenol saturated with T.E. buffer, then once or twice with chloroform iso-amyl alcohol (24:1). The layers were mixed by brief vortexing and separated by brief centrifugation. The DNA remains in the top aqueous phase in each case. Occasionally the chloroform iso-amyl alcohol extraction steps were replaced with two or three extractions with diethyl ether, in this case the DNA containing aqueous layer is on the bottom. This whole process is referred to as phenol or phenol/chloroform extraction.

Trichloroacetic acid precipitation of DNA

DNA solutions were trichloroacetic acid precipitated essentially according to the protocol in the Amersham Nick Translation Kit handbook (Amersham International plc. Amersham, U.K.) using ~30μg of phage λ DNA as carrier and Whatmann GF C glass fibre filters.

Scintillation counting of radiolabelled DNA

Labelled DNA in solution or precipitated onto glass fibre discs was counted by liquid scintillation in a Packard Tri-Carb liquid scintillation analyser, model 1600TR. Ecoscint A was used a scintillation fluid unless otherwise stated.

Dialysis of solutions of macromolecules

Dialysis tubing of various sizes (Medicell International Ltd. London, U.K.) was prepared by boiling for 20 minutes in 10 mM EDTA followed by rinsing with distilled water. The tubing was knotted at one end at least twice and then filled using two pasteur pipettetes, one as a pipette and the other as a funnel. A space was left at the top of the tubing to allow for changes in volume of the solution and the tubing sealed with again at

least two knots. The dialysis tubing was then placed in a large volume of the relevant buffer which was stirred at 4°C. Dialysis proceeded for at least 24 hours and two changes of buffer. Small proteins were dialysed in the same manner except low molecular weight cut off dialysis tubing (Spectrum) was used which was supplied ready for use. Tubing clips were occasionally used in place of knotting.

Estimation of nucleic acid concentration and purity by spectrophotometry

Using a Pye Unicam SP8-150 ultra-violet/visible spectrophotometer and 1 cm path length quartz cells, both the concentration and purity of DNA solutions were estimated.

The optical density (OD) of a DNA solution at 260 nm is ~0.02 times its concentration in micrograms per millilitre (μg/ml). A spectrum from 230 to 370 nm was used to assess purity as follows, for pure DNA:

$OD_{260} / OD_{280} \approx 1.8$

$OD_{260} / OD_{235} > OD_{260} / OD_{280}$

$OD_{320} \approx 0$

Should these three conditions not hold, then the concentration estimate will be in error.

For RNA the same method was used except that the OD_{260} of RNA is ~0.024 times its concentration in micrograms per millilitre (μg/ml) and for really pure RNA the OD_{260} /OD_{280}
ratio should be 2 or more although ~1.8 is usually taken as reasonably pure.

Estimation of DNA concentration by the spotting method

This method was taken from ZAP-cDNA Synthesis kit (Stratagene). An accurate quantitation of DNA can be obtained by UV visualisation of samples spotted on ethidium bromide (EtBr) agarose plates in comparison with DNA samples of known concentration. 100 ml of a 0.8 % (w/v) agarose and electrophoresis buffer media was prepared. The molten agarose was cooled to 50°C then 10 μl of 10 mg/ml EtBr stock solution was added. The EtBr stock solution was prepared in distilled water and was stored in the dark at 4°C. After swirling to mix the EtBr agarose was poured into 100 mm petri dishes using ~10 ml/plate. The plates were left to harden and incubated at 37°C to dry, if necessary. These plates could be stored in the dark at 4°C for up to 1 month.

Using a DNA sample of known concentration, seven serial dilutions in 100 mM EDTA were made to cover the range from 200 to 10 ng/µl. These standards may be stored at -20°C for 3 months. Using a marker, the petri dish was labelled to indicate where the sample and the standards (200, 150, 100, 75, 50, 25 and 10 ng/µl) were to be spotted. The standards were thawed and mixed thoroughly before use and 0.5 µl of each was carefully spotted onto the surface of the prepared EtBr plate. Care was taken not to dig into the surface of the plate. Capillary action was used to pull the small volume from the pipette tip to the plate surface and this avoided the formation of bubbles. The pipette tips were changed between each sample. After all of the standards had been spotted, 0.5 µl of the cDNA sample was immediately spotted onto the plate adjacent to the line of standards. The spots were allowed to absorb into the plate for 10-15 minutes at room temperature. The plate was inverted and photographed using a UV lightbox. The concentration of the unknown sample was estimated by comparing it with the standards. Standards and unknowns must be spotted within 10 minutes of each other.

The storage of bacterial strains

In the short term (up to 2 months) bacterial colonies were stored on agar plates kept inverted at 4°C, protected from light and sealed with Nescofilm (Nippon Shoji Kaisha Ltd. Osaka, Japan). Permanent storage was effected by dispersing bacterial lawns, grown from single colonies on agar plates, in 2 ml sterile portions of equal amounts of L-broth and 80% glycerol, and keeping at -80°C.

Standard enzymatic methods

Restriction endonuclease digestions

Restrictions were carried out normally in one of the five buffers supplied by Boehringer Mannheim (BCL), the choice of buffer being determined by the data poster they had supplied which was especially useful for deciding the appropriate buffer for mixed enzyme digestions. If there was not an appropriate BCL buffer then that supplied or recommended by the supplier for that enzyme was used. Normally enzymes were used at a concentration of 1-3 u/µg DNA, except where higher ratios were recommended, and were

incubated at the temperature recommended by the supplier for 1 hour. Nuclease-free BSA was added to 100µg/ml to some digestions, particularly those on a large scale or being incubated for a long time. RNAse A (10 µg per reaction), which had been boiled for 15 minutes to destroy any DNAse activity, was added to restrictions of plasmid DNA prepared by the non sequencing quality miniprep method. Genomic DNA was restricted with 5-10 u/µg DNA with the enzyme being added in two lots and incubated for 2 hours or more, great care was taken that the DNA was fully dissolved in solution.

Ligation

Ligations, the joining of dsDNA molecules in the conventional manner by covalent bonds, were carried out in PNK buffer containing 1 mM ATP, or in the buffer supplied, with T4 DNA ligase. For joining two molecules, the volume of the reaction was kept to a minimum. For circularising a molecule, a greater dilution was used. Overnight incubation at 15°C was the norm but some difficult blunt end ligations were incubated at 12°C for up to 2 days, often with the addition of more ATP and enzyme after one day.

Making double stranded DNA blunt ended

Using S1 nuclease

This method is suitable for use on dsDNA molecules with protruding 3' or 5' ends or hairpin-loops. It is known to give dsDNA molecules with ends blunt or ~1 base in length, but should be used with care as prolonged incubations with S1 nuclease can significantly shorten dsDNA molecules. S1 nuclease reactions were carried out in 200 mM sodium chloride, 50 mM sodium acetate pH 4.4, 1 mM zinc sulphate buffer at 37°C for 30 minutes and then 25°C for 30 minutes, followed by the addition of 1/10th volume of 100 mM EDTA to stop the reaction. A typical reaction, containing ~3 µg of dsDNA, average length ~1000 bp, used 1000 u of S1 nuclease in 34 µl with ~10 µg of tRNA added.

Using DNA polymerase I

This method is suitable for use on dsDNA with protruding 3' or 5' ends, but not for dsDNA with hairpin-loops. It was usually used after S1 nuclease digestion and is often referred to as "polishing" the ends of the dsDNA. The enzyme gives accurately blunt ends by its polymerase and exonuclease activities. DNA polymerase I polishing reactions were carried out in PNK buffer at 12.5°C for 30 minutes in the presence of 0.2 mM of each deoxynucleotide triphosphate and ~0.5 u of endonuclease-free DNA polymerase I per μg of dsDNA.

Using T4 DNA polymerase

T4 polymerase was used to prepare the ends of PCR fragments for blunt ended cloning. It has the same activities as DNA polymerase I but its nuclease activity is about 200 fold stronger. T4 polymerase blunt ending reactions were carried out in 33 mM Tris acetate pH 8.0, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT and 0.1 mg/ml BSA in the presence of ~0.2 mM each dNTP and using ~0.5 u/μg DNA. Incubation was at 37°C for 10 minutes followed by 65°C for 5 minutes followed by phenol extraction and ethanol precipitation.

Using Klenow DNA polymerase

Klenow polymerase is DNA polymerase I minus a fragment of the protein; this deletes the 5' to 3' exonuclease activity and hence Klenow polymerase will fill in 5' overhangs to blunt ends but does not affect 3' overhangs. It was the enzyme of choice for this reaction as the lack of nuclease activity gives more reliable and accurate ends. It was also used for end labelling DNA with the appropriate ends using the appropriate radio dNTP. 10 x Klenow buffer is; 0.5 M tris/HCl pH 7.6, 0.1 M magnesium chloride, but the enzyme works well in a wide variety of other buffers, including the standard restriction buffers, provided there is ~10 mM magnesium present. dNTPs to excess (~0.1 mM) were usually used. Incubation was at 37°C or room temperature for 15 minutes.

Phosphorylation of DNA at the 5' ends using polynucleotide kinase

Normally DNA has phosphate groups attached to its 5' terminal hydroxyl groups,

and these are necessary for successful ligation reactions. Should these phosphate groups have been removed or be missing, e.g. commercial synthetic linkers are supplied without 5' phosphate groups, polynucleotide kinase may be used to replace them, and if supplied with γ^{32}P ATP will specifically "label" the 5' end(s) of the DNA molecule. The reaction is much more efficient for 5' overhangs than for blunt ended or 5' recessed dsDNA molecules.

Phosphorylation reactions were carried out in PNK buffer in the presence of ATP. In a typical reaction 1 µl (0.88 µg) of EcoR I linkers were phosphorylated using 3 u of enzyme and 40 µCi (8 x 10^{-12} mol) of γ^{32}P ATP with incubation at 37°C for 30 minutes.

Dephosphorylation of DNA at the 5' ends using alkaline phosphatase

Calf intestine alkaline phosphatase specifically removes the phosphate groups from the 5' terminal hydroxyl groups of DNA.

Phosphatase reactions were carried out in 50 mM Tris-HCl pH 9.0, 1 mM magnesium chloride, 0.1 mM zinc chloride and 1 mM spermidine. For dsDNA molecules with protruding 5' ends incubation was for 1 hour at 37°C using 0.2 u/µg DNA of enzyme. For blunt ended molecules or those with recessed 5' ends, i.e. protruding 3' ends, incubation was for 15 minutes at 37°C, then 15 minutes at 56°C, and then a second aliquot of enzyme was added and both incubations repeated.

Polymerase chain reaction methods

Standard PCR

The polymerase chain reaction, originally Saiki *et al.* (1980) with a thermostable polymerase enzyme, was performed with the following protocol based, amongst many sources, on the data sheet supplied with Enhance Perfect Match (Stratagene); For a 50 µl reaction the following were combined;

5µl of 10 x buffer (supplied with the enzyme)
8µl of 1.25 mM each dNTP
1 µl of forwards oligo, a solution of approximately 40 µmol/ml

1µl of reverse oligo at the same concentration
Taq DNA polymerase 2.5 units
substrate DNA
dd water to 50 µl

This was mixed well, spun, 50 µl of oil was added, spun and then incubated on a PCR block.

The standard conditions for PCR were; melt 94°C for 5 minutes, anneal for 5 minutes, then 30 cycles of; extend 1 minute per kilobase at 72°C, melt 1 minute, anneal 2 minutes, followed by extend 10 minutes and cool. The annealing temperature was chosen to match the oligos being used according to the rule;

$$Tm = 2(A + T + I) + 4(G + C)$$ (Sambrook, Fritsch and Maniatis 1989)

This was calculated only for that part of an oligo which was expected to match the target sequence, i.e. an engineered on restriction site would be ignored. There were a number of variations to this protocol which were used;

Hot start; All the components except the enzyme (and optionally the oligos) were combined and heated to >80°C then the enzyme (and oligos) were added. This is supposed to reduce the production of artifacts caused by enzymatic extension from oligos mis-annealed to target DNA at low temperatures.

Enhance Perfect Match (PM); This is an additive produced by Stratagene which destabilises mismatched DNA duplexes especially around their 3' ends. As this is the site from which Taq pol extends and as Taq pol will not extend from a non duplex the effect is to suppress non specific amplification. Enhance PM was used at 1 unit per 50 µl reaction.

Taq Extender; This is an additive and a buffer system produced by Stratagene which increases the maximum length of product which can be successfully amplified. It was used as suggested by the supplier. The buffer was substituted for the normal Taq buffer and an equal number of units of Taq Extender to units of Taq pol were added. Also the concentration of dNTPs was increased to as much as 10 mM to compensate for the increased usage in the longer product.

Touchdown PCR; This is an incubation protocol whereby the annealing temperature is dropped on successive cycles so that in the early cycles the temperature is

too high for any primers to bind. As the annealing temperature drops in subsequent cycles only very well matched primers bind and are amplified from and these early products will dominate. Thus it is a method to improve specificity especially where the melting points of the primers are uncertain due to wobble bases or inosines. The usual protocol was; 2 cycles at an annealing temperature of 65°C, 2 cycles at 64°C, etcetera to 50 °C for which there were 25 cycles then 10 minutes at 72°C and cool.

RACE PCR

Both 5' and 3' RACE (rapid amplification of cDNA ends) were performed. 5' RACE was done with the 5'-AmpliFINDER RACE kit supplied by Clontech according to the instructions supplied. 3' RACE was performed by a protocol of my own devising taking advantage of the clone rescue protocol of the λZAP kit and is described and discussed in the insect haemolymph trypsin inhibitor results chapter.

TetZ reverse transcriptase PCR

TetZ (Amersham) is a thermostable enzyme which in different buffers can act as a reverse transcriptase and as a DNA polymerase. It was used according to the suppliers protocol.

Radiolabelling

Two methods of radiolabelling whole DNA were used; the PCR method is also described and discussed in the potato carboxypeptidase inhibitor results chapter.

Radiolabelling DNA by random priming

The random priming method uses random hexanucleotides to prime DNA synthesis by Klenow polymerase (Feinberg and Vogelstein 1983 and 1984) and α^{32}PdCTP is normally used as the radio nucleotide. The following procedure was used;

x μl DNA which should be about 100-250 ng

dd water to 31 µl

Heat to 100°C for 3 minutes and cool on ice.

10 µl of OLB

2 µl BSA 10 mg/ml

2 µl (2 u) Klenow polymerase

5 µl (50 µCi) α^{32}PdCTP

The components were mixed and incubated at room temperature for a couple of hours or preferably overnight. The probe was separated from the unincorporated nucleotides by G50 Sephadex column chromatography. This method is suitable for DNA greater than about 250 bp in size. For 400 Ci/mmol α^{32}PdCTP about 150 ng of probe will be synthesised for complete incorporation of the radionucleotide.

OLB was made as follows;

1) Solution O; 1.25 M Tris/HCl pH 8.0, 0.125 M magnesium chloride.

2) Solution A; 1 ml of solution O plus 18 µl of β–mercaptoethanol plus 5µl of each of 0.1 M dATP, dGTP and dTTP made up in 3 mM Tris/HCl pH 7.0, 0.2 mM EDTA.

3) Solution B; 2 M HEPES pH 6.6 with sodium hydroxide.

4) Solution C; Hexadeoxyribonucleotides (PL Biochemicals) in TE buffer at 90 OD_{260} units per ml.

5) OLB; Mix solutions A, B and C in the ratio of 100 : 250 : 150 and store at -20°C.

Radiolabelling by PCR

The method of radiolabelling by PCR is also described and discussed in the potato carboxypeptidase inhibitor results chapter.

DNA	1 µl of a 1 in 1000 dilution of a standard miniprep (see below)
10x Buffer	5 µl
1.25 mM dA, G + TTP	8 µl
5' Oligo	1 µl
3' Oligo	1 µl both at normal PCR concentration (~20 µmol in 500

µl)	
Taq pol	0.5 µl, ~2.5 u
α^{32}PdCTP	5 or 10 µl, 50 or 100 µCi
dd water	to 50 µl

The DNA was a restricted out fragment to avoid "read past" contamination of the probe with sequences from beyond the primers, it was unnecessary to isolate the DNA fragment. If a PCR product was used to make a probe it should be size fractionated on gel to reduce the non-specific labelling or be subcloned.

This PCR program has worked well for probes of 100-200 bp; 94°C 5 minutes, 50°C 5 minutes, 15 cycles of 72°C 20 seconds, 94°C 1 minute, 50°C 2 minutes, then 72°C 10 minutes and cool. This usually incorporated ~50% of the counts when checked after G50 chromatography. I.e. good hot probe though physically not very much; 156 ng of synthesised DNA for complete incorporation of the label starting with 50 µCi of α^{32}PdCTP (calculated by I. M. Evans personal communication).

End labelling DNA by filling in with Klenow polymerase

Up to 1 µg of DNA was digested with the appropriate enzyme(s) in ~25 µl of the appropriate restriction buffer. 2-10 µCi of the desired α^{32}PdNTP plus any cold dNTPs needed were added followed by 1 u of Klenow. This was mixed and incubated at room temperature or 37°C for 15 minutes. The reaction was stopped at 70°C for 5 minutes and if needed separated from the unincorporated counts by G50 Sephadex column chromatography. See also "Making double stranded DNA blunt ended; Using Klenow DNA polymerase" above.

Oligo labelling

Oligos were end labelled with T4 polynucleotide kinase in essentially the same way that DNA was phosphorylated; 1 µl of 10 pmol/µl oligo was phosphorylated with 10 µl, 100 µCi of γ^{32}PATP (20 pmol) in PNK buffer using 1 u T4 polynucleotide kinase and a total volume of 20µl. Incubation was at 37°C for 90 minutes followed by 80°C for 5 minutes. The incorporation was then checked (see below) and if the incorporation was

low then an extra unit of T4 polynucleotide kinase was added and the mix was incubated for an extra 30 minutes and then stopped again.

The reaction mix was purified by ethanol precipitation; The reaction mix was diluted to 50µl with dd water then 240 µl of 5 M ammonium acetate was added, mixed and 750 µl of ice cold ethanol was added. This was stored on ice for 30 minutes then centrifuged for 20 minutes at 12000 g. After carefully removing the supernatant with a pipette 500µl of 80% ethanol was added. This was mixed and recentrifuged and the supernatant was removed as above. The sample was vacuum dried and resuspended in 100 µl of TE buffer or dd water. This method works well for 18mers and bigger but care should be taken for smaller oligos.

The incorporation of radiolabel into an oligo was estimated by dotting 0.5 µl of the reaction mix on to each of two DE-81 cellulose filter discs and allowing them to dry at room temperature. One was kept as the total counts the other was washed four times in 250-300 ml of 0.5 M sodium phosphate pH 7.0 and once in 70% ethanol then allowed to dry. This was the incorporated counts. Both were then scintillation counted in a scintillation cocktail which does not dissolve the counts from off of the filter discs.

Gel electrophoresis of nucleic acids

Agarose gel electrophoresis of nucleic acids

DNA and RNA fragments of various sizes from 150 bp to 50 kb were analysed for size on agarose gels of concentrations ranging from 0.5-1.5% (figure 4). Horizontal slab gels submerged in buffer were used throughout. Small DNA fragments, 200-50 bp, were analysed on MetaPhor Agarose gels (FMC BioProducts) which were prepared and used as directed in the suppliers protocol. .

Agarose gels

Full sized agarose gels of approximately 20 x 15 x 0.6 cm were made using perspex moulds on clean glass plates, a thin layer of vacuum grease being used to obtain a seal. Perspex combs were suspended across the mould to produce the loading slots. Also used were commercially made gel apparatuses of various sizes which were used according to the

Concentration of agarose gel (%)	Size range of dsDNA analysed (kb)
0.5	20 - 1
0.7	10 - 0.5
1.0	7 - 0.35
1.2	7 - 0.2
1.5	5 - 0.1

Figure 4. Approximate size range of double stranded DNA fragments analysed on an agarose gel of given concentration.

manufacturers directions but with the gel buffer etcetera of my choice. All gels were normally made containing 1 μg/ml of ethidium bromide and using TBE buffer. In all cases, the gel was run in the same buffer and ethidium bromide concentration as that used to make the gel. The gels were normally run during the day at about 10 V/cm, or at about 1.6 V/cm overnight. Normally the slot size was chosen to be the smallest which could contain the sample, plus beads, so as to allow for the maximum number of slots and the easier detection of faint bands. All DNA samples were loaded with the addition of ~1/3 volume of gel loading beads. The DNA bands were detected by the orange fluorescence of the ethidium bromide DNA complex under UV light illumination at 254 nm wavelength, and were recorded by photography through a red-orange filter (Kodak 23A Wratten) using transmitted UV light with Polaroid Film type 667 3000 ASA or with a gel documentation system (Mitsubishi UVP). Gels with high levels of background fluorescence were destained at room temperature for 45 minutes in 1 mM magnesium sulphate. Exposure times varied with the efficiency of the visible light filter and the UV light source, as little as 10 ng of dsDNA could be detected.

Alkaline agarose gels

Two methods were used for alkaline agarose gels. The first method was that of Sambrook, Fritsch and Maniatis (1989) where the entire gel and running buffer were made alkaline. The second method used an ordinary agarose gel with no ethidium bromide with the DNA made single stranded by treatment with alkali immediately before being loaded (McDonnell, Simon and Studier 1977); 10μl of DNA mix was added to 6 μl of 1 M sodium hydroxide at 0°C, after 5 minutes 6 μl of 1 M HCl was added also at 0°C and the sample immediately loaded and run as normal.

These gels were stained after running with acridine orange (30 μg/ml in the running buffer) for 30 minutes at room temperature, then destained in running buffer for about an hour also at room temperature. The DNA was detected by fluorescence as for ethidium bromide staining, dsDNA fluoresces green and ssDNA red when stained with acridine orange. Alternatively ethidium bromide staining can be performed after the gel had been run using running buffer and 1.0 μg/ml of ethidium bromide, but is supposed to be less sensitive. N.B. Alkaline gels destroy RNA.

Glyoxal gels

Glyoxalation was used to denature RNA (or DNA) prior to gel electrophoresis so that the RNA would be fully linear and its rate of migration would accurately reflect its size. The procedure used was essentially that of Sambrook, Fritsch and Maniatis (1989). Deionised glyoxal 6 M was prepared as described, and was used to glyoxalate up to 20 µg of RNA in the following mixture: 6 M glyoxal 2.7 µl, DMSO 8.0 µl, 0.1 M sodium dihydrogen phosphate pH 7.0 1.6 µl, RNA and dd water to 20 µl final volume, with incubation at 50°C for 1 hour.

1% agarose gels were made up using 10 mM sodium dihydrogen phosphate pH 7.0 buffer and no ethidium bromide. The samples were loaded with the addition of 1/5th volume of loading buffer (50% glycerol, 10 mM sodium dihydrogen phosphate pH 7.0, 0.4% w/v bromophenol blue). The gel was run conventionally in the same phosphate buffer which was recirculated to maintain the pH. The RNA could be detected by staining with ethidium bromide.

Formaldehyde gels

Formaldehyde gels were the alternative protocol for denaturing nucleic acids to the single stranded form before running on gel. The procedure used was the adaption of Sambrook, Fritsch and Maniatis (1989). The running buffer was 20 mM MOPS, 8 mM sodium acetate, 1 mM EDTA made up in DEPC treated water, adjusted to pH 7.0 with sodium hydroxide and filter sterilised, usually it was made up as a 5 times stock. The gel was made with running buffer plus formaldehyde to 2.2 M and was allowed to set for at least 30 minutes in a fume cupboard as formaldehyde is toxic. RNA up to 30 µg in 4.5 µl was denatured by adding 2.0 µl of 5 x formaldehyde running buffer, 3.5 µl of stock 12.3 M formaldehyde (37% solution in water) and 10.0 µl of formamide, incubating at 65°C for 15 minutes and cooling on ice before being loaded onto the gel with the addition of 2 µl of loading dyes (50% glycerol, 1 mM EDTA, 0.25% bromophenol blue and 0.25% xylene cyanol FF). The gel was run conventionally and could be stained with ethidium bromide to visualise the RNA.

Analysis of band patterns on gels to determine fragment sizes

The sizes of dsDNA bands on agarose gels were estimated using various standard size markers; phage λ nM258 cut with restriction endonuclease Pst I or Hind III or Hind III and EcoR I, plasmid pBR322 cut with Rsa I or Alu I, plasmid pUC18 linearised with BamH I or commercially supplied size ladders. The logarithm (to the base 10) of the size of the fragment in base pairs was plotted against the distance of migration of the band in millimetres. This gave a slightly S-shaped curve which was drawn by interpolation or could be approximated closely by three straight line segments, and allows accurate estimation of dsDNA fragment sizes from 100-10,000 bp.

Other gels were analysed by similar methods. N.B. RNA runs to within 10% of ssDNA of the same size on gel.

Recovery of DNA fragments from agarose gels

The method of choice was electroelution into a dialysis bag. The band of interest was cut out of the gel using a scalpel which had been sterilised by flaming to remove any nucleases. Exposure to UV was minimised to reduce damage to the DNA or if available long wave length UV (300-360 nm) illumination was used. Prepared dialysis tubing was stored in 70% ethanol and washed with water then buffer before use. One end of the dialysis tubing was clipped shut and the gel slice placed inside then 350 µl of buffer was added. The buffer used was the same buffer which the gel had been run in but with no ethidium bromide i.e. normally TBE. Any air bubbles were then carefully excluded and the tubing clipped shut. The tube was then placed in a minigel apparatus and just covered with buffer and the DNA electroeluted at 50 mA (~90 V) for 15 minutes. The dialysis tube could be checked under UV to ensure the DNA had all migrated out of the gel slice.

The tubing was opened at one end and the gel slice was carefully squeezed out. The tubing was then carefully squeezed and the internal surfaces were rubbed together to free the DNA from the walls and resuspend it into solution. The DNA was removed from the tubing with a pipette and the tubing was rinsed with 150 µl of buffer. The DNA was then phenol extracted and ethanol precipitated usually with the addition of ~2 µg of tRNA or glycogen as carrier.

Gel Drying

Agarose gels were dried wrapped in cellophane and acrylamide gels dried sandwiched between a sheet of 3MM paper and a sheet of polythene. A commercial vacuum drying apparatus was used (BioRad Laboratories) and when needed heat at ~60°C was applied.

The transfer of nucleic acids from gel to filter membrane

Southern blotting

The method used was basically that of Southern (1975). The DNA in an agarose gel was made single stranded by gently shaking the gel in denaturing buffer twice for 15 minutes. The gel was then neutralised by gentle shaking in neutralising buffer twice for 15 minutes. A nitrocellulose filter was prepared by floating on distilled water until thoroughly wet, then soaking in 20x SSC. Alternatively nylon filters were used where their extra mechanical strength would be needed. At no time were the filters handled except with rubber gloves and at the edges. For the transfer of large fragments (greater than ~5 kb) the gel was subjected to an initial soak in 0.25 M hydrochloric acid but this was avoided if small (less than ~ 300 bp) fragments were expected to be present (Wahl, Stern and Stark 1979).

Gels were blotted in the conventional manner. The nitrocellulose filter was laid on the gel, the position and orientation of the slots marked and a sheet of 3MM paper was wetted and laid on top, care being taken to avoid "short circuit" liquid paths around the gel. Capillary action draws the 20x SSC transfer buffer up through the gel and filter into the nappies, transferring the DNA to the nitrocellulose where it sticks. The apparatus was always levelled and great care taken to exclude and air bubbles to avoid uneven transfer of DNA. The process was complete after 12-24 hours. The DNA was fixed to nitrocellulose filters by washing in 6x SSC for 5 minutes, air drying on a sheet of 3MM paper, and baking, sandwiched between two sheets of 3MM paper, at 80°C in a vacuum oven for 1-2 hours. The filter was stored dry until used, protected between sheets of 3MM paper. Nylon filters were fixed by exposure to UV light for 2 minutes wrapped in clingfilm on the transilluminator.

Alternatively a commercial vacuum blotting apparatus was used (Appligene vacuum

blotter). This uses a controlled partial vacuum to draw the denaturing, neutralising and transfer buffers through the gel and the filter membrane and is much faster in action. The buffers used are the same as for the traditional method and the pressures and times used were as recommended by the manufacturer.

Northern blotting

Northern blotting is the equivalent process to Southern blotting for RNA gels. The procedures are identical except that there is no denaturing and neutralising steps as RNA is destroyed by alkali and is run denatured already, and SSPE is some times substituted for SSC because of its greater buffering capacity. If nylon filters were used these were baked for 1 hour at 80°C after UV fixing to reverse the glyoxalation which can interfere with hybridisation.

Transformation of strains of *E. coli* with plasmids

E. coli cells were made competent for transformation with plasmids essentially by the procedure of Dagert and Ehrlich (1979). Cultures of the relevant strain were grown overnight at 37°C from single colonies and were used to inoculate, at a ratio of 1:50, 100 ml aliquots of the relevant broth, normally L-broth. These were then grown to an OD_{650} of 0.2 at 37°C on a fast shaker followed by chilling on ice. The cells were harvested by centrifugation at 3000 g, 4°C for 15 minutes, the supernatant carefully removed, and the cells resuspended in 50 ml of ice cold 100 mM calcium chloride. Great care was taken to maintain everything at 0°C. After 20 minutes the cells were isolated as before and resuspended in 1 ml of 100 mM calcium chloride at 0°C and stored on ice until use. Competent cells were kept like this for up to 30 hours, although transformation efficiencies had fallen appreciably by that time.

Transformations were effected by adding to 10 µl of a suitable dilution of DNA in 100 mM calcium chloride or other buffer, 100 µl of competent cell suspension. The mixture was incubated at 0°C for 10 minutes and at 37°C for 5 minutes, followed by the addition of 1 ml of L-broth and incubation with shaking at 37°C for 1 hour. Samples of this culture, normally 20 and 100 µl, were then spread on nutrient agar plates containing the relevant antibiotics and marker systems. The remainder of the culture was stored at

4°C for up to 30 hours before being either discarded or spread onto other plates.

Alternatively commercially prepared competent cells were bought and used according to the manufacturers instructions.

Transformation of strains of *E. coli* with replicative form M13

Replicate form (double stranded) M13mp9, and its derivatives, may be treated as a plasmid and is transformed into suitable strains of *E. coli* in the same way as for other plasmids with a few changes in method. Strains JM101 or JM109 were invariably used for M13, competent cells were prepared in the usual manner except that they were grown to an OD_{660} of 0.3-0.4 and some exponentially growing cells were transferred to a McCartney bottle containing nutrient broth and grown at 37°C during the experiment.

The transformation was effected by mixing 10 µl of a suitable concentration of DNA with 100 µl of competent cells at 0°C and leaving for at least 40 minutes. The mix was then heat shocked at 42°C for 2 minutes and the following were added in rapid succession:

10 µl 0.1 M IPTG (Isopropyl-ß-D-thiogalactopyranoside)

50 µl 2% X-Gal (5-Bromo-4-chloro-3-indoyl-ß-D-galactopyranoside) in dimethyl formamide

200 µl of growing cells and

3 ml of liquid soft agar (made with 6 gm/l instead of 15 gm/l of agar) at 42°C.

After mixing it was poured onto an agar plate, which had been allowed to warm to room temperature, and incubated at 37°C. Blue plaques were M13mp9, clear plaques M13mp9 with an insert in the cloning site.

Hybridisation of labelled nucleic acid probes to filter bound nucleic acids

This procedure was used to detect nucleic acids, both DNA and RNA, bound to filters, such as Southern blots, Northern blots or lysed colony filters. The hybridisation and all subsequent washing steps were carried out in sealed plastic bags in a shaking water bath at the relevant temperatures or in a rolling glass bottle in temperature controlled oven. Normally the hybridisation step was carried out at 65°C or 42°C for the formamide method. Almost all normal hybridisations were carried out by the normal water based

method and the formamide method was only used for RNA work where it gives better signals and for probes which were to be reused a number of times.

The filters were wetted thoroughly in 6 x SSC, some colony filters stuck to their 3MM paper protectors during the baking step and a few minutes soaking freed them. The colony filters were transferred to plastic bags and 50 ml of prewash solution added, the bags sealed and the whole incubated with shaking at 65°C or 42°C for 1-2 hours. This step was unnecessary for the blots.

The filters were next incubated for at least 1-2 hours with shaking in prehybridising solution; 5 x Denhardts solution, 5 x SSC, 0.1% SDS, 50-100µg/ml denatured salmon or herring sperm DNA, 1µg/ml polyA for cDNAs to avoid chance hybridisation to long polyA tails and 50% deionised formamide for the formamide method. ~20 ml for an 82 mm filter disc and ~40 ml for a blot filter were used. Herring sperm DNA was first denatured at 100°C for at least 6 minutes which was the method used to denature the DNA probes. Prehybridisation was usually carried out overnight.

The denatured labelled probe was then added in a small volume to the prehybridisation solution or to fresh prehybridisation solution, and hybridisation allowed to take place under the same conditions as prehybridisation according to the formula;

$$YZ/25X = C_o t_{half}$$ (Sambrook, Fritsch and Maniatis 1989)

$C_o t_{half}$ is the time in hours for half a dsDNA probe to rehybridise to itself, X is the weight of probe added in µg, Y is the complexity of the probe which is normally its length in kilobase pairs unless it consists of repetitive sequence and Z is the volume the hybridisation is carried out in. Hybridisation was normally carried out for 1-3 $C_o t_{half}$.

When the hybridisation step was finished the mix was poured off, it could be reused to probe other filters after denaturing by heating to near boiling, but the protein in the mixture tended to coagulate during this procedure so this was avoided where possible. The probes from the formamide method could be denatured by heating to 65-70°C which avoided this problem.

During the washing process care was taken to ensure that the filters never became dry. First the filters were rinsed 1 or 2 times in 2 x SSC, 0.1% SDS using ~50 ml per 82 mm disc and ~100 ml per blot to remove excess probe. Next the filters were washed to the required stringency in the appropriate dilution of SSC plus 0.1% SDS as determined by the

Southern and Northern Blot Washing Stringency

Wash at Tm -12°C

Tm =69.3 + 0.41(G+C)%

Tm decreases by 1°C for every 1% of mismatching.

$(Tm)_{M1} - (Tm)_{M2} = 18.5 \log10(M2/M1)$

Where M = ionic strength. $M = 0.5 \Sigma [i] z_i^2$ for all ions in the solution.

Assuming G+C% = 40%

Wash soln.	Temperature/°C	Washing stringency % homology
1xSSC	68°C	94%
1xSSC	65°C	91%
0.1xSSC	65°C	99%
1xSSC	55°C	81%
1xSSC	48°C	74%
3xSSC	55°C	72%
3xSSC	65.5°C	82%
3xSSC	60°C	77%
0.1xSSC	50°C	84%
0.1xSSC	42°C	76%
2xSSC	65°C	85%
2xSSC	60°C	80%
2xSSC	55°C	75%
2xSSC	50°C	70%

Tm = 81.5 + 16.6logM + 0.41(G+C%) - 500/L - 0.61(%formamide) - (%mismatch)

Where L = probe length in kb.

16.6 logM = + 4.9 for 5xSSPE

16.6 logM = - 17.8 for 0.1xSSC, 0.1%SDS

Figure 5. Washing conditions for DNA hybridisation.

following formulae:

1) Tm = 69.3 + 0.41 (G+C)% (Marmur and Doty 1962)

2) The Tm of a mismatched duplex decreases by 1°C for every increase of 1% in the number of mismatched base pairs (Bonner et al 1973).

3) $(Tm)_{\mu 2} - (Tm)_{\mu 1} = 18.5 \log_{10} \mu_2/\mu_1$

Where μ_1 and μ_2 are the ionic strengths of two solutions (Dove and Davidson 1962).

4) Each increase of 1% in formamide concentration lowers Tm by 0.7°C (McConaughty, Laird and McCarthy 1969, Casey and Davidson 1977).

Tm is the melting point of the duplex and washing was carried out at Tm-12°C. Normally filters were washed at higher ionic strengths rather than lower temperatures as better results were obtained. The filters were washed for 1 hour twice under the calculated conditions, then allowed to air dry, before being mounted on 3MM paper and covered in saranwrap, ready for autoradiography. For routine work however washing was to low stringency, usually 2 x SSC, 0.1% SDS at 65°C (figure 5).

Autoradiography

Autoradiography was used to detect ^{32}P and ^{35}S labelled nucleic acids on filters and gels. All manipulations involving undeveloped film were carried out in a darkroom using safelight illumination. The gels to be autoradiographed were always either dried wrapped in cellophane or covered with a sheet of saranwrap or polythene, except ^{35}S gels which were always dried and exposed with no wrapping, filters were always covered in saranwrap. Radioactive ink was used to mark slots etcetera, and uniquely orientate the autoradiograph.

X-ray film (Fuji RX) was sensitised by exposure to a low intensity flash of light and laid onto the gel/filter which was mounted on a glass plate. An intensifying screen (Dupont Cronex Lighting plus) was laid on top of the film followed by a second glass

plate, and the whole clamped together with bulldog clips, or elastic bands. The assembly was then wrapped in three black plastic dust bin liners and left to expose at -80°C for up to a month. Alternatively purpose made cassettes were used in place of the plates, bags and screen, with the filter or dried gel mounted on cardboard and covered with clingfilm.

The film was developed by washing in fresh Kodak X-Omat developer at room temperature for 5 minutes, cold water for 1 minute, Kodak fixer for 5 minutes, at which point the film could be exposed to light and cold water for up to 30 minutes. The films were dried at room temperature.

Restriction mapping

Plasmids containing unknown inserts were restriction mapped initially by cutting out their insert and accurately determining its size, and by finding some restriction endonucleases which cut the plasmid only once preferably close to the cloning site and the insert not at all. The insert was then mapped in a series of mixed digestions using one of these enzymes and an enzyme which does not cut the plasmid. Normally only hexanucleotide sequence recognising enzymes were used and pentanucleotide and tetranucleotide enzymes were only used for exhaustive mapping or if sites for them were suspected from other data. The sizes of all DNA fragments were determined by agarose gel electrophoresis.

Subcloning into and determining the orientation of inserts in M13

Normally inserts from pUC18 type plasmids were sub cloned into M13mp18 by taking a mixture of the plasmid and the double stranded form of M13mp18 restricting with the enzyme(s) whose site the insert was cloned into, phenol extracting, ethanol precipitating and ligating the mixture. After transformation the clear plaques could be roughly selected as large inserts give small plaques. M13mp18 also selectively clones smaller DNA fragments. The orientation of the insert in an M13 subclone could be determined absolutely by DNA sequencing or preparing the double stranded form and restricting it with the appropriate enzyme(s), or relatively by the "C test": Recombinant phages with their inserts in opposite orientations will hybridise via their inserts and the figure of eight structure resulting runs more slowly on an agarose gel than the unhybridised

forms.

Single strand M13 minipreps were run on an agarose gel containing no ethidium bromide and after running were stained for ~1 hour in normal running buffer containing lμg/ml of ethidium bromide and visualised under UV light. The rate of migration is relatively insensitive to insert size but size differences of <200 bp can be detected. Good minipreps were "C tested" by mixing equal amounts of DNA and annealing together by incubating at 65 C for 1 hour prior to running on gel against single preparations as a control.

M13 single strand miniprep

An overnight in 5 ml of 2 x YT was grown by either picking plaques direct into the broth using a sterile cocktail stick or by adding 0.2 ml of JM101 or JM109 cells to the picked plaque and broth. This was poured into an eppendorf, spun for 4 minutes and the supernatant recovered (The pellet was stored at 4°C to preserve the clone if needed). The supernatant was respun and 1 ml transferred to a fresh tube. The phage was precipitated by adding 200 μl of 20% PEG 6000, 2.5 M sodium chloride. This was left at room temperature for 20 minutes, spun for 3 minutes and the supernatant discarded. The pellet was respun for 10 seconds and all PEG was removed. The pellet was resuspended in 100 μl of TE buffer, phenol extracted, chloroform extracted. It was then ethanol precipitated by adding 10 μl of 3 M sodium acetate pH 4.8 and 300μl of ethanol, incubated at -20°C for 1 hour, spun, washed in 80% ethanol, dried and resuspended in 20μl of TE buffer. The concentration was checked on the spectrophotometer;
OD_{260} 1μl of 1μg/μl in 1 ml = 0.033.

Plasmid miniprep alkaline method sequencing quality

10 ml overnights of the desired colonies were grown up (with antibiotic etcetera) as normal in YT broth. These were spun down at 3000-4000 g for 15 minutes and dried briefly inverted over a paper towel. Each pellet was then resuspended in 200 μl of 50 mM glucose, 10 mM EDTA, and 25 mM Tris/HCl (pH 8.0) and stored on ice for 30 minutes. 400 μl of freshly prepared NaOH/SDS was added (NaOH/SDS to 10 ml of water add 0.08 gm (1 pellet) of sodium hydroxide, and 0.1 gm of SDS) mixed gently and stored on ice

for 5 minutes. The samples were transferred to 1.5 ml eppendorfs and 300 µl of ice cold acid potassium acetate was added (6 ml 5 M potassium acetate, 1.15 ml glacial acetic acid, 2.85 ml water), mixed by gently inverting and stored on ice for 30 minutes. This was then centrifuged for 30 minutes and 0.7 ml of clear supernatant taken off. 2 µl of 10 mg/ml RNAse-A (DNAse free) was added, mixed and incubated at 37° for 20 minutes. The samples were then extracted once each with phenol/chloroform (1:1) and chloroform then precipitated with 2 volumes of ethanol. Finally the pellets were rinsed once in 70% ethanol and vacuum dried briefly.

At this point the DNA could be used for all normal enzymatic manipulations but if it was intended for DNA sequencing then an extra purification step was performed.

The pellet was dissolved in 16.8 µl of dd water. 3.2 µl of 5 M sodium chloride was added, mixed, and then 20 µl of 13% PEG 8000 was added. After mixing well and incubating on ice for 20 minutes it was centrifuged at 12000 g for 10 minutes. The supernatant was removed with care and then the pellet was 70% ethanol rinsed, spun and dried as above. N.B. the purification will work if you don't bother to remove the supernatant and just add the 70% ethanol. The DNA was dissolved in 20 µl of dd water. The concentration of DNA was usually at least 250 ng/µl but it was best to estimate accurately on gel, rather than by spectrophotometry, as it contains significant amounts of RNA.

Methods for determining the titre of phage stocks

Phage λ was titred by one of two methods;

The pour plate method was best for small numbers of phage lines and for plaque lifts. A serial dilution of the phage stock was prepared in small sterile test tubes or on a microtitre plate. The microtitre plate should be blocked (see below) before use to avoid the nonspecific binding of the phage to the plastic. Dilution to 10^{-8} should be sufficient. A 0.1 ml aliquot of each dilution was taken and added to 0.1 ml of plating cells (see below) in SM buffer (although still in broth works). This was incubated at 37°C for 20 minutes to allow the phage to adsorb to the plating cells. 2.5-3 ml of top layer NZY agar was added, vortexed briefly and poured onto a labelled plate. This was swirled gently taking care not to create bubbles to spread the agar evenly across the plate. This process was repeated with each dilution. After all the plates had been poured and closed they were left

to stand at room temperature for 15 minutes until the top layer had set. They were incubated inverted overnight at 37°C.

The spotting method was best for larger numbers of phage lines and for quick approximations. This method used a microtitre plate to produce the serial dilutions and was much quicker than the pour plate method. The surfaces of the wells will adsorb phage in a nonspecific manner and so the plate was pretreated with 2% sterile BSA or gelatin for 30 min. at R.T. then allowed to dry. Plates containing host cells for spotting were prepared by drying NZY agar plates briefly if fresh and warming to 37°C before use. Either fresh or remelted NZY top agar equilibrated to 50°C was also prepared. 4 ml of plating cells per 100 ml of top layer agar was added, mixed and quickly poured in ~2.5 ml aliquots onto the plates swirling gently to get an even covering. These should be used within 3-4 hours. 50 ml of phage serial dilution was spotted onto a plate. 4 spots would fit on a normal petri dish and at least 6 by 6 on a large square plate. The plates were allowed to dry before inverting and incubating at 37°C overnight or incubated right side up taking care when moving so as to not cause the spots to run.

Plating cells were prepared by inoculating 10 ml of L-broth plus 0.4% maltose and 10 mM magnesium sulphate and growing overnight, these could be either stored on ice for 1-2 days, or spun down and resuspended in SM buffer and stored on ice, or used to start a culture if it was decided to grow one to a specific OD_{600}. An OD_{600} of 1.0 is equivalent to $8x10^8$ cells/ml. The cells were resuspended at $2x10^9$ cells/ml in SM buffer.

DNA preparation from phage λ

50 ml of L broth plus 10 mM magnesium and 0.2% maltose was inoculated with a single colony of the appropriate *E. coli* strain for the phage λ being prepared and grown at 37°C to mid phase, although overnight also worked. The OD_{600} was read, then the culture was centrifuged at 4000 g for 5 minutes, and resuspended in 10 mM magnesium sulphate at 3×10^9 cells per ml (1 OD_{600} = 8×10^8 cells per ml). 0.1 ml of cells were taken and added to 10^7 pfus of phage diluted to 0.1 ml with SM buffer and the whole incubated at 37°C for 20 minutes. This was then added to 50 ml of prewarmed LB plus 10 mM magnesium and 0.2% maltose and grown overnight on a shaker. In the morning signs of lysis were usually visible.

1 ml of chloroform was added and incubation was continued for a further 30

minutes then the debris were spun down at 4000 g for 10 minutes at room temperature. The supernatant was removed, DNAse and RNAse to 5 μg/ml each were added and the mixture was incubated at room temperature for 1 hour. Solid sodium chloride to 1 M was added and dissolved. The mixture was stored on ice for 1 hour then centrifuged at 11000 g for 10 minutes and 4°C and the supernatant recovered. Solid PEG6000 was added to 10% w/v and dissolved by slowly stirring on ice . This was incubated on ice for 1 hour then centrifuged at 11000 g and 4°C for 10 minutes. The pellet was resuspended in 0.8 ml of SM buffer and extracted with an equal volume of chloroform. The aqueous phase was recovered and diluted to 8 ml with SM buffer. This was spun down at 40000 rpm in the 10 x 10 rotor for 2 hours at 4°C and the glassy pellet was resuspended in 0.2 ml of SM buffer. EDTA to 20 mM, SDS to 0.5% and pronase to 500 μg/ml were added and the mixture was incubated at 37°C for 1 hour. This was then phenol extracted with an extra extraction with 1:1 phenol chloroform. The phage λ DNA could then be ethanol precipitated and was good for most enzymatic procedures, or alternatively could be further purified by centrifugation through a caesium chloride gradient. The yield was highly variable and crucially depended on the cell to phage ratio and the particular phage line being prepared.

Plaque lifts

Phage λ plates were plaque lifted in duplicate by laying nitrocellulose filters onto the surface of the plate, leaving for about a minute then marking with india ink and a sterile needle and lifting off. The second filter was left on for longer, usually >5 minutes, before marking and lifting off. Great care was taken to lay the filters on cleanly in one slow smooth action to avoid creases and, to compensate for the expansion of the filters as they wetted, the second filter was usually laid on at right angles to the first. The plates were best grown to no more than semi confluent or well separated plaques for isolating a single pure phage line. They should not have too much bacteria on the surface or a high background may result.

For this process plates with top layer agarose instead of agar are preferable as it is mechanically stronger and the plates should be chilled to 4°C to increase the agarose strength. The filters were processed by laying onto 3MM filter paper saturated with buffer plaque side up as follows; denaturing solution 1 minute, neutralising solution 5 minutes, 2 x SSPE 1 minute, allow to air dry on 3MM paper then wrap in more 3MM paper and

bake at 80°C in a vacuum oven for 1-2 hours. Plaque lift filters were stored dry.

Sephadex G50 and similar column chromatography

Column chromatography with Sephadex G50 and similar substances such as Sepharose CL-4B, fractionates DNA roughly by size such that the smaller the molecules the slower they pass through the column. Normally the columns were run in sterile 10 ml plastic pipettetes with a plug of siliconized glass wool at the bottom using ~6 ml of Sephadex. The columns were prepared by slurrying Sephadex in excess elution buffer (150 mM sodium chloride, 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 0.1% SDS), and heating this mix to 65°C for 2 hours, this was then stored at 4°C until used. The column was settled and run using elution buffer. The DNA was loaded on by allowing the column to run nearly dry then adding the DNA with a small volume of elution buffer. The progress of the DNA through the column could be followed by a Geiger counter for ^{32}P labelled DNA and fractions of ~0.5 ml were collected. Labelled DNA was detected by scintillation counting and is eluted in the first peak. The second peak contains unreacted nucleotides or linker fragments etcetera.

Carboxypeptidase A inhibition assays

The method used was essentially that of Peterson Holmquist and Bethune (1982) and used the substrate FAPP (N-(2-furanacryloyl)-L-phenylalanine-L-phenlyalanine).

Non-preincubation method; To 800 µl of substrate buffer mix (50 mM Tris/HCl pH 7.5, 0.45 M sodium chloride, 200 nM FAPP) equilibrated to 25°C was added a mix of 180 µl of buffer (as above minus the FAPP) also equilibrated to 25°C and 10 µl of leaf extract or inhibitor dilution (diluted or extracted in 50 mM Tris/HCl pH 7.5) in a 1 ml curvette. The OD_{330} was monitored in a temperature regulated spectrophotometer for approximately 10 minutes before the addition of bovine carboxypeptidase A 10 µl diluted in buffer to a concentration of 0.001 units per µl and kept on ice.

The preincubation method was identical except that the enzyme leaf extract and buffer were preincubated for 10 minutes at 25°C prior to adding to the substrate buffer mix and monitoring before adding the enzyme was unnecessary.

Preparation of PCI protein

PCI protein was prepared by the method of Pearce and Ryan (1983). 500 g of potato tubers, of an unknown, white-skinned, commercial variety, were cut into small pieces of approximately 50 grams and placed in a large waterbath at 80°C for 10 minutes. They were then washed under cold running water and mashed with the addition of 100 ml of 100% ethanol to a smooth paste. A further 2 litres of ethanol was mixed in over the next 15 minutes and the resulting slurry was vacuum filtered through Whatman No. 1 filter paper. The solids were resuspended in a further 200 ml of 80% ethanol and filtered as before. The combined filtrates were concentrated in a rotavap to approximately 500 ml, the solids filtered off, then dialysed in 2000 M_r cutoff dialysis tubing against several changes of water over 24 hours, and finally freeze dried to yield crude PCI.

Pure PCI was then recovered from the crude as follows. Crude PCI 75 mg was dissolved in 20 ml of 10 mM sodium citrate pH 4.3 and centrifuged 10,000 g to remove the debris. The supernatant was applied to a 17 x 1.4 cm column of sulphopropyl Sephadex C25 equilibrated with the same buffer. The column was then run with this buffer monitoring the OD_{280} until the first peak had eluted. A linear gradient of 0 - 0.3 M sodium chloride in the same citrate buffer was then used to elute the purified PCI. The resulting peak and its shoulders which are known to be isoinhibitors (Pearce and Ryan 1983) were pooled and dialysed in 2000 M_r cutoff dialysis tubing against many changes of water and then freeze dried.

Oligo hybridisation and washing

Oligos were hybridised to immobilised nucleic acids in essentially the same way as for longer probes, the same procedures being used for both DNA and RNA targets. Prehybridisation and hybridisation were done in 5 x SSC, 10 x Denhardts, 0.05% sodium pyrophosphate, 0.5% SDS and 50-100 µg/ml of denatured herring sperm DNA. The temperature was initially Tm -20°C, where Tm is calculated by the formula:

$$Tm = 2(A + T + I) + 4(G + C)$$ (Sambrook, Fritsch and Maniatis 1989)

although this often had to revised down to obtain a good signal. Washing was in:

x 1 in prehybridising solution

x 2 in 5 x SSC, 0.1% SDS

x 1 in 3 x SSC, 0.1% SDS

x 1 in 1 x SSC, 0.1% SDS (optional)

for a maximum of 10 minutes for each step and the same temperature as the hybridisation. Careful monitoring with a Geiger counter was used throughout to decide when enough washing had been done.

Quantitation of oligos

Oligo concentrations were estimated by spectrophotometry according to the relationship;

1 OD_{260} = ~33 μg/ml of oligonucleotide.

Agrobacterium methods

Transfer of constructs to *Agrobacterium* by triparental mating

Agrobacterium strain LBA4404 contains a disarmed Ti plasmid (pRAL4404) which has transfer functions, but no T-DNA. The strain is rifampicin and streptomycin resistant but kanamycin sensitive. The pBIN19-derived plasmid carries kanamycin resistance, and the T-DNA border sequences (containing the selectable marker, and the sequence for transfer to plant hosts). pRK2013 is a mobilising plasmid, which enables the transfer of the pBIN19 derivative to *Agrobacterium* . When grown together in a triparental mating, the three bacterial strains interact, and the desired transfer (pBIN19 derivative to *Agrobacterium*) takes place. It is selected for by plating out on streptomycin, rifampicin, kanamycin plates, where only *Agrobacterium* that have taken up the pBIN19 derivative can grow.

Normally the selection on appropriate antibiotic plates (and repeating for singles) is sufficient to ensure the correct strain is generated. However, the final selection should always be checked by preparing DNA and Southern blotting to ensure that the pBIN19

plasmid is intact and unrearranged.

Before starting, all strains were checked for appropriate antibiotic resistance and sensitivity by streaking on the appropriate plates (from single colonies if needed).

Agrobacterium strain (LBA4404:pRAL4404)	strepR, rifR, kanS
E. coli Mobilising strain (HB101:pRK2013)	strepS, rifS, kanR
E. coli Donor strain (DH5α:pBIN19-derivative)	strepS, rifS, kanR

N.B. *Agrobacterium* is recommended to be grown at 27°C to avoid losing the Ti plasmid. Overnight cultures of each strain were grown (YEB broth is recommended for *Agrobacterium*) under the appropriate antibiotic selection. N.B. *Agrobacterium* will take 2 days to grow!! Each overnight culture was spun down and resuspended in the same volume of 10 mM magnesium sulphate. This was repeated twice to remove the antibiotics. 200 µl of each culture was mixed in an Eppendorf tube, and poured onto a YEB plate. These were grown at 27°C for 1-2 days. Approximately 5 ml of 10 mM magnesium sulphate was poured onto each plate, the cells were resuspended by gently swirling the liquid on the plate then serially diluted (10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}) and plated out on YEB plus 500 µg/ml streptomycin, 100 µg/ml rifampicin and 50 µg/ml kanamycin. The plates were grown at 27°C for 3-4 days and the resulting colonies were checked by streaking onto selective plates, and by Southern blotting (see below).

Agrobacterium total DNA preparation

The *Agrobacterium* strain was grown up over 48 hours at 27°C in 5 ml of YEB. 1.5 ml of the culture was removed and spun down at 12000 g for 2 minutes. The supernatant was removed and the cells resuspended in 0.4 ml of pronase solution (380 µl of 50 mM Tris/HCl pH 8.0, 20 mM EDTA, 0.8% sarcosyl and 20 µl of fresh 20 mg/ml pronase). This was mixed thoroughly, incubated at 37°C for 1 hour then the lysate was sheared by adding 0.4 ml of dd water and vortexing for one minute followed by aspirating through a Gilson pipette several times. 0.4 ml of phenol was added, mixed, spun and 0.6 ml of supernatant was removed. This was repeated and 0.55 ml of supernatant was removed. 0.4 ml of chloroform was added, mixed, spun and 0.45 ml of supernatant was removed. This was repeated and 0.45 ml of supernatant was removed. The DNA was

precipitated by adding 2 volumes of 100% ethanol at room temperature, the tube was swirled to mix the two layers then shaken vigorously. The DNA was spun down and resuspended in 0.4 ml of sodium acetate pH 4.8 (with acetic acid). 2 volumes of 100% ethanol was carefully added at room temperature, the mixture incubated on ice for 10 minutes then mixed as before. The DNA was then spun down, 70% ethanol washed, vacuum dried and resuspended in 100 µl of dd water. It was sometimes necessary to perform further phenol extractions to clean up the DNA enough to restrict reliably.

Tobacco transformation

Leaf discs were cut from the young leaves (~3rd from the top) of tobacco plants (*Nicotiana tabacum*) (cultivar Samsun has been used successfully in this department), the centre rib was removed and the leaf edges were cut off. Preferably axenicly grown tobacco plants were used but if non sterile plants were used they were soaked for 15 minutes in 10% domestos then washed four times in sterile water. The discs were cultured for two days on solid MSO medium plus napthaleneacetic acid (NAA) 0.1 mg/l and 6-benzylaminopuridine (BAP) 1.0 gm/l at 25°C and 2000-4000 lux light level, 16 hours light 8 hours dark in sealed petri dishes. The construct in *Agrobacterium* was grown up in YEB plus antibiotic selection and spun down and resuspended in 20 ml of 2 mM magnesium sulphate three times then finally spun down and resuspended in 20 ml of MSO medium plus NAA and BAP as above. The leaf discs were infected by floating in the cell suspension for 1-3 minutes blotting dry on sterile 3MM paper and were then returned to their original plates.

The leaf discs were then cultured as before for 2-4 days until visible bacterial colonies appeared. Those that had developed callus were then transferred to selection on MSO plates plus NAA and BAP plus 100µg/ml of kanamycin. Culturing was continued for 2 days to 3 weeks, transferring to fresh selection plates if bacteria or fungus start to appear. When shoots became visible they were cut off the leaf discs free of callus if possible and transferred to tissue culture jars containing solid MSO plus 100 µg/ml kanamycin and 200 µg/ml augmentin. Roots developed in as little as 2-4 days. The plantlets were grown on till they were a few centimetres high and had a good root system when they were transferred to a peat based compost taking care not to damage the roots when the medium was washed off. After an initial period in compost in a sealed plastic bag

to stop wilting they were gradually opened to the air and then grown on as normal tobacco plants.

Preparation of λ ZAP II cDNA libraries

The cDNA libraries were made using the ZAP II kit supplied by Stratagene and according to the instructions in the manual supplied with it. The ZAP-cDNA synthesis kit method uses a hybrid oligo dT linker-primer, which contains an Xho I site. First strand synthesis is primed with the linker-primer and is transcribed using StrataScriptT RNase H^- reverse transcriptase and 5-methyl dCTP. The use of 5-methyl dCTP during first strand synthesis hemimethylates the cDNA, protecting it from digestion from certain restriction endonucleases such as Xho I. Therefore, on Xho I digestion of the cDNA, only the unmethylated site within the linker-primer will be cleaved. Hemimethylated DNA introduced into an $mcrA^+$ $mcrB^+$ strain would be subject to digestion by the mcrA and mcrB restriction systems. It is therefore necessary to initially infect an $mcrA^-$ $mcrB^-$ strain (e.g., the XL1-Blue MRF' strain supplied with the Uni-ZAP XR vector) when using the ZAP-cDNA synthesis kit. After passing the library through XL1-blue MRF' cells, it is no longer hemimethylated and can be grown on $mcrA^+$ $mcrB^+$ strains.

Stratagene recommends using their high-efficiency Gigapack II packaging extracts, since these packaging extracts are $mcrA^-$ $mcrB^-$ and mrr^-. Other commercially available packaging extracts can destroy hemimethylated DNA, therefore producing low titer libraries. Recombination frequency can be checked with blue-white (clear) colour selection by using XL1-Blue MRF' cells or by plating efficiencies of arms ligated to themselves, as opposed to comparing arms ligated to cDNA.

First strand cDNA synthesis begins when reverse transcriptase (RT), in the presence of nucleotides and buffer, finds a template and a primer. The template is mRNA and the primer is a 50-base oligonucleotide with the following sequence:

```
5′ GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTT 3′
     "GAGA" Sequence          Xho I        polydT
```

This oligonucleotide was designed with a "GAGA" sequence to protect the Xho I restriction enzyme recognition site and an 18 base polydT sequence. The restriction site

allows the finished cDNA to be inserted into the Uni-ZAP XR vector in a sense orientation (EcoR I-Xho I) with respect to the lacZ promoter. The polydT region binds to the 3' polyA region of the mRNA template, and StrataScript RNase H^- reverse transcriptase begins to synthesise the first strand cDNA. StrataScript RNase H-deficient (RNase H^-) reverse transcriptase is a novel Moloney murine leukemia virus reverse transcriptase (MMLV-RT) that possesses no detectable RNase H activity. Cloned StrataScript RNase H^- reverse transcriptase is purified from recombinant *E. coli* containing a genetically engineered mutant MMLV-RT gene. A point mutation has been introduced into the MMLV-RT gene, effectively abolishing the RNase H active site normally associated with the wild-type enzyme but not affecting the RT activity. Results from cDNA ladder synthesis and RT-mediated polymerase chain reaction (RT-PCR) experiments indicate that StrataScript RNase H^- reverse transcriptase produces larger yields of longer cDNA transcripts than those of wild-type MMLV-RT.

The nucleotide mixture for the first strand contains normal dATP, dGTP and dTTP plus the analog, 5-methyl dCTP. The complete first strand will have a methyl group on each cytosine base which will protect the cDNA from restriction enzymes used in subsequent cloning steps. The incorporation of 5-methyl dCTP versus normal dCTP has been tested with Klenow fragment, *E. coli* DNA polymerase I, *E. coli* DNA polymerase III, AMV-RT, MMLV-RT and StrataScript RNase H^- reverse transcriptase on various templates and has proved to be an equally acceptable substrate for the enzymes listed. During second strand synthesis, RNase H nicks the RNA bound to the first strand cDNA to produce a multitude of fragments, which serve as primers for DNA polymerase I. DNA polymerase I nick translates these RNA fragments into second strand cDNA. The second strand nucleotide mixture has been supplemented with dCTP to reduce the probability of 5-methyl dCTP becoming incorporated in the second strand. This ensures that the restriction sites in the linker-primer will be susceptible to restriction enzyme digestion. The uneven termini of the double-stranded cDNA are nibbled back or filled in with Klenow fragment, and EcoR I adaptors are ligated to the blunt ends. The adaptors have the sequence shown below.

```
5' AATTCGGCACGAG 3'
3'      GCCGTGCTC 5'
```

These adaptors are composed of 9 and 13mer oligonucleotides, which are complimentary to each other with an EcoR I cohesive end. The 9mer is kinased, which allows it to ligate to other blunt termini available in the form of cDNA and other adaptors. The 13mer is kept dephosphorylated to prevent it from ligating to other cohesive ends. After adaptor ligation is complete and the ligase has been heat inactivated, the 13mer is kinased to enable its ligation into the dephosphorylated vector arms. The Xho I digestion releases the EcoR I adaptor and residual linker-primer from the 3' end of the cDNA. These two fragments are separated on a Sephacryl column. The size-fractionated cDNA is then precipitated and ligated to the Uni-ZAP XR vector arms.

The lambda library is packaged in a high-efficiency system such as Gigapack II Gold packaging extract and is plated on the *E. coli* cell line XL1-Blue MRF'. Since most *E. coli* strains digest DNA containing 5 -methyl cytosine, it is important to plate on this $mcrA^-$ $mcrB^-$ strain.

Storage of the λ cDNA libraries

The λ cDNA libraries were stored by two methods; in glass or polypropylene bottles in the presence of 0.3% chloroform at 4°C and at -80°C with the addition of 7% DMSO.

λ ZAP II plasmid rescue and mass rescue

The Uni-ZAP XR vector has been designed to allow in vivo excision and recircularisation of any cloned insert contained within the lambda vector to form a phagemid containing the cloned insert. This in vivo excision is dependent on the DNA sequences in the lambda phage genome and on the presence of a variety of proteins, including f1 bacteriophage derived proteins. The proteins from the f1 phage recognise a region of DNA, which normally serves as the f1 bacteriophage origin of replication for positive strand synthesis. However, the origin of the plus strand replication can be divided into two overlying parts: the site of initiation and the site of termination for DNA synthesis. These two regions of the positive strand origin have been subcloned separately into the Uni-ZAP XR vector. The lambda phage (target) is made accessible to the f1-derived proteins by simultaneously infecting a strain of E coli with both the lambda vector

and the f1 bacteriophage. Inside *E. coli*, the helper proteins (i.e. proteins from f1 or M13 phage) recognise the initiator DNA that is within the lambda vector. These proteins then nick one of the two DNA strands. At the site of this nick, new DNA synthesis begins and duplicates whatever DNA exists in the lambda vector downstream (3 ') of the nicking site. DNA synthesis of a new single strand of DNA continues through the cloned insert until a termination signal positioned 3' of the initiator signal is encountered within the constructed lambda vector. The single stranded DNA molecule is circularised by a gene II product from the f1 phage forming a circular DNA molecule, which contains everything between the initiator and terminator. In the case of the Uni-ZAP XR vector, this includes all sequences of the phagemid, Bluescript SK(-), and the insert, if one is present. This conversion is the subcloning step, since all sequences associated with normal lambda vectors are positioned outside of the initiator and terminator signals and are not contained within the circularised DNA. In addition, the circularising of the DNA automatically recreates a functional f1 origin as found in the f1 bacteriophage or phagemids.

Signals for packaging the newly created phagemid are contained within the f1 terminator origin DNA sequence. They permit the circularised DNA to be packaged and secreted from the *E. coli*. Once the phagemid is secreted, the *E. coli* cells used for in vivo excision of the cloned DNA can be removed from the supernatant by heating at 70°C. The heat treatment kills all the E coli cells, while the phagemid remains resistant to the heat treatment. For production of double stranded DNA, the packaged Bluescript DNA is mixed with fresh *E. coli* cells and is spread on LB ampicillin plates to produce colonies. DNA from minipreps of these colonies can be used for analysis of insert DNA including DNA sequencing, subcloning, mapping and expression. Bluescript colonies can also be used for subsequent production of single-strand DNA suitable for dideoxy sequencing and site-specific mutagenesis.

The ExAssist/SOLR system is designed to allow efficient excision of the Bluescript phagemid from the Uni-ZAP vector, while eliminating problems associated with helper phage co-infection. The ExAssist helper phage contains an amber mutation that prevents replication of the phage genome in a nonsuppressing *E. coli* strain such as SOLR cells. This allows only the excised phagemid to replicate in the host, removing the possibility of productive co-infection from the ExAssist helper phage. Since the ExAssist helper phage cannot replicate in the SOLR strain, single-stranded rescue cannot be performed in this strain using this helper phage. Mass excision can be used to generate subtraction libraries

and subtraction DNA probes.

The plaque of interest was cored from the agar plate and transferred to a sterile microcentrifuge tube containing 500 µl of SM buffer and 20 µl of chloroform. The tube was vortexed to release the phage particles into the SM buffer. This was incubated for 1-2 hours at room temperature or overnight at 4°C (This phage stock is stable for up to 1 year at 4°C). Overnight cultures of XL1-Blue MRF' and SOLR cells in LB broth were grown at 30°C. A 1/100 dilution of the cells was made using 0.5 ml of the overnight and 50 ml of LB broth. This was grown at 37°C for 2-3 hours to mid-log phase (OD_{600} = 0.2-0.5). The XL1-Blue MRF' cells were gently spun down (1500 x g) and resuspended at OD_{600} = 1.0 for single-clone excision or at OD_{600} = 5.0 in 10 mM magnesium sulphate for mass excision.

The SOLR cells were allowed to grow to OD_{600} = 0.5-1.0, while continuing with the excision. Before the SOLR cells reach $OD_{600} \geq 1.0$, the cells were removed from the incubator and allowed to incubate at room temperature. The cells should be used the same day.

In a 50-ml conical tube was combined;

200 µl of OD_{600} = 1.0 XL1-Blue MRF' cells
250 µl of phage stock (containing >1 x 10^5 phage particles)
1 µl of ExAssist helper phage (>1 x 106 pfu/ml)

The mixture was incubated at 37°C for 15 minutes. N.B. when excising an entire library, 10- to 100-fold more of the amplified lambda phage should be excised than is found in the primary library to ensure statistical representation of the excised clones. Cells were added at a 10:1 cells to amplified lambda phage and ExAssist helper phage was added at a 10:1 phage-to-cell ratio. For example;

10^8 cells (1 OD_{600} = 8.0 x 10^8 cells/ml)
10^9 ExAssist helper phage
10^7 pfu of amplified library
Incubate at 37°C for 15 minutes

3 ml of LB broth was added (25 ml of LB for mass excision) and incubated for 2-2.5

hours at 37°C with shaking. Incubation times for mass excision in excess of 3 hours may alter the clonal representation. Single clone excision reactions could be safely performed overnight, since clonal representation was not relevant. N.B. cloudy growth was not always be seen. The cells were spun down for 15 minutes at 2000g and the supernatant transferred to a fresh tube. The tube was heated at 70°C for 15 minutes and then spun again for 15 minutes at 4000g. (N.B. to produce a subtraction library, purify single-stranded DNA. Perform the subtractive hybridisation and then transform the SOLR cells with the subtracted clones.) The supernatant was decanted into a sterile tube. This stock contained the excised phagemid Bluescript packaged as filamentous phage particles, and it could be stored at 4°C for 1-2 months. The excised phagemids were plated by adding 200 µl of freshly grown SOLR cells (OD_{600} = 1.0) to two 1.5 ml tubes, then adding 100 µl of the phage stock (1 µl for mass excision) to one tube and 10 µl of the phage stock to the other tube. The tubes were incubated at 37°C for 15 minutes. 10-50 µl from each tube was plated on LB ampicillin plates (50 µg/ml) and incubated overnight at 37°C. Due to the high-efficiency of the excision process, it was sometimes necessary to titrate the supernatant to achieve single-colony isolation.

Colonies appearing on the plate contain the Bluescript double stranded phagemid with the cloned DNA insert. Helper phage will not grow, since they are unable to replicate in Su^- (nonsuppressing) SOLR strains and do not contain ampicillin resistance genes. SOLR cells are also resistant to lambda phage infection, thus preventing lambda phage contamination after excision. From this point the Bluescript double stranded phagemid may be treated as a normal plasmid.

VCSM13 Interference-Resistant Helper Phage is recommended in place of ExAssist helper phage for the the single-stranded rescue procedure.

ELISA

ELISA, enzyme linker immunosorbent assay, is a procedure whereby antigens (or antibodies) are immobilised on a solid phase, then reacted with primary antibody. The appropriate conjugated secondary antibody is then reacted with the antigen-antibody complex. The addition of substrate, which changes colour in the presence of an enzyme, indicates a positive reaction. Ideally, the rate and degree of substrate colour change caused by the enzyme conjugate provides a measure of the amount of the target protein.

Optimisation of an ELISA to fit a specific need involves the manipulation of three variables: reagent concentration, temperature and length of incubation.

ELISA was done with either buffers and reagents made up in the laboratory or with the components supplied with the ELISA kit by Kirkegaard and Perry. All steps in the procedure were carried out at room temperature. Before use all solutions were warmed to room temperature and mixed. Plates should be specifically formulated for ELISA as polystyrene plates made for tissue culture often produced erratic backgrounds. In the first step antigen or antibody diluted in Coating Buffer (an optimised phosphate buffered saline) or PBS+T was added to the plate wells and incubated at room temperature for one hour at room temperature or overnight at 4°C. Following coating, the plate was blocked by adding BSA Diluent/Blocking Solution to the emptied wells and incubating for five minutes. This solution contains 1% bovine serum albumin (BSA) in phosphate-buffered saline. It reduces non-specific binding by covering any unreacted sites on the plate surface and protects adsorbed protein from surface denaturation. Plates could be stored at this point at 4°C by covering with a lid and sealing to avoid evaporation. If this was done the plates should be allowed to warm to room temperature before use.

Every assay should include appropriate controls to verify the performance of the test system, define the background, and establish that the material measured is the limiting component of the assay system. Both controls and samples should be diluted in BSA Diluent/Blocking Solution to keep background to a minimum. Ideally, all tests should be performed in duplicate.

The washing procedure was performed after addition of sample and conjugate and helps to reduce background colour by removing unbound reactant from the wells. Between washes plate wells were emptied by tapping the plate up side down to remove residual moisture. The wash solution contains 0.02% Tween 20, a wetting agent that minimises non specific binding. If the procedure was interrupted at any point the wells were filled with Wash Solution so that the plate did not dry out. Plates were usually washed 4-6 times at each step.

The enzyme in the secondary antibody conjugate is the detection system which by reacting with the substrate, demonstrates the presence of the target protein. Care was taken during the procedure to avoid contact with EDTA in the case of alkaline phosphatase and sodium azide in the case of peroxidase as these products inactivate the conjugated enzymes.

The chromogenic substrate provides a sensitive detection method for the enzyme in the conjugate. Generally, the colour produced is proportional to the amount of unknown in the sample. The ABTS Peroxidase Substrate, used with peroxidase-linked antibodies, produces an intense blue-green reaction product which may be visually detectable in minutes. Colour developed quickly for 1-2 hours, after which it reached a plateau for several hours. The use of Stop Solution was an optional step that interrupted further colour development in the wells, maintaining a desired level of colour intensity for visual reading. All colour reagent solutions were mixed the same day as they were used.

The basic protocol was; Binding of the antigen to the plate in 100 µl of PBS for 1 hour, washing in PBS+T, blocking in PBS + 0.1-1% BSA for 1 hour, adding primary antibody at the appropriate dilution in blocking solution and incubating for 1 hour, washing, adding the secondary antibody diluted as above and incubating for 1 hour, washing, adding the substrate mix and allowing the colour to develop, optionally adding stop solution, then monitoring on a microtitre plate reader at the appropriate wavelength.

For peroxidase labelled antibody the stop solution was 1 volume of 1% SDS and the colour was monitored at 410 nm.

Crude protein extraction from E. coli

Cells were resuspended in 1x SDS Sample Buffer, disrupted by boiling for 5 minutes and cell debris removed by centrifugation for 5 minutes at 12000 g. The product of ~1 ml of broth was resuspended in 100 µl normally. Samples were then electrophoresed with 5-15 µl sample loaded per well with β-mercaptoethanol being added to > 1%.

Disruption of E. coli by sonication

Harvested cells were resuspended in 10 mM Tris pH 8.0, 2 mM EDTA and transferred to corex tubes, such that tubes were two thirds full of liquid. An MSE Soniprep 150 sonicator was employed, using 10 second bursts, with 10 or 20 second intervals between bursts. Where possible samples were kept on ice to prevent heat gain during sonication. Cell debris was removed by centrifugation at 16000 g for 10 minutes.

SDS PAGE

	12.5% Main Gel	15% Main Gel	Stacking Gel
Acrylamides	6.25ml	7.5ml	1.25ml
Tris-HCl, pH 8.8	1.875ml	1.875ml	-
Tris-HCl, pH 6.8	-	-	2.5ml
Distilled water	6.165ml	4.915ml	5.650ml
De-gas briefly			
10% SDS	150µl	150µl	100µl
2% Ammonium Persulphate	560µl	560µl	500µl
TEMED	7.5µl	7.5µl	7.5µl

	10% Main Gel	17.5% Main Gel	7% Main Gel
Acrylamides	5.0ml	8.75ml	3.5ml
Tris-HCl, pH 8.8	1.875ml	1.875ml	1.875ml
Distilled water	7.415ml	3.665ml	8.915ml
De-gas briefly			
10% SDS	150µl	150µl	150µl
2% Ammonium Persulphate	560µl	560µl	560µl
TEMED	7.5µl	7.5µl	7.5µl

(7% gel should give 25 - 200 kDa resolution)

Acrylamides:
30g acrylamide
0.8g bisacrylamide
Make to 100ml with d.H_2O
Filter, store dark at 4°C

Stacker Buffer:
0.5M Tris HCl pH6.8
(6g Tris /100ml)
Filter and store 4°C

Stain:
0.05% Kenacid Blue R (0.5g/l)
40% Methanol
7% Glacial Acetic Acid

2xSDS Sample Buffer:
1.52g Tris pH6.8 (0.2M)
20ml glycerol (20%)
2g SDS (2%)
2mg BPB (0.002%)/100ml

10x Reservoir Buffer:
0.25M Tris (30.3g/l)
1.92M glycine (144g/l)
1% SDS (10g/l)
pH 8.3

Resolving Buffer:
3.0M Tris HCl pH8.8
(36.3g Tris/100ml)
Filter and store 4°C

Destain:
40% Methanol
7% Glacial Acetic Acid

SDS7 Markers:
66k Bovine albumin
45k Egg albumin
36k Glyceradehyde-3-phosphate dehydrogenase
29k Bovine erythrocyte carbonic acid
24k PMSF-treated trypsinogen
20k Soybean trypsin inhibitor
14k Alpha-lactalbumin

These recipes are sufficient for two minigels.

Figure 6. Quantities for making up acrylamide gels for SDS PAGE and some useful data.

Proteins were electrophoresed on polyacrylamide gels, run in the dissociating SDS-PAGE buffer system (Laemmli 1970). 8 x 10 cm gels were cast and run in an ATTO AE-6450 apparatus, they were prepared as described by Hames (1981). Samples were prepared as above and Mr standards were loaded into wells, β-mercaptoethanol was added if required for reducing (about 1µl per well) and the gels were run at 50 V. Proteins were stained with Kenacid blue overnight and destained as required. Gels were photographed or scanned or dried between cellophane under vacuum.

Method and quantities for preparing gels see figure 6.

Colony screening

E. coli colonies were streaked out in duplicate onto gridded nitrocellulose filters laid onto agar plates with the appropriate selection according to a master numbering scheme for the grid. The *E. coli* on the filters was grown up overnight as normal and one of the pair sealed and stored at 4°C. The other filter was processed by laying onto 3MM filter paper saturated with buffer colony side up as follows; 10% SDS 3 minutes, denaturing solution 5 minutes, neutralising solution 5 minutes, 2 x SSPE 5 minutes then allowed to air dry on 3MM paper before being wrapped in more 3MM paper and baked at 80°C in a vacuum oven for 1-2 hours. They were stored in the dry and if the filters stuck to the 3MM paper soaking briefly in 6 x SSC would free them.

Preparation of genomic DNA from plants

Genomic DNA was isolated essentially by the method of Ellis et al (1984); Approximately 1 gm of frozen leaf was ground with a mortar and pestle cooled with liquid nitrogen. 5 ml of extraction buffer was added (0.45 sodium chloride, 45 mM trisodium citrate, 0.1 M sodium diethyl dithiocarbamate, 0.1 M EDTA, pH 8.9, filter sterilised) and warmed to melt whilst mixing using sterile spatula. 100 µl of 20% SDS was added and mixed in. This was then extracted with 10 ml of 24:1 chloroform/iso-amyl alcohol by mixing then transferring to a corex tube mixing again by inversion and centrifuging at 4000g for 10 minutes. The upper aqueous phase was removed, transferred to a sterile 100 ml beaker, overlayed with 20 ml of ethanol and the DNA spooled out into an

eppendorf. This was washed once with 70% ethanol and vacuum dried briefly or dried with a stream of nitrogen gas. The pellet was resuspended in 500 µl of TE buffer overnight on a wheel at 4°C, extracted with 500µl phenol/cresol mix (500µl phenol, 70 µl cresol, 0.5 mg 8-hydroxy quinoline) and ethanol precipitated by adding 1 ml of room temperature ethanol and mixing by inversion. The DNA clot was removed, washed and dried as above. Finally it was resuspended in 500µl of either TE or dd water.
DNA prepared by this method was not free of RNA so if an accurate concentration was required the spotting method or the diaminobenzoic acid method was used.

Preparation of RNA

RNA is very sensitive to nucleases and great care was taken to ensure that all apparatus used with RNA was nuclease free. All glassware was baked overnight at 170°C, all plastic disposables were either supplied sterile and used directly from the pack or soaked in water plus 0.1% DEPC (diethylpyrocarbonate) at 37°C for at least 1 hour to inactivate nucleases before autoclaving. All solutions were treated similarly except those which could not be autoclaved which were filter sterilised or those containing free amine groups which decomposes DEPC and were assembled from components as nuclease free as possible. Other apparatus was freed from nuclease contamination by the chaotropic action of the guanidinium thiocyanate solution.

Preparation of total RNA

RNA was prepared by the guanidinium thiocyanate method (see Ausubel et al 1992) from both plant and insect sources; the tissue was collected directly into liquid nitrogen in pieces of no more than 2 grams each and if mechanically tough was ground in liquid nitrogen in a cooled mortar and pestle just prior to extraction. To ~2 gm of tissue 20 ml of tissue guanidinium solution was added (see below) and the whole immediately ground in the polytron at full speed for three 15 second bursts. It was important that the tissue did not melt before grinding had started. The resulting mix was centrifuged at 12000 g 12°C for 10 minutes. The supernatant was taken off, 0.1 volume of 20% sarcosyl added and heated for 2 minutes in a 65°C waterbath. Solid caesium chloride was added to 0.1 g/ml and dissolved then the solution was layered onto a 9 ml cushion of 5.7 M caesium

chloride in an ultracentrifuge tube and centrifuged over night at 113000 g and 22°C.

The supernatant was carefully aspirated off to leave the RNA pellet stuck to the bottom of the tube which was inverted to drain. the bottom of the tube with the pellet was cut off and the RNA resuspended in 3 ml of tissue resuspension buffer (5 mM EDTA, 0.5% N-lauroylsarcosine (Sarcosyl), 5% β-mercaptoethanol) overnight at 4°C. If necessary the pellet was macerated with a sterile tip and heated to 65°C briefly. The solution was extracted with 1:1 phenol/chloroform then 24:1 chloroform/iso-amyl alcohol and ethanol precipitated with 0.1 volume of 3 M sodium acetate pH 5.2 and 2.5 volumes of ethanol. After spinning down etcetera as normal it was resuspended in DEPC treated dd water and stored at -80°C or in liquid nitrogen.

Tissue guanidinium solution is made as follows; for 1 litre dissolve 590.8 gm of guanidinium isothiocyanate in ~400 ml of DEPC treated water. Add 25 ml of 2 M Tris/HCl pH 7.5 and 20 ml of 0.5 M EDTA pH 8.0. Stir overnight , adjust the volume to 950 ml and filter. Finally add 50 ml of β-mercaptoethanol.

Preparation of polyA+ RNA

PolyA+ RNA was prepared from total RNA by use of the PolyATract kit (Promega) and according to the suppliers instructions. The system uses biotin labelled oligo dT to select the polyA+ RNA and streptavidin bound to a paramagnetic particle to select the biotin. Then the polyA+ RNA is isolated from the solution by capturing the paramagnetic particles with a magnet. After a couple of washes the polyA+ RNA is released by lowering the salt concentration allowing it to unhybridise from the oligo dT into the solution. The paramagnetic particles and the biotin oligo dT are captured with the magnet and the supernatant containing the polyA+ is removed. It was stored at -80°C or in liquid nitrogen.

DNA sequencing

DNA sequencing was done on an Applied Biosystems 373 DNA sequencer by both dye primer and dye terminator chemistries according to the manufacturers protocols and on both double stranded plasmid DNA and single stranded M13 DNA.

DNA sequence handling and computer programs

DNA sequence data and other sequence data was processed on Apple Macintosh computers as was all other computing. The following software was used;

MacWrite Pro,	Word processing	Claris
Sequencher	Sequence assembly and analysis	Gene Codes Corporation
SeqEd	Sequence assembly and display	Applied Biosystems
Statview	Statistical package	Abacus Concepts Inc.
Cricket Graph III	Graphing package	Computer Associates Int.
ClarisDraw	Drawing program	Claris
Photoshop	Image manipulation	Adobe
Seqvu	Sequence display	Shareware
Endnote Plus	References	Niles and Associates Inc.
DNA Strider	Sequence analysis	Christian Marck

Also used was the gcg package of programs (gcg 1994), including fasta, which were were used to search the nation protein sequence database at Daresbury.

Integral gut membrane protein preparation

Integral gut membrane protein was prepared by the method used by Munn (1992). It works by fractionating the proteins on the basis of their solubility in solutions containing increasingly strong detergents. Guts were dissected from about 30 Diabrotica larvae which were 4th instar, approximately 15 cm long and still actively feeding. These were homogenised in ice cold PBS plus 1mM EDTA and 1mM PMSF and centrifuged for 10 minutes in a bench top centrifuge. The pellet was resuspended in the same buffer plus 0.1% Tween 20 and centrifuged as before. The pellet was resuspended in same buffer plus 2% TritonX-100 and left at 4°C for 2 hours or overnight then centrifuged as before. The integral membrane proteins remain in the supernatant which was stored at -20°C.

Aminopeptidase assays

Aminopeptidase assays were performed essentially by the method used by Munn (1992). The substrates used were the paranitroanilide derivatives of amino acids (pNA derivatives). The development of the yellow colour caused by the cleavage of the paranitroaniline from the amino acid was monitored at 405 nm on a microtitre plate reader. The assay buffer was 50 mM HEPES pH 7.0 which, while it may not be physiologically accurate for gut MAPS, was chosen as it was specified in the activity definition of the test enzyme; porcine kidney leucine aminopeptidase (Sigma). The assay was found to be quite sensitive to temperature.

To 250 µl of assay buffer in the well of a microtitre plate (flat bottomed) was added a suitable dilution of the preparation to be assayed in 10 µl of assay buffer. The plate was shaken and preincubated at the assay temperature for 10 minutes then 10 µl of 25 mM substrate was added mixed by shaking and the OD_{405} monitored at intervals.

Results; potato carboxypeptidase inhibitor (PCI)

Potato carboxypeptidase inhibitor is a small protein of molecular weight 4300 Daltons which is found mainly in potato tubers and to a lesser extent throughout the plant (Hass, Derr and Makus 1979). It has a tightly folded and cysteine cysteine bonded three dimensional structure of great stability (Leary *et al.* 1979), and inhibits carboxypeptidase A by a one to one binding mechanism (Ako, Hass, Grahn and Neurath 1976; Hass and Ryan 1981). From protein sequence data (Hass *et al.* 1979) it is known to be coded for by at least two genes. An homologous protein has also been isolated from the related species tomato (Hass and Ryan 1979).

Preparation of PCI protein

PCI protein was prepared by the method of Pearce and Ryan (1983) which exploits some novel properties of the protein namely, it is soluble in 80% ethanol and that it is an unusually stable protein surviving heating to 80°C in solution and *in situ* in potato tubers during cooking. Intact, functional PCI protein can be extracted from commercial instant mash (ibid). A sample of the purified protein prepared from an unknown variety of white potato tubers, as described in the methods section, was run on SDS PAGE and showed no high molecular weight contamination. The purified PCI protein was shown to be biologically active in assays with bovine carboxypeptidase A, and with *Heliothus virescens* gut preparations, and gave essentially identical results to a purified protein preparation obtained from Sigma.

Carboxypeptidase A inhibition assays

Carboxypeptidase A enzyme activity was assayed by the method of Peterson, Holmquist and Bethune (1982) which monitors the decrease in UV absorption at 330 nm caused by the hydrolysis of FAPP (N-(2-furanacryloyl)-L-phenylalanine-L-phenlyalanine). due to the cleaving off of the C-terminal phenylalanine. The C-terminal phenylalanine provides specificity for carboxypeptidase A (ibid) as well as giving a higher rate of hydrolysis than other amino acids. It is the amino acid of choice in other carboxypeptidase A assays which use a variety of phenylalanine derivatives (Hass and

Ryan 1981) and is used in the definition of the enzyme activity (see below).

Two methods were used; with preincubation where the enzyme and inhibitor were incubated together at 25°C prior to adding to the substrate, and without preincubation where the reaction was initiated by adding the enzyme to the rest of the temperature-equilibrated components. The OD_{330} was monitored as the reaction proceeded, and the fall in the absorbance was measured, until the slope of the absorbance/time plot could be estimated.

The diluted enzyme was unstable due to self digestion in the absence of a significant concentration of protein, such as leaf extracts, and it slowly lost activity even at 0°C during experiments. However, inhibition assays on leaf extracts without preincubation showed a more curved trace for those with inhibitor present than for those without, due presumably to the rate at which the inhibitor was able to associate with the enzyme. Normally the rate of hydrolysis was estimated from the start of the absorbance/time trace as this gave the best straight line; longer incubations showed a falling off of the rate of hydrolysis as the lower absorbance values were approached, presumably as the substrate concentration fell significantly. This effect was observed in assays with no inhibitor present and hence is not entirely due to the kinetics of association of the enzyme and inhibitor. Enzyme activity was due solely to the added enzyme preparations since in no case was any activity seen when preincubating either the leaf extracts or the inhibitor dilutions with the substrate buffer mix.

The temperature of 25°C was chosen as it matched the unit definition of bovine carboxypeptidase A; 1 unit will hydrolyse 1.0 µmol of hippuryl-L-phenylalanine per minute at pH 7.5 and 25°C (data sheet supplied by Sigma with the enzyme), and as physiologically realistic for an insect enzyme. 800 µl of 200 nM FAPP is 1.6×10^{-7} moles of FAPP, assuming the activity of the enzyme is the same for FAPP as for hippuryl-L-phenylalanine then 0.01 units of enzyme will hydrolyse the FAPP in one reaction to completion in approximately 16 minutes. This was confirmed for the initial rate of a test reaction and although the absolute absorbances were lower than those observed by Peterson, Holmquist and Bethune (1982) (~0.9 - 0.65 at completion as opposed to ~ 1.85 - 1.5.) the dynamic range was similar. The differing absolute absorbances were probably due to the presence of absorbing substances in the serum enzyme preparations they used.

Potato carboxypeptidase inhibitor protein has a specific activity definition of; 1 µg of PCI protein will inhibit 5.3 µg of bovine carboxypeptidase A by 50 % using hippuryl-L-

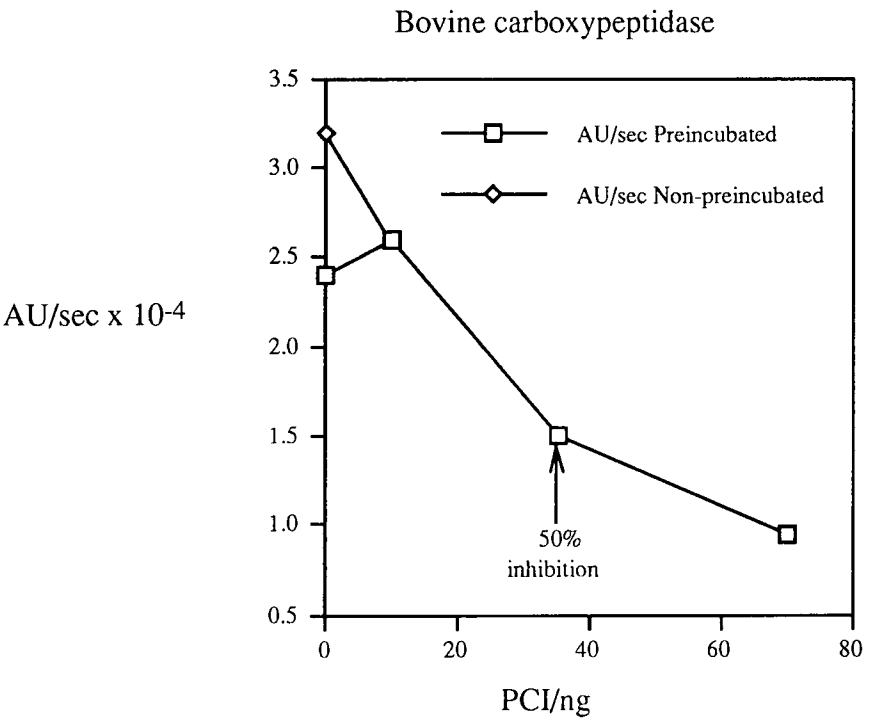


Figure 7. The inhibition of bovine carboxypeptidase A by potato carboxypeptidase inhibitor as monitored by the hydrolysis of FAPP.

phenylalanine as substrate at pH 7.5 and 25°C (data sheet supplied by Sigma with a sample of the the inhibitor). Therefore 35 ng (7 µl of a 5 ng/µl solution) should inhibit the standard assay by 50%. This was confirmed (figure 7) by a test experiment using the preincubation method for all points except the repeats of the zero inhibitor concentration point which were done by the non preincubation method. This was because of the anomalously low rate of reaction for the preincubation method zero inhibitor point which resulted from self digestion of the enzyme during preincubation.

The line obtained was quite strongly curved at the higher PCI concentrations which is similar to the data obtained by Ako, Hass, Grahn and Neurath (1976) with hippuryl-L-phenylalanine as substrate. These authors performed stoichiometric titrations of the inhibitor and enzyme using chloroacetyl-L-tyrosine as substrate instead of hippuryl-L-phenylalanine which gave essentially a straight line at the cost of insensitivity as an enzyme assay. This curved line effect might be expected to be more marked with FAPP than with hippuryl-L-phenylalanine as substrate as the enzyme should have a higher affinity for FAPP than for hippuryl-L-phenylalanine (Peterson, Holmquist and Bethune 1982).

Insect carboxypeptidase A like enzymes

Callosobruchus maculatus total gut homogenate in 1 mM DTT, and *Heliothus virescens* purified water soluble gut contents (chloroform extracted and twice 95% ammonium sulphate precipitated), were kindly supplied by Dr. K. Johnson. They were both assayed for carboxypeptidase A activity by the preincubation method. The *C. maculatus* preparation, a cloudy suspension which includes actual gut as well as contents, gave a weak activity equivalent to $4x10^{-4}$ units of bovine carboxypeptidase A for approximately one gut. This indicated that a carboxypeptidase A like enzyme was present. This activity was not further investigated due to a shortage of material, and the cloudy nature of the preparation which absorbed strongly at 330 nm. The *H. virescens* preparation showed strong activity equivalent to $4x10^{-3}$ units of bovine carboxypeptidase A for 10 µl of preparation which is approximately 4% of one gut. Thus one *H. virescens* gut contains approximately 0.1 units of carboxypeptidase A like enzyme. This is directly comparable with the results of Christeller, Laing, Marwick and Burgess (1992) who found 0.27 units for *Heliothus armigera conferta* and 0.51 units for *Heliothus punctigera* of carboxypeptidase A activity per unpurified midgut, using hippuryl DL-phenyllacetic

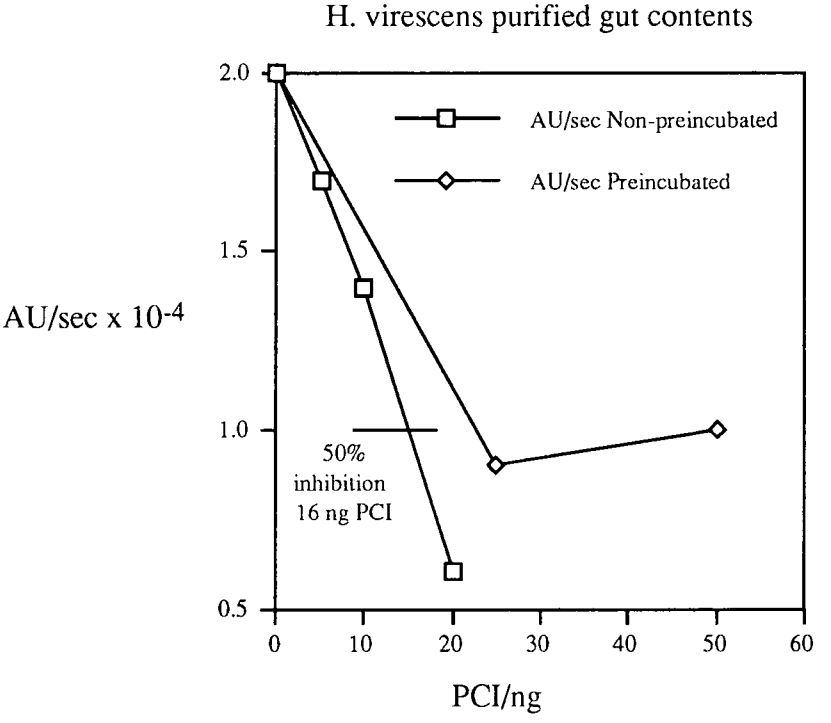


Figure 8. Carboxypeptidase activity from *Heliothus virescens* purified gut contents as monitored by the hydrolysis of FAPP and the effect of potato carboxypeptidase inhibitor on that activity.

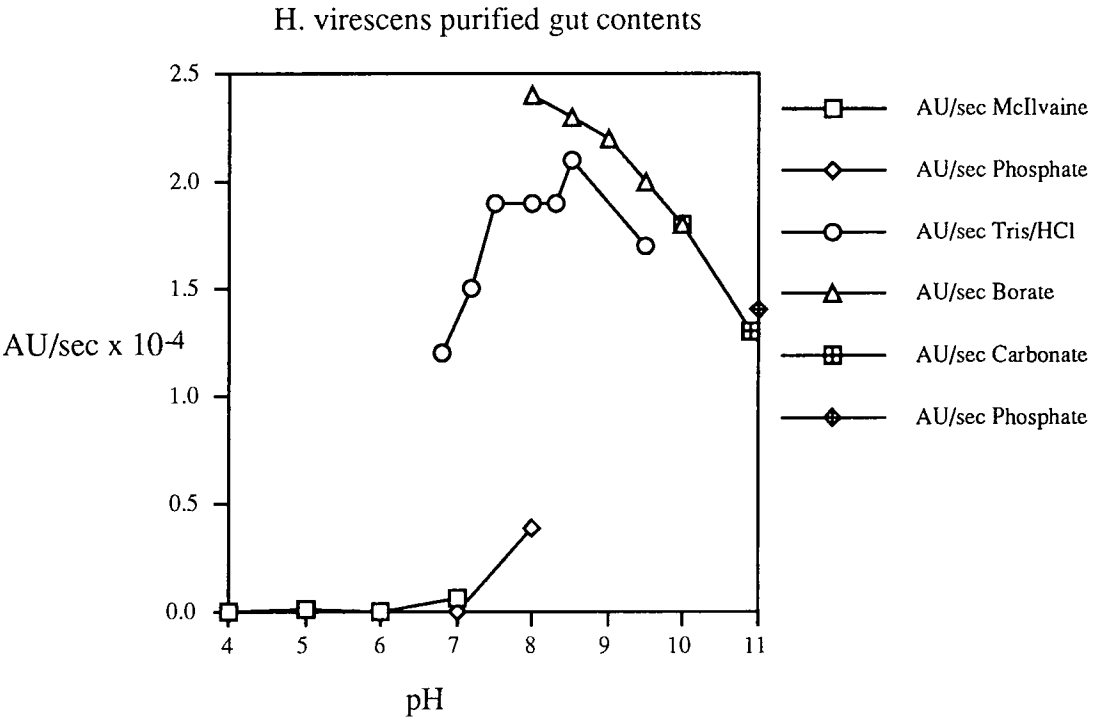


Figure 9. The effects of pH on the Heliothus virescens gut aminopeptidase activity as monitored by the hydrolysis of FAPP.

acid as substrate. These are species closely related to *H. virescens* and have carboxypeptidase A activities on the low side for the Lepidoptera surveyed with other species having up to twenty fold more activity when expressed as activity per milligramme of midgut (ibid).

The *H.virescens* gut contents preparation was used in assays with PCI by both assay methods and was shown to be susceptible to inhibition by PCI (figure 8). This is similar to the results of Hass *et al.* (1981) which showed that whole homogenate of another Lepidopteran (cabbage looper, exact species unspecified) contained a carboxypeptidase activity inhibitable by PCI. The two assay methods gave markedly differing results for the inhibition versus PCI concentration plot with no more than approximately 50% inhibition seen in the preincubation method, a straight line inhibition titration for the non-preincubation method but identical rates for the no inhibitor points from each method. The *H. virescens* gut contents preparation contained other active proteolytic enzymes including trypsin, chymotrypsin and elastase-like activities (Johnston *et al.* 1995), and these may be degrading the PCI during the preincubation process though this does not explain the ~50% plateau in inhibition.

The non-preincubation method results show that 16 ng of PCI gives 50% inhibition of 10 µl of gut preparation, equivalent to 4% of a gut contents. This is directly comparable with the purified bovine carboxypeptidase A, where 0.004 units of enzyme would require 14 ng of PCI for 50% inhibition. PCI is known to function by binding to the bovine carboxypeptidase A active site, via its carboxy terminal amino acid residues, in a one to one stoichiometric ratio (Rees and Lipscomb 1982). Thus unless the insect enzyme is greatly different from the mammalian one, which is unlikely as PCI would not be expected to inhibit an enzyme of radically different structure, the enzymes are of very similar specific activity.

The *H.virescens* carboxypeptidase A-like enzyme was further characterised by determining its pH optimum (figure 9). This was achieved by performing a series of assays in standard buffer mixes (CRC Handbook of Biochemistry 1968 and references within) plus 0.45 M sodium chloride or 50 mM Tris/HCl plus 0.45 M sodium chloride over the pH range 4 to 11. All reactions were preincubated with all components except the enzyme to check that the FAPP was stable in the buffer and pH used.

There was a strong buffer effect, with the preparation being virtually inactive in the acid to neutral buffers containing phosphate at pHs where it was active in Tris buffer, and

it was concluded that the *H. virescens* carboxypeptidase A-like enzyme activity is inhibited by $H_2PO_4^{2-}$. It can be seen that the carboxypeptidase A activity has a broad pH optimum with >50% activity in the pH range 7.0 to >11.0 with maximum activity at pH 8.0 to 8.5. This compares with the *H. virescens* trypsin like enzyme which has a similar broad pH optimum with maximum activity between pH 10.0 and 10.5 (Johnston *et al.* 1995).

Although the pH curves were similar, the actual peak in activities differed by 2 pH units. This difference may be explained by the two enzymes having evolved to have maximum activity in differing parts of the gut and/or being produced at different sites within the gut, i.e. it might be expected that carboxypeptidase activity would extend further towards the hind-gut than trypsin activity. The broad pH optima can also be explained as adaption of the enzymes to a range of gut pHs with the spread rather than the actual peak being the significant feature. This illustrates the unreliability of using enzyme pH optima as a guide to actual gut pH, particularly without localisation of the enzyme in question.

The mammalian carboxypeptidase A enzymes are known to be metallo-enzymes with a zinc atom at the active site (Allan *et al.* 1964). Deactivation to the apoenzyme occurs with removal of zinc and activity is fully restored by the addition of zinc to the metal free enzyme. The dependence of the *H. virescens* enzyme activity on zinc was demonstrated by the following experiment. 1,10 phenanthroline, which complexes zinc strongly and will remove it from the mammalian enzyme active site, when added to ~1 mM decreased the rate of FAPP hydrolysis more than 100 fold although it did not completely halt it, and the subsequent addition of zinc to ~5 mM restored much of the activity. Thus the carboxypeptidase A activity in *H. virescens* gut contents preparations is similar to the mammalian enzyme in specific activity, is a metallo-enzyme, is susceptible to inhibition by PCI and most likely is derived from a gene homologous to those for the mammalian enzymes.

Screening a potato tuber cDNA library for PCI cDNAs

PCI protein is present in many parts of the potato plant (Hass, Derr and Makus 1979) and is also induced by wounding (Graham and Ryan 1981) but is present at the highest concentrations in the tuber, so a tuber cDNA library was the library of choice to screen for PCI cDNA clones. A potato tuber cDNA library made from a Cyprus variety of new potato in λ ZAP II was available in the department and was kindly loaned by

Potato Carboxypeptidase inhibitor oligo design

```
                                    oligo #486
                                   __________________
PCI aa seq.         Q  Q  H  A  D  P  I  C  N  K  P  C  K  T  H  D  D  C  S  G  A  W  F  C  Q  A  C  W  N
PCI possible       CAACAACACGCNGACCCNATATGCAACAAACCNTGCAAAACNCACGACGACTGCTCNGGNGCNTGGTTCTGCCAAGCNTGCTGGAAC
DNA seq.             G  G  T     T     C  T  T  G     T  G     T  T  T  TAGC           T  T  G     T     T
                                       T                                 AGT
                                                                                  __________________________
                                                                                         oligo #487

                    S  A  R  T  C  G  P  Y  V  G
                   TCNGCNCGNACNTGCGGNCCNTACGTNGGN
                   AGC   A A     T        T
                   AGT   A G
```

Oligos

```
Oligo #486, 17 mer, 16 fold redundant, made in sense: TGCAACAAACCITGCAA
                                                        A  T  G     T

Oligo #487, 23 mer, 16 fold redundant, made in antisense: AACCAAACAIGCCTGACAAAACCA
                                                               G      T  G  G
```

Figure 10. The design of the oligos for potato carboxypeptidase inhibitor from its protein sequence.

Professor T. Slabas and group.

Two oligos were designed from the published protein sequences of the PCI isoinhibitors (Hass *et al.* 1975; Hass, Derr, Makus and Ryan 1979) (figure 10). The longer of the two, oligo #487, which was designed to sequence near the carboxy-terminus of the protein was a sixteen fold redundant 23mer containing one inosine. It was made in antisense and was the oligo used for all the primary screening. The shorter, oligo #486, encoded a region near the amino-terminus of the protein and was a sixteen fold redundant 17mer, contained one inosine, was made in sense and was used for confirmatory screening of the positives from oligo #487.

The melting points of the oligos were calculated from the approximate relationship;

$$Tm = 4(G+C) + 2(A+T+I)\ ^{\circ}C$$ (Sambrook, Fritsch and Maniatis 1989)

modified by the assumptions that inosine binds to a random base as strongly as adenosine to thymidine and that all mixed bases have the lower of the possible melting points, i.e. are A or T not G or C. Oligo #487 is slightly longer than the recommended maximum size this relationship is thought to hold for, oligos over 18 bases long being given too high a melting point by this method. However the alternative recommended method of calculating oligo melting points;

$$Tm = 81.5 - 16.6(\log_{10}[Na^{+}]) + 0.41(\%\ G + C) - (600/N)$$

Where N = chain length (Sambrook, Fritsch and Maniatis 1989)

gave an even higher result for Tm in 5 x SSC (hybridisation solution) of 71°C as opposed to 60°C. Hybridisation and washing were carried out at Tm - 20°C = 40°C as given in the methods.

The library was plated out at approximately $1x10^5$ plaques per large plate and two sets of duplicate lifts made. One set was washed to 3 x SSC, 0.1% SDS 40°C and the other to 1 x SSC, 0.1% SDS 40°C. Both sets of filters showed a number of positives, 68 in all and these were taken as agar plugs. The results from the more stringently washed set of filters were cleaner and a dozen of those were taken for further screening, 6 definite strong positives, lines 18, 19, 20, 23, 24 and 25, and lines 1 to 6 which had given signals of various strengths.

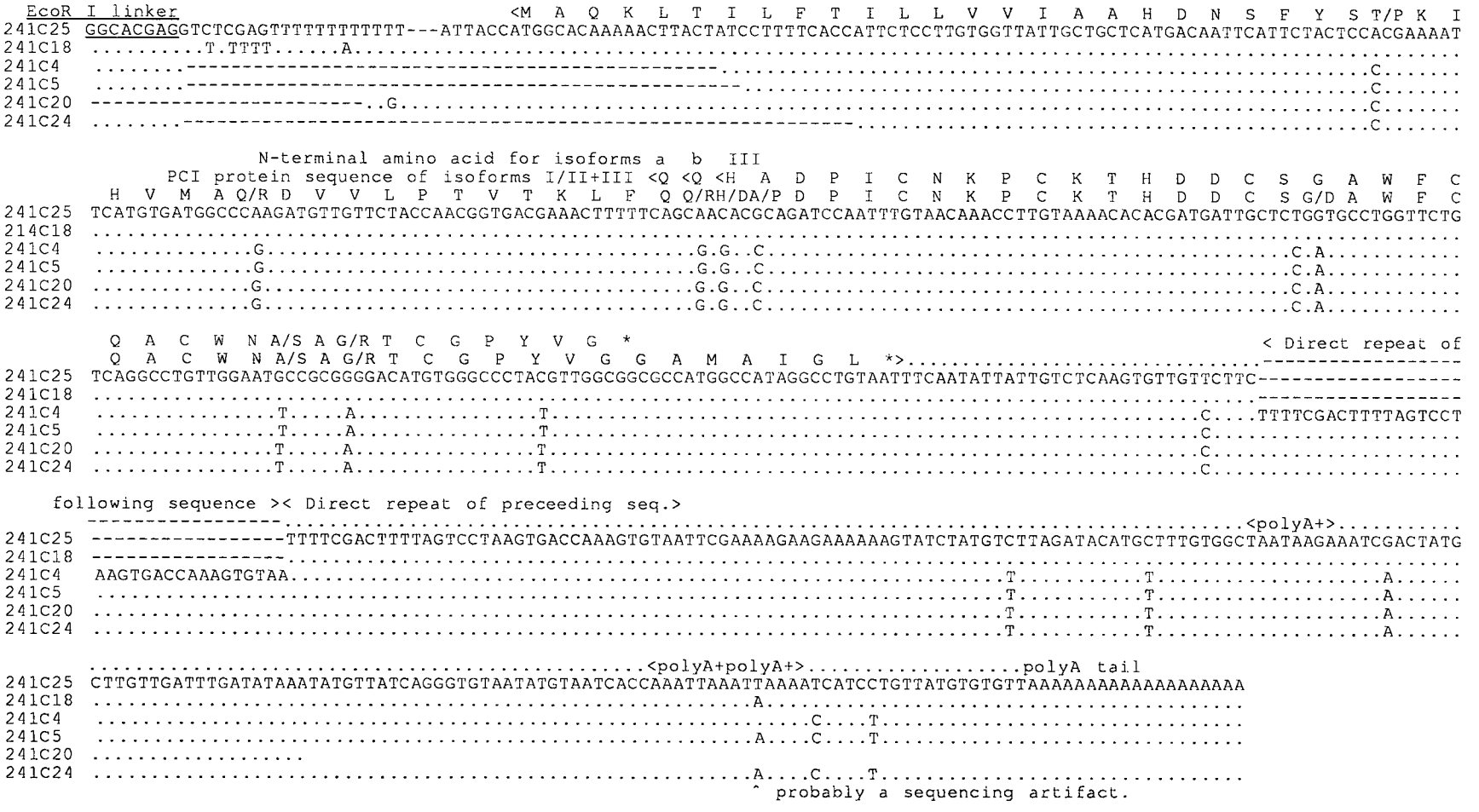


Figure 11. The DNA sequence of the isolated PCI cDNA clones compared to the published protein sequences.

Secondary screening was done on triplicate lifts of plates of appropriate dilutions from each line with two of the lifts being probed and washed as before and the third being probed with oligo #486 at ~20°C (room temperature) and with washing to 3 x SSC, 0.1% SDS ~20°C. These conditions were chosen to match the lower Tm (42°C) of oligo #486 and give approximately the same stringency conditions as for oligo #487. Six well-isolated, positive single plaques were isolated and plasmid rescue performed. The resulting plasmids were DNA sequenced in from both ends using the standard M13 sequencing primers (figure 11) and all were different cDNA clones of PCI.

The cDNAs were of two types. The only sequence differences seen between members of the same type, apart from the extent the sequence extended 5' and the length of the polyA tails, was an A to T change in the putative polyadenylation signal nearest the 3' end which was due to a persistent sequencing artifact and probably not due to a true sequence difference. The polyA tails ranged from 19 to 92 As, and the 5' end of p241C25 appears to have acquired a copy of the 3' cDNA primer (an Xho I site attached to a length of polydT). The most full length examples of each type (p241C20, p241C18 and p241C25) were fully sequenced in both directions with the aid of appropriate subclones.

The cDNA clones p241C18 and 25 encode a protein which I have called PCI I; these exactly corresponds with the isoforms from potato cultivar Russet Burbank, called Ia and Ib, isolated and protein sequenced by Hass, Derr and Makus (1979) when post-translationally processed as indicated. The clones p241C4, 5, 20 and 24 encode a protein which I have called PCI II which, with the appropriate post-translational processing is closely related to Russet Burbank isoforms IIa, IIb and III (ibid). In particular the amino acid differences between type I and II near the carboxy-terminus of the protein, WNAAGTC to WNSARTC, match the protein sequence differences observed between types Ia/b and types IIa/b and III (ibid). However, there are also a number of amino acid differences between my type II and the Russet Burbank type IIa/b and III protein sequences, namely at the N-terminus of the mature protein, QQHADP to QRDPDP, and a G to D change near the middle. These differences are presumably due to the differing potato varieties the sequences were obtained from. The abundances of the cDNA clones within the library suggest that both isoforms should be present in appreciable amounts in tubers.

The two types of cDNA are very homologous at the DNA level with only 17 base differences, i.e.~97% identity, apart from a 35 bp direct repeat in the PCI type II 3'

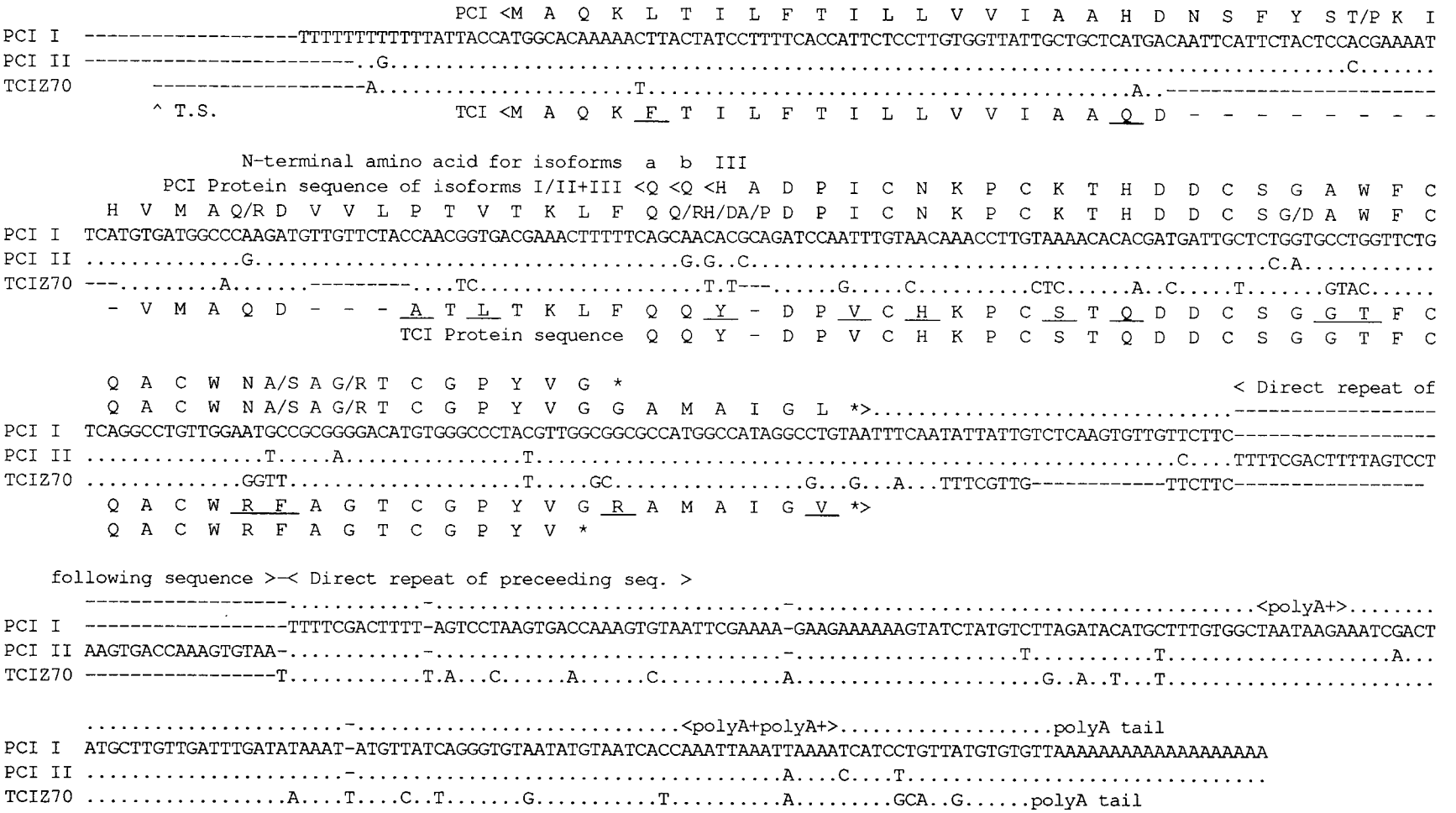


Figure 12. The potato carboxypeptidase inhibitor cDNA sequences compared with that of the carboxypeptidase inhibitor from tomato.

Figure 13. Comparative alignment of the two potato and the tomato carboxypeptidase inhibitor protein sequences.

untranslated sequence. These base differences are distributed as follows; leader sequence; 2, 98.5% identity, mature protein; 8, 93.2% identity, C-terminal extension; 0, 100% identity and 3' untranslated sequence 6, 97.6% identity. There was too little of the 5' untranslated region to make any realistic comparison, but only 1 base difference was observed in the 13 bases of overlap.

Both types have rather long leader peptides of 44 amino acids, as compared to a mature protein of 39 amino acids, if the more 5' of the two upstream in frame ATGs is assumed to be the start codon. This 5' ATG is the better match to the plant consensus start codon (Lütcke *et al.* 1987) and is upstream of the other, out of frame, ATG codons which occur. The predicted leader sequence contains a hydrophobic region typical of signals for transport across the rough endoplasmic reticulum (von Heijne 1983). Both types also have an identical 7 amino acid C-terminal peptide which is also quite hydrophobic and is cleaved during precursor processing to yield the mature peptide which ends GPYV(G). The carboxy-terminal G is slowly cleaved on exposure to carboxypeptidase A but the trimmed inhibitor has essentially the same enzyme binding capability as the original protein (Hass and Ryan 1980) and the hydrolysed glycine can remain trapped in the active site of the enzyme inhibitor complex (Rees and Lipscombe 1982). The 3' untranslated sequences are virtually identical apart from a 35 base direct repeat in the type II cDNAs which is absent in the type Is. There are three putative polyadenylation signals, the second two of which overlap and are probably the functional ones being the consensus 19 bases upstream of the polyA tail. The presence of only two cDNA types fits the speculation of Hass, Derr and Makus (1979) and Hass, Derr, Makus and Ryan (1979) that the observed protein sequences of the five isoforms they isolated could be derived from only two genes with differing post translational processing.

The sequence of a cDNA clone of an homologous carboxypeptidase inhibitor from tomato became available at this time (Martineau, McBride and Houck 1991) and the predicted mature protein amino acid sequence of this exactly matched the tomato carboxypeptidase inhibitor (TCI) protein sequence of Hass and Hermondson (1981).
When this cDNA was compared with the PCI cDNAs here (figures 12 and 13), they were found to be remarkably homologous at the DNA level, being ~90% identical both in the coding region and overall ignoring the relative deletions. The TCI cDNA extends to the more 5' ATG codon and the sequence around this region is identical in all three types of cDNA strengthening the assignment of it as the start codon. The leader sequence is highly

homologous with only two amino acid changes but with two relative deletions which shorten it to 32 amino acids, the N-terminus of the mature protein being at the same point as those potato isoforms which are least processed at the N-terminal end. The coding region shows a number of amino acid changes marked with an underline, and lacks the C-terminal glycine residue of PCI although this may be due to the exposure to carboxypeptidase A (see above). The 3' untranslated region is highly homologous apart from a small relative deletion and a few single base changes and insertions. It lacks the 35 base direct repeat of the type II PCI clones and all clones but one isolated by Martineau, McBride and Houck (1991) used the same more 3' double polyadenylation signal as PCI.

The close homologies between the potato cDNAs and the tomato cDNA reflect the recent divergence of the two species, and the exact conservation of the cysteines and last eight amino acids reflect conservation of structure and function, the "active site" being the C-terminus. The two potato cDNAs are much closer to each other and roughly equally diverged from the tomato cDNA, and presumably arose by a gene duplication event of the single ancestral gene after the potato tomato evolutionary split, there being only a single band on all tracks for a tomato genomic blot probed with the tomato cDNA (Martineau, McBride and Houck 1991). However, the distribution of the sequence differences between the potato and tomato cDNAs, ignoring the relative deletions, is roughly uniform between the coding and non coding sequences. This is unusual as coding sequence usually diverges more slowly than non coding sequence due to the selection pressure of maintaining a reading frame and not disrupting protein function. A similar, though to a lesser extent, uniform distribution of sequence differences between the two types of potato cDNA is also observed. It is not clear how to interpret these observations.

PCI plant expression constructs

It was decided to make plant expression constructs for both types of PCI cDNA in *Agrobacterium tumefaciens* Ti plasmid binary vector pROK2 (provided by M. Bevan and T. Kavanagh (Bevan 1984; Bevan, Mason and Goelet 1985)). This is related to the pROK1 vectors (Baulcombe *et al.* 1986) from which it differs mainly in having a multipurpose cloning site between the strong constitutive CAMV 35S promoter (Guilley *et al.* 1982) and the nopaline synthase gene transcription termination sequence (Bevan, Barnes and Chilton 1983). PCR primers were designed to fit the start and stop codon

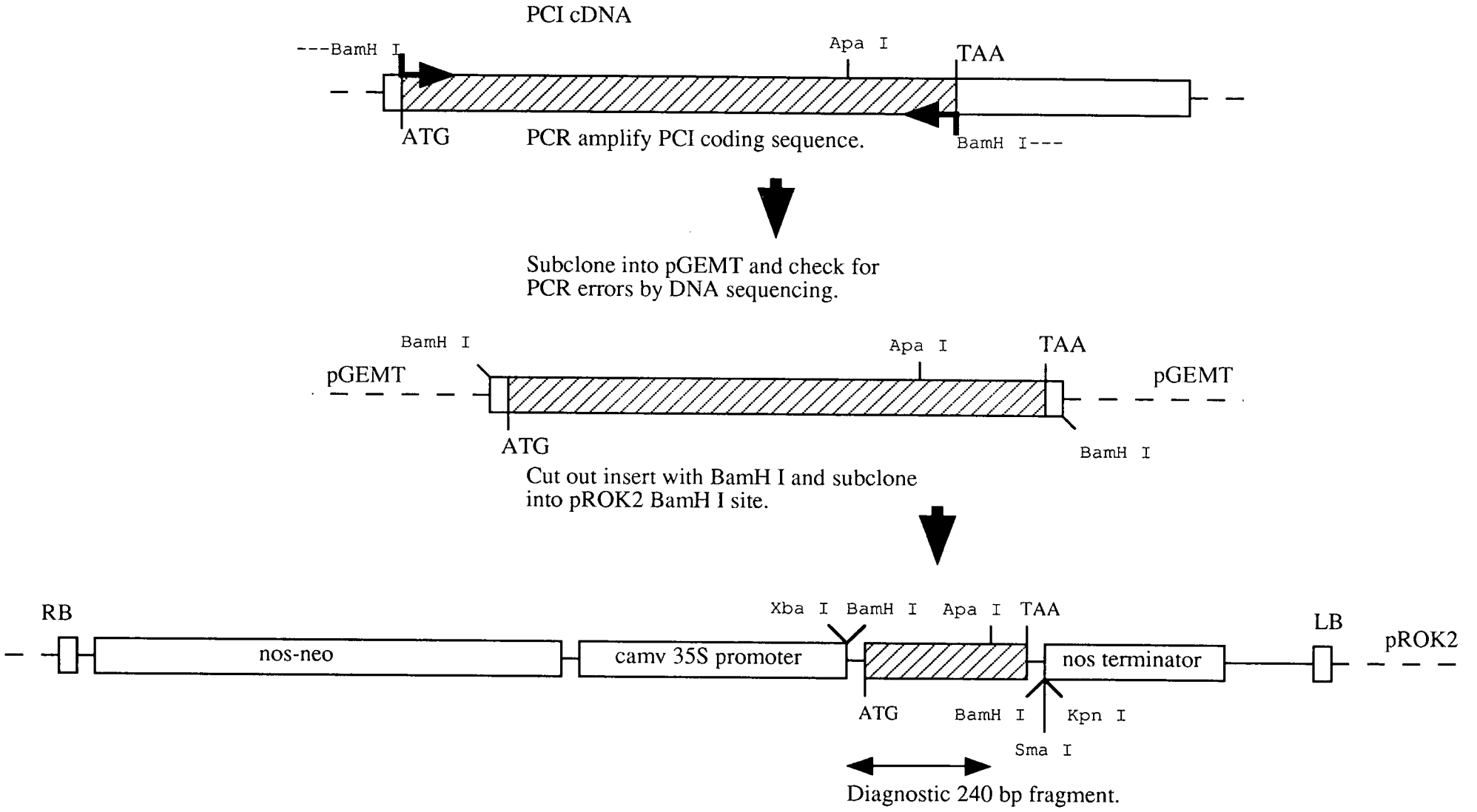


Figure 14. Schematic diagram of the construction of the potato carboxypeptidase inhibitor plant expression constructs.

regions of the PCI cDNAs, which, due to their close homology, could be done with only one pair of primers, and with engineered-in BamH I sites to facilitate cloning into the BamH I site of pROK2 (figure 14).

PCR was performed as usual using p241C18, PCI type I, and p241C20, PCI type II, as templates, an annealing temperature of 50°C, an extension time of 30 seconds and in the presence of Enhance Perfect Match (Stratagene). PCR products were run on a 1% agarose gel and the expected ~300 bp bands were seen, cut out and subcloned by the T_4 polymerase blunt ending method into the Hinc II site of pUC18. Minipreps of the white-growing, ampicillin-resistant colonies resulting from transformation into *E. coli* DH5α were sequenced to check for PCR errors, and two clones were selected: p245C1, a PCI type I, and p245D3, a PCI type II. The inserts in these plasmids were cut out using the engineered BamH I sites and cloned into the BamH I site of pROK2. After transformation into *E. coli* DH5α and selection on kanamycin, colony screening was carried out using the insert of p245C2 as probe and with washing to 2 x SSC, 0.1% SDS, 65°C. Many positives of both types were detected after autoradiography and four good growers of each type were miniprepped. Diagnostic double restrictions were performed, using Xba I, which cuts pROK2 uniquely just 5' of the BamH I site, and Apa I, which cuts the PCI fragments uniquely, generating an ~240 bp Xba I to Apa I fragment for those clones with the insert in the forwards orientation and a ~50 bp fragment for those in the reverse.

pROK2 is a low copy number plasmid and insufficient DNA is obtained from a miniprep to see small bands on gel so the diagnostic restrictions were run on gel, Southern blotted and the blot probed as above. Two of each type were demonstrated to be correctly assembled in the correct orientation and positives 245E19, PCI type I, and 245F161, PCI type II, were transferred to *Agrobacterium tumefaciens* strain LBA4404 by triparental mating using pRK2013 as the mobilising agent. Two single colonies from each class in *Agrobacterium* were picked and total DNA preparations performed. These were restricted with Xba I and Apa I, run on gel, Southern blotted and probed as above. All four were shown to be as expected, with one large fragment and one diagnostic ~240 bp fragment hybridising to the probe. Colonies 245G1, PCI type I, and 245H1, PCI type II, were preserved and carried forward to the plant transformation.

<u>The transformation of tobacco with the PCI expression constructs</u>

Leaf discs were cut from axenicly grown tobacco plants (*Nicotiana tabacum* cultivar Samsun) and were grown on for two days on sealed solid agarose plates at 24-26°C and light intensity 2000-4000 lux. They were then infected with the *Agrobacterium* containing the constructs described above and grown for a further 3 days, by which time visible callus and bacterial colonies had formed. Those leaf discs with callus were transferred to selection on kanamycin at 100 μg/ml. After about 2 weeks shoots had appeared and these were cut from the callus and transferred to rooting medium and grown on till they were a reasonable size (about 5 cm) and had a good root system. These plants were then transferred onto a peat based compost (Fisons Levington) and grown on as described in the methods.

Seven putative transgenic plants were recovered; five (G1 to 5) from the PCI type I construct, and two (H1 and 2) from the type II. All of the plants appeared morphologically normal and all but two produced seed when allowed to self pollinate, plant G1 flowered but produced empty seed pods and plant H2 could not be induced to flower.

Assaying the transgenic plants for carboxypeptidase inhibitory activity

The putative transgenics and three seed grown control plants, were screened for carboxypeptidase A inhibition by assaying crude total soluble protein extracts. These were made by taking the leaf nearest the top of the plant which was about 1 inch long, freeze drying it and extracting a weighed amount of dry leaf in 50 mM tris/HCl pH 7.5.

The first set of assays, done by the non-preincubation method but with the plant extract preincubated with the substrate, showed that all of the transgenics showed some inhibition relative to the three control plant leaf extracts and that tobacco leaf extract has no enzyme activity that can hydrolyse FAPP. As all the transgenics gave strongly curving traces a second set of assays were performed by the preincubation method with the amount of leaf extract standardised to 200 μg dry weight of leaf, which is approximately 28 μg of soluble leaf protein (Estimating leaf protein extract concentrations by this method gives more reproducible results than the Bradford assay. The approximation of 14% soluble protein in dried young leaves is rough but is fairly reproducible within species. Personal observations and A.M.R.Gatehouse personal communication). Again all the traces were

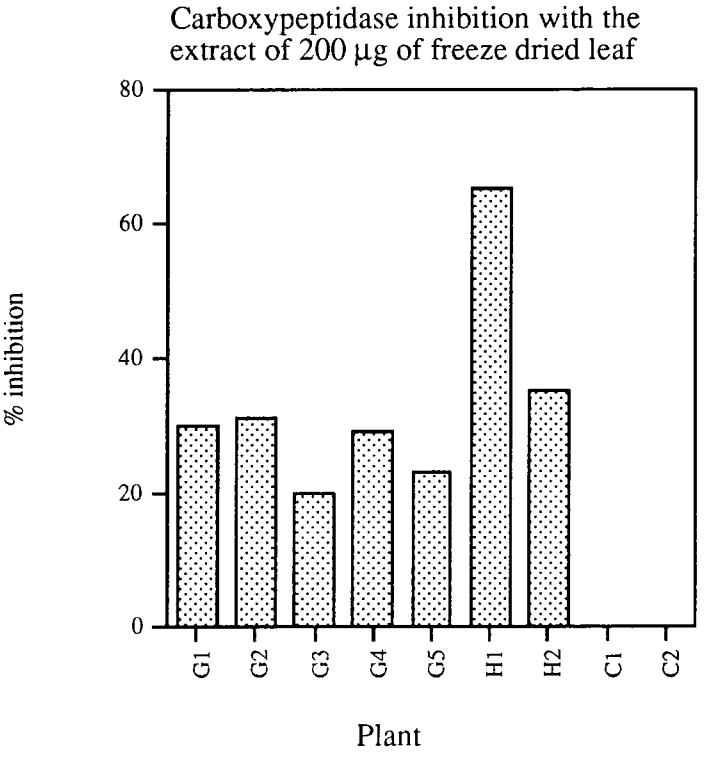


Figure 15. Inhibition of bovine carboxypeptidase A by the extract of 200 µg of freeze dried leaf taken from the primary transformants G1-5 and H1-2 and two controls C1-2. Note that the controls show no inhibition and that plant H1 is the best expresser.

noticeably curved and it could be seen that the carboxypeptidase A dilution was slowly losing activity as it was stored on ice between the assays. The inhibition percentages have been corrected for this effect. However it could be seen that all of the transgenics showed inhibitory activity, the best being plant H1 which showed approximately 65% inhibition (figure 15). For the amount of carboxypeptidase A used in this assay, 14 ng of PCI would be required give 50% inhibition, and as each assay contains approximately 28 µg of soluble leaf protein, the best expresser, plant H1, must be expressing more than 14 ng of PCI in 28 µg of soluble leaf protein, i.e. has an expression level of PCI in the range 0.05 - 0.1%.

Bioassay of PCI expressing tobacco against *Heliothus virescens*

Eight non-transformed control tobacco plants labelled C1 to 8 were grown from seed (*Nicotiana tabacum* cultivar Samsun, the line used in the transformation experiment.) and 24 experimental plants labelled E1 to 24 were grown from seed from self crossed primary transformant H1, which had been shown to be the best PCI expresser. The growth room used was set for 12 hours day i.e. full lights, 4 hours half lights and 8 hours dark at 25°C during the day and 21°C during the night with a relative humidity of 65%. When the plants were approximately 30 cm high, i.e. fairly mature but before the flower buds start forming under these growth conditions, each plant was placed into a perforated plastic bread bag which was sealed to the flowerpot but left open at the top. This system provides good insect containment without condensation and hence fungal problems. Samples of young, approximately 1 inch, leaves were taken from each plant and freeze dried for later inhibition assays. The trial was initiated by inoculating each plant with 8 neonate *H. virescens* larvae, less than 24 hours old, and sealing the top of the containment bag.

The trial was terminated after 20 days by collecting and weighing the surviving larvae and by scoring the plants for leaf damage by eye on a 1 = undamaged to 5 = heavily damaged scale. Such a visual inspection system has been shown, for a wide range of "estimators", to yield results which do not differ significantly from those obtained by quantitative image analysis of damage caused by caterpillars on tobacco plants (V. A. Hilder personal communication). The leaf samples were extracted and assayed as above for carboxypeptidase A inhibitory activity. Plants could be easily assigned to one of three

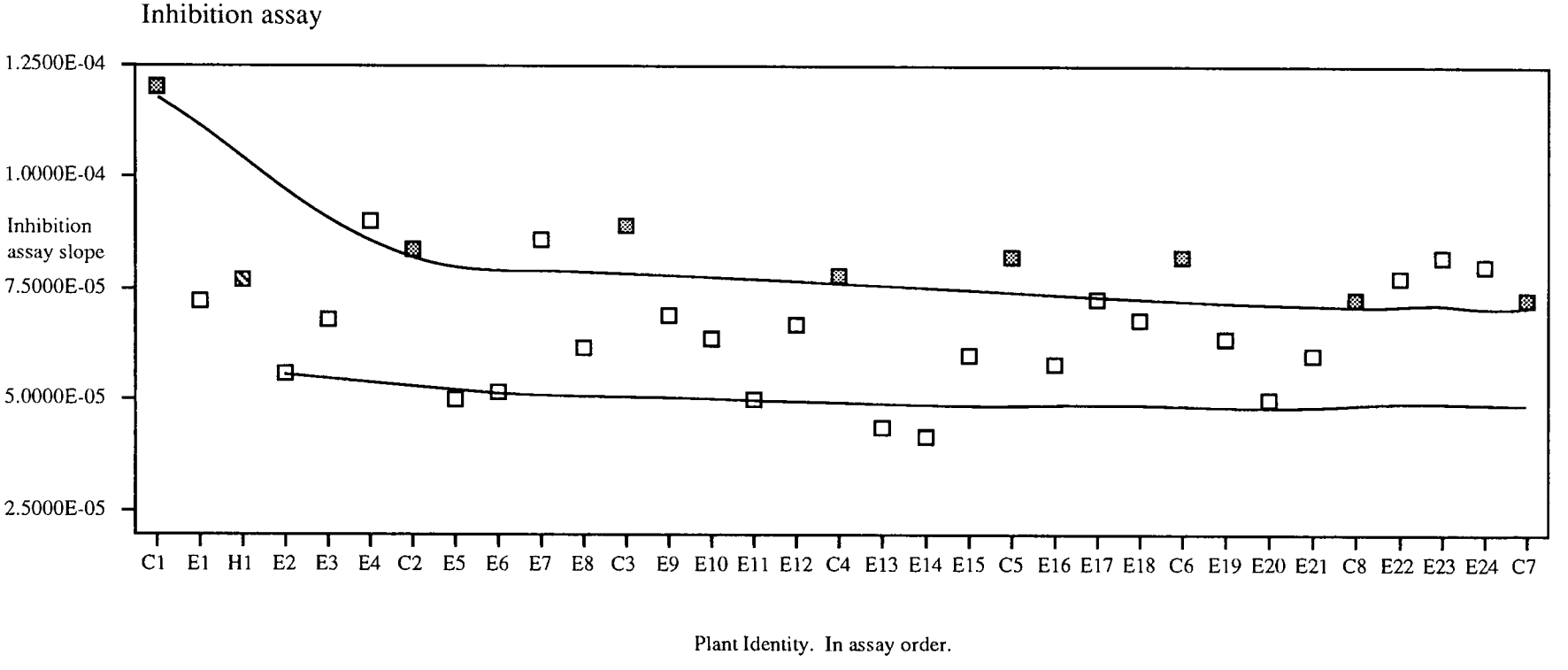
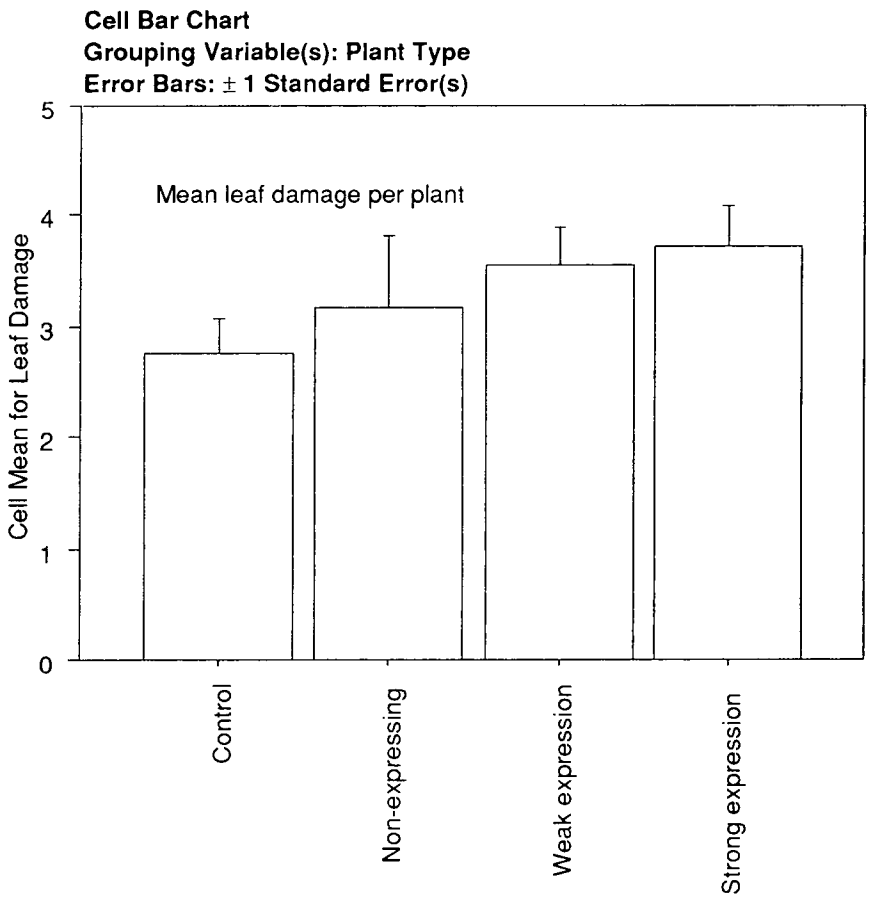


Figure 16. Scattergram of the results of the carboxypeptidase inhibition assays performed with leaf extracts from the S1 and control plants from the insect bioassay. Dark squares are the control plants and the striped square is the parent plant H1 which was not included in the bioassay.



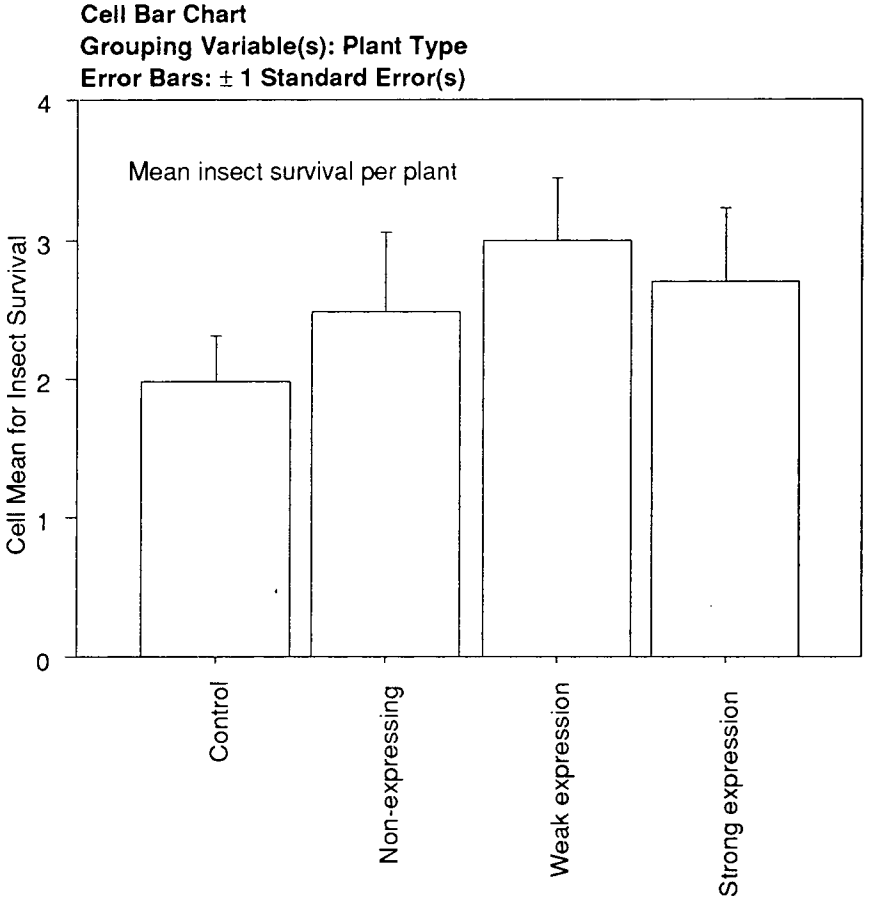
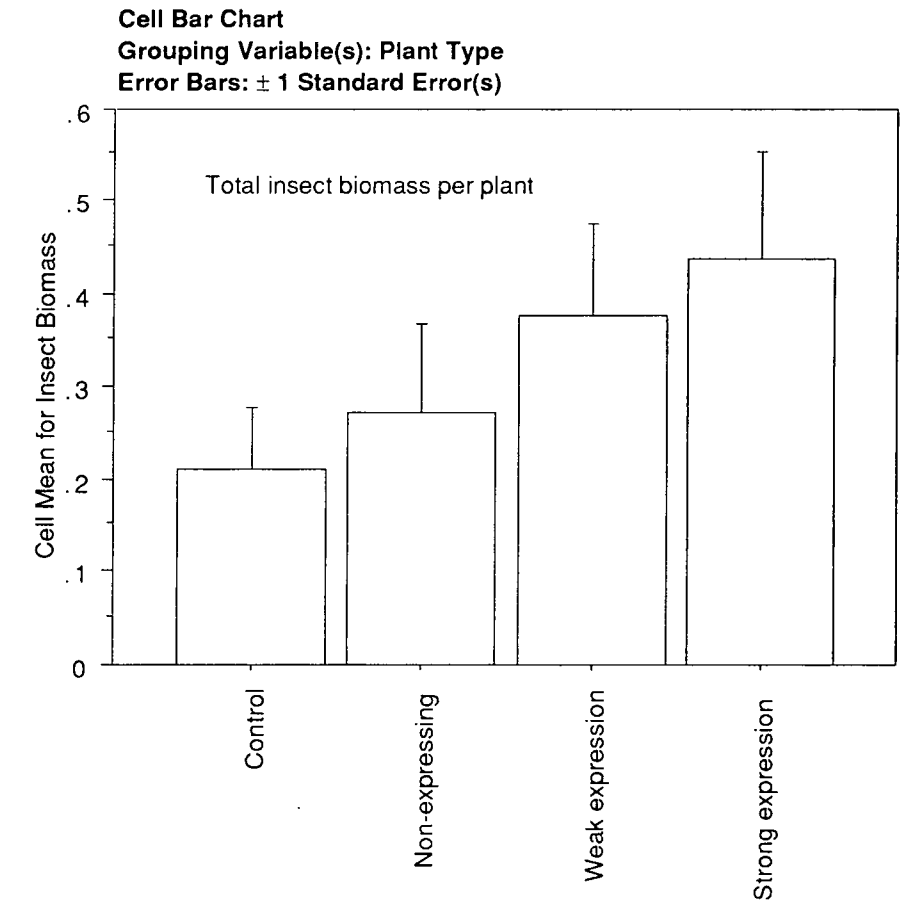


Figure 17 PCI tobacco insect bioassay.



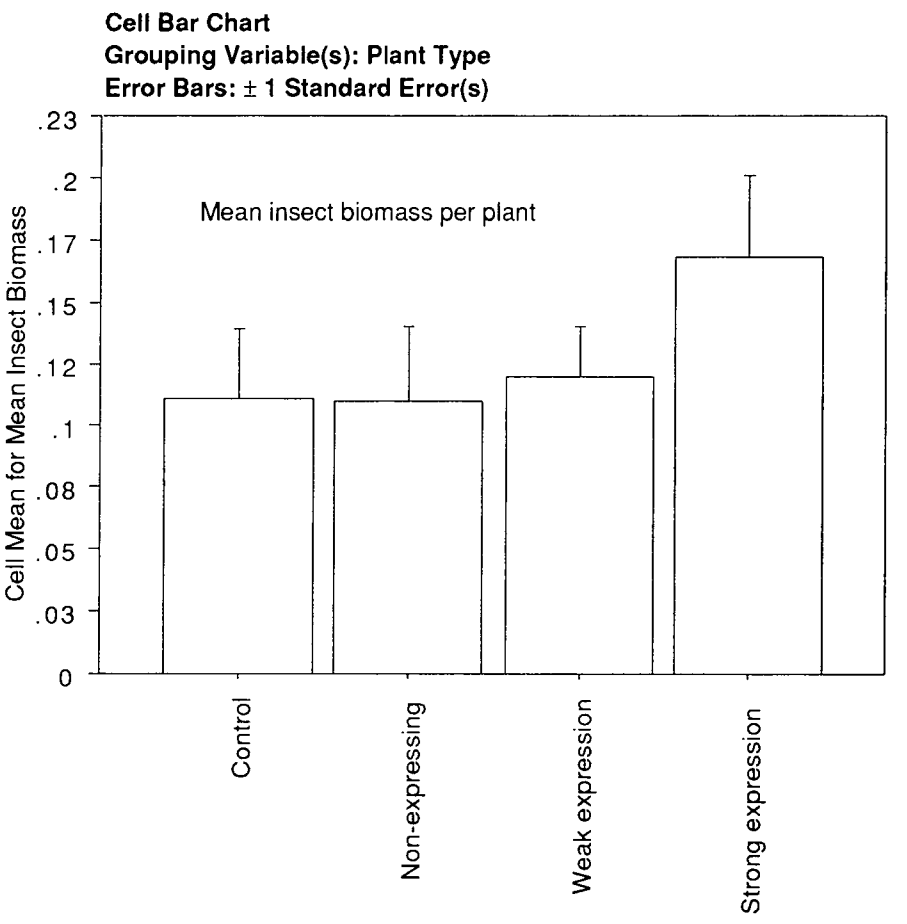
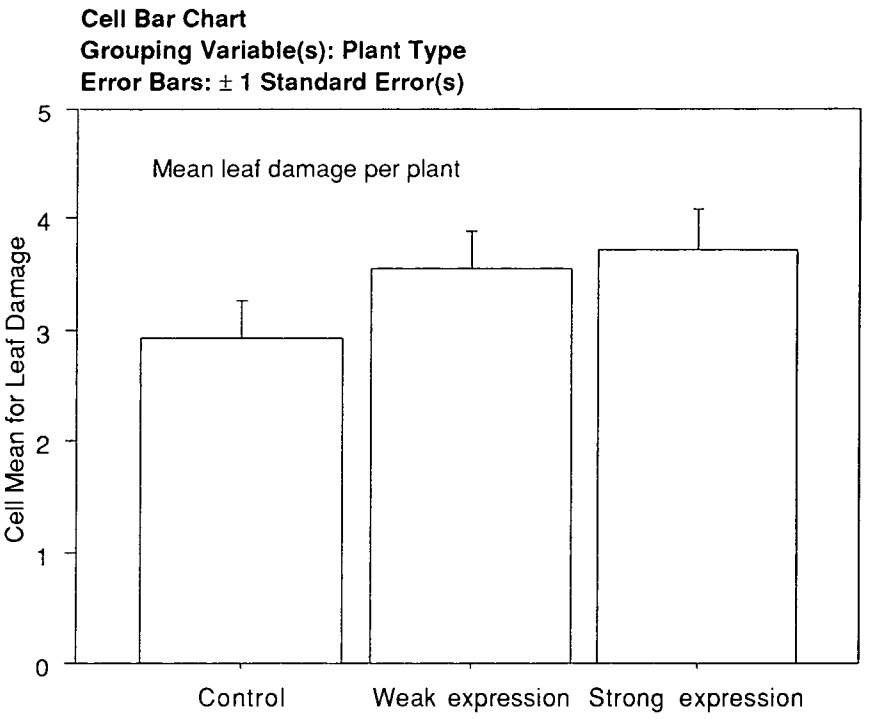


Figure 18 PCI tobacco insect bioassay.



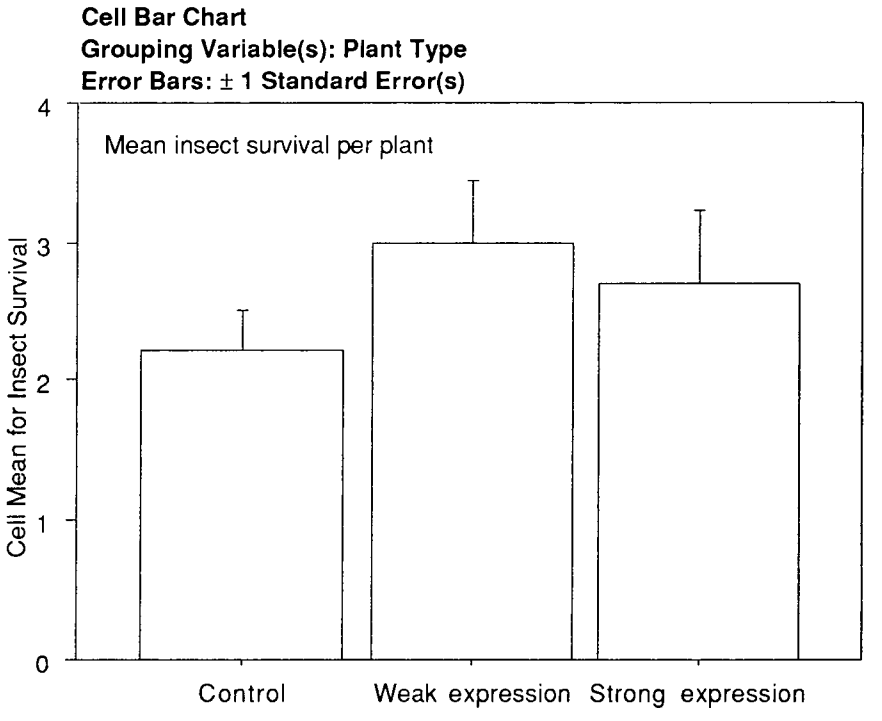
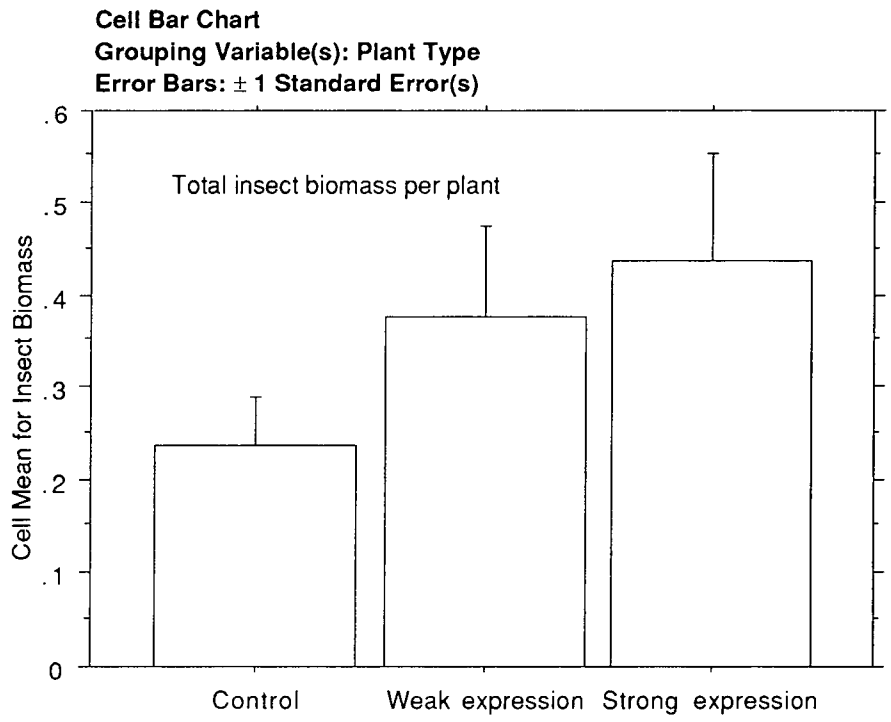


Figure 19 PCI tobacco insect bioassay.



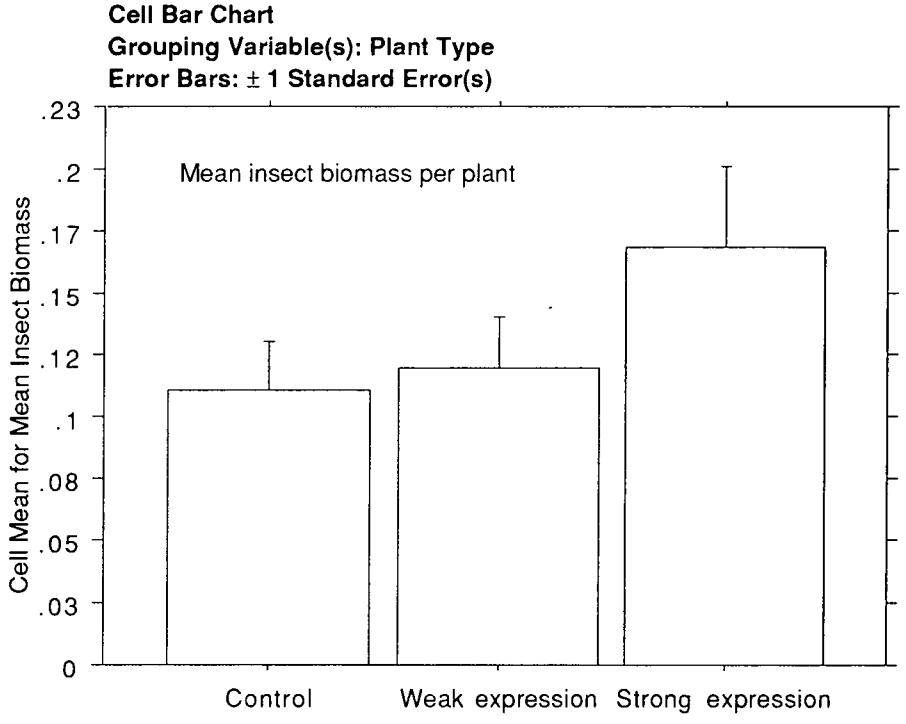


Figure 20 PCI tobacco insect bioassay.

categories corresponding to non expresser, medium expresser and high expresser on the basis of the slope of the OD versus time plot in the inhibition assay (figure 16).

All the non-transformed controls were non-expressers by this assay. The transgenic progeny plants gave six non expressers, eleven medium expressers and seven high expressers. Non-expressing transgenic progeny showed no significant difference to the non-transformed controls. The segregation ratio corresponds well to that expected for a single Mendelian locus (H_0: $\neq$ 1:2:1, χ^2 = 0.250, p >> 0.05). Plant E22 a non expresser was stunted and plant C2 flowered during the trial. The data was collated and analysed using the StatView statistical package (Abacus Concepts) (figures 17 and 18).

Segregation analysis of PCI expression levels indicated that the insertion of the construct had occurred in a single site and, therefore, that the medium expressers corresponded to hemizygous progeny and the strong expressers to homozygous progeny. The latter expressed PCI at approximately twice the level of both the former and of the hemizygous parent H1. The controls and the experimental non expressers were not significantly different in any of the measures of insect resistance applied to the data. They may therefore, be treated as a single control group and the data was analysed again with this correction (figures 19 and 20).

Transgenic progeny which did express PCI were consistently, though not always significantly, more susceptible to the *H. virescens* larvae than the controls. This result was significant for the mean total insect biomass per plant test and significant or nearly so for the mean leaf damage per plant test. There was also evidence that for medium expressers the mean insect survival increased and for strong expressers the mean insect biomass increased i.e. the survivors were larger.

Discussion

The increase in susceptibility to *H. virescens* larvae of the PCI expressing transgenics was surprising and is difficult to explain. Three possibilities may be considered.

i). That PCI at these levels is positively beneficial to the insects. It seems highly unlikely that insects would benefit from inhibition of one of their major digestive enzymes. PCI might simply be being used as a protein source, but the level of expression estimated at 0.1 - 0.2% of total soluble leaf protein would seem too low to have any effect on insect

growth or survival. It is still not even known whether expression of a foreign protein in transgenic plants is additional protein synthesis of merely substitution.

It is conceivable that PCI might be beneficial by improving the amino acid balance of the insects diet. Insect growth has been shown to be limited in some cases by shortage of specific amino acids (Murdock *et al.* 1990), often sulphur containing amino acids. PCI mature protein has 6 cysteine residues in a total of 39 amino acid residues, ~15%. This is quite sulphur rich compared to normal plant protein which contains well under 5% cysteine. If it were practicable to purify PCI on a sufficiently large scale then artificial diet bioassays could be performed to clarify these points.

ii). That PCI expression at these levels "weakens" the plants defences. PCI might inhibit some enzyme in planta and hence defence-related factors are not properly processed or regulated. However, there is no detectable carboxypeptidase A activity in the tobacco leaf soluble protein extracts, though the affected activity may be membrane bound. Hass *et al.* (1981) found that none of the carboxypeptidase enzymes of plant origin they tested were inhibited by 10 μM PCI protein including extracts of tomato fruit, the source of a closely homologous carboxypeptidase A inhibitor. PCI is found throughout the potato plant in varying amounts (Hass, Derr and Makus 1979) and is induced by wounding (Graham and Ryan 1981 and Höllander-Czytko, Andersen and Ryan 1985) which argues against it having a regulatory role. Tobacco, potato and tomato are all fairly closely related and PCI would not be expected to affect tobacco when expressed especially as it should be localised in the vacuole (Höllander-Czytko, Andersen and Ryan 1985).

iii). That insertion of the gene construct had weakened the plants defences. In this case the expression of PCI is largely irrelevant. The PCI gene may be inserted into some important defence related gene which was thereby disrupted. The even worse performance of the high expressers compared with the medium expressers would then be due to inactivation of both as opposed to only one copy of the native gene: i.e. correlation not with PCI level of expression but with the number of loci. Consistent with this hypothesis is the observation that in insect bioassays of the original cowpea trypsin inhibitor-transgenic tobacco a line of plants containing the antisense construct (intended to be the most appropriate control) performed consistently (though never significantly) worse than the non-transformed controls (V. Hilder personal communication).

Further experiments; bioassay of some different selfed transgenic lines to disentangle any gene insertion position effects, and fresh plant transformations to obtain

higher expressers (protease inhibitors seem to be effective in the 1 - 2% range in transgenic plants) would certainly clarify the situation.

The lack of early success in insect control led to further work on PCI being abandoned for commercial reasons. However, this work had important implications for the conduct of insect bioassays on transgenic plants. It highlighted the inadequacy of depending on a single transgenic line to evaluate a gene. Subsequent screening methodology was modified so that a number of lines, not just the highest expresser, were carried through to insect bioassay.

Results; insect haemolymph trypsin inhibitor (HTI)

The partial protein sequences of two trypsin inhibitor proteins from the haemolymph of *Manduca sexta* have been published (Ramesh, Sugumaran and Mole 1988) and were used to search the national protein database at Daresbury using the gcg package of programs (gcg 1994) a large number of other related protein sequences were found mainly from mammals but including some insect proteins; three chymotrypsin inhibitors from silkworm (*Bombyx mori*) (Sasaki 1978, Eguchi, Haneda and Iwamoto 1982, Sasaki and Kobayashi 1984), a hypothetical acrosin inhibitor from *Drosophila* (Schmidt *et al.* 1989) and a protein inhibitor from *Sarcophaga bullata* (flesh fly) (Papayannopoulos and Biemann 1992). The most homologous protein from other orders was bovine coloestrum trypsin inhibitor (Cechova and Muszynska 1970, Cechova, Jonakova and Sorm 1971, Cechova and Ber 1974) which, like the insect proteins, consists of only a single domain.

The sequences were aligned (figure 21) and a number of regions of conserved protein sequence were found, particularly around the cystein residues. The only region sufficiently conserved and long enough to make a reasonable oligo to was the strongly conserved sequence YGGCQGN. This is present in all the sequences with the only sequence differences occurring at the glutamine (Q) residue. The published partial protein sequence of the *Manduca* A inhibitor (MHTI A) does not extend the full length of this conserved region but the variable amino acid residue is confirmed as glutamine.

Oligo design for the HTI conserved region and for PCR

An oligo (#575) was designed to the conserved motif YGGCQGN (figure 22) which was an 8 fold redundant 20mer with 3 inosines and was made in antisense for library screening and hybridising to Northern blots. This is effectively a 17mer for the purposes of hybridisation selectivity which is on the short side for screening a cDNA library in bacteriophage. The oligo melting point was calculated by the 4(G+C)+2(A+T+I) method as 56°C (Sambrook, Fritsch and Maniatis 1989).

A second oligo (#736) was designed to the protein sequence YKPPNNI near the 5'end of the MHTI A partial protein sequence and made in sense for PCR experiments with the conserved region oligo #575. This sequence was chosen to give a primer of minimum

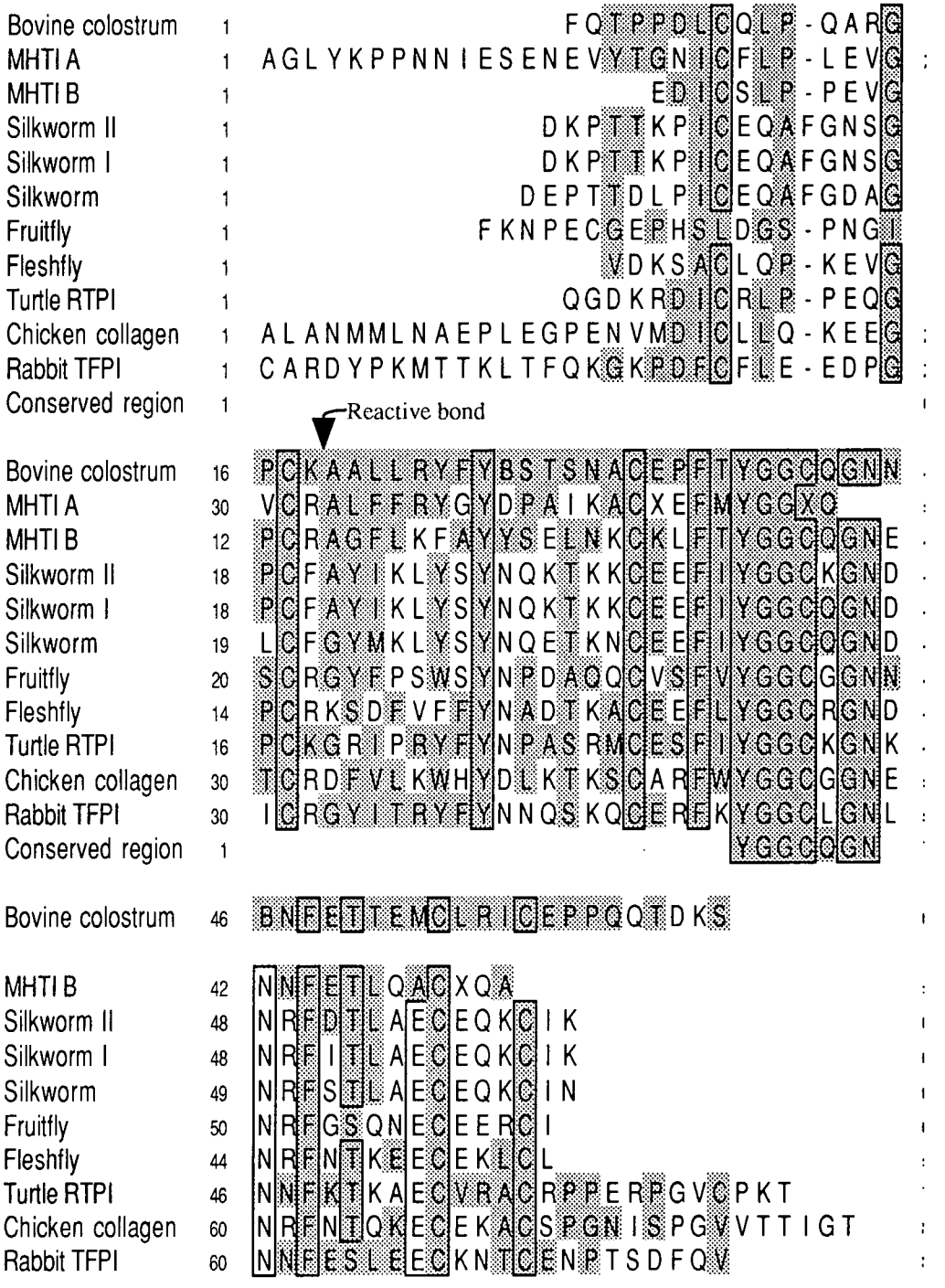


Figure 21. The alignment of the *Manduca* haemolymph trypsin inhibitor protein sequences with Bovine colostrum trypsin inhibitor and some homologous proteins and protein domains from other sources.

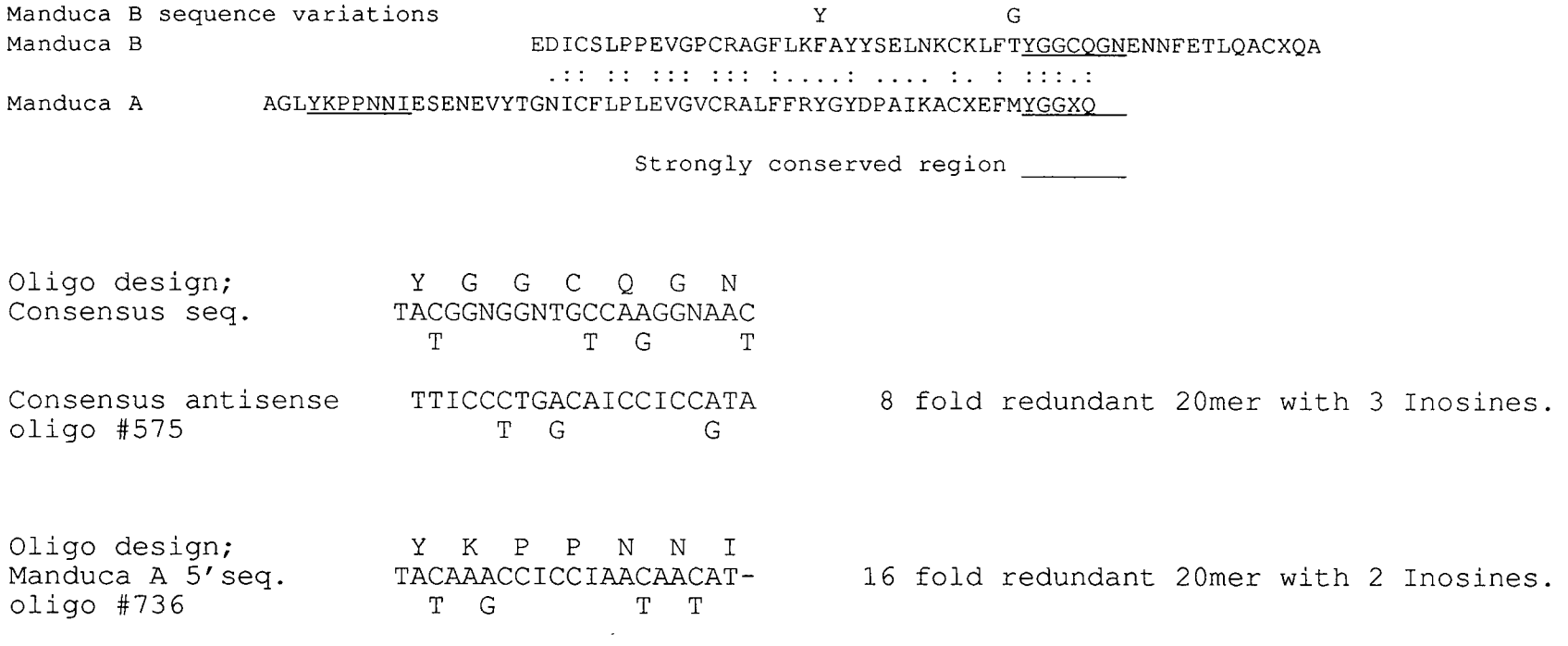


Figure 22. The design of the oligo for the *Manduca* haemolymph trypsin inhibitors strongly conserved region #575, and the MHTI A specific oligo #736.

redundancy purely for the amplification of the MHTI A sequence. It had a melting point of 48°C by the above method.

Preparation of *Manduca* RNA

Manduca sexta larvae were grown on tobacco plants until approximately 2 cm long when they were harvested into liquid nitrogen, those larvae not used immediately were stored at -80°C. RNA was prepared from total *Manduca sexta* larvae by the guanidinium thiocyanate method for tissue (Ausubel *et al.* 1992), 4 gm of larvae yielding 8.5 mg of total RNA which was fairly pure with an OD_{260} to OD_{280} ratio of 1.8. From this polyA+ RNA was prepared by the polyAttract method (Promega) 1 mg of total RNA yielding 8.8 µg of polyA+ RNA. Three samples of 0.88 µg of the polyA+ RNA was run on formamide gel and one track and the size markers stained with ethidium bromide while the rest were blotted. The stained track showed no visible bands, as would be expected for this amount of cellular RNA which had been purified free of ribosomal RNA. The quality of the polyA+ RNA was assessed when making the cDNA library (see below) and by probing one of the polyA+ blots with the consensus oligo #575. The chosen conditions were Tm - 20°C = 36°C with washing to 5 x SSC, 0.1% SDS. Weak hybridisation in the size range 1000 to 500 bases was observed with no discernible bands; 500 bases is roughly the message size that a mature protein of about 75 amino acids could be expected to have. This smear suggested that the RNA may be degraded, an alternate explanation was that the hybridisation and/or washing conditions were insufficiently stringent and that the smear represents non specific hybridisation and masks the true band. The weakness of the hybridisation suggested that the oligo screening of the cDNA library would be difficult or that the message was rare.

Preparation of a *Manduca sexta* cDNA library

The *Manduca* cDNA library was made in λZAP II using a kit (Stratagene ZAP-cDNA synthesis kit) and according to the directions in the kit manual. The first and second strand synthesises were performed in duplicate with the small scale duplicate reactions incorporating a radiolabel and these were run on gel. The cDNA was observed to have a good size range, from less than 200 to over 2000 bases and hence that the polyA+

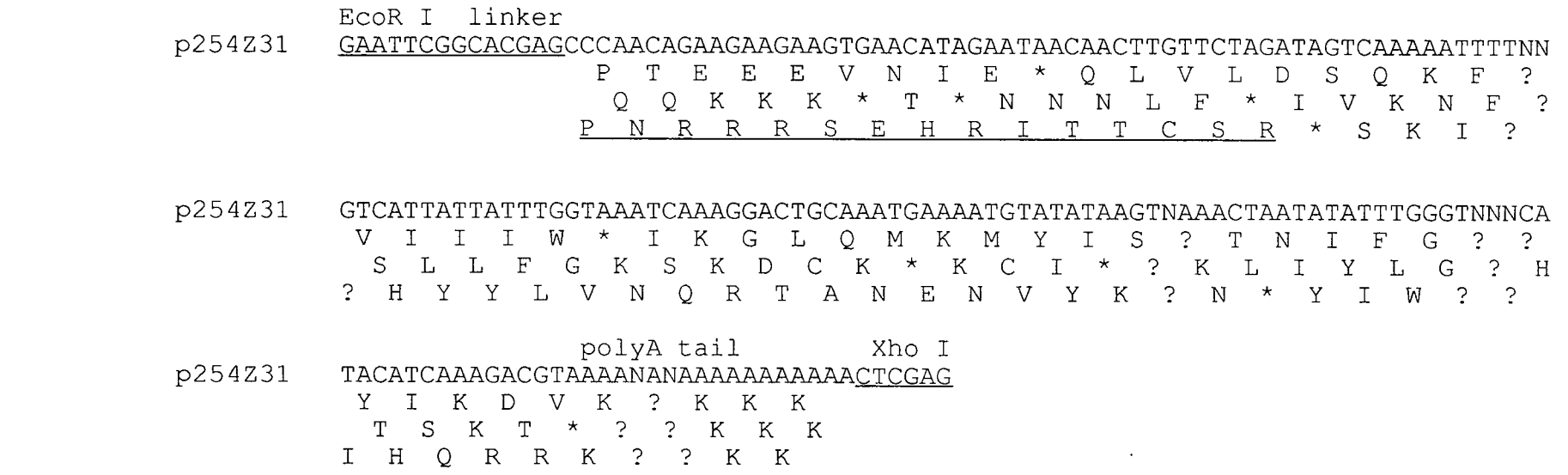


Figure 23. The DNA sequence of p254Z31 and the translation in all forwards frames. The under lined protein sequence was the ORF used for the fasta search of the protein sequence databank.

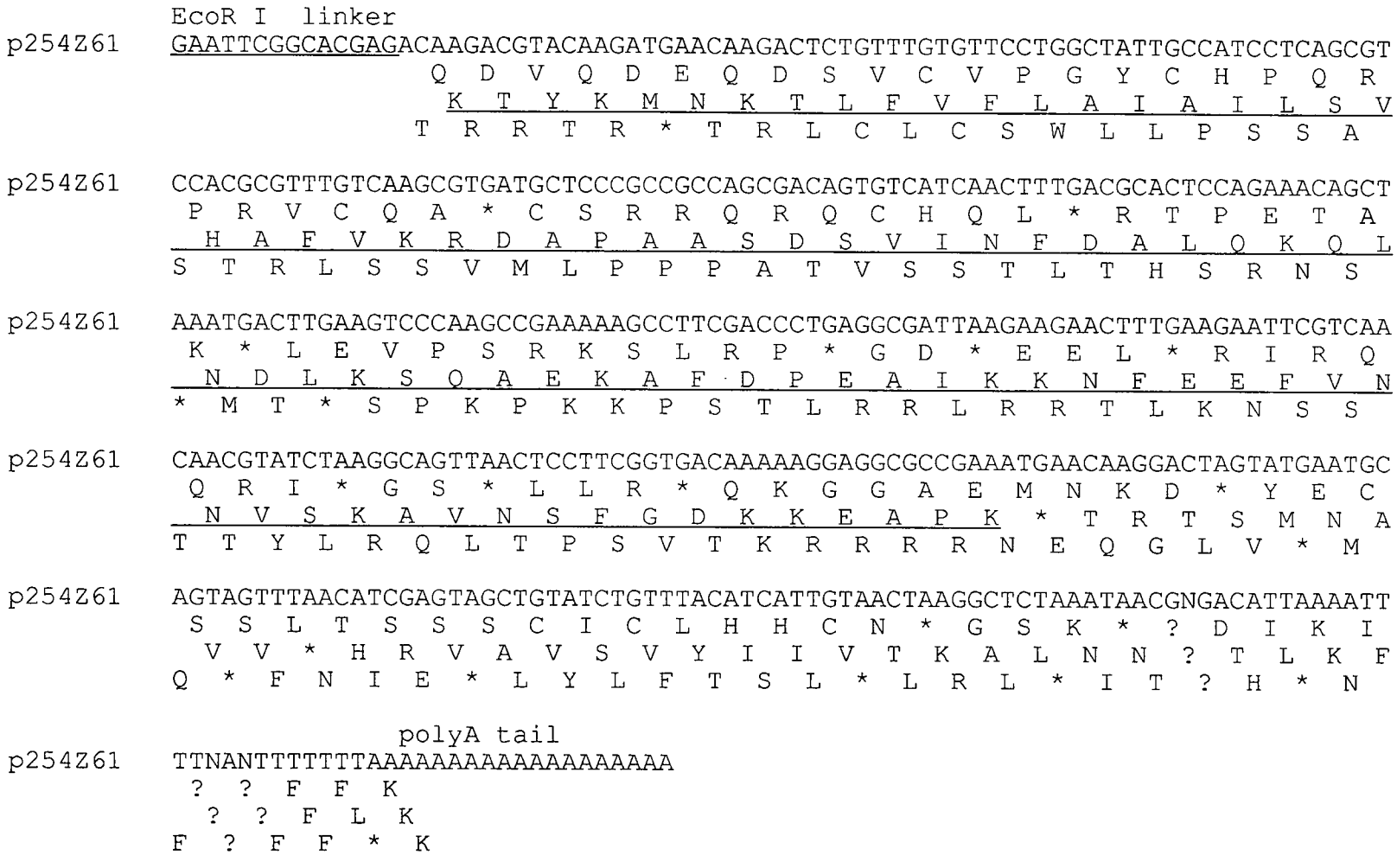


Figure 24. The DNA sequence of p254Z61 and the translation in all forwards frames. The underlined protein sequence is the long ORF.

RNA contained a reasonable proportion of high molecular weight, probably undegraded, messenger molecules. Upon quantitation of the double strand DNA by the spotting method approximately 200 ng had been synthesised and 50 ng of this was ligated into the phage λ vector arms and packaged using the Gigapack II system. A sample of the library, 0.2%, was plated out and no blue plaques were seen on X-Gal and IPTG, i.e. no clones without an insert. The library consisted of 3 x 10^6 independent clones and was split into two parts. Half was stored unamplified by the DMSO method and the other half was amplified to a titre of 4 x 10^9 pfu per ml, and stored in aliquots by both the methods recorded in the methods section and distributed around various fridges and freezers.

The quality of the library was further assessed by taking 4 single plaques at random and performing plasmid rescue on them. They had inserts of 180, 400, 1400 and 2100 bp which confirmed that the library contained some long cDNA clones. All four clones were partially DNA sequenced using the M13 reverse primer, i.e. in from the 5' (EcoR I) end.

The 180 bp cDNA (p254Z31) sequenced completely through and both cloning linker primers were present undamaged along with an 18 bp polyA tail. The only ORF of any length was 15 amino acids long and this was used to run a fasta search of the protein data bank at Daresbury but no matches were found. Thus it must be the clone of an incomplete cDNA covering the 3' end. (figure 23)

The 400 bp cDNA (p254Z61), whose insert failed to cut out with EcoR I and Xho I, also sequenced completely through to an approximately 19 bp polyA tail. The 3' linker primer did not sequence sufficiently well to confirm that its Xho I site was damaged but the 5' linker and its EcoR I site was undamaged. There was an 87 amino acid open reading frame (ORF) running in from the 5' end including a methionine codon near the 5' end and a shorter ORF of 25 amino acids also running in from the 5' end. Both of these were used to run a fasta search of the protein database at Daresbury but no significant matches were found. (figure 24)

The 1.4 kb cDNA (p254Z32) had an undamaged 5' linker and single ORF extending in from the 5' end which could be followed for 83 amino acids before the DNA sequence became too unreliable. This was used to run a fasta search of the protein database at Daresbury and a few low-significance matches were detected, however these were all proline rich proteins with the match to the proline repeat regions and so were discounted. (figure 25)

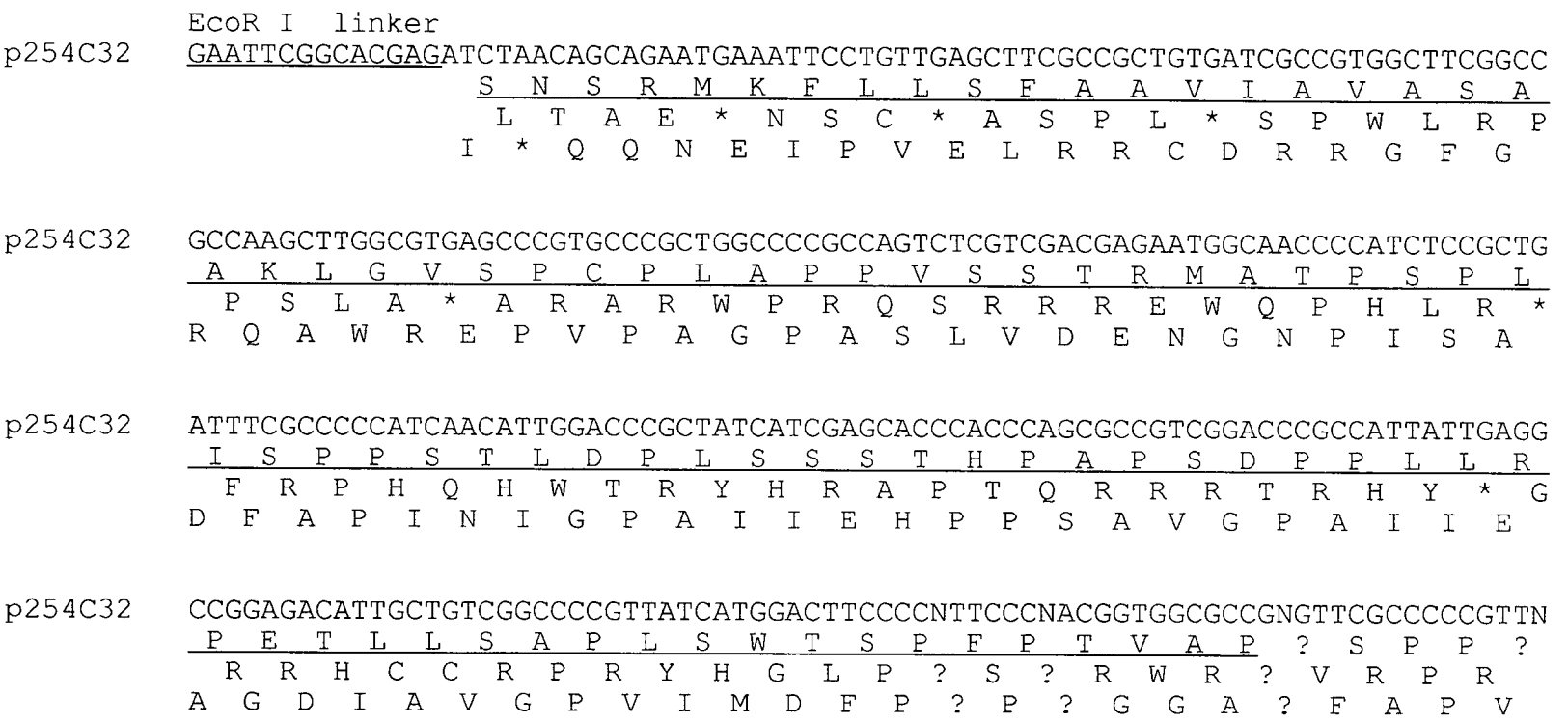


Figure 25. The partial DNA sequence of p254Z32 and the translation in all forwards frames. The long ORF is underlined.

```
  EcoR I  linker
p254C71   GAATTCGGCACGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTNTNNCGNANTCGGNNGNG
```

Figure 26. The TC artifact sequence of p254Z71, the sequence beyond this was not readable.

The 2.1 kb cDNA (p254Z71) had an undamaged 5' linker and then went into a 35 bp TCTC repeat sequence beyond which the DNA sequence was unreadable (figure 26). Artifacts of this type at the 5' end of a clone are not unknown in λZAP II libraries (Dr. N. J. Robinson personal communication), and no further work was done on this or any of the other randomly selected cDNA clones.

Screening the *Manduca* cDNA library for the HTI by oligo

An initial attempt to screen the library was made using the conserved sequence oligo #575 using the same conditions as for the Northern blot (above), namely 36°C with washing to 5 x SSC, 0.1% SDS. This left a high level of background counts uniformly over the filters and so washing was continued to 3 x SSC, 0.1% SDS. Upon autoradiography this gave no positives and a low but uniform background suggesting either over washing or too high a hybridisation temperature. The library was next screened with a mixed probe of both oligos #575 and #736 using the revised conditions of 30°C and with washing to 3 x SSC, 0.1% SDS. This gave 7 possible positives but after extensive secondary screening, these were found to be artifacts.

Synthesis of a fragment of MHTI A by PCR

It was decided that a better probe was needed and to this end a PCR experiment was set up to amplify an 164 bp fragment from the MHTI A sequence. The conserved region reverse oligo #575 and the MHTI A specific forward oligo #736 were used as primers, and a sample of the *Manduca* dscDNA produced as part of making the cDNA library as template DNA. The conditions chosen were an annealing temperature of 50°C, which matches the calculated oligo melting points, an extension time of 20 seconds, as the expected fragment should be 164 bp long, and one reaction with and one without Enhance Perfect Match (Stratagene). When run on gel the reaction without the Enhance PM gave a band of the expected size while the reaction with Enhance PM gave nothing. In order to resolve whether an artifact was produced in the absence of Enhance PM or the reaction in its presence had failed due to too high an annealing temperature the fragment was cut out of the gel, the DNA isolated, cloned into pGEMT (Promega), transformed into *E. coli* DH5α, miniprepped (as p256A3) and DNA sequenced. This showed that the PCR

Manduca inhibitor A partial protein sequence and PCR product p256A3 DNA sequence

```
MHTI A   A  G  L  Y  K  P  P  N  N  I  E  S  E  N  E  V  Y  T  G  N  I  C  F  L  P  L  E  V  G  V  C
                   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
                   Y  K  P  P  N  N  I  E  S  E  N  E  V  Y  T  G  N  I  C  F  L  P  L  E  V  G  V  C
p256A3             TACAAGCCGCCGAACAACATAGAATCTGAGAACGAAGTTTACACCGGAAATATTTGCTTCTTGCCATTGGAAGTTGGGGTATGC
5' Oligo #736      TACAAACCICCIAACAACAT
                     T  G        T  T

Mand A   R  A  L  F  F  R  Y  G  Y  D  P  A  I  K  A  C  X  E  F  M  Y  G  G  X  Q
         :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  .  :  :  :  :  :  :  .  :
         R  A  L  F  F  R  Y  G  Y  D  P  A  I  K  A  C  K  E  F  M  Y  G  G  C  Q  G
p256A3   CGAGCTCTGTTCTTTAGGTACGGATACGATCCAGCGATAAAGGCATGCAAGGAATTCATGTACGGCGGCTGCCAAGGCAA
                                                                    3' Oligo #575 TACGGIGGITGCCAAGGIAA
                                                                                    T        T  G
```

Figure 27. The sequence of p256A3, the MHTI A PCR product clone, and the comparison of its predicted protein sequence with that of MHTI A.

fragment was indeed derived from the MHTI A cDNA as the protein sequence translated from its DNA sequence exactly matched that of the MHTI A partial protein sequence. The unknown amino acid just 5' of the consensus oligo was shown to be lysine (K) in agreement with the MHTI B sequence. (figure 27)

Radiolabelling short DNA fragments

An attempt was made to screen the cDNA library using this fragment, however radiolabelling by random priming was unsuccessful, incorporating only 2.9×10^6 cpm of radiolabel, which was too low a specific activity to be of much use. Poor labelling of relatively short DNA fragments when labelled by the random priming method had been observed before in this laboratory and was thought to be due to the small number of "good" random hexanucleotide primer binding sites in so short a length of DNA. So an alternative method of labelling by PCR was devised.

This was essentially the normal PCR reaction with a small number of variations. The template DNA should be reasonably dilute; around 10 ng per reaction gave reliable results. This avoids diluting the labelled copies with excessive cold DNA and exhausting the limited amount of dCTP in the early cycles which would give rise to significant amounts of under-length linear amplification products from later cycles, and also excessive amounts of product from early cycles which amplified directly from the cold template and hence may include sequence from beyond the primers. This, if the template is a fragment cloned into a standard cloning vector, would give rise to sequences in the probe which hybridise to any vector sequence. If a specific fragment of a sequence is required to make a gene specific probe, linear amplification of sequences from beyond the primers would give rise to cross hybridisation to other homologous but different genes. The template DNA should thus be cut out of its plasmid, purification should be unnecessary if the restriction is efficient, or be a purified PCR fragment from a larger sequence. Using PCR fragments directly could give very dirty results, presumably due to the presence of artifact sequences and "read through" fragments. Running out on gel and isolating the amplified band usually gave excellent results but the most reliable method of purifying PCR fragments was found to be subcloning them.

The PCR conditions were the same as would be used for the normal PCR of the same fragment, but only 12 to 15 cycles were used. The PCR reagents were also the same

except that the nucleotide mix contained only dATP, dGTP and dTTP, the dCTP being supplied as α^{32}PdCTP, 50 or 100 μCi of 400 Ci per mM activity α^{32}PdCTP per reaction.

The labelled DNA was purified from unincorporated nucleotides on a G50 Sephadex column exactly as for normal radiolabelling and was used from then on as any other radiolabelled probe.

A sample of the MHTI A PCR fragment (~10 ng) was labelled by this PCR method and $1.2x10^8$ cpm from 100μCi were incorporated, i.e. for the approximation 1 μCi of $^{32}P = 2.3x10^6$ cpm, 52% of the available dCTP. This is directly comparable with the levels of incorporation observed for "good" templates with the random priming method. As a similar amount of DNA is synthesised, because the same amount of dCTP is supplied, starting from a smaller amount of template the specific activity of the probe will actually be slightly higher than for random primer labelling.

Screening the *Manduca* cDNA library using the MHTI A PCR fragment

A sample of the MHTI A PCR fragment labelled by the PCR method (see above) was used to screen the *Manduca* cDNA library with 60°C chosen as the hybridisation temperature because of the short probe length and with washing to 2 x SSC, 0.1% SDS. This gave very clear results with 22 positives of various intensities. Six lines were chosen at random for secondary screening, 3 strong and 3 weak. These were subjected to 2 or 3 rounds of further screening and then plasmid rescue. The resulting colonies were miniprepped and the plasmids were DNA sequenced in from either end with the M13 forwards and reverse sequencing primers.

Five of the six plasmids had inserts which are clearly derived from genes of the MHTI family and fell into two distinct groups. Sequences p256C3, 6 and 7 which were identical except at the inosines in the primer with p256A3 over the length of this clone and give a predicted protein sequence identical with the MHTI A partial sequence as far as it runs, and p256C5 and 8 which are close but not identical to p256A3 and give a predicted protein sequence different to those of MHTI A and MHTI B.

The three MHTI A cDNAs, p256C3, 6 and 7, gave differing patterns upon restriction with EcoR I and Xho I and contained inserts of differing sizes. All contained an internal EcoR I site which generated a diagnostic ~230 bp fragment to the Xho I site at the 3' end, however the size of the other fragments varied. This was explained after further

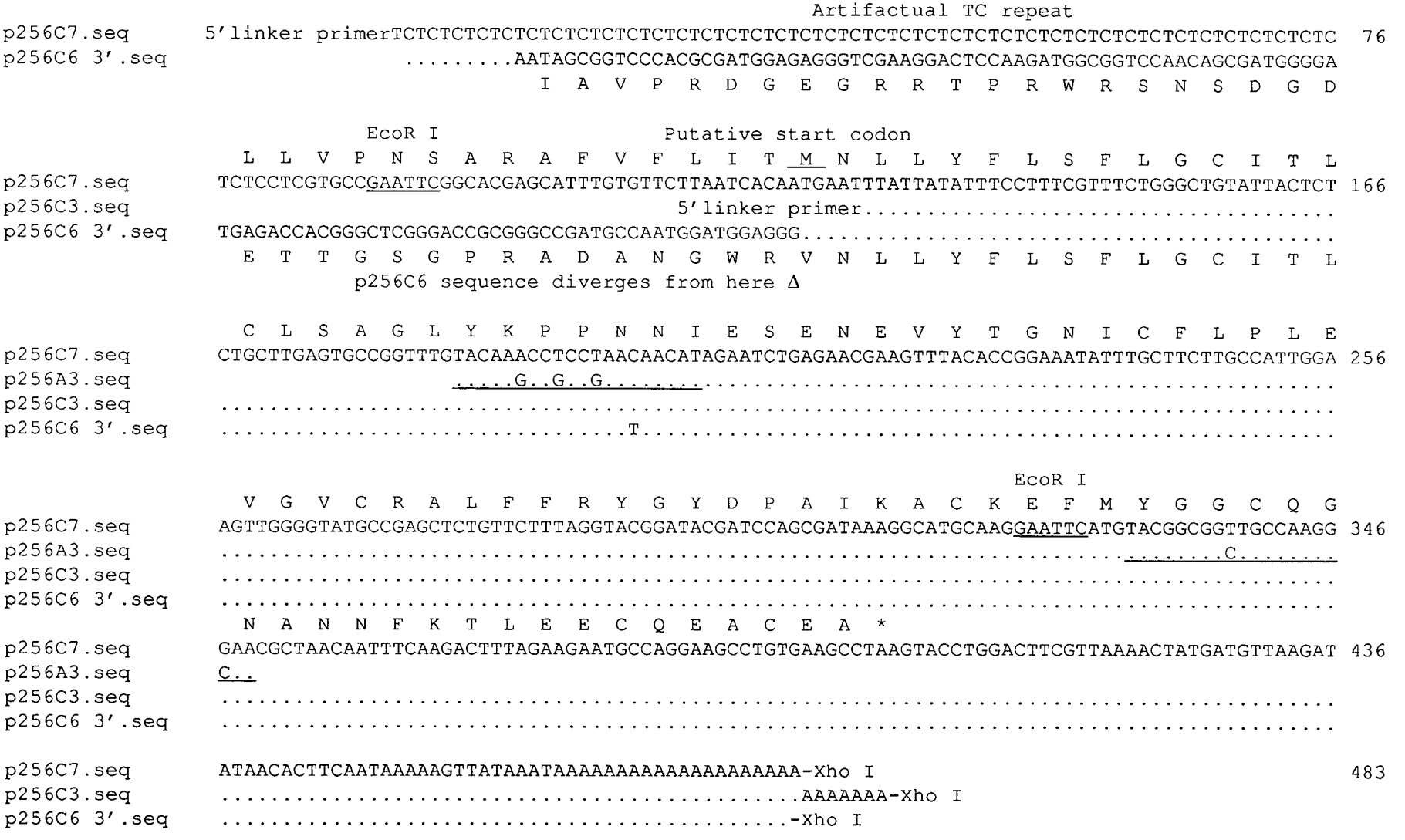


Figure 28. The DNA sequence of the MHTI A p256C cDNA clones and their translation in the MHIT A reading frame. The underlined sequences in p256A3 are the PCR primers.

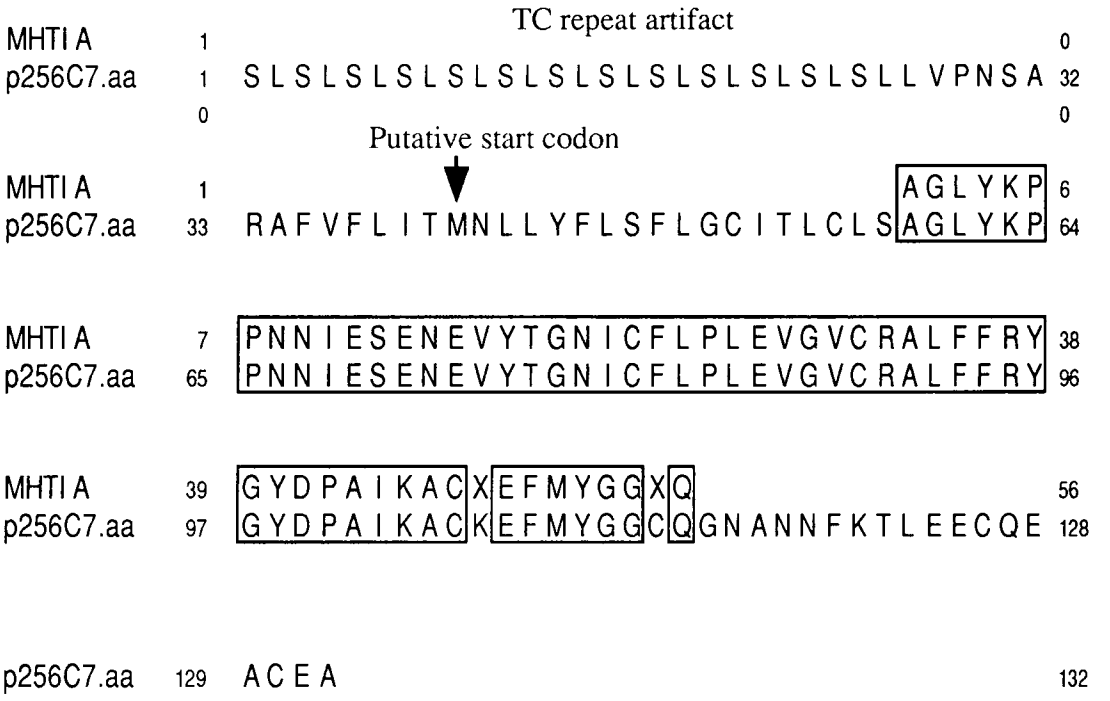


Figure 29. Comparison of the predicted protein sequence of the cDNA p256C7 and MHTI A.

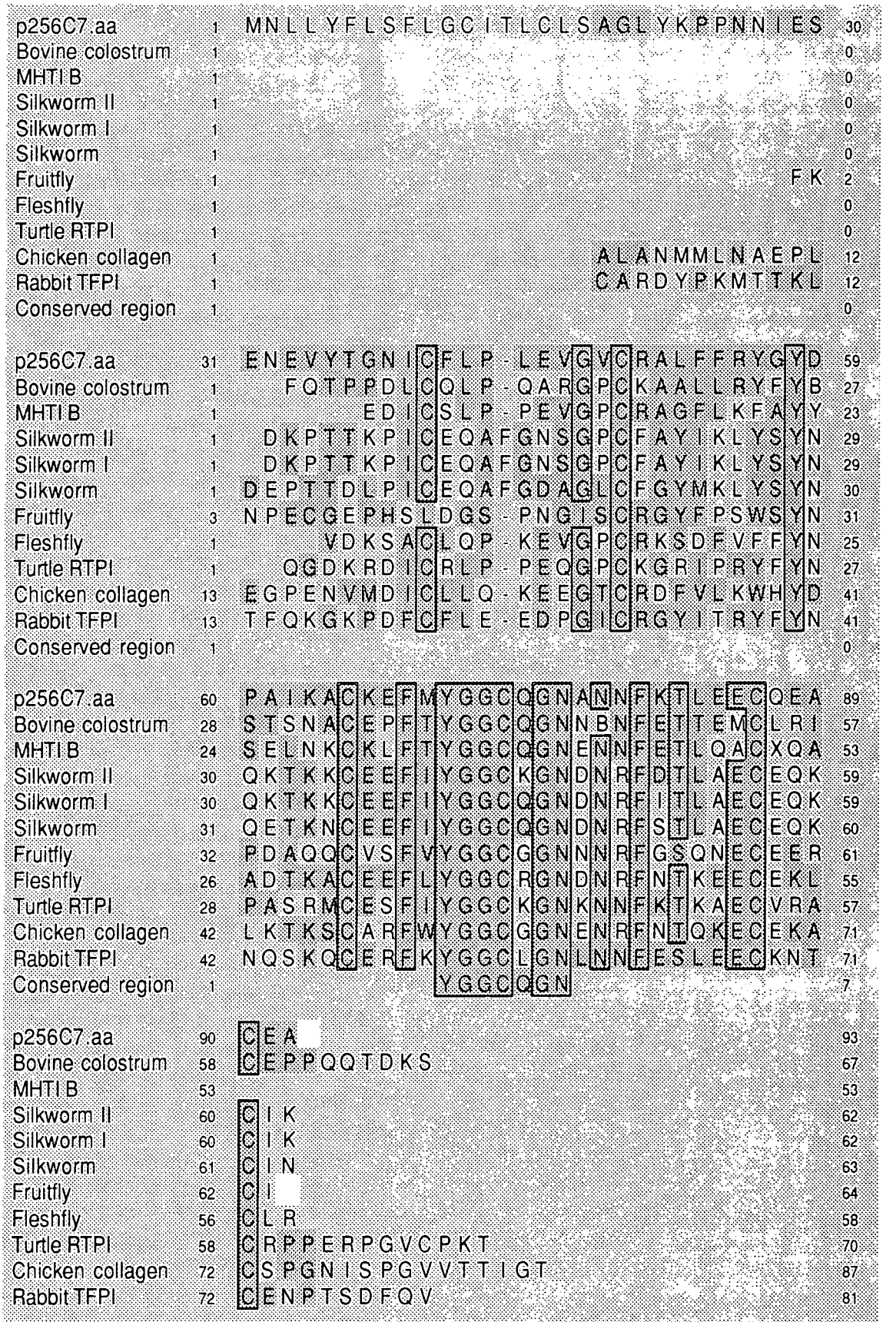


Figure 30. Alignment of the predicted protein sequence of the cDNA p256C7, MHTI A, with some homologous proteins from other sources.

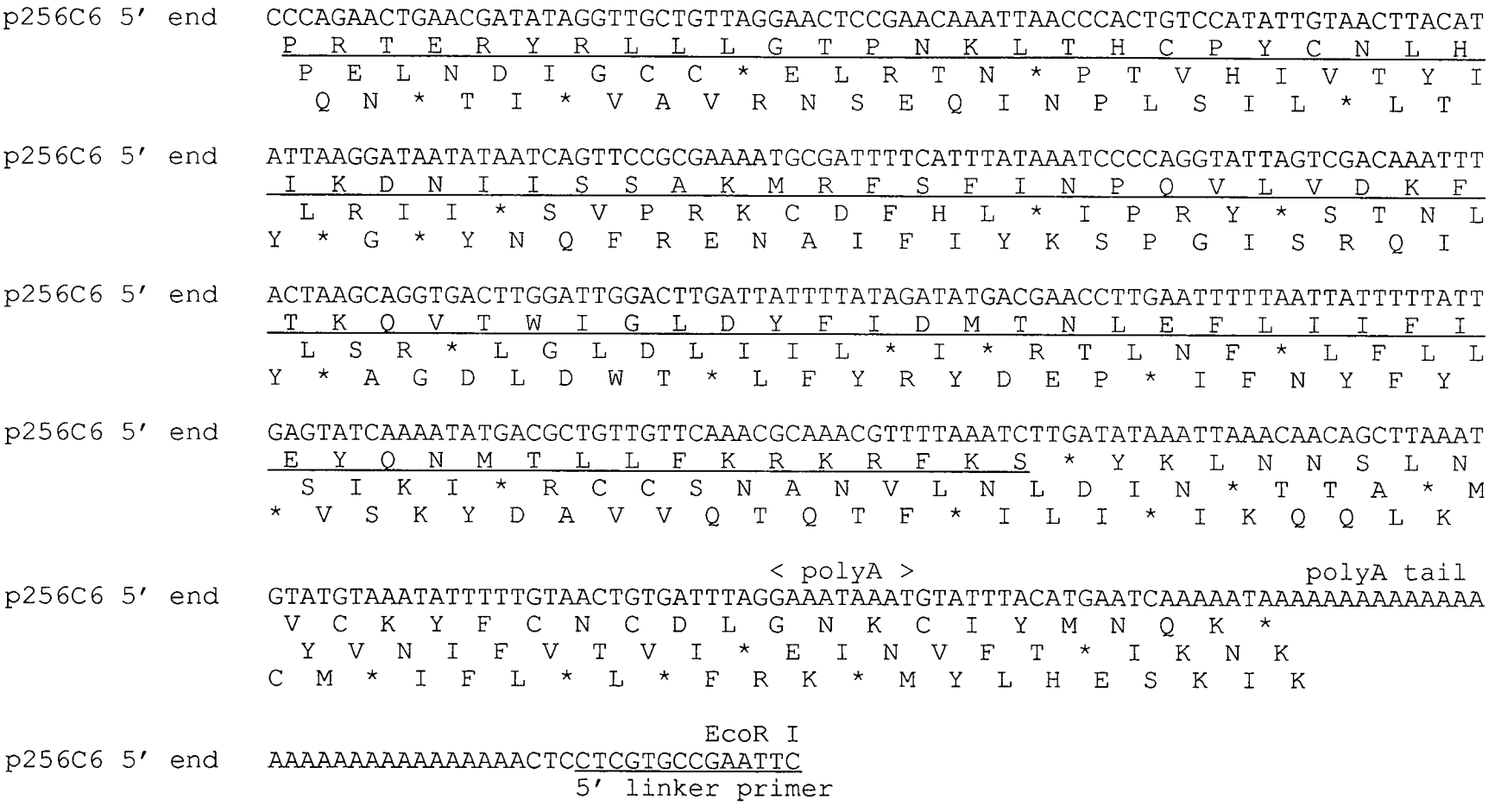


Figure 31. The DNA sequence of the 5' end of p256C6 (in reverse and complement) and the translation in all three reading frames. The long ORF is underlined as is the 5' linker primer.

DNA sequencing. (figures 28 and 29)

p256C7 extends upstream beyond the mature protein start through a methionine codon at base 123, an EcoR I site at base 89 and just beyond this goes into an 80 bp TC repeat sequence (the numbering is from the first base of the TC repeat sequence, i.e. the end of the 5' EcoR I cloning linker). This is presumably an artifact like that seen in the randomly selected 2.1 kb cDNA clone p254Z71 (see above). p256C3 ends upstream of the mature protein start but ends just down stream of the methionine codon which is seen in p256C7. The sequence of p256C6 diverges from that of p256C7 at the methionine codon (which changes the methionine codon to a valine codon ATG - GTG) and continues upstream without an in frame ATG as far as could be read leaving the MHTI A reading frame without a start codon. The insert of p256C6 when cut with EcoR I and Xho I gives the diagnostic MHTI A internal EcoR I to 3' linker-primer Xho I fragment of ~230 bp and a fragment of ~1200 bp. When sequenced from the 5' end into this 1200 bp fragment the EcoR I linker is found to be fused to a polyA tail. This suggests that the 5'end of p256C6 is a second cDNA fused in reverse orientation to a MHTI cDNA with the joining point at the point where the sequence diverges from that of p256C7. This interpretation was reinforced by the result of the Northern blot (see below).

The only other sequence difference between the cDNAs of this group is a silent base change between p256C6 and the other two cDNAs. The DNA sequence extends 3' to a stop codon giving a coding sequence which extends beyond the published MHTI partial protein sequence (Ramesh, Sugumaran and Mole 1988) and is closely homologous to the other secretory trypsin inhibitors, particularly in the conservation of the cysteine residues. (figure 30) The stop codon is followed by 48 bp of 3'untranslated sequence and then by the polyA tails of 19, 20 and 27 bp, which all start at the same position 12 bp down stream of a consensus polyadenylation signal. In light of the result of the Northern blot (see below) and the lack of any sequence differences between these cDNAs, apart from how far they extend 5', it was concluded that they had all derived from the same gene and that the 5' sequence differences were artifacts.

The 5' artifact sequence of p256C6, which was only partially sequenced, had an open reading frame which could be followed for ~80 amino acids before the DNA sequence became unreliable. This ORF was used to run a fasta search of the protein sequence database at Daresbury but no matches of any significance were found. (figure 31)

The two other MHTI type cDNAs (p256C8 and p256C5), upon restriction with

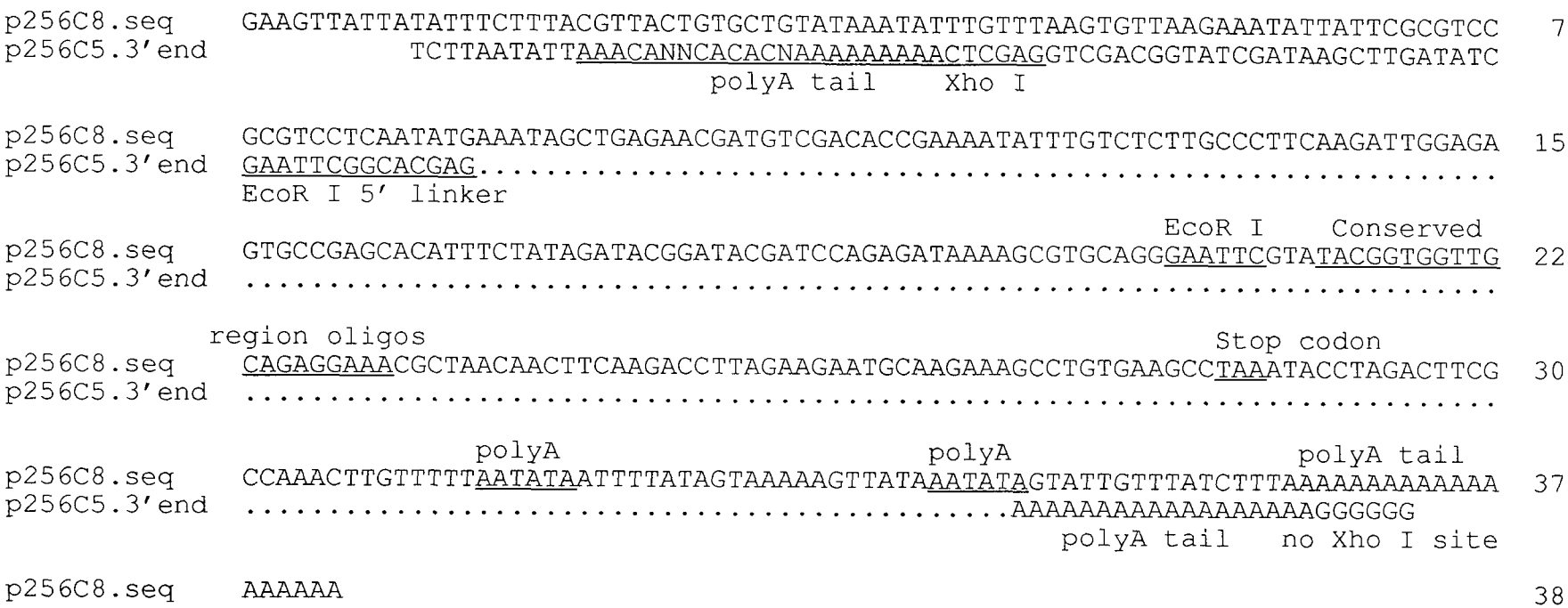


Figure 32. DNA sequence alignments of the two MPI C cDNAs p256C8 and p256C5 showing the cloning artifact at the 5' end of p256C5.

Figure 33. DNA sequence alignment of p256C7 the MHTI A cDNA and p256C8 the MPI C cDNA.

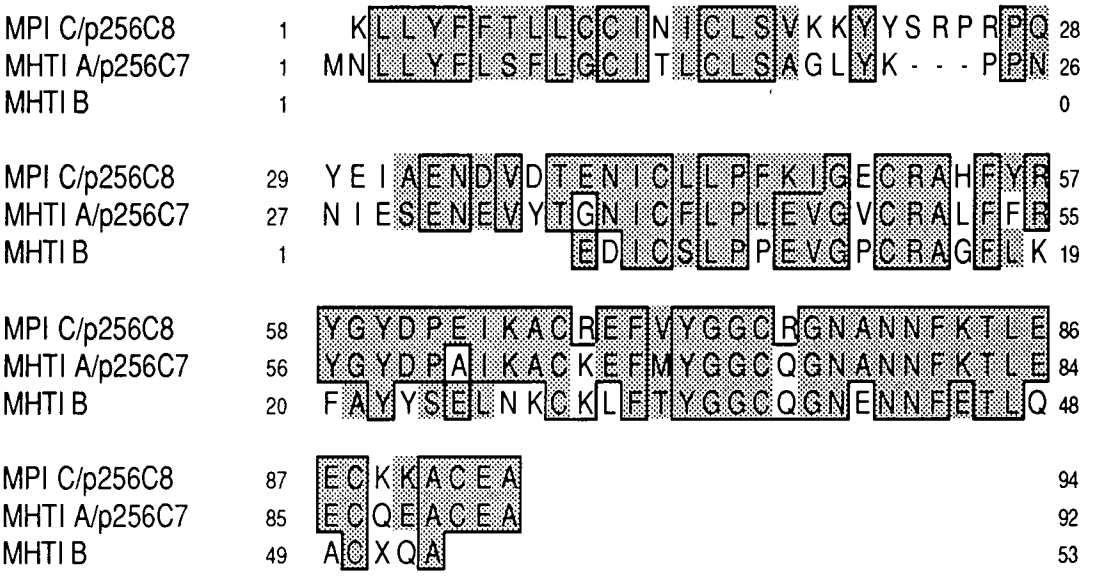


Figure 34. Alignment of the three types of *Manduca* trypsin inhibitor protein sequence.

EcoR I and Xho I, each gave a similar ~230 bp fragment which corresponded to an analogous internal EcoR I to 3' linker-primer Xho I fragment as was seen for the MHTI A cDNAs but again gave different sizes for the other fragments. These cDNAs were identical to each other in their region of overlap and closely homologous to p256C7, ~80% in the coding region ignoring insertions. The more complete, p256C8, ended at the 5' end in a position corresponding to base 125 of p256C7, i.e. the G of the methionine codon (base 123). Many of the sequence differences between p254C8 and the MHTI A cDNAs are silent or conservative amino acid changes and the homology extends some way into the 3' untranslated sequence. The site of polyadenylation is different in the two clones of this type and also differs from the p256C7 type being, in p256C8, 15 bp 3' of a imperfect polyadenylation signal AATATA and at this signal in p256C5. (figures 32 and 33)

The predicted protein sequence of the p256C8 type cDNAs, which I have provisionally called *Manduca* protease inhibitor C (MPI C), is very similar to MHTI A with all the cysteines aligning except an extra one in the MPI C leader sequence and two insertions of two and one amino acid residues just down stream of the MHTI A 5' PCR primer site. This primer would not be expected to fit the MPI C sequence. It is more homologous to MHTI A than MHTI B over the region where sequence of all three is available; A and C have identity in 40 of 50 amino acid residues compared with A to B which have identity in 28 of 50 and C to B which have identity in 25 of 50. A major difference between MHTI A and MPI C is the amino acid change of glutamine to argenine (Q to R) in the conserved sequence YGGCQGN which is also seen in the inhibitor from flesh fly (Papayannopoulos and Biemann 1992). MPI C would be predicted to be a trypsin inhibitor on the basis that its predicted reactive bond is argenine-alanine (R-A) (Laslowski and Kato 1980) at residues 51 and 52 in the p256C8 sequence. (figure 34)

Similar to the MHTI type clone p256C6, the MTI C clone p256C5 consists of two cDNAs fused together, although in this case fused head to tail. The 3' sequence is identical to p256C8 except for the polyA tail starting earlier and extending 5' some 89 bp less. Reading in the 3' to 5' direction, for the MTI C sequence and as defined by the cloning vector, the sequence then shows a 5' then a 3' linker-primer, a polyA tail and then a fresh cDNA the sequence of which could not be read from this end (figure 32).

Sequencing in from the 5'end showed a damaged 5' linker and a cDNA with an ORF which could be followed for some 80 amino acid residues. This was used to run a fasta search of the protein sequence database at Daresbury and showed strong homology to

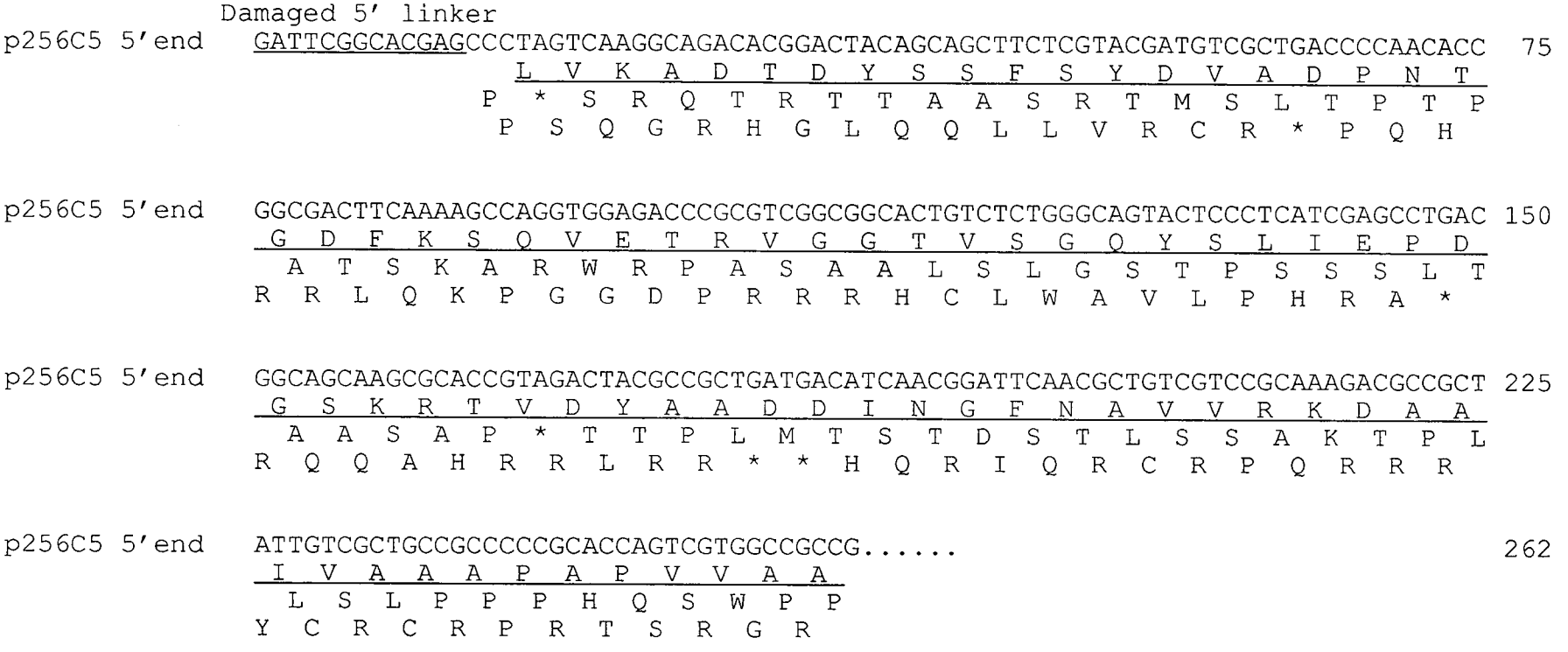


Figure 35. The DNA sequence of the 5' end of p256C5 and the translation in all three reading frames. The long ORF is underlined as is the 5' linker primer.

Alignment of p256C5 5'ORF with partial sequence of Swissprot cup8_d (P27780); pupal cuticle protein edg-84 precursor from *D. melanogaster* (59.1% identity in 66 aa overlap).

```
                                                 10        20        30
p256C5                                        LVKADTDYSSFSYDVADPNTGDFKSQVETRVGG
                                                 ::::| ||:|||:||:||| ||| |:| |:
cup8_d VKTALFVTLIGLAQAGPLPAKSSGSEDTYDSHPQY-SFNYDVQDPETGDVKSQSESRDGD
             10        20        30         40        50        60

            40        50        60        70        80
p256C5 TVSGQYSLIEPDGSKRTVDYAADDINGFNAVVRKDAAIVAAAPAPVVA
       :| |||||: ::|| :|||||:|||::|||||||:::
cup8_d VVHGQYSVNDADGYRRTVDYTADDVRGFNAVVRREPLSSAAVVVKPQATAVVPKVQLKPL
              70        80        90       100       110       120
```

Alignment of p256C5 5'ORF with partial sequence of Swissprot cu8_locmi (P11734); cuticle protein 8 from *Locusta migratoria* (58.0% identity in 69 aa overlap).

```
                                        10        20        30        40
p256C5                               LVKADTDYSSFSYDVADPNTGDFKSQVETRVGGTVSGQYS
                                        |:|:| |::|||:| | |:| |::|:| ||
cu08_l KAVVAAPAVAYAAPVAKAVVAEPVAYPKYEFNYGVHDAHTGDIKQQSEARDGDVVKGSYS
             40        50        60        70        80        90

               50        60        70        80
p256C5 LIEPDGSKRTVDYAADDINGFNAVVRKDAAIVAAAPAPVVA
       |:|||||:|||:|:||| |||||||:::::: ::| |||:
cu08_l LVEPDGSTRTVEYQADDHNGFNAVVHRTPGTHPVAVAPVAVAHAPVAVAHAPIAYH
            100       110       120       130       140
```

Alignment of p256C5 5'ORF with partial sequence of Swissprot ac20_tenmo (P26967); adult specific cuticular protein ACP-20 precursor from *Tenebrio molitor* (52.9% identity in 71 aa overlap).

```
                                              10        20        30
p256C5                                     LVKADTDYSSFSYDVADPNTGDFKSQVETRV
                                              ::: ::|  |:|:|:|::||| | |:|:||
ac20_t GLGGLGGGLGGVGVGLGGVGVVGGGHGVVDLHTPAHY-QFKYGVEDHRTGDRKQQAEVRV
             40        50        60         70        80

            40        50        60        70        80
p256C5 GGTVSGQYSLIEPDGSKRTVDYAADDINGFNAVVRKDAAIVAAAPAPVVA
       |::|:|:||| |||||: |:|:|:||| |||||||:: :  |
ac20_t GDVVKGEYSLAEPDGTVRVVKYTADDHNGFNAVVSRVGHAVHPQVLVRKAVVPVATHGVV
      90       100       110       120       130       140
```

Alignment of p256C5 5'ORF with partial sequence of Swissprot ac22_tenmo (P26968); adult specific cuticular protein ACP-22 precursor from *Tenebrio molitor* (50.8% identity in 65 aa overlap).

```
                                              10        20        30
p256C5                                     LVKADTDYSSFSYDVADPNTGDFKSQVETRV
                                              :|||:::| | :|:|||::| |||||: |:|
ac22_t GGGEGRHYEIGFGGSHFNTPVDVHHEEAIHLKAHPEYHS-DYHVADHKTKDFKSKHEVRD
     100       110       120       130       140       150

            40        50        60        70        80
p256C5 GGTVSGQYSLIEPDGSK-RTVDYAADDINGFNAVVRKDAAIVAAAPAPVVA
       | :|:| |||:||| :: |:|||::|: :|| | |:
ac22_t GYKVKGTYSLLEPDHKTVRVVDYVSDKKRGFIARVSYRKHH
      160       170       180       190
```

Figure 36. Alignment of the 5' ORF of p256C5 with part of the protein sequences of some insect cuticular proteins selected by a fasta search of the nation protein sequence databank. Other non significant matches are not shown

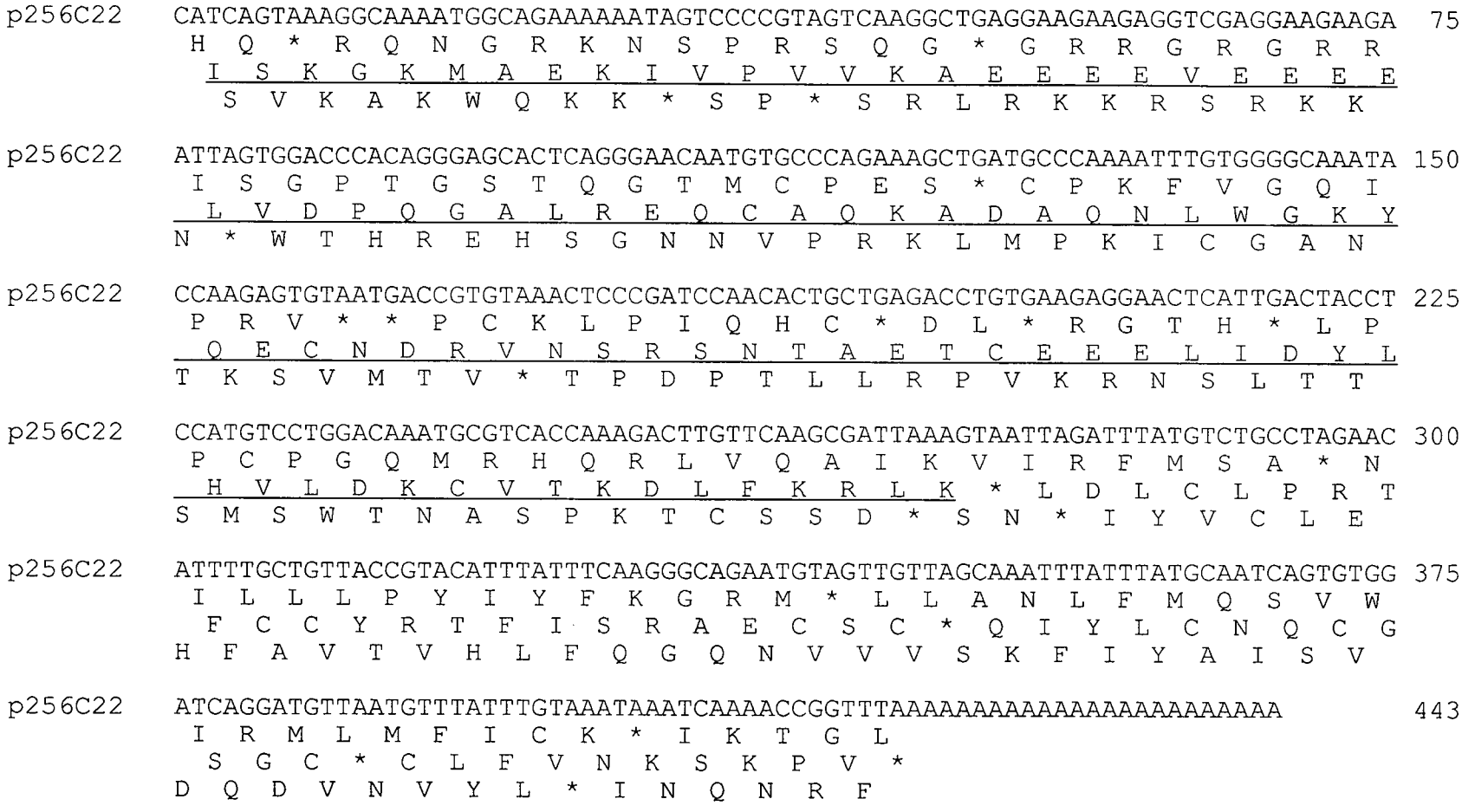


Figure 37. The DNA sequence of p256C22 and the translation in all three reading frames. The long ORF is underlined.

Alignment of p256C22 with partial sequence of Swissprot CCBO11; bovine ubiquinol-cytochrome c reductase 11kD protein (EC 1.10.2.2) (50% identity in 76 aa overlap).

```
                10        20        30        40        50        60
p256C2 ISKGKMAEKIVPVVKAEEEEVEEEELVDPQGALREQCAQKADAQNLWGKYQECNDRVNSR
                     ..X:: :::::::: ..:::::.: ..  . ... . :..:.::
CCBO11              GDPKEEEEEEEELVDPLTTVREQCEQLEKCVKARERLELCDERVSSR
                           10        20        30        40

                70        80        90
p256C2 SNTAETCEEELIDYLHVLDKCVTKDLFKRLK
       :.:.:.:.:::.:::. :.::.....:.:X
CCBO11 SQTEEDCTEELLDFLHARDHCVAHKLFNSLK
        50        60        70
```

Alignment of p256C22 with partial sequence of Swissprot s00219; human ubiquinol-cytochrome c reductase 11kD protein precursor (EC 1.10.2.2) (50% identity in 76 aa overlap).

```
                10        20        30        40        50        60
p256C2 ISKGKMAEKIVPVVKAEEEEVEEEELVDPQGALREQCAQKADAQNLWGKYQECNDRVNSR
                    .X::: :::::::: ..:::::.: ..  . ... . :..: .::
S00219 MGLEDEQKMLTESGDPEEEEEEEEELVDPLTTVREQCEQLEKCVKARERLELCDERDSSR
                10        20        30        40        50        60

                70        80        90
p256C2 SNTAETCEEELIDYLHVLDKCVTKDLFKRLK
       :.:.:.:.:::.:::. :.::.....:.:X
S00219 SHTEEDCTEELFDFLHARDHCVAHKLFNNLK
                70        80        90
```

Alignment of p256C22 with partial sequence of Swissprot RDBYUC; yeast ubiquinol-cytochrome c reductase 17 kD protein (EC 1.10.2.2) (32.5% identity in 83 aa overlap).

```
                           10        20        30        40
p256C2                ISKGKMAEKIVPVVKAEEEEVEEEELVDPQGALREQCAQKADAQNL
                                    .:.: X:::.:. ..::::.  ........:
RDBYUC EKAAEGEEKEEENGDEDEDEDEDEDDDDDDDEDEEEEEEVTDQLEDLREHFKNTEEGKAL
          40        50        60        70        80        90

         50        60                  70        80        90
p256C2 WGKYQECNDRVNSRSNTA--------ETCEEELIDYLHVLDKCVTKDLFKRLK
        .:.::..:X. . . .         :.: ::...  : ::. ..  ::..::
RDBYUC VHHYEECAERVKIQQQQPGYADLEHKEDCVEEFFHLQHYLDTATAPRLFDKLK
          100       110       120       130       140
```

Figure 38. Alignment of the p256C22 ORF with part of the protein sequences of some ubiquinol-cytochrome c proteins selected by a fasta search of the national protein databank. Other non significant matches are not shown.

insect cuticle proteins from *Drosophila melanogaster* (Fruit fly) (Apple and Fristrom 1991), *Locusta migratoria* (Locust) (Klarskov *et al.* 1998) and *Tenebrio motolitor* (the yellow meal worm) (Bouhdin *et al.* 1992, Charles *et al.* 1992). The homology scores are in the range 50 - 59% over a 65 - 71 amino acid region. The p256C5 5' sequence has no in frame start codon and by comparison with the above proteins starts down stream of the mature protein start (between 30 and 130 amino acids) but does correspond to the usually central conserved region in this diverse class of proteins. (figures 35 and 36)

During every moult in insects the epidermis synthesises a cuticle which is a complex extracellular layer consisting mainly of chitin fibres and proteins. This structure can have a wide range of mechanical properties depending on the arrangement of the fibres and the degree of sclerotization of the components and these factors are both spatially and developmentally regulated. Cuticle proteins are known to play a major role in the determination of cuticle properties, mainly by chitin protein cross linking (Charles *et al.* 1992). No further investigations of this sequence were undertaken.

The sixth cDNA isolated (p256C22) was plasmid rescued from an insufficiently plaque purified phage line and contains no sequence which hybridises to the MHTI A PCR fragment. When it was sequenced it was found to be a cDNA of 443 bp with a 24 bp polyA tail located 13 bp 3' of a consensus polyadenylation signal. It contained an open reading frame running in from the 5' end which when used to search the protein data base at Daresbury was shown to code for a protein homologous to subunit VI of ubiquinol-cytochrome c reductase also known as the mitochondrial hinge protein (Wakayabashi *et al.* 1982, Van Loon *et al.* 1984, Ohta *et al.* 1987). (figures 37 and 38)

Mitochondrial hinge protein is a subunit of ubiquitin-cytochrome c reductase (complex III or cytochrome bc_1 complex) which "hinges" cytochrome c with cytochrome c_1 and hence is part of the mitochondrial respiratory chain which is located in the mitochondrial inner membrane. It is encoded on the nuclear genome and whereas most targeting sequences are basic this has an unusual acidic targeting sequence, which is unique to mitochondrial targeting (Ohta *et al.* 1987). The partial sequence of the *Manduca* protein is very rich in glutamic acid (E) residues, including a sequence EEEEVEEEE, as are the other published sequences, but it also fits well outside these regions especially to the yeast sequence (Van loon *et al.* 1984). By comparison with this the cDNA sequence starts 50 amino acids short of the start codon i.e. is 150 - 200 bp less than full length.

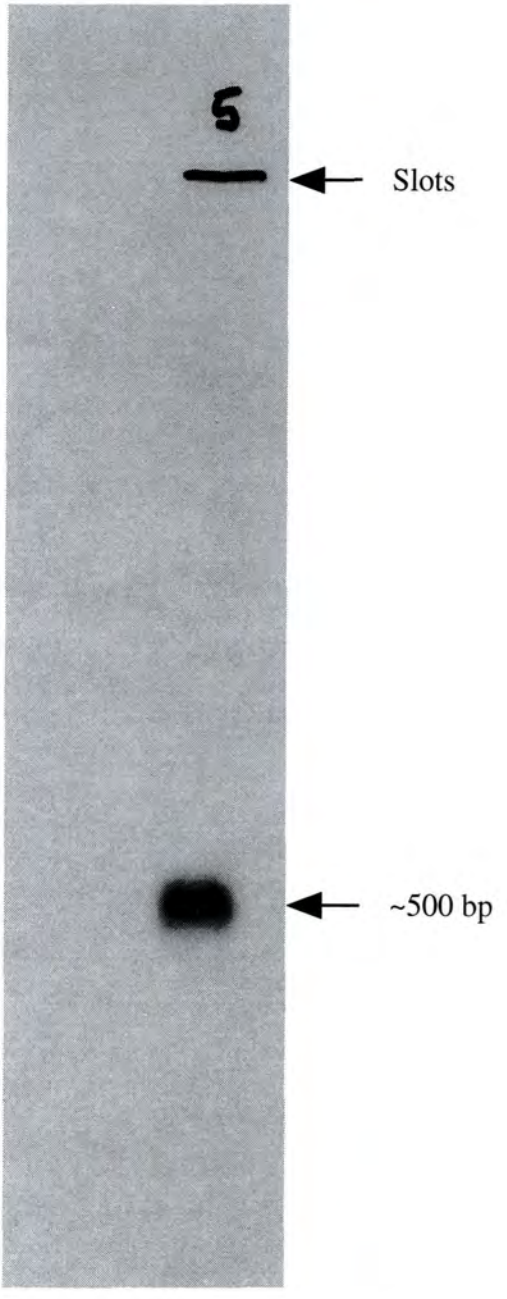


Figure 39. A Northern blot of *Manduca* polyA+ RNA probed with the *Manduca* haemolymph trypsin inhibitor A PCR product.

Probing a Northern blot with the PCR product

A Northern blot of polyA+ RNA from *Manduca* (see above) was probed with the MHTI A PCR product with prehybridisation and hybridisation at 42°C in 50% formamide and with washing to 1 x SSC, 0.1% SDS 65°C. (figure 39) This gave a single band at ~500 bases which corresponded to the lower edge of signal seen when a Northern blot was probed with the consensus oligo #575 (see above). No extra bands were seen on long exposure which confirms the artifactual natures of p256C5 and p256C6. For p256C6 not to be an artifact it should continue its ORF to a start codon. No in frame start codon occurs as far as the DNA sequence can be reliably read in from the 3' end, which is about 500 bases. Thus with a start codon and some 5' untranslated sequence the message would be greater then 500 bases long, which is incompatible with the Northern blot result.

A message of 500 bases is in excellent agreement with the p256C7 DNA sequence (ignoring the TC repeat artifact) which is shown as very near full length. Thus the methionine codon at base 123 in the p256C7 sequence is probably the start codon. This start codon is a good match to the animal start codon consensus and, in terms of expressing this insect derived protein in transgenic plants, is a fair match to the plant consensus apart from the two bases after the ATG. GC is very strongly favoured in plants, although A at either position occurs about 5% of the time and the following T is weakly favoured (Lütcke *et al.* 1987). (See figure 28)

Although the Northern blot was washed to higher stringency, about 95%, than the library screen, the message for MPI C would still be expected to show a signal on long exposure and the absence of any extra bands means that the MPI C message is of similar size to that for MHTI A and hence that p256C8 is little short of full length.

The assembly of the MHTI A expression constructs

It was thought that the expression in bacteria, and particularly in plants, of MHTI A using its own leader sequence may be problematical as the MHTI A leader would be expected to target for secretion through the cell membrane into the insect haemolymph, and plant cells have a radically different structure to insect cells in terms of the cell wall/outer membrane. For expression *in planta* this would probably mean accumulation of the

protein in the periplasmic space, i.e. transport through the membrane but failure to transport through the cell wall. *E. coli* expression is usually easier and more efficient for proteins not secreted into the medium (pET System manual Novagen) however the different nature of the prokaryote cell membrane to that of eukaryotes may be expected to make transport to the medium unlikely.

The strategy chosen to address these problems was to express the MHTI A cDNA in *E. coli* without a leader, and to make an expression construct with the leader from a plant gene replacing that of MHTI A. After surveying the easily available plant genes two good possibilities were found which were from proteins which showed high levels of tissue-specific expression and are targeted to the vacuole, where an accumulation of foreign gene products has been found normally to not affect the plant. The other criterium for selection was to preserve as much as possible of the protein sequence around the leader splice site, in particular the amino acids on either side, and not to change the leader to one of greatly different length.

The donor genes were two of the *Phaseolus vulgaris* lectin like genes, the erythroagglutinating phytohemagglutinin (PHA-E) gene (Hoffman and Donaldson 1985) and the α–amylase inhibitor (α-AI) gene (Hoffman, Ma and Barker 1982), both of which I had previously isolated from a number of *Phaseolus vulgaris* cultivars and from a Mexican wild line. These had been isolated by PCR using an oligo designed to fit the 14 bp conserved region around the start codon with a fusion to a BamH I restriction site designed to allow the expression of the gene as a lac-fusion protein in pUC18, and a second oligo designed to fit the 15 bp conserved region around the stop codon and to introduce a second BamH I site for cloning purposes.

The leader splice in all three proteins is between a serine and an alanine, amino acids 18 and 19 in MHTI A and 21 and 22 in PHA-E and α-AI. By changing the codon usage for these amino acids a unique Sac II restriction site could be created which would allow easy assembly of MHTI A with a leader sequence from either of the lectin like genes. Lacking any further selection criteria I decided to use the leader from PHA-E.

The following constructs were prepared by PCR and conventional subcloning, and checked by restriction mapping and DNA sequencing.

p259F10 which contained the whole MHTI A coding sequence including the leader in pGEMT with an engineered BamH I site at the 5' end and a Kpn I site at the 3' end.

p259B58 which contained the MHTI A coding sequence minus the leader in

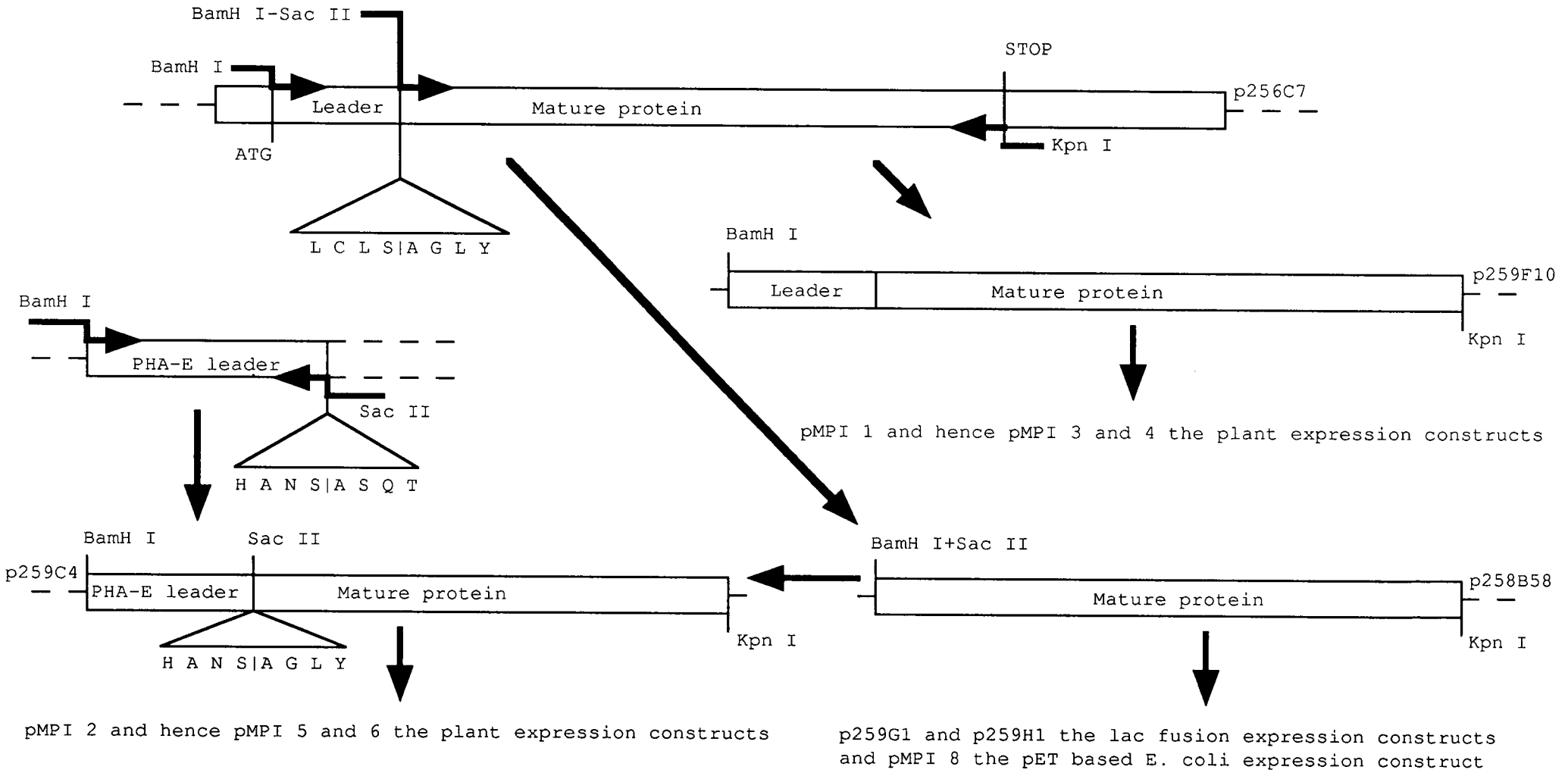


Figure 40. Schematic diagram of how the various MHTI A expression constructs were prepared.

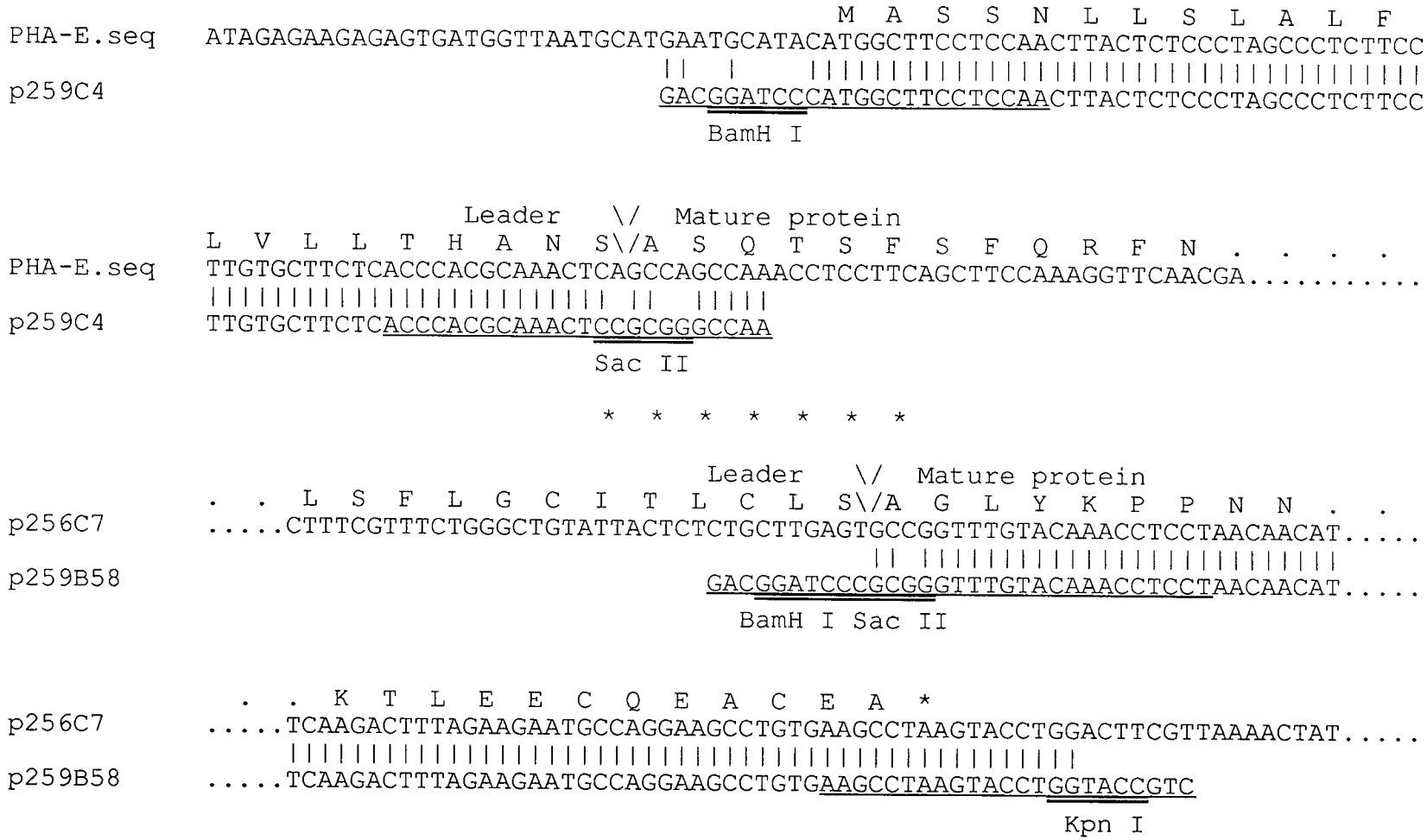


Figure 41. Primer design for the MHTI A expression constructs. Primers are underlined and restriction sites double underlined. Only the sequence around the primer sites is shown.

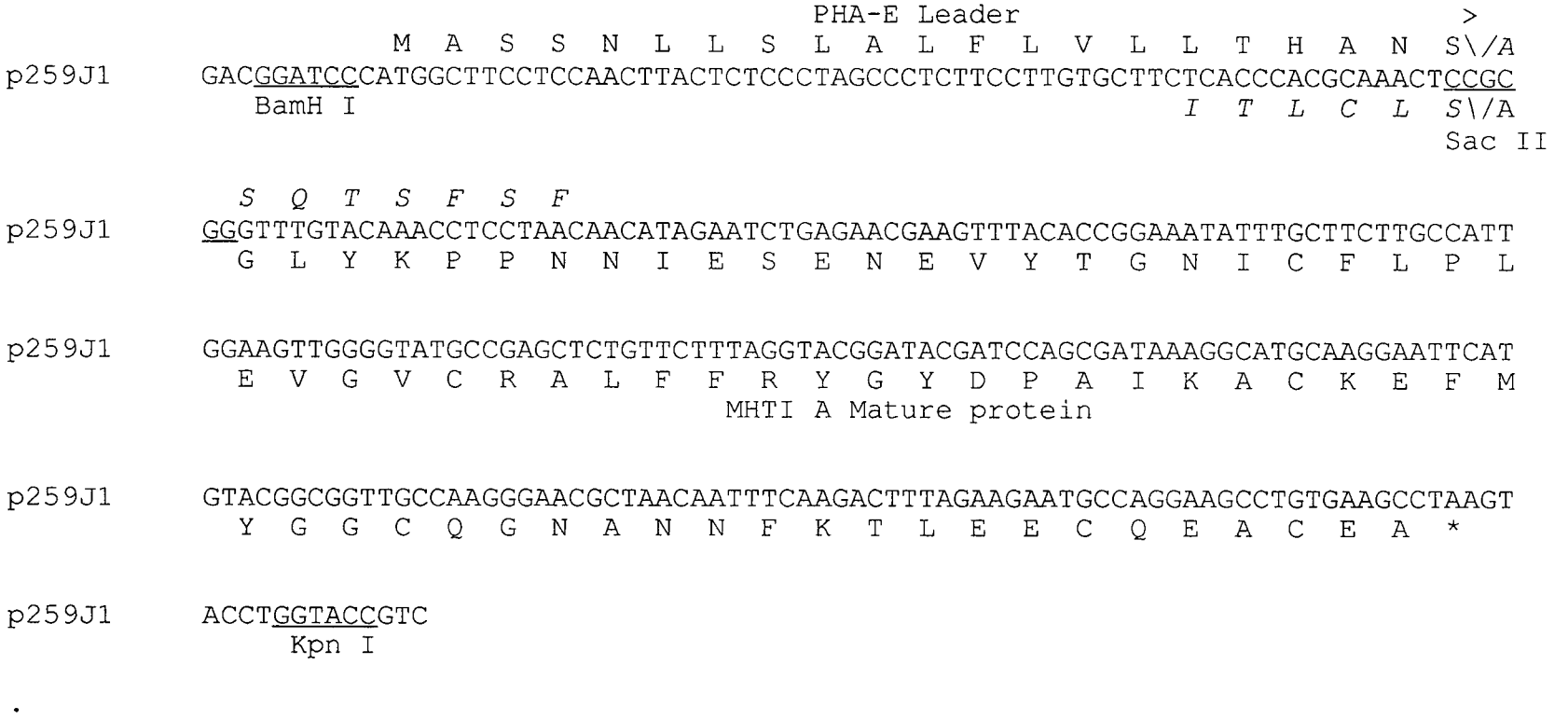


Figure 42. p259J1 the MHTI A expression construct with the PHA-E leader sequence. Restriction sites are underlined. The italics show the protein sequences of the original genes.

pGEMT, with engineered BamH I and Sac II sites at the 5' end and a Kpn I site at the 3' end.

p259G1 and H1, which were the insert of p259B58 subcloned with BamH I and Kpn I into pUC18 and 19 respectively. These were the reverse and forward lac fusion bacterial expression constructs.

p259C4 the PHA-E leader sequence in pGEMT with an engineered BamH I site at the 5' end and an engineered Sac II site at the 3' end.

p259J1 which was assembled from p259B58 and p259C4 and was the MHTI A mature protein coding sequence with the PHA-E leader.

The choice of BamH I and Kpn I sites at the 5' and 3' ends respectively was to allow easy assembly into pROK2 for transformation into plants and into pUC19 for lac fusion expression with the BamH I site designed to give the correct reading frame. (figures 40, 41 and 42)

Transgenic plants, bacterially expressed MHTI A and bioassays

It was decided to do the bacterial expression and plant transformation at AXIS where they used different vectors. So the MHTI A plasmids were sent down to AXIS where the following constructs were made by Dr. A. Merryweather:

pMPI1 is the BamH I to Kpn I fragment of p259F10 cloned into Bgl II and Kpn I polylinker sites between CaMV35S promoter and NOS terminator sequences carried on a pUC19 backbone. The promoter-MPI leader junction sequence is shown in figure 43.

pMPI2 is as above except the BamH I/Kpn I fragment came from p259J1. Therefore the leader sequence comes from PHA-E.

pMPI3 and 4 consist of the expression cassette of CaMV35S - MPI leader - MPI - NOS terminator taken from pMPI1 and cloned into the BamH I site of the pAPT5 binary vector (Newell *et al.* 1995, A. Merryweather personal communication). With respect to the Right (R) and Left (L) borders of the T-DNA, the MPI gene is expressed from L to R in pMPI3 and from R to L in pMPI4.

pMPI5 and 6 were constructed as for pMPI3 and 4 except that the expression cassette was derived from pMPI2, i.e. with the PHA-E leader. The MPI gene is expressed from L to R in pMPI5 and from R to L in pMPI6.

pMPI8 is a pET21 (Novagen) based plasmid expressing mature MHTI A protein

```
                                        Start codon
..GAGAACACGGGGGATCGGTACAGATCCCACAATG -> MPI leader
  CaMV35S prom      Bgl II/BamH I
```

The reading frame for the start codon of MHTI A in the plant expression construct pMPI1. The equivalent construct pMPI2 was made for the MHTI A with the PHA-E leader.

```
                                    Start codon
..GAGAACACGGGGGATCGGTACAGATCCAATG -> PHA-E leader
  CaMV35S prom      Bgl II/BamH I
```

* * * * *

```
T7 promoter - MPI junction sequence:
..AAGGAGATATACAT ATG GCG GGT TTG TAC AAA ACC TCC -->
  pET rbs       NdeI  Mature MPI -->

MPI COOH - T7 terminator sequence:
..TGT GAA GCC CTC GAG CAC CAC CAC CAC CAC CAC TGA GAT
  MPI COOH ->  XhoI   6 x Histidine tail      Stop
```

The reading frame at either end of the pET based MHTI A E. coli expression construct pMPI8.

Figure 43.

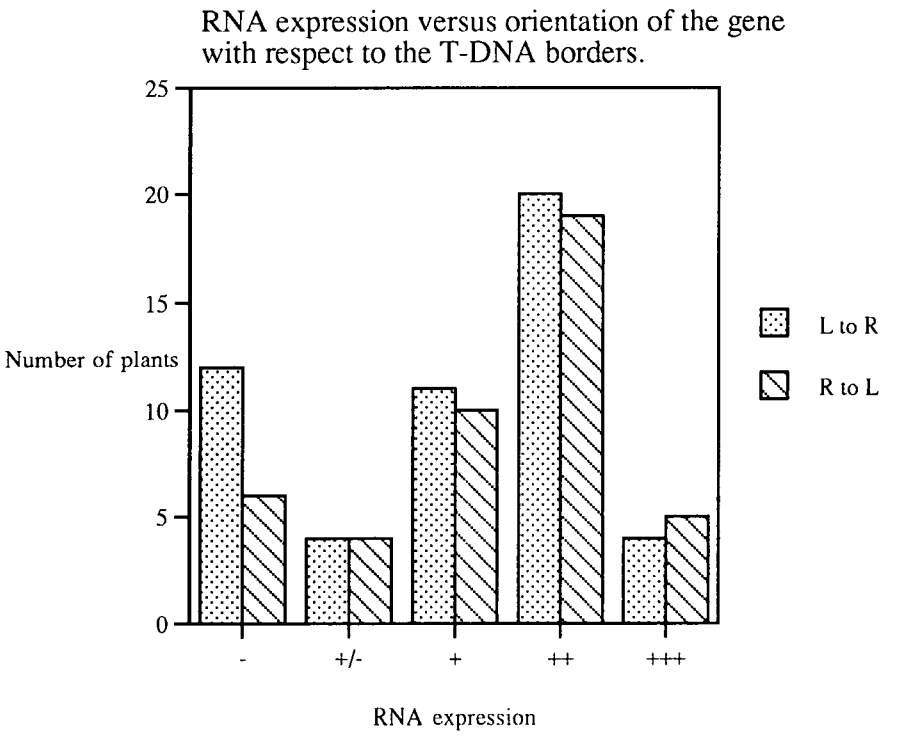

RNA expression versus orientation of the gene with respect to the T-DNA borders.
25
20
15
10
5
0
Number of plants
L to R
R to L
-
+/-
+
++
+++
RNA expression

Figure 44.

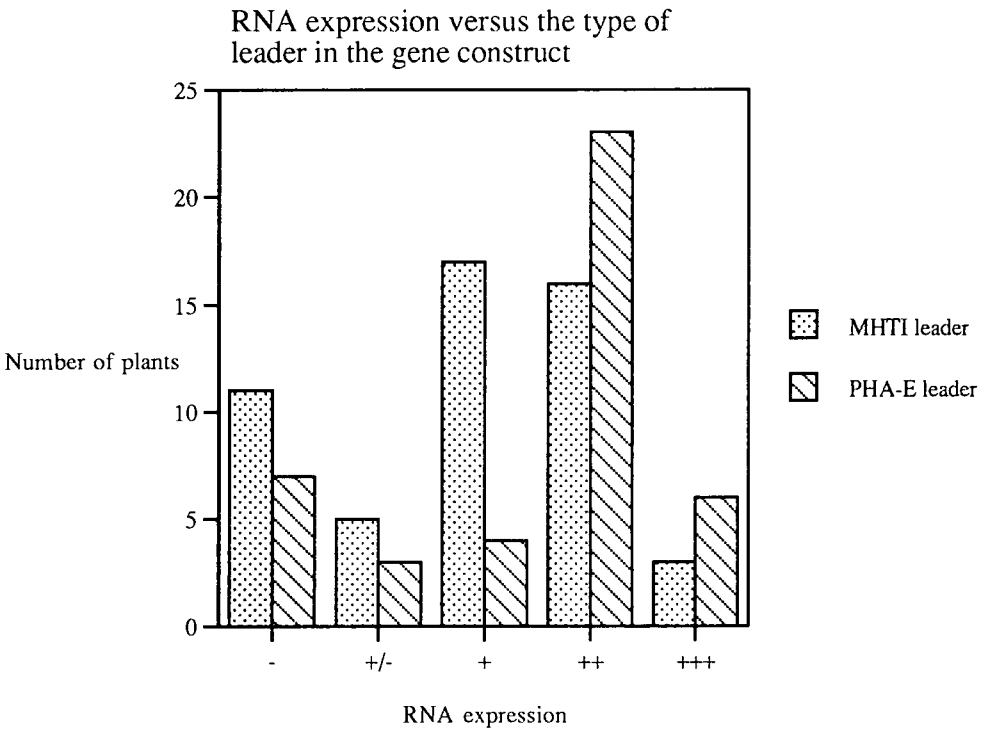

RNA expression versus the type of leader in the gene construct
25
20
15
10
5
0
Number of plants
-
+/-
+
++
+++
RNA expression
MHTI leader
PHA-E leader

Figure 45.

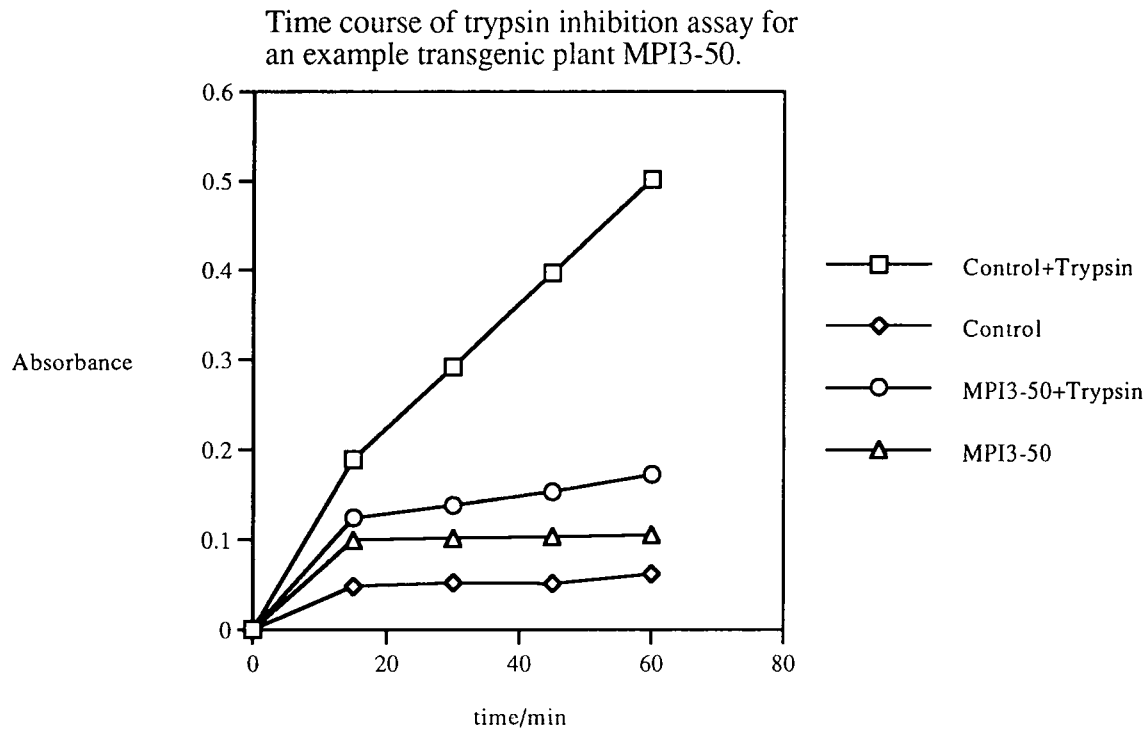

Time course of trypsin inhibition assay for an example transgenic plant MPI3-50.
Absorbance
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
time/min
Control+Trypsin
Control
MPI3-50+Trypsin
MPI3-50

Figure 46.

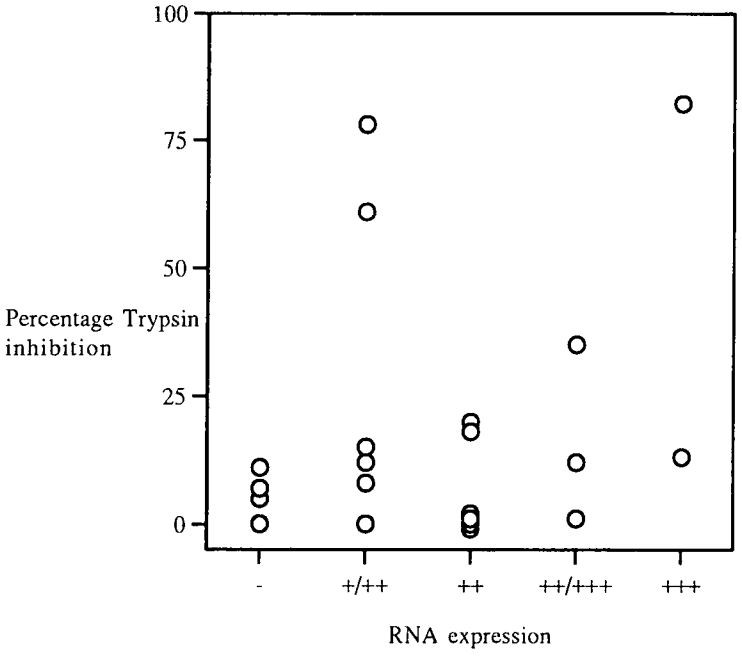


Figure 47. Scattergram to show the correlation of trypsin inhibitory activity and RNA expression of the transgenic plants.

containing no leader, and no amino terminal T7-tag sequences, but carrying a 6 x Histidine sequence at the 3' end. The fragment was PCRed to be in frame when cloned between the Nde I and Xho I sites of polypET2, which is an in house derivative of pET21c (Novagen) with an extended multipurpose cloning site. The reading frame is as shown in figure 43.

Tobacco was transformed with pMPI3, 4, 5 and 6 by *Agrobacterium* mediated gene transfer as described in the methods, and the transformed plants were screened for MHTI A expression mainly by dot blotting total RNA and probing the blots with a MHTI A cDNA (C. A. Newell personal communications). Expression level was assigned to one of five categories, '-' no expression to '+++' high expression, based on visual inspection of the blots. The distribution of expression levels showed no significant difference between expression from right to left or right to left (figure 44), but did show a clear increase in the proportion of high expressers in constructs with the PHA-E leader compared with the native insect leader (figure 45). This might result from more compatible codon usage or greater message stability with the plant derived sequences and vindicated the concern that the native insect leader would perform in plants much less favourably than a plant derived leader.

Transgenic plants showing the highest levels of expression of MHTI A were clonally propagated in tissue culture, along with control plants. Potential high expressers from lines pMPI5 and 6, and some plants taken straight from tissue culture were also assayed for trypsin inhibitory activity. That some of the transgenic plants were expressing active trypsin inhibitor was demonstrated by the *in vitro* inhibition of bovine trypsin by leaf soluble protein extracts. The time course of inhibition for 100 µg of leaf soluble protein from one transgenic line is illustrated in figure 46.

The level of trypsin inhibitory activity correlated very poorly with the results from the RNA dot blots, as may be seen in figure 47. We have recently demonstrated that transgenic expression may be much more variable within individual plants than was previously suspected (Hilder *et al.* 1995) and so estimates of expression based on single samples must be viewed with some caution.

The *E. coli* expression construct pMPI8 was used to make pure MHTI A protein (A. Merryweather personal communication) and this and the transgenic plants were sent to Horticulture Research International (HRI) for insect bioassay.

The bioassays at Horticulture Research International

The plants, the protein sample and the expression *E. coli* strain were bioassayed in various ways by D. J. Ellis at HRI.

The plant bioassays were performed on primary transformants as follows (D. J. Ellis Confidential report 1995); Plants were removed from tissue culture after rooting (approximately two weeks), and acclimatised to pots of soil in a glasshouse. Six to nine weeks after the final subculturing the plants were approximately 18 inches high and large enough to be assayed for activity against insects. For each line, five leaves on each of four plants were tested. Leaves to be tested were individually bagged with Crispywrap bags and two neonate larvae of *Manduca sexta* were placed on each leaf at the start of the assay. After 10 days the larvae were removed and weighed to obtain larval weights, and the amount of leaf area eaten per larva was measured. An untransformed, negative control line of Samsun tobacco was included with each assay along with various transgenic lines expressing other anti-insect genes as potential positive controls.

Various artificial and semi-artificial diet bioassays were also performed on a wider range of insects, however, the various artificial diet bioassays gave poor results with no significant resistance being seen for positive controls. A leaf assay method was developed using leaf discs embedded in agar. This was selected in preference to laying leaf material on top of damp filter paper, since the leaf discs remain in a healthier condition for the duration of the assay. The following method was used in all leaf assays (D. J. Ellis Confidential report 1995).

Leaf discs of 48 mm diameter are cut using a steel core-borer. 55mm plastic Petri-dishes are half-filled with sterile 1% technical No. 3 agar. On cooling, but whilst the agar is still molten, the leaf discs are laid on the surface with the upper-side down (leaf scorch can occur if the agar is too hot) and the plates are allowed to cool and set. The leaf discs may be either dipped or painted with the test sample prior to or after embedding. Etalfix (Maag), a wetter/spreader, is added to the samples to give a final concentration of 0.2% and up to 50 µl of sample is spread on any one disc. After drying neonate larvae are added, the number is dependent on species, normally 5 for *Manduca* or 10 for *Lacanobia* . The discs are incubated at 25°C, with lighting (16 hour day length) and assessed daily if possible.

The plant bioassays showed all of the problems of inherent variability which such

trials are notoriously prone to. Replicate plants derived from 45 different transgenic lines were bioassayed in 8 different experiments. However, the experimental design was seriously flawed in that too few negative control (untransformed) lines were included, an all too common fault of insect bioassays where materials are limited. A much more serious shortcoming was the use of seed grown negative control plants in most bioassays whereas the transgenic plants were always derived from tissue culture. Comparison of tissue culture and seed-derived, negative controls suggest that the insects grow to about twice the size on the former and do about twice as much damage. A similar phenomenon has been observed in this department, where *Heliothis virescens* performs much better on tissue culture derived plants than seed derived ones (V. A. Hilder personal communication) and the importance of appropriate choice of controls has been well recognised. The faster growth of caterpillars on tissue cultured tobacco is presumably due to their leaves being softer and more palatable than seed derived tissue. The consequence of this is that in the bioassays employing only tissue cultured lines (bioassays 1 and 3) MHTI A expressing lines were identified which performed significantly better than the negative controls in terms of mean larval weight (MPI6-50, MPI5-31) and mean leaf area eaten (MPI6-50, MPI5-31 and MPI3-39), as did the positive control line of GNA (Snowdrop lectin) expressing tobacco 15-GNA-35. By contrast, in the other bioassays the untransformed seed-derived plants were apparently more resistant than any of the tissue cultured lines, usually including the positive control 15-GNA-35.

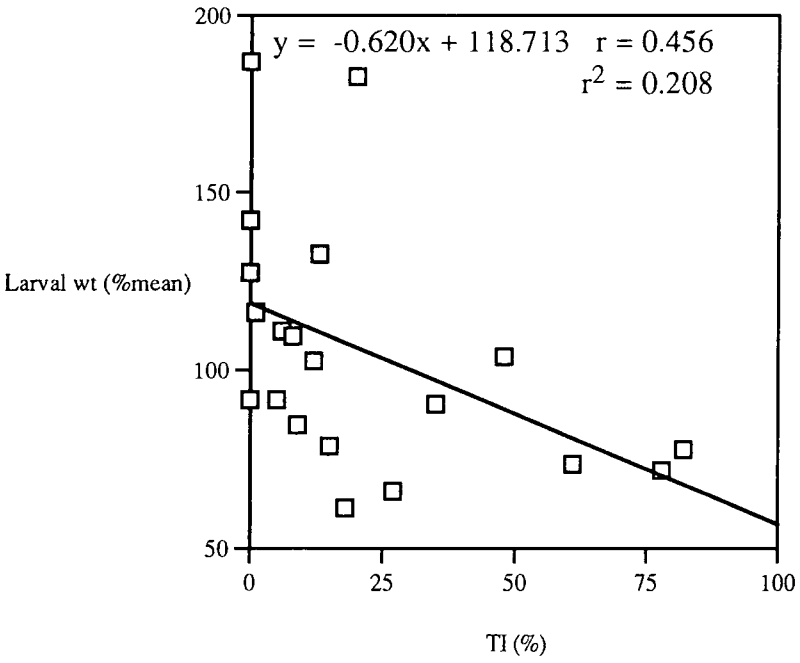
In order to rescue some information from these bioassays, the mean larval weight and mean area eaten of the replicates of each line were compared with the mean for all tissue cultured lines (excluding the positive controls) within a bioassay (See figure 48). Those lines performing better than one standard deviation from the mean; 3-29, 3-41, 4-62, 5-31, 5-40, 6-81 were identified as worthy of further testing. Note that a number of lines performed better than the positive controls; 3-25, 3-29, 4-3, 4-70, 5-14, 6-4, 6-51, 6-81 (15-GNA-35 positive control was not included in all the bioassays).

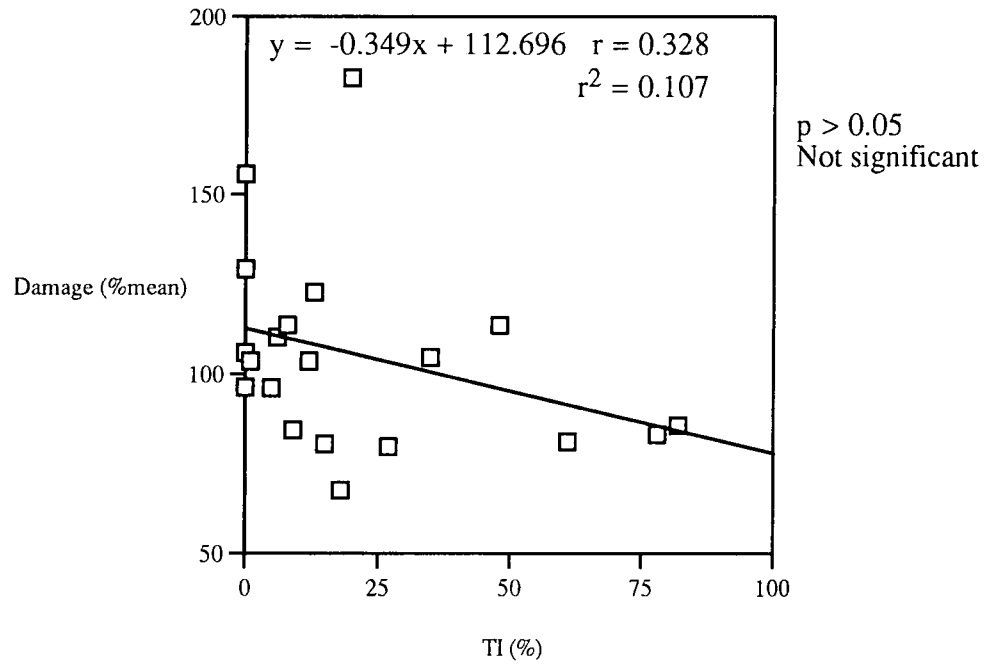
Six of the lines were tested in two separate bioassays. The results as analysed here were generally in good agreement , within ca 20% of one another. However, an exception was line MPI6-50, which in one assay appeared potentially resistant but in the other not, the results differing by >30% (damage) and >50% (larval weight).

Perhaps the most convincing evidence for insecticidal activity of a transgene is provided where one can demonstrate a correlation between expression of the gene and

Plant line	MHTI A RNA expression	Trypsin inhibitory activity	Bioassay number	Larval weight percentage of the mean	Plant damage percentage of the mean
3-16	++		5	113.1	99.1
3-25	+/++		6	70.6 <G	76.6 <G
3-29	+/++		4	53.1 *	66.1 * <G
3-36	++		4	85.2	99.5
3-37	++/+++		2	117.4	107.4
3-41	+/++		7	72.4 *	73.9
3-75	++		2	87.1	91.2
3-93	+/++		8	115.2	115.5
3-99	++		3	93.8	87.0
4-3	++/+++		2/5	89.7/72.2 <G	86.0/70.7 <G
4-44	+/++		8	93.4	92.7
4-62	++		2	67.5 *	78.3 *
4-70	++		5	68.4 <G	72.1 <G
4-74	+/++		8	107.4	94.8
4-85	++		1	96.4	93.8
4-98	+/++		8	110.4	115.5
4-124	+++		1	125.5 *	107.7
4-127	++		4	132.8 *	140.2 *
5-14	+/++	78	5	71.9 <G	83.3 <G
5-31	++	18	3	61.3 *	67.6 *
5-40	+/++	61	8	73.7 *	81.3 *
5-43	++		3	124.3 *	119.0
5-50	+/++	15	7	78.7	80.6
5-54	+++	82	2/4	92.7/77.9	87.1/84.5 <G
5-70	++	0	5	142.0 *	129.3
5-84	++	0	4	127.4	105.9
5-93	+/++	8	7	109.7	113.7
5-96	++	0	4	91.7	96.4
5-101	+/++	0	6	186.7 *	155.8 *
5-105	++	20	5	182.4 *	182.8 *
6-2	+++	13	2	132.6 *	122.8 *
6-4	+/++	9	5	84.7	84.4 <G
6-10	++/+++		2/5	99.5/91.0	111.0/109.8
6-38	++	5	3	91.6	96.2
6-50	++/+++	1	1/5	68.7/116.3 *	75.4 <G/103.9
6-51	+/++	27	6	66.0 <G	79.8 <G
6-57	++	48	3	103.9	113.7
6-68	++		4	148.4 *	115.0
6-70	+/++		7	112.9	136.0 *
6-81	++		5	58.0 <G	64.0 <G
6-92	+/++		6	89.6	98.8
6-106	•		6	87.6	89.0
6-111	+/++		7	127.0	94.8
Control	-	11/0/7/5	1/3	93.7/124.8	116.5/103.9

Figure 48. Summary of the results of the insect bioassays carried out at Horticulture Research International Littlehampton. Results marked * are more than 1 standard deviation from the assay mean. Results marked <G mean the experimental plant performed better than the 15-GNA-35 positive control plant (which expresses snowdrop lectin) where included.





$H_0 : \rho = 0 \quad r_{crit} = 0.444 \; (df = 18)$

Figure 49. Correlation between mean larval weight and trypsin inhibitory activity and mean leaf damage and trypsin inhibitory activity for all suitably assayed plants.

protective effect. All the lines selected for bioassay here were potentially high expressers and there is clearly no significant association between performance in bioassay and measured RNA expression or construct type. However, amazingly in view of the experimental deficiencies a statistically significant correlation (p H_0: $\rho = 0 < 0.05$) was found between larval weight (% assay mean) and level of trypsin inhibitor activity (figure 49)!

The potential of MHTI A to control caterpillars, including species other than *Manduca* , was further demonstrated by the leaf disc assays. The purified *E. coli* expressed MHTI A protein reproducibly gave significant protection for tobacco leaf discs against *Manduca sexta* . This protection was also observed though more weakly when the protein had been diluted, heat treated at 80°C for 10 minutes, or dialysed. Unfortunately no data on the amount of protein used is given. MHTI A was also protective, presumably at the same dosage, on cabbage leaf discs against *Lacanobia oleracia* and in a single assay against *Pieris brassicae* .

Manduca and *Diabrotica* haemolymph trypsin inhibitor cDNAs by 3' RACE PCR

RACE (rapid amplification of cDNA ends) is a technique where by using a single oligo specific to a given protein, a fragment of that protein's cDNA can be generated by PCR; the other primer is a general oligo (Frohman 1990). For 3' RACE this was originally the polyA tail i.e. amplification took place using as primers the specific oligo and oligo dT. This was improved upon by using a hybrid cDNA primer consisting of a run of oligo dT with an adapter primer on the end and performing RACE with the specific primer and an oligo of this adapter region of the cDNA primer. This allowed the melting points of the two oligos to be matched and avoided the artifacts generated by oligo dT binding to A rich regions.

The technique is performed starting with mRNA from which first strand cDNA is synthesised, the RNA and excess oligo dT etcetera is removed and RACE PCR performed on the sscDNA. After the first cycle of PCR the only DNA molecules present with polyA tails are those PCR transcripts synthesised from the specific primer and hence only these will further amplify from the general primer. The sscDNA is all complementary strand i.e. ends in polyT - adapter primer to which adapter primer will not bind. Any remaining mRNA does not function as a template for the PCR enzyme Taq polymerase (although

there are thermostable reverse transcriptases available).

The plasmid rescue procedure for λZAP II cDNA libraries works by the *in vitro* excision and recircularisation of the cloned insert and parts of the vector either side of the insert which together carry the f1 phage origin of replication, one part carrying the site of initiation for DNA synthesis the other the site of termination. Bacteriophage f1 is a filamentous phage with a single stranded DNA genome which is packaged in only one sense in the phage particle but in *E. coli* has a double stranded replicative form which is equivalent to a plasmid (pSK- in fact) and is usually treated as such (ZAP cDNA synthesis kit Stratagene). The λZAP phage requires a helper phage to supply various necessary proteins to excise and replicate as an f1 phage and in their absence remains as phage λ (or the replicative form pSK) quite stably.

The net result is that λZAP phage can be rescued partway to packaged f1 type phage particles. In this form almost all the DNA present is a specific single strand. There will be small amounts of λ DNA which will be double stranded but there will be vastly more single stranded DNA. In λZAP this single strand form, combined with the directional cloning of the kit, is specifically the non coding strand of the cDNA and an efficient protocol to produce this antisense sscDNA is given in the manual for use in making subtractive libraries.

This antisense strand cDNA, i.e. the same strand as first strand cDNA, is an ideal substrate for 3' RACE. The general primer must be chosen to match the sequence on the Xho I side of multipurpose cloning site and to read into the insert. This requirement is ideally met by the standard forwards DNA sequencing primers M13 - 20, T7 and SKT7M (see figure 50). A set of primers was also available which had been successfully used to sequence cloned cDNAs where the polyA tail was too long to read through from the normal primers or from vectors where primers were not available (Applied Biosystems); these consisted of 25 Ts then either an A, a C, or a G, i.e.a slightly more selective version of oligo dT.

A cDNA library in λZAP II of whole *Diabrotica undecimpunctata* larvae, prepared for the microsomal aminopeptidase project (see microsomal aminopeptidase results chapter), and the *Manduca* cDNA library were available and mass rescues of both were performed. In order that the rescued ssDNA was representative of the library an amount of λ phage equivalent to twenty times the library was rescued. The rescues were performed overnight to give the maximum f1 phage titre and hence yield of DNA.

```
                    Xho I pSK....                          T7 primer
polyA AAAAAAAAAACTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAGTGAG 690
                                                SKT7M primer


pSK   TCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAA                 776
                          M13 -20 primer
```

SKT7M primer	CTATAGGGCGAATTGGGT
T7 primer	AATACGACTCACTATAG
M13 -20 primer	GTAAAACGACGGCCAGT

The arrangement of the sequencing primers on pSK on the side of the multipurpose cloning site corresponding to the 3' end of a rescued λZAP clone. The "coding strand" is shown, the primers are in the opposite sense.

```
                    *  *  *  *  *  *

     Y  G  G  C  Q  G  N
TACGGIGGITGCCAAGGIAA
  T        T  G
```

Oligo #1108 made in sense to the haemolymph trypsin inhibitor conserved protein sequence for the 3' RACE experiments.

Figure 50.

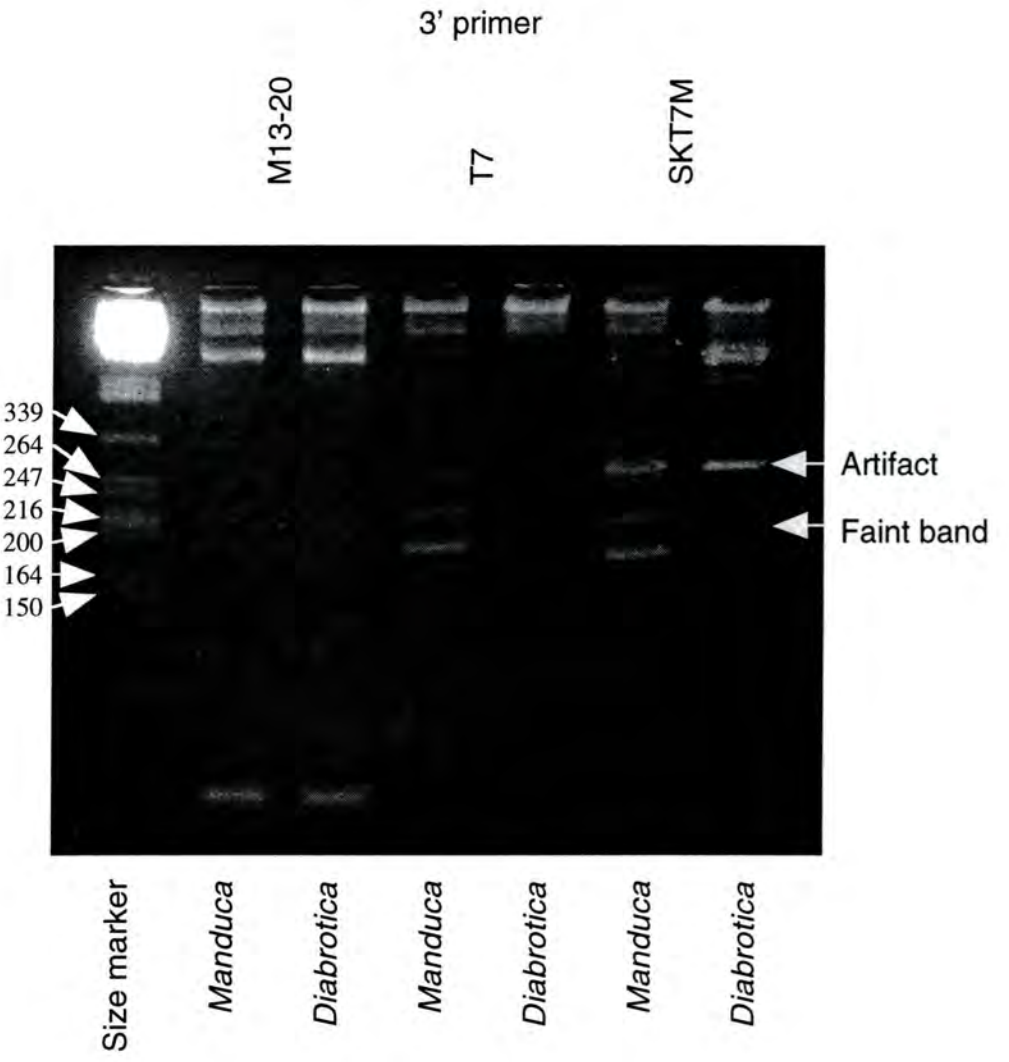


Figure 51. Gel of the 3' RACE PCR reactions on *Diabrotica* and *Manduca* single strand rescued cDNA libraries with the insect haemolymph trypsin inhibitor conserved region oligo #1108 as the specific 5' primer. The artifact band in the SKT7M tracks is marked as is the faint band in the *Diabrotica* SKT7M track. Three bands in the *Manduca* SKT7M track also occur in the *Manduca* T7 track at the expected slightly larger size.

Maintaining the representation of clones in the library was unnecessary, only some target DNA from each line need be present not a representative amount. DNA was prepared from the phage preparations by the standard M13 method of PEG precipitation after growing the phage and killing the *E. coli* with heat (see methods).

A new oligo #1108 was synthesised in the coding sense to the conserved sequence YGGCQGN, this was an 8 fold redundant 20 mer with 3 inosines and with a calculated melting point of 56°C, although this is probably an overestimate. This compares reasonably well with the available sequencing oligos which are 17 mers, except SKT7M which is an 18 mer, with no redundancy and have melting points of 54°C for SKT7M = oligo #788, 52°C for M13 - 20 = oligo #980 and 46°C for T7 = oligo #437. (figure 50)

A number of PCR experiments were run using each of the general primers on both of the rescued cDNA libraries. The PCR conditions were; hot start, 35 cycles of amplification and an annealing temperature of 50°C. It was found that the M13 - 20 primer which was the most GC rich, especially near its 3' end from which extension occurs, gave a set of bands, when run on gel, which were indistinguishable for either of the template DNAs. This was interpreted as a set of artifacts which were associated with the vector, hence the repeat of the pattern with template from differing sources, and were not investigated further.

The T7 primer also gave a number of non specific bands on gel and nothing else for *Diabrotica,* but for *Manduca* three bands in the expected size range (150 - 350 bp) were observed at ~170, ~190 and ~225 bp. (figure 51) The SKT7M primer, besides producing a number of artifacts also gave three bands from *Manduca* which were slightly larger than those obtained using the T7 primer at ~180, ~200 and ~240 bp. This was precisely the pattern expected for PCR products whose 3' primer was slightly further down stream, and strongly suggests that they should indeed be cDNA 3' fragments. There was also a ~200 bp band, which is in the expected size range, weakly visible in the *Diabrotica* track from the SKT7M primer.

A test experiment was also run using the 25TsA primer, which will fit p256C7 a MHTI A cDNA clone, on the rescued *Manduca* library ssDNA and also on the unrescued library. This gave no specific bands on gel, only a continuous streak of DNA and in view of this failure this approach was discontinued.

An attempt was also made to PCR directly from the unrescued libraries using SKT7M and T7 as 3' primers. For *Manduca* these gave the same artifacts as were

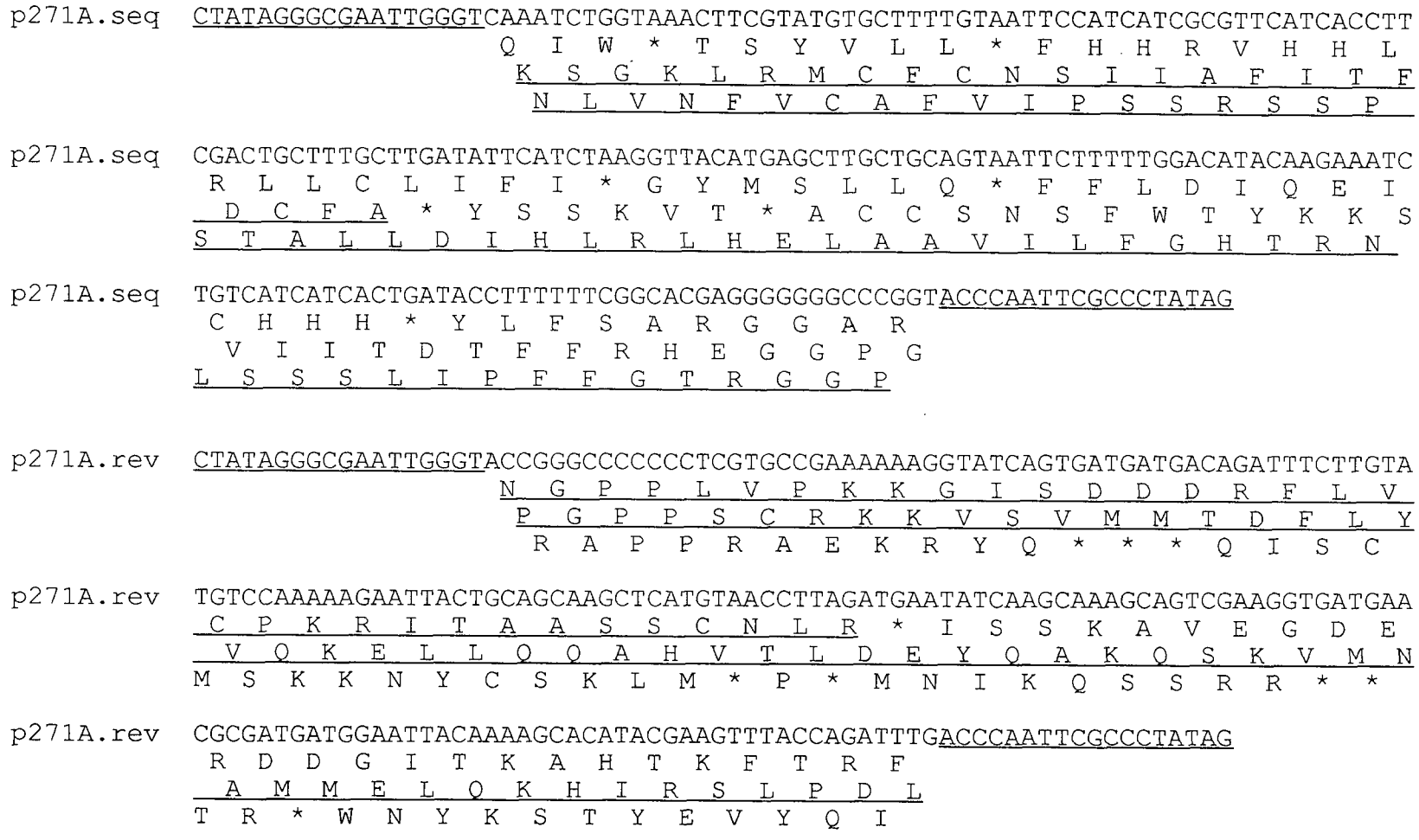


Figure 52. The DNA sequence and its translation in all reading frames of the p271A clones. The DNA sequence underlined is the SKT7M PCR primer. The protein sequences underlined are the possible ORFs.

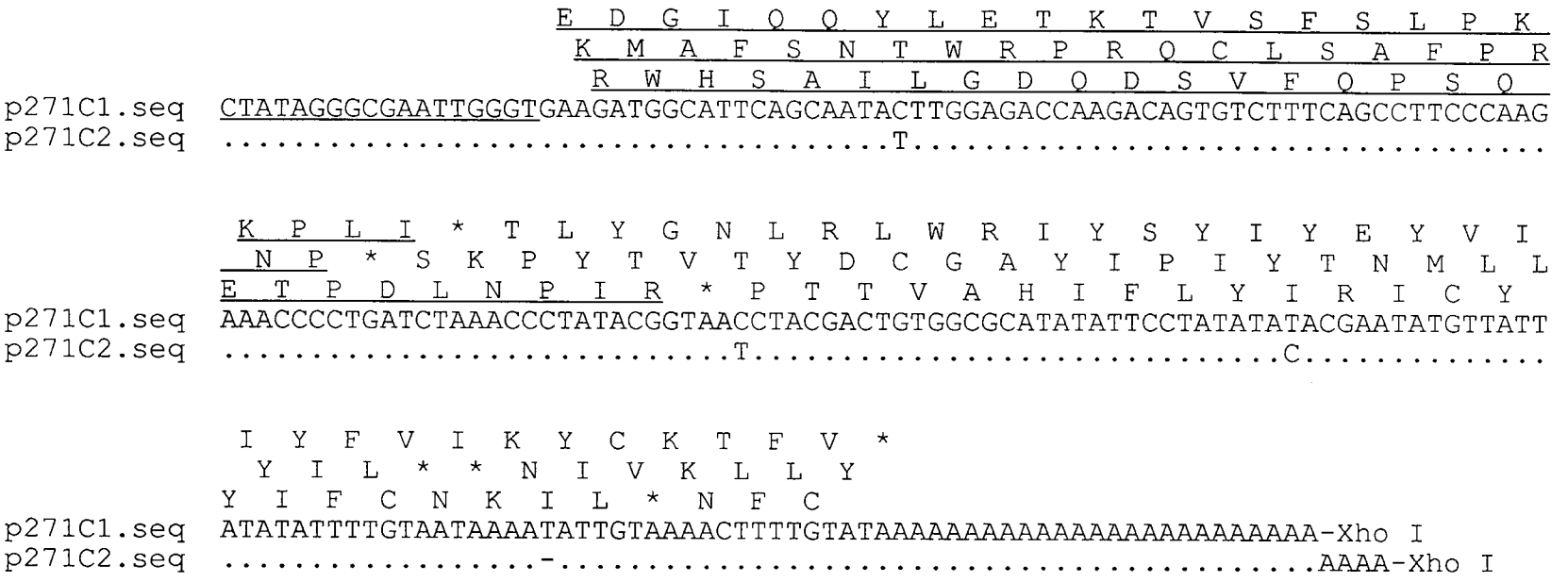


Figure 53. The DNA sequence of p271D1 and its translation in the three forwards reading frames. The DNA sequence underlined is the SKT7M PCR primer . The protein sequences underlined are the possible ORFs. Also shown are the sequence differences observed in p271C2.

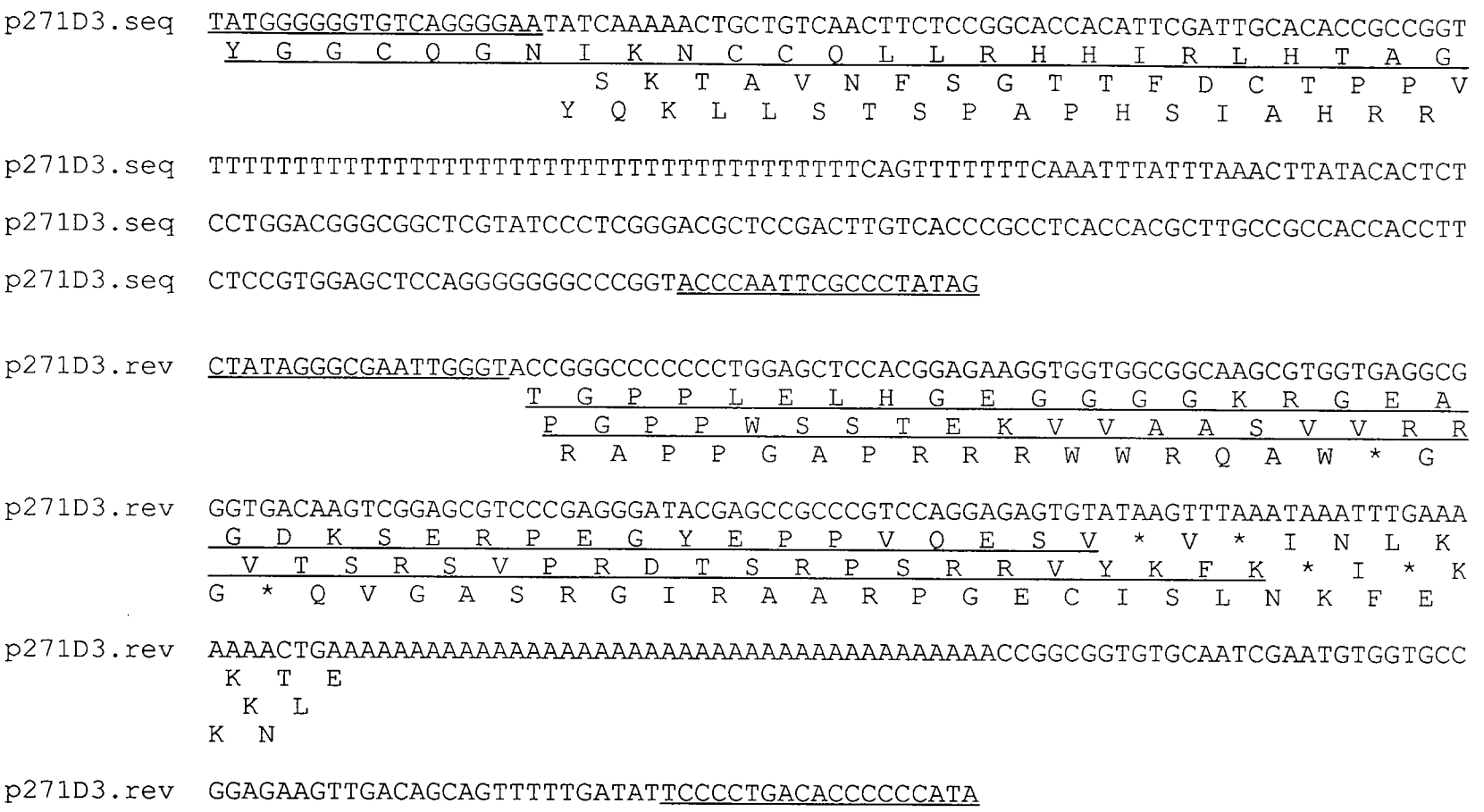


Figure 54. The DNA sequence of p271D3 and its translation in all possible reading frames. The DNA sequences underlined are the PCR primers. The protein sequences underlined are the possible ORFs which were used to search the protein databank.

observed for the rescued single strand library and showed some of the amplified bands but much fainter and less cleanly. For *Diabrotica* some very faint bands were observed in the expected size range for the T7 primer but there were no equivalent bands from the SKT7M primer and these were not investigated further.

The DNA fragments obtained using the SKT7M primer were cloned, the 200 bp band from *Diabrotica* and the three bands from *Manduca* of 180, 200 and 240 bp. The plasmids resulting from this were called p271A for the *Diabrotica* fragment, and p271B, C and D for the *Manduca* fragments in ascending size order. These were DNA sequenced to characterise them.

The p271A *Diabrotica* clones were all identical and obvious artifacts as they had the SKT7M primer at both ends, however they may derive from genuine cDNAs for, despite having no polyA tail, they do have a damaged Xho I site reading through the pSK- cloning site into the 3' SKT7M primer. There were four open reading frames two of which ran completely through the sequence, one forwards and one reverse, where as the other two run in from the ends to stop codons. All were used to run fasta searches of the protein data base at Daresbury but no significant matches were found. (figure 52)

The p271B *Manduca* clones were all identical and exactly matched the MHTI A cDNA p256C7 except at the sites of degeneracy in the conserved sequence oligo #1108 and in having a polyA tail two bases longer. Isolating clones for MHTI A in this way demonstrates that the method works.

The p271C and p271D1 *Manduca* clones were very similar and obvious artifacts as they had the SKT7M primer at both ends. They were however fragments of genuine cDNAs with a good polyA tail at the 3' end ranging from 23 to 29 bases and a Xho I site reading through the pSK- cloning site to the SKT7M primer. All four PCR products represent different individual clones from the cDNA library as they have different length polyA tails and p271C2 has a number of sequence differences mainly of the C to T PCR error type and a one base deletion near the polyA tail and hence is from a cDNA of either a gene closely homologous to that from which the rest originated, or a different allele of the same gene. There are three open reading frames reading in from the 5' end and these were used to run a fasta search of the protein data base at Daresbury but no significant matches were found. (figure 53)

The p271D3 *Manduca* clone was also an artifact although it had amplified from both primers. (figure 54) The 3' end was correct from the SKT7M primer region

Alignment of p271D3 ORF with partial protein sequence of genome polyprotein from apple stem grooving virus, Swissprot POLR ASGVP, part of the RNA directed RNA polymerase (EC 2.7.7.48) (37.5% identity in 40 aa overlap)

```
                                                  10        20        30
p271D3                                  FSFFQI-YLNLYTLLDGRLVSLGTLRLVTRLTT
                                           ||: ||:|   |: || ||:::|::: : |
polr_a SDYTAFDSSQDHVILAFEMALLQYLGVSKEFQLDYLRLKLTLGCRLGSLAIMRFTGEFCT
          1370      1380      1390      1400      1410      1420

            40
p271D3 LAATTFSVEL
       :  :||:  |
polr_a FLFNTFANMLFTQLKYKIDPRRHRILFAGDDMCSLSSLKRRRGERATRLMKSFSLTAVEE
          1430      1440      1450      1460      1470      1480
```

Figure 55.

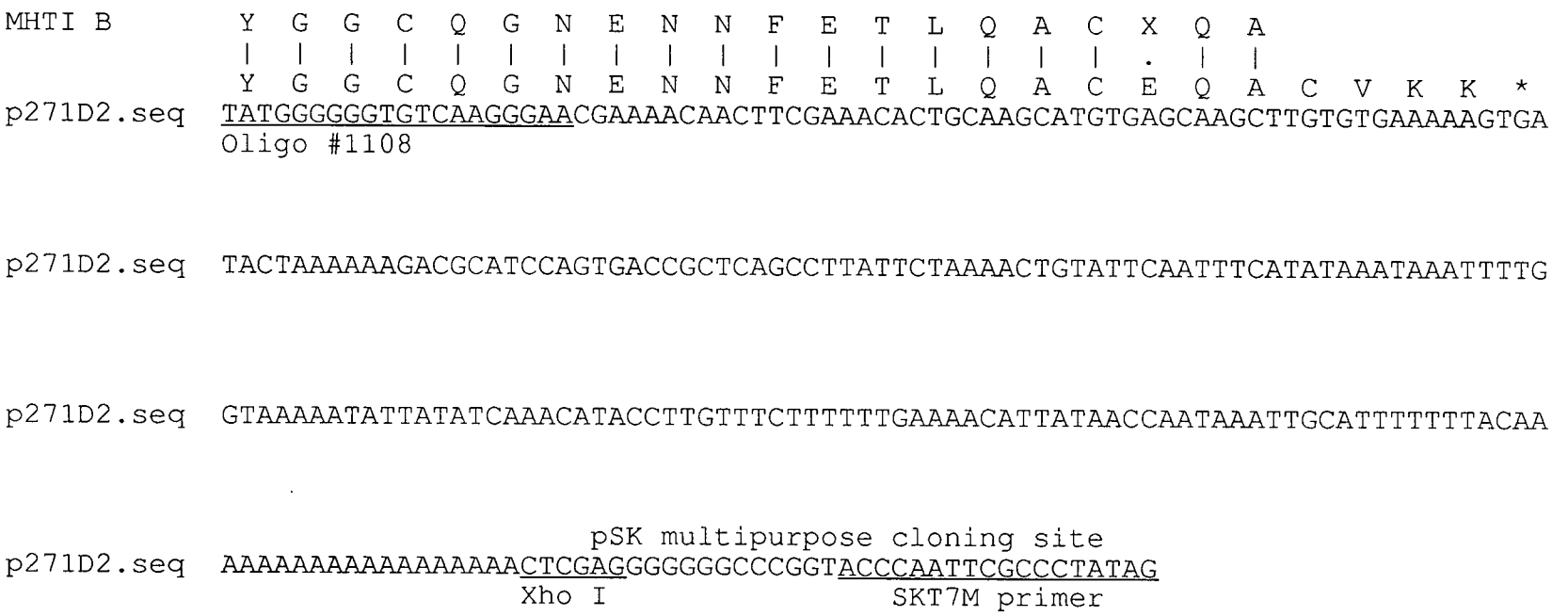


Figure 56. The DNA sequence of p271D2 and its translation in the MHTI B reading frame. The DNA sequences underlined are the PCR primers and the 3' Xho I site.

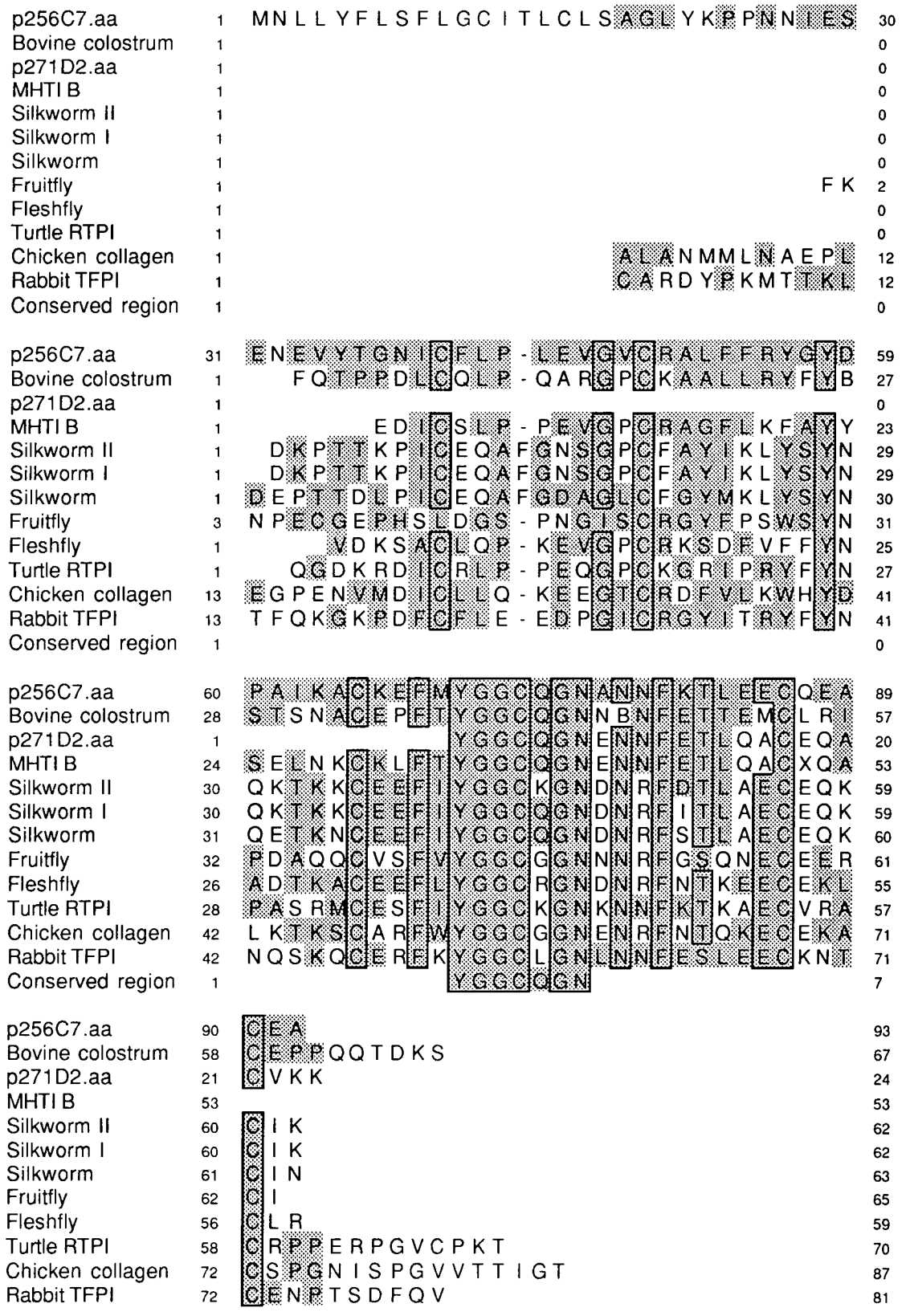


Figure 57. The alignment of the predicted protein sequence of p271D2, which is a partial cDNA for MHTI B with some homologous sequences from various sources. p256C7.aa is the predicted protein sequence of a near full length cDNA for MHTI A.

through the pSK- cloning site to the Xho I site which was damaged. After this it went into a run of sequence then a polyT run then another short run of sequence to the 5' end which was oligo #1108 and this was interpreted as a fragment of two cDNAs fused together. The sequence reading from the MHTI consensus primer shows no homology with any of the MHTI homologous proteins and scores no significant matches when used to run a fasta search of the protein data base at Daresbury. The longer more 3' region of coding sequence contained two open reading frames which were also used to run fasta searches. (figure 55) One showed some similarity to a region of a genome polyprotein from apple stem grooving virus, a single strand RNA non enveloped virus (Yoshiwaki *et al.* 1992), which is part of the active site of the RNA directed RNA polymerase. The region of homology is not highly similar, 37.5% identity in a 40 amino acid overlap, but does extend over nearly all the 43 amino acid ORF. However neither of the cysteins in the polyprotein are present in the p271D3 ORF or are conservatively substituted and the significance of this homology would require more sequence to establish. p271D3 was investigated no further.

The p271D2 *Manduca* clone was an exact match to the known partial protein sequence of MHTI B with the undetermined amino acid being shown to be glutamic acid and the sequence being extended a further four amino acids to a stop codon. (figure 56) This was in good agreement with the other insect TI protein sequences especially the completely conserved cystein residue but extended down stream one more amino acid. (figure 57) There was 130 bp of 3' untranslated sequence, longer than that seen for the other *Manduca* TI cDNAs with which there was no significant homology. The polyA tail was 19 bp in length and was 15 bp down stream of a consensus polyadenylation signal. A second consensus polyadenylation signal about 55 bp from the stop codon, a position roughly the same as in the other *Manduca* TI cDNAs, was unused though as only a single clone was isolated no conclusions can be drawn from this. No further work was done with this but screening the *Manduca* cDNA library with p271D2 to obtain a more full length clone of MHTI B should be straight forward.

This work has, therefore, produced coding sequences for both the previously described *Manduca* haemolymph trypsin inhibitors, MHTI A and MHTI B, as well as the previously unknown MTI C.

Diabrotica haemolymph trypsin inhibitor cDNAs by a nested 3' RACE PCR strategy

The conclusions drawn from the above sequences and the results of the RACE PRCs with the various primers were that; amplification of anything from *Diabrotica* with the conserved sequence oligo is very weak, which suggests that it does not match the *Diabrotica* HTI well, and hence a lower annealing temperature is required, or that *Diabrotica* does not express an equivalent protein. All of the artifacts sequenced, but one, had the SKT7M primer at both ends which suggests that the primers are not well matched, i.e. that the SKT7M primer has too high a melting point, and hence that the T7 primer which has a lower melting point would be less likely to generate artifacts.

In order to increase amplification and improve specificity, a partially-nested PCR strategy was devised. The first step used the T7 primer and the conserved sequence oligo both with and without Enhance PM, a hot start protocol was used to reduce artifacts, and an annealing temperature of 45°C was chosen to allow annealing of the specific primer to the supposedly divergent *Diabrotica* HTI cDNA. Then samples of these were used as templates in fresh PCR reactions using the conserved region oligo and SKT7M as primers and annealing at 50°C, again with and without Enhance PM and with hot start.

The principal behind this approach is that the T7 primer fits further away from the cloning site in pSK- than the SKT7M primer, and hence only molecules that amplified correctly from the T7 primer site will read through the SKT7M primer site into the cDNA and will amplify in the second round of PCR. Any artifacts amplified from a site annealing to the T7 primer other than in the correct position will not carry the SKT7M primer site and will fail to amplify in the second round. In fact the primers overlap by 7 bp but the important factor in PCR is that the 3' end of the primer should fit well, and in particular the last two bases from which the polymerase extends should fit exactly and hence amplification will not occur from this short region. Ideally a second more 3' primer should be used in the second amplification at the other end as well, but there is insufficient, strongly-conserved protein sequence to design one.

The result from this was that four bands were seen on gel in the expected size range (150 - 350 bp or there abouts) from the reaction with Enhance PM after the first PCR, and that after the second PCR a fifth band was visible and the original bands were ~15 bp smaller, as would be predicted. No bands were seen on gel from the reactions without Enhance PM other than at high molecular weight. These bands were cloned and sequenced as follows; p271E - ~360 bp, p271F - ~255 bp, p271G - ~240 bp the extra band after the

```
p271E1.seq  TATGGGGGGTGTCAGGGGAATTATTTTCATCGGAAGGAGTTTAGAAGGAAAGATGGTGTTCTGACTATTTTAGTT
            Y  G  G  C  Q  G  N  Y  F  H  R  K  E  F  R  R  K  D  G  V  L  T  I  L  V
                                    I  I  F  I  G  R  S  L  E  G  K  M  V  F  *  L  F  *  F
                                    L  F  S  S  E  G  V  *  K  E  R  W  C  S  D  Y  F  S

p271E1.seq  CTTTTTTTCAATTTAATTTAAAGTTTAATCCCATAAATTATTATTATTATGATTGAGAAACTTTTTCCATTGTCA
            L  F  F  N  L  I  *  S  L  I  P  *  I  I  I  I  M  I  E  K  L  F  P  L  S
              F  F  S  I  *  F  K  V  *  S  H  K  L  L  L  L  *  L  R  N  F  F  H  C  H
            S  F  F  Q  F  N  L  K  F  N  P  I  N  Y  Y  Y  Y  D  *  E  T  F  S  I  V

p271E1.seq  TTTTGGTTAGCGTTCGATTGATTTCTATTGGAGCGAGCTCGTAATGACGTAATTGATATGATTGATACGTACGTA
             F  W  L  A  F  D  *  F  L  L  E  R  A  R  N  D  V  I  D  M  I  D  T  Y  V
              F  G  *  R  S  I  D  F  Y  W  S  E  L  V  M  T  *  L  I  *  L  I  R  T  Y
            I  L  V  S  V  R  L  I  S  I  G  A  S  S  *  *  R  N  *  Y  D  *  Y  V  R

p271E1.seq  TATAGGATAAGTTACTTTTAGTCATATGATATTATTTTTTGTTATGTTTTTGTTGATACTTGATGATGATAGACT
             Y  R  I  S  Y  F  *  S  Y  D  I  I  F  C  Y  V  F  V  D  T  *  *  *  *  T
              I  G  *  V  T  F  S  H  M  I  L  F  F  V  M  F  L  L  I  L  D  D  D  R  L
            I  *  D  K  L  L  L  V  I  *  Y  Y  F  L  L  C  F  C  *  Y  L  M  M  I  D

p271E1.seq  TGATAAAATTAATGTATAAGCTTCTTATTAACCATTAAAGAAGCTCTTGAAGCAAAAAAAAAAAAAAAAAACTC
             *  *  N  *  C  I  S  F  L  L  T  I  K  E  A  L  E  A
              D  K  I  N  V  *  A  S  Y  *  P  L  K  K  L  L  K  Q
            L  I  K  L  M  Y  K  L  L  I  N  H  *  R  S  S  *

p271E1.seq  GAGGGGGGGCCCGGTACCCAATTCGCCCTATAG
```

Figure 58. The DNA sequence of p271E1 and its translation in all three forwards reading frames. The DNA sequences underlined are the PCR primers. The protein sequence underlined is the MHTI reading frame carried on from the consensus oligo #1108 to the next stop codon.

```
p271F16.seq  GGGGGGTGTCAGGGGAATAAAGTTAAATGCCCAAACAGATTCAACATCTCCATATGTAAGAAGTACCAAATCGAT
               G  G  C  Q  G  N  K  V  K  C  P  N  R  F  N  I  S  I  C  K  K  Y  Q  I  D
                                 I  K  L  N  A  Q  T  D  S  T  S  P  Y  V  R  S  T  K  S  M
                                    *  S  *  M  P  K  Q  I  Q  H  L  H  M  *  E  V  P  N  R

p271F16.seq  GTGCTACATAATGATAGATCGAATTTTATCTCCTCTAGACCCTTCATTGTATCCTTTAACTTTAAACTATTACCG
               V  L  H  N  D  R  S  N  F  I  S  S  R  P  F  I  V  S  F  N  F  K  L  L  P
                  C  Y  I  M  I  D  R  I  L  S  P  L  D  P  S  L  Y  P  L  T  L  N  Y  Y  R
               C  A  T  *  *  *  I  E  F  Y  L  L  *  T  L  H  C  I  L  *  L  *  T  I  T

p271F16.seq  TGAAAAGAGAGCATTAAGGGTCCTTCACGCCTACACCGATAGTGTTATCGACAAAAGAATATCTGAAAAAAAAAA
                *  K  E  S  I  K  G  P  S  R  L  H  R  *  C  Y  R  Q  K  N  I  *
                  E  K  R  A  L  R  V  L  H  A  Y  T  D  S  V  I  D  K  R  I  S  E
               V  K  R  E  H  *  G  S  F  T  P  T  P  I  V  L  S  T  K  E  Y  L

p271F16.seq  AAAAAACTCGAGGGGGGGCCCGGTACCCAATTCGCCCTATAG
```

Figure 59. The DNA sequence of p271F16 and its translation in all three forwards reading frames. The DNA sequences underlined are the PCR primers. The protein sequence underlined is the MHTI reading frame carried on from the consensus oligo #1108 to the next stop codon.

```
p271G1.seq   TATGGGGGGTGTCAGGGGAAGGTCTAGGTCAAAAAGTGGTTCTAGATCTAAAAGTCGGTCTAGGTCACGATCAAA
              Y  G  G  C  Q  G  K  V  *  V  K  K  W  F  *  I  *  K  S  V  *  V  T  I  K
                                    R  S  R  S  K  S  G  S  R  S  K  S  R  S  R  S  R  S  K
                                    G  L  G  Q  K  V  V  L  D  L  K  V  G  L  G  H  D  Q

p271G1.seq   ATCAGGAAGTAGATCTAGAAGTGCTTCGAAGTCACCTTGATATAATTGTATTTCATAATTTTAAAATTATGTCTT
              I  R  K  *  I  *  K  C  F  E  V  T  L  I  *  L  Y  F  I  I  L  K  L  C  L
              S  G  S  R  S  R  S  A  S  K  S  P  *  Y  N  C  I  S  *  F  *  N  Y  V  F
             N  Q  E  V  D  L  E  V  L  R  S  H  L  D  I  I  V  F  H  N  F  K  I  M  S

p271G1.seq   TATTGTAGACTATCATTGAAATATATGAATTAATGTTAAAAAAAAAAAAAAAAAAAAACTCGAGGGGGGGCCCGG
              Y  C  R  L  S  L  K  Y  M  N  *  C  *
               I  V  D  Y  H  *  N  I  *  I  N  V
             L  L  *  T  I  I  E  I  Y  E  L  M  L

p271G1.seq   TACCCAATTCGCCCTATAG
```

Figure 60. The DNA sequence of p271G1 and its translation in all three forwards reading frames. The DNA sequences underlined are the PCR primers. The protein sequences underlined are the ORFs used to search the protein database.

Alignment of p271G1 with partial protein sequence of E2 protein from human papillomavirus type 25, Swissprot P36787 (67.7% identity in 31 aa overlap).

```
                                                    10        20        30
p271G1                                          RSRSKSGSRSKSRSRSRSKSGSRSRSASKS
                                                ||:|:| |||:|||| ||:| |||||:|:|
ve2_hp QQTSTKGRRYERRPSSRTRRQQAQARQRRSRSKSRSRSRSQSRSRIRSRSRSRSRSESQS
       240       250       260       270       280       290

p271G1 P
       :
ve2_hp SKRRSRSRSRRKTSATRGRGPGSPTTTTSDRAARSPSTTSSATSQRSQRSRSRAGSSRGG
       300       310       320       330       340       350
```

Alignment of p271G1 with partial protein sequence of sperm specific protein PH1 from *Myrtilus edulis* (blue mussel), Swissprot Q04621 (44.8% identity in 29 aa overlap).

```
                                                    10        20        30
p271G1                                          RSRSKSGSRSKSRSRSRSKSGSRSRSASKS
                                                |:|||: :|:|:|:||:||:::|:::::|:
phi1_m PSPTRRSSKSRSKSRSRSRSASASPGKAAKRARSKTPRRGKKRARSPSKKARRRSRSTKK
               10        20        30        40        50        60

p271G1 P

phi1_m TAAKRRKRSSSPKKRRSAGKRRVRAKKKKKK
               70        80        90
```

Alignment of p271G1 with partial protein sequence of splicing factor SC35 (PR264 protein) from human, Swissprot Q01130 (72.4% identity in 31 aa overlap).

```
                                                    10        20       29
p271G1                                          RSRSKSGSRSKSRSR-SRSKSGSRSRSASK
                                                ||||:| |||:|||| ||||| ||:|| |:
sc35_h HHSRRGPPPRRYGGGGYGRRSRSPRRRRRSRSRSRSRSRSRSRSRYSRSKSRSRTRSRSR
      100       110       120       130       140       150

      30
p271G1 SP
       |:
sc35_h STSKSRSARRSKSKSSSVSRSRSRSRSRSRSRSPPPVSKRESKSRSRSKSPPKSPEEEGA
      160       170       180       190       200       210
```

Alignment of p271G1 with partial protein sequence of serine-arginine protein 55 (SRP%%) from *Drosophilia melanogaster* , Swissprot P26686 (60% identity in 30 aa overlap).

```
                                                    10        20        30
p271G1                                          RSRSKSGSRSKSRSRSRSKSGSRSRSASKS
                                                ||||:| :||:||::|:|:| ||||| |::
sr55_d RRIHLVEDRRGGRSGGGGGSGRGRSRSSSSRSRSRSRRRSRSRRSSHSRSKSRSRSKSRG
       180       190       200       210       220       230

p271G1 P

sr55_d GRSKSKSPVKSRSRSRSRSNKSRDVSKSKSKSHSRTRSRSPKRERDSRSRTRSVSKRESR
       240       250       260       270       280       290
```

Figure 61.

second PCR, p271H - ~180 bp and p271J - ~150 bp.

The p271E plasmids sequenced were identical, apart from a 1 base difference in the lengths of their polyA tails, with the SKT7M primer at one end and the consensus oligo #1108 at the other, and appeared to be derived from a genuine cDNA as there is a 19 or 20 bp polyA tail at the SKT7M primer end. (figure 58) The ORF from the consensus oligo ran for 26 amino acids before meeting a stop codon but showed no homology to the other TIs and so was considered an artifact. Of the other ORFs one was very short and the other ran for 14 amino acids. Both the sequences of any length were used to run a fasta search of the protein data base at Daresbury but no significant matches were found.

The p271F plasmids were also identical, with the SKT7M primer at one end and the consensus oligo #1108 at the other, and appeared to be derived from a genuine cDNA as there was a 16 bp polyA tail at the SKT7M primer end. (figure 59) However, again the sequence beyond the consensus primer reading frame showed no homology to the other TIs and so was considered an artifact. Only this ORF was a possible coding sequence as the other ORF ran on through the polyA tail without first having a stop codon. A fasta search was run using it and this showed some weak similarity to a surface antigen protein from some *Plasmodium* species, 46.9% identity in a 32 amino acid overlap. However from the position of the homology, ~320 amino acids from the 3' end of the *Plasmodium* protein, and the fact that the region of homology did not extend to the 5' end of the p271F sequence, discounting the MHTI consensus primer region, it was concluded that this was not significant.

The p271G plasmids were an obvious artifact as the consensus primer coding sequence ran into stop codon immediately although they were derived from a genuine cDNA with a 21 bp polyA tail and the two PCR primers at the relevant ends. (figure 60) The two ORFs, of 45 and 31 amino acids, were used to run a fasta search and the shorter showed homology to a number of diverse proteins that contained serine argenine (SR) repeat sequences and that are involved in mRNA splicing. (figure 61) One of them, though not the best match, was a *Drosophila melanogaster* nuclear protein (60% in a 30 amino acid overlap) (Roth Zahler and Stolk 1991) and similar proteins would be expected to be found in *Diabrotica* . However, the significance of these homologies is difficult to assess without more sequence data, particularly from a region outside the SR repeats, and no further work was done with these clones.

The p271H plasmids matched the *Manduca* HTI A base for base and must be due

to amplification of a trace cross contamination of the *Diabrotica* template DNA with the *Manduca* template DNA, a testament to the power of PCR.

The p271J plasmids have amplified from the consensus #1108 and SKT7M primers but lack a polyA tail. The coding sequence reading from the consensus oligo runs for 10 amino acids but shows no homology to the other TIs and so was considered an artifact. This and the other two ORFs of any length, which ran in the reverse orientation, were used to run a fasta search but no significant matches were found. (figure 62)

The failure of this method does not preclude the presence of *Diabrotica* TI proteins of the MHTI type being present, as a slightly different forwards oligo may have allowed amplification. However, it at least shows that if *Diabrotica* has TI genes of this type they are divergent at the conserved region for other insects, or are not expressed. The fortuitous isolation of a cDNA for *Diabrotica* cuticle protein (the 5' end of p256C5 see the microsomal aminopeptidase results chapter) and its close homology to cuticle proteins from insects of other orders (50 - 59% over a 65 - 71 amino acid region) suggests that *Diabrotica* proteins are not highly divergent from those of other insects. The range of MHTI like insect proteins for which sequence has been reported (see above) covers the Diptera order as well as the Lepidoptera. It seems reasonable that a class of proteins found throughout the animal kingdom should be present in *Diabrotica* .

If the failure to amplify a haemolymph trypsin inhibitor from *Diabrotica* is because it is not present in this insect then it would argue against the hypothesis that the function of HTI is to control the activity of phenoloxidase (Ramesh, Sugumaran and Mole 1988) since this biological function should occur in *Diabrotica* as well. *Diabrotica* like many other *Coleoptera* has an acid gut with thiol proteases as its major digestive enzymes, although there is about 10% of the total protease activity in *Diabrotica* guts which is inhibitable by soybean Kunitz trypsin inhibitor and cowpea Bowman-Birk trypsin inhibitor, both of which are serine protease inhibitors (Edmonds 1994). Lepidoptera (including *Manduca*) have alkaline guts and use serine proteases as their major digestive enzymes (Wigglesworth 1972). The hypothesis concerning leakage of digestive enzymes would predict that the expression of HTIs in *Diabrotica* would be unnecessary or at much lower levels than in *Manduca* , a possible explanation of the failure to amplify a HTI from *Diabrotica* .

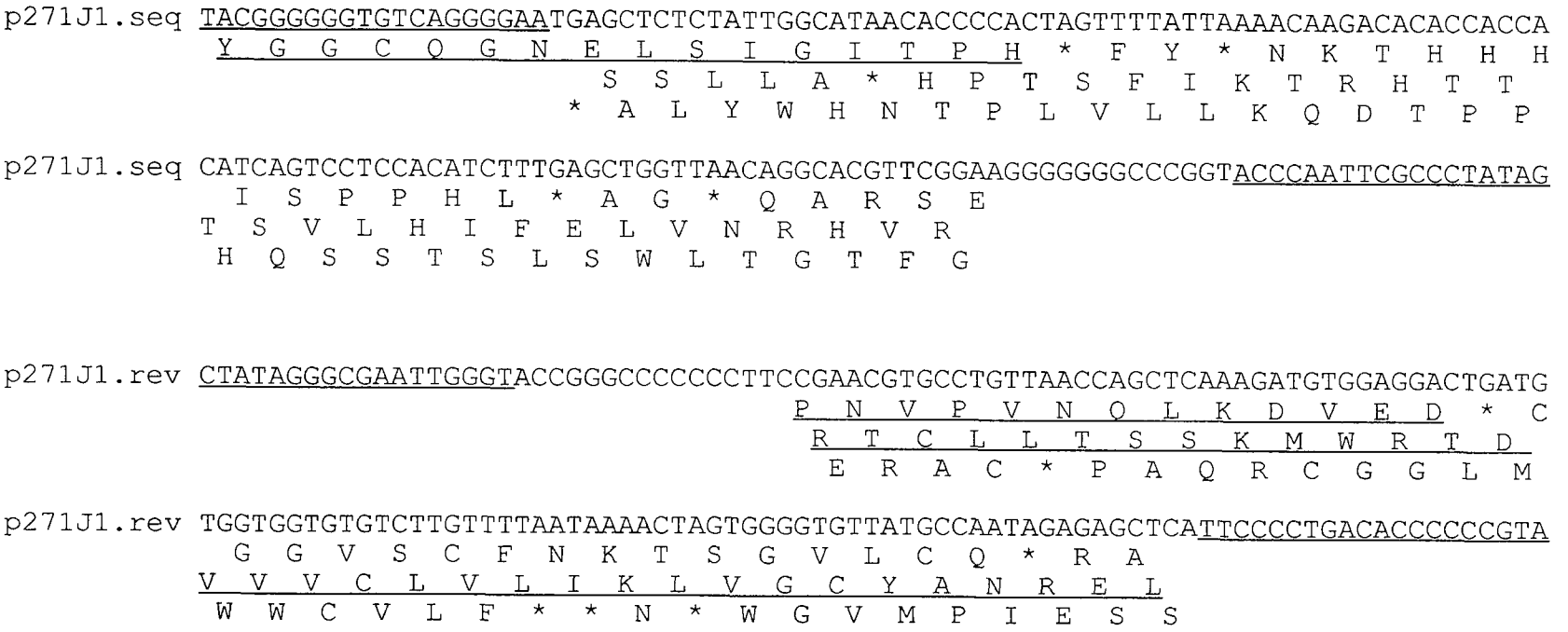


Figure 62. The DNA sequence of p271J1 and its translation in all possible reading frames. The DNA sequences underlined are the PCR primers. The protein sequences underlined are the ORFs used to search the protein database.

Results; microsomal aminopeptidase

Microsomal aminopeptidase (MAP) was selected originally as a handle to use to screen the SCAB library. The protein has a short N-terminal, membrane-spanning, region and a long C-terminal region which contains the enzyme active site and which is accessible to the gut contents. MAPs represent an important type of digestive enzyme, and it has been shown that MAP enzyme activity could be blocked by monoclonal antibody lines (Ashmun, Shapiro and Look 1992). This could be expected to be deleterious to a target insect as the inhibition of other classes of digestive enzyme has been shown to be deleterious to a range of insects (see the introduction; Enzyme inhibitors).

Diabrotica undecimpunctata (Southern Corn Rootworm, as the larva, and Spotted Cucumber Beetle as the adult (Hill 1983) was selected as the target insect primarily on commercial grounds but is also suitable as a target species as it is amenable to bioassay in artificial diets (Edmonds 1994) which could be adapted to allow screening of SCABS in microtitre plates, and is readily reared in large numbers in the laboratory.

Antibodies and gut physiological conditions

Prior to embarking on a major expenditure of resources on the antibody programme, it was considered prudent to establish certain fundamentals. For a strategy involving the feeding of antibodies to insects for the antibodies to affect the insects by binding to proteins in their gut, the antibodies must be active and reasonably stable under gut physiological conditions. In particular they must be active at insect gut pHs which range in the extremes from pH 3.5 for coleoptera to pH 11.0 for Lepidoptera (Wigglesworth 1972). Furthermore they must be either resistant to the proteases they will encounter or be protected by some co-expressed protease inhibitor(s). To this end a series of experiments were performed with various antibody preparations at a range of pH conditions and with *Diabrotica* gut preparations.

Antibodies and gut pH conditions

From previous work carried out in the department, a number of polyclonal antibody preparations raised in rabbit against various pea seed proteins were available. The

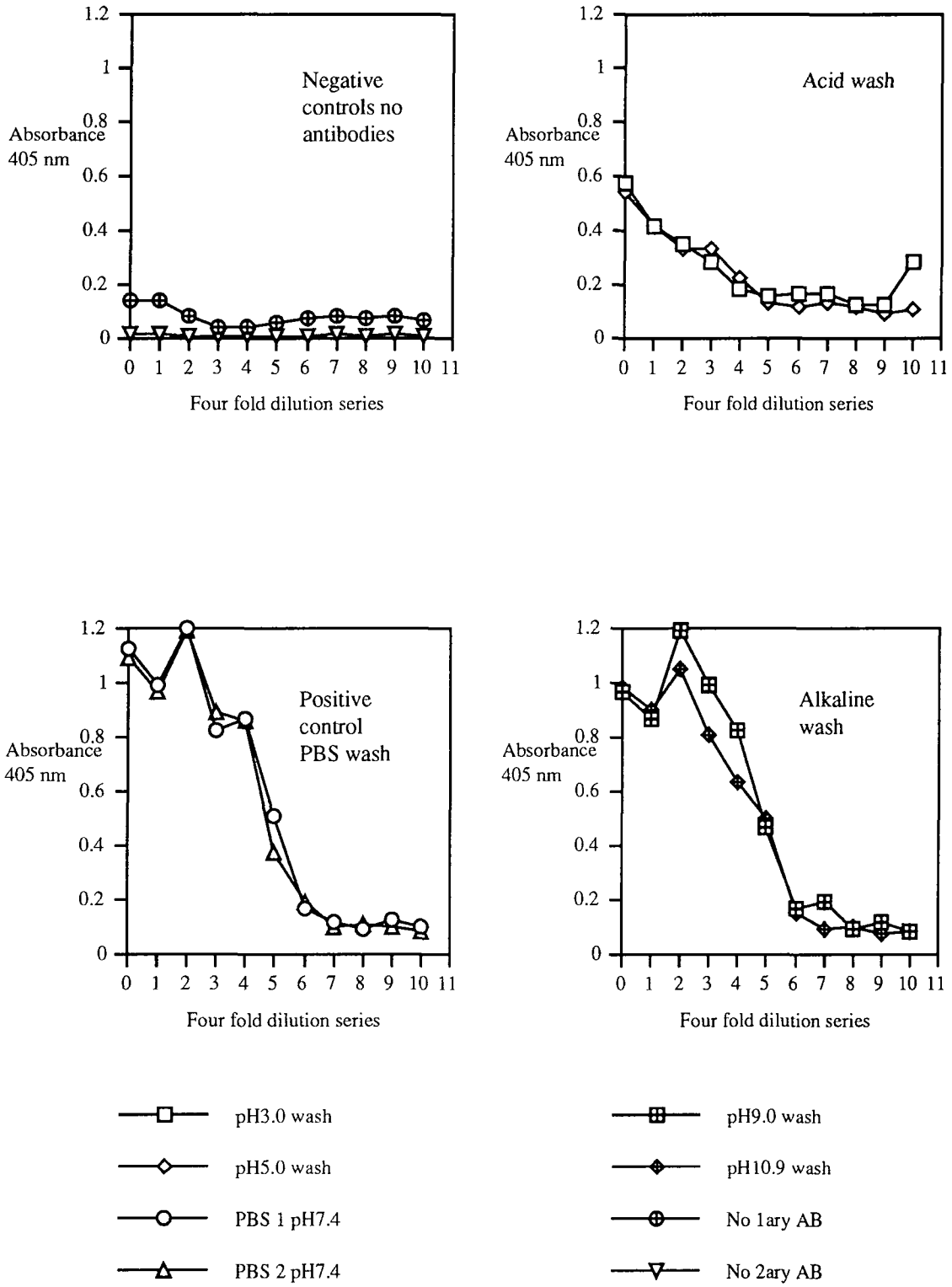


Figure 63. ELISA results from total pea protein bound as normal, incubated and washed in the given buffer then developed as normal with anti-pea legumin antibodies.

fact that the antibodies used were polyclonal was important as what needed to be demonstrated, at least initially, was that binding of at least some members of the population of antibodies occurred under the conditions chosen, i.e. that SCAB lines which are active under these conditions could be isolated in principal. The target protein was total pea seed extract in PBS plus 0.02% sodium azide to prevent bacterial and fungal growth. The other buffers (to provide a range of pHs) were made from standard buffer mixes (McIlvaine 1921; Weast 1968) plus 150 mM sodium chloride and 0.02% Tween 20, also plus 0.02% sodium azide if they were to be stored for any time. The experiments were all done by the ELISA method using peroxidase linked goat anti-rabbit secondary antibodies and ABTS as the detection system (Kirkegaard and Perry Laboratories Inc. Data sheet with their ELISA reagent kit.).

A series of experiments were performed to test whether the choice of pea seed proteins was a good one and whether the standard ELISA system needed to be modified to work reliably at non standard pHs. A times four serial dilution of total pea extract was prepared, bound to the ELISA plate, blocked as normal and washed. Using the chosen pH buffer the wells were then washed again, incubated at 37°C for 1 hour, rewashed and finally washed with PBS plus 0.02% Tween 20 (PBS + T) prior to being put through the normal ELISA process using an affinity purified anti-legumin antibody preparation. (figure 63) The results show clearly that pea seed extract has a very low level of endogenous peroxidase activity in the absence of secondary antibody (no secondary antibody curve) and the secondary antibody binds non-specifically at an acceptably low level in the absence of primary antibody (no primary antibody curve).

The curves for the PBS incubated and alkaline incubated samples are essentially identical indicating that protein is not washed off the ELISA plate and that binding is not affected by exposing the target and blocking proteins to these conditions. The acid incubated samples show decreased sensitivity but no extra non-specific signal; thus the blocking reagent is not being washed off the plate. The target protein, legumin, is known to precipitate from solution in acid conditions (Croy and Gatehouse 1985) presumably due to conformational changes which would explain the observed decreased sensitivity.

The effect of pH on the primary antibody binding was then investigated. A plate was prepared as above then washed once in the given buffer mixes which ranged from pH3.0 to 9.6. The same affinity purified anti-legumin antibody preparation was used as the primary antibody but was diluted in the given buffers and the primary washes were also

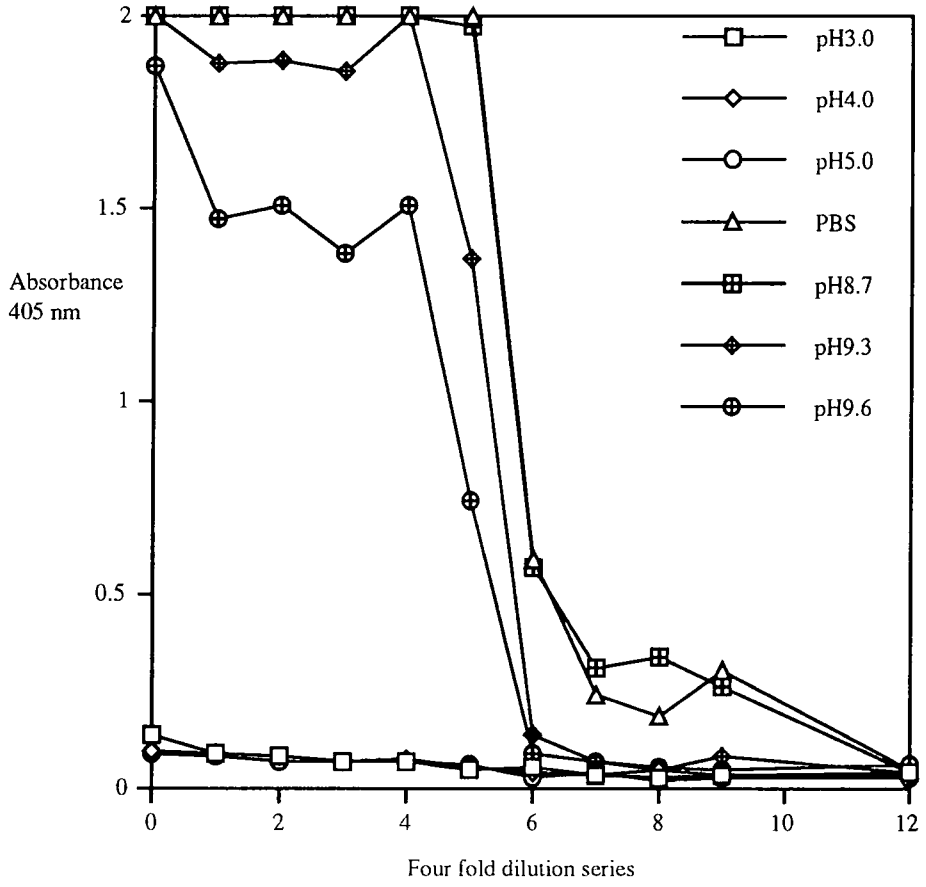


Figure 64. ELISA results from total pèa protein bound as normal, washed and the primary antibody (anti-pea legumin antibodies) bound and washed in the given buffer then developed as normal.

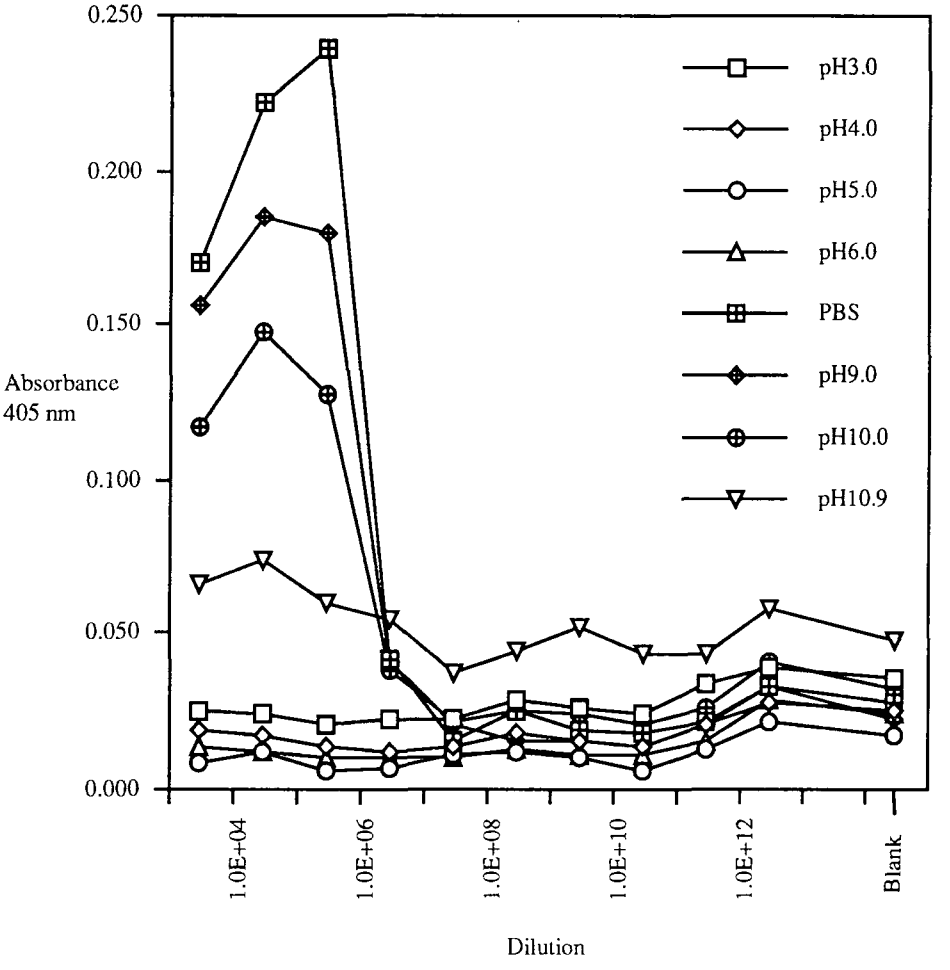


Figure 65. ELISA results from the anti-legumin antibody bound directly to the ELISA plate and blocked. The secondary antibody was added, incubated and washed in the relevant buffers and finally washed with PBS + T before being developed.

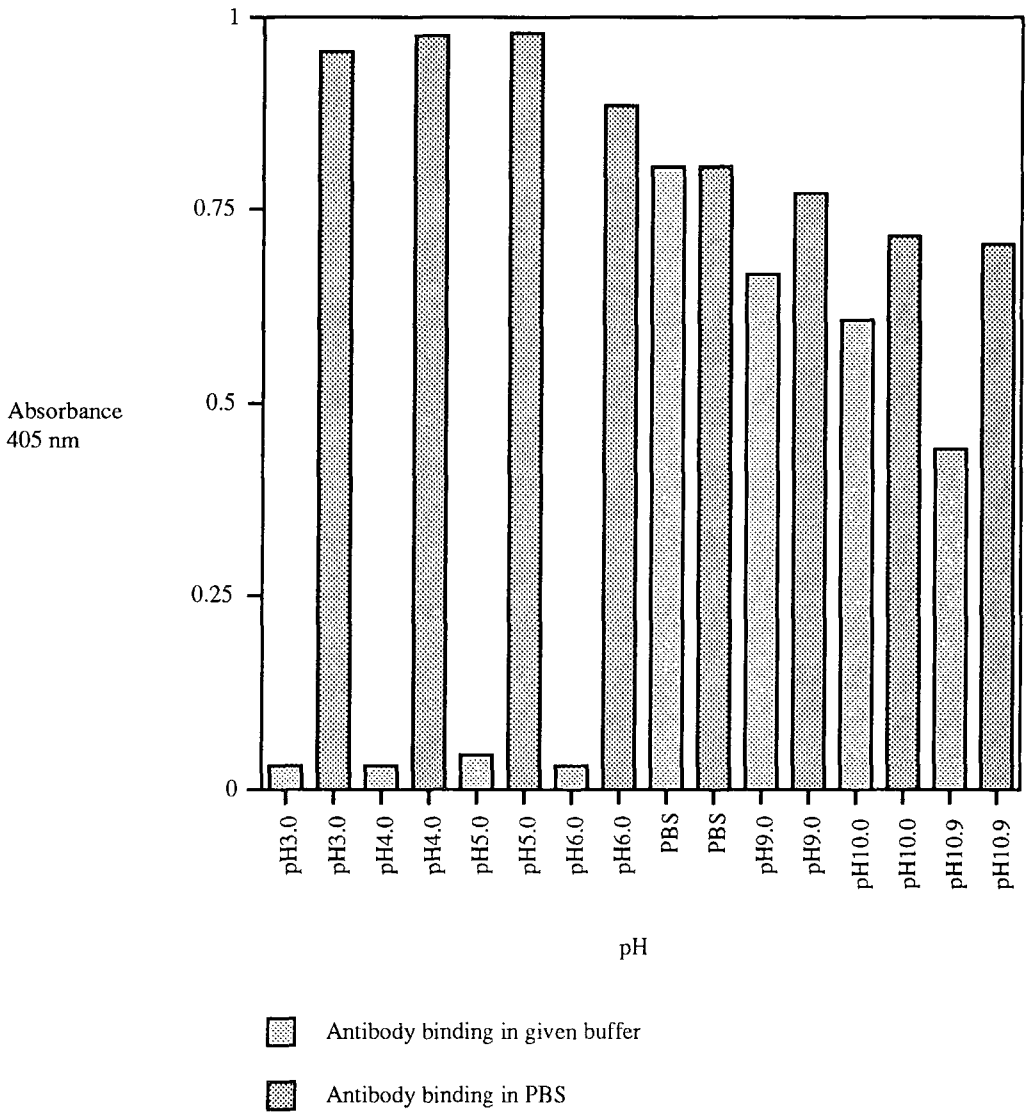


Figure 66. ELISA results from binding the secondary antibody to immobilised primary antibody in the given pH buffer and developing, then washing and binding a secondtime in PBS and developing.

in the test buffers with the final wash in PBS + T in all cases. Other conditions were as before.

A slight decrease in sensitivity was seen with increasing pH (figure 64) which presumably reflects the decreasing proportion of the antibody population active at the higher pHs. No binding at all was detected at acid pHs. This was unexpected until it was realised that affinity purified antibodies are eluted from the affinity column with an acid buffer and hence the purified antibody preparation will lack those members of the original population which do bind in acid conditions.

The experiments shown in figures 65 and 66 were performed to test the hypothesis that antibodies which bind at low pH are removed by affinity purification. The anti-legumin antibody was bound directly to the ELISA plate and blocked and the secondary antibody was added, incubated and washed with the relevant buffers and finally washed with PBS + T before being developed. The results in figure 65 showed that the peroxidase linked goat anti-rabbit also fails to bind in acid conditions, with no signal even at pH 6.0, but binds well in alkaline conditions with a good signal up to pH 10.0 and still some binding at pH 10.9. The secondary antibody was also an affinity purified preparation and this pattern was expected.

The same plate used above, where wells were probed and washed in various pH buffers, was washed with PBS + T and reprobed with the peroxidase linked goat anti-rabbit antibodies under the standard conditions, i.e. in PBS + T, then developed as normal (figure 66). All the wells gave very similar results, with good positive signals under these conditions from the wells originally treated to acid conditions as well as those treated originally to neutral and alkaline conditions. Thus, affinity purified antibodies do not bind in acid conditions and the negative results under acid conditions are not due to some permanent conformational changes or the destruction of the antibody but rather to the removal from the antibody population of that subset which do bind at low pH.

The experiments illustrated in figures 67 and 68 were performed on a four times serial dilution of total pea seed extract as the target proteins. The primary antibodies were rabbit polyclonal, non-affinity purified preparations against the following pea seed proteins; legumin, vicilin, lectin and the major albumins. All of these proteins are present in significant amounts in pea seeds and are readily soluble under the chosen extraction conditions (Croy and Gatehouse 1985). ELISA was performed with the secondary antibody steps in McIlvaine buffer (McIlvaine 1921; Weast 1968) at pH 5.0 (figure 67)

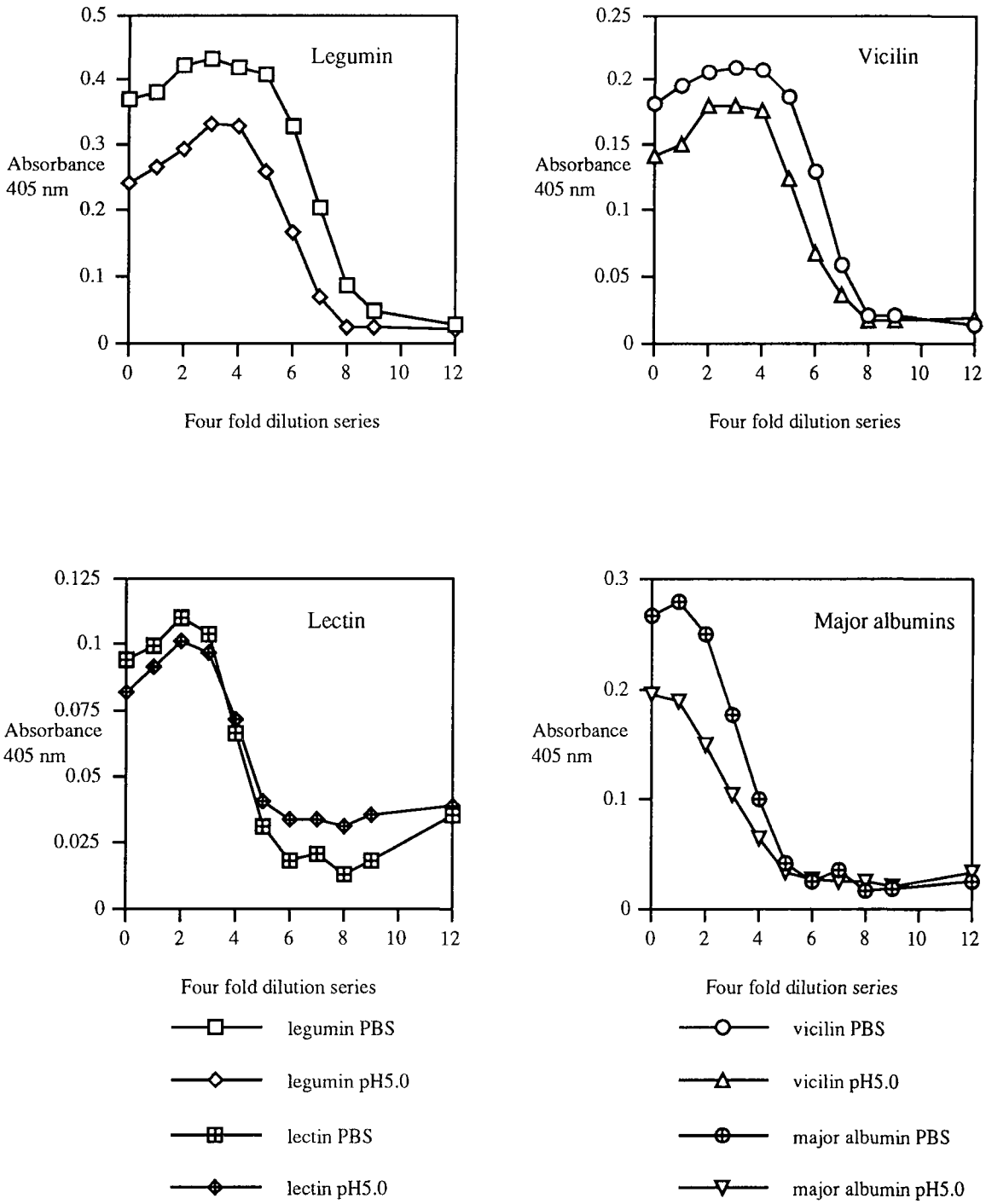


Figure 67. ELISA results from binding and washing various non-affinity purified polyclonal antibodies to immobilised total pea seed protein in pH 5.0 buffer compared with PBS then developing as normal.

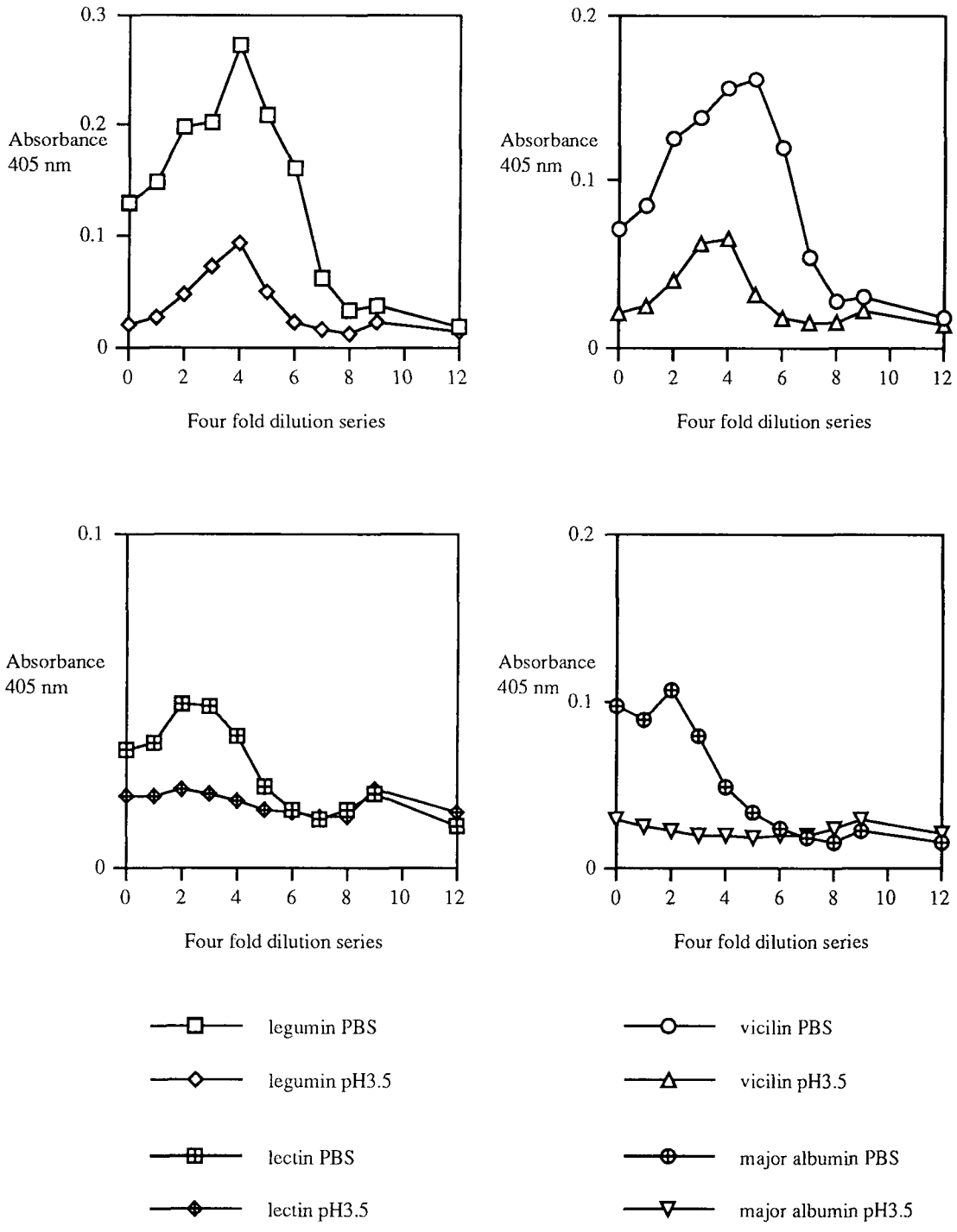


Figure 68. ELISA results from binding and washing various non-affinity purified polyclonal antibodies to immobilised total pea seed protein in pH 3.5 buffer compared with PBS then developing as normal.

and pH 3.5 (figure 68) with salt and azide as above and with PBS pH 7.4 as controls. At pH 5.0 all the antibody preparations show good binding with only a slight decrease in signal relative to the control. At pH 3.5 all the antibody preparations show greatly decreased binding relative to the control but all except possibly the major albumin antibody still show some clearly positive signal.

Thus in a polyclonal antibody preparation there are some members of the population which are able to bind to their target protein at the extreme pHs which occur in insect guts. It may therefore be concluded that SCAB lines, which are active at these extreme pHs, can be isolated.

Antibodies and gut enzymes

Experiments were performed to test the effects of *Diabrotica* gut extracts at pH 5.2 on antibody both bound and in solution. A pH of 5.2 was chosen as this is approximately the natural pH of the *Diabrotica* gut (Edmonds 1994), and the polyclonal antibodies work well at this pH. McIlvaine buffer (McIlvaine 1921, Weast 1968) plus 150mM sodium chloride and 0.02% Tween 20 pH 5.2 was used for all the gut extract incubations and primary antibody steps. The chosen antibody was the non-purified, rabbit anti-legumin preparation at 1 in 3000 dilution as used above and the target protein was total pea seed extract at approximately 1 in 100 dilution which, from the above experiments, gave the best sensitivity (being just at the top end of the linear region of the target protein dilution curve). All wells were blocked for 5 minutes with BSA before adding the secondary antibody to avoid non-specific binding to sites from which protein had been proteolyticly removed.

The gut extracts were made from approximately 4th instar *Diabrotica* larvae which were feeding, had not started to pupate and were about 15 mm long. The total gut was homogenised in water, with 10 µl of the gut preparation equivalent to 1 insect gut plus its contents.

The results (figure 69) show that *Diabrotica* gut extract does not contain significant peroxidase activity (gut extract no AB point) nor does the secondary antibody bind to gut extract (no primary AB points) and hence the other results can be interpreted unambiguously. The control curve, where the target pea seed total protein was incubated for one hour at room temperature with the relevant amount of gut extract plus buffer then

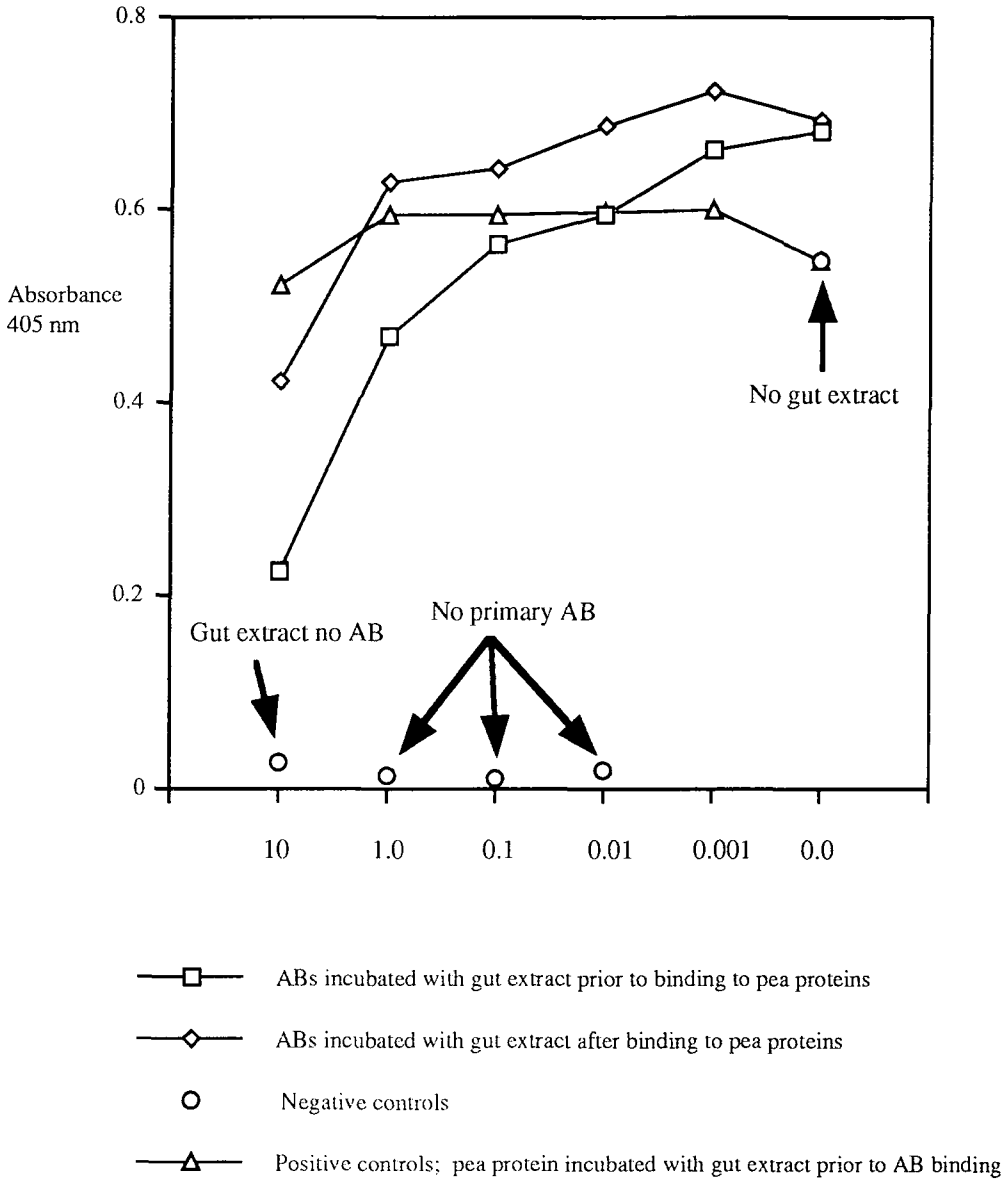


Figure 69. ELISA results showing the effects of *Diabrotica* gut extracts on rabbit anti-legumin antibody and pea seed proteins at pH 5.2

washed before the primary antibody was bound, suggests that exposure of the pea legumin target protein to *Diabrotica* gut extract causes a slight increase in binding at high dilutions of the extract but only a slight decrease in binding for 1 gut equivalent. However these effects are relatively small.

The antibodies were incubated with the gut extracts by two protocols. Preincubation, where the antibodies were incubated with the gut extract in solution for 1 hour at room temperature prior to being bound to the target protein, and post-incubation, where the antibodies were bound, washed and then incubated with the gut extracts. Both methods showed a decrease in signal with increasing gut extract concentration, which was interpreted as proteolysis and hence inactivation of the antibody. The free antibody was more sensitive than the bound antibody to this proteolysis, as would be expected, but in both cases showed reasonable resistance to inactivation with ~30% and ~60% activity respectively still present after 1 hour at the highest gut extract concentration.

Thus, while antibodies are sensitive to *Diabrotica* gut proteolytic enzymes they are not degraded very rapidly. Hence, they would be expected to survive in *Diabrotica* gut long enough to bind their target proteins, especially if they were afforded the additional protection of protease inhibitors.

In view of the stability of antibody activity in insect gut conditions it was decided to proceed with the next stage of the SCAB library production strategy. This involved the isolation of *Diabrotica* microsomal aminopeptidase cDNAs for the eventual production of protein by overexpressing them.

The construction of the *Diabrotica* cDNA library

Diabrotica undecimpunctata larvae were harvested directly into liquid nitrogen at approximately the fourth instar when they were about 15mm long, were actively feeding and had not started to pupate. Those larvae not used immediately were stored at -80°C. Total RNA was prepared by the guanidinium thiocyanate method for tissue (Asubel *et al.* 1992) from whole larvae, approximately 4 gm of larvae yielding 6.1 mg of total RNA which was reasonably pure with an OD_{260} to OD_{280} ratio of 1.7. From this polyA+ RNA was made by the polyAttract system (Promega), 4 mg of total RNA yielding 4 µg of polyA+ RNA.

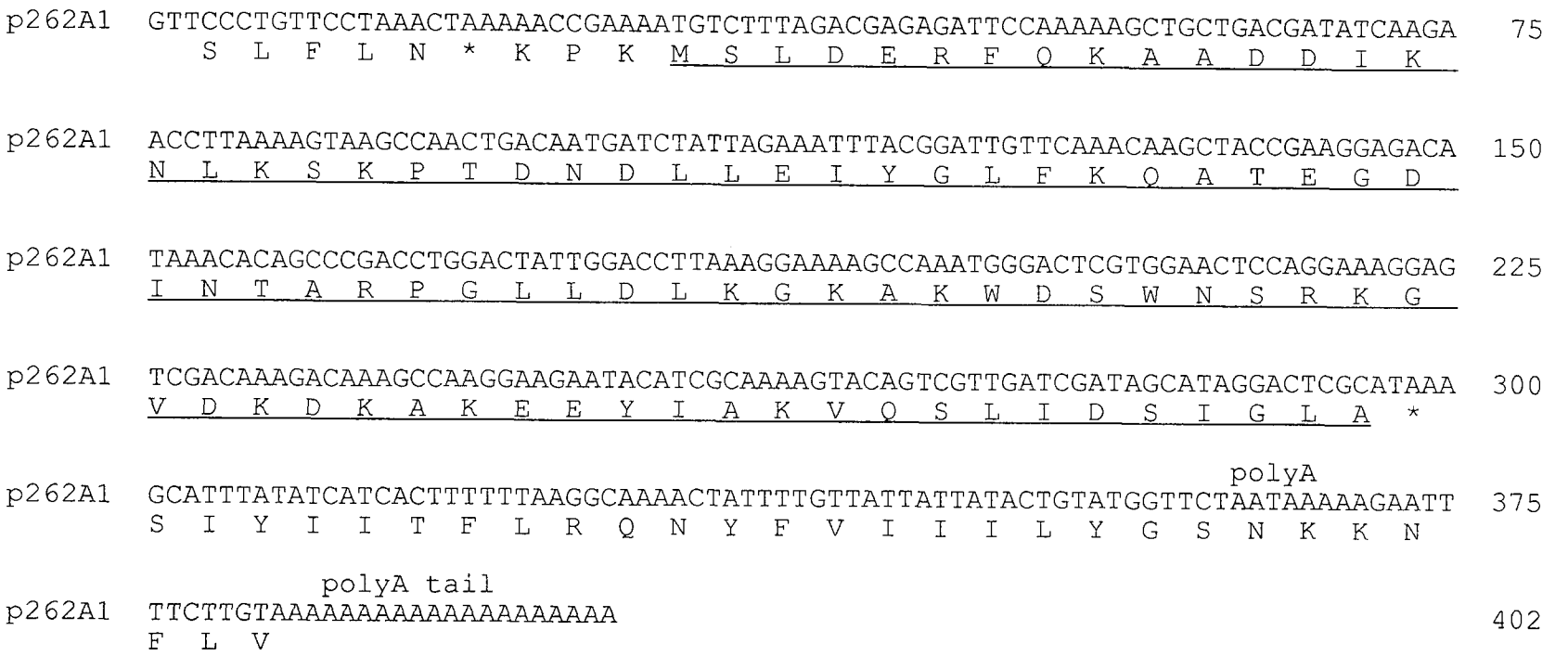


Figure 70. The DNA sequence of cDNA p262A1 and its translation in the three forwards reading frames. The underlined protein sequence is the long ORF which codes for acyl-CoA-binding protein.

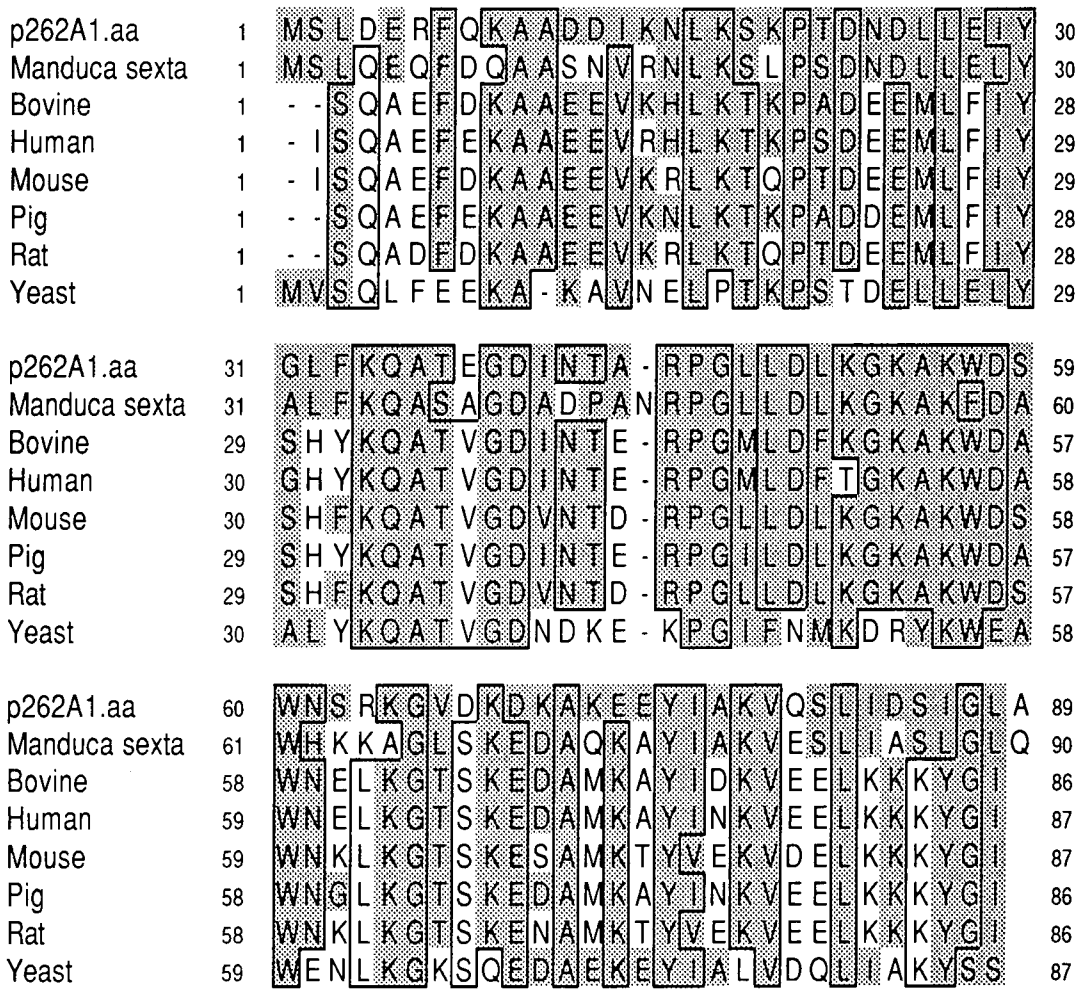


Figure 71. Protein sequence alignment of p262A1 and acyl-CoA-binding proteins from various other sources.

This polyA+ RNA was all used to make a cDNA library using a λZAP II kit (Stratagene ZAP-cDNA synthesis kit) following the kit manual. Samples of the first and second strand synthesis were made incorporating a radioactive label and were run on gel and autoradiographed. This showed the cDNA to have a size range from 100 to 2500 bp, i.e. the original RNA was essentially undegraded which was reassuring after the poor yield of polyA+ RNA. The yield of dscDNA was not quantified before ligation of the arms, but half of the product was used. After packaging a sample of the library was plated out. The library consisted of 1.0 x 10^6 independent plaques and there was 1 blue plaque, i.e. λ phage without an insert, in ~1100 plaques, i.e. approximately 0.1% of the library. The entire library was amplified and stored (in a number of aliquots by both of the methods in various fridges and freezers).

To further assess the quality of the library four single plaques were selected at random, plasmid rescued, and DNA sequenced in from both ends. Referred to as p262A1 to 4 they had inserts of 400, 400, 400 and 750 bp respectively. The deduced open reading frames were used to run fasta searches of the national protein sequence data bank at Daresbury.

p262A1 (figure 70) was a full or very near full length cDNA for acyl-CoA-binding protein, highly homologous to similar proteins from mammals, yeast and *Manduca sexta* (Snyder and Feyereisen 1993) (figure 71) to which it was most homologous, 61.1% over its full length. The 90 amino acid coding sequence runs from a methionine codon, 29 bp down stream of the 5' end which is an indifferent match to the animal consensus for start codons (Lütcke *et al.* 1987), to a stop codon, making it one amino acid shorter than the *Manduca* protein which has a one amino acid insert compared to all the other sequences, It is slightly longer than the mammalian proteins, which stop one amino acid earlier and lack the first two amino acids, and the yeast protein which stops one amino acid earlier and starts with a methionine one amino acid further down stream. The 3' untranslated sequence runs for 83 bp to a 20 bp polyA tail 16 bp down stream of a single and consensus polyadenylation signal. The cloning primers were present at both ends and undamaged.

Acyl-CoA-binding protein, also known as diazepam binding inhibitor, binds medium and long chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters. It is also able to displace diazepam from the benzodiazepam recognition site on the GABA type A receptor. It is therefore possible that

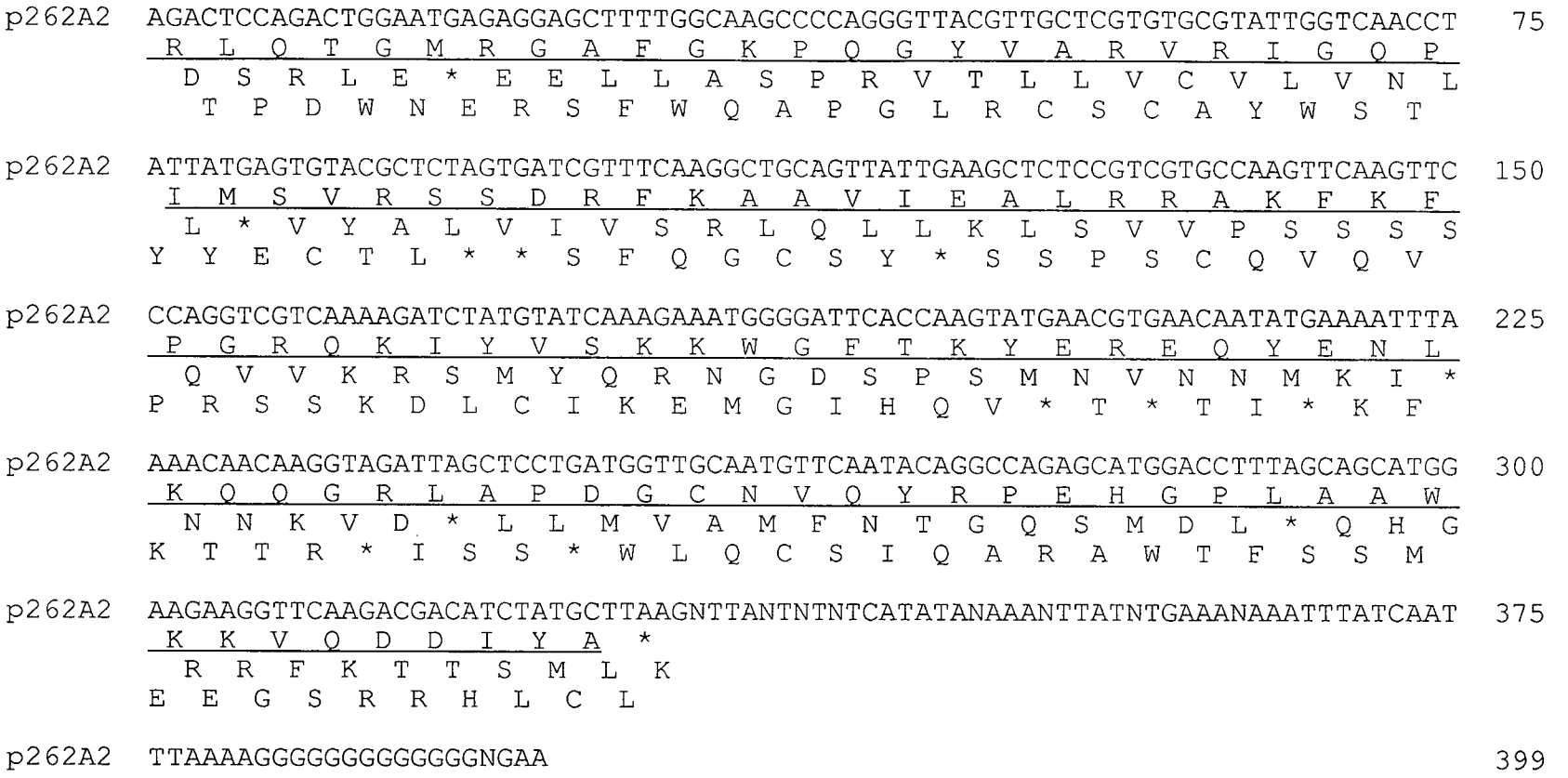


Figure 72. The DNA sequence of p262A2 and its translation in all three forwards reading frames. The underlined sequence is the long ORF which codes for a QM-like protein.

Figure 73. The alignment of p262A2 with human QM protein partial sequence (Swissprot P27635) (65.4% identity in 104 aa overlap).

```
                                                      10        20        30
p262A2                                              RLQTGMRGAFGKPQGTVARVRIGQPIMSVR
                                                    |||||||||||||||||||||:||| |||:|
qm_hum CGKDGFHIRVRLHPFHVIRINKMLSCAGADRLQTGMRGAFGKPQGTVARVHIGQVIMSIR
      80        90        100       110       120       130

               40        50        60        70        80        90
p262A2 SSDRFKAAVIEALRRAKFKFPGRQKIYVSKKWGFTKYEREQYENLKQQGRLAPDGCNVQY
       :: : |: |||||||||||||||||||:||||||||:: :::|:: :: || ||||:|:|
qm_hum TKLQNKEHVIEALRRAKFKFPGRQKIHISKKWGFTKFNADEFEDMVAEKRLIPDGCGVKY
     140       150       160       170       180       190

              100      109
p262A2 RPEHGPLAAWKKVQDDIYA
        |::|||: |: :::
qm_hum IPSRGPLDKWRALHS
     200       210
```

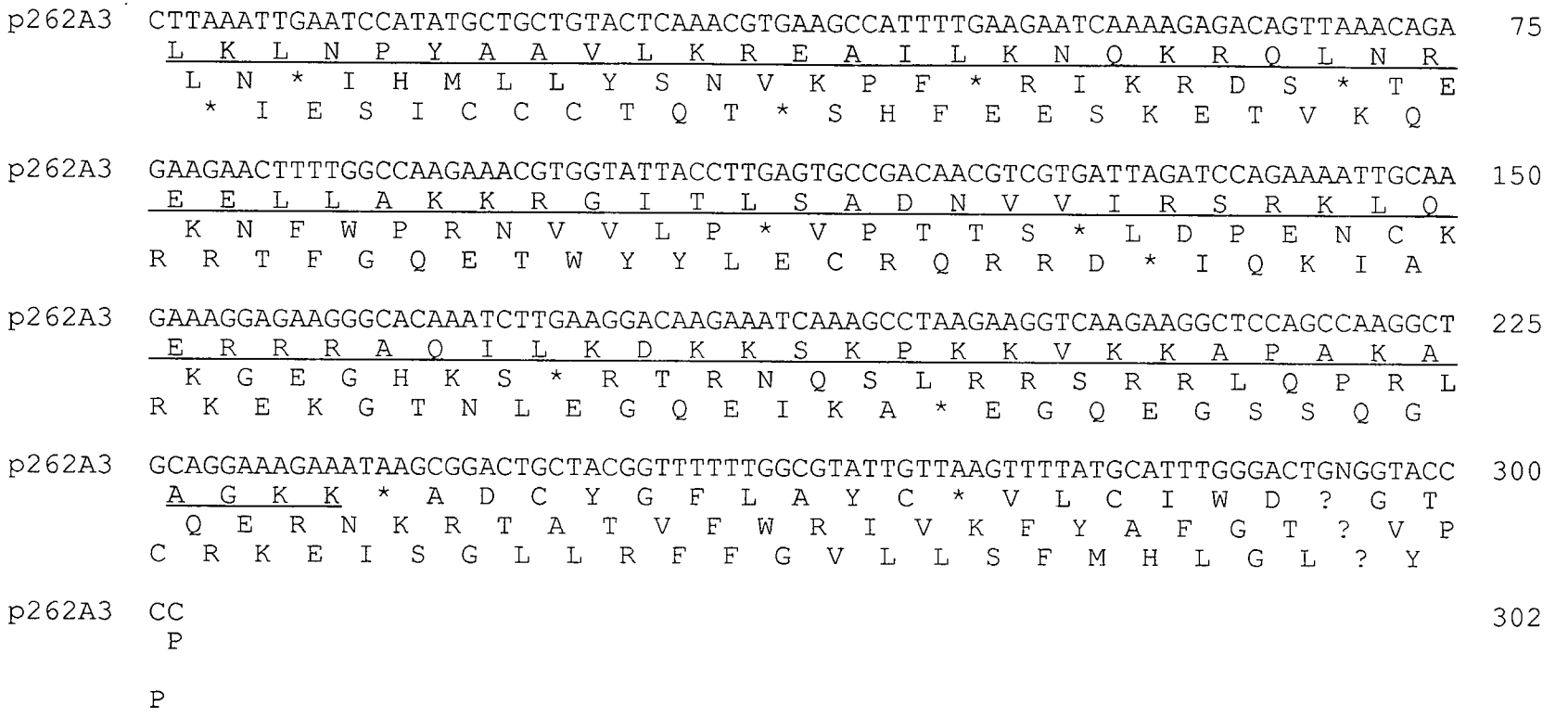


Figure 74. The DNA sequence of p262A3 and its translation in all three forwards reading frames. The underlined sequence is the long ORF which codes for a protein homologous to the L1 protein of the ribosomal 60 S subunit.

Figure 75. The alignment of p262A3 with 60S ribosomal protein L1 from *Drosophila melanogaster* (Swissprot P09180) (43.9% identity in 41 aa overlap).

```
                                                    10        20        30
p262A3                                            LKLNPYAAVLKREAILKNQKRQLNREELLA
                                                  :|||||||:|||| | | ::|| :::    |
rl1_dr EEIRKVLRDPRKRVFRSVRRLNPLTNVRQLIKLNPYAEVLKRRAALAAEKRTVAKVLAKA
      300       310       320       330       340       350

               40        50        60        70
p262A3 KKRGITLSADNVVIRSRKLQERRRAQILKDKKSKPKKVKKAPAKAAGKK
       ||:::::|: ::
rl1_dr KKQNVELAKSHFANVATKARPIAPSSGRPQEEGRRQEASGQEVNLLSPE
      360       370       380       390       400
```

it acts as a neuropeptide to modulate the action of the GABA receptor (Mogensen *et al.* 1987, Chen *et al.* 1988). As many chemical insecticides work by binding to neuropeptide receptors in insects this is an interesting observation and it might be speculated that this could be the starting point for investigation of insect control via this receptor. However, it is not known whether this GABA binding function occurs in insects.

p262A2 has an open reading frame running in from the 5' end with highly significant homology to human QM protein; 65.4% identity in an 104 amino acid overlap and two 30 amino acid runs of almost complete identity. (figures 72 and 73) The homology starts with the first codon of p262A2 which corresponds to amino acid 110 of QM protein and runs almost to the stop codon which is 4 amino acids further down stream than the end of QM protein. There is a short region of untranslated sequence beyond the stop codon and a polyA tail of about 20 bp however the DNA sequence in this region was poor and no other features could be distinguished. The comparison with QM protein suggests that p262A2 is just under half full length.

The significance of this homology is difficult to interpret as QM protein is of unknown function, and has not been reported isolated from any other sources. It is associated with the suppression of the tumourigenic phenotype in G401 Wilms tumour cell line by an unknown mechanism which appears to involve up regulation of QM protein expression. It is a member of a small multigene family in humans and homologous sequences were seen on genomic Southern blot in chimpanzee, baboon, capuchin, hamster, rat and mouse but not in chicken or *Drosophila melanogaster* , and is expressed widely in fetal and adult human tissues (Dowdy *et al.* 1991, Van Den Ouweland *et al.* 1992). However, despite the failure of the Southern blot to show any bands, the isolation of this cDNA shows that there are QM like protein genes expressed in insects. Thus it must be concluded that whatever the function of QM protein is it is also necessary in insects.

p262A3 (figures 74 and 75) has an open reading frame running in from the 5' end which has a just-significant homology to protein L1 of *Drosophila melanogaster* 60S ribosomal subunit, a member of the L1E family of ribosomal proteins (Rafti *et al.* 1988, Rafti *et al.* 1989). The homology is 43.9% in a 41 amino acid overlap starting from the 5' end of p262A3 and beyond the region of homology both sequences run to a stop at the same point. If this homology is real then p262A3 starts at amino acid 330 of the *Drosophila* protein and hence is about 25% of full length. This 41 amino acid region from

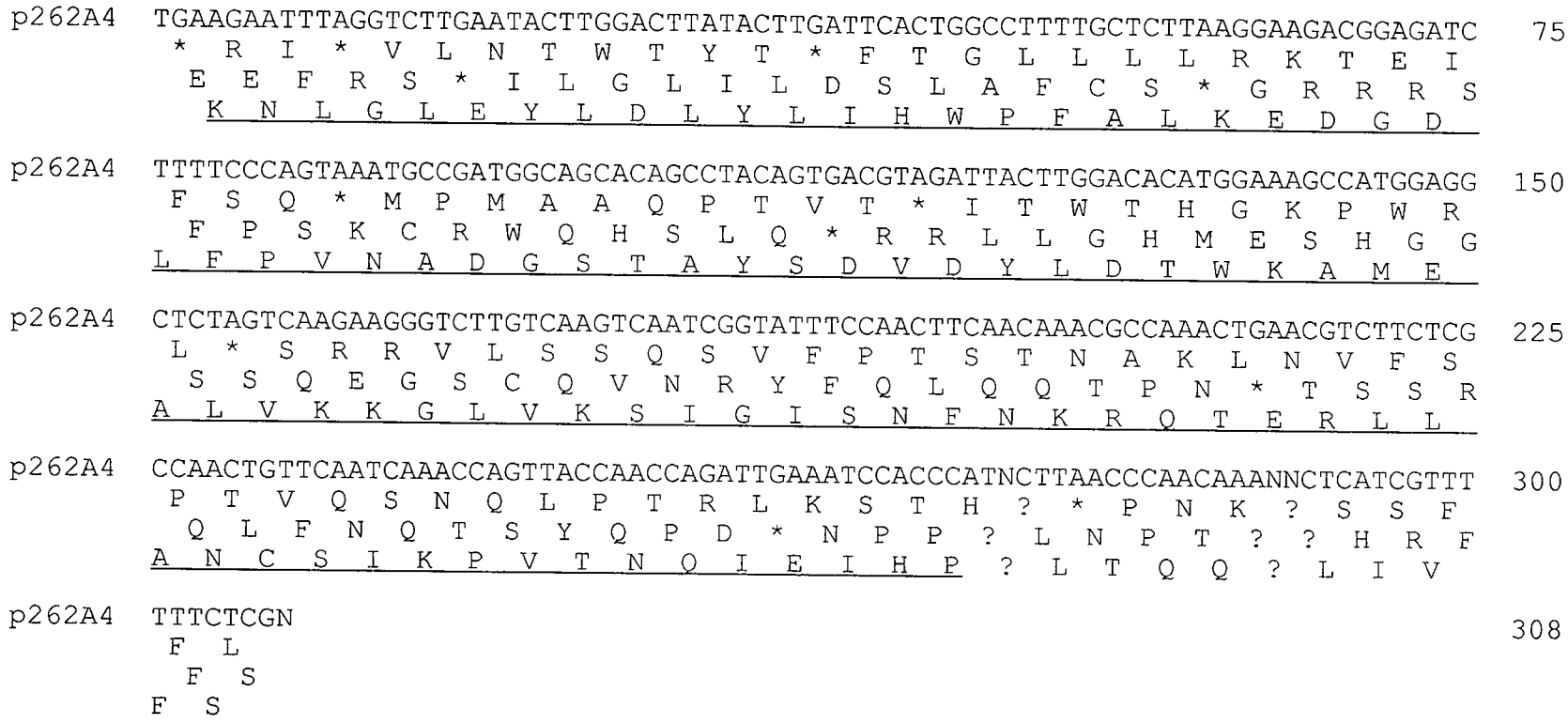


Figure 76. The DNA sequence of p262A4 and its translation in all three forwards reading frames. The underlined sequence is the long ORF which codes for a protein homologous to aldol reductase and other reductases.

The significant matches found to p262A4 by a fasta search of the national protein databank, and an example alignment.

Species	Swissprot		scores	init1	initn	opt
human	P14550	ALCOHOL DEHYDROGENASE (NADP(+)) (EC...	283	283	290
rabit	P15122	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)...	239	255	284
rat	P07943	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)	229	245	274
mouse	P21300	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)...	220	244	264
bovine	P16116	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)...	220	236	272
human	P15121	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)...	213	229	262
leishmania	P22045	PROBABLE REDUCTASE. 8/92	153	221	186
human	Q04828	TRANS-1,2-DIHYDROBENZENE-1,2-DIOL D...	186	203	218
Common frog	P02532	RHO CRYSTALLIN (FRAGMENT). 8/92	187	187	230
Bullfrog	P17264	RHO CRYSTALLIN. 8/92	187	187	230
barley	P23901	ALDOSE REDUCTASE (EC 1.1.1.21) (AR)...	132	182	228
bovine	P05980	PROSTAGLANDIN-F SYNTHASE (EC 1.1.1....	161	178	204
rat	P31210	3-OXO-5-BETA-STEROID 4-DEHYDROGENASE .	150	172	211
corynebacterium	P15339	2,5-DIKETO-D-GLUCONIC ACID REDUCTAS...	122	168	122
shuman	P17516	CHLORDECONE REDUCTASE (EC 1.1.1.225...	147	164	211
yeast	P14065	GCY PROTEIN. 10/93	89	161	179
rat	P23457	3-ALPHA-HYDROXYSTEROID DEHYDROGENASE .	160	160	208
soybn	P26690	NAD(P)H DEPENDENT 6'-DEOXYCHALCONE ...	91	150	209
yeast	P31867	NAD(P)H-DEPENDENT XYLOSE REDUCTASE ...	131	145	179
apple	P28475	NADP-DEPENDENT D-SORBITOL-6-PHOSPHA...	103	125	151

The alignment of p262A4 with human alcohol dehydrogenase (NADP(+)) (EC 1.1.1.2) (aldehyde reductase) (52.8% identity in 89 aa overlap).

```
                                                 10        20        30
p262A4                                        KNLGLEYLDLYLIHWPFALKEDGDLFPVNAD
                                              :| |||||||:|||:|:: ::: || |||
aldx_h PREELFVTSKLWNTKHHPEDVEPALRKTLADLQLEYLDLYLMHWPYAFERGDNPFPKNAD
       70        80        90       100       110       120

               40        50        60        70        80        90
p262A4 GSTAYSDVDYLDTWKAMEALVKKGLVKSIGISNFNKRQTERLLANCSIKPVTNQIEIHP
        |:: |::::| :||||:|||| ||||:::|:||||:||:: :|:  |::|:: |:| ||
aldx_h GTICYDSTHYKETWKALEALVAKGLVQALGLSNFNSRQIDDILSVASVRPAVLQVECHPY
      130       140       150       160       170       180

aldx_h LAQNELIAHCQARGLEVTAYSPLGSSDRAWRDPDEPVLLEEPVVLALAEKYGRSPAQILL
      190       200       210       220       230       240
```

Figure 77.

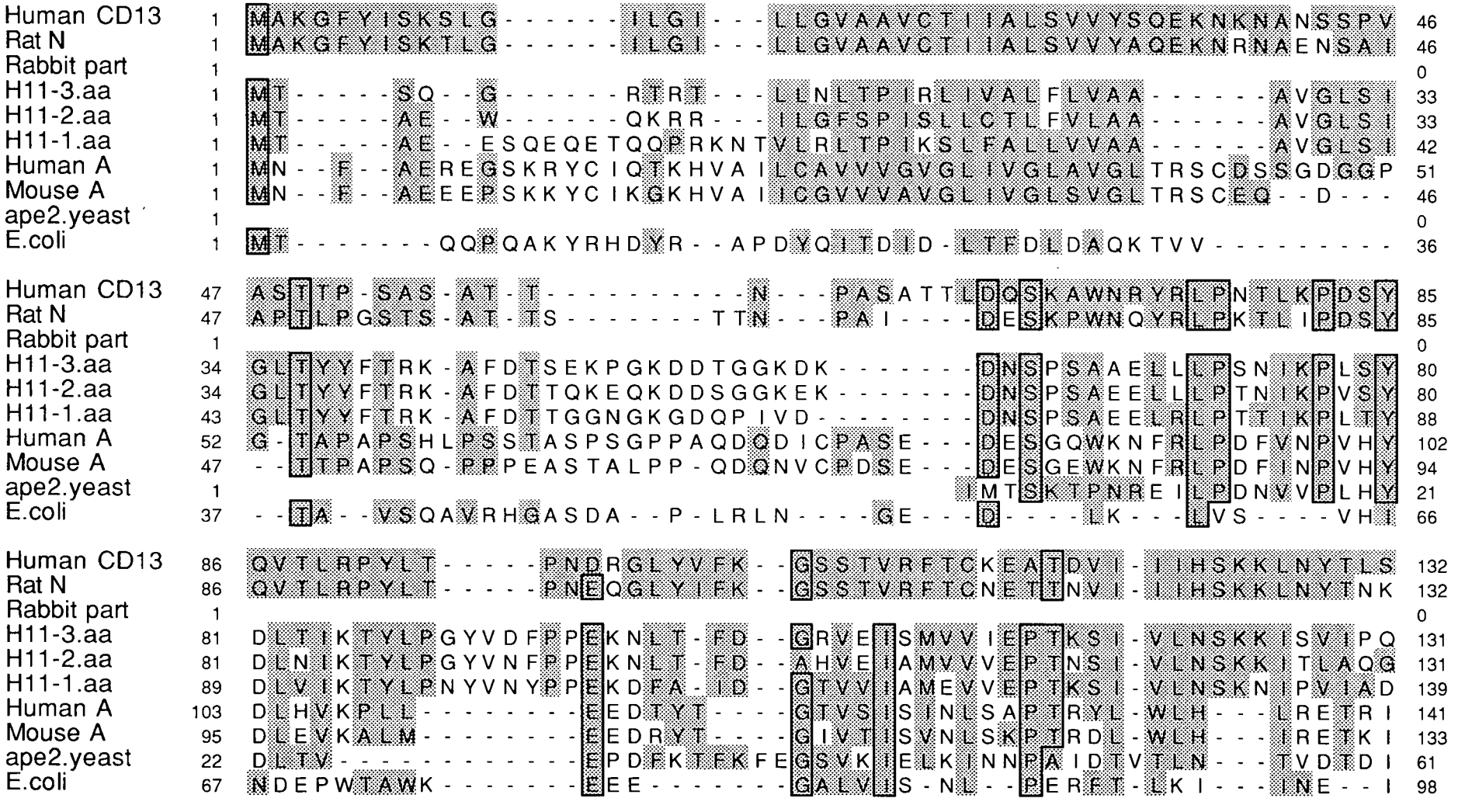
the *Drosophila* sequence was also used to search the protein database but found only its self with a significant score where as the whole *Drosophila* protein sequence found a number of other ribosomal L proteins from a range of species. Thus, either the match of p262A3 to ribosomal L1 protein is real and this 41 amino acid region must only be conserved in insects, or the match is chance and not significant. Without more sequence data no further conclusions can be drawn.

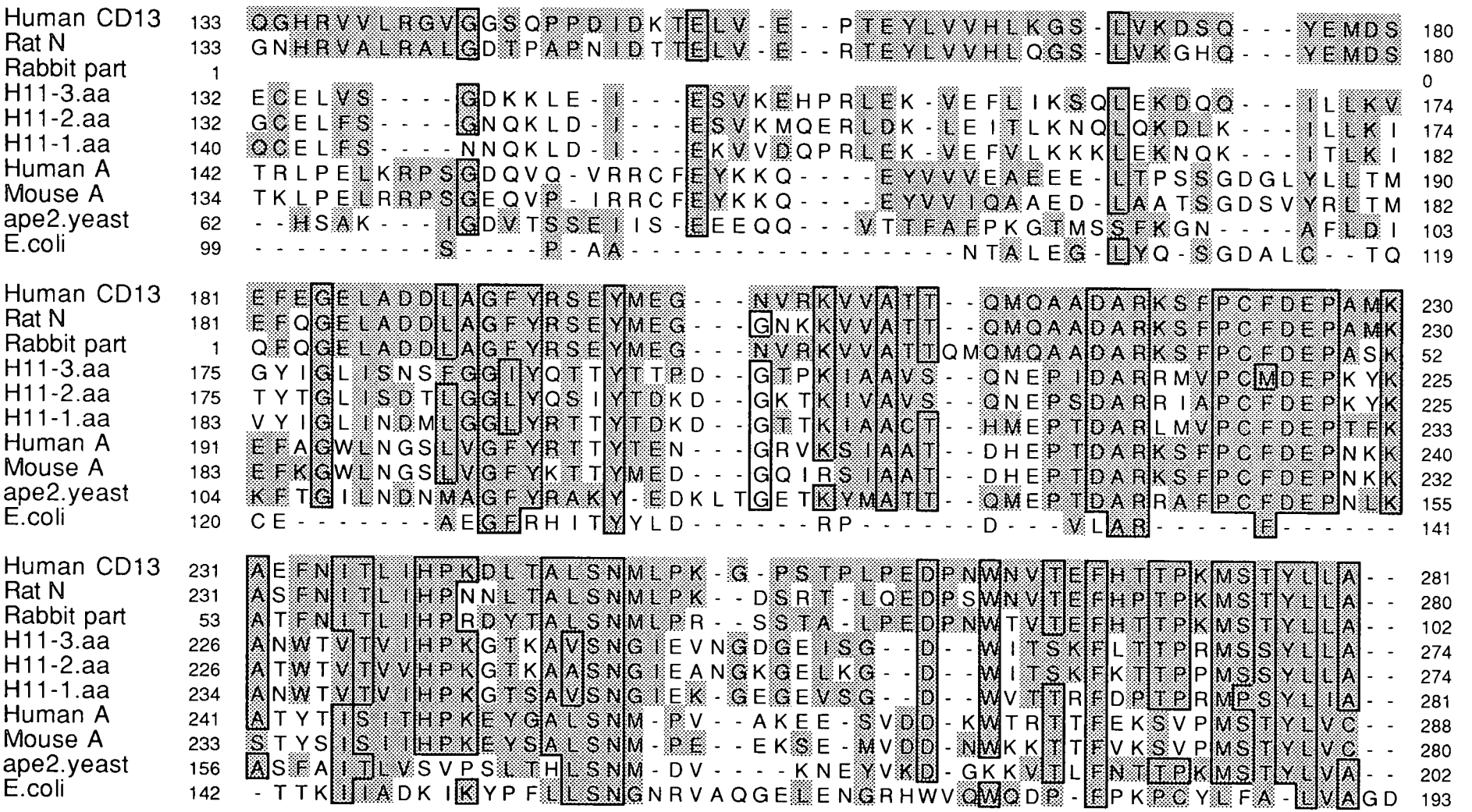
p262A4 has an undamaged 5' cloning primer and an open reading frame running in from it which is homologous, with a highly significant score, to reductases of the aldo-keto reductase superfamily (Bohren *et al.* 1989), which includes human liver alcohol dehydrogenase as well as examples from mammalian, amphibian, yeast, bacterial, trypanosome and plant species. (figures 76 and 77) The p262A4 coding sequence starts at amino acid 65 to 100 of the reductase sequence, depending on which member of the family you compare it with and hence is about 65% of full length. The region of homology, 56.5% identity in a 92 amino acid overlap for the best score, runs from the start to as far as the p262A4 sequence can be read. The 3' end of the p262A4 sequence has an undamaged cloning primer and a polyA tail but the sequence could not be read upstream of this. The p262A4 insert of ~750 bp is about the correct size to allow the open reading frame to reach the end of the consensus reductase sequence.

Aldehyde reductase [EC 1.1.1.2] and aldose reductase [EC 1.1.1.21] are cytosolic monomeric NADPH-dependant oxidoreductases having wide substrate specificities for carbonyl compounds. They were originally grouped on the basis of substrate specificities and physico-chemical properties and are now known to have a common structure and evolutionary history (Bohren *et al.* 1989) with examples known from reptiles and bacteria as well as mammals. They are also closely related to the ρ-crystalin proteins in frog eye lens and to a range of reductases important in prostaglandin and in steroid metabolism.

Thus as far a the quality of the *Diabrotica* cDNA library is concerned, from a small sample one out of four cDNAs is full length and no artifacts were observed in any sequence, suggesting that the library is of reasonable quality. These randomly selected *Diabrotica* cDNA clones have been preserved but no further work has been done on any of them.

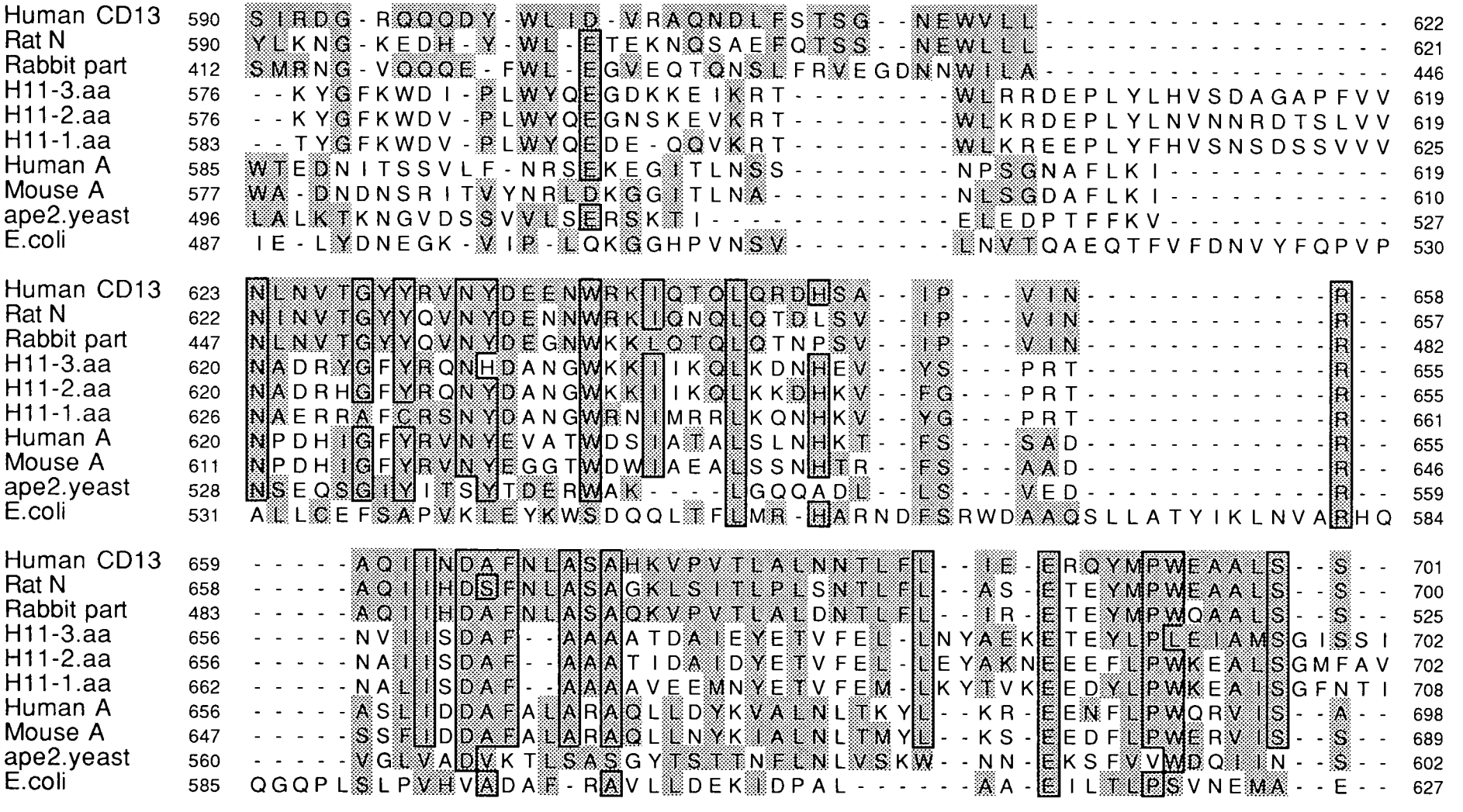
Conserved regions of microsomal aminopeptidase protein sequence, PCR oligos and PCR


Human CD13 1 MAKGFYISKSLG------ILGI---LLGVAAVCTIIALSVVYSQEKNKNANSSPV 46
Rat N 1 MAKGFYISKTLG------ILGI---LLGVAAVCTIIALSVVYAQEKNRNAENSAI 46
Rabbit part 1 0
H11-3.aa 1 MT-----SQ--G------RTRT---LLNLTPIRLIVALFLVAA------AVGLSI 33
H11-2.aa 1 MT-----AE--W------QKRR---ILGFSPISLLCTLFVLAA------AVGLSI 33
H11-1.aa 1 MT-----AE--ESQEQETQQPRKNTVLRLTPIKSLFALLVVAA------AVGLSI 42
Human A 1 MN--F--AEREGSKRYCIQTKHVAILCAVVVGVGLIVGLAVGLTRSCDSSGDGGP 51
Mouse A 1 MN--F--AEEEPSKKYCIKGKHVAIICGVVVAVGLIVGLSVGLTRSCEQ--D--- 46
ape2.yeast 1 0
E.coli 1 MT-------QQPQAKYRHDYR--APDYQITDID-LTFDLDAQKTVV--------- 36
Human CD13 47 ASTTP-SAS-AT-T---------N---PASATTLDQSKAWNRYRLPNTLKPDSY 85
Rat N 47 APTLPGSTS-AT-TS-------TTN---PAI----DESKPWNQYRLPKTLIPDSY 85
Rabbit part 1 0
H11-3.aa 34 GLTYYFTRK-AFDTSEKPGKDDTGGKDK-------DNSPSAAELLLPSNIKPLSY 80
H11-2.aa 34 GLTYYFTRK-AFDTTQKEQKDDSGGKEK-------DNSPSAEELLLPTNIKPVSY 80
H11-1.aa 43 GLTYYFTRK-AFDTTGGNGKGDQPIVD-------DNSPSAEELRLPTTIKPLTY 88
Human A 52 G-TAPAPSHLPSSTASPSGPPAQDQDICPASE---DESGQWKNFRLPDFVNPVHY 102
Mouse A 47 --TTPAPSQ-PPPEASTALPP-QDQNVCPDSE---DESGEWKNFRLPDFINPVHY 94
ape2.yeast 1 IMTSKTPNREILPDNVVPLHY 21
E.coli 37 --TA--VSQAVRHGASDA--P-LRLN----GE---D----LK---LVS----VHI 66
Human CD13 86 QVTLRPYLT-----PNDRGLYVFK--GSSTVRFTCKEATDVI-IIHSKKLNYTLS 132
Rat N 86 QVTLRPYLT-----PNEQGLYIFK--GSSTVRFTCNETTNVI-IIHSKKLNYTNK 132
Rabbit part 1 0
H11-3.aa 81 DLTIKTYLPGYVDFPPEKNLT-FD--GRVEISMVVIEPTKSI-VLNSKKISVIPQ 131
H11-2.aa 81 DLNIKTYLPGYVNFPPEKNLT-FD--AHVEIAMVVVEPTNSI-VLNSKKITLAQG 131
H11-1.aa 89 DLVIKTYLPNYVNYPPEKDFA-ID--GTVVIAMEVVEPTKSI-VLNSKNIPVIAD 139
Human A 103 DLHVKPLL---------EEDTYT----GTVSISINLSAPTRYL-WLH---LRETRI 141
Mouse A 95 DLEVKALM---------EEDRYT----GIVTISVNLSKPTRDL-WLH---IRETKI 133
ape2.yeast 22 DLTV----------EPDFKTFKFEGSVKIELKINNPAIDTVTLN---TVDTDI 61
E.coli 67 NDEPWTAWK-------EEE-------GALVIS-NL--PERFT-LKI---INE--I 98


Human CD13 133 QGHRVVLRGVGGSQPPDIDKTELV--EPTEYLVVHLKGS-LVKDSQ---YEMDS 180
Rat N 133 GNHRVALRALGDTPAPNIDTTELV--ERTEYLVVHLQGS-LVKGHQ---YEMDS 180
Rabbit part 1 0
H11-3.aa 132 ECELVS----GDKKLE-I---ESVKEHPRLEK-VEFLIKSQLEKDQQ---ILLKV 174
H11-2.aa 132 GCELFS----GNQKLD-I---ESVKMQERLDK-LEITLKNQLQKDLK---ILLKI 174
H11-1.aa 140 QCELFS----NNQKLD-I---EKVVDQPRLEK-VEFVLKKKLEKNQK---ITLKI 182
Human A 142 TRLPELKRPSGDQVQ-VRRCFEYKKQ----EYVVVEAEEE-LTPSSGDGLYLLTM 190
Mouse A 134 TKLPELRRPSGEQVP-IRRCFEYKKQ----EYVVIQAAED-LAATSGDSVYRLTM 182
ape2.yeast 62 --HSAK---IGDVTSSEIIS-EEEQQ---VTTFAFPKGTMSSFKGN----AFLDI 103
E.coli 99 ----------S----P-AA----------NTALEG-LYQ-SGDALC--TQ 119
Human CD13 181 EFEGELADDLAGFYRSEYMEG---NVRKVVATT--QMQAADARKSFPCFDEPAMK 230
Rat N 181 EFQGELADDLAGFYRSEYMEG---GNKKVVATT--QMQAADARKSFPCFDEPAMK 230
Rabbit part 1 QFQGELADDLAGFYRSEYMEG---NVRKVVATTQMQMQAADARKSFPCFDEPASK 52
H11-3.aa 175 GYIGLISNSFGGIYQTTYTTPD--GTPKIAAVS--QNEPIDARRMVPCMDEPKYK 225
H11-2.aa 175 TYTGLISDTLGGLYQSIYTDKD--GKTKIVAVS--QNEPSDARRIAPCFDEPKYK 225
H11-1.aa 183 VYIGLINDMLGGLYRTTYTDKD--GTTKIAACT--HMEPTDARLMVPCFDEPTFK 233
Human A 191 EFAGWLNGSLVGFYRTTYTEN---GRVKSIAAT--DHEPTDARKSFPCFDEPNKK 240
Mouse A 183 EFKGWLNGSLVGFYKTTYMED---GQIRSIAAT--DHEPTDARKSFPCFDEPNKK 232
ape2.yeast 104 KFTGILNDNMAGFYRAKY-EDKLTGETKYMATT--QMEPTDARRAFPCFDEPNLK 155
E.coli 120 CE-------AEGFRHITYYLD------RP------D---VLAR-----F------ 141
Human CD13 231 AEFNITLIHPKDLTALSNMLPK-G-PSTPLPEDPNWNVTEFHTTPKMSTYLLA-- 281
Rat N 231 ASFNITLIHPNNLTALSNMLPK--DSRT-LQEDPSWNVTEFHPTPKMSTYLLA-- 280
Rabbit part 53 ATFNITLIHPRDYTALSNMLPR--SSTA-LPEDPNWTVTEFHTTPKMSTYLLA-- 102
H11-3.aa 226 ANWTVTVIHPKGTKAVSNGIEVNGDGEISG--D--WITSKFLTTPRMSSYLLA-- 274
H11-2.aa 226 ATWTVTVVHPKGTKAASNGIEANGKGELKG--D--WITSKFKTTPPMSSYLLA-- 274
H11-1.aa 234 ANWTVTVIHPKGTSAVSNGIEK-GEGEVSG--D--WVTTRFDPTPRMPSYLIA-- 281
Human A 241 ATYTISITHPKEYGALSNM-PV--AKEE-SVDD-KWTRTTFEKSVPMSTYLVC-- 288
Mouse A 233 STYSISIIHPKEYSALSNM-PE--EKSE-MVDD-NWKKTTFVKSVPMSTYLVC-- 280
ape2.yeast 156 ASFAITLVSVPSLTHLSNM-DV----KNEYVKD-GKKVTLFNTTPKMSTYLVA-- 202
E.coli 142 -TTKIIADKIKYPFLLSNGNRVAQGELENGRHWVQWQDP-FPKPCYLFA-LVAGD 193


Human CD13 282 FIVSEFDYVEKQASN-GVLIRIWARPSAIAAGHGDYALNVTGPILNFFAGHYDTP 335
Rat N 281 YIVSEFKYVEAVSPN-RVQIRIWARPSAIDEGHGDYALQVTGPILNFFAQHYNTA 334
Rabbit part 103 YIVSEFTNIEAQSPN-NVQIRIWARPSAISEGHGQYALMVTGPILNFFANHYNTP 156
H11-3.aa 275 VMVSEFEYIEGETKT-GVRFRIWSRPEAKKMTQ--YALQSGIKCIEFYEDFFDIR 326
H11-2.aa 275 IIVCEFEYIEGFTKT-GVRFRIWSRPEAKRMTA--YALDAGIRCLEFYEKFFDIK 326
H11-1.aa 282 LVISEFKYIENYTKS-GVRFRIPARPEAMKMTE--YAMIAGIKCLDYYEDFFGIK 333
Human A 289 FAVHQFDSV-KRISNSGKPLTIYVQPEQKHTAE--YAANITKSVFDYFEEYFAMN 340
Mouse A 281 FAVHRFTAIE-RKSRSGKPLKVYVQPNQKETAE--YAANITQAVFDYFEDYFAME 332
ape2.yeast 203 FIVAELKYVESKNFRIPVRV--YATPGNEKHGQFA--ADLTAKTLAFFEKTFGIQ 253
E.coli 194 FDVLRDTFTT-RSGREVA-LELYVDRGNLDRAP--WAMTSLKNSMKWDERFGLE 244
Human CD13 336 YPLPKSDQIGLPDFNAGAMENWGLVTYRENSLLFDPLSSSSSNKERVVTVIAHEL 390
Rat N 335 YPLEKSDQIALPDFNAGAMENWGLVTYRESALVFDPQSSSISNKERVVTVIAHEL 389
Rabbit part 157 YPLEKSDQIGLPDFNAGAMENWGLVTYRESALLFDPLVSSISNKERVVTVVAHEL 211
H11-3.aa 327 FPLKKQDMIALPDFSAGAMENWGLITYRENSLLYDDRFYAPMNKQRIARIVAHEL 381
H11-2.aa 327 FPLEKQDMIALPDFTAGAMENWGLITYREDSLLYDEKIYAPMNKQRVALVVAHEL 381
H11-1.aa 334 FPLPKQDMVALPDFSSGAMENWGLITYREGSVLYDENLYGPMNKERVAEVIAHEL 388
Human A 341 YSLPKLDKIAIPDFGTGAMENWGLITYRETNLLYDPKESASSNQQRVATVVAHEL 395
Mouse A 333 YALPKLDKIAIPDFGTGAMENWGLVTYRETNLLYDPLLSASSNQQRVASVVAHEL 387
ape2.yeast 254 YPLPKMDNVAVHEFSAGAMENWGLVTYRVVDLLLDKDNSTLDRIQRVAEVVQHEL 308
E.coli 245 YDLDIYMIVAVDFFNMGAMENKGLNIFNSKYVLARTDTATDKDYLDIERVIGHEY 299
Human CD13 391 AHQWFGNLVTIEWWNDLWLNEGFASYV--EY---LG---ADYAEPTWNLKDLMVL 437
Rat N 390 AHQWFGNLVTVDWWNDLWLNEGFASYV--EF---LG---ADYAEPTWNLKDLIVL 436
Rabbit part 212 AHQWFGNLVTVDWWNDLWLNEGFASYV--EY---LG---ADYAEPTWNLKDLIVL 258
H11-3.aa 382 AHQWFGDLVTMKWWDNLWLNEGFARFT--EF---IG---AGQITQDDARMRNYFL 428
H11-2.aa 382 AHQWFGNLVTLKWWDDTWLNEGFATFV--EY---LG---MDEISHNNFRTQDFFL 428
H11-1.aa 389 AHQWFGNLVTMKWWDNLWLNEGFASFV--EY---IG---ADFISDGLWEMKDFFL 435
Human A 396 VHQWFGNIVTMDWWEDLWLNEGFASFF--EF---LG---VNHAETDWQMRDQMLL 442
Mouse A 388 VHQWFGNTVTMDWWDDLWLNEGFASFF--EF---LG---VNHAEKDWQMLSQVLL 434
ape2.yeast 309 AHQWFGNLVTMDWWEGLWLNEGFATWM--SW---YS---CNEFQPEWKVWEQ-YV 354
E.coli 300 FHNWTGNRVTCRDWFQLSLKEGLTVFRDQEFSSDLGSRAVNRINNVRTMRG---L 351


Human CD13 438 NDVY-RVMAVDALASSHPLSTP---ASEINTPAQISELFDAISYSKGASVLRMLS 488
Rat N 437 NDVY-RVMAVDALASSHPLSSP---ANEVNTPAQISELFDSITYSKGASVLRMLS 487
Rabbit part 259 NELH-SVMAVDALASSHPLSSP---ADEVNTPAQISELFDSITYSKGASVLRMLS 309
H11-3.aa 429 IDVLERALKADSVASSHPLSFRIDKAAEVE------EAFDDITYAKGASVLTMLR 477
H11-2.aa 429 LDGMDRGMRADSAASSHPLSFRIDKAAEVA------EAFDDISYAKGASVLTMLR 477
H11-1.aa 436 LAPYTSGITADAVASSHPLSFRIDKAADVS------EAFDDITYRKGASVLQMLL 484
Human A 443 EDVL-PVQEDDSLMSSHPIIVT---VT---TPDEITSVFDGISYSKGSSILRMLE 490
Mouse A 435 EDVF-PVQEDDSLMSSHPVVVT---VS---TPAEITSVFDGISYSKGASILRMLQ 482
ape2.yeast 355 TDTLQHALSLDSLRSSHPIEVP---VK---KADEINQIFDAISYSKGASLLRMIS 403
E.coli 352 Q---F---AEDASPMA-HPIRPD--MVI------EMNNFYTLTVYEKGAEVIRMIH 392
Human CD13 489 SFLSEDVFKQGLASYLHTFAYQNTIYLNLWDHLQEAV-NNRS-IQLPTTVRDIMN 541
Rat N 488 SFLTEDLFKKGLSSYLHTFQYSNTIYLDLWEHLQQAV-DSQTAIKLPASVSTIMD 541
Rabbit part 310 SFLTEDLFKEGLASYLHTFAYQNTIYLDLWEHLQQAV-NSQSAIQLPASVRDIMD 363
H11-3.aa 478 ALIGEEKHKHAVSQYLKKFSYSNAEATDLWAVFDEVVTDVEGPDGKPMKTTEFAS 532
H11-2.aa 478 ALIGEDNYRNAVVQYLKKFSYSNAQAADLWNVFNEVVKGVKGPDGNVMKIDQFTD 532
H11-1.aa 485 NLVGDENFKQSVSRYLKKFSYDNAAAEDLWAAFDETVQGITGPNGGPLKMSEFAP 539
Human A 491 DWIKPENFQKGCQMYLEKYQFKNAKTSDFWAALEEA---SR----LPVK--EVMD 536
Mouse A 483 DWITPEKFQKGCQIYLKKFQFANAKTSDFWDSLQEA---SN----LPVK--EVMD 528
ape2.yeast 404 KWLGEETFIKGVSQYLNKFKYGNAKTEDLWDALADA---SG----KEVR--SVMN 449
E.coli 393 TLLGEENFQKGMQLYFERHDGSAATCDDFVQAMEDA---SN----VDLS--HFRR 438
Human CD13 542 RWTLQMGFPVITV---DT--STGTLSQEHFLLDPD-SNVTRPSEFNYVWIVPIT 589
Rat N 542 RWILQMGFPVITV---NT--STGEIYQEHFLLDPT-SKPTRPSDFNYLWIVPIP 589
Rabbit part 364 RWILQMGFPVVTV---NT--TNGIISQHHFLLDPT-SNVTRPSDFNYLWIVPVS 411
H11-3.aa 533 QWTTQMGFPVISVAEFN---STT---------LKLTQSRYEAN-KDAVEKEKYRHP 575
H11-2.aa 533 QWTYQMGYPVVKVEEFN---ATA---------LKVTQSRYKTN-KDALEPEKYRNP 575
H11-1.aa 540 QWTTQMGFPVLTVESVN---ATT---------LKVTQKRYRQN-KDAKEPEKYRHP 582
Human A 537 TWTRQMGYPVLNVNGVKN--ITQKRFLLDPR-ANPSQPPSDLGYTWNIPVK 584
Mouse A 529 TWTSQMGYPVVTVSGRQN--ITQKRFLLDSK-ADPSQPPSELGYTWNIPVR 576
ape2.yeast 450 IWTKKVGFPVISVSEDGNGKITFR--QNR-YL-S--TADVKP--DEDK-TIYPVF 495
E.coli 439 -WYSQSGTPIVTVKDDYNP-ET---EQYT--LTISQRTPATPDQAEKQPLHIPFA 486


Human CD13 590 SIRDG-RQQQDY-WLID-VRAQNDLFSTSG--NEWVLL---------- 622
Rat N 590 YLKNG-KEDH-Y-WL-ETEKNQSAEFQTSS--NEWLLL---------- 621
Rabbit part 412 SMRNG-VQQQE-FWL-EGVEQTQNSLFRVEGDNNWILA---------- 446
H11-3.aa 576 --KYGFKWDI-PLWYQEGDKKEIKRT--------WLRRDEPLYLHVSDAGAPFVV 619
H11-2.aa 576 --KYGFKWDV-PLWYQEGNSKEVKRT--------WLKRDEPLYLNVNNRDTSLVV 619
H11-1.aa 583 --TYGFKWDV-PLWYQEDE-QQVKRT--------WLKREEPLYFHVSNSDSSVVV 625
Human A 585 WTEDNITSSVLF-NRSEKEGITLNSS--------NPSGNAFLKI---------- 619
Mouse A 577 WA-DNDNSRITVYNRLDKGGITLNA---------NLSGDAFLKI---------- 610
ape2.yeast 496 LALKTKNGVDSSVVLSERSKTI----------ELEDPTFFKV---------- 527
E.coli 487 IE-LYDNEGK-VIP-LQKGGHPVNSV--------LNVTQAEQTFVFDNVYFQPVP 530
Human CD13 623 NLNVTGYYRVNYDEENWRKIQTCLQRDHSA--IP---VIN----------R-- 658
Rat N 622 NINVTGYYQVNYDENNWRKLQNQLQTDLSV--IP---VIN----------R-- 657
Rabbit part 447 NLNVTGYYQVNYDEGNWKKLQTQLQTNPSV--IP---VIN----------R-- 482
H11-3.aa 620 NADRYGFYRQNHDANGWKKIIKQLKDNHEV--YS---PRT----------R-- 655
H11-2.aa 620 NADRHGFYRQNYDANGWKKIIKQLKKDHKV--FG---PRT----------R-- 655
H11-1.aa 626 NAERRAFCRSNYDANGWRNIMRRLKQNHKV--YG---PRT----------R-- 661
Human A 620 NPDHIGFYRVNYEVATWDSIATALSLNHKT--FS---SAD----------R-- 655
Mouse A 611 NPDHIGFYRVNYEGGTWDWIAEALSSNHTR--FS---AAD----------R-- 646
ape2.yeast 528 NSEQSGIYITSYTDERWAK----LGQQADL--LS---VED----------R-- 559
E.coli 531 ALLCEFSAPVKLEYKWSDQQLTFLMR-HARNDFSRWDAAQSLLATYIKLNVARHQ 584
Human CD13 659 -----AQIINDAFNLASAHKVPVTLALNNTLFL--IE-ERQYMPWEAALS--S-- 701
Rat N 658 -----AQIIHDSFNLASAGKLSITLPLSNTLFL--AS-ETEYMPWEAALS--S-- 700
Rabbit part 483 -----AQIIHDAFNLASAQKVPVTLALDNTLFL--IR-ETEYMPWQAALS--S-- 525
H11-3.aa 656 -----NVIISDAF--AAAATDAIEYETVFEL-LNYAEKETEYLPLEIAMSGISSI 702
H11-2.aa 656 -----NAIISDAF--AAATIDAIDYETVFEL-LEYAKNEEEFLPWKEALSGMFAV 702
H11-1.aa 662 -----NALISDAF--AAAAVEEMNYETVFEM-LKYTVKEEDYLPWKEAISGFNTI 708
Human A 656 -----ASLIDDAFFALARAQLLDYKVALNLTKYL--KR-EENFLPWQRVIS--A-- 698
Mouse A 647 -----SSFIDDAFFALARAQLLNYKIALNLTMYL--KS-EEDFLPWERVIS--S-- 689
ape2.yeast 560 -----VGLVADVKTLSASGYTSTTNFLNLVSKW--NN-EKSFVVWDQIIN--S-- 602
E.coli 585 QGQPLSLPVHVADAF-RAVLLDEKIDPAL------AA-EILTLPSVNEMA--E-- 627


Human CD13 702 LSYFKLMFDRSEVYGPMKNYLKKQVTPLFIHFRNNTNNWREIPENLMDQYSEVNA 756
Rat N 701 LNYFKLMFDRSEVYGPMKRYLKKQVTPLFAYFKIKTNNWLDRPPTLMEQYNEINA 755
Rabbit part 526 LNYFKLMFDRSEVYGPMKNYLSKQVRPLFEHFKNITNDWTRRPDTLMDQYNEINA 580
H11-3.aa 703 LKYFPTEPEAKPAQTYMMNILKPMYEKSSIDF-IA-NNY--RNDKLFFQINLQKD 753
H11-2.aa 703 LKFFGNEPETKPARAYMMSILEPMYNKSSIDY-IV-KNY--LDDTLFTKINTQKD 753
H11-1.aa 709 LDFFGSEPESQWASEYMRKLMKPIYDKSSIKF-IA-ENY--KKDSLFFKNNLQIA 759
Human A 699 VTYIISMFEDDKELYPMIEEYFQGQVKPIADS-LG---WNDAGDHVT-KL-LRSS 747
Mouse A 690 VSYIISMFEDDRELYPMIETYFQGQVKPVADL-LG---WQDTGSHIT-KL-LRAS 738
ape2.yeast 603 LSSMKSTWLFEPKETQDALDNFTKQLISGMTHHLG---WEFKSSDSFSTQRLKVT 654
E.coli 628 LFDII----DPIAIAEVREALTRTLATELADELLAIYNANYQSEYRVEHEDIAKR 678
Human CD13 757 -ISTACSN----GVPECEEMVSG-LF-K--QW-MENP--NNN-P------IHPNL 792
Rat N 756 -ISTACSS----GLEECRDLVVG-LY-S--QW-MNNS--DNN-P------IHPNL 791
Rabbit part 581 -ISTACSN----GIQECETLVSD-LF-K--QW-MDDP--SNN-P------IHPNL 616
H11-3.aa 754 VIDMFCAL----GSQDCRKKYKK-LF-D--DEVMNKCR-DGQ-AATECVRIAAPL 798
H11-2.aa 754 IIDAYCSL----GSKDCIKQYKD-IF-Y--DEVMPKCK-AGE-AATKCVKVSAPL 798
H11-1.aa 760 VIDTYCGL----GGKECLEEMKK-LF-D--KEVM-KCQ-PGQ-QATDCVKVTAPL 803
Human A 748 VLGFACKM----GDREALNNASS-LF-E--QW-LNGT--VSL-P------V--NL 782
Mouse A 739 ILGFACKM----GDREALGNASQ-LF-D--SW-LKGS--ASI-P------V--NL 773
ape2.yeast 655 MFGAACAA----RDADVEKAALK-MF-TDYCSG-NKEAIPALIKP-----IVFNTV 698
E.coli 679 TLRNACLRFLAFGETHLADVLVSKQFHE--ANNMTDALAALSAAVAAQLPCRDAL 731
Human CD13 793 RSTVYCNAIAQGGEEE-WDFAWEQFRNA----TLVNEADKLRAALACSKELWILN 842
Rat N 792 RSTVYCNAIASC-EEA-WNFAWATVPER----TLVNEADKLRSAVGRSNEVWILN 840
Rabbit part 617 RTTVYCNAIALGGERE-WDFAWEQFRNA----TLVNEADKLRSALACSNEVWILN 666
H11-3.aa 799 RSSVYCYGVK----EG-GDYASDKVMELYTAETLALEKDFLRLALGCHKDVTALK 848
H11-2.aa 799 RANVYCYGVQ----EG-GEEAFEKVMGLYLAEDVQLEKGILFKALACHKDVTALK 848
H11-1.aa 804 RKTVYCYGVQ----EG-GDEAFDKVMELYNAEQVQLEKDSLREALGCHKDVTALK 853
Human A 783 RLLVYRYGMQNSGNEISWNYTLEQYQKT----SLAQEKEKLLYGLASVKNVTLLS 833
Mouse A 774 RLLVYRYGMQNSGNEAAWNYTLEQYQKT----SLAQEKEKLLYGLASVKDVKLLA 824
ape2.yeast 699 ----ARVG----GAE---NY--EKVYKIYLDPISNDEKLAALRSLGRFKEPKLLE 740
E.coli 732 MQEYDDKWHQNGLVMDKWFILQATSPAANVLETVRGLLQHRSFTMSNPNRIRSLI 786

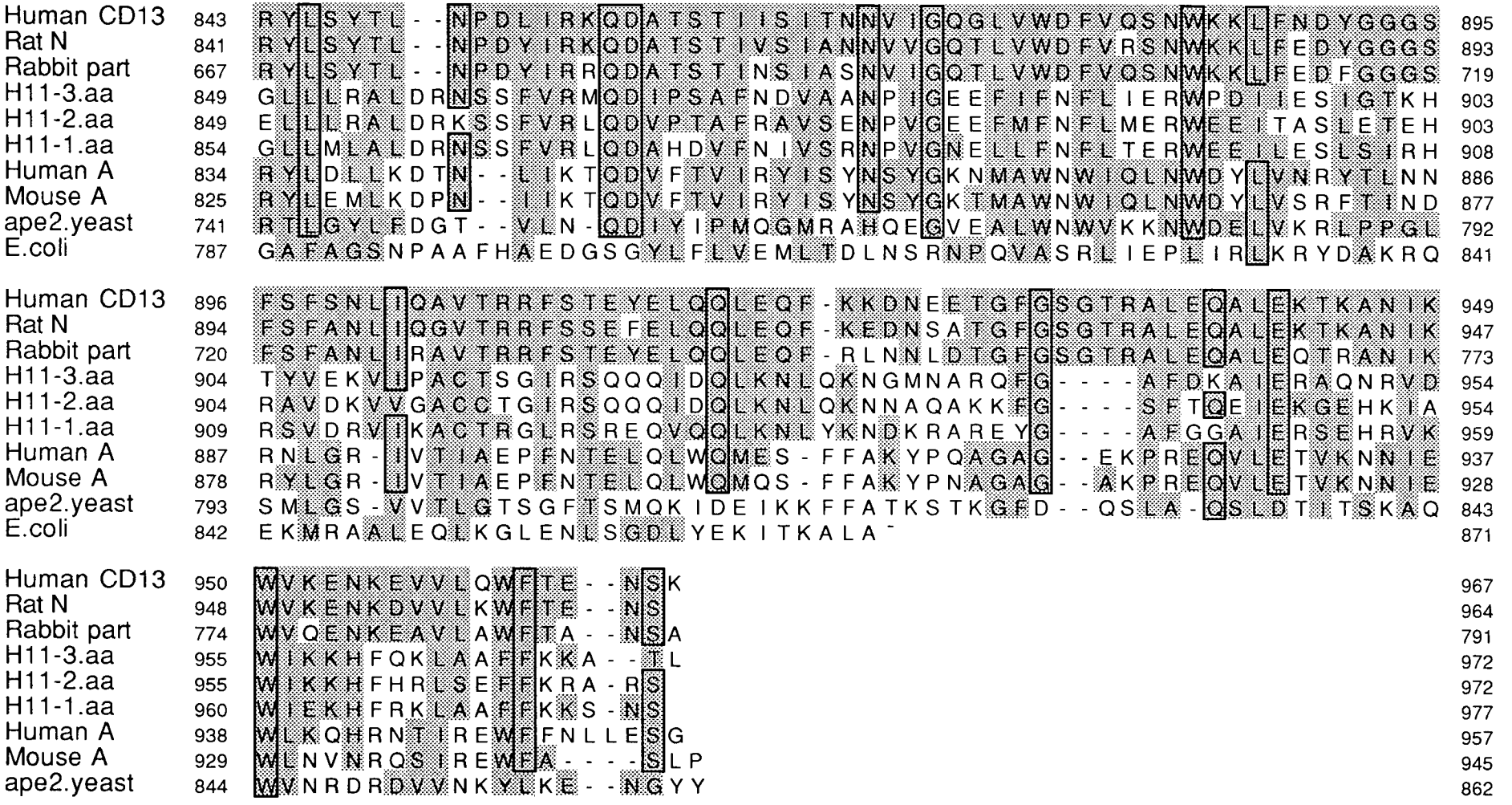


Figure 78. Alignment of the protein sequences of microsomal aminopeptidases from mammalian, nematode (H11-1, H11-2 and H11-3), yeast and bacterial sources.

Microsomal aminopeptidase 5' strongly conserved region

H11-3	334-360	I A L P D F S A G A M E N W G L I T Y R E N S L L Y D
H11-2	334-360	I A L P D F T A G A M E N W G L I T Y R E N S L L Y D
H11-1	341-367	V A L P D F S S G A M E N W G L I T Y R E G S V L Y D
Human CD13	342-368	I G L P D F N A G A M E N W G L V T Y R E N S L L F D
Rat AP-N	343-369	I A L P D F N A G A M E N W G L V T Y R E S A L V F D
Rabbit AP-N	165-191	I G L P D F N A G A M E N W G L V T Y R E S A L L Y D
Yeast ape2	262-288	V A V H E F S A G A M E N W G L V T Y R V V D L L L D
E. coli	253-279	V A V D F F N M G A M E N K G L N I F N S K Y V L A R

Oligo #952 G A M E N W G

GGIGCIATGGAAAACTGGGG

G T

20 mer, 2 inosines, 4 fold redundancy.

Figure 79. Microsomal aminopeptidase 5' oligo (#952) design.

Microsomal aminopeptidase middle and 3' strongly conserved regions

H11-3	376-409	I V A H E L A H Q W F G D L V T M K W W D N L W L N E G F A R F T
H11-2	376-409	V V A H E L A H Q W F G N L V T L K W W D D T W L N E G F A T F V
H11-1	383-416	V I A H E L A H Q W F G N L V T M K W W D N L W L N E G F A S F V
Human CD13	385-418	V I A H E L A H Q W F G N L V T I E W W N D L W L N E G F A S Y V
Rat AP-N	383-416	V I A H E L A H Q W F G N L V T V D W W N D L W L N E G F A S Y V
Rabbit AP-N	206-239	V V A H E L A H Q W F G N L V T V D W W N D L W L N E G F A S Y V
Yeast ape2	303-336	V V Q H E L A H Q W F G N L V T M D W W E G L W L N E G F A T W M
E. coli	294-327	V I G H E Y F H N W T G N R V T C R D W F Q L S L K E G L T V F R

Oligo #954 A H Q W F G N

GCNCACCAATGGTTCGGIAA

* T G T

Oligo #953 W L N E G F A

TGGCTIAACGAAGGITTCGC

T T G T

Both oligos were made in the antisense to the sequence shown, and with A/C/G/T for the N marked *.

Figure 80. Microsomal aminopeptidase middle and 3' oligos design.

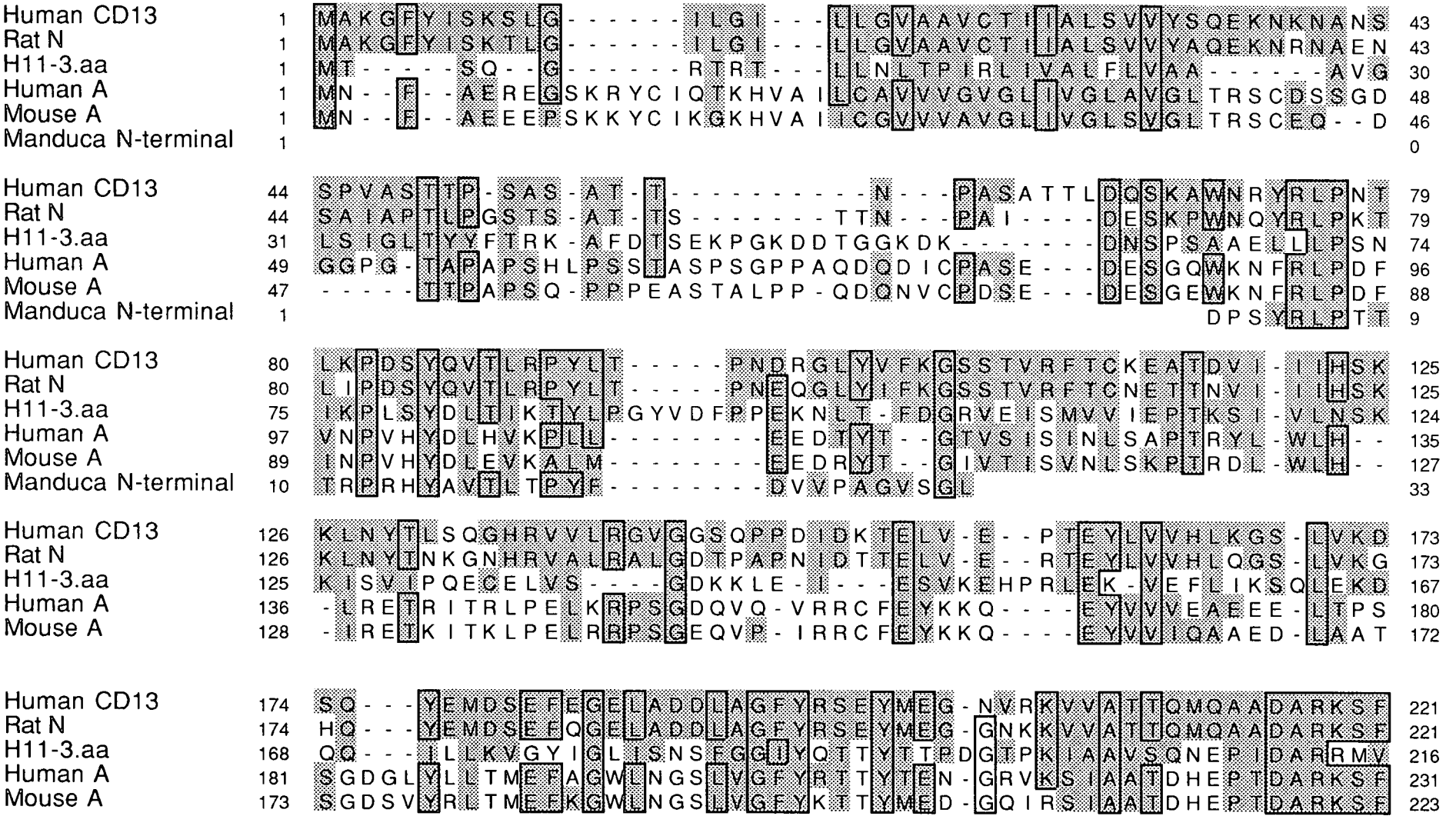


Figure 81. Alignment of the N-terminal regions of some microsomal aminopeptidases with the N-terminal sequence of a *Manduca* aminopeptidase.

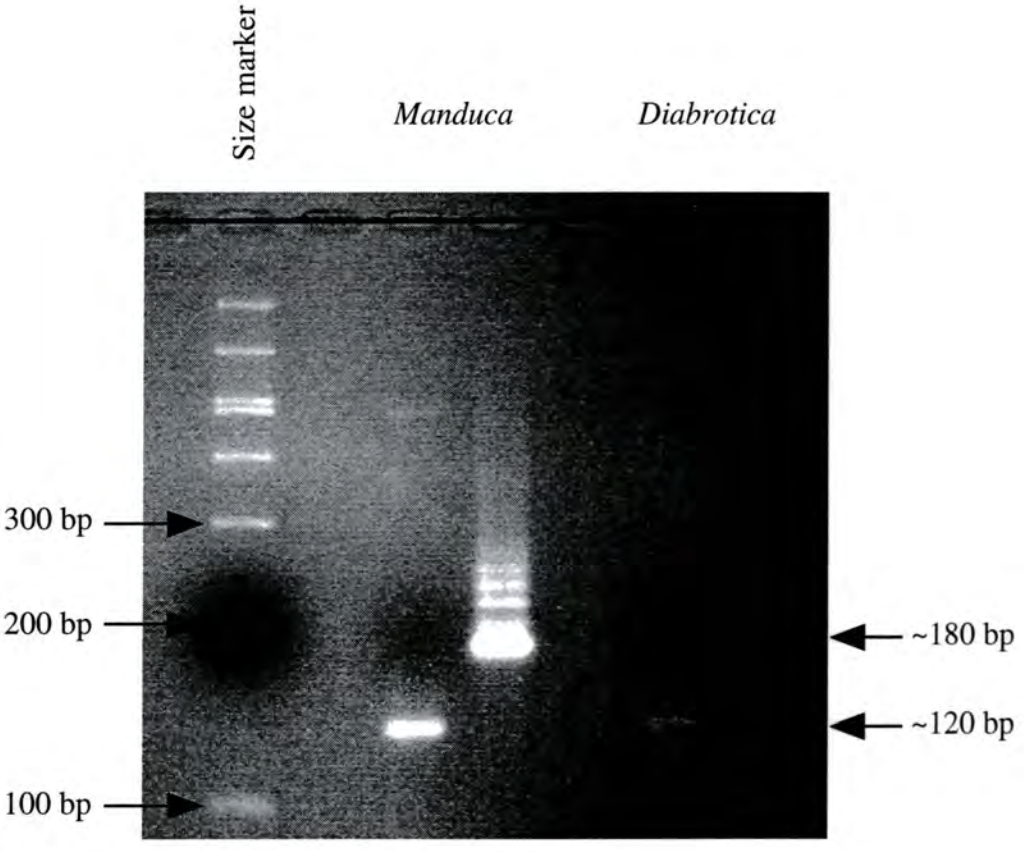


Figure 82. Agarose gel of the products of the PCR reactions on the *Manduca* and *Diabrotica* cDNA templates using the microsomal aminopeptidase strongly conserved region oligos.

The protein sequences of microsomal aminopeptidases from a number of sources downloaded from the protein data bank at Daresbury and those from the patent application (Munn 1992) were compared (figures 78, 79 and 80). A number of reasonably conserved regions were found in all the sequences except that from *E. coli* which, although still recognisably related, was much more divergent. That these regions were conserved in mammals, yeast and a nematode suggested that there was a strong chance that they would also be conserved in insects. Three regions were chosen for the design of oligos on the basis of being conserved and of minimising the degeneracy of the oligos.

A forwards oligo (#952) was made to the sequence GAMENWG, the 5' conserved region, which is at amino acid 350 of the human sequence. It was a 4 fold redundant 20 mer with 2 inosines and a melting point of 60°C (figure 79).

A reverse oligo (#953) was made to the sequence WLNEGFA, the 3' conserved region, which is at amino acid 400 of the human sequence. It was a 16 fold redundant 20 mer with 2 inosines and a melting point of 54°C (figure 80).

A reverse oligo(# 954) was made to the sequence AHQWFGN, the middle conserved region, which is at amino acid 390 of the human sequence. It was an 8 fold redundant 20 mer with 2 inosines and a melting point of 54°C (figure 80).

A forwards oligo (#951) was also designed to the sequence TPYFDVV taken from a short length of N-terminal sequence which had been reported of a *Manduca sexta* aminopeptidase N (Knight, Crickmore and Ellar 1994) (figure 81). This was a 32 fold redundant 20 mer with 2 inosines and had a melting point of 52°C. It was chosen purely as a specific PCR oligo for *Manduca* and not to match any of the weakly conserved sequences in this region.

PCR was performed using a touchdown protocol with two cycles at each annealing temperature and dropping in 1°C steps starting at 65°C and finishing with 25 cycles at 50°C. A touchdown protocol was chosen because the melting points of the oligos were uncertain due to their redundancy and the inclusion of inosine bases, and also because it is known to reduce the amplification of artifacts, a potential problem with degenerate PCR primers (Innis *et al.* 1990). The template DNAs used were samples of the remaining dscDNA after the synthesis of the *Diabrotica* and *Manduca* cDNA libraries.

Both templates gave the expected size bands of ~180 bp for the 5' and 3' oligos and ~120 bp for the 5' and middle oligos (figure 82), as either the only or the obviously strongest product when analysed on gel. The *Manduca* dscDNA was also used in PCR

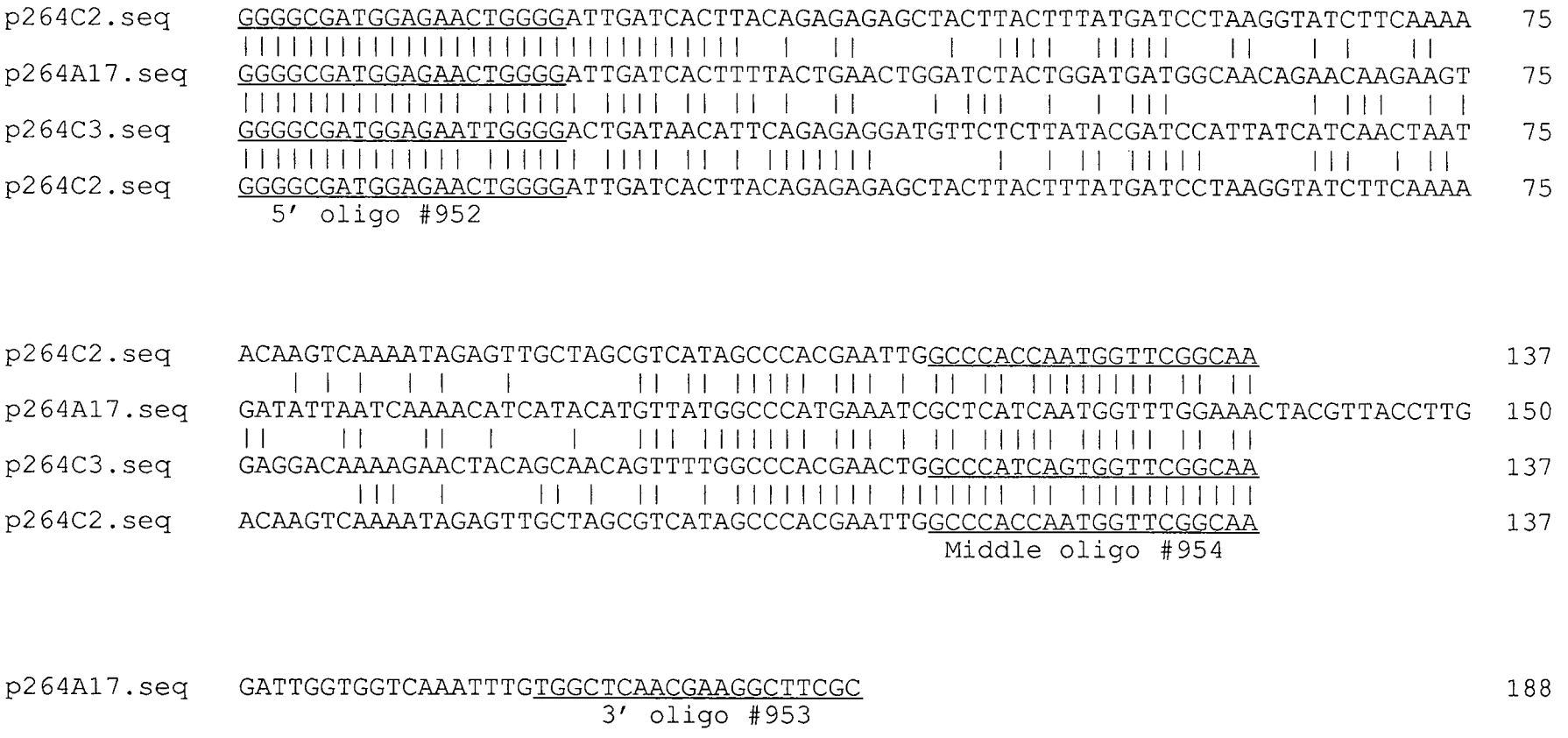


Figure 83. The DNA sequence of the *Diabrotica* microsomal aminopeptidase PCR products and their alignment to each other.

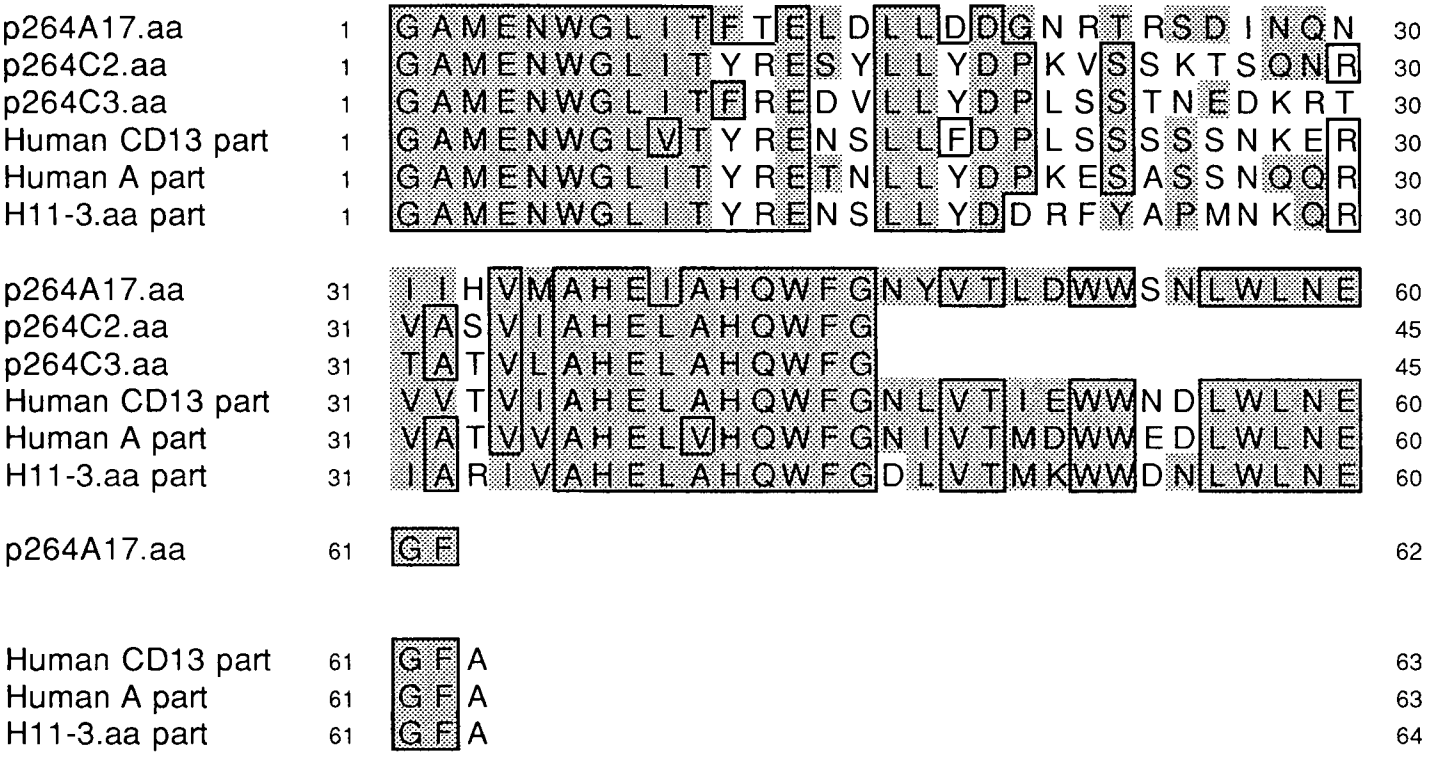


Figure 84. The predicted amino acid sequences of the *Diabrotica* microsomal aminopeptidase PCR products, their alignment to each other and to some MAPs from other sources.

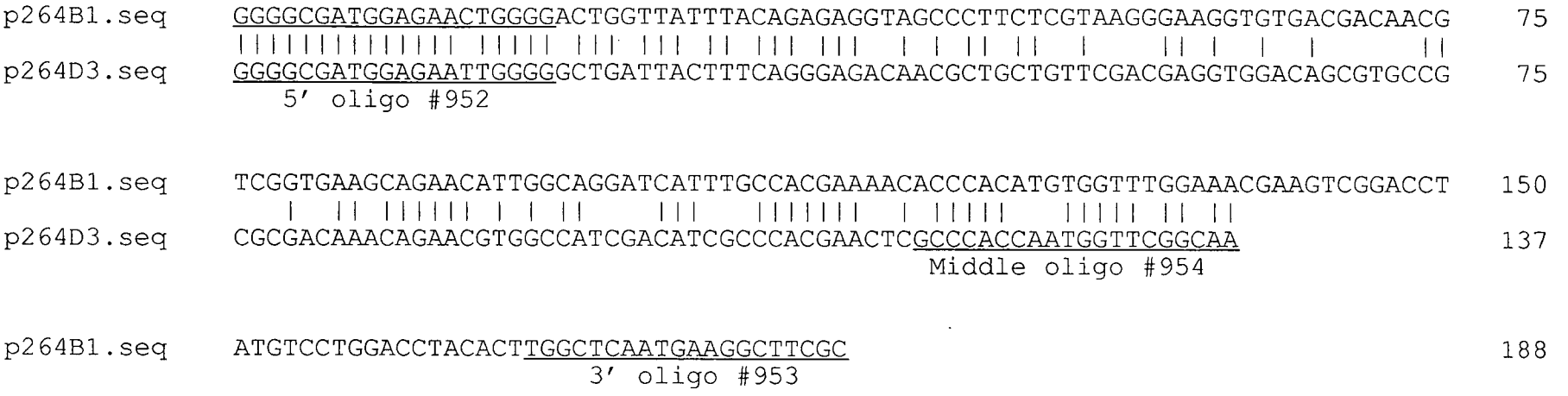


Figure 85. The DNA sequence of the *Manduca* microsomal aminopeptidase PCR products and their alignment to each other.

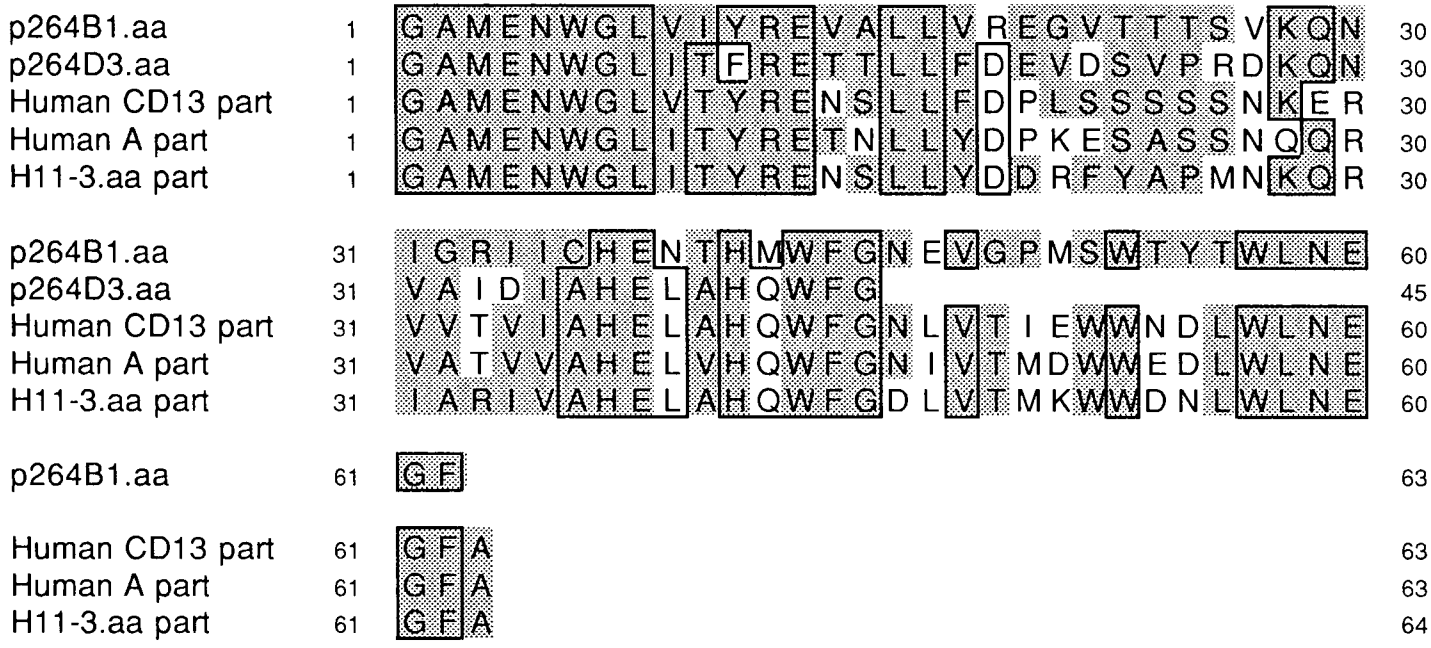


Figure 86. The predicted amino acid sequences of the *Manduca* microsomal aminopeptidase PCR products, their alignment to eachother and to some MAPs from other sources.

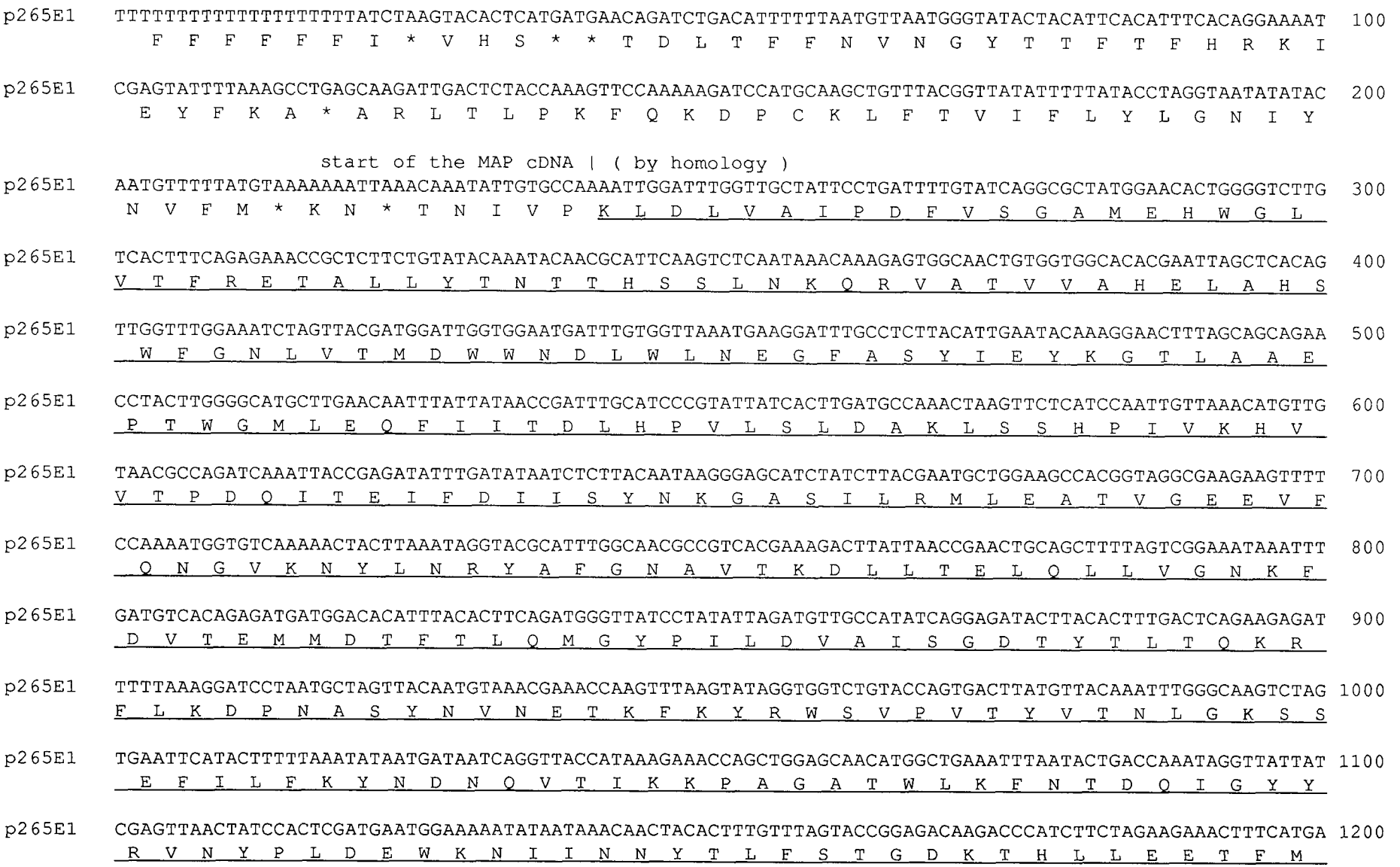
experiments using the *Manduca* N-terminal sequence primer and the middle and 3' conserved region oligos but no products were obtained. An alternate strategy using TetZ (Amersham), a reverse transcribing thermostable polymerase, and *Manduca* polyA+ RNA as template was also attempted with no success. No further attempts to obtain a *Manduca* MAP cDNA were made.

Microsomal aminopeptidase PCR products

The bands of 180 and 120 bp, obtained using primers #952 and #953, and #952 and #954 respectively from both the *Diabrotica* and *Manduca* cDNA templates, were subcloned and the inserted DNA sequenced fully in both directions.

The *Diabrotica* clones were called p264A17 for the 5' to 3' oligos 180 bp PCR product and p264C2 and p264C3 for the 5' to middle oligos 120 bp PCR products. All three were closely homologous to each other and to the published MAP protein sequences but were from different cDNAs (figures 83 and 84). The *Manduca* clones were called p264B1 for the 5' to 3' oligos 180 bp PCR product and p264D1 for the 5' to middle oligos 120 bp PCR product. Again they were closely homologous to each other and to the published MAP protein sequences but non identical (figures 85 and 86).

Fasta searches of the national protein databank were run using each of the predicted protein sequences for the PCR products and in each case the microsomal aminopeptidases were the best match. Lower but also significant scores were observed for the leucotriene A4 hydrolases as is seen for fasta searches with any of the microsomal aminopeptidase protein sequences (data not shown). The significance of the sequence differences in the conserved regions is difficult to interpret without data on enzyme specificity or further protein sequence. The *Diabrotica* sequences are very similar to the mammalian/nematode consensus with only conservative amino acid changes in these strongly conserved regions. The *Manduca* sequences are also close to the mammalian/nematode consensus, however p264B1 has significant sequence changes in the region of the zinc binding motif (Wu *et al.* 1990, Nanus *et al.* 1993, Vallee and Auld 1990) from AHELAHQWFG to CHENTHMWFG (figure 86). The histidine residues which are the zinc ligands are conserved but the substitution of cysteine for alanine, while a conservative amino acid change, could change the coordination of the zinc ion. Many classes of zinc containing enzymes coordinate the zinc through cysteines or cysteine(s) and histidine(s) (Vallee and


p265E1 TTTTTTTTTTTTTTTTTTTTATCTAAGTACACTCATGATGAACAGATCTGACATTTTTTAATGTTAATGGGTATACTACATTCACATTTCACAGGAAAAT 100
F F F F F F I * V H S * * T D L T F F N V N G Y T T F T F H R K I
p265E1 CGAGTATTTTAAAGCCTGAGCAAGATTGACTCTACCAAAGTTCCAAAAAGATCCATGCAAGCTGTTTACGGTTATATTTTTATACCTAGGTAATATATAC 200
E Y F K A * A R L T L P K F Q K D P C K L F T V I F L Y L G N I Y
start of the MAP cDNA | (by homology)
p265E1 AATGTTTTTATGTAAAAAAATTAAACAAATATTGTGCCAAAATTGGATTTGGTTGCTATTCCTGATTTTGTATCAGGCGCTATGGAACACTGGGGTCTTG 300
N V F M * K N * T N I V P K L D L V A I P D F V S G A M E H W G L
p265E1 TCACTTTCAGAGAAACCGCTCTTCTGTATACAAATACAACGCATTCAAGTCTCAATAAACAAAGAGTGGCAACTGTGGTGGCACACGAATTAGCTCACAG 400
V T F R E T A L L Y T N T T H S S L N K Q R V A T V V A H E L A H S
p265E1 TTGGTTTGGAAATCTAGTTACGATGGATTGGTGGAATGATTTGTGGTTAAATGAAGGATTTGCCTCTTACATTGAATACAAAGGAACTTTAGCAGCAGAA 500
W F G N L V T M D W W N D L W L N E G F A S Y I E Y K G T L A A E
p265E1 CCTACTTGGGGCATGCTTGAACAATTTATTATAACCGATTTGCATCCCGTATTATCACTTGATGCCAAACTAAGTTCTCATCCAATTGTTAAACATGTTG 600
P T W G M L E Q F I I T D L H P V L S L D A K L S S H P I V K H V
p265E1 TAACGCCAGATCAAATTACCGAGATATTTGATATAATCTCTTACAATAAGGGAGCATCTATCTTACGAATGCTGGAAGCCACGGTAGGCGAAGAAGTTTT 700
V T P D Q I T E I F D I I S Y N K G A S I L R M L E A T V G E E V F
p265E1 CCAAAATGGTGTCAAAAACTACTTAAATAGGTACGCATTTGGCAACGCCGTCACGAAAGACTTATTAACCGAACTGCAGCTTTTAGTCGGAAATAAATTT 800
Q N G V K N Y L N R Y A F G N A V T K D L L T E L Q L L V G N K F
p265E1 GATGTCACAGAGATGATGGACACATTTACACTTCAGATGGGTTATCCTATATTAGATGTTGCCATATCAGGAGATACTTACACTTTGACTCAGAAGAGAT 900
D V T E M M D T F T L Q M G Y P I L D V A I S G D T Y T L T Q K R
p265E1 TTTTAAAGGATCCTAATGCTAGTTACAATGTAAACGAAACCAAGTTTAAGTATAGGTGGTCTGTACCAGTGACTTATGTTACAAATTTGGGCAAGTCTAG 1000
F L K D P N A S Y N V N E T K F K Y R W S V P V T Y V T N L G K S S
p265E1 TGAATTCATACTTTTTAAATATAATGATAATCAGGTTACCATAAAGAAACCAGCTGGAGCAACATGGCTGAAATTTAATACTGACCAAATAGGTTATTAT 1100
E F I L F K Y N D N Q V T I K K P A G A T W L K F N T D Q I G Y Y
p265E1 CGAGTTAACTATCCACTCGATGAATGGAAAAATATAATAAACAACTACACTTTGTTTAGTACCGGAGACAAGACCCATCTTCTAGAAGAAACTTTCATGA 1200
R V N Y P L D E W K N I I N N Y T L F S T G D K T H L L E E T F M

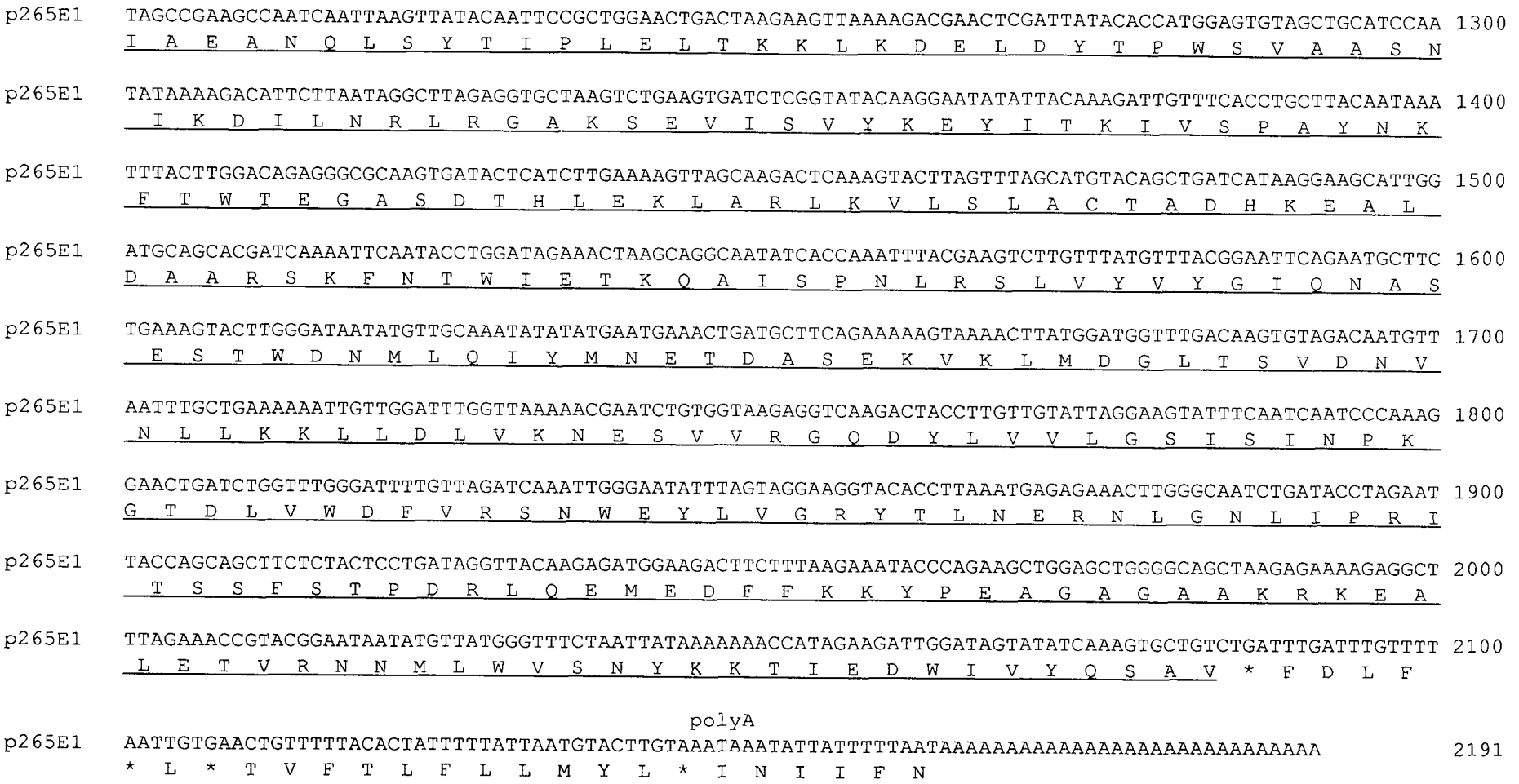


Figure 87. The DNA sequence of p265E1 including the 5' artifact cDNA. The microsomal aminopeptidase reading frame is underlined.

```
p256E1 artifact  TGGCACAATATTTGTTTAATTTTTTTACATAAAAACATTGTATATATTACCTAGGTATAAAAATATAACCGTAAA   75
                  W  H  N  I  C  L  I  F  L  H  K  N  I  V  Y  I  T  *  V  *  K  Y  N  R  K
                   G  T  I  F  V  *  F  F  Y  I  K  T  L  Y  I  L  P  R  Y  K  N  I  T  V  N
                    A  Q  Y  L  F  N  F  F  T  *  K  H  C  I  Y  Y  L  G  I  K  I  *  P  *

p256E1 artifact  CAGCTTGCATGGATCTTTTTGGAACTTTGGTAGAGTCAATCTTGCTCAGGCTTTAAAATACTCGATTTTCCTGTG  150
                  Q  L  A  W  I  F  L  E  L  W  *  S  Q  S  C  S  G  F  K  I  L  D  F  P  V
                   S  L  H  G  S  F  W  N  F  G  R  V  N  L  A  Q  A  L  K  Y  S  I  F  L  *
                 T  A  C  M  D  L  F  G  T  L  V  E  S  I  L  L  R  L  *  N  T  R  F  S  C

p256E1 artifact  AAATGTGAATGTAGTATACCCATTAACATTAAAAAATGTCAGATCTGTTCATCATGAGTGTACTTAGATAAAAAA  225
                  K  C  E  C  S  I  P  I  N  I  K  K  C  Q  I  C  S  S  *  V  Y  L  D
                   N  V  N  V  V  Y  P  L  T  L  K  N  V  R  S  V  H  H  E  C  T  *  I
                 E  M  *  M  *  Y  T  H  *  H  *  K  M  S  D  L  F  I  M  S  V  L  R  *

p256E1 artifact  AAAAAAAAAAAAAA                                                          239
```

Figure 88. The DNA sequence and its translation in all forward reading frames of the p265E1 artifact cDNA. The start of this sequence was decided by the end of the honmology of the p265E1 MAP sequence to the other MAP proteins.

Auld 1990). A change in the coordination chemistry of the active site zinc thus makes it possible that the enzyme activity of the protein corresponding to p264B1 will be significantly different to that of other aminopeptidases.

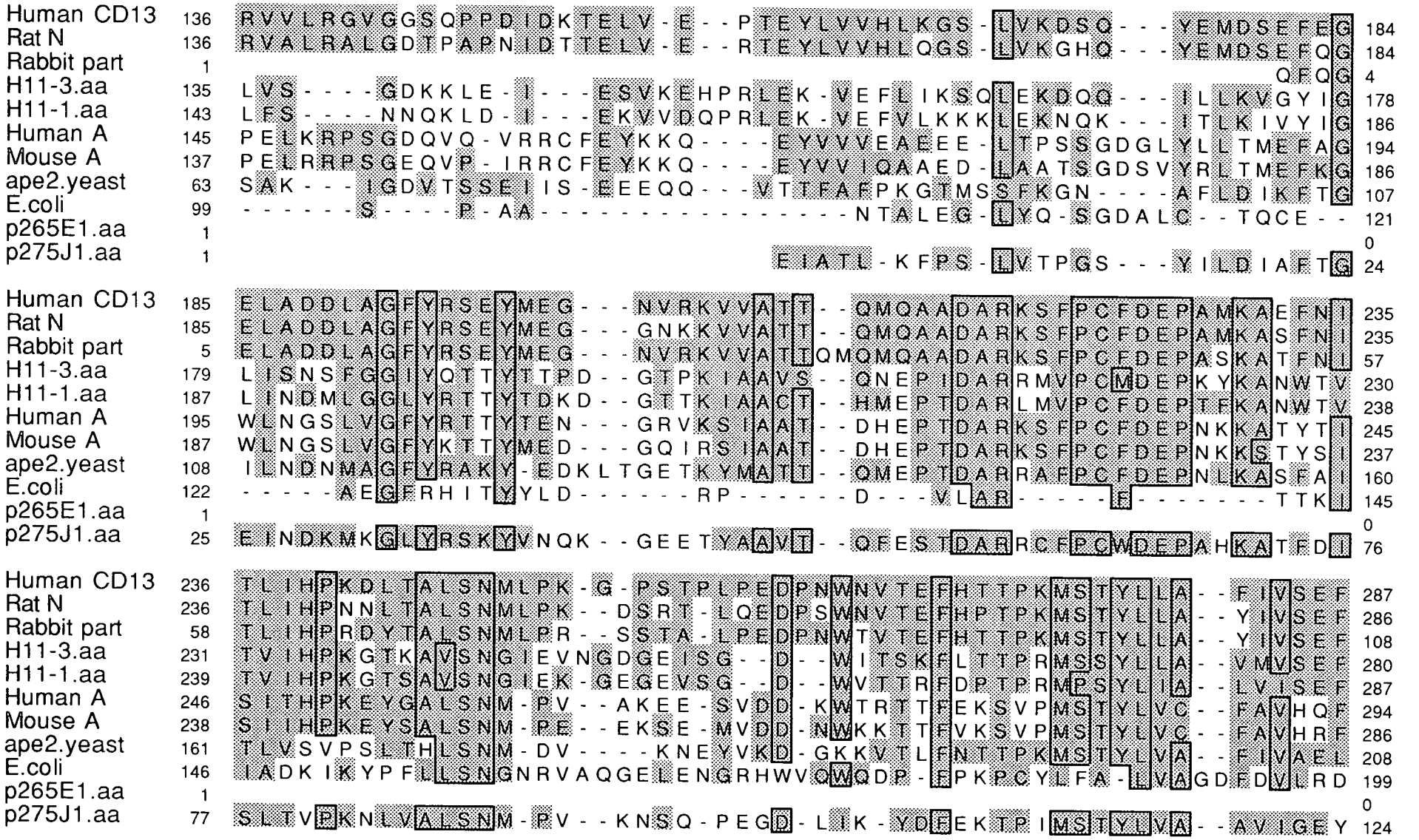
Screening the *Diabrotica* cDNA library with p264A17 the 180 bp PCR product

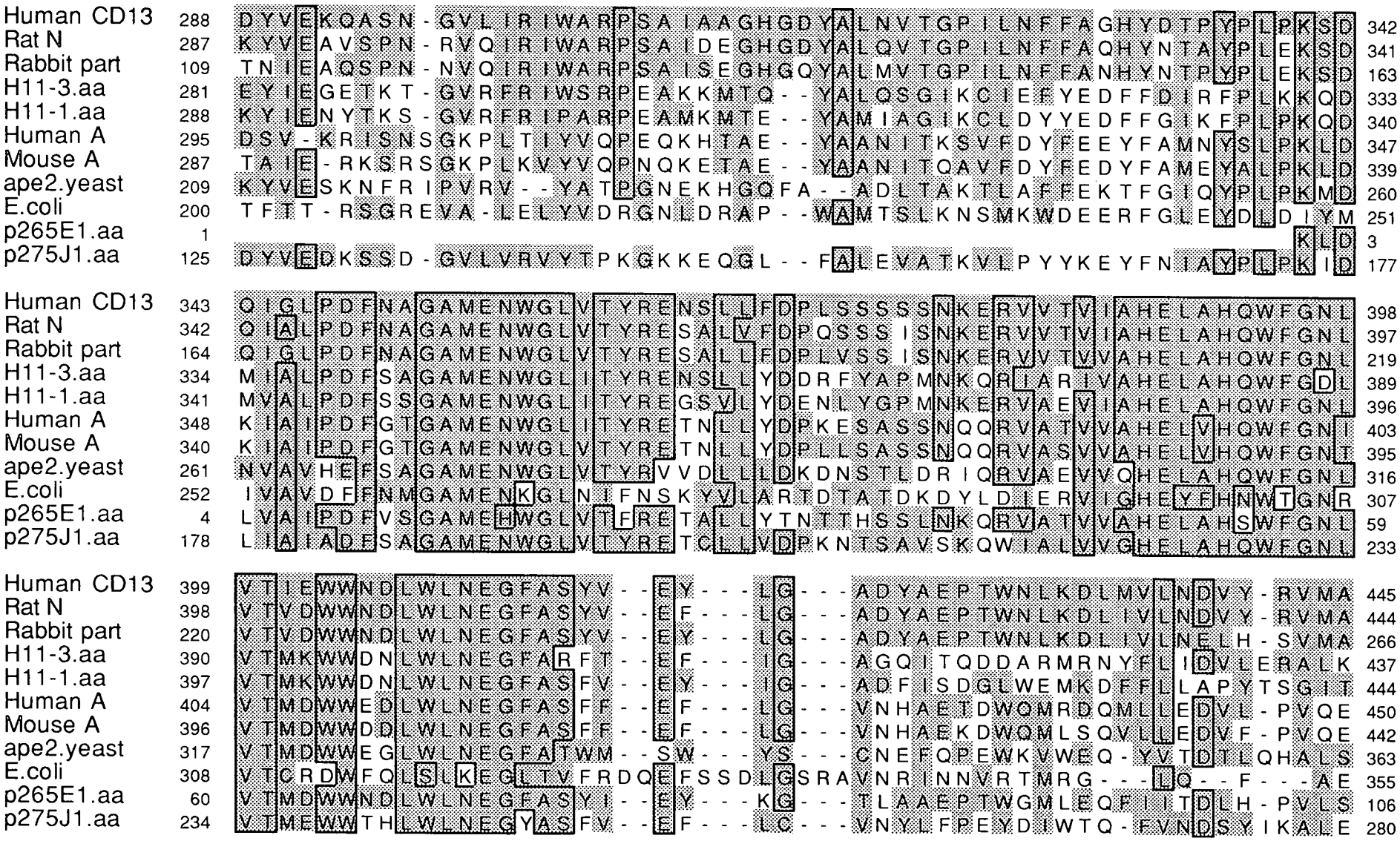
The *Diabrotica* cDNA library was screened initially with the longer of the PCR products, p264A17, which was labelled by the PCR protocol described in the methods. Experimental conditions were as usual except that the temperature for all steps was dropped to 60°C because of the short length of the probe (~180 bp); final washing was to 1 x SSC, 0.1% SDS. From four filters each carrying ~1.5 x 10^5 plaques three positives were recovered and plaque purified.

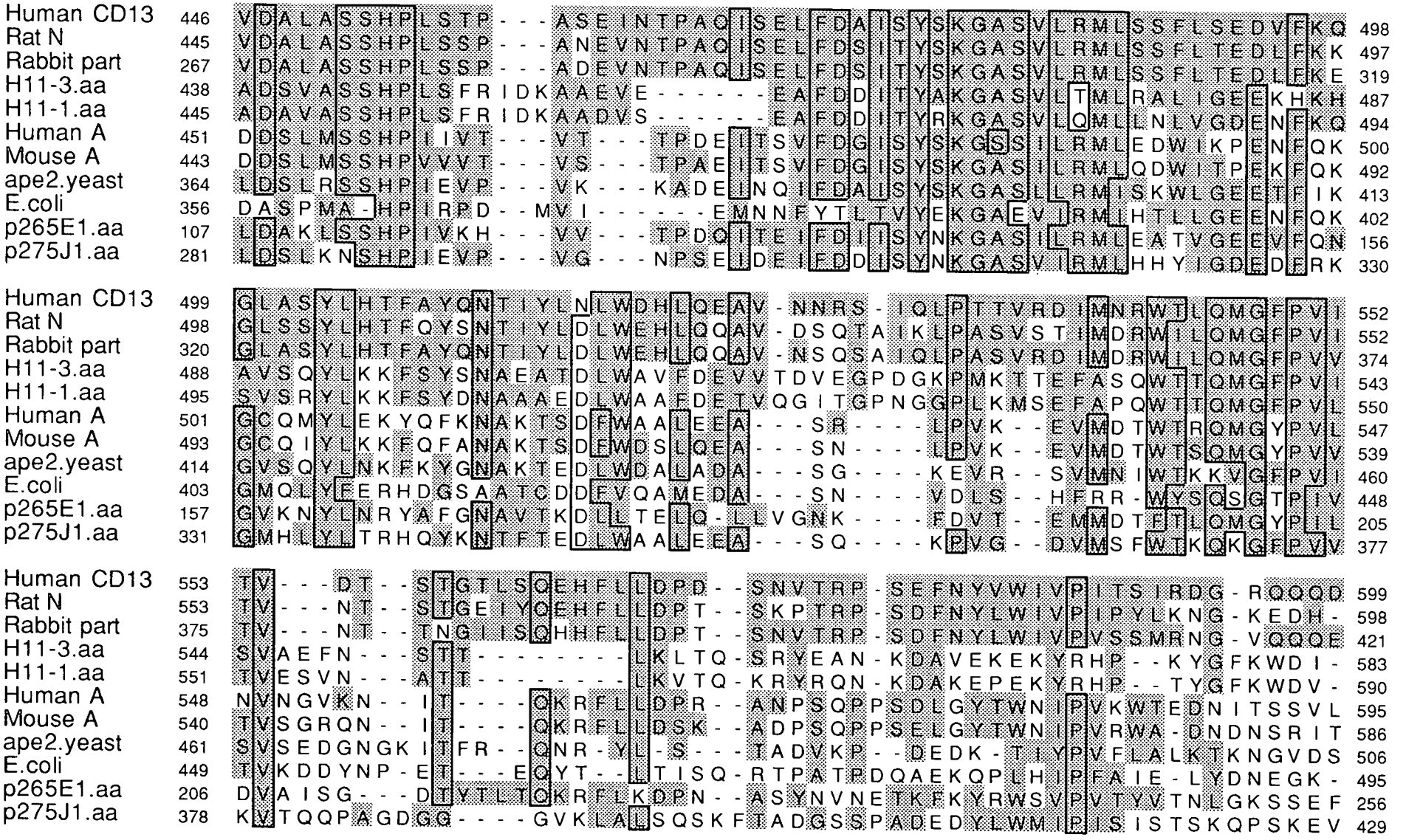
These were plasmid rescued as p265E1, p265E2 and p265G1, restriction mapped and DNA sequenced from each end. It was shown that the same cDNA clone had most probably been isolated three times as they were identical, starting at the same bases in either direction and matching base for base as far as could be read. The only exception was p265G1 which had a one base shorter polyA tail at the EcoR I end. This was probably due to the loss of a base during the single stranded replication phase of the plasmid rescue procedure, as M13 type phages with long polyA sequences in their inserts have been observed, many times previously in this department, to be unstable to partial loss of the polyA tail. Although this was not an exhaustive screen of the library, which consists of 1.0 x 10^6 independent plaques, because the same clone was isolated three times it is unlikely that an homologous clone was missed.

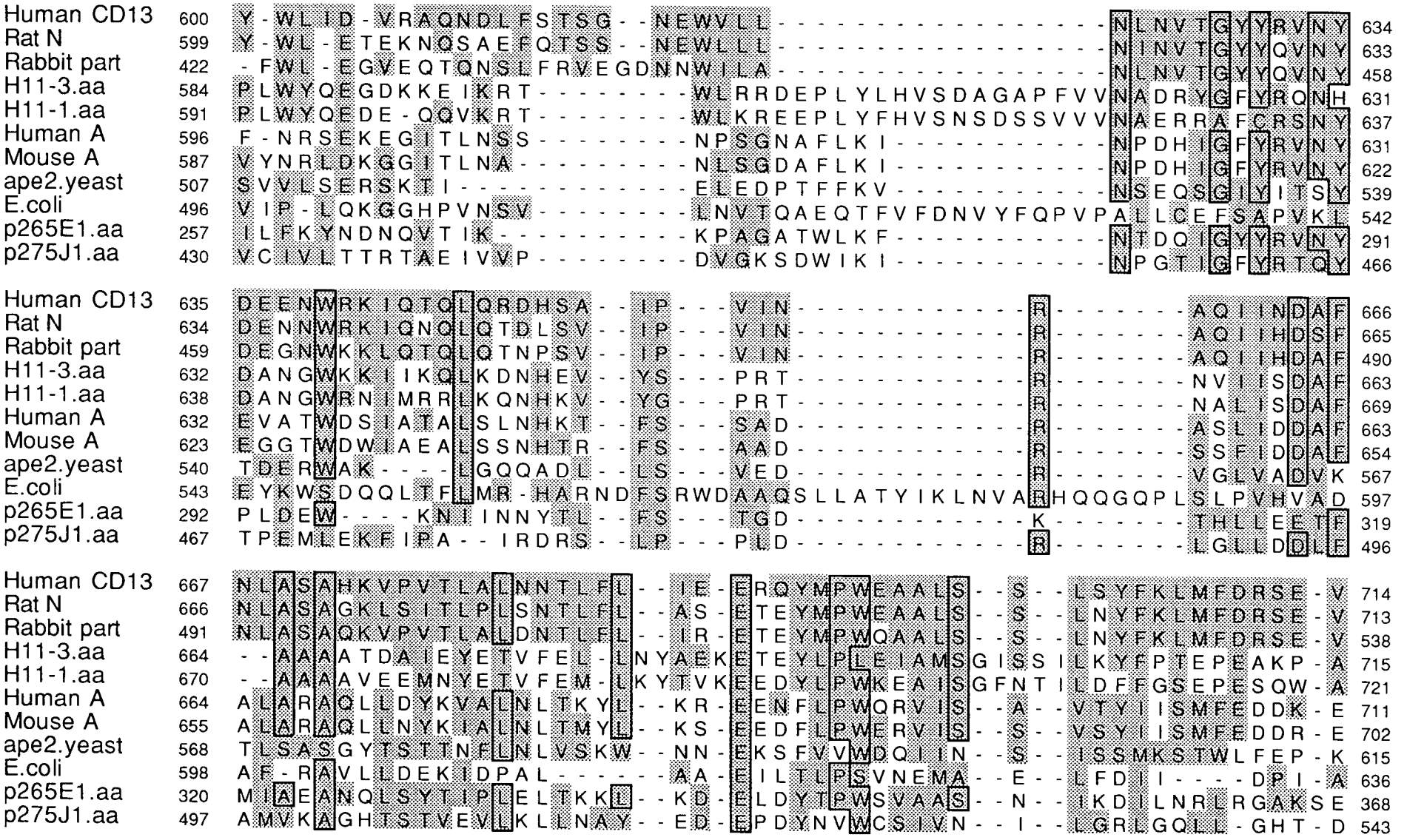
Because the clones were identical, only p265E1 was completely sequenced (figure 87). The whole insert was 2191 bp in length and it consisted of two cDNAs fused together head to head with a polyA tail at either end. The Xho I end, normally the 3' end, had an undamaged cloning linker, a 20 bp polyA tail and a cDNA sequence which ran for approximately 240 bp to the fusion point with the other cDNA. There was only one open reading frame running in from the 5' end of enough length to run a fasta search of the protein data base at Daresbury and this did not significantly match anything. This cDNA was named the p265E1 artifact sequence and no further work was done on it (figure 88).

The other cDNA, which is referred to as the p265E1 sequence or just p265E1, was in the reverse direction for λZAP II and ran from base ~240 to the EcoR I cloning


Human CD13 136 RVVLRGVGGSQPPDIDKTELV--PTEYLVVHLKGS-LVKDSQ---YEMDSEFEG 184
Rat N 136 RVALRALGDTPAPNIDTTELV--RTEYLVVHLQGS-LVKGHQ---YEMDSEFQG 184
Rabbit part 1 QFQG 4
H11-3.aa 135 LVS----GDKKLE-I---ESVKEHPRLEK-VEFLIKSQLEKDQQ---ILLKVGYIG 178
H11-1.aa 143 LFS----NNQKLD-I---EKVVDQPRLEK-VEFVLKKKLEKNQK---ITLKIVYIG 186
Human A 145 PELKRPSGDQVQ-VRRCFEYKKQ----EYVVVEAEEE-LTPSSGDGLYLLTMEFAG 194
Mouse A 137 PELRRPSGEQVP-IRRCFEYKKQ----EYVVIQAAED-LAATSGDSVYRLTMEFKG 186
ape2.yeast 63 SAK---IGDVTSSEIIS-EEEQQ---VTTFAFPKGTMSSFKGN----AFLDIKFTG 107
E.coli 99 ------S----P-AA----------NTALEG-LYQ-SGDALC--TQCE-- 121
p265E1.aa 1 0
p275J1.aa 1 EIATL-KFPS-LVTPGS---YILDIAFTG 24
Human CD13 185 ELADDLAGFYRSEYMEG---NVRKVVATT--QMQAADARKSFPCFDEPAMKAEFNI 235
Rat N 185 ELADDLAGFYRSEYMEG---GNKKVVATT--QMQAADARKSFPCFDEPAMKASFNI 235
Rabbit part 5 ELADDLAGFYRSEYMEG---NVRKVVATTQMQMQAADARKSFPCFDEPASKATFNI 57
H11-3.aa 179 LISNSFGGIYQTTYTTPD--GTPKIAAVS--QNEPIDARRMVPCMDEPKYKANWTV 230
H11-1.aa 187 LINDMLGGLYRTTYTDKD--GTTKIAACT--HMEPTDARLMVPCFDEPTFKANWTV 238
Human A 195 WLNGSLVGFYRTTYTEN---GRVKSIAAT--DHEPTDARKSFPCFDEPNKKATYTI 245
Mouse A 187 WLNGSLVGFYKTTYMED---GQIRSIAAT--DHEPTDARKSFPCFDEPNKKSTYSI 237
ape2.yeast 108 ILNDNMAGFYRAKY-EDKLTGETKYMATT--QMEPTDARRAFPCFDEPNLKASFAI 160
E.coli 122 -----AEGFRHITYYLD------RP------D---VLAR-----F--------TTKI 145
p265E1.aa 1 0
p275J1.aa 25 EINDKMKGLYRSKYVNQK--GEETYAAVT--QFESTDARRCFPCWDEPAHKATFDI 76
Human CD13 236 TLIHPKDLTALSNMLPK-G-PSTPLPEDPNWNVTEFHTTPKMSTYLLA--FIVSEF 287
Rat N 236 TLIHPNNLTALSNMLPK--DSRT-LQEDPSWNVTEFHPTPKMSTYLLA--YIVSEF 286
Rabbit part 58 TLIHPRDYTALSNMLPR--SSTA-LPEDPNWTVTEFHTTPKMSTYLLA--YIVSEF 108
H11-3.aa 231 TVIHPKGTKAVSNGIEVNGDGEISG--D--WITSKFLTTPRMSSYLLA--VMVSEF 280
H11-1.aa 239 TVIHPKGTSAVSNGIEK-GEGEVSG--D--WVTTRFDPTPRMPSYLIA--LVISEF 287
Human A 246 SITHPKEYGALSNM-PV--AKEE-SVDD-KWTRTTFEKSVPMSTYLVC--FAVHQF 294
Mouse A 238 SIIHPKEYSALSNM-PE--EKSE-MVDD-NWKKTTFVKSVPMSTYLVC--FAVHRF 286
ape2.yeast 161 TLVSVPSLTHLSNM-DV----KNEYVKD-GKKVTLFNTTPKMSTYLVA--FIVAEL 208
E.coli 146 IADKIKYPFLLSNGNRVAQGELENGRHWVQWQDP-FPKPCYLFA-LVAGDFDVLRD 199
p265E1.aa 1 0
p275J1.aa 77 SLTVPKNLVALSNM-PV--KNSQ-PEGD-LIK-YDFEKTPIMSTYLVA--AVIGEY 124


Human CD13 288 DYVEKQASN-GVLIRIWARPSAIAAGHGDYALNVTGPILNFFAGHYDTPYPLPKSD 342
Rat N 287 KYVEAVSPN-RVQIRIWARPSAIDEGHGDYALQVTGPILNFFAQHYNTAYPLEKSD 341
Rabbit part 109 TNIEAQSPN-NVQIRIWARPSAISEGHGQYALMVTGPILNFFANHYNTPYPLEKSD 163
H11-3.aa 281 EYIEGETKT-GVRFRIWSRPEAKKMTQ--YALQSGIKCIEFYEDFFDIRFPLKKQD 333
H11-1.aa 288 KYIENYTKS-GVRFRIPARPEAMKMTE--YAMIAGIKCLDYYEDFFGIKFPLPKQD 340
Human A 295 DSV-KRISNSGKPLTIYVQPEQKHTAE--YAANITKSVFDYFEEYFAMNYSLPKLD 347
Mouse A 287 TAIE-RKSRSGKPLKVYVQPNQKETAE--YAANITQAVFDYFEDYFAMEYALPKLD 339
ape2.yeast 209 KYVESKNFRIPVRV--YATPGNEKHGQFA--ADLTAKTLAFFEKTFGIQYPLPKMD 260
E.coli 200 TFTT-RSGREVA-LELYVDRGNLDRAP--WAMTSLKNSMKWDEERFGLEYDLDIYM 251
p265E1.aa 1 KLD 3
p275J1.aa 125 DYVEDKSSD-GVLVRVYTPKGKKEQGL--FALEVATKVLPYYKEYFNIAYPLPKID 177
Human CD13 343 QIGLPDFNAGAMENWGLVTYRENSLLFDPLSSSSSNKERVVTVIAHELAHQWFGNL 398
Rat N 342 QIALPDFNAGAMENWGLVTYRESALVFDPQSSSISNKERVVTVIAHELAHQWFGNL 397
Rabbit part 164 QIGLPDFNAGAMENWGLVTYRESALLFDPLVSSISNKERVVTVVAHELAHQWFGNL 219
H11-3.aa 334 MIALPDFSAGAMENWGLITYRENSLLYDDRFYAPMNKQRIARIVAHELAHQWFGDL 389
H11-1.aa 341 MVALPDFSSGAMENWGLITYREGSVLYDENLYGPMNKERVAEVIAHELAHQWFGNL 396
Human A 348 KIAIPDFGTGAMENWGLITYRETNLLYDPKESASSNQQRVATVVAHELVHQWFGNI 403
Mouse A 340 KIAIPDFGTGAMENWGLVTYRETNLLYDPLLSASSNQQRVASVVAHELVHQWFGNT 395
ape2.yeast 261 NVAVHEFSAGAMENWGLVTYRVVDLLLDKDNSTLDRIQRVAEVVQHELAHQWFGNL 316
E.coli 252 IVAVDFFNMGAMENKGLNIFNSKYVLARTDTATDKDYLDIERVIGHEYFHNWTGNR 307
p265E1.aa 4 LVAIPDFVSGAMEHWGLVTFRETALLYTNTTHSSLNKQRVATVVAHELAHSWFGNL 59
p275J1.aa 178 LIAIADFSAGAMENWGLVTYRETCLLVDPKNTSAVSKQWIALVVGHELAHQWFGNL 233
Human CD13 399 VTIEWWNDLWLNEGFASYV--EY---LG---ADYAEPTWNLKDLMVLNDVY-RVMA 445
Rat N 398 VTVDWWNDLWLNEGFASYV--EF---LG---ADYAEPTWNLKDLIVLNDVY-RVMA 444
Rabbit part 220 VTVDWWNDLWLNEGFASYV--EY---LG---ADYAEPTWNLKDLIVLNDVY-RVMA 444
H11-3.aa 390 VTMKWWDNLWLNEGFARFT--EF---IG---AGQITQDDARMRNYFLIDVLERALK 437
H11-1.aa 397 VTMKWWDNLWLNEGFASFV--EY---IG---ADFISDGLWEMKDFFLLAPYTSGIT 444
Human A 404 VTMDWWEDLWLNEGFASFF--EF---LG---VNHAETDWQMRDQMLLEDVL-PVQE 450
Mouse A 396 VTMDWWDDLWLNEGFASFF--EF---LG---VNHAEKDWQMLSQVLLEDVF-PVQE 442
ape2.yeast 317 VTMDWWEGLWLNEGFATWM--SW---YS---CNEFQPEWKVWEQ-YVTDTLQHALS 363
E.coli 308 VTCRDWFQLSLKEGLTVFRDQEFSSDLGSRAVNRINNVRTMRG---LQ--F---AE 355
p265E1.aa 60 VTMDWWNDLWLNEGFASYI--EY---KG---TLAAEPTWGMLEQFIITDLH-PVLS 106
p275J1.aa 234 VTMEWWTHLWLNEGYASFV--EF---LC---VNYLFPEYDIWTQ-FVNDSYIKALE 280


Human CD13 446 VDALASSHPLSTP---ASEINTPAQISELFDAISYSKGASVLRMLSSFLSEDVFKQ 498
Rat N 445 VDALASSHPLSSP---ANEVNTPAQISELFDSITYSKGASVLRMLSSFLTEDLFKK 497
Rabbit part 267 VDALASSHPLSSP---ADEVNTPAQISELFDSITYSKGASVLRMLSSFLTEDLFKE 319
H11-3.aa 438 ADSVASSHPLSFRIDKAAEVE------EAFDDITYAKGASVLTMLRALIGEEKHKH 487
H11-1.aa 445 ADAVASSHPLSFRIDKAADVS------EAFDDITYRKGASVLQMLLNLVGDENFKQ 494
Human A 451 DDSLMSSHPIIVT---VT---TPDEITSVFDGISYSKGSSILRMLEDWIKPENFQK 500
Mouse A 443 DDSLMSSHPVVVT---VS---TPAEITSVFDGISYSKGASILRMLQDWITPEKFQK 492
ape2.yeast 364 LDSLRSSHPIEVP---VK---KADEINQIFDAISYSKGASLLRMISKWLGEETFIK 413
E.coli 356 DASPMA-HPIRPD--MVI------EMNNFYTLTVYEKGAEVIRMLHTLLGEENFQK 402
p265E1.aa 107 LDAKLSSHPIVKH---VV---TPDQITEIFDIISYNKGASILRMLEATVGEEVFQN 156
p275J1.aa 281 LDSLKNSHPIEVP---VG---NPSEIDEIFDDISYNKGASVIRMLHHYIGDEDFRK 330
Human CD13 499 GLASYLHTFAYQNTIYLNLWDHLQEAV-NNRS-IQLPTTVRDIMNRWTLQMGFPVI 552
Rat N 498 GLSSYLHTFQYSNTIYLDLWEHLQQAV-DSQTAIKLPASVSTIMDRWILQMGFPVI 552
Rabbit part 320 GLASYLHTFAYQNTIYLDLWEHLQQAV-NSQSAIQLPASVRDIMDRWILQMGFPVV 374
H11-3.aa 488 AVSQYLKKFSYSNAEATDLWAVFDEVVTDVEGPDGKPMKTTEFASQWTTQMGFPVI 543
H11-1.aa 495 SVSRYLKKFSYDNAAAEDLWAAFDETVQGITGPNGGPLKMSEFAPQWTTQMGFPVL 550
Human A 501 GCQMYLEKYQFKNAKTSDFWAALEEA---SR----LPVK--EVMDTWTRQMGYPVL 547
Mouse A 493 GCQIYLKKFQFANAKTSDFWDSLQEA---SN----LPVK--EVMDTWTSQMGYPVV 539
ape2.yeast 414 GVSQYLNKFKYGNAKTEDLWDALADA---SG----KEVR--SVMNIWTKKVGFPVI 460
E.coli 403 GMQLYFEERHDGSAATCDDFVQAMEDA---SN----VDLS--HFRR-WYSQSGTPIV 448
p265E1.aa 157 GVKNYLNRYAFGNAVTKDLLTELQ-LLVGNK----FDVT--EMMDTFTLQMGYPLL 205
p275J1.aa 331 GMHLYLTRHQYKNTFTEDLWAALEEA---SQ----KPVG--DVMSFWTKQKGFPVV 377
Human CD13 553 TV---DT--STGTLSQEHFLLDPD--SNVTRP-SEFNYVWIVPITSIRDG-RQQQD 599
Rat N 553 TV---NT--STGEIYQEHFLLDPT--SKPTRP-SDFNYLWIVPIPYLKNG-KEDH- 598
Rabbit part 375 TV---NT--TNGIISQHHFLLDPT--SNVTRP-SDFNYLWIVPVSSMRNG-VQQQE 421
H11-3.aa 544 SVAEFN---STT-------LKLTQ-SRYEAN-KDAVEKEKYRHP--KYGFKWDI- 583
H11-1.aa 551 TVESVN---ATT-------LKVTQ-KRYRQN-KDAKEPEKYRHP--TYGFKWDV- 590
Human A 548 NVNGVKN--IT----QKRFLLDPR--ANPSQPPSDLGYTWNIPVKWTEDNITSSVL 595
Mouse A 540 TVSGRQN--IT----QKRFLLDSK--ADPSQPPSELGYTWNIPVRWA-DNDNSRIT 586
ape2.yeast 461 SVSEDGNGKITFR--QNR-YL-S---TADVKP--DEDK-TIYPVFLALKTKNGVDS 506
E.coli 449 TVKDDYNP-ET---EQYT--LTISQ-RTPATPDQAEKQPLHIPFAIE-LYDNEGK- 495
p265E1.aa 206 DVAISG---DTYTLTQKRFLKDPN--ASYNVNETKFKYRWSVPVTYVTNLGKSSEF 256
p275J1.aa 378 KVTQQPAGDGG----GVKLALSQSKFTADGSSPADEDYLWMIPIISISTSKQPSKEV 429


Human CD13 600 Y-WLID-VRAQNDLFSTSG--NEWVLL----------NLNVTGYYRVNY 634
Rat N 599 Y-WL-ETEKNQSAEFQTSS--NEWLLL----------NINVTGYYQVNY 633
Rabbit part 422 -FWL-EGVEQTQNSLFRVEGDNNWILA----------NLNVTGYYQVNY 458
H11-3.aa 584 PLWYQEGDKKEIKRT--------WLRRDEPLYLHVSDAGAPFVVNADRYGFYRQNH 631
H11-1.aa 591 PLWYQEDE-QQVKRT--------WLKREEPLYFHVSNSDSSVVVNAERRAFCRSNY 637
Human A 596 F-NRSEKEGITLNSS--------NPSGNAFLKI----------NPDHIGFYRVNY 631
Mouse A 587 VYNRLDKGGITLNA--------NLSGDAFLKI----------NPDHIGFYRVNY 622
ape2.yeast 507 SVVLSERSKTI----------ELEDPTFFKV----------NSEQSGIYITSY 539
E.coli 496 VIP-LQKGGHPVNSV--------LNVTQAEQTFVFDNVYFQPVPALLCEFSAPVKL 542
p265E1.aa 257 ILFKYNDNQVTIK--------KPAGATWLKF----------NTDQIGYYRVNY 291
p275J1.aa 430 VCIVLTTRTAEIVVP--------DVGKSDWIKI----------NPGTIGFYRTQY 466
Human CD13 635 DEENWRKIQTQLQRDHSA--IP---VIN----------R-------AQIINDAF 666
Rat N 634 DENNWRKIQNQLQTDLSV--IP---VIN----------R-------AQIIHDSF 665
Rabbit part 459 DEGNWKKLQTQLQTNPSV--IP---VIN----------R-------AQIIHDAF 490
H11-3.aa 632 DANGWKKIIKQLKDNHEV--YS---PRT----------R-------NVIISDAF 663
H11-1.aa 638 DANGWRNIMRRLKQNHKV--YG---PRT----------R-------NALISDAF 669
Human A 632 EVATWDSIATALSLNHKT--FS---SAD----------R-------ASLIDDAF 663
Mouse A 623 EGGTWDWIAEALSSNHTR--FS---AAD----------R-------SSFIDDAF 654
ape2.yeast 540 TDERWAK----LGQQQADL--LS---VED----------R-------VGLVADVK 567
E.coli 543 EYKWSDQQLTFLMR-HARNDFSRWDAAQSLLATYIKLNVARHQQGQPLSLPVHVAD 597
p265E1.aa 292 PLDEW----KNIINNYTL--FS---TGD----------K-------THLLEETF 319
p275J1.aa 467 TPEMLEKFIPA--IRDRS--LP---PLD----------R-------LGLLDDLF 496
Human CD13 667 NLASAHKVPVTLALNNTLFL--IE-ERQYMPWEAALS--S--LSYFKLMFDRSE-V 714
Rat N 666 NLASAGKLSITLPLSNTLFL--AS-ETEYMPWEAALS--S--LNYFKLMFDRSE-V 713
Rabbit part 491 NLASAQKVPVTLALDNTLFL--IR-ETEYMPWQAALS--S--LNYFKLMFDRSE-V 538
H11-3.aa 664 --AAAATDAIEYETVFEL-LNYAEKETEYLPLEIAMSGISSILKYFPTEPEAKP-A 715
H11-1.aa 670 --AAAAVEEMNYETVFEM-LKYTVKEEDYLPWKEAISGFNTILDFFGSEPESQW-A 721
Human A 664 ALARAQLLDYKVALNLTKYL--KR-EENFLPWQRVIS--A--VTYIISMFEDDK-E 711
Mouse A 655 ALARAQLLNYKIALNLTMYL--KS-EEDFLPWERVIS--S--VSYIISMFEDDR-E 702
ape2.yeast 568 TLSASGYTSTTNFLNLVSKW--NN-EKSFVVWDQIIN--S--ISSMKSTWLFEP-K 615
E.coli 598 AF-RAVLLDEKIDPAL------AA-EILTLPSVNEMA--E--LFDII----DPI-A 636
p265E1.aa 320 MIAEANQLSYTIPLELTKKL--KD-ELDYTPWSVAAS--N--IKDILNRLRGAKSE 368
p275J1.aa 497 AMVKAGHTSTVEVLKLLNAY--ED-EPDYNVWCSIVN--I--LGRLGQLL-GHT-D 543


Human CD13 715 YGPMKNYLKKQVTPLFIHFRNNTNNWREIPENLMDQYSEVNA-ISTACSN----GV 765
Rat N 714 YGPMKRYLKKQVTPLFAYFKIKTNNWLDRPPTLMEQYNEINA-ISTACSS----GL 764
Rabbit part 539 YGPMKNYLSKQVRPLFEHFKNITNDWTRRPDTLMDQYNEINA-ISTACSN----GI 589
H11-3.aa 716 QTYMMNILKPMYEKSSIDF-IA-NNY--RNDKLFFQINLQKDVIDMFCAL----GS 763
H11-1.aa 722 SEYMRKLMKPIYDKSSIKF-IA-ENY--KKDSLFFKNNLQIAVIDTYCGL----GG 769
Human A 712 LYPMIEEYFQGQVKPIADS-LG---WNDAGDHVT-KL-LRSSVLGFACKM----GD 757
Mouse A 703 LYPMIETYFQGQVKPVADL-LG---WQDTGSHIT-KL-LRASILGFACKM----GD 748
ape2.yeast 616 ETQDALDNFTKQLISGMTHHLG---WEFKSSDSFSTQRLKVTMFGAACAA----RD 664
E.coli 637 IAEVREALTRTLATELADELLAIYNANYQSEYRVEHEDIAKRTLRNACLRFLAFGE 692
p265E1.aa 369 VISVYKEYITKIVSPAYNK-FT---WTEGASDTHLEKLARLKVLSLACTA----DH 416
p275J1.aa 544 YEADYNKYQKRLLSKVYSR-LG---WTQKSGETHLDTLLRGLVLGRLAWL----DH 591
Human CD13 766 PECEEMVSG-LF-K--QW-MENP--NNN-P------IHPNLRSTVYCNAIAQGGEE 807
Rat N 765 EECRDLVVG-LY-S--QW-MNNS--DNN-P------IHPNLRSTVYCNAIASC-EE 805
Rabbit part 590 QECETLVSD-LF-K--QW-MDDP--SNN-P------IHPNLRTTVYCNAIALGGER 631
H11-3.aa 764 QDCRKKYKK-LF-D--DEVMNKCR-DGQ-AATECVRIAAPLRSSVYCYGVK----E 809
H11-1.aa 770 KECLEEMKK-LF-D--KEVM-KCQ-PGQ-QATDCVKVTAPLRKTVYCYGVQ----E 814
Human A 758 REALNNASS-LF-E--QW-LNGT--VSL-P------V--NLRLLVYRYGMQNSGNE 797
Mouse A 749 REALGNASQ-LF-D--SW-LKGS--ASI-P------V--NLRLLVYRYGMQNSGNE 788
ape2.yeast 665 ADVEKAALK-MF-TDYCSG-NKEAIPALIKP-----IVFNTV----ARVG----GAE 705
E.coli 693 THLADVLVSKQFHE--ANNMTDALAALSAAVAAQLPCRDALMQEYDDKWHQNGLVM 746
p265E1.aa 417 KEALDAARS-KF-N--TW-IETK--QAI-S------P--NLRSLVYVYGIQN-ASE 455
p275J1.aa 592 EHTISEARN-RF-D--TH-VKSE--QVL-P------A--DLRSACYK-TIL----R 626
Human CD13 808 E-WDFAWEQFRNA----TLVNEADKLRAALACSKELWILNRYLSYTL--NPDLIRK 856
Rat N 806 A-WNFAWATVPER----TLVNEADKLRSAVGRSNEVWILNRYLSYTL--NPDYIRK 854
Rabbit part 632 E-WDFAWEQFRNA----TLVNEADKLRSALACSNEVWILNRYLSYTL--NPDYIRR 680
H11-3.aa 810 G-GDYASDKVMELYTAETLALEKDFLRLALGCHKDVTALKGLLLRALDRNSSFVRM 864
H11-1.aa 815 G-GDEAFDKVMELYNAEQVQLEKDSLREALGCHKDVTALKGLLMLALDRNSSFVRL 869
Human A 798 ISWNYTLEQYQKT----SLAQEKEKLLYGLASVKNVTLLSRYLDLLKDTN--LIKT 847
Mouse A 789 AAWNYTLEQYQKT----SLAQEKEKLLYGLASVKDVKLLARYLEMLKDPN--IIKT 838
ape2.yeast 706 ---NY--EKVYKIYLDPISNDEKLAALRSLGRFKEPKLLERTLGYLFDGT--VLN- 753
E.coli 747 DKWFILQATSPAANVLETVRGLLQHRSFTMSNPNRIRSLIGAFAGSNPAAFHAEDG 802
p265E1.aa 456 STWDNMLQIYMNE----TDASEKVKLMDGLTSVDNVNLLKKLLDLVKNES--VVRG 505
p275J1.aa 627 AGGQAEYDTLLQLYRSSDLHEEKDRISRAFGATKDPNLLKKVLEFAMSDE--VRS- 679

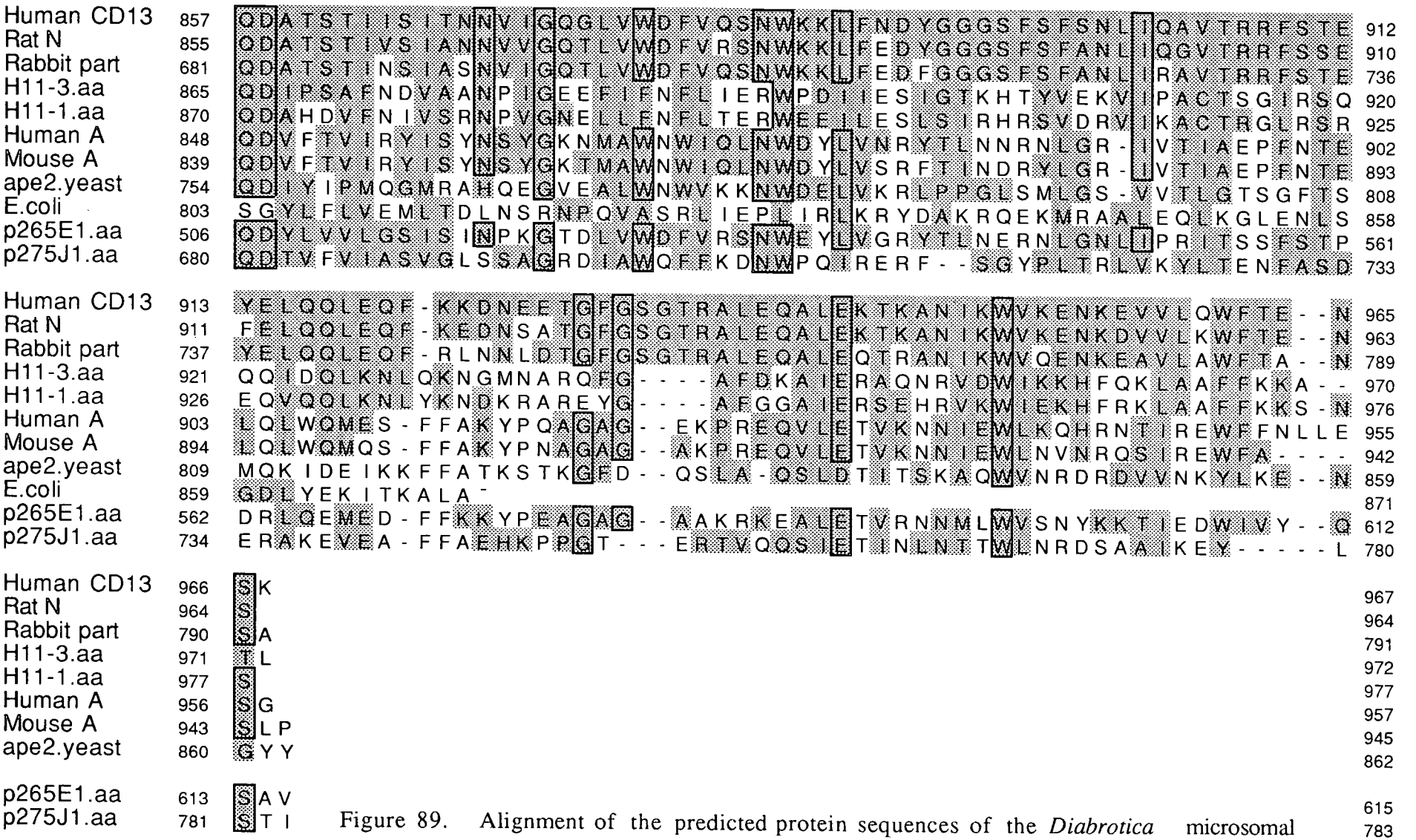


Figure 89. Alignment of the predicted protein sequences of the *Diabrotica* microsomal aminopeptidase cDNAs p265E1 and P275J1 with some aminopeptidases from mammalian, nematode (H11-1, H11-2 and H11-3), yeast and bacterial sources.

Protein	Swissprot	Identity score
mouse aminopeptidase-A	P16406	42.3% identity in 615 aa overlap
human aminopeptidase-A	Q07075	43.9% identity in 617 aa overlap
rat aminopeptidase-N	P15684	37.2% identity in 632 aa overlap
rabbit aminopeptidase-N	P15541	36.7% identity in 605 aa overlap
human aminopeptidase-N	P15144	36.5% identity in 630 aa overlap
yeast aminopeptidase II	P32454	30.3% identity in 600 aa overlap
yeast alanine/arginine aminopeptidase	P37898	36.1% identity in 363 aa overlap
Lactococcus lactis aminopeptidase-N	P37897	26.0% identity in 605 aa overlap
Lactobacillus delbrükii aminopeptidase-N	P37896	28.0% identity in 572 aa overlap
E. coli aminopeptidase-N	P04825	27.7% identity in 206 aa overlap
mouse leukotriene A-4 hydrolase	P24527	32.1% identity in 196 aa overlap
rat leukotriene A-4 hydrolase	P30349	32.1% identity in 196 aa overlap
human leukotriene A-4 hydrolase	P09960	32.1% identity in 196 aa overlap

Figure 90. The significant matches from a fasta search of the national protein databank with the predicted protein sequence of p265E1..

```
                P  K  L  D  L  V  A  I  P  D  F  V  S  G  A  M  E  H  W  G  L  V  T  F  R
p265E1 part     CCAAAATTGGATTTGGTTGCTATTCCTGATTTTGTATCAGGCGCTATGGAACACTGGGGTCTTGTCACTTTCAGA
                                                    ||  ||  |||||  |||||||  |  |||||||  |
p264A17.seq                                         GGGGCGATGGAGAACTGGGGATTGATCACTTTTACT
                                                    ||  ||  |||||  |||||||  |  ||||||  ||||
p264C2 seq                                          GGGGCGATGGAGAACTGGGGATTGATCACTTACAGA
                                                    ||  ||  |||||  |  |||||  ||  |  ||  ||||||
p264C3.seq                                          GGGGCGATGGAGAATTGGGGACTGATAACATTCAGA

                E  T  A  L  L  Y  T  N  T  T  H  S  S  L  N  K  Q  R  V  A  T  V  V  A  H
p265E1 part     GAAACCGCTCTTCTGTATACAAATACAACGCATTCAAGTCTCAATAAACAAAGAGTGGCAACTGTGGTGGCACAC
                |||   |  |||  |||  ||      |  |       ||||   |  |||  ||||   |   |  |||  ||||  ||
p264A17.seq     GAACTGGATCTACTGGATGATGGCAACAGAACAAGAAGTGATATTAATCAAAACATCATACATGTTATGGCCCAT
                ||  |  |    |  ||  |||     ||       |||||     |  |  ||  |  |||||  ||  |  ||  |  ||  |||
p264C2 seq      GAGAGCTACTTACTTTATGATCCTAAGGTATCTTCAAAAACAAGTCAAAATAGAGTTGCTAGCGTCATAGCCCAC
                ||    |  |||  |  ||         |  |     |  |  |    |  |||  ||     |||||  ||  ||||  |||
p264C3.seq      GAGGATGTTCTCTTATACGATCCATTATCATCAACTAATGAGGACAAAAGAACTACAGCAACAGTTTTGGCCCAC

                E  L  A  H  S  W  F  G  N  L  V  T  M  D  W  W  N  D  L  W  L  N  E  G  F
p265E1 part     GAATTAGCTCACAGTTGGTTTGGAAATCTAGTTACGATGGATTGGTGGAATGATTTGTGGTTAAATGAAGGATTT
                |||  |  |||||    ||||||||||    |||||  ||||||||||    ||||||||  |  ||  |||||  ||
p264A17.seq     GAAATCGCTCATCAATGGTTTGGAAACTACGTTACCTTGGATTGGTGGTCAAATTTGTGGCTCAACGAAGGCTTC
                |||||  ||  |||   |||||  ||  ||
p264C2 seq      GAATTGGCCCACCAATGGTTCGGCAA
                |||  |  ||  ||    |||||  ||  ||
p264C3.seq      GAACTGGCCCATCAGTGGTTCGGCAA

                A  S  Y  I  E  Y  K  G
p265E1 part     GCCTCTTACATTGAATACAAAGGA
                ||
p264A17.seq     GC
```

Figure 91. The sequences of the *Diabrotica* PCR products aligned with the cDNA p265E1. The identity marks are to the p265E1 sequence throughout.

linker which was undamaged. An open reading frame ran from base ~240 to base 2085 which was a TGA stop codon, some 615 amino acids. The 3' untranslated sequence ran for 77 bp to a 28 bp polyA tail 13 bp down stream of a single consensus polyadenylation signal.

This open reading frame was used to run a fasta search of the protein data base at Daresbury and selected, with very significant scores, the microsomal aminopeptidases (figures 89 and 90). The homology ran over the whole length of the p265E1 sequence to, or very nearly to, the stop codon. The highest scores and best matches, 42.3% and 43.9% identity over a 615 or 617 amino acid region, were to aminopeptidase A, also called glutamyl aminopeptidase, from mouse (Nanus *et al.* 1993) and from human (Wu *et al.* 1990) respectively. The next best matches were to aminopeptidase-Ns from some mammalian species with ~37% identity over the whole p265E1 sequence. By comparison with these homologous sequences the p265E1 sequence starts roughly at amino acid 340 and hence is 1000 to 1200 bp short of full length.

The identification of p265E1 as a cDNA for *Diabrotica* aminopeptidase A would be premature however as, although it is more homologous to the human aminopeptidase A than to the human aminopeptidase N, it is also more homologous to human aminopeptidase N than the human aminopeptidases are to each other, i.e. 34% identity over their full length. Further, the zinc binding motif, which is part of the enzyme active site (Helene, Beaumont and Roques 1991, Ashmun, Shapiro and Look 1992), is slightly different to the mammalian consensus for either type of mammalian aminopeptidase (the consensus is AHELAHQWFG with the histidines being the zinc ligands whereas p265E1 has AHELAHSWFG) (Wu *et al.* 1990, Nanus *et al.* 1993, Vallee and Auld 1990). The substitution of serine for glutamine is, however, a conservative change.

The p265E1 sequence was also compared with the MAP PCR products from *Diabrotica* and was shown (figure 91) to be homologous to all three at the DNA level but identical to none. In fact the homology at the DNA level is sufficiently poor that it was surprising that the library screening gave good strong signals. The sequence identities not including and including the oligo regions respectively, which represents the actual screening conditions, are presented in figure 94.

Screening the *Diabrotica* cDNA library with p264C2 and p264C3 the 120 bp PCR products

```
     E  I  A  T  L  K  F  P  S  L  V  T  P  G  S  Y  I  L  D  I  A  F  T  G  E  I  N  D  K  M  K  G  L  Y  R  S  K  Y  V  N
J1   GAAATTGCGACGTTAAAATTTCCCTCTTTAGTGACCCCAGGCTCCTACATTTTAGATATAGCATTTACAGGTGAAATTAATGACAAAATGAAAGGGCTGTACAGAAGTAAATATGTAAAC  120

     Q  K  G  E  E  T  Y  A  A  V  T  Q  F  E  S  T  D  A  R  R  C  F  P  C  W  D  E  P  A  H  K  A  T  F  D  I  S  L  T  V
J1   CAAAAAGGAGAGGAAACATACGCAGCGGTAACACAATTCGAATCGACAGATGCTCGTCGTTGTTTCCCATGTTGGGACGAGCCTGCACATAAAGCCACCTTTGACATAAGCCTTACTGTA  240

     P  K  N  L  V  A  L  S  N  M  P  V  K  N  S  Q  P  E  G  D  L  I  K  Y  D  F  E  K  T  P  I  M  S  T  Y  L  V  A  A  V
J1   CCTAAGAACTTAGTTGCCTTATCAAATATGCCAGTCAAAAACAGTCAACCTGAAGGTGATCTCATTAAATACGACTTCGAAAAAACCCCCATTATGTCAACGTACTTAGTAGCAGCCGTT  360

     I  G  E  Y  D  Y  V  E  D  K  S  S  D  G  V  L  V  R  V  Y  T  P  K  G  K  K  E  Q  G  L  F  A  L  E  V  A  T  K  V  L
J1   ATTGGAGAATATGATTATGTTGAAGACAAGTCTAGTGACGGTGTACTCGTTAGGGTGTATACTCCAAAAGGAAAAAAAGAACAAGGACTGTTCGCTCTTGAAGTAGCCACCAAAGTATTG  480

     P  Y  Y  K  E  Y  F  N  I  A  Y  P  L  P  K  I  D  L  I  A  I  A  D  F  S  A  G  A  M  E  N  W  G  L  V  T  Y  R  E  T
J1   CCTTATTACAAAGAGTATTTTAATATAGCTTATCCACTACCTAAGATTGATTTGATCGCTATTGCGGACTTCAGTGCTGGTGCTATGGAAAATTGGGGCCTTGTGACTTACAGAGAAACC  600

     C  L  L  V  D  P  K  N  T  S  A  V  S  K  Q  W  I  A  L  V  V  G  H  E  L  A  H  Q  W  F  G  N  L  V  T  M  E  W  W  T
J1   TGCCTTCTCGTAGATCCTAAAAACACTTCGGCTGTTAGCAAACAGTGGATTGCGCTAGTAGTGGGTCACGAACTTGCACATCAGTGGTTCGGGAATCTCGTGACTATGGAATGGTGGACC  720

     H  L  W  L  N  E  G  Y  A  S  F  V  E  F  L  C  V  N  Y  L  F  P  E  Y  D  I  W  T  Q  F  V  N  D  S  Y  I  K  A  L  E
J1   CATCTTTGGTTAAATGAGGGCTACGCTTCATTTGTAGAATTTCTCTGCGTCAACTACCTCTTCCCTGAGTATGATATATGGACTCAATTTGTTAATGATTCGTACATCAAAGCTCTAGAA  840

     L  D  S  L  K  N  S  H  P  I  E  V  P  V  G  N  P  S  E  I  D  E  I  F  D  D  I  S  Y  N  K  G  A  S  V  I  R  M  L  H
J1   TTAGACAGTTTGAAAAATTCTCATCCCATCGAAGTACCTGTTGGAAACCCCTCCGAAATTGACGAAATATTCGATGATATTAGCTACAACAAAGGAGCTTCTGTTATACGAATGTTGCAT  960

     H  Y  I  G  D  E  D  F  R  K  G  M  H  L  Y  L  T  R  H  Q  Y  K  N  T  F  T  E  D  L  W  A  A  L  E  E  A  S  Q  K  P
J1   CATTATATCGGAGATGAGGATTTCAGGAAAGGCATGCATTTGTACTTAACCAGACATCAGTACAAAAATACTTTTACAGAAGACTTGTGGGCTGCTTTAGAAGAAGCCAGTCAAAAGCCA 1080

     V  G  D  V  M  S  F  W  T  K  Q  K  G  F  P  V  V  K  V  T  Q  Q  P  A  G  D  G  G  G  V  K  L  A  L  S  Q  S  K  F  T
J1   GTTGGGGACGTTATGTCTTTTTGGACCAAACAAAAAGGATTTCCCGTTGTTAAAGTTACACAGCAACCGGCAGGTGACGGTGGTGGTGTTAAACTAGCGCTCTCGCAATCTAAATTCACC 1200

     A  D  G  S  S  P  A  D  E  D  Y  L  W  M  I  P  I  S  I  S  T  S  K  Q  P  S  K  E  V  V  C  I  V  L  T  T  R  T  A  E
J1   GCAGACGGATCAAGTCCAGCGGACGAAGATTATCTGTGGATGATACCAATTTCTATATCTACCTCTAAACAACCTTCCAAAGAAGTAGTTTGCATTGTTCTCACTACGAGAACGGCAGAA 1320

     I  V  V  P  D  V  G  K  S  D  W  I  K  I  N  P  G  T  I  G  F  Y  R  T  Q  Y  T  P  E  M  L  E  K  F  I  P  A  I  R  D
J1   ATCGTTGTACCTGATGTTGGAAAATCTGATTGGATTAAGATTAATCCAGGAACTATTGGCTTCTATCGAACCCAATATACCCCAGAAATGTTGGAGAAGTTTATACCTGCTATTAGGGAT 1440

     R  S  L  P  P  L  D  R  L  G  L  L  D  D  L  F  A  M  V  K  A  G  H  T  S  T  V  E  V  L  K  L  L  N  A  Y  E  D  E  P
J1   AGATCTCTACCACCATTGGATAGGTTAGGTTTATTAGACGACTTATTTGCTATGGTTAAAGCCGGACACACAAGTACTGTTGAAGTTCTAAAACTTCTTAACGCTTATGAAGACGAACCT 1560
```

```
     D  Y  N  V  W  C  S  I  V  N  I  L  G  R  L  G  Q  L  L  G  H  T  D  Y  E  A  D  Y  N  K  Y  Q  K  R  L  L  S  K  V  Y
J1   GATTACAACGTTTGGTGTTCAATTGTAAATATTCTCGGACGCTTAGGACAACTTTTGGGTCATACAGATTATGAAGCAGACTACAACAAGTACCAAAAGAGACTCTTGAGCAAAGTGTAC 1680

     S  R  L  G  W  T  Q  K  S  G  E  T  H  L  D  T  L  L  R  G  L  V  L  G  R  L  A  W  L  D  H  E  H  T  I  S  E  A  R  N
J1   AGCAGATTGGGATGGACACAAAAATCAGGCGAAACACATCTCGATACCCTCTTAAGAGGATTAGTGTTAGGTCGATTGGCGTGGTTAGATCACGAACATACCATATCTGAGGCCAGGAAC 1800

     R  F  D  T  H  V  K  S  E  Q  V  L  P  A  D  L  R  S  A  C  Y  K  T  I  L  R  A  G  G  Q  A  E  Y  D  T  L  L  Q  L  Y
J1   AGATTTGATACCCACGTCAAATCTGAACAAGTACTACCTGCTGATTTGAGAAGCGCTTGTTATAAAACAATTTTGAGAGCTGGTGGTCAAGCGGAATACGACACTTTATTGCAGTTGTAT 1920

     R  S  S  D  L  H  E  E  K  D  R  I  S  R  A  F  G  A  T  K  D  P  N  L  L  K  K  V  L  E  F  A  M  S  D  E  V  R  S  Q
J1   AGGTCTAGCGATCTCCACGAAGAGAAGGATAGAATTAGCAGAGCGTTTGGAGCTACCAAAGATCCAAACTTGTTAAAGAAGGTTCTGGAATTTGCTATGTCGGATGAAGTAAGATCCCAA 2040

     D  T  V  F  V  I  A  S  V  G  L  S  S  A  G  R  D  I  A  W  Q  F  F  K  D  N  W  P  Q  I  R  E  R  F  S  G  Y  L  L  T
J1   GACACAGTATTTGTTATAGCTTCGGTGGGATTGAGCAGTGCTGGAAGAGATATCGCTTGGCAATTCTTTAAAGACAACTGGCCCCAGATCAGGGAACGATTCAGCGGTTACCTTCTTACA 2160

     R  L  V  K  Y  L  T  E  N  F  A  S  D  E  R  A  K  E  V  E  A  F  F  A  E  H  K  P  P  G  T  E  R  T  V  Q  Q  S  I  E
J1   AGATTAGTGAAGTATCTGACAGAGAATTTCGCCAGTGACGAAAGAGCCAAAGAAGTTGAAGCTTTCTTCGCCGAACACAAACCCCCAGGTACAGAAAGAACTGTACAGCAGTCTATCGAG 2280

     T  I  N  L  N  T  T  W  L  N  R  D  S  A  A  I  K  E  Y  L  S  T  I  *
J1   ACAATTAATTTGAACACAACTTGGCTAAATAGAGACTCGGCTGCCATTAAAGAGTACCTCAGTACTATATAATTTACCTTGCAACTTTAAAATATTTTTGTTTTTGTTTGTTAAATATGG 2400

                                        polyA
J1   AACATATGCCAGTGACTGTACCTATGATTTGTTATTAAATAAAAGTTTTTGTAATTAGCAGAGAAATGTAGTCATATTTGATTAATCATATGGATACTGAACTTATTCCTCAGGTAAACA 2520

                     polyA     polyA                                          polyA tail
J1   AATTATTATATCATTTAAATAATGTAAAATAAACGGAAATAAAGGACTATCTGAGAAACAATTAAAAAAAAAAAAAAA                                          2598
```

Figure 92. The DNA sequence of cDNA p275J1 and its translation in the long ORF which encodes a microsomal aminopeptidase.

Protein	Swissprot	Identity score
human aminopeptiodase-A	Q07075	34.4% identity in 776 aa overlap
mouse aminopeptiodase-A	P16406	34.4% identity in 777 aa overlap
yeast alanine/arginine aminopeptidase	P37898	44.9% identity in 448 aa overlap
yeast aminopeptidase II	P32454	36.8% identity in 778 aa overlap
Lactococcus lactis aminopeptidase-N	P37897	32.0% identity in 781 aa overlap
Lactobacillus delbrükii aminopeptidase-N	P37896	30.0% identity in 776 aa overlap
rat aminopeptidase-N	P15684	31.6% identity in 770 aa overlap
human aminopeptidase-N	P15144	31.1% identity in 776 aa overlap
rabbit aminopeptidase-N	P15541	30.9% identity in 729 aa overlap
pig aminopeptidase-N part	P15145	43.9% identity in 139 aa overlap
E. coli aminopeptidase-N	P04825	27.3% identity in 319 aa overlap
human leukotriene A-4 hydrolase	P09960	24.9% identity in 353 aa overlap
mouse leukotriene A-4 hydrolase	P24527	24.9% identity in 353 aa overlap
rat leukotriene A-4 hydrolase	P30349	24.3% identity in 366 aa overlap

Figure 93. The significant matches from a search of the national protein database with the predicted protein sequence of p275J1.

The *Diabrotica* cDNA library was rescreened with a mixed probe of the two shorter PCR products, p264C2 and p264C3, which were labelled by the PCR protocol described in the methods. Experimental conditions were as above with the temperature for all steps dropped to 60°C because of the short length of the probes (~120 bp); final washing was to 1 x SSC, 0.1% SDS. Three filters carrying ~1.5 x 10^5 plaques each were screened and a single positive was detected and plaque purified. As only approximately half the library was screened it is possible that other clone(s) could be found.

After plasmid rescue was performed the clone, called p275J1, was restriction mapped and DNA sequenced in from each end and, as it was not identical with the previously isolated clone p265E1, it was then fully DNA sequenced in both directions (figure 92). The insert was 2598 bp long with an undamaged 5' cloning primer and a damaged 3' cloning primer. An open reading frame ran in from the 5' end to base 2350 which was a TAA stop codon, some 783 amino acids, and the 3' untranslated sequence ran for 231 bases to a 15 bp polyA tail. There were three consensus polyadenylation signals in the 3' untranslated sequence at 2438, 2548 and 2558 bases, the last one, 20 bp upstream of the polyA tail, being the functional one.

The open reading frame was used to run a fasta search of the protein sequence data bank at Daresbury and selected with very significant scores the microsomal aminopeptidases (figures 93 and 89). The region of homology ran over the whole length of the sequence with the best matches to an alanine/argenine aminopeptidase from the yeast *Saccharomyces cerevisiae* (Caprioglio, Padilla and Werner-Washburne 1993), 44.9% in a 448 amino acid overlap, aminopeptidase II (a leucine aminopeptidase) from the yeast *Saccharomyces cerevisiae* (García-Alvarez, Cueva and Suárez-Rendueles 1991), 36.8% in a 778 amino acid overlap, the human and mouse aminopeptidase As (Nanus *et al.* 1993, Wu *et al.* 1990), both 34.4% identity, an aminopeptidase N from *Lactococcus lactis* (Tan *et al.* 1992, Stroman 1992), 32% identity, an aminopeptidase N from *Lactobacillus delbrückii* (Klein *et al.* 1993), 30% identity and the mammalian aminopeptidase Ns at about 31% identity over the whole sequence. Again there was a slight sequence difference to the consensus in the zinc binding motif (Wu *et al.* 1990, Nanus *et al.* 1993, Vallee and Auld 1990) (the consensus is AHELAHQWFG with the histidines being the zinc ligands whereas p275J1 has GHELAHSWFG), which is part of the enzyme active site (Helene, Beaumont and Roques 1991, Ashmun, Shapiro and

Table 1

PCR product clone primers	Homology not including primers	Homology including
p264A17	58.8%	62.7%
p264C2	55.7%	60.7%
p264C3	50.5%	55.6%

Homology of the *Diabrotica* MAP PCR products to the *Diabrotica* cDNA clone p265E1.

PCR product clone primers	Homology not including primers	Homology including
p264A17	54.1%	59.0%
p264C2	54.6%	62.0%
p264C3	47.4%	59.1%

Homology of the *Diabrotica* MAP PCR products to the *Diabrotica* cDNA clone p265J1.

Figure 94.

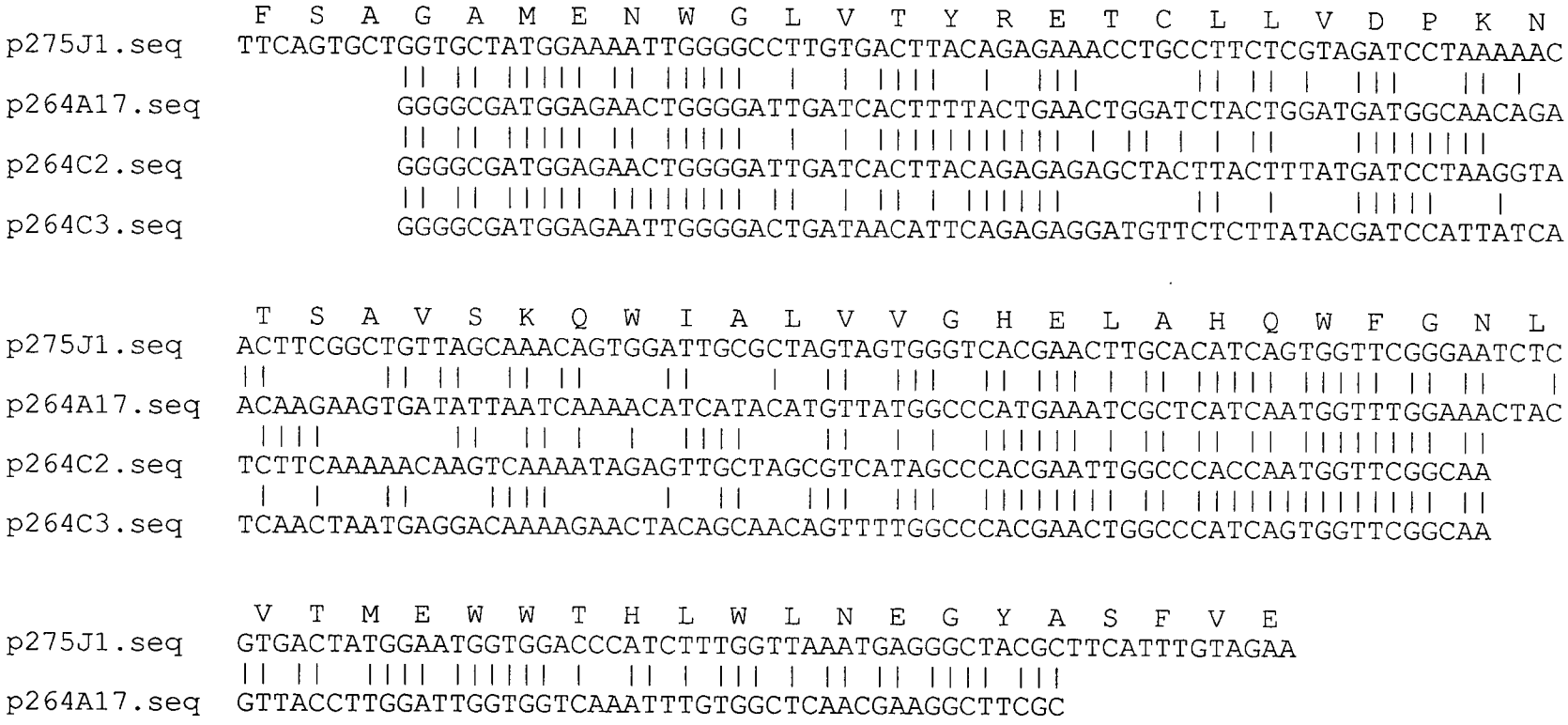


Figure 95. The sequences of the *Diabrotica* PCR products aligned with the cDNA p275J1. The identity marks are to the p275J1 sequence throughout.

Look 1992). The substitution of glycine for alanine and serine for glutamine are conservative and should not cause a big difference. The glutamine to serine was also observed in p265E1.

Assuming p275J1 is a cDNA of a microsomal aminopeptidase and has the cytosolic and transmembrane domains, unlike the yeast and bacterial genes which encode soluble enzymes which lack these N-terminal domains, by comparison with its mammalian homologs it starts at amino acid 160 to 170 and hence is ~500 bp short of full length. This is significantly nearer to full length than p265E1.

The protein sequences of p265E1 and p275J1 were truncated to the same length and when aligned were 32.4% identical at the protein level, which is slightly less identical than to the mammalian aminopeptidase As and no more homologous to each other than to genes from entirely unrelated organisms. Aligned and optimised they were 51.7% identical at the DNA level in the region of overlap. The p275J1 sequence was also compared to the *Diabrotica* MAP PCR products and was shown (figure 95) to be homologous to all three at the DNA level but identical to none. In fact, the homology at the DNA level is again sufficiently poor that it was surprising that the library screening gave good strong signals. The sequence identities not including and including the oligo regions are presented in figure 94.

These homology figures imply a cut off for library screening under these conditions between ~62% and ~59% as p275J1 was not detected in the first screen with p264A17 and p265E1 was not reisolated in the second screen with p264C2 and p264C3.

p275J1 is a cDNA of a *Diabrotica* microsomal aminopeptidase like gene but like p265E1 lacks enough of the 5' sequence to show whether it is a true microsomal aminopeptidase and is not obviously more homologous to aminopeptidases of any particular type.

Screening the *Diabrotica* cDNA library with p265E1 and p275J1 the two cDNAs

As part of the *E. coli* expression experiments (See below) a subclone of p265E1 was made called p273C2 which had had the artifact cDNA removed. The insert of this, an ~1880 bp fragment, and the ~2600 bp insert of p275J1 cut out with EcoR I and Asp718 I, were used as a mixed probe to rescreen the *Diabrotica* cDNA library to isolate a fuller length clone or a fragment which covered the 5' end of one of the cDNAs. Hybridisation

was done at 60°C and washing was to 3 x SSC, 0.1% SDS which is about 77% stringency, roughly the same as the previous screens after accounting for the shorter probes used. These relaxed screening conditions were chosen to maximise the number of homologous clones detectable without having too high a background.

Due to the long length of the probes and the need to detect plaques of low homology the hybridisations were carried out over approximately 60 hours, just under 1 x C_0t_{half} (Sambrook, Fritsch and Maniatis 1989).

Three filters carrying 1x 10^5 plaques each were probed and a total of 41 positives were detected; 13 strong and 28 weak. At this time the collapse of the Axis Genetics Ltd. insect resistance programme occurred and it was clear to me that in the limited time left characterisation of these putative positives by any method involving plaque purification would not produce results quickly enough. As a more full length cDNA was required, I devised a novel PCR based protocol to screen the primary phage plugs for their insert sizes.

Using the DNA sequencing forwards and reverse primers and the primary phage plugs as templates in standard PCR reactions with appropriate conditions, which for the sequencing primers is an annealing temperature of 50°C, and an extension time of four minutes to allow the amplification of fragments longer than the expected size of a full length MAP cDNA (3 - 3.5 kb), the inserts of the populations of phage in each primary plug would be amplified. These PCR reactions were carried out and the results were run out on gel, Southern blotted and probed with the p273C2 and p275J1 mixed probe under the same conditions as the library was screened under.

The expectation was that the insert amplified from the hybridising phage line in each primary plaque population would amplify enough to be detected as a hybridising band on the blot and that the amplified inserts from the other phage lines would not hybridise significantly. However all that was seen was variable amounts of hybridisation to low molecular weight, less than ~500 bp, smears and no definite bands.

Test PCR reactions were performed using as template the pure phage line from which p275J1 was rescued, λ275J1,1, both with and without Enhance Perfect Match (Stratagene), an additive which decreases the amplification of artifacts by destabilising mismatched DNA duplexes, and Taq Extender (Stratagene), an additive and buffer system which improves the amplification of long PCR products. The results from this were that the expected band of ~2.6 kb was by far the strongest product in the presence of

the Taq Extender and in its absence no bands were visible on gel. The Enhance Perfect Match made no apparent difference. Thus the amplification of fragments of the size range expected for a full length MAP cDNA is possible from a template of λ phage, at least under the favourable condition that the phage line is pure.

A second screening experiment was set up using both Taq Extender and Enhance Perfect Match and as a positive control a sample of λ275J1, the primary plug from which the pure λ275J1,1 line had been purified. The result was again only low molecular weight smears of hybridisation with no bands detected even for the λ275J1 positive control. Thus the amplification of short PCR fragments from the same primers as long fragments in the same reaction is able to completely suppress the amplification of the long fragments.

Due to lack of time this line of experiments was discontinued and the primary phage plugs were stored at 4°C in the usual manner where they remain for future exploitation.

Obtaining the 5' end of the cDNA p265E1 by RACE PCR

RACE, rapid amplification of cDNA ends (Frohman 1990) , is a technique whereby using a single oligo specific to a given protein, a fragment of that protein's cDNA can be generated by PCR using, as the other primer, a general oligo. For 5' RACE the procedure is to make first strand cDNA using as primer an oligo specific for the cDNA whose 5' end is wanted, and to tag the end of this cDNA in some way. Early protocols used homopolymer tailing with terminal deoxynucleotide transferase to tag the 5' end but this was improved upon by the use of specific anchor primers ligated to the 5' end of the molecule. The 5' end of the cDNA is then amplified by a PCR reaction using as template the tagged first strand cDNA, as 5' primer either an oligo of the complementary nucleotide base to the homopolymer tail or the complement of the anchor primer, and as 3' primer either the cDNA specific primer used to prime first strand cDNA synthesis or a primer designed to fit the cDNA just 5' of it which improves the specificity of the PCR reaction.

The 5' RACE was done with the 5'-Amplifinder RACE Kit (Clontech) which supplies the buffers, enzymes and anchor primers. A pair of primers were designed to the p265E1 sequence for first strand cDNA synthesis, oligo #1041, and for the PCR reaction, oligo #1040 (figure 96). For this experiment a fresh preparation of *Diabrotica* polyA+ RNA was made as for the cDNA library (see above). Because the expected distance from the primers to the 5' end of the p265E1 mRNA was 1000 to 1200 bp, which is at the

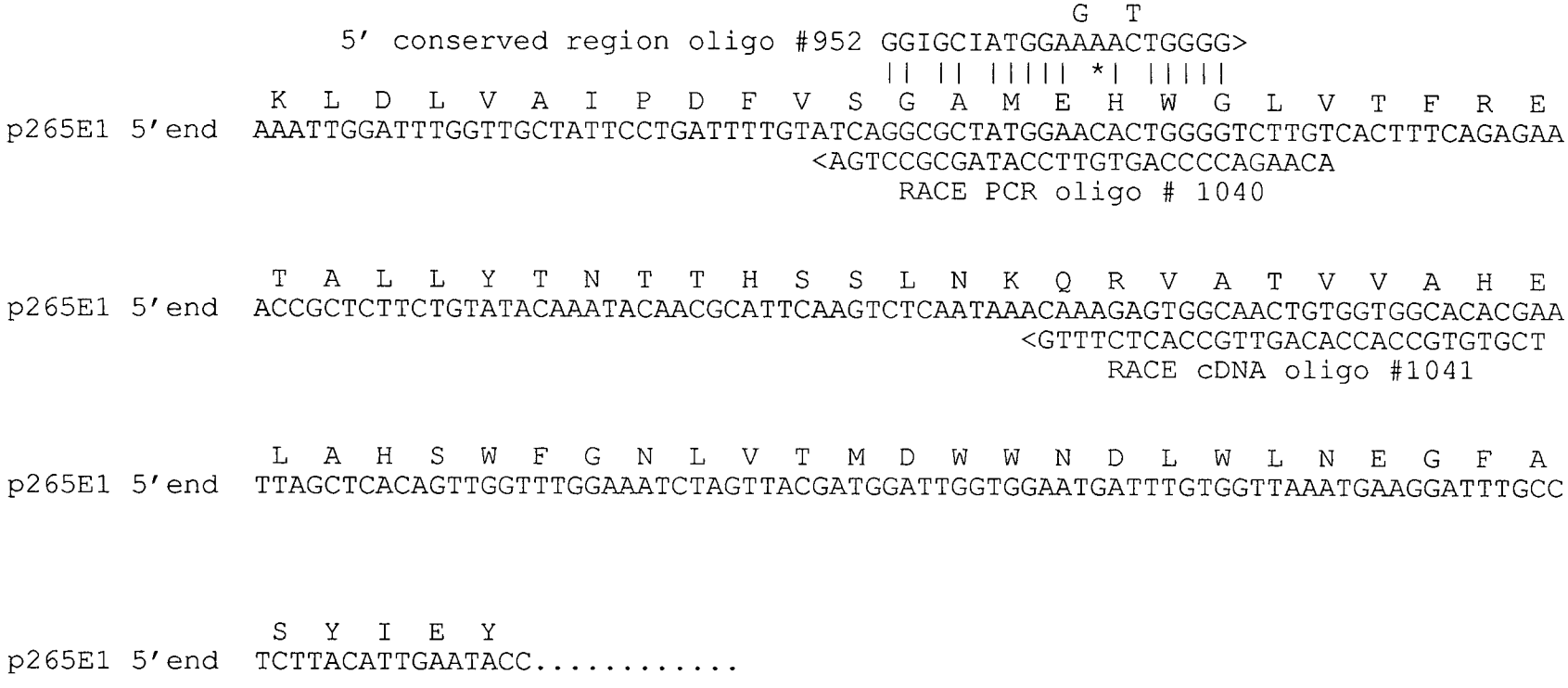


Figure 96. The RACE oligos, #1040 and #1041, their design and relation ship to each other, p265E1 and the 5'conserved region oligo #952.

upper limits of the kit's abilities, the cDNA synthesis conditions were modified to 48°C for 1 hour from 52°C for 30 minutes as recommended for longer fragments. After the first strand synthesis the RNA was hydrolysed and the DNA purified as instructed before the anchor primer was ligated to an aliquot of it.

A hot start PCR protocol was used in which all components were assembled and heated to 82°C before the primers were added and thermocycling begun. The conditions were as recommended with an annealing temperature of 60°C (Clontech 5'-Amplifinder RACE Kit manual) except that the extension time was increased to 2 minutes because of the longer product expected. The PCR was performed both with and without Enhance Perfect Match (Stratagene). After 35 cycles a sample of each reaction was run on gel but nothing was visible so a further unit of Taq polymerase was added to each reaction and a further 10 cycles of amplification performed. Samples of the reactions were again run on gel but only low molecular weight products were seen.

A positive control was devised; PCR amplification between the cDNA synthesis primer #1041 and the 5' conserved region primer from the original MAP PCR experiments oligo #952 (See above and figure 96) would be expected to produce a fragment of 113 bp, this experiment is a control for cDNA synthesis and will work on non-anchored cDNA.. An annealing temperature of 50 °C, to match the 5' conserved region primer, an extension time of 20 seconds, and Enhance Perfect Match (Stratagene) were used. This PCR experiment was performed using unpurified cDNA, purified cDNA and anchor ligated cDNA as templates and in each case gave, as by far the strongest band, the expected ~110 bp band. Thus cDNA synthesis had occurred and the problem was either the anchor primer ligation , the PCR conditions, or that the fragment to be amplified is beyond the abilities of the kit to produce.

A second aliquot of the purified cDNA was ligated to the anchor primer as a fresh template and PCR reactions were set up as before except that the annealing temperature was reduced to 55°C and the extension time increased to 3 minutes. Also as part of a nested PCR strategy two PCR reactions were set up, plus and minus Enhance PM, using the cDNA synthesis primer oligo #1041. After 35 cycles the cDNA synthesis primer reactions showed smears of DNA running up from low molecular weight but no definite bands on gel and were kept as templates for the nested PCR experiment. The nested PCR experiment was done under the same conditions with and without Enhance PM and substituting the cDNA PCR primer oligo #1040 for the cDNA synthesis primer oligo #

```
p268A1.seq   AACAACACCAACGACTTCATCAACTTCCACATCCTCTAAAGAGACCTCTACAGAAGAATCAACTACCAGCAATGA
               T  T  P  T  T  S  S  T  S  T  S  S  K  E  T  S  T  E  E  S  T  T  S  N  E
                Q  H  Q  R  L  H  Q  L  P  H  P  L  K  R  P  L  Q  K  N  Q  L  P  A  M

p268A1.seq   GCCTACTACCACCTCGACTTCTTCCACCGCTGACCCTACCACCACCGATCCTACCACTTCAGCATCGCCCTCTTC
               P  T  T  T  S  T  S  S  T  A  D  P  T  T  T  D  P  T  T  S  A  S  P  S  S
             S  L  L  P  P  R  L  L  P  P  L  T  L  P  P  P  I  L  P  L  Q  H  R  P  L

p268A1.seq   CACTCCCGAACCCACCGTCCCTTCAACCACTACCAACACTGAGCCCACCTCCACATCAAGTTCCGATGGCGATAG
               T  P  E  P  T  V  P  S  T  T  T  N  T  E  P  T  S  T  S  S  S  D  G  D  S
             P  L  P  N  P  P  S  L  Q  P  L  P  T  L  S  P  P  P  H  Q  V  P  M  A  I

p268A1.seq   CACCACTTCGTCTAGGAGCACCACTACTAGTCAGACCCCCGTTGTGAAGACCATCACTTCCTAC
               T  T  S  S  R  S  T  T  T  S  Q  T  P  V  V  K  T  I  T  S  Y
             A  P  L  R  L  G  A  P  L  L  V  R  P  P  L  *  R  P  S  L  P
```

Figure 97. The DNA sequence in from one end of p268A1 minus the PCR primer and its translation in the long ORFs. The upper is the threonine/serine rich sequence the lower the proline rich sequence.

Protein	Swissprot	Identity score
human mucin 2	Q02817	35.9% identity in 103 aa overlap
Xenopus integumentary mucin C1	Q05049	35.2% identity in 105 aa overlap
hsveb glycoprotein X precursor	P28968	36.4% identity in 99 aa overlap
Xenopus integumentary mucin A1precursor	P10667	30.4% identity in 102 aa overlap
Drosophila salivary glue protein	P13729	36.0% identity in 100 aa overlap
entfc p45 protein precursor	P13692	30.2% identity in 96 aa overlap
yeast A-agglutinin attachment subunit precursor	P32323	34.0% identity in 100 aa overlap
Drosophila salivary glue protein SGS-3	P13728	31.1% identity in 103 aa overlap
sacdi glucoamylase S1 precursor	P04065	28.0% identity in 100 aa overlap
dicdi spore germination protein	P22698	28.0% identity in 100 aa overlap
yeast glycoprotein gp100 precursor	P08640	29.0% identity in 100 aa overlap
dicdi glycoprotein gp100 precursor	Q06885	29.4% identity in 102 aa overlap
Drosophilia serine protease stubble	Q05319	29.2% identity in 106 aa overlap

Figure 98. A selection of the matches from a fasta search of the national protein databank with the threonine/serine rich ORF from p268A1.

```
p268A4.seq  CATTGGCCACCGAGGTCGACGTCTACTTCTGTGATCCGCGCTGTCCATGGCAGCGCGGGTCAAACGAAAACACCA
             H  W  P  P  R  S  T  S  T  S  V  I  R  A  V  H  G  S  A  G  Q  T  K  T  P
              I  G  H  R  G  R  R  L  L  L  *  S  A  L  S  M  A  A  R  V  K  R  K  H  Q
               L  A  T  E  V  D  V  Y  F  C  D  P  R  C  P  W  Q  R  G  S  N  E  N  T

p268A4.seq  ACAGATTGCTGCGCCAGTATCTTCCGCGTGGCACCGATCTGTCCCTGCATAGCCAGGCCAAGCTCAGCGCGATCG
             T  D  C  C  A  S  I  F  R  V  A  P  I  C  P  C  I  A  R  P  S  S  A  R  S
              Q  I  A  A  P  V  S  S  A  W  H  R  S  V  P  A  *  P  G  Q  A  Q  R  D  R
            N  R  L  L  R  Q  Y  L  P  R  G  T  D  L  S  L  H  S  Q  A  K  L  S  A  I

p268A4.seq  CAAGGCAGCTCAACGAAAGGCCGCGAAAGACCTTGCTCTATCAGACCCCAGCTGAGAAGTTCGCAGATTGTGTTG
             Q  G  S  S  T  K  G  R  E  R  P  C  S  I  R  P  Q  L  R  S  S  Q  I  V  L
              K  A  A  Q  R  K  A  A  K  D  L  A  L  S  D  P  S  *  E  V  R  R  L  C  C
            A  R  Q  L  N  E  R  P  R  K  T  L  L  Y  Q  T  P  A  E  K  F  A  D  C  V

p268A4.seq  CGGCGATCAATTGAACCCGCCGCCACAAGCGGACAATCGTCGAAAAACTAAAGGCGAGCCGCATTTTGATTCT
             R  R  S  I  E  P  A  A  T  S  G  Q  S  S  K  N  *  R  R  A  A  F  *  F
              G  D  Q  L  N  P  P  P  Q  A  D  N  R  R  K  T  K  G  E  P  H  F  D  S
            A  A  I  N  *  T  R  R  H  K  R  T  I  V  E  K  L  K  A  S  R  I  L  I
```

Figure 99. The DNA sequence in from one end of p268A4 minus the PCR primer and its translation in one direction. The underlined protein sequence has significant homology to some transposases.

1041. When analysed on gel however no bands were visible.

After 35 cycles the cDNA PCR primer reactions showed no visible bands and were further amplified with the addition of 1 extra unit of Taq polymerase for an extra 10 cycles. When analysed on gel the reaction without Enhance PM showed two high molecular weight bands at 790 and 1070 bp which is in the size range expected and so these were isolated. A smear was also visible in the 400 to 500 bp region and at low molecular weight. The reaction with Enhance PM showed a low molecular weight smear and a single band at 400 bp but as this was well below the size range expected no further investigation was performed. The isolated bands were reamplified and subcloned as p268A for the 790 bp band and p268B for the 1070 bp.

Two of each of these clones were DNA sequenced in from both ends but all were shown to be artifacts with the cDNA PCR primer oligo #1040 at both ends and the p265E1 coding sequence not persisting beyond the oligo. There were however some open reading frames in the clones and these were used to run fasta searches of the protein data base at Daresbury.

p268A1 and p268A2 were identical and had a number of ORFs of sufficient size to run fasta searches but only two ORFs gave any significant matches (figures 97 and 98). The threonine/serine rich sequence selected with good scores mucins both from humans and from *Xenopus* , a glycoprotein precursor from equine herpes virus, salivary glue proteins from *Drosophila* species and a serine protease involved in epithelial morphogenesis also from *Drosophila* called serine proteinase stubble or stubble-stubbloid protein. The homology in all cases was in the mid thirty percents and was due to the open reading frame being very rich in threonine and serine as were the selected proteins. To resolve whether these matches were of any significance would require sequence data from beyond the threonine and serine rich region which as the sequences in from either end did not meet is not available. The proline rich sequence selected a number of proteins with proline repeats and these were discounted as the other parts of the sequence did not match.

p268A4 was also DNA sequenced as it had an internal Pvu II site which none of the other A clones had and was found to have a different sequence (figure 99). It had four possible open reading frames three of which selected nothing significantly except a marginally significant match to a acetoin catabolism regulatory protein from a bacterium but the position of the region of homology, with the start of the regulatory protein being matched to the middle of the p268A4 open reading frame, makes this match unlikely to be

Protein	Swissprot	Identity score
Alcaligenes eutrophus ISE IS1086	P37248	63.5% identity in 74 aa overlap
E. coli transposase of ISE IS30	P37246	56.0% identity in 75 aa overlap
Bacteroides fragilis transposase of ISE IS4351	P37247	49.3% identity in 71 aa overlap
Strepotcoccus salivarius transposase of ISE IS1161	P37245	39.4% identity in 66 aa overlap
spiroplasma virus SPV1-R8A2 B gene 3	P15894	33.8% identity in 68 aa overlap

```
ID   TRA8_a    Alcaligenes eutrophus  insertion sequence element
IS1086
           63.5% identity in 74 aa overlap

                                                 10        20        30
p268A4                                    LATEVDVYFCDPRCPWQRGSNENTNRLLRQY
                                           ||:| ||||||::||||||: |||| |||||
tra8_a GKLQSLVAPLRQTLTYDQGREMARHAELSAATGVRVYFCDPHSPWQRGTCENTNGLLRQY
        230       240       250       260       270       280

               40        50        60        70
p268A4 LPRGTDLSLHSQAKLSAIARQLNERPRKTLLYQTPAEKFADCVAAIN
       ||:||||||::||::|:|||  ||:|||||| :::| : :|: :|:
tra8_a LPKGTDLSVYSQEELDAIADSLNGRPRKTLNWHSPLQVLAQVLANPTDRLPVQ
        290       300       310       320       330


ID   TRA8_e   E. coli transposase of insertion sequence element
IS30
           56.0% identity in 75 aa overlap

                                                  10        20        30
p268A4                                    LATEVDVYFCDPRCPWQRGSNENTNRLLRQ
                                          ::|:|:||||||::||||||:|||||  |:||
tra8_e TDKFLSLPSELRKSLTWDRGMELARHLEFTVSTGVKVYFCDPQSPWQRGTNENTNGLIRQ
        280       290       300       310       320       330

               40        50        60        70
p268A4 YLPRGTDLSLHSQAKLSAIARQLNERPRKTLLYQTPAEKFADCVAAIN
       |:|: | |: ::| :|: :| |||:|||||| ::|| | ::  ||
tra8_e YFPKKTCLAQYTQHELDLVAAQLNNRPRKTLKFKTPKEIIERGVALTD
        340       350       360       370       380
```

Figure 100. The results of a fasta search of the national protein database with a long ORF from p268A4 and the alignments of the best two matches.

real.

The fourth open reading frame selected with good scores the transposases for insertion elements from a number of bacterial species (figure 100). The region of homology ran over the whole of the open reading frame to its stop codon which was very close to the end of the transposase proteins. The best homology, 63.5% identity in a 74 amino acid overlap, was to the transposase for insertion sequence element IS1086 from *Alcaligenes eutrophus* . The other matches were to the *E. coli* transposase of insertion sequence element IS30, 56.0% identity in a 75 amino acid overlap, to the *Bacteroides fragilis* transposase of insertion sequence element IS4351, 49.3% identity in a 71 amino acid overlap, to the *Strepotcoccus salivarius* transposase of insertion sequence element IS1161, 39.4% identity in a 66 amino acid overlap, and to gene 3 of spiroplasma virus SPV1-R8A2 B, 33.8% identity in a 68 amino acid overlap. These results strongly suggest that this PCR product originates from a gut bacterium in *Diabrotica* and not *Diabrotica* itself and that that bacterium was more closely related to *Alcaligenes eutrophus* than to *E. coli* .

It is interesting that a bacterial sequence should show up from the RACE PCR, which uses as template cDNA synthesised from a supposedly specific primer, and raises the possibility of artifacts of this type in the cDNA library which was made from similarly prepared polyA+ RNA. The efficiency of incorporation of RNA sequences lacking polyA tails into cDNA libraries whose synthesis is primed with polyT is low (Sambrook, Fritsch and Maniatis 1988) and bacteria do not polyadenylate their mRNA. However a bacterial sequence with an adenosine rich run of bases at some point would copurify with polyA+ RNA and probably give rise to partial cDNAs primed from the A rich region with reasonable efficiency.

p268B1 and p268B3 were identical except for a couple of single base differences which are consistent, A/G and C/T, with the normally observed PCR errors. There were only two open reading frames and only one marginally significant match was found which, from its position and as none of the cysteins matched, was considered to be non significant.

No further work was done with any of these clones.

It was thought that the failure of the 5' RACE could be due to the size of the PCR product being too long and so a set of PCR experiments were tried using the Taq Extender buffer and additive (Stratagene). The template was the second lot of anchor cDNA

ligation product and both the cDNA PCR primer oligo #1040 and the cDNA synthesis primer oligo #1041 were used as primers with the anchor PCR primer and with and without Enhance PM (Stratagene). After 35 cycles nothing but low molecular weight smears were visible on gel from any of the four reactions. A further 10 cycles were performed but without success and the experiment was abandoned. No further work was done to obtain a full length cDNA by this method due initially to a collapse in the *Diabrotica* population at Durham and hence the inability to produce more RNA and then to the collapse of the Axis Genetics insect resistance programme.

Expression of the *Diabrotica* microsomal aminopeptidase cDNAs in *E. coli*

The *Diabrotica* MAP cDNAs were to be expressed in *E. coli* to provide a source of MAP protein. This was to be used to immunise mice for the production of a SCAB library and as a probe to pan this library and the proposed total insect gut SCAB library. The antibodies could also be used in preliminary tests to determine whether they could inhibit the MAP enzymatic activity of insect gut extracts or of the purified proteins. It was considered likely that the *E. coli* expressed MAP proteins may be active as aminopeptidase enzymes. The expressed proteins could be tested for activity against various substrates (to determine their preferred substrate and hence if they are, for example, aminopeptidase-A as opposed to aminopeptidase-N) and a Michealis Menten analysis of their kinetics could be performed to determine K_m values (which are a measure of how strongly they bind their substrates). This would more precisely characterise them and the cDNAs from which they originated.

The *E. coli* expression system used was the pET system (Novagen), chosen because it had been successfully used at Axis and they had developed in house a number of derivatives of pET plasmids with improved multipurpose cloning sites. The sequence to be expressed is cloned into the pET plasmid so as to be under the control of the strong bacteriophage T7 transcription and, optionally, translation signals. Expression is induced by providing a source of T7 polymerase in the host cell. The T7 RNA polymerase is so active and selective that almost all the cells resources are converted to target gene expression but in the absence of T7 polymerase the target gene is silent and hence can be stably propagated in *E. coli* strains lacking a T7 polymerase gene. Unless the product protein is very toxic the gene will also be stable in the uninduced state in an expression

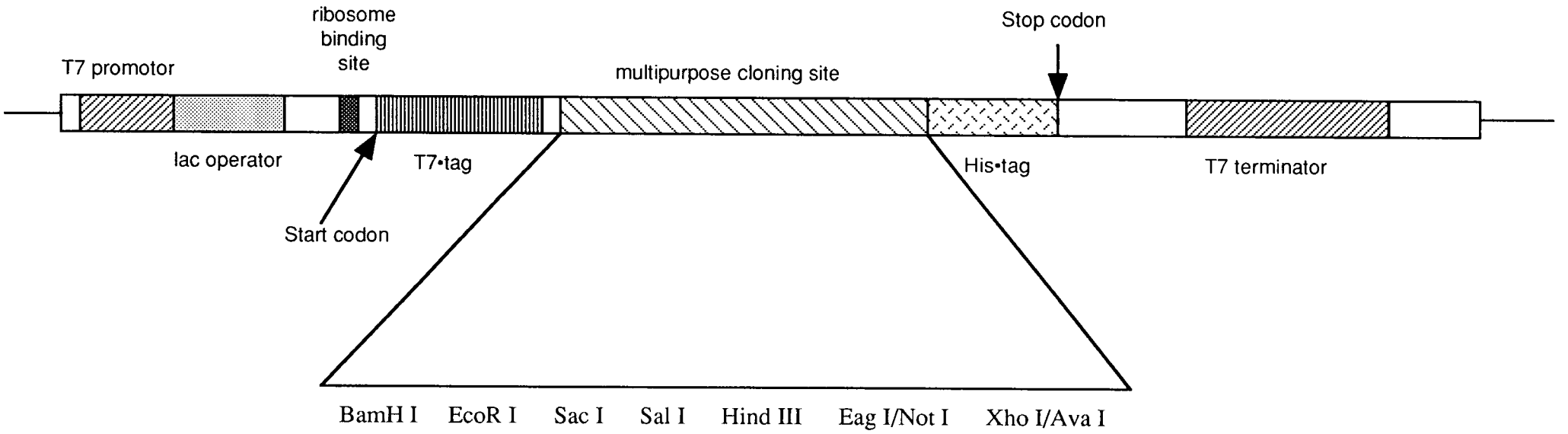


Figure 101. A schematic diagram of the pET-21 a-d series of expression vectors.

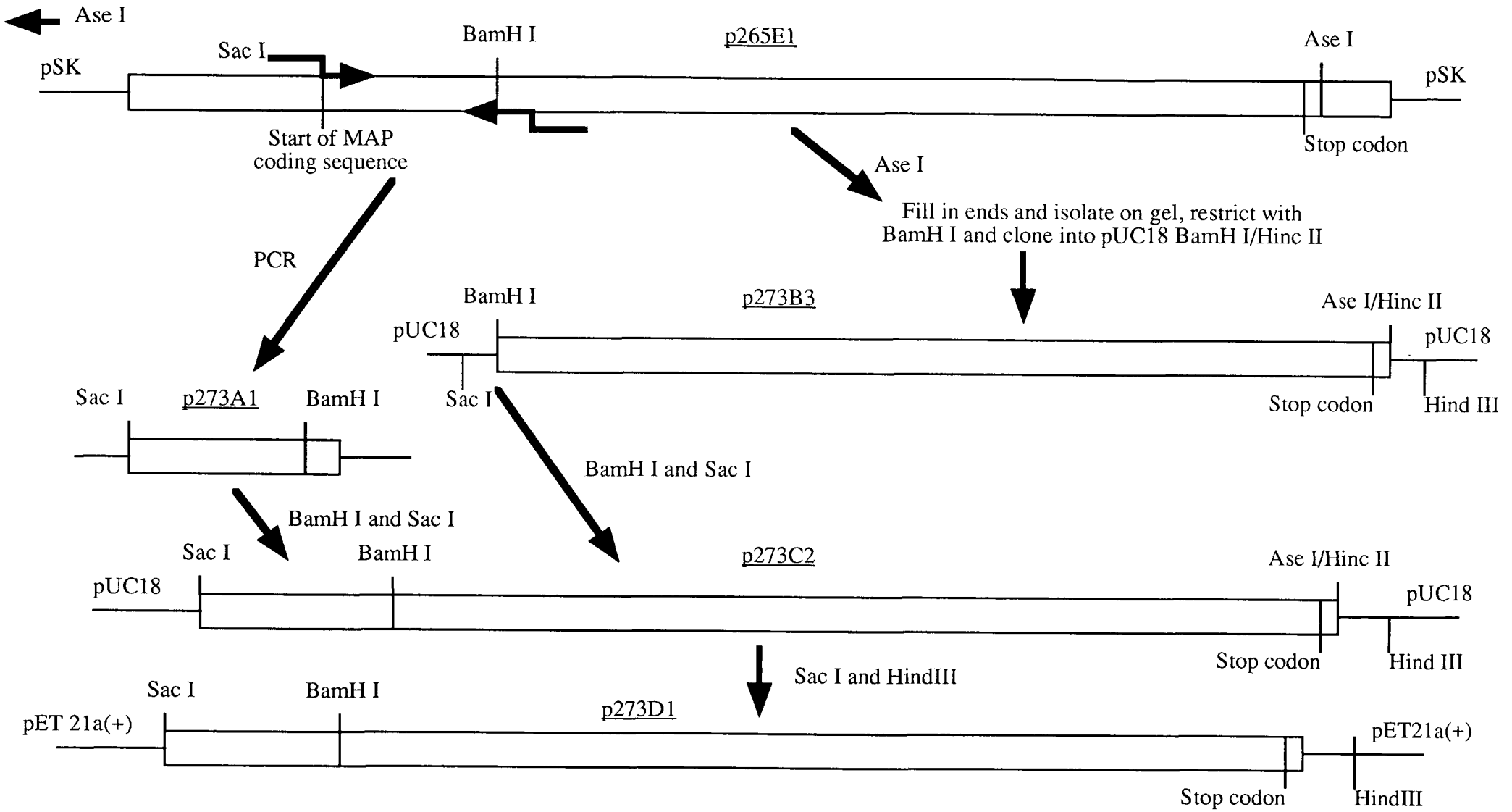


Figure 102. A schematic diagram of the construction of the p265E1 MAP cDNA E. coli expression construct.

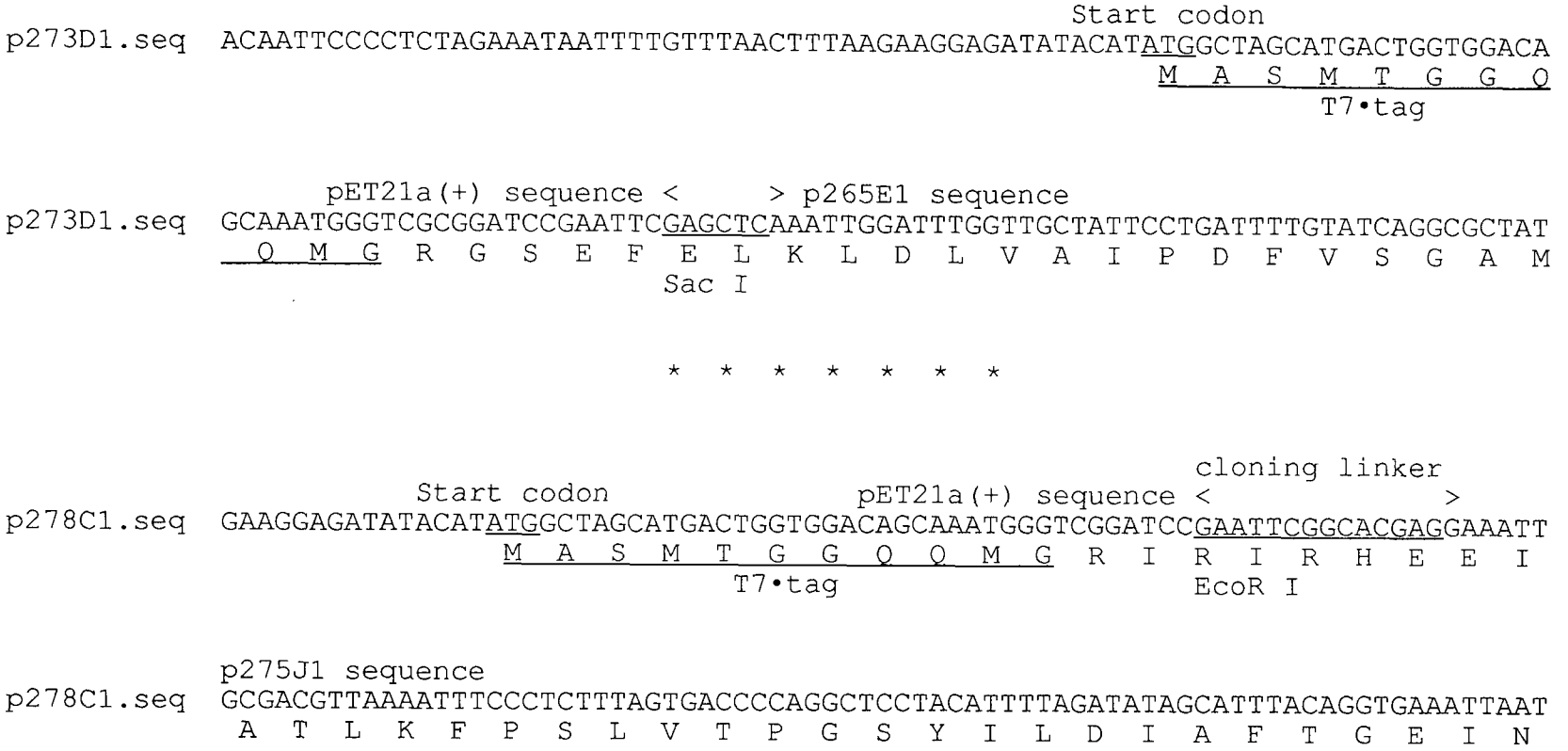


Figure 103. The reading frames about the start codon and fusion to the pET vector for the p265E1 expression construct p273D1 (top) and the p275J1 expression construct p278D1 (bottom).

strain. Expression is carried out in *E. coli* strains which carry a chromosomal copy of the T7 RNA polymerase gene under *lacUV5* control, and is induced by the addition of IPTG.

The pET plasmid range chosen, pET21a, b, c and d(+) (figure 101), carry ampicillin/carbenicillin resistance as a selectable marker and the following series of sequence elements: the T7 promoter sequence (which is the recognition site for T7 RNA polymerase and hence drives the transcription of the sequence down stream from it), the lac operator (which adds an extra layer of suppression of expression of the sequence down stream from it when the system is uninduced; this improves plasmid stability and ease of handling especially when the insert codes for a toxic protein), the highly efficient ribosome binding site from the phage T7 major capsid protein just up stream of the start codon to drive translation, the start codon (which for pET21a-c(+) is part of a Nde I site and reads directly into the eleven amino acid T7•Tag), an N-terminal fusion to the insert to which a monoclonal antibody is available (to allow the easy identification of expressed protein), a multipurpose cloning site (which differs in the reading frame for pET21a, b and c(+) plasmids relative to the T7•Tag) the His•Tag (a C-terminal fusion sequence coding for six histidines followed by a stop codon which allows easy purification as the His•Tag binds to nickel; use of this is optional and would require no stop codon at the end of the insert and the correct translational fusion), and the T7 terminator sequence (which stops the accumulation of RNA from down stream sequences, in this case the β-lactamase gene, and hence the production of large amounts β-lactamase protein)0.

The pET21d(+) vector is identical to pET21c(+) except that the start codon is part of an Nco I site. This is intended for the insertion of target genes directly into the ATG start codon which would give expression without the N-terminally fused T7•Tag.

Assembly of the p265E1 expression construct

The expression construct for p265E1, which was designed to express the whole of the long ORF homologous to MAP from the site of fusion to the artifact cDNA to the stop codon, was assembled as follows (figures 102 and 103):

The 5' artifact cDNA was removed by PCR amplifying a fragment running from the start of the MAP reading frame to the internal BamH I site. For this a 5' oligo including an engineered in Sac I site positioned to allow translational fusion of p265E1 with the reading frame of pET21a(+), and a 3' oligo just down stream of the BamH I site were used. This

fragment was subcloned, the clones checked by DNA sequencing and one without PCR errors, p273A1, was isolated.

The rest of the p265E1 cDNA was isolated as follows; p265E1 was restricted with Ase I and the ends of the fragments generated were filled in to bluntness with Klenow polymerase and dNTPs before being run on gel. The 3543 bp vector to internal fragment which contains the whole of the insert to base 2130, about 40 bp down stream of the stop codon, was cut out from the gel, isolated by electrophoresis, restricted with BamH I and shotgun cloned into the BamH I and Hinc II sites of pUC18. This generated clones with either a 2323 bp insert or the required 1220 bp Ase I to internal BamH I insert. Clones with the correct insert size were selected by gel electrophoresis and p273B3 was selected after being checked by restriction mapping and DNA sequencing in from the 3' end to check the Ase I/Hinc II fusion and the stop codon.

The p265E1 coding sequence was reassembled by restricting p273A1 with BamH I and Sac I and isolating the 560 bp insert fragment and restricting p273B3 with BamH I and Sac I and isolating the ~4000 bp fragment that consists of the vector plus the insert. These were then ligated and transformed into *E. coli* and the recovered clones analysed by restriction mapping and by DNA sequencing across the BamH I site to ensure that the coding sequence was undamaged using the 3' PCR oligo as the sequencing primer. The chosen clone p273C2 was then subcloned with Sac I and Hind III, which is part of the pUC18 multipurpose cloning site at the 3' end of the insert, into pET21a(+) and a clone, p273D1, was checked by restriction mapping and by DNA sequencing using the T7 primer, which reads across the ribosome binding site of pET through the T7•Tag and into the insert, to ensure that the reading frames had fused correctly.

Finally this was transformed into *E. coli* strain BL21(DE3), in which it could be induced for expression, to make p273E1 which was checked by restriction mapping the miniprep.

Assembly of the p275J1 expression construct

Although p275J1 had no artifact to be removed and could be cloned directly into a pET vector for expression it was thought better to remove some of the longish 231 bp 3' untranslated sequence and in particular the polyA tail; p275J1 was, therefore, restricted with Nde I, which cuts at base 2404 about 55 bp down stream of the stop codon, and the

ends filled in with Klenow polymerase and dNTPs. After phenol extraction the DNA was restricted with EcoR I, which cuts the 5'cloning linker, and the 2404 bp insert fragment was isolated on gel and subcloned into the EcoR I to Sma I sites of pUC18. The clones were checked on gel by restriction with EcoR I and Sal I, which cuts the pUC18 multipurpose cloning site 3' of the insert, and p278B2 was checked by DNA sequencing in from the ends. The insert from this was subcloned with EcoR I and Sal I into pET21c(+), which had the correct reading frame from the T7•Tag through the cDNA cloning linker into p275J1 (figure 39), to make p278C1. This was checked by restriction mapping and by sequencing with the T7 primer, which reads across the ribosome binding site of pET through the T7•Tag and into the insert, to ensure that the reading frames had fused correctly. Finally this was transformed into *E. coli* strain BL21(DE3), in which expression could be induced, to make p278D1 which was checked by restriction mapping the miniprep.

Protein expression in *E. coli* of the constructs

Overnights of the expression constructs in the expression host, p273E1 which was made from p265E1 and p278D1 which was made from p275J1, were grown with selection on ampicillin at 100 μg/ml and used to inoculate 50 ml lots of YT broth at one in one hundred dilution. The cultures were grown at 37°C in Erlenmeyer flasks on a rotary shaker at 210 rpm and no more than 20% full to give good aeration. The selection for the cultures to be induced was either 100μg/ml ampicillin or 100μg/ml carbenicillin. The cultures were allowed to grow to mid phase, which correspond to an OD_{600} of 0.6 to 1.0, and then induced with the addition of IPTG to 1 mM. If separate control cultures were not being grown a sample of the culture was taken just prior to induction and was stored on ice or processed as for the induced culture. The induced cultures were grown for a further hour under the same conditions and then harvested by cooling on ice, centrifuging, resuspending in 0.25 volumes of buffer, recentrifuging and storing at -70°C until needed.

Total cell protein was analysed by the cell cracking protocol which uses 2 x SDS sample buffer plus β-mercaptoethanol and boiling to break open the cells and denature and solubilize all the protein; this is then loaded directly on to normal SDS PAGE gels. The result was an obvious strongly stained extra protein band in the induced lanes when compared to the uninduced for both constructs. The p265E1 expression construct gave a

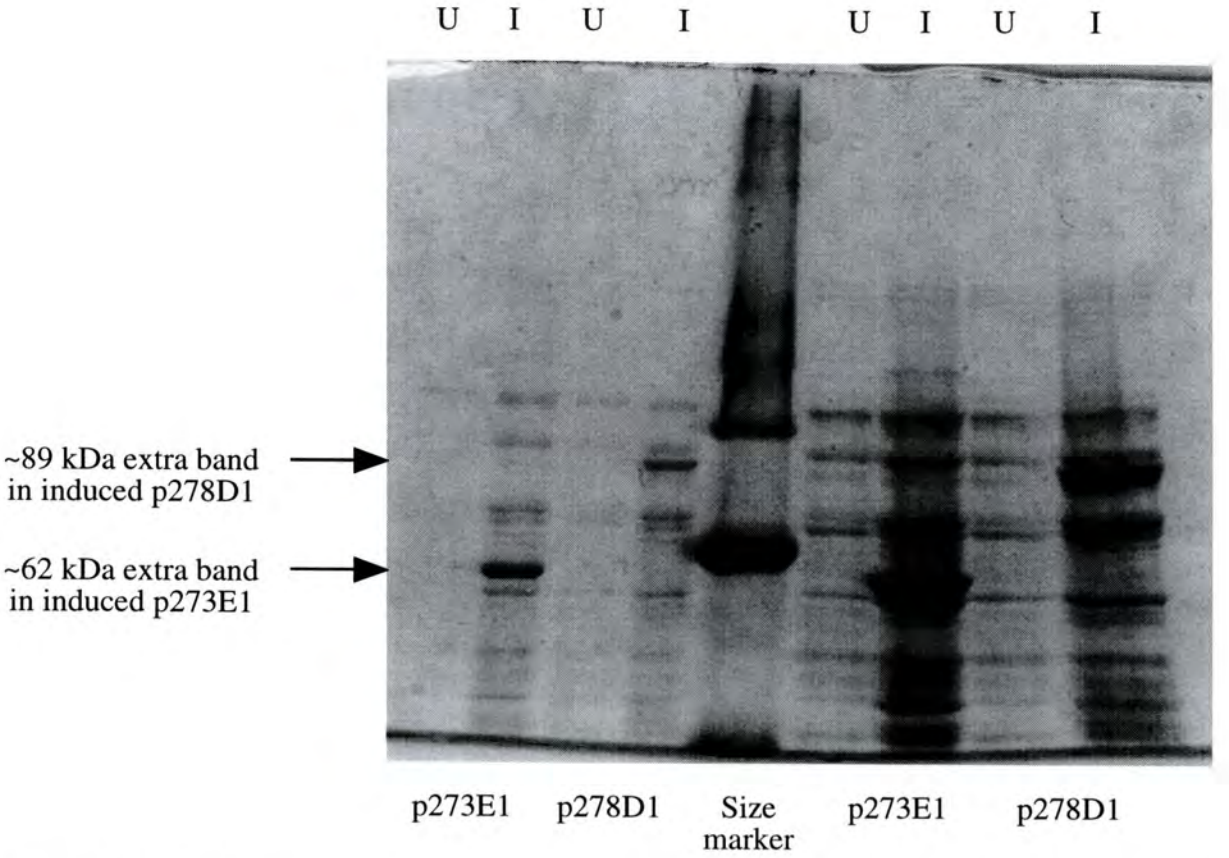


Figure 104. Total protein from *E.coli* cells carrying the microsomal aminopeptidase expression constructs. The tracks on the left are a sixth of the loading of those on the right. U = uninduced, I = induced. The uninduced tracks were run with protein from cells taken from the same culture as the induced cells and frozen just prior to induction. The lower overall levels of protein in these tracks is due to the cell concentration being lower.

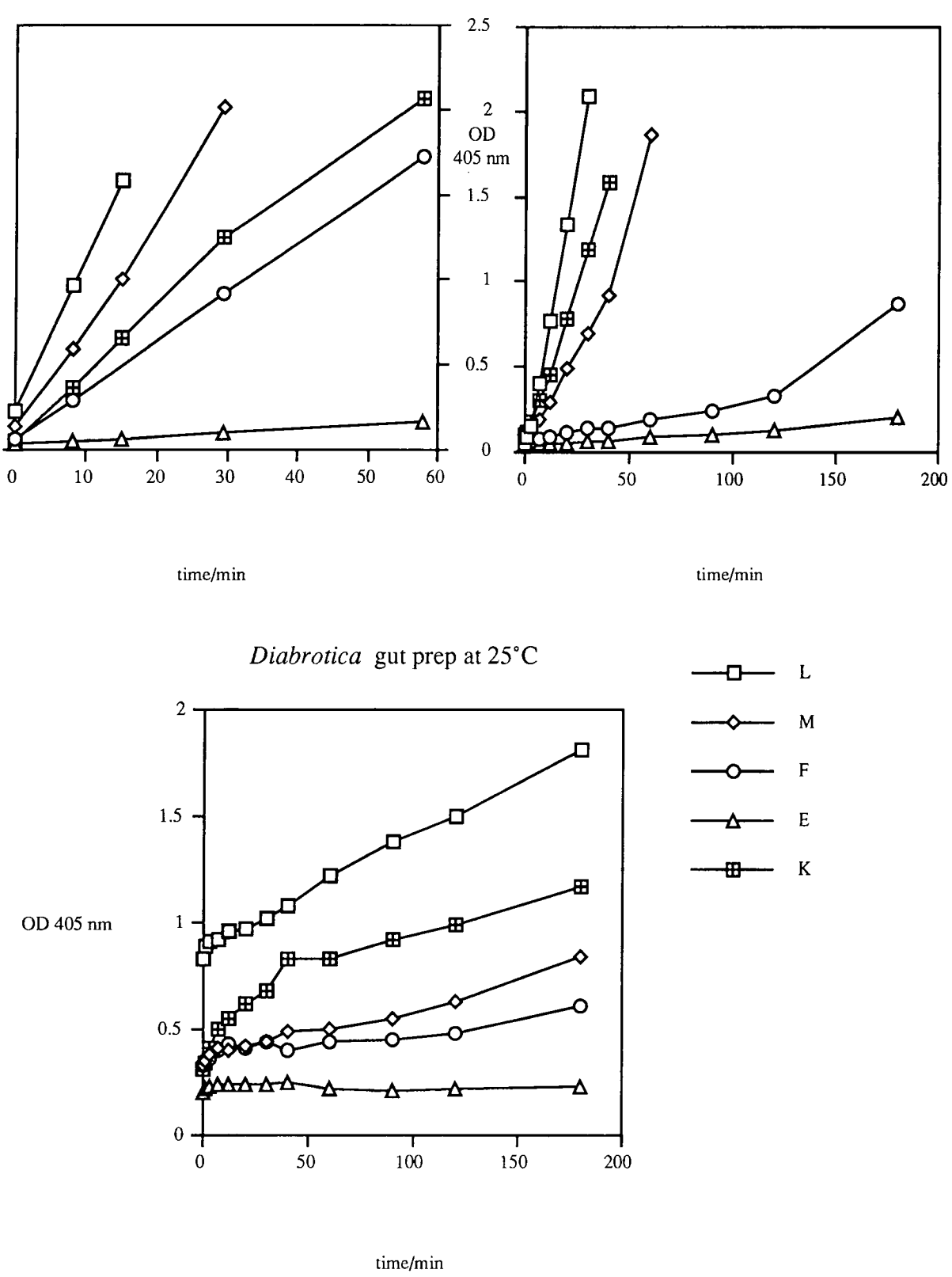


Figure 105. Pilot aminopeptidase assay results.

band at approximately 62 kDa which is slightly smaller than the 72 kDa predicted for the combined reading frame of T7•Tag plus insert of 634 amino acids. The p275J1 construct gave a band at approximately 89 kDa which matches very closely the predicted 91 kDa for the combined reading frame of T7•Tag plus insert of 801 amino acids (figure 104).

Aminopeptidase enzyme assays

Aminopeptidase assays were performed with the paranitroanilide derivatives of the following amino acids as substrates: Leucine (L), which is the amino acid against which aminopeptidase-Ns show the greatest activity, represents the neutral aliphatic side chained amino acids and is the substrate used in the definition of activity of the test enzyme (porcine kidney leucine aminopeptidase SIGMA), Phenylalanine (F), which is also neutral but has an aromatic side chain, Methionine (M), which is also neutral but has a sulphur containing side chain; there are a number of specific aminopeptidases which remove the methionine from the N-terminus of polypeptides and are not structurally related to the MAPs, Glutamic acid (E), which is the amino acid against which aminopeptidase-As show the greatest activity and represents the acidic amino acids, Lysine (K), which represents the basic amino acids.

The intention with the aminopeptidase assays was to characterise the observed aminopeptidase activities from gut preparations, isolated proteins and overexpressed gene products. This was to be accomplished firstly by determining the substrate against which the preparation was most active, i.e. is the protein/dominant activity aminopeptidase-N or aminopeptidase-A or another, and secondly by means of the observed spectrum of activity, i.e. to characterise the preparations by the ratios of the activities against the various substrates.

Due to lack of time only pilot experiments were performed (figure 105). These show that the method is very temperature sensitive. Note the change in the activity spectrum for porcine kidney leucine aminopeptidase between the results obtained at 37°C and those obtained at 25°C. At 37°C the enzyme is most active against leucine (L) then methionine (M), lysine (K), phenylalanine (F) and shows only relatively slight activity against glutamic acid (E). However, at 25°C the activity against phenylalanine has declined dramatically and the activity against lysine is now greater than the activity against methionine.

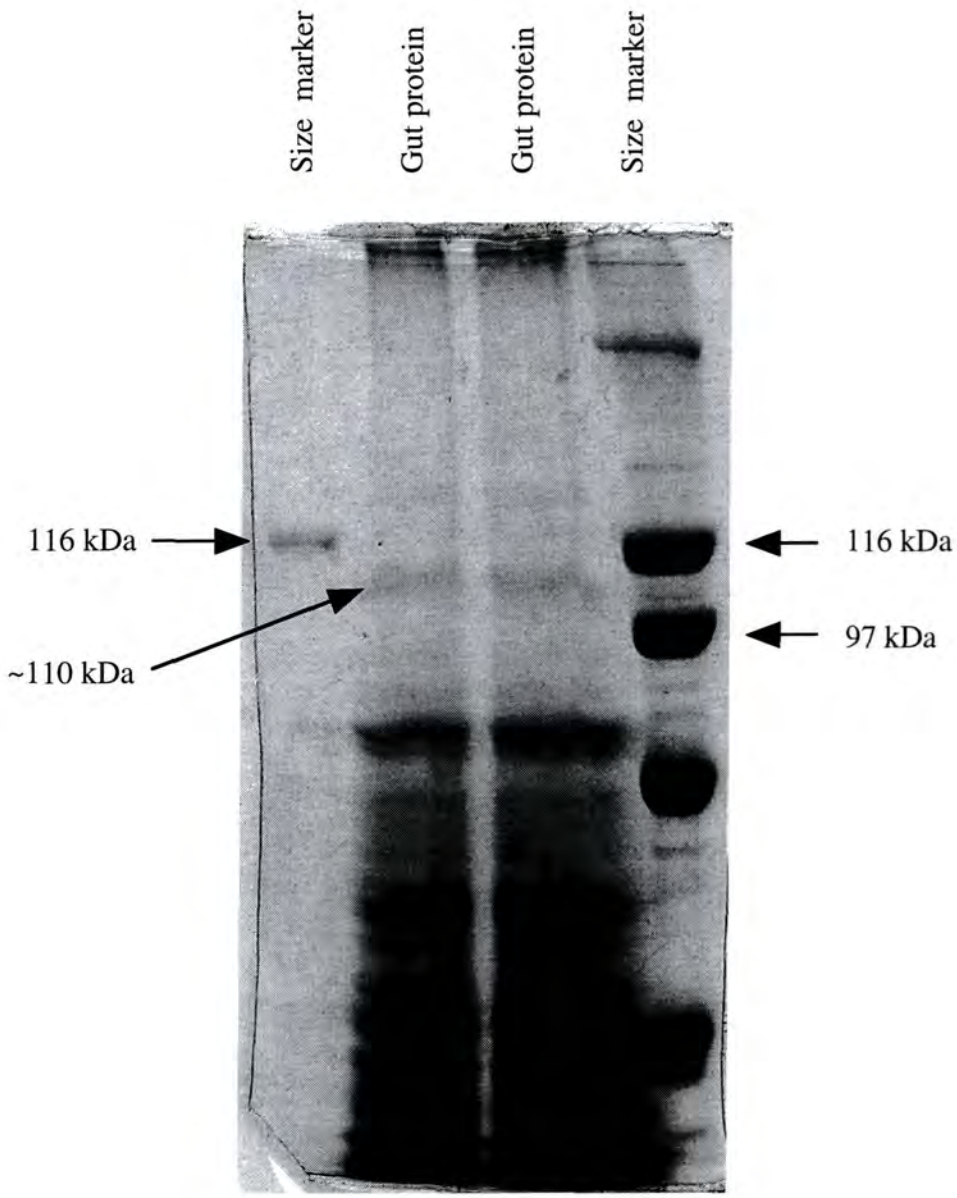


Figure 106. SDS PAGE of integral membrane proteins from *Diabrotica* gut stained with Coomassie blue. The band at approximately 110 kDa is marked and is of the same size as the band found to be microsomal aminopeptidase in similarly prepared nematode gut integral membrane protein (Munn 1992).

An assay at 25°C on a *Diabrotica* gut preparation showed aminopeptidase activity at pH 7.5 against four of the five substrates with the spectrum of activity similar to that of the porcine enzyme at this temperature, i.e. L>K>M>F>E. The cloudy nature of the gut preparations made the initial absorbances quite high and no definite activity against glutamic acid was observed. An attempt was made to assay the purified integral gut membrane protein preparation but no activity was seen either before or after the addition of zinc (results not shown). This was thought to be due to the protein having been denatured by the preparation method and by prolonged storage.

With the acquisition of a temperature-regulated block for the microtitre plate reader to solve the sensitivity to temperature problem this assay method was shown to be a workable method for assaying for aminopeptidase activities.

Integral gut membrane protein preparation

Integral gut membrane protein from *Diabrotica* larvae was prepared as described in the methods. This was used in the aminopeptidase assays (see above) and was run on SDS PAGE. The expected ~110 kDa band presumably corresponding to the MAP band (see Munn 1992) was detected as one of the major bands (figure 106). The isolation of integral membrane proteins demonstrated that a useful probe to screen the proposed SCAB libraries was obtainable from *Diabrotica* larvae.

Zinc binding site peptide

Ashmun, Shapiro and Look (1992) demonstrated that the deletion of the zinc binding motif region of a human aminopeptidase-N not only removed all enzyme activity but also removed the epitopes to which those monoclonal antibodies, raised against the intact protein, which inhibited enzyme activity, bound. Using this data and the sequence of the *Diabrotica* MAPS in this region, a synthetic peptide was designed which was chosen to cover the zinc binding (AHELAHQWFG for the consensus) motif and those amino acids around it which were most highly conserved with the sequence matching that of the *Diabrotica* cDNA p265E1. The peptide designed was seventeen amino acids long as compared with the 117 base 39 amino acid deletion introduced by Ashmun *et al.* (ibid) and corresponded to its more 3' end (figure 107). The 39 amino acid deletion of Ashmun

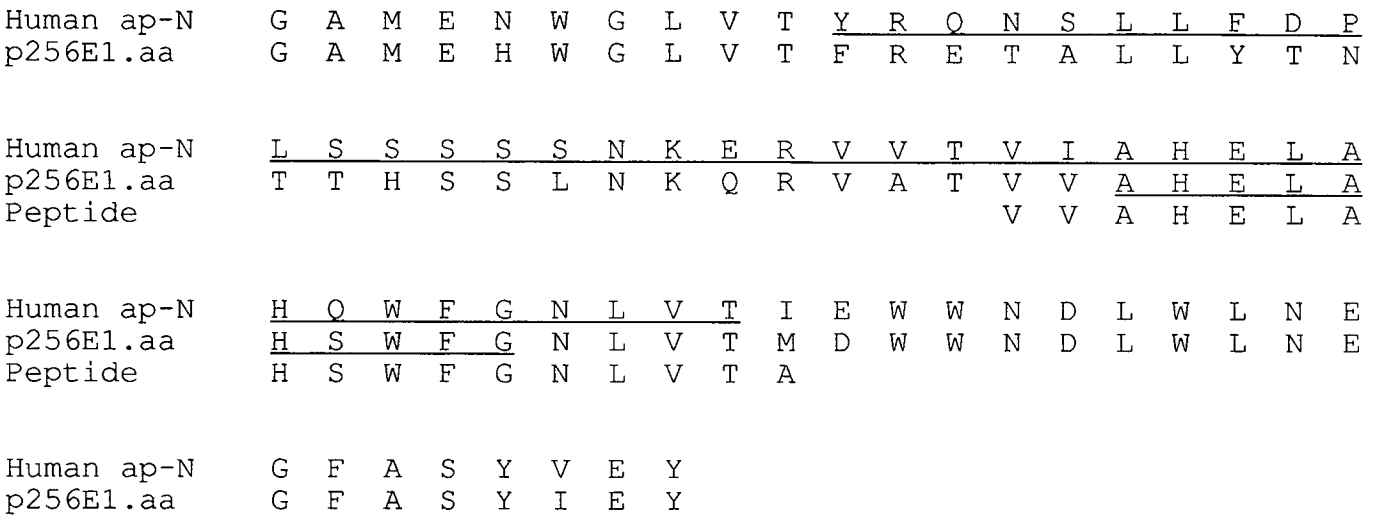


Figure 107. The amino acid sequence in the region of the 39 amino acid deletion of Ashmun, Shapiro and Look (1992) which deletes both the enzyme activity and the epitopes recognised by those monoclonal antibodies which block the enzyme activity. The underlined sequence in the human aminopeptidase-N is the 39 amino acid deleted sequence the underlined sequence in the p265E1 sequence is the zinc binding motif. The final alanine of the synthetic peptide was dictated by the availability of support resins for the synthesis machine.

et al. (ibid) runs from the tyrosine (Y) just down stream of the highly conserved region GAMENWGL 39 amino acids to just beyond the conserved valine (V) down stream of the zinc binding motif. The C-terminal alanine (A) of the peptide was dictated by the available resin for the peptide synthesis machine.

The intention was to use this synthetic peptide to raise antibodies in rabbit which, it would be expected, would block the enzymatic activity of *Diabrotica* MAPS. This could be tested for by *in vitro* assays and if successful the antibody could be fed to *Diabrotica* larvae in artificial diet bioassays to investigate any effects that the inhibition of aminopeptidase activity in their guts had on growth, development and survival. This experiment was to have been a demonstration of the ability of antibodies to affect insect metabolism in vivo and hence to further demonstrate the feasibility of the SCAB programme. Due to the collapse of the Axis Genetics insect resistance programme resources for rabbits in which to raise the antibodies were unavailable and the peptide was stored for future use.

Manduca aminopeptidase-N sequence

Knight, Knowles and Ellar (1995) have published a sequence of a *Manduca sexta* aminopeptidase-N which is a receptor for the *Bacillus thuringiensis* CryIA(c) toxin. This was compared with the MAP sequences I had obtained and was homologous to them all but was not identical to either of the *Manduca* MAP PCR products. Surprisingly it was less homologous to the *Diabrotica* MAP cDNAs than they were to the mammalian sequences (21.2% identity to p265E1 and 26.4% identity to p275J1) but the strongly conserved regions around the zinc binding motif were closely homologous. Interestingly this *Manduca* aminopeptidase-N has a divergent zinc binding motif; IVSHEIAHMWFG, somewhat like those of the PCR products p264B1 and p264A17 (see figures 83, 84, 85 and 86).

Concluding remarks

It has been demonstrated that, at least in principal, antibodies can be raised that bind to targets in the insect gut at the pH conditions which occur there. In particular they bind in the pH conditions found in *Diabrotica* gut and are reasonably resistant to

Diabrotica gut protease enzymes. This demonstrates the feasibility of using antibodies as anti-insect proteins and hence the viability of the SCAB library approach to generating novel anti-insect genes.

The MAP programme has produced two cDNAs both of which have been fully sequenced and over expressed in *E. coli* along with a number of PCR products from the highly conserved zinc binding region. The programme had reached the stage where experiments to test the ability of antibodies to be used as anti-insect proteins were feasible, as the synthetic peptide and the over expressed proteins had been produced, but unfortunately the programme was halted at this stage due to the collapse of the Axis Genetics insect resistance programme.

Final discussion

The potato carboxypeptidase inhibitor project was carried through to an insect bioassay on transgenic plants expressing the PCI gene. Unfortunately the plants expressing PCI were more susceptible to insect attack than both the controls and the homozygous non-expressing experimentals. This result does not support the hypothesis of expression of single trans-genes to increase insect resistance. Further work on this gene to increase expression levels and to determine the reason for the increased insect susceptibility in the bioassay were precluded by lack of commercial interest. However, valuable lessons were learned from the defects in the insect bioassay design.

The *Manduca* haemolymph trypsin inhibitor project was carried through to transgenic plants expressing protein from the introduced gene and both artificial diet and transgenic plant insect bioassays. For commercial reasons, the bioassays were performed elsewhere and were highly flawed in design. However, there were indications that the expression of MHTI in plants does enhance insect resistance and this was confirmed in the artificial diet bioassays with the purified *E. coli* expressed protein. These results support the hypothesis.

Thus, while only one of the two projects taken to the transgenic plant stage gave any indications of enhanced insect resistance, this approach to the production of insect resistant plants has been demonstrated to be viable. This molecular approach, where transgenic plants expressing potential anti-insect protein genes and purified *E. coli* expressed protein are employed to assay for insecticidal activity, is especially suitable where the anti-insect protein cannot be readily isolated in sufficient quantity for *in vitro* assays and artificial diet bioassays. For novel inhibitors and the single chain antibodies, the molecular approach is the only route and, while a demonstration of enhanced insect resistance by this approach was not achieved, its feasibility was demonstrated.

Both the single chain antibody and the microsomal aminopeptidase projects were abandoned at a stage prior to the demonstration of any anti-insect activity. However, the initial proving experiments had been carried out and it was demonstrated that antibody lines active at insect gut pHs and reasonably resistant to insect gut proteases could be isolated. This demonstrated the feasibility of the approach and in particular demonstrated that antibodies lines which would inhibit aminopeptidase in the insect gut could be isolated. Thus, the demonstration of potentially insecticidal antibodies required only the raising of

antibodies against the MAP active site peptide, which had been synthesised, and an insect feeding trial with these antibodies. The SCAB project was stopped at a much earlier stage, however the feasibility of using antibodies against the hypothetical gut targets had been demonstrated.

References

Ako, H., G. M. Hass, D. T. Grahn, and H. Neurath. (1976). Carboxypeptidase inhibitor from potatoes. Interaction with derivatives of carboxypeptidase A. Biochemistry, **15**, 2573-2578.

Apple, R. T. and J. W. Fristrom. (1991). 20-hydroxyecdysone is required for, and negatively regulates, transcription of Drosophila pupal cuticle protein genes. Dev. Biol., **146**, 569-582.

Applebaum, S. W. (1985). Biochemistry in digestion. In Comprehensive insect physiology, biochemistry and pharmacology. (Ed. G. A. Kerkut and L. I Gilbert), Pergamon press, Oxford.

Ashida, M. and K. Dohke. (1980). Activation of phenoloxidase by the activating enzyme of silkworm *Bombyx mori* . Insect Biochem., **10**, 37-47.

Ashmun, R. A., L. H. Shapiro, and A. T. Look. (1992). Deletion of the zinc binding motif of CD13/aminopeptidase N molecules results in loss of epitopes that mediate binding of inhibitory antibodies. Blood, **79**, 3344-3349.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith, and K. Struhl, Ed. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, 1992.

BAA. (1994). British agrochemicals association annual review and handbook 1994. BAA Ltd., Peterborough.

Baulcombe, D.C., G.R. Saunders, M.W. Bevan, M.A. Mayo, and B.D. Harrison. (1986). Expression of biologically-active viral satellite RNA from the nuclear genome of transformed plants. Nature, **321**, 446-449.

Bevan, M. (1984). Binary *Agrobacterium* vectors for plant transformation. Nucleic acids research, **12**, 8711-8721.

Bevan, M., W.M. Barnes, and M.D. Chilton. (1983). Structure and transcription of the nopaline synthase gene region of tDNA. Nucleic Acids Research, **11**, 369-385.

Bevan, M., S.E. Mason, and P. Goelet. (1985). Expression of the tobacco mosaic virus coat protein by a cauliflower mosaic virus promoter in plants transformed by *Agrobacterium* . EMBO J., **4**, 1921-1926.

Birke, Y., A. Gertler, and F. Kalef. (1963). A pure trypsin inhibitor from soybeans. Biochem. J., **87**, 281-284.

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

Bohren, K. M., B. Bullock, B. Wermuth, and K. H. Gabbay. (1989). The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences of human aldehyde and aldose reductases. J. Biological Chemistry, **264**, 9547-9551.

Boller, T., A. Gehri, F. Mauch, and U. Vögeli. (1983). Chitinase in bean leaves: induction by ethylene, purification, properties, and possible function. Planta, **157**, 22-31.

Bonaldo, P and A Colombatti. (1989). The carboxy terminus of the chicken α3 chain of collagen VI is a unique mosaic structure with glycoprotein Ib-like, fibronectin type III, and Kunitz modules. J. Biol. Chem., **264**, 20235-20239.

Bone, L. W., K. P. Bottjer, and S. S. Gill. (1988). Factors affecting the larvicidal activity of *Bacillus thuringiensis israelensis* toxin for *Trichostrongylus colubriformis* . J. Invertebr. Path., **52**, 102-107.

Bonner, T.I., D. J. Brenner, B. R. Neufeld, and R. J. Britten. (1973). Reduction in the rate of DNA reassociation by sequence divergence. J. Mol. Biol., **81**, 123.

Bouhin, H., J. P. Charles, B. Quennedey, and J. Delachambre. (1992). Developmental profiles of epidermal messanger-RNAs during the pupal adult moult of *Tenebrio molitor* and isolation of a cDNA clone encoding an adult cuticular protein - effects of a juvenile hormone analogue. Dev. Biol., **149**, 112-122.

Boulter, D., G. A. Edwards, A. M. R. Gatehouse, J. A. Gatehouse, and V. A Hilder. (1990). Additive effects of incorporating two different higher plant derived insect resistance genes in transgenic tobacco plants. Crop protection, **9**, 351-354.

Branson, T. F. (1987). The contribution of prehatch and posthatch development to protandry in the chrysomelid, *Diabrotica virgifera virgifera* . Entomol. Exp. Appl., **43**, 205-208.

Branson, T. F., P. L. Guss, J. L. Krysan, and G. R. Sutter. (1975). Corn rootworms; Laboratory rearing and manipulation. US Dept. of Agric., **ARS-NC-28**, 18.

Broadway, R. M. and S. S. Duffey. (1986). Plant proteinase inhibitors: mechanism of action and effect on the growth and digestive physiology of *Heliothus zea* and *Spodoptera exigua* . J. Insect Physiol., **32**, 823-827.

Broadway, R. M. and S. S. Duffey. (1988). The effect of plant protein quality on insect digestive physiology and the toxicity of plant proteinase inhibitors. J. insect Physiol., **34**, 1111-1117.

Broadway, R. M., S. S. Duffey, G. Pearce, and C. A. Ryan. (1986). Plant proteinase inhibitors: A defence against herbivorous insects? Ent. exp. appl., **41**, 33-.

Burgess, E. P. J., C. A. Main, P. S. Stevens, J. T. Christeller, A. M. R. Gatehouse, and W. A. Laing. (1994). Effects of protease inhibitor concentration and combinations on the survival, growth and gut enzyme activities of the black field cricket, *Teleogryllus commodus*. J. Insect Physiol., **40**, 803-811.

Burgess, E. P. J., P. S. Stevens, G. K. Keen, W. A. Liang, and J. T. Christeller. (1991). Effects of protease inhibitors and dietary protein level on the black field cricket, *Teleogryllus commodus* . Entomologia exp. appl., **61**, 123-130.

Calabrese, D. M., K. W. Nickerson, and L. C. Lane. (1980). Can. J. Microbiol., **26**, 1006-1010.

Caprioglio, D. R., C. Padilla, and M. Werner-Washburne. (1993). Isolation and characterisation of AAP1: A gene encoding an alanine/argenine aminopeptidase in yeast. J.

Biol. Chem., **268**, 14310-14315.

Casey, J. and N. Davidson. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. Nucleic Acids Res., **4**, 1539.

Cechova, D. and E. Ber. (1974). Collect. Czech. Chem. Commun., **39**, 680-688.

Cechova, D., V. Jonakova, and F. Sorm. (1971). Collect. Czech. Chem. Commun., **36**, 3352-3357.

Cechova, D. and G. Muszynska. (1970). FEBS Lett., **8**, 84-86.

Chaing, H. C. (1973). Bionomics of the northern and western corn rootworms. Ann. Rev. Entomol., **18**, 47-72.

Charles, J. P., H. Bouhin, B. Quennedey, A. Courrent, and J. Delachambre. (1992). cDNA cloning and deduced amino acid sequence of a major, glycine-rich cuticular protein from the coleopteran *Tenebrio molitor*. Temporal and spatial distribution of the transcript during metamorphosis. Eur. J. Biochem., **206**, 813-819.

Chen, Z. W., B. Agerberth, K. Gell, M. Andersson, V. Mutt, C. G. Oestenson, S. Efendic, J. Barros-Soederling, B. Persson, and H. Joernvall. (1988). Isolation and characterisation of porcine diazepam-binding inhibitor, a polypeptide not only of cerebral occurrence but also common in intestinal tissues and with effects on regulation of insulin release. Eur. J. Biochem., **174**, 239-245.

Chrispeels, M. J. and N. V. Raikhel. (1991). Lectins, lectin genes and their role in plant defence. Plant Cell, **3**, 1-9.

Christeller, J. T., A. M. R. Gatehouse, and W. A. Laing. (1994). The interaction of the elastase inhibitor, Eglin c, with insect digestive endopeptidases: Effect of pH on the dissociation constants. Insect Biochem. Molec. Biol., **24**, 103-109.

Christeller, J. T., W. A. Laing, N. P. Marwick, and E. P. J. Burgess. (1992). Midgut protease activities in 12 phytophagous lepidopteran larvae: Dietary and protease inhibitor interactions. Insect Biochem. Molec. Biol., **22**, 735-746.

Chung, A. Y. K., J. W. Ryan, J. A. Nearing, M. A. Rogoff, and P. Berryer. (1993). An improved radioassay for aminopeptidase N. FASEB J., **7**, 814.

Clore, G. M., A. M. Gronenborn, M. Nilges, and C. A. Ryan. (1987). Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. Biochemistry, **26**, 8012-8023.

Coghlan, A. (1992). Survival of the fittest molecules. New Scientist, **3 October**, 37-40.

Colepicolo-Neto, P., E. J. H. Bechara, C. Ferreira, and W. R. Terra. (1987). Digestive enzymes in close and distant genera of a same family:properties of midgut hydrolases from luminescent *Pyrophorus divergens* (Coleoptera: Elateridae) larvae. Comp. Biochem. Physiol., **87B**, 755-759.

Croy, R. R. D. and J. A. Gatehouse. (1985). Genetic engineering of seed proteins:

current and potential applications. In Plant Genetic Engineering (Ed. J. H. Dodds), Cambridge University Press,

Czapla, t. H. and B. A Lang. (1990). Effects of plant lectins on the larval development of European corn borer (Lepidoptera, Pyralidae) and Southern corn rootworm (Coleoptera, Chrysomelidae). J, Econ Entomol., **83**, 2480-2485.

Dagert, M. and S. D. Ehrlich. (1979). Prolonged incubation in calcium chloride improves the competence of *Escherichia coli* cells. Gene, **6**, 23.

de Olivera, J. T. A. "Seed lectins; the effects of dietary *Phaseolus vulgaris* lectin on the general metabolism of monogastric animals." PhD, Aberdeen University, 1986.

Deng, W, W. S. Greyburn, T. R Hamilton-Kemp, G. B. Collins, and D. F Hildebrand. (1992). Expression of soybean-embryo lipoxygenase 2 in transgenic tobacco tissue. Planta, **187**, 203-208.

Dove, W. F. and N. Davidson. (1962). Cation effects on the denaturation of DNA. J. Mol. Biol., **5**, 467.

Dulmage, H. T., Ed. Microbial control of pest and plant diseases. 1970-1980. London: Academic press, 1980.

ECPA. (1994). European crop protection association annual report 1993-1994.

Edmonds, H. S. "Antimetabolic proteins from plants and their potential use in conferring resistance against corn rootworms (*Diabrotica* sp.)." PhD, University of Durham, 1994.

Edmonds, H.S., L.N. Gatehouse, V.A. Hilder, and J.A. Gatehouse. (1995). The inhibitory effects of the cysteine protease inhibitor, oryzacystatin, on digestive proteases and on larval survival and development of the Southern Corn Rootworm (*Diabrotica undecimpunctata howardii*). Entomol. exp. appl., **In press**,

Edwards, G. A. "Plant transformation using an *Agrobacterium tumefaciens* Ti-plasmid vector system." PhD, Durham, 1988.

Eguchi, M, I Haneda, and A Iwamoto. (1982). Properties of protease inhibitors from the haemolymph of silkworms, *Bombyx mori*, *Antheraea pernyi* and *Philosamia cynthia ricini*. Comp. Biochem. Physiol., **71B**, 569-576.

Ellis, D.J. Confidential report.
The evaluation of proteins and transgenic plants for insect resistance. 1995.

Ellis, T.H.N., D.R. Davies, J.A. Castleton, and I.D. Bedford. (1984). The organisation and genetics of rDNA length variants in peas. Chromosoma, **91**, 74-81.

Enjyoji, K. I., M. Emi, T. Mukai, and H. Kato. (1992). cDNA cloning and expression of rat tissue factor pathway inhibitor. J. Biochem., **111**, 681-687.

Evans, R. J., A. Pusztai, W. B. Watt, and D. H. Bauer. (1973). Isolation and properties of protein fractions from navy beans (*Phaseolus vulgaris*) which inhibit growth of rats. Biochim. Biophys. Acta., **303**, 175-184.

Feeny, P. P. (1976). Plant apparency and chemical defence. Rec. Adv. Phytochem., **10**,

1-40.

Feinberg, A. P. and B. Vogelstein. (1983). A technique for labelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem., **132**, 6-13.

Feinberg, A. P. and B. Vogelstein. (1984). Addendum to; A technique for labelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem., **137**, 266-267.

Felsot, A. S., J. G. Wilson, D. E. Kuhlman, and K. L. Steffey. (1982). Rapid dissipation of carbofuran as limiting factor in corn rootworm (*Coleoptera; Cyrsomelidae*) control in fields with histories of continuous carbofuran use. J. Econ. Entomol., **75**, 1098-1103.

Fossum, K. (1970). Proteolytic enzymes and biological inhibitors. III. Naturally occurring inhibitors in some animal and plant materials and their effects on proteolytic enzymes of various origins. Acta. Pathol. Microbiol. Scand. Sect. B, **78**, 741.

Freitelson, J. S., J. Payne, and L. Kim. (1992). *Bacillus thuringiensis* : insects and beyond. Bio/Technology, **10**, 271-275.

Frohman, M. A. (1990). Race: Rapid amplification of cDNA ends. In PCR protocols; A guide to methods and applications (Ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White), pp.28-38. Academic Press Inc.,

García-Alvarez, N., R. Cueva, and P. Suárez-Rendueles. (1991). Molecular cloning of soluble aminopeptidases from *Saccharomyces cerevisiae* : Sequence analysis of aminopeptidase yscII, a putative zinc-metallopeptidase. Eur. J. Biochem., **202**, 993-1002.

Garczynski, S. F. and M. A. Adang. (1995). *Bacillus thuringiensis* cryIA(c) ∂-endotoxin binding aminopeptidase in the *Manduca sexta* midgut has a glucosyl-phosphatidylinositol anchor. Insect Biochem. Molec. Biol., **25**, 409-415.

Gatehouse, A. M. R., L. Barbieri, F. Stirpe, and R. R. D. Croy. (1990). Effects of ribosome inactivating proteins on insect development - differences between Lepidoptera and Coleoptera. Entomol. Exp. Appl., **54**, 43-51.

Gatehouse, A. M. R. and D. Boulter. (1983). Assessment of the anti-metabolic effects of trypsin inhibitors from cowpea (*Vigna unguiculata*) and other legumes on development of the bruchid beetle *Callosobruchus maculatus* . J. Sci. Food Agric., **34**, 345-350.

Gatehouse, A. M. R., F. M. Dewey, J. Dove, K. A. Fenton, and A. Pusztai. (1984). Effects of seed lectins from *Phaseolus vulgaris* on the development of larvae of *Callosobruchus maculatus* . Mechanism of toxicity. J. Sci. Food Agric., **35**, 373-380.

Gatehouse, A. M. R., K. A. Fenton, I. Jepson, and D. J. Pavey. (1986). The effects of α-amlyase inhibitors on insect storage pests: Inhibition of α-amlyase *in vitro* and effects on development *in vivo* .

Gatehouse, A. M. R., J. A. Gatehouse, and D. Boulter. (1980). Isolation and characterisation of trypsin inhibitors from cowpea (*Vigna unguiculata*). Phytochemistry, **19**, 751-756.

Gatehouse, A. M. R., J. A. Gatehouse, P. Dobie, A. M. Kilminster, and D. Boulter. (1979). Biochemical basis of insect resistance in *Vigna unguiculata* . J. Sci. Food Agric., **30**, 948-959.

Gatehouse, A. M. R., D. S. Howe, J. E. Flemming, V. A. Hilder, and J. A. Gatehouse. (1991). Biochemical basis of insect resistance in winged bean (*Psophocarpus tetragonolobus*) seeds. J. Sci. Food Agric., **55**, 63-74.

gcg. Program Manual for the Wisconsin Package. Version 8 Ed., Vol. Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711, 1994.

Gergrich, R. C., H. A. Scott, and J. P. Fulton. (1986). Evaluation of *Diabrotica* beetles as vectors of plant viruses. In Methods for the study of pest *Diabrotica* . (Ed. J. L. Krysan and T. A. Miller), Springer-Verlag, New York.

Gill, S. S., E. A. Cowles, and P. V. Pietrantonio. (1992). The mode of action of *Bacillus thuringiensis* endotoxins. Ann. Rev. Entomol., **37**, 615-636.

Goldberg, L. J. and J. Margalit. (1977). J. Mosquito news, **37**, 355-358.

Goldstein, I. J., R. C. Hughes, M Monsigny, T. Osawa, and N. Sharon. (1980). What should be called a lectin? Nature, **285**, 66.

Graham, J. S. and C. A. Ryan. (1981). Accumulation of a metallo-carboxypeptidase inhibitor in leaves of wounded potato plants. Biochem. Biophys. Res. Comm., **101**, 1164-1170.

Grant, G., P. M. Dorward, and A. Pusztai. (1993). Pancreatic enlargement is evident in rats fed diets containing raw soybeans (*Glycine max*) or cowpeas (*Vigna unguiculata*) for 800 days but not in those fed kidneybeans (*Phaseolus vulgaris*) or lupinseed (*Lupinus angustifolius*). J. Nutr., **123**, 2207-2215.

Green, T. R. and C. A. Ryan. (1972). Wound-induced proteinase inhibitor in plant leaves: a possible defence mechanism against insects. Science, **175**, 776-777.

Guilley, H., R.K. Dudley, G. Jonard, E. Balazs, and K.E. Richards. (1982). Transcription of cauliflower mosaic-virus DNA - detection of promoter sequences, and characterisation of transcripts. Cell, **30**, 763-773.

Gustafson, G. and C. A. Ryan. (1976). Specificity of protein turnover in tomato leaves. Accumulation of proteinase inhibitors, induced with the wound hormone, proteinase inhibitor IF. J. Biol. Chem., **251**, 7004-7010.

Hames, B. D. (1981). An introduction to polyacrylamide gel electrophoresis. In Gel electrophoresis of proteins: A practical approach. (Ed. B. D. Hames and D. Rickwood), pp.1-91. IRL Press, Oxford.

Harlow, E. and D. Lane. (1988). Antibodies a laboratory manual. Cold Spring Harbour Laboratory,

Hass, G. M. and Ryan. C. A. (1980). Carboxypeptidase inhibitor from ripened tomatoes; purification and properties. Phytochemistry, **19**, 1329-1333.

Hass, G. M., S. P. Ager, D. Le Tourneau, J. E. Derr-Makus, and D. J. Makus. (1981).

Specificity of the carboxypeptidase inhibitor from potatoes. Plant Physiol., **67**, 754-758.

Hass, G. M., H. Ako, D. T. Grahn, and H. Neurath. (1976). Carboxypeptidase inhibitor from potatoes. The effects of chemical modifications on inhibitory activity. Biochemistry, **15**,

Hass, G. M., J. E. Derr, and D. J. Makus. (1979). Distribution of carboxypeptidase isoinhibitors in the potato plant. Plant Physiol., **64**, 1029-1031.

Hass, G. M., J. E. Derr, D. J. Makus, and C. A. Ryan. (1979). Purification and characterisation of the carboxypeptidase isoinhibitors from potatoes. Plant Physiol., **64**,

Hass, G.M. and M.A. Hermondson. (1981). Amino acid sequence of a carboxypeptidase inhibitor from tomato fruit. Biochemistry, **20**, 2256-2660.

Hass, G. M., H. Nau, K. Biemann, D. T. Grahn, L. H. Ericsson, and H. Neurath. (1975). The amino acid sequence of a carboxypeptidase inhibitor from potatoes. Biochemistry, **14**, 1334-1342.

Hass, G.M. and C.A. Ryan. (1980). Cleavage of the carboxypeptidase inhibitor from potatoes by carboxypeptidase-A. Biochem. Biophys. Res. Commun., **97**, 1481-1486.

Hass, G. M. and C. A. Ryan. (1981). Carboxypeptidase inhibitor from potatoes. Meth. Enzymol., **80**, 778-791.

Haycock, K.A., J Roth, J Gagnon, W.F. Finzer, and C. Soper. StatView. 4.1 Ed., Vol. Berkley CA USA: Abacus Concepts Inc, 1992.

Heinrichs, E. A. and O. Mochida. "From secondary to major pest status: The case of insecticide-induced rice brown plant hopper, *Nilaparvata lugens* resurgence." In Proc. XV Pacific science congress in New Zealand, Year.

Helene, A., A. Beaumont, and B. P. Roques. (1991). Functional residues at the active site of aminopeptidase N. Eur. J. Biochem., **196**, 385-393.

Herrnstadt, C., G.G. Soares, E.R. Wilcox, and D.L. Edwards. (1986). A new strain of *Bacillus thuringiensis* with activity against Coleopteran insects. Biotechnology, **4**, 305-308.

Hilder, V. A., R. F. Barker, R. A. Samour, A. M. R. Gatehouse, J. A. Gatehouse, and D. Boulter. (1989). Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguicilata* Walp.). Plant Mol. Biol., **13**, 701-710.

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

Hilder, V. A., K. S. Powell, A. M. R. Gatehouse, J. A. Gatehouse, L. N. Gatehouse, Y. Shi, W. D. O. Hamilton, A. Merryweather, C. A. Newell, J. C. Timans, W. J. Peumans, E. van Damme, and D. Boulter. (1995). Expression of snowdrop lectin in transgenic tobacco plants results in added protection against aphids. Transgenic Research, **4**, 18-25.

Hill, D. S. (1983). Agricultural insect pests of temperate regions and their control. Cambridge University Press,

Hill, D. S. (1983). Agricultural insect pests of the tropics and their control. Cambridge University Press,

Hoffman, C., P. Lüthy, R. Hutter, and V Pliska. (1988). Binding of the delta endotoxin from Bacillus thuringiensis to brush border membrane vesicles of the cabbage butterfly (*Pieris brassicae*). Eur. J. Biochem., **173**,

Hoffman, L. M. and D. D. Donaldson. (1985). Characterisation of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J., **4**, 883-889.

Hoffman, L. M., Y Ma, and R. F. Barker. (1982). Molecular cloning of *Phaseolus vulgaris* lectin mRNA and use of cDNA as a probe to estimate transcript levels in various tissues. Nucleic Acids Res., **10**, 7819-7828.

Hojima, Y., H. Moriya, C. Moriwaki, and C. A. Ryan. (1979). Metalloenzyme inhibitor from kidney beans. Partial purification and characterisation. Plant Physiol., **63**, 562-566.

Holländer-Czytko, H., J. K. Andersen, and C. A. Ryan. (1985). Vacuolar localisation of wound-induced carboxypeptidase inhibitor in potato leaves. Plant Physiol., **78**, 76-79.

Huessing , J. E., L. L. Murdock, and R. E. Shade. (1991). Effect of wheat germ isolectins on development of cowpea weevil. Phytochem., **30**, 785-788.

Huessing, J. E., L. L. Murdock, and R. E. Shade. (1991). Rice and stinging nettle lectins, insecticidal activity similar to wheat germ agglutinin. Phytochem., **30**, 3565-3568.

Huessing, J. E., R. E. Shade, M. J. Chrispeels, and L. L. Murdock. (1991). Alpha-amylase inhibitor, not phytohaemeagglutinin, explains the resistance of common bean seeds to cowpea weevil. Plant Physiol., **96**, 993-996.

Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White. (1990). PCR protocols; A guide to methods and applications. Academic Press Inc.,

Ishimoto, M. and K Kitamura. (1988). Identification of the growth inhibitor on adzuki bean weevil in kidney bean (*Phaseolus vulgaris*). Japan J. Breed, **38**, 367-370.

Janzen, D. H., H. B. Juster, and I. E Leiner. (1976). Insecticidal action of the phytohaemeagglutinin in black beans on a bruchid beetle. Science, **192**, 795-796.

Jayne-Williams, D. J. and C. D. Burgess. (1974). Further observations on the toxicity of navy beans (*Phaseolus vulgaris*) for Japanese quail (*Coturnix coturnix japonica*). J. Appl. Bacteriol., **37**, 149-169.

Johnson, N., J. Narvaez, G. An, and C. Ryan. (1989). Expression of proteinase inhibitors I and II in transgenic tobacco plants: Effects on natural defence against *Manduca sexta* larvae. Proc. Natl. Acad. Sci. USA, **86**, 9871-9875.

Johnston, K. A., M. J. Lee, C. Brough, V. A. Hilder, A. M. R. Gatehouse, and J. A. Gatehouse. (1995). Protease activities in the larval midgut of *Heliothis virescens* : Evidence for trypsin and chymotrypsin-like enzymes. Insect Biochem. Molec. Biol., **25**, 375-383.

Jongeneel, C. V., J. Bouvier, and A. Bairoch. (1989). A unique signature identifies a family of zinc-dependant metallopeptidases. FEBS Lett., **242**, 211-214.

Kato, I and N Tominaga. (1979). Trypsin-Subtilisin inhibitor from red sea turtle eggwhite consists of two tandem domains--one Kunitz--one of a new family. Fed. Proc., **38**, 832-832.

Kimball, J. W. (1983). Introduction to immunology. Macmillan publishing Co. Inc. New York,

King, T. P., A. Pusztai, and E. M. W. Clarke. (1980). Immunocytochemical localisation of injested kidney bean (*Phaseolus vulgaris*) lectin in rat gut. Histochem. J., **12**, 201-208.

King, T. P., A. Pusztai, and E. M. W. Clarke. (1980). Kidney bean (*Phaseolus vulgaris*) lectin-induced lesions in rat small intestine. 3. Ultrastructural studies. J. Comp. Pathol., **92**,

Kitamura, K., A. Kikuchi, and K. Harada. (1987). Performance of near-isogenic lines lacking seed lipoxygenases. Soybean Genetics Newsletter, **14**, 109-112.

Klarskov, K., P. Hojrup, S. O. Andersen, and P. Roepstorff. (1989). Plasma-desorption mass spectrometry as an aid in protein sequence determination: Application of the method on a cuticular protein from the migratory locust (*Locusta migratoria*). Biochem. J., **262**, 923-930.

Klein, J. R., U. Klein, M. Schad, and R. Plapp. (1993). Cloning, DNA sequence analysis and partial characterisation of pepN, a lysyl aminopeptidase from Lactobacillus Delbrückii ssp. lactis DSM7290. Eur. J. Biochem., **217**, 105-114.

Knight, P. J. K., N. Crickmore, and D. J. Ellar. (1994). The receptor for *Bacillus thuringiensis* CryIA(c) delta-endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase N. Molecular Microbiology, **11**, 429-436.

Knight, P. J. K., B. H. Knowles, and D. J. Ellar. (1995). Molecular cloning of an insect aminopeptidase-N that serves as a receptor for Bacillus thuringiensis cryIA(c) toxin. J. Biol. Chem., **270**, 17765-17770.

Knowles, B. H. (1994). Mechanism of action of *Bacillus thuringiensis* insecticidal delta-endotoxins. Adv. Ins. Physiol., **24**, 275-308.

Krieg, V. A., A. M. Huger, G. A. Langenbruch, and W. Z. Schnetter. (1983). Angew. Ent., **96**, 500-508.

Krysan, J. L. (1986). Biology, distribution and identification. In Methods for the study of pest *Diabrotica* . (Ed. J. L. Krysan and T. A. Miller), Springer-Verlag, New York.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, **227**, 680-685.

Laskowski, M, Jr and R. W. Sealock, Ed. Protein proteinase inhibitors - molecular aspects. Vol. 3. The Enzymes. 1971.

Laskowski, M. Jr. and I. Kato. (1980). Protein inhibitors of proteases. Ann. Rev.

Biochem., **49**, 593-626.

Leary, T. R., D. T. Grahn, H. Neurath, and G. M. Hass. (1979). Structure of potato carboxypeptidase inhibitor: Disulfide pairing and exposure of aromatic residues. Biochemistry, **18**, 2252-2256.

Lee, M. J. and J. H. Anstee. (1995). Characterisation of midgut exopeptidase activity from larval *Spodoptera littoralis* . Insect Biochem. Molec. Biol., **25**, 63-71.

Leiner, I. E. (1980). Toxic constituents of plant foodstuffs. Academic press, New York.

Look, A. T., R. A. Ashmun, L. H. Shapiro, and S. C. Peiper. (1988). Human myeloid plasma membrane glycoprotein CD13 (gp150) is identical to aminopeptidase N. J. Clin. Invest., **83**, 1299-1307.

Lorand, L, Ed. Methods in enzymology. Vol. 45. Methods in enzymology. 1976.

Lotan, R., R. Cacan, M. Cacan, H. Debray, W. C. Carter, and N. Sharon. (1975). On the presence of two types of subunit in soybean agglutinin. FEBS Lett., **57**, 100.

Lütcke, H.A., K.C. Chow, F.S. Mickel, K.A. Moss, H.F. Kern, and G.A. Scheele. (1987). Selection of AUG initiation codon differs in plants and animals. EMBO, **6**, 43-48.

Lüthy, P. and H. R. Ebersold. (1981). *Bacillus thuringiensis* delta-endotoxins: histopathology and molecular mode of action. In Pathogenesis of invertebrate microbial diseases. (Ed. E. W. Davidson), Allanheld, Osman , New Jersey.

Marmur, J. and P. Doty. (1962). Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol., **5**, 109.

Maroux, S. and H. Feracci. (1983). Structure and biosynthesis of aminopeptidases. Meth. Enzymol., **96**, 406-423.

Martineau, B., K. E. McBride, and C. M. Houck. (1991). Regulation of metallocarboxypeptidase inhibitor gene expression in tomato. Mol. Gen. Genet., **228**, 281-286.

McConaughty, B. L., C. D. Laird, and B. J. McCarthy. (1969). Nucleic acid reassociation in formamide. Biochem., **8**, 3289.

McDonnell, M. W., M. N. Simon, and F. W Studier. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol., **110**, 119.

McGaughey, W. H. and M. E Whalon. (1992). Managing insect resistance to *Bacillus thuringiensis* toxins. Science, **258**, 1451-1455.

McIlvaine. (1921). J. Biological Chemistry, **49**, 183.

Meiners, J. P. and T. C. Elden. "Resistance to insects and diseases in *Phaseolus* . In: Advances in Legume science." In International legume conference in Kew, edited by R. S. Summerfield and A. H. Buting, , 359-364, Year.

Merryweather, A. *Manduca* protease inhibitor constructs. 1994.

Metcalf, R. L. (1986). The ecology of insecticides and the chemical control of insects. In Ecological theory and integrated pest management. (Ed. M. Kogan), John Wiley and Sons, New York.

Milstein, C. (1986). From antibody structure to immunological diversification of immune response. Science, **25??**, 1261-1268.

Mogensen, I. B., H. Schulenberg, H. O. Hansen, F. Spener, and J Knudsen. (1987). A novel acyl-CoA-binding protein from bovine liver. Effect on fatty acid synthesis. Biochem. J., **241**, 189-192.

Munn, E. A. Recombinant DNA molecules encoding aminopeptidase enzymes and their use in the preparation of vaccines against helminth infections. Ed. M. Graham, T. S. Smith, E. A. Munn, D. Knox, J. Oliver, and S. E. Newton. British Patent Application No. 9209993.6: 1992.

Murdock, L. L., J. E. Huessing, S. S. Nielsen, R. C. Pratt, and R. E. Shade. (1990). Biological effects of plant lectins on cowpea weevil. Phytochem., **29**, 85-89.

Nanus, D. M., D. Engelstein, D. A. Gastl, L. Gluck, M. J. Vidal, Morrison. M., C. L. Finstad, N. H. Bander, and A. P. Albino. (1993). Molecular cloning of the human kidney differentiation antigen gp160: Human aminopeptidase A. Proc. Natl. Acad. Sci. USA, **90**, 7069-7073.

Newell, C. A. *Manduca* protease inhibitor tobacco. 1994.

Newell, C. A., J. M. Lowe, A. Merryweather, L. M. Rooke, and W. D. O. Hamilton. (1995). Transformation of sweet potato (*Ipomoea batatas* (L.) Lam.) with *Agrobacterium tumefaciens* and regeneration of plants expressing cowpea trypsin inhibitor and snowdrop lectin. Plant Science, **107**, 215-227.

Oerke, E. C. (1994). Estimated losses due to pathogens animal pests and weeds. In Estimated losses in food and cash crops. (Ed. E. C Oerke, H. W. Dehne, F. Schonbeck, and A Weber), pp.830. Elsevier, Amsterdam.

Ohta, S., K. Goto, H. Arai, and Y. Kagawa. (1987). An extremely acidic amino-terminal presequence of the precursor for the human mitrochondrial hinge protein. FEBS Lett., **226**, 171-175.

Oslen, J., G. M. Cowell, E. Kønigshøfer, E. M. Danielsen, J. Møller, L. Laustsen, O. C. Hansen, K. G. Welinder, J. Endberg, W. Hunziker, M. Spiess, H. Sjöström, and O. Norén. (1988). Complete amino acid sequence of human intestinal aminopeptidase N as deduced from cloned cDNA. FEBS Letters, **238**, 307-314.

Papayannopoulos, I. A. and K. Biemann. (1992). Amino acid sequence of a protease inhibitor isolated from *Sarcophaga bullata* determined by mass spectrometry. Protein Sci., **1**, 278-288.

Pearce, G. and C. A. Ryan. (1983). A rapid, large-scale method for the purification of the metallo-carboxypeptidase inhibitor from potato tubers. Anal. Biochem., **130**, 223-225.

Perlak, F. J., R. L. Fuchs, D. A. Dean, S. L. McPherson, and D. A. Fischhoff. (1992).

Modification of the coding sequence enhances plant expression of insect control protein genes. Proc. Natl. Acad. Sci. USA, **88**, 3324-3328.

Peterson, L. M., B. Holmquist, and J. L. Bethune. (1982). A unique activity assay for carboxypeptidase A in human serum. Anal. Biochem., **125**, 420-426.

Peterson, V. (1984). The natural food catalogue. McDonald, London.

Powell, K. S. "Antimetabolic effects of plant proteins on homopteran insect pests." PhD, Durham, 1993.

Powell, K. S., A. M. R. Gatehouse, V. A. Hilder, and J. A. Gatehouse. (1993). Antimetabolic effects of plant lectins and plant and fungal enzymes on the nymphal stages of two important rice pests, *Nilaparvata lugens* and *Nephotettix cinciteps* . ppl., **66**, 119-126.

Powell, K. S., A. M. R. Gatehouse, V. A. Hilder, and J. A. Gatehouse. (1995). Antifeedant effects of plant lectins and an enzyme on the adult stage of the rice brown plant hopper, *Nilaparvata lugens* . Ent. Exp. Appl., **75**, 51-59.

Pusztai, A., E. M. W. Clarke, and T. P. King. (1979). The nutritional toxicity of *Phaseolus vulgaris* . Nutr. Soc., **38**, 115-120.

Rafti, F., G. Garguilo, A. Manzi, C. Malva, and F Graziani. (1989). Sequence of the ribosomal protein cDNA of *D. melanogaster* homologous to the L1 ribosomal protein gene of *X. laevis* . Nucleic Acids Res., **17**, 456-456.

Rafti, F., G. Garguilo, A. Manzi, C. Malva, G. Grossi, S. Andone, and F. Graziani. (1988). Isolation and structural analysis of a ribosomal protein gene from *D. melanogaster* . Nucleic Acids Res., **16**, 4915-4926.

Ramesh, N, M. Sugumaran, and J. E. Mole. (1988). Purification and characterisation of two trypsin inhibitors from the hemolymph of *Manduca sexta* larvae. J. Biol. Chem., **263**, 11523-11527.

Rancour, J. M. and C. A. Ryan. (1968). Isolation of a carboxypeptidase-B inhibitor from potatoes. Arch. Biochem. Biophys., **125**, 380.

Rees, D. C. and W. N. Lipscomb. (1982). Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. J. Mol. Biol., **160**, 475-498.

Richardson, M. (1977). The proteinase inhibitors of plants and micro-organisms. Phytochem., **16**, 159-169.

Richardson, M. (1991). Seed storage proteins: the enzyme inhibitors. In Methods of plant biochemistry; amino acids proteins and nucleic acids. (Ed. L. J. Rogers, P. M. Dey, and J. B. Harbourne), Academic press, London.

Ritonja, A., I Krizaj, P Mesko, M. Kopitar, M. Lukovnik, B. Strukelj, J. Pungercar, D. J. Buttle, A. J. Barrett, and V Turk. (1990). The amino acid sequence of a novel inhibitor of cathepsin-D from potato. FEBS Letters, **267**, 13-15.

Roth, M.B., A.M. Zahler, and J.A. Stolk. (1991). A conserved family of nuclear phosphoproteins localised to sites of polymerase-II transcription. J. Cell Biol., **115**, 587-

596.

Ryan, C. A. (1981). Proteinase inhibitors. In The biochemistry of plants (Ed. A. Marcus), pp.351-370. Academic press, New York.

Ryan, C. A. (1990). Proteinase inhibitors in plants - Genes for improving defences against insects and pathogens. Ann. Rev. Phytopathol., **28**, 425-449.

Ryan, J. W., A. Y. K. Chung, J. A. Nearing, F. A. Valido, C. Shun-Cun, and P. Berryer. (1993). A radiochemical assay for aminopeptidase N. Anal. Biochem., **210**, 27-33.

Saiki, R. K., D. H. Gelfand, S Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science, **239**, 487.

Sajid, M. and R. E. Issac. (1993). Neuropeptide metabolism in nematodes: Properties of a membrane-bound aminopeptidase and endopeptidase. J. Cellular Biochem., **S17C**, 119.

Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989). Molecular Cloning. A Laboratory Manual. Cold Spring Harbour Laboratory Press, New York.

Samuels, R. I., A. K. Charnley, and S. E. Reynolds. (1993). An aminopeptidase from the moulting fluid of the tobacco hornworm, *Manduca sexta*. Insect Biochem. Molec. Biol., **23**, 615-620.

Samuels, R. I. and S. E. Reynolds. (1993). Moulting fluid enzymes of the tobacco hornworm, *Manduca sexta* : Timing of proteolytic and chitinolytic activity in relation to pre-ecdysial development. Arch. Insect Biochem. Physiol., **24**, 33-44.

Sasaki, T. (1978). Chymotrypsin inhibitors from hemolymph of the silkworm, *Bombyx mori*. J. Biochem, **84**, 267-276.

Sasaki, T. and K. Kobayashi. (1984). Isolation of two novel proteinase inhibitors from hemolymph of silkworm larva, Bombyx mori. Comparison with human serum protease inhibitors. J. Biochem., **95**, 1009-1017.

Sauvion, N. "Mode d'action de deux lectines à mannose sur le puceron du pois, *Acyrthosiphon pisum* (Harris). Potentiels d'utilisation des lectines végétales dans une stratégie de création de plantes transgéniques résistantes aux pucerons." Docteur en Sciences, Lyon, 1995.

Schmidt, T., E. Stumm-Zollinger, P.S. Chen, P. Boehlen, and S.R. Stone. (1989). A male accessory gland peptide with protease inhibitory activity in *Drosophila funebris* . J. Biol. Chem., **264**, 9745-9749.

Shelton, A. M. and J. A. Wyman. "Management of diamondback moth and other crucifera pests." In Proceedings of the second international workshop in Asian vegetable research and development centre, edited by N. S. Talekar, Year.

Shukle, R. H. and L. L. Murdock. (1983). Lipoxygenase, trypsin inhibitor and lectin from soybeans: effects on larval growth of *Manduca sexta* (Lepidoptera, Sphingidae). Environ, Entomol., **12**, 787.

Snyder, m. j. and R. Feyereisen. Submission to EMBL/GENBANK/DDBJ data banks. 1993.

Southern, . (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol., **98**, 503.

Stroman, P. (1992). Sequence of a gene (lap) encoding a 95.3-kDa aminopeptidase from *Lactococcus lactis* ssp. cremoris Wg2. Gene, **113**, 107-112.

Takesue, S., K. Yokota, S. Miyajima, R Taguchi, H. Ikezawa, and Y. Takesue. (1992). Partial release of aminopeptidase N from larval midgut cell membranes of the silkworm, *Bombyx mori*, by phosphatidylinositol-specific phospholipase C. Comp. Biochem. Physiol., **102B**, 7-11.

Tan, P. S., I. J. Van Alen-Boerrigter, B. Poolman, R. J. Siezen, W. M. De Vos, and W. N. Konings. (1992). Characterisation of the *Lactococcus lactis* pepN gene encoding an aminopeptidase homologous to mammalian aminopeptidase N. FEBS Lett., **306**, 9-16.

Taylor, A. (1993). Aminopeptidases: structure and function. FASEB J., **7**, 290-298.

Taylor, A. (1993). Aminopeptidases: towards a mechanism of action. TIBS, **18**, 167-172.

Vadlamudi, R. K., T. H. Ji, and L. A. Jr. Bulla. (1993). A specific binding protein for the insecticidal toxin of *Bacillus thuringiensis* subsp. *berliner* . J. Biol. Chem., **268**, 12334-12340.

Vadlamudi, R. K., E. Weber, I Ji, T. H. Ji, and L. A. Jr. Bulla. (1995). Cloning and expression of a receptor for an insecticidal toxin of *Bacillus thuringiensis* . J, Biol. Chem., **270**, 5490-5494.

Vaeck, M., A. Reynaerts, H. Höfte, S. Jansens, M. De Beuckeleer, C. Dean, M. Zabeau, M. Van Montagu, and J. Leemans. (1987). Transgenic plants protected from insect attack. Nature, **328**, 33-37.

Valaitis, A. (1995). Gypsy moth midgut proteinases: Purification and characterisation of luminal trypsin, elastase and the brush border membrane leucine aminopeptidase. Insect Biochem. Molec. Biol., **25**, 139-149.

Vallee, B. L. and D.S. Auld. (1990). Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry, **29**, 5647-5659.

Van der Logt, C. P. E., P. H. Reitsma, and R. M. Bertina. (1991). Intron-exon organisation of the human gene coding for the lipoprotein -associated coagulation inhibitor: The factor Xa dependent inhibitor of the extrinsic pathway of coagulation. Biochemistry, **30**, 1571-1577.

Van Loon, A. P. G. M., R. J. De Groot, M. De Haan, A. Dekker, and L. A. Grivell. (1984). The DNA sequence of the nuclear gene coding for the 17-kDa subunit of the yeast ubiquinol-cytochrome c reductase: a protein with an extremely high content of acidic amino acids. EMBO J., **3**, 1039-1043.

Van Rie, J., S. Jansens, H. Höfte, D. Degheele, and H. Van Mellaert. (1989). Specificity of *Bacillus thuringiensis* delta-endotoxins: importance of specific receptors on the brush