Gene sequences encoding ribosome-inactivating proteins from soapwort (Saponaria officinalis L.)

Fordham-Skelton, Anthony P.

Thesis title.

Gene sequences encoding ribosome-inactivating proteins from soapwort (*Saponaria officinalis* L.).

A P Fordham-Skelton, PhD thesis, Department of Biological Sciences, University of Durham, February 1991.

For Anne.

GENE SEQUENCES ENCODING RIBOSOME-INACTIVATING PROTEINS FROM SOAPWORT (*SAPONARIA OFFICINALIS* L.).

A thesis submitted by Anthony Paul Fordham-Skelton BSc (Dunelm) in accordance with the requirements for the degree of Doctor of Philosophy in the University of Durham.

DEPARTMENT OF BIOLOGICAL SCIENCES. FEBRUARY 1991.

ABSTRACT.

GENE SEQUENCES ENCODING RIBOSOME-INACTIVATING PROTEINS FROM SOAPWORT (*SAPONARIA OFFICINALIS* L.).

Anthony P. Fordham-Skelton

Ribosome-inactivating proteins (RIPs) are found in a wide variety of plant species. They possess an RNA N-glycosidase activity whereby the removal of a specific adenine residue from 28 S RNA renders a eukaryotic ribosome inactive. Type II RIPS contain both an active polypeptide and a sugar-binding polypeptide. Type I RIPs are composed of a single polypeptide functionally homologous to the active type II polypeptide.

This thesis describes studies of the gene sequences of RIPs representative of each class: Ricin, a type II RIP from the castor oil plant (*Ricinus communis* L.), and saporin, a type I RIP from soapwort (*Saponaria officinalis* L.). Two ricin gene sequences were isolated from a *Ricinus* genomic library and partially characterised. One gene was a badly damaged ricin-like pseudogene whilst the other was shown to encode an active polypeptide. A second ricin sequence encoding an active polypeptide was isolated using Polymerase Chain Reaction (PCR) DNA amplification. The specificity of PCR amplification was investigated using the ricin and related agglutinin gene sequences.

Partial amino acid sequence data derived from protein sequencing of saporin-6 was used to synthesise degenerate inosine-containing oligonucleotides. These directed the PCR amplification of part of the saporin coding sequence from genomic DNA. The product was used as a saporin-specific hybridisation probe. Southern analysis of *Saponaria* genomic DNA indicated that saporin sequences comprised a small multigene family. Three independent saporin containing genomic clones were isolated from a *Saponaria* genomic library. Two clones were truncated whilst the third contained a complete saporin coding sequence.

The saporin and ricin coding sequences were expressed *in vitro* and shown to inhibit protein synthesis. Aniline cleavage assays of ribosomal RNA extracted from ribosomes exposed to the products of the RIP coding sequences were carried out. These indicated that the polypeptides encoded by the RIP gene sequences had specific RNA N-glycosidase activity.

MEMORANDUM.

Part of the work presented in this thesis has been presented in the following publication (see appendix).

Fordham-Skelton AP, Yarwood A, Croy RRD (1990) Synthesis of saporin gene probes from partial protein sequence data: Use of inosine-oligonucleotides, genomic DNA and the polymerase chain reaction. Molecular and General Genetics 221:134-138.

STATEMENT.

No part of this thesis has been previously submitted for a degree in this or any other university. I declare that, unless otherwise indicated, the work presented herein is entirely my own.

CONTENTS.

CONTENTS.

LIST OF FIGURES.

ABBREVIATIONS.

A	adenine
ATP	adenosine-5'-triphosphate
agglutinin	*Ricinus communis* agglutinin
bp	base pair
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
cpm	counts per minute
d.a.f.	days after flowering
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytidine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
dTTP	deoxythymidine-5'-triphosphate
DTT	dithiothreitol
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
E. coli	*Escherichia coli*
EDTA	ethylene diamine tetraacetic acid
EF	elongation factor
G	guanine
I	inosine
IPTG	isopropyl-2-D-thiogalactoside
kb	kilobase pair
PEG	polyethylene glycol
λ	phage lambda
min	minutes
mRNA	messenger RNA
OD	optical density
PAP	pokeweed antiviral protein
PCR	polymerase chain reaction
pfu	plaque forming units
RCA	*Ricinus communis* agglutinin
RNA	ribonucleic acid
RIP(s)	ribosome-inactivating protein(s)
RNase	ribonuclease
rRNA	ribosomal RNA

S	Svedberg
SDS	sodium dodecyl sulphate
SSC	saline sodium citrate
T	thymine
Taq	*Thermus aquaticus*
TCA	trichloroacetic acid
Tris-HCl	Tris(hydroxymethyl)methylamine hydrochloride
tRNA	transfer RNA
UTP	uridine triphosphate
UV	ultraviolet
v/v	volume for volume
w/v	weight for volume
X-gal	5-bromo-4-chloro-3-indolyl-2-D-galactoside

ACKNOWLEDGEMENTS.

I would like to thank Dr. Ron Croy for the supervision of this work and for his continued support. I would also like to thank the following people for helpful advice and many interesting and useful discussions: Dr. John Gatehouse, Dr. Marta Evans, Dr. Vaughan Hilder, Dr. Glynn Edwards, Dr. Andy Thompson, Dr. Andy Ryan, Dr. Charlie Scutt, Dr. Phil Taylor, Dr. Jenny Yarwood, Dr. Alan Yarwood, Dr Anil Shirsat, Dr Nigel Robinson, Mr. David Howe, Mr. David Bown, Mr. Neville Wilford, Mr Billy Lyndsay and Mr. Lawrence Gatehouse. Thanks are also due to Dr. Martin Hartley (Department of Biological Sciences, Warwick University) for his considerable assistance and advice during the latter stages of this work.

I would also like to thank the following people for technical assistance: Mr. John Gilroy (oligonucleotide synthesis), Miss Julia Bryden (nucleotide sequencing reactions), Mr Russell Swinhoe (RNA electrophoresis and *in vitro* translation reactions) and Mr. Paul Simpson (plasmid DNA preparations). Thanks are also due to Mr. Paul Sydney and Mr. David Hutchinson for photographic work. This work was supported by the receipt of a Quota award from the Science and Engineering Research Council.

The members of Durham City Cricket Club 2nd XI and St Cuthberts Mountaineering Club also deserve a mention for providing suitable distractions. Finally, I would like to thank all the people who have helped and encouraged me throughout my years of study.

INTRODUCTION.

1. General introduction.

The toxic plant proteins ricin and abrin were discovered in the late 19th century by examining the effects of seed extracts on blood. These were prepared from castor bean (*Ricinus communis* L.) and rosary pea (*Abrus precatorius* L.) respectively (Stillmark, 1889; Hellin, 1891; cited in Lord, 1985a). Since then it has become evident that many angiosperm plant species contain proteins similar to ricin and abrin. Their toxicity is due to their ability to inhibit protein synthesis in eukaryotes by enzymatically inactivating the 60 S ribosomal subunit. Consequently, some of these proteins are amongst the most toxic substances known; it has been suggested that for ricin its enzymatic inactivation of ribosomes is so great that the presence of a single molecule within a cell is sufficient to kill it. These plant toxins have been termed ribosome-inactivating proteins (RIPs) (reviewed by Barbieri and Stirpe, 1982; Olsnes and Pihl, 1982; Jiménez and Vázquez, 1985; Stirpe and Barbieri, 1986). RIPs have received much attention as potential immunoconjugated cytotoxic moieties for use in therapies to eliminate cancer cells (Pastan *et al.*, 1986; Lord, 1987; Vitetta *et al.*, 1987; Blakey *et al.*, 1988). However, little is known about the synthesis and role of these proteins within their plant sources. Only in the case of the best-characterised RIP, ricin, from the castor oil plant (*Ricinus communis* L.), had any information on the molecular biology of the encoding genes been obtained at the start of this project. The work described in this thesis was therefore aimed to characterise genes encoding RIPs from *Ricinus* and *Saponaria officinalis* L. (soapwort).

2. Ribosome-inactivating proteins: classification and distribution.

RIPs have been classified into one of two groups dependent upon their subunit composition (Barbieri and Stirpe, 1982). Type I RIPs possess a single active polypeptide. Type II RIPs are comprised of two disulphide bonded subunits; an active polypeptide A-chain, functionally analogous to those of the type I RIPs, and a sugar binding, lectin-like, B-chain. The presence of the B-chain is responsible for the greater toxicity of type II RIPs to intact cells and whole organisms (animals), and for other differences in functional properties.

The general properties of RIPs are summarised in the following section and the reader is directed to the reviews cited above for a more detailed survey. Much new information on the mechanism of ribosome-inactivation has become available since the publication of the reviews cited, and this is surveyed in the following sections.

2.1. Type II ribosome-inactivating proteins.

Type II RIPs include ricin from *Ricinus communis* L. (*Euphorbiaceae*), abrin from *Abrus precatorius* (*Leguminosae*), volkensin from the roots of *Adenia volkensii* (*Passifloraceae*), viscumin from the leaves of *Viscum album* (mistletoe) (*Loranthaceae*) and modeccin from the roots and fruit of *Adenia digitata* (*Passifloraceae*). These proteins are extremely cytotoxic. The ID_{50} (concentration causing 50% inhibition) of ricin for the inhibition of protein synthesis in HeLa cells is only 0.0011 nM. The other type II RIPs have similar values (Stirpe and Barbieri, 1986). The LD_{50} dose of ricin or abrin intravenously injected into mice is 2.7 μg/kg and 0.7 μg/kg respectively (Olsnes and Pihl, 1982). The extreme cytotoxicity of type II RIPs is due to the B-chains binding to specific carbohydrate residues present on the surface of cells followed by uptake into the cell and ribosome inactivation (see section 2.1.4).

2.1.1. The structure of type II ribosome-inactivating proteins.

Type II RIPs are heterodimeric glycoproteins comprised of two subunits linked by a disulphide bond and have a relative molecular mass of between 60,000 and 65,000. The relative molecular mass of the ricin A- and B-chains is 30,625 and 31,432, respectively. The subunits composing the other type II RIPs have similar molecular weights (Jiménez and Vázquez, 1985).

2.1.2. Ricin, abrin and their related agglutinins.

Ricinus and *Abrus* seeds contain lectins which have haemagglutinating activity and are serologically and structurally related to the corresponding RIPs, ricin and abrin (Pappenheimer *et al.*, 1974). These lectins share the same heterodimeric structure as ricin and abrin but non-covalent interactions between homologous subunits of the heterodimers results in dimerisation to form a tetramer composed of two heterodimers. The RIP has the structure (A/B) whilst the homologous lectin is $(A'/B')_2$

(Lord, 1985a). *Ricinus* agglutinin is not as cytotoxic as ricin *in vivo*, yet the A-chain is capable of inhibiting *in vitro* protein synthesis (Harley and Beevers, 1982). The B-chains of ricin, abrin and the agglutinins all exhibit lectin activity and bind galactose residues. The B-chain of ricin has two sugar binding sites specific for either galactose or N-acetylglucosamine (Houston and Dooley, 1982; Nicholson and Blaustein, 1972). The erythrocyte haemagglutinating activity of the lectins is due to their dimeric bivalent structure whilst the corresponding monovalent RIPs exhibit only weak cell haemagglutination (Lord, 1985a). The A- and B-chains of all the proteins are N-glycosylated with the exception of the abrin A-chain which is unglycosylated (Jiménez and Vázquez, 1985).

2.1.3. Other type II ribosome-inactivating proteins.

Volkensin and viscumin have the same galactose binding specificities exhibited by ricin and abrin. However, viscumin is toxic to HeLa cell line mutants which are insensitive to ricin intoxication. This indicates that the interaction of viscumin with the cell surface differs from that of ricin. In a similar manner modeccin also binds galactose but is toxic to cell lines insensitive to either ricin or abrin, again indicating a difference in the specificity of binding (Jiménez and Vázquez, 1985). Modeccin shows no immunological cross-reactivity with either ricin or abrin using anti-modeccin antisera (Olsnes *et al.*, 1978).

2.1.4. Type II ribosome-inactivating protein cytotoxicity.

Type II RIP toxicity to intact cells is due to the presence of the sugar binding B-chain. In the case of ricin the B-chain, whilst exhibiting specificity for galactose residues, binds with greater affinity to complex galactosides present on cell surfaces than to simple sugars (Baenziger and Fiete, 1979). After binding to cell surface galactosyl residues present on glycolipids and glycoproteins some of the molecules are endocytosed in an energy dependent manner; the inhibition of ATP production preventing internalisation (Olsnes and Pihl, 1982). Once inside the cell the disulphide bond between the A-and B-chains is cleaved and the A-chain is translocated from its intra-cellular compartment, possibly the trans Golgi cisternae (van Deurs *et al.*, 1986), into the cytoplasm causing subsequent ribosome inactivation. The A-chain of ricin is not active until it is released from the B-chain (Olsnes and Pihl, 1982; Richardson *et al.*,

1989). Dissociation of A- and B-chains may result in a conformational change in the A-chain thereby activating it. Release of the A-chain exposes a non-polar region on the A-chain, thereby aiding membrane binding and release into the cytoplasm (Houston, 1982). It has also been suggested that the B-chain in some manner facilitates the release of the A-chain into the cytoplasm (Vitetta, 1986; Simmons *et al.*, 1986). It is presumed that the other type II RIPs act in an analogous manner.

2.1.5. The primary sequences and nucleotide sequences of ricin and *Ricinus* agglutinin.

Several isoforms of both ricin and abrin can be isolated from seeds of differing origin and even from within seeds from the same source (Olsnes and Pihl, 1982). At least three forms of ricin and two of *Ricinus* agglutinin have been demonstrated, based on a comparison of apparent relative molecular weights and isoelectric points of both the intact proteins and subunits (Cawley *et al.*, 1978). The primary sequences of the A and B subunits of one form of ricin (Ricin D) have been determined (Funatsu *et al.*, 1978; Araki and Funatsu, 1985). The A-chain is a polypeptide of 265 amino acids whilst the B-chain contains 260 amino acids. A closely related variant form of ricin (Ricin E) has also been isolated (Mise *et al.*, 1977) from the seeds of small-grain *Ricinus* seeds. The B-chain primary sequence of this ricin isoform has also been determined (Araki and Funatsu, 1987).

cDNA clones for both ricin and *Ricinus* agglutinin have been isolated characterised and sequenced (Lamb *et al.*, 1985; Roberts *et al.*, 1985). The ricin cDNA sequence was deduced from two overlapping clones. The deduced amino acid sequence of the clones revealed that both ricin A- and B-chains are synthesised as a contiguous preproprotein. There is an N-terminal region of 24 amino acids which acts as a leader sequence to translocate preproricin into the endoplasmic reticulum. Between the A- and B-chain sequence there is a linking peptide of 12 amino acids. This is removed from proricin post-translationally by endopeptidases to produce the mature protein. The final destination for ricin is the matrix of the endosperm protein bodies of the seed (Lord, 1985a). The ricin cDNA sequence was confirmed independently by Halling *et al.* (1985) who isolated and sequenced a ricin-containing genomic clone, the genomic sequence did not contain introns. The coding sequence of this gene had 10 nucleotide differences when compared with the cDNA sequence. The deduced amino acid

sequence of the cDNA also had a small number of differences when compared with the protein sequence. A ricin cDNA clone isolated by Piatak *et al.* (1988) showed sequence differences in the 5′ non-translated region when compared with the ricin genomic clone of Halling *et al* (1985) indicating that the mRNA was most likely derived from a different gene. These differences are probably due to the presence of distinct genes expressing ricin isoforms, as noted above. Halling *et al.* (1985) also showed by Southern analysis of *Ricinus* genomic DNA that ricin was a member of a small multigene family.

The sequence of *Ricinus* agglutinin (Lamb *et al.*, 1985), derived from overlapping clones, revealed that the agglutinin gene shared the same coding sequence architecture as the previously reported ricin sequence. As shown for ricin, the agglutinin is also synthesised as a preproprotein containing a leader sequence and linker peptide between the A- and B-chains. The agglutinin sequence shows a high degree of similarity to the ricin sequence, the A-chains of ricin and agglutinin show 93% similarity, whilst the B-chains show 84% similarity. Thus the agglutinin gene is a member of the ricin gene family. Evidence from the restriction mapping of several agglutinin cDNA clones indicated a degree of restriction site heterogeneity between the clones. This would suggest that several slightly different agglutinin genes are expressed in the *Ricinus* endosperm.

Ladin *et al.* (1987) presented the cDNA sequence of a hybrid ricin/agglutinin coding sequence. The coding sequence was ricin-like until the last half of the B-chain at which point it became agglutinin-like. The deduced amino acid sequence was very like the primary sequence determined for the B-chain of ricin E (Araki and Funatsu, 1987). The B-chain sequence is ricin-like and then changes to agglutinin-like. These two sequences suggest that the gene for ricin E, or a very similar gene, has arisen by a recombination event between a ricin and an agglutinin gene. Ricin E is present in the seeds of small-grain castor beans along with ricin D and *Ricinus* agglutinin. Large-grain castor beans only contain ricin D and *Ricinus* agglutinin. The small-grain *Ricinus* plant is grown in temperate zones whereas the large-grain *Ricinus* plant is grown in tropical zones. *Ricinus* originates from tropical Africa and it is proposed that the recombination event

producing ricin E has occurred after the adaptation of *Ricinus* to temperate zones (Araki and Funatsu, 1987).

2.1.6. The three dimensional structure of ricin.

Montfort *et al.* (1987) have determined the three dimensional X-ray crystallographic structure of ricin at a resolution of 2.8 A. The A-chain exhibits extensive secondary structure, about 15% β structure and 30% helical structure. Three distinct domains are present. A prominent cleft is formed by the interaction of the three domains. This was proposed as the putative active site. The B-chain consists of two domains each with two disulphide bonds and a lactose binding site The domain formed by the carboxyl terminal end of the A-chain is situated between the two B-chain domains. The disulphide bridge formed between the Cys residue 259 of the A-chain and the Cys residue 4 of the B-chain is present away from this main interaction of the two subunits. As mentioned previously, the A-chain is assumed to undergo a conformational change after release from the B-chain. Recombinant ricin A-chain, expressed in *E. coli*, has been crystallised (Robertus *et al.*, 1987) to compare the structure with the native ricin molecule. No details of the recombinant ricin A-chain structure were given.

2.1.7. *In vivo* expression of recombinant ricin coding sequences.

Recombinant ricin B-chain has been expressed in *Xenopus* oocytes and monkey kidney cos-M6 cells (Richardson *et al.*, 1988; Chang *et al.*, 1987). In both cases the recombinant peptide was correctly processed, glycosylated and retained lectin activity. The B-chain was also able to reconstitute the cytotoxicity of native ricin A-chain, showing that the recombinant B-chain was able to promote A-chain translocation.

Recombinant ricin A-chain has been expressed in *E. coli*. O'Hare *et al.* (1987) produced functional ricin A-chain capable of inhibiting rabbit reticulocyte lysates to the same degree as the native A-chain. Reconstitution with native B-chain resulted in a protein toxic to vero (monkey kidney) cells. It was noted that the protein expressed at 37^0C aggregated and was less active than the A-chain expressed at 30^0C. Piatak *et al.* (1988) reported an essentially similar result. A-chain expressed at 42^0C aggregated and was less active than that expressed at 37^0C. The active form was convertible to a less active form *in vivo* but not *in vitro*

suggesting interaction with other *E. coli* proteins affects the solubility of the recombinant ricin A-chain.

2.2. Type I ribosome-inactivating proteins.

Type I RIPS are single-chain polypeptides, with relative molecular masses of between 25,000 and 30,000, they have high alkaline pI values and are mainly glycoproteins. They have been isolated from a wide variety of plant sources (Jiménez and Vázquez, 1985). The toxicity of these proteins to intact cells is, in general, much less than the corresponding type II RIPs. LD_{50} values for the toxicity of type I RIPs in mice are three orders of magnitude greater than type II RIPs (Stirpe and Barbieri, 1986). The properties of some of the better characterised type I RIPs are discussed below.

2.2.1. Trichosanthins.

Type I RIPs have been isolated from both the root tubers and seeds of *Trichosanthes kirilowii* Maximowicz (*Cucurbitaceae*). Extracts of the root tubers have long been used in Chinese medicine to induce abortions, treat choriocarcinomas and ectopic pregnancies (Chow *et al.*, 1990). The RIPs isolated from root tubers are called trichosanthins. To date, three different forms of this RIP have been isolated from root tubers (Zhang and Wang, 1986; Maraganore *et al.*, 1987; Chow *et al.*, 1990). The complete amino acid sequence has been determined for two of the trichosanthins (Zhang and Wang, 1986; Chow *et al.*, 1990) and the first[38] N-terminal amino acids of the other form isolated by Maraganore *et al.* (1987). Sequence differences between the three forms indicate that they are the products of distinct genes. The trichosanthin isolated by Chow *et al.* (1990) has been termed α-TCS to distinguish it from the other forms of trichosanthin. Casellas *et al.* (1988) have isolated trichokirin, another type I RIP from the seeds of *Trichosanthes*. A trichosanthin genomic clone has been isolated, sequenced and shown to contain the α-TCS coding sequence (Collins *et al.*, 1990).

Whilst type I RIPs generally exhibit low toxicity to intact cells trichosanthin is of interest that it is cytotoxic to specific cell types. Trichosanthins are toxic to foetal trophoblast cells from which the uterine cancer choriocarcinoma is derived (Maraganore *et al.*, 1987). McGrath *et al.* (1989) have shown that trichosanthin inhibits human

immunodeficiency virus replication (HIV) in cell lineages of lymphoid and phagocyte origin. These properties of trichosanthin suggest that the protein is specifically internalised by some cell types by an unknown mechanism. Type I RIPs isolated from another member of the *Cucurbitaceae*, *Momordica charantia*, α- and β-momorcharin also have arbotifacient activity (Yeung *et al.*, 1988).

2.2.2. Pokeweed antiviral proteins and dodecandrin.

Proteins present in the extracts of the pokeweed (*Phytolacca americana*) and the related plant *Phytolacca esculenta* have been shown to exhibit antiviral activity in several systems; the inhibition of poliovirus replication in HeLa cells (Ussery *et al.*, 1977) (pokeweed), the inhibition of tobacco mosaic virus (TMV) infectivity (Kassanis and Kleczkowski, 1948) (*P. esculenta*). The antiviral activity in pokeweed is due to the presence of type I RIPs. An antiviral protein of relative molecular mass 27,000 was isolated from the spring leaves of pokeweed and shown to inhibit *in vitro* protein synthesis and poliovirus replication in HeLa cells (Obrig *et al.*, 1973; Irvin, 1975). This protein was termed pokeweed antiviral protein (PAP). A second form of PAP (PAP II) was isolated from the summer leaves of pokeweed with a relative molecular mass of 30,000. This protein, present in small amounts in the spring leaves, becomes the major antiviral protein isolated from the summer leaves (Irvin *et al.*, 1980; Houston *et al.*, 1983). A third different PAP form, (PAP-S), was isolated from pokeweed seeds (Barbieri *et al.*, 1982). The N-terminal sequences of PAP, PAP II and PAP-S have been determined (Houston *et al.*, 1983; Bjorn *et al.*, 1984). There is a report of the cloning of the PAP gene (Lodge *et al.*, 1990) however there was no information available concerning either the organisation of the gene or its sequence. The antiviral activity of these RIPs is discussed in section 4.1.

Irvin *et al* (1980) investigated the presence of RIPs in *Phytolacca dodecandra*. This species is in the same family as *P. americana* (*Phytolaccaceae*) but is taxonomically placed in a different sub-genus (*Pircunia*) from *P. americana* (*Euphytolacca*). The two species are also geographically separated. *P. dodecandra* is found in tropical and Southern Africa whereas *P. americana* is originally from the Americas. Two protein fractions extracted from leaf tissue were shown to inhibit protein synthesis. The first fraction contained two proteins of approximately

31,000 and 32,000 relative molecular mass but were not investigated further. The second fraction contained a protein of relative molecular mass 29,000 and was called dodecandrin. The N-terminus of the protein has been sequenced (Ready *et al.*, 1984a). The relationship of dodecandrin to the pokeweed RIPs is discussed in section 2.4.

2.2.3. Luffin.

Isoforms of a type I RIP, luffin-a and luffin-b, have been isolated from the seeds of *Luffa cylindrica* (sponge gourd). The primary sequence of luffin-a has been determined (Islam *et al.*, 1990) and shows 33% similarity with ricin A-chain. The protein is only weakly cytotoxic to murine leukaemia and sarcoma cell lines but is about an order of magnitude more active than ricin A-chain when used to inhibit translation in a rabbit reticulocyte lysate system and is the strongest inhibitor of protein synthesis of all the RIPs isolated to date (Kishida *et al.*, 1983; Stirpe and Barbieri, 1986).

2.2.4. *Mirabilis* antiviral protein.

Mirabilis antiviral protein (MAP) is a type I RIP present in the leaves and roots of *Mirabilis jalapa*. MAP is also able to inhibit viral replication in a manner analogous to that of the pokeweed antiviral proteins. The primary sequence of MAP has been determined. The protein has 250 amino acids and has 24% homology with ricin A-chain (Habuka *et al.*, 1989). A synthetic gene encoding the MAP coding sequence has been synthesised and expressed in *E. coli* (Habuka *et al.*, 1989; Habuka *et al.*, 1990). The activity of MAP is discussed in section 3.3.

2.2.5. Type I ribosome-inactivating proteins isolated from *Saponaria officinalis* L. (soapwort).

The soapwort plant is a member of the *Caryophyllaceae* family. It is a deep-rooted perennial reproducing by seeds or rootstocks, it produces pale rose-coloured flowers and grows to a height of 1 to 2 feet. When the plant is crushed in water a soap-like lather is produced which has been used for centuries as a cleaning agent. The detergent and purgative effects of extracts of this species are due to the presence of saponin glucosides. The extracts of soapwort roots and leaves have been used as a herbal remedy for liver ailments, kidney stones and skin diseases as well

as for producing a "head" on beer. Its widespread use in the past has led to its colonisation of many places away outside of Europe and its classification as a weed. It can be found in waste areas throughout North America and is a problem weed in Australia, Canada and Spain (Mitich, 1990 and references cited therein).

Type I RIPs found in the seeds and leaves of *Saponaria officinalis* L. (soapwort) have been isolated and partially characterised (Stirpe *et al.*, 1985) and were termed saporins. Six distinct protein fractions were isolated from *Saponaria* seeds which exhibited RIP activity. The apparent relative molecular masses of two of the proteins were approximately 30,000. None of the saporins were glycosylated. The numerical designation of the saporins was based on the numbers of the original fraction. As such the saporin from protein peak 6 was subsequently termed saporin-6. This particular saporin comprises 7% of the total seed protein. The pI values of saporins -6 and -9 were found to be greater than 9.5. Lappi *et al.* (1985) further characterised saporin-6. Antisera raised against saporin-6 cross-reacted with saporins 5, 8 and 9 indicating that they shared the same antigenic determinants. The first 37 N-terminal residues of mature saporin-6 were determined by protein sequencing. Comparison of this sequence with the N-terminal sequences of pokeweed antiviral proteins and dodecandrin indicated a considerable degree of homology. The saporin proteins -5 and -6 were found to comprise of two slightly different forms when separated using reverse phase chromatography (Montecucchi *et al.*, 1989). The same workers have determined the N-terminal sequence of other saporin proteins (saporins -4, 5a, 5b, 6a and 6b). All shared the same sequence over the region determined except saporin-4 which was isolated from *Saponaria* leaves.

Recently, the cDNA sequence of a saporin clone very similar to saporin-6 protein was reported by Benatti *et al.* (1989). This clone was isolated from a cDNA library prepared from *Saponaria* leaves. Southern analysis of *Saponaria* genomic DNA probed with this cDNA clone indicated that the saporins comprise a small multigene family, however, the data presented was ambiguous. The several variant forms of saporin isolated does suggest that they are products of distinct but related genes, as noted previously for the ricin gene family.

2.2.6. Other type I ribosome-inactivating proteins.

RIPs have also been purified from; the latex of *Hura crepitans* L. (sandbox tree), the seeds of *Asparagus officinalis* L. (asparagus) and *Agrostemma githago* L. (corncockle). All of these are glycoproteins (Stirpe *et al.* 1983). Gelonin is a glycosylated RIP present in the seeds of *Gelonium multiflorum*.(Stirpe *et al* 1980). The leaves of carnation *Dianthus caryophyllus* (carnation) contain two glycosylated RIPs, dianthins 30 and 32, of approximate relative molecular masses of 29,500 and 31,700 respectively (Stirpe *et al.*, 1981).

Tritin is a RIP present in wheat germ (*Triticum aestivum*) with a relative molecular mass of approximately 30,000 (Roberts and Stewart, 1979; Coleman and Roberts, 1981). Other members of the *Gramineae* family also contain RIPs. Three RIPs of approximate relative molecular weight 31,000 have been isolated from barley (*Hordeum vulgare*) (Coleman and Roberts, 1982; Asano *et al.*, 1984). The primary sequence of one of the barley RIPs; Barley Protein Synthesis Inhibitor II (BPSI II) has been determined (Asano *et al.*, 1986). Rye (*Secale cereale*) and corn (*Zea mays*) also contain RIPs. The RIP isolated from rye has a relative molecular mass similar to that of tritin but the RIP from corn is significantly smaller, approximately 23,000 (Coleman and Roberts, 1982). RIPs do not appear to be present in all members of the *Gramineae*. RIPs could not be detected in either rice (*Oryza sativa*) or millet (*Setaria italica*) (Coleman and Roberts, 1982).

2.3. Ribosome-inactivating proteins of fungal and bacterial origin.

2.3.1. α-Sarcin.

α-sarcin is a small non-glycosylated toxin of (relative molecular mass 16,987) produced by *Aspergillus giganteus* which inhibits eukaryotic synthesis by enzymatically inactivating the 60 S ribosomal subunit (Jiménez and Vázquez, 1985). The primary sequence of α-sarcin has been determined (Sacco *et al.*, 1983) and exhibits 84% similarity with the related toxins mitogillin and restrictocin produced by *Aspergillus restrictus* (Lopez-Otin *et al.*, 1984). The cDNA sequence of α-sarcin has been determined and indicates that the protein is synthesised as a precursor protein containing a leader sequence of 27 amino acids (Oka *et al.*, 1990).

α-sarcin has an RNase activity when using naked RNA as a substrate,

the RNA being cleaved at multiple sites (Endo *et al.*, 1983). Its action on intact ribosomes is extremely specific. α-sarcin acts as a specific RNase and cleaves a single phosphodiester bond between G 4325 and A 4326 in the 28 S RNA of rat ribosomal 60 S subunits (Endo and Wool, 1982). Cleavage of this bond is sufficient to cause ribosome inactivation. The cleavage site is within a highly conserved sequence which is present in the corresponding position of *E. coli* 23 S RNA with only a single nucleotide change (see Fig 1.1).

2.3.2. Shiga and Shiga-like toxins.

Shigella dysenteriae type I, the causative agent of bacillary dysentery, produces a cytotoxic protein, Shiga toxin. This toxin inhibits eukaryotic protein synthesis by enzymatically inactivating the 60 S ribosomal subunit. Some strains of *E. coli* produce toxins similar to Shiga toxin. These toxins are termed Shiga-like toxins (SLT), an alternative name is Vero toxins due to their cytotoxicity to Vero cells. The Shiga-like toxins can be divided into two immunologically distinct groups. SLT-I toxins are neutralised by anti-Shiga toxin antibodies whilst SLT-II toxins are not neutralised by these antibodies. The Shiga family of toxins is composed of an enzymatically active A subunit and multiple B subunits. The B subunits are responsible for the binding of the molecule to cell surface receptors and internalisation. Following internalisation the A chain dissociates from the B chains and proteolytically nicked to produce an active A_1 fragment which is released into the cytosol (reviewed by O'Brien and Holmes, 1987).

The amino acid sequence of Shiga toxin and SLT-I A-chains are identical except for a single amino acid change explaining their immunological cross reactivity (Calderwood *et al.*, 1987; Strockbrine *et al.*, 1988). The SLT-II toxins are more distantly related to Shiga toxin, the A-chain has 56% similarity with the corresponding subunits of Shiga toxin and SLT-I (Jackson *et al.*, 1987).

The action of α-sarcin, Shiga and Shiga-like toxins is considered in relation to the mechanism of action of other RIPs in section 3.

2.4. The evolution of ribosome-inactivating proteins.

Protein sequence comparisons have provided some information about the evolution of RIP genes and their relatedness. Robertus and Ready (1984)

have shown that the B-chain of ricin has evolved by the quadruplication of a small 40 residue domain. This domain is presumed to be an ancient galactose binding subunit which has formed the ricin B-chain by undergoing gene duplications. The two homologous globular domains have identical folding topologies but the sugar binding site is at a different site in each domain. The domain is related to a similar peptide sequence present in the slime mold (*Dictyostelium discoideum*) lectin I. This lectin contains two domains which also appear to have arisen by a process of gene duplication (Robertus and Ready, 1984; Rutenber *et al.*, 1987).

Ready *et al.* (1984b) have presented N-terminal sequence comparisons of type I and type II RIPs (dodecandrin, pokeweed anti-viral proteins, ricin and modeccin A-chains). Pokeweed anti-viral protein (PAP) antiserum reacts with dodecandrin (*Phytolacca dodecandra*), to which it shows greatest sequence similarity (5 differences in the first 30 N-terminal residues), yet PAP reacts only weakly with the two other forms of PAP found in seeds and summer leaves, PAP II and PAP-S. These two forms are less related to PAP than dodecandrin.

The authors conclude that both the type I and type II RIPs compared have evolved from a common ancestor. In the case of the pokeweed RIPs it appears that a process of gene duplication followed by sequence divergence has led to the different forms of PAP expressed both at different times in the leaves and in the seeds. The evolution of the type II RIPs is presumed to have occurred by a process of gene fusion between a ribosome-inactivating polypeptide (analogous to a type I RIP) and a polypeptide with a lectin activity which becomes the subsequent B-chain of the type II RIP. Subsequent gene duplication and sequence divergence has produced the related dimerising agglutinins (eg in *Ricinus* and *Abrus)*. In some cases this has not occurred, modeccin does not have a related agglutinin whilst viscumin (Olsnes *et al.*, 1982) can act as both a heterotetrameric lectin or a heterodimeric toxin depending upon the concentration of protein. *Gelonium multiflorum* is in the same family as *Ricinus communis* (*Euphorbiaceae*) but produces a type I RIP (Stirpe *et al.*, 1980). This suggests that the gene fusion events producing type II RIPs are probably recent events.

N-terminal sequence comparisons have also shown homology between RIPs (saporins and pokeweed antiviral proteins) from *Saponaria officinalis*

(*Caryophyllaceae*) and *Phytolacca americana* (*Phytolaccaceae*) (Lappi *et al.*, 1985; Montecucchi *et al.* 1988). Similar comparisons have shown sequence relatedness between RIPs from *Cucurbitaceae* (bryodin and momordin) with RIPs from *Euphorbiaceae* (ricin A-chain and gelonin). These RIPs exhibited no significant sequence homology with the RIPs from either *Caryophyllaceae* or *Phytolaccaceae* (Montecucchi *et al.*, 1988).

Comparison of the primary sequences of ricin A-chain, barley protein synthesis inhibitor and trichosanthin with *E. coli* RNase H and retroviral reverse transcriptases have been carried out by Ready *et al.* (1988). They have demonstrated homologous domains within each of the proteins and suggest that they have all evolved from an ancient protein-folding unit that was capable of binding and modifying nucleic acids. In addition conserved residues were found within each of the RIP sequences. These were found to lie within the cleft of ricin A-chain proposed as the active site (see section 2.1.6) and have been implicated in the catalytic mechanism of these proteins. These conserved residues are also present in the A-chains of Shiga toxin and the *E. coli* Shiga-like toxins, SLT-I and SLT-II (Hovde *et al.*, 1988).

3. The mechanism of action of ribosome-inactivating proteins.

Until recently the nature of the enzymatic modification of 60 S ribosomal subunits by RIPs was unknown despite much biochemical attention (Olsnes and Pihl, 1982; Jiménez and Vázquez, 1985). The work of Endo and co-workers (Endo *et al.*, 1987; Endo and Tsurugi, 1987), and subsequently other groups, has determined that RIPs have a specific RNA N-glycosidase action. This is discussed in the following section. The nature of this action in relation to ribosome inactivation and the inhibition of protein synthesis is discussed in section 3.2.

3.1. The RNA N-glycosidase activity of ribosome-inactivating proteins.

The action of ricin, abrin and modeccin on the 28 S RNA of rat liver ribosomes was investigated by Endo *et al* (1987). rRNA extracted from ricin A-chain treated ribosomes was electrophoresed on non-denaturing gel systems. A slight decrease in mobility was noticed in one of the degradation product fragments when compared with a non-toxin treated rRNA sample. The fragment arose presumably due to RNase contamination and was located at the 3′ end of the 28 S RNA molecule. The terminal sequences at

each end of the fragment was the same in toxin treated and non-treated samples indicating that any modification must lie within the fragment. The fragment, isolated from a ricin treated rRNA sample, was treated with several RNases. Significantly, neither G-4323 or A-4324 were cleaved, suggesting a modification at this site by ricin A-chain. This modification is of great interest since A-4324 is the nucleotide adjacent to the site of α-sarcin cleavage within a conserved sequence (the phosphodiester bond between G-4325 and A-4326) (see section 2.3.1). The fragment was also observed in rRNA samples isolated from ribosomes previously treated with abrin or modeccin suggesting that these RIPs have a similar mechanism of action.

Treatment of the ricin A-chain modified fragment with either aniline at acidic pH or mild alkali resulted in the cleavage of the phosphodiester bonds 5′ and 3′ to A-4324. The authors suggested that the increased lability of these bonds was due to the specific depurination of A-4324 by ricin. Electrophoretic analysis of toxin treated rRNA samples subjected to aniline cleavage showed the presence of a RNA fragment the same size as the one produced by α-sarcin treatment of ribosomes.

It was also demonstrated that ricin could act on naked 28 S RNA, albeit at a much reduced rate, but denatured 28 S RNA could not act as a substrate. This suggested that a defined RNA secondary structure is required for recognition by ricin A-chain. *E. coli* ribosomes are not susceptible to ricin action despite having the same highly conserved sequence (see Fig 1.1). This suggests that recognition and cleavage is not merely sequence specific but is dependent upon the secondary structure of the RNA and is enhanced by its context within the ribosomal subunit (Endo *et al.*, 1987). This is further illustrated by the following observations: Endo and Tsurugi (1988) showed that ricin A-chain will depurinate naked *E. coli* 23 S RNA, albeit at high concentrations. Habuka *et al.* (1990) demonstrated that *Mirabilis* antiviral protein efficiently inhibits protein synthesis in *E. coli*, probably by ribosome inactivation. This suggests that ribosomal proteins may influence recognition of ribosomes by RIPs and that *Mirabilis* antiviral protein has a conformation capable of recognising and interacting with the conserved sequence in 23 S RNA. Endo and Tsurugi (1987) subsequently demonstrated that ricin A-chain exhibits a specific RNA N-glycosidase activity and hydrolytically cleaves the N-glycosidic bond between the ribose moiety and A-4324 of the 28 S

RNA.

Since the demonstration of the RNA N-glycosidase activity of ricin A-chain other plant RIPs have been shown to act upon rRNA of the large ribosomal subunit in an analogous manner. Stirpe *et al.*, (1988) demonstrated the production of the diagnostic aniline labile site in the 28 S RNA of rabbit reticulocyte lysates and the 26 S RNA of yeast lysates after treatment with various plant RIPs. The RIPs used were; recombinant ricin A-chain, barley protein synthesis inhibitor, bryodin, dianthins 30 and 32, dodecandrin, gelonin, momordin, saporin-6, trichokirin, tritin, viscumin and volkensin. RNA N-glycosidase action has also been independently demonstrated for the RIPs; gelonin, saporin, momordin, PAP, PAP II, PAP S (Endo *et al.*, 1988a) and *Phoradendron californicum* lectin (Endo *et al.*, 1989). May *et al.* (1989) have shown that ricin A-chain acts on an adenine residue of rabbit reticulocyte lysate 28 S RNA which is in the same homologous conserved region as A-4324 in rat liver 28 S RNA about

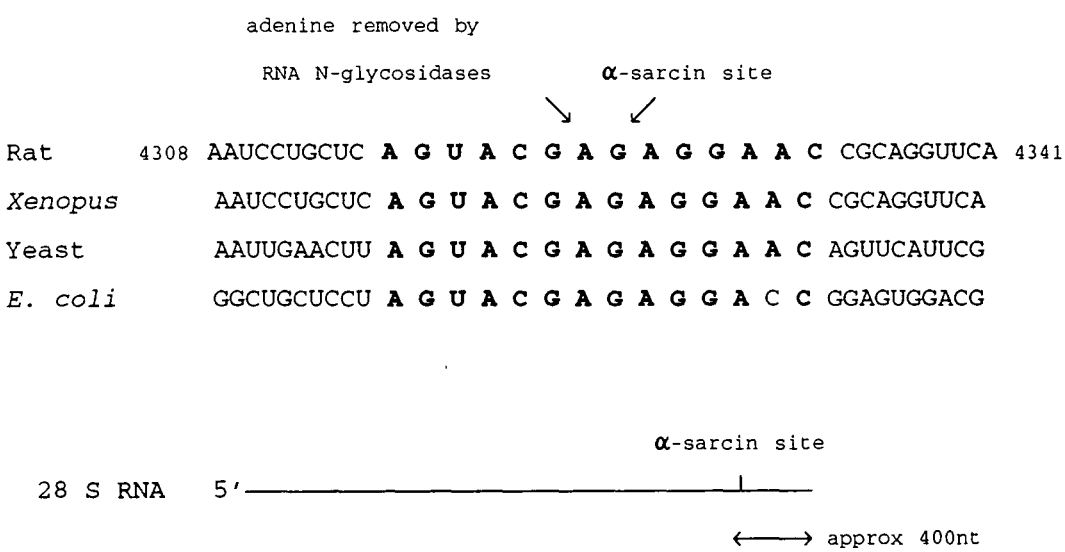


Fig 1.1. Specific nucleotides in 28 S and 26 S RNAs attacked by RIPs and α-sarcin.

The α-sarcin cleavage site is shown, determined for rat and *Xenopus*, along with the site of depurination due to RIP action. The numbering of the rat 28 S RNA is from the 5′ end (Chan *et al.*, 1983) and has a total length of 4718 nucleotides. The bases in bold type are invariant in all species. The diagram is modified from that presented by Saxena *et al.* (1989).

390 nucleotides from the 3′ end of the molecule. The yeast 26 S RNA also contains this conserved region about 300 nucleotides from the 3′ end of the molecule (Wool, 1984). Exposure of RIP treated RNA samples to aniline results in the production of a fragment of 380-400 nucleotides in length derived from 28 S RNA and about 300 nucleotides in length from the yeast 26 S RNA (Saxena *et al.*, 1989). The results obtained for the various RIPs listed above suggests that all plant RIPs have the same mode of action.

Shiga toxin and Shiga-like toxin II (SLT-II) (see section 2.3.2) have been shown to have the same specific RNA N-glycosidase activity on both rabbit and rat 28 S RNA (Endo *et al.*, 1988b). Saxena *et al.* (1989) have demonstrated the RNA N-glycosidase activity of toxins *in vivo* by injection into *Xenopus* oocytes. The toxins used were; ricin A-chain (native glycosylated form and recombinant non-glycosylated form), Shiga toxin and SLT-IIv, a variant form of SLT-II produced by *E. coli* causing edema in swine. The results of microinjection of these toxins showed that A-3732 of *Xenopus* 28 S RNA is specifically removed by all of the toxins studied, again, this is the analogous residue to A-4324 of rat 28 S RNA illustrated in Fig 1.1. Therefore, the range of ribosome-inactivating proteins exhibiting the same enzymatic action includes proteins not only from the plant kingdom but also from bacteria.

3.2. The nature of protein synthesis inhibition caused by the action of ribosome-inactivating proteins.

As noted previously the site of action of both α-sarcin and RIPs is within a purine rich 14 nucleotide single-stranded region which is almost universally conserved. This region forms an exposed stem and loop structure on the surface of the 60 S ribosomal subunit (Endo *et al.*, 1990). It has been demonstrated in *E. coli* that this conserved loop in 23 S RNA interacts with EF-Tu and EF-G, the prokaryotic equivalents of EF-1 and EF-2 (Moazed *et al.*, 1988). It is reasonable to assume that a similar interaction occurs in eukaryotic ribosomes. Terao *et al.* (1988) have proposed that the action of α-sarcin and ricin on 60 S subunits induces conformational changes at nearby but different locations within the subunit. This was determined by labelling toxin-treated ribosomal proteins with radioactively labelled N-ethylmaleimide. The extent of labelling was compared with ribosomal proteins from ribosomes that had not been subjected to either α-sarcin or ricin A-chain inactivation.

Differences in the reactivities of three ribosomal proteins to N-ethylmaleimide were noted after treatment with toxin. These were; ribosomal protein L14 after ricin treatment and proteins L3 and L4 after α-sarcin treatment. These three proteins have previously been shown to be located near each other in the large ribosomal subunit (Uchiumi *et al.*, 1985, 1986). This evidence taken with the demonstration that; high Mg^{2+} concentrations can restore activity to ricin treated ribosomes (Cawley *et al.*, 1979) and the thermal denaturation curve of ricin-trated 60 S subunits is shifted to lower temperatures (Paleologue *et al.*, 1986) does suggest that conformational changes in the 60 S subunit may be associated with ribosome inactivation.

The precise steps in protein synthesis inhibited due to the modification of 60 S ribosomes by ricin have only recently been revealed. It was known that the 60 S subunit was the only site of action of ricin A-chain (Sperti *et al.*, 1973; Montanaro *et al.*, 1973) before the discovery of RNA N-glycosidase activity.

Using a ribosome shift assay (Darnborough *et al.*, 1973) Osborn and Hartley (1990) have demonstrated the inhibition of the formation of the 80 S initiation complex due to ricin action. In this assay [^{35}S]Met-tRNA is bound initially to the 40 S pre-initiation complex, this subsequently becomes associated with the 60 S subunit to form the 80 S initiation complex. This is observed on a sucrose density gradient as a shift of [^{35}S] from the 40 S to the 80 S absorbance peak. If protein synthesis continues in the absence of inhibitors of elongation [^{35}S] activity is subsequently located in the polysome peak. Inhibition of elongation results in [^{35}S] activity accumulating in the 80 S monosome peak. When ricin A-chain or diphtheria toxin/NAD+ (elongation inhibitor) treated ribosomes are fractionated the bulk of the [^{35}S] is associated with the 40 S subunit and 80 S monosome indicating that; the binding of mRNA to the 40 S subunit is not inhibited (inhibition of mRNA binding would result in [^{35}S] confined exclusively to 40 S subunits) and that elongation is inhibited causing the accumulation of 80 S monosomes. However labelling of the 40 S subunit was greater in the ricin A-chain treated samples suggesting that ricin A-chain was inhibiting not only elongation but also the formation of the 80 S initiation complex from the 40 S pre-initiation complex and the 60 S subunit. By following the distribution of labelling after different incubation periods the authors were able to demonstrate

consistently greater activity associated with the 40 S subunit fraction. If inhibition was confined only to the elongation step this would result in similar amounts of activity associated with both ricin A-chain and diphtheria toxin treated ribosomes. Analysis of ricin A-chain and diphtheria toxin/NAD treated polyribosome profiles demonstrated the appearance of a (N + 1/2) mer species with ricin A-chain treatment. This was interpreted as polysomes bearing uncoupled 40 S subunits.

By analysing the residues attached to [^{35}S]Met labelled peptides synthesised by ricin A-chain using globin mRNA template in a rabbit reticulocyte system Osborn and Hartley (1990) demonstrated the precise step in elongation which is inhibited. The N-terminal dipeptide in both α and β globin is Met-Val. Inhibition of the binding of EF1-GTP-Val-tRNA to the peptidyl site of the ribosome would result in only Met-tRNA being associated with the 80 S monosomes, whilst inhibition of peptidyl transferase would result in both Met-tRNA and Val-tRNA being bound without the formation of a peptide bond between the two amino acids. Inhibition of translocation of the Met-Val dipeptide from the A to P site would result in the accumulation of Met-Val and confer resistance to puromycin which binds at the A site. The authors incubated ricin A-chain or diphtheria toxin treated reticulocyte lysates with globin mRNA template and [^{35}S]Met-tRNA, fractionated the incubation mixes on sucrose density gradients and after alkaline hydrolysis analysed the residues in the monosome peak using thin layer chromatography. In both ricin A-chain and diphtheria toxin treated lysates the predominant species was Met-Val. Since diphtheria toxin is known to block elongation at the translocation step (ie movement of Met-Val from the A to P site) it was concluded that the action of ricin A-chain on the 60 S ribosomal subunit results in the inhibition of EF2 dependent translocation. Further confirmation of this was the demonstration that puromycin treatment failed to release [^{35}S] from the monosome or polysome fractions. If the A site was free, puromycin would have bound with the subsequent release of the labelled dipeptide. This indicated an inhibition of translocation and the occupation of the A site by Met-Val dipeptidyl tRNA. The inhibition of translation appears to result from an inability of EF2 to bind to the modified ribosomes. Although ricin A-chain treatment subsequently inhibits both initiation and translocation their relative contributions to inhibition of protein synthesis either *in vitro* or *in vivo* is still

unknown.

Inactivated ribosomes have also been demonstrated to show a concentration dependent inhibition of binding of aminoacyl-tRNA to the A site. Inhibition being observed at low concentrations of EF1 with no observed inhibition at saturating concentrations of EF1 (Fernandez-Puertes., *et al.* 1976). Inactivated ribosomes have also been shown to exhibit reduced ability to hydrolyse GTP in the presence of EF1 and EF2 (Olsnes and Pihl, 1982) but may be due to a reduction in the binding of each of the EFs.

3.3. Studies on the putative active site of ricin.

The presence of conserved residues in the cleft of ricin A-chain, the proposed active site, has led to investigations to define which residues are critical for catalytic activity. Frankel *et al.* (1989) used an inducible yeast system to select for ricin A-chain mutants which when expressed in yeast would not cause cell death. Nine mutants were recovered, all had point mutations, leading to amino acid changes, in residues directed into the active site cleft. Further amino acid changes in this region have also been investigated (Frankel *et al.*, 1990; Bradley and McGuire, 1990). Changes in the proposed active site residues are discussed further in the context of the results presented in Chapter 5 section 5.5.2). Deletions in the coding sequence of ricin A-chain have also been studied. May *et al.* (1989) demonstrated that the deletion of a pentapeptide region, showing homology to hamster elongation factor 2, abolished activity of transcripts translated in a rabbit reticulocyte lysate system. The deletion of a second pentapeptide region, containing conserved residues in the A-chain cleft, was reported not to affect the activity of the A-chain. However it has since been reported that this result is in fact incorrect and that deletion of these residues does abolish A-chain activity (Dr. M Hartley, Department of Biological Sciences, University of Warwick, personal communication).

3.4. The action of ribosome-inactivating proteins on ribosomes from plants.

The literature contains conflicting reports concerning the susceptibility of plant ribosomes to RIP action. Batelli *et al.* (1984) reported that RIPs do not affect the ribosomes of the plant from which the RIP had been isolated and reported that in general plant ribosomes are considerably

less sensitive to RIP action. Tritin, the RIP present in wheat-germ does not inactivate wheat-germ ribosomes (Coleman and Roberts, 1981). Ricin A-chain can also be efficiently translated in a wheat-germ translation system (May *et al.*, 1989). Owens *et al.* (1973) reported that pokeweed antiviral protein (PAP) inhibits wheat-germ and cowpea ribosomes but has no effect on pokeweed ribosomes. Stirpe (1982) suggested that RIPs could only act on ribosomes from heterologous species.

Ricin has been reported to inhibit protein synthesis by *Ricinus* ribosomes, albeit at ID_{50} concentrations four orders of magnitude greater than comparable mammalian systems (Harley and Beevers, 1982; 1986). Protein synthesis in a pea polysome translation system was inhibited by both ricin and saporin (Gatehouse *et al.*, 1990a). The ID_{50} values for this system were comparable to those obtained for inhibition in a rabbit reticulocyte system (Stirpe and Barbieri, 1986).

The idea that RIPs are only active against heterologous ribosomes must be called into question. Only very low levels of protein synthesis were observed in the study of PAP action on pokeweed ribosomes (Owens *et al.*, 1973). This was also the case for the action of the seed form of PAP (PAP-S) on pokeweed ribosomes (Batelli *et al.*, 1984). Ready *et al.* (1986) reported that attempts to isolate functional pokeweed ribosomes could only demonstrate background levels of protein synthesis. The authors postulated that endogenous PAP was in fact inactivating the ribosomes. This was confirmed by Taylor and Irvin (1990) who demonstrated that the large ribosomal RNA isolated from pokeweed was specifically depurinated at a single site; aniline treatment and denaturing gel electrophoresis showed the production of a small fragment. Endogenous PAP is presumed to have acted on the pokeweed ribosomes during the extraction procedure. This would explain the low levels of protein synthesis noted previously. Similar preparations of wheat-germ RNA (containing tritin) did not contain the aniline-labile site consistent with the observed lack of activity of tritin on wheat-germ ribosomes. Considering this information it appears that the activity of RIPs to plant ribosomes is dependent both upon the type of RIP used and the source of ribosome. It is possible that other plant ribosomes are susceptible to RIPs present in the plant but this remains to be determined.

4. The role of ribosome-inactivating proteins in plants.

Whilst there has been a great deal of attention devoted to the isolation and characterisation of RIPs their function within plants is still controversial. It is possible that the type II RIPs may serve as defence proteins due to their great toxicity to whole organisms. However the role of type I RIPs is even more unclear as they are generally much less cytotoxic. As discussed earlier (section 2.2.2), the RIPs present in pokeweed have been shown to inhibit viral replication. The possible function of RIPs as antiviral factors is discussed in the next section.

4.1. Antiviral activity.

The sap of *Phytolacca esculenta* has been demonstrated to inhibit the infectivity of tobacco mosaic virus (TMV) and other plant viruses when mixed with the virus and applied to the leaves of *Nicotiana glutinosa*. A glycoprotein component of the sap was implicated in the inhibitory effect (Kassanis and Kleczkowski, 1948). Irvin *et al.* (1980) demonstrated that low concentrations of the pokeweed antiviral proteins, PAP and PAP II, mixed with TMV inhibited the formation of local lesions when applied to the leaves of *Phaseolus vulgaris*. The inhibitory action of other RIPs (ricin, abrin, modeccin and gelonin) on TMV infectivity in the leaves of *Nicotiana glutinosa* was demonstrated by Stevens *et al.* (1981). The interactions between PAP and TMV were studied by Kumon *et al.* (1990). The authors concluded that there was no requirement for association between PAP and the viral particle for inhibition of replication to occur

It has been postulated that the inhibition of viral replication is due to the entry of RIPs into the cell along with viral particles. Subsequent inhibition of protein synthesis would prevent viral replication. PAP is reported to be localised in the cell wall matrix and is active against pokeweed ribosomes (Robertus *et al.* 1986; Taylor and Irvin, 1990). This is consistent with an antiviral role for PAP. Whilst cell death due to RIP action is a reasonable hypothesis to explain the inhibition of viral replication it is reported that pokeweed is suceptible to some viral infections (Bhargava KS, 1951, cited by Ready *et al*, 1984). As such further work is required to determine if RIPs act as a "cell suicide" system when subjected to viral attack. It is possible that they are active against other pathogens. Roberts and Selitrennikoff (1986) reported that barley protein synthesis inhibitor inhibited the growth of

the fungus *Trichoderma reesi* and suggested that it may have a role in protection against fungal attack.

4.2. Anti-pest factors.

The toxic effects of ricin and saporin on insects have been studied by Gatehouse *et al.* (1990a). Feeding trials based on artificial diets containing different levels of the purified RIPs indicated that certain insect pests were susceptible to low levels of ricin and saporin. While the two *Coleopteran* species tested, cowpea seed weevil (*Callosobruchus maculatus*) and cotton bollweevil (*Anthonomus grandis*), were both highly susceptible (LD_{50} of less than 10^{-2} % w/w) the two *Lepidopteran* species, army worm (*Spodoptera littoralis*) and tobacco budworm (*Heliothis virescens*) were completely resistant. *In vitro* assays demonstrated that the resistance was probably due to proteolytic degradation of the toxins in the guts of the *Lepidoptera*.

5. Aims of the work.

The aims of this project were; to isolate, clone and sequence functional RIP coding sequences and show that these did in fact encode polypeptides with RIP activity. This would provide active coding sequences which could then be expressed in transgenic tobacco plants to determine if their expression would confer pest or viral resistance to the plants. Two representative RIPs were chosen as candidates for gene isolation and characterisation; ricin, a type II RIP and saporin, a type I RIP.

The results of this project are presented in the following chapters. Chapter 4 presents the results for the cloning of ricin gene sequences. The isolation and characterisation of genomic clones containing saporin gene sequences is presented in chapter 5. Chapter 6 describes the results of the *in vitro* expression and activity of the isolated ricin and saporin gene sequences.

During the course of this work a decision was taken to place greater emphasis on the isolation and characterisation of saporin gene sequences. Clones for ricin had already been isolated and characterised. At the commencement of this work there had been no published reports of the cloning of any other RIP genes. Hence, it was decided to direct attention mainly to the cloning of saporin gene sequences. Saporin was strategically a more difficult RIP to work on, due to; the small size of

the seeds, the indeterminate flowering and seed set, making cDNA cloning difficult, the lack of substantial protein sequence information and the lack of availability of anti-saporin antibodies. The analysis of saporin genes was potentially of greater interest from a basic molecular biology aspect as well as from an applied point of view.

2. MATERIALS.

2.1. Glassware and plasticware.

All plasticware used for DNA and RNA manipulations was siliconised with dimethyldichlorosilane (2% v/v in 1,1,1-trichloroethane), washed in distilled water, dried and autoclaved. Glassware for DNA manipulations was treated in the same manner and for RNA work, in addition, it was baked overnight at 170^{0}C after siliconisation.

2.2. Plant material.

Soapwort (*Saponaria officinalis* L.) seeds were a gift from Professor F. Stirpe and Dr L. Barbieri, University of Bologna, Italy, and were stored at 4^{0}C. Purified saporin-6 protein was from the same source and was stored, dissolved in water, at -20^{0}C.

2.3. Bacterial strains.

E. coli strains K803, LE392, P2392, BHB2688, BHB2688 and JM101 were available within the Department of Biological Sciences, University of Durham, UK. *E. coli* strain DH5α, was supplied as frozen competent cells by GIBCO-BRL Ltd, Paisley, Scotland. Strain genotypes and source references are listed in Sambrook *et al* (1989) and Fedoroff (1983).

2.4. Nucleic acids.

Total preparations of castor oil plant (*Ricinus communis* L.) and *Saponaria officinalis* DNAs, isolated from young leaves, were generously provided by Mrs E. Croy, Department of Biological Sciences, University of Durham, UK.

Plasmid pRCL59 was generously donated by Professor J.M. Lord and Dr L. Roberts, Department of Biological Sciences, University of Warwick, UK.

Plasmids pUC18 and pUC19 and replicative form M13 DNAs were obtained from Boehringer Mannheim UK, Lewes, UK.

A plasmid pUC18 derivative, pUC18Xho, was created by ligating the following palindromic oligonucleotide sequence into Xba1 restricted pUC18 ($^{5'}$CTAGCCTCGAGG$^{3'}$). The Xba1 site is lost and a Xho1 site is created, the Xho1 site is compatible with Sal1 restriction fragment "sticky ends". Both Xho1 and Sal1 sites are lost in Sal1 / Xho1 ligation reactions. The plasmid was provided by Dr R. Croy, Department of Biological Sciences, University of Durham, UK.

Ligated λ cI857 Sam7 DNA and λ EMBL3 DNA was from Stratagene, California, USA. λ DNA size markers were purchased from Northumbrian Biologicals, Cramlington, UK.

Poly $(A)^+$ RNA extracted from 16 d.a.f. developing cotyledons of pea (*Pisum sativum* L., var. Feltham First) and plasmid pJY25 were kindly supplied by Drs I. M. Evans and J. Yarwood, respectively, both of the Department of Biological Sciences, University of Durham, UK.

Oligonucleotides were synthesized on an ABI 381A DNA synthesizer by Mr J. Gilroy, Department of Biological Sciences, University of Durham, UK.

2.5. Chemical and biological reagents.

Chemical reagents were purchased from BDH Ltd., Poole, Dorset, UK, and were of "AnalaR" grade or the best available. Other reagents were obtained from the sources listed below.

Restriction endonucleases, T_4 DNA ligase, T_4 polynucleotide kinase, X-Gal, IPTG; Northumbrian Biologicals, Cramlington, UK.

Calf intestinal alkaline phosphatase, Klenow fragment of *E. coli* DNA polymerase 1 (sequencing grade), calf liver tRNA; Boehringer Mannheim UK, Lewes, Sussex.

Taq DNA polymerase (purified from *T. aquaticus* or cloned "Ampli*Taq*" enzyme), deoxynucleotide triphosphates; Perkin-Elmer / Cetus, ILS Ltd, Newbury Street, London, UK.

Acridine orange, ATP, ampicillin, bromophenol blue, bovine serum albumin, DNase, DTT, ethidium bromide, herring sperm DNA, lysozyme, mineral oil, polyvinylpyrollidone, pronase, RNase A, spermidine, xylene cyanol; Sigma Chemical Co. Poole, Dorset, UK.

Radiochemicals, nylon "Hybond-N" membranes, micrococcal nuclease treated rabbit reticulocyte lysate; Amersham International plc., Bucks., UK.

Non-nuclease-treated rabbit reticulocyte lysate; Department of Biochemistry, University of Cambridge, Tennis Court Rd, Cambridge, UK.

LambdaSorb DNA purification reagent, *Taq* DNA polymerase, non-nuclease-treated rabbit reticulocyte lysate; Promega Ltd., Enterprise Road, Southampton, UK.

Oligonucleotide purification cartridges (OPC), phosphoramidite derivatives of nucleotide bases; Applied Biosystems, Warrington, UK.

Ficoll-400, Sephadex G-50 and G-25, TransProbe SP6 transcription kit;

Pharmacia LKB, Milton Keynes, UK.

Nitrocellulose filters BA85 (0.45 μm); Schleicher and Schuell, Dassel, FRG.

3MM chromatography paper and 2.5cm GF/C filters; Whatman Ltd., Maidstone, Kent, UK.

Electrophoresis grade agarose, low melting-point agarose; GIBCO-BRL Ltd., Paisley, Scotland.

Kodak X-ray film cassettes were kindly donated by the Freeman Hospital, Newcastle upon Tyne, UK.

High gelling-temperature agarose; FMC Bioproducts, Rockland, ME, USA.

Anion exchange resin (Qiagen) tips; Diagen GmbH, Düsseldorf, FRG.

Taq DNA polymerase; Stratagene, California, USA.

Cellulose acetate filters (0.2μm); Sartorius GmbH, Göttingen, FRG.

PPO, POPOP; Koch-Light Ltd., Colnbrook, Berks., UK.

Yeast extract, Bacto-Agar; Difco, Detroit, Michigan, USA.

Trypticase peptone; Becton Dickinson, F-38240, Meylan, France.

Fuji RX X-ray film; Fuji Photo Film Co. Ltd., Japan.

3. METHODS.

3.1. Plant growth conditions.

Saponaria officinalis seeds were grown outdoors, in pots of Levington compost or similar, in the UK or in Italy. Young leaf material was harvested, frozen in liquid nitrogen, and stored at -80^{0}C prior to DNA isolation.

3.2. General techniques used for DNA manipulation and bacterial growth.

Standard DNA manipulations, ligation with T_4 DNA ligase, de-proteinisation by phenol extraction and ethanol or iso-propanol precipitations were carried out essentially as described by Maniatis *et al* (1982).

Restriction endonuclease reactions were performed using the buffers and reaction conditions recommended by the manufacturers. Typical reaction times were 2-4 hours. Restriction enzymes were inactivated by either heating to 65^{0}C for 10 mins or phenol extraction.

DNA and RNA concentrations were measured using a PYE Unicam SP8-150 UV/Vis spectrophotometer and 0.5 ml quartz cells with a 1.0 cm light path length. An absorbance at 260 nm of 1 OD unit was taken to be approximately equivalent to a double stranded DNA concentration of 50 μg/ml, an RNA concentration of 40 μg/ml and an oligonucleotide concentration of 20 μg/ml (Maniatis *et al* 1982).

The maintenance of bacterial strains on agar plates, growth in liquid culture and long term storage frozen in glycerol, was as described by Maniatis *et al* (1982).

3.3. Extraction of DNA.

3.3.1. Extraction of total plant DNA.

Saponaria officinalis DNA was prepared essentially as described by Graham (1978) and further purified by two rounds of centrifugation through ethidium bromide-caesium chloride density gradients.

3.3.2. Preparation of plasmid DNA from *E.coli*.

Growth media was inoculated with either a single bacterial colony or 10 μl of glycerol preserved culture and grown to stationary phase using appropriate antibiotic selection. Plasmid DNA was then extracted using one of the following methods.

3.3.2.1. Alkaline lysis method.

A protocol was derived from that of Birnboim and Doly (1979): Starting with 10 ml of culture, cells were pelleted using a bench top centrifuge (4,000 rpm for 15 min) and resuspended in 200 µl of lysozyme solution (25 mM Tris-HCl pH 8.0, 10 mM EDTA, 50 mM glucose, 4 mg/ml lysozyme). After 30 min incubation on ice 400 µl of 0.8% (w/v) NaOH, 1.0% (w/v) SDS was added, incubated for 5 min on ice, followed by 300 µl 3M/5M potassium acetate pH 5.6. After gentle mixing, debris was pelleted by centrifugation (12,000 rpm, 20 min) in a haematocrit micro-centrifuge for 30 min at 4^0C. 0.7 ml of the supernatant was phenol extracted and ethanol precipitated. Plasmid DNA was resuspended in 50 µl of 0.6 mg/ml DNase-free RNase made up in sterile water.

3.3.2.2. Qiagen tips method.

Plasmids were extracted from cultures according to the protocol supplied by the manufacturers. In essence this consisted of a Triton X-100 and lysozyme lysis step, removal of cellular and chromosomal debris by centrifugation, followed by binding to the Qiagen anion exchange matrix. The plasmid DNA was washed and eluted from the matrix using the recommended ionic strength buffer.

3.3.2.3. Silica fines method.

This procedure was developed by, and the silica fines supplied by, Dr N.J. Robinson, Department of Biological Sciences, University of Durham, UK (personal communication). The method is based on that described by Golden *et al* (1987). 1.5 ml of culture was spun down by centrifugation in a 1.5 ml Eppendorf tube using an MSE micro-centrifuge (13,000 rpm, 5 min) The pellet was resuspended in 0.35 ml of 8% (w/v) sucrose, 0.5% (v/v) Triton X-100, 50 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0, mixed by brief vortexing and transferred to a boiling water bath for 90 seconds. Samples were immediately centrifuged for 15 min (13,000 rpm, room temperature) and the pelleted debris was removed with a sterile toothpick. 0.7 ml of sodium iodide solution (see below) was added to the supernatant followed by 5 µl of silica fines suspension (Golden *et al* 1987) and incubated at room temperature for 15 min to bind plasmid DNA to the fines. Silica fines were pelleted by centrifugation for 30 seconds, in a micro-centrifuge, washed with 70% ethanol and dried under vacuum. DNA was eluted from the fines by resuspending in 50 µl of sterile water. Plasmids

prepared in this manner were sufficiently pure for direct use in restriction endonuclease reactions without RNase treatment.

Sodium iodide solution was prepared as follows. To 100 ml water 90.8 g Na iodide and 1.5 g Na_2SO_3 was added followed by filter sterilisation. 0.5 g Na_2SO_3 was added to saturate the solution which was then stored in the dark at 4^0C.

3.3.3. Preparation of phage λ DNA.

3.3.3.1. Extraction of λ DNA from liquid lysate cultures.

A protocol developed by Dr A. Ryan, Department of Biological Sciences, University of Durham, UK, (personal communication), was used. A single colony of *E. coli* P2392 was inoculated into 5ml of L-broth containing 0.4% maltose and grown overnight. A single well separated phage plaque was eluted into 0.5 ml of SM buffer (Maniatis *et al.*, 1982) and left overnight at 37^0C. 0.25 ml of this phage suspension was added to 0.1 ml of *E. coli* P2392 culture and left at 37^0C for 15 min. The mixture was inoculated into 400 ml of prewarmed L-broth (37^0C) containing 10 mM $MgCl_2$ in a baffled flask and shaken slowly overnight at 37^0C. 10ml of chloroform was added to complete lysis incubating for 30 min at 37^0C. The supernatant was recovered after centrifugation (4,000rpm /10min /4^0C). RNase and DNase was added each to a final concentration of 10 μg/ml and the lysate was incubated at room temp for 30 min. NaCl and PEG 6,000 were added to final concentrations of 1.0 M and 10% (w/v) respectively and dissolved by slow shaking at room temperature. The lysate was then incubated overnight on ice. The phage/PEG precipitate was pelleted (10,000rpm/4^0C/30min) and the excess supernatant removed with paper towels. The phage precipitate was resuspended in 8.0 ml of SM buffer and one volume of chloroform was added to extract the PEG. After 30 seconds vortex mixing the phases were separated by centrifugation for 10 min at 4,000rpm/4^0C. The aqueous phase was recovered and the phage were pelleted by spinning for 120min/40,000rpm/4^0C in a 10x10 ml angle head rotor using a Prepspin ultracentrifuge. Pellets were resuspended in 2.0 ml of SM. One tenth volume of 0.25 M EDTA, pH8.0 was added followed by SDS to 0.5% (w/v). Phage were then incubated in pronase (0.5 mg/ml final concentration) for 60 min at 37^0C (pronase was self-digested before use by incubating the stock aqueous solution at 37^0C for 60 min). The phage DNA preparation was then phenol extracted and ethanol precipitated. Precipitated DNA was recovered by centrifugation, washed twice in 70%

(v/v) ethanol, and resuspended in 40µl TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0).

3.3.3.2. λ DNA preparation using a phage specific antibody immunoadsorbent.

Phage λ DNA was prepared using a Promega LambdaSorb antibody preparation. The method relies on the use of rabbit polyclonal antibodies directed against phage λ linked to *Staphylococcus aureus* cells. The LambdaSorb preparation precipitates intact phage particles from a plate or liquid lysate. Phage λ clones were plated out using, *E. coli* P2392, as host in an agarose top layer at a density producing just sub-confluent growth. A crude lysate was prepared by incubating the top layer agarose with SM buffer. λ DNA was then extracted from the lysate as described in the supplied protocol. The yield of λ DNA was about 10 µg per preparation from 30 ml of plate lysate and provided a good substrate for restriction endonuclease digestion.

3.3.4. Preparation of single stranded M13 bacteriophage DNA.

Recombinant M13 DNA was extracted from cultures of transformed *E. coli* (strain JM101) grown overnight in 2xYT broth. Cells and cell debris were removed by centrifugation in an MSE micro-centrifuge (13,000, rpm, 5 min). 1.0 ml of supernatant was taken and phage precipitated by the addition of 200 µl of 20% (w/v) PEG 6000, 2.5 M NaCl followed by 30 min incubation at room temperature. Phage precipitates were pelleted by centrifugation and resuspended in TE buffer. Single stranded DNA was purified by phenol and chloroform extractions followed by ethanol precipitation. Samples were analysed on 0.7% agarose gels using non-recombinant M13 mp18 single stranded DNA as a size marker. Inserts cloned into single restriction sites in the M13 vectors had their orientations determined by "turn round assays" as described by Messing (1983).

3.4. Construction of a *Ricinus communis* gene library using phage λ EMBL3.

A *Ricinus communis* gene library was constructed essentially as described by Frischauf *et al.* (1983) using total DNA partially restricted with Sau3A. The phage vector employed (λ EMBL3) allows for the genetic selection of recombinants based on the Spi phenotype (sensitive to P2 phage interference). Wild type EMBL3 (Spi^+) carries the *red* and *gam* recombination genes on the stuffer fragment and its growth on *E. coli*

lysogenic for the prophage P2 is severely restricted. Recombinant phage (Spi^-) contain insert DNA instead of the stuffer and are not susceptible to growth inhibition on P2 lysogens (Karn *et al.* 1980; Lindahl *et al.* 1970).

All media and manipulations to maintain and propagate phage λ stocks were prepared and carried out as described by Maniatis *et al.* (1982).

3.4.1. Partial restriction of *Ricinus communis* DNA with Sau3A.

Restriction conditions producing the maximum number of DNA fragments in the 20 kb size range were established using trial incubations of DNA, with a limiting amount of Sau3A for various incubation times. (Typical reaction conditions were DNA 0.24 μg/μl, Sau3A 0.016 U/μl, final reaction volume 25 μl with incubation times of 2.5, 5.0, 7.5 and 10min). After termination of the reaction with the addition of EDTA pH 8.0 to 20 mM and heating at 70^0C for 10 min, restricted DNAs (1.2μg) were analysed on 0.4% agarose gels run at 15V overnight. Reaction conditions producing the maximal number average length fragments of approximately 20 kb were chosen, as judged by the maximal fluorescent intensity at approximately 40 kb in length (Seed, 1982). Large scale partial restriction digests were then carried out using 10 μg of genomic DNA per reaction. The selected time point chosen was 5 min with further digestions of 2.5, 7.5 and 10 min. Each sample was then processed separately to assess which would yield the most recombinant phage on subsequent ligation and packaging steps.

Partially digested DNA was purified by phenol extraction and ethanol precipitation. To prevent intermolecular ligation of insert DNA when ligating to vector DNA, the 5' phosphate groups of the insert DNA were removed using calf intestinal alkaline phosphatase, and the phosphatase was heat inactivated as described by Maniatis *et al* (1982). The efficiency of the phosphatase step was checked by taking an aliquot from each sample and carrying out the phosphatase reaction in the presence of pBR322 restricted with EcoR1. A second sample was phosphatase treated before the addition of pBR322. Both samples were then ligated with T_4 ligase, and the fraction of religated plasmid was measured by determining the number of colonies recovered for each sample when the ligation mix was used to transform *E. coli* JM101. The number of transformants recovered from the sample containing phosphatased pBR322 was between $3\text{-}7\times10^{-4}$ lower than the comparable non-phosphatased pBR322 sample. The phosphatasing of

the genomic DNA samples was of a sufficiently high efficiency, in the library construction, to reduce the number of intermolecular ligation events between insert DNA molecules to a low level frequency event compared with the ligation of single insert DNA molecules to vector.

3.4.2. Preparation of vector DNA.

EMBL3 DNA was restricted sequentially with BamH1 and EcoR1 endonucleases,phenol extracted and iso-propanol precipitated. Digestion to completion was checked by electrophoresis on 0.7% (w/v) agarose gel. Double digestion prevents the re-ligation of the stuffer fragment to the vector arms since the stuffer has EcoR1 compatible ends, whilst the arms are BamH1/Sau3A compatible. The short nucleotide sequence between the BamH1 and EcoR1 sites was eliminated during the iso-propanol precipitation step.

3.4.3. Ligation of EMBL3 arms to insert DNA.

Ligation reactions were set up using each of the genomic DNA partial digests. Ligations containing 1.5 μg EMBL3 DNA, 1.0 μg insert DNA and 1-2 Weiss units of T_4 ligase in a total volume of 5 μl were incubated at 15^0C overnight.

3.4.4. *In vitro* Packaging reactions.

The freeze thaw and sonic lysates required for *in vitro* packaging of λ DNA concatemers were prepared and used as described by Maniatis *et al* (1982) using the *E.coli* strains BHB2688 and BHB2690. The optimum sonication time was determined by calculating the packaging efficiencies of sonic lysates sonicated for different times using a fixed amount of freeze thaw lysate. DNA used in these test packaging reactions was ligated λ c1857 Sam7 DNA. Packaging reactions were further optimised by titrating the freeze thaw lysate with varying volumes of the most efficient sonic lysate preparation.

1 μl of each ligation mix was *in vitro* packaged and diluted to 500 μl with phage buffer. The packaging efficiency (pfu/μg DNA added) and number of recombinants produced was determined by plating out the packaged phage on *E. coli* indicator strains. Packaging efficiencies were determined by using *E. coli* K803. Efficiencies ranged from 9×10^5 pfu/μg for the insert DNA restricted for 2.5 min to 2×10^5 pfu/μg for the 10 min restriction.

The number of recombinant phage unable to propagate due to methylation

of the insert DNA was determined by plating out on both *E. coli* K803 (methylation insensitive) and *E. coli* LE392 (methylation sensitive). In all cases, approximately 35% of the clones were unable to grow on *E. coli* LE392 compared with growth on *E. coli* K803. To eliminate the problem of methylation-sensitive recombinants the library was amplified using *E. coli* K803 (section 3.4.5). Replication on this host removes the methylation of insert DNA and allows subsequent growth on the selective host *E. coli* P2392 (Fedoroff, 1980).

3.4.5. Production and amplification of the library.

The trial packagings described above indicated that the first two partial digestion times of 2.5 and 5 min produced the optimal number of recombinants. DNA from these digests were pooled to maximise the randomness of the insert DNA, ligated to EMBL3 arms and *in vitro* packaged as described above. Approximately 6×10^5 recombinants were obtained from several packaging reactions, as determined by titering a small aliquot on the P2 lysogen *E. coli* P2392. Loss of methylated recombinants was prevented by plating the packaged phage on large (22x22 cm) plates using *E. coli* K803 as host. The resultant phage were eluted into phage buffer and stored as an amplified library at 4^0C with the addition of chloroform to 0.3%

3.4.6. Plating out and screening of the gene library.

Phage were plated out, at a high density in an agarose top layer, on large (22x22 cm) plates using *E. coli* P2392 as host. Plaques were allowed to grow to pin-prick size, approximately 0.5 mm, at which time growth was stopped by cooling to 4^0C. Plaque lifts were taken from each plate as described in section 3.9.1. Plaque lifts taken from large (22x22 cm) plates used in the primary screen of a gene library were taken in duplicate to minimise the chances of false positives appearing in subsequent hybridisations.

3.4.7. *Saponaria officinalis* L. gene library.

A gene library of *Saponaria officinalis* L. was constructed using phage λ EMBL3 essentially as described above and was supplied by Dr P. Taylor. The library was amplified, stored and used as described for the *Ricinus* library.

3.5. Transformation of *E.coli*.

3.5.1. Transformation with plasmid DNA.

Competent *E. coli* DH5α cells were obtained commercially and transformed with plasmid DNA as described in the suppliers protocol.

3.5.2. Transformation with phage M13 mp18 / 19 replicative form DNAs.

M13 replicative form DNA was transformed into *E. coli* DH5α competent cells as for plasmid DNA except incubation at 37^0C to allow the expression of plasmid encoded antibiotic resistance was not required. *E. coli* DH5α is able to support M13 replication but lacks the F^- episome and will not therefore allow plaque formation. After transformation the cells were plated out with *E. coli* JM101, a strain which will support plaque formation, according to the method of Messing (1983). Transformed cells were added to 3.0 ml of top layer agar, at 45^0C, containing 0.3 mM IPTG, 50 μl X-Gal (2% w/v in dimethyl formamide), 200 μl of exponentially growing *E. coli* JM101 cells, mixed and plated onto agar plates.

3.6. Synthesis and purification of oligonucleotides.

Oligonucleotides, including inosine-containing oligonucleotides, were synthesized using an Applied Biosystems 381A DNA synthesizer operated with a standard synthesis programme. After cleavage from the matrix support using concentrated ammonia, the purification procedures for oligonucleotides depended upon their length and are described below. Purified oligonucleotides were stored either dry or dissolved in water at -20^0C.

3.6.1. "Trityl on synthesis".

Oligonucleotides of length 60 bases or more were synthesized as normal but the 5′ dimethoxytrityl (DMT) blocking group was not cleaved from the oligonucleotide with TCA by programming the synthesizer to omit this final step. The DMT group present at the 5′ end of the oligonucleotide then allows the full length product to be purified from shorter length incomplete oligonucleotides using a reverse phase column (Applied Biosystems) following the supplied protocol. The tritylated full length species was bound to the column, washed and then detritylated with TCA. The resulting "trityl off" species was eluted from the column using 30% (v/v) acetonitrile, not 20% (v/v) acetonitrile as specified in the protocol; this modification was found to increase yield. Fractions

containing the purified oligonucleotide were detected by measuring the absorbance at 260 nm, pooled, dried in a vacuum centrifuge and resuspended in water.

3.6.2. "Trityl off synthesis".

Oligonucleotides less than 60 bases in length were synthesized with a final TCA detritylation and then cleaved and deprotected. The resultant "trityl off" species was dried down under vacuum, twice resuspended in water and vacuum dried. Oligonucleotides treated in this manner were then used without any subsequent purification.

3.7. Radioactive labelling of DNA.

3.7.1. Labelling of DNA restriction fragments.

DNA restriction fragments purified from low melting point agarose gels (see section 3.8.2) were labelled with [α-^{32}P]dCTP using the random priming protocol of Feinberg and Vogelstein (1983). Labelled DNA was separated from unincorporated nucleotides using Sephadex G-50 equilibriated with 0.3 M NaCl, 0.1% (w/v) SDS, 50 mM Tris-HCl pH 7.5, 10 mM EDTA in a 10 ml disposable pipette. Specific activities of 10^9 cpm/μg DNA were routinely achieved.

3.7.2. End labelling of oligonucleotides.

Oligonucleotides were labelled using T_4 polynucleotide kinase and [γ-^{32}P]ATP essentially as described by Berents *et al* (1985). Typical reactions contained 10 pmol of oligonucleotide, 100 μCi [γ-^{32}P]ATP (20 pmol ATP), 40 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5mM DTT and 5-10 units of T_4 polynucleotide kinase. Reactions were incubated at 37^{0}C for 90 min, 90 μl of TE was added and the reaction mix heated at 80^{0}C for 5 min then frozen in liquid nitrogen. Labelled oligonucleotides were separated from unincorporated nucleotide as described in section 3.7.1. using a 10 ml Sephadex G-25 column .

3.8. Agarose gel electrophoresis of nucleic acids.

3.8.1 Electrophoresis of DNA.

DNA was size fractionated by electrophoresis in agarose gels basically as described by Maniatis *et al.* (1982). Agarose concentrations employed were dependant upon the size range of fragments to be separated. In general, DNA fragments less than than 1 kb in size were separated on 1.0% (w/v)

agarose gels. Fragments up to 23 kb in size were separated on 0.6-0.9% (w/v) agarose gels. Large agarose gels were cast in TAE gel buffer (40 mM Tris-acetate pH 7.7, 2 mM EDTA) whereas agarose minigels were cast in TBE buffer (89 mM Tris-borate, 2 mM EDTA). A sixth volume of gel-loading buffer (0.25% w/v each bromophenol blue and xylene cyanol, 15% w/v Ficoll 400 in sterile water) was added to each sample before loading. Visualisation of DNA under 300nm ultra-violet light was achieved by including 0.5 μg/ml ethidium bromide in both gel and running buffer. A Polaroid MP-4 land camera, fitted with a Kodak 23A Wrattan filter, was used to photograph gels whilst under UV illumination, using Polaroid type 667 film.

3.8.2 Isolation of DNA restriction fragments using low melting point agarose gels.

DNA fragments were separated by electrophoresis as described above (section 3.8.1.) but using low melting point agarose. DNA fragments of interest were excised from the gel with a sterile scalpel blade, placed in 1.5ml eppendorf tubes and heated at 70^0C for 15 min to melt the agarose. DNA was then purified by phenol extraction and ethanol precipitation after the addition of 3 volumes of 20 mM Tris-HCl pH 8.0, 1 mM EDTA.

An alternative protocol (Dr M.D. Watson, Department of Biological Sciences, University of Durham, UK, personal communication) utilising silica fines was also employed. After melting the agarose at 70^0C, samples were incubated at 37^0C for 5 min followed by the addition of 0.7 ml of sodium iodide solution and the DNA purified as described in the silica fines protocol (section 3.3.2.3.).

3.8.3 Glyoxal-agarose electrophoresis of RNA.

RNA samples, denatured by incubation with glyoxal and dimethylsulphoxide, were separated by electrophoresis through 1.5% high gelling-temperature agarose gels, using glyoxalated λ EcoR1 / Hind111 size markers. The method of McMaster and Carmichael (1977) was carried out essentially as described by Maniatis *et al.* (1982) except that after electrophoresis the track containing the size markers was excised and stained with acridine orange. After destaining, the markers were visualised and photographed under UV illumination.

3.8.4. Formamide-agarose electrophoresis of RNA.

Formamide-agarose electrophoresis of RNA was carried out essentially as described by May *et al* (1989). RNA samples were dissolved in 20µl of 60% deionised formamide in 0.1xE buffer (3.6 mM Tris-HCl, 3mM NaH_2PO_4, 2mM EDTA) (Loening, 1969) heated at 65^0C for 5 min, cooled on ice, and 3µl of loading dye (50% glycerol, 0.2% w/v bromophenol blue) was added. Samples were loaded on a large 1.2% agarose gel in 50% formamide, 0.1xE buffer, in which the wells had been previously filled with 60% formamide 0.1xE buffer. The gel was immersed in 0.1xE buffer to just below the top surface of the gel and electrophoresed at 20 mA until the dye had run half way along the gel. The gel was stained by soaking in 1 litre of 2 µg/ml EtBr in water, with gentle shaking in the dark for 30 min, followed by destaining in 1 litre of 1 mM $MgSO_4$ for 30 min. RNA was visualised by UV illumination and the gel photographed as described in section 3.8.1.

3.8.5. Autoradiography of dried agarose gels.

Agarose gels containing radio-labelled nucleic acids were sandwiched between two layers of cellophane on a glass plate. An agarose gel former was then clipped over the gel to keep the cellophane under tension. The gel was heated overnight using a desk lamp placed approximately 15 cm above the gel surface. When completely dry it was autoradiographed in the same manner as filters subjected to DNA hybridisation protocols (section 3.10.1.).

3.9. Immobilisation of DNA for filter hybridisations.

3.9.1. Phage λ plaque and bacterial colony hybridisations.

Transfer of phage from agar plates to nitrocellulose or nylon filters by plaque lifting, followed by DNA denaturation and neutralisation steps were performed as described by Maniatis *et al.* (1982). Positively hybridising clones were purified to homogeneity by sequential rounds of plaque lifting at low phage density.

Bacterial colonies were streaked in duplicate onto nitrocellulose or nylon filters on duplicate plates containing the appropriate antibiotic. After overnight growth the colonies were lysed and the liberated DNA bound and denatured as described by Maniatis *et al* (1982).

3.9.2. Southern blotting of DNA.

DNA restriction fragments separated by agarose gel electrophoresis were

transferred to nitrocellulose or nylon filters according to the method of Southern (1975). High molecular weight restriction fragments were first depurinated by soaking the gel in 1% HCl for 15 min. DNA was then denatured by treating with an excess of 0.5 M NaOH, 1.5 M NaCl three times for 15 min at room temperature with gentle shaking. The gel was neutralised in an excess of 1.5 M NaCl, 0.5 M Tris-HCl pH 7.5, 1 mM EDTA three times for 15 min with gentle shaking at room temperature. Prior to transfer the gel was soaked in 20xSSC for 30 min (1xSSC is 0.15 M NaCl, 0.015 M Na citrate). Transfer to filters was either by the method of Maniatis *et al* (1982) using a capillary blotting apparatus or using a Hybaid vacuum blotter operated as directed by the manufacturer. In both cases the transfer buffer was 20xSSC. After transfer the filter was washed briefly in 5xSSC, allowed to air dry, then baked at 80^0C for 90 min.

3.10. Hybridisation of filter-immobilised DNA to labelled DNA probes.

All hybridisation reactions were carried out in heat sealed polythene bags contained in plastic boxes. These were incubated in water baths with gentle shaking. Hybridisation solutions were equilibriated to the required temperature before use. Solutions contained dilutions of SSC, Denhardt's solution and herring sperm DNA. 1xDenhardt's solution was 0.02% (w/v) each of BSA, polyvinylpyrrolidone and Ficoll 400. Herring sperm DNA was prepared as described by Maniatis *et al.* (1982) before use.

3.10.1. Labelled DNA obtained by random priming

Filters were prehybridised for a minimum of 4 hours at 65^0C in 5xSSC, 5xDenhardt's, 0.1% (w/v) SDS and 200 µg/ml herring sperm DNA (25 ml per filter). Random primed DNA probes were denatured by incubation in a boiling water bath for 5 min followed by rapid cooling on ice. The probe was then added to the hybridisation solution (15 ml total volume) and hybridisations were carried out overnight at 65^0C in 5xSSC, 2xDenhardt's, 0.1% (w/v) SDS and 100 µg/ml herring sperm DNA.

Unless otherwise indicated filters were washed at 65^0C as follows, twice for 15 min in 2xSSC, once for 30 min in 2xSSC, 0.1% (w/v) SDS and twice for 15 min in 1xSSC, 0.1% (w/v) SDS. A final high stringency wash was twice for 15 min in 0.1xSSC, 0.1% (w/v) SDS . Filters were air dried and autoradiographed at -80^0C using preflashed X-ray film and intensifying screens in a film cassette. Modifications to this procedure are given in

the text.

3.10.2. End labelled oligonucleotides.

Hybridisation and washing temperatures for oligonucleotide hybridisations depended upon the calculated approximate melting temperature of the oligonucleotide/DNA duplex.

Filters were washed briefly in 5xSSC at room temperature and then prehybridised for 24 hours at the required temperature in 6xSSC, 10xDenhardt's, 0.5% (w/v) SDS, 0.05% (w/v) sodium pyrophosphate and 100 μg/ml herring sperm DNA (10 ml per filter). Hybridisation was carried out overnight in 5xSSC, 5xDenhardt's, 0.5% (w/v) SDS, 0.05% (w/v) sodium pyrophosphate and 100 μg/ml herring sperm DNA. The concentration of labelled oligonucleotide was no greater than 5 ng/ml to minimise background binding.

Hybridisation and subsequent filter washes were carried out at the same temperature as the prehybridisation stage. Filters were washed once for 30 min in 5xSSC, 5xDenhardt's, 0.1% (w/v) SDS, 0.05% (w/v) sodium pyrophosphate, twice for 30 min in 5xSSC, 0.1% (w/v) SDS and once for 30 min in 3xSSC, 0.1% (w/v) SDS. The required stringency of hybridisation was determined by the temperature of the final wash. Excess washing solution was removed by blotting on 3M paper. Filters were wrapped in cling film before autoradiography as described in section 3.10.1. Filters were kept moist to allow washing at a higher temperature if required.

3.11. Use of the Polymerase Chain Reaction (PCR) for *in vitro* DNA amplification.

3.11.1. Standard PCR conditions.

PCR was carried out essentially as described by Saiki *et al.* (1988). Reactions conditions were as follows; 200 μM each of dATP, dTTP, dGTP and dCTP, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, the concentration of each primer and the amount of template DNA added was dependent upon the type of PCR being carried out. *Taq* polymerase (2.5-5.0 units per reaction) was added last and the reaction mixture overlaid with mineral oil. Reaction volumes were either 50 μl or 100 μl. Initial PCR experiments were carried out using three water baths set at the required temperatures, later PCR amplifications were performed using automatic temperature cycling with a Hybaid "Intelligent Heating Block".

Optimal conditions for each PCR performed were dependent upon the

nature of the template DNA and the type of primers used. Hence the primer and template DNA concentrations, temperature/time conditions for primer annealing, DNA synthesis times at 72^0C and numbers of cycles of amplification are given in the text for specific PCR experiments.

3.11.2. Labelling of PCR products with [α-^{32}P] dCTP by re-amplification.

PCR-amplified DNA was used as a probe in filter hybridisations by re-amplifying the product in the presence of [α-^{32}P]dCTP. Approximately 10 ng of amplification product, previously isolated from a low melting point agarose gel after electrophoresis, was subjected to further rounds of amplification. Reaction conditions were as described for the standard PCR except that the number of reaction cycles was reduced to eight and [α-^{32}P]dCTP was substituted for the dCTP normally used. After completion, the reaction mixture was diluted to 300 µl with column buffer and the labelled product separated on a Sephadex G-50 column as described in section 3.7.1. To check the integrity and size of the product(s) a 1 µl aliquot of the G-50 eluate was diluted to 20 µl with water and analyzed on a 2% (w/v) agarose mini-gel. Following electrophoresis the gel was dried and autoradiographed as described in section 3.8.5.

3.12. *In vitro* synthesis of capped mRNA transcripts.

3.12.1. Production of transcripts for *in vitro* translation.

Transcripts were synthesised using a Pharmacia SP6 Transprobe kit and simultaneously capped by including m^{7}GpppG in the reaction. DNA templates, containing the SP6 promoter and methionine initiation codon, were transcribed as described in the supplied protocol. Transcription reactions contained 0.5 µg of template DNA in a total volume of 12 µl and were incubated at 37^0C for 45 min.

3.12.2. Determination of transcript yield and integrity.

Transcripts were labelled with [α-^{32}P]UTP but were not used for *in vitro* translation due to its short half-life resulting in transcript degradation. Duplicate reactions were carried out with radioactive label (1 µl [25 pmoles / 20 µCi (740 kBq)] of [α-^{32}P]UTP [≈30 TBq/mmol]) added to one reaction whilst the other used unlabelled UTP for subsequent translation *in vitro*.

The amount of transcript produced was determined by measuring the incorporation of [α-^{32}P]UTP into TCA precipitable product. {In some cases

[5,6-^{3}H]UTP (40 Ci/mmol) (1.48 TBq/mmol) was also used as a label in transcription reactions, this was supplied as an aqueous/ethanol solution and aliquots were dried down *in vacuo* and resuspended in water before use.} After incubation a 2 µl aliquot was added to 25 µl of herring sperm carrier DNA. A 5 µl aliquot was then spotted onto a cellulose acetate filter, air dried and the amount of radioactivity present (total counts) determined by liquid scintillation counting.

A second 5 µl aliquot was added to 95 µl of herring sperm DNA followed by 1.0 ml of ice-cold 10% (w/v) TCA. After incubation on ice for 10 min TCA precipitable material was recovered by filtration onto a cellulose acetate filter using a Millipore filtration apparatus. The filter was washed once with ice-cold 10% (w/v) TCA followed by two washes with ice-cold ethanol. The filter was air dried and the amount of radioactivity present (incorporated counts) determined as above.

The percentage incorporation and yield of transcript was determined assuming equal incorporation of each ribonucleotide. The capping solution used contained 0.25 mM UTP (2.5 µl in 12 µl total volume) and this was taken into account when calculating transcript yields.

Transcript size was determined by glyoxal-agarose gel electrophoresis as described in section 3.8.3. using a 1 µl aliqout of the transcription reaction. After electrophoresis the gel was dried and autoradiographed as described in section 3.8.5. Transcripts produced for *in vitro* translations were not purified prior to use.

3.13. *In vitro* translation of transcripts.

3.13.1. Translation using micrococcal nuclease treated lysate.

Rabbit reticulocyte lysate (micrococcal nuclease-treated, message-dependent) (Jackson and Hunt, 1983) was used as directed by the supplier with minor modifications. 10 µl of lysate (containing an unlabelled amino acid pool) was mixed gently with 1 µl of transcription reaction (section 3.12.1.) and 1 µl of L-[4,5-^{3}H]leucine (4.4-7.0 TBq/mmol) and then incubated at 30^{0}C for 30 min. 1 µl (1.16 µg) of pea cotyledon poly (A)$^{+}$ RNA was then added and the reaction incubated for a further 30 min.

3.13.2. Determination of radioactive incorporation into protein.

The amount of labelled leucine incorporated into protein was determined by liquid scintillation counting of TCA precipitated products. 1 µl aliquots

of the translation mix were spotted onto glass fibre GF/C filters, moistened on either side with the addition of 5 µl of bleach solution (0.6 g NaOH, 1.0 ml H_2O_2, 54 ml H_2O). A further 5 µl of bleach was added to either side and the filters allowed to dry. Filters were immersed in ice-cold 10% (w/v) TCA for at least 60 min (10 ml per filter) followed by 5% (w/v) TCA at room temperature (15 min). Filters were then transferred to boiling 5% (w/v) TCA for 10 min followed by two washes in 5% (w/v) TCA (each for 15 min) and then twice with ice-cold ethanol on a Millipore filtration apparatus. After drying, at 80^{0}C in a vacuum oven for 15 min, the amount of radioactivity on each filter was determined using liquid scintillation counting.

3.13.3. Translation using non-nuclease treated lysate.

Non-nuclease treated rabbit reticulocyte lysate, prepared as described by Jackson and Hunt (1983), was used in accordance with the supplied protocol (Dr A. Kaminsky, Department of Biochemistry, University of Cambridge, UK, personal communication). The non-nuclease treated lysate was used for *in vitro* translations in the same manner as nuclease treated lysates.

To 1.0 ml of lysate, containing 20 µM haemin and 50 µg/ml creatine kinase, 6 µl of calf liver tRNA (10 mg/ml) was added and the lysate then stored as 40 µl aliquots at -80^{0}C. 2.5 µl each of supplied 2 M KCl/10 mM $MgCl_2$, amino acid mixture and 0.2 M creatine phosphate was added to 40 µl of lysate. 3 µl of transcription reaction mix was then added and the lysate was incubated at 30^{0}C for 2 hours. Total RNA was extracted and analysed as described in section 3.14.

3.14. Aniline cleavage of depurinated RNA.

The analysis of ribosomal RNA, depurinated by incubating rabbit reticulocyte lysate with ribosome-inactivating proteins, was carried out according to May *et al.* (1989) with minor modifications.

To 50 µl of reticulocyte lysate translation mix (section 3.13.3.) 20 µl of 20% (w/v) SDS and 130 µl of water were added. 100 µl of phenol was added, the sample vortex mixed briefly, followed by 100 µl of chloroform. The sample was centrifuged in an MSE micro-centrifuge for 2 min. The aqueous phase was recovered and total RNA precipitated with the addition of 0.1 volumes 7.0 M ammonium acetate and 2.5 volumes ethanol. After incubation on dry ice for 30 min, the RNA was pelleted by centrifugation for 9 min in a haematocrit micro-centrifuge at 4^{0}C. The pellets were

washed once in 50 mM NaCl, 70% (v/v) ethanol and once in 70% (v/v) ethanol. Residual liquid was removed using a vacuum desiccator and the pellets resuspended in 10 µl of water. RNA samples were stored at -80^{0}C pending further analysis.

1-2 µl of RNA solution (3-4 µg/µl) was mixed with 20 µl of 1.0 M aniline acetate pH 4.5 and incubated in the dark either at room temperature for 3 min or 60^{0}C for 2 min. Samples were immediately cooled on ice and the RNA precipitated with ammonium acetate/ethanol as described previously. After washing and drying the RNA pellet, it was resuspended in 20 µl of 60% formamide, 0.1xE buffer and electrophoresed on formamide-agarose gels as described in section 3.8.4. An equivalent amount of control, non-aniline treated, RNA was added to 20 µl of 60% formamide, 0.1xE buffer and loaded on the gel in the same manner.

3.15. DNA sequence analysis.

Plasmid and M13 DNA sequencing was performed by the dideoxy-sequencing method of Sanger *et al.* (1977) using fluorescent dye-linked universal M13 primers and analysed using an Applied Biosystems 370A DNA sequencer. Plasmid and M13 sequencing templates were used in sequencing reactions as described in the suppliers protocol (Model 370A DNA Sequencing system, Users Manual Version 1.3A, Oct 1988).

The deduced amino acid sequences obtained by translating DNA sequences in all three reading frames was achieved using the computer programme "ISEQTR" written by Dr J.A. Gatehouse (Department of Biological Sciences, Durham University, UK).

Further analysis of DNA and deduced amino acid sequences was carried out using the MicroGenie sequence analysis software (Version 6) written by Queen and Korn (1984) and supplied by Beckman Instruments, Inc, Paolo Alto, California.

The analysis of oligonucleotides for predicted dissociation temperatures under specified ionic conditions and hybridisation temperatures was carried out using the "OLIGO" computer programme (Version 3.3) developed by Rychlik and Rhoads (1989). This programme was also used to predict both oligonucleotide self- and inter-complementarity when designing primers for use in the polymerase chain reaction.

Supplementary information to methods section.

3.3.3.2. Qiagen tips used for purification of plasmid DNA.

Cells were harvested from 50 ml of overnight culture by centrifugation at 10,000 rpm for 15 min at 4^0C. The pellet was resuspended in 3.6 ml of ice cold 50 mM Tris-HCl, pH 7.4 followed by the addition of 2.0 ml of 20 mg/ml lysozyme (Sigma grade 1 / freshly prepared) in 50 mM Tris-HCl and incubation on ice for 10 min. 1.0 ml of 0.5 M EDTA, pH 8.0 was added followed by a further 10 min incubation on ice. THe mix was centrifuged at 15,000 g for for 45 min at 4^0C and the supernatant treated with RNase A for 30 min at 37^0C at a final concentration of 20 μg/ml. Proteinase K (freshly prepared) was added to a final concentration of 10 μg/ml and incubated at 37^0C for 30 min followed by centrifugation at 15,000 rpm for 10 min. 4.5 ml of the supernatant was transfered to a new tube and 1.0 ml of 5.0 M NaCl and 0.5 ml of 1.0 M MOPS, pH 7.0 was added.

A Qiagen tip 100 was equilibriated with 2.0 ml of buffer A followed by adsorbtion of the plasmid DNA sample to the tip matrix. The tip was washed with 5.0 ml of buffer C and the DNA eluted in 2.0 ml of buffer F. Plasmid DNA was precipitated with the addition of 0.8 volumes of isopropanol and incubation on ice for 30 min. DNA was collected by centrifugation and the pellet washed twice in 70 % ethanol.

Buffer A. 400 mM NaCl, 50 mM MOPS, 15 % ethanol, pH 7.0.
Buffer C. 1000 mM NaCl, 50 mM MOPS, 15 % ethanol, pH 7.0.
Buffer F. 1500 mM NaCl, 50 mM MOPS, 15 % ethanol, pH 7.5.

3.12. *In vitro* synthesis of capped mRNA transcripts.

3.12.1. Production of transcripts for *in vitro* translation.

In vitro transcription reactions (total volume 12 μl) contained the following reagents:

Reagent buffer	2.5 μl.	Composition not given by Pharmacia.
[α-^{32}P]UTP, 400-800 Ci/mmol	1 μl.	
RNAguard	1 μl.	Ribonuclease inhibitor.
m^7GpppG cappining solution	2.5 μl.	Contains unlabelled UTP.
DNA template	1 μl.	(0.5 μg).

RNase free water to 12 μl. (DEPC treated).

Approximately 30-90 % of the RNA generated should be capped.

4. Results And Discussion. Isolation Of Ricin Gene Sequences.

4.1. Strategy for screening a *Ricinus communis* genomic library using pRCL59, a partial *Ricinus* agglutinin cDNA.

The strategy employed to screen a *Ricinus* genomic library for ricin and agglutinin genes used an agglutinin cDNA sequence as a hybridisation probe to detect genomic sequences.

The ricin and agglutinin cDNA sequences presented and compared by Roberts *et al.* (1985) show that the sequences are extremely closely related. Alignment of the two coding sequences shows 93% similarity at the nucleotide level. Either sequence used as a probe in DNA-DNA hybridisation reactions would be expected to cross-hybridise with the other related sequence. The cDNA clone pRCL59 (Roberts *et al.*, 1985) is a truncated agglutinin cDNA cloned in the PstI site of pBR322. Fig 4.1 shows its length and position compared with both agglutinin and ricin genes, selected restriction sites are also shown. The cDNA was cloned by adding poly (dC) to the 3' ends of the cDNA followed by ligation into PstI restricted pBR322 which had been poly (dG)-tailed. The cDNA insert is excisable as a PstI restriction fragment, the two internal BamHI sites common to agglutinin and pRCL59 can also be used to produce a 700 bp restriction fragment. The ricin and agglutinin coding sequences have several restriction site differences, indicated in Fig 4.1, restriction mapping of genomic clones would allow the clones to be identified as ricin- or agglutinin-like on the basis of these characteristic restriction differences.

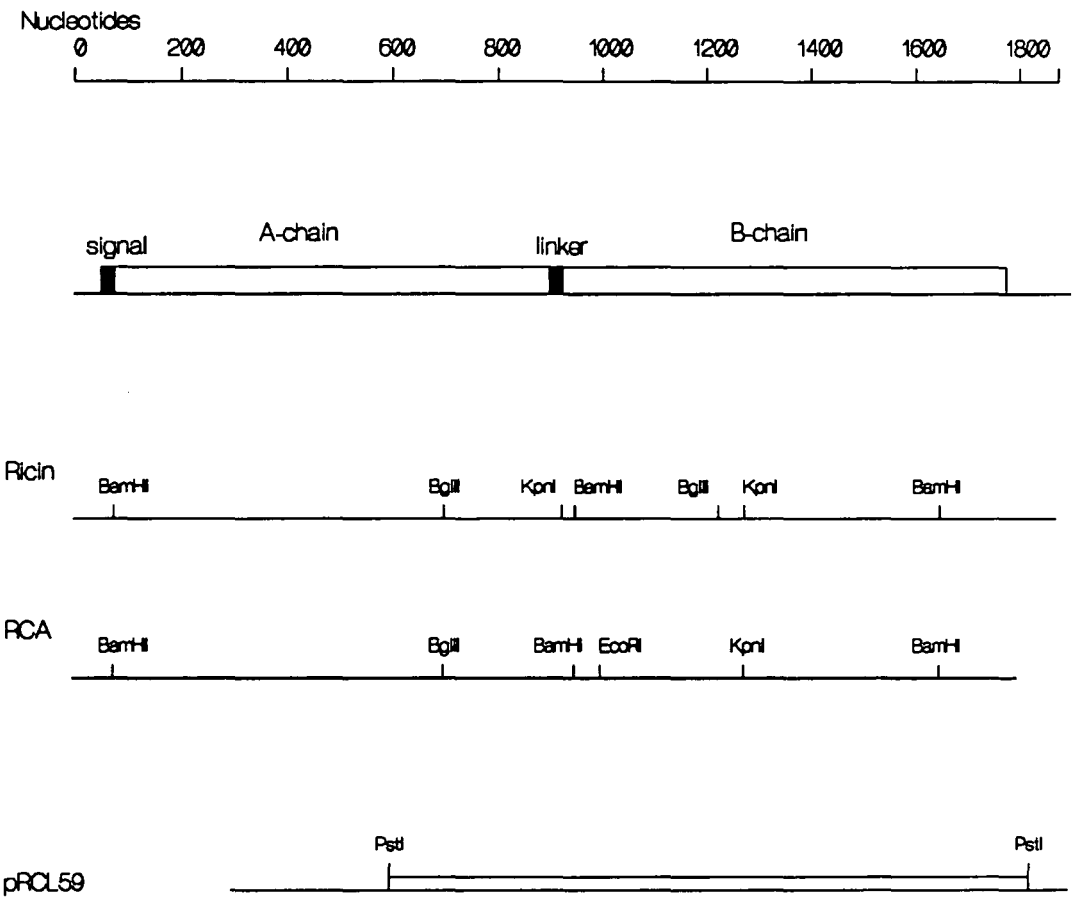
4.2. The use of pRCL59 as a probe for ricin and agglutinin genomic clones.

Southern analysis of *Ricinus* genomic DNA was performed as shown in Fig 4.2 using the PstI insert of pRCL59 as a probe. The washing stringency was initially to 1.0xSSC at 65^0C followed by a wash of 0.1xSSC at 65^0C. After each wash the filter was autoradiographed.

Under the washing stringencies employed, 1.0xSSC and 0.1xSSC at 65^0C, the probe would be expected to remain hybridising to sequences of at least 90% and 98% similarity during each respective wash. These values are an approximate indication of the washing stringencies and are derived from the formula predicting the melting temperature (T_M) of a DNA/DNA hybrid.

Figure 4.1. Comparison of selected restriction sites in the published cDNA sequences of *Ricinus communis* agglutinin and ricin.

Restriction sites present in ricin and agglutinin are shown in relation to the gene coding sequence. The alignment and extent of pRCL59, a partial agglutinin cDNA, is also shown. Sequence information was taken from Roberts *et al.* (1985) (agglutinin) and Lamb *et al.* (1985) (ricin). It should be noted that ricin lacks both EcoRI and HindIII sites, whilst agglutinin has a single EcoRI site.


Nucleotides
0
200
400
600
800
1000
1200
1400
1600
1800
signal
A-chain
linker
B-chain
Ricin
BamHI
BglII
KpnI
BamHI
BglII
KpnI
BamHI
RCA
BamHI
BglII
BamHI
EcoRI
KpnI
BamHI
pRCL59
PstI
PstI

T_M is the temperature at which 50% dissociation occurs. The empirical equation, cited by Meinkoth and Wahl (1984), used to predict DNA/DNA hybrid stabilities under differing salt and temperature conditions is as follows:-

$$T_M = 81.5^0C + 16.6 \log M + 0.41(\% G + C) - 500/n - 0.61(\% \text{ formamide})$$

M = ionic strength in mol/litre, (% G + C) = base composition, n = length of the shortest probe/target duplex.

Assuming 40% (G + C) content, the percentage mismatch allowed in a DNA duplex can be calculated (L. Gatehouse, Department of Biological Sciences, University of Durham, personal communication). The percentage values quoted for washes of 1.0xSSC and 0.1xSSC at 65^0C, 90% and 98%, should only be regarded as an indication of the relative degree of washing stringency. This is due to a number of factors: variation in probe base composition from 40% (G + C) and the potential for a non-random distribution of base mis-matches. This would allow some duplexes to be virtually 100% similar whilst others would contain many mismatches, due to the generation of a population of radioactively labelled DNA fragments of varying sizes obtained by random priming of probe DNA.

The autoradiographs shown in Fig 4.2 show that a number of hybridising restriction fragments are present, even after a high stringency wash. Several fragments retain the same relative intensity of hybridisation under both washing conditions, whereas other hybridising fragments are either reduced in relative intensity of hybridisation or not detected at all after washing in 0.1xSSC.

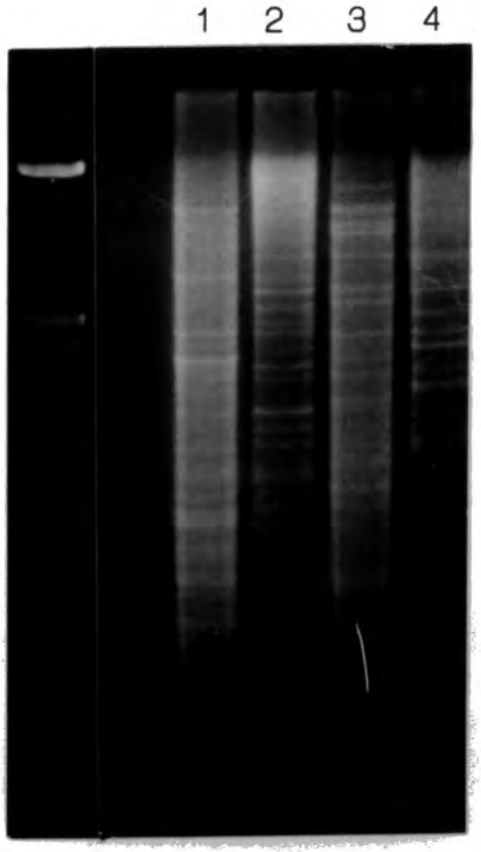
It is possible that the manner of washing, a low stringency wash followed by autoradiography, then a further high stringency wash, may not produce the same result as a duplicate experiment with one blot washed to low stringency and the other to high stringency. Some probe may bind irreversibly to the filter between the two washes in the Southern analysis shown. However the autoradiographs do show that the probe can be differentially removed from the hybridising fragments, indicating that the probe had not bound irreversibly after the first wash in 1.0xSSC. Under the lower stringency hybridisation conditions employed it is assumed that the agglutinin probe has hybridised to both ricin and agglutinin containing restriction fragments.

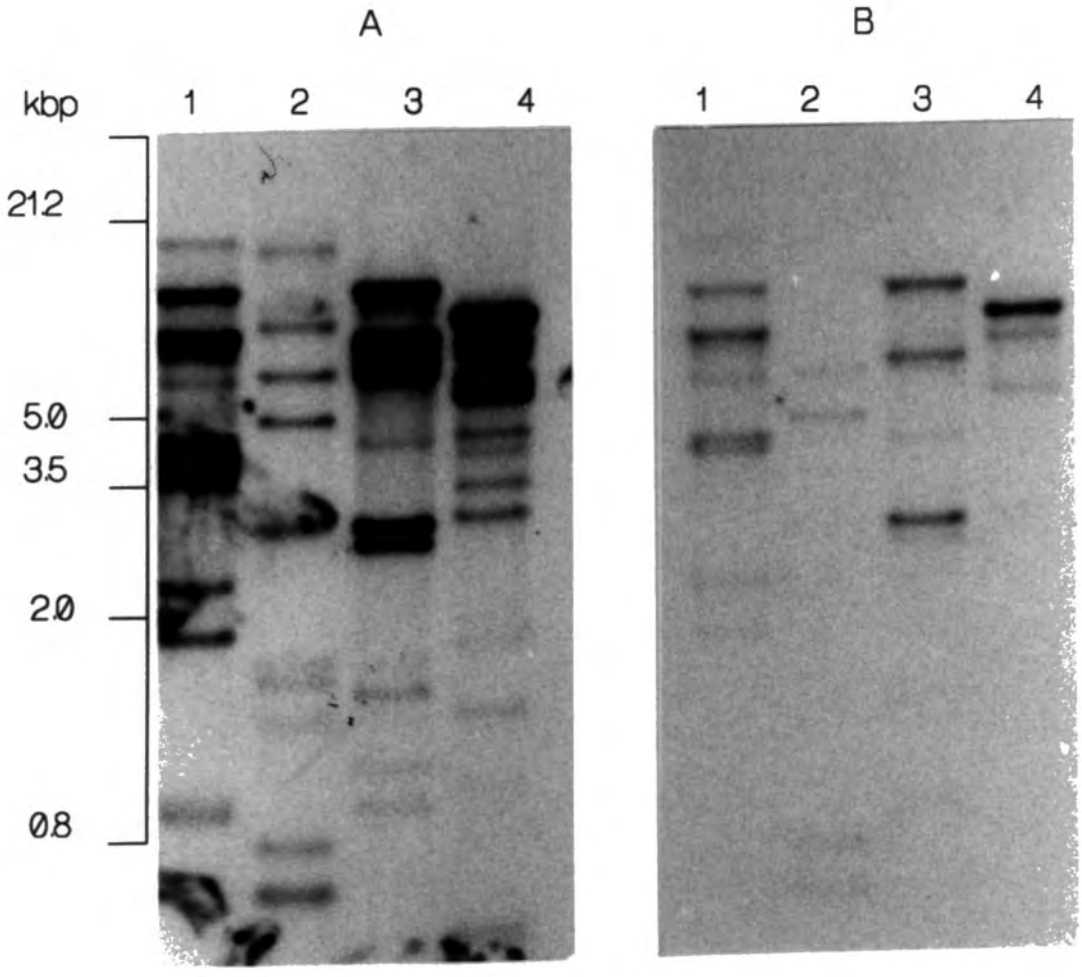
Figure 4.2. Southern blot analysis of *Ricinus communis* genomic DNA hybridised with *Ricinus communis* agglutinin cDNA, pRCL59.

Ricinus communis genomic DNA (10μg) was restricted with:-

1. EcoRI
2. BamHI
3. HindIII
4. BglII

Restriction fragments were separated by electrophoresis on a 0.7% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific fragments were detected by hybridisation with the PstI insert of pRCL59. The filter was washed twice for 30 min each at a final stringency of 1.0×SSC, 65^0C and autoradiographed whilst still moist (autoradiograph A). The filter was then washed twice for 30 min each at a stringency 0.1×SSC, 65^0C and autoradiographed (autoradiograph B). DNA size markers were λ EcoRI/HindIII.


1
2
3
4


A
B
kbp
1
2
3
4
21.2
5.0
3.5
2.0
0.8
1
2
3
4

This confirms the result of Halling *et al.* (1985) who showed by Southern analysis that ricin is a member of a multi-gene family comprising of approximately six genes. The decrease in relative hybridisation intensity observed for several restriction fragments indicates that some of the sequences are divergent from the agglutinin sequence as no hybridisation is observed after washing in 0.1xSSC. This presence of a multigene family correlates with the presence of at least two expressed ricin genes and one agglutinin gene. cDNA clones for ricin D and ricin E (Ladin *et al.*, 1987) and agglutinin (Roberts *et al.*, 1985) have been isolated. The agglutinin cDNA clones exhibited restriction site heterogeneity suggesting that they were derived from mRNAs transcribed from distinct genes.

The Southern analysis indicated that pRCL59 could be used as a heterologous probe for ricin genomic clones, provided that the stringency of post-hybridisation washes was low enough to allow the maintenance of stable hybrids between agglutinin cDNA and ricin genomic sequences.

An amplified *Ricinus* genomic library (constructed as described in section 3.4) was plated on large plates, see section 3.4.6, using the host strain *E. coli* P2392. Approximately $6x10^5$ recombinant phage λ clones were plated out, resultant plaques were transferred to duplicate nitrocellulose filters (section 3.9.1) and hybridised with the 700 bp BamHI fragment of pRCL59, radioactively labelled by random priming. The filters were washed to a final stringency of 1.0xSSC at 65^0C and autoradiographed. Positively hybridising clones, coincidental on the duplicate filter autoradiographs, were then isolated.

The positive clones were divided into three arbitrary classes on the basis of the relative intensity of hybridisation observed. 16 clones showed strong-, 10 medium- and 30 weak hybridisation to the probe. The large number of hybridising clones obtained is consistent with the existence of a multigene family but also reflects the differential growth of λ clones in the amplification step used to store the library. This can result in the over-representation of faster growing clones in the amplified library.

Four clones from each class were subjected to further rounds of plaque purification using the 700 bp BamHI fragment of pRCL59 as a hybridisation probe. Standard hybridisation conditions were used with a final high stringency post-hybridisation wash of 1.0xSSC at 65^0C. Due to the large number of positive clones obtained a preliminary characterisation of 2

clones was performed. The 2 clones, designated λRic 1ii (strong hybridisation) and λRic 3ix (medium hybridisation) were purified to homogeneity and DNA was prepared from them as described in section 3.3.3.1.

4.3. Characterisation of two λ clones positively hybridising with pRCL59.

4.3.1. Restriction analysis of λ clones λRic 1ii and λRic 3ix.

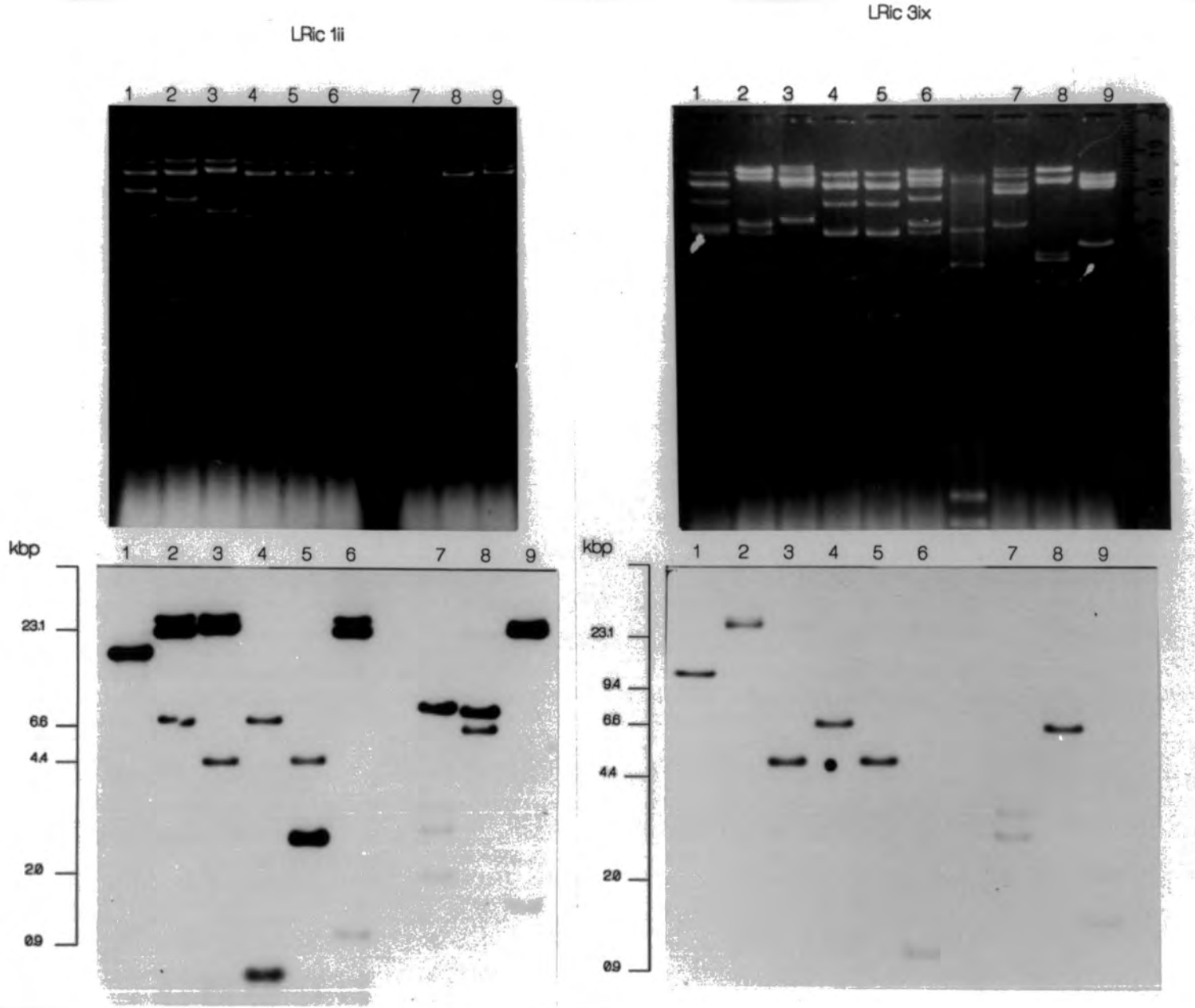
The two λ clones λRic 1ii and λRic 3ix were restricted with various restriction enzymes as shown in Fig 4.3. Restriction fragments were hybridised with the 700 bp BamHI fragment of pRCL59 after Southern transfer. The autoradiographs indicate that the two clones have three hybridising restriction fragments in common. These are a 4.5 kb EcoRI fragment, a 6.6 kb BamHI fragment and a 6.5 kb BglII fragment. The 6.6 kb BamHI fragment is present in the Southern analysis of genomic DNA (Fig 4.2). Clone λRic 1ii also contains a second more strongly hybridising region which is close to the left arm of the EMBL3 vector. BamHI restriction of λRic 1ii (lane 2) shows that the strongly hybridising sequence is not released from the left arm of the vector. Restriction with BamHI and SalI (site within the vector) releases a small, approximately 600 bp, hybridising fragment (lane 4), this is of reduced intensity due to the lower efficiency of binding of small DNA fragments to nitrocellulose. (The strongly hybridising fragments of the highest molecular weight is an artifact produced by the annealing together of the cos sites present at the termini of the arms of the vector). If either of these two clones had contained previously characterised full-length ricin or agglutinin sequences then it would have been expected to detect a hybridising 700 bp BamHI restriction fragment. This is not observed for either clone. The agglutinin cDNA sequence also contains an EcoRI site, this is absent in the ricin cDNA sequence, (Fig 4.1) this is also not present in clone λRic 1ii at the region of strong hybridisation, hence it is unlikely that this region contains a full-length gene. The restriction site heterogeneity previously noted for agglutinin cDNA clones could account for the loss of a characteristic restriction site within an agglutinin gene. The close proximity of the sequence to the left arm of the vector which strongly hybridised to the probe also indicates that the sequence is unlikely to represent a full-length and complete ricin or

Figure 4.3. Restriction analysis of λRic 1ii and λRic 3ix. λ clones hybridising with *Ricinus* agglutinin.

λ DNA was isolated from 2 clones hybridising with the 700 bp BamHI insert of pRCL59 and subjected to restriction analysis. DNA from each clone was restricted with the following enzymes:-

1. SalI	4. BamHI/SalI	7. KpnI
2. BamHI	5. EcoRI/SalI	8. BglII
3. EcoRI	6. EcoRI/BamHI	9. HindIII

Restriction fragments were separated by electrophoresis on 0.6% (w/v) agarose gels and transferred to nitrocellulose filters. Specific fragments were detected by hybridisation with the 700 bp BamHI fragment of pRCL59. Filters were washed to a final stringency of 1.0×SSC, 65^{0}C for 30 min and autoradiographed. DNA size markers were λ/HindIII and pBR322/AluI.


LRic 1ii
LRic 3ix
1 2 3 4 5 6 7 8 9
kbp
23.1
9.4
6.6
4.4
2.0
0.9

agglutinin gene. This strongly hybridising sequence is most likely to be a truncated ricin- or agglutinin-like sequence arising due to the preparation of genomic DNA during the construction of the library. Partial Sau3A restriction of the genomic DNA has led to cleavage at a site within the strongly hybridising sequence. This produces a truncated sequence upon subsequent ligation into the phage vector arms.

To further investigate the identity of clone λRic 1ii part of the weakly hybridising region was sequenced. The restriction maps of the two clones are shown in Fig 4.4 along with the subcloning strategy employed to sequence the hybridising region present in the 6.6 kb BamHI fragment of λRic 1ii. The regions showing either strong or weak hybridisation to the probe are also indicated. It should be noted that whilst the two clones share several restriction enzyme sites in common. However, the KpnI site lies at a different location in each clone and λRic 1ii contains two extra EcoRI sites in the region where several sites can be aligned. On the basis of this restriction mapping the two clones are related, but independent clones, which do not overlap.

4.3.2. Partial sequence analysis of λ clone λRic 1ii.

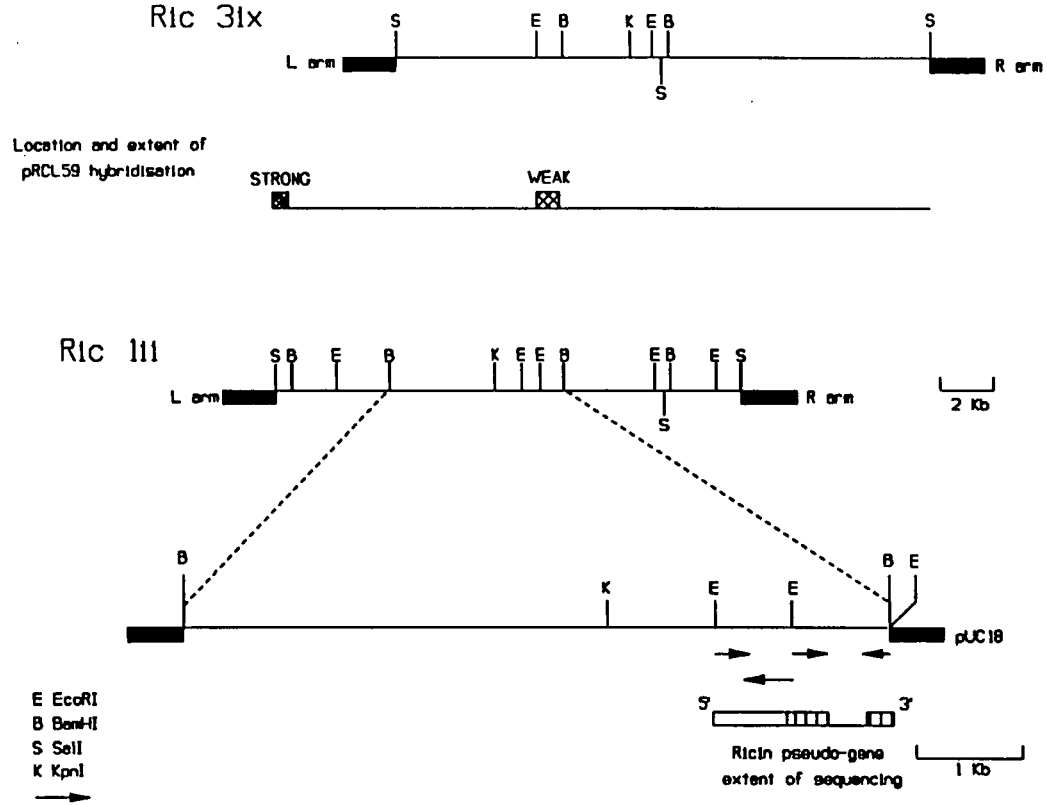
The strongly hybridising region close to the left arm of the clone was not chosen for further investigation. This was due to both the lack of characteristic restriction sites present in the sequence and its small length.

The extent of sequencing carried out on the 6.6 kb BamHI fragment is shown in Fig 4.4. The fragment was subcloned into pUC18 followed by the production of M13 subclones as shown. The EcoRI fragment was completely sequenced and designated ϕRic1. Sequencing of the EcoRI/BamHI fragment produced two sequences which did not overlap. The sequence from the EcoRI site was designated ϕRic2 and that from the BamHI site ϕRic3.

4.3.3. The genomic clone λRic 1ii contains a badly damaged ricin-like pseudogene.

The three sequences obtained from the BamHI subclone were each aligned, by matrix comparison, with the published ricin sequence (Lamb *et al.*, 1985). The result of this analysis is shown in Fig 4.5A. The matrix comparison shows the regions of homology with the ricin coding sequence and above this the diagram shows the sequenced regions aligned with a schematic

Figure 4.4. Restriction maps of clones λRic 1ii and λ Ric 3ix.
The partial restriction maps of the two λ clones isolated using pRCL59 as a probe are shown. Regions hybridising with the 700 bp BamHI fragment of pRCL59 are indicated. The map of each clone is aligned to show restriction sites in common. The 6.6 kb BamHI fragment of λRic 1ii was sub-cloned into pUC18, the direction and extent of sequencing is indicated.


Ric 31x
S
E B
K E B
S
L arm
R arm
S
Location and extent of
pRCL59 hybridisation
STRONG
WEAK
Ric 111
S B
E
B
K E E B
E B
E
S
L arm
R arm
S
2 Kb
B
B E
K
E
E
pUC18
E EcoRI
B BamHI
S SalI
K KpnI
Direction of sequencing
5'
3'
Ricin pseudo-gene
extent of sequencing
1 Kb
No sequence homology with ricin
Sequence homology with ricin

representation of the ricin/agglutinin coding sequence. It should be noted that the first 590 bp of the ϕRic1 sequence show no homology to ricin (no line present in the matrix comparison).

Fig 4.5B shows the alignment of the ϕRic sequences with the ricin cDNA sequence. The first 590 bp of ϕRic1 have been excluded from the alignment as they show no significant homology with ricin. The alignment shows that although the ϕRic sequences are related to the ricin coding sequence the clone does not contain either a ricin or an agglutinin gene.

When the sequences are translated in all 3 reading frames regions of amino acid homology to the ricin primary sequence are apparent (Fig 4.5C). Clone ϕRic1 has an open reading frame starting at nucleotide 601 with homology to ricin (and agglutinin), however there is an in-frame stop codon at nucleotide positions 678-680. The deduced amino acid sequences of clones ϕRic2 and ϕRic3 show limited regions of homology to the ricin amino acid sequence and, significantly no continuous open reading frame is present due to frameshift mutations.

On the basis of the sequences presented it is concluded that the 6.6 kb BamHI fragment obtained from λRic 1ii contains a sequence that exhibits similarity to ricin and agglutinin genes. This sequence does not encode a functional ricin or agglutinin protein and appears to be a badly damaged ricin- or agglutinin-like pseudogene. The intensity of hybridisation of the 6.6 kb BamHI restriction fragment of clones λRic 1ii and λRic 3ix, shown in Fig 4.3, is relatively weak compared with the much stronger intensity of hybridisation of the sequence adjacent to the left arm of clone λRic 1ii. This is consistent with the difference in sequence between ricin and the regions of the pseudogene sequenced. A lower degree of hybridisation to the pRCL59 probe would be expected for this divergent sequence.

The two lambda clones isolated are distinct clones both containing a weakly hybridising 6.6 kb BamHI fragment, this suggests that there are at least two divergent ricin-like sequences in the *Ricinus* genome. Eukaryotic genomes contain many pseudogenes which have arisen either by a gene duplication event, as in the globin gene cluster (Little PFR, 1982), or by the reintegration into the genome of a cDNA transcript of an RNA, as in the immunoglobulin gene family (Hollis *et al.*, 1982). Once a duplication event occurs the duplicated gene can then mutate randomly. The pseudogene sequenced has diverged significantly from ricin and

Figure 4.5. Sequence analysis of a ricin-like pseudogene present in clone λ Ric 1ii.

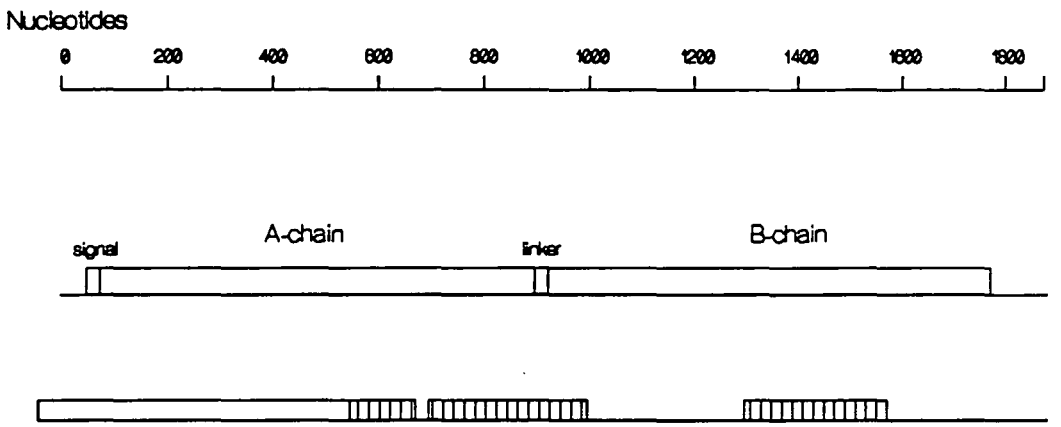
The 3 sequences obtained from the sequencing of λRic 1ii, designated ϕRic1, ϕRic2 and ϕRic3, were compared with ricin as shown in Fig 4.5 (A). Sequences showing homology with ricin were located using the Microgenie matrix comparison programme. The matrix output for each clone is shown in Fig. 4.5 (A). It should be noted that the 5′ sequence of ϕRic1 (the first 590 bp) bears no significant similarity to ricin and is excluded from the sequence alignment shown in Fig. 4.5 (B). Fig 4.5 (B) shows the alignment of sequences with ricin. The sequence of ricin from nucleotide position 560 to 1540 is shown, the numbering of the ricin sequence is as shown in Fig 4.5 (A).

..... sequence not determined

———— gaps introduced to maximise homology

The deduced amino acid sequence of each clone, in all three reading frames, is shown in Fig 4.5 (C) the sequence numbering refers to each clone, not ricin. Underlining of the sequence indicates the amino acids identical to those in ricin.

Fig 4.5A



Boxes indicate regions sequenced

Sequence homologous to ricin/RCA

Sequence showing no significant similarity to ricin/RCA

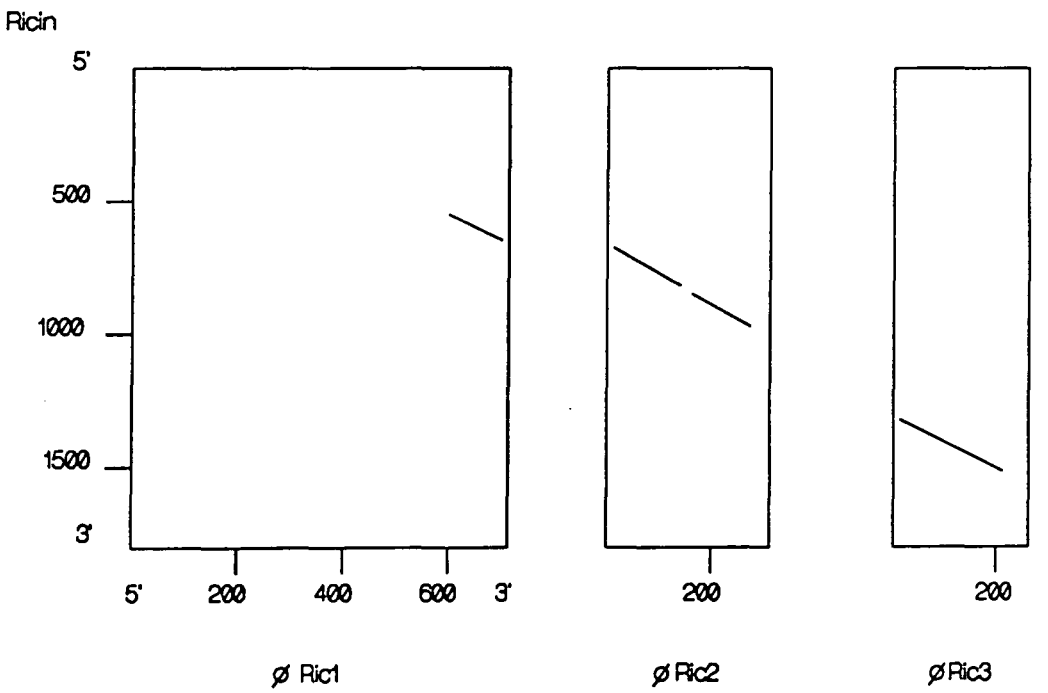
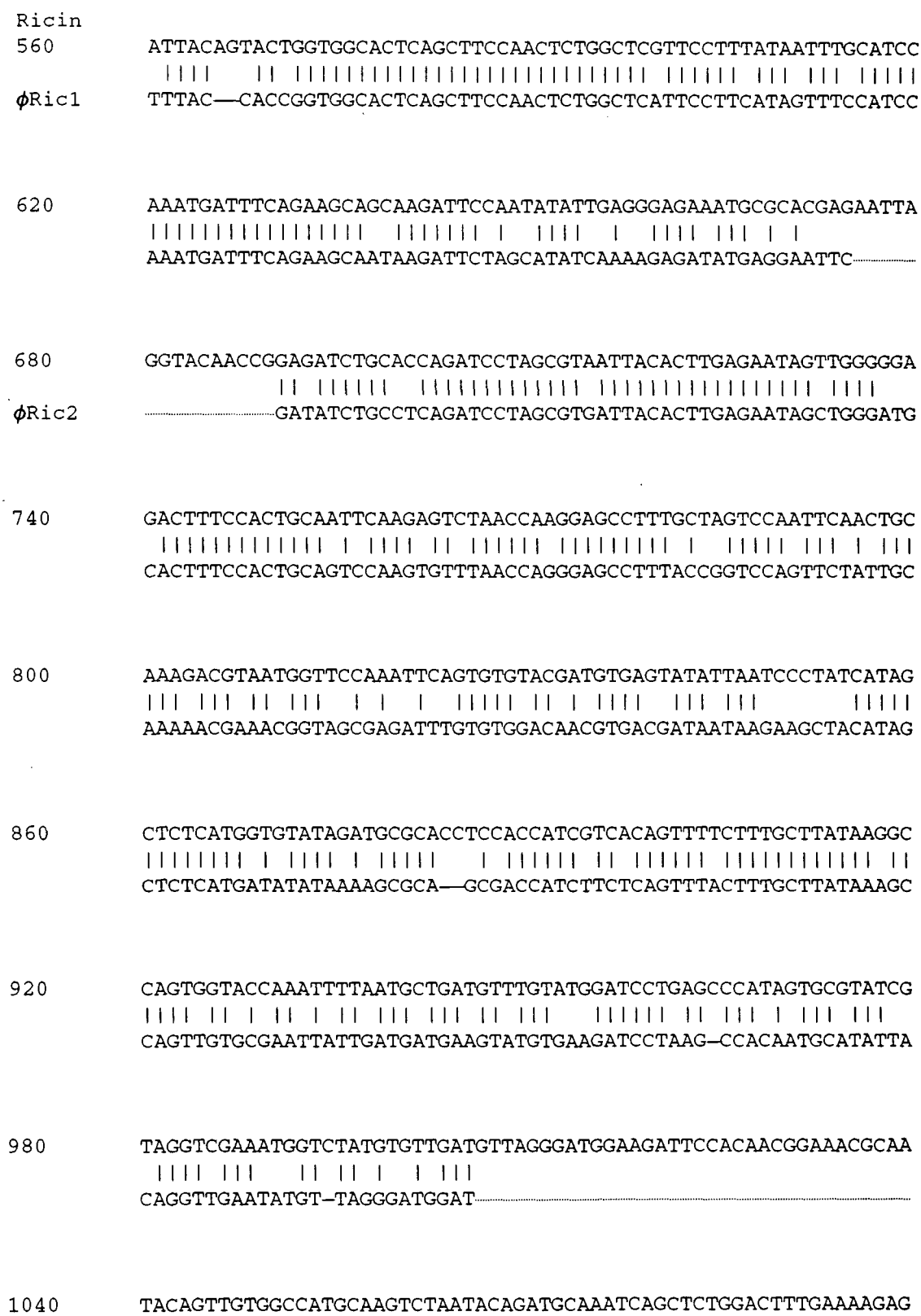


Fig. 4.5 (B). Alignment of ricin pseudo-gene sequences ϕRic1, 2 and 3 with ricin cDNA sequence.



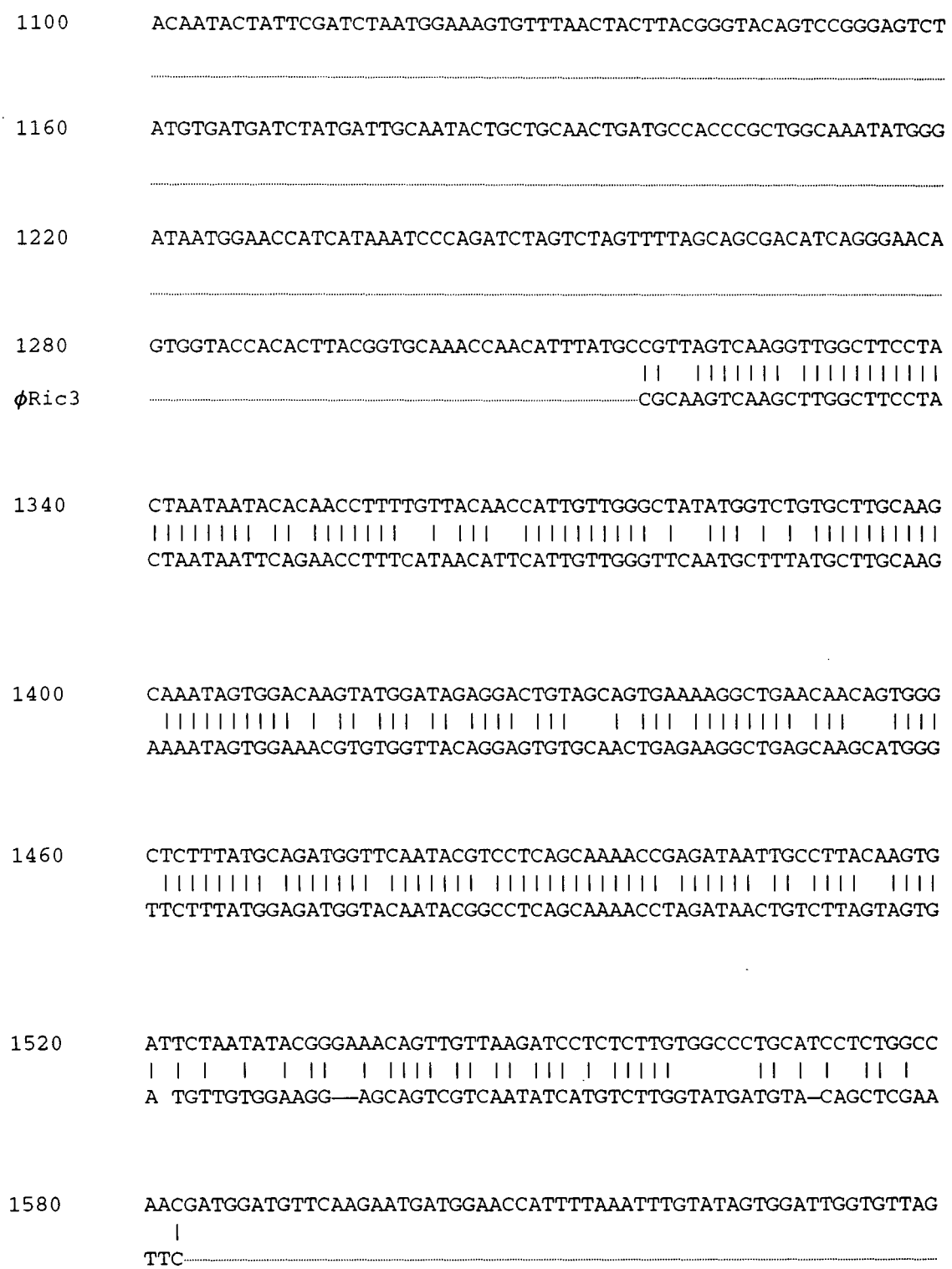

1100 ACAATACTATTCGATCTAATGGAAAGTGTTTAACTACTTACGGGTACAGTCCGGGAGTCT
1160 ATGTGATGATCTATGATTGCAATACTGCTGCAACTGATGCCACCCGCTGGCAAATATGGG
1220 ATAATGGAACCATCATAAATCCCAGATCTAGTCTAGTTTTAGCAGCGACATCAGGGAACA
1280 GTGGTACCACACTTACGGTGCAAACCAACATTTATGCCGTTAGTCAAGGTTGGCTTCCTA
ϕRic3 CGCAAGTCAAGCTTGGCTTCCTA
1340 CTAATAATACACAACCTTTTGTTACAACCATTGTTGGGCTATATGGTCTGTGCTTGCAAG
CTAATAATTCAGAACCTTTCATAACATTCATTGTTGGGTTCAATGCTTTATGCTTGCAAG
1400 CAAATAGTGGACAAGTATGGATAGAGGACTGTAGCAGTGAAAAGGCTGAACAACAGTGGG
AAAATAGTGGAAACGTGTGGTTACAGGAGTGTGCAACTGAGAAGGCTGAGCAAGCATGGG
1460 CTCTTTATGCAGATGGTTCAATACGTCCTCAGCAAAACCGAGATAATTGCCTTACAAGTG
TTCTTTATGGAGATGGTACAATACGGCCTCAGCAAAACCTAGATAACTGTCTTAGTAGTG
1520 ATTCTAATATACGGGAAACAGTTGTTAAGATCCTCTCTTGTGGCCCTGCATCCTCTGGCC
A TGTTGTGGAAGG--AGCAGTCGTCAATATCATGTCTTGGTATGATGTA-CAGCTCGAA
1580 AACGATGGATGTTCAAGAATGATGGAACCATTTTAAATTTGTATAGTGGATTGGTGTTAG
TTC

Fig. 4.5 (C). Translation in all reading frames of ricin-like pseudo-gene.

ϕRic1. Homology to ricin amino acid sequence starts at position 600.

```
        10         20         30         40         50         60
GAATTCACCCAACTCTATCAGTTCTCAATGTCTTTAAAGAACATTCACTCTTATTTTCAA

        70         80         90        100        110        120
CCAACAATTTGAATACTTCGAAAGCTTCAGATTTCTGCTTCAAAAAATAAAACCATATTT

       130        140        150        160        170        180
TGCGACTGAAATCATTAATAGACAGTAAGAAATCTTTGTTCTTACTAAAAGAAATTGGAT


       190        200        210        220        230        240
TAATTGGACCACATATGTCAGCGTGGACAAGTTCAAGAGGCTTCGATGCTCTTGAAGATG

       250        260        270        280        290        300
ATTCCTACGGAAAACTGTGTCTGAATTGCTTTCCGTACATGCATCCTTCACACAACTGAT

       310        320        330        340        350        360
AAGGATGATCAAAGAAGGTAGACCATGAACCATTCTCCCTTGAGCCAATTGCTTCAAGCT

       370        380        390        400        410        420
GTTGAAATTCAAATGACCCAGTCCTTTGTAGCCAAAGCCAGGATGAATCAGTAGCACTTG

       430        440        450        460        470        480
CCTTTAAGCACTTCATATAATCAGACTGAAGTTTGAGAATAAATATTATGTTTTTAGTCA


       490        500        510        520        530        540
TCTCAACCTTAGCAATTAGATTACCTTTGTTGTCTTTCATCACCAATACAGCATTCTTCA

       550        560        570        580        590        600
AGTAGATCTCGAATTGTTTCTCTAAAAGTTGGCCTAAACTCAACACATTCACTTTTACCA

       610        620        630        640        650        660
CCGGTGGCACTCAGCTTCCAACTCTGGCTCATTCCTTCATAGTTTCCATCCAAATGATTT
 P  V  A  L  S  F  Q  L  W  L  I  P  S  U  F  P  S  K  U  F
   R  W  H  S  A  S  N  S  G  S  F  L  H  S  F  H  P  N  D  F
T  G  G  T  Q  L  P  T  L  A  H  S  F  I  V  S  I  Q  M  I


       670        680        690        700
CAGAAGCAATAAGATTCTAGCATATCAAAAGAGATATGAGGAATTC
 Q  K  Q  U  D  S  S  I  S  K  E  I  U  G  I
   R  S  N  K  I  L  A  Y  Q  K  R  Y  E  E  F
S  E  A  I  R  F  U  H  I  K  R  D  M  R  N
```

ϕRic2

```
        10         20         30         40         50         60
GATATCTGCCTCAGATCCTAGCGTGATTACACTTGAGAATAGCTGGGATGCACTTTCCAC
 D  I  C  L  R  S  U  R  D  Y  T  U  E  U  L  G  C  T  F  H

  I  S  A  S  D  P  S  V  I  T  L  E  N  S  W  D  A  L  S  T

   Y  L  P  Q  I  L  A  U  L  H  L  R  I  A  G  M  H  F  P


        70         80         90        100        110        120
TGCAGTCCAAGTGTTTAACCAGGGAGCCTTTACCGGTCCAGTTCTATTGCAAAAACGAAA
 C  S  P  S  V  U  P  G  S  L  Y  R  S  S  S  I  A  K  T  K

  A  V  Q  V  F  N  Q  G  A  F  T  G  P  V  L  L  Q  K  R  N

L  Q  S  K  C  L  T  R  E  P  L  P  V  Q  F  Y  C  K  N  E


       130        140        150        160        170        180
CGGTAGCGAGATTTGTGTGGACAACGTGACGATAATAAGAAGCTACATAGCTCTCATGAT
 R  U  R  D  L  C  G  Q  R  D  D  N  K  K  L  H  S  S  H  D

  G  S  E  I  C  V  D  N  V  T  I  I  R  S  Y  I  A  L  M  I

T  V  A  R  F  V  W  T  T  U  R  U  U  E  A  T  U  L  S  U


       190        200        210        220        230        240
ATATAAAAGCGCAGCGACCATCTTCTCAGTTTACTTTGCTTATAAAGCCAGTTGTGCGAA
 I  U  K  R  S  D  H  L  L  S  L  L  C  L  U  S  Q  L  C  E

  Y  K  S  A  A  T  I  F  S  V  Y  F  A  Y  K  A  S  C  A  N

Y  I  K  A  Q  R  P  S  S  Q  F  T  L  L  I  K  P  V  V  R


       250        260        270        280        290        300
TTATTGATGATGAAGTATGTGAAGATCCTAAGCCACAATGCATATTACAGGTTGAATATG
 L  L  M  M  K  Y  V  K  I  L  S  H  N  A  Y  Y  R  L  N  M

  Y  U  U  U  S  M  U  R  S  U  A  T  M  H  I  T  G  U  I  C

I  I  D  D  E  V  C  E  D  P  K  P  Q  C  I  L  Q  V  E  Y


       310
TTAGGGATGGAT
 L  G  M  D

  U  G  W

V  R  D  G
```

ϕRic3.

```
    10        20        30        40        50        60
CGCAAGTCAAGCTTGGCTTCCTACTAATAATTCAGAACCTTTCATAACATTCATTGTTGG
 R K S S L A S Y U U F R T F H N I H C W

  A S Q A W L P T N N S E P F I T F I V G

   Q V K L G F L L I I Q N L S U H S L L
       70        80        90       100       110       120

GTTCAATGCTTTATGCTTGCAAGAAAATAGTGGAAACGTGTGGTTACAGGAGTGTGCAAC
 V Q C F M L A R K U W K R V V T G V C N

  F N A L C L Q E N S G N V W L Q E C A T

G S M L Y A C K K I V E T C G Y R S V Q

       130       140       150       160       170       180
TGAGAAGGCTGAGCAAGCATGGGTTCTTTATGGAGATGGTACAATACGGCCTCAGCAAAA
 U E G U A S M G S L W R W Y N T A S A K

  E K A E Q A W V L Y G D G T I R P Q Q N

L R R L S K H G F F M E M V Q Y G L S K

       190       200       210       220       230       240
CCTAGATAACTGTCTTAGTAGTGATGTTGTGGAAGGAGCAGTCGTCAATATCATGTCTTG
 P R U L S U U U C C G R S S R Q Y H V L

  L D N C L S S D V V E G A V V N I M S W

T U I T V L V V M L W K E Q S S I S C L

       250       260
GTATGATGTACAGCTCGAATTC
 V U C T A R I

  Y D V Q L E F

G M M Y S S N
```

agglutinin and shows no homology to these genes at the 5' end of the sequence. It is likely that the sequence is a vestigial sequence carrying no protein coding information (Loomis and Gilpin, 1986). As there are at least two related, but independent, copies of this sequence it is likely that the sequences have arisen by gene duplication events rather than re-integration of a processed RNA transcript. Bown *et al.* (1985) have shown that a divergent legumin pseudogene in close proximity to a transcriptionally active legumin gene (legA) had probably arisen by a tandem duplication event. The pseudogene sequence had then mutated by accumulating stop codons, deletions and frame shift errors within the coding sequence. The region hybridising strongly with the cDNA probe is more closely related to ricin and agglutinin, on the basis of intensity of hybridisation to the cDNA probe. This sequence is located 10 kb from the weakly hybridising sequence and may be the progenitor sequence for the duplication events.

4.4. Isolation of a ricin-like genomic clone using a ricin-specific oligonucleotide.

The initial lack of success in obtaining full-length, intact, ricin or agglutinin genes resulted in the adoption of a different approach in screening the *Ricinus* gene library. The nucleotide sequences of ricin and agglutinin are sufficiently divergent over part of the coding sequence to allow the synthesis of ricin- and agglutinin- specific oligonucleotide probes. This would allow ricin and agglutinin clones to be distinguished on the basis of hybridisation to the specific oligonucleotides, removing the problem of cross hybridisation between ricin and agglutinin sequences observed when using the cDNA probe. Reliance on restriction mapping of clones to indicate identity is also problematical when dealing with a large number of λ clones, some of which may contain restriction site heterogeneity.

The sequence and position of each gene-specific oligonucleotide probe is shown in Fig 4.6. Sundan *et al.* (1990) used a ricin-specific oligonucleotide probe to isolate a functional ricin coding sequence from a genomic library. The probe sequence corresponded to the same region of ricin coding sequence as shown in Fig 4.6. The probe was used to screen a partial *Ricinus* genomic library, constructed using a size selected total EcoRI digest of *Ricinus* genomic DNA. The ricin-specific probe shown in

Figure 4.6. Sequences of ricin-specific and RCA-specific oligonucleotides synthesised for use as gene-specific hybridisation probes.

The numbering refers to the position of the amino acid sequence within the B-chain of agglutinin and ricin respectively. The amino acid numbering differs between agglutinin and ricin due to a deleted alanine residue in the agglutinin A-chain at amino acid position 130 of the corresponding ricin A-chain. The oligonucleotides differ in sequence at 10 of the 21 base positions.

Ricin	Arg Asp Gly Arg Phe His Asn	aa 303-309
	5' AGG GAT GGA AGA TTC CAC AAC 3'	
	ACA GGT GAA GAA TTC TTC GAT	
Agglutinin	Thr Gly Glu Glu Phe Phe Asp	aa 302-308

Fig 4.6 was used in the same way to screen for ricin containing clones. In this case a complete genomic library was screened.

4.4.1. Oligonucleotide screening of a *Ricinus communis* genomic library.

The ricin-specific oligonucleotide was radioactively end-labelled (section 3.7.2) and used as a probe to rescreen the *Ricinus* gene library. Approximately 1×10^6 phage were plated out on two large plates and duplicate plaque lifts were taken. These were hybridised with the end-labelled ricin-specific oligonucleotide under the conditions described in section 3.10.2 with a hybridisation temperature of 35^0C. Each of the duplicate filters were subjected to a final wash of 3xSSC at 35^0C for 30 min. One of the duplicates was then washed in 3xSSC at 45^0C for a further 5 min. The 45^0C wash is approximately 20^0C below the T_M of a perfect match between oligonucleotide and target sequence in 3xSSC. After washing, the filters were autoradiographed.

18 positive clones were isolated, these exhibited coincidental hybridisation signals on each of the duplicate filters washed to differing stringencies. False positive signals are extremely likely to occur when using a short oligonucleotide probe to screen a genomic library. To eliminate these, the 18 clones were plated out and hybridised with the PstI insert of pRCL59. The filters were subjected to a final wash of 1.0xSSC at 65^0C for 30 min. Autoradiography of the filters revealed only one positively hybridising clone, designated λRic 2ii, and λ DNA was subsequently prepared from it.

4.4.2. Restriction analysis of a putative ricin containing genomic clone.

The clone λ Ric 2ii was a presumptive ricin containing clone on the basis of its hybridisation to both the ricin-specific oligonucleotide and to pRCL59. The clone was subjected to restriction and Southern analysis as shown in Fig 4.7. The Southern blot of the gel was hybridised with the PstI insert of pRCL59. The final post-hybridisation wash was at a relatively high stringency of 0.1xSSC at 65^0C for 30 min.

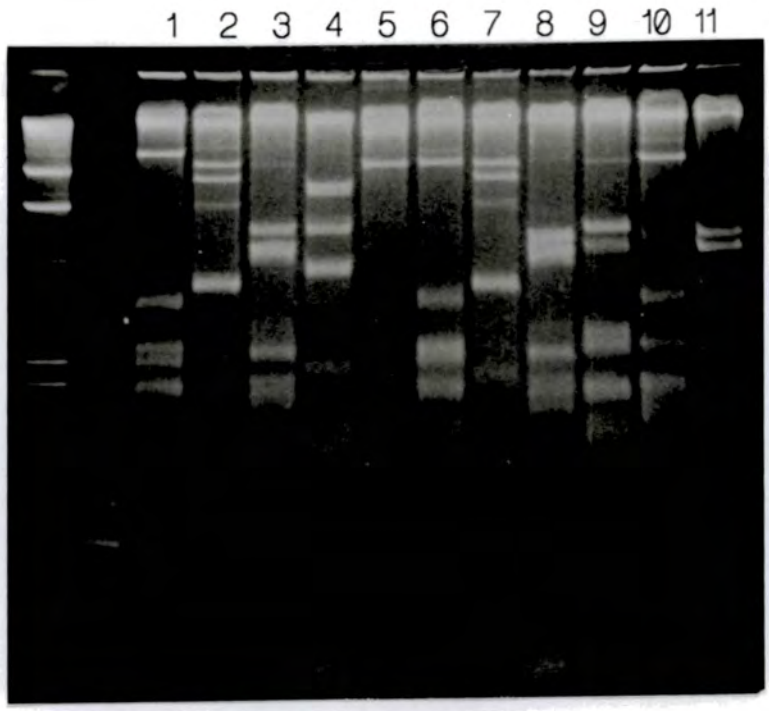
Whilst the separation of the fragments on the gel was rather indistinct, the hybridising restriction fragments on the autoradiograph were well defined and produced a strong signal even after a high stringency wash. The hybridising fragments suggested that this clone was ricin-like as opposed to agglutinin-like. (See Fig 4.1 for the comparison

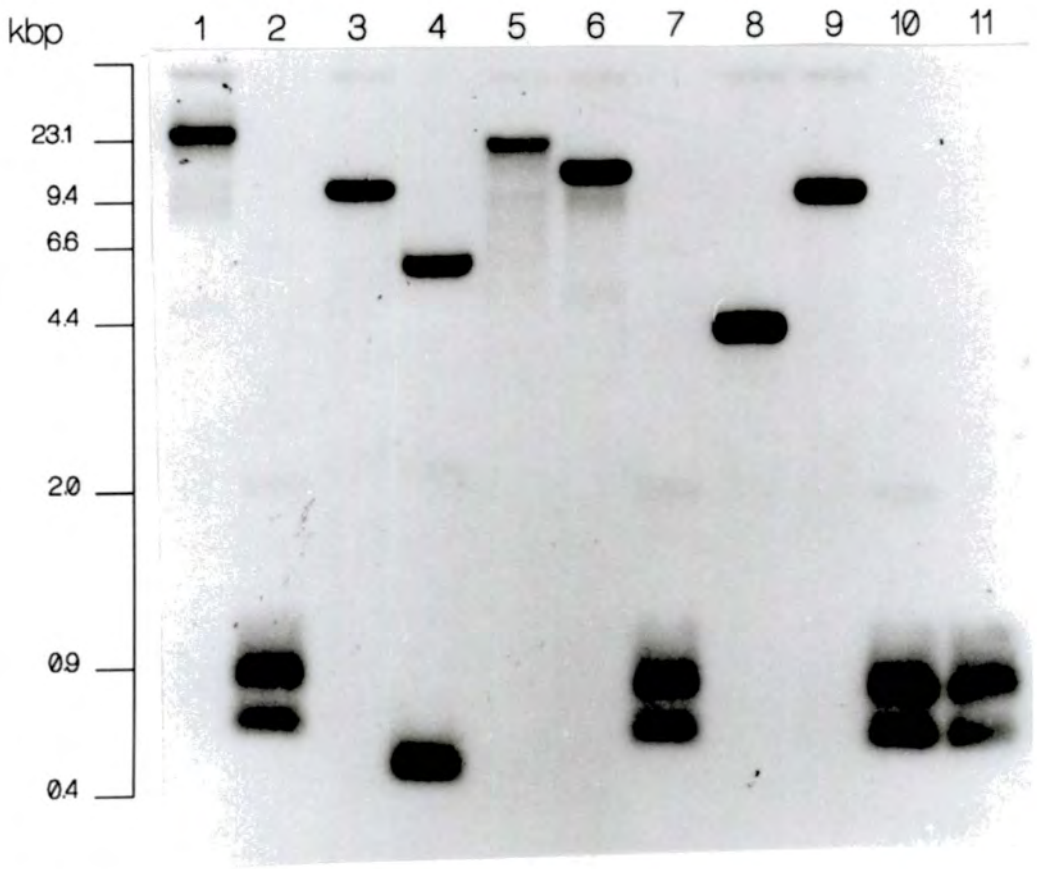
Figure 4.7. Restriction analysis of λRic 2ii. A λ clone hybridising to the *Ricinus* agglutinin cDNA pRCL59 and a ricin-specific oligonucleotide.

DNA was isolated from clone λRic 2ii and restricted with the following enzymes:-

1. EcoRI	5. SalI	9. EcoRI/HindIII
2. BamHI	6. SalI/EcoRI	10. BamHI/EcoRI
3. HindIII	7. SalI/BamHI	11. BamHI/HindIII
4. BglII	8. SalI/HindIII	

Restriction fragments were separated by electrophoresis on a 0.7% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific fragments were detected by hybridisation with the PstI insert of pRCL59. The filter was washed to a final stringency of 0.1xSSC, 65^{0}C for 30 min and autoradiographed. DNA size markers were λ/HindIII and pBR322/AluI.


1 2 3 4 5 6 7 8 9 10 11


kbp
1 2 3 4 5 6 7 8 9 10 11
23.1
9.4
6.6
4.4
2.0
0.9
0.4

of ricin and agglutinin restriction sites.) The ricin cDNA lacks an EcoRI restriction site whilst one is present in the agglutinin cDNA. In lane 1 hybridisation is to a single EcoRI fragment of approximately 23 kb in size, an agglutinin gene would have produced two hybridising fragments as an EcoRI site lies within the region of probe hybridisation. Lane 2 shows the BamHI restriction digest with 2 hybridising fragments of approximately 900 bp and 700 bp. This is consistent with both the ricin and agglutinin restriction maps. Significantly the BglII digest (lane 4) contained a hybridising fragment of approximately 500 bp and a larger fragment of about 6.0 kb. The ricin coding sequence contains two BglII sites 550 bp apart whilst the agglutinin sequence has only one BglII site.

The restriction analysis shown suggests that on the basis of a comparison of restriction sites the clone contains a ricin-like sequence. There is sufficient sequence homology between ricin and agglutinin sequences to allow cross-hybridisation. The screening of the genomic library with the oligonucleotide was facilitated by being able to eliminate false positives using a second screen with pRCL59. Had this not been available then screening using solely the oligonucleotide would have been extremely labour intensive due to the large number of false positives. The fact that only one clone hybridised to both probes suggests that the library is not complete. More than one positive signal would be expected as there is more than one ricin gene in the *Ricinus* genome.

The ricin-specific oligonucleotide used as a probe under the described conditions is ricin specific and does not cross-hybridise with agglutinin sequences. This was confirmed in hybridisations of ricin- and agglutinin-specific oligonucleotides to Southern blots of the agglutinin cDNA, pRCL59. No cross-hybridisation between the ricin-specific oligonucleotide and the agglutinin cDNA was observed, see section 4.5.2.

The genomic clone described above was not characterised further and the information presented does not confirm that it contains either an active ricin gene or a sequence encoding a functional polypeptide. However the clone was subsequently used as a template for *in vitro* expression studies, described in chapter 6, which showed that the clone probably contains a functional ricin coding sequence.

4.5. Amplification of ricin and agglutinin genomic sequences using the Polymerase Chain Reaction (PCR).

The initial gene library screening work described above was unsuccessful in producing fully characterised ricin and agglutinin gene sequences. An alternative approach was devised and carried out in parallel with the latter stages of the gene library screening described previously.

The approach entailed using the published cDNA sequences of ricin and agglutinin to design gene-specific oligonucleotide primers. These were then used to direct the specific amplification of ricin and agglutinin coding sequences using the polymerase chain reaction. Saiki *et al.* (1988a) demonstrated the successful amplification of genomic target sequences of up to 2.0 kb. However, most examples of PCR amplifications cited at this time were for target sequences of less than 1.0 kb. For example Saiki *et al.* (1988b) amplified a 536 bp region of the human β-globin gene. Kawasaki *et al.* (1988) amplified 125 bp, 200 bp and 307 bp fragments from cDNAs generated by reverse transcription of leukemia-specific mRNA sequences.

The lack of information concerning the amplification of sequences longer than 1.0 kb led to the development of a test system to demonstrate the feasibility of amplifying longer target sequences. This was of crucial importance in the PCR amplification of ricin and agglutinin coding sequences since these are 1.8 kb in length.

4.5.1. Use of plasmid pJY25 as a test template for PCR DNA amplifications.

The plasmid pJY25 (Yarwood *et al.*, 1987), contains a hybrid precursor legumin cDNA, and was chosen as a template for PCR amplifications as it contains a cDNA insert of 1.8 kb in length. This is comparable to the length of the ricin coding sequence. The cDNA insert was cloned in the BamHI site of pUC18 and is illustrated in Fig 4.8. The insert is flanked by the universal M13 forward and reverse sequencing primer annealing sites. These primers were used to direct the PCR amplification of the insert. As illustrated, the expected size of product is expected to be greater than the insert size due to the incorporation of the primers and flanking regions in the PCR product.

Fig 4.9 shows the result of a PCR reaction using pJY25 and the M13 sequencing primers. It was unknown, at this time, whether PCR

Figure 4.8. The structure of pJY25, a test template for the PCR DNA amplification of long target sequences.

The diagram shows selected restriction sites present in pJY25, a hybrid precursor legumin cDNA (Yarwood *et al.* 1987) cloned in the BamHI site of plasmid pUC18. This plasmid was chosen to test the feasibility of amplifying long (>1 kb) DNA sequences. The length of the insert is 1.82 kb and is comparable to the size of the preproricin coding sequence. The cDNA sequence is flanked by the universal M13 primer annealing sites for the forward and reverse sequencing primers. These 17 mer forward and reverse primers were used to initiate DNA synthesis on opposite strands of the plasmid DNA and amplify the insert DNA using PCR. The expected size of PCR product is shown. This is 92 bp longer than the cloned insert due to the inclusion of the primer sequences and cloning sites flanking the insert.

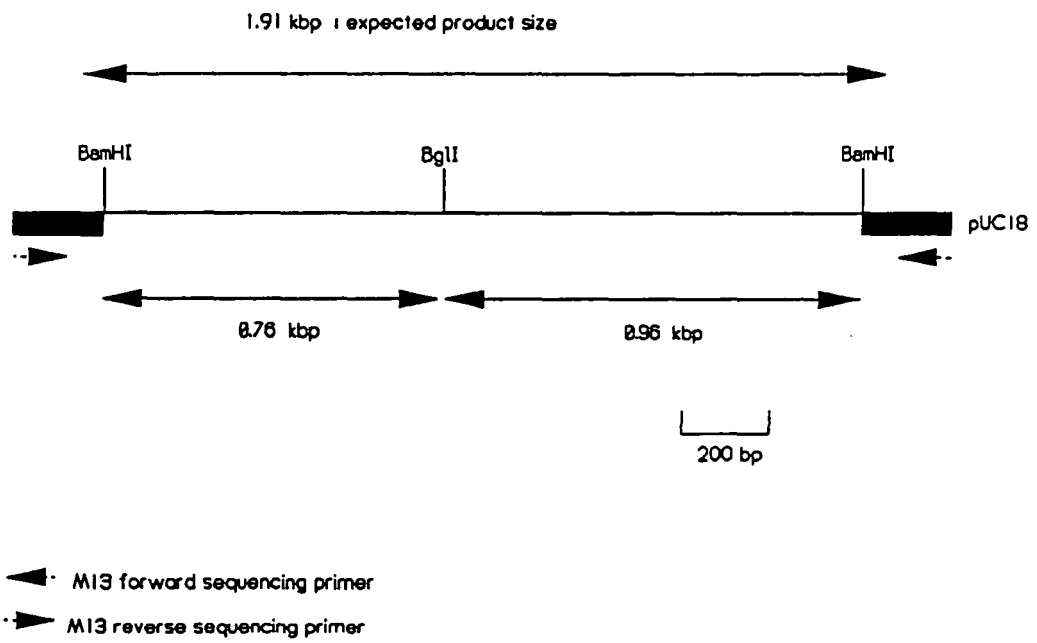


Figure 4.9. PCR amplification of a hybrid legumin cDNA coding sequence from pJY25.

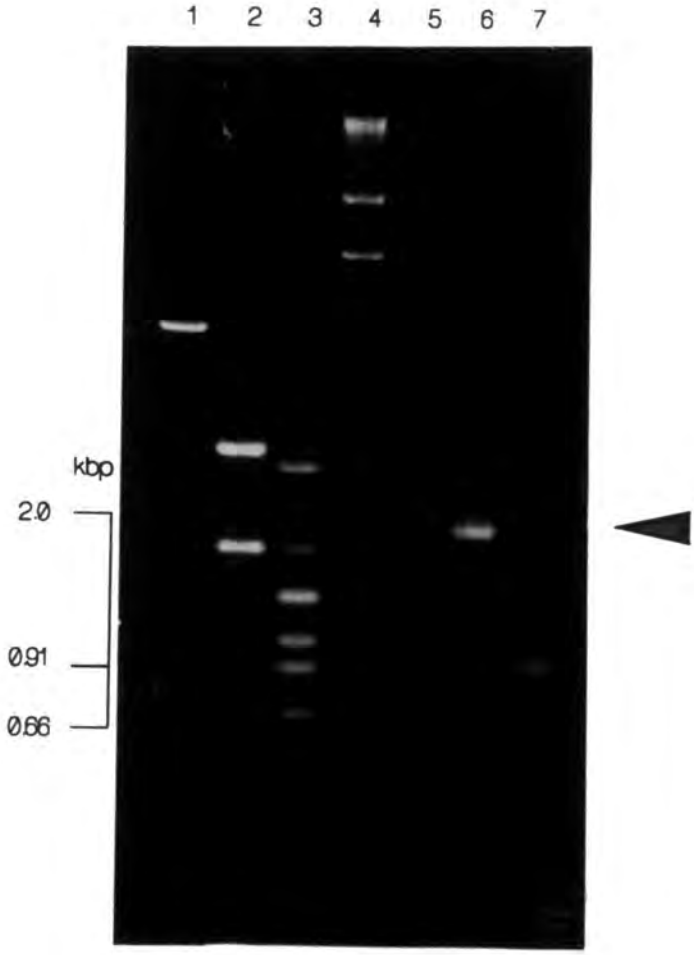
Plasmid pJY25 was linearised with restriction enzyme ScaI and subjected to PCR DNA amplification under the following conditions:-

A 100 μl PCR reaction mix contained 1 ng of plasmid template and the M13 forward and reverse sequencing primers (17 mers) were both at a concentration of 1.0 μM. After heating to 94^{0}C for 1.5 min, 2.5 units of *Taq* polymerase were added and the reaction was subjected to 25 cycles of DNA amplification using set-temperature water baths. Cycle conditions were annealing for 3 min/37^{0}C, extension 9 min/72^{0}C and denaturation 2 min 94^{0}C. The final extension time was increased to 16 min to ensure completion of strand synthesis.

PCR products were analysed by electrophoresis on a 0.8% (w/v) agarose gel with control DNA samples.

1. 0.4 μg pJY25 ScaI restricted
2. 1.0 μg —/— BamHI restricted
3. 0.4 μg —/— BamHI/BglI restricted
4. λ/HindIII size markers
5. pBR322/AluI size markers
6. 20 μl PCR reaction
7. 20 μl PCR reaction BglI restricted

The single PCR product of the predicted size (lane 6) is arrowed.


1
2
3
4
5
6
7
kbp
2.0
0.91
0.66

amplifications could be performed using circular, supercoiled plasmid templates. To eliminate the possibility of circular plasmid forms interfering in the amplification, the plasmid was linearised using restriction enzyme ScaI. This enzyme has one recognition sequence outside the pUC18 polylinker whilst the insert does not contain a ScaI site.

A PCR extension time of 9 min at 72^0C was used to ensure the completion of full-length strand synthesis, the reports cited above employed shorter times due to the smaller size of product. Lane 1 of the gel photo shows the linear form of the plasmid, lane 2 shows the plasmid restricted with BamHI, which excises the 1.8 kb cDNA insert. The product of the PCR amplification is arrowed in lane 6. It is 1.9 kb in size and is consistent with the amplification of the insert and flanking polylinker sequences.

As a further check to confirm the fidelity of the reaction the PCR product was restricted with BglI (one site within the cDNA insert). Lane 7 shows the result of this restriction: 2 fragments were produced. A control restriction of pJY25 restricted with BamHI and BglI (lane 3) was also included. This excises the insert from the vector and cleaves the insert once to produce two fragments. These are the lowest molecular weight bands (the higher molecular weight fragments of fainter intensity are partial restriction products). The fragments present in the PCR product restriction were of slightly greater size due to the incorporation of the primer and flanking sequence at either end of the insert.

This PCR experiment demonstrated that target sequences of 1.8 kb could be amplified and that a characteristic restriction site present in the cDNA template was present in the product, confirming its identity. Later PCR amplifications have demonstrated that unrestricted circular, supercoiled plasmids are also suitable templates for PCR amplifications, see PCR amplifications described in Fig 6.5, and subsequent plasmid templates used were not linearised.

4.5.2. PCR amplification of ricin gene sequences.

The success of the PCR amplification using pJY25 as a test template indicated that it may be possible to amplify ricin and agglutinin genes directly, using genomic DNA as a template. The primer sequences designed for this are described in Fig 4.10. As shown, the primers for ricin and agglutinin are extremely similar. The 5′ primer is the same for both

Figure 4.10. Sequences of primers synthesised for the PCR amplification of ricin and RCA genes.

Primers were derived from the published cDNA sequence comparisons of ricin and agglutinin (Roberts *et al.*, 1985). The 5′ primer is the same for both ricin and agglutinin as the sequence is not divergent over this region. Restriction sites were added to the 5′ end of each primer to facilitate the cloning of the PCR product. (The agglutinin sequences shown are labelled RCA.)

1. 5′ primer for ricin and agglutinin. The primer sequence starts in the 5′ untranslated region and ends at the Met initiation codon at the start of the ricin/agglutinin leader sequence.

2. 3′ primers for ricin and agglutinin. Primers were chosen at the 3′ end of the coding sequence to include the last 16 amino acids of the B-chain and the stop codon. The sequence of ricin and agglutinin is shown in this region with the DNA sequence divergence between the two genes. The actual primer sequences are the complement of the coding sequence shown.

1. 5′ Ricin and agglutinin primer.

```
                EcoR1
Primer        CCGGAATTCGGAG GAAATACTAT TGTAATATGG AT→
Ricin/RCA  5′-- AAACCGGGAG GAAATACTAT TGTAATATGG ATG TAT GCG GTG GCA ACA
                                                  M   Y   A   V   A   T
```

2. 3′ Ricin and agglutinin primers.

```
  aa 526 L   Y   P   L   H   G   D   P   N   Q   I   W   L   P   L   F   *
Ricin  CTT TAC CCT CTC CAT GGT GAC CCA AAC CAA ATA TGG TTA CCA TTA TTT TGA
RCA    GTT CAC CCT TTC CAT GGA GAC CTA AAC CAA ATA TGG TTA CCA TTA TTT TGA
  aa 525 V   H   ——  F  ————  N  -  L  ————————————————————————————————
```

Complements of the above sequences with added HindIII sites at the 5′ terminus. The agglutinin specific primer has 6 base differences compared with the ricin primer.

```
                Hindlll
Ricin   5′ GGCCAAGCTTTCAAAATAATGGTAACCATATTTGGTTTGGGTCACCATGGAGAGGGTAAAG 3′
RCA        ————————————————————————————————————A———TT————A———G—C
```

genes and the 3' primer for agglutinin differs in only 6 bases compared with the ricin 3' primer. Restriction site recognition sequences were incorporated at the 5' end of each primer to facilitate the subsequent cloning of the PCR products. The inclusion of restriction sites has previously been shown not to interfere with amplification reactions (Scharf *et al.*, 1986). EcoRI and HindIII recognition sequences were chosen as these sites are not present in the ricin coding sequence, whilst agglutinin has a single EcoRI site. If the PCR amplification using the ricin primers also amplified agglutinin sequences, due to sequence similarity, then restriction of the product with EcoRI and HindIII would allow the selective cloning of full-length ricin sequences.

The result of PCR amplifications using the ricin-specific primers with *Ricinus* genomic DNA as a template is shown in Fig 4.11. Control DNA samples were the agglutinin cDNA pRCL59 and *Ricinus* genomic DNA. The reactions were optimised with respect to magnesium ion concentration. The full-length PCR product obtained is arrowed on the gel photograph and is of the expected size. The amplification carried out at 1.5 mM Mg^{2+} did not produce a product. The reaction was successfully repeated using a different batch of *Taq* polymerase, see Fig 4.12.

When the gel was Southern blotted and hybridised with the 700 bp BamHI insert of pRCL59 strong hybridisation to the PCR product was seen (lanes 10 and 13). The insert of pRCL59 (lanes 2 and 3) also hybridised, but no hybridisation was observed in the lanes containing only genomic DNA. The product was not cleaved by EcoRI and HindIII (lanes 11 and 14). Restriction with BglII produced two hybridising fragments. This is consistent with the two BglII sites present in ricin (Fig 4.1), three fragments should be produced but only two will hybridise to the probe.

The restriction analysis suggests that the product is ricin-like, however some product was not restricted with BglII. This could be due to: the lack of sites in the product, loss of sites during the amplification, or the sample being refractory to endonuclease restriction. The PCR product appeared to be ricin-like as it was not restricted with EcoRI or HindIII. The ricin coding sequence contains neither of these sites. However this restriction is not diagnostic for ricin if the product is in fact refractory to endonuclease digestion. This, coupled with the cross-hybridisation of agglutinin sequences with ricin, indicates only that the amplified sequences are related to the ricin/agglutinin gene

Figure 4.11. Amplification of ricin gene sequences from *Ricinus communis* **genomic DNA.**

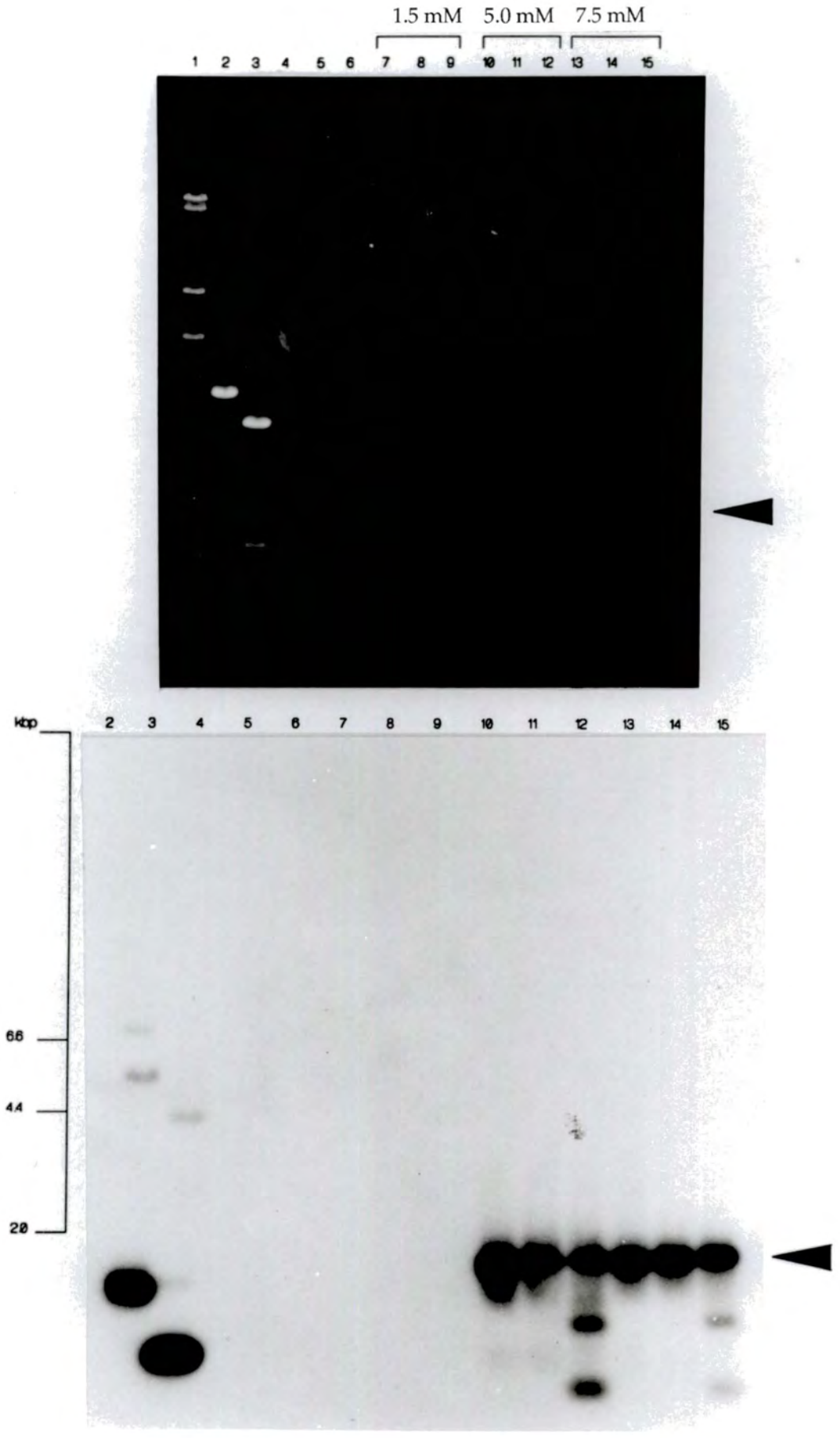
PCR primers were designed to amplify the entire coding sequence of ricin as shown in Fig. 4.10. These primers were used in a PCR amplification using *Ricinus* genomic DNA as the template. Triplicate 100µl reactions contained 1.0 µg of genomic DNA, 1.0 µM each of the 5′ and 3′ primers and either 1.5 mM Mg^{2+}, 5.0 mM Mg^{2+} or 7.5 mM Mg^{2+}. The reactions were subjected to the same temperature/time cycles as described for the amplification of plasmid pJY25, see Fig. 4.9.

After amplification the reactions were phenol extracted, ethanol precipitated and the PCR products resuspended in 20 µl of water. The products from each reaction (4 µl) were restricted with EcoRI/HindIII and BglII. The gel photo shows the analysis of each amplification reaction by electrophoresis on a 0.65% (w/v) agarose gel with control DNA samples. Full-length PCR products are arrowed. Lanes are:-

1. λ/HindIII size markers
2. 0.5 µg pRCL59 PstI restricted
3. 0.5 µg pRCL59 BamHI restricted
4. *Ricinus* genomic DNA unrestricted
5. ————/———— EcoRI/HindIII restricted
6. ————/———— BglII restricted

1.5 mM Mg^{2+}	7.	Ricin PCR product unrestricted.
Amplification	8.	————/———— EcoRI/HindIII restricted
	9.	————/———— BglII restricted
5.0 mM Mg^{2+}	10.	Ricin PCR product unrestricted
Amplification	11.	————/———— EcoRI/HindIII restricted
	12.	————/———— BglII restricted
7.5 mM Mg^{2+}	13.	Ricin PCR product unrestricted
Amplification	14.	————/———— EcoRI/HindIII restricted
	15.	————/———— BglII restricted

DNA was transferred from the gel to a nitrocellulose filter and specific fragments were detected by hybridisation with the 700 bp BamHI fragment of pRCL59. The filter was washed to a final stringency of 0.1xSSC, 65^{0}C for 30 min and autoradiographed. The full size ricin PCR product is arrowed.


1.5 mM
5.0 mM
7.5 mM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
kbp
2 3 4 5 6 7 8 9 10 11 12 13 14 15
66
44
20

family and gives no indication of the reaction specificity.

Fig 4.12 shows the effect of annealing temperature on the ricin PCR amplification. The 37^0C annealing reaction had twice the amount of product loaded on the gel. The previous amplification at this temperature was unsuccessful (Fig 4.11, lanes 7, 8 and 9). The amplification was repeated and a fifth volume of the reaction mixture was analysed by electrophoresis to ensure the detection of amplification products. The other amplification reactions were analysed using a tenth volume of the reaction mixture. Amplification is observed at the highest annealing temperature (lane 5) and the major product is the expected size of 1.8 kb. There is slightly less product "smear" at the higher annealing temperatures. The reaction products from the 60^0C annealing reaction were used in further experiments.

To determine the specificity of the ricin PCR amplification the reaction products were hybridised with the ricin and agglutinin gene-specific oligonucleotides described in Fig 4.6. Hybridisations were also carried out using the 5′ and 3′ PCR primers as probes. The results of these hybridisation experiments are shown in Fig 4.13. The four Southern blots prepared each contained the ricin PCR product and a control sample of pRCL59. Each blot was then hybridised with a different oligonucleotide probe as described in the figure legend.

The ricin-specific probe only hybridised to the PCR product, no hybridisation to the agglutinin sequence of pRCL59 was detectable, even after 72 hr exposure. The reverse situation occurs when the agglutinin specific probe is used. The probe hybridised to the agglutinin cDNA but no hybridisation to the ricin PCR product was detectable. (Exposure of each blot for 72 hr did reveal a small degree of hybridisation to the pBR322 vector sequence of pRCL59). Under the hybridisation conditions used there is no detectable cross-hybridisation between the ricin-specific probe and the agglutinin sequence.

The 3′ and 5′ primers used as probes hybridised to the PCR product as would be expected. Hybridisation to smaller sequences is also seen. These sequences are presumably non-specific amplification products as the ricin probe does not hybridise to them. Only the 3′ primer hybridises to the agglutinin sequence as the cDNA is truncated at its 5′ end.

The result of this experiment shows the specific amplification of ricin-like sequences. There is no evidence that agglutinin sequences were

Figure 4.12. Effect of annealing temperature on the amplification of ricin sequences from genomic DNA.

PCR amplifications were performed as described in Fig. 4.11 except the annealing temperatures used were either ^37°C or 50°C or 60°C. 10 µl of each reaction was analysed by electrophoresis on a 0.9% (w/v) agarose gel. The magnesium concentration was 1.5 mM (standard buffer concentration). The first reaction at 1.5 mM was not successful (Fig. 4.11), however a repeat reaction at this magnesium concentration using a different batch of *Taq* polymerase resulted in a successful amplification (Lane 3). Lanes are:-

1. λ/HindIII size markers
2. pBR322/AluI size markers
3. 20 µl ricin PCR reaction mix, 37°C annealing
4. 10 µl ————————/———————— 50°C annealing
5. 10 µl ————————/———————— 60°C annealing

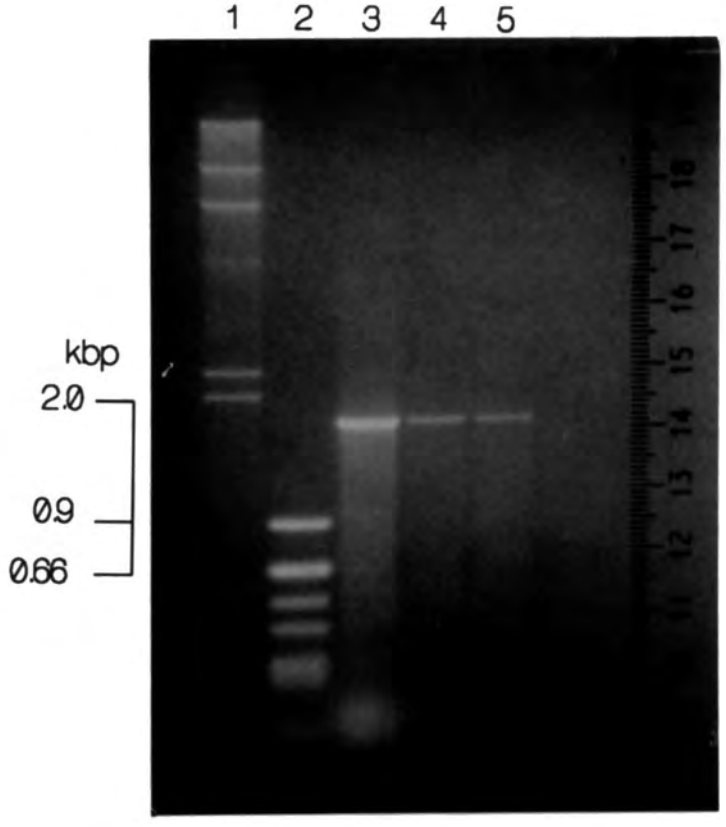

1
2
3
4
5
kbp
2.0
0.9
0.66

Figure 4.13. Analysis of ricin PCR products by hybridisation with ricin and agglutinin gene-specific oligonucleotide probes.

Ricin PCR amplifications were electrophoresed on 0.85% (w/v) agarose gels with control DNA samples (see gel photographs). DNA size markers were λ/HindIII and pBR322/AluI. Four identical Southern blots were prepared, the respective lanes contained:-

1. 0.4 μg pRCL59 PstI restricted
2. ————/———— BamHI restricted
3. 5 μl of a ricin PCR reaction using a 60^{0}C annealing step, the 1.8 kb ricin PCR product is arrowed.

Each blot was hybridised with an end-labelled oligonucleotide probe as follows:-

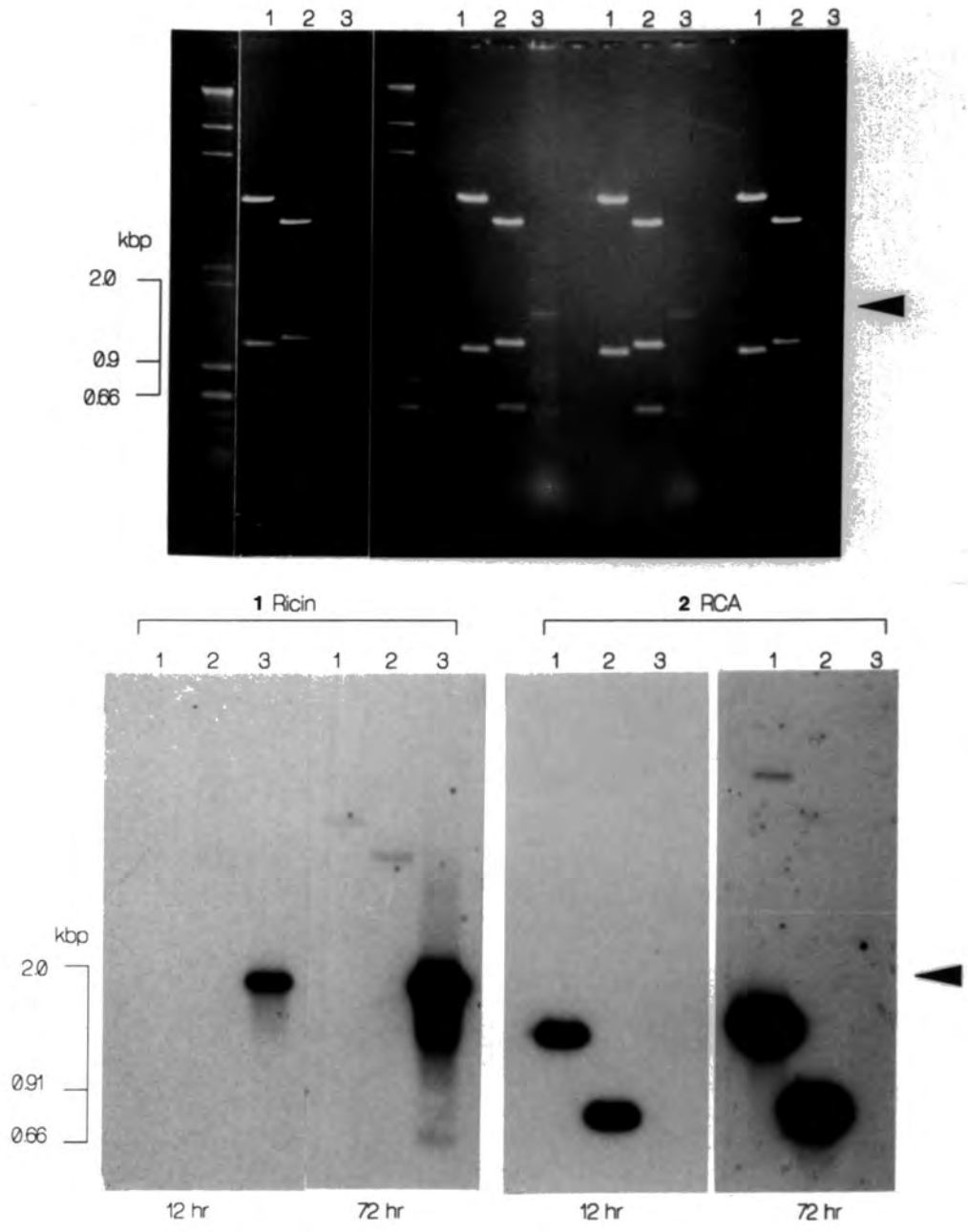
Blot 1. Ricin-specific oligonucleotide (see Fig. 4.6)

Blot 2. *Ricinus* agglutinin-specific oligonucleotide ————/————

Blot 3. 3′ ricin PCR primer (see Fig. 4.10)

Blot 4. 5′ ——————————/——————————

Hybridisation steps were carried out at 35^{0}C (low stringency) with a final post-hybridisation wash of 3×SSC, 35^{0}C for 30 min. The exposure time for each autoradiograph is indicated.


1 2 3 1 2 3 1 2 3 1 2 3
kbp
2.0
0.9
0.66
1 Ricin
2 RCA
1 2 3 1 2 3
1 2 3 1 2 3
kbp
2.0
0.91
0.66
12 hr
72 hr
12 hr
72 hr
3 3' primer
4 5' primer
1 2 3
1 2 3
kbp
2.0
0.91
0.66
12 hr
72 hr

also amplified. This indicates that the difference in sequence between the 3′ ricin and 3′ agglutinin primer is sufficient to confer ricin-specificity in the amplification reaction. The differences in the sequence of the 3′ primer are at the 3′ end of the oligonucleotide (Fig 4.10), it is from the 3′ end of the primer that DNA synthesis occurs after primer annealing to template DNA. This result is in agreement with the work of Sommer and Tautz (1989) who demonstrated that primers require at least 3 homologous nucleotides at the 3′ end to successfully prime DNA synthesis. Mismatches in the 5′ region of the primer are tolerated during PCR amplifications. The 3′ terminal base is a G in the ricin 3′ primer and a C in the corresponding agglutinin primer. The presence of the few base mismatches, especially the terminal one, are sufficient to prevent the initiation of DNA synthesis and subsequent agglutinin sequence amplification. The 5′ primer will anneal to both ricin and agglutinin sequences as they are identical over this region such that the specificity resides solely in the 3′ primer sequence.

Kwok *et al.* (1990) have shown that a single base pair mismatch at the 3′-terminal base of a primer can be tolerated in PCR amplifications. Mismatches which amplify as efficiently as a fully complementary primer-template duplex are T:C, T:G, T:T, G:G and A:C. An A:A mismatch moderately reduced the efficiency of PCR amplification whilst A:G and C:C mismatches reduced efficiency to less than 1.0%. These results are dependant upon the reaction conditions used. A G:G mismatch amplified efficiently at 800 μM dNTPS but product yield was reduced in 50 μM dNTPS. The inclusion of further mismatches within the last four 3′-terminal bases also reduces efficiency.

These results are consistent with the ricin PCR primers not amplifying the agglutinin sequence. The 3′-terminal mismatch is G:G, there are further mismatches in the primer and the dNTP concentration in the reaction was 200 μM. In the ricin PCR reaction there is no evidence for the amplification of agglutinin sequences on the basis of the hybridisation results presented. The results cited above were concerned with reduced efficiencies of amplification whereas in the case of agglutinin there is no detectable amplification. This probably reflects variation in the reaction conditions used and the combination of factors discussed above. Another possible contributory factor would be the ricin amplification itself. This would both reduce the primer concentration and

the effective amount of agglutinin target sequence as the number of ricin target sequences accumulated during the reaction.

4.5.3. PCR amplification of agglutinin gene sequences.

The primers synthesised for the PCR amplification of ricin produced a specific product, as judged by the hybridisation of the ricin- and agglutinin-specific oligonucleotides to Southern blots. The same approach was taken to assess the specificity of a PCR amplification using the agglutinin primers. The result of an agglutinin PCR amplification is shown in Fig 4.14 using genomic DNA as a template.

Lane 1 of the gel photo contains the ricin amplification products of the expected size. Lane 2 contains the agglutinin amplification products. Two products are clearly seen, one is the same size as the ricin product (1.8 kb) and the other is approximately 800 bp in size. Duplicate Southern blots of the gel were hybridised with either the ricin- or agglutinin-specific oligonucleotides as described in the figure legend.

The left hand autoradiograph shows the result of hybridisation with the agglutinin-specific oligonucleotide. Only the 1.8 kb agglutinin product hybridised (lane 2), there was no hybridisation to the ricin PCR product. This is the expected result as the previous hybridisations of specific oligonucleotides to the ricin PCR product and agglutinin cDNA showed no cross-hybridisation. This result indicates that the agglutinin PCR amplification amplified agglutinin sequences.

The right hand autoradiograph shows the result of hybridising the ricin-specific oligonucleotide to the ricin and agglutinin PCR products. Significantly the oligonucleotide hybridised to both the 1.8 kb ricin and agglutinin PCR products (lanes 1 and 2). No hybridisation to the 800 bp product was observed, this product is presumed to be a PCR artifact, probably arising due to chance non-specific priming. The identity of this 800 bp product was not investigated further, however hybridisation with the agglutinin cDNA would determine if it was a related sequence.

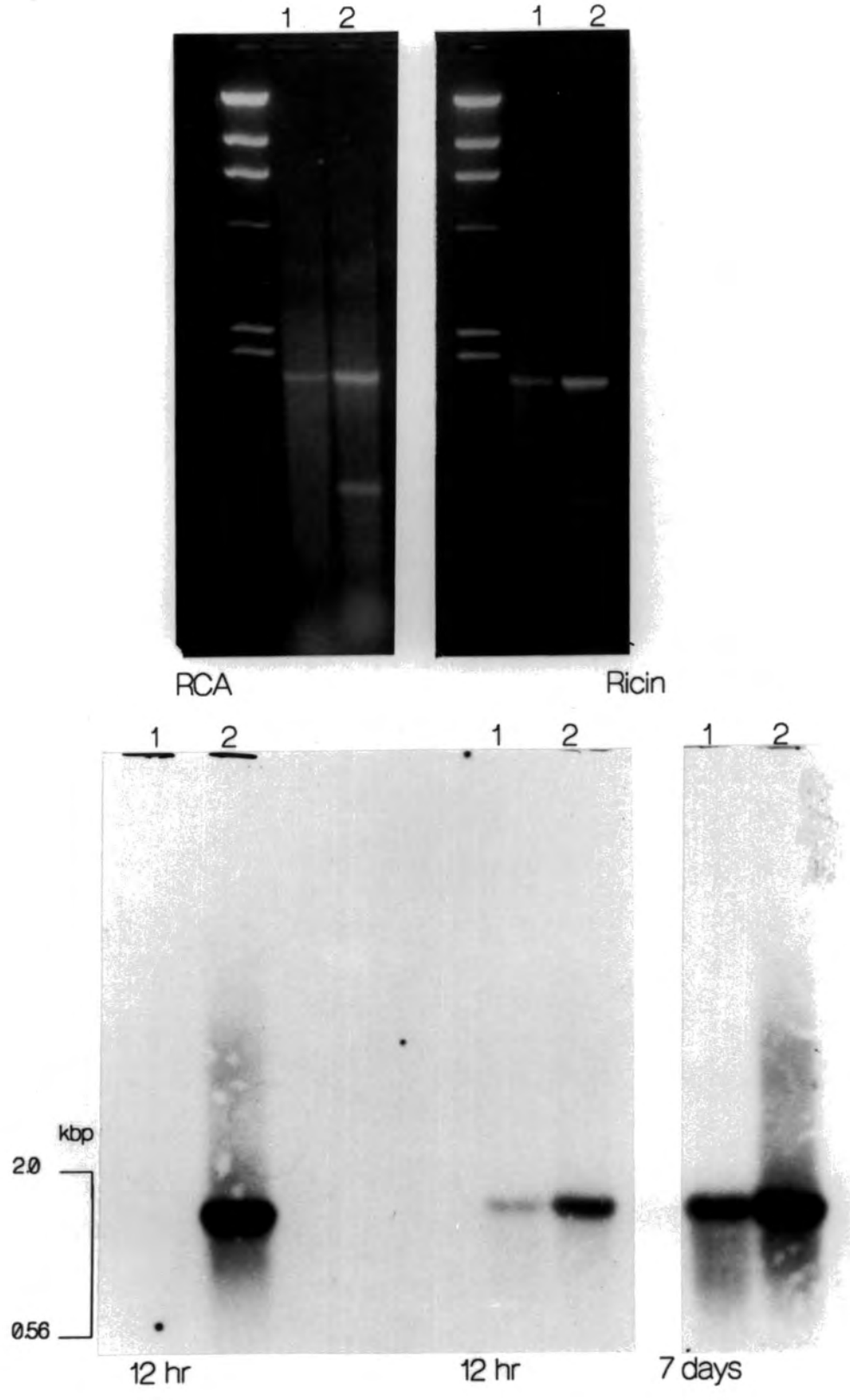
Hybridisation of the ricin-specific oligonucleotide to the agglutinin PCR product is unexpected. The previous results demonstrated that the ricin-specific oligonucleotide did not hybridise to the agglutinin cDNA. The blot hybridised with the ricin-specific oligonucleotide was washed at a higher temperature to determine if there was any differential removal of probe from the hybridising sequences (7 day exposure of autoradiograph).

Figure 4.14. PCR amplification of ricin and *Ricinus* agglutinin gene sequences.

PCR DNA amplifications were performed on *Ricinus* genomic DNA using the conditions described in Fig. 4.11 except the reactions were subjected to automatic temperature cycling with an annealing temperature of 40^0C. The two reactions performed were identical except the second reaction contained the *Ricinus* agglutinin 3' primer not the corresponding ricin primer. The 5' primer was the same in both cases, see Fig. 4.10.

20 μl of each PCR amplification was electrophoresed in duplicate on a 1.0% (w/v) agarose gel and transferred to a nitrocellulose filter. Lanes 1 and 2 contain, respectively, the ricin and agglutinin PCR amplifications.

The filters were hybridised with either the ricin- or agglutinin-specific end labeled oligonucleotides, as shown. Hybridisation was at 42^0C with final washes at this temperature in 3xSSC for 90 min, the filters were then autoradiographed. The filter hybridised with the ricin specific oligonucleotide was subsequently washed in 3xSSC for 30 min at 50^0C and autoradiographed. Exposure times for each filter are indicated. DNA size markers were λ/HindIII.


1
2
1
2
RCA
Ricin
1
2
1
2
1
2
kbp
2.0
0.56
12 hr
12 hr
7 days

Probe was still bound to both PCR products.

This result, whilst originally unexpected, is consistent with the findings of both Araki and Funatsu (1987) and Ladin *et al.* (1987). They have shown that ricin E is a product of a gene recombination event between ricin D and *Ricinus* agglutinin. Araki and Funatsu determined the B-chain amino acid sequence whilst Ladin *et al.* sequenced a cDNA clone encoding the ricin E coding sequence. The A-chain sequence of ricin E is ricin-like, as is the N-terminal sequence of the B-chain until amino acid residue 150. Subsequently the B-chain sequence is agglutinin-like. This ricin E coding sequence would be co-amplified with agglutinin using the agglutinin-specific primers as both genes are agglutinin-like at the 3' end (C-terminal). The ricin- and agglutinin-specific oligonucleotide hybridise to the B-chain coding region (see Fig 4.6).This region in the ricin E sequence is ricin-like before the change to agglutinin-like sequence. This explains the observed hybridisation of the ricin-specific oligonucleotide to the agglutinin PCR product as ricin E sequences and agglutinin sequences have been amplified.

4.5.4. Cloning the ricin PCR amplification product.

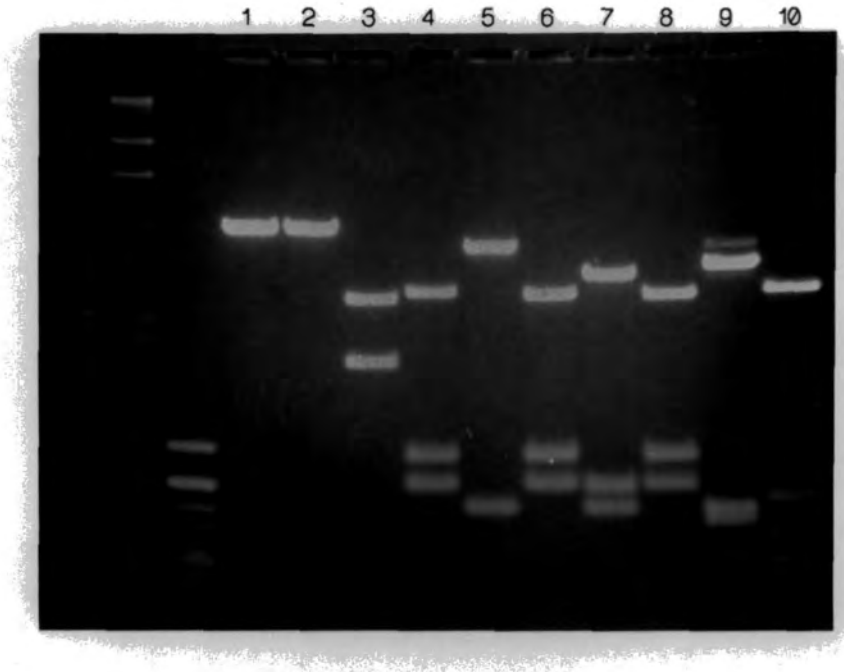
The ricin PCR product was cloned in pUC18 as described in Fig 4.15. Only one recombinant clone was obtained from 100 white colonies screened. This clone, designated pRic 1, was analysed by restriction enzyme digestion and Southern blotting. The hybridising restriction fragments are those expected for a clone containing a ricin coding sequence (see Fig 4.1). The insert is excised from the vector by EcoRI and HindIII restriction and is the same size as the ricin PCR product (lane 3). The BamHI fragments characteristic for both ricin and agglutinin sequences are present (lane 4). BglII restriction releases a fragment of 550 bp consistent with the two BglII sites in the ricin cDNA sequence. There is only one BglII site in the agglutinin sequence. These results suggest that the ricin PCR clone is a ricin-like sequence.

Partial nucleotide sequence information was obtained from the clone by plasmid sequencing using the M13 forward sequencing primer. The sequence is presented in Fig 4.16. Compared to the cDNA sequence of ricin, over the same region, there are five base differences giving rise to one amino acid change in the open reading frame. Four of the base changes are in the third base of a codon. This limited sequencing confirmed that the PCR

Figure 4.15. Restriction analysis of a cloned ricin PCR product.
The plasmid clone pRic 1 was obtained by restricting the PCR products of a ricin amplification reaction with EcoRI and HindIII, ligation into pUC18 restricted with EcoRI and HindIII and then transformation into *E. coli* DH5α. One recombinant hybridised with the PstI insert of pRCL59 when the transformants were screened by colony hybridisation. Plasmid DNA was prepared from this clone and restricted with the following enzymes:-

1. EcoRI
2. HindIII
3. EcoRI/HindIII
4. BamHI
5. BglII
6. EcoRI/BamHI
7. EcoRI/BglII
8. HindIII/BamHI
9. HindIII/BglII
10. BamHI/BglII

Restriction fragments were separated by electrophoresis on a 1.0% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific sequences were detected by hybridisation with the PstI insert of pRCL59. The final filter wash was 0.1xSSC for 30 min at 65^{0}C. DNA size markers were λ/HindIII and pBR322/Alu1.


1 2 3 4 5 6 7 8 9 10

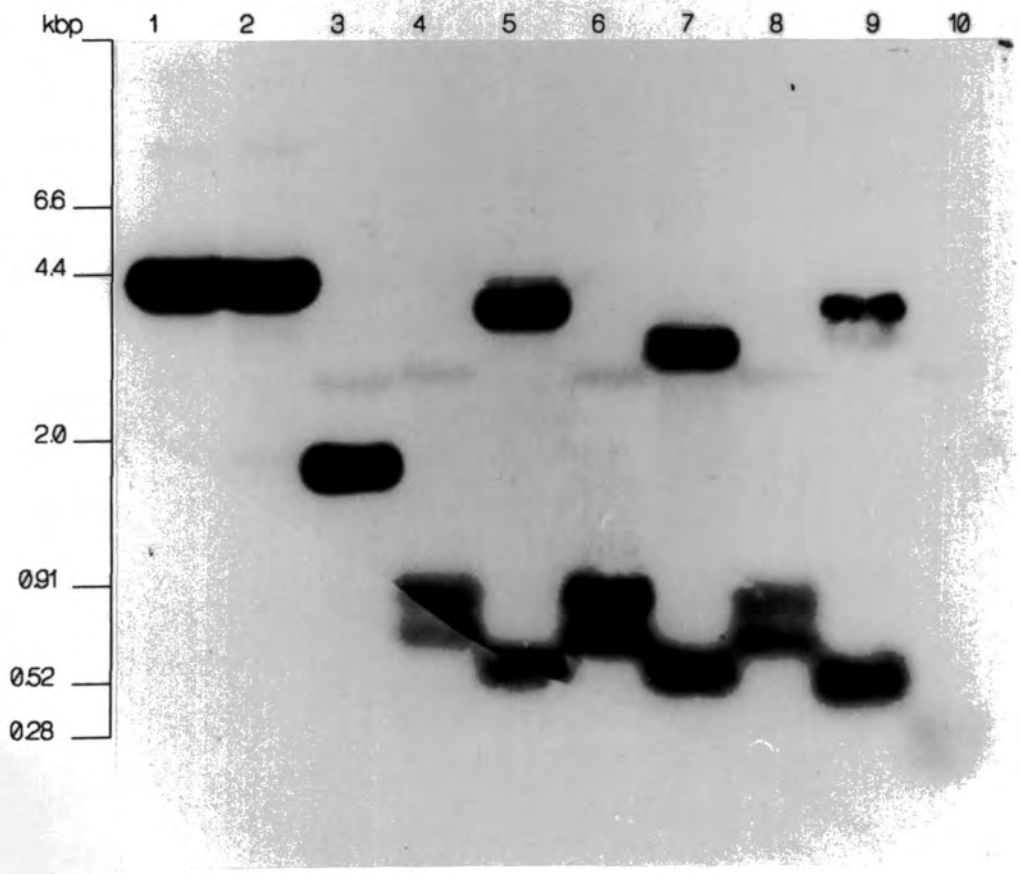
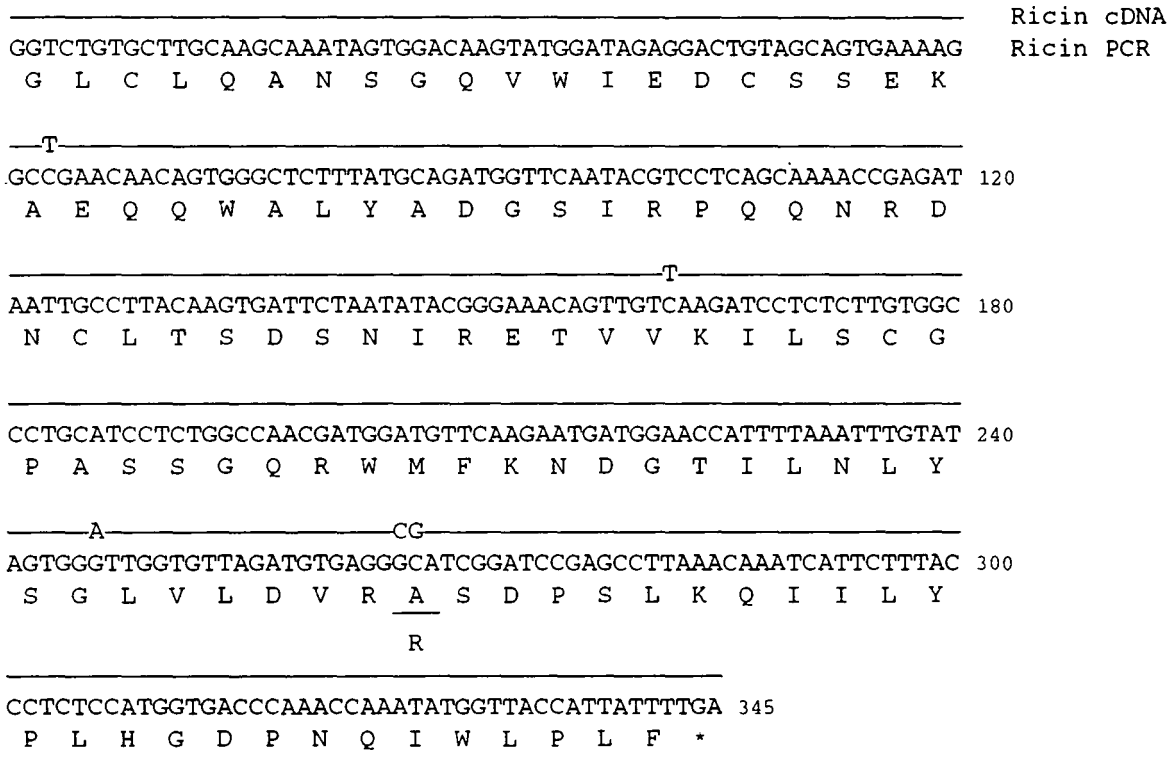

kbp
1 2 3 4 5 6 7 8 9 10
6.6
4.4
2.0
0.91
0.52
0.28

Figure 4.16. The 3' nucleotide sequence of a ricin gene sequence obtained by PCR amplification.

The clone pRic 1 was subjected to double stranded plasmid sequencing using the M13 forward sequencing primer. The nucleotide sequence shown encodes an open-reading frame corresponding to the C-terminal amino acid sequence of ricin B-chain from amino acid 427 to the stop codon (*). The alanine residue underlined in the PCR clone is an arginine residue in the corresponding ricin cDNA sequence (Lamb *et al.*, 1985). Differences in the nucleotide sequence compared with ricin are shown above the DNA sequence.



clone is a ricin-like sequence.

There are several factors which must be borne in mind when considering these results. Evidence is presented to show that the ricin PCR product is a ricin-like sequence, however there is insufficient information to say that it contains a sequence encoding a functional polypeptide. The base differences between the clone and the ricin cDNA sequence may be due to mis-incorporation of nucleotides by *Taq* polymerase during the amplification reaction. Saiki *et al.* (1988a) observed a cumulative error frequency of approximately 0.25% in a 30 cycle amplification reaction.

If the total population of product molecules are then subsequently analysed, ie by hybridisation or direct sequencing, the individual sequence errors are of no consequence. If a single clone of a PCR amplification is sequenced then the sequence will reveal its identity but the actual sequence of the clone may contain errors. To overcome this several clones produced from separate PCR amplifications must be sequenced independently.

In the case of the ricin PCR, only one clone was obtained due to the low efficiency of ligation. This most likely reflects inefficient restriction of the EcoRI and HindIII sites incorporated into the PCR product due to their close proximity to the end of the molecule (see Fig 4.10).

In order to determine if the ricin PCR clone, pRic1, is a functional coding sequence it was expressed *in vitro*. The results of this are presented in chapter 6 and show that it is probably functional.

4.6. General comments on the analysis of ricin and agglutinin gene sequences.

The experimental approaches described in this chapter to obtain coding sequences for ricin and agglutinin genes were only partially successful. Two ricin-like clones were isolated, one from a genomic library and the other using PCR amplification. Neither clone was analysed by further nucleotide sequencing of the coding regions. Having confirmed that the PCR amplifications used were specific the work on the ricin sequences was suspended in favour of investigations of saporin coding sequences. This work is described in the following chapter. The isolation of saporin genomic clones would increase the amount of data available on the structure and organisation of RIP genes, especially as saporin is a type I

RIP.

The work on ricin and agglutinin sequences was extremely valuable for the development of several techniques used subsequently to clone and isolate saporin gene sequences. Sequences were also isolated which will have utility in future research programmes. Familiarity with the handling and manipulation of gene libraries was obtained. The specificity and use of PCR amplifications was investigated and was to be of great use in further PCR strategies involving the cloning and expression of saporin gene sequences.

5. Results And Discussion. Isolation Of Saporin Gene Sequences.

5.1. Strategy for the isolation of saporin genomic clones.

The strategy employed to clone saporin genomic clones was based on the production of an amplified saporin-specific gene probe using PCR. Partial amino acid sequence data was available for the saporin-6 protein (Fordham-Skelton *et al.*, 1990) of sufficient length to design redundant oligonucleotide primers to amplify a fragment of the saporin gene coding sequence. Given that this sequence was an authentic amplification product it could then be used to screen a *Saponaria* genomic library for the corresponding full-length genomic clones.

This approach, using redundant oligonucleotide primers derived from amino acid sequences has been successful in amplifying gene fragments. Lee *et al.* (1988) amplified part of the porcine urate oxidase gene using cDNA as a template. Gould *et al.* (1989) obtained part of the iron-sulphur protein gene of succinate dehydrogenase using genomic DNA and cDNA templates from several species. In the above cases degenerate primers were able to direct specific amplification of target sequence. This is of significance since the original PCR experiments had utilised only single species primers.

A further advance was to reduce the degree of redundancy of the primers by introducing the neutral base inosine at ambiguous positions in the primer sequence (Ohtsuka *et al.*, 1985). Inosine has been incorporated into degenerate oligonucleotides used as a hybridisation probe to screen a human genomic library and obtain a clone for cholecystokinin, a brain/gut peptide (Takahashi *et al.*, 1985). Inosine containing oligonucleotides, used as PCR primers, have been used to successfully amplify; a fragment of the voltage-sensitive sodium channel gene from rat brain cDNA (Knoth *et al.*, 1988); serine protease gene fragments from genomic DNA of the parasitic nematode *Anisakis simplex* (Sakanari *et al.*, 1989) and a prokaryotic metallothionein gene amplified from *Anacystis nidulans* genomic DNA (Robinson *et al.*, 1990).

Previous attempts to screen *Saponaria* genomic and cDNA libraries with short redundant oligonucleotide probes derived from partial amino acid sequences of saporin-6 were unsuccessful (Taylor P, Department of Biological Sciences, University of Durham, unpublished results). Benatti *et al.* (1989) also reported a failure to obtain saporin clones from a leaf

cDNA library screened with a short redundant oligonucleotide probe. To this end the PCR strategy, described above, was designed to produce a probe for saporin genes which would be both sufficiently long and specific for effective use in polynucleotide hybridisation reactions.

5.2. Production of a saporin gene probe.

Partial amino acid sequences obtained for saporin-6 protein are shown in Fig 5.1. The protein sequencing was carried out by Dr A Yarwood, Department of Biological Sciences, University of Durham. N-terminal sequencing of the mature protein and isolated tryptic and peptic peptides produced an N-terminal sequence 48 residues. This confirms and extends the previously reported N-terminal sequence data of Lappi *et al.* (1985) for saporin-6. The internal, overlapping sequence, was derived from the sequences of further isolated tryptic and peptic peptides. This sequence was at an indeterminate position within the protein. Other, shorter peptide sequences, could not be aligned and are not presented.

Fig 5.1 also shows the design of degenerate, inosine-containing, oligonucleotides for use as primers in PCR amplifications. The computer programme PROBFIND (Lewis, 1986) was used to reverse translate the peptide sequences to DNA sequences and select oligonucleotides of minimal redundancy. Inosine-containing primers were synthesised corresponding to the two peptide regions shown in Fig 5.1. Inosine was incorporated at base positions with three- or four-fold redundancy in the third base position of a codon. It was also incorporated at several A/G base choices but not at C/T choices. Inosine can form hydrogen bonds with A, T or C but not efficiently with G (Ohtsuka *et al.*, 1985). At the C/T choices there is a 50% probability that a G will be present at the same position in the complementary strand. Hence it was decided to leave the C/T choice in the sequence rather than substitute with inosine. Inosine has been shown to act primarily as a guanosine (Nordmann *et al.*, 1988) and an I/G match may destabilise a primer/template duplex.

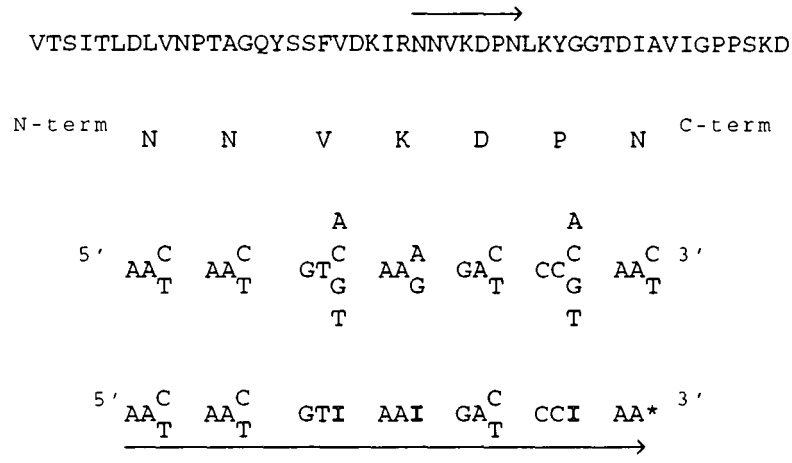
5.2.1. Amplification of saporin coding sequence fragments using PCR.

The inosine-containing primers were used in PCR amplifications with genomic DNA as a template as described in Fig 5.2. Standard PCR amplifications use a primer concentration of 10 μM each (section 3.11.1). In this case, as the primers were degenerate, a higher concentration

Figure 5.1. Design of primers for the PCR amplification of saporin gene sequences.

Inosine-containing oligonucleotide primers were synthesised corresponding to the indicated peptide sequences of the mature saporin-6 protein. The computer programme PROBFIND (Lewis 1986) was used to reverse translate the protein sequence and locate oligonucleotides of minimal redundancy. The asterisk denotes where the base of a codon has been omitted from the final primer sequence design. Inosine residues in the final primer sequences are denoted by I in bold type.

1. N-TERMINAL SEQUENCE.



2. INTERNAL SEQUENCE.

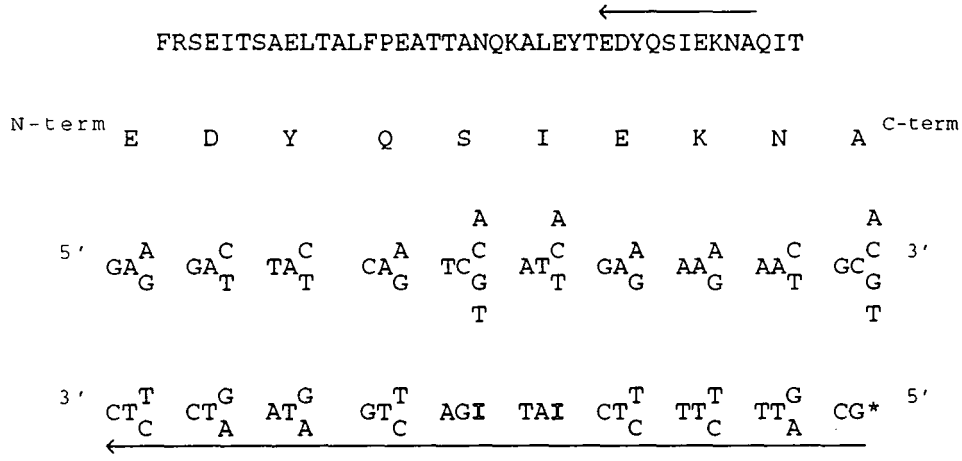
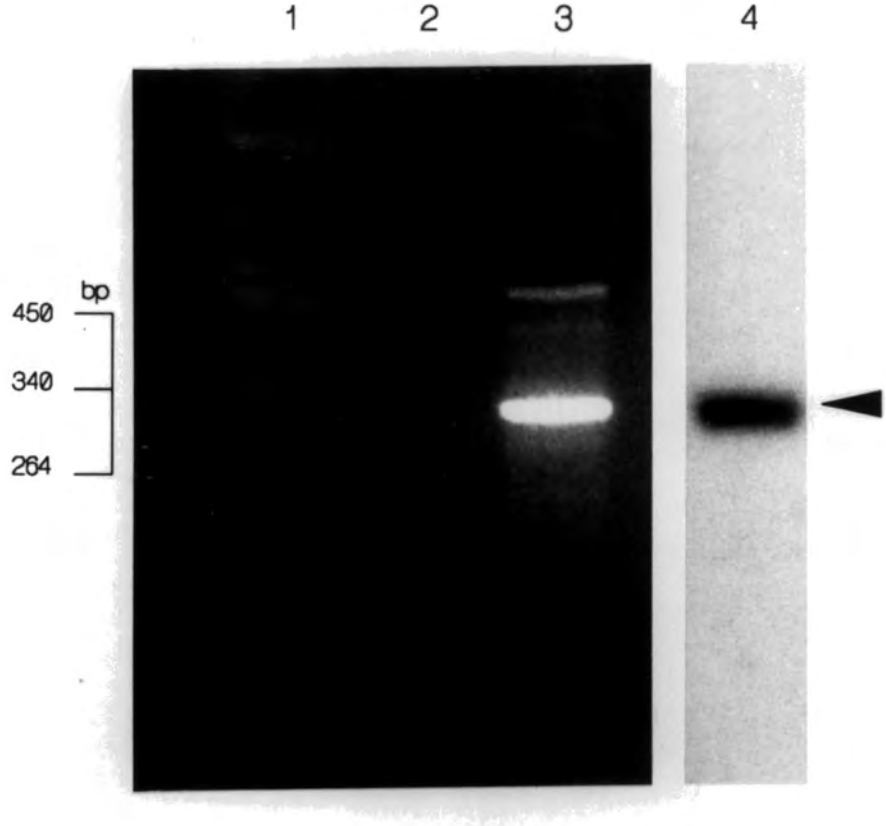


Figure 5.2. PCR DNA amplification using genomic DNA and primers derived from the partial amino acid sequence of saporin-6.

PCR amplification was carried out, using *Saponaria* genomic DNA as a template, with the inosine-containing primers described in Fig. 5.1. 100 μl reactions contained 2.5 μg of genomic DNA and 20 μg of each primer. The reaction mix was heated to 94^0C for 4 min followed by the addition of 5.0 units of *Taq* polymerase. Reactions were then cycled automatically through 30 cycles of: denaturation 91^0C/2 min, annealing 45^0C/2 min and extension 72^0C/3 min. The final extension time was increased to 15 min to ensure completion of strand synthesis. 10 μl of the reaction mix was electrophoresed on a 2% (w/v) agarose gel. Lanes were:-

1. λ/PstI size markers
2. 10 μl control PCR mix, no *Taq* polymerase added
3. 10 μl of reaction mix with *Taq* polymerase
4. Re-amplification of the 300 bp PCR product in the presence of [α-^{32}P]dCTP

The major amplification product, of a size approximately 300 bp, is indicated in track 3. The 300 bp sequence was radioactively labelled by re-amplification as described in section 3.11.2. Electrophoretic analysis of the product is shown in lane 4. The gel was dried and autoradiographed. The only labelled DNA species detected was the 300 bp sequence.


1
2
3
4
bp
450
340
264

(20μg) was used as recommended by Girgis *et al.* (1988). The major product is approximately 300 bp in size (lane 3). Under the conditions used several minor bands were reproducibly amplified but were not subjected to further analysis.

The 300 bp fragment was isolated from an agarose gel. Lane 4 of Fig 5.3 shows an autoradiograph of the products of re-amplification of the isolated 300 bp product labelled with [α-^{32}P]dCTP (section 3.11.2). The 300 bp fragment was the only labelled species detectable and indicated that it could probably be used as a hybridisation probe.

5.2.2. Verification of saporin probe identity.

The size of the fragment produced from the PCR amplification provided no indication of its identity as the distance between the two primer sequences was unknown. In the case of the ricin amplifications the expected product size of 1.8 kb was known. To confirm the specificity of the saporin amplification the 300 bp fragment was cloned and sequenced as described in Fig 5.3. The labelled 300 bp fragment was used as a probe in a colony hybridisation to successfully isolate three clones containing the saporin PCR product. (Standard hybridisation conditions were used with a final post-hybridisation wash of 0.1xSSC at 65^{0}C for 30 min).

5.2.2.1. Estimation of the saporin probe specific activity obtained by re-amplification.

An estimation of the specific activity of the probe was carried out by determining the amount of [α-^{32}P]dCTP incorporated into the labelled fragment. The first peak of radioactivity was recovered when the re-amplified product was separated from unincorporated label using a Sephadex G-50 column. An aliquot was analysed by gel electrophoresis (lane 4, Fig 5.3) and the total counts present determined by liquid scintillation counting a second aliquot. The amount of DNA synthesised in the re-amplification can be approximately determined by calculating the incorporation of label. The amount of [α-^{32}P]dCTP incorporated was 18 ng (only labelled dCTP was included in the reaction). Assuming equal base composition of the probe the total amount of DNA synthesised was approximately 4x18 = 72 ng. As the reaction contained 10 ng of unlabelled template DNA, the actual amount present was about 80 ng (0.08 μg). The total number of counts determined was 5.4x10^{7} cpm. Therefore the specific

was approximately $5.4 \times 10^7 / 0.08 = 6.75 \times 10^8$ cpm/μg.

The specific activity obtained is sufficiently high to detect single copy sequences in a complex genome. The re-amplified DNA product was full-length, unlike the population of labelled fragments generated, using random priming or nick-translation protocols. This may increase the specificity of hybridisation, especially for relatively short polynucleotide probes.

The DNA sequence and deduced amino acid sequences of the three clones containing the 300 bp amplification fragment is shown in Fig 5.3. The sequence in each case is actually 308 bp in length encoding an open reading frame of 102 amino acid residues. Two previously unaligned peptide sequences from saporin-6 can be matched to the deduced amino acid sequence between the primers. Thus the sequences are confirmed as specific products of a saporin gene sequence(s).

Two of the sequences are virtually identical (Sap 2 and Sap3) whilst Sap 1 is highly similar (Fig 5.3). Sap3 has a tripeptide region differing from the other two clones. There are several amino acid differences between the three clones, conservative amino acid substitutions are highlighted. The arginine for lysine substitution at amino acid number 67 was also noted in the protein sequencing of the internal peptide of saporin-6, reflecting a small degree of heterogeneity within the protein sample. The differences found in the sequences suggest that more than one saporin genomic sequence served as a template in the amplification reaction.

The primer sequences incorporated into the product is shown overlined. Positions in the primer sequence originally containing inosine were subsequently changed to guanosine when the sequences were cloned. This is consistent with the results of Nordmann *et al.* (1988) who reported that inosine will primarily select and base pair with cytosine when present in a template for DNA synthesis. In this case the inosine bases are present in the initial plasmid/insert ligation product replicated once transformed into a bacterial host. This result confirms that inosine primarily acts as a guanosine during base pairing. The general utility of this technique is discussed further in section 5.6.

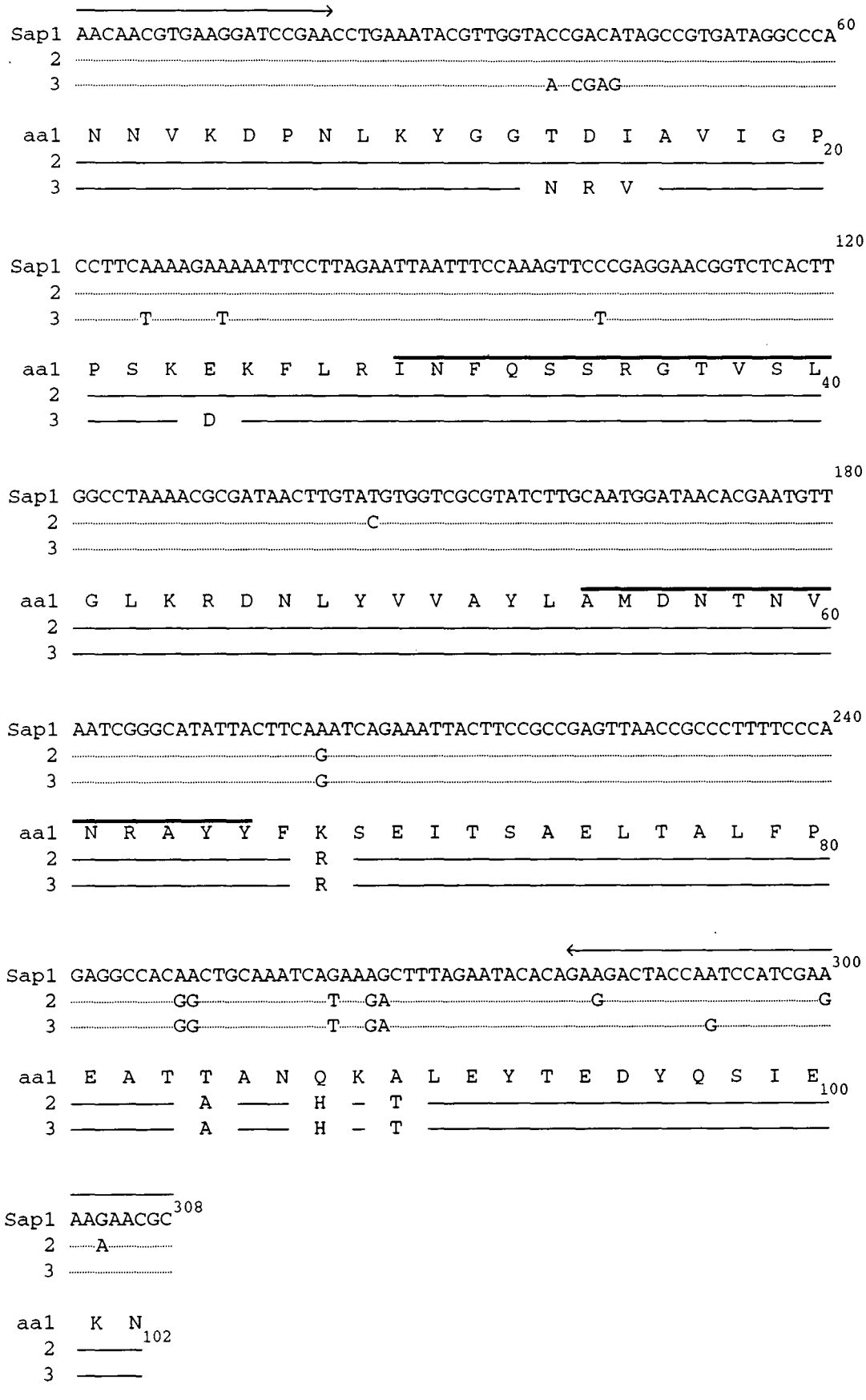
5.2.3. Southern analysis of *Saponaria* genomic DNA.

A Southern blot of *Saponaria* genomic DNA probed with the 300 bp labelled

Figure 5.3. Nucleotide sequence and deduced amino acid sequence of three plasmid clones containing the 300 bp saporin PCR product.

The 300 bp putative saporin PCR product obtained by the PCR amplification from genomic DNA, using saporin-specific primers, was cloned into the HincII site of pUC18 and transformed into *E. coli* DH5α. Three recombinant clones, containing the 300bp insert, were recovered when the transformants were colony screened with the labelled, re-amplified, 300 bp fragment. Each clone was subjected to plasmid sequencing using the M13 forward and reverse sequencing primers to confirm the identity of the saporin PCR product.

The nucleotide sequence of each of the three clones is shown. Only non-identical amino acids are given for the clones designated Sap 2 and 3. Previously unaligned peptide sequences, derived from the protein sequencing of saporin-6, are *overlined* the arrows indicate primer orientation.


Sap1 AACAACGTGAAGGATCCGAACCTGAAATACGTTGGTACCGACATAGCCGTGATAGGCCCA 60
2
3 A CGAG
aa1 N N V K D P N L K Y G G T D I A V I G P 20
2
3 N R V
120
Sap1 CCTTCAAAAGAAAAATTCCTTAGAATTAATTTCCAAAGTTCCCGAGGAACGGTCTCACTT
2
3 T T T
aa1 P S K E K F L R I N F Q S S R G T V S L 40
2
3 D
180
Sap1 GGCCTAAAACGCGATAACTTGTATGTGGTCGCGTATCTTGCAATGGATAACACGAATGTT
2 C
3
aa1 G L K R D N L Y V V A Y L A M D N T N V 60
2
3
Sap1 AATCGGGCATATTACTTCAAATCAGAAATTACTTCCGCCGAGTTAACCGCCCTTTTCCCA 240
2 G
3 G
aa1 N R A Y Y F K S E I T S A E L T A L F P 80
2 R
3 R
300
Sap1 GAGGCCACAACTGCAAATCAGAAAGCTTTAGAATACACAGAAGACTACCAATCCATCGAA
2 GG T GA G G
3 GG T GA G
aa1 E A T T A N Q K A L E Y T E D Y Q S I E 100
2 A H - T
3 A H - T
Sap1 AAGAACGC 308
2 A
3
aa1 K N 102
2
3

PCR probe is shown in Fig 5.4. The final washing stringency was sufficiently high (0.1xSSC) to allow only hybridisation of the probe to sequences with approximately 98% similarity or above. The autoradiograph shows several hybridising restriction fragments in each restriction digest. This result, together with the heterogeneity in the PCR derived sequences and the report of at least six isolated saporin fractions (Stirpe *et al.*, 1983) suggests that saporin-6 is a member of a small multigene family.

5.3. Screening of a *Saponaria* genomic library with the PCR-derived saporin gene probe.

The results of the Southern analysis of *Saponaria* genomic DNA showed that the saporin PCR sequence could be used as a hybridisation probe to specifically detect saporin-encoding sequences. A *Saponaria* genomic library (section 3.4.7), constructed using genomic DNA partially restricted with Sau3A, was screened for saporin genomic clones using the saporin-specific PCR probe.

Approximately 1.2×10^6 phage were plated on four large (22x22 cm) plates, transferred to duplicate nitrocellulose filters and hybridised with the re-amplified saporin PCR probe. Standard hybridisation conditions were used with a four post-hybridisation washes of 2xSSC at 65^0C, each for 15 min. After autoradiography positively hybridising plaques, coincidental on both filters, were isolated.

Hybridising clones were divided into three arbitrary classes on the basis of intensity of hybridisation to the probe; strong, medium and weak containing 11, 3 and 4 clones respectively. Clones from each of the hybridisation classes were subjected to further rounds of plaque purification (6 strong, 1 medium and 1 weak). The secondary screening employed a final wash stringency of 1.0xSSC, following rounds of plaque purification employed final wash stringencies of 0.1xSSC to ensure only clones highly similar to the probe would hybridise.

5.4. Isolation and characterisation of two truncated saporin genomic clones.

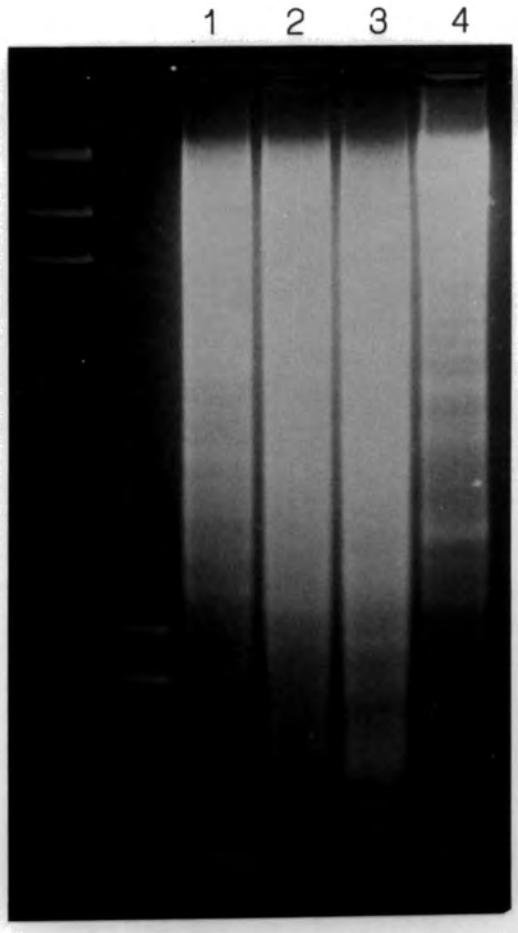
Two putative saporin-containing genomic clones were plaque purified to homogeneity and DNA prepared from them using the "lambdasorb" method described in section 3.3.3.2. Both clones, designated λSap3 and λSap4,

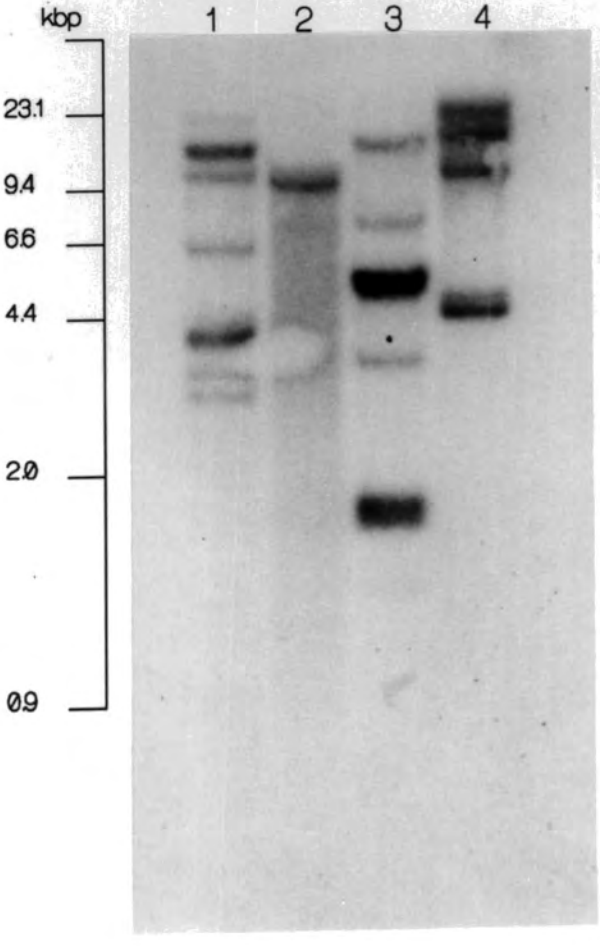
Figure 5.4. Southern analysis of *Saponaria* genomic DNA.

10 μg of *Saponaria* genomic DNA was restricted with the following enzymes:-

1. EcoRI
2. BamHI
3. Hind III
4. BglII

Restriction fragments were separated by electrophoresis on a 0.8% (w/v) agarose gel and transferred to a nylon filter. Specific fragments were detected by hybridisation with the labelled 300 bp saporin sequence obtained using PCR. The final post-hybridisation wash was at 65^{0}C for 15 min in 0.1xSSC, followed by autoradiography of the filter. DNA size markers were λ/HindIII and pBR322/AluI.


1
2
3
4


kbp
1
2
3
4
23.1
9.4
6.6
4.4
2.0
0.9

were selected from the strongly hybridising class of clones isolated from the initial library screening. DNA from each clone was subject to preliminary restriction and Southern analysis as described in Fig 5.5.

The autoradiograph shows strong hybridisation of the probe to a restriction fragment in each digest of the two clones. The stringency of washing indicated that the clones contained sequences very similar to the saporin PCR probe.

λSap3 contains an insert of approximately 14 kb whilst the insert from λSap4 is approximately 12 kb, determined by restriction with SalI which excises the insert from the vector sequences. The smallest hybridising sequence present is a 6.6 kb HindIII restriction fragment of λSap4. Further restriction analysis was carried out on each clone to determine suitable fragments for subcloning and sequencing.

Restriction mapping of the clones indicated that the hybridising regions were in each case adjacent to the left arm of EMBL3 vector. In λSap3 the probe hybridised close to the left arm as shown in Fig 5.7. In λSap4 the hybridisation was again adjacent to the left arm of the vector (Fig 5.9). The SalI site present at the end of each of the vector arms was therefore used to find suitable restriction fragments for subcloning. Double digestion with SalI and one other restriction enzyme was used to determine a suitable fragment for subcloning. This was achieved by a single digest of the clone with a restriction enzyme, producing compatible sites for cloning into pUC18, followed by a double digestion using SalI and the same enzyme. Where the hybridising fragment was reduced to a smaller size (<6 kb) upon double digestion then it was selected as suitable for subcloning.

5.4.1. Characterisation of λSap3.

The clone λSap3 was subjected to further restriction and Southern analysis (Fig 5.6). As shown in Fig 5.6, single digestion with SacI produced a 23 kb hybridising fragment (lane 7) whereas restriction with SacI and SalI produced a 1.2 kb hybridising fragment (lane 8). pUC18.XhoI, modified by the inclusion of a XhoI site (see section 2.4), was used to clone the 1.2 kb SalI/SacI fragment of λSap3.

The vector was restricted with XhoI (producing SalI compatible ends) and SacI. λSap3 was restricted with SalI and SacI and ligated into the restricted plasmid. *E. coli* DH5α transformants containing the products of

Figure 5.5. Restriction analysis of two λ clones containing saporin gene sequences.

Two λ clones which positively hybridised to the saporin PCR probe were purified to homogeneity and λ DNA was prepared. DNA from each clone, designated λSap3 and λSap4, was restricted with the following enzymes:-

1. SalI
2. EcoRI
3. BamHI
4. Hind III
5. BglII

Restriction fragments were separated by electrophoresis on a 0.8% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific fragments were located by hybridisation with the labelled saporin PCR probe. The filter was washed to a final stringency of 0.1xSSC for 30 min at 65^{0}C and autoradiographed. DNA size markers were λ/HindIII and pBR322/AluI.

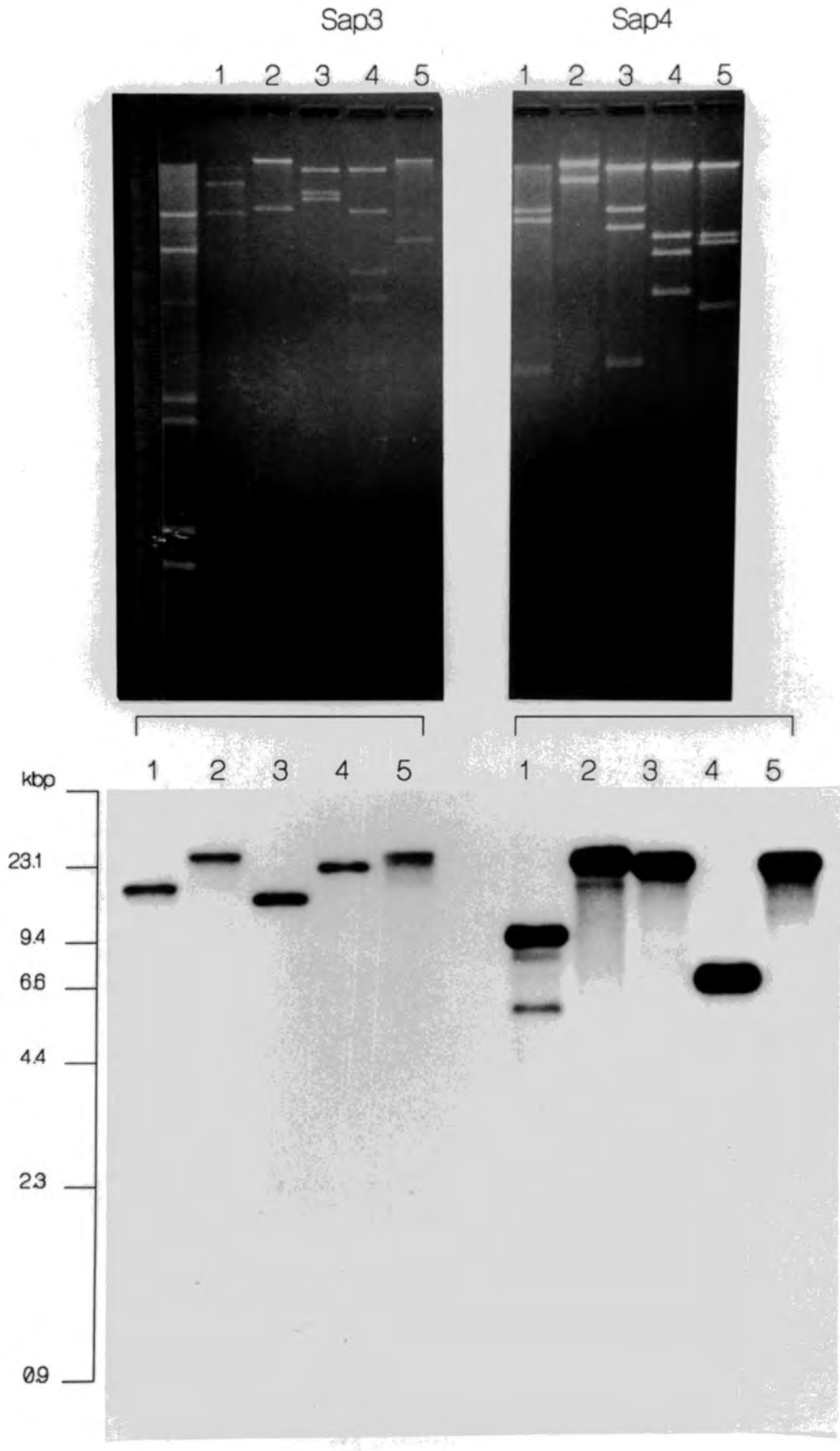

Sap3
Sap4
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
1 2 3 4 5
kbp
23.1
9.4
6.6
4.4
2.3
0.9

Figure 5.6. Further restriction analysis of λSap3.

To locate convenient fragments for the cloning and sequencing of the saporin gene sequence DNA was restricted with several enzyme combinations as follows:-

1. KpnI
2. SalI/KpnI
3. PstI
4. SalI/PstI
5. XbaI
6. SalI/XbaI
7. SacI
8. SalI/SacI

Restriction fragments were separated by electrophoresis on a 0.8% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific sequences were detected by hybridisation with the labelled saporin PCR probe as described in Fig. 5.5. The SalI/SacI fragment, arrowed on the autoradiograph, was selected as a suitable fragment for sub-cloning. DNA size markers were λ/HindIII and pBR322/AluI.

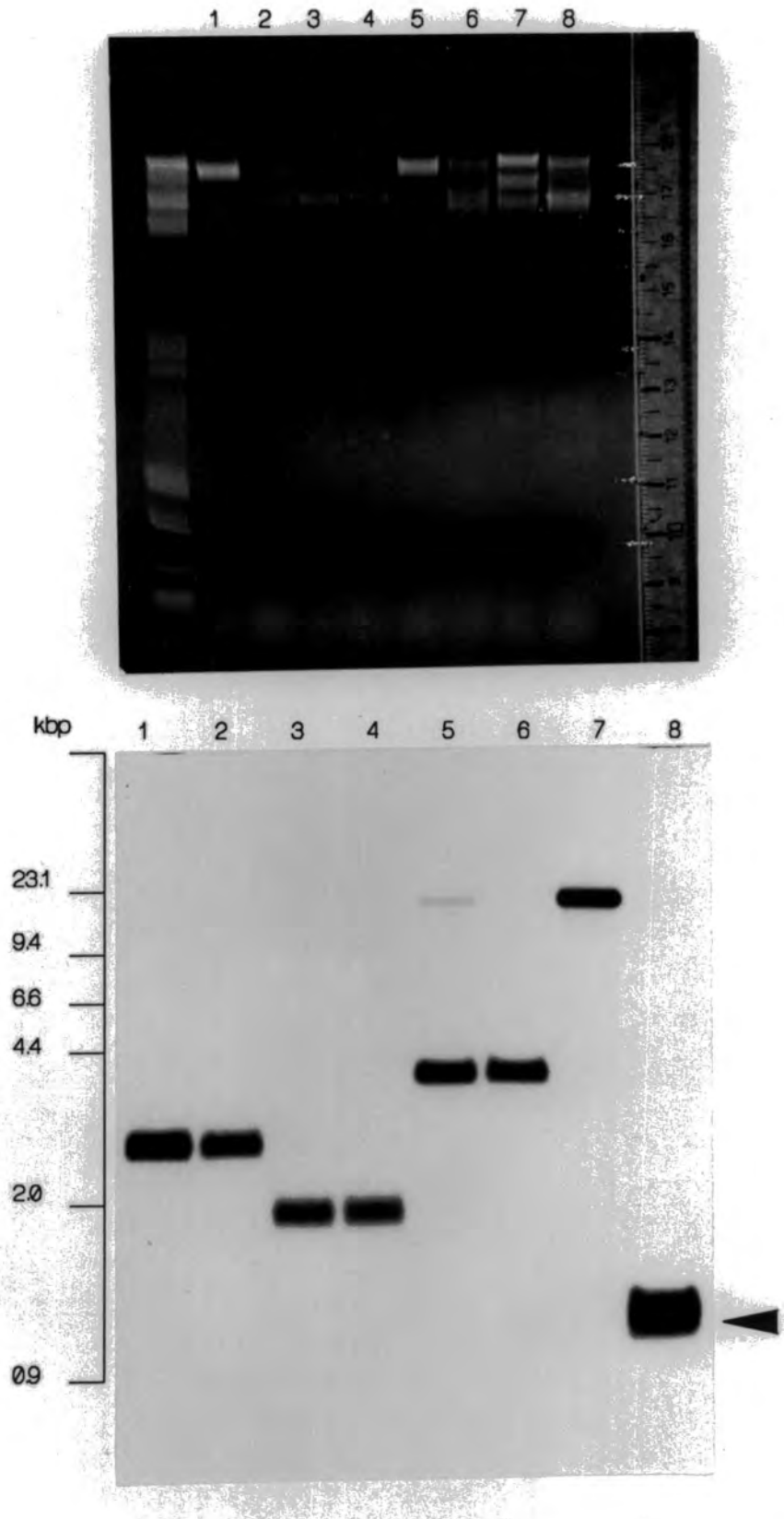
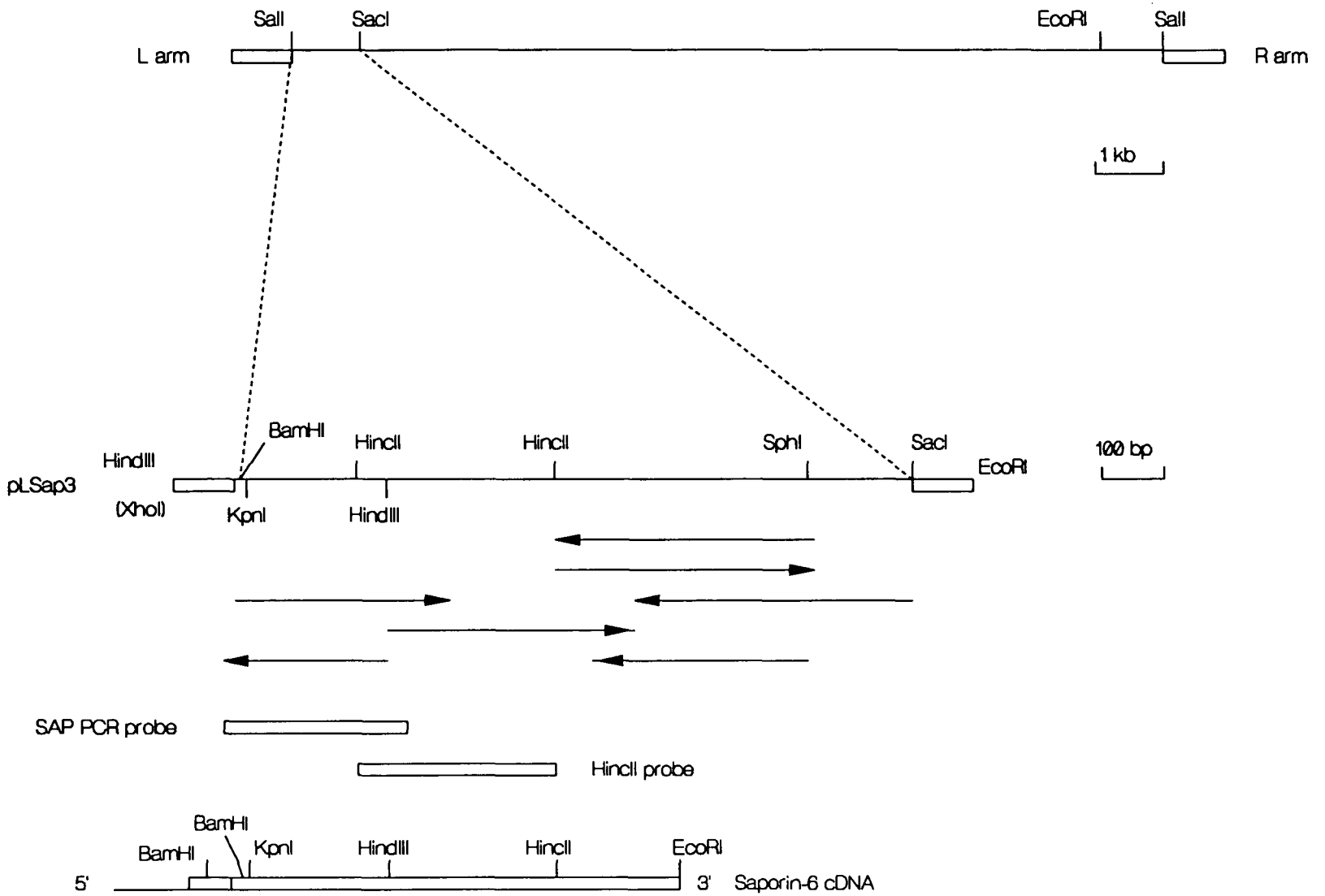

1 2 3 4 5 6 7 8
kbp
1 2 3 4 5 6 7 8
23.1
9.4
6.6
4.4
2.0
0.9

Figure 5.7. Restriction map of λSap3.

The figure shows selected restriction sites present in λSap3 with the sub-cloning and sequencing strategy used to obtain the saporin sequence present in this clone. The extent of the truncated saporin gene sequence is indicated, as is the region homologous to the saporin PCR probe used to obtain this clone. The restriction map of the partial cDNA clone for saporin-6 is shown for comparison (Benatti *et al.*, 1989). The HincII restriction fragment used subsequently as a probe to isolate a full-length saporin clone is indicated.


SalI
SacI
EcoRI
SalI
L arm
R arm
1 kb
BamHI
HincII
HincII
SphI
SacI
HindIII
EcoRI
100 bp
pLSap3
(XhoI)
KpnI
HindIII
SAP PCR probe
HincII probe
BamHI
BamHI
KpnI
HindIII
HincII
EcoRI
5'
3'
Saporin-6 cDNA

the ligation reaction were subjected to colony hybridisation using the saporin PCR probe to select positive clones containing the 1.2 kb SalI/SacI fragment. Plasmid DNA was isolated from a positively hybridising colony. This clone contained the correct size of insert and was designated pλSap3. The restriction map of pλSap3 is shown in Fig 5.7 along with the strategy used to determine its nucleotide sequence using M13 subclones. The extent and direction of sequencing of each subclone is denoted by an arrow.

The figure also indicates the region of homology to the PCR probe and a partial saporin-6 cDNA sequence determined by Benatti *et al.* (1989). The nucleotide and deduced amino acid sequence is presented in Fig 5.12 and discussed in section 5.5. Preliminary restriction mapping and plasmid sequencing of the clone showed that this clone contained a saporin coding sequence truncated at its 5'-end.

5.4.2. Characterisation of λSap4.

The strategy to subclone the hybridising fragment of λSap4 was the same as that used in producing the subclone of λSap3. Restriction and Southern analysis of λSap4 is shown in Fig 5.8. Lane 3 shows a hybridising 1.2 kb SalI/SacI selected as a suitable fragment for subcloning. In this case the probe used was the insert of pλSap3, excised as an EcoRI/SacI fragment, and labelled by the random priming method.

A SalI/SacI digest of λSap4 was cloned into the pUC18.XhoI vector. A clone containing the 1.2 kb fragment was isolated by colony hybridisation, using the labelled insert from pλSap3 as a probe, in the same manner as described previously

The restriction map of pλSap4 is shown in Fig 5.9 along with the region of homology to the saporin PCR probe and the saporin-6 cDNA sequence of Benatti *et al.* (1989). M13 subclones of pλSap4 were used to determine part of its nucleotide sequence. The extent and direction of sequencing for each subclone is denoted by the arrows.

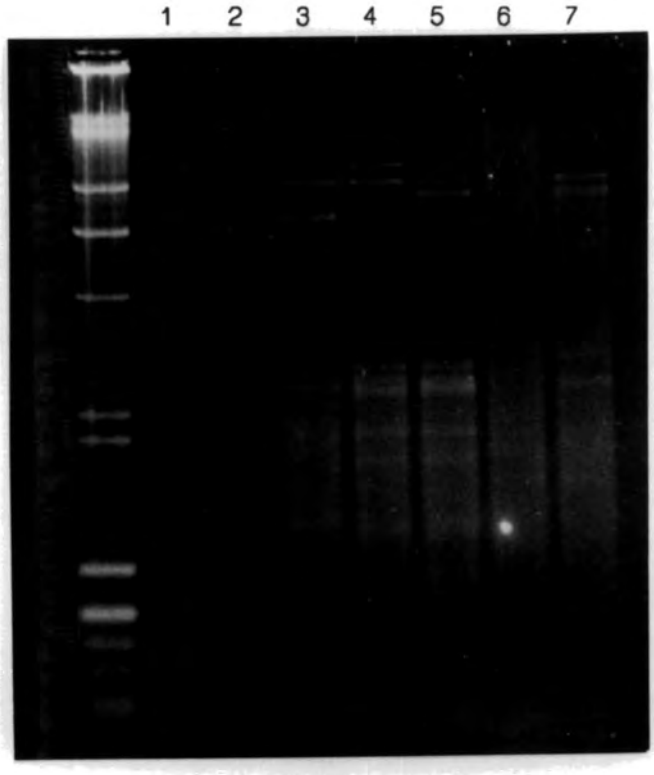
The nucleotide and deduced amino acid sequence is presented in Fig 5.13 and discussed in section 5.5. Initial plasmid sequencing of pλSap4 showed that this clone was also a truncated saporin coding sequence but in contrast to pλSap3, this clone was truncated at its 3'-end.

Figure 5.8. Further restriction analysis of λSap4.

λSap4 was subjected to restriction analysis in the same manner as described for λSap3 (Fig. 5.6) in order to locate suitable restriction enzyme fragments for sub-cloning and sequencing. The enzymes used were:-

1. SalI/EcoRI
2. SacI
3. SalI/SacI
4. PstI
5. SalI/PstI
6. SalI/KpnI
7. KpnI

Restriction fragments were separated by electrophoresis on a 0.8% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific sequences were detected by hybridisation with the EcoRI/SacI fragment of pλSap3 (see Fig. 5.7). The probe contains both the saporin coding sequence and 3′ untranslated region. The filter was washed to a final stringency of 0.1×SSC for 30 min at 65^{0}C. The arrowed SalI/SacI restriction fragment was selected as a suitable fragment for sub-cloning. DNA size markers were λ/HindIII and pBR322/AluI.


1
2
3
4
5
6
7

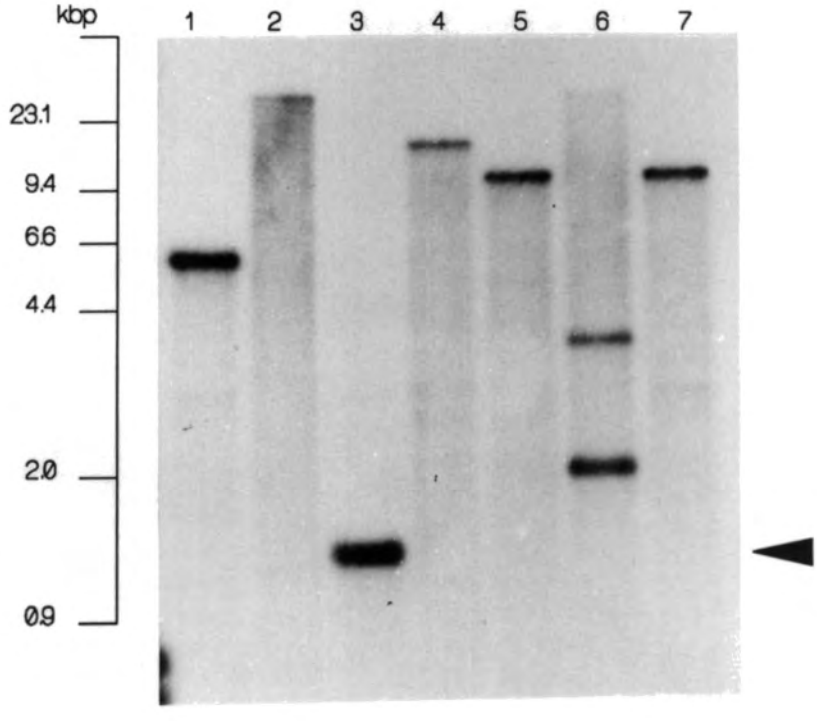
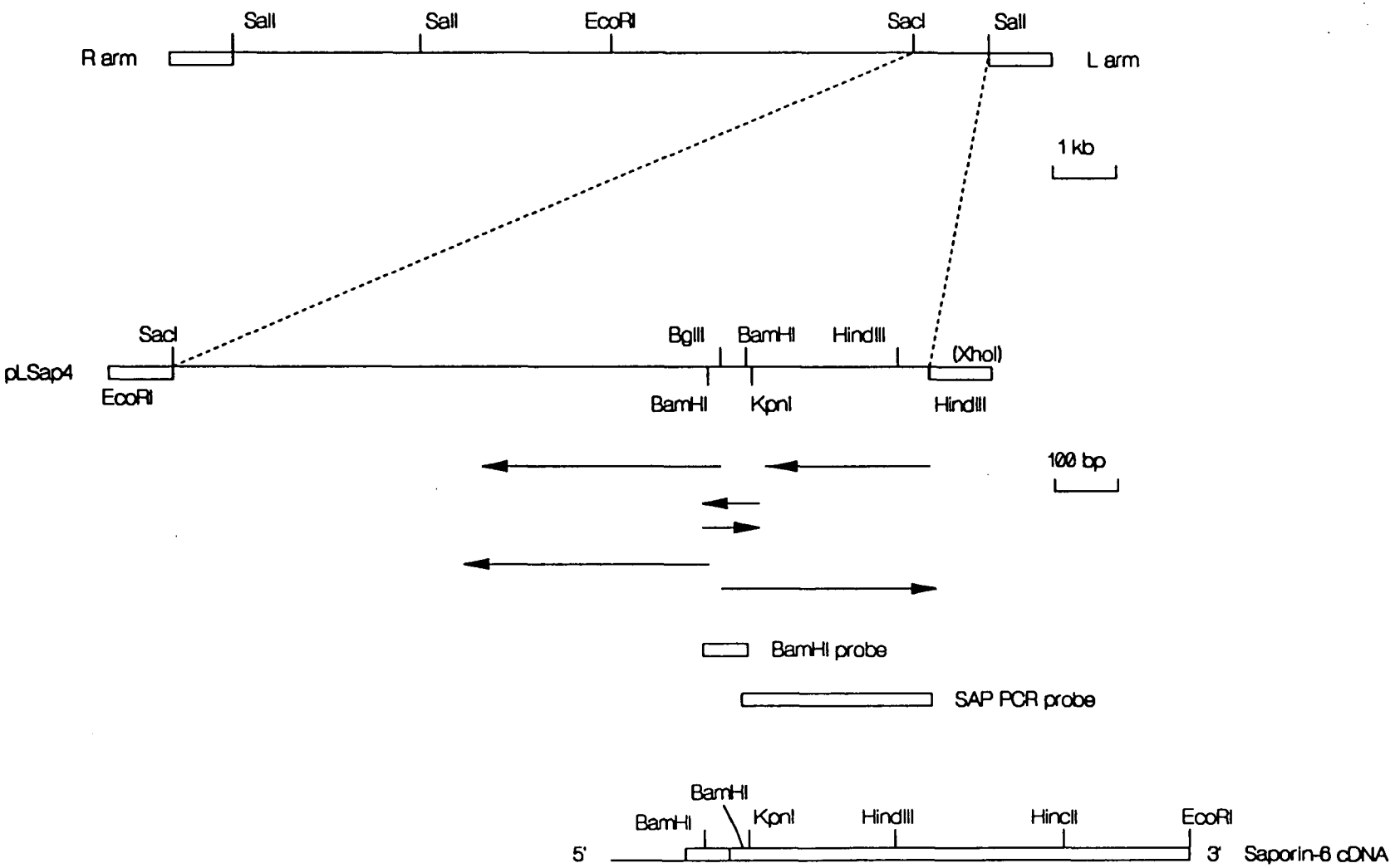

kbp
1
2
3
4
5
6
7
23.1
9.4
6.6
4.4
2.0
0.9

Figure 5.9. Restriction map of λSap4.

The restriction map shows selected restriction sites in the clone λSap4. The sub-cloning and sequencing strategies are indicated. The extent of the truncated saporin coding sequence is shown along with the region of homology to the saporin PCR probe. The restriction map of the partial saporin-6 cDNA is also shown for comparison. The BamHI restriction fragment used as a probe to subsequently isolate a full-length saporin gene is indicated.


SalI
SalI
EcoRI
SacI
SalI
R arm
L arm
1 kb
SacI
BglII
BamHI
HindIII
(XhoI)
pLSap4
EcoRI
BamHI
KpnI
HindIII
100 bp
BamHI probe
SAP PCR probe
BamHI
BamHI
KpnI
HindIII
HincII
EcoRI
5'
3'
Saporin-6 cDNA

5.4.3. Isolation and characterisation of a full-length saporin genomic clone.

Characterisation of the two saporin genomic subclones, pλSap3 and pλSap4, indicated that neither contained a full-length coding sequence. Each clone is truncated due to the partial Sau3A restriction of genomic DNA used to produce the genomic library. In each case the genomic fragment was restricted by Sau3A within the saporin coding sequence.

To obtain a full-length saporin gene the primary plaques isolated from the genomic library were re-screened with 5'- and 3'-specific probes obtained from the two truncated clones, pλSap3 and pλSap4. Positive hybridisation to both probes would indicate that the sequence was likely to be full-length.

The two probes used in this screening strategy were derived from either ends of pλSap3 and pλSap4. The probe specific for the 3'-end of the gene (corresponding to the C-terminal of the coding sequence) was a 320 bp HincII restriction fragment of pλSap3, see restriction map of Fig 5.7. The probe specific for the 5'-end of the gene (corresponding to the N-terminal of the coding sequence) was a 160 bp BamHI restriction fragment of pλSap4 shown in Fig 5.9.

11 strongly- and 3 less strongly hybridising λ clones were isolated from the initial screening of the *Saponaria* genomic library with the saporin PCR probe, of these two of the strongly hybridising clones were characterised and designated λSap3 and λSap4. The remaining twelve clones were plated out and plaques were transferred to duplicate nitrocellulose filters. These were hybridised to either the 5'- or 3'-specific probes, labelled by the random priming method. After hybridisation the filters were washed to a stringency of 0.1xSSC at 65^0C. Four clones hybridised only to the 3'-probe, one clone hybridised to both the 5'- and 3'-specific probes and seven clones did not hybridise to either probe. The clone hybridising to both probes was purified to homogeneity, DNA prepared from it, and designated λSap2. This clone contained a putative complete saporin coding sequence

The clones which did not hybridise to the probe probably contained saporin-related sequences. While these were sufficiently similar to the saporin PCR probe to hybridise during the initial screening of the library (washing stringency of 2xSSC) they did not hybridise to the saporin probe sequences under high stringency washing conditions (0.1xSSC). These

clones were not studied further.

Restriction and Southern analysis of λSap2 is shown in Fig 5.10. The autoradiographs of the Southern blot show the result of hybridising the 5'-probe (specific for the N-terminus of the coding sequence) (autoradiograph B). After removal of the probe the blot was re-hybridised with the 3'-probe (specific for the C-terminus of the coding sequence) (autoradiograph A).

Both probes hybridised to a 7.2 kb EcoRI fragment (lane 4, each blot). The 3'-probe hybridised to a 4.3 kb SalI/KpnI fragment (lane 9) whilst the 5'-probe hybridised to a 1.0 kb KpnI fragment (lanes 8 and 9) indicating that a KpnI site must be present in the clone between the two regions of probe hybridisation. A similar result is seen for the BglII restriction; the C-terminal probe hybridised to a 3.7 kb fragment (lane 10) whilst the N-terminal probe hybridised to a 10 kb fragment, indicating that a BglII site is also present between the two regions of probe hybridisation. Fig 5.9 shows the position of KpnI and BglII sites in pλSap4 and the KpnI site in the saporin-6 cDNA sequence. The hybridisation pattern of λSap2 is consistent with the position of these sites within a putative full-length saporin coding sequence present in λSap2.

The restriction map of λSap2 is shown in Fig 5.11. The subcloning and sequencing strategy for the region containing the saporin gene is also indicated. The saporin coding sequence, along with 5'- and 3'-flanking regions, was subcloned in pUC18. The subclones obtained were isolated by the "shotgun" cloning of λSap2 restricted with either BglII or EcoRI into the appropriately restricted pUC18 as described for λSap3 (section 5.4.1). Recombinant plasmids containing the 7.2 kb EcoRI fragment and 3.7 kb BglII fragment were recovered by colony hybridisation with the 3'-probe.

The recombinant plasmids obtained from this subcloning step were subjected to restriction analysis. The 3.7 kb BglII fragment was recovered as expected. The positively hybridising clone expected to contain the 7.2 kb EcoRI fragment in fact only contained an insert of 2.3 kb excisable with EcoRI. This truncated clone contained the characteristic 1.0 kb KpnI fragment, to which the 5'-specific probe hybridised, and the BglII site. This indicates that whilst the clone was truncated, probably due to EcoRI "star" activity, it did in fact contain part of the putative saporin coding sequence.

The two saporin subclones obtained by restriction of λSap2 with either

Figure 5.10. Restriction analysis of λSap2, a clone containing a full-length saporin gene.

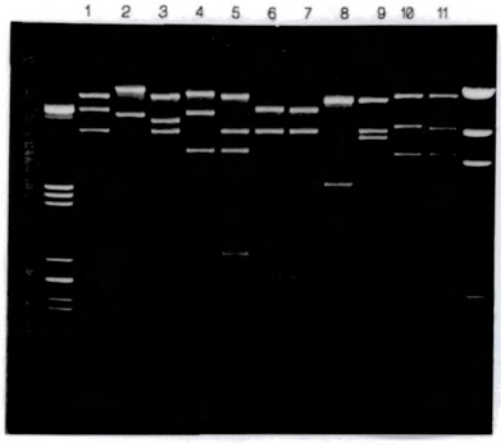
This clone was isolated using 2 probes: BamHI restriction fragment from pλSap4 and a HincII restriction fragment from pλSap3. These restriction fragments contain, respectively, coding sequence from the N- and C-terminal regions of the saporin gene sequence.

λSap2 hybridised with both probes and on the basis of this was assumed to contain a full-length saporin coding sequence. DNA was isolated from this clone and restricted with the following enzymes:-

1. SalI	5. SalI/EcoRI	9. SalI/KpnI
2. SacI	6. SalI/PstI	10. BglII
3. SalI/SacI	7. ——/——	11. SalI/BglII
4. EcoRI	8. KpnI	

Restriction fragments were separated by electrophoresis on a 0.9% (w/v) agarose gel and transferred to a nitrocellulose filter. Specific fragments were then detected by hybridisation with the 5'- and 3'-specific probes.

The filter was hybridised with the BamHI fragment of pλSap4 with a final post-hybridisation wash in 0.1xSSC for 30 min at 65^0C. The filter was autoradiographed whilst still moist (autoradiograph B). The probe was then removed by washing the filter four times for 15 min each in 0.1% (w/v) SDS at 65^0C. A second hybridisation was then carried out with the HincII fragment of pλSap3 using the same conditions as described above (autoradiograph A). DNA size markers were λ/PstI and λ/HindIII.


1 2 3 4 5 6 7 8 9 10 11

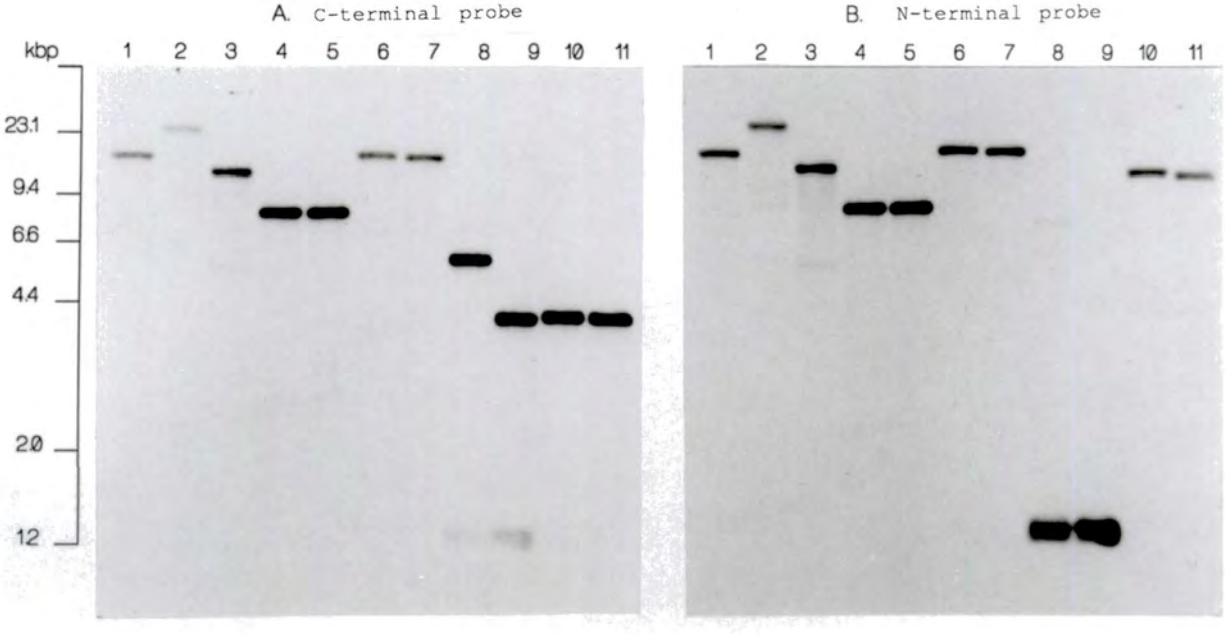
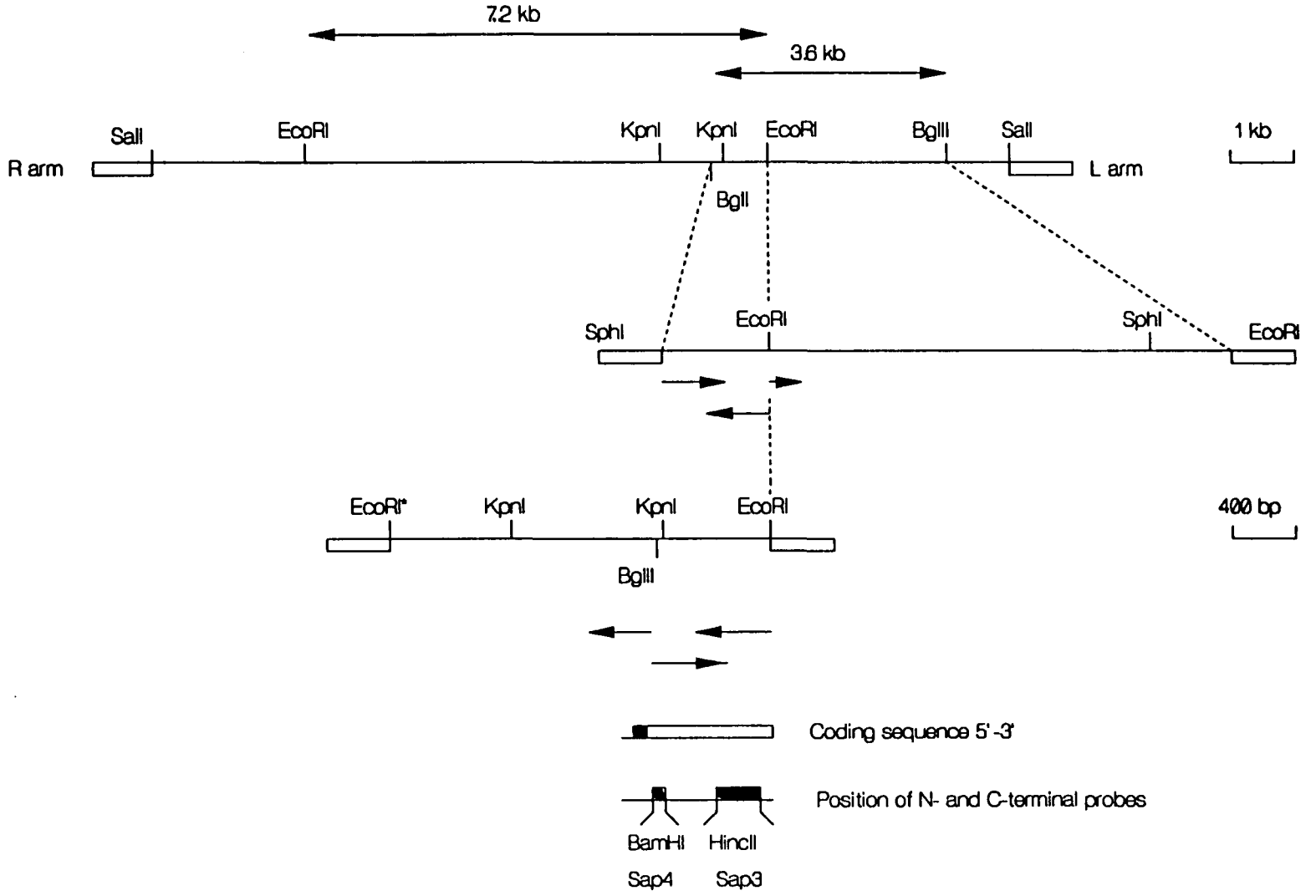

A. C-terminal probe
B. N-terminal probe
kbp
1 2 3 4 5 6 7 8 9 10 11
23.1
9.4
6.6
4.4
2.0
1.2

Figure 5.11. Restriction map of λSap2, a clone containing a full-length saporin gene.

The restriction map shows selected restriction sites within the clone and the strategy for subcloning and sequencing the saporin gene. The regions of homology to the N-terminal and C-terminal probes used to isolate this clone are shown. The position and orientation of the saporin coding sequence is indicated.


7.2 kb
3.6 kb
SalI
EcoRI
KpnI
KpnI
EcoRI
BglII
SalI
1 kb
R arm
L arm
BglI
SphI
EcoRI
SphI
EcoRI
EcoRI*
KpnI
KpnI
EcoRI
400 bp
BglII
Coding sequence 5'-3'
Position of N- and C-terminal probes
BamHI
HincII
Sap4
Sap3

EcoRI or BglII are shown in Fig 5.11. The two clones overlap in the sequence between the BglII site at the 5'-end of the coding sequence and the EcoRI site at the 3'-end of the gene. The EcoRI fragment contains the saporin promoter sequence and coding sequence extending to the EcoRI site at the genes 3'-end. The BglII fragment contains coding sequence from the BglII site at the 5'-end of the gene extending past the EcoRI site into the 3'-non-translated sequence.

Initial plasmid sequencing of the clone indicated that the EcoRI site within the coding sequence was at the same position as the EcoRI site in the saporin-6 cDNA characterised by Benatti *et al.* (1989). This site is at the extreme 3'-end of the gene. Hence the subcloned EcoRI fragment contains virtually the whole coding sequence of λSap2. The nucleotide sequence of the λSap2 gene is presented in Fig 5.14. The sequence was derived from M13 subclones of the EcoRI and BglII fragments, the extent and direction of sequencing are illustrated in Fig 5.11. The EcoRI and BglII fragments subcloned in pUC18 were designated pλSap2.E and pλSap2.B respectively.

5.5. Nucleotide sequences of members of the saporin gene family.

The screening of the *Saponaria* genomic library resulted in the isolation and characterisation of three distinct genomic clones. Each clone contained a saporin coding sequence; unfortunately two of these were truncated clones. λSap3 was truncated at the 5'-end of the coding sequence and λSap4 at the 3'-end of the coding sequence. The third clone (λSap2) contained a full-length saporin gene. The isolation of three independent genomic clones partly confirms the Southern analysis of *Saponaria* genomic DNA which had indicated that saporin genes comprise a multigene family.

There was no evidence for the presence of introns within the coding sequence of any of the clones. Introns are not present in the genomic clones of ricin (Halling *et al.*, 1985) and α-trichosanthin (Chow *et al.*, 1990). Many plant genes lack introns, for example: the zein gene family of maize (*Zea mays*) (Heidecker and Messing, 1986) and the seed lectin gene of pea (*Pisum sativum*) (Gatehouse *et al.*, 1987). It is possible that all RIP gene sequences are intron-less.

Whilst it is not known if these genomic sequences are actively expressed genes they all contain continuous open reading frames. The

Figures 5.12. Nucleotide and deduced amino acid sequence of pλSap3.
The nucleotide sequence of pλSap3 is presented in the 5′ to 3′ direction. Nucleotide numbering is shown above the sequence and the deduced amino acid sequence is numbered from the first codon present. The Sau3A restriction site at the 5′ end of the clone is indicated as are other restriction sites referred to in the text. An "in-frame" translational stop codon is denoted by an asterisk and potential polyadenylation sites (AATAAA) are shown in the 3′ nontranslated sequence.

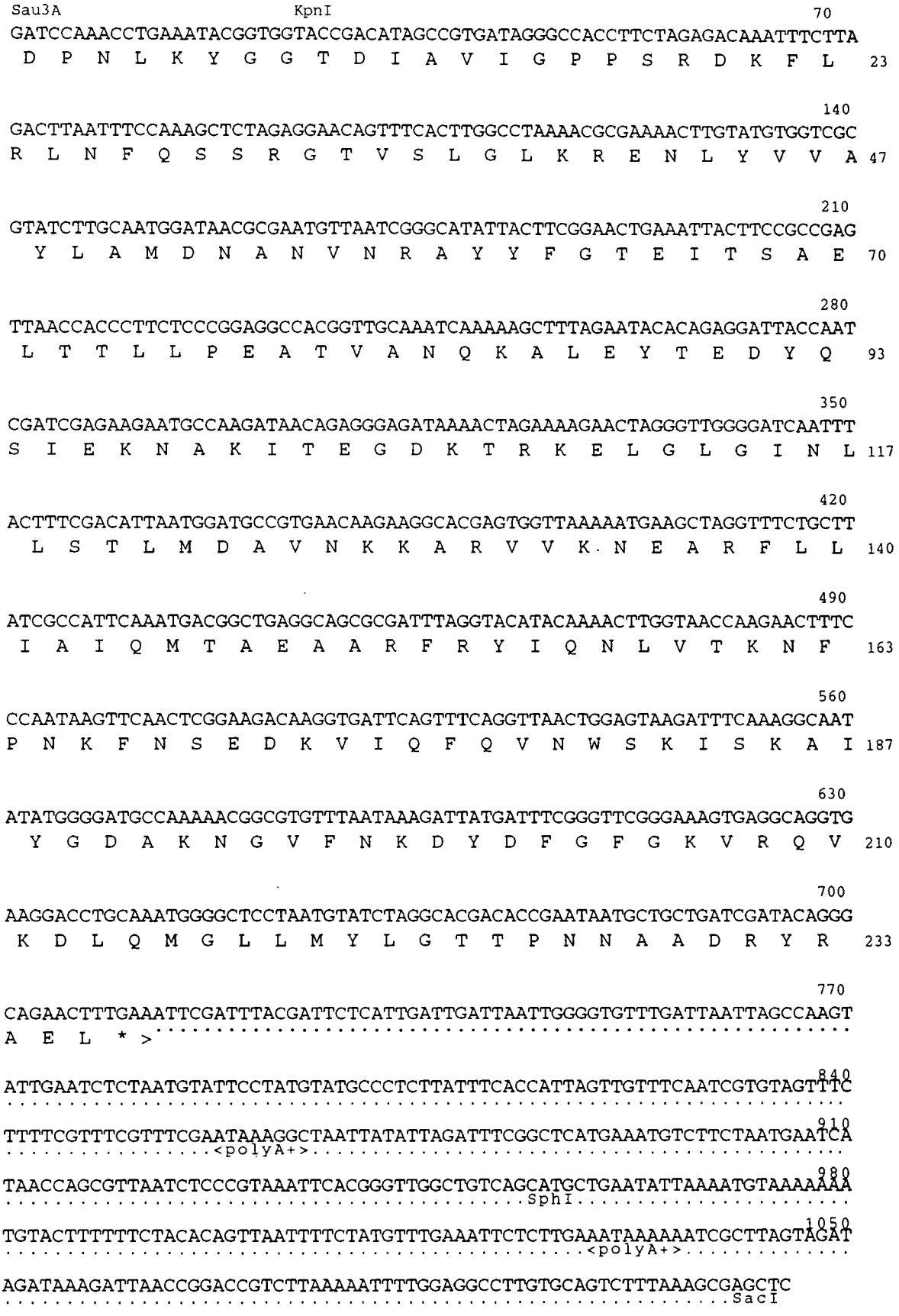

Sau3A
KpnI
70
GATCCAAACCTGAAATACGGTGGTACCGACATAGCCGTGATAGGGCCACCTTCTAGAGACAAATTTCTTA
D P N L K Y G G T D I A V I G P P S R D K F L 23
140
GACTTAATTTCCAAAGCTCTAGAGGAACAGTTTCACTTGGCCTAAAACGCGAAAACTTGTATGTGGTCGC
R L N F Q S S R G T V S L G L K R E N L Y V V A 47
210
GTATCTTGCAATGGATAACGCGAATGTTAATCGGGCATATTACTTCGGAACTGAAATTACTTCCGCCGAG
Y L A M D N A N V N R A Y Y F G T E I T S A E 70
280
TTAACCACCCTTCTCCCGGAGGCCACGGTTGCAAATCAAAAAGCTTTAGAATACACAGAGGATTACCAAT
L T T L L P E A T V A N Q K A L E Y T E D Y Q 93
350
CGATCGAGAAGAATGCCAAGATAACAGAGGGAGATAAAACTAGAAAAGAACTAGGGTTGGGGATCAATTT
S I E K N A K I T E G D K T R K E L G L G I N L 117
420
ACTTTCGACATTAATGGATGCCGTGAACAAGAAGGCACGAGTGGTTAAAAATGAAGCTAGGTTTCTGCTT
L S T L M D A V N K K A R V V K N E A R F L L 140
490
ATCGCCATTCAAATGACGGCTGAGGCAGCGCGATTTAGGTACATACAAAACTTGGTAACCAAGAACTTTC
I A I Q M T A E A A R F R Y I Q N L V T K N F 163
560
CCAATAAGTTCAACTCGGAAGACAAGGTGATTCAGTTTCAGGTTAACTGGAGTAAGATTTCAAAGGCAAT
P N K F N S E D K V I Q F Q V N W S K I S K A I 187
630
ATATGGGGATGCCAAAAACGGCGTGTTTAATAAAGATTATGATTTCGGGTTCGGGAAAGTGAGGCAGGTG
Y G D A K N G V F N K D Y D F G F G K V R Q V 210
700
AAGGACCTGCAAATGGGGCTCCTAATGTATCTAGGCACGACACCGAATAATGCTGCTGATCGATACAGGG
K D L Q M G L L M Y L G T T P N N A A D R Y R 233
770
CAGAACTTTGAAATTCGATTTACGATTCTCATTGATTGATTAATTGGGGTGTTTGATTAATTAGCCAAGT
A E L * >
840
ATTGAATCTCTAATGTATTCCTATGTATGCCCTCTTATTTCACCATTAGTTGTTTCAATCGTGTAGTTTC
910
TTTTCGTTTCGTTTCGAATAAAGGCTAATTATATTAGATTTCGGCTCATGAAATGTCTTCTAATGAATCA
<polyA+>
980
TAACCAGCGTTAATCTCCCGTAAATTCACGGGTTGGCTGTCAGCATGCTGAATATTAAAATGTAAAAAAA
SphI
1050
TGTACTTTTTTCTACACAGTTAATTTTCTATGTTTGAAATTCTCTTGAAATAAAAAATCGCTTAGTAGAT
<polyA+>
AGATAAAGATTAACCGGACCGTCTTAAAAATTTTGGAGGCCTTGTGCAGTCTTTAAAGCGAGCTC
SacI

Figure 5.13. Nucleotide and deduced amino acid sequence of pλSap4.
The nucleotide sequence of pλSap4 is presented in the 5′ to 3′ direction. Nucleotides are numbered above the sequence starting at the ATG translational initiation codon the sequence 5′ upstream from this is numbered in a negative manner. Amino acids are numbered from the first residue of the mature protein, the signal peptide residues (marked S) 5′ to the first residue of the mature protein are numbered in a negative manner. Relevant restriction sites are also indicated.

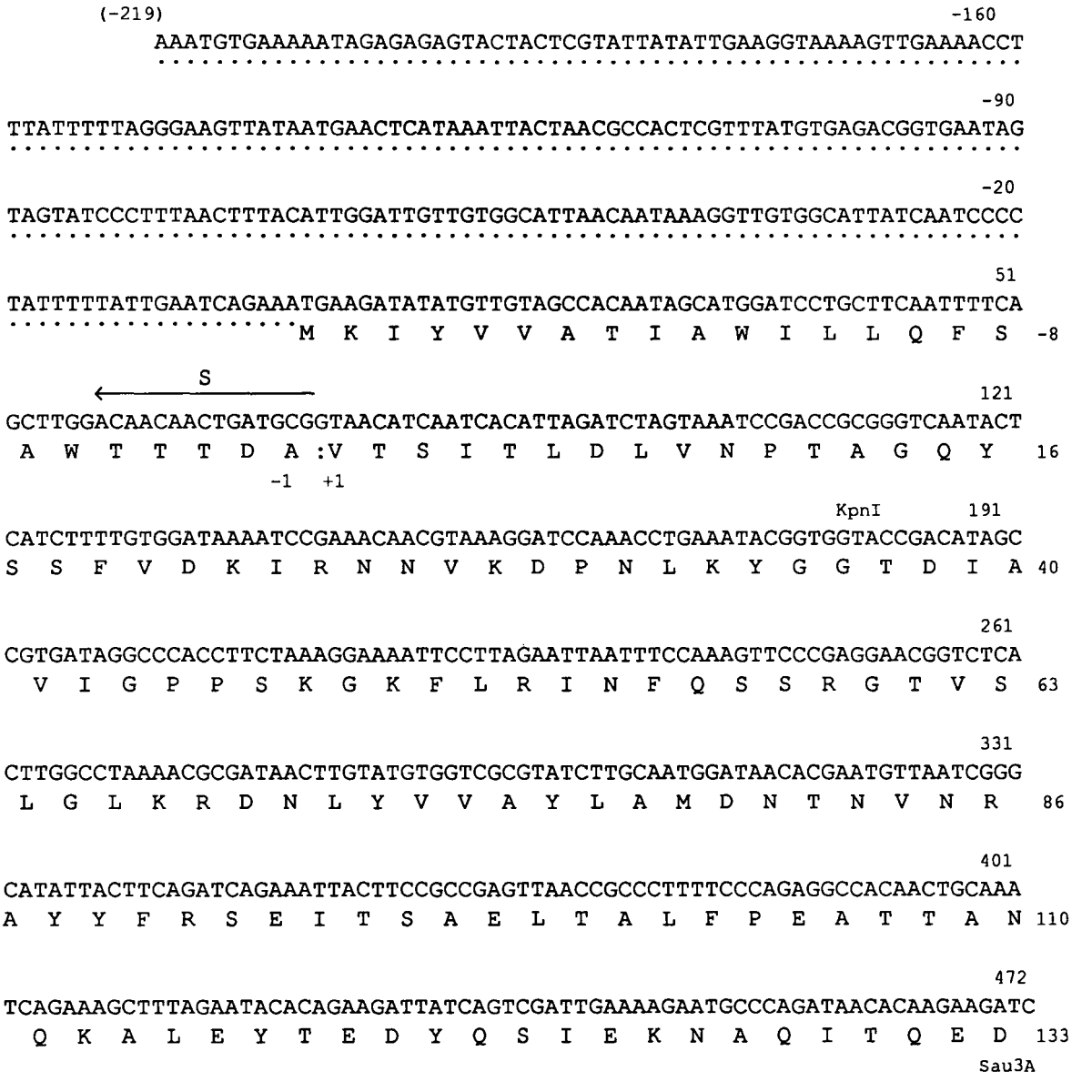
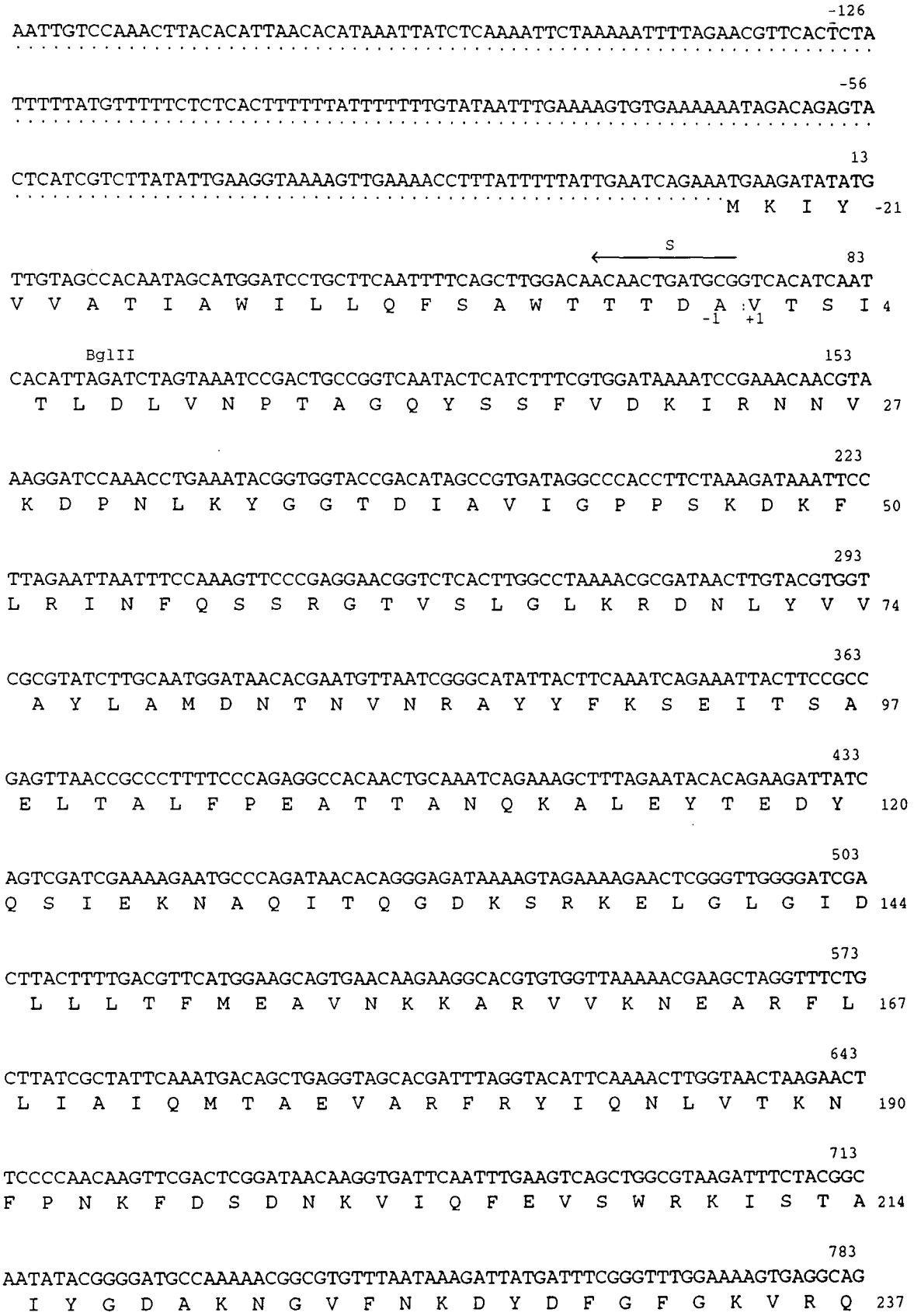

(-219)
-160
AAATGTGAAAAATAGAGAGAGTACTACTCGTATTATATTGAAGGTAAAAGTTGAAAACCT
-90
TTATTTTTAGGGAAGTTATAATGAACTCATAAATTACTAACGCCACTCGTTTATGTGAGACGGTGAATAG
-20
TAGTATCCCTTTAACTTTACATTGGATTGTTGTGGCATTAACAATAAAGGTTGTGGCATTATCAATCCCC
51
TATTTTTATTGAATCAGAAATGAAGATATATGTTGTAGCCACAATAGCATGGATCCTGCTTCAATTTTCA
M K I Y V V A T I A W I L L Q F S -8
S
121
GCTTGGACAACAACTGATGCGGTAACATCAATCACATTAGATCTAGTAAATCCGACCGCGGGTCAATACT
A W T T T D A :V T S I T L D L V N P T A G Q Y 16
-1 +1
KpnI
191
CATCTTTTGTGGATAAAATCCGAAACAACGTAAAGGATCCAAACCTGAAATACGGTGGTACCGACATAGC
S S F V D K I R N N V K D P N L K Y G G T D I A 40
261
CGTGATAGGCCCACCTTCTAAAGGAAAATTCCTTAGAATTAATTTCCAAAGTTCCCGAGGAACGGTCTCA
V I G P P S K G K F L R I N F Q S S R G T V S 63
331
CTTGGCCTAAAACGCGATAACTTGTATGTGGTCGCGTATCTTGCAATGGATAACACGAATGTTAATCGGG
L G L K R D N L Y V V A Y L A M D N T N V N R 86
401
CATATTACTTCAGATCAGAAATTACTTCCGCCGAGTTAACCGCCCTTTTCCCAGAGGCCACAACTGCAAA
A Y Y F R S E I T S A E L T A L F P E A T T A N 110
472
TCAGAAAGCTTTAGAATACACAGAAGATTATCAGTCGATTGAAAAGAATGCCCAGATAACACAAGAAGATC
Q K A L E Y T E D Y Q S I E K N A Q I T Q E D 133
Sau3A

Figure 5.14. Nucleotide and deduced amino acid sequence of pλSap2.
The nucleotide sequence of the saporin-2 gene is presented in the 5′ to 3′ direction. The nucleotide sequence was derived from sequencing the two subclones pλSap2.E and pλSap2.B (see Fig 5.11). Nucleotides are numbered above the sequence starting at the ATG translational initiation codon, sequence upstream from this is numbered in a negative manner. Amino acids are numbered from the first residue of the mature protein, the signal peptide residues (marked S) 5′ to the first residue of the mature protein are numbered in a negative manner. An "in-frame" translational stop codon is denoted by an asterisk. A potential polyadenylation site (AATAAA) is also indicated. Relevant restriction sites are also indicated.


-126
AATTGTCCAAACTTACACATTAACACATAAATTATCTCAAAATTCTAAAAATTTTAGAACGTTCACTCTA
-56
TTTTTATGTTTTTCTCTCACTTTTTTATTTTTTTGTATAATTTGAAAAGTGTGAAAAAATAGACAGAGTA
13
CTCATCGTCTTATATTGAAGGTAAAAGTTGAAAACCTTTATTTTTATTGAATCAGAAATGAAGATATATG
M K I Y -21
S
83
TTGTAGCCACAATAGCATGGATCCTGCTTCAATTTTCAGCTTGGACAACAACTGATGCGGTCACATCAAT
V V A T I A W I L L Q F S A W T T T D A V T S I 4
-1 +1
BglII
153
CACATTAGATCTAGTAAATCCGACTGCCGGTCAATACTCATCTTTCGTGGATAAAATCCGAAACAACGTA
T L D L V N P T A G Q Y S S F V D K I R N N V 27
223
AAGGATCCAAACCTGAAATACGGTGGTACCGACATAGCCGTGATAGGCCCACCTTCTAAAGATAAATTCC
K D P N L K Y G G T D I A V I G P P S K D K F 50
293
TTAGAATTAATTTCCAAAGTTCCCGAGGAACGGTCTCACTTGGCCTAAAACGCGATAACTTGTACGTGGT
L R I N F Q S S R G T V S L G L K R D N L Y V V 74
363
CGCGTATCTTGCAATGGATAACACGAATGTTAATCGGGCATATTACTTCAAATCAGAAATTACTTCCGCC
A Y L A M D N T N V N R A Y Y F K S E I T S A 97
433
GAGTTAACCGCCCTTTTCCCAGAGGCCACAACTGCAAATCAGAAAGCTTTAGAATACACAGAAGATTATC
E L T A L F P E A T T A N Q K A L E Y T E D Y 120
503
AGTCGATCGAAAAGAATGCCCAGATAACACAGGGAGATAAAAGTAGAAAAGAACTCGGGTTGGGGATCGA
Q S I E K N A Q I T Q G D K S R K E L G L G I D 144
573
CTTACTTTTGACGTTCATGGAAGCAGTGAACAAGAAGGCACGTGTGGTTAAAAACGAAGCTAGGTTTCTG
L L L T F M E A V N K K A R V V K N E A R F L 167
643
CTTATCGCTATTCAAATGACAGCTGAGGTAGCACGATTTAGGTACATTCAAAACTTGGTAACTAAGAACT
L I A I Q M T A E V A R F R Y I Q N L V T K N 190
713
TCCCCAACAAGTTCGACTCGGATAACAAGGTGATTCAATTTGAAGTCAGCTGGCGTAAGATTTCTACGGC
F P N K F D S D N K V I Q F E V S W R K I S T A 214
783
AATATACGGGGATGCCAAAAACGGCGTGTTTAATAAAGATTATGATTTCGGGTTTGGAAAAGTGAGGCAG
I Y G D A K N G V F N K D Y D F G F G K V R Q 237

```
                                                                  EcoRI  853
GTGAAGGACTTGCAAATGGGACTCCTTATGTATTTGGGCAAACCAAAGTCGTCAAACGAGGCGAATTCTA
 V   K   D   L   Q   M   G   L   L   M   Y   L   G   K   P   K   S   S   N   E   A   N   S   260

                                                                        923
CTGCTTACGCCACTACGGTCCTCTGAAGCCTACTTTACTAATCACATGAATTAATCAGCAAGTACTGCGA
T   A   Y   A   T   T   V   L   * >.....................................

                                                                        993
CTCGTACCTATGTATTCCTTTGTATTATACCAGCTTGTTTCATGTTACAATAACGTTGTAATTGGTGTGT
.....................................................................

                                                                       1063
TATAAACTTATAATGAATAAAAATTCTGTATATTAGATTATTATCGATGGCAAATTAGATTAAAGTTTGA
...............<PolyA+>.....................................................

ATAAGGTTTAAATTCCGACCTTGATTCGTATATACATATG
.....................................
```

5'-flanking sequences of λSap2 and λSap4 are A and T rich, 77% and 70% respectively. The upstream regions of ricin and α-trichosanthin genes are also A and T rich (Chow *et al.*, 1990) in common with many plant promoter sequences. The 3'-downstream sequences of λSap2 and λSap3 contain consensus polyadenylation signals (AATAAA) (Fitzgerald and Shenk, 1981; Heidecker and Messing, 1986; Waugh and Brown, 1991).

These features suggest that the gene sequences are likely to be transcribed. The deduced amino acid sequences of the clones are highly similar to both the previously published cDNA sequence of Benatti *et al.* (1989) the sequence of saporin-6 protein and the saporin PCR probe. The deduced amino acid sequence of the clones λSap2, 3 and 4 were designated Sap2, 3 and 4. This was to avoid any confusion with the isolated saporin protein fractions classified by Montecucchi *et al.* (1989). These distinct saporin fractions were designated saporin-1, saporin-5a, saporin-5b, saporin-6a and saporin-6b. No attempt has been made to assign the saporin gene sequences described here to these isolated saporin fractions.

5.5.1. Comparison of the deduced amino acid sequences of members of the saporin gene family.

The deduced primary protein sequences of the characterised saporin clones are shown aligned in Fig 5.15. The deduced primary sequence from the saporin-6 cDNA is also included for comparison.

The presumed leader peptide, 24 residues in length, is completely conserved between Sap2, Sap4 and saporin-6. N-terminal sequencing of the saporin-6 protein indicated that the mature peptide sequence commences at the valine residue marked +1 (Fig 5.1). It is assumed that this leader sequence is responsible for the targeting and transport of the protein into the lumen of the endoplasmic reticulum and is subsequently co-translationally removed. The sequence of the leader peptide obeys the "(-3, -1)" rule of von Heijne (1984) which defines the site of cleavage of the leader peptide from the polypeptide. This rule states that one of the following residues should be present at position -1; A, S, G, C, T or Q. Position -3 must not have aromatic, charged or large polar residues. The saporin leader sequence has an alanine residue at -1 and an uncharged threonine at -3, suggesting that this is a functional leader sequence. Co-translational transport of the saporin proteins may occur in a similar manner to that of ricin and agglutinin. These proteins are synthesised

Figure 5.15. Comparison of saporin polypeptide primary sequences predicted by the nucleotide sequence of members of the saporin gene family.

The deduced amino acid sequence of saporin clones Sap2, Sap3 and Sap4 are shown aligned with the deduced amino acid sequence of the Saporin-6 cDNA sequence presented by Benatti *et al.* (1989). Conservative amino acid changes are in bold type. Residue numbering is from the first residue of the mature protein. A potential N-linked glycosylation site at the C-terminus of Sap4 is also indicated in bold type. An asterisk indicates the proposed last amino acid residue of mature saporin-6.

and co-translationally secreted into the endoplasmic reticulum with the concomitant removal of the leader sequence (Lord, 1985b).

The predicted full length sequence of the mature protein Sap2 is 268 residues in length, nine residues longer than that encoded by the previously reported cDNA sequence of saporin-6. Sap4 contains the leader sequence and the first 133 residues of the mature protein whilst Sap3 is missing the leader sequence and first 28 residues at its N-terminal end. The predicted sequence of Sap3 is four residues shorter at the C-terminal end of the polypeptide than Sap2. Conservative amino acid changes are indicated.

Whilst the amino acid sequences predicted by the three clones are highly similar over most of the protein there is significant divergence at the C-terminal regions of Sap2 and Sap4. The saporin-6 cDNA sequence terminates at an EcoRI site present at the same position in λSap2, however the coding sequence continues beyond this site and indicates that the cDNA sequence is not full-length. Benatti *et al.* (1989) reported that the C-terminal sequence of mature saporin-6 is likely to end after the peptide sequence LMYLGKPK. This was based on amino acid sequencing of saporin-6 fragments generated by cyanogen bromide cleavage or treatment with pepsin or clostripain. If this is the true C-terminus of saporin-6 it suggests that the saporin proteins may be subject to post-translational processing as both Sap2 and Sap4 extend beyond the supposed C-terminal of saporin-6. This would involve the removal of the peptide sequence present at the C-terminal of each protein. Also of note in this regard is the divergence of sequence of C-terminal ends of Sap2 and Sap3. A comparison of the deduced amino acid sequence of an α-trichosanthin genomic clone (Chow *et al.*, 1990) with the primary sequence of the mature protein (Collins *et al.*, 1990) also revealed a C-terminal extension of 19 residues.

Many plant proteins are post-translationally modified at the C-terminus. A C-terminal propeptide sequence of barley lectin is required for the correct sorting of the protein to the cell vacuole (Bednarek *et al.*, 1990). The basic isozymes of tobacco β-1,3-glucanases also have a C-terminal extension of 22 residues required for the correct targeting of the proteins to the vacuoles. The sequence is subsequently cleaved from the protein. An extracellular form of the enzyme lacks this extension (Linthorst *et al.*, 1990).

The C-terminal extensions of the vacuolar isoforms of β-1,3-glucanases

of tobacco and *Nicotiana plumbaginifolia* contain asparagine-linked (N-linked) glycosylation sites. The proteins are initially synthesised as glycosylated precursors, the mature forms are produced by the removal of the glycosylated C-terminal propeptide (Shinshi *et al.*, 1988; Van den Bulcke *et al.*, 1989). Bednarek *et al* (1990) have suggested that these post-translational modifications are also required for the correct vacuolar targeting of glucanases in the same manner as the barley lectin since it also has a utilised N-linked glycosylation site in the C-terminal propeptide. The asparagine glycosylation site is (Asn-X-Ser/Thr).

The C-terminal sequences of Sap2 and Sap4 may also be propeptide sequences, however there is no solid data from the amino acid sequencing to indicate the end of the mature protein. It is tempting to speculate that the observed divergence in the C-terminal sequence may reflect differing targeting information present in the C-terminal extensions. Different information cannot reside in the N-terminal leader sequences of the saporin clones since the leader sequence is conserved in all the saporin clones. Of further interest, within the presumed C-terminal propeptide sequence is an N-linked glycosylation site (Asn Ser Thr) present in Sap2 which is not present in Sap4. This may be of importance in determining the final localisation of the saporin proteins, once sequestered into the endoplasmic reticulum, and may again reflect a difference in targeting information in the two saporin coding sequences. As Sap4 and saporin-6 show some amino acid differences it is not certain if the saporin-6 gene also has the same C-terminal sequence. The presence of a glycosylation site need not imply that it is utilised. Saporin proteins are expressed in both seeds and leaves although no studies have been undertaken to determine the cellular localisation of the proteins nor the patterns of expression of the gene family. It is possible that the saporins present in seeds are targeted to the protein bodies. In the leaves there are several possible destinations for the mature saporin proteins. They may be transported to the extracellular cell wall matrix. Pokeweed antiviral protein is found localised in the cell wall matrix of leaf mesophyll, and appears to loosely bound (Ready *et al.*, 1986). Alternatively the vacuole may be the final destination. Saporin proteins expressed from distinct genes may therefore have different destinations determined by the presence or absence of specific targeting C-terminal propeptides. If there is no targeting information present then the

protein is likely to be secreted from the cell following a default pathway (Bednarek *et al.*, 1990).

Barbieri *et al.* (1989) have established plant cell cultures derived from leaf tissue from which active RIPs can be purified. Saporin and pokeweed antiviral protein were isolated from *Saponaria officinalis* L. and *Phytolacca americana* L. cultures, respectively.

The pokeweed antiviral protein was a form related to other RIPs isolated from *Phytolacca*. It resembled the pokeweed antiviral protein PAP I found in leaves rather than the second leaf form of the protein (PAP II) (Bjorn *et al.*, 1984) or the protein isolated from seeds (PAP-S) (Barbieri *et al*,. 1982). Data was not presented concerning the identity of the saporin isolated. No activity could be found in the medium of either RIP-producing cell line suggesting that they were neither secreted or associated with the cell wall matrix.

This data may not however, reflect, the *in planta* situation. The culturing of the cells may change the pattern of saporin gene expression and final targeting of proteins, plant cells are known to undergo somaclonal variation when propagated in tissue culture. The data is also at variance with the localisation of pokeweed antiviral protein in the cell wall matrix.

There of course may be no significance in the observed C-terminal divergence. Hilder *et al.* (1989) have shown that the Bowman-Birk family of protease inhibitors from cowpea were highly divergent at both the N- and C-termini of the mature proteins yet all retained activity. This indicates that changes can be tolerated in regions removed from the active site of the inhibitor. There was no indication that the regions of sequence divergence at the C-terminal end of the proteins contained different targeting signals. This is also a possibility for the saporin genes and the C-terminal divergence may not play a role in protein targeting.

5.5.2. Alignment of saporin primary sequences with other plant RIPs and related toxins.

Comparison of the primary sequences of plant RIPs and functionally related prokaryotic toxins has shown that several amino acid residues are conserved in all toxins sequenced to date. These alignments, including the saporin primary sequences are presented in Fig 5.16. Sap4 is omitted

Sap-2	(162)	NE**A**RF-LLIAIQMTA**E**VA**R**FR----YIQNLVT	(188)
Sap-3	(162)	NE**A**RF-LLIAIQMTA**EA**A**R**FR----YIQNLVT	(188)
Saporin-6	(162)	DE**A**RF-LLIAIQMTA**EA**A**R**FR----YIQNLVI	(188)
α-Trichosanthin	(145)	NS**A**ASALMVLIQSTS**EA**A**R**YK----FIEQQIG	(172)
BPSI	(159)	QQ**A**REAVTTLLLMVN**EA**T**R**FQTVSGFVAGLL-	(189)
SLT-IA	(153)	SV**A**RA-MLRFVTVTA**EA**L**R**FR----QIQRGFR	(179)
SLT-IIA	(152)	DASRA-VLRFVTVTA**EA**L**R**FR----QIQREFR	(178)
Ricin D	(163)	TL**A**RS-FIICIQMIS**EA**A**R**FQ----YIEGEMR	(189)
Preproabrin	(140)	EK**A**RT-LIVIIQMAS**EA**A**R**YR----YISNRVG	(166)
Sap-2	(189)	KNF-PNKFDS--DNKV-IQFEVS--**W**RKI**S**TA	(214)
Sap-3	(189)	KNF-PNKFNS--ENKV-IQFQV**N**--**W**SKI**S**KA	(214)
Saporin-6	(189)	KNF-PNKFNS--ENKV-IQFEV**N**--**W**KKI**S**TA	(214)
α-Trichosanthin	(173)	KR---VDKT--FLPSLAIISLE**N**S-**W**SAL**S**KQ	(198)
BPSI	(190)	-HPKAVEKKSGKIGNE-MKAQV**N**G-**W**QDL**S**AA	(218)
SLT-IA	(180)	TTLDDLSGRSYVMTAEDVDLTL**N**--**W**GRL**S**SV	(209)
SLT-IIA	(179)	QALSE-TAPVYTMTPGDVDLTL**N**--**W**GRI**S**NV	(207)
Ricin D	(190)	TRIRYN-RRSAPDPS--VITLE**N**S-**W**GRL**S**TA	(217)
Preproabrin	(167)	VSIRTG-TAFQPDPA--NLSLE**NN**-**W**DNL**S**GG	(194)

Figure 5.16. Comparison of saporin amino acid sequences with those of several plant RIP and *Escherichia coli* toxin polypeptides.

The region of primary sequence shown corresponds to that of the major cleft present in the 3D crystal structure of ricin A chain. This cleft is the suggested active site of ricin A chain proposed by Ready *et al.*, (1988). The alignment is based on that of Hovde *et al*,. (1988) and Ready *et al.*, (1988). Primary sequences are from the following sources:-

Saporin-6: Benatti *et al.* (1989).

α-Trichosanthin: Chow *et al.* (1990).

Barley Protein Synthesis Inhibitor II (BPSI): Asano *et al.* (1986).

Shiga-like toxins (SLT-IA and SLT-IIA): Hovde *et al.* 1988.

Ricin D A-chain and preproabrin A-chain: Lord JM, Department of Biological Sciences, University of Warwick, personal communication.

Amino acid numbers refer to the positions of residues within each mature polypeptide. Conserved residues are in bold type. Sequence alignment is maximised by the introduction of gaps, denoted by dashes. Note the amino acid substitutions for Sap2 at otherwise invariant positions.

as it does not extend into the region of interest.

The saporin primary sequences also contain parts of the conserved residues present in the other proteins. Sap2 has two amino acid changes at conserved positions; a valine substituted for alanine at position 177 in the saporin primary sequence and a serine substituted for aspartic acid at position 207. The three saporin sequences have an isoleucine residue at position 211, as does the shiga-like toxin IIA at the corresponding position. The other proteins all contain a leucine residue at this position, however this residue change is a conservative one.

It is assumed that some of these residues lie within the active site of the proteins. Ready *et al.* (1988) have shown that several of the conserved residue side chains lie in close proximity to each other when mapped to the known tertiary structure of ricin determined by X-ray crystallography (Montfort *et al.*, 1987). This region forms a cleft between domains 2 and 3 of ricin-A chain and has been proposed as the ricin active site. The residues present are Ala_{165}, Glu_{177}, Asn_{209}, Trp_{211}, Leu_{214} and Pro_{229}. These six residues are invariant between ricin, trichosanthin, barley protein synthesis inhibitor and ribonuclease H (Ready *et al.*, 1988).

If other sequences are included in the alignment several of these residues are not completely invariant. SLT-IIA has a serine substitution for the alanine at position 154, corresponding to position 165 in ricin. There are two differences already noted for Sap2 and an isoleucine or leucine is present at the corresponding position 214 in ricin. The arginine residue at position 180 and the serine at position 215 in ricin is invariant in all the sequences examined. The proline residue at position 229 of ricin is not present in the saporin sequences.

Invariant residues present in all the toxin sequences shown are glutamic acid (177), arginine (180), tryptophan (211) and serine (215). (The numbering refers to the position in ricin A-chain). Conservation of these residues in toxins from diverse sources implies that they play a crucial role in determining either the conformation of the protein or have a direct role in catalysis at the active site of the protein. Hovde *et al.* (1988) have shown that the glutamic acid 167 of SLT-IA (conserved in all RIP sequences) was implicated as an active site residue. Site-directed mutagenesis was used to change this residue from glutamic to aspartic acid. The specific activity of the resultant mutant polypeptide

was 1000 fold less than the corresponding wild-type protein.

This was also demonstrated by Schlossman *et al.* (1989) using site-specific mutagenesis of ricin-A chain to change the corresponding glutamic acid 177 to aspartic acid or alanine. Changing this residue to aspartic acid reduced the specific activity by 80 fold yet changing the residue to alanine only resulted in a 20 fold reduction in specific activity. This was unexpected as the alanine substitution is less conservative than aspartic acid which retains the carboxylate side chain. The authors suggested that the amino acid changes affect the conserved arginine residue at 180. The aspartate residue changes the spatial orientation of arginine due to charge-charge interactions but the alanine has less effect as it is an uncharged residue. This implies a role for the conserved arginine residue in the catalytic mechanism of ricin action.

Frankel *et al.* (1990) provided evidence, again by site-directed mutagenesis of ricin-A chain, that changing the glutamic acid 177 to alanine led to the nearby glutamic acid 208 substituting for the carboxylate side chain and maintaining activity, albeit at a level lower than the wild-type. They also showed that a positive charge at the arginine position 180 was a requirement for both the activity of the protein and its solubility. These experiments demonstrated the involvement of some of the conserved residues in the catalytic activity of ricin and SLT-IA. The sequence alignments presented refer to the position of residues within the cleft of ricin-A chain. Collins *et al.* (1990) have shown that molecular models of α-trichosanthin and abrin A-chain proteins follow generally similar conformations when compared with ricin. The conserved residues of α-trichosanthin are clustered in an a cleft analogous to that of ricin. It is therefore likely that the saporin polypeptides will also have tertiary conformations akin to that of ricin A-chain. The presence of invariant residues within the two saporin genomic clones (Sap2 and Sap3) suggests that they may well encode active polypeptides. Experiments designed to test the activity of the characterised gene sequences are presented in chapter 6.

5.5.3. 5′-and 3′-flanking sequences in the saporin gene family.

The 5′-and 3′-flanking sequences of the saporin genomic clones were compared with each other and with the 5′-untranslated sequence of the saporin-6 cDNA. The sequence alignments are presented in Figs 5.17 a-d.

There is insufficient length of sequence to positively identify transcriptional consensus sequences in the 5′-sequences of λSap2 and λSap4 (Joshi, 1987). Analysis of the genomic clones cannot identify the position of the start of transcription of the genes. mRNA sequences would be required for determining the transcription site by primer extension analysis or SI nuclease protection studies.

Comparison of the 5′-untranslated regions of λSap2 and λSap4 of the first 207 bp upstream of the ATG translation initiation codon (Fig 5.17a) shows that the first 19 bp upstream are identical then the sequences diverge. Comparison of the 5′-untranslated regions of both λSap2 and λSap4 with the saporin-6 sequence (Figs 5.17b and 5.17c respectively) reveals no strong similarity.

Some plant multigene families have very conserved 5′-flanking sequences. The members of the legumin gene family in pea (*Pisum sativum* L. var. Feltham First) (Leg A, B and C) have an identical 321bp of 5′ non-translated sequence (Lycett *et al.*, 1985). These genes are all seed specific and subject to developmental regulation. Other gene families exhibit reduced similarity when upstream sequences are compared. The extensin cell wall protein genes *ExtA* and *ExtB* from oilseed rape (*Brassica napus* L.) have no detectable regions of homology apart from the consensus "TATA" box sequence (Gatehouse *et al.*, 1990b). Holdsworth *et al.* (1988) demonstrated that only the coding sequences of members of a wound/ripening-related gene family of tomato were homologous. No significant homology could be found in either the 5′ or 3′ flanking sequences of the genes and the genes were differentially expressed in the leaves and fruit of the plant.

There is insufficient sequence data from the 5′-flanking regions of the saporin genomic clones to make definitive statements about the organisation of putative transcriptional control sequences and the observed divergence of λSap2 and λSap4. One possibility is that the divergence may reflect a different pattern of expression for each gene. The promoter regions would then possess different regulatory elements determining for example; tissue specificity and/or timing of expression.

Comparison of the saporin genomic clones with the saporin-6 cDNA 5′-untranslated region should be treated with caution. This sequence, reported by Benatti *et al.* (1989) has 120 bp of 5′-non-translated sequence. The restriction map of the two overlapping clones from which

```
AATTGTCCAAACTTACACATTAACACATAAATTATCTCAAAATTCTAAAAATTTTAGAAC Sap2
| |  ||  |  ||  | ||| | |  ||||  |  | ||||  | |  ||  |||
ACTACTCGTA--TT--ATATTGA-AGGTAAA--AGTTGAAAACCTTTATTTTTAGGGAAG Sap4

GTTCACTCTATTTTTATGTTTTTCTC-TCACTTTTTTATTTTTTTGTATAATTTGAA--A Sap2
 |  | |  | |  ||  ||  |  |  ||||  ||||| |       | | | | |  |
TTATAATGAACTCATAAATTACTAACGCCACTCGTTTATGTGAGACGGTGAATAGTAGTA Sap4

AGTGTGAAAAAATAGACAGAGTACTCATCGTCTT-ATATTGAAG---GTAAAAGTTGAAA Sap2
    |  ||   || |  |  |  |  | |  || | | | |||   ||   | |   ||
TCCCTTTAACTTTACATTGGATTGTTGTGGCATTAACAATAAAGGTTGTGGCATTATCAA Sap4

ACCTTTATTTTTATTGAATCAGAAATG  Sap2
 ||  ||||||||||||||||||
TCCCCTATTTTTATTGAATCAGAAATG  Sap4
```

Matches = 103 Mismatches = 87 Unmatched = 14
Length = 204 Matches/length = 50.5 percent

Figure 5.17a. Comparison of the 5′ untranslated sequences of λSap2 and λSap4.

The 5′nucleotide sequences of the saporin genes λSap2 and λSap4 are shown aligned. Gaps have been introduced to maximise the alignment. Nucleotides are numbered in a negative manner from the first base 5′ to the ATG translation initiation codon (bold type).

```
-119
  ACT-TTTTTATTTTTTTGTATAATTTGAAAAGTGTGAAAAAAT-AGACAGAGTACTCATC Sap2
    |   ||| ||    ||  | | |  ||  | || | | |||| | |   | |    | |
  GTTGAATTTCTTCACTTTGACATTCAGAGCACTGGGCAGAAATCAAATCTATTTTGAAAC Sap6
-119

-63
  --GTCTTATATTGAAGGTAAAAGTTGAAAACCTTTATTTTTATTGAATCAGAAAT-GAAGATATATG Sap2
        | |||| |   |  ||| | |  |  |  ||  |       |||  | | |||
  AAAAAATCTATTTATTTTTGAAGCTTATTAATTGCATCGTACAACCTTCATCACTAGAAATG       Sap6
-58

Matches = 53      Mismatches = 61      Unmatched = 10
Length = 124      Matches/length =  42.7 percent
```

Figure 5.17b. Comparison of the 5′ untranslated sequences of λSap2 and saporin-6 cDNA.

The 5′nucleotide sequences of the saporin gene λSap2 and the saporin-6 cDNA are shown aligned. Gaps have been introduced to maximise the alignment. Nucleotides are numbered in a negative manner from the first base 5′ to the ATG translation initiation codon (bold type).

```
-120 ACGCCACTCGTTTA-TGTGAGACGGTGAATAGTAGTATCCCTTTAACTTTACATTGGATT Sap4
          | |  || | | ||| |    ||  | | |       | || | ||  ||| |
-119 GTTGAATTTCTTCACTTTGACATTCAGAGCACTGGGCAGAAATCAAATCTATTTTGAAAC Sap6

-61 GTTGTGGC-ATTAACAATAAAGGTTGTGGCATTATCAATCCCCTATTTTTATTGAATCAGAAATG Sap4
           | ||| |   |  | | | |   | | |  |||    |   ||  | | | ||||
-59 AAAAAATCTATTTATTTTTGAAGCT-TATTAAT-TGCATCGTACAACCTTCATCACT-AGAAATG Sap6

Matches = 51     Mismatches = 66     Unmatched = 5
Length = 122     Matches/length =  41.8 percent
```

Figure 5.17c. Comparison of the 5′ untranslated sequences of λSap4 and saporin-6 cDNA.

The 5′nucleotide sequences of the saporin gene λSap4 and the saporin-6 cDNA are shown aligned. Gaps have been introduced to maximise the alignment. Nucleotides are numbered in a negative manner from the first base 5′ to the ATG translation initiation codon (bold type).

```
 782 AGGTGAAGGACTTGCAAATGGGACTCCTTATGTATTTGGGC---AAACC----AAAGTCG Sap2
     |||||||||||| |||||||||| ||||| |||||| | |||   | |||    || |  |
 626 AGGTGAAGGACCTGCAAATGGGGCTCCTAATGTATCTAGGCACGACACCGAATAATGCTG Sap4

 835 TCAAACGAGGC-GAATTCTAC-TGCTTACGCCA-CTACGGTCCTCTGAAGCCTACTTTAC Sap2
       | |||  | |      || |    |  | |  |||| | |||    |  |  || |
 686 CTGATCGATACAGGGCAGAACTTTGAAATTCGATTTACGATTCTC-ATTGATTGATTAAT Sap4

 892  TAATCACATGAATTAATCAG-CAAGTACTGCGACTC----GTA--CCTATGTATTCCTTT Sap2
     |       | |||||| || |||||| ||   |||    |||  ||||||||| || |
 745 TGGGGTGTTTGATTAATTAGCCAAGTATTGAATCTCTAATGTATTCCTATGTATGCCCTC Sap4

 945 GTATTATACCA--GCTTGTTTC-ATGTTACAATAACGTTGTAATTGGTGTGTTATAAACT Sap2
      ||||  ||||    ||||||| ||  |  | |  | ||    || || |   |||||
 805 TTATTTCACCATTAGTTGTTTCAATCGTGTAGTTTCTTTTCGTTTCGTTTCGAATAAAGG Sap4

1002 TATAATGAATAAAAATTCTGTATATTAGATTATTATCGATGGCAAATTAGATTAAAGTTT Sap2
       | |  || | | ||| |   || | | | |  || |  |  ||| | |    ||  |
 865 CTAATTATATTAGATTTCGGCTCATGA-AATGTCTTCTAATG--AATCATA-ACCAGCGT Sap4

1062 GAATAAGGTTTAAATTC-CGACCTTGAT-TC-GTAT-ATACATATG  Sap2
      |||      ||||||| ||   | | | || | ||  |  ||||
 921 TAATCTCCCGTAAATTCACGGGTTGGCTGTCAGCATGCTGAATATTA Sap4

Matches = 191    Mismatches = 126    Unmatched = 30
Length = 347     Matches/length =  55.0 percent
```

Figure 5.17d. Comparison of the 3′ sequences of λSap2 and λSap3.
The 3′ sequences of the saporin genes Sap2 and Sap4 are shown aligned, gaps are included to maximise the alignment. The alignment starts in the extreme 3′ coding sequence and extends into the 3′ non-translated region. The stop codon (TGA) in each sequence is indicated by bold type. Nucleotides are numbered from the first base of ATG the translation initiation codon.

this sequence was derived indicated, that in one clone, there was a further 650 bp of sequence upstream of the start of sequencing. This implies that either the 5'-non-translated region of the saporin-6 cDNA is at least about 770 bp, or that part of this cDNA clone is in fact artifactual. Such a length of 5'-non-translated sequence would be extremely unusual. Joshi (1987) compiled 5'-non-translated sequences from plant genes and showed that variation in the length of sequence ranged from 9 to 193 nucleotides but with a length of 40 to 80 nucleotides in 53% of the genes compared. On the basis of this the sequence of the saporin-6 cDNA 5'-sequence is treated with some suspicion.

A comparison of the extreme 3'-coding and 3'-non-translated sequences of λSap2 and λSap3 is presented in Fig 5.17d. The divergence in the nucleotide sequence within the coding region is apparent and is also reflected in the amino acid sequences over this region. There is also little sequence similarity between the two genes after the stop codons. Hilder *et al.* (1989) describe similar divergence in the cDNA sequences of cowpea Bowman-Birk protease inhibitors. Divergence in the 3'-non-translated region of members of a gene family is not an unusual observation. This region can be used to produce gene-specific probes to distinguish between individual members of a gene family. Evans *et al.* (1990) used this approach to distinguish members of the extensin gene family of oilseed rape (*Brassica napus* L.) as the coding sequence of the genes is highly conserved.

5.5.4. General comments on the saporin multigene family.

The results presented above show that at least three genes comprise the saporin gene family. The saporin-6 cDNA sequence encodes a protein that is extremely similar to saporin sequences obtained by protein sequencing (Benatti *et al.*, 1989). However the cDNA clone may not encode saporin-6 protein as there are several forms of saporin that can be isolated from *Saponaria*. It is also possible that there is a degree of heterogeneity within the saporin family due to variation in the source material as *Saponaria officinalis* L. is not an in-bred line. This situation is analogous to that of ricin in that comparison of the deduced amino acid sequence with that obtained from protein sequence studies indicated a small amount of microheterogeneity (Lamb *et al.*, 1985). The three saporin genomic clones are also highly similar to both the deduced amino acid

sequence of the cDNA clone and the protein sequence of saporin-6. However, they cannot be unequivocally identified as saporin-6 genes.

The isolated protein fractions of saporin are likely to be the products of distinct genes. The saporins are *not* glycoproteins (Stirpe *et al.*, 1983) hence differential glycosylation of a single precursor does not account for the different saporin forms isolated. Montecucchi *et al.* (1989) have presented the N-terminal sequences of five distinct saporin proteins. The N-terminal sequences of saporin-5a, -5b, 6a and 6b are identical for the first 36 residues, the limit of unequivocal sequence presented. Saporin-1 shows considerable divergence when compared with the other sequences, 19 differences in the first 32 residues. All the saporins were isolated from seeds except saporin-1 which was isolated from leaf tissue. The cDNA sequence of saporin-6 was isolated using mRNA from a leaf cDNA library. This implies that at least two saporin genes are expressed in the leaves of the plant and one in the seeds.

The exact match of the sequences for saporins-5a, -5b, -6a and -6b' at the N-terminus is also seen in the deduced amino acid sequences of genomic clones (Fig 5.15). The first 45 residues are conserved but a number of amino acid changes then occur. This suggests that these saporin proteins may also have sequence differences not detected by N-terminal sequencing and would explain why the saporins can be isolated as distinct protein fractions.

It is not possible to draw any further conclusions regarding the expression of the isolated saporin genes. Further studies will be required to determine patterns of expression in different tissues, subcellular localisation of saporin proteins and potential variations in expression with seasonal change. Several forms of pokeweed antiviral protein (PAP) can be isolated from pokeweed (*Phytolacca americana* L.). Depending upon the season, two forms are present in pokeweed leaves PAP in the spring and a second form, PAP II, in the summer (Houston *et al.*, 1983; Bjorn *et al.*, 1984). A third form, distinct from the other two, is present in seeds (PAP-S) (Barbieri *et al.*, 1982). It would be of interest to examine the expression of saporins in relation to seasonal change to establish if there are similar changes in expression patterns as seen in the case of pokeweed RIPs.

5.6. General comments on the utility of the cloning strategy employed.

The successful isolation of saporin genomic clones was facilitated by adopting a cloning strategy based on the production of a saporin PCR probe. The amplified sequences obtained using primers derived from peptide sequences were shown to be derived from saporin coding sequences. Subsequent radioactive labelling of the saporin PCR product by re-amplification was demonstrated to be a viable way of producing a gene-specific hybridisation probe. The use of this probe allowed the direct screening of a genomic library to obtain the corresponding genomic clones. Using this strategy there is no requirement for either; screening genomic or cDNA libraries with highly redundant oligonucleotides or for the differential screening of cDNA libraries to obtain a cDNA sequence. This would then be used as a probe in subsequent genomic library screenings to obtain the genomic clone. A further advantage of the strategy is that only partial protein sequence information is required in order to design PCR primers.

This strategy, using redundant oligonucleotide primers of and the PCR to amplify a specific sequence from genomic or cDNA, has been used with great success and is a relatively quick procedure. Gene probes obtained in this manner and their corresponding genomic and cDNA sequences include; yeast proliferating cell nuclear antigen (Bauer and Burgers, 1990), brain-derived neurotrophic factor and neurotrophin-3 (both members of the nerve growth factor/brain-derived neurotrophic factor gene family) (Leibrock *et al.*, 1989; Hohn *et al*,. 1990). The redundancy of some of these primers meant that oligonucleotide mixtures contained as many as 262,144 unique sequences (Gould *et al.*, 1989).

The use of inosine has greatly facilitated this experimental strategy due to the reduction of base redundancy in the oligonucleotide primers used in the PCR amplifications. Whilst fully degenerate primers have been used successfully to amplify sequences from genomic DNA (Gould *et al*, 1989) initial attempts to amplify saporin gene sequences using fully degenerate primers and genomic DNA were unsuccessful (Dr. P. Taylor, Department of Biological Sciences, University of Durham, unpublished observations).

Interestingly, Chow *et al.* (1990) used the same experimental approach described here to obtain a genomic sequence for α-trichosanthin. The primary amino acid sequence of α-trichosanthin was determined (Collins *et*

al., 1990), inosine-containing primers were synthesised encoding parts of the protein and used to amplify part of the α-trichosanthin coding sequence. This sequence was used as a probe to isolate the corresponding genomic clone. The authors (Chow *et al.*, 1990) noted that whilst primers, without inosine, contained mixtures of over 4000 unique sequences and could be successfully used to amplify mammalian sequences (Larrick *et al.*, 1989), whereas primers exhibiting similar complexity were unsuccessful in amplifications with plant DNA as a template. Using degenerate primers the size of the target genome may well be of critical importance in obtaining a successful amplification event.

Redundant primers have a greater chance of annealing non-specifically with an increase in genome size. Plant genomes can be at least 10 fold greater in size when compared with, say, mammalian genomes. The use of inosine at highly degenerate base positions reduces the primer complexity and may account for the successful amplifications obtained using these primers, as opposed to fully degenerate primer combinations.

As discussed in the section dealing with the amplification of ricin gene sequences (section 4.5.4) the possibility that PCR amplified sequences may contain errors due to mis-incorporation must be continuously borne in mind. Error rates (number of mutations occurring per nucleotide per cycle) of 2×10^{-4} have been reported by Saiki *et al.* (1988a) for PCR amplifications. Similar values were reported by Keohavong and Thilly (1989) although the actual error rate of *Taq* polymerase can depend on the nature of the template sequence, reaction pH and magnesium concentration (Eckert and Kunkel, 1990).

There is also the possibility that PCR amplification can produce artifactual sequences. Shuldiner *et al.* (1989) observed that amplification of *Xenopus* preproinsulin genes I and II resulted in the correct amplification of the genes. Significantly a third extra sequence produced was a hybrid of the two closely related insulin sequences. Meyerhans *et al.* (1990) have shown that co-amplification of two closely related sequences gave rise to recombinant hybrid sequences. Co-amplification of two distinct forms of the HIV1 viral *tat* gene sequences resulted in hybrid production by recombination during the PCR amplification. This was found to be a general phenomenon as co-amplification of two forms of the HIV1 viral *env* gene also produced hybrid molecules. The authors estimated that the frequency of PCR

mediated recombination between closely related sequences was of the order of 1-5% using their experimental conditions. The production of hybrid sequences is assumed to be due to the incomplete extension during the polymerisation stage. In subsequent annealing steps the incompletely extended products can anneal to related sequences allowing the initiation of DNA synthesis. This results in the production of a hybrid molecule.

Given that there is both an error rate and artifact production associated with PCR amplification any sequences obtained by PCR amplification must be treated with a degree of caution. The fragment of saporin coding sequence obtained by PCR amplification showed sequence changes in three clones. As there are several saporin genes it is likely that they were co-amplified, and as such, would be susceptible to PCR-mediated recombination. The product sequences were not taken to represent the actual sequence of the saporin genes for the reasons discussed above. Rather, the sequence of the clones confirmed the identity of the amplification product and allowed it to be used as a specific probe for saporin gene sequences. The presence of artifacts does not detract from its use as a hybridisation probe.

Cloned sequences obtained by PCR amplification must therefore be considered separately from genomic or cDNA clones. Whilst PCR amplification has great use in cloning strategies one should recognise that there can be limitations in its use (Karlovsky, 1990).

6. Expression Of Ribosome-Inactivating Protein Gene Sequences: In Vitro Ribosome Inactivation Studies.

6.1. Rationale for the *in vitro* expression of saporin and ricin gene sequences.

The two previous chapters describe the isolation of ricin and saporin gene sequences. As noted previously the two ricin sequences, one a genomic clone (λRic 2ii) and the other, a clone obtained using PCR (pRic1), were not sequenced. It is not known therefore if these sequences encoded active ricin polypeptides. The saporin gene sequences characterised were sequenced and appeared to encode saporin polypeptides. However sequencing alone does not indicate if these are in fact fully functional saporin polypeptides. A further problem was that two of the saporin genomic clones (λSap3 and λSap4) were not complete coding sequences.

The experimental strategy employed to express RIP gene sequences and determine their activity is illustrated in Fig 6.1. The strategy is based on the use of PCR amplification to incorporate transcriptional and translational signals into a RIP coding sequence. The procedure was adapted from those described independently by Sarkar and Sommer (1989) and Mackow *et al.* (1990).

This procedure was termed by Mackow and co-workers DNA Amplification-Restricted Transcription-Translation (DARTT). Primer sequences are synthesised to amplify a defined part of a coding sequence. The primers are designed such that they also contain an RNA polymerase promoter site at the 5'-end of the primer initiating DNA synthesis from the 5'-end of the coding sequence. An in-frame methionine initiation codon is included in the primer if required. A translation stop codon can be included in the primer directing DNA synthesis from the 3'-end of the coding region if required. PCR amplification of a coding sequence using these primers will result in a product containing the correct signals to serve as a template for *in vitro* transcription reactions. RNA produced from this PCR-DNA is then able to be translated *in vitro*. After *in vitro* translation a suitable assay is required for detecting the production of functional polypeptides.

In the case of the expression of RIP gene sequences, the leader peptide sequences was removed by using primers 3' to the leader along with the introduction of a translational initiation codon. A stop codon was

Figure 6.1. Schematic representation of DNA Amplification-Restricted Transcription-Translation (DARTT). Its use in the modification of RIP gene sequences for *in vitro* transcription and translation.

The experimental approach outlined below is based on the procedures described independently by Sarkar and Sommer (1989) and Mackow *et al.* (1990). The diagram shows the rationale for using PCR primers to produce selected gene coding sequences required for *in vitro* transcription and translation. Specifically; an RNA polymerase promoter site (ie SP6 or T7), an in-frame methionine initiation codon, the required coding sequence and a stop codon were incorporated. The procedure can also be used to introduce specific 5' and 3' truncations within a coding sequence. In this case it was used to remove the leader sequences of the isolated RIP genes.

The figure shows the procedure used to modify the ricin coding sequence such that only the A-chain is expressed. The same approach was used for the saporin coding sequences except the authentic saporin stop codon was used.

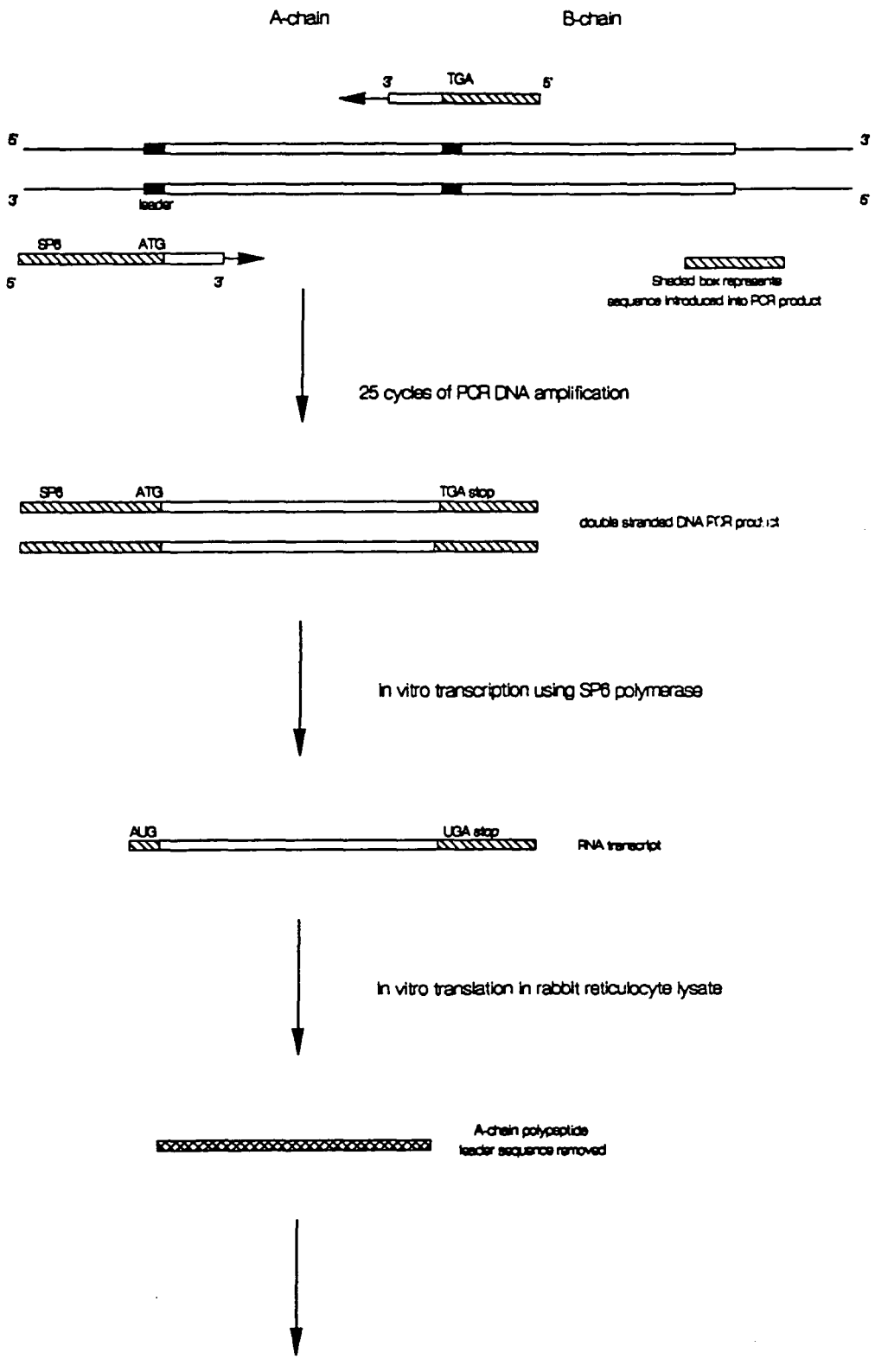
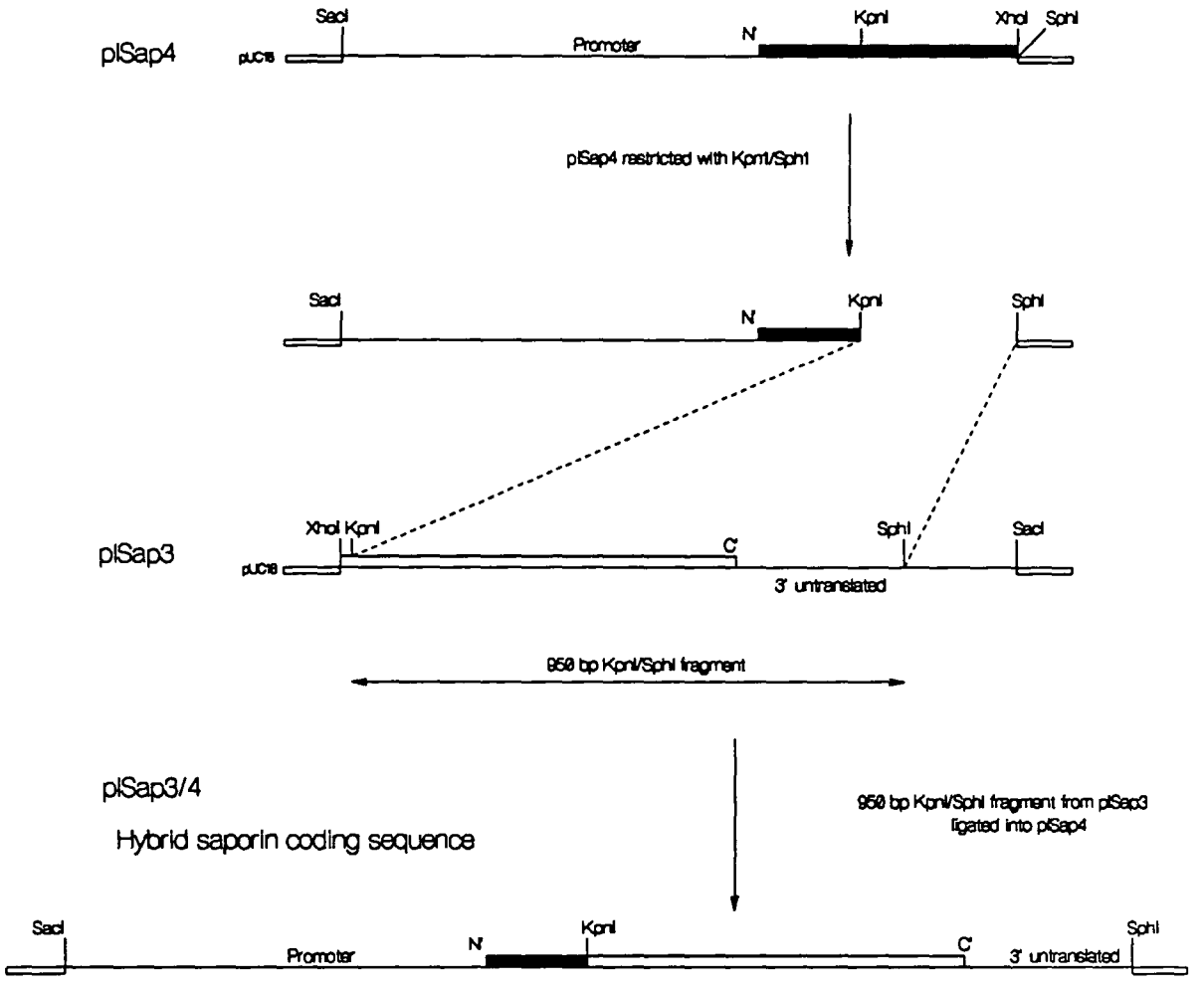

A-chain
B-chain
3'
TGA
5'
5'
3'
3'
leader
5'
SP6
ATG
5'
3'
Shaded box represents
sequence introduced into PCR product
25 cycles of PCR DNA amplification
SP6
ATG
TGA stop
double stranded DNA PCR product
In vitro transcription using SP6 polymerase
AUG
UGA stop
RNA transcript
In vitro translation in rabbit reticulocyte lysate
A-chain polypeptide
leader sequence removed
Functional polypeptide will cause ribosome inactivation

Figure 6.2. Strategy for constructing a hybrid full-length saporin coding sequence: pλSap3/4.

The diagram shows the strategy used to construct a hybrid saporin coding sequence using the characterised coding sequences of the truncated saporin clones pλSap3 and pλSap4. The common KpnI site was used to produce a gene fusion resulting in a full-length hybrid saporin gene sequence. (See Fig 6.3.)


plSap4
SacI
pUC18
Promoter
N'
KpnI
XhoI
SphI
plSap4 restricted with KpnI/SphI
SacI
N'
KpnI
SphI
plSap3
XhoI KpnI
pUC18
C'
SphI
SacI
3' untranslated
950 bp KpnI/SphI fragment
plSap3/4
Hybrid saporin coding sequence
950 bp KpnI/SphI fragment from plSap3
ligated into plSap4
SacI
Promoter
N'
KpnI
C'
3' untranslated
SphI

also introduced at the end of the A-chain sequence. The ricin primers were designed such that only the mature A-chain coding sequence was amplified and subsequently expressed. The same strategy was employed for the saporin coding sequences but there was no requirement for the introduction of a stop codon at the end of the coding sequence as saporins do not have a B-chain sequence.

Expression of functional RIP polypeptides in a rabbit reticulocyte *in vitro* translation system was then detected by assaying for the inhibition of protein synthesis and the specific depurination of 28 S rRNA molecules. The approach adopted was based on that described by May *et al.* (1989); incubation of transcripts encoding an active RIP coding sequence in an *in vitro* translation system will cause ribosome inactivation once the RIP sequence is translated. A second transcript added after initial incubation with the RIP transcript cannot then be translated. If protein synthesis is inhibited due to the specific depurination of the 28 S rRNA this can be detected by treating the purified rRNA with aniline (Endo *et al.*, 1987). The phosphodiester bonds at both sides of a depurinated site in an RNA molecule can be hydrolysed by treatment with aniline. If the aniline treated 28 S rRNA is cleaved at a specific site of depurination a diagnostic fragment is detectable, following electrophoresis of the sample. The released fragment is only observed after the specific N-glycosidase action of RIPs acting at a single site in the 28 S rRNA molecule. Aniline treatment of rRNA extracted from rabbit reticulocyte translation mixtures incubated with RIP transcripts will indicate if the RIP coding sequence produces a functional polypeptide.

The ricin clones and saporin clone, λSap2, are full-length sequences. Primers were synthesised for the ricin clones using the published cDNA sequence of ricin (Lamb *et al.*, 1985). The clone λSap2 was completely sequenced (see Fig 5.14) and virtually all the coding sequence is present on the EcoRI subclone pλSap2.E (8 residues at the extreme C-terminus are not present).

The other saporin genomic clones are truncated. pλSap4 contains only 133 amino acids of N-terminal mature coding sequence (see Fig 5.13). It is unlikely that this polypeptide sequence was of sufficient length to be active. pλSap3 contains most of the saporin coding sequence (see Fig 5.12) but is truncated by 28 amino acids at the N-terminus of the mature protein. This clone could possibly contain sufficient coding sequence to

produce a functional polypeptide. Fortunately these two clones overlap and share a common KpnI site, 116 base pairs from the start of the mature saporin coding sequence. A hybrid saporin coding sequence could therefore be constructed using these truncated clones. The resultant clone was then expressed to determine if the encoded hybrid polypeptide was active.

6.2. Construction of a hybrid saporin coding sequence: pλSap3/4.

Figure 6.2 illustrates the cloning steps required to produce the hybrid saporin clone using plasmids pλSap3 and pλSap4. The hybrid sequence, designated pλSap3/4 contains the promoter, leader sequence and first 37 amino acids of pλSap4, the remainder of the coding sequence and 3'-flanking region is derived from pλSap3. Fig 6.3 describes the construction of the hybrid clone and verification of its identity by restriction analysis. The result of the restriction analysis of four independent plasmid clones containing the hybrid saporin sequence is shown (Fig 6.3, Gel B, lanes 1-4). The expected product was observed in each case showing the successful production of the hybrid coding sequence.

6.3. Generation of RIP template DNAs for *in vitro* transcription.

The first step in the expression of the RIP coding sequences using the DARTT protocol described above was to add transcriptional and translational signals to the coding sequences. Fig 6.4 shows the design of PCR primers used to simultaneously amplify RIP coding sequences and to add sequences required for *in vitro* expression.

The ricin primer sequences were derived from the ricin cDNA sequence presented by Lamb *et al.* (1985) and were used to direct the amplification of the genomic clone λRic 1ii and the PCR derived clone pRic1. The saporin coding sequences used were; the hybrid saporin sequence of pλSap3/4, the full-length saporin sequence of pλSap2 and the truncated saporin sequence of pλSap3. The SP6 promoter sequence, an EcoRI restriction site and an in-frame ATG initiation codon were included in all the 5' primer sequences. Each primer annealed to the coding sequence of the respective RIP coding sequence template and directed the synthesis of the mature coding sequence as illustrated in Fig 6.1.

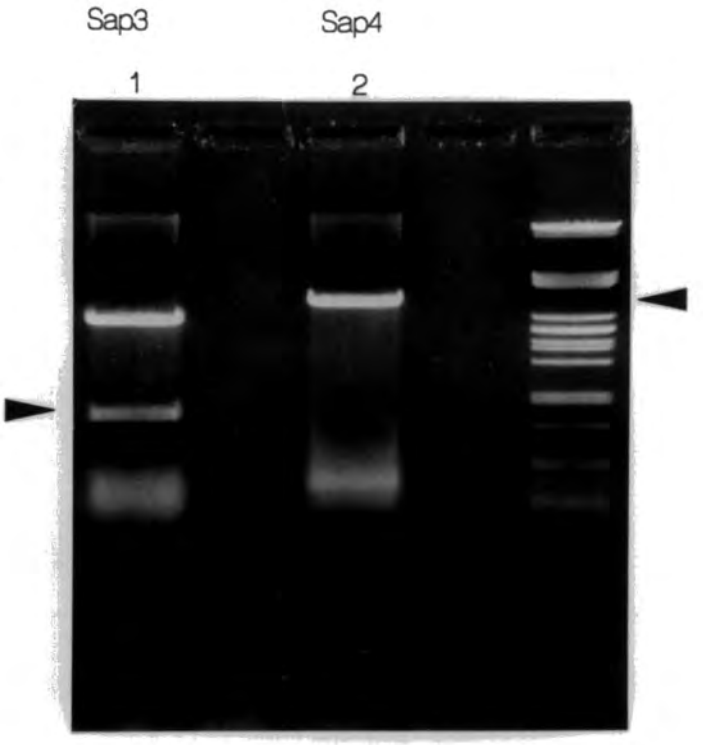
The 5' primer for pλSap3/4 and pλSap2 was the same; both clones have the same sequence over this region. The 3' primer for pλSap3/4 and pλSap3 is shared for the same reason. The 3' primer for pλSap2.E was a 17 mer

Fig 6.3. Construction of the plasmid pλSap3/4.

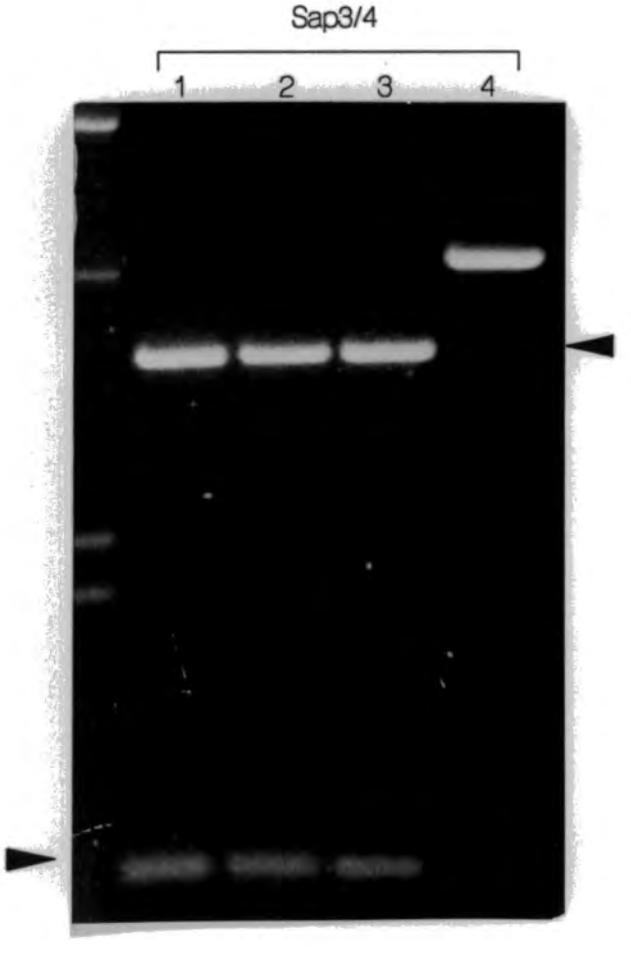
Gel A shows the result of restricting the two plasmid sub-clones pλSap3 and pλSap4 with SphI and Asp718 (Asp718 is an isoschizomer of KpnI and digests plasmid DNA more efficiently). The samples were separated by electrophoresis on a 1.0% (w/v) low-melting point agarose gel. Lanes 1 and 2 contain, respectively pλSap3 and pλSap4, the arrowed restriction fragments of the correct predicted size were then isolated. DNA size markers were λ/PstI.

The 940 bp SphI/Asp718 fragment of pλSap3, containing the C-terminal saporin coding sequence, was then ligated into the SphI/Asp718 sites of linearised pλSap4. This contained pUC18, the saporin promoter and 5' coding sequence to the KpnI (Asp718) site. The ligation mix was used to transform *E. coli* DH5α, only plasmids containing the correct ligation product should be recovered from the resultant transformants.

Gel B shows the restriction analysis of four independent clones recovered as described above. Plasmid DNA was isolated from each colony and subjected to restriction analysis. Clones 1,2 and 3 (lanes 1,2 and 3) were restricted with SphI and Asp718. Clone 4 was restricted with Asp718 only (lane 4). The restriction products were electrophoresed on a 0.7% (w/v) agarose gel with λ/HindIII DNA size markers. Clones 1,2 and 3 all contain the expected 940 bp SphI/Asp718 restriction fragment, arrowed. Clone 4, restricted with Asp718, is also of the correct predicted size (4.57 kbp) for the linear form of the hybrid produced in the ligation reaction (arrowed).



A



B

Figure 6.4. Nucleotide sequences of PCR primers used in the production of DARTT transcripts.

1. 5′ ricin primer. Coding sequence. This directs the amplification of mature protein, adds Met initiation codon, SP6 promoter and EcoRI site. The 5′ primers for all the RIP genes contained the same sequence over the SP6 promoter and EcoRI site; primer specificity is conferred by the sequence after the Met initiation codon.

2. 3′ ricin primer. Complement of coding sequence. This adds a stop codon at C-terminus of A-chain and HindIII site.

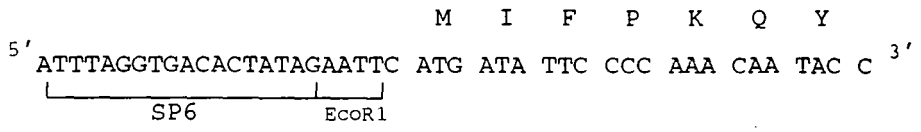
3. 5′ primer for both pλSap3/4 and pλSap2.E. Coding sequence.

4. 5′ primer for pλSap3. Coding sequence. This directs amplification of the truncated saporin coding sequence.

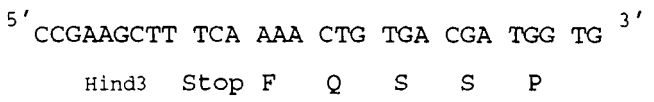
5. 3′ primer for pλSap3/4 and pλSap3. Complement of coding sequence. This primer was derived from the sequence of pλSap3 and annealed to a 19 bp sequence 4 bases 3′ to the stop codon; an SphI restriction site was also included at the 5′ end of the primer.

6. 3′ primer for pλSap2.E. A 17 mer reverse sequencing primer served as the 3′ primer. The pλSap2.E subclone (Fig 5.11) in pUC18 has an EcoRI site at the C-terminus of the coding sequence used in the sub-cloning step. The reverse primer directs DNA synthesis across this site into the coding region.

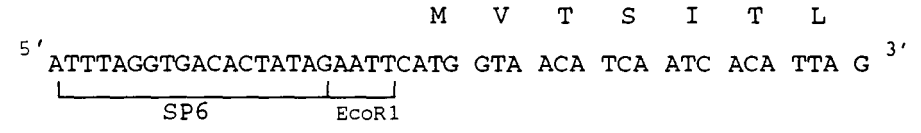
1. 5′ ricin primer.



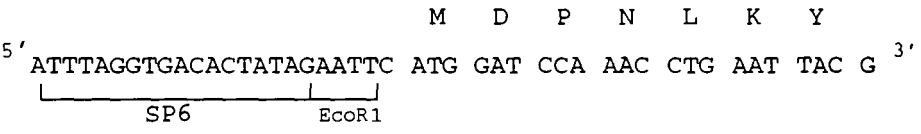
2. 3′ ricin primer.



3. 5′ primer for pλSap3/4 and pλSap2.E.



4. 5′ primer for pλSap3.



5. 3′ primer for pλSap3/4 and pλSap3.

6. 3′ primer for pλSap2.E.

reverse sequencing primer annealing 9 bases 5′ to the EcoRI site. This directed DNA synthesis into the Sap2 coding sequence across the EcoRI site at the C-terminus of the clone. In this case the small stretch of coding sequence 3′ to the EcoRI site was not included in the subsequent amplification product. The amplified product contained a translation stop codon (TAA) after the sequence Asn-Glu-Ala-Asn-Ser at the extreme 3′ end of the Sap2 coding sequence. The primers for the saporin coding sequences were designed to amplify the whole coding sequence. There is a possibility that the saporin sequences contain short C-terminal propeptide sequences not present in the mature, active protein. However these were not removed as the C-termini of the mature proteins are not known.

Restriction sites were included in all the primer sequences as shown. If the transcription reactions using the RIP template DNAs were unsuccessful the templates could then be cloned into suitable plasmid expression vectors, such as the pGEM or pBluescript series (Stratagene), using the incorporated restriction sites.

Fig 6.5 shows the production of template DNA for subsequent *in vitro* transcription reactions using RIP template DNAs. The primers contain sequences at their 5′ ends which do not hybridise to the initial template DNA. A lower annealing temperature (40^0C) was used in the first six rounds of PCR amplification to ensure the stability of primer/template duplexes. As amplification product accumulates the primers will anneal mainly to products of previous extensions and the whole of the primer sequence will anneal. As such, the annealing temperature was raised to 55^0C for the remaining rounds of amplification to increase the specificity of the amplification reaction. T_4 DNA polymerase was added at the end of the cycling reactions to fill in the ends of the amplification products and to ensure they were of full-length.

Electrophoretic analysis of the products, shown in Fig 6.5, indicated that the PCR product from each template was of the correct predicted size, non-specific products were either not detectable or of negligible amount compared with the desired product. The predicted size of each product includes the addition of the primer sequences to each end of the coding sequence. The PCR products were processed as described in Fig 6.5 and subsequently used as templates for *in vitro* transcription reactions.

Figure 6.5. The production of RIP gene constructs for *in vitro* transcription.

The primers shown in Fig. 6.4 were used to introduce the SP6 polymerase promoter, a methionine initiation codon and a stop codon into each of the RIP gene coding sequences excluding the leader sequence from each clone.

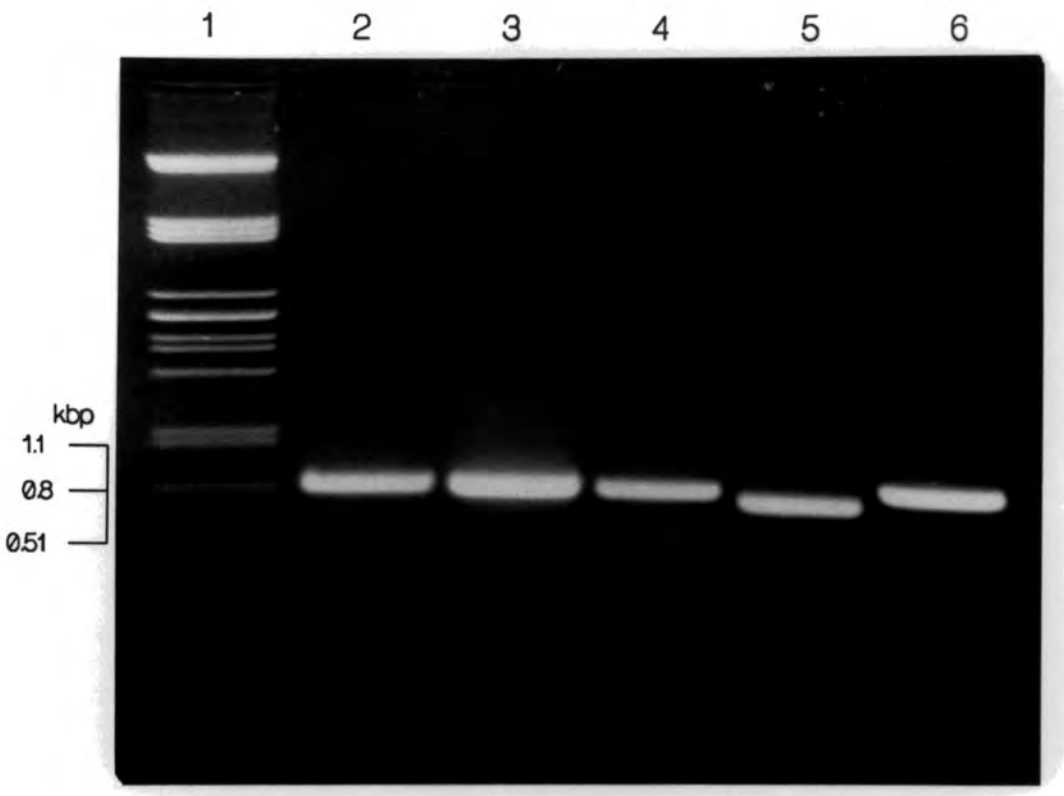
Each PCR DNA amplification contained approximately 15 ng of RIP template DNA, 1.0 μM each of 5′ and 3′ primer and 5.0 units of *Taq* polymerase in a total volume of 100μl. The reaction mixes were cycled automatically through the following temperature cycles.

1. 6 cycles of denaturation 94^0C/1 min, annealing 40^0C/1.5 min and extension 72^0C/2 min followed by:-
2. 19 cycles as described above except the annealing temperature was raised to 55^0C. The final extension time in the 19^{th} cycle was increased to 5 min.

At the end of the reaction 1.0 unit of T_4 DNA polymerase was added to each sample followed by incubation at 37^0C for 45 min. 5 μl of each reaction mix was then taken for analysis using agarose gel-electrophoresis. The remainder of each sample was phenol extracted, ethanol precipitated and resuspended in 10 μl of RNase free water for use in *in vitro* transcription reactions.

The gel photo shows the analysis of the products obtained after PCR using the DARTT primers. The expected size of product, which includes the addition of the SP6 promoter is indicated below for each template used in the PCR amplification. The size of each product estimated from the migration of the DNA size markers is in agreement with the expected size.

1. λ/PstI size markers
2. Ricin genomic clone (λRic 2ii) A-chain and promoter sequence = 836 bp
3. Ricin PCR product (pRic 1) ——————————/——————————
4. pλSap2.E, full-length saporin coding sequence and promoter = 828 bp
5. pλSap3, truncated saporin ——————————/—————— = 769 bp
6. pλSap3/4, hybrid saporin ——————————/—————— = 853 bp


1
2
3
4
5
6
kbp
1.1
0.8
0.51

6.4. *In vitro* transcription of RIP template DNAs.

RIP template DNAs produced using PCR amplification were *in vitro* transcribed and capped as described in Fig 6.6. The integrity of the RNA products was assessed by including [α-^{32}P]UTP in the reactions and analysing the products using glyoxal-agarose gel electrophoresis. The result of the transcription reactions are shown in Fig 6.6. Labelled RNA products were produced in each of the transcription reactions using the RIP template DNA. Estimates showed them to be approximately the expected size (lanes 1-5) and they were relatively undegraded. There was relatively less RNA produced from the ricin genomic clone λRic 2ii (lane 1) compared with the other reactions. No attempt was made to remove unincorporated UTP, seen as the diffuse radioactive areas at the bottom of the gel in each lane. This experiment showed that the RIP templates, generated using PCR, had incorporated the functional SP6 promoter sequences into the products and these constructs had served as templates for *in vitro* transcription reactions.

An estimate of the yield of transcript from each reaction based on the incorporation of [α-^{32}P]UTP is presented in Fig 6.7. The total amount of RNA synthesised in each reaction ranged from approximately 38 to 114 ng. the amount of RNA produced is markedly lower than the amounts obtained by Mackow *et al.* (1990). They synthesised approximately 20 μg of RNA in a 100 μl transcription reaction containing 0.1 to 1.0 μg of template DNA. The promoter sequence added to their templates was T_3 RNA polymerase. It is possible that the T_3 promoter transcribes more efficiently than the SP6 promoter sequence when the promoter sequence is at the extreme 5′ end of the template DNA.

6.5. Effect of RIP transcripts on protein synthesis in rabbit reticulocyte lysates.

The results presented previously show that the RIP gene sequences can be transcribed *in vitro*. To determine if the constructs encoded active RIP polypeptides the transcripts were translated using rabbit reticulocyte lysate. The initial investigation (derived from the protocol of May *et al.*, 1989) was designed to determine if the translated RIP polypeptides would cause inhibition of translation. This was determined by adding a second mRNA to the translation mix after an initial incubation with a RIP transcript. Measurement of the incorporation of radioactively labelled

Figure 6.6. *In vitro* **transcription and capping of RIP PCR constructs.**
In vitro transcription of RIP constructs was carried out as described in section 3.12 with the inclusion of [α-^{32}P]UTP. Transcript integrity was determined by electrophoresis of a 1 μl aliquot of the transcription reaction on a 1.5% (w/v) glyoxal-agarose gel (section 3.8.3). After electrophoresis the gel was dried and autoradiographed (75 min at room temperature). The autoradiograph shows the transcripts produced using the following RIP template DNAs produced by PCR, each containing the SP6 promoter. DNA size markers were glyoxalated λ EcoRI/HindIII restriction fragments.

1. Ricin genomic clone, λRic2ii
2. Ricin PCR clone, pRic1
3. Full-length saporin clone, pλSap2.E
4. Hybrid saporin clone, pλSap3/4
5. Truncated saporin clone, pλSap3

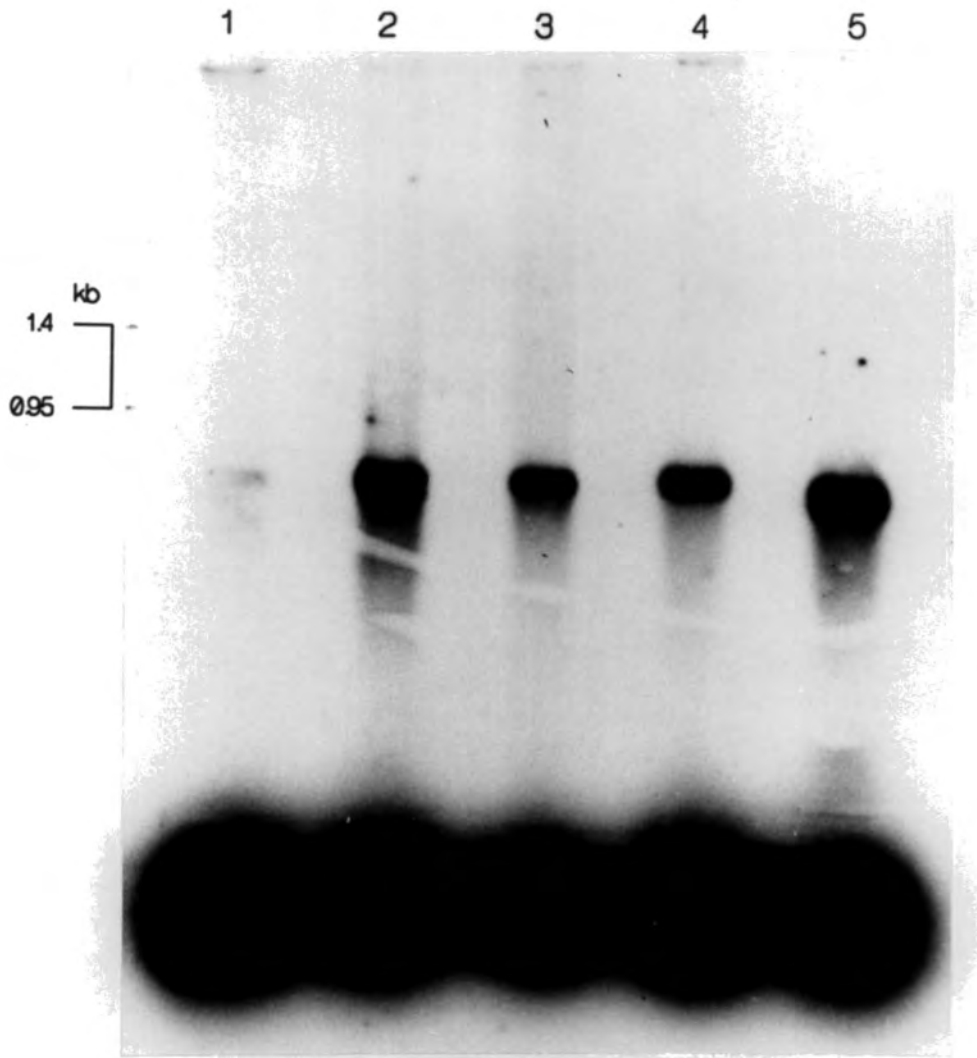

1
2
3
4
5
kb
1.4
0.95

Figure 6.7. Estimation of transcript yield from *in vitro* transcription of RIP constructs.

The yield of transcript obtained from the *in vitro* transcription of each RIP construct was determined by adding [α-^{32}P]UTP to each transcription reaction. Radioactive incorporation into TCA precipitable products was determined as described in section 3.12.2. Transcription reactions (total volume 12 μl) each contained 20 μCi [α-^{32}P]UTP (25 pmoles) and 625 pmoles unlabelled UTP present in the capping solution.

Template DNA.	ng RNA synthesised per 12 μl reaction.
Ricin genomic clone λRic 2ii	38
Ricin PCR clone pRic1	111
pλSap2.E	94
pλSap3/4	114
pλSap3	77

[^{3}H]-leucine into TCA precipitable product provided a means of determining the amount of translation. If incorporation was lowered or abolished, when compared with a control reaction not containing a RIP transcript, this indicated that a functional RIP polypeptide had been translated which inactivated ribosomes and prevented the translation of a second mRNA. The second mRNA used was poly(A)$^{+}$ RNA isolated from developing pea cotyledons.

Figure 6.8 shows the results of such experiments using each of the RIP transcripts synthesised as described previously. The experiment was carried out twice (Graphs A and B) in each case using newly synthesised transcripts to ensure that the results were reproducible. The control reaction, to which only the transcription reaction reagents were added without template DNA, had the highest incorporation of labelled leucine after the addition of poly (A)$^{+}$ RNA. This was observed in both experiments. The second control to determine the endogenous background incorporation had the lowest value in each case. Each of the translation reactions containing the RIP constructs showed significantly decreased incorporation determined after the addition of the second mRNA. The exception to this is the truncated Sap3 construct derived from pλSap3. In both experiments incorporation was only slightly lowered. This clone lacks the first 28 residues of the mature saporin coding sequence which may be a sufficiently large deletion of the polypeptide to abolish its activity.

The results of these experiments suggest that the RIP transcripts contain coding sequences producing active polypeptides capable of ribosome inactivation, the one exception being the Sap3 coding sequence. This experimental approach does not however show that protein synthesis is inhibited due to the specific depurination of the 28 S rRNA by RIP polypeptides. The experiment did however indicate that the RIP transcripts inhibited protein synthesis and as such it was worth determining if this was due to specific RIP action; it was possible for example that the observed inhibition was due to the presence of double-stranded RNA in the transcription mix. Double stranded RNA is known to inhibit protein synthesis in rabbit reticulocyte lysates (Jackson and Hunt, 1983).

Assuming this is not the case, and that the observed effect is due to specific RIP action, the experiment does not indicate the relative activities of polypeptides from each construct. In each case the amount

Figure 6.8. Inhibition of message-dependent protein synthesis in nuclease-treated rabbit reticulocyte lysate.

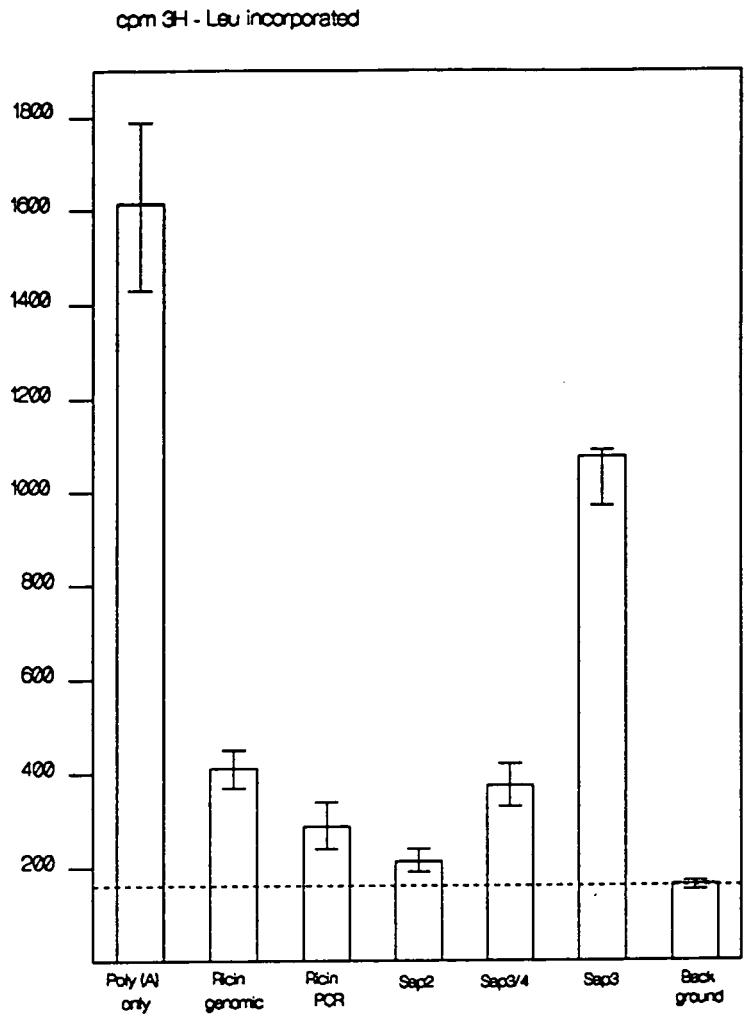
Non-labelled RIP transcripts were incubated with rabbit reticulocyte lysate as described in section 3.13.1. with the inclusion of L-[4,5-^{3}H]leucine. After 30 min incubation pea poly $(A)^{+}$ RNA was added to each translation reaction and incubation was continued for a further 30 min. Radioactive incorporation into TCA precipitable products was then determined by the method described in section 3.13.2. Two control reactions were included. A transcription reaction was set up without RIP DNA template. This was added to two translation reactions. Poly $(A)^{+}$ RNA was added to one translation reaction to determine total incorporation in the absence of RIP construct. The other had no message added and was used to determine the endogenous incorporation present.

This experiment was carried out twice, each time newly synthesised transcript being used in the translation reactions. The results are presented graphically.

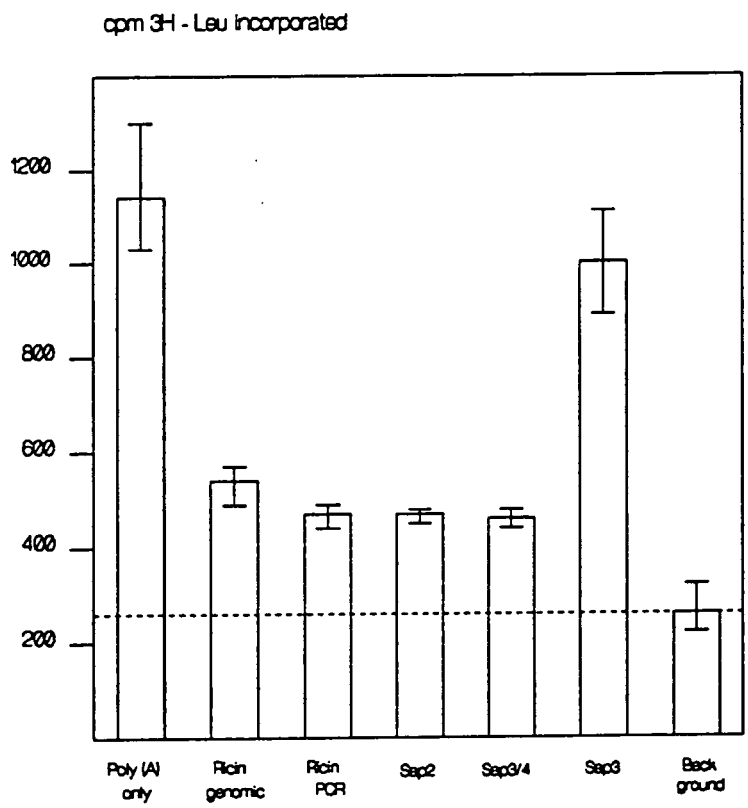
In the first experiment (graph A) duplicate 1μl samples were assayed; the mean counts are presented, the error bars showing the sampling range.

In the second experiment (graph B) triplicate 1 μl samples were taken. The results are presented as the mean counts, the error bars indicate the sampling range. The dotted line represents the background incorporation.

A


cpm 3H - Leu incorporated
1800
1600
1400
1200
1000
800
600
400
200
Poly (A) only
Ricin genomic
Ricin PCR
Sap2
Sap3/4
Sap3
Back ground

B


cpm 3H - Leu incorporated
1200
1000
800
600
400
200
Poly (A) only
Ricin genomic
Ricin PCR
Sap2
Sap3/4
Sap3
Back ground

of RIP polypeptide cannot be determined due to the nature of this "suicide system" (May *et al.*, 1989). A very active polypeptide will be synthesised to a lesser degree than a less active one yet each will inactivate ribosomes albeit at different rates. There is no meaningful way of comparing these activities, and as such, the system is qualitative but not quantitative, either translation is inhibited or it is not. A criticism of the experiment would be that the saporin-6 protein was not used in a control reaction to demonstrate inhibition of translation. This was not included because the experiment does not demonstrate specific RIP action, although it did show that RIP transcripts inhibited translation. In this case it was not felt necessary to show the inhibitory effect of saporin-6. Saporin proteins have previously been demonstrated to inhibit protein synthesis (Stirpe *et al.*, Gatehouse *et al.*, 1990a).

6.6. Modification of rabbit reticulocyte ribosomal RNA by saporin-6 protein and *in vitro* expressed RIP gene sequences.

To determine if the RIP transcripts encoded polypeptides which would inactivate rabbit ribosomes by specific depurination of 28S rRNA it was necessary to perform aniline cleavage assays. May *et al.* (1989) demonstrated that specific modification of rRNA in rabbit reticulocyte lysates by ricin A-chain polypeptides can be determined. If the 28 S rRNA is depurinated at a single specific site by RIP action a fragment is released after aniline treatment. This small fragment of approximately 390 ribonucleotides is specific for the action of ricin and other RIPs on rabbit reticulocyte ribosomes (Stirpe *et al.*, 1988). In some cases a small reduction in size of the 28 S RNA can be detected in aniline treated samples and is diagnostic for specific depurination of the 28 S RNA (Dr. M Hartley, personal communication). Thus rRNA aniline cleavage assays were performed to determine if the RIP transcripts produced in the previous experiments depurinated 28 S rRNA. RNA was isolated from lysates incubated with: RIP transcripts, saporin-6 protein or control lysates treated only with transcription mix without RIP template DNA. The RNA samples were then subjected to aniline treatment. The Sap3 construct was not included in these assays as it showed only slight inhibition of translation and therefore had little or no activity.

The initial experiments are shown in Fig 6.9. The experimental details are given in the figure legend. The control reaction was a

Figure 6.9. Susceptibility of ribosomal RNA to specific aniline-catalysed cleavage after incubation with purified saporin-6 protein or *in vitro* translated RIP transcripts.

The action of RIP constructs upon rabbit reticulocyte lysates was analysed using the aniline cleavage assay described in section 3.14.

Translation of RIP transcripts in micrococcal-nuclease treated lysates.

These translations were scaled up translation reactions described in section 3.13.3. Each translation contained 4 μl of transcription mix and 32 μl of lysate in a total volume of 40 μl. A control translation containing 1 μg of saporin-6 protein, using Promega supplied non-nuclease treated lysate as a source of ribosomes, was also included. The transcripts used were:-

1. Ricin genomic clone, λRic2ii
2. Ricin PCR clone, pRic1
3. Full-length saporin, pλSap2.E

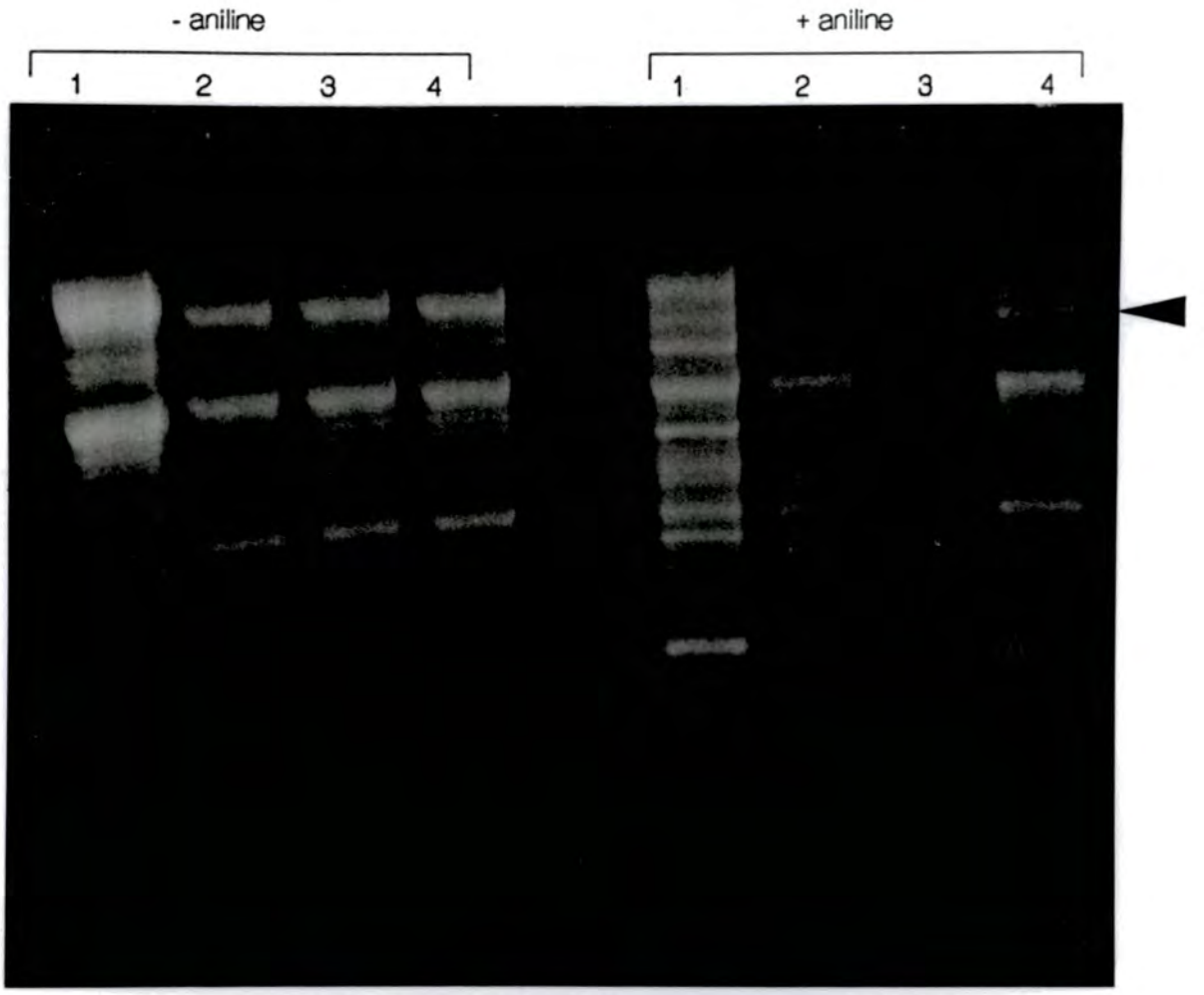
After incubation at 30^0C for 60 min the ribosomal RNA was extracted and an aliqout of each sample was subjected to aniline treatment (2 min at 60^0C in the dark) (section 3.14). The rRNA samples were then analysed by electrophoresis on a denaturing 1.2% (w/v) formamide-agarose gel (section 3.8.4) with the inclusion of non-aniline treated rRNA controls. Lanes 1-4 untreated rRNA samples, lanes 5-8 are rRNA samples treated with aniline. The arrow shows an extra fragment produced in the aniline treated samples.

Non-aniline treated samples.(-)

1. Saporin-6 treated lysate
2. Ricin genomic clone, λRic2ii transcript
3. Ricin PCR clone, pRic1 transcript
4. pλSap2.E transcript

Aniline treated samples.(+)

5. Saporin-6 treated lysate
6. Ricin genomic clone, λRic2ii transcript
7. Ricin PCR clone, pRic1 transcript
8. pλSap2.E transcript


- aniline
+ aniline
1
2
3
4
1
2
3
4

non-nuclease treated lysate treated with an excess of purified saporin-6 protein. The RIP transcripts translated in nuclease treated lysates; the two ricin clones and the full-length saporin clone Sap2.E.

The gel shows that the rRNA extracted from the nuclease treated lysates was extremely degraded before aniline treatment (lanes 2-4 of the untreated samples) in comparison with the non-nuclease treated sample (lane 1, the untreated sample). Aniline treatment of RNA extracted from the positive control results in the production of a fragment (the saporin-6 protein control, lane 1, + aniline) not seen in the untreated sample (lane 1, - aniline). Nuclease treatment of the lysate has degraded the rRNA. A fragment approximately the same size as that produced after aniline treatment of the control sample has been generated. This is present in the non-aniline treated samples (lanes 2, 3 and 4, untreated samples). This fragment prevents the detection of the fragment diagnostic for RIP action after aniline cleavage. In addition, the aniline treatment has resulted in further degradation of the ribosomal RNA, compare the rRNA treated with aniline (lane 1, + aniline) with the untreated sample (lane 1, - aniline).

An indication that the transcripts encode active polypeptides is the appearance of an extra rRNA band (arrowed) in the aniline treated samples, lanes 2-4. This is slightly smaller than the 28 S rRNA band seen above it. The appearance of this band after aniline treatment is due to the cleavage of the rRNA at the depurinated site. The smaller diagnostic fragment is released but cannot be detected, however the size of the 28 S rRNA is reduced and this difference is detectable on the gel (Fig 6.9, arrowed). This result indicated that the transcripts probably encoded functional RIP polypeptides which had specifically modified the 28 S rRNA. Unfortunately using nuclease treated lysate the results were not completely unequivocal, it would be better to demonstrate the release of the small RNA fragment. However, it is likely that the coding sequences used; the ricin genomic clone (λRic 1ii), the ricin PCR clone (pRic1) and the full-length saporin clone (pλSap2.E) do encode active RIP polypeptides when this result is considered with inhibition of translation experiments presented in Fig 6.8.

A second experiment was performed to determine if the transcripts encoded active polypeptides in which only the transcripts of Sap2 and the hybrid sequence Sap3/4 were used. Ricin A-chain has already been shown to

inactivate ribosomes using this type of assay (May *et al.*, 19889), hence it was decided to concentrate on the saporin coding sequences. To get around the problem of non-specific RNA fragments present in the nuclease treated lysate non-nuclease treated reticulocyte lysate was used.

The transcripts of Sap2 and Sap3/4 coding sequences were translated in non-nuclease treated lysates with suitable control samples as described in Fig 6.10. rRNA extracted from the lysates was then subject to aniline treatment. Non-nuclease treated lysate was used to minimise the problem of rRNA degradation observed in the previous assays using nuclease treated lysate. In addition, the conditions of aniline treatment were changed. In this experiment both the time and temperature at which the rRNA is exposed to aniline was reduced to minimise the non-specific cleavage noted previously.

Fig 6.10 shows the results of the aniline cleavage assays performed on the four separate samples. There are approximately equal loadings of RNA in each of the gel tracks. The control reaction with transcription mix only (lane 4) showed that the rRNA was relatively intact even after treatment with aniline. This is in contrast to rRNA isolated from the previous nuclease treated lysates (see Fig 6.9). Significantly a small RNA fragment was released, after aniline treatment, in the rRNA samples that had been previously incubated with either saporin-6 protein or the saporin transcripts (samples 1-3). This fragment was not detected in the control reaction (sample 4). The amount of saporin-6 protein added (1 μg) to the translation reaction was in massive excess above that required to inhibit translation, it was necessary to show that treatment of ribosomes with saporin-6 would cause depurination of the 28 S rRNA. The intensity of fluorescence of the small fragment released is approximately the same in each case. This suggests that the translated saporin proteins have depurinated about the same amount of rRNA as the saporin-6 protein sample. There is no indication in any of the samples of the proportion of ribosomes present in the translation mix which have been inactivated. As discussed earlier, this assay does not give a quantitative determination of activity.

Treatment with an excess of saporin-6 should produce virtually complete inactivation. The same saporin-6 preparation was used to inhibit translation in reticulocyte lysates at concentrations at least as low as the published ID_{50} values for saporin-6 (0.037 nM) (Gatehouse *et al.*,

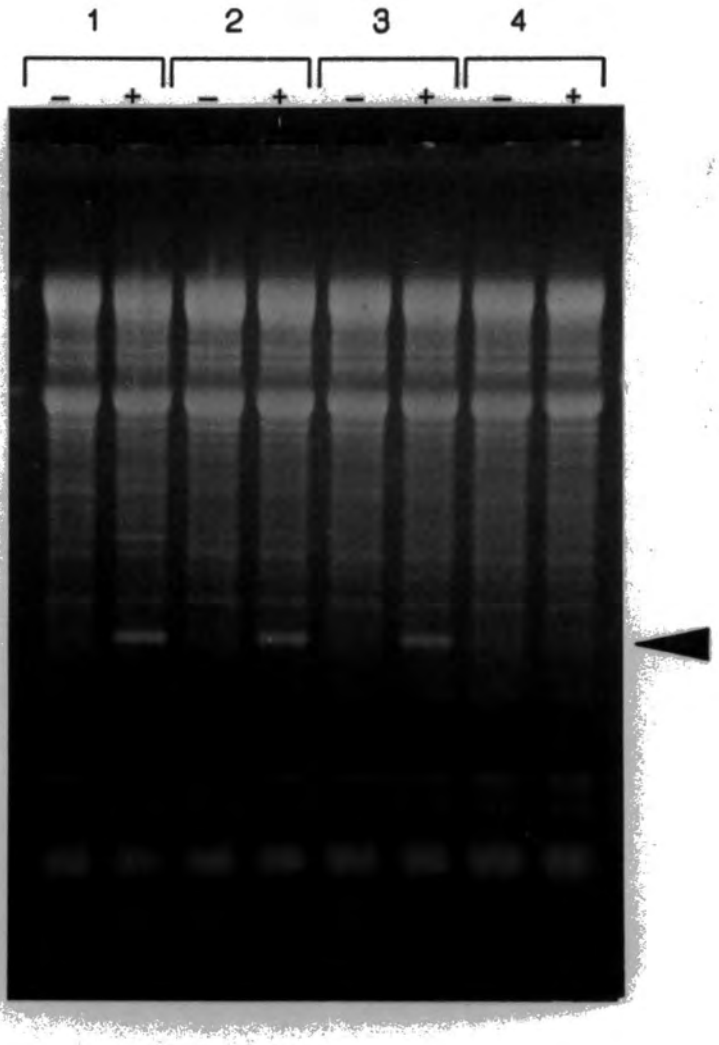
Fig 6.10. Translation of RIP transcripts in non-nuclease treated rabbit reticulocyte lysates.

Non-nuclease treated lysates (Department of Biochemistry, University of Cambridge) were used to assess the activity of RIP transcripts as described in section 3.13.3. 3 μl of transcription mix was incubated with the lysate in a total volume 40 μl, for 2 hours at 30^0C. Control reactions were lysate incubated with 1μg purified saporin-6 protein and lysate incubated with transcription mix without template DNA. The two RIP transcripts used were pλSap2.E and pλSap3/4. Ribosomal RNA was purified from each incubation and an aliquot of each sample was subjected to aniline treatment (2 min at room temperature in the dark) as described in section 3.14. Products were then analysed by formamide-agarose gel electrophoresis (section 3.8.4) with controls of untreated rRNA samples. The sample designation is shown below:-

(-) untreated control sample
(+) aniline treated sample

1. Saporin-6 treated lysate (-)
   ——/————————/———/—— (+)
2. pλSap2.E, full-length saporin (-)
   ——/————————/————————/—— (+)
3. pλSap3/4 hybrid saporin (-)
   ——/————————/———/—— (+)
4. Transcription mix, no DNA template (-)
   ————/————————/———/———/—— (+)

The RNA fragment specifically produced after aniline treatment is arrowed.


1
2
3
4
–
+
–
+
–
+
–
+

1990a; Stirpe and Barbieri, 1986).

This result demonstrated unequivocally that polypeptides produced from the saporin transcripts acted on rRNA in the same manner as the saporin-6 protein. Mature saporin-6 protein has been demonstrated to inactivate ribosomes in the same manner as ricin A-chain (Stirpe *et al.*, 1988). The conclusion to this experiment is that both the full-length saporin clone, pλSap2, and the hybrid saporin clone, pλSap3/4, encode functional polypeptides. When these are synthesised *in vitro* they adopt active conformations and are capable of specific depurination of the 28 S rRNA and ribosome inactivation.

6.7. General comments on the *in vitro* expression studies presented.

The results presented in this section provide good evidence that the saporin coding sequences isolated contain active polypeptide sequences. It is unfortunate that one of the sequences was a hybrid derived from two different genomic clones, however the hybrid sequence did show good activity.

The truncated saporin clone, pλSap3, exhibited little, if any, inhibitory activity and can be considered non-functional. The N-terminal truncation of Sap3 was 28 residues. May *et al.* (1989) demonstrated that deletion of the first 12 N-terminal residues of the mature ricin A-chain was sufficient to render the polypeptide inactive whilst deletion of 9 N-terminal residues had no affect on activity. The inactivity of Sap3 is consistent with this observation and suggests that residues at the N-terminal end of the RIP coding sequences are required for activity of the saporin polypeptide.

One further point regarding the activity of the saporin coding sequences is the presence of conserved residues at the presumed active site of RIP polypeptides. Comparison of known RIP sequences from many different sources has allowed for the identification of conserved residues implicated in the catalytic site of the proteins. This is illustrated in Fig 5.16 and discussed in section 5.5.2. The sequence of Sap2 has two amino acid substitutions at positions which are supposedly invariant. These are a valine for alanine substitution (conservative change) and a serine for aspartic acid substitution (non-conservative change) at positions 177 and 207 respectively in the mature Sap2 polypeptide. The other residues of Sap2 at invariant positions are conserved. This

suggests that these residues are not critical for Sap2 RNA N-glycosidase activity, although the specific activity of the protein is not known and may be somewhat reduced when compared with say, ricin A-chain or saporin-6. It is also possible that the residue replacements at these two positions are compensated for by other amino acid changes at other positions within the polypeptide, thus maintaining an active conformation of the protein.

Further work is required to establish if the observed differences in the Sap2 residues are of any importance. This could be accomplished by purifying the Sap2 protein and comparing its specific activity with the other characterised RIPs and by changing the two residues back to the invariant amino acids by site-directed mutagenesis followed by examination of specific activity.

The results presented do not indicate the position within the rabbit 28 S rRNA at which the *in vitro* expressed RIP polypeptides act. Ricin removes the adenine residue at position 4324 in rat 28 S rRNA (Endo *et al.*, 1987). May *et al.* (1989) reported the site modified by ricin A-chain in rabbit 28 S rRNA was at or near to the comparable site in rat 28 S rRNA. Sequence analysis of the rRNA around this site showed the rabbit sequence to be the same as the rat RNA sequence and it appears likely that the homologous adenine residue is removed. Saxena *et al.* (1989) demonstrated that Shiga toxin, Shiga-like toxin II and ricin specifically remove adenine at position 3732 of *Xenopus* 28 S rRNA. This site is analogous to the site modified in rat 28 S rRNA and the sequence around this site is identical to the rat RNA sequence. Conservation of this sequence between rat, *Xenopus* and yeast makes it extremely likely that the same sequence is present in rabbit 28 S rRNA.

An aniline-labile site in the reticulocyte 28 S rRNA is created after treatment with either saporin-6 protein or *in vitro* translated RIP transcripts. This data considered, with the observations of other workers described above, shows the specific action of the polypeptides derived from the characterised saporin genomic clones. It is sufficient to show the release of the diagnostic RNA fragment after aniline cleavage to demonstrate specific RIP action.

The strategy employed to express the RIP coding sequences was remarkably successful. Using PCR amplification to add the transcriptional and translational sequences required for the *in vitro* expression studies

presented, obviated the need for any cloning steps and allowed the length of coding sequence employed to be exactly defined. There is no requirement to rely on the presence of convenient restriction sites to, for example, remove leader sequences or insert oligonucleotides containing in-frame translational initiation codons. The results showed that a relatively small amount of transcript was produced from the modified template DNAs. No assessment can be made of the amount of protein synthesised in the *in vitro* translation reactions, however, in this case there was a sensitive method to determine the activity of the expressed RIP sequences. This meant that the low yields of transcript were of little consequence in these investigations.

The technique has great potential for the expression of defined regions of any coding sequences as long as there is a suitable sensitive assay for the polypeptides produced. Thus it is possible to use PCR primers to amplify coding sequences directly from genomic or cDNA such that the product contains all the sequences required for expression. This shows the potential for this technique in the analysis of coding sequences without the need for gene cloning. This was demonstrated by Sarkar and Sommer (1989) who expressed the catalytic domain of blood coagulation factor IX after amplification of the sequence from cDNA and incorporating the appropriate expression signals (in this case a T_7 promoter sequence was included 5′ to the amplified coding sequence). The experimental strategy is of obvious use in the analysis of regions of already coding sequence already characterised. Using the example of the RIP sequences, it was possible to express the ricin A-chain from a clone which contained the coding sequence of both A- and B-chains. As such, the technique will probably find great use in future work for the analysis of structure-function relationships of proteins for which a coding sequence has been isolated.

7. Concluding Remarks And Prospects For Future Research.

The results presented in this work describe the analysis of ricin, agglutinin and saporin coding sequences. The aim of the project was successfully accomplished, in that functional RIP coding sequences have been isolated. Three different genomic clones containing saporin genes were isolated, characterised and two were shown to encode active polypeptides. The project also led to the development of PCR strategies which were demonstrably successful and of great utility.

Without question the most important aspect of the work was the investigation of the saporin gene family. The isolation of genomic clones for saporin genes provides the basis for further investigation into a group of proteins of great interest. The major thrust of RIP research is towards the cloning and expression of the genes encoding RIPs and the mechanisms involved in ribosome inactivation. A further development of the field will be to investigate the actual *in planta* expression and function of RIPs.

Concerning the saporin gene family it would be of interest to address the following questions:

1. Are the saporin genes differentially expressed within the plant? Does the expression of the genes change with the season as in the case of the pokeweed antiviral proteins?
2. Where exactly are the different saporin proteins localised at the subcellular level? Are some of the proteins targeted to the cell wall matrix and if so do they serve an antiviral function? A further relevant point would be to determine if *Saponaria* ribosomes are themselves inactivated by saporin proteins.
3. Are the differences at the C-terminal ends of the two sequenced saporin genes involved in differential targeting of the proteins?

To attempt to find answers to some of these questions, using the saporin gene family as the experimental system the following programme could be adopted:-

Saporin gene expression in plant tissues (ie seed, leaf, stem and root) can be analysed by Northern blotting of RNA and hybridisation with saporin coding sequences. Saporins are extremely abundant in the seeds and it would be of interest to determine if saporin genes are expressed in a

similar manner to other seed proteins during embryogenesis. Further expression studies can be carried out using gene specific probes derived from the divergent regions of the 3′ untranslated sequences of the two saporin genes Sap2 and Sap4. This would indicate if the two genes are differentially expressed. The use of anti-saporin antibodies may provide information as to the subcellular localisation of the saporin proteins.

The sequence of the two saporin promoters should be completely determined and compared with other plant promoter sequences. The limited sequencing so far carried out indicated that they shared little similarity. As a hypothesis it may well be that these saporin genes are differentially expressed due to the different promoter sequences. One way of testing this would be to fuse the promoters to a suitable reporter gene, for example β-glucuronidase, transform the constructs into tobacco and compare the subsequent patterns of expression for each promoter sequence. This would indicate if the sequences confer tissue and/or developmental specificity upon the expression of saporin genes.

In a similar manner the putative C-terminal propeptide sequences could be fused to a reporter gene and expressed in transgenic plants to determine if they contain targeting information. The constructs would each have the same leader sequence and reporter gene but would either have the C-terminal extension or lack it. An alternative approach would be to use the saporin coding sequences with and without the C-terminal extensions in transgenic plants. The leader sequences would have to be required in the construct to ensure the co-translational sequestration of the saporin polypeptide into the endoplasmic reticulum. Given the availability of anti-saporin antibodies it is feasible to determine the site of saporin localisation within different tissues. It is presumed that the protein is localised in the protein bodies of seeds but where is the site of localisation in the leaves?

A further advantage of producing transgenic plants expressing active saporin polypeptides would be the ability to test if their expression conferred any viral resistance to the transformed plant. If this was the case then it would also implicate saporins as a viral defence mechanism within the plant, already a suggested role for the pokeweed antiviral proteins (Taylor and Irvin, 1990).

6. REFERENCES.

Araki T, Funatsu G (1986) Revised amino acid sequence of the B-chain of ricin D due to loss of tryptophan in the cyanogen bromide cleavage. FEBS Letts 191:121-124

Araki T, Funatsu G (1987) The complete amino acid sequence of the B chain of ricin E isolated from small-grain castor bean seeds. Ricin E is a gene recombination product of ricin D and *Ricinus communis* agglutinin. Biochim Biophys Acta 911:191-200

Asano K, Svensson B, Poulsen FM (1984) Isolation and characterization of inhibitors of animal cell-free protein synthesis from barley seeds. Carlsberg Res Commun 49:619-626

Asano K, Svensson B, Svendson I, Poulsen FM (1986) The complete primary structure of protein synthesis inhibitor II from barley seeds. Carlsberg Res Commun 51:129-141

Baezinger JM, Fiete D (1979) Structural determinants of *Ricinus communis* agglutinin and toxin specificity for oligosaccharides. J Biol Chem 254:9795-9799

Barbieri L, Aron GM, Irvin JD, Stirpe F (1982) Purification and partial characterisation of another form of the antiviral protein from the seeds of *Phytolacca americana* L. (pokeweed). Biochem J 203:55-59

Barbieri L, Stirpe F (1982) Ribosome-inactivating proteins from plants: Properties and possible uses. Cancer Surveys 1:489-520

Barbieri L, Bolognesi A, Cenini P, Falasca AI, Minghetti A, Garofano L, Guicciardi A, Lappi D, Miller SP, Stirpe F (1989) Ribosome -inactivating proteins from plant cells in culture. Biochem J 257:801-807

Battelli MG, Lorenzoni E, Stirpe F, Cella R, Parisi B (1984) Differential effect of ribosome-inactivating proteins on plant ribosome activity

and plant cells growth. J Expt Bot 35:882-889

Bauer GA, Burgers PMJ (1990) Molecular cloning, structure and expression of the yeast proliferating cell nuclear antigen gene. Nucleic Acids Res 18:261:265.

Bednarek SY, Wilkins TA, Dombrowski JE, Raikhel NV (1990) A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. The Plant Cell 2:1145-1155

Benatti L, Saccardo MB, Dani M, Nitti G, Sassano M, Lorenzetti R, Lappi DA, Soria M (1989) Nucleotide sequence of cDNA coding for saporin-6, a type-1 ribosome-inactivating protein from *Saponaria officinalis*. Eur J Biochem 183:465-470

Berent SL, Massoud M, Torczynski RM, Bragg PW, Bollon AP (1985) Comparison of oligonucleotides and long DNA fragments in DNA and RNA dot, Southern, northern, colony and plaque hybridizations. BioTechniques (May/June) 208-220

Birnboim HC, Doly J (1979) A rapid extraction procedure for screening recombinant DNA. Nucleic Acids Res 7:1513-1523

Bjorn MJ, Larrick J, Piatak M, Wilson KJ (1984) Characterisation of translational inhibitors from *Phytolacca americana*. Amino-terminal sequence determination and antibody-inhibitor conjugates. Biochem Biophs Acta 790:154-163

Blakey DC, Wawrzynczak EJ, Wallace PM, Thorpe PE (1988) Antibody toxin conjugates: a perspective. In: Waldmann H (ed) Monoclonal Antibody Therapy. Prog Allergy. Basel, Karger, vol 45, pp 50-90

Bown D, Levasseur, Croy RRD, Boulter D, Gatehouse JA (1985) Sequence of a pseudogene in the legumin gene family of pea (*Pisum sativum* L.) Nucleic Acids Res 13:4527-4537

Bradley JL, McGuire PM (1990) Site-directed mutagenesis of ricin A chain

Trp 211 to Phe. Int J Peptide Res 35:365-366

Calderwood SB, AuClair F, Donohue-Rolfe A, Keusch GT Mekalanos JJ (1987) nucleotide sequence of the Shiga-like toxin genes from *Escherichia coli*. Proc Natl Acad Sci 84:4364-4368

Cassellas P, Dussossoy D, Falasca AI, Barbieri L, Guillemot JC, Ferrera P, Bolognesi A, Cenini P, Stirpe F (1988) Trichokirin, a ribosome -inactivating protein from the seeds of *Trichosanthes kirilowii* Maximowicz. Eur J Biochem 176:581-588

Cawley DB, Hedblom ML, Houston LL (1978) Homology between ricin and *Ricinus communis* agglutinin: Amino terminal sequence analysis and protein synthesis inhibition studies. Arch Biochem Biophys 190:744-755

Cawley DB, Hedblom ML, Houston LL (1979) Protection and rescue of ribosomes from the action of ricin A chain. Biochem 18:2648-2654

Chan YL, Olvera J, Wool IG (1983) Structure of rat 28 S ribosomal ribonucleic acid inferred from the sequence of nucleotides in a gene. Nucleic Acids Res 11:7819-7831

Chang M-S, Russell DW, Uhr JW, Vitetta ES (1987) Cloning and expression of recombinant, functional ricin B chain. Proc Natl Acad Sci 84:5640 -5644

Chow TP, Feldman RA, Lovett M, Piatak M (1990) Isolation and DNA sequence of a gene encoding α-trichosanthin, a type 1 ribosome-inactivating protein. J Biol Chem 265:8670-8674

Coleman WH, Roberts WK (1981) Factor requirements for the tritin inactivation of animal cell ribosomes. Biochim Biophys Acta 654:57-66

Coleman WH, Roberts WK (1982) Inhibitors of animal cell-free protein synthesis from grains. Biochim Biophys Acta 696:239-244

Collins EJ, Robertus JD, LoPresti M, Stone KL, Williams KR, Wu P, Hwang K, Piatak M (1990) Primary amino acid sequence of α-trichosanthin and molecular models for abrin A-chain and α-trichosanthin. J Biol Chem 265:8665-8669

Darnborough C, Legon S, Hunt T, Jackson RJ (1973) Initiation of protein synthesis: evidence for messenger RNA-independent binding of methionyl-transfer RNA to the 40 S ribosomal subunit. J Mol Biol 15:4364-4369

Dranginis AM (1990) Binding of yeast a1 and α2 as a heterodimer to the operator DNA of a haploid-specific gene. Nature 347:682-685

Eckert KA, Kunkel TA (1990) High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res 18:3739-3744

Endo Y, Wool IG (1982) The site of action of α-sarcin on eukaryotic ribosomes. J Biol Chem 257:9054-9060

Endo Y, Huber PW, Wool IG (1983) The ribonuclease activity of the cytotoxin α-sarcin. J Biol Chem 258:2662-2667

Endo Y, Mitsui K, Motizuki M, Tsurugi K (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. J Biol Chem 262:5908-5912

Endo Y, Tsurugi K (1987) RNA N-glycosidase activity of ricin A-chain. J Biol Chem 262:8128-8130

Endo Y, Tsurugi K (1988) The RNA N-glycosidase activity of ricin A-chain. J Biol Chem 263:8735-8739

Endo Y, Tsurugi K, Lambert JM (1988a) The site of action of six different ribosome-inactivating proteins from plants on eukaryotic ribosomes: The RNA N-glycosidase activity of the proteins. Biochem Biophys Res Comm 150:1032-1036

Endo Y, Tsurugi K, Yutsudo T, Takeda Y, Ogasawara T, Igarashi K (1988b) Site of action of a Vero toxin (VT2) from *Escherichia coli* 0157:H7 and of Shiga toxin on eukaryotic ribosomes. FEBS Letts 171:45-50

Endo Y, Oka T, Tsurugi K, Franz H (1989) The mechanism of action of the cytotoxic lectin from *Phoradendron californicum*: the RNA N-glycosidase activity of the protein. FEBS Letts 248:115-118

Endo Y, Glück A, Chan Y-C, Tsurugi K, Wool IG (1990) RNA-protein interaction. J Biol Chem 265:2216-2222

Evans IM, Gatehouse LN, Gatehouse JA, Yarwood JN, Boulter D, Croy RRD (1990) The extensin gene family in oilseed rape (*Brassica napus* L.): Characterisation of sequences of representative members of the family. Mol Gen Genet 223:273-287

Fedoroff N (1983) Comparison of host strains for cloning maize DNA in bacteriophage λ. Plant Molec Biol Reporter 1:27-29

Fitzgerald M, Shenk T (1981) The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell 24:251-260

Fordham-Skelton AP, Yarwood A, Croy RRD (1990) Synthesis of saporin gene probes from partial protein sequence data: Use of inosine-oligonucleotides, genomic DNA and the polymerase chain reaction. Mol Gen Genet 221:134-138

Frankel A, Welsh P, Richardson J, Robertus JD (1990) Role of arginine 180 and glutamic acid 177 of toxic ricin A chain in enzymatic inactivation of ribosomes. Molec Cell Biol 10:6257-6263

Frischauf AM, Lehrach H, Poutska A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827-842

Funatsu G, Yoshitake S, Funatsu M (1978) Primary structure of Ile chain of ricin D. Agric Biol Chem 42:501-503

Gatehouse JA, Bown D, Evans IM, Gatehouse LN, Jobes D, Preston P, Croy RRD (1987) Sequence of the seed lectin gene from pea (*Pisum sativum* L.). Nucleic Acids Res 15:7642

Gatehouse AMR, Barbieri L, Stirpe F, Croy RRD (1990a) Effects of ribosome inactivating proteins on insect development - differences between *Lepidoptera* and *Coleoptera*. Entomol Exp Appl 54:43-51

Gatehouse LN, Evans IM, Gatehouse JA, Croy RRD (1990b) Characterisation of a rape (*Brassica napus* L.) extensin gene encoding a polypeptide relatively rich in tyrosine. Plant Science 71:223-231

Girgis SI, Alevizaki M, Denny P, Ferrier GJM, Legon S (1988) Generation of DNA probes for peptides with highly degenerate codons using mixed primer PCR. Nucleic Acids Res 26:10371

Gould SJ, Subramani S, Schefler IE (1989) Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. Proc Natl Acad Sci USA 86:1934-1938

Golden SS, Brusslan J, Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. In: Wu R, Grossman L (eds) Methods in Enzymology. Volume 153. Recombinant DNA. Part D. Academic Press, Inc. (London) Ltd, pp 215-231

Graham DE (1978) The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal Biochem 85:609-613

Habuka N, Murakami Y, Noma M, Kudo T, Horikoshi K (1989) Amino acid sequence of *Mirabilis* antiviral protein, total synthesis of its gene and expression in *Escherichia coli*. J Biol Chem 264:6629-6637

Habuka N, Akiyama K, Tsuge H, Miyano M, Matsumoto T, Noma M (1990) Expression and secretion of *Mirabilis* antiviral protein in *Escherichia coli* and its inhibition of *in vitro* eukaryotic and prokaryotic protein

synthesis. J Biol Chem 265:10988-10992

Halling KC, Halling AC, Murray EM, Ladin BF, Houston LL, Weaver RF (1985) Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucleic Acids Res 13:8019-8033

Harley SM, Beevers H (1982) Ricin inhibition of *in vitro* protein synthesis by plant ribosomes. Proc Natl Acad Sci 79:5935-5938

Harley SM, Beevers H (1984) Ricin inhibition of *in vivo* protein-synthesis in castor beans. Plant Sci Letts 36:1-5

Heidecker G, Messing J (1986) Structural analysis of plant genes. Ann Rev Plant Physiol 37:439-466

Hilder VA, Barker RF, Samour RA, Gatehouse AMR, Gatehouse JA, Boulter D (1989) Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.). Plant Molec Biol 13:701-710

Hohn B, Leibrock J, Bailey K, Barde Y-A (1990) Identification and characterization of a novel member of the nerve growth/brain-derived neurotrophic factor family. Nature 344:339-341

Holdsworth MJ, Schuch W, Grierson D (1988) Organisation and expression of a wound/ripening-related small multigene family from tomato. Plant Molec Biol 11:81-88

Hollis GF, Heiter PA, McBride OW, Swan D, Leder P (1982) Processed genes: dispersed immunoglobulin gene bearing evidence of RNA-type processing. Nature 296:321-325

Houston LL (1982) Transport of ricin a chain after prior treatment of mouse leukemia cells with ricin B chain. J Biol Chem 257:1532-1539

Houston LL, Dooley TP (1982) Binding of two molecules of 4- methyl -lumbelliferyl galactose to the B chains of ricin and *Ricinus communis*

agglutinin and to purified ricin B chain. J Biol Chem 257:4147-4151

Houston LL, Ramakrishnan S, Hermodson MA (1983) Seasonal variation in different forms of pokeweed antiviral protein, a potent inactivator of ribosomes. J Biol Chem 258:9601-9604

Hovde CJ, Calderwood SB, Mekalanos JJ, Collier RJ (1988) Evidence that glutamic acid 167 is an active-site residue of Shiga-like toxin. Proc Natl Acad Sci 85:2568-2572

Irvin JD (1975) Purification and partial characterisation of the antiviral protein from *Phytolacca americana* which inhibits eukaryotic protein synthesis. Arch Biochem Biophys 169:522-528

Irvin JD, Kelly T, Robertus JD (1980) Purification and properties of a second antiviral protein from *Phytolacca americana* which inactivates eukaryotic ribosomes. Arch Biochem Biophys 200:418-425

Islam MR, Nishida H, Funatsu G (1990) Complete amino acid sequence of luffin-a, a ribosome inactivating protein from the seeds of *Luffa cylindrica*. Agric Biol Chem 54:1343-1345

Jackson MP, Neill RJ, O'Brien AD, Holmes RK, Neuland JW (1987) Nucleotide sequence analysis and comparison of the structural genes for Shiga-like toxin I and Shiga-like toxin II encoded by bacteriophage from *Escherichia coli* 933. FEMS Letts 44:109-114

Jackson RJ, Hunt T (1983) Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. In: Fleischer S, Fleischer B (eds) Methods in Enzymology. Volume 96. Biomembranes. Part J. Membrane Biogenesis: Assembly and Targeting (General Methods, Eukaryotes). Academic Press, Inc. (London) Ltd, pp 50-74

Jiménez A and Vázquez D (1985) Plant and fungal protein and glycoprotein toxins inhibiting eukaryote protein synthesis. Ann Rev Microbiology 39:649-672

Joshi (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res 15:6643-6653

Karlovsky P (1990) Misuse of PCR. Trends Biochem Sci 15:419

Karn J, Brenner S, Barnett L, Cesareni G (1980) Novel bacteriophage λ cloning vector. Proc Natl Acad Sci 77:5172-5176

Kassanis B, Kleczkowski A (1948) The isolation and some properties of a virus-inhibiting protein from *Phytolacca esculenta*. J Gen Microbiol 2:143-153

Kawasaki ES, Clark SS, Coyne MY, Smith SD, Champlin R, Witte ON, McCormick FP (1988) Diagnosis of chronic myeloid and acute lymphocytic leukemias by detection of leukemia-specific mRNA sequences amplified *in vitro*. Proc Natl Acad Sci 85:5698-5702

Keohavong P, Thilly WG (1989) Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci 86:9253-9257

Kishida K, Masuho Y, Takeshi H (1983) Protein synthesis inhibitory proteins from the seeds of *Luffa cylindrica* roem. FEBS Letts 153:209-212

Knoth K, Roberds S, Poteet C, Tamkun M (1988) Highly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction. Nucleic Acids Res 16:10932

Kumon K, Sasaki J, Sejima M, Takeuchi Y, Hayashi Y (1990) Interactions between tobacco mosaic virus, pokeweed antiviral proteins, and tobacco cell wall. Phytopathol 80:636-641

Kwok S, Kellog DE, McKinney N, Spasic D, Goda L, Levenson, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nucleic

Acids Res 18:999-1005

Ladin BF, Murray EE, Halling AC, Halling KC, Tilakaratne N, Long GLL, Houston LL, Weaver RF (1987) Characterization of a cDNA encoding ricin E, a hybrid ricin-*Ricinus communis* agglutinin gene from the castor plant *Ricinus communis*. Plant Molec Biol 9:287-295

Lamb FI, Roberts LM, Lord JM (1985) Nucleotide sequence of cloned cDNA coding for preproricin. Eur J Biochem 148:265-270

Lappi DA, Esch FS, Barbieri L, Stirpe F, Soria M (1985) Characterization of a *Saponaria officinalis* seed ribosome-inactivating protein: immunoreactivity and sequence homologies. Biochem Biophys Res Comm 129:934-942

Larrick JW, Danielsson L, Brenner CA, Wallace GF, Abrahamson M, Fry KE, Borreback CAK (1989) Polymerase chain reaction using mixed primers: cloning of human monoclonal antibody variable region genes from single hybridoma cells. Biotechnology 7:934-938

Lee CC, Wu X, Gibbs RA, Cook RG, Muzny DM, Caskey TC (1988) Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. Science 239:1288-1291

Leibrock J, Lottspeich F, Hohn A, Hofer M, Hengerer B, Masiakowski P, Thoenen H, Barde Y-A (1989) Molecular cloning and expression of brain-derived neurotrophic factor. Nature 341:149-152

Lewis RM (1986) Probfind: a computer program for selecting oligonucleotides from peptide sequences. Nucleic Acids Res 14:567-570

Lindahl G, Sironi G, Bialy H, Calendar R (1970) Bacteriophage lambda; Abortive infection of bacteria lysogenic for phage P2. Proc Natl Acad Sci USA 66:587-594

Linthorst HJM, Melchers LS, Mayer A, van Roekel JSC, Cornelissen BJC, Bol JF (1990) Analysis of gene families encoding acidic and basic

β-1,3-glucanases of tobacco. Proc Natl Acad Sci 87:8756-8760

Little PFR (1982) Globin pseudogenes. Cell 28:683-684.

Lodge JK, Kaniewski W, Turner N (1990) Characterization and cloning of the pokeweed antiviral protein. J Cellular Biochem Supplement 14(E):304

Loening UE (1969) The determination of the molecular weight of ribonucleic acid by polyacrylamide-gel electrophoresis. Biochem J 113:131-138

Loomis WF, Gilpin ME (1986) Multigene families and vestigial sequences. Proc Natl Acad Sci 83:2143-2147

Lopéz-Otin C, Barber D, Férnandez-Luna JL, Soriano F, Méndez E (1984) The primary structure of the cytotoxin restrictocin. Eur J Biochem 143:621-634

Lord JM (1985a) The structure and synthesis of plant lectins. New Phytol 101:351-366

Lord JM (1985b) Precursors of ricin and *Ricinus communis* agglutinin. Eur J Biochem 146:411-416

Lord JM (1987) The use of cytotoxic plant lectins in cancer therapy. Plant Physiol 85:1-3

Lycett GW, Croy RRD, Shirsat AH, Richards DM, Boulter D (1985) The 5'-flanking regions of three pea legumin genes: comparison of the DNA sequences. Nucleic Acids Res 13:6733-6743

Mackow ER, Yamamnaka MY, Dang MN, Greenberg HB (1990) DNA amplification -restricted transcription-translation: Rapid analysis of rhesus rotavirus neutralization sites. Proc Natl Acad Sci 87:518-522

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Maraganore JM, Joseph M, Bailey MC (1987) Purification and characterisation of trichosanthin. J Biol Chem 262:11628-11633

May MJ, Hartley MR, Roberts LM, Krieg PA, Osborn RW, Lord JM (1989) Ribosome inactivation by ricin A chain: a sensitive method to assess the activity of wild-type and mutant polypeptides. EMBO J 8:301-308

McGrath MS, Hwang KM, Caldwell SE, Gaston I, Luk K-C, Wu P, Ng VL, Crowe S, Daniels J, Marsh J, Deinhart T, Lekas PV, Vennari JC, Yeung HW, Lifson JD (1989) GLQ223: an inhibitor of human immunodeficiency virus replication in acutely and chronically infected cells of lymphocyte and mononuclear phagocyte lineage. Proc Natl Acad Sci 86:2844-2848

Meinkoth J, Wahl G (1984) Hybridization of nucleic acids immobilized on solid supports. Anal Biochem 138:267-284

Messing J (1983) New M13 vectors for cloning. In: Wu R, Grossman L, Moldave K (eds) Methods in Enzymology. Volume 101. Recombinant DNA. Part C. Academic Press, Inc. London (Ltd), pp 20-78

Meyerhans A, Vartanian J-P, Wain-Hobson S (1990) DNA recombination during PCR. Nucleic Acids Res 18:1687-1691

McMaster GK, Carmichael GG (1977) Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc Natl Acad Sci 74:4835-4838

Mise T, Funatsu G, Ishiguro M, Funatsu M (1977) Isolation and characterisation of ricin E from castor beans. Agric Biol Chem 41:2041-2046

Mitich LW (1990) Bouncing Bett - the soap weed. Weed Technol 4:221-224

Moazed D, Robertson JM, Noller HF (1988) Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23 S RNA. Nature 334:362-364

Montanaro L, Sperti S, Stirpe F (1973) Inhibition by ricin of protein

synthesis *in vitro* ribosomes as the target of the toxin. Biochem J 136:677-683

Montecucchi P-C, Lazzarini AM, Barbieri L, Stirpe F, Soria M, Lappi D (1989) N-terminal sequences of some ribosome-inactivating proteins. Int J Peptide Protein Res 33:263-267

Montfort W, Villafranca JE, Monzingo AF, Ernst SR, Katzin B, Rutenber E, Xuong NH, Hamlin R, Robertus JD (1987) The three-dimensional structure of ricin at 2.8 A. J Biol Chem 262:5398-5403

Nicolson GL, Blaustein J (1972) The interaction of *Ricinus communis* agglutinins with normal and tumor cell surfaces. Biochim Biophys Acta 266:543-547

Nordmann PL, Makris CJ, Reznikoff WS (1988) Inosine induced mutations. Mol Gen Genet 214:62-67

O'Brien AD, Holmes RK (1987) Shiga and Shiga-like toxins. Microbiol Rev 51:206-220

Obrig JG, Irvin JD, Hardesty B (1973) The effect of an antiviral peptide on the ribosomal reactions of the peptide elongation enzymes, EF-I and EF-II. Arch Biochem Biophys 155:278-289

O'Hare M, Roberts LM, Thorpe PE, Watson GJ, Prior B, Lord JM (1987) Expression of ricin A chain in *E. coli*. FEBS Letts 216:73-78

Ohtsuka E, Matsuki S, Ikehara M, Takahashi Y, Matsubara K (1985) An alternative approach to deoxyoligonucleotides as hybridisation probes by insertion of deoxyinosine at ambiguous codon positions. J Biol Chem 260:2605-2608

Oka T, Natori Y, Tanaka S, Tsurugi K, Endo Y (1990) Complete nucleotide sequence of cDNA for the cytotoxin alpha sarcin. Nucleic Acids Res 18:1897

Osborn RW, Hartley MR (1990) Dual effects of the ricin A chain on protein synthesis in rabbit reticulocyte lysate. Eur J Biochem 193:401-407

Olsnes S, Sandvig K, Eiklid K, Pihl A (1978) Properties and action mechanism of the toxic lectin modeccin: Interaction with cell lines resistant to modeccin, abrin and ricin. J Supramolec Structure 9:15-25

Olsnes S, Pihl A, (1982) Toxic lectins and related proteins. In: Cohen P, van Heyningen S (eds) Molecular Actions of Toxins and Viruses. Elsevier Biomedical Press, Amsterdam, pp51-105

Olsnes S, Stirpe F, Sandvig K, Pihl A (1982) Isolation and characterisation of viscumin, a toxic lectin from *Viscum album* L. (mistletoe). J Biol Chem 257:13263-13270

Paleologue A, Reboud JP, Reboud A-M (1986) Modification of 60 S ribosomal subunits induced by the ricin A chain. FEBS Letts 208:373-378

Pappenheimer AM, Olsnes S, Harper AA (1974) Lectins from *Abrus precatorius* and *Ricinus communis*. 1. Immunochemical relations between toxins and agglutinins. J Immunol 113:835-841

Pastan I, Willingham MC, FitzGerald DJP (1986) Immunotoxins. Cell 47:641-648

Piatak M, Lane JA, Laird W, Bjorn MJ, Wamg A, Williams M (1988) Expression of soluble and fully functional ricin A chain in *Escherichia coli* is temperature sensitive. J Biol Chem 263:4837-4843

Ready MP, Adams RP, Robertus JD (1984a) Dodecandrin, a new ribosome -inhibiting from *Phytolacca dodecandra*. Biochim Biophys Acta 791:314-319

Ready MP, Wilson K, Piatak M, Robertus JD (1984b) Ricin-like plant toxins are evolutionarily related to single-chain ribosome-inhibiting proteins from *Phytolacca*. J Biol Chem 259:15252-15256

Ready MP, Brown DT, Robertus JD (1986) Extracellular localisation of pokeweed antiviral protein. Proc Natl Acad Sci 83:5053-5056

Ready MP, Katzin BJ, Robertus JD (1988) Ribosome-inhibiting proteins, retroviral reverse transcriptases and RNAse H share common structural elements. Proteins 3:53-59

Richardson PT, Gilmartin P, Colman A, Roberts LM, Lord JM (1988) Expression of functional ricin B chain in *Xenopus* oocytes. BioTechnol 6:565-570

Roberts WK, Stewart T (1979) Purification and properties of a translation inhibitor from wheat germ. Biochem 18:2615-2621

Roberts WK, Selitrennikoff CP (1986) Isolation and partial characterization of two antifungal proteins from barley. Biochim Biophys Acta 880:161-170

Roberts LM, Lamb FI, Pappin DJC, Lord JM (1985) The primary sequence of *Ricinus communis* Agglutinin. J Biol Chem 260:15682-15686

Robertus JD, Ready MP, (1984) Ricin B chain and discoidin I share a commom primitive fold. J Biol Chem 259:13953-13956

Robertus JD, Piatak M, Ferris R, Houston LL (1987) Crystallisation of ricin A chain obtained from a cloned gene in *E. coli*. J Biol Chem 262:19-20

Robinson NJ, Gupta A, Fordham-Skelton AP, Croy RRD, Whitton BA, Huckle JW (1990) Prokaryotic metallothionein gene characterization and expression: chromosome crawling by ligation-mediated PCR. Proc R Soc Lond B 242:241-247

Rutenber E, Ready M, Robertus JD (1987) Structure and evolution of ricin B chain. Nature 326:624-626

Rychlik W, Rhoads RE (1989) A computer programme for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. Nucleic Acids Res 17:8543-8551

Sacco G, Drickaner K, Wool IG (1983) The primary structure of the cytotoxin α-sarcin. J Biol Chem 258:5811-5818

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988a) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491

Saiki RK, Chang C-A, Levenson CH, Warren TC, Boehm MS, Kazazian HH, Erlich HA (1988b) Diagnosis of sickle cell anemia and β-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. New Eng J Medicine 319:537-541

Sakanari JA, Staunton CE, Eakin AE, Craik CS, McKerrow JH (1989) Serine proteases from nematode and protozoan parasite: Isolation of sequence homologs using generic molecular probes. Proc Natl Acad Sci 86:4863-4867

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463-5468

Sarkar G, Sommer S (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science 244:331-334

Saxena SK, O'Brien AD, Ackerman EJ (1989) Shiga toxin, Shiga-like toxin II variant and ricin are all single-site N-glycosidases of 28 S RNA when microinjected into *Xenopus* oocytes. J Biol Chem 264:596-601

Scharf SJ, Horn GT, Erlich HA (1986) Direct cloning and amplification of enzymatically amplified genomic sequences. Science 233:1076-1078

Schlossman D, Withers D, Welsh P, Alexander A, Robertus J, Frankel A

(1989) Role of glutamic acid 177 of ricin toxin A chain in enzymatic inactivation of ribosomes. Mol Cell Biol 9:5012-5021

Seed B, Parker RC, Davidson N (1982) Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. Gene 19:201-209

Shinshi H, Wenzler H, Neuhaus J-M, Felix G, Hofsteenge JF, Meins F Jr (1988) Evidence for N- and C- terminal processing of a plant defense -related enzyme: Primary structure of tobacco prepro-β-1,3-glucanase. Proc Natl Acad Sci 85:5541-5545

Shuldiner AR, Nirula A, Roth J (1989) Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res 17:4409

Simmons BM, Stahl PD, Russell JH (1986) Mannose receptor-mediated uptake of ricin toxin and ricin A chain by macrophages. J Biol Chem 261:7912-7920

Sommer R, Tautz D (1989) Minimal homology requirements for PCR primers. Nucleic Acids Res 17:6749

Southern E (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517

Sperti S, Montanaro L, Mattioli A, Stirpe F (1973) Inhibition by ricin of protein synthesis *in vitro*: 60 S ribosomal subunit as the target of the toxin. Biochem J 136:813-815

Stevens WA, Spurdon C, Onyon LJ, Stirpe F (1981) Effect of inhibitors of protein synthesis on tobacco mosaic virus infection. Experientia 37:257-259

Stirpe F, Olsnes S, Pihl A (1980) Gelonin, a new inhibitor of protein synthesis non-toxic to intact cells. Isolation, characterization and preparation of cytotoxic conjugates with concanavilin A. J Biol Chem 255:6947-6955

Stirpe F, Williams DG, Onyon LJ, Legg RF, Stephens WA (1981) Dianthins, ribosome-damaging proteins with antiviral properties from *Dianthus caryophyllus* (carnation). Biochem J 195:339-405

Stirpe F (1982) On the action of ribosome-inactivating proteins: are plant ribosomes species-specific? Biochem J 202:279-280

Stirpe F, Gasperi-Campani A, Barbieri L, Falasca A, Abbondanza A, Stevens WA (1983) Ribosome-inactivating proteins from the seeds of *Saponaria officinalis* L. (soapwort), of *Agrostemma githago* L. (corncockle) and of *Asparagus officinalis* L. (asparagus), and from the latex of *Hura crepitans* L. (sandbox tree). Biochem J 216:617-625

Stirpe F, Barbieri L (1986) Ribosome-inactivating proteins up to date. FEBS Letts 195:1-8

Stirpe F, Bailey S, Miller SP, Bodley JW (1988) Modification of ribosomal RNA by ribosome-inactivating proteins from plants. Nucleic Acids Res 16:1349-1357

Strockbrine NA, Jackson MP, Sung LM, Holmes RK, Neuland JW (1988) Cloning and sequencing of the genes for Shiga toxin from *Shigella dysenteriae* type 1. J Bacteriol 170:1116-1122

Sundan A, Evensen G, Hornes E, Mathiesen A (1989) Isolation and *in vitro* expression of the ricin A-chain gene: effect of deletions on biological activity. Nucleic Acids Res 17:1717-32

Takahashi Y, Kato K, Hayashizaki Y, Wakabayashi T, Ohtsuka E, Matsuki S, Ikehara M, Matsubara K (1985) Molecular cloning of the human cholecystokinin gene by use of a synthetic probe containing deoxyinosine. Proc Natl Acad Sci 82:1931-1935

Taylor BE, Irvin JD (1990) Depurination of plant ribosomes by pokeweed antiviral protein. FEBS Letts 273:144-146

Terao K, Uchiumi T, Endo Y, Ogata K (1988) Ricin and α-sarcin alter the conformation of 60 S ribosomal subunits at neighboring but different sites. Eur J Biochem 174:459-463

Ussery MA, Irvin JD, Hardesty B (1977) Inhibition of poliovirus replication by a plant antiviral peptide. Ann New York Acad Sci 284:431-440

Queen C, Korn LJ (1984) A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res 12:581-599

Uchiumi T, Kikuchi M, Terao K, Ogata K (1985) Cross-linking study on protein topography of rat liver 60 S ribosomal subunits with 2-iminothiolane. J Biol Chem 260:5675-5682

Uchiumi T, Kikuchi M, Terao K, Ogata K (1986) Cross-linking study on protein neighborhoods at the subunit interface of rat liver ribosomes with 2-iminothiolane. J Biol Chem 261:9663-9667

Van den Bulcke M, Bauw G, Castresana C, Van Montagu M, Vandekerckhove J (1989) Characterization of vacuolar and extracellular β(1,3) -glucanases of tobacco: Evidence for a strictly compartmentalized plant defense system. Proc Natl Acad Sci 86:2673-2677

van Deurs B, Tonnessen TI, Petersen OW, Sandvig K, Olsnes S (1986) Routing of internalised ricin and ricin conjugates to the Golgi complex. J Cell Biol 102:37-47

Vittetta ES (1986) Synergy between immunotoxins prepared from native ricin A chains and chemically modified ricin B chains. J Immunol 136:1880-1887

Vitetta ES, Fulton RJ, May RD, Till M, Uhr JW (1987) Redesigning natures poisons to create anti-tumor drugs. Science 238:1098-1104

von Heijne G (1984) How signal sequences maintain cleavage specificity. J Mol Biol 173:243-251

Waugh R and Brown JWS (1991) Plant gene structure and expression. In: Grierson D (ed) Plant Genetic Engineering, vol1. Blackie and Son Ltd, London, pp 1-37

Yarwood JN, Harris N, Delauney A, Croy RRD, Gatehouse JA, Watson MD, Boulter D (1987) Construction of a hybrid cDNA encoding a major legumin precursor and its expression and localisation in *Saccharomyces cerevisiae*. FEBS Letts 222:175-180

Yeung HW, Li WW, Feng Z, Barbieri L, Stirpe F (1988) Trichosanthin, α-momorcharin and β-momorcharin: identity of abortifacient and ribosome-inactivating proteins. Int J Peptide Protein Res 31:265-268

Zhang X, Wang J (1986) Homology of trichosanthin and ricin A chain. Nature 321:477-478

Mol Gen Genet (1990) 221:134–138

MGG

Synthesis of saporin gene probes from partial protein sequence data: Use of inosine-oligonucleotides, genomic DNA and the polymerase chain reaction

Anthony P. Fordham-Skelton, Alan Yarwood, and Ronald R.D. Croy
Plant Molecular Biology Group, Department of Biological Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, UK

Summary. A strategy employing the polymerase chain reaction to synthesize gene-specific probes suitable for genomic Southern analyses and for screening genomic libraries is described. The method utilizes partial amino acid sequence data from the protein of interest, genomic DNA and inosine-containing oligonucleotide primers. An example of its application for the isolation of plant gene sequences encoding saporin, a ribosome inactivating protein, is described.

Key words: Saporin/RIPs – Genes – Inosine – Oligonucleotides – PCR DNA probes

Introduction

Ribosome Inactivating Proteins (RIPs) are widely distributed throughout the plant kingdom and have been classified as either type 1 or type 2 depending upon their subunit composition (Barbieri and Stirpe 1982; Stirpe and Barbieri 1986). Type 1 RIPs including momordin (Barbieri et al. 1980a), dianthin (Stirpe et al. 1981) and gelonin (Stirpe et al. 1980) possess a single active polypeptide chain. Type 2 RIPs, such as modeccin (Barbieri et al. 1980b), ricin and abrin (Olsnes and Pihl 1976) consist of two disulphide-bonded subunits: an enzymatically active A chain, functionally homologous to type 1 RIP polypeptides, and a sugar binding, lectin-like, B chain.

Ricin possesses an RNA N-glycosidase activity and catalytically inactivates ribosomes by removing a single adenine residue from a conserved region of eucaryotic 28S ribosomal RNAs (Endo et al. 1987; Endo and Tsurugi 1987). It is likely that all plant RIPs act in this manner (Stirpe et al. 1988). To date the only RIP gene analysed is ricin, and both cDNA and genomic sequences have been determined (Lamb et al. 1985; Halling et al. 1985). *Saponaria officinalis* (soapwort) produces several immunologically related type 1 RIPs (saporins) found mainly in seed tissue but also in the leaves (Stirpe et al. 1983). The N-terminal of one saporin protein, SO-6, has been partially sequenced (Lappi et al. 1985).

Offprint requests to: A. Fordham-Skelton

As a first step in the analysis of the saporin proteins at the nucleic acid level we have used the polymerase chain reaction (PCR) to obtain probes for the saporin genes, by amplifying a fragment of the saporin coding sequence. PCR has rapidly become established as a powerful technique for gene analysis (Saiki et al. 1988). Originally used with single primer species, it is now possible to obtain specific DNA amplification using redundant primers obtained via protein sequence data. Fragments of the genes of porcine urate oxidase and, from several species, the iron-sulfur protein of succinate dehydrogenase have been successfully amplified from cDNA and genomic DNA using redundant primers (Lee et al. 1988; Gould et al. 1989).

Conventional cloning strategies used to isolate genes expressed in a temporal or tissue specific manner normally necessitate the production and differential screening of cDNA libraries to isolate full-length or partial cDNAs. Such strategies are both time consuming and labour intensive. When amino acid sequence data are available, gene-specific oligonucleotide probes can be synthesized which encode parts of the protein of interest. However the direct use of such probes for genomic library screening has had variable success depending on the degree of redundancy incurred in the reverse translation from peptide to DNA. We describe in this paper a strategy, based on the PCR, to obtain a sizeable fragment of the saporin gene which can be used to screen a genomic library for the corresponding genomic clones. The technique has wide general application for the rapid isolation of gene sequences encoding products for which only partial protein sequences are available.

Materials and methods

Materials. Saponaria officinalis plants were grown outdoors in pots of Levington compost and young leaf material was harvested for genomic DNA isolation. Restriction enzymes and T4 DNA ligase were supplied by NBL, *Taq* polymerase from *Thermus aquaticus* was from Stratagene or Perkin-Elmer/Cetus. [α-^{32}P]dCTP (14.8 TBq/mmol) and nylon (Hybond N) filters were from Amersham International. *Escherichia coli* DH5α competent cells were from GIBCO-

BRL. Anion exchange resin (Qiagen) tips for plasmid DNA purification were supplied by Diagen (FRG). Phosphoramidite derivatives of all nucleotide bases, including inosine, were obtained from Applied Biosystems.

Protein sequencing. Purified saporin-6 protein, prepared as described by Stirpe et al. (1983), was a kind gift from Professor F. Stirpe and Dr. L. Barbieri, University of Bologna. Samples (10–20 mg) of the protein were reduced, carboxymethylated and digested separately with trypsin or pepsin. The peptide mixtures were fractionated on Biogel P-6 in 0.05 M ammonium bicarbonate followed by purification of peak fractions by reverse phase HPLC on a Vydac C_{18} column (0.5 × 25 cm) using a Varian Model 5000 HPLC system and a linear gradient of 0%–70% acetonitrile in 0.1% trifluoroacetic acid. The intact protein and peptides derived from it were subjected to liquid phase manual microsequence analysis using the 4-N-N-dimethylamino-azobenzene-4 -isothiocyanate (DABITC)/phenylisothiocyanate (PITC) double coupling method as described by Yarwood (1989). Amino acid analysis of peptides was carried out using a Waters "Picotag" system (PICOTAG Manual No. 86746, Waters/Millipore, USA).

Genomic DNA preparation. Genomic DNA was prepared from young leaves of *S. officinalis* plants essentially as described by Graham (1978) and purified by two rounds of centrifugation through ethidium bromide-CsCl density gradients.

Oligonucleotide synthesis. Inosine-containing oligonucleotides were synthesized using an Applied Biosystems 381A DNA synthesizer operated with a standard synthesis program. After cleavage and deprotection the oligonucleotides were dried down under vacuum, twice resuspended in water and vacuum dried. Oligonucleotides were stored at −20° C either dry or as an aqueous solution and were used without further purification. The concentrations of the oligonucleotides were determined by UV absorbance at 260 nm.

In vitro DNA amplification. PCR was carried out essentially as described by Saiki et al. (1988). Due to primer redundancy a higher concentration of each primer was used than that suggested in the Cetus protocol (Girgis et al. 1988). Conditions for the reactions were as follows: 20 µg of each primer, 2.5 µg genomic DNA, 200 µM each of dATP, dTTP, dGTP and dCTP in a final volume of 100 µl. Reaction buffer contained 50 mM KCl, 10 mM TRIS-HCl, pH 8.3, 1.5 mM $MgCl_2$ and 0.01% (w/v) gelatin. Reaction mixtures were heated to 94° C for 4 min to denature the genomic DNA. After the addition of *Taq* polymerase (5 units per reaction) the reaction mixture was overlaid with 100 µl of paraffin oil. Reactions were cycled automatically through time/temperature cycles as follows: denaturation 91° C/2 min, annealing 45° C/2 min, and extension 72° C/3 min using a Hybaid Intelligent Heating Block repeated for a total of 30 cycles. The final extension time was increased to 15 min to ensure completion of strand synthesis. Samples were stored at 4° C prior to further analysis.

Analysis of PCR products. Ten microliters of reaction mixtures were analysed by electrophoresis on 2% agarose gels stained with ethidium bromide. Discrete PCR-amplified sequences were isolated from low melting point agarose gels, ligated directly into pUC18 restricted with *Hin*cII and used to transform *E. coli* DH5α competent cells to ampicillin resistance. Clones containing plasmids with inserts were isolated and plasmid DNA prepared using Qiagen tips according to the manufacturers recommendations.

DNA sequence analysis. Plasmid sequencing was performed by the dideoxy-sequencing method of Sanger et al. (1977) using fluorescent dye-linked universal M13 primers and analysed using an Applied Biosystems 370A DNA sequencer. Plasmids were sequenced in both directions using forward and reverse primers, as described in the supplier's protocol (Model 370A DNA Sequencing System, Users Manual Version 1.3A, Oct. 1988, pp 3.22–3.25) using Sequenase enzyme with a Klenow (DNA polymerase I) chase.

α-^{32}P labelling of PCR products by re-amplification of PCR products. Amplified DNA was used as a probe for genomic Southern blots and gene library screening by re-amplifying the product in the presence of [α-^{32}P]dCTP. Approximately 10 ng of the 300 bp initial amplification product was purified from an agarose gel following electrophoresis and subjected to further amplification. The reaction conditions were as described above, except that the number of reactin cycles was reduced to 8 and 100 µCi of [α-^{32}P]dCTP was substituted for the dCTP normally used. After completion, the reaction mixture was diluted to 300 µl with column buffer (300 mM NaCl, 50 mM TRIS-HCl, pH 7.5, 0.1% SDS w/v) and un-incorporated dCTP removed by separation on a 5 ml Sephadex G-50 column. The approximate specific activity obtained was 7×10^8 cpm/µg DNA. To check the integrity of product(s) labelled by the amplification reaction, a 1 µl aliquot of the G-50 eluate was diluted to 20 µl and electrophoresed on a 2% agarose minigel. Following electrophoresis the gel was dried and auto-radiographed.

Restriction and Southern analysis of genomic DNA, and screening of a Saponaria *genomic library.* Ten micrograms of genomic DNA were used for each restriction using incubation conditions as recommended by the manufacturers. Restriction fragments were separated on a 0.8% agarose gel, denatured and neutralized essentially as described by Maniatis et al. (1982) and the DNA was transferred to a nylon filter using a Hybaid vacuum blotting apparatus. Standard prehybridization and hybridization conditions were used (Maniatis et al. 1982) and the probe, obtained by re-amplification, was used in the same manner as those produced by nick-translation or random priming labelling protocols. The filter was washed to a final stringency of 0.1 × SSC (1 × SSC is 0.15 M NaCl, 15 mM sodium citrate) at 65° C for 15 min and autoradiographed at −80° C using a pre-flashed X-ray film and intensifying screen.

A *Saponaria* gene library was constructed in lambda EMBL3 using *Sau*3A partially digested genomic DNA essentially as described by Frischauf et al. (1983). The library was screened using standard plaque lift and hybridization methods with ^{32}P-labelled probe.

Results and discussion

Protein sequencing of saporin-6 resulted in the determination of the first 48 N-terminal amino acids of the mature protein. This confirmed the homogeneity of the protein

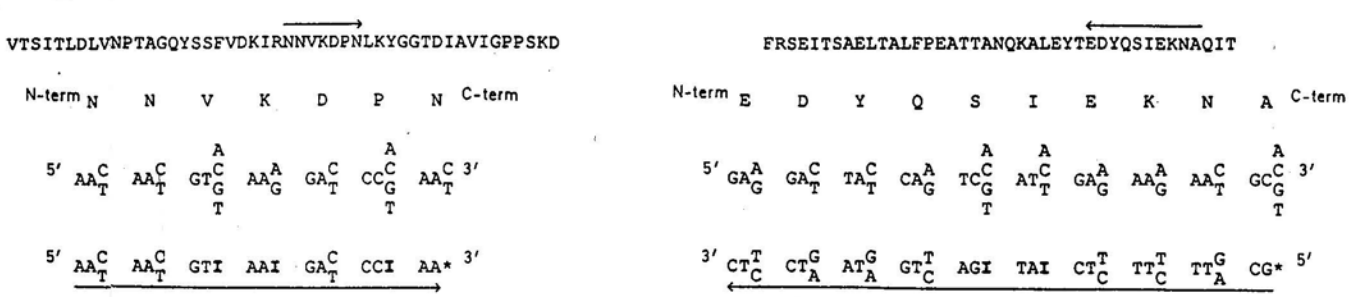


Fig. 1. Partial amino acid sequence data of the mature saporin-6 protein and deduced primer sequences used in subsequent DNA amplifications. The *arrows* indicate primer orientation. The *asterisk* indicates the third base omitted from the primer sequence

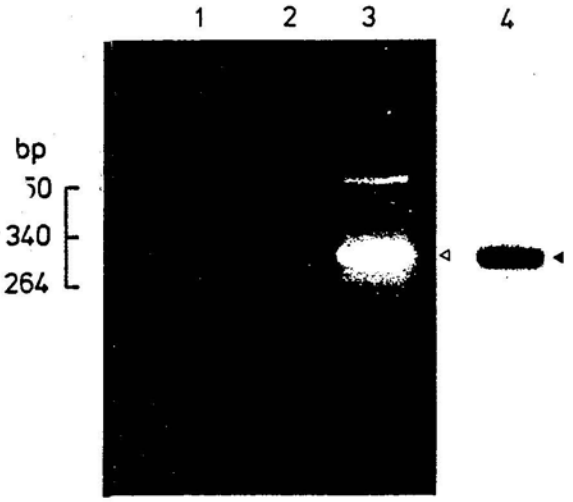


Fig. 2. Visualization of amplified saporin gene fragment on an agarose gel. Amplification conditions using genomic DNA were as described in the Materials and methods. Lane 1, λ *Pst*1 size markers; lane 2, control reaction mix without *Taq* polymerase; lane 3, reaction mix with *Taq* polymerase; lane 4, re-amplification of 300 bp polymerase chain reaction (PCR) product with ^{32}P, for use as gene probe (indicated by *arrows*)

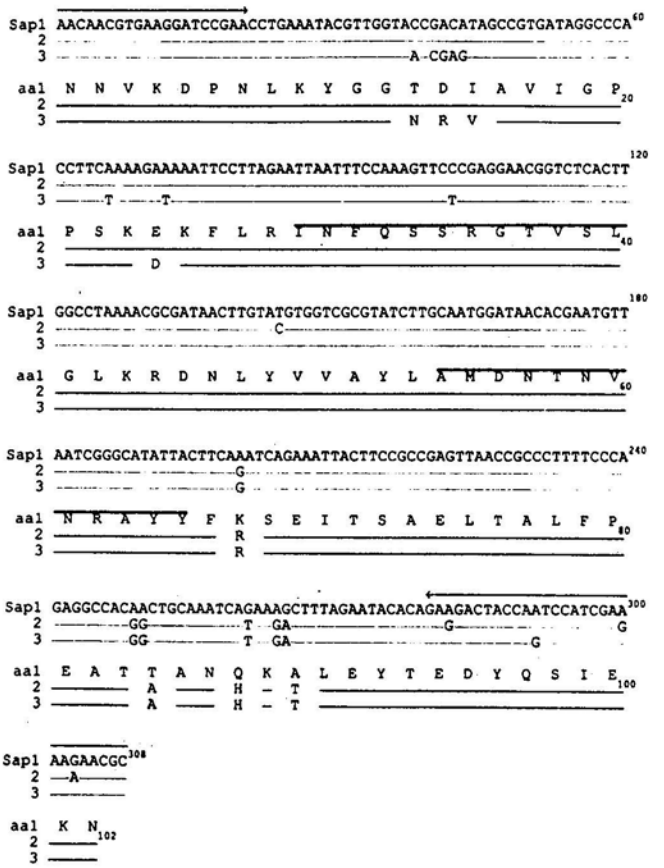


Fig. 3. Nucleotide and deduced amino acid sequences of three plasmids containing the cloned PCR product(s). Only non-identical amino acids are given for Sap 2 and 3. Previously unaligned peptide sequences are *overlined* (see text); the *arrows* indicate primer orientation

preparation used and extended the previously reported N-terminal sequence data of Lappi et al. (1985). Tryptic and peptide peptides provided further sequence data including an overlapping sequence, of 40 amino acids in length, at an undetermined position in the protein (Fig. 1). The computer program PROBFIND (Lewis 1986) was used to reverse translate the peptide sequences to DNA sequences and to select oligonucleotides of minimal redundancy. Inosine-containing oligonucleotide primers were synthesized, corresponding to the two peptide regions shown in Fig. 1; one, the sense strand, was from the N-terminal sequence, and the other, the anti-sense strand, was from the internal peptide sequence. Inosines were incorporated at base positions with three- to fourfold redundancy and at A/G base choices, but were not incorporated at C/T choices due to the higher probability of I/G base pairing (Ohtsuka et al. 985).

The analysis of the products of a PCR reaction using se primers on *Saponaria* genomic DNA is shown in g. 2. The arrowed band of 300 bp (lane 3) was the major duct in all amplification reactions. Under the conditions d a number of minor bands were detectable but these were not analysed further. Lane 4 of Fig. 2 shows an autoradiograph of the products of re-amplification of the isolated 300 bp product in the presence of [α-^{32}P]dCTP, such as the probe used for Southern blots or gene library screening; the 300 bp fragment was the only labelled species detectable.

The nucleotide and deduced amino acid sequences of three clones containing the 300 bp amplification product are shown in Fig. 3. Two of the sequences are virtually identical and highly homologous to the third. Positions in the primer sequence originally containing inosine were sub-

sequently changed to guanosine when the sequences were cloned. This is consistent with the results of Ohtsuka et al. (1985) and Nordmann et al. (1988) which showed that inosine will primarily select cytosine when present in a template for DNA synthesis.

Each sequence contains an open reading frame encoding 102 amino acids. Significantly, two previously unaligned saporin peptide sequences can be matched within the deduced amino acid sequence between the primers. This confirms that the PCR has specifically amplified part of the saporin gene. Several of the amino acid differences between the cloned sequences are conservative and Sap 3 has a region of 3 amino acids differing from the other two sequences. Evidence from the tryptic peptide sequencing indicates a corresponding heterogeneity in at least 4 of the positions in which differences were predicted by the nucleotide sequences.

A genomic Southern blot (Fig. 4) probed with the labelled, amplified product revealed several hybridizing bands in each digest under conditions that indicated sequences of low copy number. Together with the evidence for heterogeneity in the sequences and the report of at least 9 isolated saporin protein fractions (Stirpe et al. 1983), it is suggested that saporin-6 is a member of a small multigene family. The results also indicated that the amplified product might be used for the isolation of saporin genes. Thus a lambda genomic library was screened using the labelled, amplified sequences; positive clones were isolated and purified, and their identity confirmed by sequencing. The details of the structure and sequence of these saporin genes will be reported elsewhere (manuscript in preparation).

The use of the base analogue inosine has greatly facilitated this experimental strategy by significantly reducing the base redundancy in the oligonucleotide primers and thereby allowing the use of long peptide regions that would otherwise require complex oligonucleotide mixtures containing more than 500 sequences. The possibility should

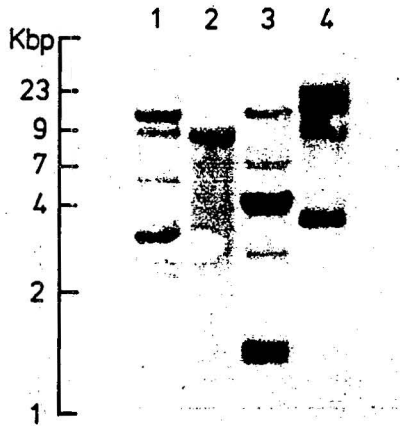


Fig. 4. *Saponaria* genomic Southern blot hybridized with ^{32}P-labelled probe. DNA was digested to completion with the following restriction enzymes: lane 1, *Eco*RI; lane 2, *Bam*HI; lane 3, *Hin*dIII; lane 4, *Bgl*II

be considered of errors arising in the PCR products due to preferential base pairing with inosine in the primers, base misincorporation by the *Taq* polymerase, or the production of hybrid genes when amplifying closely related sequences. Shuldiner et al. (1989) observed the latter effect when amplifying *Xenopus* preproinsulin genes I and II. However, in the present strategy none of these artefacts should detract from the utility of this approach to generate specific gene probes. Thus, the nucleotide sequencing confirms probe identity rather than indicating an authentic gene sequence. The length of the sequence amplified is dependent on the distance between the chosen peptide sequences and in the case of partial sequence data is purely fortuitous. The use of an N-terminal sequence and some other peptide maximizes the chance of obtaining a usefully long sequence.

In conclusion, we have demonstrated in this paper the utility of a strategy for producing gene probes suitable for screening genomic libraries, which is remarkably simple and requires only partial protein sequence data. Such a strategy for gene isolation eliminates the requirement for a spatially or temporally expressed cDNA library with its associated problems and is much easier and more versatile than direct oligonucleotide screening. Although the example used here to demonstrate the feasibility of the approach is a plant gene, the method is universal and adaptable to any system where partial protein sequence data are available.

Acknowledgements. We wish to thank Professor D. Boulter for laboratory facilities; E. Croy for *Saponaria* DNA; Dr. P. Taylor for the gift of the *Saponaria* gene library and for useful contributions to the preliminary work; J. Bryden and Dr. J.A. Gatehouse for DNA sequencing; and the University of Durham for the provision of oligonucleotide synthesis facilities. A.F-S. is financed by an SERC postgraduate studentship. This work was partly financed under the EEC Biotechnology Action Programme, contract No. BAP-0104-UK.

References

Barbieri L, Stirpe F (1982) Ribosome-inactivating proteins from plants: Properties and possible uses. Cancer Surv 1:489–520

Barbieri L, Zamboni M, Montanaro L, Sperti S, Stirpe F (1980a) Inhibition of protein synthesis in vitro by proteins from the seeds of *Momordica charantia* (bitter pear melon). Biochem J 186:443–452

Barbieri L, Zamboni M, Montanaro L, Sperti S, Stirpe F (1980b) Purification and properties of different forms of modeccin, the toxin of *Adenia digitata*. Separation of subunits with inhibitory and lectin activity. Biochem J 185:203–210

Endo Y, Tsurugi K (1987) RNA N-glycosidase activity of ricin A-chain. J Biol Chem 262:8128–8130

Endo Y, Mitsui K, Motizuki M, Tsurugi K (1987) The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. J Biol Chem 262:8128–8130

Frischauf AM, Lehrach H, Poutska A, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Girgis SI, Alevizaki M, Denny P, Ferrier GJM, Legon S (1988) Generation of DNA probes for peptides with highly degenerate codons using mixed primer PCR. Nucleic Acids Res 26:10371

Gould SJ, Subramani S, Schefler IE (1989) Use of the DNA polymerase chain reaction for homology probing: Isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. Proc Natl Acad Sci USA 86:1934–1938

Graham DE (1978) The isolation of high molecular weight DNA from whole organisms or large tissue masses. Anal Biochem 85:609–613

Halling KC, Halling AC, Murray EE, Ladin BF, Houston LL,

Weaver RF (1985) Genomic cloning and characterization of a ricin gene from *Ricinus communis*. Nucleic Acids Res 13:8019–8033

Lamb FI, Roberts LM, Lord JM (1985) Nucleotide sequence of cloned cDNA coding for preproricin. Eur J Biochem 148:265–270

Lappi DA, Esch FS, Barbieri L, Stirpe F, Soria M (1985) Characterization of a *Saponaria officinalis* seed ribosome-inactivating protein: immunoreactivity and sequence homologies. Biochem Biophys Res Commun 129:934–942

Lee CC, Wu X, Gibbs RA, Cook RG, Muzny DM, Caskey TC (1988) Generation of cDNA probes directed by amino acid sequence: Cloning of urate oxidase. Science 239:1288–1291

Lewis RM (1986) Probfind: a computer program for selecting oligonucleotides from peptide sequences. Nucleic Acids Res 14:567–570

Maniatis T, Fritschf EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Nordmann PL, Makris CJ, Reznikoff WS (1988) Inosine induced mutations. Mol Gen Genet 214:62–67

Ohtsuka E, Matsuki S, Ikehara M, Takahashi Y, Matsubara K (1985) An alternative approach to deoxyoligonucleotides as hybridisation probes by insertion of deoxyinosine at ambiguous codon positions. J Biol Chem 260:2605–2608

Olsnes S, Pihl A (1976) Abrin, ricin and their associated agglutinins. The specificity of action of animal, bacterial and plant toxins. In: Cuatrecasas P (ed) Receptors and recognition, Series B, vol 1. Chapman and Hall, London, pp 129–173

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5468

Shuldiner AR, Nirula A, Roth J (1989) Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res 17:4409

Stirpe F, Barbieri L (1986) Ribosome-inactivating proteins up to date. FEBS Lett 195:1–8

Stirpe F, Olsnes S, Pihl A (1980) Gelonin, a new inhibitor of protein synthesis, non toxic to intact cells. Isolation, characterization and preparation of cytotoxic conjugates with concanavalin A. J Biol Chem 255:6947–6955

Stirpe F, Williams DG, Onyon LJ, Legg RF, Stevens WA (1981) Dianthins, ribosome-damaging proteins with antiviral properties from *Dianthus caryophyllus* L. (carnation). Biochem J 195:399–405

Stirpe F, Gasperi-Campani A, Barbieri L, Falasca A, Abbondanza A, Stevens WA (1983) Ribosome-inactivating proteins from the seeds of *Saponaria officinalis* L. (soapwort), of *Agrostemma githago* L. (corncockle) and of *Asparagus officinalis* L. (asparagus), and from the latex of *Hura crepitans* L. (sandbox tree). Biochem J 216:617–625

Stirpe F, Bailey S, Miller SP, Bodley JW (1988) Modification of ribosomal RNA by ribosome-inactivating proteins from plants. Nucleic Acids Res 16:1349–1357

Yarwood A (1989) Manual methods of protein sequencing. In: Findlay JBC, Geisow MJ (eds) Protein sequencing. Practical Approach Series. IRL Press at Oxford University Press, Oxford, pp 119–145

Communicated by H. Saedler

Received September 11, 1989