Molecular cloning and characterisation of the gene encoding the *Ascaris* allergen ABA-1

A thesis submitted for
the degree of
Doctor of Philosophy
at the University of Glasgow

by

Heather Jane Spence

Department of Zoology,
University of Glasgow

April 1994

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Dedicated to the memory of my wee sister Sharon.

The research reported in this thesis is my own and original work except where otherwise stated and has not been submitted for any other degree.

CONTENTS

ABBREVIATIONS

ATP	adenosine triphosphate
CTP	cytidine triphosphate
GTP	guanosine triphosphate
TTP	thymidine triphosphate
DNA	2' deoxyribonucleic acid
dNTP	2' deoxy (nucleotide)
EDTA	ethylene diamine tetra-acetic acid (disodium salt)
IPTG	isopropyl B-D-thiogalactoside
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
Tris	tris (hydroxymethyl) amino ethane
dH_2O	deionised water
bp	base pair
kbp	kilobase pair (10^3 bp)
kDa	kilodalton (10^3 dalton)
mM	millimolar
µM	micromolar
mg	milligram
µg	microgram
min	minute
hr	hour
UV	ultra violet light
DNAase	deoxyribonuclease
RNAase	ribonuclease
Ig	immunoglobulin

ACKNOWLEDGMENTS

I would firstly like to thank Malcolm for his help and encouragement throughout this thesis. A special thanks goes to the Post-Docs for all their support and advice - Collette, Joyce, Philippe and Karen - and to David, Colin, Anne and Fiona for their patient technical help. Thanks to Peter Rikus for the photography presented in this thesis. Thanks also to the WUMPS for their advice and for use of their equipment.

I would like to thank everyone at the WLEP for their friendship throughout the past three years, and also for dragging me up Ben Lomond, Petanque competitions, barbeques and nights in the pub - Lisa, Andrew, Kathleen, Stephen, Jacqui, Liz, Anna, Amanda, Yu and all the people above. Thanks to Mum, Dad and Craig for always being there when I needed them.

The biggest thanks goes to Richard for putting up with me over the past three years, sharing numerous bottles of wine and for Monday night Tango lessons.

SUMMARY

A 14 kDa protein, named ABA-1, is the most abundant protein in the body fluid of adult *Ascaris*, and is released by the tissue-penetrating larval stages of the parasite during culture *in vitro*. Immunological work has shown that ABA-1 is an allergen, suggesting that the IgE-mediated hypersensitivity response to ABA-1 seen during *Ascaris* infection may contribute to the pathology of ascariasis. In order to further characterise ABA-1, a cDNA expression library was constructed using mRNA prepared from *A. suum* infective larvae. Screening with polyclonal rabbit antiserum raised to gel-excised ABA-1 protein led to the isolation of a cDNA clone with a 1255 bp insert. DNA sequencing of the entire insert revealed that it consists of two 399 bp repeats and one truncated repeat. The putative amino acid sequence derived from each of the repeats was identical to the partial N-terminal amino acid sequence of the native ABA-1 protein, which had previously been derived by direct peptide sequencing. These data can be taken to indicate that ABA-1 is translated as a polyprotein which is then processed into 14 kDa monomers.

To allow characterisation of the genomic organisation of the *aba-1* gene, two genomic *aba-1* clones were isolated from an *A.suum* genomic library. Characterisation of these clones revealed that they both contain part of the *aba-1* genomic gene and suggested that there are at least 15 repeats of 399 bp in the *aba-1* gene; in addition, there appears to be a 4 kbp intron situated at the 3' end of the gene.

Northern and Western blots were used to determine stage- and tissue-specificity of the expression of the *aba-1* gene. These data revealed that ABA-1 is expressed constitutively throughout the life of the parasite

and that it would appear to be expressed in all tissues of the adult worm, although it is only found as a polyprotein in the gut.

The gene encoding the *A. lumbricoides* ABA-1 protein was cloned and characterised by the use of the Polymerase Chain Reaction. This revealed that, like the *A. suum aba-1* gene, it is also consists of tandem repeats of approximately 400 bp, suggesting that it to is expressed initially as a polyprotein and then cleaved into 14 kDa monomers. Western blot analysis of the the native ABA-1 protein confirmed this suggestion.

Chapter 1

Introduction

1.1 Introduction

Ascaris lumbricoides, a parasitic nematode which infects humans, was initially named *lumbricus teres* by Tyson (1683), who referred to it as "that common roundworm which children usually are troubled with". Although three hundred years have past since this first description, the worm is still prevalent, with an estimated 1008 million people world-wide being infected (Crompton *et al.*, 1988; W.H.O, 1981). A further measure of the continuing prevalence of this parasite is that within the past 15 years, 153 out of 218 recognised states and countries reported ascariasis, the disease caused by *Ascaris* infection (Crompton, 1989), making it the most widespread nematode disease of humans. Although the disease is not fatal to everyone, *A. lumbricoides* infection can cause lethal pulmonary hypersensitivity responses in some humans (Oglive and de Savingny, 1982; Coles *et al.*, 1985) and can result in nutritional complications such as reduction in the ability to digest lactose and the ability to absorb vitamin A (Crompton, 1985; Carrara *et al.*, 1984; Mahalanabis *et al.*, 1979). Over the past 100 years, many researchers have been studying the immune responses of humans infected with *A. lumbricoides*, with the ultimate aim of elucidating a means of eradicating the parasite. At present the only way of treating *Ascaris* infection is by using anthelmintic drugs such as pyrantel pamoate, which causes the expulsion of adult worms from the host's gut (Crompton *et al.*, 1989). Unless such chemotherapy is continued throughout the lifespan of humans, reinfection will continually occur, implying that humans do not acquire a natural immunity to *Ascaris* infection (Elkins *et al.*, 1986).

This chapter reviews the following aspects of *Ascaris*: its life cycle; a comparison of the species *A. suum* and *A. lumbricoides;* immunological work performed to analyse the host's response to *Ascaris* infection; and

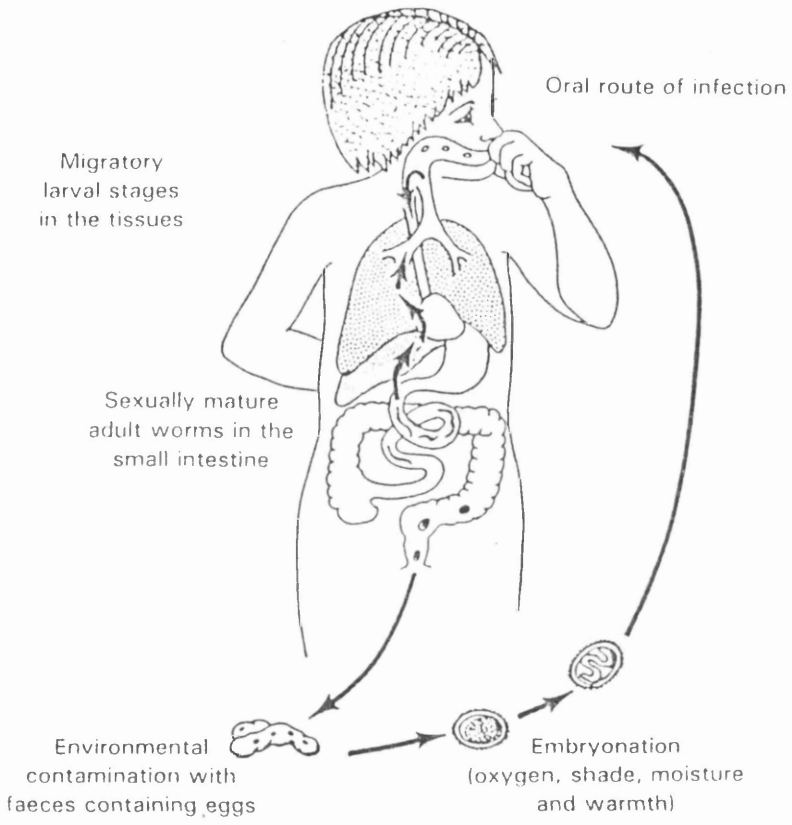
immunological work performed on the *A. suum* and *A. lumbricoides* ABA-1 proteins and their homologues in other parasitic nematodes.

1.2 The life cycle of *A. lumbricoides*

A diagrammatic representation of the cycle of *A. lumbricoides* infection in humans is shown in Figure 1.1; the information was abstracted from reviews by Pawlowski (1978, 1982), Pawlowski and Arfaa (1984), Janssens (1985), Stephenson (1987) and Crompton (1989). Infection in humans occurs by the uptake of embryonated eggs containing the second larval stage of the parasite, termed the L2 stage. This uptake results from the consumption of contaminated soil, water or food, or by inhalation of the eggs (W.H.O., 1967; Kagei, 1983). Once ingested, the embryonated eggs hatch, releasing L2 larvae into the jejunum of the host (Crompton and Pawlowski, 1985). Hatching of the eggs has been shown to be a consequence of a combination of the following conditions: elevation of the ambient temperature to 37°C, acidic pH, the correct carbonate and bicarbonate concentrations, and reducing conditions (Rodger, 1960). Newly hatched larvae in the host's jejunum penetrate the mucosa of the small intestine and migrate to the liver and then to the lungs *via* either or both of the lymphatic system or veins. It is not known how penetration of the intestinal mucosal is achieved by the larvae, but it may be envisaged that they excrete proteinases which digest the cell wall of the host's gut.

During their migration to the host's lungs, the *A. lumbricoides* larvae undergo a moult to the L3 larval stage. Upon arrival at the lungs, another moult creates the final larval stage of the nematode, the L4 stage (Watson, 1965; reviewed by Lee, 1966 and 1972). It is whilst present in the lungs that *Ascaris* can cause lethal pulmonary hypersensitivity responses

Figure 1.1. Diagrammatic representation of the life cycle of *A. lumbricoides*. Based on Crompton (1984).


Oral route of infection
Migratory larval stages in the tissues
Sexually mature adult worms in the small intestine
Environmental contamination with faeces containing eggs
Embryonation (oxygen, shade, moisture and warmth)

in their hosts (see Section 1.4.3). The exact parasitic proteins which elicit this allergic response are not known, but molecules have been isolated and characterised which can themselves generate IgE responses in the host and may therefore be central to this feature of ascariasis (reviewed by Almond and Parkhouse, 1985).

After the above development in the lungs, the L4 *Ascaris* larvae return to the small intestine of the host *via* the bronchi, trachea and oesophagus. Upon arrival at the intestine the larvae undergo their final moult to create juvenile worms which then mature into adults. This entire process of maturation, from the initial ingestion of eggs to generation of adult worms in the gut, takes approximately two months (Crompton and Pawlowski, 1985). Since the life span of the adult worms in the gut is thought to be up to 18 months, the worms can spend a total of up to 20 months within their host. It is during the adult stage of its life cycle that *Ascaris* causes the host to experience nutritional and intestinal blockage problems, which can be severe enough to cause death. These problems are primarily a consequence of the enormous bulk of foreign material that an individual is subjected to during *Ascaris* infection. Two facts serve to illustrate why this should be the case: (1) adult female worms can be up to 30 cm in length, and males up to 20 cm (Crompton and Pawlowski, 1985); (2) within the gut of the host there may be as many as 800 adult worms (Baird *et al.*, 1984).

Each mature female that is resident in the host gut can produce as many as 200,000 eggs per day, which are released into the soil with the host's faeces (Elkins *et al.*, 1986). Once situated in the soil the eggs embryonate to produce L1 stage larvae (which are enclosed within the egg) under the following conditions: moisture, oxygen, shade and favourable temperatures (14 days at 30°C, or 45 days at 17°C). After 10 days, the L1 larvae within the egg moult to produce the L2 stage larvae, which remain

coiled within the egg and have the remarkable ability to remain viable for up to 14 years before being ingested by the host (Krasnonos *et al.*, 1978).

1.3 Comparison of *A. lumbricoides* and *A. suum*

The above life cycle of *A. lumbricoides* is highly reminiscent of the *A. suum* cycle, apart from the latter's host being the pig instead of humans. Due to this similarity, and because there has been a considerable debate over whether these nematodes actually belong to separate *Ascaris* species (reviewed by Crompton, 1989), this sections compares a number of features of the two putative species.

One of the first experimental attempts to address whether this species distinction was correct involved asking whether *A. suum* can infect humans in addition to its normal host (the pig), and conversely whether *A. lumbricoides* can infect pigs as well as humans. These experiments involved infecting pigs with *Ascaris* eggs isolated from humans (de Boer, 1935; Galvin, 1968) and infecting humans with eggs isolated from pigs (Takata, 1951). The result was that parasite eggs collected from both human-derived and pig-derived *Ascaris* successfully infected either host species. Since experiments like these are no longer considered ethical they have not been repeated, but recently Lord and Bullock (1982) and Shoemaker-Nawas *et al.* (1982) reported cases where humans living in close proximity to pigs in areas of the U.S.A. where no cases of *A. lumbricoides* have occurred locally have been shown to be infected with *Ascaris*. These data have been interpreted as implying that the two species should be re-classed as one *Ascaris* species.

Physical examination of the two putative species reveals few gross structural differences. However, Sprent (1952) observed that *Ascaris*

obtained from humans contain small denticles that have concave edges, while *Ascaris* from pigs contain larger, more prominent denticles with straight edges. This work was supported by experiments performed by Crewe *et al.* (1971), Abdulrachman *et al.* (1954), Maung (1973) and Ansel *et al.* (1973). Lysek (1963) and Kurimoto (1974) suggested, however, that denticular morphology alters with age, and consequently these differences may simply be due to inconsistencies in the ages of the *A. suum* and *A. lumbricoides* worms analysed.

Kennedy *et al.* (1986) have utilised immunological techniques to compare *A. suum* and *A. lumbricoides* species by examining the excretory-secretory proteins of both putative species. Using radio-immunoprecipitation and SDS-PAGE techniques (see Section 1.4.2), these workers showed that the excretory-secretory proteins can induce immune responses in the host. This research revealed that *A. lumbricoides* L2 larvae produce a 17 kDa antigen and the L3/L4 larvae a 25.5 kDa and a 118 kDa protein for which there are no equivalent proteins excreted by *A. suum* L2 larvae and L3/L4 larvae. It was suggested that these data may represent differences between the species, but the possibility could not be excluded that these are superficial differences arising as a result of using worms from different geographical areas (*A. suum* from Scotland, and *A. lumbricoides* from India). The authors suggested that before any conclusions could be drawn a world-wide population of worms would have to be analysed.

Mutafova (1983) and He *et al.* (1986) examined the karyotypes of both species to determine if there were any chromosomal differences. They showed that although each species contains identical numbers of chromosomes (2n=20), there appears to be differences in the morphology of the chromosomes. Such a difference is seen in the morphologies of the chromosomes at the metaphase plate, where *A. lumbricoides* appears to

have five pairs of medium-size metacentric chromosomes while six pairs are present in *A. suum*. It was concluded from these data that *A. lumbricoides* and *A. suum* are indeed separate *Ascaris* species.

Recent work by Anderson *et al.* (1993) has compared the mitochondrial genes of the two putative species. This involved using PCR to amplify specific mitochondrial genes from the two species (both isolated from Guatamala), and analysing the products by gel electrophoresis after digestion with various restriction endonucleases. They found that, in general, the PCR products derived from worms isolated from the same host species (pigs or humans) were identical. They also showed that the PCR products generated from pig-derived and human-derived *Ascaris* could be distinguished. These data suggest that differences exist within the mitochondrial genomes of *Ascaris* populations that infect the two host species. The authors state, however, that this work only examined a few genes and that until more genes have been characterised no concrete conclusions can be made.

The above research suggests that it is still not clear whether *A. lumbricoides* and *A. suum* are the same or separate species. Perhaps when more extensive molecular biological research has been performed it will become clear whether they are different species. An example of such work is reported in Chapter 6, where the genes encoding the ABA-1 proteins of both *A. suum* and *A. lumbricoides* are compared. It is clear, however, that the most critical experiments required to examine this question is to determine whether the two putative species are able to cross fertilise with each other.

1.4.1 Immunological work

A major consideration in examining the host's immune response to *Ascaris*, and all parasitic helminths, is the amount of foreign material presented to the hosts. Infected individuals cannot utilise phagocytosis to respond to these parasites as they do to infection by single-cell protozoan parasites and viruses, because they are too large (Lightowlers and Rickard, 1988). Almond and Parkhouse (1985) suggested that there are three classes of parasitic helminth proteins that infected individuals are able to mount an immune response to: excretory-secretory proteins, surface proteins and somatic proteins (which can be exposed when the parasites moult, become damaged or die within the host). For many years researchers have been characterising these classes of proteins with the aim of understanding what kind of immune response the hosts mounts to them; i.e. is there an efficient immune response that leads to the expulsion of the parasite from the host, and does it provide future immunity to the parasite (reviewed by Almond, 1985)? If an isolated protein produced such a response it could be utilised as a vaccine against infection. One such protein, the *Schistosoma mansoni* gp28 excretory protein, has been shown to produce immunity to *Schistosoma mansoni* infection in experimental hosts when it is expressed as a recombinant derivative (Balloul *et al.*, 1985). This work provides hope for the identification of proteins present in other helminths that are capable of acting as vaccines.

A more fundamental aim in characterising parasitic proteins belonging to the three classes of proteins described above is to determine their function. One set of proteins whose biological role has been determined are a family of proteinases which are thought to be involved in allowing the parasite to penetrate the host's tissues (reviewed by Knox and Jones, 1991). These proteins are vaccine candidates because the parasite

would be unable to establish its infection if an immune response mounted by the host inactivated them. Another group of proteins that have been functionally characterised are those that the parasite excretes to evade or subvert the host's immune response in order to accomplish its infection (reviewed by Maizels *et al.*, 1993). An example of such a protein is superoxide dismutase, which is excreted by many helminth parasites in order to counteract the oxidative bursts produced by activated host leukocytes (reviewed by Knox and Jones, 1991). A further example is the aspartic protease inhibitor encoded by *Ascaris* which is thought to interfere with host antigen presentation in B cells (Bennett *et al.*, 1992).

The remainder of this introduction will review the research that has been performed to analyse the *Ascaris* molecules which induce an immune response in the host.

1.4.2 *Ascaris* antigens

To enable the characterisation of antigenic molecules in *Ascaris* that correspond to the three classes described in Section 1.4.1, Kennedy *et al.* (1986) have analysed the excretory-secretory products of the L2 and L3/L4 *Ascaris* larval stages and the adult *Ascaris* body fluid. These studies involved radio-labelling *Ascaris* proteins with iodine 125, immunoprecipitating them with sera from infected hosts (experimental mice, rabbits and rats infected with *A. suum*) and then analysing the products by SDS-PAGE. A range of antigenic molecules were found which had molecular weights ranging from 14 to 410 kDa. This analysis also suggested that *Ascaris* had stage-specific antigens as well as antigens that are expressed in all stages of the *Ascaris* life cycle. Explanations presented by Kennedy *et al.* (1986) and Kennedy and Knox (1988) for the stage-specific molecules were that the larvae have to migrate through different tissues

during development (L2 through the gut, and L3/L4 the lungs), and it is conceivable that they must express different proteinase at each stage to achieve this migration. Alternatively, it is suggested that the parasite varies its excretory products during development so that the host does not have time to mount an efficient immune response to them. Both these explanations may be correct, however, since many proteins can be described as being stage-specific. One protein with a molecular weight of 14 kDa was found to be expressed by the L2 larvae, L3/L4 larvae and to be present in the adult body fluid. This protein is termed ABA-1 (*Ascaris* Body fluid Allergen 1) and is discussed in Section 1.5.

To enable further characterisation of these antigens, Moore *et al.* (1992) have attempted to clone the genes which encode them, by screening cDNA libraries derived from L2 and L3/L4 stage mRNA with antiserum from infected hosts. Although still in its infancy, the aim of this work is to produce recombinant antigens that can be used to determine if any of them produce a protective response in the hosts. It is also hoped that this analysis will allow further characterisation of the molecules that *Ascaris* secretes in order to suppress or even repress the immune response mounted to the parasite.

1.4.3 IgE response to *Ascaris*

Generation of high IgE antibody levels in the host has been shown to be associated with infection by helminth parasites (Jarrett *et al.*, 1982), suggesting that this may be how an infected host tries to rid itself of such large multicellular pathogens. The exact mechanisms by which an IgE response attacks the parasite is unclear, but it is thought that the IgE molecules interact with mast cells and cause them to degranulate, releasing histamines and granules (Almond and Parkhouse, 1985). The

granules that are released are thought to attract eosinophils and neutrophils to the parasite, which presumably cause direct damage.

For many years researchers have been trying to isolate the *Ascaris* molecules which elicit such IgE responses in their hosts (Ambler *et al.*, 1974; Hussian, 1973; Dandeu and Lux, 1978; O'Donnell and Mitchell, 1978) by using assays such as Passive Cutaneous Anaphylaxis (PCA; Ambler, 1971). This assay involves sensitising an animal with test serum (e.g. serum from hosts infected with *Ascaris*) by intradermal injection and then challenging with the putative allergen. If IgE antibodies which recognise the allergen are present in the test serum the local mast cells degranulate, releasing histamine and resulting in increased vascular permeability. This is measured by the extravasation of the dye Evans Blue. Such work with *Ascaris* revealed that it contains many allergens ranging in size from 10 kDa to 350 kDa (reviewed by Almond and Parkhouse, 1985). Kennedy *et al.* (1989) have shown that the 14 kDa protein ABA-1, which is known to induce an IgG response in infected hosts (Section 1.4.2), also induces an IgE response (see Section 1.5 for further discussion). It should be noted that it is possible that some, if not all, *Ascaris* allergens may contribute to the lethal hypersensitivity responses seen in some humans during ascariasis. This response is a type I hypersensitivity response that results from IgE binding to the allergen and to mast cells, which causes the cells to degranulate and release vasoactive amines, thus inducing inflammation.

1.4.4 Genetic control of the immune response

Kennedy *et al.* (1986) have reported that experimental animals infected with *Ascaris* did not all recognise the same antigens and allergens. This restricted recognition has been shown, when analysing the excretory products of many nematode infections, to be under the control of the

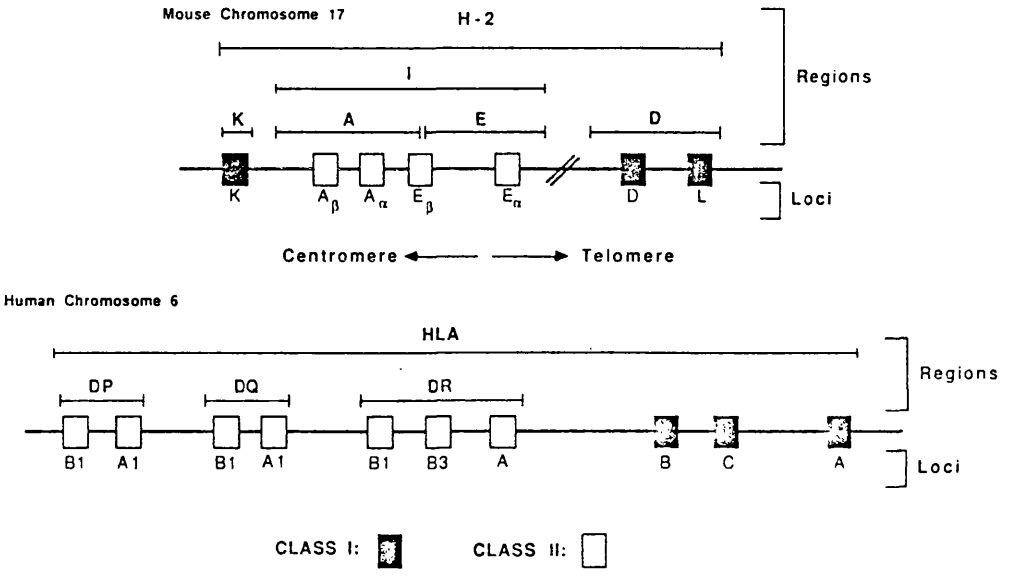
host's Major Histocompatibility Complex (MHC) genes, which play a major role in the host's immune response (reviewed by Wakelin, 1986 and Kennedy, 1989).

The MHC genes encode two classes of proteins: class I and class II. Class I molecules are found on the surface of all cells and interact with cytotoxic T cells, while class II molecules are found only on the surface of immune cells such as antigen-presenting cells, B cells and T cells. Both classes of MHC molecules interact with the antigens on the surface of cells, after the antigens have been processed by the cell, and form a complex that can interact with the relevant immune cells. It is thought that MHC class I molecules interact with endogenous antigens while MHC class II molecules interact with exogenous antigens.

The MHC genes are highly polymorphic within populations, and it is this that is thought to be reason that some hosts respond to certain antigens while others do not. The MHCs of humans and mice have been characterised and are named the human leucocyte antigen (HLA) complex and H-2 complex respectively. The HLA complex lies on chromosome 6 and is divided into 4 major regions: A, B, C and D (see Figure 1.2). Regions A, B and C are thought to encode the MHC class I molecules, while the D region encodes the class II molecules. The H-2 complex of mice is situated on chromosome 17 and is divided into 4 main regions: K, I, S and D (see Figure 1.2). Class I molecules are encoded by the K and D regions, and class II molecules by the I region.

Because the MHCs of both humans and mice have been analysed, it is possible to exploit this understanding when analysing the immune response of hosts to parasitic infections. For example, it is possible to ask whether there is a correlation between the specific MHC haplotype of a given individual and their susceptibility to infection by *Ascaris*. Since this type of analysis involves substantial field work when examining humans,

Figure 1.2. MHC genes of mice and human. The top diagram shows the MHC genes of mice (H-2) and the bottom the MHC genes of humans (HLA). Both diagrams were adapted from Kennedy (1985).


Mouse Chromosome 17
H-2
I
K
A
E
D
Regions
K
A_β
A_α
E_β
E_α
D
L
Loci
Centromere
Telomere
Human Chromosome 6
HLA
DP
DQ
DR
Regions
B1
A1
B1
A1
B1
B3
A
B
C
A
Loci
CLASS I:
CLASS II:

and it is expensive to determine the MHC type of humans, it is easier to analyse the effects of the MHC in experimental mice infected with the parasite. For these purposes many inbred strains of mice that have known MHC types and have the same background genes have been developed (see Figure 1.3: Klein and Klein, 1987).

Tomlinson *et al.* (1989) and Kennedy *et al.* (1986 and 1990) utilised this knowledge of the MHC locus in mice to determine if the IgG and IgE-restricted recognition of *Ascaris* antigens is due to the MHC haplotype of the host. They showed that this was indeed the case, since only mice with the same haplotype recognised the same set of *Ascaris* antigens. Of particular interest was that only mice with the H-2^s haplotype recognised the 14 kDa antigen, ABA-1 (see Section 1.5 for discussion of this protein). One of the implications of a restricted genetic response to the antigens of *Ascaris* is that they are unlikely to be of use as vaccines, because not all hosts will be able to mount an immune response to the antigen. It would therefore be necessary to develop strategies based on a cocktail of antigens in order to achieve broad immunity.

1.5.1 The *Ascaris* body fluid allergen, ABA-1

As reported in Sections 1.4.2 and 1.4.3, Kennedy *et al.* (1986) have shown that hosts infected with *Ascaris* mount both an IgG and an IgE response to a 14 kDa protein, named ABA-1, that is excreted by the L2 and L3/L4 *Ascaris* larval stages and is found in large abundance in the adult *Ascaris* body fluid. Kennedy *et al.* (1986 and 1990) have shown that only mice with the H-2^s haplotype respond to ABA-1.

To enable further characterisation of ABA-1, Kennedy *et al.* (1990) purified the protein from the *Ascaris* body fluid and used it for further

Figure 1.3. MHC haplotypes of classical and recombinant strains of mice.

Strain	Haplotype	K	A-beta	A-alpha	E-beta	E-alpha	D
Classical							
BALB/c	d						
NIH	q						
CBA/Ca	k						
B10.R111	r						
NZW	z						
B10.HTG	g						
B10.AKM	m						
B10.A2G	a						
B10.M	f						
B10.BR	k	k	k	k	k	k	k
C57Bl/10	b	b	b	b	b	b	b
SJL	s	s	s	s	s	s	s
Recombinant							
B10.A5R	i5	b	b	b	b/k	k	d
B10.S9R	t4	s	s	s	s/k	k	d
B10.A4R	h4	k	k	k	k/b	b	b
ATL	t1	s	k	k	k	k	d
ATH	t2	s	s	s	s	s	d
B10.A2R	h2	k	k	k	k	k	b
B10.AQR	y1	q	k	k	k	k	d
B10.D2	g1	d	d	d	d	d	b
B10HTT	t3	s	s	s	s/k	k	d

immunogenetic studies to determine exactly which locus of the MHC is responsible for the response to ABA-1. They analysed the immune response of mice with recombinant H-2 haplotypes (see Figure 1.3) to ABA-1. These data revealed that it was the I-A region that was responsible for the recognition of ABA-1. Since this region encodes MHC class II molecules (see Figure 1.2), these data suggest that ABA-1 recognition depends on whether the host's MHC class II molecules can interact with processed fragments of ABA-1 in such a way to induce an immune response *via* T-cells. These data support the finding that it is exogenous proteins that interact with the host MHC class II molecules.

Since the above work was performed on experimental animals, it was not known at this stage if the same restricted genetic response is found in humans. Haswell-Elkins *et al.* (1989) and Kennedy *et al.* (1987) have shown that humans also show a restricted IgG and IgE response to ABA-1, although the MHC types of responders and non-responders has not yet been determined. It is therefore not known if the immune response in humans to ABA-1 is under the control of the host's MHC haplotype or whether other background genes are involved. It is also not known if those individuals that respond to ABA-1 show a lesser level of *Ascaris* worms during infections than individuals that do not respond, which would indicate that the host's response to ABA-1 could control the infection.

1.5.2 Biochemical analysis of ABA-1

As described above, Christie *et al.* (1990 and 1993) have purified ABA-1 from *A. suum* adult body fluid, thereby allowing biochemical analysis of the protein. The isoelectric point of ABA-1 is estimated to be 5.2 and its molecular weight has been shown to be 14.642 kDa by mass

spectrometry, which is approximately equivalent to the mass of 14.4 kDa estimated by SDS-PAGE (Kennedy *et al.*, 1986). In contrast to these findings, McGibbon *et al.* (1990) and Greenspon *et al.* (1986) have described the molecular weight of ABA-1 to be between 9 kDa and 10 kDa by SDS-PAGE and HPLC. The reasons for, and significance of, these discrepancies are unknown. The N-terminal amino acid sequence of ABA-1 has been determined by direct sequencing of the purified protein, and is shown in Figure 1.4. In conjunction with the general amino acid composition of the protein (analysed in the same study), the N-terminal sequence has suggested that ABA-1 may be identical to another *Ascaris* allergen, named Allergen A (Almber *et al.*, 1972; Christie *et al.*,1990).

Because the above work on the ABA-1 protein was performed with *A.suum* ABA-1, Christie *et al.* (1990) also purified the *A. lumbricoides* ABA-1 homologue. The first 41 amino acid residues at the N-terminus of this protein was similarly derived and shows 100% identity to the sequence determined for *A. suum* ABA-1. These data confirm that ABA-1 is present in both species.

1.5.3 ABA-1 Homologues

Toxocara canis and *Anisakis simplex* have been shown to contain a 14 kDa protein which is recognised by sera from a host infected with *Ascaris* (Kennedy *et al.*, 1988). These data suggest that these nematodes contain a homologue of ABA-1. Christie *et al.* (1993) have determined the N-terminal amino acid sequence of the cross-reactive protein from *Toxocara* (named TBA-1) and have compared it to the ABA-1 sequence (see Figure 1.4). TBA-1 shows 65% identity to ABA-1, suggesting that these molecules are evolutionary related and therefore homologous.

Figure 1.4. Comparison of the N-terminal amino acid sequence of *A.suum* ABA-1 with the amino acid sequence of its homologues in *B. pahangi*, *D. immitis* and *T. cannis*. Where the amino acid residues of ABA-1 from *A.suum* are identical to those of the *B. pahangi*, *D. immitis* and *T. cannis* homologues they are represented as (-), while amino acid differences are shown by their single letter amino acid code.

```
Ascaris     HHFTLESSLDTHLKWLSQEQKDELLKMKKDGKAKKELEAKILHYYDELEG
Brugia      -EHS-DDYFR---S--TDA----IR---EE--Q-MDMQK---D--EN-T-
D.immitis   NDHN---YFQ-Y-S--TDA----IK---EE--S-MDIQK--FD-FES-T-
Toxocara    ------G--ESL--L-G--IV-*-----*-K-I--T-***-
```

```
Ascaris     DAKKEATE*LKG***EIL
Brugia      -G----G-K-R-----L-
D.immitis   -K--K-AA-E-QQGCLMA
```

While work for this thesis was in progress the genes encoding antigens from two filarial parasites were characterised: the *Brugia pahangi* and *Brugia malayi* gp15/400 proteins (Tweedie *et al.*, 1993), and the *Dirifilaria immitis* 15 kDa antigen (Poole *et al.*, 1992; Culpepper *et al.*, 1992). Comparing the putative amino acid sequences derived from the cDNAs encoding these proteins to the N-terminal amino acid sequence of ABA-1 suggested that they were homologues of ABA-1 (Figure 1.4). Analysis of the structure of the cDNA clones suggested that the ABA-1 homologues are encoded by genes with a highly unusual structure, consisting of numerous tandem repeats (this is considered in more detail in Chapter 3).

Work by Paxton *et al.* (1993) has shown that humans infected with *B. malayi* elicit an IgE and an IgG response to the *Brugia* gp15/400. This similarity between the antibody responses to gp15/400 during *Brugia* infection and to ABA-1 during *Ascaris* infection highlights the sequence conservation between the two homologues, and may indicate similarities in function. Although gp15/400 elicits these antibody responses, it has been shown using experimental mice with known MHC types that the IgG response to this molecule is not solely under the control of the MHC of the hosts, as is the case for the recognition of ABA-1. It is believed that some other genes outside the MHC influence the reaction (Kwan-Lim and Maizels, 1989), but these have not been characterised.

Recently, Renyolds *et al.* (1993) have reviewed all the known homologues of ABA-1 that have been isolated, and have suggested they constitute a family of nematode polyprotein allergens (which they term the NPA family). Since this article was written, cDNAs encoding a similar protein have been identified from the free-living nematode *C.elegans* (Moore *et al.*, 1994). The significance of these findings are discussed in Chapter 3.

The ABA-1 protein of Ascaris is thus important because its a major allergen, because the immune response to it appears to be under the control of the hosts MHC, and because other parasites appear to express homologues of it. The work detailed in this thesis was carried out in order to clone the gene which encodes the ABA-1 protein in *A. suum* and *A. lumbricoides*, with the long term aims of being able to characterise the gene and protein at the molecular level and also to enable further immunological work.

Chapter 2

Materials and Methods

Table 1 Bacterial strains

Strain	Genotype	Source
XL1Blue	*recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac.*	Stratagene
KW251	F^-, *supE44, supF58, galK2, galT22, metB1, hsdR2, mcrB1, mcrA-, argA81*:Tn*10, recD1014*	Promega

Table 2 Lambda clones

Name	Description	Source
λZAP II	cDNA cloning vector	Stratagene
λGEM 12	derivative of EMBL3	Promega
λHS1	1.2 kbp *A. suum aba-1* cDNA insert in λZAP II	This work
λHS2	0.6kbp *A. suumaba-1* cDNA insert in λZAP II	This work
λHS3	2.0 kbp *A. suumaba-1* cDNA insert in λZAP II	This work
λHS4	unknown *A. suum* cDNA insert in λZAP II	This work
λHSg5	*aba-1* genomic insert in λGEM12	This work
λHSg6	*aba-1* genomic insert in λGEM12	This work
λHSg7	*aba-1* genomic insert in λGEM12	This work
λHSg8	*aba-1* genomic insert in λGEM12	This work

Table 3 Plasmids

Name	Description	Source
pbluescript	derived from M13	Stratagene
pUC19	derived from pBR322	Stratagene
pT-Blue	designed to clone PCR products	Invitrogen
pHS10	1.2 kbp *A. suum aba-1* cDNA insert in pbluescript	This work
pHS20	0.6 kbp *A. suum aba-1* cDNA insert in pbluescript	This work
pHS30	2 kbp *A. suum aba-1* cDNA insert in pbluescript	This work
pHS40	0.75 kbp *A. suum aba-1* cDNA insert in pbluescript	This work
pHS50	5 kbp fragment of λHSg7	This work
pHS60	8 kbp fragment of λHSg7	This work
pHS11	pHS10 deletion	This work
pHS12	pHS10 deletion	This work
pHS13	pHS10 deletion	This work
pHS14	pHS10 deletion	This work
pHS15	pHS10 deletion	This work
pHS16	pHS10 deletion	This work
pHS17	pHS10 deletion	This work
pHS100	*A. lumbricoides aba-1* pHS10- Like repeat in pT-Blue	This work
pHS200	*A. lumbricoides aba-1* divergent repeat in pT-Blue	This work

Table 4 Chemicals

CHEMICALS	SOURCE
General chemicals, bichemicals and organic solvents	BDH and Sigma
Media	Difco and Oxoid
Agarose and antibiotics	Sigma
Radiochemicals and HybondN	Amersham

2.1 Oligonucleotide primers

HS10-NTERM	5' GGAATTCCATCATTTCACCCTTG 3'
HS10-CTERM	5' GGAATTCCCTCCTTCGTCGCGAAG 3'
Divergent-N	5' GGGAATTCCATACAATGGAACACTATC 3'
Divergent-C	5' GGGAATTCCCTCCTTCGTCGATGATG 3'
CBR2	5' CACGCTGAGGATGGAAC 3'
INTRON1	5' TCTTCTTTCAACACAGG 3'
INTRON2	5' ATTTCACTCTCACTTAAAG 3'

2.2 Bacterial strains, plasmids and transformations

The derivatives of *Escherichia coli* K-12 and the plasmids constructed for this work are listed in Table 1 and table 3 respectively. Transformations to introduce plasmid DNA into relevant *E. coli* strains were performed by standard $CaCl_2$ techniques (Cohen and Hsu, 1972 and Sambrook *et al.*, 1989).

2.3 Bacterial growth, media and conditions

Bacteria were grown in standard L-broth (Miller, 1972) and on L-agar at 37°C. Antibiotics were added where needed (see below). The bacterial strains were stored in 50% L-broth, 20% glycerol and 1% peptone at -20°C, or in thiamine-containing slopes at room temperature.

Antibiotic concentrations used for both liquid and plate selection were as follows:

Antibiotic	Stock	Solution	Selective concentration
Ampicillin	5 mg/ml	water	50 µg/ml
Tetracycline	1.25 mg/ml	10 mM HCl	12.5µg/ml

2.4.1 Parasites

Live adult *A. Suum* were collected from the Glasgow and Edinbourgh abattoirs. Worms were either dissected for the relevant parasite material and stored at -70°C until needed for the preparations of DNA, RNA and protein (see Sections 2.5, 2.6 and 2.7), or they were stored, after washing in dH_2O, in 2% formalin at 4°C for the production of *Ascaris* L2 and L3/L4 stage larvae. *A. lumbricoides* worms from Guatamala were a gift from Dr. T. Anderson.

2.4.2 Embryonation of *Ascaris* eggs

Adult worms were removed from 2% formalin and the females' uteri were dissected and placed into 5% Na Hyporchlorite (10-14% w/v chlorine) to dissolve the uterine material away from the eggs. Eggs were then washed 6 times in dH_2O (with centrifugation at 1,500 x g for 4 min each time) and embryonated by placing them in 2% formalin containing 125 units/ml of nystatin at 25°C in the dark. After 4 weeks, with agitation of the eggs once every week, the embryonation rate was checked by microscopic examination under a x10 objective; normally 60-80% of the total number of eggs were embryonated by this time. Once embryonated, the eggs were washed as above and stored under dH_2O at 4°C until they were required for hatching (Section 2.4.3), or for infection (Section 2.4.4).

2.4.3 Hatching of *Ascaris* embryonated eggs

Embryonated eggs were incubated in 5% Na Hyporchlorite at 39°C for 5-10 minutes until, when viewed under the microscope, the egg shells were seen to have been removed. At this time they were washed 6-8 times with dH_2O, as described in Section 2.4.2, to remove all traces of bleach. The eggs were then homogenised using a hand held glass tissue homogeniser (purchased from Jencons) until they could be seen to be hatched under x10 magnification, i.e. the larvae were free from their eggs. To separate the larvae from the unhatched eggs and eggshell debris, the larvae were put through a Baerman apparatus, which consists of a cotton wool plug at the surface of 12 ml of *Ascaris* medium (see Section2.4.4) in a universal (Kennedy *et al.*, 1986). The worms in the Baerman were incubated overnight in a 37°C water bath, during which time the larvae migrated through the cotton wool and were collected from the bottom of the universal. Hatched larvae were stored at -70°C until needed for DNA, RNA or protein isolation.

2.4.4 Production of lung stage larvae.

Adult, male New Zealand White rabbits weighing 2-3 kg (purchased from Interfauna) were infected orally with approximately 50,000 embryonated eggs. After 7 days the rabbit's lungs were removed and were homogenised in a blender and incubated in PBS (171 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) for 1 hr. The homogenate was sieved through a 0.45 mm Endecotts sieve (Gallenkamp) and *Ascaris* larvae were collected. After two washes in PBS, the larvae were then washed in an *Ascaris* medium (RPMI 1640 medium (purchased from Gibco) with the following: 2 mM L-glutamine, 1 mg/ml glucose, 400 ng/ml glycyl-L-Histidy-L-lysine, 16 μM glutathione, 100 i.u./ml penicillin, 100 μg/ml streptomycin and 1 mM sodium pyruvate). The larvae were

then placed in a Baerman apparatus and incubated overnight in a 37°C water bath. The next day, larvae were collected and stored at -70°C until they were needed for DNA, RNA or protein isolation.

2.5.1 Genomic DNA preparations

Two different techniques were used to isolate genomic DNA from *Ascaris*. The first technique was used to isolate DNA for construction of a genomic library (see Section 2.14), and the second was used for restriction analysis and Southern blotting (Section 2.17). Most of the genomic DNA prepared was isolated from the adult *Ascaris* male testes because this tissue contains the lowest concentration of glycogen (which co-precipitates with DNA).

2.5.2 Genomic DNA preparation (1)

This preperation is a modification of that described by Glover (1985). Approximately 1g of *A. suum* adult testes was removed from storage at -70°C and placed in a mortar, which was filled with liquid nitrogen and the tissue ground to a fine powder using a pestle. Powder was removed from the mortar and mixed with 8ml of homogenisation buffer (30 mM Tris-HCl pH8.0, 10 mM EDTA, 100 mM NaCl, 10 mM beta-mercaptoethanol, 0.62 % Triton X-100). The homogenate was centrifuged at 4,000 x g for 10 minutes at 4°C and the pellet resuspended in 1 ml of homogenisation buffer without Triton. 5ml of Nuclear lysis buffer (100 mM Tris-HCl pH8.0, 100 mM NaCl, 0.5 mg/ml Proteinase K) was added to the homogenate followed by 200 µl of 30% Sarkosyl. The solution was left at 37°C overnight, recentrifuged (as above), and the supernatant collected. The homogenate was weighed and 1.25 g of CsCl per gram of homogenate was added before being transferred to a Pollyallomer tube which was

subsequently filled with a solution of 1.25 g CsCl/ml in dH_2O. This was subjected to ultracentrifugation at 45K for 24hrs at 25°C. After centrifugation, the tube was punctured at the top with a fine needle and at the bottom with a 18G Microlance 2 needle, and fractions collected through the bottom needle. To determine which fraction contained the DNA, they were spotted onto ethidium bromide agar plates and viewed by 254 nm UV illumination. The best fractions were pooled and dialysed against TE (10 mM Tris-HCl, pH 7.4, 1 mM EDTA).

2.5.3 Genomic DNA preparation (2)

This preparation is a modification of that described by Sambrook *et al.* (1989). *Ascaris* material (adult testes and embryonated eggs) were removed from storage and ground to a fine powder in liquid nitrogen (as above). The powder was transferred to extraction buffer (10 mM Tris-HCl pH8.0, 0.1 mM EDTA, 0.5% SDS) and Proteinase K was added to a final concentration of 100 µg/ml. The solution was left at 50°C for 3 hr, at which time an equal volume of phenol (equilibrated with 500 mM Tris-HCl pH8) was added and the solution was mixed gently for 10 min, followed by centrifugation at 5, 000 x g for 10 min. The upper aqueous layer was then removed and added to an equal volume of phenol. This phenol extraction was repeated 3 times. After the final extraction, 0.2 volumes of 2.5 M Na. Acetate and 3 volumes of ethanol were added. The solution was then centrifuged at 5,000 x g for 10 min and the precipitated DNA pellet washed in 70% ethanol and then resuspended in TE pH8.0.

2.6 RNA preparation

Approximately 500 mg of *Ascaris* material (whole adults, eggs, L2 stage larvae or L3/L4 stage larvae) was ground to a powder in liquid

nitrogen with a mortar and pestle and then transferred to an 1.5 ml eppendorf containing a preheated 400 µl solution of dH_20-saturated phenol and 200 µl of 2xNETS (200 mM NaCl, 2 mM EDTA, 20 mM Tris-HCL pH7.0, 1% SDS). This solution was immediately vortexed and then cooled on ice, followed by centrifugation in a microcentrifuge at 15,000 x g for 40 seconds. The upper layer, the aqueous phase, was then transferred to a second 400 µl solution of of phenol and subsequently vortexed and centrifuged again (as above). Again the aqueous phase was removed and a final extraction was performed as above but using 200µl of phenol. 1 ml of ethanol was then added to the final aqueous phase and the RNA was precipitated at -70°C for 1 hr. The precipitated RNA was recovered by centrifugation in a microcentrifuge at 15,000 x g for 80 seconds and the ethanol was removed. The RNA pellet was left to dry at room temperature and then re-dissolved in 50 µl of dH_20 by vortexing. 50 µl of 5M LiCl was then added to the RNA followed by a 1 hour incubation on ice. The RNA was recovered by centrifugation and the RNA pellet was resuspended in 50µl of a 200 mM NaCl solution. A final ethanol precipitation was performed by the addition of 150 µl of ethanol followed by a 1 hr incubation on ice and centrifugation. The final RNA pellet was resuspened in dH_2O to a concentration of approximately 1 mg/ml.

2.7 Preparation of *Ascaris* protein homogenate.

Ascaris material (see Chapter 4) was ground to a powder in liquid nitrogen by the use of a mortor and pestle, and the powder was transferred to a solution of Tris homogenisation buffer (1 mM Tris-HCl pH 8, 2 mM EDTA, 1 mM phenylmethylsulphonyl fluoride in isopropanol, 5 µM pepstatin in methanol, 2 mM 1, 10 phenanthroline in ethanol, 5 µM leupeptin, 5 µM antipain, 25 µM N-p-tosyl-L-lysine chloromethyl ketone,

50 µg/ml N-tosyl-L-phenyl alanine choromethyl ketone). The homogenate was then vortexed and cooled on ice for 30 min followed by centifugation in a microfuge at 15,000 x g for 30 min. The upper soluble layer was then stored at -70°C until needed.

2.8 Lambda DNA preparation

Approximately $5x10^9$ pfu of phage prepared from a plate lysate (Miller, 1972) were added to a 200 ml culture of *E.coli* (OD_{600}=0.5) and incubated at 37°C until lysis occurred (3-5 hrs). After lysis, 4ml of chloroform was added and the 37°C incubation resumed for a further 10 min. DNAaseI and RNAase were both then added to a final concentration of 1 µg/ml and the lysate was incubated at room temperature for 30 min. Solid NaCl was added to a final concentration of 1M and dissolved by shaking the lysate. After a 1 hr incubation on ice the lysate was centrifuged at 10,000 x g for 10 min at 4°C. A final concentration of 10% w/v of PEG 8000 was added to the supernatent and it was precipitated on ice overnight. Precipitated phage were recovered by centrifugation at 10,000 x g for 20 min at 4°C and the pellet was resuspended in phage buffer (50 mM Tris-HCl pH7.5, 100 mM NaCl, 8 mM $MgS0_4$, 0.01% gelatin) and extracted once with an equal volume of chloroform. The phage solution was then weighed and 0.75g of CsCl was added for every 1ml of phage solution and loaded into Pollyallomer tube. The tube was filled with 0.75 g/ml CsCl in phage buffer and centrifuged at 49K for 24 hr at 20°C. The phage band was then collected and dialysed twice for 1 hr each time against 1000 x volums of 10 mM NaCl, 50 mM Tris-HCl pH8.0, 10 mM $MgCl_2$. A final concentration of 20 mM EDTA, 50 µg/ml of Proteinase K, 0.5% SDS was then added and the solution incubated at 56°C for 1 hr. This solution was then phenol extracted twice, extracted once with chloroform and then dialysed

overnight against 1000 x volumes TE. The DNA was then stored at -20°C until needed.

2.9.1 Plasmid preparations

Three methods for preparing plasmid DNA from bacterial cultures were used.

2.9.2 Plasmid preparation (1)

For rapid analysis of plasmids, the single colony lysis technique was used. In this, a single transformant colony was patched out as a 1 cm square on an L-agar plate and grown overnight at 37°C. The patch was collected and resuspended in 100-200 µl of single colony gel buffer (2% ficoll, 1%SDS, 0.01% bromophenol blue in Tris-acetate buffer). The cells were allowed to lyse at room temperature for 15 min, cell debris and chromosomal DNA were spun down in a microcentrifuge at 13,000 x g for 30 min at 4°C and 30µl of the supernatant loaded onto an agarose gel.

2.9.3 Plasmid preparation (2)

The "Magic Minipreps DNA Purification System" of Promega was used to isolate plasmid DNA for sequencing and restriction analysis. The manufacturers' instructions were followed.

2.9.4 Plasmid preparations (3)

To obtain a large quantity of highly purified plasmid suitable for long-term storage of DNA, the technique described by Birnboim and Doly (1976) was used.

2.10 *in vitro* DNA manipulations

DNA manipulations were performed essentially as described by Sambrook *et al.* (1989).

2.10.1 Restriction Digestions

Restriction digestions were performed in 10-30 µl of 1x restriction buffer, and contained 0.5-10 µg of DNA and 1-30 units of restriction enzymes. Both the enzymes and buffers were purchased from IBI, Promega or Appligene.

2.10.2 Ligation of DNA ends

DNA fragments were ligated for 2-24 hrs at either room temperature or 4°C in 20 µl of 1x ligation buffer, using 1-3 units of T4 DNA ligase (purchased from Promega).

2.10.3 Exonuclease III digestions

The Promega "Erase-a-Base System" kit was used to make Exonuclease III deletions to the *aba-1* cDNA clone pHS10 (see Chapter 3).

2.11 Gel Electrophoresis of DNA

DNA was analysed on two kinds of agarose gels: standard 0.5-1% gels, and 1% low melting point (LMP) gels. DNA samples were applied to the gels in Ficoll-containing loading buffer (Sambrook *et al.*, 1989). Standard gels were run in either TAE (40 mM Tris. Acetate (pH8.2), 20 mM Na. Acetate, 1 mM EDTA) or TBE buffer (89 mM Tris HCl pH8.0, 89 mM boric acid, 2 mM EDTA), while low melting agarose gels were only run in TAE. The preparation and running of the LMP gels was as described by the agarose manufacturers' instructions (Sigma). IBI electrophoresis kits were

used which contained 1 litre of buffer; the gels were routinely run for 15 hrs at 5 volts /cm

DNA was visualised by 254 nm UV illumination after staining with ethidium bromide. Ethidium bromide-stained gels were photographed using Polaroid type 67 land film.

2.12.Extraction of DNA from agarose gels

DNA was purified from standard agarose gels by cutting a slice of agarose from the gel containing the DNA fragment of interest and then centrifuging the gel slice at low speed through siliconised glass wool (Heery *et al.*, 1990). From LMP gels, DNA-containing gel fragments were isolated in the same way and were heated to 70°C for 10 min. An equal amount of dH_2O was then added, two phenol/chloroform extractions were performed and the DNA recovered by ethanol precipitation.

2.13 DNA sequencing

Sequencing reactions were performed on double stranded plasmids using the dideoxy sequencing reaction (Sanger *et al.*, 1977: Sambrook *et al.*, 1989). The sequenase version 2.0 kit was used (USB).

2.14 Construction of *A.suum* genomic library

2.14.1 Preparation of the genomic DNA

The lambda GEM-12 half arm site vector (purchased from Promega; Zabarovsky *et al.*, 1986) was used to construct an adult *Ascaris* testes genomic library. Restriction digestions with 2 µg of genomic *Ascaris* DNA were performed using 0.1 units of *Sau*3A for either 5, 10, 20, 40 or 80 min,

to determine which time interval produced DNA fragments in the range of 15-23 kbp (digestions were analysed on a 0.4% agarose gel). The 20 min reaction gave this range, and the reaction was then scaled up in order to digest 10 µg of *Ascaris* DNA. The digested DNA was then phenol extracted to remove protein, phenol/chloroform, extracted and finally chloroform extracted to remove the phenol. Ethanol precipitation (using 2 volumes ethanol and 0.5 volume of 7.5M ammonium acetate) was then performed and the DNA was resuspended in 35 µl of TE.

To prevent insert-insert ligations instead of vector-insert ligations, the recessed ends of the DNA prepared above was partially filled-in with 10 units of Klenow in the following buffer: 500 mM Tris-HCl pH7.2, 100 mM MgSO4, 1 mM DTT, 500 µg/ml BSA, 10 mM dATP, 10 mM dGTP. The reaction was incubated at 37°C for 30 min. Two phenol/chloroform extractions and one chloroform extraction were then performed and the DNA was ethanol precipitated (as above) and resuspended in 10 µl of TE.

2.14.2 Ligation of the genomic DNA into lambda GEM-12

Approximately 0.5 µg of the genomic DNA prepared above was ligated to 0.5 µg of pre-restricted lambda vector in 1x ligase buffer using 1 unit of ligase and overnight incubation at 4°C. Ligated DNA was packaged using the "Packagene" *in vitro* packaging system of Promega at 22°C for 2 hours following the manufacturers' instructions. 0.5 ml of phage buffer and 25 µl of chloroform was added to the packaged DNA and the titre of the library determined as described by Sambrook *et al.* (1989). Half of the library was screened for the *aba-1* gene while the other half was amplified (according to the manufacturers' instructions).

2.15 Genomic library screening

Approximately $0.25x10^6$ plaque forming units were allowed to absorb onto *E.coli* strain KW251, as described in Sambrook *et al.* (1989), and were subsequently plated onto L-agar plates and left at at 37°C for 5 hrs. A nylon membrane (Hybond N, purchased from Amersham) was then placed over the plaques for 3 min to allow DNA transfer onto it. The DNA was then denatured in 0.5 N Na OH, 1.5 M NaCl for 5 min and neutralised in 1.5 M NaCl, 0.5 M Tris-HCl pH8.0 for 5 min and washed in 20 x SSC (3M NaCL, 300 mM sodium citrate.) for 1 min. The membranes were then air dried and the DNA cross-linked onto them by exposure to 254 nm UV irradiation for 5 min. Membranes were prehybridised and then hybridised with ^{32}P-labelled probe as described for Southern blots (see Sections 2.17 and 2.18).

2.16 *Sfi*I mapping

To produce an initial restriction map of the *aba-1* genomic fragment of λHSg7 (see Chapter5) the Promega "*Sfi*I linker mapping system" was used.

2.17 Southern Blotting and hybridisation

Agarose gels were soaked in 250 mM HCl for 15 min, denatured in 1.5 M NaCl, 0.5 N NaOH for 30 min and then neutralised in 1mM Tris-HCl pH7.4, 1.5 M NaCl for 30 min. The DNA was then transferred to a nylon membrane (Hybond N, Amersham) by capillary blotting overnight in 20 x SSC. The membrane was next air dried and the DNA cross-linked to it by exposure to 254 nm UV irradiation for 5min. The blot was prehybridised for 3 hrs at 68°C in 5 x SSC, 5x Denhardt's solution (50 x

stock: 5 g of Ficoll, 5 g polyvinylpyrrolidone, 5 g bovine serum albumin made up to 500 µl in dH_2O), 0.5% SDS and 20 µg/ml of sonicated, denatured salmon sperm DNA. ^{32}P-labelled probe (see Section 2.18) was denatured by boiling for 10 mins at 100°C and was added to the prehybridisation solution and allowed to hybridise at 68°C overnight. Blots were then washed twice in 2 x SSC, 0.1% SDS at room temperature and in 0.5 x SSC, 0.1% SSC at 68°C for 10 min. The blot was then exposed to X-ray film (Kodak film) at -70°C overnight.

2.18 Radio-labelling of DNA

^{32}P-labelled gel-purified DNA fragments were prepared by "random priming" as described by Hodgson *et al.* (1987). Probes were purified from unincorporated nucleotides by chromatography in Sephadex G50 columns purchased from Pharmacia.

2.19 RNA agarose gels, Northern blots and hybridisation

Approximately 5 µg of total RNA was added to a solution containing 10 µl formamide, 3.5 µl formaldehyde, 2 µl of 10 x MOPS (200 mM morpholinoprpanesulphoric acid, 50 mM Na.Acetate, 10 mM EDTA, pH7.0). The RNA was then heated to 60°C for 15 min and loaded onto a 1% agarose gel containing 10 ml of 10 x MOPS and 16 ml of 37% formaldehyde in a total volume of 100 ml (made up with dH_2O), and was run in 1x MOPS buffer. The RNA was then transferred on to a nylon membrane (Hybond N, Amersham) by capillary blotting overnight in 20 x SSC. RNA was cross-linked to the membrane by exposure to 254 nm UV irradiation for 5 min. The blot was prehybridised for 3 hrs at 42°C in 5x SSC, 5x Denhardt's solution, 0.5% SDS, 50% formamide. Denatured ^{32}P-

labelled probe (see Sections 2.17 and 2.18) was then added and hydridisation was performed overnight at 42°C. Washes were performed after probe hybridisation as for Southern blots (Section 2.17).

2.20 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE) of protein and Western blotting

Gels were made in a Pharmacia GE/4LS slab gel apparatus according to the manufacturers' instructions. Gradient gels consisted of 120 mm of separating gel (5% to 25%) and 15 mm of stack gel of 4%, and were 0.7 mm thick. Protein samples were prepared by addition of 40 µl of SDS-PAGE sample buffer (5 g Sodium Dodecyl Sulphate, 5 ml 1M Tris pH 7.5, 2 ml 100 mM phenylmethyl sulfonyl fluoride in isopropanol, 1 ml 100 mM EDTA; 10 ml glycerol, 2 ml 0.2% Bromophenol Blue; made up to 95 ml with dH_2O, and 5% beta-mercaptoethanol was also added to allow reducing conditions). After mixing, samples were incubated in a boiling water bath for 10 min, and loaded onto gels. Gels were electrophoresed at a constant current of 8 mA overnight.

Gels were blotted onto nitrocellulose membrane in a TransBlot tank (Biorad) at 40°C for 4 hrs at 200 mA and 60 V, or overnight at 100 mA and 60V, in blotting buffer (9g Tris, 43.2 g glycine, 600 ml methanol, 3 g SDS made up to 3 l with dH_2O, pH adjusted to 8.3). Blots were then air dried and blocked in blocking solution (5% skimmed milk made up in TBS: 20 mM Tris-HCl pH8, 150 mM NaCl) for 1 hr. Three washes were performed in TBST (TBS plus 0.05% Tween 20) for 5 min each at room temperature, followed by the addition of polyclonal rabbit antiserum raised to ABA-1 recombinant protein(which was used at a concentration of 1 in 500 made up in blocking solution; gift from J. Moore). Antiserum was incubated with the blots for 1 hour at room temperature. The blots were then

washed 3 times with TBST for 5 min and secondary antibody (anti rabbit alkaline phosphatase purchased from Sigma) was added, at a concentration of 1 in 750 in blocking solution, for 1 hr. Three washes in TBST were then performed (as above) followed by a final wash in TBS for 5 min. Filters were then developed by adding a colour development solution (nitro blue tetrazolium at final concentration of 0.3 mg/ml, 5-bromo-4-chloro-3-indolyl-phosphate at a final concentration of 0.15 mg/ml).

2.21 Screening of the cDNA expression library

10^6 pfu of the L2 *A. suum* cDNA library were plated on 10 plates(10 cm by 10 cm) as described by Sambrook *et al.* (1989) and were incubated at 37°C for 5 hrs until plaques were just visible. Nitocellulose membrane (purchased from Amersham), which had been soaked in 10 mM IPTG, was put over the top of the of the bacterial lawn and the plates were incubated at 37°C for a further 2 hrs. The filters were then treated in the manner described for the Western blots (section 2.20) in order to detect which phage were expressing *A. suum* ABA-1. ABA-1-positive clones were excised from lambda phage in the form of the plasmid Bluescript following the protocols described by Stratagene.

2.22 Polymerase chain reaction (PCR)

PCR reactions were performed using approximately 500 ng of genomic DNA or 100ng of vector DNA. The reaction was performed in 1 x Taq buffer using 1 unit of Taq DNA polymerase (both the enzyme and buffer were purchased from Promega), 200 µmM each of dATP, dTTP, dGTP and dCTP and 100 ng of each oligonucleotide primer. The reaction

involved an initial DNA denaturation at 95°C for 5 min followed by 30 amplification cycles each for 1 min at 95°C, 1 min at 55°C and 1min 72°C.

Chapter 3

Molecular cloning and characterisation of the cDNA copy of the *A. suum aba-1* gene

3.1 Introduction

Immunological work has shown that a 14 kDa protein of *Ascaris suum,* named ABA-1, is the most abundant protein in the adult body fluid and is also released by tissue penetrating larvae cultured *in vitro* (Kennedy *et al.,* 1986 and 1987). It has also been shown that ABA-1 can elicit an IgE response in its host, suggesting that it is an allergen which may contribute to the hypersensitivity reaction during ascariasis (Kennedy *et al.,* 1990). Interestingly, there appears to be a genetically restricted response to ABA-1 in experimental animals which has been shown to be determined by the host's major histocompatibility complex (MHC) haplotype (Kennedy *et al.,* 1991).

At the outset of this work, genes encoding homlogues of ABA-1 had been identified from the nematodes *Brugia pahangi* (Tweedie *et al.,* 1993) and *Dirofilaria immitis* (Culpepper *et al.,* 1992 and Poole *et al.,* 1992). This homology was based on the similarity of the N-terminal amino acid sequence of the native ABA-1 protein with the predicted amino acid sequence derived from the isolated *B. pahangi* and *D.immitis* cDNA clones. It was reported that these genes have an unusual structure in that they are composed of 396 bp tandem repeats in *B. pahangi* and 399 bp repeats in *D. immitis,* with each repeat encoding a polypeptide of approximately 15 kDa. These data suggested that the homologues are translated as polyproteins which are then cleaved into monomeric proteins. This chapter reports the cloning and characterisation of 3 cDNA clones, all of which consist of part of the *A.suum aba-1* gene. This was performed in order to determine if *aba-1* has the same structure as that of the homologues.

3.2 Results

3.2.1 Cloning strategy

In order to clone the cDNA copy of the gene which encodes ABA-1, a cDNA expression library derived from mRNA isolated from *A. suum* infective larvae was constructed in the lambda expression vector Lambda ZapII (a gift from J. Moore) and was screened with polyclonal rabbit antiserum raised to gel-excised ABA-1. Before the library was screened, the antiserum was absorbed with lambda lysate and *Escherichia coli* lysate to eliminate false positives due to antibody binding to λ–derived and *E.coli*-derived proteins rather than just to ABA-1. The titre of the library was approximately $4x10^8$ plaque forming units (pfu) per ml, and in order to ensure representation of ABA-1-encoding clones, 10^6 pfu were screened. For details of the antibody screening, see Materials and Methods.

The primary antibody screen resulted in the isolation of 50 antibody positive clones. Eighteen of these clones were replated and were screened a second time, and only twelve of them reacted with the antiserum on this round. On tertiary screening a further eight clones were found not to be recognised by the ABA-1 antiserum, leaving only four putative ABA-1 expressing clones. These clones were designated λHS1, λHS2, λHS3 and λHS4.

3.2.2 Restriction analysis of the antibody positive clones

To allow further analysis of the four λ clones isolated in the antibody screening, the cDNA inserts were excised from λ ZAPII vector in the form of self-replicating plasmids which were designated pHS10, pHS20, pHS30 and pHS40 (see Materials and Methods). The benefits of this

approach are that initial restriction analysis of the clones could be performed in plasmids rather than in lambda, which would have been more difficult for two reasons: (1) the size of the genome of this vector, and (2) potential difficulties in preparing lambda DNA. DNA preparations of pHS10, 20, 30 and 40 were made from a culture of *E. coli* strain XL1blue and restriction digestions were performed using *Eco*RI and *Xho*I. These enzymes cleave within the vectors' (pBluescript) polylinker site, and therefore allow the size of the cDNA inserts in each of the clones to be determined. This revealed that pHS10, pHS20, pHS30 and pHS40 contain inserts of 1.2 kbp, 0.6 kbp, 1.7 kbp and 0.70 kbp respectively (data not shown).

3.2.3 Partial sequencing of the antibody positive clones

In order to compare the inserts present in the plasmids pHS10, 20, 30 and 40, the 5' and 3' ends of the cDNAs were sequenced by the Sanger dideoxy method using oligonucleotide primers (named T3 and -40) complementary to the vector DNA around the insert (see Figure 3.1 and Materials and Methods). Translation of the 5' sequence derived from pHS10, pHS20 and pHS30 revealed that they all encode the ABA-1 polypeptide, since the putative amino acid sequences of the clones show 100% identity to native ABA-1 N-terminal sequence (see Figure 3.2). The first amino acid encoded by both pHS10 and pHS20 corresponds to amino acid 39 of the native ABA-1 N-terminal sequence, while the first amino acid encoded by clone pHS30 corresponds to amino acid 10 (see Figure 3.2). These data imply that the complete *aba-1* gene is not present within these cDNAs, as none of them appear to encode the first amino acids of the native ABA-1 protein. Neither the 5' nor the 3' sequence of clone pHS40

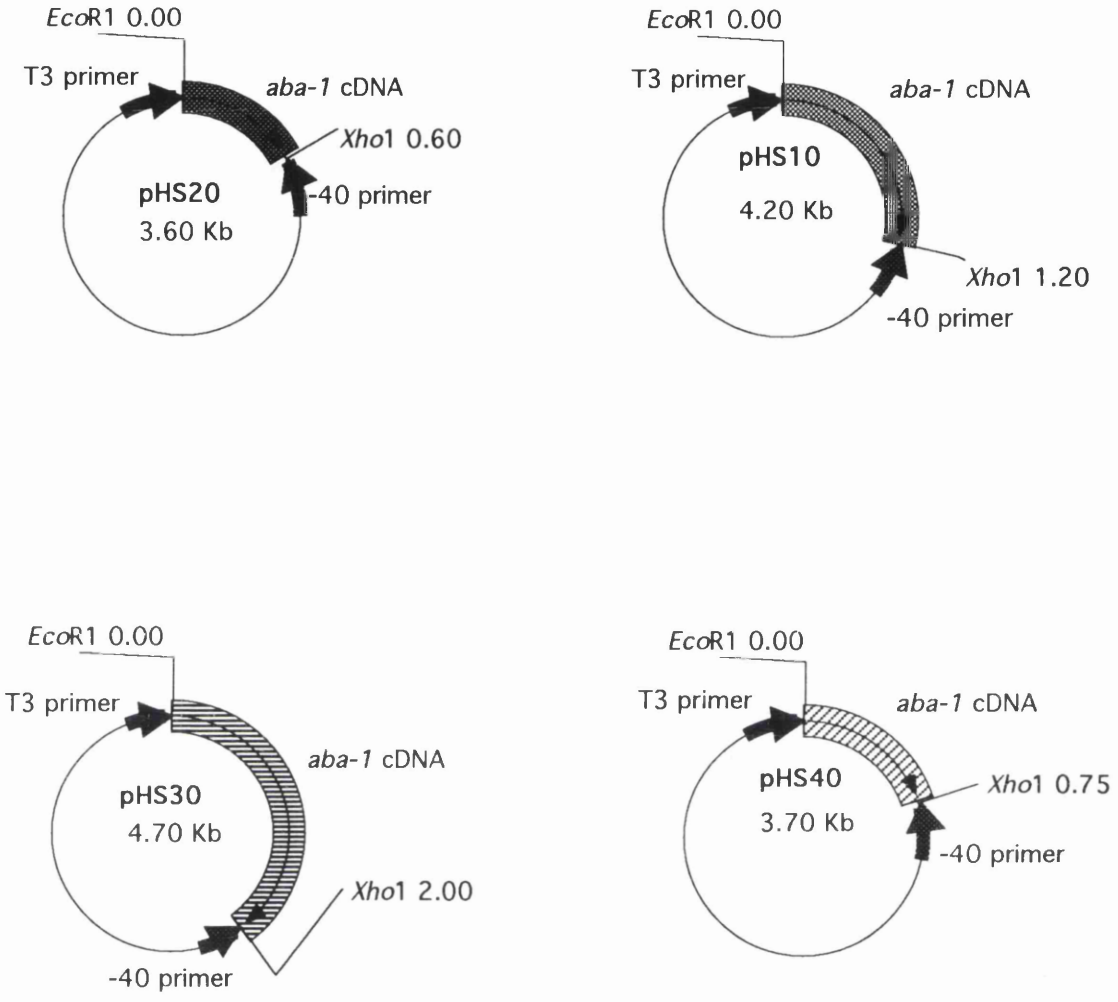


Figure 3.1. Diagramatic representation of the *aba-1*-containing cDNA clones pHS10, 20, 30 and 40. The hybridisation positions of the T3 and -40 primers, which were used to sequence the cDNAs inserts, are represented by black arrows. The sizes of the putative *aba-1* cDNA inserts within each plasmid are described in Section 3.2.2.

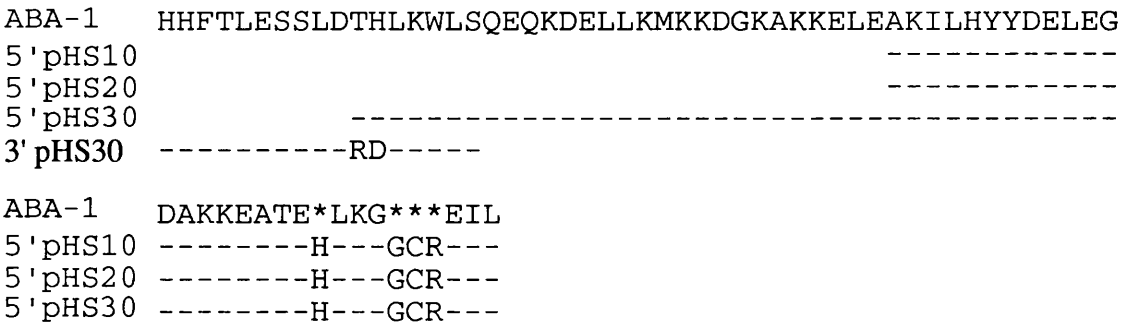


Figure 3.2. Comparison of the N-terminal amino acid sequence of ABA-1 with the putative amino acid sequences derived from the 5' regions of cDNA clones pHS10, 20 and 30, and the 3' region of pHS30. Identical amino acid residues are represented as (-), while nonidentical residues are shown by their single amino acid code. Positions of unknown amino acids in the ABA-1 N-terminal sequence(which is shown as ABA-1) are indicated by (*).

showed any sequence similarity to ABA-1, therefore suggesting that pHS40 is a false positive.

Data obtained by sequencing into the 3' region of both pHS10 and pHS20 showed that the sequences were identical, and suggest that the cDNAs contain a putative untranslated region, since the DNA sequence could not be translated to give an open reading frame any l onger than 4 amino acids. In contrast, the 3' sequence of pHS30 contained an open reading frame which shows 88% identity to the native amino acid sequence of ABA-1 (see Figure 3.2). These data suggest, therefore, that this clone does not contain the untranslated region of the *aba-1* gene and must have been internally primed at an A-rich sequence when the cDNA library was constructed. The differences in amino acid sequence between the N-terminal amino acid sequence and that derived from pHS30 are that amino acids 9 and 10 of ABA-1 are histidine and threonine whereas the corresponding residues of pHS30 are arginine and aspartic acid. Whether this has any significance, or is simply a cloning artefact, is not clear.

Analysis of the above data revealed that the *aba-1* gene of *A. suum* may be composed of repeating units like its homologues in *B. pahangi* and *D. immitis;* these genes comprise 396 bp and 399 bp tandem repeats respectively, with each repeat encoding a protein of approximately 15 kDa. The first reason for suggesting this is that, although the putative amino acid sequence encoded by clones pHS10 and pHS20 both appear to start 39 amino acids from the N-terminus of the native protein, clone pHS10 is 700 bp larger. This is surprising, since it would be expected that the clones would be the same size if they both were transcribed from the same gene. One explanation for the differences in size is that clone pHS10 contains one truncated repeat followed by complete repeats, while clone pHS20 is composed of only one truncated repeat. Alternatively, the untranslated region of clone pHS10 might simply be larger than that of pHS20. If the

latter explanation was the reason, it would imply that there may be more than one copy of the *aba-1* gene in the genome of *A. suum*, or perhaps that there are differences within the untranslated regions of individual parasites.

A second reason for suggesting that the *aba-1* gene may be composed of repeating units was that clone pHS30 encodes a protein that would appear to start at amino acid 10 and ends at amino acid 17, yet the clone is 2 kb in size. This may be interpreted as suggesting that this clone contains many complete repeats and that it starts and ends with truncated repeats.

These interpretations of this preliminary sequencing data for pHS10, 20 and 30 are cartooned in Figure 3.3, and assume that each encodes a 14 kDa protein from an approximately 400 bp gene. Section 3.2.4 addresses whether this is correct.

3.2.4 Analysis of clone pHS10

To investigate further whether the *aba-1* gene is composed of repeating units, the entire sequence of the pHS10 1.2 kbp insert was sequenced on both the coding and noncoding strands. Clone pHS10 was chosen for this analysis because it is larger than pHS20 and because it contains the untranslated region which is not present in clone pHS30. Due to the large size of this clone it was necessary to make Exonuclease III deletions into both the 5' and 3' regions to allow complete sequencing of the insert (for details of how the Exonuclease III deletions were carried out, see Materials and Methods). Figure 3.4 shows the three deletion clones (designated pHS11, 12 and 13) selected to determine the DNA sequence of the coding strand of pHS10, and the four deletions selected for analysis of the noncoding strand (designated pHS14, 15, 16 and 17). All of these

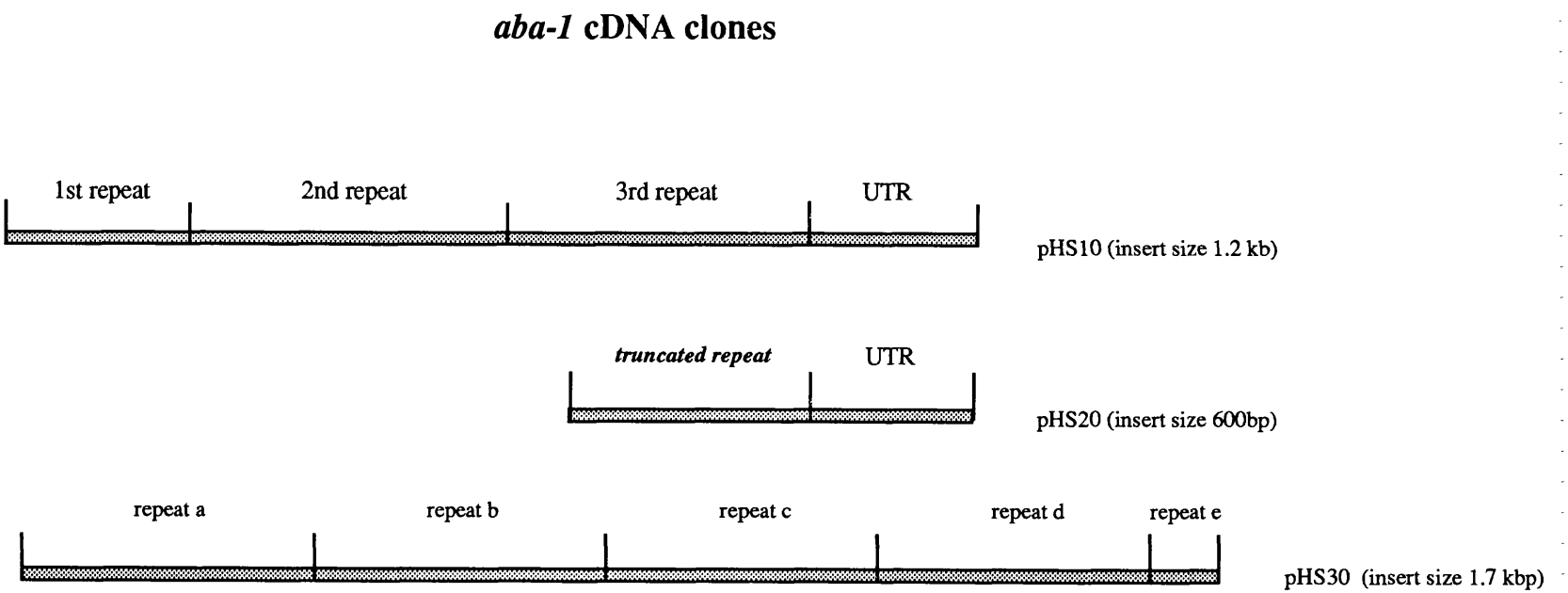


Figure 3.3. Diagrammatic representaton of the *aba-1* cDNA clones pHS10, 20 and 30. The copies of the 399 bp tandem repeats encoding ABA-1 (or part of ABA-1) are numbered 1, 2 and 3 in pHS10, and a, b, c, d and e in pHS30. The 3' untranslated regions in clones pHS10 and pHS20 are shown by UTR.

Figure 3.4. Diagrammatic representation of the exonuclease III deletion derivatives of clone pHS10. The size of the cDNA inserts remaining in the various clones are shown by a solid black line which corresponds to the pHS10 insert above. pHS11, 12 and 13 were derived by progressive deletions from a *Sac*I site, while pHS14, 15, 16 and 17 are derived by deletions from a *Kpn*I site.

deletion clones were sequenced using the relevant oligonucleotide primer; coding strand deletions used the T3 primer and noncoding strand the -40 primer (see Section 3.2.3).

The DNA sequence from clones pHS11, 12, 13, 14, 15, 16 and 17 were assembled as depicted in Figure 3.5. To eliminate any errors arising through the reading of the sequencing gels, and to determine if there were any inconsistencies between the sequence derived from the coding and non-coding strands, the BESTFIT programme (Devereux *et al.*, 1984) was used. In using this programme, the sequencing gels were read independently 5 times and entered into files, which were then compared.

Analysis of the assembled sequence of the pHS10 cDNA insert revealed that it has a total size of 1251 bp, comprising a 1185 bp open reading frame followed by a 66 bp untranslated region (both are discussed below). Two complete repeats of 399 bp, starting at positions 286 and 685 within the sequence (indicated in Figure 3.5 by an arrows), are apparent in addition to an incomplete repeat which starts at the first bp and is 285 bp in size. The incomplete repeat is lacking the first 114 bp seen at the 5' end of the two other repeats, and consequently its sequence begins at position 115 with respect to a complete 399 bp repeat. Similarity between the repeats is nearly 100%, except for 3 bp differences between the first and the last repeats at positions 81, 222 and 234, and one bp difference between repeats 2 and 3 at positions 480; note that positions 81 and 480 are identical between the first and second repeats. These data confirm that the *aba-1* gene of *A. suum* is composed of tandem repeats as seen for *B. pahangi* and *D. immitis*.

Translation of the DNA sequence of the pHS10 insert revealed a single continuous open reading frame extending for 1185 bp which generates a 375 amino acid polypeptide. Following this open reading frame is a 3' untranslated region. The large open reading frame suggests that

Figure 3.5.The DNA sequence and the putative amino acid sequence of the pHS10 insert. The junction between repeated units is indicated by arrows. The carboxy terminal region is underlined and the stop codon is indicated by an asterisk. Lower case letters indicate the putative 3' untranslated region, within which a putative polyadenylation signal is underlined. Inverted open triangles indicate base differences between the first and second repeats and the most 3' repeat. Shown at the bottom of the sequence is a cartoon diagram of clone pHS10.

```
                                                                  ▽
GCGAAAATTCTTCATTACTATGATGAACTCGAAGGAGATGCTAAAAAAGAGGCAACTGAGCACTTGAAAGGCGGATGCCGAGAAATTCTT 90
A  K  I  L  H  Y  Y  D  E  L  E  G  D  A  K  K  E  A  T  E  H  L  K  G  G  C  R  E  I  L

AAGCATGTTGTTGGGGAAGAGAAGGCAGCAGAGCTGAAGAACCTCAAAGACTCGGGAGCAAGCAAAGAAGAACTCAAAGCCAAAGTCGAA 180
K  H  V  V  G  E  E  K  A  A  E  L  K  N  L  K  D  S  G  A  S  K  E  E  L  K  A  K  V  E

                                 ▽           ▽
GAGGCGCTTCATGCAGTGACCGACGAGGAGAAGAAGCAATATATCGCCGATTTCGGACCAGCATGCAAGAAAATCTATGGTGTACATACT 270
E  A  L  H  A  V  T  D  E  E  K  K  Q  Y  I  A  D  F  G  P  A  C  K  K  I  Y  G  V  H  T

TCGCGACGAAGGAGGCATCATTTCACCCTTGAAAGTAGTCTAGATACCCATCTGAAATGGCTCAGTCAAGAACAGAAAGATGAATTGTTA 360
S  R  R  R  R  H  H  H  F  T  L  E  S  S  L  D  T  H  L  K  W  L  S  Q  E  Q  K  D  E  L  L

AAAATGAAGAAAGATGGAAAGGCAAAGAAAGAACTCGAAGCGAAAATTCTTCATTACTATGATGAACTCGAAGGAGATGCTAAAAAAGAG 450
K  M  K  K  D  G  K  A  K  K  E  L  E  A  K  I  L  H  Y  Y  D  E  L  E  G  D  A  K  K  E

                 ▽
GCAACTGAGCACTTGAAAGGCGGATGCCGAGAAATTCTTAAGCATGTTGTTGGGGAAGAGAAGGCAGCAGAGCTGAAGAACCTCAAAGAC 540
A  T  E  H  L  K  G  G  C  R  E  I  L  K  H  V  V  G  E  E  K  A  A  E  L  K  N  L  K  D

TCGGGAGCAAGCAAAGAAGAACTCAAAGCCAAAGTCGAAGAGGCGCTTCATGCAGTGACCGACGAGGAGAAGAAGCAATACATCGCCGAT 630
S  G  A  S  K  E  E  L  K  A  K  V  E  E  A  L  H  A  V  T  D  E  E  K  K  Q  Y  I  A  D

TTTGGACCAGCATGCAAGAAAATCTATGGTGTACATACTTCGCGACGAAGGAGGCATCATTTCACCCTTGAAAGTAGTCTAGATACCCAT 720
F  G  P  A  C  K  K  I  Y  G  V  H  T  S  R  R  R  R  H  H  H  F  T  L  E  S  S  L  D  T  H

CTGAAATGGCTCAGTCAAGAACAGAAAGATGAATTGTTAAAAATGAAGAAAGATGGAAAGGCAAAGAAAGAACTCGAAGCGAAAATTCTT 810
L  K  W  L  S  Q  E  Q  K  D  E  L  L  K  M  K  K  D  G  K  A  K  K  E  L  E  A  K  I  L

CATTACTATGATGAACTCGAAGGAGATGCTAAAAAAGAGGCAACTGAGCACTTGAAAGGCGGATGCCGCGAAATTCTTAAGCATGTTGTT 900
H  Y  Y  D  E  L  E  G  D  A  K  K  E  A  T  E  H  L  K  G  G  C  R  E  I  L  K  H  V  V

GGGGAAGAGAAGGCAGCAGAGCTGAAGAACCTCAAAGACTCGGGAGCAAGCAAAGAAGAACTCAAAGCCAAAGTCGAAGAGGCGCTTCAT 990
G  E  E  K  A  A  E  L  K  N  L  K  D  S  G  A  S  K  E  E  L  K  A  K  V  E  E  A  L  H

GCAGTGACCGACGAGGAGAAGAAGCAATACATCGCCGATTTTGGACCAGCATGCAAGAAAATCTATGGTGTACATACTTCGCGACGAAGG 1080
A  V  T  D  E  E  K  K  Q  Y  I  A  D  F  G  P  A  C  K  K  I  Y  G  V  H  T  S  R  R  R

AGGTATCACGCTGAGGATGGAACGGATGACATTGACGGGCTAGCGCAAAGTCGGCAGCGACGATCCGGCTTTTTCGAGAAGCTCATCGAT 1170
R  Y  H  A  E  D  G  T  D  D  I  D  G  L  A  Q  S  R  Q  R  R  S  G  F  F  E  K  L  I  D

GTGTTTGCATTTTTTTGAgtgaaattcccatgttacgcgaatcgttatgtcatatcaacagatgatagtaaataatcacaaccac      1255
V  F  A  F  F  *
```

pHS10 insert

pHS10 encodes a protein of approximately 40 kDa, despite the fact that ABA-1 is thought to be only 14 kDa in size (Christie *et al.*, 1990). Taken together with the fact that the *aba-1* cDNAs contain 339 bp repeats uninterrupted by stop codons, these data can be interpreted as suggesting that ABA-1 is initially translated from its mRNA as a polyprotein which is subsequently cleaved into 14 kDa monomers. Again this is in agreement with data from *B. pahangi* and *D. immitis*, and is considered further below.

The putative amino acid sequence derived from each repeat within pHS10 shows 100% identity to the derived N-terminal amino acid sequence of ABA-1 (Christie *et al.*, 1990). In addition, none of the DNA sequence differences between the repeats results in differences in amino acid sequence. The 5'-most repeat (which is, as stated, incomplete) starts at amino acid 39 of the native ABA-1 protein (see Section 3.2.3), while repeats 2 and 3 appear to encode the entire ABA-1 polypeptide. These data therefore provide the entire amino acid sequence of *A.suum* ABA-1 which was previously unavailable from direct protein sequencing.

All of the above sequence analysis of pHS10 gives further evidence that the *aba-1* gene is composed of repeating units of 399 bp, with each unit encoding a protein of approximately 14 kDa. Therefore it is apparent that the *aba-1* gene has a similar structure to it homologues in *B. pahangi* and *D. immitis*.

3.2.5 Analysis of putative cleavage sites within the ABA-1 polypeptide sequence

With the above hypothesis that the ABA-1 protein may be translated as a polyprotein and then cleaved into 14 kDa monomers, it

became apparent that either an enzyme must be utilised to cleave the polyprotein or that it contains some undefined self-cleaving activity. Since

ly knew the N-terminal amino acid sequence of the native ABA-1 Christie *et al.*, 1990), it was possible to predict that the cleavage of peptide would occur at the residues before the extreme N-terminal acid of the ABA-1 monomers, corresponding to the C-terminal acids encoded by each repeat. Analysis of this site showed that the inus of each repeated monomer had a run of 4 arginines, iating at positions 285, 684, and 1083 in the DNA sequence (see e 3.5). Runs of four arginines are reminiscent of the consensus ige sites for subtilisin family endoproteinases (Barr, 1991), and we fore believe this may be the means by which ABA-1 is processed. For ivage site to be recognised by a subtilisin-like endoproteinase it must mposed of at least a pair of basic amino acids; this is termed a "dibasic

The four arginines within the ABA-1 sequence has four basic ues and thus falls into the category of a tetrabasic site. These data ide further evidence that ABA-1 may be translated as a polyprotein ch is cleaved into 14 kDa monomers. It must be assumed, however, these enzymes cut on this sequence within *A.suum.*

Comparison of the predicted molecular weight of the ABA-1 ecule encoded by a single 399 bp repeat within pHS10 to the weight of native protein reveals a discrepancy which may be related to this cess. The molecular weight of the native ABA-1 protein has been ermined by mass spectrometery (Christie *et al.*, 1993) and by SDS-acrylamide gel electrophoresis (Kennedy, 1986 and 1987) to be 14.6 kDa 14 kDa respectively, but the predicted weight of the polypeptide oded by one repeat of pHS10 was 15.3 kDa. An explanation for this discrepancy is that when the putative endoproteinase cleaves the ABA-1 polyprotein at the cleavage site it removes the four arginines, or

alternatively another enzyme with exproteinase activity is involved in removing these residues. Either of these processes would result in one monomer's molecular weight being roughly the same as that predicted by mass spectrometery and SDS-PAGE. It is currently not possible to determine if this is the case, however, until the C-terminal sequence of the native protein has been determined.

3.2.6 Analysis of the carboxy-terminal region of clone pHS10

After the last repeat of clone pHS10, at position 1084 within the DNA sequence, there is no putative stop codon, but instead there is a further open reading frame of 34 amino acids (see Figure 3.5). It was originally thought that this open reading frame represents a truncated or rearranged 399 bp repeat, but the DNA and translated amino acid sequence of the region shows no homology to the other repeats. In addition, analysis of the amino acid composition of the open reading frame by Doolittle plots (Devereux *et al.*, 1984) revealed that it is substantially more hydrophobic than any sequence found within ABA-1. For these reasons, I have termed the sequence between position 1084 and 1189 of pHS10 the "carboxy-terminal region". It is interesting to note that a consensus cleavage site for subtilisin-like endoproteinases (sequence RQRR) is present in the centre of this region, between positions 1152 and 1163. Whether this is of functional significance, or whether it is an artefact, is not clear.

Identification of this region raised the following question: why is this region present; i.e. does it have a function as a 3' signal sequence which is involved in the processing of the polyprotein in the cell, or is it simply an artefact which arose during cloning as a consequence of the

unusual repeating structure of the gene? To try and answer these questions, the first approach used was to compare the amino acid sequence that this region encodes with sequences in the SwissProt FASTA data base (Devereux *et al.*, 1984). It was hoped that this region may contain a common motif that is found in other 3' signals, but no such similarities were found.

Despite the fact that no homologues of ABA-1's carboxy terminal region could be identified, it is intriguing to note that the recently cloned cDNAs encoding homologues of ABA-1 in *Dictyocaulus viviparus* (C. Britton, personal communication) and in *Caenorhabditis elegans* (J.Moore *et al.*, 1993) both contain carboxy-terminal polypeptides (10 and 9 amino acids in size respectively) following their most 3' repeats. Like the *aba-1* carboxy-terminal region, these show no similarity to the other repeats present within the isolated cDNAs. However, it should be stressed that neither of these regions display any similarity in either size, DNA sequence or translated amino acid composition to the *aba-1* carboxy-terminal region. The presence of these carboxy-terminal regions in proteins related to ABA-1 suggests that this region is unlikely to be an artefact, but the function of the region remains unknown.

3.2.7 Analysis of the untranslated region of clone pHS10

Following the carboxy-terminal region of clone pHS10 there is a putative stop codon (UGA; at position 1196 to 1198) followed by a 66 bp untranslated region which terminates in a poly A tract that is presumably derived from the poly A tail of the *aba-1* mRNA (see Figure 3.5). Although the common polyadenylation signal AAUAAA that is utilised by most organisms (Sheets *et al.*, 1991) has been identified in other *A. suum* genes

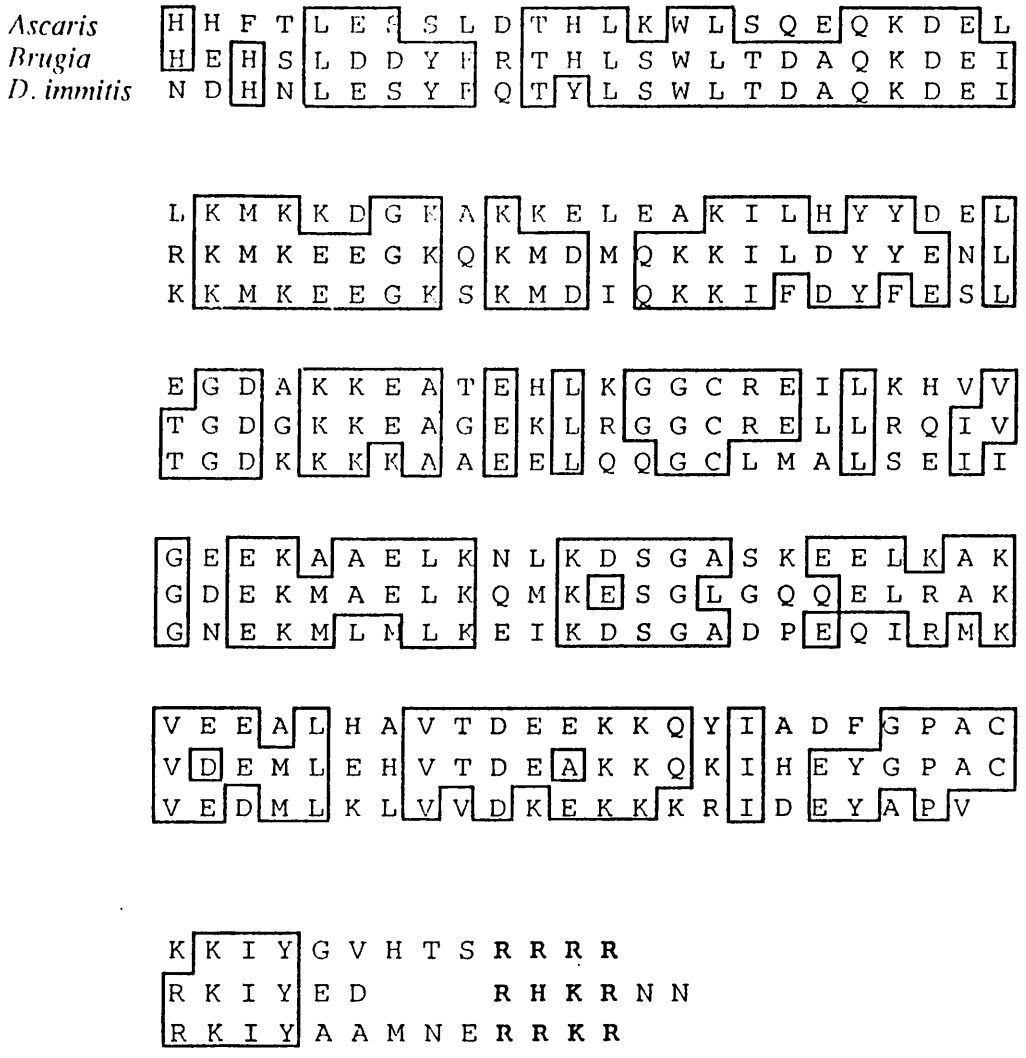
(Kingston *et al.*, 1990 and Wheellock *et al.*, 1991), it is not present in pHS10. There is, however, the sequence AGUAAA situated 13 bp upstream of the putative poly A tail. This sequence has been shown by Sheets *et al.* (1991) to act as a polyadenylation signal *in vitro*. Using site-directed mutagenesis, these workers changed AAUAAA to the following: UAUAAA, GAUAAA, CAUAAA, AUUAAA, AGUAAA or ACUAAA. Since all these derivatives allowed polyadenylation to occur efficiently, it can be concluded that the AGUAAA sequence present in the *aba-1* gene is probably a functional polyadenyalation signal sequence.

3.2.8 Homologues of ABA-1

As discussed in Chapter 1 and in Section 3.1, a number of homologues of *A. suum* ABA-1 have been cloned from different nematode species while this work was in progress. The first homologous genes cloned were those which encode the *B. pahangi* gp15/400 antigen (Tweedie *et al.*, 1993), and a *D. immitis* 15kDa antigen, named Di5 or Di22 (Poole *et al.*, 1992 and Culpepper *et al.*, 1992). Like the *aba-1* gene, the genes which encode these homologues are composed of tandem repeats; 396 bp in size in the case of *B. pahangi* and 399 bp in *D. immitis*. There is also evidence for these proteins being translated as polyproteins which are then cleaved into approximately 15 kDa monomers (Tweedie *et al.*, 1993; Poole *et al.*, 1992; Culpepper *et al.*, 1992).

A comparison of the amino acid sequences of the above homologues to ABA-1 is illustrated in Figure 3.6, with the conserved regions boxed. The percentage identity between the species was analysed using the BESTFIT programme, and was found to be an average of 45%. Since values greater than 30% are considered significant this value is

Figure 3.6. Comparison of the putative amino acid sequence derived from an *aba-1* repeat of pHS10 with the sequences of the *B. pahangi* and *D. immitis* ABA-1 homologues. Conserved regions between the three homologues are boxed.


Ascaris
Brugia
D. immitis

relatively high and confirms that these proteins are probably evolutionary related.

Whereas, for ABA-1, the putative subtilisin endoproteinase cleavage site (see Section 3.2.5) has the sequence RRRR, it is RRKR for the *D. immitis* homologue and RHKR for *B.pahangi* (underlined in Figure 3.6). Because all the homologous proteins appear to contain cleavage sites which satisfy the requirements for being subtilisin endoproteinase recognition sequences, this provides further support for the idea that ABA-1 polyprotein may indeed be processed by this mechanism.

Based on these sequence comparisons, secondary structure predictions for ABA-1 and its homologues have led to the suggestion that ABA-1 comprises a four bundle alpha helix structure (Kennedy *et al.*, 1993). Circular dichroism analysis of purified ABA-1 has confirmed the predicted high helical content of the protein (Kennedy *et al.*, 1993). It should be noted, however, that this predicted structure awaits confirmation through X-ray crystallography.

Despite the above knowledge of the structure of the genes which encode these these homologues, and the amino acid composition of the proteins they encode, the biological function of these molecules is still unknown. In an attempt to provide an insight into the biological function, the translated amino acid sequence of a single *A. suum* repeat was put through the SwissProt database, and the DNA sequence was put through the GENEMBL database. Unfortunately no similarities were found other than to the above-mentioned homologues.

3.3 Discussion

This chapter reports the cloning and characterisation of three cDNAs which encode the *A. suum* allergen, ABA-1. This was achieved by screening a cDNA expression library derived from mRNA isolated from *A.suum* infective larvae with polyclonal rabbit antiserum to gel-excised ABA-1. Partial sequencing of the three antibody positive clones, designated pHS10, pHS20 and pHS30, suggested that the *aba-1* gene is composed of tandem repeats with each repeat encoding an approximately 15 kDa polypeptide. Sequencing the entire cDNA insert of pHS10 confirmed this, since the clone contains two complete repeats of 399 bp and one incomplete repeat of 285 bp. Translation of the DNA sequence of each repeat revealed 100 % identity to the N-terminal amino acid sequence of the native ABA-1 protein (Christie *et al.*, 1990) and complete identity with each other.

The repeats within pHS10 are not interrupted by stop codons, which suggests that the clone contains a 1185 bp open reading frame. This was interpreted as meaning that ABA-1 is initially translated as a polyprotein which is subsequently cleaved into 15 kDa monomers. Additional evidence for this hypothesis is that the C-terminal four amino acids encoded by each 399 bp repeat are all arginines, which conforms to the consensus cleavage site for subtilisin-like endoproteinases. These cleavage sites are termed "tetrabasic sites", as they are composed of two pairs of basic amino acids. The most commonly used site of this type is K/R.X.K/R.R (Barr, 1991). This site has been shown through *in vitro* experiments to be utilised by the prohormone beta-nerve growth factor (Barr, 1991) and by the proprotein insulin proreceptor (Barr, 1991).

As yet, identification of subtilisin-like cleavage sites within the *A. suum* ABA-1 sequence remains a prediction, since there is no direct

experimental evidence that they are functional. It may be possible, for example, to use site-directed mutagenesis to alter the RRRR. sequence to RRDR., thus introducing an acidic residue at position 3 rather than the essential basic residue. After purification, this and other altered ABA-1 polyproteins could be analysed and compared to the native ABA-1 in *in vitro* cleavage assay using subtilisin-like endoproteinases (many of these types of enzymes have been cloned and over expressed in *E. coli*; Barr *et al.*, 1991). Such experiments would confirm the functionality of these sites, and may give an indication as to the type of endoproteinases which act on ABA-1.

Sequencing of clone pHS10 revealed that the repeats present within this cDNA showed 98% identity to one another at the DNA level, and that their translated amino acid sequence showed 100% identity. These data were originally interpreted as meaning that all the repeats present in the *aba-1* gene encode identical polypeptides. Recent data, however, have suggested that this may not be the case (J. Moore, personal communication). Further cDNA clones expressing ABA-1 were isolated from the same λZAPII cDNA expression library employed in this work, using polyclonal antiserum raised to the *A. suum* L2 larvae stage excretory products. Sequencing of these clones revealed that, like clones pHS10, 20 and 30, they are composed of repeating units of 399bp each encoding a polypeptide of 15 kDa. All but one of the polypeptides encoded by the repeating units found in these clones show 100% identity to the N-terminal amino acid sequence of the native ABA-1 protein and to the complete ABA-1 sequence derived from pHS10. A single repeat, present in clone pJM33 and termed the "divergent repeat", shows only 49% identity to the amino acid sequences encoded by all the other repeated units described. Despite these differences in sequence, it should be noted that the polypeptide encoded by this repeat retains the four C-terminal arginines

constituting the putative cleavage site and two cysteine residues thought to be important in the folding of the protein (J. Moore, personal communication). In addition, the predicted secondary structure for this divergent polypeptide is a 4 alpha helical bundle, as described for the ABA-1 polypeptides (Kennedy *et al.*, 1993). These data suggest that the divergent repeat of pJM33 is not an artifact resulting from the cDNA cloning.

Identification of this divergent repeat raises a number of questions. For example, is this repeat derived from the same *aba-1* gene as the repeats present within cDNA clones pHS10, pHS20 and pHS30, and is the polypeptide encoded by this repeat a derivative of ABA-1 with the same (or similar) function, or is it unrelated to ABA-1 and has a different function? Attempts to address these questions are reported in Chapter 5.

It is now apparent that most, if not all, nematodes contain a homologue to the ABA-1 protein (see Chapter 1). The term "nematode polyprotein allergens" (NPA) has recently been coined to describe this collection of homologues (Reynolds, 1993); this is a consequence of the fact they all appear to be translated initially as polyproteins which are then cleaved into monomers, and that the *A. suum*, *A. lumbricoides*, *B. malayi* and *D. immitis* homologues all appear to elicit an IgE response in their respective hosts. Very recently, however, an NPA homologue has been identified in the free-living nematode *C. elegans*, which, being derived from a non-parasitic orgamism, means that the term allergen would seem redundant (Moore *et al.*, 1993). Despite this, the term NPA will be used here since the function of these proteins is unclear as yet. It has been shown that the secondary structure predicted for all of the NPAs is that of a four alpha helical bundles (Kennedy *et al.*, 1993). Examples of molecules which have such a structure are cytochrome and myohemerthrin, both of which are carrier proteins (Branden and Tooze, 1991). It is therefore

conceivable that ABA-1 is also a carrier protein (this is considered further in Chapter 4 and Chapter 7).

Although no other types of proteins display any sequence similarity to ABA-1 and other NPAs, a few proteins are encoded by genes with similar organisation. Two examples are the filaggrin proteins from mouse and human (Grant *et al.*, 1989; Rothnagel *et al.*, 1987) and the *Trypanosoma brucei* proteins involved in capping microtubules (Rindisbucher *et al.*, 1993). The gene which encodes the mouse filaggrin is composed of approximately twenty 744 bp repeats and, like ABA-1, initially expresses the protein as a large polyprotein that is cleaved into monomers. As monomers, filaggrin proteins aggregate with keratin intermediate filaments and promote disulpide bond formation. The *T. brucei* microtubule capping protein is encoded by a gene which is composed of approximately thirty repeats of 600 bp and is also expressed as a polyprotein. These similarities in the gene organisation and protein expression may suggest that the nematode NPAs are structural proteins involved in intracellular or extracellular assemblages. In relation to this, the *B.pahangi* and *D. immitis* NPAs have been detected on the cuticles of the parasites, where it may be envisaged that they could interact with structural proteins. One problem, however, is that *A.suum* ABA-1 is present in the parasite's body fluid. An alternative explanation for these similarities in gene organisation could be that these are all examples of proteins that are required in large amounts in the respective organisms, and tandem repeat gene structure represents a common mechanism to achieve this high level of expression.

Some experimental approaches to determine the function of ABA-1 are reported in Chapter 4, but in the long term it is apparent that the best approach will be to use the recently identified *C. elegans* homologue (Moore *et al.*, 1994). In this organism it is possible to perform reverse

genetics and replace the wild type *aba-1* homologue with a mutant gene. Such an approach would allow a function to be determined through the phenotype exhibited by the mutant strains, although such mutants could not be obtained if the protein is essential for nematode life. Using the cloned *C. elegans* gene it should also be possible to analyse the tissue in which the protein is expressed by microinjecting plasmids containing the promoter region of the gene linked to an easily assayable protein such as Beta-galactosidase. Again, such an approach, which is not yet possible in *Ascaris,* could allow information to be derived regarding the NPA's function.

An interesting feature of the pHS10 clone is that no putative stop codon follows the most 3' *aba-1* repeat, but instead there is a further open reading frame of 34 amino acids in length. This open reading frame has been termed the carboxy-terminal region. As stated in the results, it is not known if this region has a function, or it is simply an artefact that has arisen during the cDNA cloning because of the unusual repeating structure of the *aba-1* gene. It was speculated that the region may act as a 3' signal peptide that is involved in intracellular targetting or processing. An alternative function could perhaps be as a sequence to stabilise the large polyprotein from degradation, or to aid folding. Since no homologues of this region were found in the protein data base, and because mutagenesis experiments to analyse its function are not yet possible in *Ascaris,* these possibilities remain speculations. However, the NPAs from *C.elegans* and *D. vivipourous* also contain a carboxy-terminal region after their most 3' repeat. Although these regions show no obvious similarity, it should be noted that other proteins that utilise 3' signals also show variation in size and lack sequence identity; examples are the proteases (Mottram *et al.,* 1989) and Variant Surface Antigens of Trypanosomes (Ferguson and Williams, 1988). Perhaps when more 3' regions have been isolated from

other NPA proteins a comparative sequence analysis of all known carboxy-terminal regions may be made and it may then be possible to determine a consensus motif. It is also conceivable that when further, non-NPA encoding nematode genes are cloned it might be seen that 3' signal sequences are ubiquitous. If this were the case, then the roles of the different signal sequences may then be determined.

Chapter 4

Analysis of the stage- and tissue-specificity of *aba-1* expression

4.1 Introduction

Chapter 3 reports the isolation and characterisation of three cDNAs which encode the *A. suum* allergen, ABA-1. Sequencing data from the cDNAs implied that ABA-1 is encoded by a gene which has an unusual structure, being composed of 399bp repeats that each encode an approximately 15 kDa protein. Consequently it has been suggested that ABA-1 is translated as a polyprotein, which is then cleaved into approximately 15 kDa polypeptide units. Despite this insight into the structure of the *aba-1* gene it is not known exactly how many repeats are present within the gene, because the cDNAs are incomplete and the 5' region is not represented.

This chapter reports Western and Northern blot analyses that were performed to permit an estimate of the size of the native *A.suum* ABA-1 polyprotein and the size of the *aba-1* mRNA transcript respectively. Similar approaches have been used to determine the number of repeats present within the genes that encode the *B. pahangi* and *D. immitis* homologues of ABA-1 (Tweedie *et al.*, 1993; Culpepper *et al.*, 1992; and Poole *et al.*, 1992; see also Chapter 3). Northern and Western blot analyses were also used to determine whether the *aba-1* gene of *A. suum* is expressed in a stage-specific manner. This is important as it is not known whether the protein is only needed when the parasite is within the host, or whether it is required throughout the parasite's life cycle.

Another question that is addressed in this chapter is in which tissues of the adult parasite *aba-1* is expressed. This analysis is possible in *A. suum* due to the large size of the adults, which allows easy dissection of different tissues. A similar analysis was performed by Kingston *et al.* (1989)

to determine the site of expression of two collegan genes. It was thought that this analysis might give an insight as to the function of ABA-1.

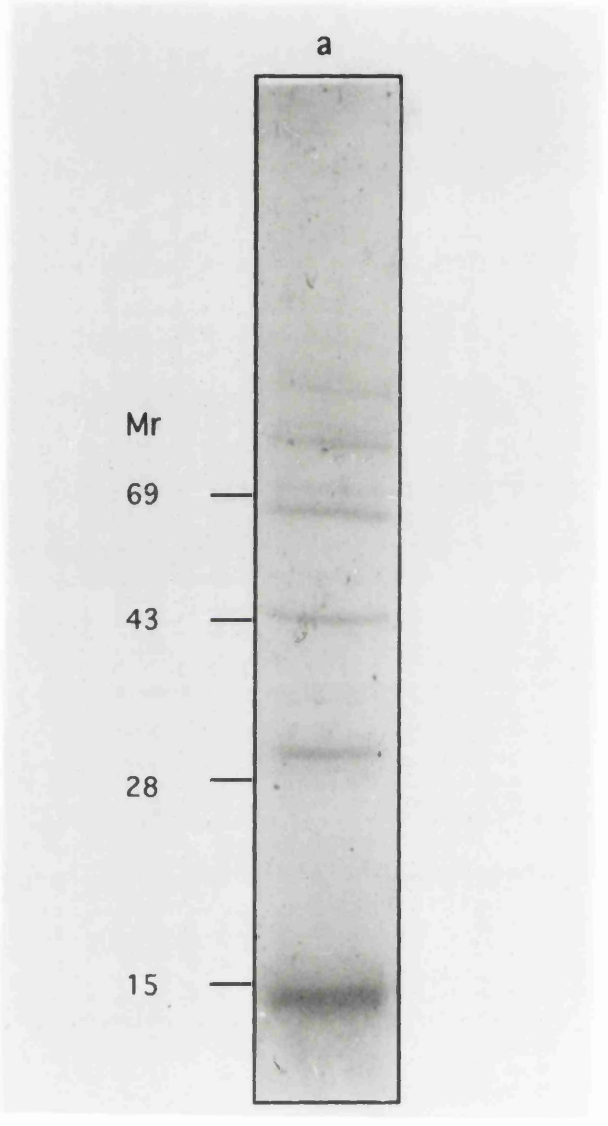
4.2 Results

4.2.1 Western blot analysis of the native ABA-1 polyprotein

The approach that was adopted to determine whether the native ABA-1 protein is found as a polyprotein in the worm was to analyse its structure in the adult *Ascaris* body fluid (ABF), where it is available in large amounts (Christie *et al.*, 1990). ABF was isolated from adult worms and a Bradford assay was carried out to determine the protein concentration. Approximately 500 µg of ABF was used for the Western blot (see Materials and Methods for details of Western blots) shown in Figure 4.1, which was probed with polyclonal rabbit antiserum raised to the recombinant ABA-1 protein (see J. Moore *et al.*, 1993 for a description of the recombinant protein and the antiserum). Polyclonal antiserum to the recombinant protein was used in preference to antiserum to gel-excised ABA-1, because it should contain no antibodies against other parasite antigens which are present in the gel-purified preparation. A control was performed by probing the blot with normal rabbit serum, which did not recognise any proteins from the *A. suum* ABF (data not shown).

Analysis of this Western blot revealed that the ABA-1 antiserum does not only recognise a 14.4 kDa protein (the known size of the ABA-1 protein; Kennedy *et al.*, 1986), but it also recognises bands of approximately a 30 kDa, 45 kDa, 60 kDa, 75 kDa, 90 kDa, 105 kDa and 120 kDa (see Figure 4.1). One interpretation of this is that these bands represent proteins containing 2, 3, 4, 5, 6, 7 and 8 units of ABA-1 respectively, which are the

Figure 4.1. Western blot of the adult body fluid of *A. suum* probed with polyclonal rabbit antiserum raised to the recombinant ABA-1 protein. Protein was electrophoresed under reducing conditions on a 5-25% SDS gradient polyacrylamide gel. The positions of molecular size markers (in kDa) run on the same gel are indicated to the left


a
Mr
69
43
28
15

products of incomplete or partial cleavage of the ABA-1 polyprotein. Another explanation is that the bands are a result of non-specific binding of the antibody; the fact that the control blot was completely clear suggests this is not the case, however. Because the gel in Figure 4.1 contained SDS and beta-mercaptoethanol, the possibility that the higher forms of ABA-1 present may have arisen through aggregation can be considered unlikely.

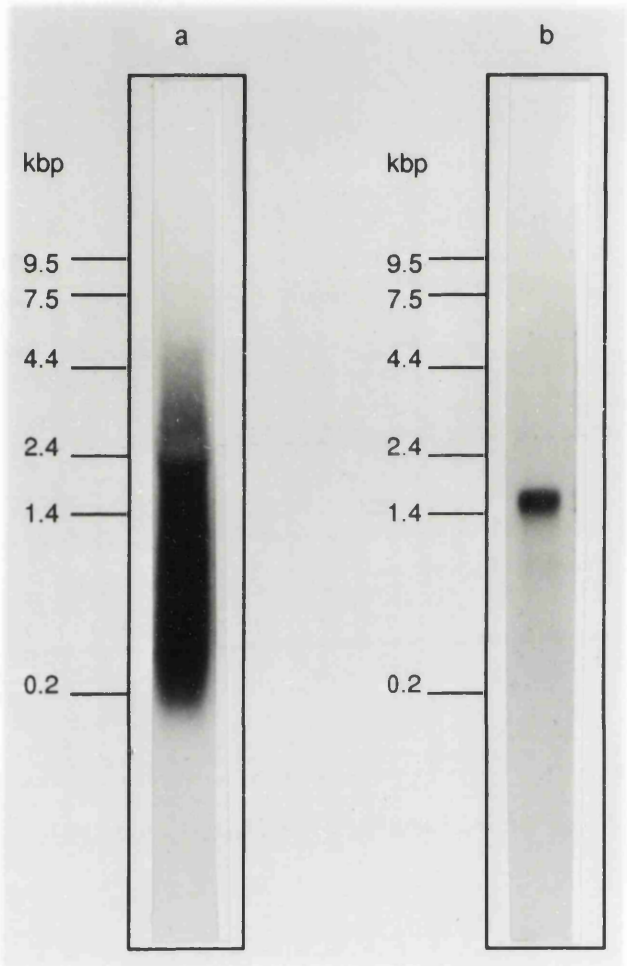
This result provides evidence for the ABA-1 protein being initially translated as a polyprotein and then being cleaved into monomers. As such, it provides conformation that the repeat structure of the *aba-1* gene discussed in Chapter 3 is not an artefact that occured during the synthesis or replication of the cDNAs. It would appear from the data that there are at least eight copies of the 399 bp repeat present in the *aba-1* gene, since a protein of 120 kDa is detected which corresponds to eight times the size of a single 15 kDa monomer (see below). Note that it is possible that larger ABA-1 polyproteins may be present, but undetected in this preparation, and the gene may therefore contain more than 8 repeats.

An intruiging feature of this Western analysis is that the final product of cleavage of the polyprotein is approximately 14.4 kDa, while the putative amino acid sequence derived from the cDNA clones which encode ABA-1 predicts a protein of approximately 15 kDa. One explanation for this is that following the putativeendoproteinase cleavage of the ABA-1 polyprotein an additional process removes the C-terminal four arginines. This is discussed in more detail in Chapter 3.

4.2.2 Northern blot analysis of the *aba-1* mRNA transcript

Further attempts to determine the number of repeats present within the *aba-1* gene, and also to confirm its unusual structure, were

Figure 4.2. Northern blot analysis of *A. suum* adult mRNA. Total RNA was isolated from adult *A. suum* and separated on a 1% denaturing agarose gel. This was then blotted onto Hybond N. The blots were probed in (A) with the pHS10 *aba-1* insert, and in (B) with the *A. suum* Cysteine proteinase gene. RNA size markers in kb are indicated to the left.


a
b
kbp
9.5
7.5
4.4
2.4
1.4
0.2
kbp
9.5
7.5
4.4
2.4
1.4
0.2

performed by Northern blot analysis. Total RNA was isolated from adult *A. suum* and approximately 10 µg was used for the Northern blot shown in Figure 4.2 A, which was probed with the cDNA insert of clone pHS10 (see Materials and Methods for mRNA isolation and Northern blots, and Chapter 3 for a description of pHS10).

The autoradiograph does not show a clear band, but instead there is a smear which covers a size range of 4.5 kb to 0.5 kb. This suggests that either there are a range of different sized *aba-1* mRNA transcripts in the RNA preparation, or that the *aba-1* mRNA (and therefore perhaps the entire RNA isolate) is degraded. To test the latter possibility, a second Northern was performed using the same RNA preparation and was probed with the cDNA insert of clone pTP4, which contains an *A. suum* proteinase gene (see Figure 4.2 B; pTP4 was a gift from K. McCurroch). In this control, a clear band of 1.5 kb was seen. Because this correlates with the size of the mRNA for this gene, as predicted from its cDNA (K. McCurrach, personal communication), it may be concluded that the smear seen in Figure 4.2 A is not a consequence of general degradation of the *A. suum* RNA.

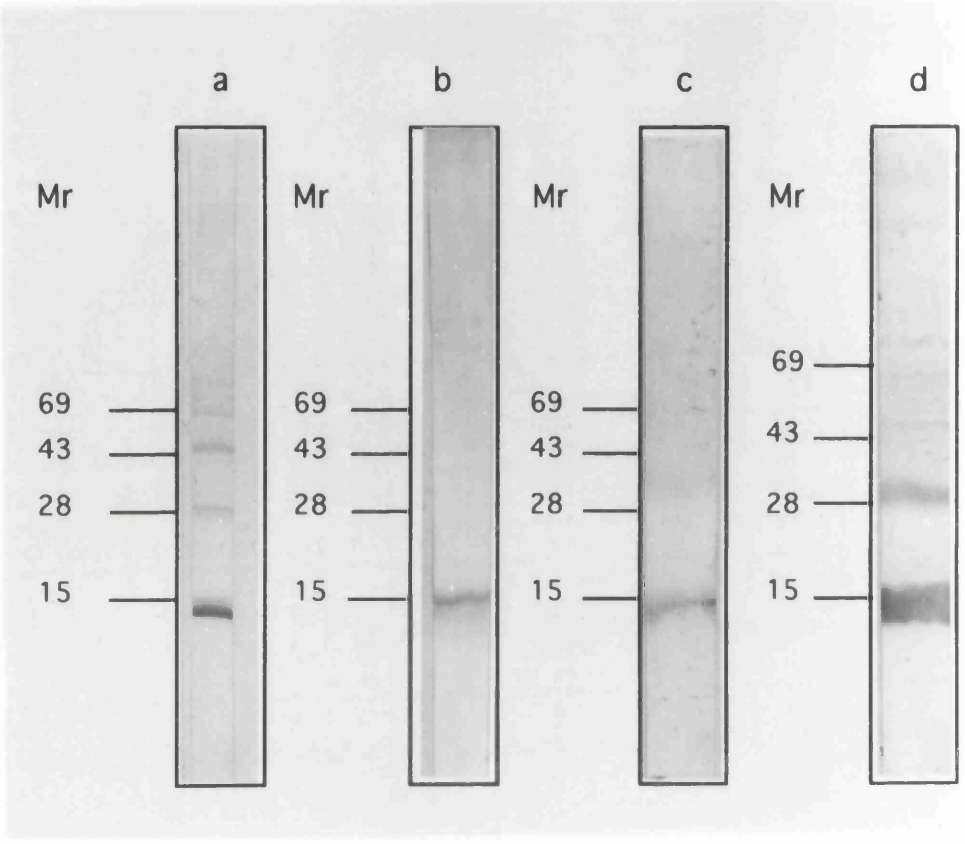
In Section 4.2.1 the Western analysis predicts that the *A. suum aba-1* mRNA must contain at least eight repeats of 399 bp, which would result in a transcript of at least 3.2 kb. However, no clear band of this size is visible (see Figure 4.2 A). Indeed, no discrete mRNA species were detected, suggesting that the *aba-1* transcript has been degraded either during this preparation or during the parasite's growth. Why this should be the case is not clear, but could perhaps be related simply to the very large size of the putative *aba-1* message. A similar result has been reported for the *A. suum* ABA-1 homologues of *B. pahangi* and *D. immitis*. In these organisms the predicted transcript for their respective genes should contain at least 20 repeats and be approximately 8 kb in size, but no discrete mRNA of the

expected size was detected (Tweedie *et al.*, 1993; Culpepper *et al.*,1992; Poole *et al.*, 1992). For these reasons, it was not possible for Northern blot analysis to be used to determine the number of 399 bp repeats within the *aba-1* gene.

4.2.3 Stage-specificity of ABA-1 expression.

To determine the stages at which the ABA-1 protein of *A. suum* is expressed during the parasite's life cycle, Western blot analysis was carried out using homogenate isolated from the following life cycle stages of *A. suum*: eggs from female adults (these eggs would have been approximately at the 2 cell stage); L2 infective stage larvae; L3/L4 larvae; and adults. Approximately 500 µg of eggs and adult homogenate, and 100 µg of L2 and L3/L4 homogenate, were used per gel (less of the L2 and L3/L4 preparations were used due to the restricted amount of material available). To exclude the possibility of contamination between lanes, each homogenate was loaded on to separate gels and blotted. The Western blots were then probed with polyclonal rabbit antiserum raised to the recombinant ABA-1 protein, which resulted in the recognition of 14.4 kDa proteins in all protein samples, as well as bands of approximately 30 kDa, 45 kDa, 60 kDa, 75 kDa and 90 kDa in the adult and egg homogenates (see Figure 4.3). The presence of a 14.4 kDa polypeptide in all stages suggests that there is no stage-specificity of ABA-1 expression, but that the protein is constitutively expressed throughout the parasite's life. Detection of the other proteins recognised by the antiserum in the egg and adult homogenates are thought to represent the incomplete cleavage products of the ABA-1 protein (see Section 4.2.1). It should be noted that there appears

Figure 4.3. Western blot analysis of the stage-specificity of the expression of the ABA-1 protein. Western blots of protein isolated from *A. suum* eggs (A), L2 stage larvae (B), L3/L4 larvae (C) and adult (D), were probed with polyclonal rabbit antiserum raised to the recombinant ABA-1 protein. Protein samples were electrophoresed under reduced conditions on a 5-25% gradient SDS-polyacrylamide gel. The positions of molecular size markers (in kDa) run on each gel are shown to the left.


a
b
c
d
Mr
Mr
Mr
Mr
69
43
28
15
69
43
28
15
69
43
28
15
69
43
28
15

to be no detection of these putative partial cleavage products in the L2 and L3/L4 larvae homogenates, perhaps because less homogenate was used.

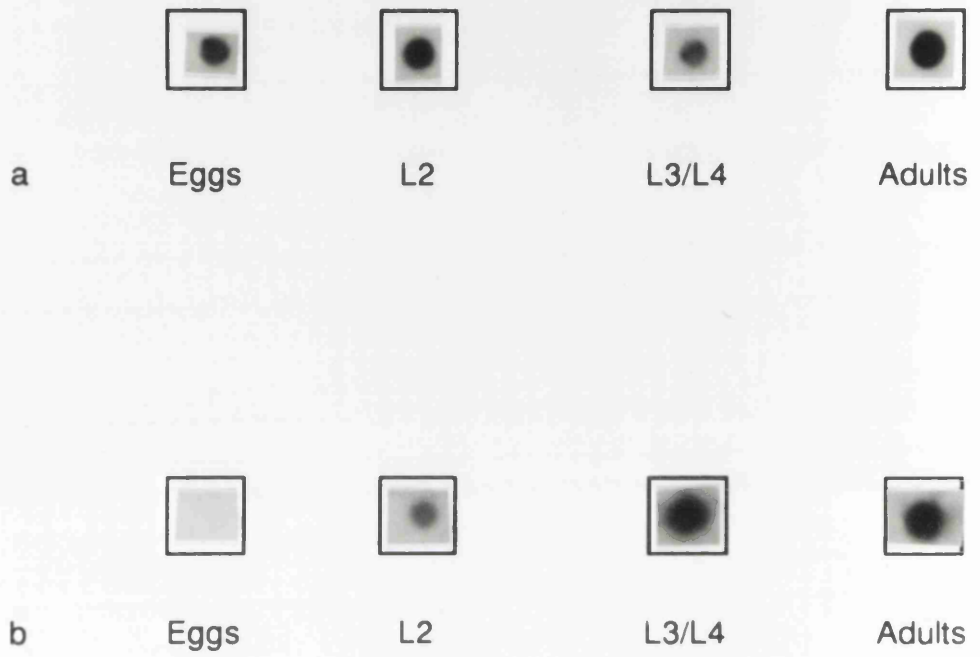
An interesting result seen in this Western blot is that the ABA-1 protein is expressed as early as the two-cell embryo stage. Detection of ABA-1 in the eggs is intriguing because if ABA-1 is utilised to evade the host's immune response, or is involved in tissue penetration, then it should not be required at such an early stage of the parasite's development. It should be noted, however, that although care was taken to isolate the eggs from the female reproductive tissue, it is possible that the ABA-1 detected at this early stage may be containment from adult tissue. However, the conclusion that the protein is expressed early in the parasite's life is supported by the existence of a homologue of *aba-1* isolated from a cDNA library constructed from mRNA isolated from early embryos of *C.elegans* (J.Moore, personal communication). The lack of stage-specificity of ABA-1 expression may suggest that ABA-1 is needed throughout the life of *A. suum.*

4.2.4 Dot blot Northern analysis of the stage-specificity of *aba-1* expression

To attempt to analyse further whether the *aba-1* gene is constitutively expressed throughout the *A. suum* life cycle, dot blot Northern analyses were carried out. Dot blots were used in preference to standard Northern analysis because, as described in Section 4.2.2, the *aba-1* mRNA appears to be prone to degradation.

Total RNA was isolated from the following *A.suum* life cycle stages: eggs isolated from adults; L2 and L3/L4 larvae; and adults. Approximately 5 μg of each preparation was used for the individual dot blots, which were

Figure 4.4. Dot blot Northern analysis of the stage-specificity of *aba-1* expression. Approximately 5 μg of total RNA isolated from *A. suum* eggs, L2 larvae, L3/L4 larvae and adults was blotted onto Hybond N and probed in (A) with the pHS10 insert, and in (B) with the *A. suum* cysteine proteinase gene.


a
Eggs
L2
L3/L4
Adults
b
Eggs
L2
L3/L4
Adults

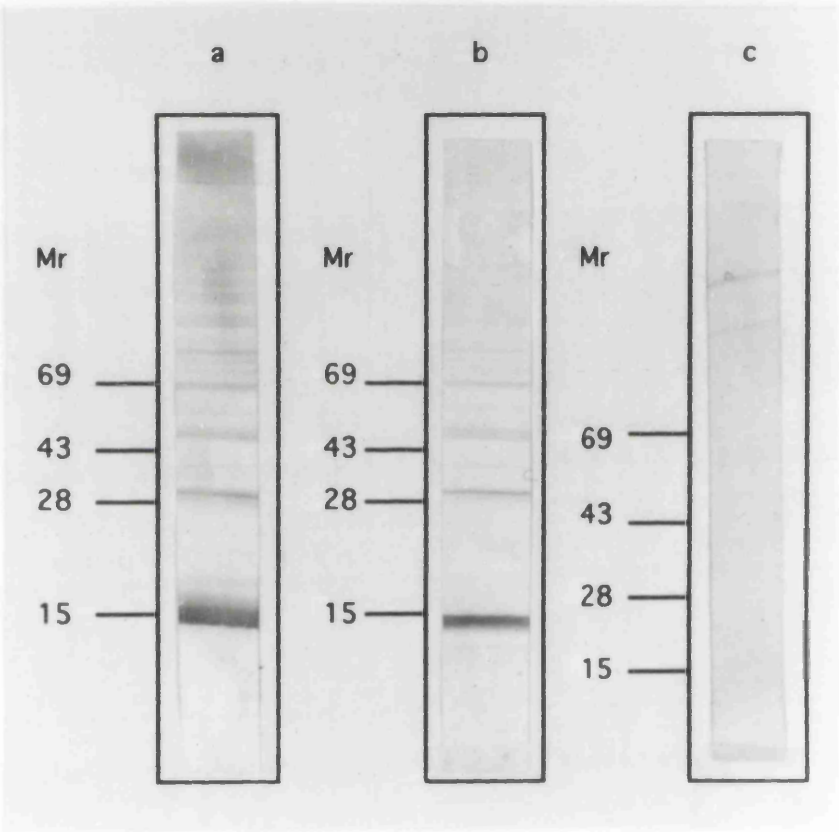
probed with the cDNA insert of pHS10 (see Materials and Methods). The results of these blots are shown in Figure 4.4 A.

It would appear from the autoradiograph that the probe hybridised to all of the RNA samples, implying that *aba-1* is expressed constitutively throughout the parasite's life cycle. To attempt to rule out the possibility that this hybridisation was a result of the *aba-1* probe non-specifically recognising degraded mRNA, the dot blots were also probed with the *A. suum* proteinase gene (Figure 4.4B; see Section 4.2.2 for details of the probe). Because this probe hybridised to only the L2, L3/L4 and adult preparations, it may be concluded that the RNA used is not generally degraded and that this gene, unlike *aba-1*, is expressed stage-specifically.

4.2.5 Western blot analysis of the tissue specificity of *A. suum* ABA-1 expression

To obtain an idea of the tissues in which the *aba-1* gene is expressed, adult *A.suum.* were dissected and protein homogenate was isolated from gut, body wall (this preparation contains cuticle, hypodermis and muscle) and the reproductive tissue. Approximately 500 µg of protein was loaded onto separate gels. As described in Section 4.2.3, each sample was loaded on to separate gels to avoid contamination between wells. After blotting the proteins onto nitrocellulose, the membranes were probed with polyclonal rabbit antiserum raised to the ABA-1 recombinant protein (see Figure 4.5). A control was also performed in which individual homogenates were probed with normal rabbit serum, but no proteins were detected (data not shown). Bands of 14.4 kDa, 30 kDa, 45 kDa 60 kDa, 75 kDa, 90 kDa, 105 kDa and 120 kDa were detected in the reproductive tissue and in the body wall tissue. Note that, as discussed in Section 4.2.1, the

Figure 4.5. Western analysis of the tissue-specificity of ABA-1 expression. Western blots of protein isolated from adult *A. suum* reproductive tissue (A), body wall (B) and gut (C), were probed with polyclonal rabbit antiserum raised to the recombinant ABA-1 protein. Protein samples were electrophoresed under reducing conditions on a 5-25% gradient SDS-polyacrylamide gel.The positions of molecular size markers (in kDa) are indicated.


a
b
c
Mr
69
43
28
15
Mr
69
43
28
15
Mr
69
43
28
15

antiserum recognises proteins that correspond to multiples of 15 kDa monomers. It is again likely that these represent the products of incomplete cleavage of the ABA-1 polyprotein.

In contrast to the pattern seen for the reproductive tissue and body wall, only two high molecular weight proteins were detected in the gut tissue (Figure 4.5). This could be interpreted as meaning that ABA-1 is either not expressed in the gut or the polyprotein is only partially processed. Because the control Western showed no such bands, it is unlikely that these proteins are a result of non-specific hybridisation. However, it is not possible to exclude the possibility that the polyclonal rabbit antiserum used recognises the consensus cleavage site of ABA-1 (RRRR; see Section 3.25) and that the large bands represent cross-reactivity to non-ABA-1 proteins containing this sequence. For these reasons it cannot be said with certainty whether or not ABA-1 is expressed in the gut. Despite this, the higher bands may indeed be derived from ABA-1 and consequently this would suggest that the polyprotein is not cleaved into 14.4 kDa monomers in this tissue. Why this would be the case is not clear, but may be related to the function of the protein.

4.2.6 Dot blot Northern analysis of the tissue-specificity of *aba-1* expression

Although the experiments in Section 4.2.5 detect the presence of ABA-1 in different tissues of the parasite, it is possible to explain this as being due to distribution of the protein after its synthesis at a single site. For this reason, dot blots of *A. suum* RNA were also prepared to attempt to establish the expression pattern of the *aba-1* transcript. Messenger RNA was isolated from the adult *A. suum* gut, body wall, and female

Figure 4.6. Dot blot Northern analysis of the tissue-specificity of *aba-1* expression. Approximately 5 μg of total RNA isolated from adult *A.* suum reproductive tissue, body wall and gut, was blotted onto Hybond N and probed in (A) with the pHS10 insert, and in (B) with the *A. suum* cysteine proteinase gene.

a

body wall

reproductive tissue

gut

b

body wall

reproductive tissue

gut

reproductive tissue. Blots derived from these preparations were probed with the cDNA insert of pHS10, and the results are shown in Figure 4.6 A. It appears from the autoradiograph that the probe hybridised to all the RNA samples, impling that *aba-1* is expressed in all the tissues.

To determine if this result is not merely due to non-specific hybridisation of the *aba-1* gene to degraded RNA, the blots were also probed with the *A. suum* proteinase gene (Figure 4.6 B; see Section 4.2.2 for details of the probe), as was performed in Section 4.2.4. Because the proteinase gene appears only to be expressed in the gut and body wall, it may be concluded that the RNA preparations used were not generally degraded and that this gene is expressed in a tissue-specific manner.

These data do not correlate with the Western blot analysis descrided in Section 4.2.5, since the *aba-1* transcript was present in all tissue preparations, whereas the 14.4 kDa ABA-1 protein was detected in the body wall and reproductive tissue but not the gut. Because the *aba-1* mRNA is expressed in the gut, this supports the idea that the higher forms of the protein observed in this tissue represent unprocessed forms of the ABA-1 polyprotein.

4.3 Discussion

This chapter reports experimental approaches taken to analyse the structure of the native ABA-1 protein, and also analysis of the stage- and tissue-specificity of its expression. The first questions addressed were whether there is evidence for the native ABA-1 being present as a polyprotein in *A. suum* and also what the total size of the polyprotein is. Western blots reported in Section 4.2.1 showed that polyclonal rabbit antiserum raised to ABA-1 recognised a series of polypeptides ranging in

size from 14.4 kDa to 120 kDa in the *A. suum* body fluid with each band being approximately 15 kDa larger than the smaller species. These data provide evidence for ABA-1 being expressed as a polyprotein which is then cleaved into approximately 14.4 kDa monomers. The largest protein detected was a polyprotein of at least 120 kDa, and suggests that the *aba-1* gene is composed of at least 8 repeats of 399 bp. This estimated size of the native ABA-1 polyprotein is smaller than that predicted for the NPA proteins of both *B. pahangi* and *D. immitis*, which are thought to be at least 200 kDa (Tweedie *et al.*, 1993; Culpepper *et al.*,1992; Poole *et al.*, 1992).

Although these data allow an estimate to be derived for the number of repeats present within the *aba-1* gene, the actual number is not known and may be greater than eight. It was hoped that Northern blot analysis of the *aba-1* mRNA transcript may have given some insight into the full size, but unfortunately this was not possible because of degradation. Similar degradation problems were reported when the size of the RNA transcripts that encode the homologous NPA from *B. pahangi* was determined by Northern blots (W. Paxton, personal communication). This type of mRNA degradation appears not to be confined to homologues of ABA-1, however, since analysis of the mRNAs encoding the Filaggrin proteins of mice (Rothnagel *et al.*, 1987) and the microtubule-associated cap proteins of Trypanosomes (Rindisbucher *et al.*, 1993) encountered the same difficulties. These genes appear to have the same repetitive structure as the genes of the NPAs, suggesting that it may be this which causes problems in mRNA preparations. The RNA degradation might also be a consequence of large putative size of all of these mRNAs. Chapter 5 describes further attempts to analyse the number of repeats present within the *aba-1* gene by examining genomic clones which encode ABA-1.

Immunological work has shown that the ABA-1 protein is present in the excretory products of cultured tissue penetrating larvae (L2, L3 and

L4 stage larvae) and that it is found in large abundance in the adult *Ascaris* (Kennedy *et al.*, 1986). Despite this, it was not known at which stage in the development of the parasite the *aba-1* gene is expressed for the first time. Western blots and dot blot Northerns detailed in this chapter suggest that the gene is expressed as early as the two-cell embryo stage. These data were interpreted as meaning that the ABA-1 protein is expressed at every stage of the parasite's development. Further evidence for this is that ABA-1 homologues in *B. pahangi* and *D. immitis* have been shown to be expressed in both the early larval and adult stages of development (Tweedie *et al.*, 1993; Culpepper *et al.*, 1992; Poole *et al.*, 1992). In addition, a cDNA encoding an ABA-1 homologue in *C. elegans* has been identified in the early embryo stages (Moore *et al.*, 1993). A possibility that arises from this work on stage-specificity of ABA-1 expression is that if the protein is present in the two-cell embryo stage, then this might imply that the adult female *A. suum* supplies the *aba-1* transcript directly to the embryo, rather than the embryo's own gene being transcribed at this stage. This has been observed for a number of *Ascaris* genes (F. Muller, personal communication).

Tissue specificity of ABA-1 expression was examined in this chapter, both at the protein level (by Western analysis) and at the mRNA level (by Northern analysis). It appeared from this work that the *aba-1* gene is expressed in the reproductive tissue, body wall and gut of the adult worm, while the 14.4 kDa monomeric form of ABA-1 was only detected in the reproductive tissue and body wall. In the gut of the worm only larger, unprocessed ABA-1 polyprotein was found (see below). It should be noted that body wall preparation used here encompasses muscle tissue, hypodermis and the cuticle, and consequently it is possible that ABA-1 may be found in any or all of these. Recently, however, Fetterer and Wasiuta (1987) have developed a technique to separate these tissues, and

further experiments could be carried out to determine which body wall tissues express the protein. Despite this, the Western and Northern analysis presented correlates with electron microscopy experiments performed by Fraser (1990) and by Kennedy (unpublished) which imply that ABA-1 is dispersed throughout most of the larval parasite tissue except the cuticle.

An interpretation of the data described above is that *aba-1* is a "housekeeping gene", encoding a protein with an active role in all cells throughout development. Furthermore, taken together with the predicted 4 alpha-helical bundle structure of ABA-1 and other NPAs, it may be that these proteins are small carrier proteins within cells (Kennedy *et al.*, 1993). This is speculation at this time, however, and does not agree with the following observations: ABA-1 is present in large quantities in the *A. suum* adult body fluid; all the ABA-1 homologues appear to be excreted/secreted by their respective parasites; and the hosts of the parasites elicit an IgE response to NPAs. One explanation for the presence of ABA-1 in the body fluid could be that rupturing of the adult *A. suum* cells during preparation could result in contamination by the protein. Similarly, ABA-1 and the other NPAs may not be genuine excretory/secretory products, but death or damage to the parasite during growth *in vitro* and *in vivo* could lead to leakage of ABA-1, thereby leading the hosts to elicit an immune response to the proteins. Definitive examination of the role of ABA-1 will probably only be possible by constructing *C. elegans* mutants in which the gene is disrupted (see Chapter 3 discussion).

At this time, the significance of the fact that the ABA-1 polyprotein appears not be cleaved into 14 kDa monomers in the gut of adult *A. suum* is not clear. Whether this means that the protein is non-functional in this

tissue, or whether it is needed as a polyprotein rather than a 14 kDa protein, awaits experimental examination.

Chapter 5

Analysis of the genomic copy of the *A. suum aba-1* gene

5.1 Introduction

Cloning and characterisation of the cDNA clone pHS10, which encodes the *A. suum* allergen ABA-1, revealed that it consists of two complete repeats of 399 bp and one truncated repeat of 282 bp. Each complete repeat was shown to encode an approximately 15 kDa polypeptide that has 100% sequence identity to the N-terminal amino acid sequence of the native ABA-1 protein. These data suggested that the ABA-1 protein is expressed as a large polyprotein that is subsequently cleaved into approximately 15 kDa monomer proteins. Chapter 4 reports the approaches taken to analyse the size of the polyprotein, which was estimated to be at least 120 kDa, suggesting that there are at least 8 repeats of 399 bp within the *aba-1* gene. Although the above analysis has given insight into the structure of the gene, the number of 399 bp repeats within the gene is still unknown.

This chapter reports Southern blot analysis of the *A. suum aba-1* gene, and isolation and characterisation of *aba-1* clones from an *A. suum* genomic DNA library. Both these analyses were performed to verify that the genomic copy of the *aba-1* gene is composed of 399 bp repeats like the cDNA clone pHS10, and also to give an insight as to how many repeats are present within the gene. Analysis of the genomic *aba-1* clones will also provide information on whether intronic sequences are present within *aba-1*, which is of interest because the few genes that have been isolated from *Ascaris* (Kingston *et al.*, 1989) all appear to have very small introns of only a few hundred base pairs (an unusual feature in comparison to many other eukaryotic genes; Watson *et al.*, 1987). In addition to this, it was hoped that isolation of a genomic clone containing the 5' region of the *aba-1* gene would allow characterisation of DNA sequences involved in

the gene's expression, for example TATA box and enhancer elements, and peptide sequences involved in intracellular routing and processing.

5.2 Results

5.2.1 Genomic organisation of the *aba-1* gene

To analyse the genomic organisation of the *aba-1* gene, Southern blot analysis of total genomic DNA, isolated from adult *A. suum* testes, was carried out. The genomic DNA was digested with two restriction enzymes that cut within the 399 bp repeats of the *aba-1* cDNA clone (*Nru*I and *Xba*I), and two which do not (*Eco*RI and *Sau*3A), and the blot was probed with the pHS10 insert (Figure 5.2A; see Figure 5.1 for a restriction map of pHS10, and Materials and Methods for genomic DNA isolation and Southern blotting). Lanes B and D of Figure 5.2A show digestion of the genomic DNA with *Nru*I and *Xba*I respectively, illustrating that in both cases the pHS10 probe hybridised to DNA fragments of approximately 400 bp. This may be interpreted as meaning that the 400 bp fragment corresponds to a 399 bp *aba-1* repeat, therefore suggesting that the genomic copy of the *aba-1* gene is composed of repeating units as seen in the cDNA clone pHS10. The probe also hybridised to fragments of 3 kbp and 21 kbp in the *Nru*I digest, and to an 18 kbp fragment in the *Xba*I digest (Figure 5.2A, lanes B and D). It is likely that these bands represent DNA fragments which resulted from digestion at *Nru*I and *Xba*I sites present at the extreme 5' or 3' ends of the *aba-1* gene and in the flanking genomic sequences. An 800 bp fragment is also present in the *Nru*I digestion (Figure 5.2A, lane B) which probably represents two repeats of 399 bp, suggesting either that one of the 5' repeats does not contain an *Nru*I site or that the restriction digestion was incomplete.

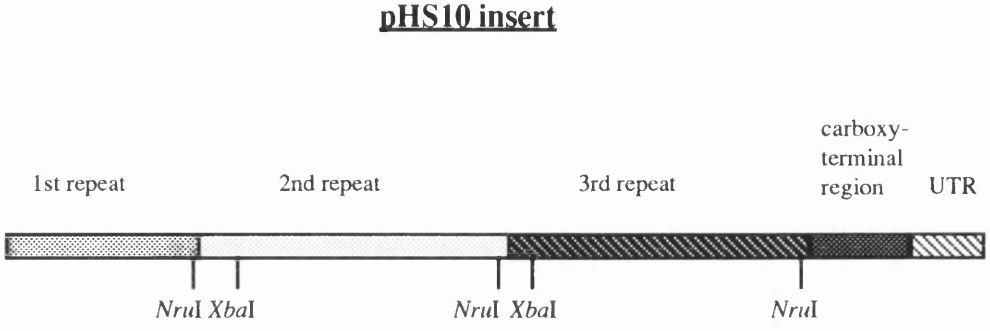
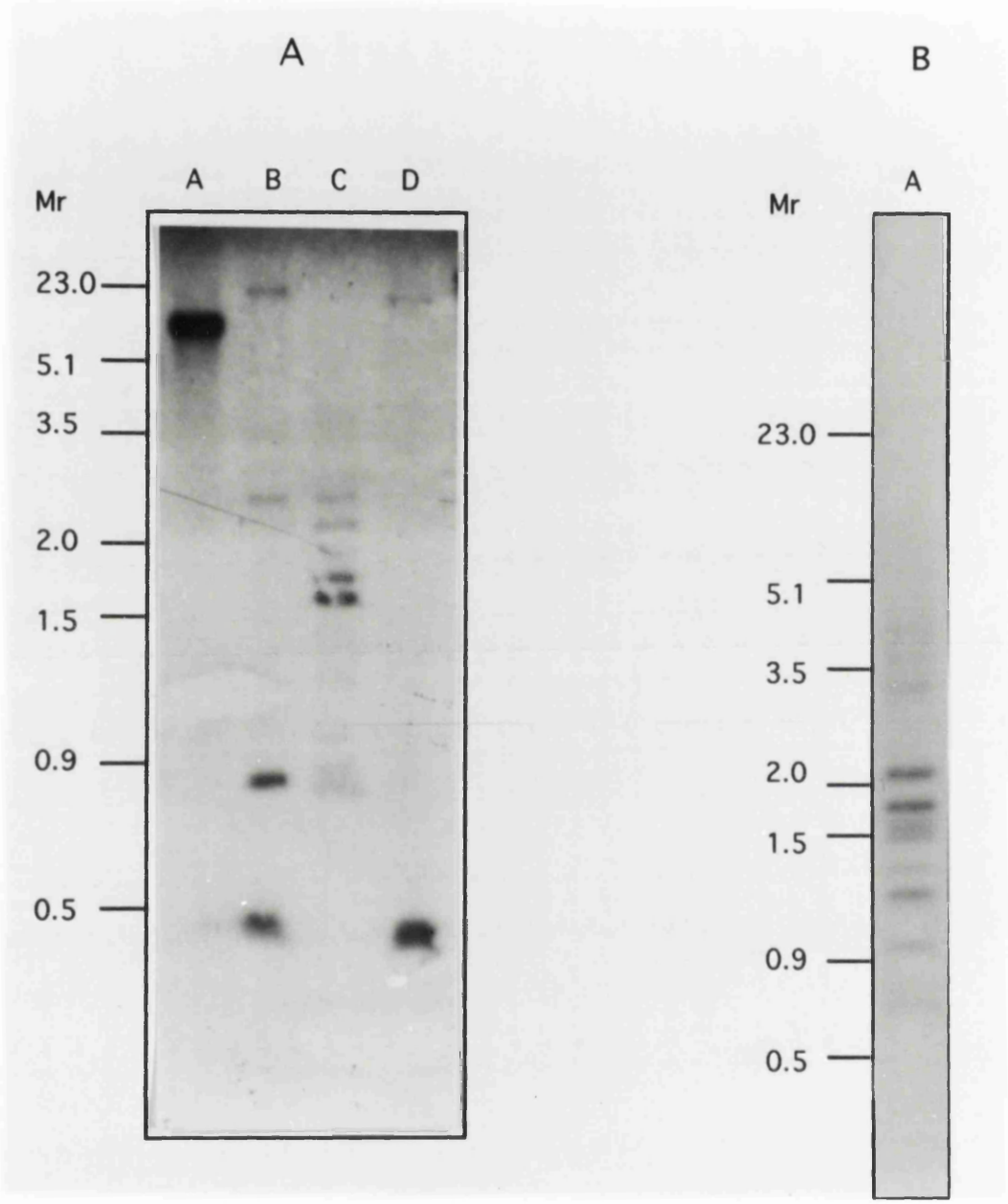


Figure 5.1. Partial restriction map of the pHS10 insert. The three tandem repeats encoding ABA-1 are indicated (the 1st repeat is truncated), as is the carboxy-terminal region and the 3' untranslated region (UTR) of the gene. The positions of *Nru*I and *Xba*I recognition sequences are shown.

Figure 5.2. Southern blot analysis of the genomic organisation of the *aba-1* gene. (A) *A. suum* adult testes genomic DNA was digested with *Eco*RI (A), *Nru*I (B), *Sau*3A (C) and *Xba*I (D) and then electrophoresed on a 0.8% agarose gel. **(B)** *A. suum* L2 larvae genomic DNA was digested with *Sau*3A and then electrophoresed on a 0.8% agarose gel. In both (A) and (B) the gels were Southern blotted and probed with the pHS10 insert. λ *Hind*III/ *Eco*RI markers (in kbp) run on the same gel are shown at the left of the figures.


A
B
A
B
C
D
Mr
23.0
5.1
3.5
2.0
1.5
0.9
0.5
Mr
A
23.0
5.1
3.5
2.0
1.5
0.9
0.5

Analysis of the genomic DNA restricted with *Eco*RI, which does not cut within the 399 bp repeat of the cDNA clone, shows that the probe hybridises to a single DNA fragment of approximately 9 kbp (see Figure 5.2A, lane A). This band is thought to represent a genomic DNA fragment that contains the full length, or near full length, *aba-1* gene. In contrast, digestion with *Sau*3A (Figure 5.2A, lane C), which also does not cut within the 399 bp repeats of pHS10, resulted in a series of fragments ranging from 0.8 kbp to 3 kbp. One explanation for this result is that repeating units which lie 5' to those contained within clone pHS10 do contain *Sau*3A sites. Alternatively it is conceivable that DNA sequence polymorphisms exist within the *aba-1* gene between worms; such polymorphisms are more likely to create *Sau*3A sites than *Eco*RI sites because of their smaller size (4 bp rather than 6 bp). To analyse the latter explanation, Southern blot analysis was performed with *Sau*3A-digested genomic DNA isolated from L2 *A.suum* infective larvae, which represents a larger population of worms than the above experiment (which used only 4 individuals); the blot was again probed with pHS10 (Figure 5.2 B). A series of DNA fragments were detected which ranged in size from 0.5 kbp to 4kbp, implying that a larger range of *aba-1* DNA fragments are created by *Sau*3A digestion than when a smaller population of worms were used. These data may suggest that individual worms do indeed display sequence polymorphisims within their *aba-1* 399 bp repeats.

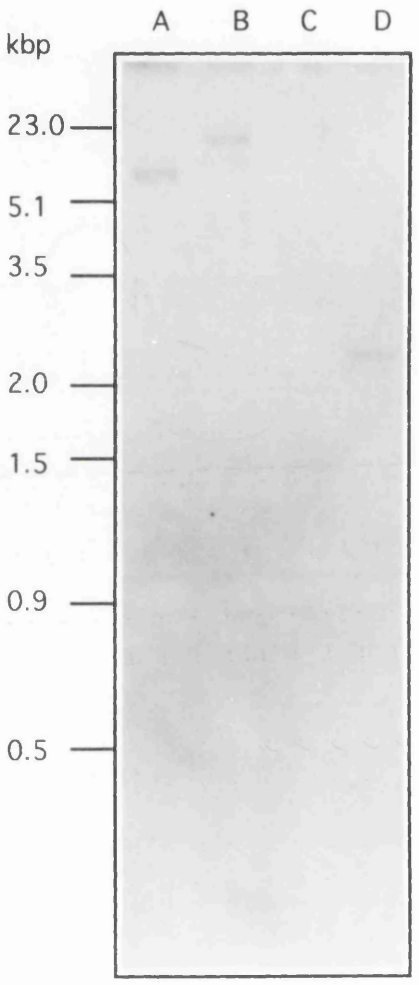
In conclusion, this Southern analysis is consistent with the *aba-1* gene being composed of 399 bp repeats, as is the pHS10 cDNA. It also implies that the entire *aba-1* gene may be 9 kbp in length and that DNA sequences differences are present in the individual *aba-1* repeats and perhaps between different worms. Note that such DNA sequences variations were seen in the *aba-1* insert analysed in Chapter 3 (Section 3.2.4).

5.2.2 Genomic organisation of the divergent repeat

Chapter 3 discusses the isolation of an ABA-1 expressing cDNA clone, named pJM33, by J. Moore which contains 5 repeats of 399 bp, the most 5' of which displays only 49% DNA sequence identity to the repeats of clone pHS10 (Moore *et al.*, 1993). It was not known at this stage if this repeat, termed "the divergent repeat", is situated within the same gene as the repeats present within pHS10, since pJM33 does not contain the 3' untranslated region of the *aba-1* gene. To determine if this is the case, the restriction pattern of the divergent repeat and a pHS10 repeat were compared by removing the pHS10 probe bound to the Southern blot used in Figure 5.2A and reprobing it with the divergent repeat (a gift from J. Moore). Figure 5.3, lane A, shows that *Eco*RI digestion generates a DNA fragment of approximately 9 kbp, which is identical to that recognised by the pHS10 probe (see Figure 5.2A, lane A). Since *Eco*RI does not cut within either of the repeats, the 9 kbp fragment again represents the full length, or near full length, *aba-1* gene and therefore suggests that both the pHS10 repeats and the divergent repeat of pJM33 lie within the same *aba-1* gene. This may be concluded because the fragments must be created by digestion with *Eco*RI at the same genomic sites surrounding the *aba-1* gene; if the repeats were in different genes it is very unlikely that the same sized *Eco*RI genomic DNA fragments would be created.

Figure 5.3, lane B, shows that the divergent repeat recognises a DNA fragment of approximately 21 kbp in the *Nru*I digestion, which is again identical to that recognised by the pHS10 insert probe. This fragment is thought to have resulted from digestion at the extreme 5' *Nru*I site within the *aba-1* gene and a flanking genomic *Nru*I site, therefore again providing evidence that the divergent repeat is situated within the same gene as the pHS10 repeats since the flanking sequences would appear to be identical.

Figure 5.3. Southern blot analysis the genomic organisation of the *aba-1* gene. *A. suum* genomic DNA was digested with *Eco*RI (A), *Nru*I (B), *Sau*3A (C) and *Xba*I (D) and run on an 0.8% agarose gel. The gel was capillary blotted on to Hybond N and probed with the divergent repeat. λ HindIII/ EcoRI markers (in kbp) electrophoresed on the same gel are shown at the left of the figure.


A
B
C
D
kbp
23.0
5.1
3.5
2.0
1.5
0.9
0.5

Comparison of the *Nru*I digest probed with pHS10 and with the divergent repeat shows that the latter probe does not hybridise to the 400 bp *aba-1* repeat fragment under high stringency conditions, and therefore highlights the sequence divergency of this repeat when compared to the pHS10-like repeats.

These Southern data suggest that the divergent repeat is present within the same *aba-1* gene as the repeats of clone pHS10, and, because pHS10 contains the 3' untranslated region, the divergent repeat must be situated 5' to the repeats which are present in pHS10. These data imply that the divergent repeat is not an artefact produced when the cDNA library was constructed, and has a biological relevance to *aba-1*.

5.2.3 Construction and screening of an *A.suum* genomic library

To allow the characterisation of genomic *aba-1* clones, an *A.suum* genomic library was constructed in the lambda vector λGEM-12 using DNA prepared from adult testes. Figure 5.4 illustrates diagramatically the construction of this library, details of which are given in Materials and Methods. The titre of this library was approximately 0.5 x 10^6 pfu/ml. Approximately 0.25 x 10^6 pfu of the library were amplified and stored, and the remaining 0.25 x 10^6 pfu were screened with the pHS10 insert. Unamplified library was screened for clones containing *aba-1* in preference to the amplified library to eliminate (as much as possible) potential recombination between the repeats of the *aba-1* gene.

The primary screen of the library resulted in the isolation of 20 *aba-1*-positive clones, 10 of which were replated and screened a second time. Of these 10 clones, only 7 remained positive. A further 4 clones were revealed

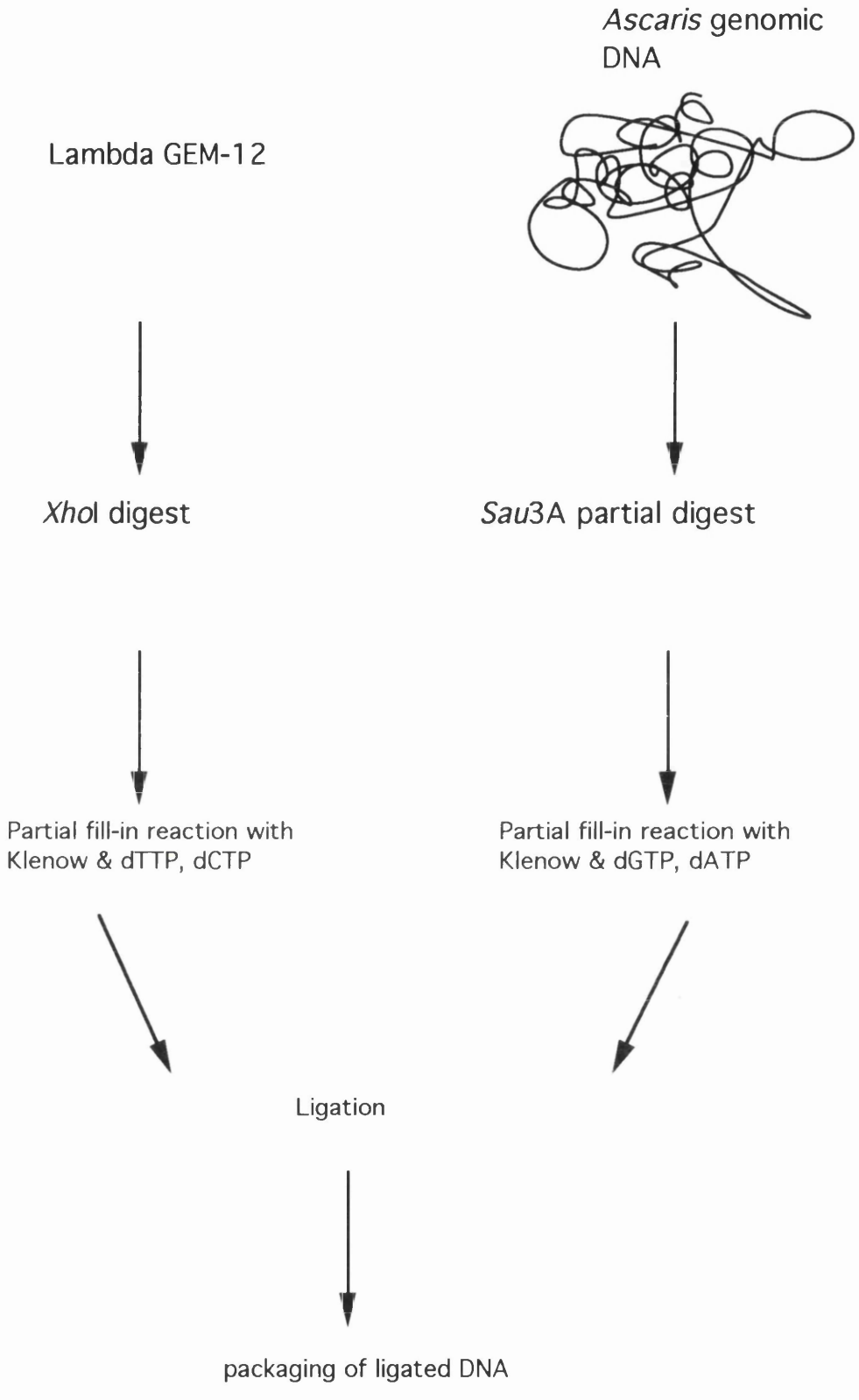


Figure 5.4. Schematic diagram of the construction of the *A. suum* genomic library.

to be false positives in a tertiary screening, resulting in the isolation of 3 putative *aba-1* clones, which were designated λHSg5, λHSg6 and λHSg7.

5.2.4 Initial characterisation of the putative *aba-1* genomic clones

Lambda DNA was prepared from clones λHSg5, λHSg6 and λHSg7 and digested with the enzymes *Sac*I, *Xho*I, *Bam*HI, *Eco*RI and *Sfi*I (all of which do not cut within a 399 bp repeat of *aba-1* cDNA clone pHS10), and by *Xba*I (which cuts within the repeats; Figure 5.1). Figure 5.5 shows a Southern blot of these DNA restriction digestions which was probed with the cDNA insert from clone pHS10. Digestion of λHSg6 and λHSg7 with *Xba*I (Figure 5.5, lanes B and N) created, in both cases, a fragment of approximately 400 bp, which is the size of the *aba-1* repeat, confirming that these clones contain the *aba-1* gene. In contrast, when clone λHSg5 was restricted with *Xba*I (Figure 5.5, lane H) it did not result in a band of 400 bp, suggesting that the *aba-1* gene is not present within its insert. Digestion of λHSg6 and λHSg7 with *Sfi*I, *Eco*RI, *Bam*HI, *Xho*I and *Sac*I (Figure 5.5, lanes A, C, D, E, F, M, O, P, Q and R) illustrates that the pHS10 probe recognisd the same DNA fragments in each clone, suggesting that they both contain *aba-1* DNA. Because the two genomic clones were indistinguishable in this analysis, only λHSg7 was characterised further.

5.2.5 Initial restriction mapping of clone λHSg7

Since the *A.suum* library from which clone λHSg7 was isolated was constructed in λGEM-12, a rough restriction map of the clone was produced by the *Sfi*I mapping technique (see Materials and Methods); this

Figure 5.5. Southern blot analysis of the genomic clones λHSg5, λHSg6 and λHSg7. These λ clones were digested with various enzymes (listed below) and were electrophoresed on a 0.8% agarose gel. The gel was capillary blotted onto Hybond N and probed with the *aba-1* insert of pHS10.

lane	clone	Digestion
A	λHSg6	*Sfi*I
B	λHSg6	*Xba*I
C	λHSg6	*Eco*R1
D	λHSg6	*Bam*H1
E	λHSg6	*Xho*I
F	λHSg6	*Sac*I
G	λHSg5	*Sfi*I
H	λHSg5	*Xba*I
I	λHSg5	*Eco*RI
J	λHSg5	*Bam*HI
K	λHSg5	*Xho*I
L	λHSg5	*Sac*I
M	λHSg7	*Sfi*I
N	λHSg7	*Xba*I
O	λHSg7	*Eco*RI
P	λHSg7	*Bam*HI
Q	λHSg7	*Xho*I
R	λHSg7	*Sac*I

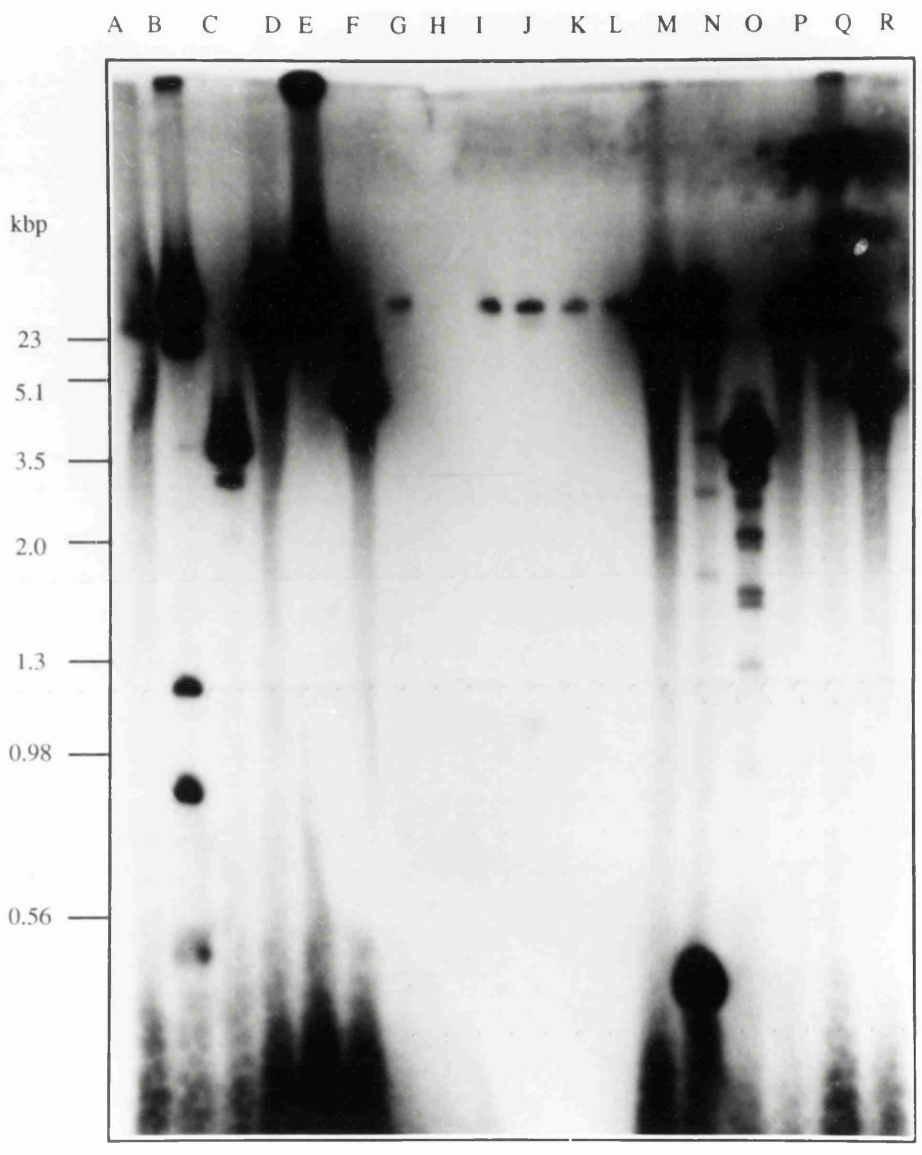
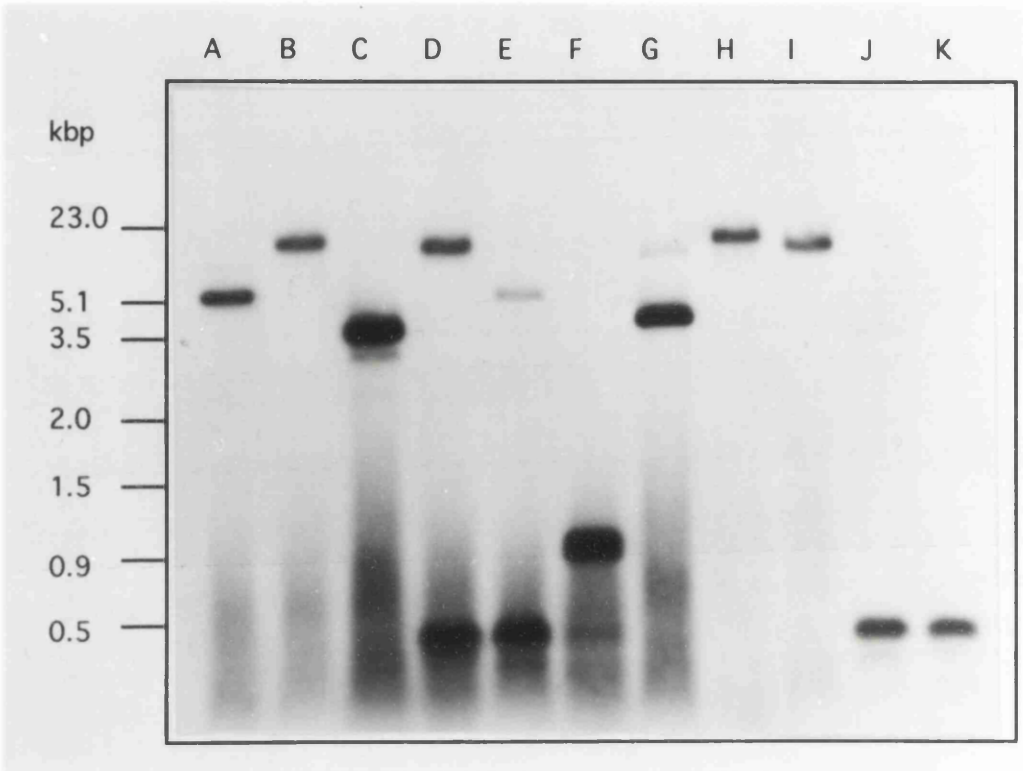

A B C D E F G H I J K L M N O P Q R
kbp
23
5.1
3.5
2.0
1.3
0.98
0.56

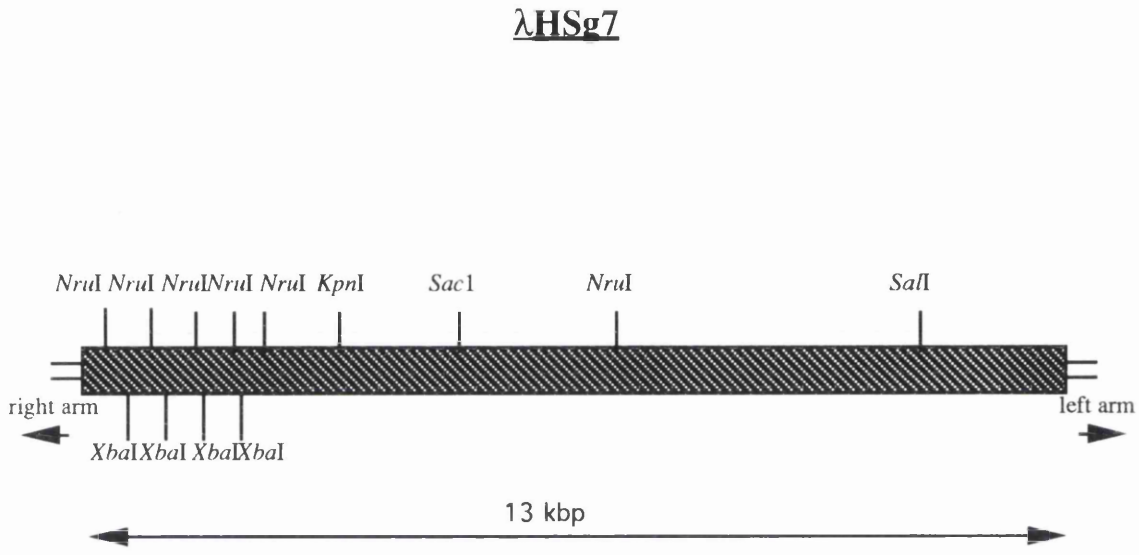
Figure 5.6. Southern blot analysis of λHSg7. λHSg7 was digested with the following enzymes: *Sac*I (A); *Bam*HI (B); *Eco*RI (C); *Xba*I and *Bam*HI (D); *Nru*I and *Bam*HI (E); *Sau*3A (F); *Kpn*I and *Bam*HI (G); *Xho*I and *Bam*HI (H); *Sal*I (I); *Bam*HI, *Nru*I and *Kpn*I (J); *Sac*I and *Nru*I (K) The digested DNA was electrophoresed on a 0.8% agarose gel, capillary blotted onto Hybond N and probed with the pHS10 insert. λ *Hind*III/ *Eco*RI markers (in kbp) run on the same gel are shown at the left of the figure.


A
B
C
D
E
F
G
H
I
J
K
kbp
23.0
5.1
3.5
2.0
1.5
0.9
0.5

is possible due to the *Sfi*I site in the λ vector's polylinker (data not shown). To verify the rough map produced by *Sfi*I mapping, and to gain insight as to where the *aba-1* gene lies within the λHSg7 insert, a Southern blot of the clone's DNA was prepared and probed with the pHS10 insert (Figure 5.6). λHSg7 was digested with the following enzymes in this Southern blot: *Sac*I; *Bam*HI; *Eco*RI; *Xba*I and *Bam*HI; *Nru*I and *Bam*HI; *Sau*3A; *Kpn*I and *Bam*HI; *Xho*I and *Bam*HI; *Sal*I; *Bam*HI ,*Nru*I and *Kpn*I; *Sac*I and *Nru*I. The restriction map compiled from these data is shown in Figure 5.7.

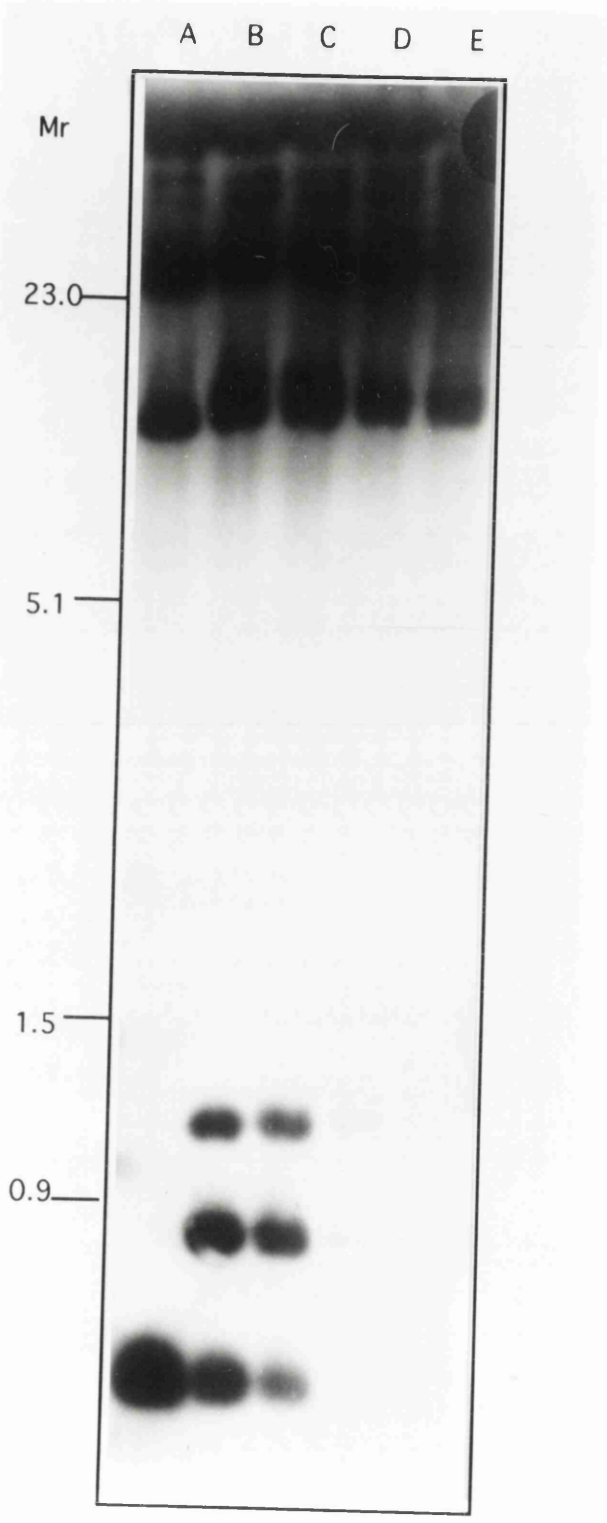
Analysis of these data implied that the λHSg7 insert comprises 13 kbp of *A. suum* genomic DNA, 8 kbp of which hybridised to clone pHS10. Approximately 2 kbp of this 8 kbp region was thought to contain repeating units of 399 bp like those found in the cDNA clone pHS10, based on the fact that this area showed a similar restriction pattern to the cDNA clone (compare Figure 5.7 to Figure 5.1). To determine the number of 399 bp repeats present within this area, Southern blot analysis was performed (using pHS10 as a probe) after partially digesting λHSg7 with *Xba*I (which cuts once within each repeat of pHS10; Figure 5.8). DNA fragments of approximately 400 bp, 800 bp and 1200 bp were detected in this analysis, suggesting that λHSg7 contains 4 *aba-1* repeats which are similar, if not identical, in size to those found in pHS10.

At this stage it was not clear what the remaining 6 kbp of the 8 kbp pHS10-hybridising region was composed of. Two possibilities were considered: firstly, this region might have consisted of 3' untranslated region of the *aba-1* gene; or, secondly, it contained a more 5' region (possibly encompassing the start of the gene and further repeats). Since it was predicted from the Southern analysis of genomic DNA descrided in Section 5.2.2 that the divergent repeat from clone pJM33 was situated 5' to the pHS10 repeats, a Southern blot was performed on λHSg7 which was probed with the divergent clone (data not shown). No DNA fragments



5.7. Initial restriction map of λHSg7 derived from S*fi*I mapping and Southern blotting. The *aba-1* genomic insert is shown as a stripped box, and the left and right arms of the λ vector are indicated.

Figure 5.8. Southern analysis of partially-digested λHSg7. λHSg7 was partially digested with the following concentrations of *Xba*I for 1 hr: 5 units (A); 1 unit (B); 0.5 units (C); 0.2 units (D); 0.1 units (E). Samples were electrophoresed on a 0.8% agarose gel which was blotted onto Hybond N. The blot was probed with the pHS10 insert. λ HindIII/ EcoRI markers (in kbp) run on the same gel are shown at the left of the figure.


A
B
C
D
E
Mr
23.0
5.1
1.5
0.9

were recognised in this experiment, suggesting that the divergent repeat does not hybridise to λHSg7. This therefore implies that the 6 kbp region comprises the 3' untranslated region of the *aba-1* gene, or alternatively that it does not contain sufficient 5' sequence to encompass the divergent repeat.

5.2.6 Subcloning of λHSg7

To analyse the insert of clone λHSg7, it was decided to subclone DNA fragments of it into an appropriate plasmid to allow sequencing. This subcloning involved digesting λHSg7 with *Sac*I to create two DNA fragments of 5 kbp and 8 kbp, both of which pHS10 hybridises to. The fragments were then ligated into the vector pUC19, creating the plasmids pHS50 and pHS60 respectively (see Figure 5.9). Clone pHS50's insert was thought to consist of 4 repeating units and a further region of 3 kbp which also hybridises to pHS10; clone pHS60 contained an insert of 8 kbp, 3 kbp of which showed hybridisation. It should be noted that difficulty was found in subcloning the 5 kbp fragment, perhaps due to its repeating nature.

5.2.7 DNA sequence analysis of the *aba-1* genomic inserts of clones pHS50 and pHS60

To analyse the DNA sequence of the inserts of subclones pHS50 and pHS60, the Sanger Dideoxy sequencing method was employed (see Materials and Methods) using oligonucleotide primers that hybridise to the pUC19 polylinker sequence, thereby allowing the sequence of the extreme 5' and 3' regions of the inserts to be determined. DNA sequence

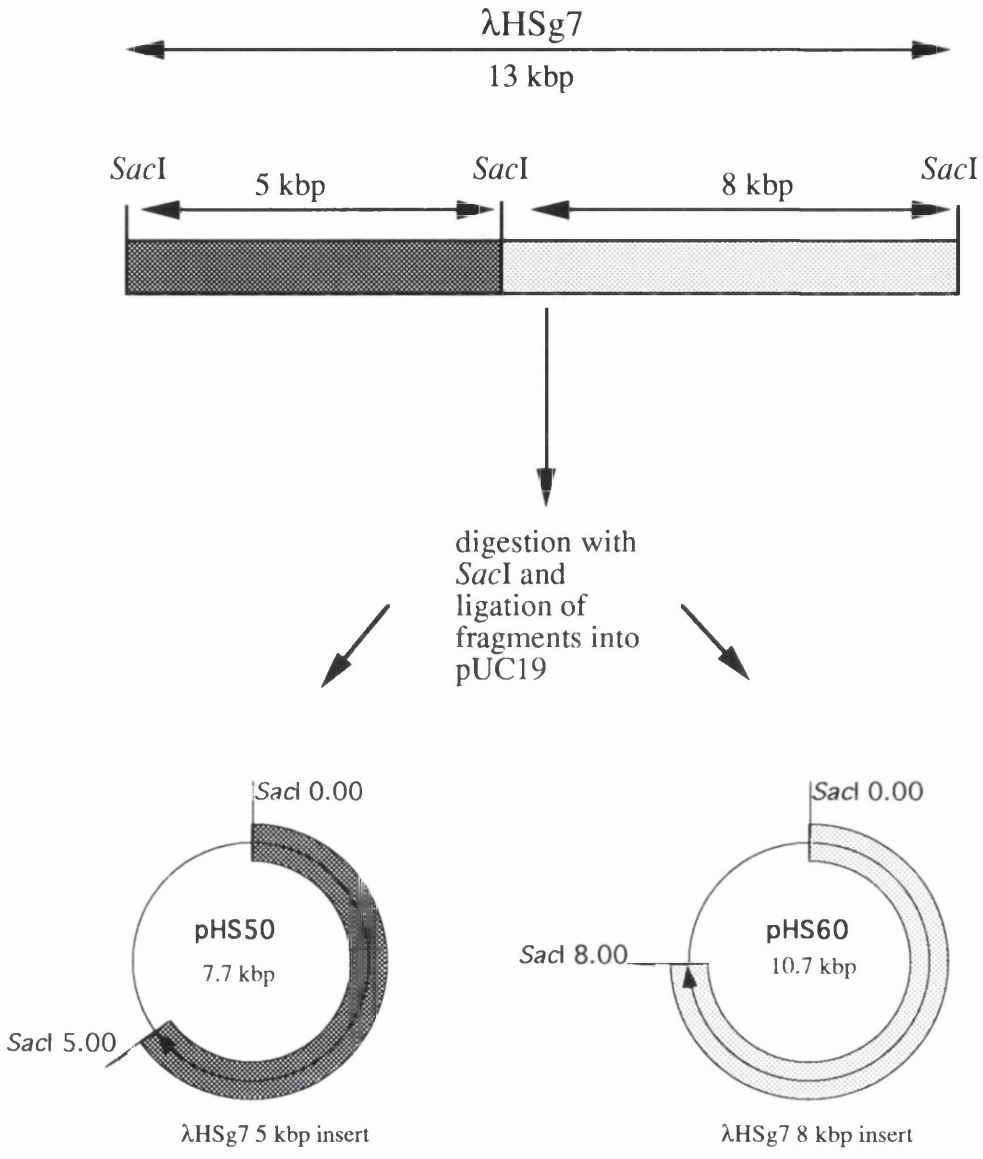
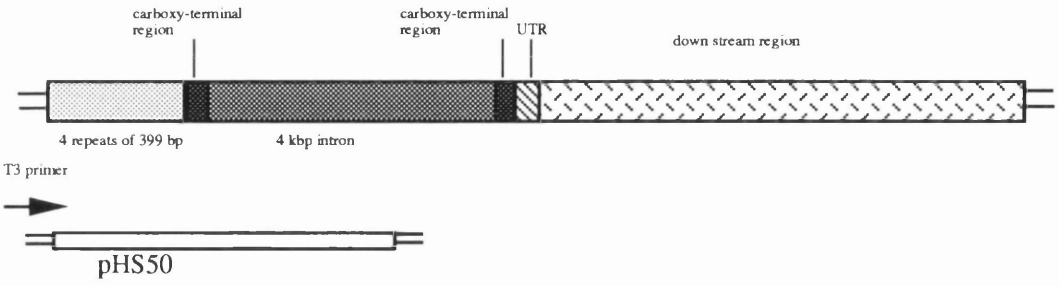


Figure 5.9. Diagrammatic representation of the subcloning of λHSg7.

λHSg7 was digested with *Sac*I, resulting in an 8kbp and a 5kbp fragment. These fragments were ligated into pUC19, which also was digested with *Sac*I, producing clones pHS60 and pHS50.

from the extreme 5' region of pHS50 revealed that it encodes the ABA-1 protein, since its translated amino acid sequence showed 90% identity to the translated sequence of pHS10 (see Figures 5.10 and 5.11; see Section 3.2.4). Further analysis of this extreme 5' sequence revealed that the genomic insert of clone λHSg7 encodes a truncated repeat of the *aba-1* gene, in which the first 105 bp are missing. These data confirm that clone λHSg7 contains the *aba-1* gene.

Sequencing of the extreme 5' and 3' regions of the insert in pHS60, and the 3' sequence of pHS50's insert, revealed that these regions do not show similarity to the *aba-1* cDNA. These data produced a discrepancy, since it was known that the pHS10 clone hybridised to both the pHS60 insert and to the entire pHS50 fragment (see Section 5.2.6). To try to understand this discrepancy, an oligonucleotide primer (UTR1) was designed, based on the cDNA sequence (Section 3.2.4), to hybridise to the 3' untranslated region of the *aba-1* gene (see Figure 5.12) and therefore allow sequencing into the repeats. Both pHS50 and pHS60 were sequenced with UTR1 to determine which area of the λHSg7 genomic insert contained the 3' untranslated region. The sequencing reaction only worked on clone pHS60, implying that the 3' untranslated region lies within this clone and not within pHS50. A further feature of the aba-1 gene was revealed by this sequencing reaction: directly 5' to the untranslated region there is 81 bp of DNA sequence that shows 100% identity to the 3' end of the carboxy-terminal region of the cDNA clone pHS10 (see Section 3.2.4), but it is followed by unknown sequence instead of the expected remainder of the carboxy-terminal region and then *aba-1* repeating units (see Figures 5.13 and 5.14). An interpretation of these data is that the carboxy-terminal region is interrupted by an intron in the genomic *aba-1* gene. Support for this hypothesis was provided by the fact that there is a consensus intron/exon acceptor splice site motif (Hammond and Bianco *et al.*, 1993)




```
cccgggaattctcGATCTTCAAGCTAAAATTCTTCATTACTAT
               D  L  Q  A  K  I  L  H  Y  Y

GACGAACTCGAAGGAGATGCTAAAAAGGAGGCAACTGAGCATT
D  E  L  E  G  D  A  K  K  E  A  T  E  H  L

TGAAGGACGGATGCCGCGAAATTCTTAAGCACGTTGTTGGGGA
  K  D  G  C  R  E  I  L  K  H  V  V  G  E

AGAGAAGGAAGCAGAGCTGAAGAAACTCAAAGACTCGGGAGCA
 E  K  E  A  E  L  K  K  L  K  D  S  G  A

AGCAAAGAGGAAGTCAAAGCCAAAGTCGAAGAGGCACTTCATG
S  K  E  E  V  K  A  K  V  E  E  A  L  H  A

CAGTAACCGACGAGGAGAAGAAGCAATATATCGCCGATTTTCG
  V  T  D  E  E  K  K  Q  Y  I  A  D  F  R

ACCAGCATGCAAGAAAATCTTTGGTGCAGCACATACTTCGCGA
 P  A  C  K  K  I  F  G  A  A  H  T  S  R

CGAAGGAGGCATCATTTCACCCTTGAAAGTAGTCTAGATACCC
R  R  R  H  H  F  T  L  E  S  S  L  D  T  H

ATCTGAAATGGCTCAGTCAAGAACAGAAAGA
  L  K  W  L  S  Q  E  Q  K
```

Figure 5.10. DNA sequence of the extreme 5' region of the insert of pHS50, and the putative amino acid sequence derived from it. Lower case letters represent lambda vector sequence. The schematic diagram at top illustrates the region of pHS50 that has been sequenced and what part of the λ HSg7 this corresponds to. The position that the sequencing primer T3 hybridises to vector sequence in pHS10 is shown by an arrow.

```
pHS10 HHFTLESSLDTHLKWLSQEQKDELLKMKKDGKAKKELEAKILH
pHS50                                        D-E-----

pHS10 YYTDELEGDAKKEATEHLKGGCREILKHVVGEEKAAELKNLKD
pHS50 ------------------D--------------E----K---

pHS10 SGASKEELKAKVEEALHAVTDEEKKQYIADFGPACKKIYGV-H
pHS50 -------V----------------------R------F-AA-

pHS10 TSRRRR
pHS50 ------
```

Figure 5.11. Comparison of the putative amino acid sequences derived from the cDNA insert of pHS10 and the genomic insert of pHS50. Where the amino acid residues in the pHS10 sequence are identical to pHS50 they are represented by (-), while amino acid differences are shown by their single letter amino acid code.

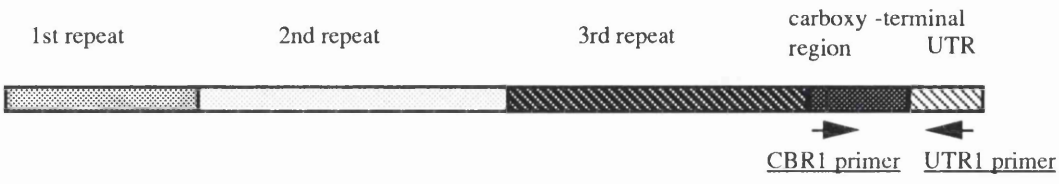
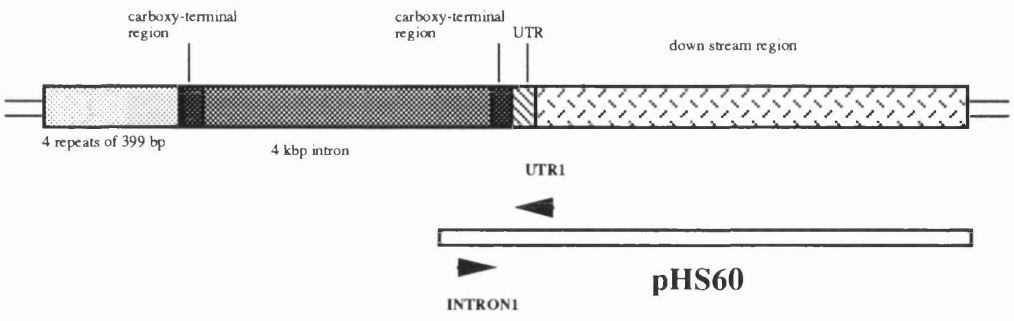


Figure 5.12. Diagrammatic illustration of where oligonucleotide primers UTR1 and CBR1 bind within the cDNA insert of clone pHS10. CBR1 and UTR1 were used to sequence the *aba-1* genomic DNA sequences within clones pHS50 and pHS60 (see Figures 5.16 and 5.14 respectively).



ggatgtaaatttcaagtattttgagcattaaccaattctt

gggagtgaattcatgctaaaaataatgcataaaaagtctt

ctttcaacacaggtaggtattgaaatcttcaactatacaa

ttccagGATGACATTGACGGGCTAGCGCAAAGTCGGCAGC
D D I D G L A Q S R Q R

GACGATCCGGATTTTTCGAGAAGCTCATCGATGTGTTTGC
R S G F F E K L I D V F A

ATTTTTTTGAGTGAATTCCCATGTTACGCGAATCGTTATG
F F * UTR Sequence→

TCATATCAACAGATGATAGTAAATAATCACAACCACATGT

TCACATGAAGACAAACCATTTGCATTTGCGCTTATAA

Figure 5.13. DNA sequence derived from clone pHS60 using primers UTR1 and INTRON1. Lowercase letters represents intronic sequence and the underlined region represents the putative splice acceptor sequence. The untranslated region (UTR) of the gene is indicated. The schematic diagram at top illustrates the region of pHS60 sequenced (the positions that the primers bind are shown as arrows) and how this relates to the *aba-1* genomic clone λHSg7.

A

```
pHS10   RRRRYHAEDGTDIDGALAQSRQRRSGFFEKLIDVF
pHS60               -------------------------

pHS10   AFF*
pHS60   ---*
```

B

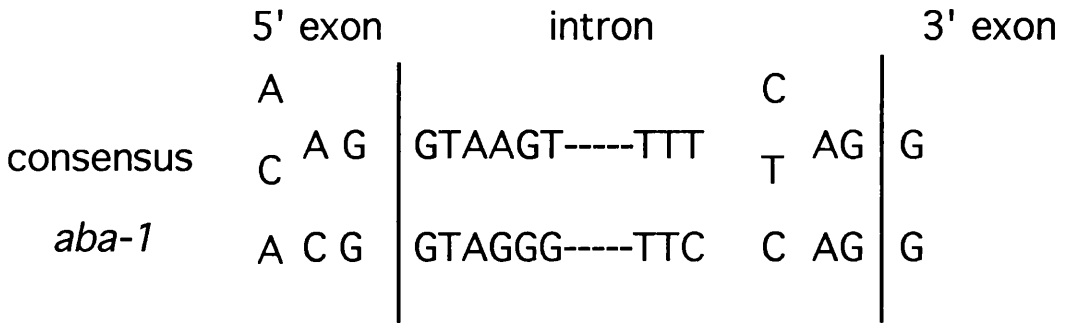
```
pHS10    cactttaaggcatgttacgcgaatcgttatgtcatatcaaca
pHS60    ------------------------------------------

         gatgatagtaaataatcacaaccacaaaaaaaaaaaaaaaaa
pHS10    ---------------------------tgttcacatgaagcac
pHS60
pHS10    aaaaaaaaaaaaaaaaaaaaaaaaa
pHS60    aaaccatttgcatttgcgcttataa
```

Figure 5.14. (A) Comparison of the amino acid sequences encoded by the carboxy-terminal regions of the inserts of pHS10 and pHS60. (B) Comparison of the DNA sequences of the untranslated regions of pHS10 and pHS60. In both diagrams, positions of sequence identity are represented by (-). In (A) the sequence comparison starts at the 12th amino acid of the carboxy-terminal region

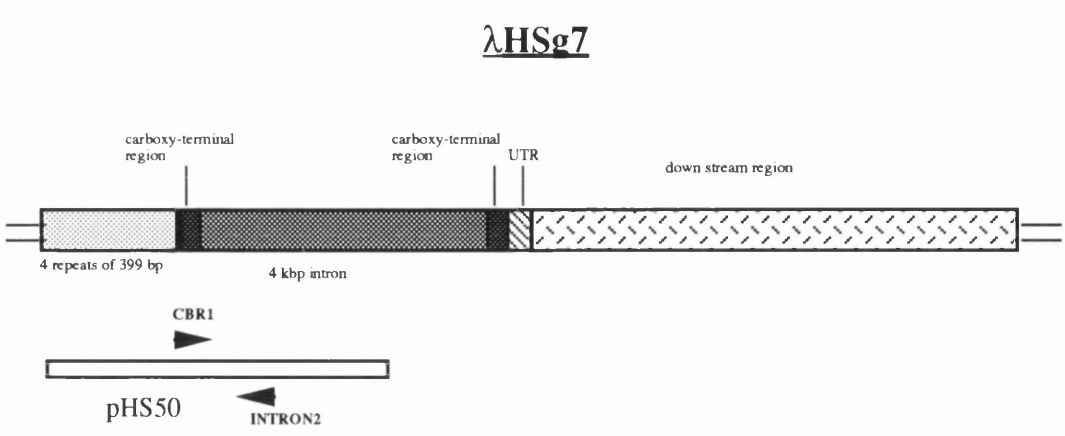
present at the predicted intron-exon boundary (see Figure 5.15). To verify this finding the putative intron/exon area was sequenced on both the coding and non-coding strands. This was achieved by designing an oligonucleotide primer, named INTRON1, that hybridised to the intron sequence and therefore allowed sequencing into the 3' untranslated region. The sequence derived from INTRON1 revealed that it was 100% complementary to that derived from UTR1, providing more evidence for the existence of an intron at this position within the *aba-1* gene.

To analyse the 5' region of the putative intron, an oligonucleotide primer, named CBR1, was designed to hybridise to the first 19 bp of the carboxy terminal region (see figure 5.12). This primer was used to sequence both pHS50 and 60, but sequencing only worked on pHS50, implying that the 5' region of the carboxy terminal region was situated in this clone. Analysis of the sequence derived from CBR1 revealed that it showed no similarity to the pHS10 insert, suggesting that it was the 5' sequence of the intron discussed above. To verify this, a further primer (INTRON2) was designed which hybridised to the intronic sequence and allowed sequencing back into the exon. Sequence obtained from INTRON2 showed 100% identity to the 5' 21 bp of the carboxy-terminal region and to the most 3' repeat of pHS10 (see Figures 5.16 and 5.17). Analysis of this putative exon-intron boundary (Figure 5.15) showed a consensus exon/intron donor site (Bianco *et al.*, 1993). Taken with the above data, this suggests that there is an intron present between the 21st and 22nd bp of the carboxy terminal region of *aba-1*.

At this stage the size of the intron was not known, and therefore a polymerase chain reaction was performed using one primer (pHS10N) designed to hybridise to the extreme 5' region of a repeat from pHS10, and a second primer (UTR1) which hybridises to the untranslated region (see Figure 5.18 A). This reaction resulted in a DNA fragment of approximately



5.15. Putative consensus exon/intron and intron/exon sites. The figure compares exon/intron and intron/exon boundaries of the *A. suum aba-1* gene to the nematode consensus sites derived by Hammond and Bianco (1992).



GAGGAGAAGAAGCAATACATCGCCGATTTTGGACCAGCATGCAA
E E K K Q Y I A D F G P A C K

GAAAATCTATGGTGTACATACTTCGCGACGAAGGAGGTATCACG
K I Y G V H T S R R R R Y H A

CTGAGGATGGAACGgtaggg cgttactttgaaacagatctgaag
E D G T intronic sequence →

tggctctctgcaaatgaaattgaggagatgaggaaaatccgtag

tgggatgtaagcaaa

Figure 5.16. DNA sequence and putative amino acid sequence derived from the genomic clone pHS50 using primers CBR1 and INTRON2. Lowercase letters illustrate intronic sequence, while the sequence underlined represents the putative splice-donor sequence at the exon/intron boundary. The schematic diagram at top illustrates the region of pHS10 sequenced and how this relates to the *aba-1* genomic DNA in clone λHSg7; the positions that the oligonucleotide primers CBR1 and INTRON2 hybridise to pHS50 are shown by small arrows.

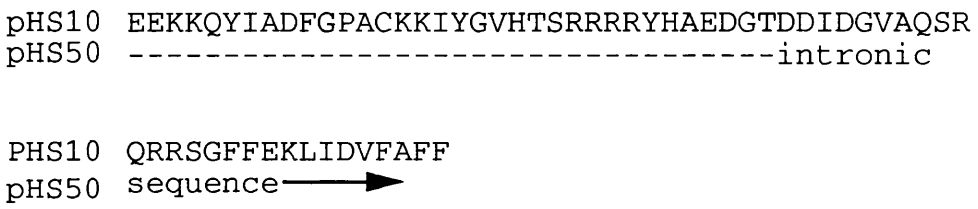
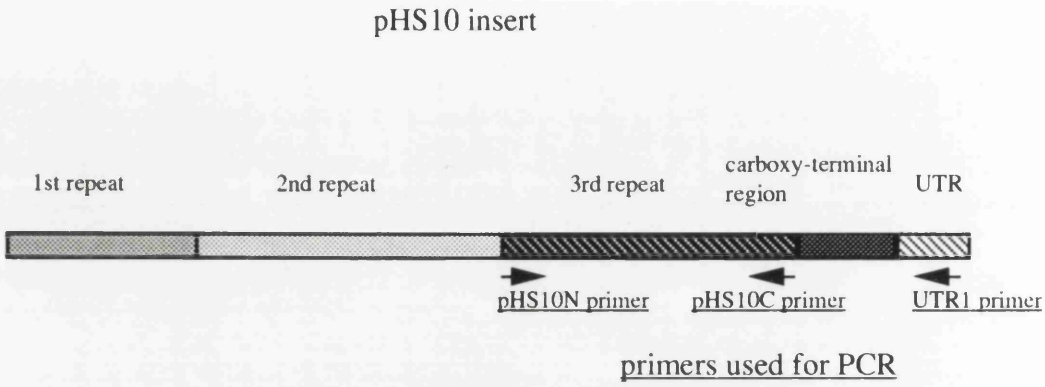
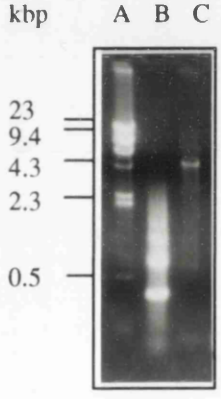


Figure 5.17. Comparison of the putative amino acid sequences derived from cDNA clone pHS10 and genomic clone pHS50. Where the amino acid sequence encoded by pHS10 is identical to that of pHS50 it is represented as (-), while differences are shown by the amino acid residues' single letter code. Also highlighted is the position within the pHS50 carboxy terminal open reading frame where the intronic DNA sequence begins.

Figure 5.18. PCR reactions performed on the *A.suum* genomic DNA clone λHSg7. The upper schematic diagram shows the regions that the PCR primers used to characterise the λHSg7 intronic sequence are predicted to bind to the *aba-1* insert of cDNA clone pHS10. The lower gel shows the PCR products derived from the primers detailed in the upper diagram acting on λHSg7. PCR reaction products were electrophoresed on a 0.8% agarose gel. Lane (A) is λ HindIII/ EcoRI markers (in kbp); lane (B) is a control PCR using primers HS10N and HS10C, which results in the amplification of 399 bp *aba-1* repeats; lane (C) is a PCR reaction using primers HS10 and UTR1, which was performed to determine the size of the intron present in λHSg7.


pHS10 insert
1st repeat
2nd repeat
3rd repeat
carboxy-terminal region
UTR
pHS10N primer
pHS10C primer
UTR1 primer
primers used for PCR

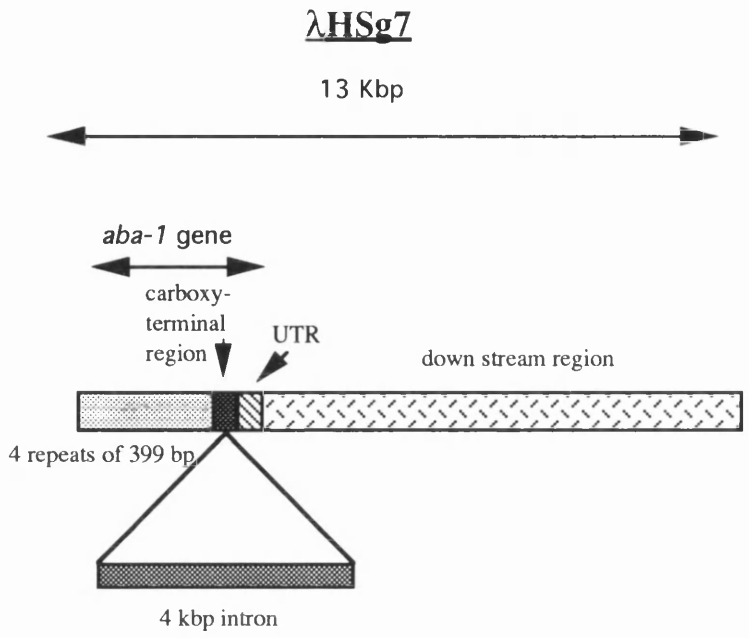

kbp
A B C
23
9.4
4.3
2.3
0.5

4.5 kbp, suggesting that the intron is approximately 4 kbp in length (see Figure 5.18 B).

The sequencing data described above indicate that the extreme 5' region of the λHSg7 insert consists of a truncated repeat of the *aba-1* gene followed by four complete repeats. After the most 3' repeat there is 21 bp of the carboxy terminal region followed by an intron of approximately 4 kbp. After the intron the remaining 81 bp of the carboxy terminal region is present and is followed by the 3' untranslated sequence. The organisation of this clone is cartooned in Figure 5.19. It should be noted that these data confirm the authenticity of the carboxy terminal region, which was thought could have been an artefact of the cDNA cloning procedure (see Section 3.2.6).

5.2.8 Isolation of clone λHSg8

It was apparent from the data reported in Section 5.2.7 that the genomic clone λHSg7 did not contain the entire *aba-1* gene, since the 5' region is not present. This implied that the total number of 399 bp repeats present within the *aba-1* gene could not be accurately determined and also that the putative 5' promoter sequences could not be characterised using λHSg7. Attempts were therefore made to isolate genomic clones containing the complete *aba-1* gene, which involved re-screening the 10 *aba-1*-positive plaques (isolated after probing the unamplified *A. suum* genomic library with the pHS10 insert; see Section 5.2.3) with the divergent repeat. This approach was used as it was thought that screening with the divergent clone would hopefully isolate clones containing 5' sequences, since previous data suggested that the divergent repeat is found upstream of the repeats present within pHS10 (see Section 5.2.2). The 10 positive



5.19. Schematic representation of the organisation of the *aba-1* genomic clone λHSg7. The figure shows that clone λHSg7 consists of 4 repeats of 399 bp followed by 21 bp of the carboxy-terminal region (which is interrupted by a 4 kbp intron). Following the intronic sequence is the remaining 81 bp of the carboxy-terminal region and then the untranslated region (shown as UTR).

plaques isolated from the initial screening were used in preference to the amplified library as it was thought that the amplified library may contain clones within which the *aba-1* repeats had recombined. After secondary and tertiary screenings, one further genomic clone, designated λHSg8, was isolated.

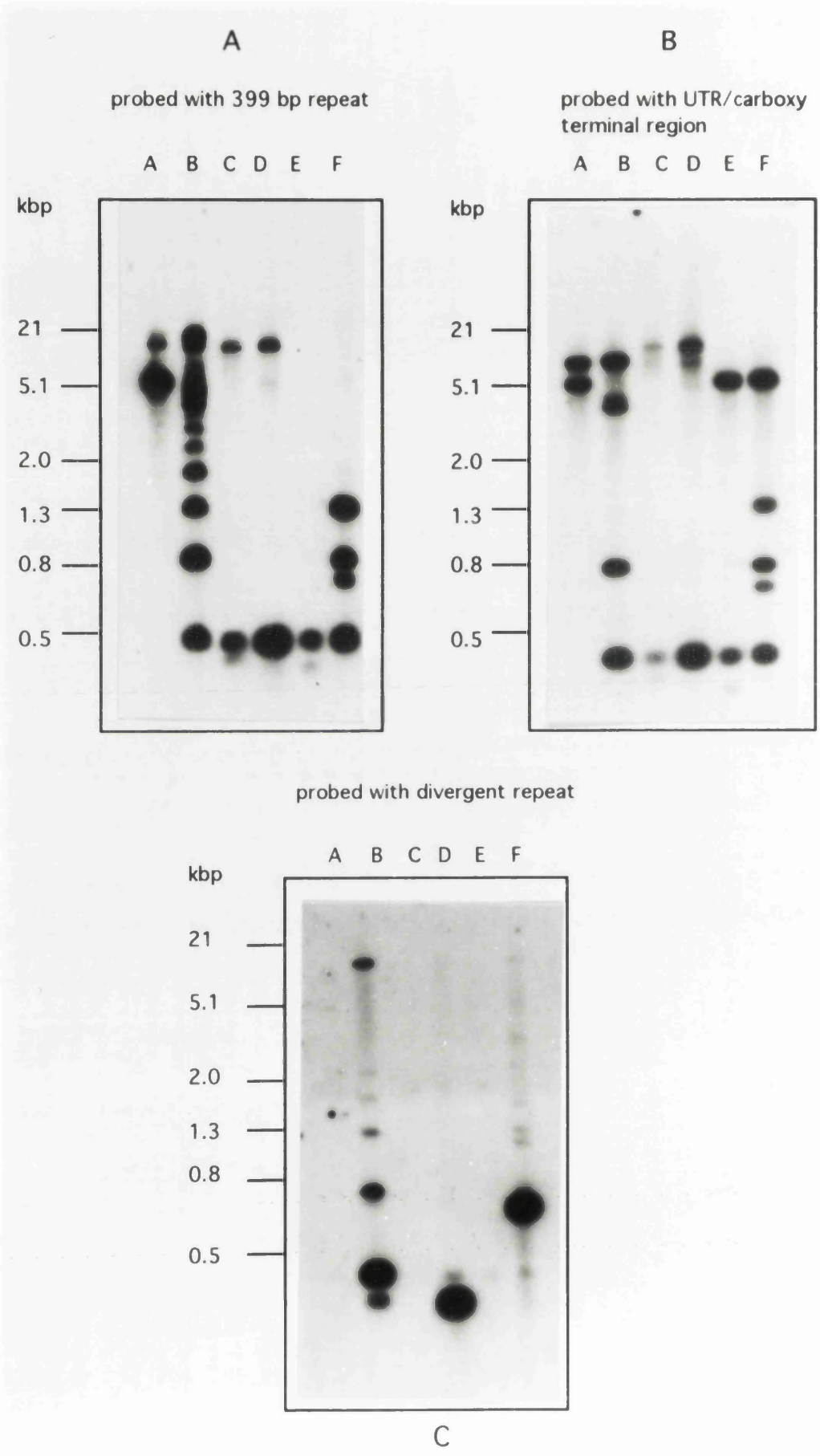
5.2.9 Characterisation of genomic clone λHSg8

Lambda DNA was prepared from genomic clone λHSg8 to allow analysis of the clone's insert, and Southern blots were performed after the DNA had been digested with restriction enzymes which do, and do not, cut within the 399 bp repeats of cDNA clone pHS10 (Figure 5.20). These blots were probed with 3 different *aba-1* sequences: (A) a single 399 bp repeat of pHS10; (B) the 3' sequence of pHS10 consisting of the carboxy terminal region and untranslated region; and (C) the divergent repeat (Figure 5.21). Different probes were used because it was thought that they would determine if repeating units like those found in clone pHS10 and the genomic clones λHSg6 and λHSg7 were present, and where the divergent repeat was situated, as well as determining if the 3' untranslated region of the *aba-1* gene was present. Clone λHSg7 was analysed in conjunction with λHSg8 to allow comparison of their restriction digestion profiles and thus aid characterisation of clone λHSg8.

These blots revealed that λHSg8 contains a genomic fragment of 19 kbp. DNA fragments of 400 bp (which is the same size as the repeats present within pHS10 and genomic clones λHSg6 and λHSg7) were recognised by the 399 bp repeat of pHS10 when the clone was digested with either *Xba*1 or *Nru*1 in conjunction with *Bam*HI (Figure 5.20A; lanes D and). Both *Xba*I and *Nru*I cut within a 399 bp repeat of *aba-1*, therefore

Figure 5.20. Southern blot analysis of λHSg8 and λHSg7. λHSg8 and λHSg7 were digested with various restriction enzymes (listed below) and then electrophoresed on a 0.8 % agarose gel. The gel was capillary blotted onto Hybond N and probed with a 399 bp *aba-1* repeat **(A)**, the *aba-1* carboxy-terminal and UTR region **(B)**, and the *aba-1* divergent repeat of clone pJM33 **(C)**.

Lane	clone	Digest
A	λHSg7	*Sac*I
B	λHSg8	*Sac*I
C	λHSg7	*Xba*I and *Bam*HI
D	λHSg8	*Xba*I and *Bam*HI
E	λHSg7	*Nru*I and *Bam*HI
F	λHSg8	*Nru*I and *Bam*HI


A
probed with 399 bp repeat
A B C D E F
kbp
21
5.1
2.0
1.3
0.8
0.5
B
probed with UTR/carboxy terminal region
A B C D E F
kbp
21
5.1
2.0
1.3
0.8
0.5
probed with divergent repeat
A B C D E F
kbp
21
5.1
2.0
1.3
0.8
0.5
C

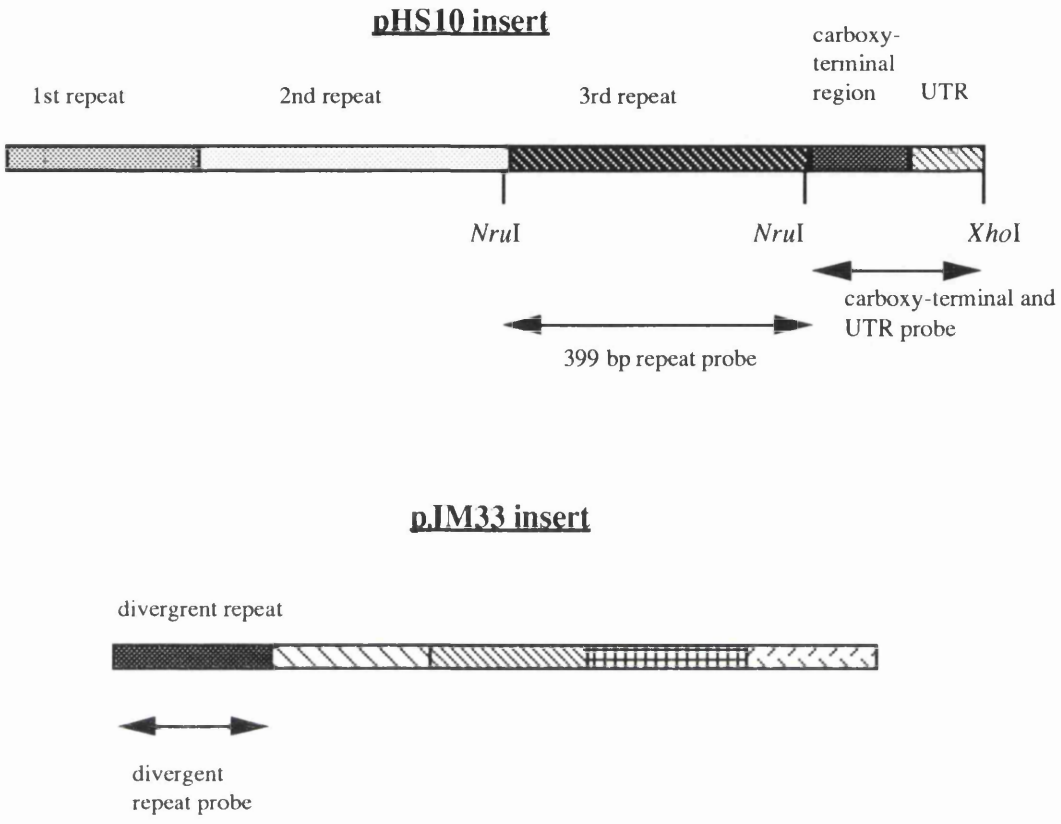


Figure 5.21. Diagrammatic representation of the probes used for the Southern blots shown in Figure 5.20. The extent of the sequences included in each of the probes, and the region of the cDNA clones from which they are derived, are indicated by arrowed lines. The 399 bp probe was isolated by *Nru*I digestion of pHS10, and the carboxy-terminal and untranslated region (UTR) probe was isolated by *Nru*I and *Xho*I digestion of the same cDNA clone. The divergent probe was a gift from J. Moore.

confirming that λHSg8's genomic insert contains the *aba-1* gene. Note that double digestions with *Bam*HI were used here because this enzyme cleaves within the λ multiple cloning sites (unlike *Nru*I and *Xba*I), but not within the *aba-1* repeats, and therefore releases the *aba-1* genomic insert (data not shown). Digestion of the clone with *Sac*I also revealed a fragment of 400 bp, despite the fact there are no *Sac*I recognition sequences present in any of the *aba-1* repeats so far characterised (Figure 5.20, lane B). This therefore implies that 5' to the repeats characterised in clones pHS10 and λHSg7 there are repeats which contain *Sac*I sites, and suggests that λHSg8 contains 5' sequences of the *aba-1* gene that are not present in any of the other genomic or cDNA clones isolated.

Analysis of the blots probed with the divergent repeat (Figure 5.20 C) revealed that the divergent repeat is situated at the extreme 5' region of λHSg8's genomic sequence. Figure 5.20 C also shows that the divergent repeat is not present in clone λHSg7, illustrating that λHSg8 contains more 5' sequence of the *aba-1* gene.

Figure 5.20 B illustrates that the 3' probe hybridises to λHSg8, suggesting that this clone contains the extreme 3' region of the *aba-1* gene. The presence of the 5 kbp *Nru*1 fragment described in Section 2.2.5 suggests also that this clone contains the intron which is present within the insert of λHSg7 (Figure 5.20 B, lane F). This may be concluded because it is known that this fragment contains the intronic sequence of the λHSg7 clone (see Figure 5.19). It should be noted that the 3' probe appeared to recognise the 400 bp fragments (see Figure 5.20 B, lanes A, B, C, D, E and F) which are thought to represent the *aba-1* repeating units; the reason for this may be that the probe contains 10 bp of the 399 bp repeat sequence.

To determine exactly how many repeating units are present within λHSg8, Southern blots of DNA partially digested with *Xba*I (which cuts once within the 399 bp repeat of pHS10) were probed with a single 399 bp

repeat from pHS10 (see Figure 5.22). Bands of approximately 400 bp, 800 bp, 1200 bp, 1600 bp, 2000 bp, 2400 bp, 2800 bp, 3200 bp, 3600 bp and 4000 bp were detected, which probably represent 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 repeating units of the *aba-1* gene respectively. These data therefore suggest that clone λHSg8 consists of at least 10 repeating units. Because the experiments in Section 5.2.5 suggested that the four 399 bp repeats in λHSg7 do not contain *Sac*I recognition sequences, but *Sac*I appeared to be able to digest the repeats within λHSg8 (Figurex 5.20), this partial restriction digestion analysis was repeated using *Sac*I rather than *Xba*I (Figure 5.23). The same range of DNA fragments, representing 1 to 10 *aba-1* repeats, were again seen. When combined with the evidence that a complete *Sac*I digest of λHSg8 also creates a DNA fragment of approximately 3 kbp (which is thought to contain 5 repeats without *Sac*I recognition sites) it may be concluded that the *aba-1* sequence of this clone contains approximately fifteen 399 bp *aba-1* repeats. Figure 5.24 shows a schematic diagram of λHSg8.

Figure 5.25 compares the *aba-1* sequences present in the genomic clones λHSg8, λHSg7, and in the cDNA clone pHS10. This illustrates that the 3' noncoding region of the *aba-1* gene has been fully characterised and contains a 4 kbp intron, but the 5' region has not been isolated, although we now have evidence that the *aba-1* gene is composed of at least 15 repeating units which are uninterrupted by intronic sequences. In addition, the divergent repeat lies 5' to the pHS10-like repeats.

Figure 5.22. Southern analysis of partially-digested λHSg8. λHSg8 was partially digested with the following concentrations of *XbaI* for 1 hr: 0.2 units (A); 0.5 unit (B); 1 units (C); 5 units (D). Samples were electrophoresed on a 0.8% agarose gel and blotted onto Hybond N. The blot was probed with the pHS10 insert. λ *Hind*III/ *Eco*RI markers (in kbp) electrophoresed on the same gel are shown at the left of the figure.

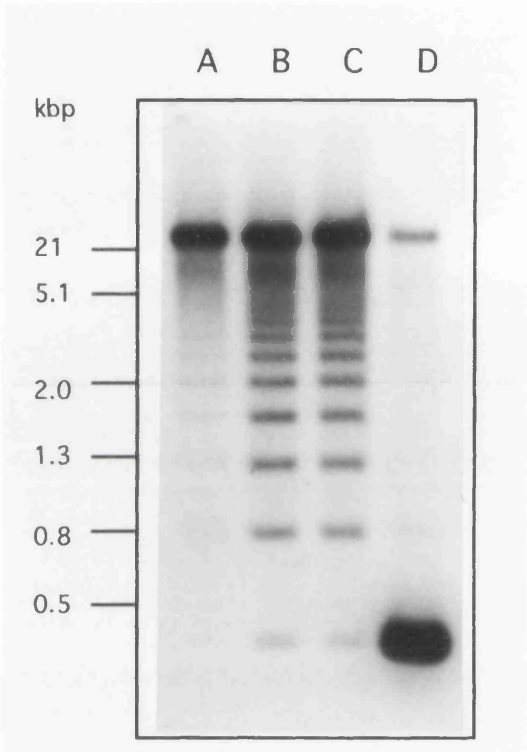
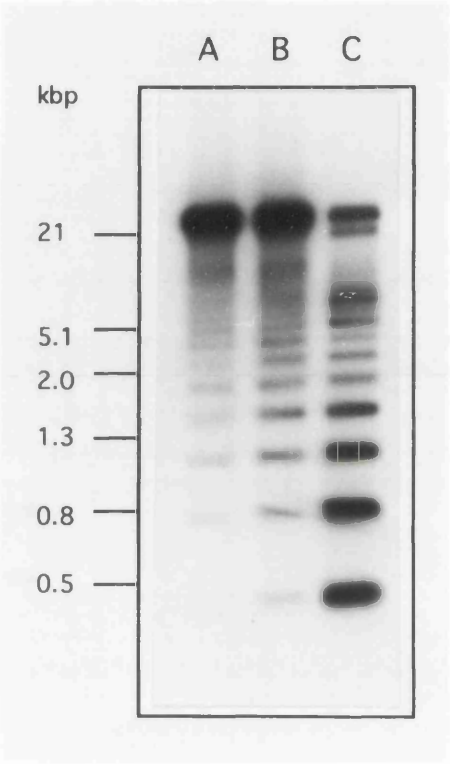

A
B
C
D
kbp
21
5.1
2.0
1.3
0.8
0.5

Figure 5.23. Southern analysis of partially-digested λHSg8. λHSg8 was partially digested with the following concentrations of *Sac*I for 1 hr: 0.5 units (A); 1 unit (B); 5 units (C). Samples were run on a 0.8% agarose gel and blotted onto Hybond N. The blot was probed with the pHS10 insert. λ *Hind*III/ *Eco*RI markers (in kbp) run on the same gel are shown at the left of the figure.


A
B
C
kbp
21
5.1
2.0
1.3
0.8
0.5

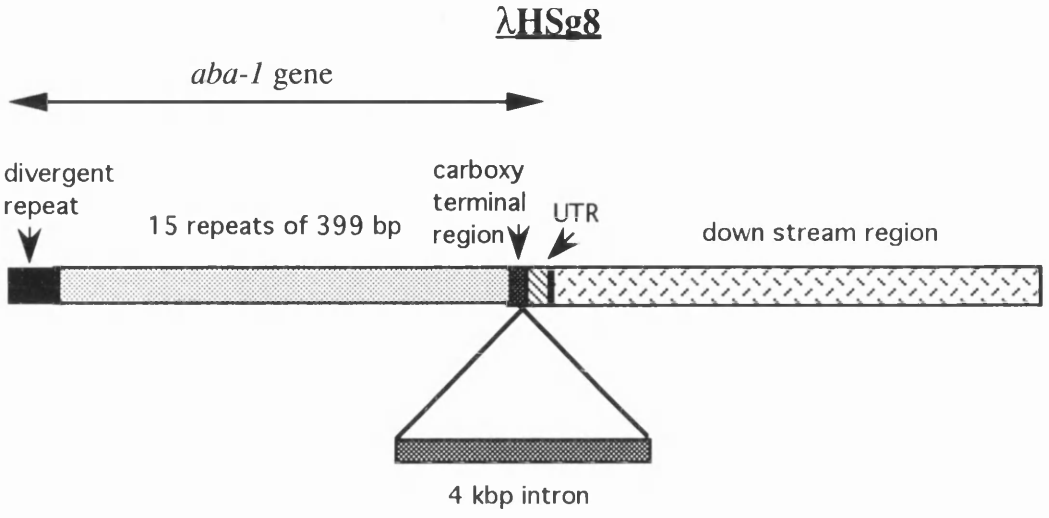


Figure 5.24. Schematic representation of the organisation of the *aba-1* genomic clone λHSg8. λHSg8 is 19 kbp in size, and contains approximately 15 copies of the *aba-1* 399 bp repeats and a 5' repeat which is believed to be the divergent repeat characterised in the cDNA clone pJM33. In addition, the clone contains the carboxy-terminal region (interrupted by a 4 kbp intron), the 3' untranslated region of the *aba-1* gene (UTR) and the downstream region that contains sequence unrelated to *aba-1*.

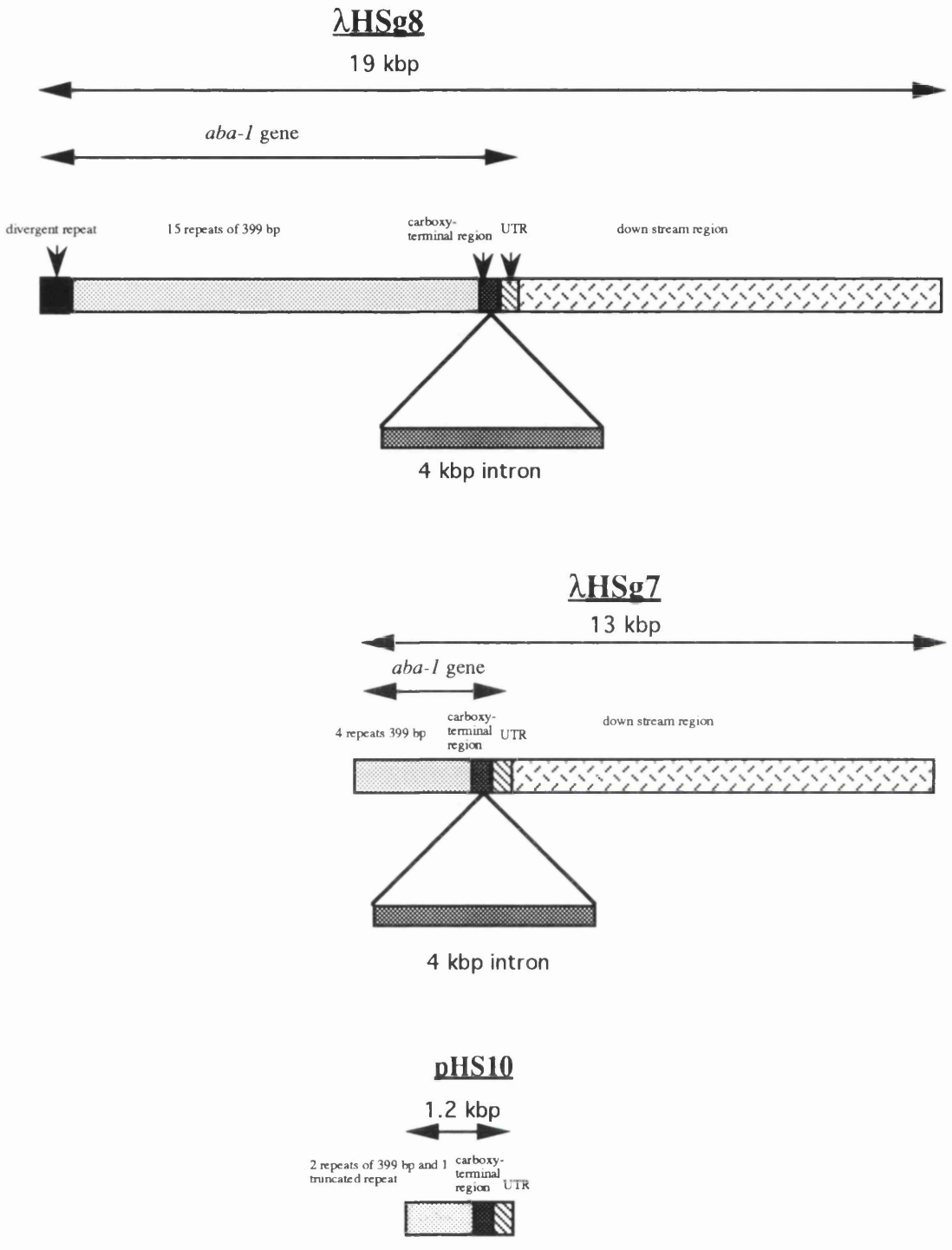


Figure 5.25. Diagrammatic comparison of the *aba-1*-containing clones pHS10, λHSg7 and λHSg8. pHS10 is an *aba-1* cDNA clone described in Figure 5.1, while λHSg7 and λHSg8 are genomic *aba-1* clones that are described in Figure 5.19 and 5.24 respectively.

5.3 Discussion

This chapter reports two experimental approaches taken to analyse the genomic organisation of the *A. suum aba-1* gene. The first approach involved Southern blot analysis of total genomic DNA of *A. suum* and revealed that, like the cDNA clone of pHS10, the genomic copy of the *aba-1* gene is composed of repeating units of 399 bp. Consequently, the repeating nature of the cDNA can be said not to be an artefact produced when the cDNA library was constructed. Southern data also suggested that the *aba-1* gene lies within a 9 kbp *Eco*RI DNA fragment, implying that it could be composed of as many as 18 repeats. This estimate is similar to that predicted for the number of repeats present in the *B. pahangi* and *D. immitis* homologues of *aba-1* (Tweedie *et al.*, 1993; Culpepper *et al.*, 1992; and Poole *et al.*, 1992), though it should be stressed that in no case has the exact number been determined.

An intruiging result was seen with the Southern analysis involving *Sau*3A digestion of *A. suum* DNA, since a series of fragments were seen rather than the single band that would have been predicted given that this enzyme does not cut within a 399 bp repeat of the cDNA in pHS10. These data were interpreted as meaning that either the extreme 5' repeats (which have not been characterised) contain *Sau*3A sites not present in the 3' repeats or that individual worms contain polymorphisms within their repeats. To determine if the latter explaination is correct, future work could involve isolating genomic DNA from individual worms and then performing Southern analysis with various enzymes, including *Sau*3A. If individual worms show different restriction profiles this would suggest that there are indeed sequence polymorphisms between individual worms, although the reason why different worms should contain 399 bp repeats with unique restriction sites is unclear at this stage. It is possible

that this may be a consequence of the fact that such a large number of ABA-1 encoding units are subject to mutations over time that do not affect function and are consequently not repaired.

Southern blot analysis also revealed that the divergent repeat (which only shows 49% identity to pHS10 repeat; see Chapter 3) is situated within the same gene as the pHS10-like repeats. These data reveal that this divergent clone is not an artefact, but why this repeat is so different in sequence to the other pHS10-like repeats remains an unanswered question. Similarly, it is not clear whether the ABA-1 protein encoded by the divergent repeat differs in its function to the pHS10-encoded ABA-1 proteins. Perhaps when further biochemical work has been carried out to determine the function of ABA-1 these questions can be addressed.

The second approach taken to analyse the genomic organisation of the *A. suum aba-1* gene involved the characterisation of two genomic clones, λHSg7 and λHSg8. These clones were found to contain repeating units of 399 bp: four in λHSg7 and approximately 15 in λHSg8. Analysis of the restriction profiles of both clones and partial sequencing of the repeats present in clone λHSg7 revealed their DNA sequences show between 86% and 100% identity to those found in the cDNA clone pHS10. These data provide further evidence that the repeating structure of the cDNA pHS10 is not an artefact. The finding that genomic clone λHSg8 contains approximately 15 repeats provides further evidence that the *aba-1* gene contains as many repeating units as its homologues described in *B. pahangi* and *D. immitis* (Tweedie *et al.*, 1993; Culpepper *et al.*, 1992; and Poole *et al.*, 1992).

Sequencing of clone λHSg7 revealed that the carboxy terminal region of the *aba-1* gene, which was identified in the cDNA copy of the gene (see Chapter 3), is present in the genomic copy of the *aba-1* gene. These data imply that the carboxy-terminal region is not an artefact but

may indeed have some functional relevance (see Chapter 3 for discussion of the function of this region). Furthermore, the presence of this sequence directly downstream of the most 3' *aba-1* repeat implies that this region is not spliced on to the end of the mRNA transcript and is not encoded at a separate genomic site.

Of particular interest is that the carboxy terminal region of the genomic copy of *aba-1* is interrupted by an approximately 4 kbp intronic sequence. Evidence for the presence of intronic sequence is that at the proposed 5' exon/intron boundary and the downstream intron/exon boundary, putative consensus splice-donor and splice-acceptor sites are present (see Figure 5.15). These sites are found in many eukaryotic genes that utilise *cis*-splicing to remove introns and create intact mRNAs (Mount *et al.*, 1982). Hammond and Bianco (1992) have compared the splice-donor and splice-acceptor sites of a total of 55 nematodes genes and produced consensus sequences for these sites (Figure 5.15). Comparison of the putative *aba-1* splice-donor and splice-acceptor sequences to these consensus reveals that they have 66% and 86% identity respectively, although it does not seem likely that the lack of complete identity is significant. Support for this conclusion is that the putative splice-donor/splice-acceptor sequences found at the exon/intron and the intron/exon boundaries of Ascaris collagen genes also do not fit their respective consensus sequences (Kingston *et al.*, 1989).

Although we have found evidence that the *aba-1* gene utilises *cis*-splicing during its expression, it is conceivable that it may utilise *trans*-splicing also. *Trans*-splicing is thought to remove intronic sequences called outrons present at the extreme 5' regions of 90% of nematode genes and concomitantly inserts splice leader sequences (Hannon *et al.*, 1990). Therefore it is possible that the *aba-1* gene utilises *trans*-splicing to splice out putative 5' introns while it also uses *cis*-splicing to splice remove

downstream intronic sequences. At present we cannot say if this is the case since neither the extreme 5' sequence of the genomic copy of the *aba-1* gene, nor the extreme 5' region of its mRNA transcript have been characterised. Analysis of this region would be possible using an oligonucleotide primer designed to hybridise to the splice leader sequence that is found at the extreme 5' region of all *trans*-spliced mRNA (Hannon *et al.*, 1990) in conjunction with an *aba-1* specific primer to perform a Polymerase Chain Reaction.

Another interesting feature of the *aba-1* intronic 4 kbp sequence is that it is considerably larger than most of the nematode introns studied to date, most of which have an average size of a few hundred base pairs (for review see Hammond and Bianco, 1993). At present the significance of this finding is not known, but perhaps when more genomic copies of nematode genes have been characterised it will become clear. It will also be of interest to determine if the homologues of the *aba-1* gene in other nematodes have an intronic sequence at the same 3' position and whether they are equally large; this would be especially true in *C. elegans* where introns have consistently found to be very small (approximately 50 bp).

Although a lot of additional information has been obtained from analysis of clones λHSg7 and λHSg8, neither of them contain the extreme 5' region of the *aba-1* gene, meaning that the following questions still remain unanswered: exactly how many 399 bp repeats are present; are there further divergent repeats 5' to the one characterised (and if so do they all have the same sequence); and does the 5' upstream region contain novel promoter and enhancer sequences? To enable these question to be addressed a genomic clone would have to be isolated that contains yet more 5' sequence. Two approaches could be taken to achieve this. Firstly, another genomic library could be constructed and screened with the divergent clone. It would be necessary to construct another genomic

library for this because the possibility of recombination having occurred between the *aba-1* repeats in the amplified remainder of the library used in this work cannot be excluded. The divergent clone should again be utilised for screening because we know that it lies in the 5' region of the gene; i.e. it is positioned at least 14 repeats from the 3' untranslated region. The second possible approach would be to use the 5' "Rapid Amplification of cDNA Ends" (RACE) PCR technique which can be used to examine the extreme 5' regions of mRNA molecules. To perform this analysis, an oligonucleotide primer designed to hybridise within the divergent repeat should be used because this is the most 5' sequence characterised and is sufficiently different to the other repeats to allow specific PCR amplification.

Chapter 6

Characterisation of the *A. lumbricoides aba-1* gene

6.1 Introduction

For many years there has been considerable debate over whether *A. lumbricoides* and *A. suum,* which are thought to infect humans and pigs respectively, are genuinely separate *Ascaris* species (reviewed by Crompton, 1989). The main reasons for this debate are that they are indistinguishable on physical examination, and experiments have suggested that *A. suum* can in fact infect humans and *A. lumbricoides* can infect pigs (see Chapter 1 for further discussion). Over the past 20 years immunological, biochemical and molecular biology work has been performed to try and elucidate whether or not these species are distiguishable from one another (see Chapter 1). Despite this body of research, the debate continues today.

Kennedy *et al.* (1986 and 1989) have reported that the 14 kDa ABA-1 protein of both *A. suum* and *A. lumbricoides* is excreted by all developmental stages of both putative species and induces both an IgG and IgE response in their respective hosts. N-terminal amino acid sequencing of both species' ABA-1 proteins revealed that they have 100% identity in this region, confirming that they are homologues. This data would appear to support the hypothesis that *A. suum* and *A. lumbricoides* are one species, but only a very small part of the protein has been characterised.

To analyse whether there are sequence differences between the remainder of the ABA-1 protein, and between the *aba-1* genes, this chapter reports Southern blot analysis of the *A. lumbricoides aba-1* gene, molecular cloning and characterisation of part of this gene by PCR and comparison of its sequence to the *A. suum* sequence (which has been extensively characterised; see Chapter 3, 4 and 5). Western blot analysis of

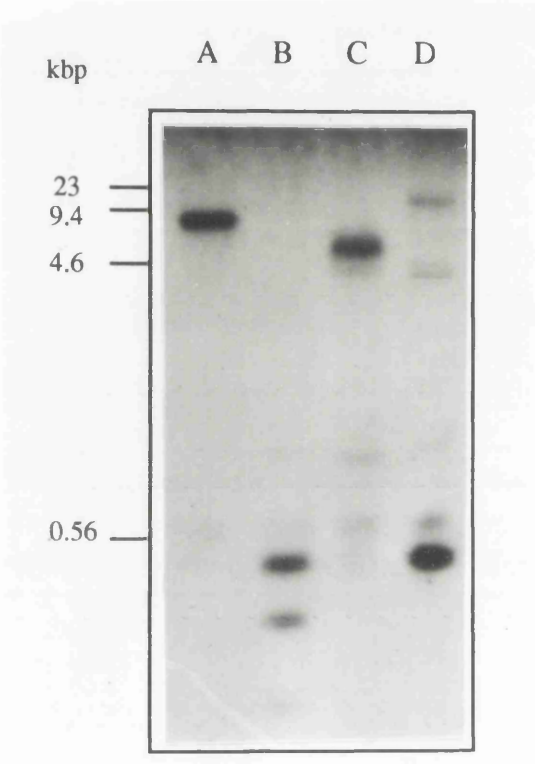
the native *A. lumbricoides* adult body fluid is also reported in this chapter to determine if, like the *A. suum* protein, *A. lumbricoides* ABA-1 is initially expressed as a polyprotein *in vivo* (see Chapter 4).

6.2 Results

6.2.1 Southern blot analysis of the *aba-1* gene of *A. lumbricoides*

To determine whether the *A. lumbricoides aba-1* gene has a similar genomic organisation to its homologue in *A. suum*, genomic DNA was isolated from embryonated eggs and digested with restriction enzymes that either cut or do not cut within the 399 bp repeats of the *A. suum aba-1* gene. A Southern blot of this DNA probed with the insert of cDNA clone pHS10 (see Chapters 3 and 4) is shown in Figure 6.1. Digestion with *Nru*I (Figure 6.1, lane D) and *Xba*I (Figure 6.1, lane B), both of which cut once within a pHS10-type 399 bp repeat of the *A. suum aba-1* gene, resulted in an *A. lumbricoides* DNA fragment of 400 bp. Since this fragment is approximately the same size as one repeat from *A. suum*, these data suggest that the *A. lumbricoides* gene may also be composed of repeating units of approximately the same size. It should be noted that a DNA fragment of approximately 200 bp was also recognised by the probe when the DNA was digested with *Xba*I one (Figure 6.1 lane B). At this stage there is not enough information to determine what part of the *aba-1* gene this fragment represented. Restriction digestion with *Eco*RI (Figure 6.1, lane A), which does not cut within the characterised *A. suum aba-1* repeats, resulted in a single DNA fragment of approximately 9 kbp, which probably represents the full length, of near full length, *A. lumbricoides aba-1* gene. Southern blots of *A. suum* genomic DNA digested with *Eco*RI and probed with pHS10 also resulted in a single DNA fragment of approximately 9 kbp

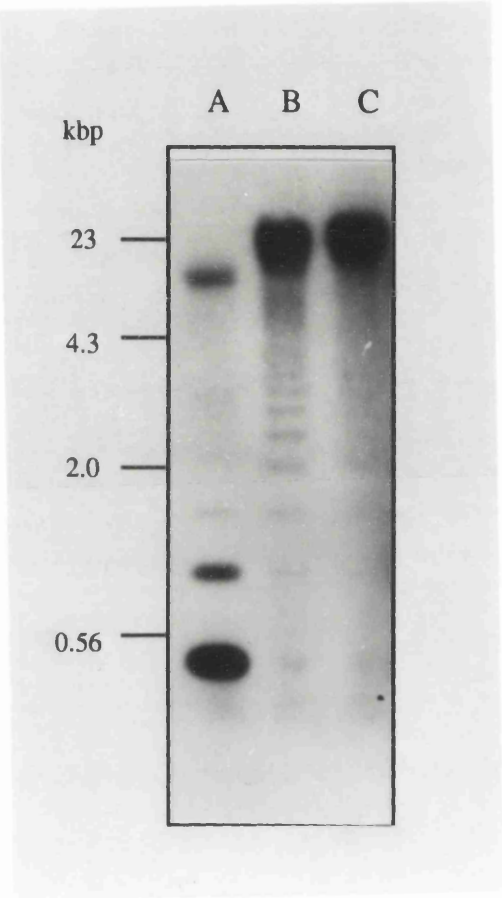
Figure 6.1. Southern blot analysis of the genomic organisation of the *A. lumbricoides aba-1* gene. *A. lumbricoides* genomic DNA was digested with *Eco*RI (A), *Xba*I (B), *Sau*3A (C) or *Nru*I(D) and electrophoresed on a 0.8% agarose gel. The gel was capillary blotted onto Hybond N and probed with the *aba-1* insert of pHS10. λ *Hind*III/ *Eco*RI markers (in kbp) run on the same gel are shown at the left of the figure.


kbp
A
B
C
D
23
9.4
4.6
0.56

(Chapter 5, Figure 5.1), suggesting, firstly, that the homologous genes contain a similar number of 399 bp repeats, and, secondly, that there is no *Eco*RI site within either of the *aba-1* genes. In contrast, digestion of genomic DNA from the two putative species with *Sau*3A (which does not cut within the characterised repeats of the cDNA clone pHS10) did not result in the same restriction pattern: a series of fragments ranging in size from approximately 700 bp to 5000 bp were seen in *A. suum* (Chapter 5, Figure 5.1), but only a single fragment of 5000 bp was found in *A. lumbricoides* (Figure 6.1, lane C). Although there are no *Sau*3A sites within the characterised *A. suum aba-1* repeats, previous Southern analysis suggested that the 5' repeats in the gene may contain *Sau*3A sites and/or that there may be polymorphisims between *aba-1* genes of different worms. These data suggested that this is not the case for *A. lumbricoides*, since there appears to be no *Sau*3A sites within the *aba-1* gene, or at least within the sample of worms from which the DNA was obtained.

The above Southern analysis suggested that, like the *A. suum aba-1* gene, the *A. lumbricoides aba-1* gene is composed of approximately 400 bp repeating units and that the two genes have a similar overall size. Despite this knowledge, the actual number of repeats present within the *A. lumbricoides aba-1* gene was unkown. A Southern blot of *A. lumbricoides* genomic DNA partially digested with *Xba*I was therefore prepared and probed with the pHS10 insert (Figure 6.2). DNA fragments of approximately 400 bp, 800 bp, 1200 bp, 1600 bp, 2000 bp, 2400 bp, 2800 bp, 3200 bp, 3600 bp and 4000 bp were detected, which were interpreted as representing 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 repeating units of the *aba-1* gene respectively. These data imply that the *A. lumbricoides aba-1* gene is composed of at least 10 repeats of approximately 400 bp, which is comparable with similar analysis performed in *A.suum* (see Chapter 5).

Figure 6.2. Southern analysis of partially-digested *A. lumbricoides* genomic DNA. Genomic DNA was partially digested with the following concentrations of *Xba*I for 1 hour: 5 units (A); 1 unit (B); 0.5 units (C). Samples were electrophoresed on a 0.8% agarose gel and capillary blotted onto Hybond N. The blot was probed with the pHS10 insert. λ *Hind*III/ *Eco*RI markers (in kbp) run on the same gel are shown at the left of the figure.

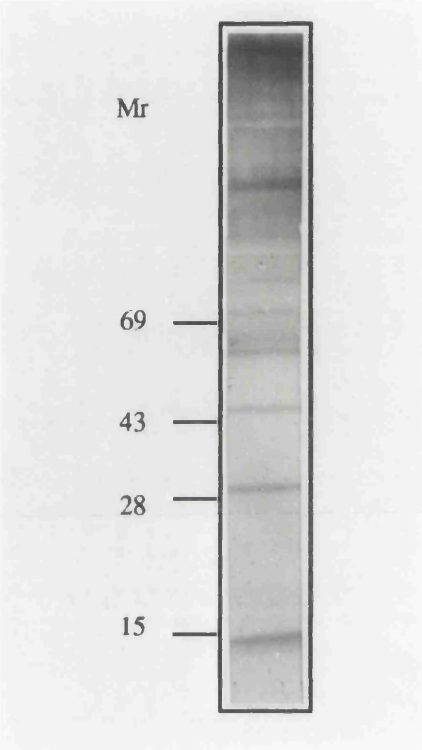

A
B
C
kbp
23
4.3
2.0
0.56

6.2.2 Western blot analysis of the native *A. lumbricoides* ABA-1 protein

Because the Southern analysis reported in Section 6.2.1 suggested that the *A. lumbricoides aba-1* gene is composed of repeating units of approximately 400 bp, it implied that, like the *A. suum aba-1* gene, it may encode a polyprotein which is then cleaved into 14 kDa monomers. To determine if this is the case, the structure of the ABA-1 protein in the *A. lumbricoides* adult body fluid was examined by Western blots.

Approximately 500 µg of *A. lumbricoides* adult body fluid (ABF) was used for the Western blot shown in Figure 6.3, which was probed with polyclonal rabbit antiserum raised to the *A. suum* recombinant pHS10-type ABA-1 protein (a gift from J. Moore). Note that a control (which is not shown) was also performed where the same blot was probed with normal rabbit serum, and no ABF proteins were detected. Protein species of approximately 14 kDa, 30 kDa, 45 kDa, 60 kDa, 75 kDa, 90 kDa, 105 kDa and 120 kDa were detected by the polyclonal antiserum to recombinant ABA-1, which were interpreted as representing polyproteins containing 1 to 8 monomers of *A. lumbricoides* ABA-1 (Figure 6.3). Like in *A. suum*, these proteins are thought to have resulted from incomplete cleavage of the ABA-1 polyprotein, and thus provide evidence for the *A. lumbricoides* ABA-1 protein being expressed as a polyprotein. The largest protein that was recognised (120 kDa) is approximately 8 times the size of the putative ABA-1 monomer, which suggests that there are at least 8 repeats of approximately 400 bp in the *aba-1* gene. It should be noted, however, that this experiment and the one reported in Section 6.6.1 detected a minimum estimate of the possible number of DNA or protein repeats, and therefore the *aba-1* gene and ABA-1 polyprotein initially expressed may be substantially larger than this estimate.

Figure 6.3. Western blot of the adult body fluid of *A. lumbricoides* probed with polyclonal rabbit antiserum raised to the *A. suum* recombinant ABA-1 protein. The protein sample was electrophoresed under reducing conditions on a 5-25% gradient SDS-polyacrylamide gel. The positions of molecular size markers (in kDa) electrophoresed on the same gel are indicated to the left.


Mr
69
43
28
15

6.2.3 Cloning of the *aba-1* gene of *A. lumbricoides* by PCR

To enable cloning of one repeating unit of the *A. lumbricoides aba-1* gene, a polymerase chain reaction was performed on genomic DNA isolated from embryonated eggs. The PCR reaction used oligonucleotide primers designed to hybridise to the extreme 5' and 3' regions of a 399 bp unit of the *A. suum* ABA-1 expressing cDNA clone pHS10 (see Materials and Methods and Figure 6.4). This approach was adopted because of the fact that the N-terminal sequence of the native *A. lumbricoides* ABA-1 protein shows 100% sequence identity to the *A. suum* ABA-1 native sequence, and also because the Southern analysis reported in Section 6.2.1 suggested that the two homologous genes contain very similar DNA sequence. It was therefore anticipated that the oligonucleotide primers, which contain *A. suum aba-1* sequence, should hybridise efficiently to the *A. lumbricoides* sequence. PCR using these primers resulted in amplification of DNA fragments 400 bp, 800 bp and 1200 bp in size, which were interpreted as representing 1, 2 and 3 copies of the *aba-1* repeats respectively (Figure 6.5A, lane D). A Southern blot of the PCR products probed with the pHS10 insert also revealed fragments of 400 bp, 800 bp and 1200 bp (Figure 6.5 B, lane A), suggesting that these DNA species are indeed derived from the A. *lumbricoides aba-1* gene. It should be noted that a control PCR was performed using the same primers but with no genomic DNA (Figure 6.5A, lane A). This reaction created no amplification products, suggesting that the DNA fragments produced above did not arise from any contaminant DNA from other organisms, for example *A. suum.*

To enable further characterisation of the 400 bp PCR product, it was subcloned in the plasmid vector pT7Blue, producing pHS100. Sanger Dideoxy sequencing was then performed using oligonuclotide primers

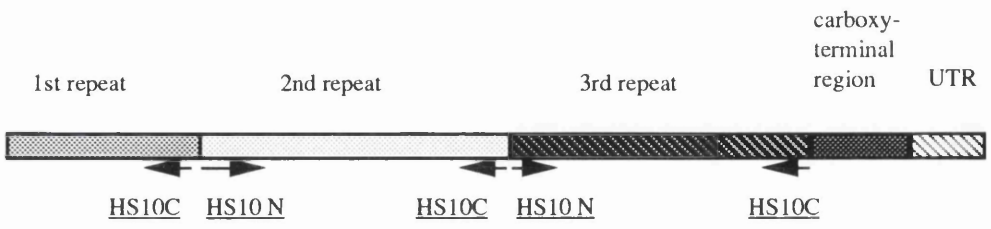


Figure 6.4. Diagrammatic illustration of where the oligonucleotide primers HS10N and HS10C are predicted to bind with the *aba-1* sequence of the *A. suum* cDNA clone pHS10. The sequences of the primers HS10N and HS10C (shown as arrows) are complementary to the 5' and 3' ends respectively of individual 399 bp *aba-1* repeats, and were used for PCR reactions on *A. lumbricoides* genomic DNA.

Figure 6.5 (A). PCR reactions performed on *A. lumbricoides* genomic DNA. PCR reactions were performed on *A. lumbricoides* genomic DNA and were electrophoresed on an 0.8% agarose gel. The primers used are described in Figures 6.4 and 6.8. The lane order is as follows:

Lane (A): control reaction with primers HS10N and HS10C where no DNA was added.

Lane (B) reaction with primers JMREPN and HS10C

Lane (C) reaction with JMREPN and JMREPC

Lane (D) reaction with primers HS10N and HS10C

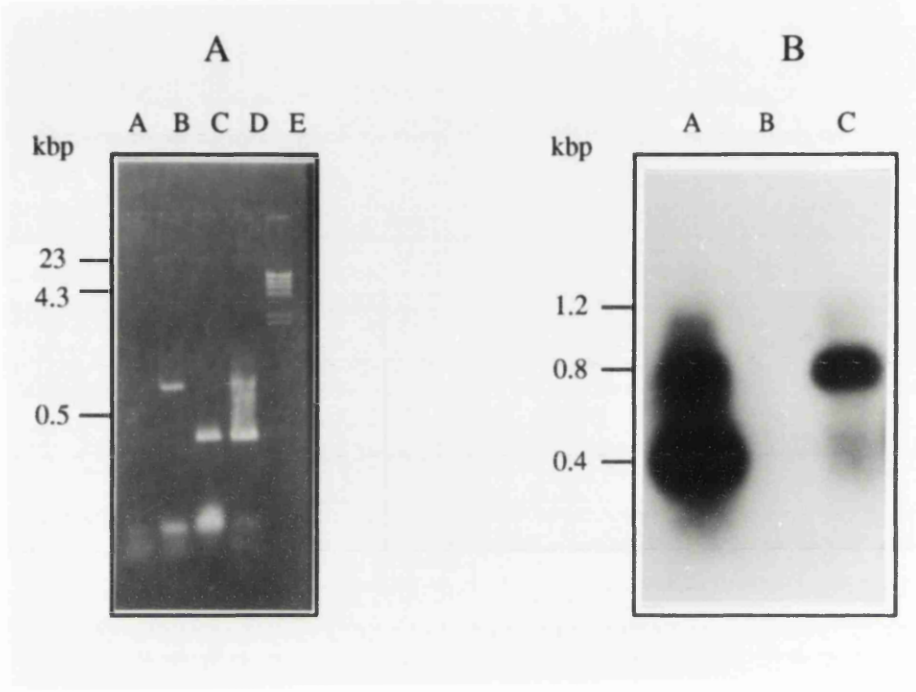
Lane (E) λ *Hind*III markers

Figure 6.5 (B). Southern blot analysis of the PCR reactions performed on *A. lumbricoides* DNA. PCR products were electrophoresed on a 0.8% agarose gel, capillary blotted onto HybondN and probed with the *aba-1* insert of clone pHS10.

Lane (A) PCR with primers HS10N and HS10C

Lane(B) PCR with primers JMREPN and JMREPC

Lane(C) PCR with primers JMREPN and HS10C.


A
A B C D E
kbp
23
4.3
0.5
B
A B C
kbp
1.2
0.8
0.4

```
CATCATTTCACCCTTGAAAGTAGTCTAGATACCCATCTGAAATGGCT
H  H  F  T  L  E  S  S  L  D  T  H  L  K  W  L

CAGTCAAGAACAGAAAGATGAATTGTTGAAAATGAAGAAAGATGGAA
 S  Q  E  Q  K  D  E  L  L  K  M  K  K  D  G  K

AGACAAAGAAAGATCTTCAAGCTAAAATTCTTTATTACTATGACGAA
   T  K  K  D  L  Q  A  K  I  L  Y  Y  Y  D  E

CTCGAAGGAGATGCTAAAAAGGAGGCAACTGAGCATTTGAAGGACGG
L  E  G  D  A  K  K  E  A  T  E  H  L  K  D  G

ATGCCGCGAAATTCTTAAGCATGTTGTTGGGGAAGAGAAGGAAGCAG
 C  R  E  I  L  K  H  V  V  G  E  E  K  E  A  E

AGCTGAAGAAACTCAAAGACTCGGGAGCAAGCAAAGAGGAAGTCAAA
   L  K  K  L  K  D  S  G  A  S  K  E  E  V  K

GCCAAAGTCGAAGAGGCACTTCATGCAGTAACCGACGAGGAGAAGAA
A  K  V  E  E  A  L  H  A  V  T  D  E  E  K  K

GCAATATATCGCCGATTTCGGACCAGCATGCAAGAAAATCTTTGCTG
 Q  Y  I  A  D  F  G  P  A  C  K  K  I  F  A  A

CAGCACATACTTCGCGACGAAGGAGG
   A  H  T  S  R  R  R  R
```

Figure 6.6. DNA sequence and putative amino acid sequence derived from the *A. lumbricoides* clone pHS100. pHS100 contains *A. lumbricoides* genomic DNA amplified by PCR using the the oligonucleotide pirmers HS10N and HS10C.

```
pHS100  HHFTLESSLDTHLKWLSQEQKDELLKMKKDGKTKKDLQAKILYYYDEL
pHS10   --------------------------------A--E-E----H-----
ABA-1   --------------------------------A--E-E---

pHS100  EGDAKKEATEHLKDGCREILKHVVGEEKEAELKKLKDSGASKEEVKAK
pHS10   -------------G--------------A----N----------L---


pHS100  VEEALHAVTDEEKKQYIADFGPACKKIFAAAHTSRRRR
pHS10   ----------------------------YGV -------
```

Figure 6.7. Comparison of the putative amino acid sequence derived from pHS100 with that of pHS10, and with the N-terminal amino acid sequence of the native *A. lumbricoides* ABA-1 protein. Where the amino acid sequence of *A. suum* ABA-1 (derived from the cDNA clone pHS10) and the *A. lumbricoides* native ABA-1 sequence correspond to that derived from pHS100 the amino acid residues are represented by (-); differences are shown by the residues' single letter code.

that allowed sequencing of both the coding and non-coding DNA strands of the insert (some of this sequencing was performed by Yu Xia). Figure 6.6 details the DNA sequence determined by this approach after the BESTFIT programme was used to exclude any potential sequencing errors (Devereux *et al.*, 1984). Comparing the DNA sequence of the pHS100 DNA insert (excluding PCR primers) to the pHS10 insert (containing part of the *A. suum aba-1* gene) revealed 92% identity. Analysis of the putative amino acid sequence encoded by the pHS100 insert revealed a putative 134 amino acid protein, which shows 92% identity to the native N-terminal ABA-1 *A. lumbricoides* sequence and 92% identity to the putative amino acid sequence encoded by the ABA-1-expressing clone pHS10 (see Figure 6.7). These data therefore confirm that a repeating unit of the *A. lumbricoides* gene has been successfully cloned. Further analysis of the sequence (Figure 6.7) revealed that the last four amino acids are arginines (positions 132-135), which is equivalent to the putative subtilisin endoproteinase cleavage site discussed in Chapter 3 and may therefore indicate that the *A. lumbricoides* gene is translated as a polyprotrein which is then cleaved into monomers.

6.2.4 Analysis of an *A. lumbricoides* divergent repeat similar to that found at the 5' region of the *A. suum aba-1* gene

To determine if the *A. lumbricoides aba-1* gene contains a divergent repeat within its *aba-1* gene, like the one present in the *A. suum aba-1* cDNA clone pJM33 (see J.Moore, 1993), PCR was performed on its genomic DNA using oligonucleotide primers designed to hybridise to the extreme 5' and 3' regions of the divergent repeat of the *A. suum aba-1* gene (Figure 6.8). A PCR product of approximately 400 bp was detected, which was

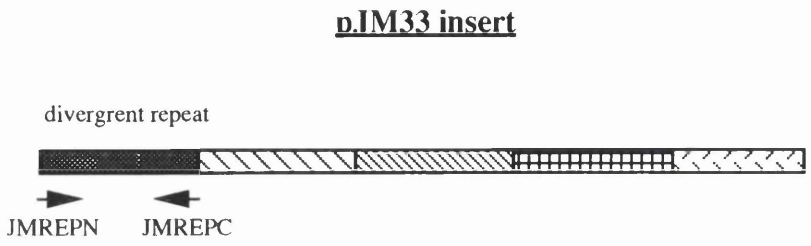


Figure 6.8. Diagrammatic illustration of where the oligonucleotide primers JMREPN and JMREPC are predicted to bind within the *aba-1* sequence of *A. suum* cDNA clone pJM33. The sequences of the primers JMREPN and JMREPC (shown as arrows) are complementary to the 5' and 3' ends of the divergent repeat present near the 5' end of the *A. suum aba-1* gene, and were used for PCR on *A. lumbricoides* genomic DNA

```
ACAATGGAACACTATCTCAAAACCTATCTGAGCTGGCTGACAGAAGA
T  M  E  H  Y  L  K  T  Y  L  S  W  L  T  E  E

ACAAAAAGAAAAGCTGAAAGAAATGAAAGAGGCAGGCAAAACGAAGG
 Q  K  E  K  L  K  E  M  K  E  A  G  K  T  K  A

CAGAGATCCAACATGAAGTGATGCACTACTACGATCAACTGCATGGT
   E  I  Q  H  E  V  M  H  Y  Y  D  Q  L  H  G

GAAGAAAAACAACAAGCAACAGAAAAGCTCAAAGTGGGCTGCAAAAT
E  E  K  Q  Q  A  T  E  K  L  K  V  G  C  K  M

GCTCCTGAAAGGAATCATCGGCGAGGAAAAGGTAGTTGAGCTGAGGA
 L  L  K  G  I  I  G  E  E  K  V  V  E  L  R  N

ACATGAAGGAAGCAGGAGCAGACATTCAAGAACTTCAACAAAAGGTT
   M  K  E  A  G  A  D  I  Q  E  L  Q  Q  K  V

GAGAAGATGCTTTCCGAGGTCACAGACGAAAAGCAGAAAGAAAAAGT
E  K  M  L  S  E  V  T  D  E  K  Q  K  E  K  V

CCACGAGTATGGACCCGCATGCAAAAAGATCTTCGGTGCGACAACAC
 H  E  Y  G  P  A  C  K  K  I  F  G  A  T  T  L

TGCAACATCATCGACGAAGGAGG
   Q  H  H  R  R  R  R
```

Figure 6.9. DNA sequence and putative amino sequence derived from the *A. lumbricoides* clone pHS200. pHS200 contains an *A. lumbricoides aba-1* divergent repeat which is 402 bp in length. This fragment was isolated using PCR on *A. lumbricoides* genomic DNA with the oligonucleotide primers JMREPN and JMREPC.

interpreted as representing a divergent repeat of the *A. lumbricoides aba-1* gene (see Figure 6.5, lane C). To exclude the possibility of this DNA being an amplification artefact, a control PCR was performed using the same primers, but without genomic DNA (data not shown). Since no DNA amplification was seen it may be concluded that this reaction product is genuine.

To further characterise this DNA fragment, it was subcloned into pT7Blue, producing the plasmid pHS200, and was sequenced by the Sanger dideoxy method using oligonucleotide primers that allowed sequencing of the coding and non-coding strands (some of this sequencing was performed by Yu Xia). The sequence from both coding and noncoding strands were compared by the BESTFIT programme in order to exclude any sequencing errors (see Figure 6.9 for the assembled DNA sequence). Comparative DNA sequence analysis of the insert of pHS200 with the *A. suum* divergent repeat revealed 100% identity. This result suggests that the *A. lumbricoides aba-1* gene contains a divergent repeat like that present at the 5' end of the *A. suum aba-1* gene. Whether the *A. lumbricoides* divergent repeat is also found at the 5' end of the gene, or is present nearer the 3' end, cannot be deduced from this experiment.

A further PCR reaction was carried out to characterise the genomic organisation of the *A. lumbricoides aba-1* gene; this reaction utilised the divergent N-terminal primer (JMREPN; Figure 6.8) and a primer which hybridises to the extreme 3' region of the *aba-1* repeats present in pHS10 (HS10C; Figure 6.4). This reaction resulted in the amplification of a DNA species of approximately 800 bp (Figure 6.5, lane B). These data suggest that the divergent repeat and pHS10-like repeats lie beside each other in the *A. lumbricoides aba-1* gene, the same organisation as has been described for the *A. suum aba-1* gene (Chapter 5).

6.3 Discussion

This chapter reports experimental approaches performed to analyse the *A. lumbricoides aba-1* gene and the ABA-1 protein which it encodes. The data presented suggest that the *aba-1* gene is composed of at least 10 repeating units of approximately 400 bp, and initially encodes a large polyprotein which is then cleaved into 14 kDa monomers. These data imply that it has a similar, if not identical, structure and pattern of expression as its *A. suum* homologue (see Chapters 3, 4 and 5).

Although the above similarities were found, analysis of the DNA sequence and the putative amino acid sequence encoded by one repeat of the *A. lumbricoides aba-1* gene suggested that there is only 92% identity with the *A. suum* pHS10-like repeats at both the DNA and amino acid sequence levels. These data may represent significant differences between the *A. lumbricoides aba-1* gene and its homologue in *A. suum*, and may therefore provide evidence for these being different *Ascaris* species. Despite these differences, there are problems with this interpretation, as this work only characterises one repeat of the *A. lumbricoides aba-1* gene, which may contain up to 20 repeats. It is possible that the *A. lumbricoides aba-1* gene does contain repeats which show 100% identity to the *A. suum* repeats present in pHS10, and it should be noted that even within the repeats present within pHS10 DNA sequence differences have been detected (Chapter 3). Another problem is that although the characterised *A. lumbricoides* repeat encodes a polypeptide that only shows 92% identity to the N-terminal amino acid sequence of the native ABA-1 protein, there may again be other repeats present within the *A. lumbricoides* gene that encode a polypeptide with 100% identity to the native ABA-1 protein. Since this analysis used worms that were isolated from different geographical locations (*A. suum* from Scotland, and *A. lumbricoides*

from Guatamala), these differences could also be accounted for by geographical isolation, as discussed in Chapter 1. In conclusion, it is not possible to determine from the above data on the *aba-1* genes of *A. lumbricoides* and *A. suum* whether they present evidence for the organisms belonging to the same or separate species.

One way of examining this question further would be to sequence the entire *aba-1* genes of both putative species. This is not, however, a very realistic suggestion as the total sizes of the *aba-1* genes have not been determined and may be as large as 8 kbp in length. Perhaps a better approach would be to to the analyse the *aba-1* genes of both species using Southern blots of their genomic DNA restricted with several more restriction enzymes than have been used here. It should be noted that species from the same geographical locations would have to be used to prevent the possibility of the differences resulting from geographical divergences.

Chapter 5 discussed the finding of a divergent *A.suum aba-1* repeat (J. Moore *et al.*,1993) and which demonstrated only 49 % identity to the the *aba-1* repeats present within the *A. suum* cDNA clone pHS10. This chapter reports the finding that *A. lumbricoides* also contains a divergent *aba-1* repeat. Sequencing of this repeat suggests that it displays 100% identity at both the DNA and amino acid level to the *A. suum aba-1* divergent repeat. This implies that this repeat was not an artefact produced during the synthesis of the cDNA library. However, it is still not clear whether the divergent repeat encodes an ABA-1-like protein which has an identical, or separate functions to the proteins encoded by the pHS10-like repeats. This question cannot be addressed until the function of the molecule has been determined, which will most likely be achieved through the use of reverse genetics with the *C. elegans* NPA homologue (see Chapter 3 for discussion).

Chapter 7

Concluding remarks

The aim of this work was to clone the gene which encodes the ABA-1 protein of *Ascaris*. This was performed primarily because ABA-1 is a major allergen, and because there is a restricted immune recognition to this molecule which has been shown to be under the control of the hosts MHC. Cloning was successfully achieved, and both cDNA and genomic copies of the *aba-1* gene of *A. suum*, and genomic copies of the *A. lumbricoides aba-1* gene, were characterised. These data revealed that the *aba-1* gene has an unusual structure in that it is composed of tandem repeats of approximately 400 bp, which suggested that ABA-1 is initially translated as a large polyprotein and is then cleaved into 14 kDa monomers. Detection of the native ABA-1 proteins as polyproteins by the use of Western blots confirmed this suggestion.

Although the structure of the *aba-1* gene has been characterised, the biological function of ABA-1 still remains unknown. It is now apparent that homologues of ABA-1 are present in a number of parasitic nematodes, and also in the free-living nematode *C. elegans* (Moore *et al.* 1993). This latter finding perhaps suggests that the function of ABA-1-type proteins is not parasite-specific. Work presented in this thesis has provided some insight into its function since we now know that it is constitutively expressed throughout the life of the parasite and that it appears to be expressed in all tissues of the parasite, although it seems only to be present as a polyprotein in the gut. These data suggest that ABA-1 may be needed in all *Ascaris* cells. However, it is not clear why ABA-1 polyprotein is not processed in the gut tissue of the parasite or why such large amounts of ABA-1 are found in the body fluid of the parasite. Secondary structure predictions from ABA-1 primary sequence suggest that the protein may have a 4 alpha helical bundle structure (which is the case for several carrier proteins ; Branden and Tooze, 1991). ABA-1 could

therefore be involved in carrying some by-product(s) of cellular metabolism to the gut where it could be released with the faeces out of the parasite, or alternatively it could be used to transport essential nutrients derived from the host for the parasite's use. This hypothesis would readily explain why ABA-1 is found in such large abundance in the body fluid, but does not explain why it is found only as a polyprotein in the gut.

Another question that arises from the prediction that ABA-1 is a carrier protein is, what does the protein carry? Kennedy *et al.* (personal communication) have used the parasite-derived and recombinant ABA-1 to perform biochemical analyses, which have suggested that it may carry fatty acids. This work is still in its infancy, but it may show that the above speculations as to the function of ABA-1 are correct.

Another means of determining the function of ABA-1 that is currently being pursued is utilising reverse genetics in *C. elegans*. A mutant could, for example, be created by insertional mutagenesis that does not express the *C. elegans* ABA-1 homologue, and any putative phenotype of such a mutant will provide a clue to the function of the protein.

Cloning of the gene encoding ABA-1 should now permit work to be performed to investigate the features of the protein that makes it an allergen. An example of such research is that now the full amino acid sequence of of ABA-1 is known, it will be possible to compare its amino acid sequence to other characterised allergens. This could potentially reveal some conserved amino acid motifs which may be responsible for the production of IgE responses in infected hosts. Another aspect of this analysis is that recombinant ABA-1 protein, which has been over-expressed in *E. coli,* has been shown to be a functional allergen by PCA studies (J. Moore *et al.*, 1994). N-terminal and the C-terminal portions of the protein have recently been expressed in isolation in order to determine which part of the ABA-1 sequence is responsible for this allergic

response. It is hoped that this research will eventually determine the precise epitope(s) of ABA-1 that induce allergic responses in the host. This could be beneficial because identification of such allergenic regions may allow them to be removed, or altered, in any recombinant ABA-1 protein which may be tested as a vaccine in order to avoid such allergic responses. Additionally, it may be possible to find such regions in other antigens that have been characterised.

Through the cloning of the genes encoding *A. lumbricoides* and *A. suum* ABA-1, and the ability to sub-clone these genes into defined vectors for expression in *E. coli* or other hosts, it is possible to purify ABA-1 in large quantities. These recombinant proteins can be used for further immunlogical experiments to determine if the MHC-restricted immune response to ABA-1 seen in mice is also seen in humans, and what features of the protein contribute to this restriction. These experiments are of importance because the more knowledge that can be derived about the restricted recognition of parasite antigens by human hosts will be of an advantage to those trying to develop vaccines against *Ascaris* and other parasites.

Bibliography

Abdulrachman, S. and Lie Kian Joe (1954). Morphological differences between *Ascaris* from man and pigs. *Documenta de Medicina Geographica et Tropica*. **6**: 342-344.

Ambler, J., Miller, J. N. and Orr, T. S. C. (1974).Some properties of *Ascaris summ* Allergen A. *Int. Archs Allergy Appl. Immunology*. **46**: 427.

Ambler, J., Croft, A. R., Doe, J. E., Gemmel, D. K., Miller, J. N. and Orr, T. S. C. (1973). Biological techniques for studing the allergenic components of nematodes. I Detection of allergenic components in *Ascaris suum* extracts. *Journal of Immunological Methods*. **1**: 317-328.

Amond, N. M. and Parkhouse, R. M. E. (1985). Nematode antigens. *Current Topics in Microbiology and Immunology*. **120**: 173-203.

Anderson, T. J. C., Romero-abal and Jaenike, J. (1993). Genetic structure and epidemiology of *Ascaris* populations: patterns of host affiliation in Guatemala. *Parasitology*. **107**: 319-334.

Ansel, M. and Thibaut, M. (1973). Value of the specific distinction betweeen Ascaris lumbricoides Linne, 1758 and Ascaris suum Goeze, 1782. *International Journal for Parasitology*. **3**: 317-319.

Baird, J. K., Mistrey, M., Pimsler, M. and Connor, D. H. (1986). Fatal human ascariasis following secondary massive infection. AM. J. Trop. Med. Hyg. **35(2)**: 314-318.

Balloul, J. M., Sondermeyer, P., Dreyer, D., Capron, M., Grzych, J. M., Poerce, R. J., Carvallo, D., Lecocq, J. P. and Capron, A(1987). Molecular cloning of a protective antigen of schistosomes. *Nature*. **326**: 149-153.

Barr, P. J. (1991). Mammalian subtilisins:the long-sought dibasic processing endoproteases. *Cell*. **66**: 1-3.

Bennett, K., Levine, T., Ellis, J. S., Peanasky, R. J., Samloff, I. M., Kay, J. and Chain B. M. (1992). Antigen processing for presentation by class II

major histocompatibility complex requires cleavage by cathepsin E. *European Journal of Immunology*. **22**: 1519-1524.

Birnboim H.C., and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**: 1513-1523.

Carrera, E., Nesheim, M. C. and Crompton, D. W. T.(1984). Lactose maldigestion in *Ascaris*-infected preschool children. *American Journal of Clinical Nutrition*. **39**: 255-264

Christie, J. F., Dunbar, B., Davidson, I. and Kennedy, M. W. (1990). N-terminal amino acid sequence identity between a major allergen of *Ascaris lumbricoides* and *Ascaris suum*, and MHC-restricted IgE responses to it. *Immunology*. **69**: 596-602.

Christie, J. F., Dunbar, B. and Kennedy, M. W. (1993). The ABA-1 allergen of the nematode *Ascaris suum*: epitope stability, mass spectrometery, and N-terminal sequence comparison with its homologue in *Toxocara canis*. *Clinical Experimental Immunology*. **92**: 125-132.

Cohen S. N. and Hsu, L. (1972). Non chromosomal antibiotic resistence in bacteria: genetic transformation of *Escherichia coli* by R factor DNA. Proc. Natl. Acad. Sci. USA. 69:2110-2114.

Coles, G. C. (1985). Allergy and immunopathology of ascariasis. In: *Ascariasis and its Public Health Significance* edited by D. W. T. Crompton, M. C. Nesheim and Z. S. Pawlowski (London and Philadelphia: Taylor and Francis), pp. 167-184.

Crewe, W. and Smith, D. H. (1971). Human infection with pig *Ascaris* (*A. suum*). *Annals of Tropical Medicine and Parasitology*. **65**: 85.

Crompton, D. W. T. and Pawlowski, Z. S. (1985). Life history and development of Ascaris Lumbricoides and the persistence of human ascariasis. In: *Ascariasis and its public health significance.* Edited by D. W. T. Crompton, M. C. Nesheim and Z. S. Pawlowski (London and Philadelphia: Taylor and Francis), pp9-23.

Crompton D. W. T. (1988). The prevalence of Ascariasis. *Parasitology Today*. **4 (6)**: 162-169.

Crompton, D. W. T. (1989). In: Biology of Ascaris lumbricoides in *Ascariasis and its Prevention and control*. Edited by D. W. T. Crompton, M. C. Nesheim and Z. S. Pawlowski (London and Philadelphia: Taylor and Francis), pp9-44.

Culpepper, J., Grieve, R. B., Friedman, L., Mika-Grieve, M., Frank, G. R. and Dale, B. (1992). Molecular characterisation of a *Dirifilaria immitis* cDNA encoding a highly immunoreactive antigen. *Molecular and Biochemical Parasitology*. **54**: 51-62.

Dandeu, J. P and Lux. M. (1978). Purification and characterisation of two proteins from *Ascaris suum* extract, antigenically different but bearing common allergenic epitopes. *Immunology Communications*. **7**: 393-415.

de Boer, E. (1935). Experimentelle onderzoek betreffende *Ascaris lumbricoides* van Mensch en varken. *Tijdscrift voor Diergeneeskunde* . **62**: 673-695.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acid Research*. **12**: 387-395.

Elkins, D. B., Haswell-Elkins, M. R. and Anderson, R. M. (1986). The epidemiology and control of intestinal helminths in the Pulicat Lake region of Southern India. I. Study design and pre- and post-treatment observations on *Ascaris lumbricoides* infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. **80**: 774-792.

Ferguson, M. A. J. and Williams, A. F. (1988). Cell-surface anchoring of proteins *via* glycosyl-phosphatidylinositol structures. *Annu. Rev. Biochem*. **57**: 285-320.

Fetterer, R. H. and Wasiuta, M. (1987). *Ascaris suum*: Partial isolation and characterisation of Hypodermis from the adult Female. *Experimental Parasitology*. **63**: 312-318.

Galvin, T. J. (1968). Development of human and pig *Ascaris* in the pig and rabbit. *Journal of Parasitology*. **54**: 1085-1091.

Glover, D. M. (1985). DNA cloning Volume 1 a practical approach. IRL press limited.

Greenspon, L. W., White, J., Shieds, R. L., Fugner, A. and Gold, W. M. (1986). Purification of *Ascaris suum* antigen: its allergenic activity *in vitro* and *in vivo*. *Journal of Allergy and Clinical Immunology*. **77**:443-451.

Hannon, G. J., Maroney, P. A., Denker, J. A. and Nilsen, T. W. (1990). *Trans*-splicing of nematode pre-messenger RNA in vitro. *Cell*. **61**: 1247-1255.

Hammond, M. P., and Bianco, A. E. (1992). Genes and genomes of parasitic nematodes. *Parasitology Today*. **8**: 299-305.

Haswell-Elkins, M. R., Kennedy, M. W., Maizels, R. M., Elkins, D. B. and Anderson, R. M. (1989). The antibody recognition profiles of humans naturally infected with *Ascaris lumbricoides*. *Parasite Immunology*. **11**: 615.

He, L., Min, X. T., Liu, G. Z (1986). Preliminary karyotype studies on *Ascaris lumbricoides* and *Ascaris suum* from Guangzhon. *Journal of Parasitology and Parasitic Diseases*. **4**: 206-208.

Heery, D. M., Gannon, F. and Powell, R. (1990). A simple method for sub-cloning DNA fragments from gel slices. *Trends in Genetics*. **6:** 173.

Hussain, R., Bradbury S. M. and Strejan G. (1973). Hypersensitivity to Ascaris antigens. VIII. Characterisation of a highly purified allergen. *Journal of Immunlogy*. **111**: 260.

Janssens, P. G. (1985). Chemotherapy of gastrointestinal nematodiasis in man. In:*Chemotherapy of Gastrointestinal Helminths. Handbook of Experimental Pharmacology, 77*. Edited by H. Vanden Bossche, D. Theinpont and P.G. Janssens (Berlin: Springer Verlag), pp. 183-406.

Jarrett, E. E. E. and Miller, H. R. P. (1982). Production and activities of IgE in helminth infection. *Prog. Allergy.* **31**: 178.

Kagei, N. (1883). Techniques for the measurement of environmental pollution by infective stage of soil-transmitted helminths. In:*Collected Papers on the Control of Soil-transmitted Helminthiases,* Vol 2, Edited by M. Yokogawa *et al.* (Tokyo: Asian Parasite Control Organization), pp. 27-46.

Klein, J, and Klein, D. (1987). Mouse inbred and congenic strains. *Methods in Enzymology.* **150**: 163-196.

Kennedy, M. W., Gordon, A., Tomlinson, L. A. and Qureshi, F. (1986). Genetic (major histocompatibility complex?) control of the antibody repertoire to the secreted antigens of Ascaris. *Parasite Immunology.* **9**: 269-273.

Kennedy, M. W. and Qureshi, F. (1986). Stage-specific secreted antigens of the parasite larval stages of the nematode *Ascaris. Parasite Immunology.* **58**: 512-522.

Kennedy, M. W., Qureshi, F., Haswell-Elkins, M. and Elkins, D. B. (1987). Homology and Heterology between the secreted antigens of the parasitic larval stages of *Ascaris lumbricoides* and *Ascaris suum. Clinical and Experimental Immunology.* **67**: 20-30.

Kennedy, M. W., Tierney, J., Ye, P., McMonagle, F. A., McIntoch, A.E., McLaughlin, D. and Smith, J. (1988). The secreted and somatic antigens of the third stage larva of *Anisakis simplex,* and antigenic relationship with *Ascaris suum, Ascaris lumbricoides,* and *Toxocara canis. Molecular and Biochemical Parasitology.* **31**:35-46.

Kennedy, M. W. (1989). Genetic control of the immune repertoire in nematode infections. *Parasitology Today.* **5**: 316-324.

Kennedy, M. W., Tomlinson, L. A., Fraser, E. M. and Christie, J. F. (1990). The specificity of the antibody responses to internal antigens of *Ascaris*:

heterogeneity in infected humans, and MHC (H-2) control of the repertoire in mice. *Clinical Experimental Immunology*. **80**: 219-224.

Kennedy, M. W., Fraser, E. M. and Christie, J. F. (1991). MHC class II (I-A) region control of the IgE antibody repertoire to the ABA-1 allergen of the nematode *Ascaris*. *Immunology*. **72**: 577-579.

Kennedy, M. W., McCruden, A.B., Brass, A., Cooper, A., Price, N.C., Kelly, S.M., Christie, J. F. and Dunbar, B., Spence, H.J. and Moore, J. (1993). Structural features of the ABA-1 allergen of *Ascaris*. *Journal of Cellular Biochemistry*. Abstract Supplement **17C**: 108.

Kingston, I. B., Wainwright, S. M. and Cooper, D. (1989). Comparison of collagen genes sequences in *Ascaris suum* and *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology*. **37**: 137-146.

Knox. D. P. and Kennedy, M. W. (1988). Proteinases released by the parasitic larval stages of *Ascaris suum* and their inhibition by antibody. *Molecular and Biochemical Parasitology*. **28**:207-216.

Knox, D. P. and Jones, D. G. (1991). Parasite enzymes in the diagnosis and control of ruminant nematodiasis. In: Parasitic nematodes-antigens, membranes and genes. Edited by Kennedy, M. W. (London and Philadelphia: Taylor and Francis), 170-194.

Krasnonos, L. I., (1978). Many-year viability of ascarid eggs (Ascaris lumbricoides) in soil of Samarkand. *Meditsinskaya Parazitologicya i Parasitaryne Bolezni*, **47**: 103-105. Cited in Helminthological Abstracts, (1979), 2623.

Kurimoto, H. (1974). Morphological, biochemical and immunological studies on the differences between *Ascaris lumbricoides* Linnaeus, 1758 and *Ascaris suum* Goeze, 1782. *Japanese Journal of Parasitology*. **23**: 251-267.

Kwan-Lim, G. Maizels, R. M. (1990) MHC AND non-MHC restricted recognition of filarial surface antigens in mice transplanted with adult Brugia malayi parasites. *Journal of immunology* . **145**: 1912-1920.

Lee, D. L. (1966). The structure and composition of the helminth cuticle. *Advances in Parasitology*. **4**: 1987-254.

Lee, D. L. (1972). The structure of the helminth cuticle. *Advances in Parasitology*. **10**: 347-379.

Lithtowlers, M. W. and Rickard, M. D. (1988). Excretory-secretory products of helminth parasites: effects on the hosts immune responses. *Parasitology*. **96**: S123-S166.

Lord, W. D. and Bullock, W. L. (1982). Swine *Ascaris* in humans. *New England Journal of Medicine*. **306**: 1113.

Lysek, H. (1963). Contribution to the morphological problem of differences between *Ascaris lumbricoides* Linne, 1758 and *Ascaris suum* Goeze, 1782. *Acta Societatis Zoologicae Bohemoslovenicae*. **27**: 97-101.

Mahalanabis, D., Jalan, K. N and Maitra, T. K. (1976). Vitamin A absorption in ascariasis. *American Journal of Clinical Nutrition*. **29**: 1372-1375

Maizels, R. M., Bundy, D.A.P., Selkirk, M. E., Smith, D. F. and Anderson, R. M. (1993). Immunological modulation and evasion by helminth parasites in human populations. *Nature*. **365**: 797-805.

Maung, M. (1973). *Ascaris lumbricoides* Linne, 1758 and *Ascaris suum* Goeze, 1782: morphological differences between specimens obtained from man and pigs. *Southeast Asia Journal of Tropical Medicine and Public Health*. **4**: 41-45.

McGibbon, A. M., Christie, J. F, Kennedy, M. W and Lee, T. D. G. (1990). Identification of the major *Ascaris* allergen and its purification to homogeneity by HPLC. *Molecular and Biochemical Parasitology*. **39:** 163.

McKinley-Grant, L. J., Idler, W. W., Bernstein, I. A., Parry, D. A. D., Cannizzaro, L., Croce, C. M., Huebner, K., Lessin, S. R. and Steinert, P. M. (1989). Characterization of a cDNA clone encoding human filaggrin and

localization of the gene to chromosome region 1q21. *Proc. Natl. Acad. Sci. USA.* **86**: 4848-4852.

Moore, J., McCurrach, K. J. and Kennedy, M. W. (1992). Molecular analysis of Ascaris secreted-excreted antigens. *Journal of Cellullar Biochemistry.* Abstract supplement **16A**: 149.

Moore, J., Kennedy M. W. and Blaxter, M. (1994) A homolog of a major class of allergens of parasitic nematodes is in the*C.aenorhabditis elegans* genome. *The Worm Breeder's Gazette.* October 1994.

Mount, S. M. (1982). A catalogue of splice junctions sequences. *Nucleic Acid Research.* **10**: 459-472.

Mutafova, T. (1983). Comparative caryological studies of *Ascaris lumbricoides* and *Ascaris suum. Helminthologia.* **15**: 48-56.

O'Donnell, I. J. and Mitchell, G. F. (1978). An investigation of the allergens of *Ascaris* using a radioallergosorbent test (RAST) and sera of naturally infected humans: comparison with an allergen for mice identified by passive cutaneous anaphylaxis tests. *Aust. Journal of Biological Science.* **31**: 459-487.

Ogilive, B. M. and de Savigny, D. (1982). Immune response to nematodes. In: *Immunology of Parasitic Infections.* Edited by S. Cohen and K. S. Warren (Oxford and London: Blackwell), pp715-757.

Miller, J. (1972). Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Mottram, J. C., North, M. J., Barry, J. D. and Coombs, G. H. (1989). A cysteine proteinase cDNA from *Trypanosoma brucei* predicts an enzyme with an unusual C-terminal extension. *FEBS letters.* **258**: 211-215.

Mount, S. M. (1982). A catalogue of splice junctions sequences. Nucleic Acid Research. 10: 459-72.

Pawlowski, Z. S. (1978). Ascariasis. In: *Clinics in Gastroenterology*, Vol. 8, *Intestinal Parasites*. Edited by P. D. Marsden (London, Philadelphia, Toronto: W. B. Saunders), pp. 167-178

Pawlowski, Z. S. (1982). Ascariasis: host-pathogen biology. *Reviews of Infectious Diseases.* **4**: 806-814.

Pawlowski, Z. S. and Arfaa, F. (1984). Ascariasis. In: *Tropical and Geographical Medicine.* Edited by K. S. Warren and A. A. F. Mahmoud (New York: McGraw-Hill), pp. 347-358.

Paxton, W. M. and Yazdanbakhsh, M. (1993). Primary structure and IgE response to the repeat subunit of gp15/400 from humans lymphatic filarial parasites. *Infect. Immun.* **61**: 2827-2833.

Poole, C. B., Grandea, A. G., Maina, C. V., Jenkins, R. E., Selkirk, M. E. and McReynolds, L. A. (1992). Cloning of a culticular antigen that contains multiple tandem repeats from the filarial parasite *Dirofilaria immitis. Proc. Acad. Science. USA.* **89**: 5986-5990.

Reynolds, L. A., Kennedy, M. W. and Selkirk, M.E. (1993). The Polyprotein Allergens of Nematodes. *Parasitology Today.* **9, No.11**: 403-406.

Rindisbacher, L., Hemphill, A. and Seebeck, T. (1993). A repetitive protein from *Trypanosoma brucei* which caps the microtubules at the posterior end of the cytoskeleton. *Molecular and Biochemical Parasitology.* **58**: 83-96.

Rogers, W. P. (1960) The physiology of infective processes of nematode parasites; the stimulus from the animal host. *Proc. Roy. Soc. London* B. **152**: 367-386.

Rothnagel, J. A., Merhrel, T., Idler, W., Roop, D. R. and Steinert, P. M. (1987). The gene for mouse epidermal filaggrin precursor. *The Journal of Biological Chemistry. 262*: 15643-15648.

Sambrook, J.,Fritsch, E. F. and Maniatis, T. (1989). In: Molecular Cloning: a laboratory manual. Cold Spring Harbor Press, New York.

Sanger, F., Nicklen, S. and Coulsen, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**: 5463-5467.

Sharp, P. A. (1987). Trans splicing: variation on a familiar theme. *Cell.* **50:** 147-148.

Sheets, M. D., Ogg, S. C. and Wickens, M. P. (1990). Point mutations in AAUAAA and the poly-(A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acid Research.* **18,** 5799-5805.

Shoemaker-Nawas, P., Frost, F., Kobayashi, J. and Jones, P. (1982). *Ascaris* infection in Washington. *Western Journal of Medicine.* **136**: 436-7.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology.* **98**: 503-517.

Sprent, J. F. A. (1952). Anatomical distinction between human and pig strains os *Ascaris. Nature, London* **170**: 627-638.

Stephenson, L. S. (1987). *Impact of Helminth Infections on Human Nutrition* (London and Philadelphia: Taylor and Francis).

Takata, I. (1951) Experimental infection of man with *Ascaris* of man and the pig. *Kitasato Archives of Experimental Medicine,* **23**: 49-59.

Tomlinson, L. A., Christie, J. F., Fraser, E. M., McLaughlin, D., McIntoch, A.E. and Kennedy, M. W. (1989). MHC restriction of the antibody repertoire to secretory antigens, and a major allergen, of the nematode parasite *Ascaris.The Journal of Immunology.***143:** 2349-2356.

Tyson, E. (1683). Lumbricus teres, or some anatomical observations on the roundworm bred in human bodies. *Philosophical Transactions of the Royal Society of London* **13:** 152-161.

Tweedie, S. A. R., Paxton, W. A., Ingram, L., Maizels, R. M., McReynolds, L. A. and Selkirk, M. E. (1993). A surface-associated glycoprotein (gp15/400) from *Brugia* filarrial parasites is composed of multiple tandemly-repeated units and processed from a 400 kDa precursor. *Experimental Parasitology*. **76**: 156-164.

Wakelin, D. (1985). Genetic control of immunity to helminth infections. *Parasitology Today*. **1 (1)**: 17-23.

Watson, B. D. (1965). The fine structure of the body wall and growth of the cuticle in the adult nematode *Ascaris lumbricoides*. *Quaterly Journal of Microscopical Science*. **106**: 83-91.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987). *Molecular Biology of the Gene*. The Benjamin/Cummings Publishing Company, Inc.

Wheelock, M. J., Komuniecki, R., Duran, E. and Johnson, K. R. (1991). Characterisation of cDNA clones for the beta subunit of pyruvate dehydrogenase from *Ascaris suum*. *Molecular and Biochemical Parasitology*. **45**: 9-18.

World Health Organisation. (1967). *Control of Ascariasis*. Technical Report Series, No. 379.

World Health Organisation. (1981). *Intestinal, Protozoan and Helminthic Infections*. Technical Report Series, No. 666.

Zabarovsky, E. R. and Allikmets, R. L. (1986). An improved technique for the efficient construction of gene libraries by partial filling-in of cohesive ends. *Gene*. **42**: 119-123.