ANTIGEN ENCODING GENE FRAGMENTS OF *CRYPTOSPORIDIUM PARVUM*

Nicola C. Lally

A thesis submitted for the degree of PhD

University of Edinburgh
September 1992

ABSTRACT.

Cryptosporidium parvum is an obligate intracellular protozoan which infects the gastrointestinal tract of a wide range of mammalian species. It is a common cause of diarrhoeal illness in humans and neonatal ruminants. Despite the medical and veterinary importance of *C. parvum* studies of this organism at the genetic level have begun only recently. This is due to the lack of interest shown in the parasite until it was recognised as a cause of human and animal disease, and also to the difficulty in producing sufficient parasite material in order to carry out such studies.

The aim of this study was to identify, by screening a DNA library with anti-*C. parvum* antisera, genes or gene fragments encoding antigens of *C. parvum*. A *C. parvum* λgt11 expression library was constructed using *Eco*RI-digested genomic DNA prepared from *in vitro*-excysted oocysts. Screening the library resulted in the isolation of two immunopositive clones. λCPR1, recognised by rat serum raised against excysted *C. parvum* oocysts, and λCPS10, recognised by serum from a gnotobiotic lamb experimentally infected with *C. parvum*. The DNA inserts from these clones (CPR1 and CPS10 respectively) were subcloned into the pMS plasmid expression vectors, and the recombinant peptides expressed by the resulting subclones analysed by Western blotting.

Subclones containing CPS10 expressed a peptide which was recognised by some, but not all, lambs infected with *C. parvum*. When CPR1 was subcloned into pMS1S, the resulting subclone expressed a 200kDa β-galactosidase fusion protein. This fusion protein was partially purifIed and used to raise polyclonal antiserum in a rabbit. Western blotting indicated that this serum recognised a 190kDa peptide constituent of the *C. parvum* oocyst wall.

The CPR1 DNA fragment was sequenced in both directions and found to consist of 2359 nucleotides, 2358 of which form a continuous open reading frame. The DNA sequence has a relatively low G+C content (39.1%) and there is a corresponding bias towards the use of codons ending in A or T (82.1%) within this open reading frame.

The deduced peptide sequence has an unusual amino acid composition, with high proportions of cysteine, proline, glutamine and histidine. The cysteine residues are found in three distinct cysteine-rich regions, which contain repeat units based on conserved cysteine residues. The corresponding cysteine residues in the native *C. parvum* 190kDa peptide are likely to participate in disulphide cross-linking which may be of structural importance in the oocyst wall.

ACKNOWLEDGEMENTS.

My appreciation and thanks are due to the following:

Dr G. D. Baird for supervision throughout this project, and for advice on writing this thesis.

Mr S. E. Wright for teaching me all the parasitological techniques used in this work, for carrying out all the Home Office licensed animal prodedures described in this thesis, and for his advice and help throughout my four years spent at Moredun.

The late Dr D. A. Blewett, and Prof. J. G. Scaife for advice and encouragement, especially during the early stages of the work.

Dr R. G. Ridley for practical advice on various techniques.

Mr B. J. Easter for photography.

Dr F. Wright for advice on the use of sequence analysis software.

S. J. McQuay for advice on the use of sequence analysis software, and for helpful discussions on protein sequence analysis.

My parents for support throughout my academic career.

Most of all, thanks to Stuart.

ABBREVIATIONS.

AIDS	Acquired immune deficiency syndrome
ATP	adenosine triphosphate
BCIP	5-bromo-4-chloro-3-indolyl-phosphate (ρ-toluidine salt)
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
CIP	calf intestinal phosphatase
ddNTP	dideoxynucleoside triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
DMSO	dimethyl sulphoxide
EDTA	ethylenediaminetetra-acetic acid
GCG	Genetics Computer Group
HBSS	Hanks' balanced salt solution
hr	hour(s)
Ig	immunoglobulin
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase pairs
kDa	kilodaltons
L-broth	Lauria broth
Mab(s)	monoclonal antibody(ies)
min	minutes(s)
MRI	Moredun Research Institute
mRNA	messenger RNA
mw(s)	molecular weight(s)
NBRF	National Biomedical Research Foundation
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PAM	accepted point mutation
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming units
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
sec	second(s)
TBS	tris buffered saline
TEMED	N,N,N',N',tetramethylethylenediamine
tris	tris(hydroxymethyl)methylamine
UV	ultraviolet
v/v	volume/volume
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-thiogalactopyranoside

CONTENTS.

TABLES.

CHAPTER 1. INTRODUCTION.

1.1 *Cryptosporidium parvum*.

Cryptosporidium parvum is a protozoan parasite which infects the gastrointestinal epithelium of a variety of mammalian species. It was first described infecting the small intestine of laboratory mice (Tyzzer 1912).

Little interest was shown in the parasite in the 50 or so years following its discovery, as it was not considered to be of medical or veterinary importance. The first human cases of cryptosporidiosis were reported in 1976 (Meisel *et al* 1976, Nime *et al* 1976), and more recently *C. parvum* has been recognised as a cause of severe life-threatening diarrhoeal illness in immunocompromised patients, especially those with the Acquired Immune Deficiency Syndrome (AIDS). In addition, it is now recognised as a frequent cause of diarrhoea in immunocompetent people.

Veterinary interest in *Cryptosporidium* was initiated in 1976 when it was first reported in association with bovine diarrhoea (Panciera *et al* 1976). Since then, *C. parvum* has been reported infecting a wide range of mammals, and is recognised as an important cause of diarrhoea in neonatal ruminants.

1.2 Taxonomy of *Cryptosporidium*.

The taxonomic classification of the genus *Cryptosporidium* is shown in table 1.1.

The classification of species within the genus *Cryptosporidium* has been the subject of some debate and the question is still not resolved. Tyzzer (1907) first described *Cryptosporidium* in the gastric glands of mice and suggested the name *Cryptosporidium muris*, but did not describe characteristics to establish a new genus until 1910. In 1912 he described the morphology and lifecycle of a second species, *Cryptosporidium parvum*, which he observed in the small intestine of laboratory mice. The latter species differed from *C. muris* in that it localized in the small intestine of the murine host, whereas *C. muris* reproducibly colonized the stomach; it also produced smaller oocysts (Tyzzer 1912).

Slavin (1955) described mortality and morbidity associated with *Cryptosporidium* infection in turkeys. He described a parasite indistinguishable from *C. parvum* but named the species infecting turkeys *Cryptosporidium meleagridis*. Between 1968 and 1981 *Cryptosporidium* was reported infecting various species of fish, birds, reptiles and mammals (Fayer & Ungar 1986), and each isolate was named on the assumption that each host species was infected with a separate species of *Cryptosporidium*. This was subsequently found not to be the case. Tzipori *et al* (1980) successfully infected lambs, pigs, rats, mice and guinea-pigs

Table 1.1 Taxonomic classification of *Cryptosporidium.*
From Fayer & Ungar (1986).

Taxon	Name	Biological characteristics
Phylum	Apicomplexa	Obligate intracellular parasites;invasive forms with apical complex
Class	Sporozoasida	Locomotion of invasive forms by body flexion, gliding or undulation
Subclass	Coccidiasina	Lifecycle with merogony, gametogony, and sporogony
Order	Eucoccidiorida	Merogony present; parasitizes vertebrate hosts
Suborder	Eimeriorina	Male and female gametes develop independently
Family	Cryptosporidiidae	Homoxenous; developmental stages just under the host cell membrane; oocyst without sporocysts and with four sporozoites; microgametes without flagella
Genus	*Cryptosporidium*	With the characteristics of the family (single genus family)

with a *Cryptosporidium* isolate from calves. They also infected mice with isolates from lambs and humans, and on the basis of these findings, proposed that *Cryptosporidium* be considered a single species genus. Levine (1984) reviewed the taxonomy of *Cryptosporidium*. He considered four species to be valid, *Cryptosporidium crotali*, infecting reptiles, *Cryptosporidium nasorum* infecting fish, *C. meleagridis* infecting birds and *C. muris* infecting mammals. He considered *C. parvum* to be a synonym of *C. muris*. This classification is now also considered invalid. The original description of *C. crotali* based on the findings of cysts in the faeces of a snake, was probably a description of sporocysts of *Sarcocystis* sp. (Current *et al* 1986). In addition, Upton and Current (1985) showed that there were two distinct species of *Cryptosporidium* infecting calves, i.e *C. parvum* and *C. muris*, thus confirming Tyzzer's original findings.

Evidence also suggests that there are at least two species of *Cryptosporidium* infecting birds. Although *C. meleagridis* is morphologically indistinguishable from *C. parvum*, cross transmission experiments have shown that *Cryptosporidium* isolated from birds can be transmitted to other bird species, but not to mammals (Fayer & Ungar 1986). This suggests that *C. parvum* and *C. meleagridis* are distinct species. Another species infecting birds was described by Current *et al* (1986). This parasite, named *Cryptosporidium baileyi*, was isolated from the bursa of Fabricius of naturally infected chickens. The authors considered *C. baileyi* to be a separate species because it

differed from *C. meleagridis* in oocyst morphology and site of infection. Oocysts of *C. baileyi* are larger than those of *C. meleagridis*, and experimental infections with *C. baileyi* in turkeys resulted in only mild infections confined to the bursa of Fabricius, whereas Slavin (1955) reported heavy infections in the lower third of the small intestine of turkeys infected with *C. meleagridis*.

Clearly there is a need for further information on the species of *Cryptosporidium* infecting the various vertebrate classes. *C. parvum* is the species which causes human cryptosporidiosis, and is the most common species found infecting other mammals. Consequently it is the most studied species of *Cryptosporidium* and is the subject of the present study. Further discussion of *Cryptosporidium* and cryptosporidiosis in this thesis will relate primarily to *C. parvum*.

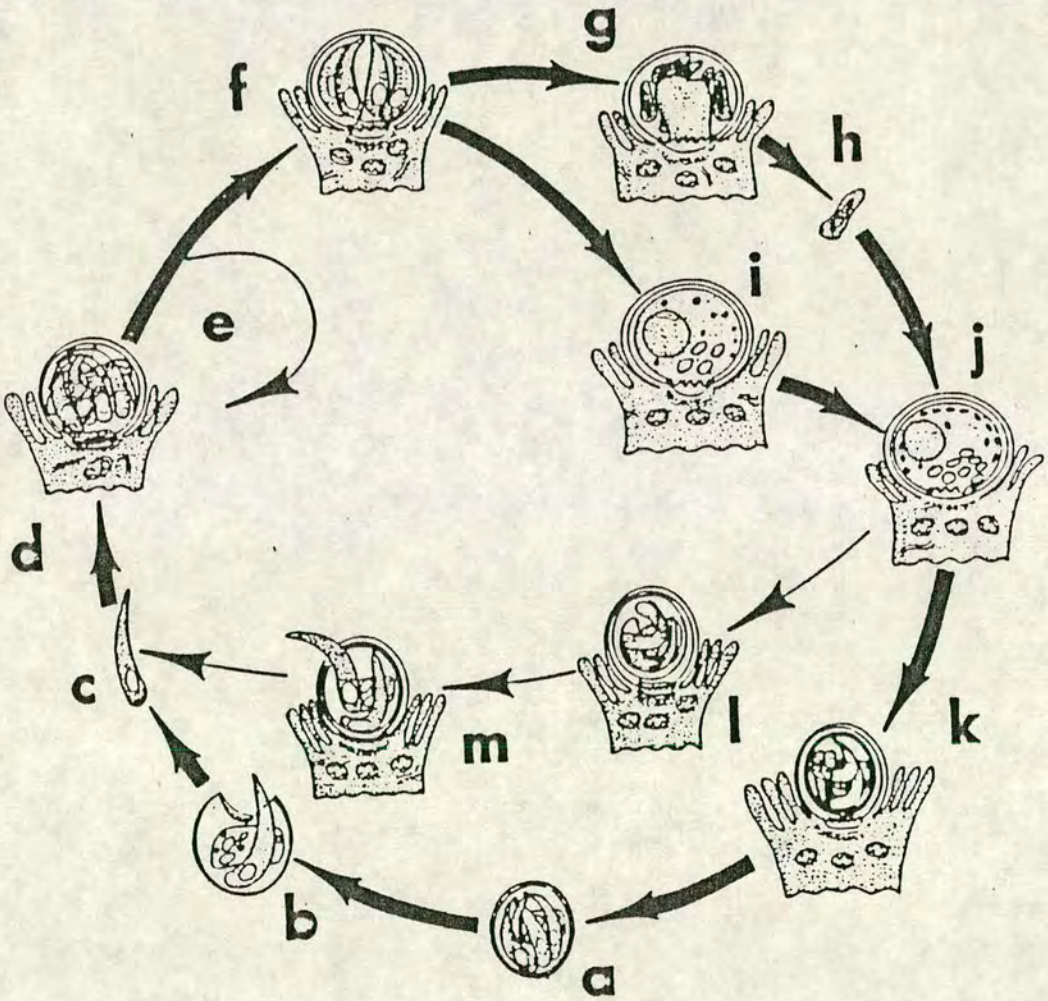
1.3 Life cycle and ultrastructure of *Cryptosporidium parvum*.

The life cycle of *Cryptosporidium parvum* is shown diagrammatically in Fig. 1.1

Much of the life cycle of *Cryptosporidium* was first described by Tyzzer (1910, 1912). Since then many of his observations have been confirmed, and further details of the life cycle have been described with the aid of ultrastructural studies. Oocysts of *Cryptosporidium parvum* are spherical to ovoid, 4-5μm in

Fig 1.1 Life cycle of *Cryptosporidium parvum.*

a) Sporulated, thick walled oocyst in faeces. b) Excystation in the intestine. c) Free sporozoite in the small intestine prior to penetration of an enterocyte. d) Mature type I meront with 6 or 8 merozoites. e) Recycling of type I merozoites to produce additional type I meronts. f) Type II meront with 4 merozoites that develop into the sexual stages. g) Microgamont, with approximately 16 microgametes. h) Free microgamete that fertilizes macrogamete (i) to form a zygote (j). Most of the zygotes (about 80%) form thick-walled oocysts that sporulate within the parasitophorous vacuole (k) before being passed out in the faeces. The remainder form thin-walled oocysts that excyst within the same host resulting in autoinfection.



diameter, and passed in the faeces fully sporulated and infective. When ingested the oocysts are stimulated to excyst by the action of trypsin and bile salts (Current & Haynes 1984). Just prior to excystation the sporozoites show rapid tumbling movements within the oocyst. Excystation results in the release of four sporozoites which move over the surface of the gut epithelium by flexing and gliding movements. As a sporozoite approaches a host cell, a thin host derived cytoplasmic extension, bounded by host membrane appears around the anterior of the parasite. This eventually extends to surround the sporozoite, forming a parasitophorous vacuole (Lumb *et al* 1988a). Vacuolation at the anterior end of sporozoites and merozoites during invasion has been observed (Lumb *et al* 1988a, Vetterling *et al* 1971), suggesting emptying of rhoptries and/or micronemes, as occurs during invasion of host cells by other apicomplexa (Perkins 1992).

All intracellular stages of *C. parvum* occupy a position in the host cell confined to the microvillous border, this has been described as intracellular but extracytoplasmic (Goebel & Braendler 1982, Pearson & Logan 1983). After invasion, the elongate sporozoite transforms into a spherical trophozoite. At the region of attachment the inner membranes of host origin, and the plasma membrane of the parasite, fuse and undergo extensive folding to form the so-called feeder organelle (Marcial & Madara 1986). During formation of the feeder organelle the trophozoite undergoes nuclear division and becomes a meront. Two further divisions, accompanied by cytoplasmic budding and infolding of

the parasite membrane, result in the formation of eight merozoites which are attached to a small residual body (Bird & Smith 1980, Vetterling *et al* 1971). Rupture of the mature meront releases the eight merozoites which are very similar to sporozoites. Current and Reese (1986) distinguished sporozoites from merozoites by the former containing amylopectin granules near the centre of the cell and the latter having a centrally placed nucleus, though Bird and Smith (1980) considered the merozoite nucleus to be at the posterior of the cell. Some workers (Snodgrass *et al* 1984, Bird & Smith 1980) thought there was only one type of meront, which when mature contained eight merozoites. However, it is now widely accepted that there are two types; i.e. type I meronts, which appear first in the lifecycle and contain eight merozoites when mature, and type II meronts, which contain only four. The former produce type I merozoites, which may develop into either type I or type II meronts, whereas type II meronts produce type II merozoites which go on to develop into gamonts. Evidence for this includes the continued presence of type I meronts over an eight day post infection period in enterocytes of mice, and the observation that type I meronts were always more numerous than type II meronts (Current & Long 1983).

Macrogamonts have a large nucleus, cytoplasm containing clearly defined rough endoplasmic reticulum, amylopectin granules near the parasite's base, wall-forming bodies, and a large lipid containing vacuole (Bird & Smith 1980, Goebel & Braendler 1982). As the macrogamont matures, the wall-forming bodies become more

numerous in the peripheral cytoplasm, and two types, which subsequently form the two layers of the oocyst wall, can be distinguished by electron density (Current & Reese 1986).

Microgamonts look very similar to macrogamonts until nuclear division begins. Microgametogenesis produces 16 microgametes by budding from a large residual body. Mature microgametes are wedge shaped, with a tapered posterior and expanded anterior end. The nucleus is situated posteriorly, and cytoplasmic microtubules run parallel to the long axis (Goebel & Brandler 1982). Microgametes released into the gut lumen have been reported to attach, by their anterior end, only to host cells containing a mature macrogamete, suggesting the involvement of specific receptors (Current & Reese 1986). Fertilization occurs by microgamete penetration of the macrogamete.

Following fertilization, oocyst formation begins. The oocyst wall consists of two layers limited by three membranes. It is believed to be formed by fusion of the wall-forming bodies with these membranes, and emptying of their contents between the membranes (Current & Reese 1986). Sporulation then occurs, forming four sporozoites. The method by which oocysts are released from the parasitophorous vacuole is not known, however, it may simply occur by degradation of the host cell after it has been shed into the lumen. Two types of oocyst have been described which differ in the thickness of their walls (Current & Reese 1986); thin-walled oocysts which readily excyst within the same host causing autoinfection, and thick-walled oocysts

which pass out in the faeces, thus completing the lifecycle.

The pre-patent period in experimental infections in calves, lambs and mice, and in cell culture is 3-4 days (Blewett 1988b, Angus et al 1982, Current & Long 1983, Current & Haynes 1984, Current & Reese 1986). This short pre-patent period, plus the potential for auto-infection by recycling of type I meronts and thin-walled oocysts contributes to the very high reproductive potential of the parasite. In one study lambs which were fed approximately 5-10 oocysts per day, each shed to the order of 10^{10} oocysts during a 20 day period (S. E. Wright, Moredun Research Institute, personal communication).

1.4 Clinical features and pathology of cryptosporidiosis.

The most common clinical feature of cryptosporidiosis in both humans and animals is diarrhoea. Other symptoms reported in both immunocompetent and immunocompromised persons include abdominal pain, weight loss, nausea, vomiting, mild fever and headache (Crawford & Vermund 1988). In most immunocompetent patients, diarrhoeal illness due to cryptosporidial infection lasts from three to 12 days. The illness usually resolves spontaneously, without intervention, though occasionally fluid replacement therapy is required, and in a few cases diarrhoea lasts more than two weeks (Current & Garcia 1991). Malnutrition may contribute to increased severity of cryptosporidial infection (Smith & van den Ende 1986, Bogaerts *et al* 1984), and

cryptosporidiosis is thought to be an important contributor to morbidity and mortality due to diarrhoeal illness in developing countries (Current & Garcia 1991).

Cryptosporidiosis in immunocompromised patients is typically more severe and prolonged than in the immunocompetent, and in cases where the cause of immunosuppression cannot be removed, may be a major factor leading to the death of the patient. In severly immunocompromised patients, for example those with AIDS, fluid loss is often excessive, three to six litres per day being not uncommon (Current & Garcia 1991). In addition, infection is not limited to the ileum, as is the case with most infections in immunocompetent persons, but may spread thoughout the gastrointestinal tract. In addition, infection of the gall bladder, pancreatic duct and the respiratory tract has been reported in a small proportion of immunocompromised patients with cryptosporidiosis (Fayer & Ungar 1986). In patients with immune deficiencies the length and severity of cryptosporidiosis is dependent on the ability to reverse immunosuppression. Patients with reversible immunosuppression include those on immunosuppressive therapy for cancer or transplants, malnourished individuals, and those with concurrent viral infections such as measles, and cytomegalovirus (DeMol *et al* 1984, Weinstein *et al* 1981). Despite the wide host range of *C. parvum*, clinical cryptosporidiosis in animals is primarily a problem in neonatal ruminants. It has been reported causing diarrhoeal illness in lambs (Tzipori *et al* 1981, Snodgrass *et al* 1984), bovine calves (Tzipori *et al* 1983, Heine *et al* 1984b),

red deer calves (Blewett 1988) and goat kids (Tzipori *et al* 1982a). The clinical features of cryptosporidiosis in these animals include diarrhoea, inappetence, depression, weight loss and dehydration. Morbidity and mortality are very variable, and may be dependent on other factors such as cold weather and concurrent infections (Angus 1989).

The mechanisms by which *C. parvum* infection causes diarrhoea are incompletely understood. In both immunocompetent and immunocompromised humans with cryptosporidiosis, blunting and fusion of villi, and infiltration of the lamina propria with inflammatory cells have been reported (Crawford & Vermund 1988). The resulting decrease in surface area may contribute to malabsorption resulting in diarrhoea. The occurrence of a secretory "cholera-like" diarrhoea in immune deficient patients with cryptosporidiosis suggests a toxin mediated hypersecretion into the gut. However, such a toxin has not been identified (Current & Garcia 1991).

Intestinal lesions associated with cryptosporidiosis have been described in lambs (Angus *et al* 1982, Snodgrass *et al* 1984, Tzipori *et al* 1981), calves (Heine *et al* 1984b, Tzipori et al 1983) and piglets (Argenzio *et al* 1990, Tzipori *et al* 1982b). In all these reports the ileum was most heavily infected, and showed the greatest extent of histological changes, while the colon and jejunum were sometimes affected. Microscopical examination revealed stunting and fusion of villi and infiltration of the lamina propria with inflammatory cells.

Decreased lactase activity was reported in calves (Tzipori *et al* 1983), lambs (Tzipori *et al* 1981) and piglets (Tzipori *et al* 1982), and reduced glucose-dependent absorption of sodium and water was reported in piglets (Argenzio *et al* 1990). The reduced absorptive surface area resulting from villous changes, and the decreased activity of membrane bound enzymes both probably contribute to malabsorption and diarrhoea associated with the infection.

1.5 Epidemiology of cryptosporidiosis.

Various reviews have summarized the results of epidemiological surveys of cryptosporidiosis in humans (Fayer & Ungar 1986, Crawford & Vermund 1988, Tzipori 1988, Current & Garcia 1991). These authors are in agreement that human cryptosporidiosis has a worldwide distribution, is more common in children than in adults (highest prevalence is reported in children under two years old), and that it is more prevalent in developing countries than in industrialized countries. The most recent of these reviews (Current & Garcia 1991) found that in industrialized countries e.g. North America and Europe, prevalence varied from 1-3%, and in underdeveloped countries values varied from 5% in parts of Asia, to 10% in some African countries. Serological studies also indicate that cryptosporidiosis is more common in underdeveloped countries. Seroprevalence rates reported in Europe and North America range from 25-35% whereas studies of seroprevalence in several South

American countries found that 64% of the population was seropositive (Current & Garcia 1991).

C. parvum has been reported in variety of mammalian species other than humans. However, the literature indicates that clinical cryptosporidiosis is most commonly a problem in neonatal ruminants, especially lambs and bovine calves. Consequently, more information is available on the infection in these species, than in other mammalian species. Reports of herd outbreaks and surveys indicate that bovine cryptosporidiosis has a worldwide distribution (Fayer & Ungar 1986). In surveys of calf diarrhoea in the UK and in Holland, Cryptosporidium was second only to rotavirus as the most commonly detected enteropathogen (Reynolds *et al* 1986, De Visser *et al* 1987). Cryptosporidiosis has been reported in lambs in Australia, Scotland, Germany and North America (Fayer & Ungar 1986). It has been suggested that cryptosporidiosis in lambs is less common and more sporadic than in calves (Angus 1989). In both lambs and calves, cryptosporidiosis appears to be age-related, with animals less than three weeks old most commonly infected.

Transmission of cryptosporidiosis is via the oocyst which is passed fully sporulated and infective in the faeces of the host. Sources of human infection include zoonotic transmission. Calves have been implicated as a source of human infection (Reese *et al* 1982, Anderson 1982), and in North Wales a study has noted an association between human cryptosporidiosis and contact with sheep (Casemore 1990). The occurrence of urban infections and

transmission within families and at daycare centres provides evidence of person to person transmission (Casemore 1990). Outbreaks of human cryptosporidiosis attributed to contaminated water supplies have been reported in Sheffield, Oxfordshire, Ayrshire, New Mexico and Georgia (Current & Garcia 1991). Possible sources of water contamination include the use of human faeces as fertilizer, contamination of water courses with raw sewage, and excretion of large numbers of oocysts by animals on pasture. Following rainfall, oocysts may enter water courses and find their way into drinking water supplies. The higher rates of infection in underdeveloped countries compared with industrialized countries is probably due to poor sanitation, resulting in the contamination of drinking water with human and animal faeces.

Bovine and ovine infections are primarily transmitted from one calf/lamb to another (Henriksen 1989, Angus 1990). Overcrowding and poor hygiene increase the risk of clinical infection in bovines (Henriksen 1989). Since adult ruminants seem not to be susceptible to *C. parvum* infection, the question arises as to how cryptosporidiosis is transmitted from one year to the next. Evidence suggests that *C. parvum* oocysts are killed by freezing (Sherwood *et al* 1982), so environmental contamination with oocysts from the previous year seems unlikely to be a source of infection for neonatal ruminants. Since *C. parvum* is not host specific, potential sources of infection include infected cats, dogs and wild rodents. In one study, 30% of 115 wild mice trapped at calving and calf rearing sites were infected with

Cryptosporidium (Klesius *et al* 1986). Another possible source of infection for neonatal ruminants is the excretion of small numbers of oocysts by adult ruminants with low grade asymptomatic infections. This has been reported for both cattle (Mann *et al* 1987) and sheep (Papadopoulou *et al* 1989), though it is not known how common such infections are.

1.6 Treatment and control of cryptosporidiosis.

Many drugs have been tested against crytosporidiosis, but with little success. Fayèr *et al* (1990b) list 94 therapeutic and preventive agents which have been tested for efficacy against cryptosporidiosis in humans or animals. Of these, few have shown any efficacy. Anecdotal success has been claimed for diloxanide furoate, furazolidone, quinine plus clindamycin, amprolium and interleukin-2. However, the value of such anecdotal reports is doubtful, especially where success is claimed in treating cryptosporidiosis in AIDS patients, in whom symptoms and oocyst shedding are known to wax and wane periodically, and in some cases to resolve spontaneously (Fayer & Ungar 1986).

Conflicting results have been reported for the macrolide antibiotic spiramycin. Two studies, which attempted treatment of cryptosporidial diarrhoea in immunocompromised patients with spiramycin, both claimed partial success (Moskovitz *et al* 1988, Portnoy *et al* 1984). By contrast, another study of

cryptosporidial diarrhoea in infants found that patients treated with spiramycin showed no improvement compared with those receiving a placebo (Wittenberg *et al* 1989). Spiramycin trials in animals indicate that this drug may have its effect in reducing the level of *C. parvum* infection indirectly, by causing vacuolation of enterocytes (Angus *et al* 1989). More recently, a related drug, azithromycin, showed promising results when tested against cryptosporidiosis in dexamethasone-immunosuppressed rats (Rehg 1991a). Prophylactic administration of the drug prevented infection. In addition, when given therapeutically to rats with an established infection, all signs of the disease disappeared after 9 days. However, the infection recurred after treatment was stopped, indicating that residual infection persists, but is undetected.

Of 15 anti-coccidials tested against cryptosporidiosis in mice, none prevented infection, but five reduced the numbers of oocysts shed (Angus *et al* 1984). The most effective of these (arprinocid), was ineffective in mice when given therapeutically, and failed to control cryptosporidiosis in lambs.

Several independent studies have reported promising results with halofuginone lactate. Natural infection with *Cryptosporidium* was successfully cured in 147 of 150 calves after 5-6 days of treatment (Villacorta *et al* 1991). Following withdrawal of the drug, animals which had received lower doses began to re-excrete oocysts. Halofuginone lactate was also found to be effective in

treating cryptosporidiosis in lambs (Naciri & Yvore 1989) and to inhibit development of *C. parvum* in cell culture (McDonald *et al* 1990). A recent evaluation of the anti-cryptosporidial efficacy of 23 sulphonamides concluded that five were effective in reducing the severity of infection and should be investigated further (Rehg 1991b).

Because of the lack of an effective drug for therapy or prophylaxis against cryptosporidiosis, a considerable amount of interest has been shown in the possibility of immunotherapy against the disease. Attempts to treat cryptosporidiosis by passive transfer of immune colostrum have given mixed results. Treatment of three immunocompromised patients with hyperimmune bovine colostrum led to clinical cure in two, and clinical and parasitologic cure in the third (Tzipori *et al* 1987). Remission of cryptosporidiosis in a child with agammaglobulinaemia following treatment with hyperimmune bovine colostrum has also been reported (Tzipori *et al* 1986). By contrast, another study which attempted treatment of three immunocompromised patients with immune bovine colostrum was unsuccessful (Saxon & Weinstein 1987). The ability of bovine dialyzable leucocyte extract (DLE) to protect against cryptosporidiosis in AIDS patients was tested in a double blind trial (McMeeking *et al* 1990). Six of seven patients given immune DLE gained weight and had less diarrhoea, but only one of seven receiving control DLE showed any clinical improvement.

Since there is currently no effective prophylactic or

therapeutic agent against cryptosporidiosis, control depends on reducing the risk of exposure to infective oocysts. Cryptosporidiosis in farm animals can be controlled by management strategies and disinfection. For example, by movement of stock into clean areas prior to parturition, since it is only young animals that are susceptible to clinical cryptosporidiosis.

Human cryptosporidiosis can be controlled by the use of disinfection procedures to reduce the risk of tranmission in domestic and hopital situations.

Blewett (1989a) tested 35 disinfectants for activity against *C. parvum* oocysts. Only three showed clear-cut efficacy at 22^{o}C. Others affected oocysts to a greater or lesser extent if used at higher temperatures, excessive concentrations or for very long exposure periods. These latter chemicals would not therefore be of use in routine disinfection procedures. Of the three chemicals which were effective at 22^{o}C, Exspor, a chlorine dioxide based disinfectant, was inhibited by protein, so its use may be limited. It was concluded that the other two, 10 volume H_2O_2, and oocide (Antec International), are suitable for domestic and agricultural use, respectively. The latter is applied in two stages, first a solution containing an ammonia source, a wetting agent, and a biocide, followed by an activator which initiates release of ammonia.

Cryptosporidium infections transmitted by contaminated water

supplies could be prevented by improved water treatment procedures. Current procedures including filtration and chlorination are not 100% effective (Smith 1990). Experiments have shown that exposure to UV light, or to high concentations of ozone (Peeters *et al* 1989, Casemore 1990) are lethal to oocysts. However, there are practical and financial problems associated with the use of such methods to treat large volumes of water.

Further investigation into the survival of oocysts under various environmental conditions, such as in water supplies, on pasture, and in slurry and other farm waste, may provide useful information on the epidemiology and transmission of cryptosporidiosis, which in turn may have important implications for future control strategies.

1.7 Immunology of *Cryptosporidium parvum* infection.

Results from surveys and experimental infections in laboratory animals suggest that cryptosporidiosis is age dependent (Fayer & Ungar 1986, Sherwood *et al* 1982). This age resistance is not entirely exposure related, since adult specific-pathogen-free mice are refractory to infection with *C. parvum*, whereas neonates are susceptible. In ruminants where infections are reported in adult animals, they are typically low grade and asymptomatic. However, experimental infections in calves show that both age and previous exposure affect their susceptibility

to cryptosporidiosis (Harp *et al* 1990). Age dependence in human cryptosporidiosis is less marked. Surveys show that infection is more common in children (Current & Garcia 1991, Fayer & Ungar 1986), but this may be exposure dependent, since clinical cryptosporidiosis has been reported in both immunocompetent and immunocompromised humans of all ages.

It seems likely that both non-specific and acquired immunity are involved in resistance to cryptosporidiosis. Various aspects of immunity to cryptosporidiosis have been studied using animal models in attempts to elucidate the mechanisms of protection against cryptosporidiosis.

1.7.1 Humoral immune responses to cryptosporidiosis.

Humoral responses to *Cryptosporidium* infection have been described in normal and immuno-compromised humans, and in a number of other mammalian species (Casemore 1987, Ungar *et al* 1986, Campbell & Current 1983, Mead *et al* 1988b, Hill 1989, 1990,Harp *et al* 1990, Whitmire & Harp 1991, Gardner *et al* 1991). However, the role of antibody in protection against cryptosporidiosis remains unclear.

Ungar *et al* (1986) measured *Cryptosporidium* specific IgG and IgM in immunocompromised and immunocompetent humans infected with *Cryptosporidium*. A specific serum antibody response to *Cryptosporidium* was present in AIDS patients, as well as in immunocompetent persons. Similar observations were reported by

Campbell and Current (1983), who found high *Cryptosporidium* specific antibody titres both in immunocompetent subjects who had recovered from cryptosporidiosis, and in AIDS patients with persistent infections; suggesting that an antibody response is not sufficient to clear the infection. These workers also found that two subjects with normal T cell function and hypogamma-globulinaemia had persistent cryptosporidiosis, and concluded that both humoral and cell-mediated immunity are necessary for recovery from cryptosporidiosis.

Absence of a role for antibody in protection against *Cryptosporidium* infection in mice is suggested by the work of Taghi-Kilani *et al* (1990). B cell depleted (anti-μ-treated) neonatal BALB/c mice did not differ from age matched controls in onset, peak or duration of cryptosporidiosis. Additionally, adult anti-μ-treated mice could not be infected with ten times the usual dose for infecting neonates, implying that the age related resistance to cryptosporidiosis is also independent of antibody.

Similarly, there is little evidence for a protective role for specific antibody in *Cryptosporidium* infection in ruminants. Harp *et al* (1990) looked at the effects of age and previous exposure on susceptibility of calves to cryptosporidiosis. They found that calves raised in isolation remained susceptible to infection up to three months of age, whereas animals which had been previously infected were resistant to reinfection. In calves infected at one week or one month of age, there was no

significant increase in *Cryptosporidium* specific antibody titre following challenge exposure. Despite this lack of a secondary antibody response, the calves were resistant to reinfection. However, all of these studies have looked at serum antibody. It may be more appropriate to look at secretory antibody present in the gut lumen, since this may be effective against extracellular stages of the parasite. Hill (1989) demonstrated *C. parvum* specific antibody in faeces and gut mucus, as well as in serum of experimentally infected, colostrum deprived lambs. IgA was the only isotype detected in faeces and gut mucus, and a rising titre of faecal IgA coincided with a decline in oocyst shedding. Hill suggested that secretory IgA may play an important role in recovery from cryptosporidiosis by agglutinating sporozoites and merozoites, and/or blocking receptor sites, and therefore preventing them from infecting host cells.

The role of passively transferred immunity to cryptosporidiosis via colostral antibody has been investigated by many workers. Epidemiological studies on cryptosporidiosis in infants in Guatemala (Cruz *et al* 1988), Costa Rica (Mata *et al* 1984) and Liberia (Hojilyng *et al* 1986) indicated that breast-fed infants were less likely to suffer from cryptosporidiosis than non breast-fed infants in the same populations. However, the possibility that increased prevalence of cryptosporidiosis in non breast-fed infants could be due to exposure to contaminated water and/or feeding utensils, rather than to a lack of breast milk, cannot be excluded. Attempts have been made to treat persistent cryptosporidiosis in immunocompromised humans using

hyperimmune bovine colostrum, but the results have not been consistent (Ungar *et al* 1990b, Tzipori *et al* 1986, 1987, Saxon & Weinstein 1987).

Passively transferred immunity to cryptosporidiosis has also been investigated in experimental infections in animals. Neonatal mice suckling dams that had recovered from cryptosporidiosis, or dams whose *Cryptosporidium* specific antibody level had been artificially raised by boosting with *Cryptosporidium* oocysts, were equally susceptible to cryptosporidiosis, compared with mice suckling naive dams (Arrowood *et al* 1989, Moon *et al* 1988). Attempts to protect calves from *Cryptosporidium* infection using hyperimmune colostrum have been more successful. Fayer *et al* (1989a) found that *Cryptosporidium* infected calves fed colostrum from cows that were hyperimmunized against *Cryptosporidium* had less diarrhoea, and shed less oocysts, than infected calves fed normal colostrum. They demonstrated high levels of *Cryptosporidium* specific IgG, IgM and IgA in the colostrum. The same colostrum also partially protected neonatal BALB/c mice against cryptosporidiosis, and neutralized the infectivity of sporozoites for neonatal mice *in vitro* (Fayer *et al* 1989b). These results implied that high levels of *Cryptosporidium* specific antibody in colostrum could partially protect against cryptosporidiosis. However, it is possible that the anti-parasitic activity may have been mediated by colostral factors other than antibody, such as cytokines or cells. Fayer *et al* (1990a) tested the ability of *Cryptosporidium*-specific

antibody purified from bovine colostrum, to protect neonatal mice against cryptosporidiosis. Mice treated with whole colostral whey or purified immunoglobulins had significantly reduced infection scores when compared with untreated mice. These results showed that colostral antibody alone could partially protect mice against cryptosporidiosis. In order to find out which antigens of the extracellular stages of the parasite are recognised by protective antibodies, several workers have carried out similar experiments using monoclonal antibodies (Mabs) instead of colostrum. Perryman *et al* (1990) showed that neonatal mice treated with Mabs which recognised sporozoite surface antigens were partially protected against oral challenge with *Cryptosporidium* oocysts. In addition, sporozoites treated with these Mabs *in vitro* lost their infectivity for neonatal mice. Subsequently, Bjorneby *et al* (1990) demonstrated that these Mabs also neutralized the infectivity of merozoites. In a similar study, Arrowood *et al* (1989) found that following infection with *Cryptosporidium* oocysts, neonatal mice treated with a mixture of Mabs recognising sporozoite surface antigens had significantly lower infection scores than control mice.

1.7.2 Cell-mediated responses to cryptosporidiosis.

The importance of T cells in resistance to, and recovery from, cryptosporidial infection is indicated by the fact that humans with T cell deficiencies have longer and more severe infections than those with normal T cell function (Crawford & Vermund

1988). Infections of this nature also occur in experimentally infected athymic rats and mice (Gardner *et al* 1991, Heine *et al* 1984a, Ungar *et al* 1990a), and in mice with experimentally induced T cell deficiencies (Ungar *et al* 1990a, Rehg *et al* 1987, Brasseur *et al* 1988). Chronically infected athymic mice clear the infection within a week of reconstitution with lymphoid cells from normal histocompatible mice which have recovered from *Cryptosporidium* infection (Ungar *et al* 1990a). In order to find out which T cell subset was involved in recovery from cryptosporidiosis, Ungar *et al* (1990a, 1991) investigated the effects of treatment with anti-CD4 and anti-CD8 Mabs on the susceptibility of mice to cryptosporidiosis. Neonatal mice treated with an anti-CD4 Mab developed persistent infections with diarrhoea and weight loss, similar to those observed in athymic mice. Anti-CD8 treated mice recovered at the same rate as untreated mice (Ungar *et al* 1990a). These results indicated that CD4+, but not CD8+, T cells are important in recovery of neonatal mice from cryptosporidiosis. However, further experiments indicated that CD4+ were not the only lymphocyte subset involved in protection against cryptosporidiosis. Adult anti-CD4 treated mice were much more resistant to cryptosporidiosis than their neonatal counterparts, developing chronic infections characterized by sparse oocyst shedding and no clinical symptoms (Ungar *et al* 1990a, 1991). This contrasts with adult athymic mice, in which severe infections with weight loss and diarrhoea occur. This suggests that some other lymphocyte subset, which is CD4- and absent in athymic mice, is important in preventing establishment of severe cryptosporidiosis, but is insufficient

to resolve a chronic infection.

A role for interferon-γ (IFN-γ) in controlling the severity of *Cryptosporidium* infection was indicated in experiments in which mice were treated with an anti-IFN-γ Mab (Ungar *et al* 1991). Treated mice shed significantly more oocysts for a longer period of time than untreated mice, but the infection resolved spontaneously after about two weeks. When anti-IFN-γ, and anti-CD4 treatment was combined, adult mice developed prolonged severe infections comparable to those seen in athymic mice. It would appear that both CD4+ T cells and IFN-γ play separate roles in protection against cryptosporidiosis, with the former controlling duration, and the latter controlling severity of infection (Ungar *et al* 1991). Harp and Moon (1991) used W/Wv (mast-cell deficient) mice to examine the role of mast cells in immunity to cryptosporidiosis. Normal mice and W/Wv mice infected at one week old showed no differences in susceptibility to, or recovery from *Cryptosporidium* infection, and both groups were resistant to reinfection. Adult W/Wv mice were more susceptible to primary infection than normal age matched controls, but both groups recovered at a similar rate. The authors concluded that recovery from, and subsequent immunity to cryptosporidiosis is independent of mast cells in neonatal mice, but the age resistance of adult mice to cryptosporidiosis may be partially dependent on mast cells, or other cell types affected by the W locus.

The emerging picture of immunity to cryptosporidiosis is one in

which several different immune mechanisms are involved, both specific and non-specific. It would appear that if one particular immune mechanism is inactivated, the infection can be controlled by other means. But in cases where the immune system is severely compromised, such that multiple effector mechanisms are affected, the infection may become prolonged and life-threatening.

1.8 Studies of *Cryptosporidium parvum* antigens.

Antigens of *Cryptosporidium parvum* that are recognised by sera from infected animals and humans have been investigated by many workers using immunoblotting techniques. While there is considerable variation between antigen profiles recognised by different sera, several workers have found consistent recognition of certain antigen bands with comparable molecular weights (mws).

Recognition of a 23kDa antigen of sporulated oocysts by serum antibody from 93% of infected humans was reported by Ungar and Nash (1986). Hill *et al* (1990) found that a 23kDa antigen of excysted oocysts was consistently recognised by serum and faecal IgA from experimentally infected lambs, and Lumb *et al* (1988a) reported that a 23kDa oocyst antigen was also recognised by nine of ten humans, and all of four goats infected with *Cryptosporidium*. Subsequently, Lumb *et al* (1989) showed that serum raised to this 23kDa antigen also recognised

Cryptosporidium antigens with mws of 37, 49, 58, 68, 120, 140, and 160kDa. They also used this serum in immunoelectron microscopy experiments to localize the 23kDa antigen to the sporozoite and merozoite surface. Mead *et al* (1988b) found that a 20kDa antigen of purified sporozoites was recognised by all sera in a study of infected humans, horses and calves. They concluded that this was the same as the 23kDa antigen described by Ungar and Nash (1986). In another study Arrowood *et al* (1989) showed that Mabs recognising a 20kDa antigen on the surface of sporozoites also bound to the surface of merozoites. Whitmire and Harp (1990) also described recognition of a 20kDa antigen by infected calves, and believed it to be the same antigen as the 20 and 23kDa antigens described by the aforementioned authors. However, in another study which looked at the antigens recognised by hyperimmune bovine colostral antibodies, Tilley *et al* (1990a) described recognition of two separate antigen bands, one with mw of 23kDa, and another with mw of 20-21.5kDa. These authors suggest that other workers are confusing at least two separate antigen bands. In support of this, Tilley *et al* (1990b) described two separate proteins of *C. parvum* sporozoites which are labelled by ^{125}I surface labelling, one with mw 18-20kDa, the other with mw 23kDa. They also demonstrated that the 18-20kDa protein bound a lectin specific for β-D-galactose, whereas the 23kDa peptide did not bind any of the lectins tested.

In addition to the 20 and 23kDa immunodominant antigens, another surface antigen present on both sporozoites and merozoites has

been described by Tilley *et al* (1991). A Mab recognising a 15kDa antigen on Western blots of *C. parvum* sporozoites bound to the surface of both sporozoites and merozoites. Competition assays indicated that N-acetyl-glucosamine, and to a lesser extent N-acetylgalactosamine were important constituents of the epitope bound by this Mab. The authors believe this 15kDa glycoprotein to be the same molecule as both a 14.5-16.5kDa antigen recognised by bovine colostral antibodies (Tilley *et al* 1990a), and a 12-17kDa antigen recognised by serum antibody and faecal IgA from infected lambs (Hill *et al* 1990). However, caution must be exercised when comparing antigens described in different studies, as bands with similar mws can be confused, as seems to have occurred in the case of the 20 and 23kDa antigens. Other low mw antigens described include an 11kDa band recognised by infected calves (Whitmire & Harp 1990), and a 9-10kDa band recognised by IgA from hyperimmune bovine colostrum (Tilley *et al* 1990a). Again, one could speculate as to whether these are the same molecule, but without further characterization one cannot draw any firm conclusions. When comparing results from different studies it may be worth considering how the animal has been exposed to *Cryptosporidium* antigens, as different routes of inoculation may result in recognition of different antigens. The literature indicates that considerably more antigen bands are recognised by animals which have been inoculated parenterally with oocyst or sporozoite preparations, than by animals which have been infected naturally, or by oral inoculation with oocysts (Hill 1989, Hill *et al* 1990, Whitmire & Harp 1991, Luft *et al* 1987). Additionally, work by Luft *et*

al (1987) indicates that the relative antigenicities of carbohydrate and protein epitopes of *Cryptosporidium* antigens may be dependent on the route of inoculation.

Evidence for the presence of a non-protein antigen on the surface of sporozoites and merozoites of *C. parvum* is put forward by Riggs *et al* (1989). A Mab which binds to the surface of merozoites and sporozoites also recognises a sporozoite antigen which migrates in the dye front during SDS-PAGE. This antigen is not labelled with ^{125}I or [^{35}S]-methionine, and is resistant to digestion with proteinase K, indicating a non-protein composition. In addition, it is eluted in the void volume of a Bio Gel column with an exclusion limit of 500kDa, indicating a very large molecule. Other surface antigens common to merozoites and sporozoites include proteins with mws of 98, 55 and 28kDa, all of which are recognised by a single Mab (Riggs *et al* 1989). Other workers have described Mabs which exhibit polar reactivity in immunofluorescence experiments. A Mab which binds to the anterior of sporozoites also recognises several bands ranging from 25-200kDa on Western blots (Arrowood *et al* 1989). In another study Bonnin *et al* (1991) described two Mabs, both of which bound to the micronemes of sporozoites and merozoites. These Mabs bound to a series of bands with mws between 40 and 210kDa. Several of these bands were recognised by both Mabs, others were recognised by only one. Periodate treatment of blots prior to incubation with Mabs prevented binding of one of the Mabs, indicating that it recognised a carbohydrate epitope.

The importance of carbohydrate moieties in immunogenicity of high mw *Cryptosporidium* antigens was also indicated in a study by Luft *et al* (1987). These workers found that recognition of several sporulated oocyst antigens with mws ranging from 72->100kDa, by mice inoculated with *Cryptosporidium* oocysts, was considerably reduced by pretreatment of the sporulated oocysts with mixed glycosidases.

1.9 *Cryptosporidium parvum* isolate variation.

Several studies aimed at identifying antigenic differences between *C. parvum* isolates have been carried out.

Using a panel of 16 anti-*C. parvum* Mabs, McDonald *et al* (1991) demonstrated that each of five *C. parvum* isolates showed a unique pattern of reactivity. Nina *et al* (1992) subsequently examined the antigenic profiles exhibited by diffferent *C. parvum* isolates by immunoblotting using monoclonal and polyclonal antibodies raised against the parasite. The presence or absence of certain bands appeared to be correlated with the host of origin, suggesting the existence of strains associated with particular host species. In a similar study Nichols *et al* (1991) used an anti-*C. parvum* Mab to probe Western blots of different *C. parvum* isolates from humans, lambs and calves. Each isolate gave one of four banding patterns. One pattern was unique to the lamb isolates, each calf isolate gave one of the

three remaining patterns, and two of these patterns were also exhibited by the human isolates.

Another method which has been used to investigate possible differences between *C. parvum* isolates, is restriction fragment length polymorphism analysis. Southern blots of *Eco*RI digested *C. parvum* DNA probed with pv47-2 (a probe containing repeated sequences (Longmire *et al* 1990)) revealed that DNA from three bovine isolates each gave an identical banding pattern, which was different to the patterns given by DNA from three human isolates (Ortega *et al* 1991). Furthermore, a 4.3kb fragment was present in each of the human isolates which was absent in each of the bovine isolates. One of the bovine isolates was passaged twice through calves, after which its restriction profile was unchanged. However, it would perhaps have been more interesting to passage one of the human isolates and re-examine its restriction profile.

Further investigation into possible differences between *C. parvum* isolates is required. Successive passage and mixed infection studies would be of particular interest in order to determine whether 'strain characteristics' are stable, and to find out whether cross fertilization occurs between 'strains'.

1.10 Genetic studies of *C. parvum*.

Little is known of the molecular biology of *Cryptosporidium*.

Studies of the organism at the molecular level have begun only recently, due, in part, to the lack of interest in *Cryptosporidium* before it was recognised as an important cause of diarrhoea in both humans and other animals. Another factor may be the difficulty in obtaining sufficient quantities of the parasite to carry out molecular studies.

The first published work on the molecular biology of *Cryptosporidium* involved field inversion gel electrophoretic separation of chromosome-sized DNA from two species of *Cryptosporidium* (Mead *et al* 1988a). Electrophoresis of *C. baileyi* gave six chromosomal bands, with apparent sizes ranging from 1,400kb to over 3,300kb. Electrophoresis of five *C. parvum* isolates each gave five chromosomal bands, with sizes falling in the same range as observed for *C. baileyi*. However, only two of the five bands coincided with bands of *C. baileyi*; conversely, there were no differences in migration patterns observed between the five isolates of *C. parvum*. A more recent study (Kim *et al* 1992), also employing gel electrophoresis of chromosomal DNA, confirmed the presence of five discrete chromosomal bands of *C. parvum*. However, these authors estimated the sizes of the bands at between 900 and 1400kb. Adding together the sizes of the observed bands, gives an approximate estimate of the size of the *C. parvum* genome at around 5.7Mb. However, the relative intensity of ethidium bromide staining of individual bands indicated that some may contain multiple chromosomes of nearly identical sizes. The above estimate is therefore likely to be an underestimate of the

true size of the *C. parvum* genome. Despite this, it seems likely that the genome of *C. parvum* is small in comparison with closely related Apicomplexa. For example, the genome of *Plasmodium falciparum* is approximately 20Mb (Goman *et al* 1982), and those of *Eimeria tenella* and *Toxoplasma gondii* have been estimated at 70Mb (Clarke *et al* 1987) and 97 Mb (Cornelissen *et al* 1984), respectively.

Johnson *et al* (1990) used ribosomal RNA sequence comparison to investigate the phylogenetic relationship of *Crptosporidium* to other Apicomplexa. They concluded that *Cryptosporidium* was not especially closely related to the other members of the Apicomplexa used in the study (*Plasmodium berghei, Sarcocystis gigantea, and Toxoplasma gondii*), but of these, it was closest to *P. berghei*. This contradicts the widely accepted classification of *Cryptosporidium* which places it in the suborder eimeriorina, together with *Eimeria, Sarcocystis and Toxoplasma*. This classification is supported by a phylogenetic analysis based on ultrastructural observations (Barta 1989), which provided evidence that *Cryptosporidium* is more closely related to *Eimeria* and *Sarcocystis* than to *Plasmodium*.

The overall G+C content of the *C. parvum* genome is unknown, but available sequence data indicate that it is likely to be relatively low. For example, the coding sequence of a *C. parvum* actin gene contains 38% G+C (Kim *et al* 1992), with only 28% G+C in the third codon position. In addition, 702bp of DNA flanking this gene contain only 27% G+C. Other *C. parvum* DNA

sequence data also point to a low G+C content for the *C. parvum* genome. This includes 2.3kb of DNA sequenced as part of a project to identify suitable sequences for diagnosis of Cryptosporidiosis by PCR, which contained 35% G+C (Laxer *et al* 1991). Another group sequenced inserts from random clones picked from a *C. parvum* λZAP library (Dykstra *et al* 1991). They found that the average G+C content of the sequences was 32.6%.

Several studies carried out since 1988 have identified protein-encoding DNA sequences of *C. parvum*. A *C. parvum* actin gene was isolated using a heterologous cDNA probe (Nelson *et al* 1991, Kim *et al* 1992), and the *C. parvum* gene encoding thymidylate synthase-dihydrofolate reductase was isolated using conserved sequences to amplify the gene by PCR (Gooze *et al* 1991). In addition, *C. parvum* antigen-encoding DNA sequences have been isolated by screening expression with anti-*C. parvum* antiserum (Dykstra *et al* 1991, Petersen *et al* 1992).

1.11 Aims of the present study.

A great deal of interest in *Cryptosporidium* has been shown by the medical and veterinary professions since it was recognised as a cause of diarrhoeal illness in both humans and animals. Over the last ten years numerous papers have been published on various aspects of *Cryptosporidium* and cryptosporidiosis, including work on epidemiology, immunology, clinical and pathological observations, ultrastructure and chemotherapy.

However, our knowledge of the genetics of *Cryptosporidium* is still very poor. In October 1988, when the present study was initiated, the only published work on the genetics of the parasite was an investigation into the size and number of chromosomes of two species of *Cryptosporidium* (Mead *et al* 1988a, section 1.10).

The aim of the present study was to identify, by screening *C. parvum* DNA libraries with anti-*C. parvum* antisera, genes, or gene fragments encoding antigens of the parasite. This approach would enable characterization of *C. parvum* proteins and their amino acid sequences. In addition, isolation and expression of *C. parvum* antigen-encoding genes may facilitate the study of immunity to cryptosporidiosis, and may ultimately contribute to the development of immunological control strategies against cryptosporidiosis, such as passive transfer of immunity.

Identification and sequencing of protein-encoding genes of *C. parvum* could also provide information on aspects of the parasite genome such as G+C content and codon usage. Such information may be of use in subsequent molecular studies of *C. parvum*. In addition, identification of *C. parvum* DNA sequences may be of use in developing DNA-based methods for detection and typing of the organism, which could be used in epidemiological investigations.

CHAPTER 2. MATERIALS AND METHODS.

2.1 Frequently used solutions.

TE: 10mM Tris, 1mM EDTA; adjusted to pH 8.0 with HCl.

TBS: 50mM Tris, 150mM NaCl; adjusted to pH 8.0 with HCl.

TBST: TBS containing 0.5% (v/v) Tween-20.

20x SSC: 3M NaCl, 0.3M sodium citrate; adjusted to pH 7.0 with NaOH.

SM: 50mM Tris, 100mM NaCl, 10mM $MgSO_4$, 0.1% (w/v) gelatin; adjusted to pH 7.5 with HCl.

RNase A: 10mg ml^{-1}. Heated to 100°C for 15 min, cooled to room temperature and stored in aliquots at -20°C.

Phenol: Equilibrated with 1M Tris-HCl pH 8.0, then with 0.1M Tris-HCl pH 8.0. 8-Hydroxyquinoline added to 0.1% (w/v) and aliquots stored at -20°C.

Chloroform: Chloroform and isoamyl alcohol mixed in a ratio of 24:1 (v:v); stored in a dark bottle at room temperature.

TEG: 25mM Tris-HCl pH 8.0, 10mM EDTA, 50mM glucose.

HBSS: 0.406mM $MgSO_4$, 0.491mM $MgCl_2$, 1.26mM $CaCl_2$, 137mM NaCl, 5.37mM KCl, 1.07mM Na_2HPO_4, 0.441mM KH_2PO_4, 0.002% (w/v) phenol red.

PBS: 137mM NaCl, 26.8mM KCl, 8.10mM Na_2HPO_4, 1.47mM KH_2PO_4.

2.2 Growth media.

The quantities of all media components are given in grams per litre final volume of medium, except in cases where final concentrations of components are stated.

L-broth: 10g Difco Bacto tryptone, 5g Difco yeast extract, 5g NaCl.

L-agar: as for L-broth, but including 15g Difco agar.

L-agarose: as for L-broth, but including 15g BDH agarose 15.

L-top agar: as for L-broth, but including 7.5g Difco agar.

L-top agarose: as for L-broth, but including 7.5g BDH agarose 15.

Minimal agar: 6g Na_2HPO_4, 3g KH_2PO_4, 1g NH_4Cl, 0.5g NaCl, 15g agar, 1mM $MgSO_4$, 0.1mM $CaCl_2$, 1mM thiamine-HCl, 0.2% (w/v) glucose.

Ampicillin: When required, media were supplemented with 50μg ml^{-1} ampicillin, (final concentration).

2.3 *Escherichia coli* strains.

Escherichia coli strains used in this study are listed in table 2.1

2.4 Cloning Vectors.

The cloning vectors used in this study are listed in table 2.2.

2.5 Antisera.

The antisera used in this study, and methods used to raise them are detailed in tables 2.3-2.5.

N.B. All Home Office licensed animal procedures described in this thesis were carried out by Mr S. E. Wright.

2.6 Parasitological techniques.

2.6.1 Maintenance of *C. parvum*.

The isolate used in this study was obtained from red deer calves, and was associated with severe diarrhoea (Blewett, 1988). This isolate has been maintained by 3-6 monthly passage in male lambs or bovine calves. Calves or lambs, 5-10 days old,

Table 2.1 *Escherichia coli* strains.

Strain	Genotype	Reference
JM109	*recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi* Δ*(lac-proAB) F'[traD36 proAB+ lacI*^{q} *lacZ* Δ*M15]*	Yannisch-Perron *et al* (1985)
NM522	*supE thi* Δ*(lac-proAB)hsd5 F' [proAB+ lacI*^{q} *lacZ* Δ*M15]*	Gough & Murray (1983)
Y1089	*araD139* Δ*(lac)U169* Δ*(lon) rpsL hflA150[chr*::Tn*10(tet*^{r}*)]* (pMC9)	Young & Davis (1983b)
Y1090	*araD139* Δ*(lac)U169* Δ*(lon) rpsL supF trpC22* ::Tn*10 (tet*^{r}*)* (pMC9)	Young & Davis (1983b)

Table 2.2 Cloning vectors.

Vector	Reference
λgt11	Young & Davis (1983a)
pMS	Scherf *et al* (1990)
pBR322	Bolivar *et al* (1977)
Bluescript SK+	Short *et al* (1988)

Table 2.3 Antisera.

Antiserum	Source
Rat anti-*C. parvum* antiserum *	MRI
Serum from gnotobiotic lambs infected with *C. parvum* **	MRI
Mouse anti-β-galactosidase serum	Sigma
Rabbit anti-P786 antiserum ***	MRI
Anti-rat HRPO conjugate	Dako
Anti-sheep HRPO conjugate ****	MRI
Anti-rabbit HRPO conjugate	Dako
Anti-mouse alkaline phosphatase conjugate	Sigma

* See table 2.4
** See table 2.5
*** See section 4.3
****Preparation of this antiserum is detailed in Hill (1989)

Table 2.4 Protocol used to raise anti-*C. parvum* antiserum in Wistar rats.

Age (days)	Inoculum	Route of inoculation
5	10^6 oocysts	oral
16	10^6 oocysts	oral
38	10^6 excysted oocysts*	intramuscular (2 sites)
45	10^6 excysted oocysts*	subcutaneous (2 sites)
52	10^6 excysted oocysts*	intramuscular (2 sites)
60	The rats were anaesthetised with halothane (May & Baker) and exsanguinated by cardiac puncture. Blood samples were pooled and serum prepared from them stored at -20°C.	

*Oocysts were excysted *in vitro*, resuspended and freeze-thawed twice in 100µl of water. The volume was then brought to 500µl with PBS.

Table 2.5 Gnotobiotic lamb serum.

Lamb	Inoculum	Blood samples taken at (age)
96	10^6 oocysts inoculated orally at 10 days of age	10 days and 26 days, except lamb 96 from which samples were taken at 10 and 21 days.
108		
146		
150		
151		
152		
64	Approximately 50 oocysts per day given orally with feed, from 0 to 22 days of age.	22 days
66		
68		
69		
74		
76		
81		
84		
89		

were infected with 10^7 or 10^6 oocysts, respectively. At the same time each animal was fitted with a harness and detachable faecal bag to allow collection of total daily faecal output. Microscopical examination of faecal samples was carried out daily, using a haemocytometer. For this purpose faeces were diluted in 0.16% (w/v) malachite green containing 1% (w/v) sodium dodecyl sulphate (SDS) and samples found to contain more than 10^6 oocysts ml^{-1} were kept for subsequent extraction of oocysts.

2.6.2 Purification of *C. parvum* oocysts.

Oocysts were extracted and purified from faeces by sedimentation and differential centrifugation. An aqueous dilution of faeces (5-20%) was acidified with 2% (v/v) sulphuric acid to a final pH of 5-6, which caused most of the faecal solids to flocculate and sediment. The fluid phase was decanted and centrifuged at 3000g to pellet solids including oocysts. The pellet was then resuspended in a small volume of water and washed in tap water by repeatedly centrifuging at 500g and retaining the pellet until the supernatant was clear. After the final wash the oocysts were suspended in 1% (w/v) SDS for an hour at room temperature, then washed by resuspending in tap water and repeatedly centrifuging at 500g until the supernatant was clear. Purified oocysts were stored in water at $4^{o}C$.

2.6.3 *In vitro* excystation of *C. parvum* oocysts.

The required number of oocysts were sedimented by centrifugation, and resuspended in HBSS containing 1% (w/v) trypsin (from beef pancreas, BDH chemicals), adjusted to pH 2.5-3 with HCl. The mixture was incubated at 37°C for an hour. Subsequently, the oocysts were pelleted by centrifugation at 500g then resuspended in HBSS containing 0.05% (w/v) sodium deoxycholate (BDH) and 0.2% (w/v) sodium hydrogen carbonate, pH 8.0-8.4. This mixture was incubated at 37°C for 30 min after which a small drop of the mixture was placed on a microscope slide. A coverslip was placed on top and the sample examined by phase contrast microscopy using a X40 objective lens. The numbers of intact oocysts, oocyst shells and sporozoites in several fields of view were counted (until the sum exceeded 250), and from this the excystation percentage and sporozoite ratio were calculated, i.e.

$$\text{Excystation percentage} = \frac{\text{No. of shells}}{\text{No. of intact oocysts + No. of shells}} \times 100$$

$$\text{Sporozoite ratio} = \frac{\text{No. of sporozoites}}{\text{No. of shells}}$$

Typical values for freshly prepared oocysts were >90% excystation, and >2.5 for the sporozoite ratio.

2.6.4 Separation of *C. parvum* sporozoites and oocyst shells.

The required number of oocysts were excysted *in vitro* and after the second incubation the excystation mixture was resuspended in 1ml of PBS. This was then layered onto the top of two identical 11ml Percoll (Pharmacia) gradients. The continuous gradients contained from 15 to 50% or 15 to 90% (v/v) of percoll diluted in 0.85% (w/v) NaCl (separations carried out using both these concentration gradients were found to be satisfactory). The gradients were centrifuged at 1000g for 20 min. After centrifugation approximately 0.5ml fractions were taken, starting from the bottom of each tube. This was done using a 1ml syringe, to which was attached a three way valve. One outlet was attached to a long metal probe which was inserted into the tube containing the gradient. To the other outlet was attached a piece of tubing through which each fraction was emptied into an eppendorf tube. A small drop from each fraction was examined using phase contrast microscopy to find out which fractions contained sporozoites, and which contained oocyst shells. Appropriate fractions were pooled, diluted with four volumes of PBS, and the contents pelleted by centrifugation at 1000g for 5 min. Samples to be subsequently analysed by SDS-PAGE were resuspended in SDS-PAGE sample buffer and boiled for 5 min before being stored at -20^{o}C.

2.6.5 Preparation of *Sarcocystis* spp. cystozoites.

Sheeps hearts, obtained from culled ewes, were trimmed of fat and veins, cut into pieces and washed well to remove as much blood as possible. The pieces were then minced using a household meat mincer and incubated for 2 hr at 37^{o}C in 0.85% (w/v) NaCl, 0.26% (w/v) pepsin A, and 0.7% (v/v) HCl with constant stirring. The mixture was strained through a sieve and the liquid centrifuged at 1000g for 10 min. The supernatent was discarded and the pellets resuspended in 30% (v/v) Percoll, 0.85% (w/v) NaCl. This was then centrifuged at 1000g for 10 min and the pellet containing *Sarcocystis* spp. cystozoites resuspended in 0.85% (w/v) NaCl.

2.7 Bacteriological techniques.

2.7.1 Growth and maintenance of *E. coli*.

Bacteria were grown on solid and liquid media using standard techniques as described by Maniatis *et al* (1982). Where appropriate, medium was supplemented with ampicillin, IPTG and X-gal. *E. coli* strains/clones were maintained using standard techniques as detailed by Maniatis *et al* (1982). Short term maintenance (2-3 weeks) was accomplished on the surface of agar media plates stored at 4^{o}C. Long term storage was at -70^{o}C in medium containing 15% (v/v) glycerol.

2.7.2 Preparation of competent cells.

Competent bacterial cells were prepared by either of two methods.

i) Fresh competent cells were prepared by the calcium chloride procedure described in Maniatis *et al* (1982), based on the original method of Mandel and Higa (1970).

An overnight culture of *E. coli* JM109 or NM522 was diluted 1:50 in fresh L-broth and grown to an OD_{600} of 0.5. The cells were then cooled on ice for 15 min before being harvested by centrifugation at 3000g for 10 min and resuspended in 0.5 volumes of cold ($4^{o}C$), sterile 50mM $CaCl_2$. The cells were then left on ice for 30 min then pelleted as before and resuspended in 0.075 volumes of cold sterile 50mM $CaCl_2$. The cells were left on ice for at least 1 hr before being used. Competent cells prepared in this way were kept for a maximum of 2 days.

ii) Competent cells were also prepared and stored at $-70^{o}C$ for future use (Chung *et al* 1989).

An overnight culture was diluted 1:50 in L-broth and grown to an OD_{600} of 0.3-0.4. Cells were cooled on ice for 15 min and then harvested at 3000g for 10 min. Next, they were resuspended in 0.1 volumes of ice cold L-broth containing 10% (w/v) PEG (molecular weight approximately 8000), 5% (v/v) DMSO and 25mM

$MgCl_2$. Cells were dispensed into 200ml aliquots and stored at -70°C for future use.

2.7.3 Preparation of phage plating cells.

A single colony of Y1090 was picked into 10ml of L-broth containing 50mg ml^{-1} of ampicillin, and 0.4% (w/v) maltose, and grown overnight with shaking at 37°C. 1ml of the overnight culture was added to 50ml of fresh medium of the same composition and incubated at 37°C with shaking until the cells had grown to an OD_{600} of 0.5. The culture was then cooled on ice and the cells harvested by centrifugation at 3000g for 10 min. Subsequently, the cells were resuspended in 15ml of ice cold sterile 10mM $MgSO_4$, and were then ready for infection with phage.

2.8 Phage techniques.

2.8.1 *In vitro* packaging of phage DNA.

Phage DNA was packaged into phage heads using commercially avialable packaging extracts (Amersham). Extracts A and B were thawed on ice. Immediately after thawing, 10µl of extract A and 15ml of extract B were each added to the DNA to be packaged. The contents of the tube were gently mixed, then incubated at 20°C for 2 hr. The volume was then made up to 0.5ml, 10ml of chloroform was added and the phage stock stored at 4°C.

2.8.2 Plating lgt11.

10-100ml of an appropriate dilution of bacteriophage stock were added to 100ml of Y1090 plating cells, and the cells then incubated at 37°C for 15 min. 4ml of liquid top agar at 45°C containing 50ml ml^{-1} ampicillin were then added to the cells and the mixture poured onto the surface of a dry L-agar plate containing 50mg ml^{-1} ampicillin. The top agar was allowed to set, after which the plates were incubated at 43°C overnight.

2.8.3 Picking plaques.

Individual plaques were cored out of agar using sterile pasteur pipettes and each added to eppendorf tubes containing 0.5ml of SM buffer and a drop of chloroform. The tubes were then stored at 4°C. Single plaques contained approximately 10^6-10^7 pfu.

2.8.4 Plate lysates.

Approximately 10^5 pfu (5-50ml) from a single plaque suspension were used to infect fresh Y1090 plating cells and plated out as described in section 2.8.2. After incubation overnight at 43°C confluent lysis had occurred and the phage were eluted in 3ml of SM buffer by gentle shaking for 2 hr at room temperature. Bacterial debris were then removed by centrifugation at 8000g for 10 min. The lysate was then stored over chloroform at 4°C, or used to prepare phage DNA as described in section 2.9.6.

2.8.5 Preparation of lgt11 lysogens.

A single colony of *E. coli* Y1089 was picked into 10ml of L-broth containing 50mg ml^{-1} ampicillin and 0.4% (w/v) maltose and grown overnight at 37°C with shaking. 1ml of the overnight culture was added to 50ml of fresh medium of the same composition, and incubated at 37°C with shaking until the cells had grown to an OD_{600} of 0.5. 1M $MgCl_2$ was then added to the cell culture to give a final concentration of 10mM, and cells were dispensed into 100µl aliquots on ice. A 100ml aliquot was infected with 10-50ml of a dilution of the appropriate phage stock, containing approximately 10^8 pfu, and incubated at 32°C for 20 min. After incubation, the cells were plated on L-agar plates containing 50µg ml^{-1} ampicillin, at a density of approximately 250 per 90mm plate, and incubated at 32°C overnight. 20 colonies were picked from each plate and spotted onto each of two L-agar plates containing 50µg ml^{-1} ampicillin. One plate was incubated at 32°C and one at 43°C. Lysogens were identified by their ability to grow at 32°C but not at 43°C.

2.9 DNA techniques.

2.9.1 Preparation of *C. parvum* DNA.

C. parvum oocysts were excysted *in vitro* and then the excystation mixture was washed by centrifuging at 500g for 5 min and resuspending the pellet in PBS. The mixture was then recentrifuged, and the sporozoites lysed by resuspending in 50mM

Tris-HCl pH 8.0, 50mM EDTA, 1% (w/v) SDS. Proteinase K was added to a final concentration of 100μg ml^{-1} and the mixture incubated at 50°C for 3 hr. Following incubation, the lysate was extracted with phenol, phenol:chloroform 1:1 (v:v) then with chloroform, and the nucleic acids precipitated with ethanol and redissolved in TE. RNase A was then added to a final concentration of 50μg ml^{-1} and the solution incubated at 37°C for 1 hr. Finally, the solution was re-extracted with phenol, phenol:chloroform, and chloroform, and the nucleic acids precipitated with ethanol. The pellet was redissolved in TE and stored at -20°C.

2.9.2 Preparation of *Toxoplasma gondii* and *Sarcocystis* spp. DNA.

Extraction of DNA from *T. gondii* tachyzoites (provided by Mr S. E. Wright, Moredun Research Institute) and *Sarcocystis* spp. cystozoites was carried out as described for *C. parvum* sporozoites.

2.9.3 Preparation of *E. coli* genomic DNA.

A 10 ml overnight culture of *E. coli* JM109 was harvested by centrifuging at 3000g for 10 min. The cells were resuspended in 2ml of TEG and lysed by the addition of 150μl of a freshly prepared solution of lysozyme (7mg ml^{-1}) and 400μl of 10% (w/v) SDS. The mixture was incubated at 37°C for 1 hr, then extracted twice with an equal volume of phenol:chloroform 1:1 (v:v). The nucleic acids were precipitated by the addition of 0.1 volumes of 3M CH_3COONa, and 2.2 volumes of ethanol, spooled, washed with

70% ethanol and dissolved in 100µl of TE. RNase was added to a final concentration of 50µl ml^{-1} and the solution incubated at 37°C for 1 hr. Finally, the solution was re-extracted with phenol:chloroform 1:1 (v:v), re-precipitated with ethanol and redissolved in TE.

2.9.4 Small scale preparation of plasmid DNA.

Double stranded plasmid DNA was prepared from 10ml overnight cultures grown under antibiotic selection by the method described in Maniatis *et al* (1982), based on the original method of Birnboim and Doly (1979).

1.5ml of the culture was pelleted by centrifugation in a microcentrifuge and the cells resuspended in 100µl TEG (25mM Tris-HCl pH 8.0, 10mM EDTA, 50mM glucose). 200µl of a freshly prepared solution of 0.2M NaOH, 1% (w/v) SDS was then added and the contents of the tube mixed by inverting several times. Genomic DNA, cellular debris and protein were then precipitated by the addition of 150µl of a solution containing 3M potassium and 5M acetate pH 4.8. These debris were pelleted by centrifugation at 10,000g for 5 min and the supernatant containing the supercoiled plasmid DNA was extracted with phenol:chloroform (1:1) and then precipitated by the addition of 1ml of ethanol. The DNA was redissolved in 50µl of TE containing 20µg ml^{-1} RNase A.

2.9.5 Large scale preparation of plasmid DNA.

A single bacterial colony was inoculated into 5ml of L-broth containing 50μg ml^{-1} ampicillin, and grown with shaking overnight at 37°C. Subsequently, 500ml of L-broth containing 50μg ml^{-1} ampicillin were inoculated with the 5ml overnight culture and grown overnight with shaking at 37°C. Bacterial cells were harvested by centrifugation at 4000g for 10 min at 4°C. Cells from 250ml of culture were resuspended in 20ml TEG and lysed by the addition of 2ml of a freshly prepared solution of lysozyme (10mg ml^{-1}) followed by 40ml of a freshly prepared solution of 0.2M NaOH, 1% (w/v) SDS. The mixture was incubated on ice for 10 min and then chromosomal DNA, high molecular weight RNA and protein/membrane/SDS complexes were precipitated by the addition of 20ml of a solution containing 3M potassium and 5M acetate pH 4.8. The precipitate was pelleted by centrifugation at 22000g for 20 min at 4°C, and the supernatant recovered. Plasmid DNA was subsequently precipitated by the addition of 0.6 volumes of propan-2-ol and pelleted by centrifugation at 15,000g for 15 min at room temperature. The pellet was then washed with 70% (v/v) ethanol and redissolved in 8ml of TE. To every ml of this solution was added exactly 1g of caesium chloride. 0.8ml of ethidium bromide solution (10mg ml^{-1}) was then added for every 10ml of the DNA/caesium chloride solution, resulting in a final density of 1.55g ml^{-1}. This solution was centrifuged at 3000g at room temperature to remove protein and cellular debris. The supernatant was then

transferred into Beckman "quickseal" tubes and centrifuged to equilibrium at 165,000g in a Beckman Ti70 rotor for 48 hr at 20°C. After centrifugation, two bands located near the centre of the gradient were visible in ordinary light. The upper band, consisting of chromosomal DNA and nicked plasmid DNA, was collected first to prevent contamination of the lower band, using an 18-gauge hypodermic needle. The lower band, containing closed circular plasmid DNA, was then collected in the same way. Ethidium bromide was removed from the plasmid DNA solution by extracting with water-saturated butan-1-ol, then the DNA was precipitated by the addition of 2 volumes of water and 6 volumes of ethanol.

2.9.6 Preparation of phage DNA.

An aliquot of bacteriophage suspension (approx 10^5 pfu) was plated on 85mm plates with *E. coli* Y1090, and incubated overnight resulting in confluent lysis of the bacterial lawn. Bacteriophage particles were eluted in 4ml of SM, and bacterial debris removed by centrifugation at 8000g for 10 min at 4°C. Bacterial nucleic acids were removed by incubation with RNase A and DNase I, both at 1μg ml^{-1}, for 30 min at 37°C. The bacteriophage particles were then precipitated by the addition of 20% (w/v) PEG, 2M NaCl in SM, followed by incubation on ice for 1 hr. The precipitated phage particles were pelleted by centrifugation at 10,000g for 20 min at 4°C and then resuspended in 0.5ml of SM. After this the bacteriophage particles were lysed by the addition of 5μl of 10% (w/v) SDS and 5μl of 0.5M

EDTA and subsequent incubation at 68°C for 15 min. The solution was then purified by extraction with phenol, phenol:chloroform 1:1 (v:v), then chloroform, and the bacteriophage DNA precipitated by the addition of an equal volume of propan-2-ol. Precipitated phage DNA was pelleted by centrifugation at 10,000g and washed with 70% (v/v) ethanol before being dissolved in 50µl of TE containing 20µg ml^{-1} RNase A.

2.9.7 Restriction enzyme digestion.

Restriction enzyme digestion was routinely carried out in a volume of 10-20µl. Each reaction was buffered with one of four buffers (table 2.6), which was added to the reaction at a 10x concentration. Restriction enzyme was added to a final concentration of approximately 10units μg^{-1} DNA and the reaction volume made up with sterile distilled water. The mixture was then incubated at the appropriate temperature for 90 min.

2.9.8 Ligation of DNA.

Ligation reactions were prepared by mixing the appropriate amounts of vector DNA, insert DNA and 10x concentrated ligation buffer (750mM Tris-HCl pH 7.8, 200mM $MgCl_2$, 10mM spermidine, 0.1% (w/v)BSA, 200mM dithiothreitol (DTT), 10mM ATP). The reaction volume was made up with sterile distilled water and 1-3 units of T4 ligase was added. Reactions were carried out in a volume of 10-20µl and incubated at room temperature for 2 hr, or at 16°C for 16-20 hr.

Table 2.6 Restriction enzyme buffers.

Buffer	Final concentration	Restriction enzymes
1	100mM Tris-HCl, pH7.5; 50mM NaCl; 6mM $MgCl_2$; 0.01% (w/v) BSA	EcoRI, EcoRV, PstI. PvuI, XbaI
2	20mM Tris-HCl, pH7.5; 50mM NaCl; 6mM $MgCl_2$; 0.01% (w/v) BSA, 10mM 2-mercaptoethanol	BglII, HindIII, SphI
3	33mM Tris-CH_3COO, pH 7.9; 66mM CH_3COOK; 10mM$(CH_3COO)_2Mg$; 0.01% (w/v) BSA; 10mM 2-mercaptoethanol	AluI, KpnI, SacI, SmaI
4	10mM Tris-HCl, pH8.0; 150mM NaCl; 6mM $MgCl_2$; 0.01% (w/v) BSA; 10mM 2-mercaptoethanol	BamHI

2.9.9 Dephosphorylation of DNA.

The terminal 5′ phosphates were removed from DNA by treatment with calf intestinal phosphatase (CIP). To the DNA was added 5µl of 10x CIP buffer (0.5M Tris-HCl pH 9.0, 10mM $MgCl_2$, 1mM $ZnCl_2$, 10mM spermidine), and water to 48µl. 0.01 unit of CIP was then added and the reaction incubated at 37°C for 30 min. After this, a second aliquot of 0.01 unit CIP was added and the incubation continued for a further 30 min. 40µl of water, 10µl of STE buffer (100mM Tris-HCl pH 8.0, 1M NaCl, 10mM EDTA) and 5µl of 10% (w/v) SDS was then added and the reaction heated to 68°C for 15 min in order to inactivate the CIP. The reaction mixture was then extracted twice with phenol:chloroform 1:1 (v:v) and twice with chloroform, and the DNA precipitated with ethanol.

2.9.10 Blunt ending DNA molecules with T4 polymerase.

To the DNA to be blunt ended ws added 2µl of 10x T4 polymerase buffer (0.33M Tris-CH_3COO pH 8.0, 0.66M CH_3COOK, 0.1M $(CH_3COO)_2Mg$, 5mM DTT, 1mg ml^{-1} BSA) and the volume made up to 19µl with sterile distilled water. 1µl(1 unit) of T4 polymerase was then added and the reaction incubated at 37°C for 5 min to allow the 3′-5′ exonuclease acivity to remove 3′ overhangs. After this incubation 2µl of a 2mM solution of the appropriate dNTP (2mM dGTP for DNA molecules with *Kpn*I and *Sph*I ends) was added and the reaction incubated at 37°C for a further 10 min resulting in an equilibrium between the 3′-5′ exonuclease

activity and the 5′-3′ polymerase activity and producing a blunt ended molecule.

2.9.11 Transformation of competent *E. coli*.

Up to 50μg of plasmid DNA was added to each 200μl aliquot of competent cells on ice. The DNA and cells were then incubated on ice for 30 min and subsequently heat-shocked at 42°C for 90 sec before being returned to ice for a further 5 min. 1ml of L-broth was then added, and the cells incubated at 37°C for 1 hr to allow expression of antibiotic resistance. 100 and 200μl aliquots were then spread on L-agar plates containing 50μg ml^{-1} ampicillin and incubated at 37°C overnight.

2.9.12 Agarose gel electrophoresis of DNA.

Nucleic acid grade "ultrapure" agarose was dissolved in TAE buffer (40mM tris-acetate, 2mM EDTA, pH 7.5-7.8), and ethidium bromide was added to the agarose solution to a final concentration of 0.5μg ml^{-1} before pouring the gel. 0.2 volume of DNA loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 15% (w/v) Ficoll Type 400; Pharmacia) was added to DNA samples before loading, and gels were electrophoresed in TAE buffer at 1-4 volts cm^{-1} until the required resolution was achieved. *Hin*dIII restricted bacteriophage λ DNA, or 1kb ladder markers (BRL) were co-electrophoresed to provide molecular weight standards. DNA fragments were visualised by fluorescence of bound ethidium bromide in UV light of 302nm

wavelength.

2.9.13 Southern transfer of DNA to nylon filters.

Transfer of DNA from agarose gels to nylon filters was carried out using the method of Smith and Summers (1980).

The gel was incubated in 0.25M HCl for 15 min then rinsed in distilled water and incubated in denaturing solution (1M NaOH, 1.5M NaCl) for 30 min. Next, the gel was placed in neutralizing solution (1M CH_3COONH_4, 0.02M NaOH) for 1 hr. Transfer to "Hybond N" nylon membrane (Amersham) was carried out by placing the membrane, previously soaked in neutralizing solution, on top of the gel, followed by three sheets of Whatman 3MM filter paper, also soaked in neutralizing solution. Finally, a 3cm stack of paper towels and a weight were placed on top of the filter paper. Transfer was allowed to proceed for 12-24 hours and the DNA was then fixed to the membrane by UV irradiation, (0.4Joules cm^{-2}).

2.9.14 Polyacrylamide gel electrophoresis of DNA.

Polyacrylamide gel electrophoresis (PAGE) of DNA was based on the method of Herring *et al* (1982).

7.5% polyacrylamide gels were prepared by mixing appropriate volumes of 30% acrylamide stock solution (29.2% (w/v) acrylamide, 0.8% (w/v) bisacrylamide), 5x Loening 'E' buffer

(180mM tris 7-9, 150mM NaH_2PO_4, 5mM EDTA), and distilled water. TEMED and ammonium persulphate were added to the gel mix to give final concentrations of 0.1% (v/v), and 0.1% (w/v) respectively, immediately before the gel was poured. Gels were cast and run in the "mini protean" gel system (Biorad) at 200 volts (constant voltage setting).

2.9.15 Silver staining DNA in polyacrylamide gels.

Nucleic acid fragments were stained using a modification of the silver-staining technique described by Herring *et al* (1982).

Following electrophoresis, the gel was fixed by soaking in a solution containing 10% (v/v) ethanol and 0.5% (v/v) ethanoic acid. The gel was then stained in 11.2mM silver nitrate for 10 min. After a quick wash in distilled water, the bands were developed by soaking the gel in 0.75M NaOH containing 0.25% (v/v) formaldehyde for 2-3 minutes. When the bands had developed to the desired intensity, the developing solution was removed and the reaction stopped by washing the gel in 75mM Na_2CO_3.

2.9.16 Recovery of DNA from agarose gels.

DNA fragments were recovered from agarose gel by the technique of Vogelstein and Gillespie (1979) using the commercially available "Gene Clean" kit (Stratech Scientific Ltd).

Briefly, an agarose gel slice containing the required DNA

fragment was excised from the gel. To the gel slice was added three volumes of 3M NaI and the mixture incubated at 50°C to melt the agarose. "Glass milk" was then mixed with this solution and the mixture was left at room temperature for 5 min to allow the DNA to bind to the glass. The DNA-glass matrix was pelleted and washed three times with "NEW wash" and the DNA was then eluted in 10µl of sterile distilled water at 50°C for 5 min.

2.9.17 Preparation of digoxigenin-labelled DNA probes.

Probe DNA was labelled with digoxigenin using the commercially available "DNA labelling and detection kit, non-radioactive" (Boehringer Mannheim).

Purified DNA was labelled by random-primed incorporation of digoxigenin-labelled deoxyuridine-triphosphate according to the manufacturer's instructions, as follows. The DNA was denatured by boiling for 10 min, then incubated with 2µl random hexanucleotide mix, 2µl dNTP labelling mix and 2 units of Klenow enzyme in a total volume of 20µl for 2-6 hr at 37°C. Random hexanucleotide mix contained 62.5 A_{260} units ml^{-1} of random hexanucleotides in a solution of 0.5M Tris-HCl pH 7.2, 0.1M $MgCl_2$, 1mM DTE, 2mg ml^{-1} BSA. The dNTP labelling mix contained dATP, dGTP and dCTP each at 1mM, plus 0.65mM dTTP, and 0.35mM digoxigenin labelled dUTP.

2.9.18 Preparation of radiolabelled DNA probes.

DNA to be radiolabelled was denatured by boiling for 10 min. To the DNA was then added 3µl of a solution of dGTP, dCTP and dTTP each at 0.5mM, 2µl of random hexanucleotide mix (Boehringer Mannheim), 2µl(2µCi) of [32P]dATP and 1µl(2units) of Klenow enzyme. The reaction volume was made up to 20µl with sterile distilled water and the reaction incubated at 37°C for 2-6 hr.

2.9.19 Hybridization of digoxigenin-labelled DNA probes and detection of bound probe DNA.

Hybridizations were carried out as described in the protocol supplied with the DNA labelling kit.

Filters were prehybridized by incubating with hybridization solution (5x SSC, 5% (w/v) "blocking reagent", 0.1% (w/v) N-lauroylsarcosine, sodium salt, 0.02% (w/v) SDS, 50% (v/v) formamide) for 1 hr at 42°C. This solution was then removed and replaced with 2-4ml of fresh hybridization solution containing the freshly denatured probe DNA. Hybridization was at 42°C for 16-20 hr. Filters were then washed twice for 15 min at room temperature with 2x SSC, 0.1% (w/v) SDS, followed by two further washes for 15 min with 0.1x SSC, 0.1% (w/v) SDS at 68°C.

Immunological detection of bound probe DNA was carried out according to the manufacturers' instructions with minor

modifications, as follows. Filters were washed in TBS (1 min), then incubated in TBS containing 0.5% (w/v) "blocking reagent" for 30 min. The filters were then washed again before being incubated in antibody-conjugate diluted 1:5000 (v:v) in TBS, for 1 hr. Unbound antibody-conjugate was removed by washing twice for 15 min in TBS. The filters were then equilibrated in 100mM Tris-HCl, pH 9.5, 100mM NaCl, 50mM $MgCl_2$, before adding the colour solution and leaving in the dark until the colour had developed to the required intensity. (Colour solution was prepared by adding 45µl "NBT solution" (75mg ml^{-1} nitroblue tetrazolium in 70% (v/v) dimethylformamide) and 33µl "X-phosphate solution" (50mg ml^{-1} 5-bromo-4-chloro-3-indolyl phosphate (ϱ-toluidine salt) in dimethylformamide) to 10 ml of 100mM Tris-HCl, pH 9.5, 100mM NaCl, 5mM $MgCl_2$). The reaction was stopped by washing in TE, then the filters were air dried.

2.9.20 Hybridization of radiolabelled DNA probes and detection of bound probe DNA.

Southern blots were prehybridized for 1hr at 42°C in a solution of 5x Denhardt's solution (0.1% (w/v) of each of ficoll, polyvinylpyrrolidone and BSA), 2x SSC, 0.1mg ml^{-1} freshly denatured salmon sperm DNA, 0.5% (w/v) SDS, and with or without 50% formamide. Before hybridization the probe was denatured by boiling for 10 min then the probe was added to 2-4ml of the prehybridization solution and this was added to the Southern blot and incubated at 42°C for 16-20 hr. After hybridization, filters were washed twice in 2x SSC, 0.5% SDS, and with or

without 50% formamide, for 30 min at 42°C. This was then repeated, but this time without formamide. Finally, filters were washed in 2x SSC for 30 min at 42°C. Hybridized probe was detected by exposure of the filter to medical X-ray film for at least 16 hr at -70°C, followed by developing and fixing of the film in CDL8 developer and CF40 fixer (Photosol).

2.9.21 Di-deoxy chain termination sequencing of DNA.

All sequencing reactions were carried out using the commercially available T7 sequencing kit (Pharmacia). The kit provided all components required to carry out DNA sequencing reactions by the di-deoxy chain termination method (Sanger *et al* 1977) except [^{35}S] dATP, which was obtained from Amersham. Preparation of double stranded plasmid template DNA was carried out using the methods described in sections 2.9.4 and 2.9.5, all subsequent steps were carried out according to the manufacturer's instructions. Approximately 1μg of plasmid DNA was denatured by adding NaOH to a final concentration of 0.4M and incubating at room temperature for 10 min. The DNA was then precipitated by the addition of 0.7 volumes of water, 0.3 volumes of 3M CH_3COONa and 6 volumes of ethanol. Precipated DNA was washed with 70% (v/v) ethanol and redissolved in 10 μl of distilled water. 2μl of annealing buffer (a buffered solution containing $MgCl_2$ and DTT), and 2μl of primer (1.6mM) were added to the DNA, mixed and incubated at 37°C for 20 min to allow primer to anneal to the template DNA. The annealed template and primer were left at room temperature for at least 10 min before proceeding to the

labelling and termination reactions. An "enzyme premix" was prepared by mixing 1µl of distilled water, 3µl of "labelling mix A" (dCTP, dGTP and dTTP in solution), 1µl (1µCi) of [^{35}S] dATP, and 2µl of T7 polymerase (diluted to 1.5units μl^{-1} in "enzyme dilution" buffer (a buffered solution containing BSA and DTT)) for each template to be sequenced. 6µl of enzyme premix was added to the annealed template and primer and incubated at room temperature for 5 min during which time newly synthesized DNA was labelled by incorporation of radiolabelled dATP. Chain termination was effected by adding 4µl from this reaction to each of four tubes containing 2.5µl of G, A, T and C mix respectively (each mix contained dGTP, dATP, dTTP and dCTP, additionally the G mix contained ddGTP, the A mix ddATP, the T mix ddTTP and the C mix ddCTP) and incubating at 37^{o}C for 5 min. 5µl of "stop solution" (deionised formamide solution containing EDTA, xylene cyanol and bromophenol blue) was added to each tube and the tubes stored at -20^{o}C.

6% acrylamide gel solution was prepared by mixing 210g urea, 72.5ml 40% acrylamide stock solution (38% (w/v) acrylamide, 2% (w/v) bis-acrylamide) and 50ml 10x TBE (1M Tris, 0,865M boric acid, 20mM EDTA) and bringing the volume to 500ml with distilled water. Sequencing gels were prepared by adding 1ml of 10% (w/v) ammonium persulphate and 50µl TEMED to 100ml of 6% gel solution immediately before pouring the gel mix into the assembled glass plates. Gels were electrophoresed using the S2 (BRL) or STS-45 (IBI) apparatus, at 50W or 70W constant power setting, respectively. The gels were pre-run for 30 min before loading

freshly denatured samples (heated to 80°C for 10 min). Further samples were loaded when the bromophenol blue in the previous samples had migrated off the end of the gel, and gels were run for a maximum of 7 hr.

After electrophoresis, gels were fixed by soaking for 15 min in 10% (v/v) methanol, 10% (v/v) ethanoic acid. Gels were then transferred to Whatman 3MM filter paper and dried on a gel drier.

Detection of [^{35}S]-labelled nucleic acids in sequencing gels was carried out by exposure of gels to medical X-ray film (Fuji) in autoradiography cassettes. Dried gels were placed in the cassette in direct contact with the film, and left overnight at room temperature. X-ray film was developed and fixed as for Southern blot autoradiographs (section 2.9.20).

2.10 Protein and immunological techniques.

2.10.1 Immunoscreening of λgt11 recombinants.

Y1090 plating cells were infected with λgt11 recombinants and plated at the required density as described in section 2.8.2. Plates were incubated at 43°C until plaques were just visible (3-3.5 hr). Expression of recombinant peptides was then induced by overlaying each plate with an 82mm nitrocellulose filter (Schleicher and Schuell) which had been previously soaked in 10mM IPTG. The plates were then moved to a 37°C incubator and

left for a further 3.5 hr. Filters were removed from the plates and rinsed in TBS and then blocked by incubating overnight in TBS containing 10% (v/v) pig serum and 0.02% (w/v) NaN_3. After blocking, the filters were washed (3x 10 min) in TBST and then incubated with primary antibody for 90 min. Filters were then washed in TBST before being incubated in horseradish peroxidase (HRPO)-conjugated secondary antibody for 90 min. The filters were washed again in TBST before adding freshly prepared substrate solution (0.2mg ml^{-1} diaminobenzidine dissolved in TBS, to which 0.5µl ml^{-1} 30% (w/v) H_2O_2 had been added). When the colour had developed to the required intensity, the substrate was washed off with distilled water and the filters air-dried. Dried filters and original plates were aligned to identify and pick positive plaques.

2.10.2 Preparation of induced cell lysates.

Crude cell lysates for SDS-PAGE were prepared from induced cultures of λgt11 lysogens and pMS plasmid clones using a method based on that described by Carroll and Laughon (1984).

Lysogens: overnight cultures of the lysogens were diluted 1:100 in L-broth containing 50µg ml^{-1} ampicillin and grown for 2 hr at 30°C with shaking. IPTG was added to the cultures to a final concentration of 1mM to induce expression from the *lac* promotor, and the cultures were moved to a 45°C water bath for 15 min. The cultures were then placed at 37°C and grown for a further 1-2 hr before harvesting the cells by centrifuging at 10,000g for 1 min

and lysing in 200µl SDS-PAGE sample buffer per 1.5ml of culture. Samples were stored at -20°C before being fractionated by electrophoresis on a 7.5% (w/v) SDS-polyacrylamide gel.

pMS clones: growth and induction of expression from the *lac* promoter by pMS clones was as described for λgt11 lysogens, except that growth was at 37°C, not 30°C, and cells were not heated to 45°C.

2.10.3 SDS-polyacrylamide gel electrophoresis.

Discontinuous SDS-PAGE was carried out using a method based on that of Laemmli (1970).

7.5 or 10% (w/v) resolving gels were prepared by combining appropriate volumes of 4x resolving gel buffer (1.5M Tris-HCl, pH 8.8), 30% acrylamide stock solution (29.2% (w/v) acrylamide, 0.8% (w/v) bis-acrylamide) and 10% (w/v) SDS (final concentration 0.1%). Ammonium persulphate and TEMED were added to give final concentrations of 0.05% (w/v) and 0.05% (v/v), respectively, immediately before casting the gel. 4% (w/v) stacking gels were prepared in the same way, except that 4x stacking gel buffer (0.5M Tris-HCl, pH 6.8) was substituted for 4x resolving gel buffer, and TEMED was added to a final concentration of 0.1% (v/v). Samples to be analysed by SDS-PAGE were mixed with at least 4 volumes of SDS-PAGE sample buffer (0.125M tris-HCl pH 6.8, 5% (v/v) β-mercaptoethanol, 10% (v/v) glycerol, 2% (w/v) SDS, 0.05% (w/v), 0.05% (w/v) bromophenol

blue, 0.05% (w/v) xylene cyanol), and boiled for 5 min immediately before loading the gel. Biorad mw markers or Sigma prestained mw markers were co-electorphoresed to provide mw standards. Gels were run using Tris-glycine-SDS running buffer (25mM Tris, 192mM glycine, 0.1% (w/v) SDS) in the "mini-protean" gel system (Biorad) at a constant voltage setting of 200 volts.

Gels were stained by gentle agitation in 30% (v/v) methanol, 10% (v/v) ethanoic acid containing 0.1% (w/v) Coomassie brilliant blue R150. Gels were destained in several changes of 10% (v/v) methanol, 10% (v/v) ethanoic acid, until background staining was reduced to the desired intensity.

2.10.4 Western blotting.

Proteins were transferred from polyacrylamide gels to "immobilon" PVDF (polyvinylidene difluoride) transfer membrane (Millipore) using a method based on that of Towbin *et al* (1979).

Gels to be blotted were first soaked in blot transfer buffer (25mM Tris, 192mm glycine) for 15 min to remove excess SDS. A sandwich was then assembled with the following successive layers: a porous fibre pad, two sheets of Whatman 3MM filter paper, the polyacrylamide gel, the transfer membrane, two sheets of Whatman 3MM filter paper and finally another porous pad. All components were prewetted with transfer buffer. The sandwich was then placed in the "mini-protean trans-blot cell" (Biorad) and

the tank filled with transfer buffer cooled to 4°C. Transfer was carried out at a constant voltage setting of 100 volts for 1 hr. Staining with Ponceau S was carried out to check that the proteins had transferred, and to locate the positions of tracks so that the membrane could be cut into strips for incubation in different antisera. The membrane was placed in a solution of 0.2% (w/v) Ponceau S, 3% (v/v) trichloroacetic acid, 3% (v/v) sulphosalicylic acid for 5 min, after which the membrane was stained a uniform pink. The staining solution was then removed and the membrane rinsed in TBS until the background had faded sufficiently to see the protein bands. The membrane was then cut into the required strips and the remaining Ponceau S removed by washing in several changes of TBST.

2.10.5 Immunodetection of proteins on Western blots.

All steps were carried out at room temperature with mild agitation. The transfer membrane was rinsed in TBS and then blocked in TBS containing 10% (v/v) pig serum and 0.02% (w/v) NaN_3. After blocking, the membrane was washed (3x 10 minutes) in TBST and then incubated with the primary antibody diluted in TBS for 90 min. The membrane was washed again in TBST before being incubated in the secondary antibody for 90 min. Finally, the membrane was washed again in TBST before adding freshly prepared solution of the appropriate substrate. For alkaline phosphatase conjugated antibodies, the substrate solution was prepared by adding 66µl of 5% (w/v) nitroblue tetrazolium in 70% (v/v) dimethyl formamide (DMF), and 33µl of 5% (w/v)

5-bromo-4-chloro-3-indolyl-phosphate (ρ-toluidine salt) in DMF to 10ml of 100mM Tris-HCl, pH 9.5, 100mM NaCl, 5mM $MgCl_2$. For HRPO conjugated antibodies, the substrate solution was prepared by adding 2mg diaminobenzidine and 5µl of 30% (w/v) H_2O_2 to 10ml of TBS.

When the colour had developed to the required intensity the substrate solution was removed and the membrane washed with water (HRPO conjugates), or 20mM EDTA (alkaline phosphatase conjugates) and then air dried.

CHAPTER 3. ISOLATION AND EXPRESSION OF TWO *CRYPTOSPORIDIUM PARVUM* GENE FRAGMENTS.

3.1 Introduction and preliminary considerations.

The aim of this project was to identify and characterize antigen encoding genes of *C. parvum*. The approach taken was to construct an expression library of *C. parvum* DNA in the widely used expression vector λgt11, and to identify antigen-expressing clones by screening with anti-*C. parvum* antisera. This same approach has been used to identify antigen-encoding genes of several other parasites (Clarke *et al* 1987, Donelson *et al* 1988, Vogel *et al* 1988, Iams *et al* 1990). The vector λgt11 is derived from bacteriophage λ (Young & Davis 1983a). Foreign DNA is inserted into a unique *Eco*RI site situated 53bp upstream from the *lac Z* termination codon. Peptides encoded by insert DNA may be expressed as fusion proteins with β-galactosidase. Expression of recombinant proteins, which may be deleterious to the host cell, is controlled by growth in *E. coli* strains (Y1090, Y1089) which express large amounts of the *lac* repressor. When required, the repressor is inactivated with IPTG, allowing expression of the recombinant protein. Another important feature of *E. coli* strains Y1090 and Y1089 is that they are deficient in the *lon* protease. This reduces the likelihood of breakdown of foreign proteins. Expression libraries in λgt11 can be prepared from genomic DNA or from complementary DNA (cDNA), which is obtained by reverse transcription of messenger RNA (mRNA). There are

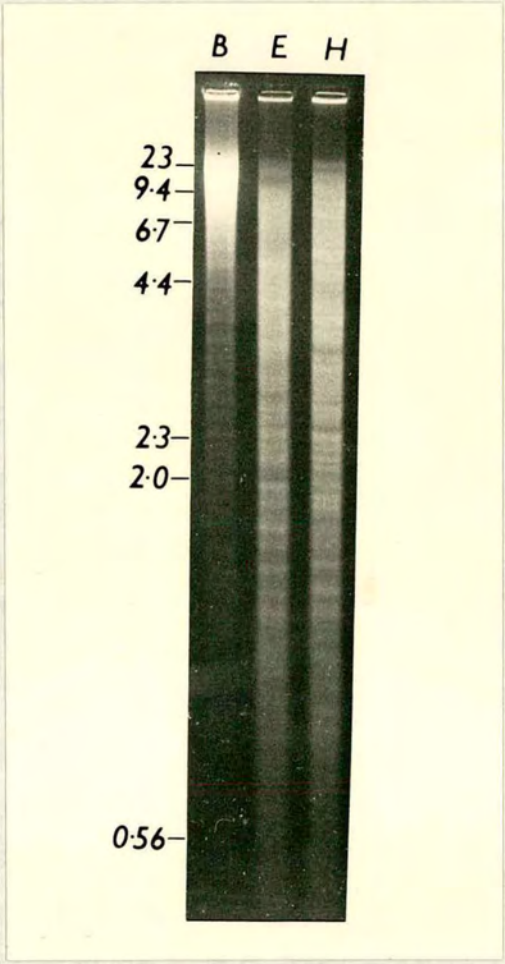
advantages and drawbacks associated with each of these approaches. For example, an expression library prepared from genomic DNA will contain non-coding, as well as coding DNA. This may be a problem when working with large genomes which contain a high proportion of non-coding DNA, such as higher eukaryotes. More clones may have to be screened in order to find the clone of interest. Another disadvantage is the possibility of coding sequences being interrupted by introns, and therefore not being expressed. Complementary DNA libraries have the advantage of only containing coding sequences. However, they only contain the sequences of those genes which are being transcribed at the time of mRNA extraction. Apicomplexans have complex lifecycles and many of their antigens are stage-specific (Kemp *et al* 1987). Therefore a cDNA library prepared from a particular lifecycle stage will only express a proportion of the antigen-encoding genes present in the genome. The only stage of *C. parvum* which could be prepared in sufficient quantities for nucleic acid extraction was the sporulated oocyst stage. Sporozoites could then be obtained by *in vitro* excystation of the oocysts. It was therefore decided to construct an expression library from *C. parvum* genomic DNA. The genome of *C. parvum* is relatively small (see section 1.10), and published gene sequences of other Protozoa suggest that introns are not as common in this group as in multicellular organisms. Therefore the drawbacks of genomic DNA libraries discussed above are not likely to be a problem in the case of *C. parvum*. Ideally, fragments of genomic DNA used to make an expression library should be generated in a completely random manner, such as by mechanical shearing, or by

sonication. Another method that is often used, though is not strictly random, is partial digestion using a restriction enzyme with a 4bp recognition site. These methods attempt to ensure that every gene has a chance of being cloned, and that all six possible reading frames are likely to be represented. However, each of these methods requires relatively large amounts of DNA, and since *C. parvum* DNA was in short supply, an alternative method of generating fragments for cloning into λgt11 was used. Restriction digests of *C. parvum* DNA revealed that *Eco*RI digestion resulted in a distribution of fragments, the majority of which fall in the size range 0-9kb (Fig 3.1). The λgt11 expression vector can accept inserts of up to 7.2kb (Huynh *et al* 1985), though clones with inserts as large as 11kb have been identified (K. Stevenson, Moredun Research Institute, pers. comm.). Use of *Eco*RI-digested *C. parvum* genomic DNA would therefore result in the possibility of the majority of the genome being cloned. Another advantage of using *Eco*RI-digested DNA is that further manipulation of the DNA is unnecessary in order to clone the fragments into the *Eco*RI site of λgt11. It was therefore decided to prepare a λgt11 expression library from *Eco*RI-digested *C. parvum* genomic DNA.

Following identification of antigen-expressing clones by immunoscreening, the DNA inserts and the recombinant proteins expressed by these clones were characterized by gel electrophoresis and Western blotting.

Fig 3.1 Restriction digests of *C. parvum* genomic DNA.

Genomic DNA from *C. parvum* was digested with *Bam*HI (B), *Eco*RI (E) and *Hin*dIII (H) and electrophoresed through a 0.8% agarose gel.



3.2 Construction of λgt11 expression library.

$5x10^9$ *C. parvum* oocysts were first excysted *in vitro*. Subsequently, DNA was extracted from the sporozoites by incubation with proteinase K and SDS, followed by further purification by phenol extraction and ethanol precipitation (section 2.9.1). The amount of DNA obtained was estimated at approximately 40μg by examination of electrophoresed DNA stained with ethidium bromide. *C. parvum* DNA was digested with *Eco*RI and electrophoresed through a 0.8% agarose gel. DNA fragments up to approximately 8kb in size were then excised and recovered from the gel using the "geneclean" procedure. Aliquots of this DNA were ligated with *Eco*RI digested, dephosphorylated λgt11 arms (Amersham) in two separate ligation reactions (1 and 2). These reactions each contained 1μg of λgt11 DNA, and were carried out in a final volume of 10μl. In addition, reaction 1 contained 40ng of *C. parvum* DNA, and reaction 2, 80ng. After ligation, each reaction was packaged *in vitro*, then serial dilutions of the packaged phage were plated with *E. coli* Y1090 in order to calculate the recombinant titre. The top agar was supplemented with colour selection reagents (1mM IPTG, and 0.02% X-gal) in order to distinguish recombinant (colourless) from non-recombinant (blue) plaques. The results of the titration are shown in table 3.1.

From the figures in table 3.1 it was calculated that packaged DNA from ligation 1 contained approximately 1.4×10^4 recombinant clones, and ligation 2 contained approximately 2.4×10^5 recombinants.

Table 3.1 Bacteriophage titration.

Plate	Ligation	Dilution factor	No. of plaques blue	colourless
1	1 (40ng DNA)	10^3	6	43
2	"	10^4	3	2
3	"	10^5	0	1
4	"	10^3	3	39
5	"	10^4	1	1
6	"	10^5	0	0
7	2 (80 ng DNA)	10^3	28	325
8	"	10^4	9	34
9	"	10^5	0	2
10	"	10^3	50	550
11	"	10^4	3	74
12	"	10^5	0	7

The volume of each ligation reaction was made up to 0.5ml with SM before preparing the serial dilutions. Two sets of serial dilutions were made per ligation, and 5µl from one of the dilutions was used per plate.

3.3 Identification of immunopositive λgt11 clones.

The antisera used to identify antigen-expressing clones were obtained from two sources. Anti-*C. parvum* antiserum was raised in a litter of rats by oral infection with oocysts, followed by parenteral inoculation with excysted oocysts (Table 2.4). Antiserum raised in this way recognises a wide range of oocyst and sporozoite antigens (Hill 1989). This was an important consideration, since, due to limitations imposed by a short supply of *C. parvum* DNA, it was almost certain that not all antigen-encoding genes of *C. parvum* would be represented in the library (section 3.1). In addition to the rat antiserum, sera from gnotobiotic lambs which were experimentally infected with *C. parvum* were also available. Serum from one of these lambs (151, table 2.5) was also used to screen recombinants in the hope that one or more antigens that were recognised by the host during infection could be identified.

Phage from ligation 2 were plated at a density of approximately 5×10^3 plaques per 80mm plate, and plaques were screened with rat anti-*C. parvum* antiserum. Three immunopositive clones were identified, denoted λCPR1, λCPR4.1 and λCPR4.2 The plaques were picked, replated and confirmed by rescreening, and this process repeated until each clone was purified. After screening, the amplified library was eluted in SM and stored over chloroform. The new titre was found to be approximately 2.5×10^8 pfu/ml

with 90% recombinants. Approximately 10^5 plaques from this amplified library were screened with serum from a gnotobiotic lamb experimentally infected with *C. parvum* (lamb 151 table 2.5). One immunopositive clone (λCPS10) was identified using this serum. The positive plaque was picked, rescreened, and the clone purified as before.

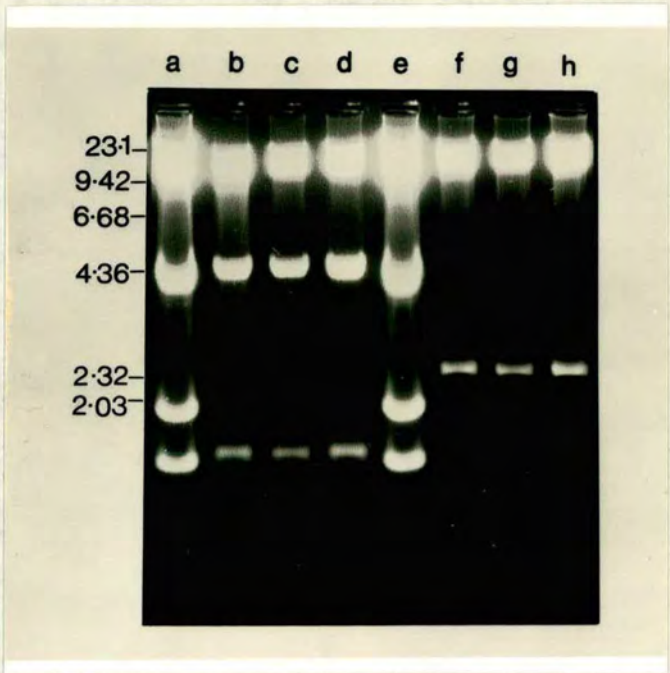
3.4 Characterization of immunopositive λgt11 clones.

3.4.1 Demonstration of inserts in immunopositive λgt11 clones.

Recombinant phage DNA was prepared from each of clones λCPR1, λCPR4.1, λCPR4.2 and λCPS10. The DNA was digested i) with *Eco*RI, and ii) simultaneously with *Kpn*I and *Sac*I (Figs 3.2 and 3.3). *Eco*RI digestion of DNA from each of clones λCPR1, λCPR4.1 and λCPR4.2 released a fragment of approximately 2.4kb. Digestion with *Kpn*I and *Sac*I generated an identical banding pattern to that obtained by digestion of λgt11, except that the 2.08kb band was missing. This was explained by the presence of the 2.4kb insert in the *Eco*RI site within this 2.08kb fragment, which resulted in the enlarged fragment (now approximately 4.48) co-migrating with the 4.43kb band of λgt11 (Fig 3.2). These results therefore indicated that clones λCPR1, λCPR4.1 and λCPR4.2 each contained an insert of approximately 2.4kb (denoted CPR1, CPR4.1 and CPR4.2 respectively).

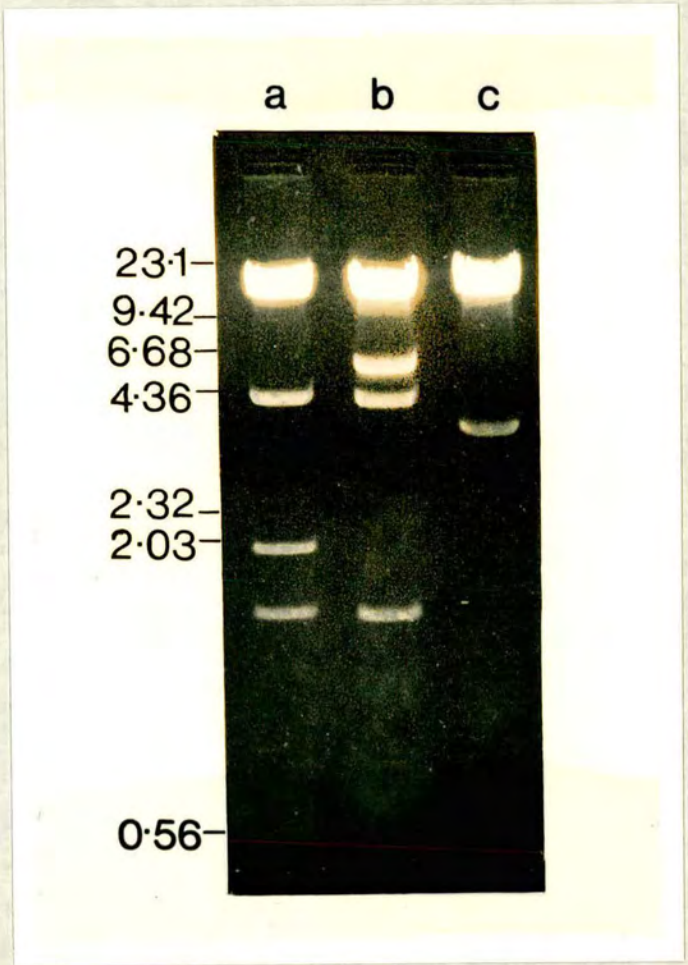
*Eco*RI digestion of DNA from clone λCPS10 released a fragment of

Fig 3.2 Insert analysis of clones λCPR1, λCPR4.1 and λCPR4.2.



Track	DNA	Restriction enzymes
a	λgt11	KpnI/SacI
b	λCPR1	KpnI/SacI
c	λCPR4.1	KpnI/SacI
d	λCPR4.2	KpnI/SacI
e	λgt11	KpnI/SacI
f	λCPR1	EcoRI
g	λCPR4.1	EcoRI
h	λCPR4.2	EcoRI

Fig 3.3 Insert analysis of clone λCPS10.



Track	DNA	Restriction enzymes
a	λgt11	KpnI/SacI
b	λCPS10	KpnI/SacI
c	λCPS10	EcoRI

approximately 3.5kb. *Kpn*I/*Sac*I digestion resulted in the generation of a fragment of approximately 5.6kb in place of the 2.08kb fragment produced by *Kpn*I/*Sac*I digestion of λgt11 DNA (Fig 3.3). These results therefore indicated that λCPS10 contained an insert of approximately 3.5kb (CPS10).

3.4.2 Identification of recombinant peptides expressed by λgt11 lysogens.

Clones λCPR1, λCPR4.1, λCPR4.2, λCPS10 and a non-recombinant λgt11 clone were lysogenized in *E. coli* Y1089. Crude cell lysates were then prepared from induced cultures of each of these lysogens, and aliquots of the lysates analysed by Western blotting (Figs 3.4 and 3.5). Lysogens λCPR1, λCPR4.1 and λCPR4.2 each expressed peptides which appeared on Western blots as a ladder-like series of bands with mws ranging from 116kDa to approximately 180kDa. These bands were recognised both by rat anti-*C. parvum* antiserum, and by mouse anti-β-galactosidase serum (Fig 3.4). Since the β-galactosidase fusion protein occurred as a series of bands, and not as a single band, it was concluded that these probably represented breakdown products of the full length fusion protein whose mw was unknown. If the entire 2.4kb insert consisted of a continuous open reading frame which was being fully expressed, then the mw of the full length protein would be estimated to be around 200kDa. This estimate allows 116kDa for the β-galactosidase portion of the protein, and assumes that each amino acid has a mw of 0.1kDa, and that no

Fig 3.4 Western blot analysis of lysogens λCPR1, λCPR4.1 and λCPR4.2.

Induced lysates prepared from lysogens λCPR1 (1), λCPR4.1 (4.1) and λCPR4.2 (4.2) were electrophoresed through a 7.5% polyacrylamide gel and Western blotted. Blots were probed with rat anti-*C. parvum* antiserum (b) or mouse anti-β-galactosidase serum (a). Lysates prepared from a λgt11 lysogen (gt11) and *E. coli* Y1089 (Y1089) were also Western blotted.

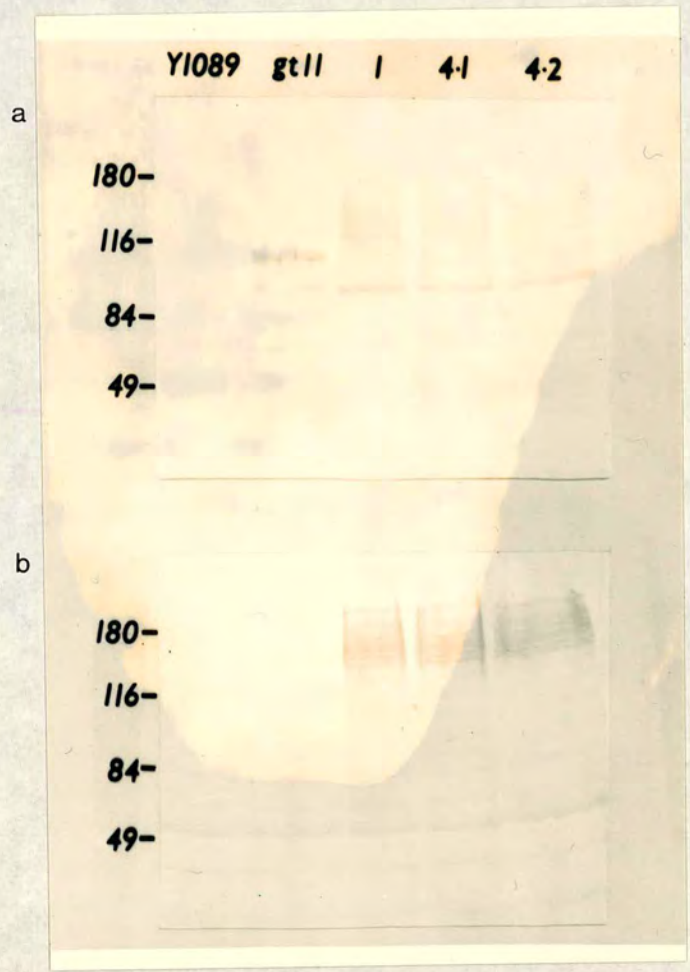
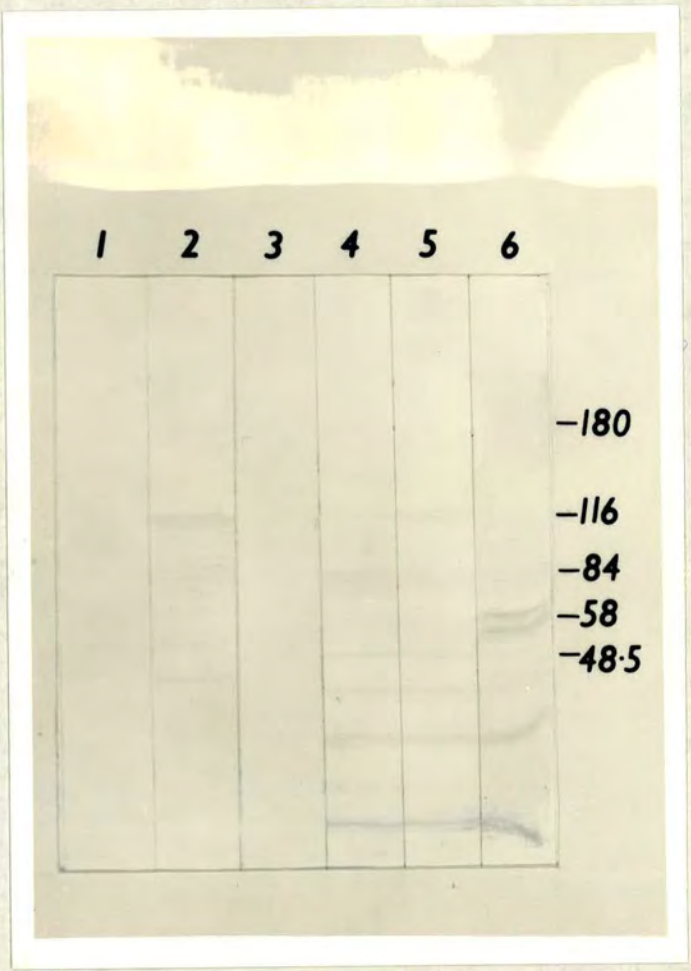


Fig 3.5 Western blot analysis of lysogen λCPS10.



Track	Lysate	Antiserum
1	Y1089	Mouse anti-β-galactosidase
2	λgt11 lysogen	"
3	λCPS10 lysogen	"
4	Y1089	*C. parvum* infected lamb (151)
5	λgt11 lysogen	"
6	λCPS10 lysogen	"

post translational modification occurs.

Lysogen λCPS10 was grown for only 1 hr following induction, instead of the routine 2 hr induction carried out for all other lysogens. This was because induction of this clone for longer than 1 hr resulted in lysis of the cells. This indicated that the recombinant peptide expressed by this clone had a deleterious effect on the *E. coli* host cells. Western blotting of cell lysates from lysogen λCPS10 revealed two bands with mws around 55kDa that were recognised by the gnotobiotic lamb serum, but not by mouse anti-β-galactosidase serum (Fig 3.5). There are three simple explanations as to why peptides encoded by inserts in λgt11 could occur as free proteins instead of as fusions with β-galactosidase. Firstly, if an insert contains a gene, or gene fragment, together with its own promoter, and if the promoter is capable of forming an initiation complex with *E. coli* RNA polymerase, then the foreign gene may be transcribed independently of the *lac Z* gene in λgt11. Secondly, the *lac Z* gene of λgt11 and the insert may be transcribed as a single unit, but if the ORF encoded on the insert DNA begins downstream of the 5′ end of the insert, or is out of frame with the β-galactosidase ORF, then β-galactosidase and the recombinant protein may be translated separately, but from the same mRNA. Thirdly, the recombinant peptide may be expressed as a β-galactosidase fusion protein, but the foreign protein may be cleaved off after translation. However, in all of these cases one would have expected β-galactosidase to be expressed as well as the foreign protein, and this does not appear to have been

the case following induction of the λCPS10 lysogen. It is possible that a mutation may have occurred in the *lac Z* gene of this clone.

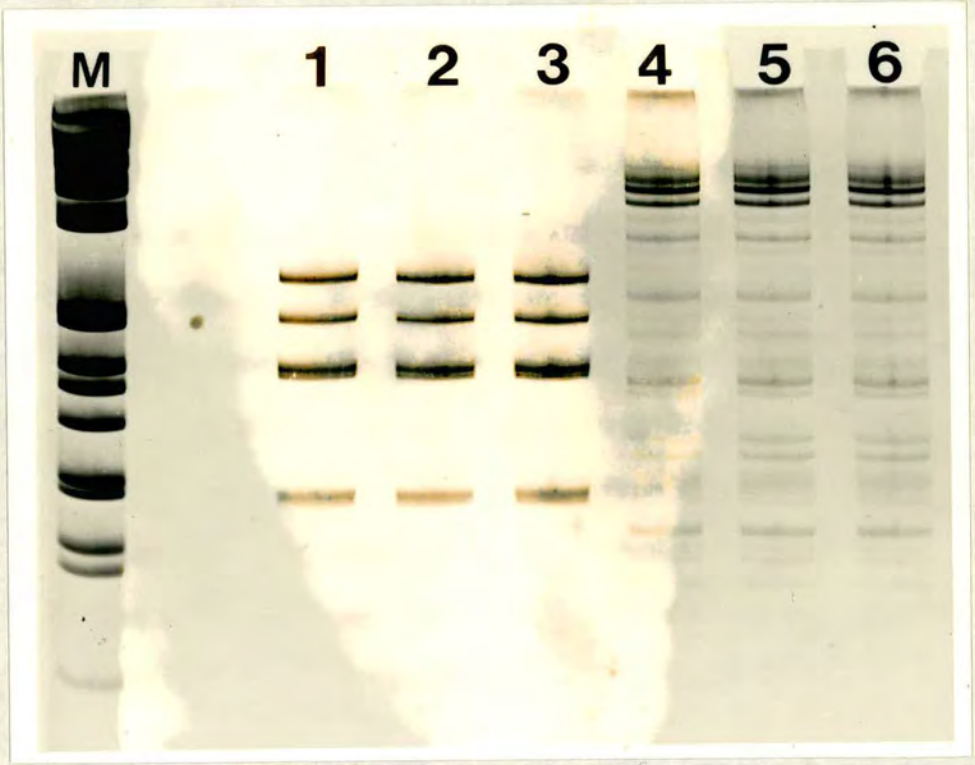
3.4.3 Demonstration that clones λCPR1, λCPR4.1 and λCPR4.2 all contain identical inserts.

Since clones λCPR1, λCPR4.1 and λCPR4.2 all contained inserts of the same size and, in addition, all expressed β-galactosidase fusion proteins with the same mws, it seemed very likely that they all contained the same insert. In order to find out whether this was the case, the insert from each of these clones was isolated by *Eco*RI digestion, electrophoresis, excision from the gel and "genecleaning". Each isolated fragment was then digested separately with *Rsa*I and *Alu*I. The digestion products were electrophoresed through a polyacrylamide gel which was then silver stained. The inserts from all three clones gave identical banding patterns for each restriction enzyme (Fig 3.6), confirming that clones λCPR1, λCPR4.1 and λCPR4.2 did indeed contain identical inserts. From this point onwards, therefore, further work was carried out only on clones λCPR1 and λCPS10.

3.5 Characterization of the recombinant peptides expressed by plasmid subclones.

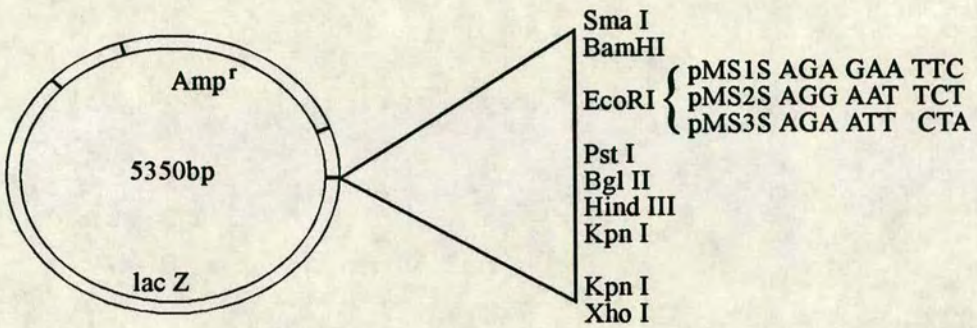
The next step in characterization of these antigen-expressing clones was to subclone their inserts into the pMS plasmid

Fig 3.6 *Alu*I and *Rsa*I restriction profiles of CPR1, CPR4.1 and CPR4.2.



Track	DNA	Restriction enzyme
1	CPR1	AluI
2	CPR4.1	AluI
3	CPR4.2	AluI
4	CPR1	RsaI
5	CPR4.1	RsaI
6	CPR4.2	RsaI

Fig 3.7 pMS plasmid cloning vectors.



Multiple cloning site.

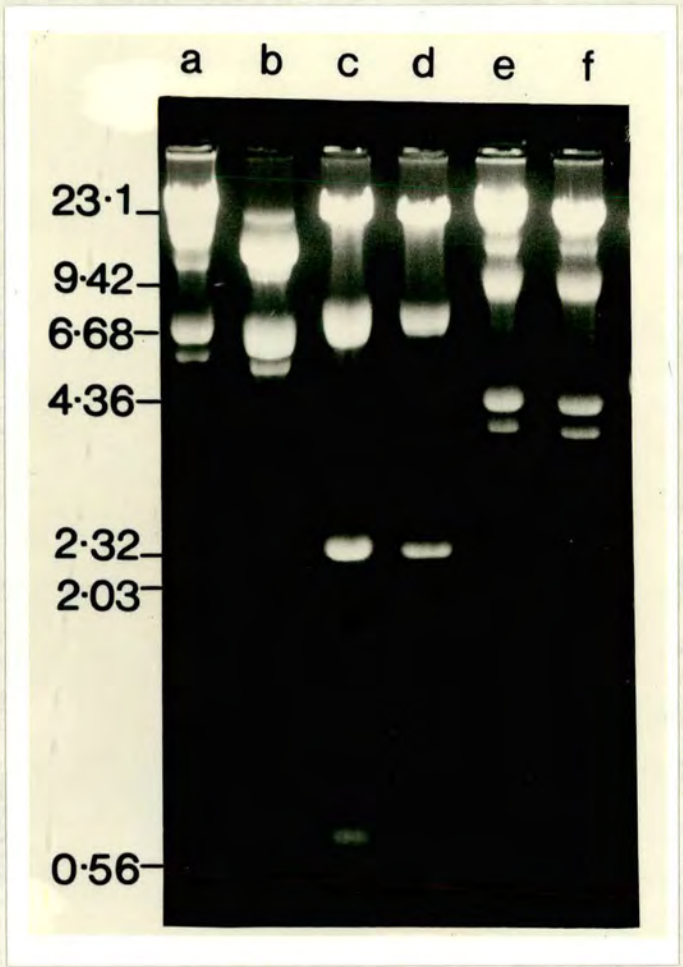
```
5'          Xho I    KpnI                                    KpnI
  CAA AAA CTC GAG GTA CCG ATC GAA GGA CGT ACG GTA CCA
                        Ile  Glu  Gly  Arg
                        Factor Xa site

Hin III      Bgl II  Pst I      pMS1S A GAA TTC    BamHI
 AGC TTA CAG ATC TGC AG         pMS2S G AAT TCT    GGA TCC
                                pMS3S A ATT CTA

          stop region                Sma I                    3'
 AGT AAT TAA TAG ATA ATA GCC CGG GCT TCC AAA TCG AGT
```

expression vectors (Fig 3.7). These are a set of three plasmids (pMS1S, pMS2S and pMS3S) which are derived from pMSgt11 (Scherf *et al* 1990) and are identical except for a small region in their multiple cloning sites. Each has a unique *Eco*RI site, situated approximately 50bp downstream from the 3′ end of the *lac Z* gene, that is in a different reading frame in each of the three plasmids. The *Eco*RI site in pMS1S is in frame with that in λgt11. There were three main reasons for using the pMS plasmids for sub-cloning. Firstly, foreign proteins are often more stable in *E. coli* when fused with the complete β-galactosidase protein, than when fused with the shortened molecule lacking the 18 carboxy-terminal residues as occurs in λgt11. It was hoped that cloning the CPR1 insert in pMS would allow the full length protein to be identified. Secondly, it is often possible to obtain larger quantities of recombinant fusion proteins when expressed from plasmid vectors than from λgt11, since expression can be carried on for a longer period of time than with lysogens. This may be important at a later stage of the work since larger quantities of recombinant protein may be required to raise antisera. Thirdly, cloning CPS10 into each of the three reading frames could have facilitated expression of its encoded peptide as a β-galactosidase fusion.

3.5.1 Sub-cloning CPR1 and CPS10 in pMS.

Before sub-cloning CPR1 and CPS10 it was desirable to identify one or more restriction sites within the inserts to allow their

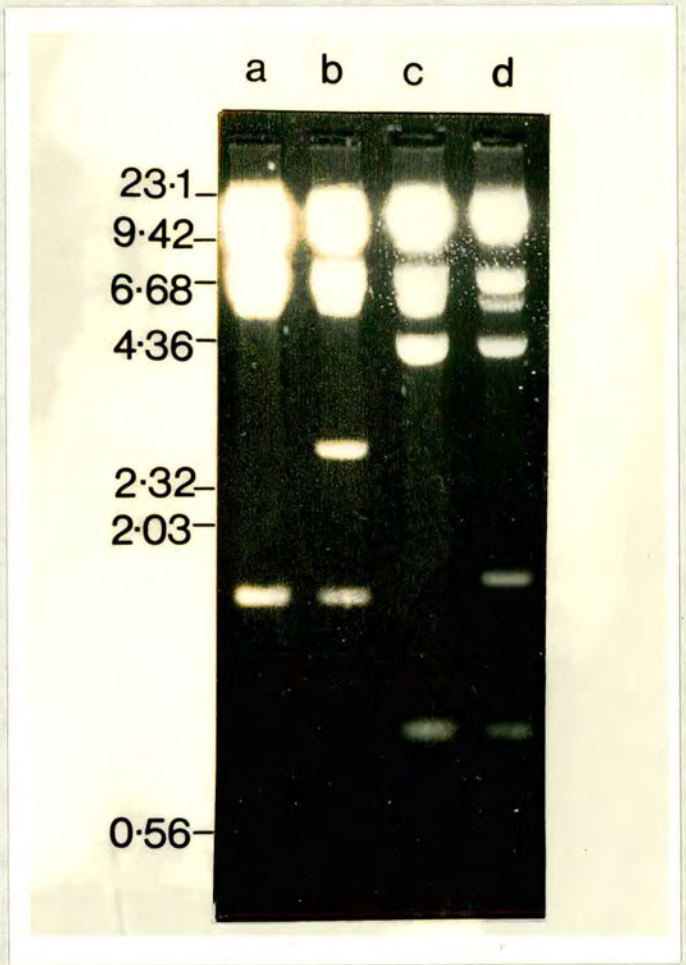
Fig 3.8a Restriction analysis of CPR1.



Track	DNA	Restriction enzymes
a	λgt11	BamHI
b	λCPR1	BamHI
c	λgt11	BglII
d	λCPR1	BglII
e	λgt11	HindIII
f	λCPR1	HindIII

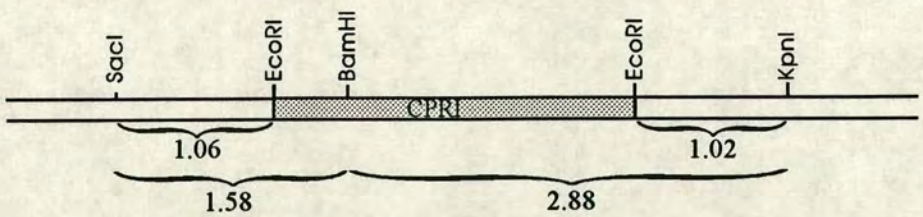
The *Eco*RI site in λgt11 is situated on a 20.44kb *Bam*HI fragment. Cleavage at this site would produce fragments of 14.07 and 6.37kb. The banding pattern resulting from *Bam*HI restriction of λCPR1 DNA (track b) suggested that the CPR1 contained a *Bam*HI site. There was no evidence of sites for *Bgl*II or *Hin*dIII in CPR1 (tracks d and f).

Fig 3.8b Restriction analysis of CPR1.



Track	DNA	Restriction enzymes	Size of 'extra' band (kb)
a	λgt11	BamHI/KpnI	-
b	λCPR1	BamHI/KpnI	2.88
c	λgt11	BamHI/SacI	-
d	λCPR1	BamHI/SacI	1.58

The *Bam*HI/*Kpn*I and *Bam*HI/*Sac*I digests indicated that the *Bam*HI site in CPR1 was situated approximately 1.86kb from its 3' end and 0.52kb from its 5' end respectively. i.e.

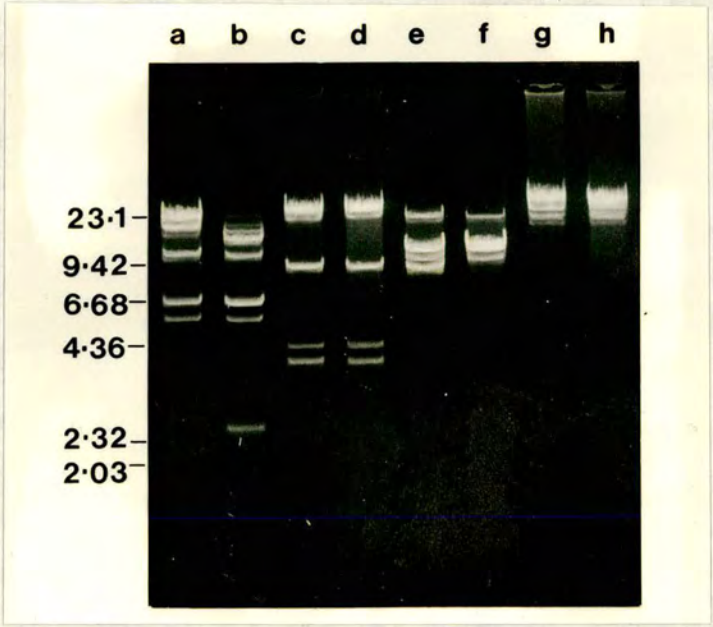


orientation to be determined following sub-cloning. Restriction digest analysis (Fig 3.8a, 3.8b) revealed that CPR1 had a *Bam*HI site approximately 0.5kb downstream from its 5′ end. CPS10 was found to contain three *Bam*HI sites, one approximately 200bp downstream from its 5′ end, and one approximately 100bp upstream from its 3′ end. The third *Bam*HI site cleaved the remainder of the insert between the aforementioned *Bam*HI sites, into two fragments of approximately 2.5 and 0.7kb (Fig 3.9a, 3.9b).

Insert DNA for sub-cloning was prepared by digesting λCPR1 and λCPS10 DNA with *Eco*RI, electrophoresing through a 0.8% agarose gel and recovering the inserts using the "geneclean" procedure.

E. coli clones harbouring plasmids pMS1S, pMS2S and pMS3S were obtained from Professor J. G. Scaife, King's Buildings, Edinburgh University. Large scale plasmid DNA preparations were carried out to obtain sufficient of each vector DNA for subcloning. The vector DNA was digested with *Eco*RI, and dephosphorylated using calf intestinal phosphatase. The CPR1 and CPS10 fragments were ligated with each of the three pMS plasmids. Ligations were carried out in a total volume of 10μl, and following transformation of competent *E. coli* NM522 cells, transformants were identified by growth on L-agar supplemented with 50μg ml^{-1} ampicillin. Plasmid DNA was prepared from these transformants following overnight growth in 10ml cultures. The plasmid DNA was then digested with *Bam*HI in order to identify clones containing inserts in the correct orientation. The CPR1

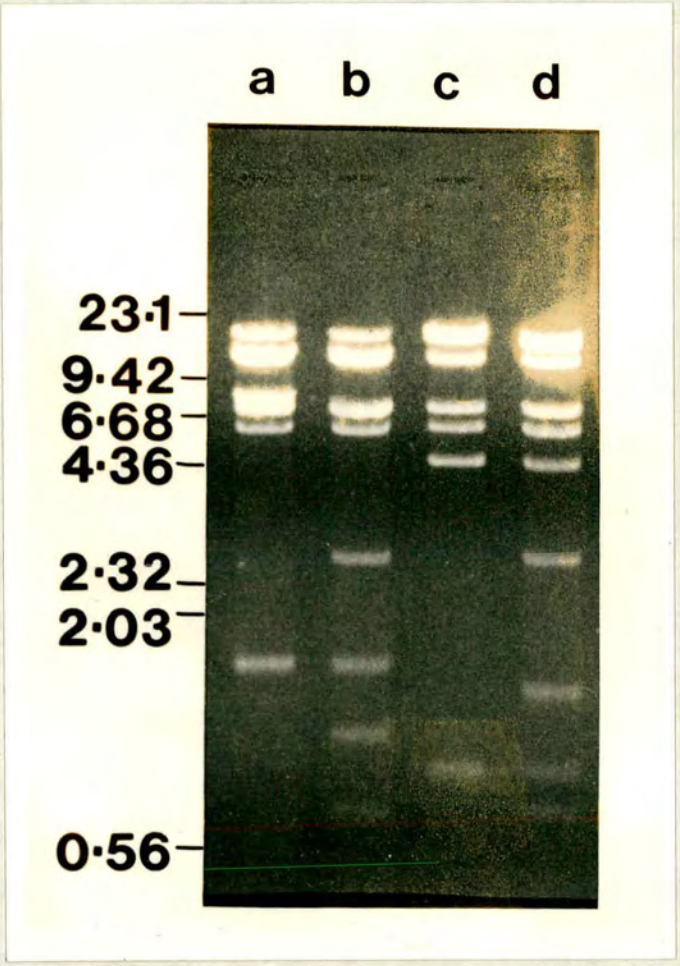
Fig 3.9a Restriction analysis of CPS10.



Track	DNA	Restriction enzymes
a	λgt11	BamHI
b	λCPS10	BamHI
c	λgt11	HindIII
d	λCPS10	HindIII
e	λgt11	PvuI
f	λCPS10	PvuI
g	λgt11	XbaI
h	λCPS10	XbaI

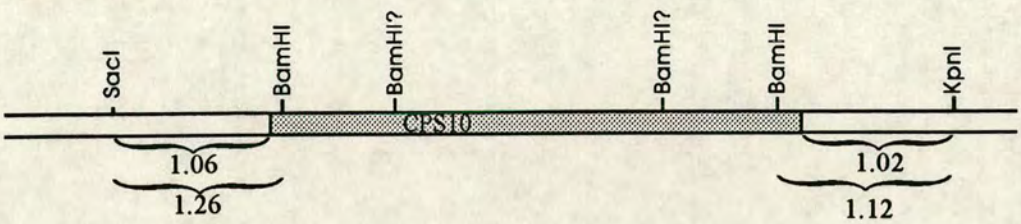
The presence of an extra band of approximately 2.5kb produced by *Bam*HI restriction of λCPS10 DNA (track b) indicates that there are likely to be at least two *Bam*HI sites in CPS10. Further restriction digests were carried out to ascertain the positions of *Bam*HI sites in CPS10 (Fig 3.9b).

Fig 3.9b Restriction analysis of CPS10.



Track	DNA	Restriction enzymes	Sizes of 'extra' bands (kb)
a	λgt11	BamHI/KpnI	-
b	λCPS10	BamHI/KpnI	2.5, 0.7, 1.12
c	λgt11	BamHI/SacI	-
d	λCPS10	BamHI/SacI	2.5, 0.7, 1.26

The *Bam*HI/*Kpn*I and *Bam*HI/*Sac*I digests indicated that there were *Bam*HI sites in CPS10 approximately 100bp from its 3' end and 200bp from its 5' end respectively. An additional *Bam*HI site cuts the remaining 3.3kb into fragments of approximately 2.5kb and 0.7kb. i.e.



insert was successfully cloned into each of pMS1S, pMS2S and pMS3S, and clones containing the insert in the same orientation as that in λgt11 were used for expression and analysis of the fusion protein.

*Bam*HI digests of the sub-clones containing CPS10 showed that part of the insert had been deleted in all of the clones examined. In each case the deletion had occurred in the 2.5kb *Bam*HI fragment reducing it to approximately 1.9kb. This may have been due to recombination between the CPS10 DNA and *E. coli* NM522 DNA, since unlike many *E. coli* strains used for cloning, NM522 is not recombination deficient. Support for this view is given by subsequent cloning work carried out at Moredun Research Institute, by Z. Jurisic (visiting worker from Belgrade). During this work the CPS10 insert was ligated into the plasmid vector Bluescript SK+ and the recombinant plasmid was successfully transformed into *E. coli* JM109, without loss or rearrangement of DNA. Unlike strain NM522, JM109 contains the *RecA1* mutation.

Table 3.2 shows the plasmid sub-clones used for SDS-PAGE Western blot analysis.

3.5.2 Western blot analysis of recombinant peptides expressed by pMS sub-clones.

Western blotting of lysates prepared from induced cultures of

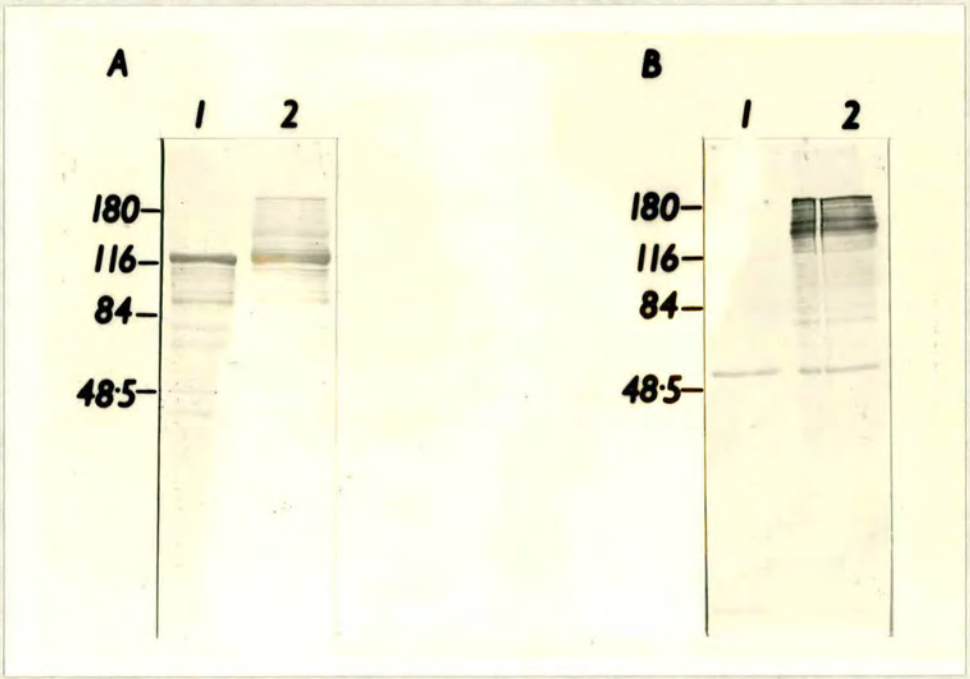
Table 3.2 pMS plasmid subclones.

Plasmid	Vector	Insert	Orientation of insert
pCPR1/1S	pMS1S	CPR1	+
pCPR1/2S	pMS2S	CPR1	+
pCPR1/3S	pMS3S	CPR1	+
pCPS10/1S	pMS1S	$CPS10^{d}$	+
pCPS10/2S	pMS2S	$CPS10^{d}$	+
pCPS10/3S	pMS3S	$CPS10^{d}$	+
$pCPS10/1S^{2}$	pMS1S	$CPS10^{d}$	-

+ same orientation as that in λgt11, with respect to lac Z
- opposite orientation to that in λgt11
$CPS10^{d}$ denotes CPS10 with a deletion in the 2.5kb *Bam*HI fragment as described in section 3.5.1

Fig 3.10 Western blot analysis of pCPR1/1S.

Lysates prepared from induced cultures of pCPR1/1S (2) and a clone harbouring pMS1S (1), were electrophoresed through a 7.5% polyacrylamide gel and Western blotted. Blots were probed with mouse anti-β-galactosidase antiserum (A) or rat anti-*C. parvum* antiserum (B).



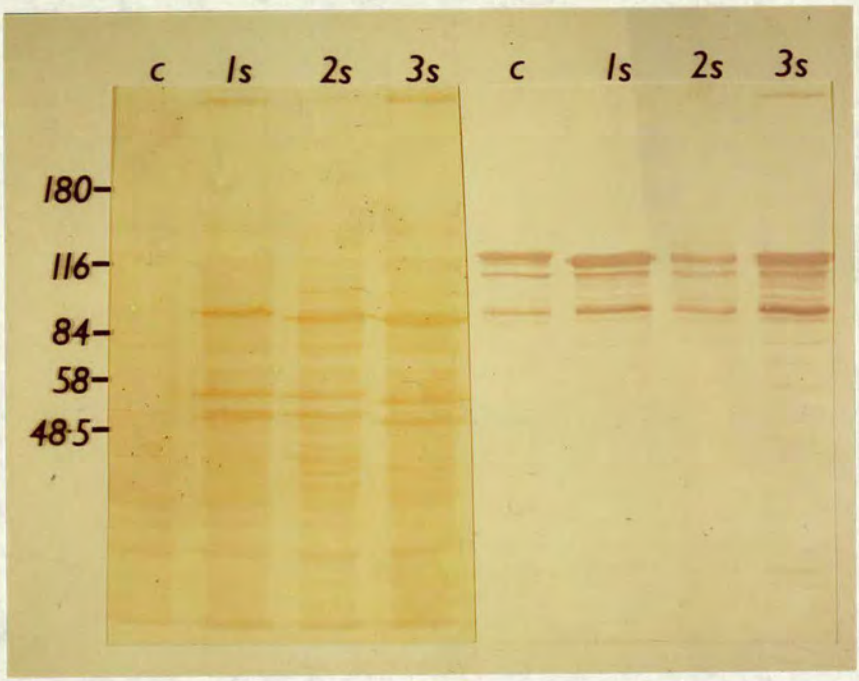
clones pCPR1/1S, pCPR1/2S and pCPR1/3S showed that a β-galactosidase fusion protein recognised by rat anti-*C. parvum* antisera was expressed by clone pCPR1/1S (Fig 3.10). This fusion protein appeared as a series of bands with mws ranging from approximately 116 to 200kDa. Plasmids pCPR1/2S and pCPR1/3S both failed to express the fusion protein, indicating that the CPR1 insert is likely to contain an ORF which is in frame with the *Eco*RI site in both λgt11 and pMS1S. A mw of 200kDa for the full length fusion protein indicates that the entire CPR1 insert is likely to consist of a single ORF.

Peptides recognised by serum from lamb 151 were expressed by each of clones pCPS10/1S, pCPS10/2S and pCPS10/3S. A band at approximately 80-90kDa and two bands between 48.5 and 58kDa were present on Western blots of induced lysates prepared from each of these clones (Fig 3.11). Bands recognised by the anti-β-galactosidase serum with apparent mws of 80-90kDa and 116kDa were also present. However, it seemed unlikely that the 80-90kDa band recognised by the lamb serum was the same band as that recognised by the anti-β-galactosidase serum since a clone harbouring the pMS1S plasmid also produced a peptide of 80-90kDa recognised by the anti-β-galactosidase serum. It therefore appeared that the recombinant peptide encoded by CPS10 existed in the cell as a free protein and not as a β-galactosidase fusion, in each of clones pCPS10/1S, pCPS10/2S and pCPS10/3S.

A lysate prepared from clone pCPS10/1S^2 was Western blotted to find out whether the recombinant peptide recognised by the lamb

Fig 3.11 Western blot analysis of pCPS10/1S, pCPS10/2S and pCPS10/3S.

Lysates prepared from induced cultures of pCPS10/1S, pCPS10/2S and pCPS10/3S (tracks labelled 1s, 2s and 3s respectively), and a clone harbouring pMS1S (c) were electrophoresed through a 7.5% polyacrylamide gel and Western blotted. Blots were probed with serum from a gnotobiotic lamb (151) infected with *C. parvum* (left panel) or mouse anti-β-galactosidase serum (right panel).



serum was still expressed when the insert was in the opposite orientation with respect to the *lac Z* gene. This was not the case. This indicated that the DNA encoding the recombinant peptide was not transcribed from its own promotor since this would not have been affected by its orientation with respect to the *lac Z* gene. The DNA encoding the recombinant peptide seemed likely to be transcribed as a single unit with the *lac Z* gene. This was supported by the observation that uninduced cultures of clones pCPS10/1S, pCPS10/2S and pCPS10/3S appeared to express less β-galactosidase and less recombinant protein than induced cultures grown for the same period of time (Fig 3.12).

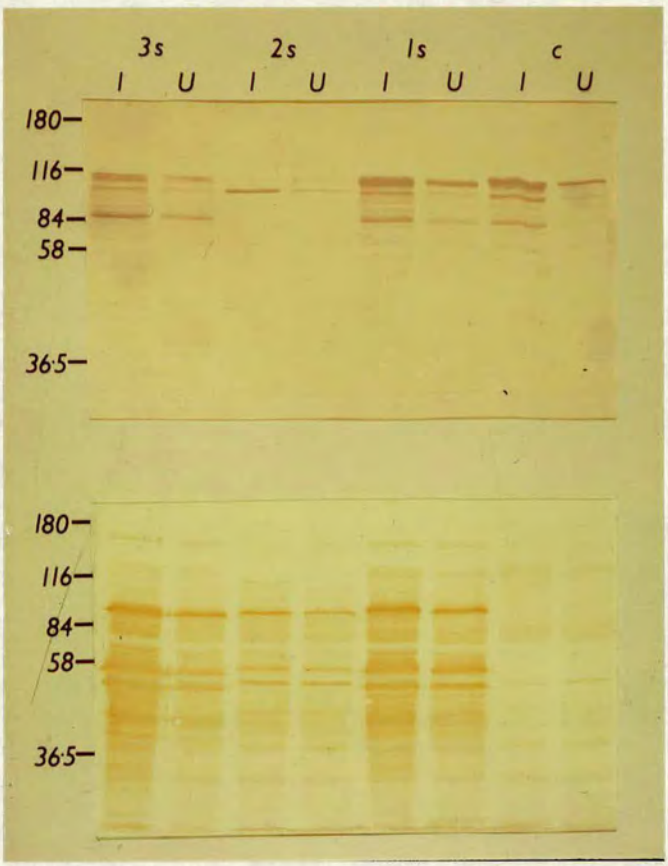
In addition Western blots were carried out to determine whether the recombinant peptide expressed by pCPR1/1S was recognised by the serum from lamb 151, and whether the peptide expressed by pCPS10/1S was recognised by the rat serum. In both cases no cross reaction was demonstrated (not shown).

3.5.3 Recognition of the CPS10 recombinant peptide by a panel of sera from lambs infected with *C. parvum*.

The recognition of *C. parvum* antigens by serum from infected animals appears to be rather variable (section 1.8). In view of this, it was of interest to find out which of a panel of 16 sera from *C. parvum*-infected lambs recognised the CPS10 recombinant protein on Western blots of pCPS10/1S lysates. Six of the sera were taken from gnotobiotic lambs which had each

Fig 3.12 Effect of induction on expression of recombinant peptide by pCPS10/1S, pCPS10/2S and pCPS10/3S.

Lysates prepared from induced (I) and uninduced (U) cultures of pCPS10/1S (1s), pCPS10/2S (2s) and pCPS10/3S (3s) and a clone harbouring pMS1S (c), were electrophoresed through a 7.5% polyacrylamide gel and Western blotted. Blots were probed with mouse anti-β-galactosidase antiserum (top) or serum from a gnotobiotic lamb (151) infected with *C. parvum* (bottom).



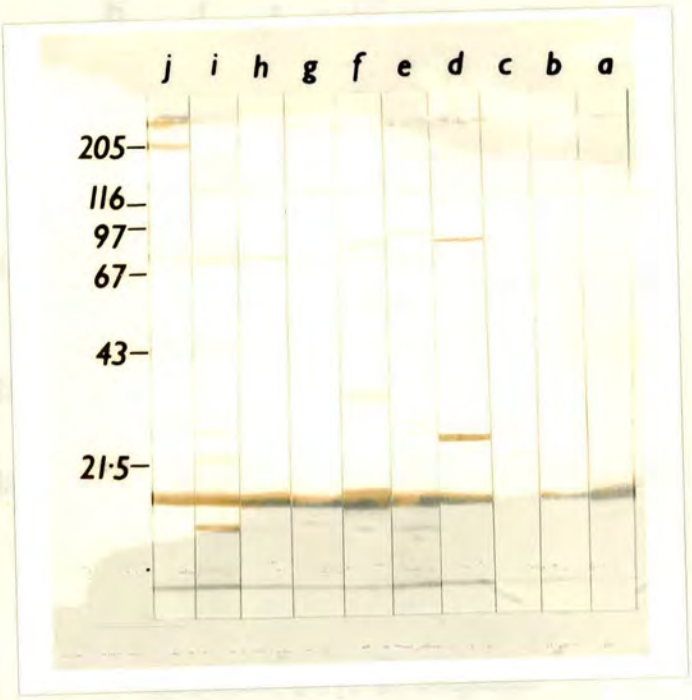
received a single oral dose of 10^6 oocysts at 10 days of age. Serum samples were taken on day 16 post infection, with the exception of that from lamb 96 which was taken on day 11 post infection, when the animal died. The remaining 10 sera were from gnotobiotic lambs, each of which received a trickle infection of approximately 50 oocysts per day for a period of 22 days. Each of these lambs became infected and exhibited a similar oocyst shedding pattern to that observed for donor lambs (Hill 1989), which are routinely infected with 10^6 purified oocysts, except that the duration of the prepatent period increased from 3 days to 5-7 days in these 10 lambs (S. E. Wright, Moredun Research Institute, unpublished).

Figs 3.13 and 3.14 show that the recognition of *C. parvum* oocyst antigens by these antisera shows a considerable degree of variation. In particular, the sera from the 10 lambs given a trickle infection recognised more antigen bands than the sera from the lambs infected with a single dose of 10^6 oocysts. There was also a considerable degree of variation in the antigens recognised by different individuals within this group of 10 lambs. Of the six lambs infected with a single dose of 10^6 oocysts, more antigens were recognised by serum from lamb 151 (lane j) than by any of the other five lambs. This was the serum originally used to identify clone λCPS10.

The results obtained when these these sera were used to probe Western blots of lysates prepared both from clone pCPS10/1S and a clone harbouring pMS1S as a control, are shown in Figs 3.15

Fig 3.13 Recognition of *C. parvum* oocyst antigens by serum from infected lambs I.

$3x10^6$ excysted *C. parvum* oocysts were electrophoresed through a 10% polyacrylamide gel, Western blotted, and strips probed with serum from gnotobiotic lambs infected with *C. parvum*.



Track	Lamb
a	64
b	66
c	68
d	69
e	74
f	76
g	81
h	83
i	84
j	89

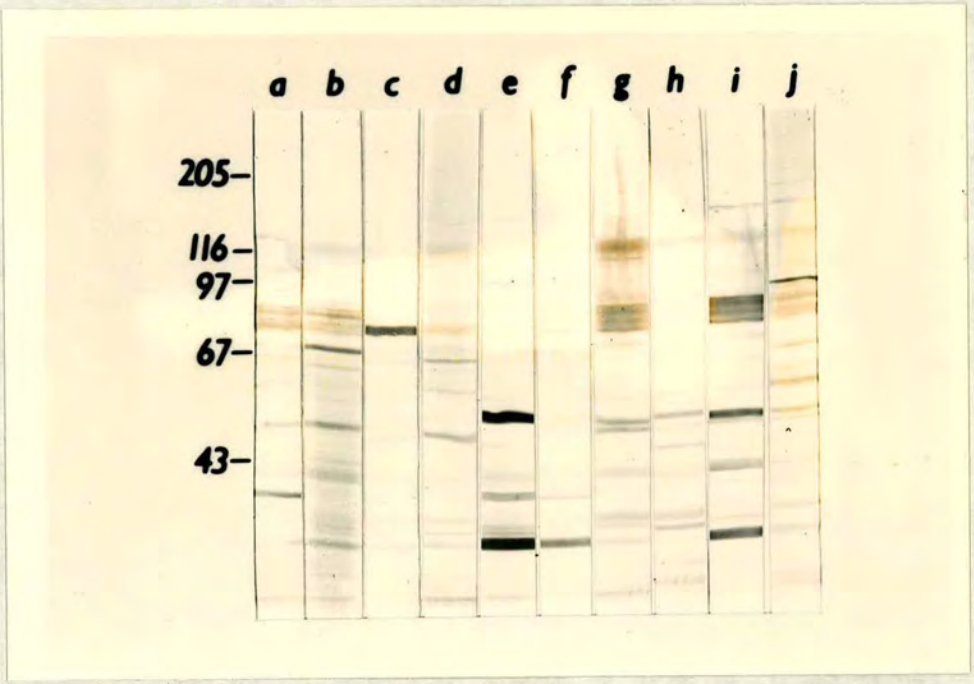
Each lamb received a trickle infection of approximately 50 oocysts per day from 0-22 days of age. Serum samples were from day 22.

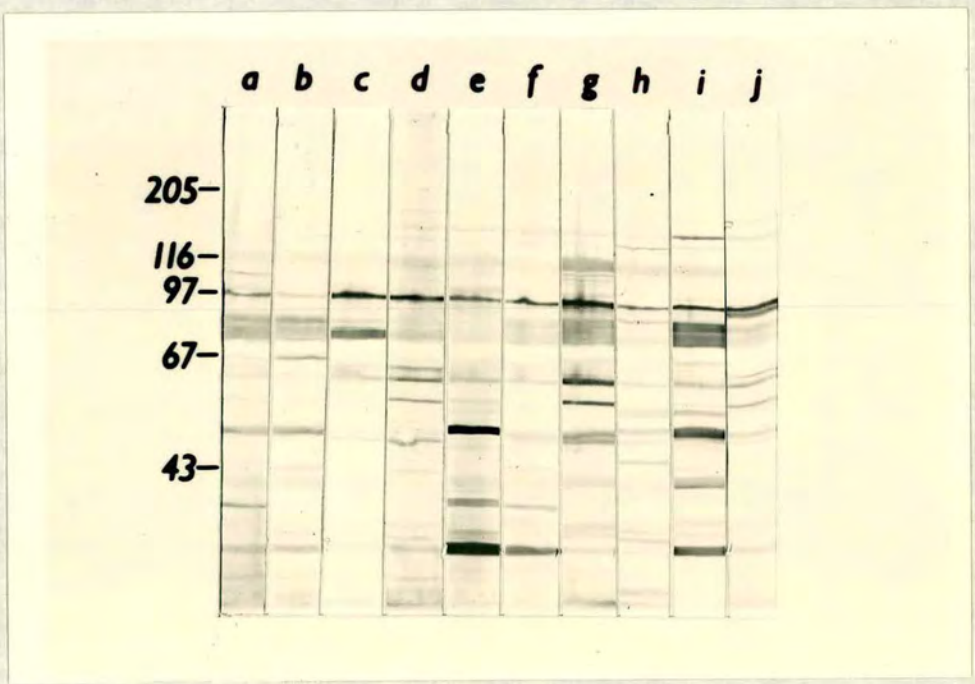
Facing page. Lysates prepared from induced cultures of a clone harbouring pMS1S (top) and pCPS10/1S (bottom) were electrophoresed through a 7.5% polyacrylamide gel, Western blotted, and strips probed with serum from gnotobiotic lambs infected with *C. parvum*.

Track	Lamb
a	64
b	66
c	68
d	69
e	74
f	76
g	81
h	83
i	84
j	89

N.B This is the same serum used to probe a blot of excysted *C. parvum* oocysts (Fig 3.13).

Fig 3.15 Recognition of the recombinant peptide expressed by pCPS10/1S by serum from lambs infected with *C. parvum* I.



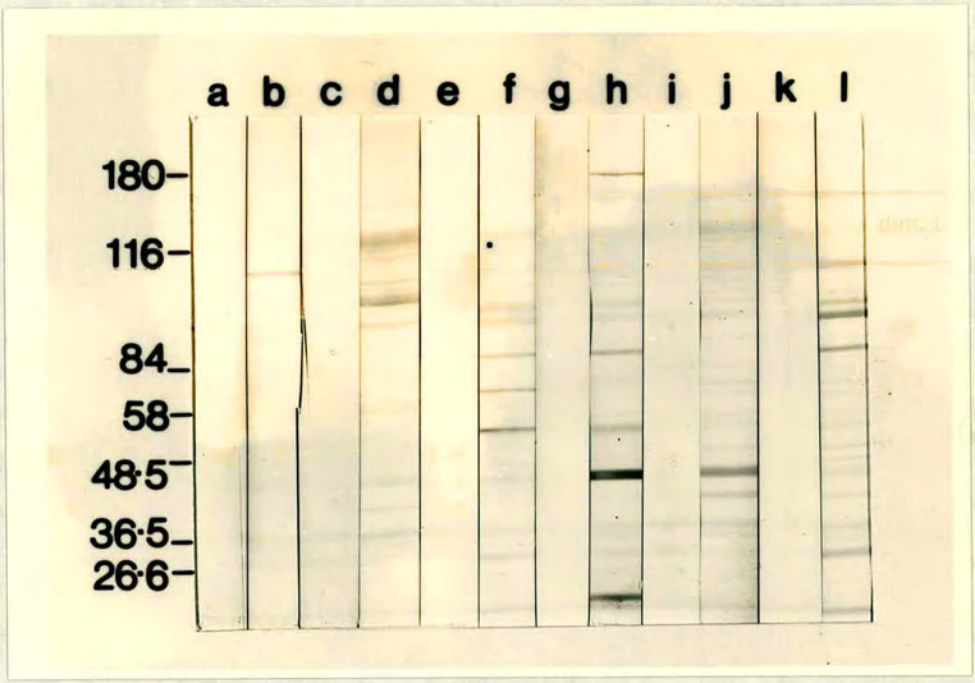


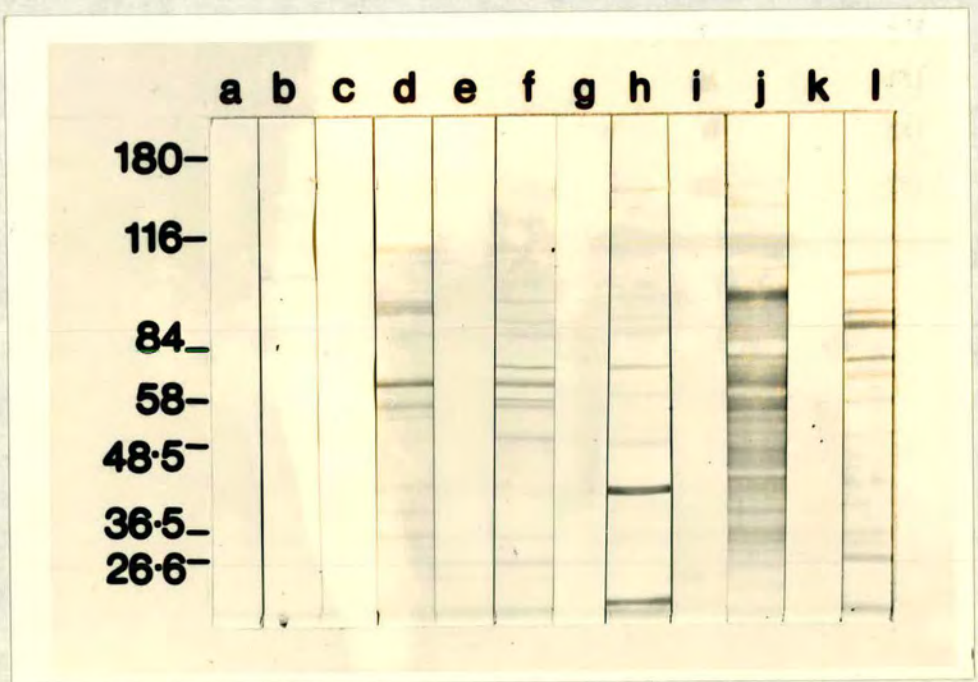
Facing page. Lysates prepared from induced cultures of a clone harbouring pMS1S (top) and pCPS10/1S (bottom) were electrophoresed through a 7.5% polyacrylamide gel, Western blotted, and strips probed with serum from gnotobiotic lambs infected with *C. parvum*.

Track	Lamb	Serum sample (day post infection)
a	96	0
b	96	11
c	108	0
d	108	16
e	146	0
f	146	16
g	150	0
h	150	16
i	151	0
j	151	16
k	152	0
l	152	16

N.B. Serum used to probe tracks b, d, f, h, j and l is the same as that used to probe a blot of excysted *C. parvum* oocysts (Fig 3.14).

Fig 3.16 Recognition of the recombinant peptide expressed by pCPS10/1S by serum from lambs infected with *C. parvum* II.





and 3.16. Of the sera from the six lambs infected with 10^6 oocysts, only serum from lamb 151 (lane j) recognised the peptide encoded by CPS10. This is evident from the fact that this antiserum was the only one which detected extra bands in the pCPS10/1S lysate that were not present in the pMS1S lysate (Fig 3.16). It seems likely that the protein encoded by CPS10 corresponds to one of the bands recognised by serum from lamb 151, but not by the other sera in this group (Fig 3.14), since in addition to being the only serum from this group to recognise the recombinant protein expressed by clone pCPS10/1S, it also recognised more *C. parvum* antigen bands than sera from the other five lambs in this group. Of the sera from the lambs which experienced a trickle infection with *C. parvum*, all 10 recognised the peptide encoded by CPS10, as is evident from Fig 3.15. This shows that all 10 sera recognised a band of approximately 97kDa on a Western blot of the pCPS10/1S lysate which was not recognised on the Western blot of the control (pMS1S) lysate (Fig 3.15).

CHAPTER 4. FURTHER CHARACTERIZATION OF CPR1 AND ITS ENCODED PEPTIDE.

4.1 Introduction.

From here onwards it was decided to characterize further only CPR1 and its encoded peptide. Due to limitations of time it was not possible to continue work also on CPS10.

For further characterization of the peptide encoded by CPR1, it was necessary to prepare antiserum that was specific for the fusion protein. Polyclonal serum was raised in a rabbit and this was used on Western blots to identify the corresponding *C. parvum* antigen.

Further characterization of CPR1 involved Southern blotting to demonstrate i) that it is indeed derived from *C. parvum* DNA, and ii) it occurs as a single copy in the genome. Preliminary Southern blotting experiments also indicated that CPR1 did not hybridize with genomic DNA from *Toxoplasma gondii* or *Sarcocystis* spp., indicating that the CPR1 sequence is not homologous with sequences from these parasites.

4.2 Partial purification of the CPR1 fusion protein.

Recombinant proteins expressed in *E. coli* often accumulate as insoluble inclusion bodies (Marston 1986). Using cell lysis techniques it is therefore often possible to solubilize the

majority of the cell components, and recover the insoluble fusion protein inclusion bodies by centrifugation. The cells may be lysed using enzymes such as lysozyme, in conjunction with detergents such as deoxycholate. Alternatively, the cells may be broken by mechanical means, for example by using a French press or a homogenizer, or by sonication. The following method, adapted from Marston (1987), was used to carry out partial purification of the fusion protein expressed by pCPR1/1S.

A 100ml culture of an clone CPR1/1S clone was grown overnight in L-broth supplemented with 50μg/ml ampicillin and 0.1mM IPTG. The cells were harvested by centrifugation at 6000rpm(4300g) for 15 min at 4^{o}C and subsequently resuspended in 15ml of ice cold lysis buffer (50mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA). To this was added 50μl of 100mM phenylmethylsulphonyl fluoride and 0.75ml of a freshly prepared solution of lysozyme (10mg ml^{-1}). After mixing well the suspension was left on ice for 20 min with occasional swirling. Following addition of 30mg of sodium deoxycholate, the mixture was incubated at 37^{o}C for 10 min, during which time it became very viscous. Next, 300μl of DNase I (1mg ml^{-1}) was added, and incubation at 37^{o}C continued for a further 30 min, after which the mixture was no longer viscous. This cell lysate was then centrifuged at successively higher speeds to determine the optimum conditions for recovery of the fusion protein. The cell lysate was centrifuged at 1000rpm (120g), 2500rpm (750g), 5000rpm (3000g) and finally at 10,000rpm (12,000g). The pellet from each centrifugation step was resuspended in 2.5ml of lysis buffer, 200μl of which were

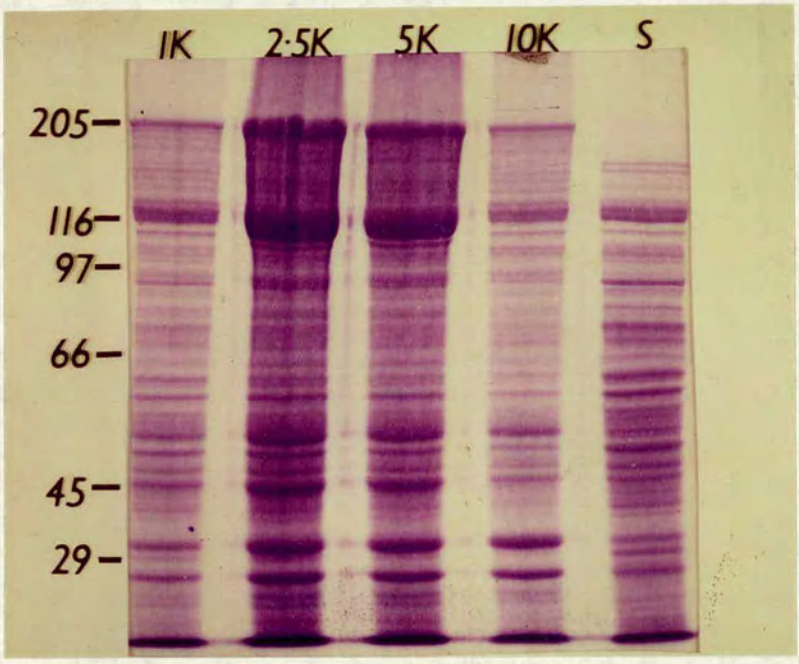
added to 800µl of SDS-PAGE lysis buffer, as was 200µl of the final supernatant. Subsequently, 10µl aliquots from each of these samples were analysed by SDS-PAGE (Fig 4.1).

From Fig 4.1 it can be seen that the pellets recovered after centrifugation at different speeds did not appear to differ in the proportions of different proteins present, only in the total amount of protein; i.e. it was not possible to use differential centrifugation to pellet the insoluble fusion protein, whilst leaving other cell debris in suspension. In each track there was a distinct band with mw of approximately 200kDa, with a ladder-like series of bands with mws between 116 and 200kDa. This seemed to confirm the indication of the results in Chapter 3 that the full length fusion protein was approximately 200kDa and that bands between 116 and 200kDa were its breakdown products. It is evident from the track containing a sample from the final supernatant that a large proportion of *E. coli* proteins were solubilized, including a proportion of the β-galactosidase. However, none of the full length 200kDa fusion protein appeared to have been solubilized.

Further purification of the CPR1 fusion protein was attempted, by washing the pellet with solutions of Triton X-100, SDS, and urea. These attempts were unsuccessful. However, washing in 8M urea did result in a considerable degree of solubilization of the 200kDa fusion protein.

Fig 4.1 Partial purification of fusion protein expressed by pCPR1/1S.

An overnight induced culture of pCPR1/1S was harvested by centrifugation, and cells lysed with lysozyme and deoxycholate. The lysate was centrifuged at successively higher speeds and samples from each pellet and the final supernatant analysed by SDS-PAGE. Tracks labelled 1K, 2.5K, 5K and 10K contain samples taken from the pellet after centrifugation at 1000, 2500, 5000 and 10000 rpm respectively. The track labelled S contains a sample from the final supernatant.



4.3 Preparation of fusion protein-specific antiserum.

Polyclonal rabbit serum was raised against the fusion protein by injection of polyacrylamide gel slices containing the fusion protein. Partially purified fusion protein was prepared as described in section 4.2, except that instead of a series of spins at different speeds, the insoluble fusion protein was sedimented by centrifugation at 5000rpm for 20 min at 4^{o}C. The pellet was washed in 5ml of lysis buffer, recentrifuged, and the pellet resuspended once again in 5ml of lysis buffer. 200µl of this suspension was mixed with 800µl of SDS-PAGE sample buffer, then boiled for 5 min, and 100µl was then loaded into each of two 6cm x 0.75mm block wells in 7.5% SDS polyacrylamide gels. After electrophoresis, a thin strip of gel was cut from the sides of each gel and stained with Coomassie blue to locate the position of the fusion protein. A horizontal strip of gel approximately 3mm wide containing the 200kDa fusion protein was then cut from each gel. The two strips were cut into small pieces and homogenized with 1ml of PBS and 1ml of Freund's complete adjuvant. This preparation was then injected subcutaneously into a rabbit, at four separate dorsal sites on day 0. At 30 days the above procedure was repeated, but using only one polyacrylamide gel, and the gel containing the fusion protein was homogenized with 0.25ml of PBS and 0.5ml of Freund's incomplete adjuvant. This homogenate was injected into the rabbit at two dorsal sites. Finally, this same procedure was repeated on day 85. Blood samples were taken from the rabbit's ear vein on days 0, 10, and 42. On day 104 the rabbit was

anaesthetized with Halothane (May and Baker) and blood taken by cardiac puncture.

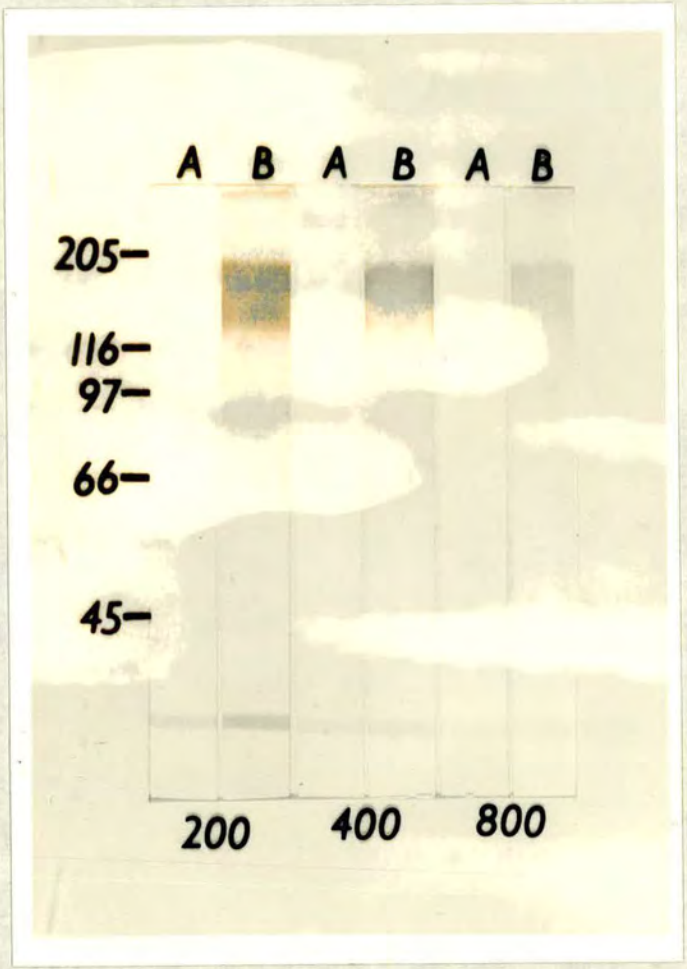
(N.B. All Home Office licensed animal procedures were carried out by Mr S. E. Wright).

Before use, serum prepared from the rabbit blood was diluted by a factor of 10 with 50% (v/v) pig serum in PBS. Antibodies recognising β-galactosidase and other *E. coli* proteins were then absorbed out by adding 100µl of a lysate prepared from an induced culture of *E. coli* NM522 harbouring pMS1S, to each 6ml of diluted serum. The serum/lysate mixture was incubated at room temperature for one hour, and debris then pelleted by centrifugation at 10,000g. The *E. coli* pMS1S/lysate used for preabsorption was prepared in the same way as the partially purified fusion protein used for inoculation of the rabbit.

Fig 4.2 shows a Western blot of partially purified CPR1 fusion protein probed with prebleed serum and with serum taken at 42 days post inoculation. This blot shows that both the prebleed serum and the serum taken at 42 days faintly recognised an *E. coli* protein of low mw. In addition, the serum taken at 42 days strongly recognised the 200kDa fusion protein and its breakdown products. However, it did not recognise a band at 116kDa, indicating that antibodies recognising β-galactosidase had been effectively absorbed out.

Fig 4.2 Western blot showing recognition of the CPR1 fusion protein by polyclonal antiserum raised in a rabbit.

A partially purified lysate prepared from clone pCPR1/1S (section 4.3) was electrophoresed through a 7.5% polyacrylamide gel, Western blotted and strips probed with serum from a rabbit. Serum samples were taken before inoculation with the fusion protein (A) and 42 days post-inoculation (B). Serum dilutions are indicated below the blot.



4.4 Identification of the *C. parvum* antigen encoded by CPR1.

The polyclonal rabbit serum raised against the 200kDa fusion protein was then used to probe a Western blot of excysted *C. parvum* oocysts (Fig 4.3). There was no recognition of any *C. parvum* antigen bands by the prebleed serum, but the serum taken at 42 days strongly recognised a band with mw of approximately 190kDa. There was also faint recognition of several bands of lower mw. It is possible that these may be breakdown products of the 190kDa protein, or alternatively they may be proteins which share common epitopes.

To find out whether the 190kDa antigen recognised by the rabbit serum was a constituent of the sporozoites or of the oocyst wall, Western blotting experiments were carried out using samples which were enriched for sporozoites or oocyst shells. These samples were prepared by centrifugation of *in vitro*-excysted oocysts on a Percoll gradient (section 2.6.4).

Fig 4.4 shows photographs taken of a) the excystation mixture prior to centrifugation on the Percoll gradient, b) pooled fractions containing oocyst shells and c) pooled fractions containing sporozoites. It can be seen that a substantial degree of separation was achieved. However a few unexcysted oocysts can be seen in the sporozoite sample.

Samples of oocyst shells and sporozoites were analysed by

Fig 4.3 Identification of the *C. parvum* antigen encoded by CPR1.

$3x10^6$ *C. parvum* oocysts were excysted *in vitro*, electrophoresed through a 7.5% polyacrylamide gel, Western blotted and strips probed with polyclonal rabbit serum raised against the fusion protein expressed by pCPR1/1S (section 4.3). Tracks labelled A were probed with serum taken from the rabbit prior to inoculation with the fusion protein. Tracks labelled B were probed with serum taken at 42 days post-inoculation. Serum dilutions are indicated below the blot.

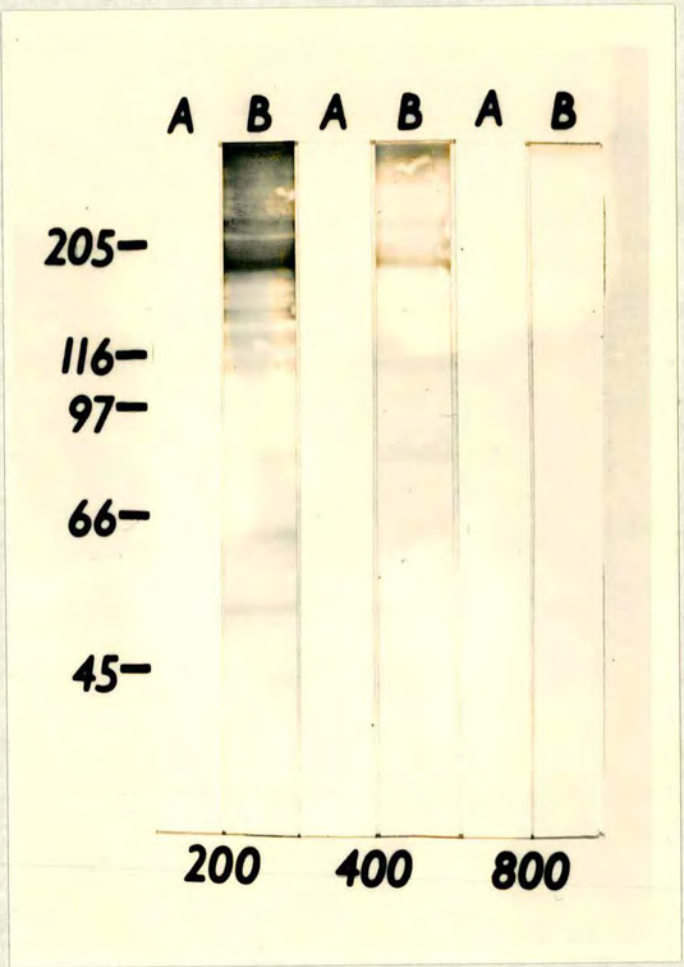


Fig 4.4 Separation of *C. parvum* oocyst shells and sporozoites by Percoll density gradient centrifugation.

Samples which were enriched for sporozoites or oocyst shells were prepared by excysting *C. parvum* oocysts *in vitro* (a), followed by centrifugation on a Percoll density gradient (section 2.6.4). 0.5ml fractions were collected, and appropriate fractions pooled in order to obtain a sporozoite enriched sample (b) and an oocyst shell enriched sample (c).

(a)

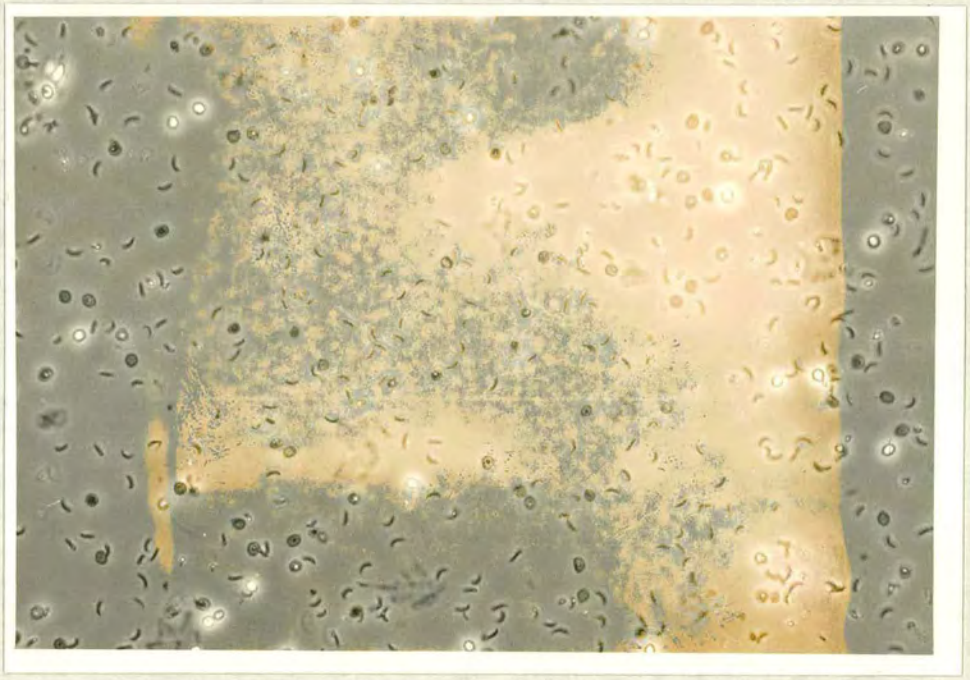
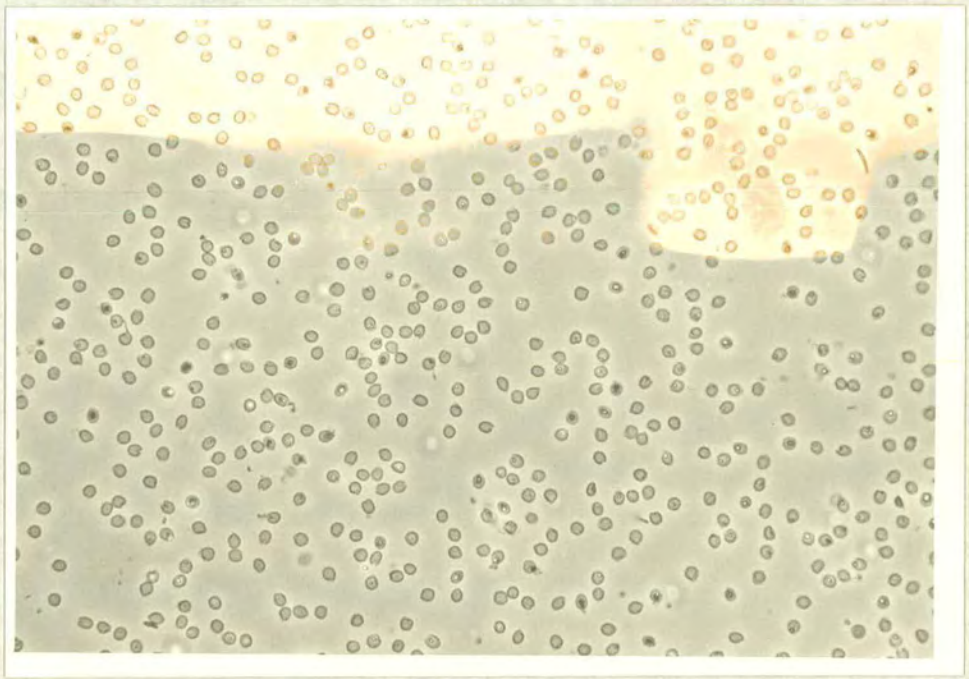


Fig 4.4 continued.

(b)

(c)



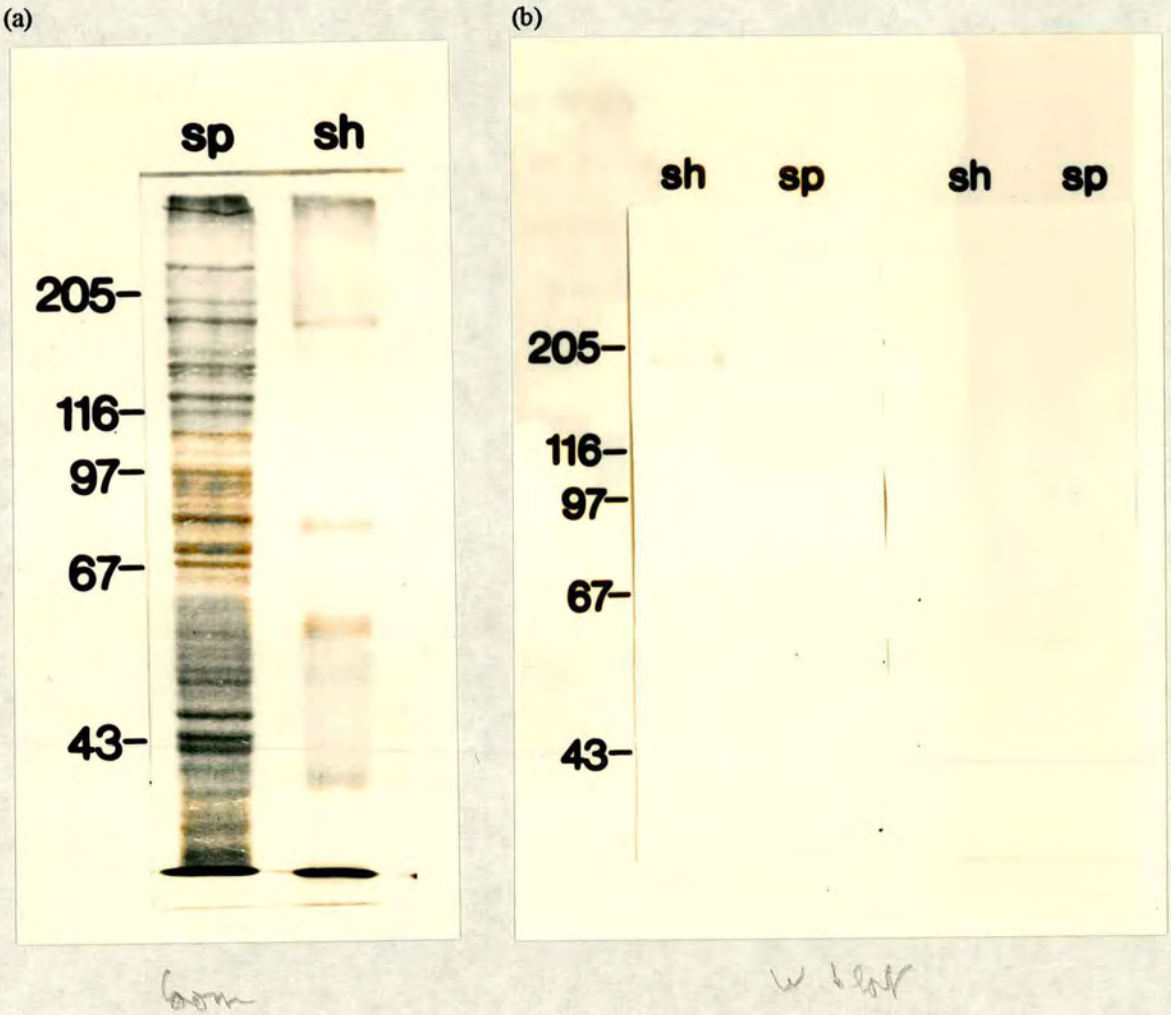
SDS-PAGE and Western blotting. Fig 4.5a shows a polyacrylamide gel stained with Coomassie blue, containing samples of oocyst shells and sporozoites. In the track containing oocyst shells, only four bands were visible, with mws of approximately 190, 85, 55 and 30kDa. By contrast, the sporozoite track contained numerous bands with mws up to >200kDa. Identical samples were Western blotted and probed with the rabbit serum (Fig 4.5b). No bands in either track were recognised by the prebleed serum. Rabbit serum taken after inoculation with the fusion protein recognised the 190kDa band in the oocyst track. There was also faint recognition of the other bands in the oocyst shell track, though this is not visible in Fig 4.5b. There was also faint recognition of a band of approximately 190kDa in the sporozoite track. This may have been due to incomplete separation of sporozoites and oocysts, since a few intact oocysts were visible in the sporozoite sample (Fig 4.4). These results indicate therefore,that the *C. parvum* antigen encoded by CPR1 is likely to be a constituent of the oocyst wall.

4.5 Confirmation that CPR1 is derived from *C. parvum*.

Since the DNA used to make the *C. parvum* DNA library was derived from oocysts which had been prepared from sheep faeces, it is possible that it could contain DNA sequences other than from *C. parvum*; the most likely contaminants being DNA from sheep or *E. coli*. On the basis of the Western blotting experiments, it seemed highly improbable that CPR1 represented a sequence of sheep or *E. coli* DNA. Nevertheless, a Southern blot was carried

Fig 4.5 Western blot of *C. parvum* sporozoite and oocyst shell enriched samples.

Sporozoite and oocyst shell enriched samples, which were prepared by Percoll density gradient centrifugation of excysted *C. parvum* oocysts (section 2.6.4, fig4.4) were electrophoresed through a 7.5% polyacrylamide gel and stained with Coomassie blue (a). Duplicate samples were Western blotted (b)and probed with polyclonal rabbit serum raised against the fusion protein expressed by pCPR1/1S (left panel), or serum from the same rabbit prior to inoculation with the fusion protein (right panel).



out to confirm that CPR1 was from *C. parvum* and not from sheep or *E. coli*.

Sheep DNA (obtained from H. Wright, Moredun Research Institute), *E. coli* DNA from strain JM109 and *C. parvum* DNA were each digested with *Eco*RI, electrophoresed through a 0.8% agarose gel and Southern blotted. For use as a probe the CPR1 insert was first subcloned into pBR322. CPR1 derived from pCPR1/1S could not be used, in view of the possible risk of contamination with pMS DNA containing *lac* sequences that would hybridize with *E. coli* DNA. Therefore, plasmid DNA prepared from a pBR322/CPR1 clone was restricted with *Eco*RI and electrophoresed through a 0.8% agarose gel. The CPR1 band was recovered by the "geneclean" procedure, then labelled with digoxigenin and used to probe the Southern blot (Fig 4.6). The CPR1 probe hybridized to a single band in the track containing *C. parvum* DNA; there was no hybridization to sheep or *E. coli* DNA. This result provided strong evidence that CPR1 does indeed represent a DNA sequence of *C. parvum*.

4.6 CPR1 is single copy in the genome.

Digoxigenin-labelled CPR1 DNA was also used to probe a Southern blot of *C. parvum* DNA digested with different restriction enzymes. Fig 4.7 shows that CPR1 hybridized to a single *Hin*dIII fragment, and to two *Bam*HI fragments of *C. parvum* DNA. This is as would be expected if CPR1 is present as a single copy in the genome, since it contains one *Bam*HI site, but no *Hin*dIII sites.

Fig 4.6 Southern blot showing hybridization of CPR1 to *C. parvum* DNA, but not to sheep or *E. coli* DNA.

Samples of genomic DNA from sheep (S), *E. coli* (E) and *C. parvum* (C) were digested with *Eco*RI, electrophoresed through a 0.8% agarose gel and Southern blotted. The blot was probed with digoxigenin labelled CPR1 DNA.

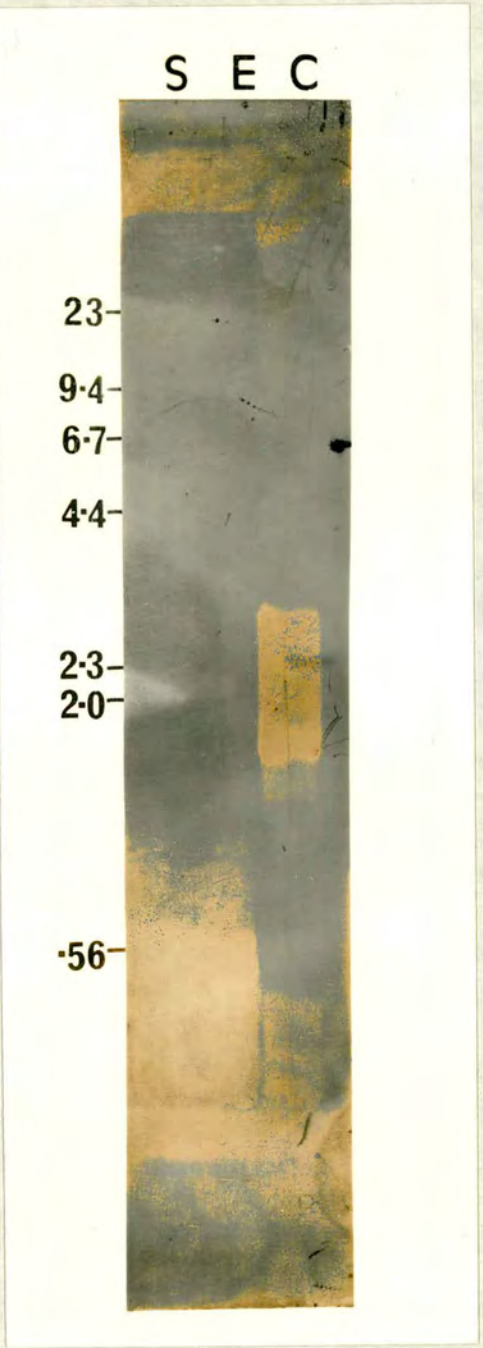
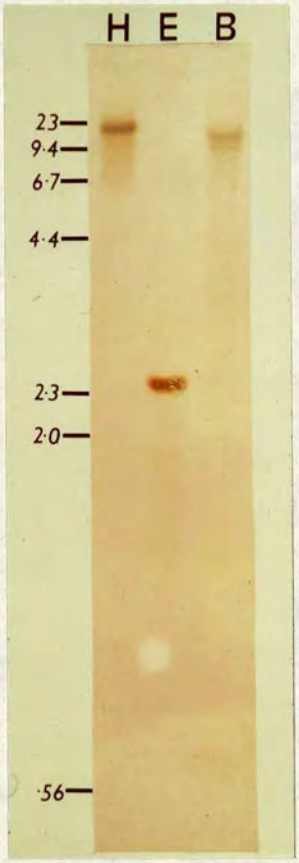


Fig 4.7 Southern blot indicating that CPR1 is present as a single copy in the *C. parvum* genome.

Genomic DNA from *C. parvum* was digested with *Hin*dIII (H), *Eco*RI (E) and *Bam*HI (B), electrophoresed through a 0.8% agarose gel and Southern blotted. The blot was probed with digoxigenin labelled CPR1 DNA.



4.7 CPR1 does not hybridize with *T. gondii* or *Sarcocystis* DNA.

CPR1 appears to be part of a gene encoding a *C. parvum* oocyst wall protein (section 4.4). It would be of interest to determine whether related Coccidia contain DNA sequences which are homologous with CPR1, and whether such sequences are expressed, producing proteins which are homologous with the protein encoded by CPR1. The only Coccidia which could be obtained with ease were *T. gondii* and *Sarcocystis* spp. DNA prepared from each of these parasites was digested with *Eco*RI and electrophoresed together with *Eco*RI-digested *C. parvum* DNA. The Southern blot prepared from this gel was then probed with [^{32}P]-labelled CPR1 DNA. Hybridization was carried out at 42^{o}C in the presence of 50% (v/v) formamide. It can be seen from Fig 4.8 that the CPR1 probe hybridized to an *Eco*RI fragment of approximately 2.4kb in the track containing *C. parvum* DNA, but there was no hybridization with DNA from *T. gondii* or *Sarcocystis*. The same blot was subsequently reprobed with radiolabelled CPR1 in the absence of formamide, and with no formamide in the washing solutions. Hybridization and washing under these less stringent conditions would allow homologous sequences with a higher percentage of mismatching to form duplexes, since a lack of formamide increases the T_m (melting temperature) of DNA duplexes.

The following equation gives an approximation of the melting temperature of DNA duplexes (Meinkoth & Wahl 1984).

$$T_m = 81.5^{\circ}C + 16.6\log_{10}[Na^+] + 0.41(\%G+C) - 0.61(\%formamide) - {}^{500}/_{L}$$

(L=length of hybrid in base pairs, for the probe used in the above experiments L is likely to be between 100bp and 2300bp).

However, there was still no hybridization of the CPR1 probe with DNA from *T. gondii* or *Sarcocystis*.

(N.B. Experiments described in section 4.7. Preparation of the probe, Southern blotting and hybridization were carried out by Miss P. Dalgleish, Moredun Research Institute.)

Fig 4.8 Southern blot showing hybridization of CPR1 to *C. parvum* DNA, but not to *T. gondii* or *Sarcocystis* spp DNA.

Samples of genomic DNA from *C. parvum* (C), *T. gondii* (T) and *Sarcocystis* spp. (S) were digested with *Eco*RI, electrophoresed through a 0.8% agarose gel and Southern blotted. The blot was probed with [^{32}P]-labelled CPR1 DNA.

CHAPTER 5. DNA SEQUENCING, AND DNA AND PROTEIN SEQUENCE ANALYSIS.

5.1 Introduction.

In order to sequence the entire CPR1 insert, which was estimated to be approximately 2.4kb in length, it was decided to subclone it into the phagemid vector Bluescript SK+ (Stratagene). Following restriction mapping, it was possible to construct a series of clones in which various parts of the insert had been deleted. This allowed sequencing of the entire insert in both directions using M13 universal and reverse primers, and in addition, four specially designed oligonucleotide primers. Extensive database searches and sequence analysis were then carried out on the DNA sequence and its deduced amino acid sequence.

5.2 Construction of deletion subclones and DNA sequencing.

CPR1 DNA for sub-cloning was obtained by *Eco*RI digestion of plasmid DNA from clone pCPR1/1S, followed by electrophoresis, excision of the 2.4kb CPR1 band and recovery of the DNA using the "geneclean" procedure. An aliquot of this was then ligated with *Eco*RI digested, dephosphorylated Bluescript SK+ DNA. Following transformation of competent JM109 cells, clones were selected which contained the recombinant plasmid with CPR1 in each of the two possible orientations. Clone pBSCPR1[1], with CPR1 in the same orientation, with respect to the *lac Z* gene, as in

λCPR1, and pBSCPR1^2 with CPR1 in the opposite orientation. The orientations were verified by restriction digests with *Eco*RI and *Bam*HI, since CPR1 contains a *Bam*HI restriction site approximately 0.5kb from its 5′ end (section 3.5.1). Further restriction digests showed that the CPR1 insert also contained recognition sites for *Sma*I and *Pst*I. The approximate positions of these restriction sites in CPR1 are shown in Fig 5.1. Since Bluescript SK+ contains unique sites for each of *Bam*HI, *Sma*I and *Pst*I in its multiple cloning site, it was possible to construct clones in which specific parts of the insert had been deleted. This was achieved by restriction of plasmid DNA from clones pBSCPR1^1 and pBSCPR1^2 with the required enzyme, followed by religation and transformation. Deletion subclones denoted pBSB0.5, pBSB1.9, pBSS0.9, pBSS1.5 and pBSP1.5 are shown in diagrammatic form in Figs 5.2 and 5.3.

During sequencing, recognition sites for *Eco*RV and *Sph*I were identified; these were situated at 280 and 1476 nucleotides from the 5′ end of CPR1, respectively. Restriction digests showed that the *Sph*I site was unique, but that there was an additional *Eco*RV site at approximately 1.7kb from the 5′ end of CPR1. Deletion subclones pBSE0.3, pBSSE0.8, and pBSBE0.2 were constructed by restriction of plasmid DNA from clones pBSCPR1^1 pBSS1.5 and pBSB0.5 respectively, followed by religation and transformation. These clones are shown in Fig 5.4.

Since Bluescript SK+ does not contain a *Sph*I site in its multiple cloning site, it was necessary to use a slightly

Fig 5.1 Approximate positions of restriction sites in CPR1.

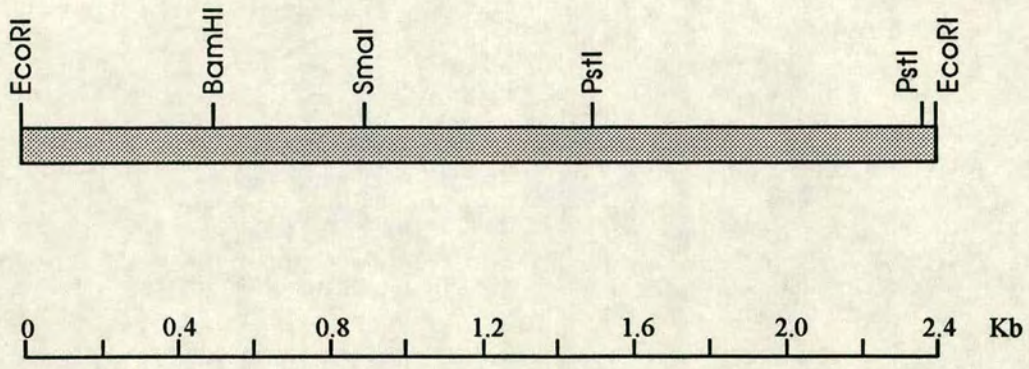


Fig 5.2 Bluescript SK+ subclones used for sequencing I.

Plasmids pBSB1.9 and pBSS1.5 were constructed by restriction and religation of pBSCPR1.[1]

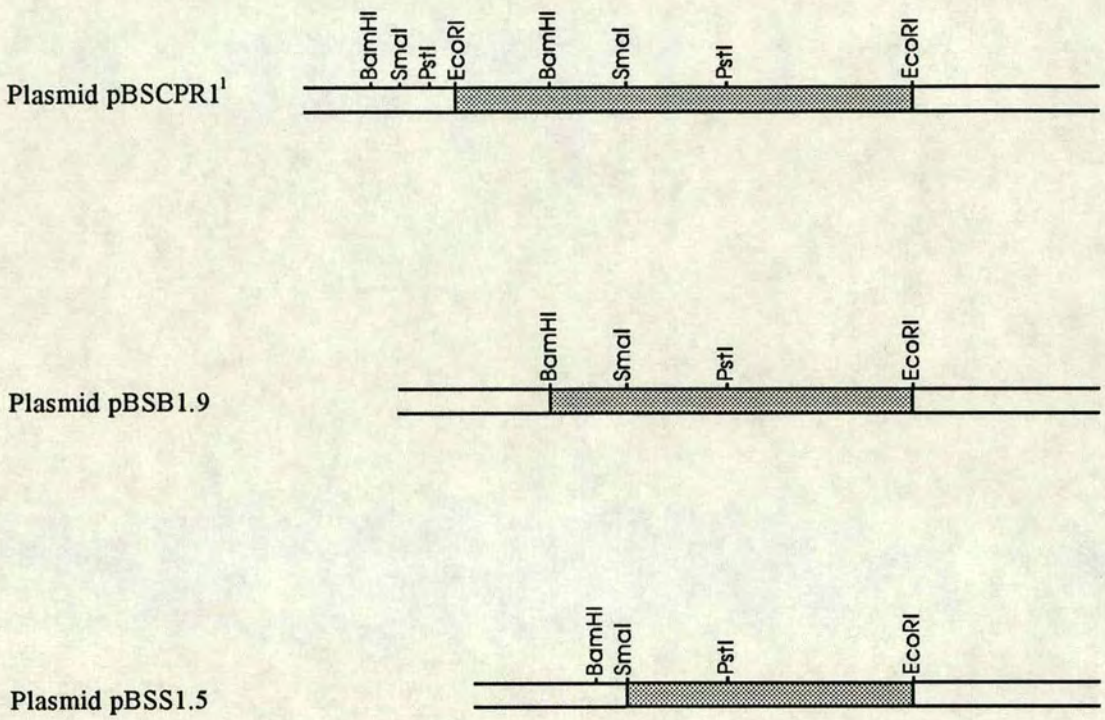


Fig 5.3 Bluescript SK+ subclones used for sequencing II.

Plasmids pBSP1.5, pBSS0.9 and pBSB0.5 were constructed by restriction and religation of pBSCPR1.[2]

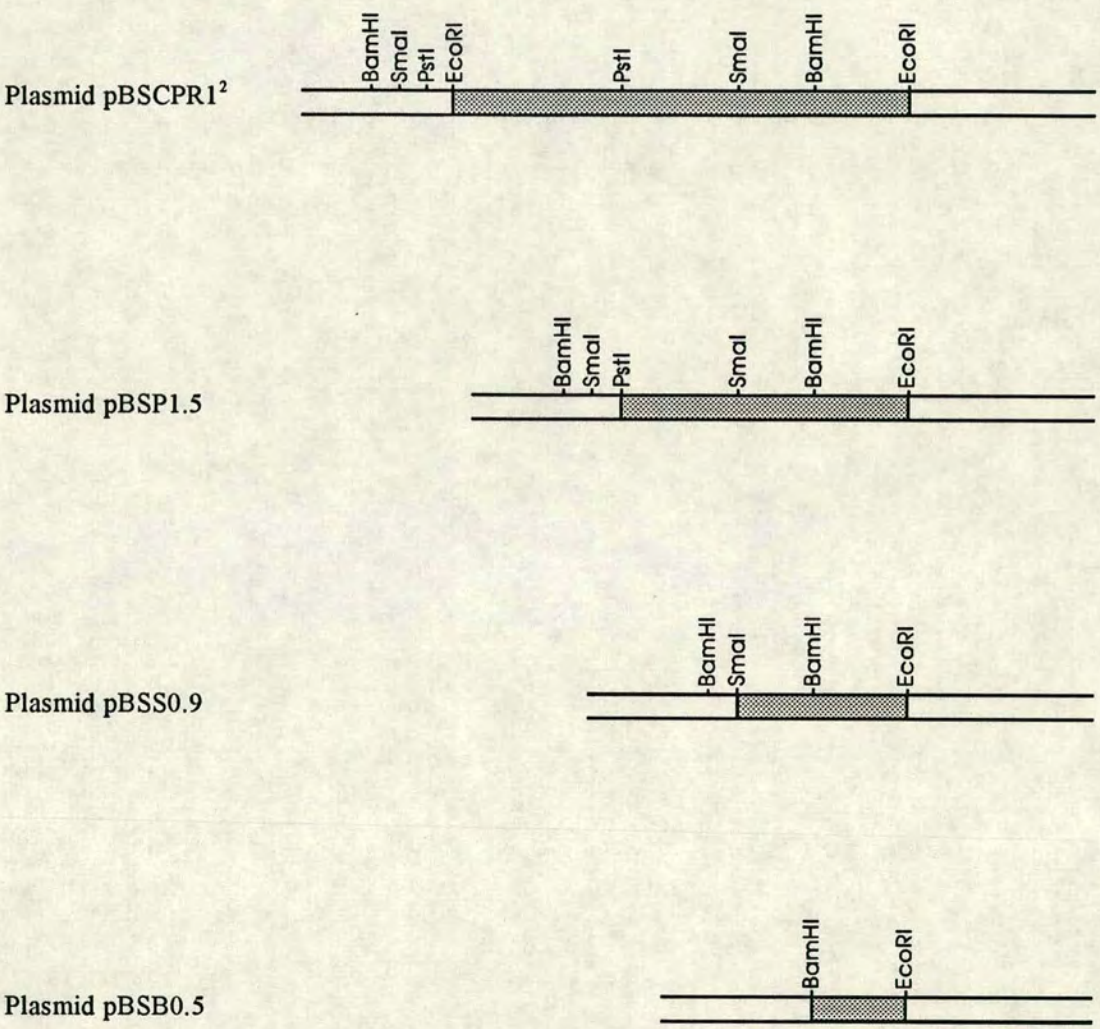
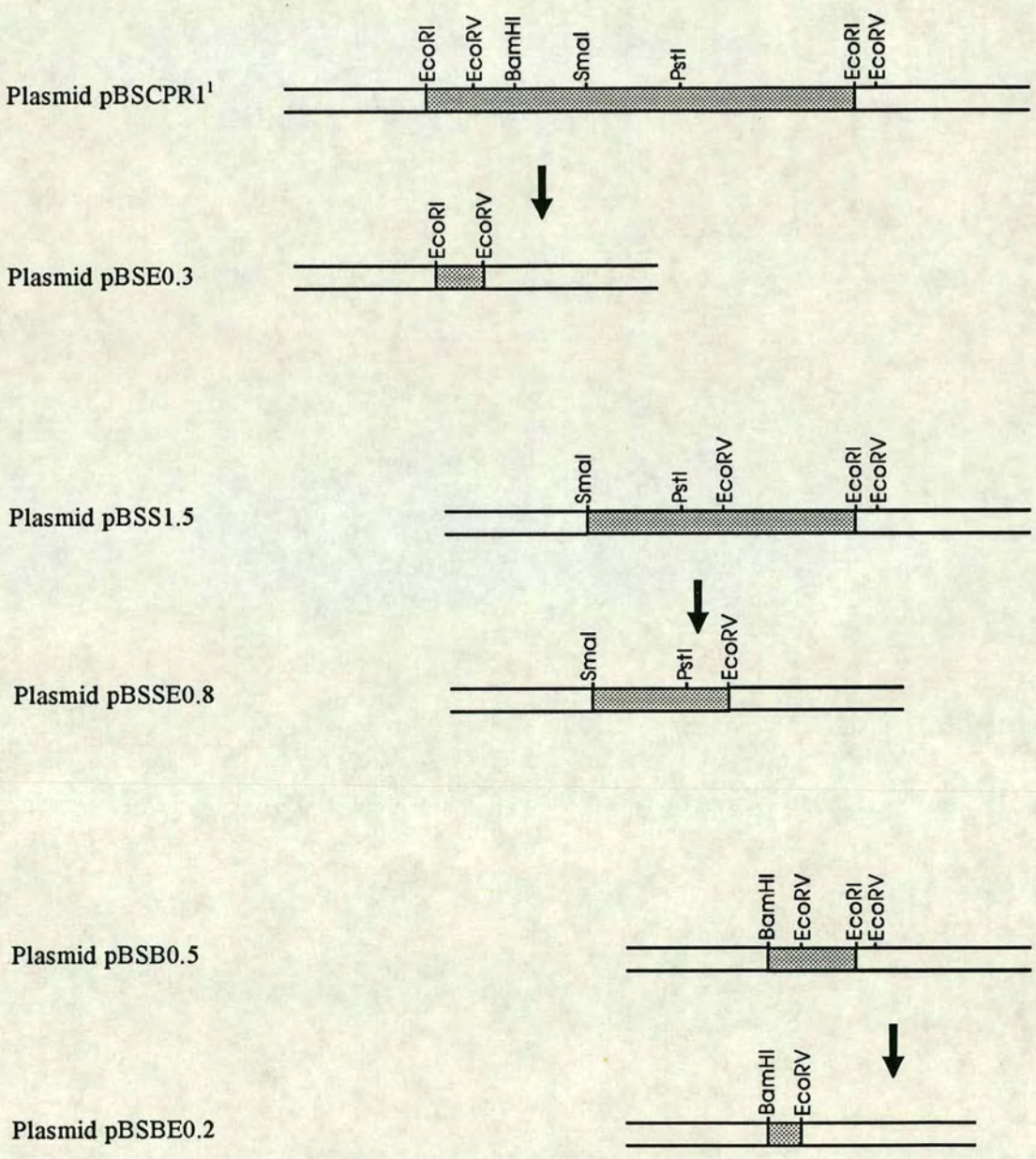


Fig 5.4 Bluescript SK+ subclones used for sequencing III.

Plasmids pBSE0.3, pBSS0.8 and pBSBE0.2 were constructed by restriction and religation of pBSCPR1[1], pBSS1.5 and pBSB0.5 respectively.



different method in order to make use of this recognition site in constructing a deletion subclone. This was achieved by first digesting clone pBSCPR1[1] with *Sph*I and *Kpn*I simultaneously. This resulted in excision of a fragment from the 5′ end of CPR1 through to nucleotide 1476, plus approximately 50bp of the Bluescript SK+ MCS. The resulting linearized plasmid, with uncomplementary *Kpn*I and *Sph*I sticky ends was then blunt ended using T4 DNA polymerase. Religation, followed by transformation, produced clone pBSSp0.9, containing a fragment of CPR1 extending from nucleotide 1476, to its 3′end. Construction of this clone is shown diagrammatically in Fig 5.5.

Using the abovementioned subclones, and M13 universal and reverse primers, it was possible to sequence most of CPR1 in both directions. In order to fill in the gaps, four specially designed oligonucleotide primers were also used, the sequences of which are shown in table 5.1. Using these primers it was possible to complete sequencing CPR1 in both directions. The sequence data obtained from each sub-clone, and the primers used, are given in table 5.2.

5.3 Composition of CPR1 and its encoded peptide (P786).

CPR1 was found to consist of 2359 nucleotides, 2358 of which encode an ORF in which the *Eco*RI site is in frame with that in λgt11 and pMS1S. The DNA sequence of CPR1 and the deduced amino acid sequence of this ORF (P786) are shown in Fig 5.6. The DNA sequence has a relatively low G+C content (39.1%) and there is a

Fig 5.5 Construction of plasmid pBSSp0.9.

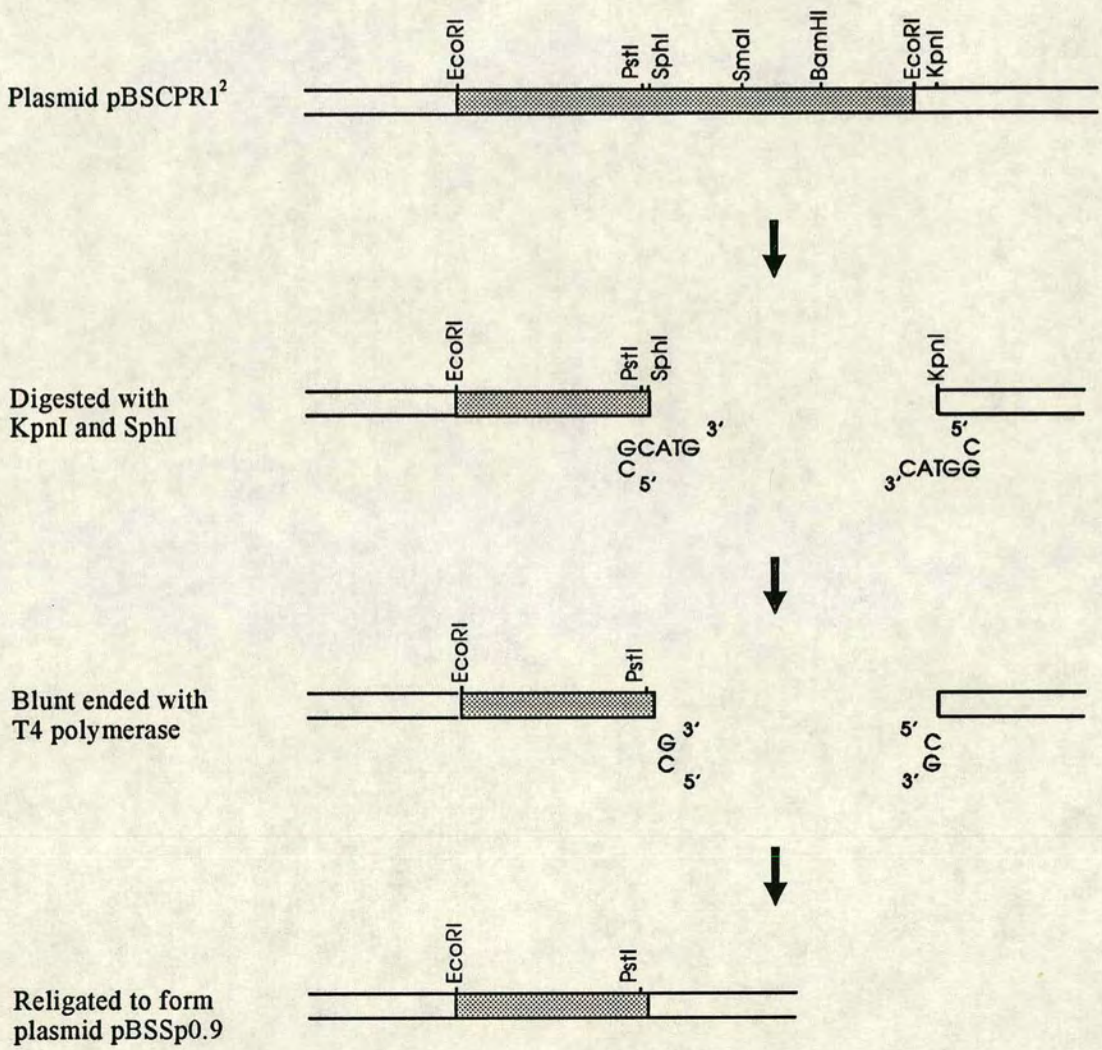


Table 5.1 Oligonucleotide primers used for sequencing.

Primer	Sequence 5' to 3'	
266	GAGCCTCTTTATGATGTT	CPR1 nucleotides 1178-1196 rc
267	CACCTGATGTTAAAGCAC	CPR1 nucleotides 1019-1036
594	TTGGACGGGAACAGAAAC	CPR1 nucleotides 1981-1998 rc
595	CTGTCCTGTTGGATCAAG	CPR1 nucleotides 1803-1820
M13U	GTAAAACGACGGCCAGT	M13 universal primer
M13R	AACAGCTATGACCATG	M13 reverse primer

rc denotes reverse complement.
CPR1 oligonucleotide primers were obtained from Oswell DNA Service, Edinburgh University.

Table 5.2 CPR1 sequence obtained from each subclone and primer combination.

Plasmid	Primer	Sequence	Direction
pBSCPR1[1]	M13U	1-309	Forward
pBSBE0.2	M13R	278-555	Forward
pBSB1.9	M13U	552-910	Forward
pBSS1.5	M13U	840-1182	Forward
pBSCPR1[1]	267	1069-1491	Forward
pBSSp0.9	M13R	1476-1890	Forward
pBSCPR1[1]	595	1841-2359	Forward
pBSE0.3	M13R	1-283	Reverse
pBSB0.5	M13U	267-555	Reverse
pBSS0.9	M13U	512-840	Reverse
pBSCPR1[1]	266	833-1155	Reverse
pBSP1.5	M13U	1082-1484	Reverse
pBSSE0.8	M13R	1303-1598	Reverse
pBSCPR1[1]	594	1588-1955	Reverse
pBSCPR1[1]	M13R	1841-2359	Reverse

Fig 5.6 DNA sequence of CPR1 and its deduced amino acid sequence (P786).

```
   1  gaattcgaatgcccaccaggtacaattttaaaagatgatcaatgtcaatcgatagaaaga
      E  F  E  C  P  P  G  T  I  L  K  D  D  Q  C  Q  S  I  E  R
  61  gttgatacaatttgtccaccagggtttgtagataatggcgaagattgtgtccaatttttct
      V  D  T  I  C  P  P  G  F  V  D  N  G  E  D  C  V  Q  F  S
 121  gcaccagagaaaatttgcccccaaggattttctctttccggaaaacaatgtgttaaaaca
      A  P  E  K  I  C  P  Q  G  F  S  L  S  G  K  Q  C  V  K  T
 181  gaatctgctccaagattaacagaatgcccaccaggtacaaccttggaaaataacagttgt
      E  S  A  P  R  L  T  E  C  P  P  G  T  T  L  E  N  N  S  C
 241  atttcatatgaactagaagatgccatttgtccacctggatatctcgacaatggatcagac
      I  S  Y  E  L  E  D  A  I  C  P  P  G  Y  L  D  N  G  S  D
 301  tgtgttcagttttctcaaccagaaaaggagtgtccaacaggttttgtattaattggaaaa
      C  V  Q  F  S  Q  P  E  K  E  C  P  T  G  F  V  L  I  G  K
 361  caatgtacccaaactactcaagctccaccacaaccagagtgtcctccaggtacaaacctg
      Q  C  T  Q  T  T  Q  A  P  P  Q  P  E  C  P  P  G  T  N  L
 421  gtaaatggacaatgccaaaaagttgaaaggataaatatggtatgtccaactggttttatt
      V  N  G  Q  C  Q  K  V  E  R  I  N  M  V  C  P  T  G  F  I
 481  gataatggtacaaattgtgcttctttctccgcaccaaacagagaatgcccacctggatat
      D  N  G  T  N  C  A  S  F  S  A  P  N  R  E  C  P  P  G  Y
 541  acactttctggatcccaatgcgagcaaataaaagaagcacctcctgtttcagaatgtcca
      T  L  S  G  S  Q  C  E  Q  I  K  E  A  P  P  V  S  E  C  P
 601  ccaggatataaacttcaaggaaatcaatgtactgcactaaaaatgatcgatgctatctgc
      P  G  Y  K  L  Q  G  N  Q  C  T  A  L  K  M  I  D  A  I  C
 661  ccagatggatttttaccaaatggagacgattgtatccaattttctcctgcttcaactgta
      P  D  G  F  L  P  N  G  D  D  C  I  Q  F  S  P  A  S  T  V
 721  tgtcctactggattcactctacaaaatcaacagtgtgttcaaacaactacctcaccaaaa
      C  P  T  G  F  T  L  Q  N  Q  Q  C  V  Q  T  T  T  S  P  K
 781  acaccagaatgtcctccaggttctgcgttggatggagactcgtgcacaagacttgttccc
      T  P  E  C  P  P  G  S  A  L  D  G  D  S  C  T  R  L  V  P
 841  ggggctcttcaatacgtttgtcctgttggtactagagaggggacgtttgcgtagagaga
      G  A  L  Q  Y  V  C  P  V  G  T  R  E  G  D  V  C  V  E  R
 901  tcgattagttcgcctgttttggaatgcccacctggttattcattggaaacaggtaaacaa
      S  I  S  S  P  V  L  E  C  P  P  G  Y  S  L  E  T  G  K  Q
 961  tgtgttagaagaagccaatatgactgttcagtaacaacttatgttacagagtgtaaaaca
      C  V  R  R  S  Q  Y  D  C  S  V  T  T  Y  V  T  E  C  K  T
1021  cctgatgttaaagcactaagaagattagcagctgcaaaagaaacatcaacagtttatgaa
      P  D  V  K  A  L  R  R  L  A  A  A  K  E  T  S  T  V  Y  E
1081  acatctgagatacaaaatccaggacatcatcatggtcattctcatgggcattcacattca
      T  S  E  I  Q  N  P  G  H  H  H  G  H  S  H  G  H  S  H  S
1141  caagttataccaattcaaacccagaatatacatacacaacatcataaagaggctccaagg
      Q  V  I  P  I  Q  T  Q  N  I  H  T  Q  H  H  K  E  A  P  R
1201  ccaatttgtgaagatgttccaaaaattaccccaaaaacttgtacaaaagctgattctgtc
      P  I  C  E  D  V  P  K  I  T  P  K  T  C  T  K  A  D  S  V
1261  ccagctgtgcctatttgcgagaacaatgctgaacttgtaggaaaagaatgtgtattaaca
      P  A  V  P  I  C  E  N  N  A  E  L  V  G  K  E  C  V  L  T
1321  aattactacccattagaagcaatttgtcaagatggaacaagatcaaaagagtgtgctaag
      N  Y  Y  P  L  E  A  I  C  Q  D  G  T  R  S  K  E  C  A  K
1381  tttgtaaaaactccacctactttaaaatgtccgccaggttctgtagatgtaggatctcaa
      F  V  K  T  P  P  T  L  K  C  P  P  G  S  V  D  V  G  S  Q
1441  tgtcaagttaacaaatattcaccatatgatcttgcatgccctgcaggatatgcattggtt
      C  Q  V  N  K  Y  S  P  Y  D  L  A  C  P  A  G  Y  A  L  V
```

Fig 5.6 continued.

```
1501  ggagacaaatgcgctaccacaagagaaaaagtttgcccgaatgaaagttgccaaagagtt
      G  D  K  C  A  T  T  R  E  K  V  C  P  N  E  S  C  Q  R  V
1561  gtaactgcgcctgtttctttaacttgtccccctggatatcaccaaatagatgaagttatg
      V  T  A  P  V  S  L  T  C  P  P  G  Y  H  Q  I  D  E  V  M
1621  aatatttctgctcatccacaccacagacacttagctggggttcaatctacttctcaaaag
      N  I  S  A  H  P  H  H  R  H  L  A  G  V  Q  S  T  S  Q  K
1681  ggatattctcatggacataaatatactcctgtaatttctcagccaccacaaccagttcca
      G  Y  S  H  G  H  K  Y  T  P  V  I  S  Q  P  P  Q  P  V  P
1741  gttgttgctcctattcagcaaatgaaatgcatccatgcagaccatgctccatataatctt
      V  V  A  P  I  Q  Q  M  K  C  I  H  A  D  H  A  P  Y  N  L
1801  atctgtcctgttggatcaagacttgtagcggataaatgtgttacatattcggataaaata
      I  C  P  V  G  S  R  L  V  A  D  K  C  V  T  Y  S  D  K  I
1861  tgtccaaatggtaattgcgagcgtatatataatgagcctgctgaattagtatgccctcca
      C  P  N  G  N  C  E  R  I  Y  N  E  P  A  E  L  V  C  P  P
1921  ggattctcatcatctaaaccaattcagccaataagccattctcatattaaccatccaaat
      G  F  S  S  S  K  P  I  Q  P  I  S  H  S  H  I  N  H  P  N
1981  gtttctgttcccgtccaaccacaaactattaaccaaccacaagtaattcaacaaagacaa
      V  S  V  P  V  Q  P  Q  T  I  N  Q  P  Q  V  I  Q  Q  R  Q
2041  gtaaattatcagccacaagtaattcatcaaacacaggaaattttaacaacttatccaact
      V  N  Y  Q  P  Q  V  I  H  Q  T  Q  E  I  L  T  T  Y  P  T
2101  ccagtttaccaaaccggcacaatttatcaaggacatcatcatcatcatcatcatcatcac
      P  V  Y  Q  T  G  T  I  Y  Q  G  H  H  H  H  H  H  H  H  H
2161  agaaatctagcttcccctgagtgcattaagacaatttcagtaccttatattttaaaatgc
      R  N  L  A  S  P  E  C  I  K  T  I  S  V  P  Y  I  L  K  C
2221  gaatctccatttatttttagatggcgacaaatgtatcgaaaaaacagaaaaaatttgtcta
      E  S  P  F  I  L  D  G  D  K  C  I  E  K  T  E  K  I  C  L
2281  caaggtgactgcagaaaacaagtcgtcgttccaccaactctttcatgtccacaaggttac
      Q  G  D  C  R  K  Q  V  V  V  P  P  T  L  S  C  P  Q  G  Y
2341  agaaatgccaacggaattc  2359
      R  N  A  N  G  I
```

corresponding bias towards the use of A or T in the third codon position within this ORF (Table 5.3); 17.9% of bases in the third codon position are G or C, compared with 53.1% and 46.3% in the first and second postions respectively.

The deduced amino acid sequence (P786) has an unusual amino acid composition in comparison with the amino acid proportions averaged over the entire National Biomedical Research Foundation (NBRF) protein sequence database (figures are for NBRF protein sequence database version 32.0 and were obtained by S. J. McQuay, Biocomputing Research Unit, University of Edinburgh). In particular, 57 (7.3%) of the 786 residues are cysteine, compared with 1.9% for the NBRF database as a whole. There are also unusually high proportions of proline (11.1% compared with 5.2%), glutamine (7.4% compared with 4.1%) and histidine (4.1% compared with 2.3%). The distribution of cysteine and histidine residues within the sequence is also noteworthy. Cysteine residues are present in four distinct cysteine-containing regions which span residues 1-338, 403-529, 590-638 and 728-786. All but two of the 32 histidine residues are found outwith these cysteine-containing regions.

5.4 Arrangement of repeat units in P786.

Dotmatrix analysis (Maizel & Lenk 1981) was carried out using GCG.COMPARE and GCG.DOTPLOT. (The prefix "GCG." indicates programs which form part of the GCG (Genetics Computer Group) sequence analysis software package (Devereux *et al* 1984), which

Table 5.3 Codon usage by CPR1.

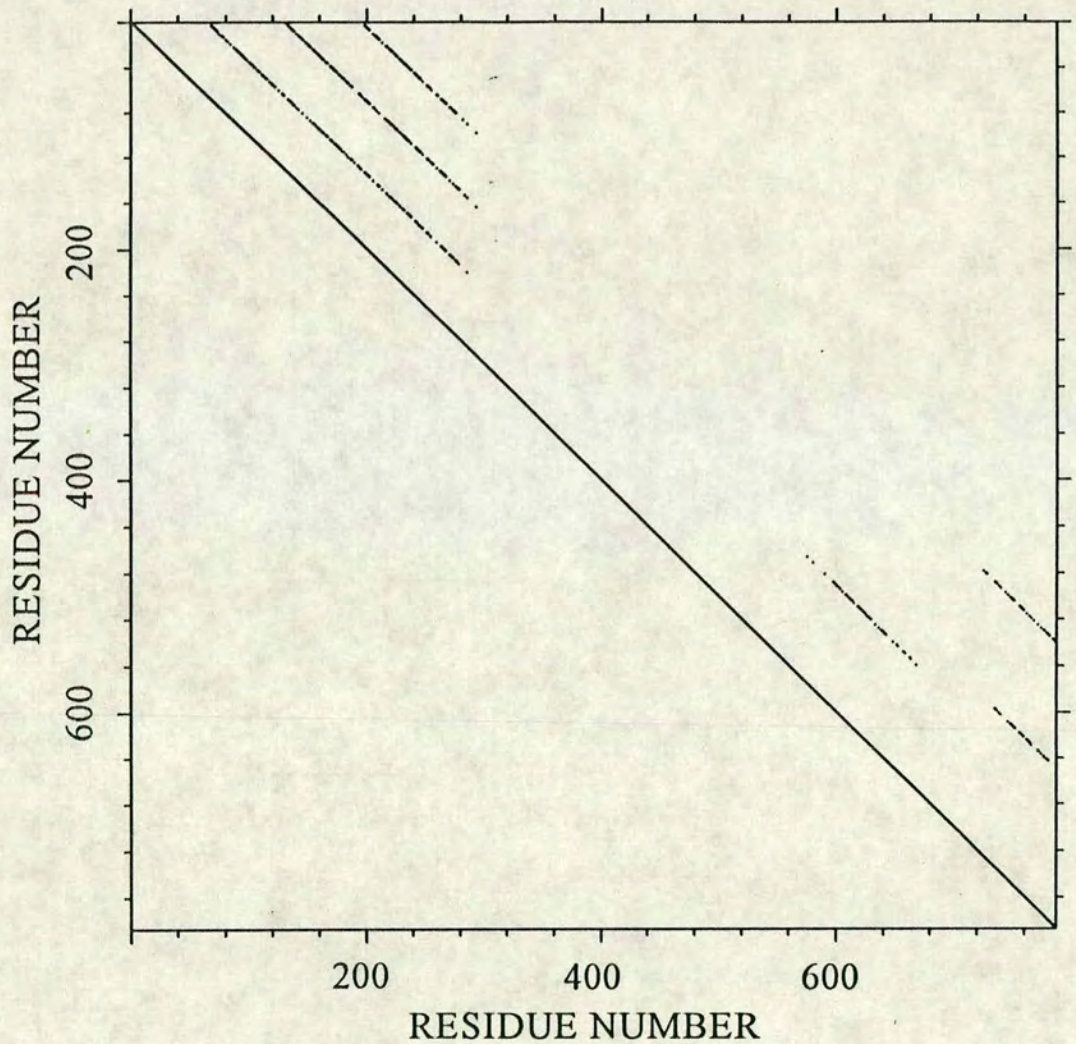
10	UUU Phe	23	UCU Ser	21	UAU Tyr	36	UGU Cys
4	UUC Phe	4	UCC Ser	5	UAC Tyr	21	UGC Cys
14	UUA Leu	17	UCA Ser	0	UAA ***	0	UGA ***
5	UUG Leu	5	UCG Ser	0	UAG ***	0	UGG Trp
10	CUU Leu	24	CCU Pro	27	CAU His	1	CGU Arg
1	CUC Leu	4	CCC Pro	5	CAC His	0	CGC Arg
6	CUA Leu	57	CCA Pro	50	CAA Gln	0	CGA Arg
1	CUG Leu	2	CCG Pro	8	CAG Gln	0	CGG Arg
29	AUU Ile	21	ACU Thr	24	AAU Asn	3	AGU Ser
6	AUC Ile	7	ACC Thr	8	AAC Asn	2	AGC Ser
10	AUA Ile	30	ACA Thr	35	AAA Lys	19	AGA Arg
4	AUG Met	0	ACG Thr	4	AAG Lys	2	AGG Arg
33	GUU Val	19	GCU Ala	21	GAU Asp	15	GGU Gly
5	GUC Val	2	GCC Ala	10	GAC Asp	3	GGC Gly
20	GUA Val	12	GCA Ala	31	GAA Glu	27	GGA Gly
1	GUG Val	3	GCG Ala	14	GAG Glu	5	GGG Gly

is available on the SEQNET molecular biology computer facility). When applied to protein sequences the GCG.COMPARE program uses a symbol comparison table which is based on an empirical study of related proteins (Dayhoff *et al* 1978). Different scoring tables are based on different accepted point mutation (PAM) values. PAM is a measure of divergent evolution, 1 PAM = 1 substitution per 100 aligned residues. (GCG.COMPARE uses the scoring table based on 250 PAMS).

Dotmatrix analysis of P786 indicates that the sequence is distinctly repetitive. The dotplot in Fig 5.7 shows that there is a strongly repetitive region within the first cysteine-containing region, indicated by the diagonal lines in the top left hand corner of the dotplot. A further repeating element in the C-terminal half of the sequence is indicated by the diagonal lines in the bottom right hand corner. Examination of the amino acid sequence reveals that the repeat regions are each based on conserved cysteine residues. Of the 32 cysteines in the N-terminal cysteine-containing region (residues 1-338), the first 26 are arrayed in 13 consecutive copies of a repeat unit consisting of the motif CPXG (7X) C, followed by either 9 or 11 residues. The first 12 of these repeats are arranged as a series of four higher order repeats, each of 65 amino acids. Optimal alignment of these four 65-amino acid repeats shows that, as well as complete conservation of the cysteine residues, 16 of the remaining 59 amino acids are also conserved in all four repeats. Of these conserved residues the asparagine-glycine sequence may be of particular interest. Under acidic conditions

Fig 5.7 Dot matrix analysis of P786.

The P786 sequence was compared with itself using a window of 50 and a stringency of 32. The solid diagonal line represents the line of identity where each 50 residue segment is compared with itself. The dotted lines represent internal repeats which are discussed in the text.



this dipeptide can be converted to an imide structure, which may be hydrolysed under alkaline conditions, resulting in the formation of a β-aspartyl peptide bond (Aswad & Johnson 1987).

The additional repeating element, clearly visible on the dotplot (Fig 5.7), consists of three non-contiguous repeats, each of 53 amino acids. In common with the 65-amino acid repeats described above, these three repeats show complete conservation of the positions of the cysteine residues between them. Each of the 53-amino acid repeats falls within one of the three remaining cysteine-containing regions.

Fig 5.8 shows a diagrammatic representation of the arrangement of repeat units in P786, plus alignments of the 53- and 65-amino acid repeats described above.

5.5 Database searching.

In order to find out whether CPR1 or its encoded peptide (P786) was homologous with any sequences in DNA and protein sequence databases, extensive database searching was carried out using software and databases available on the SEQNET molecular biology computer facility, based at the SERC Daresbury Laboratory, Warrington, UK.

5.5.1 DNA sequence database searching.

CPR1 was compared with the EMBL DNA sequence database using the

Fig 5.8 Diagrammatic representation of P786, and optimal alignments of repeats.

The diagram shows the relative positions of the repeats and the histidine-containing regions in P786. a_9 and a_{11} indicate the repeat units CPXG (7X) C (9X), and CPXG (7X) C (11X), respectively. Alignments of repeats A and B are shown, and their locations indicated on the diagram. The solid vertical bars indicate the locations of histidine residues.

A

```
    3-67 E C P P G T I L K D D Q C Q S I E R V D T I C P P G F V D N G E D C V Q F S A P E K I C P Q G F S L S G K Q C V K T E S A P R L T
  68-132 E C P P G T T L E N N S C I S Y E L E D A I C P P G Y L D N G S D C V Q F S Q P E K E C P T G F V L I G K Q C T Q T T Q A P P Q P
 133-197 E C P P G T N L V N G Q C Q K V E R I N M V C P T G F I D N G T N C A S F S A P N R E C P P G Y T L S G S Q C E Q I K E A P P V S
 198-262 E C P P G Y K L Q G N Q C T A L K M I D A I C P D G F L P N G D D C I Q F S P A S T V C P T G F T L Q N Q Q C V Q T T T S P K T P
```

B

```
 481-515 C Q V N K Y S P Y D L A C P A G Y A L V G D K C A T T R E K V C P N E S C Q R V V T A P V S L T C P P G Y
 590-642 C I H A D H A P Y N L I C P V G S R L V A D K C V T Y S D K I C P N G N C E R I Y N E P A E L V C P P G F
 728-780 C I K T I S V P Y I L K C E S P F I L D G D K C I E K T E K I C L Q G D C R K Q V V V P P T L S C P Q G Y
```

FASTA DNA and protein sequence database searching program (Pearson & Lipman 1988). This program is an improved version of FASTP (Lipman & Pearson 1985), which runs a modification of the original algorithm described by Wilbur and Lipman (1983).

The ten highest scoring alignments obtained by comparing CPR1 with the EMBL DNA sequence database involved between 30 and 198bp of CPR1 aligned with database sequences with between 55 and 94% identity. There was a strong negative correlation between the length of the alignment and the percentage identity, as might be expected. Nine of these alignments included the region of CPR1 encoding the run of nine histidine residues or its reverse complement. Similarites between DNA sequences encoding such heavily biased amino acid sequences do not necessarily indicate evolutionary relationships between the DNA sequences, nor a similar function for the encoded proteins. None of the database DNA sequences with which alignments were identified by the DNA database search was considered to be homologous with CPR1, i.e. share a common ancestor.

With DNA sequences which are known to be coding for proteins, it is more appropriate to search for sequence similarites at the protein level. Homologous protein sequences are likely to be more conserved than the genes that encode them since selection acting at the protein level may conserve the amino acid sequence, whilst silent mutations, or mutations resulting in conservative substitutions, may result in more substantial changes in the DNA sequence. Differences in G+C content and

codon usage between different species may cause substantial divergence at the DNA level between homologous genes, while maintaining relatively similar amino acid sequences. Nevertheless, even in the case of protein-encoding DNA sequences, it is advisable to do at least one DNA database search, since a homologous gene could have been sequenced and submitted to a database, without the submitter realising that it encoded a protein.

5.5.2 Protein sequence database searching.

Protein sequence database searching was carried out using the 'PROSRCH' protein sequence database searching program (Coulson *et al* 1987), which is run at the Biocomputing Research Unit, University of Edinburgh. This program uses the best local similarity algorithm of Smith and Waterman (1981) and aims to find out whether any part of the query sequence is similar to any part of any sequence in the database. The search program collects alignments between the query sequence and database sequences, and sorts them into score order. Scores are calculated using the scoring system of Dayhoff *et al* (1978), which has different scoring tables for different PAM values (see page 144).

P786 was compared with the OWL protein sequence database using the PROSRCH program. The OWL protein sequence database is a composite database made by merging all the other protein sequence databases, plus some translated DNA sequences from the

GENBANK DNA sequence database. Searches were run using 100 and 250 PAM tables. In addition to comparing the entire P786 sequence with the database, the searches were repeated with shorter segments of the amino acid sequence. This was done to remove the histidine-rich regions of P786 from the search as these were always involved in the highest scoring alignments and could have masked interesting alignments with other regions of the sequence.

When the OWL protein sequence database was searched using the entire P786 sequence as a query, the highest scoring alignments involved the histidine-rich regions of P786 and histidine-rich proteins such as those found in *Plasmodium* spp. Such alignments between similarly biased sequences are not readily informative as they often do not indicate a related function (McQuay 1991). Database searching with segments of P786 minus the histidine-rich regions identified alignments between residues 285 and 315 of P786, and other proteins with a high proportion of glutamine such as wheat gliadin and certain homeotic proteins. Again the relevance of such alignments is unclear. In addition, proteins with a high proportion of cysteine, or having a similar spacing of cysteine residues to that seen in P786 were also evident among the alignments identified. These proteins included human thrombospondin, human von Willebrand factor, thrombomodulin (human, bovine, mouse, rabbit), *Plasmodium falciparum* merozoite surface antigen, tumour necrosis factor (TNF) receptor precursor and low density lipoprotein (LDL) receptor precursor. However, none of the database sequences

which were identified using the PROSRCH database searching program was considered to be homologous to P786.

(N.B. Database searches using P786 as a query to search the NBRF protein sequence database were also carried out by S. J. McQuay, Biocomputing Research Unit, University of Edinburgh. These searches did not reveal any sequences of interest which were not identified by the OWL database searches.)

5.5.3 Profile analysis.

Another method which was used to search for protein sequences with similarities to P786, is profile analysis (Gribskov *et al* 1990). This is a sequence comparison method for finding and aligning distantly related proteins. Again, it uses the symbol comparison tables of Dayhoff *et al* (1978), and in addition, an optimal alignment of a group of similar protein sequences. Construction of the 'Profile' involves first aligning the sequences, then representing all the information in the multiple alignment as a table of position-specific symbol comparison values and gap weights. Every possible residue has a value in each row of the profile. The comparison of a residue to any row of the profile gives a specific value, or 'profile comparison value'. The best alignments of a sequence to a profile are found by aligning the sequence with the profile, such that the sum of the profile comparison values minus the sum of the gap weights is maximal. The position-specific gap weights penalize gaps more heavily in conserved regions than in more

variable regions.

It was decided to use profile analysis to search protein sequence databases for sequences which may be homologous or analogous to repeats found in P786. Searching with a profile ensures that all the information in each of the aligned protein sequences is represented. This is not the case when searching with a consensus sequence. Another advantage is that the position-specific comparison values and gap weights ensure that database sequences which are similar to the more conserved regions of the profile score more highly than those which are similar to the more variable regions. Profiles were constructed using the two sets of repeats (A and B) shown in Fig 5.8. Each of these profiles was used to search the NBRF protein sequence database using GCG.PROFILESEARCH.

The 10 highest scoring alignments identified by searching with the profile constructed from repeats A included:- two finger proteins (human); mouse complement factor H-related proteins A2, B1 and C1; a scorpion neurotoxin; *Thermus aquaticus* ferredoxin; gene nrdC protein and thioredoxin from phage T4; and a hypothetical mitochondrion protein from *Paramecium tetraurelia*. These alignments had scores between 11.89 and 15.53, and had between 6 and 10 identities with the profile consensus sequence.

The 10 highest scoring alignments identified by searching with the profile constructed using repeats B included:- six finger proteins (five human, one mouse); *Streptomyces griseus*

ferredoxin; rat trypstatin; a fowlpox virus protein; and mouse developmental control protein Krox-9. These alignments scored between 11.50 and 14.13, and had between 4 and 7 identities with the profile consensus sequence.

None of the above mentioned proteins identified by profile analysis was considered to be homologous with P786.

5.5.4 Identification of proteins sequences containing the CPXG (7X) C motif.

The P786 amino acid sequence contains 16 copies of the CPXG (7X) C motif. Twelve of these are found in repeats 'A', and a further three in repeats 'B' (Fig 5.8). The frequency of occurrence of this motif, and its presence in repeat units, may indicate that it plays some important structural or functional role. If this is the case, identification of other proteins with many copies of the same motif may give some indication of its possible significance.

GCG.FINDPATTERNS was used to search the NBRF protein sequence database for proteins containing CPXG (7X) C. Only five proteins had more than one copy of the motif. These were:- leech antistasin, human von Willebrand factor and human low density lipoprotein receptor-related protein (each with two copies); midge balbiani ring 3 protein and human fibrillin (each with three copies).

Perhaps the most interesting of these is human fibrillin. Although the CPXG (7X) C motif is present only three times, there are nine copies of CPXG (9X) C, and one of CPXG (8X) C. These occur in larger repeat units, which contain many cysteine residues whose positions are highly conserved. In this respect the arrangement of repeats in human fibrillin is similar to that seen in P786. Further investigation into the structure of these two proteins, and identification of other proteins with many copies of CPXG (7-9X) C may reveal some structural or functional importance for this motif.

CHAPTER 6. DISCUSSION.

6.1 Introduction.

Up until 1988, when the present study commenced, there was no published work on the molecular biology or genetics of *Cryptosporidium*. This reflected a lack of interest in the parasite until the mid 70's, when it was recognised as a cause of diarrhoea in both humans and animals. In addition, technical difficulties with producing sufficient parasite material for perparation of DNA probably also hindered progress in this field.

The aim of this project was to identify genes encoding antigens of *C. parvum*. Other studies have isolated *C. parvum* protein-encoding genes by screening genomic DNA libraries with heterologous DNA probes. For example, Kim *et al* (1992), identified a *C. parvum* actin gene by screening with chicken β-actin cDNA, and Nelson *et al* (1991), identified *C. parvum* tubulin genes by screening with oligonucleotide probes corresponding to conserved α and β tubulin sequences. *C. parvum* genes encoding thymidylate synthase-dihydrofolate reductase, and topoisomerase have been identified by PCR using conserved oligonucleotide primers, followed by hybridization (Gooze *et al* 1991, Dykstra *et al* 1991).

Anti-*Cryptosporidium* antiserum has also been used by other workers to identify *C. parvum* antigen-encoding genes (Dykstra *et*

al 1991, Petersen *et al* 1992). The latter study was particularly aimed at identifying apical complex and pellicle antigens of the sporozoite and/or merozoite, which may be targets for neutralizing antibody. Five clones expressing such antigens were isolated, and are being further characterized.

The present study used anti-*C. parvum* antiserum to isolate two clones which express *C. parvum* antigens. The DNA insert from one of these clones (CPR1), and its expressed peptide (P786) have been characterized in some detail, and the entire DNA sequence of the insert has been obtained. Work presented in this thesis indicates that this DNA fragment represents part of a gene encoding a 190kDa oocyst wall protein of *C. parvum*. The deduced amino acid sequence encoded by this gene fragment has an unusual composition, with high proportions of cysteine, proline, histidine and glutamine. The sequence is also markedly repetitive, with repeat units based on conserved cysteine residues.

6.2 Oocyst wall proteins of *C. parvum*.

The *Cryptosporidium* oocyst wall consists of two layers which are formed by the fusion of wall forming bodies with three membranes (Current & Reese 1986). Ultrastructural studies of oocyst wall formation in other coccidia indicate that a similar process occurs in *Eimeria* spp. (Wang 1982), *Sarcocystis* spp. (Vetterling *et al* 1973) and *Toxoplasma gondii* (Ferguson *et al* 1975). Like other coccidia, the oocyst stage of *C. parvum* is

highly resistant to adverse conditions, including exposure to many commonly used disinfectants (Blewett 1989b). This resistance is due to the protection afforded by the oocyst wall, since once released from the oocyst the sporozoites are easily destroyed.

Little is known of the biochemical composition of coccidial oocyst walls. Different workers have come to different conclusions about their composition. Strout *et al* (1963) concluded that lipid accounts for one third of the coccidial oocyst wall, while Landers (1960) concluded that the wall consisted mainly of polysaccharide and protein.

Birefringence studies were carried out by Monne and Honnig (1954) in an attempt to determine the composition of the oocyst walls of several species of *Eimeria* and *Isospora*. They concluded that the coccidial oocyst wall consisted of two layers, with the inner layer (endocyst) consisting of a lipid-protein matrix, and the outer layer (ectocyst) of quinone-tanned protein.

Ryley (1980) reported the presence of an outer ectocyst, which was removed by treatment with sodium hypochlorite, and comprised approximately 20% of the dry weight of the oocyst wall. Analysis of the inner layer (endocyst) indicated that it contained approximately 70% protein, 30% lipid and 1.5% carbohydrate. However, the presence of an outer sodium hypochlorite soluble layer is in doubt. Stotish *et al* (1978), using electron microscopy, observed no differences between *E. tenella* oocysts

treated with 5% sodium hypochlorite, and those which had not been exposed to the chemical. They suggested that the so-called ectocyst, referred to by Monne and Honnig (1954) and Ryley (1980) was actually debris from inadequate purification, as first suggested by Nyberg and Knapp (1970). Stotish *et al* (1978) also analysed the composition of *E. tenella* oocysts. Their results indicated that the wall consisted of 67% protein, 14% lipid and 19% carbohydrate. They put forward a model of the oocyst wall which consisted of an outer 10nm thick layer of lipid, and an inner 90nm thick layer consisting of disulphide-linked glycoprotein.

Work by Jolley *et al* (1976, 1979) also indicated that the oocyst walls of *E. tenella* and *E. steidae* contained many disulphide bonds, and in addition, that reduction of these bonds to sulphydryl groups occured during the excystation process. If the oocyst wall of *C. parvum* has a similar structure to those of *Eimeria* species then the protein corresponding to P786 may form part of a disulphide-linked inner glycoprotein layer. P786 contains many cysteine residues. These residues may participate in disulphide bonding and have a role in maintaining the integrity of the oocyst wall, as is believed to be the case in *Eimeria* species (Jolley *et al* 1976, 1979).

Two previous studies have looked at the protein constituents of the *C. parvum* oocyst wall by SDS-PAGE analysis of oocysts, following surface labelling of the oocysts with ^{125}I (Lumb *et al* 1988a, Tilley *et al* 1990b). The former study identified 6-10

^{125}I labelled bands, while the latter identified 17. Differences between oocyst wall bands identified in these studies, and in the present study may be due in part to the effects of trypsin and deoxycholate on oocyst wall proteins during *in vitro*-excystation. None of the bands identified by ^{125}I surface labelling had a mw of 190kDa, which could indicate that the *C. parvum* protein corresponding to P786 was not actually exposed on the surface of the oocyst wall. Lumb *et al* (1988a) also carried out SDS-PAGE analysis of oocyst shells which had been separated from sporozoites following *in vitro*-excystation. 21 bands were observed on Coomassie blue stained gels, one of which had a mw of 190kDa, and therefore might be the same protein as the 190kDa antigen corresponding to P786. Differences in other bands observed may have resulted from differences in the method of extraction of the oocysts from faeces. In our laboratory, routine oocyst extraction involves exposure of the oocysts to dilute sulphuric acid (0.02%) and SDS (1%). It is also possible that there are isolate-specific differences in oocyst wall proteins. Thus, Lumb *et al* (1988a) found that of 10 ^{125}I labelled oocyst surface proteins of *C. parvum*, four were not present in all of the isolates studied.

Searching of the DNA and protein sequence databases did not identify any sequences which were considered to be homologous to CPR1 or its encoded peptide P786. This is perhaps as might have been expected, since there are no gene or protein sequences of oocyst wall proteins of enteric coccidia in these databases. It would be of interest to determine whether P786 or its DNA

sequence has homologues in related species of coccidia. One might expect some of the oocyst wall proteins of the coccidia to be conserved, since the oocyst walls of all species have at least one important function in common, i.e. protection of the oocyst contents, and furthermore, all have to withstand similar environmental conditions.

The Southern blotting experiments described in Chapter 5 attempted to find out whether *T. gondii* or *Sarcocystis* spp. contained DNA sequences which were homologous with CPR1. No hybridization of CPR1 to the genomic DNA of *T. gondii* or *Sarcocystis* spp. was detected, even when hybridization and washing was carried out at low stringency which would have allowed hybrids to form between sequences containing up to approximately 35% mismatching. It would have been of interest to repeat this experiment using genomic DNA from *Eimeria* spp. Immunoblotting experiments suggest that *C. parvum* oocysts share common antigens with several *Eimeria* species of sheep (Ortega-Mora *et al* 1992).

6.3 G+C content and codon usage in *C. parvum*.

There is considerable variation in the genomic G+C contents of different unicellular organisms, and it is thought that this is due to a species-specific mutation pressure which has a directionality toward higher or lower G+C content (Sueoka 1988). The observation that organisms with higher or lower values of genomic G+C content have more extreme values of G+C content at

the third codon position (Muto & Osawa 1987, Sharp & Devine 1989, Saul & Battistutta 1988) can be explained in terms of directional mutation pressure. Thus, since most mutations in the third codon position do not result in a change in the amino acid encoded, directional mutation in favour of a high or low G+C content, results in a bias towards the use of codons ending in G and C, or A and T respectively.

Values for the G+C content of various *C. parvum* DNA sequences, including data from the present work, are shown in table 6.1. The relatively low G+C contents of these DNA sequences (ranging from 27.2% to 39.1%), plus the relatively lower G+C contents at the third codon position of coding regions (17.9% versus 39.1% (this study), and 28.5% versus 38.5% (Kim *et al* (1992)) are consistent with a model of the *C. parvum* genome in which directional mutation pressure tends towards decreased G+C content, resulting in a bias towards the use of codons ending in A or T.

Directional mutation pressure is not, however, the only factor affecting synonymous codon usage. Translational selection, i.e. selection for codons corresponding to the most abundant and/or efficiently translated tRNAs, also influences codon usage (Ikemura 1985). Codon usage patterns in a given organism appear to result from a balance between translational selection and mutation pressure, with highly expressed genes being more influenced by translational selection, and low or moderately expressed genes reflecting the influence of directional mutation

Table 6.1 Percentage G+C contents of various *C. parvum* DNA sequences.

DNA sequence	Reference	% G+C
CPR1	This work	39.1
CPR1 (3rd codon position)	This work	17.9
Actin gene (protein coding region)	Kim *et al* (1991)	38.5
Actin gene (flanking regions)	Kim *et al* (1991)	27.2
Actin gene (3rd codon position)	Kim *et al* (1991)	28.5
Sequence used for PCR study	Laxer *et al* (1991)	35
Clones picked at random from *C. parvum* DNA library	Dykstra *et al* (1991)	32.6

pressure (Sharp & Devine 1989, Bulmer 1988, Sharp & Li 1986, Shields & Sharp 1987).

Examination of codon usage by the *C. parvum* actin gene and by CPR1 (Table 6.2) can give us some indication of the codon usage patterns in *C. parvum*. As already stated above, codon usage in both these genes shows an overall bias towards the use of codons ending in A or T. However, there are additional biases in the codon usage by the actin gene, that are not evident in codon usage by CPR1, and these are probably the result of translational selection. Evidence of translational selection in the actin gene includes i) preferential use of codons ending in C and G to encode phenylalanine and lysine respectively; ii) the use of GGT to encode 26 of 30 glycine residues, the remaining four being encoded by GGA (3) and GGC (1); iii) the preferential use of ATC rather than ATA (9:1) to encode the 10 isoleucine residues that are not encoded by ATT (16). On the basis of the data in table 6.2, possible optimal codons for *C. parvum* appear to be TTC (phenylalanine), AAG (lysine) and GGT (glycine). However, additional sequence data will be required to confirm these and to identify optimal codons for the other amino acids.

It is not surprising that the actin gene appears to be subject to translational selection, since it is likely to be highly expressed. We do not know whether the protein encoded by CPR1 is highly expressed. As an oocyst wall protein it is only likely to be expressed by the macrogamete and/or the developing zygote

Table 6.2 Codon usage by CPR1 and by the *C. parvum* actin gene.

Codon usage by the *C. parvum* actin gene (bold) is shown alongside codon usage by CPR1.

3	10	UUU	Phe	**13**	23	UCU	Ser	**10**	21	UAU	Tyr	**4**	36	UGU	Cys
9	4	UUC	Phe	**1**	4	UCC	Ser	**5**	5	UAC	Tyr	**2**	21	UGC	Cys
13	14	UUA	Leu	**5**	17	UCA	Ser	**1**	0	UAA	***	**0**	0	UGA	***
7	5	UUG	Leu	**1**	5	UCG	Ser	**0**	0	UAG	***	**4**	0	UGG	Trp
2	10	CUU	Leu	**4**	24	CCU	Pro	**6**	27	CAU	His	**7**	1	CGU	Arg
3	1	CUC	Leu	**0**	4	CCC	Pro	**3**	5	CAC	His	**1**	0	CGC	Arg
3	6	CUA	Leu	**16**	57	CCA	Pro	**9**	50	CAA	Gln	**0**	0	CGA	Arg
0	1	CUG	Leu	**0**	2	CCG	Pro	**1**	8	CAG	Gln	**0**	0	CGG	Arg
16	29	AUU	Ile	**12**	21	ACU	Thr	**6**	24	AAU	Asn	**3**	3	AGU	Ser
9	6	AUC	Ile	**2**	7	ACC	Thr	**2**	8	AAC	Asn	**1**	2	AGC	Ser
1	10	AUA	Ile	**13**	30	ACA	Thr	**8**	35	AAA	Lys	**10**	19	AGA	Arg
10	4	AUG	Met	**0**	0	ACG	Thr	**14**	4	AAG	Lys	**2**	2	AGG	Arg
15	33	GUU	Val	**12**	19	GCU	Ala	**14**	21	GAU	Asp	**26**	15	GGU	Gly
3	5	GUC	Val	**2**	2	GCC	Ala	**6**	10	GAC	Asp	**1**	3	GGC	Gly
10	20	GUA	Val	**6**	12	GCA	Ala	**24**	31	GAA	Glu	**3**	27	GGA	Gly
0	1	GUG	Val	**0**	3	GCG	Ala	**10**	14	GAG	Glu	**0**	5	GGG	Gly

during oocyst formation. As a result the influence of translational selection on codon usage by this gene will be limited to these stages of the parasite lifecycle.

6.4 P786 and parasite repeat antigens.

Perhaps the most striking feature of the peptide sequence encoded by CPR1 is the presence of amino acid repeats. It is of interest therefore to consider the many amino acid repeats found in proteins of other protozoan parasites and their possible functions.

Proteins with repetitive amino acid sequences appear to be very common among protozoan parasite antigens, especially those of *Plasmodium* (Schofield 1991, Kemp 1987). Such repeats can be found in proteins from different developmental stages, and with different functions. For example, *Plasmodium* antigens containing amino acid repeats include the soluble S-antigens released during schizogony (Cowman *et al* 1985), sporozoite and merozoite surface antigens (McCutchan *et al* 1985, Dame *et al* 1984, Holder *et al* 1985), erythrocyte surface antigens (Favaloro *et al* 1986, Triglia *et al* 1987), rhoptry antigens (Keen *et al* 1990)and even a phylogenetically conserved heat shock protein (Yang *et al* 1987). Amino acid repeats have also been found in antigens of other protozoan parasites including *Leishmania* (Wallis & McMaster 1987), Trypanosoma cruzi (Hoft *et al* 1989, Ibanez *et al* 1988) and *Eimeria acervulina* (Jenkins 1988). The repeat

structures differ in size, number, and composition among different proteins, and in addition, allelic variants of a protein may show different repeats within otherwise conserved amino acid sequences; this occurs in the *Plasmodium cynomolgi* CS protein (Galinski *et al* 1987) and in the *Plasmodium falciparum* S-antigen (Saint *et al* 1987).

Despite the great variation in repeat units, the observation that a number of common features are shared by most of these repeat regions has prompted attempts to formulate a universal theory to explain their occurrence.

The first feature in common is of course repetitiveness. Secondly, there is a marked bias in amino acid composition of the repeat regions. Amino acids represented by A, D, E, G, N, P, Q, S and V are relatively common in repeat units. Those represented by C, F, I, L, M, W, and Y are rarely present (Schofield 1991). The third feature is the immunodominance of the repeat regions. Natural antibody responses to the CS protein, S-antigens, ring-infected erythrocyte surface antigen (RESA) and falciparum interspersed repeat antigen (FIRA) are directed predominantly against the repeat regions (Zavala *et al* 1985, Anders *et al* 1986, Cowman *et al* 1984).

It has been suggested that the repeat units may act as ligands for host structures such as red blood cells, hepactocytes and other host cells (Godson *et al* 1983, Nussenzweig & Nussenzweig 1989). Repeated binding units would presumably allow high

aviditiy interaction between parasite and host cell receptor. This hypothesis does not, however, explain some of the features of these repeat regions. For example, why should the amino acid composition of the repeats be biased towards certain amino acids? Furthermore, and perhaps more importantly, why do allelic variants of a protein possess completely different repeat structures? If such repeats were involved in receptor-ligand interaction we would expect their amino acid sequences to be constrained by selection acting on the protein sequence.

Another theory which attempts to explain the presence of repeats is that they function as an immune evasion mechanism. As already mentioned above, the repeat regions of *Plasmodium* antigens are highly immunodominant. In addition, antibodies recognising the repeat regions of *Plasmodium* antigens are extensively cross reactive, both with other repeats within the same antigen, and with repeats found in other proteins (Anders 1986). It has been proposed that the existence of such a network of cross-reacting epitopes induces proliferation of a higher than normal proportion of somatic mutants, resulting in hypergammaglobulinaemia, and inhibiting affinity maturation of the immune response to protective epitopes (Anders 1986). Another way in which repeat epitopes may participate in immune evasion has been put forward by Schofield (1991). He suggests that repeat epitopes within an antigen crosslink antibody on the surface of B cells, resulting in T-independent activation of B cells. A T-independent response is generally considered to be

inferior to a T-dependent response, usually lacking in class switching and affinity maturation, and having poor T and B cell memory (Roitt *et al* 1985).

Another feature of certain repeat antigens of *Plasmodium*, which appears to support the theory of immune evasion, is their unusual genetics and evolutionary history. Evidence suggests that the evolution and maintenance of the repeats in the circumsporozoite protein must be through a mechanism acting at the DNA level, which eliminates or spreads mutations, with the result that the repeats either remain conserved or evolve in large jumps (Galinski *et al* 1987). Existence of such a mechanism explains why in some cases, circumsporozoite protein repeats from different strains are more conserved at the DNA than at the protein level. Additional support for the view that repeats are maintained at the DNA level comes from a study of *Plasmodium falciparum* S-antigens (Saint *et al* 1987). S-antigens are secreted into the parasitophorous vacuole, and are released into the plasma on shizont rupture. There is a large degree of serological diversity among S-antigens of different strains of *P. falciparum*. One source of this diversity appears to be a variation in the reading frame. It has been demonstrated that two serologically distinct S-antigens expressed by different *P. falciparum* strains result from translation of almost identical DNA sequences, but in different reading frames (Saint *et al* 1987). So again, as with the circumsporozoite protein, the repeats are maintained at the DNA level. These observations are consistent with a mechanism of immune evasion in which the

repeat structures divert the immune response as a result of their immunodominance, cross reaction, and ability to act as B cell epitopes. In this context, the actual amino acid sequence of the repeats seems to be unimportant, provided that it has these properties.

However, the view that all parasite antigen repeats can be explained in terms of an immune evasion strategy is perhaps rather too simplistic. Ridley (1991) points out that many *Plasmodium* genes have been isolated by screening DNA expression libraries with sera from immune individuals. This approach has led to the selective identification of genes encoding proteins with immunodominant epitopes, many of which contain repetitive regions. He suggests that although some repeat antigens may protect the parasite from immune attack, they may also have a variety of functions which are not yet appreciated. For example, interaction with macromolecules such as cytoskeletal proteins and receptor molecules. In support of this view, a repeat epitope protein has been identified in *Trypanosoma brucei rhodesiense*, that is believed to have a specific cytoskeletal function (Duncan *et al* 1991). Such repeat epitopes are unlikely to serve as an immune evasion mechanism in African trypanosomes, since these parasites have evolved a mechanism of antigenic variation which is extremely successful in evading the host immune response.

It seems unlikely that the repeats in P786 participate in immune evasion. They do not share several of the characteristics which

are common to parasite antigen repeats that appear to be involved in immune evasion. For example, while the P786 repeat units contain certain residues whose positions are completely (cysteine) or highly (proline, glycine, leucine) conserved, other intervening residues are much more variable. Repeat units in other parasite repeat antigens tend to consist of many tandem copies of more or less identical repeats. Secondly, P786 contains a high proportion of cysteine residues which are particularly concentrated in the repetitive regions. By contrast, repeats thought to be involved in immune evasion rarely contain cysteine residues (Schofield 1991, Kemp *et al* 1987). Finally, the P786-β-galactosidase fusion protein was not recognised by serum from any of 16 lambs which had been infected *per os* with *C. parvum* oocysts, whereas parasite repeat antigens which appear to be involved in immune evasion elicit high levels of antibody in infected individuals (Schofield 1991, Anders 1986). As a component of the oocyst wall, the native 190kDa *C. parvum* antigen corresponding to P786 is probably not well presented to the immune system during infection, since the oocyst stage is not invasive.

6.5 Cysteine-rich repeats and P786.

Cysteine-rich repeats have been identified in several plasma proteins, including thrombospondin (Lawler & Hynes 1986), von Willebrand factor (Verweij *et al* 1986) and many components of the complement system (Reid & Day 1989). They have also been demonstrated in the extracellular domains of a number of cell

surface proteins, such as nerve growth factor (NGF) receptor (Johnson *et al* 1986), epidermal growth factor (EGF) receptor (Ullrich *et al* 1984) and low density lipoprotein (LDL) receptor (Yamamoto *et al* 1984), and in human fibrillin, a glycoprotein component of connective tissue (Maslen et al 1991).

The common feature of all these proteins appears to be the extracellular location of the cysteine-rich domains. Being highly reactive, cysteine residues readily form disulphide bonds in the oxidizing conditions of the extracellular environment (Muskal *et al* 1990). Such bonds may be important in maintaining the structure of the protein. Indeed, several of the above-mentioned proteins are known to adopt a structure which is stabilized by disulphide bonding. For example, human von Willebrand factor forms multimers which are linked by disulphide bonds (Verweij *et al* 1986). Furthermore, the binding sites of LDL receptor and NGF receptor both contain disulphide-bonded cysteine residues which are essential for maintaining the structure of these sites (Yamamoto *et al* 1984, Yan & Chao 1991).

The mature *C. parvum* protein corresponding to P786 may consist of structurally repeated regions stabilized by the presence of many disulphide bonds. In support of this, it has been found that a spore coat protein of the slime mould *Dictyostelium discoideum* contains EGF-like repeats in which the positions of the cysteine residues are highly conserved (Widdowson *et al* 1990). These residues are highly disulphide-crosslinked in the mature spore coat. The spore coat of *D. discoideum* can be

considered analogous to the oocyst wall of *C. parvum*, since each forms the outer covering of an environmentally resistant dispersal stage of the organism. The presence of proteins containing repeats based on conserved cysteine residues in both of these structures indicates that the cysteine residues are likely to perform a similar function in both cases. Further investigation into the structure of the protein components of the *C. parvum* oocyst wall may reveal important information regarding the ability of the oocyst to withstand exposure to environmental conditions and its resistance to many chemical disinfectants.

6.6 Concluding remarks.

This study has identified two antigen-encoding gene fragments of *C. parvum*. One of these, CPS10, encodes an antigen which is recognised by serum from some, but not all, lambs infected with *C. parvum*. Due to limitations of time, extensive characterization of CPS10 and its encoded peptide was not carried out. Possible future work on this gene fragment would involve identifying the location of the encoded native protein in the parasite, by Western blotting and immunofluoresence. If the antigen was found to be located on the sporozoite surface, or in the apical complex, it would be of particular interest to determine whether antibodies raised against this antigen were protective. Antibodies recognising sporozoite and merozoite surface antigens of *C. parvum* have been found to reduce infection in both *in vivo* and *in vitro* assays (sections 1.7 and

1.8).

The other gene fragment, CPR1, encodes part of a 190kDa oocyst wall protein. Sequencing of this gene fragment revealed that the encoded peptide sequence is markedly repetitive, with repeat units based on conserved cysteine residues. It is likely that the cysteine residues participate in disulphide-crosslinking, which may be of structural importance in the oocyst wall. The presence of disuphide-crosslinking in the 190kDa *C. parvum* could be demonstrated by comparing Western blots of oocyst shell samples which have been electrophoresed with and without β-mercaptoethanol in the sample buffer. In addition, further work on CPR1 would include isolation of the rest of the gene by screening DNA libraries with CPR1 as a hybridization probe.

Preliminary Southern blotting experiments (section 4.7) indicated that CPR1 did not hybridize with DNA from *T. gondii* or *Sarcocystis* spp. An extension of this work would involve carrying out similar experiments using DNA from *Eimeria* species and from different species of *Cryptosporidium* to determine whether these contain sequences which are homologous to CPR1.

Part of the CPR1 DNA sequence may prove to be a suitable target for use in a PCR detection technique for *C. parvum*. Using primers 267 and 594 (table 5.1), a 980bp DNA fragment was amplified from pBSCPR1[1] DNA, and from *C. parvum* genomic DNA (N. C. Lally & G. D. Baird unpublished). Further

experiments would be required to determine the sensitivity and specificity of amplification, and to evaluate the suitability of such a technique as a detection method for *C. parvum*.

REFERENCES.

ANDERS, R. F. (1986) Multiple cross-reactivities amongst antigens of *Plasmodium falciparum* impair the development of protective immunity against malaria. Parasite Immunology 8: 529-539.

ANDERSON, B. C. (1982) Cryptosporidiosis: a review. Journal of the American Veterinary Medical Association 180: 1455-1457.

ANGUS K. W. (1989) Mammalian cryptosporidiosis: a veterinary perspective. In: Cryptosporidiosis: Proceedings of the First International Workshop. (Angus, K. W. & Blewett, D. A. eds.), pp 43-53.

ANGUS, K. W. (1990) Cryptosporidiosis is ruminants. In: Cryptosporidiosis of man and animals. (Dubey, J. P., Fayer, R. & Speer, C. A. eds.) pp 83-103. CRC Press.

ANGUS, K. W., TZIPORI, S., & GRAY, E. W. (1982) Intestinal lesions in specific-pathogen-free lambs associated with a *Cryptosporidium* from calves with diarrhoea. Veterinary Pathology 19: 67-78.

ANGUS, K. W., HUTCHISON, G., CAMPBELL, I. & SNODGRASS, D. R. (1984) Prophylactic effects of anticoccidial drugs in experimental murine cryptosporidiosis. The Veterinary Record 114: 66-68.

ANGUS, K. W., HUTCHISON, G. & BLEWETT, D. A. (1989) Spiramycin trials in animals. Abstract, in: Cryptosporidiosis: Proceedings of the First International Workshop. (Angus, K. W. & Blewett, D. A. eds.), p117.

ARGENZIO, R. A., LIACOS, J. A., LEVY, M. L., MEUTEN, D. J., LECCE, J. G. & POWELL, D. W. (1990) Villous atrophy, crypt hyperplasia, cellular infiltration, and impaired glucose-NA absorption in enteric cryptosporidiosis of pigs. Gastroenterology 98: 1129-1140.

ARROWOOD, M. J., MEAD, J. R., MAHRT, J. L. & STERLING, C. R. (1989) Effects of immune colostrum and orally administered antisporozoite monoclonal antibodies on the outcome of *Cryptosporidium parvum* infection in neonatal mice. Infection and Immunity 57: 2283-2288.

BARTA, R. (1989) Phylogenetic analysis of the class Sporozoea (phylum Apicomplexa Levine, 1970): edidence for the independent evolution of heteroxenous lifecycles. The Journal of Parasitology 75: 195-206.

BIRD, F. G. & SMITH, M. D. (1980) Cryptosporidiosis in man: parasite lifecycle and fine structural pathology. Journal of Pathology 132: 217-233.

BIRNBOIM, H. C. & DOLY, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research 7: 1513-1523.

BJORNEBY, J. M., HUNSAKER, B. D., RIGGS, M. W. & PERRYMAN, L. E. (1991) Monoclonal antibody immunotherapy in nude mice persistently infected with *Cryptosporidium parvum*. Infection and Immunity 59: 1172-1176.

BJORNEBY, J. M., RIGGS, M. W. & PERRYMAN, L. E. (1990) *Cryptosporidium parvum* merozoites share neutralization-sensitive epitopes with sporozoites. The Journal of Immunology 145: 298-304.

BLEWETT, D. A. (1988) Cryptosporidiosis in red deer. In: The Management and Health of Farmed Deer. (Reid, H. W. ed.), pp 43-47, Kluwer Academic Publishers, London.

BLEWETT, D. A. (1989a) Disinfection and oocysts. In: Cryptosporidiosis: Proceedings of the First International Workshop. (Angus, K. W. & Blewett, D. A. eds.), pp 107-115.

BLEWETT, D. A. (1989b) Quantitative techniques in *Cryptosporidium* research. In: Cryptosporidiosis: Proceedings of the First International Workshop. (Angus, K. W. & Blewett, D. A. eds.), pp 85-95.

BOGAERTS, J., LEPAGE, P., ROUVRAY, D. & VANDEPITTE, J. (1984) *Cryptosporidium* spp., a frequent cause of diarrhoea in central Africa. Journal of Clinical Microbiology 20: 874-876.

BOLIVAR, F., RODRIGUEZ, R. L., GREENE, P. J., BETLACH, M. C., HEYNECKER, H. L., BOYER, H. W., CROSA, F. H. & FALKOW, S. (1977) Construction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene 2: 95-113.

BONNIN, A., DUBREMETZ, J. F. & CAMERLYNCK, P. (1991) Characterization of microneme antigens of *Cryptosporidium parvum* (Protozoa, Apicomplexa) Infection and Immunity 59: 1703-1708.

BRASSEUR, P., LEMETEIL, D. & BALLET, J. J. (1988) Rat model for human cryptosporidiosis. Journal of Clinical Microbiology 26: 1037-1039.

BULMER, M. (1988) Are codon usage patterns in unicellular organisms determined by selection-mutation balance ? Journal of Evolutionary Biology 1: 15-16.

CAMPBELL, P. N. & CURRENT, W. L. (1983) Demonstration of serum antibodies to *Cryptosporidium* sp. in normal and immunodeficient humans with confirmed infections. Journal of Clinical Microbiology 18: 165-169.

CARROLL, S. B. & LAUGHON, A. (1987) Production and purification of polyclonal antibodies to the foreign segment of β-galactosidase fusion proteins. In: DNA Cloning (Glover, D. M. ed.), vol. III, pp 89-111, IRL Press, Oxford.

CASEMORE, D. P. (1987) The antibody reponse to *Cryptosporidium*: development of a serological test and its use in a study of immunologically normal persons. Journal of Infection 14: 125-134.

CASEMORE, D. P. (1990) Epidemiological aspects of human cryptosporidiosis. Epidemiology and Infection 104: 1-28.

CHUNG, C. T., NIEMELA, S. L. & MILLER, R. H. (1989) One step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences USA 86: 2172-2175.

CLARKE, L. E., MESSER, L. I., GREENWOOD, N. M. & WISHER, M. H. (1987) Isolation of λamp3 genomic recombinants coding for antigens of *Eimeria tenella*. Molecular and Biochemical Parasitology 22: 79-87.

CLARKE, L. E., TOMLEY, F. M., WISHER, M. H., FOULDS, I. J. & BOURSNELL, M. E. G. (1990) Regions of an *Eimeria tenella* antigen contain sequences which are conserved in circumsporozoite proteins from *Plasmodium* spp. and which are related to the thrombospondin gene family. Molecular and Biochemical Parasitology 41: 269-280.

COHEN, S. N., CHANG, A. C. Y. & HSU, L. (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proceedings of the National Academy of Sciences 69: 2110-2114.

CORNELISSEN, A. W. C. A., OVERDULVE, V. P. & VAN DER PLOEG, M. (1984) Determination of nuclear DNA of five eucoccidian parasites, *Isospora* (*Toxoplasma*) *gondii*, *Sarcocystis cruzi*, *Eimeria tenella*, *E. acervulina*, and *Plasmodium berghei*, with special reference to gamontogenesis and meiosis in *I.* (*T*) *gondii*. Parasitology 88: 531-553.

COULSON, A. F. W., COLLINS, J. F. & LYALL, A. (1987) Protein and nucleic acid sequence database searching: a suitable case for parallel processing. The Computer Journal 30: 420-424.

COWMAN, A. F., COPPE., R. F., SAINT, R. B., FAVALORO, J., CREWTHER, P. E., STAHL, H-D., BIANCO, A. E., BROWN, G. V., ANDERS, R. F. & KEMP, D. J. (1984) The ring-infected erythrocyte surface antigen (RESA) polypeptide of *Plasmodium falciparum* contains two separate blocks of tandem repeats encoding antigenic epitopes that are naturally immunogenic in man. Molecular Biology and Medicine 2: 207-221.

COWMAN, A. F., SAINT, R. B., COPPEL, R. L., BROWN, G. V., ANDERS, R. F., & KEMP, D. J. (1985) Conserved sequences flank variable tandem repeats in two S-antigen genes of *Plasmodium falciparum*. Cell 40: 775-783.

CRAWFORD, F. G. & VERMUND, S. H. (1988) Human Cryptosporidiosis. CRC Critical Reviews in Microbiology 16: 113-159.

CRUZ, J. R., CANO, F., CACERES, P., CHEW, F. & PAREJA, G. (1988) Infection and diarrhoea caused by *Cryptosporidium* sp. among Guatemalan infants. Journal of Clinical Microbiology 26: 88-91.

CURRENT, W. L. & GARCIA, L. S. (1991) Cryptosporidiosis. Clinical Microbiology Reviews 4: 325-358.

CURRENT, W. L. & HAYNES, T. B. (1984) Complete development of *Cryptosporidium* in cell culture. Science 224: 603-605.

CURRENT, W. L. & LONG, P. L. (1983) Development of human and calf *Cryptosporidium* in chicken embryos. The Journal of Infectious Diseases 148: 1108-1113.

CURRENT, W. L. & REESE, N. C. (1986) A comparison of endogenous development of three isolates of *Cryptosporidium* in suckling mice. Journal of Protozoology 33: 89-108.

CURRENT, W. L., UPTON, S. J. & HAYNES, T. B. (1986) The lifecycle of *Cryptosporidium baileyi* n. sp. (Apicomplexa, Cryptosporidiidae) infecting chickens. Journal of Protozoology 33: 289-296.

DAME, J. B., WILLIAMS, J. L., MCCUTCHAN, T. F., WEBER, J. L., WIRTZ, R. A., HOCKMEYER, W. T., MALOY, W. L., HAYNES, J. D., SCHNEIDER, I., ROBERTS, D., SANDERS, G. S., REDDY, E. P., DIGGS, C. L. & MILLER, L. H. (1984) Structure of the gene encoding the immunodominant surface antigen on the sporozoite of teh human malaria parasite *Plasmodium falciparum*. Science 225: 593-599.

DAYHOFF, M. O., SCHARWTZ, R. M. & ORCUTT, B. C. (1978) in: Atlas of protein sequence and structure. Volume 5, Supplement 3. (Dayhoff, M. O. ed.), National Biomedical Research Foundation, Washington DC.

DEMOL, P., MUKASHUMA, S., BOGAERTS, J., HEMELHOF, W. & BUTZLER, J. P. (1984) *Cryptosporidium* related to measles diarrhoea in Rwanda. Lancet 2: 42-43.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Research 12: 387-395.

DE VISSER, N. A. P. C., BREUKINK, H. J., VAN ZIJDERVALD, F. G. & DE LEEUW, P. W. (1987) Enteric infections in veal calves: a longitudinal survey on four veal calf units. Veterinary Quarterly 9: 289-296.

DONELSON, J. E., SUKE, B. O. L., MOSER, D ., ZENG, W., ERONDU, N.E., LUCIUS, R., RENZ, A., KARAM, M. & FLORES, G. Z. (1988) Construction of *Onchocerca volvulus* cDNA libraries and partial characterization of the cDNA for a major antigen. Molecular and Biochemical Parasitology 31: 241-250.

DUNCAN, L. R., GAY, L. S. & DONELSON, J. E. (1991) African trypanosomes express an immunogenic protein with a repeating epitope of 24 amino acids. Molecular and Biochemical Parasitology 48: 11-16.

DYKSTRA, C. C., BLAGBURN, B. L. & TIDWELL, R. R. (1991) Construction of genomic libraries of *Cryptosporidium parvum* and identification of antigen-encoding genes. Journal of Protozoology 38: 76S-78S.

FAVALORO, J. M., COPPEL., R. L., CORCORAN, L. M., FOOTE, S. J., BROWN, G. V., ANDERS, R. F. & KEMP, D. J. (1986) Stucture of the RESA gene of *Plasmodium falciparum*. Nucleic acids research 14: 8265-8277.

FAYER, R. & UNGAR, B. L. P. (1986) *Cryptosporidium* spp. and Cryptosporidiosis. Microbiological Reviews 50: 458-483.

FAYER, R., ANDREWS, C., UNGAR, B. L. P. & BLAGBURN, B. (1989a) Efficacy of hyperimmune bovine colostrum for prophylaxis of cryptosporidiosis in neonatal calves. The Journal of Parasitology 75: 393-397.

FAYER, R., PERRYMAN, L. E. & RIGGS, M. W. (1989b) Hyperimmune bovine colostrum neutralizes *Cryptosporidium* sporozoites and protects mice against oocyst challenge. The Jounal of Parasitology 75: 151-153.

FAYER, R., GUIDRY, A & BLAGBURN, B. L. (1990a) Immunotherapeutic efficacy of bovine colostral immunoglobulins from a hyperimmunized cow against cryptosporidiosis in neonatal mice. Infection and Immunity 58: 2962-2965.

FAYER, R., SPEER, C. A. & DUBEY, J. P. (1990b) General biology of *Cryptosporidium*. In: Cryptosporidiosis of man and animals. (Dubey, J. P., Fayer, R. & Speer, C. A. eds.), pp 1-29. CRC Press.

FERGUSON, D. F. P., HUTCHISON, W. M. & SIIM, J. C. (1975) The ultrastructural development of the macrogamete and formation of the oocyst wall of *Toxoplasma gondii*. Acta Pathologica Microbiologica Scandinavica Section B 83: 491-505.

FINCH, P. W., WILSON, R.E., BROWN, K., HICKSON, I. D., TOMKINSON, A. E. & EMMERSON, P. T. (1986) Complete nucleotide sequence of the *Excherichia coli recC* gene and of the *thyA-recC* intergenic region. Nucleic Acids Research 14: 4437-4451.

GALINSKI, M. R., ARNOT, D. E., COCHRANE, A. H., BARNWELL, J. W., NUSSENZWEIG, R. S. & ENEA, V. (1987) The circumsporozoite gene of the *Plasmodium cynomolgi* complex. Cell 48: 311-319.

GARDNER, A. L., ROCHE, J. K., WEIKEL, C. S. & GUERRANT, R. S. (1991) Intestinal cryptosporidiosis: pathophysiologic alterations and specific cellular and humoral immune responses in rnu/+ and rnu/rnu (athymic) rats. American Journal of Tropical Medicine and Hygiene 44: 49-62.

GODSON, G. M., ELLIS, J., SDVEC, P., SCHLESINGER, D. H., & NUSSENZWEIG, V (1983) Identification and chemical synthesis of a tandemly repeated immunogenic region of *Plasmodium knowlesi* curcumsporozoite protein. Nature 305: 29-33.

GOEBEL, E. & BRAENDLER, U. (1982) Ultrastructure of microgametogenesis, microgametes and gametogony of *Cryptosporidium* sp. in the small intestine of mice. Protistologica XVIII: 331-344.

GOMAN, M., LANGSLEY, G., HYDE, J. E., YANKOVSKY, N. K., ZOLG, J. W. & SCAIFE, J. G. (1982) The establishment of genomic DNA libraries for the human malaria parasite *Plasmodium falciparum* and identification of individual clones by hybridisation. Molecular and Biochemical Parasitology 5: 391-400.

GOOZE, L., KIM, K., PETERSEN, C., GUT, J. & NELSON, R. G. (1991) Amplification of a *Cryptosporidium parvum* gene fragment encoding thymidylate synthase. Journal of Protozoology 38: 56S-58S.

GOUGH, J. A. & MURRAY, N. E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules. Journal of Molecular Biology 166: 1-19.

GRIBSKOV, M., LUTHY, R. & EISENBERG, D. (1990) Profile analysis. In: Methods in Enzymology 183: 146-159. Academic Press Inc.

HAMPTON, J. C. & ROSARIO, B. (1966) The attachment of protozoan parasites to intestinal epithelial cells of the mouse. The Journal of Parasitology 52: 939-949.

HARP, J. A. & MOON, H. W. (1991) Susceptibility of mast cell-deficient W/Wv mice to *Cryptosporidium parvum*. Infection and Immunity 59: 718-720.

HARP, J. A., WOODMANSEE, D. B. & MOON, H. W. (1990) Resistance of calves to *Cryptosporidium parvum*: effects of age and previous exposure. Infection and Immunity 58: 2237-2240.

HEINE, J., MOON, H. W. & WOODMANSEE, D. B. (1984a) Persistent *Cryptosporidium* infection in congenitally athymic (nude) mice. Infection and Immunity 43: 856-859.

HEINE, J., POHLENZ, F. L., MOON, H. W. & WOODE, G. N. (1984b) Enteric lesions and diarrhea in gnotobiotic calves monoinfected with *Cryptosporidium* species. The Journal of Infectious Diseases 150: 768-775.

HENRIKSEN, S. A. (1989) Epidemiology of cryptosporidiosis in calves. In: Cryptosporidiosis: Proceedings of the First International Workshop. (Angus, K. W. & Blewett, D. A. eds.), pp 79-83.

HERRING, A. J., INGLIS, N. F., OJEH, C. K., SNODGRASS, D. R. & MENZIES, J. D. (1982) Rapid diagnosis of rotavirus infection by direct detection of viral nucleic acid in silver stained polyacrylamide gels. Journal of Clinical Microbiology 16: 473-477.

HILL, B. D. (1989) The immunology and pathobiology of cryptosporidiosis. PhD thesis, University of Edinburgh.

HILL, B. D., BLEWETT, D. A., DAWSON, A. M. & WRIGHT, S. (1990) Analysis of the kinetics, isotype and specificity of serum and coproantibody in lambs infected with *Cryptosporidium*. Research in Veterinary Science 48: 76-81.

HOLDER, A. A., LOCKYER, M. J., ODINK, K. G., SAMDHU, J. S., RIVEROS-MORENO, V., NICHOLLS, S. C., HILLMAN, Y., DAVEY, L. S., TIZARD, M. L. V., SCHWARZ, R. T. & FREEMAN, R. R. (1985) Primary structure of the precursor to the three major surface antigens of *Plasmodium falciparum* merozoites. Nature 317: 270-273.

HOFT, D. F., KIM, K. S., OTSU, K., MOSER, D. R., YOST, W. J., BLUMIN, J. H., DONELSON, J. E. & KIRCHHOFF, L. V. (1989) *Trypanosoma cruzi* expresses diverse repetitive protein antigens. Infection and Immunity 57: 1959-1967.

HOJLYNG, N., MOLBAK, K. & JEPSEN, S. (1986) *Cryptosporidium* spp., a frequent cause of diarrhoea in Liberian children. Journal of Clinical Microbiology 23: 1109-1113.

HUYNH, T. V., YOUNG, R. A. & DAVIS, R. W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. In: DNA Cloning (Glover, D. M. ed.), vol III, pp 49-78, IRL Press, Oxford.

IAMS, P., HALL, R., WEVSTER, P. & MUSOKE, A. J. (1990) Identification of λgt11 clones encoding the major antigenic determinants expressed by *Theileria parva* sporozoites. Infection and Immunity 58: 1828-1834.

IBANEZ, C. F., AFFRANCHINIO, J. L., MACINA, R. A., REYES, M. B., LEGUIZAMON, S., CAMARGO, M. E., ASLUND, L., PETTERSSON, U. & FRASCH, A. C. C. (1988) Multiple *Trypanosoma cruzi* antigens containing tandemly repeated amino acid sequence motifs. Molecular and Biochemical Parasitology 30: 27-34.

IKEMURA, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13-34.

JENKINS, M. C. (1988) A cDNA encoding a merozoite surface protein of the protozoan *Eimeria acervulina* contains tandem-repeated sequences. Nucleic Acids Research 16: 9863.

JOHNSON, A. M., FIELKE, R., LUMB, R. & BAVERSTOCK, P. R. (1990) Phylogenetic relationships of *Cryptosporidium* determined by ribosomal RNA sequence comparison. International Journal for Parasitology 20: 141-147.

JOHNSON, S., LANAHAN, A., BUCK, C. R., SEHGAL, A., MORGAN, C., MERCER, E., BOTHWELL, M.& CHAO, M. (1986) Expression and structure of the human NGF receptor. Cell 47: 545-554.

JOLLEY, W. R., BURTON, S. D., NYBERG, P. A. & JENSEN, J. B. (1976) Formation of sulfhydryl groups in the walls of *Eimeria steidae* and *E. tenella* oocysts subjected to *in vitro* excystation. The Journal of Parasitology 62: 199-202.

JOLLEY, W. R., ALLEN, J. V. & NYBERG, P. A. (1979) Micropyle and oocyst wall changes associated with chemically mediated *in vitro* excystation of *Eimeria steidae* and *Eimeria tenella*. International Journal for Parasitology 9: 199-204.

KEEN, J., HOLDER, A., PLAYFAIR, J., LOCKYER, M. & LEWIS, A. (1990) Identification of the gene for a *Plasmidium yoelii* rhoptry protein. Multiple copies in the parasite genome. Molecular and Biochemical Parasitology 42: 241-246.

KEMP, D. J., COPPEL, R. L. & ANDERS, R. F. (1987) Repetitive proteins and genes of malaria. Annual Review of Microbiology 41: 181-208.

KIM, K., GOOZE, L., PETERSEN, C., GUT, J. & NELSON, R. G. (1992) Isolation, sequence and molecular karyotype analysis of the actin gene of *Cryptosporidium parvum*. Molecular and Biochemical Parasitology 50: 105-114.

KLESIUS, P. H., HAMES, T. B. & MALO, L. K. (1986) Infectivity of *Cryptosporidium* spp. isolated from wild mice for calves and mice. Journal of the American Veterinary Medical Association 189: 192-193.

LAEMMLI, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

LANDERS, E. J. (1960) Studies on excystation of coccidial oocysts. The Journal of Parasitology 46: 195-200.

LAWLER, J. & HYNES, R. O. (1986) The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium binding sites and homologies with several different proteins. Journal of Cell Biology 103: 1635-1648.

LAXER, M. A., TIMBLIN, B. K. & PATEL, R. J. (1991) DNA sequences for the specific detection of *Cryptosporidium parvum* by the polymerase chain reaction. American Journal of Tropical Medicine and Hygiene 45: 688-694.

LEVINE, N. D. (1984). Taxonomy and review of the coccidian genus *Cryptosporidium* (Protozoa, Apicomplexa). Journal of Protozoology 31: 94-98.

LIPMAN, D. J. & PEARSON, W. R. (1985) Rapid and sensitive protein similarity searches. Science 227: 1435-1441.

LONGMIRE, J. L., KRAEMER, P. M., BROWN, N. C., HARDEKOPF, L. C. & DEAVEN, L. L. (1990) A new multilocus DNA fingerprinting probe-pv47-2. Nucleic Acids Research 18: 16-17.

LUFT, B. J., PAYNE, D., WOODMANSEE, D. & KIM, C. W. (1987) Characterization of the *Cryptosporidium* antigens from sporulated oocysts of *Cryptosporidium parvum*. Infection and Immunity 55: 2436-2441.

LUMB, R., LANSER, J. A. & O'DONOGHUE, P. J. (1988a) Electrophoretic and immunoblot analysis of *Cryptosporidium* oocysts. Immunology and Cell Biology 66: 369-376.

LUMB, R., SMITH, K., O'DONOGHUE, P. J., & LANSER, J. A. (1988b) Ultrastructure of the attachment of *Cryptosporidium* sporozoites to tissue culture cells. Parasitology Research 74: 531-536.

LUMB, R., SMITH, P. S., DAVIES, R., O'DONOGHUE, P. J., ATKINSON, H. M. & LANSER, J. A. (1989) Localization of a 23,000 MW antigen of *Cryptosporidium* by immunoelectron microscopy. Immunology and Cell Biology 67: 267-270.

MAIZEL, J. V. & LENK, R. P. (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proceedings of the National Academy of Sciences USA 78: 7665-7669.

MANDEL, M. & HIGA, A. (1970) Calcium dependent bacteriophage DNA infection. Journal of Molecular Biology 53: 159-162.

MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory. Cold Spring Harbor. N. Y.

MANN, E. D., SEKLA, L. H., NAYER, G. P. S. & KOCHIK, C. (1986) Infection with *Cryptosporidium* spp. in humans and cattle in Manitoba. Canadian Journal of Veterinary Research 50: 174-178.

MARCIAL, M. A. & MADARA, J. L. (1986) *Cryptosporidium*: cellular localization, structural analysis of absorptive cell-parasite membrane-membrane interactions in guinea pigs, and suggestion of protozoan transport by M cells. Gastroenterology 90: 583-594.

MARSTON, F. A. O. (1986) The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. The Biochemical Journal 240: 1-12.

MARSTON, F. A. O. (1987) The purification of eukaryotic polypeptides expressed in *Escherichia coli*. In: DNA Cloning (Glover, D. M. ed.), vol. III, pp 59-88, IRL Press, Oxford.

MASLEN, C. L., CORSON, G. M., MADDOX, B. K., GLANVILLE, R. W. & SAKAI, L. Y. (1991) Partial sequence of a candidate gene for the Marfan syndrome. Nature 352: 334-337.

MATA, L., BOLANOS, H., PIZARRO, D. & VIVES, M. (1984) Cryptosporidiosis in children form some highland Costa Rican rural and urban areas. American Journal of Tropical Medicine and Hygiene 33: 24-29.

MCCUTCHAN, R. F., LAL, A. A., DE LA CRUZ, V., MILLER, L. H., MALOY, W. L., CHAROENVIT, Y., BEAUDOIN, R. L., GUERRY, P., WISTAR, R., HOFFMAN, S. L., HOCKMEYER, W. T., COLLINS, W. E. & WIRTH, D. (1985) Sequence of the immunodominant epitope for the surface protein on sporozoites of *Plasmodium vivax*. Science 230: 1381-1383.

MCDONALD, V., STABLES, R., WARHURST, D. C., BARER, M. R., BLEWETT, D. A., CHAPMAN, H. D., CONNOLLY, G. M., CHIODINI, P. L. & MCADAM, K. P. W. J. (1990) *In vitro* cultivation of *Cryptosporidium* parvum and screening for anticryptosporidial drugs. Antimicrobial Agents and Chemotherapy 34: 1498-1500.

MCDONALD, V., DEER, R. M. A., NINA, J. M. S., WRIGHT, S., CHIODINI, P. L. & MCADAM, K. P. W. J. (1991) Characteristics and specificity of hybridoma antibodies against oocyst antigens of *Cryptosporidium parvum*. Parasite Immunology 13: 251-259.

MCMEEKING, A., BORKOWSKY, W., KLESIUS, P. H. BONK, S., HOLZMAN, R. S. & LAWRENCE, H. S. (1991) A controlled trial of bovine dialyzable leukocyte extract for cryptosporidiosis in patients with AIDS. The Journal of Infectious Diseases 161: 108-112.

MCQUAY, S. J. (1991) The interpretation of alignments between proteins of unusual sequence composition. Biochemical Society Transactions 19: 523-524.

MEAD, J. R., ARROWOOD, M. J., CURRENT, W. L. & STERLING, C. R. (1988a) Field inversion gel electrophoretic separation of *Cryptosporidium* spp. chromosome-sized DNA. The Journal of Parasitology 74: 366-369.

MEAD, J. R., ARROWOOD, M. J. & STERLING, C. R. (1988b) Antigens of *Cryptosporidium* sporozoites recognised by immune sera of infected animals and humans. The Journal of Parasitology 74: 135-143.

MEAD, J. R., ARROWOOD, M. J., SIDWELL, R. W. & HEALEY, M. C. (1991) Chronic *Cryptosporidium parvum* infections in congenitally immunodeficient SCID and nude mice. The Journal of Infectious Diseases 163: 1297-1304.

MEINKOTH, J. & WAHL, G. (1984) Hybridization of nucleic acids immobilized on solid supports. Analytical Biochemistry 138: 267-284.

MEISEL., J. L., PERERA, D. R., MELIGRO, C. & RUBIN, C. E. (1976) Overwhelming watery diarrhoea associated with *Cryptosporidium* in an immunosuppressed patient. Gastroenterology 70: 1156-1160.

MONNE, L. & HONNIG, G. (1954) On the properties of the shells of the coccidian oocysts. Arkiv for Zoologi, Stockholm 7: 251-256.

MOON, H. W., WOODMANSEE, D., HARP, J. A., ABEL, S. & UNGAR, B. L. P. (1988) Lacteal immunity to enteric cryptosporidiosis in mice: immune dams do not protect their suckling pups. Infection and Immunity 56: 649-653.

MOSKOVITZ, B. L., STANTON, T. L. & KUSMIEREK, J. J. E. (1988) Spiramycin therapy for cryptosporidial diarrhoea in immunocompromised patients. Journal of Antimicrobials and Chemotherapy Suppl. B., 22: 189-191.

MUSKAL, S. M., HOLBROOKE, S. R. & KIM, S-H. (1990) Prediction of the disulfide-bonding state of cysteine in proteins. Protein Engineering 3: 667-672.

MUTO, A. & OSAWA, S. (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proceedings of the National Academy of Sciences USA 84: 166-169.

NACIRI, M & YVORE, P. (1989) Treatment of cryptosporidiosis in lambs using halofuginone lactate. Abstract. In: Cryptosporidiosis: Proceedings of the First International Workshop. K. W. Angus and D. A. Blewett, Eds., 8-9th September 1988, Edinburgh. p120.

NELSON, R. G., KIM, K., GOOZE, L., PETERSEN, C. & GUT, J. (1991) Identification and isolation of *Cryptosporidium parvum* genes encoding microtubule and microfilament proteins. Journal of Protozoology 38: 52S-55S.

NICHOLS, G. L., MCLAUCHLIN, J. & SAMUEL, D. (1991) A technique for typing *Cryptosporidium* isolates. Journal of Protozoology 38: 237S-240S.

NIME, F. A., BUREK, J. D., PAGE, D. L., HOLSCHER, M. A. & YARDLEY, J. H. (1976) Acute enterocolitis in a human being infected with the protozoan *Cryptosporidium.* Gastroenterology 70: 592-598.

NINA, J. M. S., MCDONALD, V., DEER, R. M. A., WRIGHT, S. E., DYSON, D. A. CHIODINI, P. L. & MCADAM, K. P. W. J. (1992) Comparative study of the antigenic composition of oocyst isolates of *Cryptosporidium parvum* from different hosts. Parasite Immunology 14: 227-232.

NUSSENZWEIG, V. & NUSSENZWEIG, R. S. (1989) Rationale for the development of an engineered sporozoite malaria vaccine. Advances in Immunology 45: 283-334.

NYBERG, P. A. & KNAPP, S. E. (1970) Effect of sodium hypochlorite on the oocyst wall of *Eimeria tenella* as shown by electron microscopy. Proceedings of the Helminthological Society of Washington 37: 29-32.

ORTEGA, Y. R., SHEEHY, R. R., CAMA, V. A., OISHI, K. K. & STERLING, C. R. (1991) Restriction fragment length polymorphism analysis of *Cryptosporidium parvum* isolates of bovine and human origin. Journal of Protozoology 38: 40S-41S.

ORTEGA-MORA, L. M., TRONCOSO, J. M., ROJO-VAZQUEZ, F. A. & GOMEZ-BAUTISTA, M. (1992) Cross-reactivity of polyclonal serum antibodies generated against *Cryptosporidium parvum* oocysts. Infection and Immunity 60: 3442-3445.

PANCIERA, R. J., THOMASSEN, R. W. & GARDNER, F. M. (1971) Cryptosporidial infection in a calf. Veterinary Pathology 8: 479-484.

PAPADOPOULU, C., XYLOURI, E., MANTZIOS, A., SPYROPOULOS, G. & STOFOROS, S. (1989) Cryptosporidiosis in farm animals in Greece. Abstract, in: Cryptosporidiosis: Proceedings of the First International Workshop. p123.

PEARSON, W. R. & LIPMAN, D. J. (1988) Improved tools for biological sequence comparison. Proceedings of the National Academy of Sciences USA 85: 2444-2448.

PEARSON, G. R. & LOGAN, E. F. (1983) Scanning and transmission electron microscopic observations on the host-parasite relationship in intestinal cryptosporidiosis of neonatal calves. Research in Veterinary Science 34: 149-154.

PEETERS, J. E., MAZAS, E. A., MASSCHELEIN, S. J. DE MATURANA, I. V. M. & DEBACKER, E. (1989) Effects of disinfection of drinking water with ozone or chlorine dioxide on survival of *Cryptosporidium parvum* oocysts. Applied Environmental Microbiology 55: 1519-1522.

PERKINS, M. E. (1992) Rhoptry organelles of apicomplexan parasites. Parasitology today 8: 28-32.

PERRYMAN, L. E., RIGGS, M. W., MASON, P. H. & FAYER, R. (1990) Kinetics of *Cryptosporidium parvum* sporozoite neutralization by monoclonal antibodies, immune bovine serum, and immune bovine colostrum. Infection and Immunity 58: 257-259.

PETERSEN, C., GUT, J., LEECH, J. H. & NELSON, R. G. (1992) Identification and initial characterization of five *Cryptosporidium parvum* sporozoite antigen genes. Infection and Immunity 60: 2343-2348.

PORTNOY, D., WHITESIDE, M. E., BUCKLEY, E. & MACLEOD, C. L. (1984) Treatment of intestinal cryptosporidiosis with spiramycin. Annals of Internal Medicine 101: 202-204.

REESE, N. C., CURRENT, W. L., ERNST, J. V. & BAILEY, W. S. (1982) Cryptosporidiosis of man and calf: a case report and results of experimental infection in mice and rats. American Journal of Tropical Medicine and Hygiene 31: 226-229.

REHG, J. E. (1991a) Activity of azithromycin against cryptosporidia in immunosuppressed rats. The Journal of Infectious Diseases 163: 1293-1296.

REHG, J. E. (1991b) Anticryptosporidial activity is associated with specific sulfonamides in immunosuppressed rats. The Journal of Parasitology 77: 238-240.

REHG, J. E., HANCOCK, M. L. & WOODMANSEE, D. B. (1987) Characterization of cyclophosphamide-rat model of cryptosporidiosis. Infection and immuity 55: 2669-2674.

REID, K. B. M. & DAY, A. J. (1989) Structure-function relationships of the complement components. Immunology Today 10: 177-180.

REYNOLDS, D. J., MORGAN, J. H., CHANTER, N., JONES, P. W., BRIDGER, J. C., DEBNEY, T. G. & BUNCH, K. J. (1986) Microbiology of calf diarrhoea in southern Britain. The Veterinary Record 119: 34-39.

RIDLEY, R. G. (1991) Proteins of unusual sequence composition from the malarial parasite *Plasmodium falciparum*. Biochemical Society Transactions 19: 525-528.

RIGGS, M. W., MCGUIRE, T. C., MASON, P. H. & PERRYMAN, L. E. (1989) Neutralization-sensitive epitopes are exposed on the surface of infectious *Cryptosporidium parvum* sporozoites. The Journal of Immunology 143: 1340-1345.

RIGGS, M. W. & PERRYMAN, L. E. (1987) Infectivity and neutralization of *Cryptosporidium parvum* sporozoites. Infection and Immunity 55: 2081-2087.

ROITT, I. M., BROSTOFF, J. & MALE, D. K. (1985) Immunology. Gower Medical Publishing, London.

RYLEY, J. F. (1980) Recent developments in coccidian biology: where do we go from here ? Parasitology 80: 189-209.

SAINT, R. B., COPPEL, R. L., COWMAN, A. F., BROWN, G. V., SHI, P-T., BARZAGA, N., KEMP, D. J. & ANDERS, R. F. (1987) Changes in repeat number, sequence, and reading frame in S-antigen genes of *Plasmodium falciparum*. Molecular and Cellular Biology 7: 2968-2973.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989) Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbour Laboratory. Cold Spring Harbour. N. Y.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977) DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA 74: 5463-5467.

SAUL, A. & BATTISTUTTA, D. (1988) codon usage in *Plasmodium falciparum*. Molecular and Biochemical Parasitology 27: 35-42.

SAXON, A. & WEINSTEIN, W. (1987) Oral administration of bovine colostrum anti-cryptosporidia antibody fails to alter the course of human cryptosporidiosis. The Journal of Parasitology. 73: 413-415.

SCHERF, A., MATTEI, D. & SCHREIBER, M. (1990) Parasite antigens expressed in *Escherichia coli*. A refined approach for epidemiological analysis. Journal of Immunological Methods 128: 81-87.

SCHOFIELD, L. (1991) On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. Parasitology Today 7: 99-105.

SHARP, P. M. & DEVINE, K. M. (1989) Codon usage and gene expression in *Dictyostelium discoideum*: highly expressed genes do 'prefer' optimal codons. Nucleic Acids Research 17: 5029-5039.

SHARP, P. M. & LI, W-H. (1986) Codon usage in regulatry genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Research 14: 7737-7749.

SHERWOOD, D., ANGUS, K. W., SNODGRASS, D. R. & TZIPORI, S. (1982) Experimental cryptosporidiosis in laboratory mice. Infection and Immunity 38: 471-475.

SHIELDS, D. C. & SHARP, P. M. (1987) Synonymous codon usage in *Bacillus subtilis* reflects both translational selection and mutational biases. Nucleic Acids Research 15: 8023-8040.

SHORT, J. M., FERNANDEZ, J. M., SORGE, J. A. & HUSE, W. D. (1988) λZAP: a bacteriophage λ expression vector with *in vivo* excision properties. Nucleic Acids Research 16: 7583-7600.

SLAVIN, D. (1955) *Cryptosporidium meleagridis* (sp. nov). Journal of Comparative Pathology 65: 262-266.

SMITH, H. V. (1990) Environmental aspects of *Cryptosporidium* species: a review. Journal of the Royal Society of Medicine 83: 629-631.

SMITH, G. E. & SUMMERS, M. D. (1980) The bidirectional transfer of DNA and RNA to nitrocellulose of diazobenzyloxymethyl-paper. Analytical Biochemistry 109: 123-129.

SMITH, G. & VAN DEN ENDE, J. (1986) Cryptosporidiosis among black children in hospital in South Africa. Journal of Infection 13: 25-30.

SMITH, T. F. & WATERMAN, M. S. (1981) Identification of common molecular subsequences. Journal of Molecular Biology 147: 195-197.

SNODGRASS, D. R., ANGUS, K. W. & GRAY, E. W. (1984) Experimental cryptosporidiosis in germfree lambs. Journal of Comparative Pathology 94: 141-152.

STOTISH, R. L., WANG, C. C. & MEYENHOFER, M. (1978) Structure and composition of the oocyst wall of *Eimeria tenella*. Journal of Parasitology 64: 1074-1081.

STROUT, R. G., BOTERO, H., SMITH, S. C. & DUNLOP,W. R. (1963). The lipids of coccidial oocysts. Journal of Parasitology 49 (Suppl.): 20.

SUEOKA, N. (1988) Directional mutation pressure and neutral molecular evolution. Proceedings of the National Academy of Sciences USA 85: 2653-2657.

TAGHI-KILANI, R., SEKLA, L. & HAYGLASS, K. T. (1990) The role of humoral immunity in *Cryptosporidium* spp infection. Studies with B-cell depleted mice. The Journal of Immunology 145: 1571-1576.

TILLEY, M., FAYER, R., GUIDREY, A., UPTON, S. J. & BLAGBURN, B. L. (1990a) *Cryptosporidium parvum* (Apicomplexa: Cryptosporidiidae) oocyst and sporozoite antigens recognized by bovine colostral antibodies. Infection and Immunity 58: 2966-2971.

TILLEY, M., UPTON, S. J., BLAGBURN, B. L. & ANDERSON, B. C. (1990b) Identification of outer oocyst wall proteins of three *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) species by ^{125}I surface labeling. Infection and Immunity 58: 252-253.

TILLEY, M., UPTON, S. J., FAYER, R., BARTA, J. R., CHRISP, C. E., FREED, P. S., BLAGBURN, B. L., ANDERSON, B. C. & BARNARD, S. M. (1991) Identification of a 15-kilodalton surface glycoprotein of sporozoites of *Cryptosporidium parvum*. Infection and Immunity 59: 1002-1007.

TOWBIN, H., STAEHELIN, T. & GORDON, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences USA 76: 4350-4354.

TRIGLIA, T., STAHL, H. D., CREWTHER, P. E., SCANLON, D., BROWN, G. V., ANDERS R. F. & KEMP, D. J. (1987) The complete structure of the knob-associated histidine-rich protein from *Plasmodium falciparum*. EMBO Journal 6: 1413-1419.

TYZZER,E. E. (1907) A sporozoan found in the peptic glands of the common mouse. Proceedings of the Society for Experimental Biology and Medicine 5: 12-13.

TYZZER, E. E. (1910) An extracellular coccidium, *Cryptosporidium muris* (gen. et sp. nov.), of the gastric glands of the common mouse. Journal of Medical Research XXXIII: 487-509.

TYZZER, E. E. (1912) *Cryptosporidium parvum* (sp. nov.) a coccidium found in the small intestine of the common mouse. Archiv fur Protistenkunde 26: 394-418.

TZIPORI, S. (1988) Cryptosporidiosis in perspective. Advances in Parasitology 27: 63-129.

TZIPORI, S., ANGUS, K. W., CAMPBELL, I. W. & GRAY, E. W. (1980) *Cryptosporidium*: evidence for a single-species genus. Infection and Immunity 30: 884-886.

TZIPORI, S., ANGUS, K. W., CAMPBELL, I. & CLERIHEW, L. W. (1981) Diarrhoea due to *Cryptosporidium* infection in artificially reared lambs. Journal of Clinical Microbiology 14: 100-105.

TZIPORI, S., LARSEN, J., SMITH, M. & LEUFL, R. (1982a) Diarrhea in goat kids attributed to *Cryptosporidium* infection. The Veterinary Record 111: 35-36.

TZIPORI, S., SMITH, M., MAKIN, T. & HALPIN, C. (1982b) Enterocolitis in piglets caused by *Cryptosporidium* sp. purified from calf faeces. Veterinary Parasitology 11: 121-126.

TZIPORI, S., SMITH, M., HALPIN, C. M., ANGUS, K. W., SHERWOOD, D. & CAMPBELL, I. (1983) Experimental cryptosporidiosis in calves: clinical manifestations and pathological findings. The Veterinary Record 112: 116-120.

TZIPORI, S., ROBERTON, D. & CHAPMAN, C. (1986) Remission of diarrhoea due to cryptosporidiosis in an immunodeficient child treated with hyperimmune bovine colostrum. British Medical Journal 293: 1276-1277.

TZIPORI, S., ROBERTON, D., COOPER, D. A. & WHITE, L. (1987) Chronic cryptosporidial diarrhoea and hyperimmune cow colostrum Lancet 2: 344-345.

ULLRICH, A., COUSSENS, L., HAYFLICK, J. S., DULL, T. J., GRAY, A., TAM, A. W., LEE, J., YARDEN, Y., LIBERMANN, T. A., SCHLESSINGER, J., DOWNWARD, J., MAYERS, E. L. V., WHITTLE, N., WATERFIELD, M. D. & SEEBURG, P. H. (1984) Human epidermal growth factor cDNA sequences and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309: 418-425.

UNGAR, B. L. P. & NASH, T. E. (1986) Quantification of specific antibody response to *Cryptosporidium* antigens by laser densitometry. Infection and Immunity 53: 124-128.

UNGAR, B. L. P., SOAVE, R., FAYER, R. & NASH, T. E. (1986) Enzyme immunoassay detection of immunoglobulin M and G antibodies to *Cryptosporidium* in immunocompetent and immunodedicient persons. The Journal of Infectious Diseases 153: 570-578.

UNGAR, B. L. P., BURRIS, J. A., QUINN, C. A. & FINKELMAN, F. D. (1990a) New mouse models for chronic *Cryptosporidium* infection in immunodeficient hosts. Infection and Immunity 58: 961-969.

UNGAR, B. L. P., WARD, D. J., FAYER, R. & QUINN, C. A. (1990b) Cessation of *Cryptosporidium*-associated diarrhea in an AIDS patient after treatment with hyperimmune bovine colostrum. Gastroenterology 98: 486-489.

UNGAR, B. L. P., KAO, T., BURRIS, J. A. & FINKELMAN, F. D. (1991) *Cryptosporidium* infection in an adult mouse model: independent roles for IFN-γ and CD4+ T lymphocytes in protective immunity. The Journal of Immunology 147: 1014-1022.

UPTON, S. J. & CURRENT, W. L. (1985) The species of *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) infecting mammals. The Journal of Parasitology 71: 625-629.

VERWEIJ, C. L., DIERGAARDE, P. J., HART, M. & PANNEKOEK, H. (1986) Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably longer than the mature vWF subunit. The EMBO Journal 5: 1839-1847.

VETTERLING, J. M., TAKEUCHI, A. & MADDEN, P. A. (1971) Ultrastructure of *Cryptosporidium wrairi* from the guinea pig. Journal of Protozoology 18: 248-260.

VETTERLING, J. M., PACHECO, N. D. & FAYER, R. (1973) Fine structure of gametogony and oocyst formation in *Sarcocystis* sp. in cell culture. Journal of Protozoology 20: 613-621.

VILLACORTA, I., PEETERS, J. E., VANOPDENBOSCH, E., ARES-MAZAS, E. & THEYS, H. (1991) Efficacy of halofuginone lactate against *Cryptosporidium parvum* in calves. Antimicrobial Agents and Chemotherapy 35: 283-287.

VOGEL, M., GOTTSTEIN, B., MULLER, N. & SEEBUCK, T. (1988) Production of a recombinant antigen of *Echinococcus multilocularis* with high immunodiagnostic sensitivity and specificity. Molecular and Biochemical Parasitology 31: 117-126.

VOGELSTEIN, B. & GILLESPIE, D. (1979) Preparative and analytical purification of DNA from agarose. Proceedings of the National Academy of Sciences USA 76: 615-619.

WALLIS, A. E. & MCMASTER, W. R. (1987) Identification of *Leishmania* genes encoding proteins containing tandemly repeating peptides. Journal of Experimental Medicine 166: 1814-1824.

WANG, C. C. (1982) Biochemistry and physiology of Coccidia. In: The Biology of the Coccidia. (Long, P. L. ed.), pp 167-216. University Park Press, Baltimore.

WEINSTEIN, L., EDELSTEIN, S. M., MADARA, J. L., FALCHUK, K. R., MCMANUS, B. M. & TRIER, J. S. (1981) Intestinal cryptosporidiosis complicated by disseminated cytomegalovirus infection. Gastroenterology 81: 584-591.

WHITMIRE, W. M. & HARP, J. A. (1991) Characterization of bovine cellular and serum antibody responses during infection by *Cryptosporidium parvum*. Infection and Immunity 59: 990-995.

WIDDOWSON, D. C. C., PROFFITT, J. A., JAGGER, P. S., RICHARDS, A. J. & HAMES, B. D. (1990) Developmental expression and characterization of the gene encoding spore coat protein SP60 in *Dictyostelium discoideum*. Molecular Microbiology 4: 951-960.

WILBUR, W. J. & LIPMAN, D. (1983) Rapid similarity searches of nucleic acid and protein data banks. Proceedings of the National Academy of Sciences USA 80: 726-730.

WITTENBERG, D. F., MILLER, N. M. & VAN DEN ENDE, J. (1989). Spiramycin is not effective in treating *Cryptosporidium* diarrhoea in infants: results of a double-blind randomized trial. The Journal of Infectious Diseases 159: 131-132.

YAMAMOTO, T., DAVIS, C. G., BROWN, M. S., SCHEIDER, W. J., CASEY, M. L., GOLDSTEIN, J. L. & RUSSEL, D. W. (1984) The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39: 27-38.

YAN, H. & CHAO, M. V. (1991) Disruption of cysteine-rich repeats of the p75 nerve growth factor receptor leads to loss of ligand binding. Journal of Biological Chemistry 266: 12099-12104.

YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

YANG, Y. F., TAN-ARIYA, P., SHARMA, Y. D. & KILEJIAN, A. (1987) The primary structure of a *Plasmodium falciparum* polypeptide related to heat shock proteins. Molecular and Biochemical Parasitology 26: 61-67.

YOUNG, R. A. & DAVIS, R. W. (1983a) Efficient isolation of genes by using antibody probes. Proceedings of the National Academy of Sciences USA 81: 1194-1198.

YOUNG, R. A. & DAVIS, R. W. (1983b) Yeast RNA polymerase II genes: isolation with antibody probes. Science 222: 778-782.

ZAVALA, F., TAM, J. P., HOLLINGDALE, M. R., COCHRANE, A. H., QUAKYI, I., NUSSENZWEIG, R. S. & NUSSENZWEIG, V. (1985) Rationale for development of a synthetic vaccine against *Plasmodium falciparum* malaria. Science 228: 1436-1440.

Molecular and Biochemical Parasitology, 56 (1992) 69–78

MOLBIO 01828

A 2359-base pair DNA fragment from *Cryptosporidium parvum* encoding a repetitive oocyst protein

Nicola C. Lally[a], G. David Baird[a], S.J. McQuay[b], Frank Wright[c] and Joanna J. Oliver[a]

[a]*Moredun Research Institute, Edinburgh, UK;* [b]*Biocomputing Research Unit, University of Edinburgh, Edinburgh, UK; and* [c]*Scottish Agricultural Statistics Service, Edinburgh, UK*

(Received 24 March 1992; accepted 22 July 1992)

A *Cryptosporidium parvum* λgt11 expression library was constructed using *Eco*RI-digested genomic DNA extracted from in vitro-excysted oocysts. Screening of this library with rat anti-*Cryptosporidium* antiserum led to the isolation of a clone containing a 2359-bp *Eco*RI fragment. When this fragment was ligated into the *Eco*RI site of plasmid vector pMS1S, the resulting clone expressed a 200-kDa β-galactosidase fusion protein. Western blot analysis using serum raised against this fusion protein indicated that the *Eco*RI fragment represented part of a gene encoding a 190-kDa oocyst wall protein of *C. parvum*. Sequencing of the fragment revealed a continuous open reading frame encoding 786 amino acids. The DNA sequence is relatively low in G + C (39.1%), and the third codon position contains only 17.9% G + C. The deduced peptide sequence has unusually high proportions of cysteine, proline, glutamine and histidine. Another striking feature of the amino acid sequence is the presence of distinctly repetitive regions based on conserved cysteine residues.

Key words: *Cryptosporidium parvum*; Gene cloning; Oocyst wall protein; Cysteine-rich repeat

Introduction

Cryptosporidium parvum (Apicomplexa: Cryptosporidiidae) is an obligate intracellular parasitic protozoan which infects the gastrointestinal tract of a wide range of mammalian species, including man [1]. Despite the medical and veterinary importance of *C. parvum*, studies of this organism at the genetic level have only begun in recent years. The earliest published work on the molecular biology of *Cryptosporidium* involved field inversion gel electrophoresis of chromosome-sized DNA of two species of *Cryptosporidium* [2]. *C. parvum* appeared to contain five chromosomes ranging in size from 1400 kb to over 3300 kb. The phylogenetic relationship of *Cryptosporidium* to a range of taxonomic groups has also been investigated using small subunit ribosomal RNA sequence data [3]. Other recent genetic studies have identified, by hybridization with heterologous DNA probes, *C. parvum* genes encoding actin [4], topoisomerase [5] and dihydrofolate reductase-thymidylate synthase [6]. In addition, anti-*Cryptosporidium* antiserum has been used to identify *C. parvum* antigen-encoding genes [5].

The present communication describes the use of anti-*Cryptosporidium* antiserum to isolate a 2359-bp gene fragment from *C. parvum* that encodes approximately half of a 190-kDa oocyst wall protein. The encoded amino acid sequence is markedly repetitive and contains a high proportion of cysteine.

Correspondence address: Nicola C. Lally, Moredun Research Institute, 408 Gilmerton Road, Edinburgh EH17 7JH, UK. Tel: 031-664-3262; Fax: 031-664-8001.

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ data base with the accession number M95743.

Abbreviations: GCG, Genetics Computer Group; IPTG, isopropyl-β-D-thiogalactopyranoside; NBRF, National Biomedical Research Foundation; ORF, open reading frame.

Materials and Methods

Parasites. An isolate of *C. parvum*, originally recovered from red deer calves [7], was maintained by passage in 5-day-old lambs or bovine calves. Oocysts were extracted from faeces as described by Hill et al. [8].

Construction and screening of λgt11 expression library. Excystation was carried out by incubating oocysts for 1 h at 37°C in 1% (w/v) trypsin in HBSS (0.41 mM $MgSO_4$/0.49 mM $MgCl_2$/1.26 mM $CaCl_2$/137 mM NaCl/5.37 mM KCl/1.07 mM Na_2HPO_4/0.44 mM KH_2PO_4/0.002% (w/v) phenol red) adjusted to pH 2.5–3.0 with HCl, followed by centrifugation at 500 × *g*, resuspension in 0.05% (w/v) sodium deoxycholate, 0.2% (w/v) sodium hydrogen carbonate in HBSS, pH 8–8.4, and incubation at 37°C for 30 min. After centrifugation, genomic DNA was extracted from the sporozoites by lysis in 50 mM Tris-HCl, pH 8.0/50 mM EDTA, 1% (w/v) SDS, and digestion with proteinase K (100 μg ml^{-1}), at 50°C for 3 h followed by phenol/chloroform extraction and ethanol precipitation. Purified genomic DNA was digested with *Eco*RI and then ligated into dephosphorylated λgt11 arms. After in vitro packaging, the recombinant phage were plated with *Escherichia coli* Y1090 and screened with rat anti-*Cryptosporidium* antiserum by standard procedures [9].

Sub-cloning and purification of the fusion protein. The 2359-bp *Eco*RI fragment (CPR1) was excised from purified recombinant phage DNA and sub-cloned into the *Eco*RI site of plasmids pMS1S, pMS2S and pMS3S, forming recombinant plasmids pCPR1/1S, pCPR1/2S and pCPR1/3S, respectively. The pMS plasmids were constructed by M. Schreiber (Institut fur Genetik, Cologne) and are derivatives of pMSgt11 [10]. Each of pMS1S, 2S and 3S contains a unique *Eco*RI site approximately 50 bp downstream from the 3′ end of the *lacZ* gene. In pMS2S and pMS3S the *Eco*RI site is shifted by one and two nucleotides respectively, with respect to its position in pMS1S, allowing cloning of any DNA fragment into each of the three reading frames. In addition each of the pMS plasmids contains a transcription termination region 9–22 bp downstream from the *Eco*RI site. Induction of expression with isopropyl-β-D-thiogalactopyranoside (IPTG), and preparation of crude cell lysates for SDS-PAGE analysis, were carried out in *E. coli* strain NM522 by the method of Carroll and Laughon [11].

Preparation of partially-purified fusion protein expressed from recombinant plasmid pCPR1/1S was carried out by the method of Marston [12]. Briefly, overnight cultures which had been induced with IPTG, were pelleted by centrifugation, then lysed by incubation with 0.5 mg ml^{-1} lysozyme and 2 mg ml^{-1} sodium deoxycholate. The mixture was then centrifuged at 3000 × *g* to pellet the inclusion bodies.

Antisera. Rat anti-*Cryptosporidium* antiserum was raised by oral inoculation of each of a litter of rats with 10^6 *C. parvum* oocysts at 5 and at 16 days of age. In addition, the rats were inoculated with 10^6 in vitro-excysted, freeze-thawed (3×) *C. parvum* oocysts at 38 and 52 days by intramuscular injection, and at 45 days by subcutaneous injection. The rats were exsanguinated at 60 days and their sera pooled. Polyclonal antiserum was raised against the fusion protein expressed from plasmid pCPR1/1S. To achieve this, the partially-purified fusion protein was electrophoresed through a 7.5% (w/v) polyacrylamide gel and the region containing the fusion protein excised, homogenised in Freund's complete adjuvant and inoculated subcutaneously into a rabbit. The same procedure, but substituting Freund's incomplete for Freund's complete adjuvant, was repeated on days 30 and 85. Before use the serum was diluted 1:10 and preabsorbed with a lysate prepared from a clone of *E. coli* NM522 harbouring pMS1S. This effectively removed antibodies recognising β-galactosidase and other *E. coli* proteins. Mouse anti-β-galactosidase antiserum was obtained from Sigma.

SDS-PAGE and Western blotting. Proteins were separated by SDS-PAGE under reducing conditions [13], transferred electrophoretically to polyvinylidene difluoride membranes (Immobilon, Millipore) and then incubated with the required antibody probe. Binding of primary antibody was detected by incubation in anti-mouse immunoglobulin conjugated to alkaline phosphatase (Sigma), or anti-rabbit or anti-rat immunoglobulins conjugated to horseradish peroxidase (Dako), followed by colour development in the appropriate substrate solution.

Southern blotting. Genomic DNA was digested with *Eco*RI, electrophoresed through a 0.8% (w/v) agarose gel and transferred to Hybond-N hybridisation membrane (Amersham). The blot was then probed with DNA labelled by random-primed incorporation of digoxigenin-labelled dUTP. The final wash was performed in 0.1 × SSC (15 mM NaCl/1.5 mM sodium citrate, pH 7.0) at 65°C.

Density gradient separation of sporozoites and oocyst shells. Following in vitro-excystation and washing with PBS (137 mM NaCl/26.8 mM KCl/8.1 mM Na_4HPO_4/1.47 mM KH_2PO_4), sporozoites, oocyst shells and unexcysted oocysts were separated by centrifugation (1000 × *g*, 20 min) on 10 ml 15–50% (v/v) Percoll (Pharmacia) gradients. Fractions (0.5 ml) were collected and examined by phase contrast microscopy. Appropriate fractions were pooled, diluted with PBS, and pelleted by centrifugation at 1000 × *g*.

DNA sequencing. The CPR1 insert was subcloned into Bluescript SK+ (Stratagene) and DNA from this recombinant phagemid was restricted and religated to form additional recombinant clones containing deletions of the CPR1 insert. Dideoxy chain-termination DNA sequencing [14] was then performed on double stranded template DNA using M13 universal and reverse primers and T7 polymerase. Additional oligonucleotide primers were designed to sequence the entire CPR1 insert.

Protein sequence analysis. Most of the sequence analysis was carried out using the SEQNET molecular biology computer facility based at the SERC Daresbury Laboratory, Warrington, UK. Programs in the Genetics Computer Group (GCG) sequence analysis software package [15] were used to calculate dot-matrix analyses, to search for short defined peptides and for profile analysis [16].

An exhaustive search of the entire National Biomedical Research Foundation (NBRF) protein sequence database, version 32.0 (40 298 proteins, 11 831 134 residues), was carried out using the derived protein sequence (P786) as a query for the 'prosrch' protein sequence database searching program [17] implemented on the computing facilities in the Biocomputing Research Unit, University of Edinburgh.

Results

Isolation of the CPR1 gene fragment and characterisation of its encoded peptide. Screening of 2 × 10^5 recombinant phage with rat anti-*Cryptosporidium* antiserum led to the isolation of three immunoreactive clones, each containing an identical insert of approximately 2.4 kb (designated CPR1). Following induction with IPTG, plasmid pCPR1/1S, containing CPR1, expressed a β-galactosidase fusion protein of approximately 200 kDa, recognised by the anti-*Cryptosporidium* antiserum (Fig. 1). Plasmids pCPR1/2S and pCPR1/3S both failed to express the fusion protein, indicating that the CPR1 insert was likely to contain an open reading frame (ORF) which was in frame with the *Eco*RI site in both λgt11 and pMS1S. Since β-galactosidase accounts for 116 kDa of the fusion protein, this result indicated that the entire CPR1 insert was likely to consist of a single open reading frame.

Polyclonal rabbit serum raised against the recombinant fusion protein strongly recognised a 190-kDa protein on a Western blot of proteins from excysted *C. parvum* oocysts (Fig. 2). To determine whether this 190-kDa protein

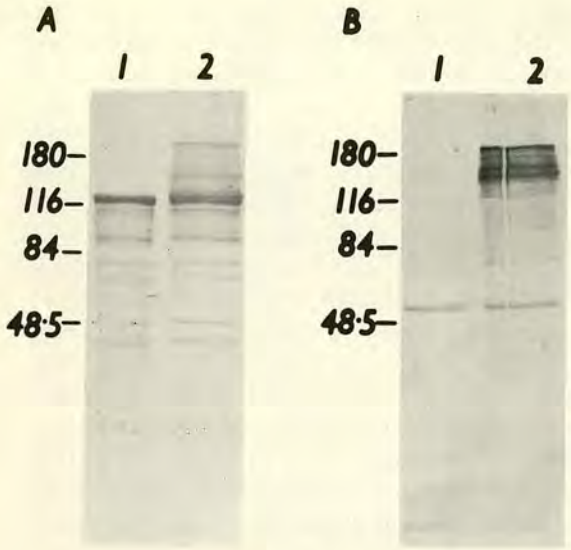


Fig. 1. Identification of the recombinant fusion protein expressed by pCPR1/1S. Induced lysates prepared from clones harbouring plasmids pMS1S (1) and pCPR1/1S (2) were fractionated on a 7.5% SDS-PAGE gel and Western blots probed with mouse anti-β-galactosidase serum (A) or rat anti-*Cryptosporidium* antiserum (B). Positions and sizes of molecular weight markers are indicated.

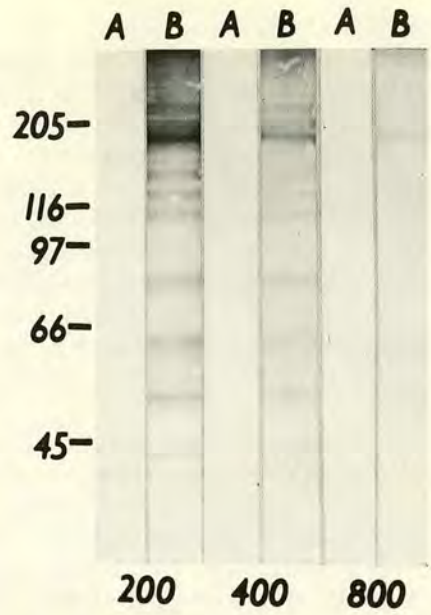


Fig. 2. Recognition of a 190-kDa *C. parvum* antigen by antiserum raised against the fusion protein expressed by pCPR1/1S. Proteins extracted from in vitro-excysted *C. parvum* oocysts were fractionated on a 7.5% SDS-PAGE gel and Western blotted. Strips were probed either with polyclonal rabbit antiserum raised against the fusion protein expressed by pCPR1/1S (B), or with serum from the same rabbit prior to inoculation with the fusion protein (A). Positions of molecular weight markers are indicated on the left. Serum dilutions are indicated below the blot.

was a constituent of the sporozoites or of the oocyst wall, immunoblotting experiments were carried out using samples enriched for either sporozoites or oocyst shells. When analysed by SDS-PAGE, only four bands were visible in the samples enriched for oocyst shells, with apparent sizes of 190, 85, 55 and 30 kDa (Fig. 3A). By contrast, the sporozoite enriched samples contained numerous bands (Fig. 3A). When a Western blot of the same samples was probed with the fusion protein-specific antiserum, strong reactivity with the 190-kDa band in the oocyst shell track was observed. There was also some indication of reactivity within the sporozoite track (Fig. 3B). These results suggest that the CPR1 gene fragment encodes a peptide constituent of the oocyst wall. The presence of a reacting band in the sporozoite track may have been due to incomplete separation of sporozoites and oocysts, since a few intact oocysts were visible in the sporozoite-enriched sample (not shown).

Characterisation and sequencing of the CPR1 insert. To eliminate the possibility that the CPR1 insert was derived from sheep or *E. coli* DNA, purified CPR1 insert DNA was labelled with digoxigenin and used to probe a Southern blot of *Eco*RI-digested genomic DNA from *C. parvum* sheep and *E. coli*. The probe hybridized to a single 2.4-kb fragment of *C. parvum* DNA, but no hybridization was detected to sheep or *E. coli* DNA (not shown). Sequencing of CPR1 revealed that it consisted of 2359 bp. All six possible reading frames were examined for ORFs. As expected, considering the experimental results, one reading frame contained a continuous ORF in which the *Eco*RI site was in frame with that in pMS1S and λgt11. The other reading frames each contained between 5 and 19 stop codons. The CPR1 DNA sequence and the deduced amino acid sequence of the ORF (P786) are shown in Fig. 4.

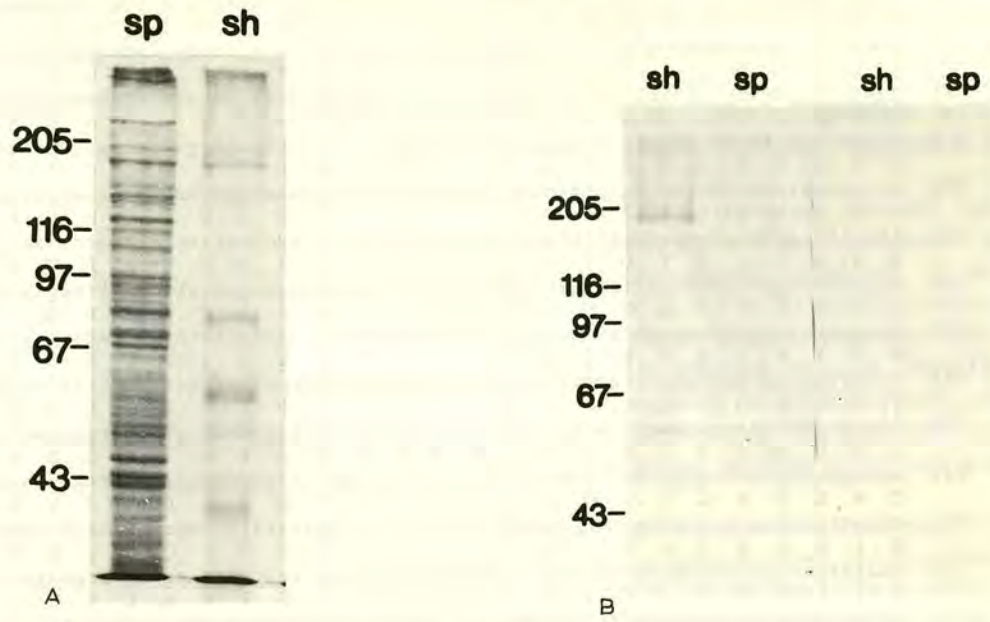


Fig. 3. SDS-PAGE and Western blot analysis of samples enriched for oocyst shells or sporozoites. (A) Proteins extracted from samples enriched for sporozoites (sp) or oocyst shells (sh) were fractionated on a 7.5% SDS-PAGE gel and stained with Coomassie blue. (B) Duplicate samples were Western-blotted and probed with rabbit serum raised against the fusion protein expressed by pCPR1/1S (left panel) or serum taken from the same rabbit prior to inoculation with the fusion protein (right panel).

Protein sequence analysis. The P786 deduced amino acid sequence has an unusual amino acid composition in comparison with the amino acid proportions averaged over the entire NBRF protein sequence database. In particular, 57 (7.3%) of the 786 residues are cysteines, compared with 1.9% for the NBRF database as a whole. There are also unusually high proportions of proline (11.1% compared with 5.2% in NBRF), glutamine (7.4% compared with 4.1%) and histidine (4.1% compared with 2.3%). It is also noteworthy that the cysteine and histidine residues are not evenly distributed throughout the sequence, which can be further described as consisting of alternating cysteine-present/histidine-absent and cysteine-absent/histidine-present regions.

Dot matrix analysis indicates that the amino acid sequence is distinctly repetitive (Fig. 5). In particular, there is a strongly repetitive region within the first cysteine-rich domain spanning residues 1-338 of the available sequence, indicated by the diagonal lines in the top left corner of the dot plot. A further repeating element in the C-terminal half of the sequence is indicated by the diagonal lines in the bottom right-hand corner.

The repeated regions in the P786 sequence are each based on conserved cysteine residues. In the N-terminal cysteine-containing region (residues 1–338), the first 26 cysteines are arrayed in 13 consecutive copies of a repeat unit which consists of the motif CPXG (7X) C, followed by either 9 or 11 residues (denoted a_9 and a_{11}, respectively, in Fig. 6). The first 12 of these repeats are arranged as a series of four higher order repeats, each of 65 amino acids (labelled A in Fig. 6). After the 13th copy of the motif, the remaining 6 cysteines in this region follow on as a fainter echo of the original motif, i.e., CPXG (6X) C (11X) CPXG (8X) C (7X) C (8X) C. The four higher order repeats show complete conservation of the cysteine residues, and in addition 16 of the remaining 59 amino acids are conserved (Fig. 6). Of these conserved residues, the asparagine-

addition, glutamine-rich and cysteine-rich proteins were reported, although such alignments between similarly biased sequences are not readily informative as they often do not indicate a related function [20]. Further use of programs available in the GCG package and at the Biocomputing Research Unit, University of Edinburgh, identified protein sequences containing the (CPXG) 7X C motif. These database sequences included several with more than one copy of the motif: i.e., midge balbiani ring 3 protein, human fibrillin, human von Willebrand factor, leech antistasin and human low density lipoprotein receptor-related protein.

Discussion

C. parvum oocyst wall proteins. The finding of only four visible bands following SDS-PAGE of oocyst shells is at odds with results from other studies. Two studies which utilized ^{125}I-labelling of untreated, intact oocysts, followed by SDS-PAGE of the extracted proteins, identified 17 [21] and 6–10 [22] oocyst surface proteins, respectively. Differences between these and the present study may be due to the effects of trypsin and deoxycholate on oocyst wall proteins during in vitro excystation. However, one of these studies [22] also examined oocyst wall proteins following in vitro excystation and separation on a density gradient. In this study, 7 major bands were identified, one of 190 kDa, which might correspond to the 190-kDa protein identified in the present work. Another difference between these earlier studies and the present work is in the method of extraction of oocysts from the faeces. In our laboratory, routine oocyst extraction involves exposure of the oocysts to dilute sulphuric acid and 1% (w/v) SDS. Other groups have used different extraction procedures. It is also possible that there are isolate-specific differences in oocyst wall proteins. Thus, Lumb et al. [22] found that of 10 ^{125}I-labelled surface proteins of *C. parvum* oocysts, four were not present in all of the isolates studied. Differences in the antigenic composition of sporozoites of different *C. parvum* isolates have also been demonstrated using a panel of monoclonal antibodies [23]. Failure to identify proteins which are homologous to P786, following exhaustive database searching, suggests that P786 represents part of a protein which has been hitherto unrecognised. Molecular studies of other coccidia, such as *Eimeria* and *Toxoplasma*, have been concerned with genes encoding antigens of the motile stages [24–27]. There are no genes encoding oocyst wall proteins of gastrointestinal coccidia in current sequence databases, and an extension of the present work would therefore be to determine whether such proteins are conserved among different coccidial genera.

Codon usage and G+C content in C. parvum. The CPR1 DNA sequence has a relatively low G+C content (39.1%). The G+C content of the first second and third codon positions (53.1%, 46.3% and 17.9%, respectively) of the CPR1 ORF, and inspection of the codon usage data (not shown) confirms that this low G+C content is not due to biased amino acid composition, but rather to preferential use of G+C-poor codons at synonymous sites. This adds to other evidence that the *C. parvum* genome as a whole is low in G+C content, which consists of a similar codon usage bias in the *C. parvum* actin gene [4], comparison of restriction digests using enzymes with G+C-poor and G+C-rich recognition sites [4] and available sequence data [3,4,5,28].

P786 and parasite repeat antigens. A striking feature of P786 is the presence of highly conserved repeats. Proteins with repetitive domains appear to be very common in protozoan parasite antigens [29], particularly in *Plasmodium* [30], and it has been suggested that such domains have evolved as a mechanism of immune evasion [29,31]. However, the repeats found in P786 appear to be different from these other parasite antigen repeats in several respects. Firstly, while the P786 repeat units contain certain residues whose positions

H.W., ed.), pp 43–47. Kluwer, London.
8 Hill, B.D., Blewett, D.A., Dawson, A.M. and Wright, S. (1990) Analysis of the kinetics, isotype and specificity of serum and coproantibody in lambs infected with *Cryptosporidium parvum*. Res. Vet. Sci. 48, 76–81.
9 Huynh, T.V., Young, R.A. and Davis, R.W. (1985) Constructing and screening cDNA libraries in λgt10 and λgt11. In: DNA Cloning (Glover, D.M., ed.), Vol. I, pp. 49–78. IRL Press, Oxford.
10 Scherf, A., Mattei, D. and Schreiber, M. (1990) Parasite antigens expressed in *Escherichia coli*. A refined approach for epidemiological analysis. J. Immunol. Methods 43, 333–341.
11 Carroll, S.B. and Laughon, A. (1987) Production and purification of polyclonal antibodies to the foreign segment of β-galactosidase fusion proteins. In: DNA Cloning (Glover, D.M., ed.), Vol. III, pp. 89–111. IRL Press, Oxford.
12 Marston, F.A.O. (1987) The purification of eukaryotic polypeptides expressed in *Escherichia coli*. In: DNA Cloning (Glover, D.M., ed.), vol. III, pp. 59–88. IRL Press, Oxford.
13 Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. Nature 227, 680–685.
14 Sanger, F., Nicklen, S. and Coulson, A.R. (1970) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
15 Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.
16 Gribskov, M., Luthy, R. and Eisenberg, D. (1990) Profile analysis. Methods Enzymol. 183, 146–159.
17 Coulson, A.F.W., Collins, J.F. and Lyall, A. (1987) Protein and nucleic acid sequence database searching: a suitable case for parallel processing. Comput. J. 30, 420–423.
18 Aswad D.W. and Johnson, B.A. (1987) The unusual substrate specificity of eukaryotic protein carboxyl methyltransferases. Trends Biochem. Sci. 12, 155–158.
19 Dayhoff, M.O. (1978) Atlas of Protein Sequence and Structure. Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington.
20 McQuay, S.J. (1991) The interpretation of alignments between proteins of unusual sequence composition. Biochem. Soc. Trans. 19, 523–524.
21 Tilley, M., Upton, S.J., Blagburn, B.L. and Anderson, B.C. (1990) Identification of outer oocyst wall proteins of three *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) species by ^{125}I surface labelling. Infect. Immun. 58, 252–253.
22 Lumb, R., Lanser, J.A. and O'Donoghue, P.J. (1988) Electrophoretic and immunoblot analysis of *Cryptosporidium* oocysts. Immunol. Cell. Biol. 66, 369–376.
23 McDonald, V., Deer, R.M.A., Nina, J.M.S., Wright, S., Chiodini, P.L. and McAdam, K.P.S.J. (1990) Characteristics and specificity of hybridoma antibodies against oocyst antigens of *Cryptosporidium parvum* from man. Parasite Immunol. 13, 251–259.
24 Clarke, L.E., Tomley, F.M., Wisher, M.H., Foulds, I.J. and Boursnell, M.E.G. (1990) Regions of an *Eimeria tenella* antigen contain sequences which are conserved in circumsporozoite proteins from *Plasmodium* spp. and which are related to the thrombospondin gene family. Mol. Biochem. Parasitol. 41, 269–280.
25 Ossorio, P.N., Schwartzman, J.D. and Boothroyd, J.C. (1992) A *Toxoplasma gondii* rhoptry antigen associated with host cell penetration has unusual charge asymmetry. Mol. Biochem. Parasitol. 50, 1–16.
26 Jenkins, M.C., Lillehoj, H.S. and Dame, J.B. (1988) *Eimeria acervulina*: DNA cloning and characterisation of recombinant sporozoite and merozoite antigens. Exp. Parasitol. 66, 96–107.
27 Prince, J.B., Araujo, F.G., Remington, J.S., Burg, J.L., Boothroyd, J.C. and Sharma, S.D. (1989) Cloning of cDNAs encoding a 28 kilodalton antigen of *Toxoplasma gondii*. Mol. Biochem. Parasitol. 34, 3–14.
28 Laxer, M.A., Timblin, B.K. and Patel, R.J. (1991) DNA sequences for the specific detection of *Cryptosporidium parvum* by the polymerase chain reaction. Am. J. Trop. Med. Hyg. 45, 688–694.
29 Schofield, L. (1991) On the function of repetitive domains in protein antigens of *Plasmodium* and other eukaryotic parasites. Parasitol. Today 7, 99–105.
30 Kemp, D.J., Coppel, R.L. and Anders, R.F. (1987) Repetitive genes and antigens of malaria. Annu. Rev. Microbiol. 41, 181–208.
31 Anders, R.F. (1986) Multiple cross-reactivities amongst antigens of *Plasmodium falciparum* impair the development of protective immunity against malaria. Parasite Immunol. 8, 529–539.
32 Lawler, J. and Hynes, R.O. (1986) The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. J. Cell Biol. 103, 1635–1648.
33 Verweij, C.L., Diergaarde, P.J., Hart, M. and Pannekoek, H. (1986) Full-length von Willebrand factor (vWF) cDNA encodes a highly repetitive protein considerably longer than the mature vWF subunit. EMBO J. 5, 1839–1847.
34 Reid, K.B.M. and Day, A.J. (1989) Structure-function relationships of the complement components. Immunol. Today 10, 177–180.
35 Johnson, D., Lanahan, A., Buck, C.R., Sehgal, A., Morgan, C., Mercer, E., Bothwell, M. and Chao, M. (1986) Expression and structure of the human NGF receptor. Cell 47, 545–554.
36 Ullrich, A., Coussens, L., Hayflick, J.S., Dull, T.J., Gray, A., Tam, A.W., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayers, E.L.V., Whittle, N., Waterfield, M.D. and Seeburg, P.H. (1984) Human epidermal growth factor receptor cDNA sequences and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309, 418–425.
37 Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984) The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell 39, 27–38.
38 Maslen, C.L., Corson, G.M., Maddox, B.K., Glanville, R.W. and Sakai, L.Y. (1991) Partial sequence of a candidate gene for the Marfan syndrome. Nature 352, 334–337.
39 Yan, H. and Chao, M.V. (1991) Disruption of cysteine-rich repeats of the p75 nerve growth factor receptor leads to loss of ligand binding. J. Biol. Chem. 266, 12099–12104.
40 Widdowson, D.C.C., Proffitt, J.A., Jagger, P.S., Richards, A.J. and Hames, B.D. (1990) Developmental expression and characterization of the gene encoding spore coat protein SP60 in *Dictyostelium discoideum*. Mol. Microbiol. 4, 951–960.

are completely (cysteine) or highly (proline, glycine, leucine) conserved, other intervening residues are much more variable. Repeat units in other parasite repeat antigens consist of many tandem copies of more or less identical repeats. Secondly, P786 contains a high proportion of cysteine. By contrast, repeats thought to be involved in immune evasion rarely contain cysteine residues [29]. Thirdly, the P786-β-galactosidase fusion protein was not recognised by serum from any of 16 lambs which had been infected per os with *C. parvum* oocysts (unpublished), whereas parasite repeat antigens which appear to be involved in immune evasion elicit the production of high levels of antibody in infected individuals [29,31]. As a component of the oocyst wall, the native 190-kDa *C. parvum* protein corresponding to P786 is probably not well presented to the immune system, since the oocyst stage is not invasive. Such considerations would therefore appear to rule out the involvement of this protein in any immune evasion mechanism.

Occurrence and function of cysteine-rich repeats. Cysteine-rich repeats have been identified in several plasma proteins, including thrombospondin [32], von Willebrand factor [33] and many components of the complement system [34]. They have also been demonstrated in the extracellular domains of a number of cell surface proteins, such as nerve growth factor receptor [35], epidermal growth factor receptor [36] and low density lipoprotein receptor [37], and in human fibrillin, a glycoprotein component of connective tissue [38]. The common feature of all these proteins appears to be the extracellular location of the cysteine-rich domains, which may have a role in stabilising protein structure through disulphide bonds. Indeed, several of the above mentioned proteins are known to adopt structures which are stabilised by disulphide bonding [33,37,39]. The mature *C. parvum* protein corresponding to P786 may consist of structurally repeated regions stabilised by the presence of many disulphide bonds. In support of this, it has been found that a spore coat protein of the slime mould *Dictyostelium discoideum* contains epidermal growth factor-like repeats in which the positions of the cysteine residues are highly conserved [40]. These residues are highly disulphide-cross-linked in the mature spore coat. The spore coat of *D. discoideum* can be considered analogous to the oocyst wall of *C. parvum*, since each forms the outer covering of an environmentally resistant dispersal stage of the organism. The presence of proteins containing repeats based on conserved cysteine residues in both of these structures indicates that the cysteine residues are likely to perform a similar function in both cases.

Acknowledgements

The authors gratefully acknowledge the advice and support provided by Mr. S.E. Wright, Dr. R.G. Ridley and Prof. R.P. Ambler, and by the late Dr. D.A. Blewett and Prof. J.G. Scaife. This work was financially supported by the Scottish Office Agriculture and Fisheries Department, and additionally by a research studentship from the Agriculture and Food Research Council (NCL) and a research scholarship from the Darwin Trust of Edinburgh (SJM).

References

1 Current, W.L. and Garcia, L.S. (1991) Cryptosporidiosis. Clin. Microbiol. Rev. 4, 325–358.
2 Mead, J.R., Arrowood, M.J., Current, W.L. and Sterling, C.R. (1988) Field inversion gel electrophoretic separation of *Cryptosporidium* spp. chromosome-sized DNA. J. Parasitol. 74, 366–369.
3 Johnson, A.M., Fielke, R., Lumb, R. and Baverstock, P.R. (1990) Phylogenetic relationships of *Cryptosporidium* determined by ribosomal RNA sequence comparison. Int. J. Parasitol. 20, 141–147.
4 Kim, K., Gooze, L., Petersen, C., Gut, J. and Nelson, R.G. (1991) Isolation, sequence and molecular karyotype analysis of the actin gene of *Cryptosporidium parvum*. Mol. Biochem. Parasitol. 50, 105–114.
5 Dykstra, C.C., Blagburn, B.L. and Tidwell, R.R. (1991) Construction of genomic libraries of *Cryptosporidium parvum* and identification of antigen-encoding genes. J. Protozool. 38, 76S–78S.
6 Gooze, L., Kim, K., Petersen, C., Gut, J. and Nelson, R.G. (1991) Amplification of a *Cryptosporidium parvum* gene fragment encoding thymidylate synthase. J. Protozool. 38, 56S–58S.
7 Blewett, D.A. (1988) Cryptosporidiosis in red deer. In: The Management and Health of Farmed Deer (Reid,