
Université d'Ottawa · University of Ottawa

# IDENTIFICATION AND MOLECULAR CHARACTERISATION OF A PUTATIVE GENE FOR CELL DETACHING FACTOR FROM *TRICHOMONAS VAGINALIS*

A Thesis Submitted to the

School of Graduate Studies

University of Ottawa

In Partial Fulfilment of the Requirements for the

Degree of Master of Science

Department of Biochemistry, Microbiology, and Immunology

Faculty of Medicine

By

Kiera L. Delgaty

# ABSTRACT

*Trichomonas vaginalis* (Tv) is the etiologic agent of trichomoniasis, a sexually transmitted disease of significant socio-economic importance. Cell-free filtrate from Tv growth media causes rounding and detachment of adherent monolayers of human epithelial cells in culture. Cell detaching factor (CDF), a 200 kDa glycoprotein in the purified preparation of the cell-free filtrate, is believed to play an important role in the pathogenesis of Tv. The identification of the gene for CDF would therefore allow further characterisation of the protein and its expression. A cDNA clone (CDF-2) was identified in a Tv cDNA library by immunological screen with rabbit anti-serum prepared against a purified preparation of CDF. This clone contained an open reading frame (ORF) that was believed to represent a partial coding sequence for CDF. A genome walk initiated from CDF-2 using Rapid Amplification of Genomic DNA Ends (RAGE) identified a contiguous 6.5 kb ORF. Northern analysis, which identified a 6.5-7 kb transcript, supported the size of the ORF, while Rapid Amplification of cDNA Ends (RACE) identified a putative 5' end of the ORF and identified a 31 bp 3' untranslated region. The ORF is predicted to encode a 232 kDa polypeptide containing 42 EEKPKL repeats, but has no sequence homology with any previously reported protein. The full-length ORF was amplified from Tv genomic DNA by PCR and cloned into the plasmid vector pCR2.1-TOPO. The ORF was transcribed *in vitro* using T7 RNA polymerase and translated in rabbit reticulocyte lysate in the presence of $^{35}$S-methionine. Several high molecular weight proteins (>200, 165, 150, 130, 118, 80, 47 kDa) were produced. Two polypeptides of 205 and 220 kDa were immunoprecipitated from the *in vitro* translation mix by the anti-CDF serum. These data suggest that the gene encoding Tv CDF has been identified.

# ACKNOWLEDGEMENTS

I would like to thank Dr. Gary Garber for allowing me the opportunity to work in his laboratory, and for his financial support throughout the project. He has provided many opportunities that have rounded out my training, and through my experiences in his lab, I have grown both personally and professionally. Most of all, I wish to thank him for believing in me.

Special thanks to Dr. Ken Dimock, who has supervised and guided my research through to success. He has always been enthusiastic, but patient, and he has given me support, reassurance, encouragement, and gentle nudges when I needed them.

I gratefully acknowledge the support of the other members of my Thesis Advisory Committee: Dr. J. R. Dillon who taught me the value of critical thinking, and Dr. Kathryn Wright for her friendly encouragement.

Thanks also to Andre Bergeron for all of his technical assistance, to Dr. Donal Hickey for his help in interpreting the BLAST data, and to Nancy Delcellier for her mentoring.

To the former and present members of the Garber/Cameron lab (Shannon, Isabelle, Dino, Wally, Debbie, Josee, Renuka and Kelley (who is an honorary member): You have all touched my heart in individual ways. Thank you so much for your encouragement, advice, friendship and good times, and for putting up with me during the bad times.

Finally, special thanks to my husband and my family for their strength and support throughout the last four years.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| A | adenosine |
| Ac | acetate |
| 3AEC | 3-amino-9-ethyl carbazole |
| Amp | ampicillin |
| $A_{xx}$ | absorbance at xx= nm |
| AP | adaptor primer |
| bp | base pair |
| BSA | bovine serum albumin |
| BLAST | basic local alignment search tool |
| C | cytosine |
| CDF | cell detaching factor |
| cDNA | complimentary DNA |
| cfu | colony forming units |
| CPE | cytopathic effect |
| cpm | counts per minute |
| DEPC | diethyl pyrocarbonate |
| DMSO | dimethyl sulfoxide |
| DTT | dithiothreitol |
| ds | double stranded |
| EDTA | ethylenediaminetetraacetic acid |
| EIA | enzyme immuno-sorbent assay |

| | |
|---|---|
| ELISA | enzyme-linked immunosorbent assay |
| EMBL | European Molecular Biology Laboratory |
| FBS | fetal bovine serum |
| G | guanine |
| gDNA | genomic DNA |
| GSP | gene specific primer |
| GST | glutathione S-transferase |
| HRPO | horse radish peroxidase |
| INR | initiator |
| IPTG | isopropyl β-D-thiogalactopyranoside |
| kb | kilobase |
| kDa | kilodalton |
| LB | Luria broth |
| M | adenosine or cytosine |
| N | any four of the nucleotides |
| NADH | nicotinamide adenine dinucleotide (reduced) |
| NADPH | nicotinamide adenine dinucleotide phosphate (reduced) |
| NCBI | National Center for Biological Information |
| NGSP/nGSP | nested gene specific primer |
| NIH | National Institutes of Health |
| Oligo | oligonucleotides |
| ORF | open reading frame |
| PAGE | polyacrylamide gel electrophoresis |

| | |
|---|---|
| PBS | phosphate buffered saline |
| PBST | phosphate buffered saline with Tween 20 |
| PCR | polymerase chain reaction |
| PEG | polyethylene glycol |
| pfu | plaque forming units |
| PolyA$^+$ | polyadenylated |
| RACE | rapid amplification of complimentary DNA ends |
| RAGE | rapid amplification of genomic DNA ends |
| rpm | revolutions per minute |
| RT | reverse transcriptase |
| SDS | sodium dodecyl sulphate |
| SM | suspension medium |
| SSC | salt sodium citrate buffer |
| SSPE | salt sodium phosphate EDTA buffer |
| STD | sexually transmitted disease |
| TAE | tris-acetate EDTA buffer |
| TBE | tris-borate EDTA buffer |
| T | thymidine |
| TLCK | Na-P-tosyl-L-Lysine chloromethyl ketone HCl |
| TYI | Diamond's TYI-S-33 medium |
| UTR | untranslated region |
| u | units of enzyme activity |
| VEC | vaginal epithelial cell |

| | |
|---|---|
| W | adenosine or thymidine |
| Y | cytosine or thymidine |
| X-gal | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |

# CHAPTER 1: INTRODUCTION

## WHY IS *T. VAGINALIS* IMPORTANT?

*Trichomonas vaginalis* is a parasitic protozoan that is the etiologic agent of trichomoniasis, a sexually transmitted disease that has important medical, social, and economical implications. Recent data estimate that the annual incidence of *T. vaginalis* infection is more than 170 million cases world-wide. In North America, more than eight million new cases are reported yearly (WHO, 1995). These estimates are likely low, since up to 50% of cases of *T. vaginalis* infection are asymptomatic and remain untreated, allowing further transmission (Fouts & Kraus, 1980).

The disease encompasses a broad range of symptoms ranging from a state of severe inflammation and irritation of the vagina with a frothy malodorous discharge vaginal to a relatively asymptomatic carrier state (Wolner-Hanssen *et al.*, 1989). The infection is linked to perinatal complications (such as premature rupture of the membranes, premature labour and low birth-weight infants), as well as cervical cancer, pelvic inflammatory disease, various inflammatory diseases of the male genito-urinary tract, infertility in both women and men, and HIV transmission (Petrin *et al.*, 1998)

It is not known why the infection is worse in females than in males. Exactly how the organism establishes infection in what seems to be an unsuitably acidic milieu of the vagina is unknown. Furthermore, it is not yet understood how the organism persists in the ever-changing vaginal environment, through hormonal fluxes and menstrual flow. Pathogenesis research has focused on the initial events of infection that permit the organism to invade the urogenital tract.

## TAXONOMY AND STRUCTURE

*T. vaginalis* is an aerotolerant anaerobic flagellated protozoan parasite that was first described by Donné (1836). According to Marquardt and Demaree (1985) and Honigberg (1990), *T. vaginalis* is classified in the following groups: Phylum Sarcomastigophora, Honigberg and Balamuth 1963; Sub-phylum Mastigophora, Diesing 1866; Class Zoomastigophorea, Calkins 1909; Superorder Parabasalidea, Honigberg 1973; Order Trichomonadida, Kirby 1947, emend Honigberg 1974; Family Trichomonadidae, Chalmers and Pekkola 1918, amend Kirby 1947; Subfamily Trichomonadinae, Chalmers and Pekkola 1918, emend Honigberg 1963. Several distinguishing features of *T. vaginalis* have been described (Honigberg & King, 1964; Warton & Honigberg, 1979). The organism possesses 5 flagella, 4 of which are located at its anterior portion. The fifth flagellum is incorporated within the parasite's undulating membrane that is supported by a slender non-contractile costa. Movement of the flagella and the undulating membrane give this parasite a characteristic quivering motility. The nucleus in *T. vaginalis* is located in its anterior portion, and, like other eucaryotes, it is surrounded by a porous nuclear envelope. A slender hyaline, rod-like structure, called an axostyle, commences at the nucleus and bisects the protozoan longitudinally, protruding through the parasite's posterior end, and terminating in a sharp point. This structure is thought to anchor the parasite to vaginal epithelial cells.

While *T. vaginalis* possess typical eucaryotic endoplasmic reticula and Golgi apparatus, the organism lacks mitochondria and peroxisomes of higher eucaryotes. Instead, unique double membrane—bound organelles called hydrogenosomes, which are visible under the light microscope as paracostal and paraxostylar granules, are considered

to be the anaerobic equivalent of mitochondria in *T. vaginalis*. Hydrogenosomes are the site for anaerobic oxidation of pyruvate to acetate, coupled to ATP production via substrate level phosphorylation; electrons are disposed of via the formation of molecular hydrogen (Muller, 1980; Muller, 1993). Key components in this process involve hydrogenase as well as iron-sulphur proteins: pyruvate:ferredoxin oxidoreductase, which uses ferredoxin as an both electron donor and an electron acceptor. Like mitochondria, hydrogenosomes have a double membrane envelope, divide autonomously by fission, import proteins post-translationally, produce ATP by substrate phosphorylation (Muller, 1993; Bradley *et al.*, 1997; Palmer, 1997), and have transmembrane electrochemical potential (Biagini *et al.*, 1997). While similar in structure and function to mitochondria, hydrogenosomes are markedly different in that they do not contain genetic material, they lack cytochromes, and they do not employ the citric acid cycle (Biagini *et al.*, 1997; Bui & Johnson, 1996). Instead, they contain enzymes that are typically found in anaerobic bacteria and are capable of producing molecular hydrogen. The presence of pyruvate:ferredoxin oxidoreductase in eucaryotes is restricted to a few amitochondriate groups, but similar enzymes (pyruvate:flavodoxin oxidoreductase) have been identified in anaerobic archaebacteria and eubacteria. Furthermore, the combination of pyruvate:ferredoxin oxidoreductase, ferredoxin, and hydrogenase is found only in a small number of anaerobic bacteria (Hrdy & Muller, 1995; Muller, 1993). These paradoxes have lead to three hypotheses to explain the origin of the hydrogenosome (Biagini *et al.*, 1997): (1) that hydrogenosomes arose independently of mitochondria through the endosymbiosis of an ancient eucaryote with hydrogenase-containing anaerobic bacterium, (2) that mitochondria and hydrogenosomes originated from a common

progenitor organelle, or (3) that hydrogenosomes are mitochondria that have lost respiration and a separate genome.

*T. vaginalis* was originally thought to have emerged at the base of the eucaryotic tree before mitochondrial endosymbiosis. This evidence was based on small subunit ribosomal RNA phylogeny, homologies of *T. vaginalis* cyclin-dependant kinase and cell division control genes (Riley *et al.*, 1993; Riley & Krieger, 1995), and the enzymes involved in energy metabolism (Hrdy & Muller, 1995; Muller, 1993). However literature that identifies similarities between the hydrogenosome and the mitochondrion with respect to the molecular, physiological and morphological attributes support the second and third hypotheses (Biagini *et al.*, 1997; Palmer, 1997). Recent phylogenetic analyses based on the presence of mitochondrial heat shock proteins and ValRS proteins in *T. vaginalis* (Germot & Philippe, 1999; Roger *et al.*, 1996; Horner *et al.*, 1996; Hashimoto *et al.*, 1998; Bui & Johnson, 1996), malate dehydrogenase (Roger *et al.*, 1999), and aminoacyl synthetase (Brown & Doolittle, 1995), suggest that mitochondrial symbiosis occurred before the *T. vaginalis* lineage diverged from the main eucaryotic tree.

***Reproduction and Life cycle***

It is thought that *T. vaginalis* lacks a cystic stage, since it has only been reported to exist in the trophozoite form (Honigberg & Brugerolle, 1990). Like many other protozoa, the organism generally reproduces by "cryptopleuromitosis"—longitudinal binary fission without the disappearance of the nuclear membrane (Muller, 1990). The organism averages 10 μm long by 7 μm wide (Honigberg & King, 1964), and is polymorphic depending on its environmental conditions. *In vitro*, the protozoan tends to be uniformly ovoid or pear-shaped (Arroyo *et al.*, 1993), but in the presence of vaginal

epithelial cells, the parasite takes on a more amoeboid appearance (Arroyo *et al.*, 1993; Heath, 1981). Several oversized round forms of the trichomonad are known to exist in dividing growth phase culture: those without flagella; those with flagella and a dividing nucleus; and those with flagella and multiple nuclei. Some believe that these forms arise from divisions of the mastigont systems without cytokinesis as a result of certain unfavourable growth conditions, and that they do not represent stages in the life cycle (Honigberg & Brugerolle, 1990). More recent studies suggest that they may be developmental stages preceding the appearance of mononuclear flagellates since they appear to divide by amitotic budding, rather than by the mitotic division of ovoid cells (Abonyi, 1992; 1995). Under unfavourable growth conditions, *T. vaginalis* cells can round up and internalise the flagella. Some believe these forms to be pseudocysts, but it is more likely that they are degenerate forms of *T. vaginalis* since they have not been reported to give rise to normal motile forms (Fari *et al.*, 1985; Honigberg & Brugerolle, 1990).

## EPIDEMIOLOGY

*T. vaginalis* infection has been identified in every continent, and in all racial groups and socio-economic strata. The true prevalence of *T. vaginalis* infection is unknown, since it is not reportable. However, studies indicate that the prevalence rate may be anywhere between 0 and 65% depending on the type of population studied (Lossick, 1990). In 1995, the World Health Organisation estimated that more than 170 million cases occur world-wide each year and that eight million of those occur in North America (Gerbase *et al.*, 1998; WHO, 1995). These estimates may be low, since up to one half of all infections are asymptomatic and are never diagnosed. Asymptomatic *T.*

*vaginalis* infections are of epidemiological importance since the organisms may be transmitted unknowingly through unprotected sexual contact.

*T. vaginalis* is an obligate parasite. Although the trichomonad can survive outside the host for several hours, non-venereal transmission is extremely rare (Heine & McGregor, 1993). The protozoan primarily adheres to the squamous epithelium of the genital tract, but does not invade the tissues. The genital infection can be multifocal; parasites may be found in the vagina, cervix, urethra, bladder, as well as the Skene's, Bartholin's, and periurethral glands of women, and found in the urethra, prostate, epididymis, and semen of men (Krieger, 1995).

## *T. VAGINALIS* INFECTION

### *Clinical manifestations*

*T. vaginalis* infection in women usually occurs during the reproductive years, and rarely occurs before the menarche or after menopause. The incubation period can be from 3 to 28 days, and the infection itself can last weeks to months. Clinical presentation ranges from asymptomatic to acute or chronic disease, although one third of asymptomatic women develop clinical trichomoniasis within six months (Heine & McGregor, 1993). Acute trichomoniasis is characterised by severe pruritis, vaginitis, vulvitis, dyspareunia, and dysuria (Heine & McGregor, 1993; Wolner-Hanssen *et al.*, 1989). Vaginal and vulvar erythema and colpitis macularis (strawberry cervix) are described as pathognomonic for *T. vaginalis* infection, but are generally not observed without the aid of a colposcope (Wolner-Hanssen *et al.*, 1989). The discharge can be profuse, mucopurulent, frothy, malodorous, and yellow or greenish in colour (Wolner-

Hanssen *et al.*, 1989). The pH of the vagina becomes elevated from the normal 4.5 to greater than 5.0.

Symptoms and parasitic load are cyclical and worsen with menstruation, but often improve during pregnancy (Heine & McGregor, 1993; Rein & Chapel, 1975). The former phenomenon may be due, in part, to a rise in pH caused by the presence of menstrual blood; this creates a milieu which is more ideal for trichomonad reproduction (Goode *et al.*, 1994). The presence of increased amounts of iron in menstrual fluid may contribute to the activation of various pathogenic mechanisms (discussed in greater detail in the section on pathogenesis) (Petrin *et al.*, 1998). While the effect of the hormonal cycle on the parasite is largely undefined, estrogen has been shown to have an inhibitory effect on cell detaching factor, a protein involved in pathogenesis (Garber *et al.*, 1991). *T. vaginalis* is associated with many perinatal complications such as premature rupture of the membranes and low birth-weight infants (Hardy *et al.*, 1984; Minkoff *et al.*, 1984; Cotch, 1990), various inflammatory conditions of the female genital tract, cervical erosion, cervical cancer, infertility, (Heine & McGregor, 1993; Wolner-Hanssen *et al.*, 1989; Rein & Chapel, 1975; Rein, 1990) and HIV transmission (Laga *et al.*, 1993; Laga *et al.*, 1994; Cameron & Padian, 1990).

*T. vaginalis* infection is slowly being recognised as an important cause of morbidity in men (Krieger *et al.*, 1993; Krieger *et al.*, 1993; Krieger, 1995). A recent American study has concluded that one third to one half of all male trichomonal infections show signs of mild symptomatic disease such as urethral discharge, urethritis and prostatitis (Krieger, 1995). It is thought that in men, the infection is transient and lasts about 10 days, although the prevalence of spontaneous resolution (13%-36%) and

the consequences of chronic infection remain poorly defined. *T. vaginalis* has also been linked to a variety of complications and inflammatory diseases of the male genitourinary tract, as well as infertility (Krieger, 1995).

***Diagnosis***

Diagnosis is often difficult because the symptoms of trichomoniasis mimic those caused by candidiasis and bacterial vaginosis (Rein & Holmes, 1983). Diagnosis is therefore made on the basis of clinical presentation, examination, and the identification of trichomonads in the urogenital secretions. While more sophisticated techniques have been developed to detect *T. vaginalis* by direct immunoflourescence (Krieger *et al.*, 1988), monoclonal antibody dot blot assays (Gombosova & Valent, 1990), latex agglutination (Carney *et al.*, 1988), oligonucleotide probes (Briselden & Hillier, 1994; Muresu *et al.*, 1994), EIA (Lisi *et al.*, 1988), and PCR (Lin *et al.*, 1997; Heine *et al.*, 1997), their costs are often prohibitive. Wet-mount and the more sensitive "gold standard" broth culture continue to provide cost-effective, efficient, and accurate diagnoses (Carr *et al.*, 1998; Bhatt *et al.*, 1996; Lossick & Kent, 1991). These methods are hampered, however, by the quantity and viability of the organisms. For wet mount to be accurate, the organisms must be motile. Culture sensitivity unfortunately requires a large inoculum. Two recent culture assays, the Plastic Envelope Method (Beal *et al.*, 1992) and the InPouch TV culture system (Levi *et al.*, 1997) which combine culture and visual examination, and also provides a transport medium, have made these gold standards more convenient. Further, analysis of tampon collected specimens (Tabrizi *et al.*, 1997) is promising for diagnosis in remote communities. Fluorescent staining and microscopy can be more convenient than wet-mount since it does not require immediate

observation, but its diagnostic value is similar. It is difficult to identify *T. vaginalis* in Gram stains, but they are useful in eliminating other causes of vaginitis. *T. vaginalis* has been identified in Pap smears, but this method is not considered reliable enough as a diagnostic tool (Rein, 1990). Recent research has therefore focused on developing a test that is cost-effective, fast, accurate, convenient, and that can be universally applied in all communities for both diagnosis and screening; much debate continues in the literature as to whether existing methods meet these expectations.

## TREATMENT

### *Drug regimen*

The treatment of *T. vaginalis* infection was revolutionised in the early 1960's with the introduction of metronidazole, a 5-nitroimidazole. Metronidazole is the only drug currently licensed for use in the Canada and the United States, but other members of the nitroimidazole class are available in other parts of the world. These include tinidazole, ornidazole, secnidazole, nimorazole, and flunidazole; all of these drugs are structurally related, but they vary from each other in pharmacokinetic and therapeutic activities (Gillis & Wiseman, 1996; Lau *et al.*, 1992; Rossignol *et al.*, 1984; Mathisen & Finegold, 1992). The preferred regimen for metronidazole is a single 2 gram oral dose (CDC, 1998; Division of STD Prevention & Control, 1998) because less drug is required than for other regimens and patient compliance is better (Lossick, 1990). Alternatively, 250 mg two to three times a day for 7 days is equally effective, particularly when side effects or prompt reinfection have been associated with the single dose regimen (Lossick & Kent, 1991; Division of STD Prevention & Control, 1998). Current guidelines

recommend concurrent treatment of both partners, regardless of symptoms, to prevent reinfection and reduce transmission (CDC, 1998; Division of STD Prevention & Control, 1998). Treatment success rates range from 90% to 95% and increases when the partner is treated simultaneously (CDC, 1998). Treatment of *T. vaginalis* infection during pregnancy can reduce associated perinatal risks, but metronidazole therapy is reserved until the second trimester, since the effects of this drug on the fetus are not clear. Clotrimazole can provide palliative relief during the first trimester (Division of STD Prevention & Control 1998; Lossick, 1990)

***Mechanism of action of metronidazole***

Metronidazole enters the cell through passive diffusion. The drug itself is inactive but is rapidly reduced anaerobically in the hydrogenosome by pyruvate-ferredoxin oxidoreductase to form nitro radical intermediates. This reductive activation causes a concentration gradient, which allows more of the drug to enter the cell. The short lived, highly reactive nitro radicals cause damage to the cellular DNA, causing death (Moreno *et al.*, 1984; Muller, 1981; Ings *et al.*, 1974; Edwards, 1993)}. The response is rapid; cell division and motility cease within 1 h and cell death occurs within 8 h as seen in cell culture.

***Refractory trichomoniasis—Treatment failure***

While treatment of *T. vaginalis* infection with metronidazole is curative in most cases, treatment failure can be a serious problem because the organisms are so difficult to eradicate. Before the era of antimicrobial resistance, such failures were rare, and were described in detail in case studies (Lumsden *et al.*, 1988). Today, resistance is well

documented, and it is believed that 5% of isolates display some degree of metronidazole resistance (Narcisi & Secor, 1996). Treatment protocols for refractory cases and resistant organisms are described (Lossick, 1990; Ahmed-Jushuf *et al.*, 1988; Lewis *et al.*, 1997). Organisms with high levels of resistance may require prolonged treatment with very high, toxic doses of therapy, often administered in an oral/vaginal combination or sometimes intravenously to reduce gastrointestinal side effects (Lossick, 1990); this is effective only 80% of the time (Narcisi & Secor, 1996). Unfortunately, there are no good alternatives to metronidazole for the treatment of refractory *T. vaginalis* infection. Alternative local therapies (creams, pessaries, and douches) are described in numerous case studies (Lewis *et al.*, 1997; British Co-operative Clinical Group, 1992) but are often anecdotal in nature and may be palliative rather than curative. It has become apparent that new chemotherapeutic agents for the treatment of *T. vaginalis* infection are needed.

***Mechanisms of resistance***

Ferredoxin and pyruvate:ferredoxin oxidoreductase play a critical role in transferring an electron to the nitro group of metronidazole to form the free radical. Oxygen is an important competitor for any available electrons. It is hypothesised that resistant strains of *T. vaginalis* have poorly functional hydrogenosomal oxidases (oxygen scavengers) resulting in an increased amount of oxygen concentration in resistant strains. This presence of oxygen results in little reduction of the nitro group, since it competes with the drug for electrons. Alternatively oxygen may oxidise the nitro-free radicals to the original compound and be reduced to become a superoxide anion with decreased amount of cytotoxic nitro-radical production (Lewis *et al.*, 1997; Ellis *et al.*, 1994). Oxygen was also found to inhibit the accumulation of [$^{14}$C]-labelled metronidazole in

resistant isolates. This process is known as futile cycling and the damage to the cell is limited due to the presence of intracellular oxygen (Perez-Reyes *et al.*, 1980). This deficiency is attributed to decreased hydrogenase activity and lower $H_2$ production. In aerobic assays investigators have demonstrated minimal lethal concentrations (MLC) of metronidazole *in vitro* tend to be higher in resistant strains than in susceptible ones; the degree of resistance varies among different isolates (Muller & Gorrell, 1983).

Researchers have been able to induce *T. vaginalis* aerobic resistance to metronidazole by culturing parasites in TYM medium containing sublethal concentrations of the drug (Tachezy *et al.*, 1993). These organisms were found to retain their pyruvate:ferredoxin oxidoreductase activity; the cause for aerobic resistance in these organisms has been attributed to the affinity of trichomonad oxidases to $O_2$. In another strain, a point mutation in the ferredoxin gene caused reduced transcription, thereby decreasing the amount of ferredoxin in the cell, and decreasing the ability of the cell to activate the drug (Quon *et al.*, 1992).

Little has been published on anaerobic resistance mechanisms in *T. vaginalis*. Kulda *et al.* (1993) were able to induce anaerobic resistance to metronidazole *in vitro* by cultivating parasites over a period of 12-21 months under increased drug pressure (1-100mg/mL). These strains lacked pyruvate:ferredoxin oxidoreductase and hydrogenase activity and failed to take up any [$^{14}$C]-metronidazole. This finding indicated that the pyruvate oxidising pathway that is necessary for drug activation was nullified, causing a subsequent cessation of metronidazole metabolism. This form of resistance is unlikely to occur in wild type strains; aerobic resistance is considered to be more clinically relevant.

***Use of susceptibility tests***

Much of the controversy regarding the possibility of truly resistant *T. vaginalis* came about because researchers relied on the anaerobic susceptibility assay, which was later found to be a poor measure of resistance. Metronidazole is used therapeutically under aerobic conditions, and this is thought to be a better measure of resistance in clinically resistant organisms. Thus, *in vitro* susceptibility tests for *T. vaginalis* are best reflected by aerobic cultivation. *T. vaginalis* should be tested for susceptibility to metronidazole since this can explain clinical failure when compliance is not an issue, and may dictate the next course of treatment. Unfortunately, the minimum lethal concentrations in susceptibility assays do not reflect the serum drug concentrations needed for cure (Lossick *et al.*, 1986), but can provide assistance in estimating the dosage of drug likely to be curative (Lossick & Kent, 1991; Muller *et al.*, 1988).

## HOST DEFENSE MECHANISMS

Like many protozoal infections, *T. vaginalis* infection does not appear to confer immunity, since repeat infections are common. Further, there exists no evidence to suggest that such repeated infections induce acquired resistance, a phenomenon observed in some protozoal infections (Ackers, 1990). Serologic tests have confirmed the presence of specific anti-*T. vaginalis* antibody during infection, but antibody titres dwindle after eradication of the organism (Sharma *et al.*, 1991). The role of cell-mediated immunity also remains undefined (Abraham *et al.*, 1996). Humoral immunity to *T. vaginalis* infection may be more important in stimulating non-specific immune responses which play an important role in restricting the organism to the genito-urinary tract (Ackers, 1990). Non-specific responses to *T. vaginalis* infection include activation of complement

via the alternative pathway (Gillin & Sher, 1981; Demes *et al.*, 1988) and chemotaxis followed by phagocytosis by neutrophils (Shaio & Lin, 1995). Macrophages, which are known to be important in combating intracellular protozoan parasites, have been shown to be efficient killers of *T. vaginalis in vitro* (Mantovani *et al.*, 1981), but their role in combating this organism in the genital tract is unknown.

It is not clear why this infection in males is transient and clinically much milder than that seen in women. The presence of zinc in the prostatic fluid has been shown to be cytotoxic to *T. vaginalis* (Krieger & Rein, 1982; Daly *et al.*, 1990), and may be a factor in inhibiting infection. Also, it is hypothesised that the reducing environment of the vagina is adequate for the activation of various proteinases involved in pathogenesis that are thought to cause symptoms. In contrast, the male genital tract is oxidative, and may inhibit such mechanisms (Alderete & Provenzano, 1997). The ecology of the vaginal tract during the reproductive years is thought to play a large role in the suitability of the vagina as a niche, since the infection in the premenarchal girl and post-menopausal woman is generally mild and transient (Rein, 1990). The presence of *Lactobacillus acidophilus* in the vagina, the pH of the vagina, estrogen fluctuation, and iron availability specifically in menstrual flow in women of reproductive age are all thought to play a role in pathogenesis. In low iron conditions, the parasite is thought to migrate deeper into the female genital tract toward richer sources of iron, causing a persistence of infection despite the flushing action of mucosal secretions, desquamation of the mucosal epithelium and nutrient limitations and fluxes (Rein, 1990). Exactly how the organism establishes infection in what seems to be an unsuitably acidic milieu is unknown. For these reasons, pathogenesis research has focused on the initial events of infection that

permit the organism to invade the urogenital tract and persist in the ever-changing vaginal microenvironment. During menstruation, the environment undergoes a dramatic change with the influx of erythrocytes, host macromolecules, and serum constituents, as well as large changes in pH. It is remarkable that the parasite not only survives these dramatic environmental changes, but infection persists. The pathogenic mechanisms involved in the parasite's ability to deal with stresses in its host environment are also discussed.

## PATHOGENESIS

The host-parasite interrelationship is very complex and the broad range of clinical symptoms cannot be attributed to a single pathogenic mechanism. Many mechanisms are thought to be involved in the pathogenesis of *T. vaginalis*. These include interaction of *T. vaginalis* with the vaginal flora, immune evasion, cell-to-cell adhesion, haemolysis, and the excretion of soluble factors such as extracellular proteinases and cell detaching factor. The surface of the trichomonad plays a major role in adhesion, host-parasite interaction, and nutrient acquisition, with surface proteins and glycoproteins having functions in this regard. Molecular analysis of these membrane macromolecules and virulence factors has really just begun. A clearer understanding of the pathogenesis of *T. vaginalis* could lead to the development of new intervention strategies, better treatment, and better disease control.

### *Interaction with vaginal flora*

The establishment of *T. vaginalis* in the vagina is puzzling indeed, since the normal pH of the vagina is a very acidic 4.5, while this organism thrives in a less acidic

pH of >5. The rise in pH in the vagina following *T. vaginalis* infection is also marked with a concomitant reduction (or complete loss) of *L. acidophilus* and an increase in anaerobic bacteria. However, the relationship between lactobacilli and *T. vaginalis* is not completely understood. It appears that *in vitro*, under a controlled pH, lactobacilli do not affect the growth of *T. vaginalis*; however, the parasite seems to have a deleterious effect on *L. acidophilus* (McGrory *et al.*, 1994). Several mechanisms have been proposed: *T. vaginalis* has been observed to phagocytize bacteria (Garcia-Tamayo *et al.*, 1978; Street *et al.*, 1984), including lactobacilli (Rendon-Maldonado *et al.*, 1998). Another hypothesis is that products, such as CDF or proteinases secreted by *T. vaginalis*, may destroy the lactobacilli (McGrory & Garber, 1992).

***Immune evasion***

Like many protozoan parasites, the ability of *T. vaginalis* to evade the host immune system is an important aspect of pathogenesis. Avoidance of complement is one such tactic that is used by *T. vaginalis*. It has long been known that *T. vaginalis* activates the alternative pathway of complement (Gillin & Sher, 1981) but *T. vaginalis* has taken advantage of a niche in which there is little complement present. Cervical mucus is surprisingly deficient in complement (Demes *et al.*, 1988; Alderete *et al.*, 1995). Menstrual blood represents the only source of complement available to the vagina. Interestingly, its complement activity is about half that of venous blood and about 1/3 of menstrual blood samples have no complement activity at all (Demes *et al.*, 1988; Alderete *et al.*, 1995), although menstrual blood has an appreciable complement-mediated cytotoxicity toward *T. vaginalis*. While reduction in parasite numbers is seen during menses, trichomonal infection persists (Demes *et al.*, 1988), and it is believed that

the virulence factors discussed below appear to contribute to the exacerbation of symptoms seen at this time.

It was found that iron was a contributing factor in complement resistance. Demes *et al.* (1988b) found that fresh isolates of *T. vaginalis* differ in their susceptibility or resistance to complement-mediated lysis in serum and it appears that complement-resistant fresh isolates become susceptible to complement after prolonged *in vitro* cultivation. This is consistent with the hypothesis that phenotypic variation (discussed below) allows the trichomonad to avoid lysis by complement (Alderete *et al.*, 1995). Resistance to complement is dependent upon a high concentration of iron (Alderete *et al.*, 1995), a nutrient which is indeed abundant during menses. It appears that iron upregulates the expression of cysteine proteinases, which have been found to degrade the C3 portion of complement on the surface of the organism; this allows the organism to evade complement destruction (Alderete *et al.*, 1995).

Like many other protozoan parasites, *T. vaginalis* employs phenotypic variation as a mechanism of immune evasion. Alderete *et al.* (1992) found that two classes of markers are alternately expressed on the surface of the organism: the highly immunogenic glycoproteins (P270) and the adhesins (AP65, AP51, AP33, and AP23) (Alderete, 1988). While all isolates (type I and type II) synthesise P270, only type II organisms can undergo phenotypic variation between cytoplasmic and cell surface expression of P270 (Alderete, 1988). Thus the phenotypes are described as A+B- (P-270 positive) and A-B+ (P-270-negative) (Alderete, 1988). The positive phenotype lacks adhesins and cannot cytoadhere or parasitise host cells (Alderete, 1988). Only the negative phenotype, which expresses the adhesins, has the ability to cytoadhere

(Alderete, 1988; Alderete *et al.*, 1986). It seems that *in vivo*, an antigenic shift occurs from the positive to the negative phenotype (Alderete *et al.*, 1986; Alderete, 1987; Alderete *et al.*, 1987). After prolonged cultivation *in vitro*, the organisms shift toward the positive phenotype (Alderete, 1988), reinforcing the idea that the organism presents the negative phenotype as required under environmental pressure.

The P-270 glycoprotein has been shown to have a single repetitive DREGRD epitope, which is important for antibody binding (Dailey & Alderete, 1991), and it is found that some organisms bearing the P270 antigen on the cell surface are susceptible to antibody-mediated, complement-independent lysis *in vitro* (Alderete & Kasmala, 1986). The lack of P-270 surface expression and the low antigenicity of the adhesins permit the negative phenotype organisms to avoid antibody attack (Alderete, 1987). The AP65 adhesin protein appears to mimic the structure of malic enzyme (Alderete *et al.*, 1995; Engbring *et al.*, 1996), while AP51 shows partial homology to the β-subunit of succinyl-CoA synthetase (Alderete *et al.*, 1998), which may account for their poor immunogenicity. This molecular mimicry is yet another example of how the trichomonad can escape detection by the immune system (Engbring *et al.*, 1996).

Another high molecular weight immunogen, P-230, is present on the surface of all parasites, but undergoes conformational changes that allow it to evade antibody (Alderete *et al.*, 1987). This was termed epitope phenotypic variation. Vaginal IgG, which is reactive to P-230 of *T. vaginalis* was not cytolytic, even in the presence of complement (Alderete *et al.*, 1991). Furthermore, the antibody response was restricted to only one, or a few epitopes on the 230 kDa protein (Alderete *et al.*, 1991), allowing the organism to evade opsonization.

In addition to these mechanisms described above, *T. vaginalis* has numerous other ways of evading the immune system. Provenzano & Alderete (1995) and Min *et al.* (1997) have reported that cysteine proteinases secreted by *T. vaginalis* degrade IgG, IgM, and IgA, which allows the organism to destroy antibody, thereby avoiding antibody attack. *T. vaginalis* also secretes copious amounts of highly immunogenic soluble antigens (Alderete & Garza, 1984). A continuous release of these antigens may neutralise antibody, thus short-circuiting specific anti-*T. vaginalis* defence mechanisms (Alderete & Garza, 1984). As well, *T. vaginalis* can coat itself with host plasma proteins, preventing the host immune system from recognising the parasite as foreign, thus preventing complement-mediated lysis (Peterson & Alderete, 1982). Finally, *T. vaginalis* was documented, by electron microscopy, to phagocytose leukocytes (Rendon-Maldonado *et al.*, 1998).

***Adherence and Adhesins***

Adhesion of trichomonads to the epithelial cells in the vaginal environment is a critical step in the pathogenesis of the parasite (Alderete & Garza, 1985; Alderete & Pearlman, 1984; Krieger *et al.*, 1985; Heath, 1981). Attachment to cells is time, temperature, and pH dependent. *T. vaginalis* appears more inclined to parasitise vaginal epithelial cell lines than other cell types *in vitro*. This is not surprising since epithelial cells are likely to be the principle cell type with which the parasite would interact *in vivo* (Alderete & Garza, 1985).

The adhesion of the parasite to the epithelial cell seems to be mediated by four adhesion proteins: AP65, AP51, AP33, and AP23 (Arroyo *et al.*, 1992; Engbring *et al.*, 1996), which act in a specific receptor-ligand fashion (Alderete & Garza, 1988; Arroyo *et*

*al.*, 1993; Arroyo *et al.*, 1992). At least three genes encode AP65 molecules (Alderete *et al.*, 1995; O'Brien *et al.*, 1996) which are similar to malic enzyme (Engbring *et al.*, 1996; O'Brien *et al.*, 1996). The adhesins are alternately expressed on the surface with P270 (Alderete, 1988; Alderete & Garza, 1988; Alderete *et al.*, 1992) as discussed above. Gene expression of the four adhesins is co-ordinately regulated at the transcriptional level by iron, and such regulation may be a mechanism by which the parasite adapts to the constantly changing environment of the vagina (Lehker *et al.*, 1991; Alderete *et al.*, 1995). Contact-initiated ameboid transformation, which is marked by the production of pseudopodia followed by the upregulation of adhesin synthesis, suggests evidence for a sophisticated signal transduction system (Arroyo *et al.*, 1993). It is thought that the adhesins are concentrated on the side opposite the undulating membrane and the recurrent flagellum since the parasite attaches itself to epithelial cells on that side (Alderete & Garza, 1985). Microfilaments become concentrated in the parasite on the side that is in contact with the vaginal epithelium (Rasmussen *et al.*, 1986).

Adhesion seems to also require the presence of cysteine proteinases (Arroyo & Alderete, 1989). Pretreating trichomonads with Na-P-tosyl-L-lysine chloromethyl ketone HCl (TLCK), a cysteine proteinase inhibitor, caused a marked decline in the parasite's ability to adhere to epithelial cells. When cysteine proteinase was added to TLCK-treated cells, the parasite's ability to attach itself to the host was restored (Arroyo & Alderete, 1989). This indicates that proteinase action on the parasite's surface is a prerequisite for host attachment.

While the adhesins are concentrated on the side opposite the undulating membrane, laminin-binding proteins are ubiquitous on the entire surface of *T. vaginalis*

(Costa e Silva Filho *et al.*, 1988). Laminin, a glycoprotein localised in the basement membrane of the epithelium, promotes cell adhesion, differentiation, morphological variation, and motility in normal cells, and it has been shown to have chemotactic properties (Costa e Silva Filho *et al.*, 1988). *T. vaginalis* has been observed adhering to laminin-coated plastic (Costa e Silva Filho *et al.*, 1988), and endocytosing laminin-covered polystyrene particles (Benchimol *et al.*, 1990).

*T. vaginalis* also has receptors for another extracellular matrix adhesion glycoprotein, fibronectin, which is secreted in both the basement membrane and serum. Although trichomonads can become coated with host fibronectin (and other serum proteins) (Peterson & Alderete, 1982), it is unclear whether fibronectin receptors function in nutrient acquisition (Peterson & Alderete, 1982; Peterson & Alderete, 1984), adherence (Thomas *et al.*, 1985; Alderete *et al.*, 1994; Peterson *et al.*, 1987), or a combination of the two (Gold & Ofek, 1992).

The group of Honigberg (Choromanski *et al.*, 1985; Warton & Honigberg, 1983; Warton & Honigberg, 1980) characterised lectin-binding carbohydrates on the surface of the trichomonad cell, after Cappuccinelli *et al.* (1975) reported their role in adherence to glass. The presence of surface carbohydrates (D-lactose and N-acetyl-D-glucosamine) appears to be correlated with virulence (Choromanski *et al.*, 1985; Warton & Honigberg, 1980; Kon *et al.*, 1988). Lectins that bind to N-acetyl-D-glucosamine and α-D-mannose were recently shown to inhibit *T. vaginalis* binding to host epithelial cell (Mirhaghani & Warton, 1998). Surface carbohydrate moieties also seem to be involved in *T. vaginalis* haemolysis of erythrocytes (de Carli *et al.*, 1994), phagocytosis of the target cells (Benchimol & de Souza, 1995), and may be associated with drug susceptibility (Dias

Filho *et al.*, 1992). On the other hand, endogenous lectins on the surface of the trichomonad cell may be important in adhesion (Roussel *et al.*, 1991; Martinotti *et al.*, 1986).

Neuraminidase seems to be found both on the surface of the trichomonad cell and secreted in the culture media (Costa e Silva Filho *et al.*, 1989; Meysick *et al.*, 1996). Cleavage of sialic acid on the surface of host cells may be important for adhesion (Bonilha *et al.*, 1995; Meysick *et al.*, 1996), but sialic acid present on the surface of *T. vaginalis* does not seem to be involved (Bonilha *et al.*, 1995).

There appears to be a number of factors involved in adherence. However, adherence does not correlate directly with virulence, since virulent strains isolated from symptomatic patients exhibited wide differences in their ability to adhere to host cells (Krieger *et al.*, 1990). This illustrates the complex host-parasite interrelationships of *T. vaginalis*.

***Haemolysis***

*T. vaginalis* lacks the ability to synthesise de novo purines, pyrimidines, and fatty acids, and must rely on its environment to supply these nutrients. Since the vaginal mucosa may be a poor nutritional milieu for microbes, erythrocytes may be a prime source of nutrient acquisition. In addition to lipids, iron is an important nutrient for *T. vaginalis*, which may also be acquired via the phagocytosis and lysis of erythrocytes (Fiori *et al.*, 1993; Rendon-Maldonado *et al.*, 1998; Lehker *et al.*, 1990).

Lysis of the erythrocytes appears to be mediated by protein receptors on the surfaces of both erythrocytes and parasites (Fiori *et al.*, 1993), and empirical evidence suggests that perforin-like proteins may be involved. Five adhesin molecules are believed

to be associated with haemolysis; two of them specific for erythrocytes (Fiori *et al.*, 1993; Arroyo *et al.*, 1992). Haemolysis is thought to occur in three steps. A specific ligand-receptor interaction allows the trichomonad to attach itself to the erythrocyte. This is followed by the release of perforin-like proteins (possibly cysteine proteinases) that form pores in the erythrocyte membrane. Cysteine proteinase inhibitors greatly reduced erythrocyte lysis, which suggests that cysteine proteinases may be a lytic factor involved in haemolysis (Dailey *et al.*, 1990). Finally *T. vaginalis* detaches itself from the cell and cell lysis occurs. Recently, a 60 kDa cysteine proteinase was shown to degrade haemoglobin (Min *et al.*, 1997). *In vitro*, the rate of haemolysis is greatest at the normal vaginal pH of 4.5 suggesting that the conditions within the vaginal microenvironment (Fiori *et al.*, 1993) are suitable for such activity. Haemolytic activity appears to be correlated with virulence (Krieger *et al.*, 1983).

***Proteinases***

Characteristics of proteinases have been well summarized by North (1994, 1982) and extensive work has been done in isolating and purifying the proteinases of *T. vaginalis* (Lockwood *et al.*, 1987; Irvine *et al.*, 1993). *T. vaginalis* has been found to have between 11 and 23 distinct cysteine proteinase activities, most of which are lysosomal in nature (Neale & Alderete, 1990; Arroyo & Alderete, 1989; Provenzano & Alderete, 1995; Garber & Lemchuk-Favel, 1989; Garber & Lemchuk-Favel, 1994). It is not surprising that these enzymes play a role in the pathogenesis of the parasite. Irvine *et al.* (1997) have used various cysteine proteinase inhibitors in an attempt to determine which cysteine proteinases are essential to parasite survival. Certainly, the reducing environment of the vagina provides the appropriate milieu for the activation of cysteine

proteinases (Alderete *et al.*, 1991). A number of functions of cysteine proteinases have been discussed. Cysteine proteinases are thought to function as probable lytic factors in the haemolysis of erythrocytes, and they degrade haem. In addition to this, cysteine proteinase activity is required for the adherence of *T. vaginalis* to epithelial cells (Arroyo & Alderete, 1989). Finally, *T. vaginalis* cysteine proteinases have been shown to degrade host immunoglobulins (Provenzano & Alderete, 1995), a mechanism of immune evasion.

***Cell Detaching Factor***

Although contact-dependent mechanisms play a significant role in the pathogenesis of *T. vaginalis*, contact-independent mechanisms are also involved. The first to report on contact-independent mechanisms was Hogue (1943), who noted that cell-free filtrates had adverse effects on cells in culture, similar to the organism itself. Others (Alderete & Garza, 1984; Nielsen & Nielsen, 1975; Farris & Honigberg, 1970; Honigberg, 1978; Honigberg & Ewalt, 1961) have also proposed the idea that some soluble cytotoxin may play a role in pathogenesis. Haemolysis and cytotoxicity, for example, cannot be explained solely by the contact-dependent mechanisms, since these effects can be seen in the absence of cell-to-cell contact (Garber & Bowie, 1990; Roussel *et al.*, 1991). While pH and acidic metabolites can be partly responsible for these effects (Garber & Bowie, 1990; Pindak *et al.*, 1993), the organism has been shown to produce other factors which cause cytopathic effect.

A secreted factor, cell detaching factor (CDF), found in cell-free filtrate of *T. vaginalis* growth media, was shown to cause cytopathic effects in cell culture (Garber *et al.*, 1989; Pindak *et al.*, 1986). When the cell-free filtrate of a *T. vaginalis* culture is applied *in vitro* to a cell culture monolayer, the cells of the monolayer detach and clump

together, but remain viable. The detachment of the cell monolayers *in vitro* is thought to be analogous to the sloughing of vaginal epithelial cells seen in the vaginal mucosa during acute infection (Garber *et al.*, 1989). Also supporting the theory that CDF is probably a factor in pathogenesis, is the observation that *Pentatrichomonas hominis*, a non-pathogenic species, does not show CDF activity (Garber *et al.*, 1989).

CDF, which is thought to be an extracellular factor (Garber *et al.*, 1989), was found to be a 200 kDa glycoprotein (Garber *et al.*, 1989) that is heat and acid labile (Garber *et al.*, 1989; Pindak *et al.*, 1986). The concentration of CDF in the filtrates varied with three factors: the duration of *T. vaginalis* growth prior to filtrate preparation, the initial inoculum size, and the pH of the filtrate prior to harvesting (Garber *et al.*, 1989). Activity was also found to be affected by pH: Pindak *et al.* (1986) found the optimum pH of the cell-free filtrate to be around 7.2. Garber *et al.* (1989), however, found that purified CDF was active within pH 5.0 to 8.5, with the optimum activity at pH 6.5, and inactive below pH 4.5. This is of clinical relevance since the normal pH of the vagina is 4.5 but is greater than 5.0 during trichomoniasis. The rise in vaginal pH during trichomoniasis may, therefore, be crucial in the pathogenesis of the organism.

CDF levels have been shown to correlate with the severity of the clinical symptoms of vaginitis. Increasing production of CDF was associated with increased severity of clinical disease (Garber & Lemchuk-Favel, 1990). A human serum reactive to *T. vaginalis* inactivates CDF activity (Garber *et al.*, 1989), suggesting that local vaginal antibodies could decrease CDF effects (Garber & Lemchuk-Favel, 1990). It is not certain whether it is the regulation of CDF production (i.e. its concentration), its activity, its immunogenicity, or a combination of the three that plays a role in the severity of

symptoms. Indeed, all of the pathogenic mechanisms (i.e. contact-dependent, contact-independent, and host immune response) are likely important in the virulence of this disease.

CDF production is likely to be influenced by the concentration of estrogen in the vagina. *In vitro*, the production of CDF by *T. vaginalis* has been shown to decrease in the presence of β-estradiol. The maximal decrease was shown at a β-estradiol concentration of $10^{-7}$ to $10^{-8}$ M, which is of clinical relevance, since human systemic levels of β-estradiol are in the range of $10^{-8}$ to $10^{-9}$ M (Garber *et al.*, 1991). This finding may explain some of the etiology of the disease; i.e. the worsening of symptoms around the time of menses when estrogen levels are lowest. As well, it may explain why the application of estradiol pellets intravaginally seems to ameliorate the symptoms, without eradicating the infection (Lirosi & Guarascio, 1972).

Several investigators have been unable to demonstrate a cytopathic effect from cell-free filtrates of *T. vaginalis* culture (Rasmussen *et al.*, 1986; Alderete & Pearlman, 1984; Christian *et al.*, 1963; Krieger *et al.*, 1985; Kulda, 1967). *T. vaginalis* is known to excrete lactic and acetic acids in cell culture, and so unless the pH of the cell culture is maintained, the pH drops below 4.5 (Garber & Lemchuk-Favel, 1990; Garber *et al.*, 1989; Pindak *et al.*, 1986). The intolerance of CDF to acidic pH may explain why these investigators could not demonstrate its activity (Garber *et al.*, 1989; Pindak *et al.*, 1986; Garber & Bowie, 1990).

## EXPERIMENTAL RATIONALE/OBJECTIVES

While *T. vaginalis* is recognised to have important social and medical implications, very little is understood about its pathogenesis. Cell detaching factor is

thought to play an important role in the pathogenic mechanism of *T. vaginalis*; understanding its function may help elucidate how the organism invades an inhospitable female genital tract, which then modifies and adapts to it. It is thought that this process is responsible for the differences in presentation of the disease between men and women. Since estrogen has been shown to influence CDF, expression studies of estrogen on CDF may further elucidate its pathogenic role.

The goal of this work is to identify and characterise the gene for CDF and its regulatory elements. Such knowledge will provide the groundwork for study of the regulation and expression of CDF, and ultimately, allow for further investigation into its pathogenic role. As a first the objective of this project was to sequence and characterise the gene for CDF.

Initial groundwork for this project involved screening a λgt11 cDNA expression library with rabbit anti-serum prepared against a purified preparation of CDF. Two clones were identified from this screening: CDF-1 and CDF-2. The CDF-1 clone was excluded as a candidate for CDF; it was identified as a gene for n-acetylneuraminidate lyase, a protein that did not possess CDF properties (Meysick, 1996).

CDF-2, a 3062 bp cDNA clone, was characterised as having a 3060 bp open reading frame containing no apparent start or stop codons and no initiator elements (Meysick, 1996). In a non-redundant database search (BLAST n), the ORF was reported to have no sequence homology with any known nucleic acids. In a non-redundant BLAST x search, the translated ORF was reported to have very weak sequence homology to human nucleolar phosphoprotein p130, and the similarity was primarily to repetitive regions. The clone was characterised as having 37 EEKPKL repeated polypeptides in

one of two major motifs: EEKPKL$^{G}_{S}$LNL$^{G}_{P}$KSPS or EEKPKLQLGGIKL. Further, a sequence encoding a 79 amino acid tandem repeat was also found between nucleotides 1344-1871. Despite the clone's lack of homology to known sequences, the considerable length of the ORF supported its potential to encode a polypeptide; CDF-2 clone was thought to be an incomplete cDNA clone that coded for CDF. Further characterisation of the clone and identification of the complete open reading frame at that time, however, were hampered by the repetitive nature of the sequence (Meysick, 1996).

Based upon the hypothesis that the CDF-2 cDNA clone represented a partial coding sequence of CDF, the objective of this master's project was to identify the missing genetic components in the sequences flanking CDF-2 and determine if the complete sequence represented by CDF-2 was the gene for CDF. The first objective was to be done by identifying genomic DNA clones by screening an *Mbo*I EMBL3 genomic DNA library with CDF-2. From this, sequence analysis was to be used to identify the open reading frame, the translational start and stop codons, and the transcriptional initiation and termination sites. In actuality, several approaches had to be taken, including screening of the genomic DNA library, inverse PCR, and RAGE PCR before DNA clones could be identified and sequenced. As the project proceeded, time constraints necessitated re-evaluation of the second objective. At that time, the second objective was scaled down; it was decided to determine if the complete CDF-2 clone had the potential to encode the CDF gene. This objective was to be completed by cloning the ORF, and performing *in vitro* transcription and translation to determine if the resultant protein displayed properties that were consistent with CDF.

# CHAPTER 2: MATERIALS AND METHODS

## INTRODUCTION

This chapter describes the methods of all the procedures used to identify and characterise a candidate gene for CDF, starting from a sequence of the cDNA clone, CDF-2, which was isolated from a *T. vaginalis* cDNA expression library by immunodetection using rabbit anti-serum prepared against a purified CDF preparation. Originally, the whole clone was used as a probe to screen a *T. vaginalis* EMBL3 genomic DNA library. However, in light of the fact that this work had been plagued for many years by the cross-reactivity caused by the repetitive sequences of the CDF-2 clone, it was later decided that a more specific probe could be prepared by isolating the first 700 base pairs (bp) of the clone, which was non-repetitive DNA. The attempt to identify the flanking regions of CDF to elucidate the start and stop codons by first screening a genomic library and later by rapid amplification of genomic DNA ends, and the use of Southern blotting and restriction mapping as a diagnostic tool for these methods, is described herein. Furthermore, the use of Northern Blot analysis, Primer extension analysis, and rapid amplification of cDNA ends to characterise the cDNA molecule are also described. Finally, *in vitro* transcription of the full length genomic DNA open reading frame, and the translation of the resulting RNA molecule in order to characterise the protein produced by the identified sequence are described.

## PARASITES

Two pathogenic, moderately virulent, clinical isolates of *T. vaginalis*, which were

grown to axenic culture from the vaginal secretions of women with symptomatic trichomoniasis, were used during this study. Isolate 202 was used for all experiments except for construction of five of the six Genome Walker DNA libraries, which were a generous gift from Mr. Dino Petrin. These five libraries were constructed using DNA from isolate DG. Both isolates were classified as moderately virulent based upon clinical presentation of the disease and the subcutaneous mouse assay (Garber & Lemchuk-Favel, 1990).

Organisms were grown from frozen stock in 16 x 125 screw cap tubes containing 10 mL of Diamond's TYI-S-33 medium (TYI) (pH 6.2) (Diamond *et al.*, 1978) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco/BRL, Burlington, ON), 100 units/mL penicillin, 100 µg/mL streptomycin (Penicillin/Streptomycin, Gibco/BRL), and 2.5 µg/mL Amphotericin B (Fungizone, Gibco/BRL). Cultures were incubated at 37°C in 5% $CO_2$, and passaged at $10^5$-$10^3$ organisms/mL every one to three days respectively. Frozen axenic stocks were prepared by adding 10% FBS and 10% dimethylsulfoxide (DMSO) to log phase growth cultures, that were then mixed, aliquoted, and frozen at −80°C or in liquid $N_2$.

## BACTERIOPHAGE METHODOLOGY

### *Preparation of plating cells*

*Escherichia coli* strain KW251 was used to titre EMBL3 bacteriophage, while strain Y1090 was used to titre λgt11 phage. Bacteria were prepared by streaking cells from frozen stock onto Luria Bertani (LB) agar plates (0.5% NaCl, 0.5% yeast extract, 1% tryptone, pH 7.5, 1.5% agar for plates) and incubating overnight at 37°C. Overnight

cultures were prepared by inoculating fresh single colonies into 5 mL of LB broth + 0.2% maltose and 10 mM $MgSO_4$ and incubated at 37°C overnight with shaking.

***Bacteriophage plating and titration***

For titration of the EMBL3 *T. vaginalis* genomic DNA library and CDF-1 and CDF-2 λgt11 phage isolates (both used as controls), 10 µL of phage stock was serially diluted 10-fold in suspension medium (SM) (100 mM NaCl, 50 mM Tris-HCl pH 7.5, 8 mM $MgSO_4$, 0.01% gelatin). A 100 µL volume of diluted phage was mixed with 200 µL of bacteria and the phage were allowed to adsorb to the bacteria cells for 30 min at 37°C. This phage/bacteria suspension was added to 3 mL of melted LB Top Agarose (LB broth containing 0.7% agarose) pre-equilibrated to 55°C, mixed, and the suspension poured onto 100 mm LB plates pre-warmed to 37°C, which were then incubated 9 hr at 37°C. Library and recombinant phage stock titres were calculated by counting the plaque forming units (pfu) per mL, and multiplying by the dilution factor. To purify individual recombinant phage, plaques were isolated in agar plugs using a Pasteur pipette. Plugs were placed in 100 µL of SM, and the phage were allowed to diffuse from the agar at 4°C overnight. Eluted phage isolates were stored at 4°C.

***Liquid Lysates***

When phage stocks were required, the eluted phage was used to prepare liquid lysates using the liquid culture method. A 500 µL volume of overnight *E. coli* KW251 culture was added to the phage eluate and the mixture was incubated for 20 min at 37°C to allow for phage adsorption to the bacteria. This mixture was used to inoculate 100 mL of pre-warmed LB medium + 10 mM $MgSO_4$, and the culture was incubated at 37°C with

shaking until cell lysis occurred (about 8 hr). The cellular debris was pelleted by centrifugation at 2100 x g (Heraeus OmnifugeRT, C1726 rotor) for 10 min at 4°C, and the supernatant containing phage was collected and stored at 4°C.

## *T. VAGINALIS* EMBL3 GENOMIC DNA LIBRARY SCREENING

### *Plating of the Library*

An EMBL3 phage library containing 15-23 kb fragments from a partial *Mbo* I digest of high molecular weight *T. vaginalis* DNA was screened for potential clones. A 0.5 mL volume of *E. coli* KW251 cells were infected with $2x10^4$ pfu of EMBL phage. Infected cells were plated on 150 mm plates following the procedure outlined above, except that the amount of top agarose was increased to 7 mL. The plates containing the phage libraries were incubated at 4°C for at least 1 hr before plaque lifts were performed.

### *Plaque Lifts*

Plaque lifts were performed by placing dry, positively charged nylon membranes (NEN Colony/Plaque Screen Hybridization Transfer Membranes, 137mm, Mandel Scientific, Guelph, ON) on the agar plates. After 3 min, the membranes were gently removed, soaked twice in 0.5 N NaOH for 2 min to denature the phage DNA, and neutralised twice in 1.0 M Tris-HCl for 2 min, blotting the membranes on filter paper between each step. Membranes were air-dried on filter paper for several hours, and the DNA fixed by baking the membranes at 60°C for 1 hr. Plates were sealed with parafilm and stored at 4°C until the results of the screening were known.

***Screening by DNA Hybridisation***

The library was screened by hybridisation using $^{32}$P-radiolabelled CDF-2U as a probe. The membranes were prehybridised in hybridisation buffer (1% Dextran sulphate, 1% SDS, 1M NaCl) at 65°C for 1-5 hr in a Model 2000 Micro Hybridisation Oven (Robbins Scientific, North York, ON). Denatured probe and salmon sperm DNA were then added to the hybridisation solution to final concentrations of ~$10^6$ counts per minute per mL (cpm/mL) and 50 µg/mL respectively, and the hybridisations were allowed to proceed at 65°C overnight. The membranes were then washed according to the protocols provided by NEN: twice with 2X SSC for 5 min at room temperature, twice with 2X SSC/ 1.0% SDS for 30 min at 60°C, and twice with 0.1X SSC for 30 min at room temperature. The membranes were then wrapped in plastic wrap to keep them moist, and exposed to Cronex 4 X-ray film (Picker International, Brampton, ON) overnight at -80°C.

Autoradiographs were aligned with the plates, which allowed for the identification of the plaques producing positive signals. Plaques were isolated in agar plugs, eluted in SM medium and plated as above. Secondary, tertiary, and quaternary screening of the phage isolates were carried out until the isolates were considered pure.

## ISOLATION OF *T. VAGINALIS* NUCLEIC ACIDS

***Isolation of High Molecular Weight Genomic DNA***

High Molecular Weight Genomic DNA was isolated in early experiments using a guanidium thiocyanate/CsCl method (Wang & Wang, 1985). In later experiments, a shorter, simpler method using Qiagen Genomic-tips 500/G Column (Qiagen,

Mississauga, ON) was employed. In both procedures, $5x10^8$ and $2.5x10^8$ log phase organisms, respectively, were harvested by centrifugation at 5000 x g for 15 min at 4°C (Beckman J2-21M centrifuge, JA10 rotor) and washed twice in cold PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4.7H2O$, 1.4 mM $KH_2PO_4$, pH 7.3).

Using the method of Wang and Wang (1985), the pellet was lysed at 25°C in 10 volumes of 4 M guanidium thiocyanate solution (Chirgwin *et al.*, 1979) (4M guanidium thiocyanate, 0.5% Sarkosyl, 25 mM sodium citrate pH 7.0, 0.1 M 2-β-mercaptoethanol, final pH 7.0). The lysate was vortexed for 10 sec and cesium chloride was added to a final concentration of 0.4 g/mL. The mixture was layered on a 1.5 mL cushion of 5.7 M CsCl/ 0.1 M EDTA pH 7.0 and centrifuged overnight at 34000 rpm at 20°C (RC70 Sorvall Ultracentrifuge, TH 695 rotor). Fractions were collected, diluted two-fold with TE (10 mM Tris-HCl pH 8.0, 1mM EDTA), and extracted repeatedly with equal volumes of chloroform:isoamyl alcohol (49:1) until a clear interface was achieved. The DNA was precipitated with 2 volumes of ethanol and 1/10 volume of 5M sodium acetate pH 5.5, and pelleted by centrifugation at 12,000 x g (Beckman J2-21M, JA20 rotor). Pellets were dissolved in TE (10 mM Tis-HCl pH 7.5, 1mM EDTA pH 8.0) buffer and the samples were treated with RNase A (200 μg/mL) for 30 min at 37°C. Aliquots from each fraction were analysed by electrophoresis on 0.5% agarose gels. Those samples found to contain high molecular weight DNA were pooled, and the DNA was re-precipitated, dissolved in TE buffer and quantitated by absorbance at 260 nm ($A_{260}$) using the Genequant Spectrophotometer (Pharmacia Biotech, Baie D'Urfé, PQ). The $A_{260}/A_{280}$ ratio was also determined to assess the purity of the DNA.

For DNA that was isolated using the Qiagen Genomic-tips 500/G Column

(Qiagen, Mississauga, ON), the manufacturer's protocol for tissue was followed. The pelleted organisms were resuspended in General Lysis Buffer G2 (to which RNase A and Proteinase K had been added to final concentrations of 200 μg/mL and 1mg/mL respectively), and allowed to lyse at 50°C for 2 hr. The cell lysate was then loaded onto the Qiagen Genomic-tips 500 Column, which had been pre-equilibrated with Equilibration Buffer QBT. The column was washed with 2x15 mL of Wash Buffer QC, and the DNA was eluted with 15 mL of Elution buffer QF. The DNA was then isopropanol precipitated at 7500 x g for 15 min at 4°C (Beckman Avanty J-25 Centrifuge, 25.5 rotor), washed in cold 70% ethanol, briefly air-dried, dissolved in 300 μL of water (50°C, 1h), and stored at 4°C. The DNA was analysed by agarose gel (0.5%) electrophoresis and quantitated by absorbance at 260 nm. The $A_{260}/A_{280}$ ratio was also determined to assess the purity of the DNA.

***Isolation of Total T. vaginalis RNA***

Total RNA was isolated from $4x10^8$ log phase organisms using TriPure reagent (Boehringer Mannheim, Laval, PQ) according to the protocol provided by the manufacturer. Briefly, the cells were harvested by centrifugation at 5000 x g for 15 min at 4°C (Beckman J2-21M centrifuge, JA10 rotor), and lysed in 3-6 mL of TriPure reagent. The solution was incubated for 5 min at 25°C to dissociate the nucleoprotein complexes. Chloroform (0.2 volumes) was added to the cell lysates, which was then vigorously shaken for 15 sec, and centrifuged at 12,000 x g for15 min at 4°C (Beckman J2-21M, JA20 rotor). The colourless aqueous phase was collected, from which the RNA was precipitated by isopropanol and pelleted by centrifugation at 12,000 x g for 10 min at

4°C. The RNA pellet was washed with 70% ethanol, dried briefly, dissolved in DEPC-treated $H_2O$, divided into aliquots, and stored at −80°C. The RNA was analysed on a 1% denaturing glyoxal agarose gel, and quantitated by absorbance at 260 nm using. The $A_{260}/A_{280}$ ratio was also measured to analyse the purity of the RNA.

***Isolation of Polyadenylated RNA***

Polyadenylated (polyA$^+$) RNA was purified from 700 µg of total RNA using the Oligotex Spin Column following the manufacturer protocol (Qiagen). Oligotex beads capture polyA$^+$ RNA by hybridisation with covalently linked $dC_{10}T_{30}$ oligonucleotides. Once hybridisation has occurred, the beads are washed free of other RNA species and then the polyA$^+$ RNA is eluted in a low salt buffer, and stored at −80°C. The polyA$^+$ RNA was analysed on a glyoxal agarose gel and quantitated by spectrophotometry at $A_{260}$.

## RNA ANALYSIS

***Denaturing Glyoxal Gel Electrophoresis (McMaster & Carmichael, 1977)***

To 5 µL of RNA sample, 15 µL of sample buffer (13.3 mM $NaH_2PO_4/Na_2HPO_4$ pH 7.0, 0.13% SDS, 66% DMSO, 20% deionized glyoxal) was added. Samples were heated at 60°C for 15 min to denature the RNA, cooled on ice for 2 min, and then loaded onto a 1% agarose gel containing 10 mM $NaH_2PO_4/Na_2HPO_4$ buffer, pH 7.0. Gels were run with re-circulated running buffer (10mM $NaH_2PO_4/Na_2HPO_4$, pH 7.0) in a Hoefer HE 33 Mini electrophoresis unit (Pharmacia) at 40 volts for 2 to 2.5 hr. After electrophoresis, the gels were soaked in running buffer containing either acridine orange

(30 μg/mL) or ethidium bromide (0.5 μg/mL) for 30 min to stain the RNA, and then destained in fresh running buffer for 30 min. RNA was visualised and photographed under short-wave UV light. RNA that was to be transferred to membrane was not stained; rather, duplicate samples were stained.

## DNA ANALYSIS

### *DNA agarose gel electrophoresis*

Samples of DNA were mixed with 6X loading buffer (30% glycerol, 0.25% bromophenol blue, and 0.25% xylene cyanol). Electrophoresis was carried out in 0.5-2% agarose gels buffered with 1X Tris-Borate-EDTA (TBE). The Hoefer HE 33 Miniphor unit (Pharmacia) was used for mini-gels and the LKB 2012 Maxiphor (Pharmacia) unit was used for midi-gels. Mini-gel electrophoresis was carried out at 100 volts for 1-2 hr, while midi gel electrophoresis was carried out at 100 volts for 2-3 hr. For separation of high molecular weight DNA, midi-gel electrophoresis was carried out at 30-40 volts overnight at 4°C or 65 volts for 4-6 hr at 25°C. The gels were soaked in ethidium bromide (0.5 μg/mL) for 10-20 min to stain the DNA, and then destained for 15-30 min in distilled water. DNA was visualised and photographed under short wave ultraviolet light.

### *Isolation and Purification of DNA Fragments from Agarose Gels*

Ethidium bromide stained DNA fragments were located using short wave ultraviolet light. The band of interest was cut from the gel and placed in a tared microfuge tube. DNA larger than 500 bp was purified using the BIO 101 GeneClean II

Kit (BioCan Scientific, Mississauga, ON), while DNA up to 500 bp was purified using the GFX PCR DNA and Gel Band Purification Kit (Pharmacia) following the instructions provided by the kit manufacturers.

## RESTRICTION AND MODIFICATION OF DNA

### *Enzymes*

Unless otherwise indicated, all restriction endonucleases were supplied by the following manufacturers: Amersham Pharmacia Biotech (Baie d'Urfé, PQ), New England Biolabs (Mississauga, ON), Promega Corporation (Fisher Scientific, Nepean, ON), MBI Fermentas (Flamborough, ON), Gibco-BRL, and all modifying enzymes were supplied by Promega, Gibco-BRL, Pharmacia, Perkin-Elmer/Applied Biosystems (Mississauga, ON), Pharmacia, New England Biolabs. The PCR enzyme rTth XL (PE/Applied Biosystems) was used for all PCR applications except RACE, which required Advantage II (Clontech, Palo Alto, CA) as per kit instructions.

### *Restriction digests*

Typically, 2-5 units of enzyme were used per µg of DNA in 20-100 µL of total reaction volume. All plasmid DNA digests were performed for 1-5 hr, while high molecular weight genomic DNA was digested overnight, in the appropriate buffers at the temperatures recommended by the suppliers. When required, restriction digests were terminated by heat inactivation or by phenol/chloroform extraction and ethanol precipitation. Phenol/chloroform extraction was done by first adding an equal volume of UltraPure Phenol:Chloroform:Isoamyl alcohol (25:24:1) (Gibco/BRL), vortexing gently for 10 sec, and centrifuging at 13,000 rpm (Heraeus Biofuge A Microcentrifuge, Baxter-

Canlab, Montreal, PQ) for 3 minutes. The aqueous supernatant was collected and extracted with an equal volume of chloroform. The DNA was then precipitated by adding 2 volumes ice-cold ethanol + 1/10 volume 5 M NaAc, and pelleted by microcentrifuation at 13,000 rpm for 20 minutes. The DNA pellet was washed in ice-cold 70% ethanol, and centrifuged again at 13,000 rpm for 15 minutes. Finally, the pellet was dried briefly at 50°C and dissolved in water or TE buffer, depending upon the application.

## PREPARATION OF RADIOLABELLED PROBES

### *Probe Design and Preparation*

For the genomic library screening, Southern blots, and Northern blots, $^{32}$P-radiolabelled CDF-2U was used as a probe (Figure 1). CDF-2U represents a unique sequence of CDF-2, between positions 98-743, which is free of repetitive elements. CDF-2U was excised from the PGEM 7Zf(+) recombinant plasmid containing CDF-2 (pCDF-2) by digesting the plasmid with *Nla* III, separating the digested fragments on a 1.5% agarose gel, and purifying the 645 bp fragment from the gel using the Geneclean II Kit. CDF-2U was further cleaved in half by double digestion of pCDF-2 with *Nla* III and *Taq* I. The two resulting fragments, 370 bp and 270 bp, were separated on a 1.7% agarose gel, and purified with the GFX PCR DNA and Gel Band Extraction Kit (Pharmacia). For use as probes, the DNA fragments CDF-2U, 373 and 272 were then radiolabelled using the Prime-a-Gene system (Promega) with [5'α $^{32}$P]dCTP.

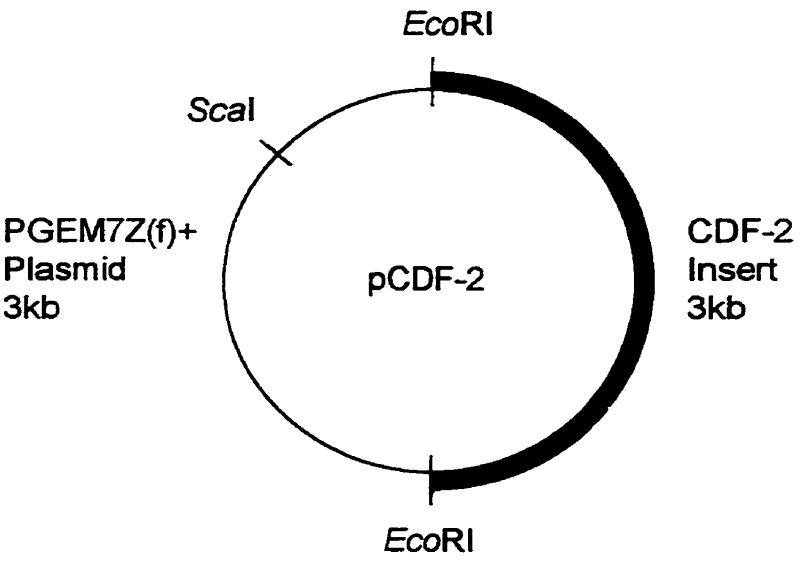
### *Radiolabelling of DNA fragments*

DNA was radiolabelled with the Prime-a-Gene System (Promega) which is based

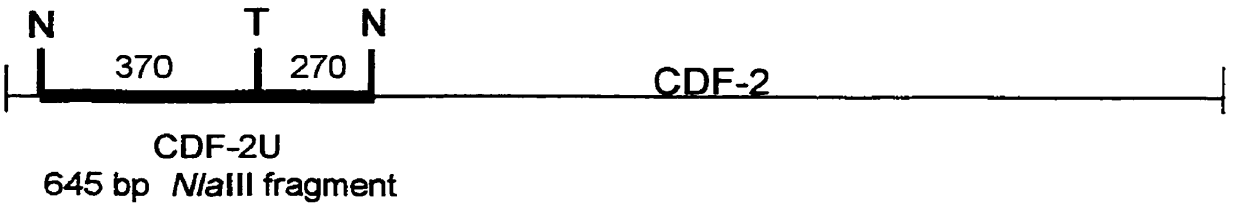
**Figure 1: Generation of probes**

CDF-2 was generated from the digest of pCDF-2 with *Eco*RI and *Sca*I to release the insert from the pGEM7Z(f)+ plasmid (Panel A). CDF-2U was prepared from an *Nla*III (N) digest of pCDF-2 to release a 645 bp fragment of CDF-2 near its 5' end which was free of repetitive sequence (Panel B). Probes 270 and 370 were prepared by digesting CDF-2U with *Taq*I (T) to divide the fragment into two (Panel B). All probes were randomly primed $^{32}P$ labelled.

**A**


EcoRI
ScaI
PGEM7Z(f)+
Plasmid
3kb
pCDF-2
CDF-2
Insert
3kb
EcoRI

**B**


N
T
N
370
270
CDF-2
CDF-2U
645 bp NlaIII fragment

upon the method of Feinberg & Vogelstein (1983) and utilises hexanucleotides for randomly priming DNA synthesis. Briefly, 25-50 ng of DNA was denatured for 2 min at 95-100°C, and snap-cooled on ice. The DNA was then added to a reaction mixture containing final concentrations of 1X labelling buffer (50 mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 2 mM DTT, 0.2 M HEPES pH 6.6, and 260 μg/mL random hexanucleotides), 400 μg/mL BSA, 20 μM each of dATP, dGTP, dTTP, and 50 μCi [$\alpha$-$^{32}$P]dCTP (3000 Ci/mmol, Amersham Pharmacia Biotech, Baie d'Urfé, PQ) in a 50 μL volume. Five units of Klenow DNA Polymerase was added, and the reaction was incubated at 37°C for 1-16 hr. After the reaction was terminated by heating at 95-100°C for 2 min, EDTA was added to a final concentration of 20 mM, and the reaction mix was passed through a Sephadex G-50 column (Pharmacia) to remove unincorporated nucleotides. Fractions containing the probe were pooled, counted in an LKB 1214 Rackbeta Liquid Scintillation Counter (Pharmacia), and then stored at -20°C.

***Radiolabelling of oligonucleotides, primers and markers***

Oligonucleotides, primers and markers were end labelled with [$\gamma^{32}$P]ATP using T4 polynucleotide kinase (Promega). Reaction mixes were assembled as follows: 10 pmol of oligonucleotide or 250 ng of dephosphorylated phiX174 *Hinf* I DNA marker, 1μL 10X PNK buffer, 3 μL [$\gamma$-$^{32}$P] ATP (3000 Ci/mmol, 10 mCi/mL), 1 μL PNK (10 units/μL) and water to a final volume of 10 μL. The reaction was incubated at 37°C for 10-20 min and terminated by heating to 90°C for 2 min. Percent incorporation was determined by dot-blot. Aliquots of 3 μL of a four-fold dilution of the nucleic acid mix were blotted twice onto each of two quarters of an 82 mm positively charged nylon

membrane (Colony/Plaque Screen Membranes, NEN). After the blots were dried, one of the two blots was washed twice in 0.5 M $Na_2HPO_4$ (pH 6.8) to remove the unincorporated [$\gamma$-$^{32}$P] ATP. The radioactivity of each membrane was measured (in cmp) using the LKB 1214 Rackbeta Liquid Scintillation Counter and the percent incorporation determined by calculating the ratio of the amount of activity of washed/unwashed membranes.

## SOUTHERN BLOTTING AND HYBRIDISATION ANALYSIS:

High molecular weight genomic DNA from *T. vaginalis* strain 202 was digested with various restriction enzymes and the fragments were separated by 0.5-1.0 % agarose gel electrophoresis. The DNA was transferred to 1.2 µm Biodyne A nylon membrane (Pall Gelman Corporation, Montreal, PQ) by overnight capillary transfer in 20X SSC according to the manufacturer's instructions for Conventional Southern Transfer or the Pall Improved High Salt Denaturation Method. Prior to conventional transfer, the DNA was depurinated, denatured, and neutralised by soaking the gels for 20-40 min in each of 0.25 M HCl, 0.5 M NaOH/1.5 M NaCl, and finally 0.5 M Tris/HCl pH 7.4, 1.5 M NaOH respectively. If the Pall Improved High Salt Denaturation method was followed, the neutralisation step was omitted. The DNA was fixed onto the membranes by baking at 80°C for 40 min, and/or by irradiation under short-wave UV light for 3 min.

Pre-hybridisation and hybridisation steps were performed using the Robbins Model 2000 Micro Hybridisation Oven and the protocol provided by Pall. Initially a Denhardt's based hybridisation buffer was used (5X Denhardt's Solution, 5X SSC, 0.1% SDS, and 100 µg/mL denatured sonicated salmon sperm DNA (Gibco/BRL)). Later, in an attempt to improve background, a PEG based buffer was used (10% PEG (MW 8000),

1.5X SSPE, 7% SDS, and 100 μg/mL denatured sonicated salmon sperm DNA). Prehybridisation steps were performed at 65°C for 1-3 hr, and, following the additions of ~ $5x10^5$ cpm/mL labelled probe to the buffer, hybridisations were performed at 65°C overnight. Prior to addition to the hybridisation solution, the probe solutions were first denatured by heating for 10 min at 95°C and chilling on ice for 10 minutes. Following hybridisation, the membranes were washed as follows: three 5 min washes in 2X SSC at room temperature, two 20 min washes in 0.2X SSC at room temperature, and one to three 20 min washes in 0.2X SSC at temperatures increasing to 65°C as required. Southern blots were exposed to Cronex 4 X-Ray film overnight or up to several days at −80°C.

## NORTHERN BLOTTING AND HYBRIDISATION ANALYSIS

After RNA was separated on a denaturing glyoxal agarose gel, the unstained RNA was transferred to a Pall 0.45 μm Biodyne B nylon membrane by capillary overnight transfer in RNase-free 20X SSC following the manufacturer's instructions. The RNA was fixed to the membrane by baking at 80°C for 30-45 min. The Northern blot prehybridisation and hybridisation steps were performed in a formamide buffer containing 50% formamide, 5X SSPE, 5X Denhardt's, 0.1% SDS, and 100μg/mL denatured sonicated salmon sperm DNA using the Robbins Micro Hybridisation oven. Following prehybridisation for 1-3 hr at 42°C, $5x10^5$—$1x10^6$ cpm/mL of denatured probe was added, and hybridisation was then carried out overnight at 42°C. The membranes were washed as follows: three 5 min washes in 2X SSC at room temperature, three 15 min washes in 0.2X SSC/ 0.1% SDS at room temperature, followed by one 15 min wash

in 0.2X SSC at room temperature. Northern Blots were exposed to Cronex 4 X-ray film overnight or up to several days at −80°C.

## PLASMID METHODOLOGY

### *Cloning and Transformations*

Unless otherwise stated, all cloning was performed exclusively from PCR products using the TOPO-TA Cloning Kit (Invitrogen, Carlsbad, CA) with pCR2.1-TOPO TA vector and TOP10 One Shot Chemically Competent *E. coli*. The vector is supplied linearised with single 3' thymidine (T) overhangs for TA cloning, and covalently bound topoisomerase for spontaneous ligation. Ligation reactions were prepared by adding 0.5 to 2 μL of fresh or fresh frozen PCR product, 2 μL of TOPO vector, and dd$H_2O$ up to 5μL, and incubating the reaction for 5 min at room temperature. The cells for the transformations were supplied in 50 μL aliquots at transformation efficiencies of $1x10^9$ cfu/μg PUC DNA. The cells were thawed on ice, and the transformation reactions were prepared by adding 2 μL of 0.5 M β-mercaptoethanol, and 2 μL of the TOPO ligation reaction to the cells. The transformation mix was incubated on ice for 30 min, heated at 42°C for 30 sec to shock the cells, and then cooled in ice for 2 min. A 250 μL volume of SOC (Invitrogen) medium was then added to the cells and the mixture was incubated at 37°C for 30 min with shaking. Thirty-five to 100 μL of transformation mix was then spread onto LB agar plates containing 50 μg/mL ampicillin (Amp), 0.5 mM isopropyl-ß-D-thiogalactoside (IPTG), and 40 μg/mL 5-bromo-4-chloroindolyl-ß-galactoside (X-GAL), which were incubated overnight at 37°C. Ten white colonies were

picked from each transformation, and cultured overnight in 3 mL of LB/Amp broth at 37°C with shaking. Larger cultures were prepared by inoculating 30-100 mL of LB/Amp broth with 0.5 to 1 mL of overnight culture, which was then incubated overnight at 37°C with shaking.

Frozen stocks of transformed *E. coli* DH5αF' containing the pCDF-2 plasmid were used for pCDF-2 plasmid DNA extractions. The host *E. coli* cells were streaked from frozen stock onto LB/Amp agar plates. Fresh single colonies were inoculated into 5 mL of LB/Amp broth and incubated at 37°C overnight with shaking. Larger cultures were prepared by inoculating 500 mL of LB/Amp broth with 1 mL of overnight culture, which was then incubated overnight at 37°C with shaking.

***Small Scale Isolations of Plasmid DNA from E. coli (Mini-preps)***

Plasmid DNA mini-prep isolations were prepared using a modified alkaline lysis method originally described by Birnboim & Doly (1979). Bacteria cells were grown overnight at 37°C with shaking in 3 mL of LB/Amp broth. The cells from 1.5 mL of culture were pelleted by centrifugation for 30 sec at 13,000 rpm (Heraeus Biofuge A Microcentrifuge, Baxter-Canlab, Montreal, PQ) and resuspended in 150 μL of ice cold Buffer I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, 20 μg/mL DNase-free RNase A). Bacteria were lysed by the addition of 150 μL of fresh Buffer II (1% SDS, 0.2N NaOH), and the lysate mixed gently by inversion. Samples were incubated for 5 min at room temperature, and then neutralised with 150 μL of 5M KAc, pH 4.8. The suspension was mixed thoroughly, placed on ice for 5 min, and centrifuged at 13,000 rpm for 10 min. The aqueous supernatant was extracted with an equal volume of chloroform,

and the plasmid DNA was precipitated from solution by the addition of 1 mL of ethanol, and pelleted by centrifugation at 13,000 rpm for 15 min. After the supernatant was drawn off, the pellet was air-dried briefly, dissolved in 50 μL of water and stored at −20°C. The DNA was analysed by agarose gel electrophoresis.

***Ultrapure Plasmid DNA Mini-prep Extractions***

Plasmid DNA was ultra-purified for sequencing or PCR applications from 1.5-3.0 mL of *E. coli* culture using the Qiagen-tip 20 Gravity Flow column following the alkaline lysis procedure provided in the Plasmid Mini Kit. Alternatively, after performing the alkaline lysis plasmid DNA extraction described above, 1 mL of Equilibration Buffer QBT was added to the dissolved DNA. The solution was then loaded onto a pre-equilibrated Qiagen-tip 20 column, and the procedure continued according to the manufacturer's instructions. Pelleted DNA was dried briefly and dissolved in 20-50 μL water or 10 mM Tris-HCl pH 8.0, analysed by agarose gel electrophoresis, and quantitated by absorbance at 260 nm.

***Large Scale Plasmid DNA Extractions***

Large scale plasmid DNA isolations were performed using either the Qiagen-tips Plasmid Isolation Kit or the Macherey-Nagel Nucleobond AX Plasmid Isolation Kit (Vector Biosystems, Toronto, ON). Midi-prep plasmid extractions of pCDF-ORF were prepared from 30 mL of *E. coli* culture. Cells were harvested by centrifugation at 2500 x g for 15 min at 4°C (Jouan CR3i centrifuge, T40 rotor), and the plasmid DNA was extracted using the Qiagen-tip 100 following the instructions provided in the kit. Maxi-prep plasmid DNA extractions of pCDF-2 were prepared from 500 mL of culture. Cells

were harvested by centrifugation at 6000 x g for 10 min at 4°C (Beckman J2-21M centrifuge, JA10 rotor), and the pCDF-2 plasmid DNA was extracted using the Nucleobond AX-500 Cartridge, using the protocol provided by the manufacturer. Both kits use modifications of the alkaline lysis method of isolating the plasmid DNA, followed by binding of the plasmid DNA onto an anion-exchange resin. The DNA is washed of impurities, and is eluted with a high salt, high pH (8.5) buffer. The DNA was analysed by agarose gel electrophoresis, and quantitated by absorbance at 260 nm.

## OLIGONUCLEOTIDES

All oligonucleotide primers (oligos) were designed according to Loffert *et al.*, (1997) using the computer program Primer Designer. Melting temperatures were calculated by nearest neighbour analysis using the Oligo Calculator at the Genosys website (www.genosys.com). All oligonucleotides designed for PCR were purchased through Bio/Can Scientific (Mississauga, ON), whose supplier changed from Genosys Biotechnologies (The Woodlands, Tx) to Operon Technologies (Alameda, Ca) in June 1999. All oligos designed for PCR were synthesised by Genosys except the oligos designed for RACE. All oligos designed for sequencing were synthesised and purified by the University of Ottawa Biotechnology Research Institute. Sequences for all oligonucleotides are listed in Appendix 1.

## INVERSE PCR

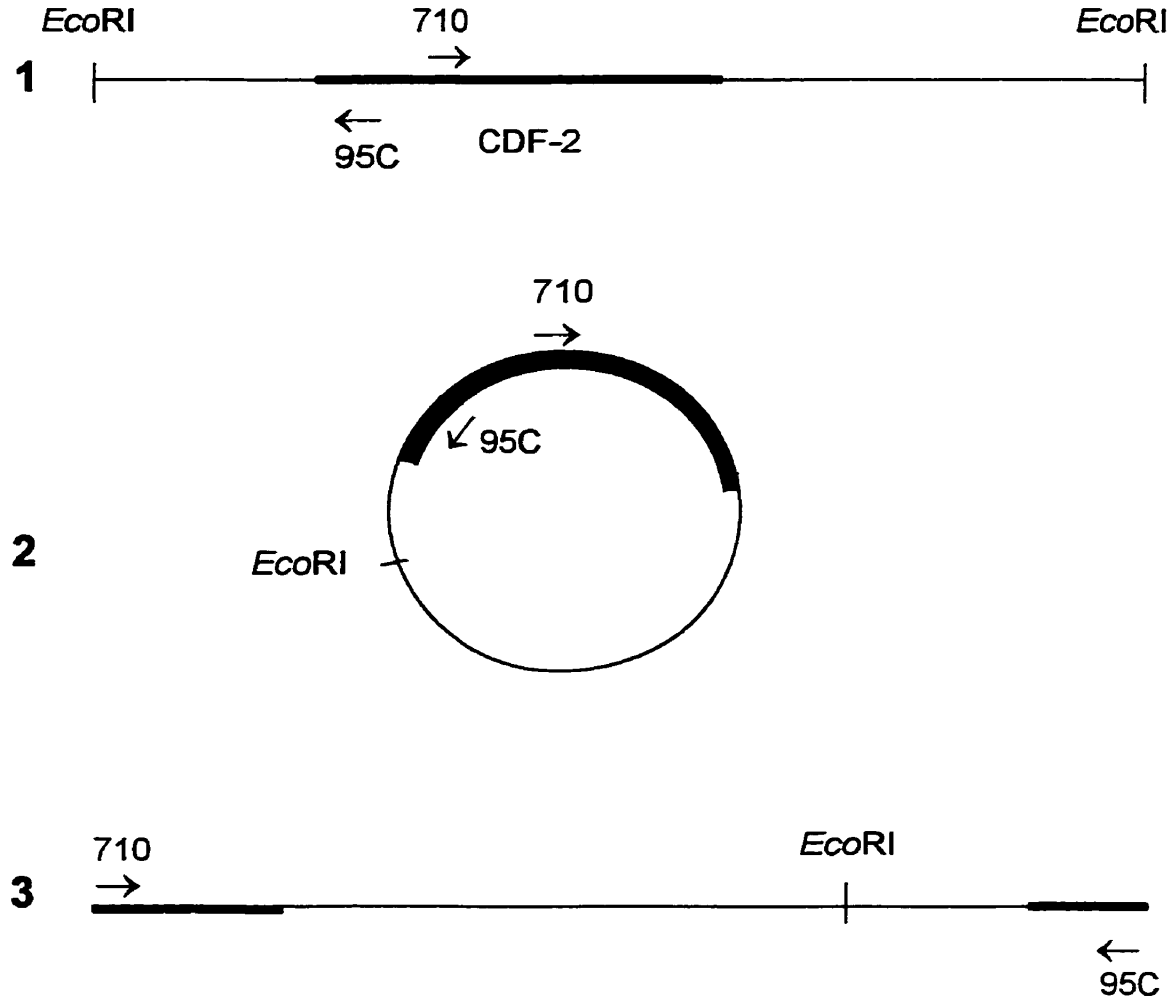
For amplification of the regions flanking CDF-2, *T. vaginalis* 202 genomic DNA was digested with *Eco*RI or *Taq*I, self-ligated into circles, and then amplified with tail-to-tail primers. The experiment was designed and carried out according to the methods of

Hartl & Ochman (1994). A diagram of the planned experiment is shown in Figure 2. Primers were designed from the non-repetitive region of CDF-2 according to general guidelines (Loffert *et al.*, 1997). Primers 95C/710 were designed to amplify the 8.7 kb *Eco*RI fragment, while primers 95C/184 and 631C/710 were designed to amplify the 1.6 kb and 4 kb *Taq*I fragments. Before experiments using genomic DNA began, the primers were first tested in PCR reactions using pCDF-2, linearised pCDF-2, and linearised/re-ligated pCDF-2 as templates to ensure that the procedure was working, that appropriate products were being amplified, and to determine appropriate annealing temperatures. *T. vaginalis* high molecular weight genomic DNA from isolate 202 was digested to completion with *Eco*RI or *Taq*I and cleaned using the GeneClean II Kit. The DNA (0.2, 0.5, 1.0, and 5.0 ng/μL) was then self-ligated into circles using 7 u/μL T4 DNA ligase at 16°C overnight, as recommended in inverse PCR protocols. The ligated DNA (0.5 to 1.0 ng) was then used directly in 50 μL PCR reactions, which were assembled as follows: water, 12.1 μL rTthXL buffer (3.3X), 2.2-2.7 μL $Mg(Oac)_2$ (25 mM), 4.0 μL NTP (10 mM), 1.0 μL of each primer (30 pmol), and DNA (1-5 μL). A separate enzyme mix was prepared consisting of 5.75 μL H2O, 3.0 μL rTth XL buffer, and 1.25 μL rTth (2.5 u)(Perkin/Elmer). The DNA/buffer mix was heated to 94°C for 10 min. before the enzyme mix was added. The reactions were cycled at 94°C for 30-45 sec., 44-58°C for 1min., and 72°C for 5 min. for 30 cycles, and then extended at 72°C for 10 min. Controls consisted of optimised amplifications with pCDF-2, master mixes containing DNA only without primers and master mixes containing primers only without DNA. Products were analysed on a 1.0 % agarose gel.

**Figure 2: Inverse PCR Amplification of *Eco*RI and *Taq*I Restriction Fragments**

*T. vaginalis* 202 high molecular weight genomic DNA was digested with either *Eco*RI or *Taq*I; the 8.7 kb *Eco*RI restriction fragment containing CDF-2 was the target template (A1). Opposing primers (95C and 710) were designed from within the non-repetitive sequence of CDF-2 (A1). The digested DNA was self-ligated into circular form (A2) and inverse PCR was performed. Expected PCR products will be linear with opposing primer-binding sites now located at the end of the amplicon facing inward. The restriction site will be found in the middle of the amplicon (A3). The same experimental design was used as an approach to amplify the 1.6 kb and 4 kb *Taq*I restriction fragments. For this experiment, primers 95C and 184 were designed to amplify the 1.6 kb fragment, while primers 631C and 710 were designed to amplify the 4 kb fragment (B).

A


EcoRI
710
EcoRI
1
95C
CDF-2
710
95C
2
EcoRI
710
EcoRI
3
95C

B


TaqI
TaqI
TaqI
184
710
95C
631C

## SEQUENCING

All plasmids that were to be sequenced were first purified using the Qiagen-tip 20 Gravity Flow column. All automated sequencing was performed by the University of Ottawa Biotechnology Research Institute using the ABI BigDye Terminator Cycle DNA Sequencing Kit with AmpliTaq DNA polymerase (Applied Biosystems) in a 373 Stretch thermocycler (Perkin Elmer/ Applied Biosystems). Primers used for sequencing include T7, SP6, M13 Reverse, and gene-specific primers: 675, 1246, 2169, 2192C, 3039, 3876, 4501, 5142, 5841, 7201C.

## SEQUENCE ANALYSIS

Sequence analysis was carried out using the IBI Pustell Sequence Analysis program in order to generate restriction maps, identify open reading frames, potential amino acid sequence, and motifs of interest within the sequence. Database searches were carried out using the National Institutes of Health (NIH) National Center for Biotechnology Information (NCBI) Basic Local Alignment Tool (BLAST) programs (Altschul *et al.*, 1990; Altschul *et al.*, 1997). Searches were carried out via the Internet at www.ncbi.nlm.nih.gov. Both BLASTn (nucleotide) and BLASTx (nucleotide sequence that has been translated) searches of the nucleotide and polypeptide non-redundant databases (Genbank, EMBL, DDBJ, and PDB) were performed.

Protein sequence analysis was carried out using a variety of programs accessed through the CMS Molecular Biology Resource website located at the University of California at San Diego (http://www.sdsc.edu/ResTools/biotools/biotools1.html). These programs included TMPred (Hofman & Stoffel, 1993) and SOUSI (Hirokawa *et al.*,

1998) for transmembrane spanning motifs and hydrophobicity plots, and PSORT (Nakai & Kanehisa, 1992), SignalP (Nielson *et al.*, 1997), and PROSITE (Bairoch *et al.*, 1997) (via PredictProtein http://www.EMBL-Heidelberg.DE/predictprotein/) for sequence motifs and sorting signals.

## RAPID AMPLIFICATION OF GENOMIC DNA ENDS (RAGE)

An adaptor-ligated PCR based GenomeWalker Kit (Clontech) was used to identify and sequence the unknown regions of DNA flanking CDF-2 (Siebert *et al.*, 1995). This technique enabled walking along uncloned genomic DNA fragments until the desired sequences were found. Before PCR could be performed, however, uncloned *T. vaginalis* libraries were prepared.

### *Preparation of T. vaginalis Uncloned Genomic DNA Libraries*

The libraries were prepared by digesting the genomic DNA with various restriction enzymes, blunt-end repairing where necessary, and ligating adaptors onto the ends of the resulting restriction fragments. Five original libraries were a generous gift from Mr. Dino Petrin, who prepared them according the protocols provided by Clontech using *T. vaginalis* DG high molecular weight genomic DNA. A sixth library was later prepared from *T. vaginalis* 202 genomic DNA by the same method, which is briefly outlined below.

Briefly, 2.5 µg aliquots of high molecular weight genomic DNA were digested at 37°C overnight with 80 units of the various restriction enzymes (Clontech), and appropriate restriction enzyme buffer in a 100 µL reaction. The original five enzymes provided with the kit were blunt-end cutters: *Dra* I, *Eco*R V, *Pvu* II, *Sca* I, and *Stu* I,

while the sixth library was prepared using *Eco*R I (Pharmacia). *Eco*R I recessed ends were blunt-end repaired directly in the restriction digest reaction mix following the method of Cobianchi & Wilson (1987) by adding 25 μM dNTP (final conc.) and 5 units of Klenow fragment (Pharmacia), and incubating at room temperature for 30 min. The digested DNA was extracted as described in the section on "Restriction Digests" with equal volumes of phenol/choloroform/isoamyl alcohol (Gibco/BRL), followed by chloroform, ethanol precipitated with 20 μg of glycogen, and dissolved in 20 μL of TE (10 mM Tris-HCl/ 0.1 mM EDTA) pH 7.5. The GenomeWalker adaptor was then attached to the ends of the digested DNA (0.4 μg) using T4 DNA ligase according to the kit instructions. The adaptors are partially double-stranded; one end is blunt-ended and phosporylated at the 5' end so that it is ligation-ready for both ends of the restriction fragment. Once the ligation was complete, the DNA was diluted 10-fold with TE (10/1, pH 7.4). The result of the ligation step was an uncloned library of adaptor-ligated restriction fragments.

***Amplification of Genomic DNA Ends***

These genomic DNA libraries were then used as templates for this specialised form of adaptor-ligated PCR which uses a combination of "vectorette" (adaptor-ligated) PCR with suppression PCR (Siebert *et al.*, 1995). The amplification of the genomic DNA ends (RAGE) begins from gene-specific primers in the known portion of the restriction fragments, and extends into the unknown region of adjacent DNA, to the adaptor ligated at the end of the restriction fragment (Figure 3). The exposed end of the adaptor is constructed so that it has a 36 bp recessed 3' end, which is terminated by an amine group. This blocks polymerase catalysed extension of the lower recessed end, and

prevents the generation of the primer binding site unless the upper complementary DNA strand is first extended by a distal gene-specific primer (Figure 4). Thus, in the first cycle, amplification begins at the gene-specific primer and proceeds uni-directionally, so that the formation of two adaptor-primer-binding sites on the general population of fragments does not occur. If the 3' end of an adaptor does extend to create a molecule that has the full-length adaptor sequence on both ends, non-specific adaptor to adaptor amplification will be suppressed by touchdown cycles. During this technique, annealing/extension temperatures ($T_{anneal}$=72°C) are performed several degrees higher than the melting point of the primers ($T_{melt}$=68°C) through the initial cycles. At this temperature, any molecule which has the full primer sequence on both ends will self-anneal end to end in a pan-handle shape (Figure 4). This intra-molecular annealing of adaptor ends suppresses non-specific amplification, allowing critical levels of gene-specific product to accumulate. Touchdown PCR also increases the binding specificity of the gene-specific primer.

The gene specific primers (GSP-1) were designed so that they were 26-30 bp in length, and had melting temperatures (nearest neighbour analysis) of 68-70°C. They were designed from regions of non-repetitive sequence, within 100-300 bp away from the end of the known sequence. Nested primers (nGSP-2) were prepared, where possible, so that there was approximately 20 nucleotides between the two primers.

The PCR reactions were prepared as follows: 1 μL of adaptor-ligated DNA (~5 nmol) was added to 0.5 mL PCR tubes (Perkin-Elmer) containing 26.8 μL of water, 4.0 μL of Clontech Tth PCR buffer (10X), 1 μL of dNTP (10mM), 2.2 μL of $Mg(OAc)_2$ (25 mM), 1 μL of AP1 (10 μM), and 1μL of GSP1 (10 μm). A separate enzyme mix was

**Figure 3: The Suppression PCR Effect**

The Genome Walker Adaptor, which is ligated to both ends of the restriction fragment, has an amine block on the 3' end of the lower strand. This prevents extension of the 3' end of the adaptor, so that an AP1 binding site cannot be created until the upper stand has first been extended from a gene specific primer. In rare cases when the 3' end of the Genome Walker Adaptor does get extended, and can serve as a template for non-specific adaptor to adaptor amplification, the high annealing temperature favours self annealing, preventing any amplification. Reprinted with permission from Clontech.

**GenomeWalker Adaptor**

*Mlu* I *Sal* I *Srf* I *Sma* I/*Xma* I

5'–GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT–3'

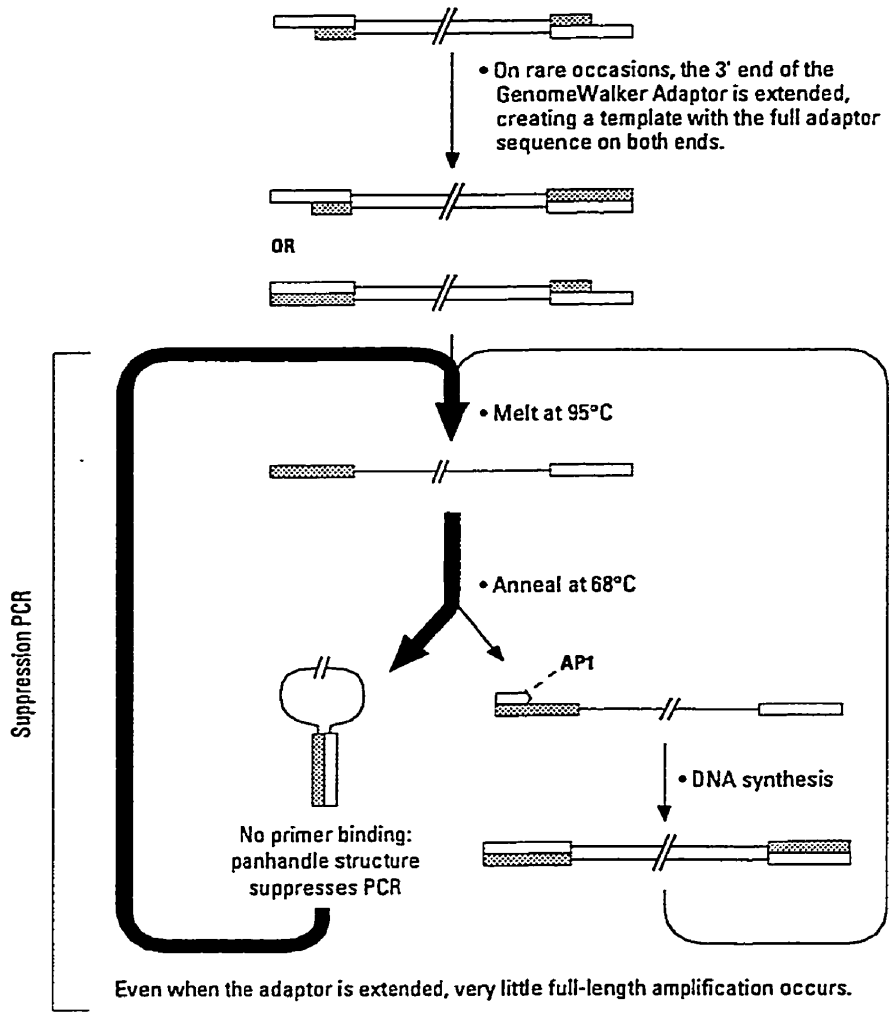
3'–$H_2N$-CCCGACCA-$PO_4$–5'

**Adaptor primer 1** (AP1; 22-mer)

5'–GTAATACGACTCACTATAGGGC–3'

**Nested adaptor primer 2** (AP2; 19-mer)

5'–ACTATAGGGCACGCGTGGT–3'

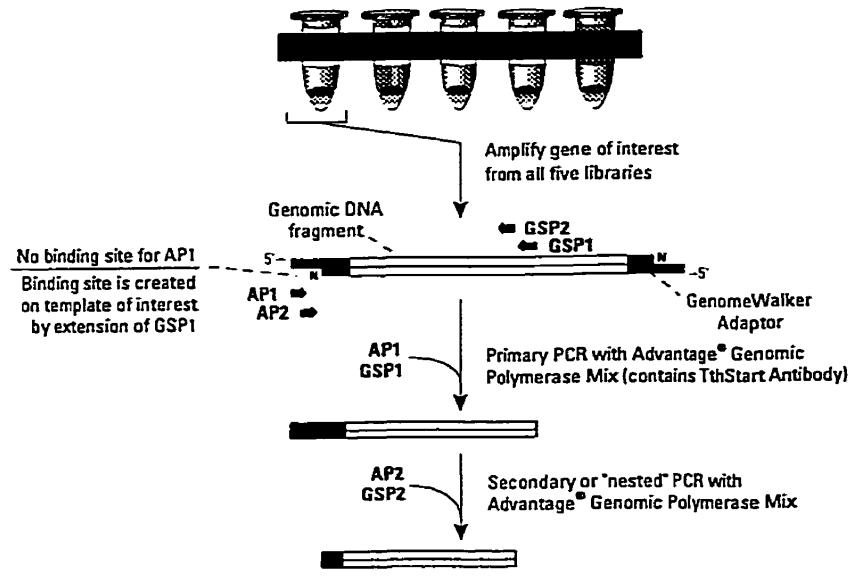



Even when the adaptor is extended, very little full-length amplification occurs.

**Figure 4: Genome Walking**

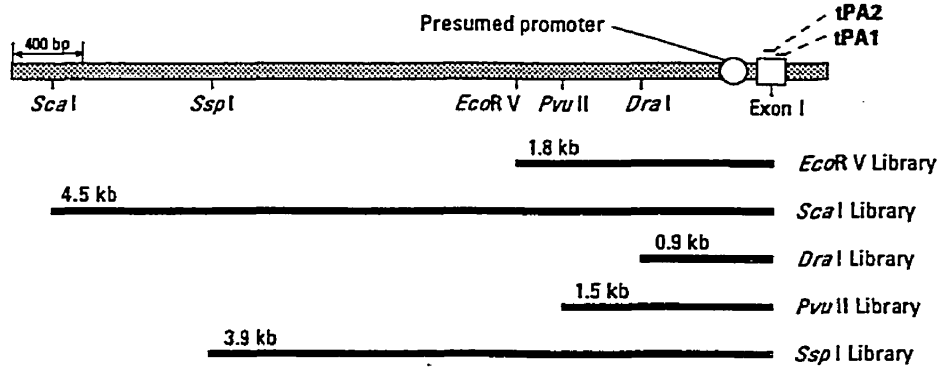
(A) Restriction fragments are amplified from the gene specific primer to the adaptor primer at the end of the restriction fragment. (B) Amplicons from different libraries overlap and can be mapped into contiguous sequences (contigs). The next segment of unknown sequence can be determined in walking steps. Reprinted with permission from Clontech.

**A**




**B**

**Map of tPA locus and expected PCR products**



prepared, containing (per reaction) 8 μL of water, 1μL of 10X Tth PCR buffer (Clontech), and 1 μL of rTthXL, 2.5 units/μL (Perkin-Elmer). A manual hot start was performed, where the DNA mix was heated to 94°C, at which point 10 μL of enzyme mix was added. The cycling parameters in a Perkin Elmer Thermal Cycler 480 were as follows: 94°C for 25 sec, 72°C for 3 min (7 cycles); 94°C for 25 sec, 67°C for 3 min (32 cycles); 67°C for 7 min (1 cycle); 94°C for 25 sec, 67°C for 3:00 (5 cycles); and 67°C for 7 min (1 cycle).

Once the primary PCR amplification was complete, the reaction mix was diluted 50 fold, and a second, nested round of PCR was performed using nested gene-specific (nGsp-2) and adaptor primers (AP-2). The reactions were assembled as above, and cycled as follows: hot start, 94°C for 25 sec, 72°C for 3 min (5 cycles); 94°C for 25 sec, 67°C for 3 min (20 cycles); and a final extension of 67°C for 7 min.

The PCR products were then analysed on a 1.5 agarose gel electrophoresis. Those PCR mixes found to contain long products in good yield, with few background products were cloned using the TOPO-TA Cloning system (Invitrogen).

The positive control entailed: library construction and PCR amplification with the DNA supplied with the system to ensure that all steps in the kit were working, and that the cycling parameters worked with the enzyme and thermocycler. The negative control entailed PCR amplification of the reaction mixes with each primer pair but with template DNA omitted to ensure all master mixes were not contaminated. Eventually, second and third controls were added which consisted of testing the reaction mixes containing DNA with no primers and with each primer individually to ensure that the products were not resulting from the DNA alone or from only one primer.

The process was initiated from CDF-2, walking 5' upstream and 3' downstream of the known sequence using Gsp1-163C/nGsp2-89C and Gsp1-3876/nGsp2-3889 respectively. As new sequence was characterised, the process was repeated (genome walking), until adequate sequence was obtained, using the following primers: Walk 3 downstream from new sequence obtained from walk 2: Gsp1-201C/nGsp2-73C. Walk 4 downstream from new sequence obtained from Walk 3: Gsp1-5142/nGsp2-5160. Walk 5 downstream from new sequence obtained from walk 4: Gsp1-7224 /nGsp2-7284.

## RAPID AMPLIFICATION OF cDNA ENDS (RACE)

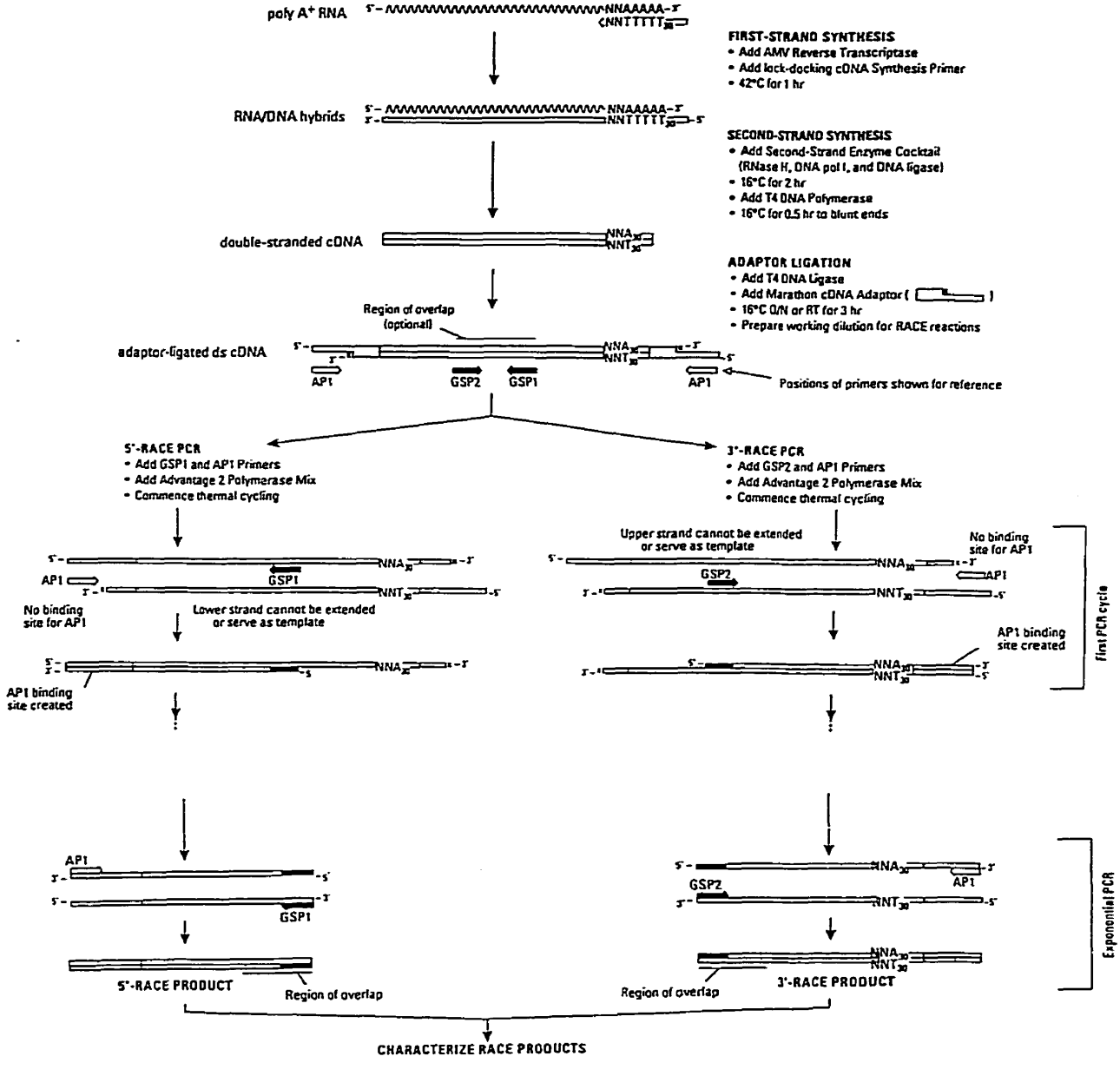
Double stranded cDNA synthesis and rapid amplification of cDNA ends was performed using the Marathon cDNA Amplification Kit (Clontech). This procedure, designed by Chenchik *et al.* (1995) involves preparation of the double stranded cDNA from total or polyA$^+$ RNA (first and second strand synthesis), ligation of the adaptors onto the ends of the cDNA molecules, and finally, the PCR amplification of the cDNA ends (Figure 5).

### *Double-stranded cDNA synthesis*

Polyadenylated RNA (1μg) and total RNA (7.5 μg) were used as templates for double stranded cDNA. The first strand of cDNA was synthesised by AMV reverse transcriptase. Synthesis was initiated by a lock-docking oligo (dT) primer containing two degenerate nucleotides at the 3' end, which place the primer at the start of the polyA$^+$ tail, eliminating 3' heterogeneity. Second strand synthesis was carried out according to the method of Gubler & Hoffman (1983) in the presence of *E. coli* polymerase I, *E. coli* DNA ligase, and *E. coli* RNase H. Using this method, the RNase H nicks the RNA in the

**Figure 5: Flow chart of the Marathon RACE Procedure**

First strand synthesis is primed by a lock-docking degenerate oligo(d)T primer in the presence of AMV Reverse Transcriptase. The second strand is synthesised by adding a 3-enzyme cocktail: RNase H which nicks and degrades the RNA strand, DNA polymerase I which synthesises nascent DNA using nicked RNA as primers, and DNA ligase which ligates the DNA fragments into a single strand. Less than 15% of all second strand synthesis is primed by hairpin loop formation. Finally T4 DNA polymerase is added to blunt the ends of the double stranded cDNA and adaptors are ligated on to the ends. At this point, an uncloned library of adaptor-ligated cDNA molecules has been prepared. cDNA molecules are amplified using gene specific and adaptor primers. Experiments carried out in this project were amplified a second time with nested gene specific and adaptor primers as in the genome walking procedure. Reproduced with permission from Clontech.


poly A+ RNA
FIRST-STRAND SYNTHESIS
• Add AMV Reverse Transcriptase
• Add lock-docking cDNA Synthesis Primer
• 42°C for 1 hr
RNA/DNA hybrids
SECOND-STRAND SYNTHESIS
• Add Second-Strand Enzyme Cocktail (RNase H, DNA pol I, and DNA ligase)
• 16°C for 2 hr
• Add T4 DNA Polymerase
• 16°C for 0.5 hr to blunt ends
double-stranded cDNA
ADAPTOR LIGATION
• Add T4 DNA Ligase
• Add Marathon cDNA Adaptor ( )
• 16°C O/N or RT for 3 hr
• Prepare working dilution for RACE reactions
Region of overlap (optional)
adaptor-ligated ds cDNA
AP1
GSP2
GSP1
AP1
Positions of primers shown for reference
5'-RACE PCR
• Add GSP1 and AP1 Primers
• Add Advantage 2 Polymerase Mix
• Commence thermal cycling
3'-RACE PCR
• Add GSP2 and AP1 Primers
• Add Advantage 2 Polymerase Mix
• Commence thermal cycling
Upper strand cannot be extended or serve as template
No binding site for AP1
GSP2
AP1
GSP1
No binding site for AP1
Lower strand cannot be extended or serve as template
AP1 binding site created
AP1 binding site created
first PCR cycle
AP1
GSP1
GSP2
AP1
5'-RACE PRODUCT
Region of overlap
Region of overlap
3'-RACE PRODUCT
Exponential PCR
CHARACTERIZE RACE PRODUCTS

RNA:DNA hybrid. This nicked or degraded RNA acts to prime second strand synthesis, while the DNA polymerase fills in the gaps; this is a continuous process until all of the RNA along the strand is degraded and the entire second DNA strand is formed. DNA ligase then acts to seal the fragments together into a complete strand. While the majority of second strand synthesis is primed in this way, up to 15% of the second strand synthesis is primed by hairpin loop. Following the generation of the double stranded cDNA, the molecule is blunt-ended with T4 DNA polymerase, and the Marathon cDNA adaptors are attached using DNA ligase. The Marathon cDNA adaptors have the same features as those used in the Genome Walker Kit. The result of the ligation step is an uncloned library of adaptor ligated double stranded cDNA.

***RACE PCR***

The RACE PCR protocol incorporates the same features as the RAGE PCR described above. Like the Genome Walker Kit, similar adaptor and adaptor primer design, together with the utilisation of touchdown PCR minimise non-specific amplification. Amplification of the 3' and 5' cDNA ends was carried out using Advantage 2 Polymerase Mix (Clontech), which contains Advantaq DNA polymerase, a small amount proof-reading enzyme, and TaqStart antibody for automatic hot starts. Gene specific primers were designed from non-repetitive sequences within the open reading frame of CDF, approximately 550 bp downstream from the 5' end and 1450 bp upstream from the stop codon. The 5' RACE products were generated using the antisense GSP1-163C with the primary adaptor primer (RACE-AP1), followed by a second round of nested PCR amplification with the anti-sense nGSP2-89C and the nested adaptor primer (RACE-AP2). A second 5' RACE experiment used primers GSP1-1458C/nGSP2-

1433C. The 3' RACE product was generated using the sense primer GSP1-5862 and RACE-AP1 followed by a second round of nested PCR amplification with nGSP-5880 and RACE-AP2.

The double stranded cDNA was diluted 50 fold in Tricine-EDTA buffer, and added to the PCR reaction mix according to the kit protocol. The reactions were cycled in a Perkin Elmer Thermal Cycler 480 as follows: 94°C for 1 min (1 cycle), 94°C for 30 sec, 72°C for 4 min (5 cycles), 94°C for 30 sec, 70°C for 4 min (5 cycles), and 94°C for 30 sec, 68°C for 4 min (20 cycles). The PCR mix was then diluted 50 fold, and nested PCR reactions assembled and cycled according to the kit protocol.

Positive controls included performing cDNA synthesis, adaptor ligation and PCR amplification with the polyA$^+$ RNA provided in the kit to ensure all steps in the kit were functioning. Once this was determined, all further RACE experiments used pre-prepared adaptor-ligated cDNA supplied in the kit as template for 5' and 3' RACE to assess the success of each experiment. One negative control entailed PCR amplification of the reaction mixes with each primer pair but with template DNA omitted to ensure all master mixes were not contaminated. The second, negative control used template DNA with only the GSP primers, to ensure that the amplified sequences were not a result of only one primer.

RACE PCR products were analysed on a 1.2% agarose gel. Those reaction mixes found to contain single product bands (or smears thereof) were used for cloning.

## LONG-PCR AMPLIFICATION OF CDF-ORF

The CDF-ORF sequence was amplified from fresh high molecular weight genomic DNA from *T. vaginalis* 202 following modified protocols from Cheng *et al.*

(1994) and the Genome Walker Kit. Primers were designed 9 bp upstream of the start codon and 78 bp downstream from the 3' end of the cDNA at positions 868 and 7451C (primers ORF-868/ORF-7451C), with melting temperatures of 69°C (calculated by Genosys) so that touchdown PCR could be used to prevent mis-priming at secondary priming sites. Given that most thermostable polymerases incorporate 40-50 bases/sec, the annealing/extension time was calculated using the following formula: 1 min + (2.5sec/100)(6600 bases)= ~ 4 min (P. Estep, personal commmunication). The reaction mix was assembled as follows: 26.8 μL of water, 4 μL of 10X Clontech Tth buffer, 2.2 μL of $Mg(OAc)_2$ (25 mM), 4 μL of dNTP mix (2.5 mM), 1 μL of each primer (10 pmol/μL), 1 μL of genomic DNA (5 ng/μL). A separate enzyme mix was prepared, containing (per reaction) 8 μL of water, 1μL of 10X Tth PCR buffer (Clontech), and 1 μL of (2.5 units/μL) rTthXL (Perkin-Elmer). A manual hot start was performed, where the DNA mix was heated to 94°C, at which point 10 μL of enzyme mix was added. The cycling parameters were adapted from RAGE PCR, as follows: 94°C for 1 min (hot start), 94°C for 25 sec, 70°C for 4 min (7 cycles); 94°C for 25 sec, 65°C for 4 min (39 cycles); 67°C (extension) for 7 min. Each primer was tested independently in the same reaction conditions to ensure that products were not formed as a result of mis-priming by one primer.

## RNA PRIMER EXTENSION ANALYSIS

Primer extension was performed using synthetic anti-sense oligonucleotides (Tv-PE1, Tv-PE2) that were complementary to the regions of the *T. vaginalis* CDF cDNA at positions 89C and 163C from the 5' end of CDF-2.

Originally, primer extension was performed using the AMV-Reverse Transcriptase Primer Extension System (Promega) and accompanying protocol. Briefly, the primers and markers were first end-labelled with [$\gamma$-$^{32}$P] ATP using T4 polynucleotide kinase. Radiolabelled primer (100 fmol) was then annealed to either 100 µg of total or 0.6 µg of polyA$^+$ RNA in AMV primer extension buffer at 64°C or 58°C for 30-60 min. The primer was extended with 1 unit of AMV Reverse Transcriptase in a 20 µL of reaction mix containing 1X Primer Extension Buffer and 1.4 µL of 40 mM sodium pyrophosphate at 42°C for 30-60 min. Two times loading dye was then added, the nucleic acids denatured at 94°C for 10 min, and ¼ (10 µL) of the mix loaded onto a sequencing gel. Products were separated on a 5% polyacrylamide sequencing gel (19:1 acrylamide/bisacrylamide, 1X TBE, 48% urea) using a Gibco-BRL model S2 sequencing apparatus at 50 watts (Bio-Rad PowerPac 3000) for approximately 2.5 hr until the leading band ran off the gel. The gel was wrapped in plastic and exposed to a phosphor screen (Eastman-Kodak, New Haven, Ct) for 4 days, after which the image was read in a Storm 860 phosphoimager (Molecular Dynamics, Sunnyvale, Ca). Gels and phosphoimaging was performed by Ms. Angela Graham.

To improve the success of the experiment, the protocol was adjusted so that the primer and RNA were hybridised in formamide buffer. One hundred fmol (~1x10$^6$ cpm) of end-labelled primer was co-precipitated with 100 µg of total RNA with ethanol. The nucleic acid pellet was then dissolved in 30 µL of formamide buffer (40 mM PIPES, 1 mM EDTA, 0.4 M NaCl, 80% deionized formamide), the RNA denatured at 90°C for 10 min, and the primers annealed to the RNA at 40°C overnight. The nucleic acids were phenol/chloroform extracted and ethanol precipitated as described in the section on

"Restriction Digests", and the pellets dissolved in 17.4 μL of 1X AMV Primer Extension Buffer to which 1.6 μL of sodium pyrophosphate and 1 μL of AMV reverse transcriptase were added. Following primer extension as above, the nucleic acids were phenol/chloroform extracted and ethanol precipitated again, and dissolved in 10 μL of 2X loading buffer. Nucleic acids were denatured, loaded and electrophoresed as above.

## *IN VITRO* TRANSCRIPTION

A run-off *in vitro* transcript was prepared by *in vitro* transcription of the linearised pCDF-ORF plasmid template by T7 RNA polymerase (Promega) according to the protocol described by the manufacturer. Briefly, the plasmid was linearised at the 3' end of the CDF-ORF insert by restriction with *Bam*H I, extracted with phenol/chloroform and ethanol precipitated as described in the section on "Restriction digests", and dissolved in DEPC-treated water. The transcription reaction was prepared to a final volume of 100 μL by adding 5 μg of linearised plasmid template to a 37°C mixture containing final concentrations of the following: 1X transcription buffer, 10 mM DTT (Promega), 0.8 units/mL RNA-guard ribonuclease inhibitor (Pharmacia), 0.5mM each rNTP (Pharmacia), and 0.4 units/mL T7 RNA polymerase (Promega). The reaction was incubated at 37°C for 3 hr, after which, 5 μL was assayed on a neutral RNase-free TAE (0.04 M Tris-acetate, 2 mM EDTA) agarose gel, and promptly frozen at −80°C.

## *IN VITRO* TRANSLATION

RNA transcripts were translated *in vitro* using the Flexi Rabbit Reticulocyte Lysate system (Promega) using the protocols provided by the manufacturer. Briefly, the *in vitro* translation mix was prepared as follows: 33 μL of Rabbit Reticulocyte Lysate,

1μL of amino acid mix minus methionine (1mM), 3 μL of [$^{35}$S]methionine (1200 Ci/mmol, 10 mCi/mL), 1 μL of MgOAc (25 mM), 1.4 uL of KCl (2.5 mM), 1μL of DTT (100 mM), 40 units of RNasin ribonuclease inhibitor, and 8 μL of *in vitro* translation mix. A parallel mix that omitted the transcript was prepared for the negative control. The translation reaction was incubated at 37°C for up to 7 hr, after which the products were separated directly by a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), or first immunoprecipitated, and then separated. *In vitro* translated protein products were detected by autoradioblot or by immunoblot.

## PROTEIN ANALYSIS

### *Immunoprecipitation*

Rabbit reticulocyte lysates containing *in vitro* translation products were diluted to 350 μL with RIPA buffer (50 mM Tris-HCl pH 7.2, 150 mM NaCl, 1mM PMSF, 1% sodium deoxycholate, and 1% (v/v) Triton-X). A 1:15 dilution of rabbit anti-serum raised against a purified preparation of CDF and 100μL of protein A Sepharose 4 Fast Flow beads (Pharmacia) diluted 2/9 (approximately 90 μL of packed beads/mL, with 7.7 mg of IgG binding capacity and 1.3 μg of protein A per mL of drained gel) were added to the lysates. After the samples were allowed to mix overnight at 4°C, the beads were pelleted by centrifugation at 13,000 rpm (microfuge) and washed 5 times in 1 mL each of RIPA buffer. After the final wash, the beads were resuspended in 30 μL of 2X sample buffer (0.125 M Tris pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, and 0.01% bromophenol blue), denatured 6 min at 95°C, centrifuged at 13,000 rpm, and loaded onto an SDS-PAGE gel. Negative controls included immunoprecipitation of the *in vitro*

translation mix with pre-immune serum, as well as immunoprecipitation of blank *in vitro* translation mix (transcript omitted) using anti-CDF immune serum.

***Polyacrylamide Gel Electrophoresis (Laemmli, 1970)***

Proteins were separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE). Samples were first mixed with 2X sample buffer, denatured 6 min at 90-95°C, and centrifuged at 13,000 rpm for 30-60 sec. Separating gels consisted of 0.375 M Tris-HCl (pH 8.8), 0.1% SDS, and 5-7.5% acrylamide (Protogel, 30% acrylamide, 0.8% bis-acrylamide, National Diagnostics, Atlanta, Georgia). Stacking gels consisted of 0.125 M Tris-HCl pH 6.8, 0.1% SDS, and 4.5% acrylamide (Protogel). Electrophoresis was carried out at 125 V in electrophoresis buffer (25 mM Tris, 192 mM glycine, and 0.1% SDS) using the Bio-Rad (Mississauga, ON) Mini-Protein II Electrophoresis cell system. Following electrophoresis, the proteins were stained by soaking gels in Coomassie Brilliant Blue (40% methanol, 7% acetic acid, and 0.1% Coomassie Brilliant Blue), followed by a destaining step in 40% methanol, 7% acetic acid. Prior to drying, the gels were soaked for 10 min in a final solution containing 7% acetic acid, 7% methanol, and 1% glycerol to prevent cracking. Gels were dried under vacuum pressure at 60°C for 40 min (Bio-Rad model 583 gel dryer), cooled, and exposed to Cronex X-ray film overnight to several days at −80°C.

***Western Blot and Immunodetection***

Following SDS-PAGE electrophoresis, proteins were transferred from gels onto Hybond-C Nitrocellulose membranes (Amersham). Prior to electroblotting, the membrane and gel were equilibrated in fresh transfer buffer (25 mM Tris, 192 mM

glycine, 20% (v/v) methanol) for 15 min. The transblotting cassette was assembled according to the instructions provided by BioRad, and the transfer performed at 100 volts for 1 hr in ice-cold buffer. For determination of molecular size, pre-stained high range protein molecular weight standards (Bio-Rad) were used.

Immunological screening of the Western blots was carried out using rabbit anti-serum raised against a purified preparation of CDF. Optimisation experiments for the Western blots (performed with a purified preparation of CDF and blotted with the rabbit anti-serum raised against CDF) were based on those previously published (Garber *et al.*, 1989) and modified until optimal. Membranes were blocked by incubation in 5% skim milk twice for 30 min each at 37°C, and washed 5 times at room temperature for 5 min each in PBST (PBS containing 0.1% Tween 20) in the Robbins Scientific Microhybridization Chamber. The blots were then adsorbed for 1-2 hr at 37°C with anti-CDF anti-serum diluted 1:100 in 5% skim milk in PBS. Following five quick washes in PBST, the blots were incubated at 37°C for 1 hr in a 1:2000 dilution of BioSource Goat Anti-Rabbit Immunoglobulin (G&I) Horse-Radish Peroxidase Conjugate (IgG-HRPO) (Immunocorp, Montreal, PQ) in 1% skim milk in PBS. The washing was repeated with PBS alone, and the blots were developed in 3-amino-9-ethyl carbazole (3AEC) containing 1 μL/mL of 30 % $H_2O_2$ until the desired intensity was reached, and then washed 3 times in water. 3AEC was prepared by mixing 45 mL of N,N dimethylformamide containing 180 mg of AEC and 105 mL of 0.05 M sodium acetate pH 5.5, and filtering through a 0.22 μm filter.

# CHAPTER 3: CHARACTERISATION OF A CANDIDATE GENE FOR CDF

The goal of this project was to determine whether the CDF-2 cDNA clone, previously identified by immunoscreen of a cDNA expression library, was a potential candidate gene for CDF. While the clone was characterised as having a 3060 bp open reading frame (ORF), it lacked several genetic components such as transcriptional start and stop sites and an initiator element. Despite its lack of homology with known nucleic acids and peptides, the clone was hypothesised to be a partial coding sequence for CDF. To answer this hypothesis, the sequences flanking CDF-2 ORF were characterised.

The first stage of this project was to identify the sequences flanking CDF-2, so that the entire ORF, including the translational start and stop codons, as well as the transcriptional initiation and termination sequences could be characterised. This was to have been done by screening an EMBL3 genomic DNA library. In actual fact, several approaches, including screening the genomic DNA library, performing inverse PCR, and finally walking along the genome from CDF-2 using RAGE PCR, were attempted before sequence could be obtained.

The second stage of this project was to determine whether the characterised gene was truly a candidate for CDF. *In vitro* transcription and translation experiments were performed to answer the questions: "Does the characterised gene produce an appropriately sized transcript?" and "Does the transcript produce a protein that can be identified using rabbit serum raised against a purified preparation of CDF?"

This chapter discusses the three approaches used to identify sequences flanking CDF-2, the characterisation of that sequence, the construction and cloning of the full

length open reading frame, its use in *in vitro* transcription, and the characterisation of the resulting *in vitro* translation products.

***Screening of the EMBL3 T. vaginalis Genomic DNA Library***

The EMBL3 phage library was plated on *E. coli* KW251, the plaque DNA was transferred to nylon discs, and the library screened by hybridisation with $^{32}P$ radiolabelled probes. Initially, screening of the genomic library by hybridisation was done using the full length $^{32}P$ radiolabelled CDF-2 as a probe. This screening identified over 95 positive phage clones in the primary screening, and all were rejected after subsequent screens showed them to be negative. It was thought that the repeated elements in the sequence were contributing to false positives.

In an effort to reduce the number of false positives, CDF-2U, a 645 bp fragment free of repetitive elements, was isolated from CDF-2 and used to re-probe the library (Figure 1). This second library screening identified 5 positive clones in the primary round. Following secondary, tertiary and quaternary screens, however, the clones could not be conclusively determined negative or positive. The addition of more accurate controls to the experiment (hybridisation of pure populations of CDF-1 and CDF-2 plaque lifts with radiolabelled CDF-2) resolved this issue. The candidate clones were found to be the same intensity as the negative controls, clearly identifying them as negative. At this point the library screening was abandoned.

***Northern Blotting***

To confirm that CDF-2 sequences were transcribed, and to determine the size of the transcript represented by CDF-2, total and polyA$^+$ *T. vaginalis* 202 RNA species were

separated on a 1% glyoxal agarose gel. The RNA in the gels was transferred to nylon membranes, and probed with $^{32}$P-radiolabelled CDF-2U. These Northern blot analyses identified a 6.5-7 kb band (Figure 6) in both the total and polyA$^+$ RNA (Figure 6). This band was found to be DNase-resistant and RNase labile (data not shown), ruling out any possibility of contamination of the RNA by DNA. These results indicate that there is a transcript associated with the CDF-2 cDNA clone, and that its length is estimated to be approximately 6.5-7 kb.

***Southern Blotting***

In an effort to identify suitably sized target restriction fragments for cloning, Southern blot hybridisation and restriction mapping were employed. For Southern blot analysis, high molecular weight *T. vaginalis* 202 genomic DNA was digested with various restriction enzymes or combinations thereof, separated on a 0.5% agarose gel, and transferred to nylon membrane. The blots were probed with $^{32}$P-radiolabelled CDF-2U. Southern blots of DNA digested with *BamH*I, *Bgl*II, *Sac*I, *Sal*I, *Sca*I, *Xho*I probed with CDF-2U identified no bands below the size of the high molecular weight DNA (data not shown). Restriction fragments identified by the hybridisation with CDF-2U include: a 3 kb *Dra*I fragment, a 2 kb *Ase*I fragment (Figure 7, Panel A), two *Taq*I fragments (1.6 and 4.2 kb), an 8.7 *Eco*RI fragment, an 8.2 *Rsa*I fragment, two major *Pst*I fragments (>23 kb and 7.9 kb), and a 6.8 kb *Bst*BI fragments (Figure 7, Panel B). The orientation of the *Taq*I fragments was verified by probing with 270 bp and 370 bp *Nla*III/*Taq*I fragments of CDF-2U (Figure 7, Panel C). Using known restriction sites within the CDF-2 sequence and the data obtained from Southern Blotting, a restriction map was constructed (Figure 7, Panel D). The *Bst*BI sites were difficult to map due to poor signal strength and size

**Figure 6: Northern Blot Analysis of *T. vaginalis* 202 Total and Polyadenylated RNA**

A 6.5-7 kb band was identified in both the total and polyA$^+$ RNA. 2.5 μg of polyA$^+$ RNA (mRNA), 9 μg of total RNA, and 35 μg total RNA were separated on a 1% glyoxal agarose gel, and transferred to a nylon membrane. The blot was probed with randomly primed $^{32}$P labelled CDF-2U.

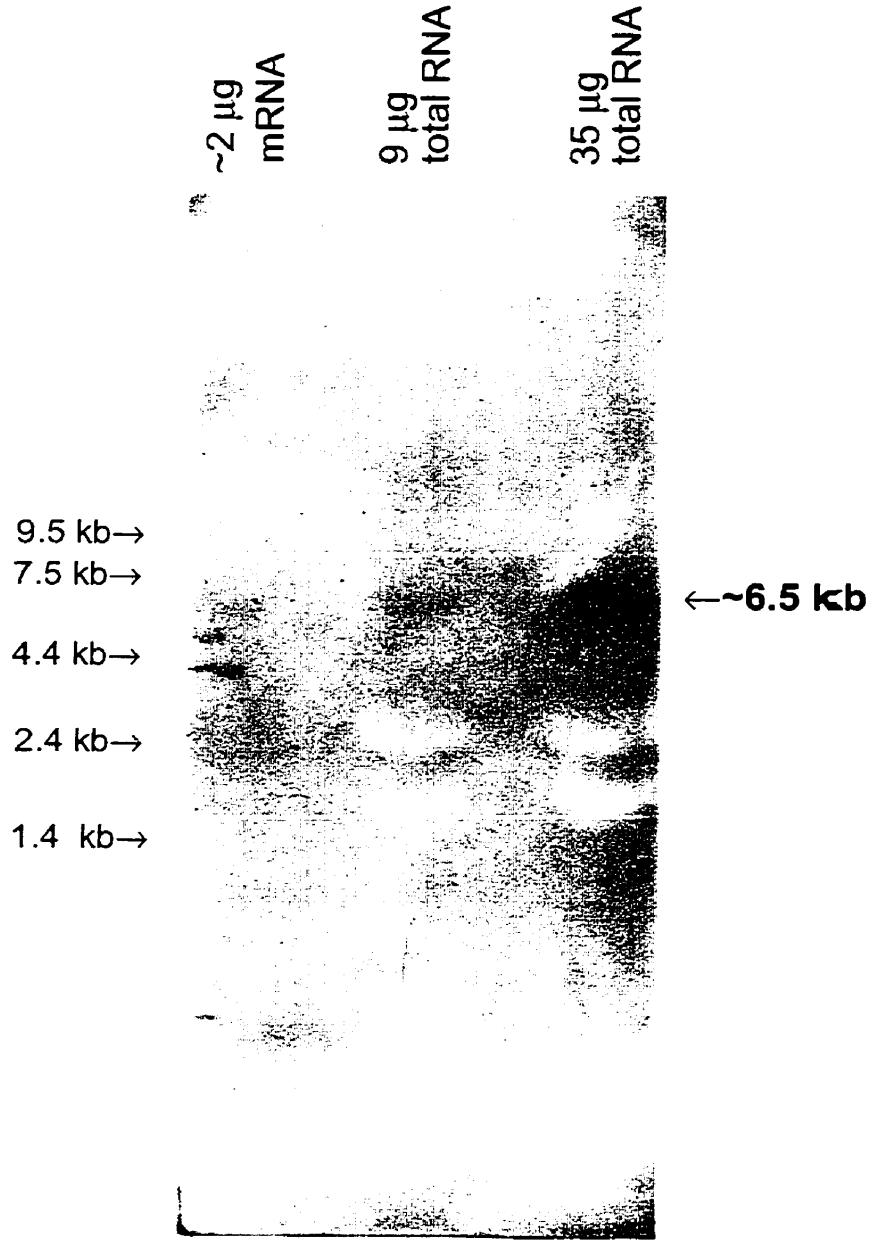
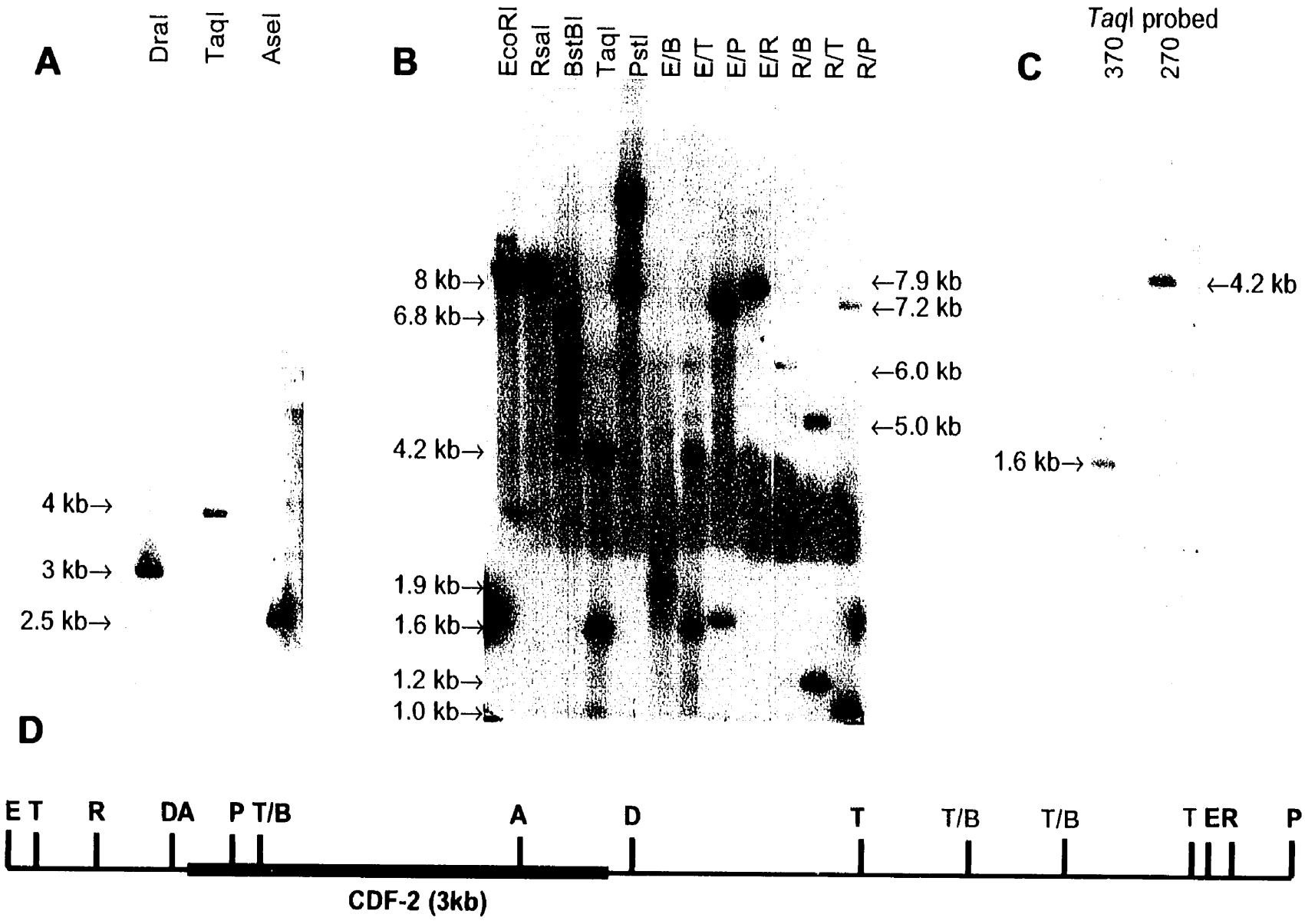

~2 µg mRNA
9 µg total RNA
35 µg total RNA
9.5 kb→
7.5 kb→
4.4 kb→
2.4 kb→
1.4 kb→
←~6.5 kb

**Figure 7: Southern Blot Analysis and Restriction Mapping of *T. vaginalis* 202 High Molecular Weight Genomic DNA**

*T. vaginalis* 202 high molecular weight genomic DNA was digested and double digested with *Dra*I (D), *Taq*I (T), *Ase*I (A), *Eco*RI (E), *Rsa*I (R), *Bst*BI (B), and *Pst*I (P) and separated on 0.5% (Panels A and B) and 1.2% (Panel C) agarose gels. The DNA was transferred to nylon membranes by conventional Southern transfer. The blots were probed with randomly primed $^{32}$P labelled CDF-2U (Panels A and B), the 370 bp portion of CDF-2U or the 270 bp portion of CDF-2U (Panel C). CDF-2U is 645 bp long and is situated within the 5' third of CDF-2. TaqI cuts CDF-2U into 370 (5') and 270 (3') fragments. For a diagram see Figure 1. Using the information obtained from these Southern blots, a restriction map was constructed (Panel D). *Taq*I/*Bst*BI sites in grey are sites that exist in the sequence but are beyond the reach of the probe and could not be picked up by Southern Blot.


A
DraI
TaqI
AseI
4 kb→
3 kb→
2.5 kb→
B
EcoRI
RsaI
BstBI
TaqI
PstI
E/B
E/T
E/P
E/R
R/B
R/T
R/P
8 kb→
6.8 kb→
4.2 kb→
1.9 kb→
1.6 kb→
1.2 kb→
1.0 kb→
←7.9 kb
←7.2 kb
←6.0 kb
←5.0 kb
C
TaqI probed
370
270
←4.2 kb
1.6 kb→
D
E T
R
DA
P T/B
A
D
T
T/B
T/B
T ER
P
CDF-2 (3kb)

inconsistencies in the fragments produced. Because there is a unique *Bst*BI site within CDF-2, two bands were expected by hybridisation of the Southern blot with CDF-2U, while one distinct band (6.8 kb) and several faint bands (4.3, 4.6, 5.1, 5.4, 6.0 kb) were produced. Comparison of the positions of the *Bst*BI sites determined from sequence analysis indicated that one of the two expected fragments should have been 4.3 kb (the size of the upstream fragment could was not be determined). The location of *Bst*BI sites further downstream indicates that the other faint bands, and possibly the 6.8 kb band, produced on the Southern blot were likely due to incomplete digestion by *Bst*BI. Another inconsistent result was the *RsaI*/*Taq*I double digest. The 5.0 kb band in the *Rsa*/*Taq*I double digest is too big, given the location of the *Taq*I site in CDF-2U and the deduced *Taq*I sites by other digests. Later comparison of the Southern Blot with the restriction map generated following sequence analysis indicates that this may also be due to incomplete digestion. An 8.7 kb *Eco*RI fragment, an 8.2 *Rsa*I fragment, and two *Taq*I fragments (1.6 and 4 kb) were identified as potential candidates for inverse PCR cloning.

***Inverse PCR***

In an attempt to clone the 8.7 kb *Eco*RI fragment and the two *Taq*I fragments, the inverse PCR strategy was used. Following restriction digestion, the DNA was self-ligated into circular form with T4 DNA ligase, and then used as a template for inverse PCR using opposing primers. The *Eco*RI fragment was to be amplified using primers 95/710, while the *Taq*I fragments were to be amplified using 95/184 and 631/710 (designed from positions 95, 184, and 710 of CDF-2) (Figure 2). Despite several attempts with varied PCR reaction components, such as template and magnesium concentrations, and varied cycling parameters such as annealing temperature and time,

number of cycles, and length of extension time, no clean, distinct, appropriately sized amplicons could be produced.

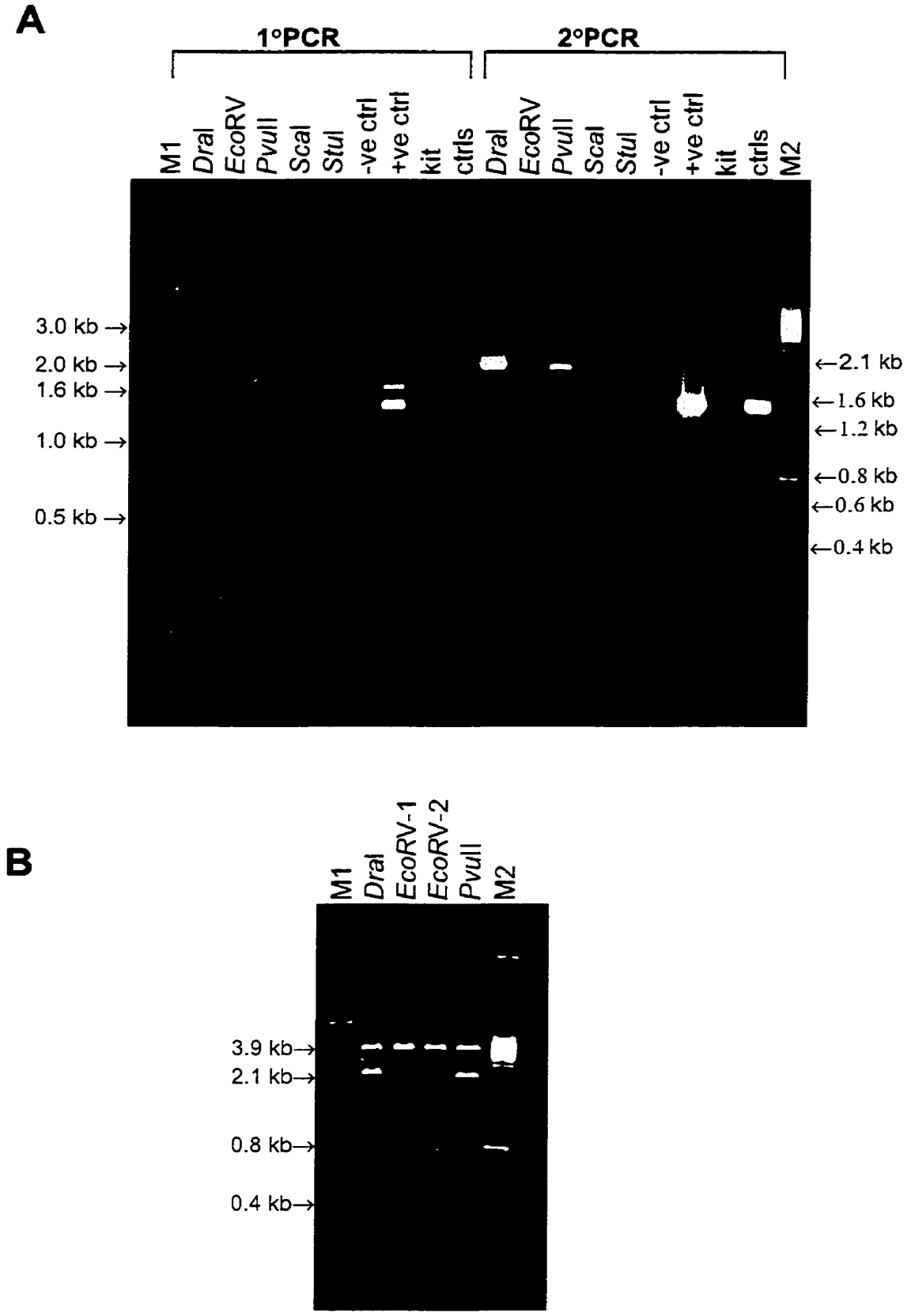
***Genome Walking and PCR Amplification of the Full-length ORF***

A third and final approach to clone the regions flanking CDF-2 was to prepare uncloned genomic DNA libraries, which would serve as template DNA for RAGE PCR. Five original genomic DNA libraries had been prepared by Mr. Dino Petrin by digesting high molecular weight *T. vaginalis* genomic DNA from isolate DG with *Dra*I, *Eco*RV, *Pvu*II, *Sca*I, and *Stu*I, and ligating adaptors onto the ends of the restriction fragments to yield five uncloned adaptor-ligated genomic DNA libraries. A sixth library was similarly prepared using *T. vaginalis* 202 high molecular weight DNA digested with *Eco*RI and first blunt-ending the restriction fragments before the adaptors were ligated. The restriction fragments that contained the CDF-2 sequence were amplified using gene-specific primers GSP1 (designed from CDF-2) and adaptor primers. Secondary nested PCR was then performed on the primary PCR products using nested gene-specific primers nGSP2. The PCR products were analysed by gel electrophoresis (Figure 8, Panel A). As sometimes occurs, PCR products were not generated in concentrations high enough to be seen in the primary PCR, despite performing extra cycles as per kit instructions. Analysis of the secondary PCR products was done to determine which were suitable for cloning. Cloned products were also analysed by gel electrophoresis to select those for sequencing (Figure 8, Panel B).

Once the clones from each step were sequenced and arranged in contiguous sequence (contigs), new primers were designed for the next walk. The process was initiated from CDF-2, walking 5' upstream and 3' downstream of the known sequence

**Figure 8: Primary and Secondary (Nested) RAGE PCR Products**

Panel A shows the products generated by nested RAGE PCR from the *T. vaginalis* DG high molecular weight Genome Walker libraries (*Dra*I, *Eco*RV, *Pvu*II, ScaI, and *Stu*I) using primers Gsp1-5142 and nGsp2-5160 in conjunction with AP1 and AP2 respectively. The primers were designed from the sequence flanking CDF-2, which was obtained in Walk 2. Primary and secondary (nested) PCR was performed using rTth XL (Perkin-Elmer/Applied Biosystems) and the cycling parameters previously described. Products were only visible after nested PCR was performed. The process from creating the libraries to performing the amplification of the libraries was tested using human genomic DNA included in the kit, which was then digested with *Dra*I, adaptor-ligated and amplified using primers PCP1/AP1 and nPCP2/AP2 provided in the kit (+ve ctrl). The negative control (-ve ctrl) to rule out products derived from contamination consisted of performing the primary and secondary PCR with GSP/AP primers only in the absence of DNA. The kit also included a pre-made human library, which was amplified using PCP1/AP1 and PCP2/AP2 to ensure that the enzyme was active and cycling conditions were appropriate. Similarly, PCR was performed on the human primers alone (in the absence of DNA) to ensure that no contamination existed. These last two controls were labelled "kit controls". The PCR products from the *Dra*I, *Eco*RV, and the *Pvu*II libraries were cloned directly from PCR mix into pCR2.1 TOPO. Panel B shows isolated plasmid clones (digested with *Eco*RI to release the insert) from each of amplifications (*Dra*I: 2.1 kb, *Eco*RV-1: 0.4 kb, *Eco*RV-2: 0.8 kb, *Pvu*II: 2.1 kb), which were sequenced. PCR products (5μL) and cloned products were analysed on 1.5% agarose gels. Molecular weights are indicated by a 1 kb Ladder (M1)(Gibco-BRL) 100 bp Ladder (M2)(Amersham Pharmacia Biotech).


A
1°PCR
2°PCR
M1
DraI
EcoRV
PvuII
ScaI
StuI
-ve ctrl
+ve ctrl
kit
ctrls
DraI
EcoRV
PvuII
ScaI
StuI
-ve ctrl
+ve ctrl
kit
ctrls
M2
3.0 kb →
2.0 kb →
1.6 kb →
1.0 kb →
0.5 kb →
←2.1 kb
←1.6 kb
←1.2 kb
←0.8 kb
←0.6 kb
←0.4 kb
B
M1
DraI
EcoRV-1
EcoRV-2
PvuII
M2
3.9 kb→
2.1 kb→
0.8 kb→
0.4 kb→

using Gsp1-163/nGsp2-89 and Gsp1-3876/nGsp2-3889 respectively. Each walk identified new sequence from which the primers for the next walk were designed. The walking progressed using the following primers: Gsp1-201/nGsp2-73 (walk 3); Gsp1-5142/nGsp2-5160 (walk 4); Gsp1-7224 /nGsp2-7284 (walk 5). A map of the contiguous sequence, position of primers, and sequential arrangement of the amplicons from each walk are shown in Figure 9.

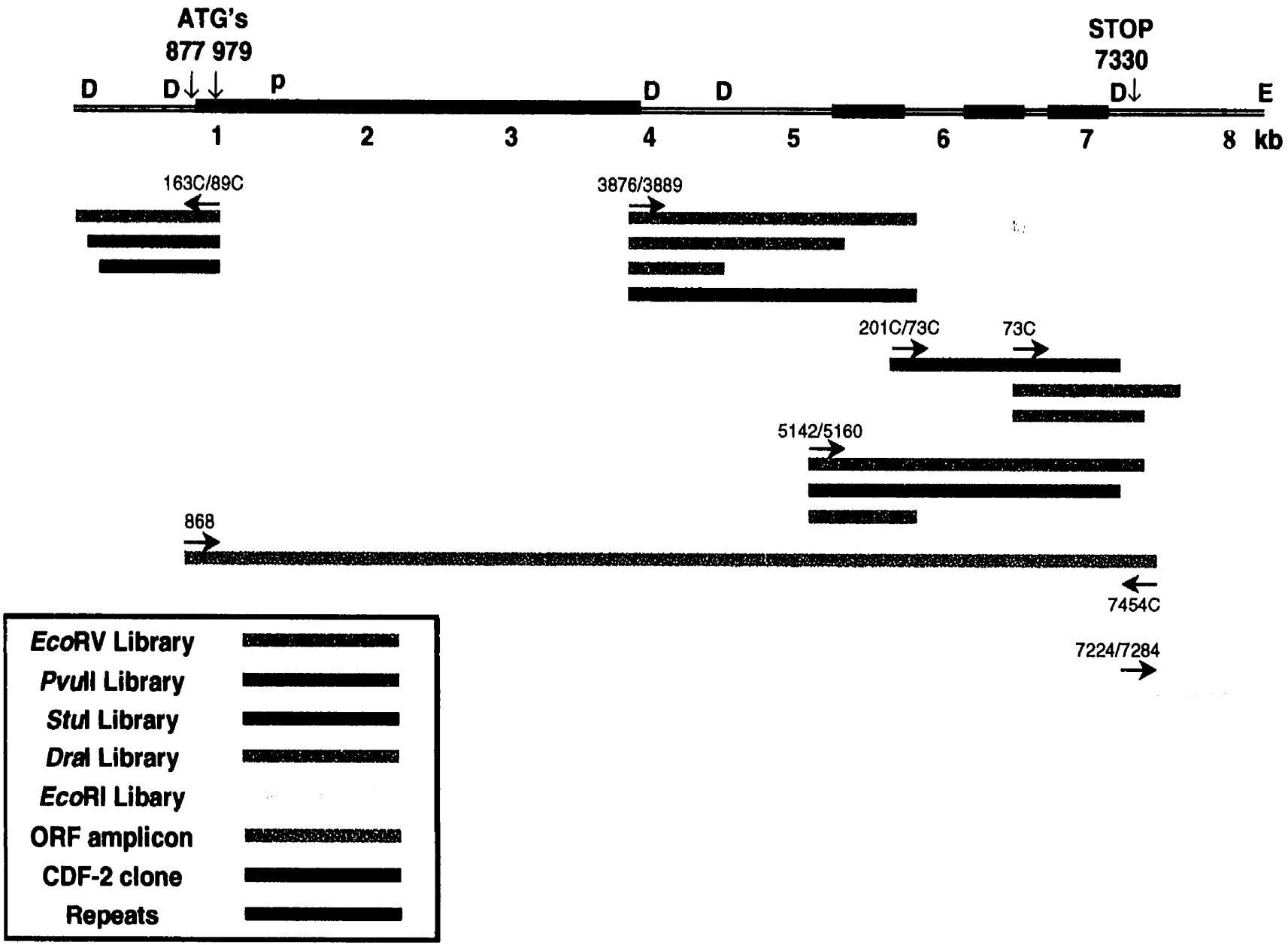
All new sequence data, which extended 883 bp upstream and 3645 bp downstream of CDF-2, was obtained using this genome walking technique. Several clones per walk were sequenced in their entirety so that when the clones were arranged contiguously, both the sense and antisense strands of the entire contig were sequenced in at least triplicate.

Interestingly, when piecing the amplicons together into a contiguous sequence, it was discovered that several clones did not contain the expected restriction sites (*Eco*RV, *Pvu*II, *Stu*I and *Sca*I used to construct the libraries) at their ends, nor were the restriction sites present when the complete sequence of the contig was determined (Figure 9). Of the five original libraries donated by Mr. Dino Petrin, only the library prepared by digestion with *Dra*I, a frequent cutter, produced amplicons with *Dra*I sites on their ends. It was also noted that there were two different, but true, binding sites for primer 73C, and that there were three repeats of approximately 450 bp present within the sequence (Figure 9). Together, these data then raised the questions about the integrity of the DNA used to produce the original libraries and whether the amplicons were truly contiguous.

The contiguity of the sequence was indeed confirmed by comparing the predicted restriction sites from the sequence with the restriction map obtained from the Southerns.

**Figure 9: Genome Walking Map**

This map shows the sequential steps made by RAGE and the construction of the 8.25 bp contiguous sequence. The step-wise progression of acquiring new sequence was initiated from CDF-2, walking 5' upstream and 3' downstream of the known sequence using Gsp1-163C/nGsp2-89C and Gsp1-3876/nGsp2-3889 respectively. As new sequence was characterised, the process was repeated (genome walking), until adequate sequence was obtained, using the following primers: Walk 3 Gsp1-201C/nGsp2-73C. Walk 4 Gsp1-5142/nGsp2-5160. Walk Gsp1-7224 /nGsp2-7284. The six restriction-fragmented genomic DNA libaries are colour coded to indicate from which libraries the clones were obtained. *Pvu*II (P), *Dra*I (D), and *Eco*RI (E) sites are indicated. There are no *Eco*RV or *Stu*I, or *Sca*I sites in the contiguous sequence. The open reading frame (ORF) was amplified end to end using primers 868 and 7454C. The start codons at positions 877 and 979 as well as the stop codon at position 7330 of the ORF are indicated on the map, as well as positions of the 450 bp repeated elements.

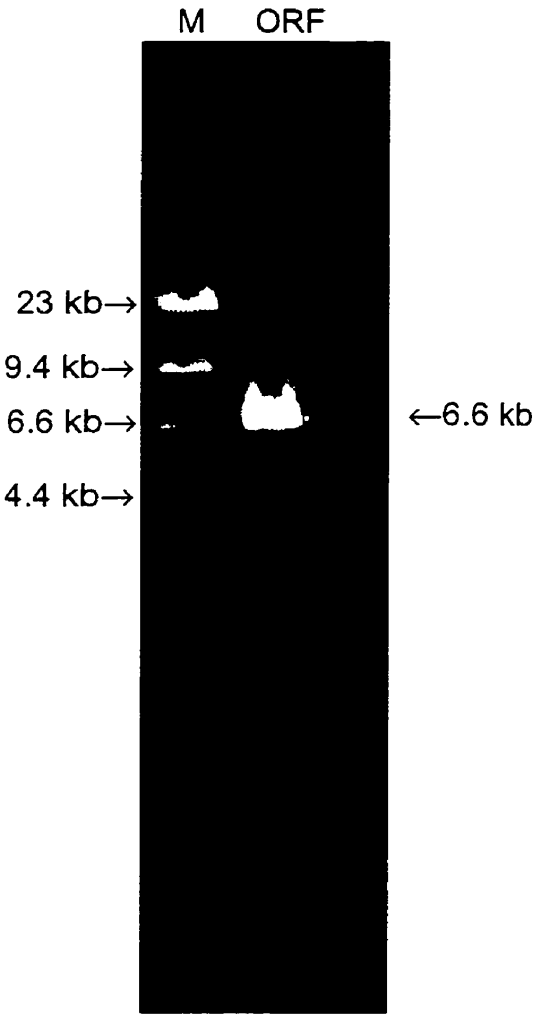

ATG's
877 979
STOP
7330
D
D
p
D
D
D
E
1
2
3
4
5
6
7
8 kb
163C/89C
3876/3889
201C/73C
73C
5142/5160
868
7454C
7224/7284
EcoRV Library
PvuII Library
StuI Library
DraI Library
EcoRI Libary
ORF amplicon
CDF-2 clone
Repeats

Unfortunately, the distal 5' *Eco*RI and *Taq*I sites, as well as the distal 3' *Rsa*I and *Pst*I sites of the restriction map (Figure 7, Panel D) could not be confirmed because they extended beyond the known sequence. The sequence was proved to be contiguous by end to end amplification of the entire ORF with ORF-868/ORF-7451C (Figure 10), which was then cloned into pCR2.1 (pORF). A series of restriction digests of pORF with *Acc*I, *Ase*I, *Bcl*I, *Bsp*DI, *Bst*BI, *Bst*XI, *Dra*I, *Hin*cII, *Pst*I, *Psp*1406I, and *Sty*I further confirmed the predicted pattern of restriction sites within the sequence. Because *Bcl*I and *Sty*I were blocked by methylation of the recombinant plasmid, these restriction sites were confirmed by digestion of the uncloned PCR product (data not shown).

It is suspected that the DNA used for the construction of the first five DG libraries was of poor quality, and perhaps nicked, damaged, or broken before the adaptors were ligated. When the *Eco*RI, *Eco*RV, *Pvu*II, and *Stu*I libraries were prepared using fresh *T. vaginalis* 202 high molecular weight genomic DNA, and the walks repeated with the new libraries, only the *Eco*RI library yielded amplicons that had *Eco*RI restriction sites in the expected locations (Figure 9). For the other three libraries, either no products, or only barely visible high molecular weight products, were amplified (data not shown). The difference between the amplification patterns of the freshly prepared libraries versus the original libraries is striking. The observation that, of the original libraries, only the frequent cutter, *Dra*I, produced a library from which amplified fragments terminate with *Dra*I sites is significant as well. Together, these data suggest that *Eco*RV, *Pvu*II and *Stu*I restriction fragments containing CDF-2 were too large to be amplified (the limit of the kit is about 6.5 kb) and that amplification of short, nicked DNA was favoured in the original libraries. Despite the fact that the original libraries were not ideal for RAGE PCR, the

**Figure 10: PCR Amplification of the CDF-ORF**

The 6.6 kb ORF was amplified from *T. vaginalis* 202 high molecular weight genomic DNA by long range PCR using primers ORF-868 and ORF-7451C which are situated just upstream and downstream of the presumed start codon and the polyadenylation site respectively. The PCR products (5 μL) were visualised on a 0.8% agarose gel. Molecular weights are indicated by the Lambda DNA (λ) digested with *Hin*dIII marker (M) (Gibco/BRL).


M
ORF
23 kb→
9.4 kb→
6.6 kb→
4.4 kb→
←6.6 kb

sequence obtained from them seems consistent, contiguous, and reliable. This is supported by the fact that amplification of the *Eco*RI library yielded expected sized products with *Eco*RI sites at the expected locations, and yielded contiguous sequence.

***RACE PCR***

To identify the 5' untranslated region and the 3' polyadenylation site, 5' and 3' Rapid Amplification of cDNA ends (RACE) was performed (Figure 11). 5' RACE PCR, which was originally initiated at position 163C and 89C (of CDF-2), produced amplicons of expected size (90 bp) from libraries prepared from *T. vaginalis* DG total RNA (data not shown). A second experiment, using primers initiated at 1458C and nested at 1433C, also produced amplicons of expected size (0.5 kb) in the cDNA libraries prepared from *T. vaginalis* DG total RNA and *T. vaginalis* 202 total and polyA$^+$ RNA. The products from both experiments were cloned for sequencing. Two of three clones ended exactly at the CDF-2 5' terminus (position 884), while the third added two bases (CT) upstream of the end of the CDF-2 clone (Figure 12, Panel A).

The 3' RACE PCR, which was initiated at position 5862 and nested at 5880, produced a 1.5 kb amplicon in the cDNA libraries prepared from *T. vaginalis* 202 total and polyA$^+$ RNA (Figure 11). The latter was cloned for sequencing. Three of the 3' RACE PCR clones contained a polyA$^+$ tract beginning at position 7364. Two other clones from the 3' RACE PCR were identical to the other clones (and the sequence obtained from RAGE) from positions 5880 to 7174 (clone 3-6) and 5880 to 7215 (clone 3-7) (Figure 12, Panel A). Clone 3-6 then included 50 bp of unknown sequence, while clone 3-7 then included 240 bp of sequence that appeared to be from a non-contiguous region 6540-6780. The positions at which clones 3-6 and 3-7 become different from the

**Figure 11: RACE PCR Products**

Panel A shows primary 5' and 3' RACE PCR products amplified using respective primers GSP1-1458C/RACE-AP1 and GSP1-5862/RACE-AP1 from cDNA generated from *T. vaginalis* DG polyA$^+$ RNA and *T. vaginalis* 202 polyA$^+$ RNA and total RNA. The sizes of the 5' and 3' products are expected to be 580 bp and 1470 bp respectively, based on the distances of the primers from the start and stop codons. The Marathon system also included Human polyA$^+$ RNA from which cDNA was prepared and amplified in parallel with the *T. vaginalis* cDNA to ensure that all steps of the kit were functioning. This human cDNA was used for 5' RACE using TRF5'/RACE-GSP1, 3' RACE using TRF3'/RACE-GSP1 and an end to end internal amplification using TRF5'/TRF3'. Panel B shows the secondary (nested) 5' and 3' RACE PCR of the *T. vaginalis* cDNA's using nested gene specific and adaptor primers nGSP2-1433C/AP2 and nGSP2-5880/AP2 respectively. The expected sizes of the 5' and 3' nested products are expected to be 556 bp and 1450 bp respectively, based on the distances of the primers from the start and stop codons. RACE PCR products (5μL) were analysed on a 1.2 agarose gel. Molecular weights are indicated by λ *Hin*dIII Marker (Giboc-BRL) and the Ladder Mix (MBI Fermentas), incorporating 100 bp and 1 kb ladders.

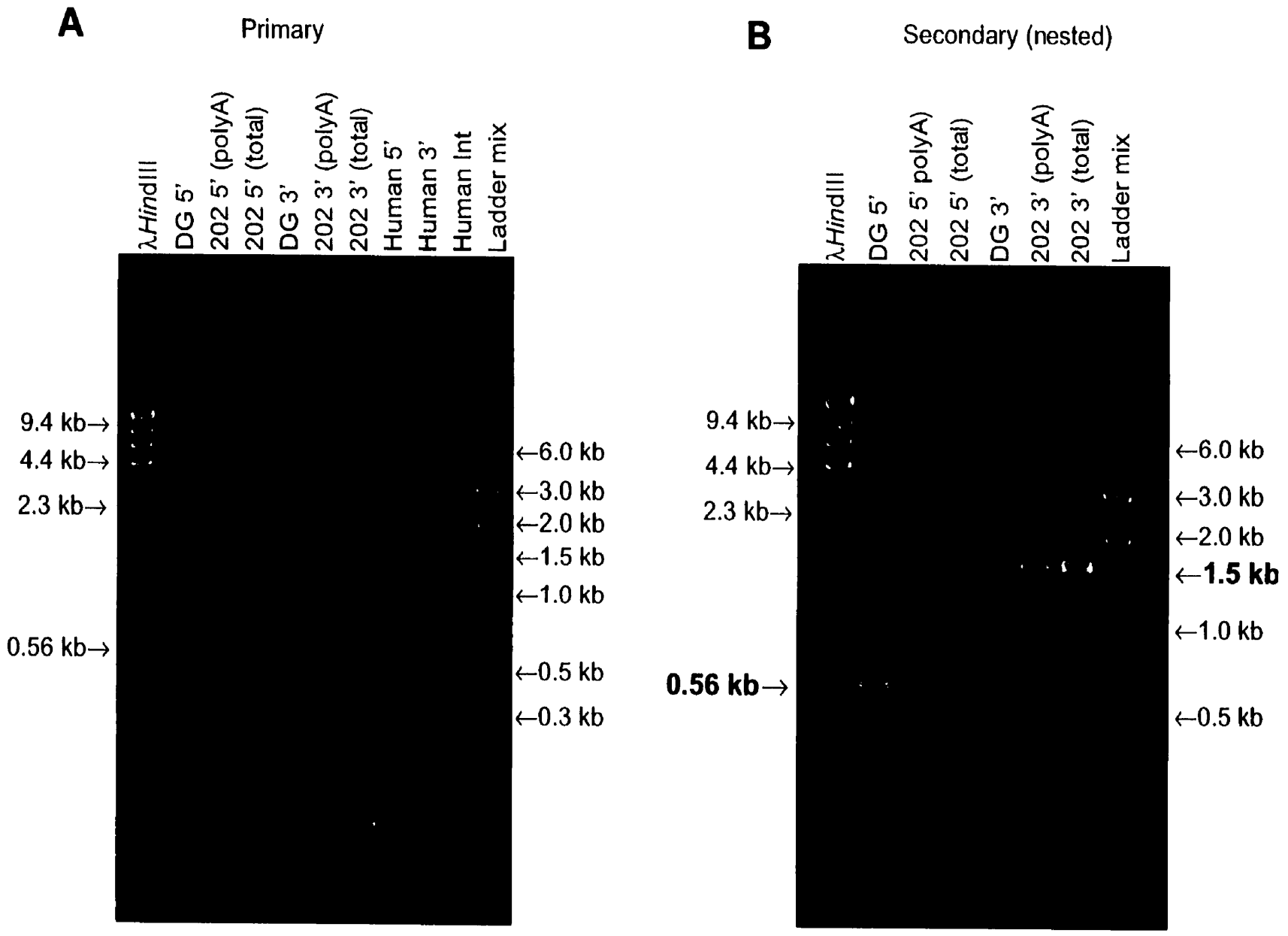
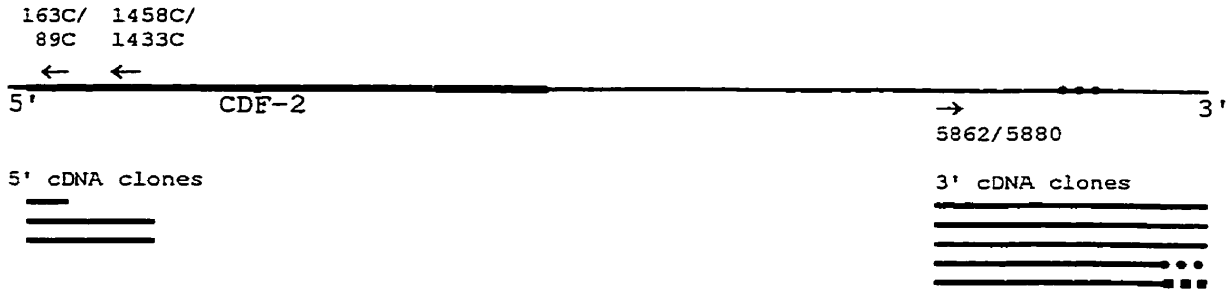

A
Primary
λHindIII
DG 5'
202 5' (polyA)
202 5' (total)
DG 3'
202 3' (polyA)
202 3' (total)
Human 5'
Human 3'
Human Int
Ladder mix
9.4 kb→
4.4 kb→
2.3 kb→
0.56 kb→
←6.0 kb
←3.0 kb
←2.0 kb
←1.5 kb
←1.0 kb
←0.5 kb
←0.3 kb
B
Secondary (nested)
λHindIII
DG 5'
202 5' polyA)
202 5' (total)
DG 3'
202 3' (polyA)
202 3' (total)
Ladder mix
9.4 kb→
4.4 kb→
2.3 kb→
0.56 kb→
←6.0 kb
←3.0 kb
←2.0 kb
←1.5 kb
←1.0 kb
←0.5 kb

**Figure 12: Characterisation of the RACE PCR clones and 5' and 3' untranslated regions.**

Three 5' RACE PCR amplicons were characterised (Panel A) from two different RACE PCR experiments. Two clones ended at position 884 (end of CDF-2), while a third added bases CT at position 882 (Panel B). All three clones were of expected size (90 bp and 550 bp). Five 3' RACE PCR amplicons were characterised (Panel A). All three of the clones were of the 1.5 kb expected size. Three of the five clones had polyA$^+$ tails at the 3' ends. Two other clones were identical to the first three (and the sequence obtained from RAGE) from positions 5880 to 7174 (clone 3-6) and 5880 to 7215 (clone 3-7). Clone 3-6 then included 50 bp of unknown sequence (square dots), while clone 3-7 then included 240 bp of sequence that appeared to be from a non-contiguous region 6540-6780 (round dots) (Panel A). There are two possible start sites **(ATG)** for the ORF at positions 877 and 979 (Panel B). The untranslated region is unlikely to extend past position 802 due to the presence of stop codons in all three reading frames (boxed). Possible *T. vaginalis* promotor sequences (INR) **TCCAT** AND **TCAAT** are not within the usual 15-20 base pairs upstream of the start codon (Panel B). The presence of possible TATA-like boxes near the start codon at position 877 are underlined. The stop codon is found at position 7330 **(TAA)** (Panel C). The 3' untranslated region is 31 bases long. A polyA$^+$ tail is found at position 7364, and a putative destabilising element is found frcm 7354-7363 (underline), flanking the 5' end of the polyA$^+$ tail.

# A RACE PCR Primers and Clones




# B 5' Untranslated Regions

```
800   .AG TAA ATC AAG TTC CAT AAT ATT TTA AAA ATT GTG GAC TCT ACC AAC

847   GAT AAG AGA TTT TGT AAA AAA AAA CTT AGG ATG GAC TCT CAG GCA GCG
                                          Met Asp Ser Gln Ala Ala

890   CCG AAA CGA AGA ATA CAT TTG TTT CGT CAG AAA CAA AAC ATA GAT TGG
      Pro Lys Arg Arg Ile His Leu Phe Arg Gln Lys Gln Asn Ile Asp Trp

944   GGG TTC AAT CCA GAT CCC TTC AAG ATA CCG CCA GAC ATG GCG ...
      Gly Phe Asn Pro Asp Pro Phe Lys Ile Pro Pro Asp Met Ala ...
```

# C 3' Untranslated Region

```
7321  ... TTA AAG AAA TAA AGT AAA AAA GAG TTT TTT ATA TTA TTA TTT TAA
      ... Leu Lys Lys End                                        ↑

7366  AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA
```

three other clones (7174 and 7215 respectively) reflect the end of the last repeat in the open reading frame (Figure 9). Neither 3-6 or 3-7 contained the stop codon at position 7330 nor did they have a polyA$^+$ tract (Figure 12, Panel A). The data suggest that there is a polyadenylation site at position 7364. It is uncertain whether 3-6 and 3-7 are products of artifact, or whether they represent true 3' RACE products that reflect alternative processing or termination. The high A/T content, and long stretches of adenine residues is believed to account for false binding of the oligo d(T) primers during cDNA synthesis, resulting in the partial CDF-2 clone described by (Meysick, 1996).

***Primer Extension Analysis***

To determine whether the ATG at position 877, the ATG at position 979, or some other ATG is the true start codon, and to elucidate the exact 5' end of the transcript, including its untranslated region, Primer Extension Analysis was performed. Despite several attempts using different annealing temperatures, different annealing buffers, and even redesigned primers, Primer Extension Analysis failed to identify the 5' end of the transcript and thus the true 5' untranslated region.

***Sequence Analysis***

Using standard T7, M13 Reverse, SP6, and synthetic oligonucleotide primers, pCDF-2 and the cloned amplicons were sequenced by automated ABI BigDye Terminator Cycle Sequencing (Applied Biosystems). These primers included: pCDF-2: T7, 675, 1246, 2169, 2192C, SP6 pORF: M13 Reverse, T7, 3039, 3876, 5142, 5841, 7201C, 3'RACE M13 Reverse, T7, 73C All other clones, T7 and M13 reverse, and where required 4501, 5841, and 7201C. Automated Sequencing typically yielded 700-

900 bp per reaction (Appendix A). The complete nucleotide sequence is presented in Appendix B and is published under Genbank Accession Number AF257323.

Sequence analysis revealed a potential translational start codon at position 877, in addition to a second, in frame, potential start codon at position 979 (Figure 12, Panel B). Sequence analysis of the region upstream of position 877 indicates that there are no other possible translational start sites in the +1 reading frame which would maintain an open reading frame (a methionine exists at position 694, but it is followed by several stop sites). Further, the presence of a stop codon in the +1 reading frame at position 802 (and other similar stop codons in the other two reading frames), suggest that the coding region could not begin any further upstream (Figure 12, Panel B). It is uncertain which of the two methionines is the true tanslational start site. While neither of the ATG start sites are surrounded by the strict Kozak sequence (GCC)GCCRCCATGG (Kozak, 1987), both start sites are surrounded by acceptable eucaryotic Kozak sequences as defined by Lackie and Dow (1999) (RNNMTGG). This data indicates that either of the start sites are likely to be functional, although the Kozak sequence associated with the start codon at position 979 fits the consensus sequence better, and may therefore be better functioning. On the other hand, because *T. vaginalis* genes are known for their unusually short 5' UTR's, the location of the 5'UTR may have provided enough information to identify the correct start codon. Primer extension analysis would have been useful in this regard.

Because primer extension was not successful, it is was not possible to determine the end of the transcript or the 5' untranslated region. 5' RACE PCR suggests that the transcript begins at least 2 nucleotides upstream of the CDF-2 clone (position 882). Most *T. vaginalis* genes identified to date have demonstrated a short 5' UTR which includes the

13 bp transcription initiation (INR) promotor consensus sequence (TCA$_{+1}$YWWYWMWMTW) typically located 6 to 20 nucleotides upstream from the translational start codon (Liston & Johnson, 1999; Liston & Johnson, 1998). The underlined portion is a highly conserved region surrounding the transcriptional start site; this TCA or TTCA motif has been reported to also function as a ribosome binding domain (Musatovova & Alderete, 1998). Following the format of most published *T. vaginalis* genes, it is expected that the transcript will end within 20 or so base pairs upstream from the ATG, and that the untranslated region will contain the INR consensus sequence. A search near both start codons for the *T. vaginalis* INR revealed TTCA motifs about 30 bp upstream from the ATG at position 979, and about 60 bp upstream of the ATG at position 877 (Figure 12, panel B). However, the 10 bp downstream from either motif did not conform to the published consensus sequence. It is possible that these sequences are nevertheless INR promotors since other genes have not conformed well to the published consensus sequence (Bagnara *et al.*, 1996). While TATA-like boxes are not usual in *T. vaginalis*; some *T. vaginalis* genes have been identified which have A/T rich regions, described as TATA-like, at the -30 position, which enhance promotor activity (Liston & Johnson, 1999). The A/T rich sequences starting near positions 815 and 856 may have such a function (Figure 12, Panel B).

The ORF extends from position 877 to a stop codon at position 7330. Following this stop codon, there exist many stop codons in all three frames. Through 3' RACE, the polyadenylation site was found at position 7364 (Figure 12, Panel C). No polyadenylation signal, AAUAAA, (Proudfoot & Brownlee, 1976) was identified in the 3'UTR, however, a destabilising element similar to those described by Musatovovoa and Alderete (1998)

was identified at the 3' end of the 3'UTR. These destabilising elements have the sequence UUAUUUAUU and are believed to cause the mRNA to have a short half-life by causing rapid de-adenylation (Zubiaga *et al.*, 1995). It is unclear whether this destabilising element plays any role in the function of this transcript.

Nucleotide and codon usage patterns were analysed and compared with the observations of Meade *et al.* (1997) who analysed the gDNA and cDNA usage patterns for 29 *T. vaginalis* protein coding sequences. Like other *T. vaginalis* genes, there was a very strong A/T bias in the 3' UTR (90%). There was a very strong A/T bias observed in the regions upstream of position 877 (77%), but not in the region upstream of position 979 (50%). However, the A/T content of the coding region was found to be much higher (66%) than the average 51% previously reported. Wang & Wang (1985) reported that *T. vaginalis* has an A/T rich genome. Codon usage in this gene, as reported Meade *et al.* (1997) for *T. vaginalis* sequences, was biased toward purine bases in the first position (57%) and A/T bases in the second position (62%). This gene, unlike the report by Meade *et al.* (1997), was not biased toward pyrimidine bases in the third position. This was mainly due to the high A/T content of the sequence. Meade *et al.* (1997) reported four amino acids in particular (Gly, Leu, Val, and Ser) to show a very strong bias towards pyrimidines in the third position. The sequence characterised here, however, showed a strong bias toward A/T in the third position (ranging from 77-92%) for these amino acids; the A was frequently presented instead of the C. Of amino acids utilising only pyrimidine bases in the third position, the bias for C>T was observed only for Tyr, and Phe, not for Asn, His, or Cys. Similar to the report of Meade *et al.* (1997), there was a strong bias (74%) for T>C for Asp. For amino acids containing third position purine

bases only, A was favoured (79-82%); Meade *et al.* (1997) had reported a G bias for Gln and Lys. Meade *et al.* (1997) also observed a strong bias toward TAA stop codons—the stop codon for this gene is TAA. Thus, codon usage was similar to that described by Meade *et al.* (1997), but skewed towards codons incorporating T and A, due to the A/T richness of the gene.

When used in a non-redundant database search (Altschul *et al.*, 1990) (Genbank, EMBL, DDBJ, and PDB), the full 8.25 kb sequence showed no homology to other nucleic acids. However, when the translated ORF was used in a non-redundant database search (Altschul *et al.*, 1997) (Genbank coding sequence translations, PDB, SwissProt, Spupdate and PIR), several peptide sequences were found to have significant probability scores. The most significant peptides included *Caenorhabditis elegans* UNC-89 (Benian *et al.*, 1996) (probability $e^{-17}$), *Saccharomyces cerevisiae* nuclear pore protein Nsp1p (Galibert *et al.*, 1996) (probability $e^{-17}$) and nucleoskeletal-like protein (probability $e^{-16}$) (Hurt, 1988), and human zonadhesion (Glockner *et al.*, 1998) (probability $e^{-16}$). While the similarity scores seemed high at first, a more detailed look at the alignments revealed that the matches resulted from repeated sequences. A BLASTp search of the non-redundant databases with the EEKPKL repeat revealed no homology with any known peptide sequences. The matches were therefore dismissed as not significant.

The predicted protein contains 2151 amino acid residues, and has a calculated molecular weight of 232 kDa (Figure 13). The hexapeptide sequence, EEKPKL, is repeated 42 times in the polypeptide (Figure 13). In addition to the 37 EEKPKL repeats in the CDF-2 clone previously described (Meysick, 1996), one other was found within the CDF-2 sequence, and an additional 4 were found in the sequence downstream.

**Figure 13: Amino Acid Sequence of the Translated ORF**

The polypeptide consists of 2151 residues, producing a molecule of 232 kDa. There are 42 EEKPKL repeats (double underlined) which are found in two major motifs EEKPKL$^{G}_{S}$LNL$^{G}_{P}$KPSPS or EEKPKLQLGGIKL (single underline). The polypeptide also has two tandem repeats of 103 amino acids between residues 440 and 623 (yellow and green highlight respectively) and a similar truncated repeat of 84 amino acids at residues 1022 to 1105 (green highlight). Further, there are 3 tandem repeats of 158-162 amino acids located between residues 1458 and 2114 (blue highlights). The putative mitochondrial pre-sequence FRQ/KQ is located at position 14-18 (bold/boxed). The inside-outside transmembrane helix is located at position 1642-1660 (boxed). Areas of sequence variability between isolates DG and 202 are indicated in red letters. This sequence is published under Genbank accession number AF257323.

```
   1 MDSQAAPKRR IHLFRQKQNI DWGFNPDPFK IPPDMADCVV TRNPVLSVLQ NQQRQITNEP
  61 YIPKQETPAP LNASVSQNQQ LTINNSASNI AVQQKSEQPA GQDNKLSTIG QFPSIENKDA
 121 NNQTAPTISL SLKNPLKPAD NLLAKSGQEN ENKAESNTDS KPKLTAMNNQ FSLKPSTPNS
 181 EAKPAIGGNG LGLGMKLSSS KPLSNISEAA QDKPATGGLS LKLPSSKPLS NIQAASEEKS
 241 QPSTGLKLGM KLPSNQASEE KPKLGLNLPK SPSNSTEEKP KLGLNLGKSP STEEKPKLQL
 301 GGIKIGQTPS NSTEEKPKLG LNLPKSPSTE EKPKLSLNLG KSPSTEEKPK LSLNLGKSPS
 361 NQSTEEKPKL GLNLPKSPSS QTSEEKPKLS LNLGKSPSNQ STEEKPKLQL GGIKLGQTQS
 421 NQKSEEKPKL SLNLGKSSTE EKPKLSLNLG KSPSNQSTEE KPKLGFNLPK APSNQTEEKP
 481 KLGTGGISLN LGNKPQSEEK PKLSLGGIKL AQSPSNSN[redacted]
 541 [redacted]
 601 [redacted]QLGGIKLN LGNKPQTETQ TEEKPKLQLG GIKLGQSPSN
 661 STEEKPKLQL GGIKLNLGSK PQTEEKPKLQ LGGIKLGTGG ISLNLGNKPQ SEEKPKLQLG
 721 GIKLGNSQPN QPLEKPKSGI NLNLGKSQPS SEEKPKLGLN LGKSPSNSTE EKPKLGTGGI
 781 SLNLGNKPQT EEKPKLSLNL PKSPSNQNQS TEEKPKLQLG GLKLNLGNKP QTETQTEEKP
 841 KLQLGGIKLN LGSKSQTEEK PKFQLGGIKL GQSPSNSTEE KPKLSLGGIK LAQSPSNEEK
 901 PKLSPNLPKS PSNQQTEEKP KLQLGGLKLN LGSKSQTEEK PKLSEGGIKL GNVSSSQTSD
 961 EKPKLGLGGI SLNFGNKQQT EEKPKLSSSQ NVEKPTLGLG GITLGQQTSE ENKPKLGLNL
1021 P[redacted]
1081 [redacted]HLQSD EKPKFSLGGM KLGKVPSNES QEPPKFTLNI
1141 PKLPSNQAEQ NDGKPKLGLG GISLNLSQKK EENEEKPKLG IAPKPIQSNQ NKIEEKPKVT
1201 NSVTKTEEKA ISAPGVKLDL KLPKKGFGLM RQPSKSLEVS NDSEQQNLSS FGQIHVTLPE
1261 KKEENQPLSS FGQIHVTLPE KTQKKEENQP LSSFGQIHVT LPQKGQNEKV TEQKVTDEKS
1321 LSSFGQIHVT IPEKLANKQE EKPLSTFEQI HVSIPEKTQK QEEKISSFGQ IHVSIPEKLQ
1381 KQEEKTSSFG QIHVTIPEKL QKQEGKSVID KLPEKTQELI KNMPSEINVK IPEQVHSQQS
1441 NPLVPLGVSL NDVAKSQ[redacted]
1501 [redacted]
1561 [redacted]
1621 LNLNLNNKTP LSNAPKPIAL SLSNDGKKPS LAGLNISKGS SLASVSINAL DKSGDDVPKL
1681 ALSIDKLRKE PPPPPTPPAV TKPALATEKE IRESSIADVT PLAMITKINS AKKPDFSALK
1741 PNLGNKQQSS SSNEQ[redacted]
1801 [redacted]
1861 [redacted]SNK
1921 GQVQGALTIK PIPLSKGDKP KADFQSLKLG KSLS[redacted]
1981 [redacted]
2041 [redacted]
2101 [redacted]QLNLNG NGLPKPGLPK PLPFNKPGIQ RLNFSAINLK K
```

The polypeptide also has two tandem repeats of 103 amino acids between residues 440 and 623 and a similar but truncated repeat of 84 amino acids at residues 1022 to 1105. Further, there are 3 tandem repeats of 158-162 amino acids located between residues 1458 and 2114 (Figure 13).

No signal peptides, which would direct the protein to the endoplasmic reticulum, could be identified using either the SignalP (Neilson *et al.*, 1887) program or the methods of McGeoch (1985) and VonHeijne (1986) using the PSORT (Nakai & Kanehisa, 1992) program. PSORT, however, did identify a putative mitochondrial pre-sequence by the method of Gavel and VonHeijne (1990) at position 14-18 (FRQ/KQ) (Figure 13), and 9 vacuolar targeting sequences (KLP/K/N/I) from positions 1221 to 2033. While T. vaginalis is a mitochondriate, the presence of mitochondrial pre-sequences has also been noted by Lahti *et al.*, (1992) and Bradley *et al.* (1997) in several hydrogenosomal enzyme sequences. These sequences are believed to direct the protein to the hydrogensome. It is not known whether this element has any function in the CDF-2 ORF. PROSITE (Bairoch *et al.*, 1997) identified 37 N-glycosylation sites and 77 phosphorylation sites (1 cAMP/cGMP dependent protein kinase phosphorylation site, 26 protein kinase C phosphorylation sites, 50 casein kinase II phosphorylation sites). While TMPred (Hofmann & Stoffel, 1993) identified one inside to outside transmembrane helix at position 1642-1660 (Figure 13), the protein was not predicted to be a transmembrane protein. SOUSI (Hirokawa *et al.*, 1998) calculated an average hydrophobicity of –0.976 with no hydrophobic regions, and thus predicted the protein to be soluble. These results suggest that the protein is post-translationally processed, that it is transported through the Golgi apparatus, and that it is a soluble protein rather than a surface protein. Whether the

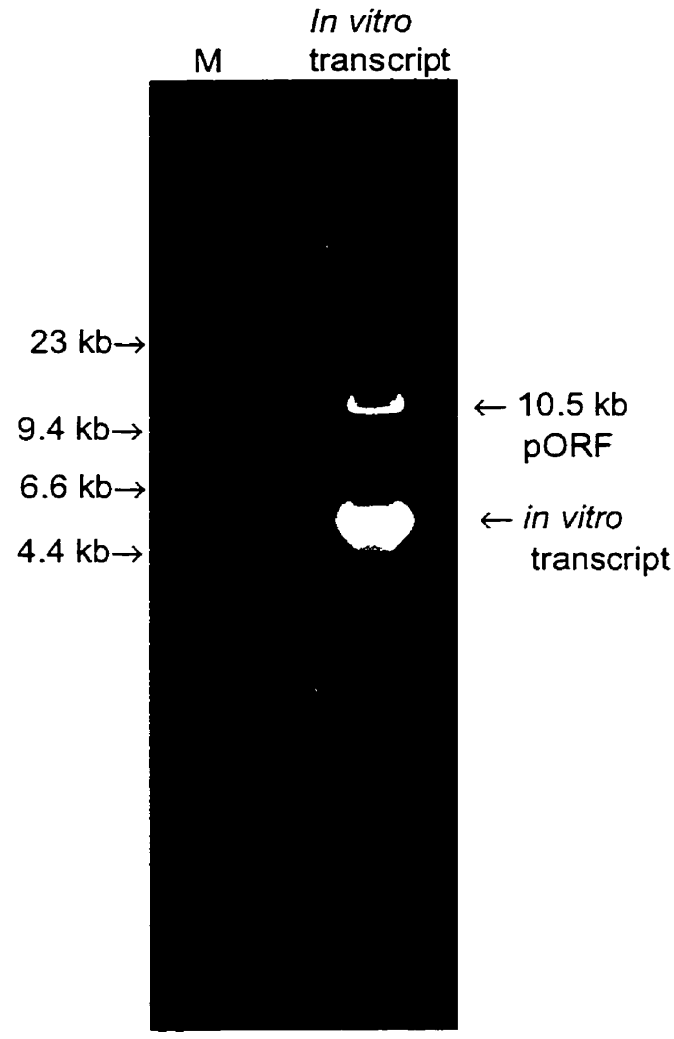
protein is located intracellularly in the hydrogenosome or vacuole, or whether it is secreted extracellularly is unknown. This analysis does not predict the function of the protein. While the closest homologues were structural in nature, the mitochondrial pre-sequence suggests that it may function within the hydrogenosome; on the other hand, the vacuolar targeting sequences may suggest the protein may function as lytic enzyme carried in the vacuoles.

***In vitro Transcription/Translation***

To determine whether the identified sequence could produce a polypeptide that could be recognised by rabbit antiserum prepared against a purified preparation of CDF, the ORF was transcribed and translated *in vitro*. The full-length ORF, which was cloned into pCR2.1 (pORF), was transcribed *in vitro* using the T7 RNA polymerase initiation site on the vector (Figure 14). The RNA was then translated *in vitro* using rabbit reticulocyte lysate in the presence of $^{35}S$ methionine. The protein species were separated by SDS-PAGE, transferred by electroblotting onto a nitrocellulose membrane, and immunoblotted with rabbit anti-serum raised against a purified preparation of CDF. Western blot identified an immunologically reactive protein at approximately 200 kDa, but similar bands appeared in the negative controls (data not shown). Despite attempts to increase the amount of product by increasing both the *in vitro* transcript in the *in vitro* translation reaction and the protein load in the gel, the results were ambiguous. Inability to clearly detect the *in vitro* translated products by immunoblot with the anti-CDF serum was likely related to inadequate amounts of protein in the gels required for such detection. For this reason, gels were overloaded purposefully, in order to increase the likelihood of detection by both Western Blot and autoradiograph.

**Figure 14: *In vitro* transcription of CDF-ORF**

CDF-ORF, which was cloned into pCR2.1 TOPO, was transcribed *in vitro* from the linearised plasmid using T7 RNA polymerase (Promega). Prior to transcription, the 10.5 kb recombinant plasmid was linearised with *Bam*HI which cut the plasmid in the multiple cloning region of the vector, 59 bp downstream of the end of the insert. The *in vitro* products are visualised on a 0.8% non-denaturing agarose gel. Both the 10.5 kb linearised plasmid and the approximately 6.5 kb *in vitro* transcript are visible from 2.5 µL of the *in vitro* transcription mix which was electrophoresed on a 0.7% agarose gel. Molecular weights are indicated by the λ *Hin*dIII marker (M) (Gibco/BRL).

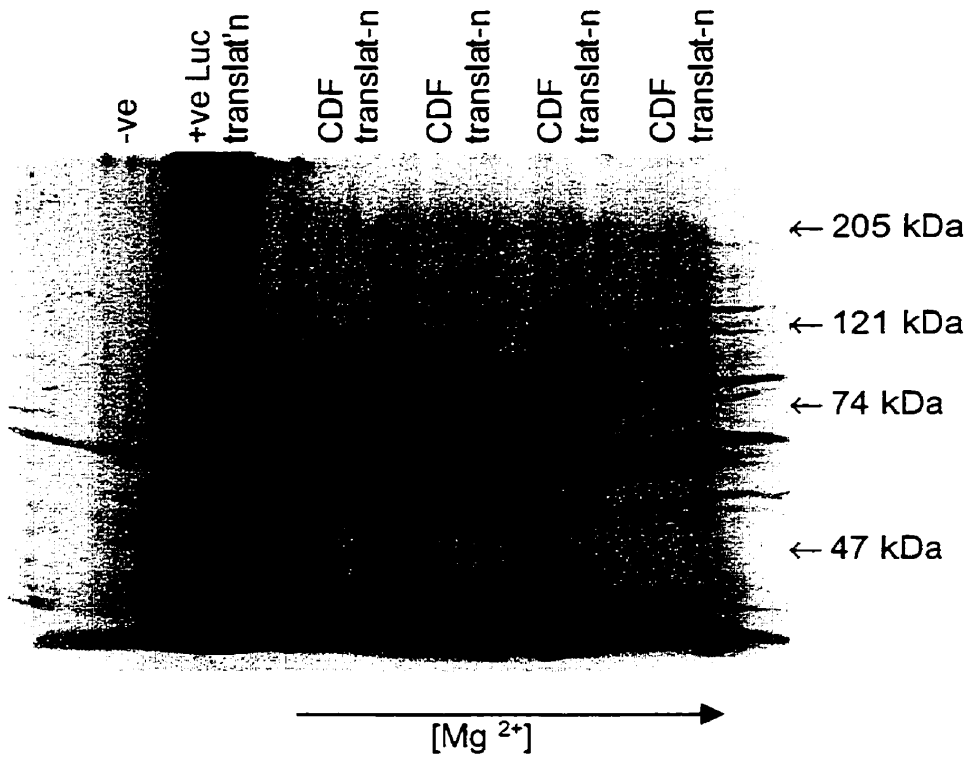

M
In vitro transcript
23 kb→
9.4 kb→
6.6 kb→
4.4 kb→
← 10.5 kb pORF
← in vitro transcript

To increase the specificity and sensitivity, immunoprecipitation with rabbit anti-serum was performed after translation. Null *in vitro* translation mix immunoprecipitated with immune rabbit anti-serum, and test *in vitro* translation mix immunoprecipitated with pre-immune rabbit serum served as negative controls. The immunoprecipitated products were separated on an SDS-PAGE gel, which were dried and exposed to X-ray film. While several radiolabelled *in vitro* translation products (>200, 165, 150, 130, 118, 80, 47 kDa) were identified (Figure 15, Panel A), only two polypeptides of approximately 205 and 220 kDa reacted strongly to rabbit antiserum raised against CDF (Figure 15, Panel B). It is not possible to determine at this point whether there are two true products, representing polypeptides initiated at different sites, or whether the smaller product is due to incomplete transcription, or whether they are different proteins. The first is unlikely, however, since the two possible initiation codons are 100 bp apart, which would translate to a difference of 3.4 kDa, much smaller than the 15 kDa observed. Why the many species of *in vitro* translated products present in 15A (which supposedly represent the CDF-2 ORF *in vitro* transcript) could not be immunoprecipitated remains unexplained. However, both pre-mature termination and initiation of translation at multiple start codons may explain why so many products are observed in the *in vitro* translation mix before immunoprecipitation.

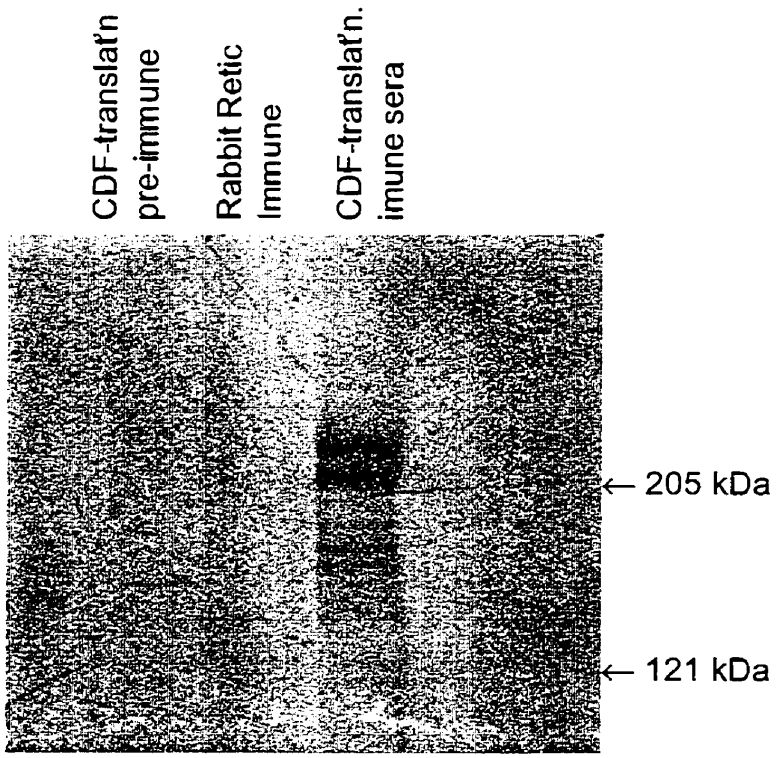
**Figure 15: *In vitro* translated products and Immunoprecipitation**

The CDF *in vitro* transcript was translated *in vitro* in the presence of $^{35}$S-methionine in a rabbit reticulocyte lysate system (Promega) with varying magnesium concentrations to optimise conditions. The proteins in 30 µL (2/3) of the *in vitro* translation mixes were then separated on a 7.5% SDS-PAGE gel and exposed to film. The gel was overloaded intentionally in order to increase the likelihood of detecting the *in vitro* translated products. Several *in vitro* translated products were identified, including bands at 200, 165, 150, 130, 118, 80, 47 kDa (Panel A). No radiolabelled proteins were identified in the negative control (-ve) (*in vitro* translation mix with no in transcript added). A 61 kDa Luciferase translation product, along with several smaller products was found in the positive control (LUC). When the CDF *in vitro* translation mix was immunoprecipitated with CDF-immune rabbit sera, separated on a 5% SDS-PAGE gel, and again exposed to film, only two products at 220 kDa and 205 kDa could be seen (Panel B). Based on the size of the ORF, the *in vitro* translation product is expected to be 230 kDa. No products were immunoprecipitated by pre-immune rabbit serum from the CDF translation mix nor were any products immunoprecipitated by CDF-immune rabbit serum from a null *in vitro* translation mix. The positions of the High Range Protein Standards (205, 121, 74, 47 kDa) (BIO-RAD) are shown.

A


-ve
+ve Luc translat'n
CDF translat-n
CDF translat-n
CDF translat-n
CDF translat-n
← 205 kDa
← 121 kDa
← 74 kDa
← 47 kDa
[Mg 2+]

B


CDF-translat'n pre-immune
Rabbit Retic Immune
CDF-translat'n. imune sera
← 205 kDa
← 121 kDa

# CHAPTER 4: DISCUSSION AND CONCLUSIONS

## DISCUSSION

The goal of this project was to determine whether the CDF-2 cDNA clone, previously identified by immunoscreen of a cDNA expression library with rabbit anti-serum raised against CDF, was a candidate gene for CDF. In order to answer this question, the sequences flanking CDF-2 were identified and characterised. This was to have been done by screening an EMBL3 genomic DNA library. In actual fact, several approaches, including screening a genomic DNA library, performing inverse PCR, and finally walking along the genome from CDF-2 using RAGE PCR, were attempted before the sequence could be obtained. Once the complete open reading frame, including the translational stop codon, polyadenylation site, and potential start sites were identified, the next step was to determine whether the characterised gene could encode a polypeptide with properties consistent with CDF. *In vitro* transcription and translation experiments were performed to help answer this question.

The identification of sequences flanking CDF-2 was hampered by a variety of problems. Undoubtedly, the repeated elements present in CDF-2 and its flanking regions turned out to be the most significant problem.

The repetitive sequences were likely responsible for hybridisation of the CDF-2 probe with "negative" clones during the initial screening of the genomic DNA library. Other explanations that could account for the inconsistency of the screenings using the CDF-2U probe include inadequate wash stringency and autoradiographic overexposure (to increase sensitivity). Further, while many more clones were screened than were

statistically required (Seed *et al.*, 1982) (the equivalent of over 28 genomes were screened), it is possible that the library did not adequately represent the *T. vaginalis* genomic DNA sequences. The original library was constructed using 17-23 kb fragments of an *Mbo*I partial digest of high molecular weight *T. vaginalis* genomic DNA. The characterised sequence contains, as expected, many *Mbo*I sites. The possibility remains that the DNA may have been overdigested, and that fragments containing CDF-2 sequences were excluded from the library in fractions containing smaller DNA fragments.

The repetitive nature of the *T. vaginalis* genome was also likely responsible for the failure of the inverse PCR experiments. Although the primers for inverse PCR were designed according to established guidelines (Loffert *et al.*, 1997), experience of later PCR techniques would prove that non-specific primer binding, as well as the existence of more than one potential primer binding site due to large repeated elements, were serious complicating factors for many PCR experiments involving *T. vaginalis* genomic DNA. The outcome of the inverse PCR experiments may have been improved by redesigning the primers to have higher melting temperatures, increasing the annealing temperature, including initial touchdown cycling, and performing secondary PCR reactions with nested primers.

Despite the difficulties arising in the first two approaches for obtaining the sequence that flanks CDF-2, this task was accomplished by walking along the genome through RAGE PCR. This technique was challenging, however. Primer design was extremely complicated due to the repetitive sequences, and the high A/T content of the sequence made design of primers with $T_m$>70°C difficult. The existence of more than

one true primer-binding site for primer 73C within the long repeats made results difficult to interpret. While the DNA in the original libraries may have contained single or double stranded breaks, the sequence obtained by RAGE PCR was shown to be contiguous and reliable. Four sets of data support this statement. First, the restriction map of the contig is consistent with the restriction map deduced by Southern blotting. Second, the ORF sequence was amplified end-to-end, producing an amplicon of expected size, which was then cloned. Third, restriction digests of the recombinant ORF plasmid were consistent with the restriction sites identified in the contig sequence. Finally, the sequence of the ORF was later confirmed to be as expected by spot sequencing of pORF.

Ironically, even though freshly prepared genomic DNA seems to be most appropriate for this technique, PCR performed on the freshly prepared libraries created with the blunt-cutting enzymes provided in the kit (*Eco*RV, *Pvu*II, and *Stu*I) did not yield any suitable amplicons for cloning. For this technique to perform optimally, tailoring the choice of restriction enzymes to the characteristics of the genome (and region of genome) of interest may work best. For example, *T. vaginalis* DNA is particularly A/T rich. The libraries that worked really well with the *T. vaginalis* DNA were created with *Dra*I, which has an A/T rich recognition site, and *Eco*RI, for which restriction cleavage sites had been well mapped out by Southern blotting. In spite of Clontech's recommendations to use only restriction enzymes producing blunt ends, the *Eco*RI library seemed to work rather well. Another restriction enzyme that would have worked well is *Rsa*I based on the data from the Southern blotting. Other good choices would have been enzymes that have A/T rich recognition sequences.

Two possible translational start codons were identified, as well as the translational

stop codon, identifying a full-length ORF. At this point, primer extension analysis was performed in hopes of establishing the location of the end of the transcript. The primer extension analysis was a key experiment in this regard. Based on the typical structure of the *T. vaginalis* 5' UTR (Liston & Johnson, 1999; Liston & Johnson, 1998), much information about the location of the INR, and the probable location of the true start codon could have been obtained simply by the identification of the 5' UTR.

Sequence analysis alone could not determine which translational start codon was the true start codon. The ATG at position 877 is the most 5' translational start codon that would maintain an open reading frame. This, coupled with the presence of a stop codon at position 802 in the +1 reading frame (and others in the other two reading frames) suggests that the coding region cannot extend past this point. While this analysis defines the 5' limit of the coding region, it does not define which of the two possible translational start codons is the actual one. Because the *T. vaginalis* 5' UTR's are so short, and the start codon is expected to be within 20 bp of the end of the transcript, 5' RACE data should help in this regard, since it should identify the end of the transcript. The 5' RACE PCR products, however, extend only to position 884, just 2 bp upstream of the 5' end of CDF-2, and just 7 bases downstream of the start codon at 877. When preparing the Marathon RACE cDNA libraries, 0-20 bp are removed from the cDNA ends during the blunt-end step with T4 DNA polymerase (Clontech Laboratories, 1999). This suggests a model where the ATG at position 877 is the true translational start codon, and the 5' UTR and start codon were nibbled by T4 DNA polymerase. Inability to obtain the 5'UTR and start codon via 5'RACE PCR has been experienced by other investigators of *T. vaginalis* (J. Alderete, personal communication). Alternatively, secondary structure exists at this

location of the RNA molecule (possibly preventing cDNA extension) and may also explain why the CDF-2 cDNA molecules consistently end at that location. Because the ATG at position 877 is more proximal to the end of the transcript (as determined by 5'RACE), and because of the typical structure of *T. vaginalis* 5'UTR's usually place the end of the transcript within 20 bases of the start codon, the end of the transcript is not likely to extend 100 bases upstream of the start site at position 979.

While the 5' RACE PCR suggests that the ATG at position 877 may be the true start codon, the more proximal location of potential INR promotors in the sequence in relation to ATG at position 979 point more toward this ATG as the true start codon. Furthermore, the more favourable Kozak consensus sequence also points to the start codon at position 979 as the true translational start site. *In vitro* transcription/translation experiments, where the possible translational start sites are mutagenised one by one may help determine which ATG is the true start codon.

Future work should include experiments that prove the location and functionality of the potential INR promotor. Clearly, the primer extension analysis will need to be revisited in order to obtain more data about the 5' UTR. This will include redesigning the primers for the experiment so they are closer to the ATG sites and so that they have lower $T_m$, so that lower annealing temperatures can be used. Multiple extensions, performed in a thermocycler (Hu *et al.*, 1998), could improve the number of copies of extension product, and could therefore increase the sensitivity of the technique, while increasing the melting temperature will ensure complete melting of the RNA. Another approach might be to use reverse transcriptase (RT) PCR using different sets of primer pairs (one which binds within 200-300 bp of the end of the cDNA and different primers which bind

increasingly further upstream) to determine how far upstream the sequence is transcribed. A third approach might be to perform inverse RT-PCR on polyA$^+$ RNA that has been self-ligated into a circle, providing that the transcript is not 5' capped. While some authors have reported the absence of a 5' cap, Liston and Johnson (1999) have reported that *T. vaginalis* transcripts are capped. Finally, the S1 nuclease mapping analysis used by Johnson *et al.* (1990) has been used successfully to obtain the 5' end of the transcript. Whatever the method, care will have to be taken to ensure that the RNA has first been completely melted to ensure that any secondary structure does not prevent extension. Once the end of the transcript has been located, a more definite statement can be made about the location of the INR promotor and the true start site. Promotor function experiments for *T. vaginalis* are described (Liston & Johnson, 1999) and involve ligating the promotor (and mutated promotor) onto reporter constructs, transfecting the plasmids into *T. vaginalis*, and studying the promotor activity of the normal vs. mutant reporter constructs.

The 3' RACE PCR experiment was successful in determining the position of the polyA$^+$ tail, 32 bp downstream from the translational stop codon. These data are likely reliable, since 3 out of 5 clones were identical. Clones 3-6 and 3-7 (discussed in the Results), however, are not expected and are difficult to explain. Typically, sequence that is derived from non-contiguous regions in genomic DNA but is found to be contiguous in cDNA is usually explained by splicing events during transcription. While *T. vaginalis* was recently discovered to have a gene encoding spliceosomal component (pre-mRNA processing 8 PRP8) (Fast & Doolittle, 1999), all genes characterised from its genome (and all other parabasilid genomes) to date have been devoid of introns and trans-spliced

5' mini-exons (Liston & Johnson, 1998). This does not exclude the possibility of splicing events, however, as it is unknown whether the lack of introns and mini-exons reflects an absolute absence of cis and trans splicing or whether it indicates that splicing is rare (Liston & Johnson, 1998). Other reasons usually offered to explain non-contiguous sequence include: that the gene is part of a multi-gene family, the open reading frame is polycistronic, or that there exists several polyadenylation sites. The first is more likely than the second, however, as *T. vaginalis* transcripts studied to date are known to be monocistronic, and members of multi-gene families do not usually reside within the same large (10-15 kb) genomic clone (Liston & Johnson, 1998). The third is unlikely, as the clones did not contain polyA$^+$ tails. The positions at which the clones end their homology coincide with the ends or near the ends of long repeats. It is possible that the clones represent multiple alleles, as described for the variant-specific surface protein of *Giardia lamblia* (Yang *et al.*, 1994). Most likely, however, is that clones 3-6 and 3-7 are products of artifact. RACE PCR is known to have many artifacts, particularly because of the many manipulations required in transforming polyA$^+$ RNA into cDNA. The long length of the described transcript renders it even more susceptible to artifact.

Several pieces of evidence suggest that the ORF associated with CDF-2 is a likely candidate gene for cell detaching factor. As discussed above, the sequence is contiguous. The size of the open reading frame is consistent with the size of cell detaching factor, as is the size of the transcript identified in the Northern Blot (all *T. vaginalis* genes characterised to date lack introns, thus the gene and the transcript are expected to be the same). The molecular weights of the largest *in vitro* translation products are consistent with the molecular weight of CDF. Finally, the *in vitro* translated products share similar

epitopes to CDF—that is they can be immunoprecipitated by anti-serum raised against CDF.

Cell Detaching Factor was approximated to be about ≥200 kDa (Garber *et al.*, 1989). Based on the molecular weight of CDF, and depending upon post-transcriptional modifications, processing, and the degree of glycosylation, the transcript was predicted to be up to 5.4 kb in size. The Northern blot identified a 6.5-7 kb transcript, which is consistent with the sequence analysis of the sequenced 6.5 kb open reading frame. The molecular weight of the polypeptide sequence encoded by the open reading frame was calculated to be 232 kDa, and the molecular weights of the immunoprecipitated *in vitro* translation products were estimated to be approximately 205 and 220 kDa. The molecular weights of both the encoded polypeptide sequence and the immunoprecipitated *in vitro* translated products are consistent with the approximated molecular weight of CDF. It is not known if the shorter of the two immunoprecipitated protein species represents a protein truncated by premature termination during *in vitro* translation. It is unlikely that the two proteins were derived from different transcriptional initiation sites (at least for the two start codons described because the size difference between them is too big). However, both pre-mature termination and initiation of translation at multiple start codons may explain why so many products are observed in the *in vitro* translation mix before immunoprecipitation. Finally, in consideration of concerns previously raised about the specificity of the rabbit anti-serum raised against a purified CDF preparation (Meysick, 1996), it may be possible that the two immunoprecipitated proteins represent different species. Why the many species of *in vitro* translated products present in 15A (which supposedly represent the CDF-2 ORF *in vitro* transcript) could not be

immunoprecipitated remains unexplained.

Of interest is an observation that the sequence obtained by RAGE PCR from *T. vaginalis* DG had three small areas of sequence which differed slightly from sequence obtained by RAGE PCR and 3'RACE PCR of *T. vaginalis* 202. These areas located near the ends of the long repeats consist of: an addition of 9 nucleotides (CT AAA GAA G) at position 6331 resulting in an addition of alanine-lysine-glutamatic acid, a substitution of 9 nucleotides (TT AAT AAA G) for 24 nucleotides (GCG AAA ACA GAA CAA CCA AAA ACA) at position 6499, resulting in a substitution of Val-Asn-Lys for Ala-Lys-Thr-Glu-Gln-Pro-Lys-Thr, and an addition of 12 nucleotides (C CAA AAC CAA AA) at position 6981 resulting in an addition of Gln-Asn-Gln-Asn (Figure 13). It is interesting that these additions and substitutions fall within multiples of 3, so that the reading frame remains intact. These additions do not change the length of the open reading frame; however, it is not known what impact they have on the structure and/or function of the protein. These additions and substitutions are located within the three large repeats.

A commonly seen phenomenon in protozoan proteins is gene size polymorphism between isolates due to a varied number of tandem repeats within a particular gene. This causes size polymorphism in the protein for which the gene codes. Once such example of this is the P270 protein described by (Alderete, 1999), while another example is the gamma-GCS gene from *Plasmodium falciparum* (Luersen *et al.*, 1999). Another possibility is that the two sequence variations represent multiple alleles within the genome. Multiple alleles of the variant-specific surface protein from *Gardia lamblia* have been shown to represent different copies of the gene, which vary in the number of tandem repeats (Yang *et al.*, 1994). Since the length of the open reading frame remains

intact, simple genetic variation between the two *T. vaginalis* strains remains most likely. Sequencing of more full-length clones from both genomic and cDNA from these isolates, careful comparison between genomic DNA and cDNA, as well as comparing sequence differences among several isolates will provide more information regarding these observations. Further, protein structure analysis will help determine whether the amino acid changes will affect the structure, and possibly function of the protein.

Sequence analysis has indicated a number of tandem EEKPKL repeats, as well as large tandem repeats (100 bp and 450 bp) within this gene. While tandem repeats are well known among protozoan genes, it is still uncertain as to the function of these repeats in this gene. These repeats are not likely to be mini/microsatellites, which can be very short (2-6 bp) and very A/T rich in protozoa (Bishop *et al.*, 1998; Field & Wills, 1996), because they are also transcribed. The most commonly described function of tandem repeats in protozoa is immunomodulation. Alderete (1999b) has described a P270 immunogenic protein of *T. vaginalis*. P270 has a 333 bp tandem repeating element which contains a DREGRD epitope. Size polymorphism in P270, due to varying numbers of the tandem repeated element, are thought to be involved in immune evasion (Alderete, 1999b). A similar example exists in the merozoite surface protein of *P. falciparum*. Variation in the number of tandem repeats of the MSP protein is thought to allow the parasite to avoid high avidity antibody binding (Ranford-Cartwright *et al.*, 1996). Another example of immunomodulation is the immunodominant epitope of the B13 antigen of *Trypanosoma cruzi* (Duranti *et al.*, 1999). This protein has a repeated motif containing the core hexapeptide AAAGDK, which is believed to allow the protein to switch from a random to a helical conformation at the antigen:antibody interface; this

transformation correlates directly with the protein's antigenicity. Short amino acid tandem repeats in *P. falciparum* are also believed cause ineffective T-cell-independent antibody response by the host (Verra & Hughes, 1999).

Other functions have also been described. Tandem repeats in *T. cruzi* Antigen 13 and Shed-Acute-Phase-Antigen stabilise the trans-sialidase catalytic activity in blood. This stabilisation effect of the tandem repeats was found to be less in the intracellular protein and more in the shed proteins. The repeats are also associated with increasing the persistence of chimeric glutathione-S-transferase (Buscaglia *et al.*, 1999). In *P. falciparum*, the hexapeptide repeats Ala-His-His-Ala-Ala-Asp in the histidine rich protein II function to bind haem (Pandey *et al.*, 1997). In *Leishmania mexicana*, large tandem repeats in the secreted acid phosphatases are believed to function as targets for glycosylation (Stierhof *et al.*, 1998). Several other functions were reviewed by Meysick (1996). Whatever the function may be of these EEKPKL and long tandem repeats in the putative CDF gene sequence will remain a mystery until analysis of protein structure and function sheds more light.

This project will not be complete until it has been shown that the CDF-2 ORF is the gene which codes for CDF. One strategy is to sequence the native protein, and compare the sequence of the native protein to the encoded CDF-2 protein. The more important question to answer, however, is: does the CDF-2 protein encode the activity observed in *T. vaginalis* culture supernatants. Until now, there has been a lack of models in which to express recombinant *T. vaginalis* proteins in order to study their function. To date, all recombinant *T. vaginalis* genes have been expressed as in *E. coli* systems. Such proteins have all been small metabolic enzymes. This system is unlikely to work for this

gene, since it is post-translationally processed (phosphorylated and glycosylated, and may require proteolytic cleavage), and it is very large. A more suitable system for expression of this protein might be yeast or baculovirus in insect cells. Recently, a model for transfection of *T. vaginalis* has been described (Delgadillo *et al.*, 1997), which has been used to study functionality of *T. vaginalis* promotors (Liston & Johnson, 1999). This model could also be used to produce recombinant protein in either a knock-out strain or perhaps in *Pentatrichomonas hominis*, which lacks CDF activity. Whatever the system, the CDF-2 ORF coding region will have to be inserted into an expression vector which is capable of handling such a large insert. Since most eucaryotic expression systems allow the protein to be tagged using histidine (rather than by GST-fusion), the recombinant protein can be recovered by means of a nickel column. The protein can then be tested for CDF-activity. Two experiments can be devised which will show whether the recombinant protein possesses the activity encoded by CDF: does the recombinant protein cause monolayer cell detachment and can the CDF rabbit anti-serum block this activity? Several authors have accomplished the task of comparing function of the the recombinant protein to the native protein simply by preparing cell lysates of transfected and non-transfected host cells (*E. coli*), and comparing activity through a specific biochemical assay. Such a strategy is not recommended in this situation, however, since cell detachment is not a phenomenon that can be measured through a defined biochemical assay. Several conditions are known to cause cell detachment, including toxicity of the culture. Using cell lysate, it would be impossible to determine whether cell detachement is caused by the recombinant protein or by something else in the lysate. For this reason, it would be better to isolate the recombinant protein for analysis.

While protein sequence analysis was able to determine that the CDF-2 ORF protein is likely to be soluble, and may be associated with the hydrogenosome or vacuoles, its location in the cell through processing and whether the protein is excreted extracellularly remains unknown. Ultimately, it will be relevant to determine the location of the protein within the cell (perhaps by means of intracellular staining) and determine what function or relevance it has to the organism. Whether the protein is a hydrogenosomal metabolic enzyme or a lytic enzyme carried in vacuoles, or whether it has some other function to *T. vaginalis* remains unknown. Extensive biochemical analysis will be required to answer these questions. To date, however, not enough is known about native CDF in regards to its location and function/relevance of the protein to the organism. It is hypothesised that the protein degrades the basement membrane of the cells in the vaginal tract in order to reach better nutrient supplies within the lower levels of the epithelium. A more fundamental understanding of the native protein with regards to its location and biochemical interactions within (and outside) the organism must be reached before it can be compared with the protein encoded by CDF-2 ORF. Eventually, it may be of interest to compare not only the function of the recombinant protein to the native protein, but also the structure, by means of spectroscopy and crystallography.

Undoubtedly, there were several pitfalls in the original approach used to clone the CDF gene, which can now be addressed using the new information obtained from this work. That two unrelated clones were obtained from the screening of the cDNA library with the rabbit anti-serum prepared against a purified preparation of CDF raises concerns about the specificity of the antiserum. Meysick (1996) showed that the antiserum did in

fact have multispecificities, which could lead to the identification of irrelevant clones. Further, she attributed her failure in identifying the true CDF gene to inadequate representation of the CDF mRNA in the cDNA library. In consideration of these two problems, several items need consideration if screening of a cDNA library is to be attempted a second time. In immunoblot assays, the assay is only as good as the antibody, and the antibody is only as good as the original protein used for its preparation. The antibody could be made more specific by first re-purifying the CDF protein to ensure a highly pure product, and then preparing a panel of monoclonal antibodies. Such antibodies are not a sensitive as polyclonal serum, but are certainly more specific.

The problem of adequate representation of the protein in the expression library is more difficult to assess. Because much manipulation occurs to transform the mRNA into cDNA, which is then cloned and expressed, loss of representation may occur at any step along the way. Poor representation of the mRNA can be attributed to low number of mRNA molecules in the RNA mix before the preparation of the cDNA, problems accurately reverse-transcribing the polyA$^+$ RNA into cDNA, problems packaging the cDNA into the expression vector, and problems expressing the recombinant protein. Several of these factors may have come into play when the original cDNA expression library was prepared. If this technique is to be used again, a new library should be prepared taking into account several changes. To increase the transcription of the CDF mRNA, the *T. vaginalis* cells should be grown in the presence of eucaryotic cells. Experimentation with different cDNA preparation kits should be performed to ensure that there is adequate representation of the transcript. Some pitfalls experienced during the course of this research have included false priming of the oligo d(T) primer at A rich

regions of the transcript and early termination of the cDNA molecule, possibly due to secondary structure. Finally, a different expression vector other than λgtll is recommended for two reasons: its packaging limit is 7 kb, which may limit its usefulness in packaging a transcript approaching that size, and it has difficulty expressing proteins with repeats (Promega Technical Support, personal communication).

Given that there are so many problems with the cDNA library screening, it might be necessary to abandon this approach in favour of others. One approach is to sequence the native CDF protein. If only some of the protein is sequenced, degenerate oligonucleotides can be prepared from this sequence and can be used either to probe a genomic DNA library or used as primers for RAGE PCR. If the entire protein is sequenced, a BLAST search can be performed, which may pull out not only an identity, but also a gene sequence.

## CONCLUSIONS

The first objective of this study, to determine and characterise the full-length gene surrounding CDF-2 was accomplished. The second objective, however, was not completed. To date, the data supports the hypothesis that the full-length CDF-2 clone could be a candidate for CDF, but does prove it. There is no hard evidence to either confirm or refute that CDF-2 encodes a protein that possesses the activity that is observed in *T. vaginalis* supernatants.

Several pieces of evidence suggest that the ORF associated with CDF-2 is a likely candidate gene for cell detaching factor. It is transcribed, and the size of the open reading frame is consistent with the approximate size of the transcript identified in the Northern Blot, as well as the CDF protein. The molecular weight of the *in vitro* translated product

is also consistent with the molecular weight of CDF. Finally, the *in vitro* translated products are antigenically similar to CDF—that is they can be immunoprecipitated with anti-CDF rabbit serum. It is not known at this time whether the two proteins (205 and 220 kDa) identified in the immunoprecipitation represent different proteins, or truncated versions of the same protein (produced as an artifact of *in vitro translation*), or whether they represent a variable length protein.

What remains to be determined is the exact location of the 5' UTR, and in doing so, the INR and the true translational start site will be identified. This can be accomplished through primer extension, RT-PCR, or inverse RT-PCR, or S1 nuclease assay. More study regarding the location of the true translational start codon is required. Mutagenesis experiments of the potential ATG's, followed by *in vitro* transcription/translation may help in this regard. Once a putative INR has been identified, its function can be verified by studying the effect of mutagenesis on its ability to direct transcription and translation either in HeLa cell extracts or in *T. vaginalis*. More study regarding the sequence variation that was found between the DG and 202 strains will be required. This can be accomplished by comparing sequence of full-length ORF and cDNA clones from several different strains of *T. vaginalis*. Finally, protein function studies are required in order to prove that CDF-2 encodes the CDF activity seen in *T. vaginalis* filtrates. Further, protein function studies will allow for more study of the repeated elements. Other strategies for determining whether the CDF-2 ORF encodes the CDF protein include comparing not only function, but sequence and structure of the two proteins. More fundamental information regarding the location and biochemical interactions of native CDF within and outside the organism is required before further

analsysis on the recombinant protein is to proceed. Ultimately, the goal is to study gene expression and regulation of CDF, and this can be accomplished using the recombinant protein in carefully controlled assays.

# REFERENCES

Abonyi, A. (1992). *In vivo* observations of the division of the oversized round forms of *Trichomonas vaginalis* in culture. *Angewandte Parasitologie* **33**, 65-68.

Abonyi, A. (1995). Examination of nonflagellate and flagellate round forms of *Trichomonas vaginalis* by transmission electron microscopy. *Applied Parasitology* **36**, 303-310.

Abraham, M.C., Desjardins, M., Filion, L.G. & Garber, G.E. (1996). Inducible immunity to *Trichomonas vaginalis* in a mouse model of vaginal infection. *Infection and Immunity* **64**, 3571-3575.

Ackers, J.P. (1990). Immunologic Aspects of Human Trichomoniasis. In *Trichomonads Parasitic in Humans.* (Honigberg, B.M., ed.), Springer Verlag, New York, pp. 36-52.

Ahmed-Jushuf, I., Murray, A.E. & McKeown, J. (1988). Managing trichomonal vaginitis refractory to conventional treatment with metronidazole. *Genitourinary Medicine* **64**, 25-29.

Alderete, J.F. (1987). *Trichomonas vaginalis* NYH286 phenotypic variation may be coordinated for a repertoire of trichomonad surface immunogens. *Infection & Immunity* **55**, 1957-1962.

Alderete, J.F. (1988). Alternating phenotypic expression of two classes of *Trichomonas vaginalis* surface markers. *Reviews of Infectious Diseases* **10 Suppl 2**, S408-S412

Alderete, J.F. (1999a). Iron modulates phenotypic variation and phosphorylation of P270 in double-stranded RNA virus-infected *Trichomonas vaginalis*. *Infection and Immunity* **67**, 4298-4302.

Alderete, J.F. (1999b). The *Trichomonas vaginalis* phenotypically varying P270 immunogen is highly conserved except for numbers of repeated elements. *Microbial Pathogenesis* **27**, 93-104.

Alderete, J.F., Arroyo, R., Dailey, D.C., Engbring, J., Khoshnan, M.A. & Lehker, M.W. (1992). Molecular analysis of *Trichomonas vaginalis* surface protein repertoires. *Molecular & Cell Biology of Human Diseases Series* **1**, 173-202.

Alderete, J.F., Arroyo, R. & Lehker, M.W. (1994). Identification of fibronectin as a receptor for bacterial cytoadherence. *Methods in Enzymology* **236**, 318-333.

Alderete, J.F., Demes, P., Gombosova, A., Valent, M., Yanoska, A., Fabusova, H., Garza, G.E. & Metcalfe, E.C. (1987). Phenotypes and protein-epitope phenotypic variation among fresh isolates of *Trichomonas vaginalis*. *Infection & Immunity* **55**, 1037-1041.

Alderete, J.F., Engbring, J., Lauriano, C.M. & O'Brien, J.L. (1998). Only two of the *Trichomonas vaginalis* triplet AP51 adhesins are regulated by iron. *Microbial Pathogenesis* **24**, 1-16.

Alderete, J.F. & Garza, G.E. (1984). Soluble *Trichomonas vaginalis* antigens in cell-free culture supernatants. *Molecular & Biochemical Parasitology* **13**, 147-158.

Alderete, J.F. & Garza, G.E. (1985). Specific nature of *Trichomonas vaginalis* parasitism of host cell surfaces. *Infection & Immunity* **50**, 701-708.

Alderete, J.F. & Garza, G.E. (1988). Identification and properties of *Trichomonas vaginalis* proteins involved in cytadherence. *Infection & Immunity* **56**, 28-33.

Alderete, J.F. & Kasmala, L. (1986a). Monoclonal antibody to a major glycoprotein immunogen mediates differential complement-independent lysis of *Trichomonas vaginalis*. *Infection & Immunity* **53**, 697-699.

Alderete, J.F., Kasmala, L., Metcalfe, E. & Garza, G.E. (1986b). Phenotypic variation and diversity among *Trichomonas vaginalis* isolates and correlation of phenotype with trichomonal virulence determinants. *Infection & Immunity* **53**, 285-293.

Alderete, J.F., Newton, E., Dennis, C., Engbring, J. & Neale, K.A. (1991a). Vaginal antibody of patients with trichomoniasis is to a prominent surface immunogen of *Trichomonas vaginalis*. *Genitourinary Medicine* **67**, 220-225.

Alderete, J.F., Newton, E., Dennis, C. & Neale, K.A. (1991b). The vagina of women infected with *Trichomonas vaginalis* has numerous proteinases and antibody to trichomonad proteinases. *Genitourinary Medicine* **67**, 469-474.

Alderete, J.F., O'Brien, J.L., Arroyo, R., Engbring, J.A., Musatovova, O., Lopez, O., Lauriano, C. & Nguyen, J. (1995). Cloning and molecular characterization of two genes encoding adhesion proteins involved in *Trichomonas vaginalis* cytoadherence. *Molecular Microbiology* **17**, 69-83.

Alderete, J.F. & Pearlman, E. (1984). Pathogenic *Trichomonas vaginalis* cytotoxicity to cell culture monolayers. *British Journal of Venereal Diseases* **60**, 99-105.

Alderete, J.F. & Provenzano, D. (1997). The vagina has reducing environment sufficient for activation of *Trichomonas vaginalis* cysteine proteinases. *Genitourinary Medicine* **73**, 291-296.

Alderete, J.F., Provenzano, D. & Lehker, W. (1995). Iron mediates *Trichomonas vaginalis* resistance to complement lysis. *Microbial Pathogenesis* **19**, 93-103.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3402

Arroyo, R. & Alderete, J.F. (1989). *Trichomonas vaginalis* surface proteinase activity is necessary for parasite adherence to epithelial cells. *Infection & Immunity* **57**, 2991-2997.

Arroyo, R., Engbring, J. & Alderete, J.F. (1992). Molecular basis of host epithelial cell recognition by *Trichomonas vaginalis*. *Molecular Microbiology* **6**, 853-862.

Arroyo, R., Gonzalez-Robles, A., Martinez-Palomo, A. & Alderete, J.F. (1993). Signalling of *Trichomonas vaginalis* for amoeboid transformation and adhesion synthesis follows cytoadherence. *Molecular Microbiology* **7**, 299-309.

Bagnara, A.S., Tucker, V.E., Minotto, L., Howes, E.R., Ko, G.A., Edwards, M.R. & Dawes, I.W. (1996). Molecular characterisation of adenosylhomocysteinase from *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology* **81**, 1-11.

Bairoch, A., Bucher, P., & Hofmann, K. (1997) The ProSite database, its status in 1997. *Nucleic Acids Research* **25**, 217-221.

Beal, C., Goldsmith, R., Kotby, M., Sherif, M., el-Tagi, A., Farid, A., ZakariaS. & Eapen, J. (1992). The plastic envelope method, a simplified technique for culture diagnosis of trichomoniasis [published erratum appears in J Clin Microbiol 1993. 31:174]. *Journal of Clinical Microbiology* **30**, 2265-2268.

Benchimol, M., Batista, C. & de Souza, W. (1990). Fibronectin- and laminin-mediated endocytic activity in the parasitic protozoa *Trichomonas vaginalis* and *Tritrichomonas foetus*. *Journal of Submicroscopic Cytology & Pathology* **22**, 39-45.

Benchimol, M. & de Souza, W. (1995). Carbohydrate involvement in the association of a prokaryotic cell with *Trichomonas vaginalis* and *Tritrichomonas foetus*. *Parasitology Research* **81**, 459-464.

Benian, G.M., Tinley, T.L., Tang, X. & Borodovsky, M. (1996). The *Caenorhabditis elegans* gene unc-89, required fpr muscle M-line assembly, encodes a giant modular protein composed of Ig and signal transduction domains. *Journal of Cell Biology* **132**, 835-848.

Bhatt, R., Abraham, M.C., Petrin, D. & Garber, G.E. (1996). New concepts in the diagnosis and pathogenesis of *Trichomonas vaginalis*. *Canadian Journal of Infectious Disease* **7**, 321-325.

Biagini, G.A., Finlay, B.J. & Lloyd, D. (1997). Evolution of the hydrogenosome. *FEMS Microbiology Letters* **155**, 133-140.

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

Bishop, R., Morzaria, S. & Gobright, E. (1998). Linkage of two distinct AT-rich minisatellites at multiple loci in the genome of *Theileria parva*. *Gene* **216**, 245-254.

Bonilha, V.L., Ciavaglia, M., de Souza, W. & Costa e Silva Filho, F. (1995). The involvement of terminal carbohydrates of the mammalian cell surface in the cytoadhesion of trichomonads. *Parasitology Research* **81**, 121-126.

Bradley, P.J., Lahti, C.J., Plumper, E. & Johnson, P.J. (1997). Targeting and translocation of the proteins into the hydrogenosone of the protist Trichomonas: similarities with mitochondrial protein import. *EMBO Journal* **16**, 3484-3493.

Briselden, A.M. & Hillier, S.L. (1994). Evaluation of affirm VP Microbial Identification Test for *Gardnerella vaginalis* and *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **32**, 148-152.

British Co-operative Clinical Group. (1992). An investigation, by questionnaire, of cases of recalcitrant vaginal trichomoniasis seen in genitourinary medicine clinics in the United Kingdom. *International Journal of STD & AIDS* **3**, 24-27.

Brown, J.R. & Doolittle, W.F. (1995). Root of the universal tree of life based on ancient aminoacyl- tRNA synthetase gene duplications. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 2441-2445.

Bui, E.T. & Johnson, P.J. (1996). Identification and characterization of [Fe]-hydrogenases in the hydrogenosome of *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology* **76**, 305-310.

Buscaglia, C.A., Alfonso, J., Campetella, O. & Frasch, A.C. (1999). Tandem amino acid repeats from *Trypanosoma cruzi* shed antigens increase the half-life of proteins in blood. *Blood* **93**, 2025-2032.

Cameron, D.W. & Padian, N.S. (1990). Sexual transmission of HIV and the epidemiology of other sexually transmitted diseases. *AIDS* **4 Suppl 1**, S99-103.

Cappuccinelli, P., Cagliani, I. & Cavallo, G. (1975). Involvement of a surface concanavalin A-binding glycoprotein in the adhesion of *Trichomonas vaginalis* to substrates. *Experientia* **31**, 1157-1159.

Carney, J.A., Unadkat, P., Yule, A., Rajakumar, R., Lacey, C.J. & Ackers, J.P. (1988). New rapid latex agglutination test for diagnosing *Trichomonas vaginalis* infection. *Journal of Clinical Pathology* **41**, 806-808.

Carr, P.L., Felsenstein, D. & Friedman, R.H. (1998). Evaluation and management of vaginitis. *Journal of General and Internal Medicine* **13**, 335-346.

Centers for Disease Control and Prevention. (1998). Guidelines for the treatment of sexually transmitted diseases. *MMWR* **47**, 1-118.

Chenchik, A., Moqadam, F. & Siebert, P.D. (1995). Marathon cDNA amplification: A new method for cloning full-length cDNAs. *CLONTECHniques* **X**, 5-8.

Cheng, S., Fockler, C., Barnes, W.M. & Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5695-5699.

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. & Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonucleic acid. *Biochemistry* **18**, 5294-5299.

Choromanski, L., Beat, D.A., Nordin, J.H., Pan, A.A. & Honigberg, B.M. (1985). Further studies on the surface saccharides in *Trichomonas vaginalis* strains by fluorescein-conjugated lectins. *Zeitschrift fur Parasitenkunde* **71**, 443-458.

Christian, R.T., Miller, N.F., Ludovic, P.P. & Riley, G.M. (1963). A study of *Trichomonas vaginalis* in human cell culture. *American Journal of Obstetrics and Gynecology* **85**, 947-954.

Clontech Laboratories. Marathon cDNA Amplification Kit User Manual PT1115-1. PR8Y870. 1999. Palo Alto, Ca.

Cobianchi, F. & Wilson, S.H. (1987). Enzymes for modifying and labelling DNA and RNA. *Methods in Enzymology* **152**, 94-110.

Costa e Silva Filho, F., Breier-Saraiva, E.M., Tosta, M.X. & de Souza, W. (1989). *Trichomonas vaginalis* and *Tritrichomonas foetus* secrete neuraminidase into the culture medium. *Molecular and Biochemical Parasitolgy* **35**, 73-78.

Costa e Silva Filho, F., de Souza, W. & Lopes, J.D. (1988). Presence of laminin-binding proteins in trichomonads and their role in adhesion. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8042-8046.

Cotch, M. F. Vaginal infections and prematurity study group. (1990). Carriage of *Trichomonas vaginalis* (Tv) is associated with adverse pregnancy outcome. Abst. 681. Program and Abstracts of the 30th Interscience Conference on Antimicrobials and Chemotherapy, 1990. Washington, D.C., American Society for Microbiology.

Dailey, D.C. & Alderete, J.F. (1991). The phenotypically variable surface protein of *Trichomonas vaginalis* has a single, tandemly repeated immunodominant epitope. *Infection & Immunity* **59**, 2083-2088.

Dailey, D.C., Chang, T.H. & Alderete, J.F. (1990). Characterization of *Trichomonas vaginalis* haemolysis. *Parasitology* **101 Pt 2**, 171-175.

Daly, J.J., Sherman, J.K., Haley, T. & Hostetler, T.L. (1990). Difference in effect of dog seminal fluid and human seminal fluid and semen on *in vitro* survival of *Trichomonas vaginalis*. *Sexually Transmitted Diseases* **17**, 106-109.

de Carli, G.A., Brasseur, P., Rott, M., da Silva, A. & Wendorff, A. (1994). Hemolytic activity of various strains and clones of *Trichomonas vaginalis* and *Tritrichomonas suis*. *Revista Latinoamericana de Microbiologia* **36**, 207-211.

Delgadillo, M.G., Liston, D.R., Niazi, K. & Johnson, P.J. (1997). Transient and selectable transformation of the parasitic protist *Trichomonas vaginalis*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4716-4720.

Demes, P., Gombosova, A., Valent, M., Fabusova, H. & Janoska, A. (1988a). Fewer *Trichomonas vaginalis* organisms in vaginas of infected women during menstruation. *Genitourinary Medicine* **64**, 22-24.

Demes, P., Gombosova, A., Valent, M., Janoska, A., Fabusova, H. & Petrenko, M. (1988b). Differential susceptibility of fresh *Trichomonas vaginalis* isolates to complement in menstrual blood and cervical mucus. *Genitourinary Medicine* **64**, 176-179.

Diamond, L.S., Harlow, D.H. & Cunnick, C.C. (1978). A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **72**, 431-432.

Dias Filho, B.P., Andrade, A.F., de Souza, W., Esteves, M.J. & Angluster, J. (1992). Cell surface saccharide differences in drug-susceptible and drug- resistant strains of *Trichomonas vaginalis*. *Microbios* **71**, 55-64.

Division of STD Prevention & Control, Bureau of HIV AIDS STD and TB. Canadian STD Guidelines--1998 Edition. (1998). Health Canada, Ottawa, Ontario.

Donne, M.A. (1836). Animacules observes dans les matieres purulentes et le produit des secretions des organes genitaux de l'homme et de la femme. *C.R.Acad.Sci.* **3**, 385-386.

Duranti, M.A., Franzoni, L., Sartor, G., Benedetti, A., Iwai, L.K., Gruber, A., Zingales, B., Guzman, F., Kalil, J., Spisni, A. & Cunha-Neto, E. (1999). *Trypanosoma cruzi*: conformational preferences of antigenic peptides bearing the immunodominant epitope of the B13 antigen. *Experimental Parasitology* **93**, 38-44.

Edwards, D.I. (1993). Nitroimidazole drugs--action and resistance mechanisms. I. Mechanisms of action. *Journal of Antimicrobial Chemotherapy* **31**, 9-20.

Ellis, J.E., Yarlett, N., Cole, D., Humphreys, M.J. & Lloyd, D. (1994). Antioxidant defences in the microaerophilic protozoan *Trichomonas vaginalis*: comparison of metronidazole-resistant and sensitive strains. *Microbiology* **140**, 2489-2494.

Engbring, J.A., O'Brien, J.L. & Alderete, J.F. (1996). *Trichomonas vaginalis* adhesins display molecular mimicry to metabolic enzymes. *Advances in Experimental Medicine and Biology* **408**, 207-223.

Fari, A., Trevoux, R. & Verges, V. (1985). Diagnosis and significance of non-motile forms of *Trichomonas vaginalis*. *Abstracts International Symposium on trichomonads and Trichomoniasis, Prague* **1985**, 28a-28a.

Farris, V.K. & Honigberg, B.M. (1970). Behaviour and pathogenicity of *Trichomonas vaginalis* Donné in chick liver cells. *Journal of Parasitology* **56**, 849-882.

Fast, N.M. & Doolittle, W.F. (1999). *Trichomonas vaginalis* possesses a gene encoding the essential spliceosomal component, PRP8. *Molecular and Biochemical Parasitology* **99**, 275-278.

Feinberg, A.P. & Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6-13.

Field, D. & Wills, C. (1996). Long, polymorphic microsatellites in simple organisms. *Proceedings of the Royal Society of London B Biological Sciences* **263**, 209-215.

Fiori, P.L., Rappelli, P., Rocchigiani, A.M. & Cappuccinelli, P. (1993). *Trichomonas vaginalis* haemolysis: evidence of functional pore formation on red cell membranes. *FEMS Microbiology Letters* **109**, 13-18.

Fouts, A.C. & Kraus, S.J. (1980). *Trichomonas vaginalis*: Reevaluation of its clinical presentation and laboratory diagnosis. *Journal of Infectious Diseases* **141**, 137-143.

Galibert, F., Alexandraki, D., Baur, A., Boles, E., Chalwatzis, N., Chuat, J.C., Coster, F., Cziepluch, C., De, H.M., Domdey, H., Durand, P., Entian, K.D., Gatius, M., Goffeau, A., Grivell, L.A., Hennemann, A., Herbert, C.J., Heumann, K., Hilger, F., Hollenberg, C.P., Huang, M.E., Jacq, C., Jauniaux, J.C., Katsoulou, C. & Karpfinger-Hartl, L. (1996). Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome X. *EMBO J.* **15**, 2031-2049.

Garber, G.E. & Bowie, W.R. (1990). The effect of *Trichomonas vaginalis* and the role of pH on cell culture monolayer viability. *Clinical & Investigative Medicine - Medecine Clinique et Experimentale* **13**, 71-76.

Garber, G.E. & Lemchuk-Favel, L.T. (1989). Characterization and purification of extracellular proteases of *Trichomonas vaginalis*. *Canadian Journal of Microbiology* **35**, 903-909.

Garber, G.E. & Lemchuk-Favel, L.T. (1990). Association of production of cell-detaching factor with the clinical presentation of *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **28**, 2415-2417.

Garber, G.E. & Lemchuk-Favel, L.T. (1994). Analysis of the extracellular proteases of *Trichomonas vaginalis*. *Parasitology Research* **80**, 361-365.

Garber, G.E., Lemchuk-Favel, L.T. & Bowie, W.R. (1989). Isolation of a cell-detaching factor of *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **27**, 1548-1553.

Garber, G.E., Lemchuk-Favel, L.T. & Rousseau, G. (1991). Effect of beta-estradiol on production of the cell-detaching factor of *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **29**, 1847-1849.

Garcia-Tamayo, J., Nunez-Montiel, J.T. & de Garcia, H.P. (1978). An electron microscopic investigation on the pathogenesis of human vaginal trichomoniasis. *Acta Cytologica* **22**, 447-455.

Gavel, Y. & von Heijne, G. (1990). Cleavage-site motifs in mitochondrial targeting peptides. *Protein Engineering* **4**, 33-37.

Gerbase, A.C., Rowley, J.T., Heymann, D.H.L., Berkeley, S.F.B. & Piot, P. (1998). Global prevalence and incidence estimates of selected curable STDs. *Sexually Transmitted Infections* **74**, S12-S16

Germot, A. & Philippe, H. (1999). Critical analysis of eukaryotic phylogeny: a case study based on the HSP70 family. *Journal of Eukaryotic Microbiology* **46**, 116-124.

Gillin, F.D. & Sher, A. (1981). Activation of the alternative complement pathway by *Trichomonas vaginalis*. *Infection & Immunity* **34**, 268-273.

Gillis, J.C. & Wiseman, L.R. (1996). Secnidazole: A review of its antimicrobial activity, pharmacokinetic properties, and therapeutic use in the management of protozoal infections and bacterial vaginosis. *Drugs* **51**, 621-638.

Glockner, G., Scherer, S., Schattevoy, R., Boright, A., Weber, J., Tsui, L.C. & Rosenthal, A. (1998). Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes. *Genome Research* **8**, 1060-1073.

Gold, D. & Ofek, I. (1992). Adhesion of *Trichomonas vaginalis* to plastic surfaces: requirement for energy and serum constituents. *Parasitology* **105**, 55-62.

Gombosova, A. & Valent, M. (1990). Dot-immunobinding assay with monoclonal antibody for detection of *Trichomonas vaginalis* in clinical specimens. *Genitourinary Medicine* **66**, 447-450.

Goode, M.A., Grauer, K. & Gums, J.G. (1994). Infectious vaginitis: selecting therapy and preventing recurrence. *Postgraduate Medicine* **96**, 85-98.

Gubler, U. & Hoffman, B.J. (1983). A simple and very efficient method for generating complimentary DNA libraries. *Gene* **25**, 263-269.

Hardy, P.H., Hardy, J.B., Nell, E.E., Graham, D.A., Spence, M.R. & Rosenbaum, R.C. (1984). Prevalence of six sexually transmitted disease agents among pregnant inner-city adolescents and pregnancy outcome. *Lancet* **2**, 333-337.

Hartl, D.L. & Ochman, H. (1994). Inverse Polymerase Chain Reaction. In *Protocols in Gene Analysis* (Harwood, A.J., ed.), Humana Press, Totowa, New Jersey, pp. 187-196.

Hashimoto, T., Sanchez, L.B., Shirakura, T., Muller, M. & Hasegawa, M. (1998). Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6860-6865.

Heath, J.P. (1981). Behaviour and pathogenicity of *Trichomonas vaginalis* in epithelial cell cultures: a study by light and scanning electron microscopy. *British Journal of Venereal Diseases* **57**, 106-117.

Heine, P. & McGregor, J.A. (1993). *Trichomonas vaginalis*: a reemerging pathogen. *Clinical Obstetrics & Gynecology* **36**, 137-144.

Heine, R.P., Wiesenfeld, H.C., Sweet, R.L. & Witkin, S.S. (1997). Polymerase chain reaction analysis of distal vaginal specimens: a less invasive strategy for detection of *Trichomonas vaginalis*. *Clinical Infectious Diseases* **24**, 985-987.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S. (1998). SOSUI: Classification and secondary structure prediction system for membrane proteins. *Bioinformatics (formerly CABIOS)* **14**, 378-379.

Hofmann, K. & Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **347**, 166.

Hogue, M.J. (1943). The effect of *Trichomonas vaginalis* on tissue culture cells. *American Journal of Hygiene* **37**, 142-152.

Honigberg, B.M. (1978). Trichomonads of importance in human medicine. In *Parasitic protozoa* (Kreiger, J.P., ed.), Academic Press, Inc., New York, pp. 275-454.

Honigberg, B.M. & Brugerolle, G. (1990). Structure. In *Trichomonads Parasitic in Humans* (Honigberg, B.M., ed.), Springer-Verlag, New York, pp. 5-35.

Honigberg, B.M. & Ewalt, A.C. (1961). Preliminary observations on pathogenicity of *Trichomonas vaginalis* for cell cultures. *International Congress on Protozoology* **1**, 568-569.

Honigberg, B.M. & King, V.M. (1964). Structure of *Trichomonas vaginalis* Donne. *Journal of Parasitology* **50**, 345-364.

Horner, D.S., Hirt, R.P., Kilvington, S., Lloyd, D. & Embley, T.M. (1996). Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proceedings of the Royal Society of London B Biological Sciences* **263**, 1053-1059.

Hrdy, I. & Muller, M. (1995). Primary structure and eubacterial relationships of the pyruvate: ferredoxin oxidoreductase of the amitochondriate eukaryote *Trichomonas vaginalis*. *Journal of Molecular Evolution* **41**, 388-396.

Hu, Y.W., Balaskas, E., Kessler, G., Issid, C., Scully, L.J., Murphy, D.G., Rinfret, A., Giulivi, A., Scalia,V., & Gill, P. (1998). Primer specific and mispair extension analysis (PSMEA) as a simple approach to fast genotyping. *Nucleic Acids Research* **26**, 5013-5.

Hurt, E.C. (1988). A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. *EMBO J.* 7, 4323-4334.

Ings, R.M., McFadzean, J.A. & Ormerod, W.E. (1974). The mode of action of metronidazole in *Trichomonas vaginalis* and other microorganisms. *Biochemical Pharmacology* **23**, 1421-1429.

Irvine, J.W., Coombs, G.H. & North, M.J. (1993). Purification of cysteine proteinases from trichomonads using bacitracin-Sepharose. *FEMS Microbiology Letters* **110**, 113-119.

Irvine, J.W., North, M.J. & Coombs, G.H. (1997). Use of inhibitors to identify essential cysteine proteinases of *Trichomonas vaginalis*. *FEMS Microbiology.Letters.* **149**, 45-50.

Johnson, P.J., d'Oliviera, C.E., Gorrell, T.E., & Muller, M. (1990). Molecular analysis of the hydrogenosomal ferredoxin of the anaerobic protist *Trichomonas vaginalis*. *Proceedings of the National Academy of Sciences of the USA* **87**, 6097-6101.

Kon, V.B., Papadimitriou, J.M., Robertson, T.A. & Warton, A. (1988). Quantitation of concanavalin A and wheat germ agglutinin binding by two strains of *Trichomonas vaginalis* of differing pathogenicity using gold particle-conjugated lectins. *Parasitology Research* **75**, 7-13.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research* **15**, 8125-48.

Krieger, J.N. (1995). Trichomoniasis in men: old issues and new data. *Sexually Transmitted Diseases* **22**, 83-96.

Krieger, J.N., Jenny, C., Verdon, M., Siegel, N., Springwater, R. & Critchlow, C.W. (1993). Clinical manifestations of trichomoniasis in men. *Annals of Internal Medicine* **118**, 844-849.

Krieger, J.N., Poisson, M.A. & Rein, M.F. (1983). Beta-hemolytic activity of *Trichomonas vaginalis* correlates with virulence. *Infection & Immunity* **41**, 1291-1295.

Krieger, J.N., Ravdin, J.L. & Rein, M.F. (1985). Contact-dependent cytopathogenic mechanisms of *Trichomonas vaginalis*. *Infection and Immunity* **50**, 778-786.

Krieger, J.N. & Rein, M.F. (1982). Canine prostatic secretions kill *Trichomonas vaginalis*. *Infection and Immunity* **37**, 77-81.

Krieger, J.N., Tam, M.R., Stevens, C.E., Nielsen, I.O., Hale, J. & Kiviat, N.B. (1988). Diagnosis of trichomoniasis. Comparison of conventional wet-mount examination with cytologic studies, cultures, and monoclonal antibody staining of direct specimens. *Journal of the American Medical Association* **259**, 1223-1227.

Krieger, J.N., Verdon, M., Siegel, N. & Holmes, K.K. (1993). Natural history of urogenital trichomoniasis in men. *Journal of Urology* **149**, 1455-1458.

Krieger, J.N., Wolner-Hanssen, P., Stevens, C. & Holmes, K.K. (1990). Characteristics of *Trichomonas vaginalis* isolates from women with and without colpitis macularis. *Journal of Infectious Diseases* **161**, 307-311.

Kulda, J. (1967). Effects of different species of trichomonads on monkey kidney cell culture. *Folia Parasitol.* **14**, 295-310.

Kulda, J., Tachezy, J. & Cerkasovova, A. (1993). *In vitro* induced anaerobic resistance to metronidazole in *Trichomonas vaginalis*. *Journal of Eukaryotic Microbiology* **40**, 262-269.

Lackie, J.M., & Dow, J.A.T. Eds. (1999). The Dictionary of Cell and Molecular Biology. Third Ed., Academic Press, London. http://www.mblab.gla.ac.uk/dictionary/

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

Laga, M., Alary, M., Nzila, N., Manoka, A.T., Tuliza, M., Behets, F., Goeman, J.S.L.M. & Piot, P. (1994). Condom promotion, sexually transmitted diseases treatment, and declining incidence of HIV-1 infection in female Zairian sex workers. *Lancet* **344**, 246-248.

Laga, M., Manoka, A., Kivuvu, M., Malele, B., Tuliza, M., Nzila, N., Goeman, J.B.F., Batter, V., Alary, M. & *et al.* (1993). Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: results from a cohort study. *AIDS* **7**, 95-102.

Lahti, C.J., d'Oliviera, C.E., & Johnson, P.J. (1992). β-succinyl-coenzyme A synthetase from *Trichomonas vaginalis* is a soluble hydrogenosomal protein with an amino-terminal sequence that resembles mitochondrial pre-sequences. *Journal of Bacteriology* **174**, 6822-6830.

Lau, A.H., Lam, N.P., Piscitelli, S.C., Wilkes, L. & Danziger, L.H. (1992). Clinical pharmacokinetics of metronidazole and other nitroimidazole anti-infectives. *Clinical Pharmacokinetics* **23**, 328-364.

Lehker, M.W., Arroyo, R. & Alderete, J.F. (1991). The regulation by iron of the synthesis of adhesins and cytadherence levels in the protozoan *Trichomonas vaginalis*. *Journal of Experimental Medicine* **174**, 311-318.

Lehker, M.W., Chang, T.H., Dailey, D.C. & Alderete, J.F. (1990). Specific erythrocyte binding is an additional nutrient acquisition system for *Trichomonas vaginalis*. *Journal of Experimental Medicine* **171**, 2165-2170.

Levi, M.H., Torres, J., Pina, C. & Klein, R.S. (1997). Comparison of the InPouch TV culture system and Diamond's modified medium for detection of *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **35**, 3308-3310.

Lewis, D.A., Habgood, L., White, R., Barker, K.F. & Murphy, S.M. (1997). Managing vaginal trichomoniasis resistant to high-dose metronidazole therapy. *International Journal of STD and AIDS* **8**, 780-784.

Lin, P.R., Shaio, M.F. & Liu, J.Y. (1997). One-tube, nested-PCR assay for the detection of *Trichomonas vaginalis* in vaginal discharges. *Annals of Tropical Medicine and Parasitology* **91**, 61-65.

Lirosi, G. & Guarascio, A. (1972). Effects of hormonal changes in the vaginal environment in the treatment of vaginitis especially due to Trichomonas. *Minerva Ginecol.* **24**, 23-27.

Lisi, P.J., Dondero, R.S., Kwiatkoski, D., Spence, M.R., Rein, M.F. & Alderete, J.F. (1988). Monoclonal-antibody-based enzyme-linked immunosorbent assay for *Trichomonas vaginalis*. *Journal of Clinical Microbiology* **26**, 1684-1686.

Liston, D.R. & Johnson, P.J. (1998). Gene transcription in *Trichomonas vaginalis*. *Parasitology Today* **14**, 261-265.

Liston, D.R. & Johnson, P.J. (1999). Analysis of a ubiquitous promoter element in a primitive eukaryote: early evolution of the initiator element. *Molecular Cell Biology* **19**, 2380-2388.

Lockwood, B.C., North, M.J., Scott, K.I., Bremner, A.F. & Coombs, G.H. (1987). The use of a highly sensitive electrophoretic method to compare the proteinases of trichomonads. *Molecular & Biochemical Parasitology* **24**, 89-95.

Loffert, D., Karger, S., Berkenkopg, M., Seip, N. & Kang, J. (1997). PCR optimization: primer design. *Qiagen News* **5**, 3-6.

Lossick, J.G. (1990a). Epidemiology of urogenital trichomoniasis. In *Trichomonads parasitic in humans* (Honigberg, B.M., ed.), Springer-Verlag, New York, New York, pp. 311-323.

Lossick, J.G. (1990b). Treatment of sexually transmitted vaginosis/vaginitis. *Reviews of Infectious Diseases* **12 Suppl 6:S665-81**, S665-S681

Lossick, J.G. & Kent, H.L. (1991). Trichomoniasis: Trends in diagnosis and management. *American Journal of Obstetrics and Gynecology* **165**, 1217-1222.

Lossick, J.G., Muller, M. & Gorrell, T.E. (1986). *In vitro* drug susceptibility and doses of metronidazole required for cure in cases of refractory vaginal trichomoniasis. *Journal of Infectious Diseases* **153**, 948-955.

Luersen, K., Walter, R.D. & Muller, S. (1999). The putative gamma-glutamylcysteine synthetase from *Plasmodium falciparum* contains large insertions and a variable tandem repeat. *Molecular and Biochemical Parasitology* **98**, 131-142.

Lumsden, W.H., Robertson, D.H., Heyworth, R. & Harrison, C. (1988). Treatment failure in *Trichomonas vaginalis* vaginitis. *Genitourinary Medicine* **64**, 217-218.

Mantovani, A., Polentarutti, N., Peri, G., Martinotti, M.G. & Landolfo, S. (1981). Cytotoxicity of human peripheral blood monocytes against *Trichomonas vaginalis*. *Clinical & Experimental Immunology* **46**, 391-396.

Marquardt, W.C. & Demaree, R.S.Jr. (1985). Introduction to the Protozoa. In *Parasitology* (Marquardt, W.C. and Demaree, R.S.Jr., eds.), Macmillan Publishing Company, New York, pp. 1-25.

Martinotti, M.G., Martinetto, P. & Savoia, D. (1986). Adherence of *Trichomonas vaginalis* to cell culture monolayers. *European Journal of Clinical Microbiology* **5**, 320-323.

Mathisen, G.E. & Finegold, S.M. (1992). Metronidazole and other nitroimidazoles. In *Treatment of Infectious Diseases* (Gorbach, J., Bartlett and Blackow, eds.), W.S. Saunders Company, Philadelphia, pp. 260-265.

McGeoch, D.J. (1985). On the predictive recognition of signal peptide sequences. *Virus Research* **3**, 271-286.

McGrory, T. & Garber, G.E. (1992). Mouse intravaginal infection with *Trichomonas vaginalis* and role of *Lactobacillus acidophilus* in sustaining infection. *Infection and Immunity* **60**, 2375-2379.

McGrory, T., Meysick, K.C., Lemchuk-Favel, L.T. & Garber, G.E. (1994). The interaction of *Lactobacillus acidophilus* and *Trichomonas vaginalis in vitro*. *Journal of Parasitology* **80**, 50-54.

McMaster, G.K. & Carmichael, G.G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 4835-4838.

Meade, J.C., Shah, P.H. & Lushbaugh, W.B. (1997). *Trichomonas vaginalis*: analysis of codon usage. *Experimental Parasitology* **87**, 73-74.

Meysick, K. C. (1996). Molecular Characterization and Expression of a N-acetylneuraminate Lyase Gene from *Trichomonas vaginalis*. University of Ottawa, Ottawa, Canada.

Meysick, K.C., Dimock, K. & Garber, G.E. (1996). Molecular characterization and expression of a N-acetylneuraminate lyase gene from *Trichomonas vaginalis*. *Molecular and Biochemical Parasitolgy* **76**, 289-292.

Min, D.Y., Ryu, J.S., Park, S.Y., Shin, M.H. & Cho, W.Y. (1997). Degradation of human immunoglobulins and cytotoxicity on HeLa cells by live *Trichomonas vaginalis*. *Korean Journal of Parasitology* **35**, 39-46.

Minkoff, H., Grunebaum, A.N., Schwarz, R.H., Feldman, J., Cummings, M.C.W., Clark, L., Pringle, G. & McCormack, W.M. (1984). Risk factors for prematurity and premature rupture of membranes: a prospective study of the vaginal flora in pregnancy. *American Journal of Obstetrics & Gynecology* **150**, 965-972.

Mirhaghani, A. & Warton, A. (1998). Involvement of *Trichomonas vaginalis* surface-associated glycoconjugates in the parasite/target cell interaction. A quantitative electron microscopy study. *Parasitology Research* **84**, 374-381.

Moreno, S.N., Mason, P.R. & Docampo, R. (1984). Distinct reduction of nitrofurans and metronidazole to free radical metabilites by *Tritrichomonas foetus* hydrogenosomal and cytosolic enzymes. *Journal of Biological Chemistry* **259**, 8252-8259.

Muller, M. (1980). The hydrogenosome. *Symposia of the Society of General Microbiology* **30**, 127-142.

Muller, M. (1981). Action of clinically utilized 5-nitroimidazoles on microorganisms. *Scandinavian Journal of Infectious Diseases - Supplementum* **26**, 31-41.

Muller, M. (1990). Biochemistry of *Trichomonas vaginalis*. In *Trichomonads Parasitic in Humans* (Honigberg, B.M., ed.), Springer-Verlag, New York, pp. 53-83.

Muller, M. (1993). The hydrogenosome. *Journal of General Microbiology* **139**, 2879-2889.

Muller, M. & Gorrell, T.E. (1983). Metabolism and metronidazole uptake in *Trichomonas vaginalis* isolates with different metronidazole susceptibilities. *Antimicrobial Agents & Chemotherapy* **24**, 667-673.

Muller, M., Lossick, J.G. & Gorrell, T.E. (1988). *In vitro* susceptibility of *Trichomonas vaginalis* to metronidazole and treatment outcome in vaginal trichomoniasis. *Sexually Transmitted Diseases* **15**, 17-24.

Muresu, R., Rubino, S., Rizzu, P., Baldini, A., Colombo, M. & Cappuccinelli, P. (1994). A new method for identification of *Trichomonas vaginalis* by fluorescent DNA in situ hybridization. *Journal of Clinical Microbiology* **32**, 1018-1022.

Musatovova, O. & Alderete, J.F. (1998). Molecular analysis of the gene encoding the immunodominant phenotypically varying P270 protein of *Trichomonas vaginalis*. *Microbial Pathogenesis* **24**, 223-239.

Nakai, K. and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**, 897-911.

Narcisi, E.M. & Secor, W.E. (1996). *In vitro* effect of tinidazole and furazolidone on metronidazole-resistant *Trichomonas vaginalis*. *Antimicrobial Agents and Chemotherapy* **40**, 1121-1125.

Neale, K.A. & Alderete, J.F. (1990). Analysis of the proteinases of representative *Trichomonas vaginalis* isolates. *Infection & Immunity* **58**, 157-162.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1-6.

Nielsen, M.H. & Nielsen, R. (1975). Electron microscopy of *Trichomonas vaginalis* Donne: interaction with vaginal epithelium in human trichomoniasis. *Acta Pathologica et Microbiologica Scandinavica - Section B, Microbiology* **83**, 305-320.

North, M.J. (1982). Comparative biochemistry of the proteinases of eucaryotic microorganisms. *Microbiological Reviews* **46**, 308-340.

North, M.J. (1994). Cysteine endopeptidases of parastic protozoa. *Methods in Enzymology* **244**, 523-539.

O'Brien, J.L., Lauriano, C. & Alderete, J.F. (1996). Molecular characterization of a third malic enzyme-like AP65 adhesin gene of *Trichomonas vaginalis*. *Microbial Pathogenesis* **20**, 335-349.

Palmer, J.D. (1997). Organelle genomes: going, going, gone! *Science* **275**, 790-791.

Pandey, A.V., Joshi, R., Tekwani, B.L., Singh, R.L. & Chauhan, V.S. (1997). Synthetic peptides corresponding to a repetitive sequence of malarial histidine rich protein bind haem and inhibit haemozoin formation *in vitro*. *Molecular and Biochemical Parasitology* **90**, 281-287.

Perez-Reyes, E., Kalyanaraman, B. & Mason, P.R. (1980). The reductive metabolism of metronidazole and ornidazole by aerobic liver microsomes. *Molecular Pharmacology* **17**, 244

Peterson, K.M. & Alderete, J.F. (1982). Host plasma proteins on the surface of pathogenic *Trichomonas vaginalis*. *Infection & Immunity* **37**, 755-762.

Peterson, K.M. & Alderete, J.F. (1984). Selective acquisition of plasma proteins by *Trichomonas vaginalis* and human lipoproteins as a growth requirement for this species. *Molecular & Biochemical Parasitology* **12**, 37-48.

Peterson, K.M., Baseman, J.B. & Alderete, J.F. (1987). Molecular cloning of *Treponema pallidum* outer envelope fibronectin binding proteins, P1 and P2. *Genitourinary Medicine* **63**, 355-360.

Petrin, D., Delgaty, K., Bhatt, R. & Garber, G. (1998). Clinical and microbiological aspects of *Trichomonas vaginalis*. *Clinical Microbiology Reviews* **11**, 300-317.

Pindak, F.F., Gardner, W.A., Jr. & Mora de Pindak, M. (1986). Growth and cytopathogenicity of *Trichomonas vaginalis* in tissue cultures. *Journal of Clinical Microbiology* **23**, 672-678.

Pindak, F.F., Mora de Pindak, M. & Gardner, W.A., Jr. (1993). Contact-independent cytotoxicity of *Trichomonas vaginalis*. *Genitourinary Medicine* **69**, 35-40.

Proudfoot, N.J. & Brownlee, G.G. (1976). 3' non-coding region sequences in eukaryotic messenger RNA. *Nature* **263**, 211-214.

Provenzano, D. & Alderete, J.F. (1995). Analysis of human immunoglobulin-degrading cysteine proteinases of *Trichomonas vaginalis*. *Infection & Immunity* **63**, 3388-3395.

Quon, D.V., d'Oliveira, C.E. & Johnson, P.J. (1992). Reduced transcription of the ferredoxin gene in metronidazole- resistant *Trichomonas vaginalis*. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4402-4406.

Ranford-Cartwright, L.C., Taylor, R.R., Asgari-Jirhandeh, N., Smith, D.B., Roberts, P.E., Robinson, V.I., Babiker, H.A., Riley, E.M., Walliker, D. & McBride, J.S. (1996). Differential antibody recognition of FC27-like Plasmodium falciparum merozoite surface protein MSP2 antigens which lack 12 amino acid repeats. *Parasite Immunology* **18**, 411-420.

Rasmussen, S.E., Nielsen, M.H., Lind, I. & Rhodes, J.M. (1986). Morphological studies of the cytotoxicity of *Trichomonas vaginalis* to normal human vaginal epithelial cells *in vitro*. *Genitourinary Medicine* **62**, 240-246.

Rein, M.F. (1990). Clinical manifestations of urogenital trichomoniasis in women. In *Trichomonads Parasitic in Humans* (Honigberg, B.M., ed.), Springer-Verlag, New York, pp. 225-234.

Rein, M.F. & Chapel, T.A. (1975). Trichomoniasis, candidiasis, and the minor venereal diseases. *Clinical Obstetrics & Gynecology* **18**, 73-88.

Rein, M.F. & Holmes, K.K. (1983). Non-specific vaginitis, vulvovaginal candidiasis and trichomoniasis: Clinical features, diagnosis and management. In *Current Clinical Topics in Infectious Diseases* (Remington, J. and Swartz, M.N., eds.), Blackwell Scientific Publ., New York, pp. 281-315.

Rendon-Maldonado, J.G., Espinosa-Cantellano, M., Gonzalez-Robles, A. & Martinez-Palomo, A. (1998). *Trichomonas vaginalis*: *in vitro* phagocytosis of lactobacilli, vaginal epithelial cells, leukocytes, and erythrocytes. *Experimental Parasitology* **89**, 241-250.

Riley, D.E., Campbell, L.A., Puolakkainen, M. & Krieger, J.N. (1993). *Trichomonas vaginalis* and early evolving DNA and protein sequences of the CDC2/28 protein kinase family. *Molecular Microbiology* **8**, 517-519.

Riley, D.E. & Krieger, J.N. (1995). Molecular and phylogenetic analysis of PCR-amplified cyclin- dependent kinase (CDK) family sequences from representatives of the earliest available lineages of eukaryotes. *Journal of Molecular Evolution* **41**, 407-413.

Roger, A.J., Clark, C.G. & Doolittle, W.F. (1996). A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 14618-14622.

Roger, A.J., Morrison, H.G. & Sogin, M.L. (1999). Primary structure and phylogenetic relationships of a malate dehydrogenase gene from *Giardia lamblia*. *Journal of Molecular Evolution* **48**, 750-755.

Rossignol, J.-F., Maisonneuve, H. & Cho, Y.W. (1984). Nitroimidazoles in the treatment of trichomoniasis, giardiasis, and amebiasis. *International Journal of Clincial Pharmacology* **22**, 63-72.

Roussel, F., de Carli, G.A. & Brasseur, P. (1991). A cytopathic effect of *Trichomonas vaginalis* probably mediated by a mannose/N-acetyl-glucosamine binding lectin. *International Journal for Parasitology* **21**, 941-944.

Seed, B., Parker, R.C. & Davidson, N. (1982). Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digest. *Gene* **19**, 201-209.

Shaio, M.F. & Lin, P.R. (1995). Influence of humoral immunity on leukotriene B4 production by neutrophils in response to *Trichomonas vaginalis* stimulation. *Parasite Immunology* **17**, 127-133.

Sharma, P., Malla, N., Gupta, I., Ganguly, N.K. & Mahajan, R.C. (1991). Anti-trichomonad IgA antibodies in trichomoniasis before and after treatment. *Folia Microbiologica* **36**, 302-304.

Siebert, P.D., Chenchik, A., Kellogg, D.E. & Lukyanov, K.A. (1995). An improved method for walking in uncloned genomic DNA. *Nucleic Acids Research* **23**, 1087-1088.

Stierhof, Y.D., Wiese, M., Ilg, T., Overath, P., Haner, M. & Aebi, U. (1998). Structure of a filamentous phosphoglycoprotein polymer: the secreted acid phosphatase of *Leishmania mexicana*. *Journal of Molecular Biology* **282**, 137-148.

Street, D.A., Wells, C., Taylor-Robinson, D. & Ackers, J.P. (1984). Interaction between *Trichomonas vaginalis* and other pathogenic microorganisms of the human genital tract. *British Journal of Venereal Diseases* **60**, 31-38.

Tabrizi, S.N., Paterson, B., Fairley, C.K., Bowden, F.J. & Garland, S.M. (1997). A self-administered technique for the detection of sexually transmitted diseases in remote communities. *Journal of Infectious Diseases* **176**, 289-292.

Tachezy, J., Kulda, J. & Tomkova, E. (1993). Aerobic resistance of *Trichomonas vaginalis* to metronidazole induced *in vitro*. *Parasitology* **106**, 31-37.

Thomas, D.D., Baseman, J.B. & Alderete, J.F. (1985). Fibronectin mediates *Treponema pallidum* cytadherence through recognition of fibronectin cell-binding domain. *Journal of Experimental Medicine* **161**, 514-525.

Verra, F. & Hughes, A.L. (1999). Biased amino acid composition in repeat regions of Plasmodium antigens. *Mol.Biol.Evol.* **16**, 627-633.

Von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14**, 4683-4690.

Wang, A.L. & Wang, C.C. (1985). Isolation and characterization of DNA from *Tritrichomonas foetus* and *Trichomonas vaginalis*. *Molecular & Biochemical Parasitology* **14**, 323-335.

Warton, A. & Honigberg, B.M. (1979). Structure of trichomonads as revealed by scanning electron microscopy. *Journal of Protozoology* **26**, 56-62.

Warton, A. & Honigberg, B.M. (1980). Lectin analysis of surface saccharides in two *Trichomonas vaginalis* strains differing in pathogenicity. *Journal of Protozoology* **27**, 410-419.

Warton, A. & Honigberg, B.M. (1983). Analysis of surface saccharides in *Trichomonas vaginalis* strains with various pathogenicity levels by fluorescein-conjugated plant lectins. *Zeitschrift fur Parasitenkunde* **69**, 149-159.

Wolner-Hanssen, P., Krieger, J.N., Stevens, C.E., Kiviat, N.B., Koutsky, L., DeRouen, T., Hillier, S. & Holmes, K.K. (1989). Clinical manifestations of vaginal trichomoniasis. *Journal of the American Medical Association* **261**, 571-576.

World Health Organization. (1995). *"An overview of selected curable sexually transmitted diseases" in Global Program on AIDS*. World Health Organization Geneva, pp. 2-27.

Yang, Y.M., Ortega, Y., Sterling, C. & Adam, R.D. (1994). *Giardia lamblia* trophozoites contain multiple alleles of a variant- specific surface protein gene with 105-base pair tandem repeats. *Molecular and Biochemical Parasitology* **68**, 267-276.

Zubiaga, A.M., Belasco, J.G., & Greenberg, M.E. (1995). The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Molecular and Cellular Biology* **15**, 2219-2230.

## APPENDIX A: Primers and Adaptors

Inverse PCR Primers

95C 5'-GTC TGG CGG TAT CTT GAA GG-3'

184 5'-AAC AAG AGA CAC CAG CAC CA-3'

631C 5'-TTG TCT TGA GCT GCT TCG GA-3'

710 5'-ATC GCA ACC ATC TAC AGG AC-3'

Sequencing Primers

| | |
|---|---|
| T7 | 5'-TAA TAC GAC TCA CTA TAG GG-3' |
| M13Rev | 5'-CAG GAA ACA GCT ATG AC-3' |
| SP6 | 5'-GAT TTA GGT GAC ACT ATA G-3' |
| 675 | 5'-CCA TTA TCA AAT ATC CAA GC-3' |
| 1246 | 5'-CAC AAT CAA ATT AAA AAT CA-3' |
| 2169 | 5'-AAT TCT CAA CCA AAT CAA CC-3' |
| 2192C | 5'-CAA TGG TTG ATT TGG TTG AG-3' |
| 3039 | 5'-AAT CAA ACT CGG TAA TTC TCA ACC AA-3' |
| 3876 | 5'-TGG AAT CAC ATT AGG GCA ACA AAC ATC-3' |
| 4501 | 5'-AAA GCA ATT TCC GCA CCA GG-3' |
| 5142 | 5'-CAT GCC TTC AGA AAT AAA TGT CAA GAT TCC-3' |
| 5841 | 5'-TAT TAG CAA AGG ATC ATC TC-3' |
| 7201C | 5'-AGG CTC GGA AGA CCC ATT TC-3' |

Primer Extension Primers

| | |
|---|---|
| PE1-Gsp1-163C | 5'-GTG ATT TGG CGC TGC TGA TTC TGT A-3' |
| PE2-GSP2-89C | 5'-CGG TAT CTT GAA GGG ATC TGG ATT GA-3' |

Genome Walker Adaptors and Primers

Genome Walker Adaptor

5'-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT-3'
3'-$H_2N$-CCCGACCA-$PO_4$-5'

Genome Walker Adaptor Primers

AP1 5'-GTA ATA CGA CTC ACT ATA GGG C-3'

AP2 5'-ACT ATA GGG CAC GCG TGG T-3'

Genome Walker Gene Specific Primers

| | |
|---|---|
| Gsp1-163C | 5'-GTG ATT TGG CGC TGC TGA TTC TGT A-3' |
| GSP2-89C | 5'-CGG TAT CTT GAA GGG ATC TGG ATT GA-3' |
| Gsp1-3876 | 5'-TGG AAT CAC ATT AGG GCA ACA AAC ATC-3' |
| Gsp2-3889 | 5'-GGG CAA CAA ACA TCA GAA GAG AAC AAA-3' |
| Gsp1-201C | 5'-TAG TTC ATC AGG CAA ATT CGA AGT TCC-3' |
| Gsp2-73 | 5'-AAA CAC CAC TTT CTA ATG CAC CGA AA -3' |
| Gsp1-5142 | 5'-CAT GCC TTC AGA AAT AAA TGT CAA GAT TCC-3' |
| Gsp2-5160 | 5'-TCC AGA ACA AGT TCA CTC ACA ACA AAG TAA-3' |
| Gsp1-7224 | 5'-ACT TCA GTT GAA CCT TAA TGG AAA TGG TC-3' |
| Gsp2-7284 | 5'-TTT CAA CAA GCC AGG TAT TCA GAG GTT G-3' |
| ORF-868 | 5'-AAA CTT AGG ATG GAC TCT CAG GCA GC-3' |
| ORF-7451C | 5'-TGT TAT CAT GCA TTT CCT ACT GAT CAA TGA-3' |

Marathon RACE Adaptors and Primers

Marathon RACE 1st Strand Synthesis Primer

$NNTTTT_{30}$

Marathon cDNA Adaptor

5'-CTAATACGACTCACTATAGGGCTCGAGCGGCCGCCCGGGCAGGT-3'
3'-$H_2$NCCCGTCCA-$PO_4$-5'

Marathon RACE Primers

| | |
|---|---|
| RACE-AP1 | 5'-CCA TCC TAA TAC GAC TCA CTA TAG GGC-3' |
| RACE-AP2 | 5'-ACT CAC TAT AGG GCT CGA GCG GC-3' |

RACE Gene Specific Primers

5'- RACE

| | |
|---|---|
| GSP1-163C | 5'-GTG ATT TGG CGC TGC TGA TTC TGT A-3' |
| nGSP2-89C | 5'-CGG TAT CTT GAA GGG ATC TGG ATT GA-3' |
| GSP1-1458C | 5'-TCC TAA TCC AAG ACC GTT TCC TCC GA-3' |
| nGSP2-1433C | 5'-ATT GCT GGT TTC GCC TCA GAG TTT GG-3' |

3' RACE

| | |
|---|---|
| GSP1-5862 | 5'-TGC ATC TGT ATC AAT CAA GCG ATT GGA-3' |
| nGSP2-5880 | 5'-CGC ATT GGA CAA ATC AGG AGA TGA CG-3' |

# APPENDIX B: CHARACTERISED SEQUENCE

```
   1 AACAAAAAAA GTCACAAAAC CCGTCATATC CTACTACTCG GATCTGAATC ATGAGAATTT
  61 GATCGCATTC ATTCTACACT ATCCTGTGAT AATGGTGCAT CATGTTATTT CATTTAAATG
 121 GAAATTTCCG TTTAACCAAT ATGATTTGTG ACGTTGATAT TGTCCAGCAG CGAATTGAAC
 181 GATTCGTACA AGCAGATCTA TACAGATATT ACACAACTCT ATAACTACTG AATATCAGAT
 241 AGCTCATTAT CCACGAGTTT TTATTAATTT CTTCGTGCCA AAATGTGACA ATTTTGAATA
 301 TGACAGTGTT TATTTATTCT GCCGTCGCAC ATATAAAATT GAGAAAAATG AGATCATGAA
 361 AACGCACTTA TTAAAAGACA AAAATTGGAT ACTGATTATG CGCTTAACAC ACCAAGAAAC
 421 AATTTGCTCA TCATTATGGC TTTTGTAAAT TCATTATCTG AGATTAAATT GAATAAATTT
 481 AATGATTCAT ATTTGTTATC CAATTCGTGT TGCATGTCTA CGCATAGATG AGAGATGATG
 541 CATTACATTA GGACCAATTC AAGCATTAAT CAGAATGGAA ATTGAATGAG TGGAATATCA
 601 TTTGAATTTT TCAGGTTAGT ATCAATAAAA CTACAACAAC TTAGCTAAAC CAAGGACTTT
 661 TATATGACAT AAATACCTAA TTCGTGATGA ATCATGTAAA ATTTAGGAGA TTTTGCTAAA
 721 ATTTCCACCT TCTATCATAA AACTTATTGC TTTTTACTTA AAAATTCCAT CATATAAGTC
 781 ACGCTTTTTA ATTATTTACA GTAAATCAAG TTCCATAATA TTTTAAAAAT TGTGGACTCT
 841 ACCAACGATA AGAGATTTTG TAAAAAAAAA CTTAGGATGG ACTCTCAGGC AGCGCCGAAA
 901 CGAAGAATAC ATTTGTTTCG TCAGAAACAA AACATAGATT GGGGGTTCAA TCCAGATCCC
 961 TTCAAGATAC CGCCAGACAT GGCGGACTGT GTTGTAACGC GTAATCCTGT TTTGTCTGTT
1021 TTACAGAATC AGCAGCGCCA AATCACTAAC GAACCATACA TTCCTAAACA AGAGACACCA
1081 GCACCATTAA ATGCTTCTGT TTCTCAAAAT CAGCAATTAA CAATTAACAA TTCTGCATCT
1141 AATATCGCTG TCCAACAAAA ATCTGAACAA CCTGCAGGTC AAGATAATAA ATTATCCACT
1201 ATTGGCCAAT TTCCAAGTAT TGAAAATAAA GATGCAAATA ATCAAACAGC TCCAACAATA
1261 TCATTATCAC TCAAGAATCC TCTTAAGCCA GCTGATAACT TACTTGCTAA ATCAGGCCAA
1321 GAAAATGAAA ATAAGGCCGA GTCAAATACA GATTCGAAGC CAAAGCTTAC TGCAATGAAC
1381 AACCAATTTA GCTTGAAGCC ATCTACACCA AACTCTGAGG CGAAACCAGC AATCGGAGGA
1441 AACGGTCTTG GATTAGGAAT GAAACTTTCT TCTTCAAAGC CATTATCTAA CATTTCCGAA
1501 GCAGCTCAAG ACAAACCAGC TACAGGCGGT CTTTCATTGA AGCTTCCTTC TTCAAAGCCA
1561 TTATCAAATA TCCAAGCTGC TTCTGAAGAG AAATCGCAAC CATCTACAGG ACTGAAACTT
1621 GGCATGAAGT TACCTTCAAA CCAAGCTTCA GAAGAAAAAC CTAAACTTGG CTTGAACCTT
1681 CCTAAATCTC CTTCTAATTC AACAGAAGAG AAACCTAAAC TTGGCTTGAA CCTCGGTAAA
1741 TCTCCATCTA CAGAAGAAAA ACCTAAGCTT CAATTAGGTG GAATCAAAAT CGGACAGACA
1801 CCATCTAATT CAACAGAAGA GAAACCAAAA TTAGGATTAA ATCTTCCTAA ATCTCCATCA
1861 ACAGAAGAAA AGCCAAAGTT AAGTTTGAAC CTTGGTAAAT CTCCTTCTAC AGAAGAAAAG
1921 CCAAAACTTT CTCTAAATCT TGGTAAATCT CCTTCTAATC AATCAACAGA AGAAAAACCA
1981 AAATTAGGAT TAAATCTTCC TAAATCTCCA TCAAGTCAAA CTTCTGAAGA AAAGCCAAAG
2041 TTAAGTTTGA ACCTTGGTAA ATCTCCTTCA AATCAGTCTA CAGAAGAGAA ACCAAAACTT
2101 CAATTAGGTG GAATCAAACT CGGACAGACA CAATCAAATC AAAAATCAGA AGAGAAACCG
2161 AAATTAAGCT TAAATCTTGG AAAATCATCA ACAGAAGAGA AACCAAAACT TTCTCTAAAT
2221 CTTGGCAAAT CTCCTTCTAA TCAATCTACA GAAGAAAAAC CAAAGCTAGG ATTTAATCTT
2281 CCTAAAGCTC CATCAAATCA AACAGAAGAA AAACCAAAAT TAGGAACAGG TGGAATTTCA
2341 TTGAACTTAG GAAATAAACC ACAATCAGAA GAGAAACCAA AACTTTCATT AGGTGGAATT
2401 AAACTCGCAC AATCTCCATC AAATTCAAAT GAAGAAAAAC CAAAACTTTC TCTAAATCTT
2461 CCTAAATCTC CATCAAATCA ATCAACAGAA GAGAAACCAA AGCTAGGATT TAATCTTCCT
2521 AAAGCTCCAT CAAATCAAAC AGAAGAAAAA CCAAAATTAG GAACAGGTGG AATTTCATTG
2581 AACTTAGGAA ATAAACCACA ATCAGAAGAG AAACCAAAAC TTTCATTAGG TGGAATTAAA
2641 CTCGCACAAT CTCCATCAAA TTCAAATGAA GAAAAACCAA AACTTTCTCT AAATCTTCCT
2701 AAATCTCCAT CAAATCAATC AACAGAAGAG AAACCAAAGC TTCAACTTGG TGGAATTAAA
2761 TTAAACTTAG GAAATAAACC ACAGACAGAA ACACAAACAG AAGAGAAGCC AAAACTCCAA
```

2821 TTAGGAGGCA TCAAATTAGG TCAATCTCCT TCTAATTCAA CAGAAGAGAA GCCAAAACTT
2881 CAATTAGGAG GAATCAAACT AAACTTAGGA AGTAAACCAC AAACAGAAGA GAAACCAAAG
2941 CTTCAATTAG GTGGAATCAA ATTAGGAACA GGTGGAATTT CATTGAACTT AGGAAATAAA
3001 CCACAATCAG AAGAGAAACC AAAACTTCAA CTTGGTGGAA TCAAACTCGG TAATTCTCAA
3061 CCAAATCAAC CATTGGAAAA ACCAAAGTCA GGAATTAATC TCAACTTAGG AAAATCTCAA
3121 CCTTCTTCTG AAGAGAAACC AAAACTTGGA TTGAATCTCG GTAAATCTCC ATCAAATTCA
3181 ACAGAAGAGA AACCGAAATT GGGAACAGGC GGAATTTCAT TAAACTTAGG AAATAAACCA
3241 CAAACAGAAG AGAAACCAAA ACTTTCTCTA AATCTTCCTA AATCTCCATC AAATCAAAAT
3301 CAATCAACAG AAGAAAAACC AAAACTCCAA TTAGGAGGAT TGAAATTAAA CTTAGGAAAT
3361 AAACCACAGA CAGAAACACA AACAGAAGAG AAACCAAAGC TTCAACTCGG TGGAATTAAA
3421 TTAAACTTAG GAAGTAAATC ACAGACAGAA GAGAAACCTA AATTCCAATT AGGAGGAATC
3481 AAATTAGGTC AATCTCCTTC TAATTCAACA GAAGAAAAAC CAAAGCTTTC ATTAGGTGGT
3541 ATCAAACTTG CACAATCTCC ATCAAATGAA GAGAAGCCAA AACTTTCTCC AAATCTTCCT
3601 AAATCTCCAT CAAATCAACA AACAGAAGAA AAGCCAAAAC TCCAATTAGG TGGACTGAAA
3661 TTAAACTTAG GAAGTAAATC ACAGACAGAA GAAAAGCCAA AACTTAGTGA AGGCGGAATC
3721 AAATTAGGAA ATGTTTCTTC AAGTCAAACT TCTGATGAAA AACCAAAACT TGGTTTAGGA
3781 GGAATTTCTC TTAACTTCGG AAATAAACAG CAAACAGAAG AAAAACCAAA ATTATCTTCA
3841 TCACAAAACG TTGAGAAACC AACACTTGGA TTAGGTGGAA TCACATTAGG GCAACAAACA
3901 TCAGAAGAGA ACAAACCAAA ACTTGGATTA AATCTTCCTA AATCAAATCA AACAGAAGAA
3961 AAACCGAAAT TAGGAACAGG TGGAATTTCA CTAAACTTAG GAAATAAACC ACAATCAGAA
4021 GAGAAACCAA AGCTTCAATT AGGTGGAATT AAACTCGCAC ATTCTCCATC AAATCAATCA
4081 ACTGAAGAGA AGCCAAAACT TTCTTTAAAT CTTGGTAAAT CTCCATCAAA TCAATCAACT
4141 GAAGAGAAGC CAAAACTTGG ATTGGGAGGA CTGGAAACTA AACCTTGGAA ATCATCTCCA
4201 TCAGATGAAA AACCAAAGTT CAGTCTTGGA GGAATGAAAC TCGGAAAAGT TCCTTCAAAC
4261 GAATCACAAG AGCCACCGAA ATTCACTCTA AACATTCCTA AACTTCCTTC AAACCAAGCT
4321 GAACAAAATG ATGGAAAACC AAAGCTTGGA TTAGGCGGAA TCTCTTTGAA TCTTTCGCAG
4381 AAGAAAGAAG AAAATGAAGA GAAGCCAAAA CTCGGAATTG CTCCAAAACC AATTCAATCA
4441 AATCAAAATA AAATTGAAGA AAAACCAAAA GTGACAAATT CTGTCACAAA AACAGAGGAA
4501 AAAGCAATTT CCGCACCAGG AGTCAAGCTT GATTTAAAGC TTCCTAAGAA AGGATTTGGA
4561 TTAATGAGAC AACCATCTAA ATCATTAGAA GTTTCAAATG ATTCTGAACA ACAAAATTTG
4621 TCAAGTTTCG GACAAATCCA CGTAACTCTT CCAGAAAAGA AAGAAGAGAA CCAACCATTG
4681 TCTAGCTTTG GTCAGATCCA TGTAACTCTT CCAGAGAAAA CACAAAAGAA AGAAGAGAAT
4741 CAGCCATTAT CCTCTTTCGG ACAAATTCAT GTTACATTGC CTCAAAAAGG TCAAAATGAA
4801 AAAGTGACAG AACAAAAAGT GACAGATGAA AAATCGTTAT CCTCATTTGG ACAAATACAT
4861 GTAACAATTC CAGAGAAACT AGCAAATAAA CAAGAAGAAA AGCCACTTTC TACATTTGAA
4921 CAAATTCACG TTTCAATTCC TGAAAAAACA CAAAAACAAG AAGAAAAAAT TTCCAGTTTT
4981 GGACAAATCC ATGTTTCAAT TCCTGAAAAA CTCCAGAAAC AAGAAGAAAA GACATCATCA
5041 TTTGGCCAAA TACATGTAAC AATTCCTGAA AAACTCCAGA AACAAGAAGG AAAATCAGTA
5101 ATAGATAAAC TTCCAGAAAA AACTCAGGAA TTAATTAAAA ACATGCCTTC AGAAATAAAT
5161 GTCAAGATTC CAGAACAAGT TCACTCACAA CAAAGTAATC CATTAGTTCC TTTGGGTGTT
5221 TCCTTAAATG ATGTTGCTAA ATCCCAAAAT CCAACACAAA ATAAACCAGA TTTAAGTTCA
5281 TTGAAACCAC CACAAAAATC AGAAGAAAAC AAACCATCAT TAGGACTTAA ATTACCAAAT
5341 AAACCTAATT TCAGTTCATT AAAACCACCA CAGAGTCAAA CAGAGTCAAA ACCAAATAAT
5401 TTAAGTTCAT TGTTACCTCC AAAACAAACA TTAGGACTTA AACTTCCAAC TTCATCTAAT
5461 TCAGGTCTTA AATTACCAAT TAAACTTCCA AACCAAAACC GAAACCAAAA CCAAGAAGCT
5521 AAAGAAGTTA AAGAAACTAA AGAAGTTAAA GCTAAAGAAG AACCAAAGAA AGAACTTGGC
5581 TTCAAGTCCG ATTTCAAGAT CGAAGGTGCA AGTTTCAAAG GCATCAATCT TAGTTCATCA
5641 GGCAAATTCG AAGTTCCAAG TGGAAAAAGT TTTGATGACA TGATCAAACA AAAAGTTAAT
5701 AAAGCAGAAG AAAACAAACC AAAATTCAGT TTTAACTTGA ACTTAAATCT GAACAATAAA
5761 ACACCACTTT CTAATGCACC GAAACCAATT GCACTTTCTC TTTCAAATGA TGGAAAGAAA
5821 CCAAGTTTGG CCGGATTAAA TATTAGCAAA GGATCATCTC TTGCATCTGT ATCAATCAAC
5881 GCATTGGACA AATCAGGAGA TGACGTTCCA AAACTCGCTC TTTCCATTGA CAAACTTAGA

```
5941 AAGGAACCTC CGCCACCTCC TACTCCTCCT GCTGTGACTA AACCAGCTCT TGCAACAGAG
6001 AAAGAGATAA GAGAATCATC AATTGCTGAT GTAACTCCTT TGGCAATGAT TACTAAGATA
6061 AATTCAGCTA AGAAACCTGA TTTCAGTGCA TTGAAACCAA ATCTAGGAAA CAAACAACAG
6121 AGTTCTTCTT CTAATGAACA AAAACCACTC AACAAACCTA ATTTTAGTGC ATTAAAACTA
6181 CCACAGAAAT CAGAAGAAAC AAAACCAAAT AACTTAAGTT CATTGTTACC TCCAAAACAA
6241 ACATTAGGAC TTAAACTTCC AACTTCATCT AATTCAGGTC TTAAATTACC AATTAAACTT
6301 CCAAACCAAA ACCAAAACCA AAACCAAGAA GTTAAAGAAT CTAAAGAAGT TAAAGCTAAA
6361 GAAGAACCAA AGAAAGAACT TGGCTTCAAG TCCGATTTCA AGATCGAAGG TGCAAGTTTC
6421 AAAGGCATCA ATCTTAGTTC ATCAGGCAAA TTCGAAGTTC CAAGTGGGAA AAGTTTTGAT
6481 GACATGATCA AACAAAAAGC GAAAACAGAA CAACCAAAAA CAGAAGAAAA CGAACCAAAA
6541 TTCAGTTTTA ACTTGAACTT AAATCTGAAC AATAAAACAC CACTTTCTAA TGCACCGAAA
6601 CCAATTGCAC TTTCTCTTTC AAATGATTCA AACAAAGGAC AAGTGCAAGG TGCATTGACA
6661 ATTAAACCAA TTCCTCTTTC AAAAGGAGAT AAACCAAAAG CAGATTTTCA GTCACTTAAA
6721 TTAGGAAAAT CTCTTTCAAA TCCAACACAA AATAAACCAG ATTTAAGTTC ATTGAAACCA
6781 CCACAAAAAT CAGAAGAAAA CAAACCATCA TTAGGACTTA AATTACCAAA TAAACCTAAT
6841 TTCAGTTCAT TAAAACCACC ACAGAGTCAA CAAGAGTCAA AACCAAATAA CTTAAGTTCA
6901 TTGTTACCTC CAAAACAAAC ATTAGGACTT AAACTTCCAA CTTCATCTAA TTCAGGTCTT
6961 AAATTACCAA TTAAACTTCC AAACCAAAAC CAAGAAGCTA AAGAAGTTAA AGAAACTAAA
7021 GAAGTTAAAG CTAAAGAAGA ACCAAAGAAA GAACTTGGCT TCAAGTCCGA TTTCAAGATC
7081 GAAGGTGCAA GTTTCAAAGG CATCAATCTT AGTTCATCAG GCAAATTCGA AGTTCCAAGT
7141 GGTAAAAGTT TTGATGACAT GATCAAACAA AAAGCGAAAA CAGAACAACC AAAAACAGAA
7201 GAAAACAAGC CAAAACTTCA GTTGAACCTT AATGGAAATG GTCTTCCGAA GCCTGGACTT
7261 CCAAAACCTC TGCCTTTCAA CAAGCCAGGT ATTCAGAGGT TGAATTTCTC TGCTATTAAT
7321 TTAAAGAAAT AAAGTAAAAA AGAGTTTTTT ATATTATTAT TTTCACATTT CAAAAACTAT
7381 TTTTTATATA CATTCTAAAT TATATTACAA TATCATTGAT CAGTAGGAAA TGCATGATAA
7441 CATGATTATT TAAATCAAAA TAACATTTTA AAGAGAGATT CTCAAACTTT CAAGCAGAAA
7501 CTGCAAAAAA ATTATGATAA TTCGGATAGA TATATGCATA TGCATTAGTT TATCATGAAA
7561 TTTGACTAGT ATATCAGGTC ATTGAGTAAT GCTGTAAAAA TTTAGAAATT GTAATTATTT
7621 TTACTTTAAG TTAAGATAAA AGTATAGTAA ACTACTACTA GATAGAGAAA ATGGCGGCTG
7681 ACATTTGTTA GATTCTGGAT ATTGAGAAGA AAATATATAG GTTATATTAT TTTCTTTGAA
7741 ATCAGCTCTT TCAAAAGCTA AAGTTGGTGC AATGCTTAGT TTCATCTGTC CAAAACTTCT
7801 AAAAACAATA ATCTCTTTTA TCTTGCTCCC AACATCATAT AAGAAAGCAG GGTCAATTGT
7861 CGAAACTGAA TCAGGAATGA TAATTCTCTC TAAATAGACA AATCGTGTCA TTGAAACTTG
7921 AAATAATTTA ACTACGGTGT TTGAAATTAT ATAAGTGGTA TCTTTTTTGA ATCTTGGATA
7981 AGCAATTAAA GTCTTCATAT CCTTGGAAAA TATTACACCA TTTTGCGAGC AAAAACATTT
8041 ATTATCTTTA TCCACTTCAA ATTCTTCTAC ATGCGATTGA TCAAATGGAT TATCTATAAA
8101 TTTAGTAACA TTTTTTGGAA TAAATAGATG TTTCAAAAAG GATGTGAATA AACATCCTTC
8161 GTTATCAATA GTAGTAAGCG TTTCAGGAAG AGACACATTT ATTAAGTTCT TAAAGTTATA
8221 AAATGTGTTA GCAGGGAGGA ATT
```