CLONING AND CHARACTERIZATION OF A SINGLE-STRANDED DNA-BINDING PROTEIN OF *Leishmania major*

by

JOHN ROBERT WEBB

B.Sc., The University of Western Ontario, 1986
M.Sc., The University of Western Ontario, 1988

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Genetics Programme)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
SEPTEMBER 1993
© John Robert Webb, 1993
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Medical Genetics

The University of British Columbia
Vancouver, Canada

Date Oct 7 / 1993
Protozoan parasites of the genus *Leishmania* are the etiologic agents of a spectrum of important human diseases collectively referred to as leishmaniasis. The major surface protein on all species of *Leishmania* is a highly abundant 63 kDa glycoprotein referred to as GP63. GP63 has been characterized as a cell surface protease, however, its exact role in the *Leishmania* life cycle is not clear. The genes encoding GP63 are arranged in the *Leishmania* genome as a species-specific combination of direct head to tail tandem repeats and single dispersed gene copies. In the present study, a single repeat unit of the *Leishmania donovani* GP63 tandem array was cloned and sequenced. Alignment of the *L. donovani* GP63 gene sequence with the previously determined GP63 gene sequences from two related species, *L. major* and *L. chagasi*, revealed that GP63 is highly conserved across species. Consistent with the observed protease activity of GP63, the predicted amino acid sequence of GP63 from all three species contained a conserved motif shared by a number of zinc metalloproteases. In addition, alignment of the untranslated regions of the three GP63 genes revealed that the immediate 5' untranslated region is highly conserved within and across species. This region did not contain any sequences characteristic of higher eukaryotic promoter elements, however, it did contain an area of conserved hexanucleotide direct repeats. To determine whether these direct repeats (CTCGCC) represented a potential site of protein-DNA interaction, a λgt11 expression library of *L. major* was screened with a radiolabelled oligonucleotide probe to detect clones expressing functional DNA-binding proteins. A gene was isolated which encoded a novel DNA-binding protein, referred to as HEXBP. The deduced amino acid sequence of HEXBP revealed that it is a 28 kDa protein containing nine 'CCHC-type' zinc finger
Abstract

motifs. The CCHC motif, Cys-X₂_Cys-X₄-His-X₄-Cys, is invariant with regards to the number and spacing of cysteine and histidine residues and is shared by a number of nucleic acid-binding proteins. In accordance with the activity exhibited by other CCHC-containing proteins, HEXBP was characterized as a single-stranded nucleic acid-binding protein. Additional analyses indicated that HEXBP bound single-stranded DNA in a sequence specific manner and that the conserved 5' untranslated region of GP63 gene contained multiple HEXBP binding sites.

To determine the cellular function of HEXBP, a HEXBP-deficient mutant of *L. major* was generated using the technique of double homologous gene replacement. Initial characterization of this mutant suggested that HEXBP was not essential for the expression of GP63 by *in vitro* cultivated promastigotes. Although the HEXBP-deficient mutant did not exhibit any gross phenotypic changes, further characterization will likely provide insight into the function of the HEXBP single-stranded DNA-binding protein.

In addition, a plasmid construct was identified that lead to stable transformation of *Leishmania* when electroporated into promastigotes as an intact circular plasmid. This construct conferred selectable drug-resistance to transfectants and was found to replicate as an extrachromosomal circular concatamer. The construct was modified to produce a functional *Leishmania* expression vector called pLEX. Initial characterization of the transcriptional regulation of pLEX suggests that it also represents a potentially useful model system for studying the process of polycistronic gene expression in kinetoplastid protozoans.
TABLE OF CONTENTS

ABSTRACT..................................................................................................................ii
TABLE OF CONTENTS .................................................................................................iv
LIST OF TABLES ..........................................................................................................viii
LIST OF FIGURES ......................................................................................................ix
LIST OF ABBREVIATIONS .......................................................................................xii
ACKNOWLEDGMENTS ..............................................................................................xiv

I. INTRODUCTION .....................................................................................................1
   A. LEISHMANIASIS-AN OVERVIEW .................................................................1
   B. GP63 - THE MAJOR SURFACE GLYCOPROTEIN ....................................11
   C. GENE EXPRESSION IN KINETOPLASTID PROTOZOANS ......................19
   D. THE PRESENT STUDY .....................................................................................24

II. MATERIALS AND METHODS .............................................................................26
   A. LEISHMANIA .................................................................................................26
      1. Leishmania strains used in the present study .........................................26
      2. In vitro maintenance of Leishmania promastigotes ..................................26
   B. BACTERIAL STRAINS, VECTORS AND MEDIA .......................................27
      1. Bacterial Strains ......................................................................................27
      2. Vectors ....................................................................................................27
   C. DNA ISOLATION ..........................................................................................28
      1. Isolation of Plasmid and Phage DNA .....................................................28
      2. Isolation of Leishmania Genomic DNA ..................................................28
   D. RNA ISOLATION ..........................................................................................29
### Table of Contents

1. Isolation of *Leishmania* RNA ................................................. 29

---

E. PROTEIN ISOLATION ............................................................. 30

1. Extracts of $\lambda$ gt11 Lysogens .......................................... 30
2. Extracts of pET-3a Clones .................................................. 30
3. Total Cell Extracts of *Leishmania* Promastigotes ...................... 31

---

F. GEL ELECTROPHORESIS .......................................................... 31

1. Non-denaturing Agarose Gel Electrophoresis ............................. 31
2. Southern Blot Hybridization Analysis ..................................... 31
3. Formaldehyde Agarose Gel Electrophoresis and Northern Blot
   Hybridization Analysis ...................................................... 32
4. SDS-polyacrylamide Gel Electrophoresis and Western Blotting ......... 33
5. Denaturing Polyacrylamide Gel Electrophoresis ........................... 33

---

G. GENERAL MOLECULAR BIOLOGY TECHNIQUES ............................... 35

1. Restriction Enzyme Digestion and Preparation of DNA fragments
   for Subcloning ..................................................................... 35
2. Ligation and Transformation of Bacteria .................................... 36
3. Polymerase Chain Reaction .................................................... 37
4. Radioactive Labeling of DNA .................................................. 37
5. Preparation of Single-stranded DNA Using $\lambda$ Exonuclease .......... 38

---

H. DNA SEQUENCE ANALYSIS ......................................................... 38

1. Preparation of Overlapping Deletion Clones ............................. 38
2. Sequencing of Single-Stranded Templates .................................. 39
3. Sequencing of Double-stranded Templates .................................. 39
4. Sequencing of PCR Products .................................................. 39
Table of Contents

5. Preparation of A + G Chemical Sequencing Ladders ..................40

I. LIBRARY SCREENING..............................................................40

1. Screening for λgt11 Clones Expressing Functional DNA-
   binding Proteins........................................................................40
2. Preparation and Screening of a Size-selected Leishmania 'sub'-
   library .................................................................................41

J. DNA-BINDING ASSAYS ............................................................42

1. Electrophoretic Mobility Shift Assay ........................................42
2. Southwestern Blot Analysis......................................................42
3. UV Cross-linking Analysis .....................................................43
4. DNase I Protection Assays ....................................................43

K. TRANSFORMATION OF LEISHMANIA........................................44

1. Electroporation of Leishmania Promastigotes .........................44
2. Cloning of Leishmania Transfectants ....................................45

III. THE GP63 GENE OF Leishmania donovani ............................47

A. RESULTS ................................................................................47

1. The Sequence of the Leishmania donovani GP63 gene ..............47
2. Arrangement of the L. donovani GP63 Gene Locus .................51
3. Heterogeneity Within the Propeptide-Coding Region of the L.
   donovani GP63 Genes ..............................................................54

B. DISCUSSION ..........................................................................66

IV. THE HEXBP GENE OF Leishmania major ..............................72

A. RESULTS .................................................................................72

1. Library Screening .................................................................72
## Table of Contents

2. The Sequence of the *Leishmania major* HEXBP Gene .......................... 74
3. Detection of HEXBP mRNA in Diverse Species of *Leishmania*.............. 76
4. The single-stranded DNA-binding Activity of HEXBP......................... 76
5. DNAse I Footprint Analysis of HEXBP Single-stranded DNA-Binding Activity ................................................................................. 81

### B. DISCUSSION.............................................................................. 107

#### V. ANALYSIS OF *Leishmania major* HEXBP-DEFICIENT MUTANTS GENERATED BY DOUBLE TARGETED GENE REPLACEMENT.......................... 117

#### A. RESULTS ............................................................................. 117

1. Replacement of the First HEXBP Allele........................................... 117
2. Replacement of the Second HEXBP Allele....................................... 119
3. Characterization of the HEXBP Deletion Mutant CC1-14-4/D............. 120

#### B. DISCUSSION.............................................................................. 131

#### VI. DEVELOPMENT OF AN EXPRESSION VECTOR FOR THE STABLE TRANSFECTION OF *LEISHMANIA* AND APPLICATION TO THE STUDY OF POLYCISTRONIC GENE EXPRESSION ................................... 139

#### A. RESULTS ............................................................................. 139

1. Transfection of *Leishmania* with the Circular Plasmid Construct pMHBH\text{Hyg}.................................................................................. 139
2. Modification of pMHBH\text{Hyg} to generate the *Leishmania* expression vector pLEX. ................................................................. 142

#### B. DISCUSSION.............................................................................. 153

#### VII. SUMMARY AND FUTURE EXPERIMENTS............................. 158

#### VIII. REFERENCES ....................................................................... 161
LIST OF TABLES

Table 1. Clinical Manifestations and Geographic Distribution of Leishmaniasis........10
Table 2. Oligodeoxyribonucleotides used in the present study..............................46
LIST OF FIGURES

Figure 1. The nucleotide and predicted amino acid sequence of the *L. donovani* GP63 gene.................................................................58

Figure 2. Southern blot hybridization of *L. donovani* genomic DNA.........................60

Figure 3. Genomic organization of the *L. donovani* GP63 gene array.........................61

Figure 4. Restriction enzyme mapping of the lambda clones λ LdGP63-7-4 and λ LdGP63-8-4........................................................................62

Figure 5. Analysis of the propeptide-coding region of *L. donovani* GP63 genes by PCR amplification and Southern blot hybridization.......................64

Figure 6. Analysis of the GP63 propeptide-coding regions in λ LdGP63-7-4 and λ LdGP63-8-4 by Southern blot hybridization........................................65

Figure 7. Sequence alignment of the GP63 gene 5' untranslated region.......................88

Figure 8. Nucleotide and predicted amino acid sequence of the *L. major* DNA-binding protein HEXBP........................................................89

Figure 9. Comparison of proteins containing the CCHC zinc finger motif..................90

Figure 10. Analysis of HEXBP gene expression in diverse species of *Leishmania*......92

Figure 11. Electrophoretic mobility shift assays of HEXBP single-stranded vs. double-stranded DNA-binding activity.........................................93

Figure 12. Sequence specificity of HEXBP binding to single-stranded oligodeoxyribonucleotides................................................................94

Figure 13. Comparison of the DNA-binding activity of bacterially synthesized fusion and non-fusion HEXBP by Southwestern blot analysis...............95
Figure 14. Competitive gel mobility shift assay of non-fusion HEXBP DNA-binding activity.................................................................96

Figure 15. Stability of HEXBP single-stranded DNA-binding activity at increasing ionic concentrations.................................................................97

Figure 16. Detection of HEXBP DNA-binding activity in *L. major* promastigote extracts.................................................................98

Figure 17. DNAse I protection of gp63-5'-50(-) by HEXBP.................................99

Figure 18. DNAse I protection of the BSHEX-327 and BS-374 single-stranded probes by HEXBP.................................................................100

Figure 19. Binding of the gp63-5'-462 single-stranded probe by HEXBP in electrophoretic mobility shift assays.................................................................102

Figure 20. DNAse I protection of the gp63-5'-462 single-stranded probe by HEXBP - I.................................................................103

Figure 21. DNAse I protection of the gp63-5'-462 single-stranded probe by HEXBP - II.................................................................105

Figure 22. Restriction maps of the HEXBP gene locus and plasmid constructs used in homologous gene replacement experiments.................................................................123

Figure 23. Southern blot hybridization analysis of Hyg\textsuperscript{R} transfectants.................................125

Figure 24. Southern blot analysis of Hyg\textsuperscript{R}/Neo\textsuperscript{R} transfectants.................................................................126

Figure 25. Detection of a HEXBP-related sequence by genomic Southern blot analysis.................................................................127

Figure 26. Northern blot hybridization analysis of HEXBP gene expression in Hyg\textsuperscript{R}/Neo\textsuperscript{R} clones.................................................................128
List of Figures

Figure 27. Northern blot hybridization analysis of GP63 expression in HEXBP-deficient clones..........................129
Figure 28. Western blot analysis of GP63 expression in HEXBP-deficient clones. .....130
Figure 29. Southern blot hybridization analysis of clones transfected with the construct pMHBHyg - I.................................................................146
Figure 30. Southern blot hybridization analysis of clones transfected with the construct pMHBHyg - II.................................................................147
Figure 31. Limited DNAse I digestion of the pMHBHyg extrachromosomal element in L. major transfectants.................................................................148
Figure 32. Restriction maps of the Leishmania expression vector pLEX and its derivatives. .................................................................149
Figure 33. Southern blot hybridization analysis of clones transfected with derivatives of the Leishmania expression vector pLEX ........................................151
Figure 34. Northern blot hybridization analysis of clones transfected with derivatives of the Leishmania expression vector pLEX ........................................152
Figure 35. Schematic model of HEXBP interacting with its binding site in the GP63 gene 5' untranslated region. .............................................................160
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>BCIP</td>
<td>5-bromo-4-chloro-3-indolylphosphate p-toluidine salt</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>DEPC</td>
<td>diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNAse I</td>
<td>deoxyribonuclease I</td>
</tr>
<tr>
<td>DTI</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-[2-hydroxyethyl]-1-piperazine-ethanesulfonic acid</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-D-thiogalactopyranoside</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>MOPS</td>
<td>3-(N-morpholino)-propanesulfonic acid</td>
</tr>
<tr>
<td>NBT</td>
<td>nitroblue tetrazolium chloride</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I wish to express my appreciation and gratitude to my colleagues and co-workers in Rob McMaster's lab, all of whom contributed in some manner to the completion of this thesis. In particular, I wish to acknowledge Dr. L.L. Button, who initially cloned the *L. donovani* GP63 gene, Charlotte Morrison, who contributed significantly to the *L. major* HEXBP project, and Dr. P.J. Joshi who contributed to the expression vector project. Special thanks to my supervisor Rob McMaster for his continuing guidance and support during the course of this project. I would also like to thank the members of my supervisory committee, George Spiegelman, Frank Tufaro and Paul Goodfellow for their advice and encouragement. I gratefully acknowledge the University of British Columbia for financial support in the form of a University Graduate Fellowship.

Last but not least, I wish to thank my wife Susan for her endless support and encouragement and for her motivation during the writing of this thesis. I would also like to acknowledge the contributions of my daughter Carly who shed new light on this research by continually rearranging my autorads during the preparation of figures.
I. INTRODUCTION

A. LEISHMANIASIS-AN OVERVIEW

Protozoan parasites of the genus *Leishmania* are the etiologic agents of a spectrum of human diseases collectively referred to as leishmaniasis. The parasites are transmitted via the bite of blood-feeding insects, specifically, sandflies of the subfamily *Phlebotominae*. Leishmaniasis is prevalent in many tropical and subtropical areas of the world where it has an estimated annual incidence of approximately 400,000 cases/year (Meshnick and Marr, 1992). Although the taxonomy of the genus *Leishmania* is currently in a state of flux due to the relatively recent application of molecular probes to categorize isolates according to species and strain, the organisms can be grouped according to geographical localization and the clinical syndromes with which they are most commonly associated (see Table 1). It is generally accepted that there are three major forms of disease in humans (for review see Chang *et al.* 1985). Cutaneous leishmaniasis causes a non-disseminating cutaneous lesion which is generally self-healing within a few months to a year, leaving the host immunologically protected against subsequent infection. Occasionally, cutaneous leishmaniasis develops into diffuse cutaneous leishmaniasis which is characterized by metastasis of skin lesions over a much larger area. Cutaneous leishmaniasis is endemic in many areas of Southern Europe, Asia and Northern Africa where it is caused primarily by *L. major* and *L. tropica* and to a lesser extent by *L. infantum*. In the New World (Mexico, Central and South America) cutaneous leishmaniasis is caused by members of the *L. mexicana* complex. Mucocutaneous leishmaniasis, or espundia, usually begins with the formation of a single lesion, similar to those observed in cutaneous leishmaniasis, but
secondary lesions subsequently develop, usually in the mucocutaneous membranes of the nasal passages. Left untreated, mucocutaneous leishmaniasis can lead to severe disfigurement through extensive destruction of the lips, palate, nose and pharynx. Mucocutaneous leishmaniasis occurs primarily in the New World where it is caused by members of the \textit{L. braziliensis} group of organisms. Visceral leishmaniasis is found in both the Old and New Worlds and is characterized by extensive systemic infection which eventually leads to hepatomegaly, splenomegaly and severe anemia. The disease is generally fatal if left untreated and most often affects young children under the age of five years. Visceral leishmaniasis is caused by \textit{L. donovani} and \textit{L. infantum} in the Old World and by \textit{L. chagasi} in the New World. There is some debate as to whether \textit{L. chagasi} actually represents a distinct species or whether it is a geographic isolate of an Old World species, hence it is designated in the literature as both \textit{L. chagasi} and \textit{L. donovani chagasi}.

All known species of \textit{Leishmania} are digenetic, spending part of their life cycle as a free-living form within the gut of the insect vector and part as an obligate intracellular parasite within the vertebrate host (for review see Chang \textit{et. al.}1985; Jeronimo and Pearson, 1992). The insect stage of \textit{Leishmania} is an elongated, motile, flagellated form called a promastigote. Promastigotes are amenable to \textit{in vitro} culturing and can be maintained indefinitely in most standard tissue culture media (M199/DMEM + 10% fetal calf serum) at 26°C. The cultures will grow to densities exceeding $1 \times 10^8$ cells/ml and therefore represent an excellent source of working material. For these reasons the promastigote life stage is generally the best characterized in terms of biochemistry and molecular biology. Promastigotes differentiate morphologically and biochemically as they develop within the gut of the sandfly vector and eventually a highly infectious form moves anteriorly to inhabit the proboscis. These changes are reflected \textit{in vitro} as promastigotes
Introduction
devlop from low to high virulence as culture density increases (Sacks and Perkins, 1984; da Silva and Sacks, 1987; Walters et. al.1989; McConville et. al.1992). When infected sandflies harboring virulent promastigotes take a blood meal from an appropriate vertebrate host, the promastigotes are inoculated at the site of feeding. Promastigotes then specifically enter host macrophage cells where they transform into the amastigote form. Amastigotes are a non-motile, non-flagellated form, although they retain what appears in electron micrographs to be a 'micro-flagellum' (Chang et. al.1985). The amastigotes reside within a membrane bound organelle known as the parasitophorous vacuole, which is formed by lysosome-phagosome fusion. Unlike other intracellular pathogens of macrophages (Legionella, Toxoplasma, Mycobacterium and Listeria), Leishmania do not block the formation or acidification of the phagolysosome and one of the unique characteristics of Leishmania is their ability to survive and replicate within the harsh environment of this organelle (Mukkada et. al.1989; Bogdan et. al.1990). Due to the intracellular location of amastigotes they are not as amenable to analysis in the laboratory as the promastigote stage, however, amastigotes can be isolated from infected mice (Glaser et. al.1990) or from macrophage cell lines infected in vitro (Ogunkolade et. al.1990). Alternatively, promastigotes of several species of Leishmania can be transformed into extracellular axenic 'amastigote-like' forms by manipulation of the culture temperature and conditions (Doyle et. al.1991; Bates et. al.1992). The intracellular amastigotes divide by binary fission within the phagolysosome and eventually the host macrophage is lysed, releasing amastigotes which go on to infect other macrophages. The specific tissue tropisms of the various species, which ultimately results in distinct diseases, is not currently understood. Ingestion of infected macrophages or amastigotes by a sand fly completes the life cycle. Generally, animal populations serve as reservoirs for Leishmania and humans are infected as a
secondary host. In particular, rodents, foxes, squirrels and especially dogs have been shown to serve as significant reservoirs for various species of *Leishmania* (Chang *et al.* 1985).

All forms of leishmaniasis are treated by chemotherapy using the pentavalent antimonial compounds Pentostam (Sodium Stibogluconate - Wellcome) and Glucantime (Meglumine Antimoniate - Rhodia) (for review see Marsden and Jones, 1985). Treatment is generally effective although some species respond to treatment better than others. Recurrent disease is common among some species of *Leishmania* and requires secondary treatment with antimonials or amphotericin B. Side effects of treatment are generally dose-related in accordance with the toxicity of antimonial compounds. The development of drug resistant variants is low, however several drug resistant clones have been isolated (Berman *et al.* 1989; Ullman *et al.* 1989). Although the exact mechanism of action is not clear, it has been suggested that pentavalent antimonials inhibit some aspect of glycolysis in the amastigote (Berman *et al.* 1989; Berman and Grogl, 1989; Hart *et al.* 1989). It may be relevant that *Leishmania* and other kinetoplastid protozoans contain a unique organelle called a glycosome which contains all of the enzymes of the glycolytic pathway (Hart and Opperdoes, 1984). Since antimonial compounds were not found to inhibit glycososomal enzyme activity *per se*, it has been speculated that the integrity of the glycosome may be a site of antimonial action (Berman and Grogl, 1989; Hart *et al.* 1989). Recent studies of purine metabolism in *Leishmania* have led to the use of allopurinol as an alternative form of chemotherapy against leishmaniasis (for review see Meshnick and Marr, 1992). This is based on the finding that *Leishmania* only generate purine nucleotides via the purine salvage pathway, they are incapable of *de novo* purine synthesis. In addition, the *Leishmania* purine salvage enzyme hypoxanthine guanine phosphoribosyl transferase is
Introduction

relatively non-specific and readily utilizes allopurinol (an analogue of hypoxanthine) as a substrate. The analogue is then activated to the nucleotide analogue of IMP which is an inhibitor of succinyl-AMP synthase, effectively preventing the formation of AMP and subsequent nucleic acid synthesis.

The nature of the immune response directed against *Leishmania* has been extensively characterized, both in terms of developing a safe and effective vaccine to protect against leishmaniasis and as a model system to study T cell-mediated immunity. As described above, the most common form of leishmaniasis, cutaneous leishmaniasis, is self-healing and cell-mediated immunity is known to play the predominant role in resolution of the disease (for review see Muller *et al.* 1989b). In contrast there is little evidence that humoral immunity has any effect on the outcome of infection. Moreover, *L. major* infections in mice, particularly Balb/c mice, have proven to be a particularly useful model system to characterize host immunological responses against *Leishmania* (for review see Liew, 1989; Muller *et al.* 1989a; Muller *et al.* 1989b; Scott, 1989; Titus *et al.* 1992). Most strains of mice develop small cutaneous lesions at the injection site after syringe infection with virulent *L. major* promastigotes. These lesions generally heal within several weeks rendering protective immunity against subsequent infection. In contrast, Balb/c mice are extremely susceptible to infection with *L. major* and following injection of promastigotes they develop extensive metastatic lesions. The infected mice fail to resolve the infection and eventually succumb to a fatal disseminating disease. The importance of T cells in resistance to murine cutaneous leishmaniasis was demonstrated in early experiments utilizing athymic mutant nu/nu mice (Mitchell *et al.* 1980; Mitchell *et al.* 1982). Athymic mice of both resistant and susceptible strains were found to be extremely susceptible to infection with *L. major*. This susceptibility could be completely reversed in resistant mouse strains by
adoptive transfer of T cells from normal syngeneic mice. Further studies showed that CD4+ T cells from resistant mice that had recovered from infection could confer protective immunity upon naive mice (Liew, 1986). The importance of CD4+ T cells in immunity to murine cutaneous leishmaniasis was further demonstrated by treating mice with repeated doses of anti-CD4 monoclonal antibody which resulted in the development of severe infections in normally resistant mice (Titus et al. 1987). Interestingly, in susceptible strains of mice, CD4+ T cells were also shown to confer susceptibility to infection with L. major, a finding which led to the use of murine cutaneous leishmaniasis as a model system for the analysis of the recently described subsets of CD4+ T cells (for review see Liew, 1989; Scott, 1989; Locksley and Scott, 1991). Isolated clones of these two classes of mouse CD4+ helper T cells differ with regards to the profile of cytokines that they produce. TH1 cells secrete IFN-γ and IL-2 whereas TH2 cells secrete IL-4, IL-5 and IL-10. In resistant strains of mice, infection with L. major leads to the proliferation of cells of the TH1 class and recovery from infection whereas infection in susceptible mice leads to the proliferation of cells of the TH2 class and exacerbation of the disease. The profile of cytokines produced by the two classes of cells correlates well with the progression of the disease since IFN-γ, produced by cells of the TH1 lineage, is a potent activator of macrophages. The role of IL-4, produced by cells of the TH2 lineage, in conferring susceptibility to leishmaniasis is not as clear however it has been suggested that IL-4 functions by inhibiting the activation of macrophages by IFN-γ and by recruiting large numbers of macrophages to the site of infection. The cytokines produced by each class also act by stimulating further proliferation of the same class and inhibiting the proliferation of the opposite class. Although the reasons for inducing the proliferation of one class over the other are complex, it has been suggested that it may be due, at least in part, to differential antigen presentation (Scott et al. 1988).
Therefore, the development of potential recombinant vaccines against leishmaniasis must take into account the complexities of inducing proliferation of different T cell subsets. Although the CD4+ helper T cell dichotomy is applicable to a certain extent to human cutaneous leishmaniasis, T cell subsets are not nearly as definitive as in the mouse model and most human CD4+ T cells exhibit characteristics of both subclasses. In addition, tumor necrosis factor (TNF) has also been shown to play a pivotal role in mediating protection against cutaneous leishmaniasis (Titus et al. 1992).

Although distinct from susceptibility to cutaneous leishmaniasis, the development of visceral leishmaniasis has been proposed to occur through susceptibility/resistance to infection with *L. donovani* (for review see Blackwell et al. 1991). Resistance to infection with *L. donovani* is thought to be controlled by a gene called *Lsh*. Furthermore, it has been suggested that *Lsh* is likely identical to a gene conferring innate resistance to infection by *Mycobacterium* (*Bcg*) since both map to identical positions on mouse chromosome 1 and both control natural resistance to intracellular parasites. Recently, a candidate for the *Bcg* gene was cloned and sequenced which encoded a putative membrane-spanning transporter protein (Vidal et al. 1993). The expression of this protein was restricted to cells of the reticuloendothelial system and was speculated to potentially function in the transport of oxidative intermediates, essential for the killing of intracellular organisms by macrophages.

*Leishmania* are members of the order *Kinetoplastida*, which includes the trypanosomes, the etiologic agents of sleeping sickness in Africa and Chaga's disease in South America. The order is so named because all members contain a unique organelle called the kinetoplast (for review see Simpson, 1987). This organelle is the equivalent of a specialized type of mitochondrion that is greatly distended at one end due to the presence of
a large amount of DNA. The kinetoplast DNA is arranged as a large network of concatenated circular molecules of two classes. Maxicircles (20 to 40 kbp in size) are the functional equivalent of mitochondrial genomes from higher eukaryotes and contain the genes encoding mitochondrial proteins and mitochondrial rRNAs. There are approximately 50 copies of the maxicircle per kinetoplast. In contrast, minicircles (1 to 2.5 kbp in size) are unique to the kinetoplastids. Minicircles exhibit significant size and sequence heterogeneity within a single species (approximately 300 different classes in *Trypanosoma brucei*) and several thousand copies are usually present per kinetoplast. Until recently the function of the minicircle was an enigma, however they are now known to be involved in the process of RNA editing (which will be discussed more extensively in a subsequent section describing gene expression in kinetoplastid organisms). Other members of the order *Kinetoplastida* include *Crithidia*, which infects only insects, and *Leptomonas*, which is a parasite of invertebrates.

The genetics of kinetoplastid protozoans are poorly understood in comparison to other lower eukaryotic organisms such as *Saccharomyces* and issues such as the ploidy of the organisms have only recently been addressed. The chromosomes of kinetoplastids do not condense at any stage of the mitotic cycle and therefore cannot be directly analyzed by conventional cytological methods. The only means of characterizing the chromosome content is by pulsed field gel electrophoresis (PFGE) techniques which have been used to identify at least 20 different chromosomes in *Leishmania* (for review see Bastien *et al.* 1992; Lighthall and Giannini, 1992). However, the *Leishmania* genome exhibits a remarkable degree of plasticity, and chromosome size polymorphisms are common between various species and strains (Bishop, 1990). Hybridization studies with chromosome-specific probes demonstrated that at least some of the chromosomes are
diploid, however chromosome identification is often confused by size polymorphisms between homologous chromosomes. The molecular mechanisms responsible for extensive size polymorphisms are generally not well understood, however, they are usually attributable to intrachromosomal amplification and deletion events (Iovannisci and Beverley, 1989; Bastien et. al.1990). In particular, the sub-telomeric 'barren' regions of chromosomes have been shown to be especially conducive to genetic rearrangement (Bastien et. al.1992). Probably the best evidence for diploidy in the Leishmania genome comes from recently developed deletion mutants generated by homologous gene replacement. Deletion mutants are obtained after two rounds of gene replacement, implying the presence of two genomic alleles (Cruz et. al.1991). However, the unstable nature of the genome is again manifested in the generation of aneuploidy and tetraploidy in some mutant cell lines (Cruz et. al.1993). Reproduction in Leishmania is generally assumed to be clonal, however, recent studies report evidence of sexual recombination (Pages et. al.1989; Kelly et. al.1991). Evidence supporting genetic exchange via sexual recombination has also been reported in the Trypanosomes (Jenni et. al.1986; Paindavoine et. al.1986).
Table 1. Clinical Manifestations and Geographic Distribution of Leishmaniasis
(adapted from Jeronimo and Pearson, 1992)

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Parasite</th>
<th>Geographical Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visceral leishmaniasis (kala azar)</td>
<td><em>Leishmania donovani</em></td>
<td>Indian subcontinent, China</td>
</tr>
<tr>
<td>-General involvement of the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reticuloendothelial system</td>
<td><em>Leishmania infantum</em></td>
<td>Middle East, Africa, China</td>
</tr>
<tr>
<td>(spleen, liver, bone marrow)</td>
<td><em>Leishmania chagasi</em></td>
<td>Latin America</td>
</tr>
<tr>
<td>New World Cutaneous leishmaniasis</td>
<td><em>Leishmania mexicana</em></td>
<td>Mexico, Central America</td>
</tr>
<tr>
<td>-single or a limited number of</td>
<td><em>Leishmania amazonensis</em></td>
<td>Amazon basin</td>
</tr>
<tr>
<td>skin lesions</td>
<td><em>Leishmania pifanoi</em></td>
<td>Venezuela</td>
</tr>
<tr>
<td></td>
<td><em>Leishmania braziliensis</em></td>
<td>Brazil, Peru, Ecuador, Bolivia, Paraguay, Argentina</td>
</tr>
<tr>
<td>New World Mucocutaneous leishmaniasis</td>
<td><em>Leishmania braziliensis</em></td>
<td>Multiple areas of South America</td>
</tr>
<tr>
<td></td>
<td><em>Leishmania panamensis</em></td>
<td>Central America</td>
</tr>
<tr>
<td>Old World Cutaneous leishmaniasis</td>
<td><em>Leishmania major</em></td>
<td>Middle East, Central Asia, Africa, Indian Subcontinent</td>
</tr>
<tr>
<td>-single or a limited number of</td>
<td><em>Leishmania tropica</em></td>
<td>Middle East, West Asia, Indian subcontinent</td>
</tr>
<tr>
<td>skin lesions</td>
<td><em>Leishmania aethiopica</em></td>
<td>Ethiopian highlands</td>
</tr>
</tbody>
</table>
**Introduction**

**B. GP63 - THE MAJOR SURFACE GLYCOPROTEIN**

The two most abundant molecules on the surface of *Leishmania* promastigotes are the major surface glycoprotein, GP63, and lipophosphoglycan (LPG). LPG is a heterogeneous glycoconjugate containing a variable number of repeated phosphorylated disaccharides of galactose and mannose linked to a heptasaccharide core (for review see Turco and Descoteaux, 1992). The entire structure is attached to the cell surface membrane via a glycosyl-phosphatidylinositol (GPI) linkage (for a review of the GPI structure see Ferguson and Williams, 1988). LPG has been implicated in a wide range of cellular processes including attachment and entry into the host macrophage cell and survival within the phagolysosome (Turco and Descoteaux, 1992). As described above, promastigotes grown *in vitro* become increasingly infective as they progress from log to stationary phase of growth. The increased level of virulence of stationary phase cells has been correlated with modification of LPG structure (McConville *et. al.*1992). It has been reported that the structurally modified form of LPG present on infectious stationary phase promastigotes mediates resistance to complement lysis while at the same time activating deposition of C3b on the parasite surface (Turco and Descoteaux, 1992). One of the ways in which promastigotes reportedly gain entry into the macrophage is via binding and internalization of surface bound C3b by macrophage complement receptor 1 (CR1) (DaSilva *et. al.*1989). LPG continues to be detectable for at least 48 hours after entry into the macrophage and the dense coat of LPG likely forms a protective barrier between the surface of the parasite and the digestive enzymes of the phagolysosome. In addition, LPG has been shown to inhibit lysosomal enzymes *in vitro* and to dampen the effect of macrophage activation (Turco and Descoteaux, 1992). At present it is not known whether amastigotes continue to synthesize
LPG. It has been estimated that there are approximately $1.25 \times 10^6$ copies of LPG on the surface of the parasite, covering approximately 25% of the total surface area (Orlandi and Turco, 1987).

Second in abundance to LPG is GP63, the major surface glycoprotein of *Leishmania*. It has been estimated that there are approximately $5 \times 10^5$ copies of GP63 on the promastigote surface, representing approximately 0.5 to 1% of the total parasite protein (Bordier, 1987; Schneider *et al.* 1993). GP63 was identified as the major surface protein of *Leishmania* by several different groups during early surface iodination experiments (Ramasamy *et al.* 1983; Bouvier *et al.* 1985; Colomer-Gould *et al.* 1985; Etges *et al.* 1985). As its name implies, it is a glycosylated protein, and is reported to contain mannose, N-acetylglicosamine and N-acetylgalactosamine (Russell and Wilhelm, 1986; Olafson *et al.* 1990). The molecular weight of the mature, cell-surface form of GP63 was estimated by SDS-PAGE to be approximately 63 kDa, but when chemically deglycosylated or synthesized in the presence of the glycosylation inhibitors, its molecular weight was reduced to approximately 54 kDa (Bordier, 1987). Like LPG, the protein is attached to the surface of the parasite by a glycosyl-phosphatidylinositol linkage but the phosphatidylinositol portion is slightly modified as compared the GPI of LPG (Schneider *et al.* 1990).

GP63 has been detected on the surface of promastigotes from all known species of *Leishmania* as well as some related non-pathogenic kinetoplastids (Bouvier *et al.* 1987; Etges, 1992; Inverso *et al.* 1993). Although there are conflicting reports in the literature regarding the expression of GP63 during the amastigote life stage, several groups have detected GP63 expression in amastigotes at either the RNA or protein level, albeit at reduced levels compared to promastigotes (Colomer-Gould *et al.* 1985; Chang *et al.* 1986;
Introduction

Button et al. 1989; Medina-Acosta et al. 1989; Frommel et al. 1990). In addition, the GP63 protein expressed in the amastigote stage of *L. major* appeared to have a slightly higher molecular weight than the protein expressed in promastigotes (Frommel et al. 1990). Similarly, GP63 from *L. mexicana* amastigotes was shown to be structurally distinct from promastigote GP63 and evidence suggested that the majority of amastigote GP63 was not accessible to surface radioiodination nor did it contain a GPI anchor attachment (Medina-Acosta et al. 1989). The latter finding may be related to the recent discovery of structurally distinct GP63 genes in *L. mexicana* (Medina-Acosta et al. 1993b) which is discussed in greater detail below. Conversely, Schneider et al. (1992) reported finding dramatically reduced synthesis of GP63 mRNA and protein in *L. major* amastigotes. Interestingly, a GP63 gene homologue has recently been identified in the related kinetoplastid protozoan *Crithidia fasciculata*, which is a monogenetic parasite with no vertebrate host (Inverso et al. 1993). This finding suggests that GP63 is particularly relevant to survival in the gut of the insect vector.

Although the exact biological function of GP63 remains to be determined, it has been implicated in a wide range of processes. A large number of studies have suggested that GP63 is either directly or indirectly involved in receptor-mediated uptake of promastigotes by host macrophages (for review see Bordier, 1987). In particular, the CR1 and CR3 complement receptors seem to be critical for promastigote uptake (Russell and Wilhelm, 1986; Da Silva et al. 1989). Uptake via the complement receptors would correlate with initial intracellular survival since these receptors are reported to be poor in triggering a respiratory burst (Bogdan et al. 1990). However, the process of parasite binding and uptake is complex and likely involves both GP63 and LPG as well as serum components (complement in particular) and multiple receptors. The observed correlation between levels
of GP63 expression and virulence (Wilson et al. 1989; Liu and Chang, 1992) suggests that GP63 may be essential for uptake of promastigotes by macrophages, however, a definitive causal relationship remains to be established.

GP63 has also been demonstrated to have a protease activity (Etges et al. 1986; Chaudhuri and Chang, 1988; Bouvier et al. 1989; Chaudhuri et al. 1989; Ip et al. 1990) however there are conflicting reports in the literature regarding the pH optimum for activity. Chaudhuri and Chang (1988) reported that the proteolytic activity of GP63 was optimum at pH 3.0 to 4.0 and suggested that GP63 mediated protection of the parasite within the phagolysosome by inactivating host degradative enzymes. Conversely, Etges et al. (1986) and Ip et al. (1990) described the pH optimum of GP63 as being neutral to slightly alkaline, implying that it likely has important functions outside of the host macrophage phagolysosome.

Button and McMaster (1988) were the first to report the cloning and sequencing of a complete GP63 gene from Leishmania. The gene was isolated by screening an L. major genomic library with a degenerate synthetic oligodeoxynucleotide, synthesized on the basis of amino-terminal sequence data obtained from GP63 isolated from the surface of L. major promastigotes. Several different classes of positive clones with unique restriction maps were isolated, implying that the protein was encoded by a multi-gene family. The complete nucleotide sequence of one class of GP63 gene was determined and was found to contain a single open reading frame encoding a protein with a predicted amino acid sequence 602 residues in length. Amino acid sequence analysis of GP63 purified from the surface of L. major promastigotes indicated that the amino terminus of mature GP63 corresponded to residue 101 of the predicted amino acid sequence. This finding implied that GP63 was synthesized as a precursor protein with a 100 residue amino-terminal extension. Consistent
with reports describing GP63 as a cell surface protease, this 100 residue amino-terminal extension was reported to contain both a prepeptide, or transport signal region, and a propeptide, or protease regulatory region. As described above, GP63 is covalently attached to the cell surface membrane via a glycosyl-phosphatidylinositol linkage and a hydrophobic carboxyl-terminal region of the L. major GP63 protein was predicted to be cleaved off during attachment of the GPI moiety. The precise cleavage site for GPI attachment has since been determined by carboxy-terminal amino acid sequence analysis of mature GP63 isolated from the surface of L. major promastigotes (Schneider et. al.1990). After completion of the described post-translational cleavages, mature GP63 from L. major was predicted to have a molecular weight of 51 kDa, however, mature GP63 isolated from the surface of promastigotes has a molecular weight of approximately 63 kDa. Glycosylation at the two potential asparagine-linked glycosylation sites is assumed to account for the observed molecular weight difference (Button and McMaster, 1988). Alignment of the predicted L. major GP63 protein sequence with other known protease sequences revealed the presence of a structural motif that is shared by a number of zinc endopeptidases (Bouvier et. al.1989). This motif represents the zinc-binding domain and catalytic site of zinc metalloproteases. The zinc metalloprotease activity of GP63 has since been well characterized and although the in vitro cleavage of artificial substrates has been shown to occur in a site specific manner (Bouvier et. al.1990; Ip et. al.1990) the natural substrate of GP63 has not yet been determined. In addition, it has been suggested that a single cysteine residue located in the putative propeptide region plays a regulatory role in L. major GP63 metalloprotease activity (Bouvier et. al.1990). This proposed regulatory activity is based on a process known as the 'cysteine switch mechanism' in which a single cysteine residue
binds to and thereby inactivates the zinc-binding site of a zinc metalloprotease when the protein is in the propeptide configuration (Van Wart and Birkedal-Hansen, 1990).

The original predicted sequence of *L. major* GP63 was reported to contain an RGD sequence motif (Asp-Gly-Arg) thought to represent a ligand for the macrophage CR3 receptor on the surface of promastigotes (Button and McMaster, 1988; Ouaissi, 1988). However, the presence of this motif was later found to be attributable to a sequencing error (Button and McMaster, 1990; Miller *et al.*1990). Recently, another GP63 sequence motif thought to be important for macrophage binding has been identified (Soteriadou *et al.*1992). This motif is similar enough to the RGD motif to be immunologically cross-reactive and peptides carrying the motif can effectively compete for uptake of promastigotes by the CR3 receptor of host macrophages.

Southern blot analysis of the *L. major* GP63 gene locus indicated the presence of multiple GP63 genes arranged as a direct head to tail tandem array (Button and McMaster, 1988). A complete restriction map of the locus was subsequently generated which revealed the presence of five GP63 genes arranged in tandem and a sixth gene separated from one end of the tandem array by approximately 8 kbp (Button *et al.*1989). The *L. major* GP63 gene for which the complete sequence was determined corresponded to the first gene of the tandem array. The GP63 locus was mapped to a 700 kbp chromosome by Southern blot analysis of chromosomes separated by pulsed field gel electrophoresis. Chromosomes of identical size were detected in several other species of *Leishmania* using the GP63-specific hybridization probe. Expression of GP63 was characterized by Northern blot hybridization analysis and a prominent 3 kb GP63 transcript was detected in *L. major* promastigote total RNA (Button *et al.*1989). In addition, a minor transcript of approximately 6 kb was observed in total RNA from *L. major* promastigotes. A prominent 3 kb transcript was also
detected in *L. major* amastigote total RNA but was less abundant than in promastigotes. A similar sized transcript was also observed in total RNA from *L. mexicana* promastigotes and amastigotes.

The nucleotide sequences of GP63 genes from several diverse species of *Leishmania* have now been reported and, in general, the characteristics initially described for the *L. major* gene are conserved across species (Miller *et. al.*1990; Webb *et. al.*1991; Ramamoorthy *et. al.*1992; Steinkraus and Langer, 1992; Medina-Acosta *et. al.*1993b). The major exception is the discovery of heterogeneity within the carboxyl-terminal domains of GP63 genes from within a single species. Both *L. donovani chagasi* (Ramamoorthy *et. al.*1992) and *L. mexicana* (Medina-Acosta *et. al.*1993b) were shown to contain distinct classes of GP63 genes which could be discriminated on the basis of sequence divergence within the carboxyl-terminal coding regions. The first class of gene, called 'L1/S1' in *L. donovani chagasi* or 'C2/C3' in *L. mexicana*, was very similar to the reported *L. major* GP63 gene. However, the second class of gene, called the 'constitutive' GP63 gene in *L. donovani chagasi* or 'C1' in *L. mexicana* encoded a protein with a carboxyl-terminal sequence that diverged from the *L. major* sequence immediately upstream of the GPI attachment site. It has been proposed that GP63 proteins containing this 'alternate' carboxy-terminal domain are attached to the surface via a classical transmembrane domain rather than a GPI linkage (Ramamoorthy *et. al.*1992) and that this difference could account for the differential localization of GP63 observed in *L. mexicana* amastigotes (Medina-Acosta *et. al.*1989; Medina-Acosta *et. al.*1993b). In accordance with this proposal, expression of the various gene classes is reported to be developmentally regulated. Although all three GP63 gene classes were expressed in *L. mexicana* promastigotes, transcripts from the C1 class were significantly enriched in amastigotes whereas transcripts
from the C2/C3 gene classes were not detectable in amastigotes (Medina-Acosta et al. 1993b). The expression pattern of the three *L. donovani chagasi* GP63 gene classes in amastigotes was not determined, however, the constitutive GP63 gene class was expressed at a much lower level in promastigotes than the L1/S1 gene class (Ramamoorthy et al. 1992).

The arrangement of the GP63 gene locus has now been determined for several different species of *Leishmania*, all of which contain loci that are significantly more complex than the *L. major* GP63 gene locus (Miller et al. 1990; Hanekamp and Langer, 1991; Webb et al. 1991; Steinkraus and Langer, 1992; Medina-Acosta et al. 1993b). Nonetheless, the arrangement of GP63 genes as direct tandem arrays is a recurrent theme across species, implying the importance of tandem gene organization for proper expression of GP63.

With the exception of the heterogeneous carboxy-terminal domain of the molecule, GP63 is highly conserved across clinically and geographically diverse species of *Leishmania*. Furthermore, since GP63 is reportedly expressed in both the promastigote and amastigote life stages it represents a potential candidate molecule for use as a subunit vaccine to protect against leishmaniasis. Accordingly, immunization with GP63 was reported to provide partial protection against the development of cutaneous leishmaniasis in mice (Handman and Mitchell, 1985; Russell and Alexander, 1988). Furthermore, susceptible strains of mice that were orally immunized with the *AroA* -vaccine strain of *Salmonella typhimurium* transformed with the full length *L. major* GP63 gene preferentially developed protective immunity against leishmaniasis via expansion of CD4+ TH1 cells (Yang et al. 1990). In addition, chemical interference of the protease activity of GP63 may be useful as a chemotherapeutic approach in treating active leishmaniasis.
C. GENE EXPRESSION IN KINETOPLASTID PROTOZOANS

Gene expression in *Leishmania* and other related members of the clinically important order *Kinetoplastidae* includes several unusual and interesting molecular phenomena, such as RNA editing, polycistronic transcription and *trans*-splicing. RNA editing (for review see Weiner and Maizels, 1990; Hajduk *et al.*.1993) occurs in the mitochondria of kinetoplastid protozoans which, as described above, contain two classes of circular DNA molecules, the maxicircle and the minicircle. The maxicircle is the functional equivalent of the mitochondrial genome from other eukaryotic organisms, however, kinetoplastid mitochondrial genes, or 'cryptogenes', are informationally incomplete. Cryptic transcripts derived from the kinetoplast maxicircle are modified through a process known as RNA editing in which uridine residues are added to, or deleted from, specific positions of the primary transcript to produce a mature mRNA containing a translationally competent open reading frame. The specificity of editing is determined by short guide RNAs (gRNAs) which are encoded by the minicircles and by specific regions of the maxicircle. The editing process involves extensive use of non-Watson Crick G:U base pairing, however, the exact mechanisms of editing are not completely understood at this time.

The *trans*-splicing reaction is the formation of a mature mRNA via the splicing of two separately encoded precursor gene products, the spliced leader RNA (SL RNA) and the mRNA precursor (for review see Laird, 1989; Huang and Hirsh, 1992). The SL RNA is a short, non-polyadenylated, capped transcript (140 nucleotides) which is encoded by a large family of clustered genes. *Trans*-splicing entails the transfer of a 35 to 39 nucleotide spliced leader (SL), derived from the 5' terminus of the SL RNA, to an internal 3' spliced
leader acceptor site on the mRNA precursor. This reaction is a general feature of gene expression in kinetoplastids and it results in all mature mRNAs having an identical 35 to 39 nucleotide spliced leader sequence at their 5' terminus. Trans-splicing reaction intermediates have been isolated which indicate that the process is analogous to the more familiar cis-splicing reaction required for intron removal in higher eukaryotes. Specifically, a Y-shaped intermediate that is susceptible to debranching enzyme has been isolated, which suggests the presence of an internal 2'-5' phosphodiester bond. This Y-shaped intermediate is therefore presumed to represent the functional equivalent of the lariat structure observed in cis-splicing. Moreover, several of the components required for cis-splicing (U2, U4 and U6 snRNAs) have been identified in kinetoplastids (Mottram et al. 1989; Gunzl et al. 1992). Interestingly, the spliced leader sequence itself functionally replaces the U1 snRNA which is otherwise absent in kinetoplastids (Bruzik and Steitz, 1990). The recent finding that kinetoplastid transcripts can be trans-spliced in mammalian cells that have been transformed to produce spliced leader RNA provides further evidence supporting the mechanistic similarity between cis and trans-splicing (Bruzik and Maniatis, 1992).

Although the exact function of the trans-spliced leader sequence is not clear, it is thought to be required for the formation of translatable mRNA in kinetoplastids, either by contributing a 5' cap structure to an otherwise non-capped mRNA precursor (Perry et al. 1987) or by affecting mRNA stability (Huang and Van der Ploeg, 1991b). In addition, protein-coding genes in kinetoplastids are often arranged as tandem arrays which are transcribed into polycistronic transcripts (Laird, 1989; Huang and Hirsh, 1992). The trans-splicing reaction may provide a mechanism for processing polycistronic precursors into individual mature mRNAs. Interestingly, classical cis-spliced introns have not been identified in any of the genes isolated from kinetoplastids thus far, and are therefore assumed to be absent.
However this is not an indication that cis-splicing and trans-splicing are mutually exclusive events since both processes are known to occur simultaneously in nematodes and trematodes (for review see Donelson and Zeng, 1990).

The process of splicing a leader sequence to an internal 3' acceptor site on the protein-encoding precursor mRNA is accompanied by the displacement of the original precursor 5' terminus. Consequently, the point of transcriptional initiation is unknown for many kinetoplastid genes and identification of associated basal promoter elements or gene specific regulatory elements remains elusive. The kinetoplastid transcriptional promoters which have been best characterized to date are those of the VSG (variant surface glycoprotein) and procyclin or PARP (procylic acidic repetitive protein) gene complexes of trypanosomes, however, both of these complexes are transcribed by an alpha-amanitin resistant class of RNA polymerase, likely RNA polymerase I (pol I) (Sherman et. al.1991; Zomerdijk et. al.1991a; Zomerdijk et. al.1991b; Rudenko et. al.1992). Conversely, the majority of kinetoplastid genes are assumed to be transcribed by RNA polymerase II (pol II) as they are transcribed in an alpha-amanitin sensitive manner, analogous to the RNA pol II transcribed genes of higher eukaryotes (Kooter and Borst, 1984). The only kinetoplastid pol II promoter which has been characterized to date is the T. brucei actin gene promoter (Ben Amar et. al.1988; Ben Amar et. al.1991). Transcription of the actin genes is reported to initiate approximately 4 kbp upstream of the actin gene locus, which contains 2 to 4 tandemly repeated actin genes, dependent upon the species. The essential region of the promoter was defined by deletion and was located approximately 600 bp upstream of the putative transcriptional start site. The nucleotide sequence of the promoter region was determined however no readily discernible regulatory sequence motifs were identified.
Interestingly, trypanosomes have been found to contain two genes encoding the largest subunit of RNA pol II (Evers et al. 1989; Smith et al. 1989). Although the two genes are highly similar, they encode proteins that differ by four amino acids. In addition, both genes encode proteins with carboxy-terminal extensions that are unique from the highly conserved pol II carboxy-terminal extension found in all other eukaryotes. The significance of the unique pol II CTD in trypanosomes and the presence of two pol II genes is currently not known.

Promoter identification in kinetoplastids is further complicated by the fact that many of the genes which have been isolated to date are present as multiple copies arranged in tandem arrays (Landfear et al. 1983; Clayton, 1985; Tschudi et al. 1985; Meade et al. 1987; Ben Amar et al. 1988; Button et al. 1989). Current evidence suggests that these arrays are transcribed in a polycistronic manner using tacit single upstream promoters (Imboden et al. 1987; Ben Amar et al. 1988; Muhich and Boothroyd, 1988; Tschudi and Ullu, 1988; Ben Amar et al. 1991). The resulting precursor molecules would be cleaved, either co-transcriptionally or post-transcriptionally, likely through the events of trans-splicing and polyadenylation (Huang and Van der Ploeg, 1991a; Ullu et al. 1993), to produce the mature, monocistronic mRNAs which are readily detectable by Northern blotting or primer extension. Interestingly, recent advances in the development of constructs for transfection of kinetoplastids have shown that reporter genes flanked only by complete or partial intergenic regions derived from naturally occurring tandem gene arrays are efficiently expressed at high levels (Laban and Wirth, 1989; Laban et al. 1990; ten Asbroek et al. 1990; Curotto de Lafaille et al. 1992). Although the point of transcriptional initiation in these constructs remains to be determined, it is apparent that there is no strict requirement for the presence of a polycistronic upstream promoter for efficient transcription.
of the reporter gene in these systems. Alternatively, it is possible that the tandem array intergenic regions used in these constructs have intrinsic promoter activity in addition to supplying the 3' trans-spliced leader acceptor site and polyadenylation site required for post-transcriptional processing. In support of this possibility, several kinetoplastid tandem array gene families have been shown to differentially express the individual members of the array (Meade et. al.1989; Bakalara et. al.1991; Medina-Acosta et. al.1993b). A third alternative is that pol II transcriptional initiation is somewhat promiscuous in kinetoplastids, occurring at multiple, loosely-defined positions including intergenic regions or cryptic plasmid sequences within transfection constructs. Indeed, multiple transcriptional initiation sites would likely be well tolerated in kinetoplastids since the 5' ends of transcripts are ultimately homogenized via the process of trans-splicing regardless of the point of transcriptional initiation. However, differential expression of genes in a developmentally regulated fashion implies that some level of control is required, either at the transcriptional or post-transcriptional level or both.

Expression of the GP63 gene locus has been characterized to a limited extent at the transcriptional level, however, like most other kinetoplastid protein-coding genes, the point of transcriptional initiation and the identity of promoter elements remains obscure. As described in the preceding section, the GP63 genes of *L. major* are arranged as a direct head to tail tandem array with a 3 kbp repeat unit length (Button and McMaster, 1988; Button et. al.1989). Each repeat unit is comprised of a 1.8 kbp protein coding region and a 1.2 kbp intergenic region. A consensus 3' spliced leader acceptor site is located 129 bp upstream of the ATG translational initiation codon (Button and McMaster, 1988) and use of this site has been confirmed by sequence analysis of GP63 cDNA clones from other species of *Leishmania* (Ramamoorthy et. al.1992). The *L. major* GP63 transcript detected
by Northern blot was also approximately 3 kb in length, implying that the 3' trans-spliced leader acceptor site and polyadenylation site are located very close together in the intergenic region. In addition, a minor GP63-specific transcript of approximately 6 kb was detected in L. major total RNA during Northern blot analysis and it was speculated that this transcript represented a potential polycistronic precursor derived from the GP63 tandem array (Button et al. 1989). Alternatively, this transcript may be derived from the single dispersed gene copy located immediately downstream of the main tandem array. Although the GP63 gene locus is presumed to be transcribed in a polycistronic manner in accordance with other kinetoplastid tandem gene arrays, recent results suggest that expression of individual GP63 genes or groups of genes is developmentally regulated (Ramamoorthy et al. 1992; Medina-Acosta et al. 1993b). At present there is no direct evidence to suggest whether expression is regulated at the transcriptional or post-transcriptional level, however, it is interesting to speculate that the intergenic region may contain the information required for appropriate expression.

D. THE PRESENT STUDY

The GP63 gene from the causative agent of Old World visceral leishmaniasis, L. donovani, was isolated and sequenced and compared to the GP63 genes from L. major and L. chagasi with the aim of identifying regions of similarity relevant to potential vaccine development and to identify regions of conservation which may be important for the function of GP63. Subsequently, a gene encoding a DNA-binding protein which interacted with oligodeoxynucleotides derived from the 5' untranslated region of the GP63 gene was isolated and sequenced. This protein, HEXBP, was found to contain nine zinc finger
motifs of the CCHC class, and was characterized as a single-stranded DNA-binding protein. Homologous gene replacement techniques were used to generate a clone of *L. major* which was deficient for the HEXBP DNA-binding protein. Analysis of the HEXBP-deficient mutant indicated that HEXBP is not required for expression of GP63 by promastigotes grown *in vitro*. The final chapter describes the development of a stable, selectable vector for the expression of cloned genes in *Leishmania*. 
II. MATERIALS AND METHODS

A. LEISHMANIA

1. Leishmania strains used in the present study

The following strains of Leishmania were used in this study where indicated. L. donovani (strain LV9, WHO designation - MHOM/ET/67/HU3), L. major (NIH S strain) (Wallis and McMaster, 1987), L. major CC1 (a clonal derivative of strain LT252, WHO designation - MHOM/IR/83/IR) (Kapler et al. 1990), L. mexicana mexicana (WHO designation - MNYC/B2/62/M379).

2. In vitro maintenance of Leishmania promastigotes

All species of Leishmania were maintained in vitro as promastigotes in M199 media (Gibco, Grand Island, New York) containing 10% fetal calf serum (Hyclone, Logan, Utah), 40 mM HEPES (pH 7.4), 50 units of penicillin per ml and 50 ug streptomycin per ml. Promastigotes were grown at 26°C in a non-humidified incubator. Cultures were maintained at densities ranging from 1 x 10^5 cells per ml to 5 x 10^7 cells per ml depending on the species and the application. Aliquots of the various species and strains described in the current study were stored in liquid nitrogen in M199 media containing 10% glycerol.
**B. BACTERIAL STRAINS, VECTORS AND MEDIA**

1. **Bacterial Strains**

   *E. coli* DH5α was routinely used as a host strain for maintaining and preparing plasmid DNA for subcloning, restriction mapping and DNA sequence analysis. *E. coli* strains DH5αF' (BRL, Gaithersburg, Maryland) and AA102 (Ahmed, 1987) were used for preparing the M13 and pAA3.7x constructs used in sequencing analysis. *E. coli* strains Y1089 and Y1090 (Young and Davis, 1983; Huynh *et al*. 1985) were used for plating the *L. major* λgt11 genomic expression library and for preparing lysogens expressing the β-galactosidase/HEXBP fusion protein. *E. coli* strain BL21(DE3)pLysS (Rosenberg *et al*. 1987; Studier *et al*. 1990) was used for expressing non-fusion HEXBP from the pET-3a-derived constructs.

2. **Vectors**

   The plasmids pUC18, pUC19 and pBluescript were used for routine subcloning, restriction mapping and double-stranded DNA sequence analysis. Bacteriophages M13mp18 and M13mp19 were used to produce single-stranded DNA templates for sequence analysis. The plasmid pAA3.7sx (Ahmed, 1987) was used for generating transposon insertion mediated deletion constructs. An *L. major* genomic expression library was prepared in λgt11 as previously described (Wallis and McMaster, 1987). Non-fusion HEXBP was expressed using constructs derived from the pET-3a bacterial expression plasmid (Rosenberg *et al*. 1987; Studier *et al*. 1990).
C. DNA ISOLATION

1. Isolation of Plasmid and Phage DNA

Small scale isolations of plasmid DNA for use in restriction enzyme mapping and double-stranded DNA sequencing were performed using an alkaline lysis/CsCl-based mini-prep method (Saunders and Burke, 1990). Larger scale isolations were performed using alkaline lysis followed by CsCl density gradient ultracentrifugation (Maniatis et al. 1989). Phage DNA was prepared by the plate lysis technique followed by CsCl density gradient centrifugation (Maniatis et al. 1989).

2. Isolation of Leishmania Genomic DNA

Genomic DNA was prepared from Leishmania promastigotes as follows. Promastigotes from fifty ml of a logarithmic stage culture were collected by centrifugation, washed twice with PBS and lysed by gently resuspending in 0.5 ml of lysis buffer (10 mM Tris-HCl, pH 8.0, 100 mM EDTA, 0.5% SDS, 20 μg RNAse A per ml, 100 μg proteinase K per ml). The lysate was incubated overnight at 55°C with occasional gentle mixing. The DNA was then extracted once with phenol, once with phenol/chloroform and once with chloroform. A one tenth volume of 10 M ammonium acetate and two volumes of ice cold ethanol were added and the DNA was allowed to precipitate at room temperature for approximately one hour. The stringy mass of genomic DNA was spooled out of the lysate on the end of a glass pasteur pipette, washed in three successive washes of absolute ethanol and resuspended in 0.5 ml of sterile water. Complete resuspension of the DNA usually took approximately 24 to 48 hours at 4°C.
Materials and Methods

D. RNA ISOLATION

1. Isolation of *Leishmania* RNA

Total RNA was isolated from *Leishmania* promastigotes using the single step acid guanidinium thiocyanate-phenol extraction protocol (Chomczynski and Sacchi, 1987). Briefly, promastigotes from a 50 ml culture were collected by centrifugation, washed twice with PBS and lysed by resuspending in 0.5 ml of ice cold solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0, 0.5% sarkosyl, 100 mM 2-mercaptoethanol). The lysate was incubated on ice for approximately 15 minutes with occasional mixing and 100 ul of 2 M sodium acetate, pH 4.0, 1 ml of water saturated phenol, and 0.2 ml of chloroform were sequentially added with mixing after each addition. The final lysate was shaken vigorously for approximately 10 seconds and incubated on ice for 15 minutes. The phases were separated by centrifugation at 10,000 x g for 20 minutes at 4°C and the aqueous phase (containing the RNA) was removed to a fresh tube, being careful not to disturb the phenol interface. The RNA was precipitated by adding 1 vol. of ice cold isopropanol and overnight incubation at -20°C. The RNA was sedimented by centrifugation in an Eppendorf microfuge for 30 minutes at top speed. The RNA pellet was washed twice in 70% ethanol, dried briefly and resuspended in 300 ul of ice cold solution D. The RNA was reprecipitated by addition of an equal volume of ice cold isopropanol and incubation at -20°C for at least 1 hour. The RNA was sedimented and washed twice with 70% alcohol as described above and resuspended in DEPC-treated water to approximately 1 µg/ul.
Materials and Methods

E. PROTEIN ISOLATION

1. Extracts of λgt11 Lysogens

Lysogens of λgt11 and λgthex were prepared in E. coli strain Y1089 as previously described (Young and Davis, 1983; Huynh et al. 1985). Two ml cultures of lysogens were incubated at 32°C until the OD<sub>600</sub> reached approximately 0.5. At this point the cultures were transferred to a 42°C incubator for 20 minutes to inactivate the temperature sensitive phage repressor. The cultures were then returned to 37°C and IPTG was added to 10 mM to induce the expression the β-galactosidase/HEXBP fusion protein. The induced cultures were incubated at 37°C for 2 to 3 hours to allow for sufficient accumulation of fusion protein. Cells were then collected by a brief centrifugation, washed twice in ice cold extract buffer (50 mM Tris-HCl, pH 7.5, 1 mM DTT and 1 mM PMSF) and resuspended in a final volume of 100 ul extract buffer. Cells were lysed by several successive freeze thaw cycles followed by a brief sonication. Lysates were cleared by a brief spin at 10,000 x g, aliquoted and immediately frozen in a dry ice/ethanol bath.

2. Extracts of pET-3a Clones

Early to mid log phase cultures of the pET-3a-derived construct pMHB3A were induced with 1 mM IPTG to express non-fusion HEXBP. Two ml cultures were induced for 2 to 4 hours. Cells were then collected by a brief centrifugation, washed twice in ice cold extract buffer (50 mM Tris HCl, pH 7.5, 1 mM DTT and 1 mM PMSF) and resuspended in a final volume of 100 ul extract buffer. Cells were lysed by several successive freeze thaw cycles followed by a brief sonication. Lysates were cleared by a brief spin at 10,000 x g, aliquoted and immediately frozen in a dry ice/ethanol bath.
3. Total Cell Extracts of *Leishmania* Promastigotes

*L. major* promastigote total protein extracts were prepared by resuspending $1 \times 10^6$ logarithmic stage promastigotes in 100 ul of extract buffer (50 mM Tris-HCl, pH 7.5, 1 mM DTT and 1 mM PMSF) and disrupting the cells by sonication. Extracts were cleared by a 5 minute spin at 10,000 X g and were frozen in dry ice/ethanol and stored at -70°C.

F. GEL ELECTROPHORESIS

1. Non-denaturing Agarose Gel Electrophoresis

DNA fragments were routinely separated by electrophoresis on non-denaturing agarose gels using 1 x TAE electrophoresis buffer (40 mM Tris-acetate, pH 8.5, 2 mM EDTA). DNA fragments less than 0.5 kbp (PCR products) were separated on 1% agarose/3% Nusieve (FMC Bioproducts, Rockland, Maine) gels. Fragments larger than 0.5 kbp were separated on 0.8% to 1.0% agarose gels and genomic DNA was separated on 0.6% agarose gels. DNA was detected by incorporating ethidium bromide in the gel (0.5 ug/ml) or by post-electrophoresis staining in ethidium bromide (0.5 ug/ml).

2. Southern Blot Hybridization Analysis

DNA to be analyzed by Southern blot hybridization was separated by electrophoresis on non-denaturing TAE agarose gels lacking ethidium bromide. DNA was detected by post-electrophoresis staining in ethidium bromide. The DNA was then transferred to Hybond-N membrane (Amersham, Oakville, Ontario) according to manufacturers instructions. Briefly, the gel was incubated in Southern blot denaturing
Materials and Methods

solution (1.5 M sodium chloride, 0.5 M sodium hydroxide) for 30 minutes followed by neutralizing solution (1.5 M sodium chloride, 0.5 M Tris-HCl, pH 7.2, 1 mM EDTA) for 30 minutes. The DNA was then transferred to the nylon membrane by overnight capillary transfer using 20 x SSC (3.0 M sodium chloride, 0.3 M sodium citrate). DNA was then covalently attached to the membrane by exposure on a UV lightbox (302 nm) for approximately 5 minutes. Filters were incubated in prehybridization solution (6 x SSC, 5 x Denhardt's solution, 0.5% SDS and 20 μg heat denatured salmon sperm DNA per ml) at 65°C for a minimum of 3 hours. Prehybridization solution was then removed and replaced by fresh prehybridization solution containing a radiolabelled hybridization probe. Hybridization was allowed to proceed overnight at 65°C. The following day the hybridization solution was removed and the filters were washed at 65°C with 2 changes of 2 x SSC (10 minutes each), one change of 2 x SSC containing 0.1% SDS (30 minutes), and one change of 0.2 X SSC (10 minutes). Filters were then placed in a sealed bag (to prevent complete dehydration) and analyzed by autoradiography. Membranes to be rehybridized were stripped of existing probe by washing in 0.4 M sodium hydroxide for 30 minutes at 45°C followed by 0.1 x SSC, 0.1% SDS, 0.2 M Tris-HCl, pH 7.5 for 30 minutes at 45°C. Complete removal of probe was confirmed by autoradiography.

3. Formaldehyde Agarose Gel Electrophoresis and Northern Blot Hybridization Analysis

RNA to be analyzed by Northern blot hybridization was separated by electrophoresis on 1% agarose gels containing 2.2 M formaldehyde using MOPS/EDTA running buffer (20 mM MOPS, pH 7.0, 5 mM sodium acetate, 1 mM EDTA). RNA was electrophoresed in the absence of ethidium bromide. A commercially available RNA ladder
(BRL, Gaithersburg, Maryland) was used as a molecular weight marker and was detected by post-electrophoresis staining in ethidium bromide (0.5 ug/ml). Samples were prepared for electrophoresis by combining 4.5 ul of RNA, 2 ul of 10 x MOPS/EDTA, 2 ul of 10 x gel loading buffer (50% glycerol, 1 mg Bromophenol blue per ml, 1 mg Xylene Cyanol per ml), 10 ul of formamide and 3.5 ul of formaldehyde. RNAs were denatured in a boiling water bath for 3 minutes and cooled on ice immediately prior to being loaded on the gel. Samples were electrophoresed until the Bromophenol blue dye front was approximately 1 cm from the bottom of the gel. Lanes containing molecular weight markers were then removed from the gel for staining and the remainder of the gel was washed in several changes of electrophoresis buffer over approximately 15 minutes to remove formaldehyde. RNA was then transferred to Hybond-N nylon membrane by capillary transfer as described above for Southern blotting. Membranes containing RNA were dried, UV-treated, prehybridized, hybridized and washed exactly as described above for Southern blot hybridization. Membranes to be rehybridized were stripped of existing probe by washing in 0.4 M sodium hydroxide for 30 minutes at 45°C followed by 0.1 x SSC, 0.1% SDS, 0.2 M Tris-HCl, pH 7.5 for 30 minutes at 45°C. Complete removal of probe was confirmed by autoradiography.

4. SDS-polyacrylamide Gel Electrophoresis and Western Blotting

Protein samples were fractionated by SDS-PAGE on a mini-PROTEAN II system according to manufacturers instructions (BIO-RAD, Richmond, California). Acrylamide concentrations ranged from 7.5% to 12.5% depending upon the application. Proteins were either directly stained using Coomassie Blue or blotted onto membranes for subsequent characterization. Gels used for staining were incubated overnight in 40% methanol/10%
acetic acid containing 0.1% Coomassie Blue. Gels were destained by washing in several changes of 40% methanol/10% acetic acid, 20% methanol/10% acetic acid and 7% acetic acid. Gels used for blotting were washed for approximately 30 minutes in several changes of Bjerrum and Schafer-Nielson transfer buffer (48 mM Tris-HCl, pH 9, 39 mM glycine, 20% methanol, 0.0375% SDS). Proteins were then transferred to either nitrocellulose (Schleicher and Schuell, Keene, New Hampshire) for Southwestern blot analysis, or Immobilon-P (Millipore, Bedford, Massachusetts) for Western blot analysis, using a semi-dry trans blot apparatus according to the manufacturer's instructions (BIO-RAD, Richmond, California). Transfer was conducted at 13 volts for 45 minutes using a small amount of Bjerrum and Schafer-Nielson transfer buffer. Membranes used for Western blot analysis were then blocked overnight with TBS (20 mM Tris-HCl, pH 7.5, 0.9% sodium chloride) containing 5% BSA. Blocked membranes were then incubated for one hour with polyclonal antisera as indicated, diluted 1:1000 in TBS containing 1% BSA, 0.05% Tween-20. Membranes were then washed with several changes of wash buffer (TBS containing 0.1% BSA and 0.05% Tween-20) and incubated for one hour with alkaline phosphatase-conjugated secondary antisera (BIO-RAD, Richmond, California), diluted 1:3000 in 1% BSA, 0.05% Tween-20. After extensive washing in wash buffer membranes were transferred to alkaline phosphatase buffer (100 mM Tris-HCl, pH 9.5, 100 mM sodium chloride, 5 mM magnesium chloride) and developed using BCIP/NBT according to the manufacturer's instructions (BRL, Gaithersburg, Maryland). Membranes used for Southwestern blot analysis were treated as described below under DNA binding assays.
5. Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gel electrophoresis was used for DNA sequencing (described below), DNase I footprint analyses (described below), and purification of synthetic oligodeoxynucleotides. In all cases electrophoresis was conducted using a Sequi-Gen sequencing gel system according to manufacturers instructions (BIO-RAD, Richmond, California). In general, gels with an acrylamide content of 6 or 8% were used for most routine sequencing applications, footprinting and oligonucleotide purification. Footprinting experiments utilizing shorter probes (50 nucleotides) were conducted using 16% acrylamide gels. Regardless of acrylamide content, gels contained 7 M urea as a denaturant and used TBE electrophoresis buffer (89 mM Tris-HCl, pH 8.3, 89 mM boric acid, 2 mM EDTA). Subsequent to electrophoresis, sequencing and footprinting gels were dried down onto Whatmann 3MM filter paper and subjected to autoradiography. Synthetic oligodeoxynucleotides were gel-purified on a preparative scale and detected by UV-shadowing. DNA-containing bands were excised and DNA was eluted from the gel by overnight incubation in 0.5 M ammonium acetate at 37°C. Oligonucleotides were then purified using SEP-PAK reverse phase chromatography columns (Millipore, Bedford, Massachusetts) as previously described (Atkinson and Smith, 1984).

G. GENERAL MOLECULAR BIOLOGY TECHNIQUES

1. Restriction Enzyme Digestion and Preparation of DNA fragments for Subcloning

Small scale restriction enzyme digestions (containing less than 5 µg of DNA) for restriction mapping, probe preparation and subcloning were performed using standard
Materials and Methods

Restriction enzyme digestions involving genomic DNA were conducted in a larger volume (generally 100 ul) and digestion was allowed to proceed overnight. Restriction endonuclease (approximately 3 U/ug of DNA) was added once to the initial reaction and again after approximately 3 or 4 hours of digestion. Subsequent to digestion the DNA was extracted with phenol/chloroform, ethanol precipitated and resuspended in a volume of TE suitable for loading on a non-denaturing agarose gel.

Digests involving multiple restriction enzymes were conducted simultaneously if 10x buffers were compatible. DNA was ethanol precipitated and resuspended if digests required different 10x buffers.

2. Ligation and Transformation of Bacteria

For routine subcloning all DNA fragments were gel purified. Ligation reactions contained DNA (10 to 1000 ng), 1x ligation buffer (50 mM Tris-HCl, pH 7.5, 5 mM magnesium chloride, 5 mM DTT, 50 ug BSA per ml), ATP (500 uM) and T4 DNA ligase (0.1 to 5 U). Ligations were conducted overnight at 15°C. Ligation reactions were terminated by adding 2 ul of 0.5 mM EDTA and 18 ul of water. Transformations were
performed using subcloning efficiency *E. coli* DH5α or DH5αF' competent cells according to manufacturers instructions (BRL, Gaithersburg, Maryland). Where required, transformants containing recombinant products could be discriminated from those containing non-recombinant products on the basis of blue white color selection by plating on media containing IPTG and X-gal.

3. Polymerase Chain Reaction

DNA was amplified by the polymerase chain reaction using a Perkin Elmer Cetus thermal cycler and AmpliTaq DNA polymerase (Cetus, Norwalk, Connecticut). Reaction conditions (i.e., templates, primers and thermocycle profiles) were as indicated for individual experiments. All PCR reactions were conducted in a final volume of either 50 or 100 ul and were supplemented with 10% DMSO. Subsequent to amplification, the reaction contents were transferred to a fresh vessel and residual oil was removed by extraction with chloroform.

4. Radioactive Labeling of DNA

Radiolabelled restriction fragments used as probes in Southern and Northern blot hybridizations were generated by the random primer method using the Klenow fragment of *E. coli* DNA polymerase I and [α32P]dCTP (Feinberg and Vogelstein, 1983). Synthetic oligodeoxynucleotide probes and enzymatically generated single-stranded probes used in DNA binding assays were end-labeled using T4 polynucleotide kinase and [γ32P]ATP (Ausubel et. al.1987). All radiolabelled probes were purified by passage over NACS reverse phase chromatography columns (BRL, Gaithersburg, Maryland) to separate probe
from free nucleotides. Except where indicated, all probes were denatured by a 5 minute incubation in a boiling water bath immediately prior to use.

5. Preparation of Single-stranded DNA Using λ Exonuclease

Double-stranded DNA was converted to single-stranded DNA using λ exonuclease for two purposes. λ exonuclease is a 5' to 3' exonuclease that digests one strand of a DNA duplex, initiating at a 5' phosphorylated terminus (Maniatis et al. 1989). Firstly, direct sequencing of PCR products (without subcloning) was performed using PCR products converted to single-stranded form using λ exonuclease (Higuchi and Ochman, 1989). This conversion was achieved by phosphorylating the 5' terminus of one of the PCR primers using T4 polynucleotide kinase prior to amplification and treating the amplification product with λ exonuclease. Secondly, the single-stranded probes BSHEX-327, BS-374 and gp63-5'-462 used in DNA binding assays were generated from PCR products by digestion with λ exonuclease. In the latter case the 5' phosphorylated terminus was generated by restriction enzyme digestion of the amplification product.

H. DNA SEQUENCE ANALYSIS

1. Preparation of Overlapping Deletion Clones

Overlapping deletion derivatives of fragments subcloned into M13mp18 and M13mp19 (Messing, 1983) were obtained by limited digestion with exonuclease III (Henikoff, 1984). Overlapping deletion derivatives of fragments subcloned into pAA3.7sx were generated by random transposon insertion (Ahmed, 1987).
2. Sequencing of Single-Stranded Templates

Single-stranded M13 templates were prepared from infected *E. coli* DH5αF' cells using standard methods (Maniatis *et. al*. 1989). Templates were sequenced using the dideoxy chain termination method (Sanger *et. al*. 1977) with T7 DNA polymerase and 7-deaza-dGTP (Pharmacia, Uppsala, Sweden). All sequencing reactions using M13 templates were conducted using the M13 universal sequencing primer (Pharmacia, Uppsala, Sweden).

3. Sequencing of Double-stranded Templates

Double-stranded sequencing was performed using supercoiled plasmid constructs isolated on both large and small scales (described above). Double-stranded templates were sequenced using T7 DNA polymerase, as described above for single-stranded templates, after being denatured with alkali and snap cooled according to manufacturers instructions (Pharmacia, Uppsala, Sweden). Sequencing reactions utilizing double-stranded templates were performed using sequence-specific synthetic oligonucleotide primers in addition to universal sequencing primers.

4. Sequencing of PCR Products

PCR amplification products were converted from duplex form to single-stranded form using λ exonuclease as described above. Single-stranded products were then sequenced as described above for single-stranded M13 templates using sequence-specific synthetic oligonucleotide primers.
5. Preparation of A + G Chemical Sequencing Ladders

Adenosine plus guanosine chemical sequencing ladders of the single-stranded probes BSHEX-327, BS-374 and gp63-5'-462 were generated for use as molecular weight markers in DNAse I footprint assays. Ladders were generated using standard chemical sequencing methods (Maxam and Gilbert, 1980). Briefly, formic acid (25 ul) was added to 10 ul (0.1 pM) of end labeled probe and incubated at room temp for 5 minutes. The DNA was then ethanol precipitated and cleaved at modified purines by incubating in 10% piperidine for 30 minutes at 90°C. Piperidine was removed by several successive rounds of lyophilization and the DNA was resuspended in formamide loading buffer at approximately 10,000 cpm/ul.

I. LIBRARY SCREENING

1. Screening for λgt11 Clones Expressing Functional DNA-binding Proteins

An L. major (strain NIH S) λgt11 genomic expression library (Wallis and McMaster, 1987) was screened for the presence of clones expressing functional DNA-binding proteins according to the method of Singh et. al. (1988, 1989) as modified by Vinson et. al. (1988) except that 1X binding buffer contained 50 mM Tris HCl, pH 7.5, 50 mM sodium chloride, 5 mM magnesium chloride, 1 mM EDTA and 1 mM DTT. Screening of Leishmania genomic expression libraries is a feasible and practical approach since none of the kinetoplastid protozoan genes isolated to date contain conventional cis-spliced introns. The primary screen was comprised of 6 plates each containing 3 x 10^4 plaques per plate. The double-stranded oligonucleotide probe used for library screening was generated
by annealing the two complementary synthetic single-stranded oligonucleotides HEX50(+) and HEX15(-) (see Table 2) and primer extension of the annealed product using the Klenow fragment of DNA polymerase I in the presence of \( [\alpha^{32}\text{P}]dCTP \). Positive clones were isolated by four rounds of plaque purification including two rounds of low density screening (<100 plaques per plate).

2. Preparation and Screening of a Size-selected *Leishmania* 'sub'-library

*L. major* genomic DNA was subjected to restriction endonuclease digestion using SstI and XbaI as described above and fractionated by electrophoresis on a non-denaturing agarose gel. The region of the gel corresponding to approximately 7 kbp was excised and the DNA was isolated using QUIEX (Quiagen, Chatsworth, California). The pool of 7 kbp SstI/XbaI fragments was subcloned into SstI/XbaI digested pUC19 to generate an *L. major* 'sub-library'. This SstI/XbaI 'sub-library' was plated at low density (<100 colonies per plate), transferred in duplicate to Hybond-N nylon membranes and screened by colony blot hybridization according to the manufacturer's instructions (Amersham, Oakville, Ontario). The probe used for colony blot hybridization was EcoRI/SalI restriction fragment of \( \lambda \) gtHEX, radiolabelled by the random primer method as described above. Positive colonies were picked directly off of the plate and purified by a second round of low density colony blot hybridization.
Materials and Methods

J. DNA-BINDING ASSAYS

1. Electrophoretic Mobility Shift Assay

Binding reactions for electrophoretic mobility shift assays contained 10,000 cpm of end-labeled probe, 1 ug of heat denatured poly d(I-C) as a non-specific competitor, unlabeled specific competitor as indicated and either 5 ug of total bacterial extract or 20 ug of L. major promastigote total cell extract in a final volume of 15 ul of 1X Binding Buffer (10 mM HEPES, pH 7.9, 4% Ficoll 400, 5 mM DTT and 0.5 mM zinc chloride). Binding reactions were incubated on ice for 90 minutes. Samples were loaded onto 4% non-denaturing polyacrylamide gels (gels were pre-electrophoresed for approximately 15 minutes) and fractionated using 0.25 X TBE electrophoresis buffer. Following electrophoresis, gels were dried and subjected to autoradiography overnight.

2. Southwestern Blot Analysis

For Southwestern blot assays, five ul of IPTG-induced bacterial extract was separated by electrophoresis on 10% SDS PAGE gels and proteins were transferred to nitrocellulose filters using a semi-dry trans blot apparatus according to manufacturers instructions (BIO-RAD, Richmond, California). Filters were then probed to detect DNA binding activity in the same manner as described above for library screening except that 1X binding buffer contained 50 mM Tris HCl, pH 7.5, 50 mM sodium chloride, 5 mM magnesium chloride, 0.5 mM zinc chloride and 1 mM DTT. Filters were incubated overnight at 4°C in a heat-sealed plastic bag with 10 ml of 1X binding buffer containing end-labeled oligonucleotide probe. The next morning filters were washed at room
temperature for approximately 15 minutes using several changes of 1X binding buffer. Filters were then dried and subjected to autoradiography overnight.

3. UV Cross-linking Analysis

Binding reactions for UV cross-linking assays were identical to those described above for electrophoretic mobility shift assays. Binding reactions were incubated on ice for 90 minutes and spotted onto a sheet of plastic wrap placed on the surface of a UV transilluminator (302 nm). Samples were UV irradiated for a total of 4 minutes, spotted samples were recovered from the saran and the spot was rinsed with an equal volume of 2X SDS PAGE sample buffer (125 mM Tris HCl, pH 6.8, 20% glycerol, 4% SDS, 10% β-mercaptoethanol, 0.1% Bromophenol Blue). Sample and rinse were then pooled and resolved by SDS-PAGE. Following electrophoresis, gels were stained with Coomassie blue, dried and subjected to autoradiography overnight.

4. DNAse I Protection Assays

Binding reactions for DNAse I protection assays were identical to those described above for electrophoretic mobility shift assays except that they contained 20,000 cpm of end-labeled probe. Reactions were incubated on ice for 90 minutes and equilibrated to room temperature immediately prior to DNAse I digestion. DNAse I (10,000 U/ml) was diluted 1:20 in 40 mM magnesium chloride, 20 mM calcium chloride and 5 ul was added in quick succession to each binding reaction. Reactions were mixed briefly and digestion was allowed to proceed at room temperature for 3 minutes unless otherwise indicated. Digestion was stopped by the addition of 70 ul of ice cold DNAse I Stop Buffer (DNAse I Stop Buffer was made by combining 645 ul of 95% ethanol, 50 ul of saturated ammonium
Materials and Methods

acetate and 5 μl of 1 mg/ml yeast tRNA), mixing and incubating in a dry ice/ethanol bath for 15 minutes. Precipitated DNA was sedimented by centrifugation, washed with 70% ethanol and resuspended in 5 μl of formamide loading buffer. Maxam and Gilbert A+G chemical sequencing ladders were prepared from each probe as described above for use as molecular weight markers. Digested fragments were resolved on either 16% or 6% denaturing polyacrylamide gels as indicated and detected by autoradiography.

K. TRANSFORMATION OF LEISHMANIA

1. Electroporation of Leishmania Promastigotes

Promastigotes in mid-logarithmic growth phase were transfected using previously established protocols (Kapler et al. 1990). Briefly, promastigotes were collected by centrifugation, washed twice with ice cold PBS and resuspended at 1 x 10^8 cells per ml in electroporation buffer (EPB is 21 mM HEPES, pH 7.5, 137 mM sodium chloride, 5 mM potassium chloride, 0.7 mM sodium phosphate, 6 mM glucose). Promastigotes in EPB (0.4 ml) and appropriate DNAs in sterile TE were mixed in electroporation cuvettes (0.2 cm gap width) and electroporated using a Gene Pulser Apparatus according to the manufacturer's instructions (BIO-RAD, Richmond, California). Electroporations were performed with the apparatus set at a voltage of 0.45 kV, and a capacitance of 500 μFd. Electroporated cells were immediately placed on ice for 10 minutes and then removed to a 25 ml tissue culture flask. Cells were incubated for 24 to 48 hours in 10 ml of drug-free media to allow for expression of transfected genes.
2. Cloning of *Leishmania* Transfectants

Transfectants were collected from drug free-media by centrifugation and plated onto semi-solid M199 plates (containing 1% agar). Transfectants were selected by growth on plates containing 8 μg G418 per ml (Geneticin, BRL, Gaithersburg, Maryland) and/or 32 μg Hygromycin B per ml (Sigma, St. Louis, Missouri). Colonies arising on drug selection plates (generally after 6 to 18 days) were picked into 1 ml of liquid culture containing 4 μg G418 per ml or 16 μg Hygromycin B per ml and expanded in the continued presence of selective drug.
### Table 2. Oligodeoxyribonucleotides used in the present study

<table>
<thead>
<tr>
<th>Name</th>
<th>Length</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEX50(+)</td>
<td>50</td>
<td>TGCACAAGCC(CTCGCC)_{5}ACCACACCCC</td>
</tr>
<tr>
<td>HEX15(-)</td>
<td>15</td>
<td>GGGGTGTGGTGCGAG</td>
</tr>
<tr>
<td>HEX50(-)</td>
<td>50</td>
<td>GGGGTGTGGT(GGCGAG)_{5}GGCTTGTGCA</td>
</tr>
<tr>
<td>HEX15(+)</td>
<td>15</td>
<td>TGCACAAGCCCTCGC</td>
</tr>
<tr>
<td>GP63-Pro48</td>
<td>18</td>
<td>CGCGGCGCCGAGAGTAAC</td>
</tr>
<tr>
<td>GP63-Pro58</td>
<td>18</td>
<td>GCTGCCCGGCGCGATC</td>
</tr>
<tr>
<td>gp63-5'-50(+)</td>
<td>50</td>
<td>CGCCGTGCAACAAGCC(CTCGCC)_{2}ACCACACCCCAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGCCCACAGCG</td>
</tr>
<tr>
<td>gp63-5'-50(-)</td>
<td>50</td>
<td>CGCTGTGGGCAGTGCGGTGTG(TGCGAG)_{2}GGC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TTGTGCACCGGC</td>
</tr>
<tr>
<td>gp63-pro-50</td>
<td>50</td>
<td>GACGTGAACTGGGGCGCGCTCGCATCGCGTGTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCACGGAGGACCTGAA</td>
</tr>
</tbody>
</table>
III. THE GP63 GENE OF *Leishmania donovani*

This chapter describes the cloning and characterization of a GP63 gene from *L. donovani*, the causative agent of Old World visceral leishmaniasis. The *L. donovani* gene was compared with the GP63 genes from two other species of *Leishmania*, *L. major* and *L. chagasi*, with the aim of identifying regions of similarity relevant to potential vaccine development and to identify regions of conservation which may be important for the function of GP63. In addition, the *L. donovani* GP63 gene locus was characterized in terms of gene copy number and organization.

A. RESULTS

1. The Sequence of the *Leishmania donovani* GP63 gene

The *L. donovani* GP63 gene was isolated from an *L. donovani* λEMBL3 genomic library (generously provided by J.M. Kelly, London School of Hygiene and Tropical Medicine) by plaque hybridization using a radiolabelled restriction fragment of the *L. major* GP63 gene as a hybridization probe. Four positive clones (λLdGP63-4-4, -7-4, -8-4 and -11-4) were isolated and characterized by Southern blot hybridization analyses (Button *et al.* 1989). Previous analysis of the *L. donovani* GP63 gene locus by genomic Southern blot analysis indicated that the GP63 genes were arranged as a tandem array with a 3 kbp repeat unit length (Button *et al.* 1989). A 3 kbp SalI restriction fragment that corresponded to a single GP63 repeat unit was isolated from clone λLdGP63-7-4. This fragment hybridized to probes specific to both the coding and non-coding regions of the *L. major*
Chapter III - Results

GP63 gene (data not shown). The 3 kbp SalI fragment was isolated by gel-purification and cloned into the SalI site of M13mp19 to generate the construct Md7sd5 and into the SalI site of the plasmid pAA3.7x (Ahmed, 1987) to generate the construct pAALdGP63. The nucleotide sequence of the entire 3 kbp SalI fragment was determined using Sanger dideoxy chain termination sequencing (Sanger et. al. 1977) of overlapping deletion derivatives of Md7sd5 and pAALdGP63.

The nucleotide and predicted amino acid sequence of the *L. donovani* GP63 gene from Md7sd5/pAALdGP63, aligned with the previously published *L. major* (Button and McMaster, 1988) and *L. chagasi* (Miller et. al. 1990) GP63 gene sequences is shown in Fig. 1. The 3 kb SalI fragment spanned a complete repeat unit of the GP63 tandem array, beginning at a SalI restriction site at nucleotide position 7 of the first gene and ending at the identical position in a second GP63 gene located directly downstream. The nucleotide sequence of the 5' untranslated region of the first gene was determined by direct sequencing of *λ* LdGP63-7-4 using synthetic oligodeoxynucleotide sequencing primers and was found to be identical to the 5' untranslated region of the downstream GP63 gene for at least 100 bp. Optimal alignment of the 5' untranslated regions of the three GP63 gene sequences shown in Fig. 1 required the insertion of a 6 bp gap in the *L. chagasi* sequence and a 12 bp gap in the *L. major* sequence. These gaps reflected a species-specific difference in the number of hexanucleotide direct repeats beginning at position -42. The hexanucleotide repeat (CTCGCC) was present as four direct copies in the *L. donovani* gene, three direct copies in the *L. chagasi* gene and two direct copies in the *L. major* gene. The remaining 198 bp of 5' untranslated sequence shown in Fig. 1 was highly conserved across the three species. The GP63 gene 5' untranslated region contained no readily apparent eukaryotic RNA polymerase II promoter elements with the exception of a single
Chapter III - Results

potential Sp1 core binding site (Dynan and Tijan, 1983) at position -97. A consensus transspliced leader acceptor site (Laird, 1989) was located at position -140. In contrast to the 5' untranslated region, the sequence of the GP63 gene 3' untranslated region was conserved across species for a much shorter distance. The 3' untranslated sequences of the *L. donovani* and *L. major* genes diverged after approximately 75 bp. However, the sequence of the 3' untranslated was highly conserved between *L. donovani* and *L. chagasi*, and this conservation was maintained throughout the remainder of the intergenic region (not shown in Fig. 1).

The nucleotide sequence of the protein-coding region of the *L. donovani* GP63 gene exhibited 93.9% identity with the *L. chagasi* GP63 gene and 88.3% identity with the *L. major* GP63 gene. Nucleotide differences among the three species were predominantly grouped in clusters as opposed to being randomly distributed throughout the protein-coding region. Clusters of nucleotide differences which resulted in five or more contiguous amino acid substitutions in the predicted protein sequence occurred at nucleotide positions 243, 351, 807, 1239 and 1455. The protein coding region of the *L. donovani* GP63 gene was 36 bp shorter than that of *L. major* and optimal alignment required the insertion of a 39 bp gap in the *L. donovani* sequence at position 243 and a 3 bp gap in the *L. major* sequence at position 1254. Similarly, the *L. donovani* protein coding region was 27 bp shorter than that of *L. chagasi*, and optimal alignment again required the insertion of a 30 bp gap in the *L. donovani* sequence at position 243 and a 3 bp gap in the *L. chagasi* sequence at position 1254.

GP63 has been reported to have a protease activity (Bordier, 1987; Chaudhuri and Chang, 1988; Bouvier et. al.1989; Ip et. al.1990; Medina-Acosta et. al.1993a) and the predicted GP63 protein sequence from all three species contained a consensus motif
(beginning at residue 251) that was homologous to the zinc histidyl ligand of Zn$$^{++}$$ metalloproteases (Bouvier et. al.1989). Amino-terminal amino acid sequence analysis of GP63 isolated from the surface of L. major promastigotes indicated that GP63 is synthesized as a precursor containing a 100 residue amino terminal extension (Button and McMaster, 1988). Consistent with the identification of GP63 as a cell surface protease it has been suggested that the 100 residue amino-terminal extension contains both a prepeptide region (transport signal peptide) and a propeptide region (protease regulatory peptide) (Button and McMaster, 1988; Button et. al.1991). The 39 residues of the putative prepeptide region and the residues constituting the putative pre/pro cleavage site are highly conserved across the three Leishmania species. Conversely, the propeptide region of GP63 exhibited species-specific size heterogeneity. The predicted propeptide region of L. donovani GP63 (48 residues) was 13 residues shorter than that of L. major GP63 and 10 residues shorter than that of L. chagasi GP63. These size differences are likely the result of insertional/deletional events since nucleotide sequences flanking the gaps are highly conserved. As demonstrated for other Zn$$^{++}$$ metalloproteases, a single cysteine residue was located in the putative pro-region of the GP63 protein from all three species. This cysteine is likely involved in the regulation of metalloprotease activity (Van Wart and Birkedal-Hansen, 1990). The amino-terminus of mature GP63 correlated with a cleavage of the precursor between Val (100) and Val(101) (Button and McMaster, 1988). The predicted L. donovani GP63 protein also contained Val-Val at this position (corresponding to nucleotide 261 in Fig. 1) suggesting that the same processing site is likely used in both species.

GP63 is anchored to the plasma membranes via a glycosyl-phosphatidylinositol (GPI) linkage (Bordier, 1987). A 22 to 25 residue carboxyl terminal fragment of GP63 is
cleaved during the addition of the GPI moiety and the amino acid sequence of this region is conserved between *L. major* and *L. donovani* with the exception of a single amino acid difference (Thr versus Ala) at one of the predicted cleavage sites. The substituted residue does however conform with the consensus sequence for lipid attachment described for a number of GPI-linked surface proteins (Ferguson and Williams, 1988). The predicted *L. donovani* GP63 sequence contained only a single potential *N*-linked glycosylation site (residue 287) whereas the *L. major* and *L. chagasi* predicted protein sequences each contained two sites (Button and McMaster, 1988; Miller et al. 1990).

2. Arrangement of the *L. donovani* GP63 Gene Locus

*L. donovani* genomic DNA was analyzed by Southern blot hybridization to determine the structure of the GP63 gene locus. Restriction enzymes that were predicted from DNA sequence analysis to cut once within the protein-coding region of the gene (SalI, NotI, BglII, NcoI, ClaI and XbaI) each produced a predominant 3 kbp fragment on genomic Southern blots when hybridized with a probe that mapped to the 3' end of the GP63 protein-coding region (Fig. 2). The same 3 kbp fragment was detected when the blot shown in Fig. 2 was stripped and rehybridized with a probe specific to the 5' end of the GP63 protein-coding region (data not shown), in accordance with the presence of tandemly repeated gene copies. However, restriction enzymes which generated a 3 kbp repeat unit fragment also produced several additional hybridizing fragments, suggesting that the *L. donovani* GP63 gene locus was more complex than a simple tandem array. In addition, restriction enzymes that were predicted from the GP63 gene sequence to have multiple recognition sites within the GP63 protein-coding region (PvuII, SstI and PstI) produced
several unexpected hybridizing fragments in genomic Southern blot hybridizations, implying that sequence heterogeneity existed amongst the GP63 genes of *L. donovani* (data not shown).

To estimate the minimum number of genes in the *L. donovani* GP63 tandem array, *L. donovani* genomic DNA was subjected to partial digestion with NcoI (predicted from sequence analysis to have a single recognition site within each gene). Hybridization of the partially digested DNA with a GP63-specific probe revealed a ladder containing at least 7 bands with 3 kbp spacing (Fig. 3, lane D) suggesting that there are at least 7 genes in the tandem array portion of the *L. donovani* GP63 gene locus. When *L. donovani* genomic DNA was digested to completion with EcoRI, a single hybridizing fragment of approximately 40 kbp was detected (Fig. 3, lane B) implying that all of the *L. donovani* GP63 genes are linked on a single EcoRI fragment of approximately 40 kbp. However, the presence of multiple hybridizing bands of similar size (40 kbp) cannot be ruled out due to the limited resolution within this portion of the gel. Previous analyses of *L. donovani* chromosomes separated by pulsed field gel electrophoresis demonstrated that all of the *L. donovani* GP63 genes were contained on a single 700 kb chromosome band (Button *et al.* 1989).

Initial attempts to map the complete *L. donovani* GP63 gene locus by genomic Southern blot hybridization were frustrated by the complexity of the locus, therefore, the four positive λ EMBL3 clones described above (λLdGP63-4-4, -7-4, -8-4 and -11-4) were characterized with the aim of identifying a contiguous restriction map spanning some or all of the GP63 locus. Southern blot hybridization of single, double and partial restriction enzyme digests of clones λLdGP63-7-4 and λLdGP63-8-4 with probes specific for both the 5' and 3' ends of the GP63 protein-coding region were used to
compile the restriction maps shown in Fig. 4C. Digestion of $\lambda$ LdGP63-7-4 DNA with restriction enzymes predicted to have a single recognition site within each GP63 gene generated 3 kbp fragments that hybridized strongly with probes specific to both the 5' and 3' ends of the GP63 protein-coding region (Fig. 4A and 4B, respectively). This restriction pattern was similar to that previously observed in the Southern blot analysis of genomic DNA, implying that $\lambda$ LdGP63-7-4 contained at least one repeat unit of the GP63 tandem array. $\lambda$ LdGP63-7-4 also contained a partial GP63 repeat unit represented by the 2.2 kbp SalI and 4.4 kbp ClaI fragments which were linked to one arm of the EMBL3 vector. In addition, $\lambda$ LdGP63-7-4 contained a single GP63 gene which was separated from the 3' end of the tandem array by 1.7 kbp, as indicated by the 3.5 kbp fragment in NolI, NcoI and ClaI digests which hybridized to both 5' and 3' specific probes. Southern blot hybridization analyses of clones $\lambda$ LdGP63-4-4 and $\lambda$ LdGP63-11-4 indicated that these clones contained identical inserts and therefore likely represent duplicates of the same clone (data not shown). The insert of $\lambda$ LdGP63-4-4/11-4 was smaller than that of $\lambda$ LdGP63-7-4 and contained only the last two genes of the tandem array and the single dispersed gene copy (Fig. 4C, genes C, D and E). $\lambda$ LdGP63-4-4/11-4 was therefore used to confirm the map of $\lambda$ LdGP63-7-4 but did not add any new information. In contrast, $\lambda$ LdGP63-8-4 contained GP63 genes which did not map to the 3 kbp tandem array. $\lambda$ LdGP63-8-4 contained two complete GP63 coding regions that were separated by an intergenic region that was 3.7 kbp in length, as indicated by the 5.5 kbp SalI and ClaI fragments that hybridized strongly to both 5' and 3' specific probes (Fig. 4A and 4B). The restriction maps of the two GP63 genes of $\lambda$ LdGP63-8-4 were similar to the map of the dispersed gene from $\lambda$ LdGP63-7-4 (Fig. 4C, gene E). The restriction maps of these lambda clones was confirmed by Southern hybridization analysis of double restriction digests and partial
restriction digestions (data not shown) and by the presence of corresponding fragments in the Southern blots of genomic DNA (Fig. 2). Together the inserts of \( \lambda \) LdGP63-7-4 and \( \lambda \) LdGP63-8-4 spanned approximately 30 kbp of DNA and therefore likely encompass the majority of the \( L. \) donovani GP63 gene locus. However, this analysis did not provide any direct evidence supporting linkage of the inserts of \( \lambda \) LdGP63-7-4 and \( \lambda \) LdGP63-8-4 in the \( L. \) donovani genome. Furthermore, several hybridizing fragments detected during Southern hybridization of \( L. \) donovani genomic DNA (Fig. 2) were not present in either of \( \lambda \) LdGP63-7-4 or \( \lambda \) LdGP63-8-4, implying that the \( L. \) donovani GP63 locus contained additional GP63 genes not represented on the maps shown in Fig. 4C.

In summary the \( L. \) donovani GP63 locus consists of a tandem array containing a minimum of seven directly repeated genes with a 3 kbp repeat unit length. A single GP63 gene is separated from the 3' end of the tandem array by 1.7 kbp and there are at least two additional GP63 genes which are dispersed from the tandem array portion of the locus.

### 3. Heterogeneity Within the Propeptide-Coding Region of the \( L. \) donovani GP63 Genes

As indicated in Fig. 1, the propeptide-coding region of the \( L. \) donovani GP63 gene was 39 bp shorter than the propeptide-coding region of the \( L. \) major GP63 gene. To determine whether this size difference was common to all of the \( L. \) donovani GP63 genes or whether it was restricted to a subset of genes, a DNA fragment encompassing the propeptide-coding region was amplified by PCR using either \( Leishmania \) genomic DNA or cloned GP63 genes as templates. Single fragments of the expected size (\( L. \) donovani - 169
bp, *L. major* - 208 bp) were amplified from plasmid clones containing the *L. donovani* GP63 gene (Md7sd5) or *L. major* GP63 gene (pBS10Rb.1) for which the complete sequence had been determined (Fig. 5A, lanes 1 and 2 respectively). A single fragment of 208 bp was also generated when genomic DNA from *L. major* was used as a template for amplification (Fig. 5A, lane 8). In contrast, a doublet was evident in the amplification products of \( \lambda \)LdGP63-8-4, \( \lambda \)LdGP63-11-4 or *L. donovani* genomic DNA (Fig. 5A, lanes 5 to 7) suggesting that there was size heterogeneity within the propeptide-coding region of the *L. donovani* GP63 genes. The smaller fragment of the doublet (169 bp) was of the size predicted based on the sequence of the *L. donovani* GP63 gene. The second fragment of the doublet was larger than the predicted *L. donovani* amplification product (169 bp) but smaller than the predicted *L. major* amplification product (208 bp). Both amplification products were isolated by gel purification and characterized by DNA sequence analysis. The DNA sequence of the smaller PCR product (Fig. 5B, PCR LOW) was identical to the propeptide-coding region of the *L. donovani* GP63 gene which was initially sequenced (Fig. 1). The sequence of the larger PCR product (Fig. 5B, PCR HIGH) corresponded to a GP63 propeptide-coding region which was 30 bp longer than that of the GP63 gene shown in Fig. 1. Interestingly, the sequence of this 'alternate' propeptide-coding region was identical to the sequence of the *L. chagasi* GP63 propeptide-coding region. These results demonstrated the presence of two classes of GP63 genes in the *L. donovani* genome that are distinguishable by the length of the propeptide-coding region. The doublet band was also observed when cDNA prepared from *L. donovani* promastigote RNA was used as template for PCR amplification (Fig. 5A, lane 9) implying that both classes of gene are actively transcribed in promastigotes. Whether the transcripts from both gene classes are translated and processed to yield functional GP63 protein.
remains to be determined. These genes were designated as GP63-Pro48 and GP63-Pro58 according to the number of amino acids predicted to constitute their propeptide regions.

Clones λ LdGP63-7-4 and λ LdGP63-8-4 were subjected to Southern blot analyses using oligodeoxynucleotide probes capable of discriminating between GP63-Pro58 and GP63-Pro48 genes in order to map the two GP63 gene types. Initially, Southern blots of the PCR amplification products shown in Fig. 5A were hybridized with the GP63-Pro58 and GP63-Pro48 specific oligodeoxynucleotides to confirm the specificity of these probes (Fig. 5C and 5D respectively). Southern blots containing λ LdGP63-7-4 and λ LdGP63-8-4 DNA (previously shown in Fig. 4) were then stripped and alternately rehybridized with the class-specific oligodeoxynucleotide probes (Fig. 6, upper panels). These analyses demonstrated that λ LdGP63-7-4 and λ LdGP63-8-4 contained both GP63-Pro48 and GP63-Pro58 gene types since the specific oligodeoxynucleotide probes hybridized to distinct fragments in each case. Although both GP63-Pro48 and GP63-Pro58 genes were detectable in λ LdGP63-7-4 by Southern blot analysis, only GP63-Pro48 genes were detected in the PCR amplification products of λ LdGP63-7-4 (Fig 5 A and C lane 4). This was likely due to sequence divergence at the 3' primer annealing site of the GP63-Pro58 genes in this clone. GP63-Pro58 genes could also be defined by the presence of a second NotI restriction site located in the extra 30 bp of the 'alternate' pro-coding region. Using these data the GP63 genes of λ LdGP63-7-4 and λ LdGP63-8-4 could be assigned as being either GP63-Pro58 or GP63-Pro48 gene types (Fig. 6, lower panel).

The last gene of the 3 kbp tandem array and the gene separated from the 3' end of the tandem array by 1.7 kb (Fig. 6, genes D and E) were both GP63-Pro58 genes whereas genes B and C of the tandem array were both GP63-Pro48 genes. Gene A of the tandem array has been truncated 3' of the propeptide-coding region during cloning and therefore its
gene type could not be determined. $\lambda$LdGP63-8-4 contained one GP63-Pro58 gene (labeled F) and one GP63-Pro48 gene (labeled G).
Figure 1. The nucleotide and predicted amino acid sequence of the *L. donovani* GP63 gene.

The nucleotide and predicted amino acid sequence of the *L. donovani* GP63 gene (L.d. DNA) is shown aligned with the reported *L. chagasi* (L.c. DNA) and *L. major* (L.m. DNA) GP63 gene sequences. The sequences are displayed in 5' to 3' orientation with the first base of the ATG initiation codon of the *L. donovani* GP63 gene labeled as position 1. The slash after nucleotide number 7 represents the SalI cleavage site utilized during subcloning (nucleotides shown 5' of this position represent the 5' non-translated region of a GP63 gene found directly downstream in the tandem array). Dots indicate nucleotides which are identical to the *L. donovani* sequence. Dashes indicate the absence of nucleotides at identical positions. The corresponding predicted amino acid sequences are found below the nucleotide sequences. The cleavage site which gives rise to the mature amino-terminus of *L. major* GP63 is indicated by an arrow at nucleotide position 261. The proposed cleavage site implicated in glycosyl-phosphatidylinositol anchor attachment at the carboxyl terminus is indicated by an arrow at nucleotide position 1695. The proposed trans-spliced leader acceptor site is indicated by an arrow at nucleotide position -140. An Sp1 consensus binding site is indicated by the line over nucleotides -97 to -102. The zinc binding site implicated in metalloprotease activity is indicated by the line over nucleotides 751 to 765. The potential glycosylation site for the *L. donovani* GP63 is indicated an asterisk at nucleotide positions 861. The complete sequence of the *L. donovani* GP63 gene has been submitted to GenBank™ under accession number M60048.
Figure 2. Southern blot hybridization of *L. donovani* genomic DNA.

*L. donovani* genomic DNA (5 μg per lane) was digested with the restriction enzymes indicated and analyzed by Southern blot hybridization with a probe specific for the 3' end of the GP63 gene (SstI/XbaI fragment of Md7sd5, bases 988 to 1769 of the GP63 coding region). Agarose gel electrophoresis and Southern blot hybridization conditions were as described in Materials and Methods. The sizes of co-electrophoresed DNA markers are indicated to the left in kilobasepairs (kb).
Figure 3. Genomic organization of the *L. donovani* GP63 gene array. *L. donovani* genomic DNA was digested to completion with EcoRI (lane B) or partially digested with decreasing amounts of NcoI (lanes C to E) and analyzed by Southern blot hybridization using a probe specific for the 3' end of the GP63 gene (see Fig. 2). The sizes of co-electrophoresed high molecular weight DNA markers (lane A) are indicated to the left in kilobasepairs (kb).
Figure 4. Restriction enzyme mapping of the lambda clones \( \lambda \text{LdGP63-7-4} \) and \( \lambda \text{LdGP63-8-4} \).

DNA isolated from the lambda clones \( \lambda \text{LdGP63-7-4} \) and \( \lambda \text{LdGP63-8-4} \) (250 ng per lane) was digested to completion with the enzymes indicated above each lane and subjected to Southern blot hybridization analysis using probes specific for A) the 5' end of the GP63 gene (Sau3A fragment of Md7sd5, bases 528 to 855 of the GP63 coding region) or B) the 3' end of the GP63 gene (SstI/XbaI fragment described in Fig. 2). Electrophoresis and Southern blot conditions were as described in Materials and Methods. The sizes of co-electrophoresed DNA markers are indicated to the left in kilobasepairs (kb). C) Restriction enzyme maps of Md7sd5 and the EMBL3 clones \( \lambda \text{LdGP63-7-4} \) and \( \lambda \text{LdGP63-8-4} \). GP63-coding regions are depicted as arrows indicating the gene orientation. The map labeled Md7sd5, shows the restriction map of the GP63 repeat unit for which the complete sequence was determined (see Fig. 1). Solid bars above the Md7sd5 map denote the positions of 5' and 3' restriction fragments used as hybridization probes in the Southern blot analyses. Restriction enzyme sites represented: S = SalI, Sa = Sau3a, N = NotI, H = Hinfl, Ps = PstI, Pv = PvuII, B = BglII, Ss =SstI, X = XbaI, K = KpnI, Nc = NcoI, C = ClaI. Arrows lettered A, B, C and D are members of the GP63 gene tandem array, having a 3 kbp repeat length. Gene E is separated from the end of the tandem array by 1.7 kbp. Genes F and G are dispersed from the tandem array and have a repeat unit length of 5.5 kbp. The scale at the bottom of the figure corresponds to the maps of the lambda clones whereas the scale at the top of the figure corresponds to the map of Md7sd5.
Chapter III - Results

A

B

C

5' probe

3' probe

Md7sd5

λLdp63-7-4

λLdp63-8-4

63
Figure 5. Analysis of the propeptide-coding region of *L. donovani* GP63 genes by PCR amplification and Southern blot hybridization.

A region encompassing the carboxyl-terminus half of the GP63 propeptide-coding region was amplified by PCR from the following sources: 1) Md7sd5 (M13 clone containing a single *L. donovani* GP63 gene), 2) pBS10Rb.1 (plasmid clone containing a single *L. major* GP63 gene), 3) λLdGP63-4-4, 4) λLdGP63-7-4, 5) λLdGP63-8-4, 6) λLdGP63-11-4, 7) *L. donovani* genomic DNA, 8) *L. major* genomic DNA, 9) *L. donovani* promastigote cDNA, 10) *L. donovani* promastigote cDNA prepared from RNA which had been treated with RNAse and 11) *L. major* promastigote cDNA. Panel A) PCR products were separated by electrophoresis on 1% agarose/3% Nusieve gels and detected by staining with ethidium bromide. The sizes of the PCR products are indicated in bp. Panel B) DNA sequence analysis of the 169 bp (PCR LOW) and the 199 bp (PCR HIGH) PCR amplification products from lane 7 above. Sequences are shown aligned with the propeptide-coding region from the *L. donovani* GP63 gene (Md7sd5), the *L. chagasi* GP63 gene (L.c.) and the *L. major* GP63 gene (L.m.). Panels C and D) Southern blot analyses of the agarose gel shown in panel A hybridized alternately with oligonucleotide probes specific for GP63-Pro58 genes (Panel C) and GP63-Pro48 genes (Panel D).
Figure 6. Analysis of the GP63 propeptide-coding regions in λLdGP63-7-4 and λLdGP63-8-4 by Southern blot hybridization.

Upper panels: Southern blots of λLdGP63-7-4 and λLdGP63-8-4 shown in Fig. 4 were stripped and alternately rehybridized with oligodeoxynucleotide probes specific for GP63-Pro48 and GP63-Pro58 genes as indicated. Lower panels: restriction maps of λLdGP63-7-4 and λLdGP63-8-4 showing assignment of GP63 genes as either GP63-Pro48 (solid arrows) or GP63-Pro58 (striped arrows) gene types.
B. DISCUSSION

GP63 is the predominant protein on the surface of Leishmania promastigotes, accounting for as much as 1% of the protein in a promastigote total cell extract (Bordier, 1987). Although the role of GP63 in the life cycle of Leishmania remains to be established, the functional significance of the molecule is underscored by its abundance and the high level of GP63 gene conservation amongst pathologically and geographically diverse species of Leishmania. This was clearly evident in the results of the current study which revealed an overall DNA sequence identity of greater than 88% between the GP63 genes of L. donovani, the agent of Old World visceral leishmaniasis, and L. major, the agent of Old World cutaneous leishmaniasis. Despite minor alterations in sequence, the overall structural organization of the GP63 protein appeared to be strictly conserved and regions of the gene that encoded amino-terminal pre and propeptide domains and carboxyl-terminal GPI attachment sites were maintained across species boundaries. Furthermore, GP63 has been characterized as a zinc metalloprotease (Bordier, 1987) and the presumptive zinc-binding domain of L. major GP63 (His-X-Met-X-His) (Bouvier et. al.1989) was conserved among L. major, L. chagasi and L. donovani.

The most significant difference observed between the three GP63 genes compared in the present study was a region of size heterogeneity within the propeptide-coding region. The propeptide-coding region of the L. donovani GP63 gene was 39 bp shorter than that of the L. major gene and 30 bp shorter than that of the L. chagasi gene (Fig 1). Since this difference occurred within a region of the protein that is presumably cleaved off during post-translational processing, short deletions or insertions may be tolerated if they do not affect the production or activity of the mature protein. PCR and Southern blot analyses
demonstrated that there are actually two classes of GP63 genes in the *L. donovani* genome that can be discriminated on the basis of the size of their propeptide-coding region. Designated as GP63-Pro58 and GP63-Pro48, these genes code for proteins having propeptide regions of 58 residues and 48 residues respectively. Both classes of genes are expressed at the RNA level in promastigotes, however, there is no direct evidence to suggest whether or not both messages are translated into functional protein. Despite the size differences within the propeptide-coding region, both GP63-Pro58 and GP63-Pro48 gene types encode the Val-Val residues known to constitute the propeptide cleavage site in *L. major* GP63 (Button and McMaster, 1988; Bouvier *et al.* 1989) and both encode a single cysteine residue which has been suggested to play a regulatory role in *L. major* GP63 metalloprotease activity (Bouvier *et al.* 1990). This proposed regulatory activity is based on a process known as the 'cysteine switch mechanism' in which a single cysteine residue binds to and thereby inactivates the zinc-binding site of a zinc metalloprotease when the protein is in the propeptide configuration (Van Wart and Birkedal-Hansen, 1990). Based on the conservation of these residues it is likely that the products of both gene types are capable of being processed to generate functional mature protein. In contrast to the *L. donovani* GP63 genes, the multiple copies of *L. major* GP63 genes were shown to be homogeneous with regards to the length of the propeptide-coding region.

Interestingly, nucleotide substitutions that occurred within the protein-coding regions of the *L. donovani, L. major* and *L. chagasi* GP63 gene sequences tended to occur in clusters rather than random positions and these clusters resulted in amino acid substitutions of up to 6 contiguous residues. These pockets of sequence diversity may represent regions of GP63 that are not structurally or functionally critical and have therefore diverged due to a lack of selective pressure to maintain the sequence. Alternatively, clusters
of sequence diversity may represent regions of the protein required for an as yet undetermined species-specific function.

The organization of the GP63 locus as a tandem array of genes with species-specific sequence differences provides an intriguing model to study the evolution of GP63 since it implies that the locus could have evolved via two distinct possible pathways. In the first pathway the ancestral *Leishmania* species would have contained multiple tandemly-linked copies of GP63 genes that diverged within each species during subsequent speciation events. The species-specific gene types could then be maintained within each species through a process of gene homogenization, such as gene conversion, resulting in multiple gene copies that are highly conserved within a species, but are at the same time unique to that species. A second possible pathway begins with an ancestral *Leishmania* species that contained only a single GP63 gene copy. The sequence of this single gene copy could have diverged during the evolution of separate *Leishmania* species prior to the formation of a tandem gene array within each *Leishmania* species via a process such a parallel gene amplification. The latter hypothesis would best conform with the presence of multiple copies of two distinct GP63 gene types within a single species, as was demonstrated for *L. donovani* in the present study.

Comparison of the GP63 gene sequences from two Old World species, *L. major* and *L. donovani* with the reported sequence for a New World species, *L. chagasi*, suggested that *L. donovani* and *L. chagasi* are more closely related to each other than either is to *L. major*. This was most clearly evident in the 3' untranslated region where there would be less selective pressure to maintain sequence conservation. In addition, the extra 30 bp present in the propeptide-coding region of the *L. donovani* GP63-Pro58 gene was identical in sequence to the propeptide-coding region of the *L. chagasi* GP63 gene.
Due to the high level of conservation observed between \textit{L. donovani} and \textit{L. chagasi} GP63 gene sequences and because \textit{L. chagasi} is the only New World strain that causes visceral leishmaniasis, it is likely that \textit{L. chagasi} resulted from the direct transmission of \textit{L. donovani} from the Old World to the New World and subsequent speciation rather than from the speciation of New World \textit{Leishmania} species.

The \textit{L. major} GP63 locus, which contains five GP63 genes in a direct head to tail tandem array and an additional single dispersed gene copy (Button and McMaster, 1988; Button et. al. 1989), is the simplest of the GP63 loci to be characterized to date. As demonstrated in the present study, the \textit{L. donovani} GP63 locus is considerably more complex than that of \textit{L. major}, containing at least seven GP63 genes arranged in a tandem array and at least three genes which are dispersed from the tandem array. In addition, the tandemly linked GP63 genes of \textit{L. donovani} are not homogeneous as the last two genes of the array were of the GP63-Pro58 class whereas the GP63 genes located directly upstream in the array were of the GP63-Pro48 type. Both GP63-Pro48 and GP63-Pro58 gene types were also present as dispersed gene copies. Thus it appears that although the GP63 genes are highly conserved both within and across species, in \textit{L. donovani} there are GP63 genes encoding proteins of different sizes. Size heterogeneity amongst GP63 transcripts has also been observed in \textit{L. chagasi} (Wilson et. al. 1989) and recently, GP63 genes encoding proteins with divergent carboxy terminal domains have been cloned from \textit{L. chagasi} (Ramamoorthy et. al. 1992), \textit{L. mexicana} (Medina-Acosta et. al. 1993b) and \textit{L. guyanensis} (Steinkraus and Langer, 1992). These divergent GP63 genes are differentially expressed across species and their transcription may be developmentally regulated.

The presence of multiple gene copies arranged as directly repeated tandem arrays is a common feature in the kinetoplastid protozoans (Thomashow et. al. 1983; Tschudi et.
Chapter III - Discussion

Current evidence suggests that these arrays are transcribed in a polycistronic manner using tacit single upstream promoters to facilitate the synthesis of highly abundant proteins (Imboden et al. 1987; Ben Amar et al. 1988; Muhich and Boothroyd, 1988; Cross, 1990), however, polycistronic precursors have not been identified, possibly because they are rapidly (probably co-transcriptionally) processed into mature monocistronic mRNAs via the events of trans-splicing and polyadenylation (Huang and Van der Ploeg, 1991a). The only genes which have been directly shown to be transcribed in a polycistronic fashion are those of the VSG expression site in Trypanosomes, which are transcribed by an \( \alpha \)-amanitin resistant RNA polymerase, likely RNA pol I (Kooter and Borst, 1984; Zomerdijk et al. 1991b). Although there is no direct evidence to support polycistronic transcription of the GP63 tandem array, RNA species that are larger than single gene repeat units have been detected by Northern blot hybridization (Button et al. 1989; Medina-Acosta et al. 1993b). These large transcripts could be indicative of some type of polycistronic precursor or alternatively they might simply represent a transcript originating from one of the divergent genes described above. Interestingly, the GP63 gene alignment presented in this study indicated that the immediate 5' untranslated region of the GP63 gene was conserved across species to a much higher extent than was the 3' untranslated region. This conservation extended for approximately 200 bp and included the putative 3' trans-spliced leader acceptor site and its associated polypyrimidine tract as well as a number of pyrimidine-rich direct repeats located midway between the 3' trans-spliced leader acceptor site and the translational initiation codon. Based on the very high conservation within a non-protein coding region and on the finding that the intergenic region fulfills a minimal requirement for efficient transcription in transfection studies (Laban and Wirth, 1989; Laban et al. 1990; ten Asbroek et al. 1990;
Curotto de Lafaille et al. 1992), the 200 bp conserved element of the GP63 gene was considered to be a candidate as a site of protein/DNA interaction. The following chapter describes the functional cloning and characterization of a gene encoding an *L. major* DNA-binding protein which interacts with oligodeoxynucleotides derived from within the 200 bp conserved element of the GP63 intergenic region.
IV. THE HEXBP GENE OF *Leishmania major*

This chapter describes the functional cloning and characterization of a novel single-stranded DNA-binding protein from *L. major*. The DNA-binding characteristics of this protein, called HEXBP, were analyzed in detail and the results demonstrated that HEXBP binds single-stranded DNA in a sequence specific manner. HEXBP was shown to bind single-stranded oligodeoxynucleotides derived from the antisense strand of the immediate 5' untranslated region of the GP63 gene and the potential consequences of this binding are discussed.

A. RESULTS

1. Library Screening

Previous comparison of the GP63 genes from diverse species of *Leishmania* revealed that the 5' untranslated region of the GP63 gene is highly conserved across species whereas the sequence of the 3' untranslated region diverges a short distance past the stop codon (Miller *et. al.*1990; Webb *et. al.*1991; Ramamoorthy *et. al.*1992). Shown in Figure 7 is an alignment of the *L. donovani*, *L. chagasi* and *L. major* GP63 gene 5' untranslated regions beginning immediately upstream of the translational initiation codon (numbered +1) and ending just beyond the putative 3' *trans*-spliced leader acceptor site. This non protein-coding region of the GP63 gene exhibits greater than 90% nucleotide identity across species, suggesting that it likely plays a role in expression of the GP63 gene. In addition, the conserved 5' untranslated region precedes both terminal and internal
Chapter IV - Results

genes of the GP63 tandem array as the *L. major* sequence shown in Fig. 7 is from the first
gene of the GP63 locus (Button *et. al.*1989) whereas the *L. donovani* and *L. chagasi*
sequences represent intergenic regions within the GP63 gene tandem array (Miller *et. al.*1990; Webb *et. al.*1991). Computer analyses of this highly conserved region did not
reveal any sequence identity with the transcriptional control elements characteristic of the
pol II promoters of higher eukaryotes, however, this region did contain a number of
conserved, pyrimidine-rich hexanucleotide direct repeats (shown boxed in Fig. 7) located
between the putative 3' *trans*-spliced leader acceptor site and the translational initiation
codon. The number of hexanucleotide direct repeats varied according to species (two in *L.
major*, three in *L. chagasi* and four in *L. donovani*) however the sequence of individual
repeat units (CTCGCC) was invariant across species.

To determine whether the hexanucleotide direct repeats located upstream of the
GP63 gene represented a potential site of protein-DNA interaction, an *L. major* λgt11
genomic DNA expression library was screened to detect clones expressing functional
DNA-binding proteins (Singh *et. al.*1988; Vinson *et. al.*1988). The probe used for library
screening was a double-stranded oligodeoxynucleotide composed of five contiguous
hexanucleotide direct repeats surrounded on either side by 10 bp of flanking sequence. This
probe was generated by annealing the two synthetic oligodeoxynucleotides HEX50(+) and
HEX15(-) (see Table 2) and extension of the annealed primer/template complex in the
presence of [α\(^{32}\)P]dCTP. From a screen of 120,000 plaques, a single clone, called
λgtHEX, which remained strongly positive on secondary and tertiary screenings was
chosen for further analysis.

Lysogens of λgtHEX were prepared in *E. coli* Y1089 host cells as previously
described (Young and Davis, 1983; Huynh *et. al.*1985) for characterization of the β-
galactosidase fusion protein. Upon induction with IPTG, λgtHEX lysogens synthesized a 125 kDa β-galactosidase/HEXBP (hexamer binding protein) fusion protein of which 104 kDa could be accounted for by the vector-encoded β-galactosidase. The 21 kDa carboxyl-terminal, HEXBP portion of the fusion protein was therefore encoded by a continuation of the β-galactosidase open reading frame for approximately 600 bp into the 5 kbp *L. major* insert of λgtHEX. To facilitate isolation of the full length HEXBP gene, Southern blots of *L. major* genomic DNA were probed with a 2 kbp EcoRI/SalI restriction fragment of λgtHEX, containing the 600 bp HEXBP-coding region. A 7 kbp SstI/XbaI restriction fragment was identified which encompassed 4 kbp of sequence upstream and 3 kbp of sequence downstream from the point of β-galactosidase/HEXBP fusion in clone λgtHEX (data not shown). A 'sub-library' of size selected, SstI/XbaI digested *L. major* genomic DNA was prepared as described in Materials and Methods and colony blots of this library were hybridized with the 2 kbp EcoRI/SalI fragment of λgtHEX to identify clones containing a HEXBP gene. A positive clone, called pMHB7sx, was isolated and the insert was further subcloned into M13mp18 and M13mp19 for sequence analysis.

2. The Sequence of the *Leishmania major* HEXBP Gene

The complete nucleotide sequence of the *L. major* HEXBP gene is shown in Figure 8. The deduced 271 residue sequence of the HEXBP protein is shown above the single open reading frame. Sequence analysis of λgtHEX DNA identified the point of β-galactosidase/HEXBP fusion as nucleotide 202 of the HEXBP open reading frame (indicated by an arrow at nucleotide 792 in Fig. 8). The HEXBP protein had a predicted
molecular mass of 28,223 Da and contained nine conserved cysteine-rich motifs (underlined in Fig. 8), each 14 amino acids in length. A search of the EMBL databank revealed that the cysteine-rich motifs of HEXBP are common to a number of nucleic acid binding proteins (for review see Summers, 1991) and are representative of a class of domain known as the 'CCHC' or 'retroviral-type' zinc finger (Fig. 9). The consensus for this motif, Cys-X2.Cys-X4-His-X4-Cys, is invariant with regards to the number and spacing of cysteine and histidine residues at positions 1, 4, 9 and 14. In addition, the consensus CCHC motif contains highly conserved glycines at positions 5 and 8, an aromatic residue at position 2 or 3, a hydrophobic residue at position 10 and a positively charged residue at position 12. All nine of the cysteine-rich motifs of HEXBP conform with all aspects of the CCHC consensus and can therefore be classified as CCHC zinc finger motifs. The individual CCHC motifs of HEXBP exhibit further conservation in the form of a positively charged residue at position 3, a positively or negatively charged residue at position 6, a serine at position 11 and a negatively charged residue at position 13. The spacer regions separating the individual CCHC motifs of HEXBP ranged in size from 12 to 29 amino acids. With the exception of spacers 5 and 6, the sequence of individual spacer regions was not conserved. Spacer regions 5 and 6 were identical in size and sequence except for the last residue. All 8 spacer regions initiated with a proline and all were generally rich in glycine and charged residues.
3. Detection of HEXBP mRNA in Diverse Species of *Leishmania*

To characterize the expression of the HEXBP gene, *Leishmania* promastigote total RNA and *Trypanosoma brucei* polyA+ RNA were analyzed by Northern blot hybridization using a radiolabelled HEXBP-specific probe. A strongly hybridizing transcript, 3.2 kb in size, was detected in *L. major* and *L. mexicana* promastigote RNA and to a lesser extent in *L. donovani* promastigote RNA (Fig. 10, lanes 1, 2 and 3 respectively). A weakly hybridizing transcript of approximately 5 kb was also observed in *L. major* and *L. mexicana* RNA. No hybridizing species were evident in polyA+ RNA from *T. brucei*. A transcript size of over 3 kb implies that the HEXBP mRNA contains extensive untranslated regions, however, it has not yet been determined whether these are found as upstream or downstream extensions of the 813 bp HEXBP protein-coding region.

4. The single-stranded DNA-binding Activity of HEXBP

To characterize the DNA-binding activity of the β-galactosidase/HEXBP fusion protein, electrophoretic mobility shift assays were performed using total cell extracts of IPTG-induced λgtHEX lysogens as a source of protein. The β-galactosidase/HEXBP fusion protein was capable of binding to end-labeled, synthetic single-stranded oligodeoxynucleotides (Fig. 11, lanes 6 and 7) but was incapable of binding to the equivalent non-denatured, double-stranded oligodeoxynucleotide probe (Fig. 11, lane 8). DNA-binding activity was detected if the double-stranded oligodeoxynucleotide probe was denatured by heating immediately prior to being added to the binding reaction (Fig. 11, lane 76)
9). The identity of β-galactosidase/HEXBP as the protein responsible for the mobility shift was confirmed by the lack of any similar activity in the extracts of IPTG-induced wild type λ gt11 lysogens (Fig. 11, lanes 2 through 5). The single-strand specific binding activity of the β-galactosidase/HEXBP fusion protein was surprising in light of the fact that λ gtHEX had originally been cloned using an oligodeoxynucleotide probe which was presumed to be double-stranded. However, the single-stranded binding activity of HEXBP is consistent with previous studies demonstrating that proteins containing the CCHC zinc finger motif function by binding to single-stranded nucleic acids (Summers, 1991).

To determine whether binding of single-stranded nucleic acids by the β-galactosidase/HEXBP fusion protein was sequence specific, Southwestern hybridization assays were performed using several different end-labeled oligodeoxynucleotide probes. Proteins in total cell extracts of wild type λ gt11 and λ gtHEX lysogens were separated by SDS-PAGE and either stained with Coomassie blue, or transferred to nitrocellulose membranes and assayed for DNA-binding activity as described above for library screening. Triplicate blots were assayed for binding to three synthetic oligodeoxynucleotide probes; 1) gp63-5'-50(+), equivalent to the sense strand of the L. major GP63 gene 5' untranslated region, nucleotides -69 to -20, 2) gp63-5'-50(-), the reverse complement of gp63-5'-50(+) and 3) gp63-pro-50, derived from within the GP63 protein-coding region, sense strand, nucleotides 307 to 356). Synthesis of the β-galactosidase/HEXBP fusion protein was induced to sufficiently high levels for the protein to be clearly visible on Coomassie blue stained gels of total lysogen extracts (indicated by an arrow in Fig. 12, lane 3). A protein specific to λ gtHEX lysogen extracts and similar in size to the β-galactosidase/HEXBP fusion protein bound to the gp63-5'-50(-) probe (Fig. 12, lane 5) but not to the gp63-5'-50(+) or gp63-pro-50 probes (Fig. 12, lanes 7 and 9 respectively). The gp63-5'-50(+)
probe was bound by a 75 kDa protein common to both wild type λgt11 and λgtHEX lysogen extracts (Fig. 12, lanes 6 and 7). This protein likely represents an *E. coli* DNA-binding protein which was capable of binding to this particular probe. These results suggested that the β-galactosidase/HEXBP fusion protein binds specifically to the antisense strand of the GP63 gene 5' untranslated region. The hexanucleotide repeats on the antisense strand (GGCGAG) are purine rich as opposed to the pyrimidine-rich repeats shown in Figure 7.

Recombinant HEXBP was synthesized using the pET-3a bacterial expression system to assess the DNA binding activity of the full length non-fusion protein and to avoid any potential influence on DNA binding activity by the β-galactosidase portion of the fusion protein. Shown in Fig. 13 is a Southwestern blot assay of recombinant fusion and non-fusion HEXBP incubated with the end-labeled oligodeoxynucleotide probe gp63-5'-50(-). DNA-binding proteins of the expected sizes were observed in lanes containing extracts of cells expressing a β-galactosidase/HEXBP fusion protein (125 kDa), non-fusion HEXBP (28 kDa) or HEXBP/gene 10 fusion protein (30 kDa) (Fig. 13, lanes 2, 4 and 6 respectively). No significant background binding activity was observed in these extracts or in the extracts of control cells containing the equivalent vectors only (Fig. 13, lanes 1, 3 and 5). Interestingly, bands which were the equivalent molecular weight of homodimers were also detected in extracts of cells expressing the non-fusion HEXBP and HEXBP/gene 10 fusion proteins but not the β-galactosidase/HEXBP fusion protein. These dimers likely represent incomplete reduction and inappropriate intermolecular disulfide crosslink formation between one or more of the 27 cysteine residues of HEXBP, however, the potential significance of dimer formation was not directly addressed in this study. The large size of the β-galactosidase/HEXBP fusion protein likely prevented dimer
formation. Further characterization of HEXBP synthesis in the pET-3a system showed that non-fusion HEXBP was expressed in a soluble form whereas the HEXBP/gene 10 fusion protein was localized primarily to the inclusion body fraction (data not shown). Total extracts of cells expressing the non-fusion HEXBP were therefore used as a source of protein for all remaining experiments.

The DNA-binding activity of recombinant non-fusion HEXBP was also characterized by competitive electrophoretic gel mobility shift assays. As previously observed with the β-galactosidase/HEXBP fusion protein, the synthetic single-stranded oligodeoxynucleotide gp63-5'-50(-) was efficiently bound by the non-fusion HEXBP (Fig. 14, lane 3). Furthermore, the binding activity of non-fusion HEXBP was also sequence specific as it could not be competed out when excess unlabeled gp63-pro-50 oligonucleotide, single-stranded M13mp18 DNA, *L. major* total RNA or non-denatured *L. major* genomic DNA were used as competitors (Fig. 14, lanes 5 through 8, respectively). Excess unlabeled gp63-5'-50(-) oligonucleotide was the only effective competitor of binding (Fig. 14, lane 4). The identity of HEXBP as the protein responsible for the specific mobility shift was confirmed by the lack of any similar mobility shift in lanes containing extracts of control cells (Fig. 14, lane 2).

To determine the optimal ionic concentration for HEXBP DNA binding activity, the formation of the HEXBP/oligodeoxynucleotide complex was assayed under conditions of varying ionic strength. Electrophoretic mobility shift assays were used to demonstrate that the total amount of gp63-5'-50(-) oligonucleotide probe bound by HEXBP remained relatively constant when either KCl or NaCl concentrations were varied from 0 to 250 mM (Fig. 15, lanes 3 to 7 and 8 to 12 respectively). Interestingly, at 100 and 150 mM KCl the protein-DNA complex changed from a predominantly monomer form to what appeared to
be some type of multimeric complex. Identical experiments using the β-galactosidase/HEXBP fusion protein provided similar results with regards to stability of the complex, however, the fusion protein was not capable of forming a multimeric complex at 150 mM KCl (data not shown). These results are consistent with the sequence-specific binding of DNA by HEXBP since weak, non-specific interactions are, in general, dissociable under conditions of increasing ionic strength (Lane et. al.1992). In contrast to HEXBP DNA-binding, the binding of polynucleotides by the CCHC-containing retroviral nucleocapsid protein (NCP) has been previously demonstrated to be NaCl concentration dependent (Karpel et. al.1987).

The detection of HEXBP in L. major total protein extracts using gel mobility shift or Southwestern blot assays was complicated by the presence of multiple DNA-binding proteins, however, a protein of the expected size and binding specificity was observed using a UV cross-linking assay. In this assay, binding reactions were identical to those used for electrophoretic mobility shift assays except that complexes were covalently cross-linked using short wave (302 nm) UV light. The resulting end-labeled oligonucleotide/protein complexes were then analyzed using SDS-PAGE and autoradiography. UV cross-linking of bacterially synthesized non-fusion HEXBP and end-labeled gp63-5'-50(-) oligonucleotide resulted in the formation of a prominent complex with a molecular weight equivalent to the sum of the two components (44 kDa) (Fig. 16, lane 3). In addition to two lower molecular weight complexes, the 44 kDa complex was also evident when L. major promastigote total cell extracts were used in binding reactions (Fig. 16, lane 4). Furthermore, the 44 kDa complex, but not the lower molecular weight complexes, could be specifically competed out by the addition of excess unlabeled gp63-5'-50(-) oligonucleotide (Fig. 16, lane 7). Addition of gp63-pro-50 oligonucleotide or single-
stranded M13mp18 DNA as competitor did not significantly affect the formation of any of the three complexes (Fig. 16, lanes 5 and 6 respectively).

5. DNAse I Footprint Analysis of HEXBP Single-stranded DNA Binding Activity

Analysis of the HEXBP/single-stranded DNA complex by DNAse I protection assay was initially performed using the end-labeled oligodeoxynucleotide gp63-5'-50(-) probe and varying the duration of digestion with DNAse I in the presence or absence of HEXBP. After 1, 2 or 4 minutes of digestion, the DNAse I cleavage ladders produced in the presence of control extracts or extracts containing HEXBP were not significantly different from a DNAse I cleavage ladder produced in the total absence of protein (Fig. 17, compare lanes 3 to 8 with lane 2). At these time points the majority of the probe remained full length irrespective of the extract used. However, after 8 minutes of DNAse I digestion, gp63-5'-50(-) probe incubated with control extract was cleaved to a much greater extent than at earlier time points and much less full length probe was detectable (Fig. 17, lane 9). This likely represented a lag in the initial rate of DNAse I digestion. In contrast, the DNAse I cleavage ladder produced after 8 minutes of digestion in the presence of a HEXBP-containing extract was indistinguishable from the cleavage ladders produced at earlier time points and the majority of probe remained essentially intact (Fig. 17, lane 10). Extending the time of DNAse I digestion to 16 minutes did not result in any further cleavage of the gp63-5'-50(-) probe when incubated in the presence of HEXBP (Fig. 17, lane 12). DNAse I digestion in the presence of control extract for 16 minutes resulted in cleavage of the
gp63-5'-50(-) probe into increasingly smaller fragments, most of which were less than 10 bases in length and not resolvable on this gel (Fig. 17, lane 11). Protection of the gp63-5'-50(-) probe could be competed out by addition of excess unlabeled gp63-5'-50(-) oligodeoxynucleotide (Fig. 17, lane 14) however addition of an equivalent amount of unrelated oligodeoxynucleotide (gp63-pro-50) had no effect on protection (Fig. 17, lane 13). These results suggested that HEXBP was capable of protecting the entire gp63-5'-50(-) oligodeoxynucleotide from digestion with DNAse I. Protection of such a large region was surprising considering the relatively small size of the HEXBP protein (28 kDa). Interestingly, although the majority of gp63-5'-50(-) was fully protected from DNAse I by HEXBP and remained intact after 16 minutes of DNAse I digestion, a limited amount of cleavage of the probe was evident at all time points. However, the ladder of fragments produced by limited DNAse I cleavage was essentially identical at all time points subsequent to 2 minutes, indicating that these cleavage fragments were resistant to any further digestion by DNAse I.

To address the possibility that HEXBP bound DNA in a non-sequence specific manner and thereby fortuitously protected the full length gp63-5'-50(-) oligodeoxynucleotide from DNAse I digestion, a longer single-stranded probe was synthesized in which gp63-5'-50(-) was flanked by sequences derived from the pBluescript plasmid vector. This probe was generated by cloning the gp63-5'-50(-) oligodeoxynucleotide into pBluescript to generate the plasmid construct pBSHEX45. PCR was then used to amplify a 359 bp product from pBSHEX45 that spanned the gp63-5'-50(-) insert. A 327 bp PvuII fragment of the PCR product was digested with λ exonuclease as described in Materials and Methods and the resulting 327 base single-stranded molecule (BSHEX-327) was end-labeled using T4 polynucleotide kinase for use
as a probe in DNase I protection assays. An equivalent 374 base single-stranded probe containing only vector sequences (BS-374) was synthesized for use as a control probe.

DNase I protection assays using total cell extracts of bacteria expressing *L. major* HEXBP resulted in the protection of specific regions within the BSHEX-327 single-stranded probe (Fig. 18A, lane 3). No similar protection was observed in control assays using total cell extracts of bacteria containing the pET-3a vector alone (Fig. 18A, lane 2), confirming the identity of HEXBP as the protein responsible for protection from DNase I. HEXBP protected two regions of BSHEX-327 from DNase I digestion, indicated by the bars labeled A and B to the right of Fig 18A. Protected site A was approximately 69 bases in size and spanned the entire region corresponding to the gp63-5'-'50(-) insert. Protection of this 69 base region correlated with full protection of the entire gp63-5'-'50(-) oligodeoxynucleotide as described above. In addition, protected site A of BSHEX-327 extended past the boundaries of the gp63-5'-'50(-) insert for about 2 bases in the 5' direction and 22 bases in the 3' direction. The second region of BSHEX-327 protected from DNase I digestion by HEXBP (site B) was located 3' of site A and was approximately 26 bases in size. This site contained pBluescript vector sequence only and shared no apparent sequence identity with protected site A. DNase I hypersensitive sites were observed at the 3' boundaries of both protected regions. Protection at both sites could be competed out by addition of excess unlabeled gp63-5'-'50(-) oligodeoxynucleotide (Fig. 18A, lane 5) however addition of an equivalent amount of unrelated oligodeoxynucleotide (gp63-pro-50) had no effect on formation of either of the protected sites (Fig. 18A, lane 4).

Further evidence supporting the sequence specific interaction of HEXBP with single-stranded DNA was provided by substituting the control probe BS-374 (containing vector sequence only) for the BSHEX-327 probe. Digestion of BS-374 in the presence of
HEXBP resulted in the specific protection of the same 26 base vector sequence described above as site B of BSHEX-327 (Fig. 18B, lane 3). No other region of BS-374 was protected by HEXBP. Protection of site B on BS-374 could also be competed out by the addition of excess gp63-5'-50(-) oligodeoxynucleotide but not by the addition of an unrelated oligodeoxynucleotide (Fig. 18B lanes 5 and 4 respectively). Control assays using extracts of cells containing the pET-3a vector alone confirmed the identity of HEXBP as the protein responsible for specific protection of the BS-374 probe (Fig. 18B, lane 2).

Protection of specific regions of the BSHEX-327 probe from digestion with DNAse I confirmed that recombinant HEXBP binds single-stranded DNA in a sequence specific manner and implied that the region of the GP63 gene 5' flanking region encompassed by the oligodeoxynucleotide gp63-5'-50(-) is a potential site of HEXBP/DNA interaction in vivo. To determine whether this was the only HEXBP binding site within the GP63 gene 5' flanking region, a single-stranded DNA probe that spanned 462 bases of the antisense strand of the GP63 gene 5' untranslated region (bases -396 to 66 relative to the ATG translational initiation codon) was synthesized. This probe, called gp63-5'-462, was generated from the plasmid pLMS10-1-3 (containing a single GP63 coding region and 1400 bp of 5' untranslated sequence) (Button and McMaster, 1988), using the PCR approach described previously for the generation of the BSHEX-327 probe.

Binding to the gp63-5'-462 probe was initially characterized using electrophoretic gel mobility shift assays. The addition of decreasing amounts of HEXBP to binding reactions resulted in a concomitant increase in the mobility of protein/DNA complexes on non-denaturing PAGE gels (Fig. 19, lanes 2 to 8). The protein concentration dependent mobility of complexes was interpreted as evidence that the gp63-5'-462 probe contained multiple HEXBP binding sites (slower migrating complexes represent high binding site
occupancy at higher HEXBP concentrations whereas faster migrating complexes represent reduced binding site occupancy at lower HEXBP concentrations). Formation of complexes could be completely inhibited by the addition of 1000 ng of gp63-5'-50(-) oligodeoxynucleotide (equivalent to approximately 5000 fold molar excess) (Fig. 19, lane 10) however the addition of 10 ng (50 fold molar excess) or less of unlabeled gp63-5'-50(-) had no effect (Fig. 19, lanes 12 and 13). Addition of 100 ng of gp63-5'-50(-) as competitor resulted in the formation of a complex having intermediate mobility (Fig. 19, lane 11). The latter result was interpreted as further evidence for the presence of multiple HEXBP binding sites within the gp63-5'-462 probe.

Binding to the gp63-5'-462 probe was further characterized using DNAse I protection assays. Protection assays were performed using three different dilutions of HEXBP-containing bacterial extracts. At the lowest concentration of protein tested (1:100 dilution of the extract), the DNAse I cleavage patterns produced in the presence of control extracts (containing vector only) were indistinguishable from those produced in the presence of HEXBP-containing extracts (Fig. 20, lanes 2 and 3). At a 1:10 dilution of extract, multiple regions of protection within the gp63-5'-462 probe were evident in reactions containing HEXBP (Fig. 20, lane 5). These regions were designated as protected sites A through D and each protected site was approximately 25 to 30 bases in size. Site A overlapped with the previously described protected site A of the BSHEX-327 probe and included a portion of the sequence encompassed by the gp63-5'-50(-) oligodeoxynucleotide. In contrast to the complete protection of the gp63-5'-50(-) sequence observed in the previous experiments, the 3' end of the gp63-5'-50(-) sequence was protected to a much greater extent than was the 5' end. Sites B through D were located 3' of site A and all sites were approximately equidistant from one another (approximately 20
to 30 bases). Increasing the HEXBP concentration five fold did not result in any further protection of the gp63-5'-462 probe (Fig. 20, lane 7). Protection at all sites could be competed out by the addition of excess unlabeled gp63-5'-50(-) oligodeoxynucleotide (Fig. 20, lane 8). Addition of excess unrelated oligonucleotide (gp63-pro-50) had no effect (Fig. 20, lane 9). None of the protected sites were evident when assays were conducted in the presence of control extracts (BL21 (DE3) cells containing the pET-3a vector only) (Fig. 20, lanes 2, 4 and 6).

To determine whether protected sites A through D of gp63-5'-462 represented HEXBP binding sites of significantly different affinity, DNAse I protection assays were conducted in the presence of increasing concentrations of unlabeled gp63-5'-50(-) oligodeoxynucleotide as a specific competitor. Protection of sites A through D by HEXBP was unaffected by the addition of either 5 or 50 fold molar excess of gp63-5'-50(-) oligodeoxynucleotide (Fig. 21, lanes 4 and 5 respectively). This implied that the amount of HEXBP protein in these assays was in excess. However, in the presence of 500 fold molar excess of gp63-5'-50(-) oligodeoxynucleotide, protection at all four sites was eliminated (Fig. 21, lane 6). An increase in the amount of gp63-5'-50(-) competitor to 5000 fold molar excess had no further effect (Fig. 21, lane 7). These results implied that the HEXBP binding affinity was similar at all four sites, since deprotection occurred at all sites at an equal concentration of competitor. The exact sequence of protected sites A and B were determined by comparison to a Maxam and Gilbert A+G sequencing ladder. Protected site A contained two hexanucleotide direct repeats (GGCGAG) followed immediately by two guanines. Protected site B contained two imperfect direct repeats (which matched the repeats of site A at five of six positions) followed immediately by two guanines. The result was a shared sequence element which was identical at 12 out of 14 positions forming the
consensus sequence $G\ G\ C\ G\ A/G\ G\ G\ C/A\ G\ A\ G\ G\ G$. Sites C and D were located beyond the readable portion of the sequencing ladder and the sequence of these sites was not determined.
Figure 7. Sequence alignment of the GP63 gene 5' untranslated region.

The sequences of the 5' untranslated regions of GP63 genes from *L. donovani* (upper), *L. chagasi* (middle) and *L. major* (lower) were aligned with respect to the ATG translational initiation codon. Conserved nucleotides are indicated by a dot, absent nucleotides are indicated by a dash. A region of tandemly repeated hexanucleotides is shown in the boxed area beginning a nucleotide position -43. The conserved hexanucleotides are present as four repeat units in *L. donovani*, three repeat units in *L. chagasi* and two repeat units in *L. major*. The putative 3' trans-spliced leader acceptor site is indicated by an arrow at nucleotide position -141.
Figure 8. Nucleotide and predicted amino acid sequence of the *L. major* DNA-binding protein HEXBP.

Nucleotides are numbered with position 592 corresponding to the first nucleotide of the ATG translational initiation codon. The predicted amino acid sequence of HEXBP is shown above the 813 bp open reading frame. The nine CCHC zinc finger motifs of HEXBP are underlined. The point of β-galactosidase/HEXBP fusion in clone λgtHEX is indicated by an arrow at nucleotide position 792. The HEXBP sequence has been submitted to GenBank™ under accession number M94390.

---

<table>
<thead>
<tr>
<th>Nucleotide Sequence</th>
<th>Predicted Amino Acid Sequence</th>
</tr>
</thead>
</table>
| AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Figure 9. Comparison of proteins containing the CCHC zinc finger motif. Conserved cysteine and histidine residues are shown in bold type. FeLV, RSV, HIV-1, Copia and CaMV are from the nucleocapsid proteins of Feline Leukemia Virus, Rous Sarcoma Virus, Human Immunodeficiency Virus, transposable element Copia and the Cauliflower Mosaic Virus respectively (for review see Summers, 1991). Xpo is a developmentally regulated gene from *Xenopus* (Sato and Sargent, 1991), SLU7 is involved with 3' splice site selection during *cis*-splicing in *Saccharomyces* (Frank and Guthrie, 1992) and CNBP is human cellular nucleic acid binding protein (Rajavashisth *et al*. 1989).
Chapter IV- Results

FeLV  Cys Ala Tyr Cys Lys Glu Lys Gly His Trp Val Arg Asp Cys

RSV  Cys Tyr Thr Cys Gly Ser Pro Gly His Tyr Gln Ala Gln Cys
  Cys Gln Leu Cys Asn Gly Met Gly His Asn Ala Lys Gln Cys

HIV-1  Cys Phe Asn Cys Gly Lys Glu Gly His Ile Ala Arg Asn Cys
  Cys Trp Lys Cys Gly Lys Glu Gly His Gln Met Lys Asp Cys

Copia  Cys His His Cys Gly Arg Glu Gly His Ile Lys Lys Asp Cys

CaMV  Cys Trp Ile Cys Asn Ile Glu Gly His Tyr Ala Asn Glu Cys

Xpo  Cys Tyr Ser Cys Gly Lys Tyr Gly His Ile Ala Arg Phe Cys

SU7  Cys Arg Asn Cys Gly Glu Ala Gly His Lys Glu Lys Asp Cys

QNEP  1 Cys Phe Lys Cys Gly Arg Ser Gly His Trp Ala Arg Glu Cys Pro
  2 Cys Tyr Arg Cys Gly Glu Ser Gly His Leu Ala Lys Asp Cys Asp
  3 Cys Tyr Asn Cys Gly Arg Gly Gly His Ile Ala Lys Asp Cys Lys
  4 Cys Tyr Asn Cys Gly Lys Pro Gly His Leu Ala Arg Asp Cys Asp
  5 Cys Tyr Ser Cys Gly Glu Phe Gly His Ile Gln Lys Asp Cys Thr
  6 Cys Tyr Arg Cys Gly Glu Thr Gly His Val Ala Ile Asn Cys Ser
  7 Cys Tyr Arg Cys Gly Glu Ser Gly His Leu Ala Arg Glu Cys Thr

HEXEP  1 Cys Arg Asn Cys Gly Lys Glu Gly His Tyr Ala Arg Glu Cys Pro
  2 Cys Phe Arg Cys Gly Glu Glu Gly His Met Ser Arg Glu Cys Pro
  3 Cys Phe Arg Cys Gly Glu Ala Gly His Met Ser Arg Asp Cys Pro
  4 Cys Tyr Lys Cys Gly Gln Glu Gly His Leu Ser Arg Asp Cys Pro
  5 Cys Tyr Lys Cys Gly Asp Ala Gly His Ile Ser Arg Asp Cys Pro
  6 Cys Tyr Lys Cys Gly Asp Ala Gly His Ile Ser Arg Asp Cys Pro
  7 Cys Tyr Lys Cys Gly Glu Ser Gly His Met Ser Arg Glu Cys Pro
  8 Cys Tyr Lys Cys Gly Lys Pro Gly His Ile Ser Arg Glu Cys Pro
  9 Cys Tyr Lys Cys Gly Glu Ala Gly His Ile Ser Arg Asp Cys Pro
Figure 10. Analysis of HEXBP gene expression in diverse species of *Leishmania*.
Total RNA (6 μg per lane) of *L. major* (lane A), *L. mexicana* (lane B) and *L. donovani* (lane C), isolated from log phase promastigotes or polyA+ RNA from procyclic stage *T. brucei* (1 μg) was separated by electrophoresis through formaldehyde/agarose denaturing gels and transferred to Hybond-N for Northern blot hybridization analysis. The blot was hybridized with a probe specific for the 5' end of the HEXBP protein-coding region (394 bp EcoRV/NotI restriction fragment of pMHB7sx, nucleotides 578 to 971 in Fig. 2) labeled by random hexanucleotide priming. The size of the *L. major* HEXBP RNA (indicated on the left in kb) was determined by comparison to a co-electrophoresed RNA molecular weight ladder (BRL, Gaithersburg, Maryland).
Figure 11. Electrophoretic mobility shift assays of HEXBP single-stranded vs. double-stranded DNA-binding activity.

Total protein extracts of IPTG-induced wild type $\lambda$ gt11 (lanes 2 through 5) or $\lambda$ gtHEX lysogens (lanes 6 through 9) were assayed for binding to single-stranded (HEX50(-)) or double-stranded (HEX50(+) / HEX50(-)) oligonucleotide probes using the electrophoretic mobility shift assay. Extracts were incubated with either single-stranded oligonucleotides (lanes 2, 3, 6 and 7) or double-stranded oligonucleotides (lanes 4, 5, 8 and 9) prior to electrophoresis on 4% nondenaturing polyacrylamide gels. Oligonucleotide probes were either added directly to the binding reactions (lanes 2, 4, 6 and 8) or were denatured by heating in a boiling water bath prior to addition to the binding reactions (lanes 3, 5, 7 and 9). Arrows marked F or B on the right side of the figure indicate free and bound oligonucleotides respectively. Lane 1 is a control showing single-stranded probe incubated in the absence of protein extract.
Figure 12. Sequence specificity of HEXBP binding to single-stranded oligodeoxynucleotides.
Total protein extracts of IPTG-induced wild type λ gt11 or λ gtHEX lysogens fractionated on 10% SDS-polyacrylamide gels were either stained for total protein using Coomassie Blue (panel A) or were blotted to nitrocellulose filters for analysis of DNA-binding specificity (panels B, C and D). Proteins blotted to filters were denatured in a solution of 6 M GuHCl and subsequently renatured by a series of rapid serial dilutions to remove the GuHCl. Filters were then incubated overnight in a solution containing one of the following end-labeled oligonucleotides; gp63-5'-50(-) (panel B), gp63-5'-50(+) (panel C), or gp63-pro-50 (panel D). The filters were washed briefly and subjected to autoradiography overnight to detect proteins which have bound radiolabeled oligonucleotides. The presence of the β-galactosidase/HEXBP fusion protein on the stained gel is indicated by the arrow beside panel A. The sizes of co-electrophoresed molecular weight markers (in kDa) are indicated to the left of each panel.
Figure 13. Comparison of the DNA-binding activity of bacterially synthesized fusion and non-fusion HEXBP by Southwestern blot analysis. Total protein extracts of IPTG-induced λ gt11 lysogens (lane 1) λ gtHEX lysogens (producing β-galactosidase/HEXBP fusion protein) (lane 2), BL21(DE3) cells containing the pET-3a vector alone (lane 3) BL21(DE3) cells containing the construct pHB3A (producing non-fusion HEXBP) (lane 4) BL21(DE3) cells containing the pET-3a vector alone (lane 5) or BL21(DE3) cells containing the construct pHB3Afus (producing HEXBP/gene 10 fusion protein) (lane 6) were separated by SDS-PAGE and transferred to nitrocellulose filters. Single-strand specific DNA-binding proteins were detected by hybridizing filters with the end-labeled oligodeoxynucleotide probe gp63-5'-50(-). The positions of co-electrophoresed protein molecular weight markers (in kDa) are indicated on the left.
Figure 14. Competitive gel mobility shift assay of non-fusion HEXBP DNA-binding activity.

Binding of the full length non-fusion HEXBP to end-labeled gp63-5'-50(−) oligonucleotide was determined using electrophoretic mobility shift assays. Total protein extracts of bacterial cells expressing non-fusion HEXBP were used a source of protein (lanes 3-8). Excess unlabeled competitor was added to the binding reactions as follows; 500 ng of gp63-5'-50(−) (lane 4), 500 ng of gp63-pro-50 (lane 5), 400 ng of single-stranded M13 DNA (lane 6), 1 μg of *L. major* total RNA (lane 7) or 1 μg of *L. major* genomic DNA (lane 8). Binding reactions containing no added protein (lane 1) or total protein extracts of bacterial cells containing the pET3A vector only (lane 2) were included as negative controls.
Figure 15. Stability of HEXBP single-stranded DNA-binding activity at increasing ionic concentrations.
The binding of non-fusion HEXBP to end-labeled gp63-5'-50(-) oligodeoxynucleotide under varying ionic conditions was determined using electrophoretic mobility shift assays. Binding reactions were supplemented with either NaCl or KCl to the following final concentrations; 50 mM, 100 mM, 150 mM, 200 mM or 250 mM NaCl (lanes 3-7 respectively) and 50 mM, 100 mM, 150 mM, 200 mM or 250 mM KCl (lanes 8-12 respectively). Binding reactions containing no added protein (lane 1) were included as a negative control.
Figure 16. Detection of HEXBP DNA-binding activity in *L. major* promastigote extracts.

Total protein extracts of bacterial cells expressing non-fusion HEXBP (lane 3) or *L. major* promastigotes (lanes 4-7) were used a source of protein for UV cross-linking assays. Binding reactions containing end-labeled gp63-5'-50(-) oligonucleotide were prepared as previously described for electrophoretic mobility shift assays. Subsequent to binding, complexes were covalently cross-linked using short wave (302 nm) UV light and analyzed on 12% SDS PAGE gels. Excess unlabeled competitors were added to the binding reactions as follows; 500 ng of gp63-pro-50 (lane 5), 400 ng of single-stranded M13 (lane 6), 500 ng of gp63-5'-50(-) (lane 7). Binding reactions containing no added protein (lane 1) or total protein extracts of bacterial cells containing the pET-3a vector only (lane 2) were included as negative controls.
Figure 17. DNAse I protection of gp63-5‘-50(-) by HEXBP.
The end-labeled synthetic oligodeoxynucleotide probe gp63-5‘-50(-) was incubated with control extracts (BL21 (DE3) cells containing the pET-3a vector only) (lanes 3, 5, 7, 9 and 11) or with extracts of cells expressing recombinant HEXBP (lanes 4, 6, 8, 10, 12-14) as previously described for gel mobility shift assays. Binding reactions were treated with DNAse I for 1 minute (lanes 3 and 4), 2 minutes (lanes 2, 5 and 6), 4 minutes (lanes 7 and 8), 8 minutes (lanes 9 and 10) or 16 minutes (lanes 11 to 14) and cleavage products were resolved on 16% denaturing polyacrylamide gels. Excess unlabeled oligodeoxynucleotides gp63-pro-50 (lane 13) or gp63-5‘-50(-) (lane 14) were added as specific competitors. Undigested probe was run as a control (lane 1) and a Maxam and Gilbert A+G sequencing ladder of gp63-5‘-50(-) (lane M) was used as a marker.
Figure 18. DNAse I protection of the BSHEX-327 and BS-374 single-stranded probes by HEXBP.

Binding of HEXBP to the long single-stranded oligonucleotide probes BSHEX-327 (panel A) and BS-374 (panel B) was assayed by DNAse I protection. The position of the internal gp63-5' -50(-) sequence is indicated by the bar to the left of panel A (striped boxes indicate the location of the hexamer repeats). Regions of the probes protected from DNAse I digestion are indicated by the boxes marked A or B. The sequences of the protected sites A and B of the BSHEX-327 probe are as shown (sequences corresponding to the oligodeoxynucleotide gp63-5'-50(-) are underlined, hexamer repeats are shown in italics). The sequence of site B on the BS-374 probe is identical to site B of the BSHEX-327 probe. The end-labeled probes were incubated in the absence of protein (lane 1), with control extracts (BL21 (DE3) cells containing the pET-3a vector only) (lane 2) or with extracts of cells expressing recombinant HEXBP (lanes 3 to 5). Excess unlabeled oligodeoxynucleotides gp63-pro-50 (lane 4) or gp63-5' -50(-) (lane 5) were added as specific competitors. Maxam and Gilbert A+G sequencing ladders of BSHEX-327 and BS-374 (lane M) were used as markers.
Chapter IV - Results

A

5'-AATCCATACTTCTCGTCTCTCCG-3'

5'-ATGGTCATAGCTGTTTCCTGTGTG-3'

B

5'-CCCGCTGTGGGCAGTGriGfITGIGGTGGCGAG GGCGAG

GriCTTGIfiCAGGTACCCAOCTTTIGTICCCM

11
Figure 19. Binding of the gp63-5'-462 single-stranded probe by HEXBP in electrophoretic mobility shift assays.
End-labeled gp63-5'-462 probe was incubated in the absence of protein (lane 1) or with extracts of cells expressing recombinant HEXBP (lanes 2-13). Extracts were added to the binding reactions undiluted (lanes 2 and 9 to 13) or were diluted in extract buffer to a maximum of 1:500 as indicated. Excess unlabeled oligodeoxynucleotide gp63-5'-50(-) was added in the amounts indicated as a specific competitor. Protein/DNA complexes were resolved on 3% nondenaturing polyacrylamide gels and detected by autoradiography.
Figure 20. DNase I protection of the gp63-5'-462 single-stranded probe by HEXBP - I.

The end-labeled, single-stranded probe gp63-5'-462 was incubated in the absence of protein (lane 1), with control extracts (BL21 (DE3) cells containing the pET-3a vector only) (lanes 2, 4 and 6) or with extracts of cells expressing recombinant HEXBP (lanes 3, 5 and 7 to 9). Excess unlabeled oligodeoxynucleotides gp63-5'-50(-) (lane 8) or gp63-pro-50 (lane 9) were added as specific competitors. The position of the internal gp63-5'-50(-) sequence is indicated by the bar to the left (striped boxes indicate the location of the hexamer repeats). Regions of the probe protected from DNase I digestion are indicated by the boxes marked A to D. A Maxam and Gilbert A+G sequencing ladder of gp63-5'-462 (lane M) was used as a marker.
Chapter IV - Results
Figure 21. DNAse I protection of the gp63-5'-462 single-stranded probe by HEXBP - II.
The end-labeled, single-stranded probe gp63-5'-462 was incubated in the absence of protein (lane 1), with control extracts (BL21 (DE3) cells containing the pET-3a vector only) (lane 2) or with extracts of cells expressing recombinant HEXBP (lanes 3 to 7). Unlabeled oligodeoxynucleotide gp63-5'-50(-), 1 ng (lane 4), 10 ng (lane 5), 100 ng (lane 6) or 1000 ng (lane 7) was added as a specific competitor. The position of the internal gp63-5'-50(-) sequence is indicated by the bar to the left (striped boxes indicate the location of the hexamer repeats). Regions of the probe protected from DNAse I digestion are indicated by the boxes marked A to D. The sequences of the protected sites A and B are as shown (hexamer repeats (site A) and hexamer repeat-like sequences (site B) are shown in italics). A number of non-contiguous partially protected bands which may constitute another protected site are indicated by arrowheads. A Maxam and Gilbert A+G sequencing ladder of gp63-5'-462 (lane M) was used as a marker.
B. DISCUSSION

*L. major* HEXBP represents an important addition to the family of proteins containing the CCHC-type zinc-binding domain since it contains the highest number of CCHC motifs known to occur in a single protein. The CCHC zinc finger motif was first identified in the retroviral nucleocapsid protein (NCP), a proteolytic product of the GAG polyprotein (Henderson *et al.* 1981; Summers, 1991), and has subsequently been found as either one or two copies in the NCPs of every known retrovirus. Although the exact role of NCP in the retrovirus life cycle remains controversial, it is known to be capable of binding single-stranded nucleic acids and it has been implicated in the processes of RNA genome dimerization, RNA packaging and in the annealing of the replication primer tRNA prior to reverse transcription (Katz and Jentoft, 1989; Summers, 1991). The CCHC motif is also present in the NCP of the *Drosophila* transposable element copia (Mount and Rubin, 1985), the coat protein of the cauliflower mosaic virus (Covey, 1986) and in the product of a developmentally regulated gene (Xpo) from Xenopus (Sato and Sargent, 1991). In addition, a gene has recently been cloned from *Saccharomyces cerevisiae* which encodes a CCHC-containing protein involved in the mRNA cis-splicing process (Frank and Guthrie, 1992). This protein, SLU7, interacts with the U5 snRNA and is required for selection of an appropriate 3′ splice site. *L. major* HEXBP shares the greatest similarity with a human single-stranded DNA binding protein called CNBP (cellular nucleic acid binding protein). CNBP contains seven CCHC-type zinc finger motifs (Rajavashisth *et al.* 1989) and exhibits a sequence specific single-stranded DNA binding activity that is very similar to the *in vitro* DNA binding activity of HEXBP observed in the present study. CNBP binds to an oligodeoxynucleotide which corresponds to one strand of the consensus sequence known
as the sterol regulatory element and is thought to function as a transcriptional repressor of several genes involved in sterol metabolism. A protein that is structurally similar to CNBP has recently been cloned from the yeast *Schizosaccharomyces pombe* (Xu et al. 1992). This protein (called Byr3) contained 7 CCHC zinc fingers and acted as a suppressor of sporulation defects in *ras1*-deficient strains of *S. pombe*. Human CNBP was shown to partially fulfill the role of Byr3 in Byr3-deficient strains of *S. pombe* implying that the proteins are functionally similar.

Interestingly, *L. major* HEXBP and human CNBP were cloned using the same library screening technique, using what were assumed to be double-stranded oligodeoxynucleotide probes. However, both proteins were subsequently shown to bind only single-stranded nucleic acids. Although this finding may be artifactual due to the fortuitous use of probes which were partially single-stranded during library screening, the extent of sequence similarity between HEXBP and CNBP suggests that given the appropriate conditions these and other CCHC-containing proteins may have the potential to destabilize or melt DNA duplexes. Interestingly, retroviral NCP was recently reported to unwind secondary structure in RNA (Khan and Giedroc, 1992) and promote the renaturation of complementary DNA in solution (Dib-Hajj et al. 1993). Whether these properties are also shared by HEXBP or CNBP remains to be determined.

The ability of the NCP CCHC motif to bind zinc is well established (Roberts et al. 1989; Green and Berg, 1990; South et al. 1990a; Fitzgerald and Coleman, 1991; Mely et al. 1991; Summers et al. 1992; Surovoy et al. 1992; Mely et al. 1993; Surovoy et al. 1993) however there are conflicting reports in the literature regarding the requirement of zinc for retroviral NCP activity. This likely reflects the bimodal (sequence specific vs. non-sequence specific) manner in which NCP binds nucleic acids. For example, within the
mature retroviral particle, NCP is associated with the RNA genome in a non-specific histone-like complex (Smith and Bailey, 1979; Karpel et al. 1987; Bowles et al. 1993) thus for many years it was assumed that retroviral NCP was a non-sequence specific nucleic acid binding protein. Non-sequence specific binding of nucleic acid by NCP has been reported to occur in a zinc independent fashion (Karpel et al. 1987; Jentoft et al. 1988). In addition, it has been reported that mutation or chemical modification of the cysteine residues of Moloney murine leukemia virus (MoMuLV) NCP had no effect on \textit{in vitro} RNA dimerization or primer tRNA annealing (Cornille et al. 1990; Prats et al. 1991) however the requirement of sequence specific nucleic acid recognition for completion of these events is not clear. Conversely, mutagenesis of basic residues that flanked the CCHC zinc-binding motif severely inhibited these processes (Prats et al. 1991; De Rocquigny et al. 1992; Housset et al. 1993), which suggests that they are mediated by non-sequence specific binding mechanisms likely involving electrostatic interactions.

Alternatively, NCP or perhaps a proteolytic precursor to NCP, has also been implicated in the sequence specific process of packaging the retroviral RNA genome into the capsid (Gorelick et al. 1988; Aldovini and Young, 1990; Gorelick et al. 1990; Bowles et al. 1993; Rice et al. 1993) and several independent mutagenesis studies have found that abolishing the zinc-binding capability of NCP by site-directed mutagenesis of any one of the cysteine residues results in defective packaging of the genomic RNA into capsids (Gorelick et al. 1988; Jentoft et al. 1988; Meric and Goff, 1989; Gorelick et al. 1990). This process would likely require sequence specificity in order to discriminate the retroviral RNA molecules from the pool of host cell RNAs and in fact \textit{in vitro} mutagenesis of the CCHC motif of Moloney murine leukemia virus NCP results in the formation of defective viral particles containing inappropriate host cell RNAs (Gorelick et al. 1988; Meric and
Goff, 1989). In addition, it has recently been reported that the NCP of HIV-1 exhibits sequence specific binding to single-stranded DNA in vitro and that this sequence specificity is dependent upon the ability to bind zinc (Summers et. al.1992; South and Summers, 1993; Surovoy et. al.1993). However, the relevance of sequence specific single-stranded DNA binding to the retroviral life cycle is unclear at this time.

The requirement of zinc for HEXBP DNA-binding activity was not directly addressed in the current study and although 0.1 mM ZnCl₂ was present during all DNA-binding assays presented, other binding studies performed in the absence of exogenously added zinc yielded similar results (data not shown). Considering the very high affinity of the CCHC motif for zinc (dissociation constants of up to 10⁻¹²) (Mely et. al.1991; Summers, 1991; Mely et. al.1993) it is likely that the zinc binding sites of HEXBP were already occupied prior to the preparation of the crude extracts used in the present study. Therefore the addition of exogenous zinc would have little effect on DNA binding activity if it is indeed a requirement for HEXBP activity.

The structure of the CCHC motif of retroviral NCP has been extensively characterized using NMR-based methods (South et. al.1990b; Omichinski et. al.1991; South et. al.1991; Summers, 1991; Morellet et. al.1992; Summers et. al.1992; Mely et. al.1993; South and Summers, 1993; Surovoy et. al.1993) and all results indicate that the CCHC motif represents a structurally unique class of zinc finger domain. The crystal structures of the classical zinc finger domains from TFIIIA, glucocorticoid receptor and GAL4 proteins complexed with their respective binding sites have recently been determined and all share the common property of exposing an α helix within the major groove of double-stranded DNA (Luisi et. al.1991; Pavletich and Pabo, 1991; Marmorstein et. al.1992). Conversely, the CCHC motif has a much more compact structure with extensive
Chapter IV- Discussion

internal hydrogen bonding and no α helices (South et. al.1990b). Conservative substitution of several positions within the motif in addition to the invariant cysteine and histidine residues implies that this structure is likely shared by all CCHC consensus zinc finger motifs (see Figure 9). In particular, glycine residues are almost always found immediately following the second cysteine and immediately preceding the histidine. One of the two residues between the first and second cysteine is usually aromatic and a hydrophobic residue is generally found immediately following the histidine. Structural analyses of the HIV-1 NCP revealed that these residues formed a hydrophobic patch on the surface of the protein that was essential for nucleic acid binding (Summers et. al.1992). In addition to these conserved residues, all nine CCHC motifs of the HEXBP protein contained an arginine residue at the third position after the histidine and a proline immediately following the last cysteine. The significance of conservation at these positions is unknown at this time.

The spacer regions found between the CCHC motifs of HEXBP are generally rich in glycine, serine and alanine and are therefore predicted to be very flexible due to limited steric constraints. However, the spacer regions also contained a high proportion of charged residues (both positive and negative) which could potentially contribute to electrostatic interactions with a single-stranded nucleic acid ligand. This type of interaction would be reminiscent of the limited electrostatic interaction observed between the linker region of the GAL 4 zinc finger protein and the double-stranded GAL 4 binding-site (Marmorstein et. al.1992). The repetitive structure of L. major HEXBP together with the predicted flexibility of the spacer regions implied that the protein is likely capable of forming multiple contacts with an appropriate ligand via the multiple CCHC zinc fingers and therefore its respective binding site would be expected to be repetitive. However, there is a disparity between the
presence of nine CCHC motifs on the HEXBP protein and only two hexamer repeat units within the 5' region of the \emph{L. major} GP63 gene to which it presumably binds. Since only two of the nine CCHC motifs of HEXBP are identical (CCHC fingers 5 and 6, Figure 9) it is conceivable that the remaining motifs each have distinct sequence specificity. Such an arrangement might allow binding to a range of sequences with increasing stability depending upon the number of fingers which are simultaneously bound at a given site.

The addition of HEXBP to the family of proteins which contain the CCHC-type zinc finger domain provides considerable support for the proposal that proteins containing this motif function by binding to single-stranded nucleic acids (Summers, 1991). Single-stranded DNA-binding proteins generally bind in a non-sequence dependent manner (Chase and Williams, 1986) and therefore the sequence specific binding exhibited by \emph{L. major} HEXBP was initially unexpected. However, the sequence specific single-stranded DNA binding activity of HEXBP was confirmed by a number of different procedures including Southwestern blotting (Figure 12), competitive gel shift assay (Figure 14), UV cross-linking (Figure 16) and DNAse 1 protection (Figures 18, 20 and 21). The sequence specific binding of single-stranded DNA by HEXBP was most clearly evident in the protection of a specific region of the artificial single-stranded template BSHEX-327 (Figure 18). The protected region of this probe had clearly defined boundaries at both ends and protection was efficiently competed out by the addition of a synthetic oligonucleotide that corresponded to the protected site but not by identical amounts of an unrelated oligodeoxynucleotide. Interestingly, both protected sites were flanked at their 3' boundaries by DNAse 1 hypersensitive sites. DNAse 1 hypersensitivity is usually an indication of altered confirmation in double-stranded DNA, likely DNA bending (Nielson, 1990; Harrington, 1992), however the significance of its occurrence in single-stranded DNA is
unknown. Furthermore, HEXBP was shown to specifically protect at least four discrete regions of a probe corresponding to the antisense strand of the GP63 gene 5' untranslated region from digestion with DNAse I. The presence of multiple protected sites could be the direct result of a single molecule of HEXBP contacting multiple positions along the probe or alternatively could represent one molecule of HEXBP bound at each protected site. Considering the size of the HEXBP molecule (28 kDa) and the average size of the protected sites on the gp63-5'-462 probe (26 bases) it seems unlikely that a single HEXBP molecule could bind all four of the gp63-5'-462 protected sites unless there was extensive looping out of the intervening sequences located between each binding site. Furthermore, mobility shift experiments showed a progressive decrease in the mobility of the gp63-5'-462 probe as protein concentration was increased (Figure 19), implying the presence of multiple HEXBP molecules in the complex. Nonetheless, the presence of nine CCHC zinc finger motifs in the HEXBP molecule suggested that the protein is capable of forming multiple contacts with an appropriate ligand. It is therefore likely that each region of the gp63-5'-462 probe protected from DNAse I digestion represents one molecule of HEXBP bound at that position via the interaction of at least one, or more likely, several CCHC zinc finger domains with their respective DNA recognition sites.

Sites A and B of the gp63-5'-462 probe which were protected from DNAse I digestion by HEXBP (Figure 21) shared a repetitive, guanine-rich consensus sequence which was identical at 12 out of 14 positions (G G C G A/G G G G C/A G A G G G). Although the sequences of the two additional HEXBP binding sites (sites C and D) were not precisely determined, the entire sequence of the gp63-5'-462 probe was scanned and no further 14 base consensus sequences were identified. However, the region of the probe encompassing protected sites C and D did contain several purine-rich regions which might
Chapter IV- Discussion

exhibit enough similarity to the 14 base consensus to constitute a HEXBP binding site. High purine content may therefore be necessary but is not sufficient for HEXBP binding since several other purine-rich regions of the gp63-5'-462 probe were not protected from DNAse I digestion. Interestingly, the human DNA-binding protein CNBP was reported to bind similar oligodeoxynucleotides containing the guanine rich strand of the sterol response element (Rajavashisth et. al. 1989). Although the binding sites of other CCHC-containing proteins are not well defined it may be that CCHC binding sites are in general guanine-rich.

A limited degree of sequence variability at HEXBP binding sites might be expected based on the amino acid sequence of the HEXBP protein. As mentioned above, all nine CCHC zinc fingers of HEXBP are invariant with regards to the number and position of cysteine and histidine residues, however, only two of the fingers are completely identical in sequence. All other fingers are similar but contain at least one unique amino acid substitution within the motif. The minimum amino acid sequence information required for nucleic acid binding by retroviral NCP is present within the boundaries of the CCHC motif as shown by the use of synthetic peptides during in vitro studies (Delahunty et. al. 1992; South and Summers, 1993). If the specificity of binding is also defined by the amino acids located within the boundaries of the CCHC motif then it is possible that each of the eight unique fingers of HEXBP correlates with a unique binding site. The overall effect would be for HEXBP to have a broad range of binding site specificity dependent upon which fingers and how many fingers contact the DNA at any given site. This type of broad range specificity may be evident in the specific binding of HEXBP to a region of vector sequence on the BSHEX-327 and BS-374 probes. This site likely represents a cryptic binding site for one or more of the fingers of HEXBP. As described above it is not conclusively known whether each DNAse I protected site is a manifestation of binding by a single zinc finger of
HEXBP or whether multiple fingers make contact within one discrete protected region. Definitive answers to these questions will likely require the use of synthetic peptides which correspond to individual CCHC fingers of the HEXBP protein. Attempts to define specific contacts within the HEXBP/DNA complex through the use of DMS (dimethyl sulfate) or DEPC (diethyl pyrocarbonate) interference footprinting (Sturm et. al.1987) were uninformative as these modifications did not interfere with HEXBP binding (data not shown). This result was not unexpected since DMS and DEPC modifications generally disrupt protein/DNA complex formation by protruding into the major groove of duplex DNA. In addition, it cannot be entirely ruled out that HEXBP binds to nucleic acids having some form of secondary structure. However, computer analyses of the oligodeoxynucleotide probes used in the present study did not reveal any potentially significant areas of secondary structure and denaturing of probes by boiling prior to adding them to the binding reactions had little effect on binding activity (Webb and McMaster, 1993).

On the basis of the results obtained, a model of HEXBP binding can be invoked in which each site of protection along the gp63-5'-462 probe represents the presence of one molecule of HEXBP. Each HEXBP protein would be bound at this site through the interaction of at least one, or more likely, several zinc fingers with their respective DNA recognition sites. Interestingly, the CCHH class of zinc finger motif has been shown, in co-crystallization studies, to bind three nucleotides per finger with multiple fingers determining the overall sequence of the recognition site (Pavletich and Pabo, 1991). Although the CCHC motif is an entirely unique class of zinc finger motif, the correlation between the presence of nine zinc fingers and a DNAse I protected site of between 25 and 30 nucleotides may be relevant. Interestingly, the synthetic 50mer gp63-5'-50(-) was
completely protected from DNAse I digestion either on its own or when cloned into a vector background. Based on the sizes of individual protected sites within the gp63-5'-462 probe (26 bases) it is likely that gp63-5'-50(−) contains at least two HEXBP binding sites. Furthermore, the results presented confirm that the hexanucleotide direct repeat region of *L. major* GP63 gene 5' untranslated region functions, *in vitro*, as site of protein DNA interaction. These hexanucleotide direct repeats, located midway between the putative 3' *trans*-spliced leader acceptor site and the translational initiation codon, are conserved in both terminal and internal genes of the GP63 multigene tandem array in at least three different species of *Leishmania* and the repeats are also found 5' of the dispersed GP63 gene copies in *L. major* and *L. donovani* (B. Voth and J. Webb, unpublished). To address the functional significance of these repeats and to determine the consequences of *in vivo* binding by HEXBP a HEXBP-deficient strain of *L. major* was generated by double targeted gene replacement. The preliminary characterization of this mutant is described in the following chapter.
V. ANALYSIS OF Leishmania major HEXBP-DEFICIENT MUTANTS GENERATED BY DOUBLE TARGETED GENE REPLACEMENT

Recently, procedures have been developed for the transient and stable transfection of Leishmania and other kinetoplastids by electroporation (Bellofatto and Cross, 1989; Laban and Wirth, 1989; Kapler et al. 1990; Laban et al. 1990; LeBowitz et al. 1990). In addition, it has been demonstrated that gene replacement by homologous recombination can be achieved in kinetoplastids with an efficiency approaching 100% (Cruz and Beverley, 1990; Lee and Van der Ploeg, 1990; ten Asbroek et al. 1990; Cruz et al. 1991; Eid and Sollner-Webb, 1991; Tobin et al. 1991). To elucidate the cellular function of HEXBP, a HEXBP-deficient strain of L. major was generated by double targeted gene replacement.

A. RESULTS

1. Replacement of the First HEXBP Allele

A 4 kbp SalI restriction fragment of the L. major HEXBP locus that encompassed the complete HEXBP protein-coding region, 1.7 kbp of 5' flanking and 1.3 kbp of 3' flanking region (Fig. 22A) was chosen as a suitable target fragment for homologous gene replacement. The equivalent fragment was excised from the plasmid pMHB7sx and subcloned into the SalI site of pUC19 to generate the construct pMHB4s (Fig. 22B). The construct pMHBHyg (Fig. 22C) was derived from pMHB4s by replacing the HEXBP protein-coding region with the bacterial hygromycin phosphotransferase (HygR) gene, conferring resistance to the aminoglycoside antibiotic Hygromycin B. The 4 kbp insert of
pMHBHyg was excised by digestion with XbaI and HindIII and 5 ug of linear fragment was used to transfect *L. major* (CC1) log phase promastigotes. Seven drug-resistant (HygR) clones were obtained after selection of transfectants on plates containing Hygromycin B. All 7 drug-resistant clones were expanded in liquid culture and characterized by Southern blot hybridization analysis. As indicated on the restriction maps shown in Fig. 23B, alleles resulting from the predicted homologous replacement event could be discriminated from wild type alleles by the presence of a unique PstI restriction site located within the HygR gene. This difference could be detected by digestion of genomic DNA with PstI and XbaI and hybridization with *Probe A*, a flanking PstI/SalI restriction fragment that is external to the targeted fragment. *Probe A* was predicted to detect a 6 kbp PstI/XbaI fragment in wild type alleles and a 3 kbp PstI/PstI fragment in HygR alleles. Southern blot hybridization analysis using *Probe A* demonstrated that all 7 of the HygR clones were heterozygotes containing one wild type HEXBP allele and one HygR allele resulting from the expected homologous replacement event (Fig. 23A). Furthermore, when the Southern blots shown in Fig. 23A were stripped and rehybridized with a probe specific for the HygR gene (EcoRV/PstI fragment of pMHBHyg, see Fig. 22C), a band corresponding to the 3 kbp PstI/PstI fragment was detected in all HygR clones but not in the untransfected parental cell line (data not shown). In addition, no other bands were detected with the HygR-specific probe, indicating that random integration events had not occurred in any of the drug-resistant clones. One of these HygR clones, CC1-14-3, was chosen for a second round of targeted gene replacement.
2. Replacement of the Second HEXBP Allele

The construct pMHBNeo (Fig. 22D) was derived from pMHBHyg by replacing the HygR gene with the bacterial neomycin phosphotransferase gene (NeoR), conferring resistance to the aminoglycoside antibiotic G418. The 4 kbp insert of pMHBNeo was excised as described above for pMHBHyg and 5 ug of insert was used to transfect log phase promastigotes of the HygR clone CC1-14-3. Eight drug-resistant (HygR/NeoR) clones were obtained after selection of transfectants on plates containing Hygromycin B and G418. All 8 clones were expanded in liquid culture and characterized by Southern blot hybridization analysis. Alleles resulting from the planned homologous gene replacement event could be discriminated from wild type alleles by the presence of unique PstI sites within both the HygR and NeoR genes. Digestion of genomic DNA with PstI and XbaI and hybridization with Probe A was predicted to detect a 2.9 kbp PstI/PstI fragment in NeoR alleles and a 3 kbp PstI/PstI fragment in HygR alleles (Fig. 24B). Southern blot analysis indicated that all 8 NeoR/HygR clones contained one band corresponding to a NeoR allele and one band corresponding to a HygR allele (Fig. 24A, lanes 3 to 10). The 6 kbp PstI/XbaI fragment indicative of the wild type HEXBP allele was detected in the untransfected parental cell line CC1 and in the HEXBP heterozygous deletion clone CC1-14-3 (Fig. 24A, lanes 1 and 2 respectively) but was not present in any of the NeoR/HygR clones, implying that double targeted gene replacement had successfully occurred in all 8 clones. When the Southern blots shown in Fig. 24A were stripped and rehybridized with a probe specific for the NeoR gene (EcoRV/PstI fragment of pMHBNeo, see Fig. 22D), a band corresponding to the 2.9 kbp PstI/PstI fragment was detected in all HygR/NeoR clones but not in the untransfected parental cell line or the HEXBP heterozygous deletion.
clone CC1-14-3 (data not shown). As observed with the Hyg-specific probe after the first round of gene replacement, no further hybridization to the Neo-specific probe was detected, implying that random integration events had not occurred in any of the drug-resistant clones. One of these clones, CC1-14-4/D was chosen for further characterization.

3. Characterization of the HEXBP Deletion Mutant CC1-14-4/D

To confirm that the NeoR/HygR clones generated by homologous gene replacement were truly deficient for the HEXBP gene, the Southern blot shown in Fig. 24A was stripped and rehybridized with a random-primed restriction fragment of the \textit{L. major} HEXBP protein-coding region (\textit{Probe B}). A band corresponding to the 6 kbp PstI/XbaI fragment from a wild-type HEXBP allele was detected in the parental clone CC1 (Fig. 25, lane 1) and at approximately half intensity in the HEXBP heterozygous deletion clone CC1-14-3 (Fig. 25, lane 2) but not in any of the NeoR/HygR clones (Fig. 25, lanes 3 to 10). Furthermore, none of eight NeoR/HygR clones contained any additional hybridizing bands which might be indicative of an additional HEXBP gene(s) retained via gene amplification or modification of chromosome content. It was recently reported that double targeted gene replacement in \textit{Leishmania} occasionally results in the retention of the targeted gene through changes in chromosome content (formation of aneuploid and tetraploid cell lines) (Cruz \textit{et al.} 1993). However, two smaller (1.7 and 1.9 kbp) fragments that cross-hybridized with the HEXBP-specific probe were detectable in the genomic DNA of all clones, including the parental lines (Fig. 25, all lanes). These cross-hybridizing fragments had a greatly reduced signal intensity compared to the 6 kbp HEXBP band (signal is less intense than the single
HEXBP gene copy in the HEXBP heterozygous deletion clone CC1-14-3) and hybridized only with specific fragments from the HEXBP gene locus (data not shown) implying that they likely represent a HEXBP-related gene(s) rather than additional HEXBP genes. Furthermore, when Northern blots of promastigote total RNA were hybridized with a HEXBP-specific probe (Probe B), HEXBP mRNA was detected in the parental clone CC1 and in the HygR clone CC1-14-3 (Fig. 26, lanes 1 and 2) but not in the NeoR/HygR HEXBP homozygous deletion clone CC1-14-4/D (Fig. 26, lane 3). Interestingly, the heterozygous deletion clone CC1-14-3 (Fig. 26, lane 2) contained approximately half as much HEXBP mRNA as their wild type counterparts (Fig. 26, lane 1) indicating that expression at the remaining allele was not increased to compensate for reduced copy number. As had been previously observed (see Fig. 10) hybridization of promastigote total RNA with a HEXBP-specific probe also detected a higher molecular weight transcript of significantly lower hybridization intensity (Fig. 26, lanes 1, 2 and 3). The relative amount of the cross-hybridizing, higher molecular weight transcript did not vary between parental and HEXBP-deficient clones implying that this transcript was not derived from the HEXBP locus. Although likely, a direct correlation between this cross-hybridizing transcript and the cross-hybridizing DNA fragments observed in Fig. 25 has not been established.

To determine whether deletion of the HEXBP gene had any effect on GP63 expression, Northern blots of total RNA from promastigotes grown to varying culture densities were hybridized with a random-primed restriction fragment of the \textit{L. major} GP63 protein coding region (Probe C). There were no dramatic differences between parental (Fig. 27, lanes 1 to 4) and HEXBP-deficient (Fig. 27, lanes 5 to 8) promastigotes with regards to the total amount of GP63 mRNA detected. In addition, identical amounts of
GP63 protein were detected in total cell extracts of CC1, CC1-14-3 and CC1-14-4/D promastigotes by Western blot analysis using anti-GP63 specific monoclonal antibody (Fig. 28).

Finally, there were no obvious differences in the general morphology of HEXBP-deficient promastigotes in comparison to promastigotes of the parental clone at the level of light microscopy. In addition, parental and HEXBP-deficient clones were indistinguishable with regards to the growth kinetics and saturating densities of in vitro promastigote cultures. Interestingly, parental and HEXBP-deficient clones differed in the rate of acidification of promastigote culture media. Despite the finding that the growth curves of promastigote cultures were identical, the acidification of culture media by all 8 HEXBP deficient strains consistently lagged behind that of their wild type counterparts by approximately 2 to 3 days (data not shown). Whether this is a direct or indirect result of HEXBP deletion remains to be determined.

To determine whether HEXBP-deficiency affected the virulence of promastigotes, Balb/c mice were injected intradermally with 1 x 10^8 stationary phase promastigotes from parental (CC1) and HEXBP-deficient (CC1-14-4/D) clones. Contrary to a recent report describing L. major CC1 as an avirulent strain (Cruz et. al.1993), mice inoculated with CC1 promastigotes in the present study developed large lesions, albeit at a much slower rate than normal. Balb/c mice are very susceptible to infection with L. major and normally develop visible lesions in approximately 6 to 8 weeks. In the present study, lesions resulting from infection with L. major CC1 were not apparent until the 23rd week post-inoculation. Two mice that were injected with an equivalent number of HEXBP-deficient promastigotes showed no sign of infection until 43 weeks post-inoculation, at which time one mouse developed a visible lesion.
Figure 22. Restriction maps of the HEXBP gene locus and plasmid constructs used in homologous gene replacement experiments.

(A) Restriction map of the HEXBP gene locus of *L. major*. The HEXBP protein coding region (translated from left to right) is shown boxed. The 4 kbp region encompassed by the first SalI sites located upstream and downstream of the HEXBP protein coding region was chosen as the target fragment for homologous recombination. (B) Restriction map of the plasmid pMHB4s. The 4 kbp SalI *Leishmania* insert containing 1.7 kbp of 5' flanking sequence, the HEXBP protein coding region and 1.3 kbp of 3' flanking sequence is shown expanded. The thin circular line depicts the pUC19 plasmid vector. (C) Restriction map of the construct pMHBHyg. pMHBHyg was derived from pMHB4s by replacing the HEXBP protein coding region (in the same orientation) with the coding region of a gene conferring resistance to hygromycin B (HygR). The insert fragment used for transfection was excised using the flanking XbaI/HindIII restriction sites. (D) Restriction map of the construct pMHBNeo. pMHBNeo was derived from pMHBHyg by replacing the HygR protein-coding region (in the same orientation) with the coding region of the gene conferring resistance to neomycin (NeoR). The insert fragment used for transfection was excised at the XbaI/HindIII sites as in C.
Chapter V. Results
Figure 23. Southern blot hybridization analysis of HygR transfectants.

(A) Genomic DNA from promastigotes of the wild type parental strain CC1 (lane 1) and HygR clones 14-1 to 14-7 (lanes 2 to 8) was digested with PstI and XbaI and analyzed by Southern blot hybridization using the restriction fragment labeled Probe A as a hybridization probe. The positions of molecular weight markers are indicated to the left in kilobasepairs (kb). (B) Restriction map of the predicted *L. major* HEXBP locus after homologous gene replacement at one allele. The wild type allele is labeled HEXBP and the allele resulting from gene replacement is labeled HygR. The location of the PstI/SalI restriction fragment used as a flanking region hybridization probe in the Southern blot shown in A is indicated by the solid bar labeled *Probe A*. The restriction fragments of wild type HEXBP (6 kb) and predicted HygR alleles (3 kb) detected using *Probe A* are indicated.
Figure 24. Southern blot analysis of HygR NeoR transfectants.
(A) Genomic DNA from promastigotes of the wild type parental strain CC1 (lane 1), the HygR clone 14-3 (lane 2) and the HygR/NeoR clones 14-1/D to 14-8/D (lanes 3 to 10) was digested with PstI and XbaI and analyzed by Southern blot hybridization using the restriction fragment labeled Probe A as a hybridization probe. The positions of molecular weight markers are indicated to the left in kilobasepairs (kb). (B) Restriction map of the predicted L. major HEXBP locus after homologous recombination at both alleles. The allele resulting from the first round of gene replacement is labeled HygR and the allele resulting from the second round of gene replacement is labeled NeoR. The location of the PstI/SalI restriction fragment used as a flanking region hybridization probe in the Southern blot shown in A is indicated by the solid bar labeled Probe A. The restriction fragments of the predicted HygR (3 kb) and NeoR alleles (2.9 kb) detected using Probe A are indicated.
Figure 25. Detection of a HEXBP-related sequence by genomic Southern blot analysis.

(A) The Southern blot shown in Fig. 24A was stripped and reprobed with a HEXBP-specific probe (Probe B). The positions of molecular weight markers are indicated to the left in kilobasepairs (kb). (B) Restriction map of the wild type *L. major* HEXBP locus. The location of the EcoRV/NotI restriction fragment used as a HEXBP-specific hybridization probe in the Southern blot shown in A is indicated by the solid bar labeled *Probe B*. The restriction fragment from the wild type HEXBP allele (6 kb) detected using *Probe B* is indicated.
Figure 26. Northern blot hybridization analysis of HEXBP gene expression in HygR/NeoR clones.
Total RNA isolated from log phase promastigotes of the wild type parental strain CC1 (lane 1), the HygR clone 14-3 (lane 2) and the HygR/NeoR clone 14-4/D (lane 3) was analyzed by Northern blot hybridization using a HEXBP-specific hybridization probe (see Fig. 25. Probe B). The presence of HEXBP mRNA is indicated by the arrowhead labeled 3.2 kb.
Figure 27. Northern blot hybridization analysis of GP63 expression in HEXBP-deficient clones.
Total RNA was isolated from promastigotes of the wild type parental strain CC1 (lanes 1 to 4) or the HygR/NeoR clone 14-4/D (lanes 5 to 8) and analyzed by Northern blot hybridization using a GP63-specific hybridization probe. RNA was isolated from log phase promastigotes grown to $3 \times 10^6$ cells/ml (lanes 1 and 5), $6 \times 10^6$ cells/ml (lanes 2 and 6) and $1.2 \times 10^7$ cells/ml (lanes 3 and 7) or stationary phase promastigotes ($1.8 \times 10^7$ cells/ml, lanes 4 and 8). The presence of GP63 mRNA is indicated by the arrowhead labeled 3.0 kb.
Figure 28. Western blot analysis of GP63 expression in HEXBP-deficient clones.

Total cell lysates from log phase promastigotes of the wild type parental strain CC1 (lane 1), the Hyg\textsuperscript{R} clone 14-3 (lane 2) and the Hyg\textsuperscript{R}/Neo\textsuperscript{R} clone 14-4/D (lane 3) were characterized by Western blot analysis using the anti-GP63 monoclonal antibody CP3.235. The presence of GP63 protein is indicated by the arrowhead on the left.
B. DISCUSSION

The development of a HEXBP-deficient strain of *Leishmania* provides an important avenue for determining the role of the HEXBP single-stranded DNA binding protein. The results presented in this chapter suggest that although HEXBP is capable of binding, *in vitro*, to oligodeoxynucleotides derived from the 5' untranslated region of the GP63 gene, it is not essential for efficient expression of GP63 in promastigotes. In addition, the viability of HEXBP-deficient promastigotes implied that HEXBP is not required for promastigote survival *in vitro*.

Although the function of HEXBP and the consequences of it binding to available single-stranded targets remains undetermined at this time, other proteins containing the CCHC zinc finger motif have been implicated in a diverse collection of cellular processes. The finding that HEXBP does not play a direct role in GP63 expression at the transcriptional level was not entirely unexpected considering the relatively simple overall structure of the HEXBP protein. The nine zinc finger motifs and the accompanying short spacer regions constitute the majority of the HEXBP sequence, which would seem to preclude its involvement in any activity other than DNA-binding. In contrast, eukaryotic transcription factors are generally modular in nature, containing both a DNA-binding domain and an 'activation' domain which likely plays a role in interacting with the preinitiation complex (Frankel and Kim, 1991). Based on the structure of HEXBP, it seems more likely that its role is to simply bind at an appropriate single-stranded site, perhaps to prevent or delay the return of that site to the duplex form, to allow for the completion of an as yet undetermined secondary activity. Conversely, the binding of HEXBP to specific single-stranded regions might play a more direct role such as occluding...
the binding site of a double-stranded DNA-binding protein or providing a binding site for another single-stranded DNA-binding protein. It is also possible that HEXBP bound to single-stranded DNA functions by recruiting factors that bind through protein/protein interactions. Binding of HEXBP within a single-stranded region might also provide protection from nuclease digestion. Interestingly, a protective role has also been attributed to the retroviral NCP (Bowles et al. 1993) and to the gene 32 protein of λ phage, which contains a zinc-binding motif similar to the CCHC motif but which is arranged in the opposite orientation (CX₃HX₅CX₂C) (Chase and Williams, 1986).

Although the DNA-binding results presented in Chapter IV were generated using single-stranded synthetic oligodeoxynucleotides or nucleic acid probes that were rendered single-stranded through enzymatic means, there are intriguing implications that can be extended to the in vivo situation. Duplex DNA becomes necessarily single-stranded, in vivo, during several processes, including passage of transcription forks and replication forks. In addition, a single-stranded binding site for the HEXBP protein might become available during normal 'breathing' of duplex DNA or during recombination or transposition events. As mentioned in Chapter IV, it is also possible that HEXBP itself mediates formation of a single-stranded DNA-binding site via dissociation of double-stranded DNA. In fact, helix destabilizing activity is a characteristic property of the large family of non-sequence specific single-stranded DNA binding proteins which includes the gene 32 protein of λ phage and the E. coli single-stranded binding protein (SSB) (Chase and Williams, 1986). Although the exact biological functions of these proteins are several fold and not completely understood, they are all capable of stimulating the in vitro activity of DNA polymerases by melting out secondary structure on single-stranded templates. Whether this activity can be extended to in vivo DNA replication is not clear. However,
unlike HEXBP-deficient mutants, mutagenesis of the genes encoding non-sequence specific single-stranded DNA binding proteins generally results in a lethal phenotype.

Single-stranded binding proteins and helix destabilizing proteins have also been reported to bind at sites characterized as origins of replication in both prokaryotes and eukaryotes (Bramhill and Kornberg, 1988; Hofmann and Gasser, 1991; Seroussi and Lavi, 1993). However, in contrast to the non-sequence specific single-stranded binding proteins described above, proteins which interact with putative replication origins exhibit sequence specific single-stranded binding. The emerging model of replication initiation suggests that the replication origin is initially bound by a factor known as the initiator (large T antigen in the case of SV40 (Borowiec et al. 1990) and ORC in the case of yeast ARS sequences (Bell and Stillman, 1992)) which initiates strand separation. Single-stranded binding proteins then invade the site and participate in further unwinding of the origin region. These single-stranded DNA-binding proteins include replication protein-A (RP-A, previously known as human single-stranded binding protein) which binds to the pyrimidine-rich strand of the SV40 origin (Seroussi and Lavi, 1993), IR factor B (IRF-B) which binds to the complementary purine-rich strand (Carmichael et al. 1993) and the ARS consensus site binding protein (ACBP) which binds to the pyrimidine rich strand of the yeast ARS sequence (Hofmann and Gasser, 1991). The RP-A factor from human cells is a heterotrimeric complex composed of 70, 34 and 14 kDa subunits (Wold and Kelly, 1988). Interestingly, a homologue of human RP-A was recently isolated from the kinetoplastid protozoan *Crithidia fasciculata* and was reported to be composed of 51, 28 and 14 kDa subunits (Brown et al. 1992). Despite the fact that the middle subunit of *C. fasciculata* RP-A and *L. major* HEXBP have identical molecular weights it is unlikely that they represent the same protein since only the largest subunit of RP-A exhibits DNA-binding activity on
Southwestern blots. Furthermore the middle subunit of human RP-A has been cloned and sequenced (Erdile et al. 1990) and it exhibits no sequence identity with *L. major* HEXBP.

Recently, another sequence specific single-stranded DNA binding protein that interacts with a putative origin of replication was identified in HeLa cells (Bergemann and Johnson, 1992). This protein, called Pur factor, had an estimated molecular weight of 28 kDa and bound to the G-rich strand of a consensus sequence referred to as the PUR element, located upstream of the *c-myc* gene. Interestingly, the PUR element is identical to the hexamer repeat region of the HEXBP binding site at 13 out of 16 positions (data not shown). The same report described the identification of the PUR element at a number of putative origins of replication, particularly from lower eukaryotic organisms. However, it seems unlikely that Pur factor and HEXBP represent equivalent activities since cultures of HEXBP-deficient promastigotes divided at a normal rate, implying that DNA replication was not affected.

One further protein which might be relevant to HEXBP is another sequence specific single-stranded DNA binding protein isolated from *C. fasciculata* (Tzfati et al. 1992). This protein had an estimated molecular weight of 27 kDa and was reported to bind to the purine-rich strand of a sequence known as the universal minicircle sequence (UMS) which is thought to represent the origin of replication of the kDNA minicircle. However, kinetoplasts with a normal morphology were detected in HEXBP-deficient *L. major* promastigotes by staining with fluorescent dyes (data not shown), suggesting that HEXBP is not involved in the replication of kinetoplast DNA.

The ability of HEXBP to bind single-stranded DNA was well characterized in the present study however its potential as an RNA-binding protein cannot be completely ruled out. The inability of excess promastigote total RNA to act as a binding competitor in gel
mobility shift assays (Figure 14) implied that HEXBP does not bind RNA randomly, however, that interpretation cannot be extended to include sequence specific RNA-binding. *In vitro* synthesized RNA molecules containing the equivalent of the HEXBP binding site were not bound by HEXBP in gel mobility shift assay (data not shown) however, these experiments were not conclusive due to varying amounts of RNAse activity in the bacterial lysates used a source of recombinant HEXBP. These experiments would be more definitive if performed using a purified form of HEXBP that was free of RNAse activity.

Although the results presented in this work suggest that there is no strict requirement for HEXBP during GP63 gene expression, a nonessential role or a redundant role in GP63 expression cannot be ruled out. In the case of a nonessential role, there may be some subtle shift in GP63 expression, such as a decrease in the rate of transcriptional initiation, which might not be accurately reflected by Northern hybridization analysis of total accumulated levels of GP63 mRNA. In the case of a redundant role, some other protein(s) may be able to partially or totally fulfill the role of HEXBP in its absence. It may be relevant that sequences which cross-hybridized with HEXBP-specific probes were detected in *L. major* in the present study (see Fig. 25). These sequences were observed at both the genomic DNA level and at the RNA level, implying that there is likely a HEXBP-related protein in wild-type and HEXBP deficient clones that could potentially overlap functionally with the HEXBP protein. However, it seems more likely that HEXBP functions at a level other than GP63 transcription. One intriguing possible function for HEXBP relates to maintenance of the GP63 gene locus. As previously reported, all species of *Leishmania* contain multiple GP63 genes arranged as directly repeated tandem arrays (Button *et. al.*1989; Miller *et. al.*1990; Webb *et. al.*1991; Medina-Acosta *et. al.*1993b). HEXBP binds to oligodeoxynucleotides derived from the 5′non-coding region of the first
gene of the *L. major* GP63 gene array, however, the identical sequence is also present in the 5' untranslated region of internal genes from the *L. donovani* and *L. chagasi* GP63 gene arrays (Miller et al. 1990; Webb et al. 1991). It is therefore likely that the HEXBP binding site is present in front of all genes of the GP63 tandem array, which means that all GP63 genes are, in effect, flanked by HEXBP binding sites. Such an arrangement may be indicative of the involvement of HEXBP in some aspect of recombination. For example, HEXBP might be required for homogenization of the GP63 gene locus through a process such as gene conversion. The presence of direct repeats within the HEXBP binding site may in fact corroborate this interpretation since short direct repeats have been reported to represent recombination points in *Leishmania* (Liu et al. 1992). Alternatively, binding by HEXBP at potential recombination points might limit the possibility of inappropriate mitotic crossing over, which could otherwise result in a rapid expansion of the GP63 tandem array.

It was intriguing that HEXBP-deficient promastigotes acidified culture media at a reduced rate compared to normal promastigotes. This finding suggested that some aspect of metabolism was impaired in HEXBP-deficient strains without having a concomitant effect on the growth curve of the culture. However, since this effect could potentially be ruled out as a fortuitous characteristic of the clones used in this study, a definitive causal relationship with HEXBP will require complementation analyses using recently developed extrachromosomal expression vectors.

Similarly, the finding that the HEXBP-deficient clone generated in this study was avirulent suggests that HEXBP is essential for establishing an infection in Balb/c mice. However, this finding could again be attributed to the use of a single cloned line, therefore additional experimentation is required before HEXBP can be designated as an essential
virulence factor. Nonetheless, avirulence of the HEXBP-deficient clone is suggestive that HEXBP is involved in virulence and therefore its role in establishing infection can be discussed in a speculative manner. Particularly intriguing is the potential relevance to recent findings that expression of different classes of GP63 genes are developmentally regulated (Ramamoorthy et al. 1992; Medina-Acosta et al. 1993b). As described previously in Chapter III, the GP63 gene families of *L. chagasi* and *L. mexicana* were reported to be comprised of structurally distinct genes encoding proteins with divergent carboxyl-terminal domains. All GP63 gene classes are apparently expressed in the promastigote lifestage, albeit to different levels, whereas in *L. mexicana* only one class was expressed in the amastigote lifestage (Medina-Acosta et al. 1993b). Although the functional significance of GP63 molecules with divergent carboxyl-terminal domains is not clear, evidence suggests that expression of the various gene types is regulated at the transcriptional level. In this regard it may be relevant that the related human protein CNBP (which contains seven CCHC zinc finger domains) was reported to function as a transcriptional repressor, regulating the expression of several genes involved in sterol metabolism (Rajavashisth et al. 1989). It is therefore plausible that HEXBP could also be acting as a transcriptional repressor by inhibiting the expression of specific GP63 classes in the amastigote life stage. This hypothesis may reflect our inability to detect significant amounts of HEXBP in *L. major* promastigote extracts by Western blot analysis (data not shown). Furthermore, a GP63 gene that was structurally equivalent to the amastigote-specific GP63 gene of *L. mexicana* was recently cloned from *L. guyanensis* (Steinkraus and Langer, 1992). Although the expression of this gene at the RNA level was not reported, it was significantly divergent in the 5' untranslated region compared to previously characterized GP63 genes and contained no hexanucleotide direct repeat region. Based on the extensive sequence
divergence it is unlikely that HEXBP could bind to the 5' untranslated region of this gene. If HEXBP does indeed function as a transcriptional repressor, expression of this particular GP63 gene would likely continue in the presence of active HEXBP, which would correlate with stage-specific transcriptional regulation. Although speculative at this time, the potential role of HEXBP as transcriptional repressor merits further experimentation.

In conclusion, the development of a HEXBP-deficient strain of *L. major* provides an important tool to facilitate the analysis of HEXBP function. Considering the growing list of proteins that contain the CCHC motif, elucidating the function of *L. major* HEXBP could have widespread implications in the area of single-stranded nucleic acid binding proteins. Furthermore, once the function of HEXBP has been elucidated, complementation studies utilizing a HEXBP-deficient strain of *L. major* will provide a convenient method for determining the functional similarity of CCHC-containing proteins. The following chapter describes the development of an expression vector for *Leishmania* which could be applicable to such complementation analyses.
VI. DEVELOPMENT OF AN EXPRESSION VECTOR FOR THE STABLE TRANSFECTION OF *LEISHMANIA* AND APPLICATION TO THE STUDY OF POLYCISTRONIC GENE EXPRESSION

This chapter describes the development of a stable, plasmid-based vector for the transformation and expression of cloned gene products in *Leishmania*. The development of this vector is based on the finding that pMHBHyg, used in the homologous gene replacement experiments described in Chapter V, leads to a stable drug-resistant phenotype when transfected into *Leishmania* promastigotes by electroporation. Transfectants were characterized by Southern blot hybridization analyses to determine what conformation the construct adopted in promastigotes. A preliminary characterization of the transcriptional organization and control was obtained by Northern blot hybridization analysis.

A. RESULTS

1. Transfection of *Leishmania* with the Circular Plasmid Construct pMHBHyg.

The pUC19-based plasmid construct pMHBHyg, which was used previously as a source of linear DNA for HEXBP homologous gene replacement experiments (see Fig. 22C), contained the coding-region of the bacterial hygromycin phosphotransferase gene (HygR) flanked by 1.7 kbp of 5' and 1.3 kbp of 3' untranslated sequence from the *L. major* HEXBP gene locus. Transfection of *L. major* CC1 promastigotes with 25 ug of the intact circular form of this construct resulted in the generation of large numbers of drug-
resistant colonies after selection of transfectants on plates containing Hygromycin B. Mock transfected cells (electroporated in the absence of added DNA), or cells incubated with plasmid DNA but not electroporated, did not give rise to any colonies when plated on selective media. Eight, well separated, drug-resistant colonies (designated as CC1/10-1 to CC1/10-8) were picked from plates, expanded in liquid culture and characterized by Southern blot hybridization analysis.

Southern blots containing transfectant genomic DNA digested with restriction enzymes predicted to linearize the construct pMHBHyg were alternately hybridized with probes specific for the Hyg\textsuperscript{R} gene and the pUC19 region of pMHBHyg (Fig. 29A and B, respectively). A band corresponding in size to intact, linearized pMHBHyg was detected in all clones using both hybridization probes. The intensity of the bands varied somewhat between individual clones, likely indicating a difference in pMHBHyg copy number. No other hybridizing bands were detected. These results suggested that pMHBHyg was maintained in promastigotes as a stable, circular, extrachromosomal element. Although a similar hybridization pattern would be predicted if multiple copies of the plasmid had integrated at a chromosomal location to form a direct repeat tandem array, bands corresponding to terminal flanking fragments would also have been detected. No additional hybridizing bands were observed, even after long exposure of the autoradiographs shown in Fig. 29 (data not shown), confirming that pMHBHyg was maintained as an extrachromosomal element.

When undigested genomic DNA from the Hyg\textsuperscript{R} clones was analyzed by southern blot hybridization, pMHBHyg DNA was observed to co-migrate with high molecular weight genomic DNA rather than single copy plasmid DNA (Fig. 30). Together, these findings suggested that pMHBHyg was maintained as either a large circular concatamer
containing multiple copies of the plasmid arranged as direct head to tail tandem repeats or as a network of concatenated single plasmids. Both types of element would be predicted to give rise to the plasmid-sized bands observed in Fig. 29 when digested with a restriction enzyme having a unique site within the construct. To discriminate between these two possible forms, genomic DNA from transfectant CC1/10-2 was subjected to limited endonucleolytic digestion using DNAse I followed by Southern blot hybridization using a probe specific for the Hyg\textsuperscript{R} gene (Fig. 31). As previously observed, the intact form of pMHBHyg DNA present in undigested genomic DNA was of high molecular weight (co-migrated with high molecular weight genomic DNA) and was inefficiently transferred out of the gel by capillary Southern blot (Fig. 31, lane 1). When CC1/10-2 genomic DNA was incubated with dilute DNAse I (diluted 1:10,000 to 1 U/ml) the pMHBHyg-specific DNA shifted to a faster migrating form that was transferred out of the gel much more efficiently during capillary Southern blot (Fig. 31, lane 2). Increasing the duration of DNAse I digestion resulted in the formation of increasing amounts of this fragment (Fig. 31, lanes 3 to 5). However, DNA that was the size equivalent of single copy plasmid DNA was not detectable by this analysis. As a positive control for DNAse I activity, the same experiment was performed using genomic DNA from the \textit{L. major} wild type parental clone CC1 that had been spiked with 1 ng of pMHBHyg plasmid DNA isolated from \textit{E. coli}. As expected, two hybridizing bands were detected in the undigested, spiked CC1 DNA; a faster migrating band corresponding to supercoiled plasmid, and a slower migrating band corresponding to relaxed circular plasmid (Fig. 31, lane 6). After treatment with dilute DNAse I the supercoiled form of the plasmid disappeared and a band that was equivalent in size to linearized plasmid was formed (Fig. 31, lane 7). Linearization of the supercoiled plasmid was essentially complete after two minutes of digestion and increasing the duration
of DNase I digestion did not substantially alter the pattern of hybridizing bands (Fig. 31, lanes 8 to 10). Together these results were interpreted as evidence that pMHBHyg was maintained as a large circular concatamer rather than a concatenated network of single plasmids. Furthermore, altered electrophoretic mobility of the element after treatment with DNAse I is consistent with the cleavage of a large circular concatamer to a linearized form.

2. Modification of pMHBHyg to generate the *Leishmania* expression vector pLEX.

Successful transfection of *Leishmania* with pMHBHyg implied that this construct contained all of the information required for stable extrachromosomal replication and expression of selectable markers. pMHBHyg was subsequently modified to investigate its potential as a vector for the expression of cloned genes in *Leishmania*. As described in earlier chapters, gene expression in *Leishmania* and other kinetoplastid protozoans is generally assumed to occur in a polycistronic fashion (Imboden *et al.* 1987; Ben Amar *et al.* 1988; Muhich and Boothroyd, 1988; Cross, 1990) and arrangement of genes as direct repeat tandem arrays is a common feature (Thomashow *et al.* 1983; Tschudi *et al.* 1985; Button *et al.* 1989). Mature monocistronic mRNAs are presumed to be generated from multigene precursor transcripts through the events of polyadenylation and trans-splicing (Huang and Van der Ploeg, 1991a; Ullu *et al.* 1993). It was therefore hypothesized that a tandem array intergenic region inserted between the HygR-coding region and the HEXBP 3' untranslated region of pMHBHyg would provide all of the information required for expression of a downstream gene. An intergenic region from the *L. major* GP63 gene locus
was chosen as an appropriate intergenic region for an expression vector since the locus is well characterized (Button et al. 1989) and because GP63 is expressed at high levels in promastigotes (Bordier, 1987). A 1.3 kbp fragment spanning a complete L. major GP63 intergenic region was generated from the plasmid pLMS-7-1-3 (Button and McMaster, 1988) by PCR amplification and blunt-end cloned into the SalI site of pBluescript to generate the construct pBlue.int.s1. A fragment of pBlue.int.s1 encompassing the entire GP63 intergenic region along with a large portion of the pBluescript multiple cloning site at its downstream end was excised by digestion with XhoI and NotI. The construct pMHBHyg.m1 was generated from pMHBHyg by eliminating most of the multiple cloning region from the pUC19-derived vector backbone to allow efficient use of an internal multiple cloning site. The XhoI/NotI restriction fragment of pBlue.int.s1 was then blunt-end cloned upstream and downstream of the HygR gene of pMHBHyg.m1 to generate the constructs pLEXUL and pLEX respectively (see Fig. 32). Electroporation of L. major promastigotes with 25 ug of pLEXUL or pLEX resulted in the formation of many drug resistant colonies (>100) on M199 plates containing Hygromycin B. Mock transfected cells (electroporated in the absence of added DNA) or cells incubated with plasmid DNA but not electroporated did not give rise to any colonies when plated on selective media. These results suggested that 1) genes located directly downstream of a GP63 intergenic region/multiple cloning site would be efficiently expressed in transfectants and 2) the expression of the selectable marker (HygR) was not inhibited by the presence of the intergenic region/multiple cloning site immediately downstream of the HygR gene.

To determine whether the pLEX construct would co-express the HygR selectable marker and a downstream gene, the bacterial NeoR gene was cloned into the multiple cloning site of pLEX to generate the construct pLEXNeo (see Fig. 32). Electroporation of
Chapter VI- Results

*L. major* promastigotes with 25 ug of pLEXNeo resulted in the formation of many drug resistant colonies (>100) on plates containing either only Hygromycin B or Hygromycin B and G418. These results implied that genes cloned into the expression site of pLEX would be efficiently expressed.

Clones transfected with pMHBHyc, pLEXUL, pLEX and pLEXNeo were expanded in liquid culture in the continued presence of selective drug and were characterized by Southern and Northern blot hybridization. The presence of plasmid in all transfectants was confirmed by Southern blot hybridization of XbaI digested genomic DNA using a probe specific for the HygR gene (Fig. 33). Expression of the HygR gene in all transfectants was characterized by Northern blot hybridization using the same Hyg-specific probe. As expected, no band was observed in the *L. major* parental clone CC1 whereas a Hyg-containing transcript that was similar in size to the wild type HEXBP transcript (3.2 kbp) was observed in the HEXBP gene replacement mutant CC1/144 (Fig. 34A, lanes 1 and 2). Two distinct Hyg-specific transcripts of 4.6 and 1.9 kb were detected in total RNA of clones transfected with pMHBHyc (Fig. 34A, lane 3). Assuming that the putative *trans*-spliced leader acceptor site in the HEXBP 5' untranslated region was correctly utilized on these constructs, the 3' ends of the 4.6 and 1.9 kb transcripts would correspond to transcriptional termination and/or polyadenylation at cryptic sites in the pUC19 vector backbone. Single distinct Hyg-containing transcripts of approximately 1.9 kb were also detected in total RNA from clones transfected with pLEX and pLEXNeo (Fig. 34A, lanes 4 to 6). This transcript was exactly the size predicted assuming correct usage of the putative *trans*-spliced leader acceptor site in the HEXBP 5' untranslated region and polyadenylation within the GP63 intergenic region. Furthermore, these findings implied that the downstream *trans*-spliced leader acceptor/polyadenylation site acts as an efficient
processing point regardless of whether or not a gene is present in the downstream expression site. Interestingly, no transcripts were detectable in the RNA of clones transfected with the construct pLEXUL (Fig. 34A, lane 3) suggesting that although this clone was resistant to Hygromycin B the Hyg\textsuperscript{R} transcript was extremely labile. The Northern blot shown in Fig. 34A was subsequently stripped and rehybridized with a probe specific for the Neo\textsuperscript{R} gene. No bands were observed in the \textit{L. major} parental clone CC1 whereas a single Neo-containing transcript, similar in size to the wild type HEXBP transcript (3.2 kbp), was observed in the HEXBP gene replacement mutant CC1/144 (Fig. 34B, lanes 1 and 2). As expected, no Neo-containing transcripts were detected in RNA from clones transfected with pMHBHyg, pLEXUL or pLEX (Fig. 34B, lanes 3, 4 and 5 respectively). Neo-containing transcripts of 4.6 kb and 1.9 kb were detected in the RNA of the clone transfected with pLEXNeo and selected for growth in the presence of Hygromycin B and G418 (Fig. 34B, lane 7). More importantly, the same transcripts were detectable in the RNA of the clone transfected with pLEXNeo and selected for growth only in the presence of Hygromycin B (Fig. 34B, lane 6). This latter result is important in terms of the usefulness of pLEX as an expression vector since it implies that genes cloned into the downstream expression site are expressed regardless of selective pressure.
Figure 29. Southern blot hybridization analysis of clones transfected with the construct pMHBHyg - I.

*L. major* CC1 promastigotes were transfected with circular pMHBHyg and selected for growth in the presence of Hygromycin B. Eight *Hyg* clones were obtained and characterized by Southern blot hybridization analysis. Genomic DNA from clones CC1/10-1 to CC1/10-8 (lanes 1 to 8) was digested with XbaI and hybridized with A) a probe specific for the *Hyg* gene or B) a probe specific for the pUC19 vector backbone. The positions of molecular weight markers are indicated to the left in kilobasepairs (kbp).
Figure 30. Southern blot hybridization analysis of clones transfected with the construct pMHBHyg - II.

Undigested genomic DNA from the Hyg$^R$ L. major CC1 promastigote clones CC1/10-1 to CC1/10-8 (lanes 1 to 8) was electrophoresed on 0.6% agarose gels and analyzed by Southern blot hybridization using a probe specific for the pUC19 vector backbone. As suggested by the extensive hybridization observed at the top of the gel, a large proportion of the pMHBHyg-specific DNA remained very close to the sample well. The location of co-electrophoresed molecular weight markers are indicated to the left in kilobasepairs (kbp).
Figure 31. Limited DNAse I digestion of the pMHBHyg extrachromosomal element in \textit{L. major} transfectants.
Genomic DNA from the Hyg$^R$ \textit{L. major} CC1 promastigote clone CC1/10-2 (lanes 1 to 5) or the parental cell line CC1 spiked with 1 ng of the construct pMHBHyg isolated from \textit{E. coli} (lanes 6 to 10) were subjected to limited digestion with DNAse I in the presence of Mn$^{++}$ to induce random double-stranded breaks. Samples were incubated in the absence of DNAse I (lanes 1 and 6) or with 1 ul of DNAse I (diluted 1:10,000 to 1 U/ml) for 2 minutes (lanes 2 and 7), 4 minutes (lanes 3 and 8), 8 minutes (lanes 4 and 9) or 16 minutes (lanes 5 and 10). DNAse I digestion was stopped by heating to 80$^\circ$C for 10 minutes, DNAs were separated by electrophoresis on 0.8% agarose gels and analyzed by Southern blot hybridization using a probe specific for the pUC19 vector backbone. The positions of molecular weight markers are indicated to the left in kilobasepairs (kbp).
Figure 32. Restriction maps of the *Leishmania* expression vector pLEX and its derivatives.

The circular constructs used to transfect *L. major* promastigotes are depicted as circular maps showing the relevant restriction enzyme recognition sites. All constructs were derived from pMHBHyg which was previously described as a source of linear fragment for homologous recombination experiments (see Fig. 22). pLEXUL and pLEX were derived from pMHBHyg by inserting a complete GP63 intergenic region upstream and downstream of the HygR gene. pLEXNeo was derived from pLEX by inserting a NeoR gene downstream of the GP63 intergenic region. The pUC19 vector backbone is shown in white, the HEXBP 5' and 3' untranslated regions are stippled, the GP63 intergenic region is striped and the HygR and NeoR genes are shown as black arrows indicating the gene orientation.
Figure 33. Southern blot hybridization analysis of clones transfected with derivatives of the *Leishmania* expression vector pLEX. Genomic DNA was isolated from the HEXBP-deficient clone CC1/144-D (lane 1), and drug-resistant clones transfected with the circular constructs pMHBHyg (lane 2), pLEXUL (lane 3), pLEX (lane 4) and pLEXNeo (lanes 5 and 6). The pLEXNeo transfectant shown in lane 5 was selected for growth only on Hygromycin B whereas the transfectant shown in lane 6 was selected for growth on Hygromycin B and G418. DNAs were digested with XbaI and subjected to Southern blot hybridization analysis using a probe specific for the Hyg<sup>R</sup> gene. The positions of molecular weight markers are indicated to the left in kilobasepairs (kbp).
Figure 34. Northern blot hybridization analysis of clones transfected with derivatives of the *Leishmania* expression vector pLEX.

Total promastigote RNA isolated from the *L. major* parental cell line CC1 (lane 1), the HEXBP-deficient clone CC1/144-D (lane 2), and drug-resistant clones transfected with the circular constructs pMHBHyg (lane 3), pLEXUL (lane 4), pLEX (lane 5) and pLEXNeo (lanes 6 and 7) was subjected to Northern blot hybridization analysis using probes specific for A) the Hyg\(^R\) gene or B) the Neo\(^R\) gene. All transfectants were selected for growth in Hygromycin B except for transfectants CC1/144-D (lane 2) and pLEXNeo (lane 7) which were selected for growth in Hygromycin B and G418. The positions of molecular weight markers are indicated to the left in kilobases (kb)
B. DISCUSSION

The development of a novel expression vector for the stable and selectable transformation of *Leishmania* will likely have widespread applicability to the study of kinetoplastid molecular biology. Although the level of expression from pLEX-based constructs was not directly evaluated in this study, expression of a Neo\(^R\) gene cloned into the expression site was sufficiently high to confer resistance to G418. In addition, since pLEX appears to be maintained episomally in *Leishmania*, it may be possible to increase the copy number of pLEX by increasing the level of selective drug (Hygromycin B) in the growth media. Similar increases in copy number have been observed for two other kinetoplastid expression vectors (LeBowitz *et. al.*1990; Kelly *et. al.*1992). Furthermore, pLEX and derivatives thereof can also function as shuttle vectors since they contain a pUC19 vector backbone and can therefore replicate as a selectable plasmid in *E. coli*. Also, the construct pLEXNeo represents a potentially useful tool to study the mechanisms of polycistronic gene expression in kinetoplastid organisms since it is structurally analogous to a 'mini' polycistronic transcription unit. Precursor molecules and intermediates of polycistronic transcription are generally not detectable in kinetoplastid total RNA and are therefore regarded as being either exceedingly rare or processed co-transcriptionally. It may be relevant that in addition to the abundant 1.8 kb transcript, several weakly-hybridizing high molecular weight bands were detectable in the RNA of clones transfected with pLEX and its derivatives (Fig. 34). If these high molecular weight bands represent authentic RNA processing precursors, they would be of obvious use in characterizing the mechanism and resolution of polycistronic transcription.
Chapter VI - Discussion

Two other constructs designed to function as stable expression vectors in *Leishmania* have been previously described. The first, pX, was derived from a 30 kbp extrachromosomal circular element that is amplified in some methotrexate-resistant *L. major* cell lines (Kapler *et al.*1990; LeBowitz *et al.*1990). This element, known as the R region, encodes at least 10 transcripts including one which codes for the bifunctional enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS). pX is a deletional derivative of the R element in which the coding-region of the DHFR-TS gene has been replaced with a gene encoding resistance to G418 (NeoR). Genes of interest can be cloned into an expression site that is located downstream of the DHFR-TS expression site and constructs are introduced into *Leishmania* by electroporation. Transfectants are selected for growth in the presence of G418. Genes encoding *E. coli* β-galactosidase and an *L. amazonensis* surface protein (GP46A) were cloned into the expression site of pX to test its suitability as an expression vector and both proteins were efficiently expressed in transfectants (LeBowitz *et al.*1990).

The second expression vector, pTEX, is derived from the *T. cruzi* glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene tandem array and functions in both *T. cruzi* and *Leishmania* (Kelly *et al.*1992). The construct contains in the following order; the 5' untranslated region of the GAPDH locus, a multiple cloning site, the intergenic region of the GAPDH locus (the wild type locus has two genes arranged as a direct tandem array), a NeoR gene (replacing the second GAPDH gene), the 3' untranslated region of the GAPDH locus and a pBluescript vector backbone. The construct pTEX is therefore structurally similar to pLEX in that it represents a simple polycistronic transcription unit. A gene encoding bacterial chloramphenicol acetyltransferase (CAT) was cloned into the expression site of pTEX and was found to be expressed at high levels in transfectants.
The intrinsic ability of pX, pTEX and pLEX to exist as autonomously replicating extrachromosomal elements suggests that they all contain a putative origin of replication. This would not be surprising for pX since it was originally derived from a larger extrachromosomal element, however, it is intriguing that pTEX and pLEX exhibit similar activity. Several other constructs designed for homologous recombination experiments in kinetoplastids have also been shown to exist as autonomously replicating extrachromosomal elements (Laban et al. 1990; ten Asbroek et al. 1990; Cruz et al. 1991; Tobin et al. 1991; Curotto de Lafaille et al. 1992). Although the exact sequences responsible for replication of these constructs is not known, it is unlikely that sequences found on the vector backbone are involved since not all constructs lead to stable transformation (Lee and Van der Ploeg, 1990; ten Asbroek et al. 1990; Eid and Sollner-Webb, 1991). Perhaps the kinetoplastid equivalent of the yeast autonomously replicating sequence (ARS) (Bell and Stillman, 1992) is closely associated with protein-coding regions and is therefore fortuitously present on constructs used for gene replacement. The proximity of replication origins and transcription units is certainly not unprecedented as two well-characterized mammalian chromosomal replication origins are closely associated with the dihydrofolate reductase gene (Burhans et al. 1990) and the c-myc gene (Bergemann and Johnson, 1992). In addition, it may also be relevant that kinetoplastid protozoans naturally maintain a very large pool of extrachromosomal elements in the form of kinetoplast minicircles and maxicircles (Simpson, 1987). Maintenance of the large and complex kinetoplast DNA network likely requires an extensive array of highly abundant enzymatic activities. Perhaps maintenance of pTEX and pLEX as extrachromosomal elements reflects the fortuitous use of cryptic kinetoplast DNA replication origins by enzymes that are present in vast excess. Interestingly, putative minicircle replication origins have been
Chapter VI- Discussion

identified in the kinetoplastid protozoan *Crithidia fasciculata* and are thought to be comprised of a 12 nucleotide conserved sequence known as the universal minicircle sequence (UMS) (Tzfati et. al.1992). A search of the construct pLEX revealed several sequences which overlap with the UMS for at least 8 out of 12 positions (data not shown).

It has been demonstrated in the present study that pLEX-based plasmids can be used as a shuttle vector that can be transferred from *E. coli* to *Leishmania*. The ability to transfer pLEX in the opposite direction (i.e. from *Leishmania* to *E. coli*) has not yet been tested however can be predicted based on previous results using pX or pTEX. Constructs derived from pX were recovered from *Leishmania* transfectants and reintroduced into *E. coli* to produce ampicillin-resistant colonies (Kapler et. al.1990) whereas ampicillin-resistant colonies could not be produced using pTEX (Kelly et. al.1992). This activity was correlated with the presence of single copy plasmids in transfectants as pX was maintained almost entirely as a single copy plasmid whereas pTEX was maintained as a large circular multicopy concatamer. The results presented in the current study suggest that pLEX is similar to pTEX in that it is maintained as a large circular concatamer, therefore, it is unlikely that pLEX can be reintroduced into *E. coli* to produce ampicillin-resistant clones.

The usefulness of pLEX as a general purpose *Leishmania* expression vector has been subsequently enhanced by substituting the HygR gene with a gene encoding resistance to the aminoglycoside Nourseothricin (P. Joshi and J. Webb, unpublished results). Nourseothricin resistance represents a novel selectable marker in kinetoplastids and pLEX constructs carrying the NouR gene can therefore be used to complement double deletion mutants generated by homologous gene replacement that are already resistant to Hygromycin B and G418.
Chapter VI- Discussion

In summary, the construct pLEX should have wide applicability as a stable *Leishmania* expression vector. pLEX has already been used to successfully transfect several diverse species of *Leishmania* and its suitability for use in transfection of *Trypanosomes* is currently being investigated. The construct also represents a unique opportunity to study the regulation of transcription and post-transcriptional processing which are, in general, poorly understood in kinetoplastid organisms.
Chapter VII- Summary

VII. SUMMARY AND FUTURE EXPERIMENTS

The majority of the results presented in this thesis can be summarized in the form of a schematic model as shown in Fig. 35. The model depicts a region of strand separation in the 5' untranslated region of the GP63 gene. This region of non-coding DNA was shown in Chapter III to be highly conserved across geographically and clinically diverse species of Leishmania. Strand separation in this region is due to binding activity of the sequence specific single-stranded DNA-binding protein HEXBP. HEXBP was shown in Chapter IV to contain 9 CCHC-type zinc-binding domains which are thought to represent a single-strand specific zinc finger motif. DNAse I footprinting analyses indicated that the antisense strand of the GP63 untranslated region contained multiple HEXBP binding sites. As indicated in Fig. 35, each discrete site binds one molecule of HEXBP and binding at each site is likely mediated through multiple zinc finger contacts. The production of a HEXBP-deficient mutant of Leishmania by double targeted gene replacement is described in Chapter V. Characterization of the HEXBP-deficient clone suggested that HEXBP was not essential for expression of the GP63 gene in Leishmania promastigotes grown in vitro. Although the precise function of HEXBP remains to be determined, potential roles in the processes of transcription, DNA replication and recombination were presented and discussed.

The availability of a HEXBP-deficient mutant of Leishmania will be of obvious use in the elucidating the precise function of HEXBP and possibly other CCHC-containing proteins. In particular, experiments utilizing these mutants in combination with the Leishmania expression vector described in Chapter VI will likely provide definitive answers with regards to the role of HEXBP. The construct pLEXNeo, which was designed to test the function of the pLEX expression vector, provides a convenient starting
point for analysis since it contains the wild type HEXBP binding site inserted upstream of a selectable marker, the NeoR gene. Transfection of parental and HEXBP-deficient clones with the construct pLEXNeo and selection of transfectants on Hygromycin B and G418 should provide a quick, definitive indication of whether HEXBP is involved in the regulation of transcription. If a differential level of drug-resistance is observed then the function of HEXBP can be further characterized by generating derivatives of pLEXNeo in which the HEXBP binding site has been deleted or modified by in vitro mutagenesis. Constructs containing the HEXBP-binding site should also be informative if HEXBP is involved in recombinational processes. Double transfection of parental and HEXBP-deficient clones with competent and non-competent expression constructs containing selectable marker genes flanked by HEXBP binding sites could be selected for growth on both selectable markers to detect recombination events between the two constructs. Likewise, if the HEXBP binding-site represents an origin of replication then transfection of parental and HEXBP-deficient clones with constructs containing the origin and a selectable marker should provide a convenient assay for competent replication. Regardless of what the putative function of HEXBP turns out to be, the function can be verified by transfecting the HEXBP-deficient clone with an expression vector carrying a competent HEXBP gene in order to reconstitute the mutant phenotype.
Figure 35. Schematic model of HEXBP interacting with its binding site in the GP63 gene 5' untranslated region.
A single gene from the *Leishmania* GP63 tandem array is shown in the 5' to 3' orientation. The ATG indicates the GP63 gene translational initiation codon and the sense/antisense orientation. Separation of the duplex strands in the 5' untranslated region indicates the single-stranded region to which HEXBP binds. Multiple molecules of HEXBP are depicted as circles bound to the antisense strand of the 5' untranslated region.
VIII. REFERENCES


References


References


References


References


References


