SEQUENCE ANALYSIS OF THE *LEISHMANIA MEXICANANA* AMASTIGOTE SPECIFIC GENE A600

by

MIN ZHAO

B.Sc., Beijing University, P.R. China, 1987

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

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(_genetics Graduate Program_)

We accept this thesis as conforming To the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December, 2001

© Min Zhao, 2001
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.

Department of Medical Genetics

The University of British Columbia
Vancouver, Canada

Date Jan. 3, 2002
ABSTRACT

Various species of the genus *Leishmania* infect millions of people, causing a wide spectrum of diseases collectively termed as leishmaniasis, with a wide distribution in tropical and sub-tropical areas. The parasite alternates between the promastigote form predominately in the insect vector and the intracellular amastigote form in the vertebrate host, including human. Accumulating evidences suggest that developmental stage-specific genes are responsible for morphological, physiological and biochemical differences between the two life stages. Particularly, the amastigote-specific genes are of great importance in term of the host-parasite relationship. The a600 gene, previously identified in our lab, is amastigote-specific in *Leishmania mexicana*. In the present study, the gene was sequenced using Nested Deletion. Sequence analysis indicated that the a600 gene encodes a novel 93-amino acid polypeptide. Predictions based on the amino acid sequence indicated that the putative polypeptide is likely a membrane protein as a transmembrane helix region is predicted. However, the putative polypeptide was neither detected in cell lysate nor in the medium of the promastigotes transfected with the *Leishmania* expression vector pLexSat containing coding region of a600 attached with the Flag sequence. Nevertheless Northern Blot showed the presence of the RNA transcribed from the insert, suggesting the involvement of post-translational regulation. Sequence variation was also observed at the 3'-untranslated region of the a600 gene, suggesting the a600 a multiple copy gene family. The approaches for further experiment were discussed.
Abnormalities in the inflammatory response are associated with the pathogenesis of several diseases caused by Leishmania. Leishmania species are classified into three distinct lineages: New World, Old World, and American. The New World lineages are further grouped into the Panamericana species complex, which includes Leishmania braziliensis and Leishmania amazonensis. These species are responsible for cutaneous leishmaniasis, an infectious disease characterized by the formation of skin lesions. The Old World lineages include Leishmania donovani and Leishmania infantum, which are responsible for visceral leishmaniasis, a severe and potentially fatal disease. The American lineages include Leishmania panamensis and Leishmania guyanensis, which are responsible for American cutaneous leishmaniasis, a disease similar to cutaneous leishmaniasis caused by L. braziliensis and L. amazonensis.
2.3.1 Plasmid DNA isolation ................................................................. 17
2.3.2 Restriction enzyme digestion ...................................................... 17
2.3.3 Agarose gel electrophoresis ....................................................... 17
2.3.4 Extraction of DNA fragments from agarose gel ......................... 17
2.3.5 Ligation and transformation ..................................................... 17

2.4 Genomic DNA isolation from \textit{L. mexicana} .................................................... 18

2.5 Nested deletion and sequencing a600 gene ................................. 18

2.6 Analysis of sequencing data and search for the open reading frame ........ 21

2.7 Characterization of A600 protein .................................................. 21

2.7.1 Cloning A600 coding region with incorporated Flag sequence into \textit{Leishmania} expression vector ........................................... 23
2.7.2 Transfection of \textit{L. mexicana} promastigote cells by electroporation ........ 24
2.7.3 Preparation of cell lysate from \textit{Leishmania} culture .................... 25
2.7.4 Affinity purification ............................................................... 25
2.7.5 Separation of Peptides by Electrophoresis ................................. 26
2.7.6 Western analysis ................................................................. 26
2.7.7 Immunofluorescence labeling of \textit{Leishmania} ......................... 27

2.8 Northern Blot Analysis for pLexSat Expression ............................ 28

2.8.1 Hybridization probe .............................................................. 28
2.8.2 Northern hybridization analysis ............................................... 29

\textbf{CHAPTER III. RESULTS} ......................................................... 30

3.1 DNA sequence determination of the 1.8Kb fragment of a600 gene .......... 31

3.1.1 Restriction digestion of the 1.8 Kb fragment ................................. 31
3.1.2 DNA sequences of the 1.8 kb fragment and the entire a600 gene .......... 35
3.1.3 Sequence variation in the 3' UTR of the a600 gene ....................... 35
3.2 DNA sequencing analysis ...................................................................... 40
3.3 Characterization of A600 protein ......................................................... 51
  3.3.1 Localization of A600 protein by Western analysis ......................... 51
  3.3.2 Localization of A600 protein by immunofluorescence labeling .......... 52
3.4 Confirmation of the transfection ............................................................ 55
3.5 Detection of mRNA transcribed from insert a600 ORF by northern blot .... 55

CHAPTER IV DISCUSSION ........................................................................ 59
4.1 The a600 gene encodes a novel protein .............................................. 59
4.2 The a600 gene is likely a multiple copy gene family .............................. 60
4.3 The putative A600 protein is likely a membrane protein .................... 61
4.4 The fused A600 protein was not detected indicating the possible post-translational regulation of a600 gene expression ........................................ 64
4.5 Proposed experiments for future study .................................................. 67

CONCLUSION .......................................................................................... 69

REFERENCES .......................................................................................... 70
LIST OF TABLES

Table 3.1. Comparison of Leishmania codon frequency with the codon frequency of potential coding regions in a600 gene .......................... 43

Table 3.2. Output of SIGNALP Prediction .................................................. 46

LIST OF FIGURES

Figure 1.1. Life cycle of Leishmania............................................................... 5

Figure 2.1. The map of Leishmania expression vector pLEX .......................... 16

Figure 2.2. Procedures of Nested deletion for sequencing amastigote specific gene a600 ............................................................................. 20

Figure 2.3. Procedures for Localization of A600 polypeptide ...................... 22

Figure 3.1. Relative positions of s800, s600 and the 1800 bp fragments within the a600 transcript................................................................. 30

Figure 3.2. Multiple cloning sites of pGEM-T............................................. 31

Figure 3.3 Restriction digestion of the 1.8 kb fragment................................ 33

Figure 3.4. Subfragments generated by nested deletion ............................... 34

Figure 3.5. Contig alignment for the 1.8 kb fragment.................................. 37

Figure 3.6. The cDNA sequence of A600 gene ......................................... 38

Figure 3.7. Sequence alignment of m13f and a4 showing the sequence variation in 3'UTR of a600 ................................................................. 39
Figure 3.8.  Sequence identity between a600 and L. major Friedlin assembled shotgun reads from chromosome 34 ................................................................. 41

Figure 3.9.  Results of SIGNALP Prediction ........................................................................ 47

Figure 3.10. Results of the PhdTopology Refinement And Topology Prediction ........ 49

Figure. 3.11. Transmembrane Prediction using "DAS" ....................................................... 50

Figure 3.12. PCR amplified a600 coding sequence from genomic DNA ...................... 53

Figure 3.13. Western blot analysis for detection of A600 protein in transfected L. mexicana........................................................................................................ 54

Figure 3.14. Confirmation of the transfection by PCR with sat-1 gene primers .......... 57

Figure 3.15. Northern blot of transfectant RNA with a600 coding region as probe ...... 58
ACKNOWLEDGEMENT

I am mostly grateful to my supervisor, Dr. Rob McMaster, for his excellent guidance and support. His patient and everlasting encouragement are essential for me to overcome the difficulties during my study years. My sincere thanks goes to Drs. Alice Mui, who kindly provide me valuable help for localization study and monoclonal anti-FLAG antibody, and a lot more. I wish to gratefully acknowledge all colleagues in McMaster lab, Dr. Tanya Nelson, Dr. Jaime Belletin, Dr. Corinna Warburton, Dr. Nicole Lawrence, Dr. Ben Kelly, Angus Murray, Dr. Celia Alves de Almeida, for their suggestions and help of all kinds, as well as their friendship. It is an unforgettable experience to have been working with them through these years.

Drs. Dixie Mager and Dr. Fumio Takei, members of my supervisory committee, have provided me valuable advice. Their effort made this work much better. I gratefully acknowledge the Faculty of Graduate Study, University of British Colombia, for granting me the University Graduate Fellowship for two years.

My deep gratitude is due to my husband Tongli, for his exceptional support and help in all ways he can find, that finally makes this work a reality. I also wish to thank our daughter Angeni and our son Arthur, their arrival and presence with wonderful sense of love, has been a good motivation to complete this study. My special thanks goes to my parents, for their love and everlasting expectation, which I hope will be satisfied a bit by this thesis.
CHAPTER I. INTRODUCTION

1.1 The diseases caused by Leishmania

Various species of the genus *Leishmania* infect millions of people worldwide, causing a wide spectrum of diseases collectively termed as leishmaniases. The diseases range in severity from a spontaneously healing skin ulcer to overwhelming visceral disease. Leishmaniases are prevalent throughout the tropical and sub-tropical regions of Africa, Asia, the Mediterranean, Southern Europe and South and Central America. It is estimated that approximately 12 million people are currently infected and a further 350 million are at risk of acquiring leishmaniases in 88 countries. Worldwide, two million new cases occur each year, and a 10th of the world's population is at risk of infection (WHO web site, 2000).

Some recent studies (Martinez et al. 1998; Delgado et al. 1998) have shown an underestimation of the severity of these infections and of their socio-economic impact. Moreover, the incidence of leishmaniases has apparently increased during the last decade, along with increasing poverty, the emergence of drug resistance and immunosuppression (Alvar et al., 1997).

A patient with leishmaniases may present with one of three quite distinct clinical forms -- visceral, cutaneous, or mucocutaneous. Visceral leishmaniases is caused by *L. donovani, L. infantum* and *L. chagasi*. It is the most severe form of the disease. It is characterized by irregular bouts of fever, substantial weight loss, swelling of the spleen and liver, and anemia. It is fatal without treatment and may be fatal despite treatment (Wilson and Streit 1996).
Cutaneous leishmaniases presents as a skin ulcer at the site of the sand fly bite and generally heals spontaneously with a scar within three to six months. A rare form of cutaneous leishmaniases known as diffuse cutaneous leishmaniases presents with disseminated skin lesions. The cutaneous forms of leishmaniases are the most common and represent 50-75% of all new cases, which is caused by *L. major*, *L. tropica*, and *L. mexicana*.

Mucocutaneous leishmaniases, or *espundia*, produces lesions which can lead to extensive and disfiguring destruction of mucous membranes of the nose, mouth and throat cavities. It is an uncommon consequence of cutaneous leishmaniases that may present even years after the initial skin ulcer has healed (El-Hassan et al. 1995). The disease is caused by *L. braziliensis*.

With the spread of HIV, visceral leishmaniases has become increasingly prevalent, and unusual presentations often occur. Reactivation of asymptomatic or previously "healed" *Leishmania* infections is common with the onset of AIDS. *Leishmania* species that normally cause only cutaneous disease may present with visceral leishmaniases. Importantly, co-infections of *Leishmania* and HIV are often resistant to treatment and substantially accelerate the progress of AIDS (Lopez-Velez et al. 1998).

Outcome of the *Leishmania* infection is determined by interactions between the host and parasite, which are governed by the genomes of the host and parasite. Thus, the genetic information from both human and parasite and the emergence of new tools such as microarray technologies will allow us to gain an understanding of the interaction between parasite virulence factors and host response factors. Molecular knowledge of the host-parasite interaction will facilitate developing of new treatments.
1.2 Taxonomy

The genus *Leishmania* belongs to the *Trypanosomatidae* family and the order *Kinetoplastida*. The genus *Leishmania* comprises some 30 species of morphologically similar kinetoplastid protozoa, including 21 species which are responsible for a spectrum of human diseases ranging from mild to fetal infections (Shaw 1994). The genus *Leishmania* can be taxonomically separated into three main groups: the Old World subgenus *L. (Leishmania)*, the New World subgenus *L. (Leishmania)* and the New World subgenus *L. (Viannia)*.

The taxonomy of this genus is complex and often confusing. In general, species differentiation is not based on the morphology of the organism but on the pathology and symptoms of the disease, site of infection, vector species and reservoir hosts. Thus, combinations of biological (Lainson et al. 1987), immunological (Handman and Hocking 1982) and molecular criteria (e.g., Cupolillo et al. 1995) have been used to identify and classify *Leishmania* species. Five assemblages of *Leishmania* (species complexes) have been distinguished, 3 for old world and 2 for new world. In most recent classification these assemblages are assigned to one of two primary lineages, i.e., subgenus *Leishmania Leishmania* and subgenus *Leishmania Viannia* (Croan et al. 1997).

1.3 Life cycle of *Leishmania*

The *Leishmania* has two distinct stages in its life cycle (shown in Figure 1.1). In its vertebrate host the parasite resides as an amastigote within mononuclear phagocyte in parasitophorous vacuoles. Whereas in the insect vector, i.e. sand fly of the genera *Lutzomyia* and *Psychodopygus*, the parasite exists as an extracellular promastigote.
The *Leishmania* is transmitted to humans by the bite of the infected female sand fly. The sand fly is usually infected with one species of flagellate protozoa belonging to the genus *Leishmania*. About 30 species of sand flies can become infected when taking a blood meal from a reservoir host. Humans are usually accidental hosts of these parasites; natural hosts include a variety of rodents, small mammals, and dogs.

In the vertebrate host, the parasite exists under the non-mobile intracellular amastigote form. Amastigote is spherical in shape and 2 to 3 μm in diameter. The ultrastructure of the nucleus is similar to nuclei of other eukaryotic cells. The nucleus is contained within two-layer nuclear membranes punctuated by pore structures and contains a prominent nucleolus (Molyneux and Killick-Kendrick 1987). The kinetoplast, a specialized mitochondria structure that is unique in the order of *Kinetoplastida*, is found posterior to the basal body. The flagellum is contained within a flagella pocket in amastigote instead of extending outside of the plasma membrane. The sub-cellular organelles including the smooth and rough endoplasmic reticulum and Golgi apparatus that are common in other eukaryotic cells are also observed in the amastigote cytoplasm. A variety of vacuoles such as lysosomes are present.
Figure 1.1 Principal events in life cycle of *Leishmania*: (1) Sandflies inoculate promastigotes in vertebrate; (2) Attachment and engulfment of promastigotes by macrophages; (3) Fusion of phagosome and lysosome; (4) Differentiation of promastigotes into amastigotes; (5) Division of amastigotes; (6) Rupture of parasitized macrophage; (7) Infection of a different macrophage; (8) Sandfly infection after a blood meal; (9-13) Development of promastigotes in the sandfly gut; (14) Metacyclic promastigotes are inoculated into the next vertebrate. Modified from Chang (1990).
Sand flies become infected when they acquire blood with mononuclear cells that contain amastigote. During digestion of the blood meal, amastigote initiate their differentiation into the mobile extracellular promastigote form, which will attach to the midgut epithelium to avoid being excreted together with the digested blood meal (Descoteaux and Turco 1999). The number of morphological forms and the location of their development within the sand fly gut varies among the *Leishmania* species and the sand fly species (Walter *et al.* 1989; Lawyer *et al.* 1990). In general the promastigote is spindle-shaped, flagellated and highly motile. The size averages from 10 to 15μm in length and 2 to 3μm in width. The flagellum extends from the basal body out through the flagella reservoir at the anterior pole. The precise sequence of events after the sand fly inoculation of promastigote into mammals is not yet clear. Previous observation suggests that promastigotes are phagocyted by macrophages, convert into amastigotes within these cells, multiply and then infect neighbor mononuclear cells (Wilson *et al.* 1987).

*Leishmania* undergo profound changes morphologically and physiologically as they pass between the mammalian host and the insect vector. The relationship and the interaction between the parasites, their host, and their vectors is derived through long term coexistence and selected for by co-evolution (Schnur and Greeblatt 1995).

Virulence is acquired during the process that the dividing, non-infective promastigotes transform into the non-dividing, infective form (Sacks 1989). These infective promastigotes detach from the gut epithelial cells and migrate towards the anterior end of the digestive tract. Upon the subsequent blood meal, the infective form of promastigotes are inoculated into the mammalian host, where they must successfully
evade and resist massive defense mechanisms such as complement-mediated lysis, to
ultimately bind and enter mononuclear phagocytes by a receptor-mediated process
(Descoteaux and Turco 1999).

Upon entering the mononuclear phagocytes, infective promastigotes reside inside
a parasitophorous vacuole (phagosome) to avoid degradation and establish conditions
favorable to their proliferation. Ultimately, infected macrophages rupture and release the
amastigotes into the surrounding environment where they can infect neighboring
macrophages.

Promastigotes and amastigotes can be maintained independently and continuously
in laboratory in vitro and in animals, respectively. Promastigotes from all Leishmania
species can be grown in a number of culture systems for in vitro cultivation. A gradual
shift to acidic pH and an increase of temperature can initiate the conversion of
promastigote into amastigote like forms in cell-free cultures (Zilberstein et al. 1991). It
was observed that some proteins synthesized by L. mexicana promastigotes during the
transformation in culture are the same as those synthesized by the parasite in their
mammalian host (Zilberstein et al. 1991). This suggests that the in vitro conversion
represent, to certain degree, the in vivo process of differentiation.

1.4 Basic Genetics and Genomic features of Leishmania

The nuclei of Leishmania contain chromosome-sized pieces of DNA that are
referred to as chromosomes of those higher eukaryotes. In general, Leishmania is
regarded as diploid, but no sexual cycle has been observed. Replication is by mitosis
(binary fission) at both stages of the life cycle, and the chromosomes do not condense at any phase of the cell cycle (Ivens and Blackwell 1999).

The haploid genome of Old World *Leishmania* species has been shown to contain 36 chromosomes defined as physical linkage groups and it is totally conserved among all those Old World species (Wincker *et al.* 1996). When comparing the karyotype of Old World species with that of their New World counterparts, 32 of the 36 linkage groups were conserved among all species (Britto *et al.* 1998). It was found that the *Leishmania* genome size is ~35 Mb and chromosome sizes range from 0.3 to 2.8 Mb. However, different numbers of chromosomes have been found in the New World species, 35 for the subgenus *Viannia* and 34 for the species of the subgenus *Leishmania* (Britto *et al.* 1998).

*Leishmania* genes have certain unusual features that are found in the family of trypanosomatids. For example, they occur frequently in tandem repeats, transcription is polycistronic or discontinuous (reviewed by Schnur and Greeblatt 1995). The protein-coding genes are organized into large clusters with many genes adjacent on the same DNA strand (Myler and Stuart 2000). The *Leishmania* genome is relatively GC rich (63%) (Alvarez *et al.* 1994), and contains ~30% repeated sequences (Ellis and Crampton 1988). A high degree of sequence conservation for the protein-coding genes at both the DNA (91–96%) and protein (90–100%) levels among some species has been observed, whereas the intergenic regions are less well conserved (78–85%) (reviewed by Myler and Stuart 2000).

The transcription in *Leishmania* is, in general, regulated post-transcriptionally (Aly *et al.* 1994). As *Leishmania* genes are often organized in tandem repeats, many of which are transcribed polycistronically. Non-repeated genes of related function could
also occur in long transcript units. Extensive post-transcriptional processing is then required to yield mature mRNA (Ivens and Blackwell 1996). A most significant event in this process is trans-splicing of a “spliced leader RNA” (SL) to the 5’ end of all mRNA molecules. The SL is associated with translation initiation (Agabian 1990), however, the mechanism of translation initiation in *Leishmania* is still not clear. In contrast to other eukaryotes, no intron has been discovered within any of the *Leishmania* protein-coding genes (Myler and Stuart 2000).

*Leishmania* together with all other members in the order of *Kinetoplastidae* have a unique feature, that is the presence of kinetoplast. It is a specialized mitochondria-like organelle and contain substantial amount of extra-nuclear DNA (kDNA) (Stuart and Feagin 1992). There are two types of kDNA, maxicircles and minicircles. They are organized as interlocked molecules and self replicate synchronously with cell mitosis (Lighthall and Giannini 1992). The maxicircles fulfill the role of mitochondria DNA in other eukaryotes and the minicircles hold the maxicircles together in a network whose replication products are readily segregated along with the flagella basal bodies during cell division. The minicircles, if transpired, also encode guide RNAs that are crucial to the editing of faulty maxicircle transcripts. This ensures that the *Leishmania* can continuously switch its pattern of respiration in the vector and thereby survive the transmission from the mammal host to the insect vector (Hide et al. 1997). Kinetoplast gene transcripts also undergo extensive editing, including deletion and addition of uridine (Stuart and Feagin 1989).
1.5 Developmental stage-specific gene expression in Leishmania

In the course of Leishmania life cycle, the parasite differentiates from promastigote in the sand fly vector to amastigote in the mammalian host. The parasite adapts to dramatically changed environments by undergoing a profound morphological and physiological transformation. The molecular mechanisms regulating the differentiation and survival of different life stages are poorly understood. Studies have shown that this differentiation is largely due to the regulation of gene expression (Wu et al. 2000; Zhang et al. 2001). The molecular basis of the transformation between promastigote and amastigote has been studied by identification and characterization of stage-specific genes as well as genes closely linked to them.

There is considerable variation in mRNA stability for many genes during the life cycle (Aly et al. 1994). The abundance of a given transcript at a particular stage of the life cycle often indicates the relative importance of that gene (Manger et al. 1998). This is supported by several studies that showed the transcript accumulation of amastigote-specific genes in amastigotes not due to an increase of RNA synthesis, but to an increase in mRNA stability in amastigotes (Beetham et al. 1997; Charest et al. 1996; Moore et al. 1996; Wu et al. 2000). These findings also suggest that the differentiation of Leishmania from promastigote into amastigote triggers the induction and/or degradation of factors that regulate transcription. Amastigote-specific regulation may also occur at the level of translation and/or post-translation (Bhaumik et al. 1991; Wu et al. 2000). These studies further revealed that differential expression of genes involves mainly the 3' -UTR sequences of the mRNA (Beetham et al. 1997; Charest et al. 1996; Moore et al. 1996).
The importance of 3’-UTR sequences and/or inter-genic regions in the stage-specific regulation of *Leishmania* transcription is also supported by the observation that differential expression of multiple copy gene families, such as GP63 (Ramamoorthy et al. 1995), HSP70 (Quijada et al. 1997), and β-tubulin (Coulson et al. 1996).

Several *Leishmania* genes have been demonstrated to be specifically associated with the amastigote stage. The A2 gene represents the first amastigote stage-specific gene characterized in *L. donovani* (Charest and Matlashewski 1994). The A2 protein was identified and proved to be specifically expressed in amastigote cells (Zhang et al. 1996). The A2 protein is comprised predominantly of a highly conserved repetitive element. It also shares some characteristics with major antigens expressed developmentally by several other unrelated pathogenic human parasites. It has been shown that the elevated production of the A2 transcripts could be induced in cultured *Leishmania* by shifting temperature and pH to mimic the condition of the passage from the insect vector to the phagolysosomal compartment of the macrophage cell (Charest and Matlashewski, 1994). Later on it was also found that A2 developmental expression during the promastigote-to-amastigote cyto-differentiation is mediated through differential RNA stability and involves the A2 mRNA 3’-UTR region (Zhang and Matlashewski G., 1997).

Increasing numbers of molecules have been demonstrated to be specifically associated with the amastigote stage. Such as, specific glycosphingolipids from *L. amazonensis* (Straus et al. 1997), superoxide dismutase from *L. chagasi* (Paramchuk et al. 1997), the proteophosphoglycan molecule(s) in *L. mexicana* (Peters et al. 1997), and more recently, the immunologically protective P-4 antigen from *L. pifanoi* (Kar et al. 2000).
It has also been found that these amastigote-specific gene products are related to special biological functions in amastigote stage. The A2 protein is necessary for \textit{L. donovani} survival in a mammalian host (Zhang and Matlashewski G. 1997). The \textit{Leishmania} superoxide dismutase is considered to be involved in the detoxification of host cell radical oxygen intermediates known to be deleterious to the intracellular amastigote (Paramchuk et al. 1997). The proteophosphoglycan molecule appears to have a role in parasite vacuole formation within the infected macrophage (Peters et al. 1997). The P-4 nuclease, according to its location, may play roles in RNA stability (gene expression) or DNA repair (Kar et al. 2000).

In summary, accumulating evidences support that developmental stage-specific genes or the products of these genes are responsible for morphological, physiological and biochemical differences between the two life stages. Particularly, the functions of the amastigote-specific genes play important roles in parasite virulence, pathogenicity, and intracellular survival. Those amastigote-specific functions are critical in term of the host-parasite relationship.

\textbf{1.6 Rationale and the focus of the present study}

During its life cycle, \textit{Leishmania} adapts to the different environments by undergoing a series of morphological and biochemical changes. A number of genes have been found to be stage-specifically expressed. The rationale is that these genes and/or their products are responsible for the stage-specific functions. Amastigote-specific genes are of great interest, as they are believed to be involved in the interaction between the parasite and the mammalian host. Thus, investigating amastigote-specific genes and gene
families would aid our understanding of *Leishmania* virulence and pathogenesis in mammalian host and would provide new options for the control of leishmaniases.

Up to date, few of these genes have been identified, and the mechanisms of their stage-specific regulation are under investigation. A600 gene was recently identified as an amastigote-specific gene using PCR based cDNA subtractive hybridization.

The present study was to characterize the a600 gene by 1) DNA sequencing study; 2) sequence analysis such as homologous searching and secondary structure prediction of the putative gene product; and 3) investigation of the cellular location of the A600 protein, which would provide some clue for understanding its biological functions. The long-term objective is to determine the role of the a600 gene and proteins in the interaction between the parasite and its host. In the hope of fighting the Leishmaniases that is increasingly developing resistance to the currently available drugs, this study may lead to the development of new vaccines for a better control of the disease.
CHAPTER II. MATERIALS AND METHODS

2.1 Axenic culture of Leishmania cells

2.1.1 Species and strains

Leishmania mexicana (WHO designation MNYC/B2/M379), clone BK1 was used for culturing and transfection.

2.1.2 Culture of Promastigotes

L. mexicana promastigotes were cultured at 26°C in 1 X M199 medium (Gibco BRL) supplied with 10% FCS, 40mM HEPES (pH7.4), 0.1 mM Adenine, 0.0005% Hemin, 0.035% Na₂CO₃ and 0.1 mg/ml antibiotic penicillin and streptomycin. Cell stocks were stored in liquid nitrogen with 10% Glycerol. To start a culture from frozen stock, cells were thawed at 37 °C in a water bath and diluted with 10 ml pre-warmed medium followed by spinning to remove the glycerol. Cells were re-suspended in 5 ml pre-warmed medium and maintained in a non-humidified incubator with closed lids, at the density of 5 x 10⁵ to 2 x 10⁷/ml.

2.1.3 Culture of Amastigotes

Axenic culture of amastigote was converted from stationary phase culture of promastigote by transferring the promastigote culture from 26°C to 32.5°C for 24 hours and gradually shifting to acidic pH by adding one volume of 1 X UM54 medium (1 x M199 supplied with 0.25% glucose, 0.5% tripticase, 25mM HEPES, 5.14 mM glutamine 0.035% Na₂CO₃, 100 unit/ml antibiotic penicillin and 100 μg/ml streptomycin, pH 5.5)
plus 20% FCS in every 24-hour. Most of cells were converted into amastigote-like form by this process after approximately one week.

2.2 *Bacterial strains and Leishmania expression vectors*

E. coli strain DH5αF’ (BRL, Gaithersburg, Maryland) and plasmid vector pGEM-T (Promega) were used for molecular cloning.

The 7.3Kb *Leishmania* expression vector pLEXSat was constructed in our laboratory previously (Joshi et al. 1995). This construct contains the coding region of bacteria sat-1 gene that encodes the streptothricin acetyltransferase, which is able to inactivate the nourseothricin (Ns), a member of the streptothricin group of antibiotics produced by *Streptomyces noursei*. *Leishmania major* promastigotes were killed in presence of 32μg/ml and 50μg/ml of Ns in liquid and solid M199 media, respectively. The 1.7 Kb 5’ and 1.3Kb 3’ untranslated regions of *Leishmania major* HEXBP gene flanking the sat-1 gene, the intergenic region of *Leishmania major* gp63 and multiple cloning sites (MCS) from part of pBluescript MCS, hence constitute a tandemly arranged genes as shown in Figure 2.1. Such intergenic fragments from loci composed of tandemly repeated genes, as previously shown in *Leishmania* (Huang and Van der Ploeg 1991; Ullu et al. 1993), provide sufficient information required for polyadenylation and trans-splicing for expression of both upstream and down stream genes. The pLEX vectors were also tested to be expressing neo gene in *Leishmania major* promastigotes.
Figure 2.1. The map of *Leishmania* expression vector pLEX

(adopted from Joshi et al. 1995)
2.3 Routinely used molecular biology methods

2.3.1 Plasmid DNA isolation

Plasmid DNA was isolated from 3ml overnight cultures inoculated with a single colony using Qiagen kit (Qiaprep Spin Miniprep Kit). Large scale plasmid DNA isolation was carried out using Qiagen Maxi kit (Qiagen Plasmid Max Kit) from 250ml overnight culture. DNA was then quantified spectrophotometrically and by gel electrophoresis.

2.3.2 Restriction enzyme digestion

Restriction enzyme digestion of plasmid DNA used for mapping and cloning was carried out with 2-5 units of restriction enzymes per µg DNA in appropriate buffer supplied by the manufacturers at 37°C, unless otherwise instructed by the manufacturers.

2.3.3 Agarose gel electrophoresis

DNA fragments were size separated using agarose gel (0.7%, 1.0%, 1.2%) in 0.5 x TBE electrophoresis buffer (44.5 mM Tris Borate and 1.25 mM EDTA, pH 8.3). DNA bands were visualized by incorporating 0.5µg /ml ethidium bromide into gels or by staining gels in 0.5 µg/ml ethidium bromide solution.

2.3.4 Extraction of DNA fragments from agarose gel

DNA fragments were extracted from agarose gel using Qiagen gel extraction kit (Qiaex II Gel Extraction Kit) according to the manufacturer's instruction.

2.3.5 Ligation and transformation

Ligation reactions were carried out with 5 units of T4 ligase in 1 x ligation buffer (50mM Tris-HCl pH 7.5, 10mM MgCl2, 10 mM DTT, 1mM Spermidine HCl, 1mM ATP
pH 7.0, 0.5% BSA) in 10µl at 12-18°C overnight. 100ng vector DNA and 3 times insert DNA in molar excess were used in the ligation reaction.

E. coli cells were made competent using the CaCl₂ method (Lederburg and Cohen 1974). Plasmid DNA (50ng) was added into 50µl competent cells in ependorf tubes and kept on ice for 45 minutes prior to 3 minutes heat shock at 37°C. Cells were then added to 900 µl LB medium and kept at 37°C for one hour with shaking. 200µl transformed cells were plated onto LB agar plates containing 100µg/ml ampicillin and incubated overnight at 37°C. Colonies that were generated from transformants were picked and cultured in LB media containing the same antibiotic.

2.4 Genomic DNA isolation from L. mexicana

Genomic DNA was isolated from 3 x 10⁸ cells culture of promastigote according to the procedure as described by Medinaacosta and Cross (1993), and the purified DNA pellet was dissolved in distilled water. DNA was treated with 20µg/ml RNase A at 37°C for 30 minutes to remove RNA and kept at 4°C overnight before quantification by spectrophotometry and agarose gel electrophoresis.

2.5 Nested deletion and sequencing a600 gene

A cDNA fragment, s800, was previously identified to be transcribed preferentially in amastigote, and the fragment was extended toward the spliced leader and resulted in an 1800bp fragment (Belletin, 1999) in our lab. The objection of the present study was to characterize this gene by sequencing the 1800bp fragment using nested deletion.
In this approach, the Exonuclease III digestion was carried out for creating unidirectional nested deletions in double-stranded DNA for sequencing the 1.8kb fragment of \textit{a600} gene. This would enable the original 1.8 kb large fragment to be divided into smaller, partially overlapping segments, therefore the entire fragment was sequenced with the universal sequencing primer desired for the original plasmid.

The procedures for nested deletion were based on the instruction of "Exo-Size™ Deletion Kit" version 2.0 (New England Biolabs, 1998) shown in Figure 2.2 with minor modifications. 10 µg of purified pGEM-T containing the 1.8 Kb fragment was used for double restriction digestion. The linearized DNA was phenol/chloroform extracted and ethanol precipitated, and re-suspended in 1 X Exo III buffer (66 mM Tris-HCl pH 8.0, 0.66 mM MgCl$_2$).

For ExoIII nuclease digestion, a set of 9 deletions was created. Mung Bean Nuclease alone was not able to create the blunt ends, even with 3 X the amount recommended by the manufacturer, based on the fact that the plasmid was not ligated. A modification was made to carry out the T4 polymerase reaction in presence of dNTP in 1 X ligase buffer (50 mM Tris-HCl pH 7.8, 10mM MgCl$_2$, 10mM DTT, 1mM ATP and 25µg/ml bovine serum albumin) at 16°C for 20 minutes prior to the ligation. T4 polymerase was heat inactivated at 75°C for 10 minutes. Without further purification, all deletions were subjected to ligation after filling the missing bases to generate the blunt ends. The transformants that carried the re-ligated plasmid were screened according to the manufacturer's instruction. Sequencing reactions were carried out with Big Dye chemistry (ABI-CE1, Version 3.0) using M13 forward primer.
Figure 2.2. Procedures of Nested deletion for sequencing amastigote specific gene a600
2.6 Analysis of sequencing data and search for the open reading frame

Sequencing data of all sub-fragments were integrated into a single contig using Seqman (version 3.03). Both DNA and the putative protein sequences were used to search the database for similarities. SIGNALP Prediction (Nielsen et al. 1997) was used to predict the potential cleavage site. The secondary structure of the putative A600 protein was predicted by PhdTopology Refinement and Topology Prediction (Rost et al. 1996) and “DAS ” Transmembrane prediction server (Cserzo et al. 1997).

2.7 Characterization of A600 protein

The procedures of characterization of the A600 protein are shown in Figure 2.3. The 10-amino acid Flag epitope was to be incorporated to the C-termini of a600 coding region predicted from the consensus sequence, therefore the commercially available monoclonal antibody against the Flag epitope was to be used to analyze the expression of a600 coding region.
Figure 2.3. Procedures for Localization of A600 polypeptide
2.7.1 Cloning A600 coding region with incorporated Flag sequence into 
*Leishmania* expression vector

2.7.1.1 PCR amplification to obtain the a600 ORF incorporated with the sequence encoding Flag epitope

The forward primer had a Cla I site before the start codon of a600 ORF:

\[ \text{5'} \rightarrow \text{cca tcg atg ccc tct atg ctc aac} \leftarrow \text{3'}. \]

The reverse primer had a Sma I site and a Spe I site flanking the 17 nucleotides at 3' end of the a600 coding sequence and 30 nucleotides encoding Flag epitope:

\[ \text{5'} \rightarrow \text{gac tag tta ctt gtc gtc atc gtc ctt gta atc ccc ggg} \]
\[ \text{cac cat gtc cgc aga tg} \leftarrow \text{3'} \]

The Flag sequence and its translation is shown below.

\[
\begin{array}{cccccccccccc}
\text{P} & \text{G} & \text{D} & \text{Y} & \text{K} & \text{D} & \text{D} & \text{D} & \text{K} \\
\text{CCCGG} & \text{GGATTACAAG} & \text{GACGATGACG} & \text{ACAAG} \\
\text{gggcc} & \text{cctaatgttc} & \text{ctgctactgc} & \text{tgttc} \\
\end{array}
\]

100 ng of genomic DNA isolated from *L. mexicana* promastigote was used for PCR reaction. The 100μl reaction mix contained:

- **Template DNA**: 100 ng
- **10 X KlenTaq PCR reaction buffer**: 10 μl
- **10 mM dNTP**: 2 μl
- **20 μM 5’primer**: 2 μl
- **20 μM 3’primer**: 2 μl
KlenTaq Polymerase Mix  
2 μl

Milli Q H₂O  
add up to 100 μl

Reaction was hot-started at 95°C for 1 minute and run through 27 cycles of 30 seconds at 95°C, 30 seconds at 65°C and 3 minutes at 68°C.

PCR product was purified from agarose gel slice using QIAEX II Gel Extraction Kit, (QIAGEN Inc. Canada) and quantified by measuring absorbance at 260 nm.

For the 324bp PCR fragment, 10 ng purified product was employed for direct sequencing. Sequencing reaction was carried out as described in 2.5.3.

2.7.1.2 Cloning the a600 coding region into *Leishmania* expression vector

20 ng purified PCR fragment PCR fragment was first cloned into pGEM®-T vector according to the manufacturer's instruction. The cloning was confirmed by Cla I and Spe I double restriction endonuclease digestion and sequencing reaction. The clone with correct sequence was used for the sub-cloning into pLexSat.

Both insert and pLexSat plasmid DNA were digested with Cla I/Spe I. The ligation and transformation were carried out as described in 2.3.5. The sub-cloning was confirmed by Cla I/Xba I digestion and sequencing reaction. The clone that generated the expected size of insert band on agarose gel and correct sequence was used for transfection of *L. mexicana* promastigotes.

2.7.2 Transfection of *L. mexicana* promastigote cells by electroporation

Log phase cells, i.e. 1 x 10⁷/ml, were pelleted by centrifugation at 3000g and washed two times with electroporation buffer (21 mM Hepes pH 7.4, 137 mM NaCl, 5 mM KCl, 0.7 mM NaH₂PO₄ and 1.1% Glucose). Cells were re-suspended at 10⁸/ml in...
electroporation buffer, cooled on ice. Mixed 6 μl plasmid DNA with 0.4ml ice cold cell suspension in cuvettes. Electroporated at 2.25kV/cm (0.45 kV for 0.2 cm cuvette), cap extender 500 μF. after electroporation the cells were incubated on ice for 10 minutes and then transferred into T-flask using pre-warmed M199, 10%FCS. Recovered cells at 26°C overnight and plated them on 1 x M199 agar plate containing 50 μg/ml Nourseothrycin. Vector plasmid pLexSat was also used to transfect promastigotes for obtaining the negative control.

Clones of transfectant were picked up to inoculate the promastigote culture. In order to obtain high copy number of clones, drug concentration was increased from 32 μg/ml to 60 μg/ml when cells were converted into amastigotes.

2.7.3 Preparation of cell lysate from *Leishmania* culture

Promastigotes and amastigotes were harvested at 1-3 x 10^7/ml and 5 x 10^7/ml respectively. Cells were separated from the supernatant by centrifugation at 3000g for 5 minutes, the supernatants were saved for immunoprecipitation. Cells were then washed 3 times in ice cold Tris buffered saline (TBS). After the last removal of the wash, cells were re-suspended in 0.5 ml 2 x gel loading buffer (0.125M Tris/HCl pH 7.4, 2% SDS, 5% 2-mercaptoethanol, 20% glycerol, 0.0025% bromphenol blue), termed as crude cell lysate.

2.7.4 Affinity purification

Monoclonal anti-Flag antibody covalently attached to agarose beads (Anti-Flag M2 Affinity gel, A1205, Sigma) was washed in ice cold TBS in 0.2% Tween-20 by centrifugation at 5000 rpm according to the manufacturer’s instruction. A 50 % slurry of
Anti-Flag agarose beads was made in the same buffer. 20 μl of such slurry was used to affinity purify 1.5ml supernatant at 4°C overnight with gentle agitation. Agarose beads were pelleted by spinning at 5000 rpm for 30 seconds and washed 3 times in 1 ml ice cold TBS in 0.2 % Tween-20. After the removal of the last wash, 75 μl 2 x loading buffer was added to the beads and boiled for 3 minutes. Beads were then removed by centrifugation at 13,000 rpm for 2 minutes.

Crude cell lysate was diluted 1 in 10 (v/v) with TBS in 2% Zwittergent and 100 mM Iodoactamid (IAA). 50 μl of 50 % a-Flag agarose slurry was mixed with 1 ml diluted cell lysate at 4°C overnight with gentle agitation. After washing 3 times in TBS in 2 % Zwittergent, beads were re-suspended in 100 μl 2 x gel loading buffer and boiled for 3 minutes, then removed by centrifugation at 13,000 rpm for 2 minutes.

2.7.5 Separation of Peptides by Electrophoresis

High resolution SDS-PAGE was used for peptide separation. The gel system consists of a 16.5% separating gel, a 10% spacer gel and 4% stacking gel. The gel buffer is Tris-Tricine instead of Tris-Glycine. The running buffers include cathode buffer (0.1M Tris, 0.1 M Tricine, 0.1% SDS) and the anode buffer (0.2M Tris-HCl, pH8.45). Samples were electrophoresed at 150 V till the bromophenol blue run to the bottom of the gel.

2.7.6 Western analysis

2.7.6.1 Protein transfer

Both the pre-wetted Immobilon-P transfer membrane (Millipore) and the gel were equilibrated in transfer buffer (25mM Tris base pH8.5, 192 mM glycine) for 10 minutes.
Proteins were transferred into Immobilon membrane in 11 transfer buffer in small Hoeffer electronic transfer unit at 400 mA for one hour with circulation of cold water for cooling.

2.7.6.2 Antibody staining

Membrane was blocked in blocking buffer (3% BSA in TBS with 0.5 % Tween-20) at 37°C for one hour, then washed 3 times, each for 5 minutes in washing buffer (0.2% Tween-20 in TBS). Biotinylated a-FLAG monoclonal antibody (Anti FLAG Biotinylated M2 monoclonal antibody, F9291, Sigma) was diluted in TBST at 1μg/ml. The blot was incubated in diluted anti-FLAG for 3-5 hours with gentle agitation, followed by the same washing procedure. HRP/streptavidin conjugated secondary antibody was added to the membrane in TBST at 1:10,000 and incubated for one hour. Membrane was thoroughly washed 3 times with TBST, each for 15 minutes, before ECL detection.

2.7.6.3 Enhanced chemiluminescence detection (NEL 103, NEN™)

Equal volume of Enhanced Luminol Reagent and Oxidizing reagent were mixed immediately before the detection and incubated with shaking for 1 minute. The membrane was then blot dried and placed in Saran wrap. Membrane was first exposed to Kodak BioMax (BIOMAX™ MR film) for 30 seconds and then exposed to the film for an optimized period according to the first exposure.

2.7.7 Immunofluorescence labeling of Leishmania

Cells were harvested at 0.5 X 10⁷/ml and re-suspended at 10⁸/ml in PBS. Methanol at -20°C was used to fix the cells for 15 minutes. Permeabilization was made by incubating the fixed cells in 50mM glycine, 2% Triton X-100 for 15 minutes. After
checking the cell morphology on inverted scope, the permeabilized cells and intact cells were blocked with 2% BSA in PBS for 30 minutes at room temperature. Incubation with Anti Flag antibody was at 1:25 (40μg/ml) in PBS for 1 hour at room temperature; 2\textsuperscript{nd} antibody incubation was at 1:200 for 1 hour at room temperature in the dark. After washing in PBS, 15 μl antiquench was added to the cover slip. Cover slip was sealed on a microscope slide by nail polish with cells facing downward.

2.8 \textit{Northern Blot Analysis for pLexSat Expression}

2.8.1 Hybridization probe

The DNA probes used in hybridization were 294 bp PCR fragments containing predicted A600 coding region. DNA fragments were labeled to high specific activity (10^9 cpm/μg) by using N9 Random Priming method.

2.8.2 Northern hybridization analysis

Total \textit{Leishmania} RNA (9 μg) was loaded in 1% agarose gel with 2.2% formaldehyde. Wild type promastigote RNA was used as the negative control and amastigote RNA (6 μg) as the positive control. For Northern hybridization analysis, RNA was fractionated by electrophoresis on 1% agarose gel (6.5 x 10 cm) in 2.2 M formaldehyde and 1x MOPS buffer (0.02 M MOPS pH7.0, 5 mM NaOAc, 1 mM EDTA pH 8.0 in DEPC treated dH\textsubscript{2}O) at 60 V for 2-3 hours (or until the bromophenol blue dye ran to 1 cm from the bottom of the gel). RNA was transferred to Nylon (Hi-Bond) membrane by capillary blotting in 20 x SSC (3M NaCl, 0.3 M sodium citrate, pH 8.3) and cross-linked to the membrane in the UV cross-linker (Ultra-Lum\textsuperscript{TM}) for 30 seconds.
Pre-hybridization and hybridization were carried out at 65°C in 5 x SSC, 5 x Denhardt's mix (0.1 % bovine serum albumin, 0.1% Ficoll, 0.1%polyvinylpyrolidone), 0.5% SDS and 50μg/ml denatured carrier DNA for 2 hours and overnight respectively. After hybridization, membrane was washed twice at 42°C for 5 minutes in low stringency buffer (2 X SSC, 0.1% SDS), followed by two washes at 65°C for 15 minutes in high stringency buffer (0.1 X SSC in 0.1% SDS). The membrane was drained on a 3 MM Watman paper and wrapped in Saran Wrap and exposed for autoradiography with Kodak film with intensifying screen at -70 °C, then exposed to Phosphor Imager screen.
The a600 gene was recently identified as an amastigote-specific gene using PCR based cDNA subtractive hybridization. Followed by virtual northern blot, our lab recently identified two cDNA fragments s600 and s800 that were highly abundant in amastigotes, thus considered as amastigote-specific genes (Belletin 1999). The extensions of these fragments toward the spliced leader resulted in two cDNA fragments of 900 bp and 1800 bp. The combination of the present study and other’s (Belletin 1999) data revealed that s600 and s800 were two non-contiguous fragments of the same gene flanking the Hae III site with the s600 fragment near the spliced leader. Thus, the extended fragment from s600 (900 bp) was part of the fragment extended from s800 (1800 bp). Another extension of the s800 fragment toward 3' poly A resulted in an 800bp fragment (not including poly A). Therefore, the size of the entire gene was 3 Kb from spliced leader to 3' poly A and the gene was named as a600 (Figure 3.1).

![Figure 3.1. Relative positions of s800, s600 and the 1800 bp fragments within the a600 transcript.](image-url)
3.1 DNA sequence determination of the 1.8Kb fragment of a600 gene

The nested deletion approach was used to sequence the 1.8Kb fragment, as the size of this fragment was too large for direct sequencing. In this approach, two restriction enzymes were chosen to generate 5' overhang end for ExoIII digestion and 3' overhang end that is protected from the ExoIII digestion. Along the time course of ExoIII digestion, nine aliquots were taken to stop the reaction at various time points in order to generate a set of unidirectional sub-fragments. When these fragments were re-ligated into the pGEM-T, the universal sequencing primer was used to sequence all the sub-fragments and their sequences were aligned to generate the contig.

3.1.1 Restriction digestion of the 1.8 Kb fragment

The restriction enzymes used to generate the desired ends must not cut the 1.8 Kb fragment so that the entire fragment would be sequenced in the following experiment. All the enzymes whose restriction sites are included in the multiple cloning site of pGEM-T were used to digest the pGEM-T/inserted 1.8 Kb fragment in order to examine whether they have the restriction site within the insert to be sequenced. Figure 3.2 illustrates the multiple cloning sites of pGEM-T:

5' → T7 promoter – Apa I – Aat I – Sph I – BstZ I – Nco I – Sac II ... 1.8Kb fragment ... Spe I – Not I – BstZ I – Pst I – Sal I – Nde I – SacI – BstX I – Nsi I – Sp6 promoter ← 3'

Figure 3.2. Multiple cloning sites of pGEM-T.
The restriction digestion revealed that Aat II, Pst I and Sac I had restriction sites within the insert, shown in Figure 3.3. Therefore these enzymes were not chosen for double digestion to generate the overhang ends for ExoIII digestion. Based on the sequence of the restriction site of other enzymes, Sph I and Nco I were chosen to digest the construct and generate 3’overhang protected end and 5’ overhang ExoIII digested end, respectively.

Both enzymes were used to digest a small aliquot of the plasmid DNA individually prior to the double digestion to confirm that they were capable of linearizing the plasmid completely. The 12 clones carrying the plasmid with right size deletions resulted from nested deletion were cultured for DNA isolation.

The plasmid DNA of the 12 clones were isolated and the vector was removed by double restriction digestion with Apa I and Nde I. Previous experiment proved that the inserted 1.8 Kb fragment does not contain the restriction sites of these two enzymes. The unidirectional sub-fragments were loaded into 1% agarose gel together with 100 bp DNA marker. As shown in Figure 3.4, 12 clones were arranged on the gel with the decreasing size order.
Figure 3.4. Subfragments generated by nested deletion. The deletions were removed from the pGEM-T by ApaI and NdeI digestion and loaded into 1% agarose gel with the size order shown in lane 1 to 4, 6 to 9 and 11 to 14, alongside with 100 bp DNA marker at lane 5, 10 and 15. Taking the original 1.8 Kb fragment (not included in this gel) into account, the maximum size difference between any two adjacent deletions is within ~350bp and therefore it was feasible to sequence all deletions and have the sequencing data integrated into one contig.
3.1.2 DNA sequences of the 1.8 kb fragment and the entire a600 gene

The DNA Sequence of each deletion (shown in Figure. 3.4) and the original non-deleted 1.8 Kb fragment was determined by sequencing reaction. Vector’s sequences were removed from the raw data. Sequences of the sub-fragments were aligned and integrated into the contig of 1.8 kb fragment with Sequencher™ (Version 4.0.5) (Figure 3.5). For the complementary strand, part of its sequence was coincidentally determined by Jaime Belletin in our lab while he was working on another fragment. The fragments a1, a2, a3 and a4 with arrows pointing to the right shown in Fig. 3.4 were to fill the gaps to determine the complete sequence of the complementary strand. The complete sequence of the 1.8 kb fragment is shown in Figure 3.6.

Part of the sequence of the 1.8 kb fragment was found to be 100% identical to that of the fragment extended from s600 independently done by Jaime Belletin in our group. This result revealed that s600 and s800 were two non-contiguous fragments of the same gene. The s600 fragment (900 bp) was part of the fragment extended from s800 (1800 bp) flanking the Hae III site with the s600 fragment near the spliced leader.

3.1.3 Sequence variation in the 3’ UTR of the a600 gene

Sequence variation was found at the 3’ UTR after 1542 bp (Figure 3.7) when the sequence of a4 was aligned with the sequence m13f (Figure 3.5). Primer a4 was designed based on the sequence of the fragment extended from s800 towards 3’ poly A, and the sequencing reaction was carried out using the same fragment as template. Thus the sequence a4 was from a different cDNA fragment. As the varied bases from the raw data occurred at unambiguous regions and at numerous positions, thus the variation is not likely caused by ambiguous readings. This sequence variation suggests that a4 and m13f
were originated from different cDNA clones, thus indicating that a600 is multiple copy gene family.
Contig[0001]
Sequencher(tm) "a600.spf"

Figure 3.5. Contig alignment for the 1.8 kb fragment
Figure 3.6. The cDNA sequence of A600 gene. The sequence of the open reading frame and the sequence of predicted polypeptide encoded are in bold.
Figure 3.7. Sequence alignment of m13f and a4 showing the sequence variation in 3'UTR of a600. Varied bases are indicated by asterisk. Sequence of a4 is shown in bold.
3.2 DNA sequencing analysis

Using BLASTN 2.0MP-WashU (Gish 2000), two regions of the a600 gene, 124 - 364 and 449 - 1096, were found from The EMBL Nucleotide Sequence Database to have high identities with assembled shotgun reads from chromosome 34 of *L. major* Friedlin, with the identities of 74% and 72%, respectively. One of the regions (124 - 364) covered part of the ORF of a600 (Figure 3.8A), while another region (449 - 1096) was located at the 3' untranslated region (Figure 3.8B). The recently released sequence from *L. major* Friedlin is preliminary, and the biological role of these sequence regions has not been defined yet.

Sequencing analysis of a600 gene indicated four possible coding regions (ORF) existing in the a600 gene. ORF 1 had 279bp, starting from base 105 down stream of the predicted spliced leader. ORF 2, 3 and 4 had 459bp, 507bp and 249bp, starting from base 521, 528 and 1127 down stream of splice leader, respectively.

The codon frequency of each potential open reading frame was compared with the *Leishmania* codon usage (Table 3.1). ORF 1 correlates with *Leishmania* codon usage, thus likely to encode a polypeptide. Also, ORF 1 is located 105 base downstream of putative spliced leader, which is typical in *Leishmania* genes (Matthews 1994). By these criteria, ORF 1 is predicted to be the coding region of a600 gene.

The complementary strand also has four potential coding regions. The ORF 1 had 585 bases and its location was from base 178 to 762; the ORF 2 had 414 bases from base 820 to 1233; the ORF 3 had 294 bases from base 1433 to 1726; and the ORF 4 had 252 bases from base 1443 to 1694. However, none of their correspondent RNA was found as amostagote-specific so far, thus they were not further analyzed in the present study.
Figure 3.8A Sequence identity between a600 and Leishmania major Friedlin assembled shotgun reads from chromosome 34 (EMBL AL499623, Length = 1,720,777). Identities = 181/243 (74%), Strand = Plus / Plus. The subject sequence shown is between 1653259 and 1653499.
Figure 3.8B. Sequence identity between a600 and Leishmania major Friedlin assembled shotgun reads from chromosome 34 (EMBL AL499623, Length = 1,720,777). Identities = 486/668 (72%), Strand = Plus / Plus. The subject sequence shown is between 1456431 and 11457082.
Table 3.1. Comparison of *Leishmania* codon frequency with the codon frequency of potential coding regions in a600 gene.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th><em>Leishmania</em> codon usage*</th>
<th>ORF 1</th>
<th>ORF 2</th>
<th>ORF 3</th>
<th>ORF 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine (A)</td>
<td>GCG</td>
<td>0.43</td>
<td>0.67</td>
<td>0.17</td>
<td>0.21</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>0.32</td>
<td>0.33</td>
<td>0.17</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>GCA</td>
<td>0.09</td>
<td>0</td>
<td>0.58</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>GCU</td>
<td>0.16</td>
<td>0</td>
<td>0.08</td>
<td>0.5</td>
<td>0.44</td>
</tr>
<tr>
<td>Cysteine (C)</td>
<td>UGC</td>
<td>0.86</td>
<td>0</td>
<td>0.61</td>
<td>0.17</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>UGU</td>
<td>0.14</td>
<td>0</td>
<td>0.39</td>
<td>0.83</td>
<td>0.67</td>
</tr>
<tr>
<td>Aspartic Acid (D)</td>
<td>GAC</td>
<td>0.80</td>
<td>0.33</td>
<td>0</td>
<td>0.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GAU</td>
<td>0.20</td>
<td>0.67</td>
<td>1</td>
<td>0.4</td>
<td>0</td>
</tr>
<tr>
<td>Glutamic Acid (E)</td>
<td>GAG</td>
<td>0.90</td>
<td>1</td>
<td>1</td>
<td>0.33</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>GAA</td>
<td>0.10</td>
<td>0</td>
<td>0</td>
<td>0.67</td>
<td>0.57</td>
</tr>
<tr>
<td>Phenylalanine (F)</td>
<td>UUC</td>
<td>0.83</td>
<td>0</td>
<td>0.26</td>
<td>0.35</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UUU</td>
<td>0.17</td>
<td>1</td>
<td>0.74</td>
<td>0.65</td>
<td>1</td>
</tr>
<tr>
<td>Glycine (G)</td>
<td>GGC</td>
<td>0.67</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>GGU</td>
<td>0.19</td>
<td>0.33</td>
<td>0.5</td>
<td>0.45</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>GGG</td>
<td>0.10</td>
<td>0.33</td>
<td>0</td>
<td>0.09</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GGA</td>
<td>0.04</td>
<td>0</td>
<td>0.5</td>
<td>0.45</td>
<td>0</td>
</tr>
<tr>
<td>Histidine (H)</td>
<td>CAC</td>
<td>0.88</td>
<td>0</td>
<td>1</td>
<td>0.43</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>CAU</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
<td>0.57</td>
<td>0.8</td>
</tr>
<tr>
<td>Isoleucine (I)</td>
<td>AUC</td>
<td>0.80</td>
<td>1</td>
<td>1</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td>0.18</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>Lysine (K)</td>
<td>AAG</td>
<td>0.94</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>AAA</td>
<td>0.06</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>Leucine (L)</td>
<td>CUG</td>
<td>0.57</td>
<td>0.46</td>
<td>0.25</td>
<td>0.13</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td>0.22</td>
<td>0.23</td>
<td>0.3</td>
<td>0.44</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>CUU</td>
<td>0.11</td>
<td>0.23</td>
<td>0.1</td>
<td>0.25</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>UUG</td>
<td>0.06</td>
<td>0.08</td>
<td>0.15</td>
<td>0.13</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>CUA</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>UUA</td>
<td>0.01</td>
<td>0</td>
<td>0.2</td>
<td>0.06</td>
<td>0</td>
</tr>
<tr>
<td>Methionine (M)</td>
<td>AUG</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asparagine (N)</td>
<td>AAC</td>
<td>0.92</td>
<td>0.6</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>AUA</td>
<td>0.08</td>
<td>0.4</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proline (P)</td>
<td>CCG</td>
<td>0.56</td>
<td>0.8</td>
<td>0.33</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CCC</td>
<td>0.20</td>
<td>0.2</td>
<td>0.17</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CCA</td>
<td>0.11</td>
<td>0</td>
<td>0.33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CUC</td>
<td>0.13</td>
<td>0</td>
<td>0.17</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>Glutamine (Q)</td>
<td>CAG</td>
<td>0.92</td>
<td>0</td>
<td>0.67</td>
<td>0.71</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>CAA</td>
<td>0.08</td>
<td>0</td>
<td>0.33</td>
<td>0.29</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 3.1. (continued).

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>Leishmania codon usage</th>
<th>ORF 1</th>
<th>ORF 2</th>
<th>ORF 3</th>
<th>ORF 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine (R)</td>
<td>CGC</td>
<td>0.62</td>
<td>0.38</td>
<td>0</td>
<td>0.33</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>CGG</td>
<td>0.12</td>
<td>0.13</td>
<td>0.38</td>
<td>0</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>CGU</td>
<td>0.12</td>
<td>0.13</td>
<td>0.08</td>
<td>0.20</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CGA</td>
<td>0.05</td>
<td>0</td>
<td>0.15</td>
<td>0.13</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>AGG</td>
<td>0.07</td>
<td>0.38</td>
<td>0.31</td>
<td>0.2</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>AGA</td>
<td>0.03</td>
<td>0</td>
<td>0.08</td>
<td>0.13</td>
<td>0.27</td>
</tr>
<tr>
<td>Serine (S)</td>
<td>AGC</td>
<td>0.28</td>
<td>0</td>
<td>0.16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UCG</td>
<td>0.30</td>
<td>0.33</td>
<td>0.16</td>
<td>0.16</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UCC</td>
<td>0.22</td>
<td>0.11</td>
<td>0.05</td>
<td>0.05</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>UCU</td>
<td>0.10</td>
<td>0.22</td>
<td>0.53</td>
<td>0.53</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AGU</td>
<td>0.05</td>
<td>0.11</td>
<td>0</td>
<td>0.26</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>UCA</td>
<td>0.04</td>
<td>0.22</td>
<td>0.10</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Threonine (T)</td>
<td>ACG</td>
<td>0.60</td>
<td>0.5</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>ACC</td>
<td>0.25</td>
<td>0.33</td>
<td>0</td>
<td>0.13</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>0.06</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>ACU</td>
<td>0.09</td>
<td>0.17</td>
<td>1</td>
<td>0.37</td>
<td>0.25</td>
</tr>
<tr>
<td>Valine (V)</td>
<td>GUG</td>
<td>0.62</td>
<td>0.67</td>
<td>0.45</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>GUC</td>
<td>0.24</td>
<td>0.33</td>
<td>0.18</td>
<td>0.27</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>GUU</td>
<td>0.09</td>
<td>0</td>
<td>0.27</td>
<td>0.09</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>GUA</td>
<td>0.05</td>
<td>0</td>
<td>0.09</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Tryptophan (W)</td>
<td>UGG</td>
<td>1.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tyrosine (Y)</td>
<td>UAC</td>
<td>0.93</td>
<td>0.5</td>
<td>0.33</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>UAU</td>
<td>0.07</td>
<td>0.5</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

* Data source: Langford et al. (1992).
The sequence of the predicted polypeptide coded by a600 gene consists of 93 amino acids (Figure 3.6) with a theoretical molecular mass of 10.5 kDa. The sequence of predicted A600 protein was not significantly similar to any currently available protein sequences in the SwissProt data base, thus suggesting that A600 a novel protein.

The putative A600 polypeptide was analyzed using SIGNALP V1.1 (Nielsen et al. 1997). A potential cleavage sites was predicted based on sequences from eukaryotes (Table 3.2 and Figure 3.9A). The cleavage site was located between position 56 and 57. However, a positively charged n-region, that is one of common motifs of signal sequence (Nielsen et al. 1997), is not present. No peptide cleavage site was predicted based on either sequences from gram-negative prokaryotes (Table 3.2 and Figure 3.9B) or sequences from gram-positive prokaryotes (Table 3.2 and Figure 3.9C).

The secondary structure of the putative polypeptide coded by a600 gene was analyzed by several programs. A possible transmembrane helix was predicted from position 27 to 48 by PhdTopology Refinement and Topology Prediction (Rost et al. 1996) (Figure 3.10). The N-terminus of A600 was predicted to be extra-cytoplasmic and the C-terminus to be intra-cytoplasmic. A potential transmembrane segment was also predicted using “DAS ” Transmembrane prediction server (Cserzo et al. 1997) from position 28-49 (Figure 3.11). These prediction methods gave almost identical predictions.
Table 3.2. Output of SIGNALP Prediction (Nielsen et al. 1997) for A600 protein based on sequences from eukaryotes, gram-negative prokaryotes and gram-positive prokaryotes.

<table>
<thead>
<tr>
<th>Organism</th>
<th>Measure</th>
<th>Position</th>
<th>Value</th>
<th>Cutoff</th>
<th>Conclusion*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukaryotes</td>
<td>Max. C-score</td>
<td>57</td>
<td>0.570</td>
<td>0.37</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Max. Y-score</td>
<td>57</td>
<td>0.635</td>
<td>0.34</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Max. S-score</td>
<td>38</td>
<td>0.985</td>
<td>0.88</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Mean S-score</td>
<td>1-56</td>
<td>0.536</td>
<td>0.48</td>
<td>Yes</td>
</tr>
<tr>
<td>Prokaryotes</td>
<td>Max. C-score</td>
<td>67</td>
<td>0.328</td>
<td>0.49</td>
<td>No</td>
</tr>
<tr>
<td>gram-</td>
<td>Max. Y-score</td>
<td>43</td>
<td>0.244</td>
<td>0.36</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Max. S-score</td>
<td>38</td>
<td>0.978</td>
<td>0.88</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Mean S-score</td>
<td>1-42</td>
<td>0.417</td>
<td>0.54</td>
<td>No</td>
</tr>
<tr>
<td>Prokaryotes</td>
<td>Max. C-score</td>
<td>43</td>
<td>0.304</td>
<td>0.42</td>
<td>No</td>
</tr>
<tr>
<td>gram+</td>
<td>Max. Y-score</td>
<td>43</td>
<td>0.305</td>
<td>0.34</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Max. S-score</td>
<td>41</td>
<td>0.972</td>
<td>0.95</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Mean S-score</td>
<td>1-42</td>
<td>0.472</td>
<td>0.55</td>
<td>No</td>
</tr>
</tbody>
</table>

*Note: The C and Y scores should be high (above the cutoff line) for the position immediately after the cleavage site (Nielsen et al. 1997). If the sequence is predicted to be a signal peptide, the position with the maximum Y-score is the most likely cleavage site.
Figure 3.9. Results of SIGNALP Prediction (Nielsen et al. 1997). One signal peptide cleavage site was predicted between position 56 and 57 using networks trained on sequences from eukaryotes (A). No peptide cleavage site was predicted either using networks trained on sequences from gram-negative prokaryotes (B) or using networks trained on sequences from gram-positive prokaryotes (C). For best prediction, 70 amino acid of the N-terminal part was analyzed for the prediction.
Figure 3.10. Results of the PhdTopology Refinement And Topology Prediction (Rost et al. 1996). A helical transmembrane region was predicted at the position 27 - 48. PHD htm = helical transmembrane (HTM) predicted by the PHD neural network (T: HTM region); PHDThtm = topology prediction based on refined model (i: intra-cytoplasmic, T: transmembrane region and o: extra-cytoplasmic).
Figure. 3.11. Transmembrane Prediction using "DAS"
3.3. Characterization of A600 protein

As there is no A600 homologue found in the current database, the biological function of this protein can not be deduced. Since the intracellular /intercellular location of a protein may provide some clue for its function, the first attempt to characterize A600 protein was to carry out the localization study.

3.3.1. Localization of A600 protein by Western blot analysis

The approach of the localization using Western analysis is shown in Figure 2.3. The sequence of a600 coding region with the incorporated Flag sequence amplified by PCR is shown in Figure 3.12. The sequence of the PCR fragment was confirmed and inserted into pGEM-T before being sub-cloned into Leishmania expression vector pLexSat (Joshi et al. 1995). The pLexSat with the insert and the pLexSat without the insert (negative control) were used to transfect wild type of promastigote of L. mexicana. The transfectants were subsequently analyzed by western blot using anti-Flag monoclonal antibody.

The results of Western analysis are shown in Figure 3.13. In this experiment, the blots that were incubated with anti-Flag antibody were then incubated with horseradish peroxidase conjugated secondary antibody (anti-mouse IgG) for detection. In the lanes that medium was affinity-purified (lane 4, 6, 11 and 13) with anti-Flag linked sepharose beads, two bands were detected at 50 kDa and 25 kDa. They are likely to be the heavy chain and light chain of the anti-Flag antibody used in the affinity purification prior to the electrophoresis. However, no expected unique A600 band (12.5 kDa) was detected either in the lysate (lane 3, 5, 10 and 12) or in the medium (lane 4, 6, 11 and 13) of transfectants with insert. The banding patterns of either the lysate or the medium did not shown any
difference between the transfected clones with a600 insert and the control without a600 insert. These results indicate that A600 protein tagged by Flag epitope was not detectable from either the cell lysate or the medium of the transfectant. Thus, these experiments did not reveal whether or not the A600 protein is a secreted protein or a membrane protein.

3.3.2. Localization of A600 protein by immunofluorescence labeling

Since the western analysis did not detect the fused a600 protein either in cell lysate or in cell-free medium, immunofluorescence labeling was applied as an alternative to localize the A600 protein. Microscopic images did not reveal any significant difference in fluorescence staining either in intact cells or in permeablized cells comparing with the control cells (data not shown).
Figure 3.12. PCR amplified *a600* coding sequence from both genomic DNA and 1.8 Kb fragment. Sequence determination confirmed that they resulted in the same sequence. Letters in bold indicate PCR primers. Restriction sites are indicated by arrows and italic font. The Flag sequence is underlined.
Figure 3.13. Western blot analysis for detection of A600 protein in transfected *L. mexicana*. Blot A was incubated with anti-Flag mAb as primary antibody. Blot B: the control blot (with the same loading order as blot A) was incubated with an irrelevant antibody as primary antibody instead of anti-Flag mAb. The secondary antibody (anti-mouse IgG) was horseradish peroxidase conjugated. lane 1, Cell lysate from control transfectant (no insert); lane 3, 5, 8, 10 and 12, cell lysate from transfectant promastigotes (with insert); lane 2, 4, 6, 9, 11 and 13, cell free medium of correspondent transfectant clones; Lane 7, molecular weight standard.
3.4. Confirmation of the transfection

Since the A600 protein was not detected, it became necessary to confirm whether the insert was successfully introduced into the wild type promastigote cells and producing mRNA.

As the presence of the pLexSat plasmid DNA would be sufficient to confirm the successful transfection, primers for exogenous SAT gene were used to perform PCR with DNA isolated from transfectant. SAT gene (550 bp) was detected in transfectant clones shown in Figure 3.14. This result proved the transfection and maintenance of plasmid DNA in the transfectants. Thus, it would be necessary and interesting to investigate whether the ORF of the a600 introduced by pLexSat was transcribed in the transfectant cells.

3.5. Detection of mRNA transcribed from insert a600 ORF by northern blot

RNA samples from all transfectant clones and wild type promastigote (the negative control) were quantified with spectrophotometer and 9 µg of each sample was electrophoresed and blotted as described in chapter 2. RNA gel was first stained with Ethidium Bromide (EtBr) for quantification (Figure 3.15B) prior to Northern transfer. The EtBr staining revealed that the amount of all RNA samples were about identical, except the amastigote RNA (lane 11, the positive control) was about 60% of wild type promastigote RNA. The probe was radioactively labeled from PCR amplified a600 coding region. The northern blot detected 1.6 kb RNA bands transcribed from the insert (the lower bands), both the size and pattern were as expected (Figure 3.15A). RNA bands were present in the clones transfected with the plasmid containing the insert (lane 2 –8) and absent in the clones transfected with the plasmid without the insert (lane 9 and 10,
control clones). The upper bands with size of 3.0 kilo base coincided with the size of a600 gene, thus were likely the RNA transcribed from a600 endogenous gene. Although only 5 µg amastigote RNA was loaded in the gel (lane 11), based on the intensities of these bands, the abundance of a600 mRNA in amastigote was significantly higher than that in wild type promastigote RNA (quantification by phosphorimager). This confirmed the previous results from our lab that a600 gene is preferentially expressed in amastigote. Interestingly, it was observed that the abundance of the RNA transcribed from endogenous a600 gene in all the transfectants was greatly enhanced in comparison with the wild type promastigote (negative control, lane 2). This enhancement was probably resulted from the presence of largely increased concentration of Nourseothrycin (i.e., 320 µg/ml vs. 32 µg/ml). The lane 6 showed much higher intensity (Figure 15 A) although the total RNA loaded in this lane was about the same as that in other lanes. There is no certain explanation for this observation. It might be possible that this clone was harvested at a particular stage when a600 gene was most preferably expressed.
Figure 3.14. Confirmation of the transfection by PCR with SAT gene primers. Lane 1, wild type BK1; lane 2 and 12, 100 bp ladder; lane 3-9, transfectants with the insert; lane 10 and 11, transfectants without the insert; lane 13, pLexSat DNA (positive control); lane 14 pGEM-T plasmid DNA (negative control); lane 15, water.
Figure 3.15. A) Northern blot of transfectant RNA with a600 coding region as probe. B) EtBr staining of the RNA gel. Lane 1, RNA ladder (EtBr stained); lane 2, wild type promastigote (negative control); lane 3 to 8, promastigote transfected with pLexSat containing a600 coding region; lane 9 and 10, promastigote transfected with pLexSat without a600 coding region; lane 11, amastigote (positive control).
CHAPTER IV DISCUSSION

4.1. The a600 gene encodes a novel protein

Results of DNA sequence comparison using BLASTN 2.0 (Gish 2000) indicate that there are two regions of the a600 gene (124 - 364 and 449 - 1096) found to have high identities (74% and 72%, respectively) with assembled shotgun reads from chromosome 34 of L. major Friedlin. However, the recently released sequence from L. major Friedlin by the Leishmania Genome Network is still very preliminary, and the biological role of these regions have not been defined yet. By the completion of the sequencing of the entire genome of this species, homologous genes will likely be found as a high degree of sequence conservation in Leishmania species has been observed for protein coding genes (Myler and Stuart 2000).

Sequence analyses indicate that the putative polypeptide coded by a600 gene is composed of 93 amino acids with a theoretical molecular mass of 10.5 kDa. The sequence of predicted poly-peptide did not show significant similarity to any currently available protein sequences in the data base, thus suggesting that a600 encodes a novel protein. Therefore, it is not possible to propose the role of the gene and gene product according to the available knowledge about the functions of the homologous genes in other species. Nevertheless, some predictions can be made by analyzing the sequence of putative a600 gene product.
4.2. The a600 gene is likely a multiple copy gene family

In the present study, sequence variation was found at the 3'-untranslated region (3'-UTR). This suggests that the a600 may be a multiple copy gene family. A previous study (Belletin 1999) indicated that the a600 gene was a single copy gene based on the results of southern blots using the s600 cDNA fragment, which is part of the a600 gene. However, the sequence of the entire a600 gene locates the s600 fragment (from 900 to 1500 bp) at the position before the sequence variation occurred (after 1542 bp, see Figure 3.5). Additionally, sequence variation may not necessarily result in size variation significant enough to be detected by southern blot. Thus, the southern blot results should not rule out the possibility that the a600 gene have multiple copies.

Multiple copy gene families are common in *Leishmania*, particularly for stage-specific genes, such as gp63 in *L. major* that has seven homologues tandemly arranged as a gene cluster on one chromosome (Voth et al. 1998). Other examples of stage-specific gene families include A2 gene family in *L. donovani* (Charest et al. 1996) and a newly identified gene family in the amastigotes of *L. infantum* encoding a homologue of amastin surface proteins (Wu et al, 2000). Stage-specific expression of these gene families is mediated through differential RNA stability and involves the 3'-UTRs of mRNA. There are also gene families whose expression in response to the change of environmental conditions is regulated through the variation in the sequence of the mRNA 3'-UTRs, such as gp63 in *L. chagasi* (Ramamoorthy et al. 1995), hsp70 in *L. infantum* (Quijada et al. 1997). The haploid genome of *L. chagasi* contains 18 or more genes encoding gp63 that can be distinguished by their unique 3'-UTRs (Roberts et al. 1993). These genes are clustered in tandem arrangement on a single chromosome (Ramamoorthy et al. 1995). They are differentially expressed in promastigotes during
growth in vitro: some of them are expressed predominantly in stationary phase promastigotes, while some of them are expressed predominantly in logarithmic phase promastigotes (Roberts et al. 1993). The major differences among these gp63 genes and their mRNA are the unique sequences and lengths of their 3'-UTRs and downstream intergenic regions (Ramamoorthy et al. 1992). Similarly, in *L. infantum* at least six copies of the hsp70 genes were found to be arranged in head-to-tail repetition units (Quijada et al. 1997). The expression of the genes 1-5 is affected by heat shock, while the gene 6 are not. The 3'-untranslated regions of genes 1-5 are highly divergent in sequence from the gene 6. In both cases, the variation of sequence at 3'-UTR is found to be associated with the regulation of gene expression by affecting maturation, stability and translation of mRNA and other processes (Ramamoorthy et al. 1995; Quijada et al. 1997).

In the present study, sequence variation was also observed at 3'-UTR of the a600 gene. It is possible that the a600 is a gene family rather than a single copy gene, and the variation in 3'-UTR sequence may be associated with the level of abundance in the two life stages. It will be important to confirm this by further investigation.

4.3. The putative A600 protein is likely a membrane protein

The SIGNALP Prediction (Nielsen et al. 1997) predicted one potential cleavage sites on putative gene product of a600 based on sequences from eukaryotes (Figure 3.6A and Table 3.2), indicating a potential signal peptide corresponding to amino acid 1 - 56. A typical signal sequence comprises three common structure motifs: 1) an N-terminal positively charged n-region; 2) a central uncharged h-region; 3) a more polar, but uncharged, c-region where the cleavage of signal sequence occurs (von Heijne and Abrahmsen 1989; Nielsen et al. 1997). Relative to the cleavage site, the residues at
positions -3 and -1 must be small and neutral for cleavage to occur correctly (von Heijne 1985). As *Leishmania* are eukaryotes, the prediction based on sequences from eukaryotes appears to be more appropriate. However, it is uncommon, about 2.2%, that a eukaryotic signal peptide is longer than 35 residues in SWISS-PROT database (Nielsen et al. 1997). On the other hand, signal peptides from *kinetoplastid* protozoan parasites are unique and some signal peptides from *Leishmania* are more similar to those from gram-positive prokaryotes that have long signal peptide comprising long n-region and h-region. Especially, the n-region of signal peptide in *Leishmania* has greater positive charge. For example, for gp63 signal peptides from several *Leishmania* species, the net charge of the n-region ranges from +4 to +7 (Button et al. 1993; Al-Qahtani et al. 1998). For A600, there is no positively charged n-region predicted. Nevertheless, these subdomains are functionally interactive (Rusch and Kendall 1995), a signal sequence lacking an n-region may still be competent if the length/hydrophobicity of h-region is increased. This appears to be the case for A600. However, the prediction based on sequences from gram-positive prokaryotes (Figure 3.6C and Table 3.2) did not find any cleavage site. Taking the results from both predictions into consideration, if there was a cleavage site, it would likely be between 56 and 57 as predicted based on eukaryotic database.

Signal peptides play critical roles in directing most secretory and membrane proteins across the endoplasm reticulum in the exocytic pathway (Rapoport et al. 1996). On the other hand, proteins comprising a signal peptide are likely to be secretory or membrane proteins. Thus, the predicted signal peptide in A600 protein suggests that A600 is either a secretory protein or a transmembrane protein.
The analysis of the secondary structure of the putative poly-peptide, using PhdTopology refinement and topology prediction (Rost et al. 1996), indicates that the protein coded from a600 gene has a possible transmembrane helix between position 27 to 48 with the N-terminus being extra-cytoplasmic and the C-terminus intra-cytoplasmic. Results of the prediction using “DAS” Transmembrane Prediction Server (Cserzo et al. 1997) provided almost identical prediction for the length and position of the transmembrane helix as PhdTopology refinement and topology prediction did. However, the cleavage site predicted (between position 56 and 57) is located at the C-terminal end of the putative transmembrane helix region. Thus, after the cleavage of the signal peptide, transmembrane helix region will not exit anymore.

According to the predictions based on amino acid sequence of the A600 protein, two possibilities exist. First, A600 is a secretary protein, and cleavage occurs during its trafficking through the exocytic pathway across endoplasm reticulum membrane. A 37-amino acid polypeptide may be secreted through the membrane by *Leishmania* amastigotes into the parasitophorous vacuole and would be likely involved in a host-parasite interaction. However, if this was the case, nearly two third of entire poly-peptide would have to be removed from the final product. This does not appear to be common. A signal peptide with 56 residues is also rare. Thus, it is doubtful whether the A600 has a signal peptide.

The second possibility is that A600 is a membrane bound protein. Cleavage does not occur and protein translocation toward the membrane does not require a signal sequence, as other pathways of protein translocation exist that bypass the ones involving
classical signal sequences (Rapoport et al. 1996). Unfortunately, the FLAG incorporated A600 protein was not detected either by western blot or by immunofluorescence labeling.

The predictions based on amino-acid sequence, however, are not conclusive. Genes or putative protein products with no similarities in the database pose a problem. At the same time they may represent specific phylogenetic and/or pathological adaptation that suggests possible targets for intervention. It is not uncommon that genes identified from Leishmania have no similarities in the database. Preliminary results from the Leishmania genome project (Myler and Stuart 2000) have shown that the majority (69%) of the genes identified from L. major Friedlin fall into the unclassified category in terms of their biological functions. Some of these genes encode putative proteins with sequence similarity to proteins with unknown functions in other organisms, or contain uninformative sequence motifs, but most of these genes encode proteins with no identified features or similarities. It is expected that there would be several thousand genes or coding regions discovered with potentially Leishmania-specific functions at the completion of the entire genome sequence (Ivens & Blackwell, 1999).

4.4 The FLAG incorporated A600 protein was not detected indicating the possible post-translational regulation of a600 gene expression

The ultimate location of the gene product can usually provide important indication of its biological functions. Experimentally, if the gene product was a secretary protein, it should be detectable from the medium of the transfectant. Otherwise the gene product should exist in the cell lysate of the transfectant. In the present study, western blot did not detect the A600 protein either from the cell lysate or from the medium of the
transfectant culture. Attempts to localize the protein using immunofluorescence labeling method were also not successful.

The presence of the plasmid DNA containing the a600 ORF in the transfectant was confirmed by PCR amplification of exogenous SAT gene. The a600 mRNA transcribed from the plasmid was also detected in the transfectant by Northern Blot. These results confirm that the a600 coding region was successfully introduced and transcribed in the *Leishmania* promastigotes. Hence the question is why the protein was not detected.

The a600 ORF was introduced using the *Leishmania* expression vector pLexSat. The pLEX series of vectors was developed and tested for the stable expression of exogenous genes in *Leishmania* (Joshi et al. 1995). These pUC-based constructs contain one of three independent selectable markers and a multiple cloning site inserted between the upstream and downstream untranslated regions of the *L. major* HEXBP gene, which is constitutively expressed in *L. major* (Webb and McMaster 1993). Giving the fact that the 3'-UTR is involved in post-transcriptional regulation of gene expression in *Leishmania* and *HEXBP* gene is constitutively expressed in Leishmania, therefore the inserted a600 coding region is supposed to be stably expressed in both stages.

Although the controls of stage-specific gene expression are mostly post-transcriptional in *Leishmania* (Aly et al. 1994; Beetham et al. 1997; Moore et al. 1996; Ramamoorthy et al. 1995; Wu et al. 2000), evidence for translational and/or post-translational control of stage-specific gene expression has also been reported in *Leishmania* and other related species. Bhaumik et al. (1991) found that translational control plays a major role in tubulin induction during *L. donovani* differentiation and
noted that the sequence micro-heterogeneity in the 5'-untranslated region influences the translational efficiency. In the present study, as the a600 coding region was cloned into pLEX vector using the 5'-UTR of HEXBP gene, therefore the influence of the 5'-untranslated region should not be a concern.

Translational and/or post-translational controls of developmental regulation were also found for the Nrk protein kinase in *Trypanosoma brucei* (Gale et al. 1994). In this case, no differences were found in the sequences at the 5' end of the gene. Tomas and Kelly (1996) reported translational and/or post-translational control of developmental regulation for the cruzipain in *Trypanosoma cruzi*. They found that although the mRNA level is the same throughout all stages of life cycle, the protein activity is 4-5 time higher in amastigote stage. However, the mechanisms of the translational and/or post-translational control of these genes are not yet clear. They could include glycosylation or stage-specific protein processing.

For the present study, it is likely that the down-regulation of the expression of a600 coding region occurred during the post-translational process in the promastigotes. One possibility is that the polypeptide coded from the a600 ORF was degraded after the translation due to its amastigote-specific nature and the unfavorable environments in the promastigotes, which resulted in the polypeptide not detectable. If this was the case, the regulation of a600 expression would be at two levels, i.e., post-transcriptional and post-translational.

There were four possible open reading frames in the a600 gene. The ORF 1 was determined as the protein coding region based on *Leishmania* codon preference and relative location to the spliced leader. In case the putative open reading frame was
mistaken although likelihood was small, the polypeptide encoded from this ORF would likely be recognized as exogenous polypeptide and be degraded by the defensive response of the transfectant.

4.5. Proposed experiments for future study

Since the a600 polypeptide was not detected either in the cell lysate or in the medium of the transfectant, alternative approaches need to be applied to determine the biological functions of the a600 genes in future experiments. Recombinant protein or synthetic polypeptide is needed to generate antibodies against A600 to detect the intracellular location of A600 protein in *L. mexicana* amastigotes.

Determining whether the a600 gene is a multiple copy gene family is important for further studying the mechanisms regulating the differential levels of a600 mRNA at different stages, and formulating strategies for the characterization of the gene and gene product. A straightforward experiment is to PCR amplify part of the heterogeneous 3'UTRs using the primers designed based on the sequencing results of the present study. Two different 5' primers are to be designed based on the region that sequence variation starts to occur; the oligo dT18 primer is to be used as 3' primer. The PCR products that are supposed to be about 1 Kb will be sequenced partially or completely by direct sequencing of PCR product or by cloning the PCR fragment into plasmid respectively.

Double targeted gene disruption (gene knock-out) by recombination (Cruz et al. 1991) is a powerful method for altering and testing gene function through generating defined mutants for complementation using shuttle vectors. The potential problem of using this method is that many *Leishmania* proteins are encoded by arrays of multiple genes like gp63 (Button et al. 1989), hsp70 (Quijada et al. 1997), and β-tubulin (Coulson...
et al. 1996). The a600 is also likely a multiple gene family indicated by the sequence variation at the 3'-UTR in the present study. However, this method was applied to characterize the A2-A2rel gene cluster in *L. donovani* (Zhang and Matlashewski 2001). The gene targeting constructs were developed based on the 5' and 3' DNA sequences flanking the A2-A2rel gene cluster. Thus, it is possible to replace a complete gene cluster located between a pair of the 5' and 3' flanking sequences with a targeting fragment constructed consisting of a NEO reporter. For the present study, therefore, the genomic organization of a600 gene(s) needs to be characterized before designing a strategy for the target gene(s) knock-out. If the a600 gene has multiple copies, but tandemly arranged on a single chromosome as most gene families do in *Leishmania*, gene knock-out approach can be promising. After the 600 gene(s) is knocked out, differences in phenotype between the mutant and the control will provide clues for the biological function of this gene.

As the a600 gene is amastigote-specific, the function of this gene should be highly associated with amastigote related functions such as parasite virulence, pathogenicity, intracellular survival and may be critical in term of the host-parasite relationship. Knock-out of this gene or gene family would ultimately aid in understanding its specific function and may improve the current knowledge about mechanisms of transmission between the two life stages. If the a600 gene is relevant to parasite virulence, especially if it is essential for amastigotes to grow in *vivo* but not in *vitro* (Cruz et al. 1991) like the A2 gene found in *L. donovani* (Zhang and Matlashewski 1997), the gene knock out approach would generate null mutant as attenuated strain and be applied in new vaccine development.
CONCLUSION

The amastigote-specific gene a600, identified from *L. mexicana* was sequenced. The putative polypeptide predicted by DNA sequence analysis consists of 93 amino acids. No significant similarity of the putative gene product was found in current protein database suggesting the A600 a novel protein. Predictions based on the amino acid sequence indicated that the putative polypeptide is likely a membrane protein as a transmembrane helix region is predicted. However, the putative polypeptide was not detectable in either the cell lysate or the medium of the promastigotes transfected with the construct comprising the ORF incorporating Flag sequence. PCR amplification of exogenous SAT gene confirms the successful introduction of a600 coding region into the *L. mexicana* promastigote. The RNA transcript of introduced a600 ORF was detected in the transfectant by northern blot. These results indicate that the polypeptide was degraded after being translated, thus suggest that the expression of a600 gene be also under post-translational regulation.

The sequence variation observed at the 3'-untranslated region suggests that the a600 gene has multiple copies, and the sequence variation at the 3'UTR might be associated with its amastigote-specific nature. Therefore, confirmation of the a600 being a multiple copy gene family and mapping the genomic organization of this gene family on chromosomes should be set as the first priority in the future study. Then, double targeted gene disruption by recombination would be an effective approach to determine and alter the biological function of this gene.
REFERENCES


