PROTEIN TRANSLOCATION ACROSS ENDOPLASMIC RETICULUM OF A TRYPANOSOME

by

BHARGAVI PATHAM

(Under the Direction of Kojo Mensa-Wilmot)

ABSTRACT

Despite being recognized as an important organelle in the secretory pathway, not many aspects of endoplasmic reticulum (ER) protein import are understood in *Trypanosoma brucei*, the causative agent of human African trypanosomiasis (HAT). Movement of proteins into the ER requires a signal sequence and a Sec61p translocation pore. We have developed a cell-free system using *T. brucei* microsomes (TbRM) and variant surface glycoprotein (VSG) as our model substrate. TbRM could import full-length proteins post-translationally in a parasite-cytosol dependent fashion. Purified Hsp70 (Ssa1p) could replace parasite cytosol in assisting protein translocation into TbRM, suggesting a role of molecular chaperones in protein translocation into ER of trypanosomes. Protein import into TbRM was signal peptide dependent. Further, model proteins from *Escherichia coli, Saccharomyces cerevisiae and Bos taurus* were not imported into TbRM. Conversely, many trypanosomatid proteins were not imported into canine microsomes. Since signal peptides are selectively recognized by specific translocation machineries, we hypothesize that signal peptides have 'co-evolved" with their translocons.

We used the *in vitro* system to identify protein translocation blockers (PTBs), compounds that inhibit parasite protein import into TbRM. PTBs (MAL3-101, equisetin and CJ-21, 058) were trypanocidal, with IC_{50} (the concentration at which fifty percent of parasites are killed) of; 125 nM (MAL3-101), 3.3 μM (equisetin) and 7 μM (CJ-21, 058). PTBs did not affect viability of a model mammalian cell. Therefore, we have i) identified ER protein translocation machinery, as a drug target in *T. brucei*, and ii) discovered trypanocidal compounds that may act on import of proteins into the parasite ER.

Signal sequences have three conserved regions; an amino-terminal (n-region), a hydrophobic (h-region), and a carboxy-terminal (c-region). Although the hydrophobic core is important for efficient import/export of proteins, functional characteristics of h-regions are poorly understood. A bioinformatics analysis led to the discovery of highly conserved peptide motifs in h-region of signal sequences in eukaryotes (*h. sapiens*, *S. cerevisiae*, *T. brucei*), and prokaryotes (*E. coli, B. subtilis*). Peptide motifs in h-regions may significantly contribute to signal sequence activity. Finally, amino acids in h-region of signal sequences are not randomly selected. On the contrary, selection and distribution of amino acids in h-regions is highly specific to, and varies across species.

INDEX WORDS: Endoplasmic Reticulum, Human African Trypanosomiasis, Protein translocation blockers, *Trypanosoma brucei, T. brucei* microsomes, Variant surface glycoprotein

PROTEIN TRANSLOCATION ACROSS ENDOPLASMIC RETICULUM OF A TRYPANOSOME

by

BHARGAVI PATHAM

M.B.B.S, Nagpur University, India, 1996

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005

PROTEIN TRANSLOCATION ACROSS ENDOPLASMIC RETICULUM OF A TRYPANOSOME

by

BHARGAVI PATHAM

Major Professor:	Kojo Mensa Wilmot
Committee Members:	Debra Mohnen
	James Lauderdale
	Marcus Fechheimer
	Ruth Furukawa

Electronic Version Approved

Maureen Grasso

Dean of the Graduate School

The University of Georgia

August 2005

DEDICATION

This dissertation is dedicated to my parents,

Mrs. Bharathi Patham and Mr. M.S.K. Patham

for their undaunted and selfless love.

ACKNOWLEDGEMENTS

To Dr. Kojo Mensa-Wilmot, for being my source of inspiration for becoming a researcher.

To my committee members Dr. Furukawa, Dr. Kipreos, Dr. Mohnen, Dr. Lauderdale, and Dr. Striepen for their valuable inputs, constant support, and tremendous encouragement.

To Dr. Fechheimer, for setting high standards, and for trusting in my abilities more than I did.

To Dr. Wenzel, for always encouraging me to become a better teacher.

To my collaborators: Dr. Jeffrey L. Brodsky, Peter Wipf, Richard C. Davis, and Dr. Marcus Fechheimer.

To my foster family: Mariam George, Bagie George, Cigie George, and Agie George for providing home away from home.

To my friends Nandita Mullapudi, Danielle Webster, Engda Hagos, Paul Griffin, Julie Stanton, Sunayana Karumbaiah, Vikas Dhingra, Sudesh Kumar, Meera Mehta and Krupal Mehta, for all the lovely moments we shared.

To the members of my lab: Frank Hardin, Dr. Zhifeng Zheng, and Joshua Duffy for all the wonderful times we spent together.

To the staff of cell biology; Marian Thomas, Genia King, Kathy Vinson, Brett Rudolf, Andrea Leach, Allen Emory and Jeff Butt for the all their assistance throughout my graduate school.

To members of Dr. Dougan's and Dr. Tarleton's lab for all their help.

To Chaya Mukherjee, my biology teacher, who instilled the love of biology in me.

To my parents, who sacrificed their present, to give me a better future.

To my brother Dr. Bhaskar Patham, for being my pillar of strength, and for always encouraging me in my endeavors.

Finally, to my husband, Dr. Sandesh Subramanya, without whom I am incomplete, without whom I cease to exist, and without whom my life will lose all its meaning. I count my blessings every day for his presence in my world.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1. *Trypanosoma brucei*

1.1: Introduction

Trypanosoma brucei is a vector borne parasite, which is responsible for causing human African trypanosomiasis (HAT) in humans and Nagana in cattles [1]. *T. brucei* are transmitted by the tsetse fly *(Glossina spp.*) [1]. Two sub-species of *Trypanosoma brucei* cause different forms of HAT. *T. b. gambiense*, found in central, west, and some parts of east Africa, causes chronic infection. *T. b. rhodesiense*, found in southern and east Africa causes acute infection [1, 2].

1.2: Life Cycle

The life cycle of *T. brucei* alternates between the human host and tsetse fly. *T. brucei* enters the blood-stream of human hosts *via* the bite of tsetse flies. At the site of the bite, a chancre is formed. The parasites propagate *via* blood vessels and lymphatics from the site of infection, leading to generalized parasitemia. They invade various organs, and eventually cross the blood brain barrier to invade the brain [3, 4]. Cattle and other wild mammals act as reservoir hosts [5]. Tsetse flies acquire parasites by feeding on reservoir hosts, or human host. In the human host, the parasites (blood stream form) occur in various stages of differentiation; long slender (dividing), short stumpy (non-dividing), and intermediate forms [6]. Upon ingestion by the tsetse fly, the short stumpy form develops into a procyclic form that migrates to salivary glands and develops into an epimastigote form. The infective stage in *T. brucei* is the metacyclic form which develops from epimastigotes [5].

1.3: Antigenic Variation

T. brucei have evolved a higbhly effective defense mechanism to evade the host immune response. Called “antigenic variation” this phenomenon is brought about by variant surface glycoproteins (VSG), which are major surface proteins of the bloodstream form of the parasite. Approximately 10^7 variant surface glycoprotein (VSG) molecules are present on the *T. brucei* surface [7, 8]. A small sub-population of these parasites expresses an antigenically distinct VSG, thereby escaping the host immune

response. This small fraction of parasites then divide, making the destruction of entire population of parasites difficult for the host immune system [9-11].

2. Human African Trypanosomiasis (HAT)

2.1: Introduction and Epidemiology

Human African Trypanosomiasis (HAT) is a tropical vector-borne parasitic disease caused by *Trypanosoma brucei* that are transmitted to humans by the bite of tsetse flies (*glossina*) [12]. According to World Health Organization (WHO), HAT is an emerging epidemic, affecting 36 countries in sub-Saharan Africa [13]. The disease is endemic in 23 countries (including Angola, the Democratic Republic of Congo (DRC), Sudan and Central African Republic). HAT threatens an estimated 60 million people annually [13]. Approximately 300,000-500,000 people are currently infected with the parasite although only 45, 000 cases are reported each year [13].

2.2: Disease Manifestations

During its first stage (hemolymphatic stage, where the parasites travel in the blood and lymphatic system of the host), HAT is marked by fever, swollen lymph nodes, itching, and rash. Other symptoms include malaise, headache, increased heart rate, weight loss and swelling) [3, 4]. Stage II (central nervous system stage) disease starts with breach of the blood brain barrier by the parasites, invasion of central nervous system (CNS), and abnormalities in cerebrospinal fluid (CSF) [3, 4].

The clinical manifestations of stage II include: confusion, poor co-ordination, and disturbance in sleep cycle (hence the term “sleeping sickness”). Other symptoms include listlessness, loss of spontaneous speech, and involuntary movements in the extremities. Death results due to severe neurological impairment [3, 4]. CNS involvement is prominent with *T. b. gambiense* infection. *T. b. rhodesiense* infection takes an acute route, with symptoms appearing in quick succession, often making stage distinction unclear.

2.3: Treatment of Human African Trypanosomiasis

The WHO currently approves only four drugs for treatment of HAT [14-16] (Table 1). The current approach to HAT treatment is stage specific, and is based on involvement of central nervous system (CNS), making accurate diagnosis critical for disease management [2]. *T. b. rhodesiense* poses a greater challenge to treatment due to the rapid course of progression of the disease. Death could occur in few weeks after infection. On the other hand, *T. b. gambiense* infection is chronic, taking several months for progression of disease. Stage II treatment is more difficult, and has numerous limitations. The drugs that are currently recommended by WHO are listed in Table 1 [14-16].

3. Drug Discovery

3.1: Need for New Drugs against *T. brucei*

Currently approved drugs for HAT have severe limitations due to toxicity and development of therapeutic resistance [16, 17]. Administration of these drugs is often difficult, and course of treatment is long (Table 1). Patients may encounter severe, and sometimes life threatening complications involving the cardiovascular, renal and central nervous systems (Table 1) [16]. Secondary prevention in the form of vaccination against *T. brucei* does not currently exist [2]. Many studies have been undertaken for development of vaccine but are not encouraging this far, due to antigenic variation exhibited by these parasites [18]. Inexpensive drugs, which are (i) easy to administer, (ii) less toxic to patients and (iii) highly effective, are needed to control the disease [19].

3.2: Newer Drugs under Consideration

Two drugs have recently entered phase IIA/IIB trials. They are DB289 [2,5-bis(4-amidinophenyl)furan-bis-O-methylamidoxime] [20], and megazol [2-amino-5-(1-methyl-5-nitro-2-imidazolyl)-1,3,4-thiadiazole] [2]. These compounds were chosen for clinical trials due to (i) high bioavailability, (ii) easy administration (oral), and (iii) tolerance amongst patients.

4. Protein Translocation into Endoplasmic Reticulum

The classical secretory pathway in eukaryotes consists of endoplasmic reticulum (ER), Golgi complex (cis, medial, trans), lysosomes, endosomes, scretory vesicles and plasma membrane [21]. Regardless of their final destination, ER serves as the gateway of many of secretory proteins [22-24]. Additionally, ER serves as an organelle for multitude of functions, including signal peptide cleavage, N-glycosylation and protein folding [25-27].

Protein translocation into the ER can occur co-translationally [28, 29] (please refer to section 6 for a detailed description), or post-translationally [30] (Please refer to section 7 for detailed description). During co-translational translocation, polypeptides are targeted to the ER prior to completion of its translation and release from the ribosome [28, 29]. On the other hand, in post-translational translocation, the nascent protein is released from the ribosome, kept in an import-competent state by cytosolic chaperones, which then targets the polypeptide to the membrane [30].

In either case, protein translocation across membranes occur in three broad steps: i) Recognition of amino-terminal signal peptide of the polypeptide, ii) interaction of signal peptide with translocation pore (translocon) on the membrane, and iii) actual translocation of polypeptide across the translocon.

5. Signal Peptides

5.1: Introduction

Signal peptides are amino-terminal targeting sequences that direct protein import into the endoplasmic reticulum (ER) of eukaryotes (e.g., humans, *Saccharomyces cerevisiae*) [25, 26]. In Gram-negative bacteria (e.g. *Escherichia. coli*), they direct export of proteins from inner bacterial membrane to the periplasmic space [31]. In Gram-positive bacteria (e.g. *Bacillus. subtilis*), they direct export of proteins into the external milieu [31].

5.2: Signal Peptide Recognition

Signal peptide recognition is the first step in protein translocation in eukaryotes and in prokaryotes. Signal peptides are recognized by (i) SRP (co-translational ER import) [30, 32], (ii) Ssa1p (post-translational ER import) or, (iii) SecA (*E. coli* post-translational targeting).

5.3: Properties of Signal Sequences

Signal peptides lack conserved sequences, but exhibit a general tri-partite organization consisting of: (i) a positively charged amino terminal n-region, (ii) a hydrophobic core, and (iii) a polar C-terminal region [33-35]. Species-specific variations in signal sequences have been documented, but the significance of these variations has still not been elucidated. Differences exist across species in terms of the length, charge, and amino acid composition of each sub-region (n, h, c) [36].

Eukaryotic signal sequences are ~18-20 amino acids long. The n-region of eukaryotic signal sequences is short (~1-5 amino acids), and carries a charge of +2. The h-region of eukaryotic signal sequences are ~ 7-15 amino acids long, are highly hydrophobic, and have higher leucine content than prokaryotes (reviewed in [36-38]). The c-region is ~ 5 amino acids long and follow "-3 to -1 rule" (recognition site for signal peptidase) for cleavage (A-x-A).

E. coli signal sequences are ~ 22-25 amino acids long [39]. The n-region is ~ 7 residues long and carries a net charge of +2 [39]. The h-region and c-region contain ~ 12 and 6 amino acids respectively [36, 37]. The c region (~ 6 amino acids) follows "-3, -1 rule". *B. subtilis* signal peptides are longer than those in *E. coli* (reviewed in [36, 37]): n-region (of 7-12 amino acids) carries a net charge of +2 to +7 (reviewed in [36-38]). *B. subtillis* h-region is ~ 15 amino acids in length. *B. subtillis* also follows "-3, -1 rule" for the c-region. *B. subtillis* c-region are longer (~8 amino acids).

Very little primary sequence homology is seen in signal peptides across biological systems. Even within the same biological system, search of signal sequence homology has been futile (reviewed in [36-38]). However, the biological significance of the observed species-specific differences in signal sequences have not been appreciated, probably because signal sequences are sometimes interchangeable across species. Signal peptides have been documented to drive protein import into the ER of other species [40]. For example, α-mating factor from *S. cerevisiae,* and β-lactamase from *E. coli* are imported into canine microsomes [40].

Recent studies have shown that signal sequences may not always be universally interchangeable. For example, proteins with *Leishmania* signal sequence do not drive import into ER when transfected in insect Sf9 cells [41]. Similarly a *T. cruzi* signal sequence could not direct protein import in the ER of murine *Vero* cells [42]. We observed that a group of trypanosomatid signal sequences do not direct import into canine microsomes [43]. Hence, signal peptides seem to be specifically recognized by translocon.

5. 4: Role of Hydrophobicity in Signal Peptide Function

The hydrophobic core of the signal sequences is considered important for signal peptide function [44]. The choice of either a co-translational or post-translational pathway for protein import into the ER depends on the hydrophobicity of the signal peptide [44]. Peak hydrophobicity of the h-region has been used successfully to identify signal sequences from many organisms. Kyte-Doolittle hydopathy plots have been used to predict hydrophobicity of signal peptides [45]. In general, higher hydrophobicity of signal peptide favors co-translational translocation. Less hydrophobic signal peptides utilize post-translational translocation [44].

6. Co-translational Protein Translocation into ER of Eukaryotes

Co-translational ER protein import occurs in mammalian and yeast cells. The pathway utilizes: i) a signal peptide; ii) signal recognition particle; and iii) a translocation pore in the ER membrane [28, 46]. Factors involved in mammalian and *S. cerevisiae* ER translocation are listed in Table 2.

6.1: Signal Recognition Particle

Co-translational targeting of protein into ER is signal recognition particle (SRP) dependent (reviewed in [28]). SRP was discovered in mammalian cells [47, 48]. The mammalian SRP components are 9p, 14p, 19p, 54p, 68p, 72p and a 7SL RNA [28, 49]. The yeast SRP components (Srp68p, Srp72p, Srp65p, Srp54p, Srp21p, Srp14p, and scR1 RNA) are homologous to their mammalian counterparts (Table 2) [30, 32, 50]. M-domain of SRP54p recognizes and binds the signal peptide of a nascent polypeptide, as it emerges from the ribosome [51]. Apart from the signal peptide, SRP interacts directly with the ribosome [51], causing "transient delay" in translocation (elongation arrest) [52-54]. The “Alu” domain (SRP9 and SRP14 and 7SL RNA) is required for elongation arrest [54]. Finally, SRP interacts with SRP receptor (SR) located on the ER membrane.

6.2: SRP Receptor

The polypeptide emerging out of the ribosome (ribosome nascent chain or RNC) is recognized by SRP, and targeted to the ER membrane *via* interaction of the SRP and SRP receptor (SR). SR is an ER membrane bound heterodimeric receptor of SRP [55]. In mammalian cells, SR consists of the 69 kDa protein SRa [56] and the 30 kDa SRβ. SRα (a GTPase) is a peripheral membrane protein that associates with integral membrane protein SRb. The GTPase activity of SRa is required for the stable formation of the SR complex [57]. The yeast SRα /Src101p [50] and SRβ /Src102p [50] are homologous to the mammalian SR components (Table 2). Src101p is not essential for cell survival [50]. Src102p binds to Src101p, which anchors the complex to ER membrane. Interaction of targeting complex (RNC-SRP) with

the SR at the ER membrane is GTP dependent. Upon GTP binding, transfer of the signal peptide from SRP54 to the Sec61a ensues. Resulting GTP hydrolysis triggers dissociation of SRP-SR complex and the polypeptide synthesis resumes [58, 59].

6.3: Translocon

The Sec61 complex forms the core of the protein translocation channel, i.e. translocon [60]. The mammalian Sec61 complex consists of Sec61α, Sec61β, and Sec61γ [61]. *S. cerevisiae* homologs of Sec61 complex are: Sec61p [62], Sbh1p [63] and Sss1p [64] respectively. The ribosome binds to Sec61p tightly, allowing the polypeptide to enter the translocon [65, 66]. If there is successful interaction between signal peptide and Sec61p, the translocon is "gated" and the polypeptide enters the ER lumen [66]. After elongation arrest, protein synthesis resumes [67].

6.4: Steps During Co-translational Protein Import into ER

Co-translational translocation of polypeptides into ER occurs in four stages. First, the polypeptide emerges from the ribosome, its amino terminal signal sequence is recognized by SRP [51]. Second, the ribosome nascent chain complex is targeted to the ER. The targeting is mediated by interaction between i) SRP and SRP receptor [58, 59], and ii) Ribosome and Sec61α [66]. Third, the signal sequence is then transferred to the translocon [65, 66]. Finally, successful recognition of signal sequence by Sec61α results in "gating of translocation pore" and transfer of polypeptide to the ER lumen ensues [65, 66].

7. Post-translational Protein Import into ER of *S.cerevisiae*

An important basis for distinction between the co-translational and post-translational pathway is the SRP utilization. Existence of a SRP independent pathway was first discovered in yeast [30]. The co-translational pathway is SRP-dependent, while the post-translational pathway is SRP-independent. In

yeast, SRP is not essential for survival in yeast [30, 50] clearly indicating the existence of both pathways. Major factors of post-translational translocation in *S.cerevisiae* are listed in Table 3.

7.1: Translocon

Posttranslational translocation of secretory proteins occurs *via* a heptameric complex consisting of the Sec61 and Sec62/63 sub-complexes [68, 69]. The yeast Sec61 complex components are Sec61p, Sbh1p and Sss1p [64]. Sec62/63 sub complex consists of Sec62p, Sec63p, Sec71p and Sec72p [62, 70] . Sec61 complex serves as the protein conducting channel for both co-translational and post-translational protein import into yeast ER (Reviewed in [71]).

Components of the Sec62/63 sub complex are Sec62p, Sec63p, Sec71p and Sec72p. While both Sec62p and Sec63p are essential (reviewed in [72]), Sec71p and Sec72p are not essential for cell growth [70, 73] in yeast. Sec62/63 sub-complex is extensively involved in post-translational translocation. The Sec61complex alone is sufficient for co-translational translocation. The Sec61complex along with the Sec62/63 sub-complex forms the heptameric complex [62, 70], which is needed for post-translational translocation.

The signal sequence binds to the Sec61p component of the translocon in a BiP, and ATP independent fashion. This initial binding directs the polypeptide to the translocation channel. Signal sequence also binds to the Sec62p component of Sec62/63 sub-complex in one single signal recognition step [74].

7.2: Cytosolic Chaperones

Cytsolic molecular chaperones are responsible for keeping the protein in an “import competent” properly folded state [75]. Luminal chaperones are responsible for assisting import of the proteins into the ER. The molecular chaperones (both cytosolic and luminal) belong to the 70 kDa class of heat shock

proteins (Hsp70) and are essentially ATPases [76]. The yeast cytosolic Hsp70 include Ssa1-4 set of proteins [77]. Hsp70 chaperones (e.g. Ssa1p) and their Hsp40 co-chaperones (e.g. Ydj1p) play an important role in protein translocation into the yeast ER. The Hsp40 class of proteins regulate ATPase activity of Hsp70 chaperones [78]. In addition, Kar2p (ER luminal ATPase) is also essential for posttranslational translocation [63].

7.3: ER Luminal Chaperones

The luminal Hsp70 class chaperone is Kar2p (yeast)/ BiP (mammalian homolog of Kar2p) [48, 79], and its homolog namely Lhs1p. kar2p/BiP plays an important role in aiding the translocation of the preprotein into ER lumen. When the peptide emerges at the luminal side of the translocon, several BiP molecules bind to the translocating polypeptide. As a result, the peptide can no longer diffuse backwards in the channel, eventually resulting in completed translocation. The binding of BiP requires an interaction between BiP and the luminal J-domain of Sec63p [80, 81]. The Sec63p J-domain binds to ATP-bound BiP, and stimulates ATP hydrolysis resulting in ADP-bound BiP. This enables BiP to bind to the translocating polypeptide with higher affinity. BiP binds to mature part of the protein emerging from the translocon [82]. The Sec63p - BiP interaction is transient, and the J-domain- induced ATP hydrolysis does not require a peptide [82]. It is suggested that apart from acting as a ratchet, BiP may also actively pull the peptide [82]. BiP molecules dissociate from the translocated peptide upon nucleotide exchange thus liberating the polypeptide to the ER lumen, and rendering BiP ready for a new cycle of J-domain activation [82].

8. Protein Translocation into the ER in *T. brucei*

8.1: Secretory Pathway in Trypanosomatids

Despite understanding the significance of the secretory pathway, not much is known about ER protein translocation in trypanosomatids. Markers have revealed the presence of nuclear ER and cortical ER (which is the continuation of nuclear ER) [83] . Many markers for *T. brucei* ER, our orgenelle of

interest, have been identified (e.g., Bip and calreticulin) [84, 85]. The ER can account for as much as 60% of internal membranes in dividing populations of parasites [86]. Proteins synthesized in rough endoplasmic reticulum are transported to the Golgi network *via* transitional ER [87]. Proteins are exported from the Golgi to the flagellar pocket through vesicular, and cisternal transport [88]. Lysosomal proteins are exported either directly from Golgi, or indirectly after being routed from flagellar pocket *via* endosomes [89, 90].

Cell surface proteins are critical for parasite survival in trypanosomatids (*T. brucei, T. cruzi and Leishmania* spp.). Some important proteins that are transported to their destination by secretory pathway in *T. brucei* include: GPI anchored proteins like variant surface glycoprotein (VSG) [10], transferrin receptors, [91, 92], and GP63 [93]. Many *Leishmania* spp. proteins that are targeted across ER include: GPI anchored proteins (e.g., GP63 [94] PSA-1 [95], and PPG [96]); membrane proteins (e.g., amastin [97], 3' and 5' nucleotidase [98]), and secreted proteins (e.g., acid phosphatase [99]) reach the surface *via* ER. *T. cruzi* surface proteins include: (e.g., mucins [100], amastin, and trans sialidase family of proteins [101]), and membrane proteins (e.g., like Gp72 [102]).

8.2: Protein Translocation in *T. brucei*

Trypanosome signal sequences maintain a tripartite structure. Signal sequences in *T. brucei* have three sub-regions; an amino-terminal (n-region), a hydrophobic core (h-region) and a carboxy-terminal region (c-region) [43]. Little is known about ER protein import in *T. brucei*. Recently, the effect of RNA interference (RNAi) of SRP54 was studied in *T. brucei*. Under Srp54 depletion, all the proteins studied were translocated into the ER lumen and properly processed [103]. It was suggested that an “alternative pathway” (i.e, SRP-independent) pathway exists for protein translocation into *T. brucei* ER [103]. From these studies, one would surmise that *T. brucei* may utilize post-translational translocation for protein import into the ER.

Signal peptides from one species may function with the translocation machinery of another. For example, prepro α-factor (α-MF) from the yeast *S. cerevisiae* and, β-lactamase (β-lac) from *E. coli* can be imported into canine microsomes [40]. Surprisingly, several Trypanosomatid proteins from *Leishmania* (e.g., GP63) and *T. brucei* (e.g., VSG117, MVAT7, BiP) cannot lead protein import into canine microsomes.

During the course of this study, we plan to understand the mechanism of ER protein translocation in *T. brucei*. We also plan to elucidate the function of the signal sequence in protein import into *T. brucei* ER. We also wish to have a better understanding of the contributions of the n and h sub-regions of the *T. brucei* signal peptide to its activity.

Table 1: Current Drug Treatment for Human African Trypanosomiasis

Drug	Species	Indication	Mechanism of Action	Limitations
Pentamidine	***T.b. gambiense***	**Stage I**	**Inhibits transporter mediated adenosine uptake**	**Hypoglycemia, hypotension**
Suramin	***T.b. gambiense, T.b. rhodesiense***	**Stage I**	**Inhibits receptor mediated LDL uptake. Inhibits glycolytic enzymes**	**Anaphylactic shock, renal failure**
Melarsoprol	***Tb gambiense, T.b. rhodesiense***	**Stage II**	**Targets thiol containing enzymes of *T. brucei***	**Reactive encephalopatic syndrome, Increasing treatment failure**
Eflornitine	***T.b. gambiense***	**Stage II**	**Irreversible inhibitor of ornithine decarboxylase**	**Pancytopenia, convulsions**
Nifurtimox	***T.b. gambiense***	**Stage II**	**Generation of free radicals which are detrimental to parasite**	**Not registered for HAT yet**

Table 2: Factors Assisting Co-translational Import in Vertebrates and their *S. cerevisiae* homologs

Factors	Mammalian Components	*S. cerevisiae* Homologs
Translocon	**Sec61p (α, β, γ)**	**Sec61p, Sbh1p, Sss1p**
SRP	**9p, 14p, 19p, 54p, 68p, 72p**	**68p, 72p, 65p, 54p, 21p, 14p**
SRP RNA	**7SL RNA**	**scR1 RNA**
SRP receptor	**SRa, SRb**	**Src101, Src102**

Table 3: Factors Assisting Post-translational Import in *S. cerevisiae*

Factors	*S. cerevisiae*
Translocon	**Sec61p, Sbh1p, Sss1p**
Other membrane proteins	**68, 72, 65, 54, 21, 14**
Cytosolic chaperones	**Ssa1-4p**
Cytosolic co-chaperones	**Ydj1p**
Luminal chaperones	**Kar2p**

REFERENCES

1. Kuzoe, F.A.S., *Current situation of African trypanosomiasis.* Acta Trop.(Basel). 1993. **54**: p. 153-162.
2. Barrett, M. and H. Densie, *Uptake and mode of action of drugs used against sleeping sickness.* Biochemical pharmacology, 2001. **61**: p. 1-5.
3. Chimelli, L. and F. Scaravilli, *Trypanosomiasis.* Brain Pathol, 1997. **7**(1): p. 599-611.
4. Chappuis, F., et al., *Options for field diagnosis of human african trypanosomiasis.* Clin Microbiol Rev, 2005. **18**(1): p. 133-46.
5. Breidbach, T., E. Ngazoa, and D. Steverding, *Trypanosoma brucei: in vitro slender-to-stumpy differentiation of culture-adapted, monomorphic bloodstream forms.* Exp Parasitol, 2002. **101**(4): p. 223-30.
6. Cox, F.E., *History of sleeping sickness (African trypanosomiasis).* Infect Dis Clin North Am, 2004. **18**(2): p. 231-45.
7. Vickerman, K., *On the surface coat and flagellar adhesion in trypanosomes.* J.Cell.Sci., 1969. **5**: p. 163-193.
8. Gruszynski, A.E., et al., *Surface coat remodeling during differentiation of trypanosoma brucei.* J Biol Chem, 2003.
9. Cross, G.A.M., *Cellular and genetic aspects of antigenic variation in trypanosomes.* Ann.Rev.Immunol., 1990. **8**: p. 83-110.
10. Pays, E. and D.P. Nolan, *Expression and function of surface proteins in Trypanosoma brucei.* Mol.Biochem.Parasitol., 1998. **91**: p. 3-36.
11. Cross, G.A.M., L.E. Wirtz, and M. Navarro, *Regulation of vsg expression site transcription and switching in Trypanosoma brucei.* Mol.Biochem.Parasitol., 1998. **91**: p. 77-91.
12. Smith, D.H., J. Pepin, and A.H.R. Stich, *Human African trypanosomiasis: an emerging public health crisis.* Br.Med.Bull., 1998. **54**: p. 341-355.

13. Molyneux, D., *Current public health status of trypanosomiasis and leishmaniasis.* CAB International, 1997: p. 39-50.

14. Docampo, R. and S.N. Moreno, *Current chemotherapy of human African trypanosomiasis.* Parasitol Res, 2003. **90 Supp 1**: p. S10-3.

15. Fairlamb, A.H., *Chemotherapy of human African trypanosomiasis: current and future prospects.* Trends Parasitol, 2003. **19**(11): p. 488-94.

16. Nok, A., *Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human african trypanosomiasis.* Parasitol Res, 2003. **90**: p. 71-79.

17. Barrett, S.V. and M.P. Barrett, *Anti-sleeping sickness drugs and cancer chemotherapy.* Parasitol Today, 2000. **16**(1): p. 7-9.

18. Croft, S.L., L. Vivas, and S. Brooker, *Recent advances in research and control of malaria, leishmaniasis, trypanosomiasis and schistosomiasis.* East Mediterr Health J, 2003. **9**(4): p. 518-33.

19. Kaiser, A., et al., *The necessity to develop drugs against parasitic diseases.* Pharmazie, 2002. **57**(11): p. 723-8.

20. Sturk, L.M., et al., *Distribution and quantitation of the anti-trypanosomal diamidine 2,5-bis(4-amidinophenyl)furan (DB75) and its N-methoxy prodrug DB289 in murine brain tissue.* Acta Trop, 2004. **91**(2): p. 131-43.

21. Blobel, G., *Protein targeting (Nobel lecture).* Chembiochem, 2000. **1**(2): p. 86-102.

22. Stirling, C.J., *Protein targeting to the endoplasmic reticulum in yeast. 1997 Fleming Lecture.* Microbiology, 1999. **145 (Pt 5)**: p. 991-8.

23. Driessen, A.J., E.H. Manting, and C. van der Does, *The structural basis of protein targeting and translocation in bacteria.* Nat Struct Biol, 2001. **8**(6): p. 492-8.

24. Jungnickel, B., T.A. Rapoport, and E. Hartmann, *Protein translocation: Common themes from bacteria to man.* FEBS Lett., 1994. **346**: p. 73-77.

25. Walter, P. and V.R. Lingappa, *Mechanism of protein translocation across the endoplasmic reticulum membrane.* Annu.Rev.Cell Biol., 1986. **2**: p. 499-516.

26. Zheng, N. and L.M. Gierasch, *Signal sequences: the same yet different.* Cell, 1996. **86**: p. 849-852.

27. Zheng, T. and C.V. Nicchitta, *Structural determinants for signal sequence function in the mammalian endoplasmic reticulum.* J Biol Chem, 1999. **274**(51): p. 36623-30.

28. Keenan, R.J., et al., *The signal recognition particle.* Annu Rev Biochem, 2001. **70**: p. 755-75.

29. Rapoport, T.A., M.M. Rolls, and B. Jungnickel, *Approaching the mechanism of protein transport across the ER membrane.* Curr.Opin.Cell Biol., 1996. **8**: p. 499-504.

30. Hann, B.C. and P. Walter, *The signal recognition particle in S. cerevisiae.* Cell, 1991. **67**: p. 131-144.

31. Rusch, S.L. and D.A. Kendall, *Protein transport via amino-terminal targeting sequences: common themes in diverse systems.* Mol.Membr.Biol., 1995. **12**: p. 295-307.

32. Nagai, K., et al., *Structure, function and evolution of the signal recognition particle.* Embo J, 2003. **22**(14): p. 3479-85.

33. Von Heijne, G., *Signal sequences. The limits of variation.* J.Mol.Biol., 1985. **184**: p. 99-105.

34. Von Heijne, G., *Signals for protein targeting into and across membranes.* Subcell.Biochem., 1994. **22**: p. 1-19.

35. Von Heijne, G. and L. Abrahmsen, *Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts.* FEBS Lett., 1989. **244**: p. 439-446.

36. Rusch, S.L. and D.A. Kendall, *Protein transport via amino-terminal targeting sequences: common themes in diverse systems.* Mol Membr Biol, 1995. **12**(4): p. 295-307.

37. von Heijne, G. and L. Abrahmsen, *Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts.* FEBS Lett, 1989. **244**(2): p. 439-46.

38. Borchert, T.V. and V. Nagarajan, *Effect of signal sequence alterations on export of levansucrase in Bacillus subtilis.* J Bacteriol, 1991. **173**(1): p. 276-82.

39. von Heijne, G., *Signal sequences. The limits of variation.* J Mol Biol, 1985. **184**(1): p. 99-105.

40. Garcia, P.D. and P. Walter, *Full-length prepro-alpha-factor can be translocated across the mammalian microsomal membrane only if translation has not terminated.* J.Cell Biol., 1988. **106**: p. 1043-1048.

41. Ramirez, M.I., et al., *Heterologous expression of a Trypanosoma cruzi surface glycoprotein (gp82) in mammalian cells indicates the existence of different signal sequence requirements and processing.* J Eukaryot Microbiol, 1999. **46**(6): p. 557-65.

42. Ramirez, M.I., et al., *Heterologous expression of a Trypanosoma cruzi surface glycoprotein (gp82) in mammalian cells indicates the existence of different signal sequence requirements and processing.* J Eukaryot Microbiol, 1999. **46**(6): p. 557-65.

43. Al-Qahtani, A., M. Teilhet, and K. Mensa-Wilmot, *Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from Trypanosoma brucei and Leishmania.* Biochem.J., 1998. **331**: p. 521-529.

44. Ng, D.T., J.D. Brown, and P. Walter, *Signal sequences specify the targeting route to the endoplasmic reticulum membrane.* J.Cell Biol., 1996. **134**: p. 269-278.

45. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.

46. Rapoport, T.A., B. Jungnickel, and U. Kutay, *Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes.* Annu.Rev.Biochem., 1996. **65**: p. 271-303.

47. Walter, P., R. Gilmore, and G. Blobel, *Protein translocation across the endoplasmic reticulum.* Cell, 1984. **38**: p. 5-8.

48. Walter, P. and A.E. Johnson, *Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane.* Annu.Rev.Cell Biol., 1994. **10**: p. 87-119.

49. Wild, K., et al., *Towards the structure of the mammalian signal recognition particle.* Curr Opin Struct Biol, 2002. **12**(1): p. 72-81.

50. Ogg, S.C., M.A. Poritz, and P. Walter, *Signal recognition particle receptor is important for cell growth and protein secretion in Saccharomyces cerevisiae.* Mol Biol Cell, 1992. **3**(8): p. 895-911.

51. Walter, P. and G. Blobel, *Signal recognition particle: a ribonucleoprotein required for cotranslational translocation of proteins, isolation and properties.* Methods Enzymol., 1983. **96**: p. 682-691.

52. Lipp, J., B. Dobberstein, and M.T. Haeuptle, *Signal recognition particle arrests elongation of nascent secretory and membrane proteins at multiple sites in a transient manner.* J.Biol.Chem., 1987. **262**: p. 1680-1684.

53. Siegel, V. and P. Walter, *Each of the activities of signal recognition particle (SRP) is contained within a distinct domain: analysis of biochemical mutants of SRP.* Cell, 1988. **52**: p. 39-49.

54. Siegel, V. and P. Walter, *Removal of the Alu structural domain from signal recognition particle leaves its protein translocation activity intact.* Nature, 1986. **320**: p. 81-84.

55. Walter, P. and G. Blobel, *Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site- specific arrest of chain elongation that is released by microsomal membranes.* J.Cell Biol., 1981. **91**: p. 557-561.

56. Gilmore, R., P. Walter, and G. Blobel, *Protein translocation across the endoplasmic reticulum. II. Isolation and characterization of the signal recognition particle receptor.* J Cell Biol, 1982. **95**(2 Pt 1): p. 470-7.

57. Schwartz, T. and G. Blobel, *Structural basis for the function of the beta subunit of the eukaryotic signal recognition particle receptor.* Cell, 2003. **112**(6): p. 793-803.

58. Rapiejko, P.J. and R. Gilmore, *Signal sequence recognition and targeting of ribosomes to the endoplasmic reticulum by the signal recognition particle do not require GTP.* Mol.Biol.Cell, 1994. **5**: p. 887-897.

59. Rapiejko, P.J. and R. Gilmore, *Protein translocation across the ER requires a functional GTP binding site in the alpha subunit of the signal recognition particle receptor.* J.Cell Biol., 1992. **117**: p. 493-503.

60. Oliver, J., et al., *The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum.* FEBS Lett., 1995. **362**: p. 126-130.

61. Gorlich, D., et al., *A protein of the endoplasmic reticulum involved early in polypeptide translocation [see comments].* Nature, 1992. **357**: p. 47-52.

62. Panzner, S., et al., *Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p.* Cell, 1995. **81**(4): p. 561-70.

63. Panzner, S., et al., *Posttranslational protein transport into the endoplasmic reticulum.* Cold.Spring.Harbor.Symp.Quant.Biol., 1995. **60**: p. 31-40.

64. Esnault, Y., et al., *SSS1 encodes a stabilizing component of the Sec61 subcomplex of the yeast protein translocation apparatus.* J.Biol.Chem., 1994. **269**: p. 27478-27485.

65. Gorlich, D., et al., *A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation.* Cell, 1992. **71**: p. 489-503.

66. Kalies, K.U., D. Goerlich, and T.A. Rapoport, *Binding of ribosomes to the rough endoplasmic reticulum mediated by the Sec61p-complex.* J.Cell Biol., 1994. **126**: p. 925-934.

67. Crowley, K.S., G.D. Reinhart, and A.E. Johnson, *The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation.* Cell, 1993. **73**: p. 1101-1115.

68. Deshaies, R.J. and R. Schekman, *A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum.* J.Cell Biol., 1987. **105**: p. 633-645.

69. Wilkinson, B.M., A.J. Critchley, and C.J. Stirling, *Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex.* J.Biol.Chem., 1996. **271**: p. 25590-25597.

70. Deshaies, R.J., et al., *Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex.* Nature, 1991. **349**: p. 806-808.

71. Matlack, K.E., W. Mothes, and T.A. Rapoport, *Protein translocation: tunnel vision.* Cell, 1998. **92**: p. 381-390.

72. Schekman, R., *Polypeptide translocation: a pretty picture is worth a thousand words [comment].* Cell, 1996. **87**: p. 593-595.

73. Green, N., H. Fang, and P. Walter, *Mutants in three novel complementation groups inhibit membrane protein insertion into and soluble protein translocation across the endoplasmic reticulum membrane of Saccharomyces cerevisiae.* J.Cell Biol., 1992. **116**: p. 597-604.

74. Plath, K., et al., *Signal sequence recognition in posttranslational protein transport across the yeast ER membrane.* Cell, 1998. **94**(6): p. 795-807.

75. Gething, M.J. and J. Sambrook, *Protein folding in the cell.* Nature, 1992. **355**(6355): p. 33-45.

76. Rudiger, S., A. Buchberger, and B. Bukau, *Interaction of Hsp70 chaperones with substrates.* Nat Struct Biol, 1997. **4**(5): p. 342-9.

77. Deshaies, R.J., et al., *A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides.* Nature, 1988. **332**(6167): p. 800-5.

78. Caplan, A.J., D.M. Cyr, and M.G. Douglas, *YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism.* Cell, 1992. **71**(7): p. 1143-55.

79. Sanders, S.L., et al., *Sec61p and BiP directly facilitate polypeptide translocation into the ER.* Cell, 1992 Apr 17. **69**(2): p. 353-65.

80. Matlack, K.E., et al., *BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane [In Process Citation].* Cell, 1999. **97**(5): p. 553-64.

81. Lyman, S.K. and R. Schekman, *Interaction between BiP and Sec63p is required for the completion of protein translocation into the ER of Saccharomyces cerevisiae.* J Cell Biol, 1995. **131**(5): p. 1163-71.

82. Misselwitz, B., et al., *Interaction of BiP with the J-domain of the Sec63p component of the endoplasmic reticulum protein translocation complex.* J Biol Chem, 1999. **274**(29): p. 20110-5.

83. Rout, M.P. and M.C. Field, *Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component.* J Biol Chem, 2001. **276**(41): p. 38261-71.

84. Bangs, J.D., et al., *Molecular cloning and cellular localization of a BiP homologue in Trypanosoma brucei. Divergent ER retention signals in a lower eukaryote.* J.Cell Sci., 1993. **105**: p. 1101-1113.

85. Bangs, J.D., et al., *A soluble secretory reporter system in Trypanosoma brucei - Studies on endoplasmic reticulum targeting.* J.Biol.Chem., 1996. **271**: p. 18387-18393.

86. Coppens, I., et al., *Receptor-mediated endocytosis in the bloodstream form of Trypanosoma brucei.* J Protozool, 1987. **34**(4): p. 465-73.

87. Weise, F., et al., *Distribution of GPI-anchored proteins in the protozoan parasite Leishmania, based on an improved ultrastructural description using high-pressure frozen cells.* J Cell Sci, 2000. **113**(Pt 24): p. 4587-4603.

88. Webster, P. and D.J. Grab, *Intracellular colocalization of variant surface glycoprotein and transferrin-gold in Trypanosoma brucei.* J Cell Biol, 1988. **106**(2): p. 279-88.

89. Langreth, S.G. and A.E. Balber, *Protein uptake and digestion in bloodstream and culture forms of Trypanosoma brucei.* J Protozool, 1975. **22**(1): p. 40-53.

90. McConville, M.J., et al., *Secretory pathway of trypanosomatid parasites.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 122-54; table of contents.

91. Steverding, D., et al., *ESAG 6 and 7 products of Trypanosoma brucei form a transferrin binding protein complex.* Eur.J.Cell Biol., 1994. **64**: p. 78-87.

92. Steverding, D., et al., *Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei.* J.Cell Biol., 1995. **131**: p. 1173-1182.

93. El-Sayed, N.M. and J.E. Donelson, *African trypanosomes have differentially expressed genes encoding homologues of the Leishmania GP63 surface protease.* J Biol Chem, 1997. **272**(42): p. 26742-8.

94. Button, L.L. and W.R. McMaster, *Molecular cloning of the major surface antigen of leishmania.* J Exp Med, 1988. **167**(2): p. 724-9.

95. Handman, E., et al., *The Leishmania promastigote surface antigen 2 complex is differentially expressed during the parasite life cycle.* Mol.Biochem.Parasitol., 1995. **74**: p. 189-200.

96. Ilg, T., et al., *Molecular cloning and characterization of a novel repeat-containing Leishmania major gene, ppg1, that encodes a membrane-associated form of proteophosphoglycan with a putative glycosylphosphatidylinositol anchor.* J Biol Chem, 1999. **274**(44): p. 31410-20.

97. Teixeira, S.M.R., L.V. Kirchhoff, and J.E. Donelson, *Post-transcriptional elements regulating expression of mRNAs from the amastin/tuzin gene cluster of Trypanosoma cruzi.* J.Biol.Chem., 1995. **270**: p. 22586-22594.

98. Debrabant, A., M. Gottlieb, and D.M. Dwyer, *Isolation and characterization of the gene encoding the surface membrane 3'-nucleotidase/nuclease of Leishmania donovani.* Mol.Biochem.Parasitol., 1995. **71**: p. 51-63.

99. Ilg, T., et al., *Distribution of parasite cysteine proteinases in lesions of mice infected with Leishmania mexicana amastigotes.* Mol.Biochem.Parasitol., 1994. **67**: p. 193-203.

100. Schenkman, S., et al., *Mucin-like glycoproteins linked to the membrane by glycosylphosphatidyli ositol anchor are the major acceptors of sialic acid in a reaction catalyzed by trans-sialidase in metacyclic forms of Trypanosoma cruzi.* Mol.Biochem.Parasitol., 1993. **59**: p. 293-304.

101. Cross, G.A.M. and G.B. Takle, *The surface trans-sialidase family of Trypanosoma cruzi.* Annu.Rev.Microbiol., 1993. **47**: p. 385-411.

102. Haynes, P.A., D.G. Russell, and G.A.M. Cross, *Subcellular localization of Trypanosoma cruzi glycoprotein Gp72.* J.Cell Sci., 1996. **109**: p. 2979-2988.

103. Liu, L., et al., *RNA interference of signal peptide-binding protein SRP54 elicits deleterious effects and protein sorting defects in trypanosomes.* J Biol Chem, 2002. **277**(49): p. 47348-57.

CHAPTER 2

CELL-FREE PROTEIN IMPORT INTO THE ENDOPLASMIC RETICULUM OF A TRYPANOSOME: FUNCTIONAL COEVOLUTION OF SIGNAL PEPTIDES AND TRANSLOCONS [1]

[1]Patham, B., Morris, M. T., Brodsky, J. L., Wipf, P., and Mensa-Wilmot, K.

Submitted to *Molecular Biology of Cell*, 05/2005

LIST OF ABBREVIATIONS

α-MF, α-mating factor; CfRM, canine microsomes; CHX, cycloheximide; GPI, glycosylphosphatidylinositol; NP-40, nonidet P-40; PK, proteinase K; PMSF, phenylmethyl-sulfonylfluoride; RNC, ribosome nascent chain; RRL, rabbit reticulocyte lysate; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TbRM, *T. brucei* microsomes; VSG, variant surface glycoprotein.

ABSTRACT

Movement of proteins into the endoplasmic reticulum (ER) is an important stage in the biogenesis of several organelles in the secretory pathway. Signal peptides and a Sec61p protein translocon are essential for co-translational and post-translational import of proteins into the ER. In trypanosomes, many aspects of the pathways and factors important for protein import into the ER are not understood. Although it is widely held that signal sequences are universally interchangeable between species, several proteins from *Trypanosoma brucei*, causative agent of human African trypanosomiasis, could not be imported into canine ER microsomes. To explore mechanisms underlying this import defect, the interactions of a variant surface glycoprotein (VSG_117) signal peptide with canine microsomes were examined: the trypanosome signal peptide was not recognized by the canine Sec61p translocon. Thus, the ER translocon is a major determinant of species-specificity in signal peptide activity. In order to perform further studies of trypanosome ER protein import, we have developed a cell-free system using *T. brucei* microsomes. *T. brucei* microsomes (TbRM) imported proteins post-translationally. Cytosol from the parasite was required for translocation of VSGs into TbRM. Hsp70 chaperones that are stimulated by J-domain co-chaperones are important for import of proteins into TbRM. Signal peptides from other species (*e.g., Escherichia coli*, *Saccharomyces cerevisiae*, *Leishmania major* and *Bos taurus*) could not direct import of proteins into TbRM. ER translocons appear to have coevolved with signal peptides from the same species.

INTRODUCTION

Cell surface proteins of *Trypanosoma brucei* play important roles as signaling receptors, nutrient transporters and virulence factors. Movement of these proteins to the plasma membrane typically involves import into the endoplasmic reticulum (ER). An N-terminal signal sequence (reviewed in [1, 2]) and a translocon composed of Sec61p (α, , β, , γ) [3-5] are essential for protein translocation into the ER. Both co-translational and post-translational pathways are used for import of protein into the ER.

For co-translational translocation, a signal peptide binds to signal recognition particle (SRP) to form a "targeting complex" (reviewed in [6]) that "docks" on the ER membrane [7] where the nascent polypeptide chain is transferred from SRP to Sec61p [8]. "Gating" of the Sec61p pore occurs when it recognizes a functional signal peptide [9-11].

Post-translational import of proteins into the ER is documented in *Saccharomyces cerevisiae* (reviewed in [12, 13]). (In *Escherichia coli* post-translational export of proteins across the luminal membrane is extensively documented (reviewed in [14].) Unique factors needed for post-translational translocation include the membrane proteins Sec62p, Sec71p, and Sec72p [15- 18], and molecular chaperones Ssa1p (Hsc70) and Ydj1p (Hsp40) [19-22]. Whereas Sec61p is the translocon for post-translational protein import [23-25], SRP is not required for post-translational translocation [3, 21, 26-28].

The composition and internal organization of signal peptides differs among species [29,30]. However, the biological significance of the observation remained elusive, since signal sequences from one species can sometimes operate with the ER translocation machinery of another [31, 32]. Recent studies indicate that signal peptides from one species may not work well with the translocation system from another species [33-36], in accordance with the variation in signal sequence composition between

different organisms [29, 37-39]. Similarly, the Sec61p translocon is not functionally interchangeable between related yeasts [40, 41].

Signal peptides are important for entry into the secretory system [42]. SRP54p (the 54 kDa protein component of the SRP complex) is dispensable for import of several proteins into the ER of *T. brucei* [43]. In a canine microsomal system, several signal peptides from *T. brucei* failed to direct ER import of trypanosome proteins. Replacement of the signal sequence of a variant surface glycoprotein (VSG_117) with that from yeast α-mating factor (α-MF) led to import of the *T. brucei* protein [35]. Thus, the trypanosome signal peptide is incompatible with the vertebrate ER protein import machinery.

In vivo evidence confirm these *in vitro* observations. First, attempts to secrete recombinant *Leishmania* gp63 from insect Sf9 cells failed. When the trypanosomatid signal sequence was replaced with a baculovirus signal peptide gp63 was secreted [44]. Second, when gp82 from *Trypanosoma cruzi* was expressed in murine *Vero* cells the protein was not targeted to the plasma membrane where it is found in the parasite. Replacement of the *T. cruzi* signal peptide with a signal sequence from influenza virus hemagglutinin targeted gp82 to the plasma membrane of *Vero* cells [34]. Thus, both *in vitro* and *in vivo*, trypanosomatid signal peptides seem to be incompatible with translocation systems from other species.

In order to elucidate the biochemical basis of the species-specificity in signal sequence utilization between the canine translocon and *T. brucei* signal peptides, we have studied the interactions between a trypanosome signal peptide and the canine ER translocon. VSG_117 nascent chain is targeted to canine ER, but fails to interact productively with canine Sec61p. This data reveals the ER translocon as a species-specificity factor in protein import into the organelle.

Little work has been performed on trypanosome ER protein translocation, despite the obvious importance of the pathway for localizing virulence factors and signaling receptors to the plasma membrane of the parasite. For example, the properties of signal peptides have not been studied directly, and the pathways used for protein import into the ER have not been completely characterized. Initiatives to address some of these issues will be advanced rapidly if an in vitro system was available to study some of the mechanistic steps, due the exquisite level of control that investigators can apply to such a cell-free system. (See [25, 45] for reviews of such efforts as applied to yeast and canine ER protein translocation.) Since canine pancreas microsomes that are widely used to study mechanisms of polypeptide import failed to import *T. brucei* proteins [35], we have developed a cell-free system that imports proteins from the trypanosome.

In studies with *T. brucei* microsomes (TbRM), several unexpected properties of the ER protein import have been discovered. TbRM imported VSGs post-translationally, providing only the second example, after *S. cerevisiae* [3, 46], of post-translational ER protein import in a eukaryote. A signal peptide was required, as anticipated, for import of proteins into TbRM. Short polypeptides (i.e., less than 114 amino acids long) could be imported efficiently into TbRM from a rabbit reticulocyte lysate. For longer proteins (250 amino acids and longer), cytosol from *T. brucei* was required for import into TbRM. Surprisingly, proteins from other species could not be imported into TbRM. Apparently, signal peptides from *T. brucei* have coevolved with the ER translocon such that only the parasite proteins are imported into TbRM.

MATERIALS AND METHODS

DNA Templates for RNA Synthesis *In Vitro*

The DNA substrate for synthesis of VSG_117$_{86}$, VSG_117$_{114}$, VSG_117$_{260}$ and VSG_117$_{500}$ was pVSG_117 (kindly provided by Dr. Jay Bangs, University of Wisconsin, Madison). The templates were

generated by PCR using the forward primer ccc*taatacgactcactata*<u>ggg</u>aggagggtttttacc<u></u>atggactgccata caaaggag which contains a T7 promoter (italicized), and a translation enhancer (underlined) [47]. The first 21 nucleotides of VSG_117 coding sequence are in in regular style font. The forward primer for VSG_117_{500}_SP and VSG_117_{500}_SP was ccc*taatacgactcactata*<u>ggg</u>aggagggtttttacc<u></u>atggccacactgagaa aggttgc. It contained nucleotides 51-57 of the coding region (regular font). Other features are similar to those mentioned for the previous forward primer. The reverse primer for VSG_117_{86} was cgaa caacgaaggggttcttatagtgcgtagattcgtagcttcgtttc; a *Hinf*I site (in bold) and nucleotides 78-86 of the coding region of VSG_117 (normal font) are shown. The reverse primer for VSG_117_{114} was *tca*ctgaat gctttcgagtgctttttgtcgcattagggc. It contained a *XmnI* site (in bold) and nucleotides 106- 113 of the coding region (regular font). The reverse primer for VSG_117_{260} was *catcatcat*cgaaagtttcaagccgatggt. The primer contained nucleotides 254-260 of the coding region (regular font). The reverse primer for VSG_117_{500} was *catcatcat*ttcctaaaaaagcaaggc. It contained nucleotides 516-525 of the coding region (regular font). The substrate for synthesis of VSG_MVAT7 was plasmid SL/3 (gift from J. Donelson, Univ. of Iowa). The forward primer for the reaction was ccc*taatacgactcactata*<u>ggg</u>aggagggtttttacc<u></u>atgtcaacaagagtccaacaa. It contained a T7 promoter (italicized), a translational enhancer (underlined), and the first 21 nucleotides of VSG_MVAT7 coding sequence (in regular font).

The DNA amplification products were purified using "Wizard PCR preps" (Promega), extracted with phenol/chloroform/isoamyl alcohol, ethanol precipitated, and quantitated by OD_{260} absorbance. The substrate for synthesis of a ribosome-nascent chain (RNC) containing 86 amino acids from preprolactin (pPL_{86}) was produced by *Pvu*II linearization of pGEM2pPLwt [15] (provided by Dr. Tom Rapoport, Harvard Univ.). Full-length pPL was obtained by *Pst*I linearization of pGEM2pPLwt. α-mating factor and β-lactamase mRNA's were obtained from Promega (as part of a canine microsome kit).

Transcription Reactions

Two approaches were used to produce RNAs *in vitro*. In the first method, two !g of DNA template was added to a reaction containing 1 mM each ribonucleoside triphosphates, 5 mM dithiothreitol, 50 units T7 polymerase (Gibco BRL), 40 mM Tris-HCl (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine, 125 mM NaCl, and 40 units RNAsin (Promega). Reactions were carried out for 1 h at 37°C, and products purified as described [48]. In the second method for RNA synthesis, one !g of DNA template was transcribed using the AmpliscribeTM T7 Kit (Epicentre Technologies). One !l (1 MBU) of RNAse free DNAse I was added to the reaction which was incubated for 15 min at 37°C. The mixture was centrifuged at 14,000 x g (15 min, 4°C), and the pellet was washed with 70% ethanol. Purified RNA was resuspended in 40 !l RNAse free T10E1. RNA concentration was determined by UV absorbance.

Protein Synthesis *In Vitro*

One !g of RNA was translated in a reaction containing 10 !l of rabbit reticulocyte lysate (Promega), 1 mM methionine/cysteine-free amino acid mixture (Promega), and 15 !Ci [35S] redivue Promix (Amersham Bioscience) in a total volume of 20 !l.

Co-translational Protein Import Into Canine Microsomes

One !g of RNA was translated in the presence of 4 !l canine pancreas microsomes (Promega, 1 U/!l) for 90 min at 30°C. This reaction mixture was aliquoted into three equal portions and treated with one of the following; (i) 120 !g/ml protease K; (ii) 120 !g/ml protease K and Triton-X-100 (1%); (iii) control buffer (250 mM sucrose, 150 mM KOAc, 5 mM Mg(OAc)2, 1 mM DTT (SRB)) on ice for 1 h. Phenylmethylsulfonylfluoride (PMSF) (20 mM final) was added to stop the reaction. Samples were precipitated with an equal volume of cold (4°C) saturated $(NH4)_2SO4$ and resolved by SDS-PAGE (10% total acrylamide, 3% crosslinker in a Tricine-HCl system) [49]. Radioactive polypeptides were visualized by fluorography.

Binding of Ribosome-Nascent Chains (RNC) to Canine Microsomes

Targeting of RNCs to microsomes was determined by a membrane floatation assay [50] with minor modifications. Briefly, 30 !l translation reactions were adjusted to 2.1 M sucrose with 330 !l of 2.3 M sucrose in RBB (50 mM HEPES-KOH (pH 7.4), 150 mM KOAc, 5 mM $Mg(OAc)_2$, and 2 mM DTT). Samples were overlaid with an equal volume (360 !l) of 1.9 M sucrose in RBB. Equal volumes (360 !l) of sucrose-free RBB were placed on top, and the cushions centrifuged for 2 h at 186,000 x g (4°C) (Beckman TLA 100.3 rotor). After centrifugations, samples were immediately frozen in liquid nitrogen. Using a razor blade and hammer, the frozen tubes were sectioned into top and bottom fractions of approximately 540 !l each. Frozen fractions were transferred to pre-labeled microcentrifuge tubes, thawed, precipitated with an equal volume of saturated $(NH4)_2SO4$, and analyzed as described above.

Microsome Protection of Nascent Chains from Protease K Digestion

Translation reactions (see above) performed in presence or absence of canine microsomes were divided into several equal fractions. When indicated, samples were digested with protease K (120 !g/ml) for 1 h on ice. Proteins were separated by SDS-PAGE. Radioactive polypeptides were detected by fluorography using pre-flashed Hyperfilm-MP (Kodak) at –80°C. Quantitation of fluorographs was by laser scanning densitometry with an IS-1000 Digital Imaging System (Alpha Innotech, CA).

Preparation of *T. brucei* Microsomes (TbRM)

A pellet of 2 x 10^9 blood stream form *T. brucei* was washed and resuspended in 1 ml of homogenization buffer (HB) (250 mM sucrose, 50 mM KOAc, 1 mM EDTA, 1 mM DTT). Cells were lysed with twenty strokes of a tight-fitting pestle in a Dounce homogenizer. The mixture was first centrifuged (2000 x g, 10 min, 4°C) to pellet particulate material including unlysed trypanosomes. The resulting supernatant was centrifuged at 12,000 x g (20 min, 4°C) to pellet rough microsomes (TbRM) that was resuspended in 40 !l RM buffer (RMB) (250 mM sucrose, 50 mM HEPES-KOH, pH 7.6, 50

mM potassium acetate, 1 mM DTT, 5 !g/ml leupeptin). An aliquot of the pellet was diluted 50-fold in 0.1% SDS and absorbance (OD_{280}) determined. Stock TbRM was diluted with RM buffer to OD_{280} of 50 (1!l of this TbRM suspension has one equivalent of microsomes). Aliquots were frozen in liquid nitrogen, and stored at –80°C.

Preparation of Cytosol from *T. brucei*

Supernatant obtained during preparation of TbRM (see above) was centrifuged at 65,000 x g (60 min, 4°C, Beckman TLA 100.3 rotor). The supernatant obtained was concentrated twenty-fold by ultrafiltration using a Centricon-10 filter (Amicon). The retentate was retrieved, an aliquot was diluted 50-fold with 0.1% SDS, and the OD_{280} obtained. One equivalent of cytosol has OD_{280} of 50. Aliquots were quick-frozen in liquid nitrogen, and stored at -80° C.

Protein Import Into *T. brucei* Microsomes (TbRM)

Two !g of RNA encoding a truncated substrate (*e.g.,* VSG_117_{86}) was translated in 40 !l of a reaction containing 20 !l rabbit reticulocyte lysate, 60 !M amino acids (-Met, -Cys), 2.4 !Ci [35S]promix. The reaction was incubated at 37°C for 15 min. Cycloheximide (50 !g/ml, final concentration) was used to stop further translation. The reaction was aliquoted into two portions. To one aliquot *T. brucei* microsomes (TbRM) (1 equivalent) was added, and to the other portion an equal volume of RM buffer added. The reactions were incubated at 37°C for 45 min. Each of reactions was divided into three portions of 10 !l and treated with one of the following: (i) RM buffer; (ii) 300 !g/ml protease K; (iii) NP-40 (2%, final concentration) followed by protease K (300 !g /ml, final concentration), on ice for 1 hour. PMSF (20 mM final concentration) was added to stop proteolysis. Samples were precipitated with an equal volume of cold ammonium sulphate (saturated) and resolved by SDS–PAGE (4% total acrylamide, 3% cross linker in tricene-HCl system) [47]. The gels were dried and visualized with a phosphorimager (Molecular Imager FX, (BioRad). Data bands were quantitated with QuantityOne software (BioRad), and graphs plotted with Canvas 9.0 (Deneba). To import full-length proteins into

TbRM, a modification was made to the scheme described (above) for import of truncated proteins. Two !g of RNA was translated in 40 !l mixture containing 20 !l rabbit reticulocyte lysate, 60 !M amino acids (-Met, -Cys), 2.4 !Ci [35S]promix. Cytosol from *T. brucei* (1.5 equivalents) was added 20 min into the translation reaction, which was then incubated at 37°C for further 40 min. Cycloheximide (50 !g/ml, final concentration) was used to stop further translation. The mixture was then divided into two equal portions. To one aliquot, RM buffer (1.5 !l) was added whereas the other portion received TbRM (1 equivalent, in 1.5 !l of RM). The reactions were then incubated at 37°C for 1 h, after which protease protection assays were performed as described earlier.

RESULTS

Detection of Trypanosome VSG_117-RNC on Canine ER Microsomes

Based on the fact that signal sequences are recognized twice during co-translational protein import into the canine endoplasmic reticulum [51, 52], failure of canine microsomes to import *T. brucei* proteins [35] may be explained by two hypotheses. First, it is conceivable that reticulocyte ribosomes that are translating *T. brucei* proteins fail to dock at the ER. Second, trypanosome signal sequences may bind to the ER but fail to execute a posttargeting event, for example, “gate” the canine Sec61p translocon. The second model is used for rejection defective preprolactin signal peptides that are targeted to the ER [51].

To evaluate the contributions of each of the above pathways to the failure of trypanosome signal sequences to direct import of proteins into canine microsomes, we first tested whether ribosome-nascent-chains (RNCs) containing a trypanosome VSG_117 signal peptide (VSG_ 117-RNC) could be targeted to canine ER microsomes. Preprolactin RNC (pPL-RNC) was studied as a positive control [51]. VSG_117 and pPL-RNCs were produced in a rabbit reticulocyte lysate in the absence or presence of canine microsomes. The RNCs were floated through a sucrose cushion [50], allowing free RNCs to sediment to the bottom, while membrane-bound RNCs floated at the top of the sucrose cushion.

In the absence of canine microsomes, almost all the signal for both pPL and VSG_117- RNCs was recovered from the bottom fraction of the sucrose cushion (Fig. 1, compare lanes 1 & 2). When microsomes were present during translation, 30% of pPL-RNC floated in the top fraction of the cushion (Fig. 1A, compare lanes 3 & 4). Under similar conditions, 15% of VSG_117-RNC is membrane-bound (Fig. 1B, compare lanes 3 & 4). Therefore, the relative targeting efficiency of VSG_117 RNC is 50% that of pPL-RNC. We conclude that VSG_117-RNC is targeted to the canine ER. Additionally, sufficient VSG _117 RNC is detected on the ER that import of VSG would have been detected if it had taken place, because the microsomal import assay (in our hands) detects 5% import of a substrate. Nevertheless, inefficient targeting of the trypanosome RNC to the ER probably makes a contribution to poor translocation competence of the protein [35].

Interactions of a Trypanosome Nascent Chain With The Canine ER Translocon

Since VSG_117 RNC was detected on canine microsomes, we determined whether the signal peptide interacted productively with the canine Sec61p translocon. When a signal peptide engages Sec61p effectively, a "tight seal" is formed between the ribosome and the ER membrane, which precludes digestion of the NC by exogenous protease K [53]. Hence, sequestration of a NC into a "protease-protected environment" indicates that, signal peptide is recognized by Sec61p [8, 51, 53]. In fact, purified Sec61p in liposomes protects a signal peptide within a targeting complex from protease digestion [54, 55].

To determine the susceptibility of the NCs to protease K, VSG_117 and pPL-RNCs were produced in presence or absence of canine microsomes and treated with protease K. In the absence of microsomes, most of pPL-NC is degraded by protease K (Fig. 2A, compare lanes 1&2). When two equivalents of microsomes are added, 60% of pPL-NC is protected from the protease (Fig. 2A, compare lanes 3 & 4). With four equivalents of microsomes, 100% of the pPL nascent-chain is protected (Fig. 2A

compare lanes 5 & 6). These observations indicate that canine Sec61p recognizes pPL signal peptide, as reported by other investigators [56].

Protease susceptibility of VSG-NC differed markedly from that observed with pPL-NC. Full-length VSG_117-NC is completely degraded by protease K in the absence of microsomes (Fig. 2B, compare lanes 1&2). When two equivalents of microsomes was introduced into the translation reaction, only 1% (Fig. 2B and Fig. 2C) of the full-length nascent-chain is protected from protease. On addition of four equivalents of microsomes 5.8% protection of full-lenth VSG-NC was achieved (Fig. 2B lanes 5-6; Fig. 2C). Efficiency of the ribosome•ER-membrane interaction triggered by VSG-NC, as judged by the proportion of NC that is protected by two equivalents of microsomes, is approximately 60-fold less than that of pPL-NC (Fig. 2C). The simplest interpretation of the data (above) is that canine Sec61p cannot recognize the trypanosome signal sequence efficiently. Therefore, defective Sec61p•signal peptide interactions contribute significantly to inability of canine microsomes to import trypanosome proteins.

Cell-Free Protein Import by *T. brucei* Microsomes (TbRM)

We have a long-term interest in learning the pathways and mechanisms used by secretory proteins to cross the ER membrane in *T. brucei*. Those studies will be advanced rapidly by availability of an *in vitro* microsomal system with which mechanistic studies can be performed. Herein, we report the establishment of a cell-free system using *T. brucei* microsomes (TbRM) that imports trypanosome proteins. Truncated VSG_117 with $_{86}$ amino acids (*i.e.* VSG_117$_{86}$) was used as the initial substrate to establish the system, with the reasoning that the requirements for microsomal import of a truncated protein could be less stringent than those of a full-length polypeptide.

Standard assays to test microsomal ER protein import include (i) proteinase protection of translocated cargo, and (ii) loss of protease protection after detergent permeabilization of ER membranes

[53]. Detergent treatment of microsomes creates pores in the membranes that allows protease to access and digest imported proteins [14].

VSG_117_{86} mRNA was translated in rabbit reticulocyte lysate followed by addition of cycloheximide to terminate further protein synthesis. TbRM was added post-translationally for import of VSG_117_{86}, and reaction mixtures treated with proteinase K. Finally, protease digestion of VSG_117_{86} in presence of detergent and TbRM was performed. (See Fig. 3A for an outline of the experimental scheme.)

In the absence of TbRM, VSG_117_{86} was degraded efficiently by proteinase K (Fig. 3B, lane 2). TbRM protected VSG_117_{86} from proteinase K digestion (Fig. 3B, lane 4). Following detergent permeabilization of TbRM, proteinase K digested VSG_117_{86} (Fig. 3B, lane 5). From these studies, we conclude that VSG_117_{86} is imported into TbRM. Notably, protein import was post-translational, since TbRM was added after the translation had been terminated with cycloheximide. Attempts at co-translational TbRM import system were not successful, because TbRM inhibited synthesis of protein by the reticulocyte lysate.

Effect of Polypeptide Length on Import into TbRM

Our success at importing a truncated substrate (*i.e.,* VSG_117_{86}) into TbRM, encouraged us to use longer polypeptides of length 114 (VSG_117_{114}), 260 (VSG_117_{260}) and 500 (VSG_117_{500}) as substrates for import into TbRM. For this study, VSG_117_{86} was used as a control to normalize import efficiency of the different polypeptides. mRNAs for VSGs of varying length were translated in rabbit reticulocyte lysate followed by addition of cycloheximide to stop polypeptide synthesis, after which TbRM was added for import of the proteins. (See Fig. 4A for outline of procedures.)

Without TbRM, VSG_117_{114} was degraded by proteinase K (Fig. 4B, lane 2). TbRM protected VSG_117_{114} from proteinase K digestion (Fig. 4B, lane 4). Unexpectedly, VSG_117_{86} and VSG_117_{114}

were imported into TbRM with varying efficiency. Whereas eighty five percent of VSG_117_{86} was imported (Fig. 4F), only forty percent of VSG_117_{114} was translocated into TbRM (Fig. 4F). Longer substrates,that is VSG_117_{260} and VSG_117_{500}, were degraded by proteinase K in the absence (Fig. 4D & E, lane 2) or presence of TbRM (Fig. 4D & E, lane 4). Therefore, neither VSG_117_{260} nor VSG_117_{500} is imported into TbRM (Fig. 4F).

In conclusion, for a set of proteins containing the same signal peptide, length of the polypeptide affects import efficiency into TbRM. Shorter polypeptides like VSG_117_{86} (Fig. 3) and VSG_117_{114} (Fig. 4B) are efficiently imported into TbRM. Longer polypeptides, VSG_117_{260} (Fig. 4D) & VSG_117_{500} (Fig. 4E) are not imported into the parasite microsomes.

Cytosol from *T. brucei* Stimulates Protein Import Into ER Microsomes

Post-translational import of proteins into the ER requires cytosolic chaperones Ssa1p (Hsc70), Ydj1p (Hsp40) in yeast [20-22, 57]. In *E. coli* SecB is a general chaperone for protein export across the inner membrane [58-60]. Molecular chaperones enable full-length proteins to maintain an "import-competent" conformation that is necessary for translocation into the ER (reviewed in [12, 61]). Based on these considerations, we suspected that longer VSGs failed at import into TbRM (Fig. 4D & E) because of the absence of molecular chaperones.

In order to test this hypothesis, we prepared cytosol from *T. brucei* and evaluated its effect on translocation of VSG_117_{260} and VSG_117_{500} neither of which could be imported into TbRM in the absence of cytosol (Fig. 4D and Fig. 4E). Without TbRM, VSG_117_{260} was degraded by proteinase K (Fig. 5B, lanes 2 and 6). TbRM alone did not protect VSG_117_{260} from proteinase K digestion (Fig. 5B, lane 2). However, cytosol from *T. brucei* together with TbRM protected VSG_117_{260} from proteinase K digestion (Fig. 5B, lane 8).

We conclude that *T. brucei* cytosol is required for VSG_117$_{260}$ import into TbRM. VSG_117$_{500}$ was also studied as a substrate for import into TbRM (see Fig. 5A for outline of protocols). In the absence of TbRM, VSG_117$_{500}$ was degraded by proteinase K (Fig. 5C, lane 2). Further, TbRM failed to protect VSG_117$_{500}$ from protease K cleavage (Fig. 5C, lane 2). Cytosol and TbRM when added together protected VSG_117$_{500}$ from protease K (Fig. 5C, lane 5). Detergent permeabilization of TbRM allowed proteinase K to digest VSG_117$_{500}$ in presence of TbRM and cytosol (Fig. 5C, lane 6), indicating that protease-protected VSG_117$_{500}$ was imported into TbRM.

We conclude that *T. brucei* cytosol is necessary for import of full-length VSG_117 into TbRM. It was important to determine whether TbRM could import other full-length trypanosome proteins beside VSG_117$_{500}$. For this purpose, translocation of VSG_MVAT7 was studied following the approaches summarized for VSG_117$_{500}$ (see Fig. 5A). In absence of TbRM, VSG_MVAT7 was degraded by proteinase K (Fig. 5D, lane 2). Addition of cytosol and TbRM led to protection of VSG_MVAT7 from protease K (Fig. 5D, lane 4). However, in presence of detergent, proteinase K degraded VSG_MVAT7 that was protected in by TbRM and cytosol. We conclude that full-length VSG_MVAT7 is imported into TbRM. These data imply that *T. brucei* cytosol allows post-translational import of full-length proteins to TbRM.

Signal Peptide Dependence of Protein Import into TbRM

A signal peptide directs import of protein into the ER (reviewed in [62]). To test whether a signal sequence was important for translocation of protein into TbRM, we first deleted the signal peptide from VSG_117$_{86}$ coding sequence to obtain VSG_117$_{86}$_SP whose mRNA was translated and import of the protein into TbRM studied as outlined for the other proteins earlier. In the absence of TbRM, VSG_117$_{86}$_SP was cleaved by protease K (Fig. 6B, lane 2). TbRM could not protect VSG_117$_{86}$_SP from digestion by proteinase K. Therefore, VSG_117$_{86}$_SP is not imported by TbRM.

We next tested whether a signal peptide was essential for import of full-length VSG_117 into TbRM. For this objective, we deleted the signal peptide from full-length VSG_117 to obtain VSG_117_{500}_SP coding sequence, which was translated in a reticulocyte lysate. Without TbRM, VSG_117_{500}_SP was degraded by proteinase K (Fig. 6C, lane 2). TbRM and *T. brucei* cytosol could not protect VSG_117_{500}_SP from proteinase K (Fig. 6C, lane 4). These data indicate that VSG_117_{500}_SP is not imported into TbRM even in the presence of *T. brucei* cytosol. In conclusion, signal peptide is essential for protein import into TbRM.

Specificity of TbRM for Signal Peptides

It is widely held that signal sequences are interchangeable between different species. Curiously, several proteins from the trypanosomatids *T. brucei* and *Leishmania major* cannot be imported into canine microsomes [35], indicating that a vertebrate translocation system may not recognize trypanosome signal sequences (Fig. 1 & Fig. 2B). Given this background, we wanted to learn whether (or not) TbRM exhibited any specificity for signal sequences that did not originate from trypanosomes: Test proteins included *E. coli* β-lactamase, *S. cerevisiae* α-mating factor, bovine preprolactin, and *Leismania major* gp63.

All polypeptides were translated in rabbit reticulocyte lysate in presence of *T. brucei* cytosol (Fig. 7B-E, lanes 1-4). TbRM was added post-translationally (Fig. 7B-E, lanes 3-4) and translated products were treated with proteinase K (Fig. 7B-E, lanes 2, 4). Unlike the *T. brucei* proteins, all polypeptides studied were degraded by proteinase K in absence of (Fig. 7B-E, lane 2) or presence of TbRM and *T. brucei* cytosol (Fig. 7B-E, lane 4). Therefore, none of the polypeptides from other species was translocated into TbRM.

An Inhibitor of Hsp70 ATPase Blocks Protein Import into TbRM

Post-translational ER protein translocation in *S. cerevisiae* is independent of SRP [6, 28] but requires cytosolic chaperones Ssa1p (Hsp70) and the co-chaperone Ydj1p (Hsp40) [21, 57, 63]. Since protein import into TbRM is post-translational, and dependent on cytosol from the parasite, we considered a possibility that a trypanosome Hsp70 or SRP in cytosol is the factor that stimulates protein import into TbRM. It seems unlikely that SRP is the stimulatory factor because SRP is not needed for post-translational protein import into the ER [61, 64]. Further, SRP is not needed for translocation of several proteins into the ER of *T. brucei* [43]. On the other hand there is no evidence that chaperones participate in protein import into the ER of a trypanosome.

We have used a pharmacological approach to explore whether (or not) an Hsp70 is important for import of proteins into *T. brucei* ER. MAL3-101 is a small molecule inhibitor of yeast Hsp70 ATPase [44]. In addition, the compound inhibits post-translational import of prepro alpha-factor into microsomes of *S. cerevisiae*. Based on these considerations, we reasoned that MAL3-101 could inhibit protein import into TbRM if a *T. brucei* Hsp70 was important in the process.

To determine whether MAL3-101 affected protein import into TbRM, *T. brucei* cytosol was pre-incubated with MAL3-101 or an equal volume of DMSO (diluent for MAL3-101). VSG_117_{500} mRNA was translated in a reticulocyte lysate in presence of *T. brucei* cytosol. TbRM was added post-translationally, and the mixture treated with proteinase K (see Fig. 8A for flow chart of procedures). With DMSO in the reaction, TbRM imported approximately 80% of VSG_117_{500}, as measured by protection from proteinase K digestion (Fig. 8B, compare lanes 1, 2). MAL3-101 inhibited translocation of VSG_117_{500} into TbRM (by 85%) (Fig. 8B, lanes 3, 4). We conclude that an Hsp70 chaperone in *T. brucei* is important for protein import into TbRM.

Finally, we draw attention to the fact that none of the proteins imported by TbRM was processed by N-glycosylation, because the molecular weight of the protease-protected VSGs was not different from the size of the protein synthesized in the absence of TbRM (compare lanes 1 and 3 in Fig. 8, for example). Absence of glycosylation does not mean that TbRM failed to import the proteins, because N-glycosylation is catalyzed by oligosaccharyltransferase (OST) [65] whereas import of proteins into TbRM depends on a Sec61p [5]. Clearly, TbRM is deficient in OST activity. However, our TbRM import assays meet the "gold standard" for demonstrating movement of a protein into a membrane vesicle (or organelle): (i) protease protection of the imported protein, and (ii) protease digestion of the protected substrate when detergent is added to solubilize microsomes [53, 66].

DISCUSSION

Properties of Cell-free Protein Import Into *T. brucei* ER

Endoplasmic reticulum is an important gateway for proteins, including signaling receptors and nutrient transporters, that reside within or traffick through the secretory pathway. Proteins with N-terminal ER signal peptides may be targeted to the ER co-translationally by SRP (reviewed in [67]). Alternatively, molecular chaperones Hsp70 (Ssa1p) and Hsp40 (Ydj1p) aid translocation of proteins into the ER [19, 21, 27, 63] (reviewed in [2]). (Proteins that lack an N-terminal signal peptide may also be targeted by SRP to the ER post-translationally without being imported by Sec61p complex [68]. Both co-translational and post-translational pathways utilize a Sec61 protein translocation pore at the ER membrane, and a luminal Hsp70 (BiP/Kar2p)/J-domain cochaperone (Sec63p) [4, 5, 18, 69-71]. In trypanosomes, a signal peptide is important for secretion of a model reporter protein [42]. However, SRP is dispensable for protein entry into the secretory pathway [43], leading authors of that study to propose the existence of an "alternative pathway" for protein entry into *T. brucei* ER.

T. brucei microsomes import proteins postranslationally (Fig. 1 & Fig. 3) in a signal sequence dependent manner (Fig. 6B & C). Full-length proteins require cytosol for translocation into TbRM (Fig. 5). MAL3-101 an inhibitor of J-domain stimulated Hsp70 ATPase blocked import of VSG_117_{500} into TbRM (Fig. 8), consistent with participation of *T. brucei* homologs of Hsp70 (*i.e.,* Ssa1p and BiP (Kar2p) and J-domain co-chaperones (*e.g.,* Ydj1p or Sec63p) in ER protein import. Genetic studies are needed to complement the proposed the roles of cytosolic and luminal Hsp70/Hsp40 chaperones in protein import into the parasite ER. The properties of ER protein translocation are strikingly different from those of the vertebrate host. Vertebrates do not import full-length proteins post-translationally [31], and there is no evidence that cytosolic chaperones are important for movement of proteins into the ER lumen.

Properties of protein import into TbRM are reminiscent of observations made with yeast microsomes. Yeast can import proteins post-translationally [26, 28, 72]. A soluble factor stimulated import of of α-mating factor into microsomes [73] and was later identified as an Hsp70 (Ssa1p) [19, 21, 63] which interacts specifically with the J-domain protein Hsp40 (Ydj1p to facilitate ER protein import [21, 57, 74]. Apart from *S. cerevisiae*, *T. brucei* is the only other eukaryote in which post-translational ER protein import has been demonstrated, to our knowledge.

Hydrophobicities of Signal Peptides Do Not Correlate with Import Competence

Signal sequences have three sub-regions; an N-terminal (n-region), a hydrophobic core (h-region) and a C-terminal region (c-region) reviewed in [39, 62]. In several cases, high hydrophobicity correlates with robust signal sequence activity both *in vivo* and in *vitro* [75, 76]. Since several proteins failed at import into TbRM (Fig. 7), we investigated the possibility that hydrophobicity of unimported proteins differed from those of that were imported into TbRM. Kyte-Dolittle hydropathy scale [77] was used to determine peak hydrophobicity of signal sequences (Fig. 9, solid bars).

Signal sequences from yeast α-mating factor (peak hydrophobicity 2.7), bovine preprolactin (peak hydrophobicity 2.5), *E. coli* β-lactamase (peak hydrophobicity 2.01) and *L. major* GP63 (peak hydrophobicity 2.8) are as hydrophobic as the signal peptide of VSG_117 (peak hydropohobicity 1.8) (Fig. 9). Based on hydrophobicity alone, α-mating factor, preprolactin, and β-lactamase, and Gp63 signal peptides should direct translocation into TbRM, because VSG_117 which has lower peak hydrophobicity was imported into TbRM.

We conclude that h-region hydrophobicity is not the most important feature of signal peptides that dictates import of proteins into TbRM. Other investigations have also concluded that high hydrophobicity *per se* is not sufficient for a peptide to gate an ER protein pore [33, 35].

Coevolution of Translocons and Signal Sequences

ER signal sequences can function across species lines [31, 32]. Mounting experimental evidence suggests, however, that signal peptides function uniquely or sometimes more effectively in the species of origin. Several proteins from the trypanosomatids cannot direct ER protein import in other species [34, 35, 44], and some prokaryote and yeast signal sequences cannot function in vertebrates [33, 75]. These data are consistent with the knowledge that signal sequences from different species have different compositions and lengths of the sub-domains (*i.e.,* h-region, c-region and n-region) [29, 30, 38, 39]. Further, the ER receptor for signal peptides (*i.e.,* Sec61p) is not functionally interchangeable between the yeasts *Candida albicans*, *Yarrowia lipolytica* and *S. cerevisiae* [40, 41].

Our work indicates that ribosome nascent chains bearing *T. brucei* signal peptides are targeted to the ER (Fig. 1) but cannot interact effectively with the canine translocon (Fig. 2). A *T. brucei* signal sequence only works very efficiently with an ER translocon from a trypanosome. Conversely, *T. brucei* microsomes cannot import proteins from several species (Fig. 7). In totality, these data, and the work of

others cited early, suggest that the Sec61p translocation machinery has coevolved with signal peptides in the same species.

REFERENCES

1. Walter, P. & Johnson, A. E. (1994) Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane, *Annu.Rev.Cell Biol. 10*, 87-119.

2. Zheng, N. & Gierasch, L. M. (1996) Signal sequences: the same yet different, *Cell. 86*, 849- 852.

3. Panzner, S., Dreier, L., Hartmann, E., Kostka, S. & Rapoport, T. A. (1995) Post-translational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p., *Cell. 81*, 561-70.

4. Gorlich, D., Prehn, S., Hartmann, E., Kalies, K. U. & Rapoport, T. A. (1992) A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation, *Cell. 71*, 489-503.

5. Oliver, J., Jungnickel, B., Gorlich, D., Rapoport, T. & High, S. (1995) The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum, *FEBS Lett. 362*, 126-130.

6. Corsi, A. K. & Schekman, R. (1996) Mechanism of polypeptide translocation into the endoplasmic reticulum, *J.Biol.Chem. 271*, 30299-30302.

7. Kalies, K. U., Goerlich, D. & Rapoport, T. A. (1994) Binding of ribosomes to the rough endoplasmic reticulum mediated by the Sec61p-complex, *J.Cell Biol. 126*, 925-934.

8. Song, W., Raden, D., Mandon, E. C. & Gilmore, R. (2000) Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel., *Cell. 100*, 333-43.

9. Crowley, K. S., Liao, S., Worrell, V. E., Reinhart, G. D. & Johnson, A. E. (1994) Secretory proteins move through the endoplasmic reticulum via an aqueous gated pore, *Cell. 78*, 461- 471.

10. Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004) X-ray structure of a protein-conducting channel, *Nature. 427*, 36-44.

11. Clemons, W. M., Jr., Menetret, J. F., Akey, C. W. & Rapoport, T. A. (2004) Structural insight into the protein translocation channel, *Curr Opin Struct Biol. 14*, 390-6.

12. Stirling, C. J. (1999) Protein targeting to the endoplasmic reticulum in yeast. 1997 Fleming Lecture, *Microbiology. 145 (Pt 5)*, 991-8.

13. Schnell, D. J. & Hebert, D. N. (2003) Protein translocons: multifunctional mediators of protein translocation across membranes, *Cell. 112*, 491-505.

14. Driessen, A. J., Manting, E. H. & van der Does, C. (2001) The structural basis of protein targeting and translocation in bacteria, *Nat Struct Biol. 8*, 492-8.

15. Fang, H. & Green, N. (1994) Nonlethal sec71-1 and sec72-1 mutations eliminate proteins associated with the Sec63p-BiP complex from S. cerevisiae, *Mol Biol Cell. 5*, 933-42.

16. Brizzio, V., Khalfan, W., Huddler, D., Beh, C. T., Andersen, S. S., Latterich, M. & Rose, M. D. (1999) Genetic interactions between KAR7/SEC71, KAR8/JEM1, KAR5, and KAR2 during nuclear fusion in Saccharomyces cerevisiae, *Mol Biol Cell. 10*, 609-26.

17. Wittke, S., Dunnwald, M. & Johnsson, N. (2000) Sec62p, a component of the endoplasmic reticulum protein translocation machinery, contains multiple binding sites for the Seccomplex., *Mol Biol Cell. 11*, 3859-71.

18. Young, B. P., Craven, R. A., Reid, P. J., Willer, M. & Stirling, C. J. (2001) Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo, *Embo J. 20*, 262-71.

19. Chirico, W. J., Waters, M. G. & Blobel, G. (1988) 70K heat shock related proteins stimulate protein translocation into microsomes, *Nature. 332*, 805-10.

20. Caplan, A. J., Cyr, D. M. & Douglas, M. G. (1992) YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism, *Cell. 71*,

114 3-55.

21. McClellan, A. J., Endres, J. B., Vogel, J. P., Palazzi, D., Rose, M. D. & Brodsky, J. L. (1998) Specific molecular chaperone interactions and an ATP-dependent conformational change are required during post-translational protein translocation into the yeast ER, *Mol Biol Cell. 9*, 3533-45.

22. McClellan, A. J. & Brodsky, J. L. (2000) Mutation of the ATP-binding pocket of SSA1 indicates that a functional interaction between ssa1p and ydj1p is required for post-translational translocation into the yeast endoplasmic reticulum [In Process Citation], *Genetics. 156*, 501-12.

23. Pilon, M., Romisch, K., Quach, D. & Schekman, R. (1998) Sec61p serves multiple roles in secretory precursor binding and translocation into the endoplasmic reticulum membrane [In Process Citation], *Mol Biol Cell. 9*, 3455-73.

24. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J. & Rapoport, T. A. (1998) Signal sequence recognition in post-translational protein transport across the yeast ER membrane., *Cell. 94*, 795-807.

25. Johnson, A. E. & van Waes, M. A. (1999) The translocon: a dynamic gateway at the ER membrane, *Annu Rev Cell Dev Biol. 15*, 799-842.

26. Hansen, W., Garcia, P. D. & Walter, P. (19_{86}) In vitro protein translocation across the yeast endoplasmic reticulum: ATP-dependent post-translational translocation of the prepro-alphafactor, *Cell. 45*, 397-406.

27. Brodsky, J. L. (1996) Post-translational protein translocation: not all hsc70s are created equal, *Trends Biochem Sci. 21*, 122-6.

28. Brodsky, J. L., Hamamoto, S., Feldheim, D. & Schekman, R. (1993) Reconstitution of protein translocation from solubilized yeast membranes reveals topologically distinct roles for BiP and cytosolic Hsc70, *J.Cell Biol. 120*, 95-102.

29. Von Heijne, G. (1985) Signal sequences. The limits of variation, *J.Mol.Biol. 184*, 99-105.

30. Rusch, S. L. & Kendall, D. A. (1995) Protein transport via amino-terminal targeting sequences: common themes in diverse systems, *Mol.Membr.Biol. 12*, 295-307.

31. Garcia, P. D. & Walter, P. (1988) Full-length prepro-alpha-factor can be translocated across the mammalian microsomal membrane only if translation has not terminated, *J.Cell Biol. 106*, 1043-1048.

32. Wiedmann, M., Huth, A. & Rapoport, T. A. (1984) Xenopus oocytes can secrete bacterial beta-lactamase, *Nature. 309*, 637-639.

33. Zheng, T. & Nicchitta, C. V. (1999) Structural determinants for signal sequence function in

the mammalian endoplasmic reticulum., *J Biol Chem. 274*, 36623-30.

34. Ramirez, M. I., Boscardin, S. B., Han, S. W., Paranhos-Baccala, G., Yoshida, N., Kelly, J. M., Mortara, R. A. & Da Silveira, J. F. (1999) Heterologous expression of a Trypanosoma cruzi surface glycoprotein (gp82) in mammalian cells indicates the existence of different signal sequence requirements and processing., *J Eukaryot Microbiol. 46*, 557-65.

35. Al-Qahtani, A., Teilhet, M. & Mensa-Wilmot, K. (1998) Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from Trypanosoma brucei and Leishmania, *Biochem.J. 331*, 521-529.

36. Laforet, G. A., Kaiser, E. T. & Kendall, D. A. (1989) Signal peptide subsegments are not always functionally interchangeable. M13 procoat hydrophobic core fails to transport alkaline phosphatase in Escherichia coli, *J.Biol.Chem. 264*, 14478-14485.

37. Von Heijne, G. (1982) Signal sequences are not uniformly hydrophobic, *J.Mol.Biol. 159*, 537-541.

38. Von Heijne, G. & Abrahmsen, L. (1989) Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts, *FEBS Lett. 244*, 439-446.

39. Izard, J. W. & Kendall, D. A. (1994) Signal peptides: exquisitely designed transport promoters, *Mol.Microbiol. 13*, 765-773.

40. De La Rosa, J. M., Ruiz, T., Fonzi, W. A. & Rodriguez, L. (2004) Analysis of heterologous expression of Candida albicans SEC61 gene reveals differences in Sec61p homologues related to species-specific functionality, *Fungal Genet Biol. 41*, 941-53.

41. Broughton, J., Swennen, D., Wilkinson, B., Joyet, P., Gaillardin, C. & Stirling, C. (1997) Cloning of SEC61 homologues from Schizosaccharomyces pombe and Yarrowia lipolytica reveals the extent of functional conservation within this core component of the ER translocation machinery, *J.Cell Sci. 110*, 2715-2727.

42. Bangs, J. D., Brouch, E. M., Ransom, D. M. & Roggy, J. L. (1996) A soluble secretory reporter system in Trypanosoma brucei - Studies on endoplasmic reticulum targeting, *J.Biol.Chem. 271*, 18387-18393.

43. Liu, L., Liang, X. H., Uliel, S., Unger, R., Ullu, E. & Michaeli, S. (2002) RNA interference of signal peptide-binding protein SRP54 elicits deleterious effects and protein sorting defects in trypanosomes, *J Biol Chem. 277*, 47348-57.

44. Button, L. L., Wilson, G., Astell, C. R. & McMaster, W. R. (1993) Recombinant Leishmania surface glycoprotein GP63 is secreted in the baculovirus expression system as a latent metalloproteinase, *Gene. 134*, 75-81.

45. Martoglio, B. & Dobberstein, B. (1996) Snapshots of membrane-translocating proteins, *Trends.Cell Biol. 6*, 142-147.

46. Hansen, W. & Walter, P. (1988) Prepro-carboxypeptidase Y and a truncated form of preinvertase, but not full-length pre-invertase, can be post-translationally translocated across microsomal vesicle membranes from Saccharomyces cerevisiae, *J Cell Biol. 106*, 1075-81.

47. Al-Qahtani, A. & Mensa-Wilmot, K. (1996) A 5' untranslated region which directs accurate and robust translation by prokaryotic and mammalian ribosomes, *Nucleic Acids Res. 24*,
1173-1174.

48. Gilmore, R., Collins, P., Johnson, J., Kellaris, K. & Rapiejko, P. (1991) Transcription of full-length and truncated mRNA transcripts to study protein translocation across the endoplasmic reticulum., *Methods Cell Biol. 34*, 223-39.

49. Schagger, H. & von Jagow, G. (1991 Dec) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form., *Anal Biochem. 199*, 223-31.

50. Lauring, B., Sakai, H., Kreibich, G. & Wiedmann, M. (1995) Nascent polypeptideassociated complex protein prevents mistargeting of nascent chains to the endoplasmic reticulum [published erratum appears in Proc Natl Acad Sci U S A 1995 Aug 15; 92(17):8088], *Proc.Natl.Acd.Sci.U.S.A. 92*, 5411-5415.

51. Jungnickel, B. & Rapoport, T. A. (1995) A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane, *Cell. 82*, 261-270.

52. Belin, D., Bost, S., Vassalli, J. D. & Strub, K. (1996) A two-step recognition of signal sequences determines the translocation efficiency of proteins., *EMBO J. 15*, 468-78.

53. Connolly, T., Collins, P. & Gilmore, R. (1989) Access of proteinase K to partially translocated nascent polypeptides in intact and detergent-solubilized membranes., *J Cell Biol. 108*, 299-307.

54. Connolly, T. & Gilmore, R. (19_{86}) Formation of a functional ribosome-membrane junction during translocation requires the participation of a GTP-binding protein, *J.Cell Biol. 103*,
2253-2261.

55. Mothes, W., Jungnickel, B., Brunner, J. & Rapoport, T. A. (1998) Signal sequence recognition in co-translational translocation by protein components of the endoplasmic reticulum membrane, *J.Cell Biol. 142*, 355-364.

56. Neuhof, A., Rolls, M. M., Jungnickel, B., Kalies, K. U. & Rapoport, T. A. (1998) Binding of signal recognition particle gives ribosome/nascent chain complexes a competitive advantage in endoplasmic reticulum membrane interaction, *Mol.Biol.Cell. 9*, 103-115.

57. Becker, J., Walter, W., Yan, W. & Craig, E. A. (1996) Functional interaction of cytosolic hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation in vivo., *Mol Cell Biol. 16*, 4378-$_{86}$.

58. Kim, J., Lee, Y., Kim, C. & Park, C. (1992 Aug) Involvement of SecB, a chaperone, in the export of ribose-binding protein., *J Bacteriol. 174*, 5219-27.

59. Kim, J. & Kendall, D. A. (2000) Sec-dependent protein export and the involvement of the molecular chaperone SecB, *Cell Stress Chaperones. 5*, 267-75.

60. Knoblauch, N. T., Rudiger, S., Schonfeld, H. J., Driessen, A. J., Schneider-Mergener, J. & Bukau, B. (1999) Substrate specificity of the SecB chaperone., *J Biol Chem. 274*, 34219-25.

61. Wilkinson, B. M., Regnacq, M. & Stirling, C. J. (1997) Protein translocation across th membrane of the endoplasmic reticulum, *J.Membr.Biol. 155*, 189-197.

62. Martoglio, B. & Dobberstein, B. (1998) Signal sequences: more than just greasy peptides., *Trends Cell Biol. 8*, 410-5.

63. Ngosuwan, J., Wang, N. M., Fung, K. L. & Chirico, W. J. (2003) Roles of cytosolic Hsp70 and Hsp40 molecular chaperones in post-translational translocation of presecretory proteins into the endoplasmic reticulum, *J Biol Chem. 278*, 7034-42.

64. Lyman, S. K. & Schekman, R. (1996) Polypeptide translocation machinery of the yeast endoplasmic reticulum, *Experientia. 52*, 1042-1049.

65. Te Heesen, S., Knauer, R., Lehle, L. & Aebi, M. (1993) Yeast Wbp1p and Swp1p form a protein complex essential for oligosaccharyl transferase activity, *EMBO J. 12*, 279-284.

66. Walter, P. & Blobel, G. (1983) Preparation of microsomal membranes for co-translational protein translocation, *Methods Enzymol. 96*, 84-93.

67. Lutcke, H. (1995) Signal recognition particle (SRP), a ubiquitous initiator of protein translocation, *Eur.J.Biochem. 228*, 531-550.

68. Abell, B. M., Pool, M. R., Schlenker, O., Sinning, I. & High, S. (2004) Signal recognition particle mediates post-translational targeting in eukaryotes, *Embo J. 23*, 2755-64.

69. Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain., *Cell. 102*, 233-44.

70. Willer, M., Jermy, A. J., Steel, G. J., Garside, H. J., Carter, S. & Stirling, C. J. (2003) An in vitro assay using overexpressed yeast SRP demonstrates that co-translational translocation is dependent upon the J-domain of Sec63p, *Biochemistry. 42*, 7171-7.

71. Wittke, S., Lewke, N., Muller, S. & Johnsson, N. (1999) Probing the molecular environment of membrane proteins in vivo., *Mol Biol Cell. 10*, 2519-30.

72. Brodsky, J. L., Goeckeler, J. & Schekman, R. (1995) BiP and Sec63p are required for both co- and post-translational protein translocation into the yeast endoplasmic reticulum, *Proc Natl Acad Sci U S A. 92*, 9643-6.

73. Waters, M. G., Chirico, W. J. & Blobel, G. (19_{86}) Protein translocation across the yeast microsomal membrane is stimulated by a soluble factor, *J Cell Biol. 103*, 2629-36.

74. Cyr, D. M. & Douglas, M. G. (1994) Differential regulation of Hsp70 subfamilies by the eukaryotic DnaJ homologue YDJ1, *J Biol Chem. 269*, 9798-804.

75. Bird, P., Gething, M. J. & Sambrook, J. (1987) Translocation in yeast and mammalian cells: not all signal sequences are functionally equivalent, *J.Cell Biol. 105*, 2905-2914.

76. McGeoch, D. J. (1985) On the predictive recognition of signal peptide sequences, *Virus.Res. 3*, 271-2_{86}.

77. Kyte, J. & Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J.Mol.Biol. 157*, 105-132.

FIGURE LEGEND

Figure 1 VSG_117 ribosome nascent chain docks at the canine ER

VSG_117_{86} and pPL_{86} mRNA's were translated in rabbit reticulocyte lysate in the absence or presence of 4 equivalents of canine microsomes (CfRM). Ribosome nascent chains (RNCs) were floated through a 1.2M sucrose cushion, and the top (T) and bottom (B) fractions collected. Proteins were resolved by SDS-PAGE and visualized by fluorography. Abbreviations used are: VSG, variant surface glycoprotein; CfRM, canine microsomes; RNC, ribosome nascent chain; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis. pPL RNCs; lanes 1 & 2, RNCs collected from top and bottom fraction in absence of CfRM; lanes 3 & 4, RNCs collected from top and bottom fraction in presence of CfRM. VSG_117 RNCs; Lanes 1 & 2, RNCs obtained from top and bottom fraction in absence of CfRM; lanes 3 & 4, RNCs collected from top and bottom fraction in presence of CfRM.

Figure 2 Interaction of VSG_117 signal peptide with canine translocon

RNCs of pPL and VSG_117 were produced in rabbit reticulocyte lysate in the presence of increasing concentrations of canine microsomes (CfRM). Thereafter, the RNCs were exposed to proteinase K (120 !g/ml, final concentration). Radiolabeled proteins were detected by fluorography after proteins had been resolved by SDS-PAGE. (***A***) pPL RNCs; (***B***) VSG_117 RNCs. Lane 1, untreated translation product; lane 2, translation product treated with proteinase K; lane 3 & 4, translation products with 2 equivalents of CfRM without (3) or with (4) proteinase K; lane 5 & 6, translation products with 4 equivalents of CfRM without (5) or with (6) proteinase K. (***C***) Quantitation of the data (by laser scanning densitometry) from panels A and B. Protease protection is scored as a percentage of full-length nascent chain resistant to proteinase K digestion.

Figure 3 Cell-free protein Import by *T. brucei* ER

(*A*) Flowchart of protocol for post-translational import of VSG_117 $_{86}$. Depicted are the various steps and temperatures at which reactions took place. The translation product for VSG_117$_{86}$ is presented along with [14C]methylated protein markers (Amersham). ***(B*) Import of VSG_117$_{86}$ into TbRM**. VSG_117$_{86}$ mRNA was translated in rabbit reticulocyte lysate for 15 min, and treated with cycloheximide (50 !g/ml, final concentration). Reaction mixtures were incubated with TbRM (one equivalent) for 45 min (37°C) followed by proteinase K digestion (300 !g/ml, final concentration) for 60 min (on ice). Proteins were resolved by SDS-PAGE and detected by phosphorimaging. Lane 1, untreated VSG_117$_{86}$; lane 2, VSG_117$_{86}$ treated with proteinase K; lane 3, VSG_117$_{86}$ with TbRM; lane 4, VSG_117$_{86}$ with TbRM treated with proteinase K; lane 5, VSG_117$_{86}$ with TbRM permeabilized with 2% NP40 during proteinase K digestion. Rectangular brackets underneath sets of bars denote those data points that were directly compare in quantitation. The values are normalized relative to the first lane. Abbreviations: RRL, rabbit reticulocyte lysate; CHX, cycloheximide; TbRM, *T. brucei* microsomes; PK, proteinase K; PMSF, phenyl methyl sulfonyl fluoride; NP-40, nonidet P-40.

Figure 4 Effect of VSG Length on Extent of Import by TbRM

(A*)** Outline of Experimental Protocol Used in Analysis of Import of VSG_117 $_{114}$ Into TbRM. Translation product for VSG_117$_{114}$ is presented. ***(B*)** VSG_117$_{114}$. VSG_117$_{114}$ mRNA was translated in a reticulocyte lysate for 15 min. After treatment with cycloheximide (50 !g/ml, final concentration), reaction mixtures were incubated with TbRM (one equivalent) for 45 min (37°C) followed by proteinase K digestion (300 !g/ml, final concentration) for 60 min (on ice). Proteins were resolved by SDS-PAGE and detected by phosphorimaging. *Lane* 1, untreated VSG_117$_{114}$; *lane* 2, VSG_117$_{114}$ treated with proteinase K; *lane* 3, VSG_117$_{114}$ with TbRM; *lane* 4, VSG_117$_{114}$ with TbRM treated with proteinase K. ***(C) Flowchart of protocol for post-translational import of VSG_117$_{260}$ & VSG_117$_{500}$.. Translation products for VSG_117$_{260}$ & VSG_117$_{500}$. are presented. ***(D)*** Different mRNAs were translated in rabbit reticulocyte lysate for 60 min, and treated with cycloheximide (50 !g/ml, final concentration). Reaction

mixtures were incubated with TbRM (one equivalent) for 60 min at 37°C, and treated with proteinase K (120 !g/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS-PAGE and radiolabeled polypeptides were visualized by phosphorimaging. (***E)*** Effect of Protein Length oon Import into TbRM. Efficiency of protein import into TbRM measured as the percentage of fulllength polypeptide protected from proteinase K digestion. Quantitation of bands visualized with the phosphorimager were performed Quantity One software (BioRad). *Lanes 1-4*: VSG_117$_{260}$. *Lane 1*, untreated VSG_117$_{260}$; *lane 2*, VSG_117$_{260}$ treated with proteinase K; *lane 3*, VSG_117$_{260}$ with TbRM; *lane 4*, VSG_117$_{260}$ with TbRM treated with proteinase K. *Lanes 5 8*: VSG_117$_{500}$. *Lane 5*, untreated VSG_117$_{500}$; *lane 6*, VSG_117$_{500}$ treated with proteinase K; *lane 7*, VSG_117$_{500}$ with TbRM; *lane 8,* VSG_117$_{500}$ with TbRM treated with proteinase K.

Figure 5 Cell-free system for full-length protein import into ER of *T. brucei*

(A) Flowchart of Experimental Protocol. ***(B-D)*** Different mRNAs encoding VSG_117 were translated in rabbit reticulocyte lysate (with or without) *T. brucei* cytosol (1.5 equivalents) for 60 min, and treated with cycloheximide (50 !g/ml, final concentration). Reaction mixtures were incubated with TbRM (one equivalent) for 60 min at 37°C, and digested with proteinase K digestion (120 !g/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS PAGE and detected by phosphorimaging. ***(B)*** **VSG_117$_{260}$**. Lanes 1-4 (without cytosol): *Lane 1*, untreated VSG_117$_{260}$; *lane 2*, VSG_117$_{260}$ digested with proteinase K; *lane 3*, VSG_117$_{260}$ incubated with TbRM; *lane 4,* VSG_117$_{260}$ with TbRM & treated with proteinase K. Lanes 5-8 (with cytosol): *Lane 5*, untreated VSG_117$_{260}$; *lane 6*, VSG_117$_{260}$ digested with proteinase K; lane 7, VSG_117$_{260}$ with TbRM; lane 8, VSG_117$_{260}$ with TbRM & treated with proteinase K. ***(C)*** **VSG_117$_{500}$**. Lanes 1-3 (without TbRM): *Lane 1*, untreated VSG_117$_{500}$; *lane 2*, VSG_117$_{500}$ treated with proteinase K; *lane 3*, VSG_117$_{260}$ with proteinase K after NP-40 permeabilization of TbRM. Lanes 4-6 (with TbRM): *Lane 4*, untreated VSG_117$_{500}$; *lane 5*, VSG_117$_{500}$ cleaved with proteinase K; *lane 6*, VSG_117$_{260}$ digested with proteinase K after NP- 40 permeabilization of TbRM. ***(D)*** **VSG_MVAT7**. Lanes 1 & 2 (without TbRM): *Lane 1*, untreated VSG_MVAT7; *lane 2*,

VSG_MVAT7 digested with proteinase K. Lanes 3-5 (with TbRM): *Lane 3*, untreated VSG_MVAT7; *lane 4*, VSG_MVAT7 cleaved with proteinase K; *lane 5*, VSG_MVAT7 digested with proteinase K following NP-40 permeabilization of TbRM.

Figure 6 Signal peptide dependence of VSG import into TbRM

***(A)* Flowcharts** of Experimental Protocols used for Testing Import of VSG_117_{86}_SP (without cytosol) and VSG_117_{500}_SP (with cytosol) by TbRM. The chart depicts various steps and temperatures at which reactions took place. ***(B)* VSG_117_{86}_SP.** mRNA encoding VSG_117_{86}_SP was translated in rabbit reticulocyte lysate for 15 min, and treated with cycloheximide (50 !g/ml, final concentration). Reaction mixtures were incubated with TbRM (one equivalent) for 45 minutes at 37°C, and digested with proteinase K (300 !g/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS-PAGE and detected by phosphorimaging. The abbreviation SP = signal peptide. *Lane 1*, untreated VSG_117_{86}_SP; *lane 2*, VSG_117_{86}_SP cleaved with proteinase K; *lane 3*, VSG_117_{86}_SP with TbRM; *lane 4,* VSG_117_{86}_SP incubated with TbRM and digested with proteinase K. ***(C)* VSG_117_{500}_SP.** VSG_117_{500}_SP mRNA was translated in a rabbit reticulocyte lysate with *T. brucei* cytosol (1.5 equivalents) for 60 min, and treated with cycloheximide (50 !g/ml, final concentration). Reaction mixtures were incubated with TbRM (one equivalent) for 60 min at 37°}C, and digested with proteinase K (120 !g/ml, final concentration) for 60 min on ice. Proteins were separated by SDS-PAGE and radiolabeled polypeptides detected by phosphorimaging. *Lane 1*, untreated VSG_117_{500}_SP; *lane 2*, VSG_117_{500}_SP treated with proteinase K; *lane 3*, VSG_117_{500}_SP with TbRM; *lane 4*, VSG_117_{500}_SP with TbRM treated with proteinase K.

Figure 7 Signal peptide specificity of TbRM

***(A)* Flowchart**. The flowchart depicts steps and temperatures at which reactions took place. ***(B)* Import of Proteins from Different Biological Families into TbRM.** Different mRNAs were translated in rabbit reticulocyte lysate with *T. brucei* cytosol (1.5 equivalents) for 60 min. Cycloheximide (50 !g/ml, final

concentration) was added, and reaction mixtures were incubated with TbRM (one equivalent) for 60 min at 37°C. Proteinase K digestion (120 !g/ml, final concentration; 60 min on ice) was then performed. Proteins were resolved by SDS-PAGE, and radiolabeled polypeptides visualized by phosphorimaging. *Lane 1*, untreated protein; *lane 2*, protein treated with proteinase K; *lane 3*, protein with TbRM; *lane 4*, protein incubated with TbRM and then treated with proteinase K.

Figure 8 Effect of an Inhibitor of Hsp70 on VSG_117$_{500}$ Import into TbRM

(A) Protocol Used for Import of VSG_117$_{500}$. ***(B)*** VSG_117$_{500}$ mRNA was translated in rabbit reticulocyte lysate with *T. brucei* cytosol (1.5 equivalents) for 60 min, and treated with Reaction mixtures were incubated with TbRM (one equivalent) that had been treated or not incubated with MAL3-101 (0.3 !M) or equal volume of DMSO (diluent for MAL3-101). The mixture was incubated for 60 min at 37^{0}C and digested with proteinase K digestion (30 !g/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS-PAGE and detected by phosphorimaging. *Lanes* 1 & 2 (TbRM pretreated with DMSO): *Lane 1*, untreated VSG_117$_{500}$ with TbRM; *lane 2*, VSG_117$_{500}$ with TbRM & treated with proteinase K. *Lanes 3 & 4* (TbRM pretreated with MAL3-101): *Lane 3*, VSG_117$_{500}$ incubated with TbRM; *lane 4*, VSG_117$_{500}$ incubated with TbRM and treated with proteinase K.

Figure 9 Correlation of Signal Sequence Hydrophobicity and Protein Import into TbRM

Efficiency of protein import into TbRM was scored as the proportion of full-length polypeptide protected from with proteinase K in the presence of TbRM (Data is obtained from Fig. 3 and Fig. 7). Kyte-Dolittle hydropathy scale was used to calculate peak hydrophobicity of signal peptides using a window of 7 amino acids.

Figure 1:

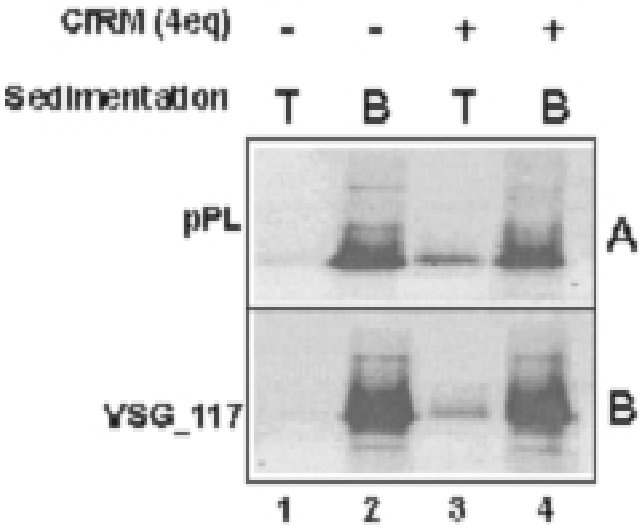
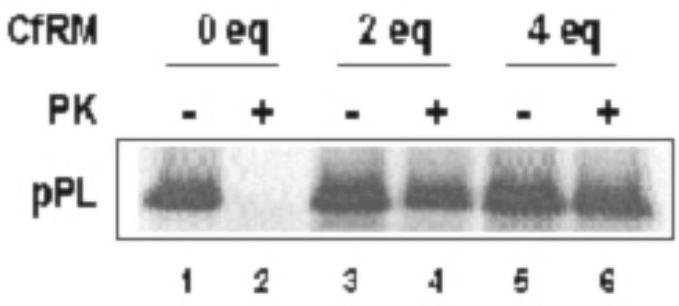
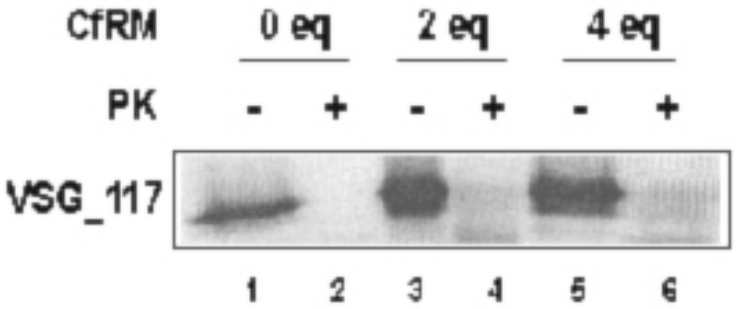


Figure 2:

A


CfRM
0 eq
2 eq
4 eq
PK
- + - + - +
pPL
1 2 3 4 5 6

B


CfRM
0 eq
2 eq
4 eq
PK
- + - + - +
VSG_117
1 2 3 4 5 6

C

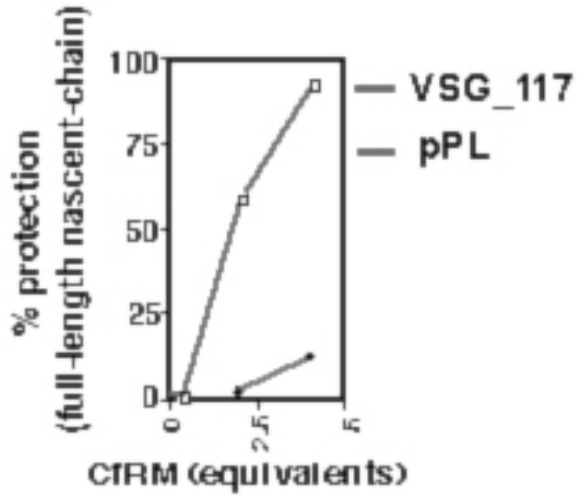
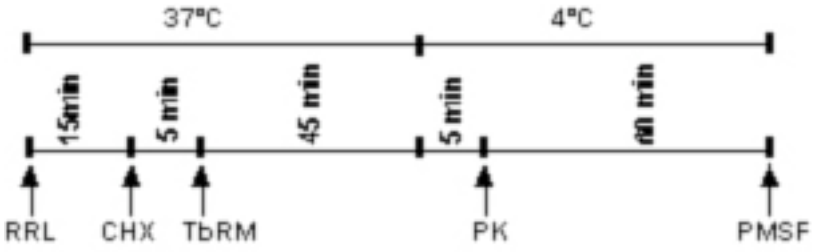

% protection
(full-length nascent-chain)
100
75
50
25
0
VSG_117
pPL
CfRM (equivalents)

Figure 3

A


37°C
4°C
15min
5 min
45 min
5 min
60 min
RRL
CHX
TbRM
PK
PMSF

B

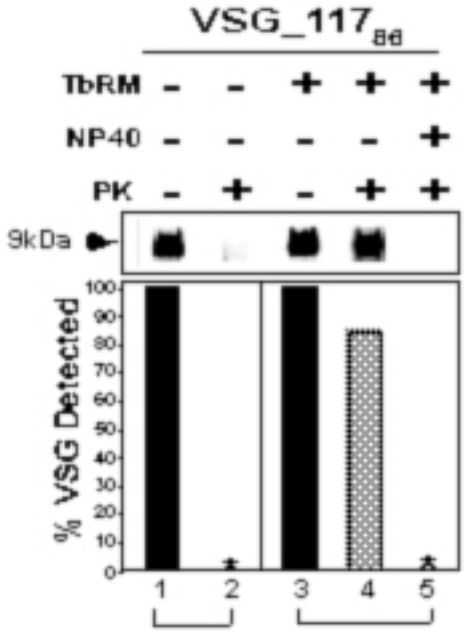
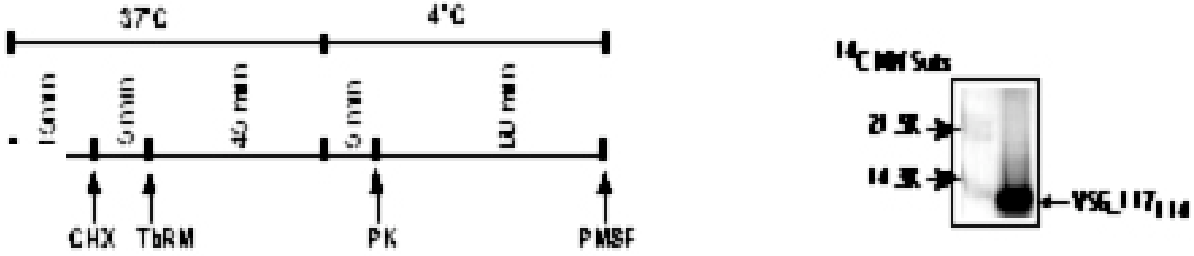
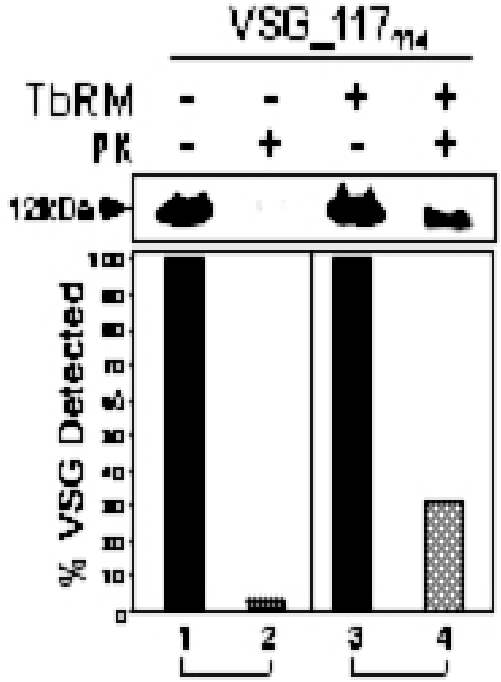

VSG_117$_{8d}$
TbRM - - + + +
NP40 - - - - +
PK - + - + +
9kDa
% VSG Detected
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5

Figure 4:

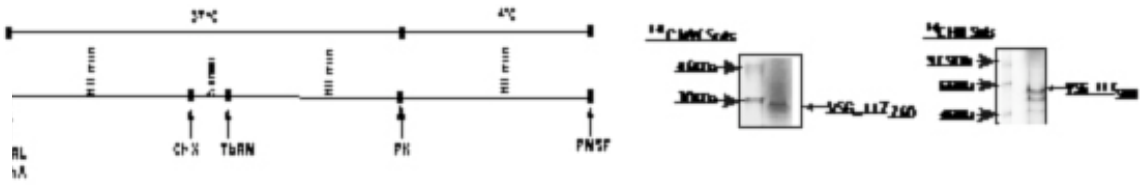
A


37°C
4°C
CHX
TbRM
PK
PMSF
VSG_117 114

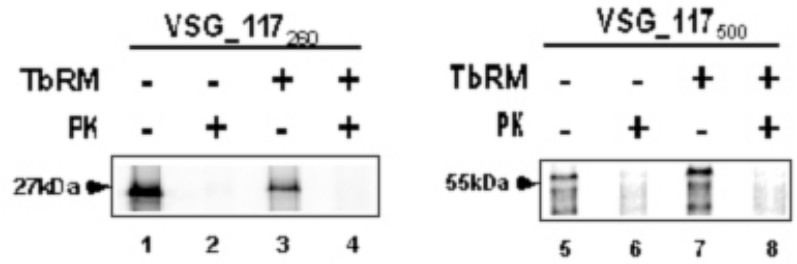
B


VSG_117 114
TbRM - - + +
PK - + - +
12kDa
% VSG Detected
1 2 3 4

C



D


VSG_117 260
TbRM - - + +
PK - + - +
27kDa
1 2 3 4
VSG_117 500
TbRM - - + +
PK - + - +
55kDa
5 6 7 8

E

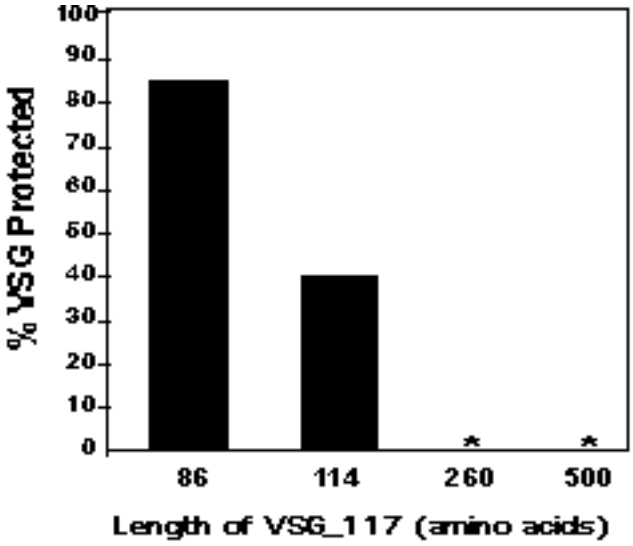
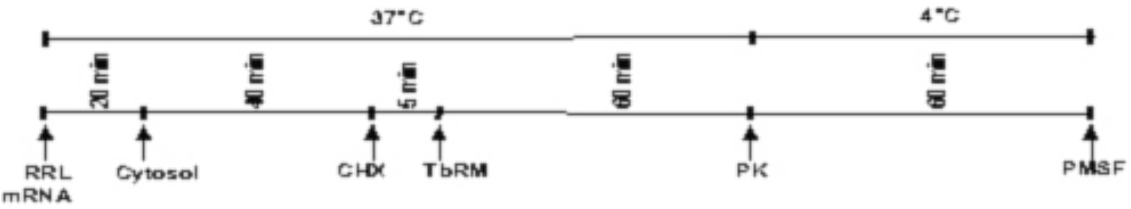
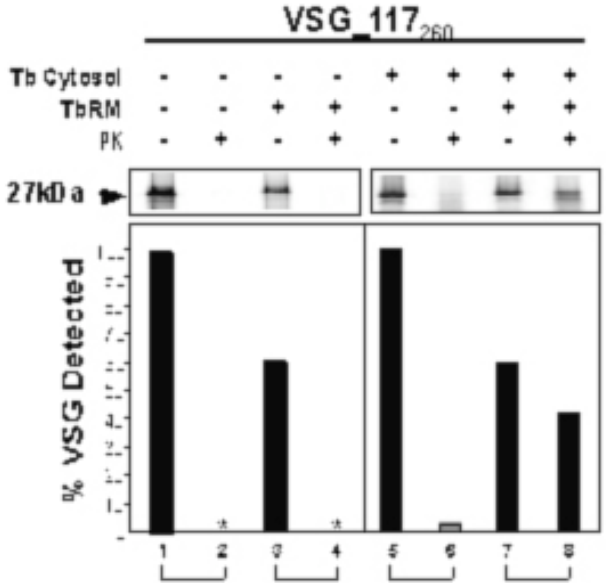

% VSG Protected
100
90
80
70
60
50
40
30
20
10
0
*
*
86
114
260
500
Length of VSG_117 (amino acids)

Figure 5

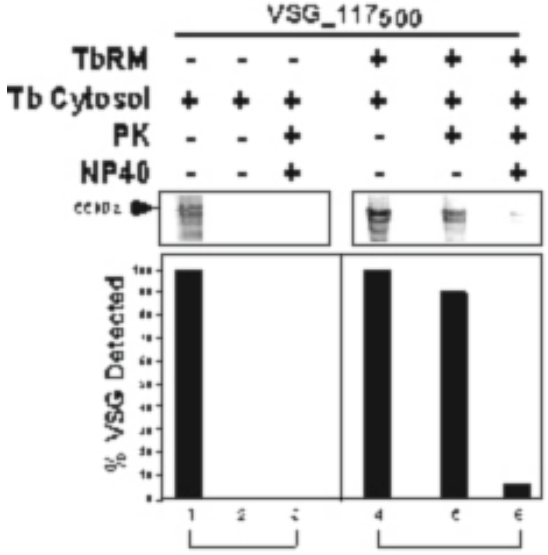
A



B



C



D

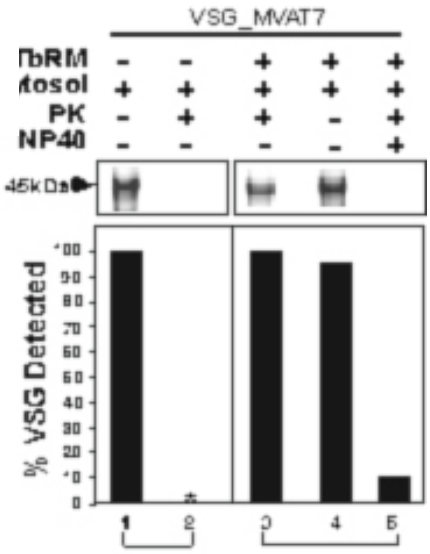
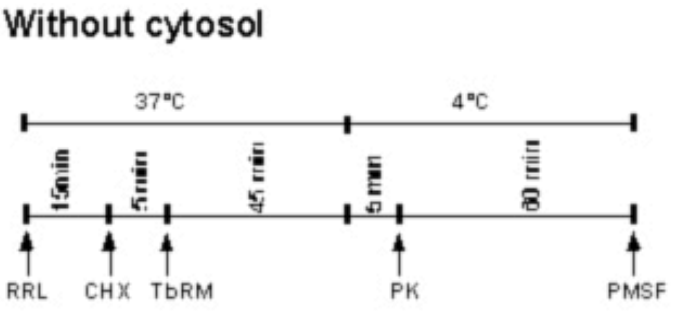
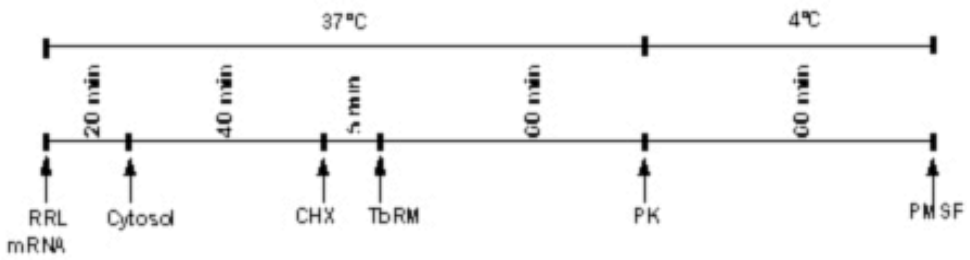


Figure 6:

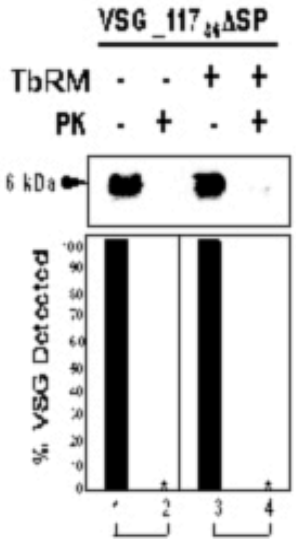
A



B



C



D

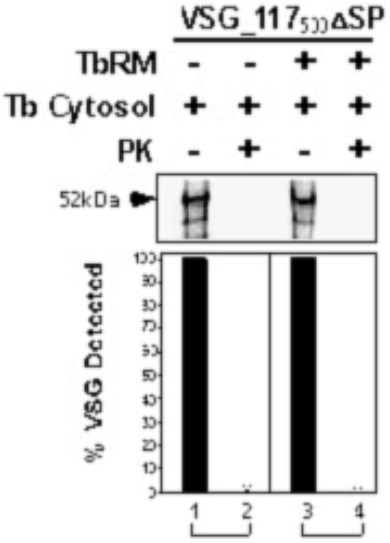
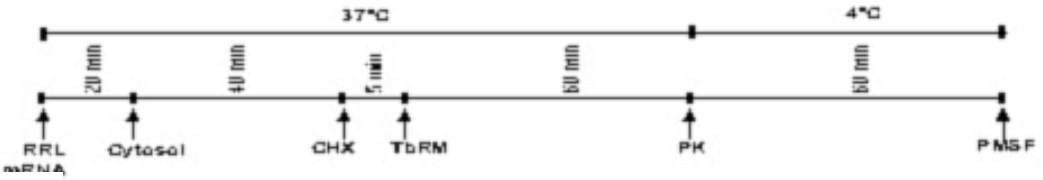


Figure 7:

A


37°C
4°C
20 min
40 min
5 min
60 min
60 min
RRL
mRNA
Cytosol
CHX
TbRM
PK
PMSF

B

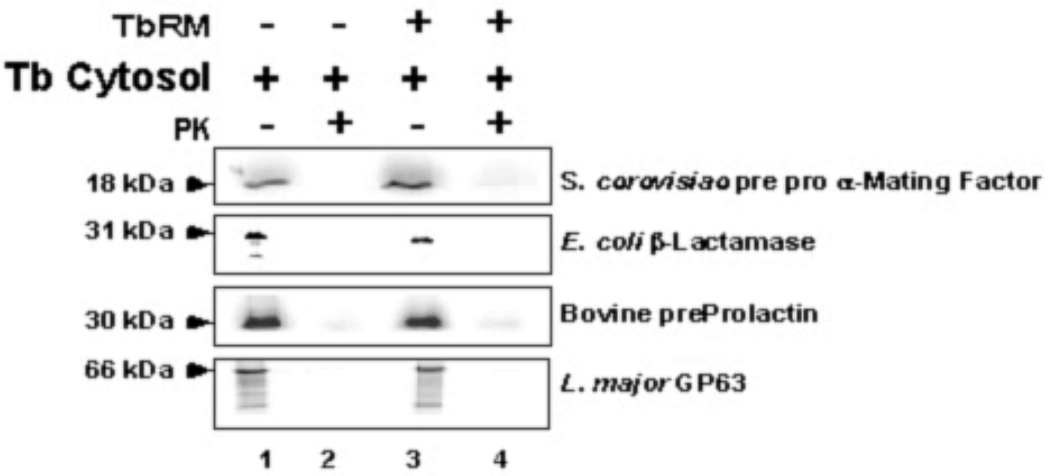
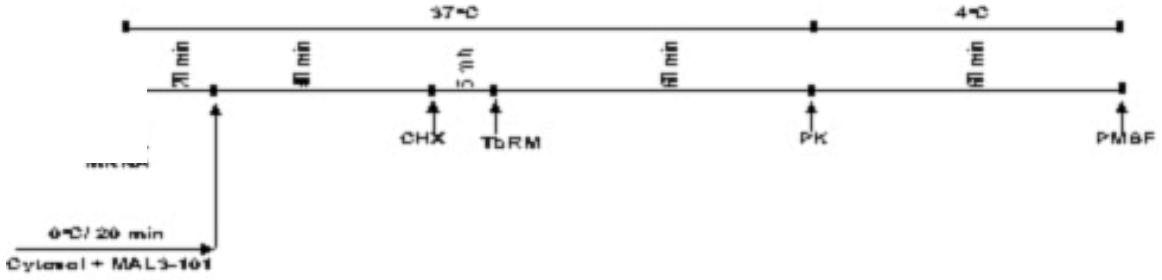

TbRM - - + +
Tb Cytosol + + + +
PK - + - +
18 kDa
S. cerevisiae pre pro α-Mating Factor
31 kDa
E. coli β-Lactamase
30 kDa
Bovine preProlactin
66 kDa
L. major GP63
1 2 3 4

Figure 8

A


37°C
4°C
CHX
TbRM
PK
PMSF
0°C/ 20 min
Cytosol + MAL3-101

B

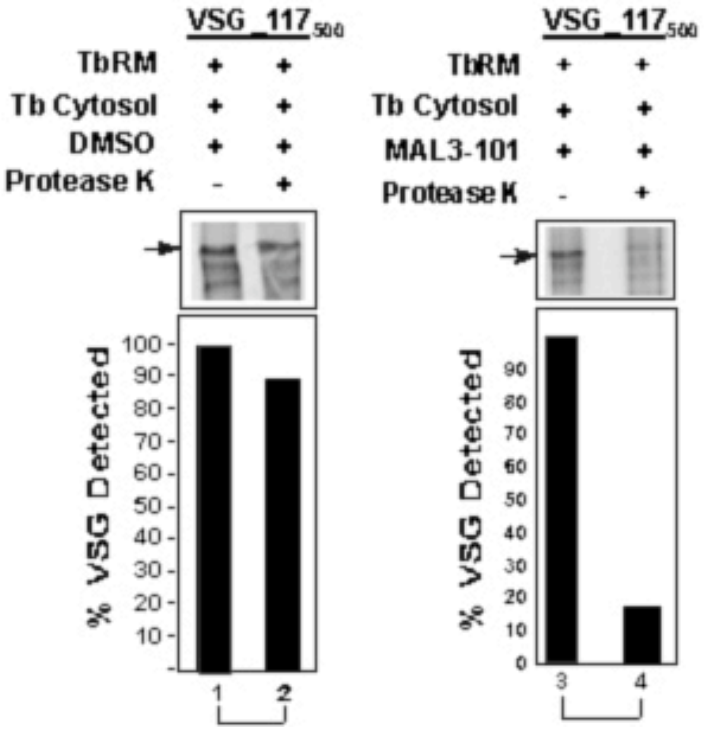
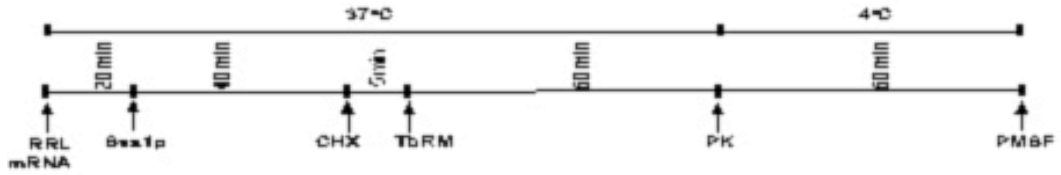

VSG_117500
TbRM + +
Tb Cytosol + +
DMSO + +
Protease K - +
% VSG Detected
100 90 80 70 60 50 40 30 20 10
1 2
VSG_117500
TbRM + +
Tb Cytosol + +
MAL3-101 + +
Protease K - +
% VSG Detected
90 80 70 60 50 40 30 20 10 0
3 4

Figure 9:

A


37°C
4°C
20 min
40 min
5 min
60 min
60 min
RRL mRNA
Ssa1p
CHX
TbRM
PK
PMSF

B

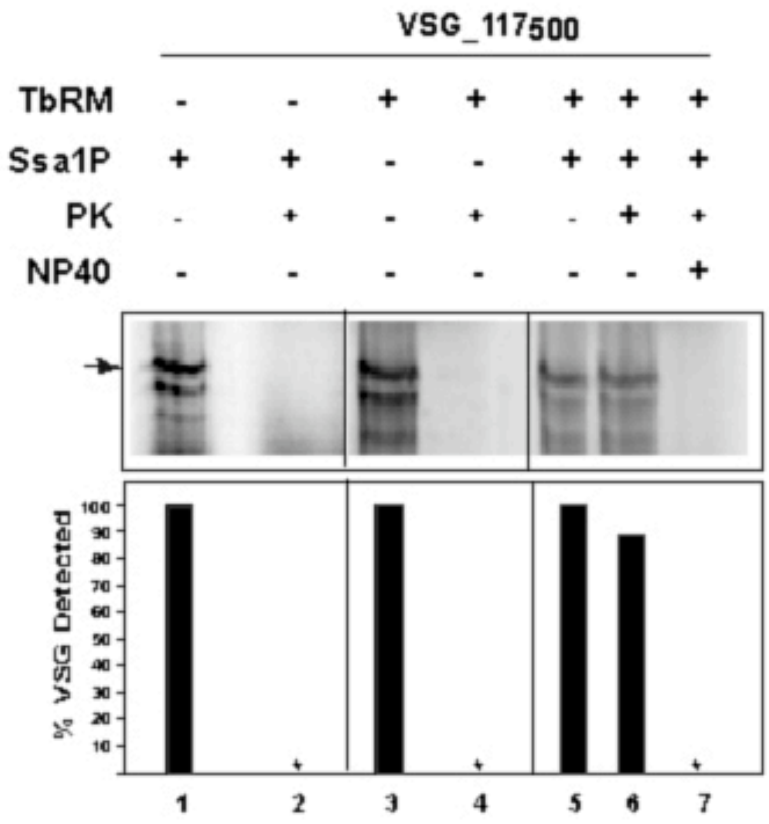
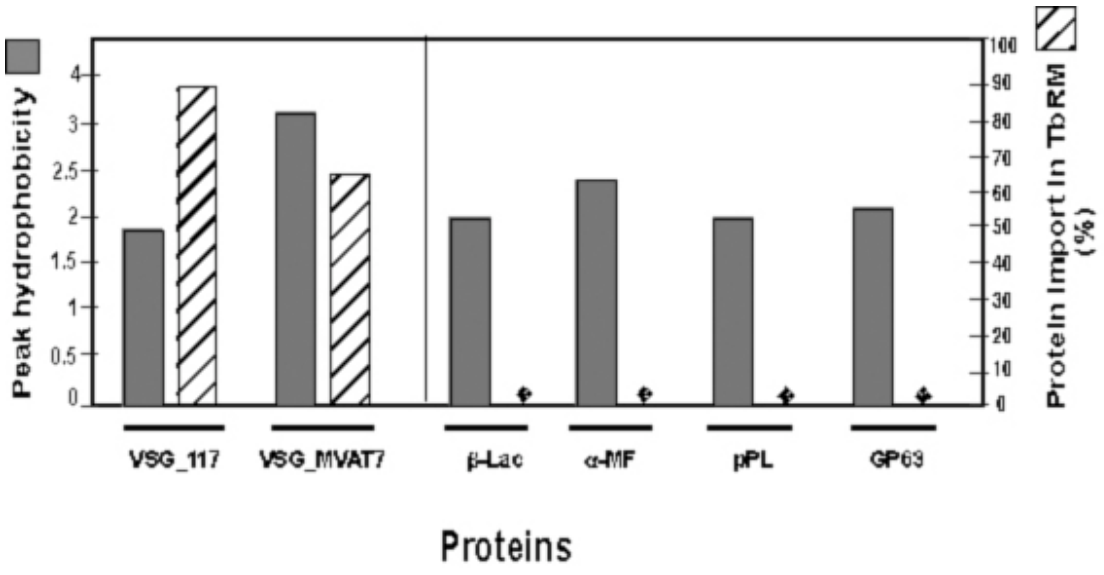

VSG_117500
TbRM - - + + + + +
Ssa1P + + - - + + +
PK - + - + - + +
NP40 - - - - - - +
% VSG Detected
100
90
80
70
60
50
40
30
20
10
1 2 3 4 5 6 7

Figure 10:


Peak hydrophobicity
Protein Import in TbRM (%)
VSG_117
VSG_MVAT7
β-Lac
α-MF
pPL
GP63
Proteins

CHAPTER 3

"PROTEIN TRANSLOCATION INTO ENDOPLASMIC RETICULUM" AS A TARGET FOR ANTI-PARASITE DRUG DISCOVERY: *TRYPANOSOMA BRUCEI* [1]

[1] Patham, B., Davis, C. R., Brodsky, J. L., Wipf, P., and Mensa-Wilmot, K.
To be submitted to *Journal of Biological Chemistry*, 05/2005

LIST OF ABBREVIATIONS

CHX, cycloheximide; DMSO, dimethyl sulfoxide; ER, endoplasmic reticulum; HAT, human African trypanosomiasis; Hsp, heat shock protein; PK, proteinase K; PMSF, phenylmethyl sulfonylfluoride; PTB, protein translocation blocker; RRL, rabbit reticulocyte lysate; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TbRM, *T. brucei* rough microsomes; VSG, variant surface glycoprotein

ABSTRACT

Caused by *Trypanosoma brucei*, human African trypanosomiasis (HAT) is an emerging disease for which new drugs are needed. Cell surface proteins are important for viability of *T. brucei*. Transport of the majority of these proteins to the plasma membrane involves transit through the endoplasmic reticulum (ER). We recently established a cell-free *in vitro* system for studying protein import into the ER of *T. brucei*. Protein translocation into *T. brucei microsomes* (TbRM) was post-translational and dependent on parasite cytosol, which could be replaced by purified chaperone Ssa1p from *Saccharomyces cerevisiae*. In the present study, we have used the *in vitro* system to identify compounds that block import of proteins into the *T. brucei* ER. Protein translocation blockers were then tested for anti-parasite activity. MAL3-101, an inhibitor of Hsp40 regulated Hsp70 ATPase, blocked import of variant surface glycoprotein (VSG) into TbRM. Two novel inhibitors of TbRM protein import namely, equisetin and CJ-21,058 were identified. In tests of trypanocidal activity, the concentration at which fifty percent of parasites were killed (IC_{50}) are as follows; 125 nM (MAL3-101), 3.3 μM (equisetin) and 7 μM (CJ-21,058). None of the compounds had a discernible effect on human HeLa cells at concentrations that killed *T. brucei*. We propose that MAL3-101, equisetin, and CJ-21,058 are lead compounds for anti-trypanosome drug development. These studies establish the trypanosome secretory system as a suitable target for drug discovery.

INTRODUCTION

Human African trypanosomiasis (HAT) occurs in 36 countries in sub-Saharan Africa and threatens an estimated 60 million people [1]. No vaccines are available yet for prevention of infection or recurrence of HAT [2]. Drugs currently in use have limitations in the form of toxicity, and increasing therapeutic resistance. As a result, new drugs are needed for treatment of HAT [3, 4]. *Trypanosoma brucei,* which causes HAT, survives in the host bloodstream by evading immune response by a process termed " antigenic variation", which involves periodic switching of surface coat proteins called variant surface glycoprotein (VSG) [5, 6]. VSGs enter the secretory pathway *via* the endoplasmic reticulum (ER) *en route* to the plasma membrane [7]. We have studied VSGs as model substrates to gain insight into translocation of proteins into the trypanosome ER [Chapter Two]. In this study, we explore the possibility of using small molecules to inhibit protein translocation into the trypanosome ER. Further, the impact of small molecules on parasite viability is also investigated.

We recently developed a cell-free system using *T. brucei* microsomes (TbRM) that imports trypanosome proteins post-translationally, in a parasite cytosol-dependent fashion [Chapter Two]. *T. brucei* cytosol could be replaced by purified Ssa1p [8], a molecular chaperone. These data indicate that protein translocation into *T. brucei* ER can be chaperone-mediated. In contrast, import of physiological proteins into the ER of mammalian host (of trypanosomes), is essentially co-translational (reviewed in [9]). Thus, biochemical differences in ER translocation pathways may exist between a trypanosome and its human host.

To exploit the presumed molecular difference between host and parasite ER protein import pathways, we first tested for compounds that could inhibit post-translational import of proteins into the trypanosome ER. Thereafter, the compounds were tested for trypanocidal activity and toxicity of vertebrate cell line. Two natural compounds of fungal origin, namely CJ-21,058 [10] and equisetin [11]

blocked import of proteins into TbRM, and exhibited anti-trypanosomal activity. Several compounds produced by combinatorial chemistry that inhibit Hsp70-Hsp40 interactions [8] were tested. Of those, MAL3-101 blocked protein translocation into TbRM and killed cultured *T. brucei* at sub-micromolar concentrations. These studies indicate that the ER protein import pathway of trypanosomes may be a target for identification of lead compounds for discovery of anti-trypanosome drugs.

MATERIALS AND METHODS

Reagents and Chemicals

Plasmid pVSG_117 [12] was provided by Jay Bangs, University of Wisconsin, Madison. CJ-21,058 and equisetin were obtained from Pfizer Inc. (New York, New York). Rabbit reticulocyte lysate and methionine/cysteine-free amino acid mixture was purchased from Promega (Madison, Wisconsin). [^{35}S] Redivue Promix^{TM} was purchased from Amersham Biosciences (Piscataway, New Jersy), etoposide from Sigma, (St. Louis, Missouri) and propidium iodide from Invitrogen (Carlsbad, California). Ampliscribe T7 *In vitro* transcription kit was purchased from Epicentre Technologies (Madison, Wisconsin).

DNA Templates and RNA Synthesis *In vitro*

The DNA substrate for synthesis of VSG_117$_{500}$ RNA was pVSG_117. A template for *in vitro* transcription was obtained by PCR using the forward primer ccc*taatacgactcactatagggaggagggtttttacc*at ggactgccatacaaaggag, which contains a T7 promoter (italicized), and a translation enhancer (underlined) [13]. The first 21 nucleotides of the VSG_117 coding sequence are in regular style font. The reverse primer for VSG_117$_{500}$ is *catcatcat*ttcctaaaaaagcaaggc that contains a termination codon (underlined) and nucleotides 516-525 of the coding region (regular font). One µg of DNA template was transcribed using the Ampliscribe T7 kit (Epicentre Technologies). RNA was purified and characterized as described [Chapter 2].

T. brucei Microsomes (TbRM) and Cytosol

T. brucei microsomes and cytosol were prepared and stored as described previously [Chapter 2].

Protein Synthesis *In vitro* and Import into TbRM

One µg of RNA was translated in a reaction containing 10 µl of rabbit reticulocyte lysate (Promega), 1 mM methionine/cysteine-free amino acid mixture (Promega), and 15 µCi [^{35}S] redivue Promix (Amersham Bioscience) at 37°C for 90 minutes in a total volume of 20 µl. Protein import assays using TbRM were performed as described [Chapter 2].

Cell Culture

Blood stream form *T. brucei* CA427 were cultured in HMI-9 medium [14] to a density of 10^6/ml. Cells were diluted to 10^4/ml, and hundred microlitres transferred to 96-well plates with or without addition of test compounds at stated concentrations. In controls, DMSO, the solvent for the compounds, was added. All experiments were performed in triplicates. Cells were counted at the end of 24 or 48 hours using a hemocytometer, and the data graphed. HeLa cell line was maintained in RPMI-1640 supplemented with 10% bovine serum, 100 U/ml penicillin/streptomycin, and 2 mM L-glutamine at 37°C in 5% CO_2 (Maselli *et al*., 2002). Cells were plated at 1 x 10^4 cells per well (96 well plate) with or without addition of test compounds. At the end of 24 hours, 1 x 10^5 cells per well were analyzed by flow cytometry (Dakocytomation CyAn).

HeLa Cell Viability Assay

Propidium iodide exclusion (reviewed in Steff *et al*., 2001; Ilic *et al*., 1998; da Costa *et al*., 1999) was used to determine HeLa cell viability. Adherent cells were rinsed with 1X phosphate buffered saline (PBS), trypsinized (37°C for 5 min), pelleted (2,000 g, 2 min) and resuspended in 1X PBS (500 µl) containing propidium iodide (1 µg/ ml), DAPI [4'-6-Diamidino-2-phenylindole] (5 µg / ml) and analyzed by flow cytometry on a Dakocytomation CyAn without gating. Single fluorophore controls were

acquired prior to the collection of the experimental data. Etoposide stock (2mM) was dissolved in DMSO, and diluted in RPMI 1640 to concentrations of 30 μM and 100 μM immediately before use.

RESULTS

Effect of MAL3-101 and MAL3-51 on Protein Translocation into TbRM

We have established a cell-free protein import system using *T. brucei* microsomes (TbRM) that appears to be specific for *T. brucei* ER signal peptides [Chapter 2]. Translocation of full-length proteins into TbRM is post-translational and requires parasite cytosol that can be replaced by purified molecular chaperone (Hsp70) [Chapter 2]. MAL3-101 is an inhibitor of intrinsic Hsp70 as well as J-domain stimulated Hsp70 in *S. cerevisiae* [8]. MAL3-101 also blocked *in vitro* import of pre-pro α mating-factor into *S. cerevisiae* microsomes [8]. MAL3-101 (0.3 μM) inhibited import of VSG_117_{500} into TbRM (Fig. 1B, compare lanes 3 and 4) [Chapter 2]. The results were encouraging, therefore we decided to test the effect of other small molecule modulators of Hsp70 ATPase activity on VSG_117_{500} import into TbRM. MAL3-51 and MAL3-90 are inhibitors of intrinsic HSP70 ATPase that do not affect J-domain stimulated ATPase activity of Hsp70 [8].

To determine whether MAL3-51 affected VSG_117_{500} import into TbRM, MAL3-51 (500 μM) was preincubated with *T. brucei* cytosol, which was added to an *in vitro* translocation reaction, in place of untreated cytosol (for outline of protocol, see Fig. 1A). Proteinase K treatment was performed to evaluate import of VSG_117_{500} into TbRM (see Fig. 1A). In absence of any inhibitor, TbRM imported approximately 80% of VSG_117_{500}, as measured by protection from proteinase K digestion (Fig. 1B, compare lanes 1& 2). In the presence of MAL3-51, 75% of VSG_117_{500} was imported into TbRM (Fig. 1B, compare lanes 5 and 6). Similarly, in presence of MAL3-90 (500 μM), 70% of VSG_117_{500} was

imported into TbRM (data not shown). Therefore, while MAL3-101 efficiently inhibited protein translocation into TbRM, MAL3-51 and MAL3-90 failed to do so.

Equisetin and CJ-21,058 Block Protein Translocation into TbRM

Post-translational protein translocation could be i) SecA mediated (e.g., *E. coli*) [15, 16], or ii) Ssa1p mediated (e.g., *S. cerevisiae*). At the time CJ-21,058 was tested, we had established that proteins were imported into TbRM post-translationally, in a parasite-cytosol dependent manner. CJ-21,058 (isolated from the soil fungus CL47745 [10]) was reported to have an antibacterial activity, that was attributed, in part, to its SecA inhibitory activity [10]. SecA mediates post-translational protein translocation in prokaryote system. Also, since some trypanosomatid (*e.g., Leishmania*) signal peptides resemble that of prokaryotes, it is possible that they utilize similar translocation machinery. We tested the effect of CJ21,058 on VSG_117_{500} import into TbRM. A protocol outlined earlier was used (see Fig. 1A). CJ-21,058 (20 μM) inhibited the translocation of VSG_117_{500} into TbRM by 65% (Fig. 1B, compare lanes 7 and 8).

Equisetin, an analog of CJ-21,058, is produced by the soil fungus *Fusarium heterosporum* [11] (Fig. 4B). Since CJ-21,058 blocked protein translocation into TbRM, we tested if equisetin had similar effects. Equisetin (25 μM) inhibited the translocation of VSG_117_{500} into TbRM by 95% (Fig. 1B, compare lanes 9 and 10). Hence, both CJ-21,058 and equisetin block protein translocation into TbRM.

Trypanocidal Effect of Protein Translocation Blockers (PTBs)

Several important proteins in *T. brucei* enter the secretory pathway *via* the ER [17]: Blocking the movement of proteins into ER would prevent them from reaching their final destination, which in turn may affect parasite survival. Therefore, compounds that block protein translocation into TbRM (PTBs) *in vitro* were tested for trypanocidal activity. In control studies, DMSO (solvent for all compounds tested) had no effect on growth of the parasite. MAL3-101 killed *T. brucei* in dose-dependent fashion (Fig. 2A).

Fifty percent of parasites were killed at 125 nM (IC_{50}), and one hundred percent of parasites were killed at 500 nM (IC_{100}). MAL3-51 did not show trypanocidal activity at concentrations as high as 5 mM (Fig. 2B). Similarly, MAL3-90 did not kill *T. brucei* at 3 mM (data not shown). Equisetin possessed trypanocidal activity (IC_{50} = 3.3 μM, and IC_{100} = 26 μM) (Fig. 2C). CJ-21,058 also had trypanocidal action (IC_{50} = 7 μM, and IC_{100} = 20 μM) (Fig. 2D).

Anti-trypanosome Compounds do not Kill a Model Mammalian Cell

A good anti-pathogen agent should have minimal effect on the host during the course of its administration. Therefore, trypanocidal compounds were tested on human HeLa cells using the concentrations that killed one hundred percent *T. brucei* (IC_{100}). Addition of MAL3-51, MAL3-101 and MAL3-90 (up to 1 □M) did not change HeLa cell viability (data not shown). CJ-21,058 (20 μM) and equisetin (26 μM) did not affect HeLa cells. Using a cell viability assay based on propidium iodide exclusion (Steff *et al*., 2001, Ilic *et al*., 1998, da Costa *et al*., 1999), no statistically significant difference was detected between drug-treated cells when compared to control HeLa cells (Fig. 3). In contrast, etoposide, a topoisomerase II inhibitor, produced significant cell death (Fig. 3).

DISCUSSION

Trypanosome Secretory Pathway as a Target for Drug Discovery

Cell surface proteins are critical for parasite survival in trypanosomatids (*T. brucei, Trypanosoma cruzi and Leishmania* species). Some important proteins that are transported to their destination by the secretory pathway in *T. brucei* include VSG [18], transferrin receptors [17] and GP63 [19]. Approximately 10^7 VSG molecules are present on cell surface contribute to antigenic variation of bloodstream form *T. brucei* [18]. Recently, RNAi of a VSG in bloodstream form resulted in cell cycle arrest and inhibition of cell growth *in vitro* [20]. We are offering evidence in our study, that inhibiting translocation of proteins into the ER is detrimental to physiology of the parasite, thus making molecular participants in ER protein translocation a target for discovery of drugs against *T. brucei.*

In vertebrates, protein import into ER is essentially co-translational and SRP-dependent [21]. On the contrary, many *T. brucei* proteins are capable of translocation across ER in SRP independent fashion [22]. We have shown earlier that *T. brucei* VSG_117 and VSG MVAT7 can be imported into TbRM post-translationally, in a signal peptide and chaperone (Ssa1p [8]) dependent fashion [Chapter Two]. Hence, both *in vitro* and *in vivo* data suggests the presence of post-translational protein translocation into ER of *T. brucei*. These biochemical differences (listed in Table 1) could be exploited to search for targets in parasite ER translocation pathway.

Protein Translocation Inhibitors (PTBs) as Effective Trypanocidal Compounds

Hsp70 (Ssa1p) molecular chaperones keep proteins in import competent state for translocation into cellular organelles [23]. Hsp70 class of proteins contains three domains namely, i) ATPase, ii) peptide binding, and iii) helical lid domain [24-27]. Helical lid domain shields the peptide-binding domain until ATP is bound to n-terminal ATPase domain. Upon binding of ATP, peptide-binding domain is uncovered, allowing proteins (that are targeted to the membrane) to bind it. ATP hydrolysis follows,

and it causes conformational changes in the chaperone, and pre-protein gets tightly bound to Hsp70 by closure of the helical lid domain [28]. Hsp40 class of proteins (e.g., Ydj1p) assist chaperone activity of Hsp70. J-domain of Hsp40 class of proteins interact with ATPase domain of Hsp70 and stimulates its ATPase activity [29].

A known Hsp70 modulator, 15 deoxyspergualin was used as a seed to search the Developmental Therapeutics program database for DSG related compounds [8, 30]. The authors found NSC 630668-R/1 (also called R/1), which bore some structural similarity to 15-DSG [30]. R/1 inhibits endogenous as well as Hsp40 stimulated Hsp70 activity [8]. R/1 also inhibits post-translational translocation into *S. cerevesiae* ER-derived microsomes [8, 30]. Further, combinatorial chemistry was used to design structural analogs of R/1 and the compounds were designated as MAL3 group of compounds [8]. Of the 31 compounds so synthesized, MAL3-101 and MAL3-39 also inhibited the Hsp40 mediated Hsp70 activity [8]. Interestingly, MAL3-101 inhibits post-translational translocation of α-mating factor into yeast microsomes [8]. Structure–activity relationship of MAL3 group of compounds on *S. cerevisiae* or *T. brucei* is yet to be elucidated [8]. However, R/1 acts by mimicking the peptide substrate and also by stimulating the homopolymerization of Hsp70 [30].

We studied the effect of small molecule modulators of Hsp70 on protein import into TbRM. MAL3-101 (R/1 analog) (structure in Fig. 4A), an inhibitor of J-domain stimulated Hsp70 ATPase [8], blocked import of VSG_117_{500} into TbRM (Fig. 1, lanes 3 and 4) [Chapter 2]. As previously noted, Ssa1p from *S. cerevisiae* could replace *T.brucei* cytosol in post-translational translocation of proteins into TbRM [Chapter 2]. These two pieces of information put together, led us to infer that *T. brucei* homologs of Hsp70 (*i.e.,* Ssa1p) and Hsp40 (*e.g.,* Ydj1p) play an important role in ER protein import in the parasite. Encouraged by these results, we studied other MAL-3 class of compounds (MAL3-51 and MAL3-90), that acted by inhibiting intrinsic Hsp70 ATPase activity [8]. Surprisingly, neither of the compounds blocked protein translocation into TbRM. Next, MAL3 class of compounds were tested for their ability to

kill cultured *T. brucei* (Figure 2). Of the three MAL3 compounds that we studied, MAL3-101 was active against *T. brucei* at sub-micromolar concentration (Fig. 2B) while having minimal toxicity on a mammalian cell line (Fig 3). The IC_{50} of MAL3-101 is comparable to that of currently used drugs against HAT (IC_{50} of suramin is 1.4-2.3 μM, pentamidine is 0.3 nM) [31]. MAL3-51 and MAL3-90 did not kill *T. brucei* even at millimolar concentrations. The difference in results between MAL3-51 (or MAL3-90) and MAL3-101 suggest that J-domain stimulated activity of a *T. brucei* Hsp70 is important for translocation of protein into TbRM. The data suggests that Hsp40 co-chaperones may be very important for post-translational protein import in *T. brucei.*

In last twenty years, ~49-62% of new chemical entities (NCE) that became lead compounds in industry were either natural products, their analogs or inspired from natural products [32, 33]. Equisetin and CJ-21,058 belong to acyl tetramic acid (pyrrolidine-2,4-dione) family of compounds (Structure in Fig. 4B) [34]. The biological activity of these compounds have been attributed to tetrameric acid, which is found in many natural compounds with proven antibiotic, antifungal and antiviral properties. Equisetin is produced by the fungus *Fusarium heterosporum* [11] while its analog, CJ-21, 058, is isolated from the fungus CL47745 [10].

Both equisetin and CJ-21,058 inhibit import of proteins into *T. brucei* ER (Fig. 1), are effective as anti-trypanosomal compounds (Fig. 2), and not toxic to HeLa cells (Fig. 3). IC_{50} of equisetin (3.3 μM) and CJ-21,058 (7 μM) is comparable to IC_{50} of drugs that are currently used (IC_{50} of suramin is 1.4-2.3 μM, pentamidine is 0.3 nM) [31]. Equisetin is presently being evaluated for treatment of HIV (human immunodeficiency virus) infection (Reviewed in [35]). Hence, "piggy-backing" the drug for anti-trypanosomal activity would be feasible [36] since limited funds are available for HAT research. Further, equisetin has been chemically synthesized [34], possibly making it easier for mass production if the need arises.

Finally, the concentration at which the compounds (MAL3-101, equisetin, and CJ-21, 058) exhibited trypanocidal effect (IC_{100}) was used to test their protein translocation blocking effect on TbRM (Table 2). Interestingly enough, we found that trypanocidal concentration and protein translocation blocking concentration of these compounds were similar. Also, MAL3-51 and MAL3-90, which did not show protein translocation blocking effect, did not kill *T. brucei* at concentrations up to 2 mM. This leads us to hypothesize that MAL3-101, equisetin, and CJ-21, 058 might be trypanocidal due to their effect on post-translcational protein translocation.

We propose that further studies i) other R/1 analogs, and, ii) equisetin derivatives (natural and synthetic) would prove valuable in our continued efforts to finding new lead compounds against *T. brucei*.

REFERENCES

1. Smith, D.H., J. Pepin, and A.H.R. Stich, *Human African trypanosomiasis: an emerging public health crisis.* Br.Med.Bull., 1998. **54**: p. 341-355.
2. Barrett, M. and H. Densie, *Uptake and mode of action of drugs used against sleeping sickness.* Biochemical pharmacology, 2001. **61**: p. 1-5.
3. Nok, A., *Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human african trypanosomiasis.* Parasitol Res, 2003. **90**: p. 71-79.
4. Docampo, R. and S.N. Moreno, *Current chemotherapy of human African trypanosomiasis.* Parasitol Res, 2003. **90 Supp 1**: p. S10-3.
5. Teixeira, S.M. and W.D. daRocha, *Control of gene expression and genetic manipulation in the Trypanosomatidae.* Genet Mol Res, 2003. **2**(1): p. 148-58.
6. Pays, E., L. Vanhamme, and D. Perez-Morga, *Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries.* Curr Opin Microbiol, 2004. **7**(4): p. 369-74.
7. McConville, M.J., et al., *Secretory pathway of trypanosomatid parasites.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 122-54; table of contents.
8. Fewell, S.W., et al., *Small Molecule Modulators of Endogenous and Co-chaperone-stimulated Hsp70 ATPase Activity.* J Biol Chem, 2004. **279**(49): p. 51131-40.
9. Lutcke, H., *Signal recognition particle (SRP), a ubiquitous initiator of protein translocation.* Eur.J.Biochem., 1995. **228**: p. 531-550.
10. Sugie, Y., et al., *CJ-21,058, a new SecA inhibitor isolated from a fungus.* J Antibiot (Tokyo), 2002. **55**(1): p. 25-9.
11. Vesonder, R.F., et al., *Equisetin, an antibiotic from Fusarium equiseti NRRL 5537, identified as a derivative of N-methyl-2,4-pyrollidone.* J Antibiot (Tokyo), 1979. **32**(7): p. 759-61.
12. Bangs, J.D., et al., *Posttranslational modification and intracellular transport of a trypanosome variant surface glycoprotein.* J Cell Biol, 1986. **103**(1): p. 255-63.

13. Al-Qahtani, A. and K. Mensa-Wilmot, *A 5' untranslated region which directs accurate and robust translation by prokaryotic and mammalian ribosomes.* Nucleic Acids Res., 1996. **24**: p. 1173-1174.

14. Hirumi, H., et al., *Short communication: cultivation of bloodstream forms of Trypanosoma brucei and T. evansi in a serum-free medium.* Trop Med Int Health, 1997. **2**(3): p. 240-4.

15. Watanabe, M. and G. Blobel, *SecA protein is required for translocation of a model precursor protein into inverted vesicles of Escherichia coli plasma membrane.* Proc.Natl.Acd.Sci.U.S.A., 1993. **90**: p. 9011-9015.

16. Knott, T.G. and C. Robinson, *The secA inhibitor, azide, reversibly blocks the translocation of a subset of proteins across the chloroplast thylakoid membrane.* J Biol Chem, 1994. **269**(11): p. 7843-6.

17. Steverding, D., et al., *Transferrin-binding protein complex is the receptor for transferrin uptake in Trypanosoma brucei.* J.Cell Biol., 1995. **131**: p. 1173-1182.

18. Pays, E. and D.P. Nolan, *Expression and function of surface proteins in Trypanosoma brucei.* Mol.Biochem.Parasitol., 1998. **91**: p. 3-36.

19. El-Sayed, N.M. and J.E. Donelson, *African trypanosomes have differentially expressed genes encoding homologues of the Leishmania GP63 surface protease.* J Biol Chem, 1997. **272**(42): p. 26742-8.

20. Sheader, K., et al., *Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes.* Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8716-21.

21. Rapoport, T.A., B. Jungnickel, and U. Kutay, *Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes.* Annu.Rev.Biochem., 1996. **65**: p. 271-303.

22. Liu, L., et al., *RNA interference of signal peptide-binding protein SRP54 elicits deleterious effects and protein sorting defects in trypanosomes.* J Biol Chem, 2002. **277**(49): p. 47348-57.

23. Brodsky, J.L., *Post-translational protein translocation: not all hsc70s are created equal.* Trends Biochem Sci, 1996. **21**(4): p. 122-6.

24. Wang, T.F., J.H. Chang, and C. Wang, *Identification of the peptide binding domain of hsc70. 18-Kilodalton fragment located immediately after ATPase domain is sufficient for high affinity binding.* J Biol Chem, 1993. **268**(35): p. 26049-51.

25. Zhu, X., et al., *Structural analysis of substrate binding by the molecular chaperone DnaK.* Science, 1996. **272**(5268): p. 1606-14.

26. Flaherty, K.M., C. DeLuca-Flaherty, and D.B. McKay, *Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein.* Nature, 1990. **346**(6285): p. 623-8.

27. Chappell, T.G., et al., *The ATPase core of a clathrin uncoating protein.* J Biol Chem, 1987. **262**(2): p. 746-51.

28. Misselwitz, B., O. Staeck, and T.A. Rapoport, *J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences.* Mol Cell, 1998. **2**(5): p. 593-603.

29. Cheetham, M.E. and A.J. Caplan, *Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function.* Cell Stress Chaperones, 1998. **3**(1): p. 28-36.

30. Fewell, S.W., B.W. Day, and J.L. Brodsky, *Identification of an inhibitor of hsc70-mediated protein translocation and ATP hydrolysis.* J Biol Chem, 2001. **276**(2): p. 910-4.

31. Merschjohann, K., et al., *In vitro effect of alkaloids on bloodstream forms of Trypanosoma brucei and T. congolense.* Planta Med, 2001. **67**(7): p. 623-7.

32. **Newman DJ,** G.M.C., **and Kenneth M. Snader**, *Natural Products as Sources of New Drugs over the Period 1981-2002. J. Nat. Prod*, 2003. **66**(1022-1037).

33. Koehn FE, C.G., *The evolving role of natural products in drug discovery.* Nat Rev Drug Discov , 2005. **3**: p. 206-220.

34. Burke, L.T., et al., *A short stereoselective total synthesis of the fusarium toxin equisetin.* Org Lett, 2000. **2**(23): p. 3611-3.

35. De Clercq, E., *Current lead natural products for the chemotherapy of human immunodeficiency virus (HIV) infection.* Med Res Rev, 2000. **20**(5): p. 323-49.

36. Nallan, L., et al., *Protein farnesyltransferase inhibitors exhibit potent antimalarial activity.* J Med Chem, 2005. **48**(11): p. 3704-13.

FIGURE LEGEND

Fig. 1. Effect of Protein Translocation Inhibitors on TbRM

(A) Protocol Used for Import of VSG_117$_{500}$. ***(B)*** VSG_117$_{500}$ mRNA was translated in rabbit reticulocyte lysate with 1.5 equivalents *T. brucei* cytosol {pretreated with MAL3-101 (0.3 μM), MAL3-51 (1 M), CJ21, 058 (20 μM) or equisetin (50 μM)} for 60 min. The reaction mixtures were treated with cycloheximide (50 μg/ml, final concentration) and incubated with TbRM (one equivalent). The mixture was incubated for 60 min at 37^{0}C and digested with proteinase K digestion (30 μg/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS-PAGE and detected by phosphorimaging. *Lanes* 1 & 2 (TbRM pretreated with DMSO): *Lane 1*, untreated VSG_117$_{500}$ with TbRM; *lane 2*, VSG_117$_{500}$ incubated with TbRM & treated with proteinase K. *Lanes 3 & 4* (TbRM pretreated with MAL3-101): *Lane 3*, VSG_117$_{500}$ incubated with TbRM; *lane 4*, VSG_117$_{500}$ incubated with TbRM and then treated with proteinase K. *Lanes 5 & 6* (TbRM pretreated with MAL3-51): *Lane 5*, VSG_117$_{500}$ incubated with TbRM; *lane 6*, VSG_117$_{500}$ incubated with TbRM and then treated with proteinase K. *Lanes 7 & 8* (TbRM pretreated with CJ-21,058): *Lane 7*, VSG_117$_{500}$ incubated with TbRM; *lane 8*, VSG_117$_{500}$ incubated with TbRM and then treated with proteinase K. *Lanes 9 & 10* (TbRM pretreated with equisetin): *Lane 9*, VSG_117$_{500}$ incubated with TbRM; *lane 10*, VSG_117$_{500}$ incubated with TbRM and then treated with proteinase K.

Figure 2: Trypanocidal Effect of Protein translocation Blockers (PTBs)

Blood stream form *T. brucei* CA427 were grown in HMI-9 media to the cell density of 10^6. The cells were then transferred to 96 well plates with the addition of different concentrations of MAL3-101, MAL3-51, CJ-21,058 or equisetin. In controls, DMSO was added. Cell density was calculated at the end of 24 hours, and the graph was plotted.

Figure 3: Anti-trypanosome Compounds do not Kill a Model Mammalian Cell

Adherent HeLa cells (treated with DMSO, MAL3-101, MAL3-51, CJ-21,058 or equisetin) were subject to viability assay, using propidium iodide (1 µg/ ml), DAPI [4'-6-Diamidino-2-phenylindole] (5 □g / ml) and analyzed by flow cytometry on a Dakocytomation CyAn without gating. Single fluorophore controls were acquired prior to the collection of the experimental data. Etoposide stock (2 mM) was dissolved in DMSO, and diluted in RPMI 1640 to concentrations of 30 □M and 100 M immediately before use.

Figure 4: Chemical Structures of MAL3-101, CJ-21,058 or Equisetin

Chemical structures of the compounds were drawn using CS ChemDraw Ultra.

A) MAL3-101, ***B)*** MAL3-51, ***C)*** CJ-21,058 and equisetin

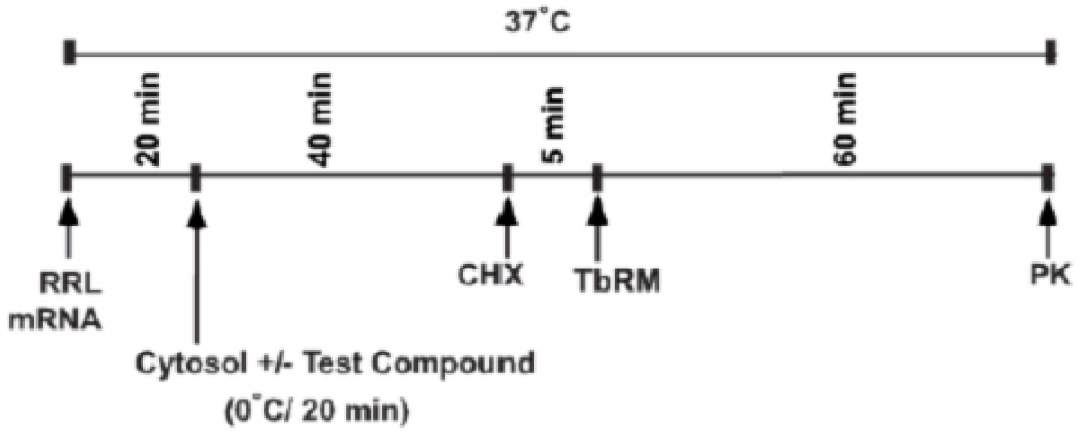
Table 1: Components of ER Protein Translocation Machinery in Host and Parasite

A list of requirements for ER protein translocation for vertebrate and *T. brucei* are compared

Table 2: Comparing the Concentration of Compounds for Protein Translocation Blocking Effect with IC_{100}

Figure 1:

A


37˚C
20 min
40 min
5 min
60 min
RRL
mRNA
Cytosol +/- Test Compound
(0˚C/ 20 min)
CHX
TbRM
PK

B

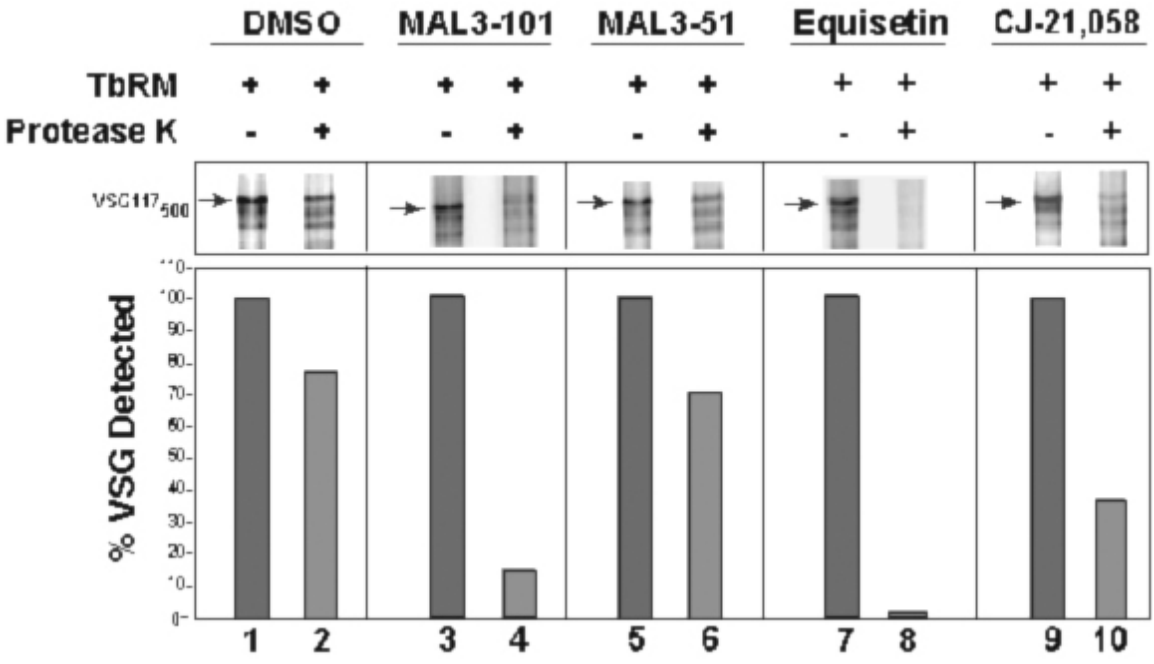
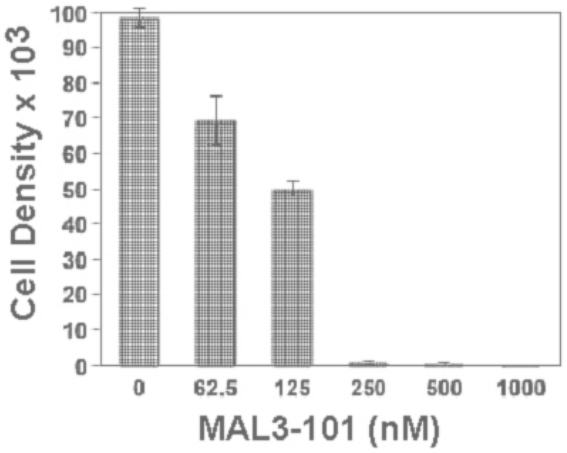
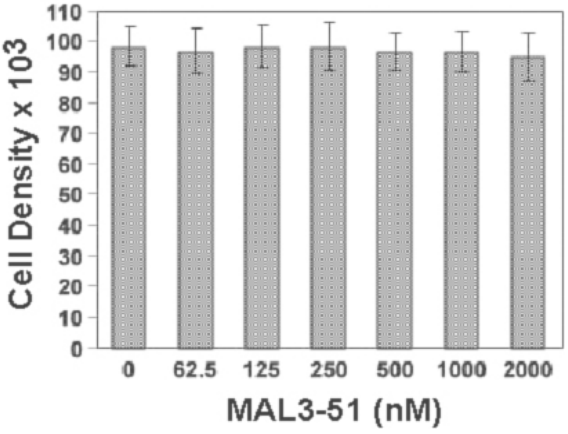

DMSO
MAL3-101
MAL3-51
Equisetin
CJ-21,058
TbRM
Protease K
VSG117500
% VSG Detected
1 2 3 4 5 6 7 8 9 10

Figure 2:

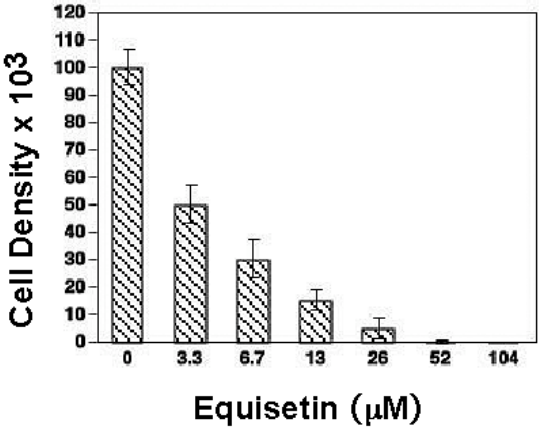
A



B



C



D

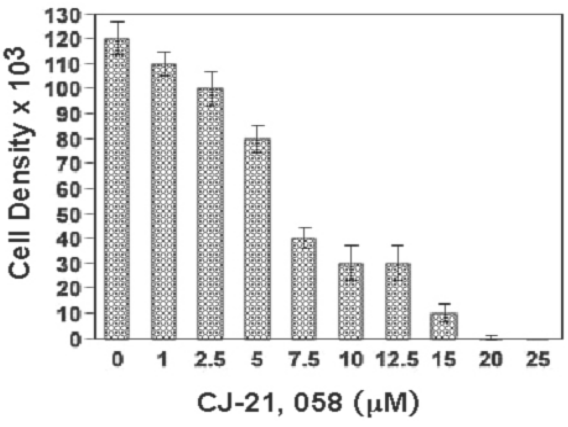


Figure 3:

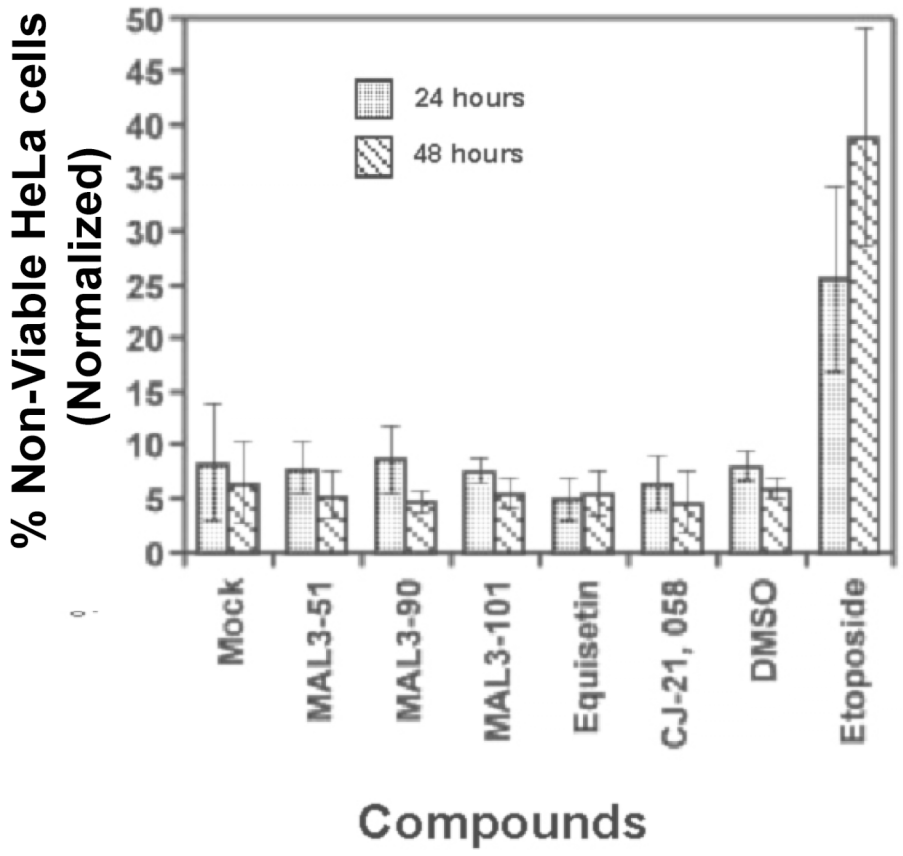
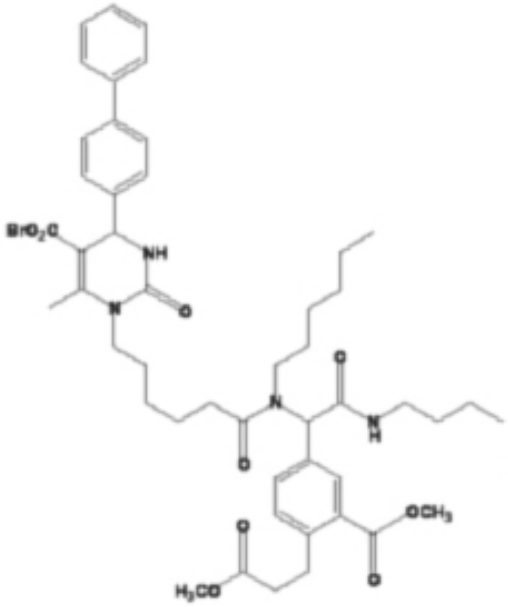
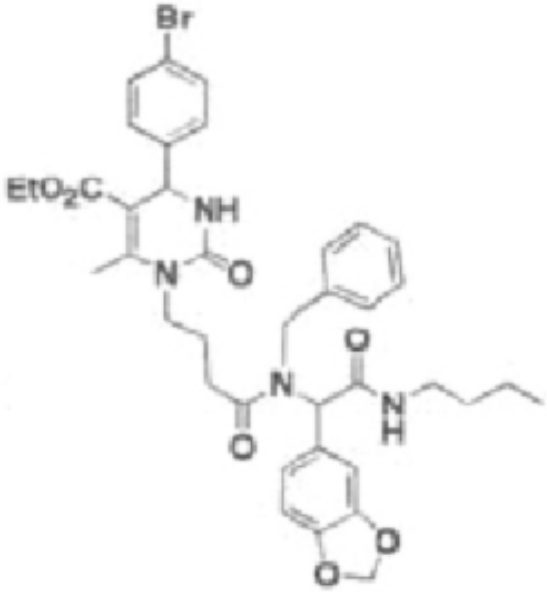

% Non-Viable HeLa cells (Normalized)
50
45
40
35
30
25
20
15
10
5
0
24 hours
48 hours
Mock
MAL3-51
MAL3-90
MAL3-101
Equisetin
CJ-21, 058
DMSO
Etoposide
Compounds

Figure 4

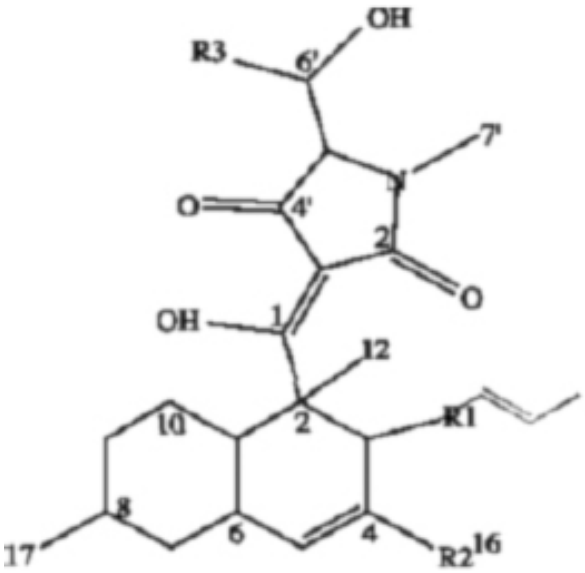
A **MAL3-101**



B **MAL3-51**



C **CJ-21,058 and Equisetin**



	R2	R3
CJ-21, 058:	Me	H
Equisetin:	H	H

Table 1: Requirements for Protein Translocation into the ER

Requirements	**Vertebrates**	***T. brucei***
Signal Peptide	Dependent	Dependent
SRP	Dependent	Independent
SR	Dependent	Not tested
Hsp70	Independent	Dependent
Hsp40	Independent	Implicated
Sec61	Different	
Signal Peptide Motifs	Different	

Table 2: Comparing the Concentration of Compounds for Protein Translocation Blocking Effect with IC_{100}

Compounds	Concentration for PTB Effect	IC_{100}
DMSO	-	-
MAL3-101	300 nM	250 nM
MAL3-51	-	Not effective up to 2 mM
MAL3-90	-	Not effective up to 2 mM
CJ-21,058	20 μM	20 μM
Equisetin	25 μM	26 μM

CHAPTER 4

ENDOPLASMIC RETICULUM SIGNAL PEPTIDES CONTAIN CONSERVED PEPTIDE MOTIFS THAT ARE SPECIES SPECIFIC [1]

[1] Patham, B., and Mensa-Wilmot, K.
To be submitted to *Journal of Biological Chemistry*, 05/2005

LIST OF ABBREVIATIONS

CHX, cycloheximide; DMSO, dimethyl sulfoxide; ER, endoplasmic reticulum; FC: fixed component; FC-1, fixed component in position 1; FC-2, fixed component in position 2; FC-3, fixed component in position 3; FC-4, fixed component in position 4; LR, linker region; LR1, linker region between FC-1 and FC-2; LR2, linker region between FC-2 and FC-3; LR3, linker region between FC-3 and FC-4; LR_{min}, minimum number of amino acids in linker region; LR_{max}, maximum number of amino acids in linker region; PK, proteinase K; PMSF, phenylmethyl sulfonylfluoride; RRL, rabbit reticulocyte lysate; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; TbRM, *T. brucei* rough microsomes; VSG, variant surface glycoprotein.

ABSTRACT

Signal peptides that direct protein import into the endoplasmic reticulum (ER) in eukaryotes and across the inner membrane of bacteria, have three conserved regions; an amino-terminal n-region, a hydrophobic h-region, and a carboxy-terminal c-region. Although the hydrophobic core is essential for efficient import/export of proteins through the appropriate translocons, the functional characteristics of the h-region is poorly understood. A bioinformatics approach was adopted to provide information of any existing patterns in the h-region of signal sequences. Analysis of the h-region of *Homo sapiens*, *Saccharomyces cerevisiae*, *Trypanosoma brucei*, *Escherichia coli*, and *Bacillus subtilis* signal sequences identified distinct peptide motifs in 75-90% of the h-regions. In a comparison of the peptide motifs, variation was found in: i) amino acid composition of motifs, ii) distance between amino acids that form the motif (*i.e.*, linker regions), and iii) length of the motif. The N-terminal 50 amino acids of human and *E. coli* cytosolic proteins failed to reveal conserved peptide motifs. The analysis revealed that the selection or distribution of amino acids in the h-region is not random as previously thought. On the contrary, we observed that the peptide motifs in the h-region of signal sequences; (i) have a well-defined organization of select hydrophobic amino acids, (ii) are highly conserved within a species. The physiological relevance of the motifs was investigated, using a trypanosome ER protein translocation system. Variant surface glycoprotein (VSG_117)) was imported into *T. brucei* microsomes (TbRM), when an unmutated h-region (which contains two overlapping motifs *i.e.*, $\mathbf{L_6}S_7\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}$, where $\mathbf{L_6}S_7\mathbf{L_8L_9}$ and $\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}$ have $\mathbf{L_8L_9}$ in common) was used. Although a signal peptide is essential for protein import into the ER, not much is known about the contribution of each sub-region to signal peptide activity. Deletion of the n-region of VSG_117 signal peptide (ΔN-VSG_117$_{500}$) did not affect import into TbRM. Mutants of VSG_117 h-region, in which the trypanosome tri-peptide motifs were disrupted without significantly altering the hydrophobicity, (*e.g.*, $\mathbf{V_8V_9}$ and $\mathbf{F_8F_9}$) of the signal peptide, failed to promote protein import into TbRM. This observation suggests that conserved peptide motifs in h region:

i) are important for signal peptide activity and, ii) do not tolerate changes in the amino acid sequence that substitute conserved hydrophobic amino acids for other hydrophobic amino acids.

INTRODUCTION

Signal peptides, together with a membrane translocon direct protein import into the endoplasmic reticulum (ER) (reviewed in [1-3]). Signal sequences are composed of three regions: (i) a positively charged n-region; (ii) a hydrophobic core; and (iii) a polar C-terminal region (reviewed in [4, 5]). In Gram-negative bacteria, (*e.g. E. coli*) signal sequences are required for export of proteins from the inner bacterial membrane to the periplasmic space (reviewed in [6, 7]), whereas in Gram-positive bacteria (*e.g. B. subtilis*) they direct secretion of protein into the external environment.

Signal peptides lack conserved primary sequence. Furthermore, in different organisms, variations exist in the length, charge and amino acid composition of each region (n, h, c) of signal peptides [8]. *E. coli* signal sequences are ~ 22-25 amino acids long [9]. The n-region is ~ 7 residues long and carries a net charge of +2 [10]. The h-region and c-region contain ~ 12 and 6 amino acids respectively [3, 8]. *B. subtilis* signal peptides are longer than those in *E. coli* (reviewed in [3, 8]): the n-region (of 7-12 amino acids) carries a net charge of +2- to - +7 (reviewed in [3, 8, 11]), and the h-region is ~ 15 amino acids in length. Leucine content of the h-region between the two types of eubacteria is similar, but *B. subtilis* has a much lower alanine content than *E. coli*. Eukaryotic signal sequences are generally slightly shorter (~ 18-20 residues) than prokaryotic sequences (~ 24-31 residues). The h-region of eukaryotic signal sequences has a higher leucine content than prokaryotes (reviewed in [3, 8, 11]).

For the purposes of directing entry into the ER, the h-region is often regarded as the most important of the three sub-regions of a signal sequence ([12-14]). It is also very well known that h-region of the signal peptide is essential for a functional signal sequence [15]. Many studies correlating hydrophobicity of the h-region to efficiency of signal sequence function have been documented [16]. In general, highly hydrophobic sequences direct co-translational import, while less hydrophobic sequences tend to drive post-translational import [16, 17].

Signal sequences from one species may not be functional in other species. A number of trypanosome signal sequences (*e.g.*, *T. brucei* VSG117, MVAT7, Bip, and *Leishmania* GP63) can not facilitate protein import into canine microsomes [18]. Conversely, several proteins across species did not get imported into *T. brucei* microsomes (TbRM) (*e.g.*, preProlactin, β-lactamase, α-mating factor) [Chapter Two]. The hydrophobicity profiles of signal peptides of all proteins were comparable, implying that incompatibility of signal peptide and translocon is not due to reduced hydrophobicity alone ([18], Chapter II]). Results from *in vivo* studies are consistent with the *in vitro* observations. For example, proteins with *Leishmania* signal sequence do not direct secretion in insect Sf9 cells [19]. Similarly a *T. cruzi* signal sequence could not direct protein import at the ER of murine *Vero* cells [20].

Several studies indicate that hydrophobicity of the h-region may not be the only factor determining efficiency of import. To garner further information about the h-region that may be important for signal peptide function; and explain species specificity, bioinformatic analysis of h-regions were performed. In this paper, we report the discovery of peptide motifs in the h-region of eukaryotic (humans, *S. cerevisiae*, *T. brucei*) as well as prokaryotic (*E. coli and B. subtils)* signal peptides. Our analysis revealed the presence of conserved tri-peptide motifs in the h-region of signal sequences for each species studied: human sequences also possessed tetra-peptide patterns. The h-region peptide motifs are highly conserved within a species and are seen in ~75-90% of signal sequences in all species we studied. Furthermore, disruption of peptide motifs in a trypanosome signal sequence (VSG_117) results in abolition of signal peptide activity. This strongly suggests that peptide motifs contribute to signal peptide function. The results show conclusively that signal peptides are not composed of random amino acids. The hydrophobicity of signal peptide is not sufficient to drive protein import [16].

MATERIALS AND METHODS

Sources Signal Sequences

H. sapiens and *E. coli* signal peptides were obtained from Signal 3P server [http://www.cbs.dtu.dk/services/SignalP/] [21, 22]. Protein sequences of *B. subtilis* were obtained from a paper by Tjalsma, H *et al.*, *(*Microbiol Mol Biol Rev, 2000. 64(3): p. 515-47) [23]. *S. cerevisiae* and *T. brucei* signal peptides were obtained from geneDB [www.genedb.org] [24]. Signal sequences chosen from GeneDB were documented to i) have a signal peptide, or ii) enter the secretory pathway. *T. brucei* VSG protein sequences were obtained from a VSG database (www.vsgdb.org).

Bioinformatic Analysis

H-regions of the signal peptides were predicted with PSORTII [http://psort.nibb.ac.jp/form2.html [25]. H-region were analyzed using PRATT [http://www.ebi.ac.uk/pratt/] [26] with the following parameters: C% (50); PL (50); PN (50); PX (5); FN (5); FL (2); FP (20); E (3) (C%, minimum percentage of sequences to match; PL, maximum pattern length; PN, maximum number of pattern symbols; PX, maximum number of consecutive x's; FN, maximum number of flexible spacers; FL, maximum flexibility; FP, maximum flexibility product; E, search greediness) [26].

PRATT Terminology

Peptide motifs are described by: i) amino acid composition of fixed components (FC), and ii) linker regions (LR) (Fig. 1). Each linker region can be represented either by a single (fixed) or multiple (flexible) amino acids [26].

Pattern Description and Terminology

The following abbreviations and parameters will be used to describe the patterns. FC, fixed component; FC-1, fixed component in position 1; FC-2, fixed component in position 2; FC-3, fixed component in position 3; FC-4, fixed component in position 4; LR, linker region; LR1, linker region

between FC-1 and FC-2; LR2, linker region between FC-2 and FC-3; LR3, linker region between FC-3 and FC-4; LR_{min}, minimum number of amino acids in linker region; LR_{max}, maximum number of amino acids in linker region.

Each fixed component can be represented either by a single (identity component) or multiple (ambiguous component) amino acids (Please refer to Fig. 1 for illustration of PRATT terminology). The terminology can be explained with the zinc-finger pattern as an example, which is written as ***C***-x(2,4)-***C***-x(3)-**[LIVMFYWC]**-x(8)-***H***-x(3,5)-***H*** [26]. A motif is comprised of fixed components that are joined by linker regions. **Fixed component (FC)** (in bold) may be either an ***identity component*** (bold and italicized) when occupied by a single amino acid (*e.g.*, C, or H), or **ambiguous component** (within square brackets) when occupied by a subset of amino acids (*e.g.*, L, V, I, M, F, Y, W or C). In the aforementioned example, FC-1 (occupied by C), FC-3 (occupied by C), FC-4 (occupied by H) and FC-5 (occupied by H) are identity components, whereas FC-2 is an ambiguous component, because it could be occupied by L, I, V, M, F, Y, W, or C.

Linker region (LR) (highlighted) separates two fixed components. For example, LR-1 is linker region between FC-1 and FC-2, while LR-2 is the spacer between FC-2 and FC-3, and so on. Linker regions affect the length of peptide motif, for example, linker region x(2,4) can accommodate a minimum of 2 amino acids (LR_{min}) and a maximum of 4 amino acids (LR_{max}) so, LR_{min} =2 and LR_{max} = 4. Linker region x(3) accommodates exactly 3 amino acids and in this case LR_{min} = LR_{max} =3.

Minimim peptide motif length is calculated by adding the total number of amino acids in fixed positions to the minimum number of amino acids that could be accommodated in all the linker regions (LR_{min}). For the pattern ***C***-x(2,4)-***C***-x(3)-**[LIVMFYWC]**-x(8)-***H***-x(3,5)-***H***, the minimum length of the motif is 21 [5 (number of fixed components) + 2 ($LR1_{min}$) + 3 ($LR2_{min}$) + 8 ($LR3_{min}$) + 3 ($LR4_{min}$)]. The maximum peptide motif length is calculated by adding number of total number of amino acids in fixed

positions to the maximum number of amino acids that could be accommodated in all the linker regions (LR_{max}). In the example of ***C***-x(2,4)-***C***-x(3)-**[LIVMFYWC]**-x(8)-***H***-x(3,5)-***H***, the maximum length of the motif is 25 consisting of [5 (fixed components) + 4 ($LR1_{max}$) + 3 ($LR2_{max}$) + 8 ($LR3_{max}$) + 5 ($LR4_{max}$)].

RESULTS

Conserved Motifs in the h-region of Human Signal Peptides

Signal sequences have three sub-regions; an amino-terminal n-region, a hydrophobic h-region, and a carboxy-terminal c-region (reviewed in [27, 28]). The h-region of the signal peptide is essential for a functional signal sequence [15]. However, alignment of sequences has failed to produce a conserved peptide motif. Furthermore, it is believed that amino acids in signal sequences are randomly chosen and distributed [35].

To determine whether conserved sequence motifs were present in h-regions, two hundred h-regions were analyzed using PRATT (http://www.ebi.ac.uk/pratt/) [26], that extracts conserved patterns in a set of sequences supplied to it [26]. Three high scoring (*i.e.,* patterns that are most commonly represented in the dataset) tetra-peptide patterns were discovered (Table 1A) that could be merged into two "unified motifs" (see Table 1A). Remarkably, seventy-five percent of human signal peptides contained at least one tetra-peptide motif.

The characteristic features of human h-region tetra-peptide motifs can be described under three sub-headings. *A)* Amino acid composition: i) FC-1 = L, ii) FC-2 = L/ A/ V/ G/ P/ I (in that order of frequency), iii) FC-3 = L, and iv) FC-4 = L. *B)* Linker region: minimum number of amino acids in linker region one ($LR1_{min}$) = 1, maximum number of amino acids in linker region one ($LR1_{max}$) = 2; $LR2_{min}$ = 0, $LR2_{max}$ = 1; $LR3_{min}$ = 0, $LR3_{max}$ = 2. *C)* Peptide motif length: the maximum length of a tetra-peptide pattern is 8 amino acids, while the minimum length is 5 amino acids (for terminology and calculations,

refer to our section on PRATT Terminology). PRATT alignment showing the tetra-peptide motif of the human h-region is shown in Table 1B.

Remaining twenty five percent of sequences examined reveal tri-peptide motifs (Table 1C). The characteristic features of human tri-peptide motifs are as follows: *A)* Amino acid composition: i) FC-1 = L, ii) FC-2 = L/ A/ V (in that order of frequency), iii) FC-3 = L. *B)* Linker region: $LR1_{min}$ = 0, $LR1_{max}$ =5; $LR2_{min}$ = 0, $LR2_{max}$ = 4. We conclude that patterns in the h-region of human ER signal peptides are highly conserved.

Conserved Peptide Motifs in the h-region of *S. cerevisiae* Signal Peptides

The h-regions of over hundred *S. cerevisiae* signal peptides were analyzed using PRATT (http://www.ebi.ac.uk/pratt/) [26] to determine the existence of conserved amino acid patterns. Ten high scoring tri-peptide patterns were discovered (Table 2A) that could be merged into five "unified motifs" (see Table 2A). Approximately ninety percent of *S. cerevisiae* signal peptides contained at least one tri-peptide pattern.

General properties of the conserved *S. cerevisiae* tri-peptide motifs are as follows: *A)* Amino acid composition: i) FC-1 = L, ii) FC-2 = L/ S, and iii) FC-3 = L/ A/ V/ G/ I (in that order of frequency). *B)* Linker region: $LR1_{min}$ = 0, $LR1_{max}$ = 5; $LR2_{min}$ = 0, $LR2_{max}$ = 4. *C)* Peptide motif length: longest tri-peptide motif is 10 amino acids, while shortest is 4 amino acids. We infer that the h-region of *S. cerevisiae* ER signal peptides contain highly conserved tri-peptide motifs.

Peptide Motifs in the h-region of *T. brucei* Signal Peptides

Signal sequences are essential for protein translocation in trypanosomes [11]. Trypanosomatid signal sequences have a tri-partite organization (n, h and c region) [29]. However, important characteristics of signal sequences in trypanosomatids have not been studied in depth.

The h-regions of fifty *T. brucei* signal sequences were analyzed using PRATT (http://www.ebi.ac.uk/pratt/) [26] in search of conserved amino acid patterns. Thirteen high scoring tri-peptide patterns were discovered, which were merged into eight unified motifs (Table 3A). Approximately eighty five percent of *T. brucei* signal peptides contained at least one tri-peptide motif.

General features of conserved *T. brucei* motifs are as follows: *A)* Amino acid composition: i) FC-1= L, V. ii) FC-2 = L, and iii) FC-3 = L/ A/ V/ P/ I (in that order of frequency). *B)* Linker region: $LR1_{min}$ = 0, $LR3_{max}$ = 2; $LR2_{min}$ = 0, $LR2_{max}$ = 3. *C)* Peptide motif length: longest tri-peptide motif could be 8 amino acids, while the shortest could be 3 amino acids. PRATT alignment of the h-region of 16 protein sequences is shown in Table 3B. We conclude that amino acid patterns in the h-region of *T. brucei* ER signal peptides are highly conserved.

Two novel features observed in *T. brucei* motifs include: i) the presence of A or V apart from L in FC-1 is allowed, ii) valine as an identity component in FC-1 (~40% sequences) making it unique to *T. brucei* (see Table 3, motif number 12). In all other species, valine is always found as an ambiguous component.

Peptide Motifs in the h-region of *T. brucei* Variant Surface Glycoproteins (VSG)

Variant surface glycoproteins (VSG) are major surface proteins (~10^7 / cell) expressed by blood stream *T. brucei*. H-regions of twenty-five VSGs (www.vsgdb.org) were analyzed as described earlier. Five high scoring tri-peptide motifs were found, that could be merged into two “unified motifs” (see

Table 4A). Ninety five percent of VSG signal peptides contained at least one tri-peptide motif. Properties of conserved tri-peptide motifs are as follows: *A)* Amino acid composition: i) L or A in FC-1, ii) L or A in FC-2 and iii) L in FC-3. *B)* Linker region: $LR1_{min}$ = 0, $LR1_{max}$ = 2; $LR2_{min}$ = 0, $LR2_{max}$ = 2. PRATT alignment of the h-region of VSG signal peptides is shown in Table 5B. H-regions of VSG signal peptides contain conserved peptide motifs.

Conserved Peptide Motifs in the h-region of Prokaryote Signal Peptides

To determine whether conserved sequence motifs were present in the h-regions in prokaryotes, Gram negative (*e.g.*, *E. coli*) and Gram positive bacterial (*e.g.*, *B. subtilis*) signal sequences were studied.

The h-regions of one hundred *E. coli* (prototype gram negative bacteria) signal sequences were analyzed using PRATT (http://www.ebi.ac.uk/pratt/) [26] to determine the existence of conserved amino acid patterns. Eleven high scoring tri-peptide patterns were discovered (Table 5A) that could be merged into four "unified motifs" (see Table 5A). Approximately, ninety percent of *E. coli* signal sequences contained at least one tri-peptide pattern.

Features of *E. coli* the h-region motifs are as follows: *A)* Amino acid composition: i) FC-1 = A/ L, ii) FC-2 = A/ L and, iii) FC-3 = A/ L/ V/ G. *B)* Linker region: $LR1_{min}$ = 0, $LR1_{max}$ = 5; $LR2_{min}$ = 0, $LR2_{max}$ = 5. *C)* Peptide motif length: longest tri-peptide pattern is 13 amino acids, while shortest is 4 amino acids. PRATT alignment of the h-region of 20 protein sequences is shown in Table 5B. We conclude that patterns in the h-region of *E. coli* signal peptides are highly conserved.

Next, we examined the thes of one hundred *B. subtilis* (prototype Gram positive bacteria) signal sequences as described earlier (http://www.ebi.ac.uk/pratt/) [26]. Eighteen high scoring tri-peptide patterns were discovered (Table 6A) that could be merged into eight "unified motifs" (see Table 6A). Over ninety percent of *B. subtilis* signal sequences contained at least one tri-peptide pattern.

General features of tri-peptides motifs in the h-region of *B. subtilis* are as follows: *A)* Amino acid composition: i) L in FC-1, ii) A, L, V, G or I in FC-2 and, iii) A, L, V or G in FC-3. *B)* Linker region: $LR1_{min} = 0$, $LR1_{max} = 5$; $LR2_{min} = 0$, $LR2_{max} = 5$. *C)* Peptide motif length: longest tri-peptide pattern is 13 amino acids, while shortest is 3 amino acids (for terminology and calculations, refer to our section on PRATT Terminology). PRATT alignment of the h-region of 25 *B. subtilis* sequences, is shown in Table 6B. We conclude that the h-region of *B. subtilis* signal peptides contain highly conserved peptide motifs.

Absence of Patterns in Cytosolic Protein Sequences

To test whether peptide motifs identified by PRATT were specific only to the h-region of signal peptides, we studied cytosolic proteins from humans and *E. coli* [21, 22] as controls. Only N-terminal 50 amino acids of cytosolic proteins were used for our analysis, since ER signal sequences are present in that region of proteins. Applying the same criteria used for analysis of signal peptides, PRATT did not reveal any common patterns of amino acids in cytosolic proteins.

The n-Region is not Essential for Trypanosome Signal Peptide Function

Signal peptides are essential for ER protein translocation in trypanosomes. [30]. However, no study has addressed the role of sub-regions of a trypanosome signal peptide in its function. We have developed a cell-free ER protein import system in *T. brucei* using *T. brucei* microsomes (TbRM) and model protein VSG_117 (Chapter II). VSG_117 is imported into TbRM post-translationally, in a cytosol-dependent fashion (Chapter II). Further. Signal peptide is essential for import of VSG_117 into TbRM (Chapter II). To address this issue, we deleted the n-region of VSG_117_{500} to create ΔN-VSG 117_{500}. mRNA of ΔN-VSG_117_{500} was translated in rabbit reticulocyte lysate. Parasite cytosol and TbRM were added post-translationally and reaction mixtures were treated with proteinase K. (See Fig. 2A for an outline of the experimental scheme).

In the absence of TbRM, ΔN-VSG_117$_{500}$ was degraded by proteinase K (Fig. 2B, lane 2). TbRM protected 65% of ΔN-VSG_117$_{500}$ from proteinase K digestion (Fig. 2B, lane 4). From these data, we conclude that ΔN-VSG_117$_{500}$ is imported into TbRM. The result indicates that n-region of *T. brucei* signal peptide is not essential for protein translocation into TbRM.

Mutation of a Conserved h-region Motif Abolishes Import of VSG_117 into *T. brucei* Microsomes (TbRM)

Of the three sub-regions of a signal peptide, h-region is considered essential for translocation purposes [16, 17]. Hydrophobicity of the h-region is important for signal peptide activity across species [16, 17]. It is also believed that amino acids in the h-region are randomly distributed [29]. However, using computational techniques, we have discovered unique peptide motifs in the h-regions of signal peptides (Table 1-7). It is not known whether the motifs have biological functions.

To test the hypothesis that h-region motifs are required for signal peptide activity, we generated a series of mutations in the h-region of the signal peptide of VSG_117. Since we know that the n-region is not essential for import of VSG_117$_{500}$ into TbRM (Fig. 2B, lanes 1-4), ΔN-VSG_117$_{500}$ was used as the substrate for PCR mutagenesis of the h-region. Numbering of residues in wild type VSG_117 h-region are as follows: $S_1T_2M_3L_4T_5L_6S_7\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}I_{12}T_{13}P_{14}$. VSG_117 h-region has two overlapping tri-peptide patterns. (i) L-L-x-[AILV] ($S_1T_2M_3L_4T_5L_6S_7\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}I_{12}T_{13}P_{14}$) (Table 8), and (ii) L-x(2,3)-L-[AILV] ($S_1T_2M_3L_4T_5\mathbf{L_6}S_7\mathbf{L_8L_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$) (residues that satisfy the conserved motifs are in bold lettering) (Table 8). Two leucines ($\mathbf{L_8L_9}$) are common to both motifs ($S_1T_2M_3L_4T_5L_6S_7\mathbf{L_8L_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$). Mutations of the h-region are depicted in Table 8.

Our first approach was to completely disrupt both motifs by mutating all the fixed components in the h-region. This was accomplished by changing all the fixed components (*i.e.*, ($\mathbf{L_6}$, $\mathbf{L_8}$, $\mathbf{L_9}$, and $\mathbf{A_{11}}$)

($S_1T_2M_3L_4T_5\mathbf{L_6}S_7\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}I_{12}T_{13}P_{14}$) to glycine ($S_1T_2M_3L_4T_5\mathbf{G_6}S_7\mathbf{G_8G_9}Y_{10}\mathbf{G_{11}}I_{12}T_{13}P_{14}$) (Table 8). mRNA of the resultant mutant ($G_6G_8G_9G_{11}$-h_{VSG_117}) was translated in presence of rabbit reticulocyte lysate. *T. brucei* cytosol and TbRM were added post-translationally, and the VSG was tested for proteinase K susceptibility (see Fig. 2A for an outline of protocol). In the absence or presence of TbRM, $G_6G_8G_9G_{11}$-h_{VSG_117} was degraded by proteinase K (Fig. 2B, lane 6 and 8). Therefore, $G_6G_8G_9G_{11}$-h_{VSG_117} is not imported into TbRM.

We surmise that either i) decreasing the hydrophobicity, or ii) disrupting conserved peptide motifs in the h-region of VSG_117 blocks signal peptide function. To determine whether it was the reduction in hydrophobicity of mutant signal peptide (due to replacement of leucine with glycine) led to decrease in VSG import, it was important to use a new substrate in which disruption of peptide motif did not reduce hydrophobicity of signal peptide.

To achieve this goal, we disrupted the h-region motifs of VSG_117 by mutating the two leucines common to both motifs (bold) ($S_1T_2M_3L_4T_5L_6S_7\mathbf{L_8L_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$). $\mathbf{L_8L_9}$ (peak hydrophobicity of leucine is 3.8 [32]) were mutated to three different amino acids of varying hydrophobicities, namely; glycine (–0.4), valine (4.2) and phenylalanine (2.8) (Kyte-Doolittle hydrophobicity scale [32]).

VSG_117 motif mutant G_8G_9-h_{VSG_117} ($S_1T_2M_3L_4T_5L_6S_7\mathbf{G_8G_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$) (Fig. 2C, lanes 3 and 4), mutant F_8F_9-h_{VSG_117} ($S_1T_2M_3L_4T_5L_6S_7\mathbf{F_8F_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$) (Fig. 2C, lanes 5 and 6), and mutant V_8V_9-h_{VSG_117} ($S_1T_2M_3L_4T_5L_6S_7\mathbf{F_8F_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$) (Fig. 2C, lanes 7 and 8) were tested for import into TbRM. The experimental protocol was similar to that used for ΔN-VSG_117_{500}, (Fig. 2C, lanes 1 and 2). Since all mutant proteins were degraded by proteinase K in presence of *T. brucei* cytosol and TbRM (Fig. 2), we conclude that mutant G_8G_9-h_{VSG_117}, mutant F_8F_9-h_{VSG_117} or mutant V_8V_9-h_{VSG_117} do not get imported into TbRM (Fig. 2C, lanes 4, 6, 8). Hence, i) peptide motifs are essential for signal peptide

function in *T. brucei*, ii) increasing the hydrophobicity cannot compensate for disruption of peptide motifs.

DISCUSSION

Discovery of Unique Peptide Motifs in the h-region of Signal Peptides

Signal sequences have three sub-regions; an N-terminal (n-region), a hydrophobic stretch (h-region) and a C-terminal region (c-region) (reviewed in [27, 28]). It is also very well documented that the h-region of the signal peptide is essential for a functional signal sequence [15]. Many studies correlating hydrophobicity of h-region to efficiency of signal sequence have been documented. Hydrophobicity of signal peptides also determines the targeting route of secretory protein [16]. In general, more hydrophobic sequences direct co-translational import, while less hydrophobic sequences tend to drive post-translational import [16, 17]. Yet, it is widely believed that "information content" of signal peptides is very low [29].

There is surmounting evidence that hydrophobicity of signal peptides may not be the only factor deciding signal peptide activity. A group of trypanosome signal peptides (*T. brucei* VSG117, MVAT7, Bip, and *Leishmania* GP63) could not direct import into canine microsomes [18]. Recently, we observed that signal peptides from other species could not lead import of proteins (preProlactin, β-lactamase, α -mating factor) into *T. brucei* microsomes. Hydrophobicity profile [32] of signal peptides of all proteins (VSG117, MVAT7, GP63, preProlactin, β-lactamase, α--mating factor) were comparable and favored efficient translocation into microsomes. ([18], Chapter II]).

Hence, we hypothesize that the h-region hydrophobicity, although important, may not be the only factor determining import of protein into ER. In light of all the given facts, we chose to bioinformatics approach to analyze the h-regions of signal peptides in various biological systems. In this paper, we report

the discovery of highly conserved peptide motifs in the h-region of signal peptide across species. Unified peptide motifs in the h-region of human signal peptides {L-x(0,2)-[AGILPV]-L-x(0,2)-L and L-x(0,5)-[ALV]-x(0,4)-L} (Table 1), *S. cerevisiae* signal peptides {L-x(0,1)-L-x(0,4)-L; L-x(4,5)-L-[AGILV]L-x(0,3)-S-x(0,4)-A; L-x(1,2)-S-[AGILV] ; L-x(0,3)-S-x(0,4)-L} (Table 2), *T. brucei* {L-x(1,3)-L-x(0,2)-L; L-x(0,3)-[AILV]-x(0,3)-L; L-x(1,3)-A-[AGILV]; L-L-[AGLV]; A-x(2,3)-L-x-[AGILPV]; V-x(0,1)-L-x(2)-[AGILV]} (Table. 3), *E. coli* L-x(0,5)-L-x(0,5)-L; L-x-[AGILV]-x(2)-L; L-x(0,3)-A-x(0,5)-L; L-x(0,5)-[IV]-x(0,5)-L; L-x(0,2)-A-[AGLV]; L-x(0,1)-L-x(0,5)-A} (Table 5) and *B. subtilis* {A-x(0,4)-L-x(0,5)-A; L-x(1,2)-A-[AGILV]; L-x(0,5)-A-x(0,5)-[AL]; L-x(0,5)-L-x(0,5)-A} (Table 6) were identified.

Next, we compared the most commonly occurring "unified motifs" in the h-region signal peptides across biological systems (Table 9A), as they are most representative (50%-75%) of each data set. This would allow us to understand the diversity in composition of amino acids in the h-region of different species that have previously been noted. Our observations are presented in Table 9B. The differences in peptide motifs across species were clear and distinct.

The first and foremost distinction came in the number of fixed components (amino acids in fixed positions) that make peptide motifs. Conserved tetra-peptide motifs (4 fixed components) are unique to human h-regions while all the h-regions (human, *S. cerevisiae, T. brucei E. coli* and *B. subtilis*) revealed the presence of tri-peptide motifs (3 fixed components).

Second, variations exist in amino acid composition of fixed components. When all unified motifs were compared across species, we observed the following (Table 9A): **A)** FC-1 =L (all species), A (*T. brucei, E. coli*), or V (only *T. brucei*); **B)** FC-2= A/L (all species); V (humans, *T. brucei, B. subtilis*); G (human*, B. subtilis*); I (human, *T. brucei, B. subtilis*); S (only *S. cerevisia*e); P (only human; **C)** FC-3= L (all species); A,V, G (all, except human); I (*S. cerevisia*e, *T. brucei, E. coli*); **D)** FC-4=L (only humans).

The third distinction involves variations in length of linker region (Table 9B, C): **A)** In humans, $LR1_{min}$=0, $LR1_{max}$=2; $LR2_{min}$=0, $LR2_{max}$=1; $LR3_{min}$=0, $LR3_{max}$=2; **B)** In *S. cerevisiae,* $LR1_{min}$=0, $LR1_{max}$=5; $LR2_{min}$=0, $LR2_{max}$=4; **C)** In *T. brucei,* $LR1_{min}$=0, $LR1_{max}$=3; $LR2_{min}$=0, $LR2_{max}$=3; **D)** In *E. coli* $LR1_{min}$=0, $LR1_{max}$=5; $LR2_{min}$=0, $LR2_{max}$=5; **E)** In *B. subtilis,* $LR1_{min}$=0, $LR1_{max}$=5; $LR2_{min}$=0, $LR2_{max}$=5 (refer to Table 9). We observed that human sequences have shortest linker regions (maximum of 2 amino acids), while prokaryotes carried longest (maximum of 5 amino acids). This falls back to the overall length of the signal peptides. Prokaryotes have the longest the h-region (~ up to 35 amino acids long) hence manage to have a longer linker region. On the other hand, eukayotic h-regions are short (~18-20 amino acids long), hence they possess relatively short linker region.

It is a popular belief that hydrophobicity of the h-region governs signal peptide activity. If it were to be so, the order of frequency of appearance amino acids in the h-region of signal peptides should reflect the peak hydrophobicity of amino acids, as calculated by Kyte-Doolittle hydropathy scale are as follows [32]; isoleucine(4.5), valine (4.2), leucine (3.8), phenylalanine (2.8), methionine (1.9), and alanine (1.8), glycine (-0.4) [32]. However, our observations did not support this prediction. Isoleucine the most hydrophobic amino acid [32], and phenylalanine are conspicuous by their absence as identity components in all motifs. Even in ambiguous positions, isoleucine and phenylalanine are very rarely found. Valine, the second most hydrophobic amino acid appears only in the ambiguous position except in *T. brucei*. Alanine, the least hydrophobic of the non-polar amino acids is the second most commonly occurring amino acid in the h-region motifs across species. Interestingly, alanine and methionine have almost similar hydrophobicities [32], but methionine is not found in the h-region motifs. Finally, leucine, is the most commonly occurring amino acid in the h-region peptide motifs across biological systems.

We predict that these peptide motifs have a significant influence on the function of signal peptide. This data clearly reveals that amino acids are neither randomly chosen, nor are they randomly distributed in the h-region of a signal peptide. To validate our perception, we revisited a particularly interesting work

in which the invertase signal peptide was replaced by random sequences from human database [29]. Approximately 20% of randomly chosen human sequences were found functional in *S. cerevisiae*, leading the authors to conclude that signal sequences are essentially made of randomly distributed amino acids. Authors document 22 functional sequences, which successfully replaced *S. cerevisiae* invertase signal sequence to drive import [29].

We analyzed the "randomly chosen functional sequences" using PRATT and obtained two unified motifs (Table 7). Not surprisingly, L-x(0,2)-S-x(0,5)-L motif was identified in 60% sequences (Table 7 and 9). A serine residue flanked by two leucines is a highly conserved peptide motif, which is unique to *S. cerevisiae* (Table 2). Next, non-functional sequences were analyzed (Table 7). We did not obtain any patterns belonging to motifs identified earlier in *S. cerevisiae* (Table 2). Hence it is no co-incidence that these "randomly chosen sequences" were either functional, or non functional in yeast. As the need for validation of our approach increased, we studied the sequences from VSG database (www.vsgdb.org) (VSGDB). In theory, the peptide motif(s) from VSGDB should be similar to the ones discovered in experimentally verified *T. brucei* signal peptides (Table 3). VSG dataset revealed two peptide motifs {L-x(0,1)-[AL]-x(0,2)-L and A-x(1,2)-L-x(1,2)-L} (Table 4) which were subsets of patterns from experimentally validated signal sequences from *T. brucei* (Table 3).

Although discovered, the significance of highly conserved peptide motifs is not yet completely unraveled. In order to understand the role of h-region peptide motifs in signal peptide activity, we disrupted both overlapping tri-peptide motifs ($S_1T_2M_3L_4T_5\mathbf{L_6}S_7\underline{\mathbf{L_8}}\underline{\mathbf{L_9}}Y_{10}\mathbf{A_{11}}I_{12}T_{13}P_{14}$) in VSG_117 signal peptide by converting two overlapping leucine residues (underlined) to two valines ($\mathbf{V_8V_9}$), two glycines ($\mathbf{G_8G_9}$), or two phenylalanines ($\mathbf{F_8F_9}$), (Table 8, Fig 2C). The results indicate that (i) disrupting the the h-region patterns is detrimental to function of signal peptide in *T. brucei*, and (ii) increasing the hydrophobicity of the h-region (LL to VV mutation) to compensate for pattern disruption does not work (Fig 2C, lanes 7 and 8).

In conclusion, we are reporting the discovery of conserved peptide motifs in the h-region of signal peptides across species. Our data indicates that hydrophobicity of a signal peptide may not be the most important factor in driving protein translocation across membrane. Further, unique peptide motifs in the h-region of signal peptides contribute to signal peptide activity. Last but not the least, amino acids are precisely positioned to attain maximum efficiency, and their distribution in the h-region of signal peptides is not random as previously thought [29].

REFERENCES

1. Walter, P. and A.E. Johnson, *Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane.* Annu Rev Cell Biol, 1994. **10**: p. 87-119.
2. Zheng, N. and L.M. Gierasch, *Signal sequences: the same yet different.* Cell, 1996. **86**(6): p. 849-52.
3. von Heijne, G. and L. Abrahmsen, *Species-specific variation in signal peptide design. Implications for protein secretion in foreign hosts.* FEBS Lett, 1989. **244**(2): p. 439-46.
4. Dev, I.K. and P.H. Ray, *Signal peptidases and signal peptide hydrolases.* J Bioenerg Biomembr, 1990. **22**(3): p. 271-90.
5. Rusch, S.L., et al., *Signal peptide hydrophobicity is finely tailored for function.* J Cell Biochem, 1994. **55**(2): p. 209-17.
6. Pugsley, A.P., *The complete general secretory pathway in gram-negative bacteria.* Microbiol Rev, 1993. **57**(1): p. 50-108.
7. Driessen, A.J., *How proteins cross the bacterial cytoplasmic membrane.* J Membr Biol, 1994. **142**(2): p. 145-59.
8. Rusch, S.L. and D.A. Kendall, *Protein transport via amino-terminal targeting sequences: common themes in diverse systems.* Mol Membr Biol, 1995. **12**(4): p. 295-307.
9. Rusch, S.L. and D.A. Kendall, *Protein transport via amino-terminal targeting sequences: common themes in diverse systems.* Mol.Membr.Biol., 1995. **12**: p. 295-307.
10. von Heijne, G., *Signal sequences. The limits of variation.* J Mol Biol, 1985. **184**(1): p. 99-105.
11. Borchert, T.V. and V. Nagarajan, *Effect of signal sequence alterations on export of levansucrase in Bacillus subtilis.* J Bacteriol, 1991. **173**(1): p. 276-82.
12. Doud, S.K., M.M. Chou, and D.A. Kendall, *Titration of protein transport activity by incremental changes in signal peptide hydrophobicity.* Biochemistry, 1993 Feb 9. **32**(5): p. 1251-6.

13. Izard, J.W., S.L. Rusch, and D.A. Kendall, *The amino-terminal charge and core region hydrophobicity interdependently contribute to the function of signal sequences.* J.Biol.Chem., 1996. **271**: p. 21579-21582.

14. Rusch, S.L., et al., *Signal peptide hydrophobicity is finely tailored for function.* J.Cell Biochem., 1994. **55**: p. 209-217.

15. McGeoch, D.J., *On the predictive recognition of signal peptide sequences.* Virus.Res., 1985. **3**: p. 271-286.

16. Ng, D.T., J.D. Brown, and P. Walter, *Signal sequences specify the targeting route to the endoplasmic reticulum membrane.* J.Cell Biol., 1996. **134**: p. 269-278.

17. Zheng, N. and L.M. Gierasch, *Signal sequences: the same yet different.* Cell, 1996. **86**: p. 849-852.

18. Al-Qahtani, A., M. Teilhet, and K. Mensa-Wilmot, *Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from Trypanosoma brucei and Leishmania.* Biochem.J., 1998. **331**: p. 521-529.

19. Button, L.L., et al., *Recombinant Leishmania surface glycoprotein GP63 is secreted in the baculovirus expression system as a latent metalloproteinase.* Gene, 1993. **134**(1): p. 75-81.

20. Ramirez, M.I., et al., *Heterologous expression of a Trypanosoma cruzi surface glycoprotein (gp82) in mammalian cells indicates the existence of different signal sequence requirements and processing.* J Eukaryot Microbiol, 1999. **46**(6): p. 557-65.

21. Bendtsen, J.D., et al., *Improved prediction of signal peptides: SignalP 3.0.* J Mol Biol, 2004. **340**(4): p. 783-95.

22. Nielsen, H., S. Brunak, and G. von Heijne, *Machine learning approaches for the prediction of signal peptides and other protein sorting signals.* Protein Eng, 1999. **12**(1): p. 3-9.

23. Tjalsma, H., et al., *Signal peptide-dependent protein transport in Bacillus subtilis: a genome-based survey of the secretome.* Microbiol Mol Biol Rev, 2000. **64**(3): p. 515-47.

24. Hertz-Fowler, C., et al., *GeneDB: a resource for prokaryotic and eukaryotic organisms.* Nucleic Acids Res, 2004. **32**(Database issue): p. D339-43.

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

26. Jonassen, I., *Efficient discovery of conserved patterns using a pattern graph.* Comput Appl Biosci, 1997. **13**(5): p. 509-22.

27. Izard, J.W. and D.A. Kendall, *Signal peptides: exquisitely designed transport promoters.* Mol.Microbiol., 1994. **13**: p. 765-773.

28. Martoglio, B. and B. Dobberstein, *Signal sequences: more than just greasy peptides.* Trends Cell Biol, 1998. **8**(10): p. 410-5.

29. Kaiser, C.A., et al., *Many random sequences functionally replace the secretion signal sequence of yeast invertase.* Science, 1987. **235**(4786): p. 312-7.

30. Bangs, J.D., et al., *A soluble secretory reporter system in Trypanosoma brucei - Studies on endoplasmic reticulum targeting.* J.Biol.Chem., 1996. **271**: p. 18387-18393.

31. Al-Qahtani, A., M. Teilhet, and K. Mensa-Wilmot, *Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from Trypanosoma brucei and Leishmania.* Biochem J, 1998. **331 (Pt 2)**: p. 521-9.

32. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J.Mol.Biol., 1982. **157**: p. 105-132.

FIGURE LEGEND

Figure 1: Pattern Description and Terminology

FC: fixed component; FC-1, fixed component in position 1; FC-2, fixed component in position 2; FC-3, fixed component in position 3; FC-4, fixed component in position 4; LR, linker region; LR1, linker region between FC-1 and FC-2; LR2, linker region between FC-2 and FC-3; LR3, linker region between FC-3 and FC-4

Table 1: Peptide Motifs in the h-region of Human Signal Peptides

(A) Highest scoring tetra-peptide motifs from PRATT analysis and unified tetra-peptide motifs**.** ***(B)*** PRATT alignment of the the h-region of human signal sequences. The first 25 hits are shown. ***(C)*** Highest scoring tri-peptide motifs from PRATT analysis and unified tetra-peptide motifs**.**

Table 2: Peptide Motifs in the h-region of *S. cerevisiae* Signal Peptides

(A) Highest scoring tri-peptide motifs from PRATT analysis and unified tri-peptide motifs

(B) PRATT alignment of the the h-region of *S. cerevisiae* signal sequences. The first 25 hits are shown.

Table 3: Peptide Motifs in the h-region of Experimentally Proven *T. brucei* Signal Peptides

(A) Highest scoring tri-peptide motifs from PRATT analysis and unified tri-peptide motifs

(B) PRATT alignment of the the h-region of *T. brucei* signal sequences. The first 25 hits are shown.

Table 4: Peptide Motifs in the h-region of VSGDB *T. brucei* Signal Peptides

(A) Highest scoring tri-peptide patterns from PRATT analysis and unified tri-peptide patterns

(B) PRATT alignment of the the h-region of VSG signal sequences. The first 25 hits are shown.

Table 5: Peptide Motifs in the h-region of *E. coli* Signal Peptides

(A) Highest scoring tri-peptide motifs from PRATT analysis and unified tri-peptide motifs

(B) PRATT alignment of the the h-region of *E. coli* signal sequences. The first 25 hits are shown.

Table 6: Peptide Motifs in the h-region of *B. subtilis* Signal Peptides

(A) Highest scoring tri-peptide motifs from PRATT analysis and unified tri-peptide motifs

(B) PRATT alignment of the the h-region of *B. subtilis* signal sequences. The first 25 hits are shown.

Table 7: Peptide Motifs in Random Sequences from Human Genome that were functional in *S. cerevisiae.*

(A) Highest scoring tri-peptide motifs from PRATT analysis and unified tri-peptide motifs

(B) PRATT alignment of the "functional sequences".

Table 8: VSG_117 h-region Mutations

The table enlists various h-region mutations in VSG_117 that were generated for this study.

Figure 2: Effect of Protein Translocation Inhibitors on TbRM

(A) Protocol Used for Import of proteins. ***(B and C)***. All the mutations are described in Table 8. Different mRNAs encoding VSG_117 were translated in rabbit reticulocyte lysate with 1.5 equivalents *T. brucei* cytosol for 60 min. The reaction mixtures were treated with cycloheximide (50 !g/ml, final concentration) and incubated with TbRM (one equivalent). The mixture was incubated for 60 min at 37^0C and digested with proteinase K digestion (30 !g/ml, final concentration) for 60 min on ice. Proteins were resolved by SDS-PAGE and detected by phosphorimaging. **(B) Lanes 1-4 (ΔN-VSG_117$_{500}$)**: Lane 1, untreated ΔN-VSG_117$_{500}$; lane 2, ΔN-VSG_117$_{500}$ treated with proteinase K; lane 3, ΔN-VSG_117$_{500}$ with TbRM; lane 4, ΔN-VSG_117$_{500}$ with TbRM treated with proteinase K digestion. **Lanes 5-8 ($G_6G_8G_9G_{11}$-h_{VSG_117})**: Lane 1, untreated $G_6G_8G_9G_{11}$-h_{VSG_117}; lane 2, $G_6G_8G_9G_{11}$-h_{VSG_117} treated with

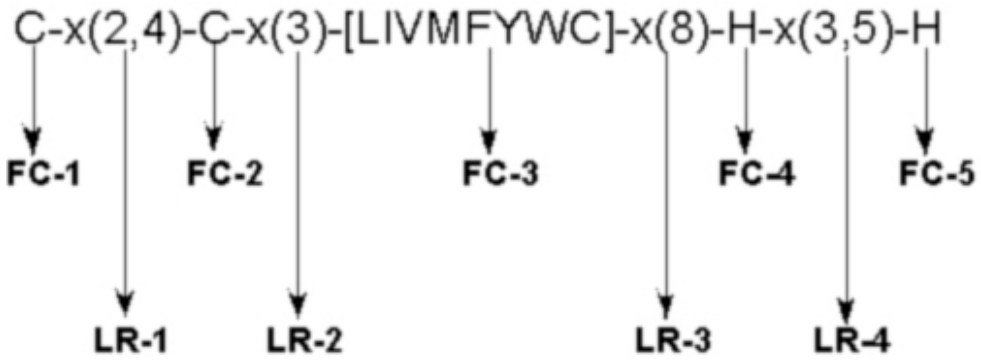
proteinase K; lane 3, $G_6G_8G_9G_{11}$-h_{VSG_117} with TbRM; lane 4, ΔNgggg-VSG_$117_{500}$with TbRM treated with proteinase K digestion. ***(C)*** **Lanes 1-2 (ΔN-VSG_117_{500})**: Lane 1, ΔN-VSG_117_{500} with TbRM; lane 2, ΔN-VSG_117_{500} with TbRM treated with proteinase K digestion. **Lanes 3-4 (G_8G_9-h_{VSG_117})**: Lane 3, mutant G_8G_9-h_{VSG_117} with TbRM; lane 2, mutant G_8G_9-h_{VSG_117} with TbRM treated with proteinase K digestion. **Lanes 5-6 (F_8F_9-h_{VSG_117})**: Lane 5, mutant F_8F_9-h_{VSG_117} with TbRM; lane 6, mutant F_8F_9-h_{VSG_117} with TbRM treated with proteinase K digestion. **Lanes 7-8 (V_8V_9-h_{VSG_117})**: Lane 7, mutant V_8V_9-h_{VSG_117} with TbRM; lane 8, mutant V_8V_9-h_{VSG_117} with TbRM treated with proteinase K digestion.

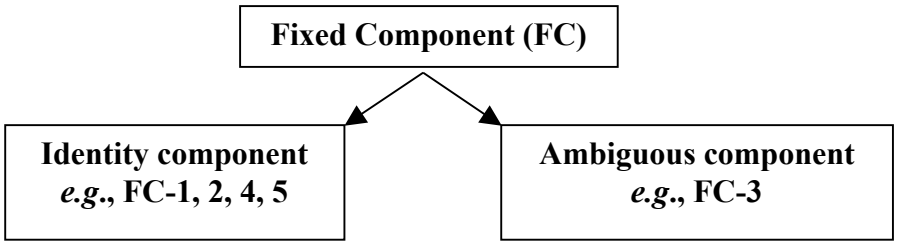
Rectangular brackets underneath sets of bars denote those data points that were directly compared in quantitation. Abbreviations: RRL, rabbit reticulocyte lysate; CHX, cycloheximide; TbRM, *T. brucei* microsomes; PK, proteinase K; PMSF, phenyl methyl sulfonyl fluoride.

Table 9: Comparison of Peptide Motifs Across Species

Comparison of all unified motifs across species. (A) Amino acid composition of most commonly occurring peptide motif in each species. (C) Linker regions of most commonly occurring peptide motif in each species. (C) A graph representing all linker regions across species.

Figure 1:


C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H
FC-1
FC-2
FC-3
FC-4
FC-5
LR-1
LR-2
LR-3
LR-4


Fixed Component (FC)
Identity component
e.g., FC-1, 2, 4, 5
Ambiguous component
e.g., FC-3

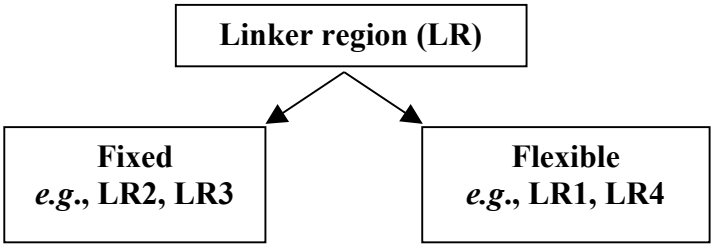

Linker region (LR)
Fixed
e.g., LR2, LR3
Flexible
e.g., LR1, LR4

TABLE 1:

A.

	Tetra-peptide Motif	Number of sequences (*)	Unified Motif
1	L-[AGLV]-x-L-x(0,1)-L	96 (164)	**L-[AGLV]-x(0,1)-L-x(0,2)-L**
2	L-[AGLV]-L-x(1,2)-L	99 (178)	
3	L-x(2)-[AGILPV]-L-x(0,2)-L	96 (160)	**L-x(2)-[AGILPV]-L-x(0,2)-L**

* Combined number of times the pattern occurs in the total sequences they are found

B.

L-[AGLV]-x-L-x(0,1)-L

```
1A01:        t LLlL-L sgala
A1AG:    ltvls LLpL-L
A2HS:        s LVlL-L cla
A2MG:    slvll LLvL-L p
  A4:     mlpg LAlL-L laa
ACHB:    mtpga LLmL-L galgp
 ALS:     ggla LAlL-L lswva
ANGI:     mvmg LGvL-L lvfvl
ANPA:          LLlL-L llppl
APA1:    aavlt LAvLfL tgsqa
 APB:       pa LLaL-L alpal
APC2:    lpalf LVlLvL g
APC3:    llvva LLaL-L asa
 APD:       vm LLlL-L salag
B2MG:      sva LAvLaL lslsg
 BAL:       lq LVvLgL tccwa
BGAL:        I LLlL-L vllll
C1QB:    ipvlm LLlL-L gli
C1QC:          LLlL-L lllal
 C1R:        w LLyL-L vpalf
CA11:          LLlL-L aatal
CAH4:       ml LAlLaL saa
CASB:    lilac LVaLaL a
CC10:    avtlt LVtLaL ccssa
DRN1:    galla LAaL-L ggavs
```

C

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-x-L-L	103 (91)	**L-x(0,5)-[L]-x(0,4)-L**
2	L-L-L	101 (218)	
3	L-L-x-L	107 (192)	
4	L-x(2)-L-x(1,2)-L	105 (184)	
5	L-x(4)-L-x(0,1)-L	100 (155)	
6	L-L-x(2,3)-L	109 (213)	
7	L-L-x(3,4)-L	105 (177)	
8	L-x(3)-L-x(0,1)-L	99 (163)	
9	L-x(3,4)-L-x-L	97 (163)	
10	L-x-L-x(2,3)-L	115 (195)	
11	L-x-L-x(3,4)-L	96 (146)	
12	L-x(3,4)-L-L	97 (174)	
13	L-x(2)-L-x(2,4)-L	100 (180)	
14	L-x(4)-L-x(1,3)-L	100 (143)	
15	L-x(4,5)-L-x(1,2)-L	100 (175)	
16	L-x(3,4)-L-x(2,3)-L	97 (172)	
17	L-[AL]-x(2)-L	107 (185)	**L-[AL]-x(2,3)-L**
18	L-[AL]-x(3)-L	97 (150)	
19	L-[ALV]-x(4)-L	100 (151)	**L-[ALV]-x(4)-L**

***** Combined number of times the pattern occurs in the total sequences they are found

TABLE 2

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-x(0,1)-L-x(0,3)-L	46 (77)	
2	L-x(0,1)-L-x(1,4)-L	43 (77)	**L-x(0,1)-L-x(0,4)-L**
3	L-x(4,5)-L-[AGILV]	44 (62)	**L-x(4,5)-L-[AGILV]**
4	L-x(0,2)-S-x(0,3)-A	48 (69)	
5	L-x(0,3)-S-x(0,2)-A	43 (66)	
6	L-x(1,2)-S-x(0,4)-A	43 (56)	
7	L-x(1,3)-S-x(0,3)-A	46 (65)	**L-x(0,4)-S-x(0,4)-A**
8	L-x(2,4)-S-x(0,4)-A	43 (69)	
9	L-x(1,2)-S-[AGILV]	43 (54)	**L-x(1,2)-S-[AGILV]**
10	L-x(0,3)-S-x(0,4)-L	43 (88)	**L-x(0,3)-S-x(0,4)-L**

* Combined number of times the pattern occurs in the total sequences they are found

B.

L-x(0,3)-S-x(0,4)-A

```
  EUG1: ivsfc LfaSftlA
   CPY: sllcg LglSttlA
  PDI1: swssl LlaSsvfA
  ERP2: vlila LvnSv--A a
  YOS9:  iiya L--S---A isali
  SED1:  lstv Ll-S---A glast
  KEX1: wlgtw LamS---A li
 SEC11: lnvcf LfaS---A ym
ASP3-1:  slnt LllSlfvA mssg
  ERJ5: lvvlg Lv-SlsyA
  ERJ5: lglvs L--Sy--A
  FET3:  mtna Ll-Si--A vllfs
  PRM4:       Lt-Si--A agvaa
  DAN1: avaaa LvaS---A t
   PAU:       Lt-Si--A agvaa
  AWA1: alalv Ly-Sqs-A lg
  SOP4: fsqiv LllS---A fiyva
  PRB1: ftlga Lg-Sis-A
  CIS3:    va LaaSv--A alsat
  CIS3: asvaa L--S---A tasa
  FRE1:  vlfc LfiSff-A tvqs
  TIP1:  iafv L--S---A iasla
 PGU1: isans LliStlcA faia
  PAU7:       Lt-Si--A agvaa
  PRY2:   vsl LaaS---A svals
```

TABLE 3:

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-L-x-[AILV]	26 (28)	**L-L-x-[AILV]**
2	L-x(2,3)-L-[AILPV]	26 (34)	**L-x(2,3)-L-[AILPV]**
3	L-x(2,3)-L-x(0,2)-L	22 (30)	**L-x(1,3)-L-x(0,2)-L**
4	L-x(1,2)-L-x(0,2)-L	20 (32)	
5	L-[AILV]-L	21 (29)	**L-[AILV]-x(0,1)-L**
6	L-[AILV]-x-L	20 (28)	
7	L-x(0,2)-A-x(1,3)-L	20 (37)	**L-x(0,3)-A-x(0,3)-L**
8	L-x(1,3)-A-x(0,2)-L	20 (38)	
9	L-x(0,2)-A-x(0,2)-L	22 (40)	
10	L-x(1,2)-A-[AILV]	21 (32)	**L-x(1,3)-A-[AILGV]**
11	L-x(2,3)-A-[AGLV]	20 (26)	
12	A-x(2,3)-L-x-[AGILPV]	20 (26)	**A-x(2,3)-L-x-[AGILPV]**
13	V-x(0,1)-L-x(2)-[AGILV]	20 (25)	**V-x(0,1)-L-x(2)-[AGILV]**

* Combined number of times the pattern occurs in the total sequences they are found

B

```
                 L-L-x-[AILV]
ESAG4_BRUCEI:   yvmyv LLlL mpypl
ESAG7_BRUCEI:    fwfv LLaL lc
ESAG6:           fwfv LLaL lg
VSG117:         mltls LLyA itp
VSG221:         slian LLtA lvalt
MITAT1.1:           g LLqV vwqpi
ILTAT1.21:        nqg LLvV iaqli
ILTAT1.24:      qtisa LLiL i
ANTAT1.1:        fitq LLvL lvtmt
MVAT4:          vvlvq LLsA q
MVAT7:          atscv LLiI gctny
BARP:           nnlwl LLtV lcta
GP63-1:         yiipc LLgL isc
MVSG4:          tagav LLlA taavl
GPEET2:           sly LLaV llfsa
P67:            savlc LLlA atltl
```

TABLE 4:

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-A-x(0,1)-L	11(11)	**L-x(0,1)-[AL]-x(0,2)-L**
2	L-[AL]-x-L	13 (15)	
3	L-L-x(0,2)-L	12 (16)	
4	L-x(0,1)-L-x(0,1)-L	11(19)	
5	A-x(1,2)-L-x(1,2)-L	11(12)	**A-x(1,2)-L-x(1,2)-L**

* Combined number of times the pattern occurs in the total sequences they are found

B

L-[AL]-x-L

```
V002-001-VSG9.0:    avs LLlL alasa
V002-008-VSG.10:      a LAtL tvltt
V002-011-VSG10.0: tltlq LLqL g
V002-015-VSG6:     cygy LAaL tfvvs
V002-015-VSG18:   lgfaq LAtL ivvls
V002-015-VSG61:   lyqml LAiL slist
V002-005-VSG4:       va LLlL lnvaa
V002-005-VSG12:      vq LAvL aatit
V002-005-VSG14Bd: ifacv LAaL q
V002-005-VSG28:       f LLaL slvfg
V002-005-VSG29:    iyiv LLaL falp
V002-005-VSG30:   alcit LAtL ilaav
```

TABLE 5:

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	A-x(0,2)-L-x(0,3)-A	56 (96)	**A-x(0,4)-L-x(0,5)-A**
2	A-x(2,4)-L-x(0,5)-A	56 (86)	
3	A-x(1,3)-L-x(0,5)-A	59 (93)	
4	L-x(1,2)-A-[AGILV]	57 (80)	**L-x(1,2)-A-[AGILV]**
5	L-x(0,1)-A-x(0,5)-L	60 (92)	**L-x(0,5)-A-x(0,5)-[AL]**
6	L-x(1,2)-A-x(0,5)-L	56 (69)	
7	L-x(0,1)-A-x(0,5)-A	56 (97)	
8	L-x(3,5)-A-x(0,5)-A	56 (101)	
9	L-x(0,1)-L-x(0,5)-A	56 (84)	**L-x(0,5)-L-x(0,5)-A**
10	L-x(2,4)-L-x(0,4)-A	57 (99)	
11	L-x(3,5)-L-x(0,5)-A	59 (99)	

* Combined number of times the pattern occurs in the total sequences they are found

B

A-x(1,3)-L-x(0,5)-A

```
 AGP: iaaav AgivLl----A snaqa
AMPC:  ttlc Al--Lit---A scstf
ARAF:    al AaigL-----A avmsq
ARGT:   sil AlslLvglstA
ASG2: mgfsg Aa--L-----A l
BLA2: iisll AtlpL-----A V
BLP2: ayvii Ac--Lsst--A lags
BTUB:       Asl-Lt----A csvta
  C2:   sll Ai--L-----A vsslv
CYPH: laama AvfaLs----A lspaa
DACC: laags AflfLf----A ptaf
ECOT:  tilp Av--Lf----A afatt
ELT1: ffill Asp-Ly----A
ELT3: ffill Asp-Ly----A
ELT6: yvlft Al--Lssly-A
ELT7: yvlft Al--Lsslc-A
FADL: salav Ava-Listq-A
FAEF:   tmm Aaa-Lvls--A lsiqs
FANC:   tll Aii-Lggm--A fattn
FECA: sllpl Ag--Lsfs--A faa
```

TABLE 6:

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-x(0,1)-L-x(1,4)-L	49 (85)	**L-x(0,1)-L-x(0,5)-L**
2	L-x(0,1)-L-x(2,5)-L	47 (77)	
3	L-x(0,1)-L-x(0,3)-L	50 (82)	
4	L-x(3,4)-L-x(0,5)-L	46 (78)	**L-x(3,5)-L-x(0,5)-L**
5	L-x(4,5)-L-x(0,5)-L	46 (67)	
6	L-x-[AGILV]-x(2)-L	45 (59)	**L-x-[AGILV]-x(2)-L**
7	L-x(0,2)-A-x(0,4)-L	45 (74)	**L-x(0,3)-[AI]-x(0,5)-L**
8	L-x(0,3)-A-x(0,3)-L	45 (84)	
9	L-x(1,3)-A-x(0,5)-L	45 (75)	
10	L-x(0,3)-I-x(0,4)-L	46 (90)	
11	L-x(0,2)-I-x(0,5)-L	46 (80)	
12	L-x(0,4)-V-x(0,4)-L	47 (92)	**L-x(0,5)-V-x(0,4)-L**
13	L-x(1,5)-V-x(0,4)-L	45 (89)	
14	L-x(1,2)-A-[AGLV]	48 (70)	**L-x(0,2)-A-[AGLV]**
15	L-x(0,1)-A-[AGLV]	45 (59)	
16	L-x(1,3)-A-x(0,5)-A	44 (60)	**L-x(1,3)-A-x(0,5)-A**
17	L-x(0,1)-L-x(0,4)-A	45 (65)	**L-x(0,1)-L-x(0,5)-A**
18	L-x(0,1)-L-x(1,5)-A	45 (72)	

* Combined number of times the pattern occurs in the total sequences they are found

B

L-x(0,5)-L-x(0,5)-L

```
  AmyE:      ts L-Lp---L fagfl
  BglS:       v L-Ll---L vtglf
  BglC:   ifitc L-Lit--L ltmgg
  CccA:   plipf L-Liav-L giglt
  CwlD:    wlsf L-LgfiiL lflfk
  DacF:         L-Lst--L ligim
  FilL:    lmii L-Lii--L iviga
  Glpq:       i baLfv--L slgll
  LipB:   afiic LsLi---L svlaa
  LYTC:   ltmcf LgLi---L fvpt
   Mdr:     fvv LgLl---L gilms
  MotB:   pyadi btLl---L alfiv
   Mpr:   ayltv LcLa---L aaavs
  MreC:         LmLl---L lciii
  NprB:      ts LlLag--L ctaaq
  PbpD:   iigwi LlLciipL faft
   Pel:   mlata LfLg---L tpagp
  PenP:   gicvg L-Lc---L sitgf
  PhoB:         L-LpiavL ssiaf
  PhrF:         LlLsc--L alstv
  PhrK:         LvLcvsiL avils
SpoIID:   svlca LiLlvptL lvipf
  TasA:   vasaa LgLa---L vggg
  TyrA:      ti L-Lag--L gligg
   Vpr:       f L-LvsfvL ffals
  YbbR:     iia L-Lfa--L llyva
```

TABLE 7:

A

	Tri-peptide Motif	Number of sequences (*)	Unified Motif
1	L-x(0,2)-S-x(0,5)-L	12 (18)	
2	L-x(0,3)-S-x(0,4)-L	12 (22)	**L-x(0,3)-S-x(0,5)-L**
3	L-x(0,3)-S-x(1,5)-L	12 (22)	
4	L-x(1,5)-F-x(0,4)-L	12 (22)	
5	L-x(0,4)-F-x(0,5)-L	12 (22)	
6	L-x(0,5)-F-x(0,4)-L	12 (26)	**L-x(0,5)-F-x(0,5)-L**

* Combined number of times the pattern occurs in the total sequences they are found

B

L-x(0,3)-S-x(0,4)-L

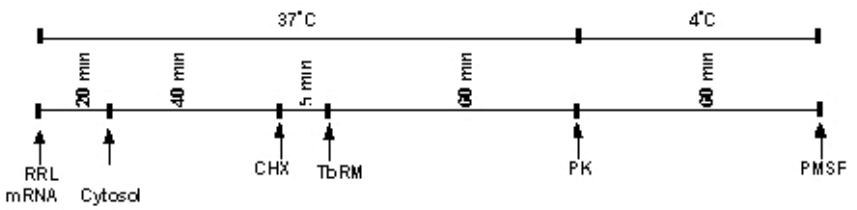
```
102:   vqwry Lg--Spqp-L pr
204:   clclh L---S----L pvplp
210:   rkpfl L---S----L wtqql
501:      pr LfycSnts-L cvlql
502:       p Lc--Svp--L fyvsv
503:   ffmnp LqlfSfi--L lktg
601:      pq L---S----L fvphf
801:      pp L---Shrg-L ptsgp
901:   fdllf LvclSi---L hteqn
903:   ppgyi L---Swii-L qshg
1001:  tcfyk L---S----L lehg
1010:  rifsi LflfSnfsfL knnnn
```

TABLE 8:

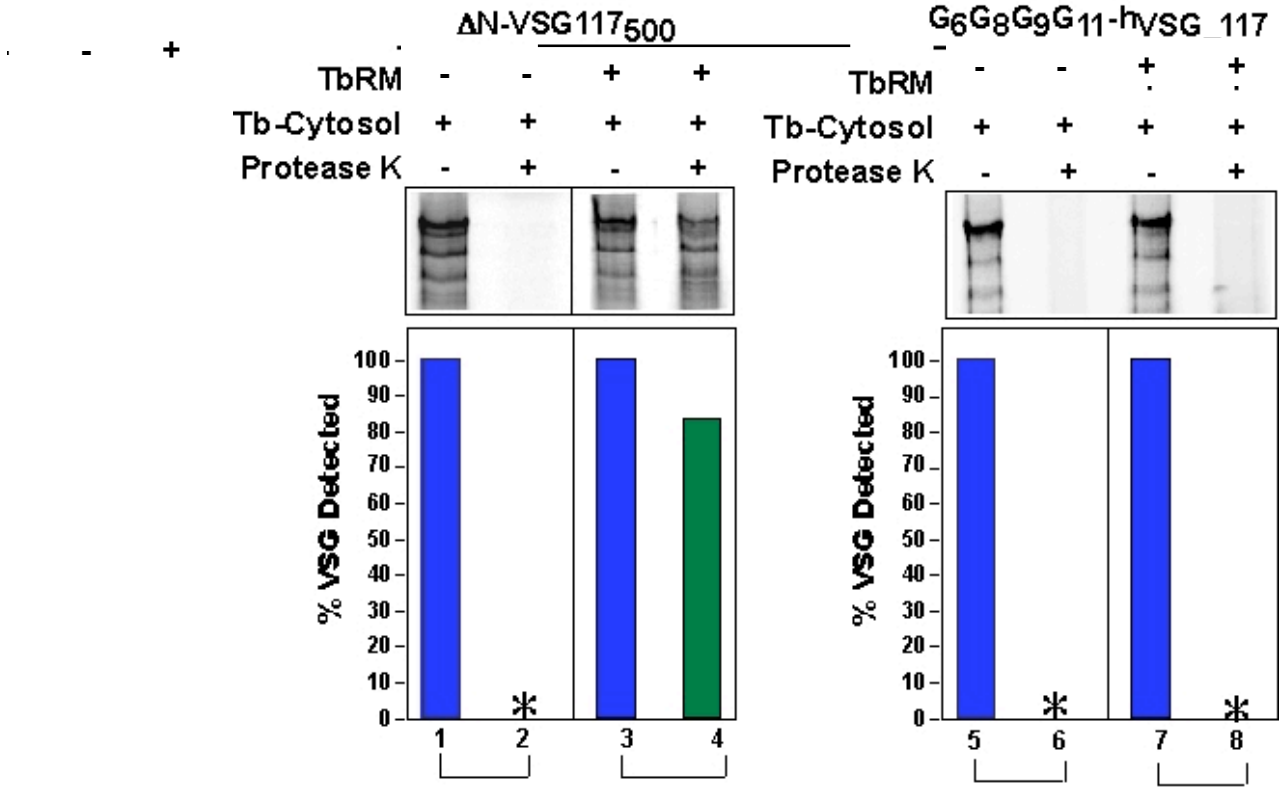
	Name	Amino Acid Sequence
Wild Type		$S_1T_2M_3L_4T_5\mathbf{L_6}S_7\mathbf{L_8L_9}Y_{10}\mathbf{A_{11}}I_{12}T_{13}P_{14}$
Mutant	$G_5G_8G_9$	$S_1T_2M_3L_4T_5\mathbf{G_6}S_7\mathbf{G_8G_9}Y_{10}\mathbf{G_{11}}I_{12}T_{13}P_{14}$
	G_8G_9	$S_1T_2M_3L_4T_5L_6S_7\mathbf{G_8G_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$
	F_8F_9	$S_1T_2M_3L_4T_5L_6S_7\mathbf{F_8F_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$
	V_8V_9	$S_1T_2M_3L_4T_5L_6S_7\mathbf{V_8V_9}Y_{10}A_{11}I_{12}T_{13}P_{14}$

FIGURE 2:

A


37°C
4°C
20 min
40 min
5 min
60 min
60 min
RRL mRNA
Cytosol
CHX
TbRM
PK
PMSF

B


ΔN-VSG117500
G6G8G9G11-hVSG 117
TbRM
Tb-Cytosol
Protease K
% VSG Detected
1
2
3
4
5
6
7
8

C

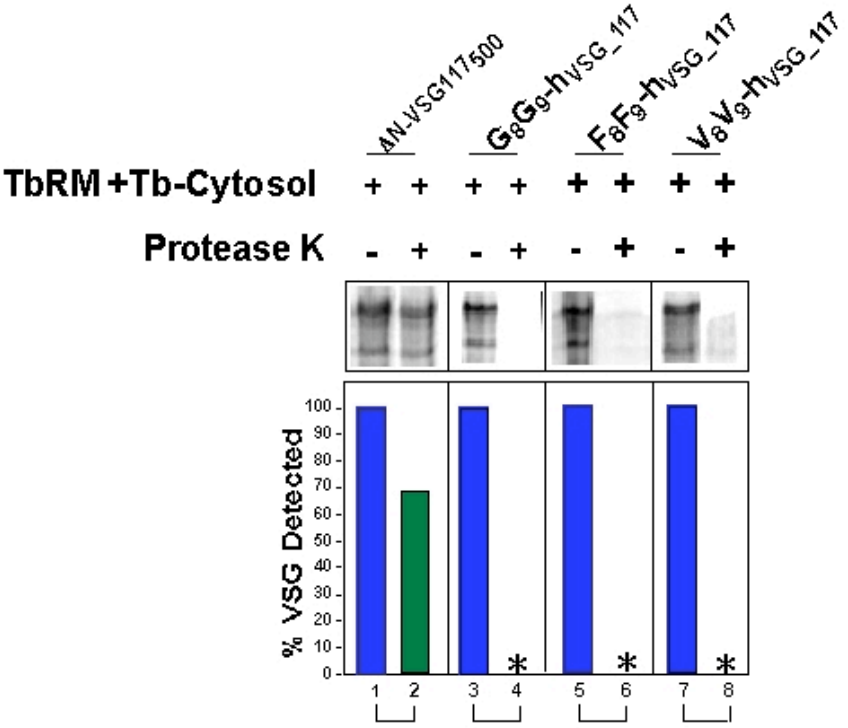

ΔN-VSG117s00
G8G9-hVSG_117
F8F9-hVSG_117
V8V9-hVSG_117
TbRM +Tb-Cytosol
Protease K
% VSG Detected
100
90
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8

TABLE 9:

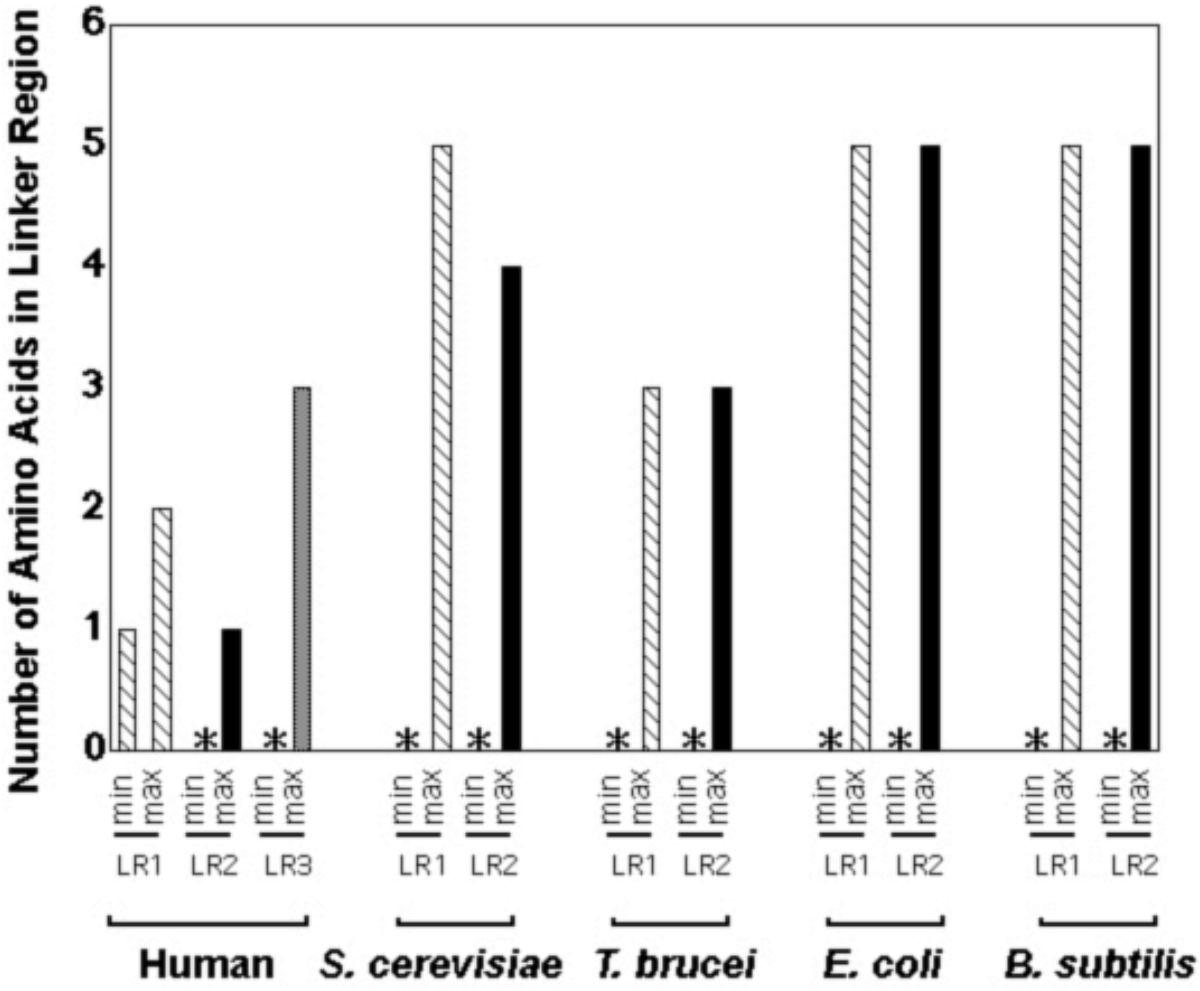
A

Fixed Component	Human	*S. cerevisiae*	*T. brucei*	*E. coli*	*B. subtilis*
FC-1	L	L	L, A, V	A. L	L
FC-2	L, A, V, G, I, P	S, L	A, I, L, V	A, L	A, G, I, L, V
FC-3	L	A, G, I, L, V	L, A, V, P, I	A, L, V, G, I	A, L, V, G,
FC-4	L	-	-	-	-

B

Linker Region	Human	*S. cerevisiae*	*T. brucei*	*E. coli*	*B. subtilis*
LR1	0	0-1	1-3	0-1	0-4
LR2	0-1	0-4	0-2	0-2	0-5
LR3	0-2	-	-	-	-

C


Number of Amino Acids in Linker Region
0
1
2
3
4
5
6
min
max
LR1
LR2
LR3
Human
S. cerevisiae
T. brucei
E. coli
B. subtilis

CHAPTER 5

DISCUSSION

1. Protein Translocation into the Endoplasmic Reticulum in a Trypanosome

The endoplasmic reticulum (ER) acts as a gateway for proteins entering the secretory pathway. Proteins may be targeted to the endoplasmic reticulum co-translationally [1] or post-translationally [2-4]. Both pathways converge at a protein translocation channel (Sec61) at the ER membrane [5, 6]. Requirements for both pathways are distinct. The co-translational pathway, in vertebrates and yeast, is dependent on signal recognition particle (SRP) [7]. In contrast, post-translational translocation is carried out with assistance of molecular chaperones (Hsp70 family e.g., Ssa1p) and cochaperones (Hsp40 family e.g., Ydj1p) [3, 8-10]. In trypanosomes, the functions of the ER have not been extensively studied. Like other systems, a signal peptide is essential for trypanosome proteins to enter the ER [11]. Inhibition of SRP expression in *T. brucei* did not significantly affect ER protein import, suggesting the existence of a post-translational pathway [12].

As a part of our study, we established the existence of a post-translational pathway in a trypanosome by developing a cell-free protein import system using *T. brucei* microsomes (TbRM) and model *T. brucei* protein (variant surface glycoprotein (VSG)). Proteins (e.g., VSG_117) are imported into TbRM post-translationally in a parasite cytosol-dependent fashion [Chapter Two]. The signal peptide is necessary for protein import. *T. brucei* cytosol could be replaced by purified *S. cerevisiae* Ssa1p, further confirming the dependence of protein import on molecular chaperones of the Hsp70 class of proteins.

2. The Secretory Pathway as a Target for anti-Parasite Drug Discovery

Recently, RNA interference of a VSG in bloodstream *T. brucei* resulted in cell cycle arrest and inhibition of cell growth *in vitro,* leading the authors to the propose that parasites with a compromised VSG coat may be easily eliminated by the host immune system. [13].

The synthetic compound MAL3-101 inhibits Ssa1p (Hsp70) ATPase activity *in vitro* in *S. cerevisiae* [14]. In *T. brucei,* MAL3-101 inhibits import of VSG_117 into TbRM [Chapter Three]. Two other compounds that inhibit protein translocation into TbRM include; equisetin [15], and CJ-21,058 [16] [Chapter Three]. Both these compounds contain a tetrameric acid as the central ring, which is known to be biologically active [15-17].

All three protein translocation blockers (PTBs) (MAL3-101, equisetin and CJ-21,058) proved to be effective trypanocidal agents that had no adverse effect on human HeLa cell viability [Chapter Three]. Structure-activity studies will be needed to ascertain whether the tetrameric ring is needed for the activity of the compounds against protein translocation into the ER of *T. brucei.*

We infer from these data that the secretory pathway of trypanosomes is a good target for discovery of anti-trypanosome compounds. Further studies on i) MAL3 compounds, and ii) equisetin derivatives, may lead to the discovery of lead compounds against HAT.

3. Identification of Conserved Peptide Motifs in Signal Peptides

Signal sequences have three conserved regions: an N-terminal (n-region); a hydrophobic core (h-region); and a C-terminal region (c-region) (reviewed in [18, 19]). Hydrophobicity of the h-region is important for function of a signal sequence [20]. Furthermore, signal sequences with higher hydrophobicity direct co-translational protein import into the ER, while less hydrophobic h-regions direct post-translational ER import [21, 22].

Signal peptides are believed to drive protein import into the ER across species [23]. Surprisingly, trypanosomatid proteins (*T. brucei* VSG117, MVAT7, BiP, and *Leishmania* GP63) are not imported into canine microsomes [24] (Fig. 1A). Conversely, yeast, bacterial and bovine proteins are not imported into

T. brucei microsomes (Fig. 1B) [Chapter Two]. The specificity of translocons for signal peptides cannot be explained by hydrophobicity alone; since hydrophobicity of all the signal peptides studied were comparable (Fig. 2) [Chapter Two].

There is increasing evidence that factors other than the hydrophobicity of the h-region influence signal peptide function [24]. Sequence alignments have been used to document that signal peptides do not have conserved sequences. Due to its direct importance in signal sequence function, we chose to focus strictly on the h-region. Our bioinformatics approach has unearthed highly conserved peptide motifs in *Homo sapiens*, *T. brucei, E. coli, S. cerevisiae,* and *B. subtilis* h-regions [Chapter Four]. Variations in peptide motifs across different species were observed in: i) amino acid sequence of motifs, and ii) length of linker regions that connect the conserved amino acid residues that constitute the motifs [Chapter Four]. We hypothesize that the conserved peptide motifs contribute to signal peptide activity. To test our hypothesis, we disrupted the conserved peptide motifs in signal sequence of a *T. brucei* model protein. Disruption in the motif led to a loss of signal peptide activity in *T. brucei,* thus validating our hypothesis [Chapter four].

Understanding *T. brucei* Signal Peptides and their Role in ER Protein Translocation

Finally, this work has led us to a better understanding of signal peptides in trypanosomes. It was known earlier that signal peptides are essential in *T. brucei* for ER protein import. *T. brucei* signal sequences maintain a classical tri-partite structure [24]. Trypanosome signal peptides direct post-translational protein import into the ER, in cytosol-dependent fashion [Chapter Two]. *T. brucei* signal peptides are essential for post-translational protein translocation into ER. Further, *T. brucei* signal sequences that efficiently direct protein import into TbRM cannot function in canine microsomes [24]. Since signal peptides are selectively recognized by the translocation machineries, we propose that signal peptides have “co-evolved” with their translocons [Chapter Two].

Previous to this study, not much was known about the role of each sub-region of *T. brucei* signal peptides in protein translocation into ER. We have now established that the n-region of *T. brucei* signal peptide is not essential for protein import into the ER [Chapter Four]. Further, bioinformatic analysis revealed the presence of "unique peptide motifs" in h-region of signal peptides in *T. brucei* [Chapter Four]. Mutations created in the *T. brucei* h-region that disrupt the conserved peptide motifs therein, abolished VSG_117 import into TbRM [Chapter Four]. Hence the h-region tri-peptide motifs are important in signal sequence activity in *T. brucei.*

Concluding Remarks

ER protein translocation in *T. brucei* can occur: i) post-translationally; ii) is chaperone dependent (Ssa1p); and iii) needs a functional signal peptide h-region. In our pursuit of drug discovery, we have found compounds that inhibit ER protein import, which are effective trypanocidal agents (e.g., MAL3-101, equisetin and CJ-21,058). Finally, we report that h-region of signal sequences from different species contains unique peptide motifs that may be crucial for their function.

REFERENCES

1. Lutcke, H., *Signal recognition particle (SRP), a ubiquitous initiator of protein translocation.* Eur.J.Biochem., 1995. **228**: p. 531-550.
2. Chirico, W.J., M.G. Waters, and G. Blobel, *70K heat shock related proteins stimulate protein translocation into microsomes.* Nature, 1988. **332**(6167): p. 805-10.
3. Ngosuwan, J., et al., *Roles of cytosolic Hsp70 and Hsp40 molecular chaperones in post-translational translocation of presecretory proteins into the endoplasmic reticulum.* J Biol Chem, 2003. **278**(9): p. 7034-42.
4. Rapoport, T.A., et al., *Posttranslational protein translocation across the membrane of the endoplasmic reticulum.* Biol Chem, 1999. **380**(10): p. 1143-50.
5. Wilkinson, B.M., et al., *Distinct domains within yeast Sec61p involved in post-translational translocation and protein dislocation.* J Biol Chem, 2000. **275**(1): p. 521-9.
6. Johnson, A.E. and M.A. van Waes, *The translocon: a dynamic gateway at the ER membrane.* Annu Rev Cell Dev Biol, 1999. **15**: p. 799-842.
7. Nagai, K., et al., *Structure, function and evolution of the signal recognition particle.* Embo J, 2003. **22**(14): p. 3479-85.
8. Panzner, S., et al., *Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p.* Cell, 1995. **81**(4): p. 561-70.
9. Misselwitz, B., O. Staeck, and T.A. Rapoport, *J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences.* Mol Cell, 1998. **2**(5): p. 593-603.
10. Lyman, S.K. and R. Schekman, *Interaction between BiP and Sec63p is required for the completion of protein translocation into the ER of Saccharomyces cerevisiae.* J Cell Biol, 1995. **131**(5): p. 1163-71.
11. Bangs, J.D., et al., *A soluble secretory reporter system in Trypanosoma brucei - Studies on endoplasmic reticulum targeting.* J.Biol.Chem., 1996. **271**: p. 18387-18393.

12. Liu, L., et al., *RNA interference of signal peptide-binding protein SRP54 elicits deleterious effects and protein sorting defects in trypanosomes.* J Biol Chem, 2002. **277**(49): p. 47348-57.

13. Sheader, K., et al., *Variant surface glycoprotein RNA interference triggers a precytokinesis cell cycle arrest in African trypanosomes.* Proc Natl Acad Sci U S A, 2005. **102**(24): p. 8716-21.

14. Fewell, S.W., et al., *Small Molecule Modulators of Endogenous and Co-chaperone-stimulated Hsp70 ATPase Activity.* J Biol Chem, 2004. **279**(49): p. 51131-40.

15. Vesonder, R.F., et al., *Equisetin, an antibiotic from Fusarium equiseti NRRL 5537, identified as a derivative of N-methyl-2,4-pyrollidone.* J Antibiot (Tokyo), 1979. **32**(7): p. 759-61.

16. Sugie, Y., et al., *CJ-21,058, a new SecA inhibitor isolated from a fungus.* J Antibiot (Tokyo), 2002. **55**(1): p. 25-9.

17. Burke, L.T., et al., *A short stereoselective total synthesis of the fusarium toxin equisetin.* Org Lett, 2000. **2**(23): p. 3611-3.

18. Izard, J.W. and D.A. Kendall, *Signal peptides: exquisitely designed transport promoters.* Mol.Microbiol., 1994. **13**: p. 765-773.

19. Martoglio, B. and B. Dobberstein, *Signal sequences: more than just greasy peptides.* Trends Cell Biol, 1998. **8**(10): p. 410-5.

20. McGeoch, D.J., *On the predictive recognition of signal peptide sequences.* Virus.Res., 1985. **3**: p. 271-286.

21. Ng, D.T., J.D. Brown, and P. Walter, *Signal sequences specify the targeting route to the endoplasmic reticulum membrane.* J.Cell Biol., 1996. **134**: p. 269-278.

22. Zheng, N. and L.M. Gierasch, *Signal sequences: the same yet different.* Cell, 1996. **86**: p. 849-852.

23. Garcia, P.D. and P. Walter, *Full-length prepro-alpha-factor can be translocated across the mammalian microsomal membrane only if translation has not terminated.* J.Cell Biol., 1988. **106**: p. 1043-1048.

24. Al-Qahtani, A., M. Teilhet, and K. Mensa-Wilmot, *Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from Trypanosoma brucei and Leishmania.* Biochem.J., 1998. **331**: p. 521-529.

FIGURE LEGEND

Figure 1: Correlation of Signal Sequence Hydrophobicity and Protein Import into TbRM

(A) Selective Recognition of Signal Peptides by Vertebrate Translocation Machinery.

(B) Selective Recognition of Signal Peptides by *T. brucei* Translocation Machinery.

Figure 2: Kyte-Dolittle hydropathy Plots of Signal Sequences

Peak hydrophobicity of signal peptides across species as calculated using the Kyte-Dolittle hydropathy scale.

FIGURE 1:

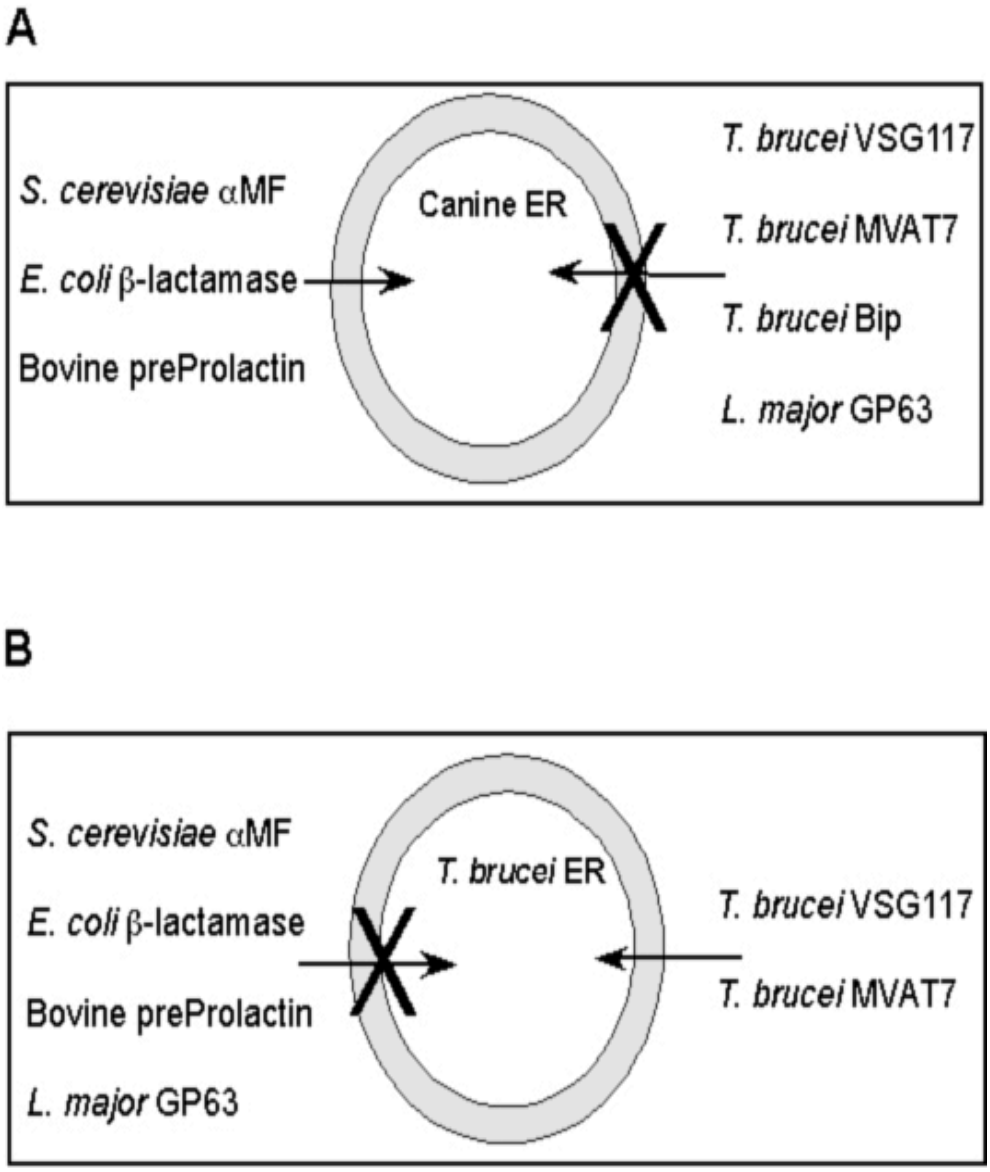
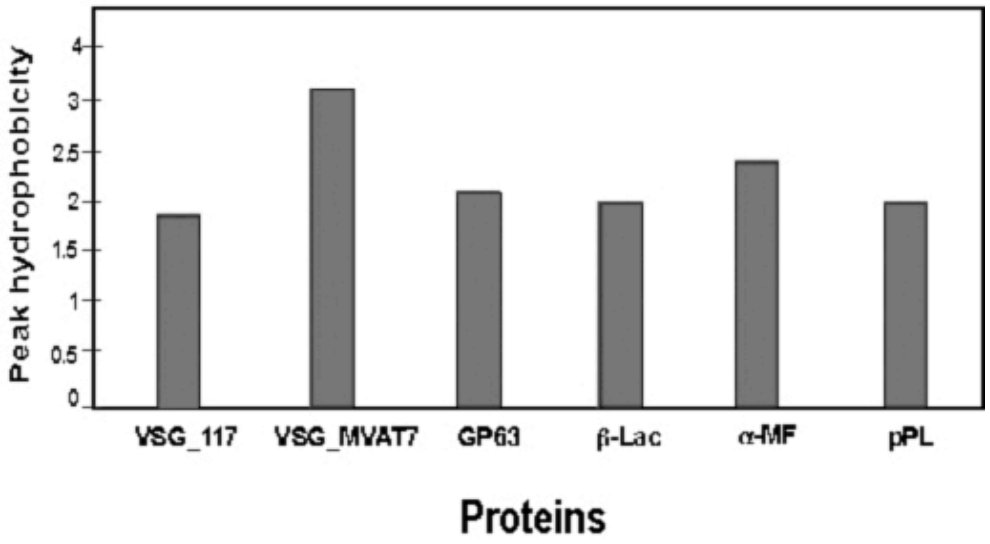

A
S. cerevisiae αMF
E. coli β-lactamase
Bovine preProlactin
Canine ER
T. brucei VSG117
T. brucei MVAT7
T. brucei Bip
L. major GP63
B
S. cerevisiae αMF
E. coli β-lactamase
Bovine preProlactin
L. major GP63
T. brucei ER
T. brucei VSG117
T. brucei MVAT7

FIGURE 2:


Peak hydrophobicity
4
3
2.5
2
1.5
1
0.5
0
VSG_117
VSG_MVAT7
GP63
β-Lac
α-MF
pPL
Proteins