Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from *Trypanosoma brucei* and *Leishmania*

Ahmed AL-QAHTANI¹, Meredith TEILHET¹ and Kojo MENSA-WILMOT²

Department of Cellular Biology, University of Georgia, 724 BioSciences, Athens, GA 30602, U.S.A.

N-Terminal signal peptides direct secretory and most membrane proteins into the exocytic pathway at the endoplasmic reticulum. Signal sequences can function across kingdoms. However, our attempts at translocating variant surface glycoprotein (VSG) 117, VSG MVAT7, VSG 221 and BiP from *Trypanosoma brucei* and gp63 from *Leishmania chagasi* into canine pancreas microsomes failed. On replacing the signal peptide of VSG 117 with that from yeast prepro- α -mating factor (pp α MF) the chimaeric protein was imported, indicating that the signal sequence of VSG 117 was incompatible with the protein-import machinery of mammalian microsomes. Replacement of the gp63-h-region with a hybrid composed of the N-terminal nine residues from the h-region of gp67 from *Autographa californica* nuclear polyhedrosis virus and the C-terminal 10 residues from the h-region of gp63

from *L. major* produced a functional signal peptide. Thus, the h-region of kinetoplastid signal peptides appears to be the sub-domain that is non-functional at the mammalian translocon. The calculated biophysical properties and computed discriminant scores (predictive of importability of signal peptides into mammalian microsomes) of the kinetoplastid signal sequences nevertheless are similar to those of pp α MF and *Escherichia coli* β -lactamase both of which were imported. These signal peptides are the first collection from one biological family that have been found to fail to function across a species barrier. They indicate that signal peptides are not as universally interchangeable as previously believed. Intriguingly, endoplasmic reticulum signal peptides from *Leishmania* and *Crithidia fasciculata* are reminiscent of signal peptides from Gram-positive bacteria.

INTRODUCTION

The endoplasmic reticulum (ER) is the entry site into the exocytic pathway for secretory and most membrane proteins. Traversing the mammalian ER membrane normally requires an N-terminal signal sequence on the protein to be translocated, the signal-recognition particle (SRP), and a translocon at the ER membrane (reviewed in [1,2]).

Signal sequences lack a consensus primary structure. Function is imparted to these peptides by conservation of three mechanistically interacting structural motifs: (i) a positively charged N-terminus; (ii) a central hydrophobic core; (iii) a more polar C-terminal region (reviewed in [3,4]). Several studies have demonstrated conclusively that signal sequences from one species can function with the translocation machinery of another species. For example, β -lactamase (β -L) from *Escherichia coli* can be secreted by *Xenopus* oocytes [5], and prepro- α -mating factor (pp α MF) of the yeast *Saccharomyces cerevisiae* can be imported co-translationally by mammalian microsomes [6].

Trypanosoma and the *Leishmania* sp. are genera within the family Kinetoplastidae, one of the most ancient eukaryotic lineages. With over 25 million people world-wide inflicted by diseases caused by these protozoans, their success as parasitic organisms is undisputed. Antigenic variation, RNA editing, trans-splicing and kinetoplast DNA aside, another engaging feature of the molecular cell biology of these kinetoplastids is their ability to blanket their plasma membrane with glycosylphosphatidylinositol (GPI)-anchored proteins (reviewed in [7]). This latter feature requires the synthesis and translocation within a few hours of millions of GPI-anchored proteins.

Each bloodstream form *Trypanosoma brucei* is covered with about 10^7 molecules of a variant surface glycoprotein (VSG).

Similarly, each promastigote (insect stage) of *Leishmania major* contains about 10^6 molecules of gp63 (leishmanolysin) [7]. Surprisingly, no body of work exists on the identification of components involved in translocation of proteins across the ER of *T. brucei* or the *Leishmania* sp., presumably because investigators have not expected to encounter anything different from that already described in mammalian cells or yeasts.

Signal sequences of several major proteins of *T. brucei* do not, we have found to our surprise, direct import of proteins into mammalian microsomes. A similar result was obtained with gp63 from *Leishmania chagasi*. To the extent that they are rejected by the mammalian protein-import machinery, we contend that signal sequences from these kinetoplastids might be unique, and we infer that signal peptides are not as universally interchangeable as previously thought. These observations are discussed in the context of emerging recognition that signal sequences are not functionally equivalent even within the same species [8,9].

MATERIALS AND METHODS

Reagents and chemicals

All reagents unless otherwise stated were of the highest purity obtainable from Sigma.

Plasmids

Episomes used in this study included pVSG117 {VSG 117 cDNA [10] cloned in pBluescript II SK (Stratagene) downstream of the

Abbreviations used: β -L, β -lactamase; ER, endoplasmic reticulum; gp63, 63 kDa glycoprotein from *Leishmania*; pp α MF, prepro- α -mating factor from *Saccharomyces cerevisiae*; SRP, signal-recognition particle; VSG, variant surface glycoprotein from *T. brucei*; GPI, glycosylphosphatidylinositol; AcNPV, *Autographa californica* nuclear polyhedrosis virus; CPY, carboxypeptidase Y; bpPL, bovine preprolactin.

¹ These authors made equal contributions to this work and must both be considered first authors.

² To whom correspondence should be addressed.

phage T3 promoter}, pBiPHA9 (*T. brucei* BiP in pBluescript SK II downstream of the T7 promoter) [11] and pVSG221ΔGPI (VSG 221 cDNA [12] truncated at residue 458 to eliminate its C-terminal GPI signal sequence and cloned in pBluescript II downstream of the T7 promoter), kindly provided by Dr. James Bangs (University of Wisconsin, Madison, WI, U.S.A.). Other plasmids were pMSPS2 (*L. chagasi* gp63 cDNA [13] cloned downstream of a T3 promoter in pBluescript SK), generously provided by Dr. Mary Wilson (University of Iowa, Des Moines, IA, U.S.A.), pSL/3, obtained from Dr. John Donelson (University of Iowa) containing VSG MVAT7 [14] cloned downstream of the T3 promoter in pBluescript KS, and pDJ100 encoding *S. cerevisiae* ppαMF downstream of an SP6 promoter [15], obtained from Dr. Randy Schekman (University of California, Berkeley, CA, U.S.A.). Two plasmids containing *L. major* gp63 with the wild-type (pAcPP63.3) or a baculovirus signal peptide (pAcPP63.5) [16] were generously provided by Dr. Robert McMaster (University of British Columbia, Canada). The 1.8 kb gp63 fragments therein were excised with *Bam*HI and ligated to pBluescript II (SK). Recombinants with gp63 cloned downstream of the T7 promoter were selected for *in vitro* translation.

Construction of ppαMF₁₋₉₁-VSG₄₅₋₂₆₀

This plasmid is designed to encode a hybrid protein containing the prepropeptide (residues 1–91) of α-mating factor of *S. cerevisiae* fused by PCR-mediated mutagenesis [17,18] with residues 45–260 of *T. brucei* VSG 117. Primers used were: F1αMF, (5'-TAAGGATCCTTAACCTAGGAGGGTTTTTACCATGAGATTTCCTTCATTTTTACTGCA-3') containing the elements necessary for efficient protein expression in rabbit reticulocyte system (italicized) [19], the coding sequence for residues 1–7 of ppαMF (in bold lettering) and a *Bam*HI site (underlined); ppαMF-VSG fusion, the 'fusion' primer (5'-GGCCGCCAGTCCGAGTGGTTATGCCAAGCTTCAGCCTC-3') which contains the complementary nucleotides encoding residues 45–51 of VSG 117 in-frame with residues 85 to 91 of ppαMF. A third primer, rVSG260 (5'-TAGGCGGATCCCTTACGAAAGTTTCAAGCCGATG-3'), contains nucleotides complementary to the coding sequence for residues 253–260 of VSG 117, a termination codon (italicized) and a *Bam*HI site (underlined) for cloning purposes.

To generate the fusion construct (ppαMF₁₋₉₁-VSG₄₅₋₂₆₀), high-fidelity PCR [20] was performed in three steps, each in 100 μl (final volume) of 10 mM Tris/HCl, pH 8, containing 50 mM KCl, 2.5 mM MgCl₂, 0.05 % Tween 20, 0.05 % Nonidet P40, 0.5 μM (final concentration) of each primer, 30 ng of template DNA, and 5 units of *Taq* polymerase (Fisher Scientific). Each amplification cycle consisted of 1 min at 95 °C, 2 min at 56 °C, and 2 min at 74 °C. An 8 min extension at 74 °C was appended to the 25th cycle. Primers and templates used at the three stages of amplification were as follows. In stage 1, F1αMF (forward) and αMF-VSG were fused (reverse) with pDJ100 to generate a megaprimer that consists of the prepropeptide fused to the coding sequence of residues 45–51 of VSG at the 3' end. This stage 1 reaction was performed for 10 cycles. In stage 2, the product of stage 1 was purified by ultrafiltration with microcon-100 (Amicon) and used as a forward primer with rVSG260 as the reverse primer and pVSG117 as template. The product of this second reaction encodes the first 91 amino acids of the prepropeptide of α-mating factor fused with residues 45 to 260 of VSG 117. It was purified by ultrafiltration with microcon-100, and used as a template for a 25-cycle PCR in which F1αMF and rVSG260 were used as primers to generate more of the full-length product, ppαMF₁₋₉₁-VSG₄₅₋₂₆₀.

Product from the last reaction was purified by ultrafiltration, digested from *Bam*HI, extracted with phenol/chloroform/isoamyl alcohol (25:24:1 by vol.), and ethanol precipitated, following standard protocols [21]. Subsequently, it was ligated overnight at 14 °C to pBluescript II SK (Stratagene). The ligation product was transformed into *E. coli* DH5α, and clones with the appropriate inset orientation with regard to the T7 promoter on the vector selected by standard protocols [21].

Cell-free protein synthesis and analysis of import into dog pancreatic microsomes

Protein was synthesized from 1 μg of plasmid, purified by anion-exchange chromatography (Qiagen tip-20; Qiagen), in a coupled transcription–translation system using rabbit reticulocyte lysate (Promega TNT). A typical 30 μl translation reaction mixture contains, as directed by the manufacturer, 15 μl of rabbit reticulocyte lysate, 1 μl of 1 mM amino acids (minus methionine), 1 μl of T7 or T3 RNA polymerase, 1 μl of ribonuclease inhibitor (RNasin, 100 U/μl) and 4 μl of [³⁵S]methionine (1000 mCi/mmol; Amersham). The reaction volume was made up to 30 μl with deionized water. When indicated, 2 μl of canine pancreatic microsomes (Boehringer) was included, as directed by the manufacturer, for import of the translation product.

The translational product was divided into three 10 μl portions which were (a) not treated, (b) treated with proteinase K (Boehringer) (150 μg/ml final concentration) or (c) digested with proteinase K (150 μg/ml) in the presence of 0.5 % Triton X-100 (final concentration). The mixtures were incubated on ice for 30 min and the reaction terminated with PMSF (5 mM final concentration), from a 100 mM stock in 95 % ethanol. Immediately after addition of PMSF, SDS/polyacrylamide gel loading buffer was added and the sample was heated at 95 °C for 5 min. Proteins in 5 μl portions of the reactions were resolved by SDS/PAGE (15 % minigel) (Bio-Rad). The gel was soaked in Entensify[™] (DuPont) for 30 min, dried and exposed to Hyperfilm[™] (Amersham) for 2–12 h at –80 °C for fluorographic detection of radiolabelled proteins.

RESULTS

Three VSGs with unrelated primary structures cannot be imported into dog pancreatic microsomes

We used rabbit reticulocyte lysate and dog pancreatic microsomes to test whether the mammalian ER import machinery could translocate proteins from the Kinetoplastidae. First, a VSG 117 cDNA was transcribed/translated *in vitro* giving rise to the expected 55 kDa product (Figure 1A, lane 1). VSG 117 thus synthesized is completely degraded by exogenously added proteinase K (Figure 1A, lane 2).

To test for import of the translation product, ER microsomes from dog pancreas were included during protein synthesis and the mixture re-examined to determine whether the product was protected from exogenous protease [22]. VSG 117 translated in the presence of dog pancreatic microsomes remained digestible by proteinase K (Figure 1A, lane 4). Addition of detergent (0.5 % Triton X-100) to solubilize the microsomal membranes giving access to proteinase K of any imported VSG made no difference to our last observation (Figure 1A, lane 5). These data indicate that VSG 117 cannot be imported into dog pancreatic microsomes.

VSGs are a group of antigenically distinct proteins each of which can serve as a major surface protein of *T. brucei*. To test the possibility that our observations with VSG 117 were an isolated case, we studied two additional VSGs, MVAT7 and 221,

Figure 1 *In vitro* translation and import of VSG 117 and VSG MVAT7 into dog pancreatic microsomes

Proteins were translated in a rabbit reticulocyte lysate in the absence or presence of dog pancreatic microsomes. **(A)** VSG 117; **(B)** VSG MVAT7. Lane 1, untreated translation product; lane 2, translation product treated with proteinase K; lane 3, translation product with microsomes; lane 4, microsome translation product treated with proteinase K; and lane 5, microsome translation product treated with proteinase K and 0.5% Triton X-100.

neither of which has protein sequence similarity to VSG 117. MVAT7 was translated efficiently in the reticulocyte lysate (Figure 1B, lane 1) and the product was digested by proteinase K (Figure 1B, lane 2). On addition of microsomes, translation was not hindered (Figure 1B, lane 3), but MVAT7 was not imported into the microsomes; it remained susceptible to exogenous proteinase K (Figure 1B, lane 4). MVAT7 was digested by proteinase K in the presence of detergent (Figure 1B, lane 5), as expected. We conclude that VSG MVAT7, like VSG 117, cannot be imported into dog pancreatic microsomes. Identical results were obtained in similar studies with VSG 221 (results not presented).

Signal sequences function across species. For example, an *E. coli* or yeast protein can be imported into dog pancreatic microsomes [5,6]. The apparent incompatibility between VSG signal peptides and the canine microsomal import machinery as documented in Figure 1 prompted us to double-check the ability of the microsomes that we were using to import and process two test proteins, namely, *E. coli* β -L and yeast pp α MF. When translated *in vitro* in the absence of microsomes, β -L was susceptible to proteinase K (Figure 2A, lanes 1 and 2). After the addition of microsomes, most of the 30 kDa β -L was processed to a 28 kDa species, presumably through the action of a signal peptidase in the lumen of the microsomes (Figure 2A, lane 3). More importantly, the processed β -L was not digested by exogenously added proteinase K, whereas the unprocessed 30 kDa protein was digested (Figure 2A, lane 4).

Figure 2 Import of yeast pp α MF and *E. coli* β -L into dog pancreatic microsomes

Proteins were translated in rabbit reticulocyte lysate in the absence or presence of dog pancreatic microsomes. **(A)** β -L; **(B)** pp α MF. Lane 1, untreated translation product; lane 2, translation product treated with proteinase K; lane 3, translation product with microsomes; lane 4, translation product treated with proteinase K; and lane 5, microsomal translation product treated with proteinase K and 0.5% Triton X-100.

Yeast pp α MF behaved similarly to *E. coli* β -L. In the absence of microsomes, its 18 kDa translation product was susceptible to proteinase K (Figure 2B, lanes 1 and 2). With translation in the presence of microsomes (Figure 2B, lane 3), a processed (glycosylated) 30 kDa protein was obtained which was no longer digested by proteinase K (Figure 2B, lane 4); detergent was now required to permeabilize the membrane so that proteinase K could digest the processed (glycosylated) protein (Figure 2B, lane 5).

These control experiments confirmed that the mammalian microsomes were not deficient in any factor required for the import process, since they were capable of importing [and processing by signal sequence cleavage (β -L) or glycosylation (pp α MF)] proteins from *E. coli* and yeast, as previously demonstrated. Above that, the control experiments suggested that the inability of the mammalian microsomes to import *T. brucei* VSGs could be a peculiarity of that 'family' of proteins. Alternatively, these data hint at the possibility that signal sequences from the Kinetoplastidae might not be compatible with the mammalian ER translocation machinery.

Figure 3 *In vitro* translation and import of gp63 from *Leishmania* and BiP from *T. brucei* into dog pancreatic microsomes

Proteins were translated in a rabbit reticulocyte lysate in the absence or presence of dog pancreatic microsomes. (A) BiP; (B) gp63. Lane 1, untreated translation product; lane 2, translation product treated with proteinase K; lane 3, translation product with microsomes; lane 4, microsome translation product treated with proteinase K; lane 5, microsome translation product treated with proteinase K and 0.5% Triton X-100.

Import failure of *T. brucei* proteins is not restricted to VSGs

T. brucei BiP and gp63 from *L. chagasi* were studied to ascertain whether the failure to import VSG was limited to that group of major surface proteins from *T. brucei*. BiP is ER-resident, and gp63 is the major cell surface protein in the *Leishmania*.

In the absence of microsomes, an 81 kDa protein was produced when a *T. brucei* BiP cDNA was transcribed/translated in the rabbit reticulocyte lysate (Figure 3A, lane 1). The translation product was degraded by proteinase K (Figure 3A, lane 2). When microsomes were present during the translation, the same 81 kDa

Figure 4 Translation and import of ppzMF₁₋₉₁-VSG₄₅₋₂₆₀ into dog pancreatic microsomes

ppzMF₁₋₉₁-VSG₄₅₋₂₆₀ was translated in a rabbit reticulocyte lysate in the absence or presence of dog pancreatic microsomes. Lane 1, untreated translation product; lane 2, translation product treated with proteinase K; lane 3, translation product with microsomes; lane 4, microsome translation product treated with proteinase K; lane 5, microsome translation product treated with proteinase K and 0.5% Triton X-100.

protein was produced (Figure 3A, lane 3) which was degraded completely after addition of proteinase K in the absence (Figure 3A, lane 4) or presence (Figure 3A, lane 5) of detergent. Hence *T. brucei* BiP could not be imported either.

Similarly, an *L. chagasi* gp63 cDNA was transcribed/translated to a 63 kDa protein (Figure 3B, lane 1), which was susceptible to proteinase K digestion (Figure 3B, lane 2). On addition of microsomes, the 63 kDa protein was produced (Figure 3B, lane 3) which remained digestible by proteinase K (Figure 3B, lane 4), or proteinase K and Triton X-100 (0.5% final concentration) (Figure 3B, lane 5). We conclude from these observations that *L. chagasi* gp63, like the proteins from *T. brucei* cannot be imported by mammalian microsomes.

Protein translocation involves several steps and many components all of which are functional in the lysates and microsomes that we used, as demonstrated in the experiments with yeast ppzMF and *E. coli* β-L (Figure 2). Hence the import deficiency observed with the *T. brucei* and *L. chagasi* proteins is specific to these protozoans. In an attempt to correct this puzzling species-specific translocation deficiency, extra ATP and/or GTP (1 mM each) was added to supplement the quantities of these nucleotides already present in the commercial translation/translocation systems. Still no import of the kinetoplastid proteins was observed with these modifications. Thereafter, we entertained the possibility that the mammalian import machinery may fail to recognize and/or use signal sequences from these kinetoplastid protozoans.

The signal sequence from yeast ppzMF can direct translocation of VSG 117

We tested whether signal peptides were the reason why the kinetoplastid proteins could not be imported into dog pancreatic microsomes. For this purpose, the signal sequence of VSG 117 (presumably in residues 1–45 [10]) was replaced with the prepropeptide of ppzMF. Residues 1–91 of ppzMF were fused in-frame to residues 45–260 of VSG 117. VSG was truncated in order to facilitate the detection of any modifications (e.g. cleavage or glycosylation) that might occur if the chimaeric protein, i.e. ppzMF₁₋₉₁-VSG₄₅₋₂₆₀, was imported.

Table 1 Sequences and predicted properties of the signal peptides of proteins used in this study

Hydrophobicity peak value was obtained from a hydropathy analysis of the protein using DNA Strider ver 1.2. When the n-region was not very long this peak value was provided in the analysis by PSORT ver 6.3 (WWW) [23]. Predicted structure was calculated with the Profile-fed neural network systems PHDsec (EMBL, Heidelberg) [24]. We employed principles originally propounded by McGeogh [25], for which algorithms have been written and modified by Nakai and Kanehisa [23] to calculate the discriminant score using the formula $[0.305 \times (\text{length of h-region}) + 6.991 \times (\text{peak hydrophobicity}) + 0.668 \times (\text{net charge of n-region}) - 18.204]$ by inspection of the first 100 amino acids in the protein. Peak hydrophobicity was calculated with the Kyte–Doolittle hydropathy analysis program [26] available in DNA Strider ver 1.2. When the natural cleavage site of the protein was unknown it was predicted using the $-3, -1$ rule of von Heijne available in PSORT ver 6.0 WWW, in order to decide the exact length of the h-region. If the protein was predicted not to have a cleavable signal sequence, the h-region was assigned as the uncharged region (UR) originally defined by McGeogh [25]. AcNPV, *Autographa californica* nuclear polyhedrosis virus; Lm, *L. major*.

(A)

Protein	n-region	h-region
<i>T. brucei</i> VSG 117	MDCHTKETLGVTQWRR	STMLTSLSLYAITP
<i>T. brucei</i> VSG MVAT 7	MSTR	VQQATSCVLLIIGCTNYAAS
<i>T. brucei</i> BiP	MSR	MWLTAAVFLTVTVAAVSA
<i>L. chagasi</i> gp63	MSVDSSSTHRHRSVAARLVR	LAAAGAAVIAAVGTAAAWA
Yeast ppαMF	MR	FPSIFTAVLFAASSALA
AcNPV gp67/Lm gp63 hybrid	M	VSAIVLYLVAVGTAAAWA
<i>E. coli</i> β-L	MMTQSIRR	SMLTVMATLPLLFSSATL

(B)

Signal peptide	Net charge of n-region	h-region					Functionality	
		Length	Hydrophobicity peak value	Predicted structure (%)			Discriminant score	Translocated experimentally
				α-Helix	β-Strand	Rest		
Yeast ppαMF	1	17	2.6	58.3	0.0	41.7	5.83	Yes
<i>E. coli</i> β-L	2	18	2.1	72.2	0.0	27.8	3.30	Yes
<i>T. brucei</i> VSG 117	2	14	2.0	0.0	52.6	47.4	1.38	No
<i>T. brucei</i> VSG MVAT7	1	20	3.2	0.0	15.0	85.0	10.94	No
<i>T. brucei</i> BiP	1	19	2.5	68.4	0.0	31.6	5.74	No
<i>L. chagasi</i> gp63	5	19	2.2	89.5	0.0	10.5	6.31	No
AcNPV gp67/Lm gp63 hybrid	0	19	3.2	90.0	0.0	10.0	9.61	Yes

In the absence of microsomes, a 34 kDa pp α MF₁₋₉₁–VSG₄₅₋₂₆₀ was produced in the rabbit reticulocyte lysate (Figure 4, lane 1) which was digested by proteinase K (Figure 4, lane 2). When microsomes were added during translation, a 37 kDa protein was obtained (Figure 4, lane 3), in addition to the 34 kDa protein observed earlier (Figure 4, lane 1). On treatment of these microsomal translation products with proteinase K, the 37 kDa presumably glycosylated α MF₁₋₉₁–VSG₄₅₋₂₆₀ was protected completely, while the unprocessed 34 kDa protein was digested (Figure 4, lane 4). (pp α MF contains three potential N-glycosylation sites at positions 23, 57 and 67.) The processed 37 kDa protein was digested only when detergent was added during the proteinase K reaction (Figure 4, lane 5). These experiments indicate the pp α MF₁₋₉₁–VSG₄₅₋₂₆₀ can be imported into dog pancreatic microsomes and processed.

Since the pp α MF signal sequence directs VSG 117 into the lumen of mammalian microsomes, we conclude that the signal peptide of VSG 117 is incompatible with the translocation machinery of canine pancreatic microsomes. This conclusion was potentially clouded by two properties of the VSG–pp α MF chimera. First, pp α MF signal peptide is 19 amino acids long (Table 1A) yet the fusion construct contains 90 amino acids from the yeast protein. Conceptually, this issue is not problematic because it is well established that sequences outside of the first 19 amino acids do not function as the signal peptide for pp α MF. Second, although the VSG 117 signal peptide is 35 amino acids long (Table 1A), we deleted 45 N-terminal amino acids from the

protein. Could the amino acids between residues 35 and 45 have been inhibitory to translocation (i.e. do kinetoplastid proteins contain information outside of their signal peptides that inhibit their import into canine microsomes)?

gp63 signal peptide does not contain a translocation inhibitor C-terminal to its h-region

We tested the possibility that kinetoplastid proteins contain information C-terminal to their signal peptides which interfere with their import into canine microsomes. Owing to the lack of sequence similarity between the proteins studied (Tables 1 and 2), such an inhibitor is likely to be conformational and can be predicted to be present in all kinetoplastid proteins. Consequently, we chose to examine gp63 in this instance. For this purpose, a chimaeric signal peptide (MVSIVLYLVAVGTAAAWA) consisting of the h-region (MVSIVLYLV) of *Autographa californica* nuclear polyhedrosis virus (AcNPV) gp67 (Table 1A) fused to the C-terminal nine residues of the gp63 h-region (AVGTAAAWA) (Table 1A) was studied. In this construct, the n-region of wild-type *L. major* gp63, which could not promote translocation (Figure 5A), is deleted but all information C-terminal to the gp63 h-region is retained. This hybrid signal peptide promoted translocation into canine microsomes (Figure 5B); a translocated and glycosylated gp63 was detectable (Figure 5B, lanes 3 and 4). A control construct in which only the n-region

Table 2 Sequences (A) and predicted properties (B) of some signal peptides from *Bos*, *Saccharomyces*, *Leishmania* and *Streptomyces*

See Table 1 for details of properties.

See Table 1 for details of properties.

(A)

Signal peptide	n-region	h-region
Bovine preprolactin	MDSKGSSQKGSR	LLLLLVVSNLLLC
Yeast carboxypeptidase Y	MK	AFTSLLCGLGLSTTLA
<i>L. donovani</i> gp63	MSVDSSSTHRHRSVAARLVR	LAAAGAAVIAAVGTAAAWA
<i>L. mexicana</i> gp63	MPVDSSSTHRHRCVAAPLVR	LAAAGAAVTAVGTAAAWA
<i>L. major</i> gp63	MSVDSSSTHRRRCVAARLVR	LAAAGAAVTAVGTAAAWA
<i>L. chagasi</i> gp63	MSVDSSSTHRHRSVAARLVR	LAAAGAAVIAAVGTAAAWA
<i>L. guyanensis</i> gp63	MSRDRSSTHRRRSVAAR	LIGFAAAGLVMAVGAAAVWA
<i>C. fasciculata</i> gp63	MHAPPTATRRSGPRRTHGIMARLVR	LAAGVLVTVLIGALTALS
<i>L. major</i> LPG 1	MAPPRWHDRRRMAIFVR	VGLYTLFLMGYVPLIIFYN
<i>L. major</i> gp46/M2	MAQCVR	LVLAAPLAAVVALLCTSSAPVARA
<i>S. cacaoi</i> β -L	MRIRPTRR	LLLGAVAPLALVPLVACGQASG
<i>Streptomyces</i> sp. cellulase	MENPRTTPTPLRRRSERRARGGR	VLTALTGVTLTAGLAIAPA

(B)

Signal peptide	Net charge of n-region	h-region		Summary (discriminant score)
		Length	Peak value	
Bovine preprolactin (bpPL)	2	13	2.5	4.57
Δ 13–17 bpPL	2	8	2.0	−0.45
Yeast carboxypeptidase Y, wild-type	1	16	1.8	−0.07
G12L yeast carboxypeptidase Y	1	16	2.5	4.82
G10L yeast carboxypeptidase Y	1	16	2.5	4.82
G10, 12L yeast carboxypeptidase Y	1	16	3.1	9.02
<i>L. donovani</i> gp63	5	19	2.2	6.31
<i>L. chagasi</i> gp63	5	19	2.2	6.31
<i>L. mexicana</i> gp63	4	19	2.1	4.94
<i>L. guyanensis</i> gp63	6	20	2.2	7.28
<i>C. fasciculata</i> gp63	7	19	3.1	13.94
<i>L. donovani</i> LPG1	6	19	2.2	6.98
<i>L. major</i> PSA-2 (gp46/M2)	2	16	3.0	8.99
<i>S. cacaoi</i> β -L	4	22	2.7	10.05
<i>Streptomyces</i> sp. cellulase	7	19	2.8	11.84

of the gp63 signal peptide was deleted was not translocated (results not presented).

Success of the above experiment has two implications. First, it rules out the possibility that proteins present in the secretory pathway of kinetoplastids contain information C-terminal to their signal sequences that inhibit translocation into canine microsomes. Second, the data implicate the h-region as the subdomain that imparts incompatibility with the mammalian translocon to kinetoplastid signal peptides. We think it unlikely that the deleted n-region played a role in inhibiting translocation of gp63, because several of the *T. brucei* signal peptides lack extended n-regions (Table 1A) yet they, like gp63, are not imported.

Discriminant scores can predict the importability of most signal sequences into canine pancreatic microsomes

Signal sequences have three subdomains [27], comprising (i) an N-terminal positively charged n-region, (ii) a central uncharged h-region and (iii) a polar c-region (reviewed in [28]). On the basis of the properties of these subdomains, a semiquantitative predictor of signal sequence function can be calculated as a 'discriminant score', first described by Nakai and Kanehisa [23]

(see Legend to Table 1). The more positive a discriminant score, the greater the likelihood that the signal peptide promotes translocation of proteins into mammalian ER. Four well-studied proteins were selected to test whether experimental support could be obtained for discriminant-score-based predictions (Tables 1B and 2B).

Yeast carboxypeptidase Y (CPY) cannot be imported into mammalian microsomes, but three site-directed mutants in the h-region (G12L, G10L and G10,12L) (Table 2B) were all successfully translocated [29]. The discriminant score of wild-type CPY was −0.07, while the G12L, G10L and G10,12L mutants had scores of 4.82, 4.82 and 9.02 respectively (Table 2B). From these tests it appears that a signal peptide with a discriminant score of 4.82 or greater is likely to direct import of proteins.

Analysis of signal peptides from three species, namely bovine preprolactin (bpPL), *E. coli* β -L and yeast ppzMF, supported our conclusions. The discriminant score of bpPL which is translocated into canine microsomes was 4.57. An import-defective mutant Δ 13–17bpPL [30] had a score of −0.45 (Table 2B). Finally, β -L from *E. coli* and yeast ppzMF had discriminant scores of 3.30 and 5.83 respectively (Table 1B), and were both imported. From these observations, we surmise that a signal sequence with a discriminant score of 3.30 or greater ought to be imported into canine microsomes.

Figure 5 A hybrid of h-regions from AcNPV gp67 and *L. major* gp63 is functional

(A) gp63 from *L. major* cannot be translocated in canine pancreatic microsomes. Native gp63 from *L. major* was studied as detailed in the legend to Figure 3. (B) A signal peptide containing h-regions from AcNPV gp67 and *L. major* gp63 promotes translocation of gp63 into canine pancreatic microsomes. The signal peptide on native gp63 was replaced by one containing sequences from AcNPV gp67 and *L. major* gp63 (Table 1). gp63 processed in the presence of microsomes (lane 3) was protected from proteinase K digestion (lane 4).

T. brucei and *Leishmania* signal peptides defy predictions based on discriminant scores

Could the rules generated from analysis of mammalian and yeast signal peptides (above) be applied to kinetoplastid signal sequences? *T. brucei* BiP had a discriminant score of +5.74 (Table 1) and was therefore predicted to have a functional signal peptide. That expectation was not supported experimentally (Figure 3). β -L from *E. coli*, on the other hand, with a discriminant score of 3.30 was imported. The ability of canine microsomes to import *E. coli* β -L without translocating BiP is intriguing, because it defies the predicted behaviour of the signal peptides which is based solely on their biochemical/biophysical properties (Table 1B). Similarly, although the discriminant score of yeast pp α MF (+5.83) was very similar to that of *T. brucei* BiP, the yeast protein was imported while the kinetoplastid protein was not.

Was the BiP signal sequence non-functional because the h-region adopted an irregular conformation? This possibility can be ruled out from examination of the secondary structures of the BiP h-region, predicted with PHDsec (EMBL Heidelberg) [24] (Table 1B). The percentage α -helical content of β -L, BiP and pp α MF were 72.2, 68.4 and 58.3 respectively. There is no correlation between the percentage of α -helix or β -sheets in the h-region and ability to be translocated, as previously documented (reviewed in [28]).

These observations about BiP hold true, in general, for VSGs 221 and MVAT7. With discriminant scores of 10.94 (Table 1B) and 11.32 respectively both were predicted to be excellent mammalian signal peptides, but neither was imported. Clearly, the N-terminus of a protein may have all the expected properties of a functional vertebrate signal peptide and still fail to be imported (Tables 1 and 2). Curiously, VSG 117 (discriminant

score = +1.38) might be intrinsically non-importable, since it scored below our empirically determined import threshold of 3.30.

Subregions of several signal peptides from *Leishmania* and *Crithidia* resemble signal sequences from Gram-positive bacteria

Species-specific differences in the organization of the subdomains of signal peptides was first noted by von Heijne [31]. Eukaryote signal sequences are the shortest. The n-region is typically four to five residues long, has a net charge of about +0.8 on average, and is not very hydrophilic. The h-region is 6–16 residues long and leucine-rich (reviewed in [4,31]). At the other extreme are the Gram-positive bacteria, e.g. *Streptococcus* and *Streptomyces* sp. A long n-region (12–15 residues) with a net charge between +3.5 and +4.3 is typical in these prokaryotes.

Four properties of the *L. chagasi* gp63 signal peptide (Table 1A) highlight its similarity to signal peptides from Gram-positive bacteria. The n-region is, first, unusually long (20 residues). Second, it has a net charge of +5. Third, the n-region is arginine-rich (Tables 2A and 2B). Finally, the h-region is long; composed of 19 amino acids. These features also make the gp63 signal peptide very similar to those from archae (see [4,31] for reviews).

gp63 is found in all *Leishmania* subspecies (reviewed in [32]). Outside of that genus, the protein is expressed in *Crithidia fasciculata*, another kinetoplastid [33]. We investigated whether the subdomain organization of the *L. chagasi* gp63 signal sequence was conserved in the Kinetoplastidae. All the gp63s from human parasites had long h-regions of 19–20 amino acids. The peak of the h-region hydrophobicity (Kyte and Doolittle scale [26]) varied between 2.1 for gp63 from *L. major* and *L. mexicana* and 3.1 for *Crithidia*. The net charge of the n-region ranged from +4 (for *L. mexicana*) to +7 for *C. fasciculata* (Table 2B). Two other proteins of *Leishmania* were included in this analysis, LPG1p [34] and gp46/m2 [35]. The average net charge of the n-region in these signal peptides is +5, highly unusual for a eukaryote. In addition, the positive charge is contributed mainly by Arg, as opposed to Lys. Thus gp63 signal sequences in the Kinetoplastidae are reminiscent of signal peptides from Gram-positive prokaryotes (Table 2A).

DISCUSSION

Signal peptides can be species-specific

Signal peptides play critical roles in directing most secretory and membrane proteins across the ER membrane. After their emergence from ribosomes, signal peptides are usually recognized by the SRP, which eventually targets ribosome–nascent chain complexes to the ER membrane (reviewed in [1,36]). At the translocon, consisting mainly of the Sec61p complex at the ribosome–ER membrane junction, the signal sequence is recognized again, this time by Sec61p, as the nascent polypeptide is ‘threaded’ through the gated aqueous pore. Dual recognition of the signal sequence (reviewed in [1,36]) appears to impart specificity to the import process, since ER signal peptides lack a consensus sequence.

Three structural motifs are found in these signal sequences: (i) an N-terminal positively charged n-region; (ii) a central uncharged h-region; (iii) a more polar, but uncharged, c-region where signal peptidase cleavage occurs [4,25,31]. These subdomains are functionally interactive (reviewed in [4]), in that a signal sequence defective by virtue of lacking an n-region, for example, may regain competence if the length and/or hydrophobicity of the h-region is increased.

All cells examined so far use these signal peptide motifs: specific sequences of amino acids are not needed. Consequently, it has been possible to interchange signal peptide function across biological kingdoms and species. For example, *E. coli* β -L can be secreted by *Xenopus* oocytes, and yeast ppzMF can be imported by the mammalian translocon in an SRP-dependent fashion. The logical conclusion from these dramatic observations is that signal peptides are interchangeable between all species.

Surprisingly, proteins from *T. brucei* (e.g. VSG 117, VSG MVAT7, VSG 221 and BiP) and *L. chagasi* (e.g. gp63) could not be imported into dog pancreatic (ER) microsomes (Figures 1 and 3). Under identical conditions, however, *E. coli* β -L and yeast ppzMF were both imported (Figure 2). Our results appear to contradict those of earlier workers who demonstrated that VSGs can be processed by canine microsomes [37]. The reader is reminded, however, that studies on translocation and processing employ very different assays and the requirements for success are different. Processing (e.g. N-glycosylation or signal peptide cleavage) unlike translocation does not require intact and appropriately oriented microsomal vesicles. In the earlier study, the investigators did not make any claims about translocation and therefore had no need to verify the integrity, by performing treatments with exogenous protease, of their microsome preparations. And they did not. Hence, the reported processing might not have required that the proteins be translocated into the lumen of the canine vesicles. Inside-out vesicles will process exposed signal peptides. As we have demonstrated in this work, translation of the trypanosomal proteins proceeds without detectable translational arrest when canine ER membranes are present: consequently, the presence of inside-out (or leaky) vesicles could easily generate processed VSGs. Reconciling the two sets of observations (i.e. theirs and ours), we surmise that canine enzymes can process *T. brucei* signal peptides if the enzymes have access to the signal sequences. This conclusion is compatible with our observation that *T. brucei* and *Leishmania* signal peptides are not translocated across the mammalian ER (see below).

When the signal sequence of VSG 117 was replaced with the signal sequence from ppzMF, the chimaeric protein was imported (Figure 4). In studies with results consistent with our findings, the signal sequence of *L. major* gp63 could not direct its secretion in cultured cells of the insect *Spodoptera frugiperda* (Sf9). Replacement of the gp63 signal sequence with that of gp67 from AcNPV resulted in secretion of the chimaeric protein [16].

These surprising observations on signal sequences from the kinetoplastid protozoan parasites suggest that substantial differences can exist in signal sequence recognition in different species. Signal sequences are recognized twice during translocation. Recognition at the ER appears to be more stringent, because a signal sequence defective in the h-region was delivered by SRP to the ER [30,38]. However, the polypeptide was not translocated. By analogy with those observations, it is possible that the kinetoplastid signal peptides either (a) fail to engage mammalian SRP, and/or (b) cannot interact productively with the mammalian translocon, most likely the Sec61p complex. Further work is clearly needed to determine exactly where the incompatibility between these kinetoplastid signal sequences and the mammalian translocation machinery resides.

Import into dog pancreas microsomes of proteins from three species, namely *S. cerevisiae*, *E. coli* and *Bos* (bovine), was accurately predicted from discriminant score analysis (Tables 1 and 2). Surprisingly, kinetoplastid proteins VSG MVAT7, VSG 221, BiP and gp63, scoring up to 300% higher in discriminant analysis than our test *E. coli*, yeast and bovine proteins, could not be imported (Tables 1 and 2; Figures 1 and 3). Apparently,

when certain species barriers are crossed, the biophysical properties of signal peptides *per se* are not sufficient to predict functionality. How could this enigma be resolved?

We propose two hypotheses which alone or together may resolve the conundrum above. First, specificity factors may be present in the mammalian translocation machinery which may reject signal peptides from 'incompatible' species (e.g. *T. brucei* or *Leishmania*). Alternatively, the kinetoplastid signal peptides may require *T. brucei*- or *L. chagasi*-encoded factors for their translocation; these essential factors might be missing from the mammalian microsomal system studied in this work.

The inability of the yeast CPY signal peptide to function with the mammalian translocation system was the first hint that signal sequences might not be totally interchangeable between eukaryotes [29]. The kinetoplastid signal peptides presented in this work firmly establish the concept that ER signal sequences are not completely interchangeable between species as previously believed.

Mutual exclusivity of the kinetoplastid signal peptides with the mammalian protein translocation machinery is consistent with recognition that, even within the same species, signal peptides are functionally diverse [8,9,29]. Mammalian signal sequences may be classified broadly into those that are dependent on translocation-chain-associated membrane protein and those that are not [9]. In *S. cerevisiae*, three classes of signal peptides have been recognized [8]: (i) SRP-dependent; (ii) SRP-independent; (iii) signal sequences that use both the SRP-dependent and SRP-independent pathways. Targeting to any of these three pathways is dictated by the h-region. Hence, in nature, the h-region bears more information than mere hydrophobicity, a conclusion that is in agreement with a report that signal sequences of identical hydrophobicity but with proline substituted at different positions within the h-region are not functionally equivalent [39].

This work was supported by grant no. AI 33383 from the National Institutes of Health, by grants from The University of Georgia Research Foundation, and by a Burroughs Wellcome Fund New Investigator Award in Molecular Parasitology to K.M.-W. We thank John Donelson, Mary Wilson and Robert McMaster for the gifts of various plasmids.

REFERENCES

- Rapoport, T. A., Jungnickel, B. and Kutay, U. (1996) *Annu. Rev. Biochem.* **65**, 271–303
- Walter, P. and Johnson, A. E. (1994) *Annu. Rev. Cell Biol.* **10**, 87–119
- Von Heijne, G. (1985) *J. Mol. Biol.* **184**, 99–105
- Rusch, S. L. and Kendall, D. A. (1995) *Mol. Membr. Biol.* **12**, 295–307
- Wiedmann, M., Huth, A. and Rapoport, T. A. (1984) *Nature (London)* **309**, 637–639
- Garcia, P. D. and Walter, P. (1988) *J. Cell Biol.* **106**, 1043–1048
- Ferguson, M. A. J. (1994) *Parasitol. Today* **10**, 48–52
- Ng, D. T. W., Brown, J. D. and Walter, P. (1996) *J. Cell Biol.* **134**, 269–278
- Voigt, S., Jungnickel, B., Hartmann, E. and Rapoport, T. A. (1996) *J. Cell Biol.* **134**, 25–35
- Allen, G., Gurnett, L. P. and Cross, G. A. M. (1982) *J. Mol. Biol.* **157**, 527–546
- Bangs, J. D., Brouch, E. M., Ransom, D. M. and Roggy, J. L. (1996) *J. Biol. Chem.* **271**, 18387–18393
- Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D. and Turner, M. (1991) *J. Mol. Biol.* **221**, 823–835
- Ramamoorthy, R., Donelson, J. E., Paetz, K. E., Maybodi, M., Roberts, S. C. and Wilson, M. E. (1992) *J. Biol. Chem.* **267**, 1888–1895
- Lenardo, M. J., Rice-Ficht, A. C., Kelly, G., Esser, K. M. and Donelson, J. E. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6642–6646
- Julius, D., Schekman, R. and Thorner, J. (1984) *Cell* **36**, 309–318
- Button, L. L., Wilson, G., Astell, C. R. and McMaster, W. R. (1993) *Gene* **134**, 75–81
- Vallete, F., Mege, E., Reiss, A. and Adesnik, M. (1989) *Nucleic Acids Res.* **17**, 723–733
- Ray, F. A. and Nickoloff, J. A. (1992) *BioTechniques* **13**, 342–346
- Al-Qahtani, A. and Mensa-Wilmot, K. (1996) *Nucleic Acids Res.* **24**, 1173–1174
- Innis, M. A., Myambo, K. B., Gelfand, D. H. and Brow, M. D. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9436–9440

-
- 21 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1988) *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, New York
- 22 Walter, P. and Blobel, G. (1983) *Methods Enzymol.* **96**, 84–93
- 23 Nakai, K. and Kanehisa, M. (1992) *Genomics* **14**, 897–911
- 24 Rost, B. and Sander, C. (1994) *Proteins* **19**, 55–72
- 25 McGeoch, D. J. (1985) *Virus Res.* **3**, 271–286
- 26 Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- 27 Haeuptle, M. T., Flint, N., Gough, N. M. and Dobberstein, B. (1989) *J. Cell Biol.* **108**, 1227–1236
- 28 Izard, J. W. and Kendall, D. A. (1994) *Mol. Microbiol.* **13**, 765–773
- 29 Bird, P., Gething, M. J. and Sambrook, J. (1987) *J. Cell Biol.* **105**, 2905–2914
- 30 Jungnickel, B. and Rapoport, T. A. (1995) *Cell* **82**, 261–270
- 31 Von Heijne, G. and Abrahmsen, L. (1989) *FEBS Lett.* **244**, 439–446
- 32 Bouvier, J., Schneider, P. and Etges, R. (1995) *Methods Enzymol.* **248**, 614–633
- 33 Inverso, J. A., Medina-Acosta, E., O'Connor, J., Russell, D. G. and Cross, G. A. M. (1993) *Mol. Biochem. Parasitol.* **57**, 46–54
- 34 Ryan, K. A., Garraway, L. A., Descoteaux, A., Turco, S. J. and Beverley, S. M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8609–8613
- 35 McMahon-Pratt, D., Traub Cseko, Y., Lohman, K. L., Rogers, D. D. and Beverley, S. M. (1992) *Mol. Biochem. Parasitol.* **50**, 151–160
- 36 Siegel, V. (1995) *Cell* **82**, 167–170
- 37 McConnell, J., Cordingley, J. S. and Turner, M. J. (1982) *Mol. Biochem. Parasitol.* **6**, 161–174
- 38 Hauser, S., Bacher, G., Dobberstein, B. and Lutcke, H. (1995) *EMBO J.* **14**, 5485–5493
- 39 Ryan, P. and Edwards, C. O. (1995) *J. Biol. Chem.* **270**, 27876–27879
-