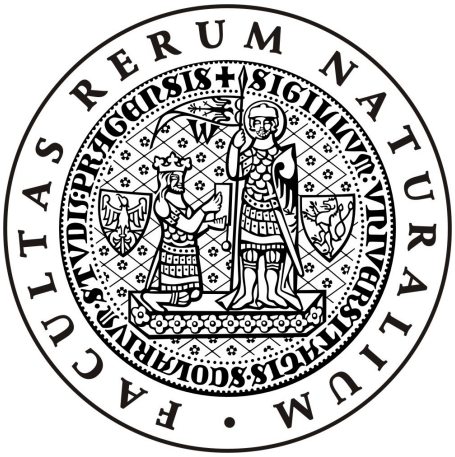
UNIVERZITA KARLOVA V PRAZE

Přírodovědecká fakulta

Katedra parazitologie

Studijní program Biologie

Studijní obor Parazitologie



Bc. Vojtěch Žárský

Protein import into mitochondria and peroxisomes of parasitic protists

Import proteinů do mitochondrií a peroxisomů parazitických prvoků

DIPLOMOVÁ PRÁCE

Školitel: Prof. RNDr. Jan Tachezy, PhD.

Praha **2012**

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V **Praze dne 23. 8. 2012**

Vojtěch Žárský

Děkuji mé rodině a přátelům za podporu během studia. Děkuji mému školiteli Janu Tachezymu za vedení během posledních let. Děkuji kolegům za milé a podnětné prostředí. Zvláště děkuji Mirce Šedinové, Evě Nývltové a Pavlu Doležalovi za spolupráci na projektech, které mohu prezentovat v této práci.

Table of Contents

1 Introduction

The presented thesis includes three related projects, that are linked by a common interest in the evolution of eukaryotic organelles and machineries that import proteins into these compartments. The first project considers the possibility of peroxisomes (eukaryotic organelles known in aerobic organisms) being conserved in two related anaerobic protists: a free-living amoeba *Mastigamoeba balamuthi* and a parasite *Entamoeba histolytica*. The most important hint for the presence of peroxisomes was the discovery of proteins that are homologous to known components of the peroxisomal protein import machinery. The second project aims to characterize the unknown protein translocase of the inner membrane (TIM) in the mitosomes (extremely reduced mitochondria) of an anaerobic protozoan *Giardia intestinalis*. We have discovered an important subunit of the mitosomal translocase (Tim44), which usually tethers the Hsp70/PAM (presequence translocase-associated motor) complex to the TIM translocon. The last project shows that the protein translocase of the outer mitochondrial membrane in trypanosomatids is related to a typical eukaryotic channel Tom40. This finding is important because the absence of Tom40 was previously considered an ancestral feature of trypanosomatids.

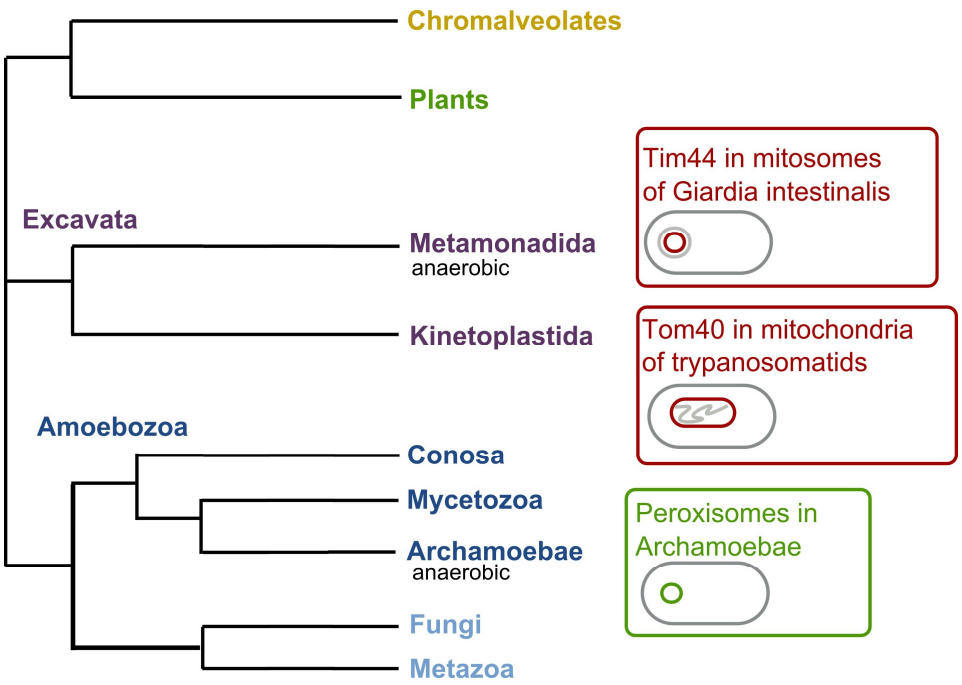


Figure 1. An overview of the topics covered by the thesis.

2 Review of literature

First I will introduce mitochondria and peroxisomes with special emphasis on the machineries and mechanisms of the protein import. Then I will shortly introduce organisms that are the subject of the thesis: parasitic protists *Entamoeba histolytica*, *Giardia intestinalis*, *Trypanosomatidae* and a free-living relative of *E. histolytica*, *Mastigamoeba balamuthi*.

2.1 Mitochondria

The key event in the evolution of eukaryotes was the acquisition of an α-proteobacterial endosymbiont which later became a fully integrated cellular organelle - the mitochondrion (Sagan, 1967). This transition included massive horizontal gene transfer of the endosymbiont genes to the host genome that was accompanied by evolution of a molecular machinery to import the nuclear-encoded mitochondrial proteins to the mitochondrion (Dolezal et al., 2006).

Nowadays mitochondria are surrounded by the inner and the outer membrane, which are probably homologous to the original bacterial membranes of the endosymbiont. The inner membrane folds to form the typical mitochondrial cristae. Mitochondria are dynamic organelles that actively divide and fuse with each other (Figure 2).

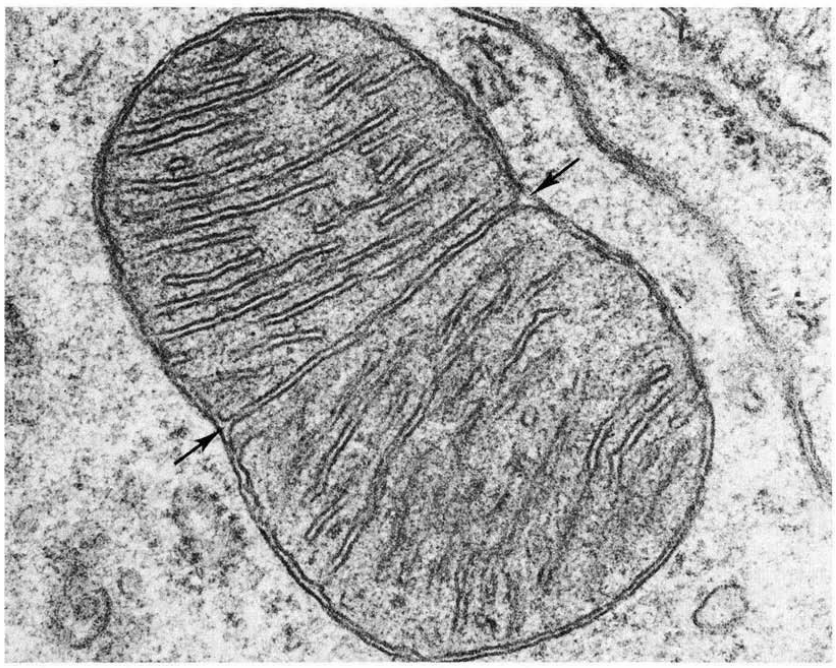


Figure 2. A dividing mitochondrion in gastric mucosa of a mole. The arrows indicate the division site. (Fawcett, 1981)

Mitochondria are the main compartment of eukaryotic oxidative metabolism. The Krebs cycle and fatty acid oxidation take place in mitochondrial matrix. The electron transport chain of mitochondria uses the sequential transfer of electrons from a donor (NADH, succinate) to an acceptor (O_2) to produce ATP. Other essential functions of mitochondria include the synthesis of heme and Fe-S clusters.

Mitochondria usually contain a small bacterial-like genome, which is a relic of the endosymbiotic ancestor, although most of the proteins of modern-day mitochondria are coded by the nuclear genome and are posttranslationally transported to the mitochondrial compartments by a specialized modular protein import machinery (Figure 3). Many proteins designated for mitochondrial import carry an N-terminal mitochondrial targeting signal whereas some carry other sequences that serve as an internal import signal.

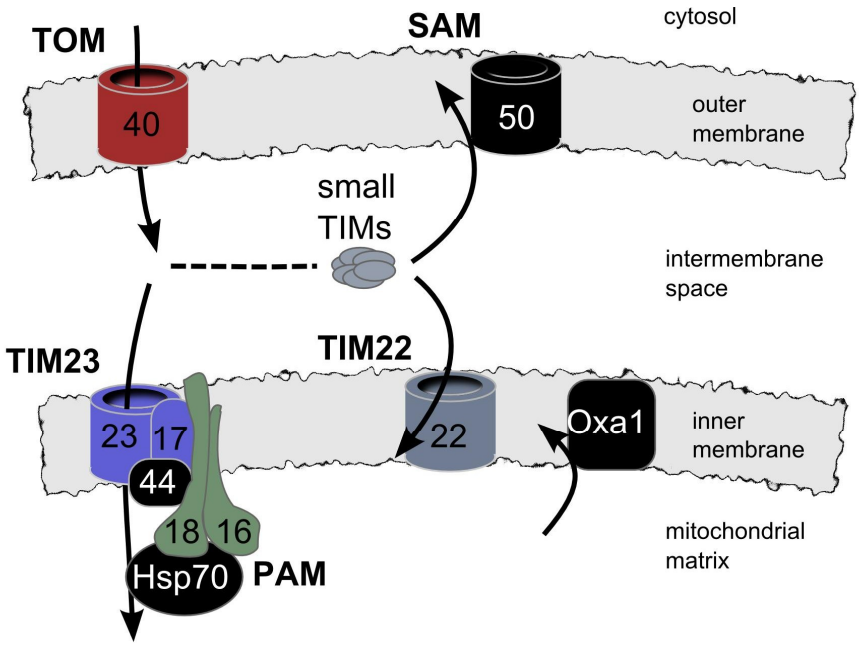


Figure 3. The mitochondrial protein import machinery consists of several complexes: TOM (translocase of the outer mitochondrial membrane) is the main translocation channel of the outer mitochondrial membrane. SAM (sorting and assembly machinery) is required for the insertion of beta barrel proteins into the outer mitochondrial membrane. TIM23 (translocase of the inner mitochondrial membrane; Tim23, Tim17 and Tim44) transports soluble proteins into the mitochondrial matrix. PAM (presequence translocase-associated motor; Pam16, Pam18 and Hsp70) is tethered to the TIM23 complex and contains the conserved mitochondrial Hsp70 ATPase. TIM22 complex facilitates the insertion of proteins into the inner mitochondrial membrane. The Oxa1 protein inserts hydrophobic mitochondria-encoded proteins into the inner mitochondrial membrane. Proteins with close homologs among bacteria are shown in black.

The outer membrane of mitochondria contains two essential complexes: TOM (translocase of the outer mitochondrial membrane) and SAM (sorting and assembly machinery). TOM complex is the main protein translocation channel of the outer mitochondrial membrane. The core translocon of TOM complex is formed by a beta-barrel protein Tom40 (Hill et al., 1998) which is homologous to VDACs (voltage-dependent anion channel) mitochondrial metabolite transporters (Zeth and Thein, 2010). The core subunit of SAM complex is Sam50 which is related to bacterial Omp85 family proteins and is important for insertion of outer membrane beta-barrel proteins (Kozjak et al., 2003).

The small TIM proteins of the mitochondrial intermembrane space (IMS) are small soluble proteins with four conserved cysteine residues. They form heterohexamers (Tim9-Tim10 and Tim8-Tim13) and facilitate import of hydrophobic proteins across the IMS (Koehler et al., 1998). Components of the disulphide relay system of the IMS (Mia40 and Erv1) drive the import of cysteine-rich proteins into the IMS (Mesecke et al., 2005).

Proteins designated for import into the mitochondrial matrix are imported through the TIM23 complex. The translocation channel of the TIM23 complex is comprised by Tim23 and Tim17 proteins. The import is driven by the peripheral PAM (presequence translocase-associated motor) complex, most notably by the Hsp70 ATP-ase. Other subunits of the PAM complex, Pam16 and Pam18 together with the TIM23 complex subunit Tim44 tether the PAM complex to the inner membrane and the TIM23 translocon (Berthold et al., 1995). The TIM22 complex inserts hydrophobic proteins into the inner mitochondrial membrane. The core subunits of the two TIM complexes; Tim17, Tim22 and Tim23 belong to the same protein family with no known homologs among prokaryotic organisms. The inner membrane protein Oxa1 is related to bacterial YidC protein family and serves as an insertase for membrane proteins that are coded by the mitochondrial genome (Bonnefoy et al., 1994).

Mitochondria of organisms adapted to anaerobic environments tend to lose typical oxygen-related functions, which is accompanied by general reduction and/or acquisition of new functions. These organelles are called Mitochondrion-related organelles (MROs). MROs emerged several times in different eukaryotic groups (Figure 4) (van der Giezen and Tovar, 2005), though common patterns in the evolution of the functions and proteome can be observed. MROs usually lack the organellar genome and respiratory complexes, although notable exceptions are known (van Hoek et al., 2000;Stechmann et al., 2008). Organisms with MROs have a high frequency of lateral gene transfers of metabolism-related genes from anaerobic bacteria.

Usually there are distinguished two types of MROs: Mitosomes and Hydrogenosomes. Mitosomes are

MROs that don't produce ATP. MROs that produce molecular hydrogen and ATP are called hydrogenosomes. The best studied hydrogenosomes are of *Trichomonas vaginalis*, a parasite of the human urogenital system. In the hydrogenosomes of *T. vaginalis* the pyruvate is broken down to CO_2 and acetyl-CoA by the enzyme pyruvate:ferredoxin oxidoreductase. The electrons generated are further passed to the electron carrier ferredoxin and finally to the enzyme hydrogenase which uses the electrons to reduce H^+ to H_2. The enzymes acetate:succinate CoA-transferase and succinate thiokinase catalyze the metabolism of acetyl-CoA into acetate and ATP (Lindmark et al., 1975).

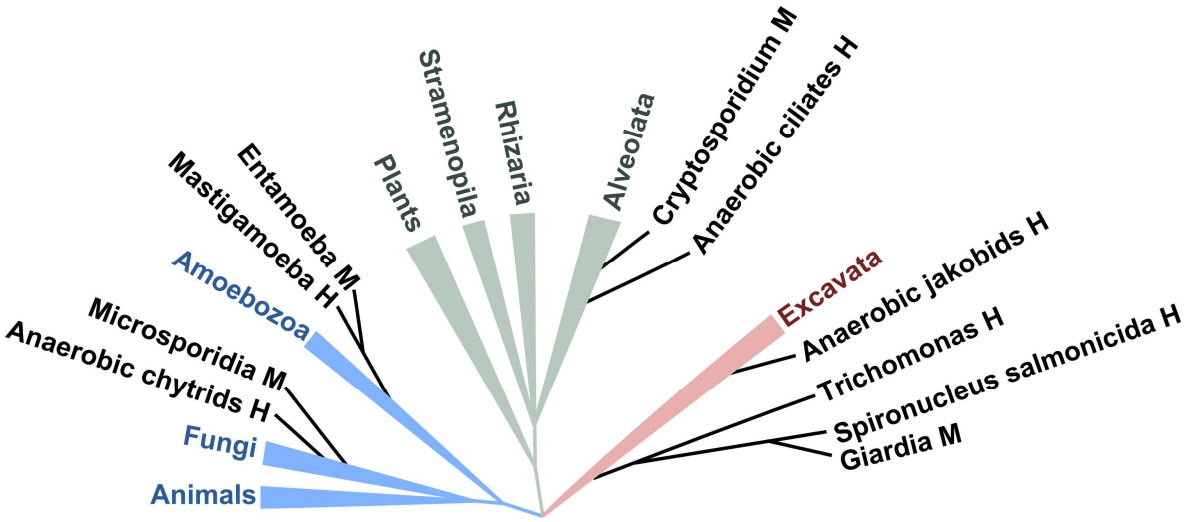


Figure 4. An overview of eukaryotic groups with MRO. M – mitosomes, H - hydrogenosomes

One of the unifying characteristics of mitochondria and MROs is the common mode of protein targeting, although in MROs and especially in mitosomes, reduced complexity of the organelles is reflected by a marked reduction of the protein import machinery (Table 1) (Dolezal et al., 2005;Lithgow and Schneider, 2010).

	Tom40	Sam50	Tim17-like	Tim44	Pam16	Pam18	Hsp70	Oxa1
Saccharomyces cerevisiae	+	+	+	+	+	+	+	+
Trypanosoma brucei	+	+	+	+		+	+	+
Trichomonas vaginalis - H	+	+	+	+	+	+	+	
Giardia intestinalis - M	+			+	+	+	+	
Entamoeba histolytica - M	+	+					+	
Cryptosporidium parvum - M	+	+	+	+		+	+	
Encephalitozoon cuniculi - M	+	+	+			+	+	

Table 1. Distribution of conserved components of the mitochondrial protein import machinery. Proteins in orange are discussed further in the thesis. The Tim17-like proteins include homologs of Tim17, Tim22 and Tim23. M – mitosome, H – hydrogenosome

2.2 Peroxisomes

Aerobic eukaryotes carry peroxisomes; organelles surrounded by a single membrane (Figure 5), compartmentalizing a variety of metabolic processes, most notably the beta-oxidation of fatty acids and detoxification of reactive oxygen species.

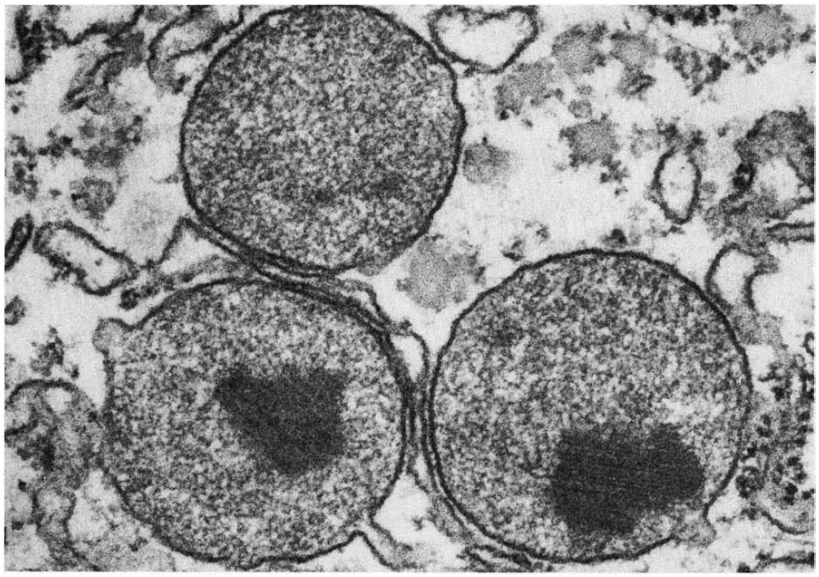


Figure 5. Rat liver peroxisomes. The crystals of the protein urate oxidase form the electron-dense structures inside the peroxisomes. (Fawcett, 1981)

Similar to mitochondria and chloroplasts, peroxisomes propagate by fission, which led some researchers to an idea that peroxisomes are of the endosymbiotic origin (Deduve, 1982). This was disproved by elegant experiments in yeast which have shown that peroxisomes can emerge *de novo* from a sub-compartment of the endoplasmic reticulum (ER) (Hoepfner et al., 2005).

Peroxisomes possess unique protein import machinery that imports proteins in a post-translational manner. It is able to transport folded proteins with bound cofactors or protein complexes (Rucktaschel et al., 2011). Soluble peroxisomal proteins are recognized by specific cytosolic receptors Pex5 and Pex7. The Pex5 receptor recognizes PTS1 (peroxisomal targeting signal), which is a sequence of three amino acids at the extreme C-terminus that has a canonical sequence Ser-Lys-Leu with several possible variations (Gatto et al., 2000). Some other proteins carry a loosely defined motif near the N-terminus called PTS2, which is recognized by the Pex7 receptor (Purdue and Lazarow, 1994). The nematode worm *Caenorhabditis elegans* and the diatoms don't use the PTS2 pathway at all (Motley et al., 2000;Gonzalez et al., 2011).

A protein (cargo) carrying the PTS1 sequence is first recognized by a soluble receptor Pex5 that further binds to a peroxisomal membrane protein Pex14 whereby creating a transient translocation pore (Meinecke et al., 2010) (Figure 6). The cargo is released to the peroxisomal lumen. Afterwards a membrane complex composed of E3 ubiquitin ligases (Pex2, Pex10 and Pex12) attaches a single ubiquitin moiety to the Pex5 receptor. The monoubiquitinated Pex5 receptor is then exported back to the cytoplasm by Pex1 and Pex6 proteins, which both carry two AAA (ATPase associated with diverse cellular activities) domains (Thoms and Erdmann, 2006). The system for recycling of the soluble import receptors is homologous to the ERAD (endoplasmic reticulum associated protein degradation) system, which exports proteins designated for proteasomal degradation from the ER to the cytoplasm (Gabaldon et al., 2006).

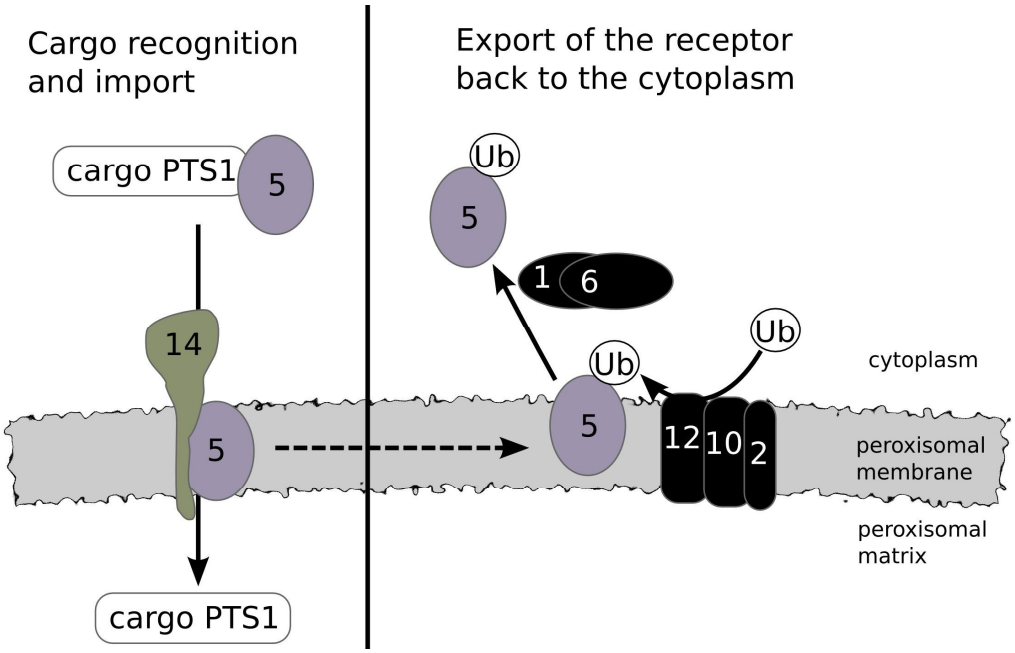


Figure 6. Import of peroxisomal matrix proteins. Pex5 recognizes the PTS1 (peroxisomal targeting signal) sequence, binds to the imported protein and together with Pex14 it forms a transient translocation channel into the peroxisomal matrix. It is then recycled back to the cytoplasm with the help of an ubiquitination machinery (E3 ligases Pex2, Pex10 and Pex12) and cytoplasmic AAA (ATPase associated with diverse cellular activities) proteins Pex1 and Pex6. Proteins homologous to the components of the ERAD (endoplasmic reticulum associated protein degradation) machinery are shown in black.

Hydrophobic proteins designated for the import into peroxisomal membrane are usually recognized by the cytosolic receptor Pex19, which is farnesylated at a conserved cysteine residue near the C-terminus (Sacksteder et al., 2000). In yeast the farnesylation is dispensable for correct function. Pex19 with the bound membrane protein binds to the peroxisomal membrane proteins Pex3 and Pex16 which assist the insertion of the imported membrane protein into the peroxisomal membrane (Figure 7). It was however shown that in some cases the peroxisomal membrane proteins are first imported to the ER and then transported to the peroxisomal membrane in a process that depends on Pex19 and Pex3 (van der Zand et al., 2010).

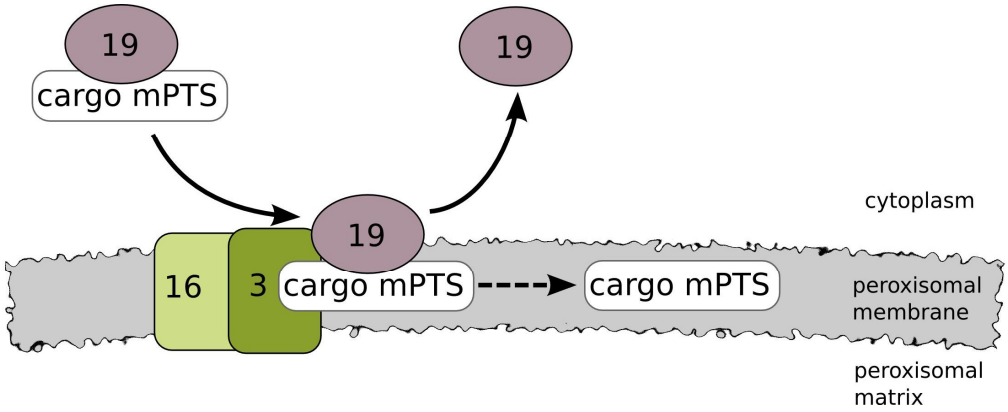


Figure 7. Import of peroxisomal membrane proteins. Insertion of membrane proteins into the peroxisomal membrane depends on the soluble Pex19 and membrane-bound Pex3 and Pex16 proteins.

Peroxisomes are organelles with a great diversity of functions. Table 2 summarizes recent view on the diversity of peroxisomes among eukaryotic groups. Typical peroxisomal metabolism is functionally related to oxygen metabolism and no peroxisomal markers were found in any anaerobic organisms so far, thus it is assumed that peroxisomes were lost in all lineages with MROs (Gabaldon, 2010).

Super-grup	Group	Peroxisomes	Special features
Opisthokonta	Ascomycota	+	glyoxysomes (compartmentalization of the glyoxylate cycle) in *Neurospora crassa*
	Microsporidia	-	
	Animalia	+	
Amoebozoa	Dictyostelia	+	
	Archamoebae	-	
SAR (Stramenopila, Alveolata, Rhizaria)	Bacillariophyceae	+	
	Oomycota	+	
	Ciliophora	+/-	peroxisomes were likely lost in the anaerobic ciliates
	Apicomplexa	+/-	peroxisomes in *Toxoplasma* only
Archaeplastida	Viridiplantae	+	glyoxysomes (compartmentalization of the glyoxylate cycle)
	Rhodophyta	+	
Excavata	Trypanosomatidae	+	glykosomes (compartmentalization of glycolysis)
	Metamonada	-	

Table 2. An overview of the peroxisomal diversity among eukaryotic groups. The groups with MROs are shown in yellow.

2.3 *Model organisms*

In the next section I will introduce the model organisms discussed in the thesis. These are mainly important human parasites (*Entamoeba histolytica*, *Giardia intestinalis* and trypanosomatids) or free-living anaerobes (*Mastigamoeba balamuthi*).

2.3.1 *Archamoebae*

Archamoebae is a monophyletic group of anaerobic protozoa that belongs to the eukaryotic "super-group" *Amoebozoa* (Bapteste et al., 2002). There are free-living genera like *Mastigamoeba* and *Pelomyxa* and parasitic genera *Entamoeba* and *Endolimax*.

Entamoeba histolytica is a world-wide parasite of humans especially common in developing countries. Humans get infected by ingestion of mature cysts. The excystation occurs in the small intestine and the released amoeboid trophozoites migrate to the large intestine, where they multiply by binary fission. Some of the trophozoites produce cysts, which are then passed with the feces to the environment, where they can survive for days. In some infections the trophozoites invade the intestinal mucosa and other organs like the liver, causing life-threatening amoebic colitis and/or amoebic liver abscess.

The cells of *E. histolytica* seem to be simplified. They lack typical mitochondria, instead they possess

tiny mitochondrial remnants - mitosomes. A recent proteomic study revealed the sulphate activation pathway as the main function of the *E. histolytica* mitosomes (Mi-ichi et al., 2009). This is likely a derived function because some of the components originated by a recent lateral gene transfer and the sulphate activation pathway localizes to the cytoplasm or chloroplasts in other eukaryotes. The endoplasmic reticulum and the golgi apparatus cannot be observed on electron micrographs, though the presence of these compartments was demonstrated using specific molecular markers (Mazzuco et al., 1997).

The genome of *E. histolytica* was sequenced, revealing loss of most of the oxygen metabolism-related proteins and acquisition of several bacterial genes through the lateral gene transfer. *E. histolytica* possesses a large repertoire of different receptor kinases pointing at complex signal-transduction pathways important for the pathogenic interactions (Loftus et al., 2005).

Mastigamoeba balamuthi is a free living protozoan found in the anoxic mud or fresh waters. The cells are amoeboid with several nuclei, sometimes carrying single flagellum. An EST (expressed sequence tag) sequencing project revealed genes typical for mitochondrion-related organelles (MROs) which seem to be more complex than the mitosomes of *E. histolytica* (Gill et al., 2007). It has been predicted that the MROs of *M. balamuthi* compartmentalize enzymes of the tricarboxylic acid cycle or the glycine cleavage system. Our laboratory participates in an ongoing genome sequencing project in collaboration with the Laboratory of Genomics and Bioinformatics, Institute of Molecular Genetics of the Czech Academy of Sciences.

2.3.2 *Giardia intestinalis*

G. intestinalis is a world-wide distributed intestinal parasite of mammals. It belongs to an anaerobic group *Metamonada* of the eukaryotic "super-group" *Excavata*. Humans get infected by ingestion of cysts. In small intestine the cysts release trophozoites with eight flagella, two symmetric nuclei and a ventral disc, which they use to attach to the intestinal mucosa. The trophozoites divide in the small intestine. Passage through the large intestine induces encystation and resistant cysts are then released with the feces.

The *G. intestinalis* genomic project revealed a very compact genome with few introns. The machinery for DNA replication, transcription, RNA processing and most metabolic pathways seem to be simplified (Morrison et al., 2007). No proteins responsible for autophagy were identified (Rigden et al., 2009).

The *G. intestinalis* cells contain MROs – mitosomes, which are important for the synthesis of iron-sulphur clusters (Tovar et al., 2003). The genomic data and a mitosomal proteomic project revealed only a very limited number of conserved components of the mitochondrial protein import machinery (Jedelsky et al., 2011). The outer membrane channel Tom40 and the components of the PAM complex (Pam16, Pam18 and Hsp70) were found, while all attempts to identify the TIM complex (translocase of the inner mitochondrial membrane) components were unsuccessful.

Several expression systems were established in *G. intestinalis*. The commonly used systems are episomal vectors with constitutive or inducible expression (Sun et al., 1998).

2.3.3 *Trypanosomatidae*

Trypanosomatidae (trypanosomatids) are parasitic flagellates belonging to the group *Kinetoplastida* of the eukaryotic "super-group" *Excavata*. Several serious human diseases are caused by trypanosomatids: sleeping sickness (*Trypanosoma brucei*), Chagas disease (*Trypanosoma cruzi*) and leishmaniasis (*Leishmania* spp.). I will use a model organism *T. brucei* as an example of trypanosomatid life cycle (Figure 8). Metacyclic trypomastigotes are injected into skin by an infected tsetse fly (genus *Glossina*) during a blood meal on a mammalian host. The metacyclic trypomastigotes enter the bloodstream, transform into bloodstream trypomastigotes and multiply by binary fission. The trypomastigotes infect a tsetse fly taking a blood meal on the infected mammalian host. The bloodstream trypomastigotes transform into procyclic trypomastigotes in the midgut of the tsetse fly and multiply. The procyclic trypomastigotes then transform to epimastigotes, migrate to the salivary glands, multiply and transform to metacylcic trypomastigotes.

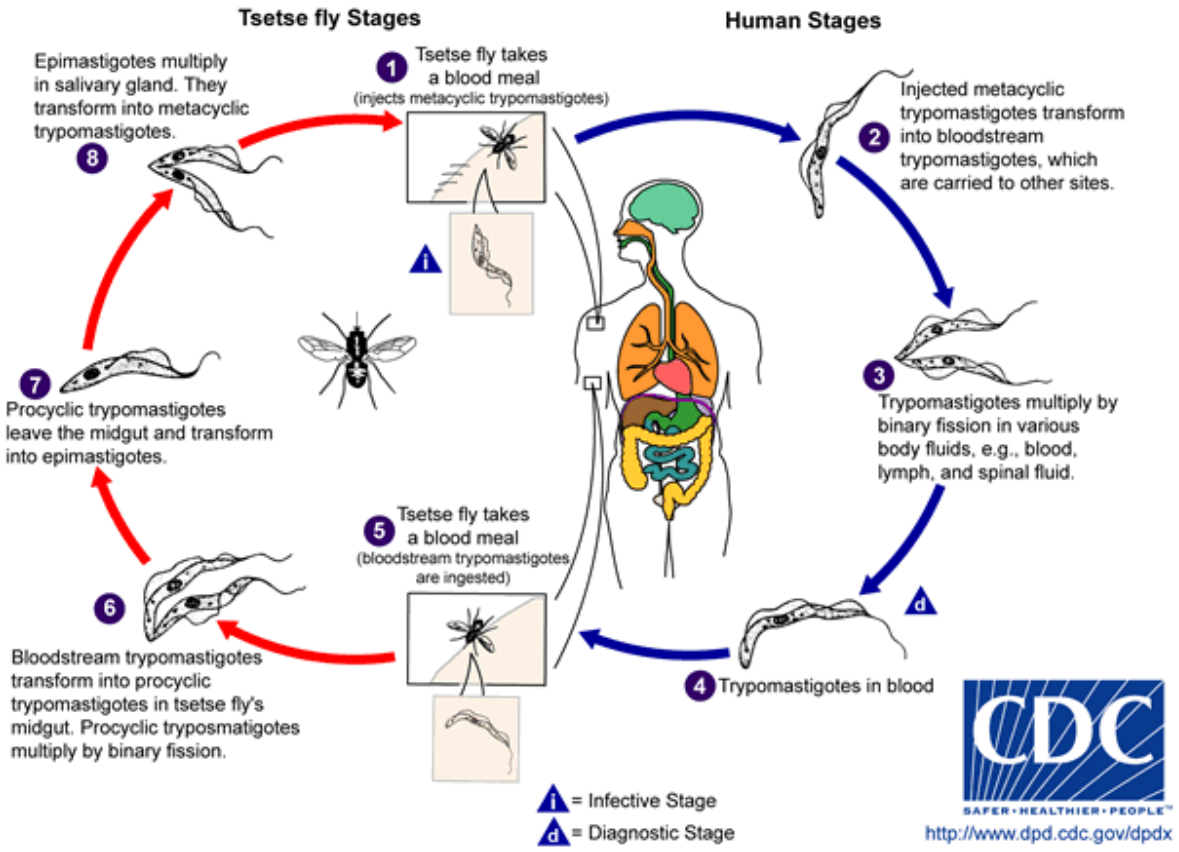


Figure 8. An overview of the *T. brucei* life cycle. The red and blue arrows represent the insect and human stage respectively. (http://www.cdc.gov)

Trypanosomatids contain a single mitochondrion with an unusual DNA-containing structure (kinetoplast) located near the basal body. The mitochondrion of *T. brucei* undergoes many changes during the life cycle. The insect stage mitochondria produce ATP by oxidative phosphorylation as other aerobic mitochondria. However the human bloodstream trypomastigotes rely on energy production by glycolysis. Many mitochondrial functions are suppressed in this stage and the membrane potential at the inner mitochondrial membrane is maintained by the ATP synthase complex at the expense of ATP (Schnaufer et al., 2005).

Genome sequencing projects of several *Trypanosoma* and *Leishmania* species (Berriman et al., 2005) provided a wealth of information about the mitochondrial protein import machineries. There was however a large difference to other aerobic mitochondria, as the mitochondrion of trypanosomatids has been predicted to contain only one TIM complex and no Tom40 homologue has been identified (Schneider et al., 2008). The absence of Tom40 was considered a primitive trait and it was proposed

that the group *Euglenozoa* (a higher taxonomy unit including *Kinetoplastida*, *Diplonemida* and *Eulgenoidea*) lies at the root of the eukaryotic tree of life (Cavalier-Smith, 2010).

In a recent publication Pusnik et al. discovered novel protein of *Trypanosoma brucei* termed ATOM (archaic translocase of the outer mitochondrial membrane) (Pusnik et al., 2011). Using sophisticated experimental methods they found out that ATOM is a beta-barrel protein of the outer mitochondrial membrane that forms the central translocation pore that imports nuclear-encoded proteins to the mitochondrion. They further concluded that ATOM is an ortholog of bacterial YtfM proteins - a subfamily of Omp85 beta-barrel proteins, supporting the basal position of *Euglenozoa*.

3 Methods

3.1 Sources of the genomic data

The sources of genomic data are shown in the Table 3. In the case of *M. balamuthi* there are no gene structure predictions yet. To produce a set of possible peptides in *M. balamuthi* the genomic sequence was translated *in silico* in all possible reading frames and the peptide sequences between the stop codons were extracted.

Organism	Strain	Source
Mastigamoeba balamuthi		img.cas.cz
Entamoeba histolytica	HM-1:IMSS	amoebadb.org
Giardia intestinalis	isolate WB	giardiadb.org
Trypanosoma brucei	TREU927	tritrypdb.org
Trypanosoma congolense	IL3000	tritrypdb.org
Trypanosoma vivax	Y486	tritrypdb.org
Trypanosoma cruzi	strain CL Brener	tritrypdb.org
Leishmania infantum	JPCM5	tritrypdb.org
Leishmania mexicana	MHOM/GT/2001/U1103	tritrypdb.org
Leishmania major	strain Friedlin	tritrypdb.org
Leishmania braziliensis	MHOM/BR/75/M2904	tritrypdb.org
Endotrypanum monterogeii	LV88	tritrypdb.org
Bodo saltans		sanger.ac.uk

Table 3. Sources of the genomic sequence data used for the analyses. The *Mastigamoeba balamuthi* genomic data is not public yet (2012).

3.2 Sequence homology detection

Several homology detection tools were used. Large-scale analyses were run and parsed using in-house Python scripts.

3.2.1 BLAST

The BLAST program is commonly used software for detection of local homology between two sequences (Altschul et al., 1997). It works well to detect close homologs. A BLAST search was used either from the NCBI web server (blast.ncbi.nlm.nih.gov) or as a stand-alone program BLAST+ ver. 2. The NCBI nr (non-redundant) protein database was the default database used for BLAST searches.

PSI-BLAST (position-specific iterative BLAST) is an extension of BLASTP program (BLASTP detects local homology between two protein sequences), which uses iterative database searches for detection of distant protein homologs. In the first round it searches protein database for homologs using BLASTP algorithm. Confident homologs are then selected and based on the frequency of amino acids on the aligned positions a position-specific scoring matrix (PSSM) is created and used for subsequent database search. This helps to detect distant homologs using information from more than one sequence.

3.2.2 HMMER

The HMMER software detects homology between a single sequence and a profile HMM (Hidden Markov model), which is an abstraction of the sequence molecular evolution based on a multiple sequence alignment (Finn et al., 2011). For distant homology searches HMMER proved to be significantly more accurate and sensitive than tools based on simple sequence-sequence comparison (BLAST, FASTA). The HMMER package ver. 3 was used.

3.2.3 HHsearch

HHsearch software detects homology between two profile HMMs and is among the most sensitive distant homology detection tools to date (Soding et al., 2005). A server version of HHsearch (HHpred) creates a profile HMM from a query sequence or sequence alignment using iterative search (HHblits) and then compares the query profile HMM to a variety of profile HMM databases. HHsearch is able to take into account determined or predicted structural features of the profile HMMs. The main output of HHsearch is the probability value, ranging from 0 to 100, 100 being unambiguous homology.

The HHpred web server (http://hhpred.tuebingen.mpg.de/hhpred) or the HHsearch stand-alone software package ver. 2 were used. The hhmake program of the HHsearch package was used to build profile HMMs from raw sequence data.

3.3 Prediction of sequence features

The alpha-helical transmembrane domains were predicted using the TMHMM web server (www.cbs.dtu.dk/services/TMHMM/) or the TMHMM ver. 2 stand-alone program. The N-terminal signal sequences were predicted using the TargetP web server (www.cbs.dtu.dk/services/TargetP/) or the TargetP ver. 1 stand-alone program. The coiled-coil domains were predicted using the COILS web server (http://www.ch.embnet.org/software/COILS_form.html). The prediction of PTS1 sequences was carried out using a Python script that was searching for (S/C/A)-(K/R/H)-(L/M) consensual sequence of the last three amino acids.

3.4 Structural modelling

The structural modelling is based on a rigorous alignment of the analysed sequence with the sequence of known structure. The structural modelling was carried out by the Modeller software ver. 9 (Marti-Renom et al., 2000). The structures were analysed and visualized using the PyMOL software ver. 1.

3.5 Buffers and solutions

PBS pH 7.4

NaCl	8g
KCl	0.2g
$Na_2HPO_4.12H_2O$	1.53g
KH_2PO_4	0.2g
H_2O	1000ml

8M Urea buffer

Urea	480g
NaH_2PO_4	1.2g
Tris-HCl	1.6g
H_2O	1000ml

ST buffer

Saccharose	85.7g
Tris base	1.21g
KCl	37mg
H_2O	1000ml

Comassie Brilliant Blue solution

Coomassie Brilliant Blue	200mg
Ethanol	225ml
H_2O	225ml
Acetic acid	50ml

Destain solution

Ethanol	250ml
H_2O	650ml
Acetic acid	100ml

BCIP/NTB

BCIP	5mg
NTP	10mg
Tris base	0.4g
$MgCl_2$	34g
H_2O	33.3g

3.6 Cultivation media

TYI-S-33

H_2O	870ml
Trypticase Peptone - BBL	20g
Yeast Extract	10g
Glucose	10g
NaCl	2g
K_2HPO_4	1g
KH_2PO_4	0.6g
L-cysteine	2g
L-ascorbic acid	0.2g
Ammonium ferric citrate (2.28g/100ml)	1ml
Inactivated bovine serum, Gibco	100ml
Bovine bile	1ml
Penicillin	600µg/ml
Amikacin	250µg/ml

Because some of the components are heat-sensitive, the medium is sterilized by filtration using the Millipore Filter Steritop 0.22µm. The final pH is set to 6.8 by a NaOH solution.

PYGC

H_2O	1000ml
Proteose-peptone	10g
Yeast extract	10g
Glucose	10.09g
L-cysteine	0.95g
NaCl	5.03g
K_2HPO_4	0.89g
KH_2PO_4	0.67g

LB

H_2O	1000ml
LB (Lysogeny broth)	20g

3.7 Cultivation of anaerobic protists

Anaerobic protists were cultivated in sealed flasks fully filled with medium. *G. intestinalis* was cultivated in TYI-S-33 medium at 37°C. *M. balamuthi* was cultivated in PYGC medium at 24°C.

3.8 Cloning and expression of M. balamuthi *gene fragments*

3.8.1 PCR amplification

Primers that were used for the PCR amplification of the *M. balamuthi* Pex14 and Nudix gene fragments are shown.

Name	Description	Sequence
MbPex14 - Forward	NdeI + MbPex14 18nt F	CATATGCCCCCCGCGCCGGCAGCG
MbPex14 - Reverse	BamHI + MbPex14 non-stop 18nt R	GGATCCGGGCTTGGCGGCGACAGC
MbNudix - Forward	NdeI + MbNudix 18nt F	CATATGGTGCGCGAGCGCTACGCG
MbNudix - Reverse	BamHI + MbNudix non-stop 18nt R	GGATCCGAGCTTGGACTTGTGGGC

M. balamuthi Pex14 and Nudix gene fragments were amplified from the *M. balamuthi* cDNA using following PCR protocol:

H_2O	15.85 µl
PFU buffer	2.5 µl
$MgSO_4$	1.5 µl
DMSO	1.25 µl
cDNA	1 µl
dNTP	0.5 µl
Taq polymerase	0.2 µl
PFU polymerase	0.2 µl
primer F	1 µl
primer R	1 µl

Cycle: 5:00 94°C, 31 x [1:00 94°C, 0:40 60°C, 1:00 72°C], 7:00 72°C

The resulting PCR product was separated on 10% agarose gel and visualized using SYBR Safe dye. The band of corresponding length was excised and the DNA product was isolated using the QIAGEN Gel Extraction Kit.

3.8.2 Cloning

The PCR product was ligated to the pGEM-T Easy vector using manufacturer protocol. The ligation product was transformed to the *E. coli* XL1-Blue competent cells that were then incubated on LB plates with ampicillin (100µg/ml) and X-Gal over night at 37°C. White colonies were screened for the insert using PCR with insert-specific primers. Positive colony was picked and incubated in 5ml LB with ampicillin (100µg/ml) over night at 37°C. The cell culture was pelleted and the plasmid was isolated using the QIAGEN Miniprep Kit. The plasmid was then treated by the NdeI and BamHI restriction enzymes. The restriction products were separated in a 10% agarose gel and visualized using SYBR Safe dye. The band of the insert length was excised and the DNA product was isolated using the QIAGEN Gel Extraction Kit. The insert was ligated into the pET42b plasmid, adding a poly-His tag to the 3' end of the open reading frame. The ligation product was transformed to the XL1-Blue competent cells that were then incubated on LB plates with kanamycin (50µg/ml) over night at 37°C. Colonies were screened for the insert using PCR with insert-specific primers. Positive colony was picked and incubated in 5ml LB with kanamycin (50µg/ml) over night at 37°C. The cell culture was pelleted and the plasmid was isolated using the QIAGEN Miniprep Kit. The correct insertion of gene fragments into the pET42b plasmid was verified by sequencing.

3.8.3 Production of recombinant protein

E. coli Rosetta cells were transformed with the pET42b expression vector carrying the right insert and incubated on LB plates with kanamycin (50µg/ml). One colony was picked and incubated in 200ml LB medium with kanamycin (50µg/ml) over night at 37°C. The culture was added to 2l LB medium and incubated on shaker at 37°C. When the absorbance reached 0.6 at 600nm IPTG was added to final concentration of 100µM. The culture was incubated for 4 hours and then pelleted.

3.9 Purification of recombinant proteins

The protein purification was done under denaturing conditions in 8M urea buffers with different pH. The recombinant His-tagged proteins were purified using following protocol:

- pellet cells and resuspend them in 15ml urea buffer pH 8
- sonicate the cells four times for 1min at amplitude 60 (Bioblock Scientific, Vibra Cell 72405)
- spin the lysate at 150000 x g for 30min
- incubate the supernatant for 15min with 1ml PerfectPro NiNTA Agarose (5 Prime) and transfer the mixture to a column
- collect the flow-through
- wash the column two times with 4ml of urea buffer pH 6.4, collect the flow-through
- wash the column four times with 500µl of urea buffer pH 5.9, collect the flow-through
- wash the column four times with 500µl of urea buffer pH 4.5, collect the flow-through

The fractions were then analysed using SDS-PAGE and Western-blot with an anti-His antibody. Fractions that contained the purified protein were separated using a preparative electrophoresis. The gel was stained in Coomassie Brilliant Blue solution and then washed in the destain solution. The corresponding band was excised and washed in PBS buffer.

3.10 Production of polyclonal antibodies

A polyclonal rabbit antibody against the *M. balamuthi* Pex14 fragment was produced by Eurogentec, Belgium. Polyclonal rat antibodies against the *M. balamuthi* Nudt fragment were produced in-house.

3.11 Cloning and expression of G. intestinalis *genes*

Primers that were used for the amplification of the *G. intestinalis* Tim44 are shown.

Name	Description	Sequence
GiTim44 - Forward	AseI + GiTim44 18nt F	ATTAATATGAAAAGTTTTACGCCC
GiTim44 - Reverse	BamHI + GiTim44 non-stop 18nt R	GGATCCATAGAAATACGTCGGCTT

G. intestinalis Tim44 gene was amplified from the *G. intetinalis* genomic DNA using the following PCR protocol:

H_2O	17.1 µl
PFU buffer	2.5 µl
$MgSO_4$	1.5 µl
cDNA	1 µl
dNTP	0.5 µl
Taq polymerase	0.2 µl
PFU polymerase	0.2 µl
primer F	1 µl
primer R	1 µl

Cycle: 5:00 94°C, 31 x [1:00 94°C, 0:40 53°C, 1:00 72°C], 7:00 72°C

The resulting PCR product was purified and ligated into pGEM-T Easy vector as described in chapter 3.8.2 Cloning.

The pGEM-T vector carrying the GiTim44 open reading frame was cut using AseI and BamHI restriction enzymes. The restriction products were separated in a 10% agarose gel and visualized using SYBR Safe dye. The band of predicted length was excised and the DNA product was isolated using the QIAGEN Gel Extraction Kit. The insert was ligated into the NdeI and BamHI restriction sites of the pONDRA plasmid. The ligation product was transformed to the XL1-Blue competent cells that were then incubated on LB plates with ampicillin (100µg/ml) over night at 37°C. Colonies were screened for the insert using PCR with insert-specific primers. Positive colony was picked and incubated in 150ml LB with ampicillin (100µg/ml) over night at 37°C. The cell culture was pelleted and the pONDRA plasmid was isolated using the Promega Wizard Plus Midiprep Kit. The correct insertion of the GiTim44 gene into the pONDRA plasmid was verified by sequencing.

The transformation of *G. intestinalis* cells was performed using the following protocol:

- start with a 50ml cell culture, discard the medium with floating cells
- add cold sterile PBS buffer and cool on ice for 30min
- spin the cells at 1000 x g 10min 4°C
- wash with sterile TYI-S-33 medium
- resuspend in small volume of sterile TYI-S-33 medium (ca. 1.5 ml)
- count the cells and dilute to the concentration of $3.3x10^7$ cells/ml
- pass the cell through a G23 needle
- transfer 300µl of cell suspension to a 0.4cm gap electroporation cuvette (Bio-Rad)
- add the plasmid (50µg) and incubate on ice for 15min
- electroporate using the exponential protocol: 350V, 1000µF, 750Ω (Bio-Rad, Gene Pulser Xcell)
- incubate on ice for 15min
- transfer the cells to a 7ml tube with TYI-S-33
- after 24h change the medium for fresh TYI-S-33 with 150µg/ml G418 (geneticin)
- change the medium every 24h for 4 days
- on the 5th day change for fresh medium with 600µg/ml G418
- change the medium every 48h for 1-2 weeks

3.12 Cell fractionation by differential centrifugation

Cells of *G. intestinalis* were fractionated using the following protocol:

- start with a 1l culture of *G. intestinalis*, discard the medium with floating cells
- add cold PBS and leave for 1h on ice
- shake the flask thoroughly
- spin the cells at 1200 x g 15min 4°C
- wash the cells in PBS

- wash the cells in ST buffer, resuspend in 50ml ST
- add the protease inhibitors TLCK (50µg/ml) and Leupeptin (10µg/ml)
- sonicate four times: amplitude 40, 1s pulses, for 30 sec (Bioblock Scientific, Vibra Cell 72405)
- spin at 680 x g 10min 4°C
- spin the supernatant at 2760 x g 20min 4°C
- spin the supernatant at 25000 x g 30 min 4°C

The resulting supernatant represents the cytoplasmic fraction and the pellet represents the large granular fraction (LGF) which is specifically enriched by mitosomes.

3.13 SDS-PAGE

Proteins were separated under denaturating conditions in vertical 13.5% polyacrylamide gel with sodium dodecyl sulfate. The samples were dissolved in the SDS sample buffer and denaturated for 5min at 100°C. To determine relative molecular weights, the Fermentas PageRuler Plus prestained protein ladder was used. The gel was stained in the solution of Coomassie Brilliant Blue R and washed in the destain buffer.

3.14 Western-blot

Specific detection of proteins was done by the Western-blot analysis using the following protocol:

- after the SDS-PAGE measure the gel and prepare 6 filter papers and a nitrocellulose membrane of the same size
- wash the gel, filter papers and the nitrocellulose membrane in the blotting buffer for 5min
- stack 3 filter papers, the nitrocellulose membrane, the gel and 3 filter papers on the Biometra blotting machine
- blot at 1.5mA per square cm of the gel for 1h
- for visualization of the blotted proteins, wash in Ponceau S (0.5%) for 1min and wash shortly in H_2O
- incubate the membrane in the blocking buffer (PBS, 2% powdered milk, 0.25% Tween) over night at 4°C or 1h at room temperature

- incubate the membrane with the primary antibody in the blocking buffer over night at 4°C or 1h at room temperature
- wash the membrane 3 times for 15min in the blocking buffer
- incubate the membrane with a secondary antibody conjugated with alkaline phosphatase in the blocking buffer for 1h at room temperature
- wash the membrane 2 times for 15min in the blocking buffer and then wash once in PBS for 10min
- incubate the membrane with the alkaline phosphatase substrate (BCIP/NBT)

3.15 Fluorescence microscopy

The following protocol was used to prepare *G. intetsinalis* cells:

- transfer a drop of concentrated cell culture on the microscope slide (Starfrost), let it dry
- transfer to cold methanol (-20°C) for 5min
- transfer to cold acetone (-20°C) for 5min
- let it dry
- block with 0.25% BSA (0.25% Gelatin Cold water fish, 0.25% BSA, 0.05% Tween20, PBS) for 1h and remove excess buffer
- incubate with the primary antibody in 0.25% BSA for 1h
- wash three times with PBS
- incubate with the secondary antibody 0.25% BSA for 1h
- wash three times with PBS
- cover with Vectashield mouning medium with DAPI and cover with a cover slip
- seal the cover slip with nail polish

4 Results

4.1 Peroxisomal proteins in Archamoebae

During an analysis of the genomic sequences of a free-living anaerobe *M. balamuthi*, we discovered several peroxisomal markers that weren't found in sequence data of any other anaerobic protist before. In subsequent analyses, more peroxisomal markers were recovered from the genomic sequences of *M. balamuthi* and its parasitic relative *E. histolytica*. In following section, we propose a possibility that peroxisomes or peroxisome-like organelles are present in the anaerobic group *Archamoebae*.

4.1.1 Putative peroxisomal markers in the genomes of *E. histolytica* and *M. balamuthi*

The machineries responsible for peroxisomal protein import and the biogenesis of the organelles (the PEX proteins) are highly conserved among all peroxisomes studied so far, therefore they are suitable for *in silico* prediction of peroxisomes (Gould et al., 1990;Gabaldon et al., 2006). Peroxisomal enzymes aren't reliable markers, because the enzymatic content of peroxisomes is known to vary considerably even among closely related lineages. However the presence of a typical peroxisomal enzyme carrying peroxisomal targeting sequence is a strong hint for the presence of a functional peroxisomal protein import and peroxisomes.

In order to detect peroxisomal markers, we created a dataset of established PEX proteins from a wide range of eukaryotes. We then used several tools (BLAST, HMMER) to search for orthologs of peroxisomal proteins among genomic sequences of *E. histolytica* and *M. balamuthi*. The result summarized in the Table 5 shows phyletic distribution of PEX homologues among *M. balamuthi*, *E. histolytica* and selected model organisms. It shows that PEX proteins are well conserved in *M. balamuthi*. Distribution of PEX proteins identified in *E. histolytica* is intriguing, as all the components responsible for recycling of import receptors are missing.

		membrane protein import			matrix protein import				receptor recycling					division
	PEX	3	16	19	14	13	5	7	2	10	12	1	6	11
Amoebozoa	**Dictyostelium**	+	+	+	+	+	+	+	+	+	+	+	+	+
	Mastigamoeba		+	+	+	+	+	+		+	+	+	+	+
	Entamoeba		+	+	+		+							+
Viridiplantae	**Arabidopsis**	+	+	+	+	+	+	+	+	+	+	+	+	+
Excavata	**Trypanosoma**		+	+	+	+	+	+	+	+	+	+	+	+
Fungi	**Saccharomyces**	+		+	+	+	+	+	+	+	+	+	+	+
	Neurospora	+	+	+	+	+	+	+	+	+	+	+	+	+
Metazoa	**Homo**	+	+	+	+	+	+	+	+	+	+	+	+	+
	Caenorhabditis	+		+	+	+	+			+	+	+	+	+
	Drosophila	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 5. Distribution of the PEX proteins among several eukaryotes is shown. *Archamoebae* are highlighted by the orange colour. The full names of the organisms are: *Dictyostelium discoideum*, *Mastigamoeba balamuthi*, *Entamoeba histolytica*, *Arabidopsis thaliana*, *Ttrypanosoma brucei*, *Saccharomyces cerevisiae*, *Neurospora crassa*, *Homo sapiens*, *Caenorhabditis elegans* and *Drosophila melanogaster*.

The BLAST support and domain organization of the recovered PEX homologues are shown in the Table 6 and Figures 9, 10 and 11. Some of the domains are not exclusive for the PEX proteins (e.g. the TPR domains of Pex5), so the BLAST analysis was necessary to correctly classify some of the putative PEX proteins.

	M. balamuthi		**E. histolytica**	
	BLAST hit	E-val.	BLAST hit	E-val.
Pex16	NP_001045732.1 Oryza sativa	3e-10	XP_001624274.1 Nematostella vectensis	0.21
Pex19	XP_001783277.1 Physcomitrella patens	1e-04	ABK23361.1 Picea sitchensis	7e-16
Pex14	XP_003630034.1 Medicago truncatula	0.082	XP_002121505.1 Ciona intestinalis	0.002
Pex13	NP_495513.1 Caenorhabditis elegans	1e-05	-	-
Pex5	XP_637903.1 Dictyostelium discoideum	9e-79	XP_970686.2 Tribolium castaneum	2e-08
Pex7	EFN60167.1 Chlorella variabilis	2e-21	-	-
Pex10	XP_002028288.1 Drosophila persimilis	4e-16	-	-
Pex12	EFA82231.1 Polysphondylium pallidum	6e-14	-	-
Pex1	XP_002116199.1 Trichoplax adhaerens	9e-56	-	-
Pex6	EGG18802.1 Dictyostelium fasciculatum	5e-45	-	-
Pex11	XP_002678589.1 Naegleria gruberi	6e-05	XP_001927674.1 Sus scrofa	1.3

Table 6. Recovered PEX homologs were searched against the NCBI non-redundant protein database. The best hits are shown.

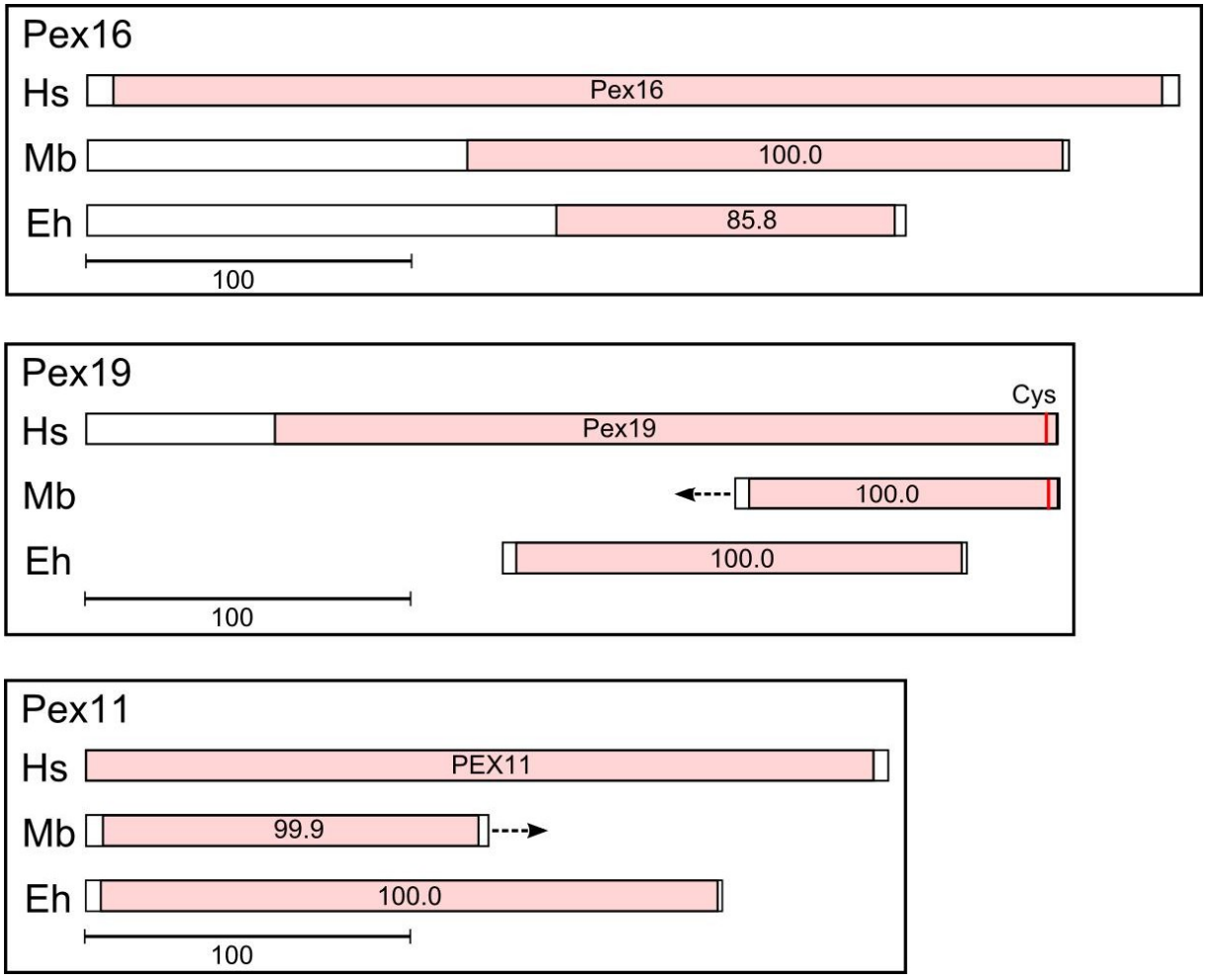


Figure 9. PEX proteins important for membrane protein import and for the peroxisome division. Domains of the human homologue are described by the Pfam name. The numbers in the domains of *M. balamuthi* and *E. histolytica* domains denote the probability value of the HHpred search. The dashed line represents interrupted sequence in *M. balamuthi* homolog. A conserved cysteine is shown in red. The scale represents 100 amino acids. Hs - *H. sapiens*, Mb - *M. balamuthi*, Eh - *E. histolytica*

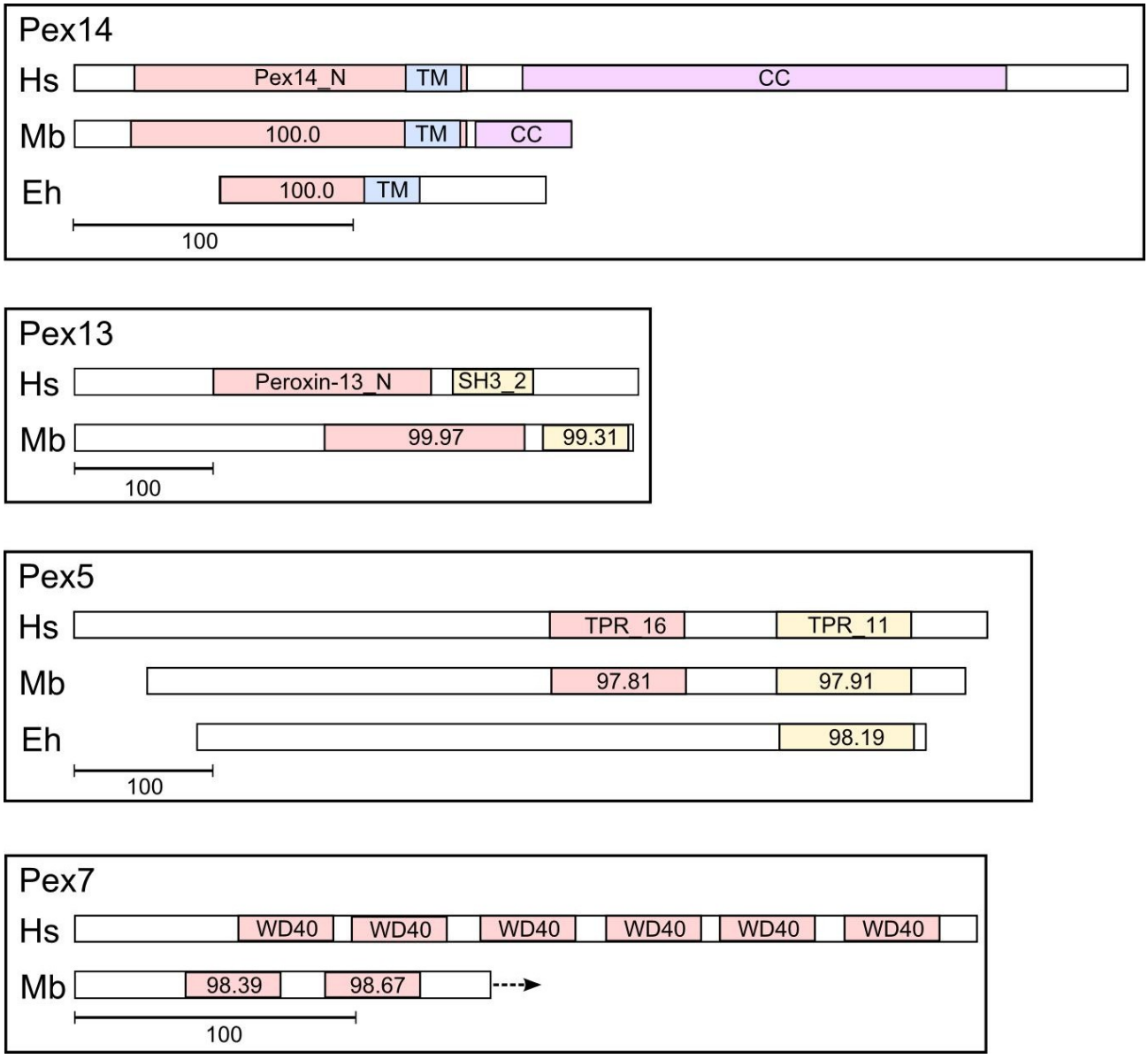


Figure 10. PEX proteins important for matrix protein import. Domains of the human homologue are described by the Pfam name. The numbers in the domains of *M. balamuthi* and *E. histolytica* domains denote the probability value of the HHpred search. The dashed line represents interrupted sequence in *M. balamuthi* homolog. The scale represents 100 amino acids. TM - transmembrane helix, CC - coiled coil, Hs - *H. sapiens*, Mb - *M. balamuthi*, Eh - *E. histolytica*

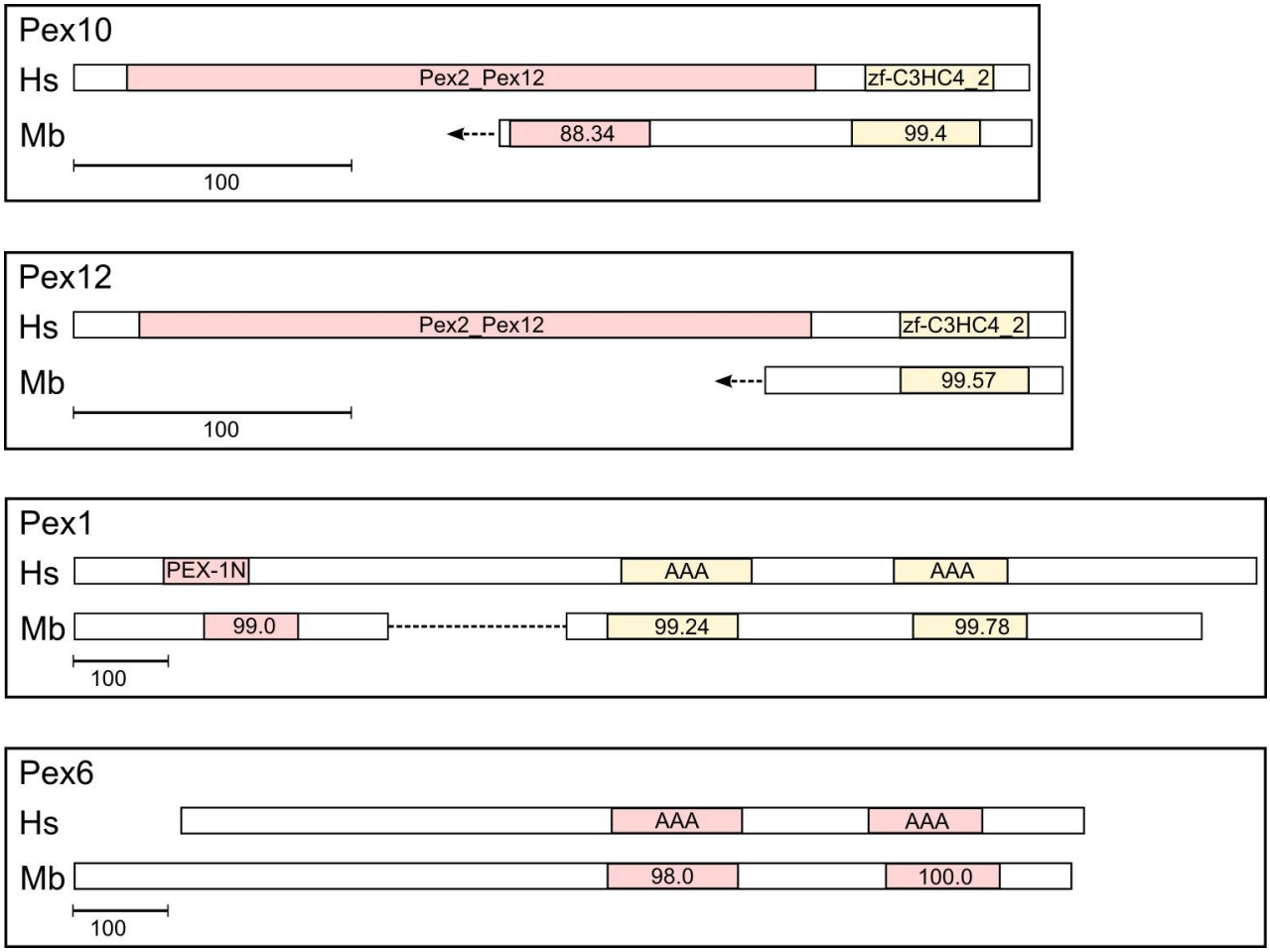


Figure 11. PEX proteins important for recycling of the import receptors. Domains of the human homologue are described by the Pfam name. The numbers in the domains of *M. balamuthi* and *E. histolytica* domains denote the probability value of the HHpred search. The dashed line represents interrupted sequence in *M. balamuthi* homolog. The scale represents 100 amino acids. Hs - *H. sapiens*, Mb - *M. balamuthi*, Eh - *E. histolytica*

We focused on the central component of the protein import machinery – the Pex14 protein, which docks the cytoplasmic import receptors and upon binding with the cargo-receptor complex, it forms a transient translocation pore. The protein is composed of three parts: a conserved N-terminal helical domain, a hydrophobic domain and a coiled coil domain of variable length. A structure of the N-terminal domain of the human Pex14 bound to the Pex5 cytosolic receptor has been solved (Neufeld et al., 2009). Our multiple sequence alignment shows that all the residues indispensable for the Pex14-Pex5 interaction (F35, V41, F52, K56) are conserved in Pex14 homologs of *M. balamuthi* and *E. histolytica* (Figure 12).

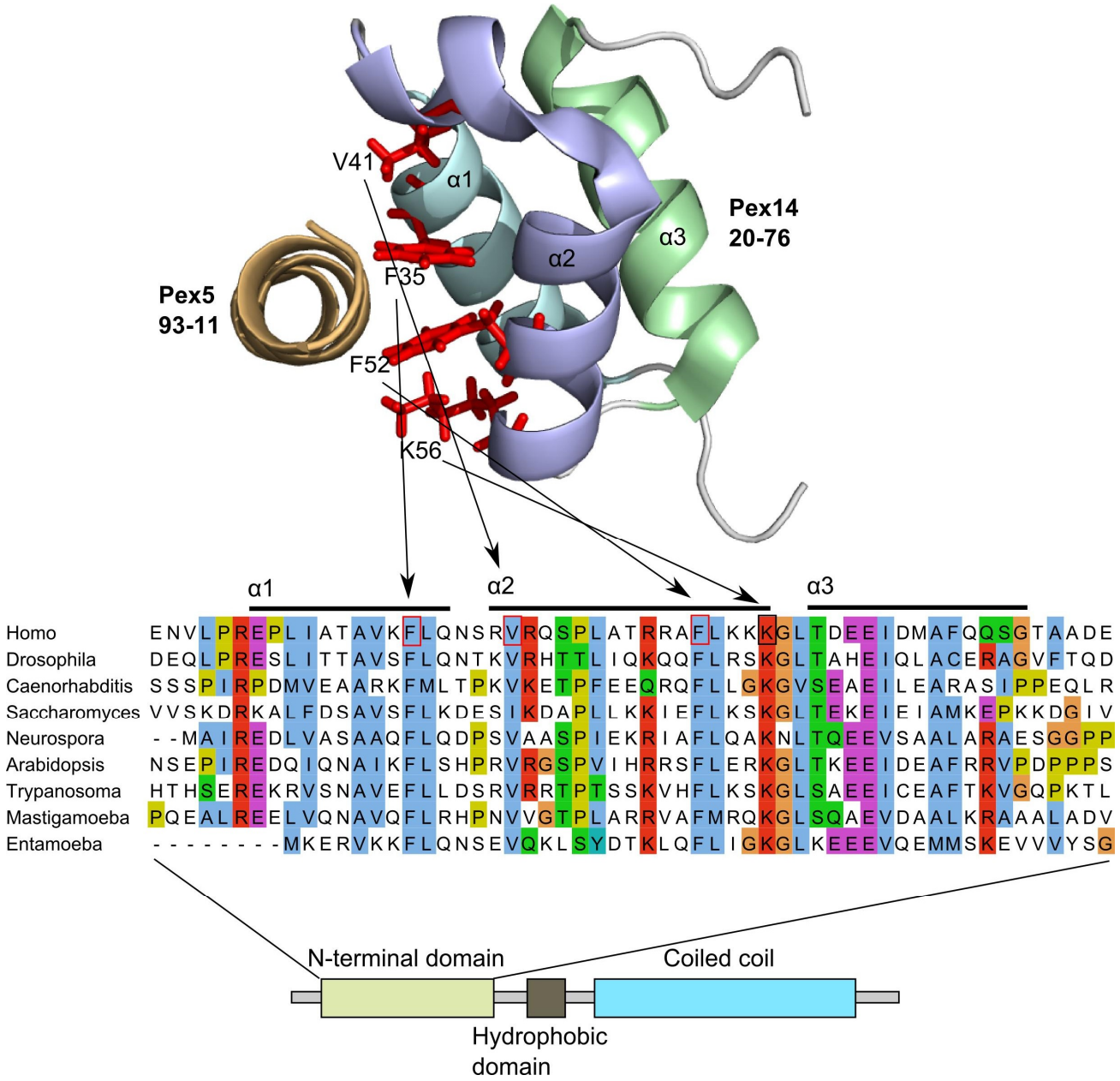


Figure 12. Multiple sequence alignment of the conserved N-terminal domain of Pex14. The structure above is of the human Pex14 and Pex5 proteins (PDB id: 2W84). Residues that are indispensable for the Pex14-Pex5 binding interface are highlighted.

We further searched for typical peroxisomal enzymes and metabolite transporters. No such proteins were identified in *E. histolytica*. Several putative peroxisomal proteins (other than the PEX proteins) were identified in *M. balamuthi* (Table 7). The Peroxisomal leader-peptide processing peptidase (Ppp) cleaves the N-terminal PTS2 signal sequence which is consistent with the presence of the Pex7 PTS2

receptor in *M. balamuthi* (Authier et al., 1995). The AbcD transporters are peroxisomal membrane proteins that facilitate transport of fatty acids or coenzyme A-attached fatty acids (Hettema et al., 1996). The Pmp34 protein is related to the mitochondrial carrier protein family. The human homologue was shown to be localized to peroxisomes and transport CoA, NAD+ and FAD (Agrimi et al., 2012). A nudix (nucleoside diphosphate linked to X) hydrolase homolog with a typical PTS1 sequence on the C-terminus was identified. Nudix hydrolases are known to hydrolyse organic pyrophosphates as ADP-ribose, CoA or NADH (McLennan, 2006). The Nudix hydrolase found in *M. balamuthi* is similar to the Nudt6 human protein, which is of unknown function.

Name	Function	BLAST hit id	E-value	Peroxisomal targeting signal
AbcD	imports fatty acid-CoA	AAL78682.1	4e-47	
Pmp34	transports CoA, NAD+, FAD	XP_002439245.1	2e-17	
Nudix hydrolase	hydrolizes organic pyrophosphates	XP_003389450.1	8e-36	PTS1, -SKL
Ppp	cleaves the N-terminal PTS2 sequence	XP_002609926.1	1e-6	PTS1, -SKL

Table 7. Putative peroxisomal proteins (non-PEX) identified in *M. balamuthi*. A PTS1 sequence at the C-terminus was identified in the two matrix proteins (Nudix hydrolase, Ppp). The best BLAST hits against the NCBI non-redundant database are shown.

4.1.2 Preparation of antibodies against peroxisomal markers of *M. balamuthi*

Pex14 and the Nudix hydrolase (Nudt) were selected as suitable markers for localization of the putative peroxisomes. Pex14 is a peroxisomal membrane protein and Nudt a matrix protein with a PTS1 sequence at the C-terminus. Gene fragments of 558bp and 684bp supported by homology to known proteins were used. The gene fragments were amplified from *M. balamuthi* cDNA using an optimized PCR protocol as the coding sequences of *M. balamuthi* are very GC rich (over 70% GC). The gene fragments were cloned to the pET42b expression vector (adding a C-terminal poly-His tag) and expressed in *E. coli* Rosetta cells. The proteins were purified using a nickel column under denaturating conditions. The purified Pex14 protein was sent to Eurogentec (Belgium) to raise antibodies in rabbits. We are currently working on an in-house production of rat anti-Nudt antibodies.

The anti-MbPex14 antibody and the pre-bleed serum (used as a negative control) were tested against the *M. balamuthi* cell fractions. The western-blot (Figure 13) shows a specific double band in the whole cell lysate (molecular masses ca. 22-23 kDa). There is no signal in the cytoplasm and a strong signal in

the large granular fraction with third emerging band, supporting localization in a small organelle. The bands of different lengths might be caused by alternative transcripts and/or by posttranslational processing.

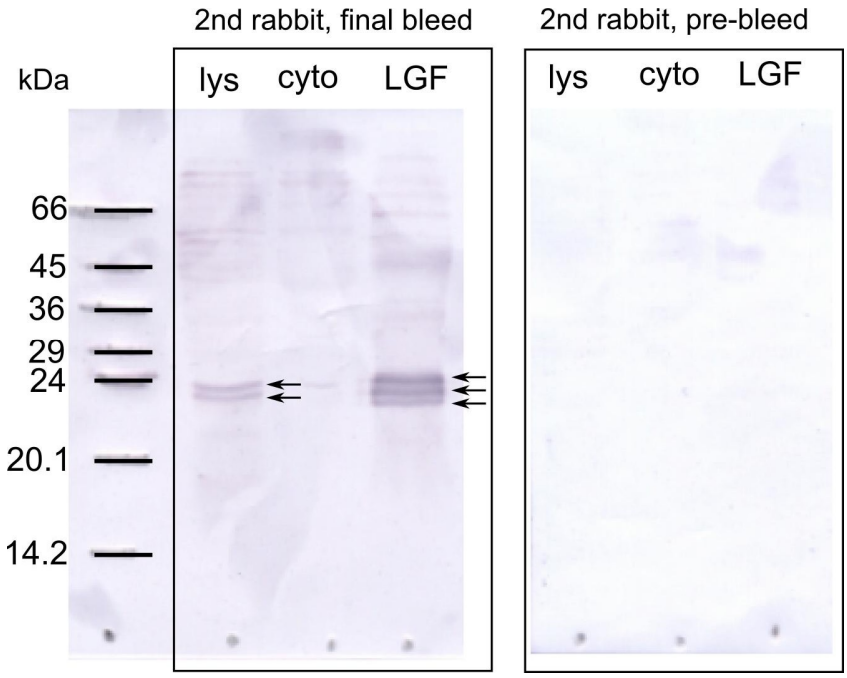


Figure 13. A Western-blot of the anti-MbPex14 antibody and the pre-bleed serum tested on *M. balamuthi* cell fractions. lys – whole cell lysate, cyto – cytoplasmic fraction, LGF – large granular fraction

4.2 Tim44 subunit of the mitochondrial translocase is conserved in G. intestinalis

In a high-throughput bioinformatics analysis of the predicted proteins of *G. intestinalis*, we discovered a highly divergent homologue of the Tim44 subunit of the TIM complex. In following experiments, we confirmed the mitosomal localization of the newly discovered Tim44.

4.2.1 Detection of the components of the mitosomal translocase in *G. intestinalis*

Predicted protein sequences of *G. intestinalis* (isolate WB) were transformed to a database of profile HMMs. We then searched the database using the HHsearch program and profile HMMs that represent

conserved components of the mitochondrial protein import machinery. All the previously established components (Tom40, Pam16, Pam18, MtHsp70) were detected and additionally a possible Tim44 homolog (GiardiaDB id: GL50803_14845) was found. The predicted protein has 286 amino acids and molecular weight of 32.99 kDa. This protein was detected during the mitosomal proteomic project and annotated as a hypothetical protein (Jedelsky et al., 2011).

BLAST searches recovered close homologs of GiTim44 in other *G. intestinalis* isolates, but no homologs from other organisms were found. We then searched an alignment of GiTim44 sequences of three *G. intestinalis* strains against several independent databases of conserved domains using the HHpred server. The two best hits were among the Pfam database family PF04280 (HHpred probability 93.1) and the SCOP database family d.17.4.13 (HHpred probability 92.5) which both represent Tim44 homologs. We consider this a strong support for the homology between the proposed GiTim44 and other Tim44 protein sequences.

We modelled possible structure of GiTim44 based on the structure of the conserved C-terminal domain of the human Tim44 (PDB id: 2CW9) (Figure 14). The C-terminal domain of Tim44 proteins forms a large pocket that facilitates interaction with the phospholipids of the inner mitochondrial membrane. Our structural modelling data shows that the GiTim44 protein might form this characteristic pocket.

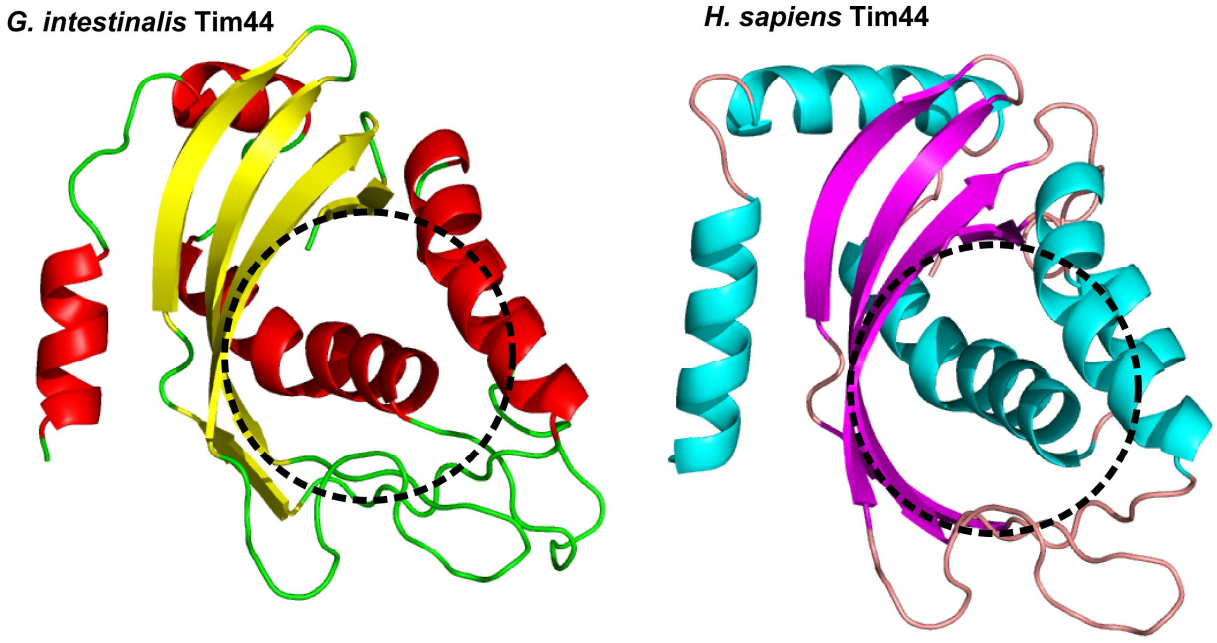


Figure 14. Structure of the C-terminal domain of *G. intestinalis* Tim44 (left) based on the structure of *H. sapiens* Tim44 (right, PDB id: 2CW9). The characteristic membrane-binding pocket is highlighted by a dashed circle.

4.2.2 Tim44 localizes to the mitosomes of *G. intestinalis*

Eukaryotic Tim44 typically localizes to mitochondria, where it is part of the TIM23 complex of the mitochondrial preprotein translocase. We utilized an expression system to assess the localization of the putative Tim44 of *G. intestinalis* (GiTim44). We first cloned the GiTim44 gene (GiardiaDB id: GL50803_14845) into the pONDRA expression vector. *G. intestinalis* cells were transformed using electroporation and selected using geneticin.

We first tested the expression of the HA-tagged GiTim44 protein (GiTim44-HA) using the western blot analysis. A specific signal of an appropriate size was found in the whole cell lysate and the large granular fraction, but no signal was detected in the cytosolic fraction (Figure 15). This supports the mitosomal localization of GiTim44.

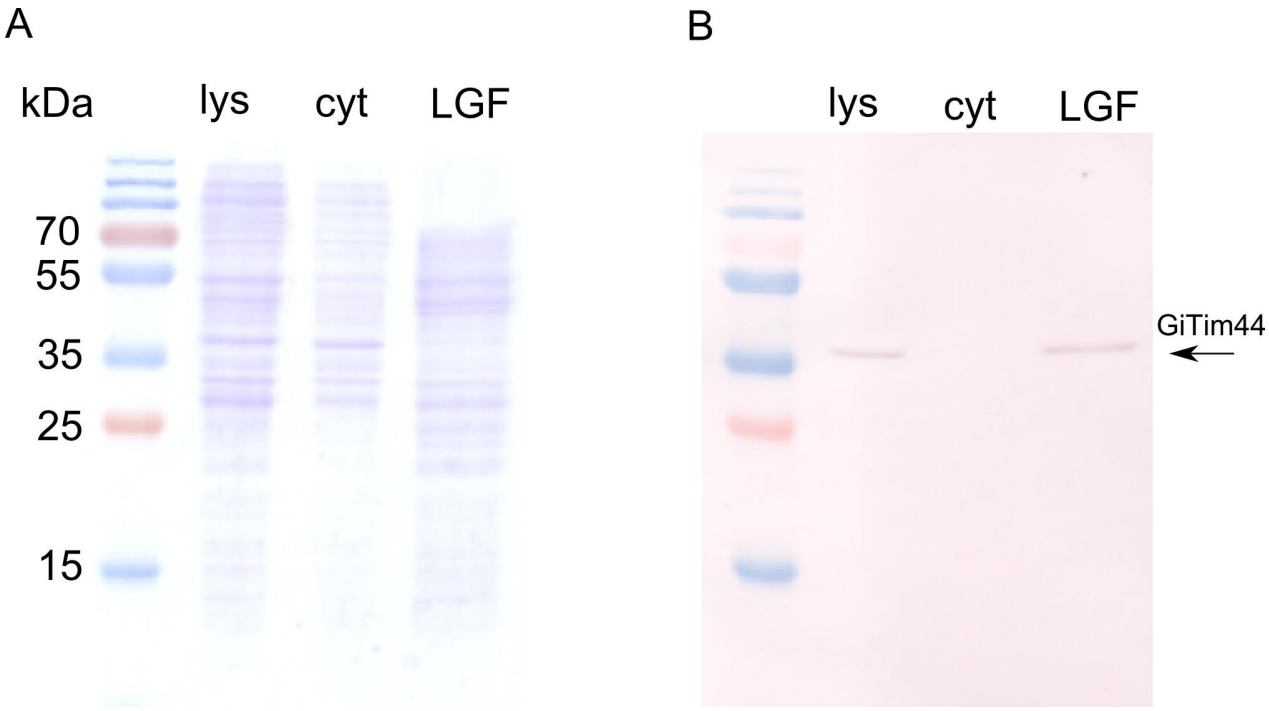


Figure 15. GiTim44 localization in cellular fractions of *G. intestinalis*. (A) An SDS-PAGE gel stained with Coomassie Brilliant Blue serves as a loading control. (B) A Western-blot incubated with an anti-HA antibody. The arrow shows a specific signal in the whole-cell lysate and the large granular fraction only. lys – whole cell lysate, cyt - cytoplasm, LGF - large granular fraction

We further used fluorescence microscopy to assess the subcellular localization of the GiTim44-HA protein (Figure 16). The anti-Cpn60 antibody recognizes the *G. intestinalis* Cpn60 chaperone which

serves as a mitosomal marker. The nuclei were stained with DAPI. The fluorescence revealed that only about 10% of the cells expressed the recombinant GiTim44-HA protein. In the cells expressing the GiTim44-HA the anti-HA antibody recognized specific vesicles that colocalized with the anti-Cpn60 signal which serves as a mitosomal marker, confirming mitosomal localization of the recombinant GiTim44-HA.

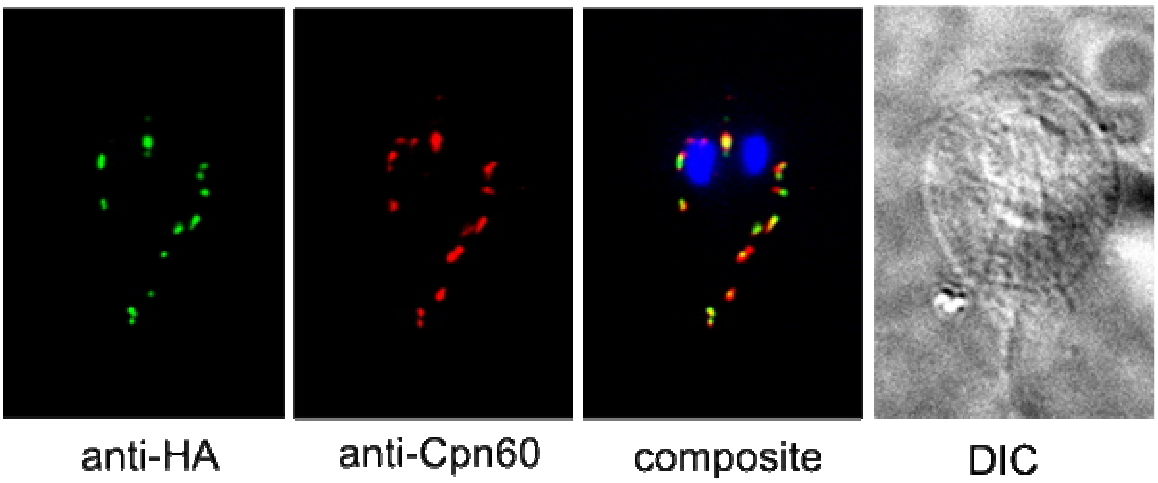


Figure 16. Fluorescence microscopy of fixed *G. intestinalis* trophozoites expressing the GiTim44-HA construct revealed colocalization of the recombinant GiTim44-HA with the mitosomal marker Cpn60. Nuclei are stained blue with DAPI. DIC - Differential interference contrast microscopy

4.3 The mitochondrial protein translocase in trypanosomatids is related to Tom40

In order to reveal the evolutionary history of the recently described trypanosomatid mitochondrial preprotein translocase ATOM (Pusnik et al., 2011), we assessed the distribution of ATOM sequences among kinetoplastids and used profile HMM-based tools to detect homology between ATOM and established protein families. Our results show that ATOM is homologous to the eukaryotic Tom40 protein family.

4.3.1 ATOM is conserved among kinetoplastids

Using BLAST, we searched sequence databases of diverse trypanosomatids on TritrypDB (http://tritrypdb.org/tritrypdb/) and the genome data of *Bodo saltans* (a free-living kinetoplastid) at sanger.ac.uk (http://www.sanger.ac.uk/resources/downloads/protozoa/bodo-saltans.html). We recovered

ATOM sequences among all the *Trypanosoma* and *Leishmania* species as reported before (Pusnik et al., 2011). Furthermore we found ATOM homologs in the genomes of *Endotrypanum monterogeii* (trypanosomatid related to the *Leishmania* clade, parasite of the sloth) and *Bodo saltans* (a free-living kinetoplastid).

4.3.2 ATOM is related to Tom40

To assess the homology of ATOM sequences to other protein families, we searched each single ATOM sequence against the Pfam database (Figure 17). The ATOM sequences from the *Trypanosoma* clade do not specifically recognize any protein family. On the other hand, sequences from the *Lesishmania* clade and especially the *Bodo saltans* ATOM are recognized as homologs of the Eukaryotic porin (PF01459) protein family with e-values between 0.2 and 0.0003. The Eukaryotic porin family represents the Tom40 and VDAC sequences. No protein families that represent the YtfM or Omp85-like proteins were found.

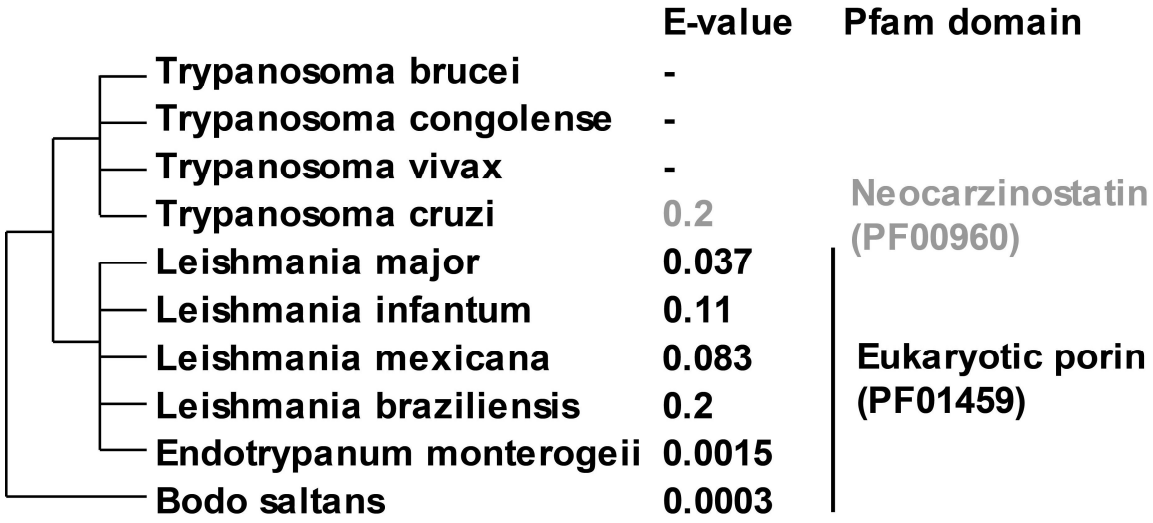


Figure 17. ATOM homologues were found in the genomes of several *Trypanosoma* and *Leishmania* species and also in the genomic sequences of *Endotrypanum monterogeii* and a free-living kinetoplastid *Bodo saltans*. HMMER searches of individual ATOM sequences against Pfam database are shown. ATOM sequences of the *Leishmania* clade and of *Bodo saltans* were recognized as being homologous to the Eukaryotic Porin family which represents both VDAC and Tom40 sequences. The Neocarzinostatin hit is likely false-positive.

We further constructed an alignment of ATOM sequences and searched the Conserved Domains

Database (CDD) using the HHpred server (Table 8). The best hit is the Tom40 protein family with significant statistical support (Probability 92.6 and E-value 0.037). The length of the stretch of recovered homology (251 positions) is also indicative of overall structure conservation. The second hit represents the Mitochondrial porin family, which includes Tom40 and VDAC sequences. The third hit represents VDAC protein family which is related to Tom40 proteins. There was no indication of homology between ATOM sequences and Omp85-like protein family. For comparison, the protein family that includes the YtfM proteins is shown at the bottom of the table.

Id	Hit nr.	Name	Probability	E-value	Aligned positions
cd07305	1	Porin3_Tom40	92.6	0.037	251
cd07303	2	Porin3	74.3	2.1	208
cd07306	3	Porin3_VDAC	39.1	18	131
cd01529	4	4RHOD_Repeats	34.8	8.7	23
cd04762	5	HTH_MerR-trunc	23.8	15	17
	...				
TIGR03303	5770	OM_YaeT	0.3	8.8e+3	46

Table 8. HMM profile of aligned ATOM sequences was searched against the Conserved Domains Database (CDD) using the HHpred server. ATOM is specifically recognized as a homologue of the Tom40 protein family. The protein family that includes the YtfM proteins is shown at the bottom.

A multiple sequence alignment of ATOM sequences revealed a conserved motif in the last beta-strand called the beta-signal (Figure 18). The beta-signal functions as a sorting signal for mitochondrial beta-barrel proteins inserted into the outer mitochondrial membrane by the SAM complex (Zeth, 2010).

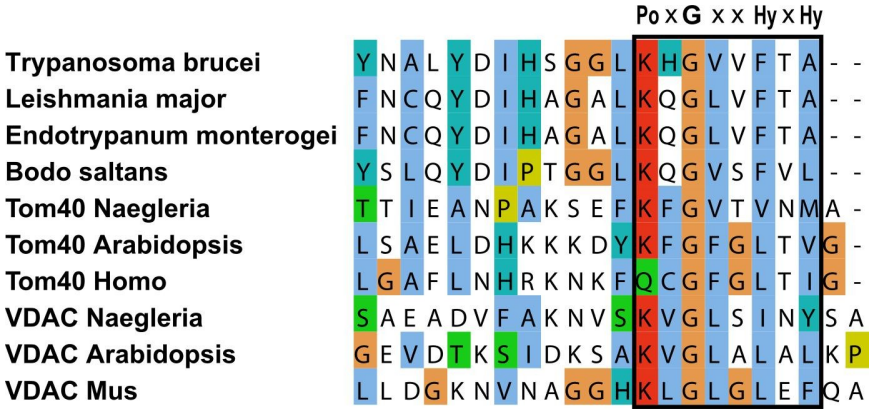


Figure 18. A sequence alignment of the C-termini of representative ATOM, Tom40 and VDAC sequences. The beta-signal is highlighted. Po – polar amino acid, x – any amino acid, G – glycine, Hy – hydrophobic amino acid

In order to get insight into the possible structural organization of ATOM, we utilized a known structure of mouse VDAC (MmVDAC) protein to model structure of ATOM under the assumption that ATOM and VDAC are distantly homologous. Because the sequence of *Bodo saltans* ATOM (BsATOM) seems to be the least divergent, we modelled the BsATOM. First we created an alignment between the BsATOM and MmVDAC using the "align2d" option of the MODELLER software, which apart from usual sequence based alignment takes into account the structural information of the template (MmVDAC). The resulting alignment was then used to model the BsATOM based on MmVDAC (PDB id: 3EMN) (Ujwal et al., 2008). The proposed model of BsATOM was visualized using the PyMOL software. The modelled structure of BsATOM reveals a 19 beta strand structure and an N-terminal alpha-helical region which is typical for the VDAC and Tom40 protein family (Figure 19). It is important to note, that this result strongly depends on the template and the alignment used, and it shows that the ATOM sequence is potentially compatible with the structure of the VDAC/Tom40 protein family.

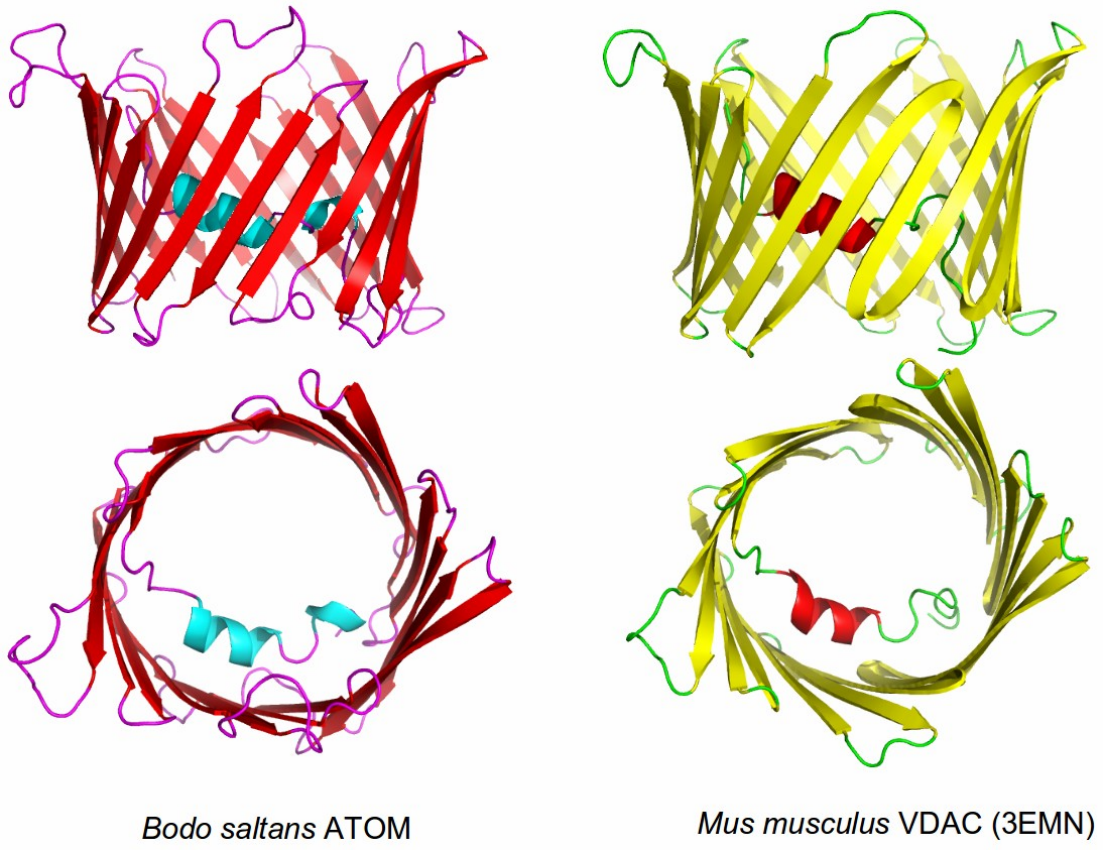


Figure 19. The structure of *Bodo saltans* ATOM was modelled based on the *Mus musculus* VDAC structure (PDB id: 3EMN) using the MODELLER software.

5 Discussion

5.1 Peroxisomal proteins in Archamoebae

Peroxisomes are eukaryotic organelles that usually compartmentalize oxygen-dependent enzymatic reactions such as the beta-oxidation of fatty acids. No peroxisomes were described in any anaerobic organism so far (Gabaldon, 2010).

Conserved eukaryotic proteins required for the protein import and biogenesis of peroxisomes are called the PEX proteins and they are the most universal molecular markers of peroxisomes (Gould et al., 1990;Gabaldon et al., 2006). We discovered several putative PEX proteins in the genomic sequences of a parasite *Entamoeba histolytica* and a free-living amoeba *Mastigamoeba balamuthi*, both obligate anaerobes of the group *Archamoebae*. Additionally several proteins possibly related to the metabolism and metabolite exchange were discovered in *M. balamuthi* (Figure 20). The Nudix hydrolase protein (NudT) belongs to a group of enzymes that hydrolyse the pyrophosphate bond of organic pyrophosphates. Interestingly all the cofactors that might be imported into the peroxisomes (using the Pmp34 and AbcD transporters) possess this bond, so the peroxisomes of *M. balamuthi* might be involved in the metabolism of these essential cofactors (FAD, CoA, NAD+).

The genome sequencing project of *M. balamuthi* is still in progress now. Currently the sequences were assembled to 25675 contigs and 1788 scaffolds with overall size of 57.64 Mb. The continuous stretches of the genomic sequence are still relatively short, which makes it difficult to predict whole-length gene structures for genes of interest. Further genomic DNA and cDNA sequencing will aid for better gene structure predictions and subsequently for better predictions of putative peroxisomal proteins based on the homology to known peroxisomal proteins and the presence of peroxisomal targeting sequences.

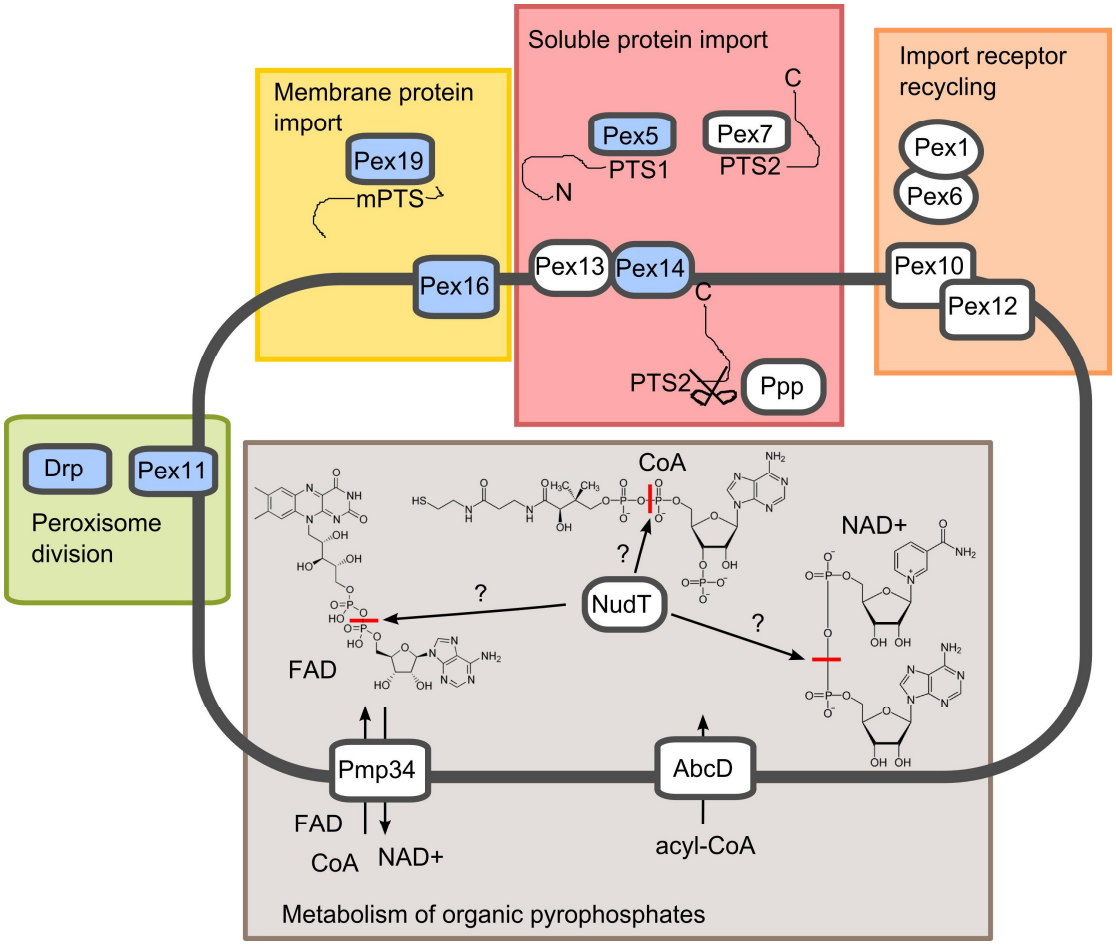


Figure 20. Predicted components of the putative *M. balamuthi* peroxisomes. Proteins shown in blue were identified in *E. histolytica*. Possible targets of the Nudix hydrolase are highlighted by arrows.

The putative peroxisomal protein import machinery of *E. histolytica* seems to be rather simplified. The lack of the Pex7 receptor points at a loss of the PTS2 import signal, as it happened independently in the nematodes and diatoms (Motley et al., 2000;Gonzalez et al., 2011). The most striking feature of the putative PEX machinery of *E. histolytica* is the absence of all the proteins required to recycle the soluble import receptors back to the cytoplasm. Interestingly all these proteins are homologous to the proteins of the ERAD machinery (Endoplasmic reticulum-associated protein degradation) (Schliebs et al., 2010). There is also similarity between the peroxisomal import-receptor recycling machinery and ERAD on the functional level, as in both cases the ubiquitination of a specific protein leads to its ATP-dependent export from the organellar lumen or membrane to the cytoplasm. An ongoing intimate connection between the ER and the peroxisomes has been shown, as peroxisomes are able to emerge *de novo* from special ER sub-compartments (Hoepfner et al., 2005) and several peroxisomal membrane

proteins cross the ER first during the import (van der Zand et al., 2010).

The status of the ER in *E. histolytica* is unresolved, because no typical ER was observed on the ultrastructure (Rosenbau and Wittner, 1970). The molecular markers of ER were however found in *E. histolytica* (Mazzuco et al., 1997). Recent 3D reconstructions of *in vivo* fluorescence images even suggest a fine net-like structure of the ER compartments in *E. histolytica*, similar to other eukaryotes (Teixeira and Huston, 2008).

One possible and exciting explanation for the simplified import machinery in *E. histolytica* is that the peroxisomes of the lineage leading to *E. histolytica* have undergone a great reduction in the functions and mechanisms of the biogenesis. These peroxisomes might have been reduced to a sub-compartment of the ER and the import receptors got translocated to the cytoplasm by modified ERAD machinery. In another model some of the ERAD components might be dually targeted to the ER and peroxisomes, where they recycle the peroxisomal import receptors. It is also possible that in *E. histolytica* the peroxisomal protein import isn't functional at all.

We have shown a possibility that peroxisomes were retained in an anaerobic lineage *Archamoebae*. In the near future we will hopefully get new insights into the nature of these compartments by localization experiments using specific antibodies and possibly by the use of fine cell-fractionation and proteomics.

5.2 Tim44 subunit of the mitochondrial translocase is conserved in G. intestinalis

Cells of *G. intestinalis* were shown to contain mitochondrial relicts called the mitosomes. We utilized a powerful tool for distant homology detection (HHsearch) (Soding et al., 2005) to find homologs of conserved components of the mitochondrial protein translocases in the sequence data of *G. intestinalis*. We confirmed previously established subunits of the TOM (translocase of the outer mitochondrial membrane) and PAM (presequence translocase-associated motor) complexes (Tom40, Pam16, Pam18 and Hsp70) and additionally we detected a possible homolog of Tim44 which we termed GiTim44.

Further bioinformatics analyses showed a moderate statistical support for GiTim44 being true homologue of the Tim44 protein family (HHpred probability 93.1 and e-value 0.047) using the HHsearch software, however no other tools (PSI-BLAST, HMMER) were able to detect any homology between GiTim44 and the Tim44 protein family.

Eukaryotic Tim44 proteins usually contain an N-terminal mitochondrial targeting sequence (MTS) and localize to the mitochondrion. The GiTim44 doesn't have any predicted MTS, so we questioned the

subcellular localization of GiTim44. Western blot of subcellular fractions and immunofluorescent microscopy revealed a mitosomal localization of expressed GiTim44-HA which is consistent with GiTim44 being a true Tim44 homolog.

In typical aerobic mitochondria Tim44 is part of the TIM23 complex (a complex responsible for import of soluble mitochondrial proteins), where it tethers the PAM complex to the translocation channel that is likely formed by the Tim17 and Tim23 proteins, which both belong to the Tim17-like protein family (Voos et al., 1996;Dolezal et al., 2006). No Tim17-like proteins were identified in *G. intestinalis* so far, which raises an important question how proteins are transported into the mitosome. The extreme sequence divergence of the newly identified GiTim44 shows that it is possible that the mitosomes utilize a channel formed by Tim17-like proteins as other eukaryotes, but the sequences are too divergent to be detected by our bioinformatics tools. It is also possible that the mitosomal translocon of *G. intestinalis* is formed by newly acquired components or even by a translocase from another compartment (SecY as proposed by Martincova and Dolezal). Altogether, the newly identified GiTim44 is a promising tool for future attempts to isolate the mitosomal translocase (by Blue native PAGE, immunoprecipitation).

5.3 The mitochondrial protein translocase in Trypanosomatidae

The TOM (translocase of the outer mitochondrial membrane) complex seems to be conserved among most eukaryotes. The central component of the TOM complex is the Tom40 beta-barrel protein which is related to eukaryotic VDAC (voltage-dependent anion channel) proteins (Zeth, 2010). Complete genomes of several trypanosomatids revealed sequences homologous to eukaryotic VDAC proteins but no homologs of Tom40 were found (Pusnik et al., 2009). Cavalier-Smith considered the absence of Tom40 in euglenozoans (a higher taxonomic unit including the trypanosomatids) as one of the ancestral features of euglenozoans that place them to the root of the eukaryotic tree (Cavalier-Smith, 2010).

Pusnik et al. discovered and experimentally confirmed the protein translocase of the outer mitochondrial membrane which they named ATOM (archaic translocase of the outer mitochondrial membrane) (Pusnik et al., 2011). They used PSI-BLAST for detection of homologous sequences. They used the *T. brucei* ATOM as the query for the first round of BLAST search, which revealed ATOM homologues among trypanosomatids with high statistical support followed by a hit with low statistical support (E-value 0.16) within bacterial YtfM proteins (*Serratia proteamaculans*).The YtfM hit was included in the next round of PSI-BLAST, so the PSSM (position-specific scoring matrix) was build using the ATOM sequences and the YtfM hit. The second round of PSI-BLAST revealed homology to

other enterobacterial YtfM proteins with radically increased statistical support (E-value 7e-44). We think this support is artificial, as the initial statistical support for inclusion of *S. proteamaculans* is well beyond confident values and no other of the trypanosomatid ATOM sequences find any YtfM proteins in the PSI-BLAST analysis. It is also noteworthy that the CLANS analysis used by Pusnik et al. is useful to classify sequences, but it doesn't prove homology.

In our analyses we discovered new ATOM orthologs among kinetoplastids, most notably in the genomic sequence of *Bodo saltans* (a free-living kinetoplastid). We then used profile HMM-based tools to detect homologous protein families of ATOM. Two independent methods (HMMER, HHsearch) revealed no obvious homology to YtfM or any bacterial beta barrels whatsoever. Furthermore our results show that ATOM is likely a divergent homologue of the Tom40 protein family (Zarsky et al., 2012). Using structural modelling, we have also shown that ATOM is potentially compatible with the 19 beta strand structure that is typical for the Tom40/VDAC protein family and which is principally different from the 16 beta strand structure of Omp85-related proteins such as YtfM.

In reaction to our published results, Pusnik et al. correctly point out, that previous attempts to identify Tom40 in trypanosomatids failed, even though comparable tools were used successfully to identify highly diverse Tom40 sequences in *Entamoeba* and *Giardia* (Dolezal et al., 2010;Dagley et al., 2009). We think that the previous studies didn't sufficiently explore the available trypanosomatid sequences and the best-performing homology detections tools were omitted.

Pusnik et al. further argue that ATOM protein sequence has predicted secondary structure consistent with a POTRA-type amino terminal domain (this is specific for proteins of the Omp85 family) and that the predicted beta barrel domain has a comparable size to other members of the YtfM protein family. In contrary, our structural modelling data shows that the ATOM sequence is consistent with the 19 beta strand structure of the Tom40/VDAC protein family.

Pusnik et al. proposed a possible evolutionary model stating that Tom40 evolved from ATOM. This seems to be partly consistent with both conclusions of Pusnik et al. and our results as it would explain the "Tom40-type sequence features" of ATOM. However this scenario is unlikely. First, our analysis shows that there is no support for relation between ATOM and bacterial YtfM proteins whatsoever making this model redundant. Second, this scenario is unparsimonous if we consider that Tom40 and VDAC are clearly homologous protein families and that there are several VDAC homologs in each of the trypanosomatid genome.

It was proposed that the mitochondrial protein translocation machinery of trypanosomatids might

represent an ancestrally simplified primitive form (Schneider et al., 2008). Our results show that the central component of the eukaryotic TOM complex is likely present though its sequence is extremely divergent. We think it is likely that more typical eukaryotic components of the mitochondrial translocase are present in trypanosomatids and their discovery is hindered by the vast sequence divergence. We further conclude that the mitochondrial protein import machinery of trypanosomatids is rather derived than ancestrally primitive.

6 Conclusions

In this work I highlight the vast divergence of peroxisomes and mitochondria known and yet to be discovered. I also point to the unifying characteristics; the protein translocation machinery is clearly one of the basic building blocks of cellular organelles. In comparison to enzymes whose substrates might be unchanged for billions of years, protein translocases rely heavily on protein-protein interactions which have relatively loose constraints for (co)evolution. The divergence of such sequences is most striking in anaerobic and/or parasitic organisms. Results of this thesis emphasize that some protein homologs believed to be lost (or never acquired) might be present in certain organisms however too divergent to be easily identified. To counter this it is important to use diverse sequence sampling, right sequence analysis tools and utilize experimental approaches.

7 References

Agrimi, G., A.Russo, P.Scarcia, and F.Palmieri. 2012. The human gene SLC25A17 encodes a peroxisomal transporter of coenzyme A, FAD and NAD(+). *Biochemical Journal* 443:241-247.

Altschul, S.F., T.L.Madden, A.A.Schaffer, J.H.Zhang, Z.Zhang, W.Miller, and D.J.Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.

Authier, F., J.J.M.Bergeron, W.J.Ou, R.A.Rachubinski, B.I.Posner, and P.A.Walton. 1995. Degradation of the Cleaved Leader Peptide of Thiolase by A Peroxisomal Proteinase. *Proceedings of the National Academy of Sciences of the United States of America* 92:3859-3863.

Bapteste, E., H.Brinkmann, J.A.Lee, D.V.Moore, C.W.Sensen, P.Gordon, L.Durufle, T.Gaasterland, P.Lopez, M.Muller, and H.Philippe. 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences of the United States of America* 99:1414-1419.

Berriman, M., E.Ghedin, C.Hertz-Fowler, G.Blandin, H.Renauld, D.C.Bartholomeu, N.J.Lennard, E.Caler, N.E.Hamlin, B.Haas, W.Bohme, L.Hannick, M.A.Aslett, J.Shallom, L.Marcello, L.H.Hou, B.Wickstead, U.C.M.Alsmark, C.Arrowsmith, R.J.Atkin, A.J.Barron, F.Bringaud, K.Brooks, M.Carrington, I.Cherevach, T.J.Chillingworth, C.Churcher, L.N.Clark, C.H.Corton, A.Cronin, R.M.Davies, J.Doggett, A.Djikeng, T.Feldblyum, M.C.Field, A.Fraser, I.Goodhead, Z.Hance, D.Harper, B.R.Harris, H.Hauser, J.Hostetter, A.Ivens, K.Jagels, D.Johnson, J.Johnson, K.Jones, A.X.Kerhornou, H.Koo, N.Larke, S.Landfear, C.Larkin, V.Leech, A.Line, A.Lord, A.MacLeod, P.J.Mooney, S.Moule, D.M.A.Martin, G.W.Morgan, K.Mungall, H.Norbertczak, D.Ormond, G.Pai, C.S.Peacock, J.Peterson,

M.A.Quail, E.Rabbinowitsch, M.A.Rajandream, C.Reitter, S.L.Salzberg, M.Sanders, S.Schobel, S.Sharp, M.Simmonds, A.J.Simpson, L.Talton, C.M.R.Turner, A.Tait, A.R.Tivey, S.Van Aken, D.Walker, D.Wanless, S.L.Wang, B.White, O.White, S.Whitehead, J.Woodward, J.Wortman, M.D.Adams, T.M.Embley, K.Gull, E.Ullu, J.D.Barry, A.H.Fairlamb, F.Opperdoes, B.G.Barret, J.E.Donelson, N.Hall, C.M.Fraser, S.E.Melville, and N.M.El-Sayed. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science* 309:416-422.

Berthold, J., M.F.Bauer, H.C.Schneider, C.Klaus, K.Dietmeier, W.Neupert, and M.Brunner. 1995. The Mim Complex Mediates Preprotein Translocation Across the Mitochondrial Inner Membrane and Couples It to the Mt-Hsp70/Atp Driving System. *Cell* 81:1085-1093.

Bonnefoy, N., F.Chalvet, P.Hamel, P.P.Slonimski, and G.Dujardin. 1994. Oxa1, A Saccharomyces-Cerevisiae Nuclear Gene Whose Sequence Is Conserved from Prokaryotes to Eukaryotes Controls Cytochrome-Oxidase Biogenesis. *Journal of Molecular Biology* 239:201-212.

Cavalier-Smith, T. 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biology Letters* 6:342-345.

Dagley, M.J., P.Dolezal, V.A.Likic, O.Smid, A.W.Purcell, S.K.Buchanan, J.Tachezy, and T.Lithgow. 2009. The Protein Import Channel in the Outer Mitosomal Membrane of Giardia intestinalis. *Molecular Biology and Evolution* 26:1941-1947.

Deduve, C. 1982. Peroxisomes and Related Particles in Historical-Perspective. *Annals of the New York Academy of Sciences* 386:1-4.

Dolezal, P., M.J.Dagley, M.Kono, P.Wolynec, V.A.Likic, J.H.Foo, M.Sedinova, J.Tachezy, A.Bachmann, I.Bruchhaus, and T.Lithgow. 2010. The Essentials of Protein Import in the Degenerate Mitochondrion of Entamoeba histolytica. *Plos Pathogens* 6.

Dolezal, P., V.Likic, J.Tachezy, and T.Lithgow. 2006. Evolution of the molecular machines for protein import into mitochondria. *Science* 313:314-318.

Dolezal, P., O.Smid, P.Rada, Z.Zubacova, D.Bursac, R.Sutak, J.Nebesarova, T.Lithgow, and J.Tachezy. 2005. Giardia mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. *Proceedings of the National Academy of Sciences of the United States of America* 102:10924-10929.

Fawcett, DW. 1981. The Cell. *W. B. Saunders Co. (Philadelphia)*

Finn, R.D., J.Clements, and S.R.Eddy. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39:W29-W37.

Gabaldon, T. 2010. Peroxisome diversity and evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* 365:765-773.

Gabaldon, T., B.Snel, F.van Zimmeren, W.Hemrika, H.Tabak, and M.A.Huynen. 2006. Origin and evolution of the peroxisomal proteome. *Biology Direct* 1.

Gatto, G.J., B.V.Geisbrecht, S.J.Gould, and J.M.Berg. 2000. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nature Structural Biology* 7:1091-1095.

Gill, E.E., S.Diaz-Trivino, M.J.Barbera, J.D.Silberman, A.Stechmann, D.Gaston, I.Tamas, and A.J.Roger. 2007. Novel mitochondrion-related organelles in the anaerobic amoeba Mastigamoeba balamuthi. *Molecular Microbiology* 66:1306-1320.

Gonzalez, N.H., G.Felsner, F.D.Schramm, A.Klingl, U.G.Maier, and K.Bolte. 2011. A Single Peroxisomal Targeting Signal Mediates Matrix Protein Import in Diatoms. *Plos One* 6.

Gould, S.J., G.A.Keller, M.Schneider, S.H.Howell, L.J.Garrard, J.M.Goodman, B.Distel, H.Tabak, and S.Subramani. 1990. Peroxisomal Protein Import Is Conserved Between Yeast, Plants, Insects and Mammals. *Embo Journal* 9:85-90.

Hettema, E.H., C.W.T.vanRoermund, B.Distel, M.vandenBerg, C.Vilela, C.RodriguesPousada, R.J.A.Wanders, and H.F.Tabak. 1996. The ABC transporter proteins Pat1 and Pat2 are required for import of long-chain fatty acids into peroxisomes of Saccharomyces cerevisiae. *Embo Journal* 15:3813-3822.

Hill, K., K.Model, M.T.Ryan, K.Dietmeier, F.Martin, R.Wagner, and N.Pfanner. 1998. Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins [see comment]. *Nature* 395:516-521.

Hoepfner, D., D.Schildknegt, I.Braakman, P.Philippsen, and H.F.Tabak. 2005. Contribution of the endoplasmic reticulum to peroxisome formation. *Cell* 122:85-95.

Jedelsky, P.L., P.Dolezal, P.Rada, J.Pyrih, O.Smid, I.Hrdy, M.Sedinova, M.Marcincikova, L.Voleman, A.J.Perry, N.C.Beltran, T.Lithgow, and J.Tachezy. 2011. The Minimal Proteome in the Reduced

Mitochondrion of the Parasitic Protist Giardia intestinalis. *Plos One* 6.

Koehler, C.M., S.Merchant, W.Oppliger, K.Schmid, E.Jarosch, L.Dolfini, T.Junne, G.Schatz, and K.Tokatlidis. 1998. Tim9p, an essential partner subunit of Tim10p for the import of mitochondrial carrier proteins. *Embo Journal* 17:6477-6486.

Kozjak, V., N.Wiedemann, D.Milenkovic, C.Lohaus, H.E.Meyer, B.Guiard, C.Meisinger, and N.Pfanner. 2003. An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *J. Biol. Chem.* 278:48520-48523.

Lindmark, D.G., M.Muller, and H.Shio. 1975. Hydrogenosomes in Trichomonas-Vaginalis. *Journal of Parasitology* 61:552-554.

Lithgow, T., and A.Schneider. 2010. Evolution of macromolecular import pathways in mitochondria, hydrogenosomes and mitosomes. *Philosophical Transactions of the Royal Society B-Biological Sciences* 365:799-817.

Loftus, B., I.Anderson, R.Davies, U.C.M.Alsmark, J.Samuelson, P.Amedeo, P.Roncaglia, M.Berriman, R.P.Hirt, B.J.Mann, T.Nozaki, B.Suh, M.Pop, M.Duchene, J.Ackers, E.Tannich, M.Leippe, M.Hofer, I.Bruchhaus, U.Willhoeft, A.Bhattacharya, T.Chillingworth, C.Churcher, Z.Hance, B.Harris, D.Harris, K.Jagels, S.Moule, K.Mungall, D.Ormond, R.Squares, S.Whitehead, M.A.Quail, E.Rabbinowitsch, H.Norbertczak, C.Price, Z.Wang, N.Guillen, C.Gilchrist, S.E.Stroup, S.Bhattacharya, A.Lohia, P.G.Foster, T.Sicheritz-Ponten, C.Weber, U.Singh, C.Mukherjee, N.M.El-Sayed, W.A.Petri, C.G.Clark, T.M.Embley, B.Barrell, C.M.Fraser, and N.Hall. 2005. The genome of the protist parasite Entamoeba histolytica. *Nature* 433:865-868.

Marti-Renom, M.A., A.C.Stuart, A.Fiser, R.Sanchez, F.Melo, and A.Sali. 2000. Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure* 29:291-325.

Mazzuco, A., M.Benchimol, and W.DeSouza. 1997. Endoplasmic reticulum and Golgi-like elements in Entamoeba. *Micron* 28:241-247.

McLennan, A.G. 2006. The Nudix hydrolase superfamily. *Cellular and Molecular Life Sciences* 63:123-143.

Meinecke, M., C.Cizmowski, W.Schliebs, V.Kruger, S.Beck, R.Wagner, and R.Erdmann. 2010. The peroxisomal importomer constitutes a large and highly dynamic pore. *Nature Cell Biology* 12:273-U68.

Mesecke, N., N.Terziyska, C.Kozany, F.Baumann, W.Neupert, K.Hell, and J.M.Herrmann. 2005. A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. *Cell* 121:1059-1069.

Mi-ichi, F., M.Abu Yousuf, K.Nakada-Tsukui, and T.Nozaki. 2009. Mitosomes in Entamoeba histolytica contain a sulfate activation pathway. *Proceedings of the National Academy of Sciences of the United States of America* 106:21731-21736.

Morrison, H.G., A.G.McArthur, F.D.Gillin, S.B.Aley, R.D.Adam, G.J.Olsen, A.A.Best, W.Z.Cande, F.Chen, M.J.Cipriano, B.J.Davids, S.C.Dawson, H.G.Elmendorf, A.B.Hehl, M.E.Holder, S.M.Huse, U.U.Kim, E.Lasek-Nesselquist, G.Manning, A.Nigam, J.E.J.Nixon, D.Palm, N.E.Passamaneck, A.Prabhu, C.I.Reich, D.S.Reiner, J.Samuelson, S.G.Svard, and M.L.Sogin. 2007. Genomic minimalism

in the early diverging intestinal parasite Giardia lamblia. *Science* 317:1921-1926.

Motley, A.M., E.H.Hettema, R.Ketting, R.Plasterk, and H.F.Tabak. 2000. Caenorhabditis elegans has a single pathway to target matrix proteins to peroxisomes. *Embo Reports* 1:40-46.

Neufeld, C., F.V.Filipp, B.Simon, A.Neuhaus, N.Schuller, C.David, H.Kooshapur, T.Madl, R.Erdmann, W.Schliebs, M.Wilmanns, and M.Sattler. 2009. Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *Embo Journal* 28:745-754.

Purdue, P.E., and P.B.Lazarow. 1994. Peroxisomal Biogenesis - Multiple Pathways of Protein Import. *Journal of Biological Chemistry* 269:30065-30068.

Pusnik, M., F.Charriere, P.Maser, R.F.Waller, M.J.Dagley, T.Lithgow, and A.Schneider. 2009. The Single Mitochondrial Porin of Trypanosoma brucei is the Main Metabolite Transporter in the Outer Mitochondrial Membrane. *Molecular Biology and Evolution* 26:671-680.

Pusnik, M., O.Schmidt, A.J.Perry, S.Oeljeklaus, M.Niemann, B.Warscheid, T.Lithgow, C.Meisinger, and A.Schneider. 2011. Mitochondrial Preprotein Translocase of Trypanosomatids Has a Bacterial Origin. *Current Biology* 21:1738-1743.

Rigden, D.J., P.A.Michels, and M.L.Ginger. 2009. Autophagy in protists Examples of secondary loss, lineage-specific innovations, and the conundrum of remodeling a single mitochondrion. *Autophagy* 5:784-794.

Rosenbau, R.M., and M.Wittner. 1970. Ultrastructure of Bacterized and Axenic Trophozoites of Entamoeba-Histolytica with Particular Reference to Helical Bodies. *Journal of Cell Biology* 45:367-&.

Rucktaschel, R., W.Girzalsky, and R.Erdmann. 2011. Protein import machineries of peroxisomes. *Biochimica et Biophysica Acta-Biomembranes* 1808:892-900.

Sacksteder, K.A., J.M.Jones, S.T.South, X.L.Li, Y.F.Liu, and S.J.Gould. 2000. PEX19 binds multiple peroxisomal membrane proteins, is predominantly cytoplasmic, and is required for peroxisome membrane synthesis. *Journal of Cell Biology* 148:931-944.

Sagan, L. 1967. On Origin of Mitosing Cells. *Journal of Theoretical Biology* 14:225-&.

Schliebs, W., W.Girzalsky, and R.Erdmann. 2010. Peroxisomal protein import and ERAD: variations on a common theme. *Nature Reviews Molecular Cell Biology* 11:885-890.

Schnaufer, A., G.D.Clark-Walker, A.G.Steinberg, and K.Stuart. 2005. The F-1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. *Embo Journal* 24:4029-4040.

Schneider, A., D.Bursac, and T.Lithgow. 2008. The direct route: a simplified pathway for protein import into the mitochondrion of trypanosomes. *Trends in Cell Biology* 18:12-18.

Soding, J., A.Biegert, and A.N.Lupas. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* 33:W244-W248.

Stechmann, A., K.Hamblin, V.Perez-Brocal, D.Gaston, G.S.Richmond, M.van der Giezen, C.G.Clark, and A.J.Roger. 2008. Organelles in Blastocystis that blur the distinction between mitochondria and hydrogenosomes. *Current Biology* 18:580-585.

Sun, C.H., C.F.Chou, and J.H.Tai. 1998. Stable DNA transfection of the primitive protozoan pathogen

Giardia lamblia. *Molecular and Biochemical Parasitology* 92:123-132.

Teixeira, J.E., and C.D.Huston. 2008. Evidence of a continuous endoplasmic reticulum in the protozoan parasite Entamoeba histolytica. *Eukaryotic Cell* 7:1222-1226.

Thoms, S., and R.Erdmann. 2006. Peroxisomal matrix protein receptor ubiquitination and recycling. *Biochimica et Biophysica Acta-Molecular Cell Research* 1763:1620-1628.

Tovar, J., G.Leon-Avila, L.B.Sanchez, R.Sutak, J.Tachezy, M.van der Giezen, M.Hernandez, M.Muller, and J.M.Lucocq. 2003. Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. *Nature* 426:172-176.

Ujwal, R., D.Cascio, J.P.Colletier, S.Faham, J.Zhang, L.Toro, P.P.Ping, and J.Abramson. 2008. The crystal structure of mouse VDAC1 at 2.3 angstrom resolution reveals mechanistic insights into metabolite gating. *Proceedings of the National Academy of Sciences of the United States of America* 105:17742-17747.

van der Giezen, M., and J.Tovar. 2005. Degenerate mitochondria. *Embo Reports* 6:525-530.

van der Zand, A., I.Braakman, and H.F.Tabak. 2010. Peroxisomal Membrane Proteins Insert into the Endoplasmic Reticulum. *Molecular Biology of the Cell* 21:2057-2065.

van Hoek, A.H.A.M., A.S.Akhmanova, M.A.Huynen, and J.H.P.Hackstein. 2000. A mitochondrial ancestry of the hydrogenosomes of Nyctotherus ovalis. *Molecular Biology and Evolution* 17:202-206.

Voos, W., O.vonAhsen, H.Muller, B.Guiard, J.Rassow, and N.Pfanner. 1996. Differential requirement

for the mitochondrial Hsp70-Tim44 complex in unfolding and translocation of preproteins. *Embo Journal* 15:2668-2677.

Zarsky, V., J.Tachezy, and P.Dolezal. 2012. Tom40 is likely common to all mitochondria. *Current Biology* 22:R479-R481.

Zeth, K. 2010. Structure and evolution of mitochondrial outer membrane proteins of beta-barrel topology. *Biochimica et Biophysica Acta-Bioenergetics* 1797:1292-1299.

Zeth, K., and M.Thein. 2010. Porins in prokaryotes and eukaryotes: common themes and variations. *Biochemical Journal* 431:13-22.

8 Supplementary information

replicated the original experiment: SCRs to the CS+ were significantly greater than for the CS– ($p < 0.05$; Figure S2), and correlated negatively with state anxiety. On day 2, however, SCRs to the CS+ and CS– no longer differed. Unlike conscious fear learning, which is known to persist over time [5], fear acquired nonconsciously is thus subject to rapid forgetting.

Previous attempts to investigate nonconscious conditioning (for example, [8]) used backward masking to suppress briefly-presented stimuli from awareness. However, the methodological limitations of masking (see Supplemental Information), as well as insufficiently rigorous measures of awareness used in past studies [3], have left the question of whether a new fear association can be learned nonconsciously unresolved. Here we used CFS to suppress long-duration CSs from awareness reliably (as assessed by both objective and subjective measures), and found that although the overall magnitude of nonconscious fear learning is comparable to conscious learning, it is characterized by a distinct temporal pattern. Conscious fear developed progressively over time, whereas nonconscious fear was acquired rapidly and declined swiftly.

The mechanisms underlying conscious and nonconscious fear conditioning may thus fulfill complementary roles: The initial orienting response that allows a stimulus to be associated with threat may not require awareness, but the long-term retention and expression of such learning does. Both conscious and nonconscious conditioning likely involve the amygdala, a brain region critical for the acquisition and expression of fear [9]. The amygdala plays a role in the automatic detection and processing of subliminally-presented affective stimuli [4], but has a tendency to rapidly habituate, especially to emotionally-laden stimuli [10]. Such habituation may, in turn, prevent the formation of a stable fear association, which might lead to rapid forgetting in the absence of other processes that involve awareness. The neural mechanisms that distinguish learning with and without awareness are thus fertile ground for further investigation.

Supplemental Information

Supplemental Information includes two figures and supplemental experimental procedures and can be found with this article online at doi:10.1016/j.cub.2012.04.023.

Acknowledgments

This research was supported by an International Brain Research Foundation Postdoctoral Fellowship to DC, and NIH Grants RO1 MH062104 to E.A.P. and RO1 EY016200 to M.C. The authors thank Daniela Schiller for helpful discussions.

References

1. Ohman, R.J., and Soares, J.J.F. (1994). 'Unconscious anxiety': Phobic responses to masked stimuli. J. Abnorm. Psychol. *103*, 231–240.
2. Morris, J.S., Ohman, A., and Dolan, R.J. (1998). Conscious and unconscious emotional learning in the human amygdala. Nature *393*, 467–470.
3. Lovibond, P.F., and Shanks, D.R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. J. Exp. Psychol. Anim. Behav. Processes *28,* 3–26.
4. Tamietto, M., and de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. Nat. Rev. Neurosci. *11*, 697–709.
5. Schiller, D., Monfils, M., Raio C.M., Johnson D., LeDoux, J.E., and Phelps, E.A. (2010). Blocking the return of fear in humans using reconsolidation update mechanisms. Nature *463*, 49–53.
6. Lissek, S., Powers, A.S., McClure, E.B., Phelps, E.A., Woldehawariat, G., Grillon, C., and Pine, D.S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. Behav. Res. Therapy *43*, 1391–1424.
7. Bishop, S.J., Duncan, J., and Lawrence, A.D. (2004). State anxiety modulation of the amygdala response to unattended threat-related stimuli. J. Neurosci. *24*, 10364–10368.
8. Esteves, F., Parra, C., Dimberg, U., and Ohman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. Psychophysiology *31*, 375–385.
9. Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., and Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J. Neurosci. *18*, 411–418.
10. Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A., Rauch, S.L., Buckner, R.L., Strauss, M.M., Hyman, S.E., and Rosen, B.R. (1996). Response and habituation the human amygdala during visual processing of facial expression. Neuron *17*, 875–887.

[1]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA. [2]Center for Neural Science, New York University, 6 Washington Place, New York, NY 10003, USA. [3]Present address: Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.
*These authors contributed equally to the work.
E-mail: davecarmel@gmail.com

Tom40 is likely common to all mitochondria

Vojtech Zarsky, Jan Tachezy, and Pavel Dolezal*

The evolution of the mitochondrion has involved the remodelling of the two membranes that enclose this organelle. During the transformation of the endosymbiotic bacterium into a genetically dependent organelle, the flow of proteins across the membranes reversed. This change is reflected by the distinct sets of protein transport machinery that operate in bacterial and mitochondrial membranes [1]. One of the exceptions is a β-barrel assembly machine, Sam50, a member of the Omp85 superfamily of proteins, which has been retained in the mitochondrial membranes. Other core components of mitochondrial translocases, such as Tom40 in the outer membrane and the Tim17 family of proteins in the inner membrane, cannot be directly related to any bacterial proteins.
Two studies by Pusnik *et al.* recently showed that the mitochondrion of *Trypanosoma brucei* was found to be devoid of the essential Tom40 channel [2]; instead, it was found to contain an essential protein called the archaic translocase *of the* outer mitochondrial membrane (ATOM) that was directly linked to bacterial YtfM proteins, which are members of the Omp85 superfamily [3]. Thus, it was suggested by Pusnik *et al.* that ATOM and Tom40 represent mutually exclusive functional analogues of distinct origins [3]. We analysed the ATOM amino acid sequences to identify homology to known protein families and to determine the phylogenetic distribution of the closest relatives of ATOM. Surprisingly, our results clearly refute the link between ATOM and bacterial Omp85-like proteins. Moreover, we propose that ATOM is, in fact, a divergent form of the 'classical' Tom40.

Tom40 and members of the Omp85 superfamily are β-barrel transmembrane proteins [4]. They form the rigid channels in the outer membranes of bacteria, plastids and mitochondria, where they guide substrates across or into the membrane. The pore-forming β-barrel structure does not require a precise

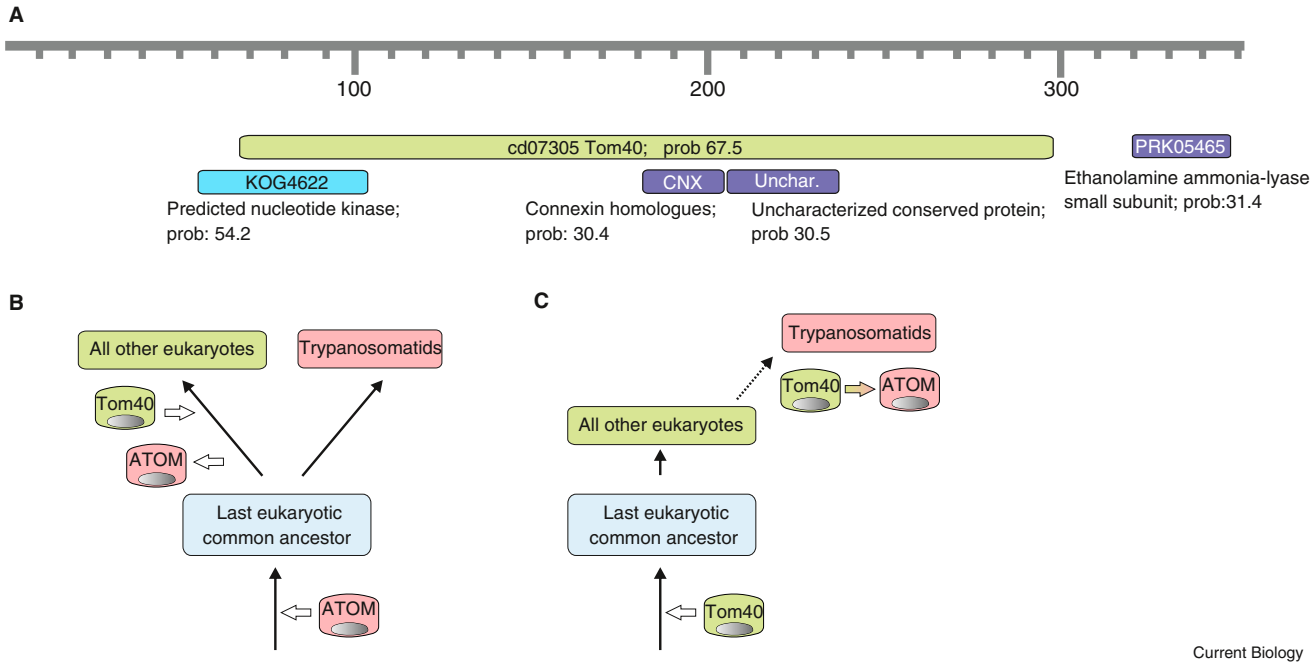


Figure 1. The ATOM is divergent Tom40.
(A) A graphical output of the HHpred search of the conserved domains database (cd_01Nov11) using the *T. brucei* ATOM sequence as the query with default parameters set. The best hit was a Tom40-specific entry (cd07305). The length of the homologous stretch indicates the conservation of the overall structure. No domains characteristic of Omp85 or Omp85-like families, which would indicate similarity between ATOM and Ytfm, were found. The full output in text format can be found in the Supplemental Information. (B) The modelled evolution of the mitochondrial outer membrane protein translocase as proposed by Pusnik *et al.* [3]. The ATOM translocase was derived from a bacterial Ytfm Omp85-like protein, and it was present in the last ancestor common to all eukaryotes. ATOM has been retained by the early branching trypanosomatids, but it was replaced by Tom40 in the lineage leading to all other eukaryotes. (C) Our model for the evolution of the mitochondrial outer membrane protein translocase proposes that an ancestral Tom40 was present in the mitochondria of the last common eukaryotic ancestor and that the ATOM proteins of trypanosomatids represent divergent Tom40 homologues.

composition of amino acid residues, meaning that the bioinformatic analyses of proteins with large evolutionary divergences cannot rely on pairwise sequence algorithms such as BLAST. In these cases, hidden Markov model (HMM)-based sequence analyses have proven to be more sensitive and specific and have led to the identification of Tom40 homologues even in the anaerobic unicellular eukaryotes that were previously considered to be ancestral amitochondriate organisms [5–7].

We used HHpred to search the alignment-based databases of conserved domains, such as the CDD, PFAM and SMART databases, using the *T. brucei* ATOM sequence as query [8]. For all of the conditions tested, we were unable to demonstrate any relationship between ATOM and the YtfM proteins in these domain databases. Moreover, the Tom40 protein family was consistently found as the best hit for the ATOM query (Figure 1).

To find the ATOM homologues among other trypanosomatids, we searched the genomic data available at TriTrypDB (http://tritrypdb.org/tritrypdb/). Using BLAST searches, close homologues of ATOM were identified in all *Trypanosoma* and *Leishmania* species. We further found an ATOM homologue in the genome of *Endotrypanum monterogeii*, an organism closely related to the *Leishmania* clade, a parasite of the sloth. When searching the Pfam database, the *Leishmania* and *Endotrypanum* sequences were recognized to contain a Porin_3 domain representing eukaryotic Tom40 and VDAC sequences with significant e-value support between 0.0015 and 0.2 (Figure S1 in Supplemental Information). No connections between the ATOM and YtfM proteins were found using this method.

The protein sequence alignment of nine available ATOM sequences revealed the presence of a conserved motif in the last β-strand. This motif functions as a sorting signal for mitochondrial β-barrel proteins when taken up by the SAM complex. A different signal was described for bacterial β-barrel proteins such as YtfM [4] (Figure S1).

The Omp85 superfamily has two signature domains — the carboxy-terminal β-barrel domain and the amino-terminal POTRA domain(s) — that participate in the assembly of the substrate precursor proteins. If ATOM were related to YtfM, the presence of residual POTRA domain(s) in ATOM would provide some support for its relationship to the Omp85 superfamily. Such support is found in plastids, where the outer membrane translocase Toc75 of the TOC translocon retained its POTRA domains [4]. However, neither we nor Pusnik *et al.* [3] were able to identify significant similarity between ATOM and POTRA domain sequences.

The absence of Tom40 in trypanosomes has been considered a primitive trait, i.e., suggesting that trypanosomes diverged from the eukaryotic tree of life before Tom40 arose. The presence of ATOM would provide additional support for the primitive character of kinetoplastid mitochondria and, as such, would

place the root of the eukaryotic tree within the superior group of Euglenozoa or between Euglenozoa and other eukaryotes [9].

Our analyses show that ATOM represents the missing Tom40 protein in the mitochondria of *T. brucei* and of other trypanosomatids with no clear link to the bacterial proteins. Given that all eukaryotes analysed to date contain a Tom40 homologue, we propose that all mitochondria of current eukaryotes descended from an ancestral Tom40-containing mitochondrial compartment (Figure 1).

Supplemental Information

Supplemental Information includes one figure and can be found with this article online at doi: 10.1016/j.cub.2012.03.057.

Acknowledgements

This work was supported by grants from the Czech Science Foundation (P305-10-0651 (P.D.) and LC07032 (J.T.)) and by a grant from the Ministry of Education of the Czech Republic (MSM 0021620858 (J.T.)).

References

1. Hewitt, V., Alcock, F., and Lithgow, T. (2011). Minor modifications and major adaptations: the evolution of molecular machines driving mitochondrial protein import. Biochim. Biophys. Acta *1808,* 947–954.
2. Pusnik, M., Charriebre, F., Mäser, P., Waller, R.F., Dagley, M.J., Lithgow, T., and Schneider, A. (2009). The single mitochondrial porin of Trypanosoma brucei is the main metabolite transporter in the outer mitochondrial membrane. Mol. Biol. Evol. *26,* 671–680.
3. Pusnik, M., Schmidt, O., Perry, A.J., Oeljeklaus, S., Niemann, M., Warscheid, B., Lithgow, T., Meisinger, C., and Schneider, A. (2011). Mitochondrial preprotein translocase of trypanosomatids has a bacterial origin. Curr. Biol. *21,* 1738–1743.
4. Zeth, K. (2010). Structure and evolution of mitochondrial outer membrane proteins of beta-barrel topology. Biochim. Biophys. Acta *1797,* 1292–1299.
5. Dagley, M.J., Dolezal, P., Likic, V.A., Smid, O., Purcell, A.W., Buchanan, S.K., Tachezy, J., and Lithgow, T. (2009). The protein import channel in the outer mitosomal membrane of Giardia intestinalis. Mol. Biol. Evol. *26,* 1941–1947.
6. Dolezal, P., Dagley, M.J., Kono, M., Wolynec, P., Likicb, V.A., Foo, J.H., Sedinovab, M., Tachezy, J., Bachmann, A., Bruchhaus, I., *et al.* (2010). The essentials of protein import in the degenerate mitochondrion of Entamoeba histolytica. PLoS Pathog. *6,* e1000812.
7. Likic, V.A., Dolezal, P., Celik, N., Dagley, M., and Lithgow, T. (2010). Using hidden markov models to discover new protein transport machines. Meth. Mol. Biol. *619,* 271–284.
8. Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33,* W244–W248.
9. Cavalier-Smith, T. (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. Biol. Lett. *6,* 342–345.

Charles University in Prague, Faculty of Science, Department of Parasitology, Vinicna 7, 128 44, Prague 2, Czech Republic.
*E-mail: paveldolezal@yahoo.com

Response to Zarsky *et al.*

Mascha Pusnik[1], Oliver Schmidt[2], Andrew J. Perry[3], Silke Oeljeklaus[4], Moritz Niemann[1], Bettina Warscheid[4], Chris Meisinger[2], Trevor Lithgow[3], and André Schneider[1,*]

Mitochondria evolved from an α-proteobacterial endosymbiont and recent phylogenetic and function-based research has demonstrated that the major pieces of the protein transport machinery were inherited from the symbiont. This includes the SAM machinery for assembly of outer membrane proteins and the TIM machinery for protein transport across, and assembly into, the mitochondrial inner membrane [1–3]. Hidden Markov model (HMM) analysis, which enables a broad, all-encompassing approach for identifying protein homologies, has been very important in detecting members of protein families that are not easily recognized by simple BLAST-based comparisons [1]; HMM searches initially failed to find a Tom40 protein in one group of eukaryotes, the kinetoplastids. These organisms, which include the experimentally-tractable *Trypanosoma brucei*, have highly developed mitochondria that have evolved from the same ancestor as mitochondria in other eukaryotes. The initial failure to identify a Tom40 homolog in *T. brucei* was both surprising and exciting.

In our paper in *Current Biology* [4] we directly assayed for protein transport function and thereby discovered the archaic protein translocase in the outer mitochondrial membrane (ATOM). In seeking related protein sequences, using E < 0.005 the PSI-BLAST search identifies only the kinetoplastid ATOM sequences. But, at a lower significance, a sub-class of Omp85 protein sequences, referred to as the YtfM/TamA group (but not Tom40 sequences) are found and the top-scoring one was manually added into the first-round outcome from the PSI-BLAST. Multiple sequence alignments using the ATOM from *T. brucei* and related organisms suggested, albeit not at statistically significant levels, an affinity to a sub-class of Omp85 proteins referred to as the YtfM group, and the putative relationship between trypanosomatid ATOMs and YtfMs was further visualized using CLANS [4]. Since YtfM is found in the α-proteobacteria, from which mitochondria evolved, one prospect was that the ATOM evolved from the YtfM in the endosymbiont's outer membrane. A second model for the evolution of the ATOM allowed for the possibility of a lateral gene transfer (LGT) early in the evolution of mitochondrial protein transport. We also raised a third model that holds Tom40 and ATOM evolved from a common ancestor. These models proposed in the original paper [4], are summarized in Figure 1. We remain open-minded on which model best explains the evolution of the pathway for protein translocation across the outer mitochondrial membrane.

In their correspondence, Zarsky *et al.* [5] argue that the ATOM is not related to YtfM-type Omp85 proteins, but is exclusively similar to the Tom40 family of proteins and that the ATOM evolved from a Tom40 progenitor. This is an attractive idea in the sense that it would be a unifying theory, with the implication being that all eukaryotes simply have a Tom40 translocase in their outer mitochondrial membrane, with some more easily recognized than others. However, two important observations need also be kept in mind.

Firstly, using HMMs based on the broad diversity of Tom40 sequences, ATOM was not initially detected in *T. brucei* [6]. This gives a context to just how divergent the ATOM and other Tom40 proteins are, given that this same type of HMM approach has succeeded in finding highly diverse Tom40 sequences in *Entamoeba* [7] and *Giardia* [8]. By broadening the search criteria with a goal to capture all members of the mitochondrial porin protein family (i.e. isoforms of Tom40 and VDAC), Flinner *et al*. [9] recently showed that *T. brucei* has two further prospective mitochondrial porins that might play a role in ion transport: their analysis did not detect ATOM.

Secondly, the ATOM protein sequence has predicted secondary structural features that seem to be consistent with a POTRA-type amino-terminal domain (data not shown) and a predicted β-barrel domain of comparable size to other members of the YtfM/TamA-family of proteins. POTRA domains are not found in Tom40 (or other mitochondrial porins), which have instead a simple amino-terminal helix [10]. With the size and characteristics of the β-barrel