A Bioinformatical Approach to Study the Endosomal Sorting Complex Required for Transport (ESCRT) Machinery in Protozoan Parasites: The *Entamoeba histolytica* Case

Israel López-Reyes[1], Cecilia Bañuelos[1],
Abigail Betanzos[2] and Esther Orozco[2,3]
[1]Instituto de Ciencia y Tecnología del Distrito Federal,
[2]Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional,
[3]Universidad Autónoma de la Ciudad de México,
México

1. Introduction

1.1 The potential of bioinformatics for the study of protein structure and function

Proteins are macromolecules formed by amino acid polymers that regulate cellular functions. Each protein is composed by the repetition and combination of 20 different amino acids, whose order is determined by the genetic code. To perform their biological functions, proteins fold into one or more specific spatial conformations, determined by non-covalent interactions such as hydrogen bonding, ionic interactions, Van der Waals forces and hydrophobic packing, and covalent interactions, such as disulfide bonds (Chiang et al., 2007).

Determining the structure and function of a protein is a milestone of many aspects of modern biology to understand its role in cell physiology. Bioinformatics is the research, development or application of computational approaches for expanding the use of biological, medical, behavioral or health-related data. It also includes those tools to acquire, store, organize, archive, analyze or visualize infomation. Over the past years, bioinformatical tools have been widely used for the prediction and study of protein biology. Moreover, bioinformatical tools have revealed the existence of protein "interactomes", demonstrating the interaction among distinct biomolecules (protein-protein, protein-lipids, protein-carbohydrates, etc.) to perform cellular processes (Kuchaiev & Przulj, 2011).

During the last decades, genome sequencing projects together with bioinformatics programs and algorithms have enormously contributed to understand protein structure, protein interactions and protein functions. At present, over six million unique protein sequences have been deposited in public databases, and this number is increasing rapidly. Meanwhile, despite the progress of high-throughput structural genomics initiatives, just over 50,000 protein structures have been experimentally determined (Kelley & Sterberg, 2009). The greatest challenge the molecular biology community is facing today is to analyze the wealth of data that has been produced by the genome sequencing projects, where bioinformatics

has been fundamental. Traditionally, molecular biology research has been carried out entirely at the laboratory bench, but the huge increase in the amount of data has made necessary to incorporate computers and sophisticated software into research.

Additionally, availability of genome databases for distinct organisms has improved our knowledge on the way to elucidate the last universal common ancestor. In conclusion, analyzing and comparing the genetic material of different species is an increasingly important approach for studying the numbers, locations, biochemical functions and evolution of genes and proteins.

In this review, we selected a particular scientific case to emphasize the usefulness and potential of bioinformatics in addressing a biological problem.

Most cellular processes use scaffold proteins to recruit other proteins and to facilitate their correct interaction and functioning. Thus, we focused on the very little studied scaffold proteins that form the Endosomal Sorting Complexes Required for Transport (ESCRT) machinery during protozoan endocytosis, a fundamental process for cell survival. Here, as a study case, we aimed to highlight the possible identity, function and interactions of ESCRT complexes in *Entamoeba histolytica,* as determined by the use of bioinformatical tools.

1.2 Role of the ESCRT in endocytosis

Endocytosis is a crucial process in multiple cellular and physiological events, including nutrient uptake, virus budding, cell surface receptor downregulation and cell signaling. It involves the internalization of molecules or particles of different sizes from the external environment, through membrane remodeling and vesicle formation events (de Souza et al., 2009). In endocytosis, a huge number of interactomes are involved. In the study of the highly complex endocytosis process, bioinformatics databases and computational tools have been of enormous value.

Several plasma membrane proteins interact with target molecules (cargo) to internalize and transport them along the endocytic pathway. Depending on their function, membrane proteins are recycled back to the cell surface or degraded at lysosomal compartments together with cargo. Delivery of endocytosed cargo for degradation occurs through the fusion of intracellular vesicles called early and late endosomes that finally reach lysosomes.

In the majority of cell types, late endosomes fuse among them to form multivesicular bodies (MVB), which are essential intermediates for nutrient, ligand and receptor trafficking (Williams & Urbé, 2007). The best characterized signal for entering cargo molecules into the degradative MVB pathway is ubiquitination. Ubiquitination is a conjugation event in which a highly conserved 76 amino acid protein called ubiquitin, is covalently attached for cargo labeling. Most of the cargo proteins that accumulate in MVB are marked by a single ubiquitin, which is recognized by a specific and conserved protein machinery termed "Endosomal Sorting Complex Required for Transport (ESCRT)" and whose function is fundamental during endocytosis (Williams & Urbé, 2007).

The ESCRT machinery was first characterized in yeast. It consists of a group of vacuolar protein sorting factors (some of them called Vps), which form different multimeric complexes (ESCRT-0, -I, -II and -III) that bind among them but also associate to accesory proteins and endosomal membrane lipids to perform the whole endocytic process (Fig. 1) (Hurley & Emr, 2006).
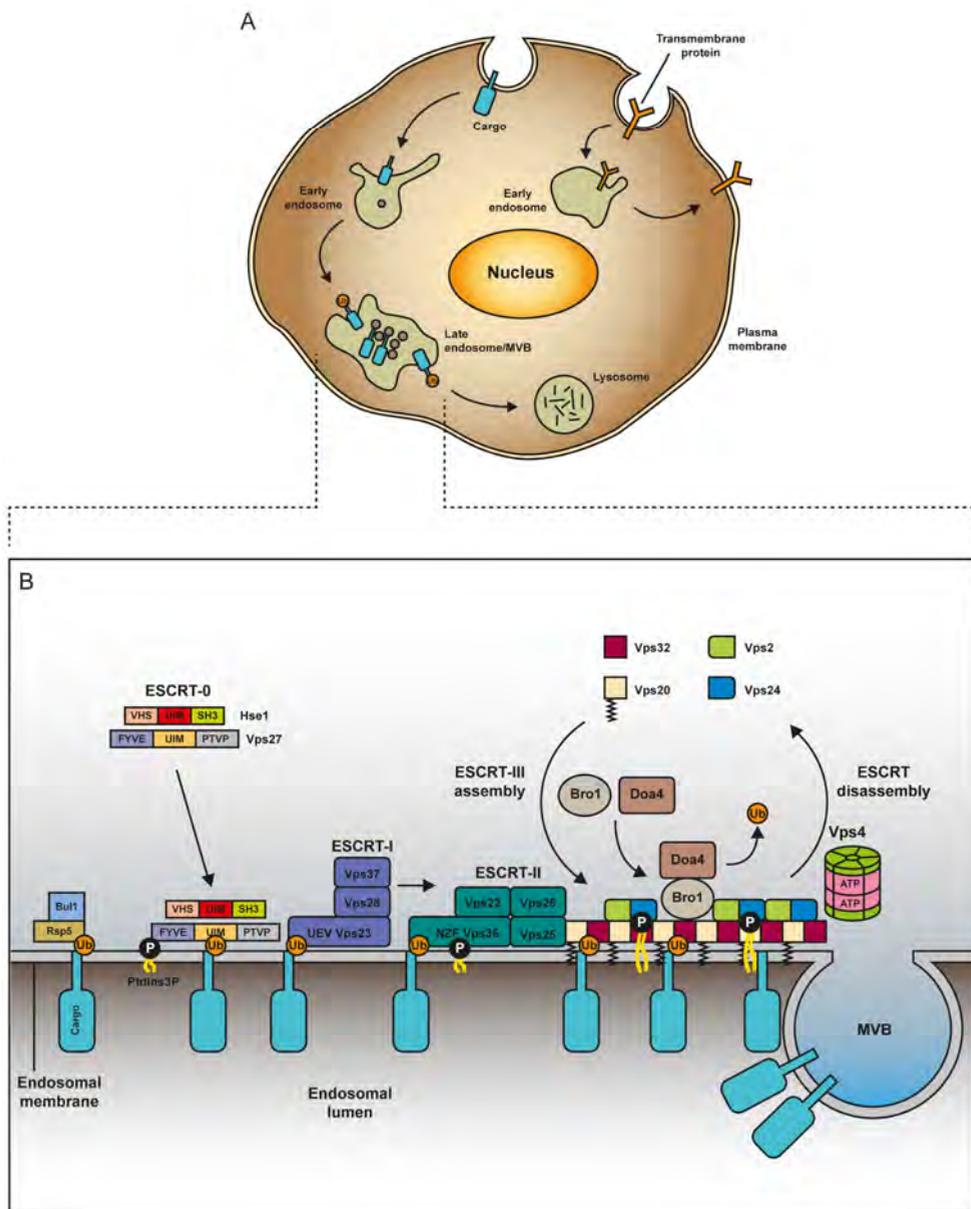
Fig. 1. The ESCRT machinery involved in the endosomal MVB pathway.

(A) Eukaryotic cells internalize cargo molecules from the external environment by endocytic processes. These molecules transit along several compartments for surface recycling or degradation. The degradation pathway involves an endomembrane system constituted of membrane bound organelles called endosomes that mature from early to late endosomes/MVB for finally cargo delivering into lysosomes. According to (B), at late endosome level, molecules to be incorporated into the degradation pathway should be tagged with ubiquitin. In yeast, ubiquitination of cargo proteins is mediated by the ubiquitin ligase Rsp5 and by Bul1. Then, the ESCRT-0 complex initiates the MVB sorting process by endosomal membrane binding through the Vps27 domain, and ubiquitin recognition of cargo by UIM domains present in Vps27 and Hse1 proteins. Subsequently, Vps27 activates the ESCRT-I complex through its interaction with Vps23. Ubiquitinated cargo is recognized by ESCRT-I (via the UEV domain of Vps23) and by ESCRT-II (via the NZF domain of Vps36). Vps36 has an extensive positively charged region with high affinity to phosphoinositides, allowing ESCRT-II attachment to endosomal membranes. Then, ESCRT-III concentrates cargo proteins into MVB. ESCRT-III also associates to accessory proteins such as Bro1 and Doa4. Importantly, the Vps4 ATPase catalyzes the dissociation of ESCRT complexes to initiate new cycles of cargo sorting and transport. Together, ESCRT-0 to -III and -accessory proteins direct cargo sorting, vesicle fusion, and MVB biogenesis (modified from Hurley & Emr, 2006).

Cargo ubiquitination is mediated by the Rsp5 ubiquitin ligase and Bul1 protein. Then, cargo sorting through the MVB pathway initiates with the association of Vps27 and Hse1 proteins to make up the ESCRT-0 complex. Vps27 has a FYVE (Fab1, YOTB, Vac1, EEA1) domain, which binds to membrane lipids, and an UIM (Ubiquitin Interaction Motif) domain that determines an important role for ESCRT-0 in the initial selection of ubiquitinated cargo at the endosomal membrane (Hurley & Emr, 2006; Williams & Urbé, 2007). Then, ESCRT-0 recruits ESCRT-I, formed by Vps23, Vps28, Vps37 and Mvb12 proteins (Curtiss et al., 2007; Katzmann et al., 2001). Vps23 also recognizes and binds ubiquitinated proteins through its terminal UEV (Ubiquitin E2 Variant) domain. ESCRT-I binds to ESCRT-II formed by Vps22, Vps25 and Vps36 proteins (Babst et al., 2002a). This later protein also displays an ubiquitin-interacting domain and a recognition region for phosphoinositides binding. Next, ESCRT-II binds to the ESCRT-III complex composed of Vps2, Vps20, Vps24 and Vps32 proteins (Babst et al., 2002b). Vps32 associates to Bro1, which recruits Doa4, an ubiquitin hydrolase that removes ubiquitin from cargo proteins prior to their incorporation into MVB (Kim et al., 2005; Odorizzi et al., 2003). One of the main functions of ESCRT-III is to concentrate the MVB cargo in the endosomal inward membrane, and to recruit Vps4, an ATPase that catalyzes the disassembly of ESCRT complexes from the endosomal membrane to initiate new rounds of cargo sorting and trafficking, and vesicle formation, and vesicle formation (Hurley & Emr, 2006; Hurley & Hanson, 2010; Williams & Urbé, 2007).

Accessory proteins such as Vta1 and Ist1, regulate Vps4 function (Dimaano et al., 2008; Shiflett et al., 2004), whereas Vps46 and Vps60 have also been suggested to bind ESCRT-III, although their precise functions have yet to be determined (Babst et al., 2002b).

1.3 Evolution of the ESCRT machinery

During the evolution from prokaryotic to eukaryotic organisms, some properties were lost while others were acquired. Among the latter is the ability of eukaryotic cells to incorporate macromolecules, complexes and other cells through endocytosis (de Souza et al., 2009). Comparative genomics and phylogenetic studies have determined that the basic features of intracellular trafficking systems arose very early in eukaryotic evolution (Dacks & Field,

2007). Similarly, evidence for the existence of MVB-like organelles in diverse primitive eukaryotes has also been reported (Allen et al., 2007; Tse et al., 2004; Yang et al., 2004).

Lysosomal targeting of ubiquitinated cargo by ESCRT complexes is conserved in animals and fungi (Leung et al., 2008). Extensive experimental and bioinformatical comparative analysis of genomic data indicate that ESCRT factors are well conserved across the eukaryotic lineage (Williams & Urbé, 2007). ESCRT-I, -II and –III as well as -accessory proteins are almost completely retained in all studied taxa, indicating an early evolutionary origin and a near-universal system for cargo trafficking through the MVB pathway. Particularly, all eukaryotic organisms studied to date have at least an ESCRT-III protein, suggesting that the minimal ESCRT necessary for MVB formation might be ESCRT-III (Williams & Urbé, 2007). In addition, the number of components of ESCRT-III is greatly expanded in mammals in comparison to yeast, being Vps46 the most frequent ESCRT-III multicopy gene product (Dacks et al., 2008).

A common ancestry within the same ESCRT complexes or among them, has been reported for Vps20, Vps32 and Vps60 proteins (sharing a Snf7 domain), and Vps2, Vps24 and Vps46 proteins (sharing a Vps24 domain). All these proteins are highly similar at sequence level and are encoded by multicopy genes, probably due to gene amplification events (Leung et al., 2008). In terms of biological conservation, it seems that several ESCRT components had to be expanded to provide functional redundancy. Thus, this redundancy would preserve ESCRT functions in the endocytic MVB pathway even if losses of components were presented along evolution.

Significantly, the Vps4 ATPase responsible for recycling ESCRT components, is present in all taxa, indicating a highly conserved mechanism for delivering energy in the system. This is consistent with recent evidence for an archael origin for Vps4 (Obita et al., 2007).

The most prominent evolutionary variation in the MVB pathway is the restriction of ESCRT-0 to animals and fungi, suggesting that a distinct mechanism for ubiquitin labeling, signal recognition and endosomal membrane binding likely operates in the rest of eukaryotic organisms (Leung et al., 2008).

1.4 Endocytosis and the MVB pathway in parasitic protozoa

Protozoa are a diverse group of single cell eukaryotic organisms, in some of them are pathogens. Parasitic infections due to protozoa affect millions of people worldwide, causing a wide range of diseases, high rates of morbidity and mortality each year and an immense economic burden for public health (Geoff, 1997).

In pathogenic protozoa, endocytosis is a basic mechanism for ingesting host macromolecules and it has thus been associated to parasite virulence. Previous work based on ultrastructural, cytochemical, biochemical and molecular studies has shown that protozoan parasites possess the structural compartments and proteins necessary to perform endocytosis (de Souza et al., 2009). The extent of endocytic activity varies among different protozoa and even across various developmental stages. In addition, in trypanosomatids, the endocytic process is highly active in a well-defined region of the parasite cell surface called the flagellar pocket (Ghedin et al., 2001). However, only very few studies have been published to characterize the endocytic MVB pathway in protozoan parasites, some of them are summarized below.

Giardia lamblia is a protozoan parasite that causes diarrheal infections. It is also one of the most primitive organisms, with a substantially different endomembrane morphology as

compared to higher eukaryotes. Although the morphology of membrane-bound vesicles in *Giardia* has been previously described, there exists few information about vesicle budding and fusion (Lanfredi-Rangel et al., 1998). Recently, it was reported that a putative gene encoding a FYVE domain-containing protein homologous to yeast Vps27 is expressed in *G. lamblia.* This protein binds to endosomal membrane phospholipids suggesting the presence of a MVB pathway in this parasite (Sinha et al., 2010). However, very little is known about the ESCRT machinery in *Giardia* (Leung et al., 2008).

Leishmania major, a flagellated parasite provoking leishmaniasis disease, presents a plasma membrane invagination (flagellar pocket) where the flagellum emerges. This site contains a complex and highly polarized MVB-like network where endocytosis and exocytosis occur for crucial exchanges such as nutrient uptake. In this parasite, a Vps4 homologue (LmVps4) has been characterized using a Vps4 dominant negative mutant in which the highly conserved E residue required for ATP hydrolysis was substituted by a Q amino acid at position 235. The LmVps4 mutant protein was accumulated around endocytic vesicular structures and this provoked a defect in cargo protein transport to the MVB-lysosomes, as it has been reported for yeast and mammalian Vps4 mutants (Babst et al, 1998; Fujita et al., 2003). Additionally, LmVps4 is probably involved in *Leishmania* pathogenicity, since the Vps4 mutant protein also impaired parasite differentiation and virulence (Besteiro et al., 2006).

Trypanosomes infect a variety of hosts and cause several diseases, including the fatal human diseases known as sleeping sickness and Chagas disease. In this group of flagellate protozoa, the trafficking system has been previously characterized (Field et al., 2007). Trypanosomes contain glycosil-phosphatidylinositol-anchored proteins and morphologically-related MVB structures, and also exhibit ubiquitin-dependent internalization of transmembrane proteins for degradation (Allen et al., 2007; Chung et al., 2004). The functional conservation of the ESCRT system has been confirmed in *Trypanosome brucei*. Despite extreme sequence divergence, epitope-tagged *Trypanosome* TbVps23 and TbVps28 proteins localize to the endosomal pathway. Knockdown of TbVps23 partially prevents degradation of ubiquitinated proteins. Therefore, despite the absence of an ESCRT-0 complex, the MVB pathway seems to function in this parasite, similarly to the yeast and human systems (Leung et al., 2008).

Members of the *Apicomplexan* phylum of intracellular parasites, such as *Plasmodium falciparum* and *Toxoplasma gondii,* responsible for malaria and toxoplasmosis, respectively, contain morphologically unique secretory organelles termed rhoptries that are essential for host cell invasion, and also display internal membrane-resembling MVB structures (Coppens & Joiner, 2003; Hoppe et al., 2000). In *T. gondii*, it has been hypothesized that the MVB pathway could intersect with the rhoptry biogenesis one. To explore this, wild type (PfVps4) and mutant (PfVps4E214Q) *P. falciparum* Vps4 proteins were independently overexpressed in *T. gondii*. As expected, PfVps4 was located in *T. gondii* vesicular structures, whereas PfVps4E214Q was found in aberrant organelles where rhoptries proteins were also present, indicating that the secretion pathway could be disrupted by the altered Vps4 protein. These findings suggest that MVB formation may occur in *T. gondii* and *P. falciparum* and that it could be affecting the secretory route too (Yang et al., 2004).

During host cell infection, *P. falciparum* lives within a special compartment known as the parasitophorous vacuole. For the parasite to survive and multiply, molecules from the host cell cytoplasm cross the parasitophorous vacuole membrane and trigger signals for the endocytic process. Despite the scarce information being available for supporting a feasible relationship between the MVB pathway and the mechanism of nutrient uptake and intracellular

phagotrophy (the ability to ingest portions of host cytoplasm) through the parasitophorous vacuole, it may be possible that these two processes are related (de Souza et al., 2009).

E. histolytica, which causes amoebiasis, destroys almost all human tissues through macromolecules participating in adherence, contact-dependent cytolysis and proteolytic and phagocytic activities. A well-characterized protein involved in these key events is EhADH112 (García-Rivera et al., 1999). Interestingly, this protein is located at MVB-like structures in *E. histolytica* trophozoites and is structurally related to Bro1 (Bañuelos et al., 2005), an accessory protein that interacts with the ESCRT-III complex in yeast. Recently, our research group reported the presence of a set of 19 putative ESCRT proteins in this parasite and characterized a yeast Vps4 homologue by analyzing its ATPase function and relationship to parasite virulence in wild type and mutant cells (López-Reyes et al., 2010). Results derived from these studies strongly suggest that *E. histolytica* possesses a well conserved ESCRT machinery.

2. Experimental approaches for the identification and characterization of ESCRT proteins

The ESCRT components involved in mediating endosomal MVB sorting of ubiquitinated proteins have been identified and characterized by several methodologies.

Initially, over 70 *vps* genes required for the vacuolar transport of proteins were identified by genetic screening in yeast (Bonangelino et al., 2002; Bowers et al., 2004). At this moment, only 20 of these genes are known to be functionally involved in yeast MVB formation.

In addition, the structure and function of putative binding domains present in ESCRT components have been characterized using recombinant proteins and site-directed mutagenesis. In particular, ubiquitin recognition and binding to ESCRT complexes by proteins such as Hse1 and Vps27, Vps23 or Vps36 were elucidated by using crystallographic structures of recombinant proteins that associate or not, to ubiquitin. The same methodologies have been used for characterizing lipid binding domains such as the FYVE motif, present in Vps27, and for positively charged regions with affinity to phosphoinositides, such as those exhibited by Vps36 and Vps24 (Misra & Hurley, 1999; Pornillos et al., 2002; Stahelin et al., 2002; Sundquist et al., 2004).

The yeast two-hybrid system is an assay to examine protein interactions. This system includes the construction of a bait protein containing a DNA binding domain, which hybridizes to a prey protein with an activation domain. The expression of the reporter gene means that the proteins of interest interact with each other since the activation domain promotes the transcription of the reporter gene (Gietz et al., 1997). On the other hand, pull-down assays are performed either to prove a suspected interaction between two proteins or to investigate unknown proteins or molecules that may bind to a protein of interest (Kaltenbach et al., 2007). Alternatively, affinity purification of histidine- or glutathione-succinyl-transferase-(GST)-tagged bait proteins can be performed via immobilized affinity chromatography. The bait protein (or ligand) is captured to a solid support (beads) by covalent attachment to an activated beaded support or through an affinity tag that binds to a receptor molecule on the support (Pandeya & Thakkar, 2005).

In yeast, Bro1 binding to Vps32 was discovered by two-hybrid experiments, whereas Bro1 association to Vps4 was revealed by GST pull-down experiments. Additionally, using both methodologies, interactions between Vps20 and Vps28; Vps20 and Vps22; and Vps22 and

Vps28, were identified. Moreover, protein-protein interactions for ESCRT assembly have been evidenced by yeast-two-hybrid assays, affinity purification or both methods (Vps20 with Vps25 and Vps36; Vps27 with Hse1; Vps4 with Vps32; Vps22 with Vps25; and Vps22 with Vps36) (Bowers et al., 2004).

Another strategy to study protein functions is via dominant negative (DN) mutants. Mutations are changes in a genomic sequence and sometimes their expression is dominant over the wild-type protein synthesis in the same cell. Usually, DN mutants can still interact with the normal partner proteins thus blocking the functions of the wild-type protein. To improve our knowledge on the ESCRT model, several DN mutants for Vps proteins have been generated, including Hrs, Vps27, Vps23, Vps20 and Vps4 (Kanazawa et al., 2003; Li et al., 1999; Fujita et al., 2003).

Research using such strategies has increased our knowledge on the identity, structure, function and biological relationships of several molecules participating in the protein sorting through the endosomal MVB pathway. However, complementary experimental efforts need to be performed to better understand this cellular process.

3. Computational research on protein biology

One of the most familiar applications of bioinformatics is the comparison of the amino acid sequence from a query protein against the amino acid sequence of a protein previously characterized in structure and function, to theoretically elucidate whether they are related. This approach gives insights into functional similarities and evolutionary relationships deduced from the presence of common structural features (Söding, 2005).

Similarity and homology are two important concepts in the bioinformatical analysis of protein sequences. Similarity is a quantitative measure between two or more related amino acid sequences. By contrast, homology is a qualitative measure which indicates if two or more proteins are evolutionarily related or derived from a common ancestor (Claverie & Notredame, 2006). Protein sequences are usually submitted, annotated and stored in databases that allow their comparison and analysis by certain software.

In general, a database is a digital system that organizes, stores and easily retrieves large amounts of data. Currently, several genome and proteome databases are freely available for studying protein biology. However, the sheer amount of data makes highly difficult to manually interpret it. Therefore, databases require supplementary and incisive computational tools in order to understand the information.

One of the most recognized databases is the UniProt Knowledgebase (UniProtKB, http://www.uniprot.org/). The UniProtKB is the central hub for the collection of functional information on annotated proteins. The UniProtKB consists of a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis (UniProtKB/Swiss-Prot), and a section with computationally analyzed records that await full manual annotation (UniProtKB/TrEMBL). Manual annotation consists of a critical and continuously updated review of experimentally proven or computer-predicted data about each protein by an expert team of biologists.

The UnipProtKB captures the mandatory core data for each entry (amino acid sequence, protein name, description, taxonomic data and citation information) and supplementary information derived from experimental evidence or computational data.

More than 99% of the protein sequences provided by UniProtKB comes from coding sequences translation and related data submitted to the public nucleic acid databases, including the European Molecular Biology Laboratory (EMBL) Bank, the GenBank (USA) and the DNA DataBank of Japan (DDBJ). Taking advantage of the information as much as possible, there are a number of computational tools to finally interpret databases, some of them are briefly described below.

The Expert Protein Analysis System (ExPaSy) is a proteomics server from the Swiss Institute of Bioinformatics that analyzes protein sequences and structures and contains genome databases for several organisms ranging from *Archae* to human (http://expasy.org/ tools/#proteome). It has several tools useful to depict primary, secondary and tertiary protein structures and to determine putative postranslational modifications, among others.

The Basic Local Alignment Search Tool (BLAST) is an algorithm for comparing primary biological sequence information, such as amino acid sequences of different proteins or nucleotides of distinct DNA sequences. A BLAST search enables a researcher to compare a query sequence with data existing in sequence libraries or databases, and to identify the sequences that resemble the query sequence above a certain threshold. The main idea of BLAST is that there are often high-scoring segment pairs (HSP) contained in a statistically significant alignment. BLAST searches for high scoring sequence alignments between the query sequence and sequences from genome databases, using a heuristic approach that approximates the Smith-Waterman algorithm (Altschul et al., 1990). The BLASTP program, which compares protein queries to protein databases, is a heuristic model that attempts to optimize a specific similarity measure. The goal of this tool is to find regions of sequence similarity. These regions can yield clues about the structure and function of the novel sequence and its evolutionary history and homology by comparison to other sequences in databases (Henikoff & Henikoff, 2000). To produce a multiple sequence alignment from the BLASTP output, this program simply collects all database sequence segments that have been aligned to the query with an expectation value (*E*-value) below a threshold by a default set to 0.001. Thus, the lower the *E*-value, the greater the similarity between the input and the match sequences will be. An *E-value* < e-3 of an alignment means that the alignment is highly unique and not due to error (http://bips.u-strasbg.fr/fr/Tutorials/Comparison/ Blast/blastall.html). As an alternative for accurate searches of query sequences, the Position Specific Iterative (PSI)-BLAST program iteratively searches for one or more proteins databases to find sequences similar to one or more protein query sequences.

ClustalW is also a widely used multiple sequence alignment computer program (http://align.genome.jp/). In many cases, the input set of query sequences is assumed to have an evolutionary relationship, share a lineage and descend from a common ancestor. This algorithm is usually supplemented by the BOXSHADE application (http://www.ch.embnet.org/software/BOX_form.html). BOXSHADE is a program for creating good looking printouts from multiple-aligned protein or DNA sequences. BOXSHADE does not produce alignments by itself, it has to take as input a file preprocessed by a multiple alignment program or a multiple file editor such as ClustalW. In the standard BOXSHADE output, identical and similar residues in the multiple-alignment chart are represented by different colors or shadings.

3.1 Computational tools for predicting protein domains

Protein domains, defined as the independent folding units within a polypeptide, are also understood as the functional and evolutionarily conserved modules of protein families.

The Pfam protein family database is a large collection of multiple sequence alignments that is generated by probabilistic models known as hidden Markov models (HMM) (http://www.sanger.ac.uk/resources/databases/pfam.html). The Pfam database contains information about protein domains and families. For each family in Pfam, one can look at multiple alignments, view protein domain architectures, examine species distribution, and follow links to other databases and view known protein structures.

Despite the increasing volume of biochemical and molecular literature on protein data, Pfam contains the essential information about major protein domains for the understanding of the ever more complicated biological landscape.

Since the ClustalW and BOXSHADE programs could be useful to identify conserved residues and similar regions among amino acid sequences, they also allow the prediction of putative domains in a protein or group of proteins of interest.

3.2 Computational approaches for predicting secondary protein structure

Secondary structure refers to highly regular local sub-structures within a molecule. The secondary structure of a protein is defined by patterns of hydrogen bonds between the main-chain peptide groups, leading to several recognizable protein domains, such as alpha (α) helices and beta (β) sheets (Offer et al., 2002).

So far, several algorithms have been described for predicting secondary protein structures, one of them being Jpred (Cole et al., 2008). Jpred uses a 3-iteration PSI-BLAST search to obtain sequences from existing databases for predicting secondary structures. Jpred now includes Jnet, a neural network method also for secondary structure prediction. The Jnet algorithm works by applying multiple sequence alignments, alongside PSI-BLAST and HMM profiles (Cuff & Barton, 1999). The updated Jnet algorithm provides α-helix and β-sheet predictions at an accuracy of 81.5% (Cole et al., 2008).

3.3 Computational algorithms for predicting tertiary protein structure

The tertiary structure of a protein refers to the three-dimensional arrangement of a single protein molecule. The α-helices and β-sheets are folded into a compact structure due to non-specific hydrophobic interactions. However, this structure is stable only when the parts of a protein domain are locked into place by specific tertiary interactions, such as salt bridges, hydrogen bonds, and the tight packing of side chains and disulfide bonds (Peng & Kim, 1994).

The Protein Data Bank (PDB) contains information about experimentally-determined structures of proteins and nucleic acids, and complex assemblies (http://www.pdb.org/pdb/home/home.do). The Resource for Studying Biological Macromolecules curates and annotates the PDB data according to agreed upon standards and also provides a variety of tools and resources. Interestingly, the PDB is a repository for three dimensional structural data of proteins (typically obtained by X-ray crystallography or Nuclear Magnetic Resonance spectroscopy) submitted by biologists and biochemists from around the world.

The PDB is a key resource in areas of structural biology, such as structural genomics. Contents of the PDB are thought to be primary data, and currently there are hundreds of derived databases that categorize data differently.

The Phyre (Protein homology/analogy recognition engine) webserver is a powerful computational tool that uses profile-profile matching algorithms to considerably improve

protein predictions (Kelley & Stenberg, 2009).   The Phyre platform follows the most successful general approaches for predicting the structure of proteins, which involve the detection of homologues of a known three dimensional structure, the so-called template-based homology modeling and fold-recognition. Practical applications from three dimensional protein structure predictions include guidance on functional hypothesis, the selection of mutagenesis sites and the design of rational drugs, among others.

The Phyre server uses a library of known protein structures taken from the SCOP (Structural classification of proteins) database and augmented with newer depositions in the PDB. Sequences of each of these structures are scanned against a non-redundant sequence database and a profile is constructed and deposited in a "fold" library. The known and predicted secondary structures of these proteins are also stored in the fold library. A user-submitted sequence follows the same process. Five iterations of PSI-BLAST are used to gather both close and remote sequence homologues. The pairwise alignments generated by PSI-BLAST are combined into a single alignment with the query sequence as the master. Following the profile construction, the secondary structure of the query is predicted using three distinct programs (Psi-Pred, SSPro and Jnet). Subsequently, both profile and secondary structure, are scanned against the fold library using a profile-profile algorithm that returns a score. Scores are fitted to an extreme value distribution to generate an *E*-value. The top ten highest scores are then used to construct full three-dimensional models for the query. Where possible, missing or inserted regions caused by deletions or insertions in the alignment are repaired using a loop library and reconstruction procedures.

An alternative program widely used to model tertiary protein structures is SWISS-MODEL. SWISS-MODEL is a fully automated protein structure homology-modeling server accessible via the ExPASy web server or from the DeepView program (http://swissmodel.expasy.org/). The purpose of this server is to make protein modeling accessible to all biochemists and molecular biologists worldwide by providing tools for protein structure accurate predictions.

Once a tertiary structure has been modeled, it is sometimes necessary to get access into a model viewer. Jmol is a free open-source viewer for chemical three dimensional structures that is written in Java (so it runs on Windows, Mac OS X, Linux and UNIX systems). Jmol returns a representation of a molecule that may be used as a teaching tool, or for research e.g. in chemistry and biochemistry. The most notable feature is an applet that can be integrated into web pages to display molecules in a variety of models: "ball and stick", "space filling", "ribbon", etc. (http://jmol.sourceforge.net/download/).

4. ESCRT protein survey in protozoan parasites with bioinformatical tools

By using a bioinformatical screening and comparative genomic analysis, we confirmed in this work the presence of ESCRT representatives in unrelated groups of unicellular parasites of medical importance belonging to the following taxa: *Entamoebidae (Entamoeba), Diplomonadida (Giardia), Alveolata* of the phyllum *Apicomplexa (Toxoplasma* and *Plasmodium),* and *Kinetoplastida (Trypanosoma* and *Leishmani*a).

First, we obtained yeast or mammalian amino acid sequences for ESCRT-0 to -III and -associated proteins from the UniProtKB database. Then, the retrieved sequences were used as probes to screen the Eukaryotic Pathogen database (EuPathDB version 2.9, http://eupathdb.org/eupathdb/). The EuPathDB has been developed as a Bioinformatics Resource Center and constitutes an integrated genome database covering eukaryotic

pathogens of the genera *Cryptosporidium, Giardia, Entamoeba, Leishmania, Plasmodium, Toxoplasma, Trichomonas* and *Trypanosoma,* among others. This portal offers an entry point to all these resources, and the opportunity to leverage orthology (structural correspondence or similarity of genes or proteins in different species due to a common ancestor origin) for searches across genera in an interface that is functional, user-friendly and sophisticated.

Using yeast ESCRT protein sequences as queries in the EuPathDB resource for each parasite genome, the BLASTP program reported several amino acid sequences for each pathogen. When no matches were found, human corresponding ESCRT protein sequences were used as queries. Putative parasite ESCRT homologous sequences were selected with the following criteria: *i)* at least 20% identity and 35% similarity to the query sequence, *ii)* E-value lower than 0.002, and *iii)* absence of stop codons in the coding sequence. Furthermore, all recovered sequences were subjected to reverse BLAST analysis in the ExPaSy server to identify related proteins from genome databases. A candidate was taken into consideration if reverse BLAST recovered the original query within the top five hits. Failure to complete these tests resulted in a "not determined" assignment.

BLAST results showed that all parasites studied here contain putative protein sequences representing the ESCRT-0 to -III and -accessory proteins involved in the endocytic MVB pathway. In Table 1, we summarized the results derived from our parasite ESCRT genomic survey in comparison to ESCRT members previously reported in yeast or human. The major noticeable feature was the high conservation of ESCRT components in all taxa, as previously reported (Leung et al., 2008). As noticed, *Entamoeba histolytica* and *Leishmania major* contain the most represented and conserved ESCRT machinery among parasites, with 19 ESCRT components. Meanwhile, *Trypanosoma cruzi* and *Plasmodium falciparum* displayed 15 and 14 ESCRT putative proteins, respectively. By contrast, we only found 9 out of the 20 ESCRT proteins in *Toxoplasma gondii* and *Giardia lamblia.*

Ubiquitin-label recognition is the signal for cargo protein entrance towards degradation through the endosomal pathway (Bowers et al., 2004). Rsp5 and Bul1 proteins mediate ubiquitin-attachment to cargo proteins in yeast. Here, bioinformatical approaches revealed that ubiquitination seems to be mediated by Rsp5 rather than Bul1 homologues, since Rsp5-like proteins were present in all protozoan genomes.

Unlike preceding work, we found at least one ESCRT-0 representative for each parasite, indicating that proteins recognizing ubiquitin signals could be participating in cargo sorting in these protozoa. ESCRT-I and -II were the least represented complexes among all parasites, suggesting that some taxa members could have lost specific components along ESCRT evolution. However, we cannot exclude that the lack of individual ESCRT components might be the result of malfunctionings in gene or protein detection, more than a real absence of the protein. In particular, failures have been frequently reported for *Giardia* due to difficulties to recover candidate orthologues in its extremely divergent genome.

To the best of our knowledge, there is no sequenced eukaryotic genome without an ESCRT-III-related gene. Moreover, the size of the subset of ESCRT-III-related genes is greatly expanded in higher eukaryotes such as mammals, compared to yeast. As a consequence, it has been hypothesized that the ESCRT-III complex might be the minimal ESCRT unit for MVB formation (Williams & Urbé, 2007). Consistently, our results revealed at least two ESCRT-III representatives in each parasite genome analyzed.

Regarding the ESCRT-accesory proteins, the most conserved sequences among all parasites were the Rsp5, Vps4, Vps46, Doa4, Vta1 and Bro1 homologues, in contrast to Ist1, which was only present in trypanosomatids.

Taken together, our *in silico* results support the existence of a seemingly conserved ESCRT machinery for endosomal protein trafficking through the MVB pathway in protozoan parasites.

| | | Amoebida | Diplomonadida | Eucoccidiorida | Haemosporida | Trypanosomatida | |
| | Saccharomycetales / Primates | | | | | | |
ESCRT complex	*Saccharomyces cerevisiae/ Homo sapiens*	*Entamoeba histolytica*	*Giardia lamblia*	*Toxoplasma gondii*	*Plasmodium falciparum*	*Leishmania major*	*Trypanosoma cruzi*
ESCRT-0	Hse1	+	-	-	-	+	+
	Vps27	+	+	+	+	+	+
ESCRT-I	Vps23/Tsg101	+	-	-	-	+	+
	Vps28	+	-	-	+	+	-
	Vps37/hVps37D	+	-	-	+	+	+
	Mvb12	-	-	-	-	-	-
ESCRT-II	Vps22/EAP30	+	+	-	-	+	-
	Vps25/EAP20	+	+	-	-	-	-
	Vps36	+	-	-	-	+	+
ESCRT-III	Vps2	+	+	-	+	+	-
	Vps20	+	-	-	+	+	+
	Vps24	+	+	+	+	+	+
	Vps32	+	-	+	+	+	+
Upstream ESCRT-accessory proteins	Bul1	-	-	-	-	-	-
	Rsp5	+	+	+	+	+	+
Downstream ESCRT-accessory proteins	Vps4	+	+	+	+	+	+
	Ist1	-	-	-	-	+	+
	Vta1	+	-	+	+	+	+
	Vps46	+	+	+	+	+	+
	Vps60	+	-	-	+	+	+
	Doa4	+	+	+	+	+	+
	Bro1	+	-	+	+	+	+

Table 1. Comparison of ESCRT machineries from parasitic protozoa.
The presence (+) of homologous proteins is based on data obtained by BLAST searches from protein sequence databases at NCBI, UniProtK and EuPathDB, as described in the text.
Proteins apparently absent (-) from complete genome sequencing projects are indicated.

4.1 Characterization of the ESCRT machinery in *E. histolytica*

Our previous work, using comparative genomics for predicting ESCRT proteins in *E. histolytica*, provided valuable insights into the existence of a highly conserved ESCRT machinery in this parasite. López-Reyes et al. (2010) reported a set of 19 putative ESCRT proteins representing from ESCRT-0 to -III and -associated proteins (Table 2). Moreover, earlier characterization of ubiquitin genes and -transcripts and demonstration of an ubiquitin-conjugating system, together with our finding of a putative Rsp5 ubiquitin ligase (EhRsp5) *E. histolytica* provided additional support for the presence of at least one candidate that possibly mediates ubiquitin attachment to cargo molecules prior to their internalization into endosomes (Wöstmann et al., 1996).

Previous work has provided knowledge into the architecture, membrane recruitment and functional interactions of the ESCRT machinery through multiple domains that have been shaped along evolution. These scaffolds serve as gripping tools for recognizing cargo proteins, membrane lipids, ESCRT components and accessory proteins along the MVB route (Hurley & Emr, 2006).

To dissect the presence of putative ubiquitin and phosphoinositide binding domains in *E. histolytica* ESCRT-like components, we selected ESCRT-0 to -III representatives (EhVps27, EhVps23, EhVps36 and EhVps24, respectively) presumably containing these structural features according to their yeast and human homologues and performed multiple sequence alignments with the ClustalW program.

Putative complex	*Entamoeba histolytica*		*Homo sapiens*					*Saccharomyces cerevisiae*				
	Predicted protein	Accesion number	Protein	Accesion number	E-value	S (%)	I (%)	Protein	Accesion number	E-value	S (%)	I (%)
ESCRT-0	EhHse1	C4M9E1	STAM1	Q92783	9e-08	72	50	Hse1	P38753	2e-10	70	62
	EhVps27	C4LYX5	HRS	O14964	---	---	---	Vps27	P40343	1e-11	61	49
ESCRT-I	EhVps23	C4LUR9	TSG101	Q99816	1.4e-10	51	33	Vps23	P25604	0.0052	31	25
	EhVps28	C4M842	hVps28	Q9UK41	--------	---	---	Vps28	Q02767	0.00057	45	28
								Vps37	Q99176	0.00084	46	21
	EhVps37D	C4M6J5	hVps37D	Q6P2C3	3.8e-06	50	35	nd	---------	--------	---	---
	------------	-----------	nd	-----------	--------	---	---	Mvb12	P42939	--------	---	---
ESCRT-II	EhVps22	C4LXI0	EAP30	Q96H20	9e-22	47	30	Vps22	Q12483	3.4e-15	48	25
	EhVps25	C4LZE7	EAP20	Q9BRG1	9e-08	52	28	Vps25	P47142	0.00038	46	19
	EhVps36	CALTE5	EAP45	Q86VN1	1.8e-15	49	25	Vps36	Q06696	1.1e-09	48	26
ESCRT-III	EhVps2	C4LZV3	CHMP2A	O43633	8.9e-24	55	29	Vps2	P36108	2e-05	50	29
			CHMP2B	Q9UQN3	7.9e-16	48	25	---------	---------	--------	---	---
	EhVps20	C4MAC7	CHMP6	Q96FZ7	--------	---	---	Vps20	Q04272	0.00015	54	26
	EhVps24	C4M2Y2	CHMP3	Q9Y3E7	--------	---	---	Vps24	P36095	2.2e-05	48	22
	EhVps32	C4M1A5	CHMP4A	Q9BY43	0.0012	48	24	Vps32	P39929	2.5e-12	48	25
			CHMP4B	Q9H444	9.6e-07	43	20	---------	---------	--------	---	---
			CHMP4C	Q96CF2	1.4e-06	43	20	---------	---------	--------	---	---
Upstream ESCRT-accessory proteins	-----------	-----------	nd	-----------	--------	---	---	Bul1	P48524	--------	---	---
	EhRsp5	C4M9I9	NEDD4	P46934	1.6e-50	53	33	Rsp5	P39940	7.8e-57	57	36
Downstream ESCRT-accessory proteins	EhVps4	C4LYN8	hVps4A	Q9UN37	1.1e-114	69	52	Vps4	P52917	3e-114	78	60
			hVps4B	O75351	3e-114	78	60	---------	---------	--------	---	---
			hIst1	P53990	--------	---	---	Ist1	P53843	--------	---	---
	EhVta1	C4M0R8	hVta1	Q9NP79	2.8e-05	44	23	Vta1	Q06263	--------	---	---
	EhVps46	C4LYH6	CHMP1A	Q9HD42	--------	---	---	Vps46	P69771	0.00032	48	18
			CHMP1B	Q7LBR1	--------	---	---	---------	---------	--------	---	---
	EhVps60	C4M7T5	CHMP5	Q9NZZ3	5.6e-15	45	28	Vps60	Q03390	9e-06	46	19
	EhDoa4	C4LTM2	UBP4	Q13107	1e-77	58	36	Doa4	P32571	2e-27	50	31
	EhADH112	Q9U7F6	Alix	Q9UKL5	1e-21	40	20	Rim20	Q12033	2e-12	38	21
								BRO1	P48582	0.003	50	32

Table 2. Comparison of *E. histolytica, H. sapiens* and *S. cerevisiae* ESCRT machineries. Data of conserved ESCRT proteins from yeast and human were obtained at NCBI and UniProtKB databases. Putative *E. histolytica* ESCRT proteins were retrieved by BLAST searches at EupathDB and corresponding UniProtKB accession numbers were obtained. Putative ESCRT proteins of *E. histolytica* exhibited significant *E*-values (1.1*e*-114 to 0.00032) and high similarity (20 to 62%) to yeast and human ESCRT orthologues. nd, not determined; ----, non-significant similarity or identity and *E*-values; S, similarity; I, identity (Modified from López-Reyes et al., 2010).

Our computational comparative analysis showed that the ESCRT-0 complex, lacks the characteristic VHS (Vps27, Hrs and STAM) domain of yeast Vps27, required for the protein interaction with ubiquitin (Williams & Urbé, 2007). However, EhVps27 displayed a (R/K)(R/K)HHCR motif usually found within conserved FYVE domains and necessary for phosphatidylinositol 3-phosphate (PtdIns3P) binding (Misra & Hurley, 1999). This finding was also supported by the Pfam database, which reported the presence of a putative FYVE domain in the EhVps27 amino acid sequence.

Membrane phospholipids such as PtdIns3P, have been previously implicated in the regulation of endocytosis and phagocytosis and 12 FYVE-domain containing proteins have been identified in *E. histolytica* (Nakada-Tsukui et al., 2009).

The UEV domain present in yeast Vps23 and its human homologue Tsg101 is necessary to recognize ubiquitin signals in proteins to be sorted into MVB (Pornillos et al., 2002; Sundquist et al., 2004). Despite a less conserved similarity among analyzed sequences, our bioinformatics approach suggested the presence of a putative UEV domain at the N-terminus of EhVps23, also supported by Pfam domain predictions.

According to our current investigation, EhVps36 lacks the yeast NZF and human GLUE domains previously reported in Vps36 homologues. Both domains have been implicated in ubiquitin and PtdIns3P binding, respectively. Instead, EhVps36 conserves a N-terminal positively charged amino acid region. Similarly, the EhVps24 protein exhibits a positively charged amino acid tract present in almost its full sequence. Since specific binding to phosphoinositides requires electrostatic interactions between negatively charged phosphates on lipids and positively charged amino acids in proteins, it is feasible that EhVps36 and EhVps24 associate to phosphoinositides present at endosomal membranes (Whitley et al., 2003).

Secondary structure assignments for putative ESCRT proteins of *E. histolytica* were achieved by using the Jpred program. In agreement with our previous findings, EhVps27, EhVps36 and EhVps24, and EhVps23 proteins resulted in similar arrangements to yeast Vps27, Vps36 and Vps24 proteins, and human Tsg101, respectively. Furthermore, according to both Phyre and SWISS-MODEL tertiary structure predictions, the three-dimensional structures of EhVps27 and EhVps36 matched to yeast Vps27 (PDB code: 1vfy) and Vps36 (PDB code: 1u5t) crystalline structures, respectively. In addition, the Phyre software predicted a conformational arrangement similar to human Tsg101 (PDB code: 1s1q) and CHMP3 (PDB code: 2gd5) proteins for EhVps23 and EhVps24, respectively.

Altogether, our results indicate the presence of putative structural and conformational features for ubiquitin and lípid binding in representative proteins from the *E. histolytica* ESCRT-0, -I, -II and –III complexes.

To determine the identity of putative ESCRT-accesory proteins, we first focused on EhADH112, a protein widely studied by our group and involved in *E. histolytica* adherence to and phagocytosis of host cells (García-Rivera et al., 1999). *In silico* analysis of the primary sequence of EhADH112 together with Pfam protein domain predictions, revealed that EhADH112 is structurally related to yeast Bro1 and its human homologue Alix. EhADH112 has a conserved Bro1 domain at its N-terminus. In Bro1 and Alix proteins, the Bro1 domain constitutes the interacting site for Vps32 or CHMP4B, respectively, both components of the ESCRT-III complex. Experimental approaches demonstrated that *E. histolytica* parasites overexpressing only a part of the EhADH112 Bro1 domain, reduced dramatically their ability to ingest cells, thus providing additional evidence for EhADH112 participation in phagocytosis (our unpublished results). Furthermore, immunolocalization of EhADH112 and truncated EhADH112 proteins in parasites, using both transmission electron and laser confocal microscopy, revealed that besides its detection at the plasma membrane and cytoplasmic vacuoles, EhADH112 is also in MVB-like organelles, whereas the EhADH112 mutant version accumulates in cytoplasmic vesicles. These findings led us to assign a putative role for the EhADH112 Bro1 domain to recruit proteins to the endosomal membranes forming MVB. Possibly, Vps proteins from the ESCRT-III complex or some other molecules could be involved in this event, thus affecting the *E. histolytica* phagocytosis process. In order to identify putative interacting partners for EhADH112, we used a

computational survey for yeast Vps32 or human CHMP4B homologous sequences in the *E. histolytica* genome. We found a putative EhVps32 protein whose existence in *E. histolytica* was confirmed by further experimental data (Bañuelos et al., 2007). According to multiple sequence analysis and Pfam database predictions, EhVps32 contains a Snf7 domain, present in all members of the Snf7 family. Additionally, the predicted EhVps32 secondary structure using the Jpred program, suggested that EhVps32 conserves the characteristic five α-helices present in the Snf7 family protein (Fig. 2A). Using the Phyre program, the tertiary structure for EhVps32 was modeled. Results showed that the predicted structure of EhVps32 is related to human CHMP3, a Snf7 family member (Fig. 2B). Since the crystal structure for CHMP4B has not yet been solved, the program uses by default the CHMP3 crystal structure as template due to the presence of the highly conserved Snf7 domain. Thus, tertiary structures for CHMP4B and Vps32 were also modeled using CHMP3 as template (Fig. 2B). Retrieved results showed that EhVps32 adopts a conformational structure and folding more similar to CHMP4B than to yeast Vps32 and this is in agreement with the highest similarity reported for EhVps32 to the human sequence of CHMP4B by BLAST analysis (Table 2).

To confirm the predicted interaction between EhADH112 and EhVps32 proteins, pull down experiments were perfomed. Assays demonstrated that EhADH112 binds through its N-terminus to a recombinant protein of EhVps32 fused to GST (our unpublished data).

Since yeast Vps4 and its orthologues have been previously described as key molecules for ESCRT dissociation and recycling, López-Reyes et al., (2010) characterized the EhVps4 protein in more detail. Protein domain predictions, as well as tertiary structure modeling and phylogenetic trees assayed for EhVps4 suggest, that it conserves a typical Vps4 architecture (Babst et al., 1998) and is more related to protozoan Vps4 homologues than to that of higher eukaryotes. Biochemical experiments using an EhVps4 recombinant protein and ATP as substrate, evidenced the ATPase activity of EhVps4 *in vitro*. As expected, when using a mutant version of EhVps4, in which an E residue was substituted by a Q amino acid, the ATPase activity was reduced. Furthermore, *E. histolytica* parasites overexpressing the EhVps4 mutant protein displayed reduced virulence properties, suggesting a role for EhVps4 in parasite pathogenicity, probably related to its participation in the endocytic pathway.

5. Challenges and perspectives

Our previous results obtained via bioinformatical tools and biochemical experiments, allow us to propose a model for the ESCRT machinery in *E. histolytica* (Fig. 3). Since we found the EhVps27 component of the ESCRT-0 complex, we suggest that it may initiate the MVB sorting process. Additionally, EhVps27 has a FYVE domain that possibly mediates protein binding to the endosomal membrane. However, EhVps27 lacks the UIM domain, important for the initial selection of ubiquitinated cargo, probably by EhRsp5. Perhaps, EhVps23, through its UEV motif, or another unidentified protein could be recruiting cargo proteins to endosomes. Furthermore, the EhVps23 UEV domain could associate toEhVps27 and other components of the ESCRT-I complex, which includes the EhVps28 and EhVps37 proteins. Then, ESCRT-I binds to ESCRT-II (formed by EhVps22, EhVps25 and EhVps36 proteins). Although EhVps36 does not exhibit an ubiquitin-interacting domain as yeast homologues, this protein contains a recognition region for phosphoinositides that presumably would allow ESCRT-II attachment to the endosomal membrane. Next, ESCRT-II binds to the ESCRT-III complex, which contains the overall components previously described for yeast. Interestingly, similarly to yeast Vps20, EhVps20 has a myristoylated modification that facilitates ESCRT-III insertion into the endosomal membrane. Then, ESCRT-III interaction
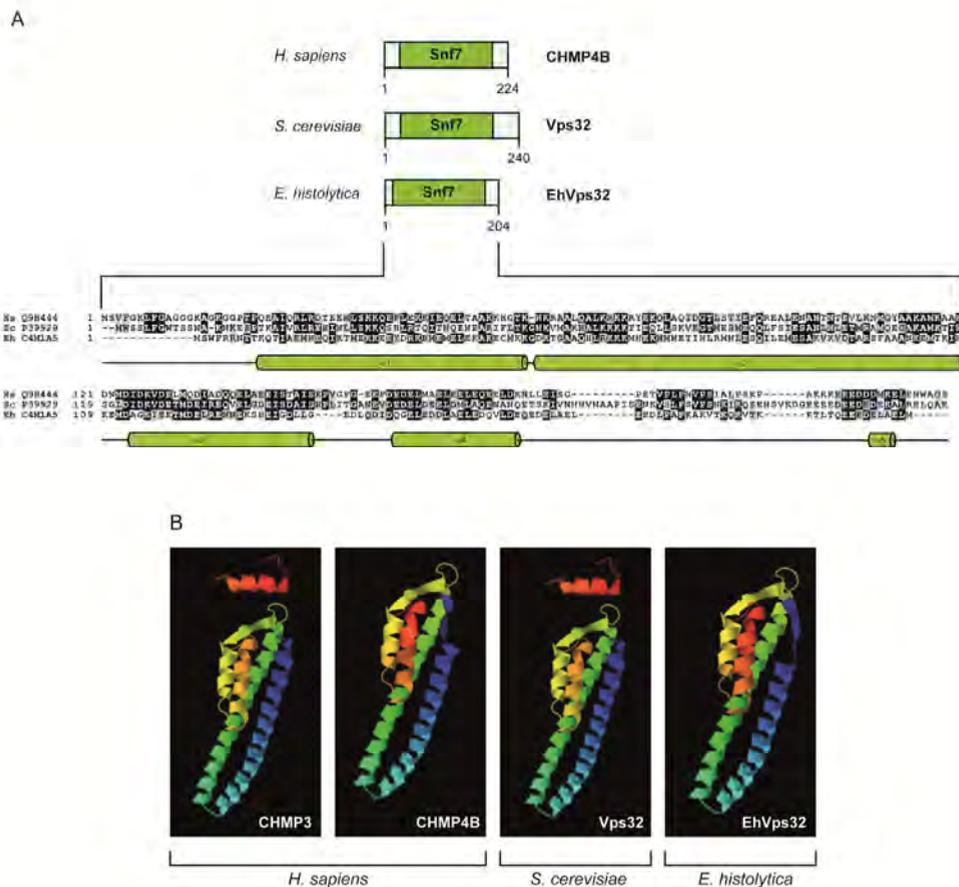
Fig. 2. Structural comparison of Vps32 homologues.
(A) At the top, a schematic representation for human CHMP4B, yeast Vps32 and *E. histolytica* EhVps32 proteins is shown. Numbers indicate amino acids for each protein. All proteins contain conserved Snf7 domains, present in the Snf7 family proteins. Vps32 orthologues belong to the ESCRT-III complex and have been described as the interacting partners of Bro1 domain-containing proteins. At the bottom, a multiple sequence alignment for Vps32 homologues is shown. Hs, *H. sapiens*; Sc, *S. cerevisiae*; and Eh, *E. histolytica*. Black boxes, identical amino acids; grey boxes, conserved substitutions; and open boxes, different residues. Numbers at left are relative to the position of the start codon in each protein. The Jpred secondary structure prediction program revealed that EhVps32 folds into five α-helices (green horizontal cylinders) as it has been reported for Vps32 homologues. (B) Tertiary protein structure for *H. sapiens*, *S. cerevisiae* and *E. histolytica* Vps32 homologues. Modeling was done using the Phyre program with the crystal structure of human CHMP3 as template.
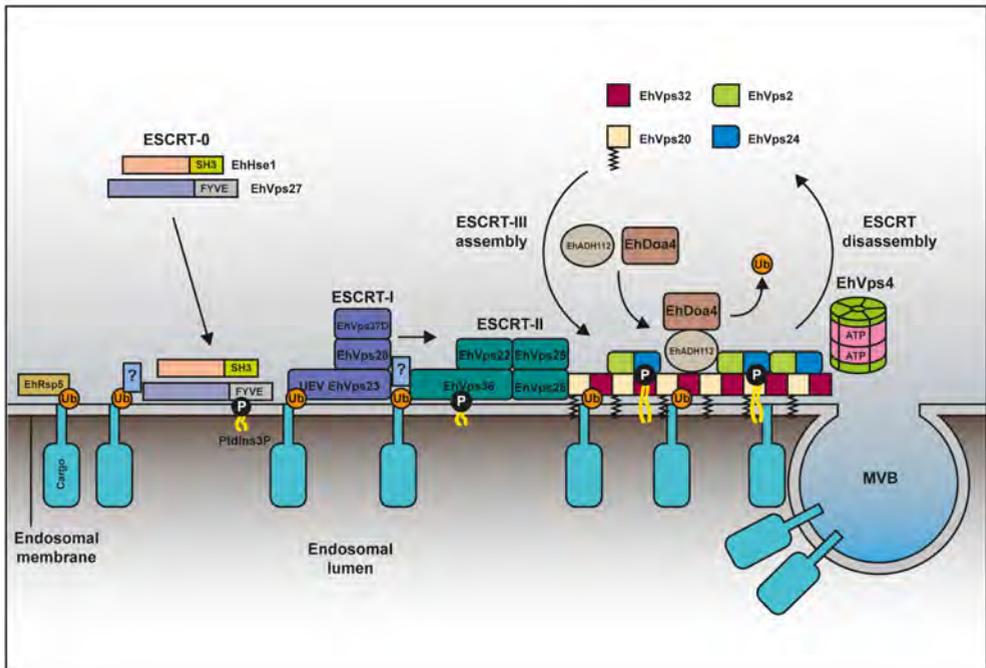
Fig. 3. Model for the role of the ESCRT machinery in *E. histolytica* within the endosomal MVB pathway.

In *E. histolytica,* the EhRsp5 protein could be responsible for cargo protein ubiquitination. Then, the EhVps27 protein could initiate the MVB process. Similar to yeast Vps27, EhVps27 has a FYVE domain that binds PtdIns3P allowing endosomal membrane attachment. However, EhVps27 lacks the UIM domain, important for ubiquitin recognition in cargo proteins. Instead, EhVps23 could be mediating this event through its UEV motif. Subsequently, EhVps27 binds to the ESCRT-I complex through EhVps23. Then, EhVps36 by its positively charged region binds to PtdIns3P, facilitating the ESCRT-II attachment to endosomal membranes. *E. histolytica* contains all ESCRT-III components which belong to the Snf7 family of proteins. In addition, it has several accessory proteins, including the EhADH112 (a Bro1 domain-containing protein), EhDoa4 (deaubiquitinating enzyme that removes ubiquitin from cargo) and EhVps4 (an ATPase) proteins. Finally, as in yeast, EhVps4 may play a critical role in catalyzing the dissociation of ESCRT from the endosomal membrane in order to start new rounds of cargo protein sorting through MVB.

with accessory proteins could be mediated by EhVps32. In fact, EhVps32 could associate to EhADH112 through its putative N-terminal Bro1 domain (our unpublished data). Besides, EhADH112 could also be recruiting another accessory molecule, the EhDoa4 ubiquitin hydrolase, removing ubiquitin from cargo prior MVB internalization. Finally, the EhVps4 ATPase might catalyze the disassembly of the ESCRT complex from the endosomal

membrane to initiate new rounds of cargo sorting and vesicle formation. Possibly, EhVta1 may have a role in regulating EhVps4 function.

Of note, *E. histolytica* possesses a conserved ESCRT machinery. However, the study related to ESCRT functions and putative interactions along the MVB pathway needs to be corroborated by experimental approaches.

6. Conclusions

Bioinformatics, the application of statistics and computer sciences to molecular biology, entails the creation and advancement of databases, algorithms, computational and statistical techniques and theory to solve formal and practical problems arising from the management and analysis of biological data. In this chapter, we used bioinformatics to analyze the ESCRT protein machinery possibly participating in parasitic protozoa endosomal pathways, with particular attention on the *E. histolytica* case.

The ESCRT machinery comprises a set of protein complexes that regulate recognition, sorting and trafficking of monoubiquitinated proteins into MVB compartments towards lysosome degradation. Previous work has shed light on molecular details underlying the assembly and regulation of ESCRT in yeast and human. Here, we took advantage from eukaryotic pathogen genome database availability and bioinformatics tools to identify proteins representing putative ESCRT components in protozoan parasites of medical importance. We found representative proteins for ESCRT-0, -I, -II, -III and -accesory proteins in almost all protozoa examined, being *E. histolytica* and *L. major* the parasites in which ESCRT components were the most represented. Despite these findings, several issues need to be experimentally addressed to finely determine the structure and function of ESCRT proteins and their putative role during endocytosis in these parasites.

In *E. histolytica,* we found a highly conserved ESCRT machinery with 19 putative components representing all complexes. These findings have been experimentally confirmed by determining the expression of most ESCRT gene transcripts (López-Reyes, et al., 2010). Furthermore, our current *in silico* results suggest that some *E. histolytica* ESCRT-0 to -III components contain putative FYVE or ubiquitin binding domains, both important to recruit cargo molecules to endosomal membranes. In addition, our computational analysis together to previous functional characterization of putative *E. histolytica* ESCRT-accessory proteins, strongly suggest the presence of a Bro1-domain containing protein (EhADH112), its putative interacting partnership, EhVps32, and an ATPase (EhVps4) that may be responsible for energy-dependent ESCRT disassembly. Of note, tertiary structure modeling of EhVps32 supported our experimental findings on EhADH112 binding to EhVps32, proving the value of bioinformatical approaches. Therefore, our overall results provide significant evidence for a conserved role of the *E. histolytica* ESCRT machinery in the MVB endocytic pathway.

In summary, bioinformatics and experimental approaches can improve our understanding on evolutionary implications of the MVB sorting pathway in *E. histolytica, L. major, T. cruzi, P. falciparum, T. gondii* and *G. lamblia* and also for elucidating its possible relationship to parasite pathogenicity and virulence.

Although some limitations exist due to incompleteness of experimental data, we conclude that computational methods have a reasonable prediction accuracy and provide invaluable basis for further experimental validation.

7. Acknowledgements

8. References

Allen, C.L., Liao, D., Chung, W.L. & Field, M.C. (2007). Dileucine signal-dependent and AP-1-independent targeting of a lysosomal glycoprotein in *Trypanosoma brucei*. *Molecular and Biochemical Parasitology*, Vol. 156, pp. 175–190, ISSN 0166-6851

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology,* Vol. 215, No. 3, pp. 403-410, ISSN 0022-2836

Babst, M., Katzmann, D.J., Estepa-Sabal, E.J., Meerloo, T., & Emr, S.D. (2002a). ESCRT-III: an endosome-associated heterooligomeric protein complex required for MVB sorting. *Developmental Cell,* Vol. 3, pp. 271–282, ISSN 1534-5807

Babst, M., Katzmann, D.J., Snyder, W.B., Wendland, B., & Emr, S.D. (2002b). Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. *Developmental Cell,* Vol. 3, pp. 283–289, ISSN 1534-5807

Babst, M., Wendland, B., Estepa, E.J., & Emr, S.D. (1998). The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *The EMBO Journal,* Vol. 17, pp. 2982–2993, ISSN 0261-4189

Bañuelos, C., García-Rivera, G., López-Reyes, I., & Orozco, E. (2005). Functional characterization of EhADH112: an Entamoeba histolytica Bro1 domain-containing protein. *Experimental Parasitology,* Vol. 110, No. 3, pp. 292-297, ISSN 0014-4894

Bañuelos, C., López-Reyes, I., García-Rivera, G., González-Robles, A. and Orozco, E. (2007). The presence of a Snf7-like protein strenghtens a role for EhADH in the *Entamoeba histolytica* multivesicular bodies pathway. *Proceedings of the 5th European Congress on Tropical Medicine and International Health*, Boeree, M.J. (ed), Vol. 978, pp. 31–35, Amsterdam, The Netherlands

Besteiro, S., Williams, R.A., Morrison, L.S., Coombs, G.H., & Mottram, J.C. (2006). Endosome sorting and autophagy are essential for differentiation and virulence of Leishmania major. *The Journal of Biological Chemistry,* Vol. 281, No. 16, pp. 11384-11396, ISSN 0021-9258

Bonangelino, C.J., Chavez, E.M., & Bonifacino, J.S. (2002). Genomic screen for vacuolar protein sorting genes in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell,* Vol. 13, pp. 2486–2501. ISSN 1059-1524

Bowers, K., Lottridge, J., Helliwell, S.B., Goldthwaite, L.M., Luzio, J.P. & Stevens, T.H. (2004). Protein–Protein Interactions of ESCRT Complexes in the Yeast *Saccharomyces cerevisiae*. *Traffic,* Vol. 5, pp.194–210, ISSN 1398-9219

Chiang, Y.S., Gelfand, T.I., Kister, A.E. & Gelfand, I.M. (2007). New classification of supersecondary structures of sandwich-like proteins uncovers strict patterns of strand assemblage. *Proteins*, Vol. 68, No. 4, pp. 915–921, ISSN 0887-3585

Chung, W.L., Carrington, M. & Field, M.C. (2004). Cytoplasmic targeting signals in transmembrane invariant surface glycoproteins of trypanosomes. *The Journal of Biological Chemistry,* Vol. 279, pp. 54887–54895, ISSN 1067-8816

Claverie, J.M. & Notredame, C. (2006). *Bioinformatics for Dummies* (2nd ed). Wiley
          Publishing, Inc. ISBN: 978-0-470-08985-9 Indianapolis, IN

Cole, C., Barber, J.D. & Barton, G.J. (2008). The Jpred 3 secondary structure prediction
          server. *Nucleic Acids Research*, Vol. 35, No. suppl. 2, pp. W197-W201, ISSN 0305-1048

Coppens, I. & Joiner, K.A. (2003). Host but not parasite cholesterol controls *Toxoplasma* entry
          by modulating organelle discharge. *Molecular Biology of the Cell,* Vol. 14, pp. 3804-
          3820, ISSN 1059-1524

Cuff, J.A. & Barton, G.J. (1999). Evaluation and improvement of multiple sequence methods
          for protein secondary structure prediction. *Proteins,* Vol. 34, No. 4, pp. 508-519,
          ISSN 0887-3585

Curtiss, M., Jones, C. & Babst, M. (2007). Efficient cargo sorting by ESCRT-I and the
          subsequent release of ESCRT-I from multivesicular bodies requires the subunit
          Mvb12. *Molecular Biology of the Cell,* Vol. 18, No. 2, pp. 636-645, ISSN 1059-1524

Dacks, J.B. & Field MC. (2007). Evolution of the eukaryotic membrane-trafficking system:
          origin, tempo and mode. *Journal of Cell Science*, Vol. 120, pp. 2977–2985, ISSN 0021-
          9533

Dacks, J.B., Poon, P.P. & Field, M.C. (2008). Phylogeny of endocytic components yields
          insight into the process of non-endosymbiotic organelle evolution. *Proceedings of the
          National Academy of Sciences of the United States of America*, Vol. 105, pp. 588–593,
          ISSN 0027-8424

de Souza, W., Sant'Anna, C. & Cunha-e-Silva, N.L. (2009). *Progress in Histochemistry and
          Cytochemistry,* Vol. 44, No. 2, pp. 67-124, ISSN 0079-6336

Dimaano, C., Jones, C.B., Hanono, A., Curtiss, M. & Babst, M. (2008). Ist1 regulates Vps4
          localization and assembly. *Molecular Biology of the Cell,* Vol. 19, No. 2, pp. 465-474,
          ISSN 1059-1524

Field, M.C., Gabernet-Castello, C. & Dacks, J.B. (2007). Reconstructing the evolution of the
          endocytic system: insights from genomics and molecular cell biology. *Advances in
          experimental medicine and biology*, Vol. 607, pp. 84–96, ISSN 0065-2598

Fujita, H., Yamanaka, M., Imamura, K., Tanaka, Y., Nara, A., Yoshimori, T., Yokota, S. &
          Himeno, M. (2003). A dominant negative form of the AAA ATPase SKD1/VPS4
          impairs membrane trafficking out of endosomal/lysosomal compartments: class E
          vps phenotype in mammalian cells. *Journal of Cell Science,* Vol. 116, Pt 2, pp. 401-
          414, ISSN 0021-9533

García-Rivera, G., Rodríguez, M.A., Ocádiz, R., Martínez-López, M.C., Arroyo, R., González-
          Robles, A. & Orozco, E. (1999). *Entamoeba histolytica*: a novel cysteine protease and
          an adhesin form the 112 kDa surface protein. *Molecular Microbiology,* Vol. 33, No. 3,
          pp. 556-68, ISSN 0950-382X

Geoff, H. (1997). The Molecular Epidemiology of Parasites, In: *Principles of Medical Biology,
          Microbiology*, Edward Bittar (ed), pp. 597-614, JAI Press Inc., ISBN: 1-55938-814-5.
          Greenwich, Conn

Ghedin, E., Debrabant, A., Engel, J.C. & Dwyer, D.M. (2001). Secretory and endocytic
          pathways converge in a dynamic endosomal system in a primitive protozoan.
          Traffic, Vol, 2, pp. 175–188, ISSN 1398-9219

Gietz, R.D., Triggs-Raine, B., Robbins, A., Graham, K. & Woods, R. (1997). Identification of
          proteins that interact with a protein of interest: Applications of the yeast two-

hybrid system. *Molecular and Cellular Biochemistry*, Vol. 172, No. 1-2, pp. 67–79, ISSN 0300-8177

Henikoff, S. & Henikoff, J.G. (2000). Amino acid substitution matrices. *Advances in Protein Chemistry,* Vol, 54, pp. 73-97, ISSN 0065-3233

Hoppe, H.C., Ngo, H.M., Yang, M. & Joiner, K.A. (2000). Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nature Cell Biology,* Vol. 2, pp. 449-456, ISSN 1465-7392

http://align.genome.jp/

http://bips.u-strasbg.fr/fr/Tutorials/Comparison/Blast/blastall.html

http://eupathdb.org/eupathdb/

http://expasy.org/tools/#proteome

http://jmol.sourceforge.net/download/

http://swissmodel.expasy.org/

http://www.ch.embnet.org/software/BOX_form.html

http://www.pdb.org/pdb/home/home.do

http://www.sanger.ac.uk/resources/databases/pfam.html

http://www.uniprot.org/

Hurley, J.H. & Emr, S.D. (2006). The Escrt Complexes: Structure and Mechanism of a Membrane-Trafficking Network. *Annual Review of Biophysics Biomolecular Structure*, Vol. 35, pp. 277–298, ISSN 1056-8700

Hurley, J.H. & Hanson, P.I. (2010). Membrane budding and scission by the ESCRT machinery: it's all in the neck. *Nature Reviews. Molecular Cell Biology,* Vol. 11, No. 8, pp. 556-566, ISSN 1471-0072

Kaltenbach, L.S., Romero, E., Becklin, R.R., Chettier, R., Bell, R., Phansalkar, A., Strand, A., Torcassi, C., Savage, J., Hurlburt, A., Cha, G.H., Ukani, L., Chepanoske, C.L., Zhen, Y., Sahasrabudhe, S., Olson, J., Kurschner, C., Ellerby, L.M., Peltier, J.M., Botas, J. & Hughes, R.E. (2007) Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genetics,* Vol. 3, No. 5, pp. e82, ISSN 1553-7390

Kanazawa, C., Morita, E., Yamada, M., Ishii, N., Miura, S., Asao, H., Yoshimori, T., & Sugamura, K. (2003). Effects of deficiencies of STAMs and Hrs, mammalian class E Vps proteins, on receptor downregulation. *Biochemical Biophysical Research Communication*, Vol. 309, No. 4, pp. 848-856, ISSN 0006-291X

Katzmann, D.J., Babst, M., & Emr, S.D. (2001). Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell,* Vol. 106, pp. 145–155, ISSN 0092-8674

Kelley, L.A. & Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols,* Vol. 4, No. 3, pp. 363-371, ISSN 1754-2189

Kim, J., Sitaraman, S., Hierro, A., Beach, B.M., Odorizzi, G. & Hurley, J.H. (2005). Structural basis for endosomal targeting by the Bro1 domain. *Developmental Cell,* Vol. 8, No. 6, pp. 937-947, ISSN 1534-5807

Kuchaiev, O. & Przulj, N. (2011). Integrative Network Alignment Reveals Large Regions of Global Network Similarity in Yeast and Human. *Bioinformatics.* Vol. Mar 16, [Epub ahead of print], ISSN 1367-4803

Lanfredi-Rangel, A., Attias, M., de Carvalho, T.M., Kattenbach, W.M. & de Souza, W. (1998). The peripheral vesicles of trophozoites of the primitive protozoan *Giardia lamblia*

may correspond to early and late endosomes and to lysosomes. *Journal of Structural Biology,* Vol. 123, pp. 225–235, ISSN 1047-8477

Leung, K.F., Dacks, J.B. & Field, M.C. (2008). Evolution of the Multivesicular Body ESCRT Machinery; Retention Across the Eukaryotic Lineage. *Traffic*, Vol. 9, pp. 1698–1716, ISSN 1398-9219

Li, Y., Kane, T., Tipper, C., Spatrick, P. & Jenness, D.D. (1999). Yeast mutants affecting possible quality control of plasma membrane proteins. *Molecular Cell Biology,* Vol. 19, No. 5, pp. 3588-3599, ISSN 1471-0072

López-Reyes, I., García-Rivera, G., Bañuelos, C., Herranz, S., Vincent, O., López-Camarillo, C., Marchat, L.A, & Orozco, E. (2010). Detection of the endosomal sorting complex required for transport in *Entamoeba histolytica* and characterization of the EhVps4 protein. *Journal of Biomedicine & Biotechnology,* Vol. 2010, pp. 890674, ISSN 1110-7243

Misra, S. & Hurley, J.H. (1999). Crystal structure of a phosphatidylinositol 3-phosphate-specific membrane-targeting motif, the FYVE domain of Vps27p. *Cell*, Vol. 97, No. 5, pp. 657-666, ISSN 0092-8674

Nakada-Tsukui, K., Okada, H., Mitra, B.N. & Nozaki, T. (2009) Phosphatidylinositol-phosphates mediate cytoskeletal reorganization during phagocytosis via a unique modular protein consisting of RhoGEF/DH and FYVE domains in the parasitic protozoan *Entamoeba histolytica*. *Cellular Microbiology*, Vol. 11, No. 10, pp. 1471-1491, ISSN 1462-5814

Obita, T., Saksena, S., Ghazi-Tabatabai, S., Gill, D.J., Perisic, O., Emr, S.D. & Williams, R.L. (2007). Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4. *Nature*, Vol. 449, pp. 735–739, ISSN 0028-0836

Odorizzi, G., Katzmann, D.J., Babst, M., Audhya, A. & Emr, S.D. (2003). Bro1 is an endosome-associated protein that functions in the MVB pathway in *Saccharomyces cerevisiae*. *Journal of Cell Science*, Vol. 116, pp. 1893–1903, ISSN 0021-9533

Offer, G., Hicks, M.R. & Woolfson, D.N. (2002). Generalized Crick equations for modeling noncanonical coiled coils. *Journal of Structural Biology*, Vol. 137, No. 1-2, pp. 41-53, ISSN 1047-8477

Pandeya, S.N. & Thakkar, D. (2005). Combinatorial chemistry: A novel method in drug discovery and its application. *Indian Journal of Chemistry,* Vol. 44B, pp. 335-348, ISSN 0019-5103

Peng, Z.Y. & Kim, P.S. (1994). A protein dissection study of a molten globule. *Biochemistry,* Vol. 33, No. 8, pp. 2136-2141, ISSN 0006-2960

Pornillos, O., Alam, S.L., Rich, R.L., Myszka, D.G., Davis, D.R. & Sundquist, W.I. (2002). Structure and functional interactions of the Tsg101 UEV domain. *The EMBO Journal,* Vol. 21, No. 10, pp. 2397-2406, ISSN 0261-4189

Shiflett, S.L., Ward, D.M., Huynh, D., Vaughn, M.B., Simmons, J.C. & Kaplan, J. (2004) Characterization of Vta1p, a class E Vps protein in *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry,* Vol. 279 No. 12, pp. 10982-10990, ISSN 0021-9258

Sinha, A., Mandal, S., Banerjee, S., Ghosh, A., Ganguly, S., Sil, A.K. & Sarkar, S. (2010). Identification and Characterization of a FYVE Domain from the Early Diverging Eukaryote *Giardia lamblia*. *Current Microbiology*, Vol. Dec 17, [Epub ahead of print], ISSN 0343-8651

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics*. Vol. 21, No. 7, pp. 951-60, ISSN 1367-4803

Stahelin, R.V., Long, F., Diraviyam, K., Bruzik, K.S., Murray, D. & Cho, W. (2002). Phosphatidylinositol 3-phosphate induces the membrane penetration of the FYVE domains of Vps27p and Hrs. *The Journal of Biological Chemistry,* Vol. 277, pp. 26379, ISSN 0021-9258

Sundquist, W.I., Schubert, H.L., Kelly, B.N., Hill, G.C., Holton, J.M. & Hill, C.P. (2004). Ubiquitin recognition by the human TSG101 protein. *Molecular Cell,* Vol. 13, No. 6, pp. 783-789, ISSN 1097-2765

Tse, Y.C., Mo, B., Hillmer, S., Zhao, M., Lo, S.W., Robinson, D.G. & Jiang, L. (2004). Identification of multivesicular bodies as prevacuolar compartments in *Nicotiana tabacum* BY-2 cells. *The Plant Cell*, Vol. 16, pp. 672–693, ISSN 1040-4651

Whitley, P., Reaves, B.J., Hashimoto, M., Riley, A.M., Potter, B.V. & Holman, G.D. (2003). Identification of mammalian Vps24p as an effector of phosphatidylinositol 3,5-bisphosphate-dependent endosome compartmentalization. *The Journal of Biological Chemistry*, Vol. 278, No. 40, pp. 38786-38795, ISSN 0021-9258

Williams, R.L. & Urbé, S. (2007). The emerging shape of the ESCRT machinery. *Nature Reviews. Molecular Cell Biology*, Vol. 8, pp. 355-368, ISSN 1471-0072

Wöstmann, C., Liakopoulos, D., Ciechanover, A. & Bakker-Grunwald, T. (1996) Characterization of ubiquitin genes and -transcripts and demonstration of a ubiquitin-conjugating system in *Entamoeba histolytica. Molecular and Biochemical Parasitology*, Vol. 82, No.1, pp. 81-90, ISSN 0166-6851

Yang, M., Coppens, I., Wormsley, S., Baevova, P., Hoppe, H.C., & Joiner, K.A. (2004). The *Plasmodium falciparum* Vps4 homolog mediates multivesicular body formation. *Journal of Cell Science*, Vol. 117, Pt 17, pp. 3831-3838, ISSN 0021-9533