# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

186,000

200M

Downloads

154

Our authors are among the

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us? Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



# Investigation on Nuclear Transport of Trypanosoma brucei: An in silico Approach

Mohd Fakharul Zaman Raja Yahya, Umi Marshida Abdul Hamid and Farida Zuraina Mohd Yusof

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/47448

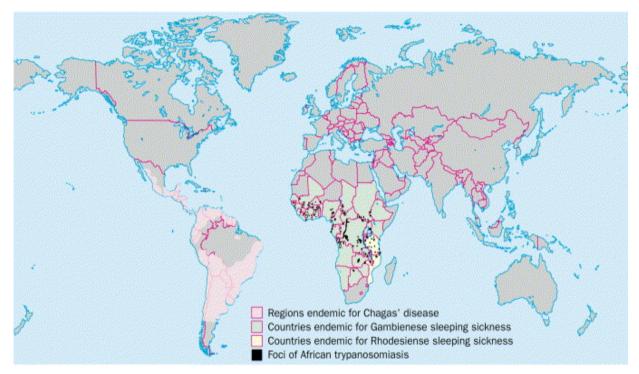
#### 1. Introduction

#### 1.1. Trypanosomiasis

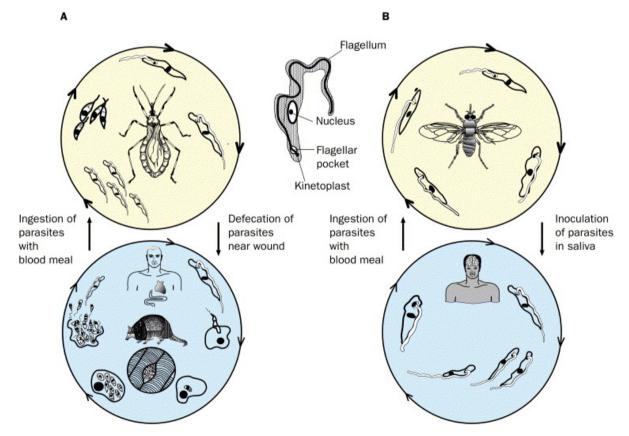
A group of animal and human diseases caused by parasitic protozoan trypanosomes is called trypanosomiases. The final decade of the 20th century witnessed a frightening revival in sleeping sickness (human African trypanosomiasis) in sub-Saharan Africa. Meanwhile, Chagas' disease (American trypanosomiasis) remains one of the most widespread infectious diseases in South and Central America. Arthropod vectors are responsible for the spread of African and American trypanosomiases, and disease restraint through insect control programs is an attainable target. However, the existing drugs for both illnesses are far from ideal. The trypanosomes are some of the earliest diverging members of the Eukaryotae and share several biochemical oddities that have inspired research into discovery of new drug targets. Nevertheless, discrepancies in mode of interactions between trypanosome species and their hosts have spoiled efforts to design drugs effective against both species. Heightened awareness of these neglected diseases might result in progress towards control through increased financial support for drug development and vector eradication [1].

Trypanosome is a group of unicellular parasitic flagellate protozoa which mostly infects the vertebrate genera. A number of trypanosome species cause important veterinary diseases, but only two cause significant human diseases. In sub-Saharan Africa, *Trypanosoma brucei* causes sleeping sickness or human African trypanosomiasis whilst in America, *Trypanosoma cruzi* causes Chagas' disease (Figure 1) [2]. Meanwhile, the life cycle of these parasitic protozoa engage insect vectors and mammalian hosts (Figure 2) [1]. All trypanosomes require more than one obligatory host to complete their life cycle and are transmitted via vectors. Most of the species are transmitted by blood-feeding invertebrates, however there





**Figure 1.** Geographic distribution of *Trypanosoma brucei* and *Trypanosoma cruzi*, showing endemic countries harboring these diseases [2].



**Figure 2.** Life cycles of (A) *Trypanosoma cruzi* and (B) *Trypanosoma brucei*. Upper cycles represent different stages that take place in the insect vectors. Lower cycles represent different stages in man and other mammalian hosts [1].

are distinct mechanisms among the varying species. In the invertebrate hosts they are generally found in the intestines as opposed to the bloodstream or any other intracellular environment in the mammalian host. As trypanosomes develop through their life cycle, they undergo a series of morphological changes [3] as is typical of trypanosomatids.

The life cycle often consists of the trypomastigote form in the vertebrate host and the trypomastigote or promastigote form in the gut of the invertebrate host. Intracellular lifecycle stages are normally found in the amastigote form. The trypomastigote morphology is unique to species in the genus Trypanosoma.

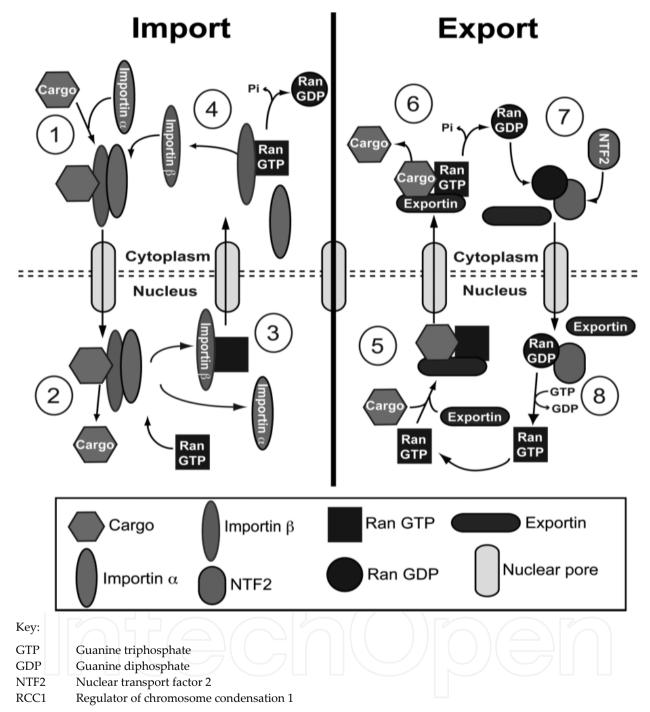
The genome organization of *T. brucei* is splitted into nuclear and mitochondrial genomes. The nuclear genome of *T. brucei* is made up of three classes of chromosomes according to their size on pulsed-field gel electrophoresis, large chromosomes (1 to 6 megabase pairs), intermediate chromosomes (200 to 500 kilobase pairs) and mini chromosomes (50 to 100 kilobase pairs) [4]. The large chromosomes contain most genes, while the small chromosomes tend to carry genes involved in antigenic variation, including the variant surface glycoprotein (VSG) genes. Meanwhile, the mitochondrial genome of the Trypanosoma, as well as of other kinetoplastids, known as the kinetoplast, is characterized by a highly complex series of catenated circles and minicircles and requires a cohort of proteins for organisation during cell division. The genome of *T. brucei* has been completely sequenced and is now available online [5].

# 1.2. Nuclear transport

Nuclear transport of proteins and ribonucleic acids (RNAs) between the nucleus and cytoplasm is a key mechanism in eukaryotic cells [6]. The transport between the nucleus and cytoplasm involves primarily three classes of macromolecules: substrates, adaptors, and receptors. The transport complex is formed when the substrates bind to an import or an export receptor. Some transport substrates require one or more adaptors to mediate formation of a transport complex. Once assembled, these transport complexes are transferred in one direction across the nuclear envelope via aqueous channels that are part of the nuclear pore complexes (NPCs). Following dissociation of the transport complex, both adaptors and receptors are recycled through the NPC to allow another round of transport to occur. Directionality of either import or export therefore depends on the formation of receptor-substrate complex on one side of the nuclear envelope and the dissociation of the complex on the other. The Ran GTPase is vital in producing this asymmetry. Modulation of nuclear transport generally involves specific inhibition of the formation of a transport complex, however, more global forms of regulation also occur [7]. The general concept of import and export process is shown in Figure 3 [8].

#### 1.3. *In silico* approach

In silico study is defined as an analysis which is performed using computer or via computer simulation. It involves the strategy of managing, mining, integrating, and interpreting



**Figure 3.** For import of molecules, cytoplasmic cargo is identified by Importin *a*, which then binds to Importin b (1). This ternary complex translocates through the nuclear membrane and into the nucleus. Once there, RanGTP binds to Importin b and causes a dissociation of the complex, which releases cargo to the nucleus (2). Import receptors are then recycled back to the nucleus (3) through binding of RanGTP and export to the cytosol. RanGTP is then hydrolyzed to the GDP-bound state and causes the release of the import receptors (4) and the cycle starts over again. Export of cargo undergoes a similar mechanism. Exported molecules will bind to the export receptor with RanGTP and exit the nucleus (5). Next RanGTP is hydrolyzed to cause release of cargo into the cytoplasm (6). NTF2 specifically identifies RanGDP and returns it to the nucleus (7) for RCC1 to then exchange it to RanGTP (8) [8].

information from biological data at the genomic, metabalomic, proteomic, phylogenetic, cellular, or whole organism levels. The bioinformatics instruments and skills become crucial for in silico research as genome sequencing projects have resulted in an exponential growth in protein and nucleic acid sequence databases. Interaction among genes that gives rise to multiprotein functionality generates more data and complexity. In silico approach in medicine is not only reducing the need for expensive lab work and clinical trials but also is possible to speed the rate of drug discovery. In 2010, for example, researchers found potential inhibitors to an enzyme associated with cancer activity in silico using the protein docking algorithm EADock [9]. About 50 % of the molecules were later shown to be active inhibitors in vitro [9]. A unique advantage of the in silico approach is its worldwide accessibility. In some cases, having internet access or even just a computer is sufficient enough. Laboratory experiments either in vivo or in vitro both require more materials. In protein sequence analysis, in silico approach gives highly reproducible results in many cases or even exactly the same results because it only relies on comparison of the query sequence to a database of previously annotated sequences. However, in sophisticated analysis such as development of the 3-D structure of proteins from their primary sequences, discrepancies in results are to be expected due to the manual optimization which must consider several crucial steps such as template selection, target-template alignment, model construction and model evaluation.

#### 1.4. Problem statements

Considering the importance of nuclear shuttling in many cellular processes, proteins responsible for the nuclear transport are vital for parasite survival. The presence of nuclear transport machinery was highlighted in the eukaryotic parasites such as Plasmodium falciparum, Toxoplasma gondii and Cryptosporidium parvum. However, the nuclear transport in T. brucei has not been established. Nuclear shuttling is one of the overlooked aspects of drug design and delivery. Exploitation of macromolecules movement across the nuclear envelope promises to be an exciting area of drug development. Furthermore, the divergence between host and parasite systems is always exploited as a strategy in drug development. Therefore, the exploitation of peculiarities of T. brucei nuclear transport machinery as compared to its host might be a promising strategy for the control of trypanosomiasis, which remains to be further investigated.

#### 1.5. Objectives

This study is carried out to investigate the nuclear transport constituents of *T. brucei* by determining the functional characteristics of the parasite proteins. This includes functional protein domain, post translational modification sites and protein-protein interaction. The parasite proteins identified to exhibit the relevant functional protein domains, post translational modification sites and protein-protein interaction, are predicted as the true components for nuclear transport mechanism. This study also aims to evaluate the unique characteristics of proteins responsible for nuclear transport machinery between the parasites and human by determining the degree of protein sequence similarity. The information on the sequence level divergence between T. brucei proteins and their human counterparts may provide an insight into drug target discovery.

#### 2. Materials and methods

Our in silico analyses were carried out using the public databases and web based programs (Table 1). The programs were employed to identify and annotate the parasite proteins involved in the nuclear transport mechanism. The identified parasite proteins were then compared with the human counterparts.

Analysis	Programme name	URL and Reference where available	
Protein sequence retrieval	National Centre for Biotechnology Information (NCBI)	www.ncbi.nlm.nih.gov/	
	Universal Protein		
	Knowledgebase/SwissProt (UniProtKB/ SwissProt)	http://www.uniprot.org/	
	TriTrypDB	http:// tritrypdb.org/ tritrypdb/	
Clustering of protein sequences Identification of protein domains	BLASTClust	www.vardb.org/vardb/analysis/blastclust.html	
	Conserved Domain Database (CDD)	http://www.ncbi.nlm.nih.gov/cdc	
	Simple Modular Architecture Research Tool (SMART)	http://smart.embl-heidelberg.de/	
	InterPro	http://www.ebi.ac.uk/interpro/	
Identification of			
post translational modification sites	PROSITE	http://prosite.expasy.org/	
Sequence similarity search	BLASTp (NCBI)	http://blast.ncbi.nlm.nih.gov/	

**Table 1.** Databases and web-based programs used in the analysis of nuclear transport of *T. brucei*.

We utilized a personal computer equipped with AMD Turion 64x2 dual-core processor, memory size of 32 gigabytes and NVIDIA graphics card to perform the analyses. Our in silico work is summarized in Figure 4.

The nuclear transport refers to a process of entry and exit of large molecules from the cell nucleus. To identify T. brucei proteins of nuclear transport, the protein sequences of other various eukaryotic organisms were retrieved in FASTA format from National Centre for Biotechnology Information (NCBI) server and Universal Protein Knowledgebase/SwissProt (UniProtKB/ SwissProt) database based on biological processes and protein name search. The number of hits obtained for the query was recorded after manual inspection. The retrieved protein sequences were clustered into groups with more than 30% similarity using

BLASTClust [10] to reduce non-redundant protein sequences. The non-redundant data set was subjected to BLASTp [11] analyses against an integrated genomic and functional genomic database for eukaryotic pathogens of the family Trypanosomatidae, TriTrypDB. The analysis was using cutoff point with E-value of less than 1e-06 and score of more than 100. Hits that pointed to the same location or overlapped location were removed manually. The identified protein sequences then were then retrieved from the TriTrypDB.

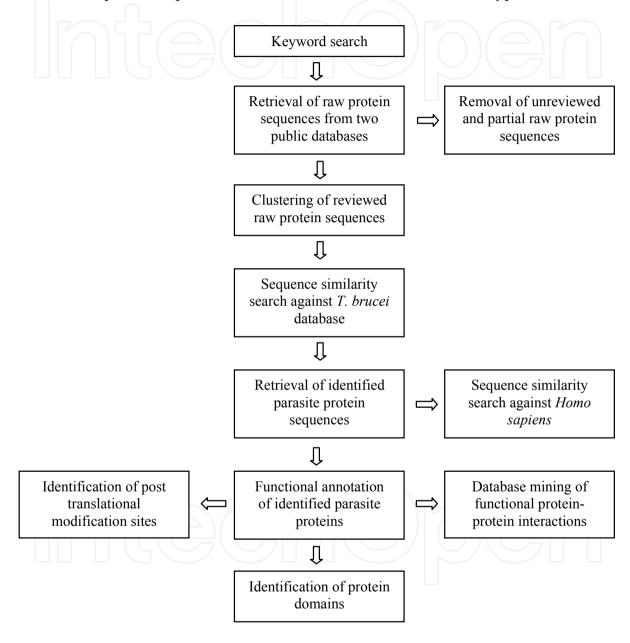


Figure 4. In silico analysis workflow.

A portion of protein that can evolve, function, and exist independently is called protein domain. It is a compact three dimensional structure, stable and distribution of polar and non-polar side chains contribute to its folding process. To determine the functional protein domains, all identified protein sequences of T. brucei from TriTrypDB were subjected to

functional annotation which makes use of Conserved Domain Database (CDD) [12], Simple Modular Architecture Research Tool (SMART) [13] and InterPro [14] programs. The protein sequences were submitted in FASTA format as queries.

Posttranslational modification (PTM) is the chemical modification of a protein after its translation. It is one of the later steps in protein biosynthesis, and thus gene expression, for many proteins. In this part of study, in relation to regulatory aspects of nuclear transport mechanism, we focused on potential glycosylation and phosphorylation sites. To analyze the post translational modification sites, all protein sequences of *T. brucei* from TriTrypDB were subjected to PROSITE [15] programme. The proteins sequences were submitted in FASTA format as queries.

Protein-protein interactions occur when two or more proteins bind together, often to carry out their biological function. Proteins might interact for a long time to form part of a protein complex, a protein may be carrying another protein, or a protein may interact briefly with another protein just to modify it. To analyze the participation of parasite proteins in proteinprotein interactions, all protein sequences of T. brucei from TriTrypDB were subjected to mining of STRING 8.2 database [16]. The STRING 8.2 database integrates information from numerous sources, including experimental repositories, computational prediction methods and public text collections. The proteins sequences were submitted in FASTA format as queries. All information on protein-protein interaction were recorded and evaluated accordingly.

The degree of similarity between amino acids occupying a particular position in the protein sequence can be interpreted as a rough measure of how conserved a particular region or sequence motif is. To compare the parasite proteins with human homologues, all protein sequences of T. brucei from TriTrypDB were subjected to BLASTp analysis against Homo sapiens proteins. The proteins sequences were submitted in FASTA format as queries. The criteria such as cutoff point with E-value of less than 1e-06 and score of more than 100 were used.

#### 3. Results and discussions

### 3.1. Parasite proteins involved in the nuclear transport machinery

Table 2 shows a summary of protein sequences used in this in silico analysis. A total of 904 and 642 protein sequences were retrieved in FASTA format from NCBI server and UniProt/SwissProt database respectively. A total of 18 protein sequences with less than 100 amino acid residues were excluded from the study as they were considered not completely functional [17]. Hence, 1528 protein sequences were used for protein sequence clustering. The 30% identity and above at the amino acid level is considered sufficient to imply functional relatedness [17]. Therefore, protein clustering with more than 30% similarity on the retrieved protein sequences produced a non-redundant data set of 248 protein sequences.

Protein sequences	Total
Raw protein sequences retrieved from NCBI and UniProtKB	1546
Raw protein sequences subjected to BLASTClust programme	1548
Non redundant protein sequences resulting from BLASTClust analysis	248
Query sequences for BLASTp analysis against TritrypDB database	248

**Table 2.** Summary of protein sequences retrieved in *in silico* analysis.

The BLASTp analyses against TriTrypDB using cut off point with E-value of less than 1e-06 and score of more than 100 for the whole 248 query protein sequences resulted in 34 hits of parasite proteins. However our approach failed to identify a Ran GTPase-activating protein (RanGAP) protein in this parasite. In reference [18] also reported that sequence similarity searches have been unable to identify a RanGAP protein in any protozoan. Keyword searches among annotated proteins in the T. gondii genome database identified one candidate which was shown to have strong similarity to Ran-binding protein 1 (RanBP1) based on sequence analysis. Perhaps the RanGAP function in apicomplexans is performed by a single protein with multiple cellular responsibilities (i.e., a fusion of Ran binding protein 1 and RanGAP). It is also possible that a completely unique parasite protein possesses the RanGAP function.

Table 3 shows the identified and characterized parasite proteins involved in the nuclear transport machinery. The functional annotation based on protein domains, showed that, out of 34, only 22 parasite protein sequences were predicted with high confidence level to be involved in the nuclear transport mechanism with the presence of relevant protein domains. This includes guanine triphosphate (GTP)-binding domain, Nucleoporin (NUP) C terminal domain, Armadillo repeat, Importin B N-terminal domain, regulator of chromosome condensation 1 (RCC1) repeat and Exportin domain (Table 4). All these protein domains were experimentally verified to regulate the nuclear transport mechanism in eukaryotes. There were seven *T. brucei* proteins that exhibited functional features of the Importin receptor. This finding is consensus with the number of Importin receptors in another eukaryotic pathogen, Toxoplasma gondii [8]. In addition, our results of other nuclear transport constituents in *T. brucei* such as RCC1, Ran, nuclear transport factor 2 (NTF2), cell apoptosis susceptibility (CAS), Exportin and Ran binding proteins were also in agreement with reference [18].

The nuclear and cytoplasmic compartments are divided by the nuclear envelope in eukaryotes. By using this compartmentalization and controlling the movement of molecules between the nucleus and the cytosol, cells are able to regulate numerous cellular mechanisms such as transcription and translation. Proteins with molecular size lower than 40 kDa are able to passively diffuse through the nuclear pore complex (NPC), whereas larger proteins require active transport through the assistance of Karyopherins, specific transport receptors that shuttle between the nucleus and cytosol. Karyopherins which are able to distinguish between the diverse proteome to target specific cargo molecules for transport, can be subdivided into those that transport molecules into the nucleus (Importins) and those that transport molecules out of the nucleus (Exportins). It has been reported that

#### 40 Bioinformatics

more than 2000 proteins are shuttled between the nucleus and the cytoplasm in yeast [19]. From our result, with the identification of Karyopherin and Nucleoporin proteins in *T. brucei*, we expect that the parasite employs the typical components for the nuclear transport machinery.

Subject sequences	E-value	Score	Functional protein domains
Tb927.3.1120	1.70E-72	718	Ran GTPase, GTP-binding domain
Tb09.211.4360	5.50E-33	348	Karyopherin Importin Beta, Armadillo repeat
Tb11.01.5940	9.30E-149	1391	Exportin-1 C terminal, Importin Beta N terminal domain
Tb11.02.0870	3.20E-16	187	Ran binding domain
Tb927.2.2240	2.40E-15	190	Exportin-like protein
Tb927.6.2640	9.10E-83	815	Karyopherin Importin Beta, Armadillo repeat
Tb927.6.4740	1.10E-75	748	CAS/CSE domain, Importin Beta N terminal domain
Tb927.7.1190	6.90E-20	172	RCC1 repeat
Tb11.03.0140	5.80E-09	107	NUP C terminal domain
Tb927.10.8170	2.10E-28	315	NUP C terminal domain
Tb927.8.3370	2.50E-48	281	Ran-binding protein Mog1p
Tb11.01.7010	8.20E-42	464	Armadillo repeat, Karyopherin Importin Beta
Tb11.02.1720	2.60E-26	276	Armadillo-like helical
Tb11.01.8030	1.70E-18	218	HEAT repeat, Armadillo repeat, Importin Beta N terminal domain
Tb11.01.7200	7.10E-07	137	Nsp1-like
Tb927.7.6320	1.20E-11	136	RCC1 repeat
Tb927.3.4600	3.70E-08	149	Armadillo-like helical
Tb09.160.2360	1.40E-36	379	WD40 repeat
Tb927.6.3870	8.50E-14	164	RNA recognition motif
Tb927.7.5760	1.30E-08	115	Nuclear transport factor 2 domain
Tb10.70.4720	4.60E-77	761	Importin Beta N terminal domain, Karyopherin domain
Tb927.8.4280	2.90E-08	112	Nuclear transport factor 2 domain

Key:	
GTP	Guanine triphosphate
CAS	Cell apoptosis susceptibility

CSE Chromosome seggregation

RCC1 Regulator of chromosome condensation 1

NUP Nucleoporin

HEAT Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast PI3-kinase TOR1

WD Trp-Asp (W-D) dipeptide

RNA Ribonucleic acid

**Table 3.** Identified and characterized *T. brucei* proteins of nuclear transport. Protein domain identification involved CDD, SMART, InterPro and PROSITE programs.

Protein domair	Accession	Description
Ran GTPase	SM00176	Ran is involved in the active transport of proteins through nuclear pores.
Ran binding domain	PDOC50196	This domain binds RanGTP and increases the rate of RanGAP1-induced GTP hydrolysis.
Armadillo	IPR000225	The Armadillo (Arm) repeat is an approximately 40 amino acid long tandemly repeated sequence motif first identified in the <i>Drosophila melanogaster</i> segment polarity gene armadillo involved in signal transduction through wingless. Animal Arm-repeat proteins function in various processes, including intracellular signalling and cytoskeletal regulation, and include such proteins as beta-catenin, the junctional plaque protein plakoglobin, the adenomatous polyposis coli (APC), tumour suppressor protein, and the nuclear transport factor
		importin-alpha, amongst others
Importin beta	IPR001494	Members of the Importin-beta (Karyopherin-beta) family can bind and transport cargo by themselves, or can form heterodimers with importin-alpha. As part of a heterodimer, Importin-beta mediates interactions with the pore complex, while Importin-alpha acts as an adaptor protein to bind the nuclear localisation signal (NLS) on the cargo through the classical NLS import of proteins.
HEAT	IPR000357	Arrays of Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast PI3-kinase TOR1 (HEAT) repeats consists of 3 to 36 units forming a rod-like helical structure and appear to function as protein-protein interaction surfaces. It has been noted that many HEAT repeat-containing proteins are involved in intracellular transport processes.
Exportin 1-like protein	pfam08389	The sequences featured in this family are similar to a region close to the N-terminus of yeast exportin 1 (Xpo1, Crm1). This region is found just C-terminal to an importin-beta N-terminal domain (pfam03810) in many members of this family. Exportin 1 is a nuclear export receptor that interacts with leucine-rich nuclear export signal (NES) sequences, and Ran-GTP, and is involved in translocation of proteins out of the nucleus.

**Table 4.** Summary of protein domains

# 3.2. Regulatory aspect of the parasite nuclear transport

Nucleoporins.

Table 5 shows the presence of phosphorylation and glycosylation sites in the parasite proteins. The phosphorylation sites were found to be present in all parasite proteins. It was

predicted that the parasite proteins could be phosphorylated at Serine, Threonine and Tyrosine amino residues. However, the O-glycosylation sites were not present in three parasite proteins, namely Tb11.02.0870, Tb927.8.3370 and Tb927.7.5760.

Subject sequences	Phosphorylation site	Glycosylation site
Tb927.3.1120	Π+	+
Tb09.211.4360	+ ( )	+ 10
Tb11.01.5940	<u> </u>	
Tb11.02.0870		
Tb927.2.2240	+	+
Tb927.6.2640	+	+
Tb927.6.4740	+	+
Tb927.7.1190	+	+
Tb11.03.0140	+	+
Tb927.10.8170	+	+
Tb927.8.3370	+	-
Tb11.01.7010	+	+
Tb11.02.1720	+	+
Tb11.01.8030	+	+
Tb11.01.7200	+	+
Tb927.7.6320	+	+
Tb927.3.4600	+	+
Tb09.160.2360	+	+
Tb927.6.3870	+	+
Tb927.7.5760	+	-
Tb10.70.4720	П+	+
Tb09.211.2550	+	
Tb927.8.4280		

**Table 5.** Phosphorylation and O-glycosylation sites in the *T. brucei* proteins. Identification of these functional sites involved ScanProsite programme.

Most of the parasite proteins were predicted to be involved in O-linked glycosylation. In eukaryotes, the O-linked glycosylation takes place in the Golgi apparatus. It also occurs in archaea and bacteria. Phosphorylation was reported to be crucial in the regulation of protein-protein interactions of the NADPH oxidase in the phagocytic cells [20]. The phosporylation-based signaling in T. brucei has been reported by reference [21]. Thus we

<sup>(+)</sup> indicates presence

<sup>(-)</sup> indicates absence

believe that the phosphorylation could also regulate the nuclear transport components of *T. brucei* to participate in various functional interactions. Meanwhile, it was suggested that O-linked glycosylation may be analogous to protein phosphorylation. According to [22], phosphorylation by proline-directed kinases share the same sites with those potentially O-glycosylated by O-linked N-acetylglucosamine transferase (OGT). From this it is possible that O-glycosylation and phosphorylation may compete for sites of modification. Therefore, it is a strong likelihood that the nuclear transport of *T. brucei* could be regulated by both phosphorylation and O-glycosylation.

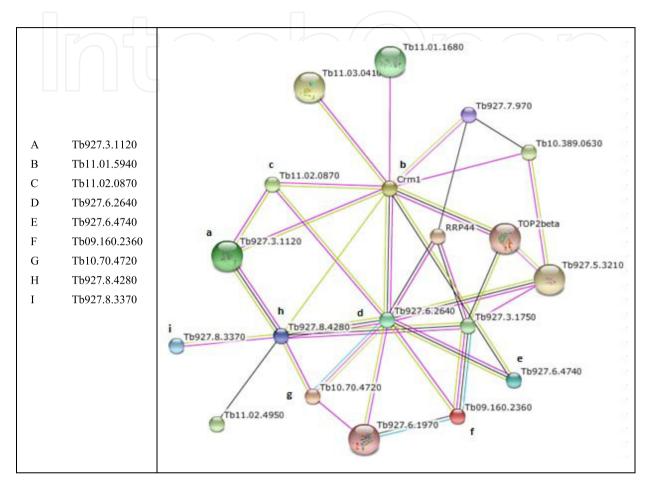
Apart from acting simply as an architectural structure which facilitates nuclear transport, the NPC may also play a more dynamic role in regulating transport. The specificity of import and export may be influenced by recognition of different substrates and alteration of the Nucleoporin expression. This would allow different interaction between the NPC and Karyopherins and modulate the nuclear import and export. However, the most common impact on nucleocytoplasmic movement stems comes from the post translational modifications of the cargo proteins themselves [23]. The post translational modification of NPC was reported by [24]. Post-translational modification of NUPs by ubiquitylation and phosphorylation can affect NUP turnover and pore disassembly, respectively. Our study identified four parasite proteins containing the Nucleoporin-related domain. We anticipate that the assembly and disassembly of the parasite Nucleoporin proteins might also be modulated by phosphorylation.

The NPC becomes an ideal target for inhibition of nuclear import or export. One of the most common features of Nucleoporins is the presence of conserved FG or FXFG repeats that bind to the Importin family members [25]. The monoclonoal antibodies such as mAb414 and RL2 can interrupt translocation through the NPC by blocking the FG and FXFG epitopes of the Nucleoporins. Consequently, several Nucleoporin proteins were identified by their reactivity against the anti-FG antibodies. Most of these FG repeat proteins exist as the cytoplasmic fibrils or projections on the nuclear side of the NPC. The monoclonal antibodies prevent cargo from associating with the edge of an NPC so it cannot cross the membrane [26]. Thus, there is a possibility that the pathogenesis of *T. brucei* could be controlled by inhibiting its Nucleoporin proteins.

# 3.3. Participation of parasite proteins in functional interaction network

Figure 5 illustrates the protein interaction data obtained from STRING 8.2 database. The mining of protein interaction data which is useful in contextual annotation of protein function showed that, out of 22 parasite homologues, only nine parasite proteins were interacting with each other. Out of the seven identified *T. brucei* Importins, only two namely Tb927.6.2640 and Tb10.70.4720 were found to be involved in that protein interaction network. This database mining approach indicated that *T. brucei* nuclear transport is typical of eukaryotic organisms. Importins initially recruit cargo at low RanGTP concentrations in the cytoplasm and release cargo at high RanGTP levels in the nucleus. Importin–RanGTP complexes return afterwards to the cytoplasm, where the Ran-bound GTP is finally

hydrolysed and Ran dissociates from the receptor. The Importin can then bind and import another cargo molecule, while nuclear transport factor 2 (NTF2) recycles RanGDP back to nucleus. The cargo binding to exportins is controlled in a reverse manner compared to Importins; they recruit cargo at high RanGTP levels in the nucleus and release cargo at low RanGTP concentrations in the cytoplasm.



**Figure 5.** Protein functional interaction network in the nuclear transport of *T. brucei*. This protein interaction data was obtained from STRING 8.2 database. The letters (a-i) indicate the parasite proteins involved in the nuclear transport.

Table 5 shows evaluation of the obtained protein interaction data of the parasite nuclear transport. There were 13 functional interactions between parasite proteins identified from the mining of STRING 8.2 database. The score values of functional interactions range from 0.45 to 0.976. The Importin alpha (Tb927.6.2640) was found to be the most interactive parasite proteins by participating in six functional interactions. Based on the relevant protein domains and previous reports, four out of 13 functional interactions were considered with high confidence level. It should be emphasized that our approach only considered the protein interaction data derived from experiments, gene fusion and text mining. To our knowledge, this is the first report of functional protein interactions in the nuclear transport of the eukaryotic parasites. Whether other eukaryotic parasites share the common protein interaction network for the nuclear transport machinery remains to be elucidated.

Subject sequence	Interacting partner	Source	Score	Confidence level	Reference
Tb927.3.1120	Tb11.01.5940	Experiment	0.45	High	Lounsbury and Macara (1997)
Tb927.3.1120	Tb11.02.0870	Experiment, Text mining	0.512	Moderate	None
Tb927.3.1120	Tb927.8.4280	Experiment	0.534	High	Fried and Kutay (2003)
Tb11.01.5940	Tb11.02.0870	Experiment,Text mining	0.88	High	Lounsbury and Macara (1997)
Tb11.01.5940	Tb927.6.2640	Experiment,Text mining,Co-expression	0.812	Moderate	None
Tb11.01.5940	Tb927.6.4740	Text mining,Co- expression	0.46	Moderate	None
Tb11.02.0870	Tb927.6.2640	Experiment, Text mining	0.453	Moderate	None
Tb927.6.2640	Tb927.6.4740	Experiment,Text mining,Co-expression	0.976	Moderate	None
Tb927.6.2640	Tb09.160.2360	Experiment, Text mining	0.647	Moderate	None
Tb927.6.2640	Tb10.70.4720	Experiment,Text mining	0.769	High	Fried and Kutay (2003)
Tb927.6.2640	Tb927.8.4280	Experiment, Text mining	0.641	Moderate	None
Tb10.70.4720	Tb927.8.4280	Experiment, Text mining	0.535	Moderate	None
Tb927.8.4280	Tb927.8.3370	Experiment	0.502	Moderate	None

Table 6. Evaluation on protein interaction data obtained from STRING 8.2 database. The evaluation was based on the identified protein domains.

To gain an insight into nuclear transport, understanding on interactions between transport receptors and proteins of the nuclear pore complex (NPC) is essential. According to [27], the fluorescence resonance energy transfer (FRET) can be employed between enhanced cyan and yellow fluorescent proteins (ECFP, EYFP) in living cells in order to explain the transport of receptor through the NPC. A FRET assay has been used to analyze a panel of yeast strains expressing functional receptor--ECFP and nucleoporin-EYFP fusions. Based on this approach, points of contact in the NPC for the related Importin Pse1/Kap121 and Exportin Msn5 were successfully characterized. That study proved the advantage of FRET in mapping dynamic protein interactions in a genetic system. In addition, both Importin and Exportin have overlapping pathways through the NPC. However, our database mining approach did not reveal any functional interaction between Nucleoporin and Karyopherin proteins of *T. brucei*.

# 3.4. Sequence similarity between parasite proteins and their human counterparts

Table 6 shows the degree of protein sequence similarity between parasite and human proteins. The similarity search for the sequence was carried out with the help of BLASTp tool. All the parasite proteins of nuclear transport machinery were found to have their

counterparts in human. The degree of sequence similarity between parasite proteins and human counterparts range from 19% to 72%. The resulting score values range from 49.3 to 558. Meanwhile, all the identified human proteins contain the same protein domains involved in the nuclear transport.

Subject sequence	Human counterparts	Score	E-value	Sequence similarity (%)
Tb927.3.1120	NP_006316.1	313	1.00E-109	72%
Tb09.211.4360	NP_694858.1	221	6.00E-62	25%
Tb11.01.5940	NP_003391.1	558	0	33%
Tb11.02.0870	AAA85838.1	79.3	5.00E-20	40%
Tb927.2.2240	AAH20569.1	79.3	2.00E-16	29%
Tb927.6.2640	NP_036448.1	360	4.00E-119	42%
Tb927.6.4740	AAC50367.1	368	9.00E-113	29%
Tb927.7.1190	AAI42947.1	453	2.00E-27	27%
Tb11.03.0140	AAH45620.2	258	2.00E-09	38%
Tb927.10.8170	NP_705618.1	134	1.00E-33	28%
Tb927.8.3370	AAF36156.1	70.9	2.00E-17	27%
Tb11.01.7010	NP_002262.3	207	2.00E-56	23%
Tb11.02.1720	NP_006382.1	156	9.00E-28	24%
Tb11.01.8030	NP_002262.3	101	3.00E-23	21%
Tb11.01.7200	CAA41411.1	59.7	4.00E-10	19%
Tb927.7.6320	NP_001041659.1	146	4.00E-17	28%
Tb927.3.4600	NP_006381.2	65.9	2.00E-12	20%
Tb09.160.2360	NP_003601.1	142	2.00E-39	30%
Tb927.6.3870	NP_001073956.2	75.5	8.00E-18	31%
Tb927.7.5760	NP_037380.1	49.3	6.00E-11	26%
Tb10.70.4720	NP_002256.2	277	2.00E-81	28%
Tb927.8.4280	NP_005787.1	73.6	1.00E-19	31%

**Table 7.** Comparison of the identified parasite proteins with human counterparts at protein sequence level. This comparison involved BLASTp programme.

A study reported by [28] showed that despite the high degree of similarity in the primary structure of human and *T. cruzi* ubiquitins, the three amino acid difference is sufficient to distinguish parasite versus host proteins. In this study, a simplified one step purification procedure to partially purify T. cruzi ubiquitin was performed. Following this preparation, ELISA and Western blots were carried out to show that chagasic sera recognise T. cruzi but not human or Leishmania ubiquitin indicating a species-specific response. Thus, it is probable that the *T. brucei* proteins could also be distinguished from human counterparts at primary sequence level by using the immunodetection method.

## 4. General discussions

### 4.1. Transport of cargoes

In RanGTPase system, Ran-binding protein 1 (RanBP1) which is cytoplasmic localized binds RanGTP and eases the RanGAP-dependent conversion of RanGTP to RanGDP [29]. This indicates that RanBP1 catalyses the cytoplasmic disassembly of RanGTP-transport receptor complexes. These complexes are kinetically so stable that RanGAP alone fails to trigger GTP hydrolysis [30-32]. RanBP2 [33] is a major constituent of the cytoplasmic filaments of NPCs and exhibits similar functions as RanBP1. It has four RanBP1 homology domains and forms a stable complex with sumoylated RanGAP [34,35], in order to dismantle the RanGTP-transport receptor complexes that exit the nucleus. Importin- and exportin-mediated transport cycles can accumulate cargoes against gradients of chemical activity, which is an energy-dependent process. The RanGTPase system hydrolyses one GTP molecule per transport cycle, and a number of evidences suggest that this contains the sole input of metabolic energy [36-39]. We have successfully identified all the required key components in the *T. brucei* nuclear transport. Whether their functionalities in vivo are consensus with the known ones still remains to be further investigated.

# 4.2. Relationship between signaling pathways and nuclear transport

Many aspects of cell physiology are greatly dependent on the signaling pathways. This includes members of the mitogen activated protein (MAP) kinase family as well as phosphatidyl inositol 3 (PI3) and adenosine monophosphate (AMP) kinases which are crucial in controlling the cell growth, proliferation, apoptosis and the response to stress. By activating the signaling pathway through multiple kinase cascades, various stressors are able to regulate the nuclear transport. For example, oxidative and heat stress activate both MAP kinase kinase (MEK)-extracellular signal regulated kinase 1/2 (ERK1/2) and PI3 kinase-Akt pathways [40]. Based on these observations and the fact that many of the transport components are modified post-translationally, it was sensible to investigate whether these modifications are regulated by stress. A study reported by [41] showed that oxidant treatment induced phosphorylation and/or GlcNAc modification of soluble transport factors and nucleoporins. Interestingly, changes in transport factor modifications are not limited to stress conditions, as modifying ERK or PI3 kinase activities in unstressed cells also affect the transport factors. This is exemplified by the regulation of RanBP3 through ERK1/2-ribosomal S6 kinase (RSK) signaling, a regulatory link which ultimately controls the Ran concentration gradient. Furthermore, phosphorylation of Nup50 which is dependent on ERK, reduces its association with importin-β1 and transportin in vitro, and ERK2 is responsible to the oxidant-induced collapse of the Ran gradient [42]. It remains unknown how much modulating individual transport factors contributes to the overall regulation of nuclear trafficking. However, it is noteworthy that the kinase inhibitor PD98059, which targets ERK1/2 and ERK5, significantly increases classical nuclear import, both under normal and stress conditions. Taken together, these results highlight a critical role of ERK activity in nuclear transport, with ERK kinases targeting both soluble factors and nucleoporins [41]. Thus, there is an urgent need to investigate the possible connection between upstream signaling apparatus with nuclear transport components in *T. brucei*.

# 4.3. *In silico* approach for drug target discovery

We have provided interpretation of heterologous data sets for nuclear transport system of T. brucei from various resources. With the availability of protein databases and computer-aided softwares, we are able to explain various functional interactions between identified parasite proteins and how these functional interactions give rise to functionality and behavior of the parasite nuclear transport. This would partially facilitate the exhausted effort to obtain system-level understanding of T. brucei pathogenesis. Our in silico approach has the potential to speed up the rate of drug target discovery while reducing the need for expensive lab work and clinical trials. The conventional approaches in vivo and in vitro have high tendencies to produce inefficient results when investigating complex large scale data such as proteins associated with nuclear shuttling of macromolecules across the nuclear envelope. Therefore, the systematic in silico approach from this study provides a tremendous opportunity of cost effective drug target discovery for the pharmaceutical industry.

# 4.4. Experimental validation of in silico data

Experimental techniques such as yeast two-hybrid assay and affinity purification combined with mass spectrometry are useful to investigate the possible protein-protein interaction. However, they have their limitations in detecting certain types of interactions. They also have technical problems to scale-up for high-throughput analysis. In conjunction with this, in silico approach may solve those problems in inferring the protein function. The scope of experimental data can be expanded to increase the confidence of certain interacting protein pairs with the availability of databases containing in silico data such as protein domain and 3D structure. The databases integrate information from various resources such as computational prediction methods and public text collections. Since in silico and experimental approaches are complementary to each other, the combination of these different approaches is very useful to obtain a more accurate picture of T. brucei nuclear transport.

#### 4.5. Our further direction

*In silico* approach offers various advantages over *in vivo* and *in vitro* approaches such as non-use of animals, low costs, and reduced execution time. This approach allows identification of proteins of interest from a particular biological study. From a protein function standpoint, transfer of annotation from known proteins to a novel target is currently the only practical way to convert vast quantities of raw sequence data into meaningful information. Many bioinformatics tools now provide more sophisticated methods to transfer functional annotation, integrating sequence, family profile and structural search methodology. Thus, in addition to data mining for protein-protein interaction, further *in silico* approach should also consider structural alignment, molecular docking and pathway modeling in order to obtain a comprehensive and more reliable insight into protein-protein interaction of *T. brucei* nuclear transport.

#### 5. Conclusion

The availability of protein databases and computer-aided softwares to identify probable components of cellular mechanisms has become a new trend in the present scientific era. We demonstrate here a computational analysis of nuclear transport in *T. brucei* as an initial step and proof of concept for further investigation. Our approach successfully identified 22 *T. brucei* proteins essential for nuclear transport. All those parasite proteins were found to contain relevant functional domains that drive the translocation of macromolecules in the parasite. The phosphorylation and O-glycosylation sites were also detected in all identified parasite proteins. This has given us an insight into the regulatory aspect of parasite nuclear transport. The database mining of protein interaction has shown that nine out of 22 parasite proteins possess relevant functional interactions for nuclear transport activities. However, more functional interactions from nuclear transport constituents of *T. brucei* are required to elucidate the exact mechanism. The homology between the parasite proteins and human counterparts was shown by BLASTp analyses. Whether there are structural differences between them remain unknown.

The nuclear transport in *T. brucei* has been characterized by using the *in silico* approach. The predicted functionalities and regulatory aspects of parasite nuclear transport constituents were in agreement with the previous reports. Moreover, the protein interaction data derived from the public database has made the participation of parasite proteins in the mechanism more convincing. Thus, we have laid a path for understanding the nuclear transport machinery in *T. brucei*. The development of drugs that target as well as alter nuclear import and export will undoubtedly become beneficial in controlling Trypanosomiasis in future. Drugs that have a direct effect on a single protein must be able to localize to the same site as the protein and interact with one or more of its domains. Alternatively, a drug that effectively blocks the target protein from reaching its proper organelle can also inhibit the protein's function.

#### **Author details**

Mohd Fakharul Zaman Raja Yahya\* and Umi Marshida Abdul Hamid School of Biology, Faculty of Applied Sciences, MARA University of Technology Shah Alam, Shah Alam Selangor, Malaysia

# 6. References

- [1] Barrett M P, Burchmore R J and Stich A (2003). The trypanosomiases. Lancet 362 (9394): 1469-80.
- [2] Miles M (2003). American trypanosomiasis (Chagas disease). GC Cook, A Zumla (Eds.), Manson's tropical disease (21st edn.), Elsevier Science, London, pp. 1325-1337.
- [3] Hecker H and Bohringer S (1977). Morphometric analysis of the life cycle of Trypanosoma brucei. Ann. Soc. belge Med. trop., vol. 57 (4-5), pp. 465-470.
- [4] Ogbadoyi E, Ersfeld K, Robinson D, Sherwin T, Gull K (2000). Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. Chromosoma 108 (8): 501-13.
- [5] Acosta-Serrano A, Vassella E, Liniger M, Renggli C K, Brun R, Roditi I and Englund P T. (2001). The surface coat of procyclic Trypanosoma brucei: Programmed expression and proteolytic cleavage of procyclin in the tsetse fly. PNAS, vol. 98(4), pp. 1513-1518, 2001.
- [6] Gorlich D and Mattaj I W (1996). Nucleocytoplasmic transport. Science vol. 271 (5255), pp. 1513-1518.
- [7] Mattaj I W and Englmeier L (1998). Nucleocytoplasmic transport: The Soluble Phase. Annual Review of Biochemistry, vol. 67, pp. 265-306.
- [8] Fried, H., and Kutay, U. (2003). Nucleocytoplasmic transport:taking an inventory. Cell Mol Life Sci 60, 1659-1688.
- [9] Röhrig U F, Awad L, Grosdidier A, Larrieu, P, Stroobant V, Colau D, Cerundolo V and Andrew J. G. (2010). Rational Design of Indoleamine 2,3-Dioxygenase Inhibitors. Journal of Medicinal Chemistry 53 (3): 1172–89.
- [10] Hayes, C N (2008). varDB: a pathogen-specific sequence database of protein families involved in antigenic variation, Bioinformatics.
- [11] Altschul S F, Gish W, Miller W, Myers E W, and Lipman D J. (1990). Basic local alignment search tool. J. Mol. Biol, vol. 215, pp. 403-410.
- [12] Marchler-Bauer A, Anderson J B, Cherukuri P F, DeWeese-Scott C, Geer L Y, Gwadz M, He S, Hurwitz D I, Jackson J D, Ke Z, Lanczycki C J, Liebert C A, Liu, Lu C F, Marchler G H, Mullokandov M, Shoemaker B A, Simonyan V, Song J S, Thiessen P A, Yamashita R A, J. J. Yin, D. Zhang, and S. H. Bryant (2005). CDD: a Conserved Domain Database for protein classification. Nucleic Acid Research, vol. 33, pp. 192-196.

<sup>\*</sup> Corresponding Author

- [13] Letunic I, Doerks T and Bork P (2009). SMART 6: recent updates and new developments. Nucleic Acid Research. vol. 37, pp. 229-232.
- [14] Hunter S, Apweiler R, Attwood T K, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn R D, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, J. McDowall, Mistry J, Mitchell J A, Mulder N, Natale D, Orengo C, Quinn A F, Selengut J D, Sigrist C J, Thimma M, Thomas P D, Valentin F, Wilson D, Wu C H, and Yeats C (2009). InterPro: the integrative protein signature database. Nucleic Acids Research, vol. 37, pp. 211-215.
- [15] Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux D S, Pagni M, and Sigrist C J A (2006). The PROSITE database. Nucleic Acid Research, vol. 34, pp. 227-230.
- [16] Jensen L J, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, and Von Mering C (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acid Research. Vol. 7 pp 412-416.
- [17] Schmid M (1998). Novel approaches to the discovery of antimicrobial agents. Curr. Opin. Chem. Biol. vol. 2, pp. 529-534.
- [18] Frankel M B and Knoll L J (2009). The Ins and Outs of Nuclear Trafficking: Unusual Aspects in Apicomplexan Parasites. DNA and Cell Biology vol. 28, pp. 277-
- [19] Macara, I G (2001). Transport into and out of the nucleus. Microbiol Mol Biol Rev 65, 570-594.
- [20] Babior B M (1999). NADPH oxidase: an update. Blood 93 (5): 1464–76.
- [21] Nett I R E, D. Martin D M A, Miranda-Saavedra D, Lamont D, Barber J D, and Mahlert A (2009). The phosphoproteome of bloodstream form Trypanosoma brucei, causative agent of African sleeping sickness. Molecular & Cellular Proteomics, vol. 8 pp. 1527-1538.
- [22] Miller, M. W., Caracciolo, M. R., Berlin, W. K., and Hanover, J. A. (1999). Phosphorylation and Glycosylation of Nucleoporins. Archives of Biochemistry and Biophysics, 367(1), 51-60.
- [23] Gasiorowski, J. Z. and Dean, D. A. (2003). Mechanisms of nuclear transport and interventions. Advanced Drug Delivery Reviews, 55, 703-716.
- [24] Schuldt, A. (2012). Post-translational modification: A monoubiquitylation pore anchor. Nature Reviews Molecular Cell Biology, 13, 66.
- [25] Radu A, Blobel G and Moore M S (1995). Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins, Proc. Natl. Acad. Sci. USA 92: 1769-1773.
- [26] Gasiorowski, J. Z. and Dean, D. A. (2003). Mechanisms of nuclear transport and interventions. Advanced Drug Delivery Reviews, 55, 703-716.

- [27] Damelin M, S. P. (2000). Mapping interactions between nuclear transport factors in living cells reveals pathways through the nuclear pore complex. Mol Cell., 5(1), 133-40.
- [28] Télles S, Abate T and Slezynger T C (1999). Trypanosoma cruzi and human ubiquitin are immunologically distinct proteins despite only three amino acid difference in their primary sequence. FEMS Immunol Med Microbio, 24(2), 123-30.
- [29] Bischoff F R, Krebber H, Smirnova E, Dong W H, and Ponstingl H. (1995). Coactivation of RanGTPase and inhibition of GTP dissociation by Ran GTP binding protein RanBP1. EMBO J, vol. 14, pp. 705–715.
- [30] Bischoff F R and Görlich D. (1997). RanBP1 is crucial for the release of RanGTP from importin  $\beta$  -related nuclear transport factors. FEBS Lett, vol. 419, pp. 249–254.
- [31] Floer M, Blobel G and M. Rexach M (1997). Disassembly of RanGTP-karyopherin β complex, an intermediate in nuclear protein import. J Biol Chem, vol. 272, pp. 19538-19546.
- [32] Lounsbury K M and Macara I G (1997). Ran-binding protein 1 (RanBP1) forms a ternary complex with Ran and karyopherin β and reduces Ran GTPase-activating protein (RanGAP) inhibition by karyopherin β. J Biol Chem, vol. 272, pp. 551-555.
- [33] Yokoyama N (1995). A giant nucleopore protein that binds Ran/TC4. Nature, vol. 376, pp. 184-188.
- [34] Mahajan R, Delphin C, Guan T, Gerace L and Melchior F (1997). A small ubiquitinrelated polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. Cell, vol. 88, pp. 97–107.
- [35] Matunis M J, Coutavas E, and Blobel G (1996). A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. J Cell Biol, vol. 135, pp. 1457–1470.
- [36] Englmeier L, Olivo J C and Mattaj I W (1999). Receptor-mediated substrate translocation through the nuclear pore complex without nucleotide triphosphate hydrolysis. Curr Biol, vol. 9, pp. 30-41.
- [37] Kose S, Imamoto N, Tachibana T, Shimamoto T, and Yoneda Y (1997). Ran-unassisted nuclear migration of a 97 kD component of nuclear pore- targeting complex. J Cell Biol, vol. 139, pp. 841–849.
- [38] Ribbeck K, Kutay U, Paraskeva E and Görlich D (1999). The translocation of transportin-cargo complexes through nuclear pores is independent of both Ran and energy. Curr Biol, vol. 9, pp. 47–50.
- [39] Weis K, Dingwall C, and Lamond A I (1996). Characterization of the nuclear protein import mechanism using Ran mutants with altered nucleotide binding specificities. EMBO J, vol. 15, pp. 7120-7128.
- [40] Kodiha M, Banski P and Stochaj U (2009). Interplay between MEK and PI3 kinase signaling regulates the subcellular localization of protein kinases ERK1/2 and Akt upon oxidative stress. FEBS Lett, 583:1987-93.

- [41] Kodiha, M., Crampton, N., Shrivastava, S., Umar, R., and Stochaj, U. (2010). Traffic control at the nuclear pore. Aging, 237-244.
- [42] Czubryt M P, Austria J A and Pierce G N (2000). Hydrogen peroxide inhibition of nuclear protein import is mediated by the mitogen-activated protein kinase, ERK2. J Cell Biol, 148:7-16.



