Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2008

Proteomic, Bioinformatic and Functional Characterization of the Nuclear Pore Complex of the African Trypanosome

Jeffrey Allen DeGrasse

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Recommended Citation
DeGrasse, Jeffrey Allen, "Proteomic, Bioinformatic and Functional Characterization of the Nuclear Pore Complex of the African Trypanosome" (2008). *Student Theses and Dissertations.* Paper 26.

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.

PROTEOMIC, BIOINFORMATIC AND FUNCTIONAL CHARACTERIZATION OF THE NUCLEAR PORE COMPLEX OF THE AFRICAN TRYPANOSOME

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Jeffrey Allen DeGrasse
June 2008

© Copyright by Jeffrey Allen DeGrasse 2008

PROTEOMIC, BIOINFORMATIC AND FUNCTIONAL CHARACTERIZATION OF THE NUCLEAR PORE COMPLEX OF THE AFRICAN TRYPANOSOME

Jeffrey Allen DeGrasse, Ph.D.
The Rockefeller University 2008

The eukaryotic genome, and its associated proteins, is intricately packaged and sequestered within the boundary of a double membrane, known as the nuclear envelope (NE). Transport across the NE is mediated by large protein assemblages known as nuclear pore complexes (NPCs). Yeast and vertebrate NPCs are comprised of about 30 proteins, termed nucleoporins (Nups), which are present in multiple copies. The origins and evolution of the nucleus and NPC are not yet clear, although it seems likely that the nucleus arose only once in eukaryotic evolution. To further our understanding of the evolution of the NPC, we characterized the NPC of a distantly related organism, relative to yeast and vertebrates. The parasitic protist *Trypanosoma brucei* is a suitable candidate for such a study due to its sequenced genome and experimental tractability, compared to other protists. In this thesis, we present the comprehensive analysis of the protein components of the trypanosome NPC.

Towards this end, we used several biochemical and proteomic strategies to identify the proteins that associate with a preparation of enriched *T. brucei* NEs. Discerning authentic trypanosome Nups from the 859 proteins identified was challenged by the large sequence divergence between yeast, vertebrates and trypanosomes. To overcome this challenge, we used a suite of rigorous bioinformatic tools, which allowed us to identify 24 putative Nups. We then confirmed fully half of the putative trypanosome Nups by fluorescent localization,

and observed that the density of trypanosome NPCs around the nucleus is less than that of yeast or vertebrates. This lower density enabled us to visualize individual NPCs and note differences in the spatial distribution of NPCs between these three species. To further characterize these putative Nups and the NPC, we employed RNAi. The results of these studies suggest that, in addition to its role in nucleocytoplasmic transport, the trypanosome NPC plays a key role in maintaining the stability and morphology of the NE.

Despite significant divergence with respect to primary structure and species-specific innovations, the trypanosome NPC contains many homologs, domains and motifs found in opisthokonts. Given these findings, it is reasonable to infer that the architecture of the NPC is conserved across Eukaryota. This suggests that the NPC of the last common eukaryotic ancestor had many features in common with NPCs of contemporary bikonts (e.g. plants and excavates) and opisthokonts (e.g. animals and fungi).

The modest Rose puts forth a thorn,
The humble sheep a threat'ning horn:
While the Lily white shall in love delight,
Nor a thorn nor a threat stain her beauty bright.
- "The Lily," by William Blake

ACKNOWLEDGEMENTS

First and foremost, this work would not be possible without the extraordinary mentorship, encouragement and patience of Professors Brian T. Chait and Michael P. Rout. Their steadfast excitement for my project was an inspiration and deeply appreciated, especially during the more difficult times of experimental research.

It is an honor and a privilege to be a part of the Laboratory for Mass Spectrometry and Gaseous Ion Chemistry. I am fortunate to continue working with such a group of talented and dedicated scientists.

I thank Doctors Markus Kalkum, Andrew Krutchinsky, Júlio Padovan, Beatrix Ueberheide, and Wenzhu Zhang for helping me understand the various aspects of mass spectrometry. I thank Dr. David Fenyö whose informatic programs allowed me to efficiently extract the gold from the copious amount of data that was generated.

I have had numerous enlightening discussions with everyone in the lab, past and present. I acknowledge Dr. Martine Cadene, Dr. Emmanuel Chang, Dr. Ileana Cristea, Mr. Herb Cohen, Dr. Steve Cohen, Ms. Sarah Flowers, Dr. Erica Jacobs, Dr. Tijana Jovanovic, Dr. Mi Jin, Dr. Ji-Eun Kim, Dr. Derek McLachlin, Ms. Kelly Molloy, Dr. Sunnie Myung, Mr. Matt Sekedat, Dr. Alan Tackett, Dr. Michelle Trester-Zedlitz, Dr. Qingjin Wang, and Ms. Eileen Woo. I thank Mrs. Gladys McMilleon for her support, humor and jelly beans.

The members of the M.P. Rout lab exhibited incredible patience helping me, a one-time analytical chemist, learn biology and biochemistry. I am especially indebted to Dr. Mario Niepel who bore the brunt of my ignorance. I also thank Dr. Vincent Archambault, Dr. Ben Timney, and Ms. Rosemary Williams for all their help, support and insightful conversations.

I am indebted to Dr. G.A.M. Cross for allowing me to work in his lab and his very helpful advice. I especially thank Mr. Nicolai Siegel for his continued help, support and friendship. I have had many rewarding discussions with Dr. Luisa Figueiredo, Dr. Christian Janzen, Mrs. Magdelena Kartvelishvili, Dr. Hee-Sook Kim, Dr. Bibo Li, Ms. Jenny Li, Dr. Veena Mandava, and Dr. Eiji Okubo.

I thank Dr. Thomas Seebeck and Dr. Michael Oberholzer for the gift of the GFP tagging vectors and Dr. Bill Wickstead for the gift the TbNup-1 antibody.

This project required extensive bioinformatic work. I thank Dr. Andrej Sali and his lab for allowing me to visit to learn the nitty-gritty of bioinformatics. Dr. Damien Devos was kind enough to work with me to develop a rigorous research strategy to tackle the challenge of identifying the TbNups. I also thank Dr. Frank Alber for his deep insight.

I thank the Field lab at the University of Cambridge for the insightful conversations, support and beer. I especially thank Dr. Mark Field, whose advice and hospitality was invaluable during this work.

At the bioimaging facility, I thank Dr. Alison North and Dr. Patrick Nahirney. Alison's patience with someone who had never attempted bio-imaging is worthy of high praise. Patrick provided the excellent TEM instruction that enabled me to observe several fascinating phenotypes induced by RNAi knockdowns.

I'd like to thank my Faculty Advisory Committee for their advice and support over the years.

I also thank Professor Joe Glavy for his critical read of this thesis, inexhaustible list of possible experiments to try, the αHsNup107 and the many excellent conversations that have shaped many of my arguments in this thesis.

All of the friendships that I have made here at Rockefeller have made this journey an extremely enjoyable and valuable one. To my poker buddies, I'm happy that I could supplement your income over the years.

Bishop Desmond Tutu once said: "You don't choose your family. They are God's gift to you, as you are to them." I dedicate my thesis to my family. The self-sacrifice and love of my parents, Quintin and Bonita DeGrasse, molded me into who I am today. My sister, Wendy, recently went back to school to earn her college degree and is an inspiration to us all. My niece, Casey, is exceptional and we are so very proud of her. There will always be a special place in my heart for Tish, whose love, support, advice and encouragement made the tough days easier and the good days even better.

This work was funded by a fellowship provided by the Training Program in Chemical Biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

BLAST	Basic Local Alignment Search Tool
DHB	2,5-Dihydroxy Benzoic acid
EM	Electron Microscopy
ER	Endoplasmic Reticulum
ESI	Electrospray Ionization
GDP	Guanosine diphosphate
GFP	Green Fluorescent Protein
GOI	Gene Of Interest
GTP	Guanosine triphosphate
HA	Hydroxyapatite
HMM	Hidden Markov Model
INM	Inner Nuclear Membrane
Kap(s)	Karyopherin(s)/Transportin(s)
LDS	Lithium Dodecyl Sulfate
LCEA	Last Common Eukaryotic Ancestor
m/z	Mass-to-charge Ratio
MALDI	Matrix Assisted Laser Desorption/Ionization
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NE	Nuclear Envelope
NEP	Nuclear Envelope Preparation
NES	Nuclear Export Sequence
NLS	Nuclear Localization Sequence
NPC	Nuclear Pore Complex
Nup(s)	Nucleoporin(s)
ONM	Outer Nuclear Membrane
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PCLF	Pore Complex Lamina Fraction
PSI-BLAST	Position-Specific Iterative BLAST
PVP	Polyvinylpyrrolidone
qPCR	Quantitative PCR
QqTOF	Quadrupole-quadrupole Time Of Flight
RNAi	RNA Interference
SDS	Sodium Dodecyl Sulfate
siRNA	Short Interfering RNA
SMB	Single Marker Bloodstream
TbNEP	*T. brucei* Nuclear Envelope Preparation
TEM	Transmission Electron Microscopy
TET	Tetracycline

CHAPTER 1 — INTRODUCTION

Unlike prokaryotes, eukaryotes sequester their genome and associated proteins within a conspicuous, complex and dynamic organelle known as the nucleus — the epicenter for information storage and retrieval (Foster and Bridger, 2005; Vandriel et al., 1991). Serving as a physical boundary between the nuclear interior (nucleoplasm) and the cytosol is a double membrane known as the nuclear envelope (NE). To translocate across the NE, molecules must pass through the nuclear pore complex (NPC), which thus functions as a gatekeeper. The composition and function of the NPC has been characterized for the metazoa (mainly vertebrates) and the fungi (principally *S. cerevisiae*) (Cronshaw et al., 2002; Rout et al., 2000).

The nucleus was the first organelle to be described and its discovery is attributed to Franz Bauer in 1802. It is the presence of a nucleus that differentiates eukaryotes from the bacteria (prokaryotes). Though the prokaryote *Gemmata obscuriglobus* contains a double membrane bounded nucleoid, it remains to be determined how this structure relates to the eukaryotic nucleus (Lindsay et al., 2001). All current evidence points towards divergent evolution of the eukaryotic tree of life, which implies that the nucleus, and the entire endomembrane system, arose only once through natural selection and was present in the Last Common Eukaryotic Ancestor (LCEA) (Martin, 2005).

The NPC may serve as a suitable structure for resolving the evolutionary relationships between eukaryotes. Since only one of the possibly many proto-

forms of the nucleus survived natural selection, it is reasonable to posit that the core mechanism to move materials across the NE was present within the LCEA (Martin, 2005). In such a scenario, the inherited components of the LCEA NPC would have diverged in such a manner that is synapomorphic, and may used to establish phylogenies.

In this thesis, we focus on the NPC in an evolutionary context. Recognizing that the fungi and metazoa are closely related in the group of eukaryotes known as opisthokonts, relative to the entire catalog of eukaryotic species, we describe a third, more distantly related, system. Here, we characterize the NPC of the protist *Trypanosoma brucei* using an integrated approach, which includes proteomics, bioinformatics, *in situ* fluorescent labeling, and RNAi. The excavates are a diverse group of protists that were once thought to represent an early branch of eukaryotes, though recent efforts to root the eukaryotic tree have called this into question (Keeling et al., 2005; Stechmann and Cavalier-Smith, 2002). With this assembled triad of NPCs, we are in a position to discuss the divergent evolution of the NPC and the core requirements of nucleocytoplasmic transport.

THE NUCLEUS AND THE NUCLEAR ENVELOPE

As depicted in Figure 1, the nucleus is bound within a continuous membrane structure known as the nuclear envelope (NE) (Callan and Tomlin, 1950). The NE is comprised of three distinct membrane domains. The outer nuclear membrane (ONM), which is continuous with the endoplasmic reticulum (ER) (Gerace and Burke, 1988; Newport and Forbes, 1987). The inner nuclear

membrane (INM) is populated with the proteins involved with chromatin regulation and nuclear stability (Dreger et al., 2001; Schirmer et al., 2003). Periodically, the INM and the ONM fuse to form annuli, which is the third membrane domain. These pores are plugged with nuclear pore complexes (NPCs) — large protein complexes that arbitrate macromolecular traffic into and out of the nucleus (Feldherr, 1962; Gorlich and Kutay, 1999; Wente, 2000). Figure 2 shows the generalized topology of the metazoan nuclear envelope. Within the nuclear face of many metazoan NEs, a dense network of coiled coil proteins known as lamins form a structural lattice (Cohen et al., 2001). Even though all three membrane domains are continuous, they each possess a discreet assemblage of proteins with specialized functions (Prunuske and Ullman, 2006).

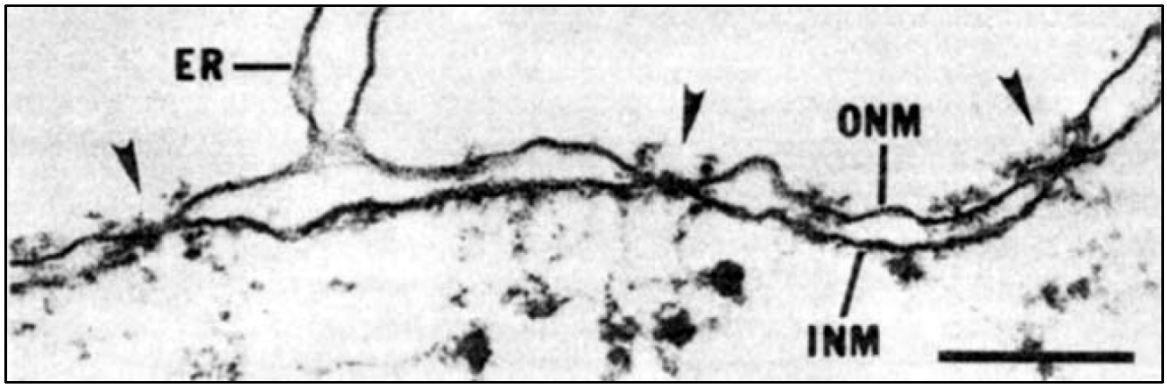


Figure 1: Thin section EM of *X. laevis* oocyte NE showing the three domains of the NE. Arrow heads indicate NPCs. Scale bar = 200nm (from Gerace and Burke, 1988).

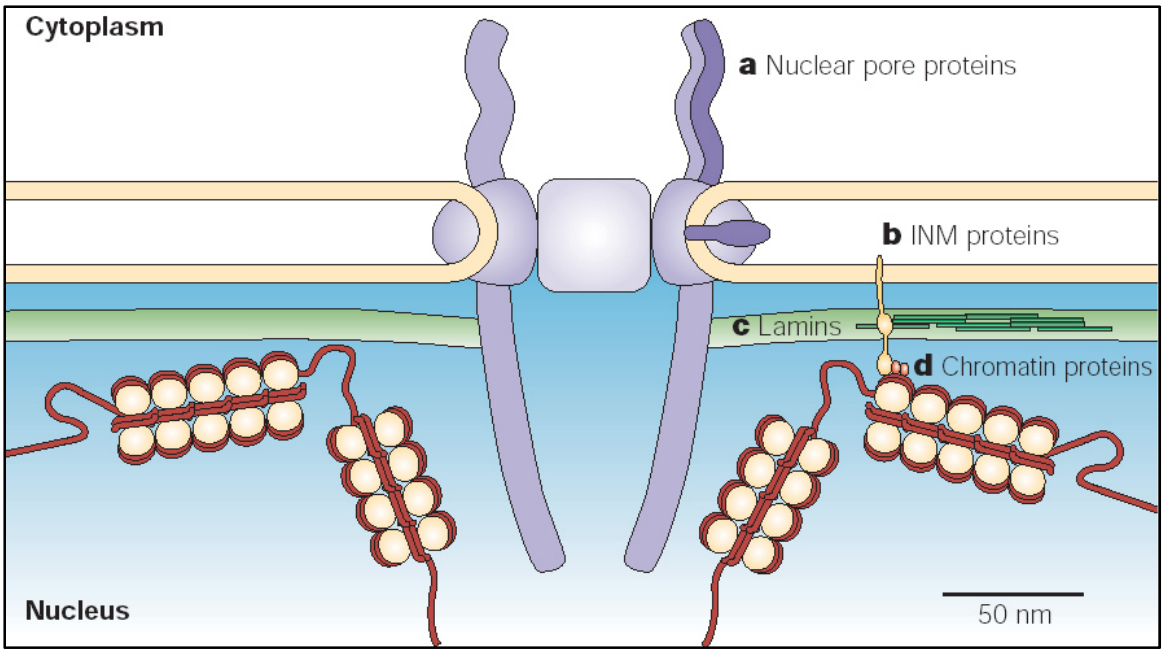


Figure 2: Topology of the generalized nuclear envelope. Also shown are examples from the major classes of nuclear envelope proteins: (a), nuclear pore proteins; (b), inner nuclear membrane proteins; (c), lamins; and (d), chromatin proteins (from Burke and Ellenberg, 2002).

THE NUCLEAR PORE COMPLEX

Continuous bidirectional trafficking of specific macromolecules across the NE is requisite for cellular function. The regulators and facilitators of nucleocytoplasmic exchange, NPCs, are large protein assemblies of discreet sub-complexes with an 8-fold degree of symmetry (Maul, 1971; Suntharalingam and Wente, 2003). The building blocks of the NPC are generally large proteins termed nucleoporins, or Nups. The majority of Nups have been characterized within the last 15 years, culminating in two large scale proteomic efforts. One study sought to identify and localize all the Nups in the *Saccharomyces cerevisiae* NPC and began work to determine its molecular architecture (Rout et al., 2000). The molecular architecture of the NPC is defined by the interconnectivity of the Nups into several major sub-complexes (for example, the seven member ScNup84/HsNup107 sub-complex), which then interact to generate the overall structure of the NPC (Fahrenkrog and Aebi, 2003). The yeast NPC, one of the largest protein structures in the cell, is a 44 MDa complex comprised of about 30 different Nups that are present in multiple copies of 8, 16 or 32 for a total of at least 456 Nups per NPC (Rout et al., 2000). The second proteomic study, focusing on the *Rattus norvegicus* NE, revealed that the mammalian NPCs are ~60 MDa and also contain about 30 distinct Nups (Cronshaw et al., 2002). Even though yeast and mammals descended through different kingdoms, their NPCs share many similarities in composition and architecture (Suntharalingam and Wente, 2003).

Over the last two decades, many electron microscopy (EM) studies have

visualized the overall structure and gross architecture of the NPC. Collectively, these studies have revealed that the morphology of the NPC is generally conserved across species (Akey and Radermacher, 1993; Lim et al., 2006; Lim and Fahrenkrog, 2006). Cryoelectron tomography coupled with three dimensional reconstruction offers the highest resolution (6 nm) structure to date of the *Dictyostelium discoideum* NPC (Beck et al., 2004; Beck et al., 2007). Shown in Figure 3, the diameter of the slime mold NPC is about 125 nm while the central pore itself is 60 nm. The structure resolves cytoplasmic filaments which point inwards toward the pore. The basket extends about 60 nm into the nucleoplasm. Overall, cargo must traverse about 110 nm through the NPC.

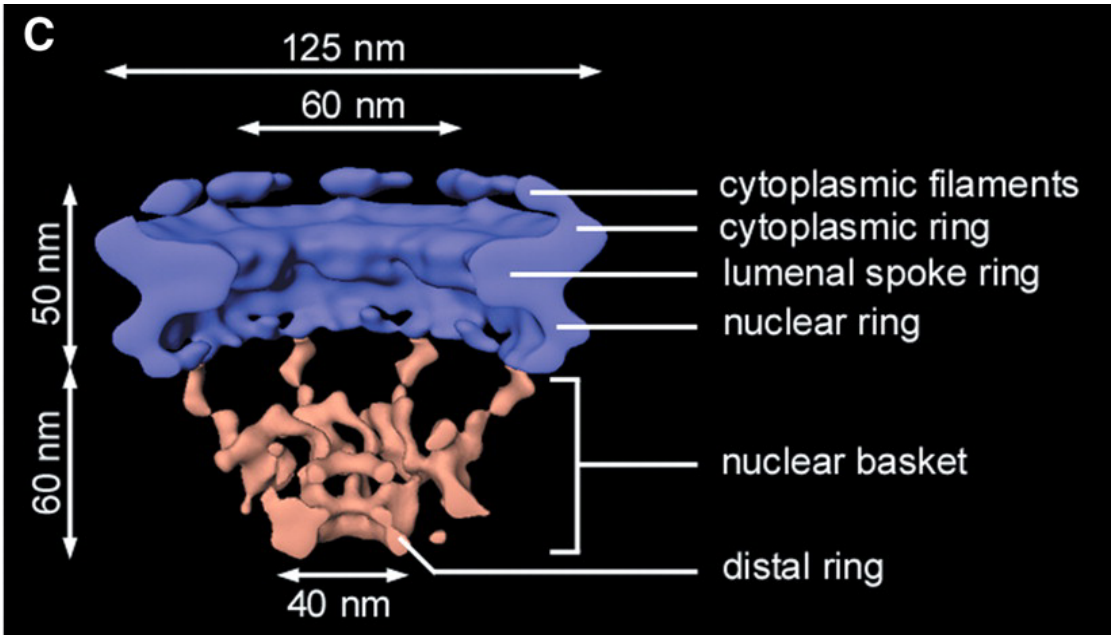


Figure 3: The reconstructed 3-dimensional structure of the *D. discoideum* NPC as determined by cryoelectron tomography (from Beck et al., 2004).

CLASSIFICATION AND CHARACTERIZATION OF NUCLEOPORINS

The molecular architecture of the NPC is modular (Devos et al., 2006). Nucleoporins form discreet sub-complexes, which in turn interact to compose three distinct groups within the NPC: transmembrane anchoring Nups, the structural scaffold Nups, and the FG-Nups, which are anchored to the scaffold (Figure 4) (Alber et al., 2007a). The FG-Nups play a functional role in nucleocytoplasmic transport and contain natively disordered phenylalanine-glycine-containing sequence repeat domains (Denning et al., 2003; Devos et al., 2006).

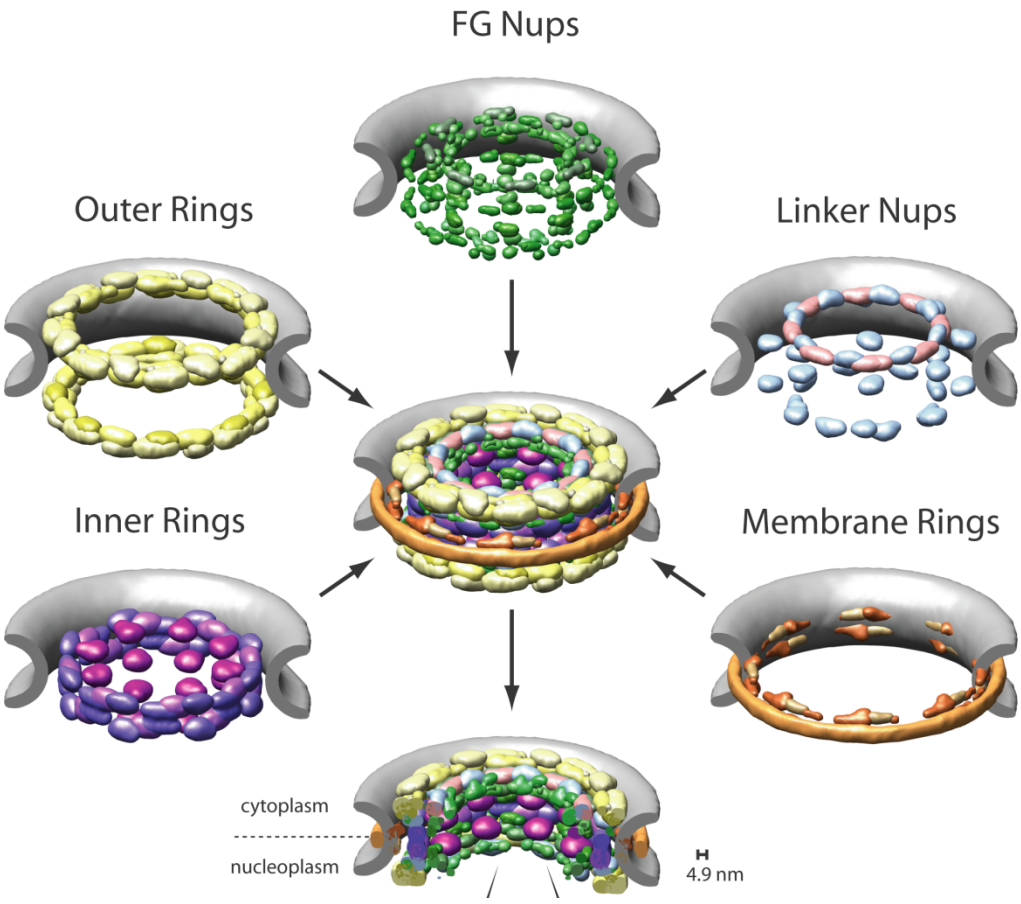


Figure 4: The structure and molecular architecture of the NPC. The membrane ring anchors the NPC into the pore within the NE. On top of membrane ring, lays the structural scaffold, which is comprised of the inner and outer rings of Nups and the linker Nups. Finally, the FG-Nups line the central channel to enable selective nucleocytoplasmic transport (from Alber et al., 2007a).

Given the low primary sequence similarity between Nups of yeast and vertebrates (for example, between *S. cerevisiae* and mammals) (Suntharalingam and Wente, 2003; Vasu and Forbes, 2001), we found that information about the secondary and tertiary structures of Nups is imperative to confidently predict orthologs in other, more distant, species. While several Nup structures have been experimentally determined at atomic resolution (Berke et al., 2004; Hodel et al., 2002; Jeudy and Schwartz, 2007; Melcak et al., 2007; Napetschnig et al., 2007; Vetter et al., 1999; Weirich et al., 2004), representing about 5% of the NPC, a wealth of information was gleaned from a homology modeling study that predicted the structures of all of the nucleoporin domains (Devos et al., 2006). Remarkably, each of the more than forty domains present in the population of Nups is predicted to adopt one of only eight different fold types. Of these, the-helix rich α-solenoid, the FG-repeat domains, and the WD-repeat β-propeller folds account for more than 80% of nucleoporin domains (Andrade et al., 2001a; Andrade et al., 2001b; Denning et al., 2003; Devos et al., 2006; Neer et al., 1994).

Appendix A lists the nucleoporins and the transiently interacting Nup Rae1/Gle2 of the well characterized human (based, in part, on the closely related rat) and budding yeast NPC. We determined the orthologs in other closely related species as well as *Arabidopsis thaliana* through BLAST analysis (and confirmed by reciprocal BLAST search and published data). Although several Nups were not identified by sequence homology, the NPCs of closely related species are generally able to be elucidated using bioinformatics. This is not the

case, however, when we extend such analyses to distantly related species, such as the Excavata.

PRIMARY FUNCTION: NUCLEOCYTOPLASMIC TRANSPORT

The NPC appears to be freely permeable to small molecules under 40 kDa (Keminer and Peters, 1999). However, larger macromolecules require a mechanism built upon transport factors (karyopherins) and a RanGTP/RanGDP gradient to transit through the NPC (Madrid and Weis, 2006; Stewart, 2007a). Karyopherins (Kaps) are highly conserved transport factors that bind their cargo through the recognition of specific nuclear localization sequences (NLS) or nuclear export sequences (NES) (Dingwall and Laskey, 1991; Lange et al., 2007; Mosammaparast and Pemberton, 2004). The structural details of the Kap-cargo-Nup interaction have yielded many clues toward the mechanism of transport (Stewart, 2007a).

As the Kap-cargo complex passes through the center channel and out of the NPC, it interacts with the FG-Nups, through their FG-repeat domains (Stewart, 2006). Phe residues occupy the deep groove between the α-helices of the HEAT domain of the Kap, while the disordered spacer region between repeating Phe residues lays across the surface of the Kap (Bayliss et al., 2000b; Denning et al., 2003). Since the FG-Nups are present in multiple copies, the yeast NPC has more the 160 FG-repeat domains representing several thousand individual FG-motifs (Rout et al., 2000). Although over half of the mass of the FG-repeats in yeast may be deleted without abolishing the selectivity and transport capacity of the NPC (Strawn et al., 2004), some FG-repeat domains are

more essential than others. It has also been shown that specific structural scaffold Nups play a key role in anchoring FG-Nups to the NPC. For example, deletion of the yeast Nup170 or Nup188 (Shulga et al., 2000) and *C. elegans* Nup205 or Nup93 (Galy et al., 2003), which are members of the structural scaffold, increases the permeability of the NPC.

The energy for this process is provided through the hydrolysis of GTP to GDP, which is associated with the small GTPase Ran (Figure 5). Vectorial transport is established in part by the RanGTP/RanGDP gradient across the NE, which is maintained by specific localization of RanGAP (hydrolyzes GTP to GDP) into the cytoplasm and RanGEF (exchanges GTP for GDP), which is bound to chromatin (Mosammaparast and Pemberton, 2004). A protein destined for the nucleus associates with a Kap through its basic NLS domain and is then shuttled through the NPC into the nucleus where the Kap is allosterically displaced by RanGTP.

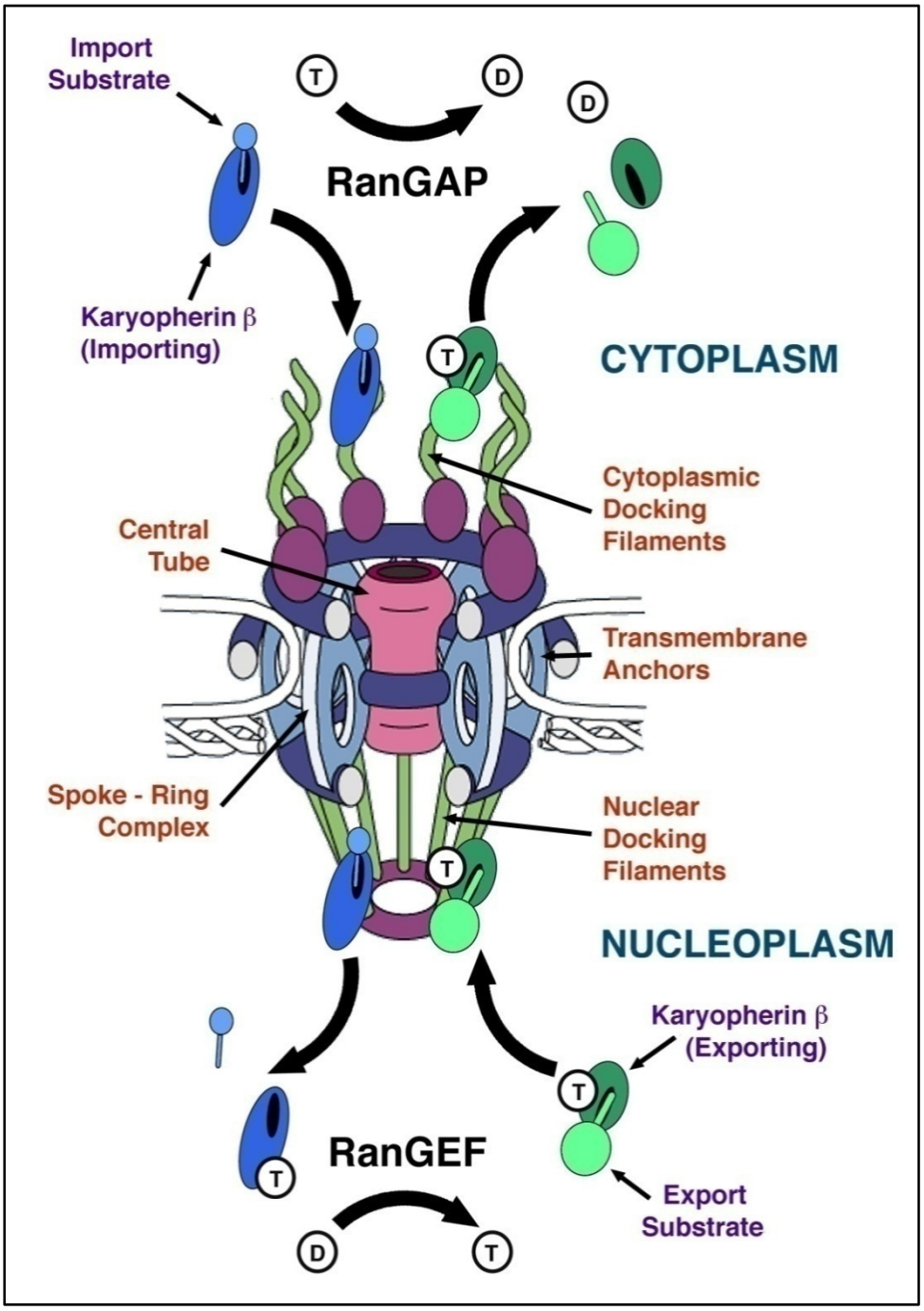


Figure 5: Generalized structure of the nuclear pore complex showing the domains of the pore complex. Also shown is the bi-directional transport pathway. T, RanGTP. D, RanGDP (from Rout *et. al.* 2001).

An exporting Kap-RanGTP complex can then bind a substrate with an NES and translocate back into the cytoplasm where RanGAP hydrolyzes GTP to GDP, dissociating the complex. Another protein, Ntf2, binds to RanGDP and recycles it to the nucleus where RanGEF converts it to RanGTP to restore the gradient.

THE NPC IS DYNAMIC AND ADAPTIVE

The NPC is not simply a static pore straddled across the NE. By altering its architecture and composition in a cell-cycle specific manner, the NPC can affect the specificity of nucleocytoplasmic transport (Tran and Wente, 2006). Several Nups are dynamic, with short residence times, while others are stable over considerable periods of time (Rabut et al., 2004). EM studies have shown that, under different transport conditions within the cell-cycle, some Nups may alter their distributions within the NPC (Fahrenkrog et al., 2002; Paulillo et al., 2005). Also, in a subset of closed mitotic systems, such as the filamentous fungus *Aspergillis nidulans*, the NPC partially dissociates and thus increases the permeability of the nucleus (De Souza et al., 2004). While remaining largely intact, the NPC of the budding yeast alters nucleocytoplasmic transport through changes to its architecture during mitosis (Makhnevych et al., 2003).

OTHER ROLES FOR THE NPC

Mediating bidirectional transport across the NE is not the sole physiological role of the NPC (Fahrenkrog et al., 2004). The NPC and its constituents have been associated with chromatin boundary activity, nascent mRNA retention, mitosis, and telomere silencing and maintenance.

The yeast NPC has been implicated in secondary roles such as chromatin organization and gene regulation (both silencing and activation). Several yeast transport factors (for example, Cse1 and Mex67, among others) block the spreading of heterochromatin and thus act as a boundary by physically tethering genomic loci to ScNup2, which transiently locates to the nuclear basket (Ishii et al., 2002). The dynamic nature of ScNup2 and associated transport factors suggests that this physical boundary activity is likewise dynamic, possibly affecting the transition between chromatin activity states (Dilworth et al., 2005). In addition, it has been shown that ScNup2 transiently interacts with numerous gene regulatory promoter regions (Schmid et al., 2006; Schmid et al., 2004). This suggests that the nucleoporin-promoter interaction, or Nup-PI, functions as an early step to gene activation (Schmid et al., 2006). Furthermore, the Rap1/Gcr1/Gcr2 transcriptional repressor/activator complex is anchored to the nuclear periphery through the ScNup84 complex (Menon et al., 2005). Taken together, the data indicates that the NPC plays a critical role in gene activation (Casolari et al., 2004). Such dynamic associations localize the boundaries of silent chromatin to the nuclear periphery while specific active transcription regions are drawn proximal to the NPC, perhaps for efficient transcriptional regulation and mRNA export.

The yeast Mlps (ScMlp1 and ScMlp2) and the mammalian ortholog, HsTpr, are large coiled coil proteins that associate with the nucleoplasmic face of the NPC and likely form the nuclear basket (Krull et al., 2004; Strambio-de-Castillia et al., 1999). It has been suggested that the Mlps play a role in telomere

anchoring and the organization of perinuclear silenced chromatin (Feuerbach et al., 2002; Galy et al., 2000). However, other studies indicate that the Mlps are not required for telomere anchoring (Andrulis et al., 2002; Hediger et al., 2002b). On the contrary, instead of anchoring telomeric regions, the Mlps may interact with the protein networks involved with the maintenance of telomere length (Hediger et al., 2002a).

Consistent with the Nup-PI observation, Mlp1, and perhaps Mlp2, may bind to and retains nascent intron-containing mRNA (Galy et al., 2004; Vinciguerra et al., 2005). Once spliced and adorned with mRNPs and transport factors (e.g. the Mex67:Mtr2 heterodimer), mRNA is exported from the nucleus in perhaps a ratcheting manner (Erkmann and Kutay, 2004; Stewart, 2007b). Away from the NPC, recent work has shown that Mlp2, but not Mlp1, co-localizes with the spindle pole body and functions to maintain the integrity of the spindle pole body (Niepel et al., 2005).

Members of the structural scaffold have been shown to have additional roles outside the NPC. The mammalian Nup107-160 scaffold sub-complex is essential for the proper assembly of NPC (Harel et al., 2003; Walther et al., 2003). During mitosis, a small percentage of this complex is localized to the kinetochores (Belgareh et al., 2001; Harel et al., 2003; Loiodice et al., 2004; Zuccolo et al., 2007) and there is evidence that Nup96, a member of the Nup107-160 complex, localizes to the mitotic spindle and spindle poles (Enninga et al., 2003). Recent work has shown that the Nup107-160 sub-complex functions in a

manner consistent with spindle checkpoint proteins, and distinct from its role at the NPC (Orjalo et al., 2006).

The Nup107-160 complex is not the only NPC constituent to function as mitotic regulators. Mammalian Nup358 moves from the cytoplasmic face of the NPC (Lim and Fahrenkrog, 2006) to the kinetochores (Joseph et al., 2002) and its deletion induces mitotic arrest (Joseph et al., 2004; Salina et al., 2003). During mitosis, the mRNA transport factor Rae1 localizes to the spindle and spindle poles (as shown in *X. laevis*) and its deletion disrupts the early stages of spindle assembly (Blower et al., 2005). Rae1 has been implicated in *S. pombe* and vertebrates as regulator of mitotic cell cycle progression (Jeganathan et al., 2005; Whalen et al., 1997). There is also evidence of similar dual functions in budding yeast. Mutations in the yeast Nup170, but not its paralog Nup157, causes defects in kinetochore formation and chromosome segregation (Kerscher et al., 2001), possibly through its interaction with the spindle assembly checkpoint proteins Mad1 and Mad2 (Campbell et al., 2001; Iouk et al., 2002; Kastenmayer et al., 2005). The functional connection between nucleocytoplasmic transport across the NPC and mitotic spindle assembly continues with the small GTPase Ran, which is heavily involved with both processes (Mosammaparast and Pemberton, 2004; Quimby et al., 2005; Quimby and Dasso, 2003).

THE NPC THROUGHOUT EVOLUTION

The nucleus is a complex structure that likely arose only once about 2 billion years ago (Martin, 2005). Interestingly, prokaryotes have yet to reveal any

obvious analogs of the nucleus (Doolittle, 1980). Even though some prokaryotes can envelope their DNA behind an internal membrane (Lindsay et al., 2001), there are no intermediate non-eukaryotes with a proto-nuclear structure known to exist today (Embley and Martin, 2006). The lack of obvious transitional species challenges efforts to model the emergence of the nucleus, though several have been proposed (Martin, 2005).

Recent studies support the vesicular (Martin, 1999) or invagination (Cavalier-Smith, 2002) models, which predict the formation of the endomembrane system, ER and NE from the plasma membrane. In the post-genomic era, genomic and proteomic data have allowed bioinformatic studies into the origin of the nucleus by investigating the NE (Bapteste et al., 2005; Mans et al., 2004) or nucleolar (a sub-nuclear region) (Staub et al., 2004) proteomes. These studies revealed that the some of the components of both the NE and the nucleolus have archaebacterial affinities, while others have eubacterial affinities (Mans et al., 2004; Staub et al., 2004). The implication of these findings is the most likely scenario that the nucleus arose after the symbiotic event that gave rise to the mitochondrion, due to the presence of both archae- and eubacterial derived genes. Also, additional computational studies showed structural (but not sequence) similarities between certain members of the NPC and coated vesicles (Alber et al., 2007b; Devos et al., 2004; Devos et al., 2006). The structural relationship between some Nups and the vesicle coat proteins suggests that the nucleus arose only after the emergence and evolution of membrane bending proteins.

Though the modern nucleus of the eukaryotes descended from the LCEA, the constituents of the NPC have diverged significantly between distant species. However, despite the large primary structure divergence between yeast and vertebrates, there is evidence that the domains and secondary structures within many Nups are conserved across Eukaryota (Cronshaw et al., 2002; Rout et al., 2000).

Here, we begin to address the extent and nature of species-specific innovations at the NPC. To do so, we first attempted to gather information about the NPCs of representative extant eukaryotes. Our early attempts to map nucleoporin orthologs across Eukaryota by sequence similarity alone were challenged by the aforementioned primary structure divergence. Indeed, such studies have presented grossly incomplete inventories of differently evolved eukaryotes, relative to the fungi and metazoa (Bapteste et al., 2005; Mans et al., 2004). An experimentally characterized representative NPC that is distantly related to the opisthokonts is requisite to begin to resolve the evolution of the NPC from the LCEA.

In this thesis, we examine differences in structure and protein composition between distantly related species in the hope of gaining a greater understanding of the function, mechanism, regulation and evolution of NPCs. Advances in the post-genomic era have given rise to a detailed picture of the NPCs of several model organisms (Appendix A). The model metazoans and yeast, however, are relatively closely related within Eukaryota (Gerhart, 1997). By applying the same rigorous interrogation to distantly related organisms (here the excavate

Trypanosoma brucei), we anticipate to gain a deeper understanding of one of the key organelles that distinguishes a major class of life.

TRYPANOSOMA BRUCEI

In order to discuss the NPC in an evolutionary context, we chose to study the NPC of a differently evolved eukaryote. The protists which comprise the taxon Kinetoplastida diverged from the rest of the eukaryotes near the root of the evolutionary tree some 1-2 billion years ago, giving rise to a large number of species, which collectively parasitize almost all vertebrates, invertebrates and some plant groups (Beverley, 2003). The kinetoplastids, including *Trypanosoma brucei*, are characterized by a single flagella and a single mitochondrion associated with a DNA containing kinetoplast (Figure 6) (Sogin et al., 1989). Since it is distantly related, the parasitic family Trypanosomatida has evolved very differently from other well studied organisms and relies on many species-specific adaptations (Simpson et al., 2006). The trypanosome nuclear envelope contains lamin-like coiled coil proteins (Ogbadoyi et al., 2000; Rout and Field, 2001), the presence of which in other eukaryotes usually signifies an open mitotic system. However *T. bruce*i undergoes a closed mitosis, during which the chromosomes do not condense (Ersfeld et al., 1999).

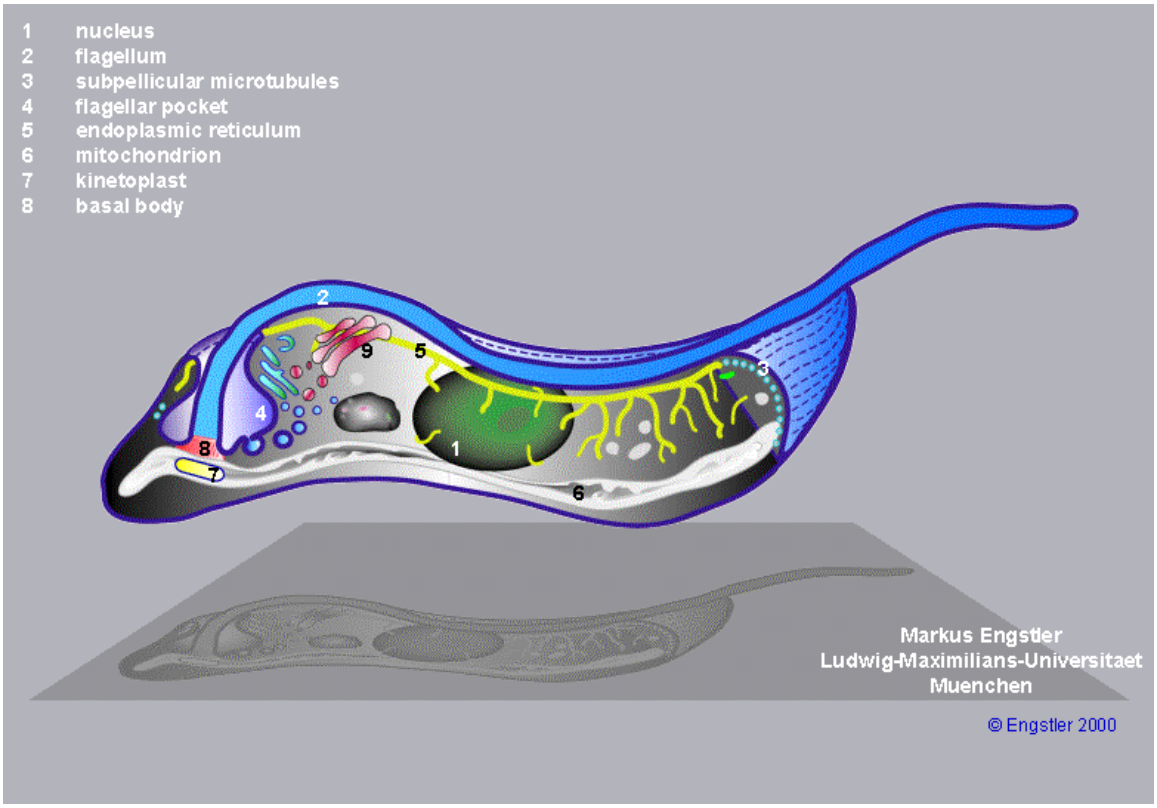


Figure 6: Cartoon schematic of *Trypanosoma brucei*.

T. brucei deviates from the eukaryotic model of transcription (Simpson et al., 2002). Polycistronic transcription of unrelated but unidirectionally clustered genes is followed by a trans-splicing event (Campbell et al., 2003; Liang et al., 2003; Siegel et al., 2005). The nascent mRNA molecules mature when functionalized by a 5' 40nt spliced leader (SL) RNA capping and 3' polyadenylation (Clayton, 2002). RNA polymerase II promoters are extremely rare (Das and Bellofatto, 2003; Gilinger and Bellofatto, 2001), and the mechanism of transcriptional regulation and initiation remains unclear although recent studies show progress in this area (Das et al., 2005; Ruan et al., 2004; Schimanski et al., 2005). The brunt of the regulation of protein levels is presumably delegated to post-transcriptional and post-translational mechanisms and higher order chromatin structure at the NE (Navarro et al., 1999). While the basic mechanisms of post-transcriptional regulation are understood, the details of the interaction between the signal cascade and modification machinery is still unclear (Akker et al., 2001). Understanding the role of the NPC with respect to the regulatory mechanisms of *T. brucei* could reveal novel features of eukaryotic protein regulation.

The 35 megabase *T. brucei* genome is divided amongst three classes of chromosomes: 11 large, <5 intermediate, and ~100 mini-chromosomes (Ersfeld et al., 1999). The 26 megabase genome (compare to the 12-14 Mb yeast genome (Goffeau et al., 1996)) within the 11 large chromosomes was recently sequenced (Berriman et al., 2005). The genome is almost devoid of introns and over 50% of the sequence encodes about 8100 open reading frames (ORFs)

(Berriman et al., 2005). The three closely related pathogens, *T. brucei*, *T. cruzi*, and *Leishmania major*, are collectively referred to as the "TriTryps" and contain about 6200 orthologous genes (> 75% of the *T. brucei* ORFs) between them (El-Sayed et al., 2005). Taking advantage of the post-genomic era, several *T. brucei* proteomic projects have been undertaken (Atwood et al., 2005; Broadhead et al., 2006; Colasante et al., 2006; Jones et al., 2006; Paba et al., 2004).

A WORLD HEALTH ISSUE

In the short history of *H. sapiens*, over 70 diverse species of protists have adapted to parasitize modern humans (Ashford and Crewe, 1998). Generally, parasites infect humans via an insect vector or consumption of contaminated food or water (Cox, 2002). *Entamoeba histolytica* (Amoebae) and *Giardia lamblia* (a diplomonad), are ingested through unsanitary conditions and may cause serious symptoms, such as diarrhea. Parasites transmitted by biting insects, such as Leishmania and Plasmodium, cause Leishmaniasis and Malaria, respectively. Three species of trypanosomes infect humans: *Leishmania*, *Trypanosoma cruzi,* responsible for Chagas' disease in South America, and *Trypanosoma brucei*, the causative agent of the African sleeping sickness.

With close to 60 million people at risk in 36 sub-Saharan African countries, African trypanosomiasis, or African sleeping sickness, is a serious world health issue (Garcia et al., 2006). The causative agents of the disease are two *Trypanosoma brucei* sub-species (Despommier, 1995). One manifests as a chronic (*T. brucei gambiense*) and the other, an acute (*T. brucei rhodesiense*), form of the disease (Fevre et al., 2006). These parasites are injected into

humans by the stinging bite of the tsetse fly (subgenus, Glossina). Figure 7 diagrams the life cycle of *T. brucei* (Centers for Disease Control and Prevention, (2003), African Trpanosomiasis Fact Sheet). Over a period of weeks to months, the long and slender flagellates continually divide and thrive in the blood stream. To evade the immune system, trypanosomes use an elaborate survival method involving gene shuffling to generate and then switch its antigenic coat proteins (Borst, 2002; Pays et al., 2004). Eventually, the flagellates pass through the blood-brain barrier. Left untreated, the disease is invariably fatal. At the height of a recent epidemic, in some communities of Angola, Democratic Republic of Congo, and Southern Sudan, sleeping sickness was the first or second greatest cause of morbidity (World Health Organization, (2006), African Trypanosomiasis Fact Sheet).

Recent efforts from private organizations and the WHO have been successful in curbing trypanosomiasis. The Programme Against African Trypanosomiasis (PAAT) estimates that a half million people are currently infected (World Health Organization, (2006), African Trypanosomiasis Fact Sheet). Currently, several drugs, such as the arsenical derivative melarsoprol, are readily available free of charge, but are marred with severe and fatal side effects. Also, these drugs require a patient to have access to a health center for regular intravenous injections. Drugs that are simpler to administer (such as oral administration) are being investigated.

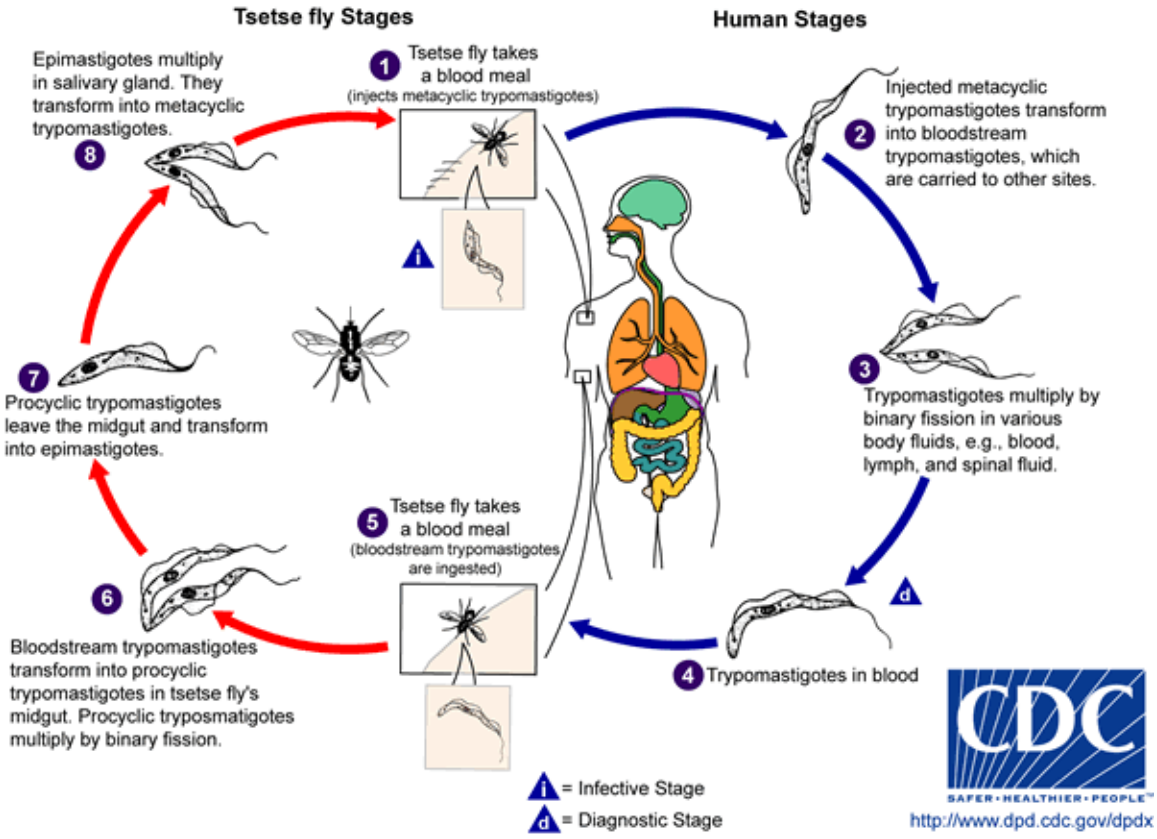


Figure 7: Cartoon schematic of the *Trypanosoma brucei* life cycle.

BIOLOGICAL APPLICATIONS OF MASS SPECTROMETRY

Mass spectrometry is central among the various techniques that we employed to identify and characterize the *T. brucei* NPC. For nearly a century, mass spectrometry (MS) has proven to be a powerful tool for elemental and small molecule analytical chemistry (Thomson, 1913). Since the development and the biological application of soft ionization techniques, such as matrix assisted laser desorption/ionization (MALDI) (Karas and Hillenkamp, 1988; Tanaka, 1988) and electrospray ionization (ESI) (Fenn et al., 1989), biological mass spectrometry has rapidly evolved over the past 25 years. Over the last decade, it has contributed significantly to the current understanding of protein interaction networks, localization, post-translational modifications (Andersen and Mann, 2006; Patterson and Aebersold, 2003), the identification of disease markers (Hanash, 2003) and drug discovery (Jeffery and Bogyo, 2003).

Large scale functional studies have been conducted based on a comprehensive analysis of yeast protein complexes (Gavin et al., 2002; Tong et al., 2002) and similar studies have been conducted with *E. coli* (Yan et al., 2002). While such endeavors produce a large quantity of data, subcellular proteome analyses provide additional insight into localization, function and interaction partners (Dreger, 2003). We have found this is to be especially true for divergent organisms because sequence similarity alone is often not sufficient to infer function. Indeed, over half of the *T. brucei* ORFs have no detectable homolog (El-Sayed et al., 2005). Direct subcellular interrogation thus aids in functional assignment. Isolating and purifying organelles has the added benefit of reducing

sample complexity by restricting the number of proteins sampled. The sequenced genomes of *S. cerevisiae* and the several vertebrates, coupled with subcellular fractionation, have allowed large-scale detailed studies of the NPC (Cronshaw et al., 2002; Rout et al., 2000), NE (Dreger et al., 2001) and nucleolus (Andersen et al., 2005) using biological mass spectrometry.

In this work, we identified most of the constituents of the NPC of an eukaryote that is distantly related to yeast and vertebrates — *Trypanosoma brucei*. We accomplished this endeavor through a combination of biochemistry, mass spectrometry, bioinformatics and molecular biology. We then made detailed comparisons between the NPCs of the metazoa, fungi and kinetoplastida. These comparisons begin to resolve the conserved aspects and the evolution of the NPC as well as elucidate the species-specific innovations that may be associated with chromatin organization and nucleocytoplasmic transport.

CHAPTER 2 — BIOCHEMISTRY AND PROTEOMICS: THE IDENTIFICATION OF PROTEINS THAT ASSOCIATE WITH THE *T. BRUCEI* NUCLEAR ENVELOPE PREPARATION

BIOCHEMISTRY AND MASS SPECTROMETRY — MATERIALS AND METHODS

We first sought to identify the constituent members of the *Trypanosoma brucei* NPC (TbNPC), and to this end, we used several biochemical and mass spectrometric techniques in order to maximize the yield. In the context of this work, we define a strategy as a specific set of biochemical and mass spectrometric techniques. Due to the complexity and dynamic range of the proteins associated with the *T. brucei* nuclear envelope preparation (TbNEP), we employed five strategies to ensure that we maximized the number of proteins identified. These strategies were initially selected to complement each method's strengths and weaknesses for a robust approach. Figure 8 summarizes the five strategies used to identify the constituents of the TbNEP, indicated by corresponding colored arrows. The adapted protocol we used to obtain the TbNEP (which includes the nuclear envelope and lipid-stripped pore complex lamina fraction, PCLF) from *T. brucei* (the black outlined box in Figure 8) is described in Appendix B.

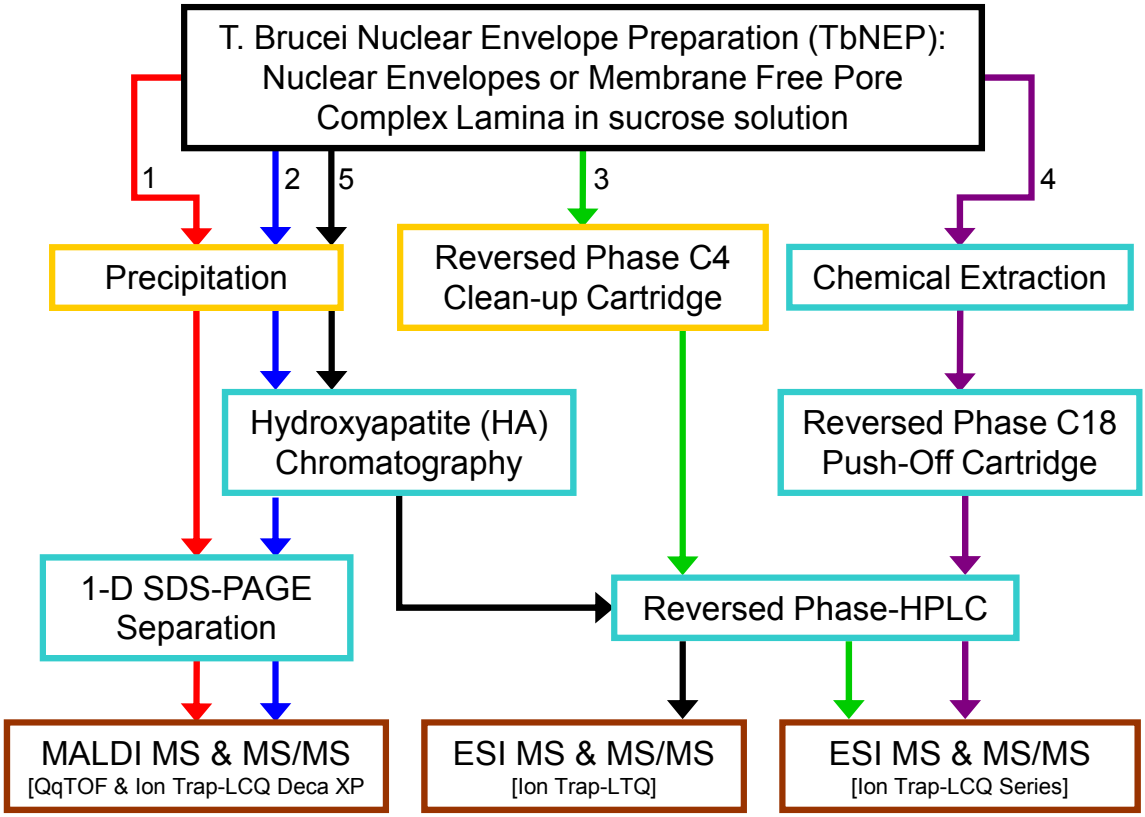


Figure 8: Summary flowchart of biochemical and mass spectrometric methods used to determine the TbNEP dataset. Strategies 1-5 are indicated by the red, blue, green, purple and black colored arrows, respectively. The black box represents the nuclear and subnuclear isolation protocol described in Appendix B. The gold boxes are protein recovery steps. Protein separation steps are colored in light blue. Finally, mass spectrometry techniques are outlined in brown. The ion trap mass spectrometers are in-house modified (LCQ Deca XP) or commercially available (LCQ and LTQ).

We chose several techniques to resolve and identify the proteins in the TbNEP. First, proteins needed to be recovered from the viscous sucrose and PVP solution. This is accomplished by either methanol precipitation or an offline C4 cleaning cartridge (gold boxes in Figure 8). To resolve the proteins, 5 separation techniques were used: SDS-PAGE, HPLC, hydroxyapatite (HA) chromatography (Gorbunoff, 1985), chemical extractions (Schirmer et al., 2003), and a novel C18 "push-off" cartridge, developed by J. C. Padovan (light blue boxes in Figure 8). Finally, the resolved proteins were identified by mass spectrometry (brown boxes in Figure 8).

STRATEGY 1 (FIGURE 8, RED ARROWS)

Methanol Protein Precipitation

To one volume of the TbNEP (1-5 ml), 5 volumes of HPLC-grade methanol were added and left to incubate for 4 hours at 4°C. The exact amount of sample material depends on the individual preparation and is empirically determined by pilot precipitations. Generally, one wants to use the maximum amount of material without saturating the SDS-PAGE gel or chromatography column. The precipitate was recovered by centrifugation (3300 (g) in a Beckman GH-3.8 for 15 minutes at 4°C). The pellet was resuspended with 500 µl of 90% methanol and transferred to a microcentrifuge vial and then left to incubate for 1 hour at 4°C. The suspension is spun one final time in a microcentrifuge (16,000 (g), 15 minutes, 4°C). The supernatant was removed to leave the protein sample pellet.

SDS-PAGE

The protein sample pellet was resuspended in 20 µl LDS sample buffer (Invitrogen, Carlsbad, CA), 8 µl 10x sample reducing agent (Invitrogen) and 52 µl water. After mixing, the solution was heated to 70°C for 10 minutes and allowed to cool to room temperature. To alkylate the reduced cysteines, 1 M iodoacetamide was added to a final concentration of 100 mM and the reaction was allowed to proceed for 30 minutes in the dark.

For strategy 1, NuPAGE® 10% and 4-12% bis-Tris gels (with MOPS running buffer) and Novex® 8% Tris-glycine gels (Tris-glycine running buffer) were used to increase resolution at specific mass ranges. For example, a Novex® 8% Tris-glycine gel offers high mass resolution while NuPAGE® 10% bis-Tris gels offer superior low mass resolution.

20 µl of alkylated sample was loaded to each gel. Electrophoresis was set at a constant 125V for 5 minutes followed by a constant 200V for 45 minutes. The gel was fixed in 50% methanol and 7% acetic acid for 15 minutes and then washed extensively. The proteins were visualized with GelCode® Blue colloidal Coomassie stain (Pierce, Rockford, IL) and documented by photography or digital flatbed scanning. Representative NE gels are show in Figure 9 and PCLF samples were similar.

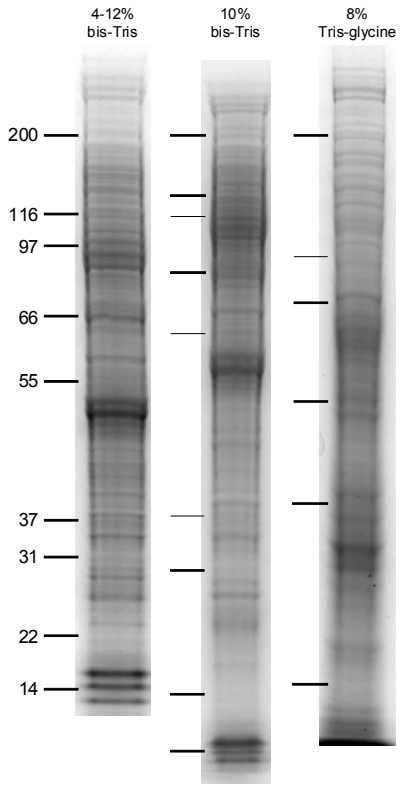


Figure 9: Representative gels of *T. brucei* NEs resolved on three different gel gradients as describe in Strategy 1.

In-Gel Digestion

On a white shallow plate or glass pane, the entire gel lane was cut into 2 mm bands using a MicroScalpel (Electron Microscopy Sciences, Hatfield, PA). Roughly thirty 2 mm bands were excised from a 10 cm gel after staining. Using the fine point tweezers (Electron Microscopy Sciences) or MicroScapel, the excised gel band was diced into 1 mm cubes and transferred to a microcentrifuge vial. As an alternative to manual slicing, one can use a Mickle® gel slicer (Brinkman Instruments, Inc.) which can mechanically slice the gel lane at 1 mm intervals.

The gel pieces were completely destained to remove all traces of stain and detergent. To the gel pieces, 500 µl of destain solution (25 mM ammonium bicarbonate in 50% acetonitrile) was added and the vials were agitated (medium setting) at 4°C with a vertical vortexer (Tomy Mixer, Tomy Seiko Co., Ltd., Tokyo, Japan). After destaining, 100 µl acetonitrile was added and then aspirated after 10 minutes. A trypsin (bovine, modified, sequencing grade, Roche Applied Science, Indianapolis, IN) aliquot was diluted in digestion buffer to a final concentration of 50 ng/µl. To the dehydrated gel pieces, ≥ 100 ng trypsin was added as well as 40 µl of 50 mM ammonium bicarbonate. The digestion mixture was incubated at 37°C for 4 hours.

POROS® R2 (Applied Biosystems, Foster City, CA) C18 resin is used to extract and recover the peptides from the gel pieces and digestion buffer. A working POROS R2 bead slurry is made as follows:

1. 500 mg of POROS R2 beads are sequentially washed with 10 ml of: (1) methanol, (2) 80% acetonitrile, and then (3) 20% ethanol.

2. Resuspend the washed beads in 20% ethanol to a final concentration of 50 mg/ml.

To 1 volume of POROS R2 bead slurry, 9 volumes of 2% trifluoroacetic acid, 5% formic acid in water was added and 40 µl of diluted POROS beads was mixed with the digest. The peptide/bead mixture was agitated (medium setting) in a vertical vortexer at 4°C for 4 hours.

The POROS beads are separated from the gel pieces with the use of ZipTips® (C18, Size P10, Millipore, Billerica, MA), which have been washed with elution solution (20% acetonitrile, 50% methanol, and 0.1% trifluoroacetic acid). The 80 µl peptide/bead mixture was transferred into a washed and conditioned ZipTip from the top and, using a syringe, the supernatant was discarded. 20 µl of 0.1% trifluoroacetic acid was added to the gel pieces and then transferred to the ZipTip from the top. The wash solution was expelled using a syringe. Wash the POROS beads on the ZipTip twice more using 20 µl of 0.1% trifluoroacetic acid.

The matrix used for MALDI mass spectrometry is 2,5-dihydroxy benzoic acid (DHB) (Lancaster, Pelham, NH). A saturated solution of DHB is prepared in elution solution at room temperature and diluted to 40% (v/v) saturated just prior to use. The peptides are slowly eluted onto a MALDI plate with 2.5 µl 40% DHB in elution solution to produce the “sample spot” (Archambault et al., 2003; Cristea et al., 2005; Tackett et al., 2005).

MALDI MS & MS/MS

Using the MALDI QqTOF mass spectrometer (Krutchinsky et al., 2000), MS data was acquired for each sample spot. The data reveals the mass-to-charge ratio (m/z) (the parent masses) for most of the peptides present in the sample. The spectra are then filtered with BackgroundFinder (written by Markus Kalkum) to remove background ions (such as trypsin autolysis peptides) and the resulting parent masses are transferred to method files that are compatible with the MALDI ion trap mass spectrometer (modified LCQ series, Thermo Fisher) (Krutchinsky et al., 2001).

Peptides are fragmented in the MALDI ion trap mass analyzer to produce daughter ions that yield mass data corresponding to the sequence, and hence identity, of the peptide. The protein may then be identified from this sequence data and the masses of the precursor peptides. We generated an in-house assembled *T. brucei* proteome database based on the raw data that was released from the *T. brucei* genome sequencing project (Berriman et al., 2005). We identified the proteins by comparing our data against this database with previously established proteomic bioinformatic tools, such as ProFound, Sonar, and X!Tandem (Craig and Beavis, 2004; Field et al., 2002; Zhang and Chait, 2000). The confidence of the identification is related to the quantity and quality of the mass data (Krutchinsky et al., 2001; Zhang et al., 2003).

STRATEGY 2 (FIGURE 8, BLUE ARROWS)

Methanol Protein Precipitation

As described in **Strategy 1**.

Hydroxyapatite Chromatography

In a 50 ml centrifuge tube, 7.5 ml Macro-Prep ceramic HA type I, 40 µm (Bio-Rad Laboratories, Hercules, CA), was washed with 20 ml of 200 mM Na_2HPO_4 (do not adjust pH). The HA was allowed to settle and the wash solution and suspended fine particles was aspirated. The wash was repeated 3 times with 20 ml loading buffer (10 mM NaH_2PO_4, pH 6.8, and 0.1 mM $CaCl_2$ in water), aspirating the loading buffer and suspended fine particles after each wash. To the final volume of HA, 4 volumes of loading buffer, supplemented with 0.1% SDS, was added.

Following methanol precipitation, the protein sample pellet was resuspended in HA sample buffer (10 mM Tris, 10 mM DTT and 2% SDS) and heated at 60°C for ten minutes. 1 volume of sample was diluted with 19 volumes of HA loading buffer. To the diluted sample, conditioned HA beads were added. Roughly 2 ml of bead slurry is required for less than 0.5 ml of sample. The mixture was incubated for 30 minutes. Keep the HA suspended by mechanical rotation or rocking.

The mixture was poured into a Poly-Prep® Chromatography Column (Bio-Rad, Hercules, CA) and the flowthrough was collected. The beads were washed

with 4 ml of 0.1% SDS in loading buffer. The wash was collected to monitor protein binding and combined with the flowthrough.

The proteins were then eluted from HA. All elution buffers are prepared from appropriate volumes of mobile phase A (1 mM DTT, 0.1 mM $CaCl_2$) and mobile phase B (1 M NaH_2PO_4, pH 6.8, 1 mM DTT). SDS is added, just before use, at a final concentration of 0.1%. The mobile phase was added in successive order to the column (300, 325, 350, 375, 400 and 500 mM NaH_2PO_4) and the eluate was collected. Generally, 4 ml of elution buffer is sufficient.

Trichloroacetic acid Precipitation

The final volume of eluate was adjusted to 10 ml with water and precipitated by sodium deoxycholic acid/trichloroacetic acid. The protein suspension was diluted with water to 1 ml (small scale) or 10 ml (large scale). Respectively, 100 µl or 1 ml of 0.3% sodium deoxycholic acid was added along with an equivalent volume of 72% trichloroacetic acid. After mixing, the solution was left to incubate at 4°C for 1 hour and then spun at maximum rotor speed (1 hour, 4°C). The pellet was washed in 1.2 ml acetone and transferred to a fresh microcentrifuge vial and stored overnight at -20°C. The pellet was recovered by centrifugation in a microcentrifuge (16,000 (g), 4°C, 1 hour).

SDS-PAGE

The same as described in **Strategy 1**, except in **Strategy 2**, 4-12% bis-Tris gels were used and the entire pellet from each fraction was loaded onto a lane within the gel. Representative gels from both the NE and PCLF are shown in Figure 10.

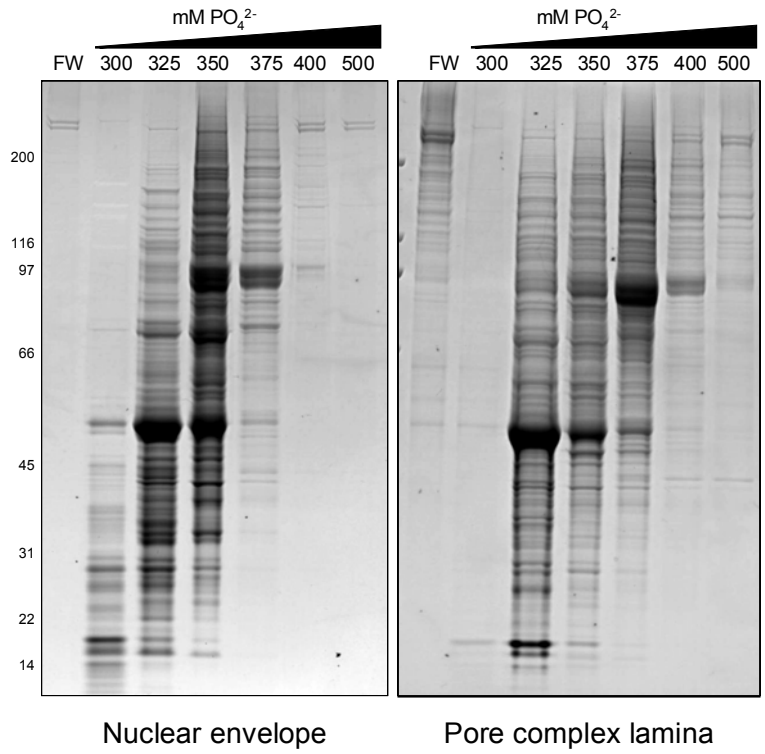


Figure 10: Representative hydroxyapatite gels. On the left is the nuclear envelope fraction and on the right is the lipid-stripped pore complex lamina fraction. FW, flowthrough and wash. Concentrations of phosphate in the elution buffer are indicated above the gel lanes.

In-Gel Digestion

As described in **Strategy 1**. All seven lanes of each gel were analyzed in their entirety by in-gel digestion and MALDI mass spectrometry.

MALDI-MS and MS/MS

As described in **Strategy 1**.

STRATEGY 3 (FIGURE 8, GREEN ARROWS)

Reversed Phase C4 Clean-Up Cartridge

To avoid the potential losses of methanol precipitation, a clean-up cartridge (Michrom Bioresources, Auburn, CA.) packed with C4 resin (Grace-Vydac, Hesperia, CA.) was used to concentrate the protein mixture. The C4 cartridge was prepared by washing 3x with 500 µl methanol, 3x with 500 µl 95% acetonitrile and 0.1% trifluoroacetic acid, and, finally, 3x with 500 µl 0.1% trifluoroacetic acid.

100 µl of the TbNEP was diluted with 500 µl of 0.1% trifluoroacetic acid and loaded onto the cartridge. The cartridge was then washed extensively with 0.5% acetic acid and 0.1% trifluoroacetic acid. After washing, the protein was eluted with 500 µl 95% acetonitrile and 0.1% trifluoroacetic acid. The eluate was then dried in a speedvac.

High Performance Liquid Chromatography

The protein sample pellet was resuspended in 50 µl of 50 mM ammonium bicarbonate. A total of 500 ng of trypsin was added and the solution was left to incubate at 37°C for 24 hours. The digestion was quenched by acidification with 0.1% acetic acid (final concentration).

10 µl (20%) of the digest was loaded onto a C18 reversed phase column (0.18 mm x 250 mm, 1.8 µl/min) and eluted under the following conditions: 25% solvent B (95% acetonitrile, 0.1% trifluoroacetic acid) for 5 minutes, 25-100% solvent B in 40 minutes. After the analytical run, the column was cleaned and conditioned with 100% solvent B for 5 minutes and then 25% solvent B for 20min. For this strategy, an Ultimate HPLC system (LC Packings-Dionex, Sunnyvale, CA) was used. Solvent A is 5% acetonitrile, 0.1% trifluoroacetic acid in water.

Electrospray ionization (ESI)-MS/MS

To ensure maximum coverage of the TbNEP, LC-MS/MS was employed as a complementary technique to MALDI MS/MS. The HPLC analytical run was online with the Finnigan LCQ series (ThermoElectron Corp., San Jose CA.) mass spectrometer. One data dependent MS/MS scan was acquired after each full (MS) scan.

Dynamic exclusion was used so that any given mass is analyzed and fragmented only once in a 30 second window. In addition to dynamic exclusion,

we used a second type of filter that uses prior knowledge of the contents of a sample. Briefly, since 20% of the sample was used for each run, the data from the previous run were analyzed. Those parent masses that were analyzed in the prior run were added to the "exclude masses" list in subsequent runs. Coupled to dynamic exclusion, this approach significantly reduced the number of redundant scans.

We generated an in-house assembled *T. brucei* proteome database based on the raw data that was released from the *T. brucei* genome sequencing project (Berriman et al., 2005). We identified the proteins by comparing our data against this database with previously established proteomic bioinformatic tools, such as ProFound, Sonar, and X!Tandem (Craig and Beavis, 2004; Field et al., 2002; Zhang and Chait, 2000).

STRATEGY 4 (FIGURE 8, MAGENTA ARROWS)

Chemical Extraction

Three separate separations were conducted on the TbNEP to reduce the complexity of the sample by enriching for transmembrane proteins (base extraction), integral proteins which are closely associated with the lamina (salt and detergent) and proteins that are peripheral to the NE (heparin) (Schirmer et al., 2003).

1. Base extraction
    a. To one volume of nuclear envelopes (in sucrose solution) 9 volumes of 100 mM NaOH with solution P (1:100) was

added and mixed completely by vortexing. Note: Solution P is made by dissolving 0.04% (w/v) pepstatin A (Sigma-Aldrich, St. Louis, MO) and 1.8% PMSF (Sigma-Aldrich, St. Louis, MO) in absolute (anhydrous) ethanol.

2. Salt and detergent extraction

    a. To one volume of nuclear envelopes (in sucrose solution) 9 volumes of salt and detergent extraction buffer (400 mM NaCl and 1% (w/v) β-octylglucoside in 25 mM Hepes, pH 7.5) with solution P (1:100) was added and mixed completely by vortexing.

3. Heparin extraction

    a. To one volume of nuclear envelopes (in sucrose solution) 9 volumes of 10 mg/ml heparin in bis-Tris/Mg buffer with solution P (1:100) was added and mixed completely by vortexing.

The extractions were left to incubate on ice for 1 hour. The extractions were then underlaid with 1 M sucrose in bis-Tris/Mg buffer with solution P (1:100) and spun at 103,460 (g) for 35 min. The extracted proteins are retained in the supernatant. The supernatant was carefully transferred to a fresh vial and precipitate the proteins with trichloroacetic acid/deoxycholate precipitation (**Strategy 2**).

Wash the enriched pellet in 1 ml acetone and transfer the suspension to a microcentrifuge vial store overnight at -20°C. Recover the pellet by centrifugation in a microcentrifuge (16,000 (g), 1 hour, 4°C).

The pellets from both the supernatant and the precipitate were checked on a gel (Figure 11). The extraction was repeated and the pellets were prepared for HPLC analysis. For HPLC analysis, the protein sample pellet was resuspended

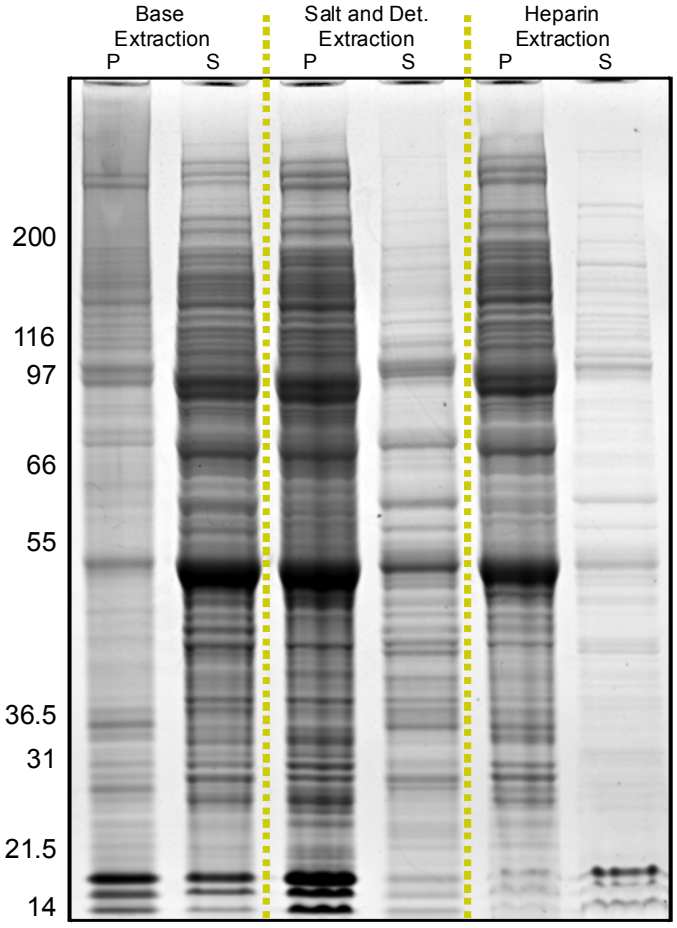


Figure 11: *T. brucei* NEs that have been subjected to chemical extraction. The three extractions (base, salt and detergent, and heparin) are separated by vertical dashed yellow lines. The pellet (P) and supernatant (S) are indicated.

in 50 µl of 50 mM ammonium bicarbonate. A total of 500 ng of trypsin was added and the solution was left to incubate at 37°C for 24 hours. The digestion was quenched by acidification with 0.1% acetic acid (final concentration).

Reversed Phase C18 "Push-Off" cartridge

5 µl of the peptide digest mixture was diluted 1 into 3 and passed through a ZipTip packed with POROS C18 resin, in a fashion analogous to **Strategy 1**. The amount of resin present is such that the binding capacity of the column is less than the total amount of peptide present in the digest.

The flowthrough is passed over a second ZipTip/POROS column with less resin than the first. The peptides bound to the first and second columns are eluted with 20 µl 95% acetonitrile, 0.1% trifluoroacetic acid. Both the eluates from columns 1 and 2 (enriched in the abundant peptides) and the flowthrough (enriched in less abundant peptides) are analyzed by HPLC. This protocol is credited to Júlio C. Padovan.

High Performance Liquid Chromatography

10 µl of the peptide mixture was loaded onto a C18 reversed phase column (0.2 mm x 50 mm, 2.2 µl/min) and eluted under the following conditions: 30% solvent B (95% acetonitrile, 0.1% trifluoroacetic acid) for 2 minutes, 30-100% B in 14 minutes. After the analytical run, the column is cleaned and conditioned at 100% solvent B for 7min and then 5% solvent B for 20min. For

this work, an Ultimate HPLC system (LC Packings-Dionex, Sunnyvale, CA) was used. Solvent A is 5% acetonitrile, 0.1% trifluoroacetic acid.

ESI-MS/MS

As described in **Strategy 3**.

STRATEGY 5 (FIGURE 8, BLACK ARROWS)

Methanol Protein Precipitation

As described in **Strategy 1**.

Hydroxyapatite Chromatography

As described in **Strategy 2**.

Trichloroacetic acid Precipitation

As described in **Strategy 2**.

High Performance Liquid Chromatography

From each fraction, the protein sample pellet was resuspended in 50 µl of 50 mM ammonium bicarbonate. A total of 500 ng of trypsin was added and the solution was left to incubate at 37°C for 24 hours. The digestion was quenched by acidification with 0.1% acetic acid (final concentration).

10 µl of the peptide solution was loaded onto a C18 column (BioBasic® PicoFrit C18 column 75 µm ID, (PFC7515-BI-5, New Objective, Woburn, MA)) and, after extensive washing with 0.1% acetic acid (solvent A), eluted with 70%

acetonitrile and 0.1% acetic acid (solvent B) in a linear gradient (0 - 100% solvent B) in 60 minutes at 20 µl/min. After the analytical run, the column was cleaned and conditioned with a blank run and 20 minutes of 100% solvent A.

ESI-MS/MS

The HPLC analytical run was online with a recently acquired Finnigan LTQ XL (ThermoElectron Corp., San Jose CA.) mass spectrometer, having enhanced performance compared with the Thermo Finnegan LCQ instrument used to acquire data in **Strategies 3** and **4**. The PicoFrit column has a 15 µm emitter tip which enables the eluate to be directly analyzed via ESI-MS/MS.

The method was designed to acquire one MS scan to determine parent masses, and then acquire 10 MS/MS scans which are dependent on the first MS scan. Dynamic exclusion was used so that any given parent mass is analyzed and fragmented only once in a 30 second time window.

We generated an in-house assembled *T. brucei* proteome database based on the raw data that was released from the *T. brucei* genome sequencing project (Berriman et al., 2005). We identified the proteins by comparing our data against this database with previously established proteomic bioinformatic tools, such as ProFound, Sonar, and X!Tandem (Craig and Beavis, 2004; Field et al., 2002; Zhang and Chait, 2000).

RESULTS AND DISCUSSION

By employing five complementary strategies, we were ensured a comprehensive analysis of the nucleoporins within the TbNEP. As described above, we compared our mass spectrometric data to our in-house assembled *T. brucei* proteome database. All identifications with an expect value worse than 10^{-5} were validated by hand if we determined that the protein identified was of interest to this body of work. We used the following criteria to ensure the confident identification of proteins. For those identifications with an expect value worse than 10^{-3}, at least one peptide must have an expect value better than 10^{-2}. In addition, in order for an identification based on just a single peptide to be considered, that expect value must be better than 10^{-3}.

In total, we identified 859 proteins in the *T. brucei* nuclear envelope preparation (TbNEP), the full inventory of which is listed Appendix C. This dataset represents ~10% of the *T. brucei* proteome (Berriman et al., 2005), and likely contains a large number of contaminants. We experimentally identified 5 TbNups and 6 NPC transport-related proteins, which were previously annotated by homology during the genome sequencing project (Berriman et al., 2005). No annotated Nups were missed. Several previously annotated transport-related proteins were missed, most likely due to fact that Kaps only interact transiently with the NPC. The number of proteins identified in this work far exceeds the number that we anticipated to be present in the NE. The first reason is that the TbNEP is an enriched, rather than a pure, preparation. The second is that we

used enhanced MS-based techniques to increase the dynamic range of our analysis to the point that we could detect proteins of low relative abundance.

With respect to the number of proteins identified, the strategies varied in their success. When we employed SDS-PAGE with MALDI-MS readout in **Strategies 1** and **2**, we identified 63 and 71 proteins, respectfully. We did not use SDS-PAGE to separate the proteins in **Strategies 3** and **4**, yet we identified 84 and 164 proteins, respectively. The most successful strategy that we employed was **Strategy 5**, which identified fully 588 proteins. In retrospect, **Strategy 5** may have been sufficient to completely analyze the TbNEP. The success of this strategy is attributed, in part, to the sequential chromatographic separation of the proteins (using hydroxyapatite chromatography) and the peptides that result from trypsin protease digestion (using C18 HPLC). Also, for **Strategy 5**, we used a linear ion trap mass spectrometer, which is more efficient and sensitive than the hyperbolic ion trap mass analyzer used in **Strategies 1-4**.

We identified all of the TbNups in two or more strategies. It is noteworthy that, although **Strategy 5** identified greater than 4X as many proteins as any other strategy, no additional Nups were identified in this most sensitive analysis. We take this finding to imply that our analysis of the NPC components is essentially saturated.

Using these five different strategies allows us to make some observations as to the relative merits and limitations of the techniques used. Complex protein samples with a large number of constituents are ill-suited to SDS-PAGE with

MALDI-MS readout (**Strategies 1** and **2**), as these methods do not have the capacity to handle the large mixtures of proteins. For example, consider that loading 650 proteins on a pre-cast gel lane, roughly 6.5 cm long, results in an average 10 proteins per linear mm. Furthermore, the distribution of proteins within the lane is not linear, with the densest region being in the middle (30-100 kDa) region of the gel. If one analyses a 1 mm gel slice from the center of the lane, then one is working with a pool of perhaps 15 or more proteins. Assuming ~50 peptides per mid-sized protein, then 750 peptides are generated from this mixture of proteins. This level of complexity is beyond the ability of our current MALDI-MS analysis techniques.

The range of protein concentration also presents a challenge if low abundance peptides cannot be detected in the mass spectrometer above the background chemical noise. One method for improving coverage in such samples is to work with more sample material (Eriksson and Fenyo, 2007). However, SDS-PAGE systems have an upper limit to the amount of protein that may be loaded. Overloading a gel leads to loss of resolution, smearing and lane collapse.

To counter the limitations of SDS-PAGE protein separation and MALDI-MS readout, we utilized gel-free methods of protein separation in combination with LC-MS. Although **Strategy 3** offered slightly superior results to SDS-PAGE with MALDI-MS readout, this methodology was not robust in my hands due to frequent clogging of the C4 cartridge. **Strategies 4** and **5** yielded the best results due to increased protein separation prior to digestion and subsequent peptide

separation. Another reason why **Strategy 5** proved superior is that the analyte was analyzed with a linear ion trap mass spectrometer, which has a higher ion capacity, sensitivity and scanning speed (which translates into more MS/MS spectra per unit time). Combined, these factors greatly contribute to the number of co-eluting peptides that can be identified in a complex sample.

In **Strategy 4**, the second best strategy, we employed a methodology developed by Júlio Padovan, known as the "push-off" cartridge. Briefly, a large peptide mixture is passed over a stationary phase (here, C18 resin) with a binding capacity less than the total amount of sample loaded. This sets up a situation in which peptides are forced to compete for the limited number of binding sites. We observed that the population of peptides within the flowthrough is different than that population bound to the resin. Thus, after passing the peptide mixture over the "push off" cartridge, the end result is that an additional separation step is introduced to the sample. One can then load more of the eluate or flowthrough onto the HPLC column, as it has been depleted of a significant population of peptides.

We demonstrate this separation in the base peak (BP) trace of the three analytical runs resulting from the base extracted TbNEP (Figure 12). Here, the sample is first passed through one "push-off" cartridge and then flowthrough is passed over a second "push-off" cartridge for further separation of peptides. The bottom panel shows the BP for the flowthrough. The middle panel shows the material eluted off the second cartridge, while the top represents the sample eluted from the first cartridge. These BP traces are different for each of the three

samples. Strikingly, the peptide at m/z ~1587 (indicated by the arrow) is an abundant peptide that has been depleted from the flowthrough, thus allowing less abundant peptides with the same retention time to be measured in the ion trap mass analyzer.

With the constituents of the TbNEP identified, we next used a series of bioinformatic screens to identify and characterize the bona fide components of the TbNPC. The next chapter will discuss the bioinformatic strategies and tools used to infer function to the set of 859 proteins associated with the TbNEP, with an emphasis on identifying the putative TbNups.

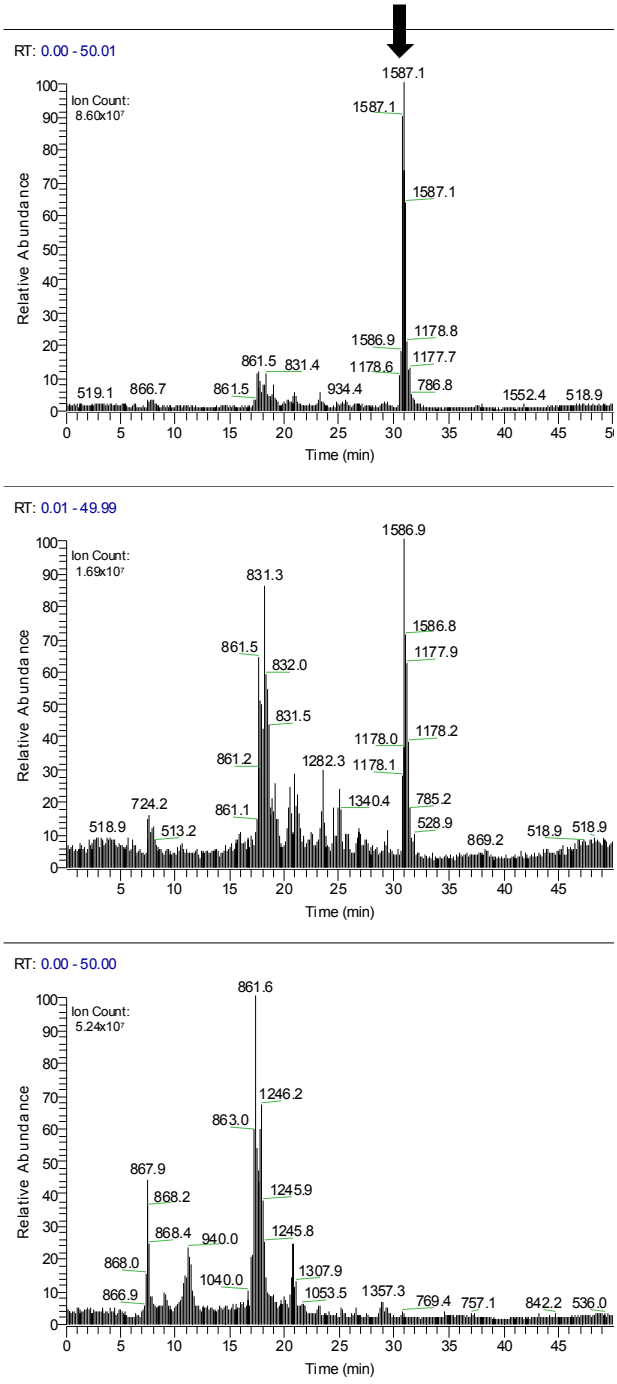


Figure 12: Separation of peptide populations in a complex mixture by using a series of “push-off” cartridges. The base peak readout (BP) shows different patterns for the material bound to the first cartridge (top), the second cartridge (middle), and peptides which did not bind to either cartridge (bottom).

CHAPTER 3 — THE BIOINFORMATIC ANALYSIS OF THE TBNEP

STRATEGY AND ALGORITHMS

Using several biochemical and proteomic techniques in five different strategies, we have identified ~860 proteins in the TbNEP. The next challenge was to identify the putative TbNups present in our dataset. Inferring protein function by homology between distantly related species is particularly difficult because significant primary structure similarity has been lost over time. To circumvent this formidable challenge, we employed a range of bioinformatic algorithms to characterize and infer function to distant orthologs and identify the putative TbNups.

To simplify the TbNEP dataset, we cross referenced each identified protein to its functional annotation page on GeneDB, the annotation database of the Wellcome Trust Sanger Institute. If a protein has been functionally annotated, and found not to be TbNPC related, we set the protein aside. There is a possibility that these proteins may functionally interact with the nuclear envelope, but further experimental studies would be required to demonstrate such an interaction or localization. In this way, we were able to set aside ~500 proteins.

We characterized the remaining ~350 functionally unannotated proteins using the following strategy:

1. We first analyzed, by pair-wise sequence alignments, the ~350 functionally unannotated proteins identified by mass spectrometry.

The sequences were queried using both PSI-BLAST (Altschul et al., 1997) and FASTA (Pearson and Lipman, 1988)) against the National Center for Biotechnology Information (NCBI) non-redundant database. The alignments were seeded with a word size of 2, and scored against the BLOSUM45 matrix. The resulting alignments were individually inspected by hand.

2. Querying the domain architectures between distant orthologs is more sensitive than querying the entire primary structure (Bateman et al., 2000). To identify domains, we conducted a Hidden Markov Model (HMMer (Eddy, 1998)) alignment, using standard parameters, to the Pfam HMM-profile database of domain families (Sonnhammer et al., 1998).

3. We also scanned both the trypanosome protein database and the TbNEP dataset for the presence of FG-repeat domains by using a simple pattern recognition algorithm written by David Fenyö (PROWL). We did not find any recognizable FG-repeat domain in the trypanosome database that was not also present in the TbNEP dataset.

4. Previous work showed that the Nups fall into 8 major fold types with characteristic secondary structure patterns (Devos et al., 2006). To identify the presence of these specific fold types, we predicted the secondary structure of the unannotated proteins (PSI-PRED

(McGuffin et al., 2000) and PROF (Ouali and King, 2000)) on all the unannotated protein sequences. When interpreting the results, one should bear in mind that these algorithms require primary structure similarity for accurate prediction. For the purposes of this study, the reduced accuracy is acceptable since we are not concerned by the details of element size and boundaries. We are concerned only whether a domain is primarily β-sheet rich or α-helix rich.

5. We also predicted the presence of several motifs with various prediction algorithms. We concentrated on the motifs that are indigenous to the NPC and NE, which include transmembrane helices (Phobius (Kall et al., 2004)), natively disordered regions (Disopred (Ward et al., 2004) and PONDR (Romero et al., 2001)), coiled coils (COILS (Lupas et al., 1991)), and putative nuclear localization sequence (NLS, Nucleo (Hawkins et al., 2007)). We accepted results that had better than an 80% predictive confidence score, based on the benchmarks of the individual algorithm.

6. For several known domains and conserved sequences, multiple sequence alignments were conducted with ClustalX v1.83 using default settings (Thompson et al., 1997). Conserved residues are indicative of a functional role within the NPC and nucleocytoplasmic transport and, whenever possible, were cross checked with the literature for any mutational analysis.

SORTING THE *T. BRUCEI* NUCLEAR ENVELOPE PREPARATION DATASET

By utilizing a suite of bioinformatic tools, we were able to characterize the proteins that were identified in the TbNEP. We excluded those proteins whose function was reasonably known and are not associated with the TbNPC. By mining the *T. brucei* protein functional annotation database (GeneDB, Wellcome Trust Sanger Institute), as well as PSI-BLAST searches to the NCBI nr database, we could exclude ~500 proteins that fall within these criteria (listed in Table 8).

At this point, sequence homology alone is sufficient to identify only five constituents of the TbNPC (TbSec13, TbNup96, TbNup158, TbNup144, and TbNup62). We concluded that information about the domains, motifs and secondary structure would be useful in identifying the remainder of the TbNups. This is based in part on recent work which has shown that the yeast and vertebrate NPCs are highly modular in nature and the entire catalog of nucleoporins utilizes only 11 domains or motifs (Devos et al., 2006).

By searching for these modules using the algorithms outlined in the previous section, rather than sequence similarity, we identified an additional 19 putative TbNups, for a total of 24 TbNups (listed in Table 5). Based on comparisons with the Nup inventories of NPCs from vertebrates and yeast, we estimate that we have identified at least 80%, by mass, of the TbNPC. We anticipate that the balance is most likely species-specific or highly divergent Nups, which would be difficult to identify within the TbNEP dataset by comparative methods. In the future, we plan to utilize the known TbNups as

"bait" to identify these elusive TbNups through interaction studies, such as immunoprecipitation.

In addition to the 24 TbNups, we identified 8 transport factors (Table 5). These proteins are somewhat easier to identify than the TbNups because of their relatively high sequence homology across Eukaryota. It is not surprising that only a subset of transport factors were identified in this work, since they are soluble proteins and only transiently interact with the NPC.

Aside from the ~500 annotated, but unrelated, proteins and the 32 proteins associated with the TbNPC, we attempted to characterize the remaining 330 unannotated proteins within the TbNEP dataset (Table 7). Using the strategies noted above, we found 23% of these proteins contain at least one coiled coil and 9% of the proteins contain a likely nuclear localization sequence (NLS). We identified at least one Pfam domain in 33% of the unannotated proteins. The details of these domains and motifs are included within Table 7.

Finally, one conspicuous group that remains elusive are membrane bound proteins, including integral membrane Nups. While 30% of the unannotated proteins within the TbNEP dataset are predicted to have at one transmembrane helix (TMH), these proteins have no obvious discernible domain structure characteristic of an NPC (e.g. the cadherin domain within ScPom152 and Hsgp210) or NE constituent protein. For example, we are unable to find, within the TbNEP dataset, a homolog of the "Sad1/Unc" domain, which is found in the SUN/Unc-84 NE proteins found in the opisthokonts, or the "LEM" domain of

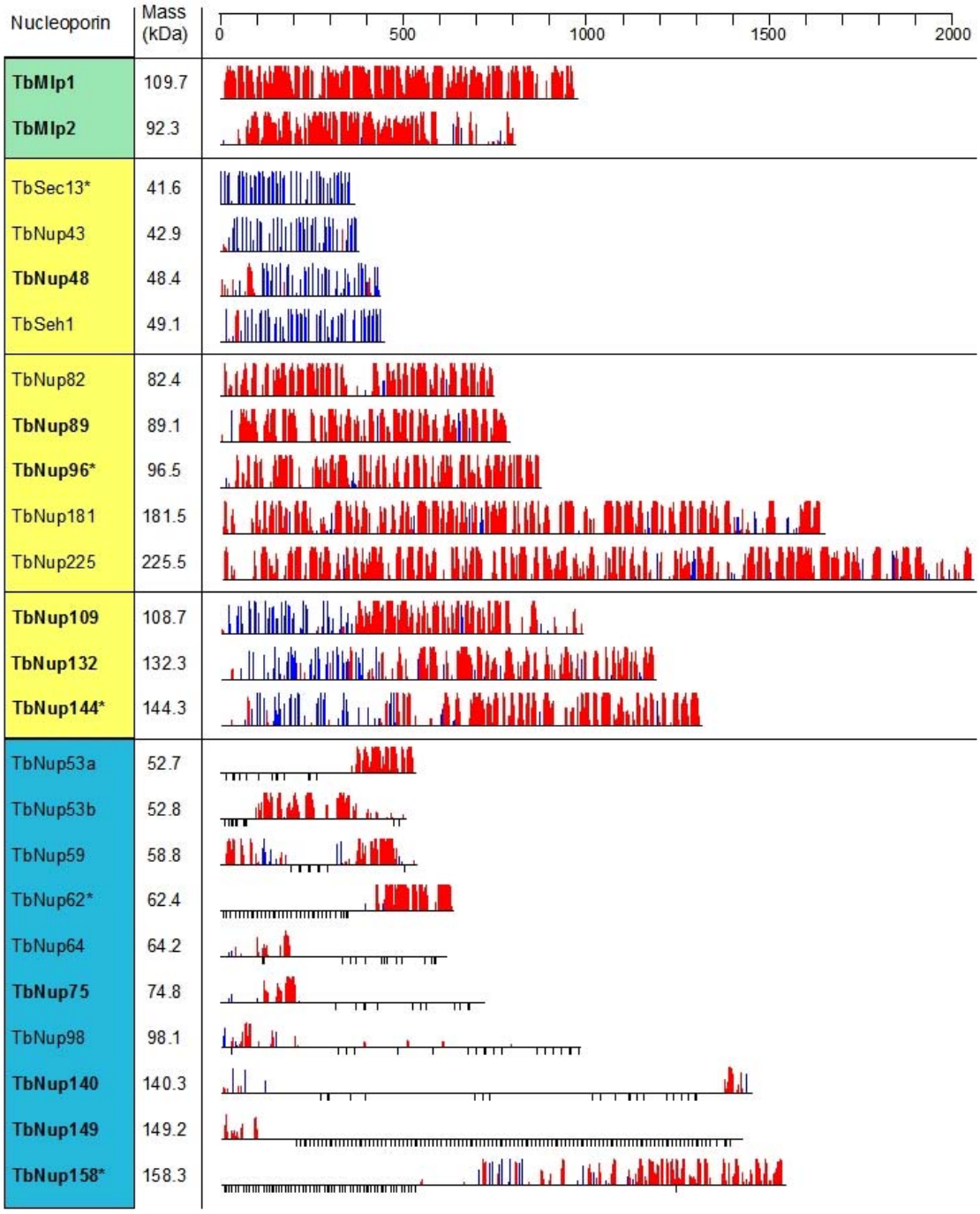
LAP2, which is also present in the NE of the opisthokonts. We searched the *T. brucei* functional annotation for any NE related domain which is not present in the TbNEP dataset. Thus far, only one protein was found to contain a "Sad1/Unc" domain (Tb927.6.1740). It is not clear why this protein was missed in our analysis.

The TbNPC lacks any obvious orthologs to the more divergent, species-specific Nups, such as HsNup358, ScNup2, HsNup214/ScNup159, and HsNup88/ScNup82. These Nups have domain and motifs structures that should be easily recognized. For example, HsNup358 has a specific domain architecture, including Zn-finger motifs, and HsNup88/ScNup82 contains a β-propeller domain. It is possible that the *T. brucei* homologs to these Nups have either been lost or have diverged to such a degree that even domain structure has adopted species-specific innovations. These Nups will need to be identified experimentally. In the future, we can target known TbNups and perform immunoprecipitations, in tandem with mass spectrometry, and identify the interacting partners of the entire TbNPC. These assays should be specific enough to confidently map the interaction network of the TbNPC and fill any gaps in our model of the TbNPC.

IDENTIFYING AND CHARACTERIZING THE *T. BRUCEI* NUCLEAR PORE COMPLEX

Figure 13 lists the putative TbNPC constituents along with their predicted secondary structures. We characterized the TbNups based on their predicted secondary structure patterns. The presence, for example, of an α-solenoid domain in a putative Nup suggests that it is a member of the structural scaffold group and is an ortholog of an α-solenoid-containing Nup found in the opisthokonts. As discussed above, we determined that sequence homology alone is not sufficient to make such characterizations with the majority of TbNups.

Figure 13: (Following page) Secondary structure map of the putative TbNups. Protein names in bold indicate those which have been localized to the TbNPC *in vivo* (Chapter 4). The asterisk indicates those Nups that share significant homology with the opisthokonts. The top ruler indicates residue number. Within a map, the horizontal black line represents the primary structure of the Nup with the N-terminus to the left. The y-axis indicates the confidence score of the secondary structure element prediction. Predicted α-helices are shaded in red, and predicted β-sheets are in blue. The black vertical lines below the primary structure indicate the FG-dipeptides. The TbNups are binned according their predicted function within the TbNPC; green, probable nuclear basket; yellow, structural scaffold; blue, FG-Nups.


Nucleoporin
Mass (kDa)
0
500
1000
1500
2000
TbMlp1 109.7
TbMlp2 92.3
TbSec13* 41.6
TbNup43 42.9
TbNup48 48.4
TbSeh1 49.1
TbNup82 82.4
TbNup89 89.1
TbNup96* 96.5
TbNup181 181.5
TbNup225 225.5
TbNup109 108.7
TbNup132 132.3
TbNup144* 144.3
TbNup53a 52.7
TbNup53b 52.8
TbNup59 58.8
TbNup62* 62.4
TbNup64 64.2
TbNup75 74.8
TbNup98 98.1
TbNup140 140.3
TbNup149 149.2
TbNup158* 158.3

THE STRUCTURAL SCAFFOLD

About a dozen Nups play a principally structural role within the NPC (Alber et al., 2007a). These Nups form an eight-fold symmetric ring upon which the FG-Nups are layered (Maul, 1971; Suntharalingam and Wente, 2003). These Nups contain a relatively simple domain architecture, either containing an α-helix rich α-solenoid domain, a β-sheet rich β-propeller domain, or both (Devos et al., 2006). A yeast model of these structural Nups is shown in Figure 14 (Alber et al., 2007a). In our search to identify TbNups from the TbNEP dataset, we identified 12 structural TbNups which are described and characterized below.

β-Propellers

β-propeller proteins are characterized by a propeller-like structure consisting of 6 or 7 blades, each of which is comprised of four anti-parallel β-sheets (Neer et al., 1994). Members of this family from human and yeast include HsSec13/ScSec13, HsSeh1/ScSeh1, HsNup37, HsNup43, and HsALADIN. HsNup37, HsNup43 and HsALADIN are found throughout the metazoa and the HsALADIN locus is linked to Allgrove syndrome in humans (Cronshaw and Matunis, 2003). HsALADIN and HsNup43 are also present in plants, a bikont. ScGle2/HsRae1 also falls into this structural category, although it only transiently interacts with the NPC (Bharathi et al., 1997; Murphy et al., 1996).

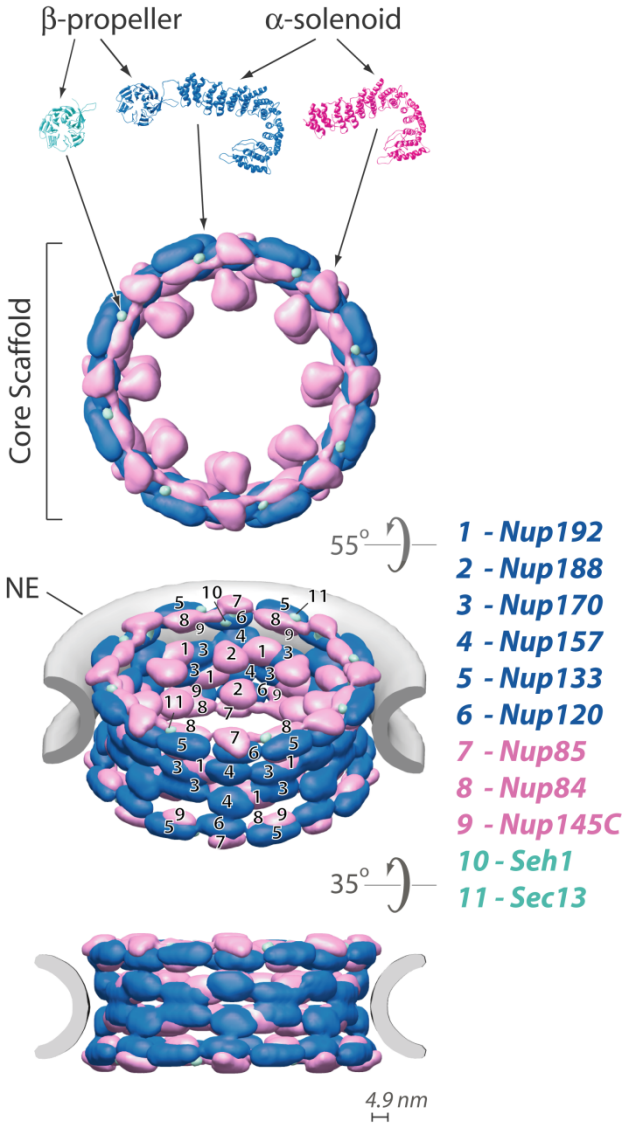


Figure 14: A model of the yeast structural scaffold. The scaffold is comprised of 3 fold types: WD-repeat rich β-propeller, α-helix rich α-solenoid, and an N-terminal β-propeller followed by a C-terminal α-solenoid (Alber et al., 2007a).

In addition to the Gle2, Sec13 and Seh1 orthologs, *T. brucei* has at least two additional β-propeller Nups. This set is more akin to the vertebrates (which have 6 β-propeller) than to yeast (3 β-propeller). The secondary structure prediction of these putative TbNups show similar patterns of β-sheets to the yeast and vertebrate homologs. To map and visualize the similarities between the β-propeller-containing Nups, we generated an alignment matrix (Figure 15). All β-propellers from *S. cerevisiae*, *H. sapiens*, *A. thaliana*, and *T. brucei* were pair-wise aligned with every other Nup using the FASTA algorithm. The Smith-Waterman scores, a measure of similarity between two aligned sequences, were then normalized to the best score that resulted from a non-paralogous pair-wise alignment. The normalized scores are visualized with a grayscale gradient, with the best score shown in black. Paralogous pairs have a normalized score greater than 1 and are also shown in black. Such a matrix is useful to identify homologs.

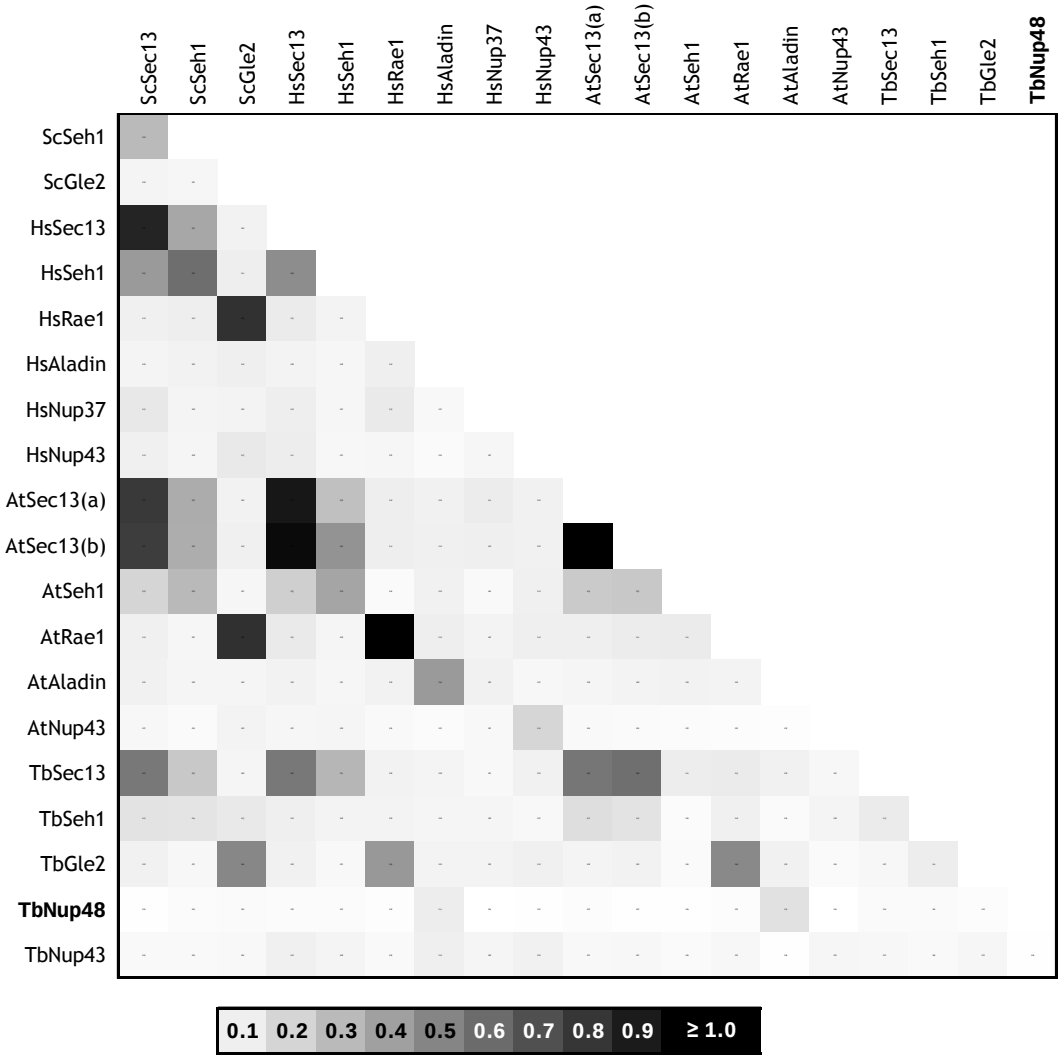


Figure 15: Pair-wise alignment matrix of the β-propeller Nups. The Nups are aligned using the pair-wise FASTA alignment algorithm with default settings. Bold names indicate those Nups which localize to the NE. The Smith-Waterman score, a measure of the similarity between two aligned sequences, is recorded at the intersection of each pair of Nups. The scores are normalized to the highest score between two homologous (but not paralogous) Nups. The normalized score is then visualized with a grayscale gradient, which is indicated below the matrix. Paralogous pairs are indicated in black but have a normalized score greater than 1. Sc, *S. cerevisiae*; Hs, *H. sapiens*; At, *A. thaliana*; Tb, *T. brucei*.

The alignment matrix confirms the highly conserved nature of Sec13, Gle2/Rae1, and, to a lesser extent, Seh1 throughout all four species. Seh1 is very similar to Sec13, but contains a characteristic extended N-terminal sequence. Interestingly, while *T.brucei* Sec13 and Gle2 are highly homologous to the yeast, vertebrate and plant orthologs, the other 3 (TbSeh1, TbNup43, and TbNup48) share little homology to plants, yeast and humans. However, the sequences contain a predicted WD-repeat sequence and secondary structure prediction indicates a high probability that those proteins would fold into a β-propeller domain. One such protein, TbNup48, which shares a very weak similarity to HsALADIN, localizes to the nuclear periphery *in vivo* (Chapter 4).

Sec13 plays a dual role within the cell (at the NPC and also as a vesicle coat component), so it is not unexpected that its primary structure has remained highly conserved throughout Eukaryota (Siniossoglou et al., 1996). Gle2/Rae1, an essential trafficker of mature mRNA, is also highly conserved. Using multiple sequence alignments, we found that all residues within Gle2 shown to be essential for proper function are conserved in trypanosomes (Moy and Silver, 2002; Murphy et al., 1996; Yoon et al., 1997). We plan to localize the remaining putative β-propeller Nup, Tb Nup43, in the future.

α-Solenoids

The second family of structural scaffold Nups contain an α-solenoid domain. α-solenoids are characterized by their stacked, anti-parallel, α-helices (Andrade et al., 2001a; Andrade et al., 2001b). The yeast and vertebrate members are ScNup84/HsNup107, ScNup85/HsNup75, ScNIC96/HsNup93, ScNup145c/HsNup96, ScNup192/HsNup205, and Nup188 (Fontoura et al., 1999; Siniossoglou et al., 1996). We identified six putative α-solenoids from within the TbNEP dataset. Figure 16 shows the alignment matrix for the α-solenoids.

The six *T. brucei* α-solenoids are predicted to be Nups based on the their similarity, both in length and domain architecture, to yeast and vertebrates. The primary structures of the trypanosome α-solenoids have diverged significantly from their vertebrate and yeast counterparts. This is especially pronounced when compared to the more conserved β-propellers. The nucleoporin interacting component, or NIC domain found in TbNIC96, is the only conserved Pfam domain found within the α-solenoids. The NIC domain shares modest homology to the bikonts. The Pfam expect values for the alignment between the NIC domain and the trypanosome and plant orthologs is 10^{-5} and 10^{-10}, respectively (compare to humans [10^{-177}], yeast [10^{-166}]). The full length alignment of ScNIC96/HsNup93 to TbNup96 is thus far less significant than the alignments between yeast, vertebrates, and plants (Figure 16). Apparently, what little homology is shared between the α-solenoids of the vertebrates, yeast and plants becomes irresolvable with the distant excavates.

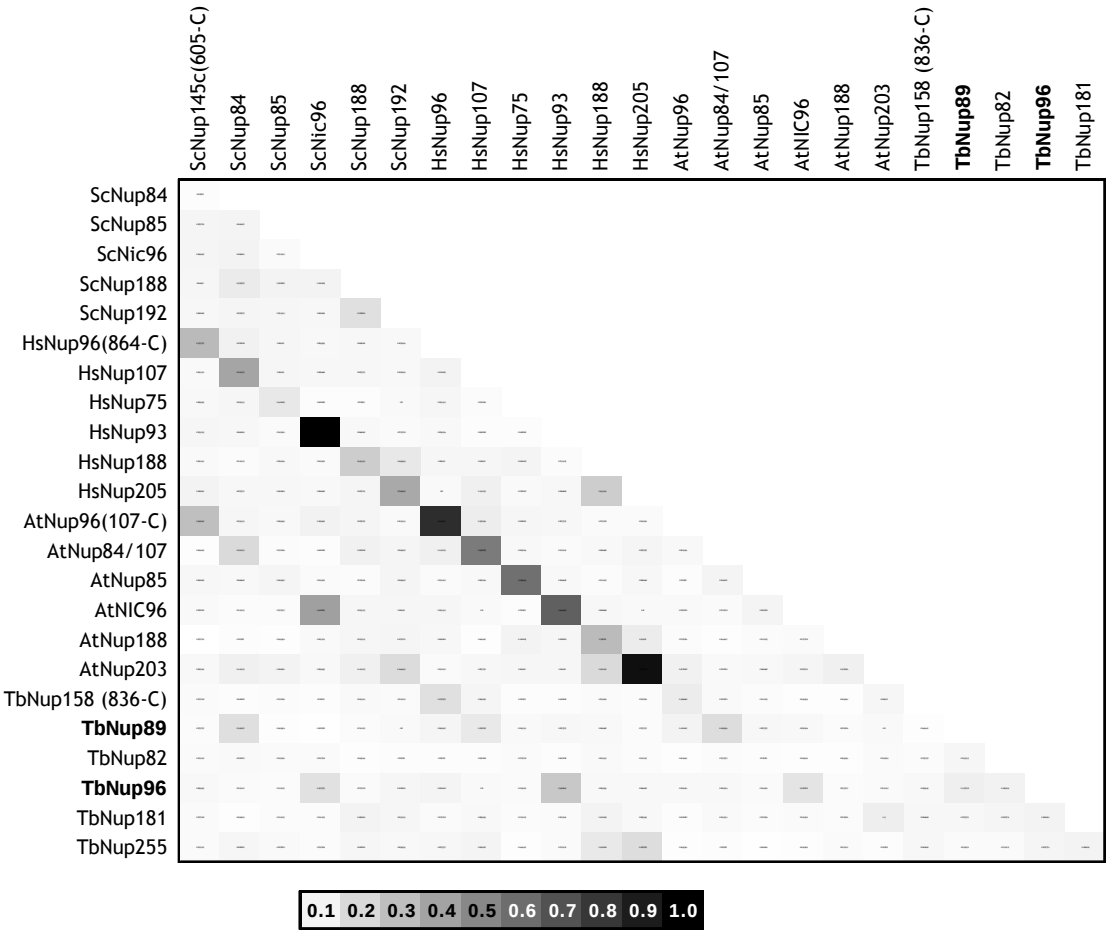


Figure 16: Pair-wise alignment matrix of the α-solenoid Nups. See Figure 15 for the description.

N-terminal β-Propeller Domain Connected to a α-Solenoid

The third and final structural scaffold family consists of an N-terminal β-propeller domain connected to an α-solenoid domain (β-α domain). Nups in this family include ScNup120/HsNup160, Nup133, and ScNup157/ScNup170/HsNup155. The pair-wise alignment matrix is shown in Figure 17. Interestingly, save for HsNup155 orthologs, these Nups share little to no homology between yeast and vertebrates and plants. However, vertebrates and plants share significant homology between orthologous pairs. We determined that *T. brucei* has three Nups of this family, which were determined primarily through their predicted secondary structure, except for TbNup144 which shares moderate but significant sequence similarity to HsNup155.

The high sequence similarity and secondary structure similarity of the Nup155 orthologs extends across the four species. While ScNup157/ScNup170/HsNup155 are homologous along their entire sequence, multiple alignments to Nup155 show high similarity primarily in the WD repeat region of the β-propeller and sporadic islands of single residue conservation within the α-solenoid interface. Also, from multiple alignments, we found that the Gle1 (mRNA export factor) binding site, proximal to the C-terminus of HsNup155, appears to be an opisthokont innovation (Rayala et al., 2004). We predicted, based on their length and domain structure, that two other members of the TbNEP dataset (TbNup132 and TbNup109) are β-α Nups. Indeed, we show that

TbNup144, TbNup132, and TbNup109 all localize to the nuclear periphery *in vivo* (Chapter 4).

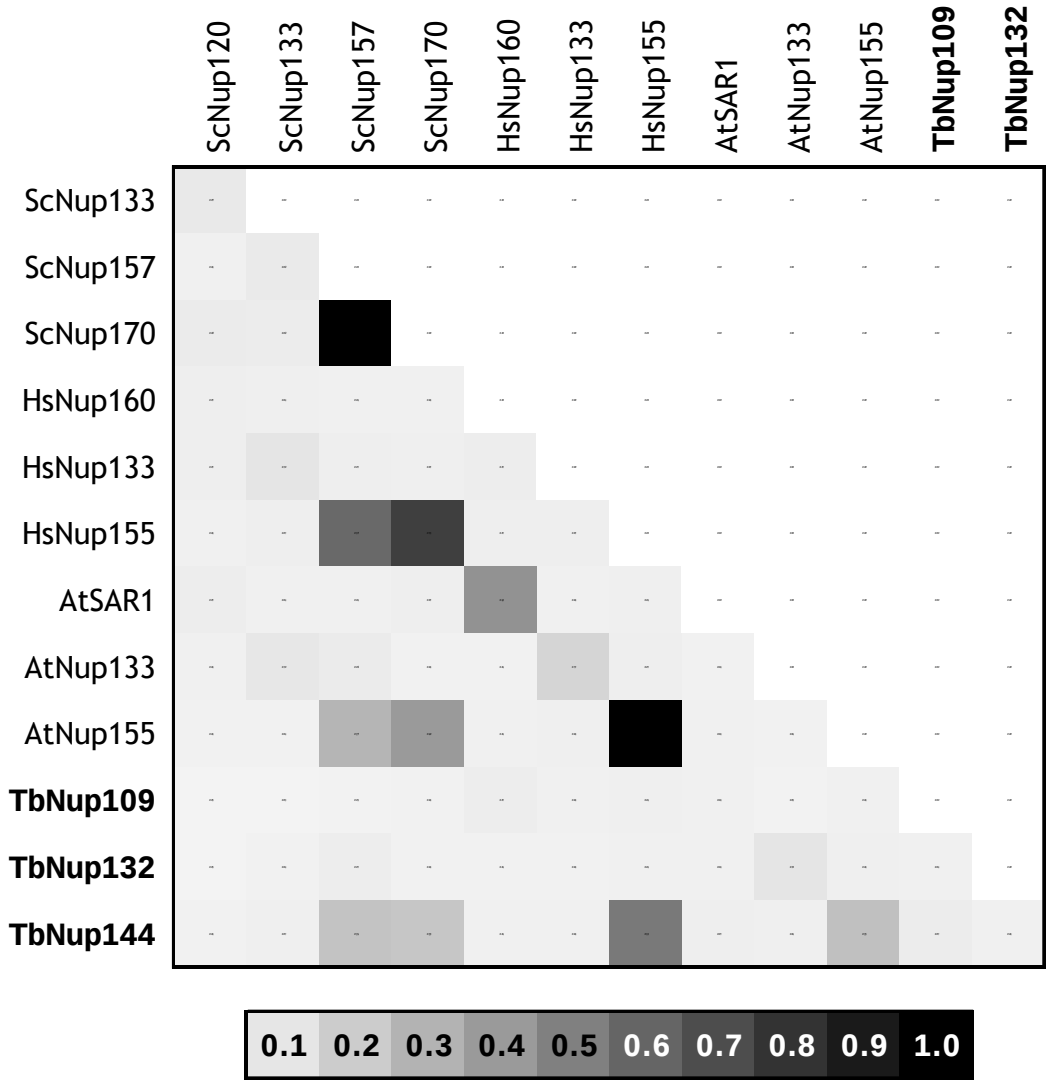


Figure 17: Pair-wise alignment matrix of the β-α Nups. See Figure 15 for the description.

THE NUCLEAR BASKET: MLPS/TPR

Large coiled coil proteins likely form the nuclear basket of the NPC and also interact with the NE and, in yeast, with the spindle pole body (Niepel et al., 2005; Strambio-de-Castillia et al., 1999). ScMlp1/2 and HsTpr are well-conserved (Figure 18). In trypanosomes, the coiled coil protein TbNUP-1 has been shown to interact with the NE and exhibit some lamin-like properties, despite the lack of lamin associated Pfam domains (Ogbadoyi et al., 2000; Rout and Field, 2001). We found by sequence similarity two putative paralogs of TbNUP-1, although whether they localize to the NE is unknown at this point.

We predicted that two proteins within the TbNEP are putative Mlp/Tpr orthologs, based on their size and coiled coil nature. Our subsequent localization supports this (Chapter 4). They exhibit very little sequence similarity to Mlp/Tpr and do not contain a Mlp/Tpr Pfam domain. This is interesting considering the high similarity of Mlp to Tpr, and that they are coiled coils, which typically offer very strong alignments (albeit, falsely).

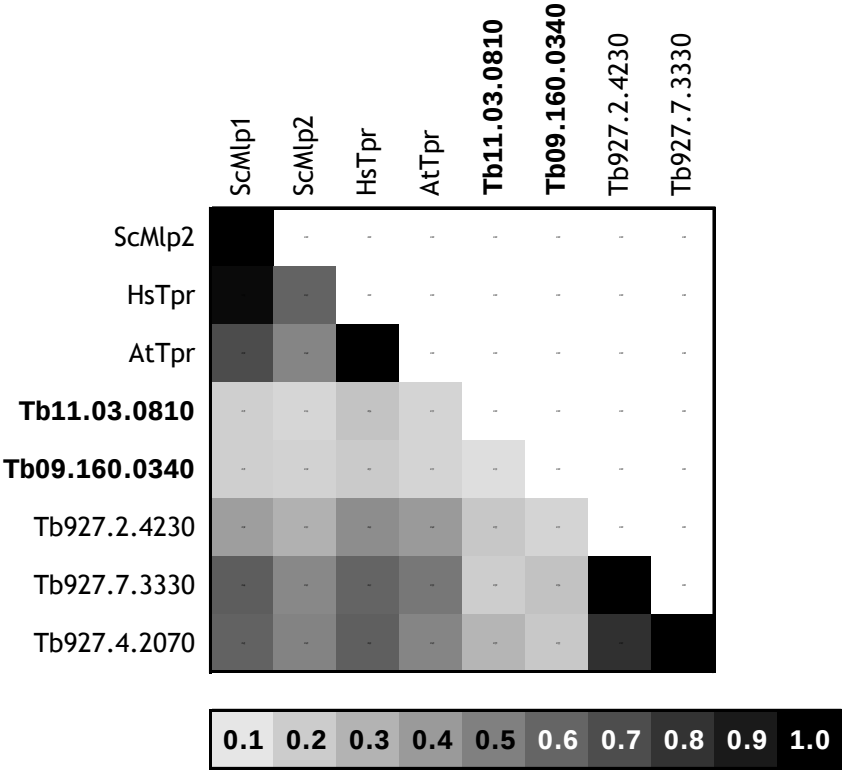


Figure 18: Pair-wise alignment matrix of coiled coil NE and NPC associated proteins. See Figure 15 for the description. TbNUP-1 is Tb927.2.4230.

THE FG-NUPS

We identified ten TbNups which contain FG-repeat domains, four of which we localized to the NE (Chapter 4). Due to the repetitive primary structure and the natively disordered secondary structure, the FG-Nups have evolved faster than the other groups of Nups (Denning and Rexach, 2007). The rapid evolution impedes any attempt to resolve homologous relationships through primary structure alone. Given that the structure of the Kaps have remained relatively unchanged throughout evolution, and that the transport mechanism is also probably conserved, there must to be a unifying theme of FG-Nups across Eukaryota (Feldherr et al., 2002; Stewart, 2007a).

There is evidence that not all FG-Nup domains are equally essential to nucleocytoplasmic transport. For example, asymmetric FG-Nups are dispensable, while only specific combinations of symmetric Nups are essential for nucleocytoplasmic transport (Strawn et al., 2004). To date, which karyopherin specifically binds to which Nup has not been experimentally decoded. However, it is probable that this code may be quite complex. For example, Kap95 (nuclear protein import) and Mex67 (mRNA export) each specifically bind to two different and discreet sub-regions within the FG-repeat domain of ScNup116 (Strawn et al., 2001). Interestingly, the only feature that distinguishes between these two chemically similar sub-domains of ScNup116 is spacer length.

Since pair-wise alignments of the FG-Nups reveal little, we compared the FG-repeat motifs of *T. brucei* to those in yeast and vertebrates. Table 1 lists FG-Nups from *S. cerevisiae*, *H. sapiens*, and *T. brucei* and their respective FG-repeat motifs. Assuming a similar molecular architecture, *T. brucei* NPCs contain more FG-dipeptide repeats and repetitive sequences than either yeast or humans, although the average spacing of the FG-repeats is smaller in trypanosomes. Both yeast and human FG-Nups have three main FG-repeat motifs, two of which they share (i.e. FXFG and GLFG). By contrast, *T. brucei* contains eight main motifs, only two of which are found in yeast and humans (FXFG and GFG) (Table 1).

Since the mechanism of nucleocytoplasmic transport is conserved, but the primary structure of the FG-motifs and FG-repeat domains are not, we analyzed the chemical nature of the spacer regions between the FG-dipeptides. To that end, we calculated the amino acid composition of each FG-repeat domain, where we define the FG-repeat domain as the region from the first Phe to the last Phe in the domain, excluding those regions with a low density of FG-dipeptides. The vertebrate, fungi and trypanosome FG-repeat domains generally have the same percentage composition of Phe residues, which is 2-3X higher than the average occurrence in proteins. In addition, the domains are generally depleted of Arg, Asp, Cys, Glu, His, Ile, Leu, Lys, Met, Trp, Tyr, and Val, and they are enriched in Ala, Asn, Gly, Pro, Ser, and Thr.

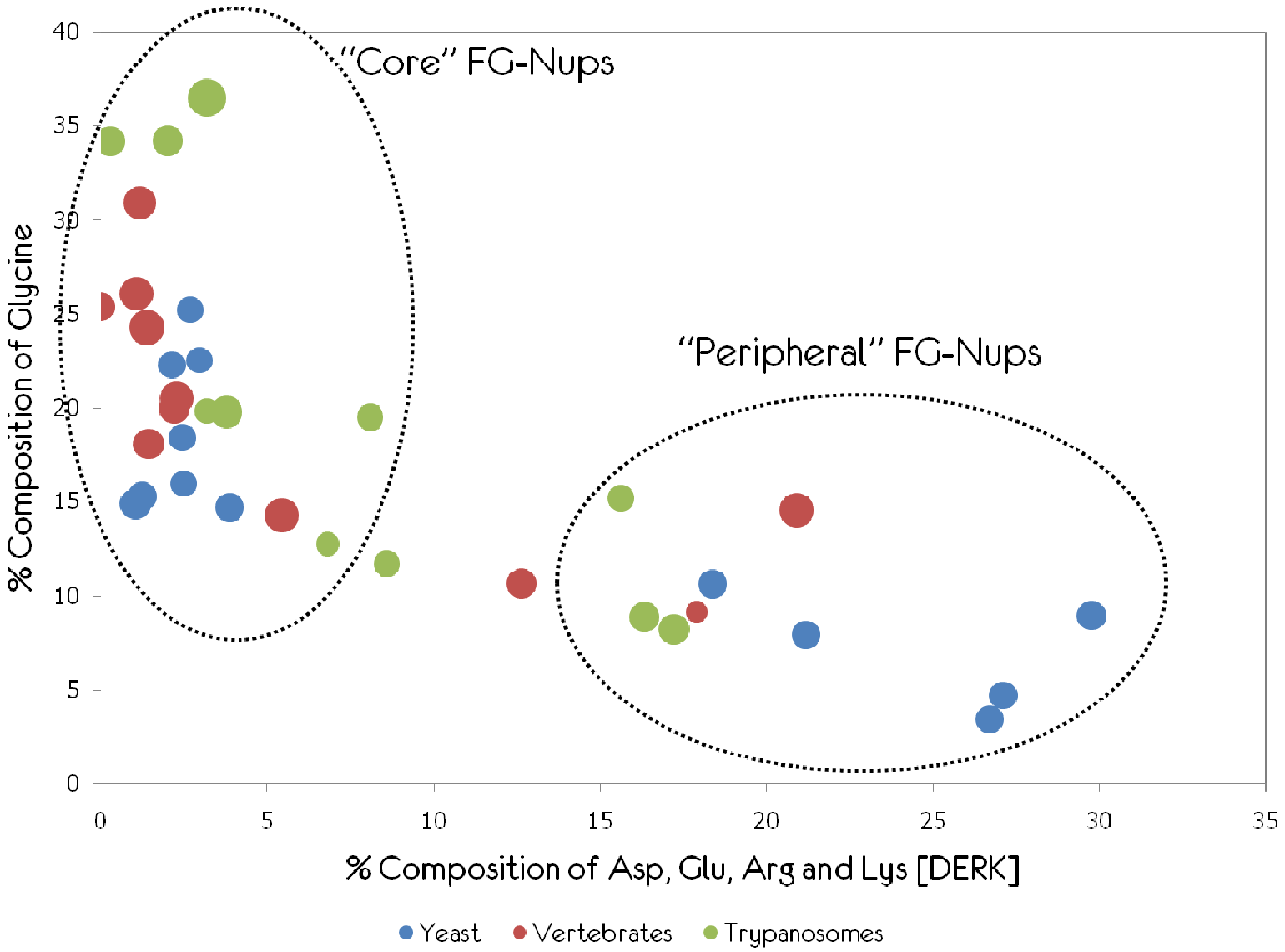
Table 1: Comparing to the FG-dipeptide motifs from *S. cerevisiae*, *H. sapiens*, and *T. brucei*. The motif that is present in the highest number in any given FG-Nup is defined as the plurality motif, the number of which is indicated to the right. An asterisk in the "Rep Seq" column indicates near-perfect sequence repeats in the FG-Repeat domain.

Nucleoporin	# of FGs	Plurality motif	# of motifs	Rep seq
ScNsp1	33	FXFG	19	*
ScNup1	17	FXFG	13	
ScNup2	12	FXFG	11	
ScNup42	29	-	-	
ScNup49	18	GLFG	8	
ScNup53	4	-	-	
ScNup57	16	GLFG	9	
ScNup59	6	-	-	
ScNup60	-	FxF	4	
ScNup100	45	GLFG	11	
ScNup116	47	GLFG	20	
ScNup145	15	GLFG	7	
ScNup159	28	-	-	*
Totals	270		102	2
HsNup50	5	-	-	
HsNup54	9	GFG	5	
HsNup58	14	-	-	
HsNup62	6	FXFG	5	
HsNup98-96	40	GLFG	9	
HsNup153	29	FXFG	15	
HsNup214	44	FXFG	7	
HsNup358	22	FXFG	13	
HsNLP1	12	-	-	
HsPom121	24	-	-	
Totals	205		54	0
TbNup53a	13	GFG	9	
TbNup53b	11	GFG	8	
TbNup59	10	FGFG	4	
TbNup62	30	GGFGA	25	*
TbNup64	13	[S/T]FG	7	
TbNup75	10	FSFG	3	
TbNup98	21	FSFG	19	*
TbNup140	101	[A/V]FGQ	97	*
TbNup149	18	VFGT	18	*
TbNup158	59	GGFGQ	37	*
Totals	286		227	5

The one amino acid whose abundance varies considerably (from 36.5% to 3.5%) between FG-repeat domains is glycine. We thus sorted the domains into descending order based on glycine content. The abundance of most other residues varies in an apparently random fashion. However, one group of depleted amino acids, the acidic and basic residues, Asp, Glu, Arg, and Lys (DERK), generally increases in abundance as the Gly abundance decreases. In other words, glycine enriched domains are DERK depleted and, conversely, DERK enriched domains are glycine depleted (below the average occurrence of glycine in proteins). This anti-correlation is visualized in Figure 19.

This trend correlates with the localization of the FG-Nups within the yeast and vertebrate NPC. In general, an FG-repeat domain that is glycine enriched and DERK deficient will be symmetrically distributed about the core of the NPC (Figure 19). Conversely, a glycine deficient and DERK enriched FG-repeat domain is more likely to be asymmetrically distributed on either the nucleoplasmic or cytoplasmic peripheral face of the NPC. Interestingly, FG-Nups that localize to the nuclear face of the NPC tend to have a higher composition of DERK, compared to cytoplasmically localized FG-Nups. FG-Nups that are distributed about the NPC in a biased fashion (localized to the core, but not symmetrically) tend to have an intermediate amount of glycine with a low DERK composition. While the motifs and primary structure are not well-conserved through evolution, this compositional anti-correlation is well-conserved, and may be a key to the specificity that some FG-Nups have for specific sub-sets of Kaps.

Figure 19: (Following page) The anti-correlation between glycine and the charged residues (Asp, Glu, Arg, and Lys, DERK). The percent composition of the glycine and the charged residues are plotted on the x and y axis, respectively. Each data point represents and FG-Nup from either *S. cerevisiae* (blue), *H. sapiens* (red), or *T. brucei* (green). The area of each data points is directly proportional to the phenylalanine concentration within the respective FG-Nup. The FG-Nups tend to cluster into two groups: the core FG-Nups that are symmetrically distributed within the NPC; and the peripheral FG-Nups, which are asymmetrically distributed. The average natural occurrence (in vertebrates) for Phe is ~4%, for Gly is ~7% and the sum natural occurrence for the charged residues is ~23%.


"Core" FG-Nups
"Peripheral" FG-Nups
% Composition of Glycine
% Composition of Asp, Glu, Arg and Lys [DERK]
0
5
10
15
20
25
30
35
40
Yeast
Vertebrates
Trypanosomes

All of the asymmetric FG-Nups, with low glycine content, are dispensable in yeast (Strawn et al., 2004). However, while not essential, a stretch of charged residues within an asymmetric FG-repeat may increase the strength and/or on–rate of the transient interaction between the FG-repeat region and the Kap. For example, a Kap-Cargo complex that is destined for the nucleus would transit through the NPC in a Brownian fashion (Rout et al., 2000). A vectorial cue, such as highly charged asymmetrical FG-Nups, would impart directionality to increase the efficiency of nucleocytoplasmic transport by guiding Kap-cargo complexes to the exit or by keeping them at the exit until dissociation occurs via ancillary factors. This correlation also provides an additional prediction for the possible distribution of FG-Nups within the TbNPC, which can be tested by immunogold EM.

In a further attempt to resolve homology relationships between FG-Nups, we compared the non-FG-repeat domains of FG-Nups by pair-wise alignments (Figure 20). However, there is little significant similarity between these domains, which is not surprising given that the anchoring domain of the FG-Nups are generally coiled coils (Melcak et al., 2007). The most similar alignments result from the zinc finger domains of HsNup358 and HsNup153.

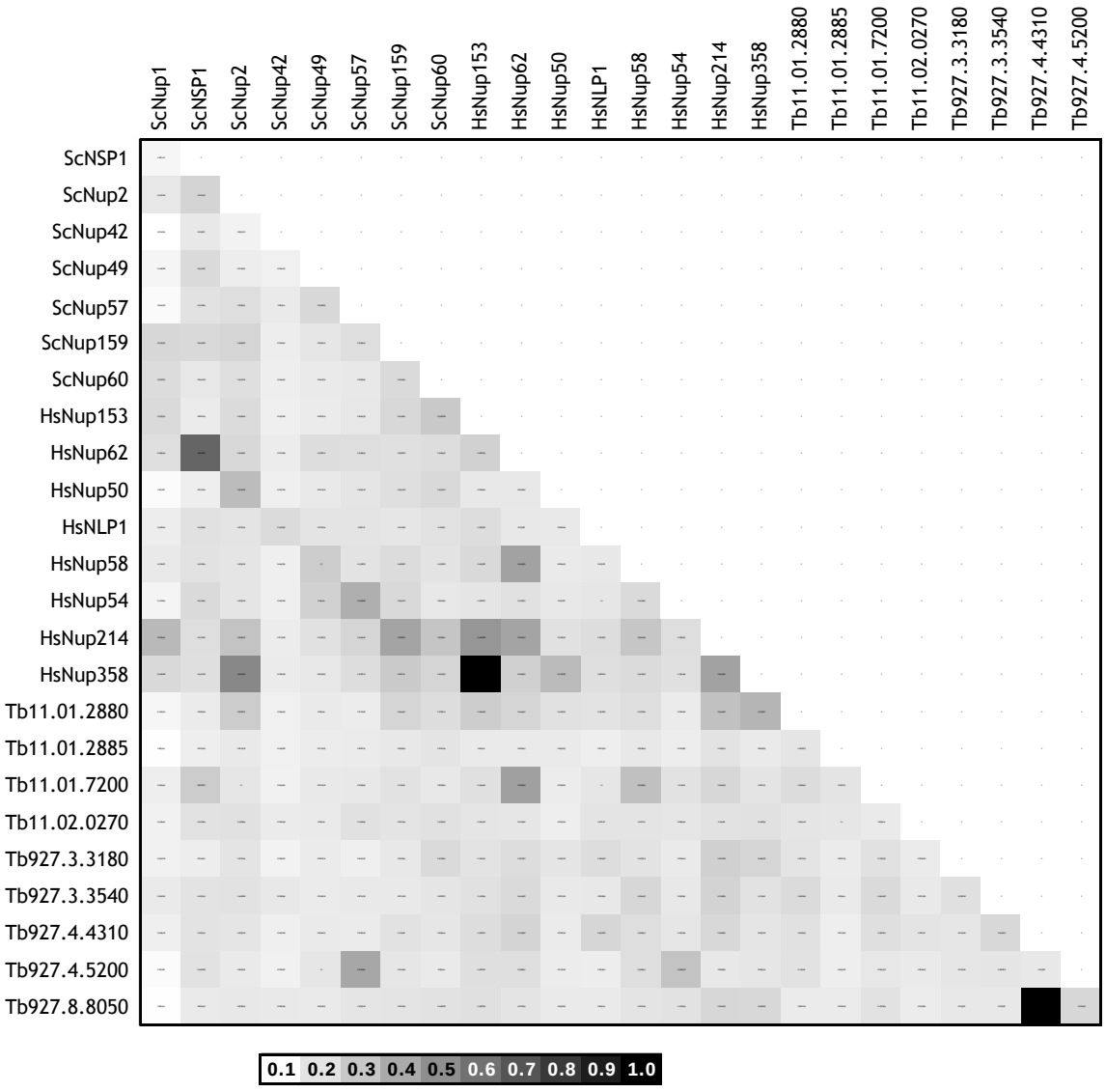


Figure 20: Pair-wise alignment matrix of the non-FG-Repeat domains from the FG-Nups. See Figure 15 for the description.

THE KARYOPHERINS

Kaps comprise the soluble part to the nucleocytoplasmic transport system. With a few exceptions, they are well-conserved across the opisthokonts and excavates (Figure 21). The least conserved transport factor is NTF2, which is responsible for Ran recycling across the NPC. The structures of several Kaps, as well as interactions with stretches of FG-repeat domains and substrates, have been solved (Bayliss et al., 2002; Bayliss et al., 2000b; Bullock et al., 1996; Grant et al., 2002; Grant et al., 2003; Latzer et al., 2007; Lee et al., 2005; Liu and Stewart, 2005; Matsuura et al., 2003; Matsuura and Stewart, 2004a; Matsuura and Stewart, 2004b; Matsuura and Stewart, 2005; Morrison et al., 2003; Saric et al., 2007).

We identified 8 Kaps or transport factors in the TbNEP dataset, including Ran, the energy carrier for nucleocytoplasmic transport. Generally, the trypanosome transport factors are more conserved than the rest of the NPC components. This is most likely due to the large number of substrates with which these molecules must interact and the conservation of the mechanism of nucleocytoplasmic transport.

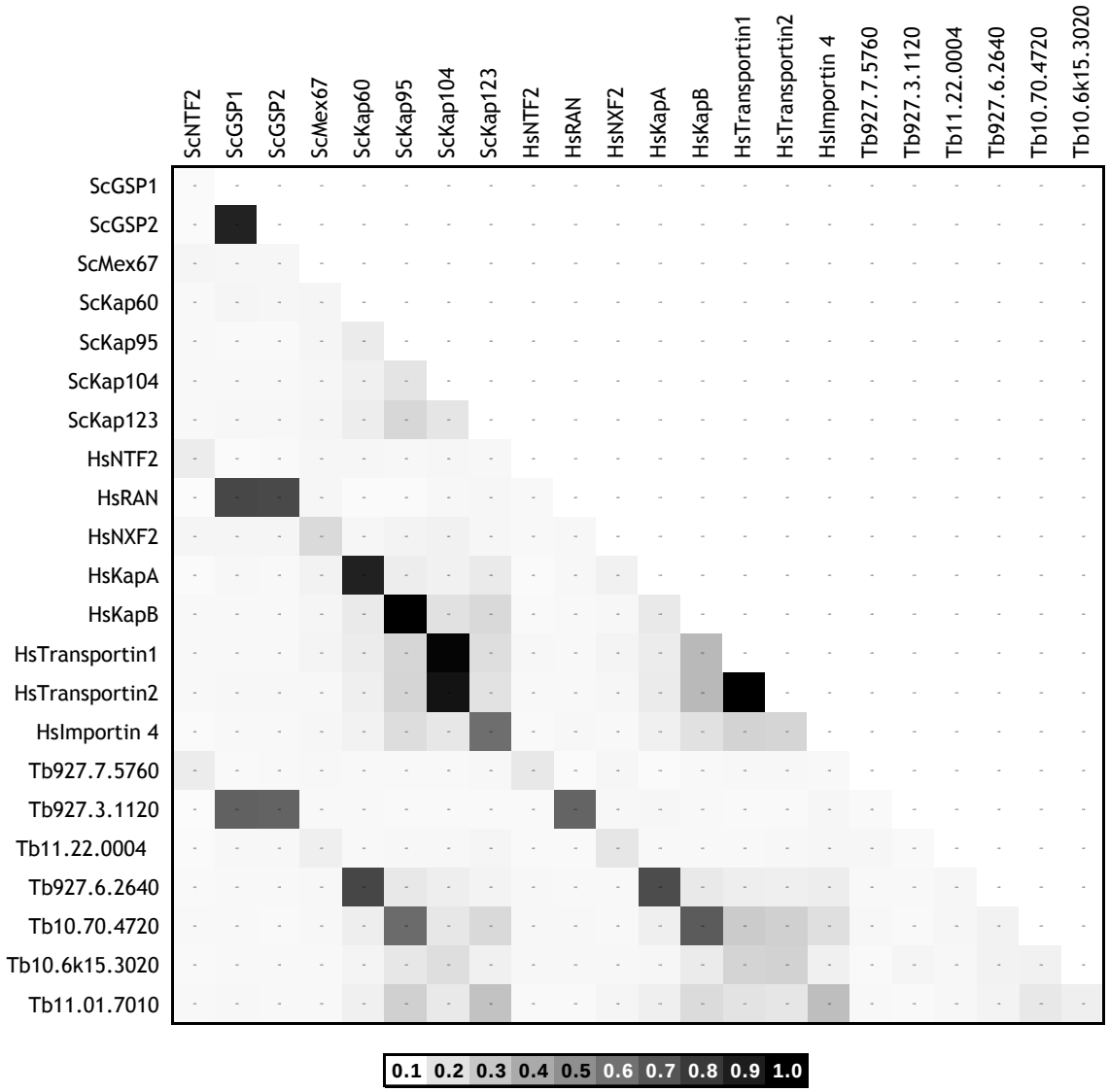


Figure 21: Pair-wise alignment matrix of the karyopherins. See Figure 15 for the description.

THE AUTOPROTEOLYTIC DOMAIN

One of the more peculiar Nups is the autoproteolytic ScNup145, and its ortholog HsNup98-96. Immediately after translation, the β-sandwich auto proteolytic domain folds and initiates cleavage at a conserved H[F/Y][S/T} tripeptide (Fontoura et al., 1999; Rosenblum and Blobel, 1999). The N-terminal protein contains an FG-repeat domain while the C-terminal molecule contains a α-solenoid structural fold and resides with the scaffold of the NPC. Using both mass spectrometry and western blotting, we find that TbNup158 does not cleave and instead functions as the full length protein. To investigate the conservation of the autoproteolytic domain, we performed a multiple alignment between 11 species (Figures 22 and 23). While the β-sandwich domain is highly conserved in *T. brucei*, as well as *G. lamblia*, both sequences lack the tripeptide catalytic site requisite to initiate cleavage. We further investigated this TbNup using fluorescent localization (Chapter 4) and RNAi (Chapter 5).

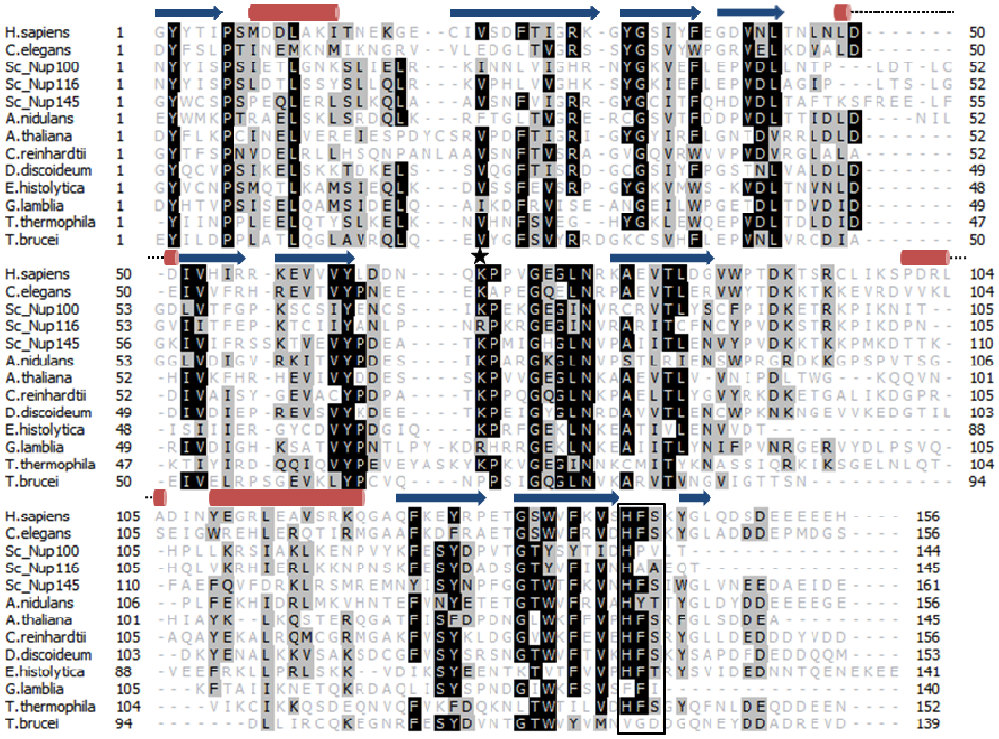


Figure 22: Multiple alignment of the autoproteolytic β-sandwich domain. Identities are indicated by black shading and similarities are shaded in gray. A conserved lysine that stabilizes the transition state of the acyl-shift cleavage is indicated with a star. The box has been drawn around the catalytic site. Above the alignment is the secondary structure information from the solved structure of HsNup98 at atomic resolution (Hodel et al., 2002).

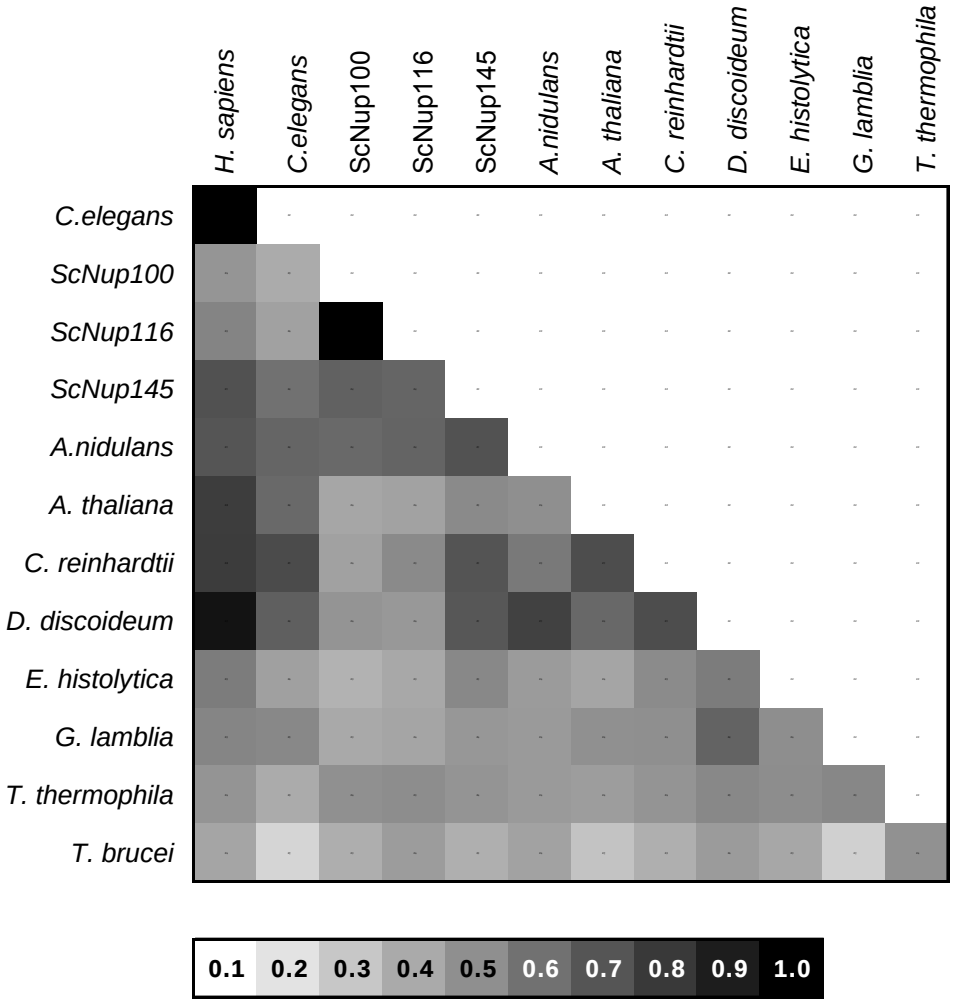


Figure 23: Pair-wise alignment matrix of the β-sandwich domain. See Figure 15 for the description.

THE NPCS OF NEIGHBORING SPECIES

The TbNPC could not have been characterized by bioinformatics alone (Bapteste et al., 2005; Mans et al., 2004). The primary structure of many of the Nups has drifted too far to retain any detectable similarity. One goal of this project was to use the excavate NPC inventory to resolve the NPCs of neighboring bikonts. However, even once the full set of TbNups was cataloged, we could not effectively characterize the entire structural scaffold of the NPC of other neighboring Eukaryotes, such a *T. gondii* and *N. gruberi* (Table 2). It is not surprising that the β-propellers are the easiest to identify by pair-wise alignment given their conservation throughout the eukaryotic tree of life.

We have identified and characterized 24 putative members of TbNPC through biochemistry, mass spectrometry and bioinformatics. We now present our continued characterization of the TbNPC and its constituent members by fluorescence localization and RNAi functional studies.

Table 2: Using the TbNPC inventory as the query sequence, the genomes of six diverse unikonts/bikonts were searched for homologous sequences. An identification is confirmed when the top pair-wise aligned sequence is properly aligned, in a reciprocal fashion, back to the *T. brucei* genome. Nf, not found. Note, a result of "not found" does not imply that that Nup does not exist in the subject species.

Domains	***T. brucei***	***E. histolytica***	***N. gruberi***	***D. discoideum***	***A. thaliana***	***C. reinhardtii***	***T. gondii***
β Propeller, α Solenoid	TbNup109	nf	nf	nf	nf	nf	nf
β Propeller, α Solenoid	TbNup132	nf	nf	nf	nf	nf	nf
β Propeller, α Solenoid	TbNup144	nf	61200	DDB0235243	AT1G14850	191024	nf
β Propeller	TbSec13	264.m00090	35757	DDB0235182	AT2G30050	139713	20.m03668
β Propeller	TbNup43	nf	60369	DDB0184087	nf	191514	52.m01597
β Propeller	TbNup48	nf	61882	DDB0233287	AT3G56900	nf	nf
β Propeller	TbSeh1/Sec13	18.m00329	81581	DDB0233985	AT5G15550	128420	nf
α Solenoid	TbNup82	nf	nf	nf	nf	nf	nf
α Solenoid	TbNup89	nf	nf	nf	nf	nf	nf
α Solenoid	TbNup96	nf	nf	nf	nf	nf	nf
α Solenoid	TbNup181	nf	nf	nf	nf	nf	nf
α Solenoid	TbNup225	nf	nf	nf	nf	nf	nf
FG Repeat, Nup98, α Solenoid	TbNup158	nf	nf	DDB0235244	AT1G59660	164461	55.m04836

CHAPTER 4 — GFP LOCALIZATION OF PUTATIVE NUCLEOPORINS

Using several biochemical and mass spectrometric strategies, we identified nearly 860 proteins that associate with the *T. brucei* nuclear envelope preparation (Chapter 2). Employing a rigorous bioinformatic strategy, we then identified and begun to characterize 24 putative TbNups, as well as 8 Kaps (Chapter 3, Figure 13, and Table 5). To further test our functional predictions, we tagged fully half of the putative TbNups with a fluorescent label. We also tagged 8 unannotated proteins because they contained secondary structure fold patterns (e.g. β-propeller, α-solenoid, transmembrane helices, Zinc-finger, and cadherin folds) that are consistent with those contained in Nups. We then observed the localization of these target proteins by fluorescent microscopy, expecting bona fide TbNups to form a punctuate pattern at the nuclear periphery.

Green fluorescent protein (GFP) is widely recognized for its use as a protein epitope tag (Chalfie et al., 1994). In recent years, GFP has been used as a tag to probe protein localizations and interactions in trypanosomes (He et al., 2004). In this work, we use a polymerase chain reaction (PCR) based approach (Shen et al., 2001) to epitope tag genes of interest (GOI) endogenously. The strategy utilizes the PCR amplification of a reporter cassette using two opposing primers which contain flanking sequences specific to 3' end of the GOI. Following transfection, the PCR product is integrated into the GOI by homologous recombination. This technique is advantageous in that it avoids the need to clone lengthy genes into exogenous expression vectors and the inherent possibility of mislocalization due to overexpression.

METHODS

It has been demonstrated that Nups are generally more tolerant of a COOH-terminal tag (Cronshaw et al., 2002; Rout et al., 2000), and so we targeted the 3' end of putative TbNups. The pMOTag series of plasmids are designed for 3' tagging (Oberholzer et al., 2006). The general scheme of these vectors is shown in Figure 24. The full reporter cassette contains the GFP tag, a splice signaling region (the intergenic region of α and β tubulin), and the antibiotic marker cassette. The forward and reverse PCR primers are designed with approximately 80 nucleotides, which are homologous to the sequences which flank the 3' end of the GOI (without the stop codon). This target specific region is followed by 20 nucleotides specific to the pMOTag vector. See Figure 24.

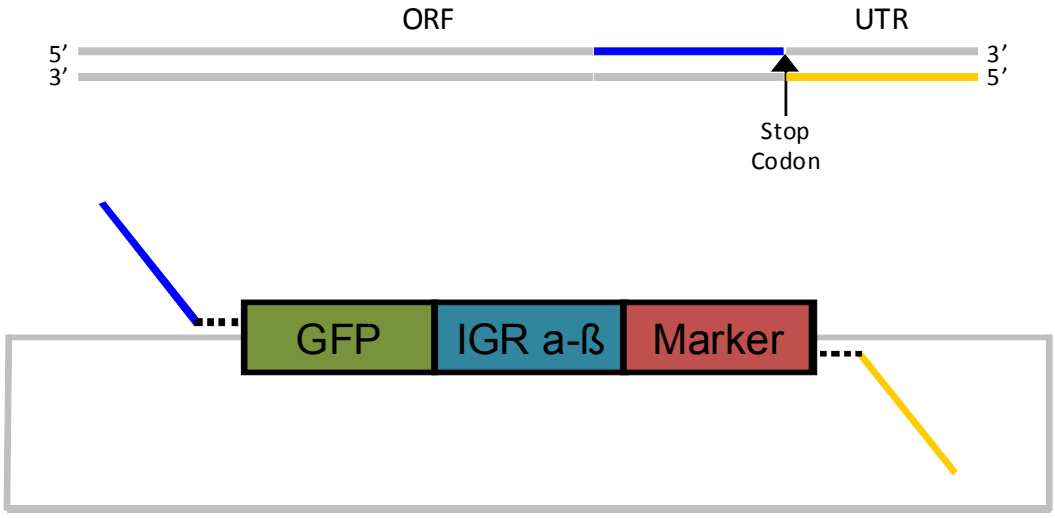


Figure 24: Schema for the pMOTag series of GFP tagging vectors. Upper panel shows the ORF plus additional downstream nucleotides (nt) which comprise the 3’ untranslated region (UTR). Blue and yellow lines indicate the location and sequence of the forward and reverse primers, respectively. These primers are approximately 80 nts in length and an additional 20 nts is added onto the primers to complement the plasmid (Bottom, dashed lines). The bottom panel shows the reporter cassette, which is comprised of the GFP epitope tag (green), the splice signaling region (blue) and the antibiotic resistance marker (red).

The product is amplified using PCR with the following program: 4 min at 94°C, followed by 30 cycles of 1 min at 94°C, 1 min at 63°C, and 2 min at 72°C. The PCR product is purified by standard ethanol precipitation. 10 µg of DNA is transfected into procyclic cells. Positive colonies are selected with hygromycin (25 µg/ml) and are assayed for proper insertion and expression using PCR and/or western blotting.

Following transfection, the stop codon of one allele has been replaced by the reporter cassette with the GFP moiety in frame with the GOI. Following translation the GFP is linked to the target protein by a short peptide linker (the linker sequence is GTGPPLE). A list of primers used in this study is included as Appendix E.

To image the GFP localization of the target protein, 10^7 cells in 1 ml are fixed with formaldehyde (2% final concentration). After five minutes, the cells are spun (3k for 3 min) and washed twice with PBS. The cells are then allowed to settle onto a silanized coverslip. After any nonattached cells are washed away, the coverslip is then mounted in 50% glycerol, 0.4 µg/ml DAPI (4',6-Diamino-2-phenylindole dihydrochloride) in PBS. Fluorescence images were acquired with the DeltaVision Image Restoration microscope (Applied Precision/Olympus).

RESULTS AND DISCUSSION

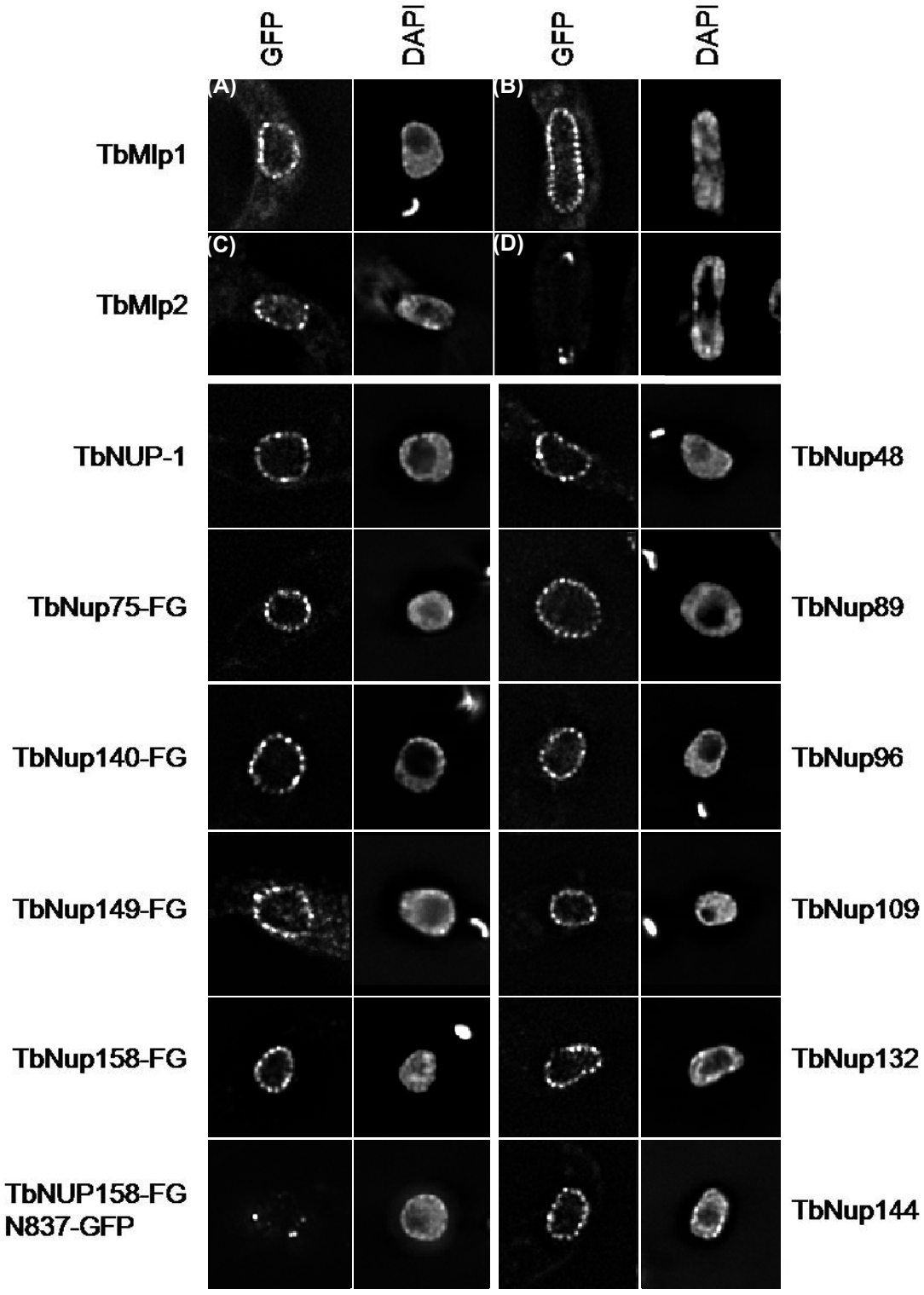
TBNUP LOCALIZATION

A summary of the results is shown in Table 3. Of the 23 targets analyzed, 12 proteins properly localized to the NE, as well as the previously described TbNUP-1 (Figure 25). The *T.brucei* nuclei are ellipsoid in shape and are about 2.5 μm in diameter along the principle axis. The fluorescence pattern of the labeled Nup is unusually punctate, compared to yeast and vertebrates. One putative Nup and two putative lamins (Tb927.4.2070, Tb10.61.0160, and Tb927.7.3330) were unsuccessfully labeled with GFP, as determined by Western blot. Proteins that do not localize to the NPC are characterized by cytoplasmic and/or nuclear staining with no punctate nuclear rim staining. However, it must be noted that lack of nuclear rim staining is inconclusive at best, and does not necessarily eliminate a candidate from consideration as a TbNup.

Table 3: Summary table of results. ND, not determined due to improper GFP insertion or expression.

GeneDB Accession	Mass (kDa)	Annotation	NE localization?
Tb11.03.0810	109.7	TbMlp1	**Yes**
Tb09.160.0340	92.3	TbMlp2	**Yes**
Tb927.2.4230	406.8	TbNUP-1	**Yes**
Tb11.02.2120	48.4	TbNup48	**Yes**
Tb11.02.0460	89.1	TbNup89	**Yes**
Tb10.6k15.3670	96.5	TbNup96	**Yes**
Tb11.01.7630	108.7	TbNup109	**Yes**
Tb927.7.2300	132.3	TbNup132	**Yes**
Tb10.6k15.2350	144.3	TbNup144	**Yes**
Tb927.8.8050	74.8	TbNup75-FG	**Yes**
Tb11.01.2885	140.3	TbNup140-FG	**Yes**
Tb11.01.2880	149.2	TbNup149-FG	**Yes**
Tb11.03.0140	158.3	TbNup158-FG	**Yes**
Tb11.03.0140N	158.3	TbNup158-FG N837-GFP	Foci at the NE
Tb927.4.1310	47.4	Unannotated	Inconclusive
Tb927.4.590	88.0	Unannotated	Inconclusive
Tb10.70.1120	109.2	Unannotated	Inconclusive
Tb927.6.1830	49.9	Unannotated	Inconclusive
Tb10.70.1110	59.9	Unannotated	Inconclusive
Tb10.70.1130	48.2	Unannotated	Inconclusive
Tb927.4.2070	511.3	Unannotated	Not Determined
Tb10.61.0160	42.9	TbNup43	Not Determined
Tb927.7.3330	502.6	Unannotated	Not Determined

Figure 25 (following page): Fluorescent microscopy images of COOH-terminal GFP labeled TbNups and corresponding DAPI fluorescence marker. Two coiled coil TbNups, TbMlp1 and TbMlp2, are visualized in under non-mitotic (panels A and C) and mitotic (panels B and D) conditions to show the differential localization of TbMlp2.


GFP
DAPI
GFP
DAPI
(A)
(B)
TbMlp1
(C)
(D)
TbMlp2
TbNUP-1
TbNup48
TbNup75-FG
TbNup89
TbNup140-FG
TbNup96
TbNup149-FG
TbNup109
TbNup158-FG
TbNup132
TbNUP158-FG
N837-GFP
TbNup144

Within the TbNEP, two proteins stood out, by their coiled coil structure, as putative Mlp/Tpr proteins. We labeled these two unannotated proteins with GFP, and found that they localize at the nuclear periphery (Figure 25 (A & C)). However, during mitosis, only TbMlp2 localizes to an area proximal to the spindle pole bodies (Figure 25 (D)). This is analogous to the likely yeast nuclear basket Nups, Mlp1/2, which are paralogous, but perform different functions at the NE. This leads us to speculate about the possible roles of the TbMlps at the nuclear periphery, such as maintenance of telomere length and telomere localization at the NE. Given the poor homology between the yeast and trypanosome Mlps, it would be interesting to identify additional interacting partners. Further colocalization studies with telomere maintenance complexes are planned.

We also carried out further investigation of TbNup158 — the non-cleaving ortholog of the autoproteolytic Nup ScNup145. In this investigation, we produce the truncated protein TbNup158(1-837)-GFP, which contains the N-terminal FG-repeat domain and the conserved β-sandwich autocatalytic domain with a GFP tag at its C-terminus. This truncated protein does not contain the C-terminal α-solenoid domain. TbNup(1-837)-GFP localized to several punctuate foci at the NE (Figure 25). These results are in contrast to that found in vertebrates. Here, a HsNup98-96(F863S/Y866R) mutant, which cannot cleave, mislocalizes to the nuclear interior, however the HsNup98(1-863) mutant properly localizes to the nuclear periphery (Fontoura et al., 1999). Evidently, the functional difference between TbNup158 and the higher eukaryotes is more complex than simply the lack of a cleavage triad. The full length trypanosome homolog of HsNup98-96,

with an identical domain structure, properly functions at the NPC. However, the abolishment of the cleavage site within the HsNup98-96 causes mislocalization and loss of function. Furthermore, the introduction of an artificial cleavage site into TbNup158, causes aggregation at distinct foci at the NE. In contrast, the introduction of an additional pool of HsNup98 into HeLa cells does not cause any disruption to NE localization (Fontoura et al., 1999).

NPC DENSITY DETERMINATION

S. cerevisiae NPCs are distributed in the NE in a non-uniform, non-random fashion with an average density of ~12.0 NPCs/μm^2 (Winey et al., 1997). Depending on the cell cycle stage, HeLa cells have an average NPC density of about 5 – 9 NPCs/μm^2 (Maeshima et al., 2006). The distribution of NPCs around a vertebrate nucleus is non-uniform, non-random, and directly dependent on the INM distribution of Lamin A/C and emerin, the presence of which leads to NPC-free islands, especially in early cell cycle stages (Maeshima et al., 2006).

In contrast to the majority of the opisthokonts, the distribution of *T. brucei* appears to be approximately uniform (Figure 26). Both the approximate uniform distribution of NPCs in *T. brucei* and their lower density, compared with yeast, allow us to readily count them using fluorescence microscopy by integrating a series of Z-axis focal plane images. The average density is 5.1 NPCs/μm^2 (N=10, σ=0.8), with an average of 93 NPCs (σ=16) per nucleus. The low density of NPCs around the nucleus explains the resolution of the punctuate pattern. A typical Z-stack of images, showing the resolution of individual NPCs, is shown in Figure 26. In this dividing nucleus, the ~150 NPCs are distributed on average ~0.5 µm apart from each other.

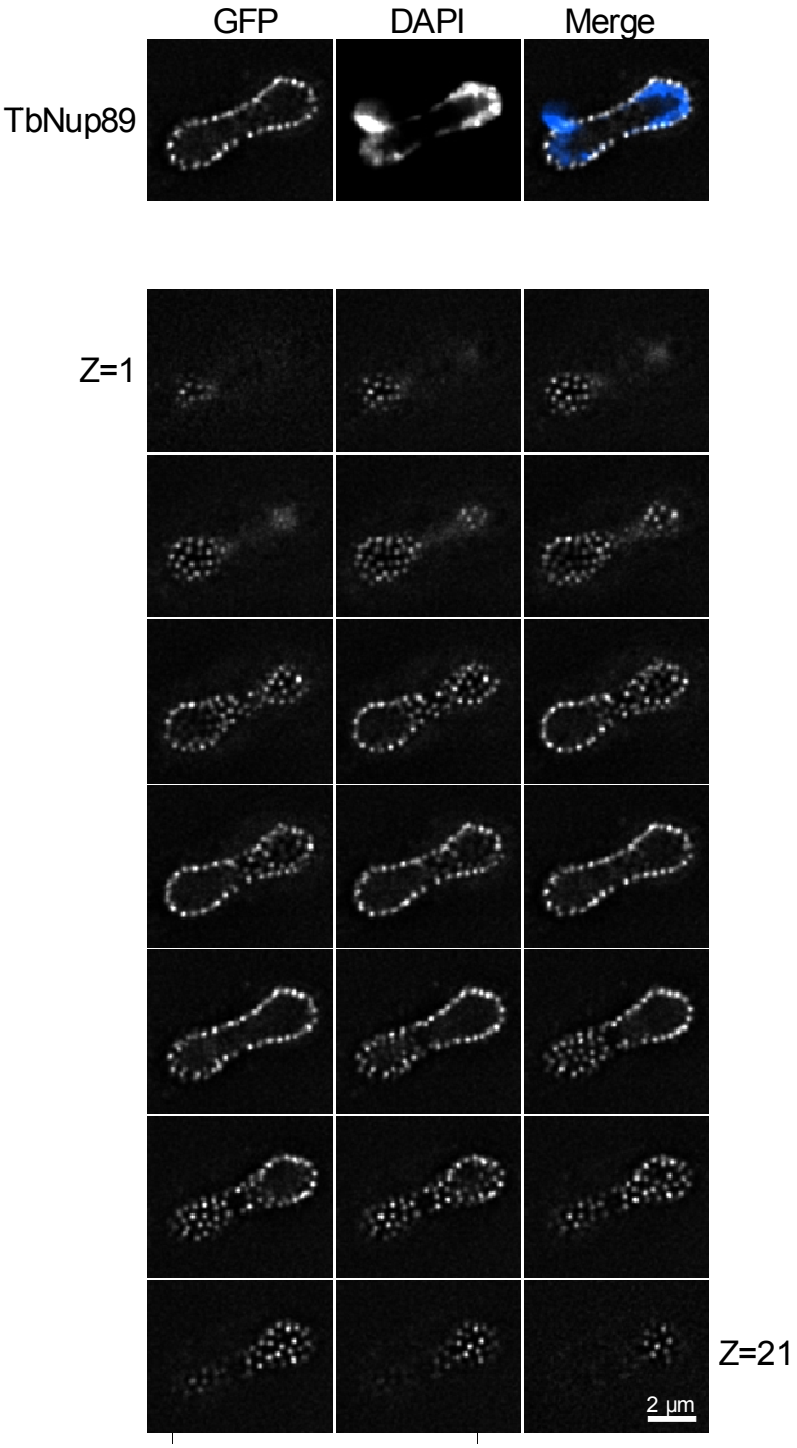


Figure 26: (Top) The center Z slice of TbNup89-GFP showing the location of DNA by DAPI staining. (Bottom) 21 Z-Stacks shown in sequential order. Each stack is 150nm is thick. There are ~150 NPS around this dividing nucleus. Note the resolution of individual NPCs.

Finally, we screened the TbNEP against a panel of available NPC antibodies with an affinity for a specific epitope within the NPC for any cross reactivity. Most of the antibodies that we screened did not cross react with the TbNPC components. However, we found that a polyclonal antibody raised against HsNup107 does cross-react with the *T. brucei* NPC (Glavy et al., 2007). Figure 27 shows the localization overlay of αHsNup107 along with both αTbNUP-1 and TbNup89-GFP (Ogbadoyi et al., 2000). αHSNup107 appears to localize in the regions devoid of αTbNUP-1 signal, which is consistent with their respective roles at the NE, since TbNUP-1 is not a nucleoporin. αHsNup107 does indeed colocalize with GFP labeled TbNup89, suggesting the αHsNup107 is cross-reacting with an epitope within the TbNPC. Although there are several candidates with similar sequences, the trypanosome antigen to HsNup107 is not obvious. αHsNup107 binds to a specific peptide (MVTNLDDSNWAAAFSSQRSG), which is not homologous with any known sequence within the TbNEP or within the *T. brucei* genome. The possibility exists that the antibody recognizes another linear epitope or a conformational epitope within the antigen.

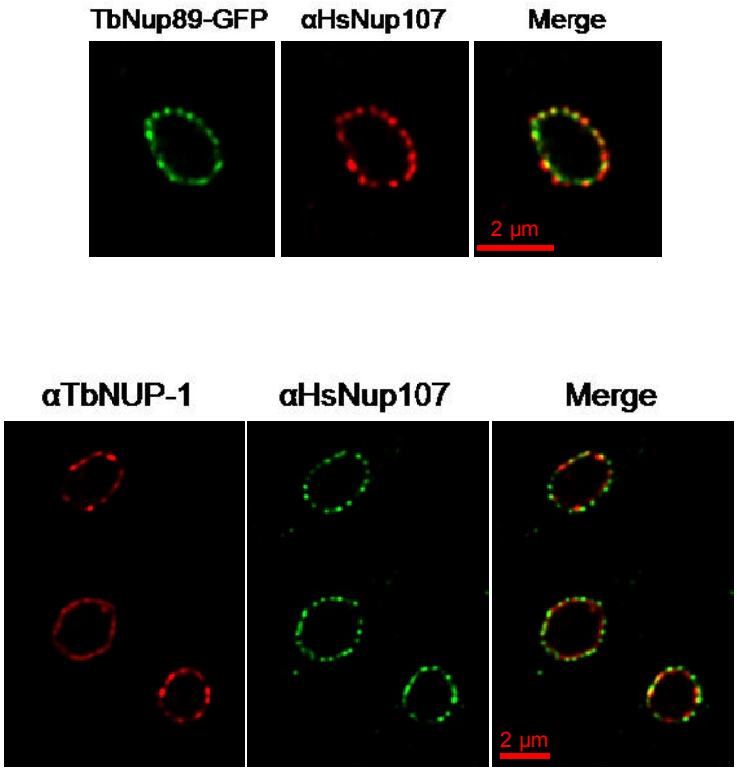


Figure 27: (Top) TbNup89-GFP and αHsNup107 colocalize at the nuclear periphery. (Bottom) αTbNUP-1 and αHsNup107 localize to the nuclear periphery, but do not colocalize.

CHAPTER 5 — KNOCKDOWN STUDIES OF PUTATIVE NUCLEOPORINS USING RNA INTERFERENCE

Using several biochemical and mass spectrometric strategies, we identified nearly 860 proteins that associate with the *T. brucei* nuclear envelope preparation (Chapter 2). Employing a rigorous bioinformatic strategy, we then identified and begun to characterize 24 putative TbNups, as well as 8 Kaps (Chapter 3, Figure 13, and Table 5). We then localized 12 TbNups to the nuclear envelope using GFP fluorescence localization. To further characterize several of the putative TbNups as well as possible additional roles of the TbNPC, we turned next to RNA interference (RNAi) for functional analysis.

RNAi employs double stranded RNA (dsRNA) to suppress the expression of a particular gene, allowing for functional analysis (Zamore, 2002). Once within the cell, dsRNAs are cleaved by Dicer, an endonuclease, to produce small interfering RNAs (siRNAs) about 21-23nt in length. siRNAs, in concert with the multiprotein complex RISC, direct the recognition and subsequent degradation of mRNA (Chi et al., 2003; Hannon, 2002; Zamore, 2002). First described in the nematode *C. elegans* (Fire et al., 1998), RNAi quickly became a powerful tool for large scale functional genomics within many organisms, such as *D. melanogaster* and *C. elegans* (Boutros et al., 2004; Chi et al., 2003; Kamath et al., 2003).

In *T. brucei*, it also been shown that dsRNA effectively silences gene expression by specifically degrading mRNA (Ngo et al., 1998). Moreover, placing RNAi constructs under the control of the TET repressor system allows for

a controllable, persistent, and heritable RNAi response (Bastin et al., 2000; LaCount et al., 2000; Shi et al., 2000; Subramaniam et al., 2006; Wang et al., 2000). RNAi has become a tool of choice over the past 5 years for probing trypanosome gene function because of its robustness and efficiency compared to traditional gene knockout studies (Beverley, 2003). Here, we characterized several putative TbNups using RNAi to determine to what extent they are essential to the viability of the cell.

METHODS

To choose appropriate candidates for RNAi studies, we surveyed the literature for the phenotypes of Nup deletions in *S. cerevisiae* (for details, see Appendix C). This survey revealed that less than a third of the Nups are essential in yeast. Based on these results, we targeted, in *T. brucei*, three well-conserved orthologs of the essential Nups: ScNup145, ScNup96, and ScSec13. Two non-essential yeast Nups that exhibit a clustering phenotype when knocked-out in yeast were also targeted: ScGle2, a transiently interacting Nup, and ScNup84, both of which are well-conserved in trypanosomes.

Detailed protocols for our RNAi studies may be found at the TrypanoFAN resource webpage (http://trypanofan.path.cam.ac.uk). The target regions were determined using the program RNAit (Redmond et al., 2003). The program optimizes both the target region and primers to eliminate any potential problems with amplification and cross reactions. The primers and target region are shown in Figure 28.

Open Reading Frame	Target	Forward Primer	Reverse Primer	Product Length	Target Region
Tb09.160.2360	TbGle2	GATACTGGGACATGAAGCAGC	GCCCAATCATAGCTACGTGC	544	422-966
Tb10.6k15.3670	TbNic96	TTCAGAGGATTGCACACGCTC	TTGGGAAGCATTCATCCGACG	511	1314-1824
Tb10.61.2630	TbSec13	TTGACATTCACCAGGGTAGCG	TGAGGTGACTCCACTGTTTGC	450	413-862
Tb11.02.0460	TbNup89	TCGCATGACTCTTGGCAAGAC	ACCTACACGCAACAAGGAACG	460	1015-1474
Tb11.03.0140	TbNup158	AATGGCAAGTCTGAGTCGGAG	TCCGCAGTTGAGTGAGGTAAC	509	3048-3556
Tb10.70.0830	Clathrin Heavy Chain	ATGCCTGTATTGAGGCCAAC	CAGGTTTTGAGGGCACGTAT	484	3821-4305

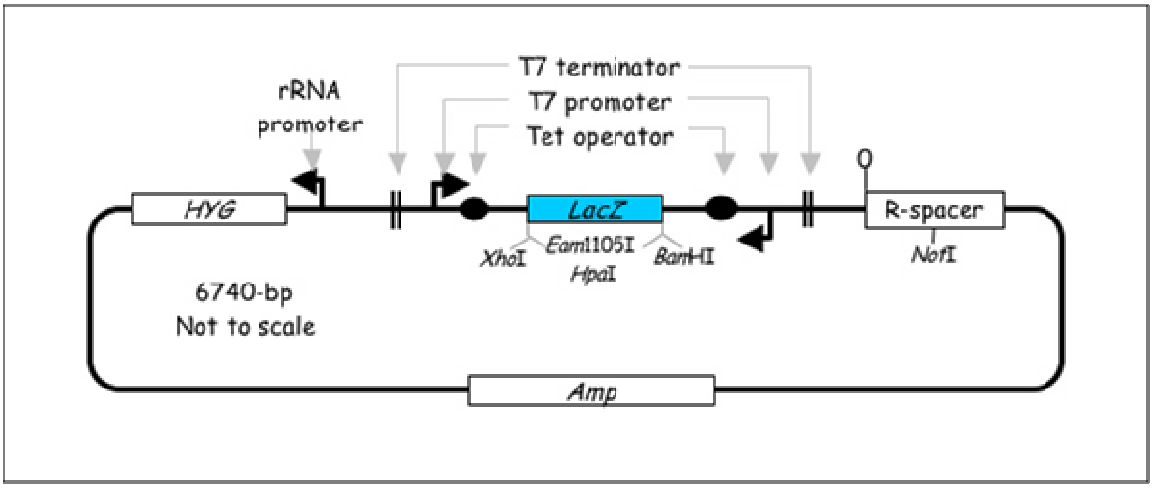


Figure 28: (Top) The PCR primers and targeted regions of the five trypanosome proteins selected for RNAi knockdown studies. Clathrin serves as a positive control. (Bottom) Schematic of the p2T7^{TABlue} plasmid. HYG, hygromycin selection cassette; R-spacer, rDNA integration region; Amp, ampicillin resistance marker.

The target regions were PCR amplified from genomic DNA. The primers and genomic template were mixed with PCR Supermix, (Invitrogen, Carlsbad, CA) and, after 5 min denaturing at 95°C, cycled 35 times with the following sequence: 30 s at 95°C, 30 s at 65°C, and 85 s at 72°C. The product was purified and inserted into p2T7^{TABlue} (Alibu et al., 2005), which was digested with Eam1105I (Fermentas, Hanover, MD) to liberate the LacZ stuffer cassette (Figure 28). The PCR product vector has adenine overhangs, which complement the thymine overhangs in the digested vector. One part digested vector and five parts PCR product were mixed and ligated with T4 ligase (Rapid DNA Ligation Kit, Roche, Indianapolis, IN).

After cloning within *E. coli*, the vector was isolated and purified using standard protocols. Prior to transfection, the plasmid was linearized with NotI (Invitrogen, Carlsbad, CA). Then 10 µg of linearized DNA was tranfected into the Single Marker Bloodstream (SMB) bloodstream form *T. brucei* cell line (T7RNAP::TETR::NEO) using standard established protocols. The SMB cell line has a T7 RNA polymerase and a TET repressor, which expresses the target cassette in the presence of doxycycline. Positive colonies were selected with hygromycin (2.5 µg/ml). After selection and cell line maturation, mRNA knockdown was induced with the addition of 1 µg/ml doxycycline. Cells were observed and counted at 6 or 12 hour time points the first day and every 24 hours post-induction. Samples of each induced cell line were prepared for thin section transmission electron microscopy analysis after 24 hours, post-induction (FEI Tecnai G2 Spirit BioTwin with Gatan 4K x 4K digital camera).

To monitor the efficiency of the RNAi knockdowns, 15 x 10^6 cells were removed at each time point. RNA was extracted using RNeasy Mini Kit (QIAGEN, Valencia, CA) and residual genomic DNA was removed by treatment with DNase (QIAGEN). After quantification, the mRNA was reverse transcribed into cDNA using polyT primers and the TaqMan® Reverse Transcription Reagents (Applied Biosystems, Foster City, CA). Quantitative PCR (qPCR) was performed with the SYBR green fluorescent marker with primers designed with the Primer Express software (Applied Biosystems). The target mRNA levels were reported relative to the endogenous control TbGAPDH (GeneDB accession Tb927.6.4300).

RESULTS AND DISCUSSION

In total, we targeted five TbNups for RNAi functional analysis: TbSec13, TbNup96, TbGle2, TbNup89, TbNup158 (see above). We monitored the efficiency of the induction of dsRNA and subsequent knockdown of the target RNA by qPCR. As shown in Figure 29, 80-90% of the target mRNA has been depleted within 24 hours. The primary assay for the RNAi experiment was cell growth. Thus we counted the cells every 6 or 12 hours for the first day, and every 24 hours thereafter, and diluted, if necessary. For each target gene, three independent cell lines were split – one culture was induced and the other was uninduced and monitored as a control. The results are shown in Figures 30 and 31.

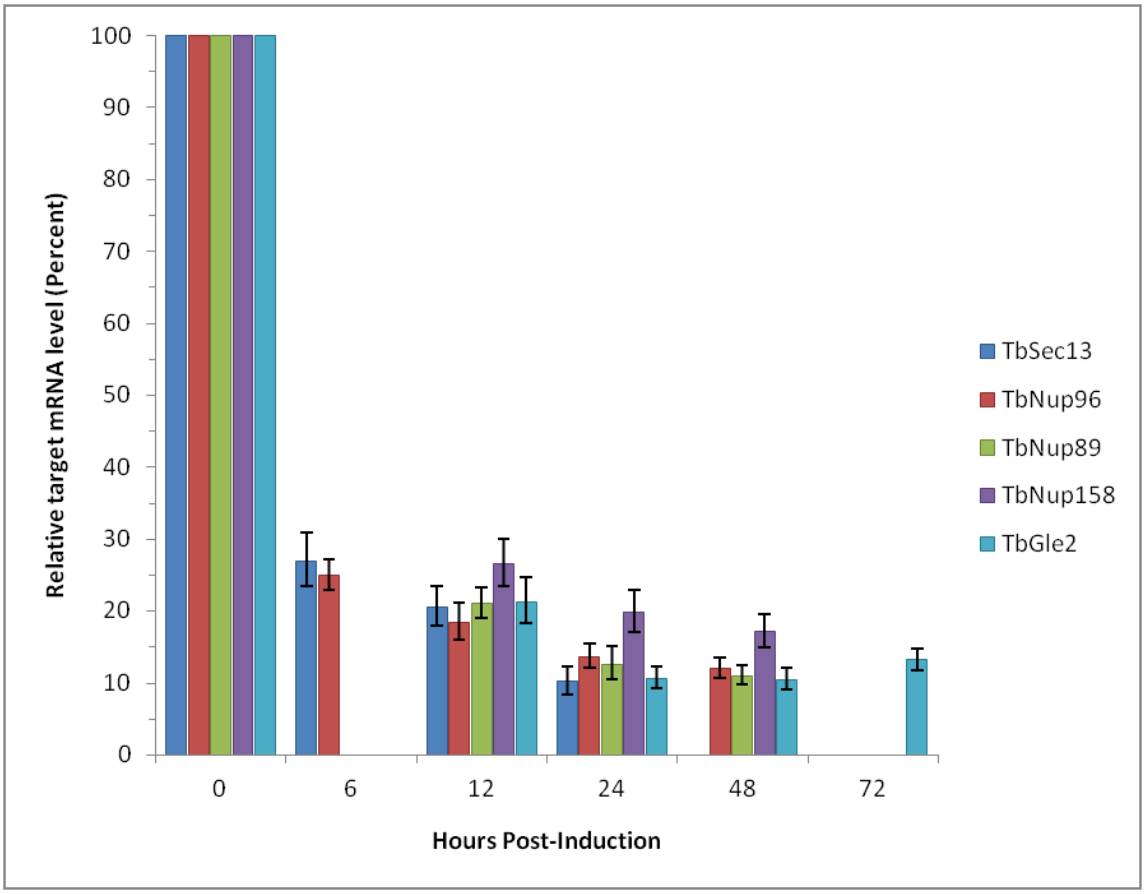


Figure 29: Results from the qPCR monitoring mRNA levels at various time points post-induction of RNAi. Percent mRNA is relative to an endogenous control (TbGAPDH) and relative to pre-induction levels, which is set to 100%. The standard error of the mean was determined from 3 RNA samples from 3 independent cell lines for each target.

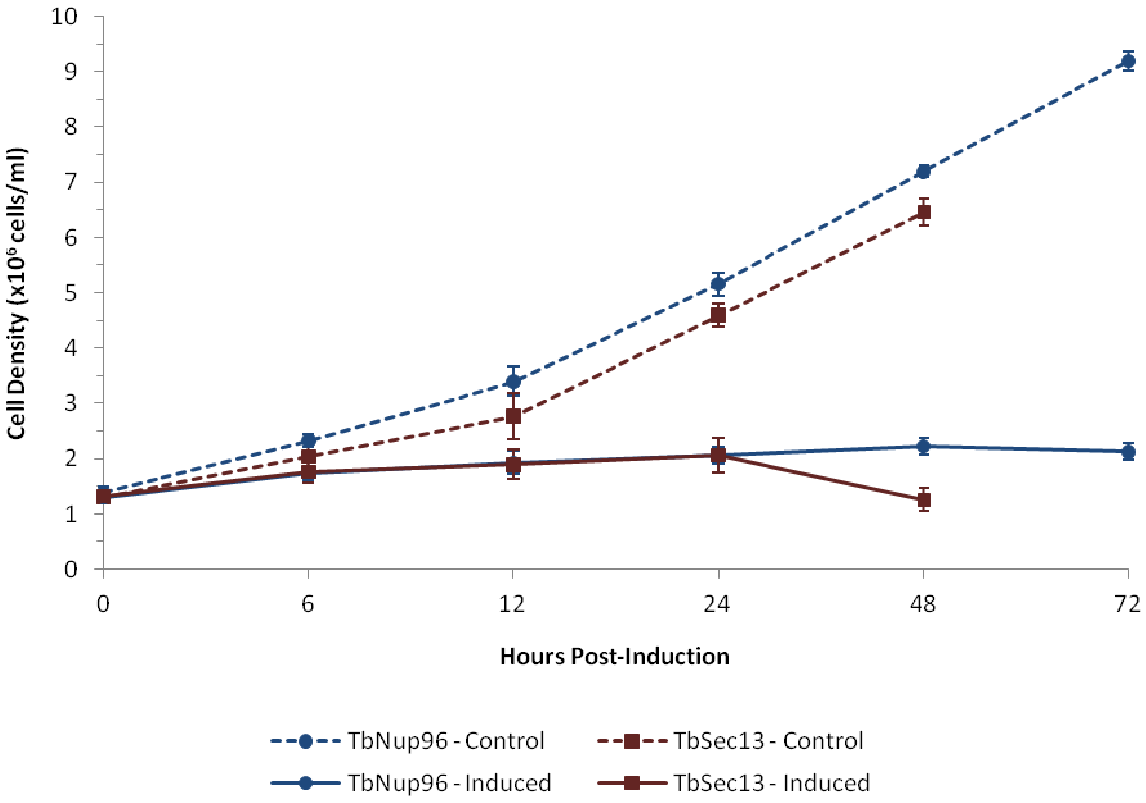


Figure 30: Growth curves of TbNup96-RNAi and TbSec13-RNAi cells. Since these knock downs were lethal within 48 hours, cell density was first measured at 6 hours post-induction. Relative to the uninduced control, cell division ceased within 6 hours post-induction. The standard error of the mean was calculated from 3 independent cell lines of each target gene.

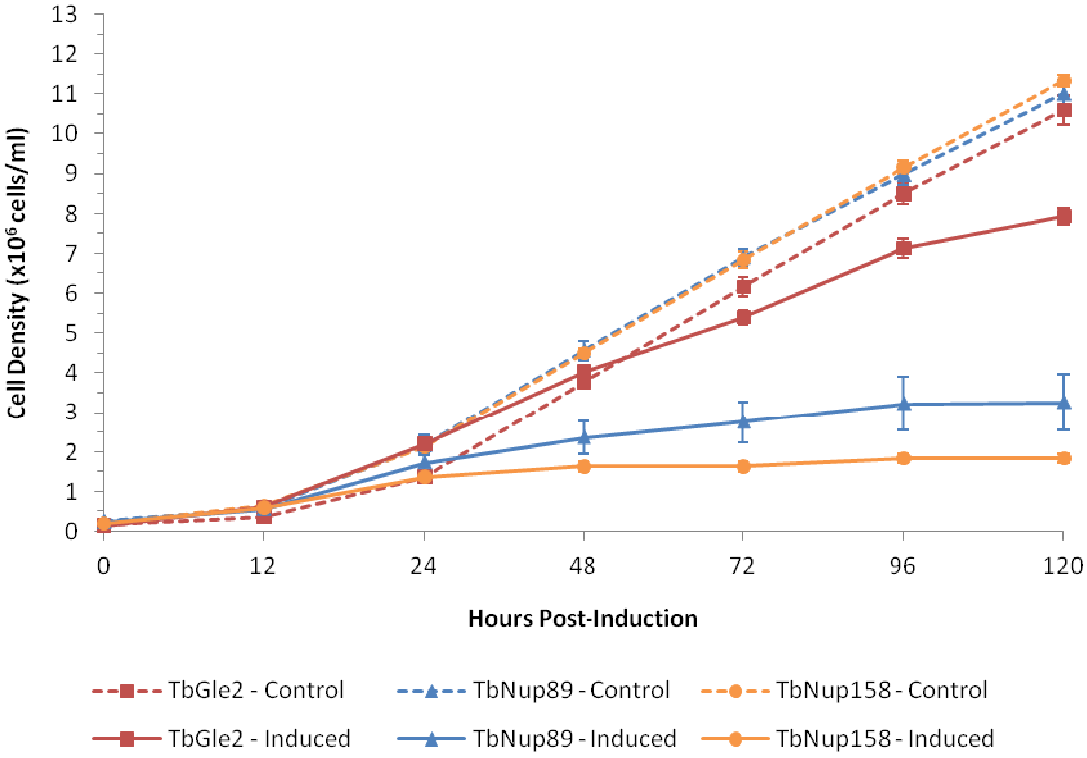


Figure 31: Growth curves of TbGle2-RNAi, TbNup89-RNAi, and TbNup158-RNAi cells. Growth at 12 hours post-induction was similar to the uninduced control. After 24 hours post-induction, TbNup89-RNAi and TbNup158-RNAi cells began to show signs of growth defects and cellular division is slowed. TbNup158-RNAi cells ceased cell division within 48 hours post-induction. TbGle2-RNAi cells and TBNup89-RNAi cells exhibited slow cell growth and eventually succumbed to pleiotropic phenotypes. The standard error of the mean was calculated from 3 independent cell lines of each target.

Knockdown of two *T. brucei* Nups, TbNup96 and TbSec13, caused immediate cell division arrest (Figure 30). After induction of TbSec13 and TbNup96 RNAi, most cells were dead by 24 and 48 hours, respectively. In the case of TbNup96, this is suggestive of cell cycle arrest due to the inability to form new NPCs during S phase. Knockdown of TbNup158 induces a growth defect for the first 24 hours, after which cell division completely arrests and no further culture growth is detected (Figure 31). Many cells were still viable, but exhibited pleiotropic effects (poly-nucleation and cell enlargement). TbNup89 and TbGle2 knockdowns induce a moderate and slight growth defect, respectively, until the 4^{th} day, when pleiotropic effects dominate and cell cycle arrests (Figure 31).

Cessation of cell growth is not an unexpected phenotype, as several fungi and vertebrate Nups have secondary functions as mitotic regulators — for example, HsNup153 (Mendjan et al., 2006). Also, several metazoan Nups, including HsNup214, cause cell cycle arrest upon knockdown (Boer et al., 1998). Although it is unclear how a knockdown of a specific TbNup will affect the cell cycle, perhaps the depletion of NPCs around the nucleus prevents the release of a check point. Since *T. brucei* undergoes a closed mitosis, the nucleus must expand during S-phase to accommodate the replicated genome and prepare for nuclear division. This would also necessitate the *de novo* formation of additional NPCs (D'Angelo et al., 2006). In this case, if the pool of TbNup96 is depleted, then perhaps nascent NPCs cannot assemble, suggesting a key function of TbNup96 in the NPC architecture. The same analysis may apply to TbSec13, but, presumably, this protein also functions as a vesicle coat protein in

trypanosomes (Siniossoglou et al., 1996). However, since Sec13 is localized to a key architectural complex with the NPC, it is possible that its knockdown directly affects NPC formation and cell cycle arrest.

Evidently, the absence of TbNup158 is tolerated for several divisions. One explanation is that the NPCs are able to form without a pool of newly synthesized TbNup158 and the cells are able to propagate with only the existing pool of TbNup158 bound to NPCs. If NPCs are formed *de novo*, as predicted, then with each division, a reduced number of NPCs will contain TbNup158. At 24 hours, or 3 nuclear divisions, perhaps the number of fully functional NPCs drops below a critical threshold and the cell cycle is arrested.

Interestingly, depletion of TbNup89 does not aggressively trigger any cell cycle checkpoints, even though the NPCs are perturbed sufficiently to cause a growth defect. The lack of an obvious phenotype (beyond slow cell growth) upon knockdown of TbGle2 may indicate that the mRNA export pathway is sufficiently redundant to tolerate the loss of this mRNA export factor.

To investigate morphological changes within the NE or nucleus, we observed the nucleus of each knockdown by transmission electron microscopy 24 hours post-induction (Figure 32). Knockdowns of TbGle2 (Figure 32, Panel B) and TbNup96 (Figure 32, Panel C) did not significantly alter the morphology of the nucleus, although TbNup96-RNAi cells were very likely to be NPC-depleted and poly-nucleated. TbSec13-RNAi had a significant effect on the morphology of the nucleus and the cell in general — there are a large number of small vesicles

and disruption to the internal membrane structures (Figure 32, Panel D). Evidently, knockdown of TbSec13 perturbs the regulation of membrane curvature within the cell. This is consistent with the known function of Sec13 as a member of the COPII vesicle coat complex.

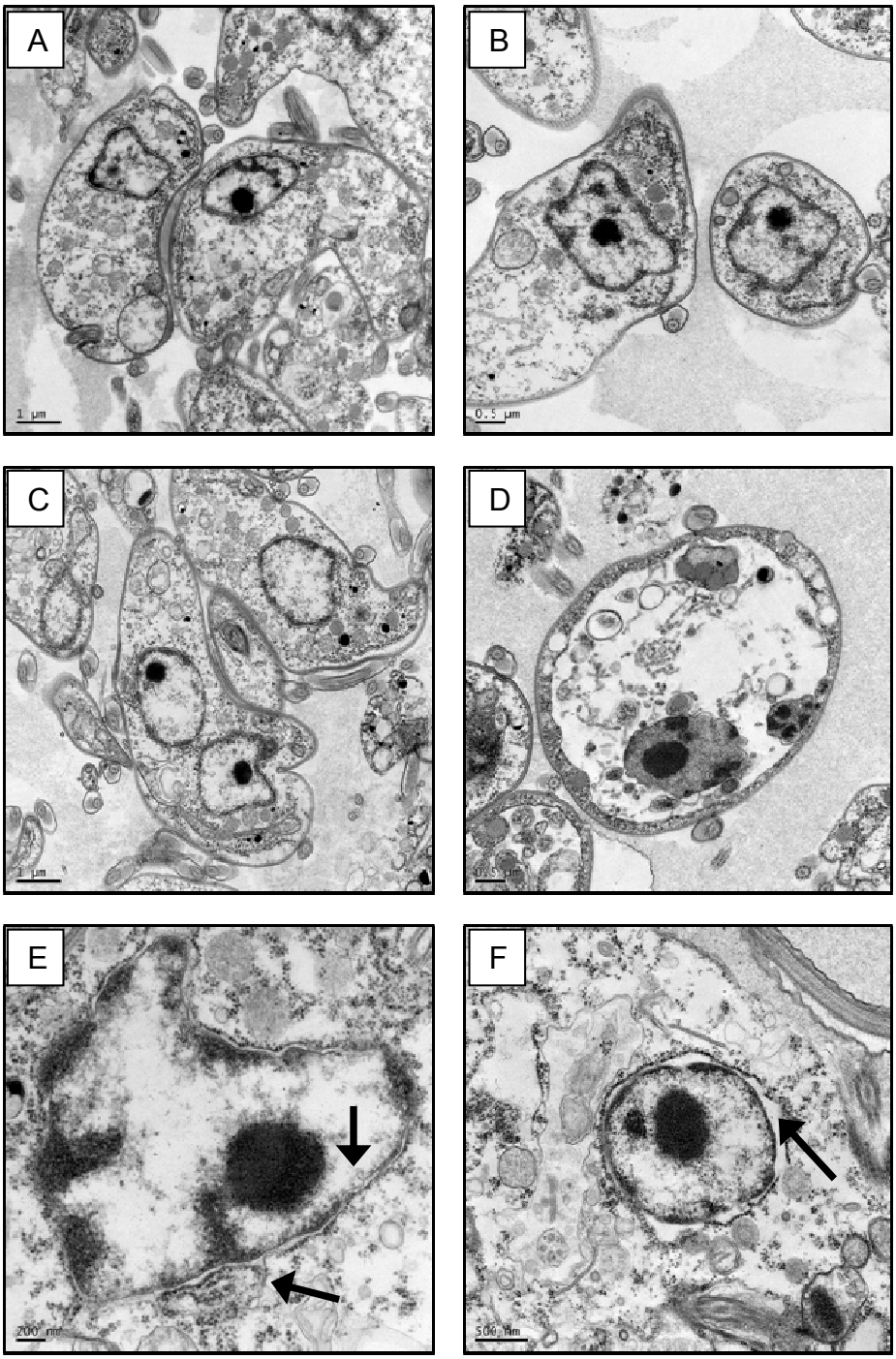


Figure 32: Transmission electron microscopy (TEM) of RNAi cell lines 24 hours post induction. Scale bars: Control (Panel A) and TbNup96-RNAi (Panel C), 1 µm; TbGle2-RNAi (Panel B), TbSec13-RNAi (Panel D), and TbNup158-RNAi (Panel F), 0.5 µm; TbNup89-RNAi (Panel E), 0.2 µm.

The knockdown of TbNup89, while not lethal, causes the deformation of the nuclear envelope (there is localized blebbing and invaginations at the INM and ONM, as seen in Figure 32 Panel E) as well as an exaggerated ER structure in some cases. Like Sec13, this suggests that TbNup89 plays a significant, if not essential, role in membrane maintenance, and its loss causes perturbation of NE curvature leading to excess ER and NE formations. This is analogous to the phenotype when ScNup53 is overexpressed in yeast. Overproduction of ScNup53 produces excess intranuclear tubular and double membrane structures, and led to the speculation that the NPC plays a critical role in NE formation and stabilization (Marelli et al., 2001).

TbNup158 plays an essential role in preserving proper membrane structure (Figure 32 Panel F). Knockdown of TbNup158 cause many nuclei to become denuded of their ONM to varying degrees (Figure 32 Panel F). The TEM field contains nuclei in various stages of ONM separation, suggesting a slow onset of this phenotype. This is perhaps either due to a significant-residual pool of TbNup158, the long half-life of TbNup158, or, most likely, a combination of both. Despite the slow onset of the observed phenotype, the knockdown of TbNup158 apparently prevents nascent NPC formation or destabilizes NPCs once the preexisting pool of TbNup158 is exhausted. The dysfunctional TbNPC no longer structurally supports the annuli across the NE. Without this support, the curvature of the annuli collapses and the pore is sealed, the result of which is the separation of the NE.

This would suggest that trypanosomes lack a protein network across the NE from the cytosol to nucleoplasm. Conceptually, such a network should maintain NE stability in the absence of NPCs, as is the case in yeast or vertebrates. This hypothesis is consistent with the lack of detectable opisthokont NE protein domains within the TbNEP (this study). It is also noteworthy that only one protein with a Sad1/Unc domain has been identified by bioinformatics within the entire trypanosome genome. The apparent lack of a trypanosome protein network would explain why knockdowns of orthologous Nups in the opisthokonts do not exhibit such severe membrane deformation phenotypes.

With the majority of the TbNPC components identified and characterized, we can now draw comparisons between the NPCs of *H. sapiens*, *S. cerevisiae*, and *T. brucei*. In the next chapter, we compare these three examples of well characterized NPCs in the context of eukaryotic evolution.

CHAPTER 6 — DISCUSSION AND FUTURE DIRECTIONS

COMPOSITION AND STRUCTURE OF THE TBNPC

Through various biochemical, proteomic and bioinformatic strategies, we have identified 24 putative TbNups and 8 putative transport factors (Chapter 3). Our analysis of the TbNPC reveals a composition homologous to the well characterized opisthokont NPCs. We estimate that we have identified the majority, if not all, of the structural scaffold TbNups (Figure 13, and Table 5). We base this estimate on the following line of reasoning. Here, we assume the TbNups have a similar stoichiometry in the NPC as do those in opisthokont NPCs. This implies that TbNups are present in a relative abundance that is within a factor of 4 of one another. Because we can recognize scaffold Nups by their characteristic secondary structure fold patterns, and because the dynamic range of our analysis is ~1000 (see Chapter 2), it seems unlikely that we have missed many scaffold Nups. We conducted a screen of all the proteins in the TbNEP dataset for secondary structure fold patterns that are characteristic of the Nups that comprise the structural scaffold. These fold patterns are the β-propeller domain, α-solenoid domain and an N-terminal β-propeller connected to an α-solenoid domains (β-α domain) (Devos et al., 2006). Our screen identified 17 proteins with such secondary structure patterns. We were able to eliminate from consideration 5 of these proteins using sequence similarity (Chapter 3) and fluorescent localization (Chapter 4), though it is possible that the introduction of

an epitope tag disrupted proper localization. This process resulted in the identification of 12 putative scaffold TbNups.

Earlier work suggests that the fold patterns discussed above are very likely to pre-date the emergence of the NPC at the LCEA (Devos et al., 2004). If true, then these fold patterns should be present within the contemporary NPC structural scaffold of all eukaryotes. The presence of these fold types within the structural scaffold of *T. brucei* supports this hypothesis.

Within the structural scaffold, the β-propellers are a particularly well-conserved family of Nups (Figure 15). Previously, it has been observed that Sec13 and the transport factor Gle2/Rae1 are well-conserved across the opisthokonts (Murphy et al., 1996; Siniossoglou et al., 1996). Here, we observed that they are conserved in trypanosomes as well. Trypanosomes, plants and animals have additional β-propellers that have apparently been lost in yeast (Appendix A). One example is an HsALADIN homolog, which is a disease locus in humans (Cronshaw and Matunis, 2003), that we observe to be present in trypanosomes. Interestingly, plant β-propellers have less sequence similarity with trypanosome (both are bikonts), than with vertebrates (an opisthokont) (Figure 15).

The number and size of the α-solenoid domains have essentially remained unchanged from yeast, plants, humans, and trypanosomes (Appendix A and Figure 13). We observe that plants share more significant sequence similarity to humans, than humans do with yeast, indicative of the relative

divergence of the fungi from the rest of the opisthokonts (Figure 16). Trypanosomes generally exhibit low sequence similarity to yeast, plants or humans, but we observe that the secondary structure is well-conserved. For example, the nucleoporin interacting component (NIC) domain of ScNIC96/HsNup93 has greatly diverged in the trypanosomes and plants, relative to the opisthokonts (Chapter 3 and Figure 16).

With respect to primary structure, HsNup155 and ScNups157/170 are the only β-α domain Nups that are conserved across Eukaryota (Figure 17). The other two yeast Nups of this family, ScNup120 and ScNup133, share little to no significant similarity to the plant, human and trypanosome sequences. Once again, the plant Nups of this family, AtSAR1 and AtNup133, share more significant similarity with HsNup160 and HsNup133, respectively, than with the yeast and trypanosome Nups (Figure 17).

The trypanosome Mlps share very little significant sequence similarity to yeast, human and plants (Figure 18). ScMlp1 and HsTpr share more similarity than ScMlp2 and Tpr, which suggests a duplication of the original ScMlp1 gene and explains the additional functional roles of ScMlp2 (Galy et al., 2004; Niepel et al., 2005). The two trypanosome Mlps that we identified appear to have undergone extensive species-specific innovation and are not paralogous (Figure 18).

In the case of the FG-Nups, we are confident that we have identified and characterized nearly all, if not all, of the FG-Nups of the TbNPC (Figure 13, and

Table 5). Since the FG-dipeptide is a requisite motif of FG-Nups, we scanned the entire *T. brucei* proteome for the presence of this motif (Chapter 3), and experimentally identified (with mass spectrometry) all of the putative FG-Nups found through the *in silico* search (Chapter 2).

Between species, the natively disordered FG-Nups have a higher rate of amino acid substitution, relative to other Nup groups (Denning et al., 2003; Denning and Rexach, 2007). The trypanosome FG-Nups also exhibit the high substitution rates observed in the opisthokonts. However, when we examined the chemical composition of the spacer region, we observed that the chemical composition is conserved (Chapter 3). Furthermore, we identified a well-conserved anti-correlation between glycine and charged residue (Asp, Glu, Arg, and Lys, DERK) composition within the FG-repeat spacer regions, and found that this glycine-DERK relationship correlates to the predicted FG-Nup localizations within the NPC (Figure 19). Amongst the asymmetrically distributed FG-Nups, those that localize to the nuclear face of the NPC tend to have a higher composition of charged residues, compared to cytoplasmically localized FG-Nups. This anti-correlation suggests that the rapid evolution within the natively disordered FG-repeat domains is constrained within a specific chemical composition that is directly related to the localization of the FG-Nups within the NPC.

One model of nucleocytoplasmic transport suggests that the Kap-Cargo complexes appear to transiently interact with FG-Nups and pass through the NPC by Brownian motion (Rout et al., 2000). In addition to the

RanGDP/RanGTP gradient, there has been speculation about the function of asymmetrically distributed FG-Nups in providing a vectorial component to nucleocytoplasmic transport, thus increasing efficiency (Bayliss et al., 2000a; Rout and Aitchison, 2000). The above anti-correlation supports such a hypothesis of a vectorial transport component in the FG-Nups. In this model, a translocating complex passes through the central channel by diffusion, at which point it interacts with the glycine rich, DERK depleted, FG-repeat domains, which are predicted to symmetrically line the channel (with respect to a plane parallel to the NE). Asymmetry could thus contribute a vectorial transport component by localizing the more highly charged, and perhaps more strongly interacting, FG-Nups at the periphery of the NPC. An additional vectorial transport component may also be present, given the higher charged residue composition within the FG-Nups that localize to the nuclear face of the NPC, relative to those at the cytoplasmic face (Figure 19).

We have begun to reveal the nature of the TbNPC through functional and localization studies (Chapters 4 and 5). Thus, through RNAi studies, we revealed that the TbNPC, through specific TbNups such as TbNup158 and TbNup89, functionally secures the ONM and INM together as well as maintains proper membrane folding and structure (Figure 28). Unlike the opisthokonts, the TbNPC distribution around the nucleus approaches uniformity in that we do not observe any evidence of clustering (which would lead to an uneven distribution of NPCs over the surface of the NE). We speculate that the distribution of TbNPC is most likely driven by the network of TbNUP-1 molecules (Figure 31). This is in

contrast to the vertebrates, whose NPC distribution is dictated by a heterogeneous lamin network (with NPC-free patches), and the fungi, whose NPC distribution is non-uniform (Maeshima et al., 2006; Winey et al., 1997). The surface density of NPC around the trypanosome nucleus (~ 5 NPCs/μm^2) is about half that of yeast (~ 12 NPCs/μm^2) and slightly less than the vertebrate average density (~ 7 NPCs/μm^2) (Maeshima et al., 2006; Winey et al., 1997). To a first approximation, the density of the TbNPC is stable as the nuclear dimensions change through the cell-cycle, unlike the vertebrates (Figures 30 & 31). With the TbNPC now well characterized, we can now draw comparisons to the opisthokont NPC.

THE EVOLUTION AND ORIGINS OF THE NPC

We find that the kinetoplastid NPC is just as complex, with respect to composition, as the opisthokont NPC. Vertebrates, yeast and trypanosomes contain similar numbers of structural scaffold Nups (14, 12 and 12, respectively) and FG-Nups (9, 12 and 9, respectively) as well as individual domains. What are lacking from our TbNPC inventory are the trypanosome nucleoporin-specific evolutionary innovations at the NPC. These TbNups, if present, are unlikely to be identified by sequence similarities. Rather, they will likely be identified in the future by immunoisolations of known TbNups. The long evolutionary distance of trypanosomes from the opisthokonts may explain our inability to identify any trypanosome domains homologous to ScNup82, ScNup159, HsNup358, HsNup214 and the membrane Nups, if these TbNups are present at all. We now

discuss the evolution of the NPC in the context of two contemporary models of eukaryotic evolution.

One model of eukaryotic evolution, the crown eukaryote model, places the root of the eukaryotic tree proximal to the diplomonads, kinetoplastids and alveolates, which are outgroups to the plants, fungi and metazoa, collectively known as the crown eukaryotes (Hedges, 2002; Templeton et al., 2004). In this model, one might predict that the LCEA had a simpler NPC, relative to the contemporary NPC, and was comprised of a small number of foundational components. Then as the eukaryotic lineages diverged, the NPC would then become increasingly complex in a species-specific manner. The crown eukaryotes would have experienced a burst of innovation and complexity relative to more simple lineages, such as the excavates. Alternatively, albeit less probable, each lineage could evolve a complex NPC, but these structures would be less likely to be related to each other. In either scenario, this model would predict significant compositional differences in the NPCs between lineages, such as different constituents within the scaffold, novel architecture, or different inventories of FG-Nups. However, analogous Nups could potentially arise through convergent evolution based on natural selection.

A second model of eukaryotic evolution, based on a systematic study of gene fusion events (Stechmann and Cavalier-Smith, 2002), does not place the excavates as an outgroup to the root of Eukaryota; rather it places the root between the bikonts (excavates and plants) and the opisthokonts (fungi and animals) (Adl et al., 2005). This rooting of the eukaryotic tree of life does not

place an emphasis on increasing complexity through evolution to resolve the systematic relationship between distant eukaryotes. The distinction is that the excavata are different from the crown eukaryotes, and such differences are the result of divergent evolution. Within this model, one would predict the NPC of the LCEA to be nearly as complete and complex (save for species-specific innovations, such as gene duplications or deletions) as the contemporary NPC across the tree of eukaryotic life.

A previous large scale bioinformatic analysis of the NPC and NE supported the crown eukaryotic model, based on the authors' conclusions (Mans et al., 2004). However this study is challenged by the poor sequence similarities discussed in Chapter 3. Thus, relative to the opisthokonts, the diplomonad and kinetoplastid components of the NPC and NE were erroneously observed to be significantly simpler. As an alternative, these workers justified a possibly complex NPC as a product of convergent evolution. Another large scale bioinformatic study adopted the Cavalier-Smith model based on studies of apparent relative evolutionary clock rates of Nups (Bapteste et al., 2005).

The results described in this thesis support the bikont-opisthokont model of the eukaryotic tree of life (Cavalier-Smith, 2002). We have determined that the kinetoplastid NPC is not significantly simpler, nor significantly different, than that of the crown eukaryotes (Chapter 3). It is possible that the similarity of the structural scaffold and the FG-Nups between the excavates and the opisthokonts could have arisen by convergent evolution, if the driving forces of evolution were similar across all eukaryotes. However, we would argue that convergent

evolution is a less probable scenario given the extraordinary number of similarities between the excavate and opisthokont NPC, which include: (1) the similar inventory, protein length and domain structure of the structural scaffold Nups (Chapter 3 and Figure 13); (2) the similar number, and conserved chemical composition, of the rapidly evolving FG-Nups (Chapter 3 and Figures 13 and 19); and (3) the presence of the highly conserved β-sandwich autoproteolytic domain. It is more probable that the large number of similarities within the NPC across clades point to a common origin of a complex NPC followed by divergent evolution. It follows that the LCEA possessed a fully complex NPC, somewhat analogous to the contemporary NPC. Furthermore, the contemporary eukaryotic NPC is a product of divergent evolution and only a minority of Nups arose, or was lost, due to species-specific innovations.

Although the bikont-opisthokont model places the root between the bikonts and opisthokonts, the alignments between the plants and the vertebrates are often fairly significant. This supports the model of a complex NPC at the LCEA and a more rapid evolutionary clock rate for the excavates, relative to plants, fungi and animals. This would explain the divergence of primary structure of excavates from both the plants and opisthokonts within the structural scaffold (Figures 15-17). The fungi have also been subjected to a rapid evolution clock, relative to animals, with its primary structure divergence and gene duplication events (Figures 15-17).

One evolutionary innovation to the early machine is the β-sandwich domain within the autoproteolytic Nup, HsNup98-96/ScNup145. Although the conserved β-sandwich domain is present, the excavates and many bikonts either do not cleave or do not express full-length products (Chapter 3). At least two scenarios exist — in one, the ancestral Nup98-96 contained a fully functional autoproteolytic domain; in another, it did not. In the former scenario, the kinetoplastids lost the autoproteolytic function while the opisthokonts and some excavates retained the function. The latter scenario would suggest a lack of innovation within the kinetoplastids (and possibly other clades). The data presented in this thesis supports the former scenario. If the LCEA expressed a full-length protein that was successively cleaved and the kinetoplastids subsequently lost this function while retaining the conserved β-sandwich domain, then it may be possible to restore the original function by introducing an artificial cleavage site. To investigate this possibility, we introduced a truncation into TbNup158 that removed the C-terminal domain. We observed that the truncated protein, TbNup158(1-837), localized to abnormal discreet foci in the nuclear periphery (Chaper 6). This suggests the presence of additional trypanosome-specific innovations in addition to the loss of the catalytic triad.

We predict that the early Nup98-96 was expressed, and functioned, as a full length protein in the LCEA NPC. The protein contained a complete autoproteolytic β-sandwich domain, which is now conserved across all of Eukaryota, and included the catalytic triad. In this scenario, several bikonts either lost the ability self cleave and/or truncated the full length protein within the

genome. Aside from the kinetoplastids, which do not possess an active cleavage cite, the diplomonads, amoebozoa (a unikont), and plants all express truncated proteins which are homologous to ScNup145N and HsNup98. These orthologs contain the FG-repeat domain at the N-terminus and the conserved β-sandwich autoproteolytic domain at the C-terminus. Interestingly, while *Giardia* lacks the HFS catalytic triad, the unikont *E. histolytica* and plants contain a HF[S/T] tripeptide close to their C-terminus.

A possible example of convergent evolution is the emergence of the TbMlps. While the two TbMlps have a coiled coiled domain and localize to the TbNPC (Figure 30), they share negligible sequence similarity and length to the ScMlps or HsTpr (Figure 18). Furthermore, the TbMlps are not the product of a gene duplication event as they are not paralogous to each other, unlike the ScMlps. However, they seemed to have evolved an analogous function to the yeast Mlps. While TbMlp1 stably localizes to the NPC, as does ScMlp1, TbMlp2 apparently localizes to the area proximal to the spindle pole body, as does ScMlp2 (Niepel et al., 2005). A detailed analysis of *S. pombe* reveals a similar situation to trypanosomes — two Mlps that arose independently, and of which only one localizes to the area proximal to the spindle pole body (Ding et al., 2000). Interestingly, TbNUP-1 which localizes to the NE between the NPCs, does share homology to the ScMlps and HsTpr (Figures 18 and 32). Perhaps TbNUP-1 is related to the nuclear basket Nups within the LCEA, but diverged to fill another function at the inner nuclear envelope.

It is of note that the *T. brucei* NPC is more akin to the vertebrate NPC than to yeast in many respects, including the existence of the HsALADIN homolog and the tendency for pair-wise alignments to be more significant between trypanosomes and vertebrates, relative to yeast (Chapter 3). While it may appear convenient to suggest that vertebrates evolved additional β-propellers due to special requirements of open-mitosis, our finding in trypanosomes complicates this simple explanation. The presence of ALADIN, a disease related locus in vertebrates, and other β-propellers within trypanosomes means that it was most likely present in the ancestral NPC and was subsequently lost in the fungi. *S. cerevisiae* and its fungi relatives appear to have simplified its inventory of β-propellers. Furthermore, the yeast primary structures of structural scaffold Nups have undergone a significant amount of species-specific adaptation relative to humans, plants and trypanosomes (Figures 15-17).

It has been postulated that the nucleus likely arose only once; no obvious intermediates are known to persist today (Martin, 2005). Thus, the evolution of the nucleus within the trunk of the eukaryotic tree of life is irresolvable. Since the ancestral NPC (LCEA) was already a complex machine at some point within the trunk, the step-wise evolution of the NPC cannot be resolved by comparative genomics until the root of the eukaryotic tree has been firmly established. However, some clues about the emergence of the nucleus have been deduced using the predicted structures of the structural scaffold Nups leading to the protocoatomer hypothesis (Devos et al., 2004). This hypothesis observes that these structural folds are also found in the eukaryotic vesicle coating complexes.

This suggests a concomitant divergent evolution of the endomembrane system and membrane curving proteins into the diverse set of proteins that are present today. We find that the presence of an extensive repertoire of β-propeller and α-solenoid domains within the TbNPC supports the protocoatomer hypothesis (Devos et al., 2004).

During evolution, the endomembrane system, which would eventually include the nucleus, likely evolved in a manner predicted by either the invagination model (Cavalier-Smith, 2002) or the vesicle model (Martin, 1999). As discussed in Chapter 1, a subset of Nups is involved in processes other than nucleocytoplasmic transport (e.g. mitotic checkpoint and chromatin interactions). We posit that their function may be remnants of ancient transitional steps as the primitive endomembrane system began to interact with chromatin and differentiate into different domains. Consistent with our data, the nucleus does not appear to be the product of a symbiotic event.

FUTURE DIRECTIONS

Our characterized inventory of the TbNPC, provides a strong basis for further experimental investigations. We have tagged several key TbNups with GFP, and so immunoisolations are now feasible. These experiments will identify trypanosome-specific Nups, which may only be very weakly, homologous to the opisthokont-specific Nups. Also, we anticipate that such immunoisolations will reveal the identity and nature of the membrane bound TbNups, if present. Also, such work will characterize the interaction network and molecular architecture within the TbNPC. A comparison of the molecular architectures of several

distantly related NPCs would shed light on the evolution of the NPC as well as additional functional features on the NPC.

Our RNAi work has yielded fascinating nuclear envelope morphology phenotypes, which should be studied further. The lack of NE integrity upon knockdown of TbNup158 is quite provocative. A likely scenario is that elimination of TbNup158 causes loss of a large sub-complex of structural Nups, perhaps the *T. brucei* analog to the central ScNup84/HsNup107 complex. The crippled NPCs can no longer secure the NE. This hypothesis should be testable by using a large-scale screen of TbNup knockdowns. A set of lynchpin TbNups can be identified as crucial for NPC and NE form and function.

Combining the RNAi constructs with the GFP tagged Nups would be a powerful assay for functional studies. For example, after induction of RNAi, one could record the fluorescence pattern at the nuclear periphery compared to the wild-type control. This would provide a direct observation into how the NE responds to the stress of the depletion of a specific TbNup.

Furthermore, one could target the GFP moiety directly with RNAi. In this experiment, the C-terminal GFP tagged TbNup would be depleted upon RNAi induction with doxycycline. The second allele would be wild-type, so no phenotype should be observed. After several divisions, the cells should have little to no GFP signal at the NE. At this point, the doxycycline would be washed away, and cells would be observed at various time points for the reemergence of the GFP signal relative to antibody staining against the NPC (using either

αHsNup107 or an antibody raised against a TbNup peptide). This would reveal information on Nup turnover and incorporation into nascent NPCs.

Finally, the pseudo-autoproteolytic Nup, TbNup158, should be explored in depth. In this work, we tagged the N-terminus of TbNup158 with GFP, and the result was localization into the discreet foci. This observation is contrasted by the opisthokonts, where the N-terminal fragment is localized, in part, to the nucleoplasm as a soluble protein. The C-term fragment should be examined in an analogous fashion to see if the α-solenoid would correctly integrate into the NPC without the N-terminal domain. Also the protein could be mutated, in a site-directed fashion, to insert the tripeptide catalytic site HFS. It would be very interesting to note whether the insertion of the tripeptide is sufficient for autoproteolysis, or if other residues (such as a stabilizing lysine residue) are also requisite.

In this thesis, we identified over ~850 proteins in the TbNEP using five biochemical and proteomic strategies (Chapter 2). We then identified and characterized 24 putative TbNups as well as 8 transport factors using an array of bioinformatic strategies (Chapter 3). We went on to characterize a subset of the TbNups by fluorescent localization (Chapter 4) and functional analysis with RNAi (Chapter 5). The result of these experiments reveals a TbNPC that approaches a uniform distribution around the NE (Figure 31) and plays a large role in stabilizing the morphology of the NE (Figure 28).

Our data supports the bikont-opisthokont model of eukaryotic evolution, which places the root of the tree between the bikonts and the opisthokonts. The TbNPC is just as complex as, and not simpler than, the opisthokont NPC, with respect to composition. This suggests that the LCEA possessed a complex NPC analogous to the contemporary NPC, which then diversified in a largely divergent manner.

APPENDIX A — TABLE OF CLOSELY RELATED NUCLEOPORIN ORTHOLOGS

To build a consensus model of the metazoan and fungi NPC, we constructed a table of Nup orthologs of closely related metazoans, fungi and plant (Table 3). Orthologs were identified by a PSI-BLAST pair-wise alignment search of *H. sapien* and/or *S. cerevisiae* query Nups against the respective protein databases of subject organisms. The top five subject alignments were then queried with PSI-BLAST back against the *H. sapien* and *S. cerevisiae* protein database. If the reciprocal BLAST alignment revealed a top match for the original Nup query sequence, then the pair was considered valid and the ortholog was recorded by its accession number.

Even within closely related species, not all Nups can be accounted for, such as the lack of HsNup37 and HsALADIN orthologs in *C. elegans*, and the absence of ScNup120 in *C. neoformans*. Also, there is evidence of two duplication events in *S. pombe* (ScNIC96 and ScNup133), but only in *S. cerevisiae* are there two paralogous orthologs each of HsNup155 and HsNup98 (ScNup157/ScNup170 and Sc100/116). The additional β-propellers of the metazoan family are absent in the fungi, but are present in the plants. This is suggestive that the fungi lost these additional Nups early in their evolution.

Table 4: Table of closely related Nup orthologs. Abbreviations: "nf", not found; "nd", not determined.

Domains	Metazoa					Fungi				Viridiplantae
	H. sapiens	*D. melanogaster*	*C. elegans*	*C. intestinalis*	*S. cerevisiae*	*A. nidulans*	*S. pombe*	*C. neoformans*	*A.thaliana*
Membrane	NDC1(37)	Q9BTX1	74868322	75015300	nf	NDC1(5)	NP_013681	AN4417	63054586	nf	nf
	Gp210(15)	NP_079199	24585893	17509225	286609		nf	nf	nf	nf	AT5G40480
	Pom121(22)	Q9Y2N3	nf	nf	nf		nf	nf	nf	nf	nf
		nf	nf	nf	nf	Pom34(34)	NP_013118	nf	nf	cn11330	nf
		nf	nf	nf	nf	Pom152(43)	NP_013848	AN3454	19112844	cn03050	nf
β Propeller	Sec13(12)	NP_899195	21356113	17544258	225177	Sec13(35)	Q04491	AN4317	19113484	cn08323	AT3G01340
	Seh1(12)	NP_001013455	24586465	3880929	285146	Seh1(35)	EDN62018	AN5889	19114663	cn02546	AT2G30050
	Nup37(6)	Q8NFH4	24649978	nf	211263		nf	nf	nf	nf	AT1G64350.1
	ALADIN(6)	NP_056480	18858057	nf	397606		nf	nf	nf	nf	nf
	Nup43(6)	Q8NFH3	24647938	17510671	293411		nf	nf	nf	nf	AT3G56900
	Rae1(28)	NP_003601	20130217	3123163	236689	Gle2(30)	NP_011033	AN1379	19113576	cn13077	AT4G30840
α Solenoid	Nup107(33)	NP_065134	119638193	17532901	270205	Nup84(3)	NP_010167	AN1190	68009955	cn08275	AT1G80670
	Nup75(6)	NP_079120	119638079	17508241	287889	Nup85(16)	EDN63361	AN9109	19112685	cn02170	AT3G14120
	Nup93(20)	NP_055484	18859665	13491276	213542	NIC96(21)	CAA51427	AN6980	19075969	cn03290	AT4G32910
	Nup188(29)	NP_056169	24653170	nf	222679	Nup188(31)	NP_013604	AN10112	19075922	nf	AT2G41620
	Nup205(20)	NP_055950	24643409	3878524	nf	Nup192(24)	NP_012495	AN10002	19114326	cn06299	AT4G38760
									19075899	cn05009	AT5G51200
β Propeller; α Solenoid	Nup160(41)	NP_056046	13124656	25150095	293089	Nup120(1)	NP_012866	AN1238	19113466	nf	AT1G33410
	Nup155(32)	NP_705618	17316372	30025080	274239	Nup157	NP_011031	AN6738	19115736	cn04508	AT1G14850
						Nup170(2)	NP_009474				
	Nup133(41)	NP_060700	119638237	68846777	261423	Nup133(9)	NP_013008	AN4293	19114827	cn12020	AT2G05120
									19112173		
FG Repeat; Nup98; α Solenoid	Nup98-96(13)	56549645	116008420	7511089	nf	Nup145(10)	CAA83584	AN5627	19115005	cn02461	AT1G80680
	Nup98		nf	nf	230079	Nup100	NP_012855	nf	nf	nf	AT1G10390
						Nup116(42)	EDN64437				AT1G59660
FG Repeat	Nup35(6)	NP_612142	18859897	115502433	294938	Nup53(27)	NP_013873	nf	19114216	nf	nd
		nf	nf	nf	nf	Nup59(27)	AAT92991	nf	19114216	nf	nd
	Nup54(11)	NP_059122	38258270	17541254	272199	Nup57(18)	NP_011634	AN1064	19112773	cn10246	nd
	Nup58(11)	NP_054808	17946038	17510275	256158	Nup49(42)	NP_011343	AN2431	19113970	cn01387	nd
	Nup62(7)	NP_036478	20130049	17507541	263011	Nsp1(23)	NP_012494	AN4499	19115070	cn04317	nd
	Nup153(40)	NP_005115	24642563	17509085	294569	Nup1(8)	NP_014741	nf	nf	cn01634	nd
	Nup214(25)	EAW87961	51316547	17505452	nf	Nup159(17)	NP_012151	AN2086	19115455	cn06023	nd
	Nlp1(38)	AAI07584	nf	nf	292234	Nup42(34)	NP_010478	AN4595	nf	cn04177	nd
	Nup50(36)	NP_009103	19921802	17556290	262262	Nup2(26)	NP_013439	AN5485	19075437	cn01191	nd
	Nup358(44)	BAA07662	45550830	32565674	nf		nf	nf	nf	nf	nd
Coiled coil	Tpr(4)	P12270	1923274	71993549	293800	Mlp1(39)	AAA34783	AN5499	19075736	cn14006	AT1G79280
		nf	nf	nf	nf	Mlp2(39)	NP_012117	nf	19115004	nf	nf
Other	Nup88(14)	NP_002523	51316527	nf	272352	Nup82(19)	EDN63515	nf	19113207	nf	AT5G0568

References for Table 4

1 Aitchison, J.D., G. Blobel, and M.P. Rout. 1995. Nup120p: A yeast nucleoporin required for NPC distribution and mRNA transport. Journal of Cell Biology. 131:1659-1675.

2 Aitchison, J.D., M.P. Rout, M. Marelli, G. Blobel, and R.W. Wozniak. 1995. 2 Novel Related Yeast Nucleoporins Nup170p and Nup157p - Complementation with the Vertebrate Homolog Nup155p and Functional Interactions with the Yeast Nuclear Pore-Membrane Protein Pom152p. Journal of Cell Biology. 131:1133-1148.

3 Bastos, R., L.R. dePouplana, M. Enarson, K. Bodoor, and B. Burke. 1997. Nup84, a novel nucleoporin that is associated with CAN/Nup214 on the cytoplasmic face of the nuclear pore complex. Journal Of Cell Biology. 137:989-1000.

4 Byrd, D.A., D.J. Sweet, N. Pante, K.N. Konstantinov, T.L. Guan, A.C.S. Saphire, P.J. Mitchell, C.S. Cooper, U. Aebi, and L. Gerace. 1994. Tpr, A Large Coiled-Coil Protein Whose Amino-Terminus Is Involved In Activation Of Oncogenic Kinases, Is Localized To The Cytoplasmic Surface Of The Nuclear-Pore Complex. Journal Of Cell Biology. 127:1515-1526.

5 Chial, H.J., M.P. Rout, T.H. Giddings, and M. Winey. 1998. Saccharomyces cerevisiae Ndc1p is a shared component of nuclear pore complexes and spindle pole bodies. Journal Of Cell Biology. 143:1789-1800.

6 Cronshaw, J.A., A.N. Krutchinsky, W.Z. Zhang, B.T. Chait, and M.J. Matunis. 2002. Proteomic analysis of the mammalian nuclear pore complex. Journal of Cell Biology. 158:915-927.

7 Davis, L.I., and G. Blobel. 1986. Identification And Characterization Of A Nuclear-Pore Complex Protein. Cell. 45:699-709.

8 Davis, L.I., and G.R. Fink. 1990. The Nup1 Gene Encodes An Essential Component Of The Yeast Nuclear-Pore Complex. Cell. 61:965-978.

9 Doye, V., R. Wepf, and E.C. Hurt. 1994. A Novel Nuclear-Pore Protein Nup133p With Distinct Roles In Poly(A)(+) Rna Transport And Nuclear-Pore Distribution. Embo Journal. 13:6062-6075.

10 Fabre, E., W.C. Boelens, C. Wimmer, I.W. Mattaj, and E.C. Hurt. 1994. Nup145p Is Required For Nuclear Export Of Messenger-Rna And Binds Homopolymeric Rna In-Vitro Via A Novel Conserved Motif. Cell. 78:275-289.

11 Finlay, D.R., E. Meier, P. Bradley, J. Horecka, and D.J. Forbes. 1991. A Complex Of Nuclear-Pore Proteins Required For Pore Function. Journal Of Cell Biology. 114:169-183.

12 Fontoura, B.M.A., G. Blobel, and M.J. Matunis. 1999. A conserved biogenesis pathway for nucleoporins: Proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. Journal of Cell Biology. 144:1097-1112.

13 Fontoura, B.M.A., G. Blobel, and M.J. Matunis. 1999. A conserved biogenesis pathway for nucleoporins: Proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. Journal of Cell Biology. 144:1097-1112.
14 Fornerod, M., J. vanDeursen, S. vanBaal, A. Reynolds, D. Davis, K.G. Murti, J. Fransen, and G. Grosveld. 1997. The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. Embo Journal. 16:807-816.
15 Gerace, L., Y. Ottaviano, and C. Kondorkoch. 1982. Identification Of A Major Polypeptide Of The Nuclear-Pore Complex. Journal Of Cell Biology. 95:826-837.
16 Goldstein, A.L., C.A. Snay, C.V. Heath, and C.N. Cole. 1996. Pleiotropic nuclear defects associated with a conditional allele of the novel nucleoporin Rat9p/Nup85p. Molecular Biology Of The Cell. 7:917-934.
17 Gorsch, L.C., T.C. Dockendorff, and C.N. Cole. 1995. A Conditional Allele Of The Novel Repeat-Containing Yeast Nucleoporin Rat7 Nup159 Causes Both Rapid Cessation Of Messenger-Rna Export And Reversible Clustering Of Nuclear-Pore Complexes. Journal Of Cell Biology. 129:939-955.
18 Grandi, P., N. Schlaich, H. Tekotte, and E.C. Hurt. 1995. Functional Interaction Of Nic96p With A Core Nucleoporin Complex Consisting Of Nsp1p, Nup49p And A Novel Protein Nup57p. Embo Journal. 14:76-87.
19 Grandi, P., S. Emig, C. Weise, F. Hucho, T. Pohl, and E.C. Hurt. 1995. A Novel Nuclear-Pore Protein Nup82p Which Specifically Binds To A Fraction Of Nsp1p. Journal Of Cell Biology. 130:1263-1273.
20 Grandi, P., T. Dang, N. Pane, A. Shevchenko, M. Mann, D. Forbes, and E. Hurt. 1997. Nup93, a vertebrate homologue of yeast Nic96p, forms a complex with a novel 205-kDa protein and is required for correct nuclear pore assembly. Molecular Biology Of The Cell. 8:2017-2038.
21 Grandi, P., V. Doye, and E.C. Hurt. 1993. Purification Of Nsp1 Reveals Complex-Formation With Glfg Nucleoporins And A Novel Nuclear-Pore Protein Nic96. Embo Journal. 12:3061-3071.
22 Hallberg, E., R.W. Wozniak, and G. Blobel. 1993. An Integral Membrane-Protein Of The Pore Membrane Domain Of The Nuclear-Envelope Contains A Nucleoporin-Like Region. Journal Of Cell Biology. 122:513-521.
23 Hurt, E.C. 1988. A Novel Nucleoskeletal-Like Protein Located At The Nuclear Periphery Is Required For The Life-Cycle Of Saccharomyces-Cerevisiae. Embo Journal. 7:4323-4334.
24 Kosova, B., N. Pante, C. Rollenhagen, and E. Hurt. 1999. Nup192p is a conserved nucleoporin with a preferential location at the inner site of the nuclear membrane. Journal Of Biological Chemistry. 274:22646-22651.
25 Kraemer, D., R.W. Wozniak, G. Blobel, and A. Radu. 1994. The Human Can Protein, A Putative Oncogene Product Associated With Myeloid Leukemogenesis, Is A Nuclear-Pore Complex Protein That Faces The

Cytoplasm. Proceedings Of The National Academy Of Sciences Of The United States Of America. 91:1519-1523.
26 Loeb, J.D.J., L.I. Davis, and G.R. Fink. 1993. Nup2, A Novel Yeast Nucleoporin, Has Functional Overlap With Other Proteins Of The Nuclear-Pore Complex. Molecular Biology Of The Cell. 4:209-222.
27 Marelli, M., J.D. Aitchison, and R.W. Wozniak. 1998. Specific binding of the karyopherin Kap121p to a subunit of the nuclear pore complex containing Nup53p, Nup59p, and Nup170p. Journal Of Cell Biology. 143:1813-1830.
28 Matsuoka, Y., M. Takagi, T. Ban, M. Miyazaki, T. Yamamoto, Y. Kondo, and Y. Yoneda. 1999. Identification and characterization of nuclear pore subcomplexes in mitotic extract of human somatic cells. Biochemical And Biophysical Research Communications. 254:417-423.
29 Miller, B.R., M. Powers, M. Park, W. Fischer, and D.J. Forbes. 2000. Identification of a new vertebrate nucleoporin, Nup188, with the use of a novel organelle trap assay. Molecular Biology Of The Cell. 11:3381-3396.
30 Murphy, R., J.L. Watkins, and S.R. Wente. 1996. GLE2, a Saccharomyces cerevisiae homologue of the Schizosaccharomyces pombe export factor RAE1, is required for nuclear pore complex structure and function. Molecular Biology Of The Cell. 7:1921-1937.
31 Nehrbass, U., M.P. Rout, S. Maguire, G. Blobel, and R.W. Wozniak. 1996. The yeast nucleoporin Nup188p interacts genetically and physically with the core structures of the nuclear pore complex. Journal of Cell Biology. 133:1153-1162.
32 Radu, A., G. Blobel, and R.W. Wozniak. 1993. Nup155 Is A Novel Nuclear-Pore Complex Protein That Contains Neither Repetitive Sequence Motifs Nor Reacts With Wga. Journal Of Cell Biology. 121:1-9.
33 Radu, A., G. Blobel, and R.W. Wozniak. 1994. Nup107 Is A Novel Nuclear-Pore Complex Protein That Contains A Leucine-Zipper. Journal Of Biological Chemistry. 269:17600-17605.
34 Rout, M.P., J.D. Aitchison, A. Suprapto, K. Hjertaas, Y.M. Zhao, and B.T. Chait. 2000. The yeast nuclear pore complex: Composition, architecture, and transport mechanism. Journal of Cell Biology. 148:635-651.
35 Siniossoglou, S., C. Wimmer, M. Rieger, V. Doye, H. Tekotte, C. Weise, S. Emig, A. Segref, and E.C. Hurt. 1996. A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. Cell. 84:265-275.
36 Smitherman, M., K. Lee, J. Swanger, R. Kapur, and B.E. Clurman. 2000. Characterization and targeted disruption of murine Nup50, a p27(Kip1)-interacting component of the nuclear pore complex. Molecular And Cellular Biology. 20:5631-5642.
37 Stavru, F., B.B. Hulsmann, A. Spang, E. Hartmann, V.C. Cordes, and D. Gorlich. 2006. NDC1: a crucial membrane-integral nucleoporin of metazoan nuclear pore complexes. Journal Of Cell Biology. 173:509-519.
38 Strahm, Y., B. Fahrenkrog, D. Zenklusen, E. Rychner, J. Kantor, M. Rosbash, and F. Stutz. 1999. The RNA export factor Gle1p is located on

the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr255p. Embo Journal. 18:5761-5777.

39 Strambio-de-Castillia, C., G. Blobel, and M.P. Rout. 1999. Proteins connecting the nuclear pore complex with the nuclear interior. Journal of Cell Biology. 144:839-855.

40 Sukegawa, J., and G. Blobel. 1993. A Nuclear-Pore Complex Protein That Contains Zinc Finger Motifs, Binds Dna, And Faces The Nucleoplasm. Cell. 72:29-38.

41 Vasu, S., S. Shah, A. Orjalo, M. Park, W.H. Fischer, and D.J. Forbes. 2001. Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. Journal Of Cell Biology. 155:339-353.

42 Wente, S.R., M.P. Rout, and G. Blobel. 1992. A New Family of Yeast Nuclear-Pore Complex Proteins. Journal of Cell Biology. 119:705-723.

43 Wozniak, R.W., G. Blobel, and M.P. Rout. 1994. Pom152 Is An Integral Protein Of The Pore Membrane Domain Of The Yeast Nuclear-Envelope. Journal Of Cell Biology. 125:31-42.

44 Wu, J., M.J. Matunis, D. Kraemer, G. Blobel, and E. Coutavas. 1995. Nup358, A Cytoplasmically Exposed Nucleoporin With Peptide Repeats, Ran-Gtp Binding-Sites, Zinc Fingers, A Cyclophilin-A Homologous Domain, And A Leucine-Rich Region. Journal Of Biological Chemistry. 270:14209-14213.

APPENDIX B — THE ISOLATION OF THE *T. BRUCEI* NUCLEUS AND SUBNUCLEAR COMPARTMENTS: THE *T. BRUCEI* NUCLEAR ENVELOPE PREPARATION (TBNEP)

Previous work done by M.P. Rout and M.C. Field established a reproducible method for the isolation and purification of whole nuclei and its subnuclear components (Rout and Field, 2001). The first step to producing subnuclear fractions is the isolation of nuclei away from the remainder of the cellular compartments (Figure 33). Once accomplished, these enriched nuclei may be further subfractionated to yield nucleoli, nuclear envelopes, or lipid-stripped nuclear envelopes (termed pore complex-lamina fraction, PCLF). The subnuclear components are of high quality and suitable for further biochemistry and mass spectrometry. Nuclei from either the vector (procyclic) or the host (blood stream form) life stage may be isolated, providing access to life-stage dependent aspects. However, the procyclic stage is somewhat more convenient as these cells can be grown to higher density in in vitro culture. We modified the protocol for use with this work and details are presented below. TbNEP

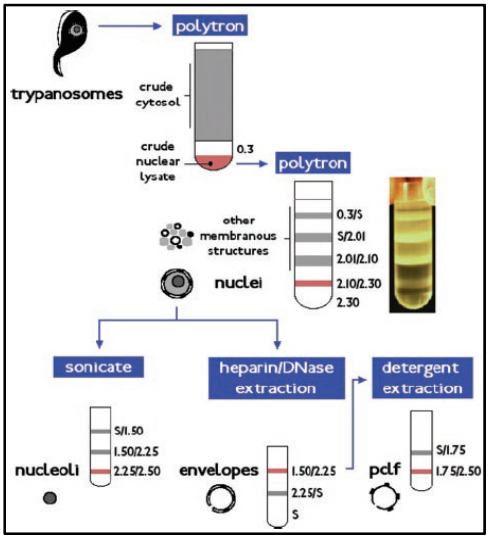


Figure 33: Flow diagram tracing the isolation and purification of nuclear and subnuclear structures by a discontinuous sucrose step gradient. Cellular material settles at the interface between sucrose densities. Fractions of interest are in red (Rout and Field, 2001).

material originating from the original protocol as well as material produce from the protocol below was used for the experiments outlined in this thesis.

MATERIALS

The Isolation of *Trypanosoma brucei* Nuclei

1. PVP solution: 8% polyvinylpyrrolidone (PVP-40, Sigma-Aldrich, St. Louis, MO), 11.5 mM KH_2PO_4, 8.5 mM K_2HPO_4, and 750 µM $MgCl_2 \cdot 6H_2O$. Adjust to pH 6.53 with concentrated H_3PO_4 (~15 µl for a 1 L solution). Store at 4°C. To minimize the risk of contamination when preparing reagents, it is imperative to use the highest reagent quality available. With the exception of polymeric, high density, acidic or basic solutions, aqueous buffers should be filtered with a 0.22 µm pore syringe or bottle filter.

2. Sucrose solutions (sucrose/PVP): In a large dish of hot water atop a stirring hot plate, the appropriate amount of sucrose is dissolved into PVP solution in a 500 ml beaker by constant stirring. Once the sucrose has completely dissolved, remove from heat, cover and allow to cool to room temperature. While continuously stirring the solution, adjust the refractive index (RI, to within 0.0003) by slowly adding PVP solution. Store in sterile tubes at -20°C.

    a. 2.01M: To 183.3 g sucrose, add PVP solution to a final weight of 338 g. Refractive Index (RI) = 1.4370.

b. 2.10M: To 193 g sucrose, add PVP solution to a final weight of 340 g. RI = 1.4420.

c. 2.30M: To 216 g sucrose, add PVP solution to a final weight of 340 g. RI = 1.4540.

3. Phosphate buffered saline (PBS Tablets, Sigma-Aldrich, St. Louis, MO). Chilled to 4°C.

4. 1 M dithiothreitol (DTT, Sigma-Aldrich, St. Louis, MO). Store at -20°C in 200 µl aliquots.

5. 10% Triton® X-100 (Sigma-Aldrich, St. Louis, MO).

6. Protease inhibitor cocktail P-8340 (PIC, Sigma-Aldrich, St. Louis, MO). Store at -20°C.

7. Solution P: Dissolve 0.04% (w/v) pepstatin A (Sigma-Aldrich, St. Louis, MO) and 1.8% PMSF (Sigma-Aldrich, St. Louis, MO) in absolute (anhydrous) ethanol. Store at -20°C. For best results, slowly dissolve each peptide inhibitor sequentially into room temperature ethanol.

8. 0.3 M sucrose/PVP. Dilute stock sucrose/PVP with PVP solution.

9. Lysis buffer (prepare fresh): Dissolve 0.05% Triton® X-100, 5 mM DTT, 1:100 solution P and 1:200 PIC into PVP solution. 20 ml of lysis buffer is equivalent to 1 volume. Protease inhibitors become

unstable at room temperature at working concentrations. On the bench, solution P must be kept on ice while the PIC may be left thawed and add the protease inhibitors to the working solutions at the last possible moment.

10. Underlay buffer (prepare fresh): Dissolve 5 mM DTT, 1:100 solution P and 1:200 PIC into 0.3 M sucrose/PVP. 10 ml of underlay buffer is equivalent to 1 volume.

11. Resuspension buffer (prepare fresh): Dissolve 5 mM DTT, 1:100 solution P and 1:200 PIC into 2.1 M sucrose/PVP. 8 ml of resuspension buffer is equivalent to 1 volume.

Subnuclear Fractionation

The Nuclear Envelope

1. 0.1 M bis-Tris-Cl, pH 6.50.

2. BT/Mg buffer: 0.01 M bis-Tris-Cl, pH 6.50, 0.1 mM $MgCl_2$.

3. Shearing buffer (prepare fresh): Dissolve 1 mM DTT, 1.0 mg/ml heparin, 20 µg/ml DNase I (Sigma-Aldrich, St. Louis, MO, store stock at -20°C), 2 µg/ml RNase A (Sigma-Aldrich, St. Louis, MO, store stock at -20°C), 1:100 solution P, and 1:200 PIC into BT/Mg buffer.

4. 2.10 M sucrose in 20% Accudenz® (Accurate Chemical & Scientific Corporation, Westbury, NY) in BT/Mg buffer. Store at -20°C.

5. 2.50 M sucrose in BT/Mg buffer. The final refractive index should be 1.4533. All sucrose BT/Mg solutions should be stored at -20°C.

6. 2.25 M sucrose BT/Mg, by stock dilution.

7. 1.50 M sucrose BT/Mg, by stock dilution.

The Pore Complex-lamina

1. Extraction buffer (prepare fresh): Dissolve 1.5% Triton® X-100, 1.5% sodium taurodeoxycholate, 1:100 solution P, and 1:200 PIC into BT/Mg buffer.

2. 2.50 M sucrose BT/Mg.

3. 1.75 M sucrose BT/Mg, by stock dilution.

METHODS

Either blood stream form (BSF) or procyclic life stage may be used with the following protocol, with similar yields. Procyclic cells are generally easier to culture since they do not require infection of animals to achieve the requisite number of cells for the isolation. At least 4 x 10^{10} cells are needed, which allows for two separate gradients with 2 x 10^{10} cells in each. One must be careful not to exceed 2 x 10^{10} cells in each gradient to maximize efficiency. Unless otherwise noted, cells and lysates must be kept on ice and pelleted in a refrigerated centrifuge at 4°C.

The Isolation of *Trypanosoma brucei* Nuclei

1. Gently pellet the cells at 1700 (g) for 10 minutes. Discard the supernatant and resuspend the pellet with roughly 25 ml of pre-chilled PBS. After a second centrifugation, resuspend the pellet in 25 ml of pre-chilled PBS and transfer the cells to a Sorvall HB-4 tube. Pellet by centrifugation once again (1790 (g), 15min, Sorvall HB-4 rotor) and discard the supernatant.

2. To lyse the cells, add 1 volume per 2 x 10^{10} cells of lysis buffer to the pellet and immediately disrupt the cells with a Polytron® (PTA-10, GlenMills, Clifton, NJ) homogenizer with 1 min bursts. As the appropriate speed setting may vary between homogenizer models, start with setting #4 and increase in increments of 0.5 until cell lysis is achieved. The lysis should be conducted in a cold room to keep

the homogenizer probe and cellular material cooled. Five minutes total homogenization time at the final setting is usually sufficient for an acceptable 70-90% cell lysis, with progress being monitored by phase contrast microscopy.

3. Once acceptable cell lysis has been achieved, underlay the equivalent of 2 x 10^{10} cells with 1 volume of underlay buffer and centrifuge for 20 minutes at 10,515 (g) in a Sorvall HB-4 rotor. Decant the supernatant (which contains the crude cytosol) and store at -80 °C.

4. The pellet should then be immediately resuspended by homogenization. Add 1 volume of resuspension buffer and homogenize with the Polytron® (setting 4.5-5) in 1 minute bursts. Monitor the progress with phase-contrast light microscopy; all cells should now be lysed (a significant proportion of the total cell lysis can actually occur at this stage), and the nuclei will be visible in the field as many small gray spheres and ovoids. Usually, 4 minutes is sufficient to achieve full dispersion.

5. Prepare the gradient. Into a Beckman SW-28 centrifuge tube, add the following: 8 ml 2.30 M sucrose/PVP, 8 ml 2.10 M sucrose/PVP, 8 ml 2.01 M sucrose/PVP. To reduce the viscosity of the sucrose solutions, allow the solutions to completely warm to room temperature. Wide-bore pipets and pipette tips (made by cutting off

~3 mm from the point of the tip) also facilitate sucrose solution handling. Before use, all sucrose solutions should be supplemented with solution P (1:100) and PIC (1:200).

6. Carefully, add the crude nuclear material on top of the gradient (the portion of the gradient which contains the crude material is designated as “S”). Afterwards, fill to within 5 mm of the brim with PVP solution to prevent collapse. In a Beckman ultracentrifuge and SW-28 rotor, spin the gradient at 141,245 (g) for 3 hours.

7. Sub-cellular material may be found at the interfaces. Each interface (PVP/S, S/2.01, 2.01/2.1, 2.1/2.3) should be collected and stored at -80°C for possible future study. Most of the nuclei settle at the 2.10/2.30 interface. The quality of the nuclei can be checked by phase-contrast light microscopy. Unloading the gradient will be easier if the interfaces are marked with a permanent marker before centrifugation. Collect the topmost fill layer and halfway through the first sucrose layer. Then, starting from the top, collect from halfway through the upper sucrose layer, through the interface, and continue to collect until halfway through the lower sucrose layer. The material at the interface may be gently dislodged with a pipette tip, if necessary.

8. The concentration of the nuclei is measured by optical density. 1 OD_{260} is equivalent to about 10^8 nuclei. Add 10 µl of nuclei to 1 ml

of 1% SDS. Measure the absorbance at 260 nm against a blank of 1% SDS. Multiply the value by 100x to obtain the OD.

Subnuclear Fractionation

Nuclear Envelope Isolation

1. To a measured volume of 300 ODs of purified nuclei, add the equivalent of 0.2 volumes of PVP solution and vortex for 1-2 minutes until the solution is homogenous.

2. In a Beckman Type 50.2Ti rotor, pellet the nuclei by centrifugation at 193,185 (g) for 1 hour. Decant the supernatant.

3. Resuspend the pellet in 3 ml of shearing buffer and shear the nuclear envelopes by vigorous vortexing for 1 full minute after the last traces of the pellet disappear.

4. After shearing, let the tube stand for 5 minutes at room temperature.

5. Add 10 ml of 2.10 M sucrose in 20% Accudenz® in BT/Mg buffer and mix well by vortexing.

6. Transfer mixture to an SW-28 centrifuge tube and overlay with 12 ml 2.25 M sucrose in BT/Mg and 10 ml of 1.50 M sucrose in BT/Mg. Top with BT/Mg to within 5 mm of the brim.

7. Spin the gradient at 141,245 (g) for 4 hours.

8. Collect all interfaces. The nuclear envelopes float up to the 1.50 M - 2.25 M interface. The quality of the nuclear envelopes can be checked by microscopy. The envelopes appear as faint “C” structures by phase contrast light microscopy.

Nuclear Pore Complex-lamina

1. To 1 volume of nuclear envelopes add the equivalent of 2 volumes of extraction buffer, and vortex for 5 min at room temperature. Allow the mixture to then incubate at room temperature for 25 minutes.

2. Prepare the gradient. In a Beckman SW-55 centrifuge tube, add 1 ml of 2.50 M sucrose in BT/Mg and then 1 ml of 1.75 M sucrose in BT/Mg.

3. Carefully overlay the extracted nuclear envelope mixture to the top of the gradient. Spin the gradient at 240,000 (g) for 30 minutes in a SW55Ti rotor.

4. Collect each interface fraction. The pore complex-lamina settles at the 1.75 M - 2.50 M interface.

APPENDIX C — THE TBNEP DATASET

The quality of the open reading frame identification may be assessed with the "protein expect" (the log of the expect value) and the total number of identified peptides. Table 4 lists the putative TbNPC associated proteins the we identified in this study. Proteins included in Table 4 with a log expect value worse than 10^{-5} have been validated by hand with the original spectra. The *T.cruzi* and *L. major* orthologs of the TbNPC associated proteins are listed in Table 5. As expected, the three kinetoplastids have remarkably similar NPC components due to their close evolutionary relationship. In Tables 6 and 7, only those identifications that meet the following criteria are included:

1. If the expect value is worse than 10^{-3}, then at least one peptide must have an expect value better than 10^{-2}.

2. If an identification is based on one peptide, that expect value must be better than 10^{-3}.

Notes for Tables:

a) As determined with Nucleo: * probability ≥ 0.70, ** probability ≥ 0.80, *** probability ≥ 0.90.

b) TMH and signal peptides were simultaneously predicted with Phobius.

c) As predicted with PEPCOIL, window=28: * probability ≥ 0.70, ** probability ≥ 0.80, *** probability ≥ 0.90.

d) HMMer against the Pfam profile databases, expect < 0.1

e) Present in T. brucei bloodstream plasma membrane fraction (D. Bridges, H. de Koning, P. Voorheis and R. Burchmore)

f) Present in T. brucei bloodstream cytoskeletal fraction (D. Bridges, H. de Koning, P. Voorheis and R. Burchmore)

g) Present in *T. brucei* flagellum proteome (Broadhead R, Dawe HR, Farr H, Griffiths S, Hart SR, Portman N, Shaw MK, Ginger ML, Gaskell SJ, McKean PG, Gull K)

h) Expressed in procyclic form (Jones A, Faldas A, Foucher A, Hunt E, Tait A, Wastling JM, Turner CM.)

i) The residue boundaries of the domains are listed along with the domain identifier: CC, coiled coil; FG, FG-repeat, the plurality motif is listed within parans.

Table 5: Putative TbNPC Associated proteins.

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	Category	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	Domains (i)	GFP localized?
Tb09.160.0340	TbMlp-2	92.3	-2.2	3	Mlp				***		Homologous to Mlp	CC: 88-200, 206-283, 294-368, 416-596	SPB during anaphase
Tb11.03.0810	TbMlp-1	109.6	-23.8	13	Mlp				***		(g), homologous to Mlp	CC: 292-336, 383-426, 436-496, 638-671, 689-748, 852-881, 884-974	Yes
Tb10.61.2630	TbSec13	41.6	-14.5	4	Nup	*				WD domain, G-beta repeat	(f,g,h)	Beta Propeller	Not Tagged
Tb10.61.0160	TbNup43	42.9	-4.1	2	Nup					WD domain, G-beta repeat	May be moved to unanno. Review and reanalyze	Beta Propeller	Inconclusive
Tb11.02.2120	TbNup48	48.4	-15.9	5	Nup						Homologous to Aladin / AAA	Beta Propeller	Yes
Tb927.4.3850	TbSeh1/Sec13	49.1	-2.1	2	Nup					WD domain, G-beta repeat		Beta Propeller	Not Tagged
Tb09.211.4780	TbNup82	82.3	-35.0	16	Nup						(e,f,g)	Alpha Solenoid	Not Tagged
Tb11.02.0460	TbNup89	89	-52.8	19	Nup					Nup84_Nup100	(e,f,h), homologous to ScNup84, VtNup84/107	Alpha Solenoid	Yes
Tb10.6k15.3670	TbNup96	96.4	-74.6	23	Nup					Nucleoporin interacting component	(e,f), homologous to ScNIC96, VtNup93	Alpha Solenoid	Yes
Tb11.01.7630	TbNup109	108.6	-21.9	9	Nup						(f,g)	Beta Propeller Alpha Solenoid	Yes
Tb927.7.2300	TbNup132	132.2	-30.8	14	Nup						(e,f,g)	Beta Propeller Alpha Solenoid	Yes
Tb10.6k15.2350	TbNup144	144.2	-70.9	27	Nup						(e,f,g,h), homologous to ScNup170, VtNup155	Beta Propeller Alpha Solenoid	Yes
Tb10.6k15.1530	TbNup181	181.4	-15.7	7	Nup						(h)	Alpha Solenoid	Not Tagged
Tb927.4.2880	TbNup225	225.4	-35.9	20	Nup			4			(e,f)	Alpha Solenoid	Not Tagged
Tb11.01.7200	TbNup53a	52.7	-27.6	8	Nup FG	**					Homologous to ScNup42, VtNup62	FG (GFG): 16-263	Not Tagged
Tb927.3.3540	TbNup53b	52.8	-36.0	9	Nup FG						(g)	FG (GFG): 10-72	Not Tagged
Tb11.02.0270	TbNup59	58.7	-24.3	6	Nup FG		Yes					FG (FGFG): 194-299	Not Tagged
Tb927.4.5200	TbNup62	62.4	-26.0	9	Nup FG				***		Homologous to ScNup57, VtNup54	FG (GGFGA): 8-349; CC: 453-486, 493-521	Not Tagged
Tb927.4.4310	TbNup64	64.1	-52.6	13	Nup FG				***			CC: 149-228; FG (FSFG): 331-583	Not Tagged
Tb927.8.8050	TbNup75	74.7	-3.2	2	Nup FG	*			***			CC: 150-237; FG (FSFG): 317-684	Yes
Tb927.3.3180	TbNup98	98	-129.9	20	Nup FG	*					(e)	FG (FSFG): 321-986	Not Tagged
Tb11.01.2885	TbNup140	140.2	-20.2	9	Nup FG						(f)	FG ([A/V]FGQ): 209-1432	Yes
Tb11.01.2880	TbNup149	149.1	-2.9	2	Nup FG							FG (VFGT): 267-388, 1007-1288	Yes
Tb11.03.0140	TbNup158	158.2	-99.7	33	Nup FG	*					(e), homologous to ScNup145, VtNup98-96	FG (GGFGQ): 5-550; Beta Sandwich: 713-851; Alpha Solenoid	Yes

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	Category	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	Domains (i)	GFP localized?
Tb927.7.5760	TbNTF2	15.8	-2.7	3	Transportin					Nuclear transport factor 2 (NTF2) domain	(h), ScNTF2, HsNTF2		Not Tagged
Tb927.3.1120	TbRTB2	24.3	-109.9	23	Transportin					Ras family; GTP-binding protein, RTB2 better alignment to GSP1/2	(e,f,h), ScGSP1/2, HsRAN(CRA_a)		Not Tagged
Tb09.160.2360	TbGLE2	38.3	-8.4	4	Transportin					WD domain, G-beta repeat	(e,f,g)	Beta Propeller	Not Tagged
Tb11.22.0004	TbMex67	56.1	-3.3	1	Transportin						(e), ScMex67, HsNXF2 (does not rBLAST)		Not Tagged
Tb927.6.2640	TbKap60	58	-18.6	6	Transportin					Armadillo/beta-catenin-like repeat, HEAT repeat (expect = 0.32)	ScKap60, HsKapA		Not Tagged
Tb10.70.4720	TbKap95	95	-8.5	4	Transportin			1		HEAT repeat (expect = 0.15)	ScKap95, HsKapB		Not Tagged
Tb10.6k15.3020	TbKap104	103.8	-2.5	2	Transportin					HEAT repeat (expect = 0.4)	(h), ScKap104 (2nd hit on rBLAST), HsRanBP6 aligns better than HsTransportin-1/2	transportin2 - like	Not Tagged
Tb11.01.7010	TbKap123	117.8	-16.6	4	Transportin				***	HEAT repeat	ScKap123, HsIMB3		Not Tagged

Table 6: *T. cruzi* and *L. major* orthologs to TbNPC associated proteins.

Domains	*T. brucei*		*L. major*	*T. cruzi*
β Propeller	TbSec13	Tb10.61.2630	LmjF32.0050	Tc00.1047053506525.20
	TbSeh1/Sec13	Tb927.4.3850	LmjF31.0200	Tc00.1047053507583.40
	TbNup43	Tb10.61.0160	LmjF19.1550	Tc00.1047053507211.30
	TbNup48	Tb11.02.2120	LmjF13.0250	Tc00.1047053506227.220
α Solenoid	TbNup89	Tb11.02.0460	LmjF33.2040	Tc00.1047053507771.90
	TbNup82	Tb09.211.4780	LmjF35.1600	Tc00.1047053510759.180
	TbNup96	Tb10.6k15.3670	LmjF36.2510	Tc00.1047053510181.50
	TbNup181	Tb10.6k15.1530	LmjF36.5890	Tc00.1047053509791.160
	TbNup225	Tb927.4.2880	LmjF34.1360	Tc00.1047053504253.10
β Propeller - α Solenoid	TbNup132	Tb927.7.2300	LmjF22.0380	Tc00.1047053511809.60
	TbNup144	Tb10.6k15.2350	LmjF36.6890	Tc00.1047053506247.70
	TbNup109	Tb11.01.7630	LmjF32.2780	Tc00.1047053507037.60
FG Repeat; Nup98; α Solenoid	TbNup158	Tb11.03.0140	LmjF27.0380	Tc00.1047053506925.440
FG Repeat	TbNup53a	Tb11.01.7200	LmjF32.2350	Tc00.1047053504159.10
	TbNup53b	Tb927.3.3540	LmjF29.0800	Tc00.1047053506591.69
	TbNup59	Tb11.02.0270	LmjF33.2370	Tc00.1047053504153.330
	TbNup62	Tb927.4.5200	LmjF31.2900	Tc00.1047053506419.60
	TbNup64	Tb927.4.4310		Tc00.1047053504717.20
	TbNup75	Tb927.8.8050		Tc00.1047053504411.10
	TbNup98	Tb927.3.3180		Tc00.1047053508153.410
	TbNup140	Tb11.01.2885	LmjF28.3010	Tc00.1047053511671.60
	TbNup149	Tb11.01.2880	LmjF28.3030	Tc00.1047053511671.50
Coiled coil	TbMlp1	Tb11.03.0810	LmjF25.0270	Tc00.1047053504109.20
	TbMlp2	Tb09.160.0340	LmjF26.2660	Tc00.1047053504769.80
Transportins	TbGle2/Rae1	Tb09.160.2360	LmjF01.0320	Tc00.1047053511577.130
	TbNTF2	Tb927.7.5760	LmjF06.1170	Tc00.1047053508173.180
	TbRTB2	Tb927.3.1120	LmjF25.1420	Tc00.1047053503539.30
	TbMex67	Tb11.22.0004	LmjF27.1690	Tc00.1047053508271.4
	TbKap60	Tb927.6.2640	LmjF30.1120	Tc00.1047053509057.20
	TbKap95	Tb10.70.4720	LmjF34.0490	Tc00.1047053504105.150
	TbKap104	Tb10.6k15.3020	LmjF36.2720	Tc00.1047053509717.70
	TbKap123	Tb11.01.7010	LmjF32.2150	Tc00.1047053511707.39

Table 7: TbNEP members which were not annotated with the *T. brucei* sequencing effort.

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb927.4.590	Conserved hypothetical	87.9	-64.9	21		Yes	1		DUF1620	(e,h), GFP labeling shows a speckled cytosolic pattern	Inconclusive
Tb927.7.3330	Conserved hypothetical	502.3	-15.2	7				***		(f), paralogous to NUP-1	Inconclusive
Tb10.70.1120	Conserved hypothetical	109.2	-1.1	2	*				Zn-finger in Ran binding protein	Zinc finger, RanBP2-type	Inconclusive
Tb927.4.1310	Conserved hypothetical	47.3	-1.1	2					Zn-finger in Ran binding protein	ZFP family member	Inconclusive
Tb927.6.1830	Conserved hypothetical	49.9	-1.1	2	*				WD domain, G-beta repeat		Inconclusive
Tb10.70.1110	Conserved hypothetical	59.9	nf		**				Zn-finger in Ran binding protein	Zinc finger, RanBP2-type	Inconclusive
Tb10.70.1130	Conserved hypothetical	48.2	nf					***			Inconclusive
Tb927.3.5370	Conserved hypothetical	34.2	-71.7	10	***					(e,f)	
Tb927.7.6670	Conserved hypothetical	332.6	-71.5	27	**	Yes	3	***		(f), large beta sheet presence, however it is very large protein	
Tb927.8.2460	Conserved hypothetical	30.8	-62.7	14		Yes	4				
Tb10.6k15.0690	Conserved hypothetical	152.8	-55.7	24						(e,f,g), possible cysteine protease	
Tb927.4.2850	Conserved hypothetical	72.5	-51.4	15	**					(e)	
Tb09.160.4710	Conserved hypothetical	170.2	-50.1	13				***		(f)	
Tb927.5.2930	Conserved hypothetical	43.2	-48.3	11						(e,f,g)	
Tb927.5.4020	Hypothetical protein	29.3	-46.2	7		Yes		***			
Tb11.02.4120	Conserved hypothetical	27.5	-45.2	9		Yes				(e,f,g)	
Tb927.2.2510	Conserved hypothetical	29.1	-44.3	13						(e,f)	
Tb927.7.5940	Conserved hypothetical	67.1	-43.7	11			14	***	Major Facilitator Superfamily		
Tb11.01.4740	Conserved hypothetical	61.1	-42.9	12		Yes				(e)	
Tb11.02.5660	Conserved hypothetical	46.2	-42.9	8			1	***		(e,h)	
Tb09.211.1240	Conserved hypothetical	38.2	-42.7	10						(e), probable ribosome-binding factor A	
Tb10.6k15.2510	Conserved hypothetical	28.8	-40.8	11						(e,h)	
Tb927.5.440	Conserved hypothetical	84.5	-40.5	12		Yes	1			(h). Probable sialidase	
Tb927.7.900	Conserved hypothetical	64.7	-39.0	13		Yes	1			(e), nicalin-like	
Tb10.6k15.1500	Conserved hypothetical	45.3	-33.6	7						(e,f,g)	
Tb927.7.2240	Conserved hypothetical	59.6	-31.4	11							
Tb11.01.0480	Conserved hypothetical	63.5	-31.0	7			9		EF hand, SNARE associated Golgi protein	Vacuole membrane protein 1 (mouse)-like	
Tb11.0400	Hypothetical protein	27.6	-29.9	11			1				
Tb09.211.1800	Conserved hypothetical	53.6	-29.2	7	**				LisH	(e,f), Nucleolar -- Confirmed by mAb	
Tb927.6.4320	Conserved hypothetical	44.3	-28.9	7			1		Eukaryotic translation initiation factor 3 subunit 7	(e)	
Tb927.8.1570	Conserved hypothetical	28.8	-28.6	9					UPF0172	Probable COX4	
Tb927.3.5620	Conserved hypothetical	112.5	-27.9	13				***	SPT16, Rtt106	(e,f), Probable transcription factor (FACT complex)	
Tb927.3.5350	Conserved hypothetical	11.5	-26.9	5			1			(e)	
Tb09.211.2530	Conserved hypothetical	41.5	-26.8	5			1			(e)	
Tb10.70.3750	Conserved hypothetical	31.5	-26.6	6			7			(e,f)	
Tb927.1.860	Conserved hypothetical	29.5	-26.6	8		Yes	1				
Tb927.3.3890	Conserved hypothetical	36.6	-26.0	5							
Tb927.4.2530	Conserved hypothetical	16.8	-26.0	7			3			(e)	
Tb927.5.1780	Conserved hypothetical	48.1	-25.6	5						(e)	
Tb927.7.6260	Conserved hypothetical	36.4	-25.4	5				***	Tetratricopeptide repeat	(e), Tetratricopeptide repeat protein	
Tb10.61.0540	Conserved hypothetical	36.3	-23.8	5	**			***		(e,f,g)	
Tb927.5.2530	Conserved hypothetical	114.8	-23.1	13					DUF699, DUF1726	(g,h)	

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb927.7.2190	Conserved hypothetical	29.2	-22.5	6		Yes	1			(e), SSR Alpha	
Tb927.6.4500	Conserved hypothetical	22.6	-22.2	8		Yes	1			(e,f)	
Tb927.4.1300	Conserved hypothetical	41.9	-22.1	7						(e,f,g,h)	
Tb927.8.1740	Conserved hypothetical	62.1	-20.9	9						(e)	
Tb10.70.7770	Conserved hypothetical	37.0	-20.2	6			1			(e)	
Tb10.70.2320	Conserved hypothetical	25.9	-20.0	5		Yes					
Tb10.6k15.1800	Conserved hypothetical	60.5	-19.2	3			3		ABC1 family		
Tb10.70.3010	Conserved hypothetical	14.7	-19.2	8		Yes	1				
Tb11.02.4620	Conserved hypothetical	88.9	-17.6	7	*				BOP1NT, WD domain, G-beta repeat	(g), BOP1-like WD-40, Erb1p homolog	
Tb10.70.1570	Conserved hypothetical	107.2	-17.5	10				***		(e,f)	
Tb927.5.1930	Conserved hypothetical	24.1	-17.4	4		Yes				(e)	
Tb927.5.3190	Conserved hypothetical	23.1	-17.4	5	**						
Tb927.3.4100	Conserved hypothetical	63.2	-17.3	7		Yes	13		Major Facilitator Superfamily	(e)	
Tb927.3.4920	Conserved hypothetical	42.7	-16.8	3					LETM1-like protein	(h), mitochondrial	
Tb10.70.2450	Conserved hypothetical	27.1	-15.9	5			3		DUF850	(e)	
Tb927.7.4760	Conserved hypothetical	58.7	-15.9	5			2				
Tb11.02.3310	Conserved hypothetical	13.5	-15.7	4			1			(e)	
Tb11.01.3860	Conserved hypothetical	18.3	-15.6	5		Yes					
Tb927.7.5840	Conserved hypothetical	26.1	-15.6	4			5	***			
Tb09.160.1070	Conserved hypothetical	41.4	-15.2	5		Yes			Methyltransferase domain		
Tb09.160.1560	Conserved hypothetical	287.0	-14.6	7			2		HEAT repeat (expect=0.48)	(h), BAP28	
Tb11.02.1320	Conserved hypothetical	39.0	-14.6	4							
Tb11.01.1090	Conserved hypothetical	23.7	-14.1	4							
Tb11.01.0490	Conserved hypothetical	60.0	-12.8	7	*			***	Sas10/Utp3/C1D family	Ribonucleoprotein, SAS10 homolog	
Tb11.01.1900	Conserved hypothetical	30.2	-12.8	6					Macro domain (ADP-ribose binding module)		
Tb11.02.2040	Conserved hypothetical	14.0	-12.3	2				***	Alba, DNA/RNA-binding protein	(h)	
Tb927.5.3250	Conserved hypothetical	172.0	-12.2	5							
Tb10.70.0870	Conserved hypothetical	127.9	-12.0	5					STAG domain	(e), possible member of cohesin complex	
Tb11.01.2430	Conserved hypothetical	450.0	-12.0	10				***			
Tb11.01.8770	Conserved hypothetical	110.0	-12.0	5					Leucine Rich Repeat	(e,f,g,h)	
Tb10.70.7760	Conserved hypothetical	46.7	-11.6	6			1				
Tb927.8.5640	Conserved hypothetical	37.0	-11.6	3							
Tb09.211.4520	Conserved hypothetical	36.7	-11.4	6							
Tb927.6.2320	Conserved hypothetical	20.1	-11.3	5		Yes	1				
Tb927.4.310	Conserved hypothetical	733.0	-10.9	10			2		SPRY domain, HECT domain	(e)	
Tb10.61.2050	Conserved hypothetical	44.1	-10.7	3			2	**		Probable DnaJ	
Tb11.02.1350	Conserved hypothetical	50.3	-10.6	5		Yes				(h)	
Tb11.01.8225	Conserved hypothetical	16.3	-10.5	3							
Tb927.7.7200	Conserved hypothetical	249.9	-10.4	8					DUF1309	(h); Possible FG - discounted after examination of amino acid composition	
Tb11.01.1540	Conserved hypothetical	118.7	-10.2	6						(h)	
Tb11.47.0022	Conserved hypothetical	20.2	-10.1	3			2				
Tb11.01.1830	Conserved hypothetical	25.3	-10.0	4					Bromodomain		
Tb11.02.5480	Conserved hypothetical	13.5	-10.0	5			4		DUF423		
Tb11.01.3290	Conserved hypothetical	68.9	-9.8	4				***		(e,h)	
Tb927.2.3800	Conserved hypothetical	55.4	-9.8	1		Yes					
Tb927.8.5580	Conserved hypothetical	526.0	-9.8	7				**			
Tb10.6k15.0500	Conserved hypothetical	33.8	-9.7	5					Tetratricopeptide repeat		
Tb11.01.5760	Conserved hypothetical	67.1	-9.7	4		Yes			Pyridine nucleotide-disulphide oxidoreductase		
Tb10.6k15.2320	Conserved hypothetical	30.8	-9.5	4							

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb11.02.0840	Conserved hypothetical	26.5	-9.4	3					RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)		
Tb927.8.3950	Conserved hypothetical	102.0	-9.4	5				***		(e), Probable Centrosomal Protein, CEP250	
Tb927.6.590	Conserved hypothetical	12.2	-9.1	1						(e)	
Tb927.7.4880	Conserved hypothetical	23.2	-9.0	3			6				
Tb927.8.3680	Conserved hypothetical	69.5	-9.0	1				***			
Tb11.02.0880	Conserved hypothetical	486.2	-8.9	8			2		SPRY domain, HECT domain		
Tb927.6.1920	Conserved hypothetical	42.3	-8.9	4	**			**		(e)	
Tb927.7.4580	Conserved hypothetical	60.5	-8.9	8	**				PPAK motif (double check)		
Tb927.2.5240	Conserved hypothetical	54.2	-8.8	3				***	WD domain, G-beta repeat	(h), Prp19/Pso4-like (spliceosomal)	
Tb927.6.4130	Conserved hypothetical	11.7	-8.7	2							
Tb927.4.2080	Conserved hypothetical	104.8	-8.5	3				***	C2 domain	(e,f,g,h), signal transduction	
Tb927.8.2470	Conserved hypothetical	11.2	-8.4	1							
Tb10.70.2970	Conserved hypothetical	24.0	-8.3	5							
Tb11.02.0800	Conserved hypothetical	166.1	-8.1	5			1	***		(e)	
Tb927.8.5560	Conserved hypothetical	59.9	-8.1	4			1			Spermidine synthase-like	
Tb927.3.1690	Conserved hypothetical	17.1	-8.0	3							
Tb927.2.3610	Conserved hypothetical	16.0	-7.9	2		Yes					
Tb927.7.840	Conserved hypothetical	14.5	-7.9	3			1				
Tb10.70.7440	Conserved hypothetical	69.0	-7.7	2	*			***			
Tb927.3.3130	Conserved hypothetical	177.9	-7.7	6		Yes		**	SAM domain (Sterile alpha motif)	(e)	
Tb927.7.6980	Conserved hypothetical	194.8	-7.7	5			1			Probable SAP 130 homolog	
Tb927.5.1630	Conserved hypothetical	17.6	-7.5	4			1		Cytochrome b5-like Heme/Steroid binding domain		
Tb927.8.940	Conserved hypothetical	237.3	-7.5	6			2			Large protein. 1 TMH on each terminus. Large amount of beta sheets. Possible candidate for a membrane nup	
Tb10.70.5220	Conserved hypothetical	64.2	-7.4	1			7		Cleft lip and palate transmembrane protein 1 (CLPTM1)	(e), Clptm1 homolog	
Tb11.01.8500	Conserved hypothetical	25.8	-7.4	2			1				
Tb927.4.3060	Conserved hypothetical	17.7	-7.4	3						(e,f)	
Tb927.4.860	Conserved hypothetical	33.1	-7.4	2			1				
Tb09.211.1390	Conserved hypothetical	79.8	-7.3	5		Yes		***	Hsp70 protein	(e)	
Tb11.03.0530	Conserved hypothetical	31.2	-7.3	1	***			***		(f,h)	
Tb927.4.1970	Conserved hypothetical	61.9	-7.2	3		Yes	4		DUF21, CBS	cyclin M2	
Tb11.02.2030	Conserved hypothetical	12.7	-7.1	1					Alba, DNA/RNA-binding protein	(h)	
Tb927.8.5520	Conserved hypothetical	19.5	-7.1	1							
Tb927.8.5760	Conserved hypothetical	53.7	-7.0	3			2		Ankyrin repeat	(e)	
Tb927.4.3450	Conserved hypothetical	13.7	-6.9	2			1				
Tb927.7.4220	Conserved hypothetical	120.9	-6.9	2					Dip2/Utp12 Family, WD domain, G-beta repeat	(e,f), Probable DIP2	
Tb11.47.0011	Conserved hypothetical	325.3	-6.8	6				***		(e)	
Tb10.70.1890	Conserved hypothetical	51.1	-6.7	1			1				
Tb11.01.3670	Conserved hypothetical	118.1	-6.7	7					HEAT repeat		
Tb10.6k15.1450	Conserved hypothetical	17.5	-6.6	3						(e,f)	
Tb09.160.0400	Conserved hypothetical	71.8	-6.5	6		Yes		***			
Tb10.61.2850	Conserved hypothetical	50.4	-6.5	2		Yes		***		(e)	
Tb11.47.0006	Conserved hypothetical	86.8	-6.5	5					DUF1126	(e,f,g)	
Tb927.1.2750	Conserved hypothetical	18.6	-6.5	3							
Tb927.7.2640	Conserved hypothetical	50.8	-6.5	5				***		(h)	
Tb927.3.4210	Conserved hypothetical	197.4	-6.4	6							
Tb927.4.720	Conserved hypothetical	19.6	-6.4	3			3				

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb927.7.5700	Conserved hypothetical	45.5	-6.4	2		Yes				(e)	
Tb927.8.1190	Conserved hypothetical	36.3	-6.4	2				***			
Tb927.8.4380	Conserved hypothetical	11.8	-6.4	1							
Tb927.8.4780	Conserved hypothetical	467.8	-6.4	7			1	***		(e,f)	
Tb10.70.6260	Conserved hypothetical	33.1	-6.3	4			1				
Tb11.01.1700	Conserved hypothetical	20.5	-6.3	1	*			***			
Tb927.8.7730	Conserved hypothetical	45.1	-6.3	4			7		Longevity-assurance protein (LAG1)		
Tb10.6k15.1040	Conserved hypothetical	57.4	-6.2	1		Yes			FAD dependent oxidoreductase	(e)	
Tb11.01.0240	Conserved hypothetical	42.7	-6.2	1		Yes					
Tb927.4.1610	Conserved hypothetical	39.6	-6.2	2						(e,f)	
Tb10.61.2900	Conserved hypothetical	23.5	-6.1	1					K+ channel tetramerisation domain		
Tb11.02.3860	Conserved hypothetical	70.7	-6.1	2			14		Major Facilitator Superfamily		
Tb927.7.1290	Conserved hypothetical	26.9	-6.1	4		Yes	1		DUF2012		
Tb11.02.4190	Conserved hypothetical	53.2	-6.0	2	***			***			
Tb09.160.1400	Conserved hypothetical	147.1	-5.9	6							
Tb11.01.2940	Conserved hypothetical	34.9	-5.9	1					WD domain, G-beta repeat	U5 snRNP-specific 40 kDa protein	
Tb11.02.4680	Conserved hypothetical	26.4	-5.8	4			1	***			
Tb11.03.0680	Conserved hypothetical	103.6	-5.8	3					WD domain, G-beta repeat	Utp13-like (ribonucleoprotein)	
Tb927.3.1080	Conserved hypothetical	28.4	-5.8	4			4		DUF1295		
Tb927.7.1260	Conserved hypothetical	26.0	-5.8	1			1				
Tb09.244.2650	Conserved hypothetical	99.9	-5.7	1					U3 small nucleolar RNA-associated protein 6		
Tb10.70.5420	Conserved hypothetical	52.1	-5.7	1						(h)	
Tb11.01.2020	Conserved hypothetical	51.8	-5.7	1							
Tb11.01.5120	Conserved hypothetical	105.3	-5.7	3			1			(e)	
Tb927.5.510	Conserved hypothetical	14.1	-5.7	1		Yes					
Tb09.211.1900	Conserved hypothetical	29.5	-5.6	1							
Tb10.70.1640	Conserved hypothetical	68.0	-5.6	5			4	***		(e)	
Tb11.01.1625	Conserved hypothetical	11.8	-5.6	1					EF hand (expect = 0.23)	(e,f,h)	
Tb11.01.2250	Conserved hypothetical	106.8	-5.6	3				***	WD domain, G-beta repeat	U3 snoRNP protein	
Tb927.3.2790	Conserved hypothetical	124.6	-5.6	4					Leucine Rich Repeat (double check)		
Tb10.26.1020	Conserved hypothetical	11.2	-5.5	2						(e)	
Tb10.389.1900	Conserved hypothetical	76.3	-5.5	5						(h)	
Tb10.70.0040	Conserved hypothetical	53.4	-5.5	3			1		DUF580		
Tb927.5.3090	Conserved hypothetical	11.6	-5.5	2							
Tb927.8.6110	Conserved hypothetical	54.6	-5.5	1					Hydroxymethylglutaryl-coenzyme A synthase N terminal	(h)	
Tb09.211.1220	Conserved hypothetical	38.8	-5.4	3			4		Tetratricopeptide repeat	(e)	
Tb10.61.2440	Conserved hypothetical	115.8	-5.3	4	***			***	Utp14 protein	Probable Ribonucleoprotein	
Tb927.4.1360	Conserved hypothetical	32.5	-5.3	3					Aldose 1-epimerase		
Tb927.8.5490	Conserved hypothetical	71.6	-5.3	1	**			**	Nucleolar protein,Nop52		
Tb09.211.1360	Conserved hypothetical	16.8	-5.2	2			4				
Tb927.7.4270	Conserved hypothetical	25.2	-5.2	1		Yes	1				
Tb11.02.4330	Conserved hypothetical	15.7	-5.1	2							
Tb09.160.0370	Conserved hypothetical	42.4	-4.9	1				***			
Tb11.01.2460	Conserved hypothetical	45.2	-4.8	2			1			(e)	
Tb927.3.3790	Conserved hypothetical	31.6	-4.8	2						(e,f), probable calcium binding	
Tb927.8.3870	Conserved hypothetical	127.6	-4.8	3	***			**		(e,f), Nucleolar -- Confirmed by mAb	
Tb927.3.1140	Conserved hypothetical	43.4	-4.7	3				*	Surp module	RNA binding	
Tb927.3.4160	Conserved hypothetical	35.6	-4.7	3					UBA/TS-N domain		
Tb927.4.2040	Conserved hypothetical	20.7	-4.7	3					Alba, DNA/RNA-binding protein	(f,h)	

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb927.5.3340	Conserved hypothetical	11.0	-4.7	1							
Tb927.6.2650	Conserved hypothetical	84.4	-4.7	3							
Tb09.211.2700	Conserved hypothetical	50.7	-4.6	2				***		(e,f,g)	
Tb10.70.5980	Conserved hypothetical	39.0	-4.6	1			1				
Tb11.01.5380	Conserved hypothetical	42.6	-4.6	4							
Tb927.1.2985	Conserved hypothetical	10.8	-4.6	1			2				
Tb927.1.4450	Conserved hypothetical	130.8	-4.6	4			8		Calcium-activated BK potassium channel alpha subunit		
Tb927.6.4440	Conserved hypothetical	37.6	-4.6	2						Possible RNA binding motif	
Tb927.8.4500	Conserved hypothetical	84.5	-4.6	4					MIF4G domain	probable eukaryotic initiation factor 4G	
Tb09.211.0580	Conserved hypothetical	251.2	-4.5	5				***		(h)	
Tb10.70.3190	Conserved hypothetical	127.0	-4.5	5		Yes			3-hydroxyacyl-CoA dehydrogenase, C-terminal domain		
Tb927.1.880	Conserved hypothetical	513.2	-4.5	3		Yes		***	Thrombospondin type 3 repeat, ATPase family associated with various cellular activities (AAA)		
Tb09.160.2370	Conserved hypothetical	23.9	-4.4	2			2			(e)	
Tb10.389.1790	Conserved hypothetical	23.9	-4.4	1						Probable RNA-binding Domain in PseudoUridine synthase	
Tb927.7.4470	Conserved hypothetical	22.3	-4.4	2			4			(e)	
Tb09.v1.0130	Conserved hypothetical	38.2	-4.3	2	**						
Tb927.6.2930	Conserved hypothetical	59.5	-4.3	1					ATPase family associated with various cellular activities (AAA)	(e)	
Tb927.7.2110	Conserved hypothetical	29.2	-4.3	1				***			
Tb927.8.1270	Conserved hypothetical	61.9	-4.3	2	**			***		(h), probable HSP20	
Tb927.8.560	Conserved hypothetical	64.5	-4.3	1			1		EF hand		
Tb11.02.5080	Conserved hypothetical	43.6	-4.2	1			1		DUF1692	(h)	
Tb927.8.5080	Conserved hypothetical	65.3	-4.2	3		Yes	1		Amidase		
Tb11.01.2740	Conserved hypothetical	71.7	-4.1	4		Yes		***			
Tb11.02.4900	Conserved hypothetical	17.0	-4.1	2		Yes			Protease associated domain		
Tb927.5.2330	Conserved hypothetical	484.6	-4.1	4	**			**	HEAT repeat (expect = 0.22)	(e,f)	
Tb927.7.2410	Conserved hypothetical	66.3	-4.1	2			2		MMR_HSR1	Sarcalumenin	
Tb927.8.6660	Conserved hypothetical	69.0	-4.1	3						(e,f,g)	
Tb10.406.0190	Conserved hypothetical	38.6	-4.0	1							
Tb10.70.4960	Conserved hypothetical	75.1	-4.0	4	***				WD domain, G-beta repeat	Similar to U3 small nucleolar RNA interacting protein 2	
Tb11.02.0210	Conserved hypothetical	50.8	-4.0	3				***		(e,f,g,h)	
Tb927.2.2950	Conserved hypothetical	107.8	-4.0	3	**	Yes		***			
Tb927.6.2140	Conserved hypothetical	28.7	-4.0	3		Yes					
Tb927.7.2200	Conserved hypothetical	8.8	-4.0	2					Tim10/DDP family zinc finger	Mitochondrial import inner membrane translocase subunit Tim10	
Tb10.70.4500	Conserved hypothetical	15.4	-3.9	1			1				
Tb927.7.700	Conserved hypothetical	74.2	-3.9	3				**		Probable Enp2p homolog	
Tb927.8.5300	Conserved hypothetical	63.8	-3.9	4				**			
Tb09.211.3140	Conserved hypothetical	109.7	-3.8	1			1			(e,f,h)	
Tb927.7.3550	Conserved hypothetical	138.0	-3.8	2	**				C2 domain	(e,f,g), involved with signal tranduction	
Tb10.389.0140	Conserved hypothetical	45.2	-3.7	2			9		CDP-alcohol phosphatidyltransferase	(e), CDP-alcohol phosphatidyltransferase	
Tb10.70.0860	Conserved hypothetical	18.5	-3.7	1							
Tb10.70.2380	Conserved hypothetical	38.3	-3.7	1						(e)	
Tb10.70.4000	Conserved hypothetical	68.6	-3.7	2		Yes	5				

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb11.01.5000	Conserved hypothetical	73.5	-3.7	3				**			
Tb11.02.2770	Conserved hypothetical	51.0	-3.7	1		Yes					
Tb927.6.3670	Conserved hypothetical	358.9	-3.7	3						(e)	
Tb927.6.4070	Conserved hypothetical	29.2	-3.7	1			2				
Tb927.8.6050	Conserved hypothetical	27.3	-3.7	3			1				
Tb11.01.3085	Conserved hypothetical	21.1	-3.6	1							
Tb11.01.5680	Conserved hypothetical	51.6	-3.6	2	**					Ptbp1-like	
Tb927.2.5820	Conserved hypothetical	117.4	-3.6	2		Yes					
Tb927.7.2170	Conserved hypothetical	168.5	-3.6	2	***			***		(e,h)	
Tb927.7.640	Conserved hypothetical	37.3	-3.6	1			3				
Tb927.8.3320	Conserved hypothetical	53.4	-3.6	2			2		IQ calmodulin-binding motif		
Tb10.70.3450	Conserved hypothetical	84.6	-3.5	1			1			acyltransferase like 2	
Tb11.01.8370	Conserved hypothetical	87.3	-3.5	4	**			***		(g)	
Tb11.02.2890	Conserved hypothetical	25.9	-3.5	3				**			
Tb927.2.2940	Conserved hypothetical	32.1	-3.5	1							
Tb927.8.1150	Conserved hypothetical	39.2	-3.5	2	*			***			
Tb09.211.1070	Conserved hypothetical	114.6	-3.4	4	**				Zinc finger C-x8-C-x5-C-x3-H type (and similar)	CDS weak match to TFIIA	
Tb09.211.4740	Conserved hypothetical	44.5	-3.4	1			1		Methyltransferase domain		
Tb11.01.4810	Conserved hypothetical	23.8	-3.4	3			1		MSP (Major sperm protein) domain	(e,f,g), vesicle-associated membrane protein	
Tb927.1.4410	Conserved hypothetical	83.0	-3.4	4	*						
Tb927.3.4380	Conserved hypothetical	53.5	-3.4	1							
Tb927.6.4140	Conserved hypothetical	13.0	-3.4	1		Yes				(e,f,g)	
Tb927.6.630	Conserved hypothetical	78.7	-3.4	3		Yes					
Tb11.01.2330	Conserved hypothetical	79.6	-3.3	3					MIF4G domain	Probable eukaryotic initiation factor 4G	
Tb11.01.7610	Conserved hypothetical	117.0	-3.3	3			2		Sec7 domain		
Tb927.7.2450	Conserved hypothetical	54.3	-3.3	1							
Tb927.8.1680	Conserved hypothetical	23.0	-3.3	1	**			***			
Tb09.211.1150	Conserved hypothetical	20.6	-3.2	2	*					(e)	
Tb10.26.0360	Conserved hypothetical	83.3	-3.2	2	**						
Tb11.01.8650	Conserved hypothetical	228.6	-3.2	4	*			***	IQ calmodulin-binding motif	(e)	
Tb11.02.0200	Conserved hypothetical	99.4	-3.2	4	*						
Tb927.2.5130	Conserved hypothetical	43.5	-3.2	3							
Tb927.3.4970	Conserved hypothetical	115.7	-3.2	2				**		(e,f,g)	
Tb927.4.1540	Conserved hypothetical	55.2	-3.2	2			1		Epimerase	(e), cinnamyl alcohol dehydrogenase	
Tb927.8.1500	Conserved hypothetical	63.1	-3.2	5	**						
Tb927.8.7280	Conserved hypothetical	27.4	-3.2	1		Yes	1			coated vesicle membrane protein	
Tb10.61.1970	Conserved hypothetical	37.6	-3.1	2	**					(e,h)	
Tb11.01.2650	Conserved hypothetical	52.6	-3.1	3		Yes					
Tb11.02.0520	Conserved hypothetical	74.8	-3.1	4			9			(h)	
Tb927.1.2190	Conserved hypothetical	187.8	-3.1	2	*				WD domain, G-beta repeat (double check)	(h), possible Mip1	
Tb927.2.5810	Conserved hypothetical	173.9	-3.1	3						(f,h)	
Tb927.5.2650	Conserved hypothetical	58.7	-3.1	1					FAD binding domain	(e,f,g,h)	
Tb927.7.3850	Conserved hypothetical	110.4	-3.1	3							
Tb927.7.7270	Conserved hypothetical	85.6	-3.1	3	**					(h)	
Tb927.8.2030	Conserved hypothetical	48.2	-3.1	1				***		(e,f,g,h)	
Tb09.160.1330	Conserved hypothetical	41.1	-3.0	2			2		ELMO/CED-12 family		
Tb09.211.1290	Conserved hypothetical	36.7	-3.0	2		Yes			Phosphoglycerate mutase family		

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb09.211.1890	Conserved hypothetical	14.6	-3.0	1			2			(h)	
Tb10.70.0640	Conserved hypothetical	54.8	-3.0	2		Yes	1				
Tb10.70.2155	Conserved hypothetical	8.6	-3.0	1							
Tb927.5.1010	Conserved hypothetical	146.7	-3.0	1	*			***		(h)	
Tb927.6.3680	Conserved hypothetical	27.8	-3.0	1			4				
Tb10.26.0680	Conserved hypothetical	14.3	-2.9	2						(e,f,g,h)	
Tb10.70.3460	Conserved hypothetical	236.9	-2.9	4				***	EF hand, WD domain, G-beta repeat	(e,f)	
Tb11.02.1290	Conserved hypothetical	32.3	-2.9	2							
Tb11.18.0012	Conserved hypothetical	122.8	-2.9	4	*				Cyclic nucleotide-binding domain	(e,f,g)	
Tb927.8.1050	Conserved hypothetical	292.3	-2.9	3	*			***	WD domain, G-beta repeat	(e,h)	
Tb09.v1.0420	Conserved hypothetical	22.9	-2.8	2			1				
Tb10.70.0625	Conserved hypothetical	12.2	-2.8	2			1				
Tb927.4.2340	Conserved hypothetical	58.8	-2.8	2						Possible PI kinase	
Tb927.8.1480	Conserved hypothetical	267.5	-2.8	4					WD domain, G-beta repeat		
Tb11.02.2450	Conserved hypothetical	44.0	-2.7	2		Yes	1				
Tb927.2.2520	Conserved hypothetical	29.2	-2.7	3						(e,f)	
Tb11.02.1880	Conserved hypothetical	21.0	-2.5	3			1				
Tb927.7.6090	Conserved hypothetical	88.1	-2.5	2				***		(h)	
Tb927.7.6590	Conserved hypothetical	97.5	-2.5	3		Yes			Apopolysialoglycoprotein (PSGP)		
Tb11.02.0445	Conserved hypothetical	13.4	-2.4	2		Yes	1				
Tb927.8.3210	Conserved hypothetical	18.7	-2.4	2						(h)	
Tb09.160.2400	Conserved hypothetical	66.9	-2.3	3				***		(f)	
Tb10.70.5410	Conserved hypothetical	41.1	-2.3	2					WD40 repeat (weak)	(f); No similarity to Nups. Sporatic interspersed alpha helcies in sec. struc. Prediction.	
Tb11.02.2870	Conserved hypothetical	66.6	-2.3	3				***			
Tb11.18.0014	Conserved hypothetical	88.7	-2.3	2					CBF/Mak21 family		
Tb927.1.670	Conserved hypothetical	65.7	-2.3	2							
Tb11.03.0900	Conserved hypothetical	113.9	-2.2	3		Yes					
Tb927.3.1560	Conserved hypothetical	42.0	-2.0	2	*						
Tb10.6k15.2360	Conserved hypothetical	14.1	-1.9	2			1				
Tb10.70.1350	Conserved hypothetical	55.3	-1.9	2			1				
Tb927.7.1060	Conserved hypothetical	53.7	-1.9	2	***						
Tb927.7.1640	GTP-binding protein	65.4	-1.9	2					GTPase of unknown function	CDS and PSI-Blast hit to EngA	
Tb927.8.6300	Conserved hypothetical	53.2	-1.9	2	**			***			
Tb08.27P2.50	Hypothetical protein	17.0	-1.8	2			2				
Tb10.70.1620	Conserved hypothetical	152.6	-1.8	2		Yes				(h)	
Tb927.2.4400	Conserved hypothetical	136.8	-1.8	2							
Tb09.211.1850	Conserved hypothetical	46.7	-1.7	2			1				
Tb09.211.3840	Conserved hypothetical	64.9	-1.7	2				***	Leo1-like	Paf1/RNA polymerase II complex	
Tb927.7.4340	Conserved hypothetical	98.0	-1.7	2	*			*	DUF1253		
Tb927.7.7210	Conserved hypothetical	77.4	-1.7	2				***			
Tb927.1.1700	Conserved hypothetical	55.0	-1.6	2	*						
Tb927.7.3040	Conserved hypothetical	69.0	-1.6	2					WD domain, G-beta repeat, RRM_1	No similarity to Nups	
Tb927.8.2570	Conserved hypothetical	48.4	-1.6	2			5	*			
Tb10.26.0240	Conserved hypothetical	44.4	-1.5	2	*			***			
Tb11.01.2520	Conserved hypothetical	66.7	-1.5	2					MA3 domain, MIF4G domain	(h)	
Tb11.01.7740	Conserved hypothetical	32.8	-1.5	2					Tetratricopeptide repeat		
Tb11.55.0017	Conserved hypothetical	21.5	-1.4	2						(e,h)	

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	# of identified peptides	NLS (a)	Signal Peptide (b)	TMH domains (b)	Coiled Coils (c)	Pfam Matches (d)	Bioinformatic Notes	GFP localized?
Tb927.4.1500	Conserved hypothetical	241.6	-1.4	2	*			*	Double-stranded RNA binding motif, DEAD/DEAH box helicase, Helicase conserved C-terminal domain, Helicase associated domain (HA2)		
Tb07.11L3.100	Conserved hypothetical	138.0	-1.3	1							
Tb927.6.1950	Conserved hypothetical	95.7	-1.3	2							
Tb927.8.6640	Conserved hypothetical	67.4	-1.3	2					Male sterility protein	(e)	
Tb09.160.3910	Conserved hypothetical	26.5	-1.2	2	**			***			
Tb09.211.4170	Conserved hypothetical	55.3	-1.2	2							
Tb11.01.7130	Tb-291 membrane assoc.	82.7	-1.1	2				***			
Tb11.02.0380	Conserved hypothetical	83.0	-1.1	2					WD domain, G-beta repeat	No similarity to Nups	
Tb11.02.1000	Conserved hypothetical	91.8	-1.1	2				***	FKBP-type peptidyl-prolyl cis-trans isomerase		
Tb927.8.6870	Conserved hypothetical	95.5	-1.1	2					IQ calmodulin-binding motif		
Tb10.70.4110	Conserved hypothetical	40.0	-1.0	2	*	Yes		***			
Tb927.3.1940	Conserved hypothetical	49.4	-1.0	2			1		DUF477		
Tb927.3.4540	Conserved hypothetical	54.3	-1.0	2							

Table 8: Annotated TbNEP members.

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb927.1.2380	alpha tubulin	49.7	-250.2	47
Tb927.5.3400	Calcium-transporting ATPase (Calcium pump)	110.2	-244.6	51
Tb927.2.4230	NUP-1 protein	406.6	-222.9	47
Tb927.1.2330	beta tubulin	49.6	-215.6	43
Tb927.3.1380	ATPase beta subunit	55.7	-180.4	33
Tb11.02.5500	glucose-regulated protein 78	71.3	-171.3	35
Tb927.4.4380	vacuolar-type proton translocating pyrophosphatase 1	85.8	-162.3	26
Tb10.70.5670	TEF1, elongation factor 1-alpha	49	-161.1	36
Tb927.7.7420	ATP synthase alpha chain, mitochondrial precursor	63.4	-146.7	35
Tb11.01.3110	heat shock protein 70	75.3	-143.4	38
Tb10.70.2650	elongation factor 2	94.2	-132.1	34
Tb927.5.1810	lysosomal/endosomal membrane protein p67	72.6	-128.0	21
Tb11.02.4150	Glycosomal Pyruvate Phosphate Dikinase	100.3	-125.4	37
Tb927.6.4280	Glyceraldehyde-3-phosphate dehydrogenase	43.8	-110.6	24
Tb927.6.3740	heat shock 70 kDa protein, mitochondrial precursor	71.4	-97.8	20
Tb09.160.3820	nop5 protein	54.9	-96.8	21
Tb11.02.1480	mitochondrial processing peptidase alpha subunit, Clan ME, Family M16	57	-94.3	19
Tb10.70.0430	chaperonin Hsp60, mitochondrial precursor	59.4	-91.6	16
Tb927.5.1210	short-chain dehydrogenase	33.8	-89.7	21
Tb10.61.1820	mitochondrial carrier protein	34	-87.9	21
Tb927.2.4210	glycosomal phosphoenolpyruvate carboxykinase	58.5	-87.4	18
Tb927.2.1080	RHS5-c	76.5	-86.5	21
Tb927.8.3750	Nucleolar Protein, SIK1	54.3	-83.3	25
H25N7.12	RHS4	97.7	-80.5	25
Tb11.01.4750	elongation factor 1 gamma	60.4	-76.5	20
Tb10.70.5800	HK2, hexokinase	51.1	-76.1	16
Tb10.61.0980	glycosomal malate dehydrogenase	33.6	-75.7	15
Tb927.4.3590	translation elongation factor 1-beta	28.3	-75.2	16
Tb927.8.1870	tGLP1, golgi/lysosome glycoprotein 1	67.5	-75.2	10
Tb11.02.0070	metallo-peptidase, Clan MF, Family M17	60.5	-74.7	17
Tb10.70.1370	fructose-bisphosphate aldolase, glycosomal	41	-73.3	15
Tb09.160.4250	tryparedoxin peroxidase	22.4	-69.8	16
Tb09.211.3560	glk1, glycerol kinase, glycosomal	56.3	-68.7	16
Tb927.8.3530	glycerol-3-phosphate dehydrogenase	37.7	-66.4	18
Tb927.5.1060	mitochondrial processing peptidase, beta subunit	54	-65.6	17
Tb10.389.1180	P-type H+-ATPase	100.1	-65.5	23
Tb09.160.2840	ACS4, fatty acyl CoA synthetase 4	77.8	-62.8	17
Tb11.02.5280	glycerol-3-phosphate dehydrogenase	66.9	-62.4	15
Tb927.7.6850	TbTS, trans-sialidase	84.4	-61.3	15
Tb927.3.3580	LPG3, lipophosphoglycan biosynthetic protein	87.7	-60.2	13
Tb927.3.4300	PFR1, 73 kDa paraflagellar rod protein	68.6	-59.7	15
Tb927.7.2930	histone H2A	14.2	-59.5	10
Tb10.v4.0045	prohibitin	32.1	-57.5	14
Tb10.70.4200	fatty acyl CoA synthetase	78.9	-56.5	17
Tb10.100.0070	ATP synthase F1 subunit gamma protein	34.3	-53.9	13
Tb09.211.1750	mitochondrial carrier protein	34.2	-51.3	12
Tb927.3.3270	TbPFK, ATP-dependent phosphofructokinase	53.4	-50.6	16
Tb927.4.470	snoRNP protein, GAR1	21.7	-50.0	19
Tb10.389.0690	mitochondrial carrier protein	33.1	-48.3	11
Tb11.01.3370	PEX11, glycosomal membrane protein	24	-47.0	11
Tb10.70.3290	ATP-dependent DEAD-box RNA helicase, DHH1	46.4	-46.7	16
Tb05.5K5.70	nucleolar RNA helicase II, Gu	68.7	-46.5	12
Tb09.160.3270	eukaryotic initiation factor 4a	45.3	-45.9	12
Tb10.70.5110	mMDH, mitochondrial malate dehydrogenase	33.1	-45.7	8
Tb09.211.4700	RISP, reiske iron-sulfur protein, mitochondrial precursor	33.6	-45.5	10
Tb09.160.4310	Glutamate dehydrogenase	112	-44.3	19
Tb927.7.4180	long chain fatty acyl elongase	33.8	-44.2	9
Tb927.5.930	FRDg, NADH-dependent fumarate reductase	123.6	-44.0	12
Tb927.2.4710	RNA-binding protein, RRM1	49.9	-43.3	8
Tb09.160.5480	adenosine transporter	50.6	-42.4	12
Tb11.02.4080	lanosterol 14-alpha-demethylase	54.3	-42.0	14
Tb09.211.0540	FBPase, fructose-1,6-bisphosphate, cytosolic	38.5	-42.0	8
Tb10.05.0230	zinc metallopeptidase	62.7	-41.0	7
Tb927.3.3330	heat shock protein 20	15.8	-39.1	9
Tb927.6.4990	ATP synthase, epsilon chain	20.1	-39.0	7
Tb927.3.1790	pyruvate dehydrogenase E1 beta subunit	37.5	-38.4	8
Tb927.7.6360	histone H2A	18.6	-37.1	13
Tb09.211.2740	Gim5B protein	25.9	-36.1	10
Tb927.5.1710	ribonucleoprotein p18, mitochondrial precursor	21.2	-35.6	6
Tb10.70.1100	translation elongation factor 1-beta	21.9	-35.5	9
Tb927.7.210	proline oxidase	63.8	-35.1	10
Tb927.8.760	Nopp44/46.	35	-35.1	7
Tb927.5.4190	histone H4	11.1	-34.8	9
Tb09.211.4511	kinetoplastid membrane protein KMP-11	11	-34.6	6
Tb10.v4.0052	microtubule-associated protein 2, MAP2	560.8	-34.1	11
Tb10.406.0350	histone H2B	12.5	-33.7	10
Tb927.6.4210	aldehyde dehydrogenase	64.5	-33.5	11
Tb10.61.2130	ATP-dependent DEAD/H RNA helicase, DBP1	71.3	-33.1	13

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb927.3.5050	60S ribosomal protein L4 (L1).	41.8	-31.9	11
Tb927.2.100	RHS1	94.8	-31.2	16
Tb927.5.900	oligosaccharyl transferase subunit	92.1	-31.2	11
Tb927.2.470	retrotransposon hot spot protein RHS4-c .	98.2	-30.6	21
Tb11.46.0001	60S acidic ribosomal subunit protein	34.6	-30.4	10
Tb10.6k15.0410	60S ribosomal protein L18	21.8	-30.4	5
Tb927.1.2530	Histone H3	14.7	-30.0	7
Tb10.61.0150	inosine-5'-monophosphate dehydrogenase	48.4	-30.0	6
Tb11.02.1085	40s ribosomal protein S4	30.6	-29.6	8
Tb927.4.1330	type IB DNA topoisomerase large subunit .	79.2	-29.0	10
Tb11.02.3210	triose-phosphate isomerase	26.8	-28.8	9
Tb10.6k15.1350	pteridine transporter	69.9	-28.6	9
Tb927.8.1610	major surface protease gp63	62.9	-28.5	8
Tb10.v4.0247	s-adenosyl-L-methionine-c-24-delta-sterol-methyl transferase a	40.1	-28.3	6
Tb927.1.420	RHS5	76.2	-28.0	19
Tb927.2.1210	RHS4	95.1	-27.5	16
Tb10.26.0560	60S ribosomal protein L6	21.1	-27.4	5
Tb927.6.1500	DHAP, alkyl-dihydroxyacetone phosphate synthase	69	-27.3	8
Tb927.8.650	cation-transporting ATPase	140.3	-27.1	10
Tb927.3.1840	3-oxo-5-alpha-steroid 4-dehydrogenase	33.3	-27.1	8
Tb10.26.1080	heat shock protein 83	80.7	-26.5	8
Tb927.8.5010	PFR2, 69 kDa paraflagellar rod protein	69.5	-26.4	7
Tb927.7.710	heat shock, 70 Kda	70.1	-26.1	13
Tb927.8.7410	calreticulin	44.9	-25.6	8
Tb927.5.3510	SMC3	136.2	-24.3	14
Tb927.5.520	stomatin-like protein	55.9	-24.3	8
Tb09.160.3670	ribosomal protein S6	13.5	-23.8	3
Tb09.244.2630	40S ribosomal protein S6	28.4	-23.5	7
Tb927.5.2080	inosine-5'-monophosphate dehydrogenase	52.2	-23.4	6
Tb927.2.2440	RPN6, proteasome regulatory non-ATPase subunit 6	57.2	-23.4	6
Tb11.01.4940	AAA ATPase	141	-23.2	13
Tb927.8.1420	acyl-CoA dehydrogenase, mitochondrial precursor	55.9	-23.2	5
Tb927.4.1790	ribosomal protein L3	54.3	-22.5	5
Tb927.7.2370	40S ribosomal protein S15	20	-22.5	4
Tb11.02.2880	DNAJ	84.6	-22.4	7
Tb927.6.1520	aquaporin 3	35	-22.2	6
Tb11.0290	RPS14, 40s ribosomal protein S14	15.5	-21.9	6
Tb09.160.2770	ACS1, fatty acyl CoA syntetase 1	78.9	-21.8	9
Tb09.160.1160	Nop86	85.8	-21.8	8
Tb927.1.4100	cytochrome C oxidase subunit IV	40.5	-21.7	7
Tb927.8.5460	flagellar calcium-binding protein TB-44A	45.6	-21.7	5
Tb927.4.4620	cytochrome c oxidase VIII (COX VIII)	18.7	-21.0	6
Tb10.26.0790	PSSA-2 procyclic form surface glycoprotein	46	-20.9	4
Tb927.6.2550	RNA-binding protein, possible PAB1	79.7	-20.7	12
Tb09.244.2760	cytosolic coat protein	24.5	-20.7	7
Tb927.8.4810	prohibitin	31.4	-20.6	10
Tb10.389.1500	short-chain dehydrogenase	36	-20.5	10
Tb10.70.3360	40S ribosomal protein S3a	29.4	-20.5	7
Tb927.3.960	protein transport protein Sec61 gamma subunit	7.6	-20.4	4
Tb927.4.395	cytoskeleton-associated protein CAP5.5	94.5	-20.3	8
Tb10.6k15.3080	dihydrolipoamide acetyltransferase precursor	48	-20.3	7
Tb10.70.3510	60S ribosomal protein L18a	20.8	-20.1	6
Tb10.6k15.3160	mammalian Fibrillarin, yNOP1	31.6	-19.9	9
Tb11.02.0250	heat shock protein, mitochondrial precursor	84.1	-19.6	6
Tb11.01.3560	vacuolar ATP synthase subunit B	55.5	-19.5	6
Tb927.7.4170	fatty acid elongase	30.4	-19.4	4
Tb10.389.0890	pyruvate dehydrogenase E1 component alpha subunit	42.4	-19.0	6
Tb927.3.3490	TDP1 high mobility group protein	30.8	-19.0	3
Tb927.4.2070	Antigenic protein, paralogous to NUP-1	511	-18.9	10
Tb927.8.2000	PPIase cyclophilin type peptidyl-prolyl cis-trans isomerase	32.8	-18.9	7
Tb10.70.6660	hypoxanthine-guanine phosphoribosyltransferase	26.3	-18.7	2
Tb927.2.340	RHS4-a	97.7	-18.3	20
Tb09.211.0560	RNA-binding protein, DRDB3	36.9	-18.1	9
Tb11.50.0007	dynein light chain	10.4	-17.9	3
Tb11.02.0750	TCP-1-zeta t-complex protein 1, zeta subunit	59.5	-17.9	3
Tb927.8.2630	kinesin	85.1	-17.7	7
Tb927.7.990	chaperone protein DNAJ	86.5	-17.7	6
Tb10.6k15.3610	delta-6 fatty acid desaturase	47.9	-17.6	6
Tb11.01.7800	nucleoside diphosphate kinase .	16.8	-17.5	5
Tb927.2.5160	DNAJ	44.7	-17.5	3
Tb09.211.1550	chaperone protein DNAJ	56.6	-17.1	9
Tb10.70.0820	universal minicircle sequence binding protein (UMSBP)	21.8	-16.9	4
Tb10.6k15.3350	40S ribosomal protein S24E	15.6	-16.8	2
Tb11.01.3610	membrane-bound acid phosphatase	53.3	-16.7	4
H25N7.01	RHS2	93.6	-16.6	11
Tb11.02.2960	mitochondrial carrier protein	29.8	-16.5	5
Tb10.61.1920	fibrillarin	31.6	-16.3	10
Tb927.3.1410	COX VII cytochrome c oxidase VII	19.2	-16.2	4
Tb927.1.2230	calpain-like protein fragment	13.5	-16.2	4
Tb927.4.2000	RuvB-like DNA helicase	52.5	-16.0	8
Tb09.160.4450	RPS3 40S ribosomal protein S3	30.3	-16.0	3
Tb09.211.0120	nascent polypeptide associated complex subunit	20.1	-16.0	3
Tb927.4.4210	ATP-dependent zinc metallopeptidase	96.1	-15.7	6

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb11.01.3180	guanine nucleotide-binding protein beta subunit	34.6	-15.4	3
Tb927.4.1080	V-type ATPase, A subunit	67.7	-15.3	5
Tb927.7.3120	Sm-D1 small nuclear ribonucleoprotein SmD1	11.7	-15.1	3
Tb11.02.1680	mannose-specific lectin	62.2	-15.0	5
Tb927.8.1890	cytochrome c1, heme protein, mitochondrial precursor	30	-14.8	7
Tb927.3.2230	succinyl-CoA synthetase alpha subunit	31.4	-14.8	3
Tb927.7.2700	B5R NADH-cytochrome b5 reductase	31.8	-14.5	5
Tb10.6k15.2020	THT2A glucose transporter	56.8	-14.5	3
Tb11.42.0003	TCP-1-beta t-complex protein 1, beta subunit	58	-14.4	6
Tb10.70.1770	Eukaryotic translation initiation factor 6 (eIF- 6)	26.9	-14.4	4
Tb10.6k15.2330	TCP-1-theta t-complex protein 1, theta subunit	58	-14.4	4
Tb10.406.0290	protein tyrosine phosphatase	29.6	-14.4	3
Tb09.211.0680	CAAX prenyl protease 1	48.8	-14.3	5
Tb09.244.2570	calcium motive p-type ATPase	113.9	-14.2	6
Tb927.5.3810	orotidine-5-phosphate decarboxylase/orotate phosphoribosyltransferase	49.9	-14.1	7
Tb927.3.3310	60S ribosomal protein L13	26.6	-14.0	5
Tb927.1.120	RHS4	97.9	-13.7	23
Tb927.3.4760	dynamin	73.1	-13.7	6
Tb09.211.3510	ATP-dependent DEAD/H RNA helicase; NPC-associated RNA helicase (Allen et al., 2001)	82.6	-13.7	5
Tb11.12.0011	ATP-dependent DEAD/H RNA helicase	45.4	-13.5	5
Tb09.160.2910	tricarboxylate carrier	35.6	-13.3	6
Tb11.03.0090	ribokinase	35.2	-13.3	3
Tb927.3.2150	protein phosphatase 2C	39.9	-13.2	6
Tb09.160.4200	60S acidic ribosomal protein	11	-13.2	4
Tb09.244.2730	60S ribosomal protein L5	34.6	-13.1	6
Tb927.4.1920	GPI transamidase component GPI16 .	75.7	-13.1	4
Tb11.0880	60S ribosomal protein L21E	17.9	-13.0	6
Tb927.8.2160	PGPA MRPA multidrug resistance protein A	174.5	-13.0	5
Tb11.02.0870	Ran-binding protein 1	17.6	-13.0	3
Tb10.61.0110	ATP-dependent zinc metallopeptidase	62.7	-13.0	3
Tb11.01.5100	paraflagellar rod component	68.3	-12.9	4
Tb927.4.3300	mitochondrial ATP-dependent zinc metallopeptidase	78.8	-12.9	3
Tb09.211.2640	60S ribosomal protein L23	14.9	-12.8	5
Tb927.5.1300	vacuolar proton translocating ATPase subunit A	89.5	-12.7	6
Tb10.61.0600	RPN9 proteasome regulatory non-ATP-ase subunit 9	45.9	-12.7	1
Tb927.8.3060	cytosolic leucyl aminopeptidase	71.2	-12.5	5
Tb927.4.3550	60S ribosomal protein L13a	39.1	-12.5	4
Tb10.70.3240	short-chain dehydrogenase	38	-12.4	3
Tb11.02.4100	pretranslocation protein (Sec61)	53.6	-12.2	4
Tb11.02.0780	squalene monooxygenase	63.7	-12.2	4
Tb927.7.3440	l/6 autoantigen	27	-12.2	4
Tb927.4.2180	60S ribosomal protein L35A	17	-12.2	3
Tb09.160.1820	cytochrome c oxidase subunit V	22.2	-12.1	4
Tb11.01.6360	metalloprotease	74.4	-12.1	4
Tb09.160.3090	heat shock protein	90.8	-12.0	7
Tb927.6.3890	replication factor C, subunit 2	38.7	-12.0	6
Tb927.6.3160	splicing factor 3A, SAP61	61.4	-12.0	5
Tb927.2.2970	mitochondrial carrier protein	33.9	-12.0	5
Tb11.01.7190	B5R NADH-cytochrome b5 reductase	33.8	-12.0	5
Tb10.70.0120	COP-coated vesicle membrane protein erv25 precursor	28.1	-12.0	4
Tb11.01.4702	cytochrome c oxidase subunit 10	13.6	-12.0	3
Tb927.4.3570	Elongation Factor 1-beta	28.3	-11.9	12
Tb927.3.5090	tryparedoxin	22.1	-11.9	4
Tb927.7.1050	40S ribosomal protein S16	16.9	-11.6	5
Tb10.70.4380	cytochrome C oxidase assembly protein	26.9	-11.6	1
Tb927.2.4550	FtsJ cell division protein	102.8	-11.5	5
Tb927.8.3380	electron transfer protein	27.2	-11.5	4
Tb927.8.1510	RNA Helicase, ATP/DEAD, DBP2	62.4	-11.4	6
Tb927.5.1200	exosome component, CSL4	32.5	-11.4	4
Tb927.7.2710	NADH-cytochrome b5 reductase	31.8	-11.3	4
Tb09.160.4560	AK arginine kinase	44.6	-11.3	4
Tb927.4.1930	RNA-binding protein	31.2	-11.3	2
Tb09.244.2600	ankyrin-repeat protein	334.2	-11.2	9
Tb11.01.2530	kinesin-like protein	69.8	-11.2	4
Tb927.2.5980	HSP100 ATP-dependent Clp protease subunit, heat shock protein 100 (HSP100)	96.9	-10.9	3
Tb10.389.1170	P-type H+-ATPase	100.5	-10.8	21
Tb10.6k15.2900	ABC transporter	75	-10.8	7
Tb09.211.2260	protein kinase	118	-10.7	4
Tb927.6.970	CP cysteine peptidase precursor	48.4	-10.7	4
Tb11.22.0012	ERF1 eukaryotic peptide chain release factor subunit 1	50.9	-10.7	3
Tb10.70.1190	VCP valosin-containing protein homolog	85.7	-10.6	5
Tb09.211.2570	TCP-1-eta t-complex protein 1, eta subunit	61.4	-10.6	4
Tb11.02.1230	B5R NADH-cytochrome b5 reductase	34	-10.6	3
Tb10.05.0220	60S ribosomal protein L10a	24.5	-10.5	1
Tb11.01.1465	nascent polypeptide associated complex alpha subunit	11.3	-10.3	2
Tb927.5.2380	hydrolase, alpha/beta fold family	34.4	-10.1	4
Tb927.3.4190	endosomal integral membrane protein	81.5	-10.1	3
Tb11.01.2280	ubiquinone biosynthesis methyltransferase	32	-10.1	3
Tb927.8.6580	succinate dehydrogenase flavoprotein	66.8	-10.0	4
Tb10.70.0830	CHC clathrin heavy chain	190.5	-9.9	7
Tb11.01.8510	TCP-1-alpha	54.5	-9.9	4
Tb927.3.2900	elongation initiation factor 2 alpha subunit	46.7	-9.9	2
Tb927.4.560	dynein heavy chain	479.5	-9.8	8

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb927.7.5170	60S ribosomal protein L23a	18.1	-9.7	3
Tb11.02.4070	60S ribosomal protein L28	16.8	-9.7	2
Tb10.406.0600	SMC2 structural maintenance of chromosome 2	133.9	-9.4	7
Tb11.01.6880	cytosolic coat protein	25.5	-9.3	3
Tb11.01.1440	aminomethyltransferase	41.1	-9.2	1
Tb927.8.900	splicing factor pTSR1	37.4	-9.0	5
Tb927.2.560	RHS4-f	95.8	-8.9	12
Tb11.47.0035	calpain-like cysteine peptidase	660.2	-8.9	7
Tb927.7.3940	mitochondrial carrier protein	36.3	-8.9	5
Tb927.4.3740	FAZ1	192.4	-8.7	7
Tb09.211.2150	poly(A)-binding protein 1	62.1	-8.7	6
Tb927.7.270	ribosome biogenesis protein	49.3	-8.7	4
Tb927.2.510	RHS4	95.1	-8.6	8
Tb11.01.6300	phosphatidylinositol 3-related kinase	320.3	-8.6	7
Tb11.55.0014	vesicular transport protein (CDC48 homologue)	78.4	-8.6	4
Tb11.02.4570	pumilio-repeat, RNA-binding protein	75.1	-8.6	4
Tb10.70.5820	hexokinase 1	51.2	-8.4	15
Tb927.1.3180	40S ribosomal protein S11	20	-8.4	3
Tb11.01.1475	40S ribosomal protein S27	9.6	-8.4	1
Tb927.3.3670	RNA-binding protein, Nop4?	66.8	-8.2	5
Tb10.70.0140	HEL64 ATP-dependent DEAD/H RNA helicase HEL64	64	-8.1	4
Tb10.6k15.3250	succinyl-CoA ligase [GDP-forming] beta-chain	54.7	-8.1	4
Tb10.6k15.2550	Replication factor C, subunit 5, RFC3	38.7	-8.1	3
Tb11.02.5250	histone H2B variant	15.7	-8.1	3
Tb11.01.7400	GPI transamidase component Tta1	41.8	-8.1	3
Tb927.8.1160	vacuolar-type Ca2+-ATPase	121.1	-8.0	6
Tb927.7.1730	60S ribosomal protein L7	27.6	-8.0	5
Tb10.70.1670	40S ribosomal protein S10	19.2	-8.0	3
Tb10.70.6360	RPN5 proteasome regulatory non-ATP-ase subunit 5	54.8	-8.0	2
Tb10.6k15.2180	COX IX cytochrome c oxidase subunit IX	13.7	-7.8	2
Tb927.8.2740	TbRBP38 TbRBP38 mitochondrial RNA binding protein	38.4	-7.8	1
Tb11.01.3550	2-oxoglutarate dehydrogenase, E2 dihydrolipoamide succinyltransferase	41.1	-7.7	5
Tb927.7.3500	glutathione-S-transferase/glutaredoxin	35.4	-7.6	4
Tb11.01.5590	NRBD2 TbP37 RNA-binding protein	30.2	-7.6	3
Tb927.3.4680	RAB GDP dissociation inhibitor alpha	49.3	-7.6	1
Tb927.6.4370	eukaryotic translation initiation factor 3 subunit 7-like protein	61.3	-7.5	6
Tb11.01.1740	2-oxoglutarate dehydrogenase E1 component	112.8	-7.5	4
Tb10.61.1330	nucleosome assembly protein	41.2	-7.5	3
Tb09.160.4380	28G16.480 succinate dehydrogenase	21.2	-7.5	3
Tb927.4.1630	RRP6 ribosomal RNA processing protein 6	82.1	-7.5	2
N19B2.040	RHS3	66.6	-7.3	7
Tb11.01.3020	40S ribosomal protein L14	21.4	-7.3	2
Tb11.18.0013	RPN3 proteasome regulatory non-ATP-ase subunit 3	38.1	-7.2	2
Tb11.01.4870	inner membrane preprotein translocase Tim17	16.2	-7.1	4
Tb09.244.2790	rRNA processing protein	28.4	-7.1	3
Tb10.6k15.0960	54 NDH2 NADH dehydrogenase	54.3	-7.1	3
Tb11.01.7170	C-14 sterol reductase	51.1	-7.0	2
Tb10.70.0170	chaperone protein DNAJ	31.1	-7.0	1
Tb10.6k15.2050	RPS12 40S ribosomal protein S12	16	-6.9	3
Tb927.3.2600	ATP-dependent DEAD/H RNA helicase	244.8	-6.8	6
Tb09.160.4600	ABC transporter	72.1	-6.8	5
Tb927.5.2290	ATP-dependent RNA helicase	249.1	-6.8	5
Tb11.01.2560	40S ribosomal protein SA	27.5	-6.8	4
Tb927.8.3150	TCP-1-gamma TCP-1-gamma t-complex protein 1 gamma subunit	60.7	-6.8	4
Tb11.01.1790	60S ribosomal protein L29	8.2	-6.8	3
Tb11.02.0290	succinyl-coA:3-ketoacid-coenzyme A transferase, mitochondrial precursor	53	-6.7	4
Tb10.70.4570	RPN11 proteasome regulatory non-ATPase subunit 11	33.8	-6.7	2
Tb927.1.220	retrotransposon hot spot (RHS) protein	94.5	-6.6	10
Tb10.100.0060	centromere/microtubule binding protein, CBF5	48.2	-6.6	2
Tb11.01.1370	eIF-3 beta; TRIP-1 eukaryotic translation initiation factor 3 subunit	37.7	-6.6	1
Tb927.2.3080	oleate desaturase	47	-6.5	5
Tb09.211.4550	60S ribosomal protein L12	24	-6.5	3
Tb927.8.730	nucleolar RNA-binding protein	28.3	-6.5	2
Tb927.6.2700	Sm-E small nuclear ribonucleoprotein Sm-E	9.6	-6.5	1
Tb11.55.0009	GBP21 mitochondrial RNA binding protein 1	23.2	-6.4	3
Tb10.70.1720	dynein heavy chain	474.4	-6.3	5
Tb927.8.7600	amino acid transporter	58.1	-6.3	4
Tb10.70.7010	60S ribosomal protein L9	21.8	-6.3	4
Tb09.211.3310	replication factor C	39.5	-6.3	3
Tb11.01.3675	40S ribosomal protein S17	16.2	-6.3	1
Tb927.8.7170	inositol polyphosphate 1-phosphatase	42	-6.2	4
Tb11.02.5210	RNA binding protein	33.5	-6.2	1
Tb11.01.4622	calmodulin	16.8	-6.2	1
Tb09.211.4600	short-chain dehydrogenase	51	-6.2	1
Tb10.6k15.2290	protein disulfide isomerase	55.5	-6.1	3
Tb927.3.4910	signal peptide peptidase	38.6	-6.1	2
Tb927.6.4340	TbLSm5 U6 snRNA-associated Sm-like protein LSm5p	12.7	-6.0	1
Tb927.4.870	dynein heavy chain	509.7	-5.8	6
Tb09.211.0340	QM 60S ribosomal protein L10	24.7	-5.8	4
Tb927.2.3370	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	43.6	-5.8	4
Tb10.389.0910	60S ribosomal protein L34	19.3	-5.8	1
Tb927.8.1940	endosomal integral membrane protein	70.4	-5.7	1
Tb11.02.5470	vacuolar type H+ ATPase subunit	26.3	-5.7	1

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb10.70.0090	dynein light chain	14	-5.6	1
Tb927.4.450	coatomer alpha subunit	132.7	-5.5	5
Tb10.05.0150	citrate synthase	53.5	-5.5	4
Tb927.5.2570	translation initiation factor	79.7	-5.5	3
Tb11.01.4701	MBAP1 membrane-bound acid phosphatase 1 precursor	59.3	-5.5	3
Tb10.61.2090	60S ribosomal protein L17	19	-5.5	2
Tb11.01.4050	heat shock protein HslVU, ATPase subunit HslU	54.8	-5.4	3
Tb927.6.3050	aldehyde dehydrogenase family	59.6	-5.3	2
Tb11.01.4620	ubiquitin ligase	203.2	-5.2	4
Tb10.61.3190	60S ribosomal protein L7	34.4	-5.2	3
Tb927.8.6000	fatty acid desaturase	49.3	-5.0	5
Tb10.6k15.0580	RPT6 proteasome regulatory ATPase subunit 6	45.7	-5.0	4
Tb927.2.4700	30M24.230 hypothetical protein, conserved	17	-5.0	2
Tb927.8.2310	(H+)-ATPase G subunit	12.7	-5.0	1
Tb10.70.1740	40S ribosomal protein S18	17.4	-4.9	3
Tb927.3.780	TbPSA7 proteasome alpha 7 subunit	25.4	-4.9	3
Tb927.8.1600	lysyl-tRNA synthetase	66.7	-4.9	2
Tb10.61.2070	RPS2 40S ribosomal protein S2	28.6	-4.8	3
Tb927.8.5120	cytochrome c	12.2	-4.8	2
Tb927.7.1300	protein disulfide isomerase	41.9	-4.8	2
Tb10.70.4800	ribosomal protein S25	12.5	-4.7	3
Tb927.7.230	40S ribosomal protein S33	11.1	-4.7	2
Tb10.61.1560	intraflagellar transport protein IFT80/CHE2	86.6	-4.7	1
Tb927.2.1180	25N24.100 retrotransposon hot spot (RHS) protein	67.7	-4.6	8
Tb927.4.750	50S ribosomal protein L7Ae	16.3	-4.6	1
Tb927.1.4490	acetyltransferase	19.6	-4.6	1
Tb927.6.3630	sphingosine phosphate lyase-like protein	59.4	-4.5	3
Tb927.4.2450	thioredoxin	44.4	-4.5	2
Tb927.2.5800	SBPase 1F7.270 sedoheptulose-1,7-bisphosphatase	35.4	-4.5	2
Tb11.01.6800	1-acyl-sn-glycerol-3-phosphate acyltransferase protein	30.3	-4.5	2
Tb10.6k15.3030	mitochondrial ATP-dependent zinc metallopeptidase	71	-4.5	2
Tb11.01.1920	60S ribosomal protein L22	15.2	-4.5	1
Tb10.61.1040	DNA-directed RNA polymerase, alpha subunit	37.3	-4.5	1
Tb927.1.2580	TbEAP1 RNasePH-like protein	42.6	-4.4	3
Tb11.02.5190	pantothenate kinase subunit	162	-4.4	3
Tb11.01.7535	60S ribosomal protein L27	15.5	-4.4	2
Tb927.5.1470	B5R NADH-cytochrome b5 reductase	32.7	-4.4	2
H25N7.20	ESAG4 protein chr intermediate-sized-chromosome	139.1	-4.3	5
Tb927.5.800	Casein Kinase	38.3	-4.3	3
Tb927.2.1920	25N14.305 expression site-associated gene (ESAG) protein	53	-4.3	1
Tb09.211.4610	vesicle-associated membrane protein	23.1	-4.3	1
Tb927.6.3150	Hydin	500.8	-4.2	5
Tb927.3.5520	RPN1 26S proteasome regulatory non-ATPase subunit	99.8	-4.2	3
Tb927.2.3030	10C8.315 ATP-dependent Clp protease subunit, heat shock protein 78 (HSP78)	90.6	-4.2	3
Tb11.02.2070	long-chain-fatty acid-CoA ligase protein	73.3	-4.2	3
Tb927.7.4160	fatty acid elongase	34.3	-4.2	2
Tb09.160.2810	ACS3 1L12.290 fatty acyl CoA synthetase 3	77.8	-4.1	16
Tb11.02.0030	dynein heavy chain	484.6	-4.1	5
Tb10.6k15.3220	MTR4 ATP-dependent DEAD/H RNA helicase	107.2	-4.1	4
Tb10.61.1870	aminopeptidase	42.6	-4.1	2
Tb10.389.0160	periodic tryptophan protein 2	102.8	-4.1	2
Tb927.7.5680	deoxyribose-phosphate aldolase	30	-4.1	1
Tb927.2.3440	28H13.70 D-alanyl-glycyl endopeptidase-like protein	29.9	-4.1	1
Tb927.2.1170	25N24.105 retrotransposon hot spot (RHS) protein	76.5	-4.0	13
Tb927.1.3230	cell division cycle protein	77.2	-4.0	3
Tb10.6k15.1000	branch point binding protein, MSL5(?)	31.6	-4.0	2
Tb11.01.5860	TCP-1-epsilon t-complex protein 1, epsilon subunit	59.3	-4.0	1
Tb927.7.2070	heat shock protein DnaJ	36.3	-3.9	4
Tb927.8.7530	3,2-trans-enoyl-CoA isomerase, mitochondrial precursor	39.7	-3.9	4
Tb927.2.3780	28H13.240 translation initiation factor IF-2	94.4	-3.9	4
Tb10.70.7840	choline dehydrogenase	56.8	-3.9	4
Tb927.7.1340	HSP10 10 kDa heat shock protein	10.6	-3.9	3
Tb927.7.1790	adenine phosphoribosyltransferase	25.6	-3.9	1
Tb10.70.4300	U2 splicing auxiliary factor	29	-3.9	1
Tb927.8.750	nucleolar RNA-binding protein	39.7	-3.8	3
Tb927.8.5090	TRP11 TRP11 DNA-directed RNA polymerase I largest subunit	195.8	-3.8	2
Tb927.3.4020	phosphatidylinositol 4-kinase alpha	255.3	-3.7	4
Tb927.4.3870	receptor-type adenylate cyclase GRESAG 4	134.8	-3.7	4
Tb11.02.4040	protein transport protein Sec31	131	-3.7	3
Tb11.02.4030	ERF3 eukaryotic release factor 3	76.8	-3.7	2
Tb11.02.1105	NT8.1 NBT1 nucleobase/nucleoside transporter 8.1	47.6	-3.7	1
Tb10.6k15.0460	chaperone protein DNAJ	29.8	-3.7	1
Tb927.2.2670	histone H4	11.2	-3.6	4
Tb927.8.4950	kinesin	165	-3.6	4
Tb10.70.4280	delta-1-pyrroline-5-carboxylate dehydrogenase	61.9	-3.6	3
Tb11.01.6590	zinc finger protein 2, ZFP2	15.7	-3.6	1
Tb927.2.5660	1F7.200 adenylate kinase	29.3	-3.6	1
Tb11.01.1320	oxidoreductase	58.5	-3.5	3
Tb10.70.7020	RPS23 40S ribosomal protein S23	15.9	-3.5	1
Tb927.2.240	3B10.75 retrotransposon hot spot (RHS) protein	76.6	-3.4	19
Tb927.7.2550	RPT5 proteasome regulatory ATPase subunit 5	49.1	-3.4	2
Tb10.61.0440	peroxisome assembly protein	43.1	-3.4	2
Tb09.160.2780	fatty acyl CoA synthetase 2	79.5	-3.3	7

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb11.55.0006	TblFT88 intraflagellar transport protein IFT88	89.6	-3.3	5
Tb11.02.1280	subtilisin-like serine peptidase	161.4	-3.3	4
Tb10.61.0990	kinesin	187.4	-3.3	4
Tb11.01.1680	polyubiquitin	76.5	-3.3	2
Tb11.01.7990	myosin	120.4	-3.2	4
Tb927.7.5230	lanosterol synthase	102.9	-3.2	3
Tb10.389.0880	heat shock protein	90.8	-3.2	1
Tb927.6.5050	V-type ATPase, C subunit	23	-3.1	2
Tb10.70.3170	60S ribosomal protein L30	11.5	-3.1	1
Tb927.7.280	cyclophilin-type peptidyl-prolyl cis-trans isomerase	24.2	-3.1	1
Tb927.8.6170	TK TK transketolase	72.5	-3.1	1
Tb927.5.1520	heat shock protein HslVU, ATPase subunit HslU	52.4	-3.0	3
Tb10.26.0140	pumillio RNA binding protein	93	-3.0	1
Tb927.5.3120	translation initiation factor	34.9	-3.0	1
Tb927.8.1990	TRYP2 TRYP2 tryparedoxin peroxidase	25.6	-3.0	1
Tb927.4.4360	monoglyceride lipase	34.9	-3.0	1
Tb09.211.1695	Sm-F small nuclear ribonucleoprotein Sm-F	8.3	-3.0	1
Tb10.70.7050	TCP-1-delta t-complex protein 1, delta subunit	58.3	-2.9	3
Tb11.01.8470	dihydrolipoyl dehydrogenase	50.4	-2.9	2
Tb927.4.3840	nucleolar protein	59	-2.9	2
Tb927.6.3840	reticulon domain protein	21	-2.8	3
Tb10.70.3950	RNA-binding protein	96.5	-2.8	2
Tb927.5.550	vacuolar ATP synthase	42.8	-2.8	2
Tb927.2.450	RHS4	97.9	-2.7	16
Tb927.6.140	RHS5	69.3	-2.7	6
Tb927.8.4170	RNA-binding protein	93	-2.7	3
Tb11.02.0090	kinesin	172.2	-2.7	3
Tb10.61.1090	histone H3 variant	16	-2.6	3
Tb927.4.4490	MRPE; PGPA multidrug resistance protein E	193.6	-2.6	2
Tb11.01.3080	heat shock protein 70	73.5	-2.5	10
Tb11.03.0670	transcription factor	67.5	-2.5	4
Tb927.3.3590	U3 small nucleolar ribonucleoprotein protein MPP10	75.2	-2.5	3
Tb11.01.3390	TOP2 DNA topoisomerase II	162.6	-2.5	3
Tb927.1.2670	axoneme central apparatus protein	56	-2.5	2
Tb927.4.3890	ATP-dependent RNA helicase	122.7	-2.5	2
Tb927.4.250	retrotransposon hot spot (RHS) protein	76.5	-2.4	16
Tb927.3.930	dynein heavy chain	531	-2.4	3
Tb927.1.1370	rRNA biogenesis protein, RRP5	75	-2.4	2
Tb09.160.4240	28G16.410 nucleosome assembly protein-like protein	47.5	-2.4	2
Tb11.02.2380	retrotransposon hot spot protein (RHS, pseudogene)	0	-2.3	8
Tb11.02.0760	dynein heavy chain	530.8	-2.2	4
Tb09.244.2560	kinesin	74.5	-2.2	2
Tb11.02.4640	TTL tubulin-tyrsoine ligase-like protein	108.3	-2.2	2
Tb927.5.890	oligosaccharyl transferase subunit	89.2	-2.1	8
Tb927.7.7450	GTP-binding protein	68.4	-2.1	3
Tb927.4.1270	RuvB-like DNA helicase, RVB1	49.8	-2.0	2
Tb10.100.0160	COXVI cytochrome C oxidase subunit VI	19.1	-2.0	2
Tb927.4.5250	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	46.5	-2.0	2
Tb09.160.3710	28G16.165 proliferative cell nuclear antigen (PCNA)	32.2	-2.0	1
Tb927.4.1020	serine-palmitoyl-CoA transferase	54	-1.9	2
Tb11.01.3690	splicing factor 3B subunit 1	121.9	-1.8	2
Tb11.27.0001	receptor-type adenylate cyclase GRESAG 4	137.5	-1.8	2
Tb927.8.5250	coatomer delta subunit	57.2	-1.8	2
N19B2.135	RHS4 retrotransposon hot spot protein, RHS4 chr intermediate-sized-chromosome	96.3	-1.6	11
Tb11.24.0012	variant surface glycoprotein (VSG, atypical)	51.6	-1.6	2
Tb10.6k15.0270	60S acidic ribosomal protein	25.6	-1.6	2
Tb927.2.5280	30J2.90 trans-sialidase	77.1	-1.6	2
Tb927.7.1600	C-1-tetrahydrofolate synthase, cytoplasmic	31.9	-1.6	2
Tb927.7.680	chaperone protein DNAJ	88.6	-1.6	2
Tb09.160.4220	General transcription factor IIB	37.6	-1.6	1
Tb927.4.4250	UDP-Gal or UDP-GlcNAc-dependent glycosyltransferase	44.5	-1.5	2
Tb10.406.0560	microtubule-associated protein, MAP	237.3	-1.4	10
Tb927.2.400	3B10.155 retrotransposon hot spot (RHS) protein	90.2	-1.4	5
Tb10.406.0050	HNRNPA, RNA-binding protein	32.3	-1.4	2
Tb11.01.7390	AMP Deaminase	163.1	-1.4	2
Tb11.01.5225	cytochrome B5	12.8	-1.4	2
Tb10.70.5480	ubiquitin-conjugating enzyme variant Kua homologue	33.7	-1.4	2
Tb927.7.320	TbRBP8, RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	20.2	-1.4	2
Tb09.211.1820	DNA polymerase epsilon catalytic subunit	254.8	-1.4	1
Tb10.100.0080	40S ribosomal protein S6	28.4	-1.3	6
Tb927.8.8330	calpain	98.3	-1.3	4
Tb09.160.4570	AK arginine kinase	41.5	-1.3	4
Tb10.70.5950	calpain	177.9	-1.3	2
Tb927.7.7500	iron/ascorbate oxidoreductase family protein	36.9	-1.3	2
Tb927.6.990	CP cysteine peptidase precursor	48.2	-1.2	4
Tb09.211.0330	chaperone protein DNAJ	35.5	-1.2	2
Tb11.02.0620	NOG1 nucleolar GTP-binding protein 1	74.7	-1.2	2
Tb927.6.4300	GAPDH glyceraldehyde 3-phosphate dehydrogenase, glycosomal	43.8	-1.1	24
Tb10.406.0390	histone H2B	12.5	-1.1	10
Tb927.8.1200	TbA2 TbA2 vacuolar-type Ca2+-ATPase 2	118.9	-1.1	5
Tb09.211.2420	p277 PRP8 protein homologue	276.8	-1.1	2
Tb927.4.5030	PP1 serine/threonine protein phosphatase PP1	39.3	-1.1	2
Tb11.01.5690	DRBD4 RNA-binding protein	54.6	-1.0	2

Accession Number	GeneDB Annotation	Mass (kDa)	Protein Expect	Total # of peptides
Tb927.8.1330	60S ribosomal protein L7a	30.8	-1.0	2
Tb09.211.0140	chaperone protein DNAJ	47.7	-0.8	2
Tb927.2.6200	adenosine transporter 2	51.2	-0.3	4
Tb927.2.6320	adenosine transporter 2	50.8	-0.2	4

APPENDIX D — *S. CEREVISIAE* DOUBLE MUTATION MATRIX

To plan which nucleoporins to target for RNAi analysis, we conducted a literature survey of the genetic interactions between yeast nucleoporins using the BIOBASE Proteome and SGD internet genomic databases. The results of this survey is summarize in Figure 34. Nups shaded in blue are FG-repeat domain containing Nups. Those shaded in green are the scaffold Nups and those Nups colored pink are the transmembrane anchoring Nups. The first column indicates those Nups which exhibit a lethal phenotype (denoted with an 'X'). If two Nups are synthetically lethal, an 'X' is placed at their intersection within the matrix. White space is the result of insufficient data, but since there have been several large scale studies, it is likely that those double knockouts are viable (Collins et al., 2007; Giaever et al., 2002; Loeillet et al., 2005; Tong et al., 2004).

It is interesting to note that the majority of reported synthetic lethal interactions are between scaffold Nups and themselves or other classes, but only 4 (including Nup145c) scaffold Nups are singularly lethal when deleted. This illustrates the robust and redundant nature of the scaffold. While much of the FG-repeat domains can be deleted (Strawn et al., 2004), over half of the FG-Nups are required for survival. For example, all of the FG-repeats on Nup159 may be removed with little effect (Strawn et al., 2004), but deletion of the entire Nup cannot be tolerated (Giaever et al., 2002). This suggests that their anchoring domains play additional roles within the complex and are vital for proper NPC architecture and formation.

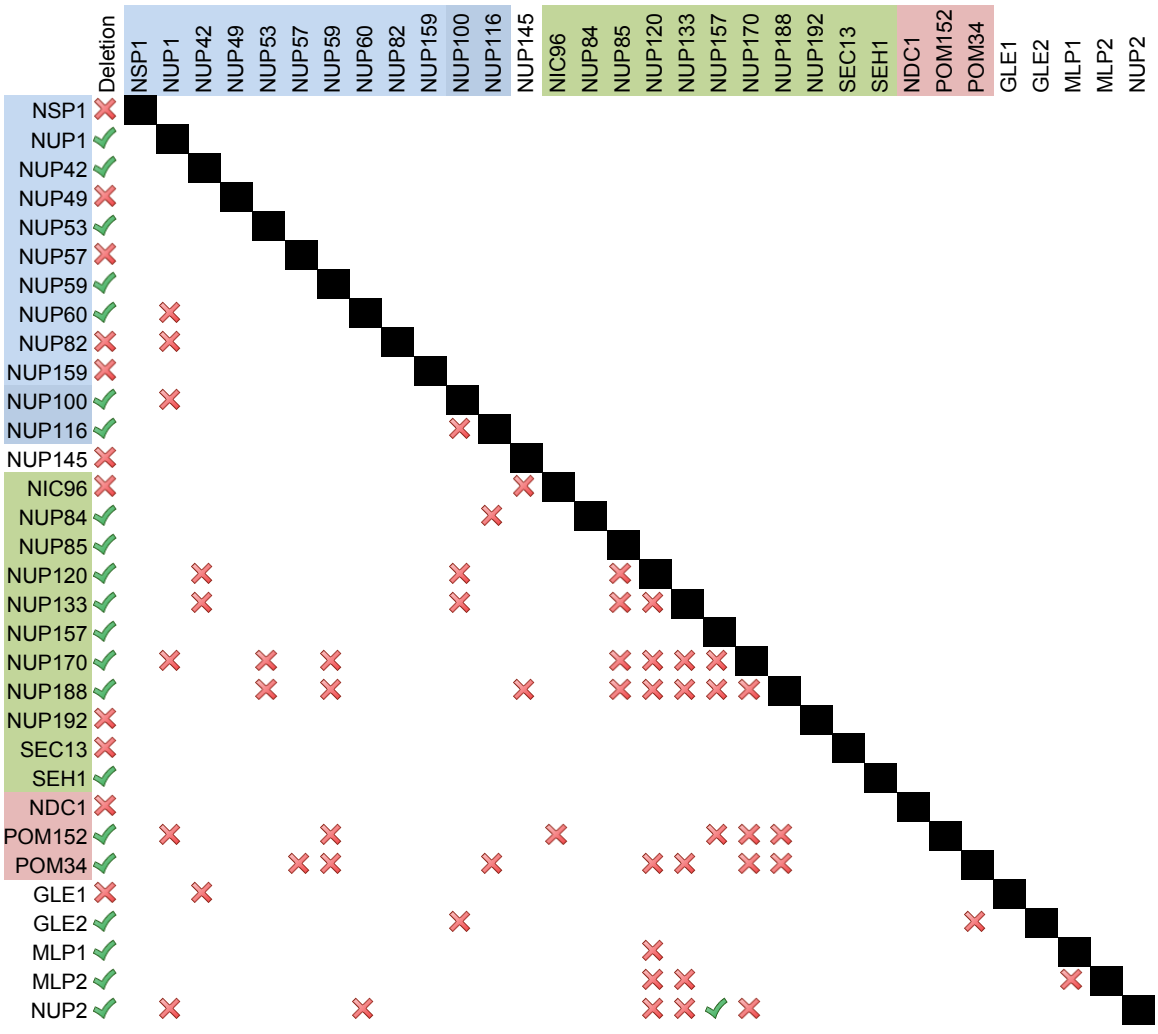


Figure 34: *S. cerevisiae* double mutation matrix. Colored in blue are FG-Nups and in green are the scaffold Nups. In pink are the transmembrane Nups. The first column, labeled “deletion” indicates which Nups causes cell death when knocked out.

APPENDIX E — PRIMER CATALOG FOR *IN SITU* GFP TAGGING

Table 9 lists the primers we used to insert the GFP epitope into various loci within the *T. brucei* genome. Primers are constructed from the 80 nucleotides which are homologous to the 3' terminal end of the gene of interest (capital letters in the following table). Lower case letters represent the primer region that is homologous with the pMOTag series of vectors. Also indicated are the masses of the proteins along with the additional mass from the GFP moiety.

Table 9: List of primers used for GFP epitope tagging in this study.

GeneDB Accesion	Mass (kDa)	MW+GFP (kDa)	Annotation	Forward Primer	Reverse Primer
Tb11.02.0460	89.1	116.6	TbNup89	GCATGATCTTTTAGTGGAAGAGGCGAAAAAGGTTGCTGTAACACAACAGCA GCTTTTTGTGACATCACAACCGAAATGGgggtaccgggcccccctcgag	TAAGGCCACCCACATCCTTCTGAGCTACATCAAGCGTACGAGCACCTCTT ACCGTCTTATCTCCACCCCACACGtggcggccgctctagaactagtggat
Tb10.6k15.3670	96.5	124.0	TbNup96	GGGTGCGTCGCTGGAAGAGGCATATGAACCGCTCTCTCTTAGATGAGCTA ATCGTTCTAGAGCGGTTGTTTGTGCAGgggtaccgggcccccctcgag	TCCCCTATTCGGTCATTCAACTCACTAGAGAAGGGGGGAGGGAGTGTTC ACCAGTATCGAAATATAACGCTtggcggccgctctagaactagtggat
Tb10.6k15.2350	144.3	171.8	TbNup144	CGGGCATTGAGCGCTGCGGAGGGATTGCTCGAACAGTGCAACCTTGCCCT GAATCGTGCCAGTTCCAGGTTTACTTTCggtaccgggcccccctcgag	CAGTTTTAGTCAATTAATATTCGCTGTCGACGCAGACAGGTCAACGCAGT GAAATAATCTATCACCGTACACtggcggccgctctagaactagtggat
Tb11.03.0140	158.3	185.8	TbNup158	TGTGAATGCGCCAAGCATTCGCACTCACTCGTCTCTGCAGGAAGCGCTTCA CCGTTTCAGTCGGGGGTTTACAAAGAGAggtaccgggcccccctcgag	TCCTTTTCATTCACCAACCCTCGTCACCTTCGCATCACCCTATATTTCCAA CCTCTGTTTGCCTTCCACTGTTCtggcggccgctctagaactagtggat
Tb09.160.0340	92.3	119.8	TbMlp2	GTAACTTGGGATAAGTTGGTTGAGCTCATTTACCCCGTTTAAAATCGGTGTC GTGGCACAGGAAGTAAAGCAGCCACAGgtaccgggcccccctcgag	TCTCATCCTTTAGAGCCTCTCTCCCCTCTTTCCTCTCCCCCCCCCCAAT TGCAGAGGGAGCATAATATATTTggcggccgctctagaactagtggat
Tb11.03.0810	109.7	137.2	TbMlp1	ATGCGACTACTGCACGTCAACAAGCAACTTGTGGAGAGAGTCAAAACCAGT CGAACTGAAGGAGAATCCCAGTCCAGTggtaccgggcccccctcgag	TACACGAATTGTCATACAACCTGACTAGCAGACGTAAGGCGCTACGAAC CTTTACTGTGGTTCAAACAAAAAtggcggccgctctagaactagtggat
Tb11.02.2120	48.4	75.9	TbNup48	GATGGCTGCACGTATGGGTCGTTGCTTGCTGTGGAAGCGGACCAGTC TGTGTCATTTGTTCCTGCGTATCATGCAggtaccgggcccccctcgag	TCTGTGCAAAAAAGGGGTGTGGCATCGTTCAATCTTCCCCCATCCTTC CGCCCGAATATCATCGTGGCCTCtggcggccgctctagaactagtggat
Tb927.8.8050	74.8	102.3	Hypothetical	AGCCAGCCGCCGATCGGGTTCAGTGCACCGACCGTAAACGCCAATTTGGG TGCCCGTAGGAGGACACCGAAGCGGTTCggtaccgggcccccctcgag	AAACAAACAAACACACACACACACACATATATATATATATATAT ATATGCGGCCAATTACATAAtggcggccgctctagaactagtggat
Tb927.4.1310	47.4	74.9	Hypothetical	GAGATCCGTGGCACTGAGGCGTCATTAGGATTATCACCAGAGCAGGAGGG TAATCTCGTGTCTATGACAAAAGAGTTTggtaccgggcccccctcgag	AGCAACATATGTCCCGACAAACCGTAACCCACACCTTCTTCCCTGGTCA CACGGCACACGGAACGCGTCACtggcggccgctctagaactagtggat
Tb927.4.590	88.0	115.5	Hypothetical	CTTACATGCAGCTTTGGTGTCCTTCCCTTGTGGCACGCTACTTTGCCATG CGTAAAGGTGTGAGACAACTGTGGCGAggtaccgggcccccctcgag	AAAACGTACACATACACACATCGCAACCAAGGCGACGACATGTTAATCCC GAGAAGTCAATATCTCTCGGGATggcggccgctctagaactagtggat
Tb10.70.1120	109.2	136.7	Hypothetical	ATTTGCAGTAACTGTGAAGCAAAGAATCTTGACGAAGCCGATGTTTGTGTT GATTGTGGCATTGCCAAATGCGACAAAggtaccgggcccccctcgag	AAAACGACCCACACGACTGACAACACGGTTAGAACTTCGGACGCTGGAC TTAATAGATAGATTTAATGAAATtggcggccgctctagaactagtggat
Tb11.01.2885	140.3	167.8	TbNup140 FG	TGTTTTTCAGTCTTCTTTTTCTAGTATTGCCGGCAATTATAAGGCCACA CCTGATTTTAAAAGTGTCTTTTCGCAGgggtaccgggcccccctcgag	TTAAGTATAATAACAAAAATATTTAAGCACCTCATTCACATGATTTTTTAGT TACCTCCCATAACCATAAACTTtggcggccgctctagaactagtggat
Tb11.01.2880	149.2	176.7	TbNup149 FG	GGAGAATTATAGGGGTAGAAATAAGTCCGGTGATAAAATAATGTGGGCTCA TTTGGAGAATGCTTTCACTAAACCCTATggtaccgggcccccctcgag	AAACCCACTAACAATTTATATCACCAGCGCTTGCAACACAGAAGGTGATTAG CATCAAAGTAACACTTCAATAAATtggcggccgctctagaactagtggat
Tb11.03.0140N	158.3	185.8	TbNup158 N837-GFP	CCGCTGCCAGAAAGAAGGTAACAGATTTGAGTCTTACGATGTGAACACGGG GACGTGGGTTTACGTGATGAATGTTGGCggtaccgggcccccctcgag	CCTTTTCATTCACCAACCCTCGTCACCTTCGCATCACCCTATATTTCCAAC CTCTGTTTGCCTTCCACTGTTCTtggcggccgctctagaactagtggat
Tb11.01.7630	108.7	136.2	TbNup109	CTGTGAGGCAATTTTGAAACGTAGCAAGAAGCCGTGCGGCGGTGTGAGAC CATGTGAGTATCATGACAGGGTGAGGAAAggtaccgggcccccctcgag	CTTATGGTCAGCTGTTTTTCTATTAACTTCTCCCTCTACCTCCACAAACCC AGTGATGTCGGACCAAACCGGCATggcggccgctctagaactagtggat
Tb927.7.2300	132.3	159.8	TbNup132	GTGTGCCGGTATTTTTGACAATGCTGAGCAACAGTTGTTGATGTCGTGGCT TCGAGCTCGCTCCGCTGGACTCGTGTCAggtaccgggcccccctcgag	AATGAGTAAATGTCGGTGGCTTGAAAATATACGGAAATTGCAAGTTATTT TAACGGTTTCCGTCACGGCGTGGTtggcggccgctctagaactagtggat
Tb927.6.1830	49.9	77.4	Hypothetical	GCACCGCAACATCCCCGGCGCCATGGAAGAAGGAAGGAACGTGGTGAAGAG GATGATGTTTTTGAGATGAATAGGCAACACAAggtaccgggcccccctcgag	TTTTGCCTTTTCCCCTTCGGTGTTTACTGTCGTTTCACCCTTTTCACGACTT ACCTTCCTTCATACGCAGTACTGtggcggccgctctagaactagtggat
Tb10.61.0160	42.9	70.4	TbNup43	CAAAGAGGGGCTATTAGCCTGCGCATCCTTCGATCAAAGTATTCGAGTTAT GCTCGTGACCAAGTTGTCGCTCGTACAAggtaccgggcccccctcgag	TCTCATGGCTCCACTCCATTTTGTCATCACTTGCAAATTGTGGTAACCTA CTGTCGTCATACCTGTTTCACCACtggcggccgctctagaactagtggat
Tb10.70.1110	59.9	87.4	Hypothetical	GGAGTGCCACAAATGTGGAGCAACGAAGACGGTAGCGGTTGCCCCGCGG CGTGCTTTTGTGCGGAAGCCCGTCAAAATTggtaccgggcccccctcgag	ACAACGTCTATCCAGCGTGATAGTTATTTGCCGCTGTTATAGCAAGCAAA GTTAAATAAACGAATAGGAATGTCtggcggccgctctagaactagtggat
Tb10.70.1130	48.2	75.7	Hypothetical	TGTTGGCAGAGTGGAGAGGCAGATGACATTGGATCTTGCACAACGTTTGC GCAACGACATTCTTGATGAAACAAAGAAAggtaccgggcccccctcgag	CTGCTGGGAGATCTTGTTTGGTAAGTTCAGATTAGAAACGACATAATACT CCACATTTGTATTTGCAAAGGAATggcggccgctctagaactagtggat
Tb927.2.4230	406.8	434.3	TbNUP-1	CACAGCGACAGGTACGGCAAGTCATGGACATACGTAGCACAAGGAAAAGG TCTCGTTCAGCCAATGCGGTCTCGgtaccgggcccccctcgag	TCTAGGTGCATGTGTAGATGAACTGCACACTTTATGCACTAATAACAGGT TTGAAGTACTTACCTGGCATCTCctggcggccgctctagaactagtggat
Tb927.4.2070	511.3	538.8	Hypothetical	AAAGTGTAATGACTTAGTTATAAAGAGACTAGAGGATTGAGGTTAAAGCTCTT CGTGAAGAACTGCGTGGAAATGAGGCGggtaccgggcccccctcgag	GTAGCAATAATACGTATTTAAAAATGTCAAAATTGTCAGCAACAAAGATGC TTACACGAACAGAAAAAAAAAGAtggcggccgctctagaactagtggat
Tb927.7.3330	502.6	530.1	Hypothetical	GGAAGGTGAAGACGATGAGGACGAGGGCGACGCAACCGGTTGCCCAACA ACGCATTTGGGAGGGCCATGGGCGCACCATggtaccgggcccccctcgag	TACATTAGACATCATTCATCGACTGTAACCTAGGTAGTGTGTATGAGATACC GTATCAATTACACACTGAGTGTCAtggcggccgctctagaactagtggat

REFERENCES

Adl, S.M., A.G.B. Simpson, M.A. Farmer, R.A. Andersen, O.R. Anderson, J.R. Barta, S.S. Bowser, G. Brugerolle, R.A. Fensome, S. Fredericq, T.Y. James, S. Karpov, P. Kugrens, J. Krug, C.E. Lane, L.A. Lewis, J. Lodge, D.H. Lynn, D.G. Mann, R.M. McCourt, L. Mendoza, O. Moestrup, S.E. Mozley-Standridge, T.A. Nerad, C.A. Shearer, A.V. Smirnov, F.W. Spiegel, and M. Taylor. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal Of Eukaryotic Microbiology*. 52:399-451.

Akey, C.W., and M. Radermacher. 1993. Architecture Of The Xenopus Nuclear-Pore Complex Revealed By 3-Dimensional Cryoelectron Microscopy. *Journal Of Cell Biology*. 122:1-19.

Akker, S.A., P.J. Smith, and S.L. Chew. 2001. Nuclear post-transcriptional control of gene expression. *Journal of Molecular Endocrinology*. 27:123-131.

Alber, F., S. Dokudovskaya, L.M. Veenhoff, W. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, A. Sali, and M.P. Rout. 2007a. The molecular architecture of the nuclear pore complex. *Nature*. 450:695.

Alber, F., S. Dokudovskaya, L.M. Veenhoff, W.H. Zhang, J. Kipper, D. Devos, A. Suprapto, O. Karni-Schmidt, R. Williams, B.T. Chait, A. Sali, and M.P. Rout. 2007b. The molecular architecture of the nuclear pore complex. *Nature*. 450:695-701.

Alibu, V.P., L. Storm, S. Haile, C. Clayton, and D. Horn. 2005. A doubly inducible system for RNA interference and rapid RNAi plasmid construction in Trypanosoma brucei. *Molecular and Biochemical Parasitology*. 139:75-82.

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

Andersen, J.S., Y.W. Lam, A.K.L. Leung, S.E. Ong, C.E. Lyon, A.I. Lamond, and M. Mann. 2005. Nucleolar proteome dynamics. *Nature*. 433:77-83.

Andersen, J.S., and M. Mann. 2006. Organellar proteomics: turning inventories into insights. *Embo Reports*. 7:874-879.

Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting. 2001a. Protein repeats: Structures, functions, and evolution. *Journal Of Structural Biology*. 134:117-131.

Andrade, M.A., C. Petosa, S.I. O'Donoghue, C.W. Muller, and P. Bork. 2001b. Comparison of ARM and HEAT protein repeats. *Journal Of Molecular Biology*. 309:1-18.

Andrulis, E.D., D.C. Zappulla, A. Ansari, S. Perrod, C.V. Laiosa, M.R. Gartenberg, and R. Sternglanz. 2002. Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Molecular And Cellular Biology*. 22:8292-8301.

Archambault, V., C.H.X. Li, A.J. Tackett, R. Wasch, B.T. Chait, M.P. Rout, and F.R. Cross. 2003. Genetic and biochemical evaluation of the importance of Cdc6 in regulating mitotic exit. *Molecular Biology Of The Cell*. 14:4592-4604.

Ashford, R., and W. Crewe. 1998. The Parasites of Homo Sapiens: An Annotated Checklist of the Protozoa, Helminths and Arthropods for Which We Are Home. Liverpool School of Tropical Medicine, Liverpool. 128 pp.

Atwood, J.A., D.B. Weatherly, T.A. Minning, B. Bundy, C. Cavola, F.R. Opperdoes, R. Orlando, and R.L. Tarleton. 2005. The Trypanosoma cruzi proteome. *Science*. 309:473-476.

Bapteste, E., R.L. Charlebois, D. Macleod, and C. Brochier. 2005. The two tempos of nuclear pore complex evolution: highly adapting proteins in an ancient frozen structure. *Genome Biology*. 6.

Bastin, P., K. Ellis, L. Kohl, and K. Gull. 2000. Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. *Journal of Cell Science*. 113:3321-3328.

Bateman, A., E. Birney, R. Durbin, S.R. Eddy, K.L. Howe, and E.L.L. Sonnhammer. 2000. The Pfam protein families database. *Nucleic Acids Research*. 28:263-266.

Bayliss, R., A.H. Corbett, and M. Stewart. 2000a. The molecular mechanism of transport of macromolecules through nuclear pore complexes. *Traffic*. 1:448-456.

Bayliss, R., S.W. Leung, R.P. Baker, B.B. Quimby, A.H. Corbett, and M. Stewart. 2002. Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *Embo Journal*. 21:2843-2853.

Bayliss, R., T. Littlewood, and M. Stewart. 2000b. Structural basis for the interaction between FxFG nucleoporin repeats and importin-beta in nuclear trafficking. *Cell*. 102:99-108.

Beck, M., F. Forster, M. Ecke, J.M. Plitzko, F. Melchior, G. Gerisch, W. Baumeister, and O. Medalia. 2004. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science*. 306:1387-1390.

Beck, M., V. Lucic, F. Forster, W. Baumeister, and O. Medalia. 2007. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature*. 449:611-615.

Belgareh, N., G. Rabut, S.W. Bai, M. van Overbeek, J. Beaudouin, N. Daigle, O.V. Zatsepina, F. Pasteau, V. Labas, M. Fromont-Racine, J. Ellenberg, and V. Doye. 2001. An evolutionarily conserved NPC subcomplex, which redistributes in part to kinetochores in mammalian cells. *Journal Of Cell Biology*. 154:1147-1160.

Berke, I.C., T. Boehmer, G. Blobel, and T.U. Schwartz. 2004. Structural and functional analysis of Nup133 domains reveals modular building blocks of the nuclear pore complex. *Journal of Cell Biology*. 167:591-597.

Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D.C. Bartholomeu, N.J. Lennard, E. Caler, N.E. Hamlin, B. Haas, W. Bohme, L. Hannick, M.A. Aslett, J. Shallom, L. Marcello, L.H. Hou, B. Wickstead, U.C.M. Alsmark, C. Arrowsmith, R.J. Atkin, A.J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T.J. Chillingworth, C. Churcher, L.N. Clark, C.H. Corton, A. Cronin, R.M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M.C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B.R. Harris, H. Hauser, J. Hostetter, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A.X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. MacLeod, P.J. Mooney, S. Moule, D.M.A. Martin, G.W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C.S. Peacock, J. Peterson, M.A. Quail, E. Rabbinowitsch, M.A. Rajandream, C. Reitter, S.L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A.J. Simpson, L. Talton, C.M.R. Turner, A. Tait, A.R. Tivey, S. Van Aken, D. Walker, D. Wanless, S.L. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M.D. Adams, T.M. Embley, K. Gull, E. Ullu, J.D. Barry, A.H. Fairlamb, F. Opperdoes, B.G. Barret, J.E. Donelson, N. Hall, C.M. Fraser, et al. 2005. The genome of the African trypanosome Trypanosoma brucei. *Science*. 309:416-422.

Beverley, S.M. 2003. Protozomics: Trypanosomatid parasite genetics comes of age. *Nature Reviews Genetics*. 4:11-19.

Bharathi, A., A. Ghosh, W.A. Whalen, J.H. Yoon, R. Pu, M. Dasso, and R. Dhar. 1997. The human RAE1 gene is a functional homologue of Schizosaccharomyces pombe rae1 gene involved in nuclear export of Poly(A)(+) RNA. *Gene*. 198:251-258.

Blower, M.D., M. Nachury, R. Heald, and K. Weis. 2005. A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. *Cell*. 121:223-234.

Boer, J., J. Bonten-Surtel, and G. Grosveld. 1998. Overexpression of the nucleoporin CAN/NUP214 induces growth arrest, nucleocytoplasmic transport defects, and apoptosis. *Molecular And Cellular Biology*. 18:1236-1247.

Borst, P. 2002. Antigenic variation and allelic exclusion. *Cell*. 109:5-8.

Boutros, M., A.A. Kiger, S. Armknecht, K. Kerr, M. Hild, B. Koch, S.A. Haas, R. Paro, and N. Perrimon. 2004. Genome-wide RNAi analysis of growth and viability in Drosophila cells. *Science*. 303:832-835.

Broadhead, R., H.R. Dawe, H. Farr, S. Griffiths, S.R. Hart, N. Portman, M.K. Shaw, M.L. Ginger, S.J. Gaskell, P.G. McKean, and K. Gull. 2006. Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature*. 440:224-227.

Bullock, T.L., W.D. Clarkson, H.M. Kent, and M. Stewart. 1996. The 1.6 angstrom resolution crystal structure of nuclear transport factor 2 (NTF2). *Journal Of Molecular Biology*. 260:422-431.

Callan, H.G., and S.G. Tomlin. 1950. Experimental Studies on Amphibian Oocyte Nuclei.1. Investigation of the Structure of the Nuclear Membrane by Means of the Electron Microscope. *Proceedings of the Royal Society of London Series B-Biological Sciences*. 137:367.

Campbell, D.A., S. Thomas, and N.R. Sturm. 2003. Transcription in kinetoplastid protozoa: why be normal? *Microbes And Infection*. 5:1231-1240.

Campbell, M.S., G.K.T. Chan, and T.J. Yen. 2001. Mitotic checkpoint proteins HsMAD1 and HsMAD2 are associated with nuclear pore complexes in interphase. *Journal Of Cell Science*. 114:953-963.

Casolari, J.M., C.R. Brown, S. Komili, J. West, H. Hieronymus, and P.A. Silver. 2004. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*. 117:427-439.

Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of protozoa. *International Journal Of Systematic And Evolutionary Microbiology*. 52:297-354.

Chalfie, M., Y. Tu, G. Euskirchen, W.W. Ward, and D.C. Prasher. 1994. Green Fluorescent Protein As A Marker For Gene-Expression. *Science*. 263:802-805.

Chi, J.T., H.Y. Chang, N.N. Wang, D.S. Chang, N. Dunphy, and P.O. Brown. 2003. Genomewide view of gene silencing by small interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America*. 100:6343-6346.

Clayton, C.E. 2002. Life without transcriptional control? From fly to man and back again. *Embo Journal*. 21:1881-1888.

Cohen, M., K.K. Lee, K.L. Wilson, and Y. Gruenbaum. 2001. Transcriptional repression, apoptosis, human disease and the functional evolution of the nuclear lamina. *Trends In Biochemical Sciences*. 26:41-47.

Colasante, C., M. Ellis, T. Ruppert, and F. Voncken. 2006. Comparative proteomics of glycosomes from bloodstream form and procyclic culture form Trypanosoma brucei brucei. *Proteomics*. 6:3275-3293.

Collins, S.R., K.M. Miller, N.L. Maas, A. Roguev, J. Fillingham, C.S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H.M. Ding, H. Xu, J.H. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S.L. Berger, P. Hieter, Z.G. Zhang, G.W. Brown, C.J. Ingles, A. Emili, C.D. Allis, D.P. Toczyski, J.S. Weissman, J.F. Greenblatt, and N.J. Krogan. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature*. 446:806-810.

Cox, F.E.G. 2002. History of human parasitology. *Clinical Microbiology Reviews*. 15:595-+.

Craig, R., and R.C. Beavis. 2004. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 20:1466-1467.

Cristea, I.M., R. Williams, B.T. Chait, and M.P. Rout. 2005. Fluorescent proteins as proteomic probes. *Molecular & Cellular Proteomics*. 4:1933-1941.

Cronshaw, J.A., A.N. Krutchinsky, W.Z. Zhang, B.T. Chait, and M.J. Matunis. 2002. Proteomic analysis of the mammalian nuclear pore complex. *Journal of Cell Biology*. 158:915-927.

Cronshaw, J.M., and M.J. Matunis. 2003. The nuclear pore complex protein ALADIN is mislocalized in triple A syndrome. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 100:5823-5827.

D'Angelo, M.A., D.J. Anderson, E. Richard, and M.W. Hetzer. 2006. Nuclear pores form de novo from both sides of the nuclear envelope. *Science*. 312:440-443.

Das, A., and V. Bellofatto. 2003. RNA polymerase II-dependent transcription in trypanosomes is associated with a SNAP complex-like transcription factor.

Proceedings of the National Academy of Sciences of the United States of America. 100:80-85.

Das, A., Q. Zhang, J.B. Palenchar, B. Chatterjee, G.A.M. Cross, and V. Bellofatto. 2005. Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Molecular And Cellular Biology*. 25:7314-7322.

De Souza, C.P.C., A.H. Osmani, S.B. Hashmi, and S. Osmani. 2004. Partial nuclear pore complex disassembly during closed mitosis in Aspergillus nidulans. *Current Biology*. 14:1973-1984.

Denning, D.P., S.S. Patel, V. Uversky, A.L. Fink, and M. Rexach. 2003. Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 100:2450-2455.

Denning, D.P., and M.F. Rexach. 2007. Rapid evolution exposes the boundaries of domain structure and function in natively unfolded FG nucleoporins. *Molecular & Cellular Proteomics*. 6:272-282.

Despommier, D.D., et.al. 1995. Parasitic diseases, New York. 196-203.

Devos, D., S. Dokudovskaya, F. Alber, R. Williams, B.T. Chait, A. Sali, and M.P. Rout. 2004. Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture. *PLoS Biology*. 2:e380.

Devos, D., S. Dokudovskaya, R. Williams, F. Alber, N. Eswar, B.T. Chait, M.P. Rout, and A. Sali. 2006. Simple fold composition and modular architecture of the nuclear pore complex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 103:2172-2177.

Dilworth, D.J., A.J. Tackett, R.S. Rogers, E.C. Yi, R.H. Christmas, J.J. Smith, A.F. Siegel, B.T. Chait, R.W. Wozniak, and J.D. Aitchison. 2005. The mobile nucleoporin Nup2p and chromatin-bound Prp20p function in endogenous NPC-mediated transcriptional control. *Journal Of Cell Biology*. 171:955-965.

Ding, D.Q., Y. Tomita, A. Yamamoto, Y. Chikashige, T. Haraguchi, and Y. Hiraoka. 2000. Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes To Cells*. 5:169-190.

Dingwall, C., and R.A. Laskey. 1991. Nuclear Targeting Sequences - A Consensus. *Trends In Biochemical Sciences*. 16:478-481.

Doolittle, W.F. 1980. Revolutionary Concepts In Evolutionary Cell Biology. *Trends In Biochemical Sciences*. 5:146-149.

Dreger, M. 2003. Subcellular proteomics. *Mass Spectrometry Reviews*. 22:27-56.

Dreger, M., L. Bengtsson, T. Schoneberg, H. Otto, and F. Hucho. 2001. Nuclear envelope proteomics: Novel integral membrane proteins of the inner nuclear membrane. *Proceedings of the National Academy of Sciences of the United States of America*. 98:11943-11948.

Eddy, S.R. 1998. Profile hidden Markov models. *Bioinformatics*. 14:755-763.

El-Sayed, N.M., P.J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E.A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D.C. Bartholomeu, B.J. Haas, A.N. Tran, J.R. Wortman, U.C.M. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J.M. Carlton, G.C. Cerqueira, T. Creasy, A.L. Delcher, A. Djikeng, T.M. Embley, C. Hauser, A.C. Ivens, S.K. Kummerfeld, J.B. Pereira-Leal, D. Nilsson, J. Peterson, S.L. Salzberg, J. Shallom, J.C. Silva, J. Sundaram, S. Westenberger, O. White, S.E. Metville, J.E. Donelson, B. Andersson, K.D. Stuart, and N. Hall. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 309:404-409.

Embley, T.M., and W. Martin. 2006. Eukaryotic evolution, changes and challenges. *Nature*. 440:623-630.

Enninga, J., A. Levay, and B.M.A. Fontoura. 2003. Sec13 shuttles between the nucleus and the cytoplasm and stably interacts with Nup96 at the nuclear pore complex. *Molecular And Cellular Biology*. 23:7271-7284.

Eriksson, J., and D. Fenyo. 2007. Improving the success rate of proteome analysis by modeling protein-abundance distributions and experimental designs. *Nature Biotechnology*. 25:651-655.

Erkmann, J.A., and U. Kutay. 2004. Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Experimental Cell Research*. 296:12-20.

Ersfeld, K., S.E. Melville, and K. Gull. 1999. Nuclear and genome organization of Trypanosoma brucei. *Parasitology Today*. 15:58-63.

Fahrenkrog, B., and U. Aebi. 2003. The nuclear pore complex: Nucleocytoplasmic transport and beyond. *Nature Reviews Molecular Cell Biology*. 4:757-766.

Fahrenkrog, B., J. Koser, and U. Aebi. 2004. The nuclear pore complex: a jack of all trades? *Trends in Biochemical Sciences*. 29:175-182.

Fahrenkrog, B., B. Maco, A.M. Fager, J. Koser, U. Sauder, K.S. Ullman, and U. Aebi. 2002. Domain-specific antibodies reveal multiple-site topology of Nup153 within the nuclear pore complex. *Journal Of Structural Biology*. 140:254-267.

Feldherr, C., D. Akin, T. Littlewood, and M. Stewart. 2002. The molecular mechanism of translocation through the nuclear pore complex is highly conserved. *Journal Of Cell Science*. 115:2997-3005.

Feldherr, C.M. 1962. Nuclear Annuli As Pathways For Nucleocytoplasmic Exchanges. *Journal Of Cell Biology*. 14:65-&.

Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse. 1989. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science*. 246:64-71.

Feuerbach, F., V. Galy, E. Trelles-Sticken, M. Fromont-Racine, A. Jacquier, E. Gilson, J.C. Olivo-Marin, H. Scherthan, and U. Nehrbass. 2002. Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast. *Nature Cell Biology*. 4:214-221.

Fevre, E.M., K. Picozzi, J. Jannin, S.C. Welburn, and I. Maudlin. 2006. Human African trypanosomiasis: Epidemiology and control. *In* Advances In Parasitology, Vol 61. Vol. 61. 167-+.

Field, H.I., D. Fenyo, and R.C. Beavis. 2002. RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database. *Proteomics*. 2:36-47.

Fire, A., S.Q. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*. 391:806-811.

Fontoura, B.M.A., G. Blobel, and M.J. Matunis. 1999. A conserved biogenesis pathway for nucleoporins: Proteolytic processing of a 186-kilodalton precursor generates Nup98 and the novel nucleoporin, Nup96. *Journal of Cell Biology*. 144:1097-1112.

Foster, H.A., and J.M. Bridger. 2005. The genome and the nucleus: a marriage made by evolution - Genome organisation and nuclear architecture. *Chromosoma*. 114:212-229.

Galy, V., O. Gadal, M. Fromont-Racine, A. Romano, A. Jacquier, and U. Nehrbass. 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell*. 116:63-73.

Galy, V., I.W. Mattaj, and P. Askjaer. 2003. Caenorhabditis elegans nucleoporins Nup93 and Nup205 determine the limit of nuclear pore complex size exclusion in vivo. *Molecular Biology of the Cell*. 14:5104-5115.

Galy, V., J.C. Olivo-Marin, H. Scherthan, V. Doye, N. Rascalou, and U. Nehrbass. 2000. Nuclear pore complexes in the organization of silent telomeric chromatin. *Nature*. 403:108-112.

Garcia, A., D. Courtin, P. Solan, M. Koffi, and V. Jamonnea. 2006. Human African trypanosomiasis: connecting parasite and host genetics. *Trends In Parasitology*. 22:405-409.

Gavin, A.C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*. 415:141-147.

Gerace, L., and B. Burke. 1988. Functional-Organization Of The Nuclear-Envelope. *Annual Review Of Cell Biology*. 4:335-374.

Gerhart, J., M. Kirschner. 1997. Cells, embryos, and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Blackwell Science, New York. 642 pp.

Giaever, G., A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A.P. Arkin, A. Astromoff, M. El Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.D. Entian, P. Flaherty, F. Foury, D.J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J.H. Hegemann, S. Hempel, Z. Herman, D.F. Jaramillo, D.E. Kelly, S.L. Kelly, P. Kotter, D. LaBonte, D.C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C.Y. Luo, M. Lussier, R. Mao, P. Menard, S.L. Ooi, J.L. Revuelta, C.J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D.D. Shoemaker, S. Sookhai-Mahadeo, R.K. Storms, J.N. Strathern, G. Valle, M. Voet, G. Volckaert, C.Y. Wang, T.R. Ward, J. Wilhelmy, E.A. Winzeler, Y.H. Yang, G. Yen, E. Youngman, K.X. Yu, H. Bussey, J.D. Boeke, M. Snyder, P. Philippsen, R.W. Davis, and M. Johnston. 2002. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*. 418:387-391.

Gilinger, G., and V. Bellofatto. 2001. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Research*. 29:1556-1564.

Glavy, J.S., A.N. Krutchinsky, I.M. Cristea, I.C. Berke, T. Boehmer, G. Blobel, and B.T. Chait. 2007. Cell-cycle-dependent phosphorylation of the nuclear pore Nup107-160 subcomplex. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 104:3811-3816.

Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S.G. Oliver. 1996. Life with 6000 genes. *Science*. 274:546.

Gorbunoff, M.J. 1985. Protein Chromatography On Hydroxyapatite Columns. *Methods In Enzymology*. 117:370-380.

Gorlich, D., and U. Kutay. 1999. Transport between the cell nucleus and the cytoplasm. *Annual Review Of Cell And Developmental Biology*. 15:607-660.

Grant, R.P., E. Hurt, D. Neuhaus, and M. Stewart. 2002. Structure of the C-terminal FG-nucleoporin binding domain of Tap/NXF1. *Nature Structural Biology*. 9:247-251.

Grant, R.P., D. Neuhaus, and M. Stewart. 2003. Structural basis for the interaction between the Tap/NXF1 UBA domain and FG nucleoporins at 1 angstrom resolution. *Journal Of Molecular Biology*. 326:849-858.

Hanash, S. 2003. Disease proteomics. *Nature*. 422:226-232.

Hannon, G.J. 2002. RNA interference. *Nature*. 418:244-251.

Harel, A., A.V. Orjalo, T. Vincent, A. Lachish-Zalait, S. Vasu, S. Shah, E. Zimmerman, M. Elbaum, and D.J. Forbes. 2003. Removal of a single pore subcomplex results in vertebrate nuclei devoid of nuclear pores. *Molecular Cell*. 11:853-864.

Hawkins, J., L. Davis, and M. Boden. 2007. Predicting nuclear localization. *Journal Of Proteome Research*. 6:1402-1409.

He, C.Y., H.H. Ho, H. Malsam, C. Chalouni, C.M. West, E. Ullu, D. Toomre, and G. Warren. 2004. Golgi duplication in Trypanosoma brucei. *Journal Of Cell Biology*. 165:313-321.

Hedges, S.B. 2002. The origin and evolution of model organisms. *Nature Reviews Genetics*. 3:838-849.

Hediger, F., K. Dubrana, and S.M. Gasser. 2002a. Myosin-like proteins 1 and 2 are not required for silencing or telomere anchoring, but act in the Tel1 pathway of telomere length control. *Journal Of Structural Biology*. 140:79-91.

Hediger, F., F.R. Neumann, G. Van Houwe, K. Dubrana, and S.M. Gasser. 2002b. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Current Biology*. 12:2076-2089.

Hodel, A.E., M.R. Hodel, E.R. Griffis, K.A. Hennig, G.A. Ratner, S.L. Xu, and M.A. Powers. 2002. The three-dimensional structure of the autoproteolytic, nuclear pore-targeting domain of the human nucleoporin Nup98. *Molecular Cell*. 10:347-358.

Iouk, T., O. Kerscher, R.J. Scott, M.A. Basrai, and R.W. Wozniak. 2002. The yeast nuclear pore complex functionally interacts with components of the spindle assembly checkpoint. *Journal Of Cell Biology*. 159:807-819.

Ishii, K., G. Arib, C. Lin, G. Van Houwe, and U.K. Laemmli. 2002. Chromatin boundaries in budding yeast: The nuclear pore connection. *Cell*. 109:551-562.

Jeffery, D.A., and M. Bogyo. 2003. Chemical proteomics and its application to drug discovery. *Current Opinion in Biotechnology*. 14:87-95.

Jeganathan, K.B., L. Malureanu, and J.M. van Deursen. 2005. The Rae1-Nup98 complex prevents aneuploidy by inhibiting securin degradation. *Nature*. 438:1036-1039.

Jeudy, S., and T.U. Schwartz. 2007. Crystal structure of nucleoporin Nic96 reveals a novel, intricate helical domain architecture. *J. Biol. Chem.*:M705479200.

Jones, A., A. Faldas, A. Foucher, E. Hunt, A. Tait, J.M. Wastling, and C.M. Turner. 2006. Visualisation and analysis of proteomic data from the procyclic form of Trypanosoma brucei. *Proteomics*. 6:259-267.

Joseph, J., S.T. Liu, S.A. Jablonski, T.J. Yen, and M. Dasso. 2004. The RanGAP1-RanBP2 complex is essential for microtubule-kinetochore interactions in vivo. *Current Biology*. 14:611-617.

Joseph, J., S.H. Tan, T.S. Karpova, J.G. McNally, and M. Dasso. 2002. SUMO-1 targets RanGAP1 to kinetochores and mitotic spindles. *Journal Of Cell Biology*. 156:595-602.

Kall, L., A. Krogh, and E.L.L. Sonnhammer. 2004. A combined transmembrane topology and signal peptide prediction method. *Journal Of Molecular Biology*. 338:1027-1036.

Kamath, R.S., A.G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D.P. Welchman, P. Zipperlen, and J. Ahringer. 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*. 421:231-237.

Karas, M., and F. Hillenkamp. 1988. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10000 Daltons. *Analytical Chemistry*. 60:2299-2301.

Kastenmayer, J.P., M.S. Lee, A.L. Hong, F.A. Spencer, and M.A. Basrai. 2005. The C-terminal half of Saccharomyces cerevisiae Mad1p mediates spindle checkpoint function, chromosome transmission fidelity and CEN association. *Genetics*. 170:509-517.

Keeling, P.J., G. Burger, D.G. Durnford, B.F. Lang, R.W. Lee, R.E. Pearlman, A.J. Roger, and M.W. Gray. 2005. The tree of eukaryotes. *Trends In Ecology & Evolution*. 20:670-676.

Keminer, O., and R. Peters. 1999. Permeability of single nuclear pores. *Biophysical Journal*. 77:217-228.

Kerscher, O., P. Hieter, M. Winey, and M.A. Basrai. 2001. Novel role for a Saccharomyces cerevisiae nucleoporin, nup170p, in chromosome segregation. *Genetics*. 157:1543-1553.

Krull, S., J. Thyberg, B. Bjorkroth, H.R. Rackwitz, and V.C. Cordes. 2004. Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Molecular Biology of the Cell*. 15:4261-4277.

Krutchinsky, A.N., M. Kalkum, and B.T. Chait. 2001. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Analytical Chemistry*. 73:5066-5077.

Krutchinsky, A.N., W.Z. Zhang, and B.T. Chait. 2000. Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *Journal of the American Society for Mass Spectrometry*. 11:493-504.

LaCount, D.J., S. Bruse, K.L. Hill, and J.E. Donelson. 2000. Double-stranded RNA interference in Trypanosoma brucei using head-to-head promoters. *Molecular And Biochemical Parasitology*. 111:67-76.

Lange, A., R.E. Mills, C.J. Lange, M. Stewart, S.E. Devine, and A.H. Corbett. 2007. Classical nuclear localization signals: Definition, function, and interaction with importin alpha. *Journal Of Biological Chemistry*. 282:5101-5105.

Latzer, J., G.A. Papoian, M.C. Prentiss, E.A. Komives, and P.G. Wolynes. 2007. Induced fit, folding, and recognition of the NF-kappa B-nuclear localization signals by I kappa B alpha and I kappa B beta. *Journal Of Molecular Biology*. 367:262-274.

Lee, S.J., Y. Matsuura, S.M. Liu, and M. Stewart. 2005. Structural basis for nuclear import complex dissociation by RanGTP. *Nature*. 435:693-696.

Liang, X.H., A. Haritan, S. Uliel, and S. Michaeli. 2003. trans and cis splicing in trypanosomatids: Mechanism, factors, and regulation. *Eukaryotic Cell*. 2:830-840.

Lim, R.Y.H., U. Aebi, and D. Stoffler. 2006. From the trap to the basket: getting to the bottom of the nuclear pore complex. *Chromosoma*. 115:15-26.

Lim, R.Y.H., and B. Fahrenkrog. 2006. The nuclear pore complex up close. *Current Opinion In Cell Biology*. 18:342-347.

Lindsay, M.R., R.I. Webb, M. Strous, M.S. Jetten, M.K. Butler, R.J. Forde, and J.A. Fuerst. 2001. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Archives Of Microbiology*. 175:413-429.

Liu, S.M., and M. Stewart. 2005. Structural basis for the high-affinity binding of nucleoporin Nup1p to the Saccharomyces cerevisiae importin-beta homologue, Kap95p. *Journal Of Molecular Biology*. 349:515-525.

Loeillet, S., B. Palancade, M. Cartron, A. Thierry, G.F. Richard, B. Dujon, V. Doye, and A. Nicolas. 2005. Genetic network interactions among replication, repair and nuclear pore deficiencies in yeast. *Dna Repair*. 4:459-468.

Loiodice, I., A. Alves, G. Rabut, M. van Overbeek, J. Ellenberg, J.B. Sibarita, and V. Doye. 2004. The entire Nup107-160 complex, including three new members, is targeted as one entity to kinetochores in Mitosis. *Molecular Biology Of The Cell*. 15:3333-3344.

Lupas, A., M. Vandyke, and J. Stock. 1991. Predicting Coiled Coils From Protein Sequences. *Science*. 252:1162-1164.

Madrid, A.S., and K. Weis. 2006. Nuclear transport is becoming crystal clear. *Chromosoma*. 115:98-109.

Maeshima, K., K. Yahata, Y. Sasaki, R. Nakatomi, T. Tachibana, T. Hashikawa, F. Imamoto, and N. Imamoto. 2006. Cell-cycle-dependent dynamics of nuclear pores: pore-free islands and lamins. *Journal Of Cell Science*. 119:4442-4451.

Makhnevych, T., C.P. Lusk, A.M. Anderson, J.D. Aitchison, and R.W. Wozniak. 2003. Cell cycle regulated transport controlled by alterations in the nuclear pore complex. *Cell*. 115:813-823.

Mans, B.J., V. Anantharaman, L. Aravind, and E.V. Koonin. 2004. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*. 3:1612-1637.

Marelli, M., C.P. Lusk, H. Chan, J.D. Aitchison, and R.W. Wozniak. 2001. A link between the synthesis of nucleoporins and the biogenesis of the nuclear envelope. *Journal Of Cell Biology*. 153:709-723.

Martin, W. 1999. A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. *Proceedings Of The Royal Society Of London Series B-Biological Sciences*. 266:1387-1395.

Martin, W. 2005. Archaebacteria (Archaea) and the origin of the eukaryotic nucleus. *Current Opinion In Microbiology*. 8:630-637.

Matsuura, Y., A. Lange, M.T. Harreman, A.H. Corbett, and M. Stewart. 2003. Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import. *Embo Journal*. 22:5358-5369.

Matsuura, Y., and M. Stewart. 2004a. Structural basis for the assembly of a nuclear export complex. *Nature*. 432:872-877.

Matsuura, Y., and M. Stewart. 2004b. Structural basis for the nuclear export of Kap60p by Cse1p: RanGTP. *Molecular Biology of the Cell*. 15:205A-205A.

Matsuura, Y., and M. Stewart. 2005. Structural basis for the nuclear export of Kap60p by Cse1p and RanGTP. *Biophysical Journal*. 88:185A-185A.

Maul, G.G. 1971. Octagonality Of Nuclear Pore Complex. *Journal Of Cell Biology*. 51:558.

McGuffin, L.J., K. Bryson, and D.T. Jones. 2000. The PSIPRED protein structure prediction server. *Bioinformatics*. 16:404-405.

Melcak, I., A. Hoelz, and G. Blobel. 2007. Structure of Nup58/45 suggests flexible nuclear pore diameter by intermolecular sliding. *Science*. 315:1729-1732.

Mendjan, S., M. Taipale, J. Kind, H. Holz, P. Gebhardt, M. Schelder, M. Vermeulen, A. Buscaino, K. Duncan, J. Mueller, M. Wilm, H.G. Stunnenberg, H. Saumweber, and A. Akhtar. 2006. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in Drosophila. *Molecular Cell*. 21:811-823.

Menon, B.B., N.J. Sarma, S. Pasula, S.J. Deminoff, K.A. Willis, K.E. Barbara, B. Andrews, and G.M. Santangelo. 2005. Reverse recruitment: The Nup84 nuclear pore subcomplex mediates Rap1/Gcr1/Gcr2 transcriptional

activation. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 102:5749-5754.

Morrison, J., J.C. Yang, M. Stewart, and D. Neuhaus. 2003. Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats. *Journal Of Molecular Biology*. 333:587-603.

Mosammaparast, N., and L.F. Pemberton. 2004. Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends In Cell Biology*. 14:547-556.

Moy, T.I., and P.A. Silver. 2002. Requirements for the nuclear export of the small ribosomal subunit. *Journal Of Cell Science*. 115:2985-2995.

Murphy, R., J.L. Watkins, and S.R. Wente. 1996. GLE2, a Saccharomyces cerevisiae homologue of the Schizosaccharomyces pombe export factor RAE1, is required for nuclear pore complex structure and function. *Molecular Biology Of The Cell*. 7:1921-1937.

Napetschnig, J., G. Blobel, and A. Hoelz. 2007. Crystal structure of the N-terminal domain of the human protooncogene Nup214/CAN. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 104:1783-1788.

Navarro, M., G.A.M. Cross, and E. Wirtz. 1999. Trypanosoma brucei variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. *Embo Journal*. 18:2265-2272.

Neer, E.J., C.J. Schmidt, R. Nambudripad, and T.F. Smith. 1994. The Ancient Regulatory-Protein Family Of Wd-Repeat Proteins (Vol 371, Pg 297, 1994). *Nature*. 371:812-812.

Newport, J.W., and D.J. Forbes. 1987. The Nucleus - Structure, Function, And Dynamics. *Annual Review Of Biochemistry*. 56:535-565.

Ngo, H., C. Tschudi, K. Gull, and E. Ullu. 1998. Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *Proceedings of the National Academy of Sciences of the United States of America*. 95:14687-14692.

Niepel, M., C. Strambio-de-Castillia, J. Fasolo, B.T. Chait, and M.P. Rout. 2005. The nuclear pore complex-associated protein, Mlp2p, binds to the yeast spindle pole body and promotes its efficient assembly. *Journal Of Cell Biology*. 170:225-235.

Oberholzer, M., S. Morand, S. Kunz, and T. Seebeck. 2006. A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. *Molecular And Biochemical Parasitology*. 145:117-120.

Ogbadoyi, E., K. Ersfeld, D. Robinson, T. Sherwin, and K. Gull. 2000. Architecture of the Trypanosoma brucei nucleus during interphase and mitosis. *Chromosoma*. 108:501-513.

Orjalo, A.V., A. Arnaoutov, Z.X. Shen, Y. Boyarchuk, S.G. Zeitlin, B. Fontoura, S. Briggs, M. Dasso, and D.J. Forbes. 2006. The Nup107-160 nucleoporin complex is required for correct bipolar spindle assembly. *Molecular Biology Of The Cell*. 17:3806-3818.

Ouali, M., and R.D. King. 2000. Cascaded multiple classifiers for secondary structure prediction. *Protein Science*. 9:1162-1176.

Paba, J., C.A.O. Ricart, W. Fontes, J.M. Santana, A.R.L. Teixeira, J. Marchese, B. Williamson, T. Hunt, B.L. Karger, and M.V. Sousa. 2004. Proteomic analysis of Trypanosoma cruzi developmental stages using isotope-coded affinity tag reagents. *Journal Of Proteome Research*. 3:517-524.

Patterson, S.D., and R.H. Aebersold. 2003. Proteomics: the first decade and beyond. *Nature Genetics*. 33:311-323.

Paulillo, S.M., E.M. Phillips, J. Koser, U. Sauder, K.S. Ullman, M.A. Powers, and B. Fahrenkrog. 2005. Nucleoporin domain topology is linked to the transport status of the nuclear pore complex. *Journal Of Molecular Biology*. 351:784-798.

Pays, E., L. Vanhamme, and D. Perez-Morga. 2004. Antigenic variation in Trypanosoma brucei: facts, challenges and mysteries. *Current Opinion In Microbiology*. 7:369-374.

Pearson, W.R., and D.J. Lipman. 1988. Improved Tools For Biological Sequence Comparison. *Proceedings Of The National Academy Of Sciences Of The United States Of America*. 85:2444-2448.

Prunuske, A.J., and K.S. Ullman. 2006. The nuclear envelope: form and reformation. *Current Opinion In Cell Biology*. 18:108-116.

Quimby, B.B., A. Arnaoutov, and M. Dasso. 2005. Ran GTPase regulates Mad2 localization to the nuclear pore complex. *Eukaryotic Cell*. 4:274-280.

Quimby, B.B., and M. Dasso. 2003. The small GTPase Ran: interpreting the signs. *Current Opinion in Cell Biology*. 15:338-344.

Rabut, G., V. Doye, and J. Ellenberg. 2004. Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nature Cell Biology*. 6:1114-U27.

Rayala, H.J., F. Kendirgi, D.M. Barry, P.W. Majerus, and S.R. Wente. 2004. The mRNA export factor human Gle1 interacts with the nuclear pore complex protein Nup155. *Molecular & Cellular Proteomics*. 3:145-155.

Redmond, S., J. Vadivelu, and M.C. Field. 2003. RNAit: an automated web-based tool for the selection of RNAi targets in Trypanosoma brucei. *Molecular and Biochemical Parasitology*. 128:115-118.

Romero, P., Z. Obradovic, X.H. Li, E.C. Garner, C.J. Brown, and A.K. Dunker. 2001. Sequence complexity of disordered protein. *Proteins-Structure Function And Genetics*. 42:38-48.

Rosenblum, J.S., and G. Blobel. 1999. Autoproteolysis in nucleoporin biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 96:11370-11375.

Rout, M.P., and J.D. Aitchison. 2000. Pore relations: nuclear pore complexes and nucleocytoplasmic exchange. *In* Molecular Trafficking. Vol. 36. 75-88.

Rout, M.P., J.D. Aitchison, A. Suprapto, K. Hjertaas, Y.M. Zhao, and B.T. Chait. 2000. The yeast nuclear pore complex: Composition, architecture, and transport mechanism. *Journal of Cell Biology*. 148:635-651.

Rout, M.P., and M.C. Field. 2001. Isolation and characterization of subnuclear compartments from Trypanosoma brucei - Identification of a major repetitive nuclear lamina component. *Journal of Biological Chemistry*. 276:38261-38271.

Ruan, J.P., G.K. Arhin, E. Ullu, and C. Tschudi. 2004. Functional characterization of a Trypanosoma brucei TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. *Molecular and Cellular Biology*. 24:9610-9618.

Salina, D., P. Enarson, J.B. Rattner, and B. Burke. 2003. Nup358 integrates nuclear envelope breakdown with kinetochore assembly. *Journal Of Cell Biology*. 162:991-1001.

Saric, M., X.D. Zhao, C. Korner, C. Nowak, J. Kuhlmann, and I.R. Vetter. 2007. Structural and biochemical characterization of the Importin-beta center dot Ran center dot GTP center dot RanBD1 complex. *Febs Letters*. 581:1369-1376.

Schimanski, B., T.N. Nguyen, and A. Gunzl. 2005. Characterization of a multisubunit transcription factor complex essential for spliced-leader RNA gene transcription in Trypanosoma brucei. *Molecular And Cellular Biology*. 25:7303-7313.

Schirmer, E.C., L. Florens, T.L. Guan, J.R. Yates, and L. Gerace. 2003. Nuclear membrane proteins with potential disease links found by subtractive proteomics. *Science*. 301:1380-1382.

Schmid, M., G. Arib, C. Laemmli, J. Nishikawa, T. Durussel, and U.K. Laemmli. 2006. Nup-PI: The nucleopore-promoter interaction of genes in yeast. *Molecular Cell*. 21:379-391.

Schmid, M., T. Durussel, and U.K. Laemmli. 2004. ChIC and ChEC: Genomic mapping of chromatin proteins. *Molecular Cell*. 16:147-157.

Shen, S.Y., G.K. Arhin, E. Ullu, and C. Tschudi. 2001. In vivo epitope tagging of Trypanosoma brucei genes using a one step PCR-based strategy. *Molecular And Biochemical Parasitology*. 113:171-173.

Shi, H.F., A. Djikeng, T. Mark, E. Wirtz, C. Tschudi, and E. Ullu. 2000. Genetic interference in Trypanosoma brucei by heritable and inducible double-stranded RNA. *Rna-a Publication of the Rna Society*. 6:1069-1076.

Shulga, N., N. Mosammaparast, R. Wozniak, and D.S. Goldfarb. 2000. Yeast nucleoporins involved in passive nuclear envelope permeability. *Journal Of Cell Biology*. 149:1027-1038.

Siegel, T.N., K.S.W. Tan, and G.A.M. Cross. 2005. Systematic study of sequence motifs for RNA trans splicing in Trypanosoma brucei. *Molecular And Cellular Biology*. 25:9586-9594.

Simpson, A.G.B., J. Lukes, and A.J. Roger. 2002. The evolutionary history of kinetoplastids and their kinetoplasts. *Molecular Biology and Evolution*. 19:2071-2083.

Simpson, A.G.B., J.R. Stevens, and J. Lukes. 2006. The evolution and diversity of kinetoplastid flagellates. *Trends In Parasitology*. 22:168-174.

Siniossoglou, S., C. Wimmer, M. Rieger, V. Doye, H. Tekotte, C. Weise, S. Emig, A. Segref, and E.C. Hurt. 1996. A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell*. 84:265-275.

Sogin, M.L., J.H. Gunderson, H.J. Elwood, R.A. Alonso, and D.A. Peattie. 1989. Phylogenetic Meaning Of The Kingdom Concept - An Unusual Ribosomal-Rna From Giardia-Lamblia. *Science*. 243:75-77.

Sonnhammer, E.L.L., S.R. Eddy, E. Birney, A. Bateman, and R. Durbin. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Research*. 26:320-322.

Staub, E., P. Fiziev, A. Rosenthal, and B. Hinzmann. 2004. Insights into the evolution of the nucleolus by an analysis of its protein domain repertoire. *Bioessays*. 26:567-581.

Stechmann, A., and T. Cavalier-Smith. 2002. Rooting the eukaryote tree by using a derived gene fusion. *Science*. 297:89-91.

Stewart, M. 2006. Structural basis for the nuclear protein import cycle. *Biochemical Society Transactions*. 34:701-704.

Stewart, M. 2007a. Molecular mechanism of the nuclear protein import cycle. *Nature Reviews Molecular Cell Biology*. 8:195-208.

Stewart, M. 2007b. Ratcheting mRNA out of the nucleus. *Molecular Cell*. 25:327-330.

Strambio-de-Castillia, C., G. Blobel, and M.P. Rout. 1999. Proteins connecting the nuclear pore complex with the nuclear interior. *Journal of Cell Biology*. 144:839-855.

Strawn, L.A., T.X. Shen, N. Shulga, D.S. Goldfarb, and S.R. Wente. 2004. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nature Cell Biology*. 6:197-206.

Strawn, L.A., T.X. Shen, and S.R. Wente. 2001. The GLFG regions of Nup116p and Nup100p serve as binding sites for both Kap95p and Mex67p at the nuclear pore complex. *Journal Of Biological Chemistry*. 276:6445-6452.

Subramaniam, C., P. Veazey, R. Seth, J. Hayes-Sinclair, E. Chambers, M. Carrington, K. Gull, K. Matthews, D. Horn, and M.C. Field. 2006. Chromosome-wide analysis of gene function by RNA interference in the African trypanosome. *Eukaryotic Cell*. 5:1539-1549.

Suntharalingam, M., and S.R. Wente. 2003. Peering through the pore: Nuclear pore complex structure, assembly, and function. *Developmental Cell*. 4:775-789.

Tackett, A.J., D.J. Dilworth, M.J. Davey, M. O'Donnell, J.D. Aitchison, M.P. Rout, and B.T. Chait. 2005. Proteomic and genomic characterization of chromatin complexes at a boundary. *Journal Of Cell Biology*. 169:35-47.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T. 1988. Protein and polymer analyses up to m/z 100000 by laser ionization time-of flight mass spectrometry. *Rapid Commun.Mass Spectrom.* 2:151-153.

Templeton, T.J., L.M. Iyer, V. Anantharaman, S. Enomoto, J.E. Abrahante, G.M. Subramanian, S.L. Hoffman, M.S. Abrahamsen, and L. Aravind. 2004.

Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Research*. 14:1686-1695.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*. 25:4876-4882.

Thomson, J.J. 1913. Bakerian Lecture: Rays of Positive Electricity. *Proceeding of the Royal Society of London. Series A.* 89:1-20.

Tong, A.H.Y., B. Drees, G. Nardelli, G.D. Bader, B. Brannetti, L. Castagnoli, M. Evangelista, S. Ferracuti, B. Nelson, S. Paoluzi, M. Quondam, A. Zucconi, C.W.V. Hogue, S. Fields, C. Boone, and G. Cesareni. 2002. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science*. 295:321-324.

Tong, A.H.Y., G. Lesage, G.D. Bader, H.M. Ding, H. Xu, X.F. Xin, J. Young, G.F. Berriz, R.L. Brost, M. Chang, Y.Q. Chen, X. Cheng, G. Chua, H. Friesen, D.S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L.Z. Ke, N. Krogan, Z.J. Li, J.N. Levinson, H. Lu, P. Menard, C. Munyana, A.B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S.L. Wong, L.V. Zhang, H.W. Zhu, C.G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F.P. Roth, G.W. Brown, B. Andrews, H. Bussey, and C. Boone. 2004. Global mapping of the yeast genetic interaction network. *Science*. 303:808-813.

Tran, E.J., and S.R. Wente. 2006. Dynamic nuclear pore complexes: Life on the edge. *Cell*. 125:1041-1053.

Vandriel, R., B. Humbel, and L. Dejong. 1991. The Nucleus - A Black-Box Being Opened. *Journal Of Cellular Biochemistry*. 47:311-316.

Vasu, S.K., and D.J. Forbes. 2001. Nuclear pores and nuclear assembly. *Current Opinion in Cell Biology*. 13:363-375.

Vetter, I.R., C. Nowak, T. Nishimoto, J. Kuhlmann, and A. Wittinghofer. 1999. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature*. 398:39-46.

Vinciguerra, P., N. Iglesias, J. Camblong, D. Zenklusen, and F. Stutz. 2005. Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. *Embo Journal*. 24:813-823.

Walther, T.C., A. Alves, H. Pickersgill, I. Loiodice, M. Hetzer, V. Galy, B.B. Hulsmann, T. Kocher, M. Wilm, T. Allen, L.W. Mattaj, and V. Doye. 2003. The conserved Nup107-160 complex is critical for nuclear pore complex assembly. *Cell*. 113:195-206.

Wang, Z.F., J.C. Morris, M.E. Drew, and P.T. Englund. 2000. Inhibition of Trypanosoma brucei gene expression by RNA interference using an integratable vector with opposing T7 promoters. *Journal of Biological Chemistry*. 275:40174-40179.

Ward, J.J., J.S. Sodhi, L.J. McGuffin, B.F. Buxton, and D.T. Jones. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal Of Molecular Biology*. 337:635-645.

Weirich, C.S., J.P. Erzberger, J.M. Berger, and K. Weis. 2004. The N-terminal domain of Nup159 forms a beta-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Molecular Cell*. 16:749-760.

Wente, S.R. 2000. Gatekeepers of the nucleus. *Science*. 288:1374-1377.

Whalen, W.A., A. Bharathi, D. Danielewicz, and R. Dhar. 1997. Advancement through mitosis requires rae1 gene function in fission yeast. *Yeast*. 13:1167-1179.

Winey, M., D. Yarar, T.H. Giddings, and D.N. Mastronarde. 1997. Nuclear pore complex number and distribution throughout the Saccharomyces cerevisiae cell cycle by three-dimensional reconstruction from electron micrographs of nuclear envelopes. *Molecular Biology Of The Cell*. 8:2119-2132.

Yan, J.X., A.T. Devenish, R. Wait, T. Stone, S. Lewis, and S. Fowler. 2002. Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. *Proteomics*. 2:1682-1698.

Yoon, J.H., W.A. Whalen, A. Bharathi, R.L. Shen, and R. Dhar. 1997. Npp106p, a Schizosaccharomyces pombe nucleoporin similar to Saccharomyces cerevisiae Nic96p, functionally interacts with Rae1p in mRNA export. *Molecular And Cellular Biology*. 17:7047-7060.

Zamore, P.D. 2002. Ancient pathways programmed by small RNAs. *Science*. 296:1265-1269.

Zhang, W.Z., and B.T. Chait. 2000. Profound: An expert system for protein identification using mass spectrometric peptide mapping information. *Analytical Chemistry*. 72:2482-2489.

Zhang, W.Z., A.N. Krutchinsky, and B.T. Chait. 2003. "De novo" peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: A preliminary study. *Journal Of The American Society For Mass Spectrometry*. 14:1012-1021.

Zuccolo, M., A. Alves, V. Galy, S. Bolhy, E. Formstecher, V. Racine, J.B. Sibarita, T. Fukagawa, R. Shiekhattar, T. Yen, and V. Doye. 2007. The human Nup107-160 nuclear pore subcomplex contributes to proper kinetochore functions. *Embo Journal*. 26:1853-1864.