Charles University, Faculty of Science

Univerzita Karlova, Přírodovědecká fakulta

Ph.D. study program: Parasitology

Doktorský studijní program: Parazitologie

Conserved mechanism for targeting of Tail-anchored proteins in eukaryotes

Konzervovaný mechanismus cílení Tail-anchored proteinů u eukaryot

Mgr. Vladimíra Najdrová

Ph.D. thesis/Disertační práce

Thesis supervisor: Mgr. Pavel Doležal, Ph.D.

Prague, 2022

Declaration of the author / Prohlášení autorky

I declare that I elaborated this PhD thesis on my own. The results presented within this thesis were accomplished by myself or in collaboration with the co-authors of the presented papers. I also proclaim that the literary sources were cited properly and neither this work nor the substantial part of it have been used to reach the same or any other academic degree.

Prohlašuji, že jsem tuto disertační práci vypracovala samostatně. Data v ní presentované jsou výsledky mé vlastní práce nebo spolupráce se spoluautory přiložených publikací. Prohlašuji, že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Mgr. Vladimíra Najdrová

Supervisor declaration / Prohlášení školitele

I declare, that data presented in this Ph.D. thesis resulted from the research activities of Mgr. Vladimíra Najdrová and her colleagues, co-authors of the enclosed publications. Her involvement in this work was substantial and that she contributed significantly to obtain the results.

Prohlašuji, že data prezentována v této disertační práci jsou výsledkem vědecké činnosti Mgr. Vladimíry Najdrové a její spolupráce se spoluautory přiložených publikací. Její zapojení v této práci bylo značné a významně přispělo k dosažení výsledků.

Mgr. Pavel Doležal, Ph.D.

Thesis supervisor

Acknowledgements / Poděkování

First, I would like to thank to my supervisor Pavel Doležal for giving me the opportunity to work in Lab in Basement team, opportunity to attend many conferences and great internship in USA, and also for his advices and support. I would like to thank to my lab colleagues for their help, team work and friendly atmosphere in lab - especially in our office. And I am grateful that colleagues can became close friends. My thanks also go to all friends and colleagues from Department of Parasitology and for all the time spent together (not only in Ryba).

I would like to thank to colleagues, which I met during my Ph.D. internship and from whom I learned a lot – not only new methods in lab, but also the overall scientific thinking itself. I thank to Shu-ou Shan, Sowmya and Un-Seng for their time, patience and advices. I would like to thank for opportunity to spend time in Bil Clemons lab with Michelle and Shyam, and I am grateful for our collaboration and that we connect the structural biology with the giardia world. And many thanks go to Nancy and Mr. Flanagan for all the beautiful time we spent together.

Chtěla bych také poděkovat své rodině za pochopení a podporu v průběhu mých studií. Mé největší poděkování patří Nikče za její bezmeznou pomoc a podporu, za motivaci v časech krize, za její trpělivost a důvěru ve mě! Děkuju Ti za všechno!

Table of content

Abbreviations

Bag6	BCL2-associated athanogene 6
BAT3	human leukocyte antigen (HLA)-B-associated transcript 3
CAML	calcium-modulating ligand
CBD	client-binding domain
CHD5	congenital heart disease protein5
EMC	ER membrane protein complex
ER	endoplasmic reticulum
ERAD	ER-associated degradation
GET	guided entry of TA protein
IMP(s)	integral membrane protein(s)
LECA	last eukaryotic common ancestor
MHC	major histocompatibility complex
NBD	nucleotide binding domain
NLS	nuclear localization signal sequence
RNAi	RNA interference
Sgt2	small glutamine-rich tetratricopeptide repeat-containing protein 2
SND	SRP-independent
SRP	signal recognition particle
TA	tail anchored
TMCO1	transmembrane and coiled coil domains1
TMD(s)	transmembrane domain(s)
TOM	outer mitochondrial translocase
TPR	tetratricopeptide repeat
TRC	transmembrane domain complex
TRM	tripartite recognition motif
UBL	ubiquitin-like
UTR	untranslated region
VSP	variant surface protein
WRB	tryptophan rich basic protein

Abstract

Approximately one-fourth of all cellular proteins represent integral membrane proteins (IMPs) that are transported through the cytosol across or into the organellar or plasma membrane. Transport of IMPs requires precise timing which needs to be precisely regulated for them to reach their final destination. Tail-anchored (TA) proteins represent specific class of membrane proteins that lack the N-terminal signal peptide, which targets the nascent polypeptide to the endoplasmic reticulum (ER) membrane for the co-translational transport. Instead, they possess single C-terminal transmembrane domain (TMD) that serves as their targeting signal. Therefore, TA proteins are transported only post-translationally when the C-terminal TMD appears from the ribosome. The Guided Entry of Tail-anchored proteins (GET) pathway is the dominant way of how TA proteins find their way into the ER membrane. It is a multistep process that is mediated by six (Sgt2, Get1-Get5) proteins in yeast and seven (plus Bag6) proteins in human, which involves recognition of a TA protein, its targeting to the ER membrane and the actual membrane insertion. In addition to the model cell systems, some of GET pathway components were studied in plants and recently in *Plasmodium falciparum*, which makes our knowledge on the distribution and the general mechanism of the pathway very limited.

We used parasitic protist *Giardia intestinalis* as model organism to study the general aspects of GET pathway and also to understand the evolution of this eukaryote-specific machinery. Detailed characterization and crystal structures of *G. intestinalis* Get3 greatly contributed to our understanding of the Get3 catalytic cycle during the transport of TA proteins. In addition, the identification of the GET pathway components in *G. intestinalis*, showed the involvement of a Bag6 homologue, which was so far known only in metazoan species. Our data show that last eukaryotic common ancestor (LECA) possessed all the GET pathway components (Sgt2, Get1-Get5, Bag6) and therefore Bag6 was secondarily lost in some lineages of eukaryotes including yeast. Our results from both the experimental and bioinformatic studies suggest that the GET pathway is an ancestral eukaryotic pathway with critical role for eukaryotes.

Abstrakt

Membránové proteiny, které tvoří přibližně jednu čtvrtinu všech proteinů v buňce, jsou transportovány skrze hydrofilní prostředí cytosolu, než jsou integrovány do membrány organel, nebo do plasmatické membrány. Jedná se o složitý proces, který vyžaduje správné načasování i souhru řady zúčastněných proteinů. Skupina transmembránových proteinů, tzv. Tail-anchored (TA) proteinů, neobsahuje klasický signální peptid umístěný na N-konci proteinu, který proteiny cílí k membráně endoplasmatického retikula. Tato informace je místo toho obsažena v jediné transmembránové doméně na C-konci TA proteinů. Z tohoto důvodu musí být TA proteiny transportovány až po ukončení translace. Hlavní způsob transportu TA proteinů je zprostředkován tzv. Guided Entry of Tail-anchored protein drahou neboli GET drahou. Tuto dráhu tvoří u kvasinek celkem šest (Sgt2, Get1-Get5) a u člověka sedm (navíc Bag6) proteinů, které mají za úkol rozeznat a navázat TA protein, přenést ho k membráně endoplazmatického retikula a zprostředkovat jeho vložení do membrány. Kromě modelových systémů, byly některé z proteinů GET dráhy studovány u rostlin a u *Plasmodium falciparum*. To do značné míry komplikuje naše pochopení obecného mechanismu dráhy i její využití napříč eukaryoty.

Rozhodli jsem se proto charakterizovat GET dráhu u parazitického prvoka *Giardia intestinalis*, který představuje divergentního zástupce eukaryot. Krystalické struktury a detailní charakterizace Get3 homologu z *G. intestinalis* doplňují potřebné informace k pochopení obecné funkce a cyklování Get3 proteinu v průběhu transportu TA proteinů. V *G. intestinalis* jsme identifikovali všechny proteiny GET dráhy, a to včetně homologu Bag6 proteinu, který byl doposud považován za inovaci přítomnou pouze u zástupců skupiny Metazoa. Na základě našich výsledků můžeme vyvodit, že všechny proteiny GET dráhy byly přítomny u posledního společného předka eukaryot (LECA) a že Bag6 protein byl u některých eukaryotických linií sekundárně ztracen. GET dráha proto představuje evolučně konzervovaný mechanismus, který je nezbytný pro správnou funkci eukaryotické buňky.

Introduction

Eukaryotic cells contain multiple membrane enclosed organelles, such as the nucleus, the endomembrane system, mitochondria, peroxisomes, and chloroplasts in plants. Integral membrane proteins (IMPs), among others, mediate communication throughout cell membrane compartments, and are involved in various signal and transport pathways (F. Li *et al.*, 2021). The number of IMPs varies across species, but they constitute approximately one-fourth of protein-coding genes (Krogh *et al.*, 2001). IMPs contain different number of TMDs formed by transmembrane helix of about 20 hydrophobic amino acids; a length spanning the lipid bilayer (White *et* von Heijne, 2005). Bacterial and archaeal IMPs are transported directly to the plasma membrane (Xie *et* Dalbey, 2008). Eukaryotic cell compartmentalization evolved mechanisms for sorting and transport of proteins into the organelles (Mayerhofer, 2016; Wiedemann *et* Pfanner, 2017; X. Xu *et al.*, 2021). However, most of the eukaryotic IMPs are first transported into ER, which is evolutionary derived from plasma membrane (Baum *et* Baum, 2014). Translation in the aqueous environment of the cytosol requires protection of the hydrophobic regions and molecular machineries for the targeting and transport of IMPs into the membrane. In case of a flaw, IMPs aggregate, mislocalize and undergo degradation (Guna *et* Hegde, 2018).

The central point of this thesis is GET pathway, a pathway that post-translationally targets and transports TA proteins, to the ER membrane. The GET pathway, together with its individual components, has been intensively studied in recent years and therefore an attempt to summarize all information is beyond the scope of this work.

In order to study post-translational protein transport to the ER, we consistently use our model organism the parasitic protist *G. intestinalis*. The next section of this thesis provides a detailed description of this model organism, and is followed by sections covering general information on TA-proteins and the concept of the GET pathway. **Publication 1** and **Manuscript 2**, in the order enclosed in this PhD thesis, cover the detailed description of the results of this first part.

In addition, our work on the targeting of TA proteins led us to different of cellular process besides the GET pathway. The mislocalization of the ER-destined TA protein to

mitosomes (reduced mitochondria in *G. intestinalis*; details are provided below) was observed during the study of mitosomal protein transport (Voleman, 2011). This prompted us to test potential role of 3' untranslated region (UTR) of mRNA in protein targeting and the role of RNA-binding proteins of the PUMILIO family. The results of this study are summarized in **Publication 3**.

1. *Giardia intestinalis*

The anaerobic protist *G. intestinalis* (syn. *G. lamblia* and *G. duodenalis*) infects a wide range of hosts and causes a disease giardiasis. *G. intestinalis* belongs to the Metamonada group, which includes both parasitic and free-living (e.g., *Trepomonas*, *Carpediemonas*) organisms and is considered one of the earliest divergent eukaryotic lineages (Adl *et al.*, 2012). *Giardia* species are classified into eight assemblages (A-H) based on genetic differences and host specificity, whereas only assemblages A or B can infect humans (Monis *et al.*, 2009). Giardiasis is a worldwide disease with annually more than 200 million symptomatic cases while around 1 billion people are estimated to be infected each year. The prevalence of giardiasis is even ten times higher in developing countries (Torgerson *et al.*, 2015).

Infection of *G. intestinalis* occurs via the oral-fecal route from contaminated water or food. Life cycle involves two major stages: infective cysts and motile trophozoites. Environmentally resistant cysts are ingested via contaminated water or food. Upon ingestion, the acidic environment in the stomach and the presence of trypsin and bile in the duodenum will initiate the infection: the trophozoites are released from the cyst (excystation) and attach to intestinal epithelial cells using their specific adhesive disc. These trophozoites actively multiply through binary fission and will eventually move further down in the small intestine where they will form infective cysts (encystation). *G. intestinalis* cysts are released from the host body through their digestive system and are immediately infectious for the next host (Ankarklev *et al.*, 2010; Einarsson *et al.*, 2016). The entire surface of the trophozoite is covered by highly immunogenic surface antigens that are called variant surface proteins (VSP). Only a single VSP protein is expressed at a time, but

expression changes from one VSP to another. This sophisticated system allows *G. intestinalis* to escape the host's immune system. This antigenic variation is seen *in vitro* and during encystation and excystation processes (Carranza *et al.*, 2002; Nash *et al.*, 1990; Svärd *et al.*, 1998).

Giardiasis can be asymptomatic, but can also cause acute or chronic diarrhoea, which could be accompanied by nausea, vomiting, or abdominal cramps (Einarsson *et al.*, 2016). In some cases, the infection may lead to years of chronic fatigue and long-term gastrointestinal disorders (Hanevik *et al.*, 2014). Children with giardiasis can experience growth failure and malnutrition. The infection usually resolves on its own, but if not, drugs containing metronidazole are currently the first-line treatment for giardiasis (reviewed in Riches *et al.*, 2020).

G. intestinalis trophozoites have a typical pear-shaped form and distinctive microtubule cytoskeleton structures. Their four flagella pairs, adhesive disk, and the median body are crucial components for motility, division, attachment to the host's intestine epithelium, encystation, and excystation (Hagen *et al.*, 2020). Additionally, the presence of mitosomes, one of the most reduced mitochondrial related organelles, makes *G. intestinalis* even more unique (Tovar *et al.*, 2003). Mitosomes are double membrane-bounded organelles without its own genome and with highly reduced translocation machinery. Mitosomes are devoid of respiration and ATP production and the only pathway retained in the mitosomes mediate the assembly of Fe-S clusters (Jedelský *et al.*, 2011; Martincová *et al.*, 2015; Nyindodo-Ogari *et al.*, 2014). Other well-defined endomembrane components present in *G. intestinalis* are the endoplasmic reticulum, peripheral vacuoles and encystations-specific vesicles. *G. intestinalis* has long been believed to lack other common subcellular compartments, such as peroxisomes and the Golgi apparatus (Faso *et* Hehl, 2011). However, recently, two peroxisomal proteins were identified *in silico* and their localization in cytosolic vesicles suggest the presence of rudimentary form of peroxisomes in *G. intestinalis* (Acosta-Virgen *et al.*, 2018). Similarly, Golgi like structures called encystation specific vesicles (ESVs), were observed in encysting cells. They originate from ER and are required for transport of cell wall material such as cyst wall proteins (Štefanić *et al.*, 2009). The observed effect on

the growth of *G. intestinalis* cells after treating them with Brefeldin A, a drug that blocks protein transport through the Golgi apparatus in animals, suggests the presence of a rudimental Golgi apparatus or an alternative pathway. Other studies have shown participation of the ER in protein sorting, thus potentially taking on Golgi apparatus functions in *G. intestinalis* (Faso *et al.*, 2013; Klausner *et al.*, 1992; Lujan *et al.*, 1995; Zamponi *et al.*, 2017).

The WB isolate (assemblage A) is the most studied *G. intestinalis* isolate and is commonly used as a model organism in the laboratory. This strain was isolated from a patient W.B. who suffered from chronic symptomatic giardiasis, after getting infected in Afghanistan. The *G. intestinalis* culture of this strain was further axenized at the NIH (Nash *et al.*, 1985; Smith *et al.*, 1982). The genetic manipulation (stable episomal transfection) of *G. intestinalis* is based on DNA electroporation and the subsequent use of antibiotics as selectable markers (Singer *et al.*, 1998; Sun *et al.*, 1998; Yee *et* Nash, 1995). Cells are maintained *in vitro* in complex media TYI-S-33 (Keister, 1983) and the protocols for excystation and encystation *in vitro* are also well established (Bingham *et* Meyer, 1979; Gillin *et al.*, 1988). Furthermore, the genome of isolate WB was the first sequenced genome of *G. intestinalis* (Morrison *et al.*, 2007) and was recently re-sequenced (F. Xu *et al.*, 2020).

G. intestinalis contains two nuclei with a compact tetraploid genome distributed on five chromosomes. Genome ploidy varies during the life cycle of *G. intestinalis* from 4N in trophozoites to 16N in the cysts (Bernander *et al.*, 2001). However, morphologically similar two nuclei may differ in genome size and number of genes, which is caused by deletions and aneuploidy (Tůmová *et al.*, 2007, 2016). Originally, *G. intestinalis* was assumed to be asexual, but genes encoding meiotic proteins have been found in the genome. Population genetic studies and the surprisingly low sequence heterozygosity in the *G. intestinalis* genome indicated the presence of an unknown system for recombination between assemblages (Ankarklev *et al.*, 2012; Cooper *et al.*, 2007; Ramesh *et al.*, 2005). DNA exchange between nuclei in one cell (diplomixis) has been suggested to occur during encystation, yet no sexual reproduction has been directly observed in *G. intestinalis* (Carpenter *et al.*, 2012; Poxleitner *et al.*, 2008).

G. intestinalis genome contains minimal noncoding regions, short 5´ and 3´ UTRs, and very few introns; specifically, eight *cis* and five *trans* introns (Elmendorf *et al.*, 2001; Franzén *et al.*, 2013; L. Li *et* Wang, 2004; Svärd *et al.*, 1999; F. Xu *et al.*, 2020). In total, 4,963 protein encoding genes and additional 306 pseudogenes were annotated in genome. Predicted proteins with unknown functions represent 42% of *G. intestinalis* of the conceptual proteomes (F. Xu *et al.*, 2020). In addition to the gene transfection technique mentioned above, a variety of tools have been developed for studying *G. intestinalis*. Homologs of Argonaut and Dicer, components of RNA interference (RNAi), were detected in *G. intestinalis* and RNAi was used for gene silencing to find new drug targets and to study antigenic variation (Marcial-Quino *et al.*, 2017; Prucca *et al.*, 2008). So far, using the RNAi as functional tool for controlling gene expression have not been utilised in *G. intestinalis*, like it is frequently done for trypanosomes (Balaña-Fouce *et* Reguera, 2007). Morpholino oligonucleotides were used to block translation of various genes (Carpenter *et* Cande, 2009). The first complete knockout of a *G. intestinalis* gene was carried out using the Cre/loxP system that enabled to recycle the selectable marker for subsequent deletion of four gene alleles (Ebneter *et al.*, 2016). The CRISPR/Cas9, CRISPRi and CasRx systems have been established in *G. intestinalis* cells for gene knockdowns (Z. Q. Lin *et al.*, 2019; McInally *et al.*, 2019; Shih *et al.*, 2021) and only recently, the efficient and easily applicable CRISPR/Cas9 mediated gene deletion system was established (Horáčková *et al.*, 2022). All these tools now allow to apply long missed functional genomics to study *G. intestinalis* biology.

Because of its genomic reduction, *G. intestinalis* allows the study of minimal sets of cellular components that are needed for general biological pathways and cellular processes. The parasitic lifestyle, unique cellular structures, and early divergent origin make *G. intestinalis* an attractive object for studying cell and evolutionary biology.

2. Integral membrane proteins (IMPs) and their ways into the ER membrane

The vast majority of the endomembrane IMPs are first transported to the ER membrane and subsequently to their final destination. Generally, the protein transport of ER-destined proteins occurs either co- or post-translationally. The best understood pathway depends on the signal recognition particle (SRP), which specifically recognizes the N-terminal signal peptide sequence on the nascent polypeptide that emerges from the ribosome during translation. Binding of SRP to the nascent peptide stops the translation and targets the entire complex to the proximity of the ER, where proteins are co-translationally transported into ER through the Sec61 translocon (reviewed by Hsieh *et al.*, 2021).

A special class of IMPs, called TA proteins, lacks the signal peptide at the N-terminus. Instead, it is the only TMD at the C-terminus of the protein that carries the targeting information (Kutay *et al.*, 1993). Therefore, this group of proteins must be targeted to the ER post-translationally without the participation of SRP. Most of the TA proteins are targeted to the membrane through a specialised GET pathway, which has been so far well studied in mammals (Favaloro *et al.*, 2008; Stefanovic *et* Hegde, 2007), yeast (Schuldiner *et al.*, 2008), and also in plants (Srivastava *et al.*, 2017; Xing *et al.*, 2017).

An alternate SRP-independent pathway (SND) has been found in yeast. Proteins with centrally positioned TMDs are captured by SND1 and directed to membrane localized SND2 and SND3, which associate with the Sec translocon (Aviram *et al.*, 2016). The mammalian homolog of SND2 participates in protein targeting into ER, but mammalian homologs of SND1 and SND3 have not yet been identified (Casson *et al.*, 2017; Haßdenteufel *et al.*, 2017). Lastly, a subset of TA proteins is transported into the ER membrane through the ER membrane protein complex (EMC), of which the details were recently discovered (Chitwood *et al.*, 2018; Guna *et al.*, 2018).

2.1 Tail-anchored proteins

TA proteins represent a special class of membrane proteins, that carry a single C-terminal TMD targeting and anchoring them to the ER membrane. TMD emerges from the ribosome when translation is completed, and therefore TA proteins must be inserted into the membrane post-translationally (Kutay *et al.*, 1993). TA proteins make up approximately 2 – 5 % of all IMPs; specifically, more than 50 TA proteins were identified in yeast, over 400 in humans and more than 500 in plants (Burri *et* Lithgow, 2004; Kalbfleisch *et al.*, 2007; Kriechbaumer *et al.*, 2009). They are distributed across various cell compartments such as the ER, mitochondria, Golgi apparatus, peroxisomes, nuclear envelope, plasma membrane, or plastids (Borgese *et al.*, 2003; Brito *et al.*, 2019). Due to the conserved membrane topology the N-terminus of TA proteins is always exposed to the cytosol where it can participate in a number of cellular processes: SNARE mediate the membrane fusion during the vesicular transport (Y. Zhang *et* Hughson, 2021), Sec61β and Sec61γ are components of the ER translocon (Gemmer *et* Förster, 2020), Tom5 and Tom6 are part of the outer mitochondrial translocase (TOM) complex (Gupta *et* Becker, 2021), and proteins from the Bcl-2 family serve as regulators of apoptosis (Wanderoy *et al.*, 2020).

Although all TA proteins share C-terminal TMDs, the hydrophobicity and the length of TMDs can differ as well as the charge of surrounding amino acids. The combination of these features determines their terminal destination and the functional requirement of the GET pathway (Beilharz *et al.*, 2003). In general, medium and highly hydrophobic TMDs are typical for the ER TA proteins, whereas long and strongly hydrophobic TMDs routes TA proteins to the Golgi apparatus or the plasma membrane. Mitochondria destined TA proteins usually carry weakly or medium hydrophobic TMDs and a positively charged C-terminal tail (Borgese *et al.*, 2007; Costello *et al.*, 2017; Figueiredo Costa *et al.*, 2018; Rao *et al.*, 2016).

Besides the terminal destination, the hydrophobicity of TMD is crucial for the initial entry in the ER membrane. Highly hydrophobic TMDs are recognized by Get3 and transported through Get1/Get2 complex (Mariappan *et al.*, 2011). However, up to one half

of TMDs are insufficiently hydrophobic to utilize the GET pathway. Experimentally, both *in vitro* and *in vivo* studies showed that TA proteins with lower hydrophobicity are inserted into the ER membrane by EMC complex. Nevertheless, Sec61β, the TA protein with moderate hydrophobicity has showed partial dependence on both insertases, suggesting overlap of these two pathways (Guna *et al.*, 2018).

Relatively little data exist about TA proteins in *G. intestinalis*. So far, 17 SNARE proteins were found in *G. intestinalis* (Elias *et al.*, 2008). One of the identified *G. intestinalis* SNARE proteins is a homolog of Sec20. It has been previously shown that Sec20 is localized in the ER where it functions in retrograde transport to the ER (Burri *et al.*, 2003; Sweet *et* Pelham, 1992). However, unexpectedly, *G. intestinalis* Sec20 homolog was localized in mitosomes (Elias *et al.*, 2008).

Therefore, we explored this atypical localization of Sec20 in more detail. The localization of Sec20 in *G. intestinalis* was determined either via its episomal expression as in the original study (Elias *et al.* 2008) or via specific polyclonal antibody raised against Sec20. Interestingly, while the episomally expressed Sec20 was always found in the mitosomes, the immunolabeling of the endogenous protein via specific antibody showed always the ER localization. Later, we studied the effect of the recombinant protein tag, the level of protein expression or the involvement of 3´UTR on Sec20 localization (Voleman, 2011). While up today we have not been able to explain the molecular basis for mistargeting of episomally encoded Sec20, these experiments prompted us to study general aspects of TA protein targeting in *G. intestinalis*.

3. GET pathway

According to current model, after the release of their hydrophobic C-tail from the ribosome, TA proteins are bound by the ribosome-associated chaperone Sgt2 (small glutamine-rich tetratricopeptide repeat [TPR]-containing protein 2) (F. Wang *et al.*, 2010). Subsequently, Sgt2 recruits Get5 subunit of heterotetrameric Get4/Get5 complex. Through this connection, Sgt2 mediates the binding of Get3 ATPase to Get4 and allows the transfer of the TA protein to Get3. Sgt2 and Get4/Get5 complex act at the beginning of the GET pathway; hence, they are sometimes called as cytosolic or pre-targeting complex of

the GET pathway (Chartron *et al.*, 2010, 2011; Gristick *et al.*, 2015). Get3 with the bounded TA protein is released from the cytosolic complex and delivers TA protein to the membrane complex consisting of Get1 and Get2. These proteins recruit TA protein from Get3 and mediate the insertion of TA protein into the ER membrane (Aviram *et* Schuldiner, 2017; Borgese *et al.*, 2019; Hegde *et* Keenan, 2011; Mariappan *et al.*, 2011; McDowell *et al.*, 2020).

In metazoan, the GET pathway is named as transmembrane domain complex (TRC) pathway and includes homologs of all known yeast GET pathway components known under different names: Sgt2 (SgtA), Get3 (TRC40), Get4 (TRC35), Get5 (Ubl4A), and Get1/Get2 (WRB/CAML). For simplicity, in this thesis the fungal nomenclature (Sgt2, Get1-Get5) will be used (Table 1). In addition to fungal homologs, the metazoan GET pathway contains Bag6 protein which together with Get4 and Get5 homologs forms a heterotrimeric Bag complex. The complex facilitates the transfer of the TA protein from Sgt2 to Get3 and represents functional parallel to the heterotetrameric fungal Get4/Get5 complex (Leznicki *et al.*, 2010; Mariappan *et al.*, 2010; Mock *et al.*, 2015; Vilardi *et al.*, 2011; Yamamoto *et* Sakisaka, 2012). Models of both GET pathways are summarized in Figure 1.

Table 1 Nomenclature of GET pathway components.

yeast	**metazoan**
Guided Entry of TA proteins – GET pathway	Transmembrane domain complex – TRC pathway
Sgt2	SgtA
Get3	TRC40
Get4	TRC35
Get5	Ubl4A
-	Bag6
Get1	WRB
Get2	CAML

Small glutamine-rich tetratricopeptide repeat-containing protein 2 (Sgt2); Guided entry of tail-anchored proteins (Get); Transmembrane domain complex (TRC); BCL2-associated athanogene 6 (Bag6); tryptophan rich basic protein (WRB); calcium-modulating ligand (CAML)

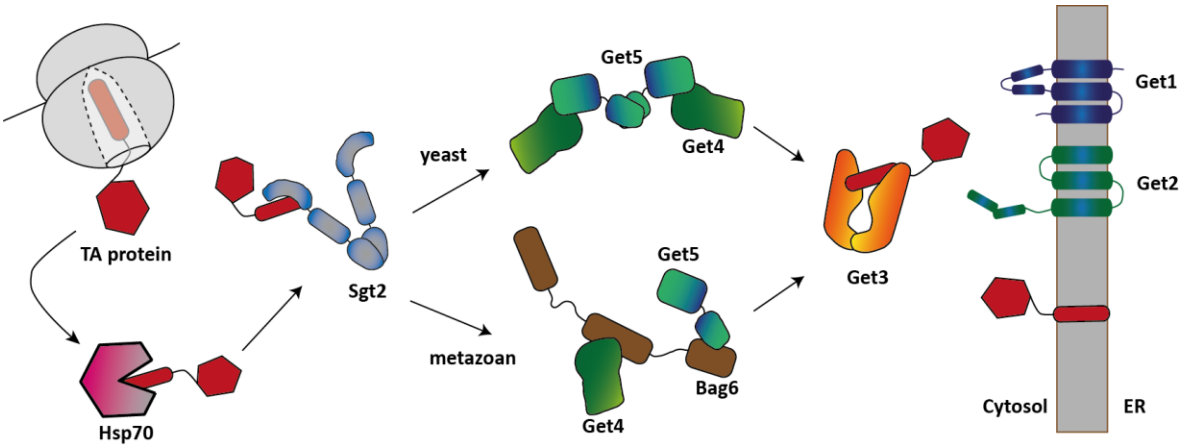


Figure 1. Scheme of yeast and metazoan GET pathway

TA protein is captured by Sgt2 chaperon with the help of Hsp70 chaperone. Subsequently TA protein is transported to Get5/Get4 complex in yeast (top) or to Get4/Bag6/Get5 complex in metazoan (bottom). TA protein is then captured by Get3 which delivers it to the membrane complex of Get1/Get2 for the insertion into ER membrane.

3.1 Chaperones and first steps of GET pathway

The existence of different ways of transporting proteins from the ribosome to their destination raises the question of how exactly proteins are sorted during translation. It is known, that SRP is ribosome binding protein and recognizes the cargo at the beginning of translation (Halic *et al.*, 2004; Hsieh *et al.*, 2021). Recognition of nascent TA proteins emerging from the ribosome is a dynamic multi-step process. Key to the process is to minimize the exposure of the hydrophobic C-tail to the aqueous environment of the cytosol and limit the protein aggregation. Capture of TA protein by chaperone Sgt2 was a generally accepted as a first step of the GET pathway. However, recent reports showed that Sgt2 itself is not sufficient to maintain the solubility of TA proteins as they require initial capture by cytosolic Hsp70 (yeast Ssa1). Upon TA protein binding Hsp70, transfer the protein to Sgt2 hereby initiating the cascade of TA protein transfers in the GET pathway. The importance of Hsp70 for TA protein targeting was confirmed by the fact that Hsp70 deletion disrupted

the delivery of TA protein into ER membrane *in vivo* (Cho *et* Shan, 2018). Additional two J-domain containing cochaperones, Ydj1 and Sis1, have been shown to play essential roles in regulating Hsp70 activity. (Cho *et al.*, 2021).

The chaperone Sgt2 forms a homodimer through its N-terminal dimerization domain, which also interacts with ubiquitin-like (UBL) domain of Get5 in yeast or UBL domain of Bag6 in mammals (Chartron, VanderVelde, *et al.*, 2012; Darby *et al.*, 2014; Winnefeld *et al.*, 2006). The central highly conserved TPR domain interacts with multiple heat shock proteins. Affinity purifications of yeast Sgt2 confirmed the interaction with Hsp70 (yeast Ssa1/Ssa2), Hsp90 (yeast Hsc82) and Hsp104 (Chartron *et al.*, 2011; Krysztofinska *et al.*, 2017; F. Wang *et al.*, 2010) and similar interactions were also observed for mammalian Sgt2 homolog (Roberts *et al.*, 2015). Mutations within TPR domain impair the TA protein loading onto Sgt2 although TMD of TA protein occurs in the C-terminal glutamine rich region of Sgt2 (Cho *et* Shan, 2018; Liou *et* Wang, 2005). Recently, structural model of Sgt2 showed that the C-terminal domain forms a hydrophobic groove which binds TMD at least 11 amino acids long. Structural similarities of Sgt2 with chaperone Sti1/HOP (STI1 domain, TPR repeats) indicated conserved role and common evolutionary origin of these chaperones (K. F. Lin *et al.*, 2021; Schmid *et al.*, 2012).

In addition to the TA protein targeting, Sgt2 is also involved in regulation of mislocalized membrane protein in cytosol (Leznicki *et* High, 2012; Wunderley *et al.*, 2014) and in retrograde protein transport by ER-associated degradation (ERAD) system for subsequent proteasomal degradation (Y. Xu *et al.*, 2012). Both yeast and human Sgt2 proteins were identified as part of polyglutamine-formed protein aggregates, which are common pathological hallmarks of neurogenerative diseases such as Parkinson´s disease or amyotrophic lateral sclerosis (S. Kubota *et al.*, 2021; Soto *et* Pritzkow, 2018).

3.2 Get4/Get5 and Bag6

Get5 was originally described as a factor required for yeast mating and Get5 disruption led to growth defect (Hu *et al.*, 2006; Iwanejko *et al.*, 1999). Get4 was identified as a protein of unknown function that interacts with Get5 in a systematic screen of ribosomal associated proteins (Fleischer *et al.*, 2006). Bioinformatic analysis of Get4 homolog from fish (Cee) revealed high conservation of the protein metazoan, yeast and also in Alveolata, but no Get4 orthologs were found in bacteria or archaea. Get4 was found to lack any known conserved domains or motifs, which complicated the initial function estimation (Fernandes *et al.*, 2008). It was the experimental characterization of Get4 (and Get5) that showed protein(s) participation in TA protein targeting in yeast and mammals (Jonikas *et al.*, 2009; Mariappan *et al.*, 2010; F. Wang *et al.*, 2010).

Crystal structures of yeast Get4/Get5 complex revealed it as a heterodimer, which is mediated by the C-terminal dimerization domain of Get5. The Get4/Get5 interface is formed by C-terminus of Get4 (helix13 and14) and the N- terminal domain of Get5. Given the interaction of Get5 via the N-terminal UBL domain with Sgt2, it constitutes a bridge between Sgt2 and Get4. Several structural characterizations also revealed Get3 binding site on N-terminus of Get4 (Bozkurt *et al.*, 2010; Chang *et al.*, 2010; Chartron *et al.*, 2010; Gristick *et al.*, 2014).

As mentioned above, in addition to Get4 and Get5, the metazoan cytosolic GET complex contains also Bag6 protein. The protein binds Get4 and Get5 independently on its C-terminal part of Bag6 (Mock *et al.*, 2015) and this binding thus does not require mutual Get4 and Get5 interaction. This difference is reflected by distinct domain structure of both metazoan Get4 and Get5. While Get5 lacks the N-terminal domain, Get4 is missing the C-terminal helices 13 and 14, which mediate yeast counterpart interaction. (Leznicki *et al.*, 2010; Mariappan *et al.*, 2010).

Human Bag6 homolog is a large protein (1132 amino acids) that includes UBL domain at the N-terminus, central proline-rich domain, nuclear localization signal sequence (NLS) and modified BAG domain (mock BAG) at the C-terminus which is responsible for binding of Get4 and Get5 (Mock *et al.*, 2015). Binding of Get4 to Bag6 conceals the NLS signal and

hence prevents Bag6 nuclear localization (Q. Wang *et al.*, 2011). From the evolutionary perspective, Bag6 has so far been identified only in metazoan.

Bag6 is a member of the BAG (Bcl-2 associated athanogene) family and it was originally described as human leukocyte antigen (HLA)-B-associated transcript 3 (BAT3) as a part of gene clusters on chromosome 6 together with human major histocompatibility complex (MHC) class III (Spies *et al.*, 1989). Later, Bag6 was found to play a role in the immune response by regulating the expression of MHC class II, Th1 cells survival and activity of natural killer cells (Kämper *et al.*, 2012; Pogge von Strandmann *et al.*, 2007; Rangachari *et al.*, 2012). Bag6 was also described as apoptosis regulator by interacting with Reaper protein, a central apoptosis effector in *Drosophila melanogaster* (Thress *et al.*, 1998). Other studies reported various Bag6 roles in meiosis and spermatogenesis (Sasaki *et al.*, 2008), gene expression (Nguyen *et al.*, 2008), microtubules organization during ciliogenesis (He *et al.*, 2020), ERAD and mislocalized proteins degradation (Hessa *et al.*, 2011; Q. Wang *et al.*, 2011). The N-terminal part (UBL domain) of Bag6 was shown to interact with the Rpn10 proteasome subunit, and the Bag complex together with the FBXO7-SCF E3 ubiquitin ligase is required for proper proteasome function (Kikukawa *et al.*, 2005). Bag complex together with Sgt2 is involved in various cellular process and defect of these protein are associated with many human diseases, such as cancer or neurodegenerative diseases (Benarroch *et al.*, 2019; Q. Wang *et al.*, 2021).

3.3 Get3

The first identified and key component of the GET pathway is cytoplasmic ATPase Get3 (Stefanovic *et* Hegde, 2007). Get3 undergoes ATP-dependent conformational changes to shuttle TA protein between the cytosolic and the ER membrane complex (Rome *et al.*, 2013; Simpson *et al.*, 2010; F. Wang *et al.*, 2010). Several solved structures of different opisthokont Get3 homologs revealed that the protein functions as a homodimer stabilized by Zn^{2+} binding via four conserved cysteines. Get3 is composed of N-terminal nucleotide binding domain (NBD) and flexible α-helical region that forms the client-binding domain

(CBD) (Bozkurt *et al.*, 2009; Gristick *et al.*, 2014; Mateja *et al.*, 2015; McDowell *et al.*, 2020; Suloway *et al.*, 2009).

Conformational changes and dynamics of Get3 function have been intensively studied and resulted into a mechanistic model which connects the ATPase cycle to protein binding. Briefly, the ATP-bound Get3 is selectively captured by cytosolic GET complex. This interaction inhibits Get3 ATPase activity and induces TA protein transfer from Sgt2 to Get3. Dissociation of Get3-TA protein from the cytosolic complex activates ATP hydrolysis (Chartron, Clemons, *et al.*, 2012; Chio *et al.*, 2017; Hegde *et* Keenan, 2011). Subsequently, TA protein is released from Get3 after binding to the Get1/Get2 membrane complex and Get3 is dissociated after new ATP molecule is bound (Gristick *et al.*, 2014; K. Kubota *et al.*, 2012; Mariappan *et al.*, 2011; Rome *et al.*, 2013, 2014; Stefer *et al.*, 2011).

Bacterial homologue of Get3 (ArsA) shares high sequence similarity with the eukaryotic protein. However, ArsA is not involved in protein transport and instead provides heavy metal resistance. Soluble ATPase ArsA in association with membrane channel ArsB mediates resistance by pumping toxic anions out of the cell (Chen *et al.*, 1986). It contains two NBD and a metal binding site with three cysteine residues that are not conserved in eukaryotic homologs (Zhou *et al.*, 2000). Archaeal homolog of Get3 exists as a tetramer and is capable of binding TA protein and delivery it to membrane. These data suggest an archaeal TA protein targeting pathway, although other archaeal orthologs of GET pathway components have not been identified yet (Borgese *et* Righi, 2010; Sherrill *et al.*, 2011; Suloway *et al.*, 2012).

Interestingly, several reports showed that Get3 serves not only in TA protein targeting but takes on additional role of an ATP-independent chaperone during stress conditions. Upon oxidative stress Get3 was observed to form reversible higher oligomeric assemblies via disulfide bridges and colocalizes with unfolded proteins and chaperons (Powis *et al.*, 2013; Ulrich *et al.*, 2022; Voth *et al.*, 2014).

3.4 Membrane insertases Get1/Get2

The first step of TA protein insertion into the ER membrane is mediated by the interaction of Get3 with the cytosolic domains of the transmembrane complex

Get1/Get2 in yeast (Schuldiner *et al.*, 2008) or WRB/CALM in mammals (Vilardi *et al.*, 2011; Yamamoto *et* Sakisaka, 2012). Positively charged amino acid residues within the soluble N-terminal domain of Get2 are required for the recognition and capture of the Get3-TA protein complex (Stefer *et al.*, 2011; Yamamoto *et* Sakisaka, 2012). Subsequently, the soluble coiled-coil domain of Get1 likely stabilizes the open form of the Get3 dimer, leading to nucleotide (ADP) release and disruption of the Get3-TA protein interaction. TA protein is then delivered to the lipid bilayer while Get3 is recycled (K. Kubota *et al.*, 2012; Mariappan *et al.*, 2011; Stefer *et al.*, 2011; F. Wang *et al.*, 2011).

Get1 and Get2 create a stoichiometric complex (Get1/Get2) with an unclear structure (F. Wang *et al.*, 2011, 2014). Three models of interaction between Get1/2 and Get3 are currently recognized: (i) static heterotetramer, (ii) static heterodimer and (iii) dynamic complex mode. In the first model, Get1/2 complex exists as a heterotetramer which binds single Get3 dimer at its opposite sites first by Get2 and later by Get1 or Get1/Get2 cytosolic domains (K. Kubota *et al.*, 2012; Mariappan *et al.*, 2011; Stefer *et al.*, 2011). In the second model, the Get3-TA protein complex is captured by one copy of Get2 within a heterodimer and this interaction is later displaces by binding of one copy of Get1 (Zalisko *et al.*, 2017). The dynamic model proposes assembly of Get1/Get2 tetramer from two heterodimer upon the arrival of Get3-TA protein complex (K. Kubota *et al.*, 2012). The presence of Get1 is required for the correct topology and for post-translational insertion of Get2 into the membrane. If Get1 is not present, Get2 forms atypical topologies that are degraded by the proteasome (Carvalho *et al.*, 2019).

Interestingly, the bioinformatic analyses revealed unexpected relationship between Get1, mitochondrial, bacterial and archaeal protein insertases that belong to the Oxa1 superfamily of protein (Oxa1/Alb3/YidC). Bacterial YidC mediates post-/co-translational protein insertion into cytoplasmic membrane mostly in concert with Sec translocase (Dalbey *et al.*, 2014). Its eukaryotic homologues are found in the inner mitochondrial membrane (Oxa1 and Cox18) and the thylakoid membrane of plastids (Alb3 and Alb4) (Petrů *et al.*, 2021). For a long time, the endomembrane system member of the Oxa1 superfamily was not known until the identification of the archaeal homolog. Ylp1 (YidC-like

protein1) is a member of the protein family named “Domain of unknown function 106” (DUF106) (Borowska *et al.*, 2015). Deep analyses of the DUF106 family revealed that this family includes not only archaeal proteins but also three eukaryotic proteins. Based on structural, functional and biochemical similarities, EMC3 (ER membrane complex subunit3), TMCO1 (transmembrane and coiled coil domains1) and Get1 protein families were assigned as members of Oxa1 superfamily. All proteins are located in ER membrane and they contain three TMD helices, cytosolic coiled coil part between the first and the second TMDs, orientation N-terminus in/C-terminus out, and hydrophilic groove for substrate binding (Anghel *et al.*, 2017).

The EMC3 protein is a component of the EMC complex that co-translationally inserts a multi-pass IMP into the ER membrane (Shurtleff *et al.*, 2018). Defect in EMC3 in humans leads to retinal bipolar cell degeneration (Shurtleff *et al.*, 2018; Zhu *et al.*, 2020). TMCO1 protein is a part of ER protein translocon, which function as insertase (McGilvray *et al.*, 2020) and also actively protects the ER from overfilling of Ca^{2+} (Q. C. Wang *et al.*, 2016). Depletion of TMCO1 causes human cerebro-facio-thoracic dysplasia (Xin *et al.*, 2010). In humans, Get1 encoding gene was located in the region of chromosome 21 associated with heart disease in individuals with Down syndrome. Therefore, Get1 is also called CHD5 (congenital heart disease protein5) (Egeo *et al.*, 1998). Cardiac abnormalities associated with Get1 homologs were also observed in other organisms: *Danio rerio* (Milan *et al.*, 2009), medaka fish (*Oryzias latipes*) (Murata *et al.*, 2009), *Xenopus* sp. (Sojka *et al.*, 2014).

Much less is known about the function and the evolution of Get2. Both yeast and mammalian Get2 homologs contain three predictable TMDs at the C-terminus, but this topology was not confirmed experimentally (Yamamoto *et* Sakisaka, 2012). While the soluble N-terminal part of Get2 is involved in the primary contact with the Get3-TA protein complex (Vilardi *et al.*, 2014), the membrane embedded C-terminus of Get2 interacts with Get1. This interaction was found for both the yeast and mammalian proteins (Colombo *et al.*, 2016; Vilardi *et al.*, 2014; Yamamoto *et* Sakisaka, 2012). Interestingly, yeast Get2 and mammalian CAML are so divergent in sequence that they were not considered as homologous proteins. Yet, recent bioinformatic search suggested a common evolutionary

origin for Get2 and CAML proteins at least across metazoan, fungi and plants (Borgese, 2020). In addition, defects in CAML were also associated human pathogenesis, specifically, depletion of the protein plays role in the proliferation of cancer cells (Shing *et al.*, 2017).

4. RNA-binding protein and their role in protein targeting

As was mentioned above (2.1.), the ER-destined Sec20 was mislocalized to mitosomes when episomally expressed in *G. intestinalis*. Our unpublished data showed that when expressed from the plasmid with the native 3´UTR partially Sec20 could be partially found in the ER (Voleman, 2011). These observations led us to investigate the potential role of 3´UTR and RNA-binding proteins in localization of proteins in *G. intestinalis*.

In addition to traditional co- and post-translational protein transport, new model based on SRP and translation independent mRNA targeting was described. Targeting of mRNA within the cell provide an effective mechanism for gene regulation, protein targeting and sorting (Cui *et* Palazzo, 2014). For example, protein p180 is ER-destined transmembrane protein, which represents a mRNA receptor mediating mRNA localization to the ER in mammals (Cui *et al.*, 2012). Surprisingly, mRNA of two TA proteins (Sec61β and nesprin-2) were partially localized to the surface of ER and probably without participation of SRP, GET pathway proteins or p180 receptor (Cui *et al.*, 2015).

The well-studied group of RNA binding proteins are Puf (Pumilio and FBF) proteins, which interacts with the 3´UTR of mRNA in the cytosol or with rRNA in the nucleolus (Qiu *et al.*, 2014; Thomson *et al.*, 2007; Zamore *et al.*, 1997). Every Puf protein contain C-terminal RNA binding domain, which is composed of Puf repeats and each of repeat binds a single base of RNA (Miller *et al.*, 2008; Y. Wang *et al.*, 2009). The RNA sequence motif recognized by Puf proteins is determined by the combination of three amino acids residues called tripartite recognition motif (TRM). This motif has been shown to be specific to particular base thus it is possible to predict RNA sequence motif that is bound by specific Puf protein (Campbell *et al.*, 2014; Dong *et al.*, 2011; C. Zhang *et* Muench, 2015).

Based on binding domain composition and biological role, Puf proteins have been classified into three groups: (i) classical Puf proteins bind 3´UTR of mRNA and contain eight

puf repeats (X. Wang *et al.*, 2002), (ii) Nop9 proteins contain 11 Puf repeats and participate in processing and folding of 18S rRNA in nucleolus (J. Zhang *et al.*, 2016), and (iii) PUM3 proteins contain 11 Puf repeats, bind double-stranded DNA or RNA, and are involved in nucleolar processing of the large subunit of the ribosome (Z. Li *et al.*, 2009; Qiu *et al.*, 2014).

Classical Puf proteins together with additional protein partners participate in mRNA control through posttranscriptional regulation and translation repression (M. Wang *et al.*, 2018), but they can mediate mRNA localization, as well. Yeast protein Puf3p was shown to facilitate the localization of nuclear coding mRNA for mitochondrial proteins in close proximity of mitochondrial membrane and play role in mitochondrial biogenesis (Saint-Georges *et al.*, 2008).

Hence, in this thesis, we have initiated the characterization of *G. intestinalis* Puf proteins in order to understand their biological role potentially also mediating proper distribution of TA proteins in the cell.

Aims and Objectives

1) Characterization of *G. intestinalis* Get3 homolog
2) Characterization of the GET pathway in *G. intestinalis*
3) Reconstruction of the evolution of GET proteins in eukaryotes
4) Identification of Puf homologs in *G. intestinalis* and elucidation of their potential role in mRNA/protein targeting

List of publications

1. Fry, M. Y., Najdrová, V., Maggiolo, A. O., Saladi, S. M., Doležal, P., Clemons, W. M. (2022). **Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3**. *Nature Structural Molecular Biology*, *29*(8), 820-830. https://doi.org/10.1038/s41594-022-00798-4

My contribution: Isolation of Get3 and identification of Get4, Get2 and Sgt2 homologs, recombinant protein production, pilot *in vitro* binding assay, characterization of TA proteins in *G. intestinalis.*

2. Najdrová, V., Dohnálek, V., Voleman, L., Doležal, P. (unpublished). **Highly diverged pre-targeting complex in *Giardia intestinalis* reveals the ancestral presence of Bag6 and the evolution of GET pathway in eukaryotes.** Manuscript

My contribution: All experiments (except phylogenetic analyses), data analyses and interpretation, writing the manuscript

3. Najdrová, V., Stairs, C. W., Vinopalová, M., Voleman, L., Doležal, P. (2020). **The evolution of the Puf superfamily of proteins across the tree of eukaryotes**. BMC Biology, *18*(1), 77. https://doi.org/10.1186/S12915-020-00814-3

My contribution: Performing all experiments (except phylogenetic analyses and deconvolution of IFA images), data analyses and interpretation and writing the associated parts of the manuscript.

Summary and conclusion

The main aim of my Ph.D. study and this thesis was the characterization of post-translational protein transport of TA proteins into the endoplasmic reticulum of *G. intestinalis*. In particular, we focused on the GET pathway, which post-translationally targets and inserts TA proteins into the ER membrane. This pathway was discovered 15 years ago and is intensively studied in yeast and mammalian cell model systems (Schuldiner *et al.*, 2008; Stefanovic *et* Hegde, 2007). Some components of the GET pathway were found in *Arabidopsis thaliana* and recently in *Plasmodium falciparum*, but less has been known about the GET pathway outside the Opisthokonta group of eukaryotes (Kumar *et al.*, 2021; Srivastava *et al.*, 2017; Xing *et al.*, 2017). Therefore, the main goals of this thesis were: i) to identify homologs of GET pathway in *G. intestinalis*, ii) to characterize cytosolic complex of the GET pathway and iii) to reveal a broader view of the evolution of the GET pathway in the eukaryotes.

The central and most conserved component of GET pathway is Get3 ATPase, which serves as a shuttle between the cytosolic complex and the ER membrane insertase (Rome *et al.*, 2013; Simpson *et al.*, 2010; Stefanovic *et* Hegde, 2007; F. Wang *et al.*, 2010). We have identified Get3 homolog in *G. intestinalis* (*Gi*Get3) with sequence identities 42.14% and 44.67% to human and yeast homologues, respectively. We have used all available proteomes and transcriptomes to search Get3 homologs across eukaryotic diversity. Following phylogenetic analyses separated eukaryotic Get3 proteins into two clades (Fry *et al.*, 2022), as previously described in the analyses with smaller datasets (Xing *et al.*, 2017). This Get3 diversification suggests that LECA contained two Get3 homologs. Using the purified recombinant protein specific polyclonal antibody was raised against *Gi*Get3 for further protein characterization of *Gi*Get3 *in vivo* and *in vitro*. Cytosolic localization with partial ER signal of *Gi*Get3 was confirmed by immunofluorescence assay and western blot analyses of cellular fractions. This localization is consistent with the known localization and function of Get3 (Rome *et al.*, 2014; Stefanovic *et* Hegde, 2007).

The ability of *Gi*Get3 to bind and hydrolyse ATP was determined by *in vitro* ATPase assay (Fry *et al.*, 2022). Analogously, the point mutation in ATP binding of *Gi*Get3 site led to

enzyme inactivation as previously described (Mateja *et al.*, 2009; Suloway *et al.*, 2009). To show if *Gi*Get3 is capable of TA protein binding, we first predicted a subset of possible *G. intestinalis* TA proteins. Co-expression of selected TA proteins with *Gi*Get3 in *Escherichia coli* and subsequent affinity purification revealed a clear interaction of two TA proteins with *Gi*Get3. The hydrophobicity values (transmembrane tendency scale) of the two identified *G. intestinalis* substrates correspond to the hydrophobicity of known TA proteins, which are transported by the GET pathway (Fry *et al.*, 2021; Rao *et al.*, 2016).

In order to identify *Gi*Get3 interaction partners, we adapted previously established method based of *in vivo* enzymatic tagging technique by biotin ligase (BirA) from *E. coli* (Martincová *et al.*, 2015). Instead of chemical crosslinking and purification under strong denaturing conditions, we used milder native conditions without crosslinkers. This approach allowed us to find additional components (*Gi*Get2 and *Gi*Get4) of the GET pathway in *G. intestinalis* that we could not identify by bioinformatic approaches. This, however, was possible for Sgt2 homologue in *G. intestinalis* (*Gi*Sgt2).

Several structural studies of opisthokont Get3 delineated current model of Get3 functional cycle, which includes nucleotide free, ATP bound, and TA protein bound states. Briefly, when the open state Get3 binds ATP, the conformation is changed to the closed state and the binding of TA protein is allowed. When Get3 with the TA protein is targeted to the ER membrane, the conformational change after ATP hydrolysis allows TA protein to be released and inserted into the ER membrane by Get1/Get2 complex (Bozkurt *et al.*, 2009; Gristick *et al.*, 2014; Mariappan *et al.*, 2011; Mateja *et al.*, 2009, 2015; Stefer *et al.*, 2011; Suloway *et al.*, 2009). However, some steps and conformational details were still incomplete. Specifically, how Get3 changes from an open to a closed conformation and how the hydrophobic groove is shielded before TA protein binding. In all current structural analyzes, these regions were observed without enough structural detail (Mateja *et al.*, 2009; Suloway *et al.*, 2009; Yamagata *et al.*, 2010).

The series of five structures of *Gi*Get3, including three functional states, complete the mechanistic picture of the Get3 cycle. Nucleotide free forms of *Gi*Get3 confirm previously describer fungal structures of open form of Get3 (Mateja *et al.*, 2009; Suloway

et al., 2009; Yamagata *et al.*, 2010). The semi-open form of *Gi*Get3, and *Gi*Get3 bound to ATP alone (without any client or binding partner), were first described conformations of any Get3 homologue. These structures also revealed how the hydrophobic binding groove is shielded in the apo Get3 states. Hence, the structural characterization of *Gi*Get3 with the bound TA protein in the post-hydrolysis state represents a missing piece of the Get3 cycle.

We have shown that apo states fluctuate in the cytosol between open, semi-open, and closed states until ATP binding stabilizes the closed conformation and creates a binding site for Get4 (Fry *et al.*, 2022). Rearrangements and subsequent opening of the binding groove allow binding of the TA protein, as previously described in yeast Get3/Get4/Get5 structure (Gristick *et al.*, 2014). Correct binding of TA protein to the Get3 binding groove induces hydrolysis, leading to opening of the Get3 dimer and subsequent release of Get4. Although Get3 slightly opens its dimer, the shielding of the TA protein by the hydrophobic groove remains undisturbed and the TA protein can be transferred to the ER membrane.

Finally, the identification of additional components of the GET pathway (*Gi*Sgt2, *Gi*Get4, and *Gi*Get2) in highly divergent organism such as a *G. intestinalis* indicates a broad conservation of the GET pathway across eukaryotes (Fry *et al.*, 2022).

These results are summarized in **Publication 1** "Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3" and accomplished first aim "Characterization of *G. intestinalis* Get3 homolog" and partly the second aim "Characterization of the GET pathway in *G. intestinalis*".

To complete the characterization of all of GET pathway components in *G. intestinalis*, the same native *in vivo* tagging method was applied for the isolation of *Gi*Sgt2 and *Gi*Get4 identified in Fry *et al.* (2022). Proteins specifically co-purified with *Gi*Sgt2 and *Gi*Get4 confirmed the interaction of known GET pathway components *Gi*Get3, *Gi*Get4, and *Gi*Sgt2 with two significantly enriched proteins of unknown function (GL50803_19378 and GL50803_5069). Subsequent inspection of these proteins revealed that they represent highly divergent homologs of *Gi*Get5 and *Gi*Bag6, respectively. Analogous purification of *Gi*Get5 and *Gi*Bag6 validated that the GET pathway in *G. intestinalis* is composed of *Gi*Sgt2, *Gi*Get3, *Gi*Get4, *Gi*Get5 and *Gi*Bag6 (Najdrová *et al.*, unpublished).

In Sgt2-specific dataset, we have also identified significantly enriched homolog of cytosolic Hsp70 (Ssa1) (Najdrová *et al.*, unpublished), which is a known member of GET pathway (Stefanovic *et* Hegde, 2007). Isolation of cytosolic Hsp70 as a protein partner of Sgt2 in *G. intestinalis* suggested that cytosolic Hsp70 could be the most upstream member of the GET pathway as it was recently shown in the yeast system (Cho *et* Shan, 2018). In the same dataset, we have also found homolog of Ydj1 the J-domain containing co-chaperone, which participates in regulation of cytosolic Hsp70 (Cho *et al.*, 2021; Cho *et* Shan, 2018). Therefore, the involvement of chaperones and co-chaperones in GET pathway seems to be evolutionary conserved.

Further analyses of *Gi*Get5 sequence shown structural conservation of the C-terminal part of human, yeast and *G. intestinalis* proteins (Najdrová *et al.*, unpublished). *Gi*Get5 lacks the N-terminal part that is present in yeast Get5 and which is herein responsible for Get4 binding. Our identification of missing N-terminal domain in *Gi*Get5 was supported by *Gi*Bag6 identification, which shows that *G. intestinalis* pre-targeting complex composition is to the human counterpart, where Get4 and Get5 homologs are bound by Bag6 separately (Bozkurt *et al.*, 2010; Chang *et al.*, 2010; Chartron *et al.*, 2010; Mariappan *et al.*, 2010; Mock *et al.*, 2015). *Gi*Bag6 sequence is also highly divergent, but sequence alignment revealed conserved C-terminal part of this protein (Najdrová *et al.*, unpublished), which is responsible for Get4 and Get5 binding (Mock *et al.*, 2015a). *Gi*Bag6 dual cytosolic and nuclear localization of *Gi*Bag6 (Najdrová *et al.*, unpublished) suggests additional functional similarities with human Bag6 function, but detailed conditions for nuclear localization must be further examined (Manchen *et* Hubberstey, 2001; Nguyen *et al.*, 2008). Another possible function of *Gi*Bag6 was outlined by detection of a small subunit of DNA-directed RNA polymerases (RPB10; GL50803_14413), Lysyl-tRNA synthetase (GL50803_16766), putative tRNA-dihydrouridine (47) synthase (GL50803_3565) and ribosomal subunit L38e (GL50803_34093). All these proteins were co-purified with *Gi*Bag6 and suggest a potential role for *Gi*Bag6 in the regulation of transcription and translation (Najdrová *et al.*, unpublished).

Comparison of datasets obtained from the purification of *Gi*SGt2, *Gi*Get4, and *Gi*Bag6, revealed a subset of proteasomal and proteasome-interacting proteins. Composition of *G. intestinalis* proteasome is known, although three (Rpn12, Rpn13, Rpn15) proteasomal proteins have not been identified yet (Jerlström-Hultqvist *et al.*, 2012). It was previously shown, that human Sgt2 binds to the Rpn13 subunit of the proteasome (Leznicki *et al.*, 2015) and Bag6 (*Xenopus*) binds to the proteasomal subunit Rpn10 (Kikukawa *et al.*, 2005). However, situation in *G. intestinalis* remains complicated, because Rpn13 subunit has not been identified yet in *G. intestinalis* (Jerlström-Hultqvist *et al.*, 2012) and Rpn10 protein was not identified as significantly enriched in any of our datasets (Najdrová *et al.*, unpublished). Regardless these facts, our data suggest that interaction of the pre-targeting complex with proteasome is a universal and conserved mechanism of the misfolded protein response.

All experimentally obtained divergent sequences of GET pathway components in *G. intestinalis* were included into HMM profiles for the analyses of the conservation of GET pathway components across eukaryotes. The central element of the GET pathway, Get3 ATPase, was identified in all eukaryotic groups (Najdrová *et al.*, unpublished), which is consistent with a previously established homology search, which shows high conservation of Get3 in most of the eukaryotic groups and presence of the homologues in archaea and bacteria (Farkas *et al.*, 2019). Similarly, our analyses confirmed that Get1 is evolutionarily highly conserved member of the Oxa1 superfamily. Surprisingly, in contrast to other eukaryotes, which contain three different paralogues of the Oxa1 superfamily in the ER (Get1, EMC3 and TMCO1) (Anghel *et al.*, 2017), Get1 is the only member of the superfamily identified in *G. intestinalis*. Our results together with the fact that the EMC complex is missing in *G. intestinalis* (Wideman, 2015), raise the question of how other membrane proteins are transported into the ER membrane and how the loss of the TMCO1 and EMC complex is replaced. The common origin of all three subfamilies (Get1, TMCO1, EMC3) suggests their potential functional overlap. Indeed, recently data showed that mitochondria targeted EMC3-EMC6 and Get1-Get2 proteins were able to partially replace the function of mitochondrial Oxa1 (Güngör *et al.*, 2022). Homology search for Get2 homologs was previously carried out among metazoan, plants, and fungi (Borgese, 2020).

Our search was applied in expanded datasets and confirmed the conservation of Get2 in all eukaryotic groups. Similarly, our search showed that Sgt2 and Get4 represent highly conserved proteins of pre-targeting complex and were identified in all eukaryotic groups (Najdrová *et al.*, unpublished).

However, reconstructing the occurrence of Get5 and Bag6 in eukaryotes has proven very difficult. In case of Bag6, despite being considered as absent outside metazoan, we have been able to define the C-terminal part of Bag6, which has also shown as the minimal domain necessary for Bag6 function (Mock *et al.*, 2015, 2017) as a conserved signature of Bag6 homologues across eukaryotic tree of life. Similarly, the C-terminal part of Get5 could be used for Get5 determination and the homology search and our homology searches revealed conservation of the protein in eukaryotes (Najdrová *et al.*, unpublished). Our data thus strongly suggest that LECA contained all GET pathway components (Sgt2, Get1-Get5, Bag6) and therefore Bag6 was secondarily lost in some lineages like yeast. Our results from both experimental and bioinformatical studies suggest that the GET pathway is an ancestral pathway and has ancient and crucial role in the biology of the eukaryotic cell.

These results are summarized in **Manuscript 2** "Highly diverged pre-targeting complex in *Giardia intestinalis* reveals the ancestral presence of Bag6 and the evolution of GET pathway in eukaryotes". These data accomplished second and third aims "Characterization of GET pathway in *G. intestinalis*" and "Reconstruction of the evolution of GET proteins in eukaryotes", respectively.

In addition to protein targeting signals, the localization of protein can also be directed by the site-specific translation that takes advantage of mRNA targeting towards the places of action (Cui e*t* Palazzo, 2014). Our preliminary data showed that episomally expressed Sec20 with native 3´UTR partially localized the protein to its correct compartment (Voleman, 2011). These observations led us to investigate the potential role of 3´UTR and RNA-binding proteins in localization of proteins in *G. intestinalis*. We have decided to examine the role of RNA binding proteins from the Puf family. Concretely, we were inspired by the ability of yeast Puf3 to localize mRNAs encoding some mitochondrial proteins on

the mitochondrial surface and thus play role in the mitochondrial biogenesis (Saint-Georges *et al.*, 2008).

We identified six Puf proteins in *G. intestinalis* and classified them as four classical Puf homologues, one Nop9, and one PUM3. All Puf proteins were episomally expressed with a BAP tag to determine their cellular localization and function. Unfortunately, the expression of almost all constructs (not Puf4) was highly unstable and diminished rapidly after establishing stable cell line. However, Puf proteins in *G. intestinalis* were analyzed by combining bioinformatic prediction and immunofluorescence assay (Najdrová *et al.*, 2020). Unexpectedly, we observed that PUM3 is localized in the nuclear periphery and Nop9 in the cytosol, which is different from other cell systems (Andersen *et al.*, 2002; Gu *et al.*, 2004; Liang *et al.*, 2018) Puf1, Puf2, and Puf4 were localized in different kinds of vesicular structures, but they did not colocalize with Sec20 signal. Only Puf3 showed predicted cytosolic localization (Najdrová *et al.*, 2020).

Due to the unstable expression of the *G. intestinalis* Puf proteins, their putative target mRNAs were predicted by bioinformatics. We have predicted cognate mRNAs for all Pufs in *G. intestinalis* genome (Najdrová *et al.*, 2020). Compared to other organisms, the number of target mRNAs is smaller in *G. intestinalis*, which could be due to extremely short UTRs in *G. intestinalis*, and which thereby limit in the available area for interaction (Elmendorf *et al.*, 2001; Hogan *et al.*, 2015). Although none of the predicted targets are TA proteins and we have not shown any role of Puf in the targeting of TA proteins, we have characterized Puf proteins in *G. intestinalis* and showed the distribution of Puf proteins in eukaryotes.

These results are summarized in **Publication 3** "The evolution of the Puf superfamily of proteins across the tree of eukaryotes" and complete the fourth aim "Determine Puf homologs in *G. intestinalis* and their potential role in mRNA/protein targeting".

List of references

Acosta-Virgen, K., Chávez-Munguía, B., Talamás-Lara, D., Lagunes-Guillén, A., Martínez-Higuera, A., Lazcano, A., Martínez-Palomo, A., Espinosa-Cantellano, M. (2018). *Giardia lamblia*: Identification of peroxisomal-like proteins. *Experimental Parasitology*, *191*, 36–43. https://doi.org/10.1016/J.EXPPARA.2018.06.006

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Gall, L. le, Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., Spiegel, F. W. (2012). The revised classification of eukaryotes. *The Journal of Eukaryotic Microbiology*, *59*(5), 429–514. https://doi.org/10.1111/J.1550-7408.2012.00644.X

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, Y. W., Steen, H., Mann, M., Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Current Biology: CB*, *12*(1), 1–11. https://doi.org/10.1016/S0960-9822(01)00650-9

Anghel, S. A., McGilvray, P. T., Hegde, R. S., Keenan, R. J. (2017). Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Reports*, *21*(13). https://doi.org/10.1016/j.celrep.2017.12.006

Ankarklev, J., Jerlström-Hultqvist, J., Ringqvist, E., Troell, K., Svärd, S. G. (2010). Behind the smile: cell biology and disease mechanisms of *Giardia* species. *Nature Reviews Microbiology 2010 8:6*, *8*(6), 413–422. https://doi.org/10.1038/nrmicro2317

Ankarklev, J., Svärd, S. G., Lebbad, M. (2012). Allelic sequence heterozygosity in single *Giardia* parasites. *BMC Microbiology*, *12*. https://doi.org/10.1186/1471-2180-12-65

Aviram, N., Ast, T., Costa, E. A., Arakel, E. C., Chuartzman, S. G., Jan, C. H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., Zimmermann, R., Schwappach, B., Weissman, J. S., Schuldiner, M. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature*, *540*(7631), 134–138. https://doi.org/10.1038/nature20169

Aviram, N., Schuldiner, M. (2017). Targeting and translocation of proteins to the endoplasmic reticulum at a glance. *Journal of Cell Science*, *130*(24), 4079–4085. https://doi.org/10.1242/JCS.204396

Balaña-Fouce, R., Reguera, R. M. (2007). RNA interference in *Trypanosoma brucei*: a high-throughput engine for functional genomics in trypanosomatids? *Trends in Parasitology*, *23*(8), 348–351. https://doi.org/10.1016/J.PT.2007.06.008

Baum, D. A., Baum, B. (2014). An inside-out origin for the eukaryotic cell. *BMC Biology*, *12*(1). https://doi.org/10.1186/S12915-014-0076-2

Beilharz, T., Egan, B., Silver, P. A., Hofmann, K., Lithgow, T. (2003). Bipartite Signals Mediate Subcellular Targeting of Tail-anchored Membrane Proteins in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *278*(10), 8219–8223. https://doi.org/10.1074/JBC.M212725200

Benarroch, R., Austin, J. M., Ahmed, F., Isaacson, R. L. (2019). The roles of cytosolic quality control proteins, SGTA and the BAG6 complex, in disease. *Advances in Protein Chemistry and Structural Biology*, *114*, 265–313. https://doi.org/10.1016/BS.APCSB.2018.11.002

Bernander, R., Palm, J. E. D., Svärd, S. G. (2001). Genome ploidy in different stages of the *Giardia lamblia* life cycle. *Cellular Microbiology*, *3*(1), 55–62. https://doi.org/10.1046/J.1462-5822.2001.00094.X

Bingham, A. K., Meyer, E. A. (1979). *Giardia* excystation can be induced *in vitro* in acidic solutions. *Nature*, *277*(5694), 301–302. https://doi.org/10.1038/277301A0

Borgese, N. (2020). Searching for remote homologs of CAML among eukaryotes. *Traffic*, *21*(10), 647–658. https://doi.org/10.1111/tra.12758

Borgese, N., Brambillasca, S., Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. *Current Opinion in Cell Biology*, *19*(4), 368–375. https://doi.org/10.1016/J.CEB.2007.04.019

Borgese, N., Colombo, S., Pedrazzini, E. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. *The Journal of Cell Biology*, *161*(6), 1013–1019. https://doi.org/10.1083/JCB.200303069

Borgese, N., Coy-Vergara, J., Colombo, S. F., Schwappach, B. (2019). The Ways of Tails: the GET Pathway and more. *The Protein Journal*, *38*(3). https://doi.org/10.1007/s10930-019-09845-4

Borgese, N., Righi, M. (2010). Remote Origins of Tail-Anchored Proteins. *Traffic*, *11*(7), 877–885. https://doi.org/10.1111/j.1600-0854.2010.01068.x

Borowska, M. T., Dominik, P. K., Anghel, S. A., Kossiakoff, A. A., Keenan, R. J. (2015). A YidC-like Protein in the Archaeal Plasma Membrane. *Structure*, *23*(9), 1715–1724. https://doi.org/10.1016/j.str.2015.06.025

Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favaloro, V., Rippe, K., Hurt, E., Dobberstein, B., Sinning, I. (2009). Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21131–21136. https://doi.org/10.1073/PNAS.0910223106

Bozkurt, G., Wild, K., Amlacher, S., Hurt, E., Dobberstein, B., Sinning, I. (2010). The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis. *FEBS Letters*, *584*(8), 1509–1514. https://doi.org/10.1016/J.FEBSLET.2010.02.070

Brito, G. C., Schormann, W., Gidda, S. K., Mullen, R. T., Andrews, D. W. (2019). Genome-wide analysis of *Homo sapiens*, *Arabidopsis thaliana*, and *Saccharomyces cerevisiae* reveals novel attributes of tail-anchored membrane proteins. *BMC Genomics*, *20*(1). https://doi.org/10.1186/S12864-019-6232-X

Burri, L., Lithgow, T. (2004). A complete set of SNAREs in yeast. *Traffic*, *5*(1), 45–52. https://doi.org/10.1046/J.1600-0854.2003.00151.X

Burri, L., Varlamov, O., Doege, C. A., Hofmann, K., Beilharz, T., Rothman, J. E., Söllner, T. H., Lithgow, T. (2003). A SNARE required for retrograde transport to the endoplasmic reticulum. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 9873–9877. https://doi.org/10.1073/PNAS.1734000100

Campbell, Z. T., Valley, C. T., Wickens, M. (2014). A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nature Structural Molecular Biology*, *21*(8), 732–738. https://doi.org/10.1038/nsmb.2847

Carpenter, M. L., Assaf, Z. J., Gourguechon, S., Cande, W. Z. (2012). Nuclear inheritance and genetic exchange without meiosis in the binucleate parasite *Giardia intestinalis*. *Journal of Cell Science*, *125*(Pt 10), 2523–2532. https://doi.org/10.1242/JCS.103879

Carpenter, M. L., Cande, W. Z. (2009). Using morpholinos for gene knockdown in *Giardia intestinalis*. *Eukaryotic Cell*, *8*(6), 916–919. https://doi.org/10.1128/EC.00041-09

Carranza, P. G., Feltes, G., Ropolo, A., Quintana, S. M. C., Touz, M. C., Luján, H. D. (2002). Simultaneous expression of different variant-specific surface proteins in single *Giardia lamblia* trophozoites during encystation. *Infection and Immunity*, *70*(9), 5265–5268. https://doi.org/10.1128/IAI.70.9.5265-5268.2002

Carvalho, H. J. F., del Bondio, A., Maltecca, F., Colombo, S. F., Borgese, N. (2019). The WRB Subunit of the Get3 Receptor is Required for the Correct Integration of its Partner CAML into the ER. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-48363-2

Casson, J., McKenna, M., Haßdenteufel, S., Aviram, N., Zimmerman, R., High, S. (2017). Multiple pathways facilitate the biogenesis of mammalian tail-anchored proteins. *Journal of Cell Science*, *130*(22), 3851–3861. https://doi.org/10.1242/jcs.207829

Chang, Y.-W., Chuang, Y.-C., Ho, Y.-C., Cheng, M.-Y., Sun, Y.-J., Hsiao, C.-D., Wang, C. (2010). Crystal Structure of Get4-Get5 Complex and Its Interactions with Sgt2, Get3, and Ydj1. *Journal of Biological Chemistry*, *285*(13), 9962-9970. https://doi.org/10.1074/jbc.M109.087098

Chartron, J. W., Clemons, W. M., Suloway, C. J. M. (2012). The complex process of GETting tail-anchored membrane proteins to the ER. *Current Opinion in Structural Biology*, *22*(2), 217–224. https://doi.org/10.1016/j.sbi.2012.03.001

Chartron, J. W., Gonzalez, G. M., Clemons, W. M. (2011). A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *Journal of*

Biological Chemistry, *286*(39), 34325–34334. https://doi.org/10.1074/jbc.M111.277798

Chartron, J. W., Suloway, C. J. M., Zaslaver, M., Clemons, W. M. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12127–12132. https://doi.org/10.1073/pnas.1006036107

Chartron, J. W., VanderVelde, D. G., Clemons, W. M. (2012). Structures of the Sgt2/SGTA Dimerization Domain with the Get5/UBL4A UBL Domain Reveal an Interaction that Forms a Conserved Dynamic Interface. *Cell Reports*, *2*(6), 1620–1632. https://doi.org/10.1016/j.celrep.2012.10.010

Chen, C. M., Misra, T. K., Silver, S., Rosen, B. P. (1986). Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *Journal of Biological Chemistry*, *261*(32), 15030–15038. https://doi.org/10.1016/s0021-9258(18)66824-3

Chio, U. S., Cho, H., Shan, S. O. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annual Review of Cell and Developmental Biology*, *33*, 417–438. https://doi.org/10.1146/annurev-cellbio-100616-060839

Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., Hegde, R. S. (2018). EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell*, *175*(6), 1507-1519.e16. https://doi.org/10.1016/j.cell.2018.10.009

Cho, H., Shan, S. (2018). Substrate relay in an Hsp70-cochaperone cascade safeguards tail-anchored membrane protein targeting. *The EMBO Journal*, *37*(16). https://doi.org/10.15252/embj.201899264

Cho, H., Shim, W. J., Liu, Y., Shan, S. O. (2021). J-domain proteins promote client relay from Hsp70 during tail-anchored membrane protein targeting. *The Journal of Biological Chemistry*, *296*, 100546. https://doi.org/10.1016/J.JBC.2021.100546

Colombo, S. F., Cardani, S., Maroli, A., Vitiello, A., Soffientini, P., Crespi, A., Bram, R. F., Benfante, R., Borgese, N. (2016). Tail-anchored protein insertion in mammals function and reciprocal interactions of the two subunits of the trc40receptor. *Journal of*

Biological Chemistry, *291*(29), 15292–15306. https://doi.org/10.1074/jbc.M115.707752

Cooper, M. A., Adam, R. D., Worobey, M., Sterling, C. R. (2007). Population genetics provides evidence for recombination in *Giardia*. *Current Biology*, *17*(22), 1984–1988. https://doi.org/10.1016/J.CUB.2007.10.020

Costello, J. L., Castro, I. G., Camões, F., Schrader, T. A., McNeall, D., Yang, J., Giannopoulou, E.-A., Gomes, S., Pogenberg, V., Bonekamp, N. A., Ribeiro, D., Wilmanns, M., Jedd, G., Islinger, M., Schrader, M. (2017). Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *Journal of Cell Science*, *130*(9), 1675-1687. https://doi.org/10.1242/jcs.200204

Cui, X. A., Palazzo, A. F. (2014). Localization of mRNAs to the endoplasmic reticulum. *Wiley Interdisciplinary Reviews. RNA*, *5*(4), 481–492. https://doi.org/10.1002/wrna.1225

Cui, X. A., Zhang, H., Ilan, L., Liu, A. X., Kharchuk, I., Palazzo, A. F. (2015). mRNA encoding Sec61β, a tail-anchored protein, is localized on the endoplasmic reticulum. *Journal of Cell Science*, *128*(18), 3398–3410. https://doi.org/10.1242/jcs.168583

Cui, X. A., Zhang, H., Palazzo, A. F. (2012). p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. *PLoS Biology*, *10*(5), e1001336. https://doi.org/10.1371/journal.pbio.1001336

Dalbey, R. E., Kuhn, A., Zhu, L., Kiefer, D. (2014). The membrane insertase YidC. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1843*(8), 1489–1496. https://doi.org/10.1016/j.bbamcr.2013.12.022

Darby, J. F., Krysztofinska, E. M., Simpson, P. J., Simon, A. C., Leznicki, P., Sriskandarajah, N., Bishop, D. S., Hale, L. R., Alfano, C., Conte, M. R., Martínez-Lumbreras, S., Thapaliya, A., High, S., Isaacson, R. L. (2014). Solution structure of the SGTA dimerisation domain and investigation of its interactions with the ubiquitin-like domains of BAG6 and UBL4A. *PloS One*, *9*(11), e113281. https://doi.org/10.1371/JOURNAL.PONE.0113281

Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M. R., Li, C., Tanaka Hall, T. M., Wang, Z. (2011). Specific and modular binding code for cytosine recognition in

Pumilio/FBF (PUF) RNA-binding domains. *The Journal of Biological Chemistry*, *286*(30), 26732–26742. https://doi.org/10.1074/JBC.M111.244889

Ebneter, J. A., Heusser, S. D., Schraner, E. M., Hehl, A. B., Faso, C. (2016). Cyst-Wall-Protein-1 is fundamental for Golgi-like organelle neogenesis and cyst-wall biosynthesis in *Giardia lamblia*. *Nature Communications*, *7*(1), 13859. https://doi.org/10.1038/ncomms13859

Egeo, A., Mazzocco, M., Sotgia, F., Arrigo, P., Oliva, R., Bergonòn, S., Nizetic, D., Rasore-Quartino, A., Scartezzini, P. (1998). Identification and characterization of a new human cDNA from chromosome 21q22.3 encoding a basic nuclear protein. *Human Genetics*, *102*(3), 289–293. https://doi.org/10.1007/s004390050693

Einarsson, E., Ma'ayeh, S., Svärd, S. G. (2016). An up-date on *Giardia* and giardiasis. *Current Opinion in Microbiology*, *34*, 47–52. https://doi.org/10.1016/J.MIB.2016.07.019

Elias, E. v., Quiroga, R., Gottig, N., Nakanishi, H., Nash, T. E., Neiman, A., Lujan, H. D. (2008). Characterization of SNAREs determines the absence of a typical Golgi apparatus in the ancient eukaryote *Giardia lamblia*. *The Journal of Biological Chemistry*, *283*(51), 35996–36010. https://doi.org/10.1074/JBC.M806545200

Elmendorf, H. G., Singer, S. M., Nash, T. E. (2001). The abundance of sterile transcripts in *Giardia lamblia*. *Nucleic Acids Research*, *29*(22), 4674–4683. https://doi.org/10.1093/NAR/29.22.4674

Farkas, Á., de Laurentiis, E. I., Schwappach, B. (2019). The natural history of Get3-like chaperones. *Traffic*, *20*(5), 311–324. https://doi.org/10.1111/tra.12643

Faso, C., Hehl, A. B. (2011). Membrane trafficking and organelle biogenesis in *Giardia lamblia*: Use it or lose it. *International Journal for Parasitology*, *41*(5), 471–480. https://doi.org/10.1016/J.IJPARA.2010.12.014

Faso, C., Konrad, C., Schraner, E. M., Hehl, A. B. (2013). Export of cyst wall material and Golgi organelle neogenesis in *Giardia lamblia* depend on endoplasmic reticulum exit sites. *Cellular Microbiology*, *15*(4), 537–553. https://doi.org/10.1111/CMI.12054

Favaloro, V., Spasic, M., Schwappach, B., Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *Journal of Cell Science*, *121*(11), 1832–1840. https://doi.org/10.1242/jcs.020321

Fernandes, J. M. O., Macqueen, D. J., Lee, H.-T., Johnston, I. A. (2008). Genomic, evolutionary, and expression analyses of cee, an ancient gene involved in normal growth and development. *Genomics*, *91*(4), 315–325. https://doi.org/10.1016/j.ygeno.2007.10.017

Figueiredo Costa, B., Cassella, P., Colombo, S. F., Borgese, N. (2018). Discrimination between the endoplasmic reticulum and mitochondria by spontaneously inserting tail-anchored proteins. *Traffic*, *19*(3), 182–197. https://doi.org/10.1111/TRA.12550

Fleischer, T. C., Weaver, C. M., McAfee, K. J., Jennings, J. L., Link, A. J. (2006). Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. *Genes Development*, *20*(10), 1294–1307. https://doi.org/10.1101/GAD.1422006

Franzén, O., Jerlström-Hultqvist, J., Einarsson, E., Ankarklev, J., Ferella, M., Andersson, B., Svärd, S. G. (2013). Transcriptome profiling of *Giardia intestinalis* using strand-specific RNA-seq. *PLoS Computational Biology*, *9*(3). https://doi.org/10.1371/JOURNAL.PCBI.1003000

Fry, M. Y., Najdrová, V., Maggiolo, A. O., Saladi, S. M., Doležal, P., Clemons, W. M. (2022). Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. *Nature Structural Molecular Biology*, *29*(8), 820-830. https://doi.org/10.1038/s41594-022-00798-4

Fry, M. Y., Saladi, S. M., Cunha, A., Clemons, W. M. (2021). Sequence-based features that are determinant for tail-anchored membrane protein sorting in eukaryotes. *Traffic 22*(9), 306–318. https://doi.org/10.1111/TRA.12809

Gemmer, M., Förster, F. (2020). A clearer picture of the ER translocon complex. *Journal of Cell Science*, *133*(3), jcs231340. https://doi.org/10.1242/JCS.231340

Gillin, F. D., Reiner, D. S., Boucher, S. E. (1988). Small-intestinal factors promote encystation of *Giardia lamblia in vitro*. *Infection and Immunity*, *56*(3), 705–707. https://doi.org/10.1128/IAI.56.3.705-707.1988

Gristick, H. B., Rao, M., Chartron, J. W., Rome, M. E., Shan, S. O., Clemons, W. M. (2014). Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4. *Nature Structural and Molecular Biology*, *21*(5), 437–442. https://doi.org/10.1038/nsmb.2813

Gristick, H. B., Rome, M. E., Chartron, J. W., Rao, M., Hess, S., Shan, S. O., Clemons, W. M. (2015). Mechanism of Assembly of a Substrate Transfer Complex during Tail-anchored Protein Targeting. *The Journal of Biological Chemistry*, *290*(50), 30006–30017. https://doi.org/10.1074/JBC.M115.677328

Gu, W., Deng, Y., Zenklusen, D., Singer, R. H. (2004). A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. *Genes Development*, *18*(12), 1452–1465. https://doi.org/10.1101/GAD.1189004

Guna, A., Hegde, R. S. (2018). Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. *Current* Biology, *28*(8), R498–R511. https://doi.org/10.1016/J.CUB.2018.02.004

Guna, A., Volkmar, N., Christianson, J. C., Hegde, R. S. (2018). The ER membrane protein complex is a transmembrane domain insertase. *Science*, *359*(6374), 470–473. https://doi.org/10.1126/science.aao3099

Güngör, B., Flohr, T., Garg, S. G., Herrmann, J. M. (2022). The ER membrane complex (EMC) can functionally replace the Oxa1 insertase in mitochondria. *PLoS Biology*, *20*(3), e3001380. https://doi.org/10.1371/JOURNAL.PBIO.3001380

Gupta, A., Becker, T. (2021). Mechanisms and pathways of mitochondrial outer membrane protein biogenesis. *Biochimica et Biophysica Acta - Bioenergetics*, *1862*(1), 148323. https://doi.org/10.1016/J.BBABIO.2020.148323

Hagen, K. D., McInally, S. G., Hilton, N. D., Dawson, S. C. (2020). Microtubule organelles in *Giardia*. *Advances in Parasitology*, *107*, 25–96. https://doi.org/10.1016/BS.APAR.2019.11.001

Halic, M., Becker, T., Pool, M. R., Spahn, C. M. T., Grassucci, R. A., Frank, J., Beckmann, R. (2004). Structure of the signal recognition particle interacting with the elongation-arrested ribosome. *Nature*, *427*(6977), 808–814. https://doi.org/10.1038/NATURE02342

Hanevik, K., Wensaas, K. A., Rortveit, G., Eide, G. E., Mørch, K., Langeland, N. (2014). Irritable Bowel Syndrome and Chronic Fatigue 6 Years After *Giardia* Infection: A Controlled Prospective Cohort Study. *Clinical Infectious Diseases*, *59*(10), 1394–1400. https://doi.org/10.1093/CID/CIU629

Haßdenteufel, S., Sicking, M., Schorr, S., Aviram, N., Fecher-Trost, C., Schuldiner, M., Jung, M., Zimmermann, R., Lang, S. (2017). hSnd2 protein represents an alternative targeting factor to the endoplasmic reticulum in human cells. *FEBS Letters*, *591*(20), 3211–3224. https://doi.org/10.1002/1873-3468.12831

He, X., Zhang, Y., Yang, L., Feng, J., Yang, S., Li, T., Zhong, T., Li, Q., Xie, W., Liu, M., Zhou, J., Li, D., Xie, S. (2020). BAG6 is a novel microtubule-binding protein that regulates ciliogenesis by modulating the cell cycle and interacting with γ-tubulin. *Experimental Cell Research*, *387*(1), 111776. https://doi.org/10.1016/J.YEXCR.2019.111776

Hegde, R. S., Keenan, R. J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. In *Nature Reviews Molecular Cell Biology*, *12*(12), 787–798. https://doi.org/10.1038/nrm3226

Hessa, T., Sharma, A., Mariappan, M., Eshleman, H. D., Gutierrez, E., Hegde, R. S. (2011). Protein targeting and degradation are coupled for elimination of mislocalized proteins. *Nature*, *475*(7356), 394–399. https://doi.org/10.1038/nature10181

Hogan, G. J., Brown, P. O., Herschlag, D. (2015). Evolutionary Conservation and Diversification of Puf RNA Binding Proteins and Their mRNA Targets. *PLoS Biology*, *13*(11), e1002307. https://doi.org/10.1371/JOURNAL.PBIO.1002307

Horáčková, V., Voleman, L., Hagen, K. D., Petru, M., Vinopalová, M., Weisz, F., Janowicz, N., Marková, L., Motycková, A., Najdrová, V., Tumová, P., Dawson, S. C., Doležal, P. (2022). Efficient CRISPR/Cas9-mediated gene disruption in the tetraploid protist *Giardia intestinalis*. *Open Biology*, *12*(4), 210361. https://doi.org/10.1098/RSOB.210361

Hsieh, H.-H., Shan, S.-O., Zimmermann, R., Lang, S., Hsieh, H.-H., Shan, S.-O. (2021). Fidelity of Cotranslational Protein Targeting to the Endoplasmic Reticulum. *International Journal of Molecular Sciences*, *23*(1), 281. https://doi.org/10.3390/IJMS23010281

Hu, Z., Potthoff, B., Hollenberg, C. P., Ramezani-Rad, M. (2006). Mdy2, a ubiquitin-like (UBL)-domain protein, is required for efficient mating in Saccharomyces cerevisiae. *Journal of Cell Science*, *119*(Pt 2), 326–338. https://doi.org/10.1242/jcs.02754

Iwanejko, L., Smith, K. N., Loeillet, S., Nicolas, A., Fabre, F. (1999). Disruption and functional analysis of six ORFs on chromosome XV: YOL117w, YOL115w (TRF4), YOL114c, YOL112w (MSB4), YOL111c and YOL072w. *Yeast*, *15*(14), 1529–1539. https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1529::AID-YEA457>3.0.CO;2-Y

Jedelský, P. L., Doležal, P., Rada, P., Pyrih, J., Šmíd, O., Hrdý, I., Šedinová, M., Marcinčiková, M., Voleman, L., Perry, A. J., Beltrán, N. C., Lithgow, T., Tachezy, J. (2011). The minimal proteome in the reduced mitochondrion of the parasitic protist *Giardia intestinalis*. *PloS One*, *6*(2), e17285. https://doi.org/10.1371/JOURNAL.PONE.0017285

Jerlström-Hultqvist, J., Stadelmann, B., Birkestedt, S., Hellman, U., Svärd, S. G. (2012). Plasmid vectors for proteomic analyses in *Giardia*: purification of virulence factors and analysis of the proteasome. *Eukaryotic Cell*, *11*(7), 864–873. https://doi.org/10.1128/EC.00092-12

Jonikas, M. C., Collins, S. R., Denic, V., Oh, E., Quan, E. M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J. S., Schuldiner, M. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science*, *323*(5922), 1693–1697. https://doi.org/10.1126/science.1167983

Kalbfleisch, T., Cambon, A., Wattenberg, B. W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic*, *8*(12), 1687–1694. https://doi.org/10.1111/J.1600-0854.2007.00661.X

Kämper, N., Franken, S., Temme, S., Koch, S., Bieber, T., Koch, N. (2012). γ-Interferon-regulated chaperone governs human lymphocyte antigen class II expression. *FASEB*

Journal: Official Publication of the Federation of American Societies for Experimental Biology, *26*(1), 104–116. https://doi.org/10.1096/FJ.11-189670

Keister, D. B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *77*(4), 487–488. https://doi.org/10.1016/0035-9203(83)90120-7

Kikukawa, Y., Minami, R., Shimada, M., Kobayashi, M., Tanaka, K., Yokosawa, H., Kawahara, H. (2005). Unique proteasome subunit Xrpn 10c is a specific receptor for the antiapoptotic ubiquitin-like protein Scythe. *FEBS Journal*, *272*(24), 6373–6386. https://doi.org/10.1111/j.1742-4658.2005.05032.x

Klausner, R. D., Donaldson, J. G., Lippincott-Schwartz, J. (1992). Brefeldin A: insights into the control of membrane traffic and organelle structure. *The Journal of Cell Biology*, *116*(5), 1071–1080. https://doi.org/10.1083/JCB.116.5.1071

Kriechbaumer, V., Shaw, R., Mukherjee, J., Bowsher, C. G., Harrison, A. M., Abell, B. M. (2009). Subcellular distribution of tail-anchored proteins in *Arabidopsis*. *Traffic*, *10*(12), 1753–1764. https://doi.org/10.1111/J.1600-0854.2009.00991.X

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/JMBI.2000.4315

Krysztofinska, E. M., Evans, N. J., Thapaliya, A., Murray, J. W., Morgan, R. M. L., Martinez-Lumbreras, S., Isaacson, R. L. (2017). Structure and interactions of the TPR domain of Sgt2 with yeast chaperones and Ybr137wp. *Frontiers in Molecular Biosciences*, *4*,68. https://doi.org/10.3389/fmolb.2017.00068

Kubota, K., Yamagata, A., Sato, Y., Goto-Ito, S., Fukai, S. (2012). Get1 stabilizes an open dimer conformation of Get3 ATPase by binding two distinct interfaces. *Journal of Molecular Biology*, *422*(3), 366–375. https://doi.org/10.1016/j.jmb.2012.05.045

Kubota, S., Doi, H., Koyano, S., Tanaka, K., Komiya, H., Katsumoto, A., Ikeda, S., Hashiguchi, S., Nakamura, H., Fukai, R., Takahashi, K., Kunii, M., Tada, M., Takeuchi, H., Tanaka, F.

(2021). SGTA associates with intracellular aggregates in neurodegenerative diseases. *Molecular Brain*, *14*(1), 59. https://doi.org/10.1186/s13041-021-00770-1

Kumar, T., Maitra, S., Rahman, A., Bhattacharjee, S. (2021). A conserved guided entry of tail-anchored pathway is involved in the trafficking of a subset of membrane proteins in *Plasmodium falciparum*. *PLoS Pathogens*, *17*(11), e1009595. https://doi.org/10.1371/JOURNAL.PPAT.1009595

Kutay, U., Hartmann, E., Rapoport, T. A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends in Cell Biology*, *3*(3), 72–75. https://doi.org/10.1016/0962-8924(93)90066-A

Leznicki, P., Clancy, A., Schwappach, B., High, S. (2010). Bat3 promotes the membrane integration of tail-anchored proteins. *Journal of Cell Science*, *123*(13), 2170–2178. https://doi.org/10.1242/jcs.066738

Leznicki, P., High, S. (2012). SGTA antagonizes BAG6-mediated protein triage. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19214–19219. https://doi.org/10.1073/pnas.1209997109

Leznicki, P., Korac-Prlic, J., Kliza, K., Husnjak, K., Nyathi, Y., Dikic, I., High, S. (2015). Binding of SGTA to Rpn13 selectively modulates protein quality control. *Journal of Cell Science*, *128*(17), 3187–3196. https://doi.org/10.1242/JCS.165209/-/DC1

Li, F., Egea, P. F., Vecchio, A. J., Asial, I., Gupta, M., Paulino, J., Bajaj, R., Dickinson, M. S., Ferguson-Miller, S., Monk, B. C., Stroud, R. M. (2021). Highlighting membrane protein structure and function: A celebration of the Protein Data Bank. *The Journal of Biological Chemistry*, *296*, 100557. https://doi.org/10.1016/J.JBC.2021.100557

Li, L., Wang, C. C. (2004). Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. *The Journal of Biological Chemistry*, *279*(15), 14656–14664. https://doi.org/10.1074/JBC.M309879200

Li, Z., Lee, I., Moradi, E., Hung, N. J., Johnson, A. W., Marcotte, E. M. (2009). Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biology*, *7*(10), e1000213. https://doi.org/10.1371/JOURNAL.PBIO.1000213

Liang, X., Hart, K. J., Dong, G., Siddiqui, F. A., Sebastian, A., Li, X., Albert, I., Miao, J., Lindner, S. E., Cui, L. (2018). Puf3 participates in ribosomal biogenesis in malaria parasites. *Journal of Cell Science*, *131*(6), jcs212597. https://doi.org/10.1242/JCS.212597

Lin, K. F., Fry, M. Y., Saladi, S. M., Clemons, W. M. (2021). Molecular basis of tail-anchored integral membrane protein recognition by the cochaperone Sgt2. *The Journal of Biological Chemistry*, *296*, 100441. https://doi.org/10.1016/J.JBC.2021.100441

Lin, Z. Q., Gan, S. W., Tung, S. Y., Ho, C. C., Su, L. H., Sun, C. H. (2019). Development of CRISPR/Cas9-mediated gene disruption systems in *Giardia lamblia*. *PloS One*, *14*(3), e0213594. https://doi.org/10.1371/JOURNAL.PONE.0213594

Liou, S. T., Wang, C. (2005). Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. *Archives of Biochemistry and Biophysics*, *435*(2), 253–263. https://doi.org/10.1016/J.ABB.2004.12.020

Lujan, H. D., Marotta, A., Mowatt, M. R., Sciaky, N., Lippincott-Schwartz, J., Nash, T. E. (1995). Developmental Induction of Golgi Structure and Function in the Primitive Eukaryote *Giardia lamblia*. *Journal of Biological Chemistry*, *270*(9), 4612–4618. https://doi.org/10.1074/JBC.270.9.4612

Manchen, S. T., Hubberstey, A. v. (2001). Human Scythe contains a functional nuclear localization sequence and remains in the nucleus during staurosporine-induced apoptosis. *Biochemical and Biophysical Research Communications*, *287*(5), 1075–1082. https://doi.org/10.1006/BBRC.2001.5701

Marcial-Quino, J., Gómez-Manzo, S., Fierro, F., Rufino-González, Y., Ortega-Cuellar, D., Sierra-Palacios, E., Vanoye-Carlo, A., González-Valdez, A., Torres-Arroyo, A., Oria-Hernández, J., Reyes-Vivas, H. (2017). RNAi-Mediated Specific Gene Silencing as a Tool for the Discovery of New Drug Targets in *Giardia lamblia*; Evaluation Using the NADH Oxidase Gene. *Genes*, *8*(11), 303. https://doi.org/10.3390/GENES8110303

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R. J., Hegde, R. S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature*, *466*(7310), 1120–1124. https://doi.org/10.1038/nature09296

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R. S., Keenan, R. J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature*, *477*(7362), 61–69. https://doi.org/10.1038/nature10362

Martincová, E., Voleman, L., Pyrih, J., Žárský, V., Vondráčková, P., Kolísko, M., Tachezy, J., Doležal, P. (2015). Probing the Biology of *Giardia intestinalis* Mitosomes Using In Vivo Enzymatic Tagging. *Molecular and Cellular Biology*. *35*(16), 2864–2874. https://doi.org/10.1128/mcb.00448-15

Mateja, A., Paduch, M., Chang, H. Y., Szydlowska, A., Kossiakoff, A. A., Hegde, R. S., Keenan, R. J. (2015). Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science*, *347*(6226), 1152–1155. https://doi.org/10.1126/science.1261671

Mateja, A., Szlachcic, A., Downing, M. E., Dobosz, M., Mariappan, M., Hegde, R. S., Keenan, R. J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature*, *461*(7262), 361–366. https://doi.org/10.1038/nature08319

Mayerhofer, P. U. (2016). Targeting and insertion of peroxisomal membrane proteins: ER trafficking versus direct delivery to peroxisomes. *Biochimica et Biophysica Acta*, *1863*(5), 870–880. https://doi.org/10.1016/J.BBAMCR.2015.09.021

McDowell, M. A., Heimes, M., Fiorentino, F., Mehmood, S., Farkas, Á., Coy-Vergara, J., Wu, D., Bolla, J. R., Schmid, V., Heinze, R., Wild, K., Flemming, D., Pfeffer, S., Schwappach, B., Robinson, C. v., Sinning, I. (2020). Structural Basis of Tail-Anchored Membrane Protein Biogenesis by the GET Insertase Complex. *Molecular Cell*, *80*(1), 72–86. https://doi.org/10.1016/j.molcel.2020.08.012

McGilvray, P. T., Anghel, S. A., Sundaram, A., Zhong, F., Trnka, M. J., Fuller, J. R., Hu, H., Burlingame, A. L., Keenan, R. J. (2020). An ER translocon for multi-pass membrane protein biogenesis. *ELife*, *9*, 1–43. https://doi.org/10.7554/ELIFE.56889

McInally, S. G., Hagen, K. D., Nosala, C., Williams, J., Nguyen, K., Booker, J., Jones, K., Dawson, S. C. (2019). Robust and stable transcriptional repression in *Giardia* using CRISPRi. *Molecular Biology of the Cell*, *30*(1), 119–130. https://doi.org/10.1091/MBC.E18-09-0605

Milan, D. J., Kim, A. M., Winterfield, J. R., Jones, I. L., Pfeufer, A., Sanna, S., Arking, D. E., Amsterdam, A. H., Sabeh, K. M., Mably, J. D., Rosenbaum, D. S., Peterson, R. T., Chakravarti, A., Kääb, S., Roden, D. M., MacRae, C. A. (2009). Drug-sensitized zebrafish screen identifies multiple genes, including GINS3, as regulators of myocardial repolarization. *Circulation*, *120*(7), 553–559. https://doi.org/10.1161/CIRCULATIONAHA.108.821082

Miller, M. T., Higgin, J. J., Tanaka Hall, T. M. (2008). Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nature Structural Molecular Biology*, *15*(4), 397–402. https://doi.org/10.1038/NSMB.1390

Mock, J. Y., Chartron, J. W., Zaslaver, M., Xu, Y., Ye, Y., Clemons, W. M. (2015). Bag6 complex contains a minimal tail-anchor-targeting module and a mock BAG domain. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(1), 106–111. https://doi.org/10.1073/pnas.1402745112

Mock, J. Y., Xu, Y., Ye, Y., Clemons, W. M. (2017). Structural basis for regulation of the nucleo-cytoplasmic distribution of Bag6 by TRC35. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(44), 11679–11684. https://doi.org/10.1073/PNAS.1702940114

Monis, P. T., Caccio, S. M., Thompson, R. C. A. (2009). Variation in *Giardia*: towards a taxonomic revision of the genus. *Trends in Parasitology*, *25*(2), 93–100. https://doi.org/10.1016/J.PT.2008.11.006

Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G., Sogin, M. L. (2007). Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science*, *317*(5846), 1921–1926. https://doi.org/10.1126/SCIENCE.1143837

Murata, K., Degmetich, S., Kinoshita, M., Shimada, E. (2009). Expression of the congenital heart disease 5/tryptophan rich basic protein homologue gene during heart

development in Medaka fish, *Oryzias latipes*. *Development Growth and Differentiation*, *51*(2), 95–107. https://doi.org/10.1111/j.1440-169X.2008.01084.x

Najdrová, V., Dohnálek, V., Voleman, L., Doležal, P. (n.d.). Highly diverged pre-targeting complex in *Giardia intestinalis* reveals the evolution of GET pathway in eukaryotes. *Unpublished*.

Najdrová, V., Stairs, C. W., Vinopalová, M., Voleman, L., Doležal, P. (2020). The evolution of the Puf superfamily of proteins across the tree of eukaryotes. *BMC Biology*, *18*(1), 77. https://doi.org/10.1186/S12915-020-00814-3

Nash, T. E., Banks, S. M., Alling, D. W., Merritt, J. W., Conrad, J. T. (1990). Frequency of variant antigens in *Giardia lamblia*. *Experimental Parasitology*, *71*(4), 415–421. https://doi.org/10.1016/0014-4894(90)90067-M

Nash, T. E., McCutchan, T., Keister, D., Dame, J. B., Conrad, J. D., Gillin, F. D. (1985). Restriction-endonuclease analysis of DNA from 15 *Giardia* isolates obtained from humans and animals. *The Journal of Infectious Diseases*, *152*(1), 64–73. https://doi.org/10.1093/INFDIS/152.1.64

Nguyen, P., Bar-Sela, G., Sun, L., Bisht, K. S., Cui, H., Kohn, E., Feinberg, A. P., Gius, D. (2008). BAT3 and SET1A Form a Complex with CTCFL/BORIS To Modulate H3K4 Histone Dimethylation and Gene Expression. *Molecular and Cellular Biology*, *28*(21), 6720–6729. https://doi.org/10.1128/mcb.00568-08

Nyindodo-Ogari, L., Schwartzbach, S., Estrano, C. (2014). *Giardia* mitosomal protein import machinery differentially recognizes mitochondrial targeting signals. *Infectious Disorders Drug Targets*, *14*(1), 23–29. https://doi.org/10.2174/1871526514666140827101555

Petrů, M., Dohnálek, V., Füssy, Z., Pavel, D. (2021). Fates of Sec, Tat, and YidC Translocases in Mitochondria and Other Eukaryotic Compartments. *Molecular Biology and Evolution*, *38*(12), 5241–5254. https://doi.org/10.1093/MOLBEV/MSAB253

Pogge von Strandmann, E., Simhadri, V. R., von Tresckow, B., Sasse, S., Reiners, K. S. S., Hansen, H. P., Rothe, A., Böll, B., Simhadri, V. L., Borchmann, P., McKinnon, P. J., Hallek, M., Engert, A. (2007). Human leukocyte antigen-B-associated transcript 3 is released

from tumor cells and engages the NKp30 receptor on natural killer cells. *Immunity*, *27*(6), 965–974. https://doi.org/10.1016/J.IMMUNI.2007.10.010

Powis, K., Schrul, B., Tienson, H., Gostimskaya, I., Breker, M., High, S., Schuldiner, M., Jakob, U., Schwappach, B. (2013). Get3 is a holdase chaperone and moves to deposition sites for aggregated proteins when membrane targeting is blocked. *Journal of Cell Science*, *126*(2), 473–483. https://doi.org/10.1242/jcs.112151

Poxleitner, M. K., Carpenter, M. L., Mancuso, J. J., Wang, C. J. R., Dawson, S. C., Cande, W. Z. (2008). Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite *Giardia intestinalis*. *Science*, *319*(5869), 1530–1533. https://doi.org/10.1126/SCIENCE.1153752

Prucca, C. G., Slavin, I., Quiroga, R., Elías, E. v., Rivero, F. D., Saura, A., Carranza, P. G., Luján, H. D. (2008). Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature*, *456*(7223), 750–754. https://doi.org/10.1038/NATURE07585

Qiu, C., McCann, K. L., Wine, R. N., Baserga, S. J., Hall, T. M. T. (2014). A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(52), 18554–18559. https://doi.org/10.1073/PNAS.1407634112

Ramesh, M. A., Malik, S. B., Logsdon, J. M. (2005). A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Current Biology, 15*(2), 185–191. https://doi.org/10.1016/J.CUB.2005.01.003

Rangachari, M., Zhu, C., Sakuishi, K., Xiao, S., Karman, J., Chen, A., Angin, M., Wakeham, A., Greenfield, E. A., Sobel, R. A., Okada, H., McKinnon, P. J., Mak, T. W., Addo, M. M., Anderson, A. C., Kuchroo, V. K. (2012). Bat3 promotes T cell responses and autoimmunity by repressing Tim-3–mediated cell death and exhaustion. *Nature Medicine*, *18*(9), 1394–1400. https://doi.org/10.1038/NM.2871

Rao, M., Okreglak, V., Chio, U. S., Cho, H., Walter, P., Shan, S. O. (2016). Multiple selection filters ensure accurate tail-anchored membrane protein targeting. *ELife*, *5*, e21301. https://doi.org/10.7554/ELIFE.21301

Riches, A., Hart, C. J. S., Trenholme, K. R., Skinner-Adams, T. S. (2020). Anti- *Giardia* Drug Discovery: Current Status and Gut Feelings. *Journal of Medicinal Chemistry*, *63*(22), 13330–13354. https://doi.org/10.1021/ACS.JMEDCHEM.0C00910

Roberts, J. D., Thapaliya, A., Martínez-Lumbreras, S., Krysztofinska, E. M., Isaacson, R. L. (2015). Structural and functional insights into small, glutamine-rich, tetratricopeptide repeat protein alpha. *Frontiers in Molecular Biosciences*, *2*, 71. https://doi.org/10.3389/fmolb.2015.00071

Rome, M. E., Chio, U. S., Rao, M., Gristick, H., Shan, S. O. (2014). Differential gradients of interaction affinities drive efficient targeting and recycling in the GET pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), E4929–E4935. https://doi.org/10.1073/pnas.1411284111

Rome, M. E., Rao, M., Clemons, W. M., Shan, S. O. (2013). Precise timing of ATPase activation drives targeting of tail-anchored proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7666–7671. https://doi.org/10.1073/pnas.1222054110

Saint-Georges, Y., Garcia, M., Delaveau, T., Jourdren, L., le Crom, S., Lemoine, S., Tanty, V., Devaux, F., Jacq, C. (2008). Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. *PloS One*, *3*(6), e2293. https://doi.org/10.1371/journal.pone.0002293

Sasaki, T., Marcon, E., McQuire, T., Arai, Y., Moens, P. B., Okada, H. (2008). Bat3 deficiency accelerates the degradation of Hsp70-2/HspA2 during spermatogenesis. *The Journal of Cell Biology*, *182*(3), 449–458. https://doi.org/10.1083/JCB.200802113

Schmid, A. B., Lagleder, S., Gräwert, M. A., Röhl, A., Hagn, F., Wandinger, S. K., Cox, M. B., Demmer, O., Richter, K., Groll, M., Kessler, H., Buchner, J. (2012). The architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. *The EMBO Journal*, *31*(6), 1506–1517. https://doi.org/10.1038/EMBOJ.2011.472

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., Schwappach, B., Weissman, J. S. (2008). The GET Complex Mediates Insertion of Tail-Anchored

Proteins into the ER Membrane. *Cell*, *134*(4), 634–645. https://doi.org/10.1016/j.cell.2008.06.025

Sherrill, J., Mariappan, M., Dominik, P., Hegde, R. S., Keenan, R. J. (2011). A conserved archaeal pathway for tail-anchored membrane protein insertion. *Traffic*, *12*(9), 1119–1123. https://doi.org/10.1111/j.1600-0854.2011.01229.x

Shih, H.-W., Alas, G. C., Rydell, D. S., Zhang, B., Hamilton, G. A., Paredez, A. R. (2021, preprint). An early signaling transcription factor regulates differentiation in *Giardia*. *BioRxiv*, 2021.05.27.446072. https://doi.org/10.1101/2021.05.27.446072

Shing, J. C., Lindquist, L. D., Borgese, N., Bram, R. J. (2017). CAML mediates survival of myc-induced lymphoma cells independent of tail-anchored protein insertion. *Cell Death Discovery*, *3*(1), 16098. https://doi.org/10.1038/cddiscovery.2016.98

Shurtleff, M. J., Itzhak, D. N., Hussmann, J. A., Schirle Oakdale, N. T., Costa, E. A., Jonikas, M., Weibezahn, J., Popova, K. D., Jan, C. H., Sinitcyn, P., Vembar, S. S., Hernandez, H., Cox, J., Burlingame, A. L., Brodsky, J. L., Frost, A., Borner, G. H. H., Weissman, J. S. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *ELife*, *7*, e37018. https://doi.org/10.7554/eLife.37018

Simpson, P. J., Schwappach, B., Dohlman, H. G., Isaacson, R. L. (2010). Structures of Get3, Get4, and Get5 provide new models for TA membrane protein targeting. *Structure*, *18*(8), 897–902. https://doi.org/10.1016/J.STR.2010.07.003

Singer, S. M., Yee, J., Nash, T. E. (1998). Episomal and integrated maintenance of foreign DNA in *Giardia lamblia*. *Molecular and Biochemical Parasitology*, *92*(1), 59–69. https://doi.org/10.1016/S0166-6851(97)00225-9

Smith, P. D., Gillin, F. D., Kaushal, N. A., Nash, T. E. (1982). Antigenic analysis of *Giardia lamblia* from Afghanistan, Puerto Rico, Ecuador, and Oregon. *Infection and Immunity*, *36*(2), 714–719. https://doi.org/10.1128/IAI.36.2.714-719.1982

Sojka, S., Amin, N. M., Gibbs, D., Christine, K. S., Charpentier, M. S., Conlon, F. L. (2014). Congenital heart disease protein 5 associates with CASZ1 to maintain myocardial

tissue integrity. *Development*, *141*(15), 3040–3049. https://doi.org/10.1242/dev.106518

Soto, C., Pritzkow, S. (2018). Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature Neuroscience*, *21*(10), 1332–1340. https://doi.org/10.1038/s41593-018-0235-9

Spies, T., Bresnahan, M., Strominger, J. L. (1989). Human major histocompatibility complex contains a minimum of 19 genes between the complement cluster and HLA-B. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(22), 8955–8958. https://doi.org/10.1073/PNAS.86.22.8955

Srivastava, R., Zalisko, B. E., Keenan, R. J., Howell, S. H. (2017). The GET system inserts the tail-anchored protein, SYP72, into Endoplasmic Reticulum Membranes. *Plant Physiology*, *173*(2), 1137–1145. https://doi.org/10.1104/pp.16.00928

Štefanić, S., Morf, L., Kulangara, C., Regös, A., Sonda, S., Schraner, E., Spycher, C., Wild, P., Hehl, A. B. (2009). Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *Journal of Cell Science*, *122*(Pt 16), 2846–2856. https://doi.org/10.1242/JCS.049411

Stefanovic, S., Hegde, R. S. (2007). Identification of a Targeting Factor for Posttranslational Membrane Protein Insertion into the ER. *Cell*, *128*(6), 1147–1159. https://doi.org/10.1016/j.cell.2007.01.036

Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y. Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F., Bernhard, F., Denic, V., Dötsch, V., Sinning, I. (2011). Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science*, *333*(6043), 758–762. https://doi.org/10.1126/science.1207125

Suloway, C. J. M., Chartron, J. W., Zaslaver, M., Clemons, W. M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(35), 14849–14854. https://doi.org/10.1073/PNAS.0907522106

Suloway, C. J. M., Rome, M. E., Clemons, W. M. (2012). Tail-anchor targeting by a Get3 tetramer: The structure of an archaeal homologue. *EMBO Journal*, *31*(3), 707–719. https://doi.org/10.1038/emboj.2011.433

Sun, C. H., Chou, C. F., Tai, J. H. (1998). Stable DNA transfection of the primitive protozoan pathogen *Giardia lamblia*. *Molecular and Biochemical Parasitology*, *92*(1), 123–132. https://doi.org/10.1016/S0166-6851(97)00239-9

Svärd, S. G., Meng, T. C., Hetsko, M. L., McCaffery, J. M., Gillin, F. D. (1998). Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia*. *Molecular Microbiology*, *30*(5), 979–989. https://doi.org/10.1046/J.1365-2958.1998.01125.X

Svärd, S. G., Rafferty, C., McCaffery, J. M., Smith, M. W., Reiner, D. S., Gillin, F. D. (1999). A signal recognition particle receptor gene from the early-diverging eukaryote, *Giardia lamblia*. *Molecular and Biochemical Parasitology*, *98*(2), 253–264. https://doi.org/10.1016/S0166-6851(98)00174-1

Sweet, D. J., Pelham, H. R. B. (1992). The *Saccharomyces cerevisiae* SEC20 gene encodes a membrane glycoprotein which is sorted by the HDEL retrieval system. *The EMBO Journal*, *11*(2), 423–432. https://doi.org/10.1002/J.1460-2075.1992.TB05071.X

Thomson, E., Rappsilber, J., Tollervey, D. (2007). Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. *RNA*, *13*(12), 2165–2174. https://doi.org/10.1261/RNA.747607

Thress, K., Henzel, W., Shillinglaw, W., Kornbluth, S. (1998). Scythe: a novel reaper-binding apoptotic regulator. *The EMBO Journal*, *17*(21), 6135. https://doi.org/10.1093/EMBOJ/17.21.6135

Torgerson, P. R., Devleesschauwer, B., Praet, N., Speybroeck, N., Willingham, A. L., Kasuga, F., Rokni, M. B., Zhou, X. N., Fèvre, E. M., Sripa, B., Gargouri, N., Fürst, T., Budke, C. M., Carabin, H., Kirk, M. D., Angulo, F. J., Havelaar, A., de Silva, N. (2015). World Health Organization Estimates of the Global and Regional Disease Burden of 11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. *PLoS Medicine*, *12*(12), e1001920. https://doi.org/10.1371/JOURNAL.PMED.1001920

Tovar, J., León-Avila, G., Sánchez, L. B., Sutak, R., Tachezy, J., van der Giezen, M., Hernández, M., Müller, M., Lucocq, J. M. (2003). Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature*, *426*(6963), 172–176. https://doi.org/10.1038/NATURE01945

Tůmová, P., Hofštetrová, K., Nohýnková, E., Hovorka, O., Král, J. (2007). Cytogenetic evidence for diversity of two nuclei within a single diplomonad cell of *Giardia*. *Chromosoma*, *116*(1), 65–78. https://doi.org/10.1007/S00412-006-0082-4

Tůmová, P., Uzlíková, M., Jurczyk, T., Nohýnková, E. (2016). Constitutive aneuploidy and genomic instability in the single-celled eukaryote *Giardia intestinalis*. *MicrobiologyOpen*, *5*(4), 560–574. https://doi.org/10.1002/MBO3.351

Vilardi, F., Lorenz, H., Dobberstein, B. (2011). WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *Journal of Cell Science*, *124*(8), 1301–1307. https://doi.org/10.1242/jcs.084277

Vilardi, F., Stephan, M., Clancy, A., Janshoff, A., Schwappach, B. (2014). WRB and CAML are necessary and sufficient to mediate tail-anchored protein targeting to the ER membrane. *PLoS ONE*, *9*(1), e85033. https://doi.org/10.1371/journal.pone.0085033

Voleman, L. (2011). The role of a SNARE proteins in the biogenesis of *Giardia intestinalis* mitosomes [Master´s thesis; supervisor: Mgr. Pavel Doležal, Ph.D.]. Charles University; http://hdl.handle.net/20.500.11956/48175.

Voth, W., Schick, M., Gates, S., Li, S., Vilardi, F., Gostimskaya, I., Southworth, D. R., Schwappach, B., Jakob, U. (2014). The protein targeting factor Get3 functions as ATP-Independent chaperone under oxidative stress conditions. *Molecular Cell*, *56*(1), 116–127. https://doi.org/10.1016/j.molcel.2014.08.017

Wanderoy, S., Hees, J. T., Klesse, R., Edlich, F., Harbauer, A. B. (2020). Kill one or kill the many: interplay between mitophagy and apoptosis. *Biological Chemistry*, *402*(1), 73–88. https://doi.org/10.1515/hsz-2020-0231

Wang, F., Brown, E. C., Mak, G., Zhuang, J., Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Molecular Cell*, *40*(1), 159–171. https://doi.org/10.1016/j.molcel.2010.08.038

Wang, F., Chan, C., Weir, N. R., Denic, V. (2014). The Get1/2 transmembrane complex is an endoplasmic-reticulum membrane protein insertase. *Nature*, *512*(7515), 441–444. https://doi.org/10.1038/nature13471

Wang, F., Whynot, A., Tung, M., Denic, V. (2011). The Mechanism of Tail-Anchored Protein Insertion into the ER Membrane. *Molecular Cell*, *43*(5), 738–750. https://doi.org/10.1016/j.molcel.2011.07.020

Wang, M., Ogé, L., Perez-Garcia, M.-D., Hamama, L., Sakr, S. (2018). The PUF Protein Family: Overview on PUF RNA Targets, Biological Functions, and Post Transcriptional Regulation. *International Journal of Molecular Sciences*, *19*(2), 410. https://doi.org/10.3390/ijms19020410

Wang, Q. C., Zheng, Q., Tan, H., Zhang, B., Li, X., Yang, Y., Yu, J., Liu, Y., Chai, H., Wang, X., Sun, Z., Wang, J. Q., Zhu, S., Wang, F., Yang, M., Guo, C., Wang, H., Zheng, Q., Li, Y., Chen, Q., Zhou, A., Tang, T. S. (2016). TMCO1 Is an ER Ca (2+) Load-Activated Ca (2+) Channel. *Cell*, *165*(6), 1454–1466. https://doi.org/10.1016/J.CELL.2016.04.051

Wang, Q., Crnković, V., Preisinger, C., Stegmüller, J. (2021). The parkinsonism-associated protein FBXO7 cooperates with the BAG6 complex in proteasome function and controls the subcellular localization of the complex. *Biochemical Journal*, *478*(12), 2179–2199. https://doi.org/10.1042/bcj20201000

Wang, Q., Liu, Y., Soetandyo, N., Baek, K., Hegde, R., Ye, Y. (2011). A Ubiquitin Ligase-Associated Chaperone Holdase Maintains Polypeptides in Soluble States for Proteasome Degradation. *Molecular Cell*, *42*(6), 758–770. https://doi.org/10.1016/j.molcel.2011.05.010

Wang, X., McLachlan, J., Zamore, P. D., Hall, T. M. T. (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell*, *110*(4), 501–512. https://doi.org/10.1016/S0092-8674(02)00873-5

Wang, Y., Opperman, L., Wickens, M., Tanaka Hall, T. M. (2009). Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(48), 20186–20191. https://doi.org/10.1073/PNAS.0812076106

Wear, M. P., Kryndushkin, D., O'Meally, R., Sonnenberg, J. L., Cole, R. N., Shewmaker, F. P., Nagai, Y. (2015). Proteins with intrinsically disordered domains are preferentially recruited to polyglutamine aggregates. *PLoS ONE*, *10*(8), e136362. https://doi.org/10.1371/journal.pone.0136362

White, S. H., von Heijne, G. (2005). Transmembrane helices before, during, and after insertion. *Current Opinion in Structural Biology*, *15*(4), 378–386. https://doi.org/10.1016/J.SBI.2005.07.004

Wideman, J. G. (2015). The ubiquitous and ancient ER membrane protein complex (EMC): Tether or not? *F1000Research*, *4*, 624. https://doi.org/10.12688/F1000RESEARCH.6944.2

Wiedemann, N., Pfanner, N. (2017). Mitochondrial Machineries for Protein Import and Assembly. *Annual Review of Biochemistry*, *86*, 685–714. https://doi.org/10.1146/ANNUREV-BIOCHEM-060815-014352

Winnefeld, M., Grewenig, A., Schnölzer, M., Spring, H., Knoch, T. A., Gan, E. C., Rommelaere, J., Cziepluch, C. (2006). Human SGT interacts with Bag-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Experimental Cell Research*, *312*(13), 2500–2514. https://doi.org/10.1016/J.YEXCR.2006.04.020

Wunderley, L., Leznicki, P., Payapilly, A., High, S. (2014). SGTA regulates the cytosolic quality control of hydrophobic substrates. *Journal of Cell Science*, *127*(21), 4728–4739. https://doi.org/10.1242/JCS.155648/-/DC1

Xie, K., Dalbey, R. E. (2008). Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases. *Nature Reviews. Microbiology*, *16*(2), 120. https://doi.org/10.1038/NRMICRO.2017.160

Xin, B., Puffenberger, E. G., Turben, S., Tan, H., Zhou, A., Wang, H. (2010). Homozygous frameshift mutation in TMCO1 causes a syndrome with craniofacial dysmorphism, skeletal anomalies, and mental retardation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(1), 258–263. https://doi.org/10.1073/pnas.0908457107

Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P., Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y. D., Grossmann, G., Grefen, C. (2017). Loss of GET pathway orthologs in *Arabidopsis thaliana* causes root hair growth defects and affects SNARE abundance. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(8), E1544–E1553. https://doi.org/10.1073/pnas.1619525114

Xu, F., Jex, A., Svärd, S. G. (2020). A chromosome-scale reference genome for *Giardia intestinalis* WB. *Scientific Data*, *7*(1), 38. https://doi.org/10.1038/S41597-020-0377-Y

Xu, X., Ouyang, M., Lu, D., Zheng, C., Zhang, L. (2021). Protein Sorting within Chloroplasts. *Trends in Cell Biology*, *31*(1), 9–16. https://doi.org/10.1016/J.TCB.2020.09.011

Xu, Y., Cai, M., Yang, Y., Huang, L., Ye, Y. (2012). SGTA Recognizes a Noncanonical Ubiquitin-like Domain in the Bag6-Ubl4A-Trc35 Complex to Promote Endoplasmic Reticulum-Associated Degradation. *Cell Reports*, *2*(6), 1633–1644. https://doi.org/10.1016/j.celrep.2012.11.010

Yamagata, A., Mimura, H., Sato, Y., Yamashita, M., Yoshikawa, A., Fukai, S. (2010). Structural insight into the membrane insertion of tail-anchored proteins by Get3. *Genes to Cells: Devoted to Molecular Cellular Mechanisms*, *15*(1), 29–41. https://doi.org/10.1111/J.1365-2443.2009.01362.X

Yamamoto, Y., Sakisaka, T. (2012). Molecular Machinery for Insertion of Tail-Anchored Membrane Proteins into the Endoplasmic Reticulum Membrane in Mammalian Cells. *Molecular Cell*, *48*(3), 387–397. https://doi.org/10.1016/j.molcel.2012.08.028

Yee, J., Nash, T. E. (1995). Transient transfection and expression of firefly luciferase in *Giardia lamblia*. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(12), 5615–5619. https://doi.org/10.1073/PNAS.92.12.5615

Zalisko, B. E., Chan, C., Denic, V., Rock, R. S., Keenan, R. J. (2017). Tail-Anchored Protein Insertion by a Single Get1/2 Heterodimer. *Cell Reports*, *20*(10), 2287–2293. https://doi.org/10.1016/j.celrep.2017.08.035

Zamore, P. D., Williamson, J. R., Lehmann, R. (1997). The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA*, *3*(12), 1421–1433.

Zamponi, N., Zamponi, E., Mayol, G. F., Lanfredi-Rangel, A., Svärd, S. G., Touz, M. C. (2017). Endoplasmic reticulum is the sorting core facility in the Golgi-lacking protozoan *Giardia lamblia*. *Traffic, 18*(9), 604–621. https://doi.org/10.1111/TRA.12501

Zhang, C., Muench, D. G. (2015). A Nucleolar PUF RNA-binding Protein with Specificity for a Unique RNA Sequence. *The Journal of Biological Chemistry*, *290*(50), 30108–30118. https://doi.org/10.1074/JBC.M115.691675

Zhang, J., McCann, K. L., Qiu, C., Gonzalez, L. E., Baserga, S. J., Hall, T. M. T. (2016). Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nature Communications*, *7*, 13085. https://doi.org/10.1038/NCOMMS13085

Zhang, Y., Hughson, F. M. (2021). Chaperoning SNARE Folding and Assembly. *Annual Review of Biochemistry*, *90*, 581. https://doi.org/10.1146/ANNUREV-BIOCHEM-081820-103615

Zhou, T., Radaev, S., Rosen, B. P., Gatti, D. L. (2000). Structure of the ArsA ATPase: The catalytic subunit of a heavy metal resistance pump. *EMBO Journal*, *19*(17), 4838–4845. https://doi.org/10.1093/emboj/19.17.4838

Zhu, X., Qi, X., Yang, Y., Tian, W., Liu, W., Jiang, Z., Li, S., Zhu, X. (2020). Loss of the ER membrane protein complex subunit Emc3 leads to retinal bipolar cell degeneration in aged mice. *PLoS ONE*, *15*(9), e0238435. https://doi.org/10.1371/journal.pone.0238435

nature structural & molecular biology

ARTICLES

https://doi.org/10.1038/s41594-022-00798-4

Check for updates

Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3

Michelle Y. Fry[1], Vladimíra Najdrová[2], Ailiena O. Maggiolo[1], Shyam M. Saladi[1], Pavel Doležal[2] and William M. Clemons Jr[1]✉

Tail-anchored (TA) membrane proteins, accounting for roughly 2% of proteomes, are primarily targeted posttranslationally to the endoplasmic reticulum membrane by the guided entry of TA proteins (GET) pathway. For this complicated process, it remains unknown how the central targeting factor Get3 uses nucleotide to facilitate large conformational changes to recognize then bind clients while also preventing exposure of hydrophobic surfaces. Here, we identify the GET pathway in *Giardia intestinalis* and present the structure of the Get3–client complex in the critical postnucleotide-hydrolysis state, demonstrating that Get3 reorganizes the client-binding domain (CBD) to accommodate and shield the client transmembrane helix. Four additional structures of *Gi*Get3, spanning the nucleotide-free (apo) open to closed transition and the ATP-bound state, reveal the details of nucleotide stabilization and occluded CBD. This work resolves key conundrums and allows for a complete model of the dramatic conformational landscape of Get3.

The delivery and insertion of integral membrane proteins (IMPs) into designated membranes is a complex process. Most IMPs are targeted cotranslationally to the endoplasmic reticulum (ER) membrane via the signal recognition particle pathway[1]. One class of IMPs, tail-anchored (TA) proteins, contain a single transmembrane domain (TMD) near their C terminus and must be targeted to cellular membranes posttranslationally, as their signal cannot access the signal recognition particle pathway[2–5]. TA proteins account for roughly 2% of all proteomes and targeting of these proteins has been studied primarily in the opishtokont clade that includes both fungi and metazoans[1,6–8].

The best-studied pathway for targeting TA proteins to the ER membrane is the conserved guided entry of TA protein (GET) pathway. The central targeting component is the homodimeric ATPase Get3, which uses multiple ATP-dependent large conformational changes to drive targeting of TA proteins[5,9–11]. Upstream from Get3 is a cochaperone, SGTA in mammals and Sgt2 in fungi, that binds specifically to TA proteins being delivered to the ER[6,12]. On binding to ATP, Get3 transitions from a 'open' to 'closed' conformation that is recognized by a loading complex, the mammalian Bag6 complex or fungal Get4–Get5, which facilitates the transfer of the TA protein (client) from the cochaperone to Get3 (refs. [13–15]). Once a client binds, ATP hydrolysis occurs driving Get3 into an 'intermediate' state releasing Get3 from the loading complex[13]. A membrane-bound receptor complex, Get1–CAML (mammals) or Get1–Get2 (fungi), recruits the client-bound Get3 to the ER, drives the release of the client from Get3 by favoring Get3 in an 'open' state and then facilitates insertion of the client into the ER membrane[16,17].

Various structures from different homologs of opisthokont Get3 provide a working model, albeit incomplete, of the TA protein targeting cycle[16–23]. In all structures, Get3 is a homodimer stabilized through a coordinated Zn^{2+} ion liganded by four conserved cysteines. Two distinct structural regions are observed, a well-ordered nucleotide binding domain (NBD) and a flexible α-helical region that forms the client-binding domain (CBD). Parts that are characteristic of G-type hydrolases are present in the NBD[10], a P-loop (formed by a deviant Walker A motif[19,24]), A-loop and Switches I and II (Fig. 1a). In the current targeting model, binding of ATP drives the transition from the 'open' to the 'closed' state causing the α-helices in the CBD to rearrange to form a hydrophobic groove that is the presumed site of TA protein binding[18].

Despite the considerable effort to characterize the dynamic conformational landscape of Get3, several key mechanistic questions remain. Various 'open' states are seen for the nucleotide-free Get3, yet a clear picture for how the dimer can transition to a 'closed' state while continuing to bury the CBD hydrophobic surfaces is missing. These regions are observed to be exposed and disordered in current structures[18,19,25]. In the closed transition-state structure[18], which is argued to mimic the ATP-bound state that is prepared for client capture, the formed hydrophobic groove remains exposed. A previous structure of a Get3–TA complex stabilized by antibodies[23] revealed that the transmembrane helix binds to the groove in the CBD but saw no conformational changes relative to either the transition-state or Get4 bound structures[18,22]. This is incompatible with both a requirement for release from the loading complex for delivery to the ER and a need for Get3 to recognize and protect the client. Critically, the structure of a posthydrolysis Get3–client complex is required. This state requires a substantial conformational change, as evidenced by single-molecule Forster resonance energy transfer (smFRET) studies that show an opening of the CBD on client binding and nucleotide hydrolysis[13]. In summary, the mechanistic picture of the nucleotide cycle is incomplete and dependent on crystal structures that likely do not reflect the states of Get3 in vivo.

Outside the opisthokont supergroup, there has been little characterization of the GET pathway. Some GET components have been demonstrably identified in Archaeplastida, *Arabidopsis thaliana*[26]

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Department of Parasitology, Faculty of Science, BIOCEV, Charles University, Prague, Czech Republic. ✉e-mail: clemons@caltech.edu

and most recently in a SAR supergroup (a clade including stramenopiles, alveolates and Rhizaria) member, the apicomplexan *Plasmodium falciparum*[27]. Plants evolved four compartment-specific paralogs of Get3. While Get3a exhibits the conserved cytosolic function[26], other proteins appear to function at the membranes of mitochondria and chloroplasts[28,29]. A deletion of the *At*Get3a gene resulted in a stunted root growth phenotype. A *S. cerevisiae* Get3 deletion phenotype was rescued by expression of *P. falciparum* Get3 (ref. [27]). These works begin to demonstrate the conservation of the GET pathway throughout eukaryotes, but conservation of mechanistic details has yet to be demonstrated.

Here, we identify key GET pathway components in *G. intestinalis* of the Excavata supergroup: homologs of Get3, Get4, Get2 and Sgt2. The protist *G. intestinalis* causes giardiasis, a neglected disease that affects up to 30% of the population in developing countries[30]. We present a series of five structures of *Gi*Get3, outside opisthokonts, that complete the mechanistic picture. This includes a cryogenic-electron microscopy (cryo-EM) structure of client-bound Get3 in the missing posthydrolysis state, which reveals the conformational changes that lead to disassociation from the loading complex and a reordering of the CBD that results in a more open groove with a shielded TMD of the bound client. Three structures of *Gi*Get3 in the apo (nucleotide-free) state illustrate the range of Get3 conformations and resolve the occluded CBD. Finally, an ATP-bound (prehydrolysis) Get3 structure shows that nucleotide binding is not sufficient for reorganizing the CBD to expose the hydrophobic groove. Altogether, these structures represent three functional states of the Get3 ATPase cycle revealing dramatic transitions in the CBD that resolves how Get3 prevents exposure of hydrophobic surfaces in the absence of client and protects the client TMD during targeting. These findings provide a comprehensive picture of the Get3 catalytic cycle, notably from a single organism, explaining the conformational changes in Get3 necessary for driving the successful targeting of TA proteins.

Results

Identifying GET pathway components in *G. intestinalis*. Initially, we began with a search for Get3 homologs in *G. intestinalis* and identified a clear single hit to the gene *GL50803_007953* (*Gi*Get3) with a sequence identity of 42.14% and 44.67% to human and yeast homologs, respectively (Fig. 1a). This search was expanded and resulted in the identification of Get3 homologs in all eukaryotic supergroups (Fig. 1b). An early evolutionary branchpoint separated Get3 homologs into two distinct clades as demonstrated previously[26], of which Excavata and Opisthokonta are both in Clade II (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). An alignment of *Gi*Get3 to other homologs reveals all conserved components including residues important to binding other GET proteins are present (Fig. 1a and Extended Data Fig. 1)[22]. To identify cellular localization, cells expressing a labeled ER-marker (HA-tagged PDI2) were immunostained with antibodies against either *Gi*Get3 or the ER-marker and visualized by fluorescence microscopy (Fig. 1c,d). *Gi*Get3 was broadly expressed, localizing primarily to the cytoplasm as expected. Western blots of the soluble or membrane fractions showed most of the Get3 was in the soluble fraction, supporting the imaging results (Fig. 1e). Some Get3 is associated with the membrane, which agrees with Get3 function and previous mammalian studies[17,31]. Hydrolase activity of *Gi*Get3 was verified by an ATPase assay using purified protein (Extended Data Fig. 2A). A D53N point mutation corresponding to an ATPase inactivating mutation in yeast (D57N) also was inactive in *G. intestinalis*[18,19].

We next sought to identify other GET components. Crude lysate of *G. intestinalis* cells expressing both biotinylated BAP-tagged Get3 and biotin ligase (BirA) was passed over a streptavidin affinity column and the resulting eluate was analyzed by mass spectrometry to identify potential Get3 binding partners. Of the highest enriched proteins, two likely GET components were identified, homologs of Get4 (*GL50803_00112893*) and Get2 (*GL50803_0017617*) (Fig. 1f and Supplementary Table 2). The putative Get4 has high structural homology, as predicted by AlphaFold2 (ref. [32]) and a 33.3% sequence identity to yeast Get4 that includes conservation of residues critical for Get3 binding, suggesting a similar mode of interaction (Extended Data Fig. 3). Get2 homologs are less conserved[33,34]; the identified Get2 homolog has a 15.1% identity to yeast Get2 with a similar predicted architecture, three TMDs and an extended N-terminal tether that contains potential Get3 binding residues (Extended Data Fig. 4a)[16]. Immunostained *G. intestinalis* cells reveal that, as expected, *Gi*Get2 localizes to the ER (Extended Data Fig. 4b,c).

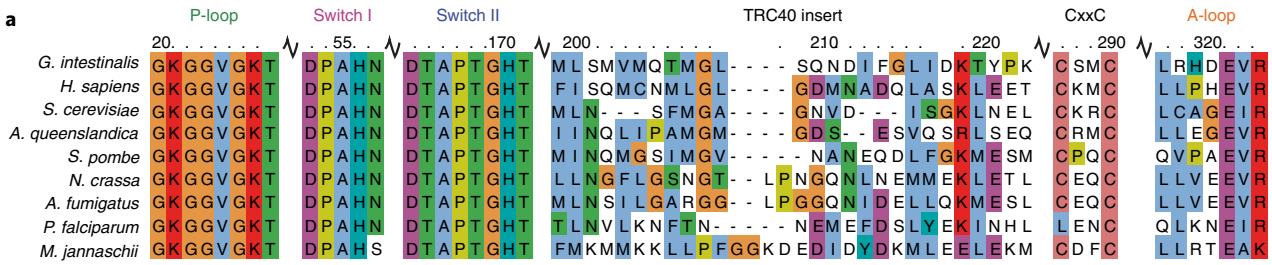
With clear evidence for a GET pathway in *G. intestinalis*, we searched more broadly for other pathway components. Using a structure-based sequence search of the *G. intestinalis* genome we identified an Sgt2 homolog (*GL50803_007287*) that has a 27.2% identity to yeast and contains the conserved domains including the N-terminal dimerization, the central cochaperone binding tetratricopeptide repeat and the predicted C-terminal 'helical-hand' for client binding (Extended Data Fig. 5)[35]. To functionally validate *Gi*Sgt2, an in vitro capture assay was used to demonstrate that *Gi*Sgt2 can capture a yeast TA protein from a yeast Hsp70 (Extended Data Fig. 2b,c)[35,36].

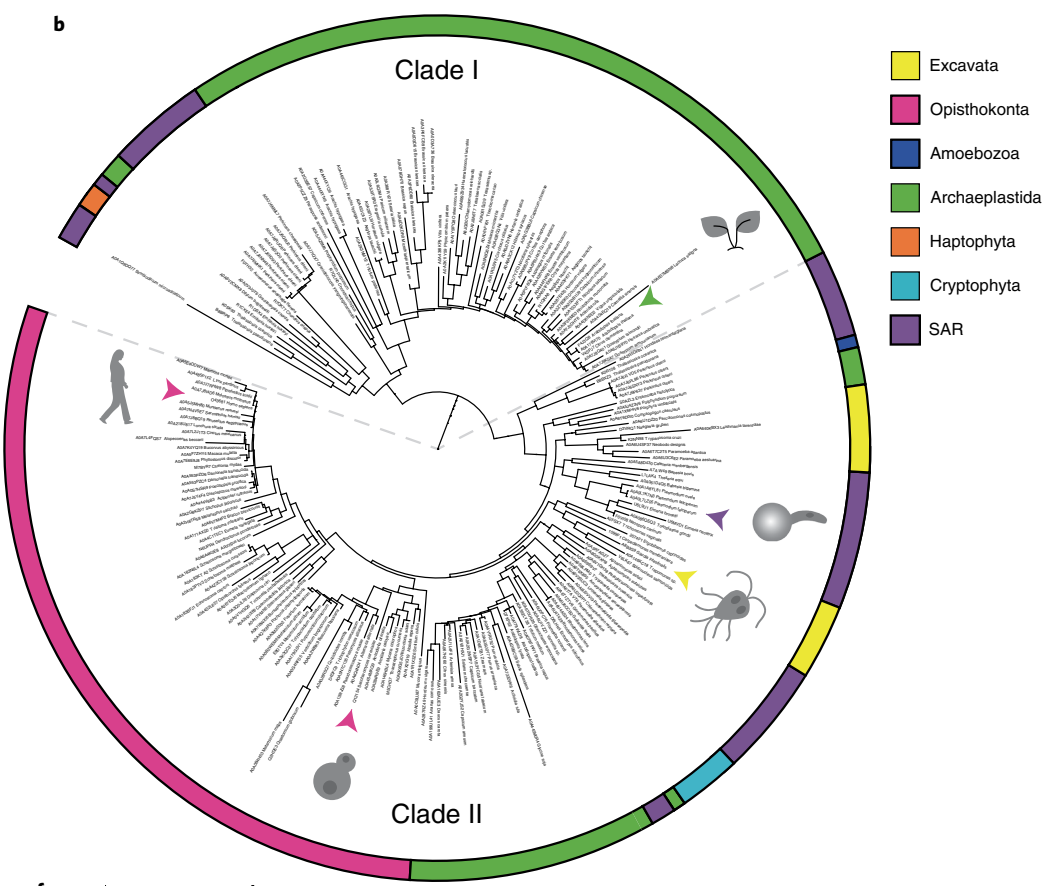
Last, we sought to identify potential clients for this newly identified GET pathway. A list of putative *Giardia* TA proteins was generated based on previously used criteria (Supplementary Table 3)[37]. Seven proteins were selected with hydrophobicities of the TMDs covering the full range of ER TA proteins (12.9 to 31.1)[1] and their TMDs were tagged with a small ubiquitin-like modifier- (SUMO-) and coexpressed with *Gi*Get3 in *Escherichia coli*. Potential complexes were purified using the affinity tag on the TA protein and Get3–TA complexes were detected by western blots using either anti-*Gi*Get3 or anti-SUMO antibodies (Extended Data Fig. 2d). Clear bands for both the TA protein and Get3 were present for *GL50803_009489* and *GL50803_0024512*, the latter annotated as a hypothetical protein in the GiardiaDB[38]. *GL50803_009489* is annotated as the v-SNARE Sec22, a known ER-targeted GET client in yeast[39]. Calculated hydrophobicities using the transmembrane tendency scale[40] of these TMDs are 25.76 and 33.05, respectively,

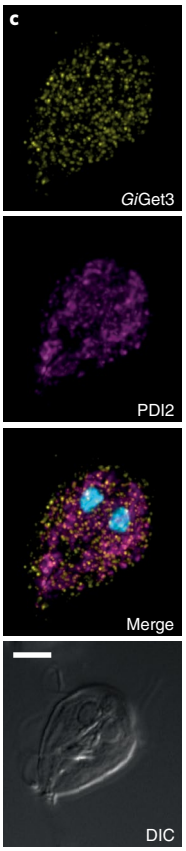
Fig. 1 | Identification of the GET pathway in *G. intestinalis*. **a**, Sequence alignments of the conserved Get3 features across several eukaryotes. Residue colors are based on ClustalX[46] and numbered by sequence position in *Giardia*. Colors for the conserved Get3 features are used throughout the text: P-loop (green), Switch I (magenta), Switch II (blue), TRC40 insert (black), CxxC (black) and A-loop (orange). **b**, A phylogenetic tree of eukaryotic Get3 homologs colored by supergroup and cartoons for organisms discussed with arrows pointing to the sequence. Supergroups are Excavata (*G. intestinalis*, yellow), Opisthokonta (*H. sapiens* and *S. cerevisiae*, magenta), Amoebozoa (blue), Archaeplastida (*A. thaliana*, green), *Hapt*ophyta (orange), Cryptophyta (cyan) and SAR (*P. falciparum*, purple). **c**, Detection of *Gi*Get3 localization through immunofluorescent projections of *z*-stacks of Giardia trophozites (brightfield image, bottom panel). Get3 (yellow) and the ER (magenta) were identified using antibodies against *Gi*Get3 and an HA-tagged PDI2. In the merged image (third panel) DAPI stains the nucleus (cyan). Scale bar, 5 µm. **d**, A 3D representation of the *z*-stacks in **c**. **e**, Western blots of *Gi*Get3 trophozite high-speed centrifugation fractions: lysate, soluble and pellet were probed with antibodies against *Gi*Get3, enolase (cytosol), PDI2 (ER)[47] and *GL50803_009296* (mitosomes)[48]. All were repeated at least five times. **f**, A volcano plot of the mass spectrometry analysis of the eluate from Get3 pull-downs (the enrichment or $\log_2$ of the fold-change of Get3 versus mock pull-down (*x* axis) and the significance or $-\log_{10}$ (*y* axis) determined using a two-sided *t*-test). Get2 and Get4 (orange) were identified in a list of 26 significant hits (right of dashed line, data available in Supplementary Table 2).

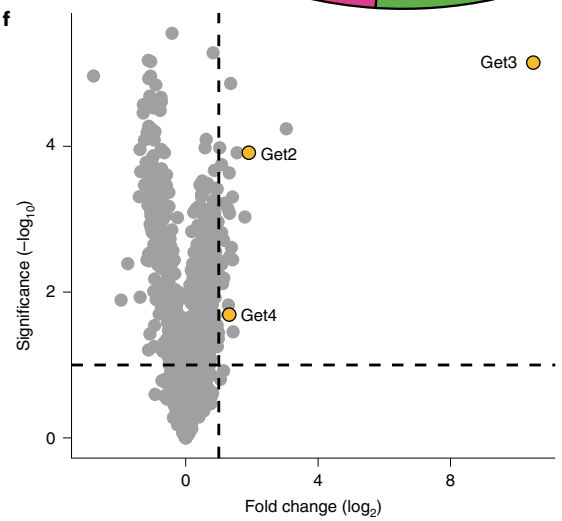
and correlate well with the expected hydrophobicity for both Get3 binding and ER targeting[1,37].

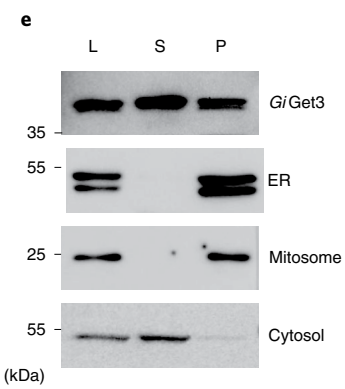
Structures of apo Get3 in a range of conformations. While structures reveal nucleotide-free Get3 adopts an open conformation, experimental evidence supports a range of conformations. To structurally characterize these states, *Gi*Get3 was cloned and expressed in *E. coli* followed by purification, where it eluted as a dimer (Extended Data Fig. 6a,b). Apo *Gi*Get3 crystallized in space group $P2_12_12$ and a data set was collected and a crystal structure was refined to 3.0 Å











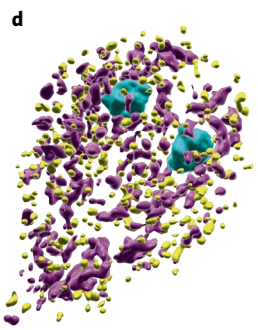


Table 1 | Data collection and refinement statistics

	Apo (PDB 7SPZ)[a]	ATP (PDB 7SPY)[a]
Data collection		
Space group	$P2_12_12$	$P3_221$
Cell dimensions		
a, *b*, *c* (Å)	54.1, 102.6, 138.8	81.0, 81.0, 130.1
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	50–3.0 (3.1–3.0)	50–2.23 (2.27–2.23)
Wavelength (Å)	0.97946	0.97946
R_{merge}	0.272 (3.09)	0.108 (1.132)
R_{pim}	0.086 (0.078)	0.035 (0.453)
I/σ	21.0 (3.8)	22.6 (1.75)
Completeness (%)	91.4 (58.8)	97.4 (79.6)
Redundancy (%)	11.9 (12.5)	8.9 (55)
Refinement		
Resolution (Å)	34.7–3.0	30.47–2.23
No. reflections	14,734	23,301
R_{work} / R_{free} (%)	29.5 / 34.8	17.9 / 20.9
No. atoms		
Protein	5,072	2,544
Ligand/ion	2	39
Solvent		178
B factors		
Protein	48.12	40.62
Ligand/ion	40.49	24.77
R.m.s. deviations		
Bond lengths (Å)	0.01	0.01
Bond angles (°)	1.62	1.74
Validation		
MolProbity score	1.20	1.13
Clashscore	1.59	1.57
Rotamer outliers (%)	0	0
$C\beta$ outliers (%)	0	0
Ramachandran outliers (%)	0	0
Ramachandran favored (%)	96	97
Ramachandran unfavored (%)	4	3

[a]Values in parentheses are for the highest-resolution shell.

resolution (full statistics provided in Table 1). The asymmetric unit contained two monomers that generated two distinct symmetry related dimer conformations, which we refer to as apo1 and apo2 (Fig. 2a, Extended Data Fig. 6c–g and Table 1). The two *Gi*Get3 monomers have an r.m.s.d. of 0.7 Å to each other and overall are structurally similar to opisthokont Get3s. Both form symmetric homodimers that contain the well-ordered NBD and α-helical CBD, consisting of helices 4–9 (H4–9). The two forms are consistent with an 'open' conformation, with apo1 most closely resembling the open conformation seen in fungal structures (Protein Data Bank (PDB) 3IBG, 2WOO and 3A36) (Fig. 2a and Supplementary Fig. 2a,c,d)[18,19,25]. The second conformation, apo2, has not yet been seen and is between the fungal 'open' and 'closed' conformations.

The range of possible conformations should include a 'closed' apo structure, which we captured using single-particle analysis cryo-EM (Extended Data Fig. 7a–f). Purified apo *Gi*Get3 was used to collect a large data set. Despite starting with >10 million particles, a small subset of 51,340 particles could be classified into a single group resulting in a map with a resolution of 8.4 Å (Extended Data Fig. 7a–f and Table 2). The reconstruction gave a model that is a hybrid of the nucleotide-bound 'closed' fungal Get3 structures and the apo1 and 2 conformations (Fig. 2b and Extended Data Fig. 7a–f). As only a small percentage of particles are in this conformation, the rest of the particles in the data set are likely in a range of open conformations. Combined with the crystal structures, these structures represent the full range of Get3 conformations in the apo state.

Additionally, the structures reveal how the CBD of Get3 is occluded in the apo state to protect the hydrophobic surfaces that bind to the client. In previous structures, much of the CBD loops are disordered with the hydrophobic surfaces exposed, likely due to the inherent flexibility of this domain and crystal packing. Here, we see that the amphipathic helix 5 (H5) packs under H8 and against H7 and H9 covering H6. This buries the hydrophobic residues in these helices resulting in an overall hydrophilic surface and masks the client-binding hydrophobic groove (Supplementary Fig. 2b). While H8 has been modeled in two fungal 'open' structures, the positioning differs from what is seen here (PDB 3IBG and 2WOO, Supplementary Fig. 2c,d). The *G. intestinalis* apo structures reveal that H8 is critical to stabilizing H5 and completes the exposed hydrophilic surface. The active site resembles the fungal open structures where the catalytic loops all adopt similar conformations (Supplementary Fig. 2b)[19]. These structures both capture the range of apo Get3 conformations and demonstrate how the hydrophobic surfaces of the CBD are occluded in the absence of nucleotide.

The groove remains occluded in the ATP-bound state. The fully closed structure requires the binding of ATP. To capture this state, crystals were obtained of a catalytically dead mutant (D53N) in the presence of ATP and $MgCl_2$ (Extended Data Figs. 2a and 8a,b). A data set was collected and refined to 2.2 Å resolution in the space group $P3_221$ and contained a monomer in the asymmetric unit that formed a symmetry related dimer (Fig. 2c and Extended Data Fig. 8c–i). This structure is the first of Get3 bound to ATP alone, that is without any client or binding partners, and adopts a 'closed' form with the two monomers rotating closer together relative to the apo-closed state. All but one stretch (residues 88–114) can be modeled in this structure, providing the most complete structure of a 'closed' Get3 (Fig. 2c and Supplementary Fig. 3a,b). In the active site of the ATP-bound structure a Mg^{2+} ion is hexavalently coordinated by the γ- and β-phosphates of ATP, three water molecules and Thr_{28}. A water molecule is coordinated by the asparagine residue that replaced the catalytic aspartate and is positioned proximal to the γ-phosphate of the ATP molecule, primed for nucleophilic attack. Lys_{22} reaches across the dimer interface interacting with the β-phosphate of the ATP molecule in the other active site (Fig. 2d). The interactions between the two active sites stabilizes the canonical 'closed' form.

Notably, there are few differences in the positioning of α-helices within the Get3 monomer compared to the apo state. In particular, H5 continues to occlude the CBD concealing the hydrophobic groove distinct from other 'closed' Get3 structures (Supplementary Fig. 3c). In the recent structure of a metazoan pretargeting complex (Get3 bound to the loading complex) H8 forms a cap for the CBD[41]. Our structure confirms that H8 is amphipathic, but in the ATP-bound state its role is to stabilize H5, which occludes the CBD, and not to cap the hydrophobic pocket. H6 adopts a conformation similar to the apo state in an upward position away from the nucleotide pocket, whereas in the structure of the transition state

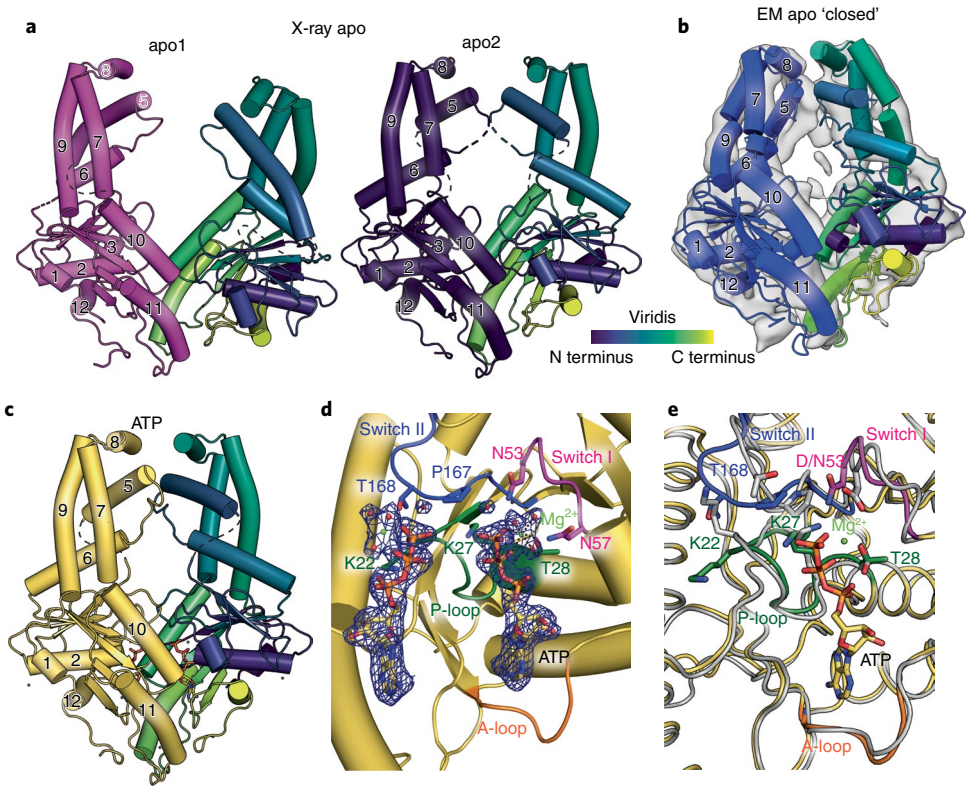


Fig. 2 | Structures of *Gi*Get3 in the 'open' and 'closed' states. a, The 3.0-Å resolution crystal structures of the two conformations of dimeric apo *Gi*Get3 referred to as apo1 (magenta) and apo2 (purple). **b**, The 8.4 Å cryo-EM map of apo *Gi*Get3 with two monomers from the apo1 crystal structure in **a** fitted. **c**, The crystal structure of $GiGet3_{D53N}$ bound to ATP at 2.2 Å. For all structures, helices are numbered based on the secondary structure prediction order and chain B is colored from N to C terminus using the viridis color map (purple to yellow)[49]. **d**, A Mg^{2+} ion in the active site of ATP-bound *Gi*Get3 is coordinated by the γ- and *β*-phosphates in the ATP molecule, three waters and Thr_{28}. **e**, A comparison of the active site from apo2 (gray) and $GiGet3_{D53N}$•ATP with the P-loop (green), A-loop (orange) and Switch I (magenta) and II (blue) highlighted. This coloring of the active site loops is preserved throughout the article.

($ADP{\cdot}AlF_4$ complex) and structures representing states further in the hydrolysis cycle (the pretargeting and Get2-fragment complexes) H6 has shifted down toward the nucleotide binding pocket, forming the bottom of the hydrophobic groove[17,18,20,22,23]. This shift is also not observed in the AMPPNP-bound *Chaetomium thermophilum*, *Ct*Get3, crystal structure (Supplementary Fig. 3d). Binding of ATP or its analogs are not sufficient to generate the reorganization of the CBD to form the hydrophobic groove, demonstrating the transition state represents Get3 in a conformation later in the cycle. The shift of H6 toward the active site likely requires H5 to rearrange from the occluded position. The previous fungal structures where H6 has shifted down likely represent the next steps that require Get4 binding to facilitate the transition state for client capture. In the ATP-bound state, ATP primes Get3 for Get4 binding by stabilizing the fully closed form, creating the Get4 binding interface and likely readies the CBD for subsequent rearrangement to form the groove for client capture. Together, the four structures of Get3 illustrate the range of conformations sampled by the ATPase detailing how regulation of client targeting is more complicated than a simple 'open' versus 'closed' conformation binary.

The structure of the posthydrolysis client-bound complex. We next sought to fill in the key missing piece of the Get3 ATPase cycle by structurally characterizing Get3 bound to a TA client in the posthydrolysis state. We turned to single-particle analysis cryo-EM to accomplish this. Wild-type *Gi*Get3 was recombinantly expressed with a His-tagged client, $BRIL\text{-}Bos1_{TMD}$, that had previously been used to investigate the binding mechanism of Sgt2 (Extended Data Fig. 9a)[35]. The complex was purified by affinity and size-exclusion chromatography (SEC) and the elution fraction correlating to 150 kDa was frozen on grids. Movies were collected and used to determine a reconstruction with an overall resolution of 4.1 Å where parts of the map had less well-defined density than others. A mask was generated that included regions with strong density and was used in a focus refinement resulting in a map with an average resolution of 3.7 Å (Fig. 3a,b and Extended Data Fig. 9b–e). This map has clear density for all of Get3 except for H8, the C terminus of H7 and the N terminus of H9, which were part of the masked low-density regions, likely due to disorder and are also regions missing in previous 'closed' structures. At this resolution (ranging from 3.6 to 7.0 Å), we were able to fully build the NBD as well as the sections of the CBD that form the hydrophobic groove (Extended Data Fig. 10a). Three new tubular densities are observed in the CBD that do not correspond to Get3 helices accounted for by previous 'closed' Get3 structures (highlighted in red and purple in Fig. 3a,b). These were modeled as poly-alanine helices as the registry could not be unambiguously assigned. The density colored in red coincides with the positioning of the TMD of the client in the prehydrolysis complex and the length is sufficient to accommodate a TMD, suggesting this density represents the Bos1 TMD. The two remaining densities in purple are symmetry related and can be accounted for by the previous loop between H4 and H5 that has transitioned to form two new features, which we refer to as H4/5 (Extended Data Fig. 1). Clear density for an ADP molecule and an Mg^{2+} ion is visible in the nucleotide binding pocket indicating that this map reflects the posthydrolysis form of the Get3–TA complex (Fig. 3c). The ADP is tightly associated with *Gi*Get3–TA complex as no nucleotide was added during purification.

Table 2 | Cryo-EM data collection, refinement and validation statistics

	Apo (EMD-25375)	Get3–TA (overall) (EMD-25374)	Get3–TA (Get3 only) (EMD-25373), (PDB 7SQ0)
Data collection and processing			
Microscope	FEI Titan Krios	FEI Titan Krios	
Voltage (kV)	300	300	
Camera	Gatan K3	Gatan K3	
Energy filter	BioQuantum	Gatan Imaging Filter Quantum	
Energy filter slit width (eV)	20	20	
Magnification (nominal)	88,000	105,000	
Defocus range (μm)	−0.5 to −2.5	−0.7 to −3.0	
Calibrated pixel size (Å/pix)	0.5295	0.433	
Electron exposure (e^-/Å^2)	50	63.2	
Exposure rate (e^-/Å^2 per frame)	1	1.58	
Number of frames per video	50	40	
Automation software	SerialEM	SerialEM	
Number of micrographs	9,300	2,732	
Initial particle images (no.)	11,596,225	1,790,962	
Final particle images (no.)	51,340	70,330	
Estimated accuracy of rations (°) (RELION)	3.355	3.461	3.302
Local resolution range	6.0–8.45	4.15–7.36	3.61–7.06
Map resolution (Å, FSC 0.143)	8.46	6.77	3.71
Model fitting			
Software (FLEX-EM)	CCPEM v.1.5.0		
Initial model used	apo1 (*Gi*Get3 chain A)		
Refinement			
Software (phenix.real_space_refine)		PHENIX v.1.16-3549	
Initial model used (PDB code)		5BW8	
Resolution of unmasked reconstruction (Å, FSC 0.5)		4.4	3.9
Resolution of masked reconstructions (Å, FSC 0.5)		4.2	3.6
Correlation coefficient (CC_{mask})		0.69	0.83
Map sharpening *B* factor (Å^2)		−145	−121
Model composition			
Nonhydrogen atoms		4,691	4,691
Protein residues		593	593
Ligand		5	5
B factors (Å^2)		Minimum/maximum/mean	Minimum/maximum/mean
Protein		52.25/165.11/81.59	52.25/165.11/81.59
Ligand		6.9/184.58/129.64	6.9/184.58/129.64
R.m.s. deviations			
Bond lengths (Å) (no. >4σ)		0.003 (4)	0.003 (4)
Bond angles (°) (no. >4σ)		1.119 (17)	1.119 (17)
Validation			
MolProbity score		1.06	1.08
Clashscore		1.97	1.97
Rotamer outliers (%)		0	0
Cβ outliers (%)		0	0
CaBLAM outliers (%)		0.67	0.67
EMRinger score		1.6	3.99
Ramachandran plot			
Favored (%)		97.4	97.4
Allowed (%)		2.6	2.6
Disallowed (%)		0	0

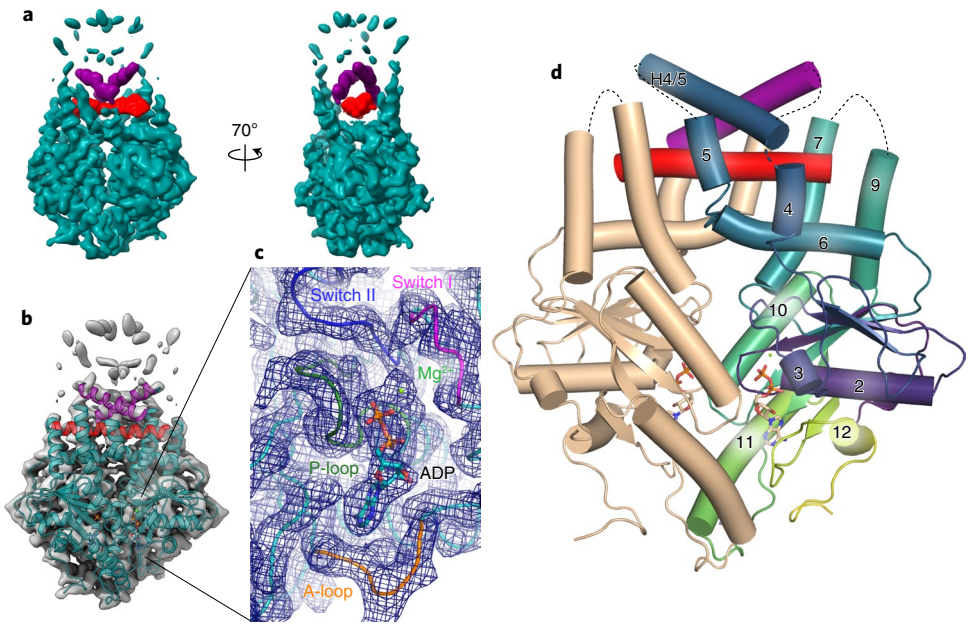


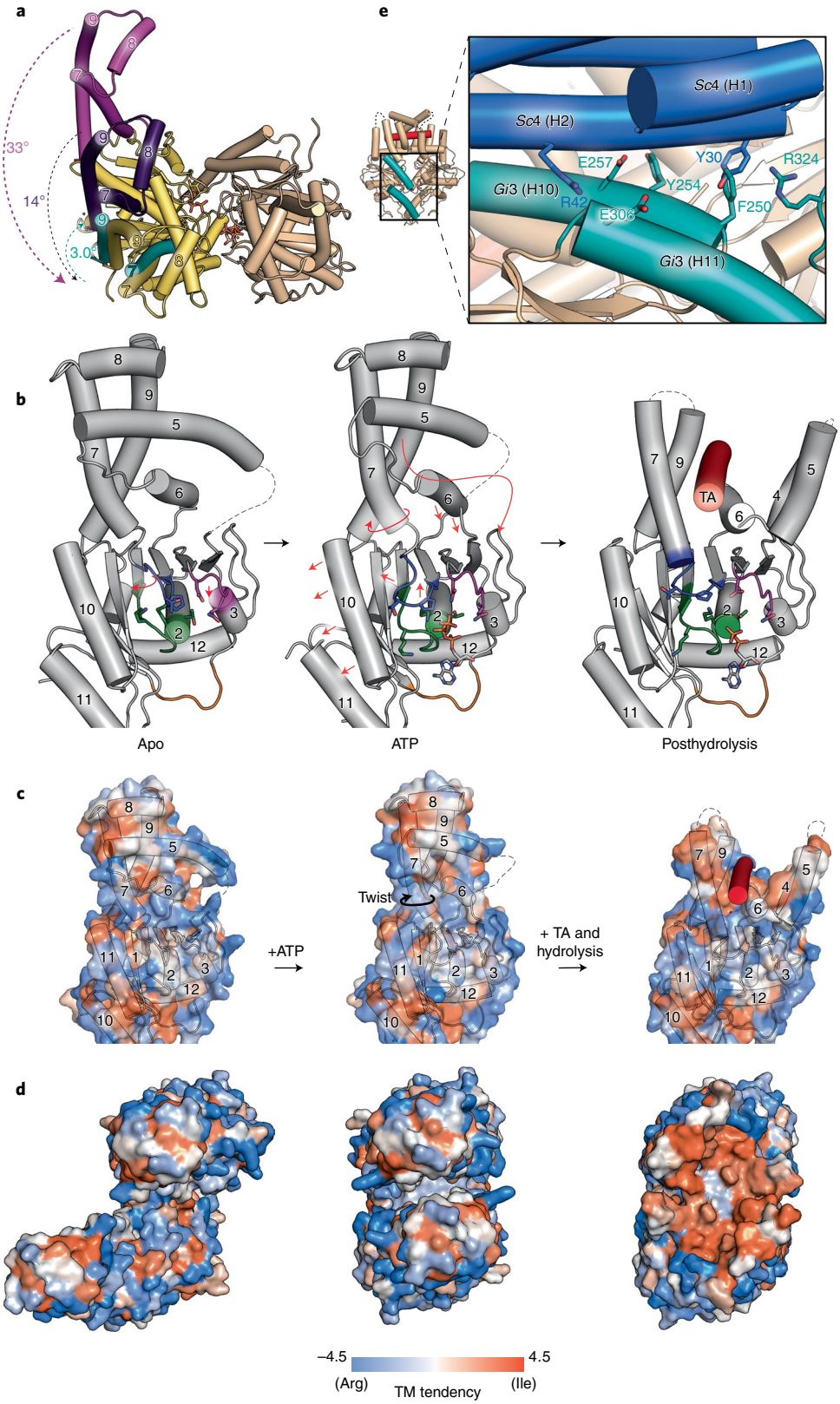
Fig. 3 | Cryo-EM structure of the posthydrolysis *Gi*Get3–TA complex. a, Two views of the unsharpened overall map of the *Gi*Get3–TA complex in the posthydrolysis state collected in Extended Data Fig. 9. Assigned density in the hydrophobic groove corresponding to the TMD (red) and H4/5 (purple) are colored. **b**, The sharpened map after a focused refinement with the molecular model of *Gi*Get3–TA complex built in. **c**, A view of the nucleotide binding pocket with the ADP molecule and the Mg^{2+} ion fit into density. **d**, The structure of the *Gi*Get3–TA complex in the 'intermediate' state, represented as a cartoon with one chain colored from the N to C terminus using the viridis color map (purple to yellow) and the other colored wheat with H4/5 colored in purple. The TMD from the TA protein is colored in red.

Additional features of the posthydrolysis complex are revealed by the structure (Fig. 3d). The Get3 dimer adopts a more open conformation consistent with the intermediate state observed from smFRET data (Fig. 4a and Supplementary Video 1)[13]. While disordered in the apo and ATP structures, in the posthydrolysis state H4 forms part of the sides of the client-binding groove in a position parallel to H7 and H9. H6 has moved to the lower position toward the active site that pulls H5 down and away from its position occluding the groove. H5 then joins H4, H7 and H9 in forming the walls of the hydrophobic groove. Relative to the transition-state structure, H4 shifts away from the center of the groove to accommodate the change in H6, resulting in an overall expansion of the CBD (Fig. 4b and Supplementary Video 1). H8, no longer stabilized by H5, becomes disordered. This is consistent with biochemical work that demonstrated that the deletion of H8 does not affect the targeting of TA proteins by Get3 and suggested another part of the protein forms the lid with client bound[42]. The most striking new feature in the posthydrolysis structure is that H4/5 docks on top of the client TMD completing the hydrophobic chamber and protecting the entire hydrophobic client TMD from solvent (Figs. 3 and 4c,d and Supplementary Fig. 5). Mechanistically, H8 likely regulates the opening of the hydrophobic groove to allow client capture, while H4/5 protects the client TMD. The posthydrolysis structure completes the overall cycle of TA protein targeting to the ER.

The binding and hydrolysis of ATP drives remodeling in Get3. The structure reveals that the various nucleotide states lead to considerable remodeling of Get3. Apo Get3 adopts a range of conformations swinging between fully open and closed. The 'closed' apo Get3 has a more expanded NBD relative to the ATP-bound state, likely to allow access of the active site by the nucleotide (Fig. 2a,b and Supplementary Fig. 6a). Upon ATP binding, the 'closed' Get3 state is stabilized, generating the binding interface for Get4 (Fig. 4a,e and Supplementary Fig. 6b). When transitioning from the apo to ATP-bound state, Switch I moves toward the nucleotide, placing the catalytic Asp_{53} above the γ-phosphate in the ATP molecule to position the water for nucleophilic attack (Figs. 2d and 4b, Extended Data Fig. 8h and Supplementary Videos 1 and 2). The backbone of the P-loop interacts with the α- and β-phosphates of the ATP molecule, wrapping around the nucleotide and Lys_{27} rotates toward the nucleotide, interacting with the β-phosphate. Switch II moves toward H10. The 'deviant Walker A' Lys_{22} in the P-loop shifts away to interact across the dimer interface with the β-phosphate of the ATP molecule of the other subunit. These changes result in the NBD of the two Get3 monomers coming closer together in a stabilized

Fig. 4 | Conformational changes induced by the nucleotide state and TA protein binding. a, A top view of the Get3 dimer with arrows and angles of rotation for each state measured from the P-loop highlighting the twist accompanying various nucleotide states. *Gi*Get3s are aligned by the P-loop in chain A. For clarity, only chain A from the ATP state of Get3 is shown (wheat) and H7-H9 of chain B are shown for the nucleotide-free (apo1 (purple) and apo2 (magenta)), ATP-bound (yellow) and posthydrolysis (cyan). **b**, Cartoon representations of the nucleotide binding cycle of *Gi*Get3 (gray): apo (left), ATP-bound (center) and posthydrolysis (right). Red arrows highlight substantial conformational changes between states. The TMD of the TA protein is colored in red. Active site features are colored as in Fig. 1a. **c**, Surface representation of the CBD of the views in **b** and helices are numbered in both **b** and **c**. A black arrow highlights the twist in H7. **d**, Top views of each state reveal the hydrophobic groove is exposed in the posthydrolysis state and H5 occludes this binding site in apo1 and ATP-bound. Hydrophobic surfaces are colored from blue (hydrophilic) to orange (hydrophobic) using the transmembrane tendency scale[40]. **e**, *Gi*Get3–TA (wheat) with helices that interact with Get4 colored in teal. *Sc*Get4 was modeled onto *Gi*Get3–TA using an alignment to the *Sc*Get3/4/5 structure (PDB 4PWX) (blue), highlighting how the movement of H10 disrupts the Get4 binding interface.

'closed' form. In this conformation, the Get4 binding interface is generated, priming Get3 for client capture. This interface is possibly further stabilized by a conserved cation-π interaction across the dimer interface between Phe_{250} and Arg_{324} (Supplementary Fig. 7). Cation-π interactions are known to stabilize protein interfaces[43]. While observed in fungal 'closed' Get3 structures (PDB 2WOJ, 4PWX and 4XTR), the cation-π is absent in the apo state (both fungal and *G. intestinalis*) (Supplementary Fig. 7a–d)[18,22,23]. Phe_{250} and Arg_{324} are conserved across Get3s and mutating these to Ala in yeast resulted in a phenotypic growth defect and decrease in Get4



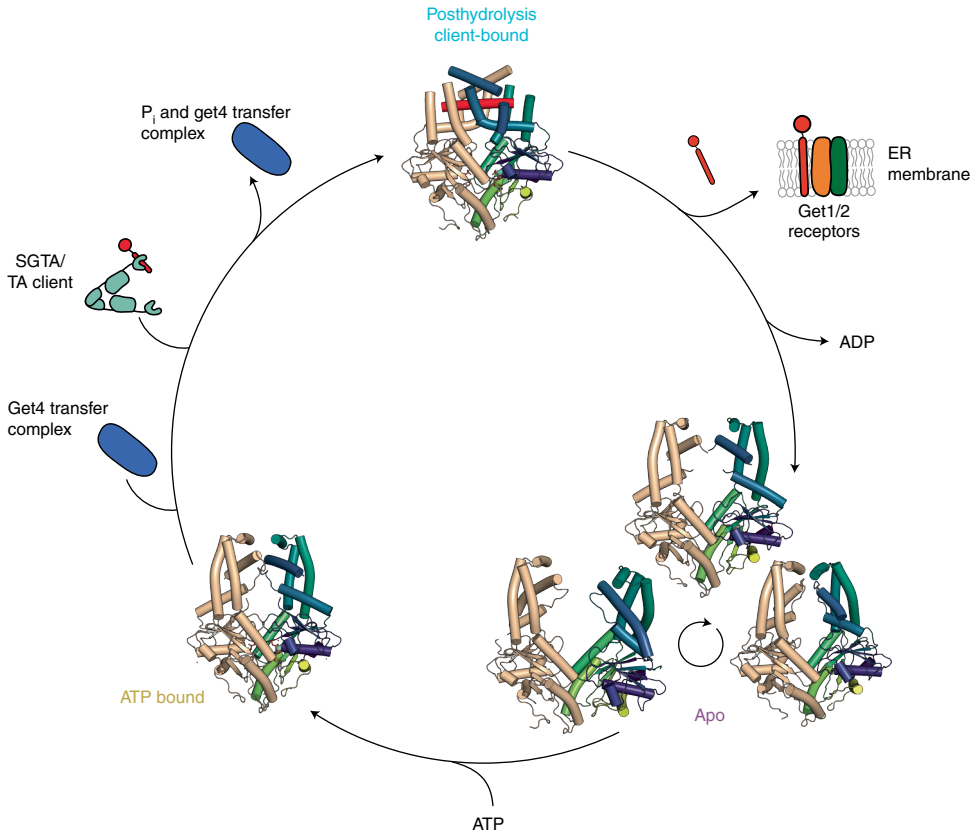


Fig. 5 | Universal model of the Get3 conformations that drive targeting. In the apo state, Get3 samples between the fully open and closed conformations. Binding of ATP stabilizes the 'closed' form, generating the Get4 binding interface and prepares for opening of the groove. Likely on Get4 binding, H5 in Get3 is released and rearrangements in the CBD reveal the hydrophobic groove, as seen in the structure of the yeast-Get3/4/5 complex[22]. Correct client binding would then cause H6 to move toward the nucleotide binding pocket, resulting in hydrolysis. Phosphate release allows the Get3 NBD to adopt a slightly more open conformation while maintaining the hydrophobic groove. In this state, the bound TA client is completely shielded by the loop connecting H4 and H5 and the Get4 binding interface is disrupted, resulting in release from the Get4 transfer complex. The Get3-TA complex is now primed for delivery to the ER membrane where the receptors facilitate ADP release and client insertion.

binding[19,22]. These interactions demonstrate how nucleotide binding regulates interactions between Get3 and Get4 even though apo Get3 adopts a 'closed' conformation.

Additional specific changes are observed in the posthydrolysis state. First, the Get3 dimer rotates to an intermediate conformation between the apo2 and the ATP-bound state (Fig. 4a,d). In this intermediate state the, cation-π interaction between Phe_{250} and Arg_{324} is preserved (Supplementary Fig. 7f). The most notable change in the active site is where Pro_{166} moves away from the nucleotide allowing the C terminus of Switch II to transition from a loop into a helix resulting in a twist in H7 and the formation of a new N terminus for the helix (Fig. 4b and Extended Data Fig. 1). This twist in H7 results in the hydrophobic residues of the helix turning inward toward the center of the client-binding groove, creating the hydrophobic interior (Fig. 4b–d). This helical transition of Switch II results in H10 moving away from the nucleotide and accommodates the downward movement of H6. H4 shifts away from the active site avoiding a clash with H6 and, together with H10 movements, results in a swelling of the Get3 monomer. As the cation-π and other interactions across the dimer interface are preserved, most of the changes in the intermediate conformation are consequences of the rearrangements to Switch II and the resulting shift in H10 (Fig. 4e).

While the Get3–Get4 binding interface has been explored biochemically and structurally, how this interface is disrupted by client recognition was unknown. Here, the observed shift in H10 and slight opening of the dimer explains the disruption of the Get4 binding interface. Aligning a Get3 monomer between the posthydrolysis Get3–TA complex and the yeast–Get3/4/5 crystal structure (PDB 4PWX) allows visualization of the Get4 interaction and shows that the Get3 H10 movement generates clashes with residues on H2 of Get4 (Fig. 4e)[22]. These changes do not disrupt the hydrophobic groove, clarifying how TA proteins remain bound to Get3 after Get4 disassociation (Fig. 4b–e and Supplementary Fig. 7).

Discussion

This study provides a comprehensive view of the catalytic cycle of Get3, demonstrating the conformational changes regulated by nucleotide binding and hydrolysis (Fig. 4 and Supplementary Video 1) and completes the model of TA protein targeting by Get3 (Fig. 5). In the apo state, Get3 fluctuates between the 'closed' and 'open' forms as highlighted by our three apo structures. H5 hides the hydrophobic surfaces of the CBD. Nucleotide binding results in a stabilized 'closed' complex where Get3 can bind to Get4 in preparation for client capture. Evidence suggests that H8 plays an important role in client loading[42] where the interaction with Get4 in the pretargeting complex may facilitate the release of H5 leading to the other CBD conformational changes. Get4 binding induces the conformational changes in Get3 to prime it for client capture, specifically the formation of the client-binding pocket with H5 moving to become parallel with H7 and H9. H6 must shift down to drive catalysis and the lower orientation is most likely stabilized by client binding. This interaction would result in ATP hydrolysis followed

by phosphate release. Get3 detaches from Get4 through conformational changes transmitted to the Get3 surface by changes in the active site. Disassociated from Get4 causes the release of H4 and the proceeding loop, resulting in the formation of H4 and H4/5. In this form, Get3 remains bounds to TA proteins completely shielding the hydrophobic TMD with the lid formed by H4/5.

The work presented here is a comprehensive picture of Get3 from a single organism. Despite the evolutionary distance, the GET pathway shows remarkable conservation across eukaryotes, highlighting the importance of this essential process. The fully masked TA-binding groove in the apo and ATP state would prevent inadvertent association with nonspecific targets, explaining why, in vitro, Get3 alone cannot capture a TA client and requires Get4 (ref. [9]). After release from Get4, the Get3–TA complex would be capable of binding to Get2 at the ER as the membrane-bound receptor does not bind across the Get3 dimer interface[17]. Once localized to the membrane, the released H8 may then bind the membrane, destabilizing the Get3–TA complex so that Get3 can be opened and release the client that is favored due to Get1 binding an open Get3. These structures demonstrate that Get3 is a dynamic protein, sampling many different conformational states during client targeting. Altogether, this work completes the composite model across multiple organisms for TA protein targeting by the GET pathway, providing detailed mechanistic insight into nucleotide regulation and demonstrating a universal Get3 mechanism in eukaryotes. Questions for the pathway remain as to how Sgt2 hands off to Get3 in a privileged transaction[44] and how the Get1–Get2 complex facilitates TA protein insertion. As the GET pathway is a drug target[45], this deeper understanding may prove critical to facilitating that potential.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-022-00798-4.

Received: 2 December 2021; Accepted: 26 May 2022;
Published online: 18 July 2022

References

1. Guna, A. & Hegde, R. S. Transmembrane domain recognition during membrane protein biogenesis and quality control. *Current Biol.* **28**, R498–R511 (2018).
2. Kutay, U., Hartmann, E. & Rapoport, T. A. A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol.* **3**, 72–75 (1993).
3. Denic, V. A portrait of the GET pathway as a surprisingly complicated young man. *Trends Biochem. Sci.* **37**, 411–417 (2012).
4. Wattenberg, B. W. & Lithgow, T. Targeting of C-terminal tail-anchored proteins: understanding how cytoplasmic activities are anchored to intracellular membranes. *Traffic* **2**, 66–71 (2001).
5. Chartron, J. W., Clemons, W. M. Jr. & Suloway, C. J. The complex process of GETting tail-anchored membrane proteins to the ER. *Curr. Opin. Struct. Biol.* **22**, 217–224 (2012).
6. Borgese, N., Colombo, S. & Pedrazzini, E. The tale of tail-anchored proteins. *J. Cell Biol.* **161**, 1013–1019 (2003).
7. Rabu, C., Schmid, V., Schwappach, B. & High, S. Biogenesis of tail-anchored proteins: the beginning for the end? *J. Cell Sci.* **122**, 3605–3612 (2009).
8. Cavalier-Smith, T. & Chao, E. E.-Y. Phylogeny and classification of phylum Cercozoa (protozoa). *Protist* **154**, 341–358 (2003).
9. Wang, F., Brown, E. C., Mak, G., Zhuang, J. & Denic, V. A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell* **40**, 159–171 (2010).
10. Simpson, P. J., Schwappach, B., Dohlman, H. G. & Isaacson, R. L. Structures of Get3, Get4, and Get5 provide new models for TA membrane protein targeting. *Structure* **18**, 897–902 (2010).
11. Rome, M. E., Rao, M., Clemons, W. M. & Shan, S. O. Precise timing of ATPase activation drives targeting of tail-anchored proteins. *Proc. Natl Acad. Sci. USA* **110**, 7666–7671 (2013).
12. Chartron, J. W., Gonzalez, G. M. & Clemons, W. M. A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *J. Biol. Chem.* **286**, 34325–34334 (2011).
13. Chio, U. S., Chung, S., Weiss, S. & Shan, S. O. A protean clamp guides membrane targeting of tail-anchored proteins. *Proc. Natl Acad. Sci. USA* **114**, E8585–E8594 (2017).
14. Chartron, J. W., Suloway, C. J. M., Zaslaver, M. & Clemons, W. M. Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl Acad. Sci. USA* **107**, 12127–12132 (2010).
15. Shan, S. O. Guiding tail-anchored membrane proteins to the endoplasmic reticulum in a chaperone cascade. *J. Biol. Chem.* **294**, 16577–16586 (2019).
16. McDowell, M. A. et al. Structural basis of tail-anchored membrane protein biogenesis by the GET insertase complex. *Molecular Cell* **80**, 72–86.e7 (2020).
17. Mariappan, M. et al. The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* **477**, 61–66 (2011).
18. Mateja, A. et al. The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* **461**, 361–366 (2009).
19. Suloway, C. J. M., Chartron, J. W., Zaslaver, M. & Clemons, W. M. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl Acad. Sci. USA* **106**, 14849–14854 (2009).
20. Stefer, S. et al. Structural basis for tail-anchored membrane protein biogenesis by the Get3-receptor complex. *Science* **333**, 758–762 (2011).
21. Bozkurt, G. et al. Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proc. Natl Acad. Sci. USA* **106**, 21131–21136 (2009).
22. Gristick, H. B. et al. Crystal structure of ATP-bound Get3–Get4–Get5 complex reveals regulation of Get3 by Get4. *Nat. Struct. Mol. Biol.* **21**, 437–442 (2014).
23. Mateja, A. et al. Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science* **347**, 1152–1155 (2015).
24. Koonin, E. V. A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol.* **229**, 1165–1174 (1993).
25. Yamagata, A. et al. Structural insight into the membrane insertion of tail-anchored proteins by Get3. *Genes Cells* **15**, 29–41 (2010).
26. Xing, S. et al. Loss of GET pathway orthologs in *Arabidopsis thaliana* causes root hair growth defects and affects SNARE abundance. *Proc. Natl Acad. Sci. USA* **114**, E1544–E1553 (2017).
27. Kumar, T., Maitra, S., Rahman, A. & Bhattacharjee, S. A conserved Guided Entry of Tail-anchored pathway is involved in the trafficking of a subset of membrane proteins in *Plasmodium falciparum*. *PLoS Pathog.* **17**, 1–39 (2021).
28. Bodensohn, U. S. et al. The intracellular distribution of the components of the GET system in vascular plants. *Biochim. Biophys. Acta: Mol. Cell Res.* **1866**, 1650–1662 (2019).
29. Anderson, S. A., Satyanarayan, M. B., Wessendorf, R. L., Lu, Y. & Fernandez, D. E. A homolog of guided entry of tail-anchored proteins functions in membrane-specific protein targeting in chloroplasts of *Arabidopsis*. *Plant Cell* **33**, 2812–2833 (2021).
30. Adam, R. D. *Giardia duodenalis*: biology and pathogenesis. *Clin. Microbiol. Rev.* https://doi.org/10.1128/CMR.00024-19 (2021).
31. Rome, M. E., Chio, U. S., Rao, M., Gristick, H. & ou Shan, S. Differential gradients of interaction affinities drive efficient targeting and recycling in the get pathway. *Proc. Natl Acad. Sci. USA* **111**, E4929–E4935 (2014).
32. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
33. Asseck, L. Y. et al. Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.2017636118 (2021).
34. Yamamoto, Y. & Sakisaka, T. Molecular machinery for insertion of tail-anchored membrane proteins into the endoplasmic reticulum membrane in mammalian cells. *Mol. Cell* **48**, 387–397 (2012).
35. Lin, K.-F., Fry, M. Y., Saladi, S. M. & Clemons, W. M. Molecular basis of tail-anchored integral membrane protein recognition by the cochaperone Sgt2. *J. Biol. Chem.* https://doi.org/10.1016/j.jbc.2021.100441 (2021).
36. Cho, H. & Shan, S. O. Substrate relay in an Hsp70-cochaperone cascade safeguards tail-anchored membrane protein targeting. *EMBO J.* https://doi.org/10.15252/embj.201899264 (2018).
37. Fry, M. Y., Saladi, S. M., Cunha, A. & Clemons, W. M. Jr Sequence-based features that are determinant for tail-anchored membrane protein sorting in eukaryotes. *Traffic* **22**, 306–318 (2021).
38. Aurrecoechea, C. et al. GiardiaDB and trichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* **37**, D526–D530 (2008).
39. Schuldiner, M. et al. The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* **134**, 634–645 (2008).

40. Zhao, G. & London, E. An amino acid 'transmembrane tendency' scale that approaches the theoretical limit to accuracy for prediction of transmembrane helices: relationship to biological hydrophobicity. *Protein Sci.* **15**, 1987–2001 (2006).
41. Keszei, A., Yip, M., Hsieh, T.-C. & Shao, S. Structural insights into metazoan pretargeting get complexes. *Nat. Struct. Mol. Biol.* **28**, 1029–1037 (2021).
42. Chio, U. S., Chung, S., Weiss, S. & Shan, S. A chaperone lid ensures efficient and privileged client transfer during tail-anchored protein targeting. *Cell Reports* **26**, 37–44.e7 (2019).
43. Salonen, L. M., Ellermann, M. & Diederich, F. Aromatic rings in chemical and biological recognition: energetics and structures. *Angew. Chem. Int. Ed.* **50**, 4808–4842 (2011).
44. Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H. & Hegde, R. S. Mechanistic basis for a molecular triage reaction. *Science* **355**, 298–302 (2017).
45. Morgens, D. W. et al. Retro-2 protects cells from ricin toxicity by inhibiting ASNA1-mediated ER targeting and insertion of tail-anchored proteins. *eLife* **8**, e48434 (2019).
46. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
47. Hehl, A. B. & Marti, M. Secretory protein trafficking in *Giardia intestinalis*. *Mol. Microbiol.* **53**, 19–28 (2004).
48. Martincová, E. et al. Probing the biology of *Giardia intestinalis* mitosomes using *in vivo* enzymatic tagging. *Mol. Cell. Biol.* **35**, 2864–2874 (2015).
49. Saladi, S. M., Maggiolo, A. O., Radford, K. & Clemons, W. M. Structural biologists, let's mind our colors. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.22.308593 (2020).

Methods

Sequence alignments. Alignments of Get3, Get4 and Sgt2 were created by downloading genes from *G. intestinalis*, *Homo sapiens*, *Saccharomyces cerevisiae*, *Amphimedon queenslandica*, *Schizosaccharomyces pombe*, *Neisseria crassa*, *Aspergillus fumigata*, *Methanocaldococcus jannaschii* and *P. falciparum* from UniProt. Sequences were aligned with PROMALS3D (ref. [50]) along with all experimentally determined structures of Get3, Get4 and Sgt2 homologs. PROMALS3D provides a way of integrating a variety of costs into the alignment procedure, including 3D structure, secondary structure predictions and known homologous positions. Alignments were visualized using Jalview[51].

Phylogenetic tree. Homologs of Get3 were identified using a local HMMER search[52] against the UniProt database[53] and against the EukProt database[54]. Annotated sequences were selected for alignment using MAFFT[55] and this alignment was used for the initial search. Sequences with positive hits were verified by reverse analysis by HHpred algorithm[56] against the Pfam database[57]. Multiple sequence alignments were enriched by newly classified homologs and used for future searches. Sequences were clustered in MMseqs2 (ref. [58]) using the easy-cluster algorithm with a minimum sequence identity cutoff of 0.5. Sequences were aligned using MAFFT[55] using the default LINS-i algorithm parameters. Poorly aligned regions with a gap threshold greater than 0.5 were automatically removed by trimAI[59]. Phylogenetic analysis was conducted with IQ-TREE2 software[60] and the best fitting model was selected using ModelFinder[61]. Computational effort was reduced by implementing the UFBoot2 (ultrafast bootstrap) method[62]. Visualization of the tree was done using the iTOL online tool[63].

Cell culture, cloning and transfection in *G. intestinalis*. The *G. intestinalis* strain WB (ATCC 30957) was grown in TYI-S-33 medium supplemented with 10% heat-inactivated bovine serum, 0.1% bovine bile and antibiotics at 37 °C (ref. [64]). The gene encoding the Get3 homolog (*GL50803_007953*) was amplified from genomic DNA and inserted into the pOndra plasmid[65] with a C-terminal BAP-tag. Then 1×10^7 cells expressing cytosolic BirA[48] were electroporated with a Bio-Rad Gene Pulser using an exponential protocol ($U = 30$ V; $C = 1{,}000$ µF; $R = 750$ Ω). The transfected cells were grown in medium supplemented with antibiotics (58 µg ml^{-1} puromycin and 600 µg ml^{-1} G418) and used for biotinylated Get3 isolation to establish protein partners.

Immunofluorescence microscopy. *G. intestinalis* trophozoites were fixed in 1% paraformaldehyde for 30 min at 37 °C, collect by centrifugation at 1,000*g* for 5 min, washed in PEM buffer (100 mM PIPES pH 6.9, 1 mM EGTA and 0.1 mM $MgSO_4$) and placed on cover slips (previously described in ref. [66]). The cells were permeabilized by 0.2% Triton X-100 for 20 min, washed three times with PEM buffer. For confocal microscopy, *G. intestinalis* cells expressing HA-tagged PDI2 were used (Fig. 1c). PDI2 was detected by primary anti-HA tag antibody at a 1:1,000 dilution (Roche) and *Gi*Get3 was detected with polyclonal anti-Get3 antibody raised in rat at a 1:1,000 dilution. Recombinant *Gi*Get3 protein was used as an antigen for in-house production of a polyclonal antibody in rats. The antibody was validated on the western blot using the recombinant protein. This band corresponded to the same protein band as a commercial anti-BAP antibody (Genescript) at a 1:1,000 dilution when detecting a BAP-tagged *Gi*Get3. As secondary antibodies, Alexa Fluor 488-conjugated goat antirat IgG and Alexa Fluor 594-conjugated donkey antirabbit IgG (Invitrogen) were used at a 1:1,000 dilution. Slides were mounted with Vectashield containing DAPI (4′,6-diamidino-2-phenylindole, Vector Laboratories) and imaged with a Leica SP8 FLIM inverted confocal microscope. The images were deconvolved using the Scientific Volume Imaging Huygens software with the CMLE algorithm (http://svi.nl). Maximum intensity projections and brightness/contrast corrections were performed in the FIJI ImageJ software[67].

For stimulated emission depletion (STED) microscopy, cells expressing HA-tagged Get3 and V5-tagged Get2 were used (Extended Data Fig. 4b,c). Rat anti-HA tag monoclonal antibody for Get3 (1:1,000 dilution) (Roche), rabbit anti-V5 tag monoclonal antibody for Get2 (1:1,000 dilution) (Abcam) and mouse anti-*Gi*PDI2 polyclonal antibody produced in house and verified through western blot of recombinant proteins and bands corresponded to the same protein band as tagged proteins detected by commercial anti-BAP antibody and against previously published work[47] were used. Thereafter, secondary antibodies Abberior STAR 580 goat antirabbit IgG and Abberior STAR 635 goat antimouse IgG (Extended Data Fig. 4b), and Abberior STAR 580 goat antirat IgG and Abberior STAR 635 goat antirabbit IgG (Extended Data Fig. 4c) were used at a 1:1,000 dilution (Abberior Instruments GmbH). The slides were mounted with Abberior mounting medium. STED microscopy was performed on a commercial Abberior STED 775 QUAD scanning microscope (Abberior Instruments GmbH) equipped with a Ti-E Nikon body, QUAD beam scanner, Easy3D STED Optics Module and Nikon CFI Plan Apo ×60 oil immersion objective (numerical aperture 1.40).

Fractionation of G. intestinalis cells. *G. intestinalis* cells were gathered in cold phosphate buffered saline (PBS), pH 7.4 by centrifugation 1,000*g* at 4 °C for 10 min, washed with 20 mM MOPS, 250 mM sucrose, pH 7.4 and again collected by centrifugation. The pellet was resuspended in 20 mM MOPS, 250 mM sucrose, pH 7.4 supplemented with protease inhibitors (cOmplete Protease Inhibitor Cocktail). The cells were lysed on ice by sonication for 2 min (1 s pulses, 40% amplitude). The lysate was subjected to centrifugation at 2,680*g* and 4 °C, for 20 min to sediment nuclei, cytoskeleton and remaining unbroken cells. The supernatant was subjected to centrifugation at 180,000*g*, for 30 min at 4 °C. The resulting supernatant corresponded to the cytosolic fraction and the high-speed pellet contained organelles including mitosomes and the ER.

Pull-down assays of Get3. *G. intestinalis* cells coexpressing biotintylated BAP-tagged Get3 and BirA were grown in TYI-S-33 medium with 50 µM biotin for 24 h before collection. The cell lysate was diluted to a final concentration of 1 mg ml^{-1} in PBS (pH 7.4) supplemented with protease inhibitors and incubated with 50 µl of streptavidin-coupled magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen) for 1 h at 4 °C with gentle rotation. Isolation was made in quadruplicates, where each sample contained 5 mg of proteins. The magnetic beads were washed three times in 50 mM HEPES pH 7.4, 150 mM potassium acetate, 5 mM magnesium acetate, 1 mM DTT and 10% glycerol then were washed three times in PBS. Beads with bound proteins were submitted to tandem mass spectrometry (MS/MS) analysis as previously described except without the detergent washing steps[68]. In brief, captured samples were released from beads by trypsin cleavage. Peptides were separated by reverse phase liquid chromatography and eluted peptides were converted to gas-phase ions by electrospray and analyzed using an Orbitrap (Thermo Scientific) followed by MS/MS to fragment the peptides through a quadropole for final mass detection. Data was analyzed using MaxQuant (v.1.6.3.4)[69] with a false discovery rate of 1% for both proteins and peptides and a minimum peptide length of seven amino acids. The Andromeda search engine[70] was used for the MS/MS spectra search against the *G. intestinalis* protein database (downloaded on 11 July, 2020)[38]. We carried out data analysis using Perseus v.1.6.1.3 (ref. [71]) and visualized as a volcano plot using the online tool VolcaNoseR[72].

Protein cloning, expression and purification in *E. coli*. Full-length *Gi*Get3 used to form Get3–TA complexes was cloned into a pET28a vector. For experiments using only *Gi*Get3 and single point mutants, the gene was cloned in a pET28a vector that was modified to have an N-terminal His-tag and SUMO fusion protein. To create the TA protein variant used for cryo-EM experiments, a BRIL fusion protein, a thermostabilized apocytochrome b562 (with mutations M7W, H102I and R106L)[73], followed by the TMD of *Sc*Bos1 was cloned into a pACYC-Duet vector modified to contain a tobacco etch virus cleavage site between an N-terminal His-tag and the fusion protein. The TMD of *Gi*TA proteins were cloned into a pET21b vector and flanked by an N-terminal 3xStrep-tag followed by a modified noncleavable SUMO fusion protein and a C-terminal opsin tag[42].

*Gi*Get3 and mutants were expressed in *E. coli* NiCo21(DE3) cells (New England BioLabs) in 2xYT broth at 37 °C and induced by 0.3 mM isopropyl-β-d-thiogalactoside (IPTG) at an optical density (OD_{600}) of around 0.6. Cells were gathered 4 h after induction. Cells were disrupted in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 20 mM imidazole and 10 mM β-ME) supplemented with protease inhibitors, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM benzamidine using a M-110 Microfluidizer Processor (Microfluidics). Lysate was centrifuged to separate the soluble and membrane fractions. Protein was purified by batch incubation with NiNTA resin at 4 °C for 1.5 h. Resin was washed with lysis buffer and *Gi*Get3 and mutants were eluted in 50 mM Tris, 300 mM NaCl, 300 mM imidazole and 10 mM β-ME, pH 7.5. The affinity tag was removed from the elution collected after nickel chromatography by an overnight ULP1 digestion in dialysis buffer (20 mM Tris, 150 mM NaCl and 10 mM β-ME). The dialyzed fraction flowed over a nickel column again to remove the His-SUMO particles and *Gi*Get3 was collected in the flowthrough. This pool was further purified using SEC through a HiLoad 16/600 Superdex 200 (GE). Protein purified for crystallography purposes were purified in 10 mM Tris, 75 mM NaCl and 10 mM β-ME while protein used in ATPase assays was purified in 50 mM HEPES, 150 mM KOAc and 10 mM β-ME.

Get3–TA complexes were formed by coexpressing TA proteins with tag-less *Gi*Get3 in *E. coli* NiCo21(DE3) cells in 2xYT broth at 37 °C and induced by 0.5 mM IPTG at an OD_{600} of around 0.7. Cells were lysed in 50 mM Tris, 300 mM NaCl and 10 mM β-ME, pH 7.5 supplemented with 1 mM PMSF and benzamidine using a M-110 Microfluidizer Processor (Microfluidics). The lysate was separated by centrifugation. Complexes used for structural determination via cryo-EM were purified by incubating the soluble fraction with NiNTA resin at 4 °C for 1.5 h. Resin was washed with 50 mM Tris, 300 mM NaCl, 35 mM imidazole and 10 mM β-ME, pH 7.5 and protein was eluted in 50 mM Tris, 300 mM imidazole and 10 mM β-ME, pH 7.5. The elutate was further purified via SEC using a HiLoad 16/600 Superdex 200 (GE) column.

*Gi*Get3–TA complexes used for pull-down experiments were purified by affinity chromatography using Strep-tactin resin (IBA Lifesciences). The resin was washed with lysis buffer and complexes were eluted in lysis buffer plus 2.5 mM desthiobiotin.

Full-length *Gi*Sgt2 was cloned into a pET33b vector that was modified to have an N-terminal His-tag and tobacco etch virus cleavage site. *Gi*Sgt2 was expressed in *E. coli* NiCo21(DE3) cells in 2xYT broth at 37 °C and induced by 0.3 mM IPTG at and OD_{600} of roughly 0.6. Cells were gathered 4 h after inductions and disrupted in lysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 mM imidazole and

10 mM β-ME) supplemented with protease inhibitors, 1 mM PMSF and 1 mM benzamidine using a M-110 Microfluidizer Processor (Microfluidics). Lysate was centrifuged to separate the soluble and membrane fractions. Protein was purified by batch incubation with NiNTA resin at 4 °C for 1.5 h. Resin was washed with wash buffer (20 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole and 10 mM β-ME) and protein was eluted in 20 mM Tris, 150 mM NaCl, 300 mM imidazole and 10 mM β-ME, pH 7.5.

Crystallization. Purified *Gi*Get3 or *Gi*Get3-D53N were concentrated to 10–12 mg ml^{-1} and crystal trays were set using the hanging-drop vapor-diffusion method by equilibrating equal volumes of protein and well liquor in VDX plates with sealant (Hampton Research). Apo *Gi*Get3 crystals formed in 0.1 M MES pH 5.3, 0.1 M $MgCl_2$ and 21% PEG3350 at 4 °C. *Gi*Get3$_{D53N}$ was incubated with 5 mM ATP and 2 mM $MgCl_2$ on ice for 1 h before setting trays and ATP-bound crystals formed in 0.1 M Tris pH 7.5, 0.2 M ammonium sulfate and 15% PEG3350 at room temperature. Crystals were cryoprotected by transfer into 30 µl of well liquor supplemented with 2 mM $MgCl_2$ and, in the case of *Gi*Get3$_{D53N}$, 5 mM ATP and increasing amounts of glycerol (10%, 15% and 20%). Crystals were incubated in each cryoprotectant drop for <5 min before flash freezing in liquid nitrogen.

Data collection, structure determination and refinement. Both the apo and ATP structures were solved using data sets collected from single crystals on beamline 12-2 at Stanford Synchrotron Radiation Lightsource (SSRL) at 12.6 keV using Blu-Ice (SSRL, https://smb.slac.stanford.edu/facilities/software/blu-ice/). Data for both the apo and ATP data sets were integrated and scaled using HKL3000 (ref. [74]) to a resolution of 3.0 and 2.23 Å, respectively. For the rest of the refinement both structures were determined using the same methods. Phases were obtained via molecular replacement in PHENIX[75–77], using the monomer of yeast–Get3 from PDB 3IBG (ref. [19]) and sequences were adjusted using Sculptor[78]. Refmac5 and phenix.refine were used for refinement and MolProbity was used to identify outliers[77,79,80]. Manual building was done using COOT with the final refinement done in Refmac5 in the CCP4 package suite[79,81]. For the apo crystal structure, the two monomers in the asymmetric unit contained residues 1–85, 106–153 and 156–347 (328 out of 357) for apo1 and residues 1–85, 118–134 and 138–345 (313 out of 357) for apo2 with a zinc ion for each on a symmetry axis. The final apo model had an R/R_{free} of 0.3 / 0.35 with no Ramachandran outliers. For the ATP crystal structure, the asymmetric unit contained a monomer with residues 1–88 and 115–346 (322 out of 357) modeled, an ATP, a Mg^{2+} ion, two coordinated zinc molecules on symmetry axes and a sulfate ion. The final ATP model had an R/R_{free} of 0.18 / 0.21 with no Ramachandran outliers. Full refinement statistics are in Table 1.

In the apo structure, each of the two modeled Zn^{2+} ions are coordinated across a dimer interface by the same cysteine pair, Cys_{287} and Cys_{290}, from two symmetry related monomers (Extended Data Fig. 6e,f). For the ATP structure, the same pair also coordinates a Zn^{2+} ion between a symmetry related dimer (Extended Data Fig. 8c,d). A second Zn^{2+} ion is present at a crystallographic interface coordinated by residues His_6 and Asp_{10} between symmetry related monomers (Extended Data Fig. 8e). All the Zn^{2+} ions were refined as half occupancy to account for their special positions. In addition, in the ATP structure the ATP molecule and Mg^{2+} ion are located in the protein active site (Extended Data Fig. 8f–g). Due to the high concentration of ammonium sulfate in the crystallization solution, extra electron density for a tetrahedral molecule was modeled as a SO_4^{2-} ion and is bound to two arginines (Arg_{291} and Arg_{291}, Extended Data Fig. 8i).

Cryo-EM grid preparation and data collection. EM samples for the *Gi*Get3–His-BRIL-Bos1$_{TMD}$ complex were collected from elution fractions from SEC. Then 3 µl at a concentration of roughly 0.73 mg ml^{-1} were placed on Holey carbon grids (Quantifoil R1.2/1.3, 300 mesh) that were glow discharged in air with a 20 A plasma current for 2 min using a Pelco easiGlow, Emeritech K100X. Grids were blotted at a force of 10 for 3.5 s and frozen in liquid ethane with the chamber at 4 °C and 100% humidity using a FEI Vitrobot Mark v4 x2. Data were collected using an automated data collection program, SerialEM[82], on a FEI Titan Krios equipped with an energy filter (20 eV slit width) at 300 keV and a Gatan K3 direct detector. Beam illumination was adjusted to a fluence of 13 e$^-$/upixels per Å. Images were collected using a defocus range of −0.7 to −3.0 µm using super resolution mode at a calibrated pixel size of 0.433 Å per pixel. Using counting mode, 1.82-s images were collected with a frame rate of 45.5 ms and dosage of 1.58 e$^-$/Å per frame.

EM samples for the purified apo *Gi*Get3 were collected from elution fractions from SEC and diluted to around 0.55 mg ml^{-1}. Then 3 µl of sample was placed on Holey carbon grids (Quantifoil R2/2, 100 mesh NH2 Finders), which were treated in the same manner as the grid prep for the complex. Data were collected using SerialEM on a FEI Titan Krios equipped with an energy filter (Gatan Imaging Filter) at 300 keV and a Gatan K3 direct electron detector. Images were collected using a defocus range of −0.5 to −2.5 µm using super resolution mode at a calibrate pixel size of 0.5295 Å per pixel. In counting mode, a total of 50 frames were collected for a total dosage of 50 e$^-$/Å^2.

Image processing. For the *Gi*Get3–TA complex data set, 2,732 videos were initially processed using cryoSPARCv.3.2.0 (ref. [83]) to produce aligned dose-weighted micrographs (Extended Data Fig. 9c,d). During motion correction, videos were downsampled to a corrected pixel size of 0.866 Å per pixel and all downstream processing was done at this pixel size. Of the 2,732 videos, 2,356 were manually selected for contrast transfer function (CTF) refinement (CTFFIND4, ref. [84]) and further processing. A small set of particles were manually picked and used for template-based picking and manually filtered to remove obvious debris, resulting in 1,790,962 particles. An initial round of two-dimensional (2D) classification was used for further particle filtration, resulting in 555,998 particles. Four ab initio models were generated using cryoSPARC and two classes were consistent with the expected shape and size of Get3 (a total of 362,614 particles). Several rounds of three-dimensional (3D) heterogeneous refinement were carried out to produce a class of 156,446 particles. 2D templates were generated using this class of particles and these templates were used for template picking. The 1,561,353 picked particles were extracted at a 4× bin, followed by 3D heterogeneous refinement to filter out bad particles and good particles were reextracted at a 2× bin. These 803,265 particles were subjected to a 3D heterogeneous refinement again. Two models had similar levels of detail and shape. The 568,836 particles that result in these two models were reextracted with no binning and underwent a round of 3D heterogeneous refinement again. This resulted in two classes with high resemblance (338,011 particles).

These particles were exported using the cryosparc2star.py program from UCSF pyem package[85] and imported into RELION v.3.1.2 (ref. [86]). Particles underwent a round of 3D homogeneous refinement and local CTF refinement. It became clear that there was weaker density above the NBD, which could suggest lower order partially due to the expected flexible BRIL. Using Segger[87] in Chimera v.1.3 (ref. [88]) we created a mask to isolate the NBD and applied a soft mask of 6 pixels extended by 4 to the particles using particle subtraction. These particles underwent a round of 3D classification into four different classes. One class, 70,330 particles, with the most detail refined to 3.86 Å. C2 symmetry was imposed and the map refined to 3.72 Å. Postprocessing was performed with a soft mask of 6 pixels extended by 4 and the *B* factor was estimated by RELION. Local resolution was estimated using RELION's own implementation. The disordered region could not be refined.

For the *Gi*Get3 apo data set, 9,300 videos were first processed in cryoSPARC v.3.2.0, producing aligned dose-weighted micrographs (Extended Data Fig. 7a). Videos were downsampled to a corrected pixel size of 1.059 Å per pixel and all future processing was done at this pixel size (Extended Data Fig. 7b,c). A subset of 7,607 videos were manually selected for particle picking. A small set of particles were manually picked and used to create 2D templates for template-based picking. Particles picked were filtered to remove debris, resulting in 11,596,225 particles that were then filtered using 2D classification resulting in 552,716 particles. Four ab initio models were generated, and one class was consistent with the expected shape and size of Get3 (174,3012 particles). Several rounds of 3D heterogeneous refinement were carried out to produce a class of 74,013 particles. 2D templates were then generated using these particles and the resulting templates were used for template picking, resulting in 17,238,072 picked particles. These particles were then extracted at 4× bin, followed by 3D heterogeneous refinement to filter out bad particles, and particles belonging to classes that resembled Get3 were then reextracted at 2× bin. These 7,599,636 particles were filtered by 3D heterogeneous refinement and particles in 3D classes that resembled Get3 were again reextracted, this time without any binning. These particles were filtered using several rounds of 3D heterogeneous refinement and the resulting 580,912 particles underwent homogeneous refinement.

These particles were exported using the cryosparc2star.py program and imported into RELION v.3.1.2. Particles underwent several rounds of 3D classification into six different classes. Five classes resembled Get3 and the combined 51,340 particles refined to 8.46 Å. Postprocessing was performed with a soft mask of 6 pixels extended by 4 and the *B* factor was again estimated by RELION. Local resolution was estimated using RELION's own implementation.

Model building into the cryo-EM map. For the *Gi*Get3–TA complexes, using phenix.dock in map two molecules were searched for in the map using the monomer of *Sc*Get3 from PDB 5BW8 as a model[22]. The *G. intestinalis* sequence was then imposed using phenix.sculptor. Manual model building was conducted in COOT and the final model was run through phenix.real space refinement. Poly-alanine sequences were built into the three tubular densities in the CBD (H4/5 and the Bos1 TMD) for residues that could not be ambiguously assigned. Residues making up H4/5 are denoted as UNK the deposited structure.

Apo1 (chain A) from the apo *Gi*Get3 crystal structure was used as a search model in the apo *Gi*Get3 map using phenix.dock in map. Two molecules were found to fit. Residues that did not fit into the density map were deleted. To provide an initial model for FLEX-EM in the CCPEM suite, missing residues were placed in the model as a random loop (residues 85–116) and the resulting model was then refined using FLEX-EM[89]. In the final model, residues that did not fit in the map were deleted (85–116). ChimeraX v.1.2 (ref. [90]) and Chimera v.1.3 were used to visualize maps and models.

ATPase assays. ATPase assays were carried out using EnzChek Phosphate Assay Kit (Thermo Fisher). Assays were carried out with 5.03 µM of *Gi*Get3 or 4.51 µM of *Gi*Get3-D53N in a buffer of 50 mM HEPES, 150 mM potassium acetate, 5 mM magnesium acetate and 10 mM β-ME, pH 7.5 at 37 °C. The reaction mixtures were

incubated in 96-well plates (Corning Costar Assay Plate) at 37 °C before initiating the reaction with ATP at concentrations of 0 µM, 37.25 µM, 62.5 µM, 125 µM, 250 µM, 500 µM, 1 mM and 2 mM. Measurements were taken of three distinct samples for each concentration by a Tecan Infinite M Nano+ at an Abs = 360 nm every 20 sections for a total of 10 min. This method was programmed using Magellan v.7.2 software (Tecan). Data were analyzed using IceKat[91].

In vitro capture assays. The in vitro transfer assays were performed as in previous reports[42,44]. Specifically, 39 µM Bos1 BPA (50 mM HEPES, 300 mM NaCl, 0.05% LDAO, 20% glycerol) was diluted to a final concentration of 0.1 µM and added to 4 µM Ssa1 supplemented with 2 mM ATP (25 mM HEPES pH 7.5, 150 mM KOAc). After 1 min, 0.3 µM of full-length *Gi*Sgt2 or mutant was added to the reaction. Samples were flash frozen after 1 min and placed under a 365-nm ultraviolet lamp for 2 h on dry ice to allow for BPA crosslinking.

Proteins were detected using western blots. Protein samples were run on an SDS–PAGE gel and then transferred onto nitrocellulose membranes by the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked for 1 h in 5% nonfat dry milk and the incubated with antibodies in TTBS buffer for 1 h at 24 °C. The primary antibodies, either an anti-penta-His mouse monoclonal (Qiagen) or anti-Strep II rabbit polyclonal (Abcam), were used at a 1:1,000 and 1:3,000 dilution, respectively. A secondary antibody antimouse or antirabbit (LI-COR Biosciences) conjugated 800 nm fluorophore at a 1:15,000 dilution was used, and the blotting signals were chemically visualized with an infrared scanner.

Visualization of *Gi*Get3 and *Gi*TA clients using western blots. For detection of *Gi*Get3 in the cytosolic or membrane fractions of *G. intestinalis*, the whole cell lysate and cytosolic and membrane pellet were run on a 12.5% SDS–PAGE gel. The gels were blotted onto 0.45-µm nitrocellulose membranes. The blots were blocked for 1 h with a 5% dry milk in TTBS buffer. Then the blots were incubated with anti-Get3 antibody (1:5,000), anti-endolase antibody (1: 10,000), anti-PDI2 antibody (1:10,000) or anti-GL80503_9296 antibody (1:2,000) for 1 h at 25 °C. The anti-endolase antibody was validated by western blot of recombinant proteins and bands corresponded to the same protein band as tagged proteins detected by commercial anti-BAP antibody. Blots were rinsed in PBS and incubated for 1 h with the appropriate secondary antibody (antirat-horseradish peroxidase (HRP) conjugate or antirabbit HRP conjugate (Thermo Fisher), 1:2,000). For immuodetection and visualization Immobilon Classico Western HRP substrate (SigmaAldrich) and Amersham Imager 600 (Vanderbilt School of Medicine) were used.

*Gi*Get3–*Gi*TA protein complexes were run on a 12.5% SDS–PAGE gel. The gels were blotted onto 0.45-µm nitrocellulose membranes that were then cut in half and blocked for 1 h with a 5% dry milk in TTBS buffer. Then the higher molecular weight upper half of the blot was incubated with an anti-Get3 antibody (1:5,000 dilution) while the lower molecular weight half was incubated with an anti-SUMO antibody at a 1:1,000 dilution (Rockland Immunochemicals, Inc.) for putative *Gi*TA proteins at 4 °C for 4 h. Blots were rinsed with TTBS and then incubated with a secondary antibody (antirat or antirabbit conjugated to an IR_{680} fluorophore (LI-COR Biosciences) at a dilution of 1:15,000. The presence of *Gi*Get3 and *Gi*TA proteins were visualized by imaging the blots at a wavelength of 680 nm.

Data availability

Atomic coordinates and structure factors for the apo *Gi*Get3 crystal structure and ATP-bound *Gi*Get3 have been deposited in the PDB under accession codes 7SPZ and 7SPY, respectively. The atomic coordinates and cryo-EM maps for the ADP-bound *Gi*Get3–TA complex have been deposited to the PDB under the accession code 7SQ0 and Electron Microscopy Data Bank (EMDB) under the accession codes EMD-25374 (overall Get3–TA complex) and EMD-25373 (NBD of the Get3–TA complex). The apo *Gi*Get3 cryo-EM map was deposited to the EMDB under the accession code EMD-25375. Source data are provided with this paper.

References

50. Pei, J., Kim, B.-H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
51. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
52. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
53. Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2020).
54. Richter, D. J., Berney, C., Strassert, J. F. H., Burki, F. & de Vargas, C. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotic life. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.30.180687 (2020).
55. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings Bioinform.* **20**, 1160–1166 (2017).
56. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
57. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2020).
58. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
59. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
60. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).
61. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
62. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
63. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
64. Keister, D. B. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans. Royal Soc. Tropical Med. Hyg.* **77**, 487–488 (1983).
65. Doležal, P. et al. Giardia mitosomes and Trichomonad hydrogenosomes share a common mode of protein targeting. *Proc. Natl Acad. Sci. USA* **102**, 10924–10929 (2005).
66. Voleman, L. et al. *Giardia intestinalis* mitosomes undergo synchronized fission but not fusion and are constitutively associated with the endoplasmic reticulum. *BMC Biol.* **15**, 27–27 (2017).
67. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
68. Najdrová, V., Stairs, C. W., Vinopalová, M., Voleman, L. & Doležal, P. The evolution of the puf superfamily of proteins across the tree of eukaryotes. *BMC Biol.* **18**, 77 (2020).
69. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
70. Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
71. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).
72. Goedhart, J. & Luijsterburg, M. S. VolcaNoseR is a web app for creating, exploring, labeling and sharing volcano plots. *Sci. Rep.* **10**, 20560 (2020).
73. Chun, E. et al. Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Cell Struct. Funct.* **20**, 967–976 (2012).
74. Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. *HKL*-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallogr. Sect. D.* **62**, 859–866 (2006).
75. Afonine, P. V. et al. Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. Sect. D.* **68**, 352–367 (2012).
76. Headd, J. J. et al. Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. Sect. D.* **68**, 381–390 (2012).
77. Afonine, P. V., Grosse-Kunstleve, R. W., Urzhumtsev, A. & Adams, P. D. Automatic multiple-zone rigid-body refinement with a large convergence radius. *J. Appl. Crystallogr.* **42**, 607–615 (2009).
78. Bunkóczi, G. & Read, R. J. Improvement of molecular-replacement models with *Sculptor*. *Acta Crystallogr. Sect. D.* **67**, 303–312 (2011).
79. Murshudov, G. N. et al. *REFMAC*5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D.* **67**, 355–367 (2011).
80. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
81. Winn, M. D. et al. Overview of the *CCP*4 suite and current developments. *Acta Crystallogr. Sect. D.* **67**, 235–242 (2011).
82. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
83. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
84. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
85. Asarnow, D., Palovcak, E. & Cheng, Y. asarnow/pyem: UCSF pyem v0.5. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2019).
86. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
87. Pintilie, G., Chen, D. H., Haase-Pettingell, C. A., King, J. A. & Chiu, W. Resolution and probabilistic models of components in cryo-EM maps of mature p22 bacteriophage. *Biophys. J.* **110**, 827–839 (2016).

88. Pettersen, E. F. et al. USCF Chimera – a visualization system of exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
89. Topf, M. et al. Protein structure fitting and refinement guided by cryo-EM density. *Structure* **16**, 295–307 (2008).
90. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
91. Olp, M. D., Kalous, K. S. & Smith, B. C. ICEKAT: an interactive online tool for calculating initial rates from continuous enzyme kinetic traces. *BMC Bioinf.* **21**, 186 (2020).
92. Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
93. Mock, J.-Y., Xu, Y., Ye, Y. & Clemons, W. M. Structural basis for regulation of the nucleo-cytoplasmic distribution of Bag6 by TRC35. *Proc. Natl Acad. Sci. USA* **114**, 11679–11684 (2017).
94. Chartron, J. W., VanderVelde, D. G. & Clemons, W. M. Structures of the Sgt2/SGTA dimerization domain with the Get5/UBL4A UBL domain reveal an interaction that forms a conserved dynamic interface. *Cell Reports* **2**, 1620–1632 (2012).

Acknowledgements

We thank A. Malyutin, S. Chen, H. Scott, G. Lander and J. Kaiser for technical assistance. We thank S.-O.u Shan, R. Voorhees, D. Rees and A. Barlow for discussion and comments. We thank Luboš, Voleman and V. Dohnálek for their help with microscopy and phylogenetic analysis, respectively, V. Mechem for help with protein purification and crystallization, A. Barlow for help with the ATPase assay and Y. Liu for help with the in vitro TA protein capture assay. Crystallography data were collected at the SSRL beamline 12-2. We are grateful to the Gordon and Betty Moore Foundation for support of the Molecular Observatory at the California Institute of Technology. SSRL operations are supported by the US Department of Energy and US National Institutes of Health (NIH). (Cryo)Electron microscopy on the *Gi*Get3–TA protein complex was done in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. A portion of this research was supported by NIH grant no. U24GM129547 and performed at the Pacific Northwest Center for Cryo-EM at Oregon Health & Science University and accessed through EMSL (grid.436923.9), a Department of Energy Office of Science User Facility sponsored by the Office of Biological and Environmental Research. Some computing resources were provided by the Extreme Science and Engineering Discovery Environment (XSEDE) resources, which is supported by the National Science Foundation grant no. Acl-1052574(108). Work in the United States was supported by the NIH grant nos. R01GM097572, R01GM125063 and DP1GM105385 to W.M.C. Work in the Czech Republic was supported by the Czech Science Foundation grant no. 20-25417S, a grant from Charles University Grant Agency (project no. 1396217) and the project 'Centre for research of pathogenicity and virulence of parasites' (grant no. CZ.02.1.01/0.0/0.0/16 019/0000759) funded by the European Regional Development Fund. S.M.S. was supported by National Science Foundation Graduate Research fellowship under grant no. 11444469. S.M.S. and M.Y.F. were supported by a NIH/National Research Service Award Training grant no. T32GM07616.

Author contributions

W.M.C., P.D. and M.Y.F. conceived the study. V.N. identified the Get3 homolog in *Giardia* and performed the verification using biochemical and cell biology methods. V.N. and S.M.S. identified other GET pathway components in *Giardia* and performed the phyologenetic analyses throughout eukaryotes. M.Y.F. performed the biochemistry to demonstrate *Gi*Sgt2 capture of TA proteins and identification of substrates in *G. intestinalis*. M.Y.F. prepared samples for structural studies, crystallized apo and ATP-bound *Gi*Get3, and performed the all cryo-EM analyses and processing. M.Y.F., A.O.M. and W.M.C. refined the crystallography data and built the atomic model into the electron density. M.Y.F., A.O.M. and W.M.C. built the model into the *Gi*Get3–TA complex cryo-EM map and M.Y.F. and W.M.C conducted the model building for the apo *Gi*Get3 cryo-EM map. M.Y.F. and W.M.C. wrote the manuscript with input from all authors. All authors edited and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

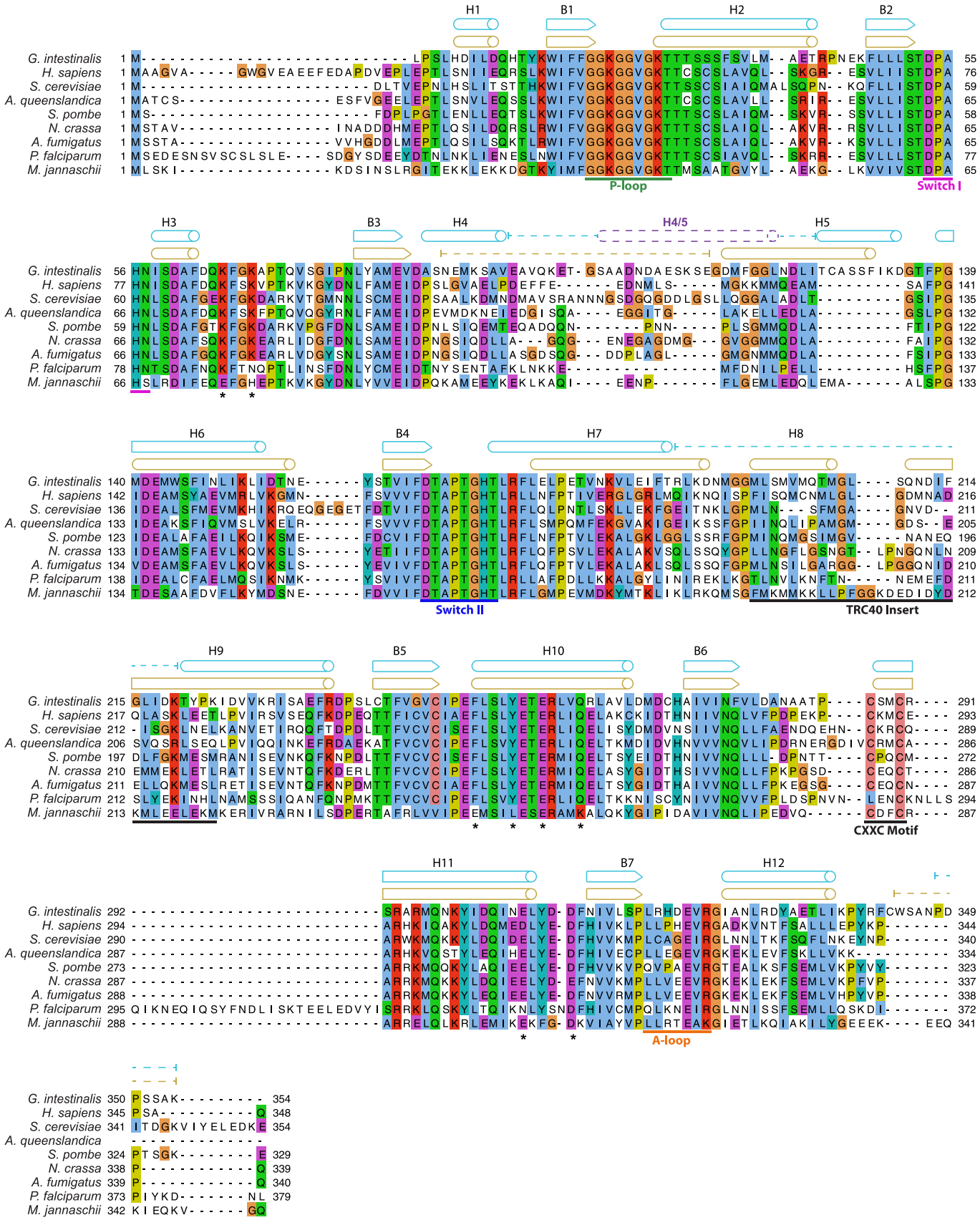
Extended data are available for this paper at https://doi.org/10.1038/s41594-022-00798-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-022-00798-4.

Correspondence and requests for materials should be addressed to William M. Clemons.

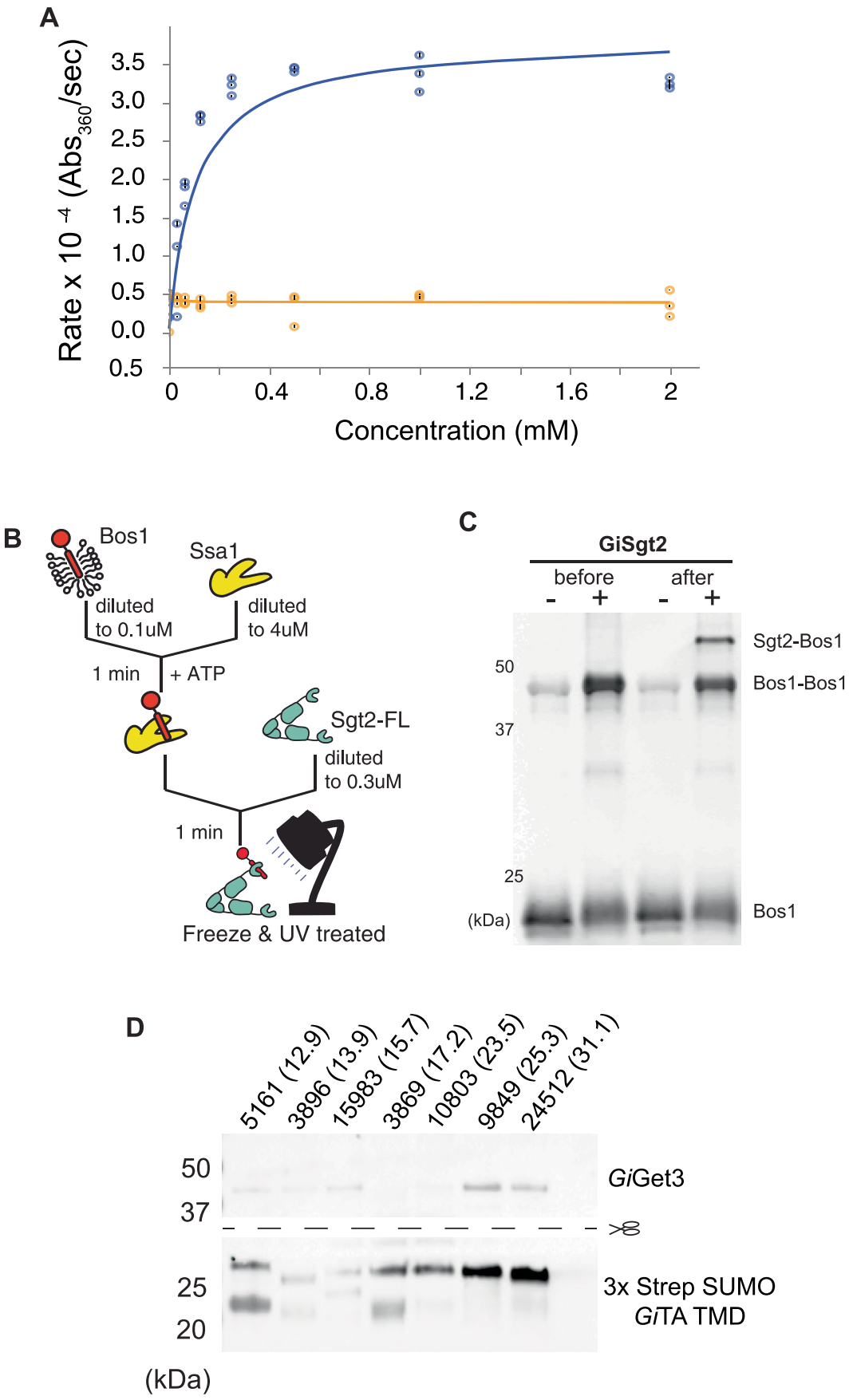
Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Florian Ullrich, in collaboration with the *Nature Structural & Molecular Biology* team.

Reprints and permissions information is available at www.nature.com/reprints.



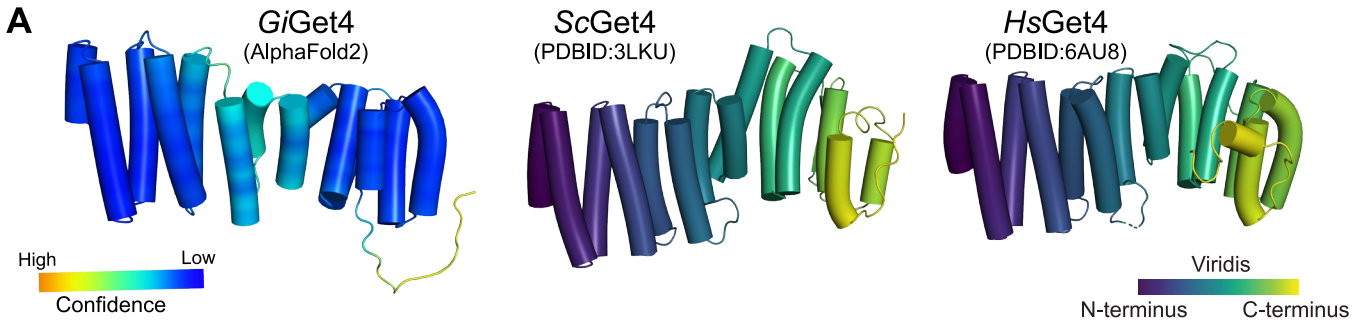
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Alignment of Get3. The full alignment of Get3 homologs partially shown in Fig. 1A. Conserved Get3 features are marked by bars below the alignment: P-loop (*green*), Switch I (*magenta*), Switch II (*blue*), A-loop (*orange*), the TRC40 insert (*black*), and the CXXC motif (*black*). Secondary structure for the Get3/TA•ATP complex (*cyan*) and $Get3_{D53N}$ •ATP (*orange*) are depicted above the alignment – cylinders for α-helices, arrows for β-sheets, and dashed lines for residues that were disordered. The region that contains H4/5 highlighted in purple. Conserved residues that were demonstrated to play a role in Get4 binding have asterisks below the alignment. Residues are colored using the ClustalX color scheme[46].

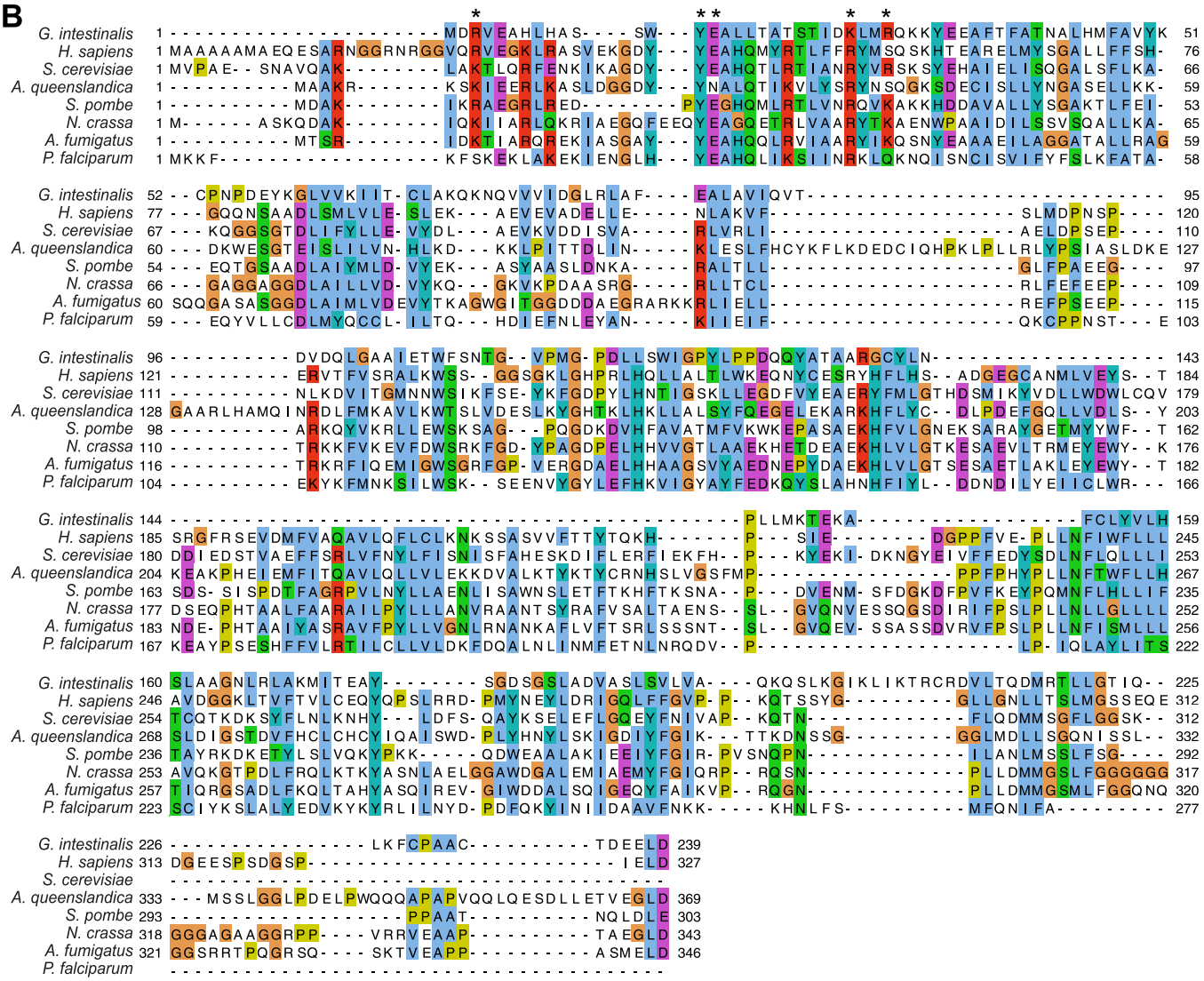


Extended Data Fig. 2 | See next page for caption.

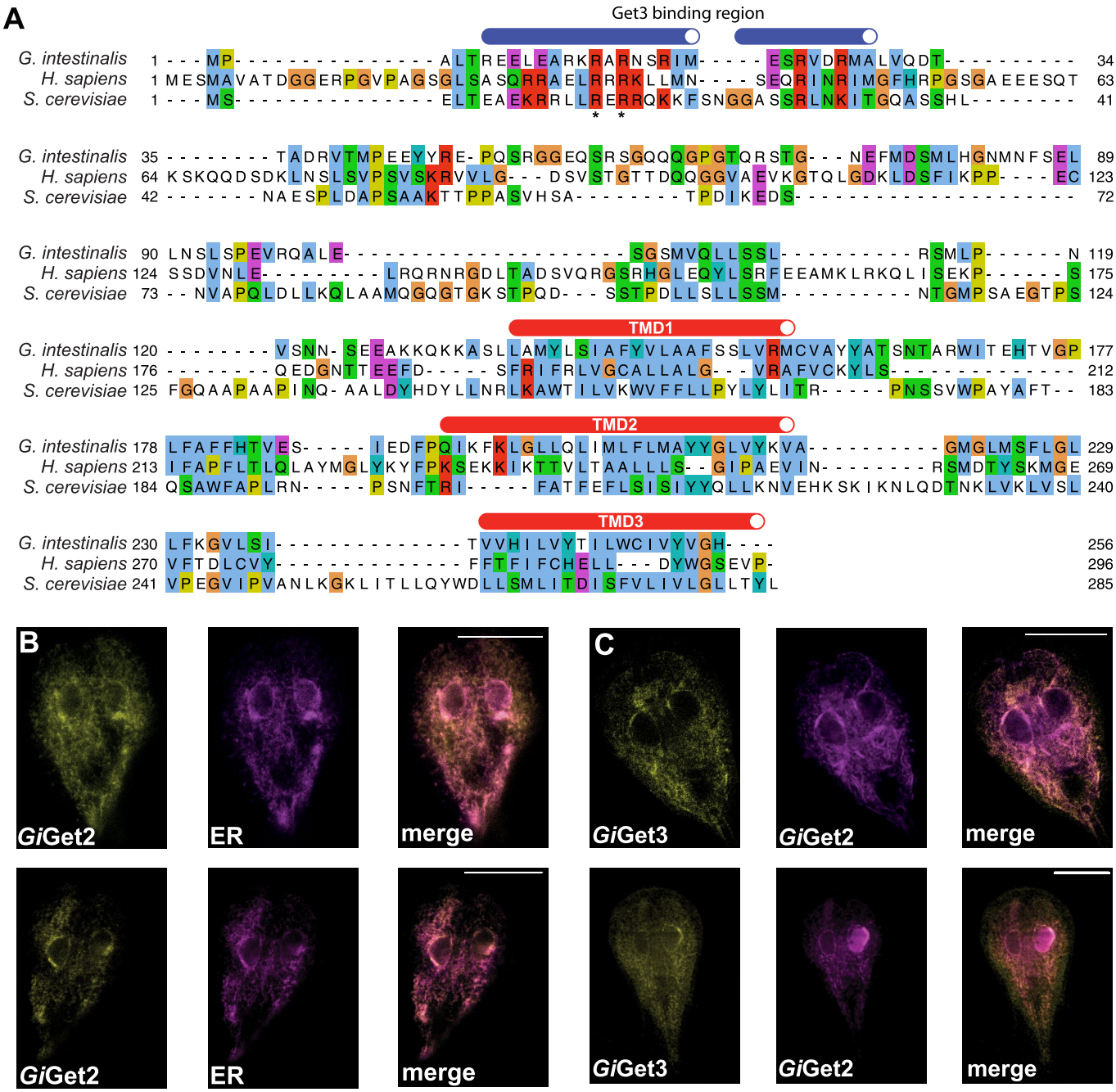
Extended Data Fig. 2 | Validation of *G. intestinalis* GET components. *A*) ATPase assays with *Gi*Get3 (*blue*) & *Gi*Get3$_{D53N}$ (*orange*). Absorbance was measured at a wavelength of 360nm at varying ATP concentrations. Experiment was done in triplicate. The lines represent the data fit using a LOESS (LOcally Estimated Scatterplot Smoothing) regression[92]. *B*) A schematic of experimental set-up which was conducted twice. *Sc*Ssa1 transfers *Sc*Bos1 with an engineered BPA crosslinking site to *Gi*Sgt2. Crosslinking is initiated by UV exposure. *C*) A western blot visualizing the crosslinked *Gi*Sgt2/*Sc*Bos1 complexes before and after transfer and with and without UV light treatment. *C*) *Gi*Get3 and putative *Gi*TA proteins were recombinantly expressed in *E. coli* then purified 2-3 times (depending on putative client) utilizing an affinity column to the Strep-tag on the TA proteins. Get3 could only be visualized if the TA protein expressed and bound to the TA protein. For western analysis, eluted samples were run on a gel and transferred to a membrane. The membrane was split in half, marked by dotted line, with the top blotted with an anti-*Gi*Get3 antibody and the bottom by an anti-SUMO antibody. Putative TA proteins tested (*GL50803_005161*, *GL50803_003896*, *GL50803_0015983*, *GL50803_003869*, *GL50803_0010803*, *GL50803_009849* and *GL50803_0024512*) are arranged by increasing TMD hydrophobicity (labeled in parentheses) using the TM tendency scale. *Gi*Get3 was clearly identified in the eluates of *GL50803_0024512* and *GL50803_009849*.



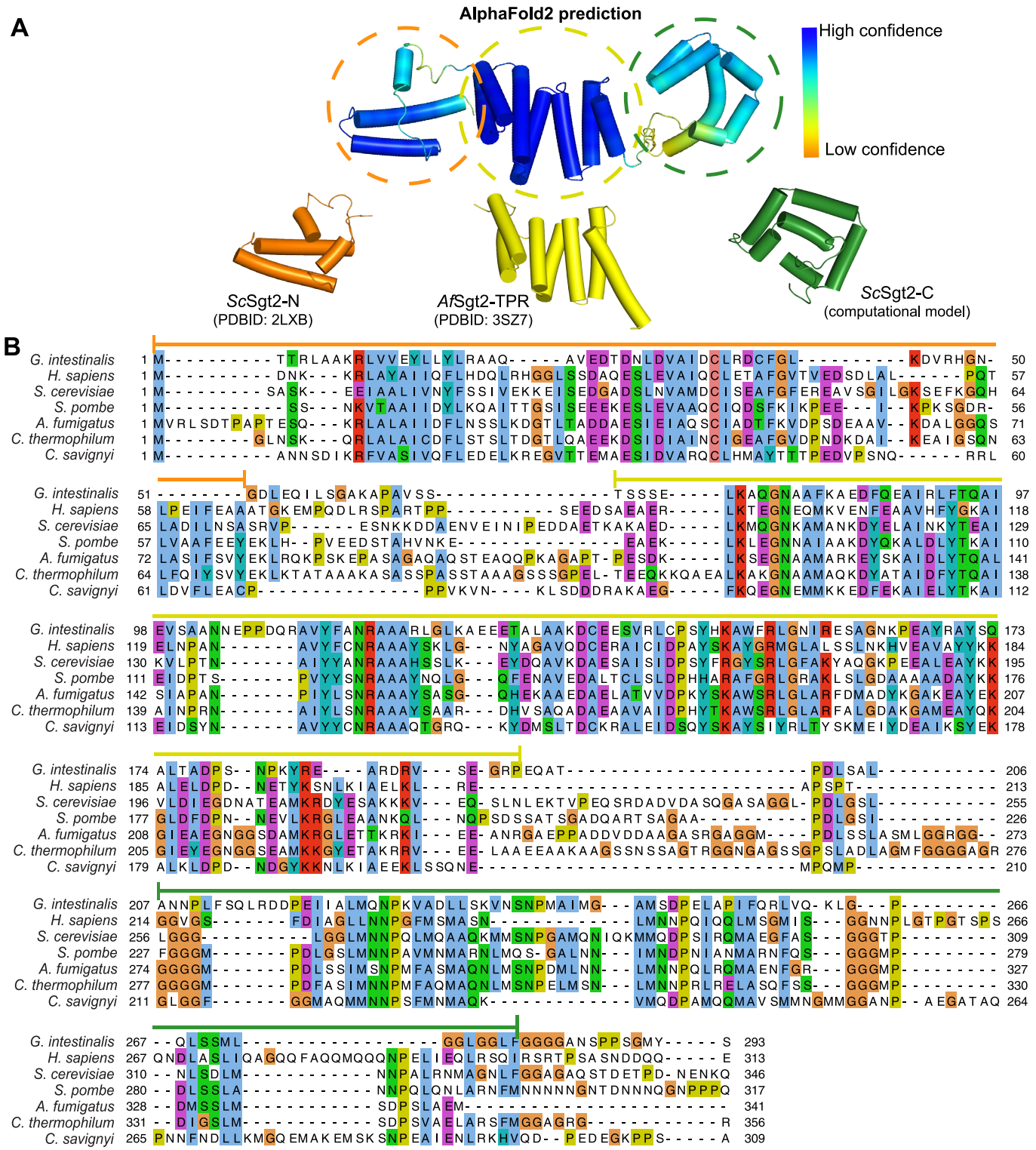
B



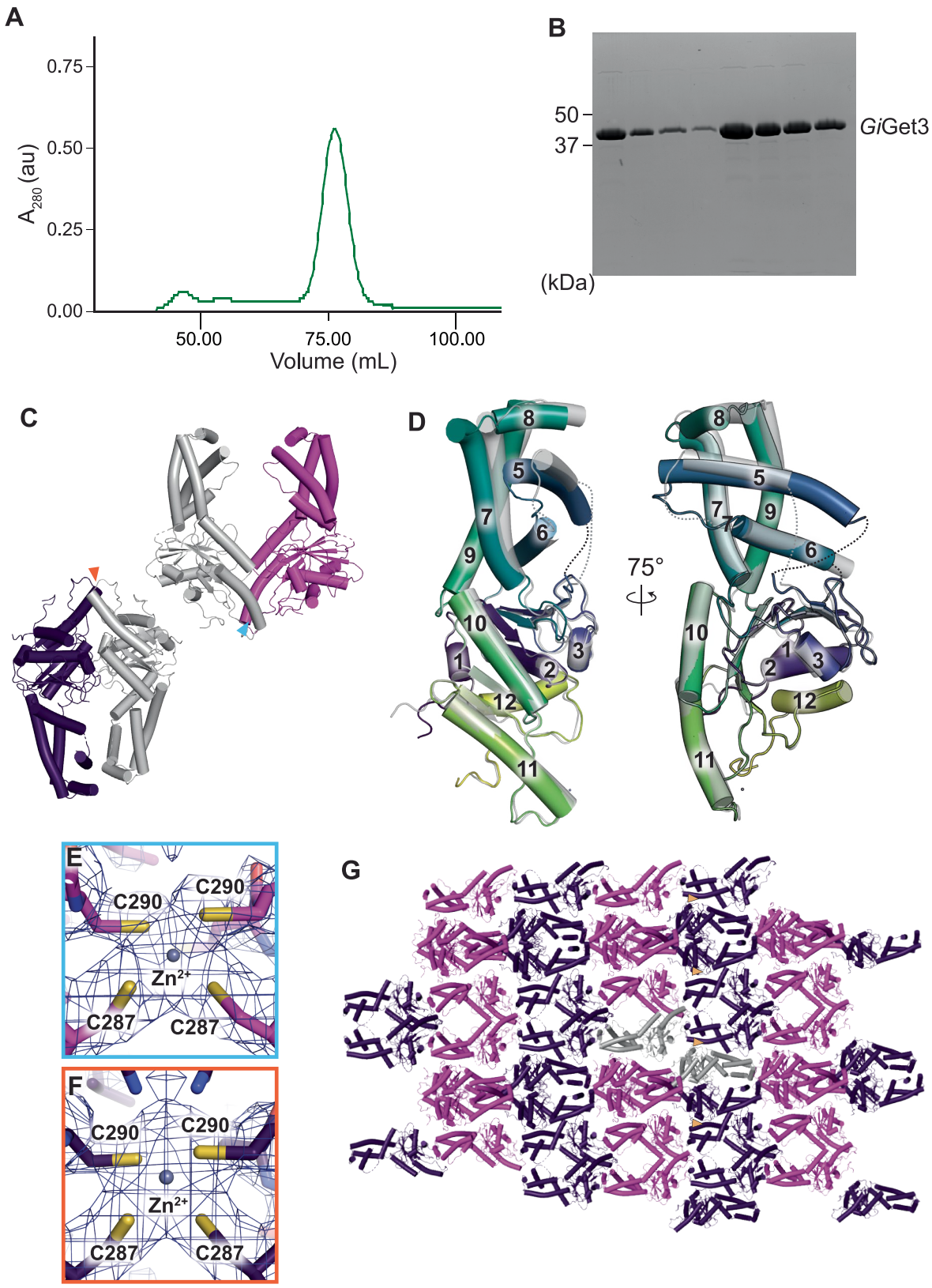
Extended Data Fig. 3 | Alignment of Get4 homologs. *A*) (*Left*) AlphaFold2 structural prediction of the identified *Gi*Get4 colored based on confidence from blue (most confident) to orange (least confident). Experimentally determined structure of the yeast Get4 (*center*)[14] and human Get4 (PDBID: 6AU8) (*right*)[93] are colored from N- to C-terminus using the viridis color scheme (*purple to yellow*). *B*) Alignment of Get4 homologs including: *G. intestinalis*, *H. sapiens*, *S. cerevisiae*, *A. queenslandica*, *S. pombe*, *N. crassa*, *A. fumigatus*, and *P. falciparum*. Conserved residues that were demonstrated to play a role in Get3 binding are highlighted with asterisks above. Residues are colored as in Extended Data Fig. 1[22].



Extended Data Fig. 4 | Identification of a Get2 homolog in *G. intestinalis*. *A*) An alignment of the identified Get2 from *G. intestinalis* with CAML from *H. sapiens* and Get2 from *S. cerevisiae*. The three predicted TMDs are shown with red cylinders above the alignment and the conserved N-terminal helices that tether to Get3 are shown as blue cylinders. The two conserved residues involved in Get3 binding are marked by asterisks below the alignment. Residues are colored using the ClustalX color scheme. Representative images from three experiments for (*B*) detection of *Gi*Get2 using Stimulated Emission Depletion (STED) microscopy images of *G. intestinalis* trophozites. Get2 with a C-terminal V5 tag was detected by an anti-V5 antibody (*yellow*) and the ER membrane was labeled by an anti-PDI2 antibody (*magenta*). Images are merged in the third panel. (*C*) The ER localization of *Gi*Get2 differs from the cellular localization of *Gi*Get3. *Gi*Get3 with a N-terminal HA-tag was labeled by an anti-HA antibody (*yellow*) and *Gi*Get2 is labeled by the same method in (*A*) (*magenta*). Images are merged in the third panel. Scale bars represent 5μm.

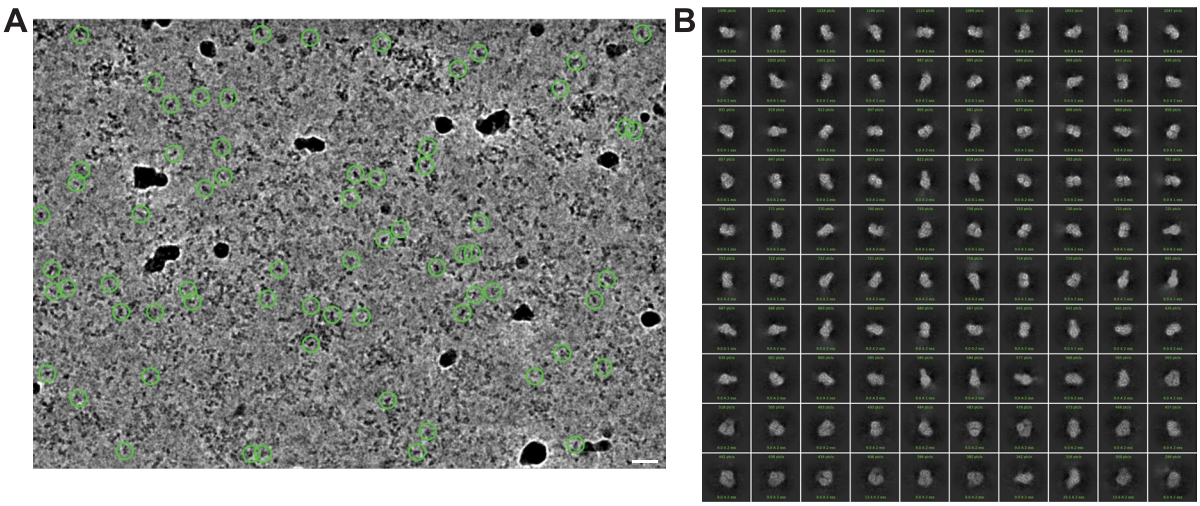


Extended Data Fig. 5 | Structures and alignment of Sgt2. *A*) AlphaFold2 structural prediction of the identified *Gi*Sgt2 colored based on confidence as in Extended Data Fig. 3A. Structural models of fungal Sgt2 domains: the N-terminal domain from *S. cerevisiae* (*orange*) (PDBID:2LXB)[94], the TPR domain from *A. fumigatus* (*yellow*) (PDBID:3SZ7)[12] and the C-domain from *S. cerevisiae* (*green*)[35]. *B*) An alignment of Sgt2 homologs: *G. intestinalis*, *H. sapiens*, *S. cerevisiae*, *S. pombe*, *A. fumigatus*, *C. thermophilum*, and *C. savignyi*. The three domains are indicated by lines above the alignment, N-terminal dimerization (*yellow*), TPR-domain (*orange*), and substrate binding C-domain (*green*). Residues are colored using the ClustalX color scheme.



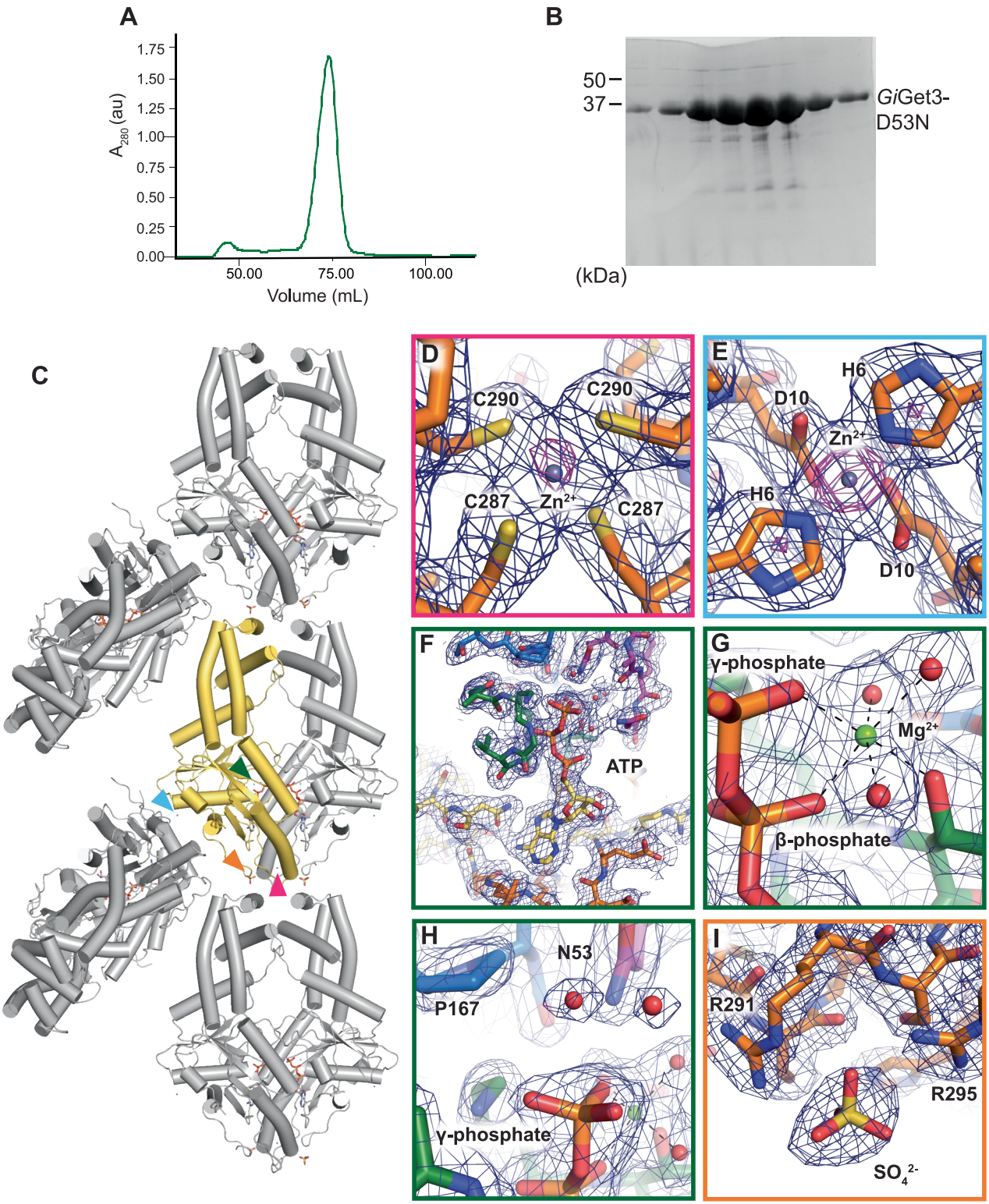
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | The crystal structure of apo *Gi*Get3. *A*) Size exclusion chromatograms of nickel eluate for Get3 and (*B*) SDS-PAGE gels of the peak highlighted in (*A*). *Gi*Get3 was purified 10-20 separate times in order to produce enough material for crystal trays and cryo-EM analysis. *B*) The asymmetric unit of apo *Gi*Get3 contains two distinct monomers of Get3 (*gray*), which are designated apo1 and apo2. Each monomer pairs with its respective symmetry equivalent in the neighboring asymmetric unit to form two distinct dimers. One dimer, apo1 (*magenta*), has the 'open' conformation. The other dimer, apo2 (*purple*), has a conformation that is slightly more closed than the previously seen 'open' conformations. *C*) Alignment of the two monomers highlighting additional slight differences between. Apo1 is colored from N to C terminus using the viridis color map and apo2 in grey. The monomers are aligned to the P-loops and there are no notable differences in the NBD. The flexibility of the CBD is demonstrated here. Most distinct is a shift in the loop between H5 and H6 (residues 121-139) with most of this region disordered in apo2. Additionally, H5 is shorter at the N-terminus in apo2 compared to apo1 due to a close contact with a symmetry mate that presumably disrupts the helix (*red arrows in F*). A 2Fo-Fc map shown as blue mesh at 1.0 sigma showing the region that includes the Zn^{2+} ions on the symmetry axes for (*D*) apo1 (outlined in *cyan*) and (*E*) apo2 (outlined in *orange*); these positions correspond to the arrows in (*B*). *F*) A single layer of the crystal lattice for apo *Gi*Get3 crystals. Grey monomers define the asymmetric unit and symmetry related dimers are colored for either apo1 (*magenta*) or apo2 (*purple*).



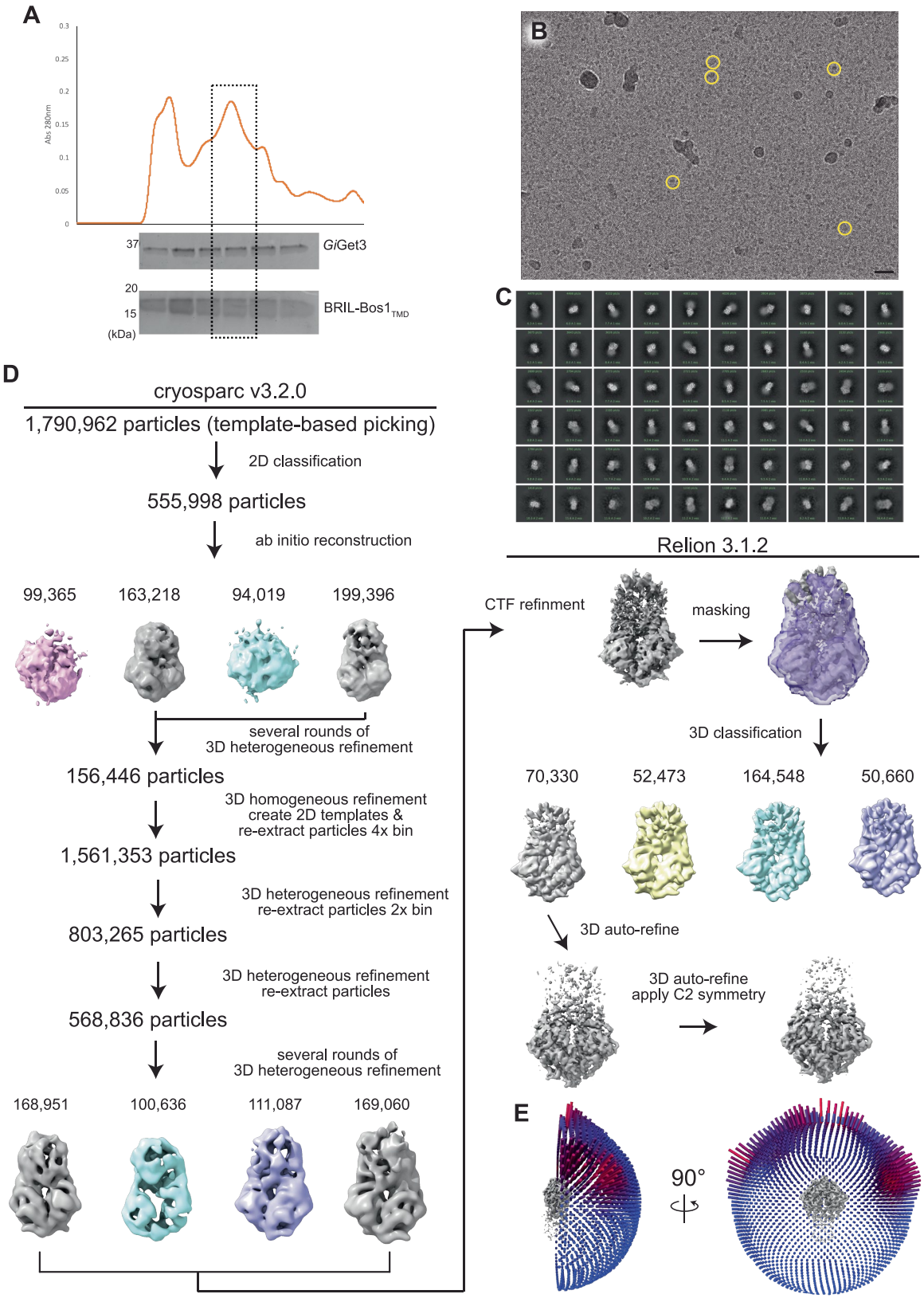
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | SPA data processing of apo *Gi*Get3. *A)* An aligned and dose weighted micrograph from the data collection (7,607 micrographs) with apo *Gi*Get3 sample particles selected with green circles. Scale bar represents 50nm. *B)* 2D class averages of particles used for the reconstruction. *C)* Processing of data through cryoSPARCv3.2.0 and RELION 3.1.2. *D)* Two views of the angular distribution of particles. Particle concentration is displayed by color and length (*blue to red*). Density filtered to 7 sigma for various helices that are shown as sticks: *E)* H7 (*cyan*) & H9 (*magenta*) and *F)* H10 (*magenta*) & H11 (*cyan*).



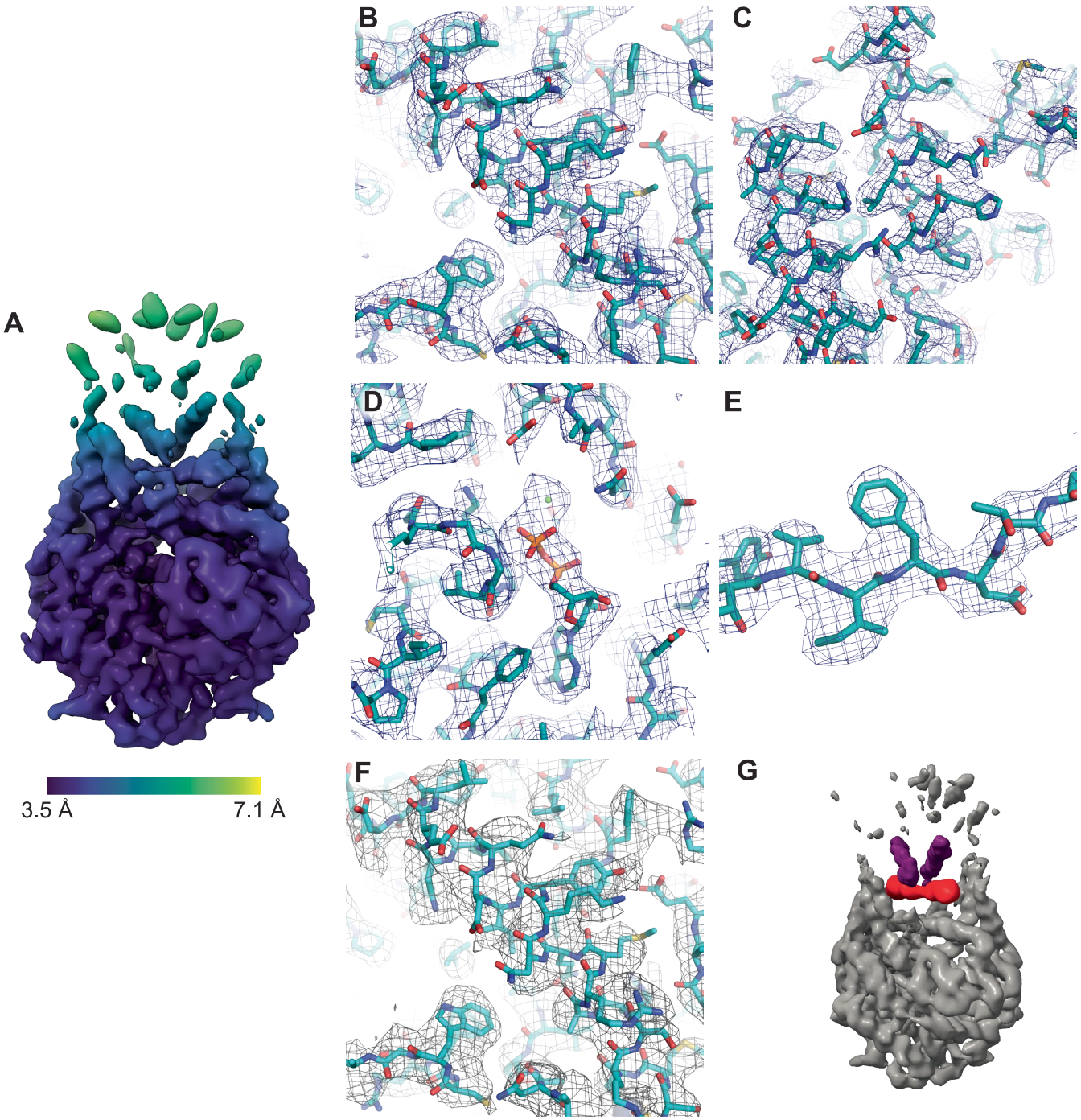
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | The crystal structure of *Gi*Get3$_{D53N}$ •ATP. *A*) Size exclusion chromatograms of nickel eluate for Get3$_{D53N}$ and (*B*) a SDS-PAGE gel of the peak highlighted in (*A*). *Gi*Get3$_{D53N}$ was purified 10-20 separate times in order to produce enough material for crystal trays. *C)* A cartoon model of the *Gi*Get3$_{D53N}$ •ATP crystals. The asymmetric unit is shown in (*yellow*) with symmetry related molecules in *grey*. Regions of interest in the crystals are shown with arrows. Corresponding close-up views are shown in outlined panels boxed corresponding to the arrows in (*C*) each shown in sticks colored as before with $2F_o$-F_c density contoured at 1.5 σ and colored blue in *C-I*. *D*) The bridging Zn^{2+} ion coordinated by four cysteines, two from each monomer (Cys_{287} & Cys_{290}, *magenta arrow*). *E*) Similar representation as in *B* for the Zn^{2+} ion coordinated via a crystal contact by the surface residues His_6, Asp_{10} and the same residues in a symmetry related molecule. In *C* & *D* the identity of the Zn^{2+} ions are confirmed by positive density (*magenta mesh*) in an anomalous double difference map obtained from data collected at two energies, 9.669keV and 9.659keV, which are just above and below the absorption K-edge of zinc. The anomalous double difference map is contoured at 6 sigma. *F-H*) Three views of the active site. *F*) The ATP molecule, surrounding loops, and coordinated waters within the active site. *G*) The residues and waters coordinating the Mg^{2+} ion. *H*) In the active site a Pro_{167} and Asn_{53} orient the catalytic water molecule above the γ-phosphate in the ATP molecule. *I*) A SO_4^{2-} ion from the buffer bound to two arginines (Arg_{291} & Arg_{295}).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | SPA data processing of *Gi*Get3-TA complexes. *A)* Size exclusion chromatogram of nickel eluate of Get3-TA complexes and an SDS-page gel of the peak from one of the 10 separate purifications of the complex. A single fraction was used for structure determination *B)* An aligned micrograph from the data collection (2,732 micrographs) with sample particles selected with yellow circles. Scale bar represents 50nm. *C)* 2D class averages of particles used for the reconstruction. *D)* Processing of data through cryoSPARCv3.2.0 and RELION 3.1.2. *E)* Two views of the angular distribution of particles. Particle concentration is displayed by color and length (*blue to red*).



Extended Data Fig. 10 | Representative resolution and density of *Gi*Get3-TA complex. *A*) The local resolution of the unsharpened map presented in Extended Data Fig. 9. Representative density of (*B*) H11, (*C*) H7 & H9, (*D*) active site with ADP molecule and Mg^{2+} ion, and (*E*) β-sheet 1. *F*) Representative density (*grey*) of the reconstruction without any symmetry imposed of H11 as shown in (*B*). *G*) The unsharpened map of the reconstruction before imposing symmetry. Density corresponding to the TA protein and H4/5 are colored as they are in Fig. 3A,B.

Highly diverged pre-targeting complex in *Giardia intestinalis* reveals the ancestral presence of Bag6 and the evolution of GET pathway in eukaryotes

Najdrová Vladimíra [1], Dohnálek Vít [1], Voleman Luboš[1], Doležal Pavel [1]

[1]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 252 50, Vestec, Czech Republic

ABSTRACT

Approximately one-fourth of proteins in cells represent integral membrane proteins. Tail-anchored proteins carry a single C-terminal transmembrane domain that anchor them to organelle membranes. This topology enables TA proteins to mediate interaction among the compartments in processes such as vesicular transport, apoptosis and protein translocation. TA proteins are targeted post-translationally to the ER membrane by the Guided Entry of TA proteins (GET) pathway, which is well studied in metazoan and yeast. Some of GET pathway components were identified in plants and recently in *Plasmodium falciparum* and *Giardia intestinalis*. Here, we identified all missing components of the GET pathway in *G. intestinalis*, including divergent Bag6 homolog, which was considered to be present only in metazoan until now. Moreover, the interactome of *Gi*GET components revealed the involvement of GET proteins in the other cellular pathways such as protein degradation. Our results from both experimental and bioinformatical studies suggest that the GET pathway is an ancestral eukaryotic pathway and has ancient and crucial role in cellular function.

Keywords: *Giardia intestinalis*, protozoan, tail-anchored protein targeting, Guided Entry of tail-anchored protein pathway, native isolation, evolution

INTRODUCTION

Tail-anchored (TA) proteins represent a special class of integral membrane proteins characterized by the presence of a single C-terminal transmembrane domain (TMD) and the absence of any N-terminal signal sequence (Kutay *et al.*, 1993). Targeting information of TA proteins is contained within the TMD and hereby hidden within the ribosome until the end of translation. Therefore, TA proteins must be targeted and inserted into the membrane post-translationally and without SRP participation (Guna *et* Hegde, 2018; Jiang, 2021). Most of the TA proteins are targeted to the endoplasmic reticulum (ER) membrane through the Guided Entry of TA proteins (GET) insertion pathway (named as TRC in metazoan) (Schuldiner *et al.*, 2008; Stefanovic *et* Hegde, 2007).

After translation, TA proteins are first bound by the ribosome-associated chaperone Sgt2 (small glutamine-rich tetratricopeptide repeat-containing protein 2) and this process is facilitated by cytosolic Hsp70 (Cho *et* Shan, 2018; F. Wang *et al.*, 2010). Subsequently, Sgt2 recruits Get5 that together with Get4 builds a heterotetrameric Get4/Get5 complex. The entire so-called pre-targeting (or cytosolic) complex then passes TA protein to Get3 ATPase (Chartron *et al.*, 2010; Chio *et al.*, 2017; Gristick *et al.*, 2015). Get3 with the bounded TA protein is released from the pre-targeting complex and delivered to the membrane complex containing Get1 and Get2 proteins. These components mediate the dissociation of TA protein from Get3 and further facilitate the insertion of TA protein into ER membrane (Aviram *et* Schuldiner, 2017; Borgese *et al.*, 2019; Hegde *et* Keenan, 2011; Mariappan *et al.*, 2011; McDowell *et al.*, 2020). The metazoan pre-targeting complex also involves Bag6 (also named Bat3 or Scythe) that combines together with Get4 and Get5 into a heterotrimeric complex (Leznicki *et al.*, 2010; Mariappan *et al.*, 2010; Mock *et al.*, 2015; Vilardi *et al.*, 2011; Yamamoto *et* Sakisaka, 2012). Get4 and Get5 bind separately to Bag6 and this is reflected by the absence of domains that mediate the interaction of the yeast counterparts (Bozkurt *et al.*, 2010; Chang *et al.*, 2010; Chartron *et al.*, 2010; Mariappan *et al.*, 2010; Mock *et al.*, 2015).

Bag6 itself is a multifaceted protein containing N- terminal ubiquitin-like domain (UBL), central proline-rich domain, bipartite nuclear localization signal (NLS), and the C-terminal Bag6 domain that is responsible for Get4 and Get5 binding (Mock *et al.*, 2015). In addition to Bag6 role in the targeting of TA proteins, it participates in other diverse

cellular processes, regulation of gene expression and apoptosis (Minami *et al.*, 2007; P. Nguyen *et al.*, 2008; Thress *et al.*, 1998). Moreover, UBL domain of Bag6 interacts with the Rpn10 proteasome subunit and together with the FBXO7-SCF E3 ubiquitin ligase is required for proper proteasome function. The Bag6 defect leads to various human diseases, including Parkinson's disease (Kikukawa *et al.*, 2005; Q. Wang *et al.*, 2021). Therefore, depending on Bag6, the TA proteins are successfully transferred from SgtA to Get3 (Mariappan *et al.*, 2010) or polyubiquitinated by the E3 ubiquitin ligase RNF126 and degraded in the proteasome (Rodrigo-Brenni *et al.*, 2014; Shao *et al.*, 2017).

Previous work showed that Get3, which is the most conserved component of the GET pathway, undergoes a catalytic cycle of protein binding and release in manner conserved from metazoans (Bozkurt *et al.*, 2009; Gristick *et al.*, 2014; Mateja *et al.*, 2009, 2015; Stefer *et al.*, 2011; Suloway *et al.*, 2009) to unicellular eukaryotes such as *G intestinalis* (Fry *et al.*, 2022). Given the significant functional differences between the pre-targeting complexes of yeast and metazoan, both members of Opisthokonta supergroup of eukaryotes, we decided to characterize the complex in *G. intestinalis* from very distant Metamonada supergroup. Via the characterization of the *in situ* interactome, cell localization and *in vitro* protein interaction assays, we could characterize *G. intestinalis* pre-targeting complex that involves *Gi*Sgt2, *Gi*Get4, *Gi*Get5 and *Gi*Bag6 quartet. We used comparative genomics to define the ancestral pre-targeting complex in the last eukaryotic common ancestor (LECA) and the evolution of the GET pathway in eukaryotes.

RESULTS

Identification of the missing pre-targeting components in *G. intestinalis*

Previous characterization of *G. intestinalis* Get3 revealed the presence of additional three components of the GET pathway (Fry *et al.*, 2022). *Gi*Get4 (GL50803_00112893) and *Gi*Get2 (GL50803_0017617) were found among *Gi*Get3-binding partners, while *Gi*Sgt2 was identified (GL50803_007287) via structure-based sequence search (Fry *et al.*, 2022). The latter was further showed to capture client TA proteins from Hsp70. These data prompted us to identify the missing components of the pre-targeting complex in *G. intestinalis,* that could either diverge beyond the homology-detection

algorithms or represent a lineage-specific components. To this aim, both known pre-targeting complex components *Gi*Sgt2 and *Gi*Get4 were separately expressed in *G. intestinalis* together with biotin ligase (BirA) and the crude cell lysates purified on streptavidin-coupled magnetic beads (Figure S1). The eluates were analyzed by mass spectrometry to identify *Gi*Sgt2 and *Gi*Get4 binding partners.

In total, 51 significantly enriched proteins were co-purified with *Gi*Sgt2 (Fig.1A; Table S1), of which *Gi*Sgt2 itself was the second most enriched protein (394-fold enrichment, FE) while a ubiquitin-like domain-containing protein (GL50803_007287) was found as the most abundant (1,197 FE). Of the known GET pathway components, the dataset contained highly enriched *Gi*Get4 and *Gi*Get3, and also cytosolic chaperone Hsp70 (Ssa1; GL50803_0088765). The rest included DnaJ subfamily A protein (Ydj1; GL50803_009808), heat shock factor binding protein (GL50803_007351), Rac/Rho-like protein (GL50803_008496), and 14 proteins with yet unknown function or domain homology. Interestingly, ten proteasome subunits of the 19S regulatory particle (RP) (Figure 2), ubiquitin (GL50803_007110) and ubiquitin-conjugating enzyme E2 (GL50803_005921) were also identified, suggesting specific co-purification of the proteasome complex.

Isolation of *Gi*Get4 returned a dataset with 67 significantly enriched hits, among which GL50803_0019378 was the top hit (11,000 FE) followed by the bait protein *Gi*Get4 and also GL50803_005069 (Fig. 1B; Table S1). Both *Gi*Get3 and *Gi*Sgt2 were also found among the significantly enriched proteins. Similarly to *Gi*Sgt2 data, a set of proteasome subunits of both the 19S RP and the 20S catalytic particle (CP) (Figure 2), ubiquitin and ubiquitin carboxyl-terminal hydrolase 14 were identified. The dataset contained additional 14 proteins of unknown function. The comparison of two experiments using hierarchical clustering showed a common co-purification of 19 proteins including known GET pathway components (Fig. 1C).

Two proteins GL50803_0019378 and GL50803_005069 of unknown function, which were placed on top of both datasets, were selected for further analyses. GL50803_0019378 represents a protein of theoretical Mw 16.74 kDa but the bioinformatic tools including highly sensitive HHpred algorithm failed to identify any homologous domains within the amino acid sequence. The structural prediction by

Alphafold2 did not result in a high confidence model, as most of the polypeptide chain remained unstructured (Fig. 3A). Only the C-terminal part of 56 (91-147) residues was predicted to fold into three helices of mutual conformation similar to the C-terminal part of yeast and human Get5 (Fig. 3B). Specific comparison with the human Get5 showed that the proteins are similar in size and share only 20,4% similarity (Fig. 3C), but it is comparable to 20.8 % similarity between yeast and human Get5 orthologs. This data indicated that GL50803_0019378 could correspond to highly diverged Get5 orthologue in *G. intestinalis*. The identified protein was hereafter referred to as *Gi*Get5 (Fig. 3C).

In case of GL50803_005069, the homology detection Hhpred algorithm showed similarity to UBL on the N-terminus of the protein (Fig. 4A). No further sequence or structural similarity has been identified but the presence of UBL on the N-terminus of the protein is characteristic feature of human Bag6 protein (Banerji *et al.*, 1990; Mock *et al.*, 2015) . Detailed comparison of the human and *G. intestinalis* proteins showed that their UBL domains are very similar when modelled by Alphafold2 (Fig. 4B) and additional similarity could be identified at the C-terminal region (Fig. 4C). In human Bag6, this region comprises Get4 binding part, the nuclear localization signal (NLS) and Bag6 domain (Mock *et al.*, 2015). The inspection of *G. intestinalis* sequence within the protein sequence alignment indicated the presence of all three components (Fig. 4C). This data suggested that GL50803_005069 represents highly divergent Bag6 homolog in *G. intestinalis*, hereafter labelled as *Gi*Bag6. Using the immunofluorescence microscopy, the BAP-tagged *Gi*Bag6 was found in spotted pattern across the cytosol and approximately 10 % of the cells showed nuclear localization (Figure S2). All cells were found to accumulate the protein in the posterior end of the cell (Fig. 4D).

Get5 and Bag6 interactomes show stable *in vivo* interaction within the pre-targeting complex.

In order to support the involvement of these newly identified highly diverged components in the *G. intestinalis* GET pathway, analogous native isolations of GiGet5 and *Gi*Bag6 were performed (Figure S1). The purification of *Gi*Get5 returned 16 significantly enriched proteins (Fig. 5A; Table S1). Bellow *Gi*Get5 that was found as the top enriched protein (5,800 FE) other GET pathway components *Gi*Get4, *Gi*Bag6, *Gi*Sgt2, and *Gi*Get3

were identified. Other hits included two variant surface proteins (VSP; GL50803_00113439; GL50803_00113450), eukaryotic translation initiation factor 2 (GL50803_0091398), DNA mismatch repair protein (GL50803_0034058), DNA-directed RNA polymerases I and III (GL50803_0010840), A49-like RNA polymerase I associated factor (GL50803_0016615), DUF4485 domain-containing protein (GL50803_005183), and four proteins of unknown function (GL50803_005324; GL50803_0086815; 32668; GL50803_004259).

The isolation of *Gi*Bag6 returned dataset of 14 significantly enriched proteins with *Gi*Bag6 as the most enriched protein (40 FE), followed by *Gi*Get4 and *Gi*Get5 (Fig 5B; Table S1). Additional enriched proteins involved four proteasomal 19S subunit proteins (GL50803_0033166; GL50803_0015454; GL50803_0086683; GL50803_0016659), putative dihydrouridine synthase (GL50803_003565), DNA-directed RNA polymerase subunit RPB10 (GL50803_0014413), ribosomal protein L28e (GL50803_0034093), lysyl-tRNA synthetase (GL50803_0016766), and three proteins of unknown function (GL50803_009605; GL50803_0014997; GL50803_0014984). These reciprocal co-purification data showed that both *Gi*Get5 and *Gi*Bag6 are specific components of *G. intestinalis* GET pathway.

The specificity of the interaction among the GET components was further demonstrated via hierarchical clustering built upon the datasets derived from all four pre-targeting complex components (Fig 5 C). Of 115 identified hits single strongly supported cluster comprising already known components of pre-targeting complex, *Gi*Sgt2 and *Gi*Get4, their newly identified partners *Gi*Get5 and *Gi*Bag6 as well as the TA-protein targeting component *Gi*Get3 (Fig. 5D). Interestingly, the cluster included additional protein of unknown function (GL50803_005324) without any similarity to known domains. The protein thus may represent *G. intestinalis* specific GET pathway component.

Evolution of the GET pathway in eukaryotes.

The identification of diverged GET pathway components in *G. intestinalis* was only possible via the isolation of relatively well evolutionarily conserved proteins Sgt2 and Get4. The protein sequences of *G. intestinalis* are generally highly diverged when compared to their opisthokont orthologues (Morrison *et al.*, 2007; Xu *et al.*, 2020) but in case of the newly identified *Gi*Get5 and *Gi*Bag6, even the most sensitive homology detection algorithms such as HHpred failed to identify these in the genomic data. This indicated that while some of the GET pathway components are well conserved across the eukaryotic diversity (Get3, Sgt2, Get4, Get1), some of the components thought to be missing in eukaryotes could diverge beyond the recognition of the bioinformatics when opisthokont sequences are used as queries.

Hence, when we set up HMM-based searches for the identification of GET pathway components across eukaryotic diversity, we thus used as diverse dataset as possible, including *G. intestinalis* proteins, to build the protein specific HMMs. The models were used against the proteomes and transcriptomes of representatives of ten eukaryotic supergroups (Fig. 6, Table S2). Of all GET pathway components proteins Sgt2, Get4, and Get3, were found as the most conserved and they could be identified in all supergroups (Fig. 6, Table S2). Similarly, homologs of the membrane complex components, Get1 and Get2, were identified in most eukaryotes (Fig. 6, Table S2). In case of Get5 and Bag6 proteins, the sensitivity of the searches was highly increased by the inclusion of *G. intestinalis* and other Metamonada sequences. This enabled to identify at least one representative of Get5 in the genome/transcriptome also in Discoba, Fungi, Metazoa, *Telonema* sp., Alveolata and Rhizaria (Fig. 6, Table S2).

In the case of Bag6, the C-terminal domain of Bag6 (Figure 4C) was identified as a key identifier for the recognition of divergent Bag6 homologs. This allowed as to identify homologs of Bag6 in Metamonada, Metazoa, Amoebozoa, Archaeplastida (including Chloroplastida, Rhodophyta, and Glaucophyta), TSAR (including *Telonema* sp., Stramenopia, Alveolata, and Rhizaria), Haptista, CRuMs, and non-classified organism *Ancoracysta twista* (Fig. 6, Table S2).

DISCUSSION

Post-translational targeting of TA-proteins has been intensively studied over the last fifteen years, which resulted in the identification of the key GET pathway components and the characterization of their interaction (Schuldiner *et al.*, 2008; Stefanovic *et* Hegde, 2007). Yeast and metazoan cell systems played crucial role in these discoveries but interestingly, significant functional differences were already found between these two systems, such as the involvement of Bag6 in humans (Mariappan *et al.*, 2010; Q. Wang *et al.*, 2011), suggesting that the components of the GET pathway may vary in different eukaryotes.

Recently, the study of the GET pathway was also extended to Archaeplastida (Xing *et al.*, 2017) and Alveolata (Kumar *et al.*, 2021). Moreover, our recent characterization of *Gi*Get3 revealed full molecular mechanism, by which this targeting factor controls ATP hydrolysis and the substrate binding and its release (Fry *et al.*, 2022). *G. intestinalis* is a member of highly diverged Metamonada supergroup of eukaryotes that are commonly adopted to anaerobic environments (Adl *et al.*, 2012) and while many aspects of their cell biology underwent unique adaptations and dramatic simplification (Adam, 2021), core components of pathway such as those involved in the protein transport have remained conserved (Fry *et al.*, 2022). For this reason, *G. intestinalis* represents a great model for the evolutionary cell biology.

Together with Get3, *G. intestinalis* was also shown to possess Get4, Get2, and Sgt2 homologs (Fry *et al.*, 2022). Both Get5 and Bag6 of the pre-targeting complex were found which prompted us to identify if Get5 and Bag6 are truly missing in *G. intestinalis* or just diverged beyond recognition. In case of Bag6, its absence would be expected because of its previous identification only in Metazoa (Leznicki *et al.*, 2010; Mariappan *et al.*, 2010; Minami *et al.*, 2007). Using *Gi*Sgt2 and *Gi*Get4 as baits resulted in specific co-purification of two proteins, subsequently designated as *Gi*Get5 and *Gi*Bag6, which represent highly divergent homologs of Get5 and Bag6, respectively. Although *Gi*Get5 sequence appears as highly divergent, structural alignment of *Gi*Get5 and other homologs revealed the conservation of the C-terminal part, that, we proposed, could serve as a general determinant for Get5 identification in eukaryotes. *Gi*Get5 lacks the N-terminal part that is present in yeast Get5 and is responsible for Get4 binding. Interestingly, this is analogous

to human GET pathway, where human Get5 also lacks the N-terminal domain. Herein, Get4 and Get5 are bound separately by Bag6, which is not present in yeast. (Bozkurt *et al.*, 2010; Chang *et al.*, 2010; Chartron *et al.*, 2010; Mariappan *et al.*, 2010; Mock *et al.*, 2015). In this light, *G. intestinalis* GET pathway is reminiscent to the human pathway.

The identification of Bag6 in *G. intestinalis* actually represents first report of Bag6 homolog outside Metazoan. When compared to human protein, *Gi*Bag6 is shorter in sequence and lacks any similarity in the central part of the protein, however both proteins share the N-terminal UBL domains and multi-functional C-terminal part (Mock *et al.*, 2015). The C-terminus of Bag6 was shown to carry NLS that targets human Bag6 to nucleus, where the protein participates in gene regulation processes (Manchen *et* Hubberstey, 2001; P. Nguyen *et al.*, 2008). Both cytosolic and nuclear localization found also for *Gi*Bag6 suggesting similar functional repertoire for Bag6 in both cellular systems. However, detailed inspection of *Gi*Bag6 targeting is required to understand if it is also Get4 binding to Bag6 that prevents the nuclear translocation of the protein via masking the NLS (Q. Wang *et al.*, 2011). Of 14 proteins co-purified with *Gi*Bag6 two possible interactors small subunit of DNA-directed RNA polymerases (RPB10; GL50803_14413) and putative tRNA-dihydrouridine synthase (GL50803_3565) point to the nuclear function of *Gi*Bag6, while Lysyl-tRNA synthetase (GL50803_16766) and ribosomal subunit L38e (GL50803_34093) suggest the potential role for *Gi*Bag6 in the regulation of translation.

In addition to canonical GET pathway components, *Gi*Sgt2 shown interaction with the cytosolic chaperone hsp70 (GL50803_88765), the chaperone Ydj1 (GL50803_9808) and the heat shock factor binding protein (GL50803_7351). These results correspond with the fact that Sgt2 itself is not sufficient to maintain the solubility of TA proteins and TA proteins recognition and transfer are facilitated by cytosolic chaperones from hsp70 family and J-domain containing cochaperones (Cho *et al.*, 2021; Cho *et* Shan, 2018).

Importantly, the purification of *Gi*SGt2, *Gi*Get4, and *Gi*Bag6 showed specific presence of multiple proteasomal proteins of both the catalytic and regulatory complexes in the datasets. According to the number of co-purified subunits, these data show strong interaction of the proteasome with *Gi*Get4 and *Gi*Sgt2 but also some interaction with *Gi*Bag6. In this regard, the role of pre-targeting complex in quality control of mislocalized membrane proteins was previously shown for Sgt2 and Bag6 (Leznicki *et* High, 2012;

Minami *et al.*, 2010; Shao *et al.*, 2017; Wunderley *et al.*, 2014). Human Sgt2 and *Xenopus* Bag6 were found to bind Rpn13 and Rpn10 subunits of the proteasome, respectively (Kikukawa *et al.*, 2005; Leznicki *et al.*, 2015). Despite the fact, different proteasome subunits were found among the significantly enriched proteins with GET components in *G. intestinalis*, our data support the interaction of the pre-targeting complex with proteasome in a cellular system entirely different from Metazoa. Thus, the interaction likely reflects ancestral mechanism of protein binding and degradation.

The GET pathway, or at least its central element, Get3, has been considered as the evolutionarily ancient feature of eukaryotes as it could be identified not only in all eukaryotic groups, but its homologues function also in bacteria and archaea (Farkas *et al.*, 2019). In this work, we show that the conservation of the GET pathway can be extended to all seven GET components. While also homologues of Sgt2, Get4 can be readily identified in all eukaryotic supergroups, remaining GET components are often difficult to identify (Chang *et al.*, 2010; Fry *et al.*, 2022; Kumar *et al.*, 2021; Mariappan *et al.*, 2010; Srivastava *et al.*, 2017). The identification of *Gi*Get5 and *Gi*Bag6 enabled us to define signature motifs at the C-terminal part of both proteins (Mock *et al.*, 2015, 2017) that proved to be very specific in identifying other diverged homologues across eukaryotic tree of life.

Similarly, the GET components of the membrane complex that is responsible for inserting the TA-protein into the membrane are also prone to diverge quickly in different lineages of eukaryotes. Get1 protein is a member of OXA1 superfamily of proteins that also includes two other ER-destined insertases, TMCO1 and EMC3 (Anghel *et al.*, 2017). While our search of Get1 homologs showed common presence of the protein to all eukaryotic supergroups, several eukaryotes including *G. intestinalis* carry only Get1 homolog. The missing homolog of TMCO1 (Transmembrane and coiled-coil domain 1) is a Ca2+ channel responsible for homeostasis in ER (Q. C. Wang *et al.*, 2016). The EMC3 is a subunit of the EMC complex (ER membrane protein complex) through which multi-pass transmembrane proteins (Chitwood *et al.*, 2018; Tian *et al.*, 2019) and some TA-proteins are inserted into the ER membrane (Guna *et al.*, 2018). Our results together with the fact that the EMC complex is missing in *G. intestinalis* (Wideman, 2015), raise the question of how other membrane proteins are transported into the ER membrane and how the loss

of the TMCO1 and EMC complex is replaced. The common origin of all three subfamilies (Get1, TMCO1, EMC3) suggests potential functional redundancy. Recent report shows that mitochondria targeted EMC3-EMC6 and Get1-Get2 proteins are able to partially rescue the function of mitochondrial Oxa1 mutant (Güngör *et al.*, 2022).

Homology search for Get2 homologs was previously carried out among metazoan, plants, and fungi (Borgese, 2020). The protein was experimentally identified in yeast and human (Schuldiner *et al.*, 2008, Yamamoto *et* Sakisaka, 2012), recently also in Alveolata (*P. falciparum*) (Kumar *et al.*, 2021) and plants (*Arabidopsis thaliana*) (Asseck *et al.*, 2021) and *G. intestinalis* (Fry *et al.*, 2022). Our search showed the presence of Get2 in an extended dataset and demonstrated the conservation of Get2 in almost all eukaryotic groups.

Altogether, these data show that LECA already contained all GET pathway components (Sgt2, Get1-Get5, Bag6) and some organisms such as yeast lost Bag6 while simultaneously extended their Get4 and Get5 proteins to allow their mutual interaction in the absence of Bag6. Our result also suggests that the interaction of the pre-targeting complex with the proteasome may represent another ancestral eukaryotic feature. Yet, more experiments are needed to understand the biological role of the interaction and its conservation across eukaryotic tree of life.

Methods

Bioinformatic searches

Seed alignments from Pfam of homologs of GET proteins were searched by local HMMER search v3.3.2 (http://hmmer.org/) (Mistry *et al.*, 2013, 2021) against UniProt protein databases released 2022-02 (Bateman *et al.*, 2021) and against the the EukProt database v2 (Richter *et al.*, n.d.). Sequences were aligned using MAFFT v.7 (Katoh *et al.*, 2019). Returned positive hits were verified using by HHpred algorithm (Zimmermann *et al.*, 2018) (https://toolkit.tuebingen.mpg.de/tools/hhpred) against the Pfam-A_v33.1 database. Multiple sequence alignments were enriched with newly classified homologs and used for subsequent searches. The distribution of protein domains was determined by combination of the Pfam and InterPro online search (El-Gebali *et al.*, 2019; Mitchell *et*

al., 2019). Protein similarity was calculated from sequence alignment by online tool Ident and Sim (Stothard, 2000).

Phylogenetic analyses

MMseqs2 (Steinegger *et* Söding, 2017) was used to cluster datasets using the easy-cluster algorithm with different minimum sequence identity thresholds that were the most suitable for each dataset. Representative sequences were aligned using MAFFT version 7 (Katoh *et al.*, 2019). The L-INS-i algorithm with default parameters was used. Poorly aligned regions were subsequently automatically removed by trimAl (Capella-Gutiérrez *et al.*, 2009) with the gappyout mode. Phylogenetic analysis was performed with IQ-TREE2 software (L. T. Nguyen *et al.*, 2015). The best fitting model was obtained with the ModelFinder (Kalyaanamoorthy *et al.*, 2017). To reduce the computing time, UFBoot method (ultra-fast bootstrap) method was employed (Hoang *et al.*, 2018).

Cell culture, cloning, and transfection

The *G. intestinalis* strain WB (ATCC 30957) was grown in TYI-S-33 medium supplemented with 10% heat-inactivated bovine serum, 0.1% bovine bile and antibiotics at 37 °C (Keister, 1983). The genes encoding *Gi*Sgt2 (GL50803_7287), *Gi*Get4 (GL50803_112893), *Gi*Get5 (Gl50803_19378) and *Gi*Bag6 (GL50803_5069) were amplified from genomic DNA and inserted into plasmid pOndra (Dolezal *et al.*, 2005) with C- or N-terminal BAP tag (Table S3). Then 1×10^7 cells expressing cytosolic BirA (Martincová *et al.*, 2015) were electroporated with Bio-Rad Gene Pulser using an exponential protocol (U = 30 V; C = 1,000 µF; R = 750 Ω). The transfected cells were grown in TYI-S-33 media supplemented with antibiotics (58 µg/ml
puromycin and 600 µg/ml G418) and used for native isolation to establish GET components in *G. intestinalis*.

Native isolation of Get proteins, proteomics.

Native isolation of biotinylated proteins was made as previously described (Fry *et al.*, 2022). In brief, lysate from *G. intestinalis* cells co-expressing BAP-tagged GET pathway component and BirA was incubated with streptavidin-coupled magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen, Waltham, MA) for 1 hr at 4°C with gentle rotation. The magnetic beads were washed 3 times in wash solution (50mM HEPES pH 7.4, 150 mM potassium acetate, 5mM magnesium acetate, 1 mM DTT, and 10% glycerol) then washed 3 times in PBS. Magnetic beads with bound proteins were submitted to tandem mass spectrometry analysis. Isolation was made in triplicates and lysate from *G. intestinalis* cells expressing only BirA was used as control for protein quantification. Data analyses was performed using Perseus 1.6.1.3 (Tyanova *et al.*, 2016) and visualized as a volcano plot using online tool VolcaNoseR with Fold change 1 and Significance threshold 2 (Goedhart *et* Luijsterburg, 2020).

Immunofluorescence microscopy

Trophozoites of *G. intestinalis* were fixed in 1% paraformaldehyde, placed on cover slip as previously described (Voleman *et al.*, 2017). BAP-tagged proteins were detected by a commercial anti-BAP mouse primary antibody (dilution 1:1000) and Get3 was detect by anti-Get3 rabbit antibody (dilution 1:1000) (Fry *et al.*, 2022). Alexa Fluor 488-conjugated donkey anti mouse (cat. nr. A-21202) and Alexa Fluor 594-conjugated donkey anti-rabbit IgG (cat. nr. A-21207) were used at a 1:1000 dilution as secondary antibodies. Slides were mounted with Vectashield containing DAPI (cat. nr. H-1200-10) and imaged with a Leica SP8 FLIM inverted confocal microscope. Maximum intensity projection and brightness/contrast corrections were performed in the FIJI ImageJ software (Schindelin *et al.*, 2012).

Figures

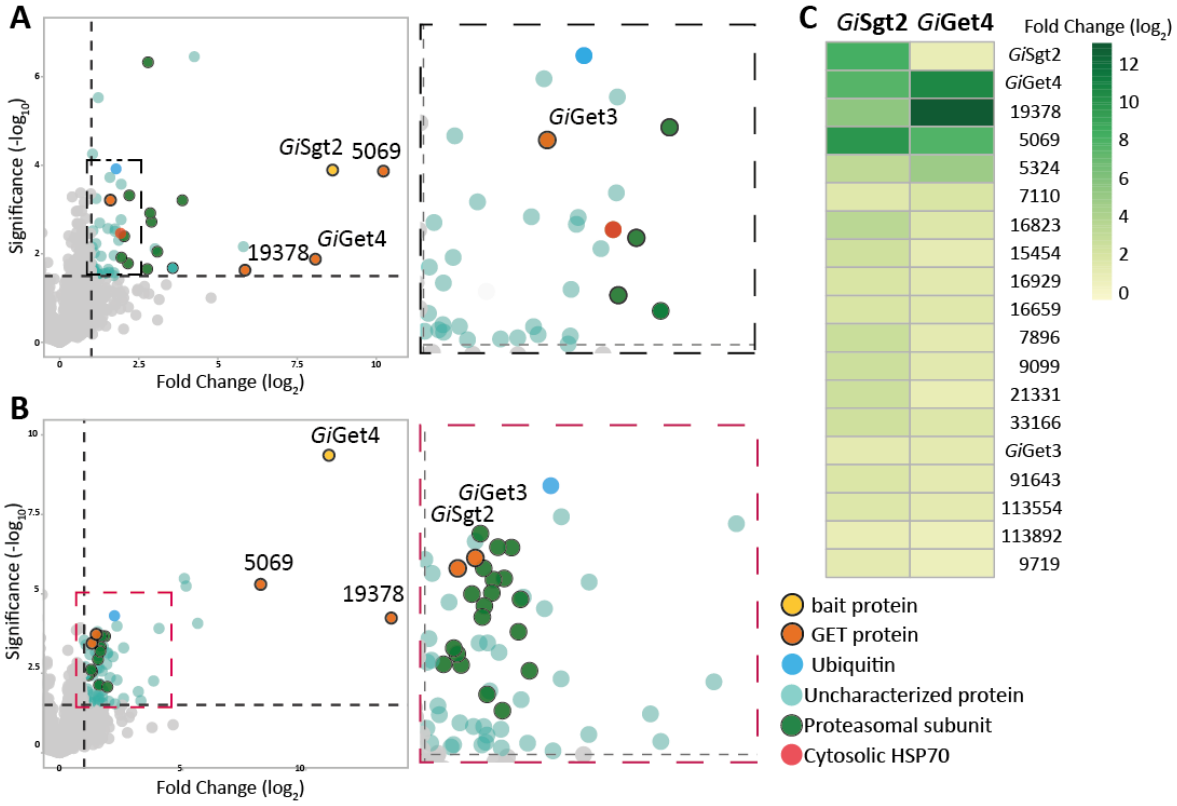
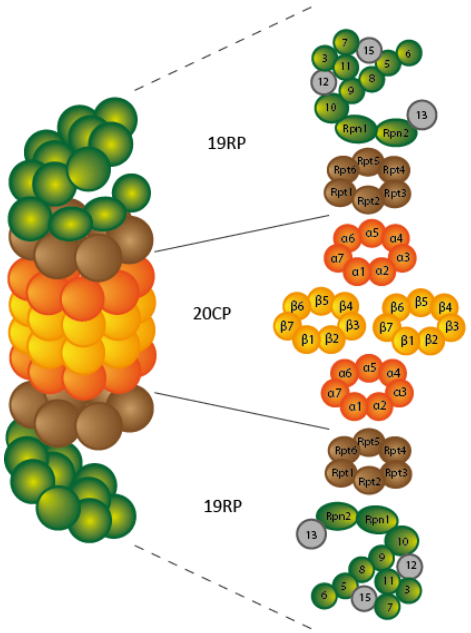


Figure 1

A. Isolation of *Gi*Sgt2 under native conditions revealed specific co-purification of 51 proteins including known GET pathway components (Get3 and Get4) and two yet unknown binding partners (GL50803_005069 and GL50803_0019378). **B,** Analogous isolation of *Gi*Get4 returned same two partner proteins among 67 identified hits. Specific co-purification of 19 S proteasome subunits was revealed for both GET pathway components. **C,** Combination of *Gi*Sgt2- and *Gi*Get4-specific dataset comprised 19 common proteins with four proteins (*Gi*Sgt2, *Gi*Get4, GL50803_0019378, and GL50803_005069) at the centre of the mutual interactome.



	Subparticle	Subunit	ORF nr.	*Gi* Sgt2	*Gi* Get4	*Gi* Bag6
19S RP	Base	Rpt1	86683	×	✓	✓
		Rpt2	113554	✓	✓	×
		Rpt3	7950	×	✓	×
		Rpt4	21331	✓	✓	×
		Rpt5	4365	×	✓	×
		Rpt6	17106	×	✓	×
		Rpn1	33166	✓	✓	✓
		Rpn2	91643	✓	✓	×
		Rpn10	15604	×	×	×
		Rpn13	NA	NA	NA	NA
	Lid	Rpn3	15454	✓	✓	×
		Rpn5	16929	✓	✓	×
		Rpn6	16659	✓	✓	✓
		Rpn7	4331	×	✓	×
		Rpn8	7896	✓	✓	×
		Rpn9	9099	✓	✓	×
		Rpn11	16823	✓	✓	×
		Rpn12	NA	NA	NA	NA
		Rpn15	NA	NA	NA	NA
20S CP	Alpha Ring	α1	7962	×	✓	×
		α2	11434	×	×	×
		α3	15099	×	×	×
		α4	11486	×	×	×
		α5	2980	×	×	×
		α6	16924	×	✓	×
		α7	14497	×	×	×
	Beta Ring	β1	13127	×	×	×
		β2	1995	×	×	×
		β3	13756	×	×	×
		β4	27059	×	×	×
		β5	12949	×	×	×
		β6	9824	×	×	×
		β7	3209	×	×	×

Figure 2

Proteasome subunits co-purified with *Gi*Sgt2, *Gi*Get4 and *Gi*Bag6. (Left) Scheme of *G. intestinalis* proteasome: subunits of base and lid of 19S subparticle are shown in green and brown, respectively. The α a β rings of 20 CP subunits are shown in orange yellow, respectively. Three subunits were not identified in *G. intestinalis* (Jerlström-Hultqvist *et al.* 2012) (grey). (Right) Proteasome subunits identified among the significantly enriched proteins co-purified with the GET pathway components.

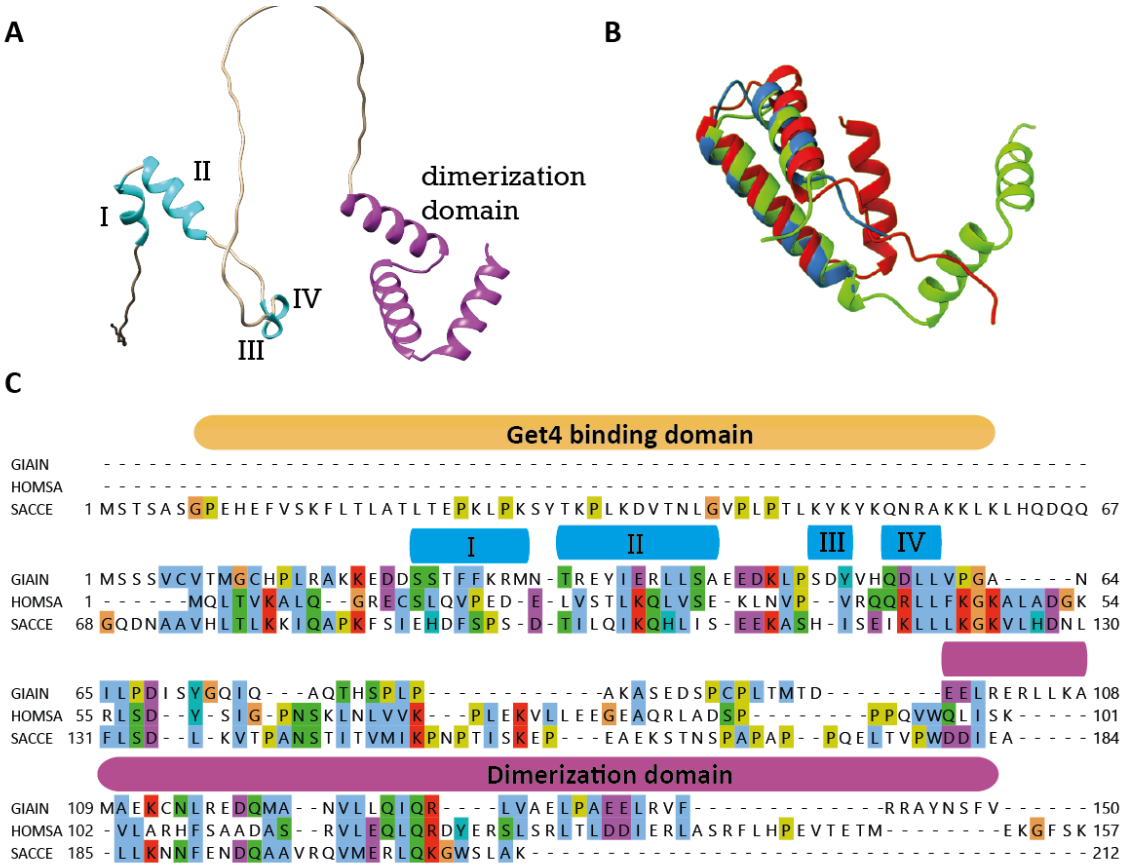
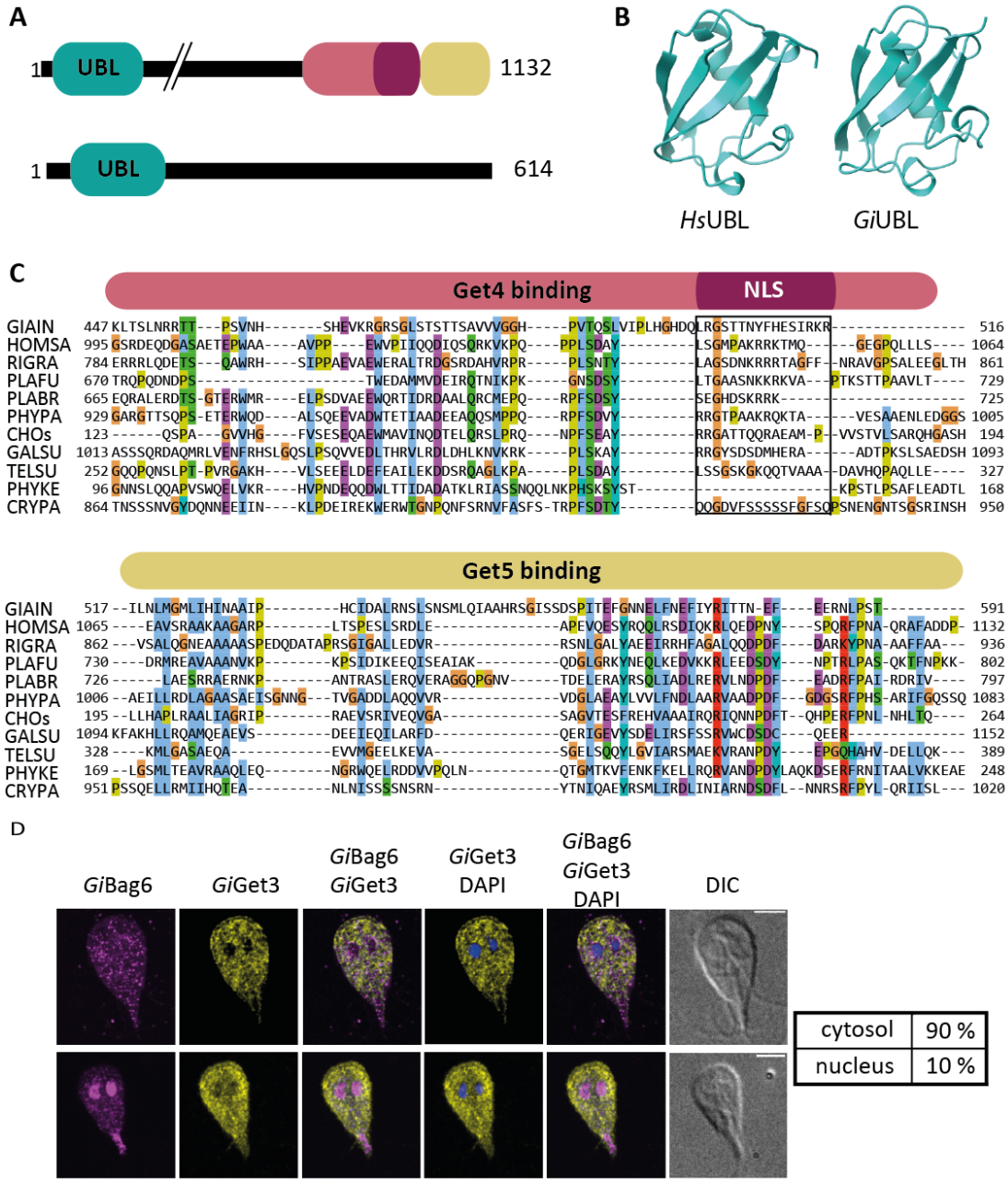


Figure3

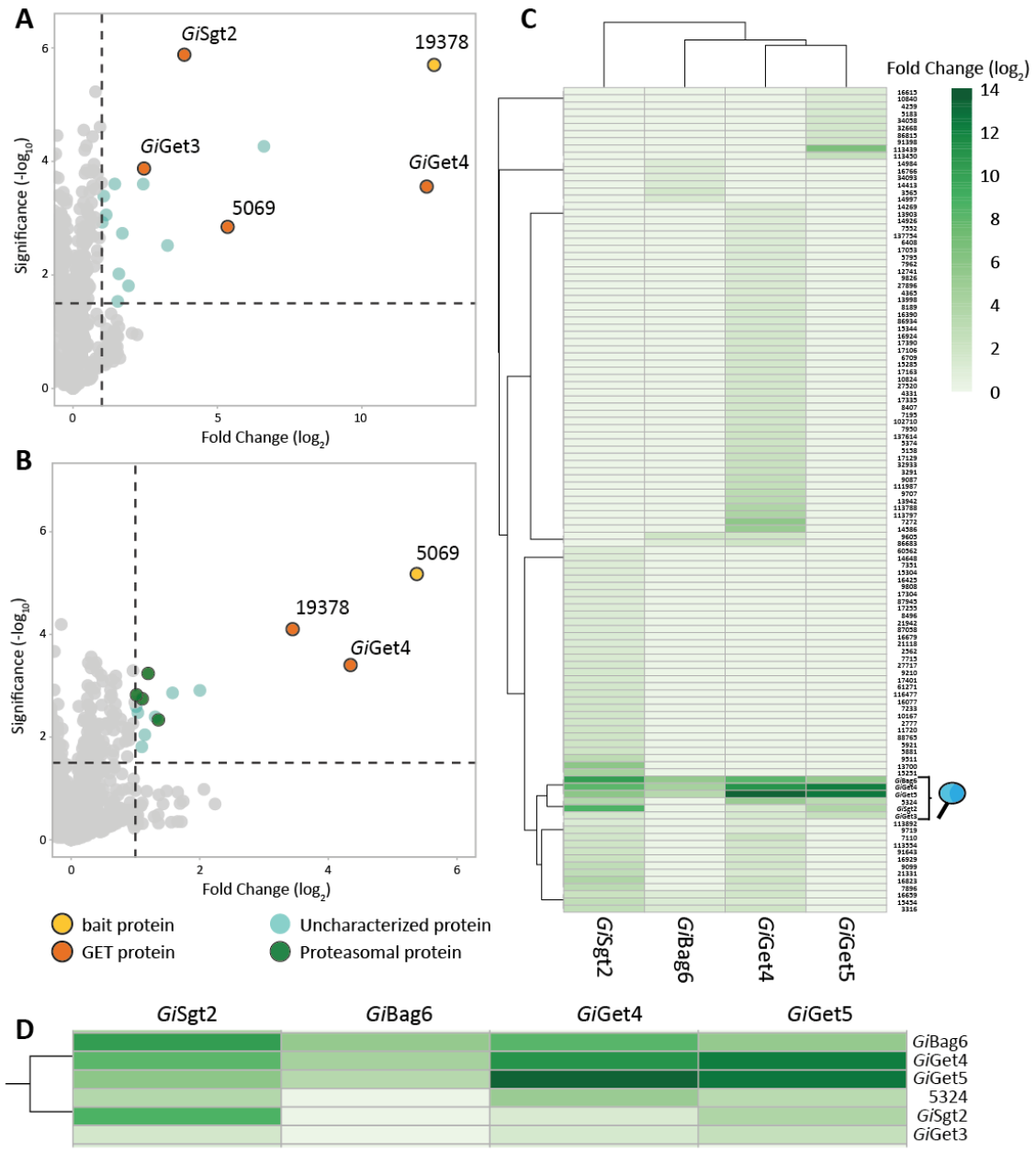
Identification of highly divergent Get5 orthologue in *G. intestinalis* A, Except for the C-terminal part of the protein, Alphafold2 prediction of GL50803_0019378 did not result in a high confidence model, as most of the protein sequence remained unstructured. **B,** Comparison of Alphafold2 structural prediction of the C-terminal dimerization domains of yeast (blue), human (green), and *G. intestinalis* (red) Get5 orthologues. **C,** The protein sequence alignment shows the missing N-terminal Get4 binding domain in human and *G. intestinalis* proteins.



433

434 **Figure 4. Identification of Bag6 homologue in *G. intestinalis***

435 **A,** Domain arrangement in human *and G. intestinalis* Bag6 sequences. **B**, Aplhafold2model
436 of UBL domain in human and *G. intestinalis* Bag6. **C**, sequence alignment of conserved C-
437 terminal part of Bag6. GIAIN - G. intestinalis; HOMSA - *H. sapiens*; RIGRA - *Rigifila ramosa*;
438 PLAFU - *Planoprotostelium fungivorum*; PLABR - *Plasmodiophora brassicae*; PHYPA -
439 *Physcomitrella patens*; CHOs - *Choanocystis sp.*; GALSU - *Galdieria sulphuraria*; TELSU -
440 *Telonema subtile*; PHYKE - *Phytophthora kernoviae*; CRYPA - *Cryptosporidium parvum*
441 **D**, Localization of BAP-tagged *Gi*Bag6. Using immunofluorescence microscopy, *Gi*Bag6 was
442 found in spotted pattern across the cytosol. Approximately 10 % of the cells showed
443 localization in nucleus and in the posterior end of the cell. *Gi*Bag6 (purple); *Gi*Get3
444 (yellow); nucleus – DAPI (blue)



445

446 **Figure 5. The interactomes of *Gi*Get5 and *Gi*Bag6 revealed stable *in vivo* interaction of**
447 **the GET pathway components.**
448 **A,** Isolation of *Gi*Get5 under native conditions revealed specific co-purification of 16
449 proteins including *Gi*Get4, *Gi*Bag6, *Gi*Sgt2, and *Gi*Get3. **B,** Analogous isolation of *Gi*Bag6
450 returned 14 proteins including *Gil*Get4 and *Gi*Get5. **C,** The isolated interactomes of four
451 GET components using hierarchical clustering. **D,** Core of the cytosolic complex of the GET
452 pathway in *G. intestinalis*, which is composed of pre-targeting proteins *Gi*Sgt2, *Gi*Get4,
453 *Gi*Get5 and *Gi*Bag6, targeting factor *Gi*Get3 and *G. intestinalis*-specific protein of unknown
454 function GL50803_005324).

455

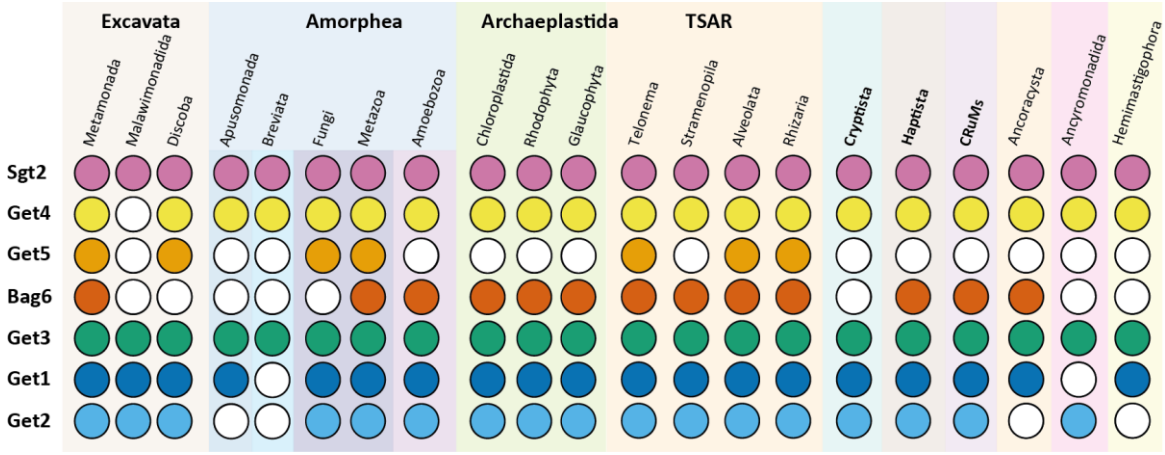


Figure 6. Presence of GET pathway in eukaryotes

The experimentally obtained divergent sequence of *G. intestinalis* GET pathway homologs facilitated the identification of GET proteins in almost all eukaryotic groups. The most conserved GET pathway components are Sgt2, Get4, Get3, Get1, and Get2, which were identified in almost all eukaryotic groups. Presence of Get5 homologs was identified in Metamonada, Discoba, Metazoa, Fungi, *Telonema*, Alveolata, and Rhizaria. Bag6 homologs were found in Metamonada, Metazoa, Amoebozoa, *Telonema*, Alveolata and Rhizaria. Our data show that LECA contained all GET pathway components.

Supplementary figures

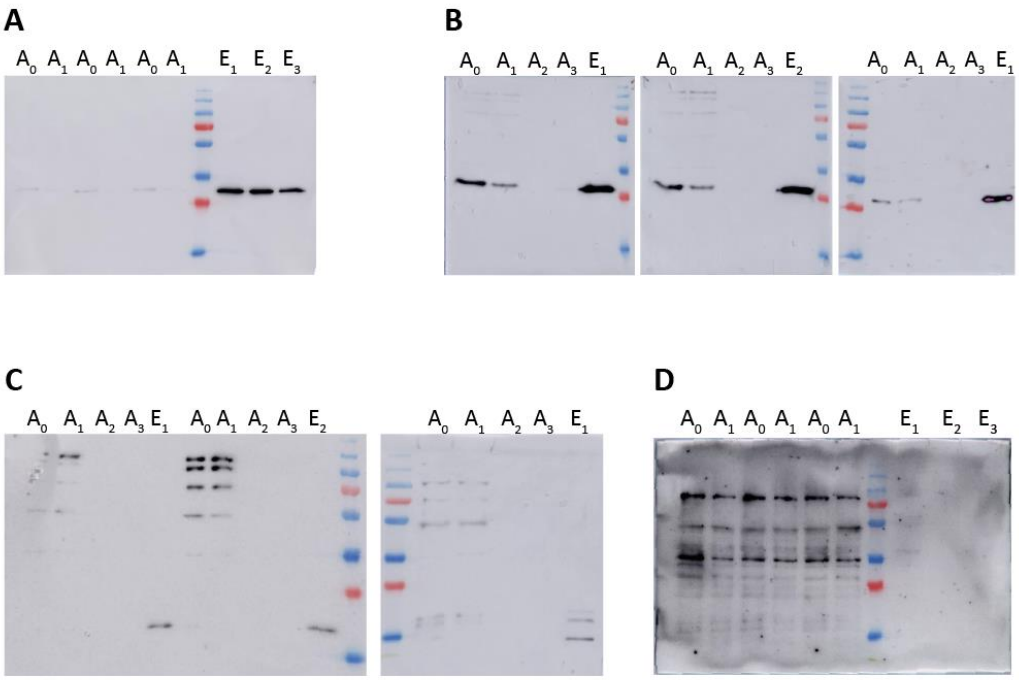


Figure S1. Native co-purification of *G. intestinalis* GET components

Lysate (A_0) of *G. intestinalis* cell co-expressing BirA and BAP-tagged *Gi*Sgt2 (**A**), *Gi*Get4 (**B**), *Gi*Get5 (**C**) or *Gi*Bag6 (**D**) was incubated with streptavidin-coupled magnetic beads. After incubation (A_1), the magnetic beads were washed 3 times in wash solution (A_2) and then washed 3 times in PBS (A_3). One fifth of the sample was boiled in sample buffer to eluate (E) bait protein with bound protein partners to test the efficiency by western blot analyses.

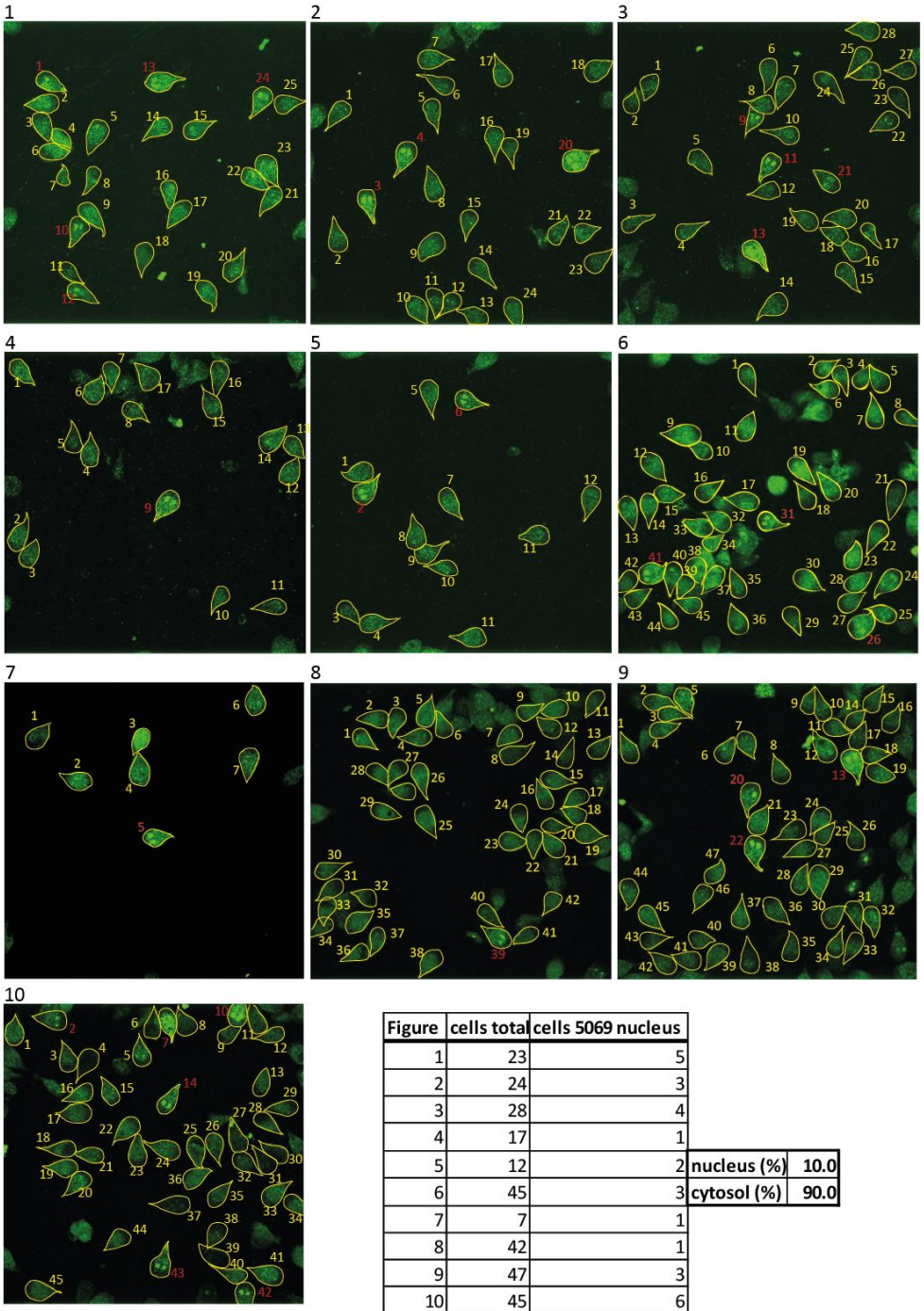


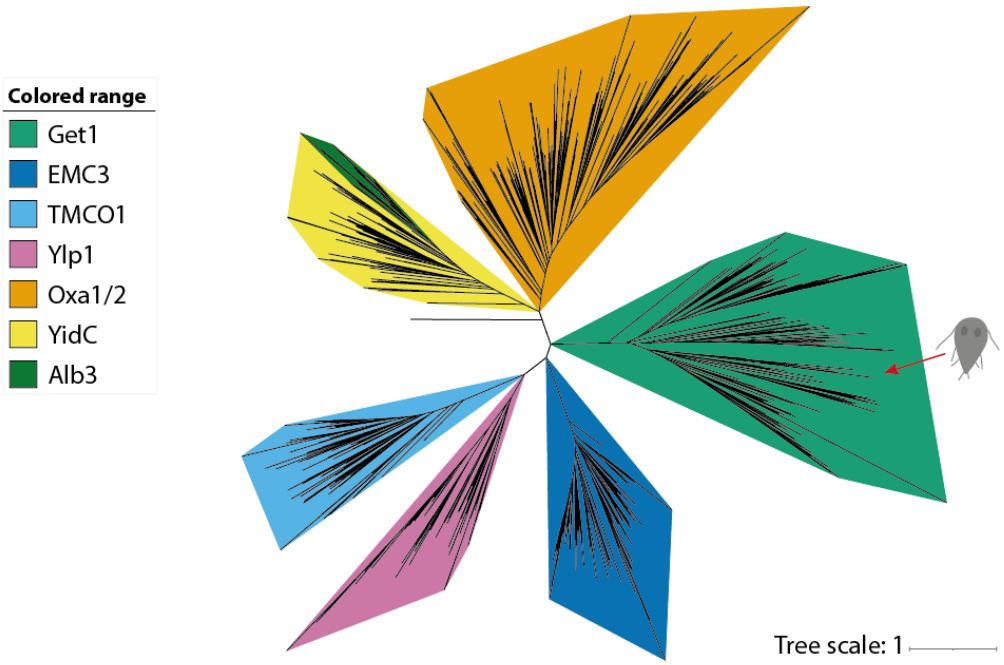
Figure	cells total	cells 5069 nucleus
1	23	5
2	24	3
3	28	4
4	17	1
5	12	2
6	45	3
7	7	1
8	42	1
9	47	3
10	45	6
	290	**29**

nucleus (%)	10.0
cytosol (%)	90.0

Figure S2. Quantification of GiBag6 dual localization

From ten figures of cells of *G. intestinalis* expressing BAP-tagged *Gi*Bag6, which were fixed and detect by fluorescence antibody, the total number of cells were counted. The cells with nuclear localization visible in all z-layers were counted as well. The resulting ration show that 10 % cells have *Gi*Bag6 localized also in nucleus.

A


Colored range
Get1
EMC3
TMCO1
Ylp1
Oxa1/2
YidC
Alb3
Tree scale: 1

B

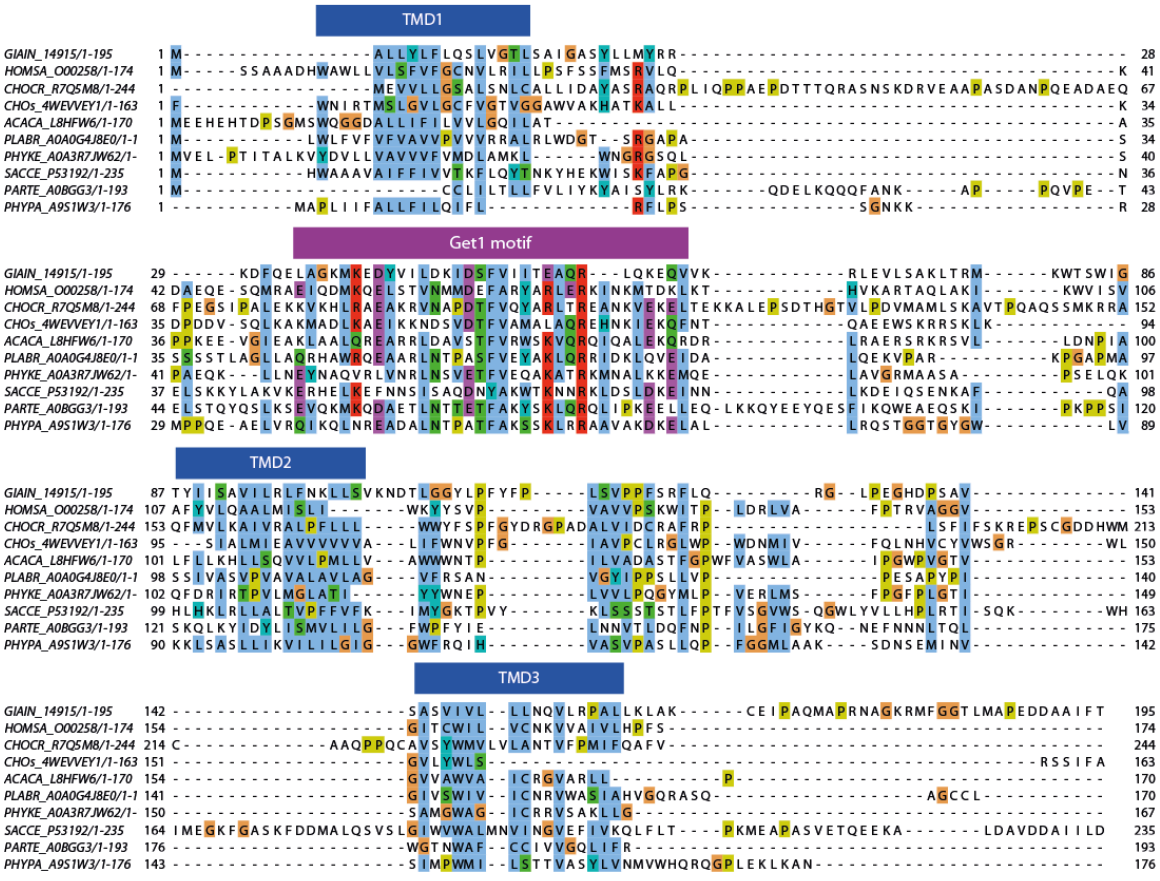

TMD1
Get1 motif
TMD2
TMD3

Figure S3. *G. intestinalis* possess only one ER-destined OXA1 family protein

A, HHsearch revealed only one OXA1 family candidate in *G. intestinalis* genome. Phylogenetic analyses of all OXA1 family members showed that *G. intestinalis* OXA1 candidate (GL50803_14915) clustered with Get1 homologs. **B,** Sequence alignment of Get1 homologs from representative organisms and *G. intestinalis* show Get1 topology, which include three TMDs and conserved soluble domain (between first and second TMDs).

List of references:

Adam, R. D. (2021). *Giardia duodenalis*: Biology and Pathogenesis. *Clinical Microbiology Reviews*, *34*(4), e0002419. https://doi.org/10.1128/CMR.00024-19

Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Gall, L. le, Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., Spiegel, F. W. (2012). The revised classification of eukaryotes. *The Journal of Eukaryotic Microbiology*, *59*(5), 429–514. https://doi.org/10.1111/J.1550-7408.2012.00644.X

Anghel, S. A., McGilvray, P. T., Hegde, R. S., Keenan, R. J. (2017). Identification of Oxa1 Homologs Operating in the Eukaryotic Endoplasmic Reticulum. *Cell Reports*, *21*(13). https://doi.org/10.1016/j.celrep.2017.12.006

Asseck, L. Y., Mehlhorn, D. G., Monroy, J. R., Ricardi, M. M., Breuninger, H., Wallmeroth, N., Berendzen, K. W., Nowrousian, M., Xing, S., Schwappach, B., Bayer, M., Grefen, C. (2021). Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(1), e2017636118. https://doi.org/10.1073/PNAS.2017636118

Aviram, N., Schuldiner, M. (2017). Targeting and translocation of proteins to the endoplasmic reticulum at a glance. *Journal of Cell Science*, *130*(24), 4079–4085. https://doi.org/10.1242/JCS.204396

Banerji, J., Sands, J., Strominger, J. L., Spies, T. (1990). A gene pair from the human major histocompatibility complex encodes large proline-rich proteins with multiple repeated motifs and a single ubiquitin-like domain. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(6), 2374. https://doi.org/10.1073/PNAS.87.6.2374

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., … Zhang, J.

(2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Borgese, N. (2020). Searching for remote homologs of CAML among eukaryotes. *Traffic*, *21*(10), 647–658. https://doi.org/10.1111/tra.12758

Borgese, N., Coy-Vergara, J., Colombo, S. F., Schwappach, B. (2019). The Ways of Tails: the GET Pathway and more. *The Protein Journal*, *38*(3). https://doi.org/10.1007/s10930-019-09845-4

Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favaloro, V., Rippe, K., Hurt, E., Dobberstein, B., Sinning, I. (2009). Structural insights into tail-anchored protein binding and membrane insertion by Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21131–21136. https://doi.org/10.1073/PNAS.0910223106

Bozkurt, G., Wild, K., Amlacher, S., Hurt, E., Dobberstein, B., Sinning, I. (2010). The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis. *FEBS Letters*, *584*(8), 1509–1514. https://doi.org/10.1016/J.FEBSLET.2010.02.070

Capella-Gutiérrez, S., Silla-Martínez, J. M., Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chang, Y.-W., Chuang, Y.-C., Ho, Y.-C., Cheng, M.-Y., Sun, Y.-J., Hsiao, C.-D., Wang, C. (2010). *Crystal Structure of Get4-Get5 Complex and Its Interactions with Sgt2, Get3, and Ydj1. Journal of Biological Chemistry*, *285*(13), 9962-9970 https://doi.org/10.1074/jbc.M109.087098

Chartron, J. W., Suloway, C. J. M., Zaslaver, M., Clemons, W. M. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12127–12132. https://doi.org/10.1073/pnas.1006036107

Chio, U. S., Cho, H., Shan, S. O. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annual Review of Cell and Developmental Biology*, *33*, 417–438. https://doi.org/10.1146/annurev-cellbio-100616-060839

621 Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., Hegde, R. S. (2018). EMC Is Required to
622 Initiate Accurate Membrane Protein Topogenesis. *Cell*, *175*(6), 1507-1519.e16.
623 https://doi.org/10.1016/j.cell.2018.10.009
624 Cho, H., Shan, S. (2018). Substrate relay in an Hsp70-cochaperone cascade safeguards tail-
625 anchored membrane protein targeting. *The EMBO Journal*, *37*(16).
626 https://doi.org/10.15252/embj.201899264
627 Cho, H., Shim, W. J., Liu, Y., Shan, S. O. (2021). J-domain proteins promote client relay from
628 Hsp70 during tail-anchored membrane protein targeting. *The Journal of Biological*
629 *Chemistry*, *296*. https://doi.org/10.1016/J.JBC.2021.100546
630 Dolezal, P., Smíd, O., Rada, P., Zubácová, Z., Bursać, D., Suták, R., Nebesárová, J., Lithgow,
631 T., Tachezy, J. (2005). *Giardia* mitosomes and trichomonad hydrogenosomes share
632 a common mode of protein targeting. *Proceedings of the National Academy of*
633 *Sciences of the United States of America*, *102*(31), 10924–10929.
634 https://doi.org/10.1073/PNAS.0500349102
635 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M.,
636 Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L.,
637 Piovesan, D., Tosatto, S. C. E., Finn, R. D. (2019). The Pfam protein families database
638 in 2019. *Nucleic Acids Research*, *47*(D1), D427-D432.
639 https://doi.org/10.1093/nar/gky995
640 Farkas, Á., de Laurentiis, E. I., Schwappach, B. (2019). The natural history of Get3-like
641 chaperones. In *Traffic* *20*(5), 311–324. Blackwell Munksgaard.
642 https://doi.org/10.1111/tra.12643
643 Fry, M. Y., Najdrová, V., Maggiolo, A. O., Saladi, S. M., Doležal, P., Clemons, W. M. (2022).
644 Structurally derived universal mechanism for the catalytic cycle of the tail-anchored
645 targeting factor Get3. *Nature Structural Molecular Biology*, 29(8), 820-830.
646 https://doi.org/10.1038/s41594-022-00798-4
647 Goedhart, J., Luijsterburg, M. S. (2020). VolcaNoseR is a web app for creating, exploring,
648 labeling and sharing volcano plots. *Scientific Reports. 10*(1), 20560.
649 https://doi.org/10.1038/s41598-020-76603-3
650 Gristick, H. B., Rao, M., Chartron, J. W., Rome, M. E., Shan, S. O., Clemons, W. M. (2014).
651 Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3

652 by Get4. *Nature Structural and Molecular Biology*, *21*(5), 437–442.
653 https://doi.org/10.1038/nsmb.2813
654 Gristick, H. B., Rome, M. E., Chartron, J. W., Rao, M., Hess, S., Shan, S. O., Clemons, W. M.
655 (2015). Mechanism of Assembly of a Substrate Transfer Complex during Tail-
656 anchored Protein Targeting. *The Journal of Biological Chemistry*, *290*(50), 30006–
657 30017. https://doi.org/10.1074/JBC.M115.677328
658 Guna, A., Hegde, R. S. (2018). Transmembrane Domain Recognition during Membrane
659 Protein Biogenesis and Quality Control. *Current Biology*, *28*(8), R498–R511.
660 https://doi.org/10.1016/J.CUB.2018.02.004
661 Guna, A., Volkmar, N., Christianson, J. C., Hegde, R. S. (2018). The ER membrane protein
662 complex is a transmembrane domain insertase. *Science*, *359*(6374), 470–473.
663 https://doi.org/10.1126/science.aao3099
664 Güngör, B., Flohr, T., Garg, S. G., Herrmann, J. M. (2022). The ER membrane complex
665 (EMC) can functionally replace the Oxa1 insertase in mitochondria. *PLoS Biology*,
666 *20*(3), e3001380. https://doi.org/10.1371/JOURNAL.PBIO.3001380
667 Hegde, R. S., Keenan, R. J. (2011). Tail-anchored membrane protein insertion into the
668 endoplasmic reticulum. In *Nature Reviews Molecular Cell Biology*. *12*(12), 787–798.
669 Nat Rev Mol Cell Biol. https://doi.org/10.1038/nrm3226
670 Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., Vinh, L. S. (2018). UFBoot2:
671 Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*,
672 *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281
673 Jerlström-Hultqvist, J., Stadelmann, B., Birkestedt, S., Hellman, U., Svärd, S. G. (2012).
674 Plasmid vectors for proteomic analyses in *Giardia*: purification of virulence factors
675 and analysis of the proteasome. *Eukaryotic Cell*, *11*(7), 864–873.
676 https://doi.org/10.1128/EC.00092-12
677 Jiang, H. (2021). Quality control pathways of tail-anchored proteins. In *Biochimica et*
678 *Biophysica Acta - Molecular Cell Research* (Vol. 1868, Issue 2). Elsevier B.V.
679 https://doi.org/10.1016/j.bbamcr.2020.118922
680 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., Jermiin, L. S. (2017).
681 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature*
682 *Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

683 Katoh, K., Rozewicki, J., Yamada, K. D. (2019). MAFFT online service: multiple sequence
684 alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*,
685 *20*(4). https://doi.org/10.1093/bib/bbx108
686 Keister, D. B. (1983). Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented
687 with bile. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *77*(4),
688 487–488. https://doi.org/10.1016/0035-9203(83)90120-7
689 Kikukawa, Y., Minami, R., Shimada, M., Kobayashi, M., Tanaka, K., Yokosawa, H.,
690 Kawahara, H. (2005). Unique proteasome subunit Xrpn 10c is a specific receptor for
691 the antiapoptotic ubiquitin-like protein Scythe. *FEBS Journal*, *272*(24), 6373–6386.
692 https://doi.org/10.1111/j.1742-4658.2005.05032.x
693 Kumar, T., Maitra, S., Rahman, A., Bhattacharjee, S. (2021). A conserved guided entry of
694 tail-anchored pathway is involved in the trafficking of a subset of membrane proteins
695 in *Plasmodium falciparum*. *PLoS Pathogens*, *17*(11), e1009595.
696 https://doi.org/10.1371/JOURNAL.PPAT.1009595
697 Kutay, U., Hartmann, E., Rapoport, T. A. (1993). A class of membrane proteins with a C-
698 terminal anchor. *Trends in Cell Biology*, *3*(3), 72–75. https://doi.org/10.1016/0962-
699 8924(93)90066-A
700 Leznicki, P., Clancy, A., Schwappach, B., High, S. (2010). Bat3 promotes the membrane
701 integration of tail-anchored proteins. *Journal of Cell Science*, *123*(13), 2170–2178.
702 https://doi.org/10.1242/jcs.066738
703 Leznicki, P., High, S. (2012). SGTA antagonizes BAG6-mediated protein triage. *Proceedings*
704 *of the National Academy of Sciences of the United States of America*, *109*(47), 19214–
705 19219. https://doi.org/10.1073/pnas.1209997109
706 Leznicki, P., Korac-Prlic, J., Kliza, K., Husnjak, K., Nyathi, Y., Dikic, I., High, S. (2015). Binding
707 of SGTA to Rpn13 selectively modulates protein quality control. *Journal of Cell*
708 *Science*, *128*(17), 3187–3196. https://doi.org/10.1242/JCS.165209/-/DC1
709 Manchen, S. T., Hubberstey, A. v. (2001). Human Scythe contains a functional nuclear
710 localization sequence and remains in the nucleus during staurosporine-induced
711 apoptosis. *Biochemical and Biophysical Research Communications*, *287*(5), 1075–
712 1082. https://doi.org/10.1006/BBRC.2001.5701

Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R. J., Hegde, R. S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature*, *466*(7310), 1120–1124. https://doi.org/10.1038/nature09296

Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R. S., Keenan, R. J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature*, *477*(7362), 61–69. https://doi.org/10.1038/nature10362

Martincová, E., Voleman, L., Pyrih, J., Žárský, V., Vondráčková, P., Kolísko, M., Tachezy, J., Doležal, P. (2015). Probing the Biology of *Giardia intestinalis* Mitosomes Using In Vivo Enzymatic Tagging. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.00448-15

Mateja, A., Paduch, M., Chang, H. Y., Szydlowska, A., Kossiakoff, A. A., Hegde, R. S., Keenan, R. J. (2015). Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science*, *347*(6226), 1152–1155. https://doi.org/10.1126/science.1261671

Mateja, A., Szlachcic, A., Downing, M. E., Dobosz, M., Mariappan, M., Hegde, R. S., Keenan, R. J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature*, *461*(7262), 361–366. https://doi.org/10.1038/nature08319

McDowell, M. A., Heimes, M., Fiorentino, F., Mehmood, S., Farkas, Á., Coy-Vergara, J., Wu, D., Bolla, J. R., Schmid, V., Heinze, R., Wild, K., Flemming, D., Pfeffer, S., Schwappach, B., Robinson, C. V., Sinning, I. (2020). Structural Basis of Tail-Anchored Membrane Protein Biogenesis by the GET Insertase Complex. *Molecular Cell*, *80*(1), 72–86. https://doi.org/10.1016/j.molcel.2020.08.012

Minami, R., Hayakawa, A., Kagawa, H., Yanagi, Y., Yokosawa, H., Kawahara, H. (2010). BAG-6 is essential for selective elimination of defective proteasomal substrates. *Journal of Cell Biology*, *190*(4), 637–650. https://doi.org/10.1083/jcb.200908092

Minami, R., Shimada, M., Yokosawa, H., Kawahara, H. (2007). Scythe regulates apoptosis through modulating ubiquitin-mediated proteolysis of the *Xenopus* elongation factor XEF1AO. *The Biochemical Journal*, *405*(3), 495–501. https://doi.org/10.1042/BJ20061886

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., Bateman, A. (2021).

744 Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–
745 D419. https://doi.org/10.1093/nar/gkaa913
746 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., Punta, M. (2013). Challenges in homology
747 search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids*
748 *Research*, *41*(12), e121–e121. https://doi.org/10.1093/nar/gkt263
749 Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D.,
750 Chang, H.-Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I.,
751 Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., … Finn, R. D. (2019).
752 InterPro in 2019: improving coverage, classification and access to protein sequence
753 annotations. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1100
754 Mock, J. Y., Chartron, J. W., Zaslaver, M., Xu, Y., Ye, Y., Clemons, W. M. (2015). Bag6
755 complex contains a minimal tail-anchor-targeting module and a mock BAG domain.
756 *Proceedings of the National Academy of Sciences of the United States of America*,
757 *112*(1), 106–111. https://doi.org/10.1073/pnas.1402745112
758 Mock, J. Y., Xu, Y., Ye, Y., Clemons, W. M. (2017). Structural basis for regulation of the
759 nucleo-cytoplasmic distribution of Bag6 by TRC35. *Proceedings of the National*
760 *Academy of Sciences of the United States of America*, *114*(44), 11679–11684.
761 https://doi.org/10.1073/PNAS.1702940114
762 Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A.
763 A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H.
764 G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning,
765 G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner,
766 D. S., Samuelson, J., Svard, S. G., Sogin, M. L. (2007). Genomic minimalism in the
767 early diverging intestinal parasite *Giardia lamblia*. *Science (New York, N.Y.)*,
768 *317*(5846), 1921–1926. https://doi.org/10.1126/SCIENCE.1143837
769 Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., Minh, B. Q. (2015). IQ-TREE: A fast and
770 effective stochastic algorithm for estimating maximum-likelihood phylogenies.
771 *Molecular Biology and Evolution*, *32*(1), 268–274.
772 https://doi.org/10.1093/molbev/msu300
773 Nguyen, P., Bar-Sela, G., Sun, L., Bisht, K. S., Cui, H., Kohn, E., Feinberg, A. P., Gius, D.
774 (2008). BAT3 and SET1A Form a Complex with CTCFL/BORIS To Modulate H3K4

775 Histone Dimethylation and Gene Expression. *Molecular and Cellular Biology*, *28*(21),
776 6720–6729. https://doi.org/10.1128/mcb.00568-08
777 Richter, D. J., Berney, C., H Strassert, J. F., Poh, Y.-P., Muñoz-Gómez, S. A., Wideman, J. G.,
778 Burki, F., de, C. (n.d.). EukProt: a database of genome-scale predicted proteins across
779 the diversity of 1 eukaryotes *2 3*. https://doi.org/10.6084/m9.figshare.12417881.v3
780 Rodrigo-Brenni, M. C., Gutierrez, E., Hegde, R. S. (2014). Cytosolic Quality Control of
781 Mislocalized Proteins Requires RNF126 Recruitment to Bag6. *Molecular Cell*, *55*(2),
782 227–237. https://doi.org/10.1016/j.molcel.2014.05.025
783 Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch,
784 S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V.,
785 Eliceiri, K., Tomancak, P., Cardona, A. (2012). Fiji: an open-source platform for
786 biological-image analysis. *Nature Methods*, *9*(7), 676–682.
787 https://doi.org/10.1038/NMETH.2019
788 Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H. D., Schwappach,
789 B., Weissman, J. S. (2008). The GET Complex Mediates Insertion of Tail-Anchored
790 Proteins into the ER Membrane. *Cell*, *134*(4), 634–645.
791 https://doi.org/10.1016/j.cell.2008.06.025
792 Shao, S., Rodrigo-Brenni, M. C., Kivlen, M. H., Hegde, R. S. (2017). Mechanistic basis for a
793 molecular triage reaction. *Science*, *355*(6322), 298–302.
794 https://doi.org/10.1126/science.aah6130
795 Srivastava, R., Zalisko, B. E., Keenan, R. J., Howell, S. H. (2017). The GET system inserts the
796 tail-anchored protein, SYP72, into Endoplasmic Reticulum Membranes. *Plant*
797 *Physiology*, *173*(2), 1137–1145. https://doi.org/10.1104/pp.16.00928
798 Stefanovic, S., Hegde, R. S. (2007). Identification of a Targeting Factor for Posttranslational
799 Membrane Protein Insertion into the ER. *Cell*, *128*(6), 1147–1159.
800 https://doi.org/10.1016/j.cell.2007.01.036
801 Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y. Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F.,
802 Bernhard, F., Denic, V., Dötsch, V., Sinning, I. (2011). Structural basis for tail-anchored
803 membrane protein biogenesis by the Get3-receptor complex. *Science*, *333*(6043),
804 758–762. https://doi.org/10.1126/science.1207125

Steinegger, M., Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. In *Nature Biotechnology* (Vol. 35, Issue 11, pp. 1026–1028). Nature Publishing Group. https://doi.org/10.1038/nbt.3988

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, *28*(6). https://doi.org/10.2144/00286IR01

Suloway, C. J. M., Chartron, J. W., Zaslaver, M., Clemons, W. M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(35), 14849–14854. https://doi.org/10.1073/PNAS.0907522106

Thress, K., Henzel, W., Shillinglaw, W., Kornbluth, S. (1998). Scythe: a novel reaper-binding apoptotic regulator. *The EMBO Journal*, *17*(21), 6135. https://doi.org/10.1093/EMBOJ/17.21.6135

Tian, S., Wu, Q., Zhou, B., Choi, M. Y., Ding, B., Yang, W., Dong, M. (2019). Proteomic Analysis Identifies Membrane Proteins Dependent on the ER Membrane Protein Complex. *Cell Reports*, *28*(10), 2517-2526.e5. https://doi.org/10.1016/j.celrep.2019.08.006

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods 2016 13:9*, *13*(9), 731–740. https://doi.org/10.1038/nmeth.3901

Vilardi, F., Lorenz, H., Dobberstein, B. (2011). WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *Journal of Cell Science*, *124*(8), 1301–1307. https://doi.org/10.1242/jcs.084277

Voleman, L., Najdrová, V., ÁstvaldssonÁsgeir, Tumová, P., Einarsson, E., Švindrych, Z., Hagen, G. M., Tachezy, J., Svärd, S. G., Doležal, P. (2017). *Giardia intestinalis* mitosomes undergo synchronized fission but not fusion and are constitutively associated with the endoplasmic reticulum. *BMC Biology*,*15*(1). https://doi.org/10.1186/S12915-017-0361-Y

834 Wang, F., Brown, E. C., Mak, G., Zhuang, J., Denic, V. (2010). A chaperone cascade sorts
835 proteins for posttranslational membrane insertion into the endoplasmic reticulum.
836 *Molecular Cell*, *40*(1), 159–171. https://doi.org/10.1016/j.molcel.2010.08.038
837 Wang, Q. C., Zheng, Q., Tan, H., Zhang, B., Li, X., Yang, Y., Yu, J., Liu, Y., Chai, H., Wang, X.,
838 Sun, Z., Wang, J. Q., Zhu, S., Wang, F., Yang, M., Guo, C., Wang, H., Zheng, Q., Li, Y.,
839 Chen, Q., Zhou, A., Tang, T. S. (2016). TMCO1 is an ER Ca2+ load-activated Ca2+
840 channel. *Cell*, *165*(6), 1454–1466. https://doi.org/10.1016/j.cell.2016.04.051
841 Wang, Q., Crnković, V., Preisinger, C., Stegmüller, J. (2021). The parkinsonism-associated
842 protein FBXO7 cooperates with the BAG6 complex in proteasome function and
843 controls the subcellular localization of the complex. *Biochemical Journal*, *478*(12),
844 2179–2199. https://doi.org/10.1042/bcj20201000
845 Wang, Q., Liu, Y., Soetandyo, N., Baek, K., Hegde, R., Ye, Y. (2011). A Ubiquitin Ligase-
846 Associated Chaperone Holdase Maintains Polypeptides in Soluble States for
847 Proteasome Degradation. *Molecular Cell*, *42*(6), 758–770.
848 https://doi.org/10.1016/j.molcel.2011.05.010
849 Wideman, J. G. (2015). The ubiquitous and ancient ER membrane protein complex (EMC):
850 Tether or not? *F1000Research*, *4*, 624.
851 https://doi.org/10.12688/F1000RESEARCH.6944.2
852 Wunderley, L., Leznicki, P., Payapilly, A., High, S. (2014). SGTA regulates the cytosolic
853 quality control of hydrophobic substrates. *Journal of Cell Science*, *127*(21), 4728–
854 4739. https://doi.org/10.1242/JCS.155648/-/DC1
855 Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P.,
856 Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y. D., Grossmann, G., Grefen, C. (2017).
857 Loss of GET pathway orthologs in *Arabidopsis thaliana* causes root hair growth
858 defects and affects SNARE abundance. *Proceedings of the National Academy of*
859 *Sciences of the United States of America*, *114*(8), E1544–E1553.
860 https://doi.org/10.1073/pnas.1619525114
861 Xu, F., Jex, A., Svärd, S. G. (2020). A chromosome-scale reference genome for *Giardia*
862 *intestinalis* WB. *Scientific Data*, *7*(1). https://doi.org/10.1038/S41597-020-0377-Y

Yamamoto, Y., Sakisaka, T. (2012). Molecular Machinery for Insertion of Tail-Anchored Membrane Proteins into the Endoplasmic Reticulum Membrane in Mammalian Cells. *Molecular Cell*, *48*(3), 387–397. https://doi.org/10.1016/j.molcel.2012.08.028

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, *430*(15). https://doi.org/10.1016/j.jmb.2017.12.007

Najdrová et al. BMC Biology (2020) 18:77
https://doi.org/10.1186/s12915-020-00814-3

BMC Biology

RESEARCH ARTICLE Open Access

The evolution of the Puf superfamily of proteins across the tree of eukaryotes

Vladimíra Najdrová[1], Courtney W. Stairs[2], Martina Vinopalová[1], Luboš Voleman[1] and Pavel Doležal[1*]

Abstract

Background: Eukaryotic gene expression is controlled by a number of RNA-binding proteins (RBP), such as the proteins from the Puf (Pumilio and FBF) superfamily (PufSF). These proteins bind to RNA via multiple Puf repeat domains, each of which specifically recognizes a single RNA base. Recently, three diversified PufSF proteins have been described in model organisms, each of which is responsible for the maturation of ribosomal RNA or the translational regulation of mRNAs; however, less is known about the role of these proteins across eukaryotic diversity.

Results: Here, we investigated the distribution and function of PufSF RBPs in the tree of eukaryotes. We determined that the following PufSF proteins are universally conserved across eukaryotes and can be broadly classified into three groups: (i) Nop9 orthologues, which participate in the nucleolar processing of immature 18S rRNA; (ii) 'classical' Pufs, which control the translation of mRNA; and (iii) PUM3 orthologues, which are involved in the maturation of 7S rRNA. In nearly all eukaryotes, the rRNA maturation proteins, Nop9 and PUM3, are retained as a single copy, while mRNA effectors ('classical' Pufs) underwent multiple lineage-specific expansions. We propose that the variation in number of 'classical' Pufs relates to the size of the transcriptome and thus the potential mRNA targets. We further distinguished full set of PufSF proteins in divergent metamonad *Giardia intestinalis* and initiated their cellular and biochemical characterization.

Conclusions: Our data suggest that the last eukaryotic common ancestor (LECA) already contained all three types of PufSF proteins and that 'classical' Pufs then underwent lineage-specific expansions.

Keywords: RNA-binding protein, RNA processing, LECA, Puf superfamily proteins, *Giardia intestinalis*

Background

The Puf superfamily (PufSF) of proteins encompasses a class of eukaryotic RNA-binding proteins (RBPs), which interact with the 3′-untranslated regions (3′-UTRs) of mRNA in the cytosol or with the precursors of rRNA molecules in the nucleolus [1–3]. The name of the protein family is derived from Pumilio and Fem-3 binding factor of *Drosophila melanogaster* and *Caenorhabditis elegans*, respectively [4, 5]. PufSF proteins are defined by the presence of Puf repeats, each of which binds a single base of an RNA molecule [6–9]. In the past two decades, three types of PufSF proteins (i.e. Puf, Nop9, and PUM3, discussed below) have been distinguished according to their biological role and structural arrangement [2, 10–12].

The so-called classical Puf proteins (Pufs) bind 3′-UTRs of mRNA and usually have eight Puf repeats that organize into a crescent-shaped structure [13]. Each Puf repeat contains a tripartite recognition motif (TRM), and it is the combinations of eight TRMs, which specifies the sequence motif at which a particular Puf binds. In vivo, a single Puf protein can recognize hundreds of transcripts and, together with additional protein partners, regulate protein translation [1]. Puf proteins can mediate the repression [14–16], the activation [17] of

* Correspondence: pavel.dolezal@natur.cuni.cz
[1]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic
Full list of author information is available at the end of the article

gene expression, or the site-specific translation [18, 19]. These proteins function in the cytoplasm.

The two other types of PufSF proteins operate predominantly within the nucleolus, where ribosome biogenesis takes place [20]. Nop9 proteins carry 11 Puf repeats organized into a U-shaped structure, which participate in the processing and folding of 18S rRNA in the nucleolus [10, 21, 22], while PUM3 proteins (including the human PUM3/PufA and yeast Puf6) bind double-stranded DNA or RNA without a sequence specificity [2, 23]. Similar to Nop9, PUM3 proteins carry 11 Puf repeats, although organized in L-shaped domain, and they are involved in the nucleolar processing of the large subunit of the ribosome [23], potentially during the processing of 7S to 5.8S rRNA [2].

Although the PufSF proteins are essential for core processes of the eukaryotic cell, such as rRNA maturation, investigations of PufSF proteins have been restricted to only a few lineages of eukaryotes (e.g. Opisthokonta and Viridiplantae). These studies recognized that the number of PufSF proteins within and between lineages of eukaryotes is highly variable [24, 25]. Due to the number of different splice isoforms and the repetitive nature of the proteins, it has been difficult to delineate the evolutionary history of this essential protein family throughout eukaryotic evolution. Moreover, given the role of PufSF proteins in regulating the translation of a complex network of transcripts in modern eukaryotes [19, 26], it is tempting to speculate that PufSF proteins played a role in the origin of cellular complexity in eukaryotes. Indeed, such information is critical to understand the true nature of the last eukaryotic common ancestor (LECA); a concept of which can be drawn upon by examining the cellular properties of diverse eukaryotes, including the highly diverged organisms [27].

Metamonada represent highly diverged unicellular anaerobic eukaryotes [28] carrying specialized mitochondria [29]. Some Metamonada are important human and animal pathogens [30]. One of the best studied species of Metamonada is the human intestinal parasite *Giardia intestinalis*. Although the RNA metabolism in *G. intestinalis* is poorly understood, we would argue that this organism can be useful in studying various aspects of eukaryotic RNA biology owing to its transcriptome streamlining and overall extreme biology. For example, unlike most eukaryotes, the processing of rRNA and the actual nature of *G. intestinalis* nucleolus are still under debate [31]. Moreover, *G. intestinalis* generates large number of sterile transcripts of unknown function, which are both capped and polyadenylated [32]. To date, only six *cis*-spliced and two *trans*-spliced transcripts have been described in *G. intestinalis* making it easier to predict the transcriptome purely from genomic data [33, 34]. The 5′-untranslated regions (5′-UTRs) of *G. intestinalis* mRNAs are efficiently capped and bound by the ribosome despite being extremely short (i.e. 0–14 nucleotides) [35, 36]. Therefore, posttranscriptional regulation of gene expression is mostly limited to the stability and sequestration of the mRNAs [37]. Thus, 3′-UTRs remain the key regions of mRNAs, which affect its stability and localization via the interaction with RNA-binding proteins [37].

Here, we report systematic bioinformatic survey of distribution of PufSF proteins with sampling across major eukaryotic supergroups. Our analyses show three groups of proteins encompassing (i) Nop9, (ii) Puf, and (iii) PUM3 homologues. In a given organism, Nop9 and PUM3 are usually represented by a single gene, while the number of Pufs is highly variable. However, the actual number of Pufs correlates with the number of transcripts of the particular lineages and thus the number of putative mRNA targets. These data also suggest that the LECA already contained one Nop9, one PUM3, and two Puf proteins and that the large copy number of Pufs in modern organisms can be explain by lineage-specific expansions.

We were able to identify all three PufSF proteins even within *G. intestinalis*. However, their initial characterization points to unique adaptations in *G. intestinalis* RNA metabolism.

Results

Classification of PufSF proteins

We initially classified the PufSF proteins across eukaryotic diversity. First, we performed a clustering analysis based on sequence similarity. This unbiased approach is based on mutual pairwise BLAST comparisons, and it is especially useful for the analysis of large protein datasets [38]. The initial dataset contained 4469 unique eukaryotic proteins carrying Puf repeat domain(s) (PF00806). No PufSF proteins were identified in Archaea or Bacteria, confirming the eukaryotic origin of the family. The clustering analysis revealed three major groups of PufSF proteins (Fig. 1, Additional file 1: Table S1): the Puf cluster, consisting of 2955 proteins containing eight Puf repeats in the C-terminal part of the protein; the PUM3 cluster, consisting of 674 PUM3 orthologues, including plant PUM24 proteins, with up to eleven Puf repeats; and the Nop9 cluster, consisting of 675 Nop9 proteins including the plant PUM23 proteins.

The proximity of the Puf and PUM3 proteins in our clustering analysis suggests that these proteins are more similar at the sequence level when compared to Nop9s, despite carrying different numbers of Puf repeats. Given the conservation of 11 Puf repeats in both the rRNA-binding proteins PUM3 and Nop9 proteins, we suspect that this was the ancestral domain arrangement of all members of the PufSF. Following adaptation to

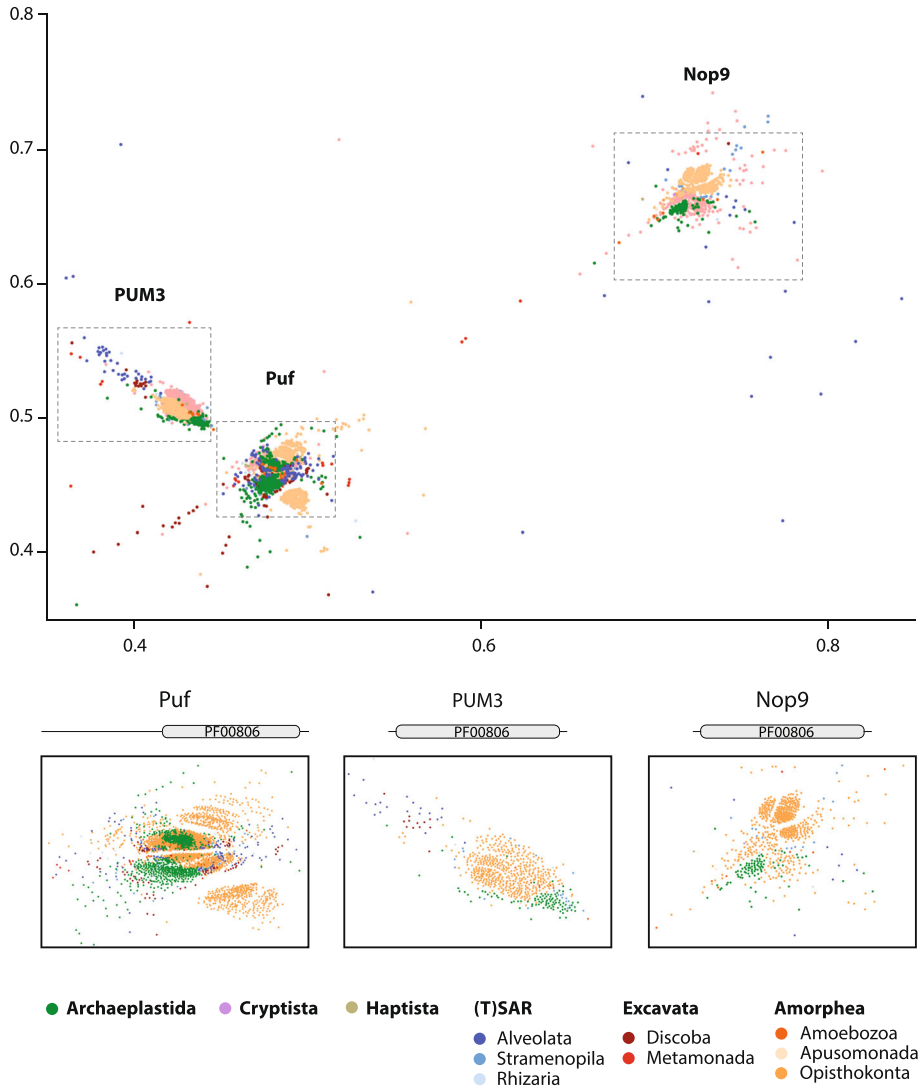


Fig. 1 Clustering analysis of Puf family proteins. PufSF proteins were analysed by CLANS. A total of 4469 proteins formed three major clusters of proteins corresponding to Nop9 homologues (Nop9 cluster), 'classical' Puf proteins (Puf cluster), and PUM3 orthologues (PUM3 cluster). The protein sequences were colour-coded according to the taxonomic affiliation to the eukaryotic supergroup. The details of each large cluster show mostly lineage-specific subclusters, except two large subclusters of Puf proteins. The diagrams above the cluster details depict the general domain arrangement of PufSF proteins; PF00806 corresponds to Puf RNA-binding repeat. The *x* and *y* axes represent relative positions of the protein sequence in 2D CLANS

interaction with mRNA molecules, the Puf proteins adopted a domain arrangement of only eight Puf domains [2, 21, 22]. The clustering approach was not sensitive enough to reveal detailed taxon-specific grouping of PufSF proteins except for the most represented eukaryotic groups of animals, plants, and fungi. The only exception was the formation of two clear large subclusters within the Puf cluster (Fig. 1) containing proteins of mixed taxonomic affiliation, which indicated the existence of at least two different Puf orthologues in the last common ancestor of eukaryotes. The position of eukaryotic supergroups encompassing protist lineages was rather dispersed across the subclusters. Metamonad proteins including the *G. intestinalis* sequences represented one of the most diverge proteins of the family.

Taxonomic distribution of PufSF proteins

The number of PufSF proteins encoded in a given genome differs significantly across the eukaryotic diversity (e.g. [4, 25]). Hence, we surveyed the distribution of PufSF of proteins on a species level across the tree of eukaryotes. To this aim, we retrieved all eukaryotic 1180 reference proteomes from UniProtKB and classified them as Puf, PUM3, or Nop9 orthologues. While the

dataset is biassed towards the proteomes from Opisthokonta and plants, it also contains curated proteomes of species from other eukaryotic supergroups. We used combination of InterPro precomputed protein families as a final determiner for the affiliation to one of the PufSF members (IPR001313—Puf repeat, IPR040000—Nop9, IPR040059—PUM3). In total, 7762 proteins were identified, and the proteins were classified according to hierarchic taxonomic groups (Fig. 2, Additional file 2: Table S2). Plotting the taxonomic distribution of the proteins showed that the highest number of PufSF proteins can be found in most of the plant species (Streptophyta), where the number of proteins ranged from 10 to 50. Extremely high number of proteins was also identified in ciliate *Paramecium tetraurelia* (43) but not in other ciliates or alveolates. In addition, all organism of analysed Euglenozoa group such as parasitic *Trypanosoma* and *Leishmania* species showed higher number of proteins (12–22). On the opposite end of the spectrum were parasites with reduced genomes, especially microsporidia or *Cryptosporidium* species with only one or two PufSF proteins. However, many animal taxa including insect and nematodes were also found to have only two or three proteins. Hence, at this point, we observed no clear relationship between the number of PufSF proteins and the biology or complexity of the surveyed organisms.

Three types of PufSF proteins have undergone different evolution in eukaryotes

Of all 7762 proteins from the reference proteomes, 1135 Nop9s, 5423 Pufs, and 1204 PUM3 proteins were identified (Additional file 2: Table S2). Interestingly, on average, each eukaryotic species contains a single Nop9 and PUM3 homologue and five Pufs (Fig. 3a). Moreover, while the number of Pufs is highly variable among lineages, the occurrence of single Nop9 and PUM3 proteins seems to be retained across eukaryotic diversity. The high number of PufSF proteins found in plants, Euglenozoa, and other species reflects lineage-specific amplification of the Puf proteins and not the ancestral state of early eukaryotes (Fig. 3a). The discrepancy observed between Nop9 and PUM3 gene copy number compared to Puf copy number might be related to different selective pressures experienced by these proteins owing to their role in the biogenesis of ribosomal RNA. On the other hand, given the Pufs' role in controlling the translation of multiple mRNAs—which will vary between organisms—it is possible that the number of Pufs will differ and could instead relate to the total number of the protein-coding genes in the cell. In order to test this hypothesis, we normalized the number of Pufs with respect to the total number of the protein-coding genes in the corresponding species (Fig. 3b) (Additional file 2: Table S2). Interestingly, the resulting ratio between Pufs and

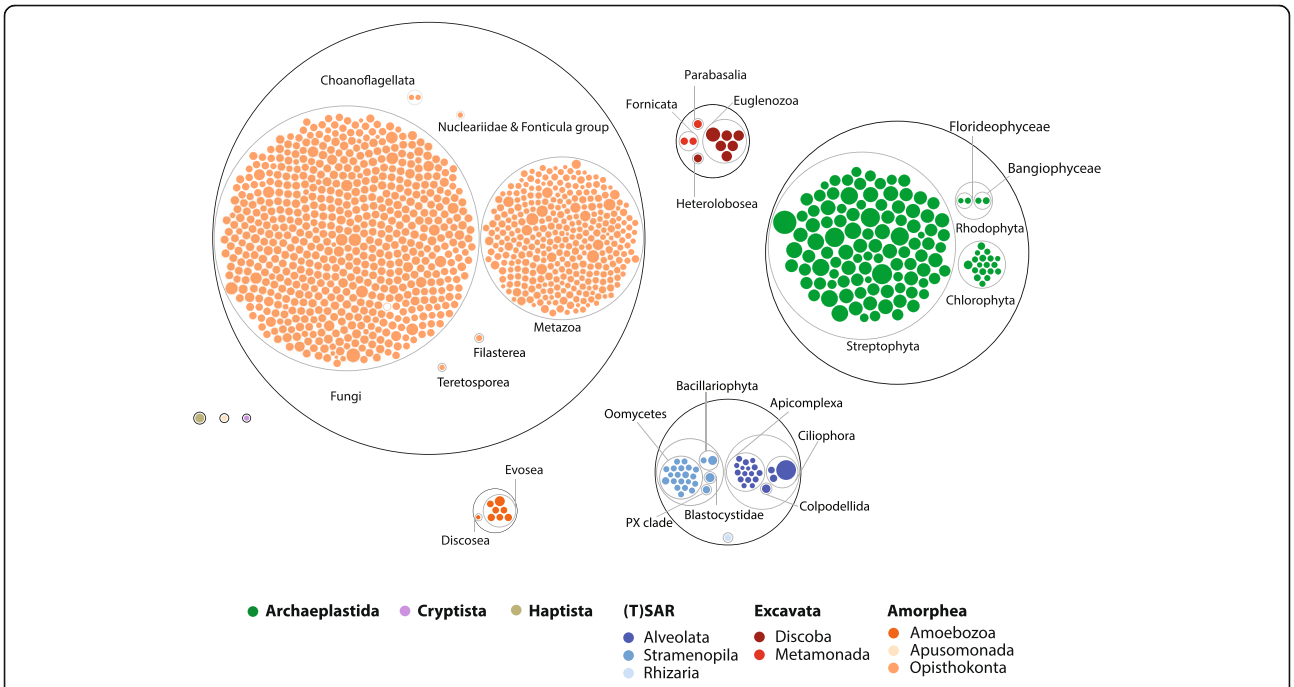


Fig. 2 PufSF proteins in major groups of eukaryotes. Puf family proteins identified in all UniProt reference proteomes were grouped according to taxonomic affiliation as specified in Additional file 2: Table S2 and showed in circle packing plot. Each dot corresponds to a genome of particular eukaryotic species, and the size of the dot represents a number of Puf family proteins in the predicted proteome and the colour-codes to the taxonomic affiliation to the eukaryotic supergroup. Grey circles depict taxonomic groups

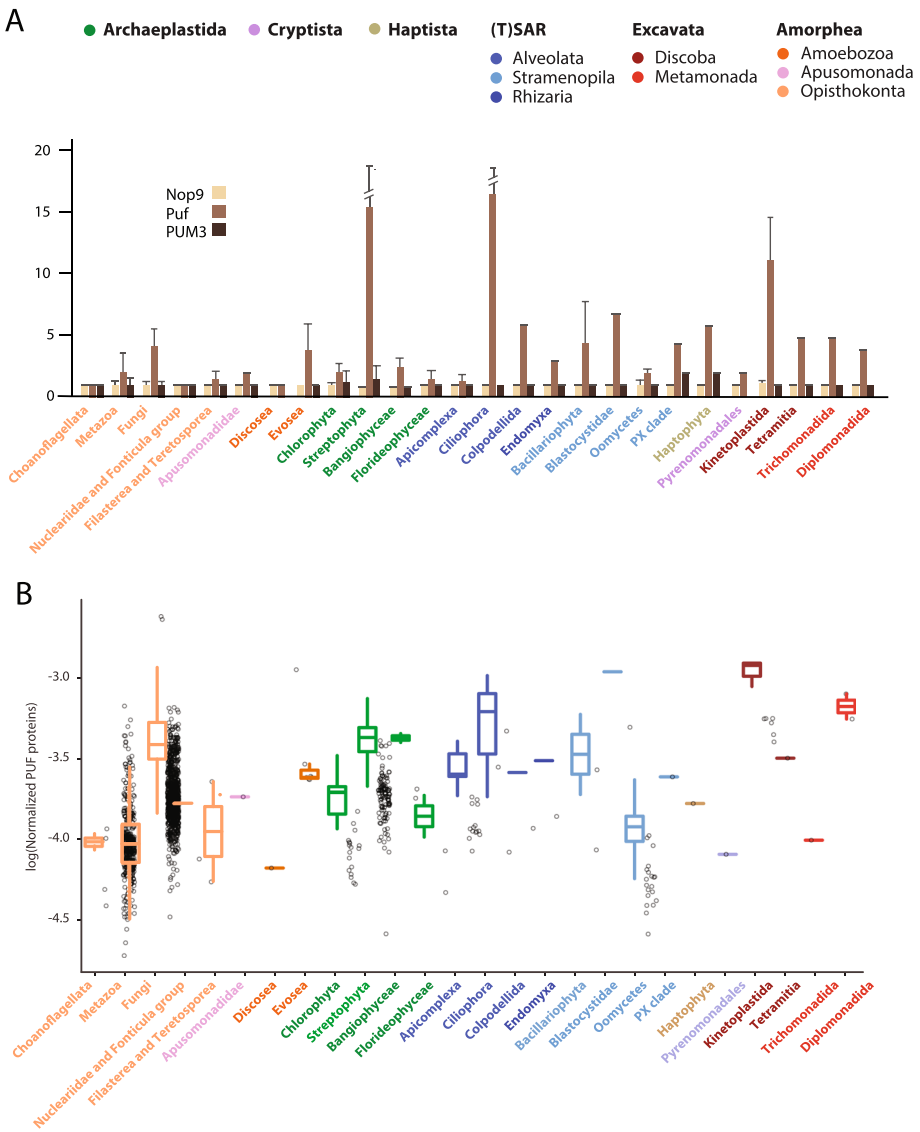


Fig. 3 Three types of PufSF proteins in major groups of eukaryotes. **a** The number of Nop9, Puf, and PUM3 orthologues was identified for each proteome in the dataset, and the values were averaged for the taxonomic group of eukaryotes. The error bars correspond to the standard deviation of variance of values within the particular taxonomic group. Error bars for Streptophyta and Ciliophora were cut for better visualization. **b** The number of Pufs normalized with respect to the total number of the protein coding; colour-codes correspond to the taxonomic affiliation to the eukaryotic supergroups

the pool of putative target transcripts seem to be very similar across eukaryotes (Fig. 3b) with the average number of one Puf for every 3.47×10^4 protein-coding genes.

Phylogenetic reconstruction of PufSF proteins

In order to get insight into the evolution of PufSF proteins, we performed the phylogenetic analyses on a dataset containing all three types of the proteins or the just a particular subset of either Puf, PUM3, or Nop9 orthologues (see the 'Materials and methods' section for more details). While the phylogenetic reconstructions proved to be problematic due to the repetitive structure of PufSF proteins, the overall tree (Fig. 4, Additional File 3: Fig. S1) shows three distinct clades corresponding to Puf, PUM3, and Nop9. Subsequent subtrees of PUM3 and Nop9 proteins, which are present as single proteins, resolved all major eukaryotic groups with some unexpected position of orthologues mainly from the Metamonada supergroup (Additional File 3: Fig. S1), most likely caused by their high sequence divergence.

Given the presence of several Pufs in most eukaryotes, we endeavoured to resolve their evolutionary relationships and specifically if the presence of multiple Pufs reflects the ancestral state of LECA or rather they are independent paralogues arisen by linage-specific gene duplication(s). The phylogenetic reconstructions of Pufs

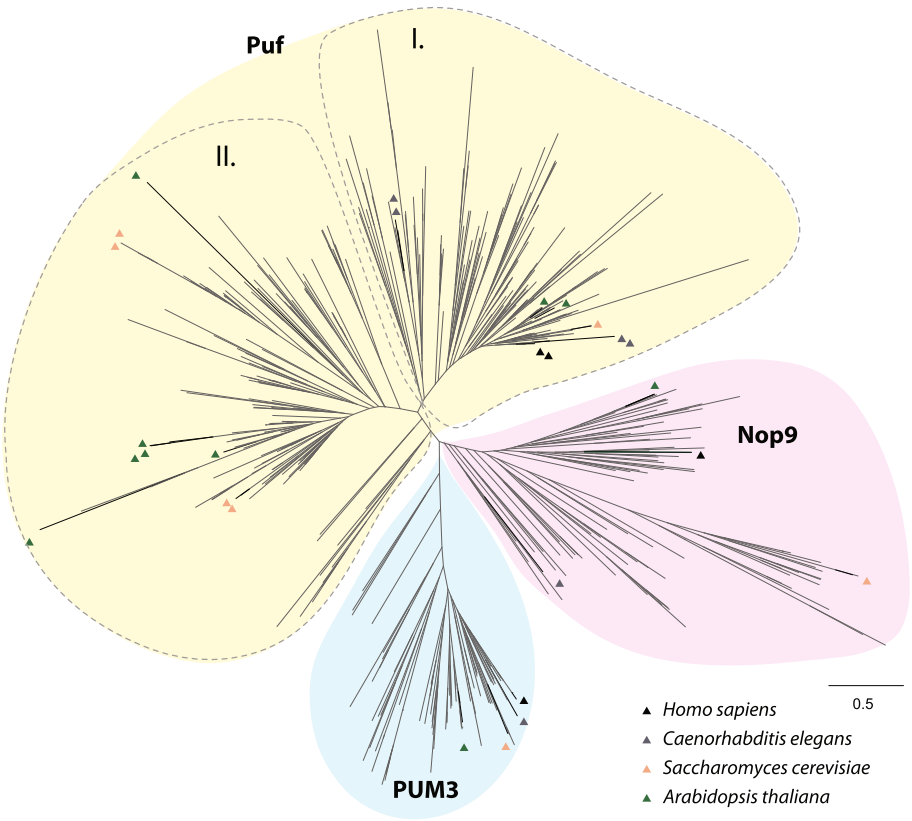


Fig. 4 Phylogenetic analysis of PufSF proteins. Maximum likelihood phylogenetic inference of PufSF proteins with sequences from *H. sapiens*, *C. elegans*, *S. cerevisiae*, and *A. thaliana* is shown as triangles with the indicated colours. For visualization purposes, support values were removed. Full phylogenies of the PUM3, Nop9, and Puf proteins can be found in Additional file 3: Fig. S1

remained very problematic, and despite using distinct alignment strategies, they did not return clear separation among individual Pufs and the eukaryotic taxa (Additional File 3: Fig. S1). However, we could identify two clear groups (labelled 'I' and 'II') (Fig. 4) that encompass the vast majority of eukaryotic taxa. This possibly reflects the presence of only two Pufs in LECA. Within group I and group II, there have been a number of lineage-specific duplications giving rise to the multitude of Pufs seen in different genomes.

PufSF proteins in the metamonad *G. intestinalis*

To test whether PufSF proteins are conserved in some of the most divergent eukaryotes, we specifically investigated the presence of PufSF proteins in *G. intestinalis*. Using a variety of sensitive sequence searching strategies (see the 'Materials and methods' section), we identified six proteins in *G. intestinalis*. The position and number of Puf repeats within the domain were predicted using HHpred and the alignments with the structurally characterized classical Pufs, Nop9, or PUM3 proteins, respectively. The classification of the proteins into the three types was confirmed by comparison with the domains defined at InterPro. Based on these classifications, we identified four *G. intestinalis* Puf homologues (GiPuf1–GiPuf4), one Nop9 (GiNop9), and PUM3 (GiPUM3) homologue (Fig. 5a). All four *G. intestinalis* Pufs were predicted to contain eight Puf repeats corresponding to eight TRMs (Fig. 5b), while both Nop9 and PUM3 homologues contained 11 Puf repeats similar to their homologues in *Saccharomyces cerevisiae* (Additional File 4: Fig. S2 and S3). The prediction of TRMs was performed by HHpred against the structurally characterized orthologues from *S. cerevisiae* and *D. melanogaster* (PDB ACNO. 5BZ1and 5KLA). However, the obtained TRMs for more divergent GiPuf1 and GiPuf2 were not in full agreement with the protein sequence alignment containing other *G. intestinalis* Puf proteins (Fig. 5b) as their Puf domain appeared shifted by two Puf repeats towards the C-terminus. At present, it is difficult to resolve if just the two Puf repeats were re-arranged or the entire domain was modified in these two proteins.

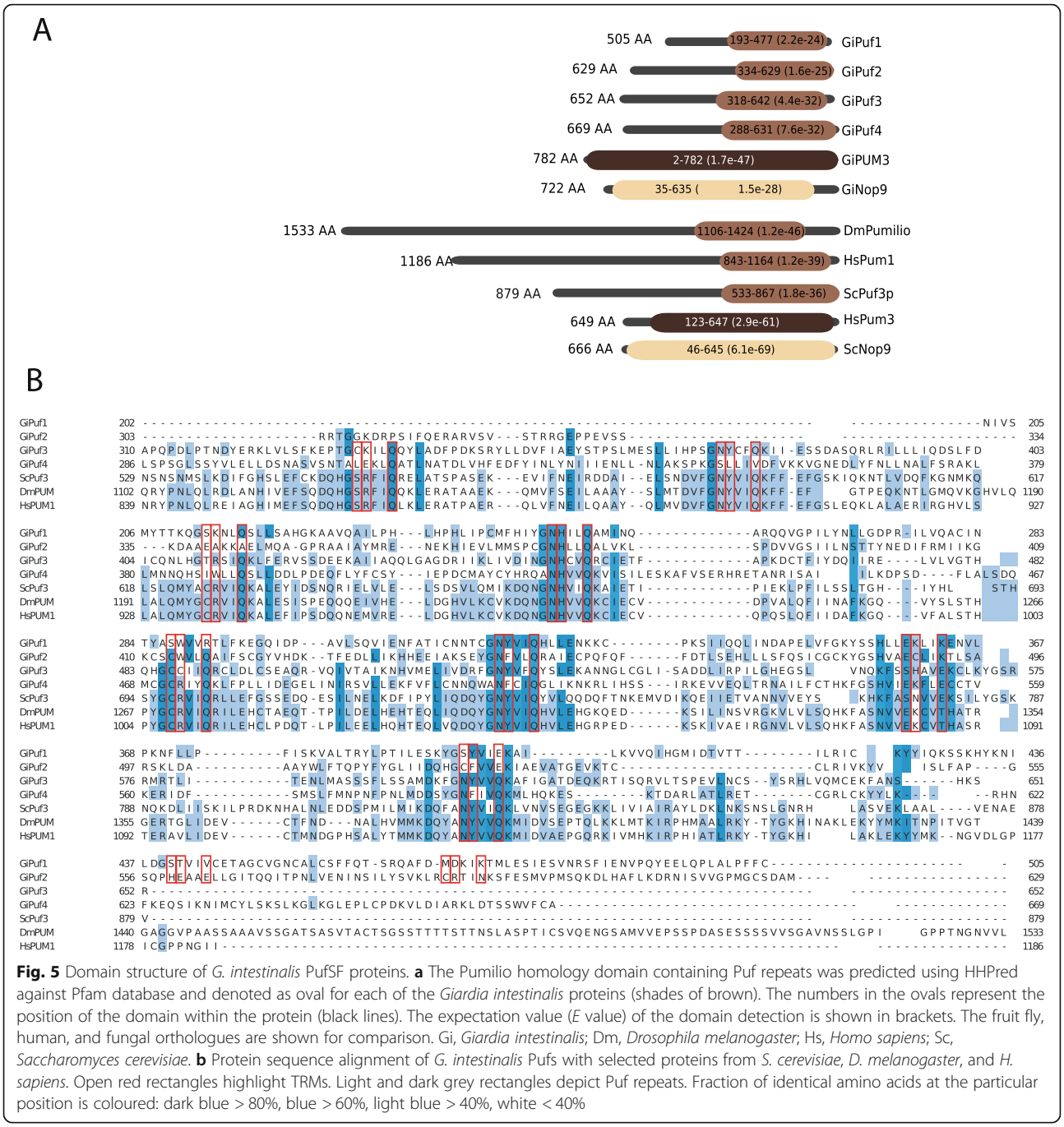


Fig. 5 Domain structure of *G. intestinalis* PufSF proteins. **a** The Pumilio homology domain containing Puf repeats was predicted using HHPred against Pfam database and denoted as oval for each of the *Giardia intestinalis* proteins (shades of brown). The numbers in the ovals represent the position of the domain within the protein (black lines). The expectation value (*E* value) of the domain detection is shown in brackets. The fruit fly, human, and fungal orthologues are shown for comparison. Gi, *Giardia intestinalis*; Dm, *Drosophila melanogaster*; Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*. **b** Protein sequence alignment of *G. intestinalis* Pufs with selected proteins from *S. cerevisiae*, *D. melanogaster*, and *H. sapiens*. Open red rectangles highlight TRMs. Light and dark grey rectangles depict Puf repeats. Fraction of identical amino acids at the particular position is coloured: dark blue > 80%, blue > 60%, light blue > 40%, white < 40%

According to the phylogenetic reconstruction, the four *G. intestinalis* Pufs grouped together with the orthologues from closely related diplomonad species *Spironucleus salmonicida* and *Trepomonas* sp. distributed in group I and group II (Additional File 3: Fig. S1). GiPuf3, the most conserved Puf homologue of *G. intestinalis*, branched with group I Puf proteins while GiPuf1, GiPuf2, and GiPuf4 affiliated with the proteins from group II (Additional File 3: Fig. S1). The latter proteins thus likely represent lineage-specific gene duplications.

Cellular localization of *G. intestinalis* PufSF proteins

In general, PufSF proteins localize to the nucleus or cytosol. In the cytosol, Puf proteins often associate with the cytoplasmic face of cellular compartments [1].

To test the cellular localizations of each PufSF protein in *G. intestinalis*, we explored bioinformatic and experimental strategies. For bioinformatic predictions, we used DeepLoc [39], which uses neural networks to assess the localization of proteins based on a training set of experimentally localized proteins on UniProt. This algorithm

predicted a cytoplasmic localization for all four *G. intestinalis* Pufs and Nop9 homologue and nuclear localization for only GiPUM3 (Fig. 6a). The cytosolic localization of the Pufs and the nuclear localization of GiPUM3 are in agreement with the expected roles of PufSF proteins, which control the stability and the localization of mRNAs in the cytosol and the nucleolar processing of 7S rRNA, respectively [2, 3]. However, given the role of Nop9 proteins in the maturation of pre-18S rRNA, the protein is expected to be in the nuclear compartment.

To test these subcellular localization predictions, we expressed all the *G. intestinalis* PufSF proteins with a C-terminal BAP (biotin acceptor peptide) tag in *G.*

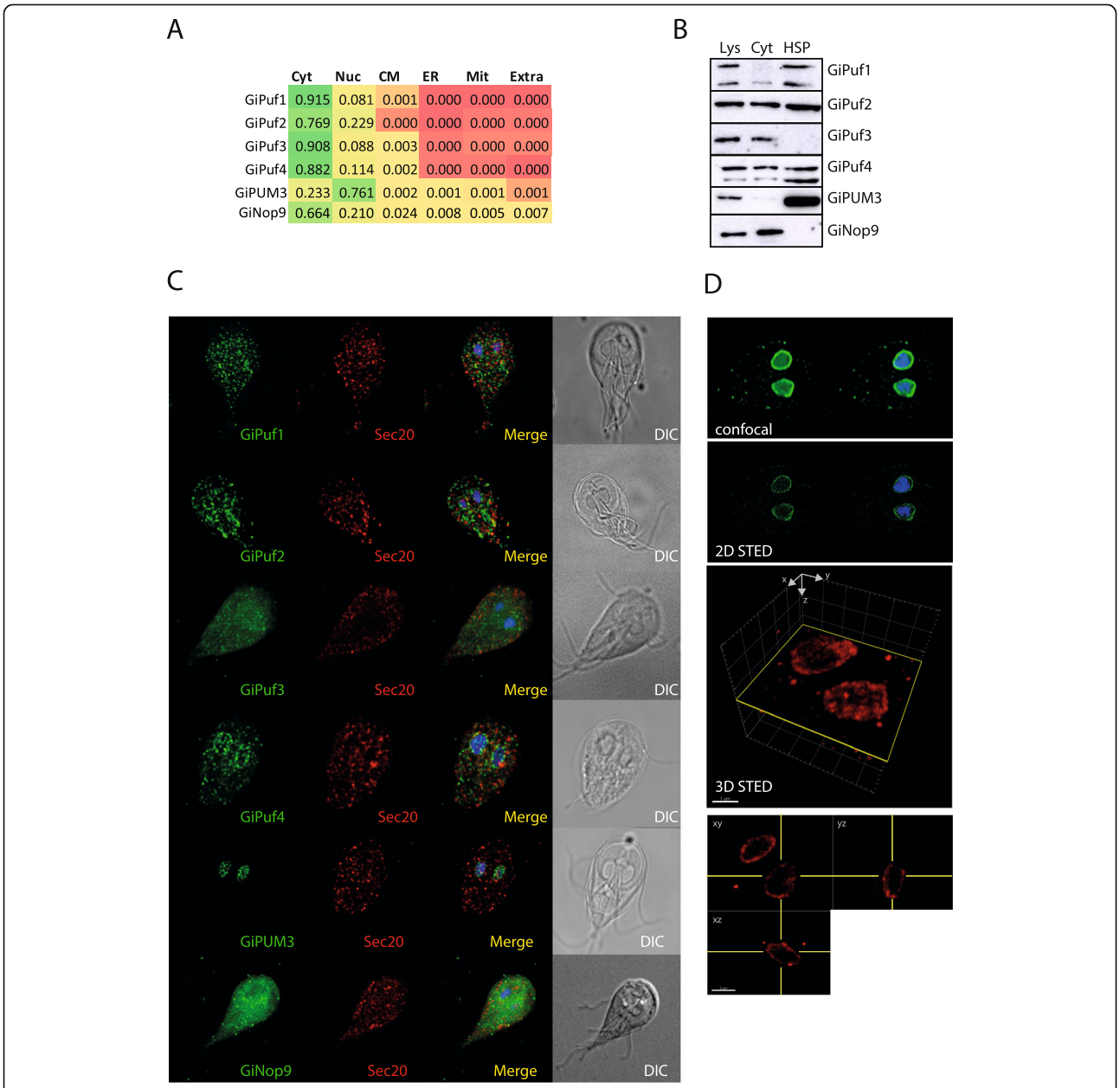


Fig. 6 Cellular localization of *G. intestinalis* Puf and Nop9 proteins. **a** The scores obtained by DeepLoc prediction indicate cytosolic localization of all but GiPUM3 protein, which is predicted as nuclear protein. Colouring gradient represents the values from 0 (red) to 1 (green). Cyt, cytosol; Nuc, nucleus; CM, cytoplasmic membrane; ER, endoplasmic reticulum; Mit, mitochondrion; Extra, extracellular. **b** Western blot analysis of *G. intestinalis* expressing BAP-tagged PufSF proteins (one of at least three independent cell experiments is show). SDS-PAGE and immunoblots show total lysate (Lys), cytosolic (Cyt), and high-speed pellet (HSP) fraction. **c** Immunofluorescence analysis of the same cell lines shows cellular localization of the proteins. PufSF proteins in green, Sec20-endomembrane system marker in red. DIC, Differential Interference Contrast. **d** Detailed imaging of GiPUM3 by confocal and 2D STED (GiPUM3 in green, DNA in blue). 3D STED of GiPUM3 with the orthogonal projections (GiPUM3 in red)

intestinalis and analysed their localizations with cell fractionation and microscopy using antibodies directed at the BAP tag. Expression of all constructs but one (GiPuf4) was highly unstable and diminished quickly after establishing stable cell lines. Therefore, *G. intestinalis* cell lysates were obtained as soon as possible and separated into two fractions: (i) the high-speed pellet, containing sedimentable membrane-bound organelles such as the nucleus, endoplasmic reticulum, peripheral vacuoles, or mitosomes, and (ii) the cytoplasmic fraction [40] (Fig. 6b). In general, the Puf proteins showed three different types of distribution: GiPuf3 and GiNop9 were found specifically in the cytosolic fraction, while GiPuf1 and GiPUM3 were present predominantly in high-speed pellet fraction. Finally, GiPuf2 and GiPuf4 showed the presence in both cytosolic and high-speed pellet fractions indicating their partial association with the cellular membranes.

In agreement with the western blot analyses, the immunofluorescence confocal microscopy of GiPuf3 and GiNop9 showed mainly cytosolic localization of the proteins with some punctate distribution in the cell that do not co-localize with the endomembrane marker Sec20 (Fig. 6c). While the data are in agreement with the bioinformatic prediction, the cytosolic presence of *G. intestinalis* Nop9 homologue remains puzzling. GiPuf1, GiPuf2, and GiPuf4 were present in different kinds of vesicular structures likely corresponding to specific regions of the endomembrane system, which however did not co-localize, with our endomembrane marker protein Sec20 (Fig. 6c). In addition, GiPuf2 and GiPuf4 showed also a perinuclear staining, which indicated that the protein is associated with the nuclear membrane. Conversely, a very specific labelling of two *G. intestinalis* nuclei was observed for GiPUM3. In other eukaryotes, PUM3 localizes to discrete nucleolar spots in the nuclear matrix [3, 10]. To determine the subnuclear localization of GiPUM3, we performed high-resolution STED microscopy. In both 2D and 3D STED microscopy, we observed GiPUM3 localizing to the periphery of the nucleus (Fig. 6d).

To determine potential Puf-interacting proteins in *G. intestinalis*, we explored the Puf-interactome using a high-resolution proximity labelling coupled to mass spectrometry. By determining potential interaction partners of the *G. intestinalis* Pufs, we could better predict their involvement in gene expression control. Unfortunately, the expression of tagged Pufs was highly unstable and diminished quickly after the cell transformation and we thus could not perform larger scale experiments required for protein- or RNA-pull down experiments. We were, however, able to generate a cell line weakly expressing BAP-tagged GiPuf4 in the presence of cytosolic biotin ligase BirA [41]. Upon crosslinking and purification of GiPuf4 on streptavidin-coupled Dynabeads, the triplicate samples were analysed by mass spectrometry. The purified GiPuf4 was found to be specifically enriched in our sample, although the experiment did not reveal any specific interacting partner protein above the statistical threshold (Additional file 5: Fig. S4, Additional file 6: Table S3). Thus, any functional predictions could not be drawn at this stage.

Prediction of binding motifs of *G. intestinalis* Pufs and their target mRNA

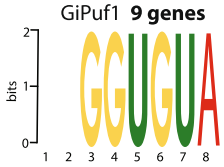
While we could not identify the interacting factors for *G. intestinalis* Pufs by mass spectrometry, we decided to predict the sets of recognized mRNAs for each homologue. The RNA sequence motif recognized by Puf/Nop9 proteins is determined by the combination of three amino acid residues, referred to as tripartite recognition motif (TRM). TRM is part of five residues in the second α-helix of each Puf repeat represented as 1-2-X-X-5 (where X is any hydrophobic residues). Within the TRM, positions 1 and 5 bind the edge of the RNA base, while the position 2 makes a stacking interaction with RNA molecule [42]. Some TRMs have been shown to be specific for particular base [11]. Hence, upon the identification of the TRMs in each Puf repeat, it is possible to predict its sequence-specific binding properties [9]. However, it should be noted that for some naturally occurring TRMs, the specificity has not been determined. By comparing the most closely related sequences to each *G. intestinalis* Puf identified with HHpred, we predicted the putative binding motif by manually checking the position of individual Puf repeats (Fig. 5b), and we could predict putative RNA-binding motifs for all *G. intestinalis* Pufs (Fig. 7a). Several predicted TRMs located in GiPuf1, GiPuf2, and GiPuf4 contained experimentally unidentified amino acid combinations which left these putative binding motifs incomplete. Interestingly, a complete binding motif predicted for GiPuf3 (5′-UGUAUUUA-3′) was found to be highly similar to 5′-UGUAUAUA-3′motif of prototypical members of the protein family such as human PUM1 or yeast Puf3 [43].

The predicted motifs of *G. intestinalis* Pufs were then used to search the dataset of theoretical 3′-UTR of all 9747 *G. intestinalis* genes retrieved from GiardiaDB. Given that the 3′-UTRs of *G. intestinalis* mRNAs are very short [30, 32, 44], the length of the UTRs was limited to 50 bases only.

Using the FIMO (Find Individual Motif Occurrence) algorithm [45], a specific set of possible cognate mRNAs for GiPuf1–GiPuf4 was retrieved (Fig. 7, Additional file 8: Table S5). Each of the *G. intestinalis* Puf proteins was predicted to interact with the different number of transcripts, and this number was also inversely proportional to the *G. intestinalis* length of the predicted binding

A

GiPuf1 **9 genes**



*	GL50803_17350	Hypothetical protein
	GL50803_114442	Hypothetical protein
	GL50803_7272	GARP-like protein 1
	GL50803_6334	Hypothetical protein
	GL50803_95023	Hypothetical protein
	GL50803_86511	Acyl-CoA synthetase
	GL50803_8460	Hypothetical protein
	GL50803_26228	Hypothetical protein
	GL50803_24537	Spindle pole protein

GiPuf2 **24 genes**

GL50803_106320	Zinc finger domain protein
GL50803_16862	Kinase, NEK
GL50803_33020	hypothetical protein
GL50803_9034	hypothetical protein
GL50803_113269	VSP
GL50803_15347	hypothetical protein
GL50803_7800	hypothetical protein
GL50803_14328	hypothetical protein
GL50803_11624	hypothetical protein
GL50803_103818	hypothetical protein
GL50803_117981	hypothetical protein
GL50803_6988	hypothetical protein
GL50803_4692	hypothetical protein
GL50803_10698	Phosphorylase kinase
GL50803_10524	hypothetical protein
GL50803_6918	DNA damage repair protein
GL50803_7556	Lysosomal acid phosphatase
GL50803_36600	hypothetical protein
GL50803_8382	Short-chain dehydrogenase
GL50803_15046	Ribosomal protein L26
GL50803_7532	Vacuolar ATP synthase
GL50803_95789	Deoxyribonuclease
GL50803_14742	Kinase, NEK
GL50803_10876	hypothetical protein

GiPuf3 **7 genes**

GL50803_115159	hypothetical protein
GL50803_16454	Protein kinase, putative
GL50803_97219	Fibrillarin-like pre-rRNA proc. p.
GL50803_89285	hypothetical protein
GL50803_14487	hypothetical protein
GL50803_23630	hypothetical protein
GL50803_6188	hypothetical protein

GiPuf4 **44 genes**

	GL50803_6530	Protein 21.1
	GL50803_16217	NAG pyrophosphorylase
	GL50803_6592	hypothetical protein
	GL50803_36515	hypothetical protein
	GL50803_19677	hypothetical protein
	GL50803_99071	hypothetical protein
	GL50803_41927	hypothetical protein
	GL50803_4959	Kinase, NEK
	GL50803_3147	Protein 21.1
	GL50803_14104	hypothetical protein
	GL50803_7760	Sentrin
*	GL50803_17350	hypothetical protein
	GL50803_11621	hypothetical protein
	GL50803_14375	Asparaginyl-tRNA synthetase
	GL50803_7018	hypothetical protein
	GL50803_6408	SnRNP Sm-like protein, putative
	GL50803_114307	Kinase, NEK
	GL50803_92246	Cation-transporting ATPase 2
	GL50803_8974	hypothetical protein
	GL50803_13851	hypothetical protein
	GL50803_221691	Histone methyltransferase
	GL50803_91643	26S proteasome subunit
	GL50803_26603	hypothetical protein
	GL50803_17143	Pyruvate kinase
	GL50803_90434	hypothetical protein
	GL50803_33301	hypothetical protein
	GL50803_3414	5-AMP-activated pr. kinase
	GL50803_8496	Rac/Rho-like protein
	GL50803_8917	Mu adaptin
	GL50803_4587	hypothetical protein
	GL50803_118786	VSP, putative
	GL50803_11521	VSP
	GL50803_103916	VSP, putative
	GL50803_23833	Vacuolar protein sorting 35
	GL50803_33232	GARP-like protein 4
	GL50803_13438	Protein 21.1
	GL50803_10231	TCP-1 chaperonin subunit zeta
	GL50803_98892	hypothetical protein
	GL50803_113723	hypothetical protein
	GL50803_13075	hypothetical protein
	GL50803_10773	hypothetical protein
	GL50803_28888	hypothetical protein
	GL50803_114636	hypothetical protein
	GL50803_7026	hypothetical protein

B

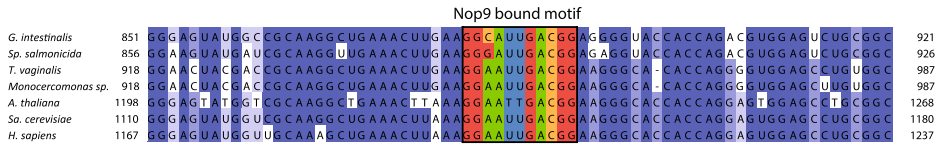


Fig. 7 (See legend on next page.)

(See figure on previous page.)
Fig. 7 Predicted binding motifs of *G.intestinalis* PufSF proteins. **a** The tripartite recognition motifs (TRMs) of each Puf repeats of *G.intestinalis* Pufs were predicted, and the resulting sequence was used to search the conceptual transcriptome. The number of putative mRNA targets, which contain the motif in the 3′-UTR, is shown in bold. Asterisk denotes the same putative mRNA recognized by two different Pufs. **b** Sequence alignment of 18S rRNA shows the conservation of the sequence recognized by Nop9 as it was experimentally identified for *S. cerevisae* Nop9. Fraction of identical nucleotides at the particular position is coloured: dark blue > 80%, blue > 60%, light blue > 40%, white < 40%

motif: where GiPuf3, GiPuf1, GiPuf2, and GiPuf4 were predicted to interact with 7, 9, 24, and 44 transcripts, respectively. These numbers are substantially lower than other Puf proteins that are predicted to interact with hundreds of RNA targets [19, 46]. We next explored the putative function and subcellular localization of the protein products of the predicted target transcripts (Additional file 8: Table S5). Interestingly, the target transcripts included other RNA-processing proteins (e.g. fibrillarin) and a component of the ERAD (endoplasmic reticulum-associated degradation) pathway (e.g. Derlin 1) (Additional file 8: Table S5).

Unlike Pufs, Nop9s use their 11 Puf repeats [21, 22] to specifically bind to the pre-18S rRNA at the central pseudoknot region [22] and regulate its processing possibly by competing with Nob1 nuclease [21]. However, how Nop9 TRMs interact with the rRNA was unknown until recently where it was shown that yeast Nop9 binds to a specific 11-nucleotide region of the pre-18S rRNA [22]. The alignment of 18S rRNA sequences showed that this region is very well conserved across eukaryotes including species from the Metamonada group (Fig. 6b) suggesting that Nop9 might also recognize this region. However, in *G. intestinalis*, the cytoplasmic (and not nuclear) localization of this protein challenges whether GiNop9 and rRNA processing occurs in the nucleus.

Finally, when GiPUM3 was aligned with its characterized human and yeast counterparts, no previously identified TRMs could be identified within the sequence. This supported the absence of recognizable binding motif for this type of PufSF proteins (Additional File 4: Fig. S3) [2, 23].

Discussion

RBPs stand at the centre of the regulation of eukaryotic protein expression by mediating a wide range of post-transcriptional processes. There are several distinctive molecular properties of PufSF proteins [4, 5]. The protein structure is composed of 8–11 units (Puf repeats), combination of which defines the sequence specificity towards their RNA targets. They have been shown to provide molecular basis for long sought processes such as site-specific translation of proteins, RNA degradation, or maturation of rRNA [1].

Here, we explored the evolution of PufSF proteins across the tree of eukaryotes. In addition to the originally characterized 'classical' Pufs [4, 5], which mediate the turnover of mRNAs via binding to their 3′-UTRs [19], recent studies in yeast model system have defined two additional types of PufSF proteins, which perform distinct cellular functions. Nop9 binds pre-18S rRNA during its maturation [3, 21, 22], and PUM3 orthologues are involved in processing 7S rRNA [2, 10, 47] and positioning of specific mRNAs [48].

Given that PufSF proteins are involved in the core processes of the eukaryotic cell (e.g. the maturation of rRNA and the translational control of mRNAs), we hypothesized that this system should be conserved across eukaryotic diversity. Indeed, we found that all majority lineages of eukaryotes encode at least one homologue of Nop9, PUM3, and two Puf proteins (Figs. 2, 3, and 4; Additional file 2: Table S2). The lack of PufSF proteins in prokaryotic genomes (including the closest prokaryotic relatives of eukaryotes—the Asgard archaea [49]) suggests that these proteins are in fact eukaryotic innovations. Moreover, their ubiquity across the tree of eukaryotes (including the highly diverged metamonads) suggests they were already present in protoeukaryotes before the emergence of LECA. To uncover the within protein family evolutionary history of the PufSF proteins, we performed a clustering analysis on the basis of sequence similarity. We found that PufSF proteins form three larger groups corresponding to the three independent biological functions performed by Puf, PUM3, and Nop9 proteins.

This suggests that the Puf and PUM3 proteins likely originated from the same ancestral gene. Interestingly, however, Nop9 and PUM3 proteins, despite belonging to separate groups of proteins within the family, share an array of 11 Puf repeats by which they interact with pre-18S [21] or 7S rRNA [2], respectively. On the basis of clustering analysis, it seems either that both Nop9 and PUM3 independently evolved into 11 Puf repeat containing proteins or that Pufs evolved from a larger 11-repeat domain ancestral protein into an 8-repeat domain protein by disposing of three Puf repeats. The latter seems to be more plausible as Pufs acquired additional long N-terminal part of mostly unknown function, which may, however, carry additional RNA-binding motifs [50].

As discussed above, it is most likely that the Nop9, PUM3, and two Puf proteins evolved during eukaryogenesis prior to the emergence of LECA, which was already indicated by the existence of four eukaryotic orthologous groups (KOGs) specific for PufSF [51]. The role of PufSF

proteins in both nuclear and cytoplasmic processes suggests that this protein family likely coincided with the evolution of a nucleic acid compartment; however, the exact timing cannot be determined with current data. Indeed, as protoeukaryotes began to complexify, it is tempting to speculate that functional networks necessary to manage an expanding genome and gene repertoire would demand additional levels of regulation—a role potentially filled by RBPs such as the PufSF proteins. Future examination of the relative timing [52] of the emergence of PufSF proteins and their experimental characterization in diverse eukaryotic lineages will be critical in assessing these hypotheses.

The classification of the protein family members also revealed that while the number of Pufs is highly variable across lineages, both Nop9 and PUM3 orthologues remained present as single copy genes with no paralogous sequences in the majority of eukaryotes. This can likely be attributed to the conserved role of the Nop9 and PUM3 proteins in the maturation of rRNA which is essential for the biogenesis of eukaryotic ribosome. Our analysis also demonstrated that the expansion of Pufs in some lineages of eukaryotes is similar to those documented before for *Arabidopsis thaliana* [25]. Interestingly, we showed that the actual number of Pufs in eukaryotic supergroups corresponds to the size of the proteome and therefore also the number of putative target mRNAs within the cell. Importantly, our estimates are based on the total number of genes present in a genome and will therefore be an underestimation of the total number of transcripts present in a cell resulting in an overestimated Puf to total gene ratio depicted in Fig. 3. However, in organisms such as *G. intestinalis* or yeast, which are intron-poor and lack methods for generating splice variants, the total number of transcripts is approximately equal to the total number of protein-coding genes. Taking this into account, the Puf to total gene ratio of *G. intestinalis* is one of the highest across eukaryotic diversity. This suggests that while the total number of transcripts might still be determinative of Puf number in other organisms, additional factors beyond transcript number seem to govern the persistence of multiple Pufs in *G. intestinalis*.

The unique cellular and molecular features of *G. intestinalis* and its specialization make it an ideal model to study the extreme limits of cell biology [53]. With respect to RNA metabolism, *G. intestinalis* is often the exception to the rule. Previous studies have demonstrated that *G. intestinalis* lacks core regulatory components of transcription [37] and has even repurposed some conserved RNA metabolic pathways [54]. Our analyses of PufSF proteins from *G. intestinalis* have added at least one piece to the puzzle of RNA metabolism where we were able to show that like other eukaryotes, *G. intestinalis* has retained genes encoding all the three core RBPs.

However, the initial characterization of the proteins has uncovered interesting deviations from the eukaryotic blueprint. *G. intestinalis* PUM3 orthologue (GiPUM3) is localized on the nuclear periphery, which is different from nucleolar localization found in all experimentally analysed organisms [10, 48, 55]. Similarly, the cytosolic localization of *G. intestinalis* Nop9 orthologue contradicts its expected role in pre-18S rRNA processing occurring in the nucleolus.

This raises a question on the nature and localization of *G. intestinalis* nucleolus. For long time, *G. intestinalis* has been thought to lack nucleolus-like structures since nucleolar markers have failed to show discrete nuclear labelling [56]. However, later hybridization of 18S rRNA and localization of fibrillarin, pseudouridine synthase, and snoRNA suggested specific distribution of the nucleoli on the periphery of both *G. intestinalis* nuclei [57–60]. It is therefore likely that the nuclear periphery functionally replaces nucleolar bodies typical of other eukaryotes [31].

Unexpectedly, we observed that GiNop9 localizes to the cytosol of *G. intestinalis*. We have a number of hypotheses that could explain this observation. First, we cannot rule out that the overexpression of the protein resulted in mistargeting of *G. intestinalis* Nop9, although we should be able to observe at least partial nuclear signal. Alternatively, GiNop9 does not participate in pre-18S rRNA processing in *G. intestinalis* or its cytosolic presence is a consequence of unique nucleolar localization in *G. intestinalis*. Nop9 is thought to prevent rRNA processing by cytosolic Nob1 nuclease by competing for binding at the same rRNA sequence [21]. Nob1 mediates the late step of pre-rRNA processing which generates 18S rRNA from 20S intermediate rRNA [61], and these interactions thus occur at the interface between the nucleus and cytosol. It is possible that in *G. intestinalis*, the relocation of the nucleolus to the nuclear periphery also relocated the latter steps of rRNA processing outside the nucleus whereby GiNop9 participates in the cytosolic maturation of the ribosome.

The bioinformatic survey for putative target mRNAs, whose 3′-UTR are bound by one of four *G. intestinalis* Pufs, returned only a small set of candidate genes compared to other eukaryotes [8]. This might stem from the extremely short UTRs in *G. intestinalis* [32, 37], which limits the actual region of interaction with RBPs. We could only confidently predict a complete binding motif (5′- UGUAUUUA-3′) for GiPuf3, and in this case, the motif was found to be highly similar to binding motifs of Pufs from other lineages of eukaryotes. The conservation of this binding motif in *G. intestinalis* was surprising given that *G. intestinalis* usually lacks many other core

regulatory elements when compared to model eukaryotes (e.g. TATA box or Inr element) [62, 63]. Given the theoretical capabilities of Pufs to bind any possible RNA sequence [64], it raises an important question on the functional advantage of this particular binding motif in RNA-protein interactions.

Obviously, there are many unknowns left concerning the function of PufSF proteins especially in non-model organisms like *G. intestinalis* and other protists. However, despite numerous differences in posttransriptional processes found among eukaryotes, this work shows that PufSF proteins have constituted key RNA-processing proteins since LECA.

Conclusions

In this study, we show that all three types of PufSF proteins are found across eukaryotic diversity. Moreover, we found that their role in both rRNA maturation (Nop9, PUM3) and mRNA translational control (Pufs) is conserved in all major lineages of eukaryotes and that they were present in the last eukaryotic common ancestor (LECA). Finally, we experimentally show the presence of three types of PufSF family proteins in a Metamonad *G. intestinalis* and highlight their several intriguing lineage-specific adaptations.

Materials and methods

Bioinformatic analyses

For the clustering analyses, the entry dataset of PufSF proteins was obtained from InterPro/UniProt database. All duplicates and sequences shorter than 500 amino acids were filtered out. Clustering analysis with CLANS was performed using MPI Bioinformatics toolkit available at https://toolkit.tuebingen.mpg.de/ [38]. The resulting 2D coordinates were plotted in GraphPad, and the dots representing the protein sequences were colour-coded according to current taxonomic descriptors of eukaryotes [65]. To identify putative Puf homologues in *G. intestinalis*, the protein alignment of PufSF was used in HHpred search (available at https://toolkit.tuebingen.mpg.de/#/tools/hhpred) against conceptual *G. intestinalis* proteome (Giardia_lamblia_ATCC_50803_31_Aug_2017) [66]. For the detection of Puf-binding motifs in 3′-UTRs, the motifs were generated by MEME (Multiple Em for Motif Elicitation available at http://meme-suite.org/tools/meme) [67] and used by FIMO (Find Individual Motif Occurrences) program (available at http://meme-suite.org/tools/fimo) [45] against our custom database of all conceptual 50-nt-long 3′-UTRs (obtained from GiardiaDB http://giardiadb.org/giardiadb/) [68]. Dataset of 18S rRNA was obtained from the high-quality ribosomal RNA database SILVA (available at https://www.arb-silva.de/) [69]. Sequences were aligned using mafft –auto [70], and alignment was coloured according to percentage identity in Jalview [71]. The proteins co-purified with GiPuf4 and identified by mass spectrometry were analysed by HHpred for homology detection and by DeepLoc (available at http://www.cbs.dtu.dk/services/DeepLoc/) for subcellular localization [39].

Classification of PufSF proteins in eukaryotes

To assess the distribution of Nop9, PUM3, and Puf sequences across the tree of eukaryotes, we used the precomputed domain annotation on InterProScan (ftp://ftp.ebi.ac.uk/pub/databases/interpro/protein2ipr.dat.gz) of all eukaryotes present in the Reference Proteome database from UniProtKB. Importantly, since we were interested in understanding how many gene expansions occurred in each genome, we ignored splice variants by selecting only one representative protein per variant. Proteins were classified based on the presence of the following IPR domains/PANTHER family/SMART domain: Nop9 (IPR040000/PTHR13102), PUM3 (IPR040059/PTHR13389), and Puf (IPR001313/SM00025).

Phylogenetic dataset construction and analyses

For phylogenetic dataset construction, a subset of representative taxa across eukaryotic diversity were selected from the entry dataset from the clustering analysis. Due to the variety of different protein lengths, only the region corresponding to the Puf repeats was analysed; for each sequences, we extracted the region corresponding to the start position of the first, until the last position of the final, Puf repeat using HMMsearch with the SMRT profile (SM00025). Sequences were aligned using mafft-linsi [70], and ambiguously aligned sites were removed using trimal with the '-gappyout' option [72]. Initial phylogenetic tree of the entire PufSF (i.e. Puf, Nop9, and PUM3 proteins) was generated using FastTree with the -lg option [73]. This tree was manually examined, and in the cases where an organism had multiple copies of a paralogue that were monophyletic, only one sequence was retained. This refined dataset was realigned as above. Maximum likelihood inference of the reduced dataset was performed using IQ-TREE v 2.0 [74] using the LG + G + F model. Given the large size of the dataset, we did not perform model testing with the mixture models for the PufSF tree. A total of 1000 SH-aLRTs (SH-like approximate likelihood ratio test; -alrt 1000), 1000 ultrafast bootstraps (-bb 1000) [75], 100 non-parametric boostraps (-b 100), and the transfer bootstrap expectation calculation (--tbe) [76] were performed and mapped onto the best-scoring maximum likelihood tree. To refine the phylogeny of Nop9, PUM3, and Puf proteins, we analysed each subclade separately. The sequences were aligned as above. Maximum likelihood inference of the reduced dataset was performed using IQ-TREE v 2.0

[74] under the best-scoring model of evolution selected by ModelFinder supplemented with the C-series mixture models using the -mset option [77]. A total of 1000 SH-aLRTs (SH-like approximate likelihood ratio test; -alrt 1000), 1000 ultrafast bootstraps (-bb 1000) [75], were mapped onto the best-scoring maximum likelihood tree. The resulting maximum likelihood tree was used as the guide tree (-ft) for rapid approximation of posterior mean site frequency (PMSF) of the C-series of mixture models [78] to generate 100 non-parametric bootstraps (-b 100) together with the transfer bootstrap expectation (--tbe). Alignment features and model parameters are specified in Additional file 7: Table S4, and datasets are available on the figshare repository (DOI:https://doi.org/10.6084/m9.figshare.12097692), see 'Availability of Data and Materials'.

Cell culture, cloning, and transfection

The *G. intestinalis* strain WB (ATCC 30957) was grown in TYI-S-33 medium [79] supplemented with 10% heat-inactivated bovine serum, 0.1% bovine bile, and antibiotics at 37 °C. The genes encoding *G. intestinalis* PufSF proteins [GiardiaDB accession numbers GL50803_17262 (GiPuf1), GL50803_17590 (GiPuf2), GL50803_17325 (GiPuf3), GL50803_4548 (GiPuf4), GL50803_16602 (GiPUM3), and GL50803_14117 (GiNop9)] and 150-bp-long 5′-UTR of GiPuf1, GiPuf2, GiPuf3, GiPuf4, and GiNop9 were amplified from genomic DNA using PCR and inserted to the plasmid pOndra [40] with the C-terminal biotin acceptor peptide (BAP). Additional file 9: Table S6 in lists all primers used in this study. *Escherichia coli* biotin ligase (BirA) in pTG vector was used for the co-expression with GiPuf4 [80]. *G. intestinalis* transfection was performed as previously described [40]. Briefly, 1×10^7 cells were electroporated with a Bio-Rad Gene Pulser using an exponential protocol ($U = 350$ V; $C = 1{,}000$ μF; $R = 750$ Ω). The transfected cells were grown in medium supplemented with antibiotics (57 μg/ml puromycin and 600 μg/ml G418).

Cell fractionation

Trophozoites of *G. intestinalis* in logarithmic growth were harvested in ice-cold phosphate-buffered saline (PBS, pH 7.4) at 1000×*g* at 4 °C for 10 min, washed in SM buffer (20 mM MOPS, 250 mM sucrose, pH 7.4), and collected by centrifugation at 1000×*g*, for 10 min at 4 °C. The pellet of the cells was resuspended in SM buffer supplemented with protease inhibitors (Roche). The cells were lysed on ice by sonication with 1-s pulses and 40% amplitude for 2 min (Sonicator ultrasonic processor Q125, Qsonica). The lysate was subjected to centrifugation at 2680×*g*, for 20 min at 4 °C to sediment nuclei, cytoskeleton, and remaining unbroken cells. The supernatant was subjected to centrifugation at 180,000×*g*, for 30 min at 4 °C. The resulting supernatant corresponded to cytosolic fraction, and the high-speed pellet (HSP) contained organelles including the mitosomes and the endoplasmic reticulum.

Crosslinking, protein isolation, and mass spectrometry (MS)

G. intestinalis cells were grown in TYI-S-33 medium enriched with 50 μM biotin for 24 h before harvesting. The cells were lysed by sonication, and the cell lysate (25 mg of total protein) was used for the protein isolation. The sample was diluted to final protein concentration of 3 mg/ml in PBS (pH 7.4) and supplemented with protease inhibitors (Roche). Crosslinker DSP (dithiobis [succinimidyl propionate]; Thermo Scientific) was added to the final concentration of 200 μM, and the sample was incubated 1 h on ice. The reaction was stopped by adding 50 mM final concentration of Tris (pH 8.0) for 15 min at room temperature. The sample was diluted 5 times with the boiling buffer (50 mM Tris, 1 mM EDTA, 1% SDS, pH 7.4), incubated at 80 °C for 10 min, and collected by centrifugation at 30,000×*g*, for 10 min at room temperature. Supernatant was diluted 1:10 in the binding buffer (50 mM Tris, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, pH 7.4). Meanwhile, 100 μl of streptavidin-coupled magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen) was washed 3 times in the binding buffer using a magnetic stand (according to the instructions of the manufacturer). The beads were mixed with the supernatant, and the sample was incubated for 16 h at 4 °C with gentle rotation. The magnetic beads were washed 3 times in the binding buffer with 0.1% sodium deoxycholate (SDC) for 5 min, once in boiling buffer for 5 min, once in washing buffer (60 mM Tris, 2% SDC, 10% glycerol) for 5 min, and finally twice in 100 mM triethylammonium bicarbonate (TEAB) with 0.1% SDC for 5 min. One tenth of the beads was used to analyse the efficiency of the procedure. Specifically, the sample was mixed with the SDS-PAGE sample buffer containing 20 mM biotin and incubated at 95 °C for 5 min to elute the proteins from the beads. The samples were resolved on 12% SDS-PAGE and transferred to nitrocellulose membrane. For immunodetection, the following antibodies were used: 1° anti-BAP (1:1000) (Genscript) and 2° anti-rabbit polyclonal antibody coupled to HRP (1:2000) (Sigma). Remaining beads with bound proteins were submitted to mass spectrometry analysis.

Tandem mass spectrometry (MS/MS) analyses

The beads were incubated at 5 mM tris (2-carboxyethyl) phosphine (TCEP) at 60 °C for 1 h to reduce the disulfide bridges and blocked with 10 mM final concentration of methyl methanethiosulfonate (MMTS) for 10 min at 37 °C. Samples were cleaved on the beads with 1 μg of

trypsin at 37 °C for 16 h. After digestion, samples were centrifuged, and supernatants were collected and acidified with trifluoroacetic acid (TFA at a final concentration of 1%). SDC was removed by extraction with ethylacetate [81]. Peptides were desalted on Michrom C18 column. Nano Reversed phase column (EASY-Spray column, 50 cm × 75 μm ID, PepMap C18, 2 μm particles, 100 Å pore size) was used for LC/MS analysis. Mobile phase buffer A was composed of water and 0.1% formic acid. Mobile phase B was composed of 0.1% formic acid in acetonitrile. Samples were loaded onto the trap column (Acclaim PepMap300, C18, 5 μm, 300 Å Wide Pore, 300 μm × 5 mm, 5 Cartridges) for 4 min at 15 μl/min. Loading buffer was composed of water, 2% acetonitrile, and 0.1% TFA. Peptides were eluted with mobile phase B gradient from 4 to 35% B in 1 h. Eluting peptide cations were converted to gas-phase ions by electrospray ionization and analysed on a Thermo Orbitrap Fusion (Q-OT-qIT, Thermo). Survey scans of peptide precursors from 400 to 1600 m/z were performed at 120 K resolution (at 200 m/z) with a 5×10^5 ion count target. Tandem MS was performed by isolation at 1.5 Th with the quadrupole, HCD fragmentation with normalized collision energy of 30, and rapid scan MS analysis in the ion trap. The MS 2 ion count target was set to 104, and the max injection time was 35 ms. Only those precursors with charge state 2–6 were sampled for MS 2. The dynamic exclusion duration was set to 45 s with a 10-ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was turned on. The instrument was run in top speed mode with 2-s cycles [82]. All data were analysed and quantified with the MaxQuant software (version 1.5.3.8) [83]. The false discovery rate (FDR) was set to 1% for both proteins and peptides, and we specified a minimum length of seven amino acids. The Andromeda search engine was used for the MS/MS spectra search against the *G. intestinalis* database (downloaded from UniProtKB in September 2017, containing 12,665 entries). Enzyme specificity was set as C-terminal to Arg and Lys, also allowing cleavage at proline bonds and a maximum of two missed cleavages. Dithiomethylation of cysteine was selected as fixed modification and N-terminal protein acetylation and methionine oxidation as variable modifications. The 'match between runs' feature of MaxQuant was used to transfer identifications to other LC-MS/MS runs based on their masses and retention time (maximum deviation 0.7 min), and this was also used in quantification experiments. Quantifications were performed with the label-free algorithms described recently [83]. Data analysis was performed using Perseus 1.5.2.4 software [84]. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE [85] partner repository with the dataset identifier PXD019608.

Immunofluorescence microscopy

G. intestinalis trophozoites were fixed in 1% paraformaldehyde for 30 min at 37 °C and collected by centrifugation at 1000×*g* for 5 min at room temperature. The cells were washed in PEM buffer (100 mM PIPES pH 6.9, 1 mM EGTA, and 0.1 mM $MgSO_4$) and placed on cover slips. The cells were permeabilized by 0.2% Triton X-100 (Sigma-Aldrich) for 20 min, washed three times with 1× PEM buffer, and incubated with primary antibodies in PEMBALG (100 mM PIPES pH 6.9, 1 mM EGTA, 0.1 mM $MgSO_4$, 1% BSA, 0.1% NaN_3, 100 mM lysine, and 0.5% cold-water fish skin gelatin) for 1 h. Cells were probed with the following primary antibodies: rat anti-HA monoclonal IgG antibody (1:1000 dilution), mouse anti-BAP monoclonal antibody (1:1000 dilution), and rabbit anti-Sec20 polyclonal [86] (1:1000 dilution). The cover slips were washed three times with 1 ml of 1× PEM and were incubated with secondary antibodies Alexa Fluor 488-conjugated goat anti-rat IgG and Alexa 594-conjugated donkey anti-rabbit IgG or Alexa594-conjugated goat anti-mouse IgG for 1 h. After 3 × 5 min washes in PEM buffer, slides were mounted with Vectashield containing DAPI (4′, 6-diamidino-2-phenylindole; Vector Laboratories) or Hoechst solution (33258). Stimulated emission depletion (STED) microscopy was performed on a commercial Abberior STED 775 QUAD scanning microscope (Abberior Instruments GmbH, Germany) equipped with Ti-E Nikon body, QUAD beam scanner, Easy3D STED Optics Module, and Nikon CFI Plan Apo Lambda objective (× 60 Oil, NA 1.40). Samples were illuminated by pulsed 561-nm and 640-nm lasers and depleted by a pulsed 775-nm STED laser of 2D donut shape (all lasers: 40 MHz repetition rate). Fluorescence signal was detected with single photon counting modules (Excelitas Technologies). Line-interleaved acquisition enabled separated detection of individual channels in spectral range from 605 to 625 nm and from 650 to 720 nm. The confocal pinhole was set to 1 AU.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12915-020-00814-3.

Additional file 1: Table S1. The clustering analysis of PufSF proteins.

Additional file 2: Table S2. Classification of PufSF proteins from UniProt referenced proteomes.

Additional file 3: Figure S1. Phylogenetic analysis NOP9, PUM3 and PUF proteins. Full phylogenies of maximum-likelihood analysis of PUM3, Nop9 and Puf proteins with SH-like approximate likelihood ratio test, ultrafast bootstrap supports, and with both non-parametric Felsenstein's Bootstrap Proportion (FBP) supports (i.e., PMSF) and Transfer Bootstrap

Expectation (TBE) supports as indicated. See Additional file 7 for alignment properties and model parameters.

Additional file 4: Figure S2. Proteins sequence alignment of Nop9 orthologues. **Figure S3.** Protein sequence alignment of PUM3 orthologues.

Additional file 5: Figure S4. Volcano plot of proteomics analysis of GiPuf4 pull down.

Additional file 6: Table S3. Proteomics analysis of GiPuf4 pull down.

Additional file 7: Table S4. Alignment properties and model parameters used for the phylogenetic analyses.

Additional file 8: Table S5. Prediction of *Giardia* Pufs target mRNA.

Additional file 9: Table S6. Primers used in the study.

Acknowledgements
We would like to thank Nikola Polanska for the help with the bioinformatic analyses.

Authors' contributions
PD, VN, and CS designed the experiments; VN, CS, LV, and MV performed the experiments; VN, CS, PD, LV, and MV analysed the data; NV, CS, and PD wrote the manuscript. All authors read and approved the final manuscript.

Funding
The project was supported by the GAUK grant 1071016 and PRIMUS grant PRIMUS/SCV34 from Charles University; by the Czech Science Foundation grant 20-25417S; by the National Sustainability Program II (Project BIOCEV-FAR, LQ1604) from the Ministry of Education, Youth and Sports of CR (MEYS); and by the project 'Centre for research of pathogenicity and virulence of parasites' (no. CZ.02.1.01/0.0/0.0/16_019/0000759) funded by the European Regional Development Fund. Microscopy was performed in the ICF co-financed by the European Regional Development Fund and the state budget of the Czech Republic, project nos. CZ.1.05/4.1.00/16.0347 and CZ.2.16/3.1.00/21515.

Availability of data and materials
All data generated or analysed during this study are included in this published article and its Additional files or deposited online. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE [85] partner repository with the dataset identifier PXD019608. Phylogenetic datasets are available on the figshare repository (DOI:https://doi.org/10.6084/m9.figshare.12097692) [87].

Ethics approval and consent to participate
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Parasitology, Faculty of Science, Charles University, BIOCEV, Průmyslová 595, 252 50 Vestec, Czech Republic. [2]Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123 Uppsala, Sweden.

Received: 16 December 2019 Accepted: 18 June 2020
Published online: 30 June 2020

References
1. Quenault T, Lithgow T, Traven A. PUF proteins: repression, activation and mRNA localization. Trends Cell Biol. 2011;21:104–12..
2. Qiu C, McCann KL, Wine RN, Baserga SJ, Hall TMT. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. Proc Natl Acad Sci. 2014;111:18554–9.
3. Thomson E, Rappsilber J, Tollervey D. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. Rna. 2007;13:2165–74.
4. Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, et al. A conserved RNA-binding protein that regulates sexual fates in the *C elegans* hermaphrodite germ line. Nature. 1997;390:477–84.
5. Barker DD, Wang C, Moore J, Dickinson LK, Lehmann R. Pumilio is essential for function but not for distribution of the *Drosophila* abdominal determinant Nanos. Genes Dev. 1992;6:2312–26.
6. Miller MT, Higgin JJ, Hall TM. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol. 2008;15:397–402.
7. Wang Y, Opperman L, Wickens M, Hall TM. Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. Proc Natl Acad Sci U S A. 2009;106:20186–91.
8. Hogan GJ, Brown PO, Herschlag D. Evolutionary conservation and diversification of Puf RNA binding proteins and their mRNA targets. PLoS Biol. 2015;13:e1002307.
9. Campbell ZT, Valley CT, Wickens M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat Struct Mol Biol. 2014;21:732–8.
10. Liang X, Hart KJ, Dong G, Siddiqui FA, Sebastian A, Li X, et al. Puf3 participates in ribosomal biogenesis in malaria parasites. J Cell Sci. 2018;131: jcs212597.
11. Zhang C, Muench DG. A nucleolar PUF RNA-binding protein with specificity for a unique RNA sequence. J Biol Chem. 2015;290:30108–18.
12. Crittenden SL, Bernstein DS, Bachorik JL, Thompson BE, Gallegos M, Petcherski AG, et al. A conserved RNA-binding protein controls germline stem cells in *Caenorhabditis elegans*. Nature. 2002;417:660–3.
13. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell. 2002;110:501–12.
14. Blewett NH, Goldstrohm AC. A eukaryotic translation initiation factor 4E-binding protein promotes mRNA decapping and is required for PUF repression. Mol Cell Biol. 2012;32:4181–94.
15. Nyikó T, Auber A, Bucher E. Functional and molecular characterization of the conserved *Arabidopsis* PUMILIO protein, APUM9. Plant Mol Biol. 2019; 100:199–214.
16. Goldstrohm AC, Hook BA, Seay DJ, Wickens M. PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol. 2006;13:533–9.
17. Suh N, Crittenden SL, Goldstrohm A, Hook B, Thompson B, Wickens M, et al. FBF and its dual control of gld-1 expression in the *Caenorhabditis elegans* germline. Genetics. 2009;181:1249–60.
18. Garcia-Rodriguez LJ, Gay AC, Pon LA. Puf3p, a Pumilio family RNA binding protein, localizes to mitochondria and regulates mitochondrial biogenesis and motility in budding yeast. J Cell Biol. 2007;176:197–207.
19. Gerber AP, Herschlag D, Brown PO. Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoSBiol. 2004;2:E79.
20. Pederson T. The nucleolus. Cold Spring Harb Perspect Biol. 2011;3:1–15.
21. Zhang J, McCann KL, Qiu C, Gonzalez LE, Baserga SJ, Hall TMT. Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. Nat Commun. 2016;7:13085.
22. Wang B, Ye K. Nop9 binds the central pseudoknot region of 18S rRNA. Nucleic Acids Res. 2017;45:gkw1323.
23. Li Z, Lee I, Moradi E, Hung NJ, Johnson AW, Marcotte EM. Rational extension of the ribosome biogenesis pathway using network-guided genetics. PLoS Biol. 2009;7.
24. Wickens M, Bernstein DS, Kimble J, Parker R. A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet. 2002;18:150–7.
25. Tam PPC, Barrette-Ng IH, Simon DM, Tam MWC, Ang AL, Muench DG. The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. BMC Plant Biol. 2010;10.
26. Galgano A, Forrer M, Jaskiewicz L, Kanitz A, Zavolan M, Gerber AP. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. PLoS One. 2008;3:e3164.
27. O'Malley MA, Leger MM, Wideman JG, Ruiz-Trillo I. Concepts of the last eukaryotic common ancestor. Nat Ecol Evol. 2019;3:338–44.
28. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol. 2019;66:4–119.
29. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The origin and diversification of mitochondria. Curr Biol. 2017;27:R1177–92.
30. Adam RD. Biology of *Giardia lamblia*. Clin Microbiol Rev. 2001;14:447–75.
31. Jiménez-García LF. The nucleolus of *Giardia lamblia*. MOJ Anat Physiol. 2017; 3:41–3.

32. Elmendorf HG, Singer SM, Nash TE. The abundance of sterile transcripts in *Giardia lamblia*. Nucleic Acids Res. 2001;29:4674–83.
33. Nixon JEJ, Wang A, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, et al. A spliceosomal intron in Giardia lamblia. Proc Natl Acad Sci U S A. 2002;99: 3701–5.
34. Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T. Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing. Curr Biol. 2011;21:311–5.
35. Li L, Wang CC. Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. J Biol Chem. 2004;279: 14656–64.
36. Hausmann S, Altura MA, Witmer M, Singer SM, Elmendorf HG, Shuman S. Yeast-like mRNA capping apparatus in *Giardia lamblia*. J Biol Chem. 2005; 280:12077–86.
37. Williams CW, Elmendorf HG. Identification and analysis of the RNA degrading complexes and machinery of *Giardia lamblia* using an *in silico* approach. BMC Genomics. 2011;12:586.
38. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004;20:3702–4.
39. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics. 2017;33:3387–95.
40. Dolezal P, Smíd O, Rada P, Zubácová Z, Bursać D, Suták R, et al. *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. Proc Natl Acad Sci U S A. 2005;102:10924–9.
41. Martincová E, Voleman L, Pyrih J, Žárský V, Vondráčková P, Kolísko M, et al. Probing the biology of *Giardia intestinalis* mitosomes using in vivo enzymatic tagging. Mol Cell Biol. 2015;35:2864–74.
42. Hall TMT. Expanding the RNA-recognition code of PUF proteins. Nat Struct Mol Biol. 2014;21:653–5.
43. Jarmoskaite I, Denny SK, Vaidyanathan PP, Becker WR, Andreasson JOL, Layton CJ, et al. A quantitative and predictive model for RNA binding by human Pumilio proteins. Mol Cell. 2019;74:966–81 e18.
44. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, et al. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science. 2007;317:1921–6.
45. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27:1017–8.
46. Kershner AM, Kimble J. Genome-wide analysis of mRNA targets for *Caenorhabditis elegans* FBF, a conserved stem cell regulator. Proc Natl Acad Sci U S A. 2010;107:3936–41.
47. Yang YT, Ting YH, Liang KJ, Lo KY. The roles of Puf6 and Loc1 in 60S biogenesis are interdependent, and both are required for efficient accommodation of Rpl43. J Biol Chem. 2016;291:19312–23.
48. Gu W, Deng Y, Zenklusen D, Singer RH. A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev. 2004;18:1452–65.
49. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Di B, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017;541:353–8.
50. Qiu C, Dutcher RC, Porter DF, Arava Y, Wickens M, Hall TMT. Distinct RNA-binding modules in a single PUF protein cooperate to determine RNA specificity. Nucleic Acids Res. 2019;47:8770–84.
51. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin E V., et al. The COG database: an updated vesion includes eukaryotes. BMC Bioinformatics. 2003;4:1–14.
52. Vosseberg J, van Hooff JJE, Marcet-Houben M, van Vlimmeren A, van Wijk LM, Gabaldón T, et al. Timing the origin of eukaryotic cellular complexity with ancient duplications. bioRxiv. 2019;:823484:1–18.
53. Ankarklev J, Jerlström-Hultqvist J, Ringqvist E, Troell K, Svärd SG. Behind the smile: cell biology and disease mechanisms of *Giardia* species. Nat Rev Microbiol. 2010;8:413–22.
54. Prucca CG, Slavin I, Quiroga R, Elías EV, Rivero FD, Saura A, et al. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. Nature. 2008; 456:750–4.
55. Andersen JS, Lyon CE, Fox AH, Leung AKL, Lam YW, Steen H, et al. Directed proteomic analysis of the human nucleolus. Curr Biol. 2002;12:1–11.
56. Narcisi EM, Glover CVC, Fechheimer M. Fibrillarin, a conserved pre-ribosomal RNA processing protein of *Giardia*. J. Eukaryot. Microbiol. 1998;45:105–11.
57. Jiménez-García LF, Zavala G, Chávez-Munguía B, Ramos-Godínez M d P, López-Velázquez G, Segura-Valdez M d L, et al. Identification of nucleoli in the early branching protist *Giardia duodenalis*. Int J Parasitol. 2008;38:1297–304.
58. Tian XF, Yang ZH, Shen H, Adam RD, Lu SQ. Identification of the nucleoli of *Giardia lamblia* with TEM and CFM. Parasitol Res. 2010;106:789–93.
59. Li L, Wang CC. Identification in the ancient protist *Giardia lamblia* of two eukaryotic translation initiation factor 4E homologues with distinctive functions. Eukaryot Cell. 2005;4:948–59.
60. Saraiya AA, Wang CC. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. PLOS Pathog. 2008;4:e1000224.
61. Lamanna AC, Karbsteina K. Nob1 binds the single-stranded cleavage site D at the 3′-end of 18S rRNA with its PIN domain. Proc Natl Acad Sci U S A. 2009;106:14259–64.
62. Teodorovic S, Walls CD, Elmendorf HG. Bidirectional transcription is an inherent feature of Giardia lamblia promoters and contributes to an abundance of sterile antisense transcripts throughout the genome. Nucleic Acids Res. 2007;35:2544–53.
63. Vanacova S, Liston DR, Tachezy J, Johnson PJ. Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. Int J Parasitol. 2003;33:235–55.
64. Filipovska A, Razif MF, Nygard KK, Rackham O. A universal code for RNA recognition by PUF proteins. Nat Chem Biol. 2011;7:425–7.
65. Burki F, Roger AJ, Brown MW, Simpson AGB. The new tree of eukaryotes. Trends Ecol Evol. 2020;35:43–55.
66. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018;430:2237–43.
67. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43:W39–49.
68. Aurrecoechea C, Barreto A, Basenko EY, Brestelli J, Brunk BP, Cade S, et al. EuPathDB: the eukaryotic pathogen genomics database resource. Nucleic Acids Res. 2017;45:D581–91.
69. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80.
71. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91.
72. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses Bioinformatics 2009;25:1972–1973.
73. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.
74. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020.
75. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35: 518–22.
76. Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556:452–6.
77. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
78. Wang H-C, Minh BQ, Susko E, Roger AJ. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst Biol. 2018;67:216–35.
79. Keister DB. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. Trans R Soc Trop Med Hyg. 1983;77:487–8.
80. Pyrihová E, Motyčková A, Voleman L, Wandyszewska N, Fišer R, Seydlová G, et al. A single Tim translocase in the mitosomes of *Giardia intestinalis* illustrates convergence of protein import machines in anaerobic eukaryotes. Genome Biol Evol. 2018;10:2813–22.
81. Masuda T, Tomita M, Ishihama Y. Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. J Proteome Res. 2008;7:731–40.
82. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, et al. The one hour yeast proteome. Mol Cell Proteomics. 2014;13:339–47.

83. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics. 2014;13:2513–26.
84. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods. 2016;13:731–40.
85. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 2016;44:11033–11033.
86. Voleman L, Najdrová V, Ástvaldsson Á, Tůmová P, Einarsson E, Švindrych Z, et al. *Giardia intestinalis* mitosomes undergo synchronized fission but not fusion and are constitutively associated with the endoplasmic reticulum. BMC Biol. 2017;15:27.
87. : Najdrova V, Stairs CW, Vinopalová M, Voleman L, Dolezal P. The evolution of Puf superfamily proteins for rRNA maturation and mRNA translational regulation across the tree of eukaryotes. Supporting datasets. 2020. Figshare. DOI: https://doi.org/10.6084/m9.figshare.12097692.

Publisher's Note