Proteomic identification of the UDP-GlcNAc : PI α1-6 GlcNAc-transferase subunits of the glycosylphosphatidylinositol biosynthetic pathway of *Trypanosoma brucei*.

Zhe Ji[1], Michele Tinti[1], and Michael A.J. Ferguson[1]*

[1]The Wellcome Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dundee, UK

*Corresponding author: majferguson@dundee.ac.uk

Graphical abstract

First step of GPI anchor biosynthesis pathway in *T.brucei* BSF is catalysed by TbGPI3 complex.

Abstract

The first step of glycosylphosphatidylinositol (GPI) anchor biosynthesis in all eukaryotes is the addition of N-acetylglucosamine (GlcNAc) to phosphatidylinositol (PI) which is catalysed by a UDP-GlcNAc : PI α1-6 GlcNAc-transferase. This enzyme has been shown to be a complex of at least seven subunits in mammalian cells and a similar complex of homologous subunits has been postulated in yeast. Homologs of most of these mammalian and yeast subunits were identified in the *Trypanosoma brucei* predicted protein database. The putative catalytic subunit of the *T. brucei* complex, TbGPI3, was epitope tagged with three consecutive c-Myc sequences at its C-terminus. Immunoprecipitation of TbGPI3-3Myc followed by native polyacrylamide gel electrophoresis and anti-Myc Western blot showed that it is present in a ~240 kDa complex. Label-free quantitative proteomics were performed to compare anti-Myc pull-downs from lysates of TbGPI-3Myc expressing and wild type cell lines. TbGPI3-3Myc was the most highly enriched protein in the TbGPI3-3Myc lysate pull-down and partner proteins TbGPI15, TbGPI9, TbGPI2, TbGPI1 and TbERI1 were also identified with significant enrichment. Our proteomics data also suggest that an Arv1-like protein (TbArv1) is a subunit of the *T. brucei* complex. Yeast and mammalian Arv1 have been previously implicated in GPI biosynthesis, but here we present the first experimental evidence for physical association of Arv1 with GPI biosynthetic machinery. A putative E2-ligase has also been tentatively identified as part of the *T. brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex.

Introduction

Trypanosoma brucei is a protozoan pathogen that undergoes a complex life cycle between its tsetse fly vector and mammalian hosts. The parasite causes human African trypanosomiasis in humans and nagana in cattle in sub-Saharan Africa.

The bloodstream form (BSF) of *T. brucei* produces a dense coat of GPI anchored variant surface protein (VSG) to protect it from the innate immune system and, through antigenic variation, the acquired immune system[1]. Other *T. brucei* surface molecules that have been shown experimentally to possess a GPI membrane anchor are the ESAG6-subunit of the BSF transferrin receptor (TfR) [2] and the procyclins, the major surface glycoproteins of the tsetse mid-gut dwelling procyclic form (PCF) of the parasite [3] . In addition, many other surface molecules with N-terminal signal peptides and C-terminal GPI addition signal peptides are predicted to be GPI-anchored in *T. brucei*, including the BSF haptaglobin-haemaglobin receptor [4] and the factor H receptor [5], the epimastigote BARP glycoprotein [6] and the metacyclic trypomastigote invariant surface protein (MISP)[7]. Thus far, GPI anchor structures have been completely or partially solved for four *T. brucei* VSGs [8–11], the TfR [2] and the procyclins [3]. As for the structure of GPIs, research on *T. brucei* was the first to yield methodologies to delineate the steps of GPI biosynthesis that were subsequently applied to mammalian cells and yeast[12–14]. However, it was the power of mammalian cell and yeast genetics that led to the identification of the majority of GPI biosynthesis genes, reviewed in [15–17].

We currently have reasonably advanced models for GPI anchor biosynthesis and processing in trypanosomes, mammalian cells and yeast and the similarities and differences in these pathways have been reviewed extensively elsewhere [15–18]. For most organisms, the functions and interactions of putative GPI pathway gene products have been inferred from experimental work in mammalian or yeast cells. In a few cases these functions have been experimentally confirmed in *T. brucei*, i.e., for the GlcNAc-PI de-N-acetylase (TbGPI12) [19], the third mannosyltransferase (TbGPI10) [20] and the catalytic (TbGPI8) [21] and other subunits (TTA1 and 2 [22] and TbGPI16) [23]) of the GPI transamidase complex.

The first step of GPI biosynthesis is the addition of GlcNAc to PI by a UDP-GlcNAc : PIα1-6 GlcNAc-transferase complex. The composition of this complex was determined in mammalian cells, where seven subunits have been identified: PIGA, PIGC, PIGH, PIGP, PIGQ, PIGY and DPM2 (Table 1) [15], in which DPM2 is the non-catalytic subunit of dolichol phosphate mannose synthetase. The complex was realised through a series of elegant functional

cloning experiments and co-immunoprecipitation experiments using individually epitope-tagged bait and prey components. A similar multi-subunit complex has been proposed in yeast where homologues for all the subunits, except DPM2, have been identified (Table 1) [17]. However, experimental evidence for the physical associations between these yeast subunits is lacking.

Here we describe epitope tagging of the putative catalytic subunit of *T. brucei* UDP-GlcNAc : PIα1-6 GlcNAc-transferase (TbGPI3), equivalent to yeast GPI3 and mammalian PIGA. Furthermore, we demonstrate its presence in a protein complex and identify its partner proteins through label-free quantitative proteomics.

Materials and Methods

Cultivation of Trypanosomes

T.brucei brucei strain 427 bloodstream form (BSF) parasites expressing VSG variant 221 and transformed to stably express T7 polymerase and the tetracycline repressor protein under G418 antibiotic selection was used in this study and will be referred as bloodstream form wild type (BSF WT). Cells were cultivated in HMI-11T medium containing 2.5 µg/mL of G418 at 37 °C in a 5% CO_2 incubator as previously described [24]. HMI-11T is a modification of the original HMI-9 [25] that uses 56 mM 1-thioglycerol in place of 200 mM 2-mercaptoethanol, and contains 10% heat inactivated fetal bovine serum (PAA) and lacks of serum plus (Hazleton Biologics, Lenexa, Kansas).

DNA Isolation and Manipulation

Plasmid DNA was purified from *Escherichia coli* (chemically competent DH5α cells) using Qiagen Miniprep kits and Maxiprep was performed by the University of Dundee DNA sequencing service. Gel extraction and PCR purification were performed using QIAquick kits (Qiagen). Custom oligonucleotides were obtained from Eurofins MWG Operon or Thermo Fisher. *T. brucei* genomic DNA was isolated from ~5 × 10^7 BSF cells using lysis buffer containing 100 mM Tris-HCl (pH 8.0), 100 mM NaCl, 25 mM EDTA, 0.5% SDS, and 0.1 mg/mL proteinase K (Sigma) by standard methods.

Generation of Gene Replacement Constructs

The tagging cassette was amplified from the pMOTag43M plasmid [26] using the forward primer: 5'-TGATTGATATTGCACCAGATTTTCCACTGGAGTTGTACTCTCGTAACCGGGAGAAGC TTCAAGTTGTGGGAAGCCCATCCgaacaaaagctgggtacc-3' and the reverse primer: 5'-CAACGCGAAACAATGACagAGAGAGAGAGAGAAGGGCGAAAACAAAAAGGATCGC GGTAGAGAGGACCCCGCCCATACCCctattcctttgccctcggac-3'. The PCR product contains 80 bp corresponding to the 3'-end of the TbGPI3 open reading frame (capital letters of forward primer) followed by a sequence encoding the 3Myc epitope tag, an intergenic region (igr) from the *T. brucei* α-β tubulin locus, the hygromycin phosphotransferase (HYG) selectable marker gene and the 3'-UTR of TbGPI3 (capital letters of reverse primer).

Transformation of BSF *T.brucei*

Constructs for *in situ* tagging were purified and precipitated, washed with 70% ethanol, and re-dissolved in sterile water. The released DNA was electroporated into BSF WT cell line. Cell culture and transformation were carried out as described previously [24,26]. After five days of selection with hygromycin, cells were sub-cloned and four independent clones were selected and cultured.

Western blot of cell lysates

To confirm the C-terminal tagging of TbGPI3 with 3Myc, cells from the four selected clones in parallel with BSF WT cells were lysed in SDS sample buffer. Aliquots corresponding to 5×10^6 cells per sample, were subjected to SDS-PAGE on NuPAGE bis-Tris 10% acrylamide gels (Invitrogen) and transferred to a nitrocellulose membrane (Invitrogen). Ponceau staining confirmed equal loading and transfer. The blot was further probed with anti-Myc rat monoclonal antibody (Chromotek, 9E1) in a 1:1,000 dilution. Detection was carried out using IRDye 800CW conjugated goat anti-rat IgG antibody (1:15,000) and LI-COR Odyssey infrared imaging system (LICOR Biosciences, Lincoln, NE).

Co-immunoprecipitation and Native-PAGE protein blotting

To investigate detergent solubilisation conditions for the immunoprecipitation of TbGPI3-3Myc complexes, aliquots of 2×10^8 cells were harvested and lysed in 500 µL of 50 mM Tris-HCl, pH 7.4, 150 mM NaCl containing different detergents; 0.5% digitonin, 1% digitonin, 1% Triton X-100 (TX-100), 1% n-octyl-beta-glucoside (NOG) or 1% decyl-β-D-maltopyranoside (DM). After centrifugation at 16,000 g, 4 °C for 20 min, aliquots of the supernatants equivalent to 2 x 10^8 cells were incubated with 10 µL anti-Myc agarose beads (Myc-Trap™, Chromotek) for 1 h at 4 °C. The beads were washed three times in 50 mM Tris-HCl, pH 7.4, 150 mM NaCl containing the corresponding detergents and bound proteins were eluted three times with 10 µL 0.5 mg/mL c-Myc peptide (Sigma M2435) in the corresponding detergent containing buffer. The combining eluates for each detergent condition, equivalent to 2×10^8 cells, were subjected to NativePAGE (Invitrogen) and transferred to a PVDF membrane (Invitrogen) followed by immunoblotting with anti-Myc antibody (Chromotek, 9E1) in 1:1,000. The blot was then developed by ECL using an HRP-conjugated secondary antibody (Sigma, A9037, 1:3,000).

Label free proteomics of TbGPI3-3Myc and BSF WT lysate pull downs

BSF WT and TbGPI3-3Myc expressing cell lines were cultured and 1×10^9 cells of each were harvested and lysed in 1 mL of lysis buffer containing 0.5% digitonin. After centrifugation 16,000 g, 4 °C for 20 min, the supernatants were mixed with 20 µL of Myc-Trap™ beads and incubated for 1 h at 4 °C. The beads were washed three times in the same buffer, and bound proteins were eluted with 1×SDS sample buffer and subjected to SDS-PAGE, running the proteins only 10 cm into the gel. Whole lanes containing TbGPI3 and wild type cell lines samples were cut identically into 3 slices and the gel pieces were dried in Speed-vac (Thermo Scientific) for in-gel reduction with 0.01 M dithiothreitol and alkylation with 0.05 M iodoacetamide (Sigma) for 30 min in the dark. The gel slices were washed in 0.1 M NH_4HCO_3, and digested with 12.5 µg/mL modified sequence grade trypsin (Roche) in 0.02 M NH_4HCO_3 for 16 h at 30 °C. Samples were dried and re-suspended in 50 µL 1% formic acid and then subjected to liquid chromatography on Ultimate 3000 RSLC nano-system (Thermo Scientific) fitted with a 3 Acclaim PepMap 100 (C18, 100 µM × 2 cm) and then separated on an Easy-Spray PepMap RSLC C18 column (75 µM × 50 cm) (Thermo Scientific). Samples (15µL) were loaded in 0.1% formic acid (buffer A) and separated using a binary gradient consisting of buffer A and buffer B (80% acetonitrile, 0.1% formic acid). Peptides were eluted with a linear gradient from 2 to 35%

buffer B over 70 min. The HPLC systems were coupled to a Q-Exactive Plus Mass Spectrometer (Thermo Scientific) equipped with an Easy-Spray source with temperature set at 50 °C and a source voltage of 2.0 kV. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD022979 [27].

Protein identification by MaxQuant

RAW data files were analysed using MaxQuant version 1.6.10.43, with the in-built Andromeda search engine [28], using the *T. brucei brucei* 927 annotated protein sequences from TriTrypDB release 46 [29], supplemented with the *T. brucei brucei* 427 VSG221 (Tb427.BES40.22) protein sequence. The mass tolerance was set to 4.5 ppm for precursor ions and MS/MS mass tolerance was set at 20 ppm (MaxQuant default parameters). The enzyme was set to trypsin, allowing up to 2 missed cleavages. Carbamidomethyl on cysteine was set as a fixed modification. Acetylation of protein N-termini, and oxidation of methionine were set as variable modifications. Match between runs was enabled, allowing transfer of peptide identifications of sequenced peptides from one LC-MS run to non-sequenced ions, with the same mass and retention time, in another run. A 20-min time window was set for alignment of separate LC-MS runs. The false-discovery rate for protein and peptide level identifications was set at 1%, using a target-decoy based strategy.

Data Analysis

Data analysis was performed using custom Python scripts, using the SciPy ecosystem of open-source software libraries [30]. The data analysis pipeline is available at GitHub https://github.com/mtinti/PIG-A and Zenodo https://zenodo.org/record/3735036 repositories, DOI:10.5281/zenodo.3735036. The MaxQuant proteinGroups.txt output file was used to extract the iBAQ scores for forward trypanosome protein sequences identified with at least two unique peptides and with an Andromeda score >4. The protein iBAQ scores were normalised for sample loading by dividing each iBAQ value by the median of all the iBAQ values in each experiment. Missing values were replaced by the smallest iBAQ value in each sample. Differential abundance analysis between the bait and control samples was performed with the ProtRank Python package [31]. Briefly, ProtRank performs a rank test between each control and bait

sample pair to output as signed-rank and false discovery rate values. The signed-rank is proportional to the significance of the differential abundance of the protein groups between the bait and control samples.

The BSF intensity rank was computed from a recent dataset published by our laboratory [32] of *T. brucei* protein half-lives computed from a label-chase experiment. In those experiments, BSF parasites were labelled to steady-state in medium SILAC culture medium (M) and then placed into light SILAC culture medium (L). Seven time points, with three biological replicates, were sampled and each mixed 1:1 with BSF lysate labelled to steady state in heavy SILAC culture medium (H) to provide an internal standard for normalisation. Here, we exploited the heavy-labelled internal standard in every sample: The log10 summed eXtracted Ion Currents (XICs) of the heavy-labelled peptides for each protein were averaged across the BSF technical replicates and used to rank a deep BSF proteome from the most abundant (rank=1) to the least abundant (rank=7125). Missing values were replaced with the highest rank (7125).

Results

Identification of putative *T. brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex components.

Conventional BLASTp searches with default settings (13) were sufficient to identify *T. brucei* homologues of PIGA(GPI3), PIGC(GPI2), PIGP(GPI19), PIGQ(GPI1) and DPM2. However, the results for PIGH(GPI15) and PIGY(ERI1) were equivocal so a Domain Enhanced Lookup Time Accelerated BLAST [33] using a PAM250 matrix was applied to find the corresponding *T. brucei* homologues (Table 1).

Table 1. Genes encoding known and putative UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex subunits in mammalian cells, yeast and *T. brucei*.

In situ epitope tagging of TbGPI3.

To investigate whether a multi-subunit UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex might exist in *T. brucei* we selected TbGPI3, which encodes a 455 amino acid protein with two predicted transmembrane domains, one near its N-terminus and one near its C-terminus [34], for

epitope tagging. We chose this PIGA(GPI3) homologue as the bait protein because PIGA has been shown to have either direct or indirect interactions with all other subunits in the mammalian UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex [15]. Alignment of putative TbGPI3, yeast GPI3 and PIGA protein sequences show that the *T. brucei* sequence has 43.9% and 50.8% sequence identity with the yeast and human sequences, respectively (Fig. S1).

In situ tagging of the TbGPI3 gene was achieved by transfecting BSF *T. brucei* with PCR products amplified from the pMOTag43M plasmid [26], (Fig. 1A). Transfected cells were selected using hygromycin and subsequently cloned by limit-dilution. Lysates of four separate clones were subjected to anti-Myc Western blotting (Fig 1B and C). *In situ* tagged TbGPI3-3Myc protein was detected in all four clones at an apparent molecular weight of ~47 kDa, somewhat lower than the predicted molecular weight of 55 kDa.

Fig.1. ***In situ* C-terminal tagging of TbGPI3 with 3Myc.** (A) Map of plasmid pMOTag43M [26] used for the *in situ* tagging of TbGPI3, and a scheme of how the PCR product generated with the indicated forward (For) and reverse (Rev) primers inserts into the 3'-end of the *TbGPI3* ORF (checked box) and 3'-UTR (striped box) in the parasite genome to effect *in-situ* tagging. HYG = hygromycin phosphotransferase selectable marker; igr = α-β tublin intergenic region. (B) Ponceau staining of denaturing SDS-PAGE Western blot shows similar loading and transfer of lysates (corresponding to 5×10^6 cells) from four *in-situ* tagged clones (lanes 1-4) and wild type cells (lane 5). (C) The identical blot was probed with anti-Myc antibody. TbGPI-3Myc is indicated by the arrow. The positions of molecular weight markers are indicated on the left of (B) and (C).

Solubilisation and native-PAGE of TbGPI3-3Myc.

The analysis of epitope-tagged membrane bound multiprotein complexes requires detergent extraction and anti-epitope pull-down under conditions that preserve intermolecular interactions within the complex. To investigate detergent extraction conditions, TbGPI-3Myc expressing cells were cultured and lysed with 0.5% digitonin, 1% digitonin, 1% TX-100, 1% NOG and 1% DM and centrifuged. The solubilised proteins in the supernatants from these treatments, along with a 1% TX-100 extract of wild type cells, were immunoprecipitated with Myc-Trap™ agarose beads that were washed three times and finally eluted with synthetic c-Myc peptide. The proteins in the eluates were separated by denaturing SDS-PAGE and by native PAGE [35] and analysed by anti-

Myc Western blot (Fig 2A and B, respectively). All detergents, apart from DM, extracted the TbGPI3-3Myc protein (Fig 2A). Of these conditions, 1% TX-100 gave the highest efficiency of extraction but duplicate samples analysed by native PAGE and anti-Myc Western blot showed that digitonin best preserved a TbGPI3-3Myc-containing complex with a native apparent molecular weight of ~240 kDa and that 0.5% digitonin gave a higher-yield of the complex than 1 % digitonin (Fig. 2B). The reason for not detecting any clear complexes in other conditions may be due to the ~240 kDa complex falling apart into multiple sub-complexes below the limits of detection.

Fig 2. **TbGPI3-3Myc is present in complexes in BSF *T. brucei*.** (A) Aliquots of 2×10^8 cells were harvested and lysed in lysis buffer containing different detergents to assess TbGPI3-3Myc solubilisation. After immunoprecipitation of the supernatants with anti-Myc agarose beads, proteins were eluted with 0.5 mg/mL c-Myc peptide solution and aliquots were subjected to SDS-PAGE followed by anti-Myc Western blotting. (B) Identical samples were also separated by native-PAGE and subjected to anti-Myc Western blotting. In both cases, lane 1 corresponds to wild type cells lysed with 1% TX-100 as a negative control and lanes 2-6 correspond to TbGPI3-3Myc clone1 lysed with 0.5% digitonin, 1% digitonin, 1% TX-100, 1% NOG or 1% DM, respectively.

Identification of *T. brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex components by quantitative proteomics.

Having established detergent solubilisation conditions that retained TbGPI13-3Myc in a complex, we performed label-free quantitative proteomics on Myc-Trap™ pull downs to identify the components within the complex. For this experiment, BSF WT and TbGPI13-3Myc expressing parasites were grown under identical conditions and the same numbers of cells were harvested and lysed in 0.5 % digitonin lysis buffer. Immunoprecipitation was performed using Myc-Trap™ beads and the proteins eluted from these two samples by c-Myc peptide were processed to tryptic peptides for LC-MS/MS analysis (Fig. 3A). The experiments were performed in biological triplicates and the data were analysed using MaxQuant software and a newly developed data analysis method written in Python called ProtRank [31], see Materials and Methods. The protein groups identified were displayed on a plot of the minus log10 value of their False Discovery Rate (y-axis) and enrichment rank (x-axis) between the bait versus control

samples (Fig. 3B). As expected, the bait protein TbGPI3-3Myc was the most highly enriched protein and its putative partner proteins TbGPI15, TbGPI9, TbGPI2, TbGPI1 and TbERI1 were also significantly enriched. Notably, TbDPM2 (dolichol-phosphate-mannose synthetase 2) was not detected. However, although TbDPM2 is annotated in the TriTrypDB database it should be noted that, like yeast [36], *T. brucei* makes a single-chain dolichol-phospho-mannose synthetase (DPM1) rather than a trimeric enzyme made of a soluble catalytic DPM1 subunit associated with small transmembrane DPM2 and DPM3 subunits, as found in mammalian cells. For these reasons, we feel that the absence of DPM protein components in the *T. brucei* complex to be expected.

Interestingly, an Arv1-like protein (hereon referred to as TbArv1, Tb927.3.2480) and a putative ubiquitin-conjugating enzyme E2 (UbCE, Tb927.2.2460) were also co-immunoprecipitated with TbGPI3 (Fig 3B). The data were also processed in a different way (see Materials and Methods) that plots the experimental rank (x-axis) against the rank order of estimated abundances of the protein groups, generated from data in (31), on the y-axis (Fig 3C). In this plot, the very low abundant TbArv1 clusters better with the canonical and similarly low abundant UDP-GlcNAc : PI α1-6 GlcNAc-transferase subunits. By contrast, although UbCE is clearly enriched in the pull-down it is a much more abundant protein, suggesting that only some fraction of it may be associated with the complex.

Fig.3. **Identification of UDP-GlcNAc : PI α1-6 GlcNAc-transferase subunits by immunoprecipitation of TbGPI3-3Myc from BSF *T. brucei* digitonin lysates.** (A) Scheme of the label-free proteomics approach to identify TbGPI13-3Myc binding partners. BSF WT and TbGPI13-3Myc expressing cell lines were cultured, harvested and lysed in 0.5% digitonin lysis buffer. Identical quantities of the supernatants were subjected to anti-Myc agarose bead immunoprecipitation and the bound proteins were eluted from the beads with c-Myc peptide. The eluted proteins were reduced, alkylated and digested with trypsin and the resulting peptides analysed by LC-MS/MS. (B) Volcano plot comparing protein groups present in the anti-cMyc immunoprecipitates from TbGPI3-3Myc expressing cell lysates versus WT cell lysates. Mean values (from biological triplicate experiments) for each protein group (dots) are plotted according to their minus log10 False Discovery Rate values (y-axis), calculated by MaxQuant, and the enrichment rank (x-axis). The enrichment rank was computed with the ProtRank algorithm using the iBAQ values calculated by MaxQuant. The higher the rank value on the x-axis, the higher the abundance in the TbGPI3-3Myc samples. The putative subunits of UDP-GlcNAc : PI α1-6

GlcNAc-transferase in *T. brucei* are highlighted in red and annotated with their corresponding names (Table 1). (C) Relative intensity plot using a new algorithm. The same data as (B) are plotted with a different y-axis, whereby each protein group is assigned an intensity rank from the most abundant protein group (1) to least abundant protein groups (7,125) based on their summed eXtracted Ion Currents (XICs) for the total BSF proteome. (Details of the mass spectrometry and data analysis are provided in in Materials and Methods.)

Discussion

The proteomics data suggest that: (i) the *T. brucei* UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex contains the expected subunits (TbGPI3, TbGPI15, TbGPI9, TbGPI2, TbGPI1 and TbERI1). (ii) like yeast, but unlike mammalian cells, DPM components are not subunits of the parasite complex. (iii) an Arv1-like protein (TbArv1) is a part of the parasite complex. (iv) a putative E2-ligase UbCE may be a part of the parasite complex.

TbArv-1 is predicted to contain four transmembrane domains and an Arv1 domain (PF04161). Previous studies in yeast have indicated that Arv1p is required for the efficient synthesis of Man_1GlcN-acylPI (mannosyl-glucosaminyl-acyl-phosphatidylinositol)[36] and has been postulated to be a GPI flippase [36] [37] helping deliver GlcN-acylPI, which is made on the cytoplasmic face of the ER, to the active site of mannosyl-transferase I (MT I) on the luminal face of the ER. The complementation of yeast Arv1 mutants by the human Arv1 [38] and recent findings that human Arv1 mutations lead to deficiencies in GPI anchoring [39] [40] strongly suggest a related role in mammalian cells and that it is a component of the mammalian UDP-GlcNAc : PI α1-6 GlcNAc-transferase complex. It is possible that TbArv1 plays an analogous role to that proposed for Arv1 in yeast and mammalian cells in the *T. brucei* GPI pathway. However, since (unlike yeast and mammalian cells) acylation of the PI moiety occurs strictly after the action of MT I in *T. brucei* [41], TbArv1 would need to facilitate the delivery of GlcN-PI rather than GlcN-acylPI to TbMT Ⅰ in this parasite. Alternatively, mammalian, yeast and *T. brucei* Arv1 proteins may play some other, perhaps regulatory, role in the UDP-GlcNAc : PI α1-6 GlcNAc-transferase reaction.

Finally, a recent study in mammalian cells showed that GPI anchor biosynthesis is upregulated in ERAD (endoplasmic-reticulum-associated protein degradation) deficient and

PIGS mutant cell lines, suggesting that the GPI anchor biosynthetic pathway is somehow linked to and regulated by the ERAD system [42] . Since ERAD involves E2-dependent ubiquitylation of misfolded proteins as they exit the ER, it is possible that the UbCE associated, in part, with UDP-GlcNAc : PI α1-6 GlcNAc-transferase might play some role in regulation of the *T. brucei* GPI pathway.

Acknowledgements: We thank Drs Alvaro Acosta Serrano, Lucia Güther and Samuel Duncan for helpful advice and the Fingerprints Proteomics Facility for expert assistance with quantitative proteomics. This work was supported by a China Scholarship Council PhD scholarship to Z.J. (No.201706310166) and a Wellcome Investigator Award (101842/Z13/Z) to M.A.J.F.

References

1. Horn D. Antigenic variation in African trypanosomes. Mol Biochem Parasitol. 2014;195(2):123–9.
2. Mehlert A, Ferguson MAJ. Structure of the glycosylphosphatidylinositol anchor of the Trypanosoma brucei transferrin receptor. Mol Biochem Parasitol. 2007;151(2):220–3.
3. Treumann A, Zitzmann N, Hülsmeier A, Prescott AR, Almond A, Sheehan J, et al. Structural characterisation of two forms of procyclic acidic repetitive protein expressed by procyclic forms of Trypanosoma brucei. J Mol Biol. 1997;269(4):529–47.
4. Lane-Serff H, MacGregor P, Lowe ED, Carrington M, Higgins MK. Structural basis for ligand and innate immunity factor uptake by the trypanosome haptoglobin-haemoglobin receptor. Elife. 2014;3:e05553.
5. Macleod OJS, Bart JM, MacGregor P, Peacock L, Savill NJ, Hester S, et al. A receptor for the complement regulator factor H increases transmission of trypanosomes to tsetse flies. Nat Commun. 2020;11(1):1–12.
6. Urwyler S, Studer E, Renggli CK, Roditi I. A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of Trypanosoma brucei. Mol Microbiol. 2007;63(1):218–28.
7. Casas-Sánchez A, Perally S, Ramaswamy R, Haines LR, Rose C, Yunta C, et al. The crystal structure and localization of Trypanosoma brucei invariant surface glycoproteins suggest a more permissive VSG coat in the tsetse-transmitted metacyclic stage. bioRxiv. 2018;
8. Ferguson M, Homans S, Dwek R, Rademacher T. Glycosyl-phosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane. Science (80-). 1988 Feb 12;239(4841):753–9.

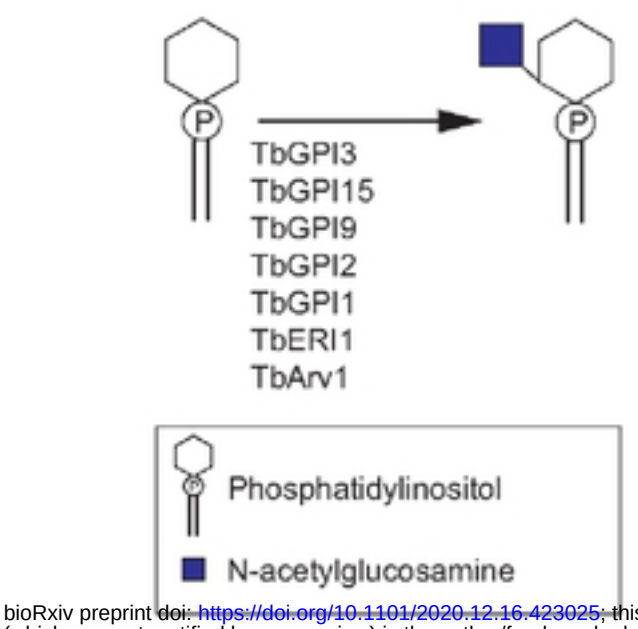
9. Mehlert A, Richardson JM, Ferguson MAJ. Structure of the Glycosylphosphatidylinositol Membrane Anchor Glycan of a Class-2 Variant Surface Glycoprotein from Trypanosoma brucei. 1998;3.
10. Mehlert A, Sullivan L, Ferguson MAJ. Glycotyping of Trypanosoma brucei variant surface glycoprotein MITat1.8. Mol Biochem Parasitol. 2010;174(1):74–7.
11. Treumann A, Güther MLS, Schneider P, Ferguson MAJ. Analysis of the Carbohydrate and Lipid Components of Glycosylphosphatidylinositol Structures. In: Glycoanalysis Protocols. New Jersey: Humana Press; 2010. p. 213–36.
12. Masterson WJ, Doering TL, Hart GW, Englund PT. A novel pathway for glycan assembly: Biosynthesis of the glycosyl-phosphatidylinositol anchor of the trypanosome variant surface glycoprotein. Cell. 1989;56(5):793–800.
13. Masterson WJ, Raper J, Doering TL, Hart GW, Englund PT. Fatty acid remodeling: A novel reaction sequence in the biosynthesis of trypanosome glycosyl phosphatidylinositol membrane anchors. Cell. 1990;62(1):73–80.
14. Menon AK, Schwarz RT, Mayor S, Cross GAM. Cell-free synthesis of glycosyl-phosphatidylinositol precursors for the glycolipid membrane anchor of Trypanosoma brucei variant surface glycoproteins. Structural characterization of putative biosynthetic intermediates. J Biol Chem. 1990;265(16):9033–42.
15. Kinoshita T. Biosynthesis and biology of mammalian GPI-anchored proteins. Open Biol. 2020;10(3):190290.
16. Kinoshita T, Fujita M. Biosynthesis of GPI-anchored proteins: Special emphasis on GPI lipid remodeling. J Lipid Res. 2016;57(1):6–24.
17. Pittet M, Conzelmann A. Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae. Biochim Biophys Acta - Mol Cell Biol Lipids. 2007;1771(3):405–20.
18. Ferguson MAJ, Kinoshita T HG. Glycosylphosphatidylinositol Anchors. In: Varki A, Cummings RD, Esko JD et al., editor. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2009. p. 137–50.
19. Chang T, Milne KG, Güther MLS, Smith TK, Ferguson MAJ. Cloning of Trypanosoma brucei and Leishmania major genes encoding the GlcNAc-phosphatidylinositol De-N-acetylase of glycosylphosphatidylinositol biosynthesis that is essential to the African sleeping sickness parasite. J Biol Chem. 2002;277(51):50176–82.
20. Nagamune K, Nozaki T, Maeda Y, Ohishi K, Fukuma T, Hara T, et al. Critical roles of glycosylphosphatidylinositol for Trypanosoma brucei. Proc Natl Acad Sci U S A. 2000;97(19):10336–41.

21. Lillico S, Field MC, Blundell P, Coombs GH, Mottram JC. Essential Roles for GPI-anchored Proteins in African Trypanosomes Revealed Using Mutants Deficient in GPI8. Gilmore R, editor. Mol Biol Cell. 2003 Mar;14(3):1182–94.
22. Nagamune K, Ohishi K, Ashida H, Hong Y, Hino J, Kangawa K, et al. GPI transamidase of Trypanosoma brucei has two previously uncharacterized (trypanosomatid transamidase 1 and 2) and three common subunits. Proc Natl Acad Sci U S A. 2003;100(19):10682–7.
23. Hong Y, Nagamune K, Ohishi K, Morita YS, Ashida H, Maeda Y, et al. TbGPI16 is an essential component of GPI transamidase in Trypanosoma brucei. FEBS Lett. 2006;580(2):603–6.
24. Wirtz E, Leal S, Ochatt C, Cross GAM. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Mol Biochem Parasitol. 1999;99(1):89–101.
25. Hirumi H, Hirumi K. Continuous Cultivation of Trypanosoma brucei Blood Stream Forms in a Medium Containing a Low Concentration of Serum Protein without Feeder Cell Layers Author (s): Hiroyuki Hirumi and Kazuko Hirumi Published by : Allen Press on behalf of The American Soc. J Parasitol. 1989;75(6):985–9.
26. Oberholzer M, Morand S, Kunz S, Seebeck T. A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. Mol Biochem Parasitol. 2006;145(1):117–20.
27. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. Nucleic Acids Res. 2019;47(D1):D442–50.
28. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301–19.
29. Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. Nucleic Acids Res. 2010 Jan;38(suppl_1):D457–62.
30. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72.
31. Medo M, Aebersold DM, Medová M. ProtRank: Bypassing the imputation of missing values in differential expression analysis of proteomic data. BMC Bioinformatics. 2019;20(1):1–12.
32. Tinti M, Güther MLS, Crozier TWM, Lamond AI, Ferguson MAJ. Proteome turnover in the bloodstream and procyclic forms of trypanosoma brucei measured by quantitative proteomics [version 1; peer review: 3 approved]. Wellcome Open Res. 2019;4:1–26.
33. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. Biol Direct. 2012;7:1–14.

452 34. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide
453 prediction method. J Mol Biol. 2004;338(5):1027–36.
454 35. Schägger H, von Jagow G. Blue native electrophoresis for isolation of membrane protein
455 complexes in enzymatically active form. Anal Biochem. 1991;199(2):223–31.
456 36. Kajiwara K, Watanabe R, Pichler H, Ihara K, Murakami S, Riezman H, et al. Yeast ARV1 Is
457 Required for Efficient Delivery of an Early GPI Intermediate to the First Mannosyltransferase
458 during GPI Assembly and Controls Lipid Flow from the Endoplasmic Reticulum. Munro S, editor.
459 Mol Biol Cell. 2008 May;19(5):2069–82.
460 37. Okai H, Ikema R, Nakamura H, Kato M, Araki M, Mizuno A, et al. Cold-sensitive phenotypes of a
461 yeast null mutant of ARV1 support its role as a GPI flippase. FEBS Lett. 2020;594(15):2431–9.
462 38. Ikeda A, Kajiwara K, Iwamoto K, Makino A, Kobayashi T, Mizuta K, et al. Complementation
463 analysis reveals a potential role of human ARV1 in GPI anchor biosynthesis. Yeast.
464 2016;33(2):37–42.
465 39. Davids M, Menezes M, Guo Y, Mclean SD, Hakonarson H, Collins F, et al. Homozygous splice-
466 variants in human ARV1 cause GPI-anchor synthesis deficiency. Mol Genet Metab.
467 2020;130(1):49–57.
468 40. Segel R, Aran A, Gulsuner S, Nakamura H, Rosen T, Walsh T, et al. A defect in GPI synthesis as
469 a suggested mechanism for the role of ARV1 in intellectual disability and seizures. Neurogenetics.
470 2020;21(4):259–67.
471 41. Guther MLS, Ferguson MAJ. The role of inositol acylation and inositol deacylation in GPI
472 biosynthesis in Trypanosoma brucei. Parasitol Today. 1995;11(9):319.
473 42. Wang Y, Maeda Y, Liu YS, Takada Y, Ninomiya A, Hirata T, et al. Cross-talks of
474 glycosylphosphatidylinositol biosynthesis with glycosphingolipid biosynthesis and ER-associated
475 degradation. Nat Commun. 2020;11(1):1–18.

476

Graphic abstract



Tabel 1

Mammalian cell gene names	Yeast gene names	*T. brucei* gene names (Gene IDs inTriTrypDB)	Predicted protein molecular weight of *T.brucei* subunits
PIGA	*GPI3*	*TbGPI3* (Tb927.2.1780)	51.6kDa
PIGH	*GPI15*	*TbGPI15* (Tb927.5.3680)	28.8 kDa
PIGP	*GPI19*	*TbGPI19* (Tb927.10.10110)	16.5 kDa
PIGQ	*GPI1*	*TbGPI1* (Tb927.3.4570)	81.5 kDa
PIGC	*GPI2*	*TbGPI2* (Tb927.10.6140)	38.5 kDa
PIGX	*FRI1*	*TbFRI1* (Tb927.4.780)	10.1 kDa
DPM2		*TbDPM2* (Tb927.9.6440)	9.3 kDa

Figure 1

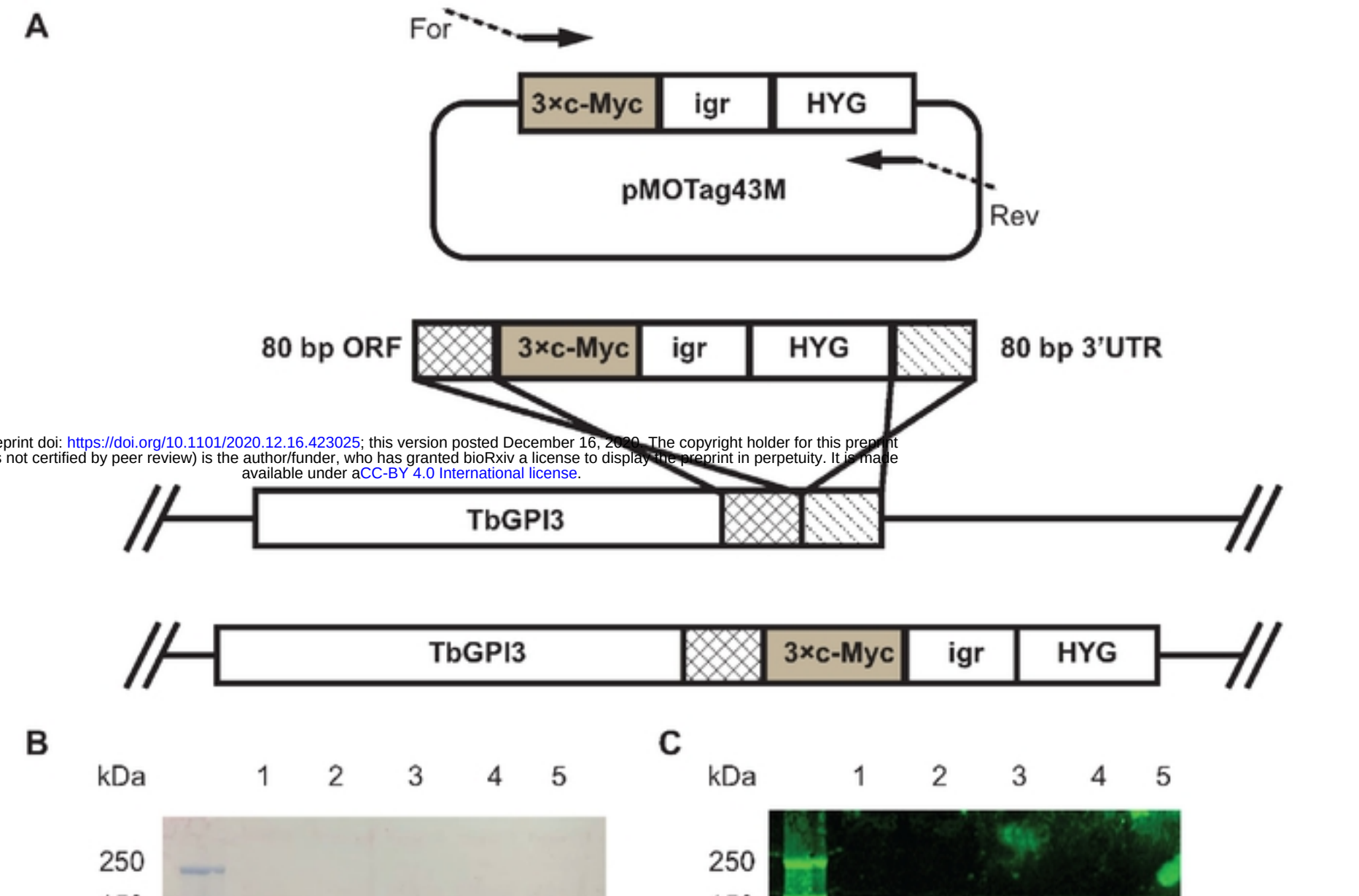


Figure 2

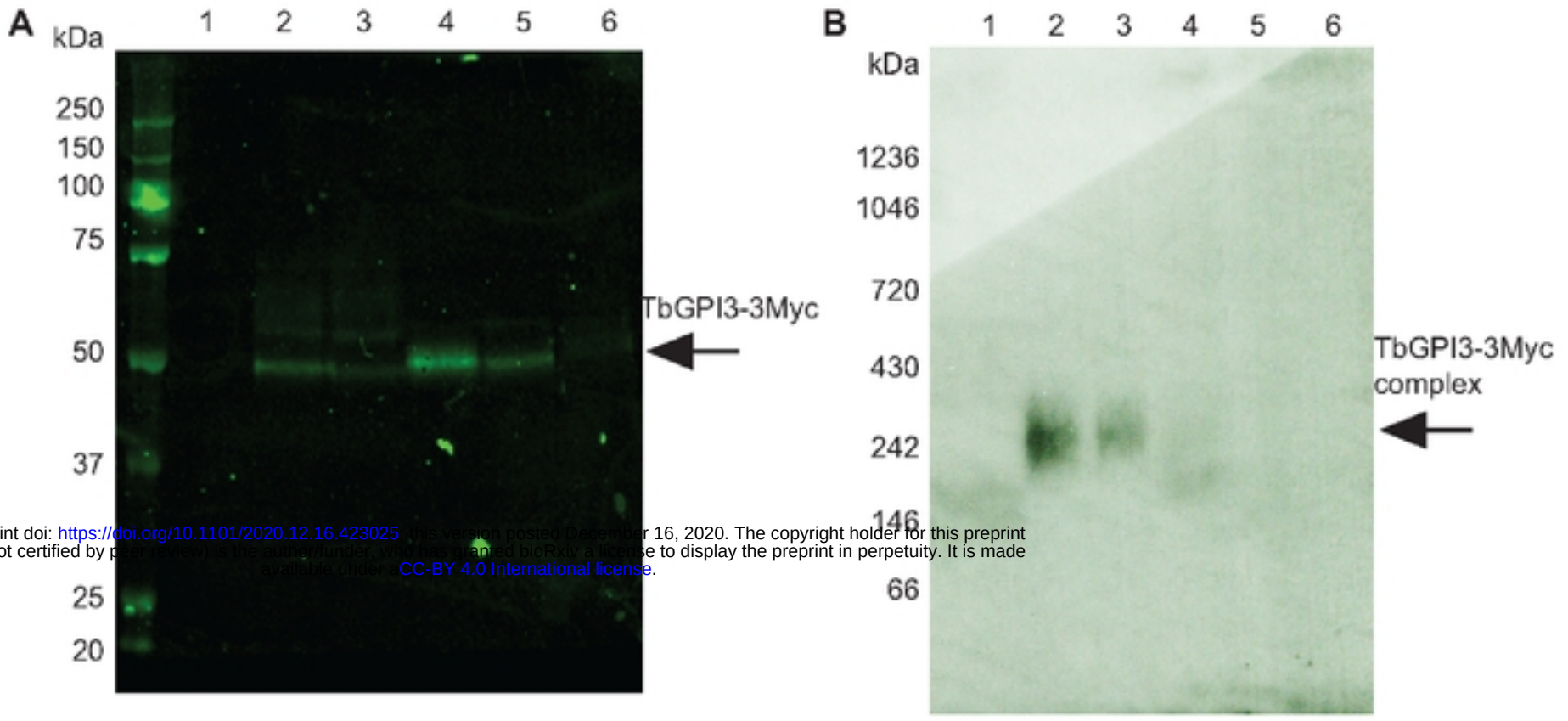


Figure 3