CellPress
OPEN ACCESS

Trends in
Parasitology

Review

Cell-to-flagellum attachment and surface architecture in kinetoplastids

Laryssa Vanessa de Liz [1], Patrícia Hermes Stoco [1] and Jack D. Sunter [2,*]

A key morphological feature of kinetoplastid parasites is the position and length of flagellum attachment to the cell body. This lateral attachment is mediated by the flagellum attachment zone (FAZ), a large complex cytoskeletal structure, which is essential for parasite morphogenesis and pathogenicity. Despite the complexity of the FAZ only two transmembrane proteins, FLA1 and FLA1BP, are known to interact and connect the flagellum to the cell body. Across the different kinetoplastid species, each only has a single FLA/FLABP pair, except in *Trypanosoma brucei* and *Trypanosoma congolense* where there has been an expansion of these genes. Here, we focus on the selection pressure behind the evolution of the FLA/FLABP proteins and the likely impact this will have on host–parasite interactions.

Highlights

Lateral attachment of the flagellum to the cell body is mediated by the flagellum attachment zone (FAZ) and defines kinetoplastid cell morphogenesis and pathogenicity.

Two interacting proteins, FLA1 and FLA1BP, dominate the attachment in the FAZ extracellular domain and we propose that FLA1 and FLA1BP structure has evolved as a consequence of surface protein coat architecture.

The FLA and FLABP gene family have duplicated and expanded in the African trypanosomes, *Trypanosoma brucei* and *Trypanosoma congolense* but not in *Trypanosoma vivax*.

A recent *T. vivax* vaccine candidate generated antibodies that bound to the cell adjacent to the FAZ and we suggest that accessibility to an invariant antigen results from *T. vivax* expressing only a single FLA1/FLA1BP pair, which needs to maintain flagellum attachment throughout its life cycle across different surface architectures.

FLA and FLABP proteins mediate the lateral attachment of the flagellum to the cell body

Kinetoplastid parasites have a single flagellum laterally attached to the cell body with the connection mediated by a large cytoskeletal structure called the **flagellum attachment zone** (**FAZ**) (see Glossary) [1,2] (Figure 1A). FAZ length varies between the different parasite species; however, in all of them the FAZ connects the cell body cytoskeleton to the cytoskeletal structures in the flagellum through the cell body and flagellum membranes [3,4] (Figure 1B,C). These linkages are essential, and disruption to the FAZ leads to defects in cell morphogenesis and a reduction in parasite pathogenicity [2,4,5]. Flagellum attachment occurs within the context of the surface coat of these parasites, variations in which underpin their interaction with respective hosts and vectors. Despite the complexity of the FAZ, with ~100 FAZ proteins identified, the only FAZ proteins with large extracellular domains required for connection are FLA1 and its binding partner FLA1BP [6–8] (Figure 1D). Although FLA1–FLA1BP interaction has been confirmed experimentally, their binding mechanism is unknown [6]. Here, we consider the functional implications of the duplication and diversification of these proteins in *Trypanosoma brucei* and *Trypanosoma congolense* and the implications for vaccine development for African trypanosomes.

FLA and FLABP gene family in *T. brucei*

FLA1 was identified in *T. brucei* due to its identity to GP72, a glycoprotein required for *Trypanosoma cruzi* flagellum attachment [7,9,10]. Depleting FLA1 by RNAi in *T. brucei* **procyclic form** (**PCF**) resulted in flagellum detachment from the cell body [11]. When RNAi was used to deplete FLA1 in **bloodstream forms** (**BSFs**) the flagellum also became detached. However, when the amount of *FLA1* mRNA was analysed by northern blot, a second higher-molecular-weight band was observed in BSFs in addition to the expected *FLA1* band. The higher-molecular-weight band corresponded to the transcript for FLA2, which is highly similar to FLA1 over the first 100 amino acids. This demonstrated the stage-specific expression of the FLA proteins, and later transcriptomic and proteomic work confirmed that *FLA2* is enriched in BSFs and that *FLA1* is enriched in PCFs [12–15]. FLA1

[1]Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil
[2]Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK

*Correspondence:
jsunter@brookes.ac.uk (J.D. Sunter).

RNAi in BSFs reduced both *FLA1* and *FLA2* mRNA, so the relative importance of FLA1/FLA2 for BSF flagellum attachment is still unclear [8]. Subsequently, an additional gene, FLA3, was identified that was 97% identical to FLA2 at the amino acid level and followed the same expression pattern, enriched in BSFs versus PCFs [6,13,14].

An immunoprecipitation approach identified FLA1BP as the interacting partner of FLA1 (Figure 1), and as with FLA1, FLA1BP mRNA and protein is enriched in PCFs [6,15]. FLA1BP is encoded by two identical genes, with one copy directly upstream of *FLA2* and the other upstream of *FLA3*. In the kinetoplastids, genes do not have their own promoters, and gene duplication is often a strategy to increase protein expression [16]. The stoichiometry of the FLA1–FLA1BP interaction is unknown. The duplicated copies of *FLA1BP* suggest that more FLA1BP is required, and therefore the FLA1–FLA1BP interaction may not have a 1:1 stoichiometry.

FLA1BP RNAi caused the detachment of the flagellum from the cell body [6]. Importantly, through a combination of FLA1/FLA1BP RNAi and eYFP-tagging, FLA1 was shown to localise to the cell body membrane of the FAZ and FLA1BP to the flagellum membrane [6]. Around the same time, two FAZ proteins, expressed only in BSFs, were identified and both also named FLA3 (Tb927.5.4570, Tb927.5.4580) [15,17]. These FLA3s identified by Woods localised to the FAZ, and their depletion resulted in flagellum detachment [17]. They are encoded by two genes on chromosome 5, and although similar, are not identical (91% identity) and are related to the FLA1BP [6,17].

Given the expression patterns and relationships between these proteins, we conjecture, as did Sun and colleagues [6], that the FLA3s (Tb927.5.4570, Tb927.5.4580) identified by Woods are the BSF equivalent of FLA1BP and will localise to the flagellum membrane in BSF cells and bind to FLA2 and FLA3 (Tb927.8.4110) on the cell body membrane and potentially FLA1, as this gene is expressed at the transcript level in BSFs. There are currently three different proteins called FLA3. We suggest re-naming the FLA3 (Tb927.8.4110) identified by Sun to FLA2 and the FLA3s (Tb927.5.4570, Tb927.5.4580) identified by Woods to FLA2BP (Figure 1E,F) [17]. The interaction between FLA2 and FLA2BP is a reasonable conjecture, yet we lack direct evidence for this; however, Nolan and colleagues have recently been able to demonstrate the binding of FLA2BP to FLA2 (D. Nolan, personal communication). While this new nomenclature simplifies matters, it must be noted that there are two copies of FLA2 and FLA2BP with differences between the copies, which may impact their function and specificity of interaction.

Structural organisation of FLA and FLABP

All the FLA/FLABP proteins identified in *T. brucei* have the same domain structure, with a large extracellular domain that contains a predicted NHL repeat domain and a transmembrane domain close to the C terminus with a short intracellular region (Figure 1). FLA1 and FLA2 are predicted to have a signal peptide and hence have a type I topology. However, the situation for the FLABPs is more complex – FLA1BP does not have a predicted signal peptide, with a transmembrane domain close to its N terminus instead. The FLA2BPs are predicted to have a signal peptide but with a low probability. When the FLABPs from a range of species are aligned, the *T. brucei* sequences have a short N-terminal extension (see Figure S1 in the supplemental information online). However, just downstream of the predicted start codon for FLA1BP and FLA2BP, there are in-frame methionine residues, and if these are used as the start of a shorter protein, then a signal peptide is more strongly predicted. There is no proteomic evidence for the N-terminal region of the FLABPs, therefore we cannot be sure which is the true start codon [18]. Given the conserved type I topology of the FLAs and FLABPs from other species, we predict that the FLABPs also have a single transmembrane domain located towards their C terminus, with an N-terminal signal peptide (Figure 1D,F).

Glossary

BARP: initially called bloodstream alanine-rich proteins, BARPs are the surface coat proteins found on the epimastigote form of *T. brucei*.
Bloodstream forms (BSFs): life-cycle stage of the African trypanosomes found in the blood of mammals, with a VSG coat.
Epimastigote: a type of trypanosome morphology defined by the positioning of the kinetoplast (mitochondrial DNA) to the anterior of the nucleus with a long-attached flagellum. This morphology is often found during the insect stages of the parasite life cycle.
Flagellum attachment zone (FAZ): large cytoskeletal structure that connects the cell body cytoskeleton to the flagellum through the cell body and flagellum membranes.
N-linked: linkage of a glycan to a protein via its *N*-acetylglucosamine to an asparagine.
P-linked: glycans bound to proteins via a phosphodiester linkage.
Procyclic form (PCF): life-cycle stage of *T. brucei* and *T. congolense* found in the midgut of the tsetse fly, with a procyclin coat.
Procyclin: major surface coat protein of procyclic forms. There are two classes of procyclins, EP and GPEET, composed of repeating units of EP or GPEET, respectively. Initially both forms are expressed and, 7–9 days after tsetse infection, only EP procyclin covers the parasite surface.
Trypomastigote: a type of trypanosome morphology defined by the positioning of the kinetoplast (mitochondrial DNA) to the posterior of the nucleus with a long-attached flagellum. This morphology is often found during the mammalian stages of the parasite life cycle.
Variant surface glycoprotein (VSG): surface coat protein of the African trypanosomes' bloodstream and metacyclic stages. The process of antigenic variation through VSG monoallelic expression, with the stochastic switching to a new VSG subtype, enables a persistent infection of the mammalian host.

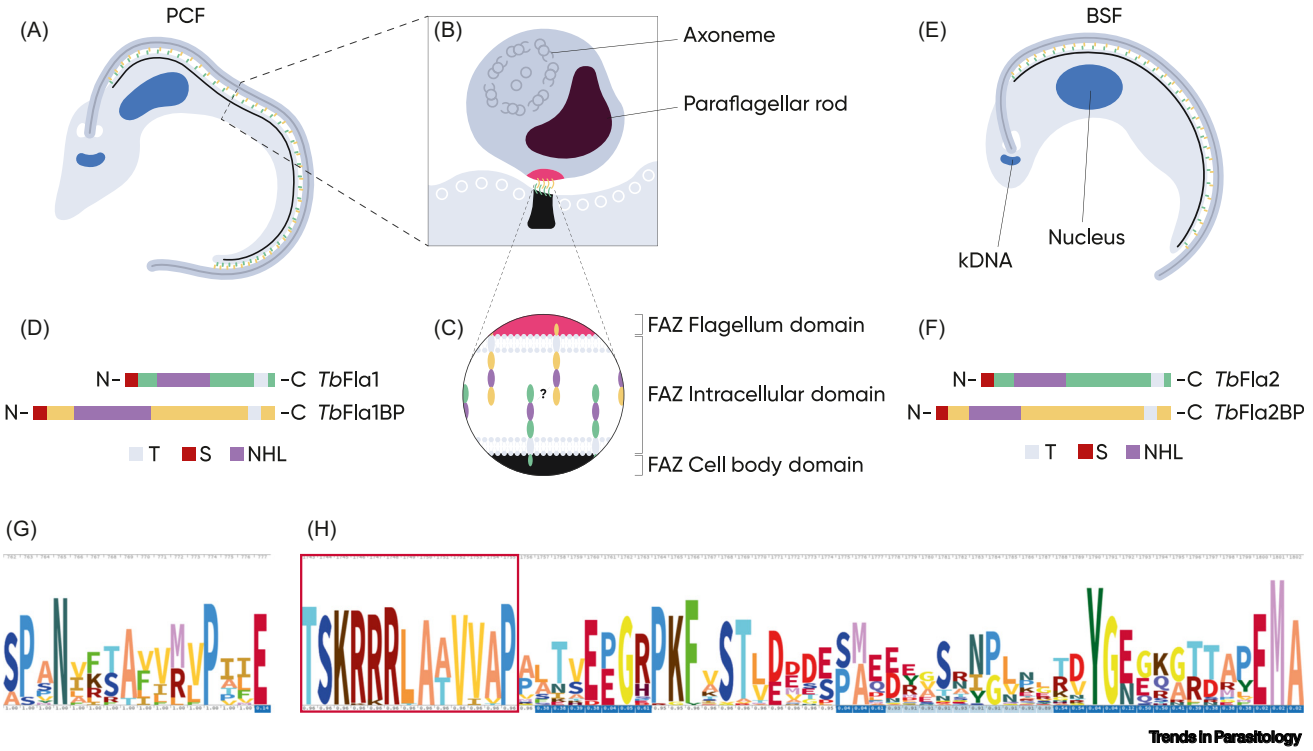

Figure 1. Flagellar adhesion glycoproteins laterally attach the flagellum to the cell body in *Trypanosoma brucei*. (A) Cartoon of a procyclic form (PCF), with FLA1 (green) and FLA1BP (yellow) localised to the cell and flagellum membranes, respectively, interacting through their extracellular region based on Sun *et al*. [6]. (B) Cartoon of the transverse section of flagellum attachment zone (FAZ), in which FLA1 and FLA1BP connect the flagellum and cell body FAZ domains. While FLA1 and FLA1BP interact through their extracellular domains, their intracellular domains connect into the cell body and flagellum FAZ domains, respectively. (C) Enlarged view from (B), highlighting FLA1 and FLA1BP. NHL protein domain is represented in purple. The specific binding mechanism of FLA1 and FLA1BP is unknown (question mark). (D) Cartoon of PCF FLA1 and FLA1BP highlighting NHL protein domain (NHL), transmembrane (T) and signal peptide (S) regions. (E) Cartoon of a bloodstream form (BSF) of *T. brucei*, with FLA2 (green) and FLA2BP (yellow) localised to the cell and flagellum membranes, respectively, interacting through their extracellular region. (F) Cartoon of BSF FLA2 and FLA2BP. (G) Logo of the intracellular protein sequence from FLAs, created with SkyLign tool. (H) Logo of the intracellular protein sequence from FLABPs showing the conserved motif (red box). The logo was created with SkyLign tool. See also Figure S1.

NHL repeats form a six-bladed β-propeller structure and are found in teneurins, adhesion receptors involved in vertebrate cell–cell interaction [19,20]. In addition to being important for metazoan multicellularity, teneurins are widely distributed in bacteria and are considered to be an ancient structure important for cell–cell interaction [19]. The teneurin NHL repeats mediate homophilic interactions; therefore, it is likely that the NHL domains of the FLAs and FLABPs enable their interaction.

While the extracellular region of FLA and FLABP is required for the FLA1–FLA1BP interaction, the intracellular domain is likely important for directing and/or anchoring the proteins into the FAZ [6]. The intracellular domain of the FLAs across different species is only ~15 amino acids long, while the FLABPs are longer at 31–48 amino acids. In the FLA intracellular domain, there is a significant sequence conservation of residues, with serine, proline, and asparagine at specific positions from the transmembrane domain (Figure 1G). For FLABP, there is a 13 amino acid region adjacent to the transmembrane domain with a highly conserved sequence, including the positively charged motif KRRR (Figure 1H). This conservation of positive residues likely occurs to maintain FLABP topology, as described by the positive-inside rule [21]. However, intriguingly, lysine residues were necessary for the localisation of calflagelin to the *T. cruzi* flagellum, and the intracellular domain of *T. brucei* FLA1BP is sufficient to direct eYFP to the flagellum [22]. Finally, all the

described FLAs and FLABPs are glycosylated and earlier work highlighted that the FAZ in both PCF and BSF was able to bind lectins (Box 1) [6,7,17,23,24]. However, the role of the FLA/FLABP glycans needs to be clarified. In sponges, glyconectin-mediated cell adhesion, essential for multicellularity, depends on glycans [25].

The early FAZ descriptions noted that the flagellum-to-cell body connection is not continuous but consists of discrete junctional complexes spaced along the FAZ and are seen in both longitudinal and transverse sections [3,26]. Later freeze-fracture and cryo-electron tomography showed that these complexes are comprised of a conglomerate of proteins in the membrane called staples [26,27]. The staples appear to act as a series of 'press-studs' connecting the flagellum to the cell body and are likely formed from FLAs and FLABPs. Understanding how these proteins are corralled into discrete microdomains along the FAZ is critical to dissecting FAZ assembly and function. The presence of these discrete staples raises the question of whether this arrangement of proteins provides a more resilient attachment between the cell body and flagellum than a narrower array of particles distributed along the length of the FAZ.

Why might cell-type-specific FLA–FLABP pairings evolve?

The trypanosomatid FLAs and FLABPs mediate the connection between the flagellum and cell body in the context of other cell-surface proteins, both variant and invariant. The *T. brucei* life cycle is characterised by different forms and different major coat proteins as it cycles between the tsetse fly vector and the mammalian host. In the mammalian host, the parasite is covered by **variant surface glycoproteins** (**VSGs**) [28] (Figure 2A), which are replaced by **procyclins** in the midgut and proventricular forms in the tsetse fly [29] (Figure 2). In the salivary glands, the epimastigotes express **BARP** [30], with the infective metacyclic forms expressing VSGs [31]. Other cell-surface proteins are present in these different life-cycle stages, but we focus on the dominant surface proteins for simplicity. If we map the expression of the FLA/FLABP variants, we see that FLA1/FLA1BP is associated with the insect stages and procyclin and BARP expression, while FLA2 and FLA2BP are associated with the mammalian stages and VSGs [12–15]. Thus, an appealing hypothesis is that the evolution of multiple FLA–FLABP combinations is related to ensuring a strong cell body-to-flagellum connection across a range of different surface environments.

In *T. brucei*, both the flagellum and cell body membranes are covered with surface proteins, and this leads to four simple scenarios through which FLA–FLABP pairings might interact so enabling the attachment of the flagellum to the cell body across the cell coat architecture (Figure 2A–C):

(i) FLA and FLABP must extend beyond the surface proteins;
(ii) Surface proteins interdigitate;
(iii) Surface proteins are excluded from the FAZ;
(iv) Surface proteins deform, enabling the membranes to be brought together for FLA–FLABP binding.

Box 1. Glycosylation of FLA and FLABP

FLA1 and FLA2BP have N-linked glycans, whereas FLA2 in BSF *T. brucei* has **N-linked** and potentially the rarer **P-linked** glycans [6,7]. Only the carbohydrate structure P-linked via a phosphodiester linkage to GP72 (FLA ortholog) in *T. cruzi* **epimastigote** forms has been characterized in detail [56]. The sugar composition of the *T. brucei* FLAs and FLABPs glycans is unknown but is likely distinct from *T. cruzi* GP72 as *T. brucei* cannot synthesize xylose, rhamnose, or galactofuranose [57]. In *T. brucei* PCFs, the depletion of GDP-mannose dehydratase (*Tb*GMD), the first enzyme in the fucose synthesis pathway, caused flagellum detachment [58]. Moreover, the presence of fucose on *T. cruzi* GP72 suggests that this sugar is potentially important for FLA function in *T. brucei* [58].

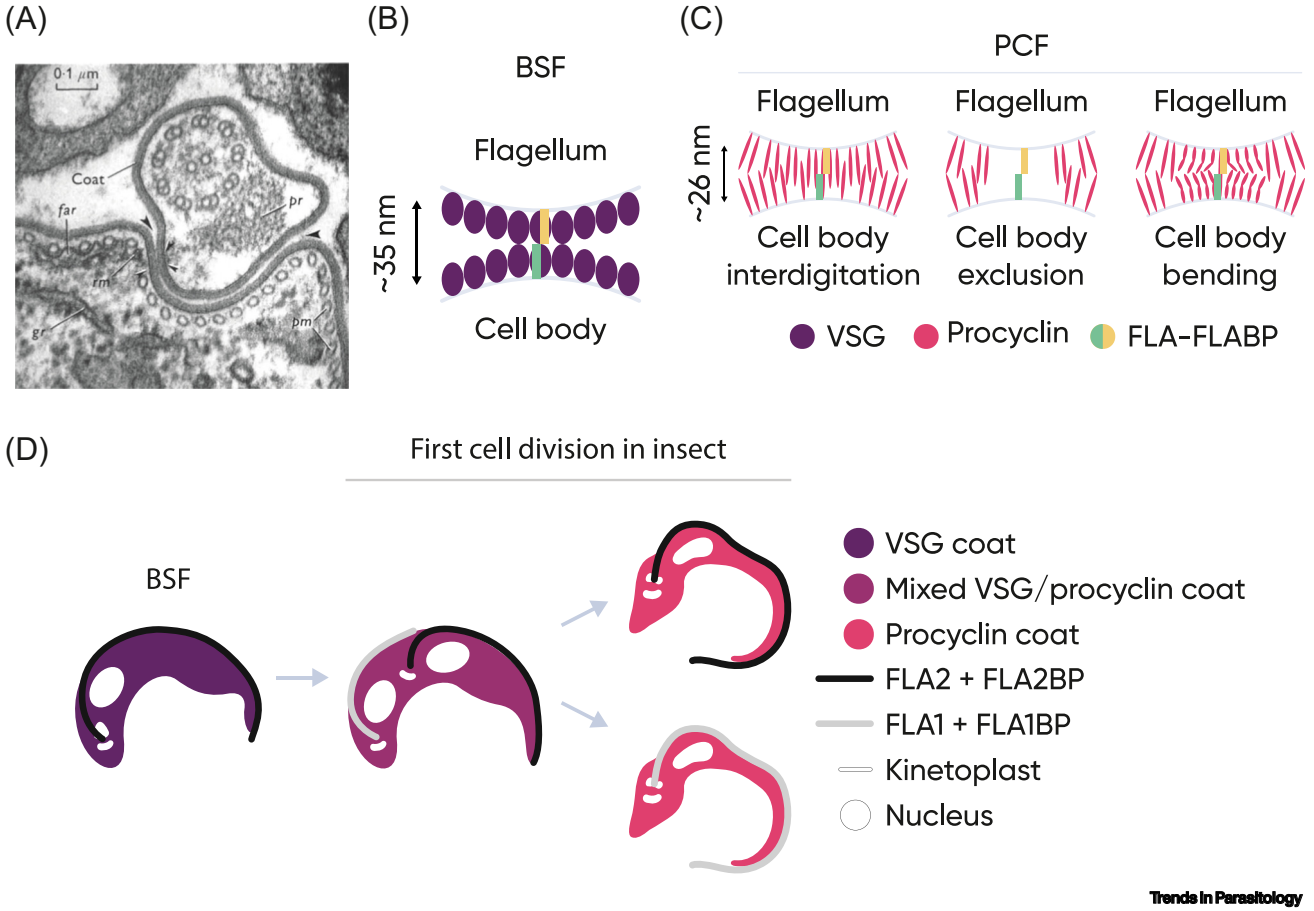

Figure 2. Surface coat architecture and FLA/FLABP evolution and expression. (A) Transverse section of a bloodstream form (BSF) from *Trypanosoma brucei rhodesiense* from Vickerman [3]. There is a uniform variant surface glycoprotein (VSG) coat covering the flagellum and cell body, even within the flagellum attachment zone (FAZ) region (small black arrowheads). (B) Cartoon of the intermembrane space between the flagellum and cell body of BSF from *T. brucei*, showing the presence of VSG in the FAZ region. (C) Representation of the space between the flagellum and cell body of procyclic form (PCF) *T. brucei*, representing the three possible models of interaction between FLAs/FLABPs and major surface coat proteins: interdigitation, exclusion, and bending. (D) Cartoon of the surface coat changes and FLA/FLABP pairings in *T. brucei* cell during its differentiation from BSFs to PCFs and its first cell division post-differentiation.

These models are not mutually exclusive, with a combination potentially occurring. Recently, it has been suggested that the gap between the flagellum and cell body membranes in a BSF cell from published micrographs is narrow and similar between BSF and PCF cells [32]. However, while acknowledging the caveats associated with electron microscopy sample preparation and imaging, our measurements of the gap in a BSF cell from electron micrographs is ~35 nm [3,33], (Figure 2A) and, given that the height of an individual VSG is ~16 nm [34], interdigitation of the VSG proteins is unlikely to occur (Figure 2A,B) [3]. Furthermore, the high density of the VSG coat means that there is little scope for an exclusion zone around the FAZ [3], and at the resolution of electron microscopy there is no visible disruption to the VSG coat (Figure 2A). This suggests that VSG is not excluded from the FAZ, and that FLA2 and FLA2BP likely extend beyond the VSG to bind [3,23].

Conversely, the gap between the flagellum and cell body membrane in PCFs is ~26 nm [26,27] and procyclins are modelled to have an elongated linear form up to 23 nm long [35]. This suggests that procyclin is either excluded from the FAZ, interdigitates, or deforms to enable

FLA1–FLA1BP binding (Figure 2C). Procyclin was not detected between the flagellum and cell body when live PCFs were labelled with anti-procyclin antibodies [36], which hints at the exclusion of procyclin; however, simply might be due to lack of antibody access.

In the above context, we note a difference between FLA1 and FLA2, and FLA1BP and FLA2BP in their overall length. FLA2s have a ~40 amino acid proline-rich insertion just after the NHL domain, and there are ~70 additional amino acids in FLA2BPs within the extracellular portion, close to the transmembrane domain. The AlphaFold model of these additional residues shows that they are flexible and unstructured, suggesting that the FLA2–FLA2BP could span a larger gap than FLA1–FLA1BP (Figure 3A,B). This correlates with the larger gap between the membranes seen in BSFs by electron microscopy [3]. Hence, evolution of the FLA/FLABP protein structures

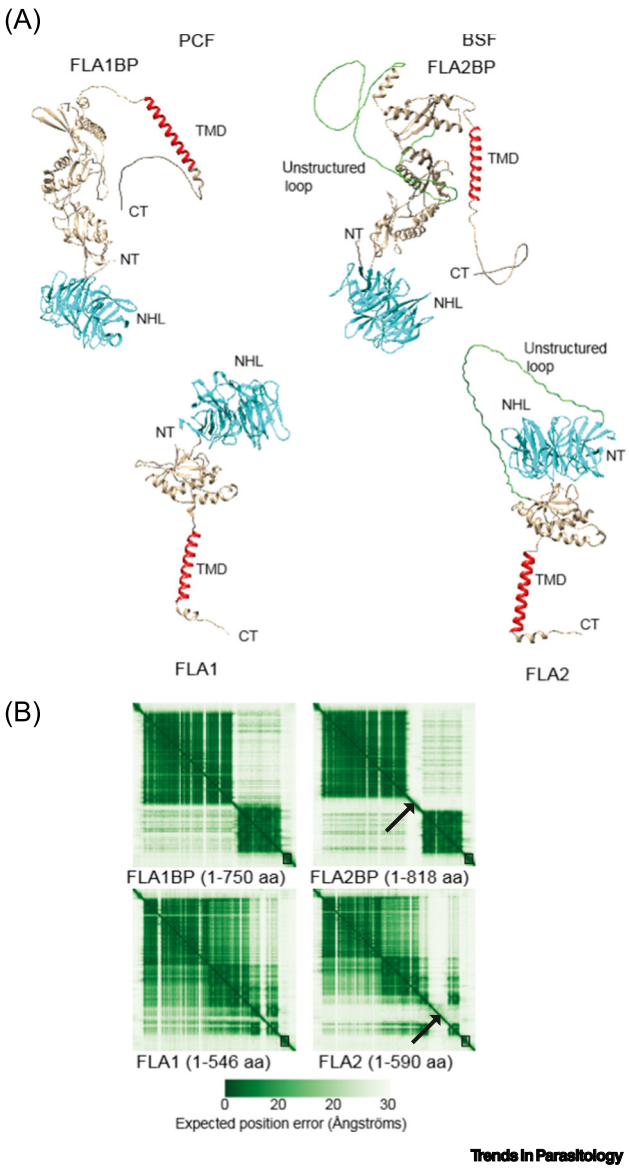

Figure 3. AlphaFold models of FLA2 and FLA2BP have an unstructured loop in the extracellular region. (A) AlphaFold models of bloodstream form (BSF) FLA2 and FLA2BP, with the unstructured loop in green, and procyclic form (PCF) FLA1 and FLA1BP [64]. The FLA/FLABP protein models are positioned next to each other for illustrative purposes and is unlikely to represent their interaction, as the binding mechanism is unknown. Abbreviations: CT, C terminus; NHL, NHL domain (blue); NT, N terminus; TMD, transmembrane domain (red). The signal peptide was removed to represent the mature protein. (B) Predicted aligned error (PAE) plots from each protein model. The amino acids of the proteins run along the vertical and horizontal axes, with the colour indicating the PAE value for that pair of amino acids. The black box indicates the predicted TMD and the arrow indicates the unstructured loop of FLA2 and FLA2BP. Note that the PAE plots indicate the regions either side of the TMD are flexible. In the predicted AlphaFold model this flexibility has resulted in the extracellular region lying alongside the TMD, which would not reflect the situation in the FAZ.

looks to have been influenced by the need to maintain attachment in different cell-surface coat architectures.

In addition to life-cycle changes in FLA and FLABP expression, there is evidence that the glycosylation patterns are also different, with N-linked glycans on FLA1 in PCFs and N/P-linked glycans on FLA2 in BSFs [7]. This will influence their interactions not only with the surrounding environment but also with their cognate-binding partner. The composition of sugars linked to *T. cruzi* GP72 also varies between life-cycle forms characterised by different surface coats [37,38]. In *T. cruzi*, these glycans are important for differentiation, and incubating epimastigotes with a monoclonal specific to the GP72 glycan blocked their differentiation to **trypomastigotes** [39].

However, this concept of a specific FLA–FLABP pairing for specific surface proteins needs to be more sophisticated to deal with life-cycle transitions. The FAZ is a stable structure, and during the cell cycle, a new FAZ is assembled alongside the new flagellum, with one daughter cell inheriting the old flagellum/FAZ and the other inheriting the new flagellum/FAZ [40]. This means in the context of differentiation, for example from BSF to PCF, after the first division post differentiation, one daughter cell will inherit a FLA1–FLA1BP-positive FAZ and the other a FLA2–FLA2BP-positive FAZ (Figure 2D). This highlights two important aspects of the FLA–FLABP connection. Firstly, the connections between the different FLA–FLABP pairs can operate in the mixed coat environment of an intermediate cell as VSG is replaced by procyclin [36]. Secondly, the FLA2–FLA2BP connection is likely maintained in a procyclin coat context until these proteins are turned over, or the cell dies.

Evolution of FLA and FLABP

The evolution of FLA and FLABP provides insights into their function, as flagellum attachment to the cell body and the switching of surface coats during the life-cycle transitions are common features of the kinetoplastid parasites. All analysed kinetoplastids except *Bodo saltans* have at least one *FLA* and *FLABP* ortholog present as a syntenic pair (Table 1, Figure 4, Table S1). Orthologs to both *Tb*FLA1 and *Tb*FLA1BP are present in *B. saltans*, but these genes are not on the same contig, and there is no conservation of synteny with the adjacent genes. Two strains of *T. cruzi* (TCC and CL) have additional *FLA* and *FLABP* genes; however, these are hybrid strains, and their genomes are a combination of two haplotypes, with each gene belonging to one haplotype [41,42].

In addition to *T. brucei*, only *T. congolense* has multiple orthologs of *FLA* and *FLABP*. In both the *T. congolense* IL3000 reference genome [43] and the more recent 2019 genome [44], there are three *FLA1* orthologs, and these are on chromosome 8, except for one gene (TcIL3000_0_31510) in the reference genome, which is on an unassigned contig. Moreover, genes adjacent to *FLA1*/*FLA1BP* are duplicated on chromosome 8 of both *T. brucei* and *T. congolense* (Figure 4A, gene numbers 47 and 48). This has the hallmarks of segmental duplication which occurred in the ancestor of *T. brucei* and *T. congolense*. However, a phylogenetic analysis of the FLA sequences shows that the *T. brucei* and *T. congolense* FLA orthologs group separately on the tree and this is potentially due to concerted evolution mediated by gene conversion after speciation (Figure 4B).

The exact number of *FLA1BP* genes in *T. congolense* is unclear, with variation between the different genome assemblies. The genome reference strain [43] has three *FLA1BP* orthologs, each on an unassigned contig. One of these (TcIL3000_0_31520) is located downstream of a *FLA1* ortholog, matching the syntenic organisation seen in other kinetoplastids. In the 2019 genome [44], there is only one *FLA1BP* ortholog, and this is located on chromosome 5 and potentially more closely related to *T. brucei FLA2BP*. However, upstream of the *FLA1* gene on chromosome 8, a pseudogene (TcIL3000.A.H_000601100) encodes a truncated *FLA1BP* ortholog.

Table 1. FLA and FLABP orthologs in Euglenozoa[a]

Species	Strain	TriTrypDB accession numbers	
		FLA	FLABP
Angomonas deanei	Crithidia deanei Carvalho (ATCC PRA-265)	ADEAN_001018200	ADEAN_001018300
Blechomonas ayalai	B08-376	Baya_022_0190	Baya_022_0200
Bodo saltans	Lake Konstanz	BSAL_89220	BSAL_00605
Crithidia fasciculata	Cf-Cl	CFAC1_040014300	CFAC1_040014200
Endotrypanum monterogeii	LV88	EMOLV88_100006800	EMOLV88_100006700
Leishmania braziliensis	MHOM/BR/75/M2904	LbrM.10.0770	LbrM.10.0760
Leishmania mexicana	MHOM/GT/2001/U1103	LmxM.10.0630	LmxM.10.0620
Leishmania tarentolae	Parrot-TarII	LtaP10.1050	LtaP10.1040
Leptomonas pyrrhocoris	H10	LpyrH10_22_0970	LpyrH10_22_0960
Paratrypanosoma confusum	CUL13	PCON_0078160	PCON_0078170
Trypanosoma brucei brucei	TREU927	Tb927.8.4010 Tb927.8.4060 Tb927.8.4110	Tb927.5.4570 Tb927.5.4580 Tb927.8.4050 Tb927.8.4100
Trypanosoma congolense	IL3000	TcIL3000_8_3780 TcIL3000_8_3810 TcIL3000_0_31510	TcIL3000_0_31520 TcIL3000_0_35140 TcIL3000_0_17090
Trypanosoma cruzi	CL Brener Esmeraldo-like	TcCLB.503571.10	TcCLB.503571.19
Trypanosoma theileri	isolate Edinburgh	TM35_000251760	TM35_000251750
Trypanosoma vivax	Y486	TvY486_0803430	TvY486_0803420

[a]Orthologous genes were identified by OrthoMCL and confirmed by reciprocal best BLAST analysis.

All the species examined except *B. saltans* have a single pair of *FLA* and *FLABP* that have maintained synteny. In *T. brucei* and *T. congolense*, this syntenic *FLA1* and *FLA1BP* pairing was on chromosome 8; therefore, the *FLABP* genes on chromosome 5 are likely paralogs derived from an ancestral copy of *FLA1BP* on chromosome 8. Overall, it appears that there has been a two-stage expansion of the *FLABP* genes in *T. brucei* and *T. congolense*, with a segmental duplication occurring on chromosome 8, with an additional duplication of *FLABP* onto chromosome 5. This represents a rapid diversification of the *FLABPs*, suggestive of an adaptive radiation. The phylogenetic analysis of the *FLABP* gene family supports the two-stage expansion as the chromosome 8 *FLABP* genes group independently from those on chromosome 5 (Figure 4C). Moreover, it is unclear if the duplication onto chromosome 5 occurred independently in *T. congolense* and *T. brucei*, or if this happened in their common ancestor (Figure 4C). If the *FLABP* duplication onto chromosome 5 occurred independently, this would support the idea that the segmental duplication on chromosome 8 occurred first as this was present in the *T. brucei* and *T. congolense* ancestor.

FLA/FLABP gene expansion is associated with maintaining three major surface coat proteins

The expansion of the *FLA/FLABP* gene family was restricted to the *T. brucei* and *T. congolense* ancestor, raising the question of what contributed to this expansion. A simple explanation based on changes to the surface coat is unlikely sufficient as many trypanosomatid species including *T. cruzi* and *Leishmania* have complex life cycles, alternating between an insect vector and a mammalian host, with associated changes in the composition of their surface coat [45]. Yet for both *T. cruzi* and *Leishmania* there is only a single FLA/FLABP pair, with little change in the transcript abundance for these proteins during the life cycle [46,47]. Nor is there a major

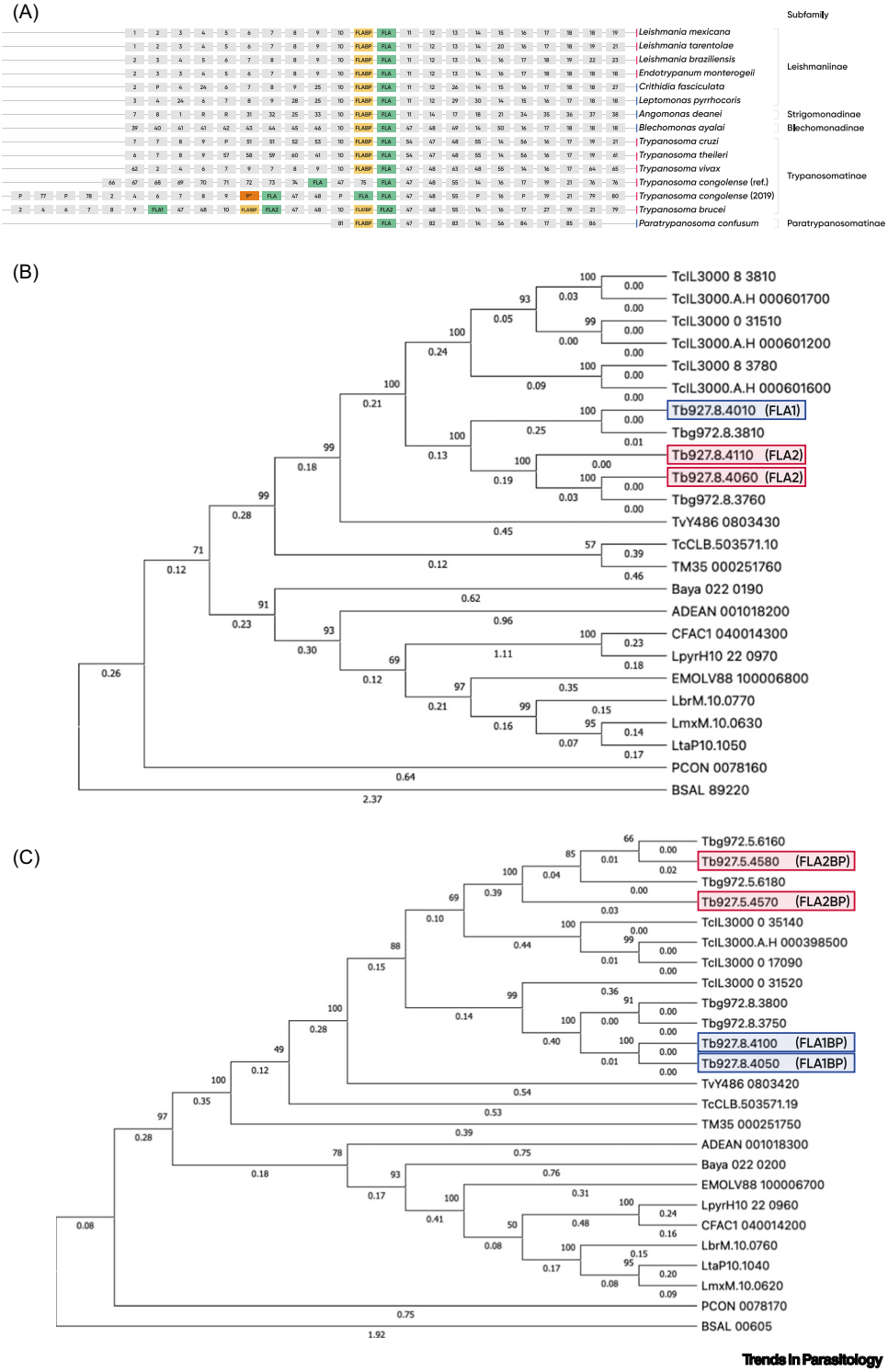

Figure 4. Genomic organisation of FLA and FLABP is conserved among trypanosomatids. (A) Schematic representation of the genomic organization of FLA and FLABP genes in comparison to chromosome eight from *Trypanosoma brucei*. Species are coloured according to their life cycle type [1]: pink, dixenic; blue, monoxenic. Numbers

(Figure legend continued at the bottom of the next page.)

correlation with overall cell shape. *T. cruzi* has an extended FAZ similar to *T. brucei* along which GP72 (*T. cruzi* FLA) is localised, while *Leishmania* has only a short region of flagellum attachment within the flagellar pocket neck region, where FLABP has been shown to localise [4,48].

A major difference between *Leishmania* and *T. cruzi* and the African trypanosomes is that the latter are exclusively extracellular parasites within the mammalian host, while the former both have an intracellular amastigote form. The ability of African trypanosomes to infect a vertebrate host as an extracellular parasite, with the associated continual exposure to the immune system, has likely contributed to the *FLA/FLABP* gene expansion. However, this expansion was restricted to the ancestor of *T. brucei* and *T. congolense* and was not seen in the related African trypanosome *Trypanosoma vivax*. All the African trypanosomes have similar life cycles; however, these parasites follow different routes through the fly (Box 2) [49]. In simple terms, *T. brucei* and *T. congolense* have three major coat proteins [VSG, procyclin, and BARP/glutamic acid/alanine-rich protein (GARP)], whereas *T. vivax* only has two (VSG and Fam50); the ancestral African trypanosome most likely had a midgut stage with a procyclin coat and this has been subsequently lost by *T. vivax* [49]. We suggest that the FLA–FLABP interaction can balance the pressure of operating in two different coat environments but not three. One solution to this problem is to simplify the life cycle, reducing the number of different coats required, as seen with *T. vivax*. Alternatively, the evolution of additional FLA–FLABP pairings can maintain flagellum attachment with an additional coat environment. However, it must not be forgotten that the surface coats of these parasites do not simply consist of a single protein and there are multiple different proteins present. For example, the surface coat of BSF *T. vivax* has ~15 protein families in addition to VSG and overall appears less dense than the *T. brucei* and *T. congolense* coat, which will impact on the selection pressure on FLA/FLABP.

In *T. brucei* BSFs, flagellum attachment is maintained by FLA2 and FLA2BP, which have evolved to operate in the VSG coat environment. The FLA/FLABP combinations expressed in the different *T. congolense* life-cycle stages are unknown and even if such regulation occurs is not clear from the available transcriptomic and proteomic datasets [50,51]. It is, therefore, not currently possible to assess which of these proteins are present in the different surface environments. In *T. vivax*, the ancestral pairing of FLA–FLABP is used, which must operate in different coat environments. Does this cause consequential compromises to the VSG organisation within the FAZ? A recent *T. vivax* vaccine candidate recognised an invariant protein that localised along the interface of the flagellum and cell body, suggesting increased accessibility to antibodies in this region [52]. Given the ~100 proteins localised to the FAZ in our recent work [8], we suggest that some of these may present additional vaccine candidates due to exposure at the FAZ interface. Moreover, the glycosylation patterns must not be discounted as these change during the life cycle of the kinetoplastid parasites, influencing the binding characteristics of the FLAs and FLABPs [7].

The surface of African trypanosomes is protein-rich, whereas in other kinetoplastid parasites such as *T. cruzi* and *Leishmania* spp. the surface is dominated by lipid phosphoglycans [53], creating a very different surface environment in which FLA–FLABP binding occurs. The expression of GP72 (*T. cruzi* FLA1) in *T. brucei* PCFs had a dominant negative effect, causing new flagellum

represent genes from different ortholog groups. Each corresponding gene and its genomic localisation is listed in Table S1. P, pseudogene; P* (orange), pseudogene with sequence similarity to FLA1BP; R, RNA coding gene. For *Trypanosoma congolense*, both reference genome (ref.) [43] and 2019 [44] genome assembling were used. (B) Phylogenetic tree of FLA proteins from trypanosomatid species. The topology was constructed using the maximum likelihood method with Mega11. The numbers next to the branches represent the percentage of trees in which the associated taxa clustered together. FLA1 (procyclic form, PCF) is highlighted in blue, while FLA2 (bloodstream form, BSF) is in red. (C) Phylogenetic tree of FLABPs proteins from trypanosomatids. FLA1BPs (PCF) are in blue, while FLA2BPs (BSF) are highlighted in red. The topology was constructed using the maximum likelihood method with Mega11.

Box 2. Surface coat changes associated with different trypanosome life cycles

After parasite ingestion, *T. brucei* and *T. congolense* differentiate into PCFs within the tsetse midgut, while *T. vivax* lacks the procyclic stage and directly differentiates into epimastigotes and metacyclic forms in the tsetse fly proboscis [59–61]. *T. congolense* PCFs subsequently reach the proboscis, proliferating as epimastigotes before differentiating into metacyclic forms [62]. Conversely, *T. brucei* migrates to the salivary glands and proliferates as an epimastigote before differentiating into metacyclic forms [63]. The different forms of the African trypanosomes in the tsetse fly have different surface coat proteins and surface architectures that will influence the evolution of FLA and FLABP (Table I). *T. brucei* and *T. congolense* PCFs have a procyclin surface coat, and as *T. vivax* does not have this stage, it does not encode any procyclin genes. The epimastigotes of all three species are associated with Fam50 surface proteins – BARP in *T. brucei* and GARP in *T. congolense*. The metacyclic forms and BSFs of all three are covered in variants of VSG.

Table I. Flagellar glycoproteins and surface coat proteins in African trypanosomes[a]

Host	Location	*T. brucei*		*T. congolense*		*T. vivax*	
		Stage	Coat	Stage	Coat	Stage	Coat
Mammal	Blood	BSF	VSG	BSF	VSG	BSF	VSG
Tsetse	Midgut	PCF	Proc	PCF	Proc	–	–
	PV	Long Tryp	Proc	Long Tryp	Proc	Long Tryp	?
	SG	Epi/Met	BARP/VSG	–	–	–	–
	Proboscis	–	–	Epi/Met	GARP/VSG	Epi/Met	Fam50/VSG

[a]Stage and coat surface of *T. brucei*, *T. congolense*, and *T. vivax* within mammals hosts and tsetse vectors. Abbreviations: BSF, bloodstream forms; Epi, epimastigote; Long Tryp, long trypomastigote; Met, metacyclic form; PCF, procyclic forms; Proc, procyclin; PV, proventriculus; SG, salivary glands; ?, unknown.

Outstanding questions

Do FLAs and FLABPs oligomerise with themselves, and how is the spacing between the 'press-studs' determined?

How does the FAZ and attached flagellum affect the flow of VSG and receptors over the surface of the parasite?

Are invariant surface proteins more accessible to antibodies within the FAZ region?

detachment, with the existing flagella remaining attached [11]. Importantly, overexpression of FLA1 in PCF *T. brucei* had no effect on flagellum attachment [6]; therefore, we speculate that GP72 interferes with FLA1 binding to FLA1BP at the proximal end of the FAZ during assembly. This shows that FLA from other species appears unable to successfully interact with FLA1BP, potentially due to the different surface architectural contexts and/or divergence of the binding interface between GP72 and *T. brucei* FLA1.

Concluding remarks

Trypanosomatids that inhabit the blood invariably have a trypomastigote morphology, with a long attached flagellum [54], and how the immune system of the host interacts with the FAZ will therefore be of significance for parasite survival (see Outstanding questions). The cell must balance its ability to evade the immune response while maintaining flagellum attachment and therefore the set of invariant proteins within the FAZ are a potential Achilles heel for the parasite. Moreover, the FAZ forms a clear specialised membrane domain that will influence movement of proteins and other molecules across the cell surface of the parasite. Earlier work showed a flow of antibody-bound VSGs towards the flagellar pocket [55], and advances in microscopy technology will now help to determine the effect of the attached flagellum on VSG movement on the cell surface.

Acknowledgments

We would like to thank Professor Edmundo Carlos Grisard (Federal University of Santa Catarina) for critical reading of the manuscript. We would also like to thank Professor Keith Gull (University of Oxford) for many useful and insightful conversations. We also thank Dr Richard Wheeler (University of Oxford) for helping with the AlphaFold models. We thank the reviewers for their insights into the evolution of FLA and FLABPs. Laryssa Vanessa de Liz was a recipient of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) scholarship – Brazilian Government Agency. Laryssa Vanessa de Liz received a travel grant from The Company of Biologists Limited, sponsored by the Journal of Cell Science. Work in the laboratory of J.D.S. is supported by the Wellcome Trust (221944/Z/20/Z) and Leverhulme Trust.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.pt.2023.02.009.

References

1. Kostygov, A.Y. *et al.* (2021) Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 11, 200407
2. Sunter, J.D. and Gull, K. (2016) The flagellum attachment zone: 'the cellular ruler' of trypanosome morphology. *Trends Parasitol.* 32, 309–324
3. Vickerman, K. (1969) On the surface coat and flagellar adhesion in trypanosomes. *J. Cell Sci.* 5, 163–193
4. Halliday, C. *et al.* (2020) Role for the flagellum attachment zone in *Leishmania* anterior cell tip morphogenesis. *PLoS Pathog.* 16, e1008494
5. Sunter, J.D. *et al.* (2019) *Leishmania* flagellum attachment zone is critical for flagellar pocket shape, development in the sand fly, and pathogenicity in the host. *Proc. Natl. Acad. Sci.* 116, 6351–6360
6. Sun, S.Y. *et al.* (2013) An intracellular membrane junction consisting of flagellum adhesion glycoproteins links flagellum biogenesis to cell morphogenesis in *Trypanosoma brucei*. *J. Cell Sci.* 126, 520–531
7. Nozaki, T. *et al.* (1996) Characterization of the *Trypanosoma brucei* homologue of a *Trypanosoma cruzi* flagellum-adhesion glycoprotein. *Mol. Biochem. Parasitol.* 82, 245–255
8. Dean, S. *et al.* (2017) TrypTag.org: a trypanosome genome-wide protein localisation resource. *Trends Parasitol.* 33, 80–82
9. Snary, D. *et al.* (1981) Cell surface antigens of *Trypanosoma cruzi*: use of monoclonal antibodies to identify and isolate an epimastigote specific glycoprotein. *Mol. Biochem. Parasitol.* 3, 343–356
10. Cooper, R. *et al.* (1993) Deletion of an immunodominant *Trypanosoma cruzi* surface glycoprotein disrupts flagellum-cell adhesion. *J. Cell Biol.* 122, 149–156
11. LaCount, D.J. *et al.* (2002) *Trypanosoma brucei* FLA1 is required for flagellum attachment and cytokinesis. *J. Biol. Chem.* 277, 17580–17588
12. Siegel, T.N. *et al.* (2010) Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* 38, 4946–4957
13. Jensen, B.C. *et al.* (2014) Extensive stage-regulation of translation revealed by ribosome profiling of *Trypanosoma brucei*. *BMC Genomics* 15, 911
14. Vasquez, J.J. *et al.* (2014) Comparative ribosome profiling reveals extensive translational complexity in different *Trypanosoma brucei* life cycle stages. *Nucleic Acids Res.* 42, 3623–3637
15. Tinti, M. and Ferguson, M.A.J. (2022) Visualisation of proteome-wide ordered protein abundances in *Trypanosoma brucei*. *Wellcome Open Res.* 7, 34
16. Reis-Cunha, J.L. *et al.* (2018) Gene and chromosomal copy number variations as an adaptive mechanism towards a parasitic lifestyle in trypanosomatids. *Curr. Genomics* 19, 87–97
17. Woods, K. *et al.* (2013) Identification and characterization of a stage specific membrane protein involved in flagellar attachment in *Trypanosoma brucei*. *PLoS ONE* 8, e52846
18. Shimogawa, M.M. *et al.* (2015) Cell Surface proteomics provides insight into stage-specific remodeling of the host–parasite interface in *Trypanosoma brucei*. *Mol. Cell. Proteomics* 14, 1977–1988
19. Jackson, V.A. *et al.* (2018) Structures of Teneurin adhesion receptors reveal an ancient fold for cell-cell interaction. *Nat. Commun.* 9, 1079
20. Chaudhuri, I. *et al.* (2008) Evolution of the β-propeller fold. *Proteins* 71, 795–803
21. Baker, J.A. *et al.* (2017) Charged residues next to transmembrane regions revisited: 'Positive-inside rule' is complemented by the 'negative inside depletion/outside enrichment rule'. *BMC Biol.* 15, 66
22. Maric, D. *et al.* (2011) Molecular determinants of ciliary membrane localization of *Trypanosoma cruzi* flagellar calcium-binding protein. *J. Biol. Chem.* 286, 33109–33117
23. Brickman, M.J. and Balber, A.E. (1990) *Trypanosoma brucei rhodesiense* bloodstream forms: surface ricin-binding glycoproteins are localized exclusively in the flagellar pocket and the flagellar adhesion zone. *J. Protozool.* 37, 219–224
24. Woods, A. *et al.* (1989) Evidence for a Mr 88,000 glycoprotein with a transmembrane association to a unique flagellum attachment region in *Trypanosoma brucei*. *J. Cell Sci.* 93, 501–508
25. Misevic, G. and Garbarino, E. (2021) Glycan-to-glycan binding: molecular recognition through polyvalent interactions mediates specific cell adhesion. *Molecules* 26, 397
26. Sherwin, T. and Gull, K. (1989) The cell division cycle of *Trypanosoma brucei brucei*: timing of event markers and cytoskeletal modulations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 323, 573–588
27. Höög, J.L. *et al.* (2012) Cryo-electron tomography and 3-D analysis of the intact flagellum in *Trypanosoma brucei*. *J. Struct. Biol.* 178, 189–198
28. Morrison, L.J. *et al.* (2009) Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell. Microbiol.* 11, 1724–1734
29. Acosta-Serrano, A. *et al.* (2001) The surface coat of procyclic *Trypanosoma brucei:* Programmed expression and proteolytic cleavage of procyclin in the tsetse fly. *Proc. Natl. Acad. Sci. U. S. A.* 98, 1513–1518
30. Urwyler, S. *et al.* (2007) A family of stage-specific alanine-rich proteins on the surface of epimastigote forms of *Trypanosoma brucei*. *Mol. Microbiol.* 63, 218–228
31. Cross, G.A.M. *et al.* (2014) Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Mol. Biochem. Parasitol.* 195, 59–73
32. Trépout, S. (2020) *In situ* structural analysis of the flagellum attachment zone in *Trypanosoma brucei* using cryo-scanning transmission electron tomography. *J. Struct. Biol. X* 4, 100033
33. Vickerman, K. (1969) The fine structure of *Trypanosoma congolense* in its bloodstream phase. *J. Protozool.* 16, 54–69
34. Bartossek, T. *et al.* (2017) Structural basis for the shielding function of the dynamic trypanosome variant surface glycoprotein coat. *Nat. Microbiol.* 2, 1523–1532
35. Ferguson, M.A.J. (2000) Glycosylphosphatidylinositol biosynthesis validated as a drug target for African sleeping sickness. *Proc. Natl. Acad. Sci. U. S. A.* 97, 10673–10675
36. Roditi, I. *et al.* (1989) Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. *J. Cell Biol.* 108, 737–746
37. Haynes, P.A. and Cross, G.A.M. (1996) Differential glycosylation of epitope-tagged glycoprotein Gp72 during the *Trypanosoma cruzi* life cycle. *Mol. Biochem. Parasitol.* 83, 253–256
38. Pech-Canul, Á. de la C. *et al.* (2017) A brief view of the surface membrane proteins from *Trypanosoma cruzi*. *J. Parasitol. Res.* 2017, 1–13
39. Sher, A. and Snary, D. (1982) Specific inhibition of the morphogenesis of *Trypanosoma cruzi* by a monoclonal antibody. *Nature* 300, 639–640
40. Abeywickrema, M. *et al.* (2019) Non-equivalence in old- and new-flagellum daughter cells of a proliferative division in *Trypanosoma brucei*. *Mol. Microbiol.* 112, 1024–1040
41. Berná, L. *et al.* (2018) Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. *Microbial Genomics* 4, e000177
42. Wang, W. *et al.* (2021) Strain-specific genome evolution in *Trypanosoma cruzi*, the agent of Chagas disease. *PLoS Pathog.* 17, e1009254

43. Jackson, A.P. *et al.* (2012) Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proc. Natl. Acad. Sci.* 109, 3416–3421
44. Abbas, A.H. *et al.* (2018) The structure of a conserved telomeric region associated with variant antigen loci in the blood parasite *Trypanosoma congolense*. *Genome Biol. Evol.* 10, 2458–2473
45. de Castro Neto, A.L. *et al.* (2022) Role of virulence factors of trypanosomatids in the insect vector and putative genetic events involved in surface protein diversity. *Front. Cell. Infect. Microbiol.* 12, 807172
46. Smircich, P. *et al.* (2015) Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics* 16, 443
47. Fiebig, M. *et al.* (2015) Comparative life cycle transcriptomics revises *Leishmania mexicana* Genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLoS Pathog.* 11, e1005186
48. Wheeler, R.J. *et al.* (2016) Flagellar pocket restructuring through the *Leishmania* life cycle involves a discrete flagellum attachment zone. *J. Cell Sci.* 129, 854–867
49. Jackson, A.P. *et al.* (2013) A cell-surface phylome for african trypanosomes. *PLoS Negl. Trop. Dis.* 7, e2121
50. Awuoche, E.O. *et al.* (2018) Expression profiling of *Trypanosoma congolense* genes during development in the tsetse fly vector *Glossina morsitans morsitans*. *Parasit. Vectors* 11, 380
51. Eyford, B.A. *et al.* (2011) Differential protein expression throughout the life cycle of *Trypanosoma congolense*, a major parasite of cattle in Africa. *Mol. Biochem. Parasitol.* 177, 116–125
52. Autheman, D. *et al.* (2021) An invariant *Trypanosoma vivax* vaccine antigen induces protective immunity. *Nature* 595, 96–100
53. Ferguson, M.A.J. (1997) The surface glycoconjugates of trypanosomatid parasites. *Philos. Trans. R. Soc. Lond. B* 352, 1295–1302
54. Wheeler, R.J. *et al.* (2013) The limits on trypanosomatid morphological diversity. *PLoS ONE* 8, e79581
55. Engstler, M. *et al.* (2007) Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* 131, 505–515
56. Allen, S. *et al.* (2013) Structure of a complex phosphoglycan epitope from gp72 of *Trypanosoma cruzi*. *J. Biol. Chem.* 288, 11093–11105
57. Turnock, D.C. and Ferguson, M.A.J. (2007) Sugar nucleotide pools of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. *Eukaryot. Cell* 6, 1450–1463
58. Turnock, D.C. *et al.* (2007) The *de novo* synthesis of GDP-fucose is essential for flagellar adhesion and cell growth in *Trypanosoma brucei*. *J. Biol. Chem.* 282, 28853–28863
59. Osório, A.L.A.R. *et al.* (2008) *Trypanosoma* (*Duttonella*) *vivax*: its biology, epidemiology, pathogenesis, and introduction in the New World - a review. *Mem. Inst. Oswaldo Cruz* 103, 1–13
60. Jackson, A.P. *et al.* (2015) Global gene expression profiling through the complete life cycle of *Trypanosoma vivax*. *PLoS Negl. Trop. Dis.* 9, e0003975
61. Bargul, J.L. *et al.* (2016) Species-specific adaptations of trypanosome morphology and motility to the mammalian host. *PLoS Pathog.* 12, e1005448
62. Peacock, L. *et al.* (2012) The life cycle of *Trypanosoma* (*Nannomonas*) *congolense* in the tsetse fly. *Parasit. Vectors* 5, 109
63. Dyer, N.A. *et al.* (2013) Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends Parasitol.* 29, 188–196
64. Wheeler, R.J. (2021) A resource for improved predictions of *Trypanosoma* and *Leishmania* protein three-dimensional structure. *PLoS ONE* 16, e0259871