Crystal structure of the TbBILBO1 N-terminal domain reveals a ubiquitin fold with a long rigid loop for the binding of its partner

Keni Vidilaseris[1, 3, 4], Nicolas Landrein[2, 3], Yulia Pivovarova[1, 3], Johannes Lesigang[1], Niran Aeksiri[1, 5], Derrick R. Robinson[2], Melanie Bonhivers[2], and Gang Dong[1]

[1] Max Perutz Labs, Medical University of Vienna, Vienna BioCenter, 1030 Vienna, Austria

[2] Microbiologie Fondamentale et Pathogénicité, CNRS UMR 5234, Université de Bordeaux, Bordeaux, France

[3] Equal contribution

[4] Current address: Research Program in Molecular and Integrative Biosciences, University of Helsinki, Helsinki, Finland

[5] Department of Agricultural Sciences, Naresuan University, Thailand

*Running title: *Crystal structure of TbBILBO1-NTD*

To whom correspondence should be addressed: Gang Dong, Center for Medical Biochemistry, Medical University of Vienna, Dr. Bohr-Gasse 9/3, 1030 Vienna, Austria, Tel.: +43-1-4277-61625; Fax: +43-1-4277- 9616; E-mail: gang.dong@meduniwien.ac.at.

Keywords: BILBO1; crystallography; cytoskeleton; flagellar pocket collar; FPC4; parasite; protein structure; protein-protein interaction; structural biology; *Trypanosoma brucei*; ubiquitin fold

Background: BILBO1 is the core component of the flagellar pocket collar (FPC) in trypanosomes and recruits other FPC proteins including FPC4 to form this essential cytoskeleton-associated structure in kinetoplastids.

Results: We determined a high-resolution crystal structure of the N-terminal domain of *T. brucei* BILBO1, which revealed a characteristic horseshoe-like surface pocket for specific binding of *T. brucei* FPC4.

Conclusion: The C-terminal loop of the TbBILBO1 N-terminal domain adopts an unusual circular conformation to provide a unique contact site for TbFPC4 binding.

Significance: The crystal structure together with mutagenesis studies reveals for the first time how two FPC proteins interact with each other.

ABSTRACT

BILBO1 was the first characterized component of the flagellar pocket collar (FPC) in trypanosomes. The N-terminal domain (NTD) of BILBO1 plays an essential role in *T. brucei* FPC biogenesis and is thus vital for the parasite's survival. Here we report a 1.7-Å resolution crystal structure of TbBILBO1-NTD, which revealed a conserved horseshoe-like hydrophobic pocket formed by an unusually long loop. Mutagenesis studies suggested that another FPC protein, FPC4, interacts with TbBILBO1 via specific binding to this pocket. Overall, we have determined the exact binding site of TbFPC4 on TbBILBO1-NTD, which may provide a basis for rational drug design in the future.

INSTRODUCTION

Trypanosoma brucei is a protist parasite causing sleeping sickness and Nagana in Sub-Saharan Africa. At the base of its single flagellum is a bulb-like invagination of the plasma membrane called the flagellar pocket (FP), which is responsible for all endo-/exocytosis in the cell [1-3]. Around the neck of the FP on its cytoplasmic face is an electron-dense structure called the flagellar pocket collar (FPC), which is essential for FP biogenesis and thus for the survival of the parasite [4].

TbBILBO1 was the first reported protein component of the FPC [5]. It consists of four structural domains: a globular N-terminal domain; two central EF-hand motifs; a long coiled-coil domain; a C-terminal leucine zipper [6]. Analysis using electron microscopy showed that TbBILBO1 forms long filaments *in vitro* [6, 7]. Based on its modular architecture and intrinsic property to form spiral-like structures when ectopically expressed *in vivo*, TbBILBO1 was proposed to act as the structural core to recruit other components during FPC assembly [8, 9]. Another confirmed FPC component besides BILBO1 is FPC4, which is a microtubule binding protein that interacts with microtubule through its N-terminal part and with TbBILBO1-NTD via its C-terminal region [8]. However, how TbFPC4 binds to TbBILBO1 remains unclear.

We have previously reported an NMR structure of TbBILBO1-NTD (aa1-110) and identified a conserved surface patch essential for the function of TbBILBO1 in the parasite [10]. However, due to the limitation of the method, that structure had a relatively low resolution, particular for the ill-defined surface patch because of the lack of assigned signals of the C-terminal loop. It was also unclear to us at that time what the binding partner of this conserved patch was.

Here we report a 1.7-Å resolution crystal structure of TbBILBO1-NTD, which, with a longer sequence (aa1-175), revealed a unique horseshoe-like pocket harboring multiple conserved aromatic residues. This pocket is formed by an extended loop at the C-terminus of TbBILBO1-NTD, spanning residues 96-115. Structure-based mutagenesis studies suggested that TbFPC4 binds primarily to the three aromatic residues in the center of the horseshoe-shaped pocket. The binding interface also extends to the close peripheral area of the pocket. Overall, our work unveils a unique structural feature of TbBILBO1-NTD, including a long rigid loop and a conserved aromatic pocket, and sheds light on how TbFPC4 is docked at the pocket to achieve high binding affinity.

EXPERIMENTAL PROCEDURES

Cloning and site-directed mutagenesis – Cloning of full-length (aa1-587) and NTD-EFh (aa1-249) constructs of TbBILBO1 has been reported previously [10, 11]. For mutagenesis studies, a new TbBILBO1-NTD construct based on the crystal structure, which contains residues 1-120), was cloned

into the expression vector pET15b (Novogen) between *Nde*I and *BamH*I sites. The resulting construct provides an N-terminal 6×His tag, which is cleavable by the thrombin protease.

For TbFPC4, we cloned the sequence encoding its C-terminal domain (CTD, aa357-404), which was reported to be sufficient for TbBILBO1 binding [8], from genomic DNA into a custom vector SUMO15b. The vector provides an N-terminal 6×His-SUMO tag that is cleavable by the Sentrin-specific protease 2 (SENP2) protease.

Two single-residue mutants of TbBILBO1-NTD, Y64A and W71A, were generated by site-directed mutagenesis using a QuikChange kit (Stratagene) according to the manufacturer's instructions. Incorporation of mutations was confirmed by DNA sequencing.

Protein expression and purification –Recombinant TbBILBO1-NTD and TbFPC4-CTD proteins were expressed in *Escherichia coli* (strain BL21-DE3). Briefly, the cloned constructs for protein expression were used to transform competent bacterial cells. The cells were grown in Luria-Bertani (LB) medium at 37°C to an OD_{600} of 0.6-0.8, and then subjected to cold shock on ice for 30 min. Protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cultures were further incubated at 16°C overnight. Cells were harvested by centrifugation in a Sorvall GS3 rotor (6,000×g, 12 min, 4°C) and then resuspended in 20 ml of lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 20 mM imidazole, 5% (v/v) glycerol) per L of cell culture.

Bacteria were lysed in an EmulsiFlex-C3 homogenizer (Avestin) and cell debris was pelleted by centrifugation (40,000×g, 40 min, 4°C). The supernatant was filtered (0.45-µm pore size, Amicon) and loaded onto a Ni-HiTrap column (GE Healthcare) pre-equilibrated with the same lysis buffer. The column with bound proteins was washed with 5 × column volume (cv) of lysis buffer, and bound protein was eluted using a linear gradient concentration of imidazole (20 - 500 mM, 20×cv) in the same lysis buffer.

The N-terminal 6×His tag of TbBILBO1-NTD and the 6×His-SUMO tag of TbFPC4-CTD were cleaved off by incubating the pooled fractions of interest with ~2% (w/w) of thrombin and ~1% (w/w) of SENP2, respectively (4°C, overnight). Target proteins were further purified on a Superdex-200 16/60 column (GE Healthcare) pre-equilibrated with a running buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. The eluted proteins were used for subsequent binding tests or native gel electrophoresis. Protein concentration was determined using Ultraviolet (UV) absorbance at 280 nm.

Crystallization and structure determination – Crystallization of TbBILBO1-NTD and diffraction data collection have been reported [11]. In summary, the thrombin-digested product of selenium-methionine (SeMet) substituted TbBILBO1-NTD-EFh, which corresponded to residues 1-175, was used to set up crystallization trials. The resulting plate-like crystals were used to collect a highly redundant

single anomalous dispersion (SAD) dataset at the absorption edge of selenium (Se, λ = 0.9792 Å) on the ID14-4 beamline (ESRF). Structure determination was carried out using the SAD method exploiting the three selenium (Se) sites in the SeMet-substituted methionine residues (M1, M63, and M107). All Se sites were ordered and could be readily located, and experimental maps were calculated using AutoSol in the software suite Phenix [12], which built the majority of the structural model automatically. The model was further manually checked and rebuilt (for loops) using the program COOT [13]. Refinement was carried out using phenix.refine [12]. The final R_{work} and R_{free} are 0.167 and 0.198, respectively.

Isothermal titration calorimetry (ITC) – ITC measurements were performed on a MicroCal™ iTC200 microcalorimeter (GE Healthcare). Protein samples of TbBILBO1-NTD and TbFPC4-CTD were dialyzed overnight against 100 × volume of the binding buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. A typical ITC titration experiment consisted of 20 injections of TbFPC4 (300 µM), first with 1 × 0.2 µl and then with 19 × 2 µl, into a reaction cell filled with 200 µl of TbBILBO1-NTD (30 µM). All measurements were carried out under constant stirring at 350 rpm, and each injection lasted for 4 s with a 180 s interval between two injections. Titration peaks were analyzed using the Origin 7.0 Microcal software and corrected for the SAM/SFG dilution heat measured by injecting cofactors into a buffer containing no protein using the same protocol described above. Non-linear least squares fitting using one binding site model was used to calculate the association constant (Ka) and stoichiometry values. Dissociation constants (Kd) were calculated according to the formula Kd = 1/Ka.

Static light scattering (SLS) - SLS measurements were carried out by coupling size-exclusion chromatography (SEC) with mass determination. A 50-µl protein sample (1-2 mg/ml) was analyzed on a Superdex S-200 10/300 GL column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM dithiothreitol, and 1% (v/v) glycerol. Data analyses were carried out using ASTRA software provided by the manufacturer.

Electrophoretic mobility shift assay (EMSA) – All EMSA experiments were carried out on 5% (w/v) native polyacrylamide gels using Tris-acetate-EDTA (TAE) buffer containing 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA. The gels were run at 150 V for 2 h at 4°C. The TbBILBO1-NTD/TbFPC4-CTD complex purified by SEC was loaded directly on the native gel in the presence of 10% (v/v) glycerol. In all other binding tests, the separately purified protein samples (~1 mg/ml) were mixed in a 1:1 ratio and incubated overnight at 4°C before loading onto the native gel. Single protein samples with the same concentration as the complexes were loaded on the same gel as controls.

T. brucei cell culture and mutant protein production - The *in vivo* work in the parasite described in this study used the procyclic (PCF) *T. brucei* 427 29.13 co-expressing the T7 RNA polymerase and the tetracycline repressor [14, 15]. Cells were cultured at 27˚C in SDM79 medium (Sigma-Aldrich) containing 10% (v/v) heat-inactivated fetal calf serum, hemin 20 µg/ml, hygromycin 25 µg/ml, neomycin 10 µg/ml. Phleomycin (5 µg/ml) was also added for the cell lines expressing the Ty1-tagged

TbBILBO1 constructs [10]. 3×10^7 cells were transfected with 10 µg of *Not*I linearized plasmids using the AMAXA electroporator (Lonza, program X-001) as described [14], with the transfection buffer [16]. Ty1-tagged TbBILBO1 mutant cell lines - single mutant Y64A and W71A, and triple mutant D65K/E66K/E67K (DEE/KKK) - were generated by site-directed mutagenesis (Quick Change Lightning Site directed Mutagenesis Agilent Kit 210518) using the pLew100-TbBILBO1 plasmid and transfection of the Tb427 29.13 cell line [10]. Incorporation of mutations was confirmed by sequencing of the plasmids and of genomic DNA after cell line selection. Expression of Ty1-tagged WT-TbBILBO1 and mutant forms was induced by adding 20 ng/ml of tetracycline in the culture medium.

Mammalian cell culture - U-2 OS cells (human bone osteosarcoma epithelial cells, ATCC Number: HTB-96 [17]) were grown in D-MEM Glutamax (Gibco) supplemented with final concentrations of 10% fetal calf serum (Invitrogen), 100 units/ml of Penicillin (Invitrogen), and 100 µg/ml of Streptomycin (Invitrogen) at 37°C plus 5% CO_2. Exponentially growing U-2 OS cells in 24-well plates with glass coverslips were lipotransfected as described in [18] with 0.5µg of DNA using Lipofectamine 2000 in OPTIMEM (Invitrogen) according to the manufacturer's instructions, and processed for immunofluorescence imaging 24 hours post-transfection. To transiently express TbBILBO1 and its mutants in the heterologous mammalian system together with TbFPC4-GFP, we generated the pcDNA3-TbBILBO1-Y64A and pcDNA3-TbBILBO1-W71A mutant plasmids using the pcDNA3-TbBILBO1 plasmid [9] and site-directed mutagenesis as described above.

Immunofluorescence - Immunofluorescence on detergent-extracted U-2 OS cells was done as described in [8].

Western blotting - Trypanosome whole cell protein lysates (2×10^6 cells per well) were separated using SDS-PAGE (10% gels) and transferred by semi-dry blotting (BioRad) for 45 min at 25V to PVDF membranes. After 1 h blocking step in 5% skimmed milk in TBS, 0.2% Tween-20 (blocking solution BS), the membranes were incubated with the primary antibodies diluted in BS – anti-Ty1 (BB2 mouse monoclonal, 1:10,000 [19]) and anti-Enolase (loading control, rabbit polyclonal, 1:10,000 [20]). After one wash in BS, one wash in 1M NaCl and one wash in BS, the membranes were incubated with HRP-coupled secondary antibodies diluted 1:10,000 in BS (anti-mouse IgG Jackson 515-035-062, anti-rabbit IgG Sigma A-9169), washed twice in BS and twice in TBS, and visualized using the Clarity Western ECL Substrate kit (Bio-Rad) with an ImageQuant LAS4000. After stripping (two washes of 5 min in 100 mM glycine-HCl pH 2.3, 1% SDS, 0.1% NP40), washing in TBS then blocking in BS, the membranes were incubated with a rabbit polyclonal anti-TbBILBO1 (rabbit polyclonal [8]) and imaged as described above.

Accession code - Coordinates and structure factors of the TbBILBO1-NTD crystal structure have been deposited in the Protein Data Bank (PDB) under accession code **6SJQ**.

RESULTS

1.7-Å resolution crystal structure of TbBILBO1-NTD – As reported previously, TbBILBO1-NTD-EFh was expressed in bacteria and purified on three consecutive chromatographic columns to obtain highly homogeneous sample [11]. The thrombin-cleaved product of SeMet-substituted TbBILBO1-NTD-EFh, corresponding to residues 1-175, was crystallized after initial screening and subsequent optimization. The resulting crystals belonged to space group $P2_1$ (a = 29.69 Å, b = 50.80 Å, c = 37.22 Å; β = 94.61°). The structure was solved to 1.7-Å resolution employing the single-wavelength anomalous dispersion method, using data collected at the absorption edge of selenium (Table 1). The final structure contained residues 2-115 of TbBILBO1-NTD together with 135 tightly bound water molecules. The refined model has R_{work} and R_{free} of 16.9% and 19.8%, respectively.

The crystal structure revealed that TbBILBO1-NTD consists of five β strands (β1 - β5) and a single long α helix (Fig. 1A). The 2Fo-Fc electron density map is excellent for all residues in the structure as well as the bound water molecules (Fig. 1B). Notably, at this resolution one can easily spot the holes in the electron density of all aromatic rings as well as the proline residues. The first two β strands, β1 & β2, are much longer than β3-β5. These five β strands together form a slightly twisted β sheet, with the α-helix packed diagonally along the back side of the sheet (Fig. 1C). Overall the structure is topologically the same as the canonical ubiquitin fold, similarly to the previously reported NMR structure [10].

The C-terminal tail of TbBILBO1-NTD forms an extended loop – The NMR structure of TbBILBO1-NTD is similar to the crystal structure reported here. However, the C-terminal extension (aa101-110) of TbBILBO1-NTD was poorly defined in the NMR structure due to the missing long-range nuclear Overhauser effect (NOE) signals for the residues in that region, and was thus proposed to be structurally dynamic (Fig. 2A). Interestingly, the crystal structure showed a rigid loop that is unusually long, spanning residues 96-115, and wrapping around β1, β2 and the C-terminal part of the α helix (Fig. 2B). Although this loop is less rigid than the structural core as demonstrated by the slightly high temperature factors of the residues in this loop (Fig. 2C), it was clearly visible in the original experimental map and could be readily refined. Interaction of the residues in the loop with the core structure is stabilized by more than 20 hydrogen bonds, either directly or mediated by water molecules, and many hydrophobic interactions (Fig. 2D).

The C-terminal loop helps to generate a hydrophobic horseshoe-shaped pocket – In the crystal structure, there is a sharp kink at the beginning of the rigid loop to bend the loop from the longitudinally orientated C-terminus of β5 (relative to the long axis of the β sheet) to a horizontally arranged conformation. At the kink point is the highly conserved residue H96, whose side chain is completely buried in a deep pocket (Fig. 3A). This rigid, curved loop generates a horseshoe-like pocket, with one side of the pocket being shaped completely by the loop (Fig. 3B). This pocket was not well defined in

the previously reported NMR structure, mostly due to the poorly defined loop that was detached from the rest of the structure (Fig. 3C). The central part of the pocket contains three aromatic residues, W71, Y87 and F89, with another aromatic residue Y64 in the proximity of the pocket (Fig. 3D). There are also two positive charged residues, K60 and K62, at one side of the pocket, and residue Q100 from the loop on the opposite side of the pocket (Fig. 3E).

TbFPC4 binds to TbBILBO1-NTD mainly by contacting three central aromatic residues in the pocket – Previous studies have shown that several residues near the hydrophobic pocket are essential for the function of TbBILBO1 [10]. Later studies demonstrated that TbFPC4 binds to a similar region of TbBILBO1-NTD [8]. Based on these studies and the newly determined crystal structure of TbBILBO1-NTD, we used structure-based mutagenesis to define how TbFPC4 is orientated on TbBILBO1 in the formed complex.

We chose four sets of residues to test based on their high conservation among kinetoplastids and close location to the pocket. The two double mutants within the pocket, Y87A/F89A and K60A/K62A, have been previously tested for their influence on TbFPC4 binding [8]. Based on the unique horseshoe-shaped conformation of the pocket, we further separately mutated two aromatic residues, including the bulky residue (W71) gating the gap of the horseshoe, and Y64, the aromatic residue just outside of the pocket (Fig. 4A). Each of the two residues were mutated to alanine to check how the mutants affect the interaction of TbBILBO1-NTD with TbFPC4.

Interaction tests by ITC experiments revealed that the W71A mutant completely abolished the interaction of TbBILBO1-NTD with TbFPC4, whereas Y64A only reduced the binding affinity by approximately threefold (Fig. 4B). To further validate the dramatic effect of those mutants that abolish the interaction, we additionally carried out SEC/SLS using purified recombinant proteins. The results showed that, in contrast to the stably formed binary complex between wild-type TbBILBO1-NTD and TbFPC4-CTD, neither of the two TbBILBO1 mutants, W71A and Y87A/F89A, could form a stable complex with TbFPC4 (Fig. 4C), which further suggests that these three aromatic residues play pivotal roles in their interaction.

Furthermore, we checked the interaction of TbBILBO1 and TbFPC4 using EMSA on native gels. As expected, the purified TbBILBO1/TbFPC4 complex showed clearly a shifted band relative to the bands of the two individual proteins (Fig. 4D & E). Similarly, the mixture of TbBILBO1-NTD and TbFPC4-CTD in an equimolar ratio also showed a shifted band corresponding to the formed complex (Fig. 4F). In contrast, mixing TbFPC4-CTD with TbBILBO1-NTD mutants W71A or Y87A/F89A resulted in two separate bands of the two individual proteins (Fig. 4F).

Influence of the TbBILBO1 mutants on TbBILBO1-TbFPC4 interaction in U-2 OS cells – It was previously shown that the strong interaction between TbBILBO1 and TbFPC4 drives the colocalization

of the two proteins when ectopically expressed in mammalian U-2 OS cells [8]. We carried out a similar assay to further check the effect of the mutations on the interaction between TbBILBO1 and TbFPC4 under *in vivo* conditions. Similar to the previous report, expression of TbBILBO1 alone in U-2 OS cells generated filamentous polymers (Fig. 5A), while expression of GFP-tagged TbFPC4 showed that GFP-TbFPC4 decorated a filamentous network that matched the previously determined colocalization with microtubules (Fig. 5B). When co-expressed with TbBILBO1, TbFPC4 was only associated with the TbBILBO1 polymers (Fig. 5C). Co-expression of TbBILBO1-Y64A and TbFPC4 showed that TbFPC4 is associated with the polymers formed by TbBILBO1-Y64A, but some microtubule interaction was also observed (Fig. 5D, inset), suggesting a decrease in the binding affinity. When co-expressed with TbBILBO1-W71A, TbFPC4 was not associated with the TbBILBO1 polymers but gave a microtubule pattern, demonstrating a complete absence of binding to the TbBILBO1-W71A form (Fig. 5E), which was consistent with the *in vitro* data (Fig. 4).

Effect of the TbBILBO1 mutants in trypanosomes – We assessed the effect of expression of the Ty1-tagged Y64A and W71A TbBILBO1 mutants, as well as another mutant D65K/E66K/E67K (DEE/KKK), in procyclic (PCF) trypanosomes. Cell growth was assayed following induction of protein expression with 20 ng/ml tetracycline. The mutant DEE/KKK was chosen because these three residues are next to the peripheral aromatic residue Y64 but farther away from the hydrophobic pocket.

We used Western blotting to assess the expression level of Ty1-tagged TbBILBO1 (Fig. 6A). Overall, induction of the expression of wild-type and mutant Y64A of $_{Ty1}$TbBILBO1 had no or little impact on cell growth (Fig. 6B, C). However, induction of TbBILBO1-W71A was lethal after 24 hours (Fig. 6D), demonstrating that this residue is essential for TbBILBO1 function in the parasite. Interestingly, despite the close location of D65/E66/E67 to the hydrophobic pocket, there was no observable effect of TbBILBO1-DDD/KKK on cell growth at all (Fig. 6E). Consistently, this mutant did not affect the binding affinity between TbBILBO1 and TbFPC4 in our ITC experiments (data not shown).

DISCUSSION

BILBO1 was the first identified component of the FPC [5]. It consists of four structural domains, including the globular NTD that was shown to be essential for the function of TbBILBO1 in the cell [10]. We have previously reported the NMR structure of TbBILBO1-NTD, which demonstrated that it adopts a ubiquitin-like fold [10]. The 1.7-Å resolution crystal structure of TbBILBO1-NTD reported here shows an overall similar conformation to the NMR structure, with an extended β sheet formed by five strands and a long α helix packed diagonally across the surface of the β sheet (Fig. 1).

In the NMR structure, the C-terminal region of TbBILBO1-NTD, spanning residues 101-110, was poorly defined. Surprisingly, in the newly determined crystal structure this region was rigidly bound to the structural core of the protein. How can we explain the discrepancy observed between these two structures?

Firstly, the construct used for NMR studies encoded residues 1-110, which was designed based on the conservation analysis and *a priori* structural prediction. Sequence alignment of BILBO1 homologs from various kinetoplastid species showed a clear boundary at the residue P110 between the highly-conserved NTD and the following variable linker [10]. Furthermore, secondary structure prediction suggested that there were no helices or strands after residue 90. Therefore, we considered only residues 1-110 of TbBILBO1 in our original structural studies, which was much shorter than the protein we later used for crystallization reported in this study, i.e. residues 1-175.

During our NMR studies, we were able to determine the core part of the structure up to residue Q100. However, no long-range NOE constraints could be confidently assigned for residues 101-110. Notably, there are four proline residues (P104, P105, P109, & P110) in that region, which could not be assigned due to their lack of backbone amide hydrogens (Fig. 2A). These were probably the main reasons for the undefined C-terminal loop in the NMR structure.

Second, the crystal structure was solved based on the naturally degraded product of a longer construct of TbBILBO1 covering residues 1-175 [11]. The extra residues in the crystallization construct might have played a role in stabilizing the loop to make it bind more tightly to the structural core of TbBILBO1. Indeed, there were multiple hydrophobic interactions as well as water-mediated hydrogen bonds between residues 110-114 and the C-terminal part of the α-helix in the crystal structure of TbBILBO1-NTD (Fig. 2D). Notably, these interactions are mediated mostly by the backbone atoms of these residues, which explains why they are not conserved in the primary sequence, but could yet still participate in the binding of the loop to the structural core of TbBILBO1-NTD.

It was striking to observe such an enormously long rigid loop in the crystal structure, particularly for such a small compact domain. To the best of our knowledge, this is the first case in which a long rigid loop nearly fully encircles a cylinder-like structure of the same protein. Such an unusual structural arrangement seems to be an evolutionary selection based on our interaction analyses between TbBILBO1-NTD and TbFPC4-CTD. The stably bound loop is obviously critical for the formation of the unique horseshoe-like hydrophobic pocket (Fig. 3), which turned out to be the exact binding site of TbFPC4 (Fig. 4). We have tried to delete parts of the C-terminus of the loop in TbBILBO1-NTD, and found that although deletion of the loop to residue Q100 did not affect the conformation of the protein, it completely abolished the interaction between TbBILBO1 and TbFPC4 (data not shown). These results suggest that, while the loop itself does not affect the overall folding and structural stability of TbBILBO1-NTD, it is essential for the interaction between TbBILBO1 and TbFPC4.

Despite the poorly defined loop in our previously reported NMR structure, we had mapped a similar conserved patch on TbBILBO1-NTD [10]. Although we did not know the binding partner of TbBILBO1-NTD at that time, our mutagenesis studies demonstrated that such patch is critical for TbBILBO1 function *in vivo*. Compared to the NMR structure, the crystal structure reported here not only provides a higher resolution structure of TbBILBO1-NTD, but also sheds light on the horseshoe-like conformation of the hydrophobic pocket, which was shown to be relatively flat and irregular in the previous NMR structure.

Regarding the binding of TbFPC4 to TbBILBO1-NTD, our mutagenesis studies demonstrated that TbFPC4 contacts all aromatic residues in the center of the hydrophobic pocket (Fig. 4 & 5). *In vivo* examination of these mutants revealed that the three central aromatic residues play essential roles in controlling cell growth (Fig. 6), which is consistent with the previously reported results [10], and further confirms the importance of the conserved hydrophobic pocket.

In our mutagenesis studies we found that residue W71, which gates the entrance to the horseshoe-like pocket on TbBILBO1-NTD, is absolutely required for both TbFPC4 binding (Fig. 4B, Fig. 5E) and cell growth (Fig. 6D). However, mutating Y64, the residue outside the horseshoe-like pocket, only mildly reduced the binding of TbBILBO1-NTD to TbFPC4 (Fig. 4B; Fig. 5D). Further, the Y64A mutant did not cause dramatic effect in cell growth when overexpressed *in vivo* (Fig. 6C). We therefore believe that the corresponding binding region on TbFPC4, which is predicted to lack any secondary structures and thus adopt an extended conformation, likely binds horizontally across the β sheet to allow it pass through the gap of the horseshoe-like pocket (Fig. 7). Nevertheless, the binding site of TbFPC4 on TbBILBO1-NTD seems to be limited to the hydrophobic pocket and neighboring residues in the close proximity, as mutations of the three negatively charged residues, D65, E66, and E67, which are slightly farther away from the pocket, did not affect either the intermolecular interaction or cell growth (Fig. 6E).

The close correlation of the hydrophobic pocket in mediating the interaction of TbBILBO1 with TbFPC4 and its requirement for normal cell growth suggests that the main function of TbBILBO1-NTD might be to act in concert with TbFPC4 to regulate the structural and/or function of the FPC in trypanosomes. Previous studies have shown that an N-terminal region of TbFPC4 binds microtubules, which prompted the proposal that TbFPC4 might serve as a linker to connect the FPC to the microtubule quartet [8].

One important and interesting lesson we have learned from the crystal structure is that one should consider including extra residues beyond a conserved region assignment based on primary sequence alignment and structural prediction. As seen in the crystal structure of TbBILBO1-NTD, some non-conserved residues might still be involved in protein-protein interaction via their main chain atoms. Obviously, it is not always easy to predict how much extra sequence should be included when cloning

a subdomain of a protein that lacks three-dimensional structural information. An alternative strategy could be to first use limited proteolysis to identify the domain boundaries and then reclone the protease resistant compact region for subsequent structural studies. Such method has been used by many groups and was also very successful in our previous structural studies of exocyst components [21]. A modified protocol for that would be to carry out *in situ* proteolysis during crystallization to let the added trace amounts of proteases trim off the flexible regions potentially hindering crystallization [22].

In summary, we have reported a high resolution crystal structure of TbBILBO1-NTD (aa1-115), which reveals an unusual rigid loop critical for defining the horseshoe-like hydrophobic pocket responsible for the binding of its partner, TbFPC4. Notably, by now there have been no reports about a similar surface pocket on any other known ubiquitin-like fold containing proteins except for TbBILBO1-NTD. Future studies will try to map the exact region on TbFPC4 and find out the molecular detail of their interaction. Given the relatively low binding affinity between TbBILBO1 and TbFPC4, with a Kd of ~5 μM, it is probably worth checking whether a modified synthetic polypeptide with significantly higher affinity than the native sequence of TbFPC4 could be found, which could be potentially used as an inhibitor to disrupt the interaction between TbFPC4 and TbBILBO1 in the parasite. Alternatively, screening small molecular libraries may be carried out to identify candidate inhibitors that are tightly bound to the hydrophobic pocket to prevent the formation of the complex between TbBILBO1-NTD and TbFPC4.

REFERENCES

1. Engstler, M., et al., *Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei.* J Cell Sci, 2004. **117**(Pt 7): p. 1105-15.

2. Engstler, M., et al., *Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes.* Cell, 2007. **131**(3): p. 505-15.

3. Grunfelder, C.G., et al., *Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels.* Traffic, 2002. **3**(8): p. 547-59.

4. Gull, K., *The cytoskeleton of trypanosomatid parasites.* Annu Rev Microbiol, 1999. **53**: p. 629-55.

5. Bonhivers, M., et al., *Biogenesis of the trypanosome endo-exocytotic organelle is cytoskeleton mediated.* PLoS Biol, 2008. **6**(5): p. e105.

6. Vidilaseris, K., et al., *Assembly Mechanism of Trypanosoma brucei BILBO1, a Multidomain Cytoskeletal Protein.* J Biol Chem, 2014. **289**(34): p. 23870-81.

7. Vidilaseris, K., et al., *Assembly mechanism of Trypanosoma brucei BILBO1 at the flagellar pocket collar*. Commun Integr Biol, 2015. **8**(1): p. e992739.

8. Albisetti, A., et al., *Interaction between the flagellar pocket collar and the hook complex via a novel microtubule-binding protein in Trypanosoma brucei.* PLoS Pathog, 2017. **13**(11): p. e1006710.

9. Florimond, C., et al., *BILBO1 is a scaffold protein of the flagellar pocket collar in the pathogen Trypanosoma brucei.* PLoS Pathog, 2015. **11**(3): p. e1004654.

10. Vidilaseris, K., et al., *Structure of the TbBILBO1 protein N-terminal domain from Trypanosoma brucei reveals an essential requirement for a conserved surface patch.* J Biol Chem, 2014. **289**(6): p. 3724-35.

11. Vidilaseris, K. and G. Dong, *Expression, purification and preliminary crystallographic analysis of the N-terminal domain of Trypanosoma brucei BILBO1*. Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 5): p. 628-31.

12. Terwilliger, T.C., et al., *Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard.* Acta Crystallogr D Biol Crystallogr, 2009. **65**(Pt 6): p. 582-601.

13. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

14. Wirtz, E., et al., *A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei.* Mol Biochem Parasitol, 1999. **99**(1): p. 89-101.

15. Wirtz, E. and C. Clayton, *Inducible gene expression in trypanosomes mediated by a prokaryotic repressor*. Science, 1995. **268**(5214): p. 1179-83.

16. Schumann Burkard, G., P. Jutzi, and I. Roditi, *Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters.* Mol Biochem Parasitol, 2011. **175**(1): p. 91-4.

17. Heldin, C.H., et al., *A human osteosarcoma cell line secretes a growth factor structurally related to a homodimer of PDGF A-chains.* Nature, 1986. **319**(6053): p. 511-4.

18. Dacheux, D., et al., *A MAP6-related protein is present in protozoa and is involved in flagellum motility.* PLoS One, 2012. **7**(2): p. e31344.

19. Bastin, P., et al., *A novel epitope tag system to study protein targeting and organelle biogenesis in Trypanosoma brucei.* Mol Biochem Parasitol, 1996. **77**(2): p. 235-9.

20. Hannaert, V., et al., *Kinetic characterization, structure modelling studies and crystallization of Trypanosoma brucei enolase.* Eur J Biochem, 2003. **270**(15): p. 3205-13.

21. Lesigang, J. and G. Dong, *Analysis of Three-Dimensional Structures of Exocyst Components.* Methods Mol Biol, 2016. **1369**: p. 191-204.

22. Dong, A., et al., *In situ proteolysis for protein crystallization and structure determination.* Nat Methods, 2007. **4**(12): p. 1019-21.

23. Wallace, A.C., R.A. Laskowski, and J.M. Thornton, *LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.* Protein Eng, 1995. **8**(2): p. 127-34.

ACKNOWLEDGEMENT

* This work was supported by funding from the Max Perutz Labs and grant P24383-B21 from the Austrian Science Fund (FWF) to GD, the Bordeaux University and the Centre National de la Recherche Scientifique (CNRS) to DRR and MB, and the LabEx ParaFrap (ANR-11-LABX-0024) to DRR. YP has been supported during 2016-2019 by the “Integrative Structural Biology” PhD program (W-1258 Doktoratskollegs) funded by the FWF. KV was supported by an OeAD graduate scholarship during 2009-2012. NA was supported by a postdoctoral fellowship provided by the Austrian Agency for International Cooperation in Education & Research (OeAD-GmbH) during January – September 2019. We thank Frédéric Bringaud (Bordeaux University) for the anti-Enolase antibody and P. Bastin (Pasteur Institute) for the anti-Ty1 (BB2) antibody. We also greatly appreciate Dr. Brooke Morriswood for critical reading of the manuscript.

FIGURE LEGENDS

FIGURE 1. **Crystal structure of TbBILBO1-NTD.** *A,* Ribbon diagram of the TbBILBO1-NTD structure with two orthogonal views. The structure is color-ramped from blue at the N-terminus to red at the C-terminus. The five β strands (β1 - β5) and the single α helix (α1) are labeled. *B,* Stereo view of a part of the 2Fo-Fc map contoured at 2 σ. *C,* Secondary structure diagram of TbBILBO1-NTD, with the residue ranges for each structural elements indicated.

FIGURE 2. **A long rigid loop is present at the C-terminus of TbBILBO1-NTD.** *A,* Superimposed 10 energy-minimized NMR conformers of TbBILBO1-NTD. Residues in the dynamic C-terminal region (aa101-110) are listed, with the four proline residues marked by arrows. *B,* Ribbon diagram of the crystal structure of TbBILBO1-NTD. The last twenty residues (aa96-115) comprising the loop are shown as sticks in purple. *C,* Ribbon diagram in the same orientation as in *B* and colored by temperature factors (B-factors), with blue being the most rigid and yellow the most dynamic. The average B-factor for the structure is 23 $Å^2$. *D,* Details of interactions between the C-terminal tail and the rest of TbBILBO1-NTD. The plot was generated using DIMPLOT in the LigPlot plus suite [23]. Residues involved in hydrogen bond formation are shown as ball-and-stick, with oxygen, nitrogen and carbon atoms colored in red, blue, and gray, respectively. Water molecules mediating inter-molecular hydrogen bond formation are shown as cyan-colored spheres. Green dotted lines indicate hydrogen bonds. Non-bonded residues involved in hydrophobic interactions are shown as spoked arcs.

FIGURE 3. **TbBILBO1-NTD has a conserved horseshoe-like surface pocket.** *A,* Crystal structure of TbBILBO1-NTD with the C-terminal loop depicted as sticks and the rest of the structure as an electrostatic surface plot. Shown on the right are two orthogonal zoom-in views of the deeply buried side chain of residue H96 at the beginning of the C-terminal loop. *B,* Full electrostatic plot with the same orientation as in *A* to show the horseshoe-like hydrophobic pocket. *C,* Superposition of the crystal structure (pink) onto the previously reported NMR structure (gray, pdb code: 2MEK). The corresponding C-terminal loops in the two structures are colored in magenta and green, respectively. *D,* Cartoon view of the crystal structure around the hydrophobic pocket of TbBILBO1-NTD, with residues in and around the pocket shown in sticks. *E,* Electrostatic plot of the pocket with the same orientation as in *D*.

FIGURE 4. **The aromatic residues of the TbBILBO1-NTD pocket are required for TbFPC4 binding.** *A,* Cartoon view of the hydrophobic pocket of TbBILBO1-NTD with all mutated residues shown in sticks and labeled. *B,* ITC results of wild-type (WT) and mutants of TbBILBO1-NTD with TbFPC4-CTD. Mutants W71A and Y87A/F89A showed no interaction, whereas mutants Y64A and K60A/K62A showed slightly reduced binding affinity. *C,* SLS profiles of individual TbBILBO1-NTD and TbFPC4-CTD, and TbFPC4-CTD mixed with WT and mutants of TbBILBO1-NTD. *D,* Purified proteins of TbBILBO1-NTD, TbFPC4-CTD, and their complex from SEC on SDS-PAGE and visualized by Coomassie staining. *E,* The same samples as in *D* on a native gel. *F,* Comparison of WT and mutants of TbBILBO1-NTD alone or in complex with TbFPC4-CTD on a native gel. In contrast to the stable complex formed between the WT TbBILBO1-NTD and TbFPC4-CTD, mutants W71A and Y87A/F89A do not form complexes with TbFPC4-CTD.

FIGURE 5. **The aromatic residues of the TbBILBO1-NTD pocket are required for TbFPC4 recruitment *in vivo*.** *A,* Ectopic expression of TbBILBO1 alone in U-2 OS cells generated filamentous polymers. *B,* Ectopic expression of TbFPC4 alone showed its tight association with microtubules. *C,* Co-expressed TbBILBO1 and TbFPC4 proteins colocalized with each other. *D,* Co-expression of TbBILBO1-Y64A and TbFPC4 showed relatively robust overlap of TbFPC4 with the TbBILBO1-Y64A polymers, but some microtubule interaction is also observed (Inset). *E,* Co-expression of TbBILBO1-W71A with TbFPC4 showed no colocalization of the two proteins; TbFPC4 showed microtubule pattern.

FIGURE 6. **Residue W71 on TbBILBO1 is essential for its function in *T. brucei*.** *A,* WB for the PCF *T. brucei* cells, non-induced (NI) or induced up to 96 hours for the expression of Ty1-tagged TbBILBO1-WT, TbBILBO1-Y64A, TbBILBO1-W71A, or TbBILBO1-DEE/KKK. Anti-Enolase was used as a loading control. *B-E,* Growth curves for the PCF *T. brucei* cells from the same experiments as in *A*. The error bars represent the standard error from three independent experiments.

FIGURE 7. **A hypothetical model depicting the interaction between TbBILBO1-NTD and TbFPC4-CTD. Tb**BILBO1-NTD is shown in a surface plot with residues essential for TbFPC4 binding highlighted in red, and those partially affecting TbFPC4 binding colored in pink. TbFPC4 is drawn as a blue line, with regions involved in critical interactions marked as asterisks. The arrow points to the part in TbFPC4 that passes through the gap at one side of the horseshoe-like pocket to reach the peripheral binding site outside of the pocket.

Table 1. Data collection and refinement statistics

Data collection	
Wavelength (Å)	0.9792
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	29.69, 50.80, 37.22
α, *β*, *γ* (°)	90, 94.61, 90
Resolution (Å)	37.1-1.70 (1.78-1.70) *
No. observed reflections	98,166 (13,228)
No. unique reflections	12,398 (1,744)
R_{meas} (all I+ & I-)	0.124 (1.127)
R_{pim} (all I+ & I-)	0.044 (0.410)
$I / \sigma I$	12.7 (2.1)
Completeness (%)	99.7 (97.9)
Redundancy	7.9 (7.6)
Anomalous completeness (%)	99.0 (96.2)
Anomalous redundancy	4.0 (3.9)
Refinement	
Resolution (Å)	19.3-1.70
No. reflections	27,060
R_{work} / R_{free} (%)	16.7/19.8
No. atoms	
Protein	981
Water	135
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.17
Ramachandran plot	
Favored (%)	99.2
Allowed (%)	0.8

*Values in parentheses are for highest-resolution shell.

Fig. 1

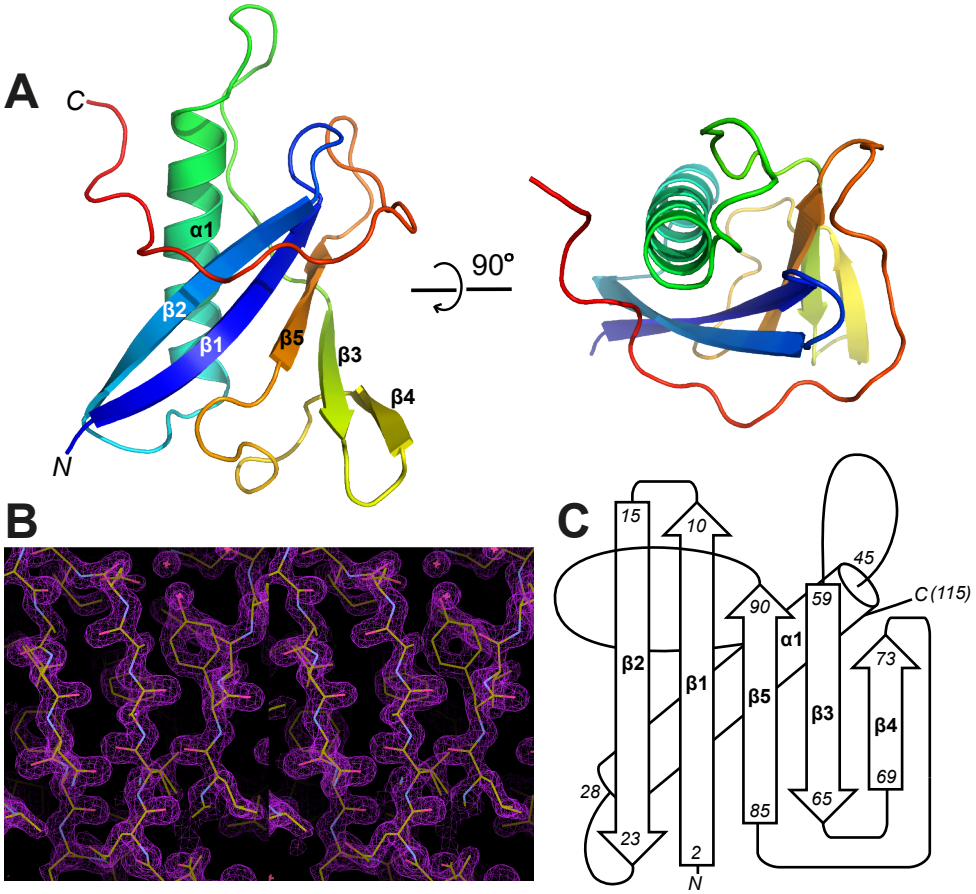

Fig. 2

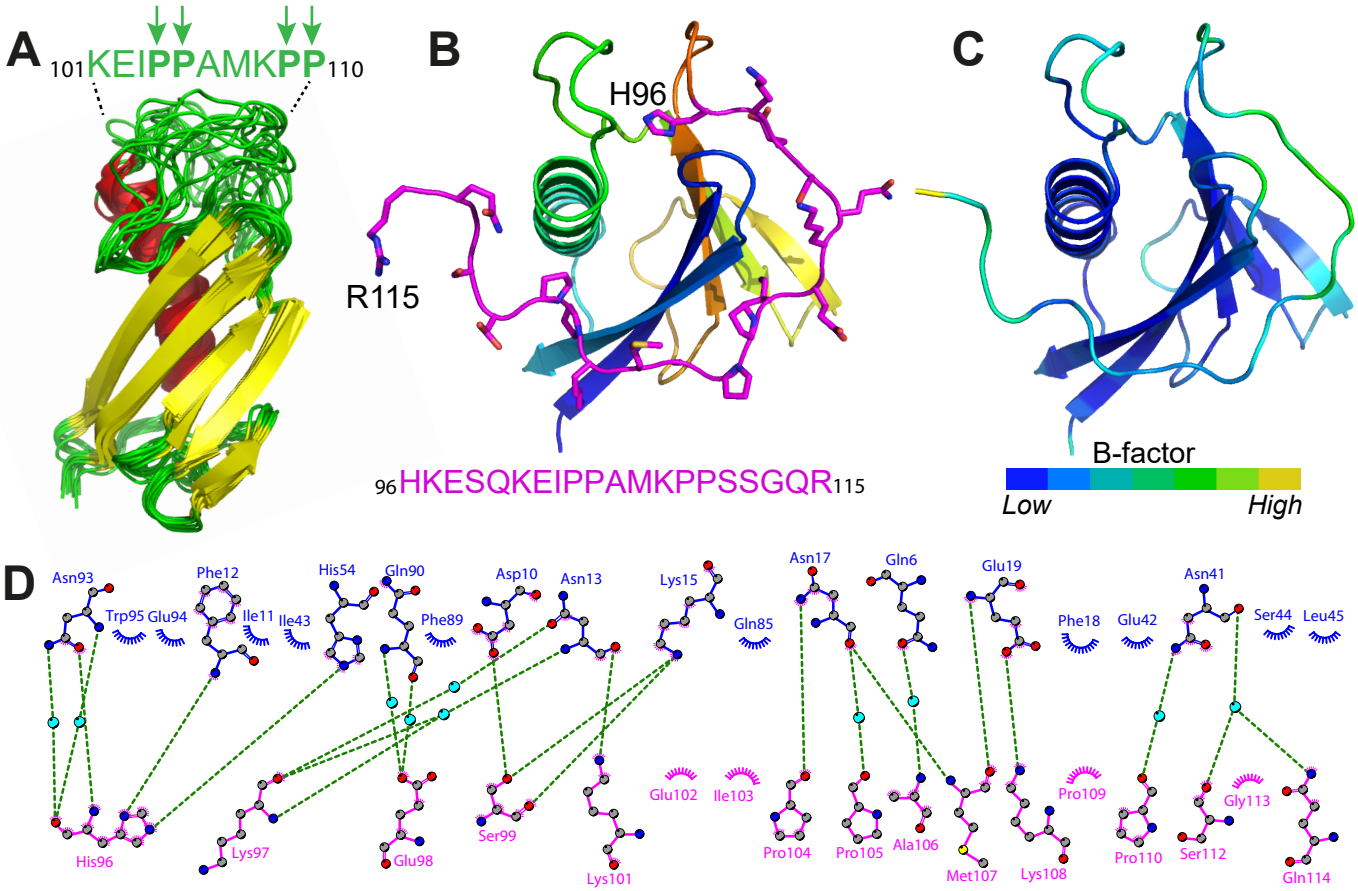

Fig. 3

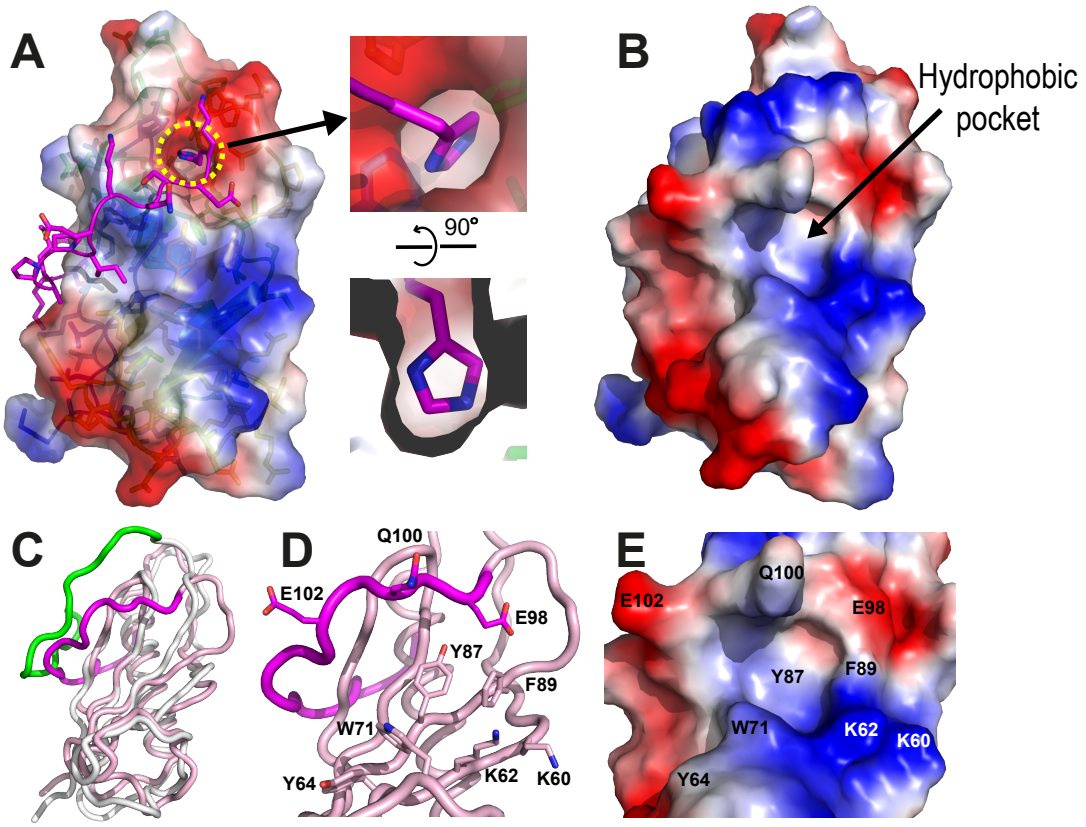

Fig. 4

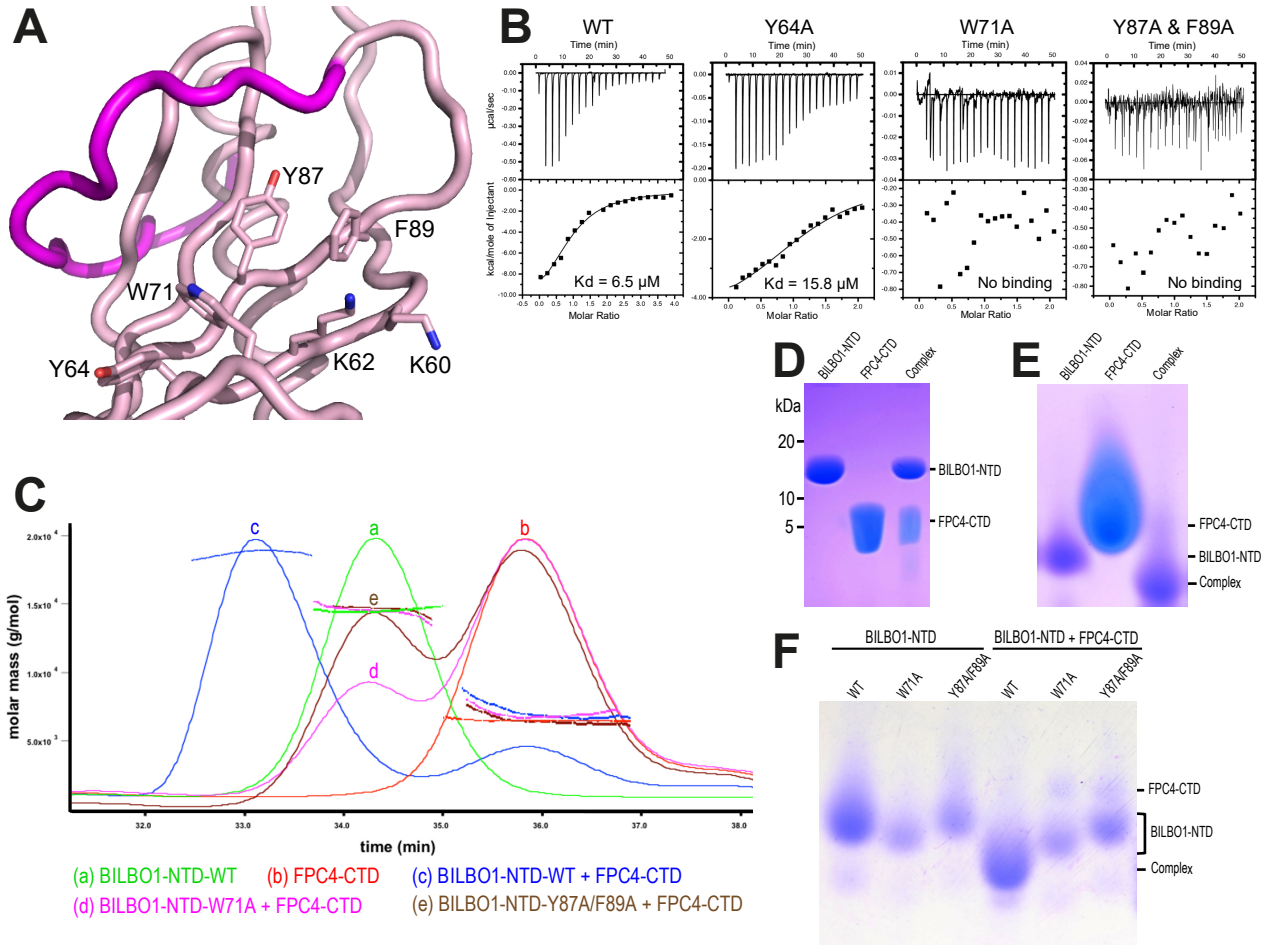

Fig. 5

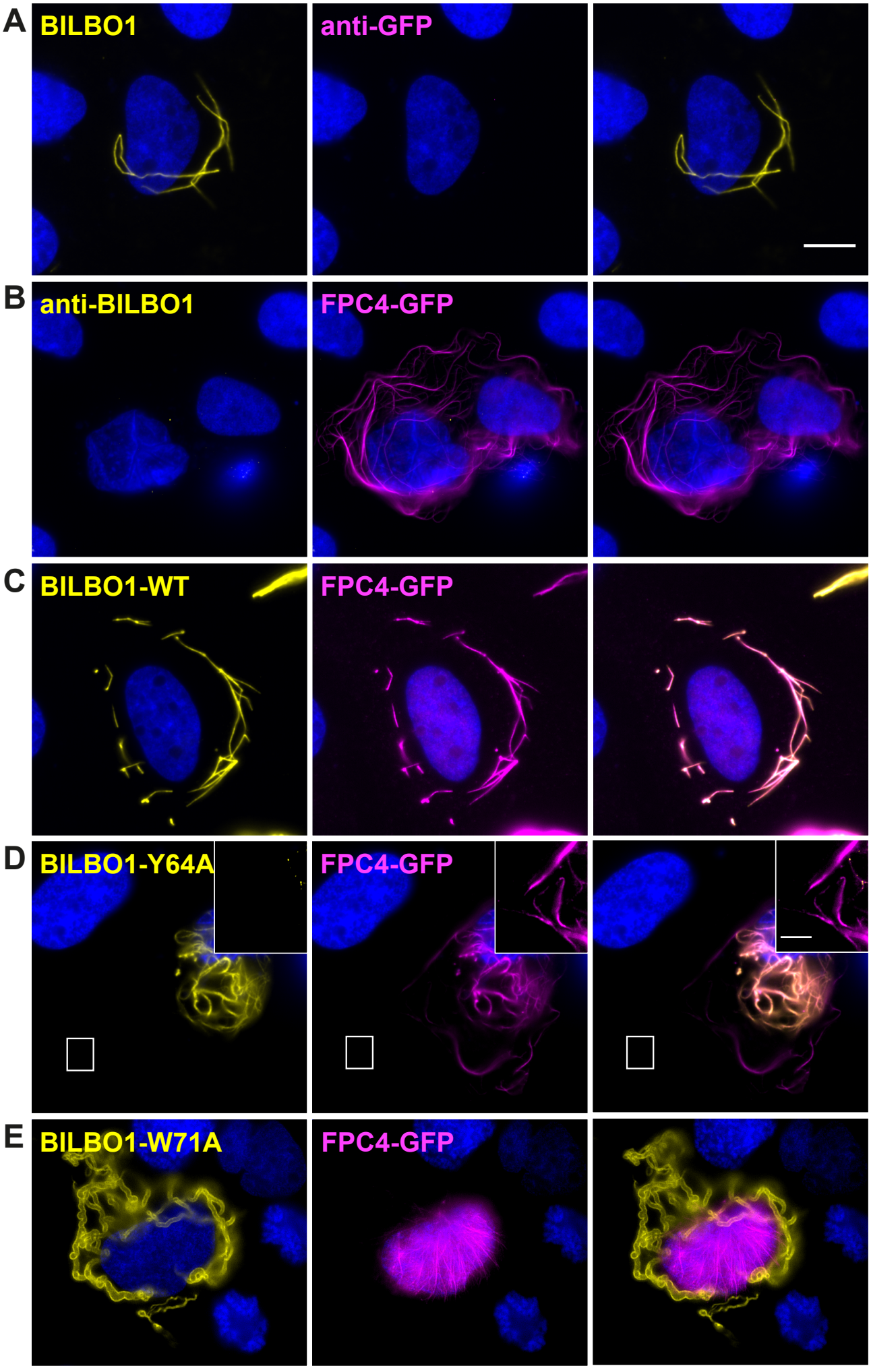

Fig. 6

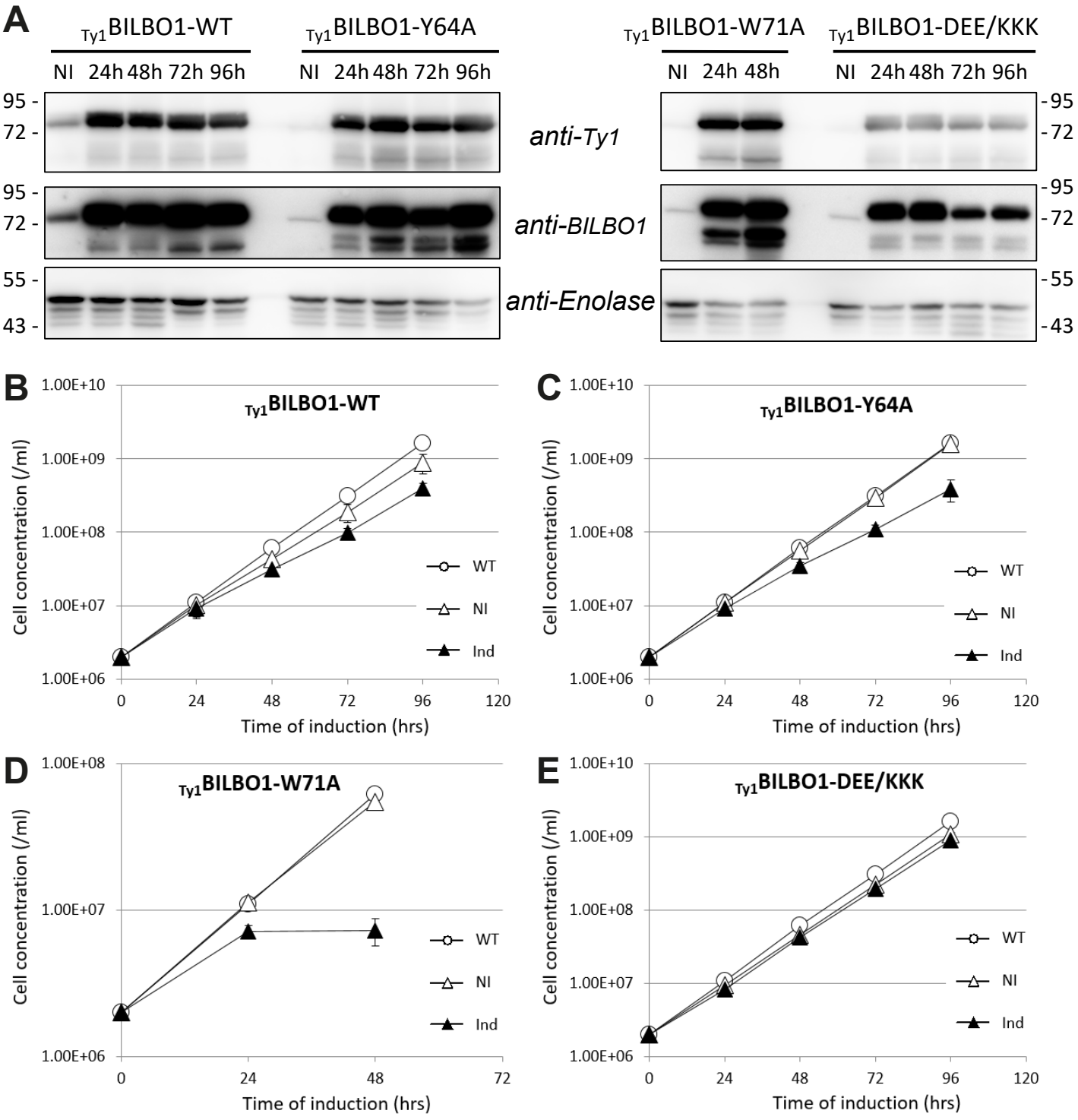

Fig. 7