universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Structure Determination of TbBL232 N-Terminal Domain by X-Ray Crystallography

verfasst von / submitted by

Rui Qiang Chen, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Dr. Gang Dong

Acknowledgements

First of all, I want to give my special thanks to Dr. Gang Dong for giving me the chance to work on this project and made it possible for me to have a pleasant time in his lab. I am grateful for the scientific guidance, continuous support, and advice he has given me during my work.

I want to thank Johannes Lesigang for teaching me everything in the lab, the inspiring conversations we had and the great time in the lab.

I want to express my gratitude to Emma Stepinac and Yulia Pivovarova for their scientific advice, the enjoyable time in- and outside the lab and for creating a wonderful place I really enjoyed working at.

Many thanks to Niran Aeksiri for giving me a helping hand on my experiments and the pleasant time we have spent in the lab.

I want to thank Paul Majneri and the whole Djinovic lab for the technical and scientific advice whenever I needed it.

I am grateful to Anna Nele Herdina for proof-reading my thesis and her helpful tips.

I am very grateful to my family for their constant support, motivation, and consideration throughout all these years.

At last, I want to thank my girlfriend Jasmin for encouraging me and giving me the strength whenever I am feeling down.

Table of content

Abstract

Trypanosoma brucei is a unicellular protozoan found in Sub-Saharan African countries. This single-celled organism leads an exclusively parasitic lifestyle and causes the lethal sleeping sickness in humans, threatening the lives of over 60 million people. *T. brucei's* cell architecture is spindle-shaped with a single flagellum attached to the posterior cell body. The flagellum originates from a specialized invagination of the plasma membrane described as the flagellar pocket (FP). Endo- and exocytosis are limited to the FP, which gives it an essential role for parasite-host interaction. The exit site of the flagellum harbours a "ring" structure called flagellar pocket collar (FPC). TbBILBO1 is the predominant component found in the FPC. It is a multidomain protein composed of an N-terminal domain (NTD), two EF-hand motifs, a large coiled-coil (CC) domain and a C-terminal leucine zipper. Flagellum pocket collar 4 (FPC4) has been identified as a binding partner, which binds to TbBILBO1 NTD via its C-terminal BILBO1 binding domain (B1BD). However, FPC4 B1BD is also targeted by other proteins such as TbBL232, the NTDs of which share a 35 % sequence homology with the TbBILBO1 NTD. The structure and the binding mechanism of TbBL232 NTD to FPC4 are poorly understood. The aim of this work was to understand how TbBL232 NTD binds to FPC4 by employing x-ray crystallography.

For that purpose, both interaction partners were co-expressed in *E. coli* (DE3) and purified using different types of High-Performance Liquid Chromatography (HPLC). Crystallization of TbBL232 NTD was achieved by in-situ proteolysis using trypsin. Protein crystal diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) and the protein model was built in COOT. The electron density map revealed a truncated TbBL232 NTD missing about 28 residues at the C-terminus as well as the FPC4 ligand. Considering the structural homology of TbBL232 NTD and TbBILBO1 NTD, the C-terminus (aa 91-119) of TbBL232 is also believed to form a long dynamic loop which might constitute a section of the hydrophobic binding site. To gain a better understanding of the importance of this loop region, short missing loop NTD constructs of TbBILBO1 and its homologs TbBL271 and TbBL232 were generated for ligand binding tests via isothermal titration calorimetry (ITC). The truncated proteins lost their capability to bind FPC4, indicating a crucial role for this previously reported dynamic region.

Zusammenfassung

Trypanosoma brucei ist ein einzelliges Protozoon, das in afrikanischen Staaten südlich der Sahara vorkommt. Dieser einzellige Organismus führt eine ausschließlich parasitäre Lebensweise und ist für die letale Schlafkrankheit bei Menschen verantwortlich, die das Leben von über 60 Millionen Menschen bedroht. *T. brucei* hat einen spindelförmigen Zellaufbau mit einem einzelnen Flagellum am hinteren Zellkörper. Die Geißel entspringt aus einer speziellen Einstülpung an der Plasmamembran die als Flagellar Pocket (FP, dt. Flagellumtasche) bezeichnet wird. Die Endo- und Exozytose finden ausschließlich im FP statt, weshalb diese Struktur eine essenzielle Rolle in der Parasit-Wirt Interaktion einnimmt. An der Ursprungsstelle des Flagellums befindet sich eine ringartige Struktur, genannt Flagellum Pocket Collar (FPC, dt. Flagellumtaschenkragen). TbBILBO1 ist die hauptsächliche Komponente, die im FPC bestimmt wurde. Es ist ein Multidomänenprotein bestehend aus einer N-terminalen Domäne (NTD), 2 EF-Hand Aminosäurenmotive, eine Coiled-Coil Domäne und ein Leucin Zipper am C-Terminus. Flagellum Pocket Collar 4 (FPC4) wurde als Bindungspartner identifiziert. Dieser bindet die NTD von TbBILBO1 mithilfe seiner C-terminale BILBO1 Bindungsdomäne (B1BD). Jedoch wird die B1BD ebenfalls von anderen Proteinen gebunden, wie bei TbBL232, dessen NTD eine 35-prozentige Sequenzhomolgie zu TbBILBO1 NTD aufweist. Die Struktur und der Bindungsmechanismus zwischen TbBL232 und FPC4 sind zum Großteil unbekannt. Ziel dieser Studie ist, zu verstehen wie die NTD von TbBL232 FPC4 bindet mittels Röntgenkristallographie.

Dafür wurden beide Interaktionspartner in *E. coli* (DE3) co-exprimiert und mithilfe von verschiedenen Hochleistungsflüssigkeitschromatographiemethoden aufgereinigt. Die Kristallisation von TbBL232 NTD und FPC4 wurde durch in-situ Proteolyse mit Trypsin realisiert. Die Proteinkristalldaten wurden in der European Synchrotron Radiation Facility (ESRF) erfasst und das Protein-Modell wurde in COOT erstellt. Die Elektronendichtekarte zeigte sowohl eine am C-Terminus um 28 Aminosäuren verkürzte Variante des TbBL232 NTD als auch das Fehlen von FPC4. Unter Berücksichtigung der strukturellen Homologie der N-terminalen Domänen von TbBL232 und TbBILBO1 wird angenommen, dass die C-terminalen Aminosäuren 91-119 von TbBL232 ebenfalls eine lange Schleifenstruktur ausbilden, die einen Teilbereich der hydrophoben Bindungstasche bildet. Um die Signifikanz dieser Region zu verstehen, wurden kurze Konstrukte von TbBILBO1 und seinen

homologen Proteinen TbBL271 und TbBL232, die keine Schleifenregionen aufweisen, für isotherme Titrationskalorimetrie-Versuche hergestellt. Die verkürzten Proteine verloren vollständig die Fähigkeit FPC4 zu binden, was auf eine essenzielle Rolle dieser Struktur hindeutet.

Introduction

Trypanosoma brucei taxonomy

Trypanosoma brucei is a unicellular protozoan. It leads an exclusively parasitic lifestyle causing devastating human and animal diseases.

Trypanosoma brucei is divided into three subspecies (Fig.1). *T. brucei brucei* infects livestock and game animals and causes the nagana disease restraining agricultural development and therefore promoting starvation and poverty in sub-Saharan Africa (Matthews, 2005). *T. brucei gambiense* is found in western and central Africa, where it causes the 95-97 % of cases of sleeping sickness. *T. brucei rhodesiense* infect animals as well as in humans in eastern and southern Africa causing 3-5 % of all cases. (Franco *et al.*, 2014; Kennedy, 2019).

Like other species in the order *Kinetoplastida*, *T. brucei* is characterized by a unique mitochondrial genome described as kinetoplast. The kinetoplast is comprised of few maxicircles and thousands of interconnected minicircles, which form a structure like a chainmail armour (Simpson, Stevens and Lukeš, 2006).

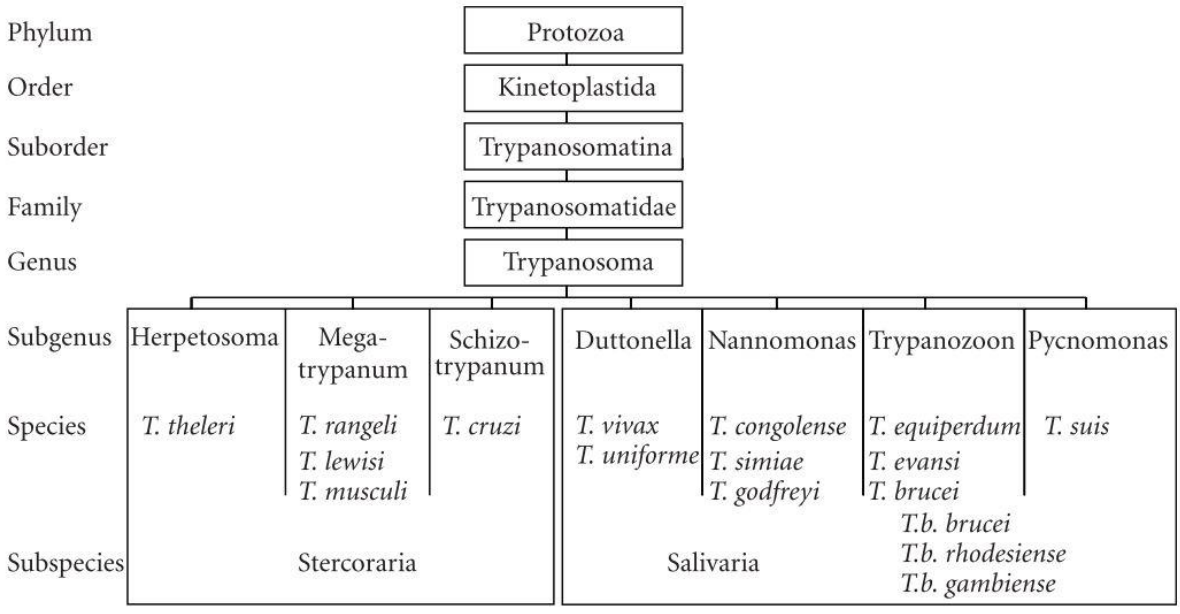


Figure 1 Taxonomy of Trypanosomes. Figure is modified from (Baral, 2010).

Lifecycle

Trypanosoma brucei undergoes a complex lifecycle in its insect and mammalian hosts. The Tsetse fly, as its insect host, takes up *T. brucei* during its blood meal (through a bite of the fly´s proboscis). *T. brucei* multiplies and evolves in the fly's midgut. The procyclic trypomastigotes will then migrate to the salivary gland, where they further proliferate into the metacyclic state and eventually get injected into the mammalian host with the next blood meal of the fly. Inside the vertebrate host, *T. brucei* transforms into the bloodstream trypomastigote form and multiplies in various body fluids, where it remains. The parasite in the bloodstream again will be taken up by the Tsetse fly during its blood meal, completing the cycle (Matthews, 2005; Baral, 2010) (Fig. 2).

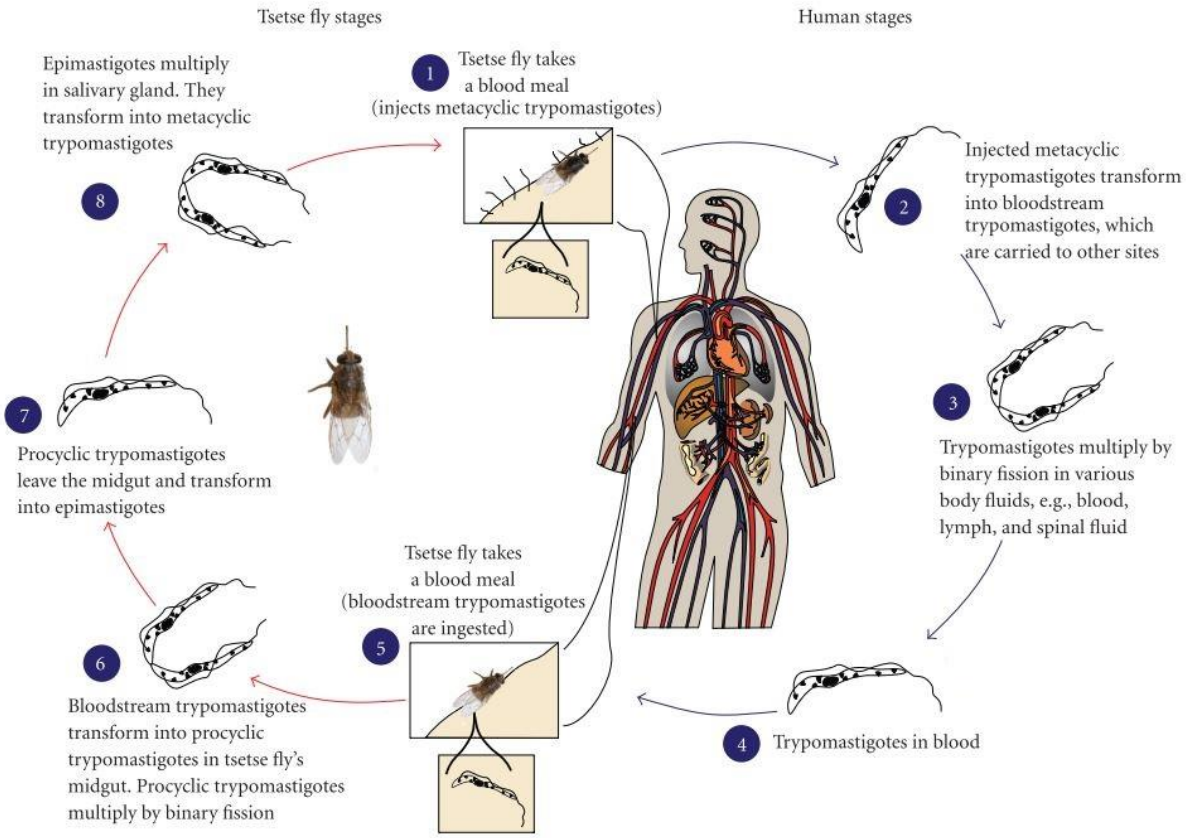


Figure 2 T. brucei life cycle is divided in two stages. In the Tsetse fly stage, upon ingestion, the bloodstream trypomastigotes undergo a transformation into the procyclic trypanomastigote form in the midgut. In this state the protozoan multiply by binary fission until leaving the midgut, where they evolve into epimastigotes. Upon reaching the salivary gland the parasite transforms into metacyclic trypanomastigotes before it is injected into the human host via the Tsetse flies bloodmeal initiating human stage of its development. In the new host the trypanosomes will adopt the bloodstream trypanomastigote form again. Figure is modified from (Baral, 2010).

Human African Trypanosomiasis (sleeping sickness)

Human African Trypanosomiasis (HAT) is a devastating disease endemic in 36 sub-Saharan countries, which poses a threat to over 60 million people. There are two distinct forms of HAT caused by two subspecies of *T. brucei*. The most common form is the chronic form, which is prevalent in west and central Africa caused by *T. brucei gambiense.* The acute form caused by *T. brucei rhodesiense* is believed to be a zoonotic disease that occasionally affects humans. Unlike *T. brucei gambiense* which primarily infects humans*,* animals serve as the main reservoir for *T. brucei rhodesiense*. Patients of both HAT types show similar symptoms, but with often different incubation period and severity.(Kaufer *et al.*, 2017)

In the early or first stage of HAT, the parasite resides in the lymphatic system and the bloodstream. This is accompanied by unspecific symptoms like fever, headache, dermatologic disorders (e.g., pruritus), lymphadenopathies, asthenia, anemia, cardiac disorders, musculoskeletal problems. As the disease progresses *T. brucei* crosses the blood-brain barrier and infects the central nervous system, which initiates the late stage of the disease. If patients are left untreated neuropsychiatric problems emerge, followed by organ failure and eventual death.The acute form of HAT manifests with poor demarcation between the stages and results in death within half a year. In comparison, the chronic form of HAT is more a progressive disease that leads to coma after a few years (Malvy and Chappuis, 2011; Franco *et al.*, 2014).

T. brucei morphology and Flagellar pocket

T. brucei's cell architecture is described as elongated spindle-shaped with a single flagellum attached to its cell body. The parasite's organelles including the nucleus, Golgi apparatus, kinetoplast, flagellar pocket, and the flagellum, are located between the posterior end and the center of the trypanosome. The mitochondrion is an elongated structure spanning from the posterior to the anterior (Matthews, 2005; Kaufer *et al.*, 2017) (Fig. 3).

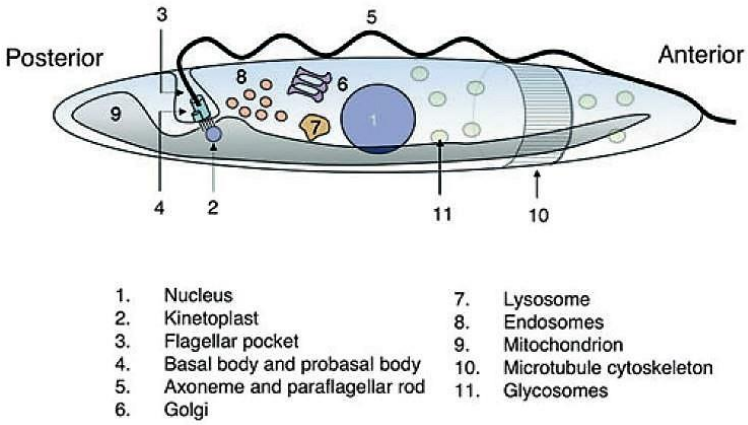


Figure 3 Simplified depiction of trypanosome cell structure. The FPC and the organelles are located at the posterior end of the parasite. Figure is modified from (Matthews, 2005).

The source of *T. brucei's* mobility is the flagellum, which is laterally attached throughout the entire cell body via specialized transmembrane junctions known as the flagellum attachment zones (FAZ). The flagellum is made up of microtubules arranged in a 9+2 axoneme structure, where nine double microtubules are positioned around a pair of singlet microtubules (Fig. 4). The flagellum originates from a specialized invagination of the plasma membrane described as the flagellar pocket (FP), located at the posterior of the cell. The parasite's secretion and endocytosis are limited to this site(Field and Carrington, 2009; Langousis and Hill, 2014). The lumen is filled with a carbohydrate-rich matrix, where important receptors are located. Mammalian host´s immunoglobulins that bind to the surface of the parasite are directed into the FP for endocytosis and degradation giving it an essential role for parasite-host interactions (Perdomo, Bonhivers and Robinson, 2016).

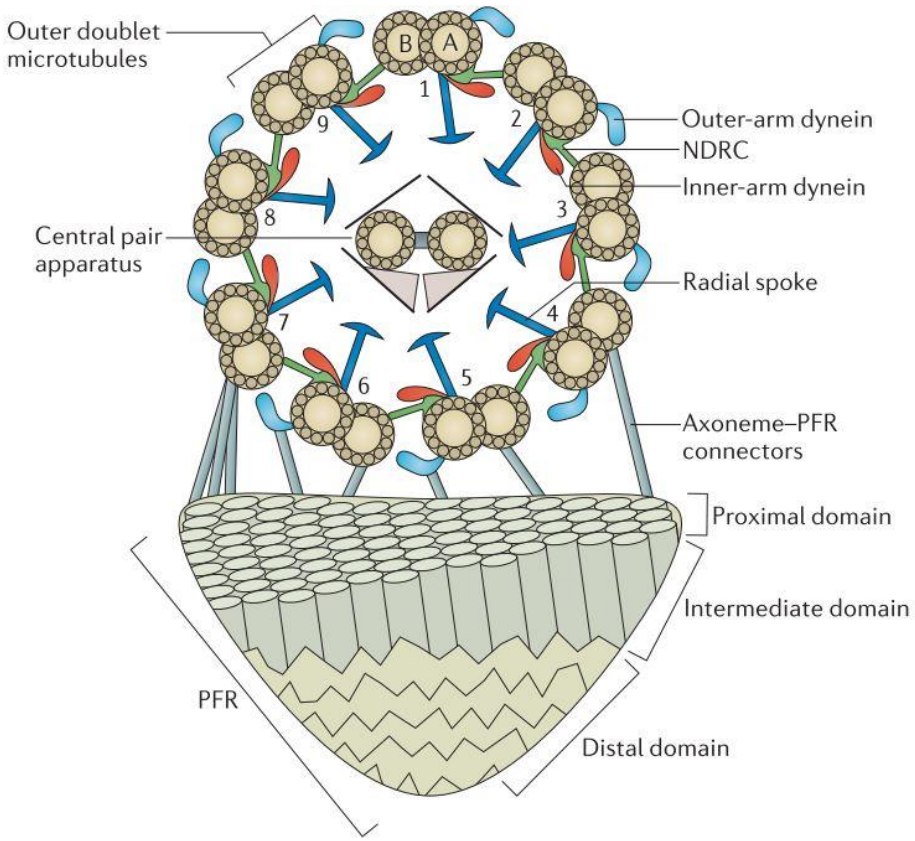


Figure 4 Cross section of the 9+2 axoneme structure. A Pair of singlet microtubules are surrounded by 9 outer microtubule pairs, connected by radial spokes. Figure is modified from (Langousis and Hill, 2014).

T. brucei flagellar pocket collar and its cytoskeleton protein TbBILBO1

The Flagellar pocket collar (FPC) forms a horseshoe shaped structure around the exit site of the flagellum, connecting the pellicular and FP membranes, thus it forms a boundary between the membranes and stabilizes the FP. Electron tomography studies of the FP indicate that there is a gap for a specialized microtubule quartet in the FPC, which spans from alongside the basal body to the cells anterior (Vidilaseris *et al.*, 2015; Perdomo, Bonhivers and Robinson, 2016).

The initial component found in the FPC was a protein termed TbBILBO1. This cytoskeletal component is a multidomain protein composed of an N-terminal domain (NTD), which adopts a ubiquitin-like fold, two EF-hand motifs, a large coiled-coil (CC) domain, and a putative leucine zipper at its C-terminus (Perdomo, Bonhivers and Robinson, 2016) (Fig.5).

TbBILBO1 forms antiparallel homodimers via the CC domain. The C-terminal leucine zipper facilitate FPC targeting and filament formation by interaction between adjacent leucine zippers of neighbouring TbBILBO1 dimers. The EF-hand motifs might adopt a role in the regulation of the FPC assembly or function upon calcium binding which

triggers conformational changes and enables interaction with binding partners (Vidilaseris *et al.*, 2015; Albisetti *et al.*, 2017). The function of the NTD of TbBILBO1 is poorly understood. Deletion or mutation of conserved residues of this domain affect cell viability, indicating a crucial role in the interaction with TbBILBO1 binding partners (Albisetti *et al.*, 2017). TbBILBO1 polymers might form a molecular framework for other FPC associated proteins to interact with (Albisetti *et al.*, 2017; Majneri, 2018).

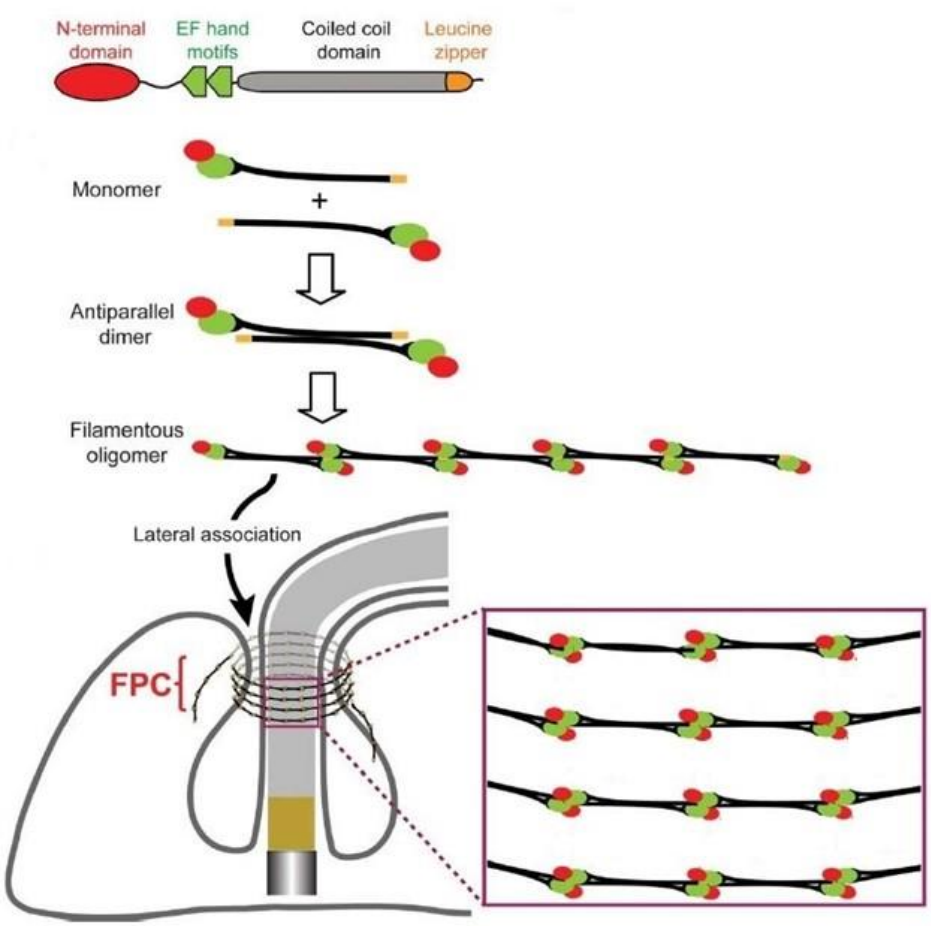


Figure 5 BILBO1 the main component of the FPC, is a multidomain protein consisting of a N-terminal domain, two EF hand motifs, a coiled-coil domain and a leucin zipper at the C-terminus. It forms antiparallel dimers via its coiled-coil domain, which then oligomerizes with further BILBO1 dimers through interaction of the C-terminal leucine zippers. The FPC is formed by long BILBO1 filaments wrapped around the flagellum pocket neck. Figure is modified from (Vidilaseris *et al.*, 2015).

FPC4 and TbBILBO1

Yeast-two-hybrid (Y2H) experiments revealed putative TbBILBO1 binding partners. One protein that received special attention is a 444 amino acid protein named Flagellar pocket collar protein 4 (FPC4). FPC4 binds the NTD of TbBILBO1 via its BILBO1 binding domain (B1BD) located at the C-terminus (Albisetti *et al.*, 2017). The FPC4 N-terminus on the other hand, interacts with microtubules, whereas the binding mechanism remains unknown (Vidilaseris *et al.*, 2020). In vivo localization reveals that FPC4 is co-localized together with TbBILBO1 as well as a Hook Complex (HC) protein named Membrane Occupation and Recognition Nexus protein 1 (MORN1). The HC is one of the two domains of the cytoskeleton-associated structure called bilobe, which overlaps the FPC. This positioning of FPC4 indicates that it is in contact with MORN1 and TbBILBO1 simultaneously, acting as a link between the FPC and the HC (Albisetti *et al.*, 2017). Recently, unpublished novel proteins which interact with FPC4 have been identified by our collaborators at the University of Bordeaux in France. The NTD of the binding partners share a 30-40 % sequence homology with the NTD of TbBILBO1, thus they were termed BILBO-like (BL) proteins.

One of the homologs is TbBL232 which is localized at the HC. This 232 amino acid (aa) protein consists of a predicted NTD domain similar to TbBILBO1 and a coiled-coil domain. The binding mechanism of BL232 to its ligand FPC4 as well as its role in the FP are poorly understood.

The aim of this work was to understand how BL232 binds to FPC4 using x-ray crystallography, and to determine the crucial amino acids involved in this protein-protein interaction.

Material and Methods

List of Buffers, media, and solutions

LB-medium(autoclaved)

25 g of Luria-Bertani broth in 1l dH_2O

Selective Agarose plates(autoclaved)

15 % (w/v) Agarose in dH_2O

30 µg/ml Kanamycin and/or

50 µg/ml Ampicillin

10x SDS-PAGE Running buffer

1.92 M Glycine

0.25 M Tris-HCl (pH=8)

1 % (w/v) SDS

dH_2O

6x SDS-PAGE loading dye

0.35 M 4x Tris-HCl SDS (pH=6.8)

30 % (w/v) Glycerol

10 % (w/v) SDS

0.6 M DTT

0.01 % (w/v) Bromphenol blue

dH_2O

SDS-PAGE destaining solution

5 % (v/v) Ethanol

5 % (v/v) Acetic acid

dH_2O

50x TAE Buffer

2 M Tris-HCl (pH=8.0)

5.7 % (v/v) Acetic acid

50 mM EDTA

Histrap running buffer A

20 mM Tris-HCl (pH=8)

100 mM NaCl

20 mM Imidazole (pH=8)

10 mM β-Mercaptoethanol

Histrap elution buffer B

20 mM Tris-HCl (pH=8)

100 mM NaCl

1 M Imidazole (pH=8)

10 mM β-Mercaptoethanol

dH_2O

Dialysis buffer

20 mM Tris-HCl (pH=8)

100 mM NaCl

10 mM β-Mercaptoethanol

Size-exclusion Chromatography buffer (Gelfiltration buffer)

10 mM Tris-HCl (pH=8)

100 mM NaCl

1 mM DTT

dH_2O

Coomassie brilliant blue staining solution

0.025 % (w/v) Coomassie brilliant blue R250

40 % (v/v) Ethanol

10 % (v/v) Acetic acid

dH_2O

6x DNA loading dye

30 % (v/v) Glycerol

0.25 % (w/v) Bromphenol blue

dH_2O

ITC buffer

10 mM Tris-HCl (pH=8)

100 mM NaCl

dH_2O

Anion Exchange Chromatography Buffer A (AEX buffer A)

20 mM Tris-HCl (pH=8)

50 mM NaCl

1 mM DTT

dH_2O

Anion Exchange Buffer Chromatography B (AEX buffer B)

20 mM Tris-HCl (pH=8)

1 M NaCl

1 mM DTT

dH_2O

Plasmid preparation

Polymerase chain reaction

The Q5® High-Fidelity PCR Kit (New England Biolabs® Inc., USA) was used to amplify DNA fragments. A PCR master mix, including 5x Q5 reaction buffer, 5x Q5 high GC enhancer, deoxyribonucleotide triphosphate (dNTP), template DNA and nuclease-free water, was prepared. The master mix was then split, and the forward and reverse primer were added to each reaction tube. Q5 polymerase was added just before the PCR run. The composition of the PCR reaction and the running protocol are shown in Tables 1 and 2, respectively.

Table 1 Q5 PCR

Q5 PCR reaction composition	
5x Q5 reaction buffer	10 µl
5x Q5 high GC enhancer	10 µl
dNTP (10 mM)	1 µl
Forward primer (20 mM)	1 µl
Reverse primer (20 mM)	1 µl
Template DNA	1 µl
Q5 Polymerase	0.5 µl
ddH2O	25.5 µl
Total reaction volume	50 µl

Table 2 Q5 PCR running protocol.

PCR running protocol		
Initial Denaturation 98°C	1 min	
Denaturation 98°C	15 sec	40 cycles
Annealing 55°C	30 sec	
Elongation 72°C	15 sec	
Final extension 72°C	2 min	

Agarose gel preparation

Following the PCR, amplified DNA was checked and purified on 1.25 % Agarose gels. 1 g of agarose was melted in 80ml of 1xTAE buffer in the microwave. 8 µl of SYBR™

safe stain were added after cooling down the agarose gel solution. Then the warm liquid gel was poured into a gel tray and left covered for 30 min to allow solidification.

50 µl of PCR product were mixed with 10 µl of 6x DNA loading dye and loaded into the wells of the agarose gel. The electrophoresis was run at 120 V for 30 min.

Gel extraction and Insert preparation

Gel blocks containing the desired DNA were cut out from the Agarose gel. Gel extractions of the amplicons were performed using the Wizard© SV Gel and PCR Clean-Up System (Promega).

42 µl of purified PCR products were double digested using 1.5 µl of NdeI and BamHI restriction enzymes from NEB© while adding of 5 µl 10x restriction buffer for 30 min.

Purification of generated inserts was done using the Wizard© SV Gel and PCR Clean-Up System (Promega, USA).

Ligation

The previously generated inserts of TbBL232 were ligated into a pET15b vector backbone including a small ubiquitin like modifier (SUMO) protein tag. The SUMOpET15b vector provides an ampicillin resistance and a 6x histidine tag for histidine affinity chromatography. The SUMO protein tag enhances the expression and solubility of the fused protein and supplies an efficient SUMO protease cleavage site (Malakhov *et al.*, 2004). The FPC4-9/1 inserts were ligated into a pET28a, which in contrast provides a kanamycin resistance cassette. Therefore, only *E. coli* BL21 cells expressing both target proteins were able to grow in the selective medium including both antibiotics.

Isothermal Titration Calorimetry (ITC) samples were ligated into a custom MalpET expression vector (kindly provided by Johannes Lesigang, Dong Lab, Max Perutz Laboratories, Austria), which contains a N-terminal maltose binding protein (MBP) tag, a kanamycin resistance cassette, a 10x histidine tag and a Tobacco etch virus (TEV) protease cleaving site.

The ligation was performed using the NEB T4 DNA ligase. Before the reaction, the vector and the double digested insert were combined at a ratio of 1:5, respectively.

After the plasmid components were mixed to a volume of 4.2 µl, 0.5 µl of 10x NEB ligase buffer was added. Finally, 0.3 µl of NEB T4 DNA ligase was subjoined and the sample was either incubated for 1 hour at RT or at 16°C overnight.

Competent cells

10ml of *E. coli* BL21 (DE3) or DH5α were incubated at 220 rpm at 37°C overnight. 2 ml of the overnight culture were transferred to 200 ml LB-medium and incubated at 37°C under constant shaking until OD_{600}=0.4-0.6 was achieved. Then bacterial pellets were generated by centrifuging at 2,700 rpm for 15 min. After that, the pellet was resuspended in 70 ml of 0.1 M $CaCl_2$ and pelletized again at 2,500 rpm for 10 min. Following the centrifugation, the pellet was dissolved in 0.1 M $CaCl_2$ again and kept on ice for 30 min. The bacteria were pelletized again at 2,500 rpm for 5 min and the resulting pellet was resuspended in 15 ml 0.1M $CaCl_2$ with 15 % (v/v) glycerol for cryoprotection.

Transformation

For transformation, 1 µl of plasmids were added to 50 µl of competent E. coli cells. E. coli DH5α were used for plasmid generation. After addition of the DNA, the bacteria were left on ice for 20 min. A heat shock at 42°C for 45 sec follows and 200 µl of LB-medium were added. The cells were then incubated at 37°C for approx. 60 min. After incubation, the transformed bacteria were plated onto selective agar plates and/or injected into liquid LB-media.

DNA Miniprep

Plasmids were multiplied by transferring a single colony to 5 ml LB-medium followed by incubation at 37°C overnight. Extraction of the DNA was done using the Wizard® Plus SV Minipreps DNA Purification System (Promega, USA). The overnight culture was pelletized by centrifugation at 16,000 rpm for 10 min. The bacterial pellet was dissolved in 250 µl of Cell Resuspension solution. After the addition of 250 µl Cell Lysis solution, the sample was mixed by inversion. Then 10 µl Alkaline Phosphatase solution was added and mixed. 5 min of incubation followed before 350 µl of neutralization solution were subjoined. The sample was centrifuged at 16,000 rpm on the table centrifuge for 10 min to obtain the clear cell lysate.

The previously generated plasmid DNA was bound by inserting the lysate into the spin column, which is sitting on the collection tube, and centrifuging it at 16,000 rpm. The flowthrough was discarded and 750 µl of wash solution was added to the spin column, followed by centrifugation for 1 min at 16,000 rpm. This washing step was repeated with 250 µl wash solution, before the removal of all washing agents by centrifuging at maximum speed for 2 min. The spin column was inserted into a sterile 1.5 ml Eppendorf tube and the DNA was first eluted with 30 µl of nuclease-free water. After the first elution, the column was centrifuged again at 16.000 rpm under addition of 25 µl nuclease-free water.

The plasmid DNA were sent to Microsynth© for sequencing to verify a proper ligation of the insert into the vector.

List of Constructs

The constructs and their amino acid sequence are listed below.

Green – Construct; Blue – FPC4; Red – SUMO-tag; Yellow – MBP-tag

HHHHHH – His-tag

GG^{v} – SENP2 cutting site

ENLYFQ^{v}G – TEV cutting site

MBP-tag

Theoretical molecular weight (MW) / Isoelectric point: (pI) 44.49 kDA / 5.71

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDG
PDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALS
LIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKY
ENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTING
PWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL
TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY
AVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNLGHHHHHHHHHHENLY
FQ^{v}GH

SUMO-tag

Theoretical molecular weight (MW) / Isoelectric point: (pI) 13.48 kDA / 5.89

MGSSHHHHHHSSGLVPRGSHNSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIF
FKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGISIQADQTPEDLDMEDNDIIEAHRE
QIGG^{v}H

Crystallization constructs

TbBL232-1/3 (aa 1-119)

Theoretical molecular weight / Isoelectric point: 14.00 kDA / 5.44

MYPLLVCADLYGEKYNLEVRFPSKPTIEDLQKRIVTVFSTEMDSRRPEGYPEADFSI
ALVHIYDDKSQQWEKLLSDAQLHEYDQLYVFQPQTRWHMDVQKNLPPPTTPSKDH
RTATPHS

FPC4-9/1 (aa 424-444)

Theoretical molecular weight / Isoelectric point: 2.45 kDA / 9.98

AKEPQRRRVLTSPVPDELLIK

ITC constructs

FPC4-5/1 (aa 394-444)

Theoretical molecular weight / Isoelectric point: 5.77 kDA / 5.09

SLSPYLRYLPSDVSGGEWDKPDVGDVLCFQAKEPQRRRVLTSPVPDELLIK

TbBILBO1-1/8 (aa 1-100)

Theoretical molecular weight / Isoelectric point: 11.79 kDA / 5.28

HMAFLVQVAADIFNNKVNFELSFPSRPSISELTRSAETAFSNEISLRRPDNVPSHKFH
SSKIKMYDEELNKWVDLIREDQLTDYCQLYVFQPPNEWHKESQ

TbBL271-1/3 (aa 1-111)

Theoretical molecular weight / Isoelectric point: 12.32 kDA / 4.91

MMGGISICVATDCDGEKVNLRFLFDAAGPSVSRLLNYSTTAFNNYFRLKGISRAFAV
NSAVVFNDVHCTDRLERTTQLLHNSQVYLFQPDTLDIPAAIPEPYEGEPLLSA

TbBL232-1/5 (aa 1-90)

Theoretical molecular weight / Isoelectric point: 10.66 kDA / 4.61

FFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGISIQADQTPEDLDMEDNDIIEAHR
EQIGG^{Y}HMYPLLVCADLYGEKCNLEVRFPSKPTIEDLQKRIVXVFSTEMDSRRPEGY
PEADFSIALVHIYDDKSQQWEKLXSDAQLHEYDQLYVFQPQ

TbBL271-1/4 (aa 1-90)

Theoretical molecular weight / Isoelectric point: 10.12 kDA / 7.75

DLIKNNPLNADPDYSIAEAAFNKGETAMTINGPWAWSNIDTSXVNYGVTVLPTFKGQ
PSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEE
LAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQT
NSSSNNNNNNNNNNLGHHHHHHHHHHENLYFQ^{Y}GHMMGGISICVATDCDGEKVNL
RFLFDAAGPSVSRLLNYSTTAFNNYFRLKGISRAFAVNSAVVFNDVHCTWDRLERT
TQLLHNSQVYLFQP

Molecular weight and the Isoelectric point were calculated using the Expasy protparam online tool (Gasteiger *et al.*, 2005). TbBILBO1 constructs, TbBL232-1/3, and FPC4 constructs were kindly provided by Johannes Lesigang. TbBL271-1/3 was kindly provided by Paul Majneri (Dong Lab, Max Perutz Laboratories, Austria).

Protein expression and Purification

Transformation and Overnight cultures

1 µl of each plasmid were added to 50 µl of E coli. BL21 (DE3) competent cells for co-expression. E coli. BL21 (DE3) were then grown overnight in 75 ml selective LB-medium with 30 µg/ml kanamycin and 50 µg/ml ampicillin at 37°C and 180 rpm.

Protein expression

The overnight culture was split into three 2L bottles of selective LB-medium each, totalling of 6L medium for each sample. After transferring the bacteria into the new media, they were incubated at 37°C and 160 rpm until OD_{600}=0.6-0.8 was achieved. In order to stop further division of the cells, they were put on ice for rapid cooling of the media. After approx. 20 min on ice, the E coli. were incubated at 18°C for about 45 min while shaking. When the media acquired the desired temperature, the culture was induced using a final concentration of 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Expression of the protein of interest was realized at 18°C and 160 rpm overnight.

Harvest and cell lysis

Cells from overnight expression were harvested by pelletizing at 4°C and 4,000 rpm for 12 min. The bacteria pellets were resuspended in 20-40 ml Histrap running Buffer A. 1 mM PMSF and 2 µg/ml DNAse have been added before the cells were broken by homogenization by a French pressure cell press. The cell lysate was centrifuged at 4°C and 15,000 rpm for 30-50 min to remove the cells debris. The soluble proteins were collected in the supernatant and filtered using a 0.45 µm syringe filter.

Histidine-tagged protein affinity chromatography (Histrap)

The histidine tag protein purification column (histrap column) was equilibrated with 5 column volumes (cv) of histrap buffer A prior to the chromatography. After loading the filtered supernatant, the column was washed with 8 cv of histrap buffer A. Following the washing step, a continuously increasing concentration gradient of histrap buffer B was applied to elute the bound proteins. Finally, the histrap column was washed with 100 % histrap buffer B. The eluted proteins were collected in 1-1.5 ml fractions and their protein compositions were analyzed using SDS-PAGE.

Fractions, which comprised the protein of interest, were combined and stored at 4°C.

Sodium dodecyl sulfate-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was the method of choice to examine protein samples in terms of presence and purity at each purification step. The samples were mixed with 6x SDS loading buffer and boiled at 96°C for 5min before loading onto self-made 15 % SDS polyacrylamide gels. The composition of the 15 % polyacrylamide gel is shown in Table 3.

The electrophoresis was run at 300 V for 30 min.

After running the SDS-PAGE, the gel was soaked in a Coomassie brilliant blue G-250 staining solution for 45 min. The gel was then destained using the destaining solution until the protein bands were clearly visible.

Table 3 SDS PAGE gel composition of the stacking and resolving gel.

Stacking gel		Resolving gel	
4x Tris-HCl pH 6.8	0.50 ml	4x Tris-HCl pH 8.8	1.00 ml
30 % (w/v) Acrylamide/Bis-acryamide solution	0.27 ml	30 % (w/v) Acrylamide/Bis-acryamide solution	2.0 ml
10 % (w/v) APS	16 µl	10 % (w/v) APS	25.00 µl
TEMED	2.55 µl	TEMED	3.0 µl
ddH2O	1.23 ml	ddH2O	1.00 ml

Cleavage of the protein tag

The Small Ubiquitin-related Modifier (SUMO) protein tag was cut using Sentrin-specific protease 2 (SENP2). 1 % (v/v) of SENP2 protease was added to the pooled histrap elutions for tag cleavage. The MBP-tag was cleaved using 2 % (w/w) of TEV protease. The mixture was transferred to a 3.5 K molecular weight cut-off (MWCO) SnakeSkin® Dialysis Tubing for overnight dialysis at 4°C in 4 L of dialysis buffer, to reduce the imidazole concentration for further Histrap experiments.

Reverse Histrap

A reverse histrap was performed to remove the SUMO-tag and isolate the protein of interest. The dialyzed sample was loaded onto a histrap column, pre-equilibrated with histrap buffer A. The flow through (FT) fraction was collected and the column was washed with 2 cv of histrap buffer A. Histrap buffer B was then applied to elute the SUMO-tag.

The eluted fractions of the reverse histrap were analyzed using SDS-PAGE.

Anion Exchange Chromatography (AEC)

The Resource Q anion exchange chromatography column (kindly provided by Djinovic lab) was washed and equilibrated with 5 cv of AEX buffer A prior to the chromatography. After loading the sample, a washing step using 4 cv of AEX buffer A was executed. Then a gradient of continuously increasing AEX buffer B concentration was applied to elute the column bound protein. The elution was collected in 1-1.5 ml fractions, which were analysed by SDS-PAGE.

Size-exclusion chromatography (SEC)

The reverse histrap FT fraction was concentrated via Amicon® Ultra centrifugal filters with a 3 kDA molecular weight cut-off (MWCO). Upon achieving 2-5 ml of sample volume, the protein solution was centrifuged for 10 min at 5,000 rpm and 4°C to eliminate contaminants and precipitated proteins. The Superdex® S75 16/600 size-exclusion chromatography column was pre-equilibrated with 1 cv gelfiltration buffer which also serves as the running buffer during the chromatography. The SEC was run at a flowrate of 1ml/min at 4°C. The eluted proteins were collected in 1-1.5 ml fractions, which were analysed by SDS-PAGE after the chromatography.

Lysine methylation

Lysine methylation was done to facilitate protein crystal formation, by reduction of charge repulsion effects. The purified protein was dialyzed in HEPES-buffer overnight at 4°C in the cold room. Following the dialysis, 20 µl of freshly prepared 1 M dimethyl-borane complex solution and 3.25 µl of 36 % formaldehyde solution was added per millilitre protein solution to the sample under the fume hood. The sample was left stirring for 2 h at 4°C. Again, the same amount of 1 M dimethyl-borane complex and 36 % formaldehyde solution was added, and the incubation was repeated. After 2 h, 10 µl of freshly prepared 1 M dimethyl-borane complex per millilitre protein solution were added and the sample was left stirring overnight at 4°C.

The materials used for Lysine methylation were disposed of as toxic waste.

Isothermal titration calorimetry

The binding affinities between TbBILBO1, TbBILBO1-like proteins and their ligand FPC4 was defined by ITC. The samples were either purified in a SEC in which the ITC buffer serves as the running buffer or were dialyzed against the ITC buffer, before they were concentrated to the intended protein concentration. Prior to the ITC measurement, a water-to-water test run was done to check the device in terms of functionality and purity. If the peak in the water titration test did not exceed -0.02 µcal/sec, the actual experiment was performed. 300 µl of TbBILBO1 or TbBILBO1-like proteins (30 µM) were injected into the sample cell, while 50 µl of the ten times higher concentrated FPC4 (300 µM) was loaded onto the syringe. The initial injection volume was 0.2 µl. The ligand was then titrated into the sample cell in 19 injections of

2 µl each, with a spacing of 150 sec between the injections. Stirring of the sample inside the cell was set to 750 rpm at 25°C.

Crystallization

Crystal screen

The purified protein samples were concentrated to 20 mg/ml for crystallization trials using the sitting-drop-vapor-diffusion method with the Crystal Screen HT™ (Hampton Research, USA) screen. 0.1 µl of buffer were mixed with protein solution in a 1:1 (drop 1) and 1:2 (drop 2) volume ratio respectively (Fig.6) on a 2-drop 96 well crystallization plate. Pipetting of the crystal plates was executed on the mosquito® HTS nanoliter pipetting robot (SPT Labtech, England). The plates were stored in a rock imager crystal hotel device at room temperature (RT) for imaging.

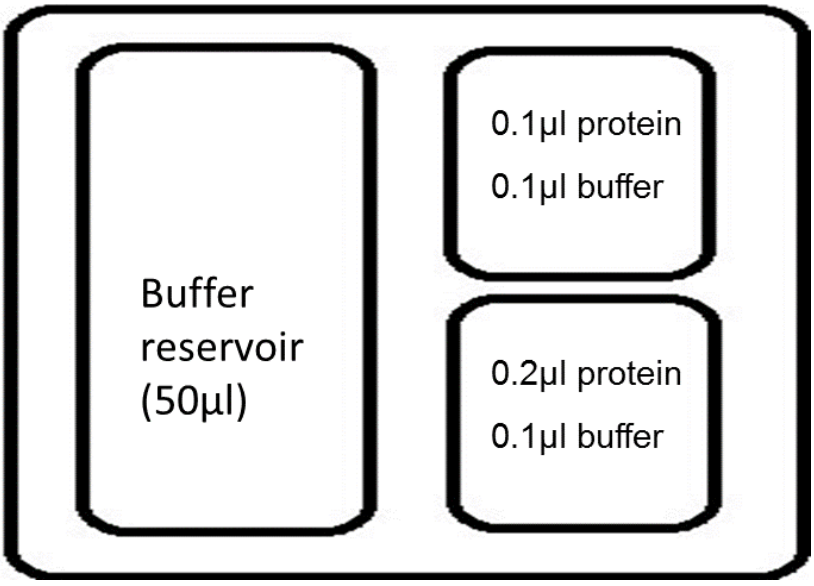


Figure 6 Well-Layout of the sitting-drop-vapor-diffusion crystal plate.

Crystal growth optimization

Optimization plate

This optimization step was performed for the lysine methylated TbBL232-1/3 construct to obtain crystals in larger quantity and size. 200 µl of the custom-made buffer were placed inside a well, which serves as the reservoir liquor. 1 µl of the

same buffer was then mixed with 1 µl of the protein sample on a thin cover slide. The cover slide was placed on top of the reservoir well upside down (Fig.7). The 48-well optimization plate was stored at either room temperature or at 4°C. Crystal growth was monitored via light microscope.

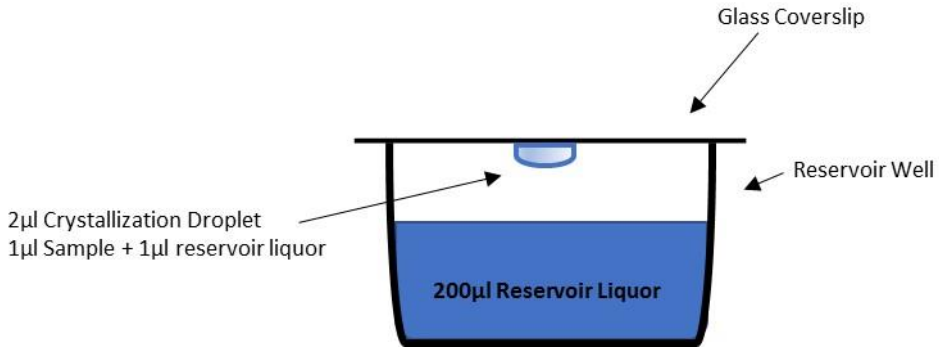


Figure 7 Layout of the hanging drop vapor diffusion method. 1 µl of the protein sample is mixed with 1 µl of the reservoir buffer on the glass coverslip. The glass sheet is then placed on top of the reservoir well.

The original conditions, where either protein nucleation or crystal forming was observed, were in the well position B4 (0.1 M HEPES buffer pH=7.5; 1.5M $LiSO_4$), D5 (0.1 M HEPES buffer pH=7.5; 10 % v/v isopropanol), D7 (30% w/v Polyethylene glycol (PEG) 1.5K), and H8 (0.1M Tris buffer pH=8.5; 20 % (v/v) Ethanol) taken from the Hampton Research Crystal Screen HT™ screen. These buffer conditions were slightly altered in pH, salt, and precipitant concentration to facilitate crystal formation and growth. The crystals found in the optimization plate did not include any protein.

In-situ proteolysis using Trypsin and Chymotrypsin

In addition to the previously described attempts trypsin or chymotrypsin digested BL232 proteins were also used for crystallization trials. Prior to preparation of the crystallization plate, 1 µl of either 1 mg/ml trypsin or 1mg/ml chymotrypsin were added to approximately 50 µl of the concentrated protein (6 mg/ml) BL-protein directly before its distribution onto the 2-drop 96 well crystallization plate by mosquito® HTS nanoliter pipetting robot. The Hampton Research Crystal Screen HT™ served as crystal screen. The plates were stored at 4° and crystal growth was observed manually via light microscope.

Protein crystals were found in the C1 condition (0.1 M Imidazole pH=6.5; 1M Sodium acetate trihydrate) in both drops.

Additive screen

Additive screening was performed for the lysine methylated BL232-1/3 construct, since crystals were detected in the D5 condition of the HR crystal screen HT plate stored at 4°C. 125 nl of the protein was mixed with the same amount of HR D5 condition buffer (D5 buffer) in the first drop with addition of 25 nl of HR Additive screen HT. 200 nl protein solution were mixed with 100 nl buffer, with addition of 30 nl additive in the second drop. The reservoir was filled with custom-made D5 buffer (0.2 M HEPES pH=7.5; 10 % (v/v) Isopropanol; 20 % (w/v) PEG 4K). Pipetting of the crystallization plate was realized by the mosquito® HTS nanoliter pipetting robot. The plate was incubated at 4°C and crystal growth was observed via light microscope. No suitable protein crystals were found.

Crystal harvest

The buffer of the C1 condition, in which protein crystals had been grown, was mixed with a final concentration of 30 % (v/v) glycerol as cryo-protectant. The crystals were first transferred to a fresh drop of buffer from the reservoir. Under the light microscope, the crystals were picked individually and soaked in a drop of buffer-glycerol solution. The cryo-protectant buffer was changed several times, by removing the buffer with a pipet and adding fresh cryo-protectant buffer, without moving the crystal. Finally, the crystal was picked up with a nylon loop and transferred into a pre-cooled crystal puck in liquid nitrogen and stored in a liquid nitrogen filled dewar.

Data collection

Crystal data were collected at the European Synchrotron Radiation Facility (ESRF) on a microfocus beamline (ID29) with tuneable wavelengths. The general strategy was to start screening with a resolution set to 2.5 to 3 Å. Two to four images were initially taken with automated centering of the crystal, automated processing of the image and extended exposure time.

The actual screening was performed with a resolution about 0.2 Å lower than the suggested value. For space groups P1 and P2 a full circular sweep of 360° with the suggested oscillation range was collected. For space groups P3, P4 and P6, if the

suggested sweep angle was higher than 90°, a sweep was collected with a sweep angle two times higher than the suggested value, without changing the suggested oscillation range. If the suggested sweep angle was lower than 90°, a half circle sweep of 180° was collected with the suggested oscillation range.

Data processing

The protein model was build using COOT, a map-fitting tool for structure determination (Emsley *et al.*, 2010). Since a molecular model of the structurally related TbBILBO1 NTD (the PDB file was kindly provided by Gang Dong, Max Perutz Laboratories, Austria) already exists, molecular replacement with Phaser was the method of choice to solve the crystal structure of the protein complex of interest (Mccoy *et al.*, 2007). The residues were manually fitted into the electron density cloud. Conformationally unfavored regions were visualized by the Ramachandran plot. The refinement step was executed using the phenix refinement tool (Afonine *et al.*, 2012).

The generated protein data bank (PDB) file was used to render a three-dimensional image in PyMOL for visualization of the TbBL232 N-terminal domain.

Results

Purification of TbBL232-1/3-Sumo15b and FPC4-9/1-Sumo28a Co-Expression

Histidine-tagged protein affinity chromatography

The initial Histrap was performed to isolate the histidine-tag carrying constructs from the purified cell lysate. The protein of interest bound to the nickel packed resin of the chromatography column and was subsequently eluted using an increasing imidazole concentration gradient. The fractionated elution displayed an UV-absorbance signal at 280 nm from fraction A9 to C3 (Fig. 8). These fractions were collected for Sumo-tag cleavage.

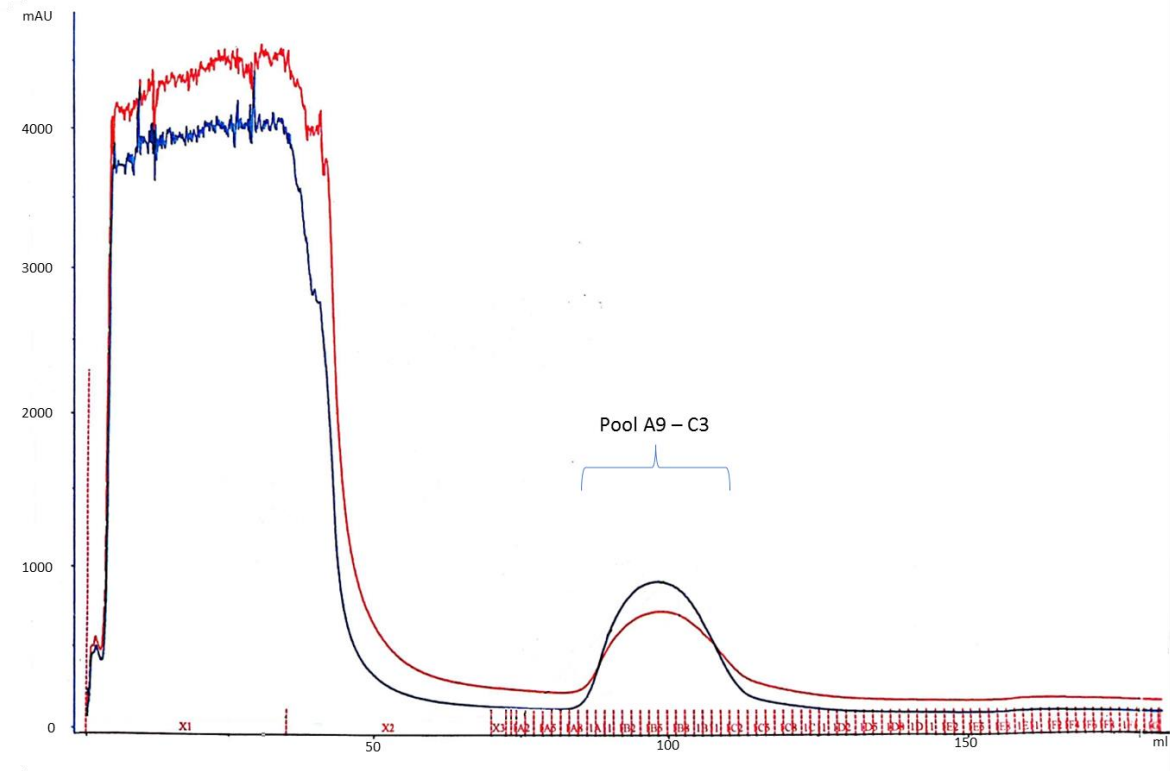


Figure 8 Chromatogram of the initial histrap of TbBL232-1/3-Sumo15b and FPC4-9/1-Sumo28a Co-expression. A single fluorescence peak (Blue: A_{280}; Red: A_{260}) was detected in the fractionation step achieved by an increasing Histrap elution buffer B concentration gradient. The fractions A9 - C3 were collected and pooled.

Sumo-tag cleavage and Reverse Histrap

The pooled fractions obtained from the initial Histrap amounted to a volume of 28 ml. Therefore, 280 µl of SENP2 protease were added to the pool and the sample was dialyzed at 4°C overnight. After dialysis and his-tag cleavage, a small aliquot was taken for later gel electrophoresis experiments together with an uncut sample, to verify the removal of the protein-tag. Precipitation was observed in the dialyzed protein solution. For this reason, it was filtered for reverse Histrap. The TbBL232-1/3-Sumo15b protein band was found in the uncut fraction at 30 kDA. The intensity of it was reduced in the cut fraction and the formation of the bands at 15 kDA indicate, that tag cleavage took place. The other protein band in the uncut fraction at 20 kDA is presumed to be FPC4-9/1-Sumo28a. The ligand was freed from its Sumo-tag upon SENP2 addition. Since both protein constructs were removed from their histidine-tags as well, TbBL232-1/3 and FPC4 were found in the flowthrough and the washout. Both fractions were combined for further purification (Fig.9).

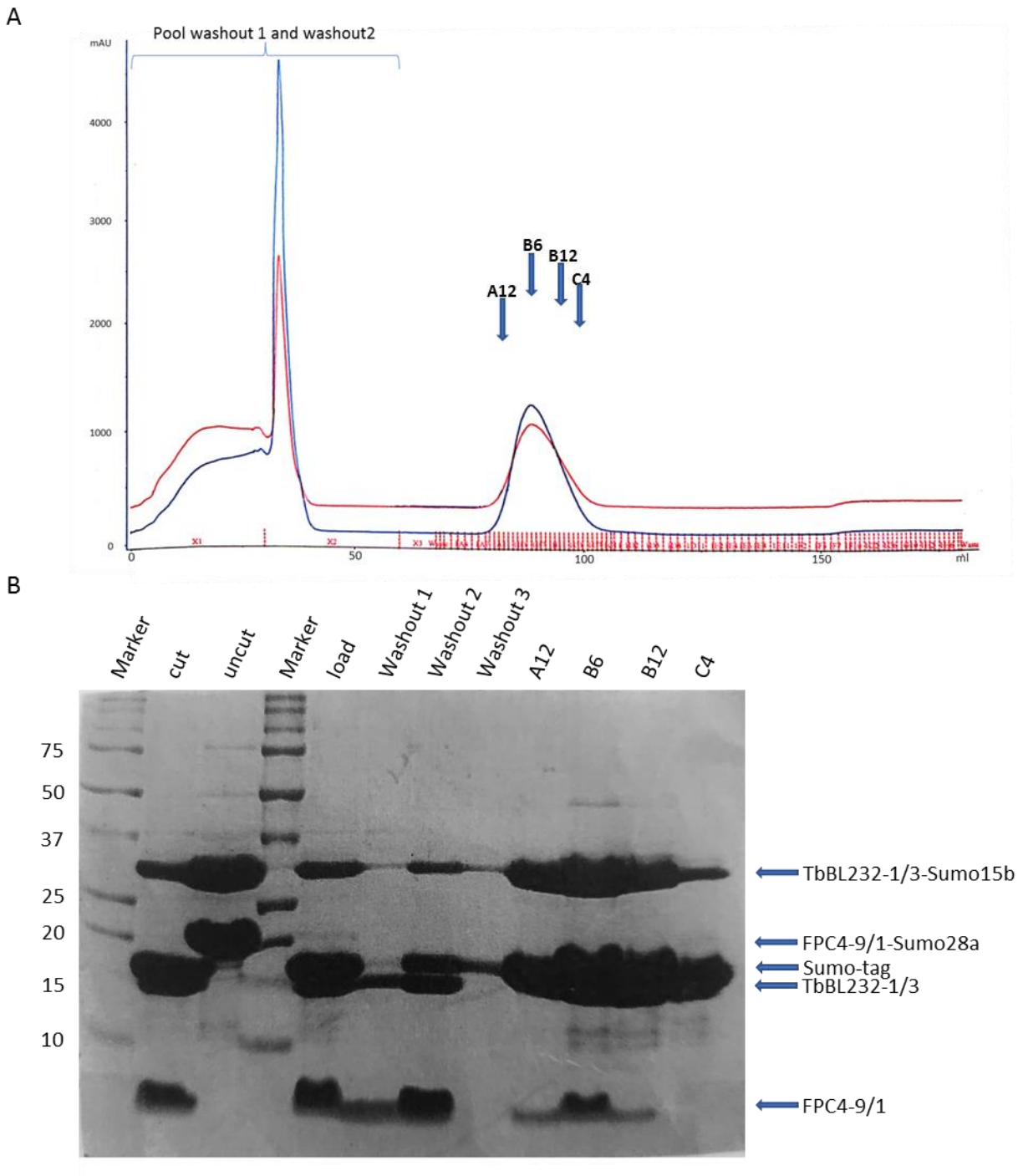


Figure 9 (A) Reverse Histrap was applied after SENP2 mediated cleavage of the Sumo-tags. The co-expressed proteins did not bind to the column and therefore, were washed out. The his-tag carrying Sumo-tags were bound by the column and thus were separated from its former fusion partner. (Blue: A_{280}; Red: A_{260})
(B) SDS-PAGE gel. Left side of the central marker displays the efficacy of the protein-tag cleavage. FPC4-9/1-Sumo28a (20 kDA) was completely digested by SENP2. However only a partition of expressed TbBL232-1/3-Sumo15b (30 kDA) was stripped of its Sumo-tag. The fraction's content of the reverse Histrap is shown on the right side of the central marker. The Sumo-tag and the uncut protein were caught on the Histrap column and eluted in the fractionation step. The SENP2 digested proteins were mostly found in the washout along with some tagged TbBL232 and Sumo protein.

Anion exchange chromatography

Prior to the AEX, the pooled flowthrough and washout fraction were diluted with an equal volume of 20 mM Tris-HCl (pH=8) buffer to reduce the final sodium chloride concentration to 50 mM. The positively charged Resource™ Q column was used, due to the low theoretical isoelectric point of TbBL232-1/3 of around 5.5 and the alkaline conditions of the buffers. This is presumed to give the TbBILBO1-like protein a negative charge. However, the target proteins were found in the flowthrough and washout fraction in a high purity, whereas the Sumo-tag and uncut TbBL232 proteins bound to the column and were eluted by the increasing AEX buffer B concentration gradient. Therefore, the flowthrough and the washout fractions were combined (Fig.10).

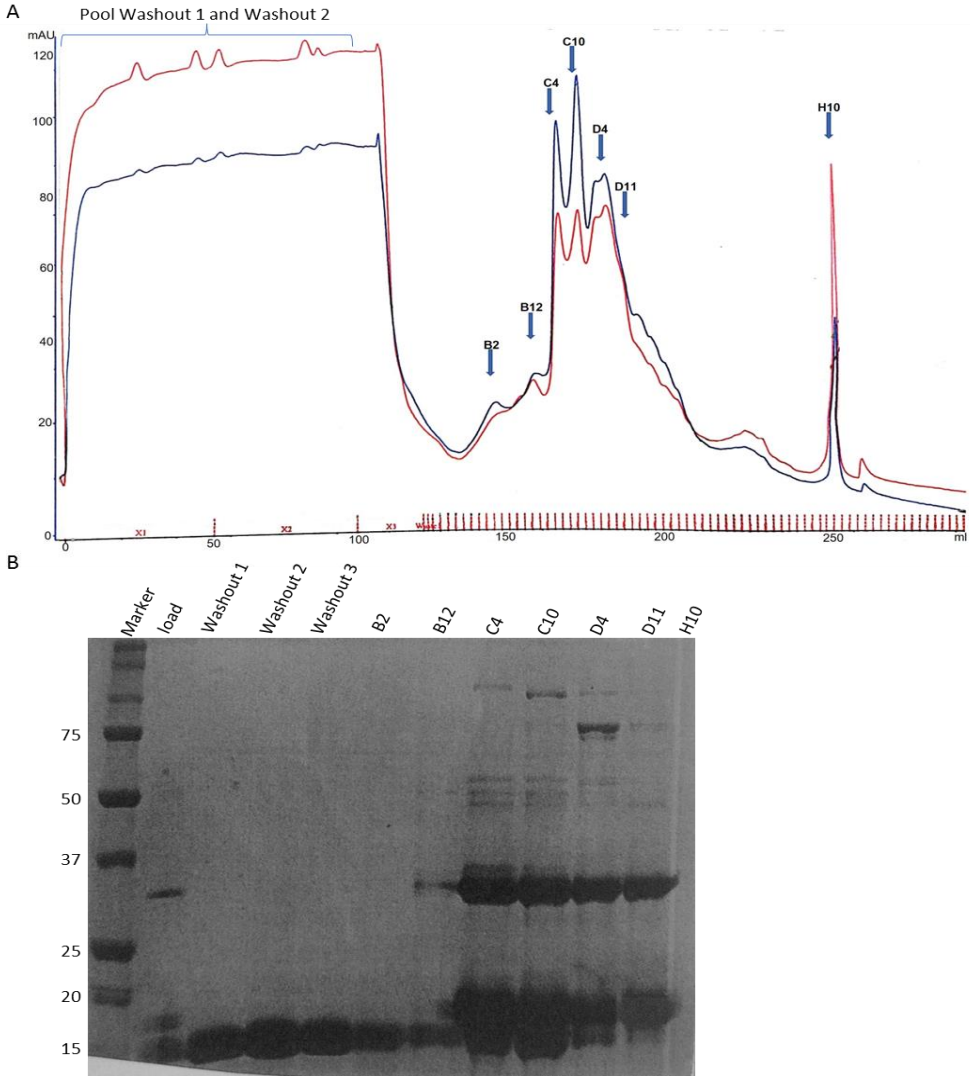


Figure 10 (A) Chromatogram of the Anion exchange chromatography. (Blue: A_{280}; Red: A_{260})
(B) The tagged TbBL232 proteins were found collectively with the Sumo-tag in the AEX buffer B gradient fractionation step at 30 kDA and 17 kDA, respectively. The untagged TbBL232 protein band was present in the washout fraction at 15 kDA with a high purity. Therefore washout 1 and 2 were combined.

Size-exclusion chromatography

SEC samples were concentrated to around 3 ml before size-dependent fractionation on the Superdex® S75 16/600 gelfiltration column. The chromatogram revealed only one elution peak indicating, that both binding partners were eluted together. Analysis via gel electrophoresis confirmed the co-elution of both targets. Hence, the fractions covered by the peak area, were combined for concentration (Fig.11).

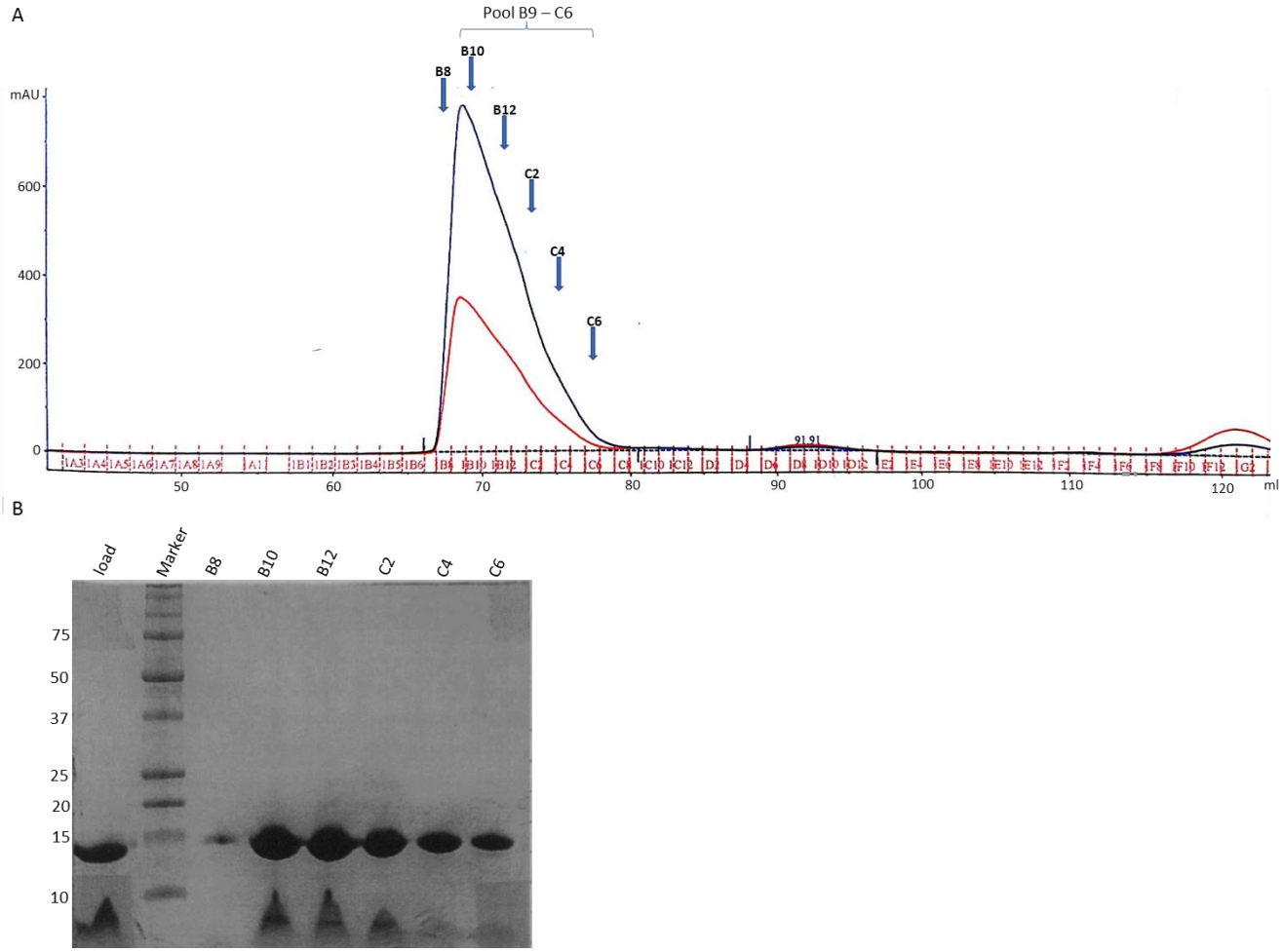


Figure 11 (A) The Co-expressed proteins eluted in together in the SEC chromatogram. (Blue: A_{280}; Red: A_{260}) (B) TbBL232 was found at the 15 kDA region whereas the distorted band below the 10 kDA marker was presumably FPC4. The fractions B9 – C6 were pooled.

Crystal structure of TbBL232 NTD

The resolution determined for the crystal was 1.2 Å. The unit cell displayed an orthorhombic system with an angle of α = β = γ = 90° and axial lengths of a = 31.82, b = 35.02 and c = 84.95. The Space group was categorized as $P2_12_12_1$ (Table 4). TbBILBO1 NTD served as a model for molecular replacement using the phaser software (Mccoy *et al.*, 2007). After the residues were fit to the electron density cloud,

refinement was achieved using the phenix.refine program suite (Afonine *et al.*, 2012). While building the model, a premature absence of the electron density cloud at the residue position 92 was observed. TbBL232-1/3 was expected to be a construct consisting of 119 aa. Therefore, 28 residues at the C-terminal domain were lost. In addition, TbFPC4 was found to be missing as well.

Table 4 Data collection and refinement statistics of TbBL232-NTD

Data collection	
Source	ID29 (ESRF)
Resolution (Å)	42.47 - 1.20
Space group	$P2_12_12_1$
Unit cell dimensions (Å, °)	a = 31.82, b = 35.02, c = 84.95
	α = β = γ = 90
No. of reflections	30156
V_M ($Å^3$ Da^{-1})	2.12
MW (Da)	11171*
*Estimated MW of the truncated TbBL232 NTD	
Refinement	
Resolution (Å)	18.16 – 1.20
R–work / R-free (%)	23.1 / 25.2
Completeness (%)	98.70
Molprobity statistics	
All-atom clashscore	1.90
Ramachandran plot	
Outliers (%)	0.00
In allowed regions (%)	1.06
In favored regions (%)	98.94
Rotamer outliers (%)	0.00

The NTD of TbBL232 was found to adopt a ubiquitin-like fold like TbBILBO1 NTD. It consists of an α-helix and five β-strands (Fig.12).

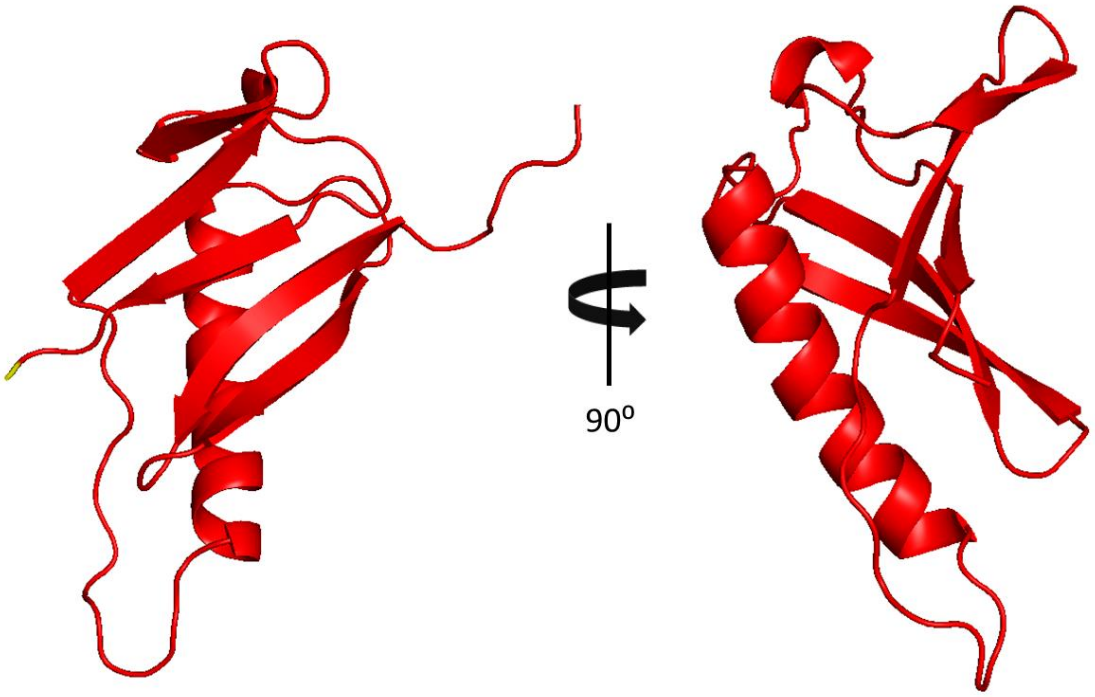


Figure 12 Simplified illustration of the TbBL232 NTD (aa 1-91). It consists of an α-helix and five β-strands. The truncated C-terminal end is highlighted in yellow. The illustration was created using the PyMOL visualization tool.

However, the C-terminal residues 96-115 of TbBILBO1 form a long folded loop that winds arounds the α-helix and the N-terminal β-sheets, which appears to be a dynamic structure in NMR spectroscopy measurement (Vidilaseris *et al.*, 2020). Previous studies have shown, that TbBL271 NTD forms a similar structure. As a closely related homolog, TbBL232 NTD is believed to exhibit a loop as well (Fig.13).

A

```
TbBL271       MMGGISICVATDCDGEKVNLRFLFDAAGPSVSRLLNYSTTAFNNYFRLKGI----SRAFA      56
TbBILBO1      -HMAFLVQVAADIFNNKVNFELSFPS-RPSISELTRSAETAFSNEISLRRPDNVPSHKFH      58
TbBL232       ---MYPLLVCADLYGEKYNLEVRFPS-KPTIEDLQKRIVTVFSTEMDSRRPEGYPEADFS      56
                    : *.:*  .:* *:.. * :  *::. * .   *.*.. :  :      .  *

TbBL271       VNSAVVFNDVHCTWDRLERTTQLLHNSQVYLFQPDTLD-------IPAAIPEPYEGEPLL      109
TbBILBO1      SSKIKMYDEELNKWVDLIREDQLTDYCQLYVFQPPNEWHKESQKEIPPAMKPPSSGQRHS      118
TbBL232       IALVHIYDDKSQQWEKLLSDAQLHEYDQLYVFQPQTRWHMDVQKNLPPPTTPSKDHRTAT      116
                   ::::    *  *    ** .  *:*:*** .        :*       . .

TbBL271       SA-       111
TbBILBO1      AGG       121
TbBL232       PHS       119
```

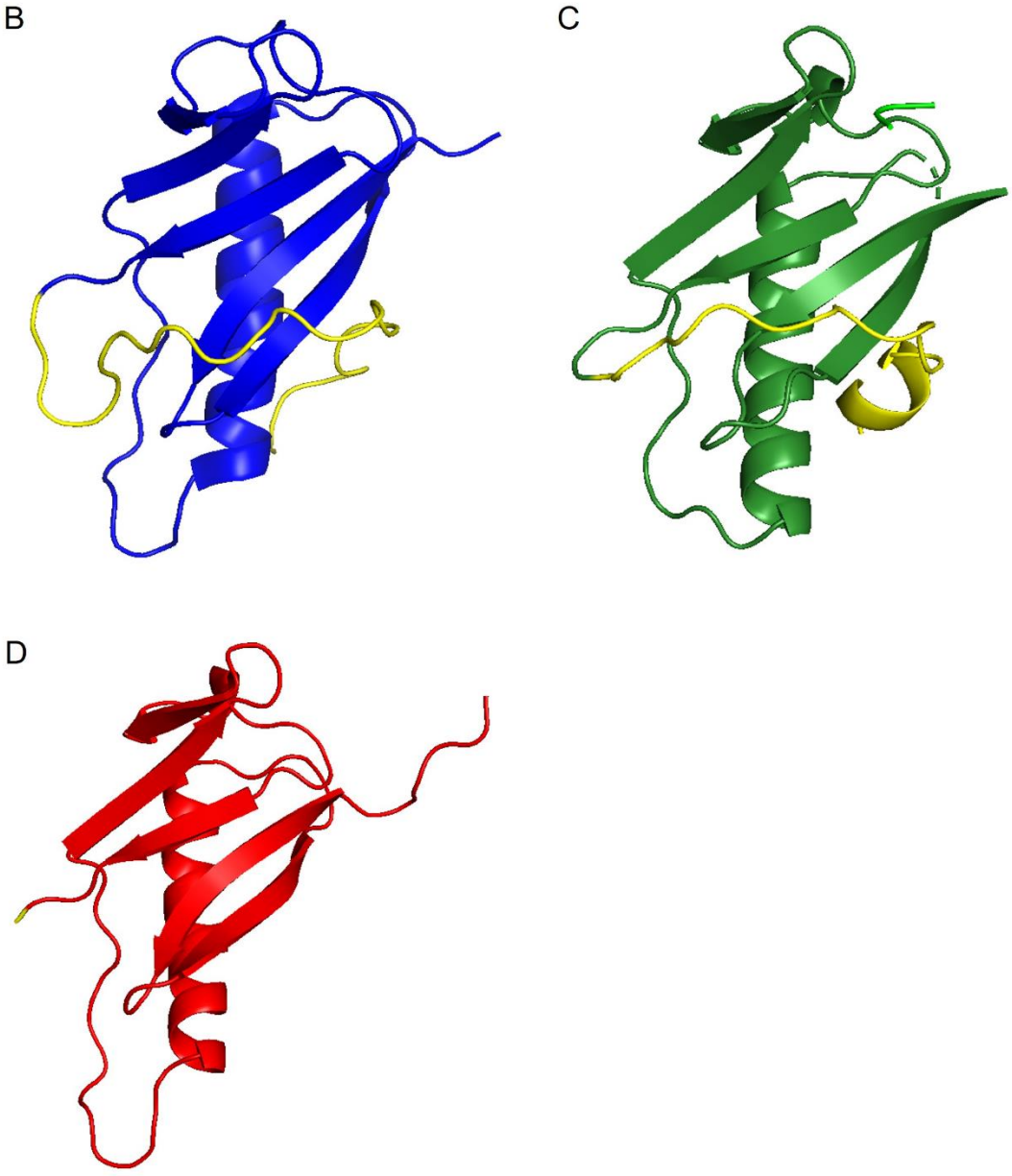


Figure 13 (A) Alignment of full length TbBL271, TbBILBO1 and TbBL232 N-terminal domains. Asterisk labelled residues display identical amino acids. Colon labelled positions display amino acid substitutions with similar chemical properties. Period labelled residues display semi-conservative substitutions.

Comparison of TbBILBO1 NTD (B) and its homologs TbBL271 (C) and TbBL232 (D). Apart from the sequence similarity, the Bilbo-like proteins also adopt a similar structure like BILBO1. The C-terminal loop region, which winds around the α-helix and the two N-terminal β is highlighted in yellow. TbBL232 is also believed to form a similar structure.

Cloning of TbBL232-1/5-Sumo15b and TbBL271-MalpET constructs for ITC measurement

Amplification of TbBL232-1/5 and TbBL271-1/4 DNA inserts were performed using the Q5® High-Fidelity PCR Kit, where the existing TbBL232-1/3-Sumo15b plasmid served as the template. Analysis of the PCR product were done using a 1.2 % (w/v) agarose gel. Appropriate DNA bands were excised from the gel and purified using the Wizard© SV Gel and PCR Clean-Up System. The extracted TbBL232-1/5 and TbBL271-1/4 DNA fragments were digested by NdeI and BamHI restriction enzymes and purified again using the Wizard© SV Gel and PCR Clean-Up System. A nucleic acid concentration of 20.7 ng/µl for TbBL232-1/5 and 11.8 ng/µl for TbBL271-1/4 was determined by measurement using the NanoDrop™ UV-Vis spectrophotometer. Tb232-1/5 inserts were ligated into the Sumo15b vector, whereas the TbBL271-1/4 insert was fused to a MalpET vector.

The novel plasmids were transformed to *E. coli* DH5α cultures for duplication. The generated Plasmids were extracted using the Wizard® Plus SV Minipreps DNA Purification System. TbBL232-1/5-Sumo15b as well as TbBL271-1/4-MalpET were eluted with a concentration of approximately 110 ng/µl, and 12 µl of each construct were sent in for sequencing. Since the novel generated constructs TbBILBO-1/8, TbBL271-1/4 and TbBL232-1/5 are missing the dynamic loop region they will be termed loopless.

Expression of ITC samples

In addition, TbBILBO1-1/8-MalpET, FPC4-5/1-MalpET (kindly provided by Johannes Lesigang) and TbBL232-1/3-Sumo15b were expressed and purified along with the two novel constructs. Purified TbBL271-1/3 protein was kindly provided by Paul Majneri. Transformation was done by transferring 1 µl of plasmid DNA to 50 µl competent *E. coli* BL21 (DE3). Upon plasmid uptake, the cells were inoculated to 25 ml LB-medium and incubated overnight. The overnight cultures of the three constructs were transferred to 2 L LB-media, respectively. The cells were left growing at 37°C until the optical density of the microbial cultures ranged between 0.6 to 0.8 before induction using IPTG. Protein expression was accomplished overnight at 18°C. The bacteria were lysed, and the obtained lysates were cleared for Histrap.

Histidine-tagged protein affinity chromatography of ITC samples

Protein isolation was achieved by Histrap. The histidine-tag carrying constructs bound to the nickel resin of the Histrap column were slowly eluted by an increasing Histrap buffer B gradient in 1.5 ml fractions.

TbBILBO1-1/8-MalpET eluted with two distinct UV-absorbance peaks of different height. SDS-PAGE analysis revealed, that both elution signals were emitted by the target construct, but the first eluting minor peak carried significantly more impurities and less protein compared to the major peak (Fig.14). Therefore, only the fractions A8 to B10 of the later elution peak were combined for further purification.

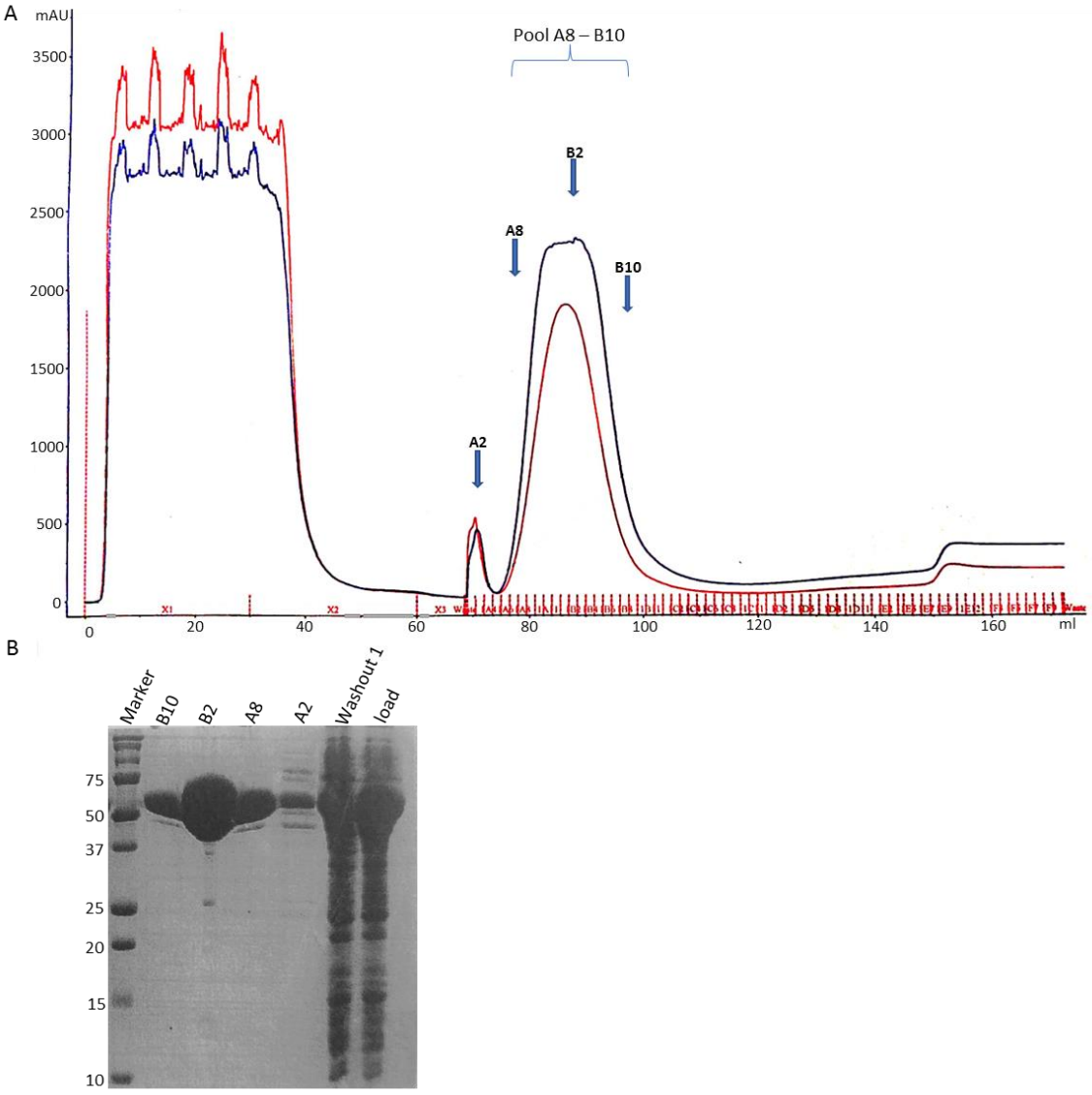


Figure 14 (A) Histrap chromatogram of TbBILBO1-1/8-MalpET. (Blue: A_{280}; Red: A_{260}) Two UV-absorbance peaks are present in the elution buffer B gradient fractionation step.
(B) SDS-PAGE analysis reveal that the fractions of the weaker absorbance peak contained TbBILBO1-1/8-MalpET (55 kDA) along with some unspecific proteins. The fractions of the later absorbance peak included the target protein at a higher purity. The fractions A8 – B10 were pooled.

Elution of TbBL271-1/4-MalpET resulted in a single weak UV-absorbance signal of less than 500 mAU, indicating a low expression or purification yield. SDS-PAGE analysis suggested that a part of the target construct remained in the flowthrough and washout fraction of the chromatography, due to a broad visible band in the first washout fraction (Fig.15). The pooled fractions of the three loopless constructs were split, and half of the sample were removed of their tags where the other half were applied directly to the SEC column.

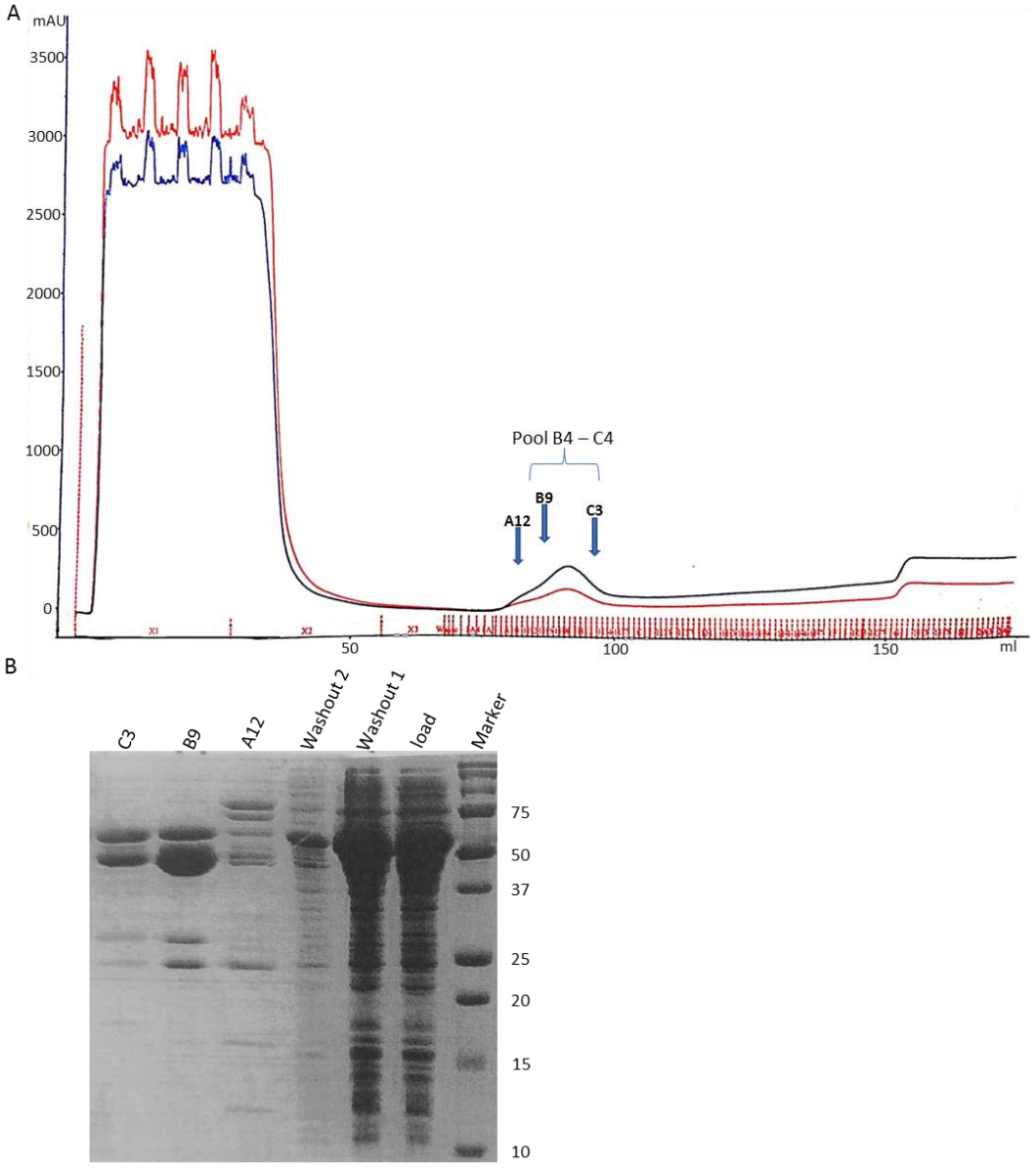


Figure 15 (A) Histrap purification of TbBL271-1/4-MalpET. (Blue: A_{280}; Red: A_{260}) The shallow absorbance peak (< 500 mAU) indicates a low expression yield or poor column binding.
(B) The target protein (54 kDA) is found in the fractions C3 and B9 along with another band at 50 kDA. However, the first Washout fraction also contains a considerable amount of TbBL271-1/4-MalpET, since a large band above the 50 kDA marker was observed. The fractions B4 – C4 were combined.

Measurement of UV-absorbance for TbBL232-1/5-Sumo15b elution fractions resulted in two distinct signal peaks. The sample material, which eluted first, absorbed visible UV light at 280 nm, whereas the later signal absorbed UV light at 260 nm. SDS-PAGE analysis showed that the protein of interest was mainly present in the major peak fractions, which eluted first. The fractions covered by the minor peak only contained little protein. Hence, the fractions A12 to B10 were pooled (Fig.16).

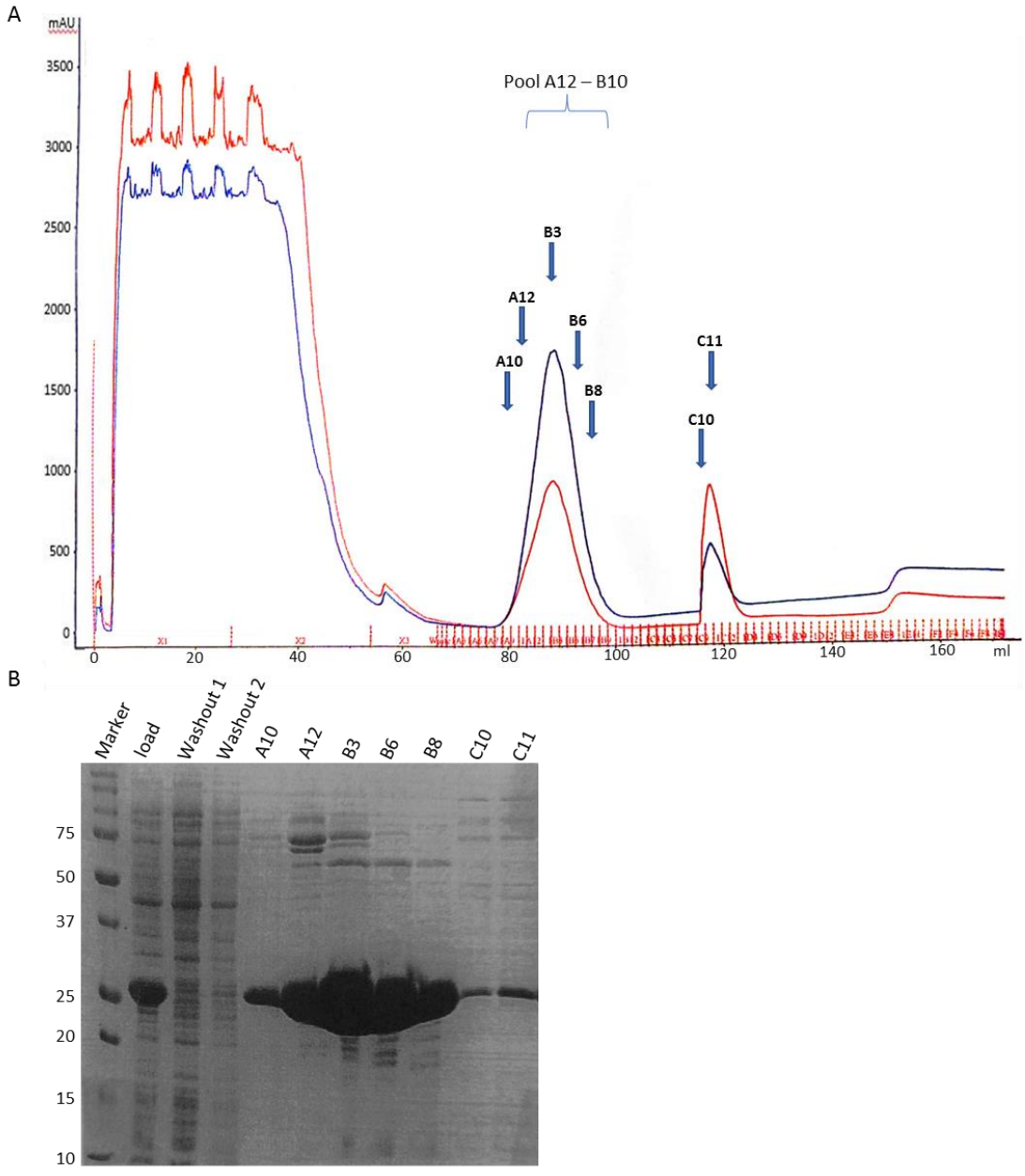


Figure 16 (A) Chromatogram of TbBL232-1/5-Sumo15b Histrap purification. (Blue: A_{280}; Red: A_{260}) Two UV-absorbance peaks were found in the plot, whereas the earlier signal showed absorption at 280 nm and the latter signal at 260 nm wavelength.
(B) The content of the fractions was analyzed via SDS-PAGE. The expressed protein was mainly found in the first peak. The fractions of the second peak included only a small amount of protein. Therefore, the fraction A12 – B10 were pooled.

TbBL232-1/3-Sumo15b and FPC4-5/1-MalpET were eluted in a single peak in the early fractions of the imidazole concentration gradient. The desired proteins were found to be in a clean state with little impurities on the SDS-polyacrylamide gel. However, the FPC4 sample formed another band below the actual 50 kDA protein construct's band, which might be a product of degradation. The peak fractions were combined and concentrated for SEC on a S200 16/600 chromatography column (Fig.17).

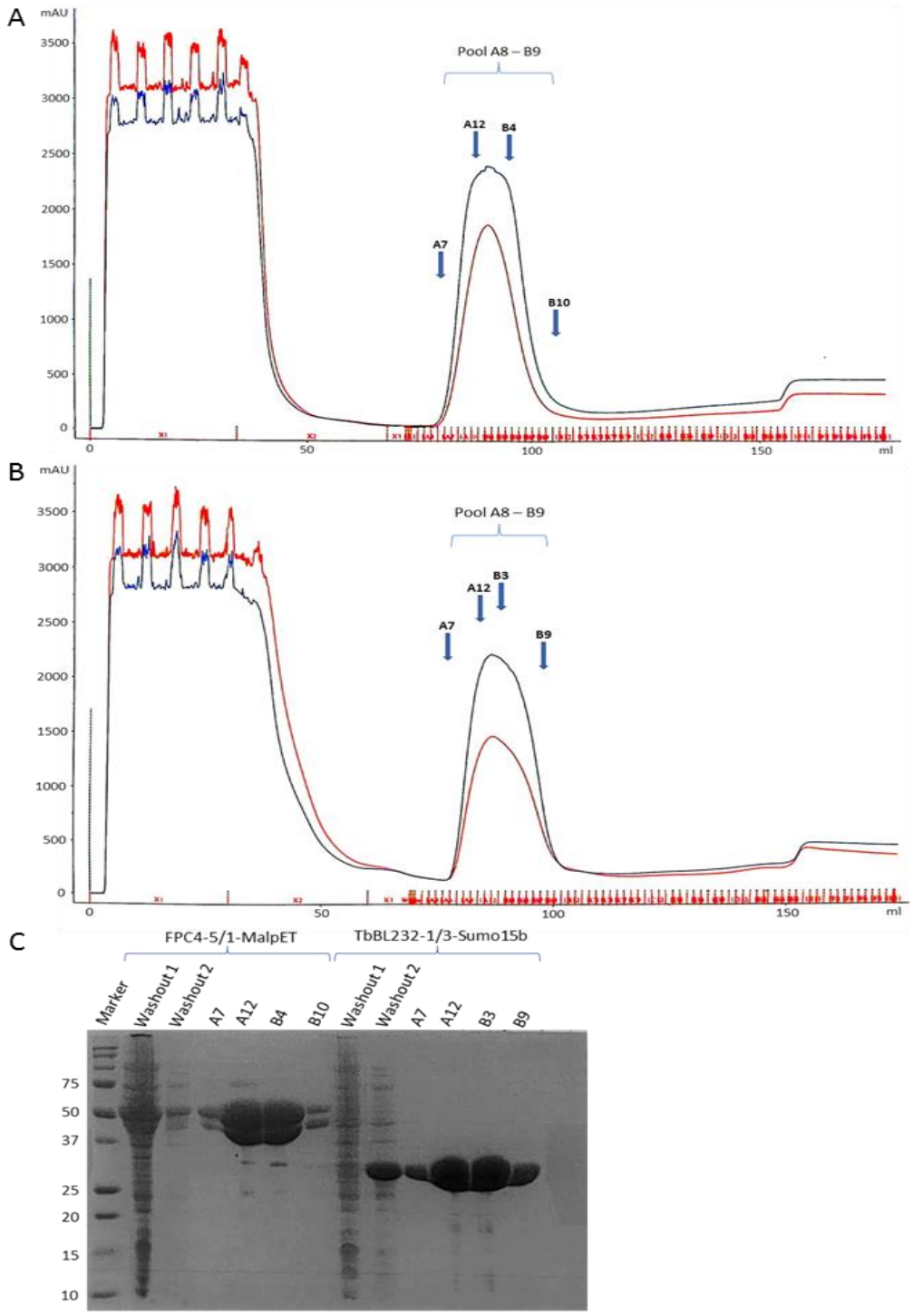


Figure 17 Histrap chromatogram of FPC4-5/1-MalpET (A) and TbBL232-1/3-Sumo15b (B). (Blue: A_{280}; Red: A_{260}) Both constructs displayed a single absorbance peak in the elution gradient at a similar Histrap buffer B concentration.
(C) FPC4 construct was loaded onto the left side of the SDS-PAGE gel. The band was located at the 50 kDA marker. However, another distinct band could be detected below FPC4, which might be due to degradation. The TbBL232 protein band was located at 30 kDA on the right side of the gel. The fractions A8 – B9 of both constructs were pooled respectively.

Tag cleavage and Reverse Histrap

TbBILBO1-1/8-MalpET, TbBL232-1/5-Sumo15b and TbBL271-1/4-MalpET were stripped of their tags. SENP2 was used to cut the Sumo15b protein-tag, whereas TEV cleaved the MalpET-tag. After tag removal and overnight dialysis, the samples were reapplied to the Histrap. The presence of the target proteins in the flowthrough and washout were checked by SDS-PAGE before pooling. Sumo-tag and MBP-tag proteins collected in the eluted fractions were taken as a non-template control for ITC measurement. TbBIL271-1/4, which was already poorly isolated on the Histrap, was no longer detectable on the gel. The MBP-tag of the missing TbBL271-1/4 was found in the acrylamide gel, indicating that the unstable loopless construct precipitated before loading onto the nickel column. Tb232-1/5 and TbBILBO1-1/8 were stripped off their tags and collected in the washout. (Fig.18). Due to the instability of the untagged TbBIL271 construct the protein tags were left on all samples for ITC measurement.

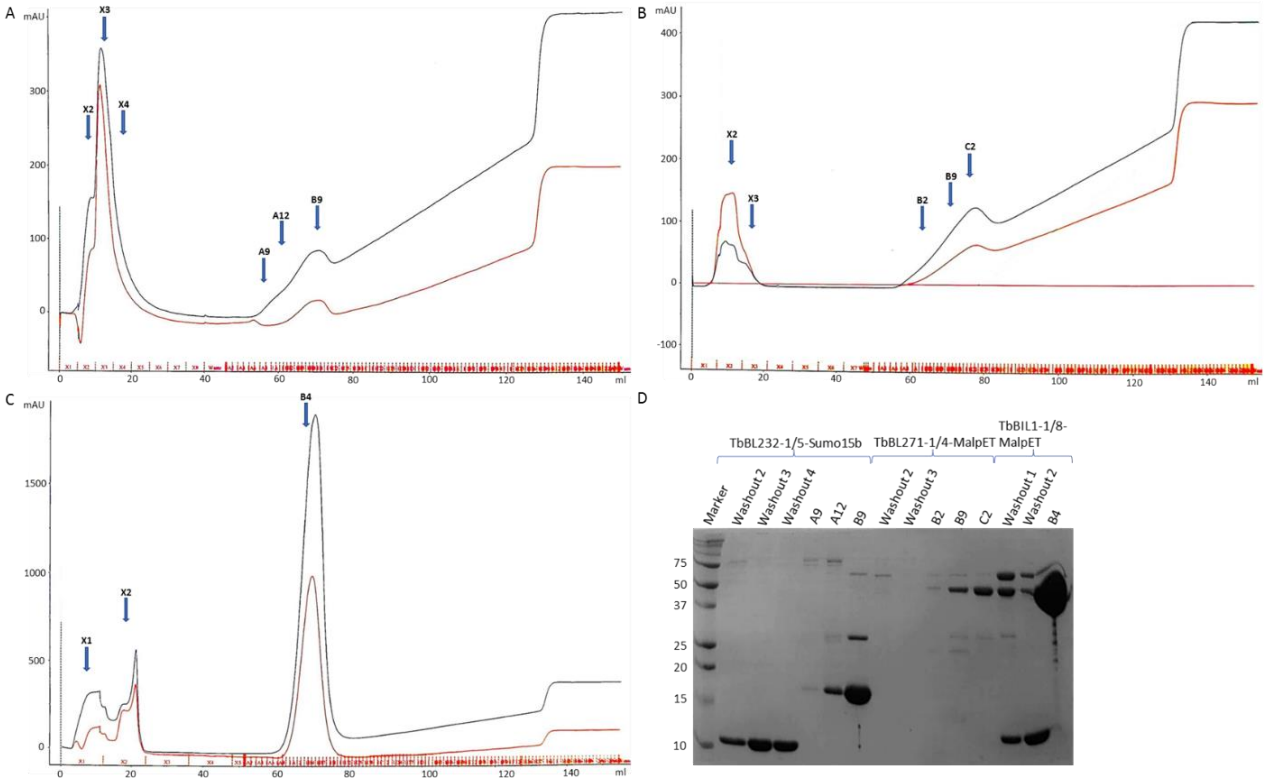


Figure 18 Reverse Histrap chromatogram of the SENP2 digested loopless TbBL232 (A), TbBL271 (B) and TbBILBO1 (C) NTDs. No absorbance at 280 nm could be detected in the TbBL271 sample. TbBL232 and TbBILBO1 eluted with a A_{280} signal of approximately 350 mAU and 550 mAU, respectively. (Blue: A_{280}; Red: A_{260}) (D) The fractions of TbBL232 (10 kDA) were loaded onto the left side of the SDS-PAGE gel. The Sumo-tag (15 kDA) was caught on the nickel resin of the column, whereas TbBL232 were found in the washout. TbBL271 could not be found in the washout fractions. TbBILBO1 (12 kDA) on the right side of the gel, eluted in the washout fraction. The MBP-tag (43 kDA) was eluted in the Histrap buffer B fractionation step.

Size-exclusion Chromatography

The samples were filtered using Amicon® Ultra centrifugal filters (Merck, Germany) with a MWCO according to the size of the construct. The concentrated probes were then either applied to Superdex® S75 16/600 or Superdex® 200 16/60 columns depending on their molecular weight and/or column availability. ITC buffer was used as the running buffer instead of the Gel filtration buffer.

The TbBILBO1-1/8-MalpET construct were loaded onto the Superdex® S75 16/600 gelfiltration column. Two absorbance peaks were observed when eluting the sample. SDS-PAGE analysis showed that both signals were emitted by the target construct.

The chromatogram of TbBL232-1/5-Sumo15b, which was loaded onto the Superdex® S75 16/600 column, displayed two signal peaks. The early peak showed absorbance at 280 nm, whereas the latter absorbs at 260 nm wavelength. Therefore, only the content of the first peak was observed by SDS-PAGE. The tagged TbBL232-1/5 was detected above the 25 kDA marker in the resolving gel.

TbBL232-1/3-Sumo15b was purified using the Superdex® 200 16/60 gelfiltration column. A strong absorbance signal followed by three weak signals was detected in the chromatogram. By screening the absorbance peak content, the presence of the target construct was detected.

FPC4-5/1-MalpET was filtrated using the Superdex® 200 16/60 column. Three absorbance peaks were observed in the fractionation step. SDS-PAGE analysis showed that all signals were given by the target constructs.

The fraction with a suitable protein concentration and purity were taken for ITC measurement (Fig.19).

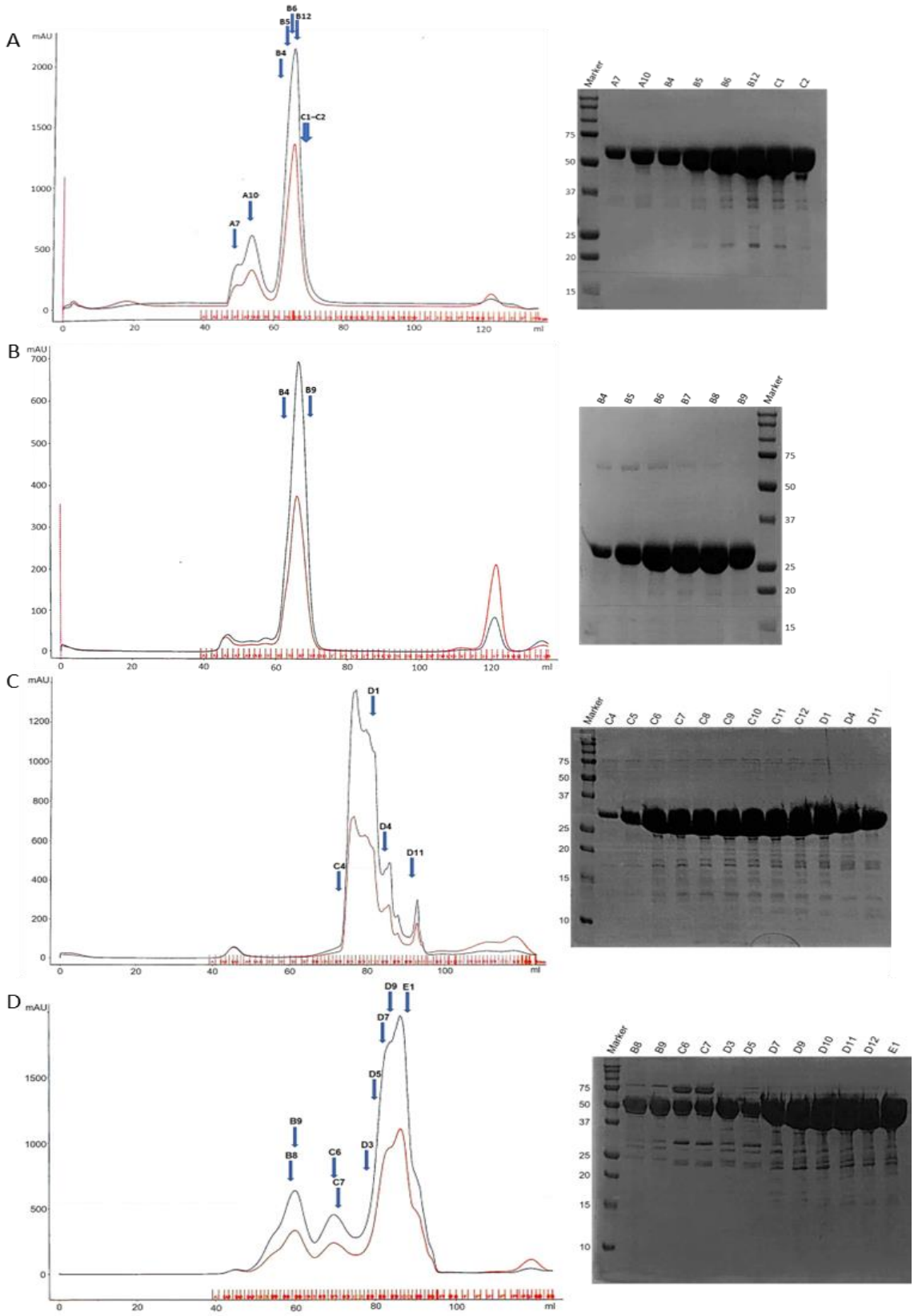


Figure 19 (A) Two absorbance peaks has been detected in the TbBILBO1-1/8-MalpET (55 kDA) elution fractions. Both signal peaks were generated by the target construct, according to SDS-PAGE analysis. A fraction covered by the second peak was taken for ITC measurement. (B) TbBL232-1/5-Sumo15b (26 kDA). A fraction of the major peak was taken for ITC. (C) TbBL232-1/3-Sumo15b (30 kDA). A fraction of the major peak was taken for ITC (D) In the fractionation of FPC4-5/1-MalpET (49 kDA) three absorbance peaks were detected. SDS-PAGE examination revealed that these absorbance signals were created by the target construct. A fraction of the third peak was taken for ITC measurement. (Blue: A_{280}; Red: A_{260})

However, purification of TbBL271-1/4-MalpET resulted in a low protein yield. Therefore, fractions B6 – B9 were pooled and concentrated for ITC measurement (Fig.20).

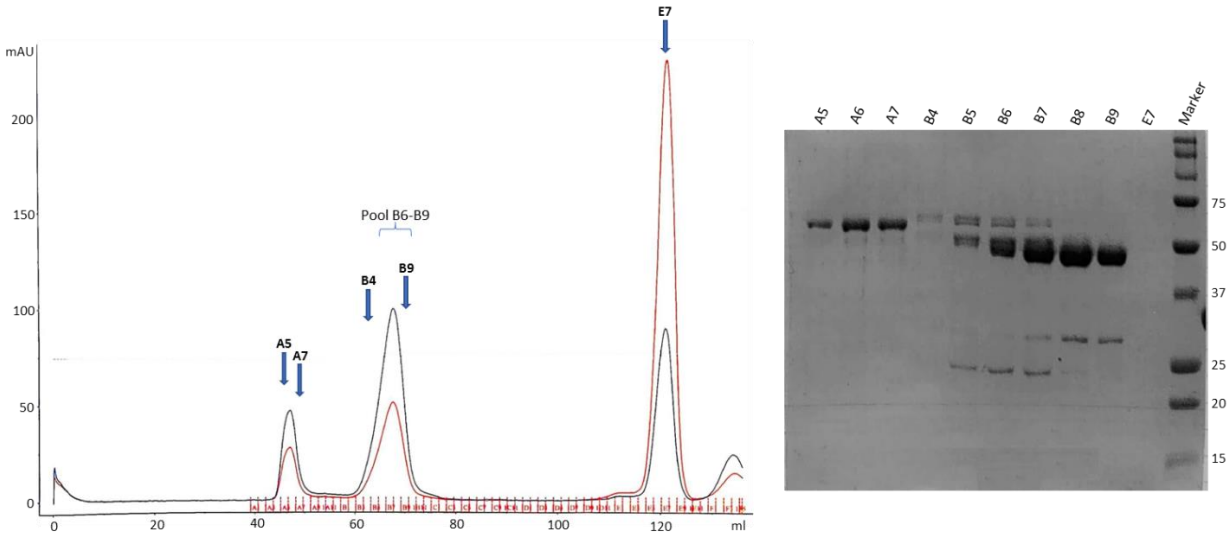


Figure 20 The SEC chromatogram of TbBL271-1/4-MalpET (52 kda) shows three absorbance peaks. (Blue: A_{280}; Red: A_{260}) The low absorbance signal and SDS-PAGE analysis indicate a low protein concentration. Due to the low concentration of the target protein, the fractions B6 – B9 were combined.

Isothermal Titration Calorimetry

ITC was the method of choice to determine if ligand binding of the loopless constructs takes place in comparison to the wildtype NTDs of TbBL271 and TbBL232. Due to the instability of untagged TbBL271 all samples as well as the ligand retained their protein tag for this binding experiment. Prior to loading the actual protein of interest, the cleaved maltose-binding protein tag was titrated against the SUMO protein. No reaction was observed in this non-template control experiment. In addition, FPC4 were titrated again the ITC buffer to ensure there was no unspecific reaction. As no binding was measured, the BILBO constructs were tested against FPC4 (Fig.21).

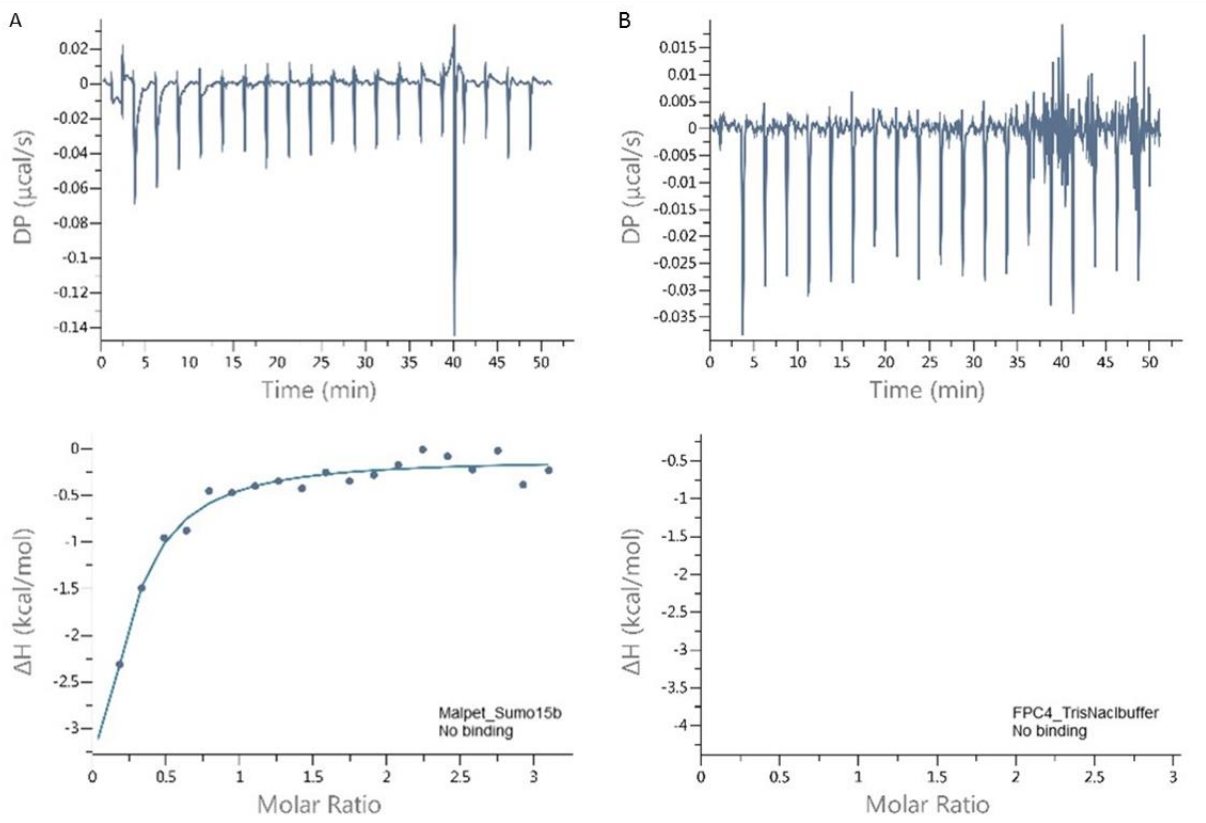


Figure 21 (A) The MBP-tag was tested against the Sumo-tag to verify that no unspecific binding had occured. No binding was detected. (B) The FPC4-5/1-MalpET ligand was titrated against the ITC buffer. No binding was detected.

TbBILBO1 NTD is known to bind FPC4-B1BD (aa 357-440) with a strong binding affinity (K_D≈5 µM) (Albisetti *et al.*, 2017). Binding of TbBILBO1 NTD to the FPC4-5/1 construct was also measured via ITC. The interaction of TbBILBO1 NTD with the short ligand was described to be stronger compared to the interaction with full-length B1BD (K_D≈1 µM) (Dong, 2017). Therefore, it was not purified and measured again.

TbBL232-1/3-Sumo15b and Tb271-1/3-MalpET bound with a high affinity to the short FPC4-5/1 fragment in the binding experiment with a dissociation constant of K_D≈3,0 µM and K_D≈5,5 µM respectively (Fig.22).

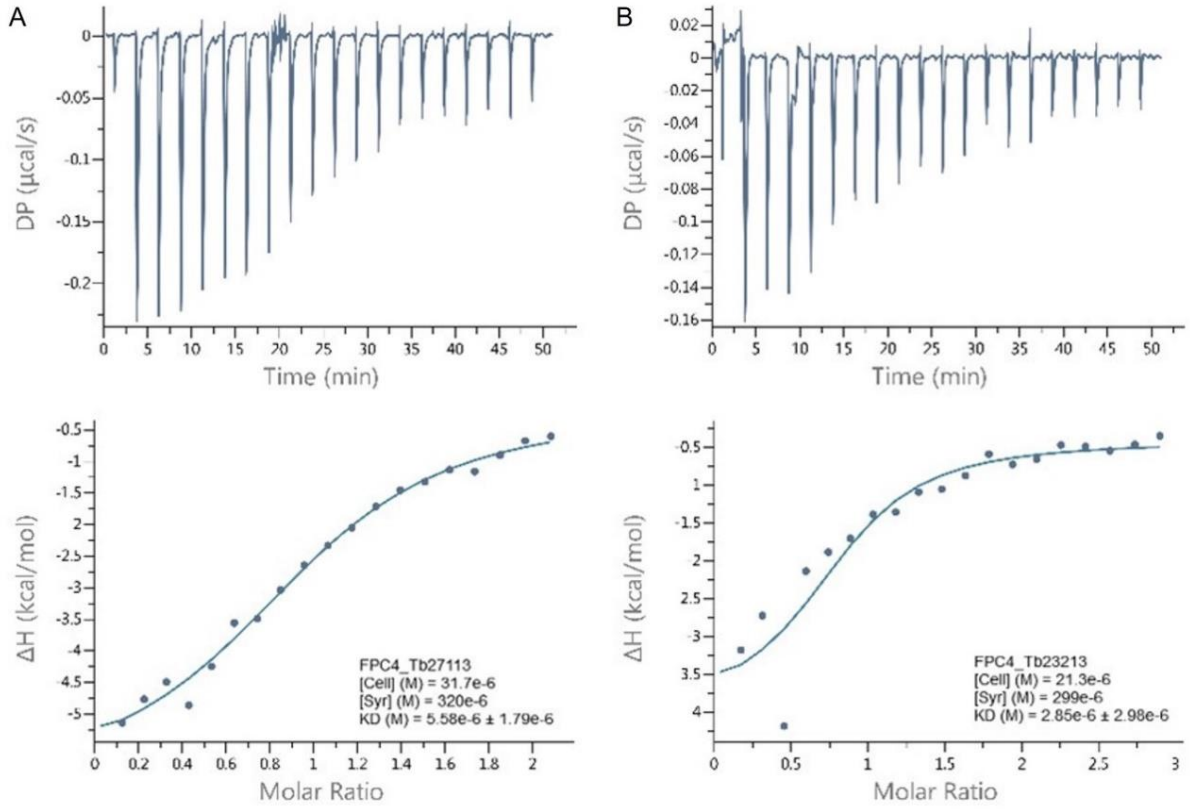


Figure 22 (A) ITC measurement of full length TbBL271 NTD and FPC4-5/1. Protein binding was detected K_D: 5.58 ± 1.79 µM. (B) Full-length TbBL232 NTD was titrated against FPC4-5/1. Protein interaction was detected K_D: 2.85 ± 2.98 µM.

However, removing the loop region of the NTD of BILBO1 and its two homologs seems to completely abolish ligand binding. None of the loop-less construct interacted with FPC4 (Fig.23).

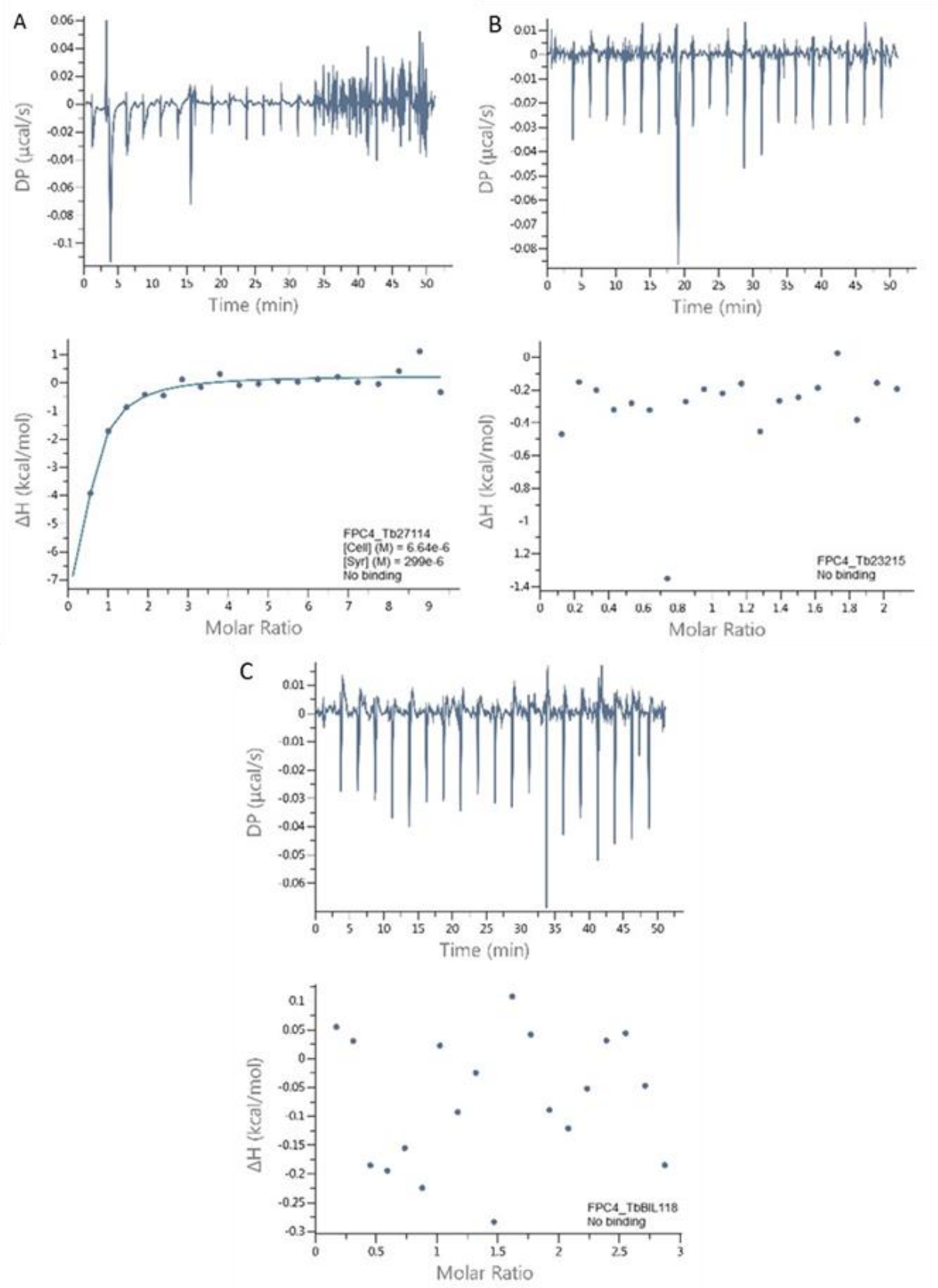


Figure 23 ITC measurement of the loopless TbBL271-1/4 (A), TbBL232-1/5 (B) and TbBILBO1-1/8 (C) constructs. None of the truncated proteins were able to interact with FPC4.

Discussion

TbBIL232 NTD was purified for crystallization experiments in its ligand bound and non-ligand bound native state to study its structure on a molecular level. In the purification process the construct remained stable, while forming protein crystals has proven to be difficult. For this reason, optimization and rescue techniques were applied.

Relying on the sitting-drop-vapor-diffusion method with the Hampton Research Crystal Screen HT™ screen alone proved to be insufficient for crystal growth. Therefore, different techniques were used to crystallize TbBL232. Optimization of the crystal screen conditions around the original condition where at least nucleation was observed did not bear any protein crystals. Lysine methylation of non-ligand bound TbBIL232 NTD could not rescue the protein of interest in crystallization attempts and combined with additive screening using the HR Additive screen HT resulted in empty salt crystals.

In-situ proteolysis, by adding low concentration of trypsin, has been shown to be the method to successfully produce protein crystals suitable for structure determination. The selective protease trypsin is known to cleave the carboxyl side of basic amino acids like lysine or arginine (Wernimont and Edwards, 2009). The idea was to change the surface properties of the protein complex to facilitate protein packing, crystal forming and growth (Wernimont and Edwards, 2009), but neither the C-terminal region (aa 92-119) nor the ligand were found inside the crystal. However, the cut-off at C-terminus of glutamine at residue 91 is not a typical target of that serin protease. The SEC chromatogram of the TbBL232 NTD and FPC4 co-expression displayed one elution peak, suggesting that both interaction partners eluted together. In the following SDS-PAGE gel, TbBL232-1/3 was found right below the 15 kDA marker band, whereas FPC4 was presumably contributing to the distorted band below the 10 kDA marker. Therefore, the loop region was presumptively lost along with FPC4 due to trypsin digestion.

Recent limited proteolysis experiments on TbBL232-1/3 revealed by adding trypsin resulted in a truncated protein with a molecular weight of approximately 11 kDA. The size of the digested product corresponds with the size of the loopless structure described in this work. The limited proteolysis experiment was performed by the means of a logarithmic dilution of trypsin. 1 µl of 1 mg/ml trypsin were added to 10 µl

of 0.5 mg/ml protein sample (10^{-1}). 1 µl of this dilution was then carried over to the next tube which contains 10 µl of protein (10^{-2}). This dilution step was repeated two more times until a 10^{-4} dilution (0.02 % (w/w) trypsin per mg TbBL232 protein) was achieved, where undigested TbBL232-1/3 proteins were observed. The protease was capable to remove 3 kDA of peptide in the early dilutions. However, adding FPC4 to the sample seems to enhance the resistance of TbBL232-1/3 to trypsin cleavage, since the full-length construct was detectable one dilution earlier (0.2 % (w/w) trypsin per mg TbBL232 protein). This observation indicates, that the unstructured CTD of TbBL232 NTD might undergo conformational changes upon FPC4 binding and adopt a more compact fold (Fig 24) (Qiu, 2019). Considering the results of the limited proteolysis experiments, the concentration of trypsin (0.33 % (w/w) trypsin per mg TbBL232 protein) used for in-situ proteolysis might be excessive and needs to be adjusted for further X-ray crystallization attempts.

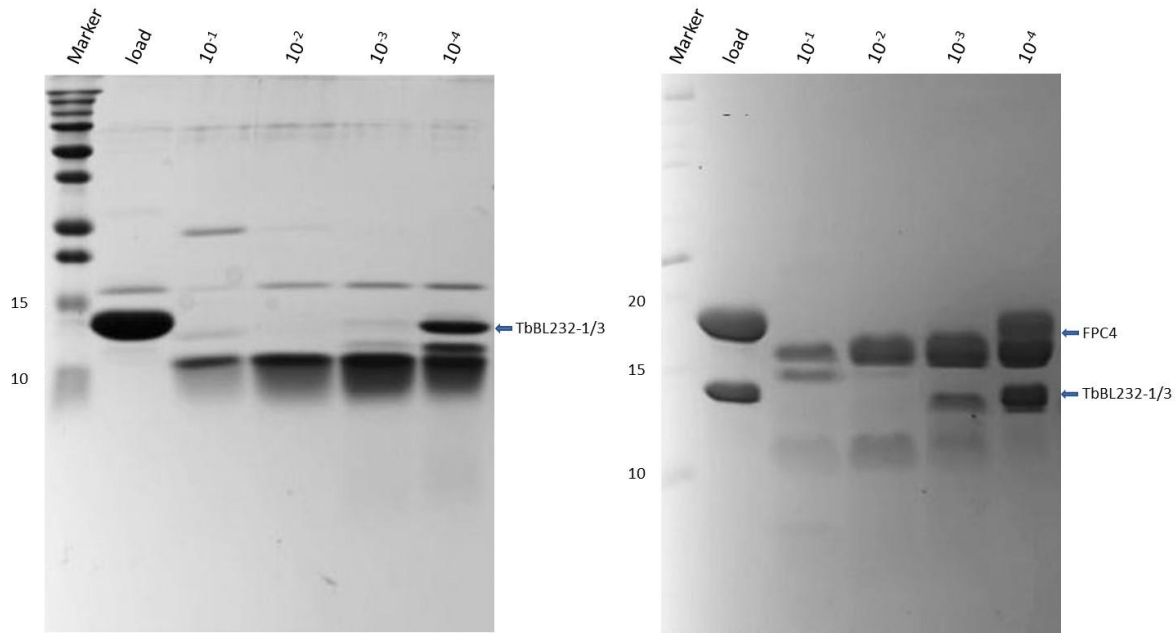


Figure 24 Limited proteolysis of TbBL232-1/3. The 10^{-1} dilution of the logarithmic serial dilution corresponds to an addition of 1 µl trypsin (1mg/ml) to the protein (0.5 mg/ml) sample. Trypsin digestion of the sole TbBL232-1/3 (14 kDA) construct results in an almost complete cleavage of the protein for the dilutions 10^{-1} – 10^{-3} into a 11 kDA sized product and visible degradation. A detectable amount of uncut protein remains in the 10^{-4} dilution. (left) FPC4 (16 kDA) was added to the second attempt of the experiment. TbBL232-1/3 can be detected in the 10^{-3} already. (right) Figure is modified from (Qiu, 2019).

ITC was the method of choice, to gain a better understanding of the importance of this dynamic structure in TbBL232 NTD and its homologs TbBILBO1 NTD and TbBIL271 NTD. To answer the question how ligand binding is affected by eliminating the loop region, the missing loop constructs were synthetized. The MalpET protein tag was not cut off for ITC measurement since the missing loop TbBIL271 NTD

construct was not stable in solution and precipitated. Therefore, the protein tags were kept for all samples. All three loopless constructs lost the capability to bind FPC4, implying a crucial role of this dynamic region in protein-ligand interaction.

The C-terminus of TbBILBO1 NTD consists of a loop region (aa 96-115) which regulates its interaction with other FPC proteins (Vidilaseris *et al.*, 2014). Recent studies have shown that the loop helps to generate a horseshoe-shaped pocket, which provides the binding site for FPC4. Approximately half of the hydrophobic pocket is made of the loop region, whereas ligand binding is mainly enabled by contacting three aromatic residues W71, Y87 and F89 in the core structure. Mutating either W71 or Y87/F89 completely abolished binding of TbBILBO1 NTD and FPC4 (Vidilaseris *et al.*, 2020). Although no specific residue of the loop region is known yet to be involved in ligand binding, the loopless construct of TbBILBO1 did not bind FPC4.

Previous studies on TbBl271 NTD revealed that the loop is involved in ligand binding. The X-ray crystallography structure unveiled a loop structure which constitutes to a considerable part of the hydrophobic pocket like the one in TbBILBO1. Taking a closer look at the binding site of Tb271 NTD on a molecular level, approximately half of the interactions with the short FPC4 (VLTSPVP) ligand were accomplished by amino acids provided by the loop region (aa 91-111) through hydrophobic interactions or hydrophobic bonds. The protein-ligand interaction involved the contact of the three highly conserved aromatic residues W70, Y86 and F88 as well as another aromatic residue F63. On the other side, mainly hydrophobic interactions with aliphatic residues of the loop were observed plus a noticeable water mediated interaction with an aspartic acid on position 94 (Fig.25) (Majneri, 2018). Therefore, this finding would correspond to the results of the ITC measurement and supports the assumption of the loop facilitating ligand binding.

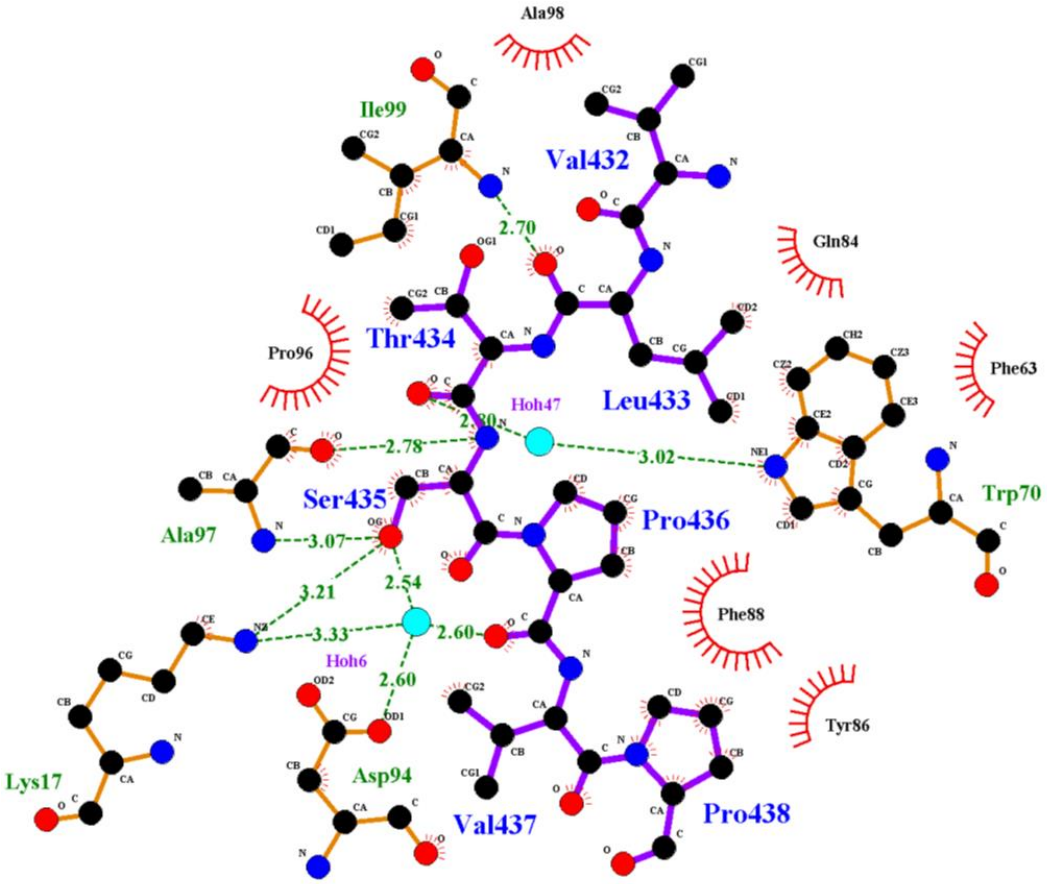


Figure 25 Visualization of the FPC4 binding site of TbBL271 NTD. The short FPC4 VLTSPVP fragment (purple) mainly interacted with the structured core TbBL271 protein (orange) through four aromatic residues F63, W70, Y86 and F88. The loop mainly interacts with FPC4 through hydrophobic residues as well as a water mediated aspartic acid. Figure is modified from (Majneri, 2018).

TbBL232 NTD presumably binds FPC4 through the four aromatic amino acids as well, since W70, Y86 and F88 are conserved in both proteins. Only the phenylalanine at position 63 in Tb271 NTD was substituted with the structurally related tyrosine at position 63 in TbBL232 NTD, which should serve as an interaction partner in the same manner. In addition to the sequence homology, they adopt a similar tertiary structure as well. Hence, TbBIL232 is also believed to manifest such a loop region, which might facilitate FPC4 interaction (Fig.26).

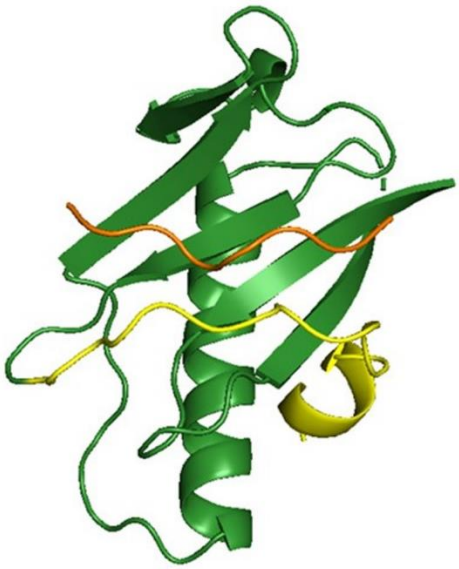
A

```
TbBL232      -MYP--LLVCADLYGEKYNLEVRFPSK-PTIEDLQKRIVTVFSTEMDSRRPEGYPEADFS      56
TbBL271      MMGGISICVATDCDGEKVNLRFLFDAAGPSVSRLLNYSTTAFNNYF---RLKG-ISRAFA      56
              *    : *.:*  *** **.. * :  *::. * :  .*.*.. :   * :*  .  *:

TbBL232      IALVHIYDDKSQQWEKLLSDAQLHEYDQLYVFQPQTRWHMDVQKNLPPPTTPSKDHRTAT      116
TbBL271      VNSAVVFNDVHCTWDRLERTTQLLHNSQVYLFQPDTL---DIPAAIPEPYEGEPLLSA--      111
             :  . :::*    *::*   :** . .*:*:***:*    *:   :* *   .    :

TbBL232      PHS        119
TbBL271      ---        111
```

B



C

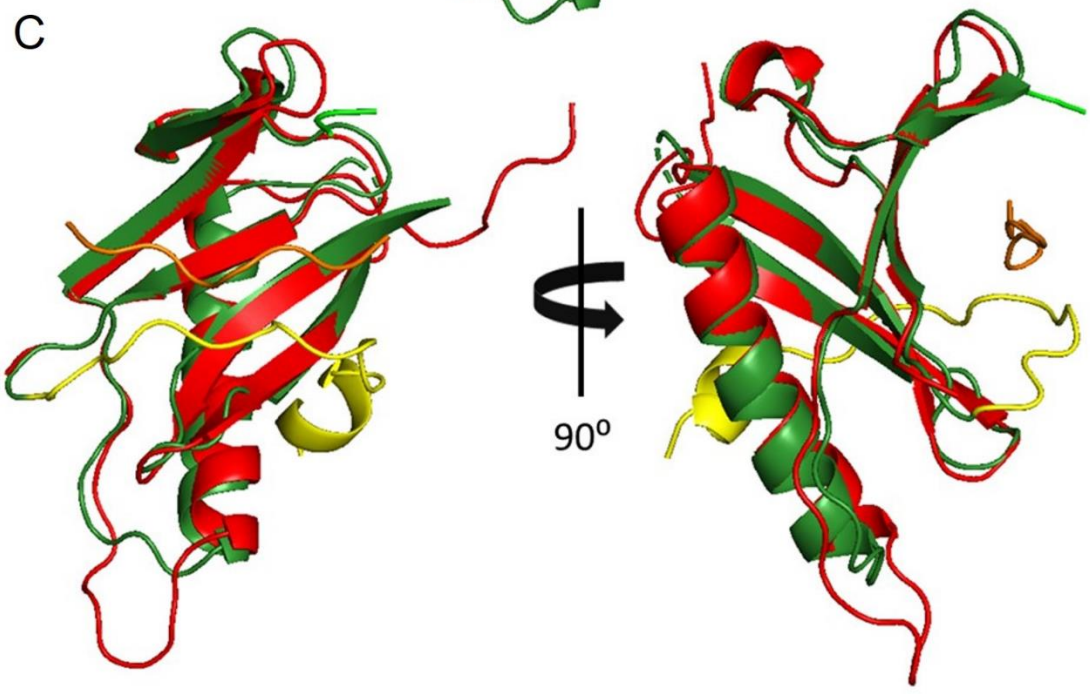


Figure 26 (A) Sequence Alignment of full length TbBL271 and TbBL232 N-terminal domains. Asterisk labelled residues display identical amino acids. Colon labelled positions display amino acid substitutions with similar chemical properties. Period labelled residues display semi-conservative substitutions. These two proteins share a sequence similarity of 28 %. (B) Tertiary structure of full length TbBL271 NTD (green) with FPC4 (orange). The loop is highlighted in yellow. (C) Overlay of TbBL271 (green) and TbBL232 (red). The dynamic loop of TbBL271 which constitutes to a considerable part of the hydrophobic pocket is highlighted in yellow. FPC4 (orange) is located inside the hydrophobic pocket.

Conclusion

The aim of solving the structure of TbBL232 NTD and unveiling the binding mechanism between the target protein and its ligand was not achieved, due to an unforeseen premature truncation of the loop in TbBL232 construct. However, ITC measurement of TbBL232 NTD and FPC4 revealed that the BILBO-like protein does bind FPC4, whereas the loopless construct does not. Comparison of TbBL232 NTD to BILBO1 NTD and TbBL271 NTD showed conserved aromatic residues involved in FPC4 binding were present in TbBL232 as well, which indicates a similar binding mechanism. This implicates a crucial role of this loop in protein binding.

FPC4 acts as a link protein between the FPC and the HC by binding BILBO1 and MORN1 (Albisetti *et al.*, 2017). Unpublished Protein localization studies using fluorescent labelling on TbBL232 suggested that it is localized in the HC. However, the role of TbBL232 in the HC and how it interacts with FPC4 remains unknown.

Further studies on the binding mechanism are necessary to understand which residues of this C-terminal region are involved in the binding of FPC4 B1BD by TbBL232 NTD. Limited proteolysis data suggests, that the NTD of TbBL232 exhibits a dynamic loop, which might change its conformation upon FPC4 binding as well as in-situ proteolysis using lower concentrations of trypsin could facilitate crystallization without cleaving the loop structure for repeated X-ray crystallography experiments.

Appendix

Sequencing results

TbBL271-1/4-MalpET

NGNTTCTTTCGGGCTTTGTTAGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAG
CTTGTCGACGGAGCTCGAATTCGGATCCTTAGTCGGGCTGAAAAAGATACACTTGGCTGTTGTGCAGAAGCT
GTGTGGTGCGTTCTAACCTGTCCCAAGTACAGTGAACATCATTGAAAACAACGGCGGAGTTGACGGCAAATG
CACGAGAGATTCCTTTCAGACGAAAGTAGTTGTTAAAGGCCGTCGTTGAATAGTTCAAAAGGCGTGAAACAG
ATGGCCCTGCTGCATCGAAAAGAAAGCGTAGGTTCACCTTCTCCCCATCACAATCCGTGGCAACACAAATAG
AGATTCCTCCCATCATATGTCCCTGAAAATACAGGTTTTCATGGTGATGGTGATGGTGATGGTGATGGTGCCC
GAGGTTGTTGTTATTGTTATTGTTGTTGTTCGAGCTCGAATTAGTCTGCGCGTCTTTCAGGGCTTCATCG
ACAGTCTGACGACCGCTGGCGGCGTTGATCACCGCAGTACGCACGGCATACCAGAAAGCGGACATCTGCG
GGATGTTCGGCATGATTTCACCTTTCTGGGCGTTTTCCATAGTGGCGGCAATACGTGGATCTTTCGCCAACT
CTTCCTCGTAAGACTTCAGCGCTACGGCACCCAGCGGTTTGTCTTTATTAACCGCTTCCAGACCTTCATCAGT
CAGCAGATAGTTTTCGAGGAACTCTTTTGCCAGCTCTTTGTTCGGACTGGCGGCGTTAATACCTGCGCTCAG

CACGCCAACGAACGGTTTGGATGGTTGACCCTTGAAGGTCGGCAGTACCGTTACACCATAATTCACTTWGCT
GGTGTCGATGTTGGACCATGCCCACGGGCCGTTGATGGTCATCGCTGTTTCGCCTTTATTAAAGGCAGCTTC
TGCGATGGAGTAATCGGGGTCAGCATTCAGGGGATTGTTTTTAATCAGGTCAA

TbBL232-1/5-Sumo15b

CANATCTCTCGAGTGCGGCCGCAAGCTTGTCGACGGAGCTCGAATTCGGATCCTTACTGGGGCTGGAATAC
GTAGAGTTGGTCATATTCATGCAGCTGCGCGTCACTAANAAGTTTCTCCCATTGTTGAGATTTATCGTCGTAA
ATGTGAACTAATGCAATGCTAAAGTCCGCCTCAGGGTACCCCTCTGGTCGCCGGGAGTCCATTTCTGTAGAA
AACACAGNAACTATTCTCTTCTGGAGATCTTCTATGGTCGGTTTTGAAGGGAACCTTACCTCCAAATTGCATTT
CTCCCCGTACAGGTCAGCGCACACGAGGAGTGGGTACATATGGCCACCAATCTGTTCTCTGTGAGCCTCAAT
AATATCGTTATCCTCCATGTCCAAATCTTCAGGGGTCTGGTCAGCTTGAATGCTAATACCGTCGTACAAGAAT
CTTAAGGAGTCCATTTCCTTACCCTGTCTTTTAGCGAACGCTTCCATCAGCCTTCTTAAAGGAGTGGTCTTTTT
GATCTTGAAGAANATTTCTGAAGATNCATCGGACACCTTTAAATTGATGTGAGTCTCAGGNTTGACTTCTGGC
TTGACCTCTGGCTTAGCTTCTTGATTGACTTCTGAGTCCGA

Optimization plate

Table 5 Optimization plate conditions. 1 µl of reservoir liquor was combined with 1 µl of protein sample. The Optimization plate was incubated at 4°C.

	1	2	3	4	5	6	7	8
A	**25µl** 1M HEPES pH 7,5 **(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O	**25µl** 1M HEPES pH 8,5 **(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **93,75µl** 4M (NH4)2SO4 **(1,5M)** **131,25µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **93,75µl** 4M (NH4)2SO4 **(1,5M)** **131,25µl** H2O	**25µl** 1M HEPES pH 8,5 **(0,1M)** **93,75µl** 4M (NH4)2SO4 **(1,5M)** **131,25µl** H2O	**25µl** 1M HEPES pH 7, 5**(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O
B	**25µl** 1M HEPES pH 7,5 **(0,1M)** **187,5µl** 2M LiSO4 **(1,5M)** **37,5µl** H2O	**25µl** 1M Tris pH 7,5 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**25µl** 1M Tris pH 8,0 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**25µl** 1M HEPES pH 8,5 **(0,1M)** **150µl** 50% PEG 1,5K **(30%)** **75µl** H2O	**150µl** 50% PEG 1,5K **(30%)** **100µl** H2O
C	**150µl** 50% PEG 1,5K **(30%)** **100µl** H2O	**150µl** 50% PEG 1,5K **(30%)** **100µl** H2O	**25µl** 1M HEPES pH 7,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 6K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 6K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 6K **(20%)** **100µl** H2O
D	**25µl** 1M HEPES pH 7,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 8K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 8K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 8K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **125µl** 50% PEG 4K **(25%)** **75µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **125µl** 50% PEG 4K **(25%)** **75µl** H2O	**25µl** 1M HEPES pH 8,0 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **125µl** 50% PEG 4K **(25%)** **75µl** H2O	**25µl** 1M HEPES pH 7,0 **(0,1M)** **37,5µl** 100% Isopropanol **(15%)** **100µl** 50% PEG 4K **(20%)** **87,5µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **37,5µl** 100% Isopropanol **(15%)** **100µl** 50% PEG 4K **(20%)** **87,5µl** H2O
E	**25µl** 1M HEPES pH 8,0 **(0,1M)** **37,5µl** 100% Isopropanol **(15%)** **100µl** 50% PEG 4K **(20%)** **87,5µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M HEPES pH 7,5 **(0,1M)** **25µl** 100% Isopropanol **(10%)** **100µl** 50% PEG 4K **(20%)** **100µl** H2O	**25µl** 1M Tris pH 7,5 **(0,1M)** **37,5µl** 100% ETOH **(15%)** **187,5µl** H2O	**25µl** 1M Tris pH 8,0 **(0,1M)** **37,5µl** 100% ETOH **(15%)** **187,5µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **37,5µl** 100% ETOH **(15%)** **187,5µl** H2O	**25µl** 1M Tris pH 7,5 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O
F	**25µl** 1M Tris pH 8,0 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O	**25µl** 1M Tris pH 7,5 **(0,1M)** **62,5µl** 100% ETOH **(25%)** **162,5µl** H2O	**25µl** 1M Tris pH 8,0 **(0,1M)** **62,5µl** 100% ETOH **(25%)** **162,5µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **62,5µl** 100% ETOH **(25%)** **162,5µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O	**25µl** 1M Tris pH 8,5 **(0,1M)** **50µl** 100% ETOH **(20%)** **175µl** H2O

Citations

Afonine, P. V. *et al.* (2012) ‘Towards automated crystallographic structure refinement with phenix.refine’, *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 68(4), pp. 352–367. doi: 10.1107/S0907444912001308.

Albisetti, A. *et al.* (2017) ‘Interaction between the flagellar pocket collar and the hook complex via a novel microtubule-binding protein in Trypanosoma brucei’, *PLoS pathogens*, 13(11), p. e1006710. doi: 10.1371/journal.ppat.1006710.

Baral, T. N. (2010) ‘Immunobiology of African trypanosomes: Need of alternative interventions’, *Journal of Biomedicine and Biotechnology*, 2010(February 2010). doi: 10.1155/2010/389153.

Dong, G. (2017) *ITC data*.

Emsley, P. *et al.* (2010) ‘Features and development of Coot’, *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Field, M. C. and Carrington, M. (2009) ‘The trypanosome flagellar pocket’, *Nature Reviews Microbiology*. Nature Publishing Group, 7(11), pp. 775–786. doi: 10.1038/nrmicro2221.

Franco, J. R. *et al.* (2014) ‘Epidemiology of human African trypanosomiasis’, *Clinical Epidemiology*, 6(1), pp. 257–275. doi: 10.2147/CLEP.S39728.

Gasteiger, E. *et al.* (2005) ‘The Proteomics Protocols Handbook’, *The Proteomics Protocols Handbook*, pp. 571–608. doi: 10.1385/1592598900.

Kaufer, A. *et al.* (2017) ‘The evolution of trypanosomatid taxonomy’, *Parasites and Vectors*. Parasites & Vectors, 10(1), pp. 1–17. doi: 10.1186/s13071-017-2204-7.

Kennedy, P. G. E. (2019) ‘Update on human African trypanosomiasis (sleeping sickness)’, *Journal of Neurology*. Springer Berlin Heidelberg, 266(9), pp. 2334–2337. doi: 10.1007/s00415-019-09425-7.

Langousis, G. and Hill, K. L. (2014) ‘Motility and more : the flagellum of Trypanosoma brucei’, *Nature Publishing Group*. Nature Publishing Group, 12(July). doi: 10.1038/nrmicro3274.

Majneri, P. (2018) *Structural studies of the interaction between two newly identified components, FPC3 and FPC4, of the Trypanosoma brucei flagellar pocket collar*.

Malakhov, M. P. *et al.* (2004) ‘SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins’, *Journal of Structural and Functional Genomics*, 5(1–2), pp. 75–86. doi: 10.1023/B:JSFG.0000029237.70316.52.

Malvy, D. and Chappuis, F. (2011) ‘Sleeping sickness’, *Clinical Microbiology and Infection*. European Society of Clinical Microbiology and Infectious Diseases, 17(7), pp. 986–995. doi: 10.1111/j.1469-0691.2011.03536.x.

Matthews, K. R. (2005) ‘The developmental cell biology of Trypanosoma brucei’, *Journal of Cell Science*, 118(2), pp. 283–290. doi: 10.1242/jcs.01649.

Mccoy, A. J. *et al.* (2007) ‘Phaser crystallographic software’. International Union of Crystallography, pp. 658–674. doi: 10.1107/S0021889807021206.

Perdomo, D., Bonhivers, M. and Robinson, D. (2016) ‘The Trypanosome Flagellar Pocket Collar and Its Ring Forming Protein—TbBILBO1’, *Cells*, 5(1), p. 9. doi: 10.3390/cells5010009.

Qiu, Y. (2019) ‘Expression , purification and limited proteolysis’.

Simpson, A. G. B., Stevens, J. R. and Lukeš, J. (2006) ‘The evolution and diversity of kinetoplastid flagellates’, *Trends in Parasitology*, 22(4), pp. 168–174. doi: 10.1016/j.pt.2006.02.006.

Vidilaseris, K. *et al.* (2014) ‘Structure of the TbBILBO1 Protein N-terminal domain from Trypanosoma brucei reveals an essential requirement for a Conserved surface patch’, *Journal of Biological Chemistry*, 289(6), pp. 3724–3735. doi: 10.1074/jbc.M113.529032.

Vidilaseris, K. *et al.* (2015) ‘Assembly mechanism of Trypanosoma brucei BILBO1 at the flagellar pocket collar’, *Communicative and Integrative Biology*, 8(1), pp. 1–3. doi: 10.4161/19420889.2014.992739.

Vidilaseris, K. *et al.* (2020) ‘Crystal structure of the N-terminal domain of the trypanosome flagellar protein BILBO1 reveals a ubiquitin fold with a long structured loop for protein binding’, *Journal of Biological Chemistry*, 295(6), pp. 1489–1499. doi: 10.1074/jbc.RA119.010768.

Wernimont, A. and Edwards, A. (2009) ‘In Situ Proteolysis to Generate Crystals for Structure Determination : An Update’, 4(4). doi: 10.1371/journal.pone.0005094.