

Towards understanding of unconventional kinetochores in kinetoplastids



Patryk Ludzia

Hertford College
Department of Biochemistry, University of Oxford
Trinity Term 2021

A thesis submitted for the degree of Doctor of Philosophy

Contents

Abstract	4
Declaration	5
Acknowledgments	7
Abbreviations	8
1. Introduction	9
1.1.The centromere	10
1.2.The kinetochore	12
1.3.Kinetoplastids	16
1.4.Kinetochores in kinetoplastids	17
1.5.KKT4 is a microtubule-binding kinetochore protein	21
1.6.KKT23 is a predicted N-acetyltransferase	28
1.7.Aims of this thesis	31
2. Materials and Methods	33
3. Structure and function of KKT4 microtubule-binding domain	80
3.1.KKT4 forms oligomers	81
3.2.NMR analysis of the microtubule-binding domain structure and dynamics	93
3.3.X-ray crystallography of the microtubule-binding domain	109
3.4.KKT4 microtubule-binding activity	120
4. DNA-binding activity of KKT4	129
4.1.KKT4 interacts with DNA with nanomolar affinity	130
4.2.KKT4 has two DNA-binding domains	132
4.3.Mutations in the second DNA-binding domain reduce the affinity to DNA	134
4.4.In vivo analysis of the DNA-binding defective KKT4 mutants	136
5. Structure and function of the KKT4 BRCT domain	138
5.1.Crystal structure of the KKT4 BRCT domain	139
5.2.NMR analysis of the KKT4 BRCT domain	143
5.3.The KKT4 BRCT domain is a phosphopeptide-binding domain	147
5.4.BRCT domain interacts with KKT4 microtubule-binding region	153
6. Structure and function of KKT23	157
6.1. The crystal structure of the KKT23 GNAT domain bound to acetyl coenzyme A	158
6.2.NMR analysis of the KKT23 GNAT domain	168
6.3.KKT23 can acetylate histones in vitro	181
6.4.Mutagenesis of the KKT23 acetyltransferase domain	188
6.5.KKT23 forms a complex with KKT3 and KKT22	192
7. Discussion	195
7.1.KKT4	195
7.2.KKT23	200
Appendix	204
Bibliography	219

Abstract

The kinetochore is the macromolecular protein machinery that drives chromosome segregation by interacting with spindle microtubules. Kinetoplastids, an evolutionarily divergent group of eukaryotes, possess a unique set of kinetochore proteins (KKT1—20, 22—25 and KKIP1—12) that lack significant similarity to known kinetochore components in other eukaryotes.

To date, KKT4 is the only component of kinetoplastid kinetochores shown to have microtubule-binding activities. However, the KKT4 structure and mechanism of action have so far remained unknown. In this thesis, I have shown that KKT4 forms oligomers and I have characterised the structure of KKT4 using X-ray crystallography, NMR spectroscopy and crosslinking mass spectrometry. I have shown that the microtubule-binding domain of KKT4 consists of a coiled-coil region followed by a positively charged disordered tail. Moreover, I demonstrated that the coiled-coil region is sufficient to interact with microtubules and the positively charged unstructured tail enhances the KKT4 affinity for microtubules. Key residues for the microtubule-binding activity have been identified using mutagenesis. I have also shown that KKT4 has DNA-binding activities *in vitro*.

The structure of the KKT4 C-terminal BRCT domain, revealed that the domain is a phosphorylation-dependent protein-protein interaction domain. The KKT4 BRCT domain interacts with a phosphopeptide derived from KKT8 and also makes direct contacts with the N-terminal region of microtubule-binding domain, suggesting possible regulation mechanism.

In the second part of this thesis, the structure and function of KKT23 acetyltransferase is reported. Using X-ray crystallography and NMR spectroscopy, the C-terminal region of KKT23 is shown to have a histone acetyltransferase domain which specifically acetylates histone H2A and H2AZ *in vitro*. The critical residues for the KKT23 catalytic activity are identified. Moreover, the N-terminal fragment of KKT23 for which no structural homolog is known, adopts a helical structure. Finally, KKT23 is shown to form a complex with KKT3 and KKT22.

Patryk Ludzia
Hertford College

Trinity Term 2021

Declaration

This thesis contains ~45 000 words, 231 pages and includes 78 figures and 28 tables. The work in this thesis was undertaken in the Department of Biochemistry at the University of Oxford. Work was completed between October 2017 and June 2021 under the supervision of Dr Bungo Akiyoshi and Prof Christina Redfield. All the work in this thesis is my own unless stated otherwise and has not been submitted for a degree at this or any other university.

NMR experiments were done under close supervision of Prof Christina Redfield. Prof Redfield helped to set up NMR experiments, processed all raw data and helped with the assignments of all spectra shown in this thesis. The spectra of KKT4^{463–645} and KKT23^{125–348} were assigned by Timothy Robinson and Charlotte Nugent as part of their Part II projects, which I co-supervised with Prof Christina Redfield.

Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) and analytical ultracentrifugation (AUC) experiments were done by Dr David Staunton at the Molecular Biophysics Facility at the Department of Biochemistry, University of Oxford.

Crosslinking mass spectrometry performed on KKT4 was done by Svenja Hester and Shabaz Mohammed. Marjorie Fournier performed mass spectrometry on acetylated histones. All mass spectrometry data were obtained at the Advanced Proteomics Facility at the Department of Biochemistry, University of Oxford.

The imaging data presented in this thesis were obtained at the Micron Advanced Imaging Consortium at the Department of Biochemistry, University of Oxford.

Below is the list of publications arising from my DPhil.

Tromer, E.C., Wemyss, T.A., **Ludzia, P.**, Waller, R.F., Akiyoshi, B.

Repurposing of Synaptonemal Complex Proteins for Kinetochores in Kinetoplastida.

Open Biology 11:210049. 210049 (2021). <https://doi.org/10.1098/rsob.210049>

Ludzia, P., Lowe, E., Marciano, G., Mohammed, S., Redfield, C., Akiyoshi, B.

Structural characterisation of KKT4: an unconventional microtubule-binding kinetochore protein. *Structure* (2021). <https://doi.org/10.1016/j.str.2021.04.004>

Ludzia, P., Akiyoshi, B., Redfield, C. ^1H , ^{13}C and ^{15}N resonance assignments for the microtubule-binding domain of the kinetoplastid kinetochore protein KKT4 from *Trypanosoma brucei*. *Biomolecular NMR Assignments* 14, 309–315 (2020). <https://doi.org/10.1007/s12104-020-09968-1>

Nerusheva, O.O., **Ludzia P.**, Akiyoshi, B. Identification of four unconventional kinetoplastid kinetochore proteins KKT22–25 in *Trypanosoma brucei*. *Open Biology* 9: 190236 (2020). <https://doi.org/10.1098/rsob.190236>

Llauró, A., Hayashi, H., Bailey, M., Wilson, A., **Ludzia, P.**, Asbury, C.L., Akiyoshi, B.

The unconventional kinetoplastid kinetochore protein KKT4 tracks with dynamic microtubule tips. *Journal of Cell Biology* 217 (11): 3886–3900 (2018). <https://doi.org/10.1083/jcb.201711181>

Acknowledgments

During my DPhil studies, I was extremely fortunate to be trained by Dr Bungo Akiyoshi and Prof Christina Redfield. Completion of my work would not be possible without their dedication, numerous discussions and overwhelming attitude to help. Working with them has been truly a highlight of my DPhil. I am sure their passion and excellence in science will continue to inspire me long after completing my Oxford journey.

I would also like to express my gratitude to all Akiyoshi lab members for their support and helpful discussions. I want to thank especially Gabriele Marcianò and Krzysztof Kuś for introducing me to the field of structural biology. Midori Ishii has been a helpful adviser in microscopy, while Hanako Hayashi provided me with fascinating discussions about KKT4. I could not wish for a better work environment that these people helped to create.

I am extremely grateful to my undergraduate supervisors, Dr Alicja Sochaj-Gregorczyk and Prof Małgorzata Zakrzewska, for their training and encouragement. I wish to thank Prof Jacek Otlewski for allowing me to join his group from the first year of my studies.

I am grateful for all the advice I got from my Committee members, Prof David Sherratt, Dr Lidia Vasilieva, Prof Matt Higgins and Prof Elspeth Garman.

I want to thank Boehringer Ingelheim Fonds for supporting me during my DPhil.

Finally, I would like to thank my family, especially my parents and grandparents. I am extremely fortunate to have their support.

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
AcCoA	Acetyl Coenzyme A
AUC	Analytical Ultracentrifugation
bp	base pair
BRCT	BRCA1 C-Terminal Domain
BS ³	bis(sulfosuccinimidyl)suberate
BTROSY	Best Transverse Relaxation-Optimised Spectroscopy
CD	Circular Dichroism
CDE	Conserved DNA element
CENP-A	Centromere Protein A
CoA	Coenzyme A
Da	Dalton
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic Acid
FHA	Fork Head Associated Domain
GNAT	Gcn5-related N-acetyltransferase
HetNOE	Heteronuclear NOE
HSQC	Heteronuclear Single Quantum Coherence
IEX	Ion Exchange Chromatography
IPTG	Isopropyl- β -D-thiogalactopyranoside
kb	kilo base
kDa	kilo Dalton
KKIP	KKT-Interacting Protein
KKT	Kinetoplastid Kinetochore Protein
LB	Luria-Bertani
MS	Mass Spectrometry
NHS	N-Hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Data Bank
ppm	parts per million
RDC	Residual Dipolar Coupling
RT	Room Temperature
SDS	Sodium Dodecylsulphate
SEC	Size Exclusion Chromatography
SEC-MALS	Size Exclusion Chromatography Coupled to the Multi-Angle Light Scattering
SeMet	Selenomethionine
TCEP	Tris(2-carboxyethyl)phosphine
TOCSY	Total Correlation Spectroscopy
UV	Ultra Violet light
WT	Wild-Type

CHAPTER 1

1. Introduction

The basic concepts of cell biology introduced to me in my early education fuelled my interest in science. Cells must duplicate to develop and maintain viability. To duplicate, they must replicate their components and equally distribute them to the progeny cells. I was fascinated by how science can explain life in a simple and logical manner. Not long after, however, I found out that these simple concepts are, in fact, very complicated problems that scientists have tried to solve for decades. One of my main interests has been mitosis, the process by which one cell becomes two, controlled by very sophisticated mechanisms. One of these mechanisms, chromosome segregation is driven by the kinetochore complex. Kinetochores are multi-protein machines that drive the separation of sister chromatids by connecting their centromeres to spindle microtubules. This allows for an accurate and equal distribution of the genetic material to progeny cells (McIntosh, 2016).

Molecular studies on kinetochores started from an exciting discovery of anti-centromere autoantibodies (ACA) in patients with scleroderma spectrum disease (CREST syndrome) (Moroi et al., 1980, Tan et al., 1980). These patients' antibodies recognised three antigens, CENP-A (Centromere Protein A), CENP-B and CENP-C, the first kinetochore proteins identified (Earnshaw and Rothfield, 1985). Two terms, kinetochores and centromeres, are widely used in kinetochore biology and need to be distinguished. Centromeres typically refer to a chromosomal locus on which kinetochores assemble and subsequently form stable attachments with spindle microtubules. Kinetochore refer to a protein complex that assembles on the centromeric locus (Cheeseman, 2014).

1.1. The centromere

Although centromeres have a central function in cell division, their structure is highly divergent and can be divided into three categories: point centromeres, regional centromeres and holocentromeres.

Point centromeres (first discovered in *Saccharomyces cerevisiae*) are a simple structure composed of three conserved DNA elements: CDEI—III. All together, they form a single CENP-A containing nucleosome and occupy a 125 bp long sequence, which is bound by the CBF3 complex (Hegemann and Fleig, 1993, Clarke, 1998, Biggins, 2013). The CDE elements are necessary for the formation of functional centromeres and even a point mutation in conserved DNA element III can disrupt the centromere function (Clarke, 1998).

A more complicated example of the centromeric structure can be found in fission yeast, *Schizosaccharomyces pombe*, where centromeres span around 40–100 kb and contain a central core of 4–7 kb, where kinetochores assemble. These centromeres are an example of the second category: regional centromeres that are the most common type of centromeres and found in most model organisms, including humans (Fukagawa and Earnshaw, 2014), *Drosophila melanogaster* (Sun et al., 2003), *Arabidopsis thaliana* (Copenhaver et al., 1995) and *Oryza sativa* (Nagaki et al., 2004). Regional centromeres typically consist of repetitive DNA sequences and involve epigenetic mechanisms to specify the site of kinetochore assembly (Fukagawa and Earnshaw, 2014).

The third type of centromeres can be found in some insects, nematodes and plant species; their centromeres span along whole chromosome arms (Fukagawa and Earnshaw, 2014); hence they are referred to as holocentromeres.

Another non-structural distinction of centromeres can be made, termed neocentromeres. In humans, in physiological conditions, centromeres are present in α -satellite repeat sequences, arrays typically found in natural centromeres (Pardue and Gall, 1970, Fukagawa and Earnshaw, 2014). However, disruption or inactivation of existing centromeres can rarely occur and have been discovered in over 100 human clinical samples (Marshall et al., 2008). In these patients, *de novo* centromeres are formed on a DNA sequence that is not associated with α -satellite arrays. These structures are termed neocentromeres and have become evidence for the hypothesis that specification of human centromeres is not achieved by the sequence but rather by epigenetic mechanisms.

A wide range of studies suggest that kinetochore position is determined by the presence of a histone H3 variant termed CENP-A, which forms specialised nucleosomes at centromeres (Palmer et al., 1991). CENP-A is required for the localisation of all kinetochore proteins in most eukaryotes (Regnier et al., 2005, Liu et al., 2006) and remains closely associated with centromeric chromatin even during replication (Jansen et al., 2007). Moreover, CENP-A is deposited to chromatin where only the pre-existing CENP-A molecules are present. Together with the fact that CENP-A is found at neocentromeres where kinetochores are formed, CENP-A is thought to be a universal epigenetic feature for specification of centromeres and kinetochore assembly in eukaryotes (Fukagawa and Earnshaw, 2014).

1.2. The kinetochore

Kinetochores are hierarchal structures that localise specifically to the centromeric region and interact with spindle microtubules during mitosis (Figure 1.1). Kinetochores can be subdivided into inner and outer kinetochores. The inner kinetochore is responsible for the interaction with DNA and serves as the platform for the outer kinetochore, which makes contacts with dynamic microtubules.

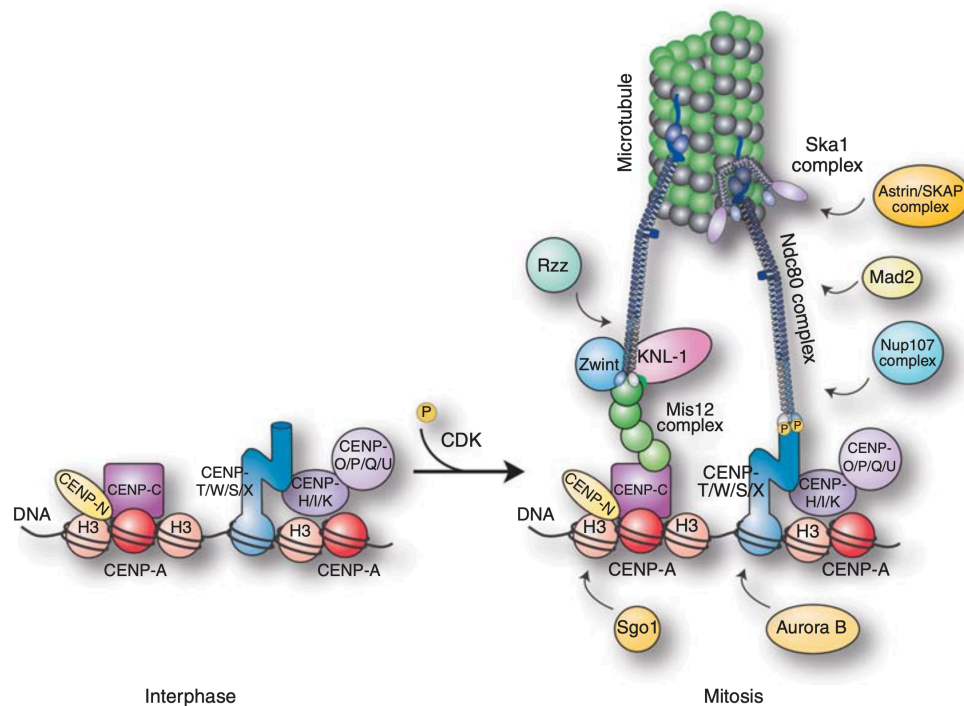


Figure 1.1. Schematic of kinetochore structure during interphase and mitosis.

Kinetochores are enormous protein assemblies composed of over 100 proteins in humans. Numerous mechanisms, including the CDK/Cyclin B-dependent phosphorylation, promote the kinetochore assembly at the onset of mitosis. This figure is adapted from a review by Iain Cheeseman (Cheeseman, 2014). The function and structure of individual proteins shown in the figure are reviewed in (Cheeseman and Desai, 2008) and (Musacchio and Desai, 2017).

Inner kinetochore

The inner kinetochore that constitutes the base of the kinetochore is built by the constitutive centromere associated network (CCAN), which is composed of 16 DNA-proximal kinetochore proteins: CENP-C, CENP-H, CENP-I, CENP-K—U, and CENP-W and CENP-X. The CCAN components can be divided into sub-groups: CENP-C, CENP-L-N, CENP-H-I-K-M, CENP-O-P-U-R and CENP-T-W-S-X complexes. Most of CCAN components constitutively localise at centromeres and require CENP-A for their assembly. The composition of CCAN varies among eukaryotes (van Hooff et al., 2017). For example, in *D. melanogaster* and *C. elegans* most of CCAN components are missing (McKinley and Cheeseman, 2016).

Outer kinetochore

The outer kinetochore components form a platform that allows for the kinetochore complex to form stable attachments with microtubules. The KMN network constitutes a core of the outer kinetochore and is formed by Ndc80, Mis12 and Knl1 complexes (Cheeseman and Desai, 2008, Musacchio and Desai, 2017). The Mis12 complex promotes KMN assembly by interacting with Ndc80 and Knl1 complexes (Petrovic et al., 2010). Its interaction with CENP-C connects the 10-subunit KMN network with the inner kinetochore (Przewloka et al., 2011, Screpanti et al., 2011, Petrovic et al., 2016).

The Ndc80 complex acts as a primary receptor for microtubules at the kinetochore (Cheeseman et al., 2006, Cheeseman and Desai, 2008, Ciferri et al., 2008, Musacchio and Desai, 2017). The complex is composed of 4 subunits (Ndc80, Nuf2, Spc24 and Spc25) that form a rod-like shaped structure with globular domains on both ends (Figure 1.2) (Ciferri et al., 2008). The structure of the Ndc80 complex consists of a long central coiled-coil region, which is preceded by two globular calponin homology (CH) domains in the Ndc80 and Nuf2 subunits. On the other end of the coiled coil, there are RWD domains in Spc24 and Spc25, responsible for interacting with the Mis12 complex and anchoring the Ndc80 complex at the kinetochore. The four subunits of the Ndc80 complex interact with each other through the tetramerisation junction in the middle of the structure (Figure 1.2). Calponin homology domains directly interact with microtubule protofilaments through their positively charged interface (Wei et al., 2007, Ciferri et al., 2008). Another microtubule-binding site is located at the N-terminus of Ndc80. The N-terminal ~80 residue long disordered tail has been shown to contribute to the interaction with microtubules both *in vitro* and *in vivo* (DeLuca et al., 2006, Wei et al.,

2007, Guimaraes et al., 2008, Miller et al., 2008, DeLuca et al., 2011, Valverde et al., 2016). However, the deletion of the disordered tail does not cause severe phenotypic consequences in contrast to the charge-reversal mutations in the CH domains (Kemmler et al., 2009, Cheerambathur et al., 2013, Lampert et al., 2013). *In vitro* studies have shown that the disordered tail can be phosphorylated by Aurora B kinase, which results in the destabilisation of contacts that Ndc80 makes with microtubules (Lampson and Grishchuk, 2017).

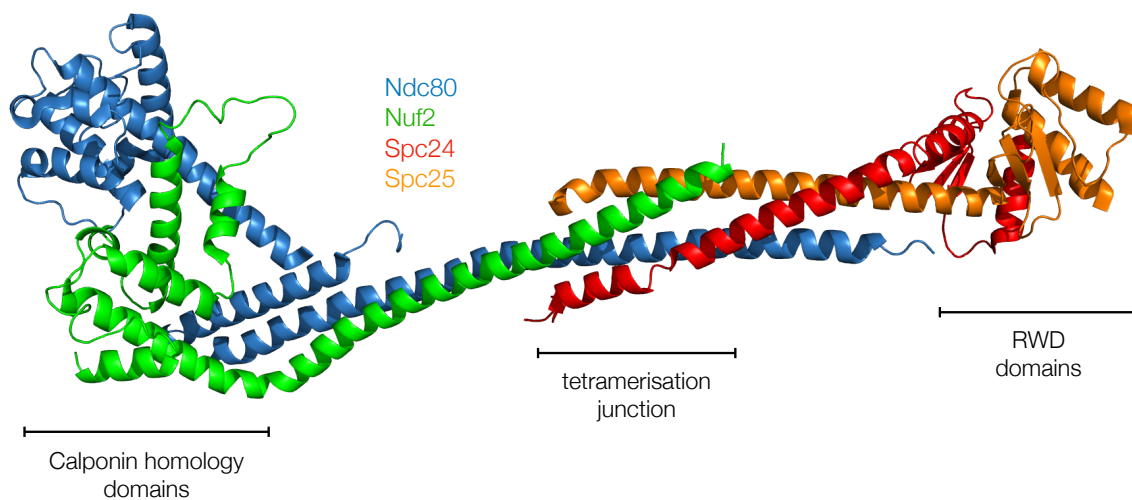


Figure 1.2. Crystal structure of the engineered "Dwarf" Ndc80 complex.

The structure shows a central coiled-coil segment that connects the N-terminal calponin homology domains (Ndc80 and Nuf2) to the C-terminal RWD domains (Spc24 and Spc25). In the centre of the structure, subunits interact through the so-called tetramerisation junction. The cartoon was prepared using coordinates deposited in the Protein Data Bank (PDB: 5TD8). For simplicity, the structure of the nanobody used to assist the crystallisation is not shown in the figure. The structure of the "Dwarf" Ndc80 complex is described in detail in (Valverde et al., 2016).

Mis12 and Knl1 contribute to the microtubule-binding activity of Ndc80 by synergistically increasing the affinity of Ndc80 for microtubules (Cheeseman et al., 2006). It is worth mentioning that although Ndc80 is the primary receptor for

microtubules at the kinetochore, there are other known microtubule-binding proteins, including the Ska and Dam1 complexes, SKAP/Astrin, CENP-E, CENP-F, MCAK, INCENP, XMAP215 and dyneins (Maiato et al., 2004, Foley and Kapoor, 2013, Musacchio and Desai, 2017).

1.3. Kinetoplastids

Kinetoplastids are an evolutionarily divergent group of unicellular flagellated eukaryotes such as free-living *Bodonida* (*Bodo saltans*) and parasitic *Trypanosomatida* (e.g. *Trypanosoma brucei*, *Trypanosoma cruzi* and *Leishmania* species) (Berriman et al., 2005, Cavalier-Smith, 2010), causing many diseases in humans such as Leishmaniasis or Trypanosomiasis and Nagana in livestock. These diseases have been classified as neglected tropical diseases by the World Health Organisation (WHO) and are a causative agent of 10 000–40 000 deaths annually. The name of kinetoplastids originates from their organelle, called the kinetoplast, which is a mitochondrial DNA cluster located at the base of the flagellum (Figure 1.3). The model organism of kinetoplastids, *Trypanosoma brucei*, has a complex digenic life cycle between the alimentary tract/salivary glands of the tsetse fly vector and the bloodstream of a mammalian host (Vickerman, 1965, Fenn and Matthews, 2007). During different stages of the life cycle, the parasite undergoes many morphological and metabolic changes (in tsetse fly as procyclic, epimastigote, metacyclic and as a bloodstream form in the mammalian host). The genome of *Trypanosoma brucei* includes ~9000 identified genes located on 11 diploid pairs of megabase chromosomes, 1–5 intermediate chromosomes and ~100 minichromosomes (Berriman et al., 2005). Interestingly, none of the *T. brucei* genes is orthologous to conventional kinetochore components identified in other eukaryotes.

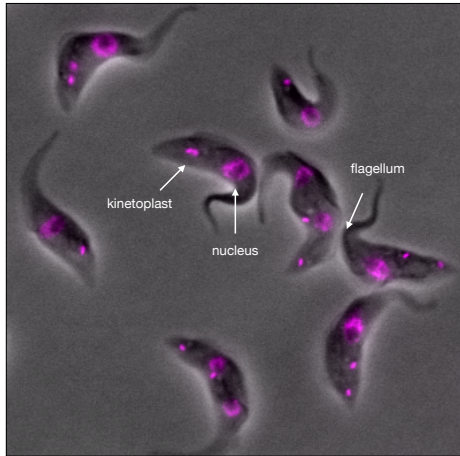


Figure 1.3. Microscopy image of *Trypanosoma brucei* cells.

The figure shows fixed cells of *Trypanosoma brucei* at different stages of the cell cycle, stained with DAPI to visualise DNA (magenta). The position of the DNA-containing organelles (nucleus and kinetoplast) and flagella is labelled and highlighted using white arrows.

1.4. Kinetochores in kinetoplastids

Kinetochore components are widely conserved among eukaryotes, especially the DNA- and microtubule-binding components, such as CENP-A and Ndc80. Their homologues can be found in nearly all sequenced eukaryotes except kinetoplastids. Neither of the CENP-A, Ndc80 or any other known kinetochore component has been found in their genome (Berriman et al., 2005). Instead, a number of unique kinetochore proteins have been discovered in *Trypanosoma brucei*, including 24 kinetoplastid kinetochore proteins (KKT1–20, KKT22–25) and 12 KKT-Interacting proteins (KKIP1–12) (Akiyoshi and Gull, 2014, Nerusheva and Akiyoshi, 2016, D'Archivio and Wickstead, 2017, Brusini et al., 2019, Nerusheva et al., 2019). These proteins are conserved among kinetoplastid species and do not show significant sequence similarity to conventional kinetochore proteins, suggesting that kinetoplastids use a distinct set of proteins to build up the kinetochore complex. All KKT proteins localise at the kinetochore in the procyclic form of *Trypanosoma brucei* cells (Akiyoshi and Gull, 2014). The RNAi-based depletion of most KKT proteins resulted in growth retardation (Akiyoshi and Gull, 2014). Moreover, upon RNAi induction, cells exhibited abnormal DNA content and morphology. Using chromatin

immunoprecipitation, KKT2 and KKT3 have been shown to be enriched in the centromeric region (Akiyoshi and Gull, 2014). All the above observations suggest that the KKT proteins are indeed kinetochore proteins in kinetoplastids.

A combination of localisation pattern and bioinformatic analysis of KKT proteins can facilitate the prediction of their possible function. Studies in conventional eukaryotic model organisms have shown that localisation pattern/recruitment of kinetochore proteins reflect their function (Cheeseman, 2014, Musacchio and Desai, 2017). For example, in humans, many of the inner kinetochore components that localise at the kinetochore constitutively have DNA-binding activities, whereas microtubule-binding components are recruited to the kinetochore in the later phases of mitosis and disassemble at the end of anaphase (Cheeseman, 2014, Musacchio and Desai, 2017). These patterns could possibly reflect the functions of kinetoplastid kinetochore proteins (Figure 1.4). KKT2, KKT3, KKT4, KKT20, KKT22 and KKT23 localise at the kinetochore constitutively; these are therefore candidates for the DNA-binding components of the kinetochore (Akiyoshi and Gull, 2014, Nerusheva and Akiyoshi, 2016, Nerusheva et al., 2019). KKT13 has a unique S-phase specific localisation, thus can be involved in the replication processes (Akiyoshi and Gull, 2014). Proteins that can be detected from the S phase until the end of anaphase (structural role/microtubule-binding activities) are KKT1, KKT5, KKT6, KKT7, KKT16, KKT17, KKT18, KKT24, KKT25, and KKIP1 (Akiyoshi and Gull, 2014). Microtubule-binding candidates could be those proteins that localise from the G2/M phase until the end of anaphase; KKT14 and KKT15. In contrast, KKT8, KKT9, KKT10, KKT11, KKT12 and KKT19 localise at the kinetochore from the S phase until anaphase onset (Akiyoshi and Gull, 2014), suggesting that they may have a regulatory role in the kinetochore function. It has been proposed that proteins that localise until the end of anaphase contribute to the core of the kinetochore complex (KKT1, 2, 3, 4, 5, 6, 7,

14, 15, 16, 17, 18, 20, 22, 23, 24, 25) (Akiyoshi and Gull, 2014). The combination of IP/MS results of individual KKT proteins and their localisation pattern can not only help in predicting the protein function but also help to establish sub-complexes that form functional units. The following sub-complexes for the *T. brucei* kinetochore have been proposed in the original work (Akiyoshi and Gull, 2014): KKT14/15 sub-complex, KKT16/17/18 sub-complex and KKT6/7/8/9/10/11/12/19 sub-complex. Since the discovery of the kinetochore in kinetoplastids, experimental data have confirmed the following complexes: KKT7-10, KKT8-9-11-12 (Ishii and Akiyoshi, 2020) and KKT16-17-18 (Tromer et al., 2021).

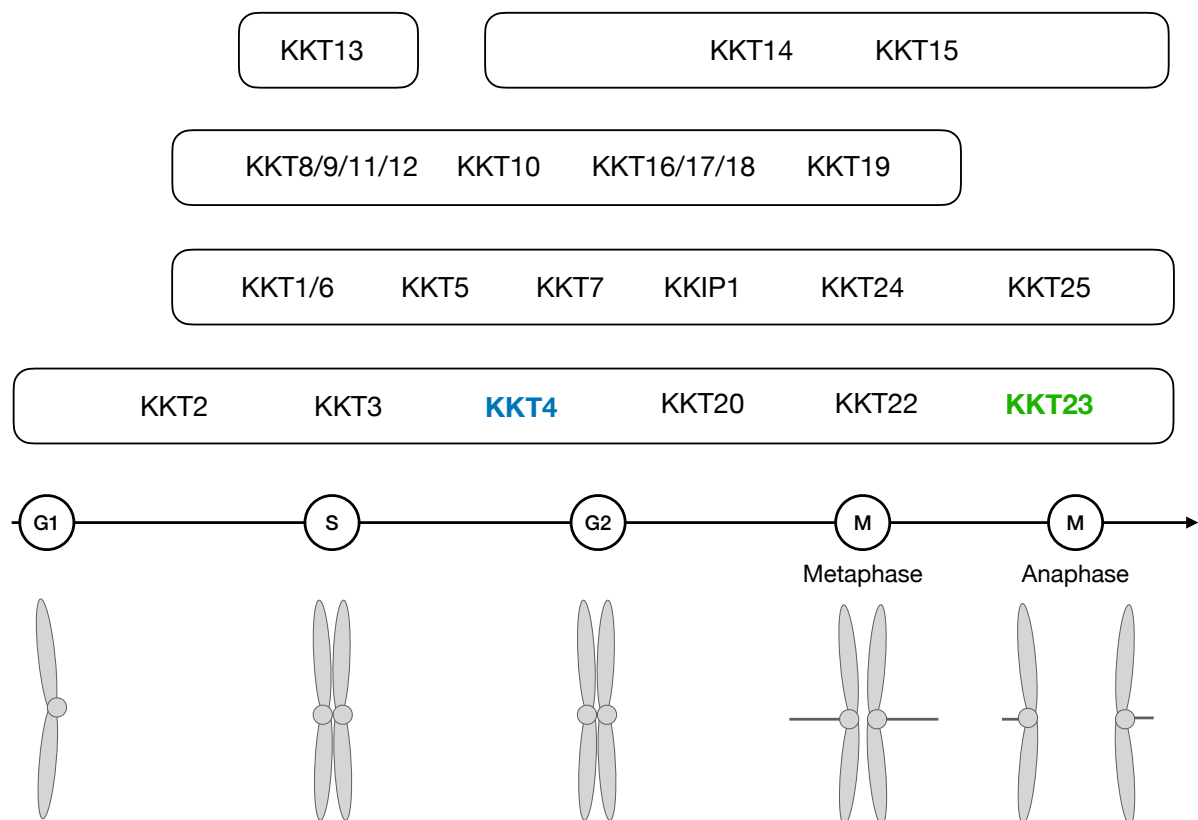


Figure 1.4. Kinetochore proteins in *Trypanosoma brucei* are recruited at different stages of the cell cycle.

The localisation pattern of kinetochore proteins in the procyclic form of *Trypanosoma brucei* is summarised in this figure. KKT4 and KKT23 localise at the kinetochore throughout the cell cycle. For simplicity, the localisation of KKIP2–12 proteins is not shown here but can be found in (Brusini et al., 2019) and (D'Archivio and Wickstead, 2017). The lower panel illustrates the cartoon of chromosome segregation during the cell cycle.

Analysis of KKT primary sequences revealed several known domains and motifs that are conserved in kinetoplastids, including BRCA1 C-terminal (BRCT) domain in KKT4, the Forkhead-associated (FHA) domain in KKT13, the WD40-like domain in KKT15, kinase domains in KKT2, KKT3, KKT10 and KKT19, cysteine-rich and polo box domains (PBD) in KKT2, KKT3 and KKT20, a putative PP1-binding motif in KKT7 and the Gcn5-related N-acetyltransferase (GNAT) domain in KKT23 (Akiyoshi and Gull, 2014, Nerusheva and Akiyoshi, 2016, Nerusheva et al., 2019). Lastly, a homology between the KKT16-17-18 complex and axial element components of the synaptonemal complex (SYCP2-3) have been identified (Tromer et al., 2021). Interestingly, BRCT, FHA and the GNAT domains are not found in any known kinetochore proteins in other eukaryotes (Cheeseman, 2014, Musacchio and Desai, 2017), highlighting the uniqueness of kinetoplastid kinetochore proteins.

In summary, kinetoplastids possess a set of unconventional kinetochore proteins with no significant similarity to known kinetochore components in other eukaryotes. A substantial amount of information about the function and structure of these proteins has been acquired since their discovery. KKT2 and KKT3 have been shown to be important for the localisation of several kinetochore proteins (Marciano et al., 2021). Their divergent polo box domains (DPBD) and central domains are sufficient for kinetochore localisation *in vivo* (Nerusheva and Akiyoshi, 2016, Marciano et al., 2021). Crystal structures of the KKT2 central domain from divergent kinetoplastid species, *Perkinsella* and *B. saltans* show that the domains have a zinc-binding fold; mutations in this region have been shown to abolish its kinetochore localisation and function of KKT2 and KKT3 in *T. brucei* (Marciano et al., 2021). Moreover, CLK-like kinases, KKT10 and KKT19, have been characterised and shown to phosphorylate KKT2, KKT4 and KKT7 (Ishii and Akiyoshi, 2020, Saldivia et al., 2020). Double

knockdown of KKT10 and KKT19 led to a significant delay in the metaphase to anaphase transition (Ishii and Akiyoshi, 2020). Finally, microtubule-binding activity was identified in KKT4, which is described in detail below.

1.5. KKT4 is a microtubule-binding kinetochore protein

One of the most fundamental kinetochore functions is to form stable attachments to spindle microtubules. One of the KIP proteins, the outer kinetochore protein KIP1 has been proposed to be a distantly related homolog of Ndc80/Nuf2 based on weak similarity in their coiled-coil regions (D'Archivio and Wickstead, 2017). However, the analysis of KIP1 primary sequence does not appear to predict a CH domain, which is essential for the microtubule-binding activity of the Ndc80 complex. Moreover, no *in vitro* or *in vivo* evidence of the microtubule-binding activities in KIP1 has been reported (D'Archivio and Wickstead, 2017). It is therefore unclear whether KIP1 is indeed an Ndc80/Nuf2-like protein.

So far only one kinetoplastid kinetochore protein, KKT4, was reported to have microtubule-binding activities (Llauro et al., 2018). Using total internal reflection (TIRF) microscopy, the full-length KKT4 has been shown to bind and diffuse along the microtubules. Moreover, KKT4 can form load bearing attachments with microtubule tips and can track depolymerising microtubules (Llauro et al., 2018). Similar behaviour was observed for conventional microtubule-binding proteins including the Ndc80 (Powers et al., 2009), the Dam1 (Westermann et al., 2006) and the Ska (Schmidt et al., 2012) complexes. Using recombinant fragments of KKT4 and a microtubule co-sedimentation assay, KKT4^{115–343} was shown to contain the microtubule-binding domain (Llauro et al., 2018). The interaction between KKT4 microtubule-binding domain (KKT4^{115–343}) and microtubules seems to be

electrostatic. Charge-reversal mutations of three residues (R123, K132 and R154) have been shown to reduce the affinity of KKT4^{115–343} for microtubules *in vitro*, although the same mutations had no visible effect *in vivo* (Llauro et al., 2018). Within KKT4^{115–343}, a shorter fragment of KKT4 (KKT4^{115–174}) has been reported to be sufficient to co-sediment with microtubules *in vitro*, albeit with a lower affinity compared to KKT4^{115–343}. Deletion of this fragment *in vivo* (KKT4^{Δ115–174}) resulted in a growth defect of trypanosomes, suggesting that residues between K115 and G174, are important for the proliferation of *Trypanosoma brucei*. However, the expression level of the deletion mutant was lower compared to that of the wild-type KKT4, making it difficult to conclude whether the observed growth defect is due to the defect in microtubule-binding or the expression level problem (Llauro et al., 2018).

Interestingly, despite being a microtubule-binding protein, KKT4 was shown to localise at the kinetochore constitutively (Figure 1.5) (Akiyoshi and Gull, 2014). Moreover, KKT4 was shown to localise closer to the putative DNA-binding component, KKT2, rather than the outer kinetochore component, KKIP1 (Llauro et al., 2018). In many eukaryotes, microtubule-binding components localise to the outer kinetochore complex only during mitosis (Cheeseman, 2014). Thus, the constitutive localisation of KKT4 is atypical for a microtubule-binding kinetochore protein. Interestingly, besides kinetochores, KKT4 also localises to the spindle (Llauro et al., 2018) but the function of KKT4 at the spindle is not known.

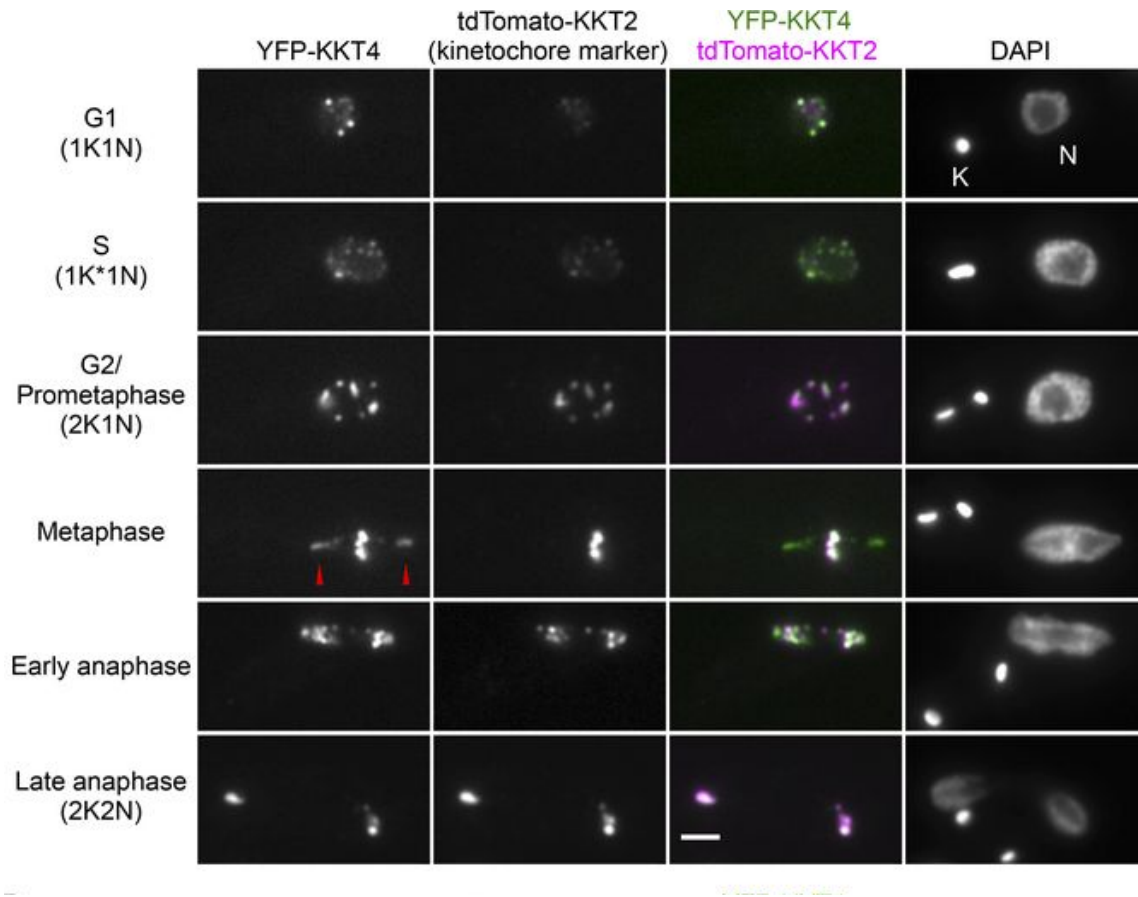


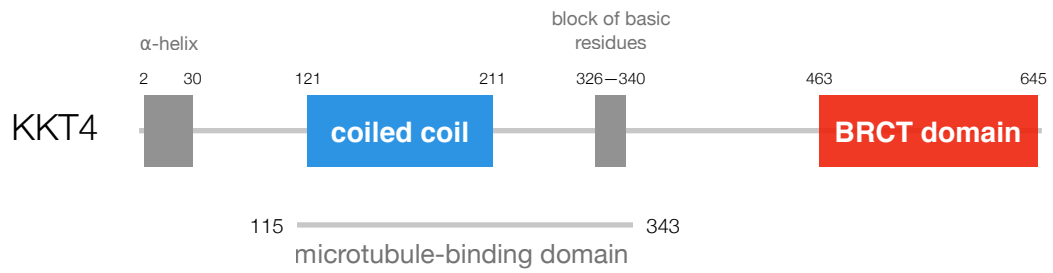
Figure 1.5. KKT4 localises at the kinetochore and spindle microtubules.

The fluorescence microscopy images show the localisation pattern of YFP-KKT4 and TdTomato-KKT2 in *Trypanosoma brucei* cells at different cell cycle stages. In G1, S and G2 phases, these kinetochore proteins form individual dots within the nucleus. During metaphase, kinetochores line up in the centre of the nucleus and at the onset of anaphase, the dots start to separate as chromosomes move away to the opposite poles of the cell. K and N indicate the kinetoplast and nucleus, respectively. Red arrows highlight the spindle localisation of KKT4. DNA was visualised using DAPI staining. The figure is adapted from a previous study (Llauro et al., 2018).

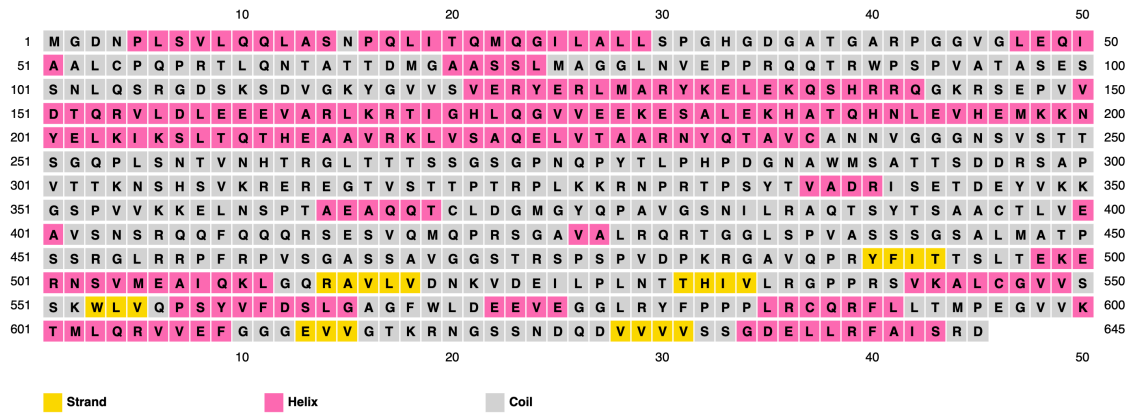
It is unknown how KKT4 activities are regulated or what (if any) are the partner proteins that interact with KKT4. Immunoprecipitation of KKT4 from trypanosomes revealed that KKT4 co-purifies with KKT1, KKT2, KKT7, KKT8, KKT9, KKT10 and KKT12. In addition to these KKT proteins, KKT4 also co-purified with several subunits of the APC/C complex, a widely conserved ubiquitin ligase complex that regulates cell cycle progression in eukaryotes (Akiyoshi and Gull, 2014). Recently, KKT4 has been shown to be closely associated with KKT20 (Nerusheva and Akiyoshi, 2016). These proteins that co-purify with KKT4 are potential regulators of KKT4 functions.

KKT4 consists of 645 amino acid residues with an estimated isoelectric point of ~9.5. KKT4 is predicted to have at least two domains, the N-terminal coiled-coil region and the C-terminal tandem BRCT domain (Figure 1.6A). The coiled coil, according to the COILS server (Lupas et al., 1991), starts at Y124 and ends with Q210. Interestingly, the coiled coil is predicted to have one or more gaps, depending on the window used in the search (Figure 1.6D). The secondary structure prediction reveals a helical regions in the N-terminal region (5—29) and within the microtubule-binding domain (121—236) and a mixture of strands and helices near the C-terminus of KKT4, consistent with the predicted BRCT domain (Figure 1.6B). KKT4 is also predicted to contain three intrinsically disordered regions, 82—114, 290—366 and 402—454 (Figure 1.6C). The following regions of KKT4 are conserved among kinetoplastids: the N-terminal predicted α -helix, predicted coiled coils, a block of basic residues, and the C-terminal BRCT domain (Figure 1.7).

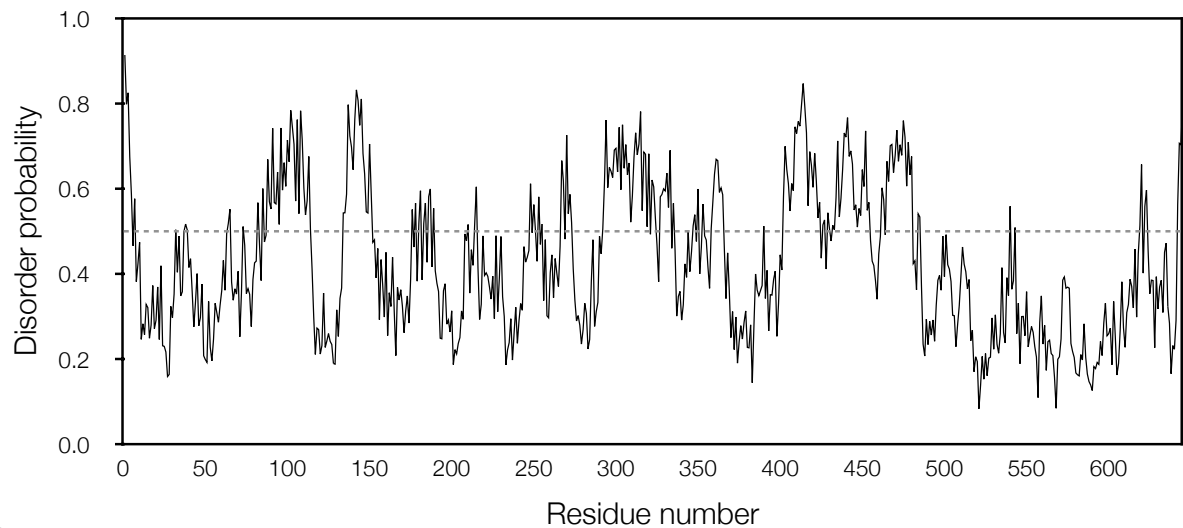
A



B



C



D

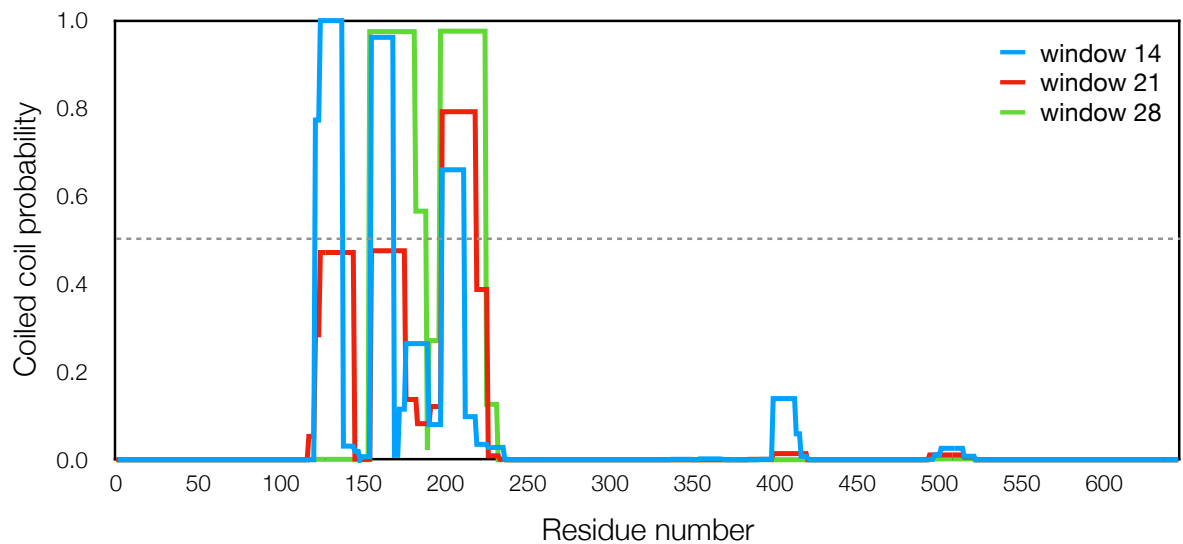


Figure 1.6. Analysis of the KKT4 sequence.

- (A) Cartoon showing the domain organisation of KKT4 (Llauro et al., 2018). The main subjects of my structural studies, the coiled-coil region and the BRCT domain are highlighted in blue and red, respectively. The position of the microtubule-binding domain is indicated as a grey line.
- (B) The secondary structure of KKT4 was analysed using PSIPRED (McGuffin et al., 2000). Each box represents an amino acid labelled with one letter code and coloured according to the predicted secondary structure (yellow for β -strand and pink for α -helix). The N-terminal region of KKT4 is mainly helical followed by ~250 disordered residues. The C-terminal region has mixed α/β topology, consistent with the BRCT domain prediction.
- (C) The disorder probability is plotted against the KKT4 sequence. The prediction was made using the DisEMBL server (Linding et al., 2003). The dashed line represents the threshold (0.5) between order and disorder.
- (D) The position of coiled coils in KKT4 was predicted using COILS (Lupas et al., 1991). The different window sizes refer to different frames of the sequence used for the search. The coiled coil is predicted between V121 and T211. This region is located within the microtubule-binding domain. Interestingly, the coiled coils in KKT4 are predicted to be disrupted with short loop insertions.

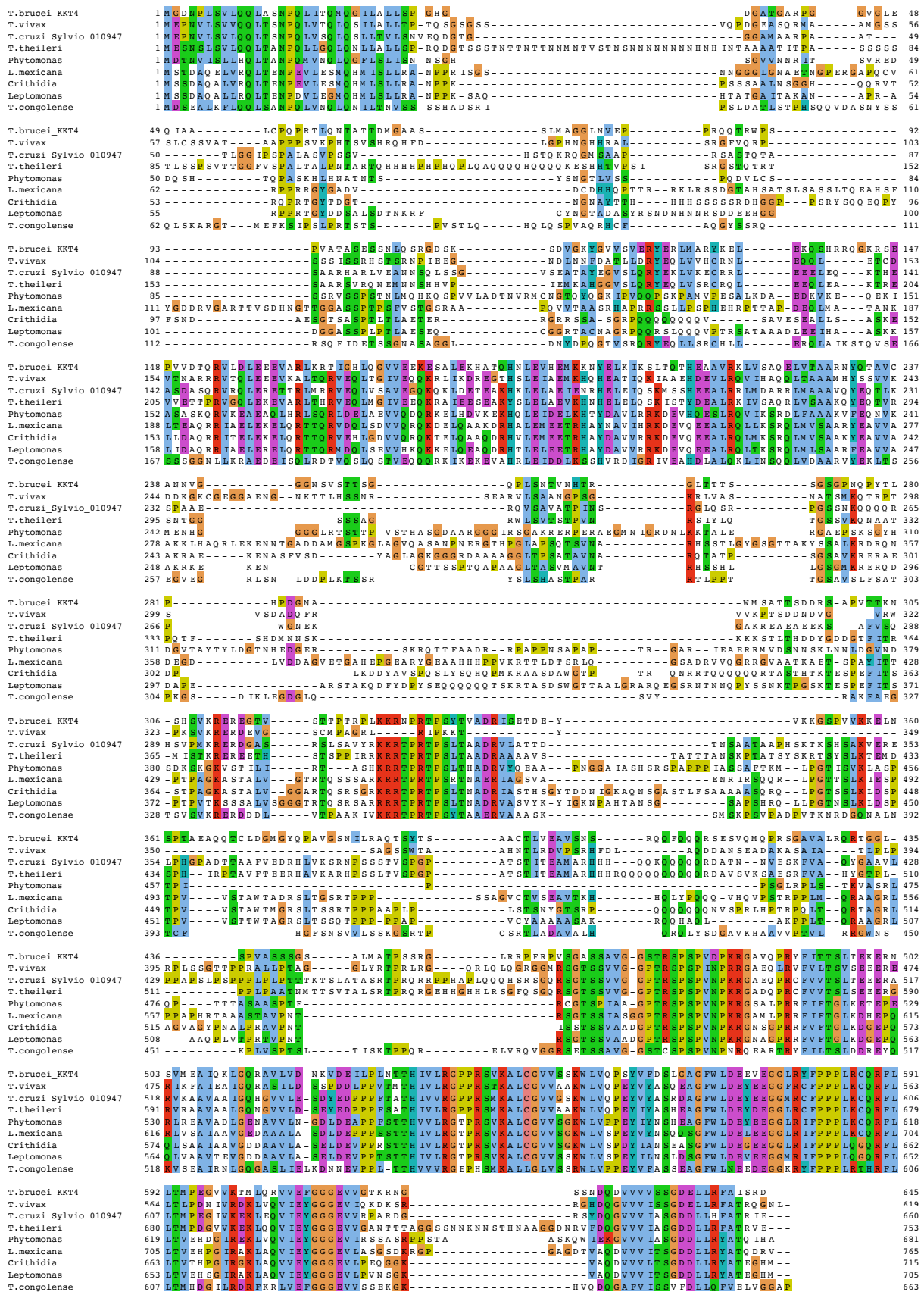


Figure 1.7. Multiple sequence alignment of KKT4 from different kinetoplastid species.

Following regions in KKT4 are conserved among kinetoplastids: the N-terminal helices, the predicted coiled-coil region, the block of basic residues and the C-terminal BRCT domain (please refer to Figure 1.6).

1.6. KKT23 is a predicted N-acetyltransferase

Immunoprecipitation of KKT3 from *Trypanosoma brucei* led to the discovery of another kinetochore protein, KKT23 (Nerusheva et al., 2019). Fluorescence microscopy of YFP-KKT23 showed that the protein forms dots that co-localise with the kinetochore marker, KKT2. Moreover, KKT23 localises at the kinetochore throughout the cell cycle and its immunoprecipitation from trypanosome cells revealed that KKT23 co-purified with many known kinetochore proteins (Nerusheva et al., 2019). These results suggest that KKT23 is a constitutive component of the kinetochore. The RNAi of KKT23 results in a growth defect (Marciano et al., 2021) and severe chromosome mis-segregation in *Trypanosoma brucei* (data not shown), showing that KKT23 is essential for cell viability .

KKT23 consists of 348 amino acid residues (pI 6.36). Its secondary structure analysis predicts helices within the N-terminal region (3–66) and the mixture of α -helices and β -stands at the C-terminal part of the protein (Figure 1.8B), most of which are conserved among kinetoplastids (Figure 1.9). The profile-profile comparison using HHpred (Zimmermann et al., 2018) revealed a Gcn5-like N-acetyltransferase (GNAT) domain (Figure 1.8A). Interestingly, N-acetyltransferase domains have not been identified in any known structural component of kinetochores in other eukaryotes (Cheeseman, 2014, Musacchio and Desai, 2017). The human acetyltransferases, including TIP60 and KAT7, have been shown to regulate kinetochore functions but are not structural components of the kinetochore complex (Mo et al., 2016, Ohzeki et al., 2016). The function of KKT23 at the kinetoplastid kinetochore remains unknown.

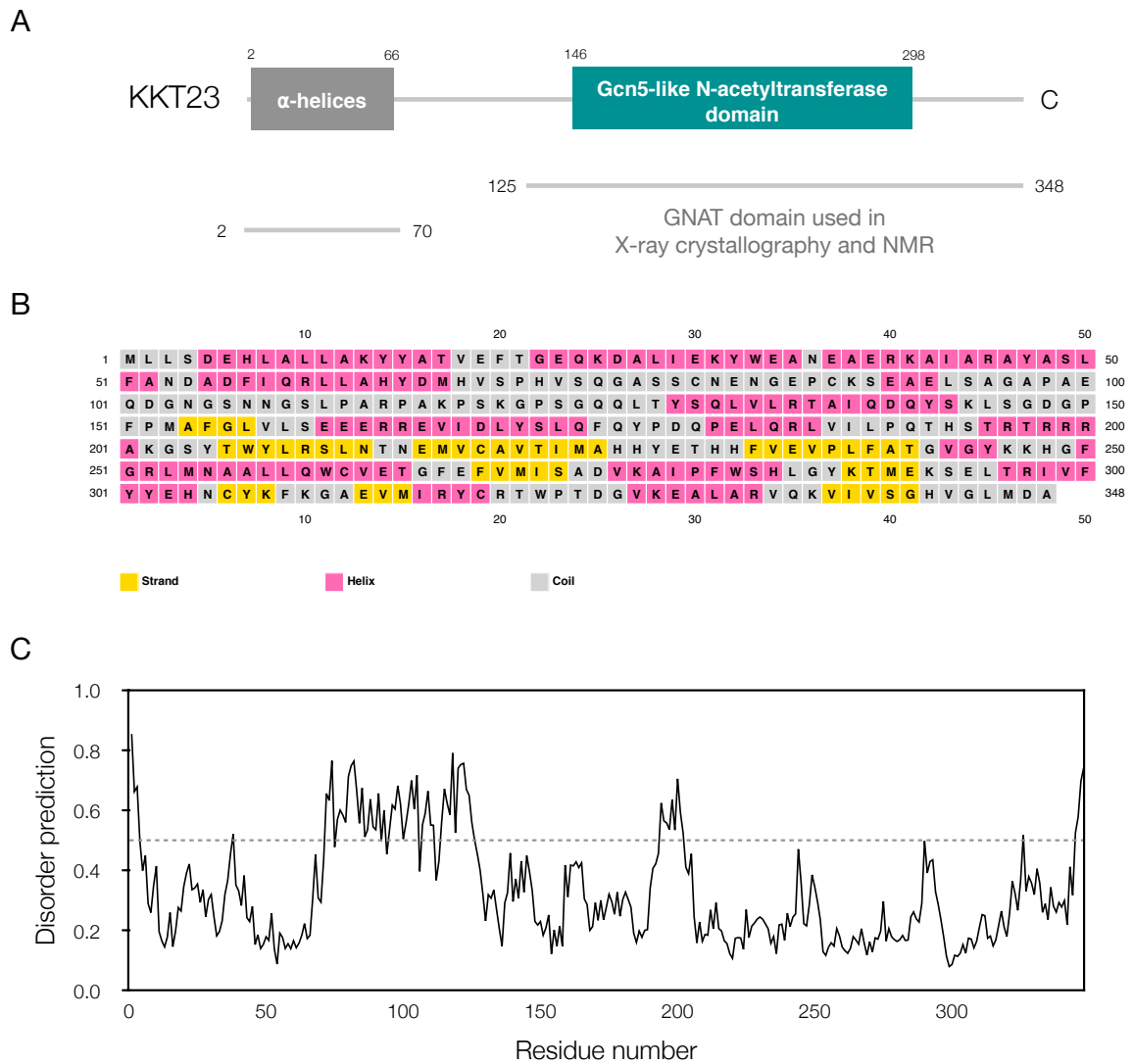


Figure 1.8. Analysis of the KKT23 sequence.

- (A) Cartoon showing the domain organisation in KKT23. The N-terminal region is predicted to be mostly helical (B), while the C-terminal region is predicted to have a Gcn5-related N-acetyltransferase domain (Nerusheva et al., 2019). Below the cartoon, KKT23 fragments that were used to study the function and structure of KKT23 are highlighted with grey lines.
- (B) The secondary structure of KKT23 was analysed using PSIPRED (McGuffin et al., 2000). Each box represents an amino acid labelled with one letter code and coloured according to the predicted secondary structure (yellow for β -strand and pink for α -helix).
- (C) The disorder probability is plotted against the KKT23 sequence. The prediction was made using the DisEMBL server (Linding et al., 2003). The dashed line represents the threshold (0.5) between order and disorder. KKT23 is mainly ordered excluding the regions between 72–126 and 194–202 that are have disorder probability values >0.5 . This is consistent with the secondary structure prediction shown in (B).

1.7. Aims of this thesis

Discovery of the unconventional kinetochore proteins in kinetoplastids was an important discovery in the field of kinetochore biology. The fact that KKT proteins lack significant similarity to other known eukaryotic kinetochore systems raises many questions and exciting possibilities. How do kinetoplastid kinetochores perform conserved kinetochore functions? How are they regulated? What are the functions and structures of individual KKT proteins? How do kinetoplastid kinetochores maintain attachments to microtubules or DNA? Studies on kinetochores in kinetoplastids have the potential to uncover new mechanisms of chromosome segregation in eukaryotes and identify new drug targets to fight against diseases caused by kinetoplastid parasites.

In the first part of my thesis, I aimed to understand how kinetoplastid kinetochores make attachments to microtubules by characterising KKT4, the only known kinetochore protein that has microtubule-binding activities.

1. A previous study (Llauro et al., 2018) revealed that KKT4 has a tendency to form oligomers even at low nanomolar concentrations; however, this has never been directly tested. Using a combination of biophysical techniques, I show that KKT4 can form tetramers in solution and identify regions that drive oligomerisation.
2. The structure of the microtubule-binding domain of KKT4 remained unknown because it has no significant similarity to other known microtubule-binding domains. Using X-ray crystallography and NMR spectroscopy I show that KKT4 microtubule-binding domain consists of a coiled-coil region followed by a positively charged disordered tail. I demonstrate that the coiled coil is sufficient to interact with microtubules *in vitro* and the disordered tail increases the affinity

of KKT4 for microtubules. Finally I identify residues critical for the microtubule-binding activity *in vitro*.

3. KKT4 is a constitutive kinetochore protein and was previously reported to co-purify with DNA during the purification from insect cells (Llauro et al., 2018). I show that KKT4 can tightly interact with DNA and determine its DNA-binding domains. I also report residues important for that activity *in vitro* and test the importance of these residues in *Trypanosoma brucei* cells.
4. The C-terminal region of KKT4 is predicted to adopt the BRCT domain fold. Using X-ray crystallography and NMR spectroscopy I show that indeed KKT4 has the tandem BRCT domains. I demonstrate that the BRCT domain interacts with a phosphopeptide derived from KKT8 and also makes direct contacts with the N-terminal fragment of the microtubule-binding domain of KKT4. The BRCT domain is likely to be a phosphorylation-dependant protein-protein interaction domain.

The second part of my thesis involves the characterisation of KKT23, a predicted N-acetyltransferase.

1. To understand the function of KKT23 I determined the structure of KKT23 C-terminal region and show that it is structurally similar to histone acetyltransferases. A complementary structural study was carried out using solution-state NMR. I also provide preliminary data on the KKT23 N-terminus structure.
2. To identify the substrates of KKT23, I established an *in vitro* assay to screen potential targets of KKT23 acetylation activity. I have shown that KKT23 can acetylate histone H2A and H2AZ *in vitro* and then identified residues in KKT23 critical for this activity.

CHAPTER 2

2. Materials and Methods

This chapter provides a detailed description of all materials and methods used to facilitate research during my doctoral studies. It is organised to follow the sequential steps from cloning, protein expression to structural and functional characterisation.

2.1. Analysis of KKT sequences

The DNA and protein sequences for KKT proteins were retrieved from the TriTryp database (Aslett et al., 2010), UniProt (UniProt, 2019) or a published study (Butenko et al., 2020). These sequences were analysed using PSIPRED (McGuffin et al., 2000) for the secondary structure prediction, DisEMBL (Linding et al., 2003) for the disorder prediction and COILS (Lupas et al., 1991) for the presence of coiled coils. Searches for the homologous proteins were done using BLAST in the TriTryp database (Aslett et al., 2010) or Jackhmmer on the UniProtKB proteome database

using a default setting (HMMER web version 2.24) (Potter et al., 2018). The multiple sequence alignments were performed with MAFFT (L-INS-i method, version 7) (Kato et al., 2019) and visualised with the Clustalx colouring scheme in Jalview (version 2.10) (Waterhouse et al., 2009). The sequence analysis mentioned above was used to facilitate the rational design of N- and C-termini boundaries of protein fragments used throughout this thesis.

2.2. DNA Handling and Cloning

A variety of cloning methods were employed in this work to obtain recombinant DNA fragments that express KKT proteins in bacterial or insect cell expression systems. Oligonucleotides were designed for the use of ligation independent cloning (LIC), NEBuilder Assembly (New England Biolabs) or classical restriction cloning. KKT fragments were amplified using the Polymerase Chain Reaction (PCR). The purified PCR products were hybridised with linearised vectors, amplified in chemically competent *E. coli* NEB5 α cells, purified and sequence verified before being used for heterologous protein expression. Further, a number of mutants were designed to analyse the functional role of specific amino acids or sequence regions.

Oligonucleotide Primer Design

Domain boundaries of protein fragments were designed based on the analysis described in section 2.1 (Analysis of KKT sequences). All fragments were amplified from the genomic DNA of *Trypanosoma brucei* or *Leishmania mexicana*. KKT4 fragments from other kinetoplastid species than *T. brucei* and *L. mexicana* were cloned using synthesised genes (Thermo Fisher Scientific). Coding sequence of the oligonucleotides were around 20 bp in length and GC content ~60%. The oligonucleotides used in this thesis are summarised in Table 2.1

Table 2.1. Oligonucleotides used throughout this work*

Primer	Description	Sequence
BA686	Forward primer for 3FLAG-KKT2	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGTTCAATGTCTCACCAGC GA
BA687	Reverse primer for 3FLAG-KKT2	GCTAGC TCATTTCGGTAGACGTTGC
BA688	Forward primer for 3FLAG-KKT3	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGATTGGCACAGTAGAACA A
BA689	Reverse primer for 3FLAG-KKT3	GCTAGC TTACACGCCACCGTAGAGCA
BA775	Forward primer for 6HIS-H3/H2A/ H2B/H4	GATCGATC TCCGGA AATAATTTTGTCTTAAGGAGATATACAT ATGGCCACTCCTAAGAGCAC
BA776	Reverse primer for 6HIS-H3/H2A/ H2B/H4	ACTGACTG ACGCGT GGATCT TGTACATTA GCTGGAAGCGTGTACACAG
BA778	Forward primer for 6HIS-H2A (LIC)	TACTTCCAATCCATG GCAACACCCAAGCAGGCAGT
BA779	Reverse primer for 6HIS-H2A (LIC)	TATCCACCTTTACTGTCA GACGCTTGGCGTCGCCTTCG
BA916	Forward primer for 6HIS-H2AZ	TACTTCCAATCCATG TCTCTACAGGTGATGATGC
BA917	Reverse primer for 6HIS-H2AZ	TATCCACCTTTACTGTCA CGAGCCCCTCTTCGATTCT
BA923	Forward primer for H2AZ to make pBA562	GATCGATCAAGCTTCGCCAGGGTTTTCCAGTCGAGCTCAATAATTTGT TTAACTTTAAGAAGGAGATATACATATGTCTCTTACAGGTGATGA
BA924	Reverse primer for H2AZ to make pBA562	CAGTCAGTGGTACCGGATCT TGTACATTACGAGCCCCTCTTCGATTCT
BA964	Forward primer for H2BV to make pBA562	GATCGATCTCCGGAATAATTTGTTAACTTTAAGAAGGAGATATACAT GCCGCCAACTAAAGGAGG
BA965	Reverse primer for H2BV to make pBA562	CAGTCAGTCTCGAGGGATCTTGTACATTACGTACCTTGGAGTACGCCT
BA989	Forward primer for KKT4 ^{2–114}	TACTTCCAATCCATG GGAGACAATCCACTGTCTGT
BA992	Reverse primer for KKT4 ^{115–645}	TACTTCCAATCCATG GTGAGCGGTGCCTCCTCTGC
BA1266	Forward primer for KKT4 ^{250–645}	TACTTCCAATCCATGACCTCAGGGCAACCGCTTTC
BA1268	Reverse primer for KKT4 ^{115–343}	TATCCACCTTTACTGTCAATCGGAAATCCTGTGGCTAC
BA1344	Forward primer for KKT1	GATCGATC CCCGGG ATGGATTCTTCGCTTCGTAC
BA1348	Forward primer for 3FLAG-KKT6	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGCATTCCATTCTACCATAC A
BA1349	Reverse primer for 3FLAG-KKT6	ATGCATGC GCTAGC TTACCTAGCTCCCAAGACAG
BA1350	Reverse primer for KKT1	GCTAGC CTAAACGATTTCACGAAT
BA1355	Forward primer for 3FLAG-KKT4	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGGGAGACAATCCACTGTCTG TG
BA1356	Reverse primer for 3FLAG-KKT4	ATGCATGC GCTAGC TTAATCACGACTTATAGCGA

Primer	Description	Sequence
BA1356	Reverse primer for KKT4	ATGCATGC GCTAGC TTAATCACGACTTATAGCGA
BA1374	Reverse primer for KKT3	CCCGGG ATGATTGGCACAGTAGAACAAA
BA1375	Forward primer for KKT4	CCCGGG ATGGGAGACAATCCACTGTCTG
BA1464	Forward primer for KKT4 ^{70–645}	TACTTCCAATCCATG GCAGCATCTTCTCTCATGGC
BA1465	Reverse primer for KKT4 ^{70–645}	TATCCACCTTTACTGTCAGGATTTTATCTTTAATTCATAGTTTTTTTCATCT CATGAACCTCTAAGTTATGTTGAGTTGC
BA1466	Reverse primer for KKT4 ^{115–246}	TATCCACCTTTACTGTCAGGAATTACCCCCGCCAACAT
BA1471	Reverse primer for KKT4 ^{115–323}	TATCCACCTTTACTGTCA TCTGGTGGGAGTCGTGGAAA
BA1472	Forward primer for KKT4 ^{115–645}	TACTTCCAATCCATGAAATATG GCGTCGTTTCTGTTGAGCG
BA1550	Reverse primer for KKT4 ^{115–174}	TATCCACCTTTACTGTCACCCCCTGTAGATGTCCTATC
BA1721	Reverse primer for KKT4 ^{2–114}	TACTTCCAATCCATGCCTGGACATGGTGATGGTGCCACTGG
BA192	Forward primer for KKT3	GATCGATC TCTAGA GGAGCAGGT ACAACTCCCCATCCACGCC
BA1958	Forward primer for KKT4 ^{101–352}	GTATTTTCAGGGATCCTCCAACCTGCAGTCCCGTGG
BA1959	Forward primer for KKT4 ^{63–352}	GTATTTTCAGGGATCCACCGCTACTACCGATATGGG
BA1960	Reverse primer for KKT4 ^{63–352}	GTATTTTCAGGGATCCACCGCTACTACCGATATGGG
BA1988	Forward primer for KKT4 ^{145–232}	TACTTCCAATCCATGAGGAGTGAACCCGTCGTCGA
BA1990	Forward primer for KKT4 ^{138–234}	TACTTCCAATCCATGTCACACCGGAGACAAGGGAA
BA1991	Reverse primer for KKT4 ^{145–230}	TATCCACCTTTACTGTCACCTAGCGGCAGTTACCAACT
BA1992	Reverse primer for KKT4 ^{145–232}	TATCCACCTTTACTGTCAGTAGTTCCTAGCGGCAGTTA
BA1993	Reverse primer for KKT4 ^{138–234}	TATCCACCTTTACTGTCACGTCTGGTAGTTCCTAGCGG
BA1998	Forward primer for <i>L. mexicana</i> KKT4 ^{146–486}	TACTTCCAATCCATGGCAGCGCCTCAGGTGGTCAC
BA2003	Reverse primer for <i>L. mexicana</i> KKT4 ^{146–486}	TATCCACCTTTACTGTCACGACGTTGTCCCAGGCAGGC
BA2024	Forward primer for FLAG-KKT7	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGAGCGGATATGGCGAATCT A
BA2028	Reverse primer for FLAG-KKT7	GCTAGCTCACC GCGTTCTCCCCCAA
BA2037	Forward primer for KKT4 ^{463–645}	TACTTCCAATCCATGAGCGGTGCCTCCTCTGCTGT
BA2079	Forward primer for KKT4 ^{119–232}	TACTTCCAATCCATGGTTTCTGTTGAGCGTTATGA
BA2095	Forward primer for KKT4 ^{114–232}	TACTTCCAATCCATGGGGAAATATGGCGTCGTTTC
BA2137	Forward primer for <i>L. mexicana</i> KKT4 ^{184–284}	GTATTTTCAGGGATCCACCGCCAACAAGCTGACGGA
BA2144	Reverse primer for <i>L. mexicana</i> KKT4 ^{184–284} (no stop codon)	CGCCGAGCTCGAATTCCTGAGCGTGAAGCTTCTTCG

Primer	Description	Sequence
BA2178	Forward primer for <i>L. mexicana</i> KKT4 ^{146–262}	TACTTCCAATCCATGCCGCGCCGCTCCTCGCTTCT
BA2181	Reverse primer for <i>L. mexicana</i> KKT4 ^{146–262}	TATCCACCTTTACTGTCAGGATTTGAGCAGCTGGCGCA
BA2229	Forward primer for KKT23 ^{2–348}	TACTTCCAATCCATGTTACTTAGTGATGAGCACCT
BA2230	Reverse primer for KKT23 ^{2–348}	TATCCACCTTTACTGTCAGGCATCCATTAAACCCACAT
BA2291	Reverse primer for <i>L. mexicana</i> KKT4 ^{184–284}	CGCCGAGCTCGAATTCCTACTGAGCGTGAAGCTTCTTCG
BA2292	Forward primer for <i>L. mexicana</i> KKT4 ^{159–359}	GTATTTTCAGGGATCCCCGCGCCGCTCCTCGCTTCT
BA2317	Reverse primer for <i>L. mexicana</i> KKT4 ^{159–359}	CGCCGAGCTCGAATTCCTAGCTGCTGTACTTGGCTGTCG
BA2411	Forward primer for <i>T. cruzi</i> sylvio KKT4 ^{117–233}	GTATTTTCAGGGATCCTCTCTTCAGCGCTATGAGAA
BA2412	Forward primer for <i>T. cruzi</i> sylvio KKT4 ^{117–233}	CGCCGAGCTCGAATTCCTAGGGGCTCTTAGTGTCTGCT
BA2436	Forward primer for KKT23 ^{128–348}	GTATTTTCAGGGATCCACGTACTCTCAATTGGTTCT
BA2437	Forward primer for KKT23 ^{150–348}	GTATTTTCAGGGATCCCCCTTCCGATGGCCTTCGG
BA2438	Reverse primer for KKT23 ^{94–348}	CGCCGAGCTCGAATTCCTAGGCATCCATTAAACCCACAT
BA2445	Forward primer for KKT4 ^{111–232}	GTATTTTCAGGGATCCAGCGATGTAGGGAAATATGG
BA2447	Reverse primer for KKT4 ^{111–232}	CGCCGAGCTCGAATTCCTAGTAGTTCCTAGCGGCAGTTA
BA2453	Forward primer for <i>T. congolense</i> KKT4 ^{134–244}	GTATTTTCAGGGATCCCCGAGGGTACAGTTAGCCG
BA2482	Forward primer for <i>T. vivax</i> KKT4 ^{117–240}	GTATTTTCAGGGATCCCCGATTGAAGAAGGTAACGA
BA2484	Reverse primer for <i>T. vivax</i> KKT4 ^{117–240}	CGCCGAGCTCGAATTCCTAGCTGCTATAATGCATTGCTG
BA2485	Forward primer for <i>L. seymouri</i> KKT4 ^{117–240}	GTATTTTCAGGGATCCACCGCATGTAATGCAGGTCTG
BA2487	Reverse primer for <i>L. seymouri</i> KKT4 ^{117–240}	CGCCGAGCTCGAATTCCTACGGACTGCTGGTTGTACCGC
BA2491	Forward primer for 3FLAG-KKT22	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGGTCGGATCCTCCTTCTC
BA2492	Forward primer for KKT22	CCCGGGATGGTCGGATCCTCCTTCTC
BA2493	Reverse primer for KKT22	GCTAGCTTAGGTCACCTTCTTGGGAAG
BA2581	Forward primer for KKT7 ^{144–512}	ACCTGTATTTTCAGGGATCCCCACGCTCACGTTCTCAAAG
BA2584	Reverse primer for KKT7 ^{144–512}	GCGCCGAGCTCGAATTCCTATGGTCGTCGCACAGTGGCGG
BA2587	Forward primer for 3FLAG-KKT23	CCCGGGATGGATTACAAGGATGACGACGATAAGGATTACAAGGATGACG ACGATAAGGATTACAAGGATGACGACGATAAGTGTCTGTCTGACGAGCA
BA2588	Forward primer for KKT23	CCCGGGATGCTGCTGTCTGACGAGCA
BA2589	Reverse primer for KKT23	GCTAGCTTAAGCGTCCATCAGGCCCA

Primer	Description	Sequence
BA2696	Forward primer for KKT4 ^{123–224}	ACCTGTATTTTCAGGGATCCCGTTATGAGCGGCTCATGGC
BA2697	Forward primer for KKT4 ^{126–224}	ACCTGTATTTTCAGGGATCCCGGCTCATGGCTCGCTACAAAG
BA2700	Reverse primer for KKT4 ^{123–224}	GCGCCGAGCTCGAATTCCTACTCCTGTGCACTTACCAAC
BA2771	Forward primer for <i>T. congolense</i> KKT4 ^{145–244}	ACCTGTATTTTCAGGGATCCCAGCTGCTGAGCCGTTGTCA
BA2772	Forward primer for <i>L. mexicana</i> KKT4 ^{168–268}	ACCTGTATTTTCAGGGATCCCCGAGCCCGCACGAGCATCG
BA2774	Forward primer for <i>T. cruzi sylvio</i> KKT4 ^{117–233}	ACCTGTATTTTCAGGGATCCGCAATGGTTCCGAAAGCGC
BA2776	Reverse primer for <i>T. congolense</i> KKT4 ^{145–244}	GCGCCGAGCTCGAATTCCTACTGCTGGCTATTAATCAGTTTC
BA2776	Reverse primer for <i>T. congolense</i> KKT4 ^{134–244}	GCGCCGAGCTCGAATTCCTACTGCTGGCTATTAATCAGTTTC
BA2777	Reverse primer for <i>L. mexicana</i> KKT4 ^{168–268}	GCGCCGAGCTCGAATTCCTACGACACCATCAGCTGGCGGG
BA2808	Forward primer for KKT4 ^{115–343}	ACCTGTATTTTCAGGGATCCAAATATGGCGTCGTTTCTGTTG
BA2810	Reverse primer for KKT4 ^{233–343}	GCGCCGAGCTCGAATTCCTATTCGAAATCCTGTGCGGTAC
BA2842	Forward primer for KKT23 ^{94–348}	ACCTGTATTTTCAGGGATCCTCGGCAGGCGCTCCCGCTGAACAGGA
BA3108	Forward primer for KKT23 ^{125–348}	ACCTGTATTTTCAGGGATCCCAACAACCTCACGTACTCTCAATTGG
BA3137	Forward primer for KKT23 ^{146–348}	ACCTGTATTTTCAGGGATCCAGTGGTGACGGCCCCCTTTCC
BA3138	Forward primer for KKT23 ^{143–348}	ACCTGTATTTTCAGGGATCCTCCAACTTAGTGGTGACGGC
BA3255	Forward primer for KKT4 ^{233–343}	ACCTGTATTTTCAGGGATCCCAGACGGCTGTTTGTGCCAA
BA3292	Forward primer for KKT23 ^{2–70}	ACCTGTATTTTCAGGGATCCTTACTTAGTGATGAGCACCT
BA3314	Reverse primer for KKT23 ^{2–70}	GCGCCGAGCTCGAATTCCTAGCTAACATGCATATCGTAGTG
BA3340	Forward primer for H2A ^{2–134}	ACCTGTATTTTCAGGGATCCGCAACACCCAAGCAGGCAGTGAAGAAGG CATCG
BA3341	Forward primer for H2A ^{16–134}	ACCTGTATTTTCAGGGATCCAGCAGCCGCTCTGTGAAGGCGGGGT
BA3342	Reverse primer for H2A ^{16–134}	GCGCCGAGCTCGAATTCCTAGACGCTTGCGTCGCCTTCGCGTGCTT
* All KKT proteins mentioned in this table refer to <i>Trypanosoma brucei</i> unless stated otherwise.		

DNA Amplification by Polymerase Chain Reaction (PCR)

DNA fragments encoding KKT truncations were amplified by PCR in 50 µl reactions using polymerase master mix of Q5 High-Fidelity (New England Biolabs) or PrimeSTAR (Takara Clontech) according to the manufacturer's instructions. The PCR reactions contained 0.5 µM of each forward and reverse primer and ~3 ng of template DNA. Mixtures were prepared at RT, using 8-well string PCR tubes. PCR products were obtained after 30 cycles in a Biometra TAdvanced thermal cycler (New Route) using the programme detailed in Table 2.2

Table 2.2. The details of the PCR program used in this work.

PCR Step	Cycles	Time [s]	Temperature [°C]
Initial denaturation	1	30	98
Denaturation	29	10	98
Annealing		15	54
Elongation		45	72
Elongation	1	180	72

PCR reactions were analysed using 1% agarose gel electrophoresis by mixing a sample volume with 6x loading dye (New England Biolabs) in a 5:1 v/v ratio. The electrophoresis was resolved in 1x Tris-Acetate-EDTA buffer and the DNA bands stained with ethidium bromide were visualised using Chemidoc MP Imager (Biorad) with a UV trans-illuminator. PCR products were purified using silica-membrane purification kit QIAquick PCR Purification (QIAGEN) and eluted with milliQ water.

Ligation-independent Cloning

The pNIC28-Bsa4/pBA219 vector was linearised using BsaI in a 50 µl reaction volume at 37°C. After 1 h digestion, the entire reaction was resolved on 1% agarose gel to separate the digested from undigested DNA. The band corresponding to the

linearised vector was cut out and purified using gel extraction kit (QIAGEN). The vector was eluted in 30 µl volume with miliQ water. The purified vector was incubated with 2 µl (10 units) of T4 DNA polymerase (Fermentas), 2 µl of dGTP (2 mM), 10 µl of 5x T4 DNA polymerase buffer (Fermentas) and 6 µl of water. The purified PCR fragment was processed the same way, except that the dCTP was used instead of dGTP. The reactions were incubated for 30 min at 22°C to allow the digestion, followed by a 20 min incubation at 75°C to inactivate the enzyme. This was to generate single stranded overhangs at the end of both the vector and the insert. The hybridisation reaction was prepared by mixing 1.5 µl of insert with 1 µl of the vector and incubated for ~30 min at room temperature. Next, 2 µl of the reaction were transformed into 100 µl of competent *E.coli* NEB5α cells following the standard protocol and plated on appropriate antibiotic selective LB plates. To grow colonies of bacteria, the plates were incubated overnight at 37°C.

NEBuilder DNA Assembly

The backbone of RSF-Duet1 vector was amplified using PCR. To remove the methylated template, the PCR product was digested with 10 units of DpnI, then purified using QIAquick PCR purification kit (QIAGEN) and eluted with miliQ water. The hybridisation reaction was performed by mixing 1 µl (50 ng) of the vector, 2 µl (~200 ng) of purified insert, 5 µl of 2x NEBuilder HiFi DNA Assembly (New England Biolabs) and 3 µl of miliQ water. The reaction was incubated 15–60 min at 50°C. Next, 2 µl of the reaction were transformed into 50 µl of competent *E.coli* NEB5α cells following the standard protocol and plated on the appropriate antibiotic selective LB plate.

Mutagenesis

To generate point mutants of KKT4 and KKT23, a PCR-based site-directed mutagenesis was used. Two overlapping primers with opposite directionality and carrying the appropriate codon change were designed. The mutated codon was followed by ~18 bp and proceeded by 6 bp to allow annealing in the PCR reaction. The PCR reactions contained 0.5 μ M of forward and reverse primer, ~3 ng of the template DNA and PrimeSTAR polymerase (Takara Clontech). The PCR products were obtained after 30 cycles in a Biometra TAdvanced thermal cycler (New Route) using the programme detailed in Table 2.3. The amplicons were incubated with 10 units of DpnI (New England Biolabs) at 37°C for 30 minutes to remove the methylated template DNA. Following the incubation, 4 μ l of the reaction were transformed into 100 μ l of competent *E.coli* NEB5 α cells following the standard protocol and plated on appropriate antibiotic selective LB plate.

Table 2.3. The details of the PCR program used for mutagenesis.

PCR Step	Cycles	Time [s]	Temperature [°C]
Initial denaturation	1	30	98
Denaturation	29	15	98
Annealing		15	55
Elongation		180	72
Elongation	1	180	72

Table 2.4 Oligonucleotides used for mutagenesis*.

Primer	Description	Sequence
BA1600	Forward primer to introduce R123E in KKT4	GGCGTCGTTTCTGTTGAGGAGTATGAGCGGCTCATGGCTCGC
BA1600	Forward primer to introduce R123E in KKT4	GGCGTCGTTTCTGTTGAGGAGTATGAGCGGCTCATGGCTCGC
BA1601	Reverse primer to introduce K132E in KKT4	GGTGTGATTGCTTCTCTAGTTCCTCGTAGCGAGCCATGAGCCGCTCAT
BA1601	Reverse primer to introduce K132E in KKT4	GGTGTGATTGCTTCTCTAGTTCCTCGTAGCGAGCCATGAGCCGCTCAT
BA1602	Forward primer to introduce K132E in KKT4	CGGCTCATGGCTCGCTACGAGGAACTAGAGAAGCAATCACACCGGAG
BA1602	Reverse primer to introduce K132E in KKT4	CGGCTCATGGCTCGCTACGAGGAACTAGAGAAGCAATCACACCGGAG
BA1603	Reverse primer to introduce R154E in KKT4	CCTCTTCAAGATCAAGGACCTCCTGCGTATCGACGACGGGTTCAC
BA1603	Reverse primer to introduce K154E in KKT4	CCTCTTCAAGATCAAGGACCTCCTGCGTATCGACGACGGGTTCAC
BA1604	Forward primer to introduce R154E in KKT4	CCCGTCGTCGATACGCAGGAGGTCCTTGATCTTGAAGAGGAAGTGGCG
BA1604	Forward primer to introduce K154E in KKT4	CCCGTCGTCGATACGCAGGAGGTCCTTGATCTTGAAGAGGAAGTGGCG
BA1632	Reverse primer to introduce R123E in KKT4	GAGCCATGAGCCGCTCATACTCCTCAACAGAAACGACGCCATATTCATGG
BA1632	Reverse primer to introduce R123E in KKT4	GAGCCATGAGCCGCTCATACTCCTCAACAGAAACGACGCCATATTCATGG
BA2205	Forward primer to introduce K310E, R311D, R313D in KKT4	CTCAGTGGAAGACGAGGACGAGGGCACTGTTTCCACGACTCC
BA2206	Reverse primer to introduce K310E, R311D, R313D in KKT4	AGTGCCCTCGTCCTCGTCTTCCACTGAGTGACTGTTTTTAG
BA2207	Forward primer to introduce K326E, K327E, R328D in KKT4	CCCCTCGAGGAGGACAATCCCCGACCCCCAGCTA
BA2208	Reverse primer to introduce K326E, K327E, R328D in KKT4	GGGATTGTCCTCCTCGAGGGGTCTGGTGGGAGTCGTGG
BA2209	Forward primer to introduce R456D, R457D, R460D in KKT4	GGACTCGACGACCCGTTTGACCCAGTGAGCGGTGCCTCCTC
BA2210	Reverse primer to introduce R456D, R457D, R460D in KKT4	CACTGGGTCAAACGGGTCGTGAGTCCACGACTACTTGGAGTTGC
BA2273	Forward primer to introduce V8A, L9A, L12A in KKT4	CTGTCTGCGGCACAGCAAGCCGCATCTAATCCCCAGCTG
BA2274	Reverse primer to introduce V8A, L9A, L12A in KKT4	AGATGCGGCTTGCTGTGCCGACAGTGGATTGTCTCC
BA2279	Forward primer to introduce I19A in KKT4	CAGCTGGCCACTCAGATGCAGGGAATT
BA2280	Reverse primer to introduce I19A in KKT4	CTGAGTGGCCAGCTGGGGATTAGATGC
BA2281	Forward primer to introduce I19A, L26A, L29A in KKT4	GGAATTGCAGCACTTGCGTCACCTGGACATGGTGAT
BA2282	Reverse primer to introduce I19A, L26A, L29A in KKT4	AGGTGACGCAAGTGCTGCAATTCCTGCATCTGAGT

Primer	Description	Sequence
BA2570	Reverse primer to delete 300–488 region in KKT4	TGCGGACCTGTCGTCGCTGG
BA2574	Forward primer to delete 300–488 region in KKT4	CGTTACTTTATTACGACGAGTCTAACA
BA2641	Forward primer to introduce R145D in KKT4	GGAAGGATAGTGAACCCGTCGTCGAT
BA2642	Reverse primer to introduce R145D in KKT4	TTCATATCCTTCCCTTGTCTCCGGTG
BA2643	Forward primer to introduce K166E in KKT4	AGACTTGAACGAACGATAGGACATCTA
BA2644	Reverse primer to introduce K166E in KKT4	CGTTCGTTCAAGTCTCGCCACTTCCTC
BA2645	Forward primer to introduce R167D in KKT4	CTTAAGGATACGATAGGACATCTACAG
BA2646	Reverse primer to introduce R167D in KKT4	TATCGTATCCTTAAGTCTCGCCACTTC
BA2648	Reverse primer to introduce K179E in KKT4	GCTCTCTTCCTCCTCAACAACCCCTG
BA2652	Reverse primer to introduce K199E in KKT4	ATAGTTTTCTTTCATCTCATGAACCTC
BA2653	Forward primer to introduce K204E in KKT4	GAATTAGAAATAAAATCCCTCACACAG
BA2654	Reverse primer to introduce K204E in KKT4	TTTTATTTCTAATTCATAGTTTTTTTT
BA2655	Forward primer to introduce K206E in KKT4	AAGATAGAATCCCTCACACAGACGCAT
BA2656	Reverse primer to introduce K206E in KKT4	GAGGGATTCTATCTTTAATTCATAGTT
BA2657	Forward primer to introduce K218E in KKT4	GTGCGGGAATTGGTAAGTGCACAGGAG
BA2658	Reverse primer to introduce K218E in KKT4	TACCAATTCCCGCACTGCGGCTTCATG
BA2659	Forward primer to introduce R230D in KKT4	GCCGCTGATAACTACCAGACGGCTGTT
BA2660	Reverse primer to introduce R230D in KKT4	GTAGTTATCAGCGGCAGTTACCAACTC
BA3065	Forward primer to introduce R126D in KKT4	TATGAGGATCTCATGGCTCGCTACAAA
BA3066	Reverse primer to introduce R126D in KKT4	CATGAGATCCTCATAACGCTCAACAGA
BA3067	Forward primer to introduce R130D in KKT4	ATGGCTGATTACAAAGAACTAGAGAAG
BA3068	Reverse primer to introduce R130D in KKT4	TTTGTAATCAGCCATGAGCCGCTCATA
BA3069	Forward primer to introduce K136E in KKT4	CTAGAGGAACAATCACACCGGAGACAA
BA3070	Reverse primer to introduce K136E in KKT4	TGATTGTTCTCTAGTTCTTTGTAGCG
BA3071	Forward primer to introduce R140D in KKT4	TCACACGATAGACAAGGGAAGAGGAGT
BA3072	Reverse primer to introduce R140D in KKT4	TTGTCTATCGTGTGATTGCTTCTCTAG
BA3073	Forward primer to introduce R141D in KKT4	CACCGGGATCAAGGGAAGAGGAGTGAA
BA3074	Reverse primer to introduce R141D in KKT4	CCCTTGATCCCGGTGTGATTGCTTCTC
BA3075	Forward primer to introduce K144E in KKT4	CAAGGGGAAAGGAGTGAACCCGTCGTC
BA3076	Reverse primer to introduce K144E in KKT4	ACTCCTTCCCCTTGTCTCCGGTGTGA

Primer	Description	Sequence
BA3084	Forward primer to introduce R164D in KKT4	GTGGCGGATCTTAAGCGAACGATAGGA
BA3085	Reverse primer to introduce R164D in KKT4	CTTAAGATCCGCCACTTCCTCTTCAAGATCA
BA3086	Forward primer to introduce K198E in KKT4	GAGATGGAAAAAACTATGAATTAAAGATAAAATC
BA3087	Reverse primer to introduce K198E in KKT4	GTTTTTTTCCATCTCATGAACCTCTAA
BA3088	Forward primer to introduce R217D in KKT4	GCAGTGGATAAGTTGGTAAGTGCACAGG
BA3089	Reverse primer to introduce R217D in KKT4	CAACTTATCCACTGCGGCTTCATGCGT
BA3090	Forward primer to introduce K115E in KKT4	GGATCCGAATATGGCGTCGTTTCTGTT
BA3091	Reverse primer to introduce K115E in KKT4	GCCATATTCGGATCCCTGAAAATACAGGTT
BA3092	Forward primer to introduce K179E in KKT4	GAGGAGGAAGAGAGCGCATTAGAGAAACACG
BA3094	Forward primer to introduce K199E in KKT4	ATGAAAGAAAACTATGAATTAAAGATAAAATCC
BA3116	Forward primer to introduce K543A in KKT4	AGCGTGGCGGCGTTGTGCGGTGTTGTG
BA3117	Reverse primer to introduce K543A in KKT4	CAACGCCGCCACGCTCCGTGGGGGACC
BA3266	Forward primer to introduce G249A in KKT23	AAGCATGCATTCGGCAGGCTAATGAAC
BA3267	Reverse primer to introduce G249A in KKT23	GCCGAATGCATGCTTCTTATACCCAAC
BA3268	Forward primer to introduce G251A in KKT23	GGCTTCGCAAGGCTAATGAACGCAGCA
BA3269	Forward primer to introduce G251A in KKT23	TAGCCTTGCGAAGCCATGCTTCTTATACCCAACC
BA3270	Forward primer to introduce R252A in KKT23	TTCGGCGCACTAATGAACGCAGCACTG
BA3271	Reverse primer to introduce R252A in KKT23	CATTAGTGCGCCGAAGCCATGCTTCTT
BA3272	Forward primer to introduce H284A in KKT23	TGGTCCGCACTTGGCTATAAACTATGG
BA3273	Reverse primer to introduce H284A in KKT23	GCCAAGTGC GGACCAAAAAGGAATCGC
BA3274	Forward primer to introduce E235A in KKT23	TTCGTGCGAGTGCCGCTGTTGCAACT
BA3275	Reverse primer to introduce E235A in KKT23	CGGCACTGCGACGAAGTGGTGCGTTTC
BA3302	Forward primer to introduce T241A in KKT23	TTCGCAGCGGGGGTTGGGTATAAGAAGCATGGCTTCG
BA3303	Reverse primer to introduce T241A in KKT23	AACCCCCGCTGCGAACAGCGGCACCTC
BA3304	Forward primer to introduce H226A in KKT23	ATGGCAGCGCACTATGAAACGCACCAC
BA3305	Reverse primer to introduce H226A in KKT23	ATAGTGCGCTGCCATAATTGTCACGGC

Primer	Description	Sequence
BA3306	Forward primer to introduce P237A in KKT23	GAGGTGGCGCTGTTGCAACTGGGGTT
BA3307	Reverse primer to introduce P237A in KKT23	GAACAGCGCCACCTCGACGAAGTGGTG
BA3308	Forward primer to introduce E235A, P237A in KKT23	GCGGTGGTGCTGTTGCAACTGGGGTT
BA3309	Reverse primer to introduce E235A, P237A in KKT23	GAACAGCACCAACCGGACGAAGTGGTG
BA3310	Forward primer to introduce W282A in KKT23	CCTTTTGGCTCCCATCTTGGCTATAAACTATGG
BA3311	Reverse primer to introduce W282A in KKT23	ATGGGACGCAAAAGGAATCGCCTTAAC
BA3312	Forward primer to introduce F239A in KKT23	CCGCTGGCGGCAACTGGGGTTGGGTAT
BA3313	Reverse primer to introduce F239A in KKT23	AGTTGCCGCCAGCGGCACCTCGACGAA
BA3344	Forward primer to introduce D275A in KKT23	TCCGCTGCCGTTAAGGCGATTCTTTTTGG
BA3345	Reverse primer to introduce D275A in KKT23	CTTAACGGCAGCGGATATCATCACAACCTCGAACCC
* All KKT proteins mentioned in this table refer to <i>Trypanosoma brucei</i> unless stated otherwise.		

Plasmid DNA amplification

Bacterial colonies were selected and used to inoculate 5 ml of 2x TY media supplemented with an appropriate selective antibiotic. The cultures were incubated overnight at 37°C while shaking at 220 rpm. The plasmid DNA was purified using the QIAprep Spin Miniprep Kit according to the manufacturer's protocol and eluted in Tris-EDTA buffer (QIAGEN). All plasmids were sequenced with Source BioScience (Oxford) using vector or insert specific primers.

Plasmids

Table 2.5 outlines all plasmids used and described in this work.

Table 2.5 List of plasmids used in this study.

Name	Description	Vector	Resistance
pBA435	6HIS-H2A	pNIC28-Bsa4	Kanamycin
pBA443	6HIS-H3/H2A/H2B/H4	pST44	Ampicillin
pBA529	6HIS-H2AZ	pNIC28-Bsa4	Kanamycin
pBA562	6HIS-H3/H2AZ/H2BV/H4	pST44	Ampicillin
pBA985	6HIS-KKT4 ^{70–645}	pNIC28-Bsa4	Kanamycin
pBA987	6HIS-KKT4 ^{115–645}	pNIC28-Bsa4	Kanamycin
pBA1064	6HIS-KKT4 ^{115–323}	pNIC28-Bsa4	Kanamycin
pBA1065	6HIS-KKT4 ^{115–343}	pNIC28-Bsa4	Kanamycin
pBA1066	6HIS-KKT4 ^{250–645}	pNIC28-Bsa4	Kanamycin
pBA1079	6HIS-KKT4 ^{115–246}	pNIC28-Bsa4	Kanamycin
pBA1171	6HIS-KKT4 ^{115–174}	pNIC28-Bsa4	Kanamycin
pBA1328	6HIS-KKT4 ^{115–343} R123E	pNIC28-Bsa4	Kanamycin
pBA1329	6HIS-KKT4 ^{115–343} K132E	pNIC28-Bsa4	Kanamycin
pBA1330	6HIS-KKT4 ^{115–343} K154E	pNIC28-Bsa4	Kanamycin
pBA1391	6HIS-KKT4 ^{63–352}	RSFDuet-1	Kanamycin
pBA1393	6HIS-KKT4 ^{101–352}	RSFDuet-1	Kanamycin
pBA1413	6HIS-KKT4 ^{2–114}	pNIC28-Bsa4	Kanamycin
pBA1436	6HIS-KKT4 ^{138–234}	pNIC28-Bsa4	Kanamycin
pBA1440	6HIS-KKT4 ^{145–230}	pNIC28-Bsa4	Kanamycin
pBA1441	6HIS-KKT4 ^{145–232}	pNIC28-Bsa4	Kanamycin
pBA1513	6HIS-KKT4 ^{463–645}	pNIC28-Bsa4	Kanamycin
pBA1537	6HIS- <i>Lm</i> KKT4 ^{146–486}	pNIC28-Bsa4	Kanamycin
pBA1551	6HIS-KKT4 ^{119–232}	pNIC28-Bsa4	Kanamycin
pBA1572	6HIS-KKT4 ^{114–232}	pNIC28-Bsa4	Kanamycin
pBA1601	6HIS-KKT4 ^{115–232}	RSFDuet-1	Kanamycin
pBA1618	<i>L. mexicana</i> 6HIS-KKT4 ^{184–284} R218Q, + 23 aa	RSFDuet-1	Kanamycin
pBA1670	<i>L. mexicana</i> 6HIS-KKT4 ^{159–262}	pNIC28-Bsa4	Kanamycin
pBA1695	6HIS-KKT4 ^{300–488} K310E, R311D, R313D	pNIC28-Bsa4	Kanamycin
pBA1696	6HIS-KKT4 ^{300–488} K326E, K327E, R328D	pNIC28-Bsa4	Kanamycin
pBA1697	6HIS-KKT4 ^{300–488} R456D, R457D, R460D	pNIC28-Bsa4	Kanamycin

Name	Description	Vector	Resistance
pBA1700	6HIS-KKT4 ^{300–488} K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D	pNIC28-Bsa4	Kanamycin
pBA1725	6HIS-KKT23 ^{2–348}	pNIC28-Bsa4	Kanamycin
pBA1737	YFP-KKT4 K310E, R311D, R313D	pDEX777	Ampicillin
pBA1741	YFP-KKT4 K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D	pDEX777	Ampicillin
pBA1781	6HIS-KKT4 ^{2–114} V8A, L9A, L12A, I19A, L26A, L29A	pNIC28-Bsa4	Kanamycin
pBA1782	<i>L. mexicana</i> 6HIS-KKT4 ^{184–284} R218Q	RSFDuet-1	Kanamycin
pBA1811	<i>L. mexicana</i> 6HIS-KKT4 ^{159–359}	RSFDuet-1	Kanamycin
pBA1821	<i>L. mexicana</i> 6HIS-KKT4 ^{159–284}	RSFDuet-1	Kanamycin
pBA1858	<i>T. cruzi sylvio</i> 6HIS-KKT4 ^{117–233}	RSFDuet-1	Kanamycin
pBA1865	6HIS-KKT23 ^{126–348}	RSFDuet-1	Kanamycin
pBA1866	6HIS-KKT23 ^{150–348}	RSFDuet-1	Kanamycin
pBA1868	6HIS-KKT4 ^{111–232}	RSFDuet-1	Kanamycin
pBA1900	<i>T. vivax</i> 6HIS-KKT4 ^{117–240}	RSFDuet-1	Kanamycin
pBA1901	<i>L. seymouri</i> 6HIS-KKT4 ^{121–262}	RSFDuet-1	Kanamycin
pBA1960	YFP-KKT4 V8A, L9A, L12A, I19A, L26A, L29A	pDEX777	Ampicillin
pBA1962	YFP-KKT4 ^{Δ300–488}	pDEX777	Ampicillin
pBA1984	6HIS-KKT7 ^{144–512}	RSFDuet-1	Kanamycin
pBA2036	6HIS-KKT4 ^{115–343} R167D	pNIC28-Bsa4	Kanamycin
pBA2037	6HIS-KKT4 ^{115–343} K218E	pNIC28-Bsa4	Kanamycin
pBA2038	6HIS-KKT4 ^{115–343} R230D	pNIC28-Bsa4	Kanamycin
pBA2042	6HIS-KKT4 ^{115–343} K204E	pNIC28-Bsa4	Kanamycin
pBA2043	6HIS-KKT4 ^{115–343} R145D	pNIC28-Bsa4	Kanamycin
pBA2044	6HIS-KKT4 ^{115–343} K166E	pNIC28-Bsa4	Kanamycin
pBA2045	6HIS-KKT4 ^{115–343} K206E	pNIC28-Bsa4	Kanamycin
pBA2054	6HIS-KKT4 ^{123–224}	RSFDuet-1	Kanamycin
pBA2058	6HIS-KKT4 ^{126–224}	RSFDuet-1	Kanamycin
pBA2064	6HIS-KKT4 ^{126–232}	RSFDuet-1	Kanamycin
pBA2128	<i>T. congolense</i> 6HIS-KKT4 ^{145–244}	RSFDuet-1	Kanamycin
pBA2129	<i>L. mexicana</i> 6HIS-KKT4 ^{166–268}	RSFDuet-1	Kanamycin
pBA2143	<i>T. congolense</i> 6HIS-KKT4 ^{134–244}	RSFDuet-1	Kanamycin

Name	Description	Vector	Resistance
pBA2151	<i>T. cruzi</i> sylvio 6HIS-KKT4 ¹¹⁷⁻²¹⁸	RSFDuet-1	Kanamycin
pBA2159	6HIS-KKT23 ⁹⁴⁻³⁴⁸	RSFDuet-1	Kanamycin
pBA2244	6HIS-KKT4 ¹¹⁵⁻³⁴³ R126D	RSFDuet-1	Kanamycin
pBA2245	6HIS-KKT4 ¹¹⁵⁻³⁴³ R130D	RSFDuet-1	Kanamycin
pBA2246	6HIS-KKT4 ¹¹⁵⁻³⁴³ K136E	RSFDuet-1	Kanamycin
pBA2247	6HIS-KKT4 ¹¹⁵⁻³⁴³ R140D	RSFDuet-1	Kanamycin
pBA2248	6HIS-KKT4 ¹¹⁵⁻³⁴³ R141D	RSFDuet-1	Kanamycin
pBA2249	6HIS-KKT4 ¹¹⁵⁻³⁴³ K144E	RSFDuet-1	Kanamycin
pBA2252	6HIS-KKT4 ¹¹⁵⁻³⁴³ K115E	RSFDuet-1	Kanamycin
pBA2253	6HIS-KKT4 ¹¹⁵⁻³⁴³ R164D	RSFDuet-1	Kanamycin
pBA2254	6HIS-KKT4 ¹¹⁵⁻³⁴³ K198E	RSFDuet-1	Kanamycin
pBA2255	6HIS-KKT4 ¹¹⁵⁻³⁴³ R217D	RSFDuet-1	Kanamycin
pBA2256	6HIS-KKT4 ¹¹⁵⁻³⁴³ K179E	RSFDuet-1	Kanamycin
pBA2257	6HIS-KKT4 ¹¹⁵⁻³⁴³ K199E	RSFDuet-1	Kanamycin
pBA2262	6HIS-KKT23 ¹²⁵⁻³⁴⁸	RSFDuet-1	Kanamycin
pBA2264	6HIS-KKT4 ⁴⁶³⁻⁶⁴⁵ K543A	pNIC28-Bsa4	Kanamycin
pBA2280	6HIS-KKT23 ¹⁴³⁻³⁴⁸	RSFDuet-1	Kanamycin
pBA2280	6HIS-KKT23 ¹⁴⁶⁻³⁴⁸	RSFDuet-1	Kanamycin
pBA2281	6HIS-KKT23 ¹⁴³⁻³⁴⁸	RSFDuet-1	Kanamycin
pBA2380	6HIS-KKT4 ²³³⁻³⁴³	RSFDuet-1	Kanamycin
pBA2403	6HIS-KKT23 ²⁻³⁴⁸ E235A	pNIC28-Bsa4	Kanamycin
pBA2414	6HIS-KKT23 ¹²⁵⁻³⁴⁸ E235A	RSFDuet-1	Kanamycin
pBA2415	6HIS-KKT23 ¹²⁵⁻³⁴⁸ G249A	RSFDuet-1	Kanamycin
pBA2416	6HIS-KKT23 ¹²⁵⁻³⁴⁸ G251A	RSFDuet-1	Kanamycin
pBA2417	6HIS-KKT23 ¹²⁵⁻³⁴⁸ R252A	RSFDuet-1	Kanamycin
pBA2418	6HIS-KKT23 ¹²⁵⁻³⁴⁸ H284A	RSFDuet-1	Kanamycin
pBA2419	6HIS-KKT23 ¹²⁵⁻³⁴⁸ T241A	RSFDuet-1	Kanamycin
pBA2420	6HIS-KKT23 ¹²⁵⁻³⁴⁸ H226A	RSFDuet-1	Kanamycin
pBA2421	6HIS-KKT23 ¹²⁵⁻³⁴⁸ P237A	RSFDuet-1	Kanamycin
pBA2422	6HIS-KKT23 ¹²⁵⁻³⁴⁸ E235A, P237A	RSFDuet-1	Kanamycin
pBA2423	6HIS-KKT23 ¹²⁵⁻³⁴⁸ W282A	RSFDuet-1	Kanamycin

Name	Description	Vector	Resistance
pBA2424	6HIS-KKT23 ¹²⁵⁻³⁴⁸ F239A	RSFDuet-1	Kanamycin
pBA2426	6HIS-KKT23 ²⁻⁷⁰	RSFDuet-1	Kanamycin
pBA2448	6HIS-H2A ²⁻¹³⁴	RSFDuet-1	Kanamycin
pBA2448	6HIS-H2A		
pBA2449	6HIS-H2A ¹⁶⁻¹³⁴	RSFDuet-1	Kanamycin
pBA2451	6HIS-KKT23 ¹²⁵⁻³⁴⁸ E235A, D275A	RSFDuet-1	Kanamycin
pBA2452	6HIS-KKT23 ¹²⁵⁻³⁴⁸ E235A, D275A	RSFDuet-1	Kanamycin

2.3. Protein expression and purification

Work in this thesis employed two expression systems to produce protein fragments for biochemical studies. The vast majority of proteins was produced using bacterial expression system, however, in cases of larger constructs, where bacterial expression was shown to be of limited solubility and yield, the use of baculovirus-mediated expression was used.

2.3.1. Bacterial expression of unlabelled proteins

Plasmids bearing a T7 lac promoter (RSF_Duet-1 and pNIC28-Bsa4) were transformed into chemically competent *E. coli* cells BL-21 (DE3) (Novagen) and grown overnight at 37°C in 2x TY medium (Table 2.6) containing an appropriate selective antibiotic. Following the overnight growth, 5 ml of the pre-culture was used to inoculate 1 l of 2x TY media containing the selective antibiotic (normally 1 l of culture was grown in 2.5L baffled flasks). Cells were grown at 37°C with shaking (200 rpm) until the optical density (OD₆₀₀) reached 0.6–0.8. The protein expression was induced with 0.2 mM IPTG (Sigma-Aldrich) and flasks were transferred to a pre-cooled 16°C or 20°C shaker. Protein expression was continued for 16–20 hours while shaking (200 rpm).

Table 2.6. Composition of the 2x TY medium.

Ingredient	Amount
Tryptone	16 g
Yeast extract	10 g
NaCl	5 g
Ultra-Pure water	up to 1 l

2.3.2. Bacterial expression of uniformly labelled proteins

For the expression of isotopically labelled proteins used in NMR studies, M9 minimal medium was used, which was supplemented with 1g/l $^{15}\text{NH}_4\text{Cl}$ and/or 4g/l [^{13}C]-*D*-glucose (Cambridge Isotope Laboratories) as the sole nitrogen and carbon sources. Table 2.7 represents the detailed composition of the M9 minimal medium. Colonies of BL-21 cells (DE3) transformed with plasmids expressing KKT4 fragments were used to inoculate 5 ml of 2x TY medium supplemented with 50 µg/ml kanamycin. Cells were grown for 6 hours at 37°C with shaking (220 rpm). Next, 1 part of this initial pre-culture was used to inoculate 100 parts of the M9 minimal medium supplemented with 50 µg/ml kanamycin, 1g/l $^{15}\text{NH}_4\text{Cl}$ and/or 4g/l [^{13}C]-*D*-glucose (Cambridge Isotope Laboratories). The pre-culture was incubated overnight at 37°C with shaking (220 rpm). Following the overnight growth, the pre-culture was used to inoculate a large-scale culture of minimal media (5 ml per 1 l of M9 media in 2.5 l flasks) and grown at 37°C with shaking until it reached an OD of 0.9–1.0. The protein expression was induced with 0.4 mM IPTG and the flasks were transferred to a pre-cooled shaker at 16°C for the next 16–20 hours.

The above protocol did not work for plasmids that express KKT4^{463–645} and KKT23 fragments (cells from the pre-culture did not recover after being inoculated into 1 l of M9 medium). Instead, the following changes were made. Transformed BL-21 (DE3) competent cells were inoculated into 40 ml of 2xTY media supplemented with 50 µg/ml kanamycin and grown overnight at 37°C with shaking. Following the overnight incubation, cells were spun down at 3400 g at 4°C. The supernatant was discarded and pellet was resuspended in 40 ml of M9 minimal media supplemented with 50 µg/ml kanamycin, 1g/l $^{15}\text{NH}_4\text{Cl}$ and/or 4g/l [^{13}C]-*D*-glucose (Cambridge Isotope Laboratories). Next, 40 ml of the resuspended culture was used to inoculate 1 l of

M9 media supplemented with 50 µg/ml kanamycin, 1g/l $^{15}\text{NH}_4\text{Cl}$ and/or 4g/l [^{13}C]-*D*-glucose (Cambridge Isotope Laboratories). The starting OD₆₀₀ was ~0.25. Cells were grown at 37°C with shaking (200rpm) until an OD₆₀₀ of 0.9–1.0. Protein expression was induced with 0.4 mM IPTG and carried out for 16–20h at 20°C with shaking (200rpm).

The cell growth and expression of deuterated KKT4 samples (^2H , ^{13}C , ^{15}N -KKT4^{145–232}) was done according to the work described in (Tugarinov et al., 2006).

Table 2.7. Composition of the M9 minimal medium.

Ingredient	Amount
NaCl	0.5 g
NH ₄ Cl/ ^{15}N -NH ₄ Cl	1 g
Na ₂ HPO ₄	6 g
KH ₂ PO ₄	3 g
1M Mg ₂ SO ₄ *7H ₂ O	2 ml
20% Glucose/ ^{13}C -Glucose	20 ml
1000x Vitamins	1 ml
100x Trace Elements	10 ml
Ultra-pure water	up to 1 l

Table 2.8. Components for 1000x vitamins.

Ingredient	Amount
Rboflavin	0.5 g
Niacinamide	0.5 g
Pyridoxine monohydrate	0.5 g
Thiamine (hydrochloride)	0.5 g
Ultra-Pure water	up to 500 ml

Table 2.9. Components for 100x trace elements

Ingredient	Amount
EDTA	5.00 g
FeSO ₄ *7H ₂ O	1.40 g
ZnSO ₄ *7H ₂ O	0.05 g
CuCl ₂	0.01 g
CoCl ₂ *6H ₂ O	0.01 g
H ₃ BO ₃	0.01 g
MnCl ₂ *4H ₂ O	1.60 g
NiSO ₄ *6H ₂ O	0.001 g
Na ₂ MoO ₄	0.001 g
Ultra-pure water	up to 1 l

2.3.3. Purification of His-tagged recombinant proteins

Small-scale purification

Small-scale purification was sufficient to purify samples for *in vitro* microtubule co-sedimentation and acetylation assays. It was also used to perform test expression of protein fragments used in this thesis. Typically 1 l of bacterial culture was grown.

Following the overnight protein expression, cells were spun at 3,400 g at 4°C and resuspended in lysis buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, and 10% glycerol) supplemented with protease inhibitors (20 µg/ml leupeptin, 20 µg/ml pepstatin, 20 µg/ml E-64 and 0.4 mM PMSF), benzonase nuclease (500U/l culture) and 0.5 mM TCEP. All subsequent extraction steps were performed at 4°C. Cells were disrupted using sonication (4 rounds of 15s pulse with a 1 min break in-between with 80% amplitude). Lysed cells were spun at 48 000 g for 30 min and the

supernatant was incubated for 1 h with TALON beads (Takara Clontech) pre-equilibrated in the lysis buffer (0.75 ml of TALON per 1 l of bacterial culture). Following the incubation, samples were centrifuged at 500 g for 2 min, supernatant (flow-through) was discarded and the beads were washed with 10 ml of lysis buffer. This was repeated 4 more times. After 5 washes, the beads were loaded on a gravity column (Biorad) and were once again washed with 5 ml of the lysis buffer. Proteins were eluted with the elution buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole, 0.5 mM TCEP), aliquots were flash-frozen in liquid nitrogen and stored at -80°C .

Large-scale purification

The following protocol was used to purify samples for X-ray crystallography, NMR, SEC-MALS, AUC and fluorescence anisotropy assays. Typically 6–12 l of bacterial culture were grown to express protein.

Cells were pelleted at 3 400 g at 4°C and resuspended in lysis buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, and 10% glycerol) supplemented with protease inhibitors (20 $\mu\text{g/ml}$ leupeptin, 20 $\mu\text{g/ml}$ pepstatin, 20 $\mu\text{g/ml}$ E-64 and 0.4 mM PMSF), benzonase nuclease (500U/l culture) and 0.5 mM TCEP. All subsequent extraction steps were performed at 4°C . Cell lysis was facilitated by mechanical cell disruption (French press, 1 passage at 20 000 psi). Lysed cells were spun at 48 000 g for 30 min and the supernatant was loaded on a gravity column loaded with TALON beads (Takara Clontech) pre-equilibrated in lysis buffer (1.5 ml of beads per 1 l of bacterial culture). After the loading, beads were washed extensively with the lysis buffer (~300 ml) without protease inhibitors and proteins were eluted with elution buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, 10% glycerol, 250

mM imidazole, 0.5 mM TCEP). To remove the N-terminal His-tag, the proteins were incubated with TEV protease in 1:50 w/w ratio and dialysed overnight against buffer containing 25 mM sodium phosphate pH 7.5, 250 mM NaCl, 5% glycerol, 5 mM imidazole and 0.5 mM TCEP. Next morning, to remove the cleaved His-tag and increase the sample purity, samples were re-loaded on TALON beads pre-equilibrated with dialysis buffer and the flow-through was collected. Next, the samples were further purified using either one-step (size exclusion chromatography) or two-step (ion exchange and size exclusion chromatography) purification. The purification strategy for individual proteins can be found in Table 2.10. To promote the binding of proteins to ion exchange columns, protein samples were diluted with 25 mM HEPES pH 7.5 with 0.5 mM TCEP (buffer A) to the final NaCl concentration of 50 mM. Ion exchange chromatography was done using either cation exchange 6 ml RESOURCE S or anion exchange 6 ml RESOURCE Q column (GE Healthcare), pre-equilibrated with 5% of buffer B (25 mM HEPES pH 7.5, and 1 M NaCl with 0.5 mM TCEP). Proteins were eluted from the column with a linear gradient from 5% to 100% of buffer B. Fractions containing desired protein were pooled, concentrated and loaded on a gel filtration column SD200 or SD75 16/60 (GE Healthcare) to further purify and buffer exchange into 25 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM TCEP (^{13}C , ^{15}N -KKT4^{463–645} and ^{13}C , ^{15}N -KKT4^{125–348} used for NMR studies, were buffer exchanged into 50 mM sodium phosphate pH 7.0, 150 mM NaCl and 0.5 mM TCEP, while the full-length KKT23^{2–348} used in *in vitro* acetylation assay was buffer exchanged into 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM DTT, 0.1 mM EDTA and 10 mM sodium butyrate). The samples were concentrated using a 3 or 10-kD MW Amicon concentrator (Millipore), and aliquots were flash-frozen in liquid nitrogen and stored at -80°C .

Table 2.10 List of proteins used in this study and their purification strategy.

Plasmid	Protein	Purification step		
		Affinity Chromatography	Ion Exchange Chromatography	Size Exclusion Chromatography
pBA435	<i>Tb</i> 6HIS-H3/H2AZ/H2BV/H4	Talon (Takara)	-	SD75 10/300
pBA443	<i>Tb</i> 6HIS-H3/H2A/H2B/H4	Talon (Takara)	-	SD75 10/300
pBA987	<i>Tb</i> KKT4 ^{115–645}	Talon (Takara)	-	SD200 16/60
pBA1065	<i>Tb</i> KKT4 ^{115–343}	Talon (Takara)	RESOURSCE S	SD200 16/60
pBA1065	<i>Tb</i> ¹⁵ N-KKT4 ^{115–343}	Talon (Takara)	RESOURSCE S	SD200 16/60
pBA1065	<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–343}	Talon (Takara)	RESOURSCE S	SD200 16/60
pBA1066	<i>Tb</i> KKT4 ^{250–645}	Talon (Takara)	-	SD200 16/60
pBA1171	<i>Tb</i> KKT4 ^{115–174}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1171	<i>Tb</i> ¹⁵ N-KKT4 ^{115–174}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1171	<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–174}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1328	KKT4 ^{115–343} R123E	Talon (Takara)	-	-
pBA1329	KKT4 ^{115–343} K132E	Talon (Takara)	-	-
pBA1330	KKT4 ^{115–343} K154E	Talon (Takara)	-	-
pBA1393	<i>Tb</i> KKT4 ^{101–352}	Talon (Takara)	-	SD200 16/60
pBA1413	<i>Tb</i> KKT4 ^{2–114}	Talon (Takara)	-	SD75 16/60
pBA1413	<i>Tb</i> ¹⁵ N-KKT4 ^{2–114}	Talon (Takara)	-	SD75 16/60
pBA1433	<i>Tb</i> KKT4 ^{343–645}	Talon (Takara)	-	SD200 16/60
pBA1441	<i>Tb</i> KKT4 ^{145–232}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1441	<i>Tb</i> ¹⁵ N-KKT4 ^{145–232}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1441	<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{145–232}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1441	<i>Tb</i> ² H, ¹³ C, ¹⁵ N-KKT4 ^{145–232}	Talon (Takara)	RESOURSCE S	SD75 16/60
pBA1513	<i>Tb</i> KKT4 ^{463–645}	Talon (Takara)	-	SD200 16/60
pBA1513	<i>Tb</i> ¹⁵ N-KKT4 ^{463–645}	Talon (Takara)	-	SD200 16/60
pBA1513	<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{463–645}	Talon (Takara)	-	SD200 16/60
pBA1537	<i>Lm</i> KKT4 ^{184–284} R218Q	Talon (Takara)	-	SD75 16/60
pBA1572	<i>Tb</i> KKT4 ^{114–232}	Talon (Takara)	-	SD75 16/60
pBA1601	KKT4 ^{115–232}	Talon (Takara)	-	SD75 16/60

Plasmid	Protein	Purification step		
pBA1601	<i>Tb</i> ¹⁵ N-KKT4 ^{115–232}	Talon (Takara)	-	SD75 16/60
pBA1601	<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–232}	Talon (Takara)	-	SD75 16/60
pBA1618	<i>Lm</i> KKT4 ^{184–284}	Talon (Takara)	-	SD75 16/60
pBA1641	<i>Tb</i> KKT4 ^{300–488}	Talon (Takara)	-	SD75 16/60
pBA1695	<i>Tb</i> KKT4 ^{300–488} K310E, R311D, R313D	Talon (Takara)	-	SD75 16/60
pBA1696	<i>Tb</i> KKT4 ^{300–488} K326E, K327E, R328D	Talon (Takara)	-	SD75 16/60
pBA1697	<i>Tb</i> KKT4 ^{300–488} R456D, R457D, R460D	Talon (Takara)	-	SD75 16/60
pBA1700	<i>Tb</i> KKT4 ^{300–488} K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D	Talon (Takara)	-	SD75 16/60
pBA1725	<i>Tb</i> KKT23 ^{2–348}	Talon (Takara)	RESOURCESCE Q	SD75 16/60
pBA1725	<i>Tb</i> ¹⁵ N-KKT23 ^{2–348}	Talon (Takara)	RESOURCESCE Q	SD75 16/60
pBA1791	<i>Tb</i> KKT4 ^{2–114} V8A, L9A, L12A, I19A, L26A and L29A	Talon (Takara)	-	SD75 16/60
pBA2036	<i>Tb</i> KKT4 ^{115–343} R167D	Talon (Takara)	-	-
pBA2037	<i>Tb</i> KKT4 ^{115–343} K218E	Talon (Takara)	-	-
pBA2038	<i>Tb</i> KKT4 ^{115–343} R230D	Talon (Takara)	-	-
pBA2042	<i>Tb</i> KKT4 ^{115–343} K204E	Talon (Takara)	-	-
pBA2043	<i>Tb</i> KKT4 ^{115–343} R145D	Talon (Takara)	-	-
pBA2044	<i>Tb</i> KKT4 ^{115–343} K166E	Talon (Takara)	-	-
pBA2045	<i>Tb</i> KKT4 ^{115–343} K206E	Talon (Takara)	-	-
pBA2151	<i>Tc</i> KKT4 ^{117–218}	Talon (Takara)	RESOURCESCE S	SD75 16/60
pBA2245	<i>Tb</i> KKT4 ^{115–343} R130D	Talon (Takara)	-	-
pBA2246	<i>Tb</i> KKT4 ^{115–343} K136E	Talon (Takara)	-	-
pBA2247	<i>Tb</i> KKT4 ^{115–343} R140D	Talon (Takara)	-	-
pBA2248	<i>Tb</i> KKT4 ^{115–343} R141D	Talon (Takara)	-	-
pBA2249	<i>Tb</i> KKT4 ^{115–343} K144E	Talon (Takara)	-	-
pBA2252	<i>Tb</i> KKT4 ^{115–343} K115E	Talon (Takara)	-	-
pBA2253	<i>Tb</i> KKT4 ^{115–343} R164D	Talon (Takara)	-	-
pBA2254	<i>Tb</i> KKT4 ^{115–343} K198E	Talon (Takara)	-	-

Plasmid	Protein	Purification step		
pBA2255	<i>Tb</i> KKT4 ^{115–343} R217D	Talon (Takara)	-	-
pBA2256	<i>Tb</i> KKT4 ^{115–343} K179E	Talon (Takara)	-	-
pBA2257	<i>Tb</i> KKT4 ^{115–343} K199E	Talon (Takara)	-	-
pBA2262	<i>Tb</i> KKT23 ^{125–348}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2262	<i>Tb</i> ¹⁵ N-KKT23 ^{125–348}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2262	<i>Tb</i> ¹³ C, ¹⁵ N-KKT23 ^{125–348}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2264	<i>Tb</i> KKT4 ^{463–645} K543A	Talon (Takara)	-	SD200 16/60
pBA2380	<i>Tb</i> KKT4 ^{233–343}	Talon (Takara)	-	SD75 10/300
pBA2403	<i>Tb</i> KKT23 ^{2–348} E235A	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2414	<i>Tb</i> KKT23 ^{125–348} E235A	Talon (Takara)	-	-
pBA2415	<i>Tb</i> KKT23 ^{125–348} G249A	Talon (Takara)	-	-
pBA2416	<i>Tb</i> KKT23 ^{125–348} G251A	Talon (Takara)	-	-
pBA2417	<i>Tb</i> KKT23 ^{125–348} R252A	Talon (Takara)	-	-
pBA2418	<i>Tb</i> KKT23 ^{125–348} H284A	Talon (Takara)	-	-
pBA2419	<i>Tb</i> KKT23 ^{125–348} T241A	Talon (Takara)	-	-
pBA2420	<i>Tb</i> KKT23 ^{125–348} H226A	Talon (Takara)	-	-
pBA2421	<i>Tb</i> KKT23 ^{125–348} P237A	Talon (Takara)	-	-
pBA2422	<i>Tb</i> KKT23 ^{125–348} E235A, P237A	Talon (Takara)	-	-
pBA2423	<i>Tb</i> KKT23 ^{125–348} W282A	Talon (Takara)	-	-
pBA2424	<i>Tb</i> KKT23 ^{125–348} F239A	Talon (Takara)	-	-
pBA2426	<i>Tb</i> KKT23 ^{2–70}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2426	<i>Tb</i> ¹⁵ N-KKT23 ^{2–70}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2426	<i>Tb</i> KKT23 ^{2–70}	Talon (Takara)	RESOURSCE Q	SD75 16/60
pBA2448	<i>Tb</i> H2A ^{2–134}	Talon (Takara)	-	-
pBA2449	<i>Tb</i> H2A ^{16–134}	Talon (Takara)	-	-
pBA2451	<i>Tb</i> KKT23 ^{125–348} E235A, D275A	Talon (Takara)	-	-
pBA2452	<i>Tb</i> KKT23 ^{125–348} E235A, D275A	Talon (Takara)	-	-

2.3.4. Baculovirus-mediated expression

To express recombinant proteins in *Spodoptera frugiperpa* (Sf9) cells, it requires three sequential transfection/infection cycles before the large-scale protein expression. Sf9 cells were cultured in SF900II medium (Gibco) and grown at 27°C with shaking (100 rpm). Cell density was calculated by manual counting using haemocytometer and the light microscope. Cells were continuously grown, oscillating between 0.6×10^6 and 4.0×10^6 cell/ml. Cell culture was diluted every 48–72 h.

Plasmid preparation

Transformation of pACEBac plasmid into DH10EmBacY competent cells is the first step of baculovirus preparation. Around 200 ng of plasmid was transformed into 100 μ l of chemically competent DH10EmBacY *E.coli* cells. Detailed transformation protocol is highlighted in Table 2.11.

Table 2.11 Transformation of pACEBac plasmid into DH10EmBacY cells

Step	Temperature [°C]	Time
1	4	45 min
2	42	1 min
3	4	2 min
4	Add 1 ml of 2x TY media	
5	37	6 hours with shaking

Transformed cells were plated onto LB plates supplemented with 15 µg/ml gentamycin, 50 µg/ml kanamycin, 10 µg/ml tetracyclin, 100 µg/ml BlueGal and 1 mM IPTG. Cells were grown for 2 days at 30°C. BlueGal is a dye used to assess the success of the transformation. White colonies result from a successful transformation, while the blue colonies indicate cells that failed to incorporate plasmid DNA. Two single white colonies were inoculated into separate tubes, each containing 20 ml of 2x TY medium supplemented with 15 µg/ml gentamycin and 50 µg/ml kanamycin. The culture was incubated at 30°C overnight with shaking (220rpm). The next day, 7.5 ml of each culture was spun down at 3000 g in RT for 15 minutes and the bacmid DNA was purified from the pellet using PureLink HiPure Plasmid Miniprep Kit (Invitrogen) according to the manufacturer's protocol. The plasmid DNA was eluted in TE buffer (Invitrogen) and stored at 4°C.

Recombinant baculovirus preparation

Sf9 cells at a density of $2-3 \times 10^6$ cell/ml were diluted into 6-well plates so that each well contained 0.8–1.0 million cells. Plates prepared this way were incubated for 1 h at 27°C to allow cells to adhere to the bottom of the wells. Following the incubation, purified plasmid DNA was mixed with a transfection agent, Cellfectin (Invitrogen), incubated for 30 min and gently pipetted into each well with adhered S9 cells. A sample with no DNA was prepared as a control of the transfection. Cells were incubated with cellfectin/DNA mix for 4–5 h at 27°C. Next, the media was removed from each well and fresh pre-warmed SF900II media without antibiotics was added. Plates prepared this way were incubated for 5 days at 27°C. After 5 days, cells were harvested, combined and spun at 500 g for 3 minutes. Supernatant, a virus stock (P1) was collected and inoculated into 30 ml of Sf9 culture at 1×10^6 cells/ml in a T75 flask. Cells were incubated for 5 days at 27°C without shaking. After 5 days,

cells were harvested and spun down at 500 g for 3 min and the virus stock supernatant (P2) was collected and stored in the dark at 4°C. Further amplification was carried out by infecting 30 ml of 1×10^6 cell/ml Sf9 cells with 1 ml of P2 viral stock. The culture was incubated at 27°C with shaking (~100 rpm) for 4 days. Following the incubation, cells were harvested by centrifugation at 500 g for 3 min and the virus stock supernatant (P3) was collected and stored in the dark at 4°C.

Table 2.12 List of baculoviruses used in this work.

Name	Description	Vector	Resistance
pBA314	3FLAG-KKT2	pACEBac2/EmBacY	Gentamycin
pBA315	3FLAG-KKT3	pACEBac2/EmBacY	Gentamycin
pBA826	3FLAG-KKT4	pACEBac2/EmBacY	Gentamycin
pBA828	3FLAG-6HIS-KKT6, KKT1	pACEBac2/EmBacY	Gentamycin
pBA942	SNAP-6HIS-3FLAG-KKT4	pACEBac2/EmBacY	Gentamycin
pBA1371	3FLAG-KKT4, KKT14, KKT15	pACEBac2/EmBacY	Gentamycin
pBA1386	3FLAG-KKT4, KKT7	pACEBac2/EmBacY	Gentamycin
pBA1388	3FLAG-KKT4, KKT14, KKT15	pACEBac2/EmBacY	Gentamycin
pBA1531	3FLAG-6HIS-KKT7	pACEBac2/EmBacY	Gentamycin
pBA1540	KKIP1	pACEBac1/EmBacY	Gentamycin
pBA1682	KKT3	pACEBac2/EmBacY	Gentamycin
pBA1684	KKT4	pACEBac2/EmBacY	Gentamycin
pBA1903	FLAG-KKT22	pACEBac2/EmBacY	Gentamycin
pBA1925	KKT22	pACEBac2/EmBacY	Gentamycin
pBA2274	KKT23	pACEBac2/EmBacY	Gentamycin
pBA2279	FLAG-KKT23	pACEBac2/EmBacY	Gentamycin

Protein expression and purification from insect cells

Typically, 500 ml of Sf9 culture at 1.2×10^6 cell/ml was used for large scale purification and 50 ml of Sf9 culture was used for small scale purification. The protein expression was started by mixing 1 part of P3 viral stock with 100 parts of insect cell culture. Protein expression was carried out for 72–96 h post the infection at 27°C with shaking (100 rpm). All purification steps were performed at 4°C. Cells were pelleted at 700 g for 10 min, washed once with PBS, and resuspended in BH0.25 buffer (25 mM HEPES, pH 8.0, 250 mM NaCl, 10% glycerol, 2 mM MgCl_2 , 0.1 mM EDTA, 0.5 mM EGTA-KOH, 0.2% NP-40) supplemented with protease inhibitors (20 $\mu\text{g/ml}$ leupeptin, 20 $\mu\text{g/ml}$ pepstatin, 20 $\mu\text{g/ml}$ E-64, and 0.4 mM PMSF), benzonase nuclease (500U/l of culture) and 0.5 mM TCEP. Cells were sonicated on ice (three rounds of 15s pulse with a 1min break in between with 80% amplitude) and the lysate was centrifuged for 30 min at 45 000 g. The supernatant was incubated with anti-FLAG M2 affinity gel (Sigma) (1 ml of beads per 1 l of cell culture) for 3 h with constant rotation. Following the incubation, the beads were extensively washed with BH0.25 supplemented with 0.5 mM TCEP. The proteins were eluted from the beads with gentle agitation in BH0.25 buffer containing 0.5 mg/ml 3FLAG peptide (Sigma). No reducing agents were included in the elution buffer. The proteins were flash-frozen in liquid nitrogen and stored at -80°C . SNAP-3FLAG-6HIS-KKT4^{2–645} used in SEC-MALS and fluorescence anisotropy assays was further purified using 1 ml HiTrap Heparin HP column pre-equilibrated with 5% of buffer B (buffer A: 25 mM HEPES, pH 7.5, with 0.5 mM TCEP; buffer B: 25 mM HEPES, pH 7.5, 1 M NaCl with 0.5 mM TCEP) and eluted with a linear gradient from 5% to 100% of buffer B. Fractions containing SNAP-tagged KKT4 (eluted at ~ 340 mM NaCl) were pooled, concentrated to 20 μM using a 10-kD MW Amicon concentrator (Millipore), flash-frozen in liquid nitrogen and stored at -80°C .

2.4. X-ray Crystallography

The following section describes the protein crystallisation, collection of the X-ray data and methods used to solve the crystal structures.

Crystallisation

All crystals were obtained in sitting drop vapour diffusion experiments in 96-well plates, using drops of overall volume 200 nl, mixing protein and mother liquor in a 1:1 v/v ratio. The images of obtained crystals can be found in Figure A1.

Crystals of *Trypanosoma cruzi* (Sylvio X10) KKT4^{117–218} (10.0 mg/ml) were grown at 18°C in Morpheus II crystallisation screen (Molecular Dimensions) solution containing 0.1 M buffer system 4 (MOPSO, Bis-Tris) pH 6.5, 50% v/v precipitant mix 7 (20% w/v PEG 8000, 40% v/v 1,5-Pentanediol) and 100 mM amino acids II (0.2 M DL-Arginine hydrochloride, 0.2 M DL-Threonine, 0.2M DL-Histidine monohydrochloride monohydrate, 0.2 M DL-Hydroxylysine hydrochloride, 0.2 M trans- 4-hydroxy-L-proline). Mother liquor served as a cryoprotectant.

Crystals of *Leishmania mexicana* KKT4^{184–284} (13.5 mg/ml) were grown at 4°C in ProPlex crystallisation screen (Molecular Dimensions) solution containing 0.1 M imidazole pH 7.0 and 50% v/v MPD. Mother liquor served as a cryoprotectant.

Crystals of *Trypanosoma brucei* KKT4^{463–645} (26.5 mg/ml) were grown at 4°C in Index crystallisation screen (Hampton Research) solution containing 0.1 M bis-Tris pH 5.5 and 2.0 M ammonium sulphate. Crystals were briefly transferred into mother liquor containing 23% glycerol prior to flash-cooling by plunging into liquid nitrogen.

Crystals of *Trypanosoma brucei* KKT4^{463–645} K543A (24 mg/ml) were grown at 4°C in JCSG+ crystallisation screen (Hampton Research) solution containing 0.2 M sodium thiocyanate and 20% w/v PEG3350. Crystals were briefly transferred into mother liquor containing 23% glycerol prior to flash-cooling by plunging into liquid nitrogen.

Prior to crystallisation, *Trypanosoma brucei* SeMet-KKT23^{125–348} and *Trypanosoma brucei* KKT23^{125–348} at 12.5 and 14.5 mg/ml, respectively, were mixed with a 2 molar excess of acetyl coenzyme A. Crystals of *Trypanosoma brucei* SeMet-KKT23^{125–348} appeared at 4°C in MIDAS crystallisation screen (Molecular Dimensions) solution containing 20% v/v jeffamine M-2070 and 20% v/v DMSO. Crystals of *Trypanosoma brucei* KKT23^{125–348} appeared at 4°C in MIDAS crystallisation screen (Molecular Dimensions) solution containing 0.1 M HEPES pH 7.0, 0.1 M sodium tartrate dibasic dehydrate and 20% SOKOLAN PA 25 CL. Prior to the cryo-cooling, crystals were briefly transferred to mother liquor supplemented with 24% glycerol.

Data collection and model building

X-ray diffraction data from *TcKKT*^{117–218} and *LmKKT*^{4184–284} were collected at the I03 and I24, beamlines, respectively, at the Diamond Light Source (Harwell, UK). The structures were solved using ARCIMBOLDO LITE optimised for coiled-coils (Rodríguez et al., 2009, Rodríguez et al., 2012), followed by initial model building with BUCCANEER (Cowtan, 2006). Next, the data files were scaled to the high-resolution limit of 1.9 Å and processed using anisotropic scaling (Strong et al., 2006). Further manual model building and refinement were completed iteratively using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010).

X-ray diffraction data from *TbKKT*^{463–645} were collected at the I24 beamline at the Diamond Light Source (Harwell, UK). The structure was solved using ARCIMBOLDO LITE (Rodríguez et al., 2009, Rodríguez et al., 2012), followed by initial model building with BUCCANEER (Cowtan, 2006). The further model building and refinement were completed using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010).

The diffraction data for *TbKKT*^{463–645} K543A were collected at the I03 beamline at the Diamond Light Source (Harwell, UK). The structure was solved by molecular replacement using PHASER (McCoy et al., 2007), where the already obtained structure of *KKT*^{463–645} (PDB: 6ZPK) was used as a model. The initial model building was completed by BUCCANEER (Cowtan, 2006), which was then followed by the iterative refinement and model building using PHENIX (Adams et al., 2010) and COOT (Emsley et al., 2010), respectively.

X-ray diffraction data for *TbSeMet-KKT*^{23^{125–348}} were collected at I03 beamline at the Diamond Light Source (Harwell, UK). Phases were solved using an automated structure solution pipeline for experimental phasing, CRANK2 software (Skubak and Pannu, 2013), followed by the initial model building with BUCCANEER (Cowtan, 2006). The further model building and refinement were completed using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010).

X-ray diffraction data from the native crystals of *TbKKT*^{23^{125–348}} were collected at I03 beamline at the Diamond Light Source (Harwell, UK). The structure was solved using PHASER (McCoy et al., 2007) with the *SeMet-KKT*^{23^{125–348}} crystal structure as a search model. The molecular replacement was followed by the initial model

building with BUCCANEER (Cowtan, 2006) and the further model building and refinement were completed using COOT (Emsley et al., 2010) and PHENIX (Adams et al., 2010), respectively. The final refinement was done using the original dataset scaled to high resolution limit of 1.75Å.

The final refinement statistics for six structures are summarised in Tables 3.3, 5.1 and 6.1. All structure figures were prepared using PyMOL (DeLano, 2002). Protein coordinates have been deposited in the RCSB Protein Data Bank (<http://www.rcsb.org/>) with accession codes: PDB:6ZPM (*Trypanosoma cruzi* sylvio KKT4^{117–218}), PDB:6ZPJ (*Leishmania mexicana* KKT4^{184–284}) and PDB:6ZPK (*Trypanosoma brucei* KKT4^{463–645}). The coordinates for *Trypanosoma brucei* KKT4^{463–645} K543A and KKT23^{125–348} structures are yet to be deposited.

2.5. NMR Spectroscopy

Solution state nuclear magnetic resonance (NMR) is one of the primary methods used to study structure and dynamics of KKT4 and KKT23 fragments throughout this thesis. This section describes in detail the sample preparation and measurements including heteronuclear NOE (hetNOE) and residual dipolar couplings (RDC) used to analyse KKT proteins.

Sample Preparation

Isotopically-labelled protein samples were used at a range of concentrations and different buffers supplemented with 5% D₂O for deuterium locking. NMR experiments were performed in D₂O-matched symmetrical Shigemi microtubes (Shigemi Inc) or shaped tubes with a minimal volume of 350 µl. The detailed composition of samples is listed in Table 2.13.

Table 2.13. List of samples used in NMR studies

Protein	Temperature [°C]	Buffer composition	Concentration [μM]
KKT4^{115–343}			
<i>Tb</i> ¹⁵ N-KKT4 ^{115–343}	20	25 mM HEPES pH 7.15, 150 mM NaCl, 0.5 mM TCEP	380
<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–343}	20	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	420
<i>Tb</i> ¹⁵ N-KKT4 ^{115–343} + <i>Tb</i> KKT4 ^{463–645}	20	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	220 μM KKT4 ^{115–343} + 220 μM KKT4 ^{463–645}
KKT4^{115–232}			
<i>Tb</i> ¹⁵ N-KKT4 ^{115–232}	30	25 mM HEPES pH 7.2, 150 mM NaCl, 0.5 mM TCEP	470
<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–232}	30	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	460
<i>Tb</i> ¹⁵ N-KKT4 ^{115–232} + <i>Tb</i> KKT4 ^{463–645}	20	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	220 μM KKT4 ^{115–232} + 220 μM KKT4 ^{463–645}
KKT4^{115–174}			
<i>Tb</i> ¹⁵ N-KKT4 ^{115–174}	20	25 mM HEPES pH 7.22, 150 mM NaCl, 0.5 mM TCEP	560
<i>Tb</i> ¹⁵ N-KKT4 ^{115–174} (RDC measurements)	20	50 mM sodium phosphate pH 7.5, 150 mM NaCl, 0.5 mM TCEP	400
<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{115–174}	20	25 mM HEPES pH 7.22, 150 mM NaCl, 0.5 mM TCEP	430
<i>Tb</i> ¹⁵ N-KKT4 ^{115–174} + <i>Tb</i> KKT4 ^{463–645} (BRCT)	20	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	220 μM KKT4 ^{115–174} + 220 μM KKT4 ^{463–645}
KKT4^{145–232}			
<i>Tb</i> ¹⁵ N-KKT4 ^{145–232}	30	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	370
<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{145–232}	30	25 mM HEPES pH 7.2, 150 mM NaCl, 0.5 mM TCEP	400
<i>Tb</i> ² H, ¹³ C, ¹⁵ N-KKT4 ^{145–232}	30	25 mM HEPES pH 7.2, 150 mM NaCl, 0.5 mM TCEP	350
<i>Tb</i> ¹⁵ N-KKT4 ^{145–232} + <i>Tb</i> KKT4 ^{463–645} (BRCT)	20	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	220 μM KKT4 ^{145–232} + 220 μM KKT4 ^{463–645}
KKT4^{2–114}			
<i>Tb</i> ¹⁵ N-KKT4 ^{2–114}	30	25 mM HEPES pH 7.16, 150 mM NaCl, 0.5 mM TCEP	400
<i>Tb</i> KKT4 ^{2–114} V8A, L9A, L12A, I19A, L26A, L29A (natural abundance)	30	25 mM HEPES pH 7.06, 150 mM NaCl, 0.5 mM TCEP	400
KKT4^{463–645}			
<i>Tb</i> ¹⁵ N-KKT4 ^{463–645}	20	50 mM sodium phosphate pH 7.0, 100 mM NaCl, 0.5 mM TCEP	474
<i>Tb</i> ¹⁵ N-KKT4 ^{463–645} (RDC measurements)	20	50 mM sodium phosphate pH 7.0, 100 mM NaCl, 0.5 mM TCEP	200
<i>Tb</i> ¹⁵ N-KKT4 ^{463–645} (RDC measurements)	20	25 mM HEPES pH 7.0, 50 mM NaCl, 0.5 mM TCEP	200
<i>Tb</i> ¹⁵ N-KKT4 ^{463–645} (KKT8 peptide titration)	20	25 mM HEPES pH 7.1, 50 mM NaCl, 0.5 mM TCEP	100 μM KKT4, 25–200 μM KKT8 peptide
<i>Tb</i> ¹³ C, ¹⁵ N-KKT4 ^{463–645}	20	50 mM sodium phosphate pH 7.0, 100 mM NaCl, 0.5 mM TCEP	500
KKT23			
<i>Tb</i> ¹⁵ N-KKT23 ^{2–348}	20, 30, 40	25 mM HEPES pH 7.3, 150 mM NaCl, 0.5 mM TCEP	245 μM + 500 μM acetyl CoA
<i>Tb</i> ¹⁵ N-KKT23 ^{2–70}	50	25 mM HEPES pH 7.3, 75 mM NaCl, 0.5 mM TCEP	230
<i>Tb</i> ¹⁵ N-KKT23 ^{125–348}	20	25 mM HEPES pH 7.15, 150 mM NaCl, 0.5 mM TCEP	350 μM + 700 μM acetyl CoA
<i>Tb</i> ¹³ C, ¹⁵ N-KKT23 ^{125–348}	20	50 mM sodium phosphate pH 7.0, 150 mM NaCl, 0.5 mM TCEP	435 μM + 500 μM acetyl CoA
<i>Tb</i> ¹⁵ N-KKT23 ^{125–348} (RDC measurements)	20	25 mM HEPES pH 7.15, 150 mM NaCl, 0.5 mM TCEP	350 μM + 700 μM acetyl CoA

Data Collection Parameters

NMR spectrometers used in this thesis range in ^1H frequencies between 750 MHz and 950 MHz. The spectrometers were equipped with a Bruker Avance III HD console and a 5mm TCI CryoProbe. The spectrometers are located in the Department of Biochemistry at the University of Oxford. All NMR data were processed using NMRPipe (Delaglio et al., 1995) and analysed using CCPN Analysis version 2 (Vranken et al., 2005).

Analysis of chemical shifts

^1H , ^{13}C and ^{15}N chemical shifts of KKT4^{115–174}, KKT4^{145–232}, KKT4^{115–343}, KKT4^{463–645} and KKT23^{125–348} were analysed using TALOS-N (Shen and Bax, 2013) and/or SSP (Marsh et al., 2006) to predict secondary structure propensities.

Backbone dynamics

The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE was measured for 0.2–0.5 mM samples of KKT4^{115–174}, KKT4^{145–232}, KKT4^{115–343}, KKT4^{463–645}, KKT23^{125–348} and KKT23^{2–70} using the TROSY-based heteronuclear NOE experiment recorded with and without ^1H saturation for either 4.0 or 4.5s at 750 MHz (Zhu et al., 2000). Data were collected at 20°C for KKT4^{115–174}, KKT4^{115–343}, KKT4^{463–645}, KKT23^{125–348} and at 30°C for KKT4^{145–232}. A temperature of 50°C was used for KKT23^{2–70}. The $\{^1\text{H}\}$ - ^{15}N NOE was calculated as the ratio of the peak intensities in the spectra recorded with and without ^1H saturation. Peak heights were determined using CCPN Analysis. Uncertainties in the $\{^1\text{H}\}$ - ^{15}N NOE values were estimated from 500 Monte Carlo simulations using the baseline noise as a measure of the error in the peak heights.

Residual Dipolar Coupling Measurement

Partial alignment of KKT4 fragments was achieved using C12E6/*n*-hexanol liquid crystals prepared as described by Rückert and Otting (Rückert and Otting, 2000). A 10% C12E6/*n*-hexanol solution was prepared in HEPES buffer and added to protein samples to achieve the desired final concentration of 3% for KKT4^{115–174}.

NMR experiments were performed at 750 MHz. ¹⁵N-¹H^N RDCs were measured at 20°C for KKT4^{115–174} using BEST TROSY and BEST semi-TROSY experiments (Schulte-Herbruggen and Sorensen, 2000, Lescop et al., 2007). 128 complex points and a sweep width of 1597.444 Hz were collected in F₁ (¹⁵N) for KKT4^{115–174}. Residual dipolar couplings were measured as the difference between the splitting observed in the isotropic and aligned data sets.

The principal components (A_{xx} , A_{yy} and A_{zz}) and orientation (θ , ϕ and ψ) of the molecular alignment tensor were fitted to minimise the χ^2 between the experimental and calculated RDCs using the *T. cruzi* X-ray coordinates. The sequence of KKT4^{115–174} was aligned with different positions in the X-ray structures in order to find the best fit of the residues into the heptad repeats. Residues with {¹H}-¹⁵N heteronuclear NOE values of less than 0.6 were excluded from the fitting procedure.

Modelling of *T. brucei* KKT4^{115–232} and KKT4^{115–645}

Homology models for the two coiled-coil regions of *T. brucei* KKT4^{115–232} were generated using Modeller 9 v24 (Webb and Sali, 2016), the X-ray structure of the *T. cruzi* KKT4^{117–218} coiled coil, and the sequence alignments of *T. brucei* and *T. cruzi* derived from the RDC data collected for KKT4^{115–174}. Modeller was run using the fully automated comparative modelling mode. Random extended structures for the

N- and C-termini of KKT4^{115–232} and for the inter-helix linker were generated using X-PLOR (Brünger, 1992); these represent possible conformations that these residues might sample in solution. These coordinates were merged with the two coiled-coil dimer models to generate an overall model for *T. brucei* KKT4^{115–232}. The model for KKT4^{115–645} was created by merging the model for KKT4^{115–232}, the X-ray structure of the BRCT domain (KKT4^{463–645}), and two random structures for disordered residues 233–473 generated using X-PLOR (Brünger, 1992). The two random structures selected to represent 233–473 were chosen from a group of ten structures to illustrate possible conformations in which the BRCT domain is in close proximity to the N-terminal coiled coil and is more distant from this region.

Interaction of the microtubule-binding and BRCT domains

¹H-¹⁵N BEST TROSY spectra were collected at 20°C for 0.2 mM KKT4^{115–174}, KKT4^{115–232}, KKT4^{145–232} and KKT4^{115–343} alone and in the presence of 0.22 mM KKT4^{463–645}. Following addition of KKT4^{463–645}, the samples were incubated at room temperature for 30 minutes. Interaction was monitored by comparing peak positions and/or peak heights.

Interaction of the BRCT domain with KKT8 phosphopeptide

¹H-¹⁵N BEST TROSY spectra were collected at 20°C for 0.1 mM KKT4^{463–645} alone and in the presence of KKT8 phosphopeptide ranging from 0.025–0.2 mM. Following addition of the peptide, samples were incubated at room temperature for 10 minutes. Interaction was monitored by comparing peak positions and/or peak heights.

2.6. Size Exclusion Chromatography Coupled with Multi-Angle Light Scattering

MALS experiments were performed during size exclusion chromatography on analytical columns equilibrated with 25 mM HEPES pH 7.5, 150 mM NaCl and 0.5 mM TCEP (for SNAP-6HIS-3FLAG-KKT4, 25 mM HEPES pH 7.5, 340 mM NaCl and 0.5 mM was used). Superose 6 HR10/300 was used for SNAP-3FLAG-6HIS-KKT4, SD200 HR10/30 for KKT4^{101–352}, KKT4^{463–645} and KKT4^{2–114}, while SD75 HR10/30 column was used for KKT4^{2–114} 6A mutant and KKT23^{2–348}. Protein elution was monitored via online static light-scattering (DAWN HELEOS 8+, Wyatt Technology), differential refractive index (Optilab T-rEX, Wyatt Technology) and UV (SPD-20A, Shimadzu) detectors. Data were analysed and molecular weight calculated using the ASTRA software package (Wyatt Technology). Concentrations of individual samples loaded on size exclusion columns are mentioned in Table 3.1.

2.7. Analytical Ultracentrifugation

The analytical ultracentrifugation experiments were carried out using a Beckman An-60 Ti rotor in a Beckman Optima XL-1 analytical ultracentrifuge at 20°C. Samples were prepared at concentrations of 0.5–2.0 mg/ml in 25 mM HEPES (pH 7.5), 150 mM NaCl and 0.5 mM TCEP (For the full-length KKT4, 340 mM NaCl was used). Scans were recorded using interference optics. Multiple scans were fitted to a continuous size distribution using SEDFIT (Schuck et al., 2013). The buffer density and viscosity were calculated with SEDNTERP (Hayes D, 1995). The sample partial specific volumes were estimated from the amino acid sequences, also using SEDNTERP.

2.8. Circular Dichroism

Prior to the measurements, protein samples were buffer exchanged using Zeba spin desalting columns (Thermo Fisher) into 25 mM sodium phosphate pH 7.0 and NaF ranging from 20 to 75 mM. Circular dichroism spectra of KKT proteins (at 0.05–0.2 mg/ml) were measured using a Jasco J-815 spectrometer (Jasco Spectroscopic Company, Japan). Experiments were performed at room temperature using a quartz glass cell with a 1 mm path length. The samples were recorded in a far-UV wavelength range (190–250 nm) with a total of three scans for each sample.

2.9. Chemical crosslinking

Prior to the experiment, bis(sulfosuccinimidyl)suberate, BS³ (Thermo Fisher) was equilibrated at room temperature for 2 h and then resuspended to 0.87 mM in distilled water. Immediately after, 2 µl of the crosslinker were mixed with 18 µl of 5–20 µM protein samples in 25 mM HEPES pH 8.0, 250 mM NaCl, 10% glycerol, 2 mM MgCl₂, 0.1 mM EDTA, 0.5 mM EGTA-KOH, 0.1% NP40 and 0.5 mg/ml FLAG peptide. The crosslinking reaction was incubated on ice for 60 min. 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and *N*-Hydroxysulfosuccinimide sodium salt (Sulfo-NHS) were resuspended in distilled water to 4 mM and 10 mM respectively. Immediately after, 1 µl of EDC and 1 µl of Sulfo-NHS were mixed with 18 µl of 5–20 µM protein samples. The crosslinking reaction was incubated at room temperature for 60 min. Following the incubation, all the crosslinking samples were boiled for 10 min and resolved on Nu-PAGE 4–12% gradient polyacrylamide gel (Invitrogen). The gels were stained using SimplyBlue (Invitrogen) and bands corresponding to crosslinked proteins were cut out and subjected to mass spectrometry.

2.10. Mass spectrometry

Mass spectrometry of crosslinked samples

Gel bands were destained by cycles of incubation with 50% acetonitrile in 100 mM triethylammonium bicarbonate (TEAB) and 100% acetonitrile. Subsequently, dried gel bands were incubated with 10 mM TCEP in 100 mM TEAB for 30 min at room temperature followed by centrifugation. Supernatant was removed. Subsequently, 50 mM 2-chloroacetamide in 100 mM TEAB was added to the gel and incubated for 30 min at room temperature in the dark. Gel bands were subsequently washed with 100% acetonitrile 3 times followed by addition of 100 ng of trypsin and digested overnight at 37°C. Subsequently, the supernatant was extracted. The gel pieces were then washed with 5% formic acid and then acetonitrile. Both washes were combined with the initial supernatant and the mixture was then dried down. Peptides were resuspended in 5% formic acid and 5% DMSO and analysed by LC-MS with an Ultimate 3000 UHPLC system (Thermo Fischer Scientific) coupled to an QExactive mass spectrometer (Thermo Fischer Scientific) through an EASY-Spray nanoelectrospray ion source (Thermo Fischer Scientific). The peptides were trapped on a C18 PepMap 100 pre-column (300 mm i.d. x 5 mm, 100Å, Thermo Fisher Scientific) using solvent A (0.1% formic acid, HPLC grade water). The peptides were separated on an in-house packed analytical column (75 mm i.d. x 50 cm packed with ReproSil-Pur 120 C18-AQ, 1.9 mm, 120Å, Dr. Maisch GmbH) using a linear gradient (length: 45 min, 15% to 50% solvent B (acetonitrile with 0.1% formic acid)). Acquisition was performed in data-dependent mode (DDA). Full scan MS spectra were acquired in the Orbitrap (scan range 350–1,500 m/z, resolution 70,000, AGC target 3×10^6 , maximum injection time 50 ms) followed by 10 MS/MS events at 30% NCE (resolution 17 500, AGC target 5×10^4 , maximum injection time 120 ms,

isolation window 1.5 m/z) with first fixed mass at 180 m/z. Charge exclusion was selected for unassigned 1+ and 2+ ions. MS data were converted into mgf format using pParse and searched by the pLink software (Chen et al., 2019) (version 1 for BS³ dataset and version 2 for EDC dataset) using FASTA databases containing KKT1–4, 6, 7–11, 14, 15, 20, and α/β tubulins without KKIP1 (for BS³ dataset) or with KKIP1 (for EDC dataset). Search parameters were as follows: maximum number of missed cleavages = 2, fixed modification = carbamidomethyl-Cys, variable modification 1 = Oxidation-Met, variable modification 2 = Glu to pyro-Glu. Crosslinks that have score $<1 \times 10^{-7}$ were visualised using xiNET (Combe et al., 2015).

Mass spectrometry of histones

To identify the acetylation sites on histones, recombinant histones were subjected to *in vitro* acetylation assay (see section 2.13) and resolved on SDS-PAGE. Gel bands corresponding to individual histones were cut out and in-gel digestion was performed with trypsin. Mass spectrometry was done as follows. Peptides were separated by nano liquid chromatography (Thermo Fisher Scientific Easy-nLC 1000) coupled with a Q Exactive mass spectrometer equipped with an Easy-Spray source (Thermo Fischer Scientific). Peptides were trapped onto a C18 PepMac100 precolumn (300 μm i.d.x5 mm, 100 Å, Thermo Fischer Scientific) using Solvent A (0.1% Formic acid, HPLC grade water). The peptides were further separated onto an Easy-Spray RSLC C18 column (75 μm i.d., 50cm length, Thermo Fischer Scientific) using a 30 minutes linear gradient (15% to 35% solvent B (0.1% formic acid in acetonitrile)) at a flow rate 200 nl/min. The raw data were acquired on the mass spectrometer in a data-dependent acquisition mode (DDA). Full-scan MS spectra were acquired in the Orbitrap (Scan range 350-1500m/z, resolution 70 000; AGC

target, 3×10^6 , maximum injection time, 50 ms). The 10 most intense peaks were selected for higher-energy collision dissociation (HCD) fragmentation at 30% of normalised collision energy. HCD spectra were acquired in the Orbitrap at resolution 17,500, AGC target 5×10^4 , maximum injection time 120 ms with fixed mass at 180 m/z. Charge exclusion was selected for unassigned and 1+ ions. The dynamic exclusion was set to 10s. MS data were converted into mgf format using pParse. Search parameters were as follows: maximum number of missed cleavages = 2, fixed modification = carbamidomethyl-Cys, variable modification 1 = Oxidation-Met, variable modification 2 = Glu to pyro-Glu, variable modification 3 = Lys to Ac-Lys.

2.11. Fluorescence Anisotropy Assay

The DNA-binding analysis of KKT4 and KKT23 was performed in the binding buffer (25 mM HEPES pH 7.5, 50 mM NaCl and 0.5 mM TCEP) using a 50 bp DNA probe BA3098 (~36% GC content), which is part of the *Trypanosoma brucei* centromere CIR147 sequence (Obado et al., 2007), labelled at the 5' end with 6-carboxyfluorescein (6-FAM). Prior to the assay, KKT proteins were buffer-exchanged into the binding buffer using Zeba spin desalting columns (Thermo Fisher). KKT samples were mixed with a DNA probe in the binding buffer to a final DNA concentration of 1 nM. Next, the proteins were serially diluted in the binding buffer containing 1 nM DNA in the 2:3 v/v ratio. The starting protein concentrations were the following: KKT4^{2–645} (0.7 μ M), KKT4^{115–645} (4 μ M), KKT4^{250–645} (10 μ M), KKT4^{343–645} (17.0 μ M), KKT4^{2–114} (1 μ M), KKT4^{463–645} (1 μ M), KKT4^{114–232} (10.4 μ M), KKT^{300–488} (1 μ M), KKT4^{300–488} mutants (1 μ M) and KKT23^{2–348} (2 μ M). The binding reactions were incubated for 30 min at room temperature and the fluorescence anisotropy was measured using a PHERAstar FS next-generation microplate reader (BMG LABTECH). Equilibrium dissociation constants (K_D) were calculated by fitting the data in SigmaPlot (Monks, 2002). The phosphopeptide-binding experiments

were done as described above using fluorescently-labelled phosphorylated (DICGISGQ(pS)VRYSLND) and unphosphorylated (DICGISGQSVRYSLND) peptide probes. KKT4^{463–645} (170 μ M) was mixed with the probe (100 nM) and serially diluted in a 2:3 v/v ratio. Incubation of the samples and measurements were done as described above.

2.12. Co-sedimentation assay

For the co-sedimentation assays, KKT fragments were buffer exchanged into the BRB80 (80 mM Pipes-KOH pH 6.9, 1 mM EGTA, and 1 mM MgCl_2) with 100 mM KCl using Zeba spin desalting column (ThermoFisher). Taxol-stabilised microtubules were prepared by mixing 2.5 μ l of 100 μ M porcine tubulin (Cytoskeleton) resuspended in BRB80 with 1 mM GTP (Cytoskeleton), 1.25 μ l BRB80, 0.5 μ l of 40 mM MgCl_2 , 0.5 μ l of 10 mM GTP, and 0.25 μ l DMSO and incubated for 20 min at 37°C. Then, 120 μ l of BRB80 containing 12.5 μ M Taxol (paclitaxel; Sigma) was added to the sample to achieve 2 μ M microtubules. To achieve higher concentrations of microtubules, the protocol described above was scaled up. For the assay, 20 μ l of protein (4 μ M) were mixed with 20 μ l of microtubules (2 μ M) and incubated for 45 min at room temperature. For a no-microtubule control, KKT4 fragments were mixed with BRB80 that contained 12.5 μ M Taxol. The samples were spun at 20 000 g at room temperature for 10 min, and the supernatant was collected. To the tube with a pellet, 40 μ l of chilled BRB80 was added with 5 mM CaCl_2 and incubated on ice for 5 min to depolymerise microtubules. Following the incubation, samples were boiled for 5 min before analysis by SDS-PAGE gels stained with InstantBlue (Invitrogen).

2.13. Acetylation assay

Prior to the assay, KKT23^{2–348} and KKT23^{125–348} were buffer exchanged into the HAT buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol, 1 mM DTT, 0.1 mM EDTA and 10 mM sodium butyrate), diluted to 1 mg/ml and mixed with acetyl coenzyme A sodium salt (Sigma Aldrich) up to 2 mM concentration. Mutants of KKT23^{125–348} eluted in 50 mM sodium phosphate pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole and 0.5 mM TCEP were diluted with HAT buffer in 1:1 v/v ratio. Histones (H3/H2A/H2B/H4 and H3/H2AZ/H2BV/H4) used as substrates in the acetylation assay were buffer exchanged to the HAT buffer using size exclusion chromatography. Histones H2A and H2AZ used in the assay were in 50 mM sodium phosphate pH 7.5, 500 mM NaCl, 10% glycerol, 250 mM imidazole and 0.5 mM TCEP. For the acetylation assay, 20 µl of the enzyme was mixed with 20 µl of the substrate sample and incubated at 30°C for 1 h. Following the incubation, samples were boiled for 3 min and resolved using SDS-PAGE electrophoresis, after which the samples were analysed by Western Blot using pan-acetyl lysine antibodies (Cell Signaling). To access the sample loading, samples were resolved using NuPage 4–12% gradient gels (Invitrogen) and stained with SimplyBlue (Invitrogen)

2.14. Western Blot analysis

For the Western Blot analysis, SDS-PAGE gels were transferred onto nitrocellulose membranes, blocked with 5% skimmed milk in PBST and incubated with primary pan-acetyl lysine antibodies overnight at 4°C (1:1000 dilution v/v in PBST, Cell Signaling). Following the incubation, the membrane was washed 3 times with PBST and incubated with anti-rabbit horseradish peroxidase-linked secondary antibodies for 2 h at RT (at 1:1000 v/v dilution in PBST). Acetylated proteins were detected by chemiluminescence using detection reagent 1 and 2 (Thermo Scientific).

2.15. Trypanosome cells

Trypanosome cell lines (listed in Table 2.14) used throughout this thesis were derived from *T. brucei* SmOxP927 procyclic form cells (TREU 927/4 expressing T7 RNA polymerase and the tetracycline repressor to allow inducible expression; (Poon et al., 2012)). *Trypanosoma brucei* cell cultures were grown at 28°C in SDM-79 medium supplemented with 10% (vol/vol) heat-inactivated fetal calf serum (Brun and Schonenberger, 1979). Cell growth was monitored using a CASY cell counter and analyser system (Roche). Endogenous N-terminal YFP tagging was performed as described in (Llauro et al., 2018). Expression of RNAi-resistant forms of KKT4 was facilitated by endogenous C-terminal YFP-tagging constructs made from pEnT5-Y vector (Kelly et al., 2007). Plasmids linearised by NotI were transfected to trypanosomes by electroporation into an endogenous locus (pEnT5-Y and pBA148 derivatives) or 177-bp repeats on minichromosomes (pBA310 derivatives). Transfected cells were selected by the addition of 25 µg/ml hygromycin, 10 µg/ml blasticidin, or 5 µg/ml phleomycin. The RNAi was induced with doxycycline at a final concentration of 0.1 µg/ml.

2.16. Microscopy of trypanosome cells

Prior to imaging, cells were fixed with formaldehyde as previously described (Nerusheva and Akiyoshi, 2016). Images were captured at room temperature on a DeltaVision fluorescence microscope (Applied Precision) installed with softWoRx version 5.5 housed in the Oxford Micron facility (<https://www.micron.ox.ac.uk/home.php>). Fluorescent images were captured with a CoolSNAP HQ camera using 60× objective lenses (1.42 NA), typically 16–21 z sections at 0.2- to 0.25-µm steps. Images were generated in ImageJ (Schneider et al., 2012).

Table 2.14. List of *T. brucei* cells lines used in this work.

Name	Description	Resistance
BAP1669	KKT4 ^{K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D}	Hygromycin, Puromycin
BAP1734	KKT4 ^{V8A, L9A, L12A, I19A, L26A, L29A}	Hygromycin, Puromycin
BAP1735	KKT4 ^{Δ300–488}	Hygromycin, Puromycin
BAP1775	RNAi, KKT4 ^{V8A, L9A, L12A, I19A, L26A, L29A}	Hygromycin, Puromycin, Phleomycin
BAP1776	RNAi, KKT4 ^{Δ300–488}	Hygromycin, Puromycin, Phleomycin
BAP1785	KKT4 ^{K310E, R311D, R313D}	Hygromycin, Puromycin
BAP1800	RNAi, KKT4 ^{K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D}	Hygromycin, Puromycin, Phleomycin
BAP1801	RNAi, KKT4 ^{K310E, R311D, R313D}	Hygromycin, Puromycin, Phleomycin

CHAPTER 3

3. Structure and function of KKT4 microtubule-binding domain

Kinetochores-based microtubule-binding proteins play a vital role in the kinetochore function and include Ndc80, Ska and Dam1 complexes, SKAP/Astrin, CENP-E, CENP-F, MCAK, INCENP, XMAP215, and dyneins (Maiato et al., 2004, Foley and Kapoor, 2013, Musacchio and Desai, 2017). In human kinetochores, microtubule attachment is dependent on the KMN network, which is composed of three subcomplexes: Ndc80, Mis12 and Knl1 complex (Cheeseman et al., 2006). A key player in the KMN network is the heterotetrameric Ndc80 complex that interacts with microtubules through a calponin homology domain (CH domain) (Cheeseman and Desai, 2008, Musacchio and Desai, 2017). A weak similarity to Ndc80 based on the coiled coil region was found in KKIP1, a structural component of kinetoplastid kinetochores. However, KKIP1 apparently lacks the CH domain and no evidence of microtubule-binding activity in KKIP1 has been reported (Brusini et al., 2019). Therefore, it is unlikely that KKIP1 is an Ndc80 homolog in kinetoplastids.

To date, KKT4 is the only known microtubule-binding component of kinetoplastid kinetochores. It has been shown to bind microtubules directly, track depolymerising microtubule tips and maintain load-bearing attachments to both growing and shortening microtubule tips (Llauro et al., 2018). Therefore, KKT4 has all functions required for a kinetochore microtubule-binding protein. Interestingly, sequence analysis of KKT4 has not revealed any significant similarity to Ndc80 or other known kinetochore or microtubule-binding protein (Akiyoshi and Gull, 2014, Llauro et al., 2018). Therefore, the structure and mechanism of how KKT4 interacts with microtubules remain unknown. To address this, I carried out extensive structural studies using NMR and X-ray crystallography. Throughout this thesis, KKT4 refers to *Trypanosoma brucei* KKT4 unless stated otherwise.

3.1. KKT4 forms oligomers

To date, no information on the oligomerisation state of KKT4 has been reported. In this section, I aim to investigate it using a combination of biochemistry and cell biology.

KKT4 can form tetramers *in vitro*

Previous single-molecule experiments suggested that KKT4 tends to oligomerise even at low nanomolar concentrations (Llauro et al., 2018). To test the ability of KKT4 to oligomerise *in vitro*, Size Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) was employed (Wen et al., 1996). SEC-MALS analysis of the full-length SNAP-3FLAG-6HIS-KKT4^{2–645} purified from insect cells revealed that KKT4 can form tetramers in solution at high protein concentrations (20 μ M) and dimers at low concentrations (4 μ M) (Figure 3.1A). The oligomerisation of KKT4 is concentration-dependant.

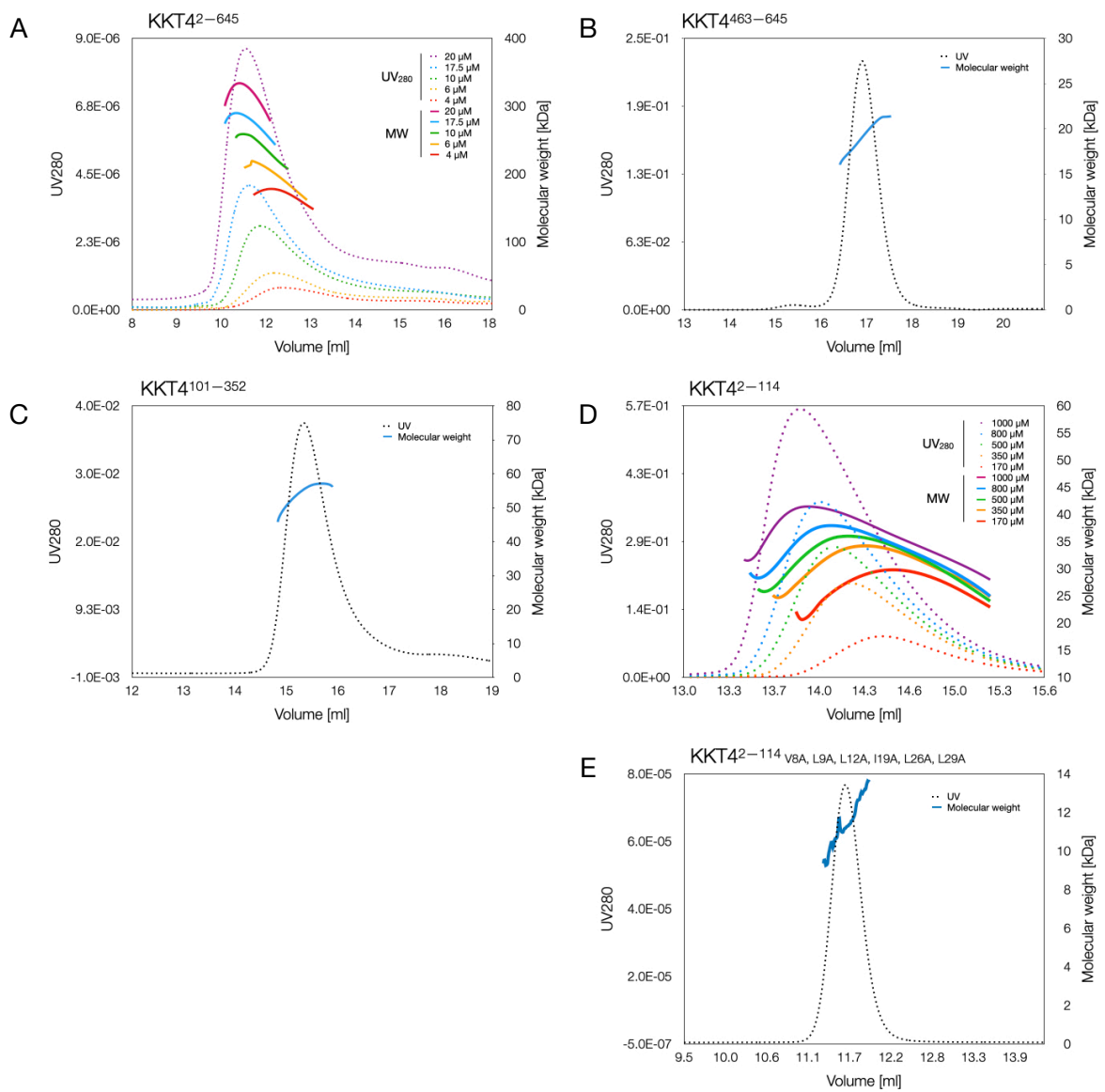


Figure 3.1. SEC-MALS analysis of KKT4 fragments

SEC-MALS elution profiles of (A) SNAP-3FLAG-6HIS-KKT4^{2–645}, (B) KKT4^{463–645}, (C) KKT4^{101–352}, (D) KKT4^{2–114} and (E) KKT4^{2–114} V8A, L9A, L12A, I19A, L26A, L29A. The UV280 signal (dotted line) is plotted against the elution volume, while the molecular weight is indicated as a thick line for each peak. Concentrations of samples and the calculated molecular weights for each construct are reported in Table 3.1.

Oligomerisation domain of KKT4

Based on the observation that the full-length KKT4 can oligomerise, a further study was carried out to identify the fragments responsible for its oligomerisation activity. The following structural regions of KKT4 are well conserved among kinetoplastids: N-terminal predicted α -helices, predicted coiled coils within the microtubule-binding domain, a block of basic residues, and the C-terminal BRCT domain (Llauro et al., 2018) (Figure 1.6 and 1.7). Several fragments of KKT4 were analysed using SEC-MALS and AUC, showing that the C-terminal BRCT domain of KKT4 (KKT4^{463–645}) was a monomer (Figures 3.1B and 3.2A), while the fragment able to interact with microtubules (KKT4^{101–352}) formed dimers (Figures 3.1C and 3.2B). In contrast, the N-terminal region of KKT4, KKT4^{2–114} was shown to form tetramers at high protein concentrations and trimers at lower concentrations (Figures 3.1D and 3.2C), similar to the full-length protein. Based on the fact that the full-length KKT4 and KKT4^{2–114} can form tetramers at high concentrations and dimers or trimers at low concentrations, the observed trimeric species likely represent an equilibrium mixture of tetramers and dimers rather than trimers themselves. The summary of all the oligomerisation experiments are listed in Tables 3.1.

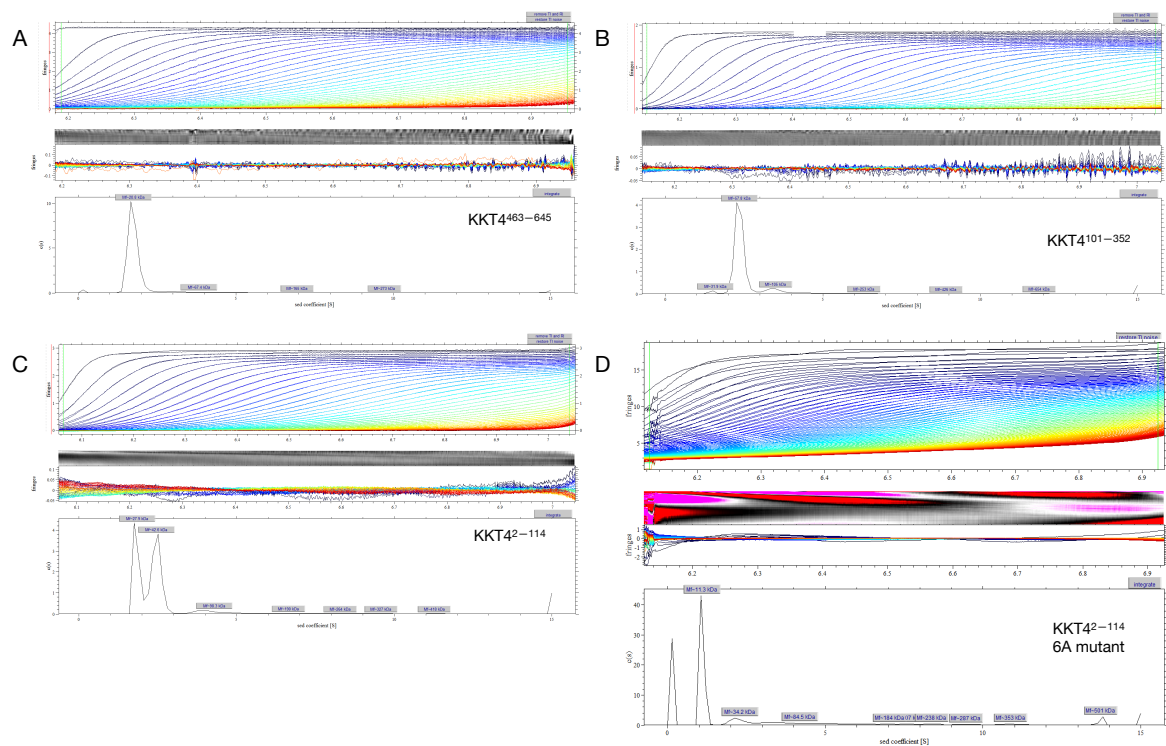


Figure 3.2. AUC analysis of KKT4 fragments.

The outputs of the AUC experiments were analysed using SEDFIT (Schuck et al., 2013) for (A) KKT4^{463–645}, (B) KKT4^{101–352}, (C) KKT4^{2–114} and (D) KKT4^{2–114} V8A, L9A, L12A, I19A, L26A, L29A. The sample concentration and the calculated molecular weight for each construct are reported in Table 3.1.

Table 3.1. Summary of SEC-MALS and AUC experiments.

KKT4 fragment	Calculated MW [kDa]	Concentration [μ M]	Observed MW [kDa]	Oligomerisation state
SEC-MALS				
KKT4 ^{2–645}	92	20/17/10/6/4	334/275/240/195/170	tetramer/dimer
KKT4 ^{463–645}	20	100	19	monomer
KKT4 ^{101–352}	28	35	55	dimer
KKT4 ^{2–114}	11	1000/800/500/350/170	38/35/32/33/28	tetramer/ trimer or tetramer/dimer
KKT4 ^{2–114} V8A, L9A, L12A, I19A, L26A and L29A	11	174	11	monomer
AUC				
KKT4 ^{463–645}	20	100	21	monomer
KKT4 ^{101–352}	28	35	58	dimer
KKT4 ^{2–114}	11	60	43 and 28	tetramer/ trimer
KKT4 ^{2–114} V8A, L9A, L12A, I19A, L26A and L29A	11	60	11	monomer

In vitro analysis of the oligomerisation-defective mutants

Tetramerisation activity of KKT4 resides within residues 2 to 114. Analysis of this fragment's primary sequence predicts a conserved helix at positions 5–29, while the remaining region (30–114) is predicted to be disordered and is not well conserved among kinetoplastids (Figure 1.6B, 1.7). To test the possibility that the helical region in the N-terminus facilitates the tetramerisation activity of KKT4, via formation of a helix bundle, several mutations (V8A, L9A, L12A, I19A, L26A and L29A) were introduced into KKT4^{2–114}, which I refer to as a 6A mutant. Analysis of

the SEC-MALS and AUC experiments showed that the 6A mutant could not oligomerise (Figures 3.1E and 3.2D). The circular dichroism spectrum of the mutant showed no features that are characteristic of any secondary structure elements, whereas the wild-type KKT4^{2–114} showed negative ellipticity with peaks at 208 and 222 nm, which is typical for α -helical structures (Figure 3.3). The introduced mutations likely destabilised the N-terminal helix and resulted in an oligomerisation defect. Taken together, these results confirm the importance of the N-terminal helical region for the tetramerisation activity of KKT4^{2–114}.

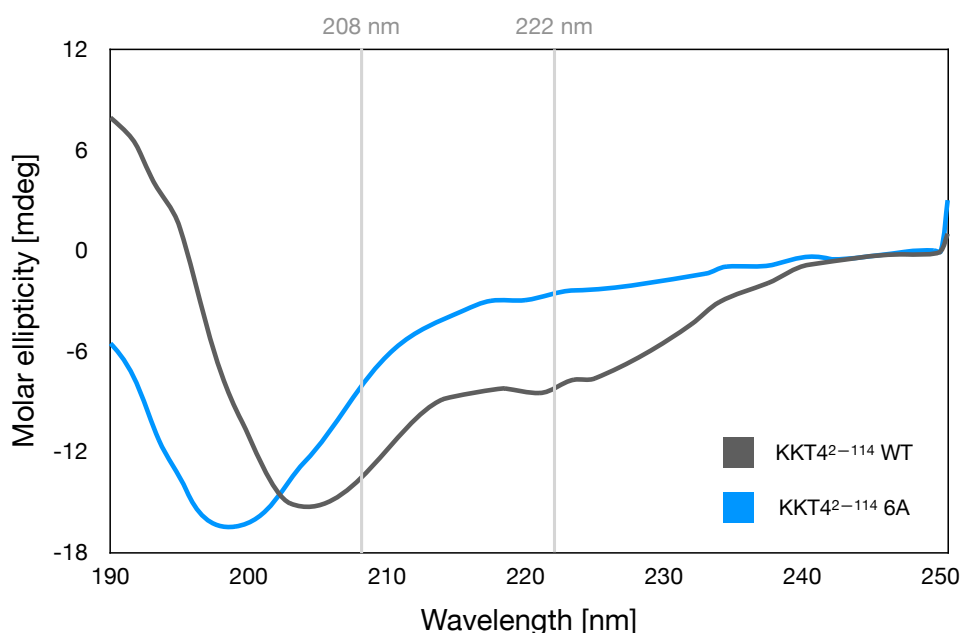


Figure 3.3. Circular dichroism spectra for KKT4^{2–114} and 6A mutant.

The spectrum of the wild-type KKT4^{2–114} (grey) shows negative ellipticity around 208 and 222 nm, typical for helical structures. These peaks are not visible in the shifted spectrum of KKT4^{2–114} V8A, L9A, L12A, I19A, L26A, L29A (blue), suggesting that the introduced mutations disrupted the secondary structure of the KKT4 fragment.

The N-terminal region of KKT4 is mainly disordered

KKT4^{42–114} is predicted to be mainly disordered with a short helical region within residues 5–29 (Figure 1.6B, C). Such a protein fragment is not suitable for crystallographic studies as the disordered region is likely to prevent crystal formation. Trials to purify a shorter, soluble KKT4 fragment composed of just the N-terminal helix (KKT4^{2–30}) were unsuccessful. Instead, to obtain structural information on KKT4^{42–114}, I employed NMR spectroscopy. The 2D ¹H-¹⁵N BTROSY spectrum showed a low level of dispersion and high-intensity peaks; these indicate a significant level of disorder present in KKT4^{42–114} (Figure 3.4). To confirm this, I probed the dynamics of KKT4^{42–114}, using the {¹H}-¹⁵N heteronuclear NOE, which is sensitive to backbone motions on a timescale (ps) faster than the overall tumbling of the molecule (ns). A {¹H}-¹⁵N NOE ratio of greater than ~0.6 is characteristic of a structured backbone, while values below ~0.6 indicate higher flexibility (Kay et al., 1989). Indeed, all visible peaks in KKT4^{42–114} spectrum displayed hetNOE ratios below 0.6, confirming they are disordered (Figure 3.5).

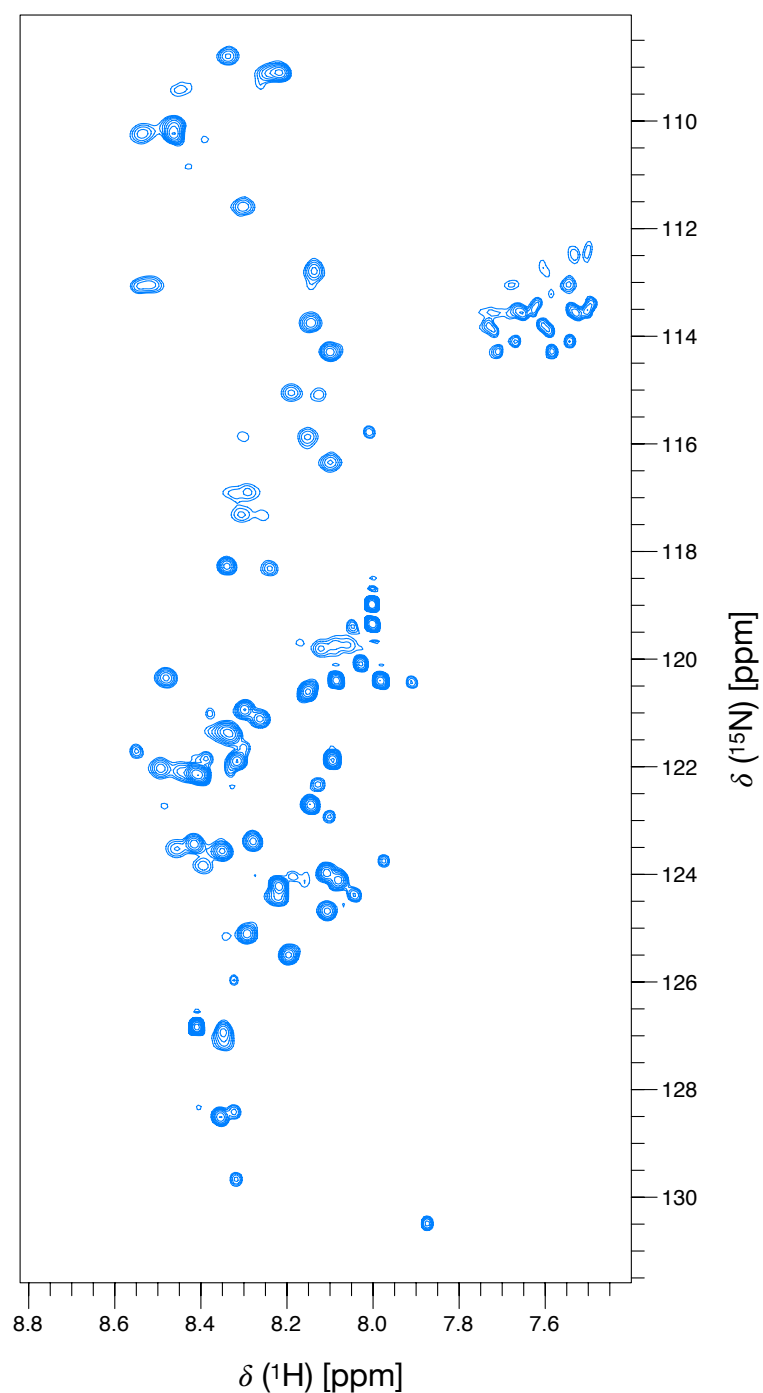


Figure 3.4. The ^1H – ^{15}N BEST-TROSY spectrum of KKT4^{2–114}.

750 MHz ^1H – ^{15}N BEST-TROSY spectrum of KKT4^{2–114} in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 30°C. Most peaks are located between 7.9–8.5 ppm, typically observed for unstructured protein regions. Peaks in the region of 112–114 ppm and upfield of ~7.6 ppm are artefacts in the BEST-TROSY arising from incomplete cancellation of signals from the side chain amides of Asn and Gln.

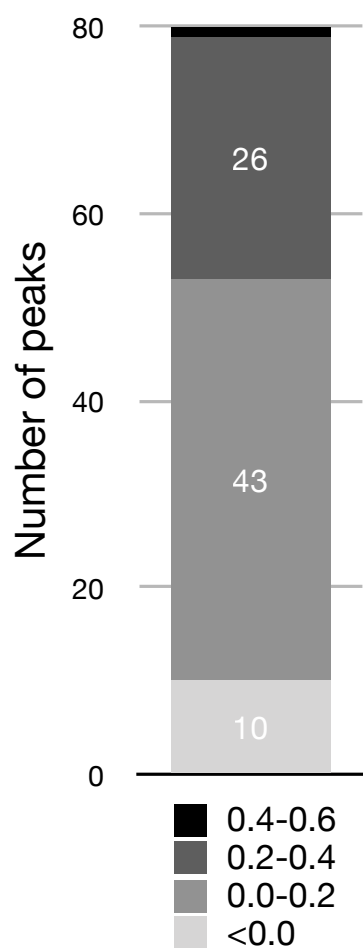


Figure 3.5. The $\{^1\text{H}\}$ - ^{15}N Heteronuclear NOE analysis of KKT4²⁻¹¹⁴.

The histogram shows the distribution of $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE ratios measured for peaks in the spectrum of KKT4²⁻¹¹⁴. The heteronuclear NOE ratios were segregated and colour coded by different shades of grey according to different values observed. All peaks displayed values <0.6 suggesting they are highly flexible, therefore, disordered. The peaks with hetNOE ratios between 0.2–0.4 most likely correspond to the region where the protein becomes less flexible; these peaks might correspond to residues that follow the predicted N-terminal helix.

The ^1H - ^{15}N BTROSY spectrum of KKT4²⁻¹¹⁴ showed ~80 peaks, which is lower number than expected for a 103 non-proline residues of the protein. The heteronuclear NOE data show that the observed peaks all correspond to residues in flexible regions of the structure. It is likely that 23 missing peaks correspond to the predicted helical region (5–29); it is possible that in the tetramer this region undergoes microsecond to millisecond motions that lead to peak broadening or that this structured region tumbles in an anisotropic manner that leads to unfavourable line widths. The natural abundance (i.e. no ^{15}N labelling) BTROSY spectrum of the 6A mutant was compared to the wild-type KKT4 (Figure 3.6). The majority of peaks align well between the wild-type and the mutant KKT4 spectra, suggesting that a significant number of residues are disordered in both constructs and that their properties were not affected by destabilising the N-terminal helix. Additional peaks

can be observed in the spectrum of 6A mutant that are not visible in the wild-type spectrum. These peaks likely arise from the N-terminal region that is disordered in the 6A mutant. In summary, the combination of circular dichroism, NMR data and secondary structure predictions suggest that KKT4^{2–114} is mainly disordered with a small helical region near the N-terminus. Detailed resonance assignments and the analysis of secondary structure using chemical shifts of KKT4^{2–114} could be done in the future to confirm these observations.

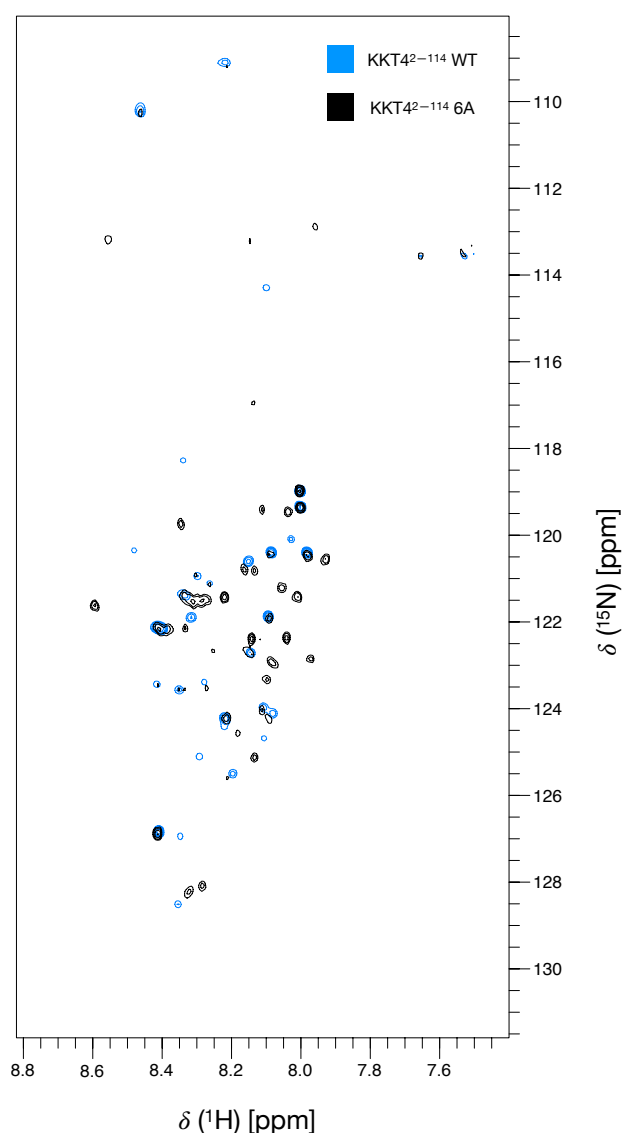


Figure 3.6. The natural abundance ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{2–114} 6A mutant.

The natural abundance 750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{2–114} 6A mutant (black) is overlaid with the spectrum of the wild-type KKT4^{2–114} (blue) from Figure 3.4. Both spectra were collected in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 30°C. The spectrum of the wild-type KKT4^{2–114} was contoured to a similar level of signal-to-noise ratio compared to the spectrum of 6A mutant. Some of the peaks observed in the spectrum of 6A mutant are identical to peaks in the spectrum of the wild-type KKT4^{2–114}. However, there are also peaks that are not visible in the wild-type spectrum. These peaks likely arise from the N-terminal region that is not structured in the mutated KKT4^{2–114}.

In vivo analysis of the oligomerisation-defective mutants

Analysis of SEC-MALS and AUC experiments showed that the KKT4^{2–114} 6A mutant (V8A, L9A, L12A, I19A, L26A, L29A) was not able to form tetramers *in vitro* due to the loss of its secondary structure (Figure 3.3). To test the relevance of these mutations *in vivo*, a *Trypanosoma brucei* cell line was generated, where the mutations mentioned above were introduced into one of the two KKT4 alleles. This allowed for phenotype analysis upon the silencing of the wild-type KKT4 using the RNAi machinery in trypanosomes. The expression of KKT4 6A mutant resulted in a growth defect upon the RNAi induction (Figure 3.7A). However, it has been observed that the signal intensity of YFP-KKT4 6A was lower compared to wild-type KKT4 (Figure 3.7B). It therefore cannot be excluded that the observed growth defect for the 6A mutant is due to its lower expression level rather than inability to form tetramers.

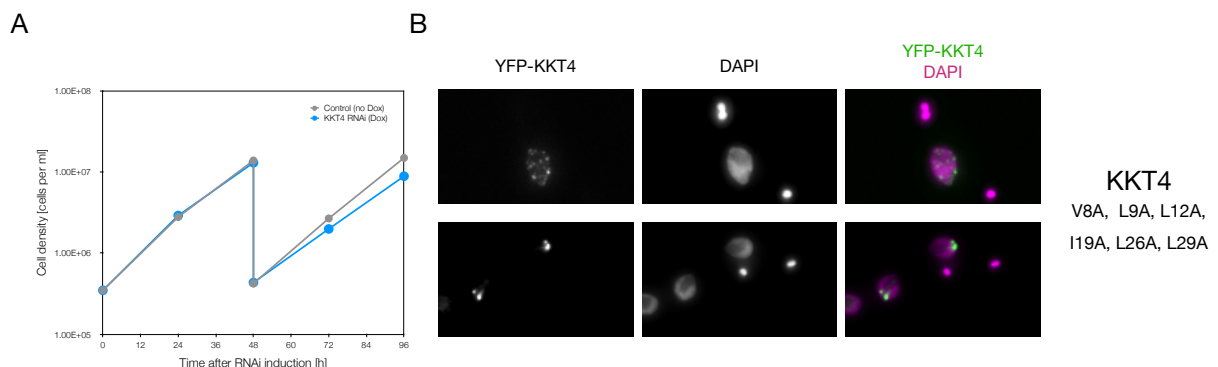


Figure 3.7. In vivo analysis of KKT4^{V8A, L9A, L12A, I19A, L26A, L29A}.

- (A) Graph represents the growth rate of *Trypanosoma brucei* cells expressing the KKT4^{V8A, L9A, L12A, I19A, L26A, L29A}, which was monitored over 96h post the RNAi induction. Cells were diluted after 48h. Expression of the KKT4 6A mutant doesn't rescue the KKT4 3' UTR-targeting RNAi phenotype, suggesting that the mutated residues are important for the proliferation of *T. brucei* cells.
- (B) Fluorescence microscopy images showing the localisation of the KKT4^{V8A, L9A, L12A, I19A, L26A, L29A} in G1 cell (top) and anaphase cell (bottom). The 6A mutant localises to the kinetochore but its signal intensity is lower compared to the wild-type KKT4 (Figure 1.5). The lower expression level is likely an effect of the growth defect observed in (A).

The data in this section show that KKT4 is an oligomeric protein that is able to form tetramers and dimers in a concentration-dependent manner. Two shorter KKT4 fragments, close to the N-terminus KKT4^{2–114} and the microtubule-binding region (KKT4^{101–352}) formed tetramers and dimers, respectively. Residues important for the oligomerisation of KKT4^{2–114} *in vitro* were essential for the growth of *Trypanosoma brucei* cells. Finally, structural data were obtained on KKT4^{2–114}, showing it is mainly disordered.

3.2. NMR analysis of the microtubule-binding domain structure and dynamics

To gain structural insight into the microtubule-binding domain of KKT4, crystallisation trials of KKT4^{115–343} were carried out but turned out to be unsuccessful. This can be explained by the presence of a long (>100 residues) unstructured tail, predicted near the C-terminus of KKT4^{115–343} (Figure 1.6C), which is likely to prevent the crystal formation. Instead, solution-state NMR spectroscopy to probe the structure and dynamics of the KKT4 microtubule-binding domain was employed.

Resonance assignments of KKT4 microtubule-binding domain

Protein NMR studies typically start with the analysis of a standard 2D ¹H-¹⁵N heteronuclear single quantum coherence (HSQC) spectrum, which requires isotopically labelled protein samples. Hence, ¹⁵N-KKT4^{115–343} was expressed and purified from *E. coli*. The ¹H-¹⁵N HSQC spectrum of KKT4^{115–343} showed less than the expected number of peaks for the 218 non-proline residues of the protein and the observed peaks were not well resolved (data not shown). To improve the spectrum quality, the ¹H-¹⁵N BEST TROSY spectrum of KKT4^{115–343} was collected. BEST TROSY experiments are used to achieve shorter inter-scan delays (0.2 instead of >1s) and record only a fraction of the full points, which results in an improvement of resolution and signal-to-noise ratio compared to the standard HSQC experiments. The ¹H-¹⁵N BEST TROSY spectrum of KKT4^{115–343} showed significant improvement of the resolution and approximately 160 peaks were observed with a large variation of intensities (Figure 3.8). The variation of peak intensities suggests a mixture of structured and unstructured regions within KKT4^{115–343} and/or differences in the backbone dynamics. Backbone and side-chain resonance assignments for the

strongest peaks were achieved using a combination of 3D triple resonance experiments, which are listed in Table S1 in the Appendix, using ^{15}N - ^{13}C double labelled protein. An example of how peaks were linked together using the $\text{C}\alpha$ and $\text{C}\beta$ atoms, and then assigned to specific residues in the protein sequence is shown in Figure 3.9.

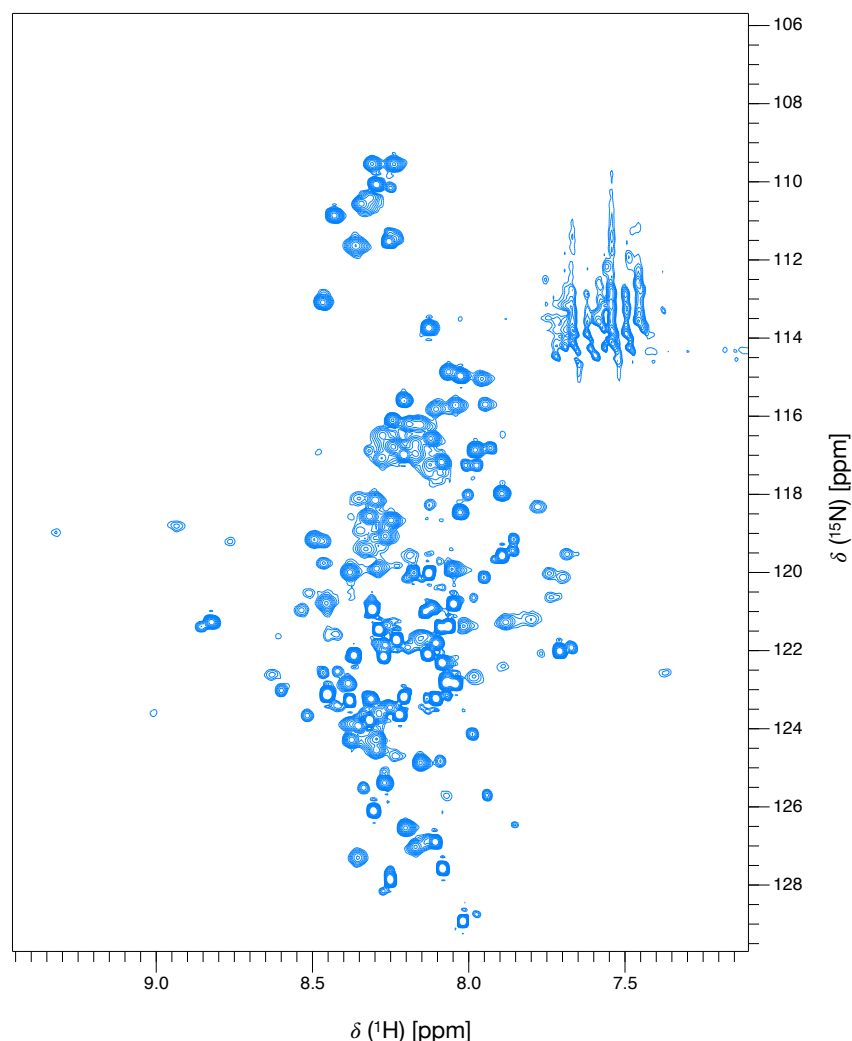


Figure 3.8. The ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–343}.

750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–343} in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 20°C. This spectrum is contoured at a low level to highlight the range of peak intensities observed for KKT4^{115–343}. The differences in intensity suggest that the KKT4^{115–343} consists of structured and unstructured regions.

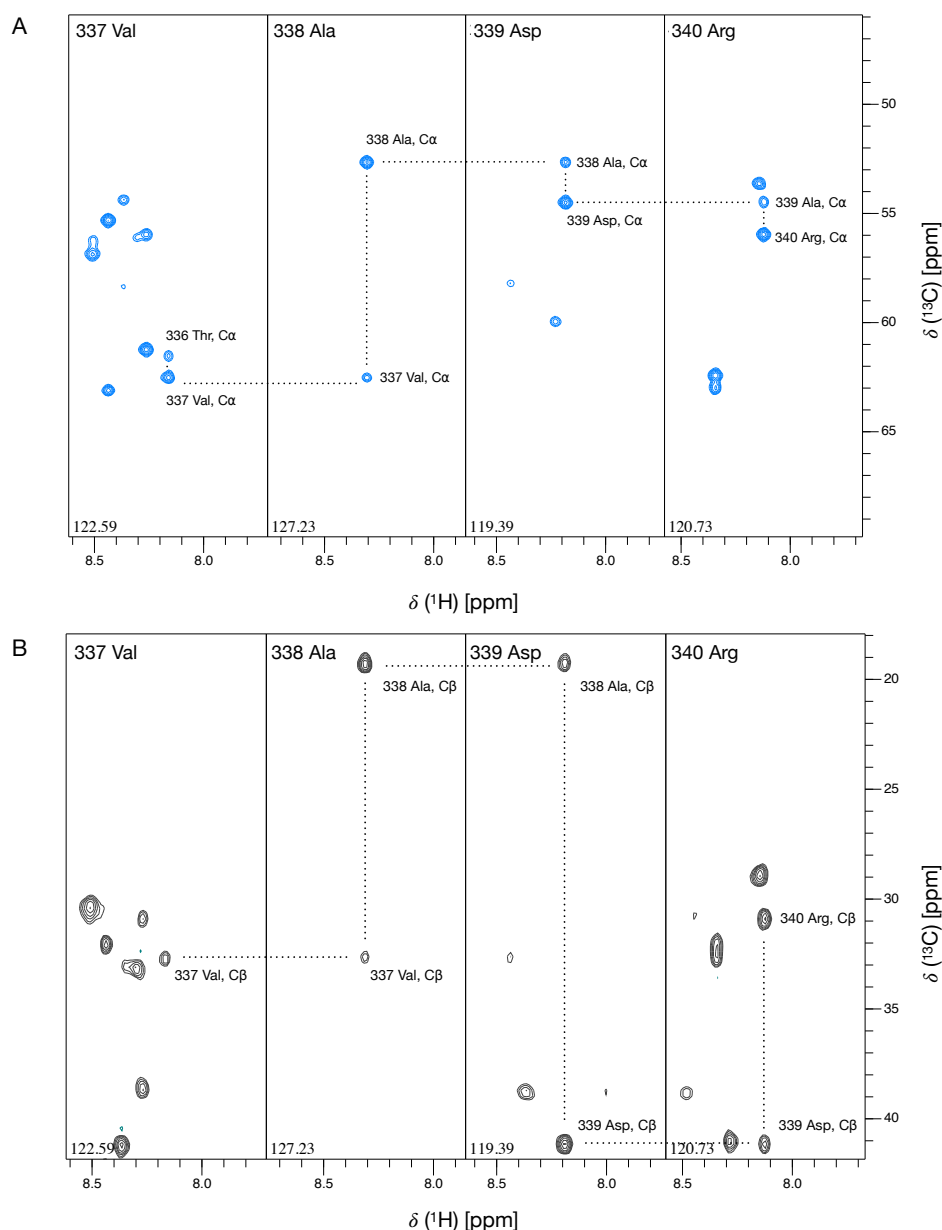


Figure 3.9. Examples of the NMR triple resonance experiments used to assign protein backbone.

- (A) Four strips from the HNCA experiment are shown for peaks that were assigned to residues 337–340 in KKT4^{115–343}. The HNCA gives 2 peaks in the centre of each strip, one for its own $\text{C}\alpha$ (residue i) and one for $\text{C}\alpha$ of the preceding residue (residue $i-1$). The peaks that link the neighbouring residues are connected with dotted lines.
- (B) The CBCANH experiment correlates $\text{C}\alpha$ and $\text{C}\beta$ of residues i and $i-1$. This means that in the centre of each strip, 4 peaks should be observable (2 for $\text{C}\alpha$ and 2 for $\text{C}\beta$). The figure is limited to the region between 20–40 ppm to show the position of $\text{C}\beta$ peaks as $\text{C}\alpha$ peaks are shown in (A). The $\text{C}\beta$ peaks are used to determine the type of amino acids. The CBCANH in combination with the HNCA (A) allows to link neighbouring residues, determine the type of amino acids and assign peaks to specific region in the protein sequence. X and Y axis in (A) and (B) represent ^1H and ^{13}C chemical shifts. The ^{15}N chemical shift (ppm) for each residue is denoted at the bottom of the strip.

Assignment of the ~90 strongest peaks in the spectrum has shown that these peaks correspond to residues between K115 and V118 near the N-terminus and those between N231 and E343, in the C-terminal domain region (Figure 3.10). In this region of the sequence, 87.4% of $^1\text{H}^{\text{N}}$ and 88.5% of ^{15}N were assigned. No correlations in triple-resonance experiments linking neighbouring residues were observed for ~30 weakest peaks in KKT4^{115–343} spectrum; the assignment of these peaks was not possible in this construct. The rest of the missing $^1\text{H}^{\text{N}}/^{15}\text{N}$ assignments correspond to either Asn/Ser/Thr residues in the disordered part of the protein with unfavourable solvent exchange rates or residues found in regions of repetitive sequence such as TTTSS. The chemical shifts for $\text{C}\alpha$ and $\text{C}\beta$ atoms in such regions are almost indistinguishable. In addition to the backbone assignments, partial side-chain ^1H and ^{13}C assignments were obtained. Detailed assignment statistics for KKT4^{115–343} can be found in Table 3.2. In summary, resonance assignments have been obtained for 115–118 and 231–343, while no peaks have been assigned to residues 119–230 in KKT4^{115–343}.

Table 3.2. Extent of assignment for KKT4 fragments.

Percent assigned						
Sample	$^1\text{H}^{\text{N}}/^{15}\text{N}$	^{13}C	$^1\text{H}\alpha/^{13}\text{C}\alpha$	$^1\text{H}\beta/^{13}\text{C}\beta$	$^1\text{H}\gamma/^{13}\text{C}\gamma$	$^1\text{H}\delta/^{13}\text{C}\delta$
KKT4 ^{115–343}	87.4/88.5	87.6	91.8/91.2	87.4/94.2	-/64.3	-/50.0
KKT4 ^{115–174}	100/98.3	90.0	100/100	68.0/87.5	38.3/50.0	43.8/45.5
KKT4 ^{145–232}	96.6/95.5	96.6	-/98.9	-/89.5	-/-	-/-

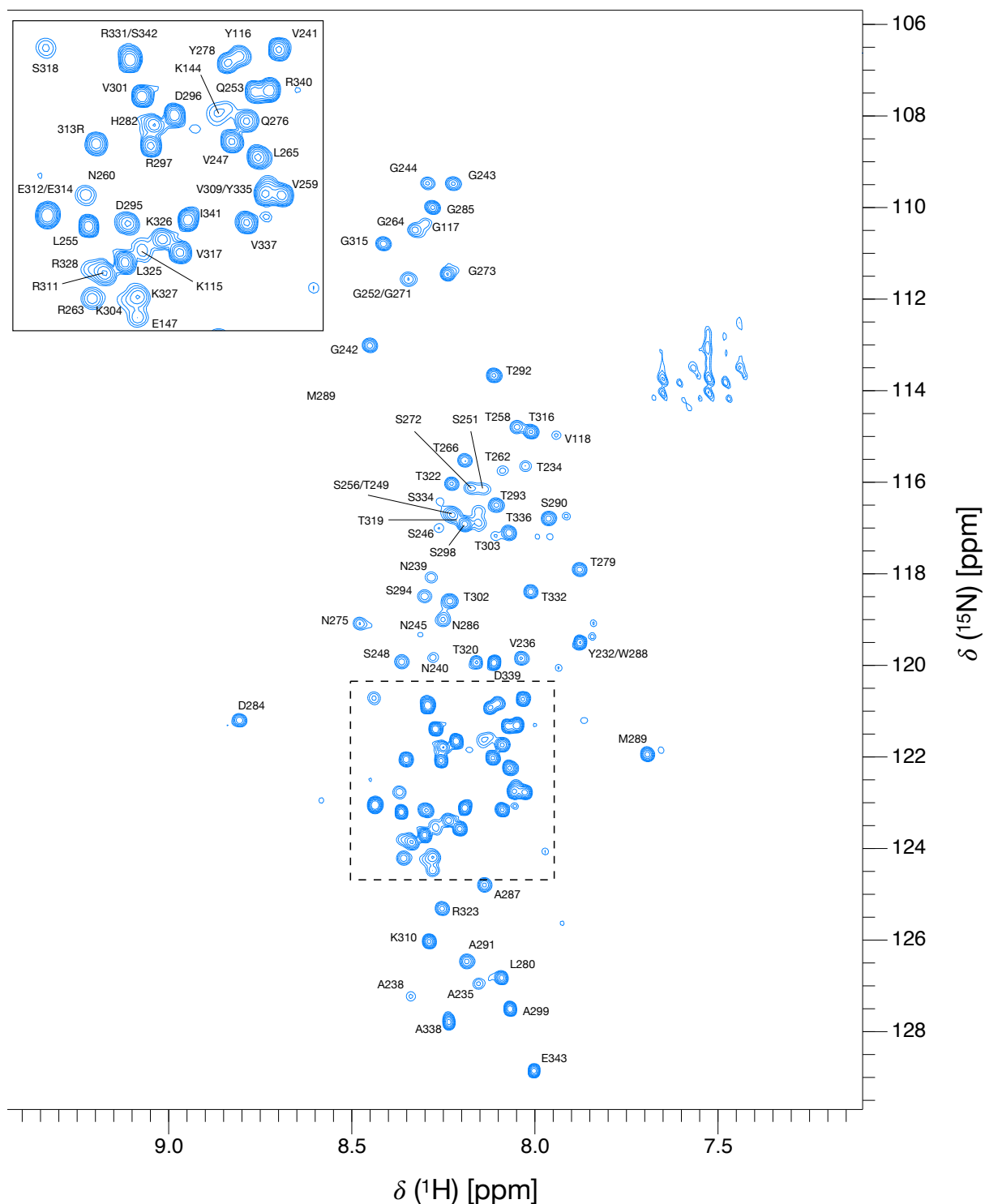


Figure 3.10. The ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–343}.

750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–343} in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 20°C. The peaks for backbone amides of the C-terminal region (residues 231–343) of KKT4^{115–343} are assigned and annotated. The inset is an expansion of the area outlined with dashed lines. Peaks in the region of 111–114 ppm and upfield of ~7.6 ppm are artefacts in the BEST-TROSY arising from incomplete cancellation of signals from the side chain amides of Asn and Gln which have not been assigned.

The C-terminal part of the KKT4 microtubule-binding domain is disordered

Chemical shifts of the protein backbone including ^1H , ^{13}C and ^{15}N , can be used to indicate the secondary structure elements in the studied protein (Spera and Bax, 1991, Wishart et al., 1991, Beger and Bolton, 1997, Cornilescu et al., 1999). TALOS-N is an NMR software that can make quantitative predictions for the secondary structure, protein backbone angles ϕ/ψ and sidechain χ^1 using a high-resolution structural database and chemical shifts known for these structures. Analysis of $^1\text{H}\alpha$, ^1HN , $^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO and ^{15}N chemical shifts using TALOS-N (Shen and Bax, 2013) predicted no secondary structure elements for residues N231 to E343 in KKT4^{115–343} (data not shown). Chemical shifts for residues N231 to E343 were also analysed using a more suitable software for analysing disordered proteins, the Secondary Structure Propensity (SSP) score (Marsh et al., 2006); no significant secondary structure propensity was found (Figure 3.11A), suggesting that the C-terminal region of KKT4^{115–343} from residue N231 onwards is indeed disordered.

To probe the dynamics of the C-terminal region of KKT4^{115–343}, the $\{^1\text{H}\}-^{15}\text{N}$ heteronuclear NOE was measured (Figure 3.11B). For KKT4^{115–343}, a $\{^1\text{H}\}-^{15}\text{N}$ NOE ratio of less than 0.6 was found for all residues from N231 to E343, confirming they are disordered. Variation of the NOE is observed along the sequence indicating some variability in the degree of flexibility. The reduced values at ~245 coincide with a Gly-Gly-Gly sequence at 242–244 likely due to the absence of a sidechain for glycine that allows this residue to sample a wider range of backbone conformations. Lower NOE ratios are also observed in the vicinity of the Gly-Ser-Gly sequence at 271–273. However, not all glycine residues are located in the most flexible regions. G285, which has an NOE ratio of 0.38, is located in a region with the highest NOE

ratios, consistent with the SSP prediction. Taken together, the chemical shift analysis and heteronuclear NOE confirmed disorder in the C-terminal part of KKT4^{115–343}.

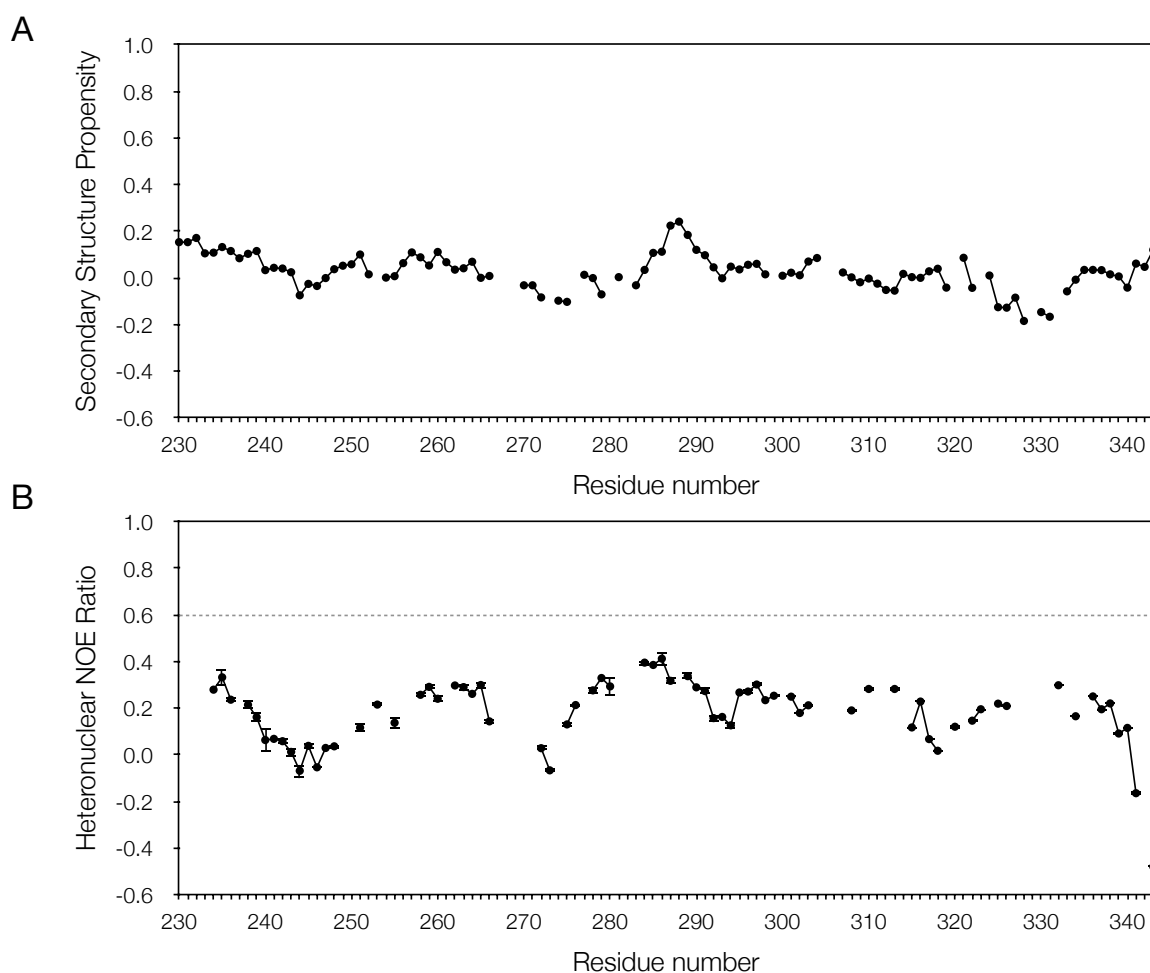


Figure 3.11. The C-terminal region of KKT4^{115–343} is disordered.

(A) Secondary structure propensity (SSP) scores (Marsh et al., 2006) for residues 230–343 of KKT4^{115–343} show no helical (+) or sheet (-) propensity greater than 0.25.

(B) The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE ratios measured for KKT4^{115–343} are less than 0.6 for residues 231–343, showing that this region is disordered. $\{^1\text{H}\}$ - ^{15}N NOE errors were estimated from 500 Monte Carlo simulations using baseline noise as a measure of peak height error.

The N-terminal region of the KKT4 microtubule-binding domain is helical

The N-terminal region of KKT4^{115–343} is predicted to adopt a coiled-coil structure (Figure 1.6D). Information for residues preceding N231 could not be obtained from the NMR spectra of KKT4^{115–343} due to weak or missing peaks for these residues. To assign the missing peaks near the N-terminus of KKT4^{115–343}, additional experiments have been performed using a shorter KKT4 fragment, KKT4^{115–232}, lacking the flexible C-terminal region. The ¹H-¹⁵N BEST TROSY spectrum of KKT4^{115–232} shows better dispersion of the ¹H^N chemical shifts suggesting that residues in this region are more structured than residues within the C-terminal region of the microtubule-binding domain (Figure 3.12). Most of the resolved peaks located ~7.4–8.0 and 8.5–9.4 ppm overlay well between the spectra of KKT4^{115–232} and KKT4^{115–343} (Figure 3.13). This suggests that removing the C-terminal part of the microtubule-binding domain did not perturb the structure close to the N-terminus. However, only ~50 peaks are visible in the spectrum of KKT4^{115–232}; this is less than the expected number for 117 non-proline residues. This can result from an unfavourable NMR relaxation time caused by the large size of the molecule or/and its non-globular shape, found in folds such as coiled coils predicted for the N-terminal fragment of the microtubule-binding domain (Figure 1.6D). To overcome this problem, two shorter, overlapping KKT4 constructs have been used, KKT4^{115–174} and KKT4^{145–232}. KKT4^{115–174} is the minimal microtubule-binding domain (Llauro et al., 2018), while KKT4^{145–232} was experimentally identified as a stable fragment in the trypsin digest of KKT4^{115–343} (see section 8.1 in Appendix). The ¹H-¹⁵N BEST TROSY spectra of both constructs showed the expected number of uniform intensity peaks and better dispersion of ¹H^N chemical shifts than those observed for KKT4^{115–343} (Figures 3.14 and 3.15). This suggests that KKT4^{115–174} and KKT4^{145–232} are mainly structured. The superposition of the NMR spectra of KKT4^{115–174} and KKT4^{145–232} with KKT4^{115–}

²³² showed good agreement in peak positions (Figure 3.16). This indicates that the two shorter constructs retain the structural and dynamical properties observed in the longer fragment, thereby justifying the dissection approach. The spectra of KKT4^{115–174} and KKT4^{145–232} were assigned using triple resonance experiments summarised in Table S1 and the detailed assignment statistics for KKT4^{115–174} and KKT4^{145–232} are included in Table 3.2.

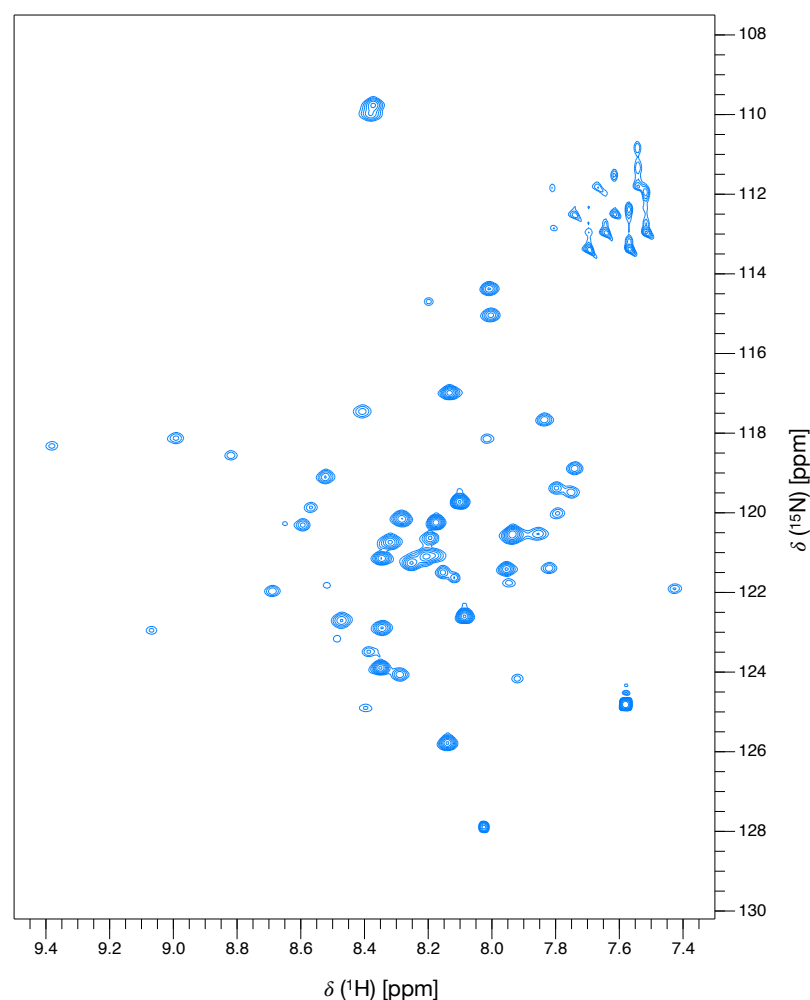


Figure 3.12. The ¹H–¹⁵N BEST-TROSY spectrum of KKT4^{115–232}.

750 MHz ¹H–¹⁵N BEST-TROSY spectrum of KKT4^{115–232} in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H₂O/5% D₂O), at pH 7.2, 20°C. The poor quality of the triple resonance data made it difficult to assign the spectrum of KKT4^{115–232}. However, the assignments can be compared and transferred using the assignments obtained for KKT4^{115–174} (Figure 3.14) and KKT4^{145–232} (Figure 3.15). Peaks in the region of 111–114 ppm and upfield of ~7.6 ppm are artefacts in the BEST-TROSY arising from incomplete cancellation of signals from the side chain amides of Asn and Gln.

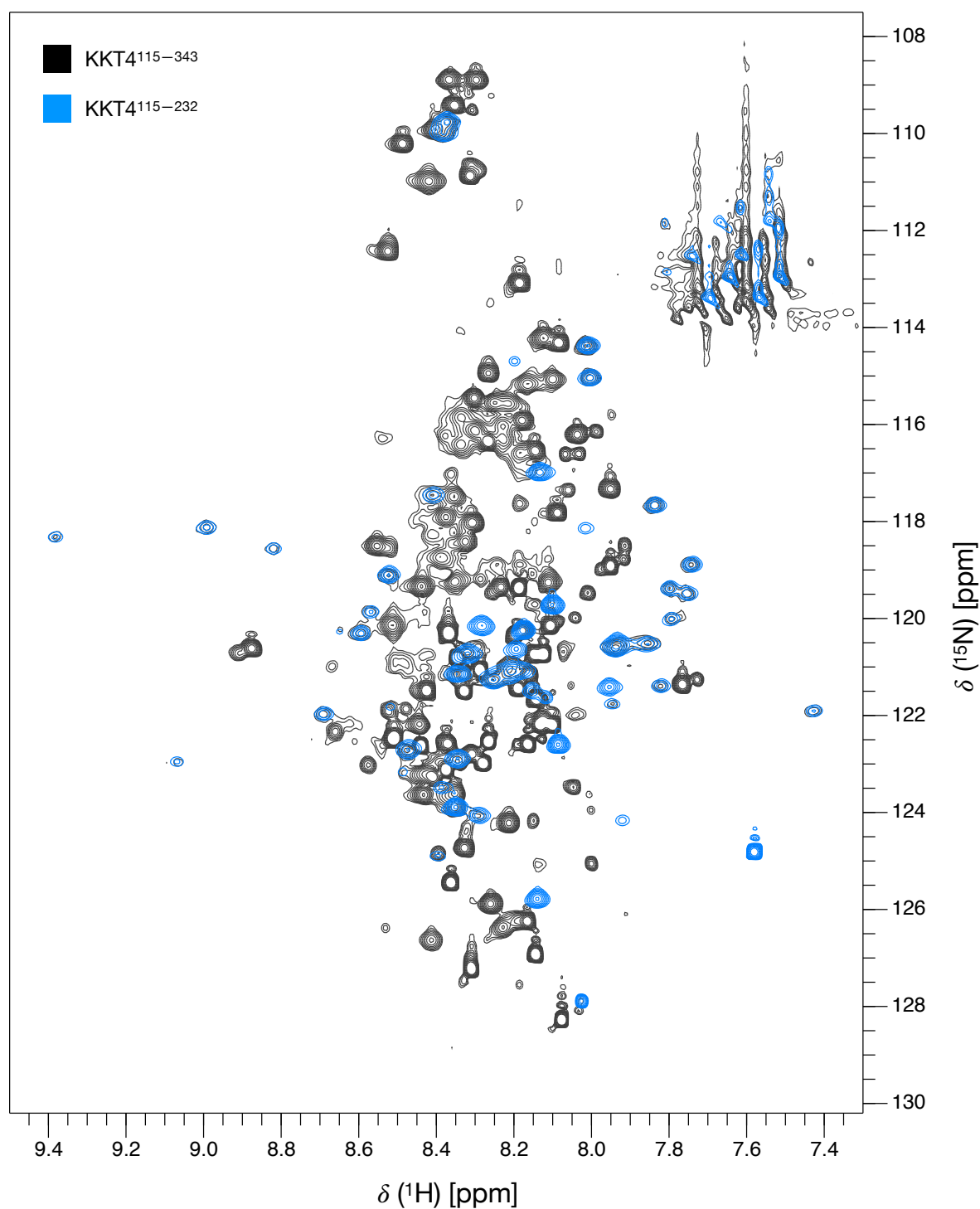


Figure 3.13. An overlay between the spectra of KKT4^{115–343} and KKT4^{115–232}.

The overlay shows that some of the resolved peaks overlay well between the KKT4^{115–343} (black) and KKT4^{115–232} (blue), suggesting that the properties of the N-terminal KKT4 microtubule-binding domain are similar in both fragments. The peaks in the spectrum of KKT4^{115–232} that are not observed in the spectrum of KKT4^{115–343} likely arise from the more structured regions, not visible in the spectrum of KKT4^{115–343}.

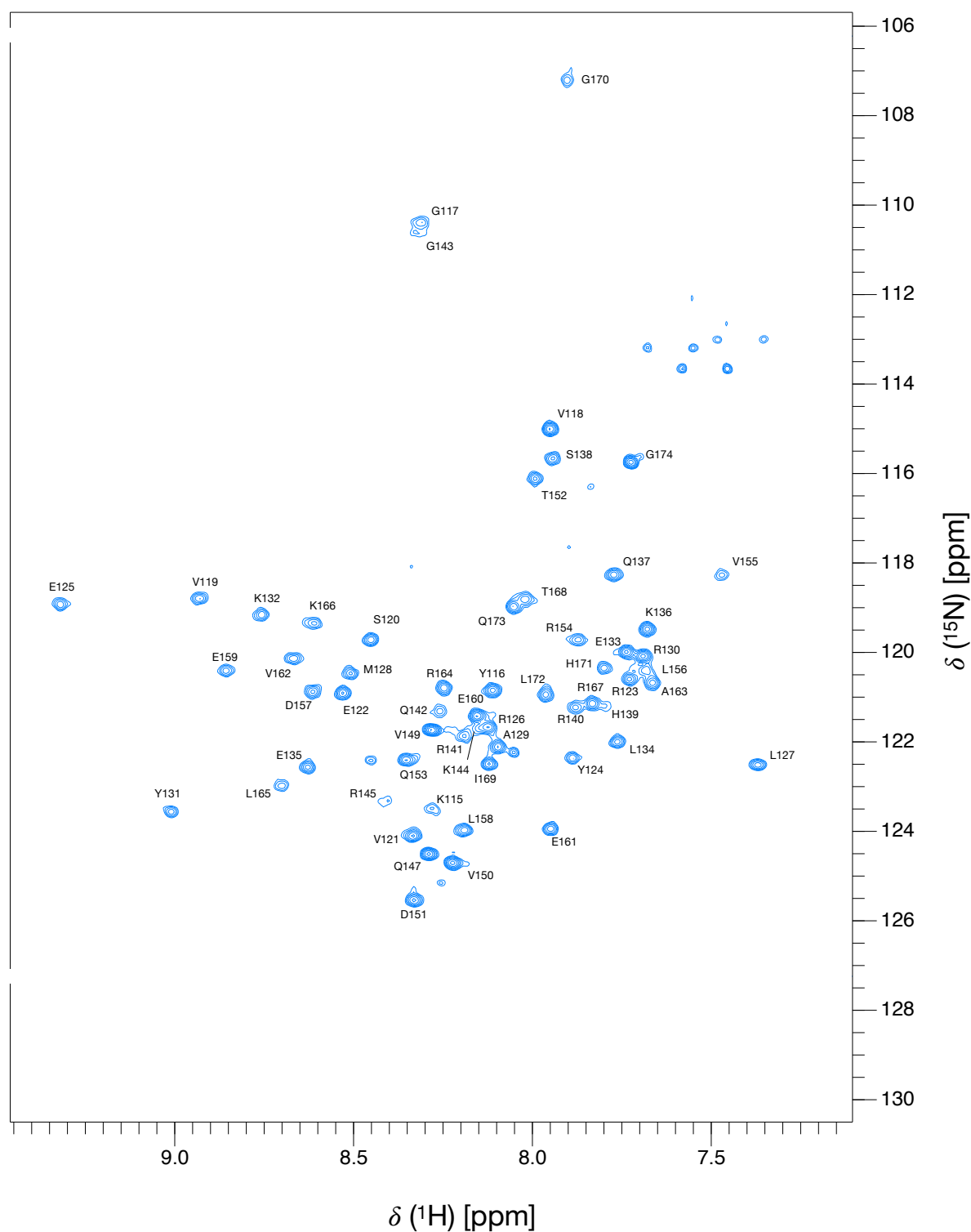


Figure 3.14. The ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–174}.

750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT4^{115–174} in 25 mM HEPES, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 20°C. The peak assignments for backbone amides are annotated. Peaks in the region of 111–114 ppm and upfield of ~7.6 ppm are artefacts in the BEST-TROSY arising from incomplete cancellation of signals from the side chain amides of Asn and Gln.

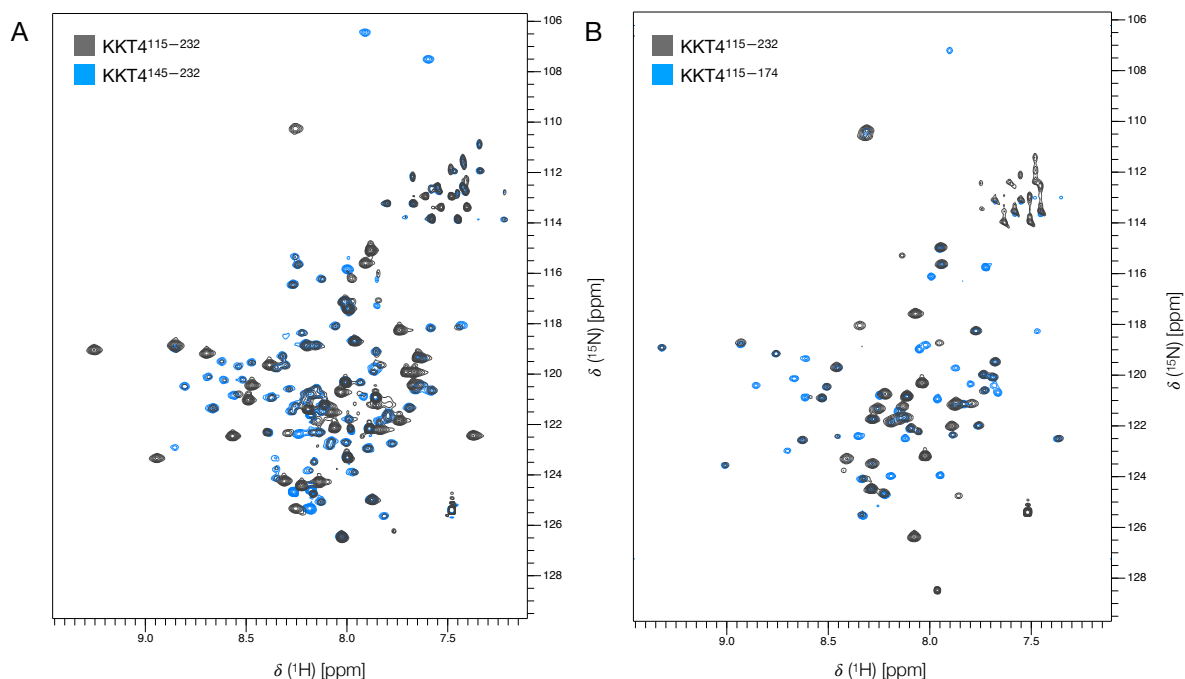


Figure 3.16. An overlay of spectra between the KKT4 fragments.

Overlay of 750 MHz ^1H - ^{15}N BEST-TROSY spectra of (A) KKT4 $^{115-232}$ and KKT4 $^{145-232}$ and (B) KKT4 $^{115-232}$ and KKT4 $^{115-174}$. Most of the peaks in the spectra of KKT4 $^{115-174}$ and KKT4 $^{145-232}$ overlay with peaks observed for KKT4 $^{115-232}$, indicating that the dissection approach used was justified. Peaks that do not overlay arise from the different C-terminal sequence in KKT4 $^{115-174}$ and N-terminal sequence in KKT4 $^{145-232}$. Some peaks visible in the spectra of the shorter KKT4 fragments are not visible in the longer construct, KKT4 $^{115-232}$.

The TALOS-N (Shen and Bax, 2013) analysis of KKT4 $^{115-174}$ and KKT4 $^{145-232}$ chemical shifts revealed significant amounts of secondary structure (Figure 3.17A, B). A continuous helix from T152 to L225 was observed for KKT4 $^{145-232}$ (Figure 3.17A), while two helices, encompassing residues V121 to H139 and T152 to L172, separated by an unstructured linker were observed for KKT4 $^{115-174}$ (Figure 3.17B). The presence of an unstructured linker in KKT4 $^{115-174}$ suggests flexibility between the two helices. To confirm this, $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE was used to probe backbone dynamics in KKT4 $^{115-174}$ and KKT4 $^{145-232}$ (Figure 3.17C). The $\{^1\text{H}\}$ - ^{15}N NOE ratios for KKT4 $^{115-174}$ and KKT4 $^{145-232}$ are consistent with the predicted secondary

structure. Typically, a structured backbone is characterised by $\{^1\text{H}\}-^{15}\text{N}$ NOE ratios of greater than ~ 0.6 . This was observed for most residues that were predicted to be helical (Figure 3.17A, B). In KKT4¹¹⁵⁻¹⁷⁴, residues 140–151, which were not predicted to be helical, had lower $\{^1\text{H}\}-^{15}\text{N}$ NOE ratios, confirming the presence of a disordered dynamic linker between the two helices (Figure 3.17C). Interestingly, higher $\{^1\text{H}\}-^{15}\text{N}$ NOE ratios were observed at the start of the 2nd helix in KKT4¹¹⁵⁻¹⁷⁴ than for this region in KKT4¹⁴⁵⁻²³². This may indicate that the presence of the full linker and/or the 1st helix contributes to the stability of the 2nd helix. The spectrum of KKT4¹¹⁵⁻³⁴³ contained peaks for residues 115–150, which correspond to the 1st helix and part of the flexible linker in KKT4¹¹⁵⁻¹⁷⁴. The $\{^1\text{H}\}-^{15}\text{N}$ NOE ratios for these residues in the KKT4¹¹⁵⁻³⁴³ construct agree with the values measured for KKT4¹¹⁵⁻¹⁷⁴, indicating that this region has the same dynamic properties in the two constructs.

In summary, the NMR analysis shows that the microtubule-binding domain of *T. brucei* KKT4 is composed of two helices, between residues 121–139 and 152–225, separated by a 12-residue flexible linker, followed by a ~ 120 residue positively charged disordered region (predicted isoelectric point for residues 226–343 is 10.1).

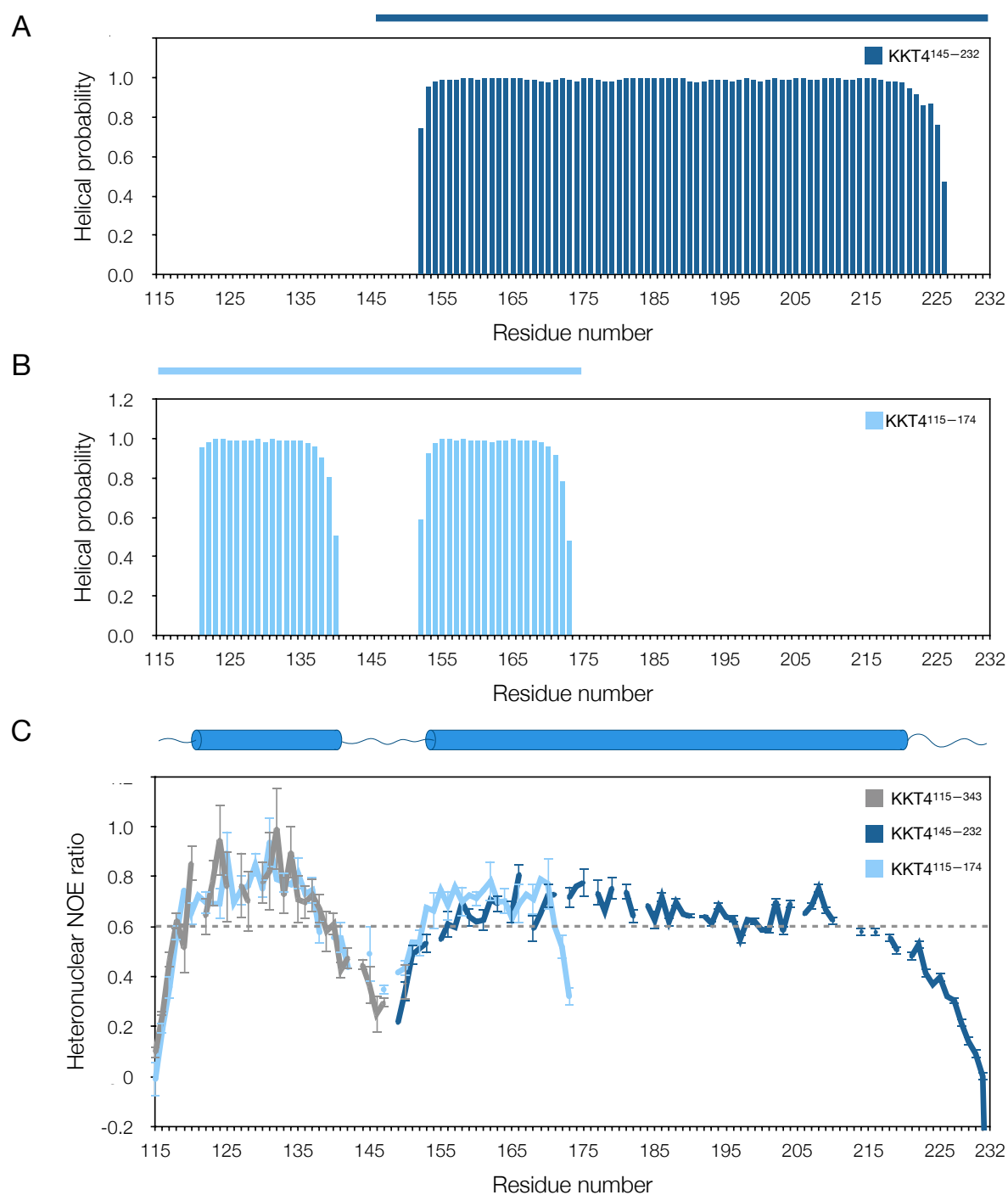


Figure 3.17. The N-terminal region of KKT4 microtubule-binding domain is mainly helical.

- (A) TALOS-N secondary structure analysis of KKT4^{145–232}. The prediction shows a continuous helix in 152–225 region.
- (B) TALOS-N secondary structure analysis of KKT4^{115–174}. The prediction shows two helices, 121–139 and 152–172, which are separated by an unstructured linker. The blue bars above the graphs in (A) and (B) indicate the length of KKT4 fragments used for data collection and TALOS-N analysis.

(C) The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE ratios measured for KKT4^{115–343} (grey), KKT4^{115–174} (light blue) and KKT4^{145–232} (dark blue). Values for the C-terminal region of KKT4^{115–343} are shown in Figure 3.11B. The heteronuclear NOE values in the region of the first helix for KKT4^{115–343} and KKT4^{115–174} agree well, suggesting that removal of the disordered C-terminus did not affect the properties of the KKT4^{115–174} N-terminal region. The lower ratios at the beginnings and ends of the helices may indicate helix fraying. The C-terminal region of KKT4^{115–174} has significantly lower heteronuclear NOE ratios compared with the same residues in KKT4^{145–232} because G174 is the artificially designed C-terminus of KKT4^{115–174}. The value for the C-terminal residue Y232 (1.16) in KKT4^{145–232} is not shown for clarity of the figure. The top panel shows a summary cartoon of the overall secondary structure of KKT4^{115–232} predicted by TALOS-N. The $\{^1\text{H}\}$ - ^{15}N NOE errors were estimated from 500 Monte Carlo simulations using baseline noise as a measure of peak height error.

3.3. X-ray crystallography of the microtubule-binding domain

The NMR analysis of the microtubule-binding domain secondary structure led to an improved construct design for the microtubule-binding domain of KKT4 that minimised the amount of disordered regions, which led to crystallisation of a few constructs. However, none of the obtained crystals of *T. brucei* KKT4 fragments was suitable for solving the protein structure due to poor diffraction patterns. As an alternative approach, several fragments of KKT4 homologues from different kinetoplastid species were generated and tested for crystallisation. This includes KKT4 fragments from *Trypanosoma cruzi* Sylvio, *Trypanosoma vivax*, *Trypanosoma congolense*, *Leishmania mexicana* and *Leptomonas seymouri*. Two fragments, *Trypanosoma cruzi* sylvio KKT4^{117–218} and *Leishmania mexicana* KKT4^{184–284}, yielded crystals that diffracted well enough to determine their crystal structures. The complete list of KKT4 fragments screened for crystallisation can be found in Table S2.

Crystal structure of *Trypanosoma cruzi* KKT4^{117–218}

The crystal structure of KKT4^{117–218} from *T. cruzi* was solved at 1.9 Å resolution (Table 3.3) using ARCIMBOLDO, an *ab initio* macromolecular phasing software (Rodríguez et al., 2012). TcKKT4^{117–218} is homologous to KKT4^{120–224} in *T. brucei* (Figure 1.7). Consistent with the coiled-coil prediction (Figure 1.6D), the structure consists of two parallel α -helices organised in a left-handed ~150 Å long coiled-coil dimer (Figure 3.18A); the helical structure starts at L118 and ends at D215. The coiled-coil consists of 8 regular heptad repeats starting at Y121 and ending at K176. Analysis with TWISTER (Strelkov and Burkhard, 2002) shows that the six central heptads are characterised by an inter-helical distance of 4.84 ± 0.06 Å and a coiled-coil pitch, which is the periodicity of the coiled coil of 136 ± 26 Å; the latter value is

close to the theoretical value of 135 Å for the coiled-coil. If the 1st and 8th heptad are included, the pitch increases to 163 ± 60 Å indicating less supercoiling in these terminal heptads. After K176, the helices move apart (inter-helical distance increases to 5.66 ± 0.15 Å) and the coiled-coil pitch increases significantly, indicating a loss of supercoiling.

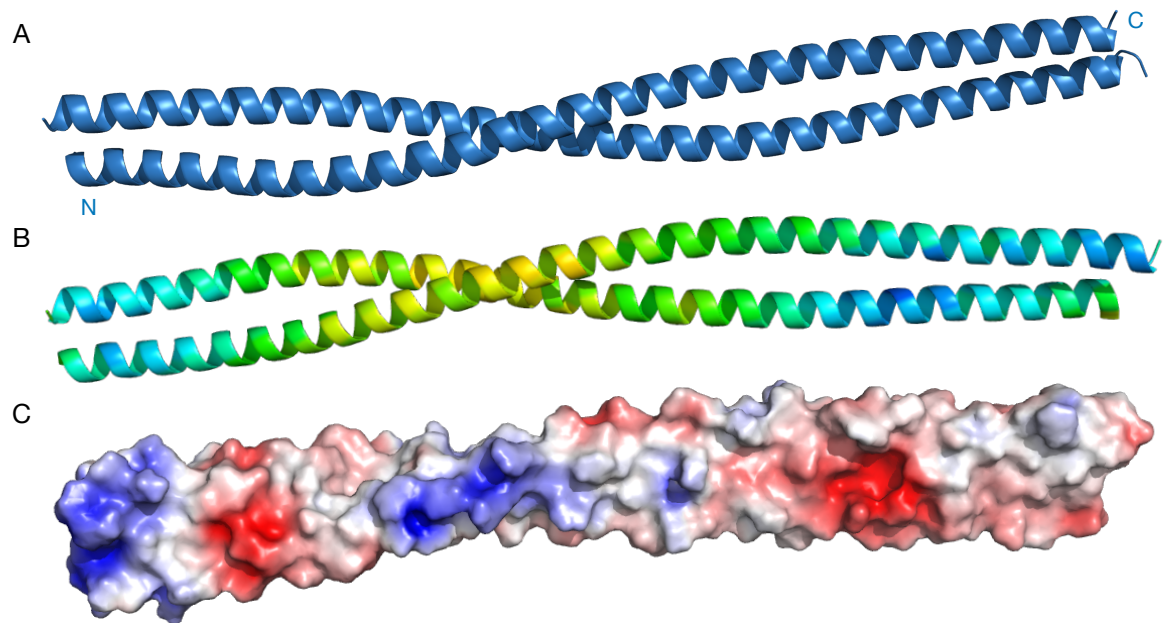


Figure 3.18. Crystal structure of *Trypanosoma cruzi* KKT4¹¹⁷⁻²¹⁸.

- (A) Ribbon model of *Trypanosoma cruzi* (Sylvio X10) KKT4¹¹⁷⁻²¹⁸. The N and C represent the N and C-terminus, respectively.
- (B) B-factors for Cα atoms have been represented for TcKKT4¹¹⁷⁻²¹⁸ using a blue to red spectrum indicating low to high values, respectively. The figures were rendered using PyMol (DeLano, 2002).
- (C) The surface electrostatic potential is coloured from red to blue, -5 kbT to +5 kbT, as calculated by APBS electrostatic plugin in Pymol (Jurrus et al., 2018).

Table 3.3. Data collection and refinement statistics*.

Data collection	<i>T. cruzi</i> KKT4 ^{117–218}	<i>L. mexicana</i> KKT4 ^{184–284}
Beamline	Diamond Light Source I03	Diamond Light Source I24
Wavelength (Å)	0.9760	0.91587
Space group (Z)	<i>P</i> 1 2 ₁ 1	<i>P</i> 1 2 ₁ 1
Unit cell	33.62 Å 25.31 Å 136.88 Å 90° 96.81° 90°	31.31 Å 37.71 Å 122.39 Å 90° 92.17° 90°
Resolution range (Å)	67.96–1.90 (1.97–1.90)	61.15–1.90 (1.97–1.90)
Unique reflections	16812 (368)	22522 (2211)
Completeness (%)	90.4 (20.1)	98.3 (97.4)
Multiplicity	6.1 (4.0)	6.5 (5.3)
<i>I</i> / σ <i>I</i>	6.1 (0.1)	7.1 (0.6)
<i>R</i> _{merge}	0.095 (7.495)	0.117 (2.537)
CC _{1/2}	0.998 (0.274)	0.997 (0.352)
Wilson B-factor (Å ²)	24.4	17.6
Refinement		
No. reflections	16788 (368)	22502 (2211)
<i>R</i> _{work}	0.241 (0.347)	0.199 (0.253)
<i>R</i> _{free}	0.258 (0.478)	0.232 (0.264)
Number of atoms	1871	1816
Protein	1670	1488
Solvent	201	328
RMS bonds (Å)	0.006	0.014
RMS angles (°)	0.97	1.45
Ramachandran favoured (%)	100.00	100.00
Ramachandran allowed (%)	0.00	0.00
Ramachandran outliers (%)	0.00	0.00
Average B-factor (Å ²)	36.0	30.0

*Statistics for the highest-resolution shell are shown in parentheses.

Crystal structure of *Leishmania mexicana* KKT4^{184–284}

To test the structural conservation of the coiled-coil in kinetoplastid species, KKT4 fragments from *Leishmania mexicana* were purified and tested for crystallisation. No diffraction-quality crystals of fragments that contained the N-terminal residues of the microtubule-binding domain (115–140 in *T. brucei*) were obtained. Instead, *LmKKT4*^{184–284}, which corresponds to residues 141–244 in *T. brucei* (Figure 1.7) was crystallised and its structure determined at 1.9 Å resolution (Table 3.3). Due to a cloning error, which failed to place a stop codon after the *LmKKT4*^{184–284} coding sequence, an additional 23 residues (EFELGAPAGRQACGRIMLKSNRK) from the vector were inserted at the C-terminus. Like *TcKKT4*^{117–218}, *LmKKT4*^{184–284} has helices arranged in a parallel coiled-coil fold (Figure 3.19A). *LmKKT4*^{184–284} has 6 regular heptad repeats from the N-terminus to Q224, and after that point, the helices move apart and lose their supercoiling, as observed for *TcKKT4*^{117–218}. The two chains in *LmKKT4*^{184–284} differ in length, with helices spanning T184 to the C-terminus in one chain and T184 to R257 in the other. The region beyond I245, where the helices have moved away from each other, had high B-factors in both chains, suggesting enhanced flexibility (Figure 3.19B). The electron density of the shorter helix disappears at Q258, most likely due to the disordered nature of the protein backbone in this region. Interestingly, the end of the longer helix, which contains extra residues from the expression vector, makes seemingly stabilising contacts with other molecules in the crystal lattice, which potentially explains why my attempts to obtain diffraction-quality crystals for the proper construct ending at Q284 failed.

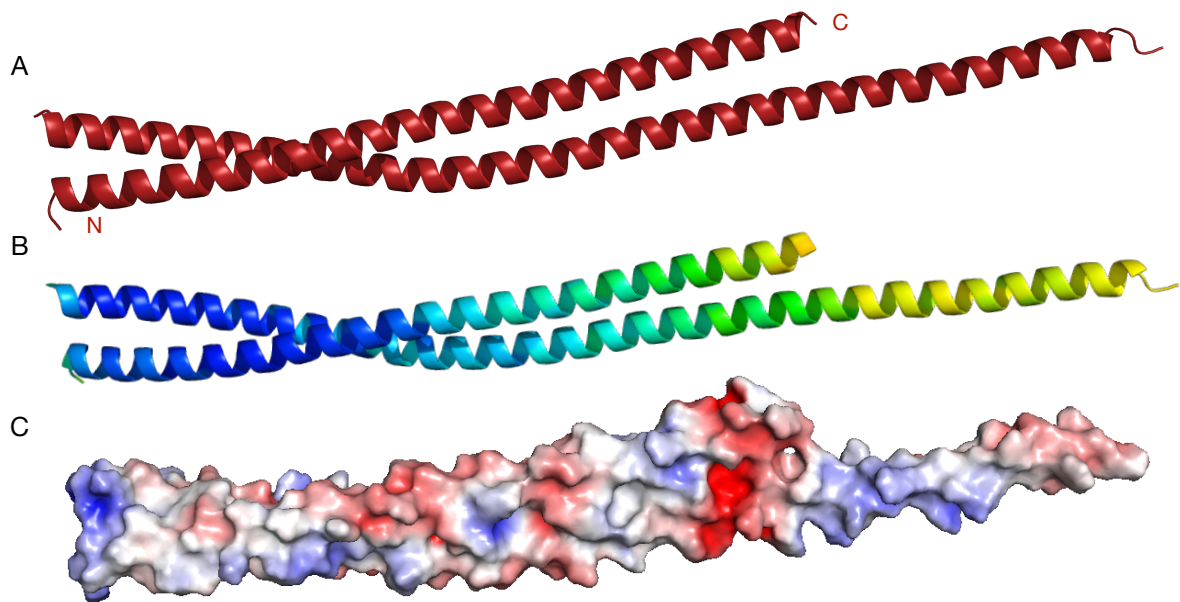


Figure 3.19. Crystal structure of *Leishmania mexicana* KKT4^{184–284}.

- (A) Ribbon model of *Leishmania mexicana* KKT4^{184–284}. The N and C represent the N and C-terminus, respectively.
- (B) B-factors for Cα atoms have been represented for *Lm*KKT4^{184–284} using a blue to red spectrum indicating low to high values, respectively. The figures were rendered using PyMol (DeLano, 2002).
- (C) The surface electrostatic potential is coloured from red to blue, -5 kbT to +5 kbT, as calculated by APBS electrostatic plugin in Pymol (Jurrus et al., 2018).

The helical regions identified in *T. brucei* KKT4 by NMR match those observed in the crystal structures of *TcKKT4*^{117–218} and *LmKKT4*^{184–284}. However, in both crystal structures, no flexible linker was found within the coiled coils that was identified in *T. brucei* KKT4 using NMR. Instead, elevated B-factors could be observed (Figure 3.18B) for the region where one expects to find an unstructured linker in the *TcKKT4* structure based on the sequence alignment (Figure 1.7). The lack of a flexible linker in *T. cruzi* and *L. mexicana* crystal structures may be either due to the stabilising contacts within the crystal lattice or structural differences of KKT4 between *T. brucei* and the other two kinetoplastids. The electrostatic surface potential in crystal structures of both *TcKKT4*^{117–218} and *LmKKT4*^{184–284} shows an equal distribution of acidic and basic regions along the length of coiled coils. A DALI search for structural homologues of *TcKKT4*^{117–218} and *LmKKT4*^{184–284} identified similarity to a number of coiled-coil proteins (Table S4). Superposition of crystal structures from *T. cruzi* and *L. mexicana* show a good structural match with an RMSD of 1.03 Å (for 123 C α) (Figure 3.20), confirming the conservation of the coiled-coil structure in these species.

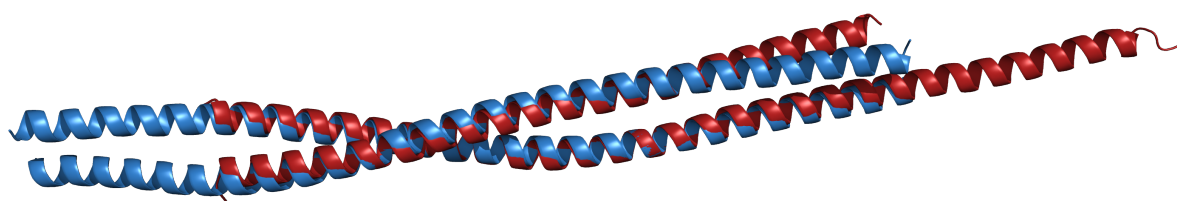


Figure 3.20. An overlay between *TcKKT4*^{117–218} and *LmKKT4*^{184–284} crystal structures.

Using the align function in Pymol (DeLano, 2002), the overlay gives an RMSD of 1.03 Å, suggesting an excellent structural similarity.

Modelling of *T. brucei* KKT4^{115–232}

Dimeric coiled-coil structures have been obtained for residues 117-218 and 184-284 of KKT4 from *T. cruzi* and *L. mexicana*, respectively. Although the crystal structure for *T. brucei* KKT4 could not be solved, the NMR analysis identified helical structure for residues 121-139 and 152-225 in *T. brucei* KKT4. It was interesting to determine if these helices adopt coiled-coil structures similar to those determined for KKT4 homologues. Although NMR chemical shifts alone cannot distinguish between undistorted and supercoiled helices, residual dipolar couplings (RDCs) measured for partially aligned protein samples can provide this information. RDCs are very sensitive to N-H bond vector orientation and are suited to characterising the bending of α -helices that occurs in coiled-coil structures. In a supercoiled helix, the helical turns at the packing interface (residues *a*, *d*, *e*, *g* in the heptad repeat) are slightly compressed while those facing outside (*b*, *c*, *f*) are stretched, which leads to a periodic variation in the RDCs within the heptad repeat (Schnell et al., 2005).

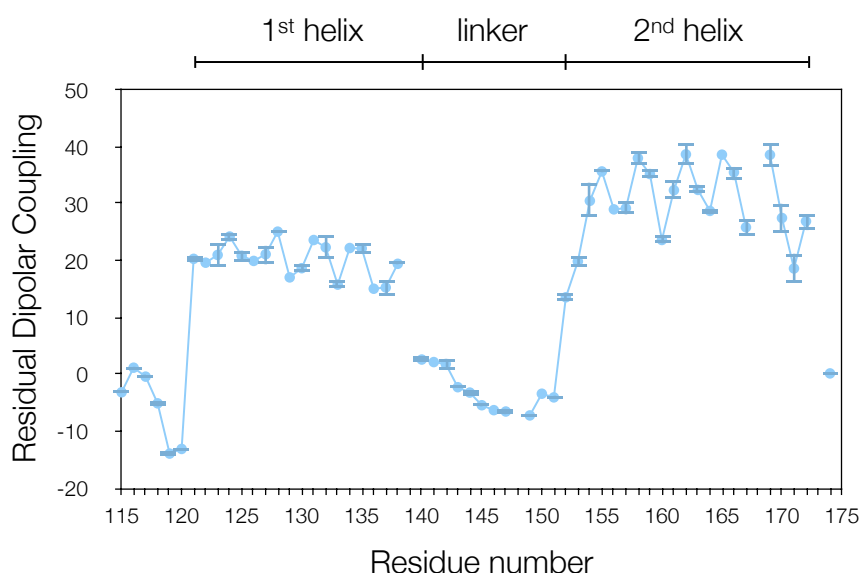


Figure 3.21. The experimental ^1H - ^{15}N residual dipolar couplings for KKT4^{115–174}

Experimental ^1H - ^{15}N residual dipolar couplings for KKT4^{115–174}, measured in 3% C12E6/n-hexanol, are plotted as a function of sequence. The RDCs for both helices show periodic variation that is consistent with a coiled-coil structure. The RDCs close to 0 for the linker region indicate that this region is dynamic. Error bars are standard deviations from three RDC measurements.

RDCs were measured for KKT4¹¹⁵⁻¹⁷⁴ partially aligned in 3% PEG-hexanol (Figure 3.21). Both helices in KKT4¹¹⁵⁻¹⁷⁴ show large positive RDCs while the linker region and the N- and C-terminal residues show RDCs close to 0. The RDCs for both helices show periodic variation, which is consistent with a coiled-coil structure. The flexible linker separating the helices is dynamic, sampling multiple conformations, and this leads to averaging of its RDCs to values close to 0. The different values of the RDCs for the two helices likely reflect the presence of the flexible linker, which means that the two helices can move independently and the molecule does not align as a single rigid unit. Using the *T. cruzi* X-ray structure as a model, different ways of fitting the helices identified by NMR into a coiled-coil structure were tested by optimising the fit between RDCs predicted from the X-ray structure and the experimental RDCs (Figure 3.22A, B). For both helices in *T. brucei*, good fits were found when they were placed within the first half of the *T. cruzi* sequence, corresponding to the regular coiled-coil structure. Poorer agreement was obtained when the *T. brucei* helices were placed within the C-terminal half of the *T. cruzi* structure that showed less supercoiling. For the 2nd helix, the fit of the experimental RDCs suggests an offset of -6 residues between the *Tb* and *Tc* sequences (Figure 3.22B), in agreement with the sequence alignment of these two proteins (Figure 1.7); this places hydrophobic residues in *T. brucei* in positions *a/d* of the heptad repeat in the *T. cruzi* structure. For the 1st helix, two solutions, with sequence offsets of -3 or 0 residues, are found from the fitting of experimental RDCs (Figure 3.22A). The sequence offset of 0 was excluded because this would place L127 and L134 in heptad position *g* away from the helix packing interface and would not give rise to the experimentally observed upfield shifts for the methyl groups that arise from packing with Y124 and Y131 at the interface (Figure 3.22C, D). The sequence offset

of -3 determined from the RDC and chemical shift data is in agreement with the alignment of the *Tb* and *Tc* sequences (Figure 1.7) in the region of the 1st helix.

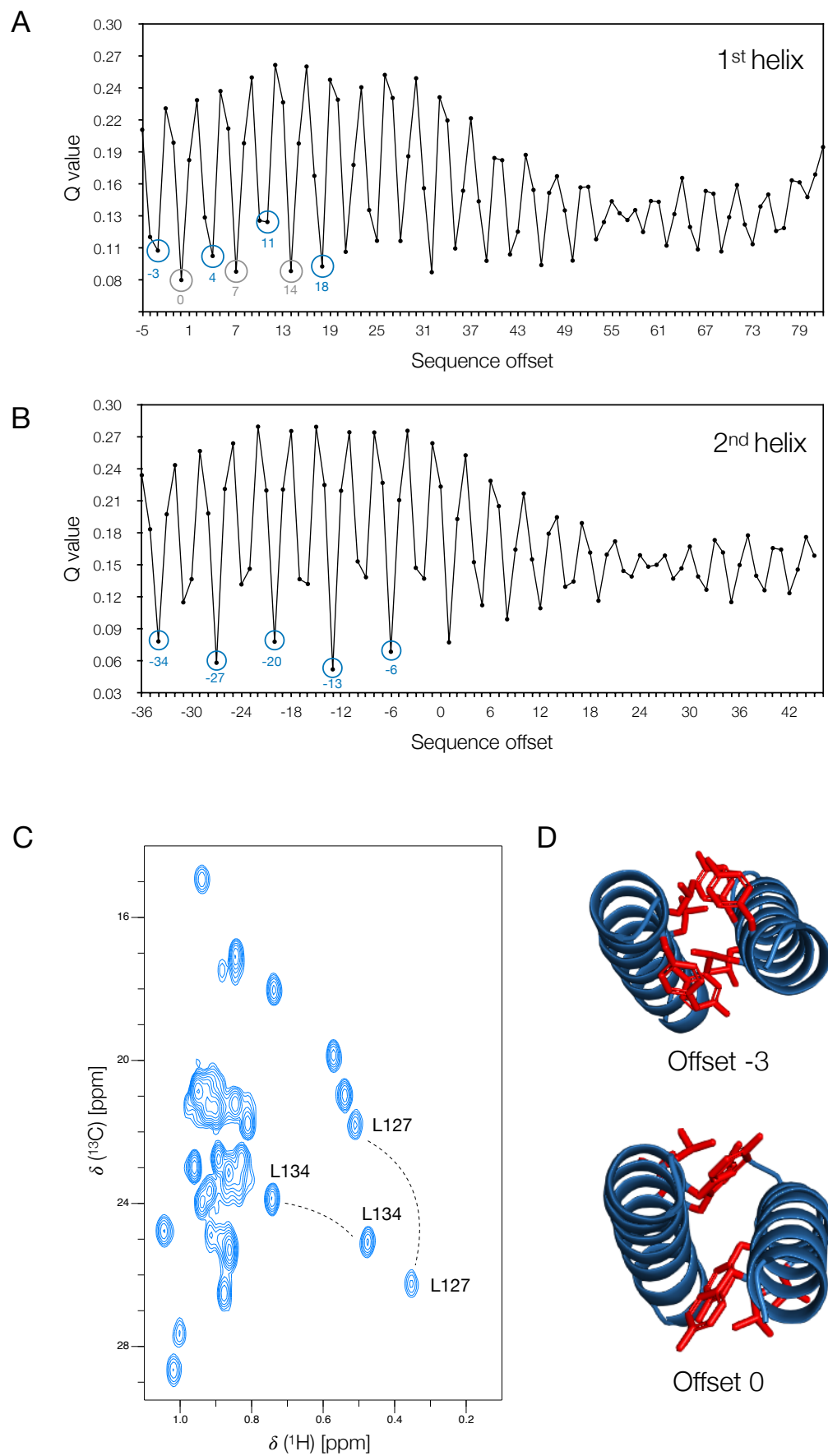


Figure 3.22. Fitting the experimental RDC data using the *TcKKT4*^{117–218} crystal structure.

The experimental RDCs for residues 126–135 from the 1st helix (A) and 154–170 from the 2nd helix (B) from *T. brucei* KKT4^{115–174} were fitted to the calculated RDCs based on the *T. cruzi* KKT4 X-ray structure. (C) The ¹H-¹³C HSQC of KKT4^{115–174} shows that both of the methyl groups of L127 and one of the methyl groups from L134 have ¹H peaks that are shifted upfield (to the right) of the peaks arising from most other L/V/I residues (around 0.7–1.05 ppm). These upfield shifts are due to the interaction of L127 and L134 with the aromatic rings of Y124 and Y131 and can be used to resolve the degeneracy between the two possible sequence offsets. Homology models for the 1st helix of *T. brucei* KKT4^{115–174} have been built with Modeller using the *T. cruzi* X-ray structure and sequence alignments based on sequence offsets of -3 and 0 residues; these are shown in (D) with the side chains of Y124/L127/Y131/L134 shown as red sticks. The -3 sequence offset places Y124/Y131 and L127/L134 in positions *a* and *d* of the heptad, respectively, where they interact closely. The 0 sequence offset places these pairs of residues in positions *d* and *g*; L127/L134 in the *g* position are not closely packed at the inter-helix interface. ¹H chemical shifts for the L127 and L134 methyl groups were then predicted for the two alternative coiled-coil models. Only the homology model generated with a sequence offset of -3 residues predicts upfield shifts for the methyl groups of L127 and L134 due to their close proximity to Y124 and Y131 within the same helix and between helices in the packing interface. The sequence alignment of the *T. brucei* and *T. cruzi* proteins (Figure 1.7) in the region of the 1st helix suggests an offset of -3 residues, which is consistent with the RDC data and chemical shift analysis of the homology models.

A homology model for the two coiled-coil regions of *T. brucei* KKT4^{115–232} was built using the *T. cruzi* coiled-coil X-ray structure, and the sequence alignments derived from the RDC data (Webb and Sali, 2016). Random extended structures for the flexible N- and C-termini and for the inter-helix linker, all identified as flexible on the basis of the heteronuclear NOE data, were generated; these represent possible conformations that these residues might sample in solution. These coordinates were

merged with the coiled-coil homology models to generate an overall model for *T. brucei* KKT4^{115–232} (Figure 3.23A). This model allows the properties of the surface of *T. brucei* KKT4^{115–232} to be investigated (Figure 3.23B).

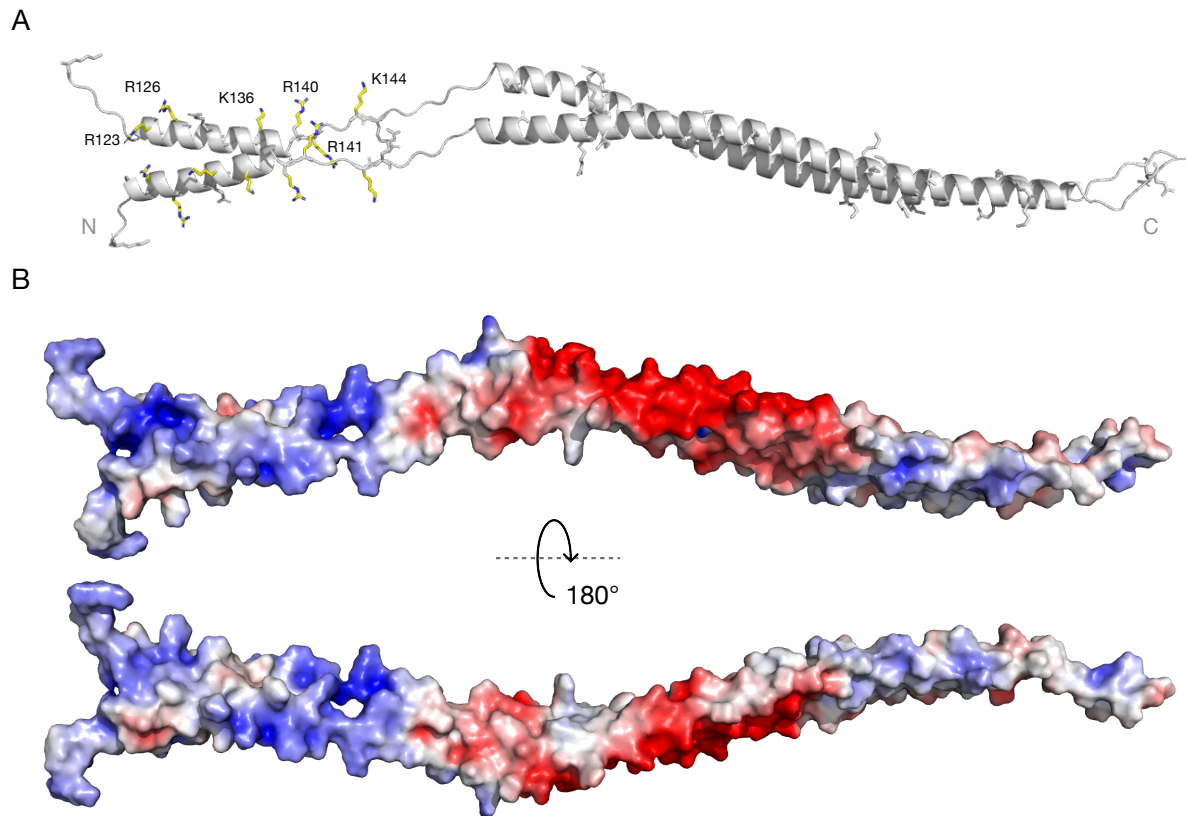


Figure 3.23. Homology model of *Trypanosoma brucei* KKT4^{115–232}.

- (A) Ribbon representation of the *T. brucei* KKT4^{115–232} structural model generated using MODELLER (Webb and Sali, 2016). The model, based on NMR, X-ray crystallography, and sequence alignment between *T. brucei* and *T. cruzi* KKT4 shows two coiled-coil segments separated by an unstructured linker and flanked by small disordered regions at the N and C termini of the molecule. Lysine and arginine side chains are presented as sticks. Residues for which charge-reversal mutations affected the microtubule-binding activity of KKT4 (see Figure 3.26) are shown in yellow and labeled.
- (B) Homology model of KKT4^{115–232} coloured by surface electrostatic potential, showing positively charged patches in the N-terminal helix. Red to blue, 5 kbT to +5 kbT, as calculated using the APBS electrostatic plugin in PyMOL (Jurrus et al., 2018).

3.4.KKT4 microtubule-binding activity

Disordered C-terminal tail enhances KKT4 affinity to microtubules

Structural analysis of the microtubule-binding domain in KKT4 revealed two regions: the N-terminal coiled coil and the C-terminal basic, unstructured tail. To test contributions of these regions for the microtubule-binding affinity of KKT4, three KKT4 fragments have been purified: a complete microtubule-binding domain (KKT4^{115–343}), the coiled-coil region (KKT4^{115–232}) and the unstructured basic tail (KKT4^{233–343}) (Figure 3.24A). The unstructured basic tail did not co-sediment with microtubules on its own (Figure 3.24B), consistent with the previous study (Llauro et al., 2018), where KKT4^{168–343} was used. In contrast to the C-terminal fragment, the coiled-coil region bound microtubules in the co-sedimentation assay (Figure 3.24B), however, with weaker affinity compared to the KKT4^{115–343}. This suggests that although the C-terminal tail is unable to interact with microtubules on its own, it enhances the affinity of the coiled-coil region for microtubules. To compare affinities between KKT4^{115–343} and KKT4^{115–232}, I repeated the co-sedimentation assay with a gradient of microtubule concentrations and measured the dissociation constants by calculating the ratios between the bound and unbound fractions of KKT4 proteins (Figure 3.24C, D). This experiment confirmed that KKT4^{115–343} has a higher affinity for microtubules compared to the KKT4^{115–232}.

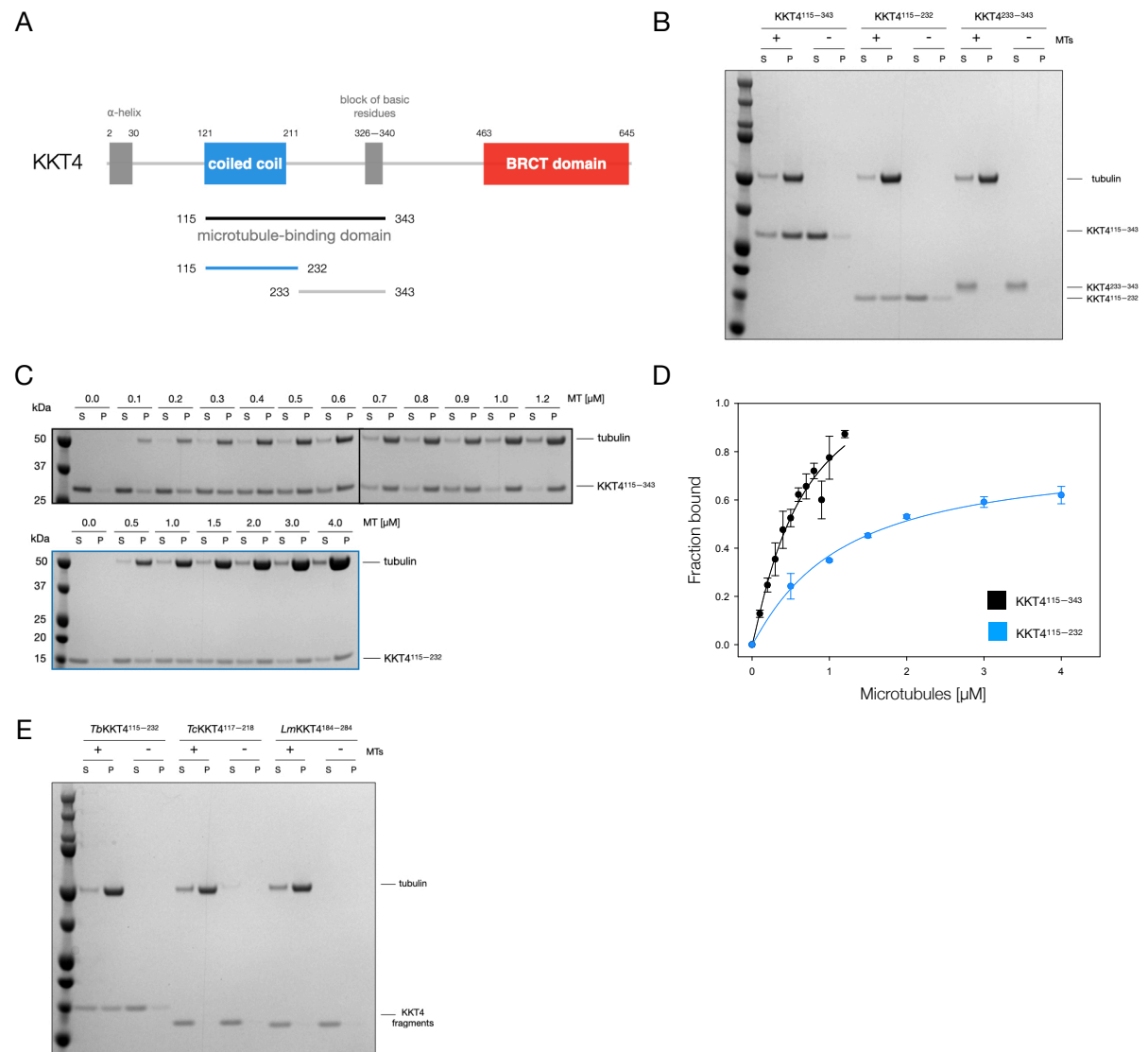


Figure 3.24. *T. brucei* KKT4 coiled coil is sufficient for microtubule interaction *in vitro*.

- (A) Cartoon representation of KKT4 fragments used for quantitative analysis of the microtubule-binding activity.
- (B) Microtubule co-sedimentation assay of KKT4¹¹⁵⁻³⁴³, KKT4¹¹⁵⁻²³² and KKT4²³³⁻³⁴³, showing that the KKT4 coiled coil (KKT4¹¹⁵⁻²³²) is sufficient to interact with microtubules (although with weaker affinity compared with KKT4¹¹⁵⁻³⁴³). The unstructured basic region (KKT4²³³⁻³⁴³) does not co-sediment with microtubules, but its presence in KKT4¹¹⁵⁻³⁴³ enhances the microtubule-binding activity of KKT4. S and P correspond to supernatant and pellet fractions, respectively.
- (C) Microtubule co-sedimentation assay of KKT4¹¹⁵⁻³⁴³ and KKT4¹¹⁵⁻²³² with increasing concentrations of microtubules.
- (D) Plot showing the fraction bound against the concentration of microtubules from (C). Error bars are standard deviations from three independent measurements.
- (E) Microtubule co-sedimentation assay of TbKKT4¹¹⁵⁻²³², TcKKT4¹¹⁷⁻²¹⁸ and LmKKT4¹⁸⁴⁻²⁸⁴. Unlike *T. brucei* KKT4, coiled-coil regions from *T. cruzi* and *L. mexicana* did not co-sediment with taxol-stabilised microtubules at the concentration tested. S and P correspond to supernatant and pellet fractions, respectively.

A previous study showed that KKT4 fragments corresponding to the microtubule-binding domain from different kinetoplastid species (*T. cruzi*, *L. mexicana*, *T. congolence* and *Phytomonas*) co-sedimented with taxol-stabilised microtubules, showing the functional conservation of the domain (Llauro et al., 2018). Analysis using *T. brucei* KKT4 shows that the coiled-coil region (KKT4^{115–232}) is sufficient to interact with microtubules (Figure 3.24B). In contrast, corresponding fragments from *T. cruzi* and *L. mexicana* (TcKKT4^{117–218}, LmKKT4^{184–284}) did not co-sediment with microtubules (Figure 3.24E). This could be explained by different affinities to microtubules of KKT4 domains in *L. mexicana* and *T. cruzi* or by minor structural differences like an unstructured linker insertion observed in *T. brucei* KKT4 by NMR, but not in *L. mexicana* or *T. cruzi* crystal structures. It might be that the basic disordered tail is essential for microtubule-binding activity in *T. cruzi* or *L. mexicana* but not in *T. brucei*.

An NMR-based approach to analyse the interaction between KKT4 and microtubules

Analysis of the KKT4 fragments suggests that the microtubule-binding interface is located within the coiled-coil region of the microtubule-binding domain, while the C-terminal tail can enhance the affinity of KKT4 to microtubules. However, it is unclear which residues on KKT4 are critical in forming attachments to microtubules. To obtain residue-specific information, I employed NMR spectroscopy using a method previously used to study the interaction between a kinetochore protein Knl1 and microtubules (Bajaj et al., 2018). In this study, a comparison of the 2D ^1H - ^{15}N HSQC spectra of ^{15}N -Knl1 with and without unlabelled taxol-stabilised microtubules has been made. I decided to apply the same approach to study the interaction between ^{15}N -KKT4^{115–343} and microtubules. In these experiments, the peaks corresponding to residues from KKT4^{115–343} that interact with microtubules should display chemical shift changes and/or intensity changes due to the altered environment, while those corresponding to residues that are not involved in the binding should remain unaffected. However, mixing equimolar amounts of taxol-stabilised microtubules and ^{15}N -KKT4^{115–343} resulted in severe precipitation; this was not observed in the *in vitro* microtubule co-sedimentation assay, where lower protein concentrations are used. Regardless of the precipitation, a 2D ^1H - ^{15}N BEST TROSY spectrum of the complex was acquired; however, no peaks were observed (data not shown). Therefore, KKT4^{115–343} in complex with microtubules is unsuitable for NMR analysis. This is likely due to large size of the complex formed between KKT4 and microtubules, resulting in signal broadening. This could suggest that the interaction between KKT4 and microtubules is a slow-exchange interaction and all of the KKT4 is bound to the microtubules. In conclusion, the application of NMR spectroscopy to study the interaction between KKT4 and microtubules was unsuccessful.

Crosslinking mass spectrometry approach to analyse the interaction between KKT4 and microtubules.

Crosslinking mass spectrometry (XL-MS) can be used to study protein-protein interactions by identifying interacting surfaces between partner proteins or within the same molecule (Mattson et al., 1993, Leitner et al., 2016). The XL-MS has been previously applied to determine the structural basis of the interaction between the Ska complex and microtubules (Abad et al., 2014). I applied crosslinking mass spectrometry to study the interaction between KKT4^{115–343} and microtubules. The purified KKT4^{115–343} was mixed with taxol-stabilised microtubules and crosslinked using a zero-length 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, commonly known as EDC in the presence of N-hydroxysulfosuccinimide (sulfo-NHS) (Figure 3.25A, B). Control reactions were prepared, where KKT4^{115–343} and microtubules were crosslinked separately. When both KKT4 and microtubules were used, a ~75 kDa band was observed after the crosslinking reactions (Figure 3.25A). This band was not found in either of the control reactions, suggesting that it likely corresponds to a KKT4/tubulin complex. To identify the inter-protein crosslinks, the gel bands corresponding to the crosslinked complex were cut out from the SDS-PAGE gel and analysed using mass spectrometry. Although MS confirmed the presence of both tubulin and KKT4 in the analysed bands, no crosslinked peptides were detected (data not shown). This can be explained by insufficient trypsin cleavage, size of the crosslinked peptides beyond the detection limit of MS or limited efficiency of the crosslinking reaction. To optimise the coverage of the detected peptides, different proteases have been tested (including chymotrypsin, Glu-C, Arg-C, Lys-C and pepsin) but crosslinked peptides were still not identified. In summary, the application of crosslinking mass spectrometry to study the interaction between KKT4 and microtubules was unsuccessful.

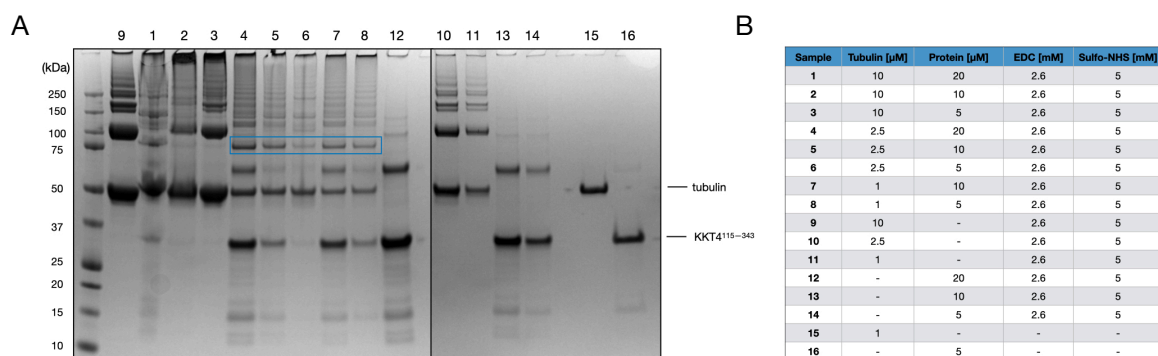


Figure 3.25. Crosslinking mass spectrometry of KKT4^{115–343} in complex with microtubules.

(A) The SDS-PAGE gel showing resolved crosslinking reactions, composition of which is reported in (B). In reactions 4–8, ~75 kDa crosslinked bands can be observed and are highlighted in the blue box. These bands are not observed in the control reactions (9–14), where KKT4 and microtubules were not combined, suggesting that crosslinks were formed between KKT4 and microtubules in these bands. The migration of proteins, prior to the crosslinking reaction is shown in lane 15 and 16.

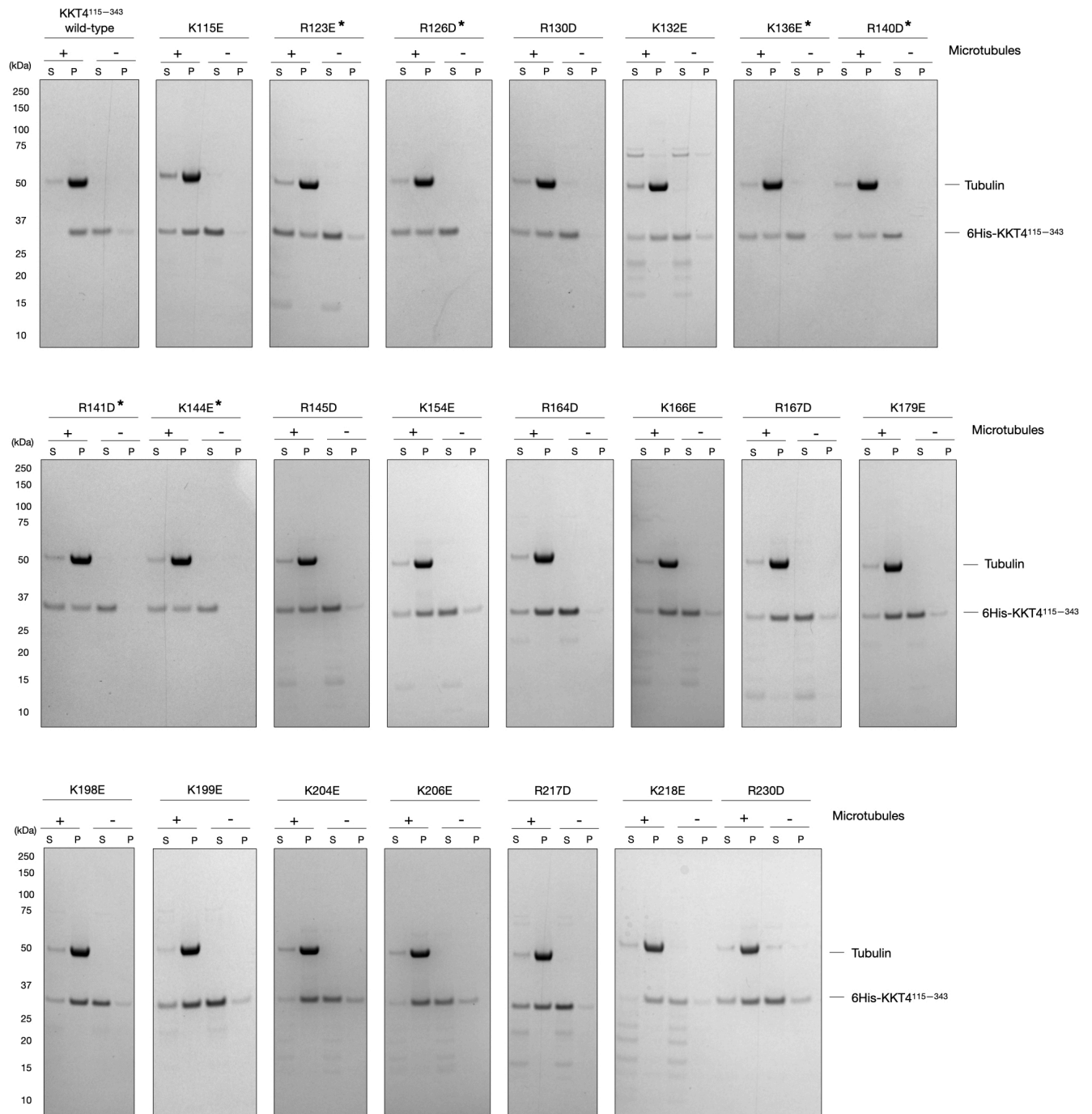
(B) Table showing the detailed composition of crosslinking reactions resolved in (A).

Microtubule-binding interface

NMR and XL-MS unfortunately failed to provide residue-specific information about the interaction between KKT4 and microtubules. Interaction with microtubules often involves positively charged surfaces of binding proteins and acidic tails of tubulin dimers (Ciferri et al., 2008). A previous study showed that replacing three basic residues with acidic residues (R123E, K132E, R154E) in KKT4 severely reduced the microtubule-binding activity of *T. brucei* KKT4^{115–343} (Llauro et al., 2018). To analyse charge distributions on the coiled-coil surface in the *T. brucei* KKT4^{115–232} homology model, the electrostatic surface potential has been calculated using the APBS electrostatic plugin in Pymol (Jurrus et al., 2018). The structural model of *T. brucei* KKT4^{115–232} revealed positively charged regions in the N-terminal fragment of the coiled-coil structure (Figure 3.23B), with basic residues exposed on the protein surface (e.g. R123, R126, R130, K132, K136, R140, K199, K204, K206, R217 and

K218) (Figure 3.23A). To identify residues that facilitate the interaction between KKT4 and microtubules, single mutants of KKT4^{115–343} have been generated, where lysine and arginine residues that are present within the coiled-coil region (115–232) of the microtubule-binding domain were replaced with glutamic or aspartic acid. The microtubule-binding activity of these mutants was tested using a co-sedimentation assay and compared to that of wild-type KKT4^{115–343}. Charge reversal mutants of residues R123, R126, K136, R140, R141 and K144 significantly reduced the KKT4^{115–343} affinity for microtubules (Figure 3.26A, B). These residues are located within the first part of KKT4^{115–174}, which was previously shown to be sufficient for the interaction with microtubules (Llauro et al., 2018). In contrast, mutation of residues located in the later part of the coiled-coil (R145, K154, R164, K166, R167, K179, K198, K199, K204, K206, R217, K218, and R230) had only mild or no effect on the affinity for microtubules (Figure 3.26A, B). 1D ¹H NMR spectra were collected for mutants that had reduced affinity to microtubules (R123, R126, K136, R140, R141, K144) to test if mutations altered the stability of the coiled-coil; these confirmed that the mutations did not affect the protein stability (data not shown). These results confirm the importance of the positively charged residues for microtubule-binding activities in KKT4 and revealed that the interaction with microtubules is located in the N-terminal basic region of the coiled-coil.

A



B

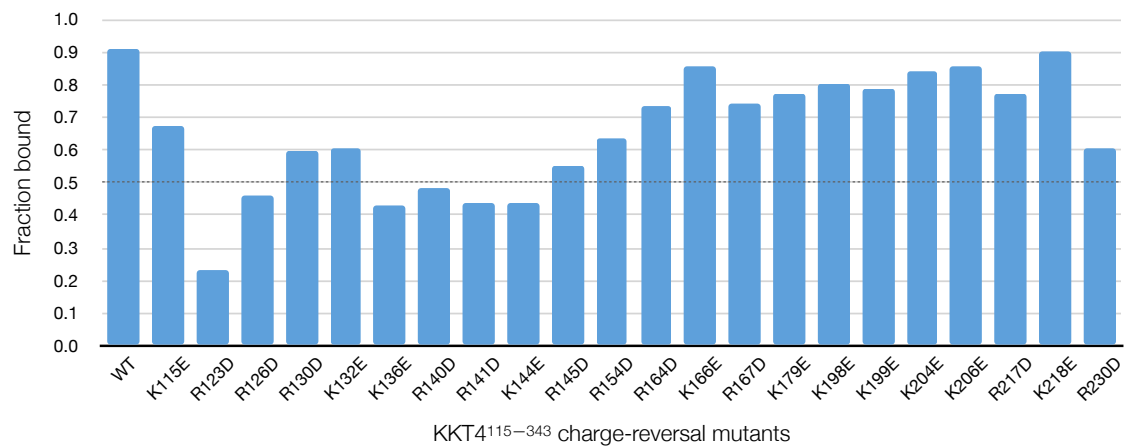


Figure 3.26. Mapping the microtubule-binding site in KKT4^{115–343}.

- (A) Microtubule co-sedimentation assay of 6His-KKT4^{115–343} charge-reversal mutants, showing that mutations in the N-terminal basic coiled-coil region reduce the microtubule-binding affinity. The mutants that show ~50% reduction in the microtubule binding are indicated with an asterisk (*). S and P correspond to supernatant and pellet fractions, respectively.
- (B) Histogram showing the ratios between bound and unbound fraction for different mutants of KKT4^{115–343} using gels shown in (A). The mutations of the N-terminal residues in the domain show the most severe defects in microtubule binding.

Chapter conclusions

- The microtubule-binding domain of *T. brucei* KKT4 is composed of two helices, encompassing residues 121–139 and 152–225, separated by a 12-residue flexible linker, followed by a ~120-residue positively charged disordered region.
- Crystal structures of *Tc*KKT4^{117–218} and *Lm*KKT4^{184–284} show coiled-coil folds.
- Based on crystal structures, NMR data and sequence alignments, a structural model of *T. brucei* KKT4^{115–232} was generated.
- Mutagenesis of KKT4^{115–343} suggests that KKT4 binds to microtubules through a basic N-terminal coiled-coil region.

CHAPTER 4

4. DNA-binding activity of KKT4

In humans, the constitutive centromere-associated network (CCAN) is built by 16 chromatin-proximal kinetochore proteins that act as a platform for the kinetochore assembly by interacting with CENP-A nucleosomes. Most of the CCAN components are constitutively localised throughout the whole cell cycle and some of them have DNA-binding activities. In contrast, microtubule-binding components of the kinetochore complex localise only during mitosis (Masumoto et al., 1989, Palmer et al., 1991, Saitoh et al., 1992, Sugimoto et al., 1994, Sullivan et al., 2013). Among KKT proteins, six proteins have constitutive localisation: KKT2, KKT3, KKT4, KKT20, KKT22 and KKT23. This suggests that KKT4 might be associated with DNA, although it has microtubule-binding activities. Moreover, the C-terminal fragment of KKT4 is predicted to be a tandem BRCT domain found in a number of prokaryotic

and eukaryotic proteins with functions in various biological processes such as DNA damage response pathways. Some of them can bind DNA or RNA (Zhang et al., 1998, Yu et al., 2003, Owen and McMurray, 2009, Leung and Glover, 2011). When purifying the full-length KKT4 from insect cells, significant DNA contamination was detected, which was removed using a heparin column and washes with high salt. However, it was unknown whether KKT4 can interact with DNA.

4.1. KKT4 interacts with DNA with nanomolar affinity

To test if KKT4 has DNA-binding activities, I employed a fluorescence anisotropy assay, where double-stranded 50 bp DNA was labelled with 6-carboxyfluorescein (6-FAM); the sequence of the DNA probe was designed using a fragment from the *Trypanosoma brucei* centromeric region (Obado et al., 2007). Interactions between the DNA probe and a protein molecule result in substantial changes in anisotropy (compared to the DNA molecule on its own) and allow the dissociation constant (K_D) of the protein-DNA complex to be measured. To test whether KKT4 has DNA-binding activities, the full-length KKT4 was purified from insect cells. The protein was found to interact with the DNA probe with an 11 nM K_D (Figure 4.1A), meaning that KKT4 can indeed interact with DNA at least *in vitro*.

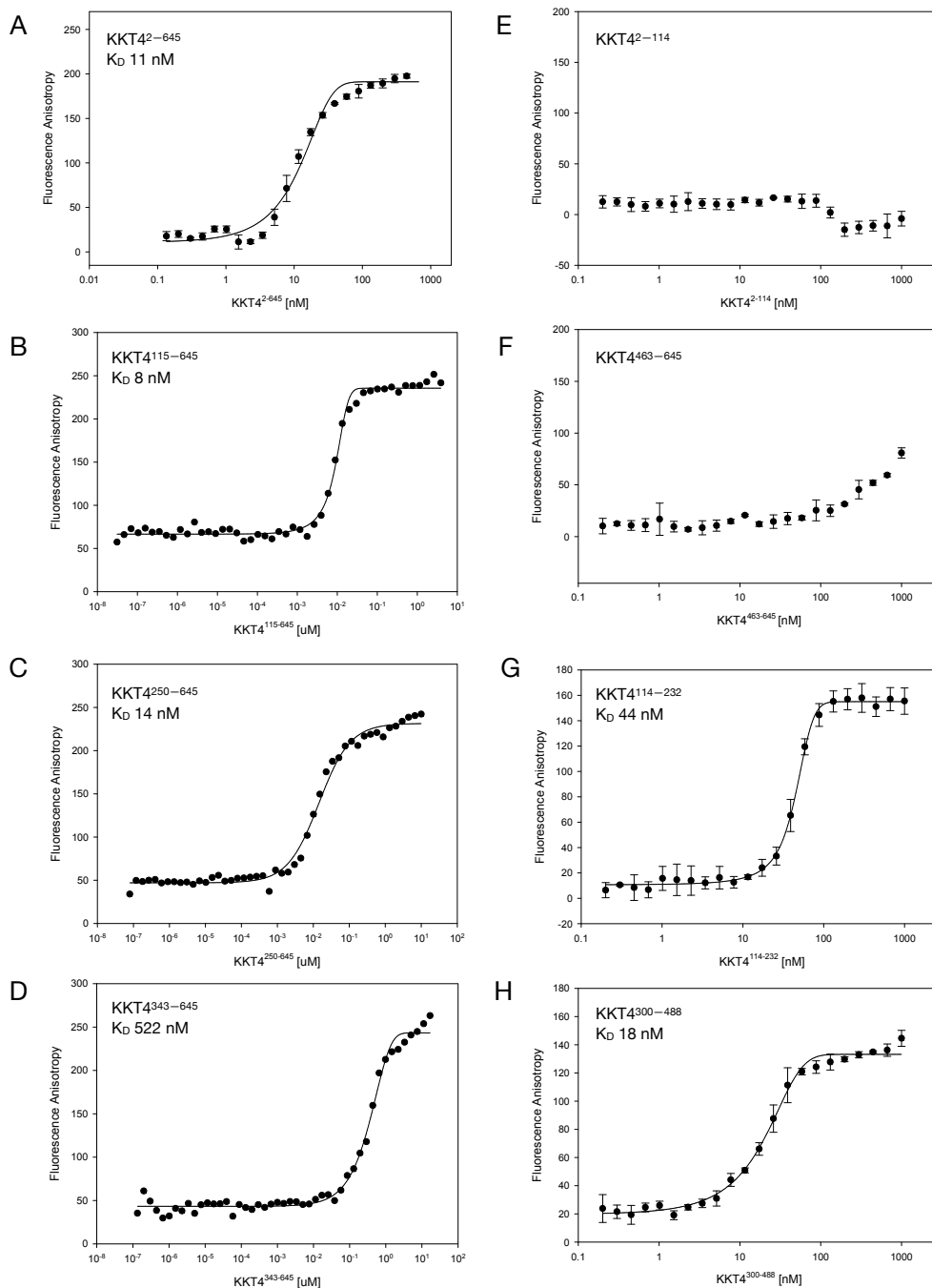


Figure 4.1. KKT4 interacts with DNA *in vitro*.

Measured anisotropy is plotted against protein concentrations of KKT4 fragments (A–H). The length of each fragment and K_D values are reported in the top left corner of each graph. In this fluorescence anisotropy assay a 50-bp DNA probe was used, which is part of the centromeric sequence (CIR147) in *T. brucei*. The K_D values were calculated using non-linear regression using SigmaPlot (Monks, 2002); the fit is shown as a solid line. The K_D was not calculated for KKT4^{463–645} and KKT4^{2–114} as the reactions did not reach a plateau; therefore, the K_D estimation would be incorrect. Error bars are standard deviations from three independent measurements. The fluorescence anisotropy was measured just once for KKT4^{115–645}, KKT4^{250–645} and KKT4^{343–645}, so no error bars are reported for these constructs.

4.2. KKT4 has two DNA-binding domains

To determine which part of KKT4 binds DNA, shorter KKT4 fragments were purified from bacteria and tested using a fluorescence anisotropy assay (Figure 4.2).

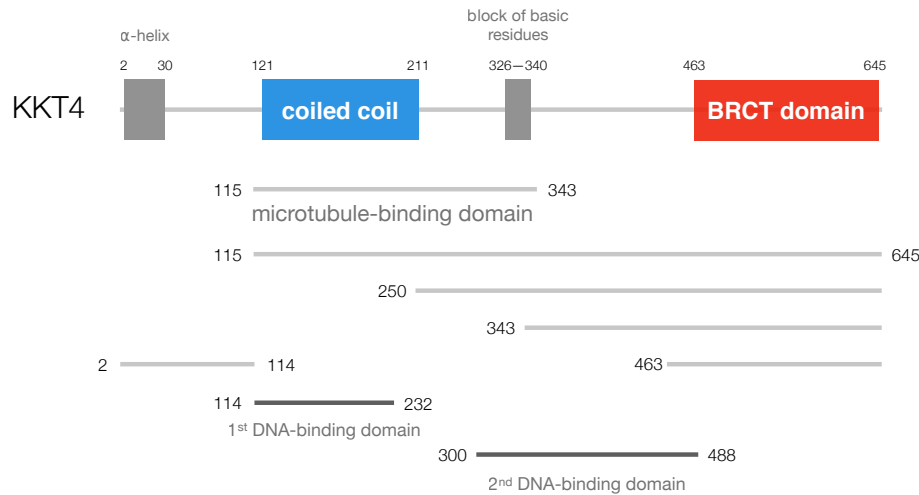


Figure 4.2. Cartoon representation of fragments used to study the DNA-binding activity of KKT4.

The C-terminal fragments of KKT4 bound DNA tightly (Figure 4.1B–D), suggesting that the N-terminal fragment of KKT4 (KKT4^{2–114}) is not essential for the DNA-binding activity. Indeed, KKT4^{2–114} did not bind DNA (Figure 4.1E). The C-terminal region of KKT4 has a predicted BRCT domain known to interact with DNA or RNA in other organisms (Leung and Glover, 2011). To test whether the BRCT domain in KKT4 is sufficient to bind DNA *in vitro*, KKT4^{463–645} was purified and tested. The KKT4 fragment did not saturate the DNA signal under the same conditions as the full-length protein (Figure 4.1F); this shows that the BRCT domain in KKT4 is either not able to bind DNA or binds DNA with significantly lower affinity compared to the full-length protein, which suggests that the DNA-binding domain is located elsewhere. In the end, two KKT4 fragments have been found that bind DNA tightly: KKT4^{114–232} and KKT4^{300–488}, with their binding affinities of 44 and 18 nM, respectively (Figure 4.1G, H). These fragments are referred to as the first and second DNA-binding domains.

Table 4.1. Summary of the fluorescence anisotropy assays.

KKT4 fragment	K _D [nM]
KKT4 ^{2–645}	11
KKT4 ^{115–645}	8
KKT4 ^{250–645}	14
KKT4 ^{343–645}	522
KKT4 ^{2–114}	-
KKT4 ^{463–645}	-
KKT4 ^{115–232}	44
KKT4 ^{300–488}	18
KKT4 ^{300–488} K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D	386
KKT4 ^{300–488} K310E, R311D, R313D	247
KKT4 ^{300–488} K326E, K327E, R328D	334
KKT4 ^{300–488} R456D, R457D, R460D	342

4.3. Mutations in the second DNA-binding domain reduce the affinity to DNA

Despite the fact that KKT4 has microtubule-binding activities, it can also strongly interact with DNA (Figure 4.1A). Mutating arginines and lysines within the microtubule-binding region would not allow the DNA- or microtubule-binding defects to be distinguished *in vivo*. Therefore, to test the relevance of the KKT4 DNA-binding activity *in vivo*, I aimed to identify residues important for the interaction within the second DNA-binding domain (KKT4^{300–488}) instead. KKT4^{300–488} did not interact with microtubules in the co-sedimentation assay (Figure 4.3).

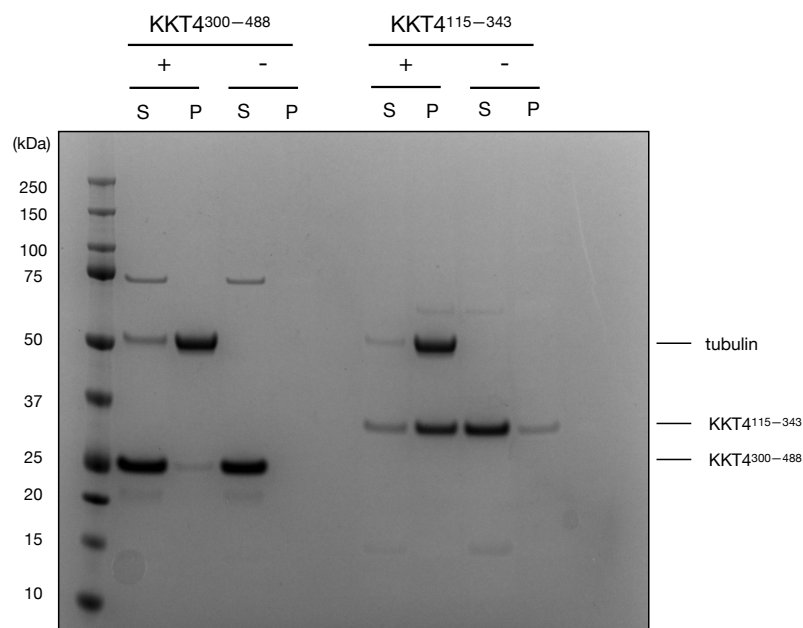


Figure 4.3. The KKT4^{300–488} does not interact with microtubules.

Microtubule co-sedimentation assay is shown for KKT4^{300–488} and KKT4^{115–343}. In contrast to the microtubule-binding domain, KKT4^{300–488} does not sediment with the taxol-stabilised microtubules.

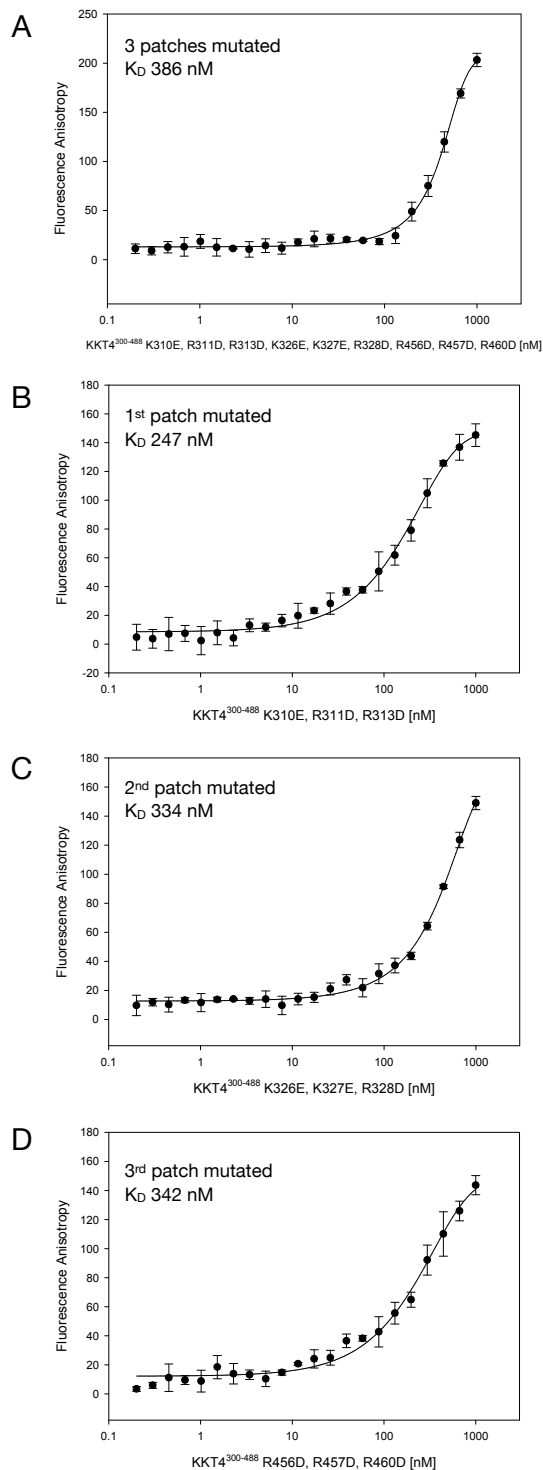


Figure 4.4. Mutations in the 2nd DNA-binding domain reduce DNA-binding affinity of the domain .

Measured anisotropy is plotted against protein concentrations of KKT4^{300–488} mutants (A–D). The mutated residues are reported below each graph. In these assays, 50-bp DNA probe was used, which is part of the centromeric sequence (CIR147) in *T. brucei*. The K_D values were calculated using non-linear regression using SigmaPlot (Monks, 2002); the fit is shown as a solid line. All mutants displayed reduced affinity to the DNA compared to the wild-type KKT4^{300–488}. Error bars are standard deviations from three independent measurements.

The isoelectric point (pI) of KKT4^{300–488} is 10.8 at physiological pH. Analysis of the KKT4^{300–488} primary sequence resulted in identification of three patches of positively charged residues; first (K310, R311, R313); second (K326, K327, R328) and third (R456, R457, R460) (Figure 1.7). To test the contribution of these positively-charged patches to the DNA-binding activity, four constructs were generated: one

where all of these positively charged residues have been mutated (KKT4^{300–488} K310E, R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D), and the remaining three where the first, second and third patch were mutated individually (Figure 4.4). All mutants displayed weaker affinity for DNA than the wild-type protein (~20-fold decrease) (Figure 4.4A–D). Mutations in the first, second and third patch reduced the K_D of the wild-type (18 nM) to 247, 334, 342 nM, respectively (Figure

4.4B–D). In summary, three patches of positively charged residues have been identified in KKT4^{300–488} and the mutations of these patches have shown reduction in affinity of KKT4 to DNA *in vitro*.

4.4. *In vivo* analysis of the DNA-binding defective KKT4 mutants

To test the relevance of KKT4 DNA-binding activities *in vivo*, *Trypanosoma brucei* cell lines with the RNAi-resistant form of YFP-tagged KKT4 were generated, where the second DNA-binding domain was either deleted or mutated in one of the KKT4 alleles (Table 2.14). Upon RNAi induction, all cell lines expressing the deletion or charge reversal mutants displayed severe growth defects (Figure 4.6A–C, left panels). However, YFP signal for all KKT4 mutants was significantly lower compared to the wild-type KKT4 (Figure 4.6A–C, right panels). It is likely that the introduced mutations affected the expression level of KKT4; therefore, no conclusion can be drawn from this experiment. More work is required to understand the importance of the DNA-binding activity of KKT4 *in vivo*.

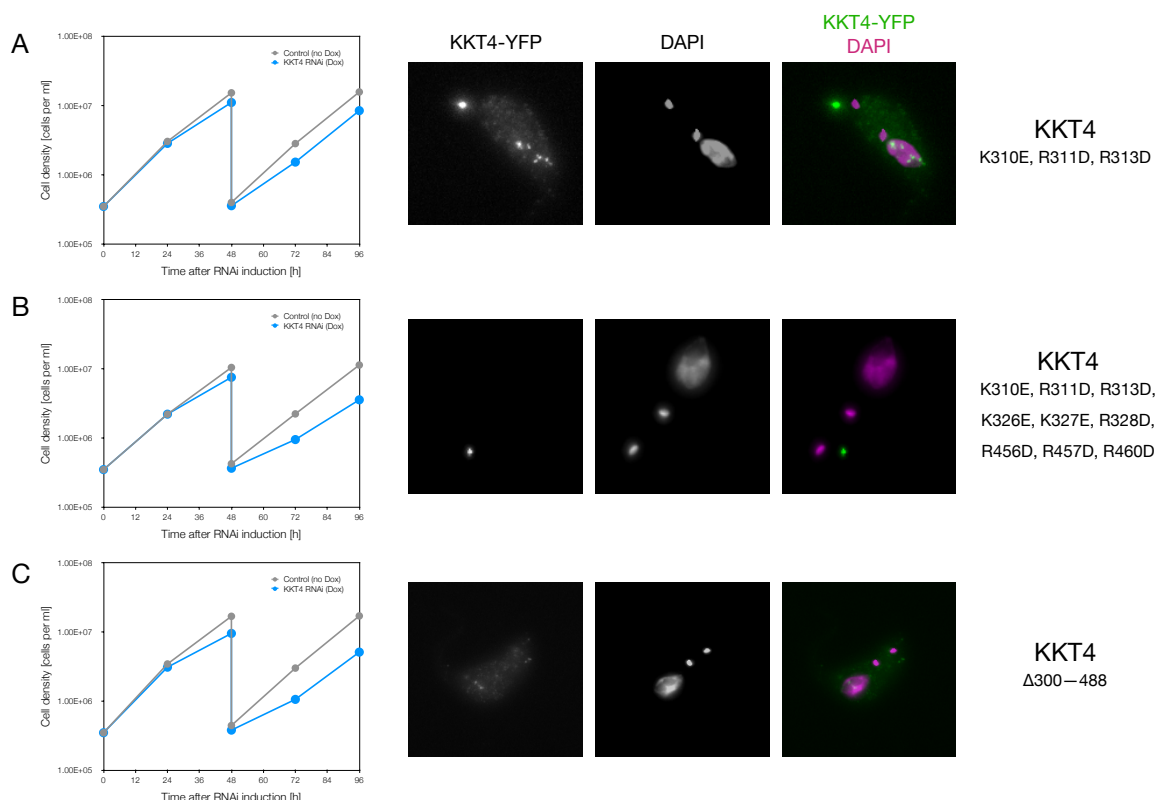


Figure 4.6. *In vivo* analysis of KKT4^{300–488} mutants.

The figure shows an *in vivo* analysis of *Trypanosoma brucei* cell lines that express KKT4 mutants: (A) K310E R311D, R313D; (B) K310E R311D, R313D, K326E, K327E, R328D, R456D, R457D, R460D and (C) Δ 300–488. The left panels show that upon the induction of RNAi, KKT4 mutants cannot support proper cell growth. However, the fluorescence microscopy images suggest that the expression of KKT4 mutants was lower compared to the wild-type KKT4. The observed growth defects in (A) are most likely the result of low levels of KKT4 inside the nucleus rather than functional effects of the mutated residues.

Chapter conclusions

- KKT4 interacts with DNA with an 11 nM K_D .
- The BRCT domain in KKT4 is not a DNA-binding domain.
- KKT4 has two fragments that bind DNA with high affinity, KKT4^{114–232} and KKT4^{300–488}.
- In contrast to the first DNA-binding domain, the second domain (KKT4^{300–488}) interacts with DNA but not microtubules.
- Three patches of positively charged residues were found in the second DNA-binding domain and their charge reversal mutations reduce KKT4^{300–488} affinity to the DNA.
- *In vivo* analysis of charge reversal mutants was not conclusive.

CHAPTER 5

5. Structure and function of the KKT4 BRCT domain

The analysis of the KKT4 primary sequence revealed a tandem BRCT domain at its C-terminus. The BRCT (BRCA1 C-Terminal) domains consist of ~90–100 amino acids with a $\beta\alpha\beta\beta\alpha\beta\alpha$ secondary structure topology (Leung and Glover, 2011). The BRCT domains are found in a number of prokaryotes and eukaryotes with various biological functions, including DNA- or RNA-binding, phosphorylation-dependent and independent interactions and poly(ADP-Ribose) binding (Leung and Glover, 2011). The KKT4 BRCT domain did not strongly interact with DNA (Figure 4.1F). In other eukaryotes, the BRCT domain has not been found in any of the known kinetochore proteins (Cheeseman, 2014, Musacchio and Desai, 2017), therefore, the structure and function of the BRCT domain in KKT4 cannot be inferred from other kinetochore proteins. To understand the function of the BRCT domain in the kinetoplastid kinetochores, I employed X-ray crystallography, NMR spectroscopy and a fluorescence anisotropy assay.

5.1. Crystal structure of the KKT4 BRCT domain

To determine the high-resolution structure of the KKT4 BRCT domain, X-ray crystallography was employed. The *Trypanosoma brucei* KKT4^{463–645} formed crystals that diffracted up to 1.6 Å resolution (Table 5.1). The overall structure consists of 6 α -helices and 8 β -strands, which form tandem BRCT domains composed of one full (BRCT1) and one partial domain (BRCT2) (Figure 5.1A). The N-terminal full domain is composed of a 4-stranded β -sheet central core, flanked by 2 α -helices on one side and 1 α -helix on the opposite side. A smaller C-terminal domain consists of 2 α -helices and 3 β -strands (missing 1 β -strands and 1 α -helix). In the linker between the two domains, a very short β -strand and a short α -helix can be observed. The electron density for residues 519–523 and 617–626 was not visible, suggesting that these regions are highly flexible or disordered. A search for structural homologs using DALI server (Holm, 2019) revealed similarity to several BRCT-containing proteins such as Topoisomerase II β binding protein 1 (TopBP1) and Breast cancer-associated 1 (BRCA1) as the top hits (Table S4). The tandem BRCT domains in these proteins have a highly conserved interface between the repeats, a central hydrophobic core known that function as a phosphopeptide-binding pocket (Shiozaki et al., 2004, Williams et al., 2004, Leung and Glover, 2011). Superposition of the KKT4 BRCT domain with the BRCT domain of *H. sapiens* BRCA1 showed a good structural match with an RMSD of 1.11 Å (for 65 Ca) (Figure 5.2), suggesting that the KKT4 BRCT domain may bind phosphopeptides. In fact, in the KKT4 BRCT structure, additional electron density was observed, arising from a sulphate ion located in the hydrophobic pocket, which can mimic a phosphate group (Figure 5.3). The sulphate ion is located within a basic surface of the BRCT domain (Figure 5.1C, B) and is coordinated by three residues: T494, S495 and K543 (Figure 5.3); these correspond to conserved amino acids, critical for interaction with phosphopeptides

in other eukaryotic BRCT domains (Williams et al., 2004). These results suggest that the KKT4 BRCT domain is likely to be a phospho-dependent protein-protein interaction domain.

Table 5.1. Data collection and refinement statistics*.

Data collection	<i>T. brucei</i> KKT4 ^{463–645}	<i>T. brucei</i> KKT4 ^{463–645} K543A
Beamline	Diamond Light Source I24	Diamond Light Source I03
Wavelength (Å)	0.969	0.9763
Space group (Z)	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Unit cell	46.37Å 61.63Å 67.78Å 90° 90° 90°	46.66Å 61.34Å 67.75Å 90° 90° 90°
Resolution range (Å)	45.60–1.57 (1.63–1.57)	37.14–1.80 (1.86–1.80)
Unique reflections	27032 (2137)	18590 (1818)
Completeness (%)	96.8 (74.7)	99.95 (99.89)
Multiplicity	10.0 (3.8)	12.0 (1.2)
I/σI	14.9 (3.1)	89.9 (0.2)
R _{meas}	0.089 (0.549)	0.030 (1.580)
CC _{1/2}	0.997 (0.496)	1.0 (0.6)
Wilson B-factor (Å ²)	15.7	22.17
Refinement		
No. reflections	26945 (2109)	18590 (1818)
R _{work}	0.173 (0.219)	0.199 (0.197)
R _{free}	0.192 (0.303)	0.249 (0.257)
Number of atoms	1494	1370
Protein	1270	1180
Solvent	224	190
RMS bonds (Å)	0.011	0.007
RMS angles (°)	1.46	0.96
Ramachandran favoured (%)	98.68	96.60
Ramachandran allowed(%)	1.32	3.40
Ramachandran outliers (%)	0.00	0.00
Average B-factor (Å ²)	22.0	26.79

*Statistics for the highest-resolution shell are shown in parentheses.

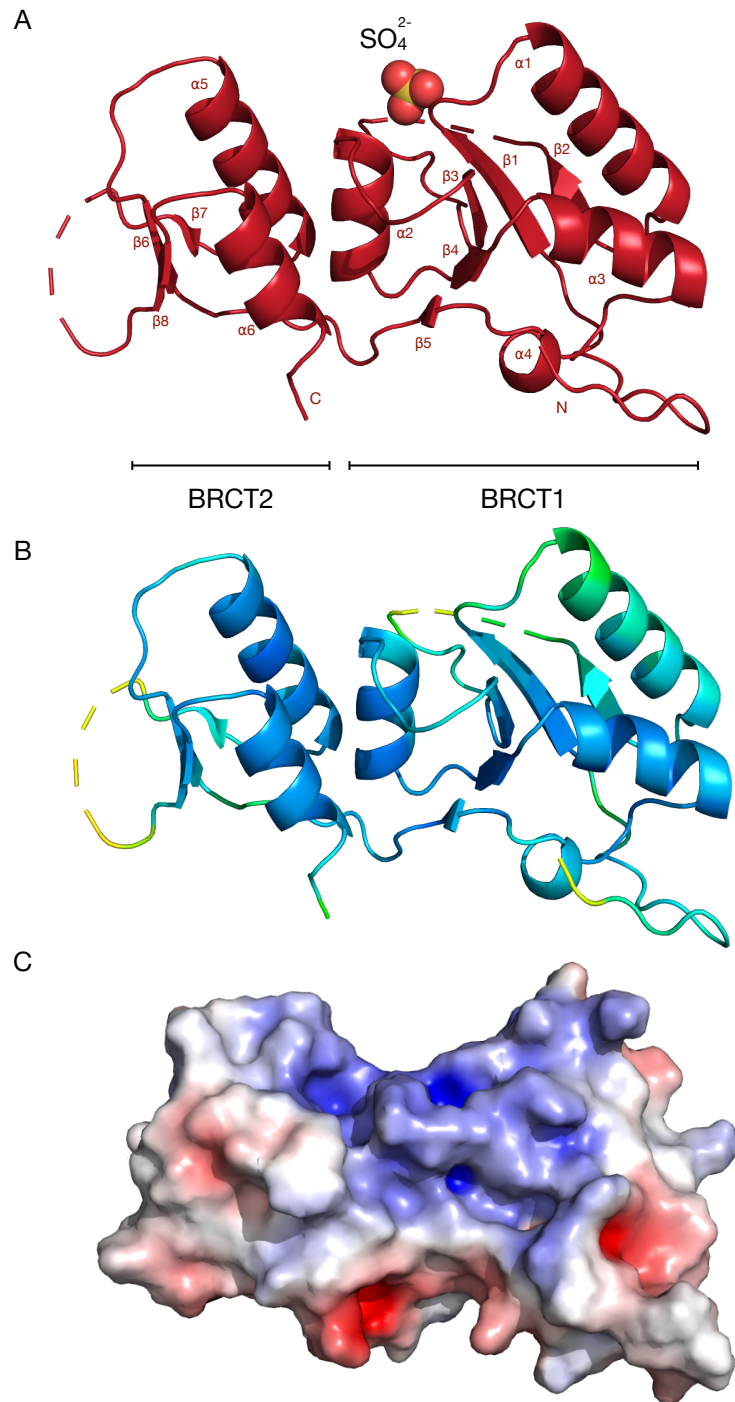


Figure 5.1. Crystal structure of *Trypanosoma brucei* KKT4^{463–645}.

- (A) Ribbon model of *Trypanosoma brucei* KKT4^{463–645}. The N and C represent the N and C-terminus, respectively. Residues for which the electron density was not observed are shown as dotted lines. The sulphate ion is represented as spheres.
- (B) B-factors for Cα atoms have been represented for TbKKT4^{463–645} using a blue to red spectrum indicating low to high values, respectively. The figures were rendered using PyMol (DeLano, 2002).
- (C) The surface electrostatic potential is coloured from red to blue, -5 kbT to +5 kbT, as calculated by APBS electrostatic plugin in Pymol (Jurrus et al., 2018).

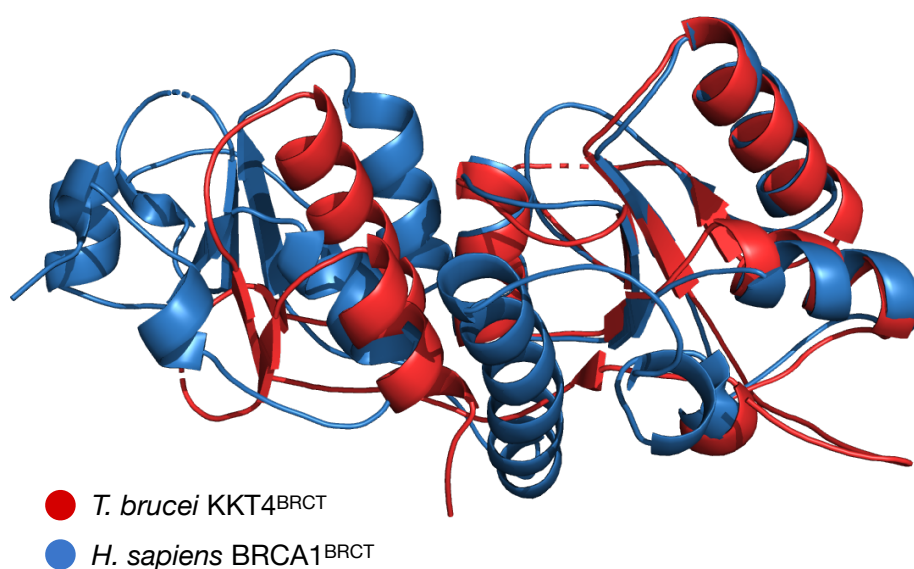


Figure 5.2. The alignment between the *Tb*KKT4 and *Hs*BRCA1 BRCT domains.

Superposition between the BRCT domain crystal structures of KKT4 (red, PDB: 6ZPK) and BRCA1 (blue, PDB: 3FA2). The alignment clearly shows that BRCT2 in KKT4 is only a partial domain compared to the domain in BRCA1. The BRCT2 in KKT4 is missing missing 1 β -strand and 1 α -helix. The RMSD of 1.11 Å was calculated using the *super* function in PyMOL for the N-terminal domain only (DeLano, 2002).

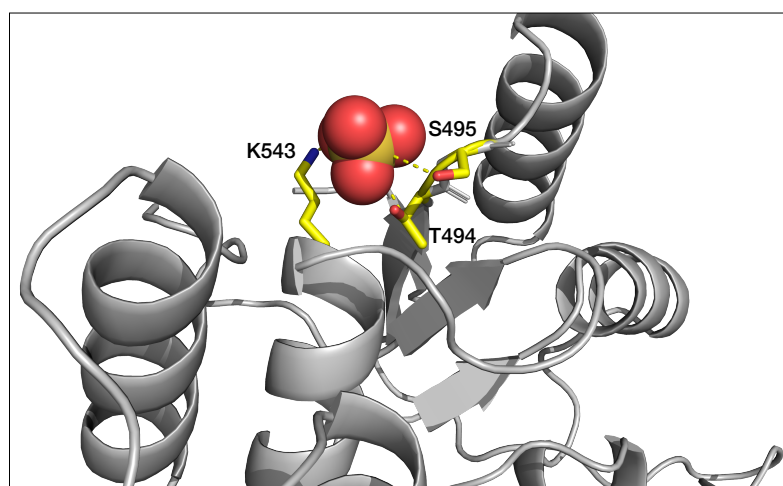


Figure 5.3. Sulphate ion is coordinated by T494, S495 and K543.

Close-up view showing the coordination of a sulphate ion by T494, S495, and K543. The side chains of these residues are shown as yellow sticks.

5.2. NMR analysis of the KKT4 BRCT domain

The crystal structure of KKT4^{463–645} revealed that the C-terminal region of KKT4 has tandem BRCT domains (Figure 5.1A). To characterise the protein in solution, NMR backbone and side-chain assignments of KKT4^{463–645} were carried out, followed by the structural and functional analysis using NMR spectroscopy.

Resonance assignments of the BRCT domain

The BRCT domain (KKT4^{463–645}) consists of 184 residues, including 12 proline residues, for which no signals are observed in a BTROSY spectrum. In the ¹H-¹⁵NH BTROSY spectrum of the KKT4 BRCT domain, approximately 200 peaks can be observed (Figure 5.4). This is more than expected for a protein containing 172 non-proline residues. Additional peaks may be the result of residues that are adjacent to proline residues that undergo cis/trans isomerisation. To obtain the complete backbone and side chain assignments, the triple resonance data were acquired and analysed (Table S1). Out of 172 non-proline residues, 159 were assigned to specific peaks in the BTROSY spectrum (Figure 5.4). The detailed assignments statistics are shown in Table 5.2.

Table 5.2. Extent of assignment for KKT4 BRCT domain.

Sample	Percent assigned						
	¹ H ^N / ¹⁵ N	¹³ C	¹ H α / ¹³ C α	¹ H β / ¹³ C β	¹ H γ / ¹³ C γ	¹ H δ / ¹³ C δ	¹ H ϵ / ¹³ C ϵ
KKT4 ^{463–645}	100/92.4	98.8	79.5/99.4	66.8/98.1	57.0/50.3	47.2/40.4	19.0/18.5

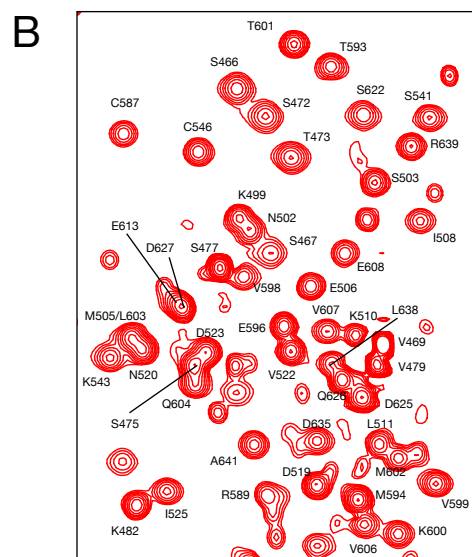
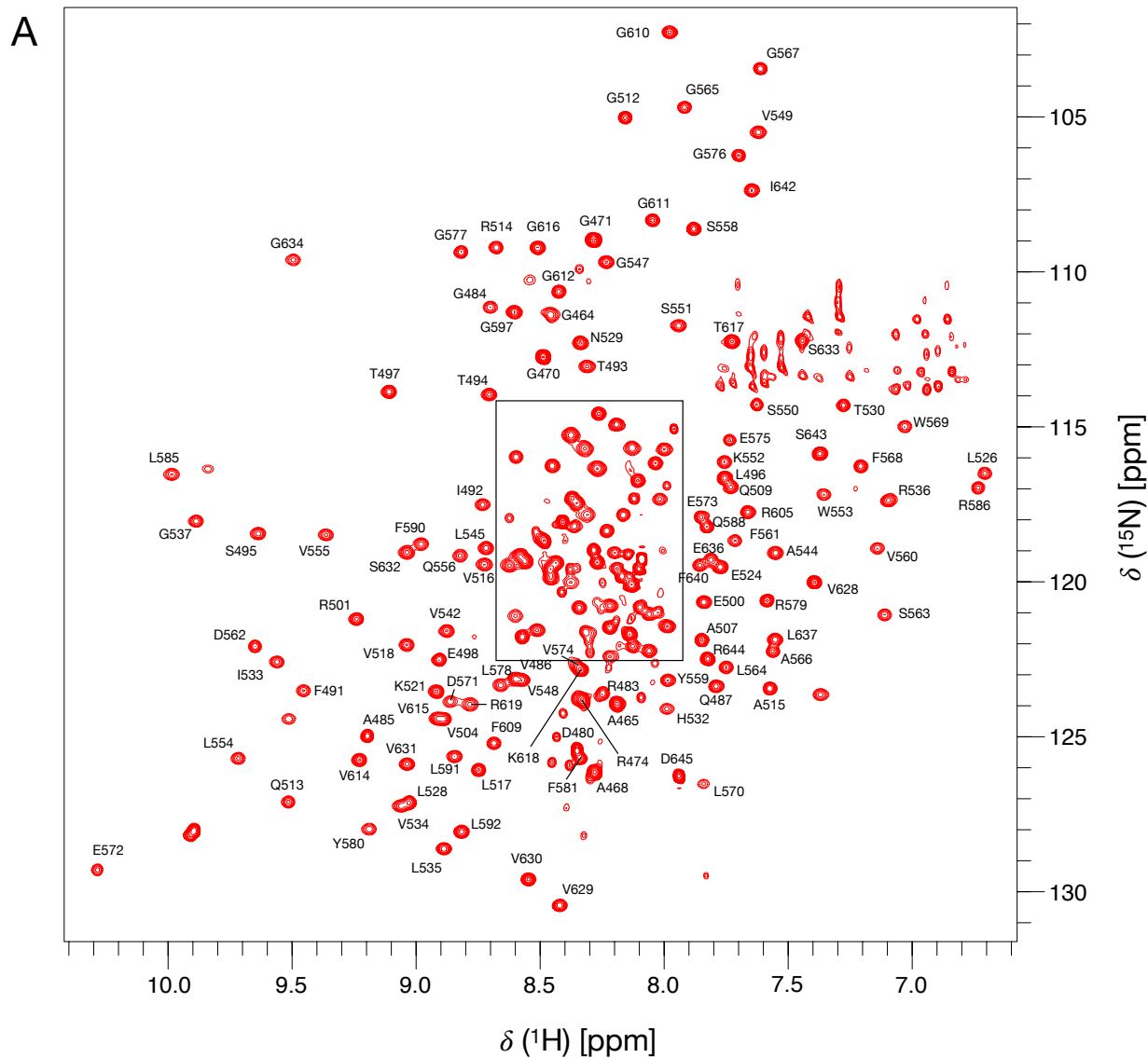


Figure 5.4. The ^1H - ^{15}N BEST-TROSY spectrum of KKT4⁴⁶³–645.

- (A) 750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT4⁴⁶³–645 in 50 mM sodium phosphate, 100 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.0, 20°C. The peak assignments for backbone amides are annotated.
- (B) The expansion of the crowded region highlighted with a black box in (A).

Secondary structure

Chemical shifts obtained for the KKT4^{463–645} backbone and side chains have been analysed using TALOS-N. The BRCT domain is a mixture of 6 α -helices and 8 β -strands (Figure 5.5A). The secondary structure elements identified using NMR are in excellent agreement with the elements observed in the crystal structure. However, there are some minor differences, including a β -strand at residues 578–580, shown to be shorter in the crystal structure (Figure 5.5B).

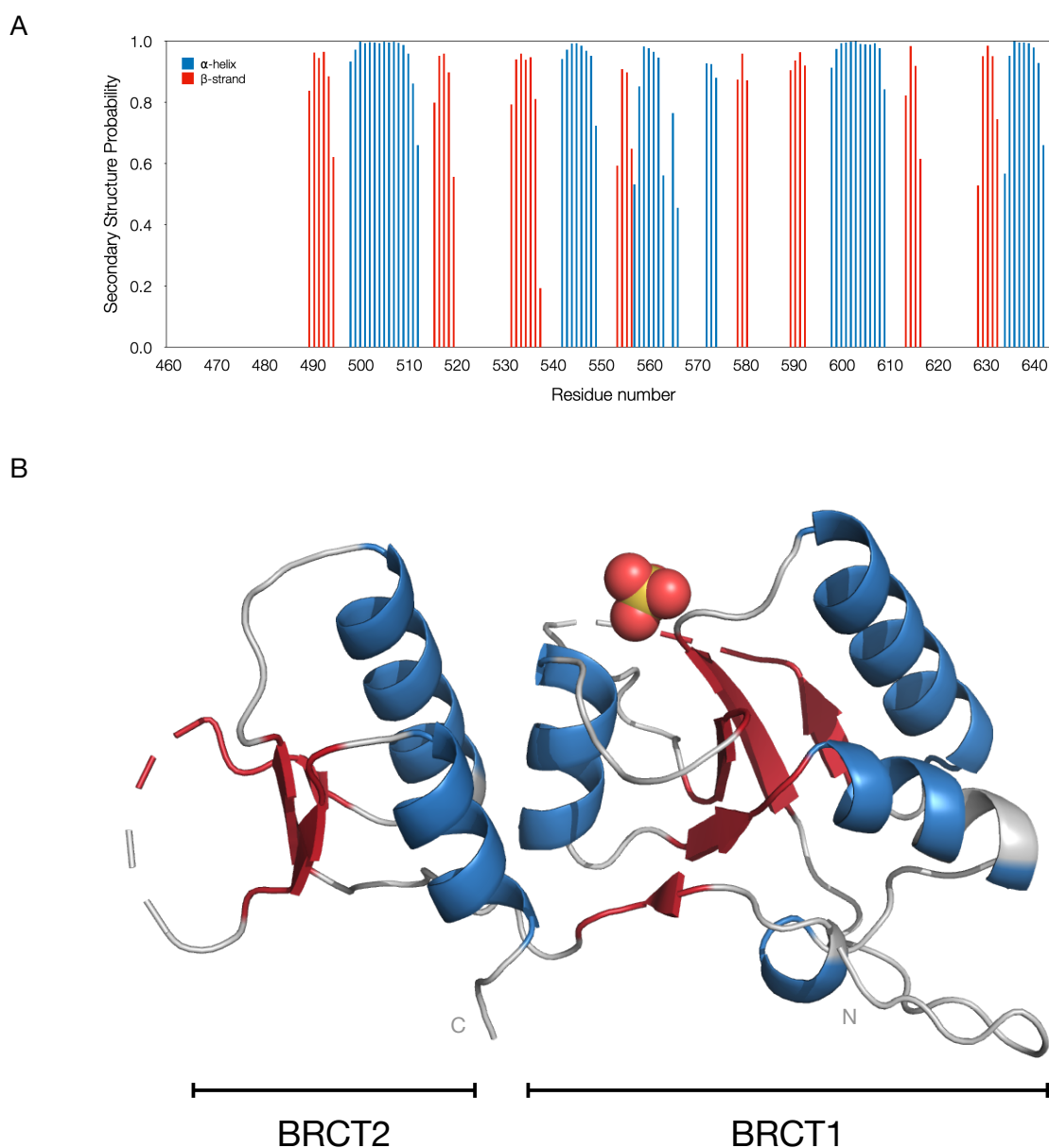


Figure 5.5. NMR analysis of KKT4^{463–645} secondary structure.

- (A) TALOS-N secondary structure analysis of KKT4^{463–645}. The prediction shows that the BRCT domain in KKT4 is composed of 6 α -helices and 8 β -strands. The secondary structure elements for which the probability was lower than 0.5 are not shown in the figure.
- (B) The secondary structure elements identified using TALOS-N (A) are coloured in the crystal structure of KKT4^{463–645}. The secondary structure observed in the crystal structure are in excellent agreement with the TALOS-N prediction. The α -helices and β -strands are coloured in blue and red, respectively.

Backbone dynamics

To study the BRCT domain backbone dynamics, the ^1H - ^{15}N heteronuclear NOE ratios were measured for KKT4^{463–645} (Figure 5.6). Peaks located in the poorly resolved regions were not included as the calculation of hetNOE ratios of the overlapping peaks is inaccurate. Most of the residues have ^1H - ^{15}N hetNOE intensity ratios of >0.7 , indicating a rigid backbone. The regions between 519–525 and 616–628 display lower ^1H - ^{15}N hetNOE ratios (<0.7), indicating higher flexibility in these regions, consistent with the crystal structure, in which the residues between 519–523 form a loop, while no electron density is observed for region 617–625. The N-terminal region of the BRCT domain (464–483) is highly flexible.

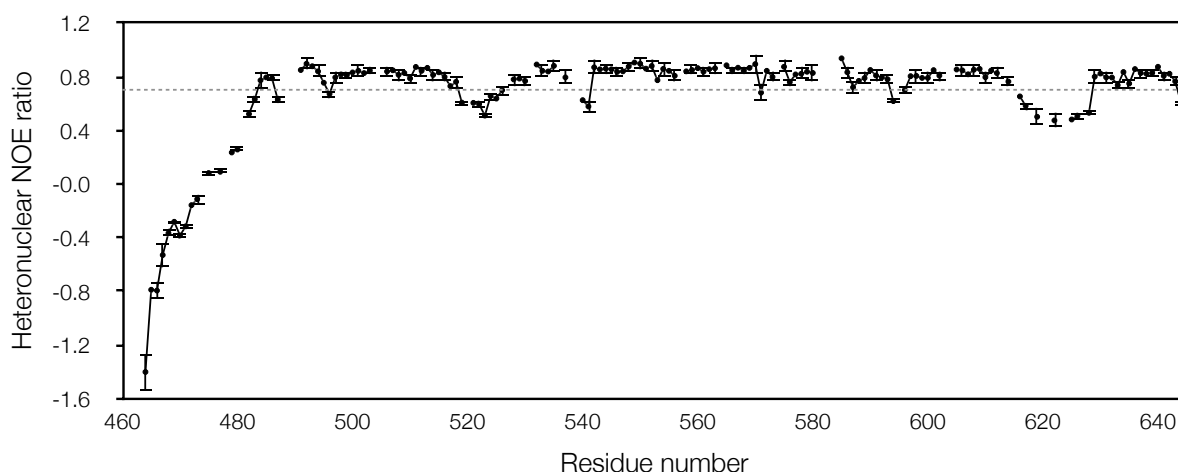


Figure 5.6. The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE measurements for KKT4^{463–645}.

The $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE ratios were measured and plotted against the sequence of KKT4^{463–645}. Most of the residues display hetNOE ratios >0.7 , indicating a rigid conformation of the BRCT domain backbone. The regions with higher flexibility (ratios <0.7), correspond to 464–483, 519–525 and 616–628. Consistent with the hetNOE analysis, no electron density was observed for these regions in the crystal structure of the BRCT domain. The $\{^1\text{H}\}$ - ^{15}N NOE errors were estimated from 500 Monte Carlo simulations using baseline noise as a measure of peak height error.

5.3. The KKT4 BRCT domain is a phosphopeptide-binding domain

The BRCT domains in other organisms are known to recognise the pS/pT-X-X-(F/Y/I/L) consensus motifs (Manke et al., 2003, Yu et al., 2003). To identify potential binding partners for the KKT4 BRCT domain, sequences of kinetochore proteins were searched for the motif mentioned above. Among those proteins that co-purified with KKT4 from the trypanosome culture (Akiyoshi and Gull, 2014), possible motifs were identified in KKT7 (SVTF, residues 65–68), KKT8 (SVRY, residues 381–384) and KKT12 (SILL, residues 192–195). All identified motifs are highly conserved among kinetoplastids (Figure 5.7A–C).

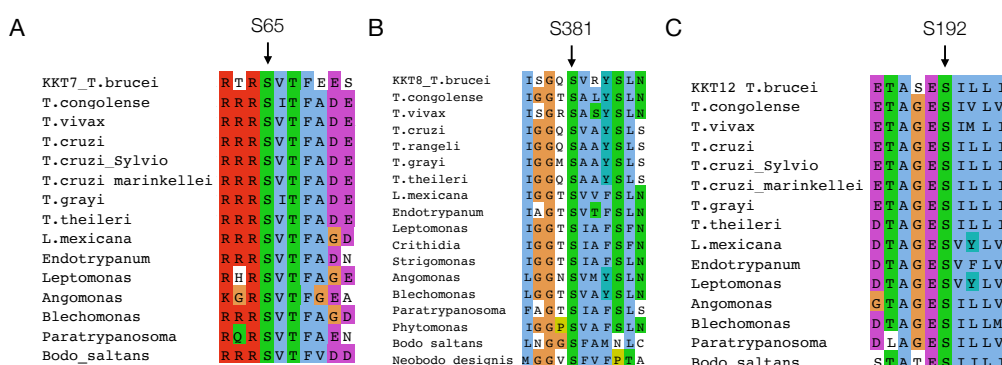


Figure 5.7. KKT7, KKT8 and KKT12 contain consensus sequences bound by the BRCT domains.

Multiple sequence alignments of (A) KKT7, (B) KKT8 and (C) KKT12 sequence regions that match the consensus sequence (pS/pT-X-X-Y/F/I/L) recognised by BRCT domains in other organisms (Leung and Glover, 2011). Based on these sequences, phosphorylated peptide probes were designed and used in the fluorescence anisotropy assay shown in Figure 5.8.

Fluorescently-labelled phosphopeptides derived from these proteins were tested for BRCT-binding using the fluorescence anisotropy assay. The peptide derived from KKT8 bound KKT4 BRCT domain with a K_D of $\sim 30 \mu\text{M}$ (Figure 5.8). The other two peptides failed to bind with a similar affinity (data not shown). Importantly, a replacement of K543 located in the putative phosphopeptide-binding site with alanine, decreased the binding affinity by roughly an order of magnitude (Figure 5.8). Similar results were obtained using a non-phosphorylated KKT8 peptide, showing that the phosphorylation increases the affinity of this interaction (Figure 5.8).

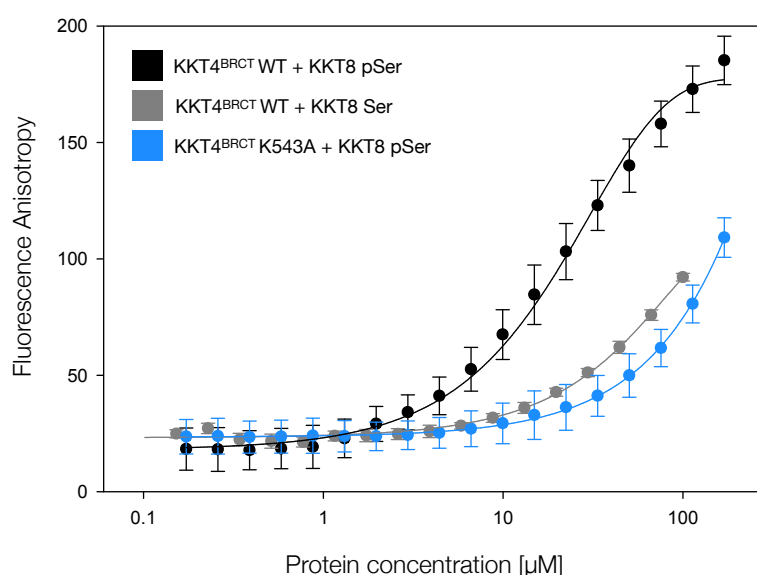


Figure 5.8. KKT4 BRCT domain binds the phosphopeptide derived from KKT8.

Fluorescence anisotropy assay using a fluorescently labelled KKT8 phosphopeptide (DICGISGQ(pS)VRYSLND) is shown for KKT4^{463–645} wild-type (black) and KKT4^{463–645} K543A mutant (blue). The binding with the non-phosphorylated KKT8 probe (DICGISGQSVRYSLND) is shown in grey. The K_D of $30 \mu\text{M}$ for the wild-type BRCT domain was calculated using non-linear regression using SigmaPlot (Monks, 2002); the fits are shown as solid lines. Error bars are standard deviations from three independent measurements.

Endogenous mutation of K543A was shown to affect the growth of trypanosomes (H. Hayashi, in preparation). Moreover, the overexpressed KKT4^{343–645} with K543A mutation failed to localise at the kinetochore in contrast to the wild-type KKT4 fragment (H. Hayashi, in preparation). K543 thus seems to be an important residue for KKT4 function. To confirm that the defects observed with the K543A mutant were not due to the altered BRCT domain structure, the crystal structure of KKT4^{463–645} K543A was determined to 1.8 Å resolution (Table 5.1, Figure 5.9). The structure containing the K543A mutation confirmed that the folding and the conformation of the BRCT domain remained intact, which is shown with the structural alignment between the wild-type and K543A mutant (Figure 5.10).

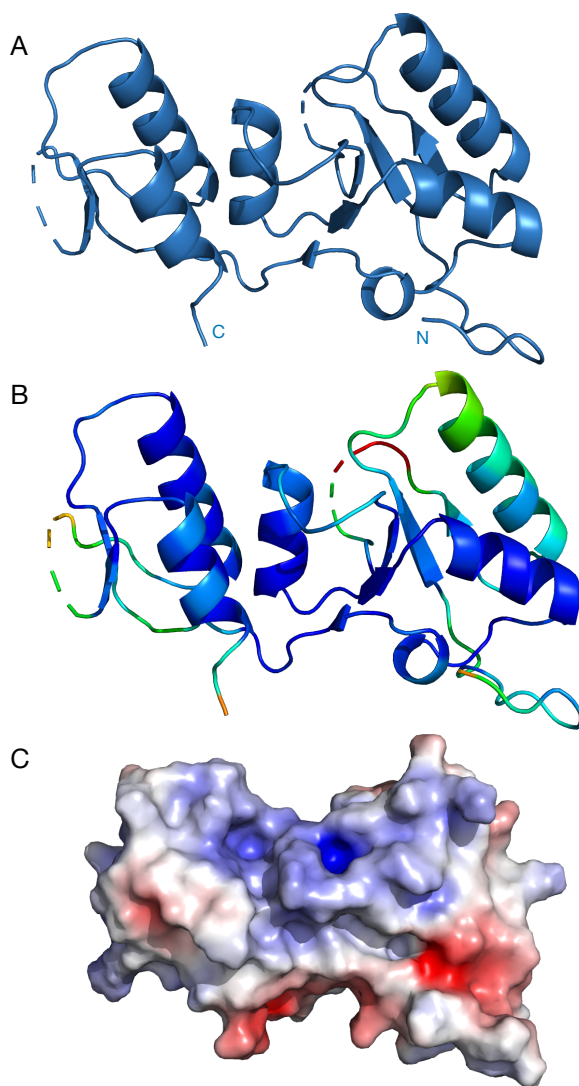


Figure 5.9. Crystal structure of *Trypanosoma brucei* KKT4^{463–645} K543A mutant.

- (A) Ribbon model of *Trypanosoma brucei* KKT4^{463–645} K543A. The N and C represent the N and C termini, respectively. Residues for which the electron density was not observed are shown as dotted lined.
- (B) B-factors for Cα atoms have been represented for *Tb*KKT4^{463–645} K543A using a blue to red spectrum indicating low to high values, respectively. The figures were rendered using PyMol (DeLano, 2002).
- (C) The surface electrostatic potential is coloured from red to blue, -5 kbT to +5 kbT, as calculated by APBS electrostatic plugin in Pymol (Jurrus et al., 2018).

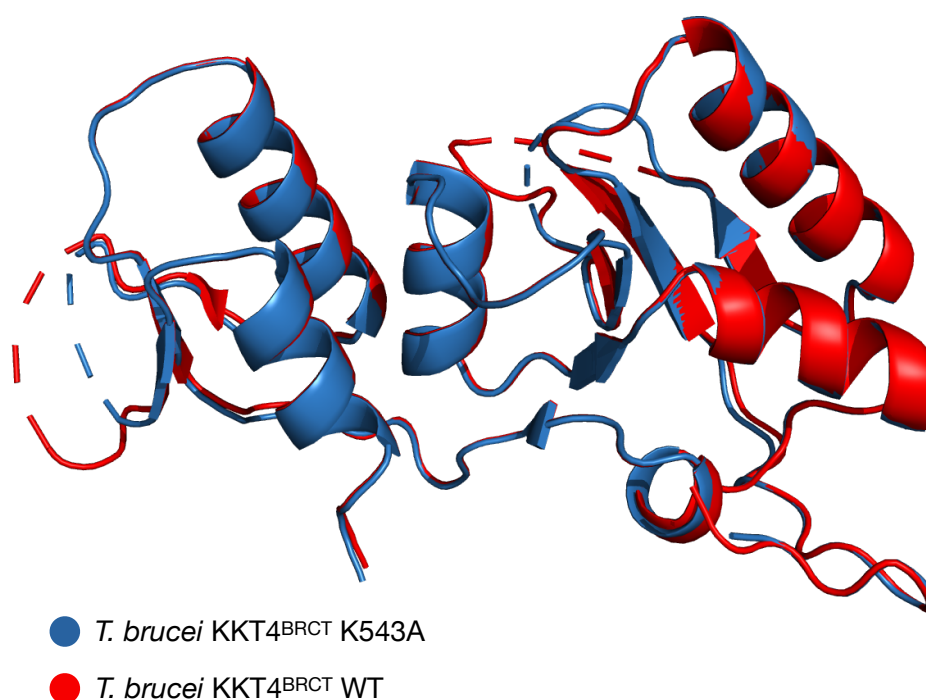


Figure 5.10. An overlay between the crystal structures of the wild-type BRCT domain and the K543A mutant.

Superposition of the wild-type KKT4 BRCT domain (red, PDB: 6ZPK) with the K543A mutant (blue). The structures show an excellent structural match suggesting that K543A mutation did not perturb the BRCT domain structure. It is noteworthy that the orientation of loops differ between the two structures, suggesting they are flexible.

To confirm the results obtained using a fluorescence anisotropy assay, the interaction between the BRCT domain and KKT8 peptide was studied using NMR. Comparison of the 2D ¹H-¹⁵N BTROSY spectrum of KKT4^{463–645} with and without KKT8 phosphopeptide revealed several residue-specific changes; these include I492, T493, T494, T497, L496, L517, V518, K521, I525, L526, G537, K543, A544, C546, G547, R579, T593, T601, E636, R639, I642 and R644 (Figure 5.11A–C). Most of these residues are located in or close to the hydrophobic groove shown to make contacts with sulphate ion in the crystal structure. Taken together, the fluorescence anisotropy assay and the NMR analysis show that KKT4 BRCT domain is a phosphopeptide-binding domain that interacts with KKT8 phosphopeptide.

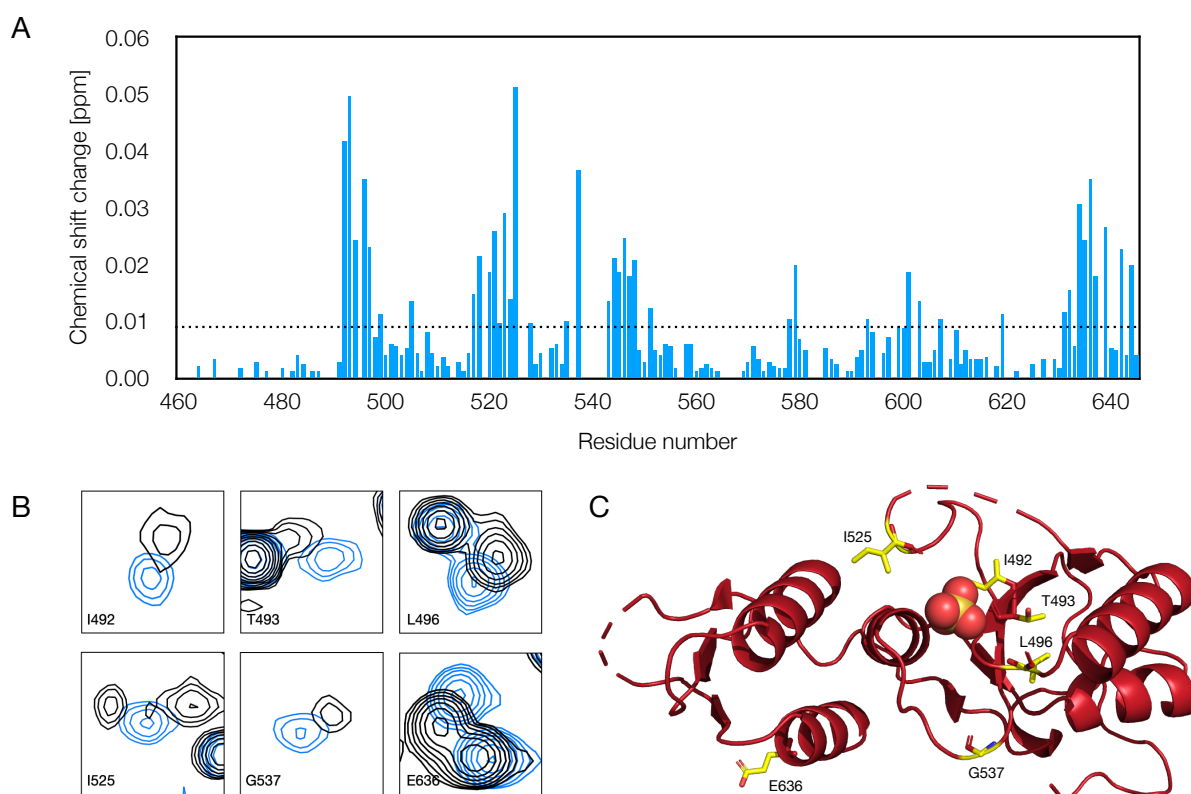
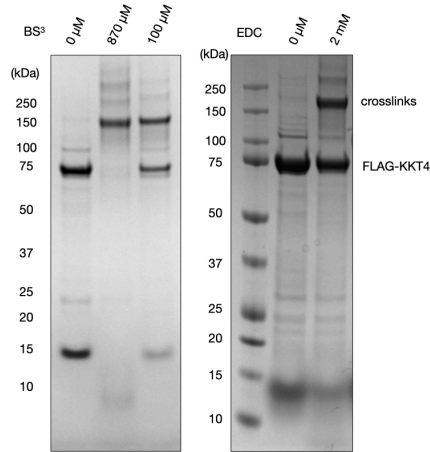


Figure 5.11. The NMR analysis of the interaction between the BRCT domain and the KKT8 peptide.

- (A) Chemical shift changes observed upon the addition of KKT8 phosphopeptide (DICGISGQ(pS)VRYSLND) to ^{15}N -KKT4⁴⁶³⁻⁶⁴⁵ have been plotted against the BRCT domain sequence. The dotted line represents the average value (0.009 ppm) of all observed chemical shift changes. The largest shifts (>0.03) are observed for residues I492, T493, L496, I525, G537 and E636. The position of these residues is shown in (C).
- (B) Expansion of peaks that showed chemical shift changes upon addition of the KKT8 peptide to the sample. Peaks coming from ^{15}N -KKT4⁴⁶³⁻⁶⁴⁵ and ^{15}N -KKT4⁴⁶³⁻⁶⁴⁵/KKT8 peptide spectra are coloured in black and blue, respectively.
- (C) Top view of the KKT4 BRCT domain crystal structure. Residues for which the biggest chemical shift changes were observed (>0.03 ppm) are shown as yellow sticks. The highlighted side chains are located around the hydrophobic groove known to be the phosphopeptide-binding site in other organisms (Leung and Glover, 2011). The regions for which the electron density is not visible are shown as dotted lines.

A



B

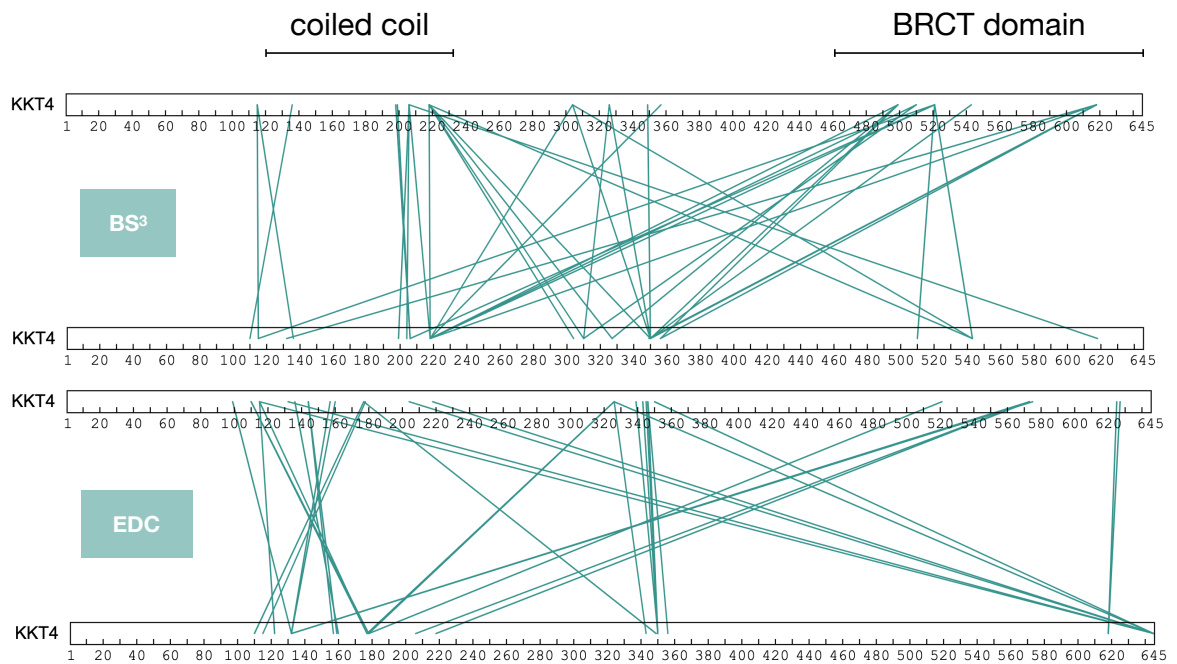


Figure 5.12. Crosslinking mass spectrometry of the full-length KKT4.

- (A) The SDS-PAGE gel showing the resolved crosslinking reactions performed on the full-length KKT4 with BS³ (left) and EDC (right).
- (B) Crosslinking mass spectrometry of the full-length KKT4 using BS³ and EDC/Sulfo-NHS. The green lines indicate pairs of the crosslinked residues. For clarity, two molecules of KKT4 are shown. It is noteworthy that crosslinks between the top and bottom KKT4 do not necessarily mean that crosslinks formed between two separate molecules because it was not possible to distinguish between inter- and intra-molecule crosslinks in this experiment (except for those inter-molecule crosslinks formed between the same residues). A complete list of identified crosslinks is shown in Tables S3.

5.4. BRCT domain interacts with KKT4 microtubule-binding region

To obtain additional structural information on KKT4, crosslinking mass spectrometry was used, where the full-length KKT4 was crosslinked using BS³, a homobifunctional crosslinker that covalently links pairs of lysines that are within 26–30 Å on the protein surface. XL-MS of KKT4 resulted in numerous crosslinks across the molecule (Figure 5.12A–B). The KKT4 fragment near the N-terminus does not contain lysines. Therefore, to increase the coverage of the crosslinked regions, a different type of crosslinker was used, a zero-length EDC and Sulfo-NHS that activates carboxyl groups for reaction with primary amines. The pattern of crosslinks was similar to that obtained with BS³ (Figure 5.12B). Interestingly, a number of crosslinks were identified between the BRCT domain and the microtubule-binding domain using both BS³ (K115/K499, K115/K521, K115/K543, K132/K618, K199/K521, K206/K521, K218/K499, K218/K510, K218/K521, K218/K618) and EDC/Sulfo-NHS (K115/D645, K132/E575, K132/D645, E178/K521, K204/D645, K206/E573, K206/D645, K218/E573, K218/D645). These results suggested that the KKT4 BRCT domain interacts directly with the microtubule-binding domain. The complete list of crosslinked residues is shown in Table S3.

To confirm that the BRCT domain interacts with the microtubule-binding domain, the 2D ^1H - ^{15}N BEST TROSY spectrum of ^{15}N -KKT4^{115–232} mixed with the unlabelled KKT4^{463–645} was collected. Comparison of the KKT4^{115–232} spectrum with and without KKT4^{463–645} revealed several residue-specific changes (Figure 5.13A, B). The peaks that showed the largest shifts correspond to K115, Y116, G117, V118, V119, S120, V121, E122 and R123 (Figure 5.13A, B). It is interesting that one of the residues that shifted, K115, was shown to crosslink with K543 from the BRCT domain. Crosslinking mass spectrometry revealed additional interaction sites at residues E178, K199, K204, K206 and K218. These residues did not give observable peaks in the spectrum of KKT4^{115–232}, making it impossible to confirm the interaction by NMR. The experiment with a shorter KKT4 construct, KKT4^{115–174}, identified the same interaction sites as in the experiment with KKT4^{115–232} (Figure 5.13A). No significant changes in chemical shifts were observed when KKT4^{145–232} was mixed with KKT4^{463–645} (Figure 5.13A). These results suggest that the interaction between the BRCT domain and the microtubule-binding domain of KKT4 depends on residues near the N-terminus of the microtubule-binding domain. This interaction could modulate the microtubule-binding activity of KKT4. To test this, KKT4^{115–645} that contains both the microtubule-binding and BRCT domains was purified. KKT4^{115–645} was shown to interact with microtubules with a similar affinity compared to the KKT4^{115–343} (Figure 5.13C, D), suggesting that the BRCT domain does not enhance or destabilise interaction between KKT4 and taxol-stabilised microtubules. In the future, it will be interesting to analyse whether the observed interaction can regulate other activities of KKT4.

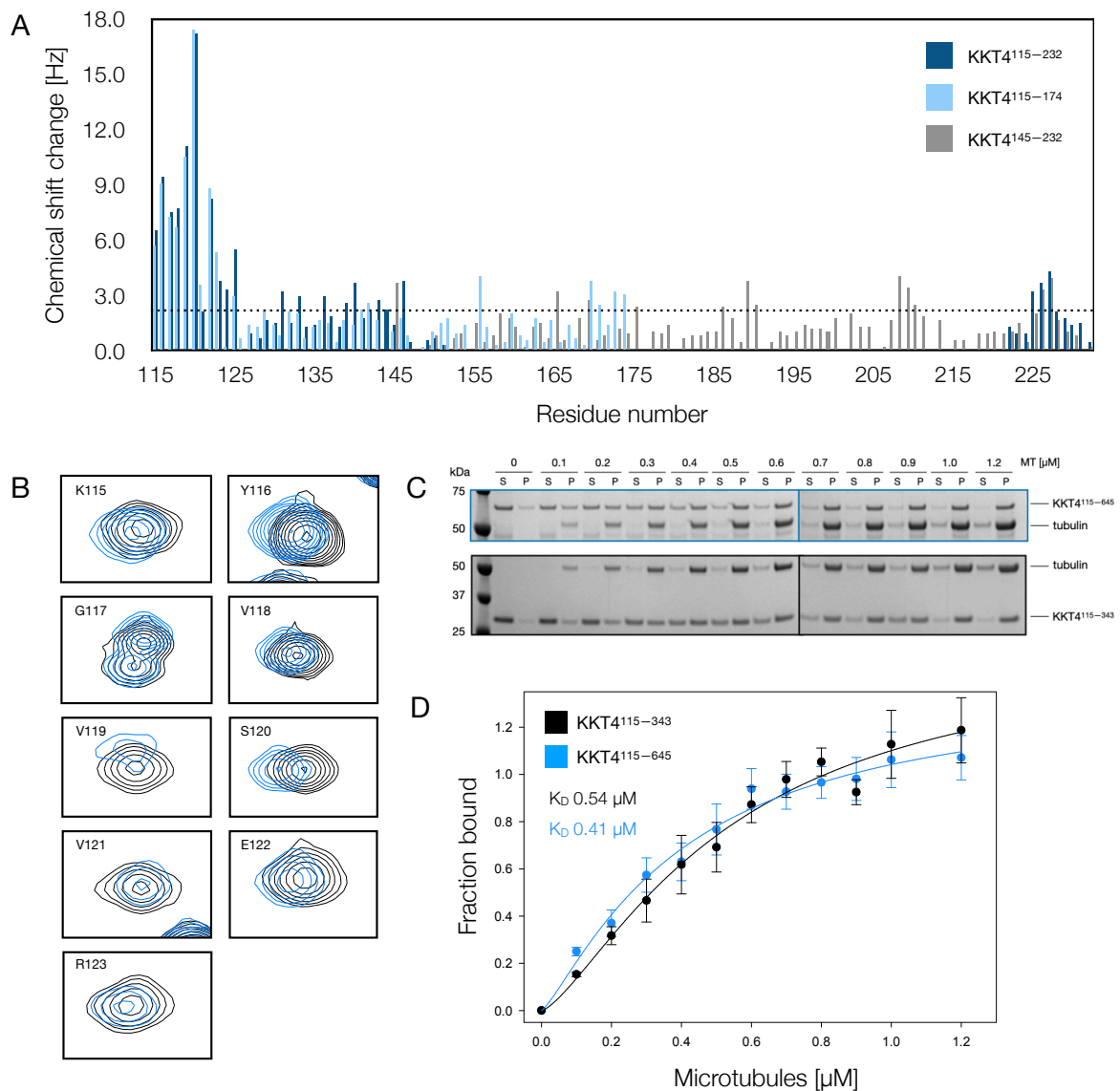


Figure 5.13. The N-terminal region of the microtubule-binding domain directly interacts with the BRCT domain.

- (A) Chemical shift changes observed upon addition of KKT4⁴⁶³⁻⁶⁴⁵ to KKT4¹¹⁵⁻²³² (dark blue), KKT4¹¹⁵⁻¹⁷⁴ (light blue), or KKT4¹⁴⁵⁻²³² (grey) have been plotted against the KKT4 sequence. The dotted line represents the average value (2.2 Hz) of all chemical shift changes observed.
- (B) Expansion of peaks that showed chemical shift changes upon addition of the KKT4 BRCT domain to the sample. Peaks coming from ¹⁵N-KKT4¹¹⁵⁻²³² and ¹⁵N-KKT4¹¹⁵⁻²³²/KKT4⁴⁶³⁻⁶⁴⁵ spectra are coloured in black and blue respectively.
- (C) Microtubule co-sedimentation assay of KKT4¹¹⁵⁻⁶⁴⁵ and KKT4¹¹⁵⁻³⁴³, showing that the presence of the BRCT domain does not significantly affect the affinity to microtubules. S and P correspond to supernatant and pellet fractions, respectively.
- (D) Plot showing the fraction bound against the concentration of microtubules between KKT4¹¹⁵⁻⁶⁴⁵ and KKT4¹¹⁵⁻³⁴³. Error bars are standard deviations from three independent measurements.

Chapter conclusions

- X-ray crystallography and NMR spectroscopy show that the C-terminal region of KKT4 has the BRCT domain that act as a phosphopeptide-binding domain.
- The KKT4 BRCT domain binds phosphopeptide derived from KKT8 and make direct contacts with the N-terminal region of the microtubule-binding domain.
- The BRCT domain does not modulate the microtubule-binding activity of KKT4.

CHAPTER 6

6. Structure and function of KKT23

KKT23 is an unconventional kinetochore protein with an unknown function. Its C-terminal region is predicted to be a GNAT domain (Figure 1.8A); therefore, it is possible that KKT23 acts as an acetyltransferase at the kinetochore. Acetyltransferases are not known to be part of the structural kinetochore in other eukaryotes (Cheeseman, 2014, Musacchio and Desai, 2017), making the function of KKT23 enigmatic. To understand the function of KKT23 at the kinetochore, I aimed to obtain detailed structural information on the protein using X-ray crystallography and NMR spectroscopy. Finally, I used an *in vitro* acetylation assay to characterise the acetyltransferase activity of KKT23.

6.1. The crystal structure of the KKT23 GNAT domain bound to acetyl coenzyme A

The analysis of the KKT23 primary sequence revealed a predicted helical region near the N-terminus (4—66), which is then followed by a stretch of unstructured residues (67—128). The C-terminal fragment of KKT23 (146—298) is predicted to be a mixture of α -helices and β -strands, consistent with the GNAT domain prediction in that region (Figure 1.8B, C). To obtain basic structural information, SEC-MALS and AUC were employed to determine the oligomerisation state of the full-length KKT23 protein (KKT23^{2–348}), which was found to be monomeric using both methods (Figure 6.1A, B).

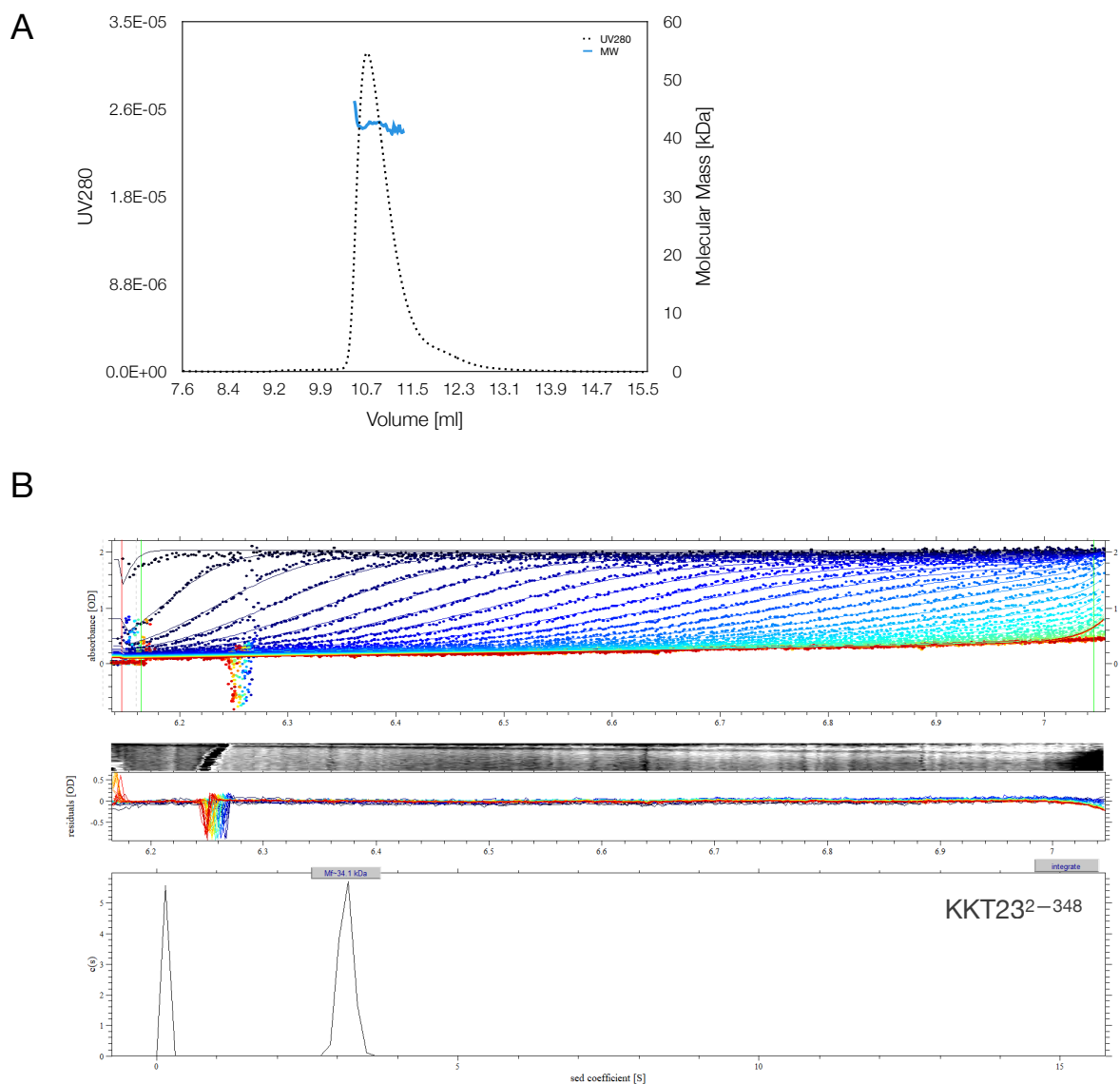


Figure 6.1. Analysis of KKT23 oligomerisation states using SEC-MALS and AUC.

(A) SEC-MALS elution profile of 50 μ M KKT23^{2–348} in the presence of two molar excess of AcCoA.

The UV₂₈₀ signal (dotted line) is plotted against the elution volume, while the molecular weight is indicated as a thick line. The estimated molecular weight (~40 kDa) matches the calculated value based on the sequence (39 kDa), suggesting that KKT23 is a monomer.

(B) AUC analysis of 50 μ M KKT23^{2–348} in the presence of two molar excess of AcCoA. The AUC results are consistent with the SEC-MALS analysis in (A), showing that KKT23 is a monomer.

To determine the high-resolution structure of KKT23, X-ray crystallography was used. However, no crystals of the full-length protein were obtained; therefore, I aimed to solve the structure of the N-acetyltransferase domain located at the KKT23 C-terminus. To increase the chance of crystallisation, multiple constructs have been designed with different N-terminal boundaries (Table S2). The natural C-terminus of KKT23 ending at A348 was left intact. Initially, none of the constructs yielded crystals. However, the crystallisation attempts of KKT23^{125–348} with acetyl coenzyme A (AcCoA), a known cofactor of acetyltransferases, led to robust crystal formation in multiple conditions, with the best crystals diffracting up to 1.8 Å resolution. Molecular replacement was used to solve the phase problem, where structures of N-acetyltransferase domains available in the Protein Data Bank (PDB) were used as models. Although unique solutions were found by PHASER (McCoy et al., 2007), the refinement statistics indicated that the searches were unsuccessful. As an alternative approach to molecular replacement, an experimental phasing approach was used, where a selenomethionine derivative of KKT23^{125–348} was purified and crystallised. The crystals of SeMet-KKT23^{125–348} diffracted up to 2 Å resolution, which allowed to solve the phase problem and eventually determine the crystal structure. Furthermore, the molecular replacement using the native KKT23^{125–348} diffraction dataset and SeMet-KKT23^{125–348} crystal structure as the model resulted in solving the 1.8 Å structure of KKT23^{125–348} from native crystals (Figure 6.2). The

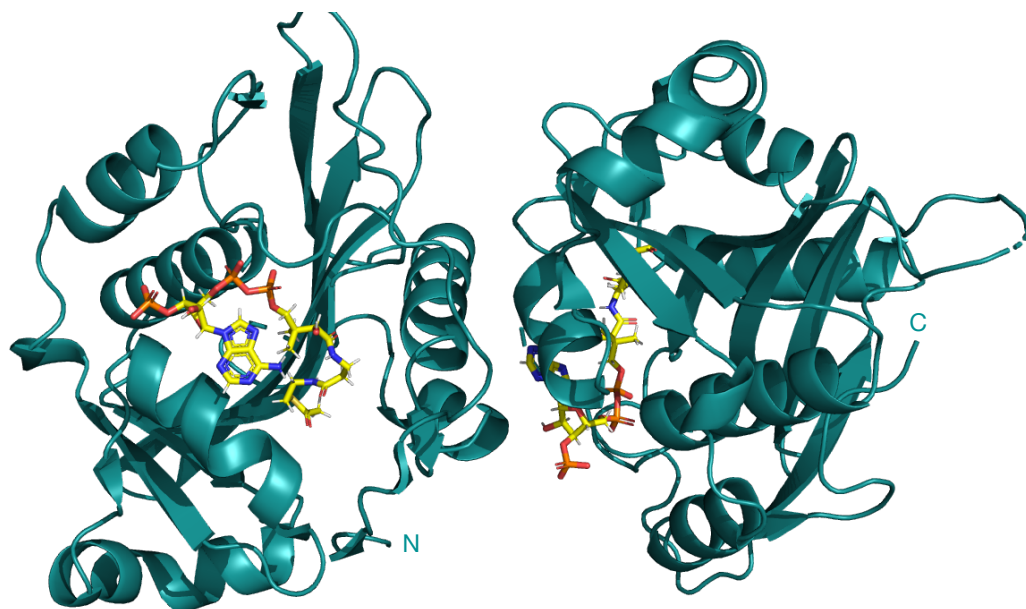
data collection and refinement statistics of both structures are summarised in Table 6.1. The native and selenomethionine structures overlay perfectly with no differences between the two (data not shown). In the forthcoming section, I use the native crystal structure to describe the KKT23 GNAT domain.

Table 6.1. Data collection, refinement statistics*.

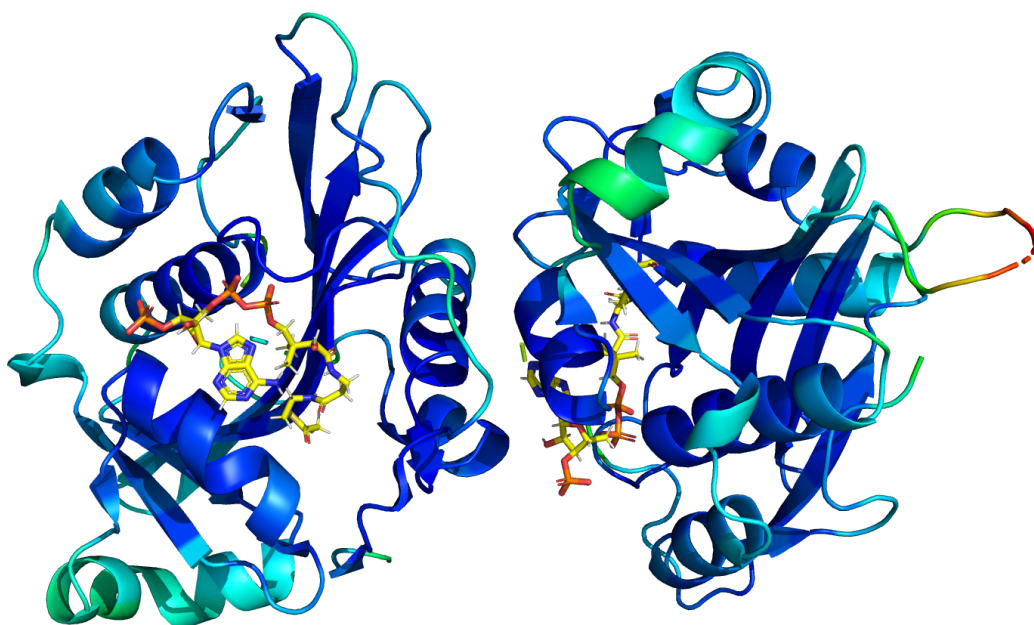
Data collection	<i>T. brucei</i> SeMet-KKT23 ^{125–348}	<i>T. brucei</i> KKT23 ^{125–348}
Beamline	Diamond Light Source I03	Diamond Light Source I03
Wavelength (Å)	0.9786	0.9763
Space group (Z)	<i>P1</i>	<i>P1</i>
Unit cell	42.30Å 44.53Å 68.97Å 100.54° 103.17° 101.33°	42.27Å 44.32Å 68.82Å 100.51° 103.00° 101.32°
Resolution range (Å)	29.47–1.95 (2.02–1.95)	65.14–1.75 (1.81–1.75)
Unique reflections	32751 (2983)	40204 (2232)
Completeness (%)	96.56 (86.99)	85.93 (47.56)
Multiplicity	3.7 (3.4)	3.7 (3.6)
<i>I</i> / σ <i>I</i>	38.7 (2.0)	45.4 (1.8)
<i>R</i> _{meas}	0.054 (0.704)	0.026 (0.741)
<i>CC</i> _{1/2}	1.0 (0.8)	1.0 (0.8)
Wilson B-factor (Å ²)	25.62	18.67
Refinement		
No. reflections	32733 (2981)	40196 (2231)
<i>R</i> _{work}	0.201 (0.273)	0.178 (0.265)
<i>R</i> _{free}	0.239 (0.3489)	0.183 (0.265)
Number of atoms	3700	4003
Protein	3423	3480
Ligands	102	102
Solvent	175	421
RMS bonds (Å)	0.008	0.007
RMS angles (°)	0.93	0.96
Ramachandran favoured (%)	96.40	97.61
Ramachandran allowed(%)	3.36	2.39
Ramachandran outliers (%)	0.24	0.00
Average B-factor (Å ²)	31.45	24.15

*Statistics for the highest-resolution shell are shown in parentheses.

A



B



C

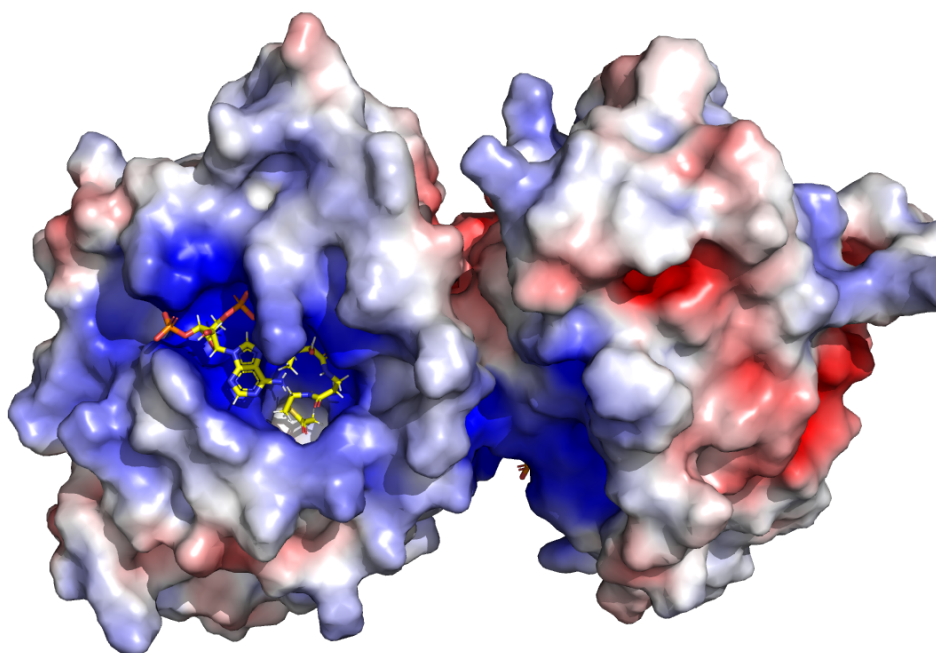


Figure 6.2. Crystal structure of KKT23^{125–348} bound to acetyl coenzyme A.

- (A) Ribbon model of molecule A (left) and B (right) of *Trypanosoma brucei* KKT23^{125–348}. Acetyl coenzyme A is shown as sticks and coloured by the elements. The N and C represent the N- and C- termini, respectively. The secondary structure elements are numbered in Figure 6.3.
- (B) B-factors for C α atoms have been represented for *Tb*KKT23^{125–348} using a blue to red spectrum indicating low to high values, respectively. The figures were rendered using PyMol (DeLano, 2002).
- (C) The surface electrostatic potential is coloured from red to blue, -5 kbT to +5 kbT, as calculated by APBS electrostatic plugin in Pymol (Jurrus et al., 2018).

Two molecules (A and B) of KKT23^{125–348} bound with AcCoA have been found in the asymmetric units of both structures. In molecule A, the electron density is missing for the N-terminal Q125 and residues 192–200, whereas in molecule B, the N-terminal regions of 125–126, 137–139 and 191–196 are missing, most likely due to being disordered and/or flexible. The overall KKT23^{125–348} structure has a mixed α/β topology containing 7 α -helices and 8 β -strands organised in 4 motifs (A, B, C and D), as shown in Figure 6.3; this motif organisation is consistent with previously reported structures of N-acetyltransferases (Dyda et al., 2000). The core of the enzyme is built by a three-stranded antiparallel β -sheet (strands 3,4 and 5) that sits below α -helix 3 (motif A and D). The β -strands β 3 and β 4 (motif D) near the N-terminus are hydrogen-bonded to each other, most likely stabilising the core region. Another part of the protein core is a β -loop- α motif built by β 6 and α 4 (motif B). In contrast to the secondary structure prediction (Figure 1.8), no secondary structure elements are observed in the crystal structure within residues 125–150 (Figure 6.2). Using the *Stride* plugin in Pymol (DeLano, 2002), additional short β -strands can be observed in the structure. Two of these strands, involving residues 129–130 and 133–134, are located in the unstructured N-terminal region and are hydrogen bonded to each other and the longer strand β 2 as shown in Figure 6.3. The

hydrogen bonds formed by these small strands could explain why the N-terminal fragment is well defined in the electron density despite the lack of defined secondary structure. The overall electrostatic potential of the protein surface was calculated using the APBS electrostatics plugin in Pymol (DeLano, 2002) and shows a mixture of basic and acidic regions with the most positively charged region occupied by AcCoA. The cartoons showing the structure coloured by the b-factors and electrostatic potential can be found in Figure 6.2B, C, respectively. The DALI search for the structural homologues of *Tb*KKT23^{125–348} crystal structure identified the human KAT2A (also known as hGcn5) histone acetyltransferase domain as the top hit (Table S4). Indeed, the alignment between KKT23^{125–348} and human Hcn5 reveals an excellent structural similarity with an RMSD of 3.84 (125 to 125 atoms) (Figure 6.4). Additional β -strand (β 8) and an α -helix (α 7) are present in the crystal structure of *Tb*KKT23^{125–348} but are not part of the consensus GNAT domain fold; however, both secondary structure elements are conserved among kinetoplastids (Figure 1.9).

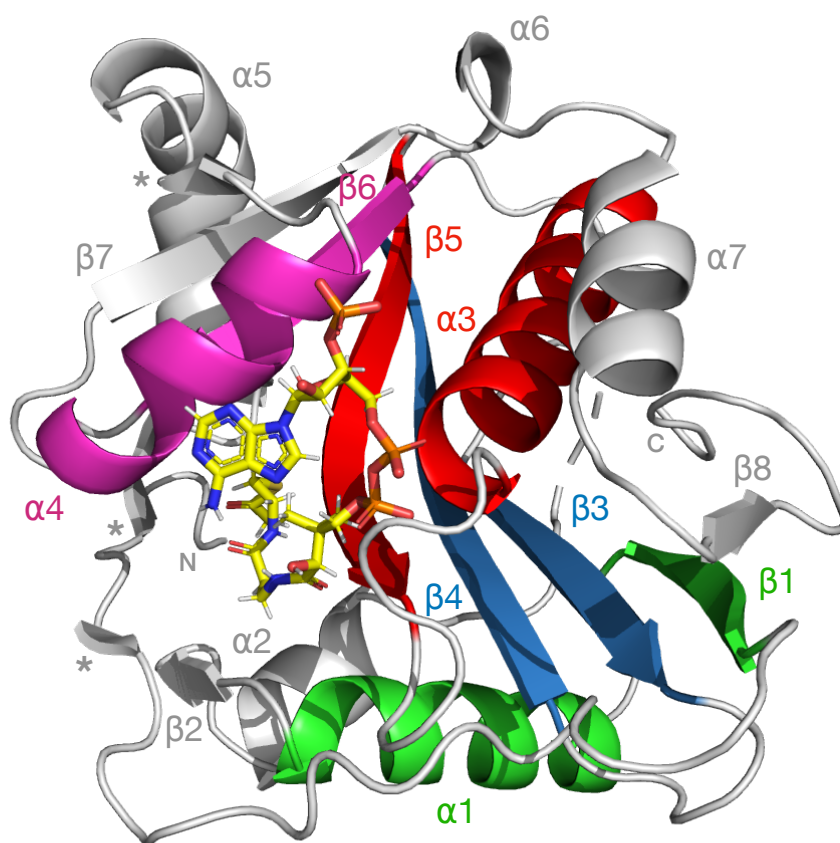


Figure 6.3. KKT23^{125–348} is composed of motifs A, B, C and D.

In the crystal structure, motifs are highlighted using different colours: motif A (red), B (pink), C (green) and D (blue). The acetyl coenzyme A is shown as sticks and coloured by the elements. The core of the enzyme is built by motifs A and D. The three asterisks represent short β -strands, which were not included in the numbering of α/β elements. The N and C represent the N- and C- termini, respectively.

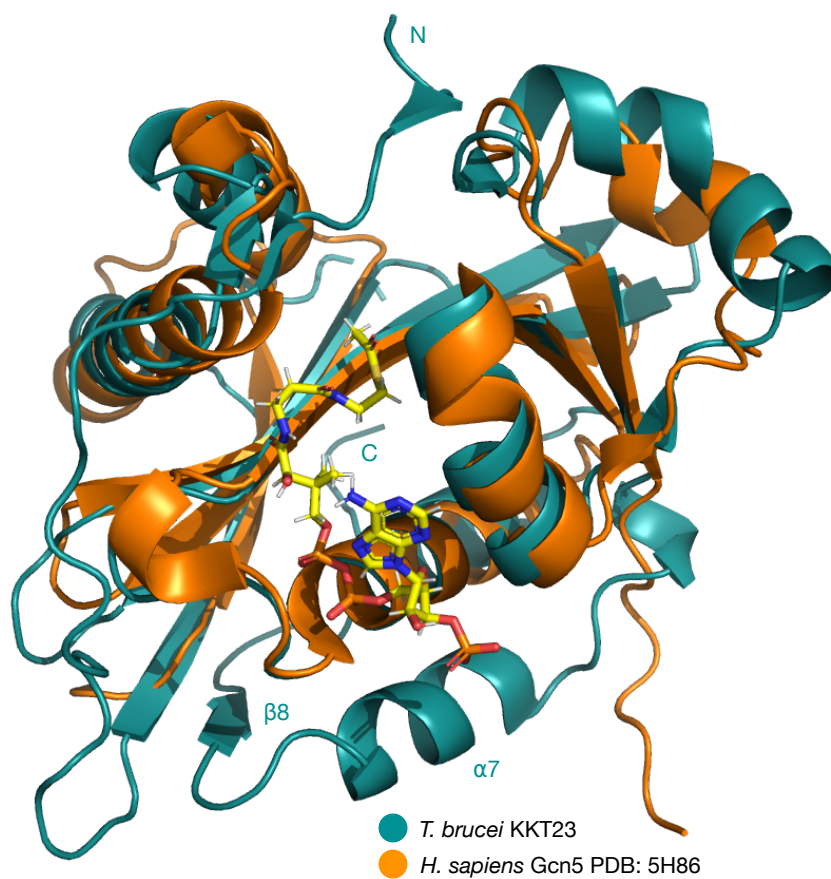


Figure 6.4. The overlay of *TbKKT23* GNAT domain and *HsGcn5* histone acetyltransferase crystal structures.

Superposition of the *TbKKT23*^{125–348} (teal) with the *HsGcn5* histone acetyltransferase domain (orange). The structures were overlaid using *align* function in Pymol (DeLano, 2002) with RMSD of 3.84. The $\alpha 7$ and $\beta 8$ are not part of the GNAT structure in *HsGcn5*. The N and C represent the N- and C- termini of KKT23^{125–348}.

According to the analysis done with PDBePISA using the crystal structure (Krissinel and Henrick, 2007), AcCoA is coordinated by neighbouring residues through hydrogen bonds and hydrophobic interactions (Table 6.2, Figure 6.5). Residues that form hydrogen bonds with AcCoA include F239, T241, K247, G249, F250, G251, R252 and H284. F239 and T241 are located on $\beta 5$; K247 occupies the loop region preceding $\alpha 3$, whereas G249, F250, G251 and R252 are located on $\alpha 3$. Lastly, H284 is located on $\alpha 4$. In summary, most residues that are hydrogen-bonded to AcCoA are part of motif A; this is consistent with the previous studies of acetyltransferases within the GNAT superfamily (Dyda et al., 2000, Marmorstein and Roth, 2001). In summary, the crystal structure of KKT23^{125–348} shows a histone acetyltransferase domain, followed by an α -loop- β region.

Table 6.2. KKT23 residues form hydrogen bonds with AcCoA.

KKT23 residues	Distance [Å]	Acetyl Coenzyme A atoms
F239 [O]	2.06	HN4
T241 [N]	2.67	09P
T241 [OG1]	2.85	04A
K247 [N]	2.78	05A
G249 [N]	2.00	01A
F250 [N]	3.45	02A
G251 [N]	2.93	02A
G251 [N]	2.94	04A
R252 [NH2]	2.73	08A
R252 [N]	2.86	02A
H284 [ND1]	2.85	07A

PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) was used to analyse the interaction between KKT23 GNAT domain and AcCoA in the crystal structure. Atoms of KKT23 residues that make contacts with AcCoA are shown in brackets.

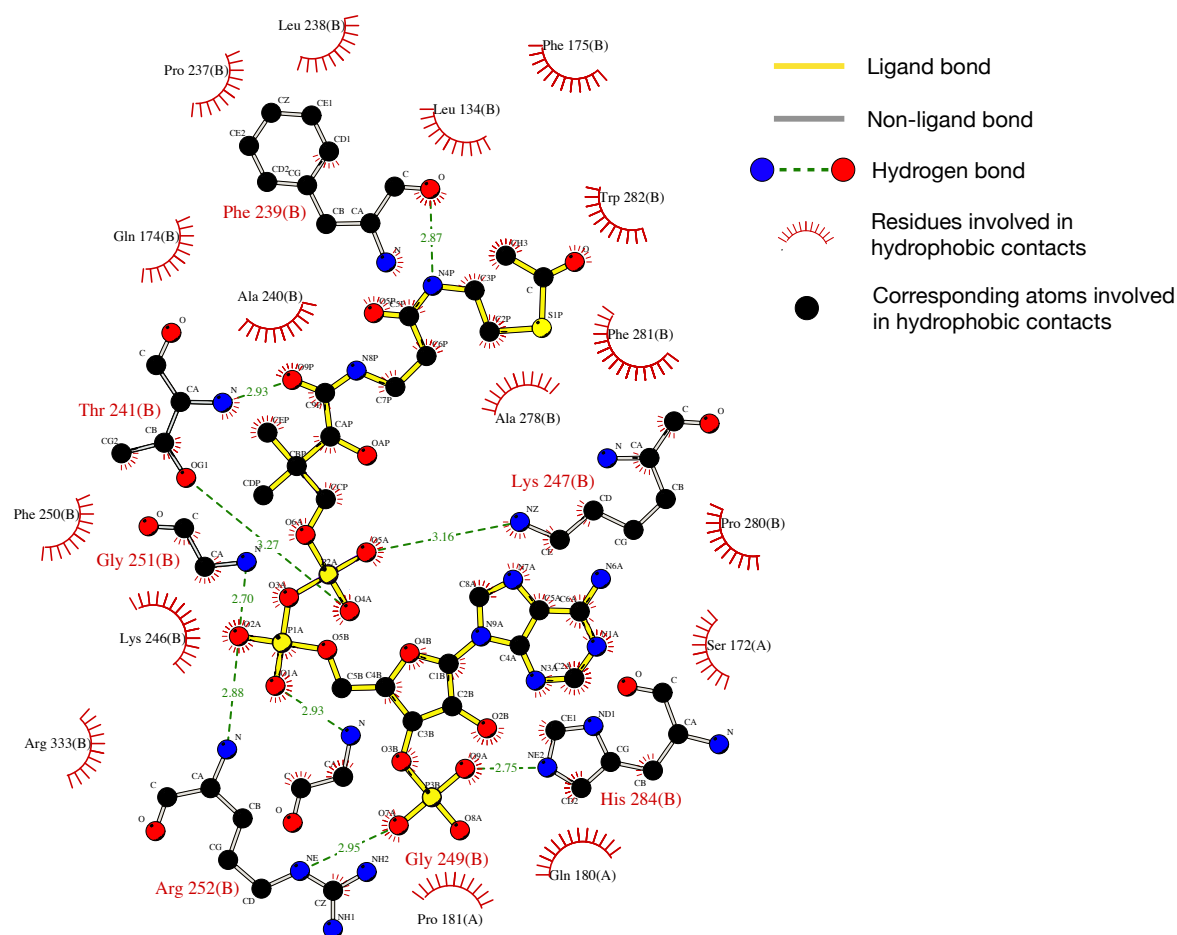


Figure 6.5. Coordination of AcCoA in the crystal structure.

This figure shows the visual representation of interactions observed between AcCoA and KKT23 residues in the crystal structure of KKT23^{125–348}. The AcCoA is coloured in yellow, while the side chains of KKT23 residues are shown as grey sticks. AcCoA is coordinated through hydrogen bonds and hydrophobic interactions. The graph shows AcCoA bound to molecule B from the crystal structure but also represents equivalent ligand in molecule A. This figure was prepared and analysed using LIGPLOT (Wallace et al., 1995). Please also see Table 6.2.

6.2. NMR analysis of the KKT23 GNAT domain

In this section, I aim to characterise the KKT23^{125–348} structure and dynamics in solution using NMR methods. Detailed backbone assignments of KKT23^{125–348} bound to AcCoA were carried out, followed by its structural and functional analysis.

Resonance assignments of KKT23^{125–348}

The 2D ¹H-¹⁵N BTROSY spectrum of KKT23^{125–348} shows fewer than the expected number of peaks (~170) for a protein that contains 216 non-proline residues (Figure 6.6). The peaks in the spectrum are well dispersed, suggesting that the protein is folded in the absence of the AcCoA ligand. To determine how AcCoA affects the structure of the GNAT domain, a 2D ¹H-¹⁵N BTROSY spectrum was collected for KKT23^{125–348} with an addition of a two-fold molar excess of AcCoA. The overlay of the KKT23^{125–348} spectra with and without AcCoA shows that very few peaks remained unaffected (Figure 6.6), suggesting that the binding of AcCoA to KKT23^{125–348} resulted in conformational changes. The change of conformation seems to affect the chemical environment of almost all residues. AcCoA contains an aromatic adenine moiety and negatively-charged phosphate groups; many of the observed chemical shift differences may arise from the aromatic ring current shifts, electrostatic and hydrogen bonding contributions to the shifts arising upon binding of AcCoA. In the presence of AcCoA, 26 additional peaks are observed, leaving ~20 peaks missing, which might result from overlap, from rapidly exchanging amides or due to specific exchange broadening. Significant degradation of the apo-KKT23 was observed over a period of 24 h. Based on the fact that AcCoA stabilises KKT23^{125–348} and that the acetyltransferase domain in complex with AcCoA is of more interest than the apoenzyme state, further NMR studies have been carried out on KKT23^{125–348} in the presence of AcCoA.

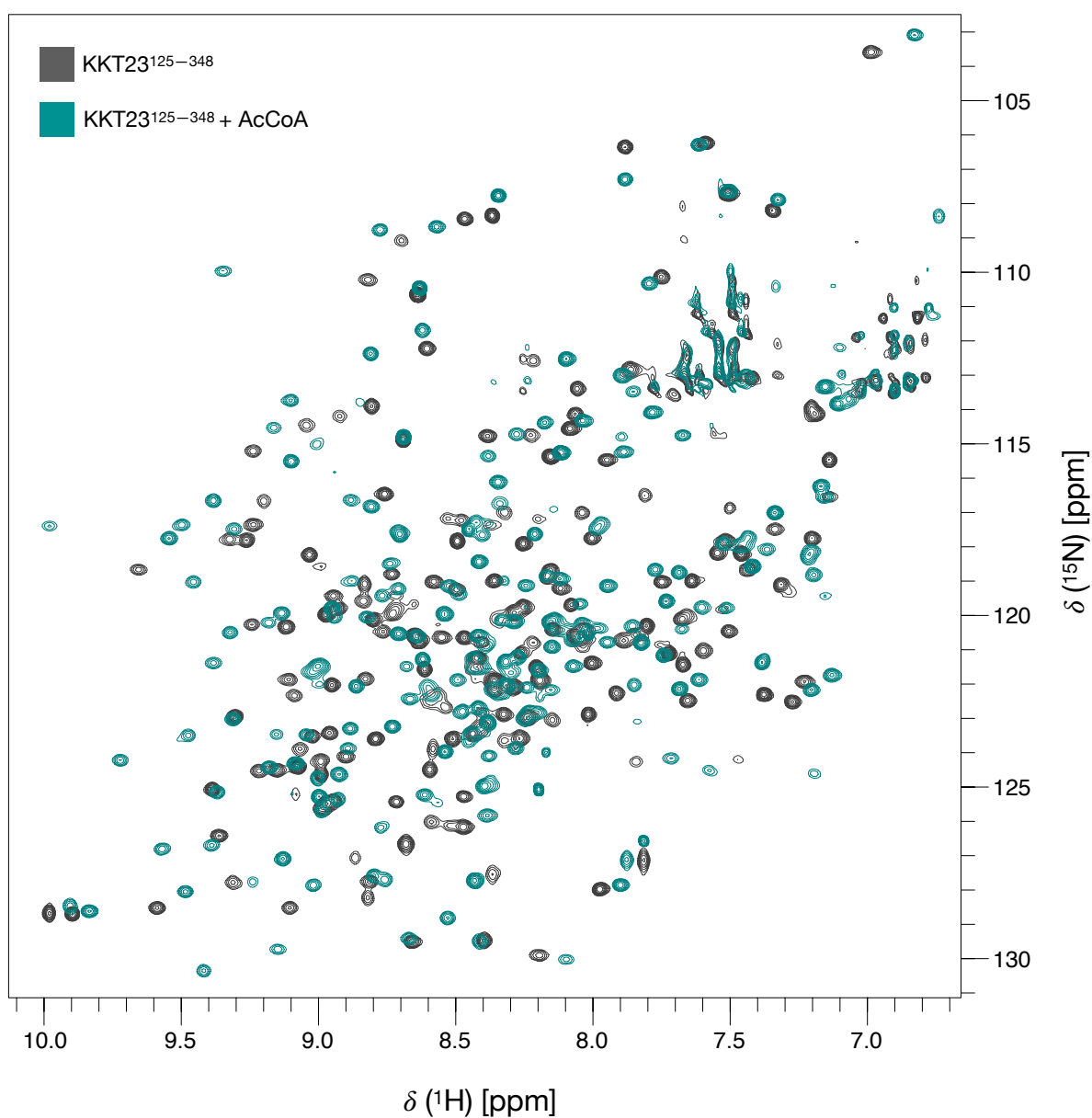


Figure 6.6. Acetyl coenzyme A changes the conformation of KKT23¹²⁵⁻³⁴⁸ upon binding.

The overlay between the 750 MHz spectra of KKT23¹²⁵⁻²³⁸ with and without addition of AcCoA presented in teal and black, respectively. The addition of AcCoA results in significant changes to the chemical shifts of most peaks.

To obtain detailed backbone and side-chain assignments of KKT23^{125–348}, a set of triple resonance experiments has been collected and analysed (Table S1). Out of 216 non-proline residues, ¹H^N-¹⁵N for 193 were assigned to specific peaks shown in the BTROSY (Figure 6.7). The missing assignments for 31 residues can be explained by the presence of 8 prolines (for which the ¹H-¹⁵N^H signal cannot be observed), poor quality of the triple resonance data, where ¹³C correlations were missing for many of the ¹H^N-¹⁵N peaks and the fact that the number of observed peaks was lower than expected. The extent of the obtained assignments is summarised in Table 6.3.

Table 6.3. Extent of assignment for KKT23 GNAT domain

Percent assigned							
Sample	¹ H ^N / ¹⁵ N	¹³ CO	¹ H α / ¹³ C α	¹ H β / ¹³ C β	¹ H γ / ¹³ C γ	¹ H δ / ¹³ C δ	¹ H ϵ / ¹³ C ϵ
KKT23 ^{125–348}	89.4/89.4	87.9	76.1/93.8	49.9/80.5	-/-	-/-	-/-

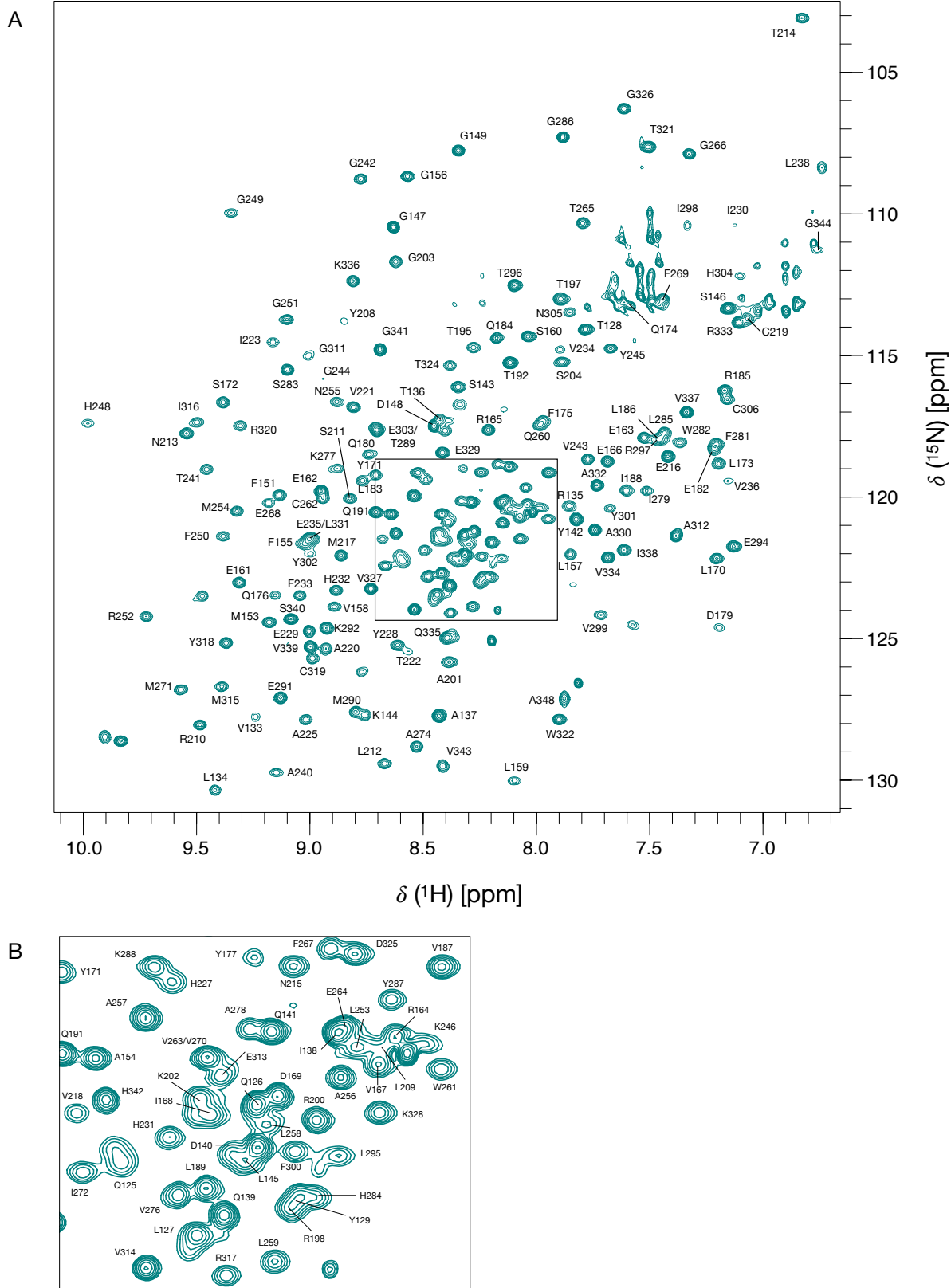


Figure 6.7. The ^1H – ^{15}N BEST-TROSY spectrum of KKT23^{125–348}.

(A) 750 MHz ^1H – ^{15}N BEST-TROSY spectrum of KKT23^{125–348} in 50 mM sodium phosphate, 150 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 20°C. The peak assignments for backbone amides are annotated.

(B) The expansion of the crowded region highlighted with a black box in (A).

Secondary structure

Using the obtained chemical shifts, TALOS-N was used to predict the secondary structure elements in KKT23^{125–348} in solution. The TALOS-N analysis shows a mixture of α -helices and β -strands (Figure 6.8A). Overall, the secondary structure elements identified by the TALOS-N are in excellent agreement with those identified in the crystal structure (Figure 6.8B). It is noteworthy that as in the crystal structure, the NMR analysis confirmed the lack of the predicted helix between residues 125–150. Moreover, the presence of short β -strands, highlighted in Figure 6.3 was confirmed with TALOS-N.

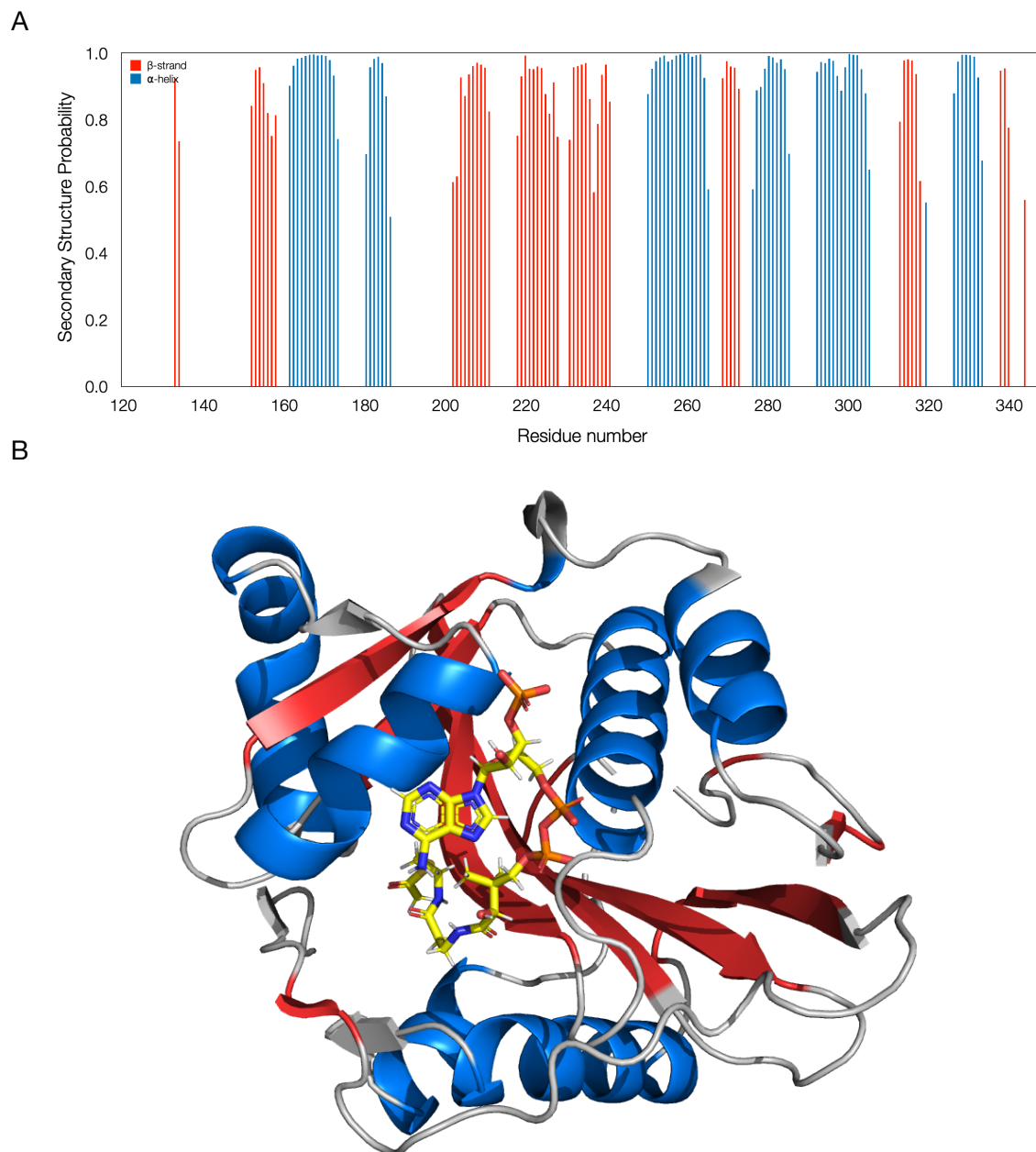


Figure 6.8. The secondary structure analysis of KKT23^{125–348} using NMR.

- (A) The TALOS-N secondary structure analysis of KKT23^{125–348} shows a mixed α/β topology. The secondary structure elements for which the probability was lower than 0.5 are not shown in the figure.
- (B) The secondary structure elements identified using TALOS-N (A) are highlighted in the crystal structure of KKT23^{125–348} showing a good agreement between the two methods. The β -strands and α -helices are coloured in red and blue, respectively. The minor differences include short β -strands highlighted with asterisks in Figure 6.3; 2 out of 3 β -strands are not predicted to adopt secondary structure according to the TALOS-N prediction.

Backbone dynamics

To study the dynamics of KKT23^{125–348}, the {¹H}-¹⁵N heteronuclear NOE ratios were measured in the presence of AcCoA. Peaks located in the poorly resolved regions were not included because the calculation of hetNOE ratios of the overlapping peaks is inaccurate. The measured {¹H}-¹⁵N NOE ratios were high (>0.7) for almost all residues (Figure 6.9). This means that the majority of the residues in KKT23^{125–348}, even those located in the loop regions, have a rigid conformation. The regions between 126–128, 137–143 and 192–201 show reduced hetNOE ratios suggesting a higher flexibility compared to the rest of the protein. This is consistent with the crystal structure and the analysis by TALOS-N, as no secondary structure was observed for these parts of the sequence. Moreover, elevated b-factors can be observed for these regions in the crystal structure (Figure 6.2B).

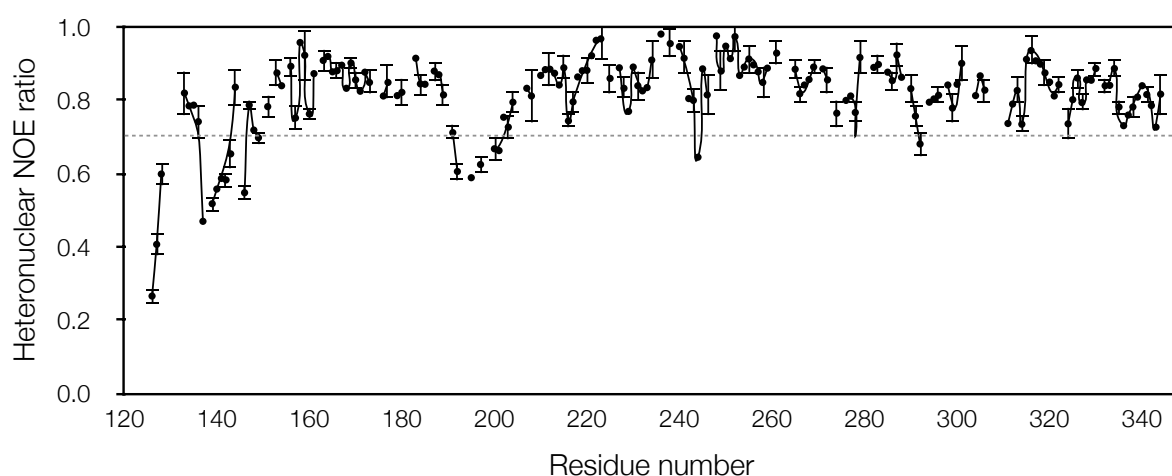


Figure 6.9. The {¹H}-¹⁵N heteronuclear NOE measurements for KKT23^{125–348}.

The {¹H}-¹⁵N heteronuclear NOE ratios were measured and plotted against the sequence of KKT23^{125–348}. Most of the residues display hetNOE ratios >0.7 indicating a rigid conformation of the protein backbone. The regions with higher flexibility (ratios <0.7) include 126–128, 137–143 and 192–201. The regions between 126–128 and 137–143 are coils according to the crystal structure, whereas no electron density is visible for residues 192–201 suggesting that all three regions are part of flexible loops. The {¹H}-¹⁵N NOE errors were estimated from 500 Monte Carlo simulations using baseline noise as a measure of peak height error.

NMR analysis of the full-length KKT23

The efforts to crystallise the full-length KKT23 were unsuccessful. Instead, to obtain some structural information on the full-length KKT23, NMR spectroscopy was employed. The 2D ^1H - ^{15}N BTROSY spectrum of KKT23^{2–348} did not result in the expected number of peaks for a protein with 332 non-proline residues (Figure 6.10). The overlay of the spectra between KKT23^{2–348} and KKT23^{125–348} shows a good alignment (Figure 6.10), suggesting that the removal of the N-terminal region did not affect the properties of KKT23^{125–348}. This means that the N-terminal region is unlikely to form contacts with the acetyltransferase domain near the C-terminus. Besides peaks that align with the spectrum of KKT23^{125–348}, ~77 peaks with strong intensities are visible in the middle of the KKT23^{2–348} spectrum (Figure 6.11). These peaks most likely correspond to the disordered region predicted between residues 66 and 127 (Figure 1.8C). Indeed, the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE confirmed these peaks are highly flexible with hetNOE ratios <0.7 (Figure 6.12). I suspect that the missing peaks in the spectrum of KKT23^{2–348} correspond to the predicted helical region located near the N-terminus of KKT23.

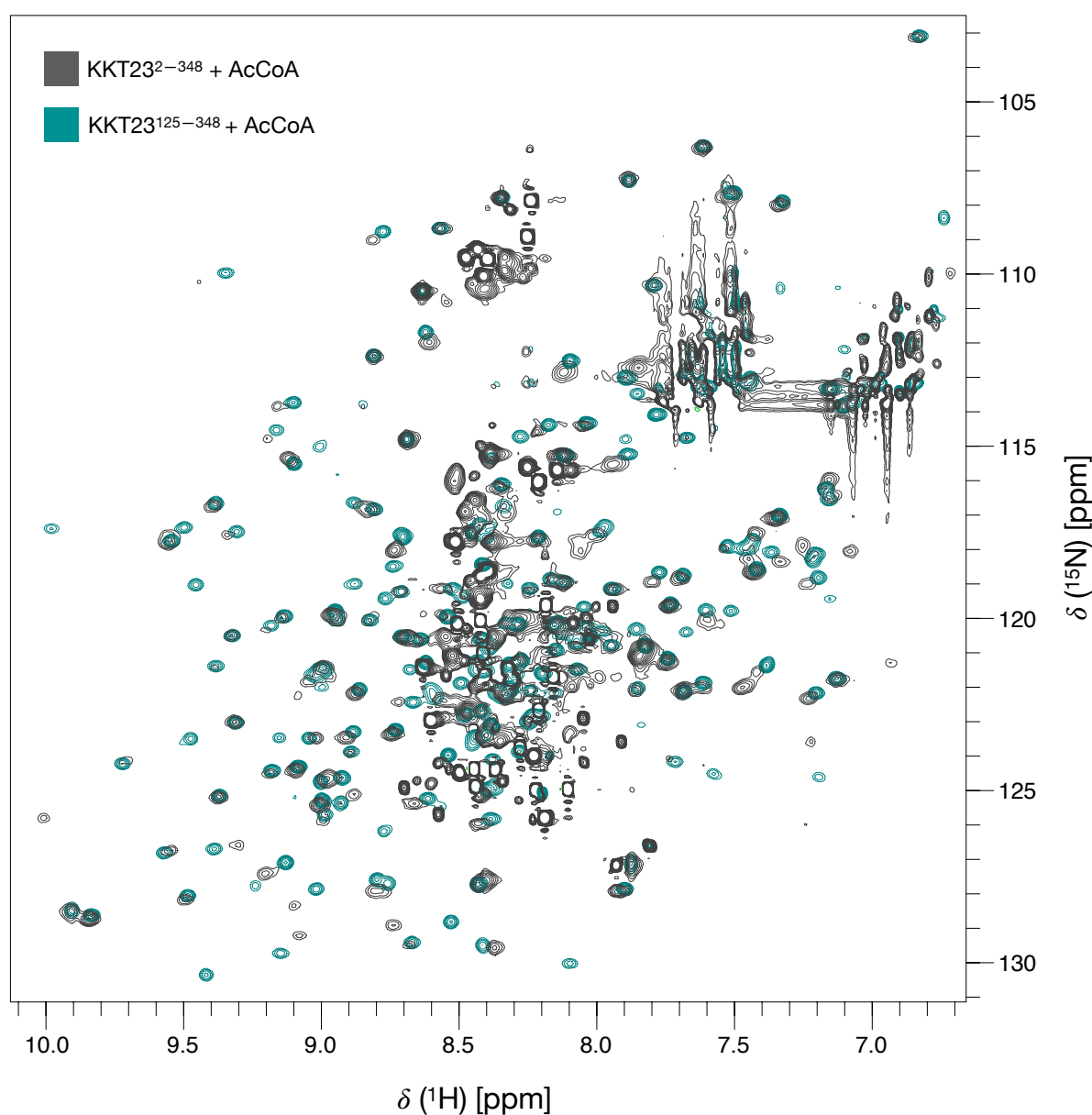


Figure 6.10. Overlay of spectra between the full-length KKT23 and KKT23 GNAT domain.

Overlay of 750 MHz ^1H - ^{15}N BEST-TROSY spectra of KKT23²⁻³⁴⁸ (black) and KKT4¹²⁵⁻³⁴⁸ (teal). Most of the peaks in the spectrum of KKT4¹²⁵⁻³⁴⁸ overlay well with the spectrum of KKT4²⁻³⁴⁸, indicating that removing the N-terminal region does not perturb the properties of KKT23¹²⁵⁻³⁴⁸. Peaks that do not overlay likely arise from the different N-terminal sequence in KKT4¹²⁵⁻³⁴⁸. Peaks located at ~8.0–8.7 ppm have high intensity and are not well dispersed, suggesting that they arise from residues that are disordered.

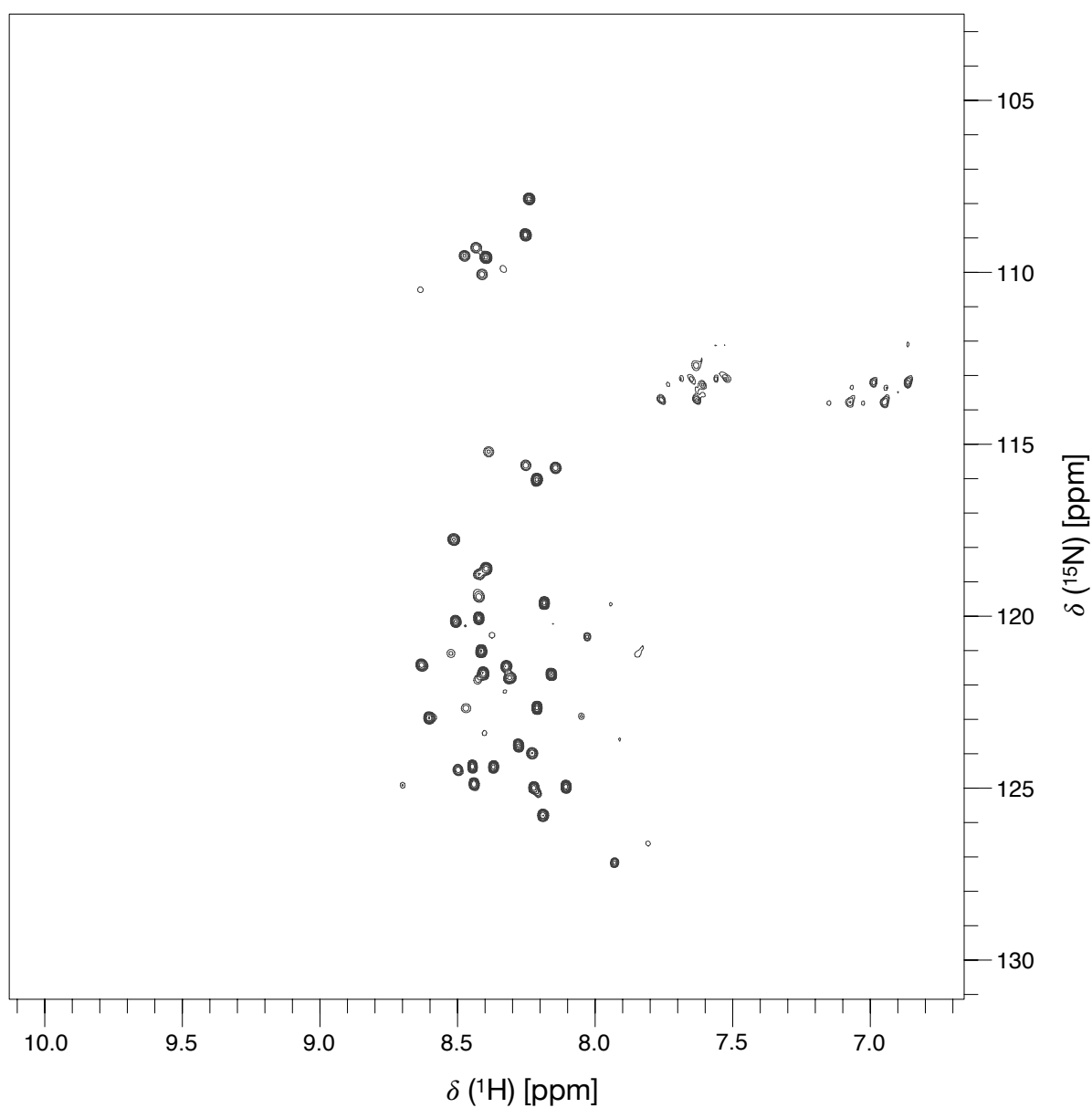


Figure 6.11. The ^1H - ^{15}N BEST-TROSY spectrum of KKT23^{2–348}.

The spectrum of the full-length KKT23 was contoured in a way to highlight the high intensity peaks around 8.0–8.7 ppm. These peaks likely correspond to disordered regions of KKT23 predicted between 72–126 (Figure 1.8C).

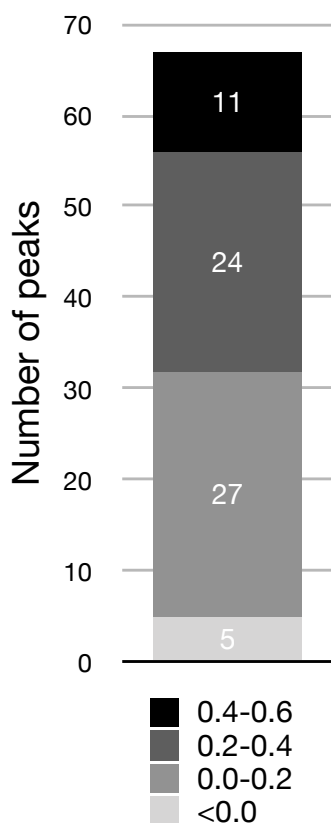


Figure 6.12. The $\{^1\text{H}\}\text{-}^{15}\text{N}$ heteronuclear NOE measurements of KKT23^{2–348}.

The histogram shows the distribution of $\{^1\text{H}\}\text{-}^{15}\text{N}$ heteronuclear NOE ratios measured in KKT4^{2–348} for 67 strongest peaks shown in Figure 6.11. The heteronuclear NOE ratios were segregated and colour coded by different shades of grey according to different values observed. All peaks displayed the hetNOE ratios <0.6 , which is characteristic for the flexible backbone. These peaks likely correspond to residues located within 72–126 region, which is predicted to be disordered (Figure 1.8C).

The N-terminal region of KKT23 is mainly helical

The N-terminal residues of KKT23 are predicted to be helical (Figure 1.8B) and is well conserved among kinetoplastid species (Figure 1.9). Moreover, no significant similarity to any known fold was found when analysing the primary sequence of this region; therefore, it is difficult to predict its structure or function. In this section, the preliminary structural data obtained on KKT23^{2–70} is presented.

To confirm the presence of helices near the N-terminus of KKT23, KKT23^{2–70} was analysed using circular dichroism. Indeed the CD confirmed that KKT23^{2–70} is helical (Figure 6.13). To obtain more detailed structural information, NMR was used to analyse KKT23^{2–70} because no obvious peaks corresponding to the N-terminal region of KKT23 were observed in the spectrum of KKT23^{2–348}. The 2D $^1\text{H}\text{-}^{15}\text{N}$ BTROSY spectrum of KKT23^{2–70} collected at 20°C identified as few as 6 peaks

(data not shown). This was unexpected for an 8 kDa protein that is likely to be monomeric. To improve the signal-to-noise and/or linewidths in the KKT23^{2–70} spectrum, BTROSY spectra were collected at 30°C, 40°C and 50°C. The best results were obtained at 50°C, where 56 peaks were observed (Figure 6.14). This is closer to the expected number of peaks for the 69 non-proline residues. The spectrum is reasonably well dispersed and peaks have uniform intensities, suggesting that most residues in the protein are folded. Indeed, the {¹H}-¹⁵N heteronuclear NOE analysis shows that all peaks have {¹H}-¹⁵N NOE ratios above 0.6, which is typical for structured regions (data not shown). In summary, the N-terminal fragment of KKT23 has a helical propensity with a rigid backbone. In the future, it will be interesting to determine the structure and function of KKT23^{2–70}.

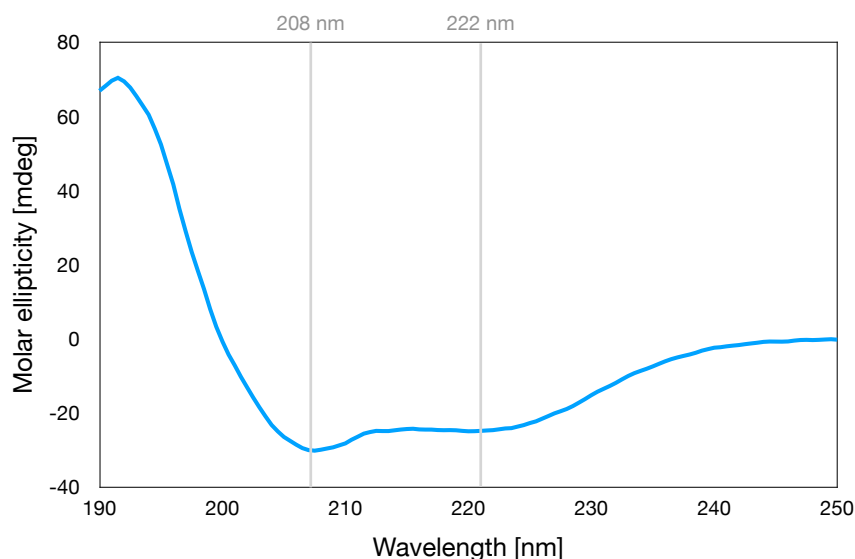


Figure 6.13. Circular dichroism spectrum of KKT23^{2–70}.

The spectrum of KKT23^{2–70} shows observable negative ellipticity at 208 and 222 nm, revealing that the N-terminal residues of KKT23 are helical.

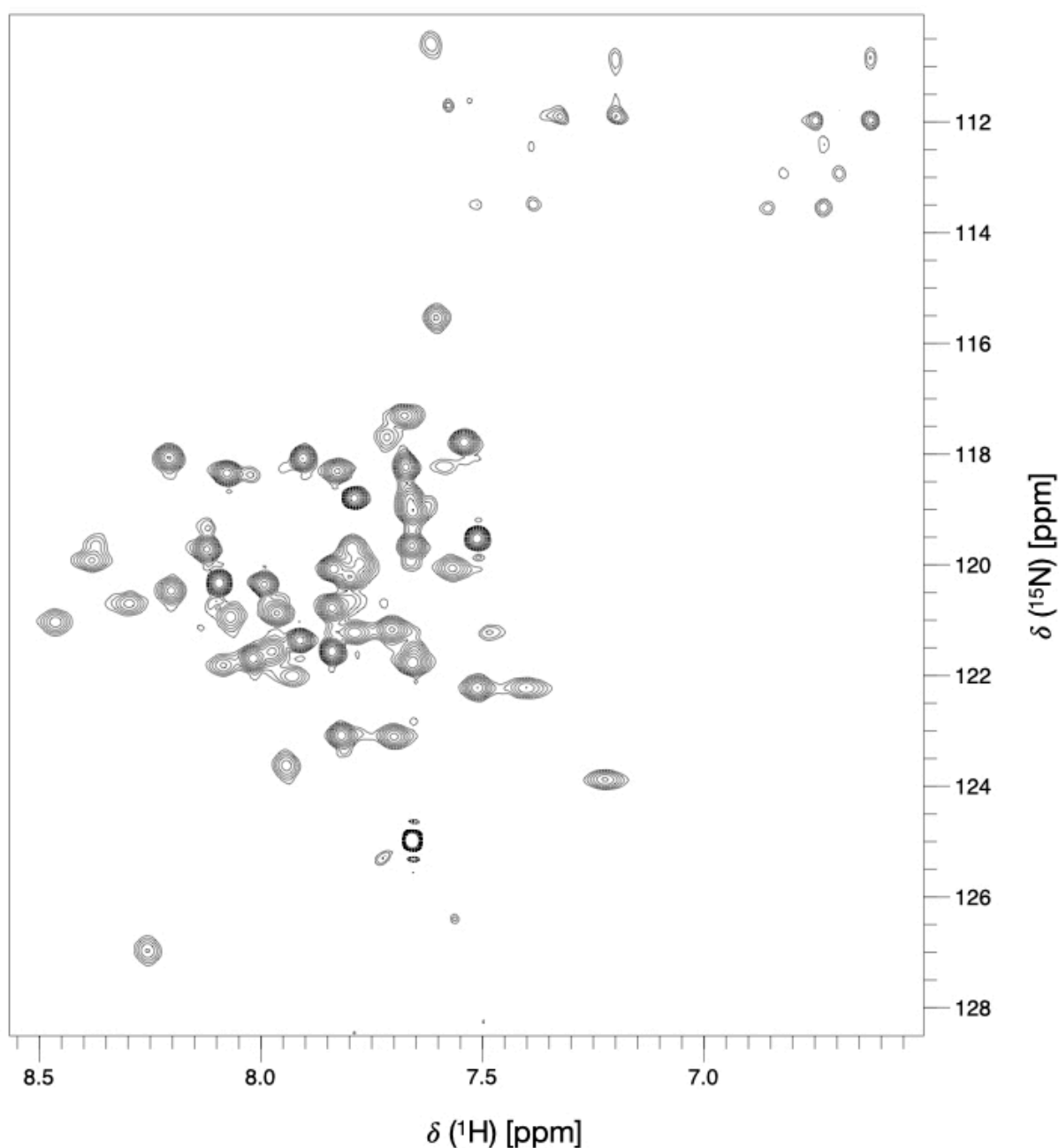


Figure 6.14. The ^1H - ^{15}N BEST-TROSY spectrum of KKT23²⁻⁷⁰.

750 MHz ^1H - ^{15}N BEST-TROSY spectrum of KKT23²⁻⁷⁰ in 25 mM HEPES, 75 mM NaCl and 0.5 mM TCEP (95% H_2O /5% D_2O), at pH 7.2, 50°C. Most of the observable peaks are well resolved and have uniform intensities suggesting that the N-terminal region of KKT23 is structured. Peaks in the region of 111–114 ppm and upfield of ~7.4 ppm are artefacts in the BEST-TROSY arising from incomplete cancellation of signals from the side chain amides of Asn and Gln.

6.3. KKT23 can acetylate histones *in vitro*

The C-terminal histone acetyltransferase domain of KKT23 has a high level of structural similarity to the human Gcn5 histone acetyltransferase, which is known to acetylate histones and non-histone targets (Orpinell et al., 2010, Fournier and Tora, 2017). KKT23 might regulate the function of other kinetochore components by acetylating them or may act as a hallmark of kinetochore assembly sites by acetylating histones specifically at the centromere. To test if KKT23 has acetyltransferase activity and to identify KKT23 substrates, I aimed to establish an *in vitro* acetylation assay, which would allow me to test if KKT23 can acetylate histones from *Trypanosoma brucei*.

Histone acetyltransferase assay

Histone acetyltransferase activity has been studied *in vitro* using a range of biochemical methods. These methods use purified core histones or synthesised histone tails as substrates and an acetyl donor, so-called hot acetyl coenzyme A, labelled with either ^3H or ^{14}C (Ait-Si-Ali et al., 1998, Mizzen et al., 1999, Berndsen and Denu, 2005). Another assay links the release of CoA to the reduction of NAD^+ to NADH, which is directly measured using a spectrophotometer (Kim et al., 2000). An alternative approach to detect acetyltransferase activity is an assay based on immunoblotting and the use of sensitive polyclonal antibodies that specifically recognise acetylated ϵ -amine groups of lysine side-chains. I decided to employ this method as it allows for quick and robust detection of acetylated samples and does not require the use of radioactive isotopes.

Histone H2A and H2AZ are the substrates of KKT23

To obtain histone substrates, *Trypanosoma brucei* histones were expressed from a polycistronic vector in bacteria and purified from a soluble fraction. The affinity purification of His-tagged H3 in the presence of H2A, H2B and H4 was followed by a size exclusion chromatography to buffer exchange the sample into the HAT buffer for the acetylation assay (see Materials and Methods). In the SEC profile, at least four peaks were detected at 280 nm; two of those peaks, around 7.5 ml and 22.5 ml, include all four histones, whereas the other two peaks were mainly contaminants or degradation products (Figure 6.15A). The first peak (7.5 ml) elutes in the void and it is heavily contaminated with DNA. The second peak (22.5 ml), also contaminated with the DNA, co-elutes with the salt, which suggests that histones interact with the resin beads during the size exclusion run. It is unknown whether the obtained histones are folded and form octamers or exist as separate unfolded proteins in these samples. Besides the core histones (H2A, H2B, H3, H4), histone variants H2AZ and H2BV were purified for the acetylation assay (Figure 6.15B). Both H2AZ and H2BV are essential for *Trypanosoma brucei* cells *in vivo* (Lowell et al., 2005); therefore, I reasoned that it is important to test whether they can be acetylated by KKT23.

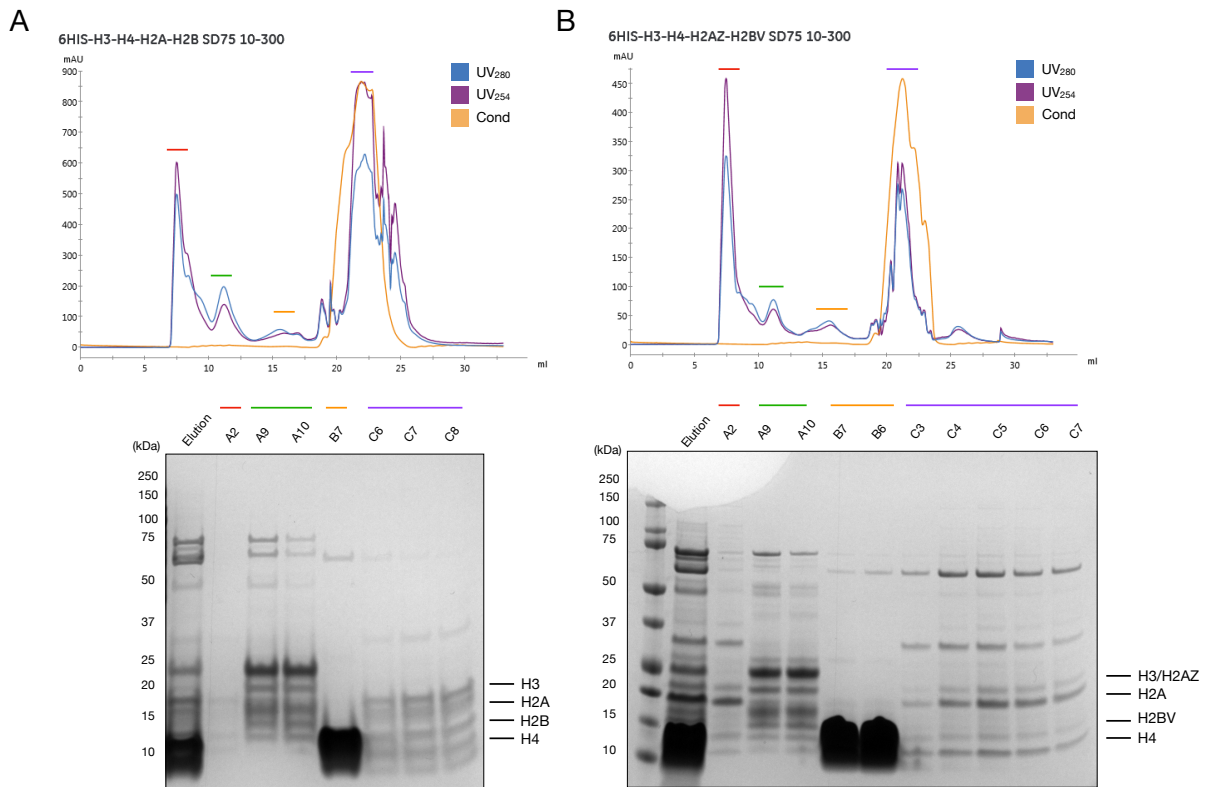


Figure 6.15. Crude purification of recombinant *T. brucei* histones.

The figure shows size exclusion profiles (top panel) and SDS-PAGE analysis (lower panel) of individual peaks for (A) 6His-H3/H2A/H2B/H4 and (B) 6His-H3/H2AZ/H2BV/H4. The size exclusion profiles are similar between the two constructs; 4 peaks can be observed in each profile. The gels in the lower panels show that the fractions from two peaks (~7.5 ml and 22.5 ml) contain histones, while other peaks are mainly contaminants. Histones from both peaks are contaminated with DNA, which was not removed despite the use of benzonase nuclease in the cell lysate and 500 mM NaCl during the purification. The collected fractions are highlighted and colour coded with the corresponding peaks observed in gel filtration.

The histone acetyltransferase assay using the full-length KKT23 and recombinant *T. brucei* histones as substrates resulted in a ~15 kDa band detected by anti-AcLys antibodies (Figure 6.16A, B). Based on its size, the band is likely histone H2A. By repeating the assay on purified histone H2A, I confirmed that histone H2A was indeed acetylated (Figure 6.16A, B). This means that KKT23 specifically acetylates histone H2A because no other histone was found to be acetylated in the assay.

Interestingly, the band corresponding to the histone variant H2AZ was also found to be acetylated using the anti-AcLys antibodies, whereas no acetylation was detected for histones H3, H4 and H2BV (Figure 6.16A, B). The experiment was repeated with the sample containing only H2AZ and it confirmed that the acetylation was detected on the H2AZ band (Figure 6.16A, B).

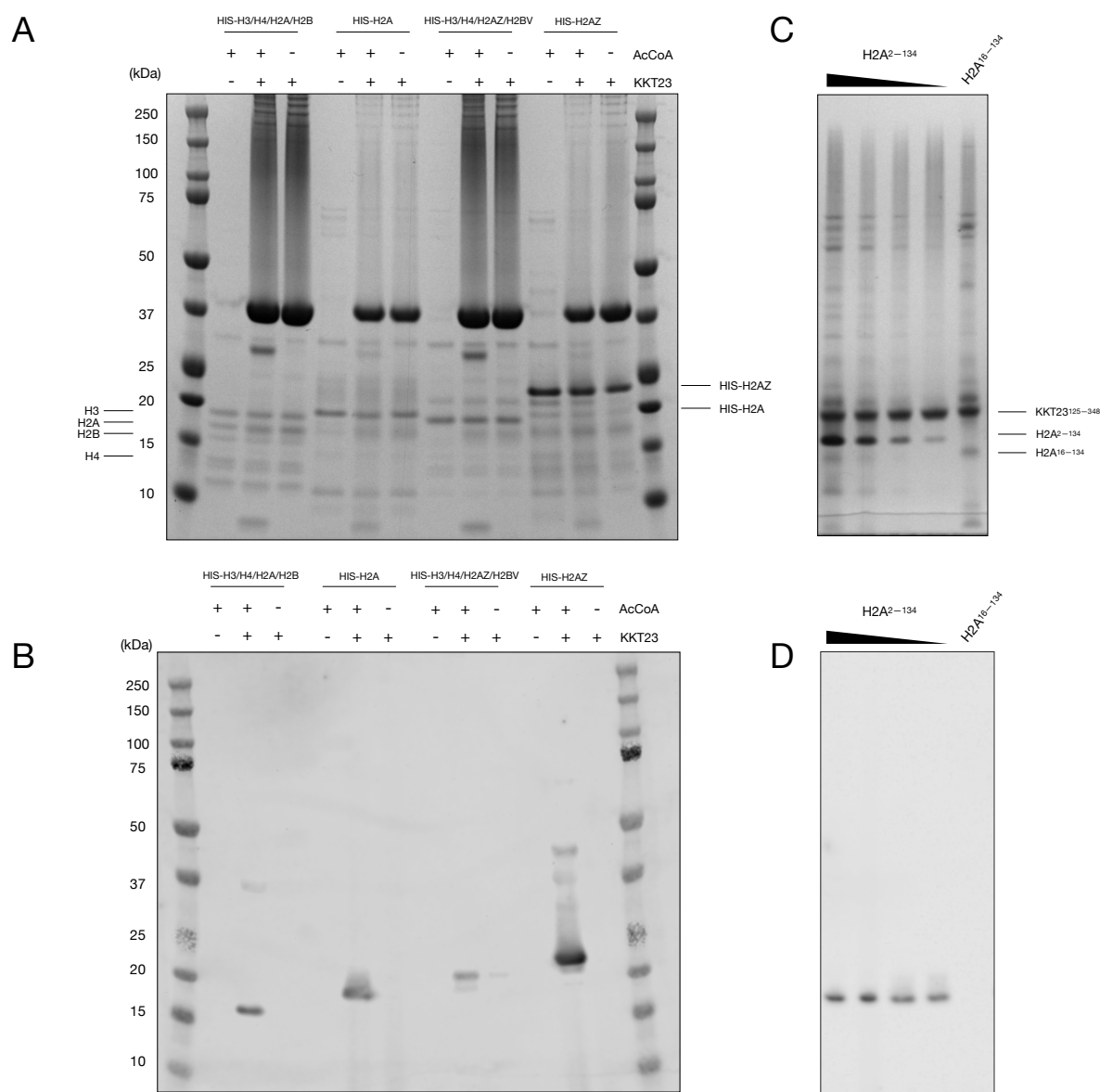


Figure 6.16. KKT23 acetylates histone H2A and H2AZ.

- (A) The SDS-PAGE showing *in vitro* acetylation assay on recombinant histones using KKT23^{2–348} acetyltransferase.
- (B) Western blot analysis of the reactions shown in (A) using anti-AcLys antibodies. The signal is detected in the region where histone H2A and H2AZ migrate, suggesting that they have been acetylated by KKT23. No acetylation signal is detected in the regions, where other histones migrate.
- (C) The SDS-PAGE showing an *in vitro* acetylation assay using the wild-type H2A (H2A^{2–134}, lanes 1–4) and a deletion mutant of H2A (H2A^{16–134}, lane 5). Gradient of H2A^{2–134} concentrations was used to match the signal intensity of H2A^{16–134}. The intensity of histones in lane 4 and 5 are comparable. In this assay, KKT23^{125–348} was used.
- (D) Western blot analysis of the reactions shown in (C) using anti-AcLys antibodies. The signal is detected for histone H2A^{2–134} but not for H2A^{16–134}. The N-terminal residues of histone H2A are required for it to be acetylated by KKT23 *in vitro*.

To identify the acetylated sites in H2A, mass spectrometry was employed. Following the acetylation assay, all samples were run on SDS-PAGE and the gel bands corresponding to the individual histones (H3, H2A, H2B, H4) were cut out and analysed using mass spectrometry. Besides the H2A band, other histones were analysed to confirm the specificity toward histone H2A observed with the anti-AcLys antibodies. Two control samples were used in the mass spectrometry analysis: the first sample did not contain KKT23 to check the baseline level of acetylated peptides detected by MS and the second did not have AcCoA to exclude the possibility that histone H2A is chemically acetylated in the assay. Unexpectedly, the MS analysis identified acetylated peptides on all histones besides the histone H4 (Table 6.4). These results were not consistent with the Western blotting results, where acetylation was detected specifically on histone H2A. Moreover, the acetylated peptides were detected even in the control samples, suggesting that the acetylation was not KKT23-related and the histones were already acetylated prior to

the assay. However, it is rather unlikely that histones were acetylated in bacteria. An alternative explanation is that the acetylation on histone peptides are false-positive results generated by MS. It is also noteworthy that the scores of identified peptide were of $\sim 10^{-7}$ magnitude, which is the borderline of statistically significant matches. The bands corresponding to histone H2AZ have not been analysed using MS.

Table 6.4. Mass spectrometry analysis of the acetylated histones.

	H2A	H2B	H3	H4
Control 1 (- KKT23 ^{125–348})	K115 (51) K128 (35)	K96 (67)	K23 (56) K32 (69) K76 (64)	-
Control 2 (- AcCoA)	K115 (49) K128 (38)	K96 (61) K104 (19)	K23 (71) K32 (76) K76 (69)	-
KKT23 ^{125–348} + AcCoA	K20 (55) K115 (47) K128 (36)	K96 (76) K104 (57)	K23 (56) K32 (69) K61 (47) K76 (83)	-

The table shows a summary of acetylation sites detected using mass spectrometry. The MASCOT scores are shown in brackets next to each site identified. Control 1 and 2 represent samples, where KKT23 or AcCoA were missing, respectively.

It is noteworthy that the N-terminal sequence of histone H2A was not covered in the MS results (data not shown). This is a common problem with N-terminal tails of histones, which contain multiple trypsin sites making them difficult to be detected in trypsin-digested samples. It is possible that the identified acetylation sites in *T. brucei* recombinant histones are the baseline acetylation and the main site is located within the N-terminal region of histone H2A. To test this, I generated the deletion mutant of H2A, where I removed 15 residues at the N-terminus (H2A^{16–134}) and compared its acetylation with that of the full-length H2A (Figure 6.16C, D). Interestingly, the mutant was no longer acetylated in my assay, suggesting that the N-terminal region of H2A contains the acetylation site. This needs to be confirmed

using MS analysis, where different proteolytic enzymes (Arg-C, Glu-C) will be used to increase the coverage of the detected peptides. In the assay mentioned above, KKT23^{125–348} was used instead of the full-length KKT23 (Figure 6.16C, D). KKT23^{125–348} was still able to acetylate histone H2A *in vitro*, suggesting that the N-terminal fragment of KKT23 (2–124) is not required for the enzymatic activity of the protein.

In summary, using the acetylation assay, I confirmed that KKT23 has acetyltransferase activity and acetylates histone H2A and H2AZ *in vitro*. More work needs to be done to confirm these results and then to identify the acetylation sites on both histones.

6.4. Mutagenesis of the KKT23 acetyltransferase domain

The closest structural homolog of KKT23^{125–348}, the human Gcn5 acetyltransferase domain (also known as KAT2A), is widely conserved among eukaryotes (Figure 6.17). To date, extensive studies have been done on Gcn5 in many model organisms (Dyda et al., 2000). Detailed work on yeast Gcn5 led to the identification of critical residues for its enzymatic activity (Kuo et al., 1998, Trievel et al., 1999); these include residues responsible for coordinating coenzyme A and residues that act as the catalytic sites. In yGcn5, the role of the catalytic site is performed by E173 (Kuo et al., 1998, Tanner et al., 1999, Trievel et al., 1999). Mutation of E173 to glutamine led to the reduction of histone acetylation *in vitro* and *in vivo* (Kuo et al., 1998, Tanner et al., 1999, Trievel et al., 1999). Corresponding residues were identified in human (E575) and murine (E568) Gcn5. Similarly to yeast Gcn5, mutating these residues led to the loss of Gcn5 enzymatic activity in these species (Kuo et al., 1998, Tanner et al., 1999, Trievel et al., 1999, Bu et al., 2007). The sequence and structural alignments of the KKT23 histone acetyltransferase (HAT) domain and HAT domains from other eukaryotic species show that E235 is likely to be the catalytic site in KKT23 (Figure 6.17). However, the alanine mutant of E235 did not lead to the loss of acetylation activity in KKT23 against histone H2A (Figure 6.18A, B). Although the single mutation is sufficient to abolish the acetylation activity in yeast and mice Gcn5, both E575 and D615 are mutated to generate the catalytically dead mutant in human Gcn5 (Orpinell et al., 2010, Fournier and Tora, 2017). Although the corresponding double mutant (E235A, D275A) of KKT23^{125–348} resulted in a reduced level of H2A acetylation *in vitro*, it did not completely abolish the enzymatic activity of KKT23 (Figure 6.18A, B). This suggests that the catalytic site in KKT23 might be different from Gcn5 in other organisms.

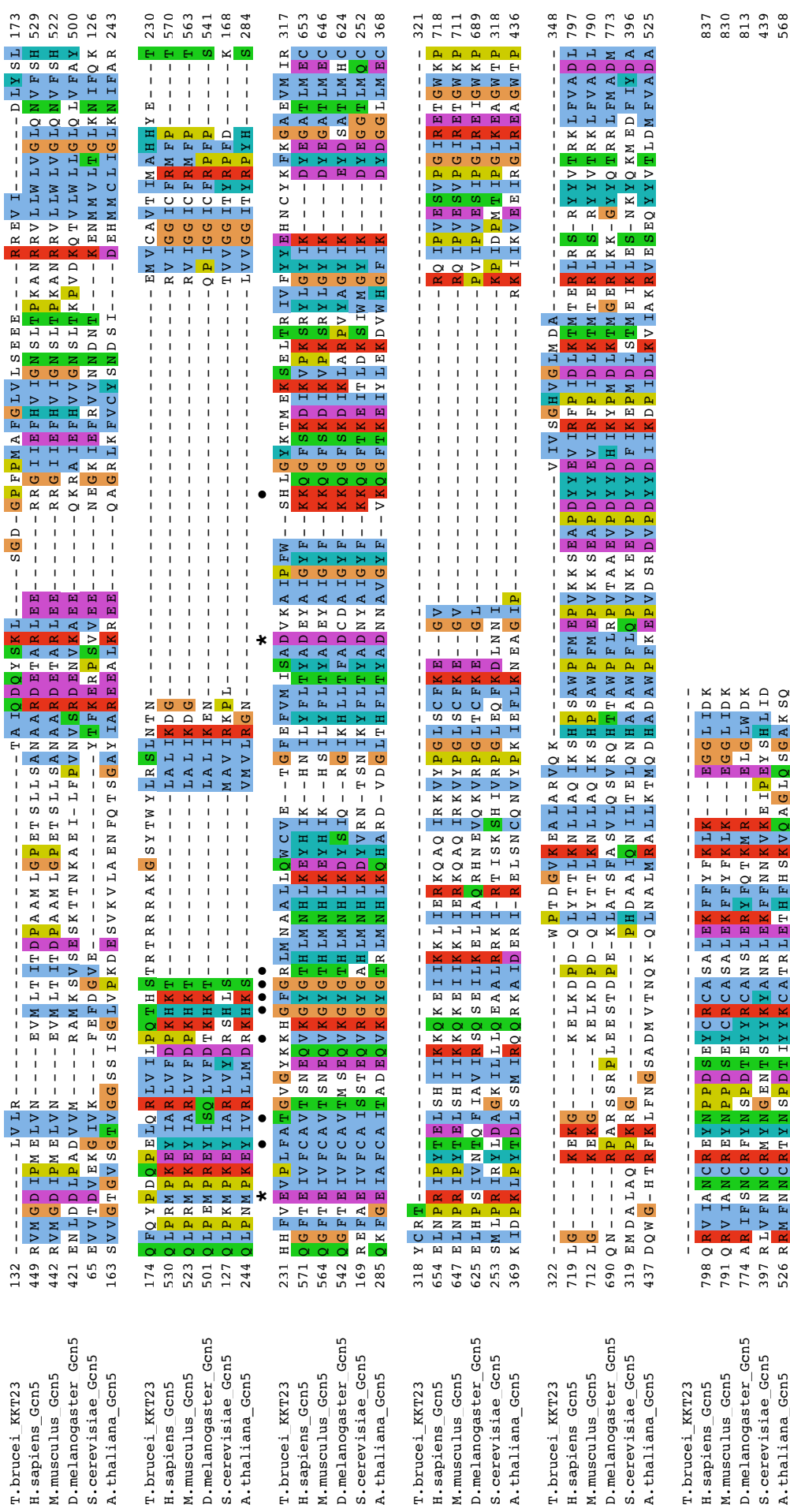


Figure 6.17. Multiple sequence alignment between the KKT23 GNAT domain and Gcn5 proteins from other eukaryotes.

The position of predicted catalytic sites (E235 and D275) is indicated by asterisks; these residues are conserved between KKT23 and Gcn5 from other species.

The residues in KKT23 that were hydrogen-bonded to AcCoA (Table 6.2) are indicated with black circles.

To identify residues critical for KKT23 enzymatic activity, point mutants of KKT23^{125–348} were generated based on the information from the crystal structure, sequence alignments and analysis from PDBePISA. H226 is conserved among kinetoplastids (Figure 1.9) and could be involved in the acetyl transfer based on its close proximity to E235. Alanine mutant of H226 was generated to test the functional relevance of this residue. P237 is highly conserved among kinetoplastid species (Figure 1.9) but is not present in Gcn5 from other eukaryotes (Figure 6.17). Alanine mutant of P237 was generated to test its importance for KKT23 function or structure. F239, T241, G249, G251 and H284 form hydrogen bonds with AcCoA in the crystal structure (Table 6.2, Figure 6.5). Mutation of these residue to alanine should destabilise the interaction between KKT23 and AcCoA, which subsequently will lead to the loss of acetyltransferase activity; this strategy was successfully used in studies on yeast Gcn5 (Kuo et al., 1998, Trievel et al., 1999). Lastly, W282 was mutated based on making hydrophobic contacts with AcCoA (Figure 6.5) and its close proximity to the acetyl group (Figure 6.18C). In summary, following residues were mutated: H226, P237, F239, T241, G249, G251, R252, W282 and H284 (Figure 6.18A–C).

Alanine mutants of H226, G251 and W282 led to a total loss of acetylation activity *in vitro*. G249A, H284A resulted in a significant reduction of the enzymatic activity, whereas E235A, E235A/P237A, P237A, F239A and T241A did not significantly affect the levels of acetylation (Figure 6.18A, B). To determine whether the mutations that led to a complete loss of the enzymatic activity of KKT23^{125–348} (H226A, G251A, W282A) affected the domain structure, 1D NMR spectra were collected for corresponding mutants and compared to the spectrum of the wild-type KKT23^{125–348}. None of the mutations affected the protein folding (data not shown), suggesting that the defect in acetyltransferase activity was not due to the loss of the protein structure.

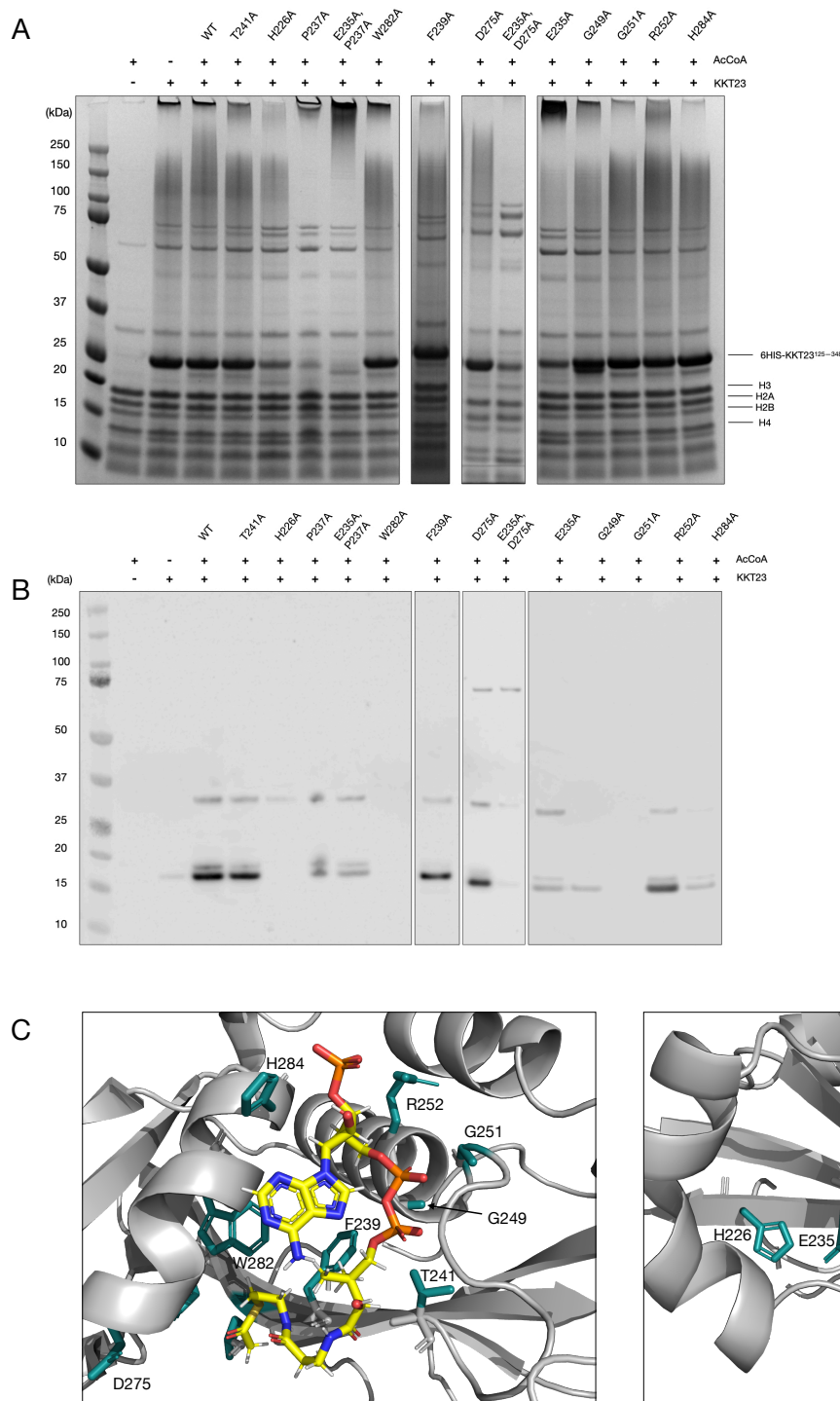


Figure 6.18. Mutagenesis of KKT23 identifies key residues required for its enzymatic activity.

(A) The SDS-PAGE analysis of *in vitro* acetylation assay performed to test the activity of KKT23^{125–348} mutants. The core histones (H3/H2A/H2B/H4) were used as substrates. In each reaction comparable amounts of substrates and KKT23 were used. Lower amounts of KKT23^{125–348} P237A, E235A/P237A and E235A/D275A mutants were loaded compared to the wild-type protein, which likely affect the lower acetylation levels observed in (B).

- (B) Western blot analysis of the reactions shown in (A) using anti-AcLys antibodies. The complete loss of acetylation activity was observed with H226A, W282A and G251A mutants. A weaker defect was observed with E235A/D275A, G249A and H284A mutants. The reduction in signal observed with E235A, E235A/P237A can be explained by the lower amount of the mutants used in the assay (A).
- (C) The side-chains of the mutated residues reported in (B) are shown in teal in the crystal structure of KKT23^{125–348}. The acetyl coenzyme A is shown as sticks coloured by elements. It is noteworthy that most of the highlighted residues are located in motif A as reported in the previous studies (Dyda et al., 2000). These residues are in close proximity to the AcCoA.

6.5. KKT23 forms a complex with KKT3 and KKT22

KKT23 was identified only in the immunoprecipitation of KKT3 and KKT22 (Nerusheva et al., 2019), suggesting that KKT3, KKT22 and KKT23 are in close proximity at the kinetochore. Moreover, it has been shown that KKT23 no longer localises at the kinetochore upon induction of KKT3 RNAi (Marciano et al., 2021). This suggests that KKT3 may directly interact with KKT23 and recruit it to the kinetochore. To test this possibility, recombinant proteins were expressed in insect cells. FLAG-KKT3 was not able to pull down either KKT22 or KKT23 (Figure 6.19A). Similarly, FLAG-KKT22 or FLAG-KKT23 were not able to pull down KKT3 (Figure 6.19A). In contrast, KKT22 and KKT23 co-purified with each other. This suggests that KKT22 and KKT23 might directly interact but neither is able to pull down KKT3 independently. To test if the presence of all three proteins is required to form a complex, FLAG-KKT22 was expressed in the presence of both KKT3 and KKT23. Interestingly, proteins robustly co-purified when combined. Similar results were obtained with FLAG-KKT23 in the presence of untagged KKT3 and KKT22 (Figure 6.19A).

To identify the interacting surfaces within the complex, crosslinking mass spectrometry was employed using BS³ and EDC (Figure 6.19B). The crosslinking with BS³ resulted in aggregates. However, the reaction with EDC resulted in a crosslinked bands of approximately 140 kDa in size. A decrease in intensities of the bands corresponding to KKT3 and KKT22 was observed, suggesting that these proteins took part in the reaction. Although the MS identified proteins of interest in the crosslinked bands, no crosslinked peptides were detected following exposure to trypsin. More work needs to be done in the future to optimise the crosslinking reaction, proteolytic cleavage and mass spectrometry. An alternative approach is to express truncations of the complex components to determine the interaction sites.

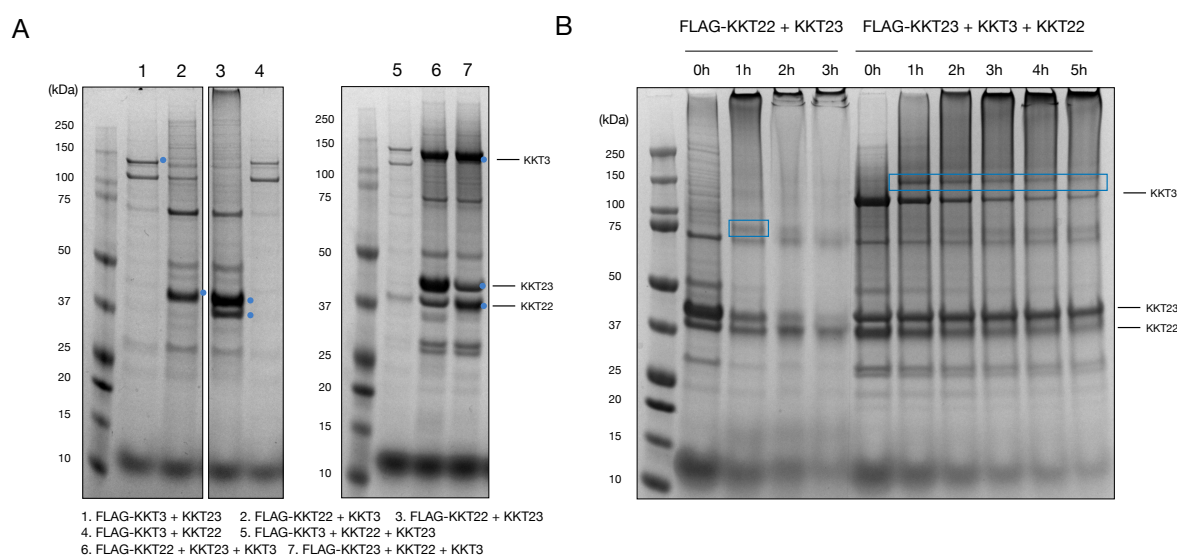


Figure 6.19. KKT23 forms a complex with KKT3 and KKT22.

- (A) Immunoprecipitation of KKT3, KKT22 and KKT23 from insect cells. KKT23 co-purified with FLAG-KKT22 (lane 3). To pull down KKT3, expression of both KKT22 and KKT23 is required (lane 6 and 7).
- (B) The crosslinking reactions using EDC/Sulfo-NHS were performed on KKT22/23 and KKT3/22/23. The crosslinked species are highlighted with blue boxes.

Chapter conclusions

- The crystal structure of the KKT23 C-terminal region reveals similarity to a histone acetyltransferase domain.
- The N-terminal region of KKT23 consists of a helical region followed by a disordered tail.
- KKT23 specifically acetylates histone H2A and the H2AZ variant.
- H226, G251 and W282 were found to be critical for the acetylation activity *in vitro*.
- KKT3, KKT22 and KKT23 form a complex *in vitro*.

CHAPTER 7

7. Discussion

7.1. KKT4

Many of the microtubule-binding proteins, such as the Ndc80, Ska and Dam1 complexes, SKAP/Astrin, CENP-E, CENP-F, MCAK, INCENP, XMAP215, and dyneins, have been extensively studied in other model organisms (Maiato et al., 2004, Foley and Kapoor, 2013, Musacchio and Desai, 2017). My work has shown that KKT4, the only known microtubule-binding component of the kinetoplastid kinetochores, has a number of unique features for a microtubule-binding protein. Overall, using crystal structures, NMR and crosslinking mass spectrometry data, I have generated a structural model of KKT4^{115–645} (Figure 7.1). It shows that the N-terminal region of microtubule-binding domain has a coiled-coil structure followed by a long unstructured tail with tandem BRCT domains at the C-terminal part of KKT4. Although a number of kinetochore proteins have coiled coils, the microtubule-binding domains are typically located elsewhere, such as the calponin-homology domains in Ndc80/Nuf2 (Wei et al., 2007, Ciferri et al., 2008). In the case of SKAP/Astrin that has predicted coiled coils, it has been shown that the coiled-coil segment

is unable to interact with microtubules on its own and that the N-terminal disordered fragment is required for its microtubule-binding activity (Frieze et al., 2016). In contrast to SKAP/Astrin, KKT4's coiled coil is sufficient to interact with microtubules. Besides the structured domains, the unstructured regions have been shown to contribute to microtubule-binding activities of many proteins (Guimaraes et al., 2008, Frieze et al., 2016, Volkov, 2020) but in most cases, the predicted disorder has not been experimentally confirmed. NMR confirmed that the C-terminal fragment of the microtubule-binding domain of KKT4 is disordered. Moreover, the C-terminal tail has been shown to contribute to the microtubule-activity of KKT4 but is not sufficient to interact with microtubules (Figure 3.24B). Using site-directed mutagenesis, I identified residues important for the microtubule-binding activity in *T. brucei* KKT4 (Figure 3.26); these residues are located within the N-terminal basic coiled-coil region (Figure 3.23). My work could be extended by implementation of cryogenic electron microscopy that could reveal the structure of KKT4 bound to microtubule protofilaments. Moreover, mutations shown to affect the microtubule-binding activities of KKT4 should be tested *in vivo*.

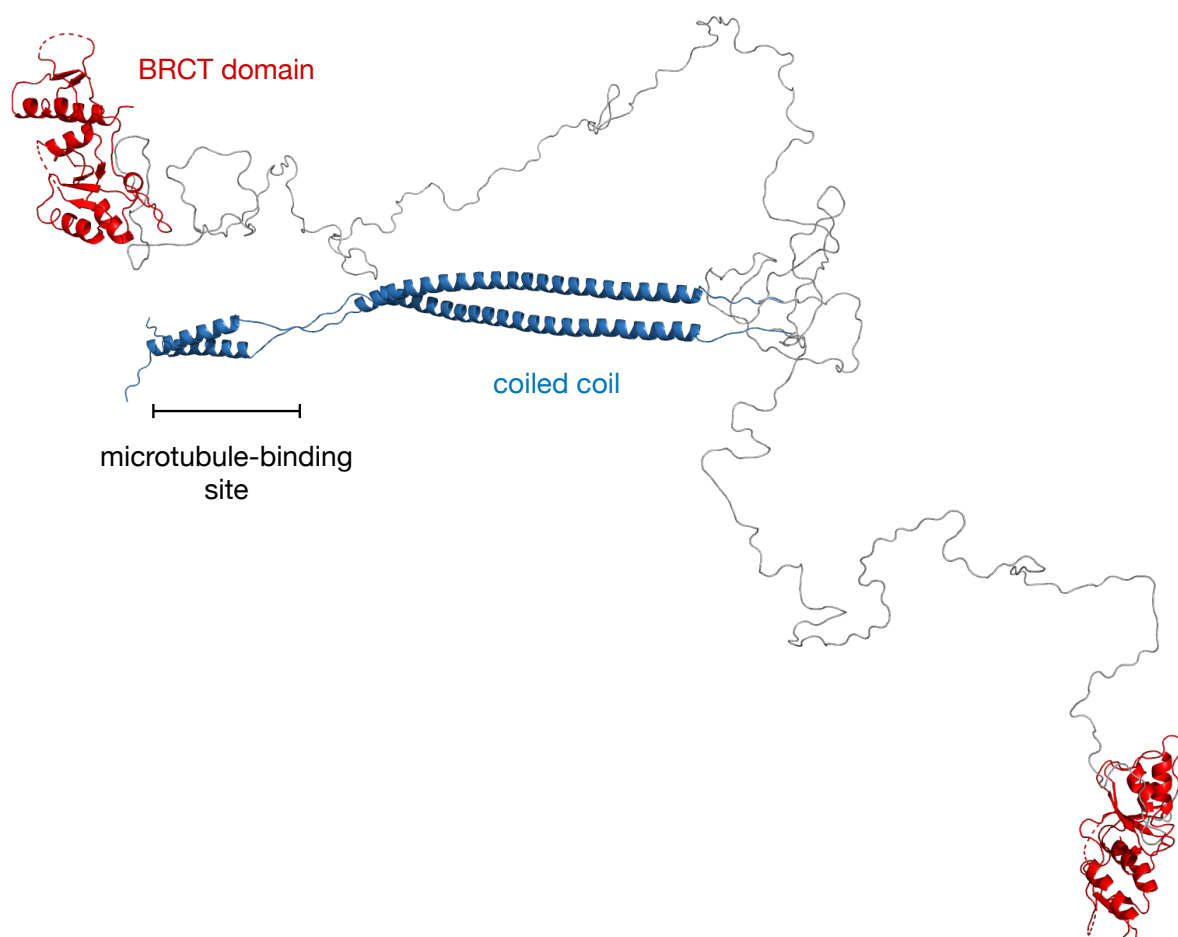


Figure 7.1. The structural model of *T. brucei* KKT4^{115–645}.

This model was generated using X-ray crystallography, NMR data and modelling. The residues that are most important for the KKT4 microtubule-binding activity are located in the region indicated by the bar. The model shows two possible orientations of the BRCT domain (shown in red) with respect to the coiled coil (shown in blue) based on random conformations for the disordered region between Q233 and T473. These disordered conformations allow the BRCT domain to interact with the N-terminal region of the coiled coil. However, alternative conformations in which the BRCT domain is distant from the coiled coil are also possible.

Interestingly, my preliminary data show that KKT7 is another KKT protein with microtubule-binding activities (section 8.3, Appendix). More work needs to be done to confirm this result and further narrow down the region of KKT7 that is sufficient to interact with microtubules. Eventually, the relevance of the microtubule-binding activity needs to be tested *in vivo* in *Trypanosoma brucei* cells. Like in KKT4, other activities associated with microtubules need to be confirmed in KKT7, including tracking depolymerising microtubules, forming load-bearing attachments or ability to diffuse along the microtubule lattice.

Using X-ray crystallography and NMR spectroscopy I have shown that the C-terminal region of KKT4 adopts the BRCT domain fold. The BRCT domain in KKT4 seems to be a phosphorylation-dependent protein-protein interaction domain. It was shown to interact with phosphorylated peptide derived from KKT8 with a moderate K_D of 30 μ M. The same peptide sequence, lacking phosphorylation, binds to the BRCT domain with a significantly weaker affinity. Besides KKT8, there might be other protein(s) that interacts with the BRCT. Using an insect-cell expression system, I have shown that KKT4 co-purifies with KKT3 (Figure A3). However, both proteins have significant DNA contamination when purified from insect cells. Based on the fact that KKT4 can strongly interact with the DNA (Figure 4.1A), it will be important to test if the KKT4-KKT3 interaction is not indirect via the DNA and if not whether the BRCT is mediating that interaction. This can be achieved using crosslinking mass spectrometry or by co-expressing KKT3 and KKT4 truncations. Moreover, NMR analysis has shown that the KKT4 BRCT domain directly interacts with the N-terminal part of the microtubule-binding domain but does not modulate the microtubule-binding activities of KKT4 (Figure 5.13C, D). Based on the fact that the observed chemical shifts in KKT4 are small, I speculate that the phosphorylation

within the N-terminal region of the microtubule-binding domain may increase the affinity between these two regions and possibly regulate other KKT4 activities. KKT4 was shown to co-purify with the APC/C subunits (Akiyoshi and Gull, 2014). The interaction between the BRCT domain and the microtubule-binding domain might be governed by the attachment status, which in turn controls APC/C activities and the cell cycle progression.

In other eukaryotes, the Aurora B kinase is the crown regulator of the kinetochore-microtubule attachments by phosphorylating microtubule-binding kinetochore proteins, including Ndc80 and the Ska complexes (Cheeseman et al., 2002, Tien et al., 2010, Chan et al., 2012, Redli et al., 2016). It is unknown if Aurora B regulates kinetochore-microtubule attachments in trypanosomes (Tu et al., 2006). The unpublished work in our lab show that *T. brucei* Aurora B phosphorylates KKT4 microtubule-binding domain suggesting that Aurora may regulate microtubule-binding activity of KKT4. It was previously shown that KKT4 is also phosphorylated by the KKT10 kinase, which promotes the metaphase-anaphase transition (Ishii and Akiyoshi, 2020). A phospho-deficient KKT4^{S477A} mutant failed to rescue the growth defect caused by KKT4 RNAi (Ishii and Akiyoshi, 2020). Although the underlying molecular mechanism remains unknown, it is noteworthy that S477 is located just prior to the BRCT domain. In this study, we identified KKT8 as a putative binding partner for the KKT4 BRCT domain. It will be important to identify which kinase(s) phosphorylates the KKT8 S381 site to promote the interaction. Because kinetochore localisation of the KKT10 kinase depends on the KKT8 complex (composed of KKT8, KKT9, KKT11, and KKT12) (Ishii and Akiyoshi, 2020), it is possible that KKT10's role in regulating the metaphase-to-anaphase transition is controlled by the

KKT4^{BRCT}-KKT8 interaction. These hypotheses will need to be tested in the future to better understand the mechanism of chromosome segregation in trypanosomes.

Interestingly, KKT4 localises constitutively at the kinetochore and was shown to localise closer to the putative DNA-binding component, KKT2 rather than the outer kinetochore component, KKIP1 (Llauro et al., 2018). KKT4 has been found to strongly interact with the centromeric DNA *in vitro* (Figure 4.1A). The DNA-binding activity has been narrowed down to two fragments: KKT4^{114–232} and KKT4^{300–488}. My *in vivo* analysis failed to establish whether the DNA-binding activity of KKT4 is important for the proliferation of trypanosome cells. It is noteworthy that no other eukaryotic kinetochore protein has been shown to interact with DNA in addition to its microtubule-binding activities. To date, KKT4 is the only kinetochore protein that is known to bind both DNA and microtubules. A functionally similar protein exists in the bacterial R1 plasmid partitioning system that uses a single component (ParR) to mediate connections between the plasmid DNA (parC) and filament-forming protein ParM. In the future, additional work needs to be done to reveal the function of KKT4 DNA-binding activity and determine its importance in trypanosomes.

7.2. KKT23

Since its discovery (Nerusheva et al., 2019), KKT23 was demonstrated to be essential for cell growth and chromosome segregation (Marciano et al., 2021). However, its structure and function remained unknown. The fact that in other eukaryotes, no kinetochore protein is known to have a GNAT domain highlights that KKT23 is a unique kinetochore protein. Moreover, KKT23 is conserved in all sequenced kinetoplastids, suggesting that KKT23 may play a fundamental role in the kinetochore function. My preliminary data show that KKT23 can bind DNA

(Figure 7.2A), which is consistent with its constitutive localisation (Nerusheva et al., 2019). The region responsible for the DNA-binding activity in KKT23 has yet to be determined. KKT23 cannot interact with microtubules (Figure 7.2B).

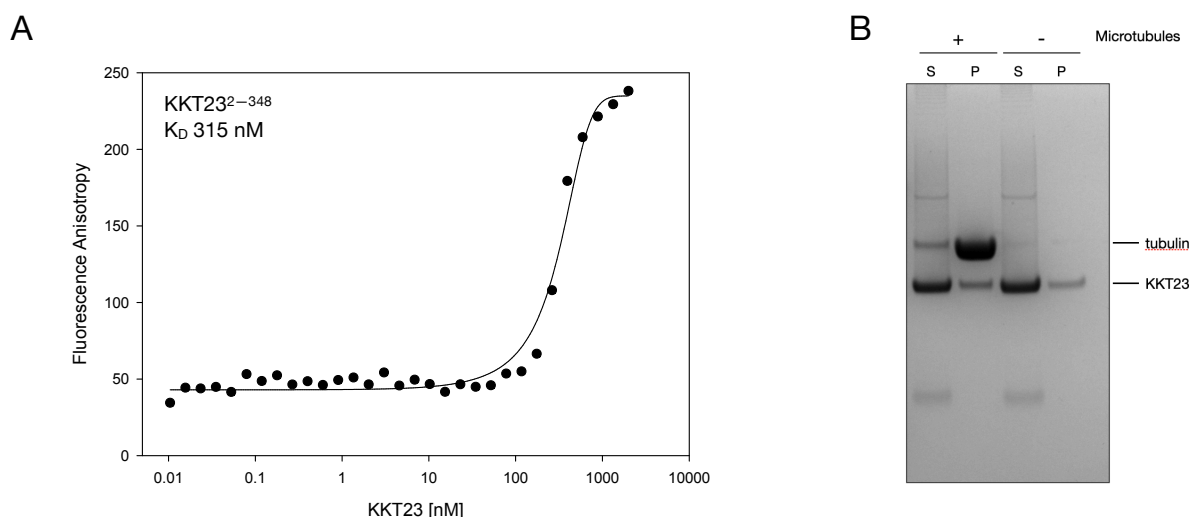


Figure 7.2. KKT23 has DNA-binding activities.

(A) Fluorescence anisotropy assay using the full-length KKT23. Measured anisotropy is plotted against protein concentration. In this assay a 50-bp DNA probe (~36% GC content) was used, which is part of the centromeric sequence (CIR147) in *T. brucei*. The K_D was calculated using non-linear regression using SigmaPlot (Monks, 2002); the fit is shown as a solid line. Microtubule co-sedimentation assay showing that KKT23 does not interact with taxol-stabilised microtubules.

The structural information obtained using X-ray crystallography and NMR confirmed that the C-terminal region of KKT23 has a histone acetyltransferase fold with human and yeast Gcn5 being its closest structural homologues (Table S4 and Figure 6.4). In line with this discovery, I show that KKT23 can acetylate histone H2A and H2AZ *in vitro*. Other histones were not acetylated in my assay, demonstrating that histone H2A is a specific substrate of KKT23 (Figure 6.16A, B). My preliminary data suggest that the acetylation site is located within the 15 N-terminal residues of H2A (Figure

6.16C, D), the region not found to be acetylated in previous studies (Janzen et al., 2006, Mandava et al., 2007, Kraus et al., 2020). I plan to identify the acetylation site(s) by mass spectrometry, raise antibodies against these sites and test whether the KKT23-dependant acetylation occurs specifically at centromeres. Moreover, the functionality of acetylation-deficient and acetylation-mimetic H2A mutants needs to be tested in *T. brucei* cells. I speculate that the acetylation activity of KKT23 could alter the chromatin structure and/or promote the assembly of kinetoplastid kinetochores. The kinetochore assembly may be achieved using acetylated histone H2A as the binding site for other kinetochore proteins. It is also possible that the acetylation of histone H2A destabilises the interaction between the DNA and histones by neutralising the positive charge of lysines, which subsequently could lead to an enhanced DNA accessibility and transcription. (Bannister and Kouzarides, 2011). Transcription has been detected at centromeres in various eukaryotes and is thought to play important roles in ensuring proper kinetochore functions (Talbert and Henikoff, 2018). In fact, small non-coding RNA has been detected for the centromeric repeat sequence of several chromosomes in *T. brucei* (Tschudi et al., 2012).

The mutagenesis identified residues (G251, H226, W282) important for the acetylation activity of KKT23 (Figure 6.18A, B). Mutation of these residues likely affects the ability of KKT23 to coordinate its cofactor, acetyl coenzyme A. Using the identified residues, it will be important to test if the acetylation activity of KKT23 is essential for trypanosomes. Mutation of the highly conserved E568 and D609 in murine Gcn5 (Bu et al., 2007) results in a catalytically dead Gcn5 that is no longer able to acetylate histones *in vitro* and *in vivo*. The same results were later observed with mutants of yeast and human Gcn5 (Tanner et al., 1999, Trievel et al., 1999,

Orpinell et al., 2010, Fournier and Tora, 2017). Surprisingly, the mutation of the corresponding residues in KKT23 (E235 and D275) failed to completely abolish the acetyltransferase activity of the protein (Figure 6.18A, B), suggesting that the catalytic site and mechanism of acetylation may differ from other known histone acetyltransferases. The crystal structure of KKT23 GNAT domain bound to both AcCoA and histone H2A should reveal the catalytic site in KKT23.

Although I was not able to establish the potential function of the KKT23 N-terminal region, my preliminary structural data suggest that it is helical (Figure 6.13). Consistent with this observation, the heteronuclear NOE analysis confirmed that KKT23²⁻⁷⁰ has a rigid protein backbone (data not shown). In the future, it will be important to solve the high-resolution structure of KKT23 N-terminal region, which may hint at its function. The data should be complimented with an *in vivo* analysis to help establish whether the N-terminal residues of KKT23 are important for an accurate cell division or whether it is required for kinetochore localisation of KKT23.

Appendix

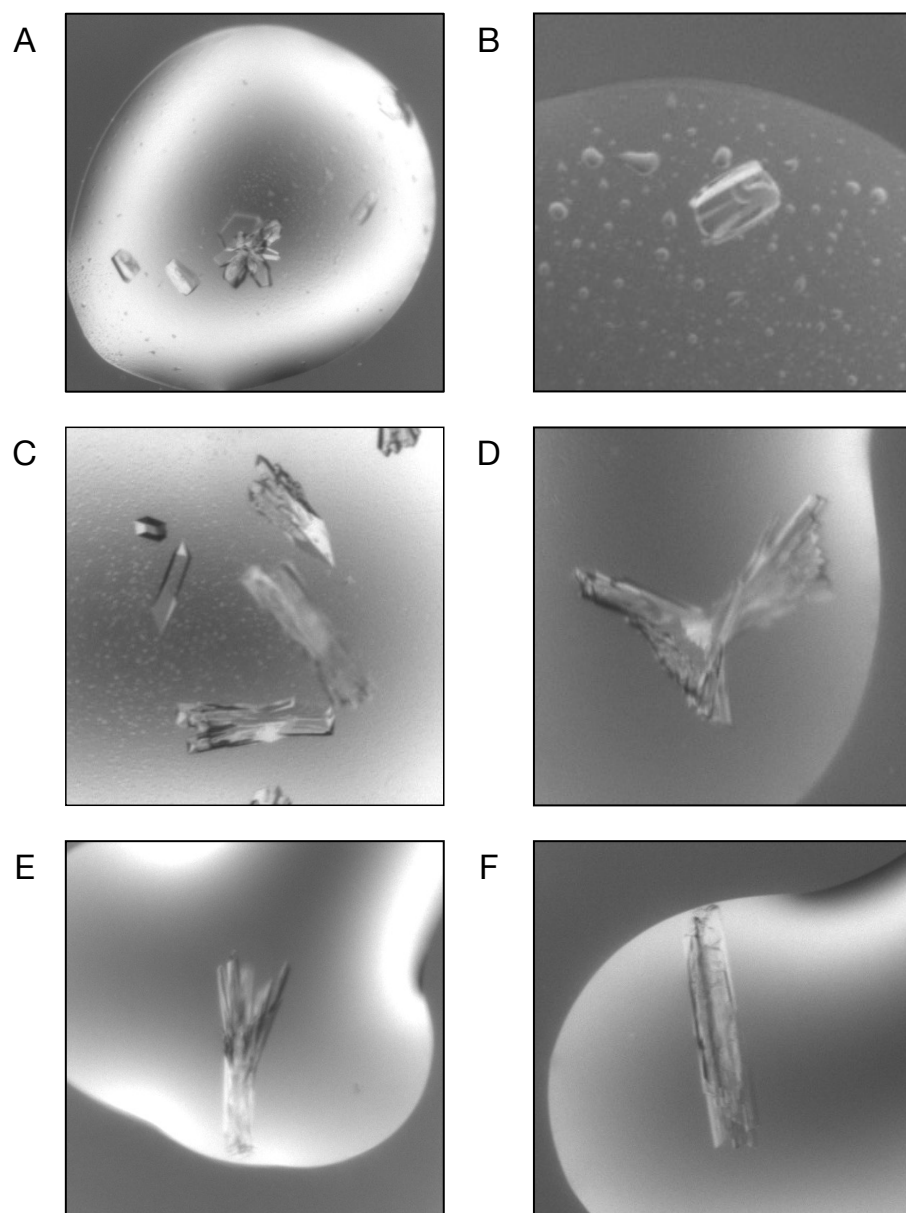


Figure A1. Images of protein crystals used to solve the structures reported in this thesis.

(A) Crystals of *Trypanosoma cruzi* KKT4^{117–218}.

(B) Crystals of *Leishmania mexicana* KKT4^{184–284}.

(C) Crystals of *Trypanosoma brucei* KKT4^{463–645}.

(D) Crystals of *Trypanosoma brucei* KKT4^{463–645} K543A.

(E) Selenomethionine crystals of *Trypanosoma brucei* KKT23^{125–348}.

(F) Crystals of *Trypanosoma brucei* KKT23^{125–348}.

Crystallisation attempts of *T. brucei* KKT4 microtubule-binding domain

Unstructured and highly flexible protein fragments are not suitable for protein crystallisation as they can prevent crystal formation. *In situ* proteolysis and *in vitro* limited proteolysis have been successfully used in protein X-ray crystallography. In these techniques, the protein of interest is incubated with low concentrations of proteolytic enzyme. This process, if successful, removes flexible and/or disordered parts of the protein resulting in construct more structured, therefore, more likely to crystallise. KKT4^{115–343} was incubated with increasing concentration of trypsin at 4°C (Figure A2). The proteolysis resulted in a ~10 kDa band, which was processed and subjected for mass spectrometry analysis to reveal the protein sequence. Peptides derived from mass spectrometry analysis revealed that the protected fragment corresponds to KKT4^{145–230}. Nine KKT4 constructs were generated that had different N- and C-termini ends. Three of them, KKT4^{138–234}, KKT4^{145–230} and KKT4^{145–232} showed highest yield during the purification, and were chosen for initial crystallisation screening. KKT4^{145–232} resulted in several crystals, formed in Natrix HT screen (Hampton Research). Crystals of KKT4^{145–232} diffracted to 4.8 Å resolution. The conditions, where protein crystallised were optimised by increasing the drop size and use of additive screen. This resulted in an improved diffraction pattern up to 2.8 Å resolution. To obtain an anomalous dataset, selenomethionine derivative of KKT4^{145–232} was crystallised. Although the crystals diffracted up to ~3 Å, the selenomethionine signal was too low to solve the phase problem (there are only 3 methionines within KKT4^{145–232}, one of which is located at the very N-terminus and is likely unstructured). Molecular replacement using available structures of regular coiled coils and crystal structures of *Tc*KKT4^{117–218} or *Lm*KKT4^{184–284} were not successful. At the same time, a different approach, an *in situ*

proteolysis was carried out. In this experiment, KKT4^{115–343} was mixed with the pre-made solution of trypsin protease; this mixed reaction was used for setting crystallisation trays. Interestingly, the protein yielded crystals in the same conditions as KKT4^{145–232}, however none of the crystals diffracted.

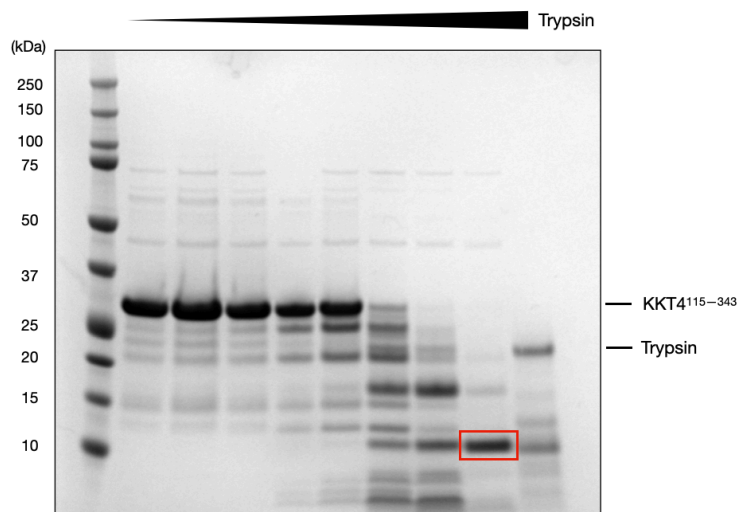


Figure A2. Limited proteolysis of KKT4^{115–343}.

Samples containing 20 µg of KKT4^{115–343} in 50 mM sodium phosphate pH 7.5, 200 mM NaCl and 0.5 mM TCEP were mixed with increasing amounts (0, 0.002, 0.004, 0.02, 0.04, 0.2, 0.4, 2 and 200 µg – lanes 2-10) of trypsin in 25 mM Tris pH 8.0, 2 mM CaCl₂ in a 1:1 v/v ratio. Following incubation on ice for 1 hour, samples were analysed using SDS-PAGE to assess the degree of proteolysis. The band highlighted in red, identified following incubation with 2 µg trypsin, was cut out and subjected to mass spectrometry analysis. This revealed that the trypsin-resistant fragment corresponds to residues 145–230 of KKT4 (KKT4^{145–230}).

Interaction partners for KKT4

So far, direct interaction partners of KKT4 are unknown. Identifying which KKT proteins can interact with KKT4 could reveal the mechanisms of how KKT4 is recruited to the kinetochore, what other proteins KKT4 may recruit or how KKT4's microtubule- and DNA-binding activities are regulated. Mass spectrometry of the immunoprecipitated YFP-KKT4 from *Trypanosoma brucei* identified several KKT proteins that co-purified with KKT4: KKT1, KKT2, KKT6, KKT7, KKT8, KKT9, KKT10, KKT11, KKT12 and KKT20 with KKT7 being the top hit (Akiyoshi and Gull, 2014); this suggests that KKT4 is likely to be in close proximity with those kinetochore components mentioned above. However, it remains unknown which KKT protein (if any) directly interacts with KKT4. To screen for potential KKT4-binding proteins, individual KKT proteins were expressed in *Spodoptera frugiperda* (Sf9) cells and tested for interaction with KKT4 using immunoprecipitation (Figure A3). KKT7 as the top hit identified in IP of KKT4 was a good candidate for KKT4 interaction partner. However, KKT7 was not pulled-down with FLAG-KKT4 (Figure A3, lane 2). Next, the following KKT proteins have been tested: KKT2, KKT3, KKT6/1 complex and KKIP1. KKT8-9-11-12 complex, shown to co-purify with KKT4 in IP from trypanosomes was not soluble in insect cells, and could not be tested. In contrast, KKT3 co-purified with KKT4 in a stoichiometric ratio (Figure A3, lane 1). The apparent interaction between KKT4 and KKT3 needs to be further investigated because both proteins strongly interact with DNA (Figure 4.1A and data not shown) and the observed co-purification may be indirect via nucleic acids. Size elution chromatography was performed to see if two proteins co-migrate in the complex. However, KKT3 tends to precipitate in the size exclusion column. Further optimisation needs to be carried out to characterise the interaction between KKT4 and KKT3.

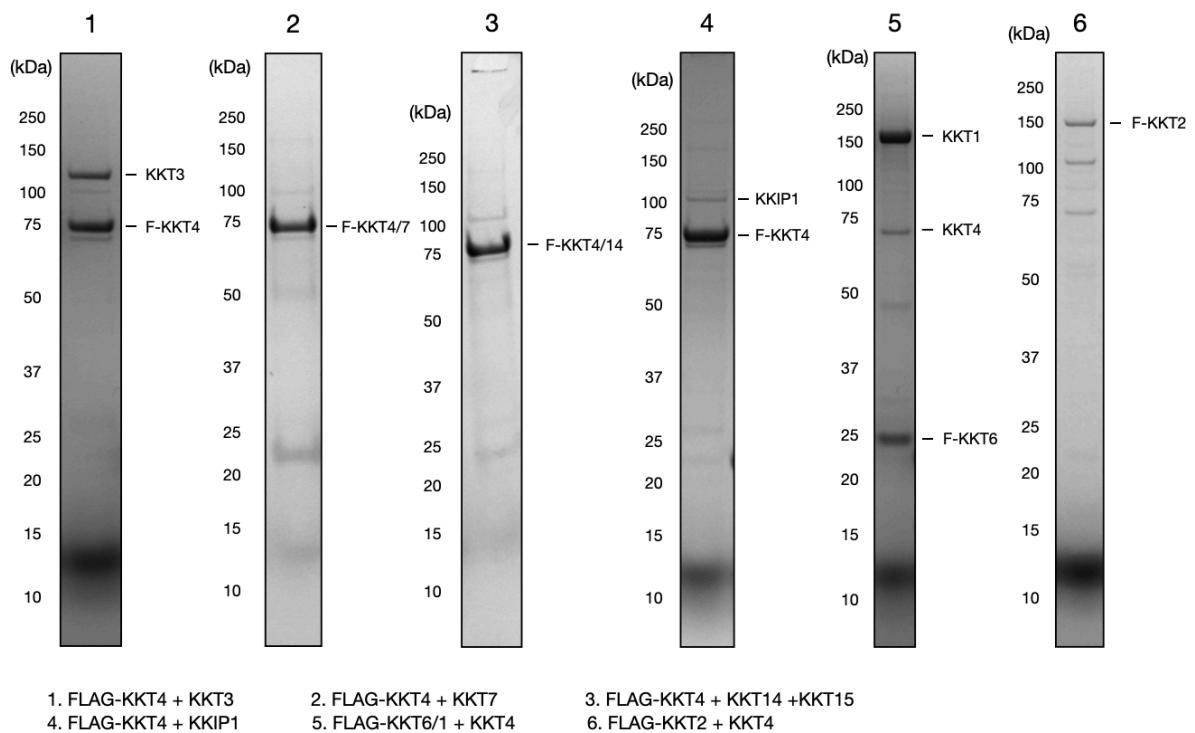


Figure A3. The immunoprecipitation of KKT4 with other KKT proteins from insect cells.

Lane 1 shows that KKT3 co-purifies with FLAG-KKT4. This co-purification might be indirect via the DNA (both proteins have DNA-binding activity). KKT7 is the same size as KKT4; therefore, these protein are indistinguishable using SDS-PAGE (Lane 2). However, MS analysis failed to identify KKT7 peptides in the sample (data not shown). KKT14 also overlaps with KKT4 (Lane 3); MS analysis confirmed that KKT4 co-purifies with KKT14 (data not shown). In lane 4, very little amount of KKIP1 was observed compared to the KKT4 making it unlikely that the two proteins interact. In lane 5, little KKT4 was detected in KKT6/1 immunoprecipitation. KKT4 did not co-purify with KKT2 (lane 6). F- in the figure indicates proteins tagged with 3FLAG.

KKT7 interacts with microtubules

Previous studies reported that no microtubule-binding activities have been observed for KKT1, KKT6, KKT10, KKT13, KKT14, KKT16, KKT17 and KKT18. To date, KKT4 is the only microtubule-binding kinetochore protein identified in kinetoplastids (Llauro et al., 2018). While searching for other microtubule-binding candidates, I purified and screened other KKT proteins using microtubule co-sedimentation assay. I have found that KKT8-9-11-12 complex, KKT14 and KKT15 failed to co-sediment with microtubules (data not shown). Excitingly, the full-length KKT7 co-sedimented with taxol-stabilised microtubules (Figure A4). Moreover, the microtubule-binding region was narrowed down to residues between 144 and 512 (Figure A4). In the future it will be important to test if KKT7 can bind and diffuse on microtubule filaments, track depolymerising microtubules or form load bearing attachments with dynamic microtubule tips. Finally, the *in vivo* relevance, molecular mechanism and regulation of the interaction between KKT7 and microtubules will be the subject of future studies.

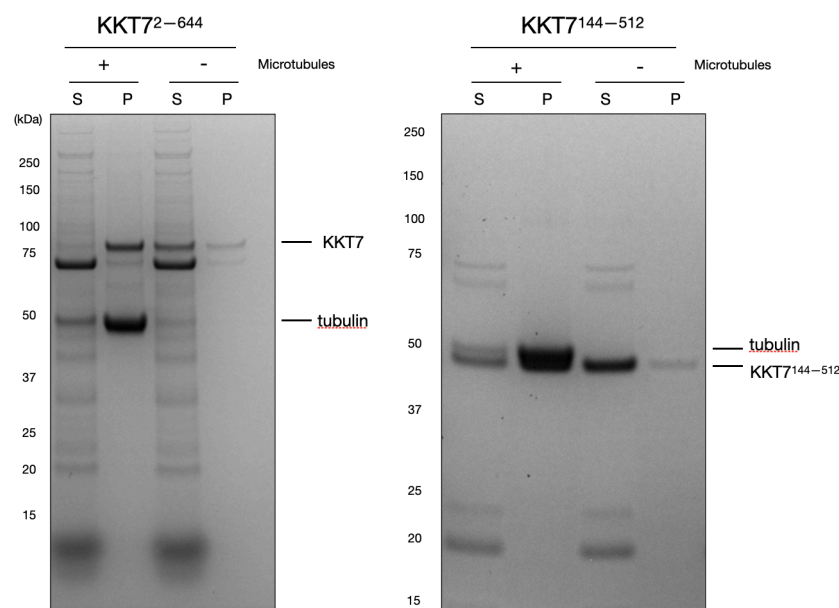


Figure A4. KKT7 co-sediments with microtubules.

Microtubule co-sedimentation assay shows that the full-length KKT7 can interact with taxol-stabilised microtubules (left). The activity was also observed for KKT7 truncation, KKT7¹⁴⁴⁻⁵¹² (right). KKT7 is another candidate for a microtubule-binding protein at the kinetochore.

Supplemental tables

Table S1. List of experiments used to obtain resonance assignments.

Experiment	KKT4 ^{115–343}	KKT4 ^{115–174}	KKT4 ^{145–232}	KKT4 ^{463–645}	KKT23 ^{125–348}
2D ¹ H- ¹³ C HSQC	+	+	+	+	+
2D ¹ H- ¹⁵ N BEST-TROSY	+	+	+	+	+
3D ¹⁵ N-edited NOESY-HSQC		+	+	+	+
3D ¹⁵ N-edited TOCSY-HSQC	+	+	+	+	+
3D CBCANH	+				
3D CBCA(CO)NH	+	+			
3D HNCO	+				
3D HN(CA)CO	+				
3D HBHA(CO)NH	+	+		+	
3D HCA(CO)N	+				
3D HCAN	+				
3D (H)N(CA)NNH	+				
3H (H)N(COCA)NNH	+				
3D (H)CC(CO)NH	+	+		+	+
3D H(CCCO)NH				+	+
3D HCCH-TOCSY		+		+	+
3D BT-HNCA	+	+	+	+	+
3D BT-HNCO		+	+	+	+
3D BT-HN(CA)CO			+	+	+
3D BT-HNCACB			+	+	+
3D BT-HN(CO)CACB			+	+	+

Table S2. Proteins used for crystallisation screening

Protein	Plasmid	Result
<i>T. brucei</i> KKT4		
KKT4 ^{470–645}	pBA985	no crystals obtained
KKT4 ^{115–645}	pBA987	no crystals obtained
KKT4 ^{115–323}	pBA1064	no crystals obtained
KKT4 ^{115–343}	pBA1065	no crystals obtained
KKT4 ^{250–645}	pBA1066	no crystals obtained
KKT4 ^{115–246}	pBA1079	no crystals obtained
KKT4 ^{115–174}	pBA1171	no crystals obtained
KKT4 ^{63–352}	pBA1391	no crystals obtained
KKT4 ^{101–352}	pBA1393	no crystals obtained
KKT4 ^{2–114}	pBA1413	no crystals obtained
KKT4 ^{138–234}	pBA1436	non-diffraction quality crystals
KKT4 ^{145–230}	pBA1440	non-diffraction quality crystals
KKT4 ^{145–232}	pBA1441	crystals: diffraction up to 2.8Å
SelMet-KKT4 ^{145–232}	pBA1441	crystals: diffraction up to 8Å
KKT4 ^{463–645}	pBA1513	structure solved PDB: 6ZPK
KKT4 ^{119–232}	pBA1551	crystals: diffraction up to 6Å
KKT4 ^{114–232}	pBA1572	crystals: diffraction up to 6Å
KKT4 ^{115–232}	pBA1601	non-diffraction quality crystals
KKT4 ^{111–232}	pBA1868	crystals: diffraction up to 6Å
KKT4 ^{123–224}	pBA2054	crystals: diffraction up to 4Å
KKT4 ^{126–224}	pBA2058	crystals: diffraction up to 4Å
KKT4 ^{126–232}	pBA2064	no crystals obtained
KKT4 ^{463–645} K543A	pBA2264	structure solved at 1.8Å
<i>L. mexicana</i> KKT4		
<i>Lm</i> KKT4 ^{146–486}	pBA1537	no crystals obtained
KKT4 ^{184–284} R218Q + 23aa	pBA1618	structure solved PDB: 6ZPJ
KKT4 ^{159–262}	pBA1670	non-diffraction quality crystals

KKT4 ^{184–284} R218Q	pBA1782	non-diffraction quality crystals
KKT4 ^{159–359}	pBA1811	non-diffraction quality crystals
KKT4 ^{159–284}	pBA1821	no crystals obtained
KKT4 ^{166–268}	pBA2129	crystals obtained, no information about diffraction
KKT4 from other kinetoplastid species		
<i>T. cruzi sylvio</i> KKT4 ^{117–233}	pBA1858	no crystals obtained
<i>T. vivax</i> KKT4 ^{117–240}	pBA1900	no crystals obtained
<i>L. seymouri</i> KKT4 ^{121–262}	pBA1901	no crystals obtained
<i>T. congolense</i> KKT4 ^{145–244}	pBA2128	no crystals obtained
<i>T. congolense</i> KKT4 ^{134–244}	pBA2143	no crystals obtained
<i>T. cruzi sylvio</i> KKT4 ^{117–218}	pBA2151	structure solved PDB: 6ZPM
<i>T. brucei</i> KKT23		
KKT23 ^{2–348}	pBA1725	no crystals obtained
KKT23 ^{128–348}	pBA1865	no crystals obtained
KKT23 ^{94–348}	pBA2159	no crystals obtained
KKT23 ^{125–348}	pBA2262	no crystals obtained
KKT23 ^{125–348} + AcCoA	pBA2262	structure solved at 1.8Å
SeMet-KKT23 ^{125–348} + AcCoA	pBA2262	structure solved at 2Å
KKT23 ^{143–348}	pBA2281	crystals obtained, no information about diffraction

Table 9.3. List of crosslinked residues identified in KKT4.

BS ³		EDC	
Links	Score	Links	Score
KKT4(115)-KKT4(115)	5.63E-12	KKT4(110)-KKT4(177)	1.36E-14
KKT4(115)-KKT4(136)	3.37E-12	KKT4(110)-KKT4(177)	3.81E-14
KKT4(115)-KKT4(136)	2.48E-10	KKT4(110)-KKT4(177)	2.62E-13
KKT4(115)-KKT4(136)	4.16E-09	KKT4(110)-KKT4(178)	1.08E-11
KKT4(136)-KKT4(110)	3.68E-08	KKT4(110)-KKT4(178)	4.44E-11
KKT4(198)-KKT4(206)	8.08E-12	KKT4(110)-KKT4(178)	1.17E-10
KKT4(198)-KKT4(206)	1.04E-08	KKT4(110)-KKT4(178)	2.87E-10
KKT4(199)-KKT4(206)	3.47E-17	KKT4(115)-KKT4(122)	6.16E-23
KKT4(199)-KKT4(206)	3.65E-15	KKT4(115)-KKT4(122)	4.22E-20
KKT4(199)-KKT4(206)	9.77E-15	KKT4(115)-KKT4(122)	9.33E-20
KKT4(199)-KKT4(206)	2.17E-10	KKT4(115)-KKT4(122)	9.48E-20
KKT4(199)-KKT4(206)	8.39E-08	KKT4(115)-KKT4(122)	1.99E-18
KKT4(206)-KKT4(199)	5.17E-10	KKT4(115)-KKT4(122)	2.12E-10
KKT4(206)-KKT4(199)	6.41E-10	KKT4(115)-KKT4(122)	1.36E-08
KKT4(206)-KKT4(199)	7.42E-09	KKT4(115)-KKT4(177)	4.36E-15
KKT4(206)-KKT4(199)	2.07E-08	KKT4(115)-KKT4(177)	6.79E-11
KKT4(206)-KKT4(204)	2.36E-08	KKT4(115)-KKT4(177)	1.22E-10
KKT4(206)-KKT4(218)	4.95E-09	KKT4(115)-KKT4(177)	1.32E-10
KKT4(206)-KKT4(618)	7.55E-09	KKT4(115)-KKT4(177)	2.38E-10
KKT4(218)-KKT4(218)	1.54E-14	KKT4(115)-KKT4(177)	5.41E-10
KKT4(218)-KKT4(218)	5.51E-09	KKT4(115)-KKT4(177)	1.15E-09
KKT4(218)-KKT4(304)	8.31E-12	KKT4(115)-KKT4(177)	1.85E-09
KKT4(218)-KKT4(304)	6.07E-11	KKT4(115)-KKT4(177)	9.00E-08
KKT4(218)-KKT4(304)	1.58E-08	KKT4(115)-KKT4(645)	1.85E-10
KKT4(218)-KKT4(304)	2.57E-08	KKT4(132)-KKT4(645)	5.12E-12
KKT4(218)-KKT4(310)	1.55E-08	KKT4(136)-KKT4(160)	2.30E-12
KKT4(218)-KKT4(327)	6.52E-10	KKT4(144)-KKT4(157)	2.32E-08
KKT4(218)-KKT4(350)	2.78E-10	KKT4(144)-KKT4(159)	2.18E-08
KKT4(218)-KKT4(350)	9.30E-09	KKT4(157)-KKT4(132)	7.39E-12
KKT4(218)-KKT4(543)	4.47E-15	KKT4(157)-KKT4(132)	1.05E-11
KKT4(304)-KKT4(218)	2.87E-11	KKT4(157)-KKT4(132)	9.89E-09
KKT4(304)-KKT4(218)	3.38E-11	KKT4(157)-KKT4(132)	2.14E-08
KKT4(304)-KKT4(218)	9.34E-10	KKT4(160)-KKT4(132)	1.05E-11
KKT4(304)-KKT4(218)	2.36E-09	KKT4(160)-KKT4(132)	1.28E-08
KKT4(304)-KKT4(218)	2.03E-08	KKT4(177)-KKT4(110)	7.93E-10
KKT4(304)-KKT4(350)	6.32E-15	KKT4(177)-KKT4(110)	3.07E-08

KKT4(304)-KKT4(350)	1.38E-12	KKT4(177)-KKT4(349)	4.58E-17
KKT4(304)-KKT4(543)	7.95E-09	KKT4(177)-KKT4(349)	1.92E-13
KKT4(326)-KKT4(310)	1.60E-08	KKT4(178)-KKT4(115)	5.67E-08
KKT4(326)-KKT4(350)	4.20E-10	KKT4(204)-KKT4(645)	2.99E-12
KKT4(326)-KKT4(350)	6.29E-09	KKT4(218)-KKT4(645)	3.75E-09
KKT4(349)-KKT4(350)	4.58E-10	KKT4(218)-KKT4(645)	1.44E-08
KKT4(357)-KKT4(218)	1.35E-25	KKT4(326)-KKT4(177)	1.53E-09
KKT4(357)-KKT4(218)	6.12E-19	KKT4(326)-KKT4(177)	3.03E-08
KKT4(499)-KKT4(218)	3.97E-15	KKT4(326)-KKT4(177)	6.74E-08
KKT4(499)-KKT4(327)	4.68E-08	KKT4(326)-KKT4(178)	2.71E-11
KKT4(499)-KKT4(350)	6.55E-10	KKT4(326)-KKT4(343)	2.08E-11
KKT4(499)-KKT4(350)	2.34E-08	KKT4(326)-KKT4(343)	8.95E-09
KKT4(499)-KKT4(356)	5.39E-11	KKT4(326)-KKT4(645)	1.31E-11
KKT4(499)-KKT4(356)	1.25E-09	KKT4(339)-KKT4(350)	1.41E-12
KKT4(510)-KKT4(218)	9.89E-23	KKT4(339)-KKT4(350)	1.11E-08
KKT4(510)-KKT4(218)	7.16E-19	KKT4(343)-KKT4(350)	3.62E-17
KKT4(510)-KKT4(218)	1.76E-12	KKT4(343)-KKT4(350)	4.08E-13
KKT4(510)-KKT4(218)	4.67E-12	KKT4(343)-KKT4(350)	5.60E-12
KKT4(510)-KKT4(218)	6.23E-10	KKT4(343)-KKT4(350)	1.50E-10
KKT4(510)-KKT4(310)	8.59E-08	KKT4(343)-KKT4(350)	3.36E-09
KKT4(521)-KKT4(115)	1.78E-13	KKT4(345)-KKT4(350)	8.72E-18
KKT4(521)-KKT4(115)	2.13E-09	KKT4(346)-KKT4(350)	1.97E-09
KKT4(521)-KKT4(206)	6.84E-21	KKT4(346)-KKT4(350)	5.80E-08
KKT4(521)-KKT4(206)	1.34E-20	KKT4(346)-KKT4(350)	8.85E-08
KKT4(521)-KKT4(206)	2.90E-16	KKT4(346)-KKT4(356)	1.01E-09
KKT4(521)-KKT4(218)	2.71E-26	KKT4(350)-KKT4(645)	5.21E-11
KKT4(521)-KKT4(218)	9.06E-26	KKT4(350)-KKT4(645)	1.10E-08
KKT4(521)-KKT4(218)	2.13E-25	KKT4(350)-KKT4(645)	4.77E-08
KKT4(521)-KKT4(218)	2.83E-24	KKT4(521)-KKT4(178)	1.25E-08
KKT4(521)-KKT4(218)	1.39E-23	KKT4(573)-KKT4(132)	6.71E-16
KKT4(521)-KKT4(218)	8.11E-23	KKT4(573)-KKT4(206)	7.13E-13
KKT4(521)-KKT4(218)	5.83E-19	KKT4(573)-KKT4(218)	7.64E-18
KKT4(521)-KKT4(218)	3.72E-14	KKT4(575)-KKT4(132)	5.95E-13
KKT4(521)-KKT4(218)	1.96E-11	KKT4(625)-KKT4(618)	9.50E-12
KKT4(521)-KKT4(510)	7.21E-11	KKT4(625)-KKT4(618)	1.29E-11
KKT4(521)-KKT4(510)	4.00E-09	KKT4(625)-KKT4(618)	2.73E-10
KKT4(521)-KKT4(543)	2.49E-18	KKT4(627)-KKT4(618)	1.07E-12
KKT4(521)-KKT4(543)	1.58E-17	KKT4(627)-KKT4(618)	9.24E-09
KKT4(521)-KKT4(543)	1.34E-15	KKT4(99)-KKT4(132)	9.06E-09

KKT4(521)-KKT4(543)	5.07E-14	KKT4(99)-KKT4(132)	8.44E-08
KKT4(521)-KKT4(543)	6.57E-14		
KKT4(521)-KKT4(543)	1.00E-13		
KKT4(521)-KKT4(543)	1.53E-13		
KKT4(521)-KKT4(543)	2.53E-13		
KKT4(521)-KKT4(543)	2.75E-13		
KKT4(521)-KKT4(543)	3.42E-13		
KKT4(521)-KKT4(543)	3.75E-13		
KKT4(521)-KKT4(543)	4.12E-13		
KKT4(521)-KKT4(543)	4.96E-13		
KKT4(521)-KKT4(543)	5.17E-13		
KKT4(521)-KKT4(543)	7.83E-13		
KKT4(521)-KKT4(543)	2.06E-12		
KKT4(521)-KKT4(543)	3.77E-12		
KKT4(521)-KKT4(543)	4.19E-12		
KKT4(521)-KKT4(543)	9.91E-12		
KKT4(521)-KKT4(543)	1.12E-11		
KKT4(521)-KKT4(543)	1.12E-11		
KKT4(521)-KKT4(543)	1.33E-11		
KKT4(521)-KKT4(543)	1.59E-11		
KKT4(521)-KKT4(543)	2.76E-11		
KKT4(521)-KKT4(543)	3.42E-11		
KKT4(521)-KKT4(543)	9.66E-11		
KKT4(521)-KKT4(543)	1.60E-10		
KKT4(521)-KKT4(543)	2.37E-10		
KKT4(521)-KKT4(543)	2.83E-10		
KKT4(521)-KKT4(543)	2.85E-10		
KKT4(521)-KKT4(543)	6.31E-10		
KKT4(521)-KKT4(543)	7.44E-10		
KKT4(521)-KKT4(543)	1.23E-09		
KKT4(521)-KKT4(543)	1.79E-09		
KKT4(521)-KKT4(543)	3.78E-09		
KKT4(521)-KKT4(543)	6.76E-09		
KKT4(521)-KKT4(543)	8.13E-09		
KKT4(521)-KKT4(543)	9.15E-09		
KKT4(521)-KKT4(543)	1.01E-08		
KKT4(521)-KKT4(543)	1.18E-08		
KKT4(521)-KKT4(543)	1.69E-08		
KKT4(521)-KKT4(543)	2.11E-08		

KKT4(543)-KKT4(350)	4.23E-14		
KKT4(543)-KKT4(350)	6.42E-10		
KKT4(618)-KKT4(132)	2.39E-08		
KKT4(618)-KKT4(218)	1.50E-16		
KKT4(618)-KKT4(218)	5.44E-13		
KKT4(618)-KKT4(349)	3.99E-08		
KKT4(618)-KKT4(350)	5.69E-12		
KKT4(618)-KKT4(350)	8.39E-12		
KKT4(618)-KKT4(356)	2.04E-11		

Table S4. Summary of the DALI results for crystal structures.

Chain	Z	lali	nres	RMSD	%id PD	PDB Description
<i>Trypanosoma cruzi</i> KKT4 ^{117–218}						
6e2j-B	7.8	2.9	103	104	15	KERATIN, TYPE II CYTOSKELETAL 1
4ytd-B	7.7	2.1	99	99	7	PROTEIN BICAUDAL D HOMOLOG 1
2xzt-A	7.7	2.4	103	108	8	IMMUNOGLOBULIN-BINDING PROTEIN EIBD
6e2j-A	7.7	2	103	107	6	KERATIN, TYPE II CYTOSKELETAL 1
6bl7-A	7.6	1.6	99	99	8	PROTEIN STU2
1d7m-A	7.6	2.3	100	101	12	CORTEXILLIN I
6f62-A	7.6	3.1	103	105	16	SYNAPTONEMAL COMPLEX PROTEIN 1
4qkw-A	7.6	2.7	101	101	9	MUSCLE-RELATED COILED-COIL PROTEIN
3u1c-A	7.6	1.9	101	101	12	TROPOMYOSIN ALPHA-1 CHAIN
2wpq-A	7.6	2.6	98	99	1	TRIMERIC AUTOTRANSPORTER ADHESIN FRAGMENT
<i>Leishmania mexicana</i> KKT4 ^{184–284}						
2xzt-A	7.9	3.7	108	108	8	IMMUNOGLOBULIN-BINDING PROTEIN EIBD
6uui-C	7.7	4.4	107	107	16	KERATIN, TYPE II CYTOSKELETAL 1
6e2j-B	7.7	3.8	104	104	8	KERATIN, TYPE II CYTOSKELETAL 1
6f62-A	7.6	2.8	105	105	8	SYNAPTONEMAL COMPLEX PROTEIN 1
6e2j-A	7.6	3	107	107	11	KERATIN, TYPE II CYTOSKELETAL 1
1d7m-A	7.4	3.5	101	101	7	CORTEXILLIN I
6fkh-b	7.4	3	110	161	7	ATP SYNTHASE SUBUNIT A, CHLOROPLASTIC
4ytd-B	7.4	2.7	99	99	7	PROTEIN BICAUDAL D HOMOLOG 1
6mi3-B	7.4	2.7	109	124	9	NF-KB ESSENTIAL MODULATOR,NF-KAPPA-B ESSENTIAL MO
2wpq-A	7.4	4.3	99	99	2	TRIMERIC AUTOTRANSPORTER ADHESIN FRAGMENT
<i>Trypanosoma brucei</i> KKT4 ^{463–645}						
3al2-A	12	3.9	139	220	23	DNA TOPOISOMERASE 2-BINDING PROTEIN 1
1l0b-A	11.9	3.5	142	190	22	BRCA1
2nte-A	11.8	3.4	142	210	17	BRCA1-ASSOCIATED RING DOMAIN PROTEIN 1
3sqd-A	11.4	3.5	137	211	17	PAX-INTERACTING PROTEIN 1
3k05-B	11.3	3.8	133	200	20	MEDIATOR OF DNA DAMAGE CHECKPOINT PROTEIN 1
2xnk-A	11.1	2.4	85	291	22	DNA TOPOISOMERASE 2-BINDING PROTEIN 1
6hm5-A	11.1	4.8	100	251	21	DNA TOPOISOMERASE II BINDING PROTEIN 1

Chain	Z	lali	nres	RMSD	%id PD	PDB Description
4bmc-A	11	2.8	87	183	16	S-M CHECKPOINT CONTROL PROTEIN RAD4
3szm-C	10.9	3.3	131	191	24	MICROCEPHALIN
1kzy-C	10.9	3.2	135	232	16	CELLULAR TUMOR ANTIGEN P53
<i>Trypanosoma brucei</i> KKT23 ^{125–348}						
5h86-A	18.2	2.1	155	167	15	HISTONE ACETYLTRANSFERASE KAT2A
1cm0-A	18.1	1.9	151	162	17	P300/CBP ASSOCIATING FACTOR
6cw2-D	17.2	2.5	161	241	14	HISTONE ACETYLTRANSFERASE GCN5
5wje-A	15.2	2.4	143	159	11	CG8481, ISOFORM B
3d8p-B	14.5	2.5	145	160	14	ACETYLTRANSFERASE OF GNAT FAMILY
2cnm-A	14.3	1.9	131	151	11	MODIFICATION OF 30S RIBOSOMAL SUBUNIT PROTEIN S18
4ua3-B	14.2	2.4	145	193	7	UNCHARACTERIZED N-ACETYLTRANSFERASE C825.04C
3efa-A	13.9	2.3	134	146	16	PUTATIVE ACETYLTRANSFERASE
2r7h-B	13.9	2.3	139	159	13	PUTATIVE D-ALANINE N-ACETYLTRANSFERASE OF GNAT FA
2q7b-A	13.9	2.6	145	164	12	ACETYLTRANSFERASE, GNAT FAMILY

Bibliography

Abad, M. A., Medina, B., Santamaria, A., Zou, J., Plasberg-Hill, C., Madhumalar, A., Jayachandran, U., Redli, P. M., Rappsilber, J., Nigg, E. A., et al. 2014. Structural basis for microtubule recognition by the human kinetochore Ska complex. *Nat Commun*, 5, 2964, 10.1038/ncomms3964.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., et al. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66, 213-221, 10.1107/S0907444909052925.

Ait-Si-Ali, S., Ramirez, S., Barre, F. X., Dkhissi, F., Magnaghi-Jaulin, L., Girault, J. A., Robin, P., Knibiehler, M., Pritchard, L. L., Ducommun, B., et al. 1998. Histone acetyltransferase activity of CBP is controlled by cycle-dependent kinases and oncoprotein E1A. *Nature*, 396, 184-6, 10.1038/24190.

Akiyoshi, B. & Gull, K. 2014. Discovery of unconventional kinetochores in kinetoplastids. *Cell*, 156, 1247-1258, 10.1016/j.cell.2014.01.049.

Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., et al. 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.*, 38, 457-462, 10.1093/nar/gkp851.

Bajaj, R., Bollen, M., Peti, W. & Page, R. 2018. KNL1 Binding to PP1 and Microtubules Is Mutually Exclusive. *Structure*, 26, 1327-1336 e4, 10.1016/j.str.2018.06.013.

Bannister, A. J. & Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell Res.*, 21, 381-95, 10.1038/cr.2011.22.

Beger, R. D. & Bolton, P. H. 1997. Protein phi and psi dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. *J Biomol NMR*, 10, 129-142, 10.1023/a:1018302105638.

Berndsen, C. E. & Denu, J. M. 2005. Assays for mechanistic investigations of protein/histone acetyltransferases. *Methods*, 36, 321-31, 10.1016/j.ymeth.2005.03.002.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science*, 309, 416-422, 10.1126/science.1112642.

Biggins, S. 2013. The composition, functions, and regulation of the budding yeast kinetochore. *Genetics*, 194, 817-846, 10.1534/genetics.112.145276.

- Brun, R. & Schonenberger 1979. Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication. *Acta Trop.*, 36, 289-92.
- BrüNger, A. T. 1992. X-PLOR, Version 3.1 : a system for X-ray crystallography and NMR, New Haven, Yale University Press.
- Brusini, L., D'archivio, S., Mcdonald, J. & Wickstead, B. 2019. Ndc80/Nuf2-like protein KKIP1 connects a stable kinetoplastid outer kinetochore complex to the inner kinetochore and responds to metaphase tension. *bioRxiv*, 764829.
- Bu, P., Evrard, Y. A., Lozano, G. & Dent, S. Y. 2007. Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. *Mol. Cell. Biol.*, 27, 3405-16, 10.1128/MCB.00066-07.
- Butenko, A., Opperdoes, F. R., Flegontova, O., Horak, A., Hampl, V., Keeling, P., Gawryluk, R. M. R., Tikhonenkov, D., Flegontov, P. & Lukes, J. 2020. Evolution of metabolic capabilities and molecular features of diplomonads, kinetoplastids, and euglenids. *BMC Biol.*, 18, 23, 10.1186/s12915-020-0754-1.
- Cavalier-Smith, T. 2010. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biology Letters*, 6, 342-345, 10.1098/rsbl.2009.0948.
- Chan, Y. W., Jeyaprasath, A. A., Nigg, E. A. & Santamaria, A. 2012. Aurora B controls kinetochore-microtubule attachments by inhibiting Ska complex-KMN network interaction. *J Cell Biol*, 196, 563-571, 10.1083/jcb.201109001.
- Cheerambathur, D. K., Gassmann, R., Cook, B., Oegema, K. & Desai, A. 2013. Crosstalk between microtubule attachment complexes ensures accurate chromosome segregation. *Science*, 342, 1239-42, 10.1126/science.1246232.
- Cheeseman, I. M. 2014. The kinetochore. *Cold Spring Harb Perspect Biol*, 6, a015826, 10.1101/cshperspect.a015826.
- Cheeseman, I. M., Anderson, S., Jwa, M., Green, E. M., Kang, J., Yates, J. R., 3rd, Chan, C. S., Drubin, D. G. & Barnes, G. 2002. Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell*, 111, 163-172, 10.1016/S0092-8674(02)00973-x.
- Cheeseman, I. M., Chappie, J. S., Wilson-Kubalek, E. M. & Desai, A. 2006. The conserved KMN network constitutes the core microtubule-binding site of the kinetochore. *Cell*, 127, 983-997, 10.1016/j.cell.2006.09.039.
- Cheeseman, I. M. & Desai, A. 2008. Molecular architecture of the kinetochore-microtubule interface. *Nat. Rev. Mol. Cell Biol.*, 9, 33-46, 10.1038/nrm2310.
- Chen, Z. L., Meng, J. M., Cao, Y., Yin, J. L., Fang, R. Q., Fan, S. B., Liu, C., Zeng, W. F., Ding, Y. H., Tan, D., et al. 2019. A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun*, 10, 3404, 10.1038/s41467-019-11337-z.

- Ciferri, C., Pasqualato, S., Screpanti, E., Varetto, G., Santaguida, S., Dos Reis, G., Maiolica, A., Polka, J., De Luca, J. G., De Wulf, P., et al. 2008. Implications for Kinetochore-Microtubule Attachment from the Structure of an Engineered Ndc80 Complex. *Cell*, 133, 427-439, 10.1016/j.cell.2008.03.020.
- Clarke, L. 1998. Centromeres: proteins, protein complexes, and repeated domains at centromeres of simple eukaryotes. *Curr Opin Genet Dev*, 8, 212-8, 10.1016/s0959-437x(98)80143-3.
- Combe, C. W., Fischer, L. & Rappsilber, J. 2015. xiNET: cross-link network maps with residue resolution. *Mol. Cell. Proteomics*, 14, 1137-1147, 10.1074/mcp.O114.042259.
- Copenhaver, G. P., Doelling, J. H., Gens, S. & Pikaard, C. S. 1995. Use of RFLPs larger than 100 kbp to map the position and internal organization of the nucleolus organizer region on chromosome 2 in *Arabidopsis thaliana*. *Plant J*, 7, 273-86, 10.1046/j.1365-313x.1995.7020273.x.
- Cornilescu, G., Delaglio, F. & Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, 13, 289-302, 10.1023/a:1008392405740.
- Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, 62, 1002-1011, 10.1107/S0907444906022116.
- D'archivio, S. & Wickstead, B. 2017. Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J. Cell Biol.*, 216, 379-391, 10.1083/jcb.201608043.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*, 6, 277-93, 10.1007/bf00197809.
- Delano, W. L. 2002. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 82-92.
- Deluca, J. G., Gall, W. E., Ciferri, C., Cimini, D., Musacchio, A. & Salmon, E. D. 2006. Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. *Cell*, 127, 969-82, 10.1016/j.cell.2006.09.047.
- Deluca, K. F., Lens, S. M. & Deluca, J. G. 2011. Temporal changes in Hec1 phosphorylation control kinetochore-microtubule attachment stability during mitosis. *J. Cell Sci.*, 124, 622-34, 10.1242/jcs.072629.
- Dyda, F., Klein, D. C. & Hickman, A. B. 2000. GCN5-related N-acetyltransferases: a structural overview. *Annu. Rev. Biophys. Biomol. Struct.*, 29, 81-103, 10.1146/annurev.biophys.29.1.81.

- Earnshaw, W. C. & Rothfield, N. 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, 91, 313-21, 10.1007/BF00328227.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 10.1107/S0907444910007493/10.1107/S0907444910007493.
- Fenn, K. & Matthews, K. R. 2007. The cell biology of *Trypanosoma brucei* differentiation. *Curr. Opin. Microbiol.*, 10, 539-46, 10.1016/j.mib.2007.09.014.
- Foley, E. A. & Kapoor, T. M. 2013. Microtubule attachment and spindle assembly checkpoint signalling at the kinetochore. *Nat. Rev. Mol. Cell Biol.*, 14, 25-37, 10.1038/nrm3494.
- Fournier, M. & Tora, L. 2017. KAT2-mediated PLK4 acetylation contributes to genomic stability by preserving centrosome number. *Mol Cell Oncol*, 4, e1270391, 10.1080/23723556.2016.1270391.
- Friese, A., Faesen, A. C., Huis in 't Veld, P. J., Fischbock, J., Prumbaum, D., Petrovic, A., Raunser, S., Herzog, F. & Musacchio, A. 2016. Molecular requirements for the inter-subunit interaction and kinetochore recruitment of SKAP and Astrin. *Nat Commun*, 7, 11407, 10.1038/ncomms11407.
- Fukagawa, T. & Earnshaw, W. C. 2014. The centromere: Chromatin foundation for the kinetochore machinery.
- Guimaraes, G. J., Dong, Y., McEwen, B. F. & Deluca, J. G. 2008. Kinetochore-microtubule attachment relies on the disordered N-terminal tail domain of Hec1. *Curr. Biol.*, 18, 1778-1784, 10.1016/j.cub.2008.08.012.
- Hayes D, L. T. P. J. 1995. Program Sednterp: sedimentation interpretation program. *Alliance Protein Laboratories, Thousand Oaks, CA*.
- Hegemann, J. H. & Fleig, U. N. 1993. The centromere of budding yeast. *Bioessays*, 15, 451-60, 10.1002/bies.950150704.
- Holm, L. 2019. Benchmarking fold detection by DaliLite v.5. *Bioinformatics (Oxford, England)*, 10.1093/bioinformatics/btz536/10.1093/bioinformatics/btz536.
- Ishii, M. & Akiyoshi, B. 2020. Characterization of unconventional kinetochore kinases KKT10 and KKT19 in *Trypanosoma brucei*. *J. Cell Sci.*, 13310.1242/jcs.240978.
- Jansen, L. E., Black, B. E., Foltz, D. R. & Cleveland, D. W. 2007. Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol*, 176, 795-805, 10.1083/jcb.200701066.
- Janzen, C. J., Fernandez, J. P., Deng, H., Diaz, R., Hake, S. B. & Cross, G. A. 2006. Unusual histone modifications in *Trypanosoma brucei*. *FEBS Lett.*, 580, 2306-10, 10.1016/j.febslet.2006.03.044.

- Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., et al. 2018. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.*, 27, 112-128, 10.1002/pro.3280.
- Katoh, K., Rozewicki, J. & Yamada, K. D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*, 20, 1160-1166, 10.1093/bib/bbx108.
- Kay, L. E., Torchia, D. A. & Bax, A. 1989. Backbone Dynamics of Proteins As Studied by ¹⁵N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry*, 28(23), 8972-8979, 10.1021/bi00449a003.
- Kelly, S., Reed, J., Kramer, S., Ellis, L., Webb, H., Sunter, J., Salje, J., Marinsek, N., Gull, K., Wickstead, B., et al. 2007. Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol. Biochem. Parasitol.*, 154, 103-9, 10.1016/j.molbiopara.2007.03.012.
- Kemmler, S., Stach, M., Knapp, M., Ortiz, J., Pfannstiel, J., Ruppert, T. & Lechner, J. 2009. Mimicking Ndc80 phosphorylation triggers spindle assembly checkpoint signalling. *EMBO J.*, 28, 1099-110, 10.1038/emboj.2009.62.
- Kim, Y., Tanner, K. G. & Denu, J. M. 2000. A continuous, nonradioactive assay for histone acetyltransferases. *Anal. Biochem.*, 280, 308-14, 10.1006/abio.2000.4546.
- Kraus, A. J., Vanselow, J. T., Lamer, S., Brink, B. G., Schlosser, A. & Siegel, T. N. 2020. Distinct roles for H4 and H2A.Z acetylation in RNA transcription in African trypanosomes. *Nat Commun*, 11, 1498, 10.1038/s41467-020-15274-0.
- Krissinel, E. & Henrick, K. 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, 372, 774-97, 10.1016/j.jmb.2007.05.022.
- Kuo, M. H., Zhou, J., Jambeck, P., Churchill, M. E. & Allis, C. D. 1998. Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes Dev.*, 12, 627-39, 10.1101/gad.12.5.627.
- Lampert, F., Mieck, C., Alushin, G. M., Nogales, E. & Westermann, S. 2013. Molecular requirements for the formation of a kinetochore-microtubule interface by Dam1 and Ndc80 complexes. *J Cell Biol*, 200, 21-30, 10.1083/jcb.201210091.
- Lampson, M. A. & Grishchuk, E. L. 2017. Mechanisms to Avoid and Correct Erroneous Kinetochore-Microtubule Attachments. *Biology (Basel)*, 610.3390/biology6010001.
- Leitner, A., Faini, M., Stengel, F. & Aebersold, R. 2016. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem. Sci.*, 41, 20-32, 10.1016/j.tibs.2015.10.008.
- Lescop, E., Schanda, P. & Brutscher, B. 2007. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J Magn Reson*, 187, 163-9, 10.1016/j.jmr.2007.04.002.

- Leung, C. C. & Glover, J. N. 2011. BRCT domains: easy as one, two, three. *Cell Cycle*, 10, 2461-2470, 10.4161/cc.10.15.16312.
- Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. 2003. Protein disorder prediction: implications for structural proteomics. *Structure*, 11, 1453-1459, 10.1016/j.str.2003.10.002.
- Liu, S. T., Rattner, J. B., Jablonski, S. A. & Yen, T. J. 2006. Mapping the assembly pathways that specify formation of the trilaminar kinetochore plates in human cells. *J Cell Biol*, 175, 41-53, 10.1083/jcb.200606020.
- Llauro, A., Hayashi, H., Bailey, M. E., Wilson, A., Ludzia, P., Asbury, C. L. & Akiyoshi, B. 2018. The kinetoplastid kinetochore protein KKT4 is an unconventional microtubule tip-coupling protein. *J Cell Biol*, 217, 3886-3900, 10.1083/jcb.201711181.
- Lowell, J. E., Kaiser, F., Janzen, C. J. & Cross, G. A. 2005. Histone H2AZ dimerizes with a novel variant H2B and is enriched at repetitive DNA in *Trypanosoma brucei*. *J. Cell Sci.*, 118, 5721-30, 10.1242/jcs.02688.
- Lupas, A., Van Dyke, M. & Stock, J. 1991. Predicting coiled coils from protein sequences. *Science*, 252, 1162-1164, 10.1126/science.252.5009.1162.
- Maiato, H., Deluca, J., Salmon, E. D. & Earnshaw, W. C. 2004. The dynamic kinetochore-microtubule interface. *Journal of Cell Science*, 117, 5461-5477, 10.1242/jcs.01536.
- Mandava, V., Fernandez, J. P., Deng, H., Janzen, C. J., Hake, S. B. & Cross, G. A. 2007. Histone modifications in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.*, 156, 41-50, 10.1016/j.molbiopara.2007.07.005.
- Manke, I. A., Lowery, D. M., Nguyen, A. & Yaffe, M. B. 2003. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*, 302, 636-639, 10.1126/science.1088877.
- Marciano, G., Ishii, M., Nerusheva, O. O. & Akiyoshi, B. 2021. Kinetoplastid kinetochore proteins KKT2 and KKT3 have unique centromere localization domains. *J Cell Biol*, 22010.1083/jcb.202101022.
- Marmorstein, R. & Roth, S. Y. 2001. Histone acetyltransferases: function, structure, and catalysis. *Curr Opin Genet Dev*, 11, 155-61, 10.1016/s0959-437x(00)00173-8.
- Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. 2006. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci*, 15, 2795-2804, 10.1110/ps.062465306.
- Marshall, O. J., Chueh, A. C., Wong, L. H. & Choo, K. H. 2008. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet*, 82, 261-82, 10.1016/j.ajhg.2007.11.009.

- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. & Okazaki, T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *Journal of Cell Biology*, 109, 1963-1973, 10.1083/jcb.109.5.1963.
- Mattson, G., Conklin, E., Desai, S., Nielander, G., Savage, M. D. & Morgensen, S. 1993. A practical approach to crosslinking. *Mol. Biol. Rep.*, 17, 167-183, 10.1007/BF00986726.
- Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography*, 10.1107/S0021889807021206/10.1107/S0021889807021206.
- Mcguffin, L. J., Bryson, K. & Jones, D. T. 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 10.1093/bioinformatics/16.4.404/10.1093/bioinformatics/16.4.404.
- Mcintosh, J. R. 2016. Mitosis. *Cold Spring Harb Perspect Biol*, 8,10.1101/cshperspect.a023218.
- Mckinley, K. L. & Cheeseman, I. M. 2016. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.*, 17, 16-29, 10.1038/nrm.2015.5.
- Miller, S. A., Johnson, M. L. & Stukenberg, P. T. 2008. Kinetochore attachments require an interaction between unstructured tails on microtubules and Ndc80(Hec1). *Curr. Biol.*, 18, 1785-91, 10.1016/j.cub.2008.11.007.
- Mizzen, C. A., Brownell, J. E., Cook, R. G. & Allis, C. D. 1999. Histone acetyltransferases: preparation of substrates and assay procedures. *Methods Enzymol.*, 304, 675-96, 10.1016/s0076-6879(99)04041-0.
- Mo, F., Zhuang, X., Liu, X., Yao, P. Y., Qin, B., Su, Z., Zang, J., Wang, Z., Zhang, J., Dou, Z., et al. 2016. Acetylation of Aurora B by TIP60 ensures accurate chromosomal segregation. *Nat. Chem. Biol.*, 12, 226-32, 10.1038/nchembio.2017.
- Monks, S. 2002. SigmaPlot 8.0. *Biotech Software & Internet Report*, 3, 141-145, 10.1089/152791602321105816.
- Moroi, Y., Peebles, C., Fritzler, M. J., Steigerwald, J. & Tan, E. M. 1980. Autoantibody to centromere (kinetochore) in scleroderma sera. *Proc Natl Acad Sci U S A*, 77, 1627-31, 10.1073/pnas.77.3.1627.
- Musacchio, A. & Desai, A. 2017. A Molecular View of Kinetochore Assembly and Function. *Biology (Basel)*, 6,10.3390/biology6010005.
- Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P. B., Kim, M., Jones, K. M., Henikoff, S., Buell, C. R. & Jiang, J. 2004. Sequencing of a rice centromere uncovers active genes. *Nat Genet*, 36, 138-45, 10.1038/ng1289.

- Nerusheva, O. O. & Akiyoshi, B. 2016. Divergent polo box domains underpin the unique kinetoplastid kinetochore. *Open Biol*, 6, 10.1098/rsob.150206.
- Nerusheva, O. O., Ludzia, P. & Akiyoshi, B. 2019. Identification of four unconventional kinetoplastid kinetochore proteins KKT22-25 in *Trypanosoma brucei*. *Open Biol*, 9, 190236, 10.1098/rsob.190236.
- Obado, S. O., Bot, C., Nilsson, D., Andersson, B. & Kelly, J. M. 2007. Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biol*, 8, 37, 10.1186/gb-2007-8-3-r37.
- Ohzeki, J., Shono, N., Otake, K., Martins, N. M., Kugou, K., Kimura, H., Nagase, T., Larionov, V., Earnshaw, W. C. & Masumoto, H. 2016. KAT7/HBO1/MYST2 Regulates CENP-A Chromatin Assembly by Antagonizing Suv39h1-Mediated Centromere Inactivation. *Dev. Cell*, 37, 413-27, 10.1016/j.devcel.2016.05.006.
- Orpinell, M., Fournier, M., Riss, A., Nagy, Z., Krebs, A. R., Frontini, M. & Tora, L. 2010. The ATAC acetyl transferase complex controls mitotic progression by targeting non-histone substrates. *EMBO J.*, 29, 2381-94, 10.1038/emboj.2010.125.
- Owen, B. & McMurray, C. 2009. Rapid method for measuring DNA binding to protein using fluorescence anisotropy.
- Palmer, D. K., O'day, K., Trong, H. L., Charbonneau, H. & Margolis, R. L. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 3734-8, 10.1073/pnas.88.9.3734.
- Pardue, M. L. & Gall, J. G. 1970. Chromosomal localization of mouse satellite DNA. *Science*, 168, 1356-8, 10.1126/science.168.3937.1356.
- Petrovic, A., Keller, J., Liu, Y., Overlack, K., John, J., Dimitrova, Y. N., Jenni, S., Van Gerwen, S., Stege, P., Wohlgemuth, S., et al. 2016. Structure of the MIS12 Complex and Molecular Basis of Its Interaction with CENP-C at Human Kinetochores. *Cell*, 167, 1028-1040 e15, 10.1016/j.cell.2016.10.005.
- Petrovic, A., Pasqualato, S., Dube, P., Krenn, V., Santaguida, S., Cittaro, D., Monzani, S., Massimiliano, L., Keller, J., Tarricone, A., et al. 2010. The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J Cell Biol*, 190, 835-52, 10.1083/jcb.201002070.
- Poon, S. K., Peacock, L., Gibson, W., Gull, K. & Kelly, S. 2012. A modular and optimized single marker system for generating *Trypanosoma brucei* cell lines expressing T7 RNA polymerase and the tetracycline repressor. *Open Biol*, 2, 110037, 10.1098/rsob.110037.
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R. & Finn, R. D. 2018. HMMER web server: 2018 update. *Nucleic Acids Res.*, 46, W200-W204, 10.1093/nar/gky448.
- Powers, A. F., Franck, A. D., Gestaut, D. R., Cooper, J., Gracyzk, B., Wei, R. R., Wordeman, L., Davis, T. N. & Asbury, C. L. 2009. The Ndc80 kinetochore complex

forms load-bearing attachments to dynamic microtubule tips via biased diffusion. *Cell*, 136, 865-75, 10.1016/j.cell.2008.12.045.

Przewloka, M. R., Venkei, Z., Bolanos-Garcia, V. M., Debski, J., Dadlez, M. & Glover, D. M. 2011. CENP-C is a structural platform for kinetochore assembly. *Curr. Biol.*, 21, 399-405, 10.1016/j.cub.2011.02.005.

Redli, P. M., Gasic, I., Meraldi, P., Nigg, E. A. & Santamaria, A. 2016. The Ska complex promotes Aurora B activity to ensure chromosome biorientation. *J Cell Biol*, 215, 77-93, 10.1083/jcb.201603019.

Regnier, V., Vagnarelli, P., Fukagawa, T., Zerjal, T., Burns, E., Trouche, D., Earnshaw, W. & Brown, W. 2005. CENP-A is required for accurate chromosome segregation and sustained kinetochore association of BubR1. *Mol. Cell. Biol.*, 25, 3967-81, 10.1128/MCB.25.10.3967-3981.2005.

Rodríguez, D., Sammito, M., Meindl, K., De Ilarduya, I. M., Potratz, M., Sheldrick, G. M. & Usón, I. 2012. Practical structure solution with ARCIMBOLDO. *Acta Crystallogr D*, 68, 336-343, 10.1107/S0907444911056071.

Rodríguez, D. D., Grosse, C., Himmel, S., González, C., De Ilarduya, I. M., Becker, S., Sheldrick, G. M. & Usón, I. 2009. Crystallographic ab initio protein structure solution below atomic resolution. *Nat. Methods*, 10.1038/nmeth.1365/10.1038/nmeth.1365.

Rückert, M. & Otting, G. 2000. Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.*, 122, 7793-7797.

Saitoh, H., Tomkiel, J., Cooke, C. A., Ratrie, H., Maurer, M., Rothfield, N. F. & Earnshaw, W. C. 1992. CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell*, 70, 115-125, 10.1016/0092-8674(92)90538-N.

Saldivia, M., Fang, E., Ma, X., Myburgh, E., Carnielli, J. B. T., Bower-Lepts, C., Brown, E., Ritchie, R., Lakshminarayana, S. B., Chen, Y. L., et al. 2020. Targeting the trypanosome kinetochore with CLK1 protein kinase inhibitors. *Nat Microbiol*, 5, 1207-1216, 10.1038/s41564-020-0745-6.

Schmidt, J. C., Arthanari, H., Boeszoermenyi, A., Dashkevich, N. M., Wilson-Kubalek, E. M., Monnier, N., Markus, M., Oberer, M., Milligan, R. A., Bathe, M., et al. 2012. The kinetochore-bound Ska1 complex tracks depolymerizing microtubules and binds to curved protofilaments. *Dev. Cell*, 23, 968-980, 10.1016/j.devcel.2012.09.012.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, 9, 671-675, 10.1038/nmeth.2089.

Schnell, J. R., Zhou, G. P., Zweckstetter, M., Rigby, A. C. & Chou, J. J. 2005. Rapid and accurate structure determination of coiled-coil domains using NMR dipolar

couplings: application to cGMP-dependent protein kinase α . *Protein Sci.*, 14, 2421-2428, 10.1110/ps.051528905.

Schuck, P., Gillis, R. B., Besong, T. M. D., Almutairi, F., Adams, G. G., Rowe, A. J. & Harding, S. E. 2013. SEDFIT-MSTAR: Molecular weight and molecular weight distribution analysis of polymers by sedimentation equilibrium in the ultracentrifuge. *The Analyst*, 139, 79-92, 10.1039/c3an01507f.

Schulte-Herbruggen, T. & Sorensen, O. W. 2000. Clean TROSY: compensation for relaxation-induced artifacts. *J Magn Reson*, 144, 123-8, 10.1006/jmre.2000.2020.

Screpanti, E., De Antoni, A., Alushin, G. M., Petrovic, A., Melis, T., Nogales, E. & Musacchio, A. 2011. Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr. Biol.*, 21, 391-8, 10.1016/j.cub.2010.12.039.

Shen, Y. & Bax, A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, 56(3), 227-241, 10.1007/s10858-013-9741-y.

Shiozaki, E. N., Gu, L., Yan, N. & Shi, Y. 2004. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. *Mol. Cell*, 14, 405-412, 10.1016/s1097-2765(04)00238-2.

Skubak, P. & Pannu, N. S. 2013. Automatic protein structure solution from weak X-ray data. *Nat Commun*, 4, 2777, 10.1038/ncomms3777.

Spera, S. & Bax, A. 1991. Empirical correlation between protein backbone conformation and C.alpha. and C.beta. ^{13}C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society*, 113, 5490-5492, 10.1021/ja00014a071.

Strelkov, S. V. & Burkhard, P. 2002. Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol*, 137, 54-64, 10.1006/jsbi.2002.4454.

Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*, 103, 8060-5, 10.1073/pnas.0602606103.

Sugimoto, K., Yata, H., Muro, Y. & Himeno, M. 1994. Human Centromere Protein C (CENP-C) Is a DNA-Binding Protein Which Possesses a Novel DNA-Binding Motif1. *The Journal of Biochemistry*, 116, 877-881, 10.1093/oxfordjournals.jbchem.a124610.

Sullivan, K. F., Hechenberger, M. & Masri, K. 2013. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *Cancer Discovery*, 3, 581-592, 10.1083/jcb.127.3.581.

Sun, X., Le, H. D., Wahlstrom, J. M. & Karpen, G. H. 2003. Sequence analysis of a functional *Drosophila* centromere. *Genome Res*, 13, 182-94, 10.1101/gr.681703.

- Talbert, P. B. & Henikoff, S. 2018. Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. *Trends Genet.*, 34, 587-599, 10.1016/j.tig.2018.05.001.
- Tan, E. M., Rodnan, G. P., Garcia, I., Moroi, Y., Fritzler, M. J. & Peebles, C. 1980. Diversity of antinuclear antibodies in progressive systemic sclerosis. Anti-centromere antibody and its relationship to CREST syndrome. *Arthritis Rheum.*, 23, 617-25, 10.1002/art.1780230602.
- Tanner, K. G., Trievel, R. C., Kuo, M. H., Howard, R. M., Berger, S. L., Allis, C. D., Marmorstein, R. & Denu, J. M. 1999. Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J. Biol. Chem.*, 274, 18157-60, 10.1074/jbc.274.26.18157.
- Tien, J. F., Umbreit, N. T., Gestaut, D. R., Franck, A. D., Cooper, J., Wordeman, L., Gonen, T., Asbury, C. L. & Davis, T. N. 2010. Cooperation of the Dam1 and Ndc80 kinetochore complexes enhances microtubule coupling and is regulated by aurora B. *J Cell Biol*, 189, 713-723, 10.1083/jcb.200910142.
- Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L. & Marmorstein, R. 1999. Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc Natl Acad Sci U S A*, 96, 8931-6, 10.1073/pnas.96.16.8931.
- Tromer, E. C., Wemyss, T. A., Ludzia, P., Waller, R. F. & Akiyoshi, B. 2021. Repurposing of synaptonemal complex proteins for kinetochores in Kinetoplastida. *Open Biol*, 11, 210049, 10.1098/rsob.210049.
- Tschudi, C., Shi, H., Franklin, J. B. & Ullu, E. 2012. Small interfering RNA-producing loci in the ancient parasitic eukaryote *Trypanosoma brucei*. *BMC Genomics*, 13, 427, 10.1186/1471-2164-13-427.
- Tu, X., Kumar, P., Li, Z. & Wang, C. C. 2006. An aurora kinase homologue is involved in regulating both mitosis and cytokinesis in *Trypanosoma brucei*. *J. Biol. Chem.*, 281, 9677-9687, 10.1074/jbc.M511504200.
- Tugarinov, V., Kanelis, V. & Kay, L. E. 2006. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nature Protocols*, 10.1038/nprot.2006.10110.1038/nprot.2006.101.
- Uniprot, C. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, 47, D506-D515, 10.1093/nar/gky1049.
- Valverde, R., Ingram, J. & Harrison, S. C. 2016. Conserved Tetramer Junction in the Kinetochore Ndc80 Complex. *Cell Rep*, 17, 1915-1922, 10.1016/j.celrep.2016.10.065.
- Van Hooff, J. J., Tromer, E., Van Wijk, L. M., Snel, B. & Kops, G. J. 2017. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep*, 18, 1559-1571, 10.15252/embr.201744102.

- Vickerman, K. 1965. Polymorphism and mitochondrial activity in sleeping sickness trypanosomes. *Nature*, 208, 762-6, 10.1038/208762a0.
- Volkov, V. A. 2020. Microtubules pull the strings: disordered sequences as efficient couplers of microtubule-generated force. *Essays Biochem*, 10.1042/EBC2019007810.1042/EBC20190078.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, 59, 687-96, 10.1002/prot.20449.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25, 1189-1191, 10.1093/bioinformatics/btp033.
- Webb, B. & Sali, A. 2016. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics*, 54, 5.6.1-5.6.30, 10.1002/cpbi.3.
- Wei, R. R., Al-Bassam, J. & Harrison, S. C. 2007. The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nat. Struct. Mol. Biol.*, 14, 54-59, 10.1038/nsmb1186.
- Wen, J., Arakawa, T. & Philo, J. S. 1996. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal. Biochem.*, 240, 155-166, 10.1006/abio.1996.0345.
- Westermann, S., Wang, H. W., Avila-Sakar, A., Drubin, D. G., Nogales, E. & Barnes, G. 2006. The Dam1 kinetochore ring complex moves processively on depolymerizing microtubule ends. *Nature*, 440, 565-9, 10.1038/nature04409.
- Williams, R. S., Lee, M. S., Hau, D. D. & Glover, J. N. 2004. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nat. Struct. Mol. Biol.*, 11, 519-525, 10.1038/nsmb776.
- Wishart, D. S., Sykes, B. D. & Richards, F. M. 1991. Simple techniques for the quantification of protein secondary structure by ¹H NMR spectroscopy. *FEBS Lett*, 293, 72-80, 10.1016/0014-5793(91)81155-2.
- Yu, X., Chini, C. C. S., He, M., Mer, G. & Chen, J. 2003. The BRCT Domain Is a Phospho-Protein Binding Domain. *Science*, 302, 639-642, 10.1126/science.1088753.
- Zhang, X., Moréra, S., Bates, P. A., Whitehead, P. C., Coffey, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J. E., Lindahl, T. & Freemont, P. S. 1998. Structure of an XRCC1 BRCT domain: A new protein-protein interaction module. *EMBO Journal*, 17, 6404-6411, 10.1093/emboj/17.21.6404.
- Zhu, G., Xia, Y., Nicholson, L. K. & Sze, K. H. 2000. Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson*, 143, 423-6, 10.1006/jmre.2000.2022.

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A. N. & Alva, V. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.*, 430, 2237-2243, 10.1016/j.jmb.2017.12.007.