

# Structure and Function Analysis of Kinetoplastid RNA Editing Ligase

---

**Daniel Moses**

Institute of Parasitology

Faculty of Agricultural and Environmental Sciences

McGill University

Montreal, Quebec

December 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree  
of Master of Science

© Daniel Moses (2022)

## Table of Contents

<b>Abstract</b>	<b>4</b>
<b>Résumé</b>	<b>5</b>
<b>Acknowledgements</b>	<b>6</b>
<b>Contribution to original knowledge</b>	<b>7</b>
<b>Contribution of authors</b>	<b>8</b>
<b>Abbreviations</b>	<b>9</b>
<b>List of Figures</b>	<b>11</b>
<b>List of Tables</b>	<b>13</b>
<b>Introduction</b>	<b>14</b>
<b>Comprehensive literature review</b>	<b>16</b>
<i>Trypanosomes and their diseases</i>	16
<i>The kinetoplast and post-transcriptional RNA editing</i>	19
<i>The RNA editing core complex (RECC)</i>	20
<i>Kinetoplastid RNA Editing Ligases</i>	21
<i>Kinetoplastid RNA Editing Proteins A<sub>1-6</sub></i>	23
<b>Methodology</b>	<b>25</b>
<i>Sequence alignment and structure predictions</i>	25
<i>Expression and purification of recombinant KREL1 and KREPA2</i>	29
<i>KREPA2 pulldown assay</i>	30
<i>Three-step ligation assay</i>	31
<i>Step 1, ligase adenylylation assay</i>	31
<i>Step 2, RNA adenylylation assay</i>	32
<i>Step 3, pre-adenylylated ligation assay</i>	32
<b>Research findings</b>	<b>34</b>

<i>Structural analysis of KREL1 and KREPA2</i>	34
<i>KREL1 and KREPA2 protein-protein interaction studies</i>	38
<i>Functional role of KREL1 domains in ligation</i>	41
<i>Mutational analysis of KREL1 NTD and VLR</i>	42
<i>Location motif VI on KREL1's CTD</i>	45
<i>Effect of KREPA2 interaction on KREL1 ligation activity</i>	47
<b>Comprehensive discussion</b>	<b>50</b>
<b>Conclusion and summary</b>	<b>56</b>
<b>Future directions</b>	<b>57</b>
<b>Appendix</b>	<b>58</b>
<b>References</b>	<b>64</b>

## Abstract

Parasitic protozoans of the Trypanosoma and Leishmania species have a uniquely organized mitochondrial genome, the kinetoplast. Most kinetoplast-transcribed mRNAs are cryptic and encode multiple subunits for the electron transport chain following maturation through a uridine insertion/deletion process called RNA editing. This process is achieved through an enzyme cascade by an RNA editing catalytic complex (RECC), where the final ligation step is catalyzed by the kinetoplastid RNA editing ligases, KREL1 and KREL2. While the N-terminal domain (NTD) of these proteins is highly conserved with other DNA ligases and mRNA capping enzymes, with five recognizable motifs, the functional role of their diverged C-terminal domain (CTD) has remained elusive. This work investigates the role of KREL1's CTD in protein interaction and ligation activity. We assayed recombinant KREL1 *in vitro* to unveil critical residues from its CTD. The data show that the  $\alpha$ -helix (H)3 of KREL1 CTD interacts with the  $\alpha$ H1 of its editosome protein partner KREPA2. Intriguingly, the OB-fold domain and the zinc fingers on KREPA2 do not appear to influence the RNA ligation activity of KREL1. Moreover, a specific KWKE motif on the  $\alpha$ H4 of KREL1 CTD is found to be implicated in ligase auto-adenylation analogous to motif VI in DNA ligases. These findings detail the mode of action of KREL1 CTD in ligation and the functional role of KREPA2's interaction. We further propose a model for KREL1's "open-close" conformational change to provide motif VI to the adenylation domain for ATP hydrolysis that KREPA2 mediates. These findings of the KREL1/KREPA2 RNA ligation mechanism are unique to kinetoplastids.

## Résumé

Les protozoaires parasites des espèces *Trypanosoma* et *Leishmania* possèdent un génome mitochondrial organisé de manière unique, le kinétoplaste. La plupart des ARNm transcrits dans le kinétoplaste sont cryptiques et codent pour de multiples sous-unités de la chaîne de transport des électrons après maturation par un processus d'insertion/délétion d'uridine appelé édition de l'ARN. Ce processus est réalisé par une cascade d'enzymes par un complexe catalytique d'édition de l'ARN (RECC), où l'étape finale de ligature est catalysée par les ligases d'édition de l'ARN des kinétoplastides, KREL1 et KREL2. Alors que le domaine N-terminal (NTD) de ces protéines est hautement conservé avec d'autres ADN ligases et enzymes de coiffage d'ARNm, avec cinq motifs reconnaissables, le rôle fonctionnel de leur domaine C-terminal (CTD) divergent est resté insaisissable. Ce travail étudie le rôle du CTD de KREL1 dans l'interaction des protéines et l'activité de ligature. Nous avons testé KREL1 recombinant *in vitro* pour dévoiler les résidus critiques de son CTD. Les données montrent que l'hélice  $\alpha$  (H)3 de la CTD de KREL1 interagit avec l' $\alpha$ H1 de son partenaire protéique d'éditosome KREPA2. De manière intrigante, le domaine OB-fold et les doigts de zinc de KREPA2 ne semblent pas influencer l'activité de ligature de l'ARN de KREL1. De plus, un motif KWKE spécifique sur l' $\alpha$ H4 de la CTD de KREL1 est impliqué dans l'auto-adénylylation de la ligase, analogue au motif VI des ADN ligases. Ces résultats détaillent le mode d'action de KREL1 CTD dans la ligation et le rôle fonctionnel de l'interaction de KREPA2. Nous proposons en outre un modèle pour le changement de conformation " ouvert-fermé " de KREL1 afin de fournir le motif VI au domaine d'adénylylation pour l'hydrolyse de l'ATP qui est médiée par KREPA2. Ces résultats du mécanisme de ligature de l'ARN KREL1/KREPA2 sont uniques aux kinétoplastides.

## Acknowledgements

This research is only made possible with the help and support of everyone around me. Thank you, Professor Reza Salavati, for the research opportunity through my undergraduate and master's degrees. I entered your laboratory with the objective of finding a motif on KREL1, and through your guidance and support, we have discovered so much more about these enzymes. You have provided me with an exceptional research experience with many opportunities and accomplishments. On my first day at the lab, I met the Ph.D. candidate, Dr. Vaibhav Mehta, who devoted his time to teaching me the necessary laboratory skills and has been my mentor since. Thank you so much for all your help, especially in the preparation of the manuscript. I would also like to thank the past and present colleagues of the Salavati lab and of the Institute of Parasitology, working alongside everyone has been a pleasure.

I am thankful to obtain a master's degree at McGill University, for which I owe gratitude to many alumni. Thank you, Prof. Igor Cestari and Dr. Vaibhav Mehta for being part of my advisory committee. Thank you to Dr. Kalle Gehring for the review of my thesis. Thank you, Prof. Marilyn Scott, for your help during the parasitology seminar series. I would also like to thank all of the professors at the Institute of Parasitology for your participation and feedback during my seminars.

I am forever grateful for all the support from my family and friends. Thank you to both of my parents for working as hard as you did to provide me with this opportunity. Thank you to my sisters, Laura, Dahlia, and Melissa, for taking an interest in my work. I would also like to thank the *The Fellowship* for their encouragement during this lifelong friendship. Also, I would like to thank my partner, for all of her help, motivation, and love.

## Contribution to original knowledge

My thesis investigates the ligation mechanism of editosome proteins KREL1 and KREPA2. I conduct a detailed structural, functional, and mutational analysis of these proteins to understand their mechanism of action and identify critical regions of interest. To date, the role of KREL1 and KREPA2 domains has yet to be described and understood entirely in the ligation mechanism.

This thesis describes five optimized and efficient *in vitro* assays to test recombinant KREL1 and KREPA2; a pulldown assay for protein-protein interaction, fluorescent-based dsRNA ligation assays for the three-step mechanism and step 3 of ligation, and radiolabelled assays for steps 1 and 2 of ligation. Collectively, I expressed 100 mutants of KREL1 and KREPA2 to test these assays and uncover the unknowns of these proteins.

My thesis work has been published in the *RNA journal*, titled “The discovery and characterization of two novel structural motifs on the C-terminal domain of kinetoplastid RNA editing ligases.” My findings detail the role of the diverged CTD of KREL1 in relation to adenylation activity and KREPA2 interaction. I have identified specific motifs as critical regions that appear as novel inhibition targets as well as provide insight into the mechanism of dsRNA ligation in kinetoplastid parasites.

First, I narrowed the specific interaction site between KREL1 CTD  $\alpha$ H3 and KREPA2  $\alpha$ H1 and further identified the electrostatic interaction that takes place. I as well ruled out the involvement of the OB-fold and ZnF domains of KREPA2 in this interaction. The activity of KREL1 was further examined in each step of the three-step ligation reaction in the presence and absence of its interacting partner, KREPA2. For the first time, I show that KREL1 NTD is active independently of its CTD in the final step of ligation, contrary to ATP hydrolysis (step 1), whereby the CTD is essential. In the presence of KREPA2, I show that its ability to enhance KREL1 activity is only in step 1 and not in the subsequent ligation steps. I further focused on uncovering the mechanism of step 1 and the role of the CTD. Through mutational studies, I discovered a functional KWKE motif VI (aa 441-444) on KREL1 CTD that is crucial for auto-adenylation, synonymous with the motif VI of DNA ligase. I also bring evidence that the OB-fold and ZnFs of KREPA2 do not have a role in KREL1 RNA ligation directly. In summary, my work has functionally annotated the domains of KREL1 and KREPA2 during the ligation mechanism.

## **Contribution of authors**

This thesis contains the contents of a manuscript that was co-authored by Dr. Vaibhav Mehta and Prof. Reza Salavati.

## Abbreviations

%SID	Percent sequence identity
aa	Amino acid
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BSF	Bloodstream form
CRB	Calmodulin resin beads
CTD	C-terminal domain
dsRNA	Double-stranded ribonucleic acid
ETC	Electron transport chain
FL	Full length
gRNA	Guide RNA
HAT	Human African Trypanosomiasis
KREL1	Kinetoplastid RNA Editing Ligase 1
KREPA2	Kinetoplastid RNA Editing Protein A2
L.m.	<i>Leishmania major</i>
NTD	N-terminal domain
OB-fold	Oligonucleotide binding fold
PAL	Pre-adenylylated ligation
PPI	Protein-protein interaction
RECC	RNA editing core complex
T.b.	<i>Trypanosoma brucei</i>
T.c.	<i>Trypanosoma cruzi</i>
T4Rnl2	Bacteriophage T4 RNA ligase 2

VLR	Variable loop region
WT	Wild type
ZnF	Zinc finger
$\alpha$ H#	Refers to the alpha helix number

## List of Figures

Figure 1. Number of HAT cases in the past 20 years recorded by WHO.....	17
Figure 2. The kinetoplastid DNA network made of mini and maxi circles.....	19
Figure 3. RNA Editing Core Complex .....	21
Figure 4. RNA ligation of KREL1.....	22
Figure 5. Full-length predicted structures of KREL1 and KREPA2. ....	26
Figure 6. Sequence alignment of KREPA2 structural domains.....	27
Figure 7. Expression of KREL1 and KREPA2 recombinant proteins.....	30
Figure 8. KREL1 adenylylation activity in pH 5-10.....	32
Figure 9. Control experiments for ligation assays. ....	33
Figure 10. KREL1 shares a conserved adenylylation domain with Human DNA ligase. ....	35
Figure 11. Sequence alignment of RNA ligases conserved motifs of the catalytic ATP binding domains.....	37
Figure 12. $\alpha$ H2- $\alpha$ H3 superimposition of KREL1 and T4Rnl2. ....	37
Figure 13. The predicted structure of KREPA2 includes an OB-fold domain and two zinc finger domains. ....	38
Figure 14. Pulldown assay of KREL1 with KREPA2.....	40
Figure 15. <i>In vitro</i> functional assays of KREL1 full-length and domain truncated mutants with and without KREPA2. ....	42
Figure 16. KREL1 ATP binding pocket mutants. ....	43
Figure 17. KREL1 Variable loop region.....	44
Figure 18. Locating motif VI on KREL1 C-terminal domain. ....	46
Figure 19. Activity of motif VI mutants in ligation and pre-adenylylated ligation assays.....	47
Figure 20. KREL1 adenylylation kinetics .....	48
Figure 21. <i>In vitro</i> functional assays of KREL1 with and without KREPA2 over time.....	49

Figure 22. KREL1-KREPA2 prediction model of interacting residues. ....	51
Figure 23. Model for ligase auto-adenylation.....	55
Figure 24. Multiple sequence alignment of RNA ligases. ....	58
Figure 25. Plasmid designs of pET30b with KREL1 insert and pLEXY with KREPA2 insert. ....	60
Figure 26. Full-length amino acid sequence of KREL1 and KREPA2 from strain Tb927 .....	61

## List of Tables

Table 1. Summary of Kinetoplastid causing disease: HAT, Chagas disease and Leishmaniasis (Field et al. 2017; Stuart et al. 2008; De Koning 2020).....	18
Table 2. Group A proteins of the RNA editing core complex .....	24
Table 3. List of KREL1 and KREPA2 mutations.....	28
Table 4. Percent identity matrix of full-length RNA ligases .....	61
Table 5. Percent identity matrix of the catalytic domain of RNA ligases .....	61
Table 6. Percent identity matrix of the helical C-terminal domain of RNA ligases.....	62
Table 7. Root-mean squared deviation matrix of the domains of RNA ligases .....	62
Table 8. Percent identity matrix of full-length Kinetoplastid A1 and A2 proteins.....	63
Table 9. Percent identity matrix of full-length Kinetoplastid A proteins of <i>Trypanosoma brucei</i> .....	63
Table 10. Root-mean squared deviation matrix of full-length Kinetoplastid A proteins of <i>Trypanosoma brucei</i> .....	63

## Introduction

The kinetoplastid parasites *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania* spp. cause endemic diseases worldwide, namely human African trypanosomiasis, Chagas disease, and leishmaniasis (Alalaiwe et al. 2018; Stuart et al. 2008). These early diverged eukaryotes harbour within their mitochondria, a uniquely organized genome of interlocked dsDNA in the form of mini and maxicircles, known as the kinetoplast (Jensen and Englund 2012). The maxicircles encode several subunits of the electron transport chain and one subunit of the mitochondrial ribosome, while minicircles transcribe short *trans*-acting template guide RNA (gRNA) for use in post-transcriptional maturation of the maxicircle transcripts (Aphasizhev and Aphasizheva 2014). Catalyzed by an ~800 kDa multi-protein RNA editing catalytic complex (RECC), this process entails insertions and/or deletions of uridine (U) residues to varying degrees in the premature transcripts (Kable et al. 1996; Stuart et al. 2005). Among the catalytic RECC components, the kinetoplastid RNA editing ligase 1 (KREL1) and the kinetoplastid RNA editing protein A2 (KREPA2) are specific for ligation post-U-deletion and are paralogous to KREL2 and KREPA1 that are specific for ligation post-U-insertion (Schnauffer et al. 2003). KREL1 is the essential ligase for parasite survival and functional editing *in vivo*, while its interacting partner, KREPA2, is critical for RECC stability and KREL1 integration into the complex (Huang et al. 2002; Guo et al. 2008; Schnauffer et al. 2001a).

KREL1's ligation mechanism is analogous to the extensively characterized DNA and RNA ligases (Doherty and Suh 2000; Sriskanda and Shuman 1998; Subramanya et al. 1996; Martins and Shuman 2004; Ho et al. 2004). Catalysis can be broken down into three key steps: (1) KREL1 auto-adenylation, where the ligase catalyzes a covalent linkage between a conserved lysine of motif I with the  $\alpha$ -phosphate of ATP to form a KREL1-AMP intermediate while releasing pyrophosphate; (2) RNA adenylation, transfer of AMP to the 5' PO<sub>4</sub> termini of the nicked RNA duplex; (3) phosphodiester bond formation, a nucleophilic attack by the 3' OH group enables the de-adenylated ligase to covalently bind the nicked ends of the RNA while releasing AMP.

Several studies outline the structural domains of KREL1 and KREPA2 and their potential role in the ligation mechanism (Deng et al. 2004a; Mehta et al. 2015; Worthey et al. 2003; Schnauffer et al. 2003). The crystallized N-terminal domain of KREL1 (PDB code 1XDN) reveals the ATP binding site, formed by beta-strands containing residues from five key motifs (I, III, IIIa,

IV and V). These functional motifs are highly conserved in nucleotidyl transferase enzymes such as DNA/RNA ligases and mRNA capping enzymes (Deng et al. 2004; Subramanya et al. 1996). Sequence alignments and experimental mutation analysis revealed a conserved sixth motif (VI) in DNA ligases critical for step 1, ligase auto-adenylation, located outside the binding pocket. The motif is found on the CTD adjacent to its oligonucleotide binding fold (OB-fold) domain (Sriskanda and Shuman 1998; Samai and Shuman 2012). This motif has been elusive in the KRELS, due to its highly diverged CTD lacking the OB-fold, unlike DNA ligase and mRNA capping enzymes (Flynn and Zou 2010; Ellenberger and Tomkinson 2008). Sequence alignment-based search tools identify bacteriophage T4 RNA ligase 2 (T4Rnl2) as the closest ortholog for the KRELS (Nandakumar et al. 2006; Ho and Shuman 2002). Although the CTD of T4Rnl2 is dispensable in RNA adenylation, the deletion of this domain in KREL1 appears detrimental to its activity (Ho et al. 2004; Mehta et al. 2015) implicating a potential motif VI-like region in its CTD.

The KREL1 CTD is mainly known for its interaction with KREPA2 as established through yeast-2-hybrid and mutational studies (Schnauffer et al. 2010; Mehta et al. 2015). The current hypothesis for KREPA2's involvement in ligation focuses on its OB-fold domain being provided *in trans* for RNA recognition and binding, in a mechanism analogous to the fused OB-fold domain in DNA ligases that interacts with nicked DNA during ligation (Håkansson et al. 1997; Subramanya et al. 1996). While this involvement has not been proven experimentally, the theory has been supported by KREPA2-mediated enhancement of KREL1 activity *in vitro* (Mehta et al. 2015). Moreover, enhancement of KREL1 auto-adenylation indicated a likely motif VI also to be provided *in trans*, albeit KREL1 still exhibits robust activity *in vitro* without its interaction partner.

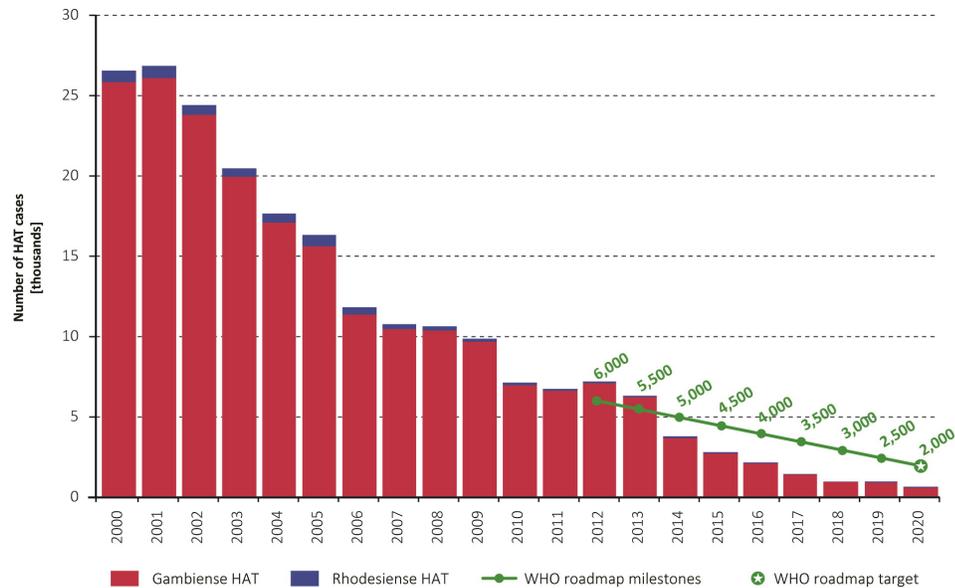
In this study, we aimed to further characterize the ligation mechanism of kinetoplastid RNA editing ligases. We performed an extensive mutational analysis involving terminal truncations, internal truncations, point mutations and group mutations of the recombinant (r) KREL1 and KREPA2 proteins of *T. brucei*. We narrowed down the KREL1-KREPA2 regions of contact and further tested the effect of this interaction in ligation *in vitro*. We also identified a diverged motif VI on KREL1 critical for step 1 activity, ligase auto-adenylation, which validates the functional role of its CTD. Our results implicate a crucial role for the motif VI in ATP hydrolysis, enhanced by KREPA2 interaction.

## Comprehensive literature review

### *Trypanosomes and their diseases*

Kinetoplastid parasites are unicellular eukaryotes with two intermediate hosts and three developmental stages. The flagellated pathogens, measuring 10-30  $\mu\text{m}$ , reside in an invertebrate vector and further transmit to a vertebrate host, including humans (Van Den Abbeele et al. 1999). Transmission to the bloodstream acts as a highway for the parasites to migrate to a preferred location of the body, including the cephalic, cardiac, and gastrointestinal systems etc. and cause morbidity (Stuart et al. 2008). Below I discuss the parasites and the diseases they cause.

Human African trypanosomiasis (HAT), also known as sleeping sickness, is caused by two subspecies of the protozoan *Trypanosoma brucei*: *T. b. gambiense*, causing the endemic disease in Central and West Africa and *T. b. rhodesiense*, causing east African trypanosomiasis (WHO 2022). A zoonotic subspecies, *T. b. brucei*, threatens agriculture and increases the challenge of containing eradication (Büscher et al. 2017). Tsetse flies of the genus *Glossina* transmit the parasite. Over 30 species have been identified, six of which act as vectors for the disease, and the two most common are *G. palpalis* and *G. fuscipes* (Wamwiri and Changasi 2016). During a blood meal of an infected tsetse fly, the metacyclic trypomastigotes are released and differentiate into its long slender bloodstream form (BSF) through extracellular signals that induce signal transduction pathways (Macleod et al. 2007; Walsh and Hill 2021). This results in the hemolymphatic stage of the disease with symptoms of headache, joint pain, and fever (Büscher et al. 2017). The parasite successfully evades the immune system through antigenic variation, causing parasitemia to persist for years (Malvy and Chappuis 2011). The BSF form can invade organs such as the central nervous system by crossing the blood-brain barrier causing chronic encephalopathy (Schultzberg et al. 1988). The second stage of the disease can result in mental disorders, sleep abnormalities and if left untreated, death (Malvy and Chappuis 2011). In the 1990s, the seemingly eradicated disease re-emerged, becoming endemic in 36 Sub-Saharan African countries with 35,000 cases yearly (Franco et al. 2022). With increased control efforts by the World Health Organization (WHO), as of 2009, the number of cases has reached below 10,000 for the first time in 50 years and continues to decrease, as described in Figure 1 (Franco et al. 2022). Today there are reports of 2000 new cases a year, with still 500,000 currently infected and at risk of death as HAT remains a unsolved health problem (Büscher et al. 2017).



**Figure 1. Number of HAT cases in the past 20 years recorded by WHO.**  
*Figure reproduced from PLOS Neglected Tropical Diseases by Franco et al. 2022.*

Chagas disease is the persistent infection of *Trypanosoma cruzi* that is endemic to 21 countries in Latin America with prevalence in North America and Europe, causing 50,000 deaths per year (WHO, 2019). The parasite is transmitted by contact with the triatomine bug. Excreted products of the bug contain trypomastigotes that can invade cells around the bite wound. These are released into the bloodstream, where they can infect cells of the heart, gastrointestinal tract, central nervous system, smooth muscle and adipose tissue (Stuart et al. 2008). It can also be transmitted congenitally, from a blood transfusion or organ transplantation, whereby screening assays aid in reducing incidence (Pérez-Molina and Molina 2018; Lidani et al. 2019).

The disease can be treated with benznidazole and nifurtimox, which have proven effective in curing the disease when administered during the acute phase (Field et al. 2017). The drugs strongly inhibit dehydrogenase activity, reducing the mitochondrial membrane potential (Boiani et al. 2010). However, the efficacy diminishes with a longer interval between infection onset and initiation of treatment. Moreover, current detection methods, such as enzyme-linked immunosorbent assay (ELISA) and hemagglutination, are inadequate in the early stages (Tarleton et al. 2007). Moreover, the initial phase may be asymptomatic, where less than 50% of individuals experience swelling at the bite location, headaches, and fever. In the chronic phase, two months post-infection, the parasites harbour within the heart and digestive tract, causing cardiac disorders

and an enlarged intestine and colon. If untreated, it may lead to death, such as heart failure in 30% of infected individuals (Stuart et al. 2008).

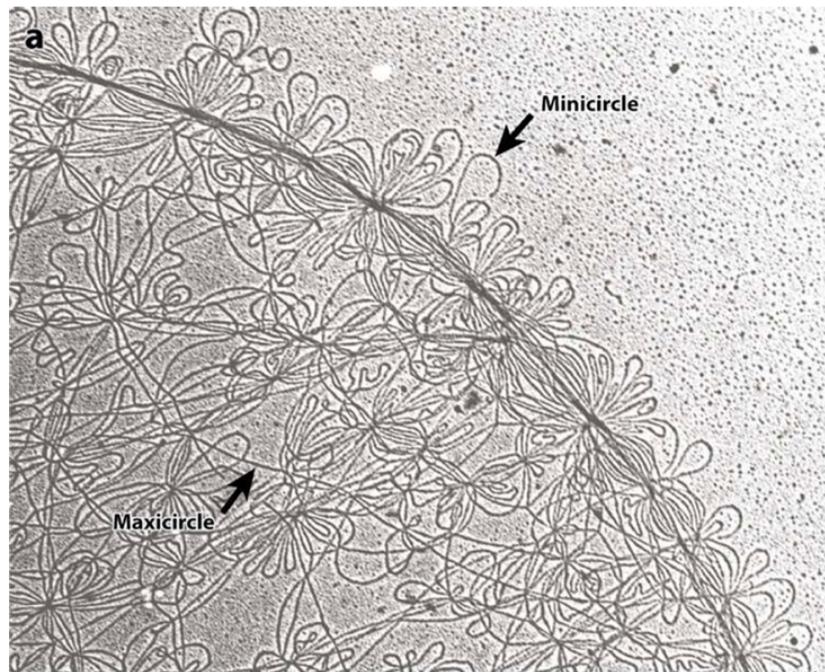
Leishmaniasis can be caused by 21 species of *Leishmania* transmitted to humans by a bite from an infected female phlebotomine sandfly (WHO, 2019). The disease, characterized by location, takes on three forms: cutaneous (CL), visceral (VL) and mucocutaneous (ML). According to reports from WHO, Afghanistan and the Syrian Arab Republic suffer the most from CL, while Ethiopia and Kenya are affected mainly by VL. The most common form of the disease, CL, leads to more than five million cases annually. The condition in this form is not fatal, however, it can leave permanent scars and disability from an excessive number of ulcers (Reithinger et al. 2007). ML targets mucous membranes of the nose, mouth, and throat, causing lesions and/or destruction of the affected area. VL caused by *L. infantum* and *L. donovani* is characterized by fever, weight loss, splenomegaly and anemia, which can be fatal if left untreated (Stuart et al. 2008). *Leishmania* spp., in all cases, shares a similar life cycle that results in the infection of macrophages (Handman and Bullen 2002). While in the case of CL, the parasite infects macrophages within the skin, *L. infantum* and *L. donovani* will target macrophages of the lymphoid tissues rendering the host immunocompromised and susceptible to secondary infection (Denkers and Butcher 2005).

**Table 1. Summary of Kinetoplastid causing disease: HAT, Chagas disease and Leishmaniasis (Field et al. 2017; Stuart et al. 2008; De Koning 2020)**

	<b>Human African Trypanosomiasis</b>	<b>Chagas disease</b>	<b>Leishmaniasis</b>
<b>Causative parasite</b>	<i>T. brucei gambiense</i> <i>T. brucei rhodiense</i>	<i>T. cruzi</i>	<i>Leishmania</i> spp. i.e. <i>L. donovani</i> & <i>L. infantum</i> causing VL
<b>Vector</b>	Tsetse fly ( <i>Glossina</i> spp.)	Kissing bug ( <i>Triatominae</i> spp.)	Phlebotomine sandfly
<b>Location</b>	Sub-Saharan Africa	South and Central America	Afghanistan, Syria, Kenya, Ethiopia, S & C america
<b>Deaths/year</b>	>2,000	>7,000	~50,000
<b>DALY</b>	203,497	252,204	1,068,537
<b>Treatment</b>	Suramin, Pentamidine, Melarsoprol, Elornithine & Fexinidazole	Benznidazole, Nifurtimox	Amphotericin, Miltefosine, Paromomycin & Pentavalent antomionials

## *The kinetoplast and post-transcriptional RNA editing*

The parasitic protozoans *T. brucei* spp., *T. cruzi* and *Leishmania* spp. belong to a group referred to as kinetoplastids. They share a unique structure within the mitochondria called the kinetoplast, a dense mass of DNA (kDNA) that encodes subunits of the ETC (Jensen and Englund 2012). While cellular ATP is generated using the oxidative phosphorylation system in the insect stage, the bloodstream form in the mammalian host produces ATP through aerobic glycolysis (Zíková 2022). Although oxidative phosphorylation is repressed in the bloodstream form, the ETC undergoes remodelling for the reverse activity of the F<sub>0</sub>F<sub>1</sub>-ATP synthase to maintain a proper mitochondrial membrane potential, further emphasizing the importance of kDNA in all stages (Schnauffer et al. 2005).



**Figure 2. The kinetoplastid DNA network made of mini and maxi circles.**

*Electron microscopy of isolated kDNA network. Figure reproduced from Frontiers in Cellular and Infection microbiology, Maldonado et al. 2012.*

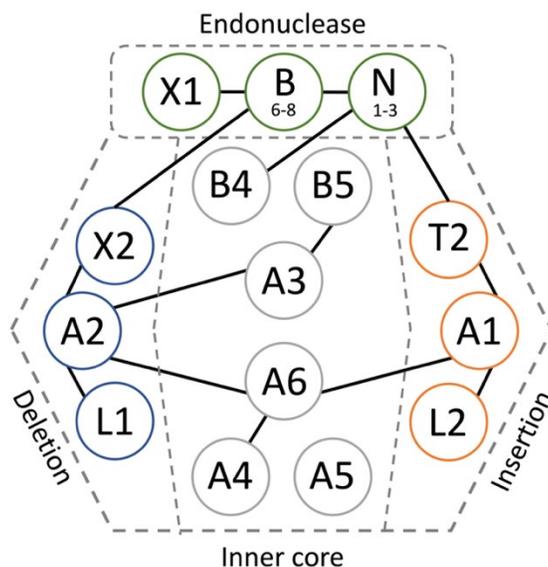
The kinetoplast comprises a large, interconnected network of circular kDNA, maxicircles and minicircles (Figure 2). Maxicircle genes encode ribosomal RNA and subunits of the ETC embedded into the mitochondrial membrane (Hajduk and Ochsenreiter 2010). Among the 18 transcripts encoded by maxicircles, 12 require a post-transcriptional modification termed RNA

editing (Read et al. 2016). This was first described in the mRNA of cytochrome-c oxidase subunit II (COII) (Benne et al. 1986). The minicircles are transcribed into short antisense RNAs with a 3'poly-uridine tail, termed guide RNA (gRNA), that directs the site-specific editing (Hajduk and Ochsenreiter 2010; Blum and Simpson 1990).

RNA editing has been observed in all kinetoplastids, while it varies in extensiveness between species and life stages (Simpson et al. 2000; Read et al. 2016). It is a post-transcriptional modification comprised of insertions and deletions of uridylates (Read et al. 2016). The gRNAs hybridize to regions of the maxicircle transcripts through complementary base pairing, forming a double-stranded RNA segment with a mismatch that dictates the editing of the pre-mRNA (Stuart et al. 1997). The insertion and deletion of uridines are catalyzed by a multiprotein complex termed the editosome (McDermott et al. 2016).

#### *The RNA editing core complex (RECC)*

The post-transcriptional RNA editing mechanism in kinetoplastids is catalyzed by the ~20S (~800 kDa) RNA editing core complex (RECC) described in (Aphasizheva et al. 2020). The RECC components are nuclear-encoded and contain a mitochondrial import signal. The complex requires the proper assembly of subcomplexes directed by protein-protein interactions to maintain its integrity (Figure 3) (Schnauffer et al. 2010; McDermott et al. 2016; Huang et al. 2002). The purification and detection of the composition of the RECC have been detailed in (Aphasizheva et al. 2014) among other complexes of the editosome, including those of accessory proteins that mediate editing fidelity and efficiency (Mehta et al. 2020). At the editing site, exonuclease/endonuclease activity cleaves the phosphodiester bond on the pre-mRNA (Aphasizheva et al. 2020). This allows for either the U-addition from 3' terminal uridylyl transferase (TUTase) or U-deletion from ExoUase activity (Ringpis et al. 2010; Stagno et al. 2010). In both cases, the nicked dsRNA is ligated by an RNA editing ligase (Salavati et al. 2012). KREPA and B proteins form the inner core of the complex essential for its integrity and the assembly of subcomplexes.



**Figure 3. RNA Editing Core Complex**

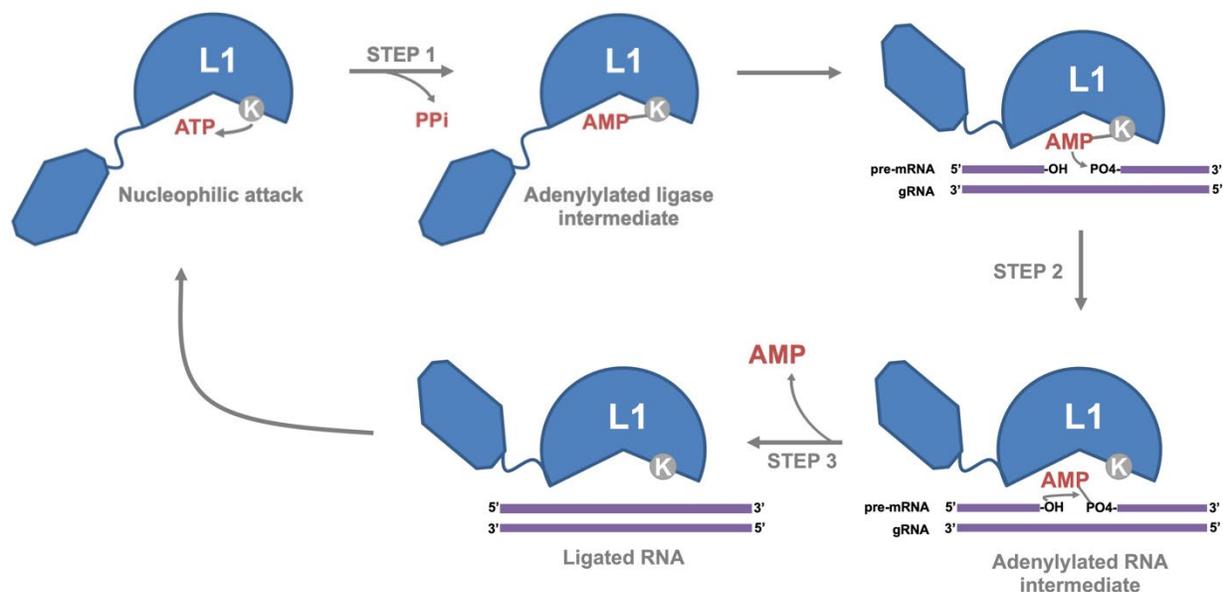
Diagram depicts the RNA editing core complex made of three subcomplexes, insertion, deletion, and endonuclease, around an inner core of A and B proteins. Protein-protein interaction is represented with a connected line.  $X_{1,2}$  (exonucleases),  $B_{1-8}$  (KREPB proteins),  $N_{1-3}$  (endonucleases),  $T_{1,2}$  (TUTases),  $A_{1-6}$  (KREPA proteins),  $L_{1,2}$  (ligases) (Aphasizheva et al. 2020).

### Kinetoplastid RNA Editing Ligases

Kinetoplastids have two ATP-dependent RNA editing ligases that function in the U-insertion/deletion post-transcriptional editing: KREL1 and KREL2. RNA editing ligases are part of an early lineage of nucleotidyl transferases such as T4 RNA ligase 2 (T4Rnl2), DNA ligase and mRNA capping enzymes (Ho et al. 2004; Doherty and Suh 2000; Håkansson et al. 1997). They became of particular interest in kinetoplastids for their essential enzymatic role in RNA editing. The kinetoplastid RNA ligases share a three-step ligation mechanism highly conserved with other nucleotide transferases (Figure 4). The hydrolysis of ATP forms a ligase-AMP intermediate that further adenylyates the 5' PO<sub>4</sub> of the pre-mRNA that drives a nucleophilic attack from the 3'OH. The phosphodiesterase bond completes the ligation of the mRNA fragments. *In vitro* and *in vivo* KREL1 knock-out studies have denoted abolished RNA editing of ETC subunits and detrimental effects on parasite viability and parasitemia in mice (Schnauffer et al. 2001). However, RNAi silencing of the homologous enzyme KREL2 of the insertion subcomplex did not result in a similar phenotype. The presence of KREL1 was sufficient to maintain cell growth and *in vivo* editing, thus

proposing a dual role of KREL1 in both insertion and deletion editing (Drozd et al. 2002; Aphasizheva et al. 2020).

In 2004, a 1.2 Å crystallized structure was published of the adenylylation domain of KREL1 *T. brucei* (PDB code 1XDN), which confirmed its annotation (Deng et al. 2004a). The ATP binding pocket is formed by five functional motifs (I, IIIa, IIIb, VI, and V) essential for ATP catalysis in all nucleotide transferases (Mehta et al. 2015; Unciuleac and Shuman 2015; Gu et al. 2016; Srisakanda and Shuman 1998). Furthermore, *in vitro* screening of compounds has shown potential in finding KREL1 competitive inhibitors (Zimmermann et al. 2016; Salavati et al. 2012). However, to date, no compound has been developed to successfully inhibit KREL1 *in vivo* and remains an area for further investigation. Interestingly, the adenylylation domain of KREL1 is not functional when independent of its CTD (Mehta et al. 2015). Although CTD has not been crystallized (Deng et al. 2004), sequence analysis to date records bacteriophage T4Rnl2 as its closest homologue (Ho and Shuman 2002). The crystallized structure of this enzyme (PDB code 2HVR) shows the domain comprised of four alpha helices (Ho et al. 2004). Contrary to KREL1, the NTD of T4Rnl2 is functionally independent in *in-vitro* dsRNA ligation (Ho et al. 2004).



#### Figure 4. RNA ligation of KREL1

Schematic of the 3-step ligation mechanism. L1 depicted in blue has an ATPase domain and a CTD attached via a variable loop region. KREL1 is adenylylated in step 1 to form the KREL1-AMP intermediate. The AMP is then transferred to the 5' end of the 3' RNA fragment of the nicked dsRNA in step 2. The final step is formation of the phosphodiester bond. The result of the ligated mechanism is a sealed dsRNA.

Mutational studies of Chlorella virus DNA ligase had narrowed an RxDK motif VI conserved in DNA ligases and mRNA capping enzymes (Sriskanda and Shuman 1998). The RxDK motif VI, located downstream (following the OB-fold domain) from the other five motifs, is essential for enzyme adenylation. While RxDK is not conserved to KRELs, the shared phenotype of CTD truncated KREL1 mutant suggests the possibility of finding a cryptic motif VI (Mehta et al. 2015).

KRELs do not have a structure similar to the OB-fold domain of DNA ligases. It has been shown that KREL1 interacts with KREPA2, and KREL2 interacts with KREPA1 in separate subcomplexes of the RECC (Schnauffer et al. 2010). Interestingly, the KREPA proteins have an OB-fold domain and in theory, can provide this to KRELs *in trans*. RNAi has supported the role of KREPA proteins in complex assembly and integrity; however, the functional role of providing this domain in ligation has yet to be proven analogous to that of DNA ligases (Huang et al. 2002).

#### *Kinetoplastid RNA Editing Proteins A<sub>1-6</sub>*

The interacting partner of KREL1 is the RNA editing protein KREPA2. The enzyme is one of six A proteins (KREPA<sub>1-6</sub>) of trypanosomes that are part of the RECC (Figure 3). The catalytic complex is divided into subcomplexes formed through PPI; KREPA1 interacts with ligase (KREL2) and exonuclease (KREX2) to form the U-insertion subcomplex, KREPA2 protein interacts with ligase (KREL1) and TUTase (KRET2) to form the U-deletion subcomplex (Aphasizheva et al. 2020). The remaining three KREPA proteins (KREPA<sub>3-6</sub>) form the inner core that interacts with each other and aid in the assembly of the subcomplexes (Aphasizheva et al. 2020). Previous studies have investigated the structure and function of these enzymes, summarized in Table 2.

The KREPA proteins share a structurally homologous oligonucleotide binding fold domain. The domain consists of six beta sheets arranged into a barrel with a winged helix, previously described (Park and Hol 2012) (PDB code 4DNI). KREPA1, 2 and 3 also have two C2H2 zinc fingers (Worthey et al. 2003). Both domains have been implicated in forming the core of the RECC (Aphasizheva et al. 2020; Schnauffer et al. 2010). This is supported by studies on all KREPAs, except for KREPA5, that provide evidence on the essentiality for the stability and integrity of the RECC that is detrimental to *in vivo* editing and the parasite viability (Drozd et al.

2002; Huang et al. 2002; Salavati et al. 2006; Kala and Salavati 2010; Guo et al. 2008; Tarun et al. 2008; Guo et al. 2010; Kala et al. 2012). Apart from PPI, KREPA proteins have been shown to possess other functions such as endo-exoribonuclease activity from KREPA3 both *in vitro* and *in vivo* (Brecht et al. 2005) as well as enhanced KREL1 ligation activity *in vitro* (Mehta et al. 2015).

**Table 2. Group A proteins of the RNA editing core complex**

<b>Enzyme</b>	<b>Domains</b>	<b>Function</b>
KREPA1	ZnF1 & ZnF2 OB-fold	Part of the stability and integrity of the RECC. Essential for RNA editing associated with U-insertion. (Drozdz et al. 2002)
KREPA2	ZnF1 & ZnF2 OB-fold	Part of the stability and integrity of the RECC. Essential for KREL1 integration and ligase activity. (Huang et al. 2002)
KREPA3	ZnF1 & ZnF2 OB-fold	Essential for the integrity of the RECC and RNA editing. (Guo et al. 2008)
KREPA4	OB-fold	Essential for editosome stability and U-insertion/deletion activity. (Kala et al. 2012; Salavati et al. 2006)
KREPA5	OB-fold	-
KREPA6	OB-fold	Essential for RNA editing and structural integrity of the editosome. (Tarun et al. 2008)

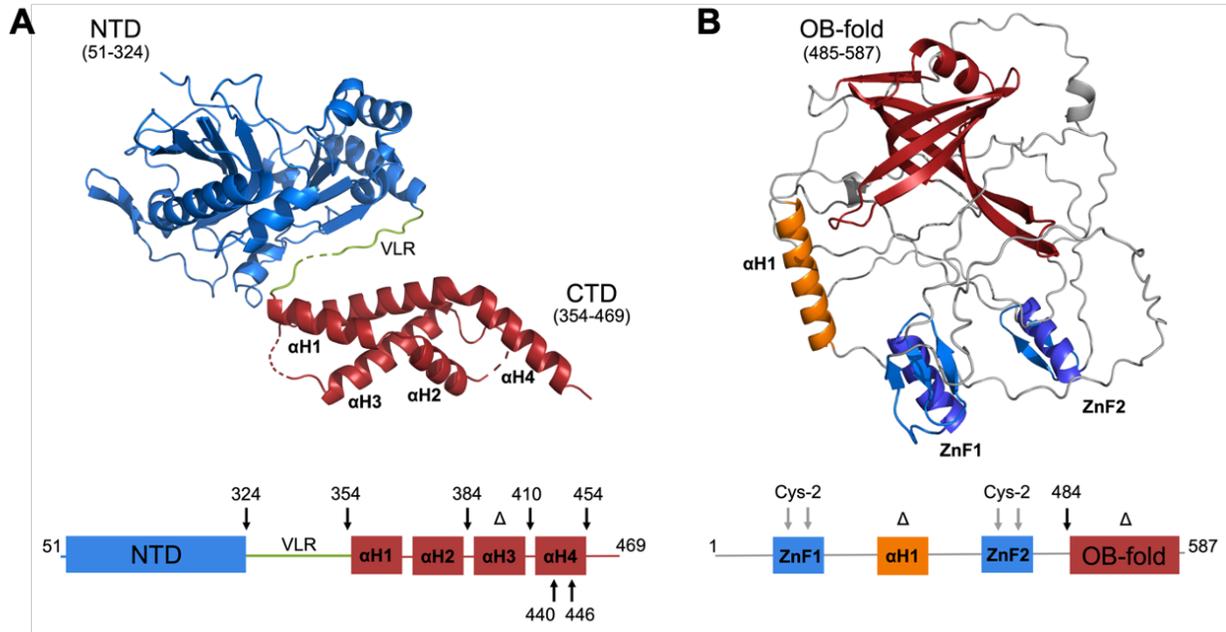
## Methodology

My thesis is comprised of a structural, functional, and mutational analysis of KREL1 and KREPA2 proteins. First, using AlphaFold2, the full-length structures of KREL1 and KREPA2 were predicted. Next, these structures were further analyzed through pairwise structural alignments with related proteins along with multiple sequence alignments to highlight conserved residues/regions. The full-length genes were then cloned into their respective plasmids and recombinantly expressed and purified. The rKREL1 and rKREPA2 were then tested for protein-protein interaction and ligation activity using five *in vitro* assays including a pulldown assay, ligation assay, adenylation assay, RNA adenylation assay and pre-adenylylated ligation assay. Alongside these experiments, a series of constructs including truncations, internal deletions, group mutations and point mutations were made to both KREL1 and KREPA2 and tested *in vitro* to identify critical regions for activity.

### *Sequence alignment and structure predictions*

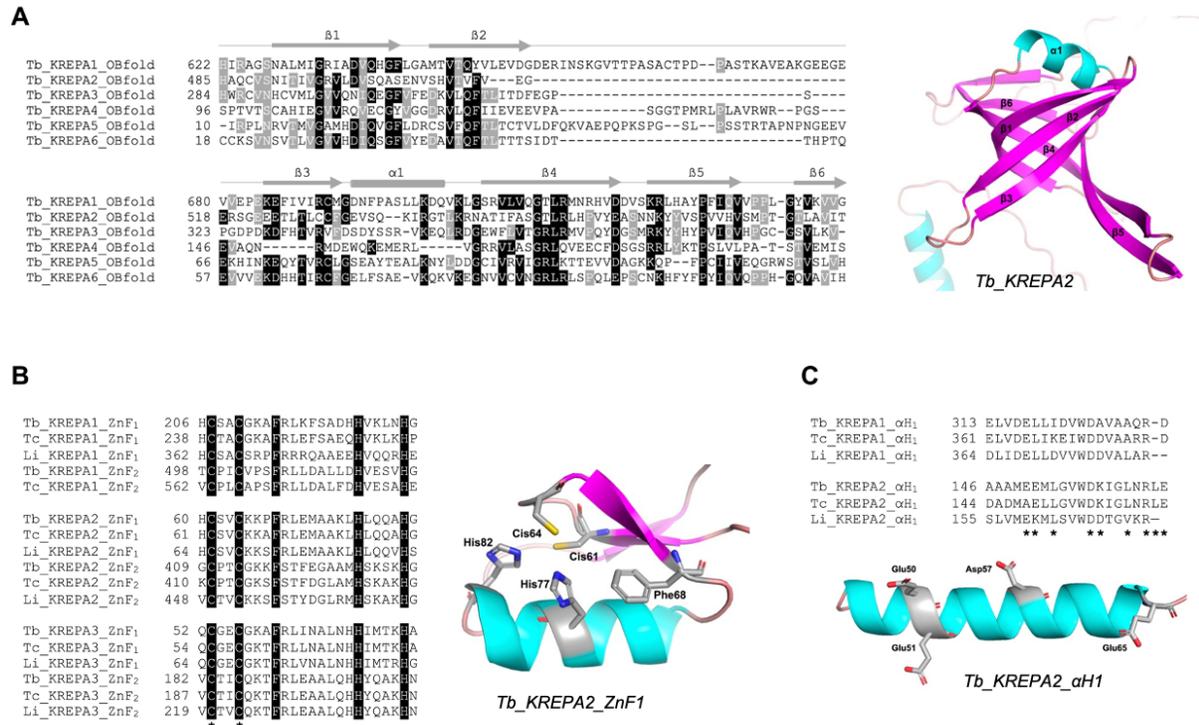
Amino acid sequences of KREL1 and KREPA2 from *T. brucei* and related species were extracted from the TriTryp database (<https://tritrypdb.org/tritrypdb/app>) and aligned to related proteins using Clustal Omega (Sievers et al. 2011). We excluded the 50 amino acid mitochondrial import signal of kinetoplastid RNA ligases (Appendix Figure 24).

Structural predictions of KREL1 and KREPA2 were achieved using the AlphaFold Colab database developed by DeepMind and EMBL's European Bioinformatics Institute (Jumper et al. 2021; Varadi et al. 2022) and used in the analysis of the structural domains (Figure 5, Figure 6). We then used RCSB Protein Data Bank PDB (<https://www.rcsb.org/alignment>) to perform a pairwise structural alignment of KREL1 and KREPA2 domains to related proteins.



**Figure 5. Full-length predicted structures of KREL1 and KREPA2.**

(A) Full-length AlphaFold structure prediction of KREL1 *T. brucei*. The structure has two domains: NTD and CTD that are connected by a variable loop region (VLR). The four-alpha helices of the CTD are labelled. A schematic of the mutations made on KREL1 is shown below the structure. The black arrows represent the site of the truncations and the  $\Delta$  represents the internal helix deletion. (B) Full-length AlphaFold structure prediction of KREPA2 *T. brucei*. The structure has an OB-fold domain, two ZnF domains and one alpha helix. A schematic of the mutations made on KREPA2 is shown below the structure. The black arrow at 484 represents the truncation made for the deletion of the OB-fold<sub>7</sub> and the grey arrows represent the alanine substitutions of the cysteine residues.



**Figure 6. Sequence alignment of KREPA2 structural domains.**

(A) Amino acid multiple sequence alignment of the oligonucleotide binding fold (OB-fold) domain of KREPA1-6 of *Trypanosoma brucei* (Tb), with predicted Alphafold structure beside it. (B) Amino acid multiple sequence alignment of the zinc finger (ZnF) domains of KREPA1-3 of *Trypanosoma brucei* (Tb), *Trypanosoma cruzi* (Tc) and *Leishmania infantum* (Li), with predicted Alphafold structure beside it. The sequence alignment shades five conserved residues, C2FH2, which are specific to this ZnF domain family. (C) Amino acid sequence alignment of the αH1 of KREPA1 and KREPA2 of Tb, Tc and Li, with predicted Alphafold structure below it. Each residue mutated for analysis is represented by an asterisk.

### Cloning of full-length KREL1 and KREPA2

KREL1 FL (51-469) of *T. brucei* was cloned in pET30b as previously described (Mehta et al. 2015) to express rKREL1 with an N-terminal 6x his-tag. *T. brucei* KREPA2 (1-587) was excised from a pSG1-KREPA2 construct and cloned into a pLEXY invitro-2 plasmid ([www.jenabioscience.com](http://www.jenabioscience.com)), then outsourced for -TAP-tag fusion plasmid by GenScript Corporation (Piscataway, NJ) (Appendix Figure 25). The expressed rKREPA2 contains a C-terminal affinity purification tag, with a calmodulin tag – TEV protease site – 6x his-tag. The advantages of using this plasmid were efficient affinity chromatography purification and western blot analysis for quantification.

*Cloning of KREL1 and KREPA2 truncations and mutants*

A summary of the truncations, point mutations, and group mutations performed on KREL1 and KREPA2 is listed in Table 3. For KREL1, DNA fragments were generated by PCR between restriction enzymes KpnI and XhoI using the wild-type KREL1 construct (Appendix Figure 25). For KREPA2, DNA fragments were generated by PCR between restriction enzymes NcoI and XbaI using the wild-type KREPA2 construct. Plasmid preparation of KREL1 and KREPA2 mutants was prepared by GenScript Corporation (Piscataway, NJ). Mutations on KREL1 are marked in the multiple sequence alignment (Appendix Figure 24), and mutations on KREPA2 are marked in the multiple sequence alignment (Figure 6).

**Table 3. List of KREL1 and KREPA2 mutations.**

	<b>KREL1 WT (51-469)</b>	<b>KREPA2 WT (1-587)</b>
<b>Truncations</b>	51-454, 51-446, 51-440, 51-430, 51-420, 51-410, 51-384, 51-354 (NTD_VLR), 51-324 (NTD), $\Delta\alpha$ H3 ( $\Delta$ 384-410)	$\Delta$ OB-fold (1-491), $\Delta\alpha$ H1 ( $\Delta$ 146-166)
<b>Point mutants</b>	K87A, K441A, W442A, W442F, K443A, E444A	
<b>Group mutants</b>	<p><b><math>\alpha</math>H3:</b> <u>Hydrophobic residues</u> L399A + L404A + L408A</p> <p><u>Positively charged residues</u> K401A + K405A + K409A</p> <p><u>Negatively charged residues</u> D402A + D406A + E410A</p> <p><b><math>\alpha</math>H4:</b> Residues KWKE (441-444) to AWAA Residues KWKE (441-444) to AAAA</p>	<p><b><math>\alpha</math>H1:</b> <u>Hydrophobic residues</u> L153A + L161A + L164A</p> <p><u>Positively charged residues</u> K158A + R163A</p> <p><u>Negatively charged residues</u> E150A + E151A + D157A + E165A</p> <p><b>ZnF mutant</b> (C61A + C64A + C410A + C413A)</p>

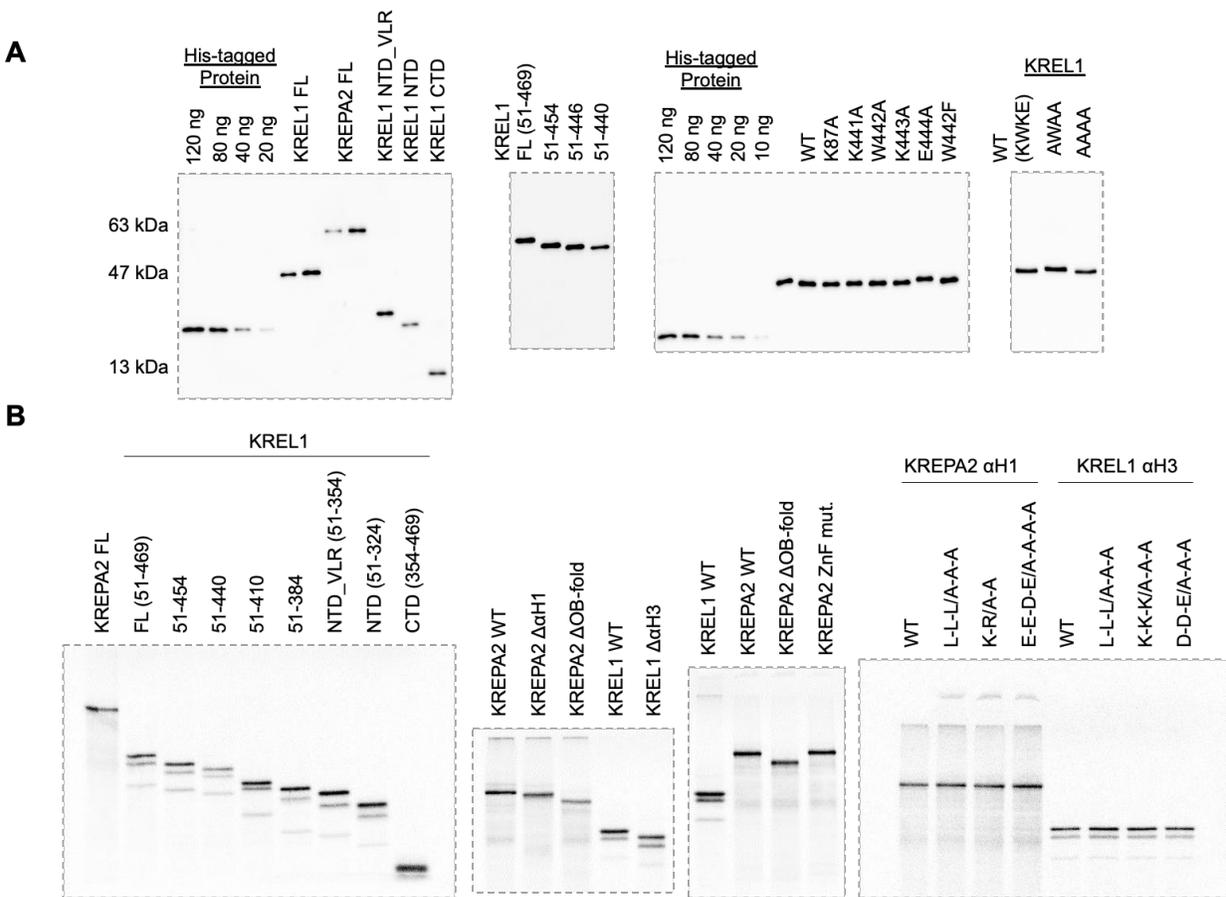
### *Expression and purification of recombinant KREL1 and KREPA2*

The rKREL1 and rKREPA2 wild-type and mutant proteins were expressed *in vitro* with a reticulocyte lysate-based cell-free coupled transcription and translation system (TnT) (Cat. # L1170, Promega). Proteins were expressed with [<sup>35</sup>S] methionine (NEG709A500UC, Perkin Elmer), and used directly for the pull-down assay. An aliquot of the expressed protein was resolved in an SDS-PAGE gel (Figure 7). For *in vitro* functional assays, proteins were expressed with TnT and purified.

For rKREL1 purification, TnT reactions post incubation were added to magnetic nickel beads (Dynabeads; 00972814, Invitrogen) in binding buffer (25 mM NaPO<sub>4</sub> [pH 8.0], 200 mM NaCl and 0.01% Tween 20) and rotated at 4°C for 1 hour. Using a magnetic rack, the supernatant was removed, and the beads were subsequently washed three times with binding buffer. The beads were then added to an elution buffer (50 mM NaPO<sub>4</sub> [pH 8.0], 300 mM NaCl, 0.01% Tween 20, and 300 mM Imidazole) and rotated at 4°C for 45 minutes.

For rKREPA2 purification, TnT reactions post incubation were added to calmodulin resin beads (CRB) in binding buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1%NP40, 2 mM β-ME, 1 mM Mg Acetate, 1 mM imidazole and 2 mM CaCl<sub>2</sub>) and rotated at 4°C for 2 hours. The beads were briefly spun down to remove the supernatant and washed with a binding buffer three times. CRB were then suspended in elution buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1%NP40, 10 mM β-ME, 1 mM Mg-acetate, 1 mM imidazole and 2 mM EGTA) and rotated at 4 °C for 1 hour.

The eluates containing either rKREL1 or rKREPA2 were buffer exchanged in 1xHHE (25 mm HEPES, 10 mm Mg(OAc)<sub>2</sub>, 1 mm EDTA, 50 mm KCl) and further concentrated using 10K Amicon Ultra centrifugal filters (R1CB94236, Sigma Aldrich). Recombinant proteins were examined by western blotting using an anti-His-tag antibody (631212, Clontech) and visualized with ChemiDoc (Bio-Rad) (Figure 7). A His-tagged protein ladder was loaded in increasing amounts of 20 ng, 40 ng, 80 ng and 120 ng to calculate a standard curve for quantifying recombinant proteins.



**Figure 7. Expression of KREL1 and KREPA2 recombinant proteins.**

(A) Western blots using an anti-his-tagged antibody against KREL1 FL, independent domains, CTD truncations, point mutants, and group mutants. (B) SDS-PAGE gels showing expression of the radiolabeled recombinant proteins expressed with S35 methionine and visualized using the PhosphorImager.

### *KREPA2 pull-down assay*

Recombinant proteins were radiolabelled with [<sup>35</sup>S] methionine using the TnT expression system. An aliquot of the reactions was loaded on an SDS-PAGE gel to visualize expression (Figure 7). Equal volumes of rKREL1 TnT reaction and rKREPA2 TnT reaction were mixed on ice for 15 min. KREPA2, with its calmodulin tag, would bind to the CRB and pull down KREL1. The mixed reactions were suspended in binding buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1%NP40, 2 mM β -ME, 1 mM Mg Acetate, 1 mM imidazole and 2 mM CaCl<sub>2</sub>) and rotated at 4°C for 2 hours with CRB. Supernatants were removed, and CRB was subsequently washed three times with a binding buffer. The supernatants were removed using ZEBRA micro spin 10 kD

columns (Cat No: P189879), and the CRB was subsequently washed three more times, each time spun through the column. CRB were then suspended in elution buffer (10 mM Tris-HCl [pH 8.0], 150 mM NaCl, 0.1% NP40, 10 mM  $\beta$ -ME, 1 mM Mg Acetate, 1 mM imidazole and 10 mM EGTA). Eluted proteins were denatured with SDS loading dye, migrated in a 10% SDS-PAGE gel, exposed to an X-ray film, and visualized using a PhosphoImager (Biorad). Analysis was conducted using Quantity-one software (Biorad).

### *Three-step ligation assay*

KREL1 ligation activity was studied *in vitro* using annealed dsRNA. 2  $\mu$ M of 5' labelled CY5 5'RNA fragment (5'- GGAAAGUUGUGACUGA-3') and 2  $\mu$ M of 3'RNA fragment (5'- pUGAGUCCGUGAGGACGAAACAAUAGAUCAAAUGUp-3',) were annealed to a 4  $\mu$ M guide RNA (5'- GUUUUGUUCUUAUGGACUCAUCAGUCAUAAUUUCCUU-3') in 2x HHE buffer.

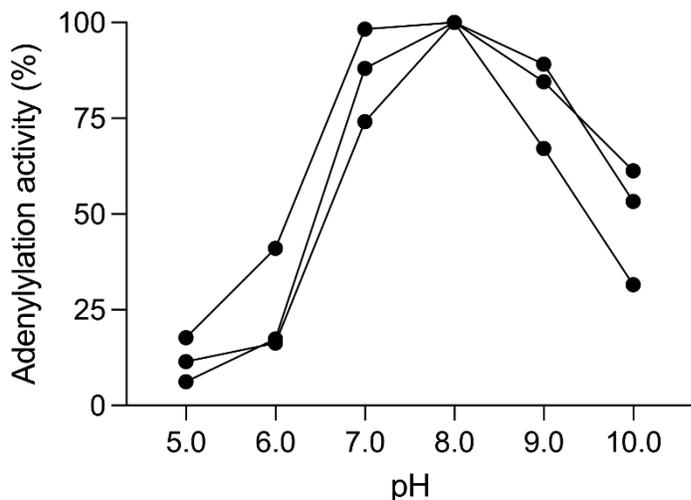
The RNA fragments were placed in a water bath at 70°C and cooled to RT. The dsRNA was then added to a ligation reaction (2x HHE [pH 8.0], CaCl<sub>2</sub>, Tx-100, ATP, RNAase inhibitor) with 1 pmol of rKREL1 +/- rKREPA2 proteins. The reactions were incubated overnight and at various time points, at 28°C in the dark, shaking at 50rpm.

The reactions were quenched by adding 20  $\mu$ M guide RNA competitor (DNA) (5'- AAAAAAAAAAGGAAAATTATGACTGAGTGAGTCCATAAGAACAAAAC-3') and an equal volume of 10M Urea in TBE and heated at 95°C for 90 seconds. The RNA was resolved on a 20% Acrylamide gel (37:1) and visualized at ~650nm for CY5 using a ChemiDoc (Biorad) machine. Analysis was conducted using ImageLab software (Biorad).

### *Step 1, ligase adenylylation assay*

KREL1 auto-adenylylation activity was studied *in vitro* with [ $\alpha$ -<sup>32</sup>P] ATP. Reactions were set up with 0.2 pmol of KREL1 +/- KREPA2 along with 1 $\mu$ Ci [ $\alpha$ -<sup>32</sup>P] ATP in a solution containing 25 mM Tris-HCl [pH 8.0], 10 mM Mg(OAc)<sub>2</sub>, 0.5 mM DTT, 1% BSA and 10% DMSO at RT. Reactions were optimized at pH 8.0 (Figure 8). The reactions were incubated for 10 min and at various time points, at RT. Reactions were quenched with SDS-loading dye, migrated on a 10%

SDS-PAGE gel, exposed to an X-ray film, and visualized using a PhosphorImager (Biorad). Analysis was conducted using Quantity One software (Biorad).



**Figure 8. KREL1 adenylylation activity in pH 5-10.**

Reactions include 25 mM Tris-HCl of pH 5-8 or 25-mM Tris-Acetate pH 9-10.

#### Step 2, RNA adenylylation assay

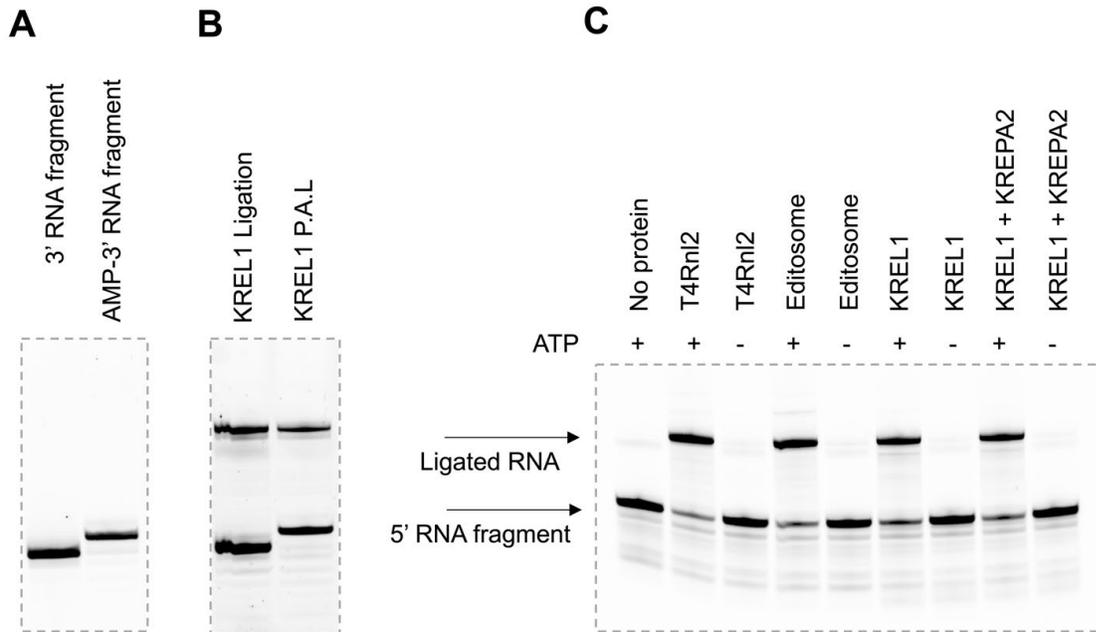
KREL1 de-adenylylation activity was studied *in vitro* in the presence of nicked dsRNA as described in the ligation assay. Adenylylated KREL1 with [ $\alpha$ - $^{32}$ P] ATP, as described in the adenylylation assay, was added to dsRNA mixtures detailed in the three-step ligation assay. The reaction rate was examined in the presence of KREPA2 and compared to the rate of KREL1 de-adenylylation. Reactions were quenched at different time points with the denaturing SDS-loading dye and visualized with methods of all radioactive assays.

#### Step 3, pre-adenylylated ligation assay

KREL1 phosphodiester bond formation activity was studied *in vitro* using a pre-adenylylated dsRNA. For this assay, the 3' RNA fragment was first adenylylated using a 5' DNA adenylylation kit (NEB, catalogue # E2610S). It contains a *Methanobacterium thermoautotrophicum* ligase that adenylylates a nucleic acid strand at the 5' end. The RNA was purified through ethanol precipitation and migrated through an acrylamide gel for verification

using a CY3 label (Figure 9). Next, 2  $\mu\text{M}$  of 5' labelled CY5 RNA fragment and 2  $\mu\text{M}$  of pre-adenylylated 3' RNA fragment were annealed to 4  $\mu\text{M}$  guide RNA in 2x HHE. The RNA fragments were placed in a water bath at 70°C and cooled to RT. The RNA mix was then added to a reaction (2x HHE [pH 8.0],  $\text{CaCl}_2$ , Tx-100, and RNase inhibitor) with 1 pmol of KREL1 +/- KREPA2 protein. It should be noted that this assay lacked the addition of ATP (Figure 9).

The reactions were incubated overnight and at various time points, at 28°C in the dark, while shaking at 50 rpm. The reactions were quenched by adding 20  $\mu\text{M}$  guide RNA competitor (DNA) and an equal volume of 10M Urea in TBE and heated at 95°C for 90 seconds. The RNA was resolved on a 20% Acrylamide gel (37:1) and visualized at ~650 nm for CY5 using a ChemiDoc (Biorad) machine. Analysis was conducted using ImageLab software (Biorad).



**Figure 9. Control experiments for ligation assays.**

(A) 3' RNA fragment migrated alongside the adenylylated 3' RNA fragment in a 20% acrylamide gel. (B) Ligation of unadenylylated RNA with ATP alongside ligation of pre-adenylylated RNA in the pre-adenylylated ligation (P.A.L.) assay in the absence of ATP. (C) Three-step ligation assay of T4 RNA ligase 2 (T4Rnl2), the whole editosome, and KREL1 with and without KREPA2, in the presence and absence of ATP.

## Research findings

### *Structural analysis of KREL1 and KREPA2*

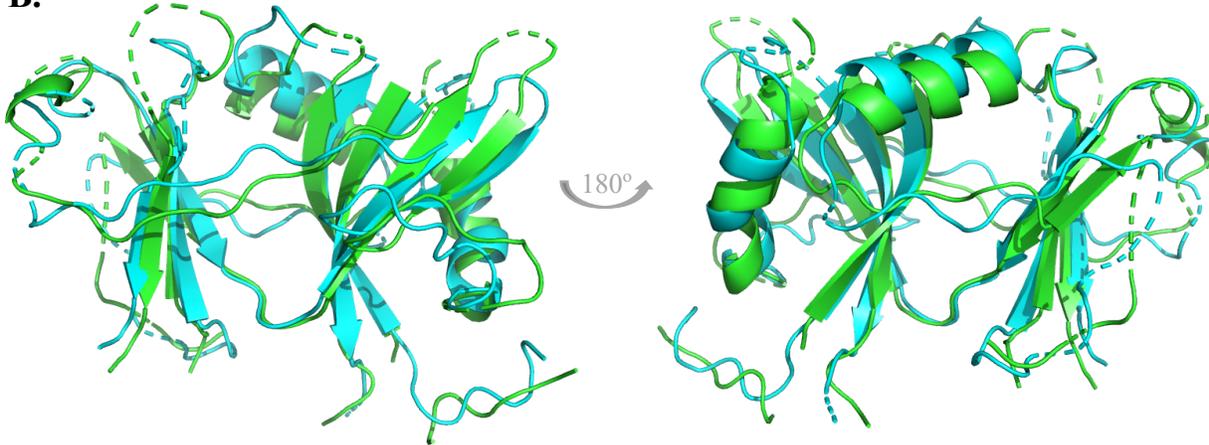
KREL1 has two domains connected by a flexible coil (Figure 5); the NTD (aa 51-324) is the ATP binding pocket, and the CTD (aa 354-469) is made of four alpha-helices. As both domains are essential for the complete ligation mechanism, I performed a structural analysis on each. In this structural analysis of KREL1, both domains are compared to KREL1 of related species, to its paralogous enzyme KREL2, to its closest analogue T4Rnl2, and a significantly diverged Human DNA ligase 1.

The ATP binding pocket at KREL1's N-terminus (PDB: 1XDN) is formed by five beta sheets representing motifs I-V (Deng et al. 2004). Amino acid residues of these motifs form critical interactions with the ATP molecule for proper catalysis (Mehta et al. 2015). Other nucleotide transferases share the five conserved motifs, such as human DNA ligase 1, although they form a significantly different adenylation domain with an RMSD value over 3.0 when superimposed (Figure 10). This contrasts with an RMSD range of 0.2-1.7 with superimpositions of the predicted AlphaFold structure of the NTD from KREL's of related species (*T. cruzi* and *L. major*) (Appendix Table 7). To date, the closest analog of KREL1 is T4Rnl2, a bacteriophage RNA ligase (PDB: 2HVR) (Ho and Shuman 2002), which reveals a higher degree of similarity between 1XDN with an RMSD of 2.1-2.4, and a domain sequence identity of less than 30%. Interestingly, its paralog, KREL2, shares a sequence identity of only 40% (Appendix Table 5). Using PyMol, I examined the superimposition of the ATP pocket of 1XDN and 2HVR that show aligned interactions between residues of motifs I-V and ATP/AMP, with an RMSD of 0.56 (Figure 11). These included E85, R111, F209, R288, K307 and K309 of KREL1, which are conserved in T4Rnl2 previously shown to abolish the ligation activity with alanine substitutions (Yin et al. 2003). One defining difference between structures 2HVR and 1XDN is the absence of a covalent bond between AMP and K87 in KREL1.

**A.**

	I	III	IIIa	IV	V
<b>L1_T_brucei</b>	85 E-KVHGTNF -153-	LVLNGELF -206-	-FFAFD----- -282-	EGVIR -304-	TIIKLR-----
BT7_DNA_Ligase	31 EIKYDGVNR - 88-	MLDGELMV -144-	HFKLVAILPLH -216-	EGLIVK -234-	GWVKMKPENEA--D
Hu4_DNA_Ligase	270 ETKLDGER -326-	LLDGEMMA -362-	CYCVFDVLMVN -426-	EGLMVK -445-	GWLKIKPYYVS--G
Hu3_DNA_Ligase	505 EIKYDGER -554-	MILDSEVL -590-	CLFVFDCTYFN -654-	EGLVIK -671-	HWLKVKKDYLNEGA
Hu1_DNA_Ligase	565 EYKYDGER -615-	FILDTEAV -655-	CLFAFDLIYLN -719-	EGLMVK -740-	NWLKIKKDYLD--G
Sce_DNA_Ligase	416 EYKYDGER -466-	FILDCEAV -506-	CLFAFDILCYN -570-	EGLMVK -592-	NWLKIKKDYLD--G
Spo_DNA_Ligase	413 EYKYDGER -463-	FILDCEAV -503-	CLFAFDLIYLN -567-	EGLMVK -589-	HWLKVKKDYLS--G
Eco_DNA_Ligase	112 ELKLDGLA -167-	LEVRGEVF -220-	TFFCYGVGVLE -284-	EGVVIK -306-	PRWAVAKKFPAQE-

**B.**



**Figure 10. KREL1 shares a conserved adenylation domain with Human DNA ligase.**

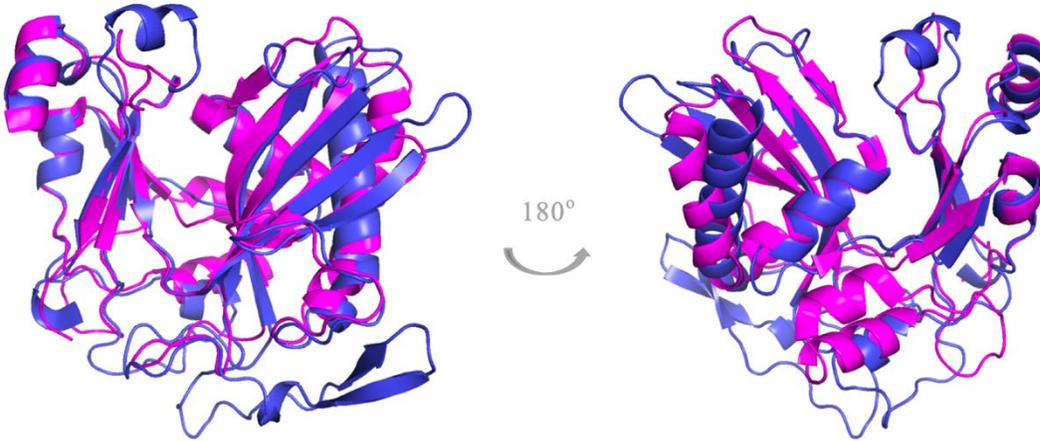
(A) Multiple sequence alignment between KREL1 of *T. brucei* with human DNA ligases 1-4 as well as other DNA ligases from *S. cerevisiae* (Sce), *Schizosaccharomyces pombe* (Spo) and *E. coli* (Eco). Alignment adapted from Shuman, 1999. (B) Pairwise structural alignment using RCSB PDB of Human DNA ligase 1 (1X9N) in green and KREL1 NTD (1XDN in cyan. RMSD of 3.07, TM-score of 0.23 and coverage of 1XDN of 62%.

The CTD of KREL1 is a four-alpha helix domain that is structurally similar to T4Rnl2, although T4Rnl2 does not retain these essential roles (Figure 11) (Ho et al. 2004). The CTD of KREL1 was not successful in being crystallized, and its structure prediction is entirely based on T4Rnl2, resulting in an RMSD score equal to that of KREL1 between related species (Appendix Table 7). This is likely inaccurate when comparing percent sequence identities (%SID) of 25% to 40-80%, respectively (Appendix Table 6). However, residues essential for the structure of T4Rnl2, such as R266 and D292 (Nandakumar et al. 2004), are conserved in KREL1 (R372/D402). These residues interact between overlapping alpha-helices ( $\alpha$ H2 and  $\alpha$ H3), and preliminary data suggests they play a role in the integrity of the domain (Figure 12). In summary, while the CTD of T4Rnl2 and KREL1 have functionally diverged, their structures appear very similar when using Alphafold predictions.

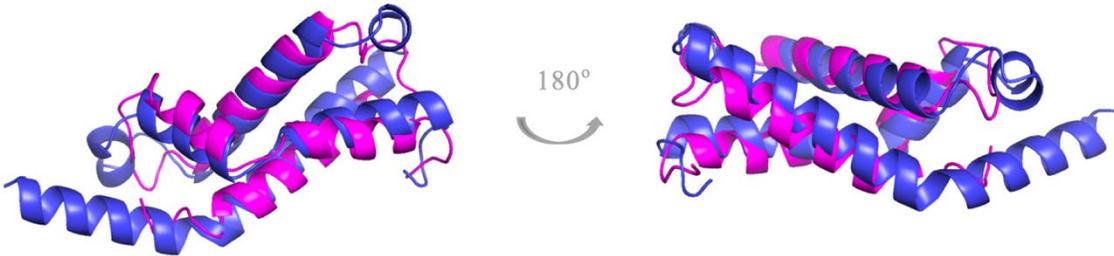
**A.**

		I		III		IIIa		IV		V
L1 <i>T. brucei</i>	85	-EKVHGTFN	-153-	LVINGELF	-206-	FFAFD	-282-	EGVVR	-304-	TIIKLR
L1 <i>T. cruzi</i>	81	-EKVHGTFN	-147-	VVINGELF	-200-	FFAFD	-276-	EGVVR	-298-	TIIKLR
L1 <i>L. major</i>	108	-EKVHGTFN	-172-	VVINGELF	-225-	YFAFD	-301-	EGVVR	-323-	TIIKLR
L2 <i>T. brucei</i>	57	-EKVHGTFN	-120-	VVINGELF	-170-	FYAFD	-245-	EGIVMK	-267-	TIIKFK
KOD1Rn11	91	-EKVDGYNF	-141-	IVLVGEMA	-168-	FFLFD	-232-	EGIVMK	-245-	-IAKYV
T4Rn12	29	-EKVHGTFN	- 93-	YQVFGETA	-115-	FYVFD	-191-	EGVVR	-213-	VAIKCK

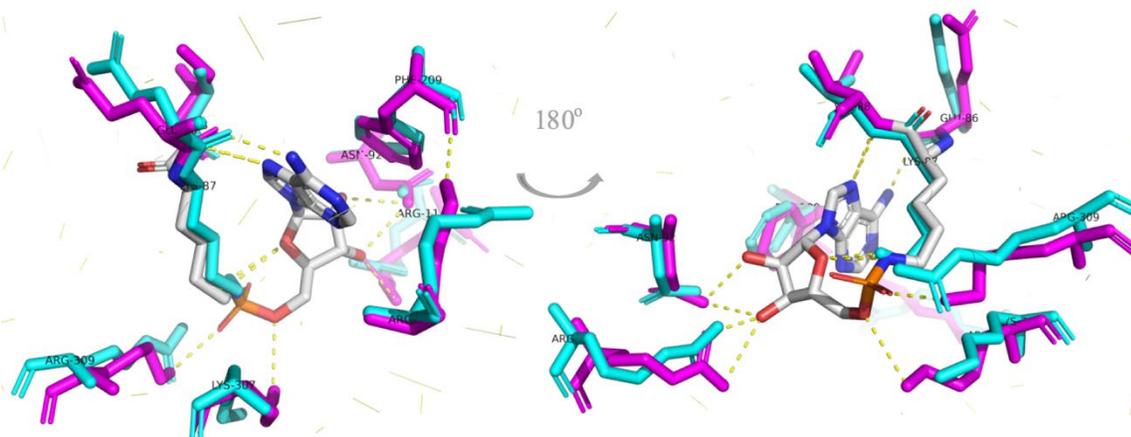
**B.**



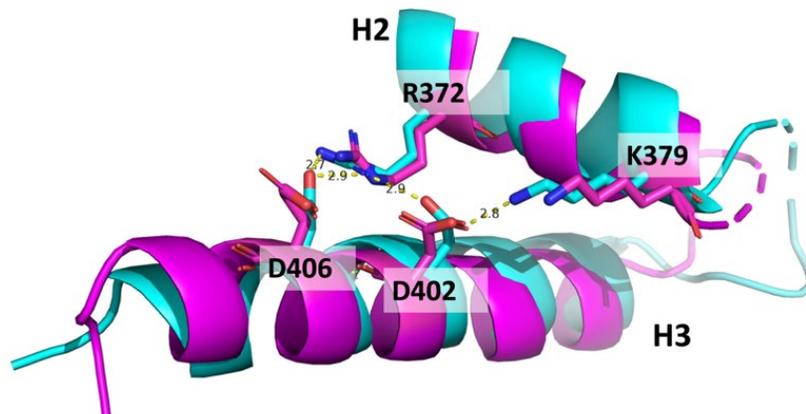
**C.**



**D.**



**Figure 11. Sequence alignment of RNA ligases conserved motifs of the catalytic ATP binding domains**  
 (A) Multiple sequence alignment of the ATP catalytic site motifs between KREL1 and KREL2 of *T. brucei*, *T. cruzi* and *L. major* against an RNA ligase from the archaeon, *Thermococcus Kodakarensis*, KOD1Rnl1 and a bacterial ligase T4Rnl2. (B) Pairwise structural alignment using RCSB PDB of T4 Rnl2 (2HVR in magenta) and KREL1 NTD (1XDN in mauve). RMSD of 3.05, TM-score of 0.6 and coverage of 1XDN of 82%. (C) Pairwise structural alignment using RCSB PDB of T4 Rnl2 (2HVR in magenta) and KREL1 CTD predicted through Alphafold colab server (mauve). Pairwise structural alignment reveals an RMSD of 2.71, TM-score of 0.61 and coverage of KREL1 of 73%. (D) Superimposition of residues in the binding pocket of T4Rnl2 in magenta and KREL1 in cyan with calculated interactions using PyMol. Distance ranges from 1.2-2.2 and RMSD of 0.56.



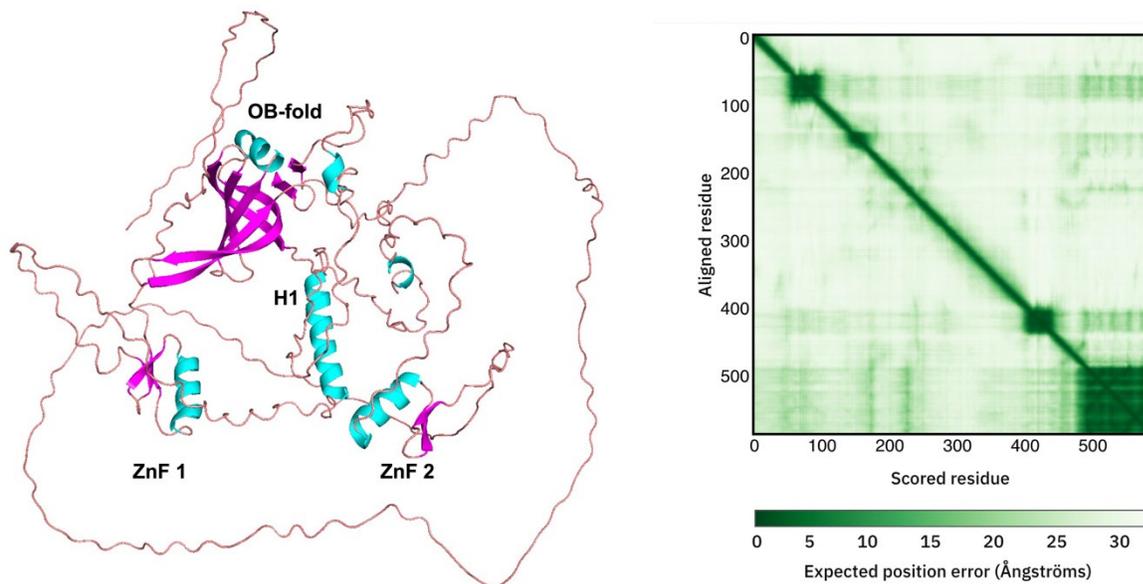
**Figure 12.  $\alpha$ H2- $\alpha$ H3 superimposition of KREL1 and T4Rnl2.**

Alignment of KREL1 (cyan) and T4Rnl2 (magenta) highlighting interacting residues R372/D406 and K379/D402 of KREL1 conserved with R266/D292 and K303/D288 of T4Rnl2.

Kinetoplastid RNA editing protein KREPA2 is among six Group A proteins of the RECC. They can be further categorized based on their domains; for example, KREPA4-6 possesses the OB-fold domain, while KREPA1-3 has the OB-fold domain and two C2H2 zinc finger domains. In my analysis of the Alphafold-predicted structures of the KREPA proteins, it was apparent that KREPA 1-2 shared an alpha helix structure between the two ZnFs, absent in KREPA3 (Figure 13). My PPI studies revealed that the  $\alpha$ Helix 1 of KREPA2 is the KREL1 binding site and therefore possibly the site of interaction between KREL2-KREPA1.

The OB-fold of the KREPA proteins is structurally conserved with five beta strands forming a barrel, with a short alpha helix superior to it, termed a “winged helix” (Fanning et al. 2006). Based on amino acid sequence comparison, they differ by ~80% while their predicted structures share an RMSD of 3.0-6.0, a similar RMSD of homologous KREPA proteins of related species (Appendix Table 9, Table 10). As previously discussed, all KREPA proteins, except for

KREPA5, have been experimentally shown to be essential for the integrity of the RECC and RNA editing due to the deletion of their OB-fold domain (Aphasizheva et al. 2020). KREPA1-3 harbours two ZnFs with a critical characteristic of C2H2 residues that are essential for its structure (Figure 6) (Guo et al. 2008, 2010).



**Figure 13. The predicted structure of KREPA2 includes an OB-fold domain and two zinc finger domains.**

*Alphafold2 w/mmseqs2 Google colab predicted the structure of RNA editing protein KREPA2 with an associated heat map.*

#### *KREL1 and KREPA2 protein-protein interaction studies*

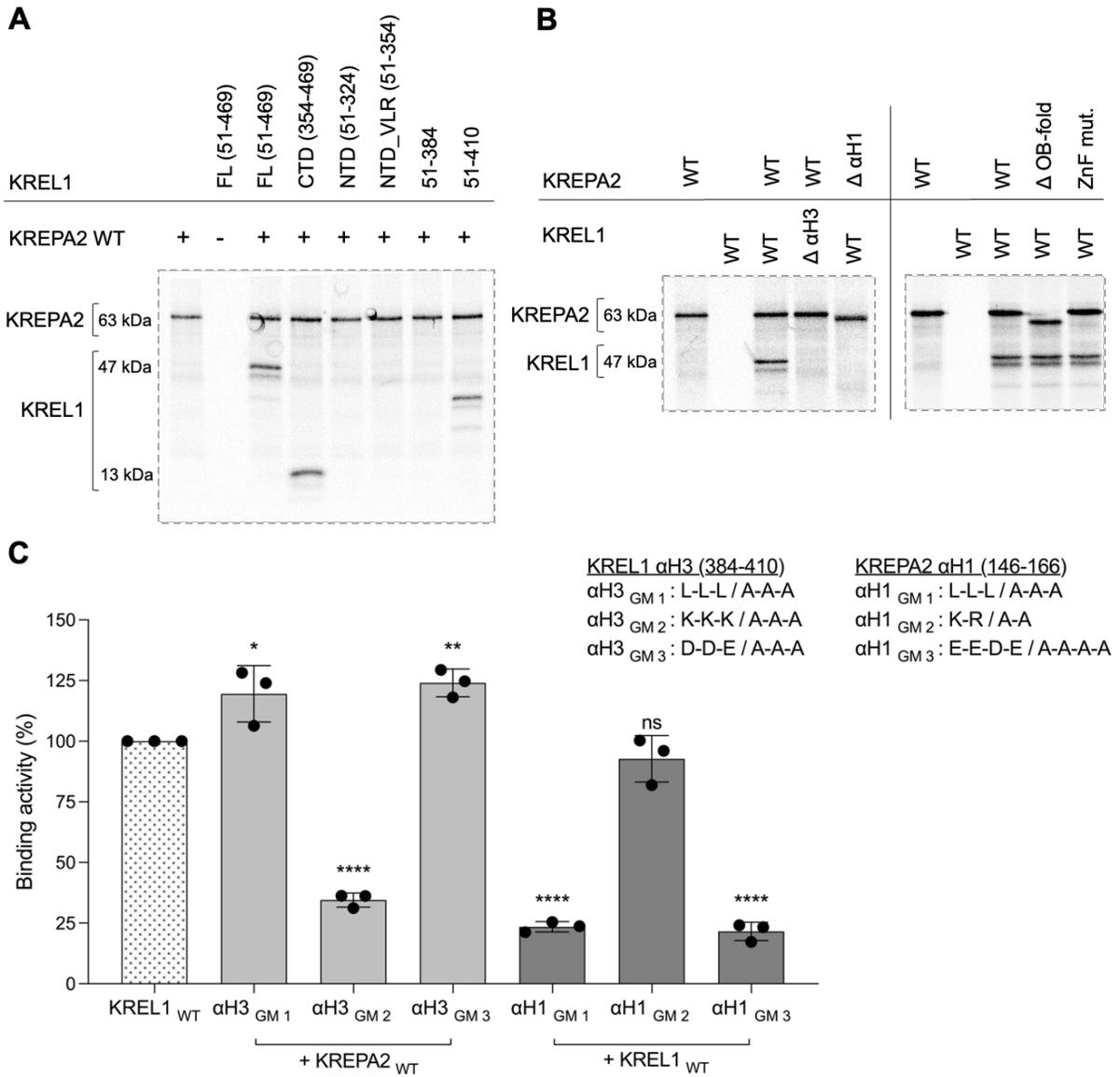
To narrow the region on KREL1 that interacts with KREPA2, we generated a panel of recombinant KREL1 proteins - full-length and truncated (Figure 5) and performed pulldown experiments with recombinant KREPA2. We expressed the NTD and CTD of KREL1 independently and showed that its CTD (aa 354-469) could interact with KREPA2 (Figure 14) as anticipated from published data (Schnauffer et al. 2010; Mehta et al. 2015). In fact, KREPA2 interaction was completely absent with KREL1 CTD truncates: NTD (aa 51-324) and NTD\_VLR (NTD with the variable loop; aa 51-354). We then expressed a series of C-terminal truncations of KREL1 dictated by recognizable conserved regions that structurally resolve into the four  $\alpha$ -helices. In the KREPA2 pulldown assay, we noticed a loss of binding with the KREL1 384 truncation (aa 51-384) that was subsequently present with the KREL1 410 truncation (aa 51-410), suggesting

that the region of KREPA2 contact is between residues 384-410 of KREL1, with the  $\alpha$ H3 of its CTD. Hence, we next confirmed the loss of KREPA2 binding with an  $\alpha$ H3 internal deletion variant of KREL1 ( $\Delta$   $\alpha$ H3).

Next, to identify the region on KREPA2 that interacts with KREL1, several versions of rKREPA2 were generated corresponding to the full-length protein, with internal truncations of ( $\Delta$ ) $\alpha$ H1 or the ( $\Delta$ )OB-fold domain, and mutated ZnFs (Figure 5). This experiment shows that KREPA2  $\Delta$  OB-fold (aa 1-484) or KREPA2 with cysteine to alanine substitutions in the ZnF domains, have no impact on KREL1 binding activity (Figure 14), ruling out these domains in mediating protein-protein interactions with KREL1. Previous yeast-2 hybrid data indicate the binding region on KREPA2 for KREL1 to be between residues 121-192 (Schnauffer et al. 2010). This region contains a predicted  $\alpha$ H1 from conserved residues 146-166 (Figure 6). A pull-down test confirmed this region is the KREL1 binding site, as rKREPA2 with an internal deletion of this  $\alpha$ H1 ( $\Delta$   $\alpha$ H1) abolished the interaction.

The interaction between the  $\alpha$ H3 of KREL1 CTD and  $\alpha$ H1 of KREPA2 was further investigated through substitution mutation analyses where the hydrophobic, positively charged, and negatively charged residues on both helices were substituted for alanine as separate group mutations. Group mutations of the positively charged lysine residues on KREL1 CTD  $\alpha$ H3 led to a  $\sim$ 70% decrease in the binding efficiency to KREPA2 (Figure 14). Similarly, group mutations of negatively charged glutamic and aspartic acid residues of KREPA2  $\alpha$ H1 led to a  $\sim$ 75% decrease in binding efficiency, revealing an electrostatic nature of KREL1-KREPA2 interaction. Mutations of the hydrophobic leucine residues of KREPA2  $\alpha$ H1 also resulted in a similar effect, although the latter was not seen in  $\alpha$ H3 of KREL1 CTD. This suggests hydrophobic interactions with adjacent helices of KREL1 CTD for further stability, but not directly with  $\alpha$ H3.

In summary, we located the region of contact between KREL1 and KREPA2 between  $\alpha$ H3 of KREL1 CTD and  $\alpha$ H1 of KREPA2. The  $\alpha$ H3 region of KREL1 is defined as a LAKD repeat motif that is completely absent in T4Rn12, but highly conserved between trypanosomatid KRELS (Appendix Figure 24). Mutational analyses further confirm this to be driven through electrostatic interactions; the lysine residues of KREL1  $\alpha$ H3 form potential interactions with the glutamic and aspartic acid residues of KREPA2  $\alpha$ H1.

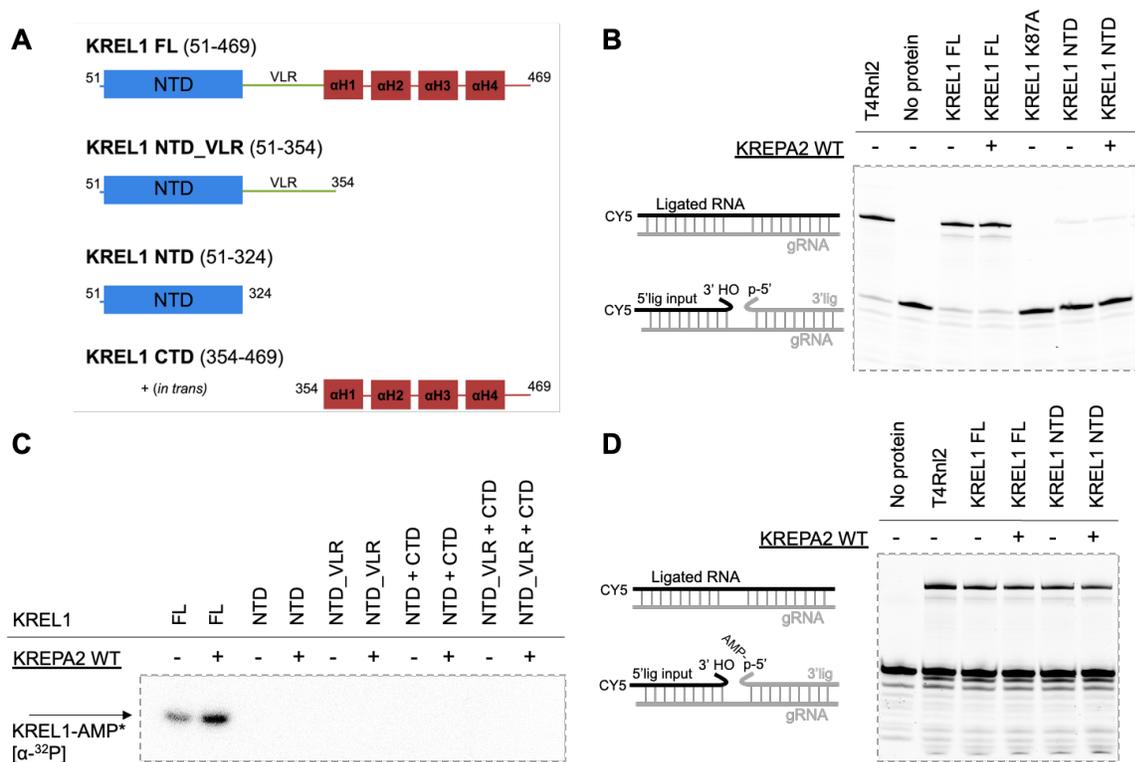


**Figure 14. Pulldown assay of KREL1 with KREPA2.**

(A) Pull-down of KREPA2 wild-type (WT) with KREL1 full-length (FL), N and C-terminal domains (NTD and CTD), and CTD truncations. (B) Pull-down of KREL1 WT and KREL1  $\Delta\alpha$ H3 of the CTD with KREPA2 WT and domain mutants; deletion of  $\alpha$ H1 ( $\Delta\alpha$ H1), deletion of OB-fold ( $\Delta$ OB-fold) and ZnF mutant which substitutes four cysteine residues to alanine (ZnFC/A). (C) Pull-down assay of group mutations (GM) of  $\alpha$ H3 of KREL1 and  $\alpha$ H1 of KREPA2. GMs are detailed next to the graph, including alanine substitutions of positively charged residues, negatively charged residues, and hydrophobic residues on each helix. Error bars represent SD obtained from three replicate experiments. \*\*\*\* $p < 0.0001$ , \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .

### *Functional role of KREL1 domains in ligation*

To investigate the essentiality of KREL1's CTD in ligation, functional activities of KREL1 full-length (FL) and the truncated NTD (aa 51-324) were compared *in vitro*. The deletion of the CTD renders KREL1 inactive in the 3-step ligation assay as well as in step 1 auto-adenylation assay (Figure 15), as observed previously (Deng et al. 2004). While the addition of KREPA2 enhances KREL1 FL activity, as shown earlier (Mehta et al. 2015), it does not salvage the lost activity of the KREL1 NTD construct, potentially due to the loss of its binding site. It was not possible to experimentally test for KREL1 step 2 activity (RNA adenylation) when the step 1 activity is abolished as a KREL1-AMP intermediate is required. However, step 3 (phospho-diester bond formation) activity can still be assessed with pre-adenylated RNA substrates. In the pre-adenylated ligation assay, KREL1's NTD (aa 51-324) is equally active as the KREL1 FL, suggesting that the CTD domain is not required for phosphor-diester formation. Point mutations of amino residues in the ATP binding pocket, including the conserved K87 of motif I, showed the essentiality of the residues from motifs I-V for steps 1 and 3 (Figure 16), explaining that the ligase-ATP/AMP interaction occurs throughout the three-step ligation mechanism. We also attempted salvaging KREL1 NTD's repressed auto-adenylation activity by providing the purified CTD *in trans*, with and without KREPA2, however, no improvement was observed. In summary, we show that the CTD of KREL1 is required for ligation activity, specifically for step 1. We also conclude that the NTD of KREL1 alone is required for the last step, by providing the AMP binding pocket interactions to catalyze the phosphor-diester bond formation.



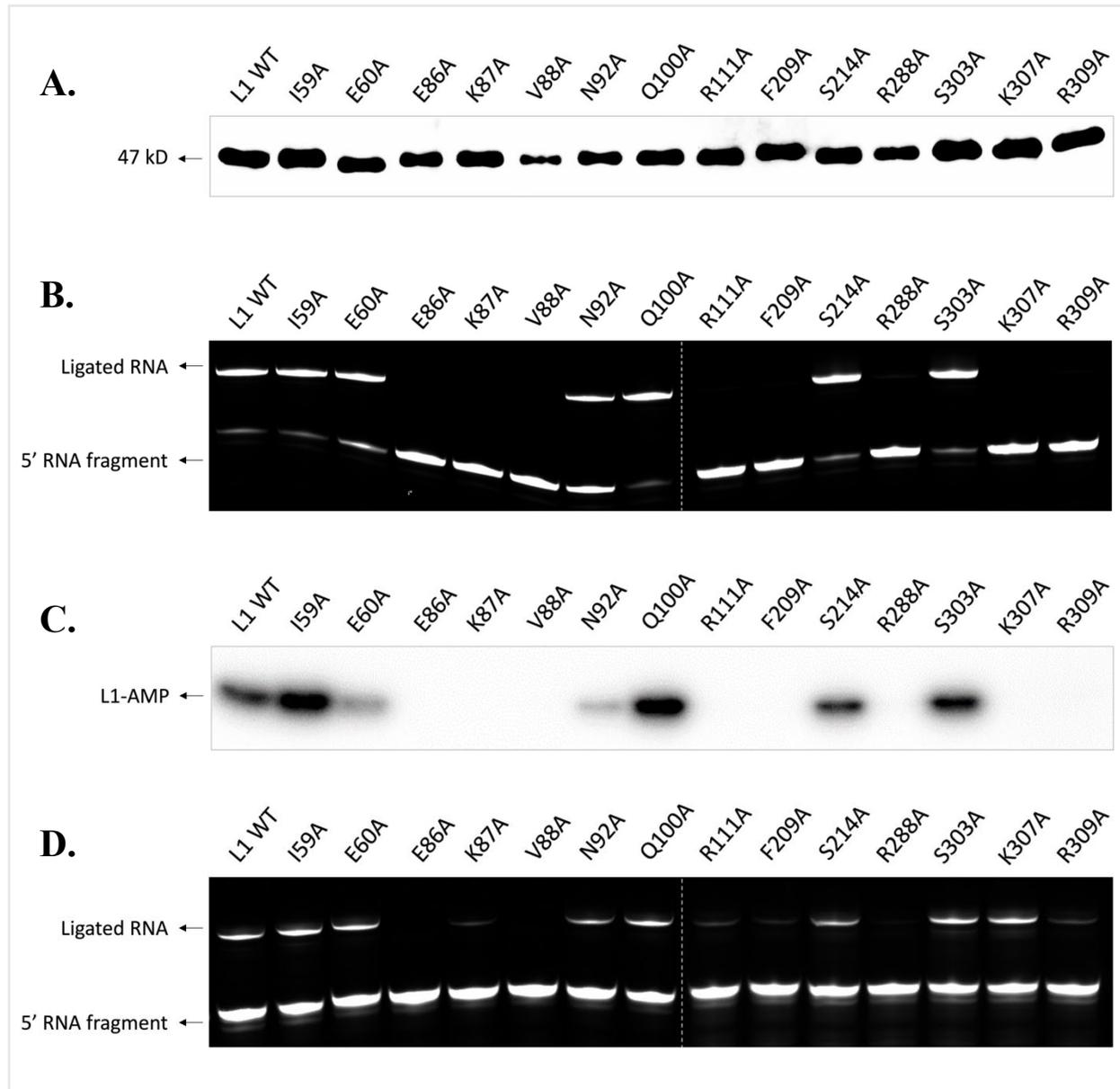
**Figure 15. *In vitro* functional assays of KREL1 full-length and domain truncated mutants with and without KREPA2.**

(A) Structural representation of the KREL1 constructs used including full-length, NTD, NTD\_VLR and the CTD provided in trans. (B) Three-step ligation assay of KREL1 FL vs. NTD, with and without KREPA2 WT. T4Rnl2 was used as a positive control, and KREL1 K87A was used as a negative control. (C) Ligase auto-adenylation assay, representative of step 1 of ligation, of KREL1 FL and domain, truncated mutants with and without KREPA2. A "+" symbol represents reactions with domains provided in trans. The KREL1-AMP product is radiolabeled with  $\alpha$ -<sup>32</sup>P represented by an asterisk. (D) Pre-adenylylated ligation assay, representative of step 3 of ligation, of KREL1 FL vs. NTD with and without KREPA2.

#### Mutational analysis of KREL1 NTD and VLR

A series of point mutations were made at the N-terminal domain to further investigate critical residues for ligation. Alanine substitutions within the ATP binding pocket belonging to motifs I-V, as well as control mutations, Q300A, S309A, and S303A, outside the pocket were recombinantly expressed and tested *in vitro* in the four functional assays (Figure 16). K87, of motif I, forms a covalent bond with the  $\alpha$ -phosphate of AMP, indicative of the loss of function with the K87A mutant. Mutants E60A, V88A, R111A, F209A, R288A, K307A, and R309A of the ATP binding pocket, previously shown to abolish activity in step 1 (adenylation), remain essential for

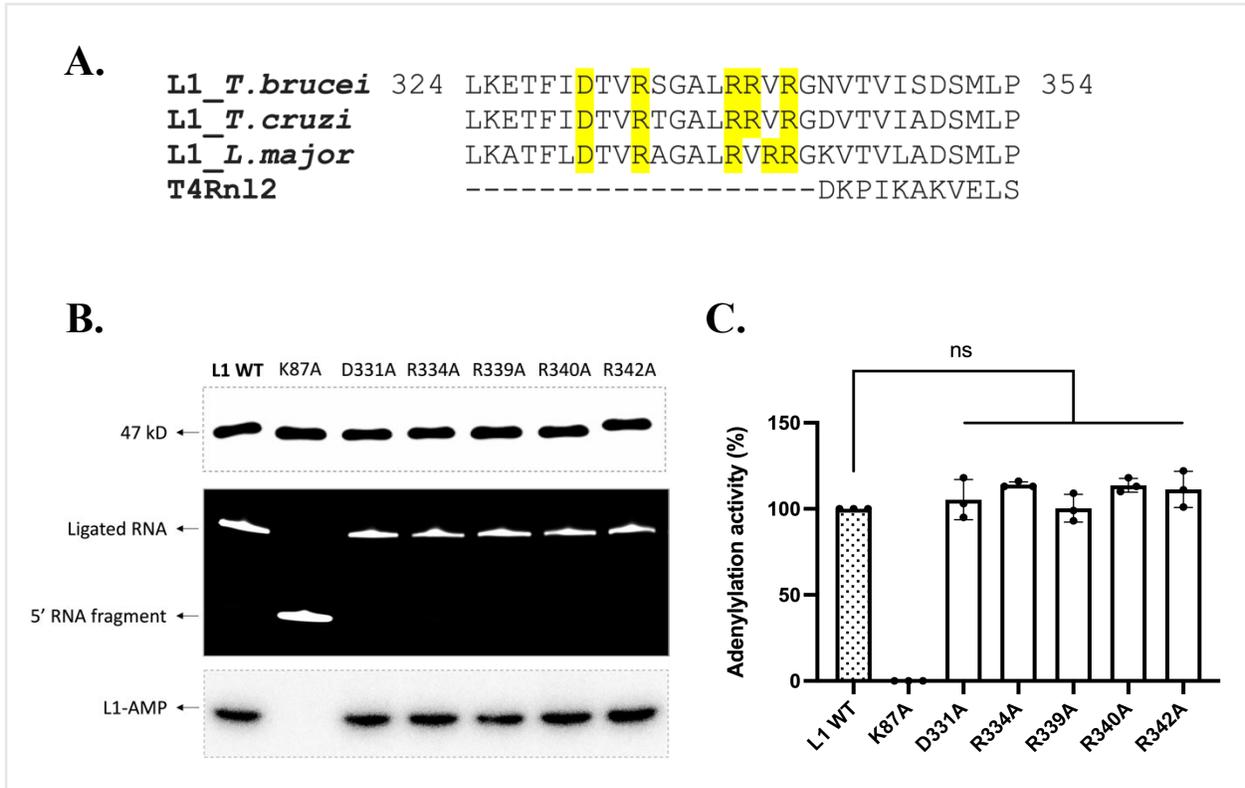
step 3, suggesting that interactions with the molecule are critical throughout the entire ligation mechanism (Figure 22). K87 is also required for step 3, even though the AMP molecule is covalently linked to the RNA substrate. Control mutants Q300A, S309A, and S303A located outside the binding pocket were active in all the *in vitro* functional assays.



**Figure 16. KREL1 ATP binding pocket mutants.**

(A) Western blot of recombinant protein expression of KREL1 WT and adenylation domain point mutants. (B) Three-step ligation assay of KREL1 WT and point mutants. Reaction separated through 20% acrylamide gel, visualized using chemidoc. (C) Radiolabelled P32 adenylation assay of KREL1 WT and point mutants. Protein separated on SDS-PAGE gel and visualized using PhosphorImager. (D) Pre-adenylylated ligation assay of KREL1 WT and point mutants. Reaction separated through 20% acrylamide gel, visualized using chemidoc.

Single point mutations were also made along the variable loop region of KREL1. The 30-residue coil connects the NTD and CTD according to the Alphafold predicted structures. All five mutants; D331A, R334A, R339A, R340A, and R342A have no significant effect on KREL1 activity.



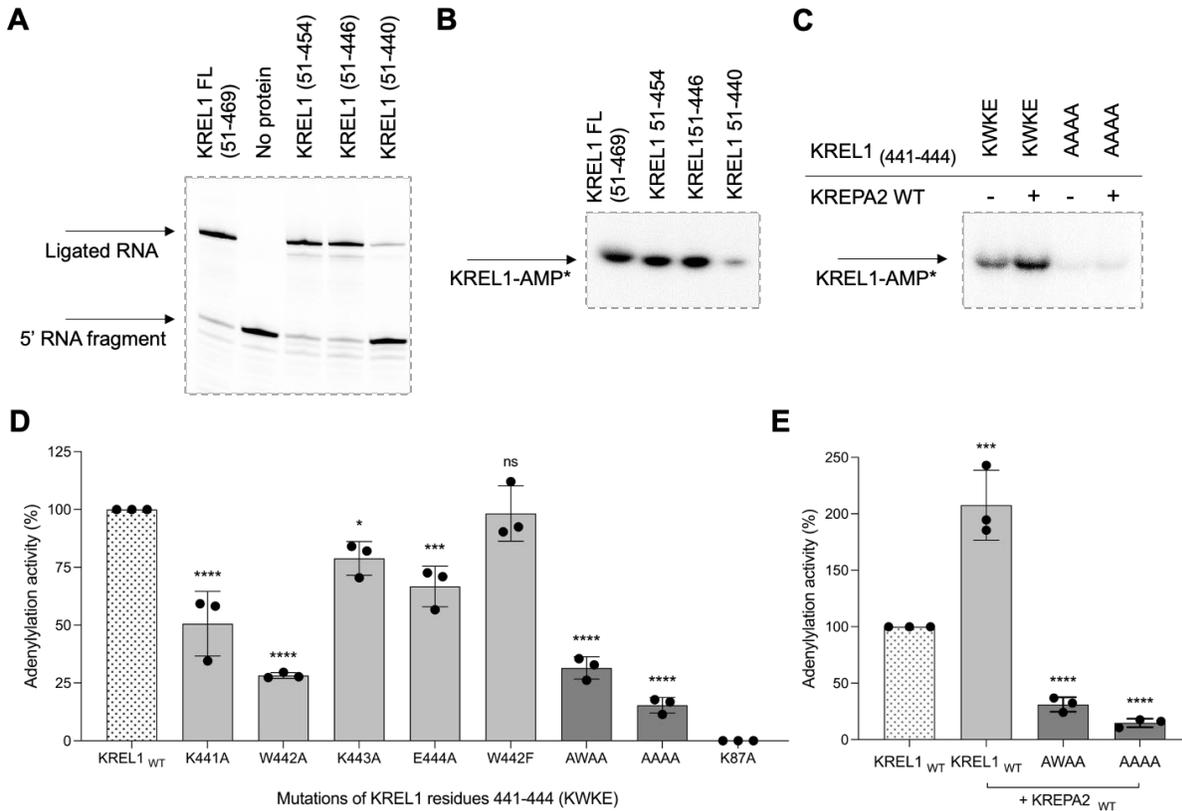
**Figure 17. KREL1 Variable loop region**

(A) Clustal Omega sequence alignment of the variable loop region of KREL1 of kinetoplastids with the loop region of T4Rnl2. (B) Functional activity of point mutants of the variable loop region. The panel includes expression, followed by the ligation activity, followed by the adenylylation activity. (C) Graphical representation of the adenylylation activity of the VLR mutants.

### *Location motif VI on KREL1's CTD*

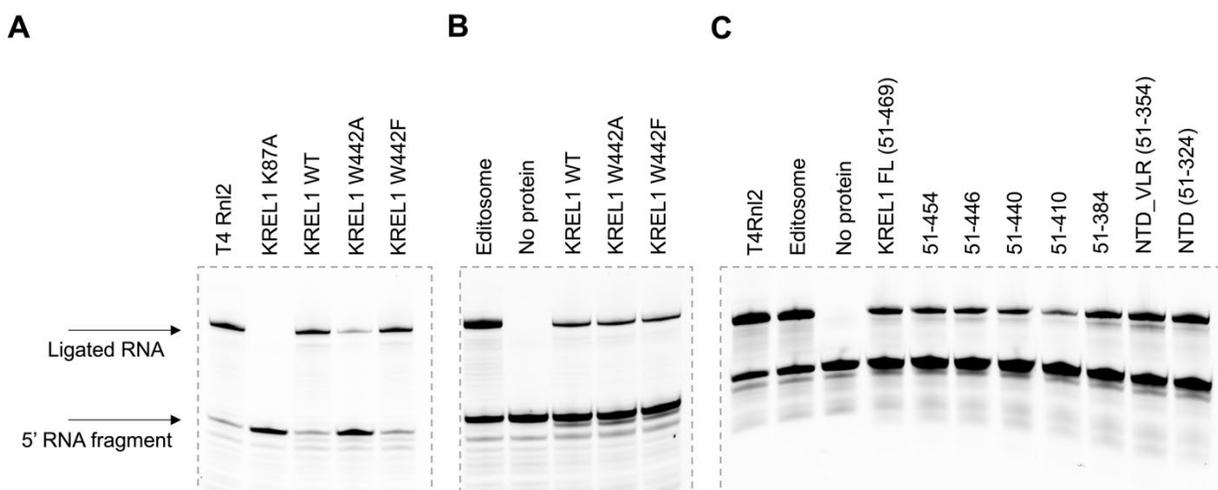
As KREL1 NTD (aa 51-324) is inactive in adenylylation activity, we aimed to narrow the CTD region responsible for its essentiality in ATP hydrolysis. First, we tested the activity of the rKREL1 CTD truncations in our *in vitro* functional assays (Figure 18). While the truncations at residues 454 and 446 were functional in ligation and auto-adenylylation activities *in vitro*, truncation at residue 440 significantly impacted both processes. These truncations were also tested in the pre-adenylylated ligation assay, for which all truncated mutants were active, confirming the specific role of the CTD for step 1 (Figure 19). The region between residues 440-446 is located on  $\alpha$ H4 of KREL1 CTD. A multiple sequence alignment of this region highlights four highly conserved residues, KWKE (441-444) between KREL1 and KREL2 of the kinetoplastid parasites. Single point mutations K441A, W442A, K443A and K444A individually result in a reduction of KREL1-AMP formation by ~50%, ~75%, ~25% and ~30%, respectively. The 75% reduction in activity from W442A is salvageable with a synonymous mutation W442F, suggesting a primary role for W442 in pi-stacking interactions. Although point mutations partially affect KREL1 auto-adenylylation, group mutations of KWKE to AAAA and AWAA almost completely abolished activity. Moreover, we tested the KWKE group mutants in the presence of KREPA2, which was unable to rescue the loss of activity of these mutants.

These data narrow the region of KREL1 CTD important for ligase auto-adenylylation at the conserved KWKE motif. The essentiality of the KWKE residues in ATP hydrolysis is synonymous with the findings of mutational studies implemented in locating motif VI (RxDK) on DNA ligases (Sriskanda and Shuman 1998), suggesting the role of the CTD of KREL1 in providing a diverged motif VI-like region during the adenylylation step.



**Figure 18. Locating motif VI on KREL1 C-terminal domain.**

(A) Three-step ligation assay with KREL1 FL and CTD truncations. (B) Adenylylation assay with KREL1 FL and CTD truncations. (C) Adenylylation assay of KREL1 WT and group mutants (alanine substitutions for residues 441-444, KWKE) with and without the addition of KREPA2. (D) Adenylylation of point mutants and group mutants of KREL1, compared to adenylylation activity of KREL1 WT. Group mutants (GM) include KWKE substituted for AWAA and AAAA. (E) Adenylylation assay of KREL1 WT and group mutants with the addition of KREPA2 WT. Error bars on each graph represent SD obtained from three replicate experiments. \*\*\*\*  $p < 0.0001$ , \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .



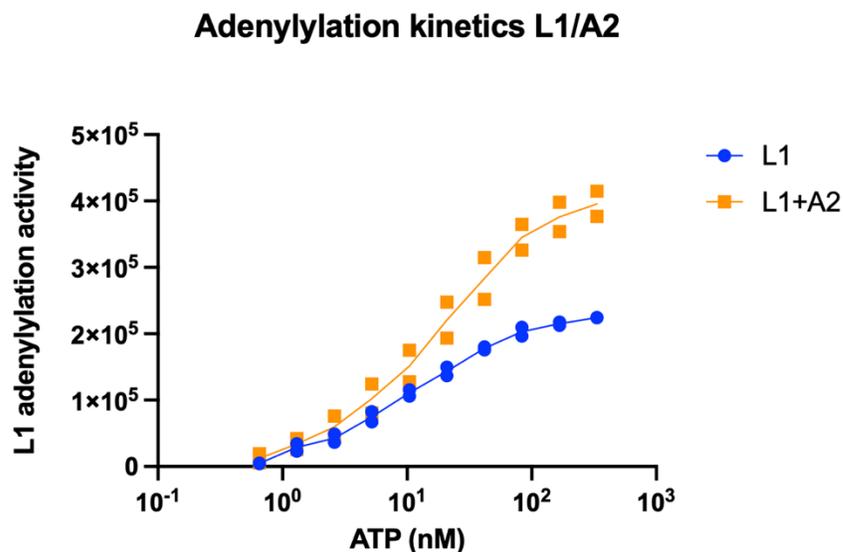
**Figure 19. Activity of motif VI mutants in ligation and pre-adenylylated ligation assays.**

(A) Three-step ligation assay of T4Rnl2, KREL1 WT and W442 mutants. (B) Pre-adenylylated ligation assay of the editosome, KREL1 WT and W442 mutants. (C) Pre-adenylylated ligation assay of the KREL1 FL (51-469) and CTD truncations.

#### *Effect of KREPA2 interaction on KREL1 ligation activity*

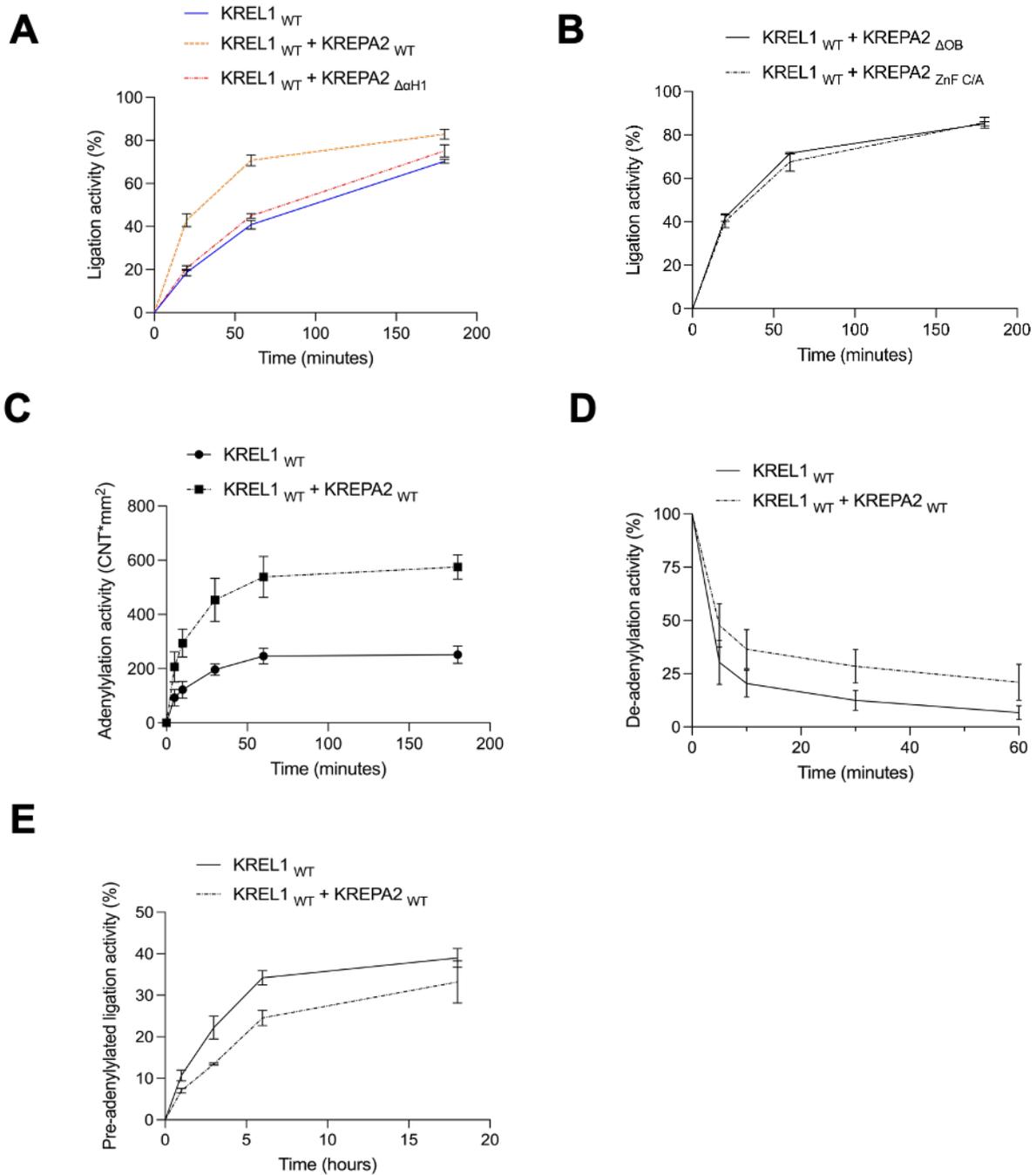
While rKREL1 alone is sufficient for ligation activity *in vitro*, we examined the effect of KREPA2 on KREL1 activity to determine the functional impact of its interaction during the ligation mechanism. We performed several time-point experiments to measure the rate of KREL1 activity at each step of ligation and compared it with the addition of its interacting partner KREPA2 (Figure 21). In the 3-step ligation mechanism, KREL1 only ligates 40% of the input RNA in the first hour. Upon the addition of KREPA2, efficiency increases by 30%. Using the KREPA2  $\Delta\alpha$ H1 mutant, we showed that the deletion of the KREL1 binding site on KREPA2 results in baseline ligation efficiency, abolishing the enhancement effect. However, the deletion of the OB-fold domain or alanine substitutions of the cysteine residues in the ZnF domains on KREPA2 does not affect the increase in the rate of KREL1 ligation activity. A similarly designed experiment used radiolabelled ATP [ $\alpha$ - $^{32}$ P] to measure KREL1-AMP formation in the step 1 ligase auto-adenylylation assay. In the presence of KREPA2, there was an increase in KREL1 adenylylation activity and the overall yield of this ligase intermediate; exhibited by the increased maximal binding capacity ( $B_{max}$ ) of the curve, with no improvement in the apparent dissociation constant,

K<sub>d</sub> (Figure 20). With K<sub>d</sub> ranges of ~11-13nM for KREL1 and ~13-17nM in the presence of KREPA2, the affinity for ATP is assumed not to be significantly improved. Step 2, ligase de-adenylation (RNA adenylylation) activity, was assessed by observing the decrease in the radiolabeled KREL1-AMP in the presence of ligatable RNA substrates. Although no significant change in the rate of ligase de-adenylation was observed with the addition of KREPA2, represented by the curve slope, an increased percentage of KREL1-AMP intermediate remained in the presence of KREPA2 (Figure 21). In the isolated step 3 ligase activity assay (pre-adenylated ligation), the ligation rate was slower in the presence of KREPA2 *in vitro*; suggesting either the need for KREPA2 to dissociate from KREL1 in the final step (3) of ligation or this is a rate-limiting step specific to the KRELS assuming KREPA2 dissociation does not occur. In summary, the interaction between KREL1 and KREPA2 increases the efficiency of KREL1 ligation, specifically for auto-adenylation. We also show that the OB-fold and ZnF domains of KREPA2 do not directly impact the RNA ligation mechanism of KREL1 *in vitro*.



**Figure 20. KREL1 adenylylation kinetics**

*KREL1 adenylylation activity with increasing ATP concentration. Reactions were quenched after 10 minutes, separated through SDS-PAGE, visualized through PhosphorImager and quantified through Chemidoc.*



**Figure 21. *In vitro* functional assays of KREL1 with and without KREPA2 over time.**

(A) Three-step ligation activity over time of KREL1 WT with and without the addition of KREPA2 WT and KREPA2  $\Delta\alpha H1$  mutant. (B) Two other conditions, KREL1 WT with KREPA2  $\Delta OB$  mutant and KREPA2 ZnF mutant (alanine substitutions of cysteine residues), were extracted from the graph of panel A for easier visualization. (C) KREL1 adenylylation activity over time with and without the addition of KREPA2 WT. (D) KREL1 de-adenylylation activity over time with and without the addition of KREPA2 WT. (E) KREL1 pre-adenylylated ligation activity over time with and without the addition of KREPA2 WT. Error bars on each graph represent SD obtained from three replicate experiments.

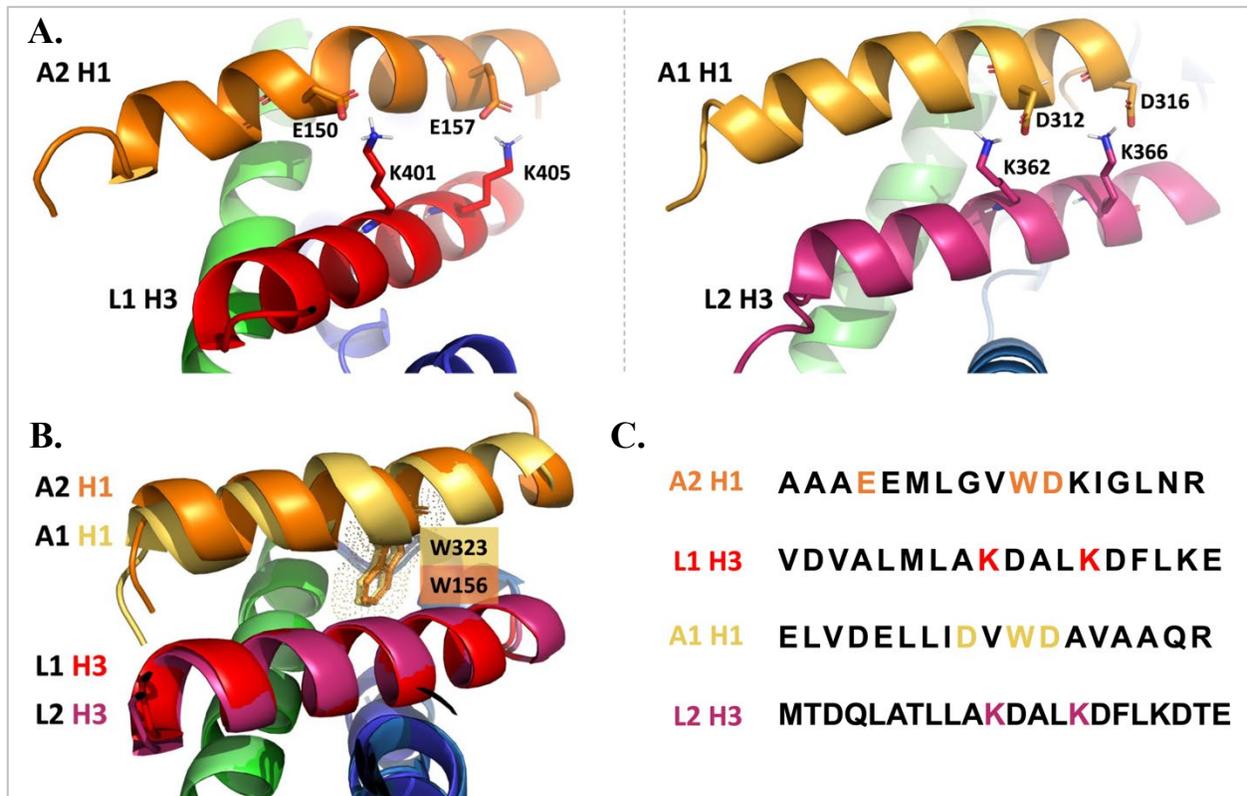
## Comprehensive discussion

KREL1 and KREPA2 are two proteins of the RNA editing catalytic complex, essential for dsRNA ligation post-U-insertion/deletion. These interacting proteins share a three-step ligation mechanism with other widely studied nucleotidyl transferases, including DNA ligases and mRNA capping enzymes. However, KREL1 and KREPA2 have structurally diverged, and their domains' specific roles have remained unknown. My thesis characterizes the interaction and function of KREL1 and KREPA2, to propose novel insights into their ligation mechanism. This work is comprised of a structural, functional, and mutational analysis of these proteins.

Our analysis of KREL1 and KREPA2 initiated from sequence alignments and predicted structures from the Alphafold Colab database (Jumper et al. 2021; Varadi et al. 2022) (Figure 5, Figure 6, Figure 24). The full-length KREL1 protein, without its mitochondrial import signal, harbours a ~34 kDa ATP-hydrolysis N-terminal domain (51-324). The ATP-binding pocket is composed of Motifs (I, III, IIIa, IV, V) that are conserved in the family of nucleotidyl transferases (Doherty and Suh 2000). The multiple sequence alignment of the KRELS from the three major kinetoplastid parasites and their closest known ortholog, bacteriophage T4 RNA ligase 2 (T4Rnl2), highlight these conserved motifs (Ho and Shuman 2002). As the predicted KREL1 structure shows, the NTD is linked to a ~13 kDa C-terminal domain (CTD) through a variable loop region (VLR) that likely provides flexibility in engaging/disengaging the interactions between the two domains supported by our inability to obtain KREL1 activity with the domains provided *in trans* (Figure 15). The CTD is comprised of four alpha-helices:  $\alpha$ H1 (aa 357-367),  $\alpha$ H2 (aa 370-380),  $\alpha$ H3 (aa 392-410) and  $\alpha$ H4 (aa 419-456). Through our mutational analysis of KREL1 in PPI and *in vitro* ligation activity, we identified two critical regions of interest on the CTD:  $\alpha$ H3 (aa 392-410) as the KREPA2 binding site and KWKE motif (aa 441-444) as the diverged motif VI.

Previous work investigated the CTD of KREL1 through LAMA analysis and MEME motif database search and found aa 381-383 region similar to a motif found in microtubule-associated tau proteins (IPB0011084D) (Worthey et al. 2003). Protein Tau binds microtubules through short sequence motif repeats, with a KxGS signature (Avila et al. 2019). This region of KREL1, KIG (aa 381-383), noted as microtubule-associated tau was, however, ruled out to be involved in the KREPA2 interaction with our pulldown assay as KREL1 384 truncation (aa 51-384) was unable to bind to KREPA2 (Figure 14). Interestingly, the stretch of residues before the KXGS motif of

Tau shares similarities to the  $\alpha$ H3 of KRELS (Avila et al. 2019). Within this region of Tau, mutational analysis of the tau protein revealed the importance of the lysine residues for this interaction (Goode et al. 1997, 2000). Consistent with this conclusion, our work supports the essentiality of the positively charged lysine residues of  $\alpha$ H3 of KREL1 CTD in the interaction with KREPA2.



**Figure 22. KREL1-KREPA2 prediction model of interacting residues.**

(A) Electrostatic interactions predicted by Cluspro protein docking server. For comparison, we included the predicted interaction between KREPA1  $\alpha$ H1 and KREL2  $\alpha$ H3. (C) Docking of W156 of KREPA2 to KREL1 CTD conserved with W323 of KREPA1 to KREL2 CTD.

KREL1-KREPA2 interaction of the U-deletion subcomplex is paralogous to the KREL2-KREPA1 interaction of the U-insertion subcomplex of the RECC. In a pulldown assay, I narrowed KREPA2  $\alpha$ H1 to be the interaction site of KREL1 CTD  $\alpha$ H3, while ruling out the OB-fold domain and ZnFs to be involved. Group mutants along the alpha-helices further characterized the electrostatic interaction between the helices. This finding was corroborated with ClusPro protein docking revealing this interaction between KREL1-KREPA2 and showing how a similar interaction might occur between KREL2-KREPA1 (Figure 22) (Desta et al. 2020; Vajda et al.

2017; Kozakov et al. 2017, 2013). In both cases, the negatively charged residues (E/D) interact with the positively charged lysines. The conservation of the lysine residues between ligases may explain the ability of KREL1 to rescue RNA editing and cell viability in RNAi silencing of KREL2 by mimicking the electrostatic interaction to partake in both subcomplexes. The difference between the negatively charged residues of KREPA1/KREPA2 of D/E may be KRELs way of differentiating between KREPA proteins of separate subcomplexes (Schnauffer et al. 2001). One residue of high conservation posing a particular interest is W323/W156 of KREPA1 / KREPA2. The residue appears structurally relevant for pi-stacking interactions with the ligase, similar to that of W442 of KREL1.

The essentiality of the CTD in KREL1 auto-adenylylation suggests a region of critical residues required for the proper hydrolysis of ATP. The region, KWKE (aa 441-444) is synonymous with the RxDK motif previously described in DNA ligases and mRNA capping enzymes that interact with ATP in the binding pocket prior to adenylylation (Sriskanda and Shuman 1998; Håkansson et al. 1997; Subramanya et al. 1996). Specifically, the arginine and lysine residues of RxDK motif VI in DNA ligases interact with the  $\beta$ - and  $\gamma$ -phosphates for the proper orientation of the ATP molecule for the covalent bond to form between the K of motif I and AMP. We suggest a similar interaction between the lysine residues of KWKE and the  $\beta$ - and  $\gamma$ -phosphates to enable the hydrolysis of ATP and the release of pyrophosphate. Highlighted in a multiple sequence alignment, these residues are highly conserved in the KRELs while loosely conserved in T4Rnl2. Although the NTD of T4Rnl2 is functionally independent of its CTD in adenylylation (Ho et al. 2004), the author notes an increase in the ideal pH condition to be more alkaline, likely aiding in the de-protonation of the lysine residue of motif I for covalent bond formation. We have shown that pH 8.0 is the optimal condition of KREL1 adenylylation, although the KREL1 NTD alone was inactive in all pH conditions. We believe the motif VI of KRELs is provided to the binding pocket to aid in ATP hydrolysis and potentially maintain the pH of the microenvironment required for covalent bond formation.

My optimized pre-adenylylated ligation assay and preliminary data of binding pocket mutants, set up a platform for future experiments related to finding competitive ATP inhibitors. As previously discussed in the structural analysis of the binding pocket, residues of motifs I-V are conserved in the analogous T4Rnl2. For example, F209, K307 and R309 of KREL1 are conserved with F119, K225 and K227 of T4Rnl2. The two lysine residues in T4Rnl2 interact with the alpha

phosphate and 5'PO<sub>4</sub> during the ligation and were also shown to be functional when mutated to arginine (Yin et al. 2003; Nandakumar et al. 2006).

KREL1's ~63 kDa interacting partner, KREPA2, contains two recognizable zinc-finger domains (ZnF1 and ZnF2) flanking a 20-residue  $\alpha$ -helix ( $\alpha$ H1) and a C-terminal OB-fold domain. The OB-fold domains conserved in all the six RECC KREPA proteins are essential for the integrity of the complex as they are potentially involved in protein-protein interactions that make up the scaffolding of the RECC (Schnauffer et al. 2010). Of the six KREPA proteins, KREPA1-3 each share two classical C2H2 zinc fingers (Cys2-His2) of 24 aa comprised of a left-handed  $\beta\beta\alpha$  structure, involved in nucleic acid binding and protein-protein binding (Matthews and Sunde 2002; Brayer and Segal 2008; Fedotova et al. 2017). For example, KREPA3 ZnFs seem to play a role in RNA editing efficiency, likely by mediating RECC-RNA interactions (Guo et al. 2010), while the KREPA2 ZnF1 was shown to be involved in complex assembly, potentially by integrating the deletion subcomplex into the RECC core similar to the OB-fold. Our mutational analysis of KREPA2 showed no significant change with the deletion of the OB-fold or mutations of the ZnFs, in both KREL1 interaction and KREL1 ligation and auto-adenylation activities (Figure 21). While these domains may not directly play a role in KREL1 RNA ligation, the question remains if these domains have a role in RNA interaction, possibly downstream of the editing site in conjunction with the other KREPA proteins of the complex.

A class of single-stranded DNA binding proteins, Replication Protein A (RPAs), may reveal overlaps with the assembly of the KREPA proteins of the RECC and their role in RNA editing. RPAs harbour an OB-fold domain fused with a ZnF domain (Bochkareva et al. 2002). The assembly of RPAs is through the interaction of the OB-folds at their winged helix allowing the beta-sheet barrels to act as a track for ssDNA during the human DNA replication (PDB: 1L1O, 6I52) (Yates et al. 2018; Fan and Pavletich 2012). This may propose a model for ssRNA integration into the RECC downstream/upstream of the sites of U-insertion/deletion. While both the OB-fold domain and the ZnF domains were ruled out for KREL1 interaction and KREL1 RNA ligation *in vitro*, future work allocated to RNA-protein interaction studies can reveal the role of these domains in ssRNA binding as in the proposed assembly of RPAs.

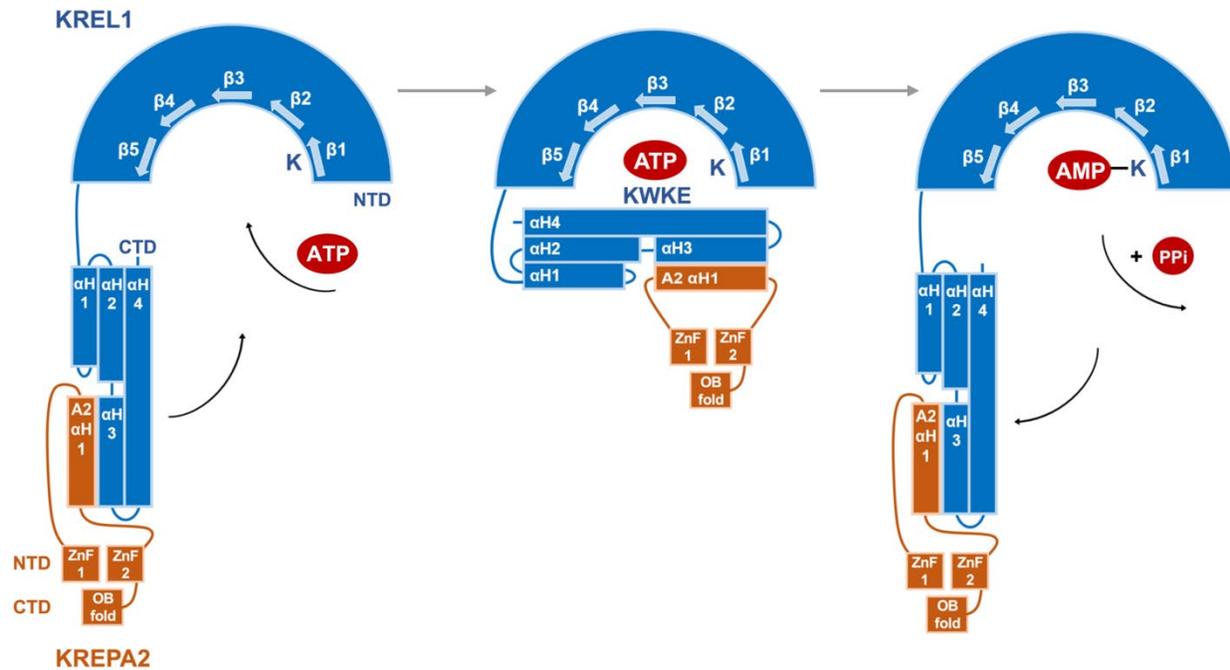
Unlike KREPA3, KREPA1 and KREPA2 harbour a highly conserved alpha-helix between the ZnF domains, containing several exposed charged residues. This region was narrowed to be

the binding site of KREL1  $\alpha$ H3. The conservation of  $\alpha$ H1 in both KREPA2 and KREPA1 proteins may indicate an analogous interaction between the paralogs KREL2-KREPA1. The high degree of similarity between these binding sites may be a possible explanation for the ability of KREL1 to rescue RNA editing and cell viability in KREL2 knockdown mutants (Schnauffer et al. 2001) through binding to KREPA1 to take part in the U-insertion subcomplex. However, no data have been shown to support this to date.

The effect of KREPA2 on KREL1 ligation activity is twofold, it increases the efficiency of KREL1 auto-adenylation, and it requires the interaction between  $\alpha$ H3 of KREL1 CTD and  $\alpha$ H1 of KREPA2. In our *in vitro* functional analysis of the adenylation step, the overall yield of KREL1-AMP formation increases in the presence of KREPA2. KREPA2's role appears to be specific to step 1 as there was no significant improvement in KREL1 activity in steps 2 and 3 in our *in vitro* functional assays. KREL1 NTD is active independent of its CTD in step 3, suggesting that KREPA2 interaction is not required and may be unfavourable. As KREPA2 interacts with the CTD of KREL1, the domain that is essential for ATP-hydrolysis, and further increases the efficiency of KREL1 adenylation activity, we suggest that KREPA2's role mediates the conformational change necessary for this step and regulates the accessibility of these domains.

Based on the results from this work, a model of the ligase-adenylate formation is compatible with the open-closed conformational changes observed with structurally resolved GTPase capping enzymes and DNA ligases (Figure 23) (Unciuleac et al. 2019; Håkansson et al. 1997). The uniqueness of KREL1 is that the conformational change required for auto-adenylation is improved with KREPA2 interaction. Initially, the KREL1 enzyme is required in an open conformation to accept ATP for adenylation. Critical residues of motif I-VI form interactions with ATP. For a covalent bond to form between the K87 of motif I and AMP, residues in motif VI (KWKE) are required to interact and orient the  $\beta$ - and  $\gamma$ -phosphates of the ATP molecule. Such interaction is provided in a closed conformation, mediated by KREPA2's interaction. This is consistent with the unprocessed ATP in the crystalized NTD of KREL1 (PDB code 1XDN) that was missing the motif VI from the CTD for proper orientation and hydrolysis with K87 of motif I. An open conformation at the final stage of ligation, phospho-diester bond formation, is suggested to interact with the dsRNA, whereby motif VI or KREPA2 are not required for activity. KREL1 CTD is always to be imagined with KREPA2  $\alpha$ H1 for RECC integration and

for optimal movement in assisting KREL1's conformational change for ligase auto-adenylation via involvement of a motif VI-like region (KWKE), critical for adenylation.



**Figure 23. Model for ligase auto-adenylation.**

*Mechanism of KREL1 adenylation, representative of step 1 of ligation, in the presence of KREPA2. (Left) KREL1 (in blue) is shown with its N-terminus with five  $\beta$  sheets that make up the ATP binding pocket. The VLR connects the NTD to a four-helix domain at the C-terminus. KREL1 begins in the open conformation with KREPA2 (in orange)  $\alpha$ H1 bound to KREL1  $\alpha$ H3 CTD. (Middle) In the presence of ATP in the binding pocket, the CTD is shown to fold onto its NTD in the closed conformation. The ligase catalyzes a covalent linkage between a lysine of motif I (depicted as the “K” protruding from  $\beta$  sheet 1) with the  $\alpha$ -phosphate of ATP. Motif VI, KWKE, on  $\alpha$ H4 is involved in adenylation to form the KREL1-AMP intermediate. (Right) The ligase opens to release pyrophosphate and proceed in the ligation reaction. KREPA2's interaction mediates the conformational changes.*

## Conclusion and summary

RNA ligation is a three-step mechanism catalyzed by an RNA editing ligase. This thesis characterizes the structure and function of KREL1 and interacting partner KREPA2 to better understand this process in kinetoplastids. A summary of the findings is demonstrated in a model (Figure 23). KREL1  $\alpha$ H3 of the CTD is the site of interaction with KREPA2  $\alpha$ H1, which mediates the open-closed conformational change providing motif VI to the adenylation domain. The *in vitro* analysis of KREL1-KREPA2 interaction and catalytic activity has provided several novel findings about the ligation mechanism of kinetoplastids.

### What is the role of KREL1's NTD?

The domain of residues 51-324, is the binding pocket for ATP hydrolysis. Contacts between residues of motifs I-V are formed throughout the three-step mechanism.

### What is the role of KREL1's CTD?

The four- $\alpha$ -helix domain of residues 354-469 provides the sixth additional motif, KWKE, required for orientation of the ATP  $\beta$  and  $\gamma$  phosphates for adenylation. Interaction likely involves the folding of the enzyme bringing the NTD and CTD together. The VLR (324-354) is directly involved in the conformational change. The CTD of KREL1 is also the site of KREPA2 interaction, specifically with  $\alpha$ H3 (aa 384-410). The physical contact supports KREPA2's role in KREL1 integration into the U-deletion subcomplex. However, we cannot ignore the enhancement of KREL1 in the presence of KREPA2, thus making this an area of investigation.

### What is the role of KREPA2 interaction?

The region of KREPA2 that interacts with KREL1 is aa 142-166, an alpha-helix. The helix-helix interaction is necessary for efficient adenylation activity (step 1); however, it has no effect on RNA adenylation (step 2) or phosphodiester bond formation (step 3). We hypothesize that the connection of KREL1 CTD between the interaction with KREL1 and the essentiality of the domain is through mediating the conformational change of KREL1 in providing the motif VI.

### What is the role of KREPA2 OB-fold and ZnF domains?

KREPA1-3 has two structural zinc finger domains, while all KREPA proteins of the RECC have a structural oligonucleotide binding fold domain. These domains are located on biological proteins involved in protein-protein interaction and protein-nucleotide interactions, of both RNA and DNA. These works have provided no evidence of the effect of KREPA2's OB-fold and ZnF domains in KREL1 interaction and KREL1 ligation activity. Therefore, their detailed function remains a direction for further investigations.

### **Future directions**

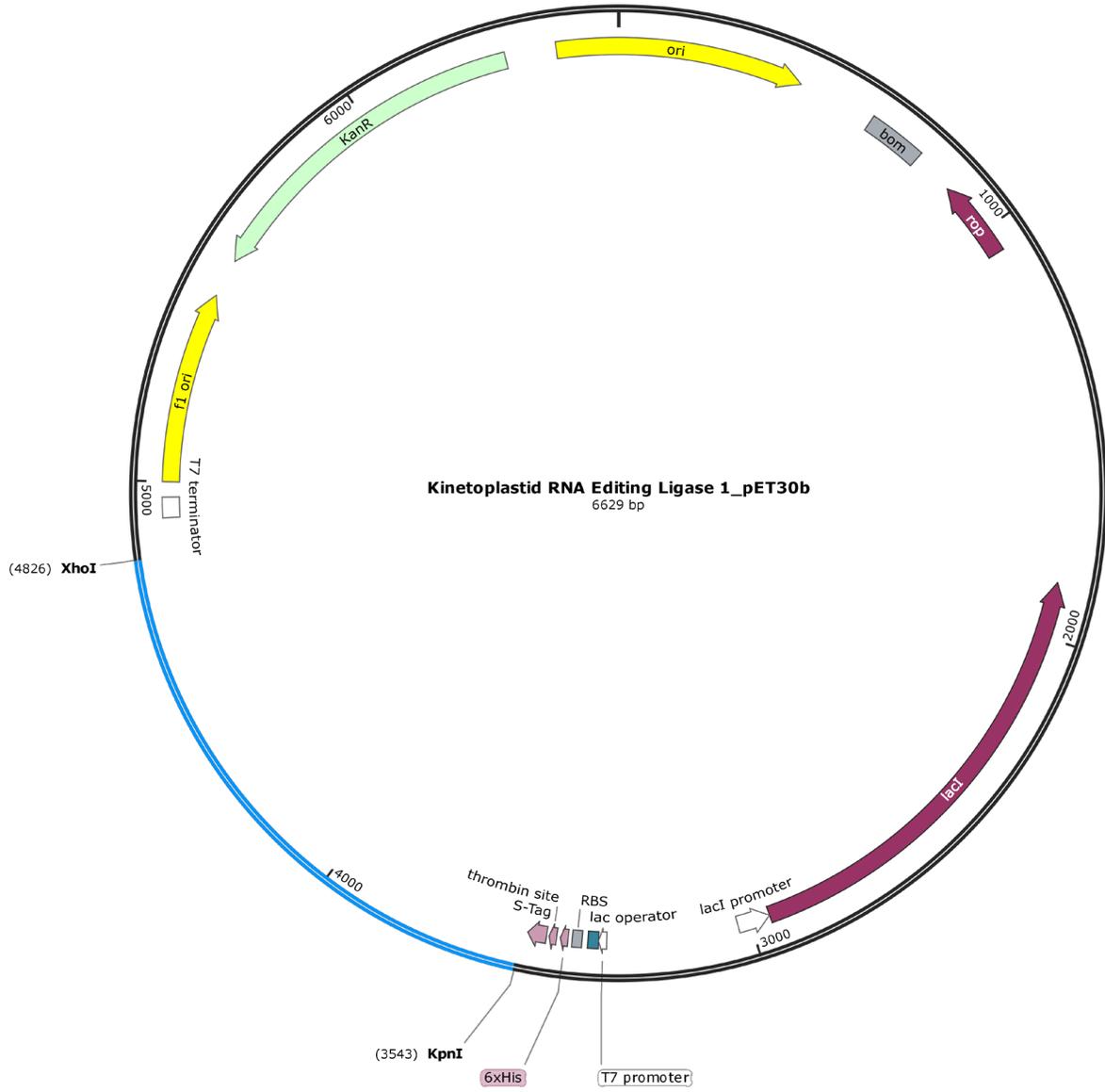
The future direction of this research can be with both *in vivo* experiments and further *in vitro* analysis. Two essential motifs were discovered on KREL1 CTD for KREPA2 interaction and ATP hydrolysis. KREL1 mutants of these motifs should be used in future *in vivo* experiments to monitor cell viability and RNA editing. Moreover, further investigating the ability of KREL1 to partake in both subcomplexes through its CTD. This may further advance drug inhibition discovery against the CTD of KREL1.

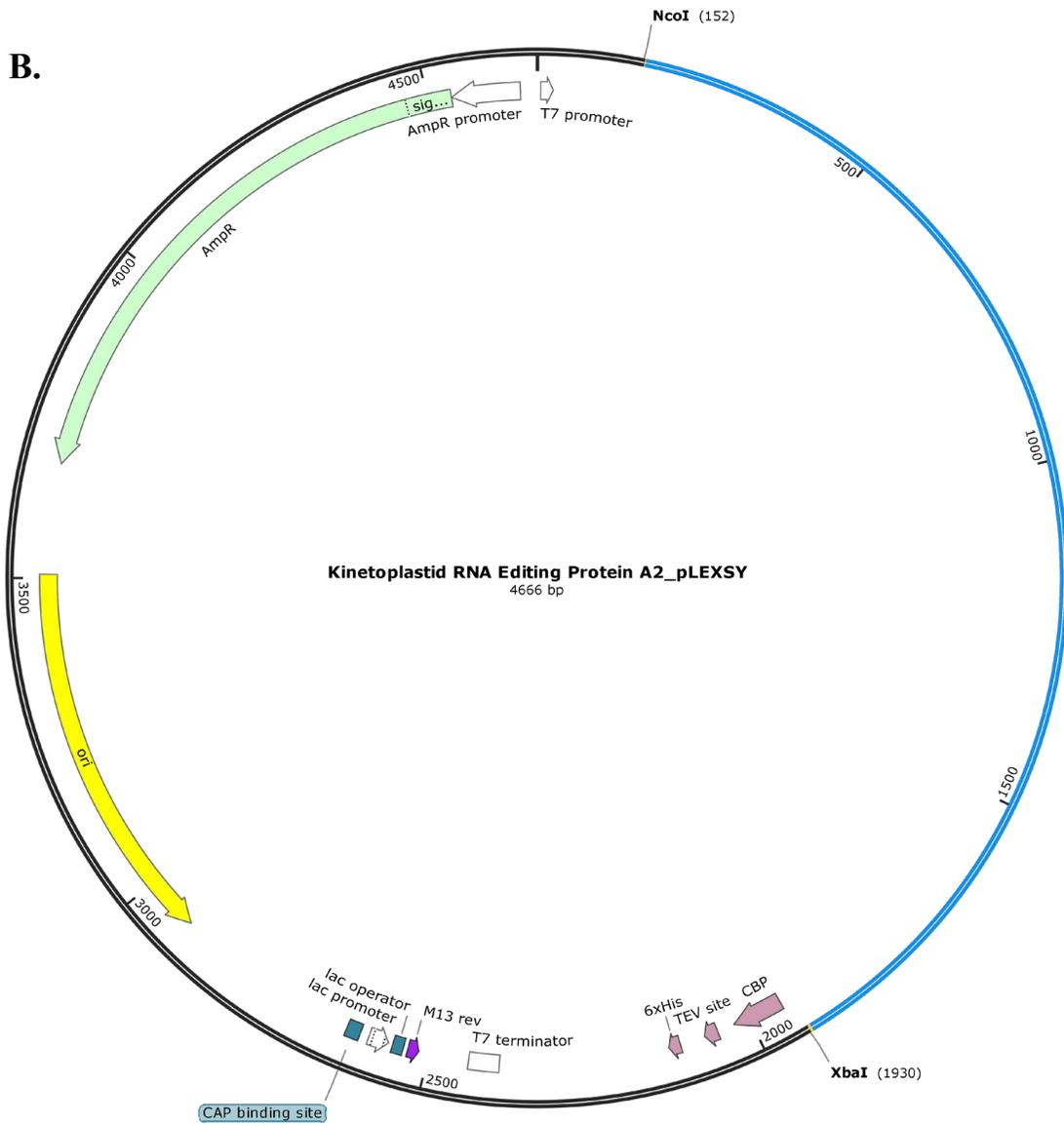
Future *in vitro* analysis should be devoted to investigating the proteins of the editosome, particularly the RNA editing catalytic complex, including how RNA integration and protein assembly occur. We must further understand the involvement of the OB-folds and ZnFs of all six KREPA proteins. To locate the regions of contact between the OB-fold domains of KREPA proteins. To assess the interaction of KREPA proteins with dsRNA/ssRNA. This work would add to our understanding of the RECC assembly and function in RNA editing.



A.

Created with SnapGene®





**Figure 25. Plasmid designs of pET30b with KREL1 insert and pLEXY with KREPA2 insert.**  
 (A) Plasmid map of KREL1 insert in pET30b plasmid with NTD 6xhistag. (B) Plasmid map of KREPA2 insert in pLEXY plasmid with NTD 6xhistag and CTD calmodulin binding peptide.

**L1\_Tb927**

**MQLQRLGAPLLKRLVGGCIRQSTAPIMPCVVVSGGGFLTPVRTYMLPN**DQSDFSPIYIEIDLPSERIQLSHKSGL  
AAQEWVACEKVHGTNFGIYLINQGDHEVVRFKRSKIMDPNENFFGYHILIDEFTAQIRILNDLLKQKYGLSRVGR  
VLNGELFGAKYKHPLVPKSEKWCTLPNGKKFPIAGVQIQREPPQYSPELHFFAFDIKYSVSGAEEDFVLLGYDEFV  
EFSSKVPNLLYARALVRGTLDECLAFDVENFMTPLPALLGLGNYPLEGNLAEGVVIRHVRRGDPAVEKHNVTIIKL  
RCSSFMELKHPGKQKELKETFIDTVRSGALRRVRGNVTVISDSMLPQVEAAANDLLLNNVSDGRLSNVLSKIGREPL  
LSGEVSQVDVALMLAKDALKDFLKEVDSLVLNNTTLAFRKLITNVYFESKRLVEQKWKELMQEAAAQSEAIPLSP  
AAPTKGE

**A2\_Tb927**

MYRRFFRSAAVKHLGTTSAQCLMRATKYPCGAMCRNASSCHPQTLGYQQRFOSTTDKKFHCVCKKPFRLMAAKLH  
LQQAHGGDGSVEAGPGPGVEASSVNVVSTPVPSPI SPVERTFVDEEERRPRMRPTPKPLHQDRDIPAAAMEEMLG  
VWDKIGLNRLEGNFVHSTMVMKVFAAPPDVSEIPLYEHVAPEGENPFDSLDMHTTGTVDATYVGHDAFAEVDLVDPF  
VAAPDQTLNPFPRAGKVRNPFTRISPTQREVVKPLQPPQKKEPLKAPVTPFGQLPMFGQTRPESASFAAAAVSVQAA  
ANPNEVSSPFAAAVSSSPFVGQVASFAPAQDVAGSPFEASPCAASSPFVTAGGQETS PFAPSTPASFGQGSFLFP  
MGTGAPGFTAITETEQRQOELEHGCPTCGKKFSTFEGAAMHSSKSHGIVLESKKVKDRLNKRGVDPDPAYVPSVVDL  
SSTSPFGTRSAIGASWAETELIPHAQCVSNITIVGRVLDVSAENSVHVTVFVEGERSGEEETLTLCCFGEVVSQKI  
RGTLLKRNATIFASGTLRLHPVYEASNNKYYVSPVVHVSMPGTGLAVIT

**Figure 26. Full-length amino acid sequence of KREL1 and KREPA2 from strain Tb927**

**Table 4. Percent identity matrix of full-length RNA ligases**

Percent Identity Matrix - created by Clustal2.1							
1: T4Rnl2_FL	100.00	25.55	26.73	25.16	26.90	25.00	24.68
2: L1_FL_Li	25.55	100.00	73.63	73.33	38.65	40.50	40.00
3: L1_FL_Tb	26.73	73.63	100.00	83.48	38.94	42.51	39.86
4: L1_FL_Tc	25.16	73.33	83.48	100.00	39.90	41.06	38.41
5: L2_FL_Li	26.90	38.65	38.94	39.90	100.00	65.14	67.55
6: L2_FL_Tb	25.00	40.50	42.51	41.06	65.14	100.00	78.12
7: L2_FL_Tc	24.68	40.00	39.86	38.41	67.55	78.12	100.00

**Table 5. Percent identity matrix of the catalytic domain of RNA ligases**

Percent Identity Matrix - created by Clustal2.1							
1: T4Rnl2_NTD	100.00	26.43	28.51	25.88	28.89	26.67	27.11
2: L1_NTD_Li	26.43	100.00	78.33	76.67	39.59	40.96	41.30
3: L1_NTD_Tb	28.51	78.33	100.00	84.71	39.94	43.00	39.74
4: L1_NTD_Tc	25.88	76.67	84.71	100.00	40.26	41.69	39.09
5: L2_NTD_Li	28.89	39.59	39.94	40.26	100.00	68.28	68.61
6: L2_NTD_Tb	26.67	40.96	43.00	41.69	68.28	100.00	80.26
7: L2_NTD_Tc	27.11	41.30	39.74	39.09	68.61	80.26	100.00

**Table 6. Percent identity matrix of the helical C-terminal domain of RNA ligases**

Percent Identity Matrix - created by Clustal2.1							
1: T4Rnl2_CTD	100.00	23.33	23.33	20.00	25.56	24.44	25.56
2: L2_CTD_Li	23.33	100.00	56.07	64.49	36.11	36.11	38.89
3: L2_CTD_Tb	23.33	56.07	100.00	71.96	39.25	41.12	39.25
4: L2_CTD_Tc	20.00	64.49	71.96	100.00	36.45	39.25	35.51
5: L1_CTD_Li	25.56	36.11	39.25	36.45	100.00	60.33	63.33
6: L1_CTD_Tb	24.44	36.11	41.12	39.25	60.33	100.00	80.00
7: L1_CTD_Tc	25.56	38.89	39.25	35.51	63.33	80.00	100.00

**Table 7. Root-mean squared deviation matrix of the domains of RNA ligases**

RMSD Matrix - created by DALI server							
L1_NTD_T.b.	0	0.2	0.7	1.7	1.6	1.5	2.2
L1_NTD_T.c.	0.2	0	0.7	1.6	1.6	1.5	2.3
L1_NTD_L.i.	0.7	0.7	0	1.5	1.5	1.4	2.1
L2_NTD_T.b.	1.7	1.6	1.4	0	0.2	1.1	2.4
L2_NTD_T.c.	1.5	1.5	1.5	0.2	0	1.1	2.4
L2_NTD_L.i.	1.6	1.6	1.5	1.1	1.1	0	2.2
T4Rnl2_NTD	2.2	2.3	2.1	2.4	2.4	2.2	0

RMSD Matrix - created by DALI server							
L1_CTD_T.b.	0	2.4	1.2	2.1	1.3	1.6	2.1
L1_CTD_T.c.	2.4	0	1.2	2	1.3	1.5	1.9
L1_CTD_L.i.	1.2	1.2	0	2	1.3	1.6	1.9
L2_CTD_T.b.	2.1	2	2	0	1.8	1.6	2
L2_CTD_T.c.	1.3	1.3	1.3	1.8	0	1	2.1
L2_CTD_L.i.	1.6	1.5	1.6	1.6	1	0	2.1
T4Rnl2_CTD	2.1	1.9	1.9	2	2.1	2.1	0

**Table 8. Percent identity matrix of full-length Kinetoplastid A1 and A2 proteins**

Percent Identity Matrix - created by Clustal2.1						
1: A2_L_infantum	100.00	40.21	38.87	22.82	20.34	22.91
2: A2_T_brucei	40.21	100.00	60.79	20.04	19.47	20.00
3: A2_T_cruzi	38.87	60.79	100.00	20.56	21.59	21.68
4: A1_L_infantum	22.82	20.04	20.56	100.00	39.28	38.59
5: A1_T_brucei	20.34	19.47	21.59	39.28	100.00	55.18
6: A1_T_cruzi	22.91	20.00	21.68	38.59	55.18	100.00

**Table 9. Percent identity matrix of full-length Kinetoplastid A proteins of *Trypanosoma brucei***

Percent Identity Matrix - created by Clustal2.1						
1: KREPA4	100.00	21.82	20.88	25.44	20.41	19.00
2: KREPA5	21.82	100.00	19.00	26.28	30.84	34.82
3: KREPA2	20.88	19.00	100.00	24.51	23.00	32.67
4: KREPA1	25.44	26.28	24.51	100.00	31.19	31.58
5: KREPA3	20.41	30.84	23.00	31.19	100.00	40.37
6: KREPA6	19.00	34.82	32.67	31.58	40.37	100.00

**Table 10. Root-mean squared deviation matrix of full-length Kinetoplastid A proteins of *Trypanosoma brucei***

RMSD Matrix - created by DALI server						
KREPA1	0	20.8	16.9	5.3	7.2	3.6
KREPA2	20.8	0	16.1	4.9	3	6
KREPA3	16.9	16.1	0	5.1	5.8	6
KREPA4	5.3	4.9	5.1	0	4.2	3.3
KREPA5	7.2	3	5.8	4.2	0	5.5
KREPA6	3.6	6	6	3.3	5.5	0

## References

- Alalaiwe A, Wang PW, Lu PL, Chen YP, Fang JY, Yang SC. 2018. Synergistic anti-MRSA activity of cationic nanostructured lipid carriers in combination with oxacillin for cutaneous application. *Front Microbiol* **9**: 1493.
- Aphasizhev R, Aphasizheva I. 2014. Mitochondrial RNA editing in trypanosomes: Small RNAs in control. *Biochimie* **100**: 125–131.
- Aphasizheva I, Alfonzo J, Carnes J, Cestari I, Cruz-Reyes J, Göringer HU, Hajduk S, Lukeš J, Madison-Antenucci S, Maslov DA, et al. 2020. Lexis and Grammar of Mitochondrial RNA Processing in Trypanosomes. *Trends Parasitol* **36**: 337–355.
- Aphasizheva I, Zhang L, Wang X, Kaake RM, Huang L, Monti S, Aphasizhev R. 2014. RNA Binding and Core Complexes Constitute the U-Insertion/Deletion Editosome. *Mol Cell Biol* **34**: 4329–4342.
- Avila J, Jadhav S, Primio C Di, Devred F, Landrieu I, Barbier P, Zejneli O, Martinho M, Lasorsa A, Belle V, et al. 2019. Role of Tau as a Microtubule-Associated Protein: Structural and Functional Aspects.
- Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted *coxII* gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**: 819–826.
- Blum B, Simpson L. 1990. Guide RNAs in kinetoplastid mitochondria have a nonencoded 3' oligo(U) tail involved in recognition of the preedited region. *Cell* **62**: 391–397.
- Bochkareva E, Korolev S, Lees-Miller SP, Bochkarev A. 2002. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J* **21**: 1855–1863.
- Boiani M, Piacenza L, Hernández P, Boiani L, Cerecetto H, González M, Denicola A. 2010. Mode of action of nifurtimox and N-oxide-containing heterocycles against *Trypanosoma cruzi*: is oxidative stress involved? *Biochem Pharmacol* **79**: 1736–45.
- Brayer KJ, Segal DJ. 2008. Keep your fingers off my DNA: Protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem Biophys* **50**: 111–131.
- Brecht M, Niemann M, Schlüter E, Müller UF, Stuart K, Göringer HU. 2005. TbMP42, a protein component of the RNA editing complex in African trypanosomes, has endo-exoribonuclease activity. *Mol Cell* **17**: 621–630.
- Büscher P, Cecchi G, Jamonneau V, Priotto G. 2017. Human African trypanosomiasis. *Lancet* **390**: 2397–2409.
- de Koning HP. 2020. The drugs of sleeping sickness: Their mechanisms of action and resistance, and a brief history. *Trop Med Infect Dis* **5**.
- Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WGJ. 2004a. High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. *J Mol Biol* **343**: 601–613.
- Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WGJ. 2004b. High Resolution Crystal Structure of a Key Editosome Enzyme from *Trypanosoma brucei*: RNA Editing Ligase 1. *J Mol Biol*

**343:** 601–613.

- Denkers EY, Butcher BA. 2005. Sabotage and exploitation in macrophages parasitized by intracellular protozoans. *Trends Parasitol* **21**: 35–41.
- Desta IT, Porter KA, Xia B, Kozakov D, Vajda S. 2020. Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure* **28**: 1071-1081.e3.
- Doherty AJ, Suh SW. 2000. Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Res* **28**: 4051–4058.
- Drozd M, Palazzo SS, Salavati R, O’Rear J, Clayton C, Stuart K. 2002. TbMP81 is required for RNA editing in *Trypanosoma brucei*. *EMBO J* **21**: 1791–1799.
- Ellenberger T, Tomkinson AE. 2008. Eukaryotic DNA ligases: Structural and functional insights. *Annu Rev Biochem* **77**: 313–338.
- Fan J, Pavletich NP. 2012. Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev* **26**: 2337–2347.
- Fanning E, Klimovich V, Nager AR. 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* **34**: 4126–4137.
- Fedotova AA, Bonchuk AN, Mogila VA, Georgiev PG. 2017. C2H2 zinc finger proteins: The largest but poorly explored family of higher eukaryotic transcription factors. *Acta Naturae* **9**: 47–58.
- Field MC, Horn D, Fairlamb AH, Ferguson MAJ, Gray DW, Read KD, De Rycker M, Torrie LS, Wyatt PG, Wyllie S, et al. 2017. Anti-trypanosomatid drug discovery: An ongoing challenge and a continuing need. *Nat Rev Microbiol* **15**: 217–231.
- Flynn RL, Zou L. 2010. Oligonucleotide/oligosaccharide-binding fold proteins: A growing family of genome guardians. *Crit Rev Biochem Mol Biol* **45**: 266–275.
- Franco JR, Cecchi G, Paone M, Diarra A, Grout L, Ebeja AK, Simarro PP, Zhao W, Argaw D. 2022. The elimination of human African trypanosomiasis: Achievements in relation to WHO road map targets for 2020. *PLoS Negl Trop Dis* **16**.
- Goode BL, Chau M, Denis PE, Feinstein SC. 2000. Structural and functional differences between 3-repeat and 4-repeat tau isoforms: Implications for normal tau function and the onset of neurodegenerative disease. *J Biol Chem* **275**: 38182–38189.
- Goode BL, Denis PE, Panda D, Radeke MJ, Miller HP, Wilson L, Feinstein SC. 1997. Functional interactions between the proline-rich and repeat regions of tau enhance microtubule binding and assembly. *Mol Biol Cell* **8**: 353.
- Gu H, Yoshinari S, Ghosh R, Ignatovkina A V., Gollnick PD, Murakami KS, Ho CK. 2016. Structural and mutational analysis of archaeal ATP-dependent RNA ligase identifies amino acids required for RNA binding and catalysis. *Nucleic Acids Res* **44**: 2337–2347.
- Guo X, Ernst NL, Carnes J, Stuart KD. 2010. The zinc-fingers of KREPA3 are essential for the complete editing of mitochondrial mRNAs in *Trypanosoma brucei*. *PLoS One* **5**: e8913.
- Guo X, Ernst NL, Stuart KD. 2008. The KREPA3 Zinc Finger Motifs and OB-Fold Domain Are Essential for RNA Editing and Survival of *Trypanosoma brucei*. *Mol Cell Biol* **28**: 6939–6953.

- Hajduk S, Ochsenreiter T. 2010. RNA editing in kinetoplastids. *RNA Biol* **7**: 229–236.
- Håkansson K, Doherty AJ, Shuman S, Wigley DB. 1997. X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* **89**: 545–553.
- Handman E, Bullen DVR. 2002. Interaction of Leishmania with the host macrophage. *Trends Parasitol* **18**: 332–334.
- Ho CK, Shuman S. 2002. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci U S A* **99**: 12709–12714.
- Ho CK, Wang LK, Lima CD, Shuman S. 2004a. Structure and Mechanism of RNA Ligase. *Structure* **12**: 327–339.
- Ho CK, Wang LK, Lima CD, Shuman S. 2004b. Structure and Mechanism of RNA Ligase. *Structure* **12**: 327–339.
- Huang CE, O’Hearn SF, Sollner-Webb B. 2002. Assembly and Function of the RNA Editing Complex in Trypanosoma brucei Requires Band III Protein. *Mol Cell Biol* **22**: 3194–3203.
- Jensen RE, Englund PT. 2012. Network news: The replication of kinetoplast DNA. *Annu Rev Microbiol* **66**: 473–491.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nat* 2021 5967873 **596**: 583–589.
- Kable ML, Seiwert SD, Heidmann S, Stuart K. 1996. RNA editing: A mechanism for gRNA-specified uridylyate insertion into precursor mRNA. *Science (80- )* **273**: 1189–1195.
- Kala S, Moshiri H, Mehta V, Yip CW, Salavati R. 2012. The Oligonucleotide Binding (OB)-Fold Domain of KREPA4 Is Essential for Stable Incorporation into Editosomes. *PLoS One* **7**: 46864.
- Kala S, Salavati R. 2010. OB-fold domain of KREPA4 mediates high-affinity interaction with guide RNA and possesses annealing activity. *RNA* **16**: 1951–1967.
- Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Vajda S. 2013. How good is automated protein docking? *Proteins Struct Funct Bioinforma* **81**: 2159–2166.
- Kozakov D, Hall DR, Xia B, Porter KA, Padhorney D, Yueh C, Beglov D, Vajda S. 2017. The ClusPro web server for protein-protein docking. *Nat Protoc* **12**: 255–278.
- Lidani KCF, Andrade FA, Bavia L, Damasceno FS, Beltrame MH, Messias-Reason IJ, Sandri TL. 2019. Chagas disease: From discovery to a worldwide health problem. *J Phys Oceanogr* **49**: 166.
- Macleod ET, Darby AC, Maudlin I, Welburn SC. 2007. Factors affecting trypanosome maturation in tsetse flies. *PLoS One* **2**.
- Malvy D, Chappuis F. 2011. Sleeping sickness. *Clin Microbiol Infect* **17**: 986–995.
- Martins A, Shuman S. 2004. An RNA ligase from Deinococcus radiodurans. *J Biol Chem* **279**: 50654–50661.
- Matthews JM, Sunde M. 2002. Zinc fingers - Folds for many occasions. *IUBMB Life* **54**: 351–355.

- McDermott SM, Luo J, Carnes J, Ranish JA, Stuart K. 2016. The Architecture of Trypanosoma brucei editosomes. *Proc Natl Acad Sci U S A* **113**: E6476–E6485.
- Mehta V, Moshiri H, Srikanth A, Kala S, Lukeš J, Salavati R. 2020. Sulfonated inhibitors of the RNA editing ligases validate the essential role of the MRP1/2 proteins in kinetoplastid RNA editing.
- Mehta V, Sen R, Moshiri H, Salavati R. 2015. Mutational analysis of Trypanosoma brucei RNA editing ligase reveals regions critical for interaction with KREPA2. *PLoS One* **10**.
- Nandakumar J, Ho CK, Lima CD, Shuman S. 2004. RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J Biol Chem* **279**: 31337–31347.
- Nandakumar J, Shuman S, Lima CD. 2006a. RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward. *Cell* **127**: 71–84.
- Nandakumar J, Shuman S, Lima CD. 2006b. RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward. *Cell* **127**: 71–84.
- Park YJ, Hol WGJ. 2012. Explorations of linked editosome domains leading to the discovery of motifs defining conserved pockets in editosome OB-folds. *J Struct Biol* **180**: 362–373.
- Pérez-Molina JA, Molina I. 2018. Chagas disease. *Lancet* **391**: 82–94.
- Read LK, Lukeš J, Hashimi H. 2016a. Trypanosome RNA editing: The complexity of getting U in and taking U out. *Wiley Interdiscip Rev RNA* **7**: 33–51.
- Read LK, Lukeš J, Hashimi H. 2016b. Trypanosome RNA editing: The complexity of getting U in and taking U out. *Wiley Interdiscip Rev RNA* **7**: 33–51.
- Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B, Brooker S. 2007. Cutaneous leishmaniasis. *Lancet Infect Dis* **7**: 581–596.
- Ringpis GE, Stagno J, Aphasizhev R. 2010. Mechanism of U-insertion RNA editing in trypanosome mitochondria: Characterization of RET2 functional domains by mutational analysis. *J Mol Biol* **399**: 696–706.
- Salavati R, Ernst NL, O’Rear J, Gilliam T, Tarun S, Stuart K. 2006. KREPA4, an RNA binding protein essential for editosome integrity and survival of Trypanosoma brucei. *RNA* **12**: 819–831.
- Salavati R, Moshiri H, Kala S, Shateri Najafabadi H. 2012. Inhibitors of RNA editing as potential chemotherapeutics against trypanosomatid pathogens. *Int J Parasitol Drugs Drug Resist* **2**: 36–46.
- Samai P, Shuman S. 2012. Kinetic analysis of DNA strand joining by Chlorella virus DNA ligase and the role of nucleotidyltransferase motif VI in ligase adenylation. *J Biol Chem* **287**: 28609–28618.
- Schnauffer A, Clark-Walker GD, Steinberg AG, Stuart K. 2005. The F1-ATP synthase complex in bloodstream stage trypanosomes has an unusual and essential function. *EMBO J* **24**: 4029–4040.
- Schnauffer A, Ernst NL, Palazzo SS, O’Rear J, Salavati R, Stuart K. 2003. Separate insertion and deletion subcomplexes of the Trypanosoma brucei RNA editing complex. *Mol Cell* **12**: 307–

- Schnauffer A, Panigrahi AK, Panicucci B, Igo RP, Salavati R, Stuart K. 2001a. An RNA Ligase Essential for RNA Editing and Survival of the Bloodstream Form of *Trypanosoma brucei*. *Science (80- )* **291**: 2159–2162.
- Schnauffer A, Panigrahi AK, Panicucci B, Igo RP, Salavati R, Stuart K. 2001b. An RNA Ligase Essential for RNA Editing and Survival of the Bloodstream Form of *Trypanosoma brucei*. *Science (80- )* **291**: 2159–2162.
- Schnauffer A, Wu M, Park YJ, Nakai T, Deng J, Proff R, Hol WGJ, Stuart KD. 2010. A protein-protein interaction map of trypanosome ~20s editosomes. *J Biol Chem* **285**: 5282–5295.
- Schultzberg M, Ambatsis M, Samuelsson E -B, Kristensson K, van Meirvenne N. 1988. Spread of *Trypanosoma brucei* to the nervous system: Early attack on circumventricular organs and sensory ganglia. *J Neurosci Res* **21**: 56–61.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
- Simpson L, Thiemann OH, Savill NJ, Alfonzo JD, Maslov DA. 2000. Evolution of RNA editing in trypanosome mitochondria. *Proc Natl Acad Sci U S A* **97**: 6986–6993.
- Sriskanda V, Shuman S. 1998. Mutational analysis of Chlorella virus DNA ligase: Catalytic roles of domain I and motif VI. *Nucleic Acids Res* **26**: 4618–4625.
- Stagno J, Aphasizheva I, Bruystens J, Luecke Hartmut H, Aphasizhev R. 2010. Structure of the mitochondrial editosome-like complex associated TUTase 1 reveals divergent mechanisms of UTP selection and domain organization. *J Mol Biol* **399**: 464–475.
- Stuart K, Allen TE, Heidmann S, Seiwert SD. 1997. RNA editing in kinetoplastid protozoa. *Microbiol Mol Biol Rev* **61**: 105–120.
- Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, Reed S, Tarleton R. 2008a. Kinetoplastids: Related protozoan pathogens, different diseases. *J Clin Invest* **118**: 1301–1310.
- Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, Reed S, Tarleton R. 2008b. Kinetoplastids: Related protozoan pathogens, different diseases. *J Clin Invest* **118**: 1301–1310.
- Stuart KD, Schnauffer A, Ernst NL, Panigrahi AK. 2005. Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* **30**: 97–105.
- Subramanya HS, Doherty AJ, Ashford SR, Wigley DB. 1996. Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* **85**: 607–615.
- Tarleton RL, Reithinger R, Urbina JA, Kitron U, Gürtler RE. 2007. The challenges of Chagas disease - Grim outlook or glimmer of hope? *PLoS Med* **4**: 1852–1857.
- Tarun SZ, Schnauffer A, Ernst NL, Proff R, Deng J, Hol W, Stuart K. 2008. KREPA6 is an RNA-binding protein essential for editosome integrity and survival of *Trypanosoma brucei*. *RNA* **14**: 347–358.

- Unciuleac MC, Goldgur Y, Shuman S. 2019. Structures of ATP-bound DNA ligase D in a closed domain conformation reveal a network of amino acid and metal contacts to the ATP phosphates. *J Biol Chem* **294**: 5094–5104.
- Unciuleac MC, Shuman S. 2015. Characterization of a novel eukaryal nick-sealing RNA ligase from *Naegleria gruberi*. *RNA* **21**: 824–832.
- Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Kozakov D. 2017. New additions to the ClusPro server motivated by CAPRI. *Proteins Struct Funct Bioinforma* **85**: 435–444.
- Van Den Abbeele J, Claes Y, Van Bockstaele D, Le Ray D, Coosemans M. 1999. *Trypanosoma brucei* spp. development in the tsetse fly: Characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118**: 469–478.
- Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, et al. 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**: D439–D444.
- Walsh B, Hill KL. 2021. Right place, right time: Environmental sensing and signal transduction directs cellular differentiation and motility in *Trypanosoma brucei*. *Mol Microbiol* **115**: 930–941.
- Wamwiri FN, Changasi RE. 2016. Tsetse Flies (*Glossina*) as vectors of human African Trypanosomiasis: A review. *Biomed Res Int* **2016**.
- WHO. 2022. Trypanosomiasis, human African (sleeping sickness). [https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-\(sleeping-sickness\)](https://www.who.int/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness)) (Accessed September 28, 2022).
- Worthey EA, Schnauffer A, Mian IS, Stuart K, Salavati R. 2003. Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res* **31**: 6392–6408.
- Yates LA, Aramayo RJ, Pokhrel N, Caldwell CC, Kaplan JA, Perera RL, Spies M, Antony E, Zhang X. 2018. A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. *Nat Commun* **2018 91 9**: 1–14.
- Yin S, Ho CK, Shuman S. 2003. Structure-function analysis of T4 RNA ligase 2. *J Biol Chem* **278**: 17601–17608.
- Zíková A. 2022. Mitochondrial adaptations throughout the *Trypanosoma brucei* life cycle. *J Eukaryot Microbiol* e12911.
- Zimmermann S, Hall L, Riley S, Sørensen J, Amaro RE, Schnauffer A. 2016. A novel high-Throughput activity assay for the *Trypanosoma brucei* editosome enzyme REL1 and other RNA ligases. *Nucleic Acids Res* **44**: e24–e24.
- 2019a. WHO | Chagas disease (American trypanosomiasis). *WHO*.
- 2019b. WHO | Leishmaniasis. *WHO*.