BIOCHEMICAL AND FUNCTIONAL INVESTIGATION OF *TB*RAP1 AND ITS ROLE IN *T. BRUCEI* VARIANT SURFACE GLYCOPROTEIN REGULATION

YANG XIAN

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Applied Biology and Chemical Technology

Biochemical and Functional Investigation of *Tb*RAP1 and its Role in *T. Brucei* Variant Surface Glycoprotein Regulation

YANG XIAN

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

JUNE 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written nor material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Signature:

Name: YANG Xian

Abstract

Protozoan parasite *Trypanosoma brucei* (*T. brucei*) causes African sleeping sickness in humans and Nagana in cattle. These diseases are difficult to cure because *T. brucei* can evade the host immune response by regularly switching its major surface antigen, Variant Surface Glycoprotein (*VSG*). *T. brucei* has a repertoire of over 2,000 *VSG* genes. These genes are expressed from one of the Expression Sites (ESs) located immediately upstream of telomeres in a strictly monoallelic manner. Previous studies have found that telomeric proteins belonging to the Shelterin complex play important roles in regulating *VSG* monoallelic expression and switching. Specifically, *Tb*RAP1, a member of the telomere Shelterin complex, is essential for repressing the expression of *VSG*s located in "silent" ESs.

We have carried out biochemical and structural studies of *Tb*RAP1 to elucidate the molecular mechanism of *Tb*RAP1-mediated *VSG* silencing.

First of all, we identified an unusual DNA binding activity in *Tb*RAP1 MybLike domain (*Tb*RAP1-ML) by *in vitro* EMSA assay and NMR titration experiments. A positively charged $_{737}$RKRRR$_{741}$ patch (R/K patch) within the MybLike domain is responsible for electrostatics-based, sequence non-specific binding to both single- and double-stranded DNA. Consequently, we have termed *Tb*RAP1 (aa 734-761) containing the R/K patch as DNA binding (DB) region. Cell-based assays reveal that this electrostatics-based DNA binding activity is essential for *Tb*RAP1's telomere localization, *VSG* silencing, telomere integrity, and cell proliferation.

Secondly, we also determined the structure of *Tb*RAP1 MybLike domain (aa 639-761) by NMR. Unexpectedly, this domain turned out to contain a canonical RNA Recognition Module (RRM) with the characteristic topology of a four-stranded anti-parallel β-sheet plus two α-helices. Structural superimposition and sequence alignment of *Tb*RAP1 RRM domain to other classic RRMs suggest that this module may have potential RNA binding activity.

To validate whether the *Tb*RAP1 RRM domain possesses RNA binding activity, we used EMSA and NMR titration methods to assess the interaction between *Tb*RAP1 RRM and various RNA substrates. Our results show that the RRM domain binds preferably to *VSG* mRNA. And the conserved F655 and F694 residues located on the canonical RNA binding site are critical for this interaction. Additionally, the positively charged DB segment also showed electrostatics-based non-specific RNA binding. Given that no RAP1 homologs have been reported to possess RNA binding activity, our finding of *Tb*RAP1's RNA binding activity through its RRM domain is completely novel.

Our research team also carried out *in vivo* studies to delineate the functional significance of the RNA binding activity in *Tb*RAP1 RRM domain. Indeed, *Tb*RAP1 associates with the active *VSG* mRNA *in vivo* as confirmed by RNA-IP experiments. Additionally, mutations that abolish *Tb*RAP1's RNA binding activity led to the expression of all *VSG*s located in silent ESs cell growth defect. These data indicate that *Tb*RAP1's RNA binding activity is essential for *VSG* monoallelic expression and cell

viability. Most importantly, mutations that abolish *Tb*RAP1's RNA binding activity caused a decrease in the mRNA level of the only active *VSG*. Thus, the novel RNA-binding activity mediated by *Tb*RAP1 RRM module may be essential for monoallelic expression of *VSG* by sustaining the high expression level of the only active *VSG* while silencing the rest.

In summary, using a combination of biochemical, structural, and functional studies, my thesis work has uncovered novel DNA- and RNA-binding activities in *Tb*RAP1. The electrostatics-based sequence non-specific DNA binding activity mediated by the DB region is critical for telomere localization of *Tb*RAP1 and *VSG* silencing. The novel RNA binding activity mediated by the *Tb*RAP1 RRM domain is essential for sustaining monoallelic expression of the only active *VSG*.

Publications arising from the thesis

Afrin, M., Gaurav, A. K., **Yang, X**., Pan, X., Zhao, Y., & Li, B. (2020). *Tb*RAP1 has an unusual duplex DNA binding activity required for its telomere localization and *VSG* silencing. *Science advances*, *6*(38), eabc4065.

Acknowledgements

I would like to express my deepest thankfulness to my supervisor Prof. Zhao Yanxiang for her professional guidance, kind support, and positive encouragement in the past years. Thanks a lot for giving me this wonderful opportunity of studying in her lab. Prof. Zhao is happy to share with us her experiences and opinions in both science and life, which is precious wealth for me. I am deeply impressed by her patience, creativity, and passion for science, and I am grateful for what I learned from her.

Afterward, I would like to thank all lab members of Prof. Zhao's team. Thanks Dr. Pan Xuehua, for his preliminary studies on *Tb*RAP1 project; thanks Dr. Li Na, who guided me the protein biochemical technology at the beginning of my Ph.D. study; thanks Dr. Wu Shuai, Dr. Qiu Xianxiu, and Ms. Zhang Xiaozhe, for teaching me useful experimental skills and helping me adapt to my life in Hong Kong; thanks Ms. Zhang Shuqi, Ms. Chen Jingyi, Ms. Gao Shan, Ms. Feng Yu, Ms. Yu Yingting and Ms. Wang Lei, for your companion and support in my study and life. I am so appreciative to meet all of you.

I am grateful to our collaborator, Dr. Li Bibo from the Cleveland State University, for giving me a great opportunity to study EMSA assays and cell-based *in vivo* assays in her lab. It is a really happy experience for three months of study in Cleveland. Special thanks to Dr. Amit Gaurav, for his generous help in my study and life during my visiting period. Many thanks to Dr. Arpita Saha, Dr. Maiko Tonini, Ms. Brittny Schnur, Ms. Marjia Afrin, Ms. Hanadi Kishmiri, Mr. Sayeed Abdus, and Mr. M A G Rabbani.

I would like to thank Prof. Zhang Mingjie (The Hong Kong University of Science and Technology) for his kind support in solving NMR structure of *Tb*RAP1 RRM and providing NMR machine for a series of NMR-related assays. Thanks Dr. Ye Fei and Dr. Ji Zeyang, lab members of Prof. Zhang Mingjie, for helping NMR data collection.

Last, I must express my appreciation to my dear parents and all my family members. Thanks a lot for your endless support and love on me during these years.

Table of contents

List of figures and tables

Figures

Tables

Abbreviations

6-FAM	6-carboxyfluorescein
aa	amino acids
β-ME	β-Mercaptoethanol
BRCT	BRCA1 C-terminus
BSA	bovine serum albumin
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
CO	crossover
Co-IP	Co-immunoprecipitation
CV	Column Value
DAPI	4′,6-diamidino-2-phenylindole
DDR	DNA damage response
DSBs	DNA double-strand breaks
E.coil	*Esherichia coli*
EMSA	Electrophoretic Mobility Shift Assay
ESs	Expression Sites
FP	Fluorescence Polarization
GC	gene conversion
HAT	human African trypanosomiasis
His6	hexa-histidine
hnRNP	heterogeneous nuclear ribonucleoprotein
hRAP1	human RAP1

HSQC	Heteronuclear Single Quantum Coherence
IF	Immunofluorescence
IgG	immunoglobulin G
kb	kilobase
KD	kilodalton
m/z	mass-to-charge ratio
MS	Mass Spectroscopy
MW	Molecular weight
NHEJ	Nonhomologous End-Joining
NLS	nuclear localization signal
NMR	Nuclear magnetic resonance
PCR	Polymerase Chain Reaction
PMSF	Phenylmethanesulfonyl fluoride
qRT-PCR	Quantitative reverse transcription PCR
RBD	RNA-binding domain
RCT	RAP1 C-terminus
RMSD	Root Mean Square Deviation
RNP	ribonucleoprotein
RRM	RNA Recognition Motif
*Sc*RAP1	*Saccharomyces cerevisiae* RAP1
*Sp*RAP1	*Schizosaccharomyces pombe* RAP1
T. brucei	*Trypanosoma brucei*
*Tb*RAP1	*T. brucei* Repressor/Activator Protein 1

*Tb*TIF2	*T. brucei* TRF-Interacting Factor 2
*Tb*TRF	*T. brucei* TTAGGG Repeat-binding Factor
TERRA	telomeric transcript
VSG	Variant Surface Glycoprotein
WHO	World Health Organization

1 Introduction

1.1 *Trypanosoma brucei* and its *VSG*-mediated antigenic variation

1.1.1 *T. brucei* is an infectious protozoan that causes sleeping sickness

Trypanosoma brucei (*T. brucei*) is one of the extracellular eukaryotic flagellate parasites that can cause infectious diseases such as Human African Trypanosomiasis (HAT) (Berriman*, et al.* 2005). HAT wildly spread in sub-Saharan Africa (Figure 1.1). Because of causing sleeping problems after infection, HAT is also called sleeping sickness. According to the report from the World Health Organization (WHO), *T. brucei* infects about 20,000 to 30,000 sleeping sickness cases per year, and it is invariably fatal without treatment (Rotureau*, et al.* 2013).

T. brucei is transmitted by an insect vector, the *Tse Tse* fly, a species of large biting flies. By circulating through the salivary glands of *Tse Tse* flies, *T. brucei* can easily spread from one infected host to another when the flies take a blood meal. Currently, there is no effective vaccine against *T. brucei*, and late-stage infection becomes incurable and fatal when the parasites penetrate the blood-brain barrier and infect the central nervous system.

Infection of sleeping sickness can be divided into two stages (Malvy*, et al.* 2011) (Figure 1.1). Firstly, *T. brucei* pathogens proliferate and develop in the gut of tsetse fly. Afterward, the infected tsetse fly transmits pathogens into humans. At the initial stage, parasites only exist in blood and the lymph systems with rapid proliferation. The common symptoms at the initial stage contain enlargement of lymphatic nodes, fevers, headaches, and pains in patients. *T. brucei* pathogens gradually penetrate the blood-brain barrier and affect the Central Nervous System (CNS), which marks entering the second stage of infection. Patients appear neurological symptoms such as confusion,

anemia, poor coordination, troubled sleeping, and speech disorders. Trypanosomes continue to spread to other healthy people by tsetse fly feeding on injected blood.

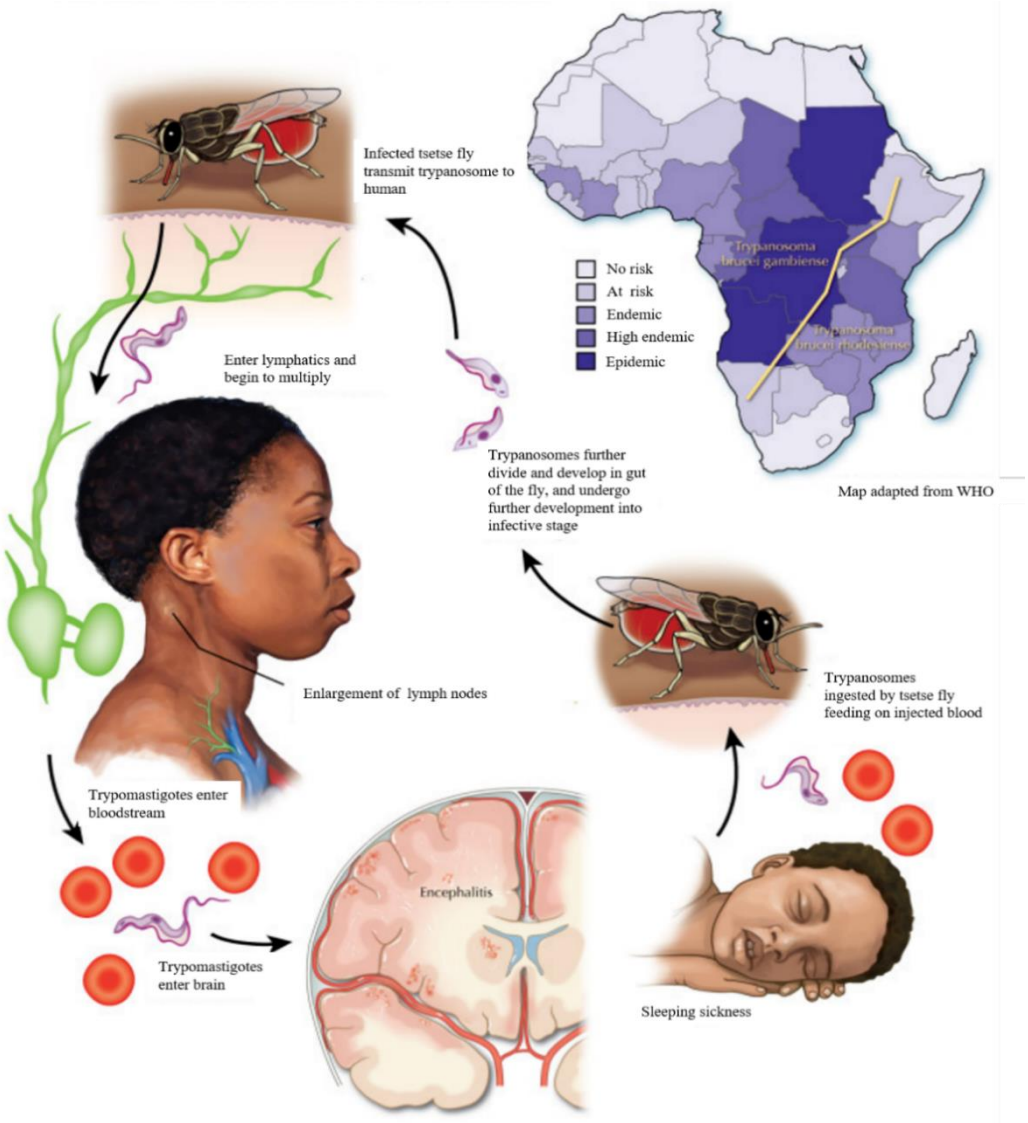


Figure 1.1 Distribution of human African trypanosomiasis HAT and the life cycle of *T. brucei* parasites. (https://www.netterimages.com/)

1.1.2 Monoallelic *VSG* expression and *VSG* switching in *T. brucei*

The biggest challenge for African trypanosomiasis therapy is the relapse due to chronic infection. The human innate immune system is usually effective in eradicating a significant portion of *T. brucei* pathogens upon initial infection. However, this parasite

regularly undergoes antigenic variation to evade the host immune system through switching the expression of its major surface antigen, Variable Surface Glycoproteins (*VSG*s) (Marcello, *et al.* 2007). Although the human innate immune system decreases the bloodstream form of parasites initially, the left pathogens change a new surface glycoprotein to avoid attacks from the host immune system (Figure 1.2 A). Consequently, pathogens begin a new round of proliferation until the new antibodies appear. Such *VSG* switching leads to long-term infections and poor prognosis in African trypanosomiasis.

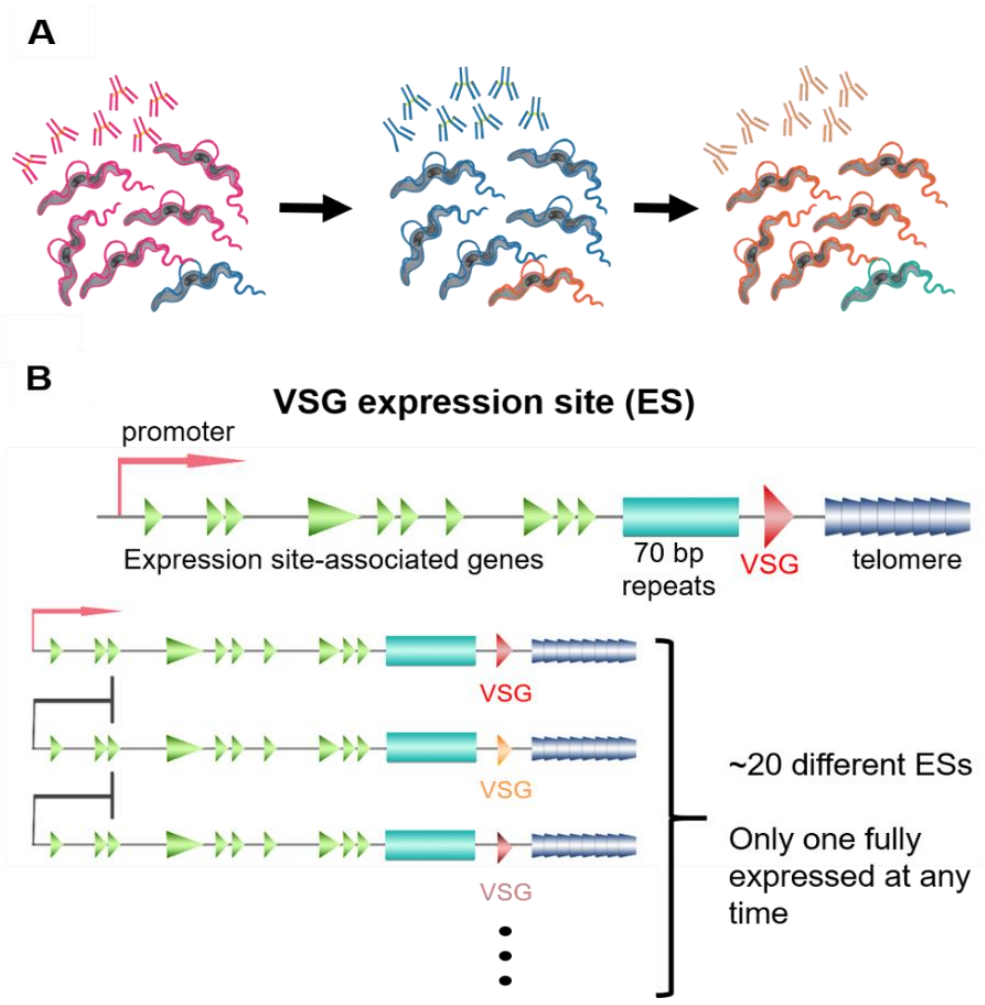


Figure 1.2 Monoallelic *VSG* expression in *T. brucei.* (A) *T. brucei* evades from the host immune system by regularly switching *VSG* (change from one color to another).

(B) *VSG* expression sites are located at subtelomeric region nearby telomere. At any time, only one *VSG* gene is well expressed.

As a major surface component, there are more than 2,500 complete or truncated *VSG* genes within *T. brucei* genome (Cross*, et al.* 2014). When the parasite resides in its insect vector, all *VSG*s are silenced. In human hosts, however, *VSG*s are transcribed by RNA polymerase I from subtelomeric *VSG* expression sites (ESs) located on the megabase chromosomes(Hertz-Fowler*, et al.* 2008) (Horn*, et al.* 2005). The *VSG* genes are always the last gene in any ES located within 2 kilobase (kb) from the telomere repeats, while the ES promoter is located at about 40 to 60 kb upstream of the *VSG* gene. *T. brucei* usually has multiple ESs which contain different *VSG* genes. For example, 13 different ESs has been identified in the *T. brucei* Lister 427 strain. And all of them share the same gene organization and keep 90% sequence similarity(Hertz-Fowler*, et al.* 2008; Müller*, et al.* 2018). At any moment, only one ES is fully active, resulting in a single type of *VSG* being expressed on the cell surface. The transcription status of ESs between activation and inactivation determines the monoallelic *VSG* expression (Figure 1.2 B). This monoallelic *VSG* expression ensures that after a *VSG* switch, the originally active *VSG* is no longer expressed on the cell surface, which contributes to its long-term infection pattern. Therefore, both *VSG* switching and monoallelic *VSG* expression are essential mechanisms for *T. brucei* pathogenesis.

1.1.3 *T. brucei* occurs transcriptional switch and DNA recombination during *VSG* switching

VSG switching occurs through two major pathways. One involves transcriptional switch and the other DNA recombination (Li 2015) (Fig. 1.3). During an in situ transcriptional switch, the originally active ES is silenced while an originally silent ES

is expressed (Cross, *et al.* 1998; Horn, *et al.* 1997; Zomerdijk, *et al.* 1992). DNA recombination-mediated *VSG* switching can occur in several ways, but crossover (CO) and particularly gene conversion (GC) are more frequent events. In CO, the active *VSG* gene can exchange place with a silent *VSG* gene in a different ES. In GC, a silent *VSG* gene is duplicated into the active ES to replace the originally active *VSG*, which is lost after the switch. The *VSG* donor in GC switches can originate from a silent ES, a minichromosome subtelomere, or a *VSG* gene array (Li 2015). Therefore, *VSG* switching frequency can be influence by factors related to homologous recombination, DNA damage repair(McCulloch, *et al.* 2015). Additionally, several telomere proteins suppress *VSG* switching(Jehi, *et al.* 2014; Jehi, *et al.* 2014; Yang, *et al.* 2009).

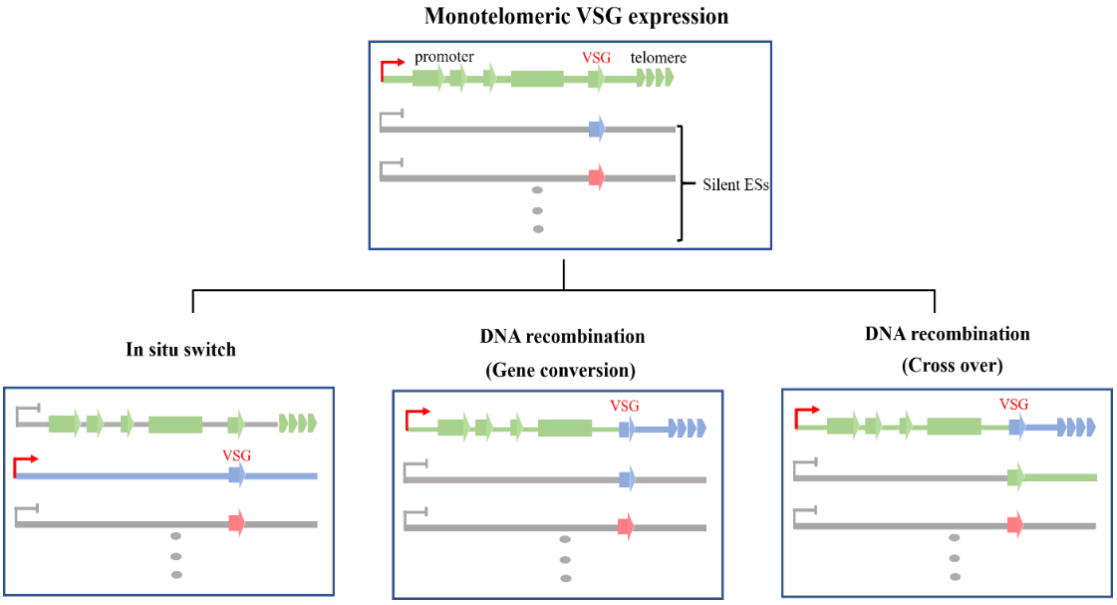


Figure 1.3 *VSG* switching mechanisms in *T. brucei*.

1.2 Structure and function of telomere

Since *VSG* genes are specifically located close to telomere region, telomere architecture needs to be opened during *VSG* switching. Therefore, telomere stability is important for maintaining monoallelic *VSG* expression.

Telomere is a nucleoprotein complex located at the two ends of a linear eukaryotic chromosome consisting of repetitive telomeric DNA sequences and a variety of telomere-binding proteins (Figure 1.4). Telomere protects chromosome from inappropriate degradation, recombination, and fusion. Consequently, telomere acts like caps at the end of shoelaces to maintain genetic information integrity (Jafri*, et al.* 2016).

1.2.1 Telomeric DNA consists of telomeric double-stranded DNA and 3' single-stranded overhang

Telomeric DNA contains a long double-stranded region with TG-rich, non-coding, tandem repeats followed by a short 3' single-stranded overhang. The duplex region forms a double-stranded helix structure based on the principle of complementary base pairing. 3' single-stranded overhang is known as G-tail or G-overhang because of the rich G base.

To protect vulnerable telomere terminus, the 3' single-stranded overhang is inserted into a homogeneous double-stranded region, forming a T-loop structure (Figure 1.4) (De Lange 2005). T-loop shields the exposed chromosome end from recognition by the DNA damage response (DDR) machinery. As a secondary DNA structure, T-loop was firstly discovered by electron microscopy in 1999 (Griffith*, et al.* 1999). The size of the T-loop varies among different species ranging from 1 to 25 kb (Xu 2011).

3' single-stranded overhang can also fold as a G-quadruplexes. In this secondary structure, four 5'-GMPs are self-assembled into a square planar structure by hydrogen bonds, known as G-tetrad. Afterward, stacking of several G-tetrads form G-quadruplexes. G-quadruplexes formation depends on the presence of alkali metal ions, mainly Na^+ and K^+(Parkinson*, et al.* 2002; Redon*, et al.* 2003). G-quadruplexes are

more likely to occur when DNA replication is slow (Schaffitzel, *et al.* 2001). The function of G-quadruplexes formation is not clear yet. Because of high stability, G-quadruplexes are difficult to resolve during replication, which in turns, stalls replication forks. On the other hand, the G-quadruplex structure may play a role in protecting telomere architecture to avoid degradation caused by nuclease(Lin, *et al.* 2013).

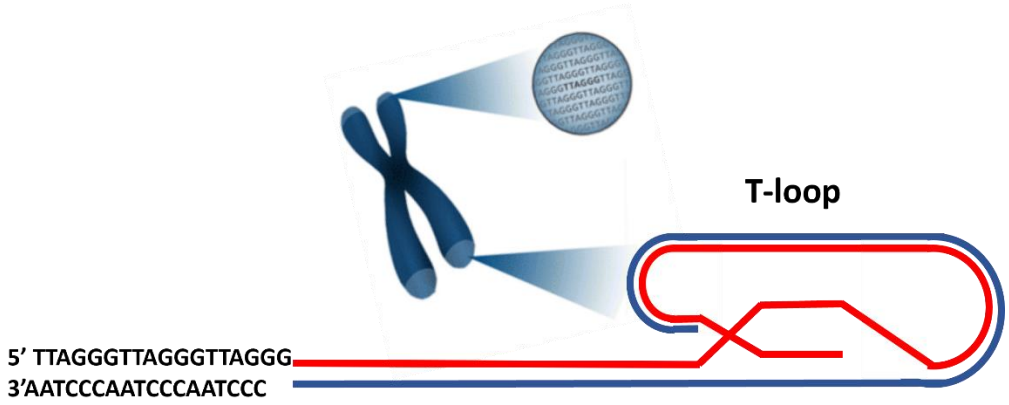


Figure 1.4 Structure of telomere in vertebrates. Telomeric DNA contains dsDNA with "TTAGGG" repeats and 3' single-stranded overhang in mammalian and *T. brucei.* The 3' single-stranded overhang forms a T-loop structure after inserted into a homogeneous double-stranded region.

Telomeric DNA is characterized by a repeated sequence in different species: "TTGGGG" repeats in *Tetrahymena* and "TTTTGGGG" repeats in *Oxytricha*. While "TTAGGG" repeat is the specific sequence in mammalian and *T. brucei* (Blackburn, *et al.* 2006).

1.2.2 Telomere shortens after each DNA replication cycle

The normal length of mammalian telomere is about 7-20 kb pairs. However, due to the "end replication" problem, the length of telomeric DNA has been predicted to become progressively shortened after each DNA replication cycle(Xu 2011).

Chromosomes are copied by semi-conservative replication machinery with a fixed direction from 5’ to 3’. The leading strand can be fully replicated from 5’ to 3’ as a template. While the lagging strand is replicated discontinuously because of the limit of replication direction. RNA primers bind to lagging strand and recruit DNA polymerases to elongate fragments, which forms Okazaki fragments. Then RNA primers are degraded and replaced by DNA under the function of DNA polymerases. While RNA primer of most distal Okazaki fragment cannot be replaced after degrading. Because DNA polymerases only recognize 3’-OH as a start site for nucleotide addition (Figure 1.5). Consequently, a small segment at the end will be lost during cell division, which results in replication-associated telomere shortening (Palm, *et al.* 2008; Shay, *et al.* 2000).

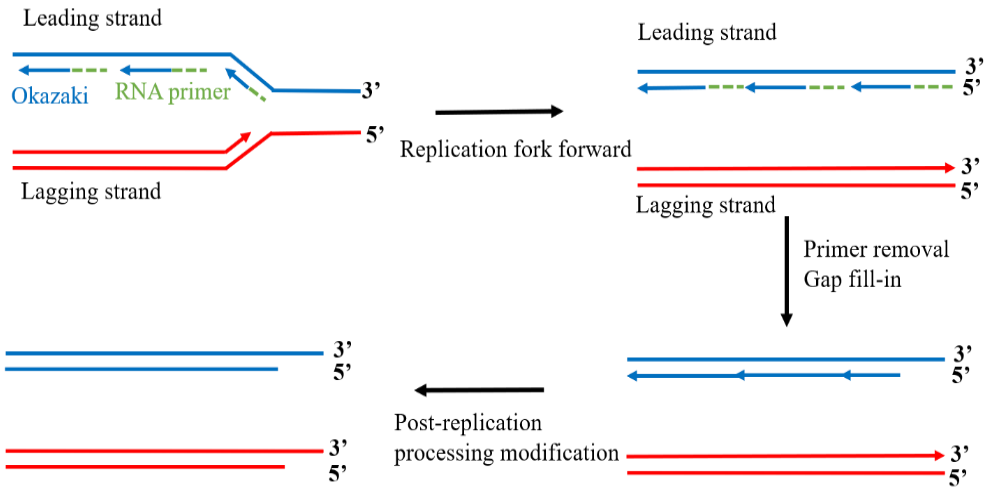


Figure 1.5 Telomere shortening and “the end replication” problem.

Telomere length is one of the key biomarkers of senescence. This process determines the loss of about 200 nucleotides for each cell cycle and when telomeres reach a threshold length(Teixeira, *et al.* 2004). Critically short telomere will be detected by DNA damage repair machinery. Afterward, cells with critically short telomere enter

replicative senescence and finally undergo apoptosis(Karlseder, *et al.* 2002). To counteract telomere shortening during each cell division, telomerase performs the function to add telomere repeats and maintain telomere length(De Vitis, *et al.* 2018).

1.3 Shelterin complex in mammalian system and *T. brucei*

1.3.1 Shelterin complex in mammalian system is important for telomere protection

Telomeric protein is another important component of telomere. In addition to maintaining telomere length, the telomeric DNA ends need to be protected from deterioration or end fusion. For this purpose, a group of proteins tightly associate with telomeric DNA and form the so-called Shelterin complex that physically shields the DNA ends from the vigilance of various DNA repair mechanisms.

In mammalian system, the telomeric protein complex contains TRF1, TRF2, RAP1, TIN2, POT1, and TPP1 (Figure 1.6). Three of them can directly interact with telomeric DNA for telomere localization. TRF1 and TRF2 bind to the duplex TTAGGG repeats (Fairall, *et al.* 2001) (Palm, *et al.* 2008). While TPP1/POT1 heterodimer interacts with 3' single-stranded telomeric repeats(Xin, *et al.* 2007). As a core component, TIN2 is located at the central position of the Shelterin complex and acts as a bridge to connect with other proteins, such as TRF1, TRF2, and TPP1 (Chen, *et al.* 2008; Kim, *et al.* 1999; Lei, *et al.* 2004) (Figure 1.6). However, RAP1 is the least understood protein within the Shelterin complex. Only TRF2(Li, *et al.* 2000) and phosphorylated peptides(Leung, *et al.* 2011) were known as its binding partners.

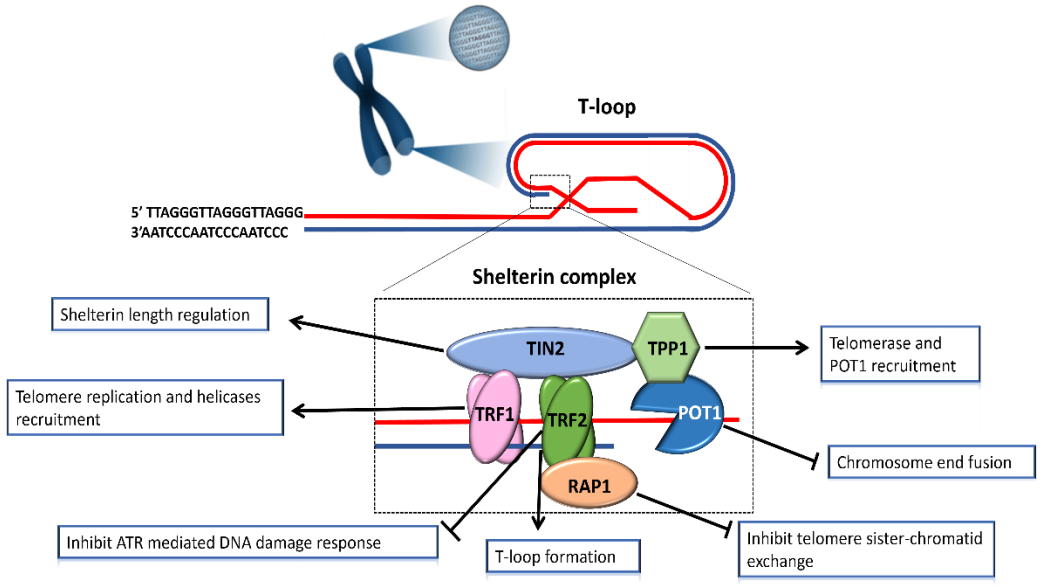


Figure 1.6 Shelterin complex in mammalian system. Mammalian Shelterin complex consists of six components (Ishikawa 2013). The six components within mammalian Shelterin complex interact with each other to maintain telomere stability.

Shelterin complex interaction with telomeric DNA plays a role in reducing DNA damage signaling and DNA repair reactions to maintain chromosome stability and integrity. Removal of Shelterin from mice causes chromosome end protection problems (Sfeir, *et al.* 2012). Many functional studies focused on single Shelterin member have been carried out. For example, researchers have found that TRF1 protects telomeres during semiconservative replication, through recruiting helicases to break secondary structures formed by the G-rich strand (Sfeir, *et al.* 2009). While TRF2 facilitates the formation of T-loops to maintain telomere stability(Doksani, *et al.* 2013). In addition, TRF2 can bind to the ATM kinase and block the ATM pathway efficiently (Denchi, *et al.* 2007). RAP1/TRF2 complex has been proved as an indispensable component for blocking the Nonhomologous End-Joining (NHEJ) pathway(Bae, *et al.* 2007). Downregulating POT1 of humans results in telomere dysfunction and causes

chromosome end fusion increase(Hockemeyer, *et al.* 2005).

1.3.2 Shelterin proteins in *T. brucei* play critical roles in *VSG* regulation

It is well-known that DNA double-strand breaks (DSBs) occur frequently at subtelomere in *T. brucei*. Fragile subtelomeric region increases *VSG* switching frequency and contributes to *T. brucei* survival, because of the specialized location of *VSG* gene. At the same time, *T. brucei*, like other organisms, needs to stabilize telomere/subtelomere architecture to keep genome integrity. Therefore, telomere region of *T. brucei* maintains a dynamic balance between stability and plasticity.

Like mammalian system, *T. brucei* also contains telomeric repetitive "TTAGGG" sequence and a Shelterin complex located at telomere to maintain stability of chromosome. The telomeric DNA of *T. brucei* contains TTAGGG-repeats (~15 kbp) (Dreesen, *et al.* 2007) with terminal T-loop structures (~1 kbp). So far, three components of Shelterin complex have been discovered in *T. brucei*, including *Tb*TRF, *Tb*RAP1, and *Tb*TIF2 (Figure 1.7 A).

Like other telomere proteins, telomeric proteins in *T. brucei* have a similar chromatin end protection function. Since *VSG* is specifically expressed from subtelomeric loci that immediately adjacent to the telomeres, those telomere-specific proteins are supposed to be critical for *VSG* variation regulation.

1.3.2.1 *T. brucei* TTAGGG repeat-binding factor (*Tb*TRF)

T. brucei TTAGGG repeat-binding factor (*Tb*TRF), is the first identified Shelterin complex component in *T. brucei*. *Tb*TRF is an integral telomeric protein throughout the

cell cycle. As a homolog of mammalian TRF2, *Tb*TRF has two functional domains including the N-terminal TRFH domain and C-term Myb domain (Figure 1.7 B). Specifically, *Tb*TRF keeps high sequence similarity to TRF2 in C-terminal Myb domains while has low similarity in their N-terminal domains.

*Tb*TRF associates with itself and forms homodimers through its TRFH domain. In previous works, our lab determined the structure of Myb domain as a canonical helix-loop-helix structure conserved to the Myb domains of mammalian TRF proteins (Jehi, *et al.* 2014). *Tb*TRF Myb domain directly binds to the telomeric duplex DNA region in a similar way to its homologs. Specifically, in *T. brucei*, this protein-DNA interaction contributes to telomere G overhang stability and consequently suppresses *VSG* switching *via* gene conversion (Li, *et al.* 2005).

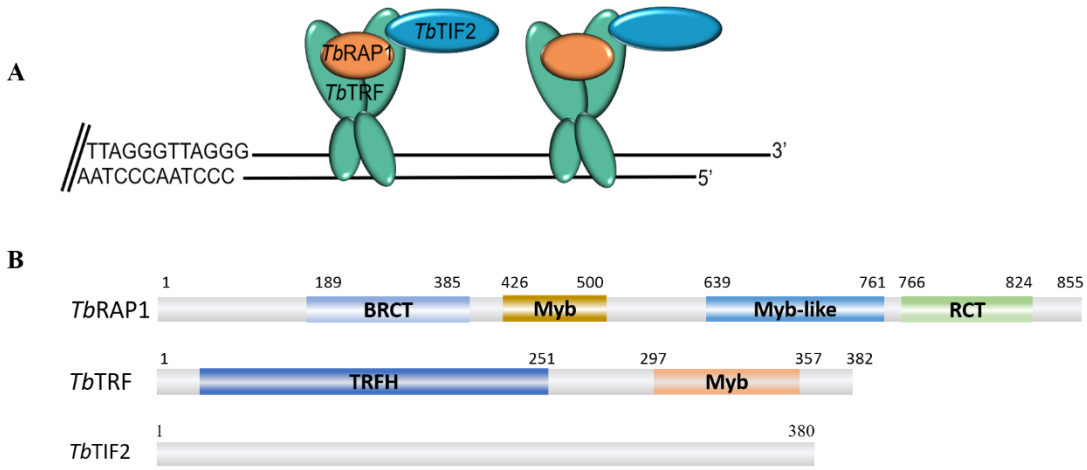


Figure 1.7 Shelterin complex in *T. brucei*. (A) Shelterin complex in *T. brucei* is consisted of three components, including *Tb*TRF, *Tb*TIF2, and *Tb*RAP1, respectively. (B) Functional domains of three telomeric proteins in *T. brucei*.

1.3.2.2 *T. brucei* Repressor/Activator Protein 1 (*Tb*RAP1)

T. brucei Repressor/Activator Protein 1 (*Tb*RAP1), the second Shelterin member identified in *T. brucei*, is a homolog of mammalian RAP1. *Tb*RAP1 was discovered through a yeast-protein interaction screen, using *Tb*TRF as bait. In immunofluorescence (IF) analyses, *Tb*RAP1 partially co-localizes with *Tb*TRF. Furthermore, the interaction between *Tb*RAP1 and *Tb*TRF was confirmed by Co-immunoprecipitation (CoIP) experiment. Therefore, *Tb*RAP1 is supposed to be a *Tb*TRF-binding partner. [24].

As a major regulator for silencing *VSG* expression, knockdown of *Tb*RAP1 led to derepression of all *VSG*s in silent expression sites (Yang*, et al.* 2009). Depletion of *Tb*RAP1 also leads to accumulation of telomeric RNA:DNA hybrids, telomeric/subtelomeric DNA damage, and higher *VSG* switching frequency(Nanavaty*, et al.* 2017). These findings suggest that *Tb*RAP1 serves as a telomere/subtelomere protector and further suppressing *VSG* expression.

1.3.2.3 *T. brucei* TRF-Interacting Factor 2 (*Tb*TIF2)

T. brucei TRF-Interacting Factor 2 (*Tb*TIF2), a functional homolog of Tin2, is identified as the third Shelterin complex member in 2014. N terminus of *Tb*TIF2 interaction with *Tb*TRF TRFH domain has been confirmed *in vivo*. Depending on this protein-protein interaction, *Tb*TIF2 is recruited to telomere and executes its suppressor role in *VSG* switching. Furthermore, *Tb*TIF2 stabilizes *Tb*TRF protein levels by inhibiting their degradation by the 26S proteasome.

*Tb*TIF2 maintains telomeric region integrity by reducing replication fork stalling

formation. More importantly, an accumulated amount of DSBs were detected in the *Tb*TIF2 depletion strain (Jehi*, et al.* 2014). *Tb*TIF2 is the first telomeric protein that shows influence on the amount of subtelomeric DSBs, indicating that *Tb*TIF2 maintains subtelomere integrity. Subsequently, subtelomeric DSBs accumulation led to a higher *VSG* switching frequency and eventual cell growth arrest, which suggests its importance on switching regulation and cell viability.

In summary, these findings demonstrate that these three telomere-specific proteins are indispensable for maintaining telomere and subtelomere region stability, as well as regulating *VSG* expression.

1.3.3 *Tb*RAP1: a critical protein in regulating *VSG* switching

*Tb*RAP1, a RAP1 ortholog in *T. brucei*, has been identified by our collaborator Prof. Li Bibo in 2009 (Yang*, et al.* 2009). *Tb*RAP1 comprises four functional domains including the N-terminal BRCT domain, Myb domain, Myb like domain, and an RCT domain, respectively (Figure 1.7 B) (Afrin*, et al.* 2020). RAP1 is the most conserved telomeric protein component, with homologs identified from protozoa to mammals (Figure 1.8 A).

Like other RAP1 orthologs, *Tb*RAP1 executes varied functions through interacting with different binding partners. Specifically, the BRCT domain is responsible for *Tb*RAP1 self-interaction. Depletion of BRCT domain leads to low *Tb*RAP1 level detected by western bolts. It is possible that *Tb*RAP1 self-interaction contributes to preventing *Tb*RAP1 degradation by proteases. *Tb*RAP1 Myb domain is required for interacting with *Tb*TRF. Depletion of Myb domain causes cell growth arrest and *VSG* derepression, indicating Myb domain is also important for normal cell proliferation and

VSG silencing. Additionally, *Tb*RAP1 interacts with importin α through its nuclear localization signal (NLS) within MybLike domain to localized in nuclear rather than the cytoplasm. RCT domain in other RAP1 orthologs is usually served as a protein-protein interaction module. With only 11.7% sequence identity between *Tb*RAP1 RCT domain and other RCTs, the function of *Tb*RAP1 RCT domain is still unknown (Afrin*, et al.* 2020).

*Tb*RAP1 association with telomeric DNA has been confirmed by Chromatin immunoprecipitation (ChIP) assay. But the mechanism of *Tb*RAP1 targeting telomere still needs further studies. To localize telomere, some RAP1s bind to telomeric DNA directly while others are recruited through interaction with nuclear proteins (Figure 1.8 B). For example, human RAP1 (hRAP1) has no binding activity with telomeric DNA, because of its negatively charged surface (Hanaoka*, et al.* 2001). Instead, hRAP1 is recruited to telomeres through interaction between its RCT domain and TRF2 (Rai*, et al.* 2016). A similar pattern has been found in *Schizosaccharomyces pombe* RAP1 (*Sp*RAP1). *Sp*RAP1 is recruited to telomere by binding partner Taz1 (ortholog of mammalian TRF1 and TRF2) *via* its RCT domain (Kanoh, *et al.* 2001). So far, only budding yeast *Saccharomyces cerevisiae* RAP1 (*Sc*RAP1) with two Myb domains was discovered directly binding activity with double-stranded telomeric DNA for telomere localization (Hanaoka*, et al.* 2001).

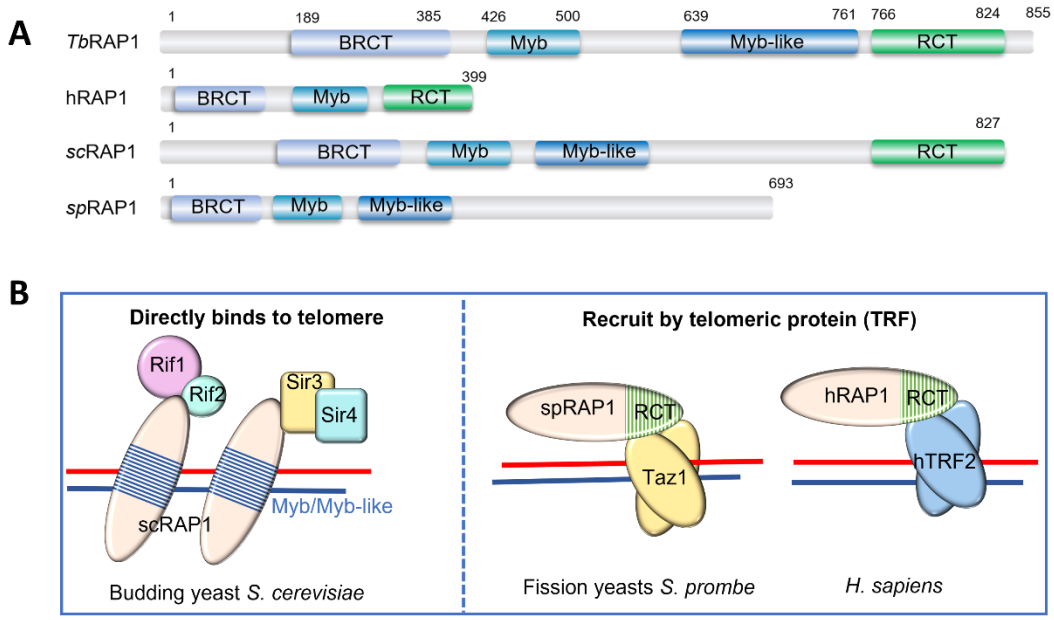


Figure 1.8 RAP1 is a highly conserved telomeric protein with homologs identified from protozoa to mammals. (A) Functional domains involved in RAP1 orthologs. (B) Two mechanisms within RAP1 orthologs to achieve telomere localization.

As an important component of Shelterin complex in *T. brucei*, many studies have confirmed that *Tb*RAP1 is essential for *VSG* silencing. Deficiency of *Tb*RAP1 causes a dramatic derepression of all ES-linked *VSG* genes more than 1000-fold. Consequently, simultaneously more than a single type of *VSG* expression can be detected at the *Tb*RAP1 depletion cell surface. In addition, *Tb*RAP1 can suppress telomeric *VSG* switching frequency mediated by gene conversion. Cell growth arrest also can be detected after removal of *Tb*RAP1, indicating its importance for cell proliferation(Yang, *et al.* 2009).

*Tb*RAP1 functions on *VSG* regulation could be mediated by telomeric transcript (TERRA). TERRA is specifically transcribed from active ES-adjacent telomeres instead of from silent ES-adjacent telomeres in *T. brucei* (Rudenko, *et al.* 1989). Recent works from our collaborator Prof. Li's lab reveals that the telomere downstream of an

active ES is transcribed into TERRA, presumably resulting from read-through beyond the *VSG* gene at the end of ES into the telomeric region. Genetic depletion of *Tb*RAP1 accumulates the amount TERRA level, indicating *Tb*RAP1 represses TERRA expression. An increased TERRA level leads to the higher telomeric RNA:DNA hybrids formation and causes more subtelomeric DSBs. As DSBs are strong triggers for *VSG* switching, more frequent *VSG* switching is also observed in *Tb*RAP1 depleted cells. (Nanavaty*, et al.* 2017). However, whether *Tb*RAP1 directly interacts with TERRA or has any RNA binding potential is still unknown.

1.4 Objectives

Telomere is important for *VSG* regulation. As a novel component within *T. brucei* Shelterin complex, *Tb*RAP1 is regarded as a critical regulator for *VSG* silencing. However, the underlying mechanism of *Tb*RAP1 suppression *VSG* switching is still unclear, especially the key structural domains for this important function have not been identified. To further investigate the role of *Tb*RAP1 in *VSG* switching, the following three objectives were carried out.

1.4.1 *Tb*RAP1 Myb-like region mediated DNA binding activity investigation and its role in *VSG* regulation

*Tb*RAP1 is a telomeric protein but also locates elsewhere in the genome(Pina*, et al.* 2003). How does *Tb*RAP1 localize to telomere for executing its *VSG* switching suppression role? As shown in Figure 1.7 B, *Tb*RAP1 contains a Myb domain and a Myblike domain that typically has DNA binding activities. Among RAP1 orthologs, some of them are recruited to telomere through strong interaction with TRF orthologs

such as hRAP1 and spRAP1. While scRAP1 directly binds to telomeric DNA to achieve telomere. Although *Tb*RAP1 binds to *Tb*TRF *in vivo*, the *Tb*TRF-*Tb*RAP1 interaction is more moderate compared to the *Tb*TRF-*Tb*TIF2 interaction.

Guided by these findings, we hypothesis that *Tb*RAP1 may have DNA binding activity and its protein-DNA interaction help *Tb*RAP1 achieve telomere localization. As *Tb*RAP1 Myb domain is required to interact with *Tb*TRF, it is unlikely this tiny region binds to both protein and DNA at the same time. Therefore, we first ruled out Myb domain as a potential DNA binding region. Consequently, we decided to focus our study on *Tb*RAP1 Myblike region's DNA binding activity instead.

To understand the function of *Tb*RAP1 Myb-like region, the DNA-binding activity of Myb-like region will be investigated *in vitro* and *in vivo*, respectively. The Electrophoretic Mobility Shift Assay (EMSA) and NMR ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) titration assay will be performed between purified Myb-like region protein and telomeric dsDNA/ssDNA for DNA-binding studies. Once the Myb-like region has been determined to have a DNA-binding function, Myb-like region mutants with weakened or abolished DNA binding will be generated to map critical residues/regions responsible for DNA-binding activity.

For *in vivo* studies, the functional importance of Myb-like domain DNA binding activity to *T. brucei* cell growth and *VSG* regulation will be tested by ChIP, IF analyses, and Co-IP. These experiments will be done in collaboration with Dr. Li's lab at Cleveland State University of Ohio, USA.

1.4.2 Structural studies of *Tb*RAP1 Myb-like region

*Tb*RAP1 Myb-like domain is proposed to be a canonical helix-loop-helix structure. To better investigate the role of *Tb*RAP1 Myb-like region in *VSG* switching, a structural determination study of Myb-like region will be carried out. We propose to use NMR spectroscopy to determine its three-dimensional structure. Because the size of the potential construct of this Myb-like region, about 14 kilodalton (KD), is well within the range of the NMR method. Isotope-labeled recombinant protein of *Tb*RAP1 Myb-like region will be obtained for NMR raw data collection. Once the structure of Myb-like region has been solved, a structural superimposition by software CCP4MG and WinCoot will be examined between *Tb*RAP1 Myb-like region and other similar domains.

1.4.3 Biochemical and functional characterization of potential RNA-binding activity of *Tb*RAP1 and its role in *VSG* regulation

Telomere site contains multiple types of RNA. The most abundant type is *VSG* mRNA, followed by RNA transcribed from ESAGs. There is no report of any RNA-binding activity for *Tb*RAP1. However, with its unique function in *VSG* silencing, it is possible *Tb*RAP1 has some novel RNA-binding activity that is specific for the *T. brucei*. Therefore, we plan to use *VSG* mRNA from the Lister 427 strain as substrate to investigate *Tb*RAP1's RNA binding activity. To detect *Tb*RAP's RNA binding sequence preference, random RNA will also be applied in experiments.

Among the multiple structural domains in *Tb*RAP1, the Myb and Myb-like domains typically function on nucleotides binding. But Myb domain of *Tb*RAP1 shares low sequence similarity to classic Myb domains and has been proved necessary for

*Tb*TRF-*Tb*RAP1 interaction. Therefore, we hypothesis that the Myb-like region, rather than the Myb domain within *Tb*RAP1, has the most potential to bind to RNA.

The interaction between Myb-like region and RNA substrates will be assessed by *in vitro* biochemical methods including EMSA and Fluorescence Polarization (FP) assay. Once RNA binding activity of *Tb*RAP1 Myb-like region is confirmed, the HSQC titration assay will be performed to identify key residues directly involved in *Tb*RAP1-RNA interaction. Finally, *in vivo* studies including ChIP, IF analyses, and Co-IP assays will be carried out to test the functional importance of *Tb*RAP1-RNA binding activity to *T. brucei* cell growth and *VSG* regulation.

2 Methodology

2.1 Gene cloning

2.1.1 Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR) was used to get target cloning genes. Full-length *Tb*RAP1 plasmid was obtained from our collaborator Dr. Li Bibo's lab. Afterward, target recombinant DNA molecules were cloned into certain vectors following the below steps.

(a) Design a pair of forward and reverse primers containing different restriction enzymes, respectively. Ensure these two restriction enzymes are available in the target expression vector.

(b) Amplification: amplify insert (recombinant DNA molecule) by PCR following protocol as below.

Reagents	**Volume for 1 reaction (μl)**
100 mM Forward primer	0.2
100 mM Reverse primer	0.2
5*HF buffer	4.0
10 mM dNTP	0.4
Phusion™ High-Fidelity DNA Polymerases	0.2
ddH_2O	15.5
Total	20.0

(c) The mixture was placed in the Thermal Cycler and the PCR procedure was set as followed. Melting temperature (Tm) was calculated based on the primer pair sequence.

Cycle Step	**Temperature**	**Duration(sec.)**
Initial Denaturation	98.0	180
Denaturation	98.0	30
Extension	72.0	45
Cycle	Go to step2	30 cycles
Final Extension	72.0	600
Hold	4.0	Forever

(d) Then the agarose gel electrophoresis was performed for PCR product purification with a DNA extraction kit (BioTeke).

(e) Digestion: purified insert and empty expression vector were digested by certain restriction enzymes. Afterward, digested insert and empty vector were purified with DNA purification kit. And then, the Nanodrop was applied for measuring the concentration of well digested insert and vector, respectively.

(f) Ligation: ligating inserts and vector together following insert-vector molar ratio 3:1 to 5:1, with T4 ligase at room temperature for 30min. Finally, ligate product was prepared for later transformation.

2.1.2 Mutant generation

To generate mutants, the overlapping extension PCR cloning was conducted. It is comprised of two steps. Firstly, using wild-type constructs as templates, a pair of primers with mutant DNA sequence was used to amplify the target mutant gene. The 1st round PCR product was purified and then was applied as primers for the 2nd round PCR. Digest methylated template DNA by enzyme DpnI (Thermo Fisher Scientific) at 37°C for 12 hours. Afterward, digested PCR product was prepared for transformation.

2.1.3 Transformation

Transformation refers to the process that introduces exogenous DNA into competent cells as below steps.

(a) Add recombinant DNA molecules (PCR products) into competent cells and incubate for 20-30 min on ice.

(b) Heat shock competent cells at 42.5°C for 90 seconds. Cool down sample on ice for 5min.

(c) Add 500 µl LB media without antibiotics, incubating for 45-60 min at 37°C, 250 rpm.

(d) Spread cells into a pre-warm LB agar plate with the corresponding antibiotic and incubated overnight.

2.2 Recombinant protein expression in *Esherichia coli* (*E.coil*)

Firstly, single colonies were picked and transferred to LB broth with appropriate antibiotics, incubating at 37℃ until the OD_{600} reached 0.6-0.8. Afterwards, the expression of target protein was induced with 0.3 mM IPTG under 30℃ for 6 hours. Cells were harvested at centrifuge with 6500 rpm, 10 min.

Cell pellets were lysed by lysis buffer with β-Mercaptoethanol (β-ME) and Phenylmethanesulfonyl fluoride (PMSF), followed by sonication (35% amplitude, 10 sec pulse, 12 sec pause, 15 min). Lastly, the sonicated samples were transferred to 50ml tubes and centrifuged at 4℃, 19,000 rpm, 2 hours. Filter supernatant with 0.22 μm membrane for protein purification process. Then SDS-PAGE was applied to test the expression level and solubility of the target protein. Staining with Coomassie brilliant blue, protein bands were visualized.

2.3 Isotopic labeled protein expression

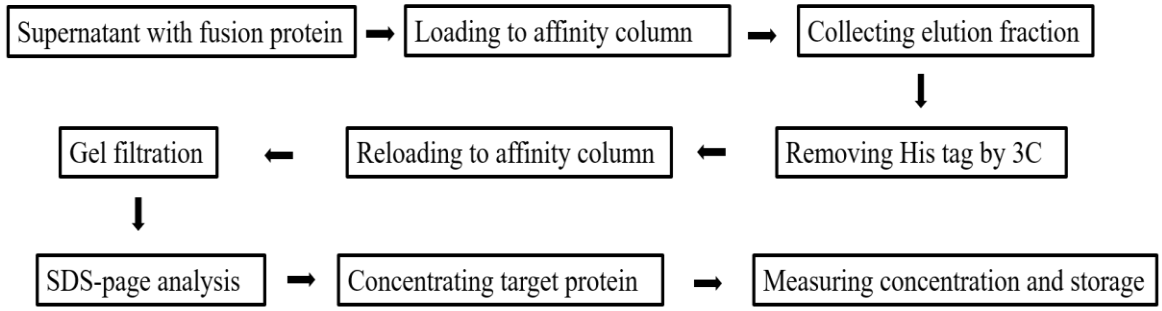
For collecting NMR spectrum, ^{15}N and/or ^{13}C labeled proteins were expressed with M9 minimal medium instead of LB medium. During expression, Ammonium Chloride (^{15}N, 98%+) and/or D-Glucose (U-$^{13}C6$, 99%) (Cambridge Isotope Laboratories, Inc.) were added as the sole nitrogen and/or carbon source. The component list of M9 medium is described as followed table:

Component	Amount (per liter)
$Na_2HPO_4 \cdot 7H2O$	12.8 g
KH_2PO4	3 g
1M MgSO4	2 ml
0.1M $CaCl_2$	1 ml
$^{15}NH_4Cl$	1 g
$^{12}C6/^{13}C6$-glucose	2 g
Minimal supplies	1 ml

^{15}N and $^{15}N/^{13}C$ isotope labeled proteins were prepared by growing *E. coli* BL21 (DE3) in M9 minimal medium. Culture cells for 10-12 hours under 37℃ until the OD_{600} value reached 1.0-1.2. Afterward, induced with 1.0 mM IPTG under 30℃ for 5-6 hours and harvested cells with centrifuging 6500 rpm, 10 min. Optimized lysis buffer were used for sonication for preparing supernatant in the protein purification process.

2.4 Protein purification

ÄKTA purifier (GE Healthcare) was used to purify target proteins. The flow chart of general purification process was described as followed:



2.4.1 Affinity chromatography

The affinity column (5 ml His Trap HP column, GE Healthcare) was washed by Milli Q water with 5 Column Value (CV) to remove ethanol and then equilibrated with 5 CV binding buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 40 mM imidazole). The supernatant with dissolved recombinant protein was then loaded into the His Trap HP column using Super-loop (GE Healthcare) with a flow rate of 1 ml/min. After loading supernatant, the column was washed with binding buffer for 10-20 CV with a flow rate of 2 ml/min to remove non-specific binding. Finally, recombinant protein was washed out by his-elution buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, 500 mM imidazole) with a flow rate of 3 ml/min. Fractions were collected during elution.

2.4.2 Removal of the fusion tag

Recombinant protein with a hexa-histidine (His6) tag was concentrated by concentrator (Amicon® Ultra Centrifugal Filters), followed by exchanging into Tris buffer (50 mM Tris, pH 8.0, 150 mM NaCl). Afterward, 1% (w/w) 3C protease was added to remove the His6 tag and its fusion partner. The digestion reaction was performed at 4℃ for 4 hours.

After digestion, the fusion protein was divided into 4 parts including target protein, His6 tag with fusion partner, incompletely digested fusion protein, as well as 3C protease. This mixture was injected into a 5 ml HisTrap HP column again to separating components with tag without tag. Theoretically, the target protein and 3C protease

would be wash out by Tris buffer, while incompletely digested fusion protein, His tag and its fusion partner would be wash out by his-elution buffer. The fraction washed out by Tris buffer was then collected and concentrated by a protein concentrator.

2.4.3 Gel filtration chromatography

The size exclusion column (HiLoad 16/600 Superdex 75 PG column, GE Healthcare) was equilibrated by Tris buffer. And then protein sample was injected into column with a flow rate of 1 ml/min. Proteins would be washed out based on their molecular weight from large size protein to the small one. Fractions were collected when UV absorbance was increased. Take out 10 µl from each collected fraction for SDS-PAGE analysis to track target protein.

2.4.4 Concentration measurement

(a) Measuring OD_{280} by Nanodrop (Thermo Fisher Scientific).

Measuring the absorbance at 280 nm is one of the most convenient and common methods to determine the concentration of proteins with tyrosine and tryptophan residues. Due to the relationship between absorbance and protein concentration is linear, the concentration of protein with a known extinction coefficient can be easily calculated.

(b) Bradford protein assay.

Bradford protein assay is a method that measuring absorbance at 595nm to detect the concentration of complex of Coomassie Brilliant Blue G-250 dye with protein. A standard curve was conducted by bovine serum albumin using the spectrophotometer

(GE Healthcare). Based on the generated standard curve, the concentration of the unknown protein sample will be determined.

2.5 Biophysical and biochemical analysis assays

2.5.1 EMSA assay

EMSA assay is a wildly used method to investigate the interaction between protein and nucleic acid. The principle of EMSA is based on the observation that protein-DNA/RNA probe complexes migrate slowly during non-denaturing polyacrylamide gel electrophoresis, comparing to the unbound (free) DNA/RNA probe. When a specific nucleic acid probe and protein are mixed and incubated, protein that binds to the nucleic acid probe will form protein-DNA/RNA probe complexes. Due to large molecular weight, complex migration is slow during polyacrylamide gel electrophoresis.

Purified varied amounts of recombinant proteins were incubated with 0.5 ng of radiolabeled DNA/RNA probe in 15 µl HEPES buffer at room temperature for 30 min. 1.5 µl EMSA loading dye (50% glycerol, 0.1% bromophenol blue, and 0.1% xylene cyanol) was added to each sample before it was electrophoresed in a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer. Drying gel followed by exposing gel to a phosphorimager. HEPES buffer: 20 mM HEPES, pH 8.0, 100 mM KCl, 1 mM $MgCl_2$, 0.1 mM EDTA; 5% glycerol; 1 mM DTT and bovine serum albumin (BSA; 100 ng/µl).

2.5.2 FP assay

FP assay is another technique for protein-ligand interaction investigation. When a fluorescently labeled molecule is excited by polarized light, its degree of polarization is inversely proportional to the rate of the molecule rotation. Based on this characteristic, FP can be used to measure the interaction of a small labeled ligand with a larger protein. Free small ligand emits depolarized light because of rapid rotation. While protein-ligand complex formation leads to polarized emission benefiting from a slow rotation of large molecular (Figure 2.1) (Moerke 2009). As a wildly used interaction investigation method, FP is safer than EMSA because of the non-radioactive materials used. Another key advantage of FP is lower amounts of sample required and more sensitive, comparing to HSQC titration.

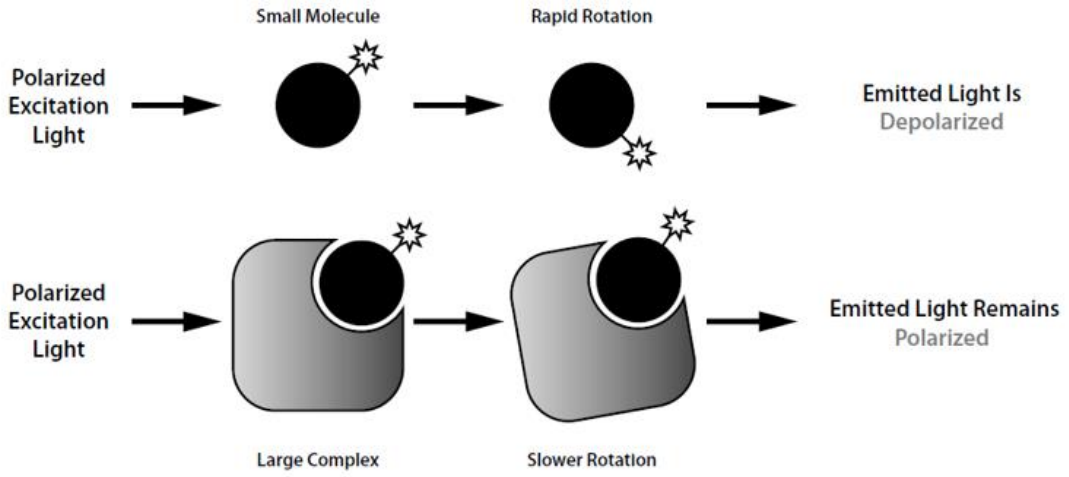


Figure 2.1 Principle of FP assay.

For detecting binding activity between *Tb*RAP1 and RNA oligos, purified proteins were dissolved in an optimized buffer. Different RNA oligos with 5' 6-carboxyfluorescein (6-FAM) were purchased from Integrated DNA Technologies. As a

single isomer derivative fluorescein, FAM is the most common fluorescent dye attachment for oligonucleotides, with an absorption wavelength of 495 nm and an emission wavelength of 517 nm. Mix 20 nM 5'6-FAM oligo with varied concentration protein together, followed with 30 min incubation at 4°C. The mixtures were transferred into a black 384-well plate. FP signals were collected at 21°C using a CLARIOstar (BMG LABTECH, Germany) multi-mode microplate reader. Choose polarization filters with excitation and emission wavelengths at 482 nm (482-16 mode) and 530 nm (530-40 mode) for data collection, respectively. Meanwhile, the adjusted polarization was set as 10 milli polarization (mP) to a reference.

2.5.3 NMR ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) titration

^{1}H-^{15}N HSQC is one type of 2-dimension spectrum, which implicates a correlation between single nitrogen and its neighboring hydrogen. ^{1}H-^{15}N HSQC shows the fingerprint of a protein, as well as correlates amino acid residues in the primary protein sequence. Not only used to determine protein structure, but ^{1}H-^{15}N HSQC also can be applied to detect protein-peptide, protein-protein, and protein-nucleic acid binding activity.

To investigate the interaction between *Tb*RAP1 and DNA, ^{15}N-labeled *Tb*RAP1-ML and *Tb*RAP1-ML-5A were prepared by growing *E. coli* BL21 (DE3) in M9 minimal medium. Cells were cultured for 10-12 hours under 37°C until OD_{600} reached 1.0-1.2. Afterward, protein expression was induced with 1.0 mM IPTG at 30°C for 5-6 hours. Cells were then harvested by centrifugation at 6500 rpm for 10 min.

The ^{1}H-^{15}N HSQC titration experiments were performed and analyzed using an AVANCE III 700 NMR spectrometer (Bruker) at 298K. The concentration of ^{15}N-labeled purified proteins was diluted to 0.1 mM in the buffet condition of sodium phosphate buffer (20 mM sodium sulfate, pH 6.5, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT) containing 10% D_2O. The information about three types of DNA substrates was described as below:

Telomeric dsDNA: 5′ TTAGGGTTAGGGTTAGGG 3′, 18bp

Random dsDNA: 5′ TGTTGAGGAGGTGGTGAT 3′, 18bp

Random ssDNA: 5′ TGTTGAGGAGGTGGTGAT 3′, 18nt

To investigate the interaction between *Tb*RAP1 and RNA, ^{15}N-labeled *Tb*RAP1-ML, *Tb*RAP1-ML-2FL, *Tb*RAP1-RRM were cultured in M9 minimal medium. The ^{1}H-^{15}N HSQC titration experiments were performed and analyzed using an AVANCE III 700 NMR spectrometer (Bruker) at 298K. The final concentration of ^{15}N-labeled purified proteins was 0.1 mM in the sodium phosphate buffer (20 mM sodium sulfate, pH 6.5, 150 mM NaCl, 1 mM EDTA, and 1 mM DTT) containing 10% D_2O. The information about three types of DNA substrates was described as below:

34-nt *VSG* mRNA: 5′AAAACUUUUUGAUAUAUUUUAACACCAAAACCAG 3′.

35-nt random RNA: 5′ACAGACCACUACAAGAUACACAGUACAACCAACCA 3′.

2.6 Structure analysis

2.6.1 Mass Spectroscopy (MS)

As a high-resolution technique for identifying Molecular Weight (MW), mass spectrometry has been widely used for protein characterization. The sample is ionized as multiple ions by ion source and introduced to mass analyzer. The mass analyzer separates the multiple ions based on their mass-to-charge ratio (m/z). Finally, these signals can be recorded by an ion detector and analyzed by computer software.

With the rapid development of analytical chemistry, mass spectrometry technique not only can be used to determine MW but also protein modifications analysis and fragment sequence identification after digested into small peptides.

To confirm the molecular weight of purified *Tb*RAP1-ML, the total 20 μL protein sample was prepared in Tris buffer (50 mM Tris, pH 8.0, 150 mM NaCl) with 0.1 mM final concentration. Agilent 6460 Liquid Chromatography-Electrospray Ionisation Triple Quadrupole Mass Spectrometer was used to collect and analyze MS spectra of *Tb*RAP1-ML.

2.6.2 Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is a reliable technique for identifying biomolecular structures. Comparing to crystallization, the sample for NMR spectroscopy is easier to prepare, which only requires to dissolve well in the liquid state. With data of multi-dimension NMR, protein structure and conformation can be determined.

Atom nuclei consist of protons and neutrons and each of these particles can spin. Different types of nuclear spin quantum numbers are different and can be divided into three situations. 1) When the sum of protons and neutrons are even, the spin quantum number is zero; 2) When the total particles are odd, the spin quantum number is a half integer, e.g., ^{1}H, ^{13}C, and ^{15}N; 3) When the total particles are even, the spin quantum number is an integer. Nuclei with half integer spin quantum number are active to absorb energy and cause energy transition. Putting sample with NMR active nuclei in a strong external magnetic field, amount of absorbed external energy can be recorded and used in the calculation for structural information. The differences in resonance frequencies are transfer into chemical shifts.

To determine the structure of *Tb*RAP1-ML, we prepared isotopic labeled protein in M9 minimal medium, with Ammonium Chloride (^{15}N, 98%+) and/or D-Glucose (U-$^{13}C6$, 99%) (Cambridge Isotope Laboratories, Inc.) as the sole nitrogen and/or carbon source. Afterwards, a series of NMR data were collected by the Varian Inova 750- and 800-MHz spectrometers. Specifically, to obtain backbone information, the spectra for HSQC(N-H), HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, HNCO, HN(CA)CO were collected. The side-chain resonances were determined through spectrum analysis from the ^{15}N TOCSY-HSQC, HCCH-TOCSY, HCCH-COSY, H(CC)(CO)NH, HBHA(CO)NH and (H)CC(CO)NH. Afterwards, the inter-proton distance restraints were confirmed by results of ^{1}H,^{15}N-NOESY-HSQC, ^{1}H,^{13}C-NOESY-HSQC, 3D

^{1}H,^{13}C-HSQC-NOESY-^{1}H,^{13}CHSQC, as well as 4D ^{1}H,^{13}C-HSQC-NOESY-^{1}H,^{13}C-HSQC.

2.7 Cell-based assays

2.7.1 ChIP assay

ChIP assay is a typical methodology to investigate protein/DNA association *in vivo*. It is wild used for confirming protein functions involved in nuclear processes such as transcription, replication, and DNA repair response. ChIP assay can be divided into 4 steps: (1) Cross-linking protein and DNA under treatment by 1% formaldehyde for 20 min at room temperature with constant mixing; (2) Sonicating for six cycles (each 30 seconds on and 30 seconds off) to isolate the DNA fragment about 500 bp on average; (3) each sample was aliquoted into several fractions, incubating with antibodies to detect protein, or immunoglobulin G (IgG) as negative control respectively; (4) Eluting immunoprecipitated products from the beads, and isolating DNA from the products followed by Southern blot hybridization or qPCR analysis.

2.7.2 IF analysis

IF analysis is a modern biology technique that analyzes the distribution of molecules, specifically proteins. IF analyses procedures are described as followed: fixing cells with 2% formaldehyde, and permeabilizing in PBS buffer containing 0.2% NP-40. Afterward, PBS with 0.2% cold fish gelatin and 0.5% bovine serum albumin (BSA) was used for blocking under room temperature for 10 min. Cells were then incubated with the primary antibody and the secondary antibody, respectively. Cells

were then washed and stained by 0.5 g/ml DAPI. Images were taken by a Delta Vision Elite deconvolution microscope.

2.7.3 Quantitative reverse transcription PCR

Quantitative reverse transcription PCR (qRT-PCR) is a common technique that measures specific mRNA level to detect gene expression. The process of qRT-PCR can be divided into two steps, including reverse transcription and quantitative RT-PCR. Firstly, RNA was extracted from cells using kit RNAstat (TEL-TEST, Inc) and purified by RNeasy kit (QIAGEN). Reverse transcription transcribed RNA into complementary DNA by using M-MLV (Promega). Afterward, quantitative RT-PCR was performed to amplify specific DNA from the last step by using Bio-Rad iTaq SYBR Green Supermix with ROX according to the manufacturer's protocol.

2.7.4 RNA sequencing analysis

The Cre expression was induced by doxycycline in *Tb*RAP1$^{F/+}$, *Tb*RAP1$^{F/mut}$, cells for 30 hours before total RNA was isolated and purified through RNeasy columns (Qiagen). All RNA samples were run on a Bioanalyzer 2100 (Agilent Technologies) using the Agilent RNA 6000 Nano Kit to verify the RNA quality before they were sent to Novogene for library preparation and RNA high-throughput sequencing followed by bioinformatic analysis.

3 *Tb*RAP1 Myb-like region mediated DNA binding activity investigation and its role in *VSG* regulation

3.1 *Tb*RAP1 Myb-like region shows sequence-non-specific interaction with dsDNA/ssDNA

*Tb*RAP1 contains four functional domains, including BCRT at N-terminus, Myb, Myb-like, and C-terminus domain. As mentioned in the introduction part, the majority of Myb domains rely on DNA binding activity to achieve telomere localization. To investigate *Tb*RAP1's DNA binding activity, the construct MybLike domain [amino acids (aa) 639 to 761] was generated, named as *Tb*RAP1-ML. After obtained *Tb*RAP1-ML through the recombinant protein purification process, EMSA assay and HSQC NMR titration assay were performed to investigate *Tb*RAP1-ML's ssDNA/dsDNA binding activity.

3.1.1 EMSA data show that *Tb*RAP1-ML has sequence non-specific dsDNA/ssDNA binding activities

Purified Trx-His6-tagged *Tb*RAP1-ML was applied for EMSA experiments. Different amounts of protein *Tb*RAP1-ML (0, 100, 250, and 500 ng, respectively) were mixed with 0.5 ng radiolabeled dsDNA oligos (TTAGGG)*12. After 30 min incubation, the mixture was loaded into a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer. Afterwards, gel shift was detected under a phosphor imager.

With the gradually increasing amount of *Tb*RAP1-ML addition, gel shifts became more obvious (Figure 3.1 A), indicating more *Tb*RAP1-ML-dsDNA complex formation.

A similar interaction pattern also can be found when using 100 nt random ssDNA sequence (Figure 3.1 B) or 100 bp random dsDNA sequence (Figure 3.1 C) as substrates. In summary, EMSA data suggest that *Tb*RAP1-ML has sequence non-specific interaction with both dsDNA and ssDNA.

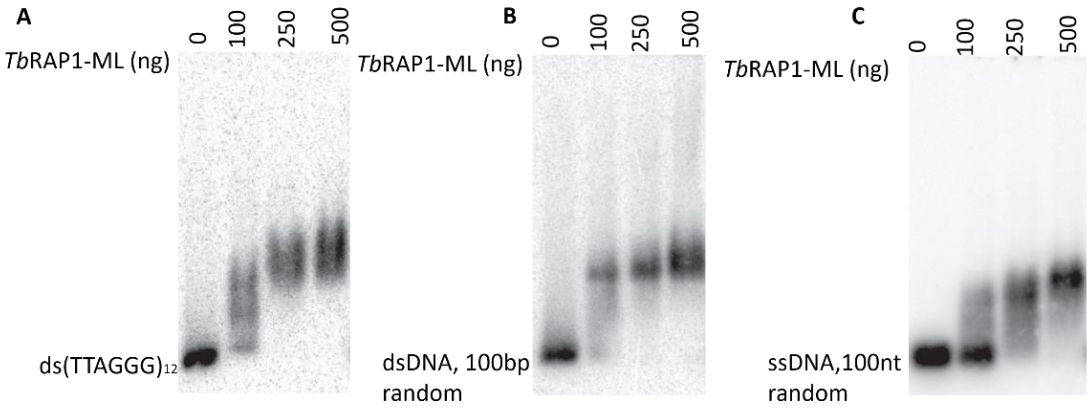


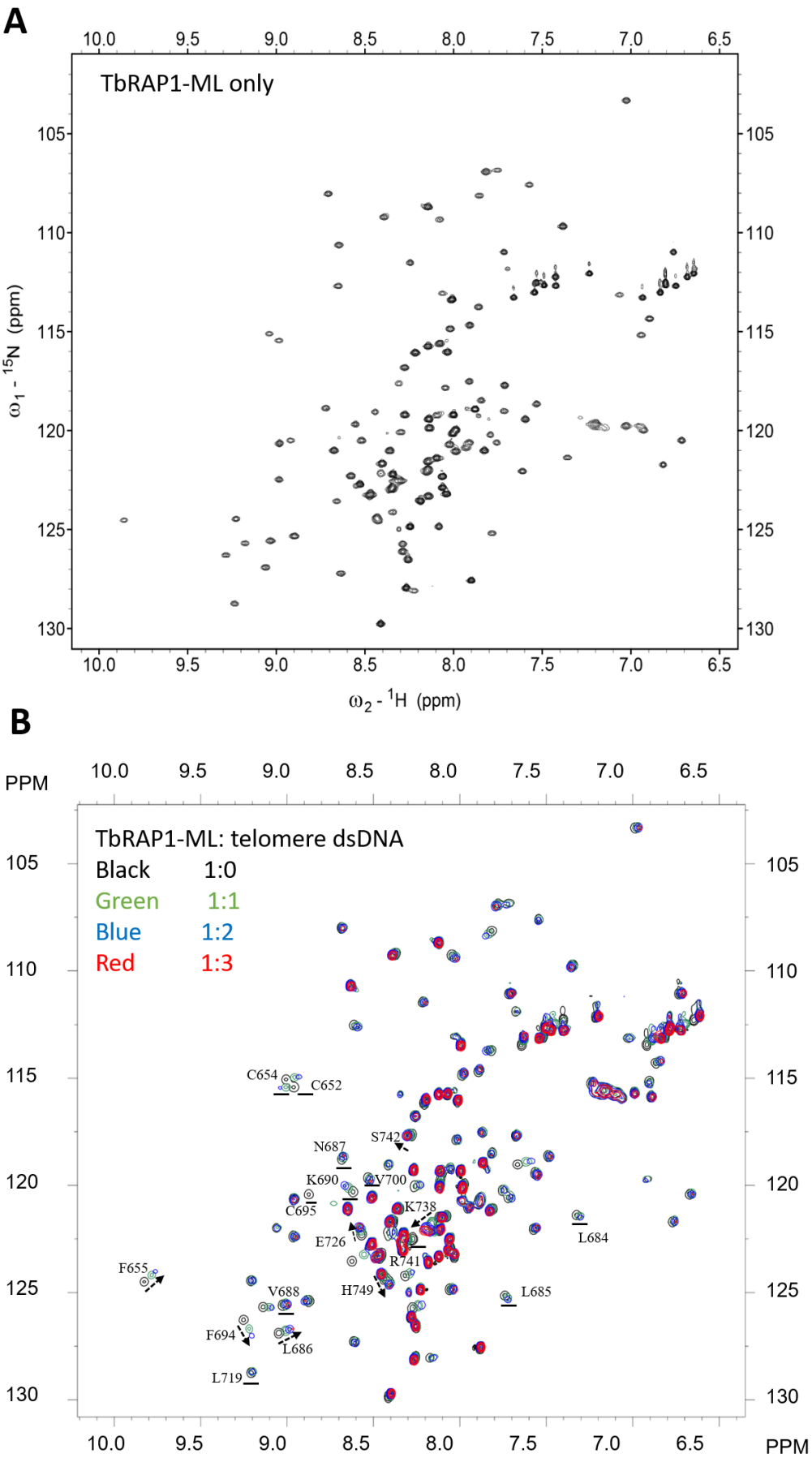
Figure 3.1 EMSA data show that *Tb*RAP1-ML has non-specific dsDNA and ssDNA binding activities. (A) EMSA experiments were performed by using a radiolabeled telomeric dsDNA containing $(TTAGGG)_{12}$ as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML. (B) EMSA experiments were performed by using a radiolabeled random dsDNA (100 bp) as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML. (C) EMSA experiments were performed by using a radiolabeled random ssDNA oligo (100 nt) as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML. The amount (ng) of protein used is indicated on top of each lane, respectively.

3.1.2 HSQC titration data confirmed the interaction between *Tb*RAP1-ML and sequence non-specific dsDNA/ssDNA

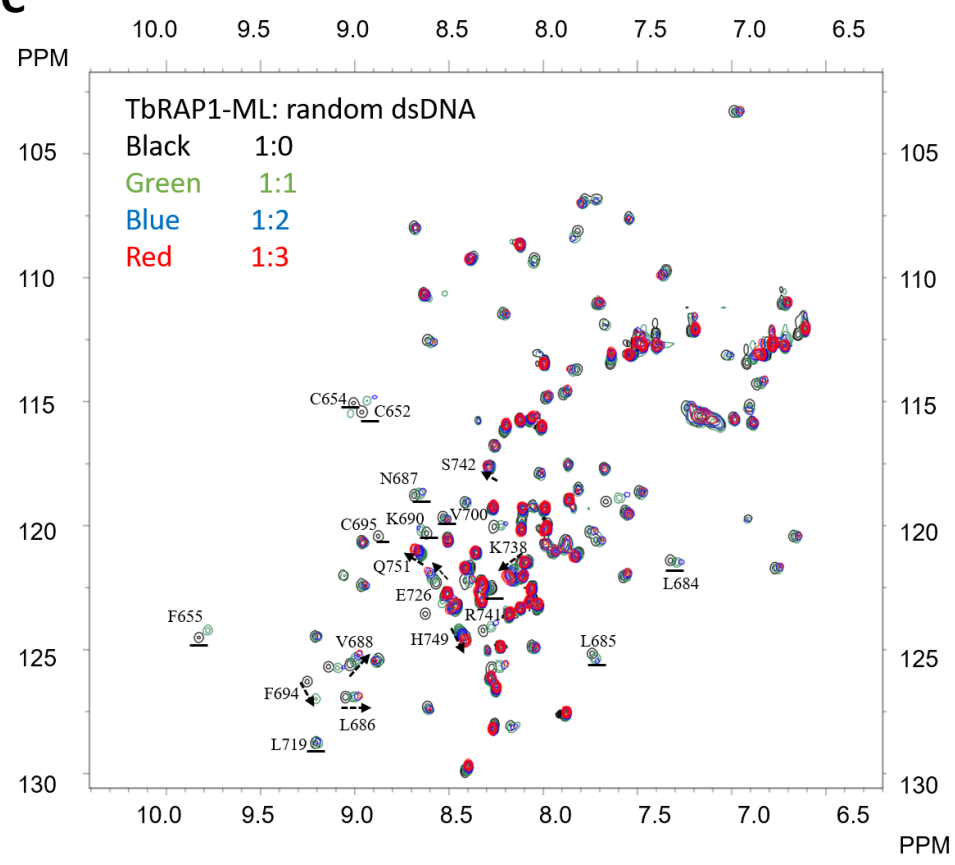
To further validate the interaction between *Tb*RAP1-ML and DNA, HSQC titration assay was carried out. First of all, ^{15}N-labeled recombinant protein *Tb*RAP1-ML was cultured in M9 minimal media containing $^{15}NH_4Cl$. To collect HSQC spectrum, purified ^{15}N-labeled *Tb*RAP1-ML was dissolved in buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10% D_2O, and the final concentration of protein is 0.1 mM.

Before NMR titration, the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled *Tb*RAP1-ML alone was collected. As shown in Figure 3.2 A, the overall homogeneity of *Tb*RAP1-ML was pretty good, and most residues were visibly and able to distinguish from each other. It indicates that *Tb*RAP1-ML is well folded under this condition. Consequently, HSQC titration assay can be carried out to investigate *Tb*RAP1-ML's DNA binding activities.

Based on EMSA results, we have noticed that *Tb*RAP1-ML binds to both dsDNA and ssDNA. Therefore, we designed three types of DNA oligos as substrates for *Tb*RAP1-ML HSQC titration, including telomeric dsDNA (18 bp), random dsDNA (18 bp), and random ssDNA (18 nt).


A
TbRAP1-ML only
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)
B
PPM
TbRAP1-ML: telomere dsDNA
Black 1:0
Green 1:1
Blue 1:2
Red 1:3
C654
C652
S742
N687
K690
V700
C695
K738
E726
R741
L684
F655
V688
H749
L685
F694
L686
L719

C


TbRAP1-ML: random dsDNA
Black 1:0
Green 1:1
Blue 1:2
Red 1:3
PPM
C654
C652
S742
N687
C695
K690
V700
K738
Q751
E726
R741
L684
F655
V688
H749
L685
F694
L686
L719

D

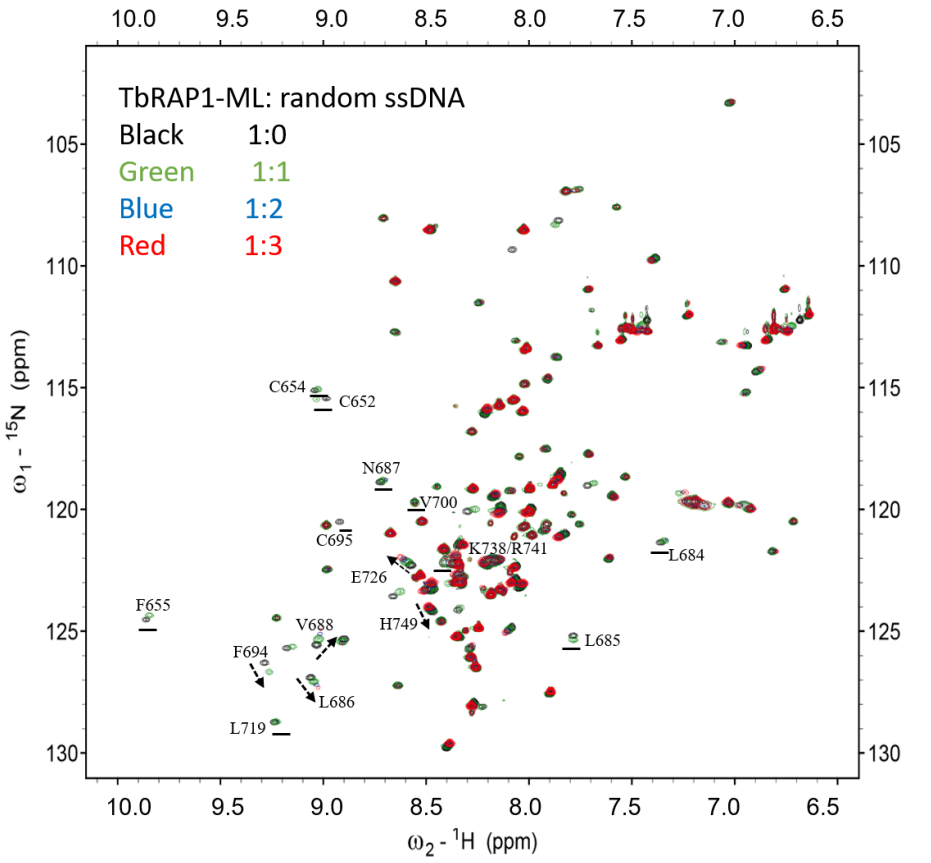

TbRAP1-ML: random ssDNA
Black 1:0
Green 1:1
Blue 1:2
Red 1:3
ω1 - 15N (ppm)
ω2 - 1H (ppm)
C654
C652
N687
V700
C695
K738/R741
L684
E726
F655
V688
H749
L685
F694
L686
L719

Figure 3.2 HSQC titration results between *Tb*RAP1-ML and varied DNA oligos suggest that *Tb*RAP1-ML has non-specific DNA binding activity. (A) ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled *Tb*RAP1-ML alone was collected. The protein concentration was 0.1 mM. Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML in the absence (black) and presence of (B) telomeric dsDNA, (C) random dsDNA, and (D) random ssDNA in 1× (green), 2× (blue), and 3× (red) molar excess. Residues with chemical shifts are labeled. Arrows indicate substantial chemical shifts after adding the DNA substrate.

In HSQC titration assay, telomeric dsDNA (TTAGGG*3, 18 bp) was gradually titrated into ^{15}N-labeled *Tb*RAP1-ML with protein-DNA molar ratio 1:1, 1:2, and 1:3, respectively. The HSQC spectra after each titration were collected and analyzed. We found that multiple residues, such as C652, C654, L684, L685, N687, V700, E726, K738, R741 and H749, showed noticeable concentration-dependent chemical shifts when more telomeric dsDNA was titrated into ^{15}N-labeled *Tb*RAP1-ML (Figure 3.2 B). These data suggest that telomeric dsDNA indeed interacts with *Tb*RAP1-ML.

Notably, the NMR signal of many residues disappeared when protein-DNA substrates molar ratio reached 1:3, including residues E652, K654, C654, F655, I693, C695, and F694. The complex formed by protein-DNA with large molecular weight may block the appearance of other signals. Furthermore, the low amount of complex formation also can cause the weak signal detection, indicating the moderate binding

affinity between *Tb*RAP1 and DNA. It is also possible excess DNA affects proper folding of *Tb*RAP1, thus negatively affecting the HSQC spectrum.

A similar interaction pattern was found when using random dsDNA (Figure 3.2 C) or random ssDNA (Figure 3.2 D) as substrates titrated into ^{15}N-labeled *Tb*RAP1-ML. HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML in the absence (black) and presence of random dsDNA/ssDNA, in 1×(green), 2× (blue), and 3× (red) molar excess were monitored. With analyzing by Sparky software, we found that addition of an increasing amount of random dsDNA/random ssDNA caused more peaks disappearance and larger chemical shifts. More importantly, the disappeared or chemical shifted residues were consistent with the telomeric dsDNA titration result. Overall, HSQC titration results indicate that *Tb*RAP1-ML shows sequence non-specific DNA binding activity.

Combined biochemistry characterization results from EMSA and NMR titration assays together, we conclude that *Tb*RAP1 Myb-like region (aa 639-761) shows sequence-non-specific interaction with both ssDNA and dsDNA.

3.2 *Tb*RAP1-ML has a higher affinity for longer DNA substrates

To further investigate characteristic of *Tb*RAP1's sequence-non-specific DNA binding activity, different lengths of DNA oligos were used as substrates in EMSA assay. *Tb*RAP1-ML showed a higher binding affinity with longer dsDNA oligos when using different length of random dsDNA (60, 80, and 150 bp) as substrates, respectively. With the shortest dsDNA (60 bp) as probe, 100 ng of *Tb*RAP1-ML caused little gel shifts (Figure 3.3 A), indicating small amounts of protein-DNA complex formations. While

100 ng of *Tb*RAP1-ML was sufficient to cause all the longest dsDNA (150 bp) shifts (Figure 3.3 A). These data reveal that the longer dsDNA has a higher binding affinity with *Tb*RAP1-ML.

Additionally, competitive EMSA experiments were carried out. Different lengths of non-radiolabeled dsDNA oligos (60, 100, 200, and 500 bp, respectively) were used to compete with radiolabeled 80 bp random dsDNA for *Tb*RAP1-ML binding (Figure 3.3 B). Specifically, 250 ng *Tb*RAP1-ML were mixed with a radiolabeled 80 bp random dsDNA as the substrate and all of *Tb*RAP1-ML occurred gel shifts. Afterward, non-radiolabeled 60, 100, 200, and 500 bp random dsDNA oligos were used as competitors, respectively. An increasing amount of competitor was added in 1×, 3×, 5×, 10× and 20× molar excess than radiolabeled 80 bp random dsDNA.

EMSA competition result showed that shifted band was partly inhibited after 10-time amount of 60 bp random dsDNA (the shortest competitor) addition. While the longer competitors required less additions to prevent the EMSA shift. For example, 10× molar excess of competitor with 100 bp random dsDNA oligos were required for inhibiting EMSA shifts completely. While only 3× molar excess of 500 bp random dsDNA oligos (the longest competitor) addition was sufficient to block all EMSA shifts. Based on these findings, we summarized that longer dsDNA oligos competed better than shorter ones (500 > 200 > 100 > 60 bp).

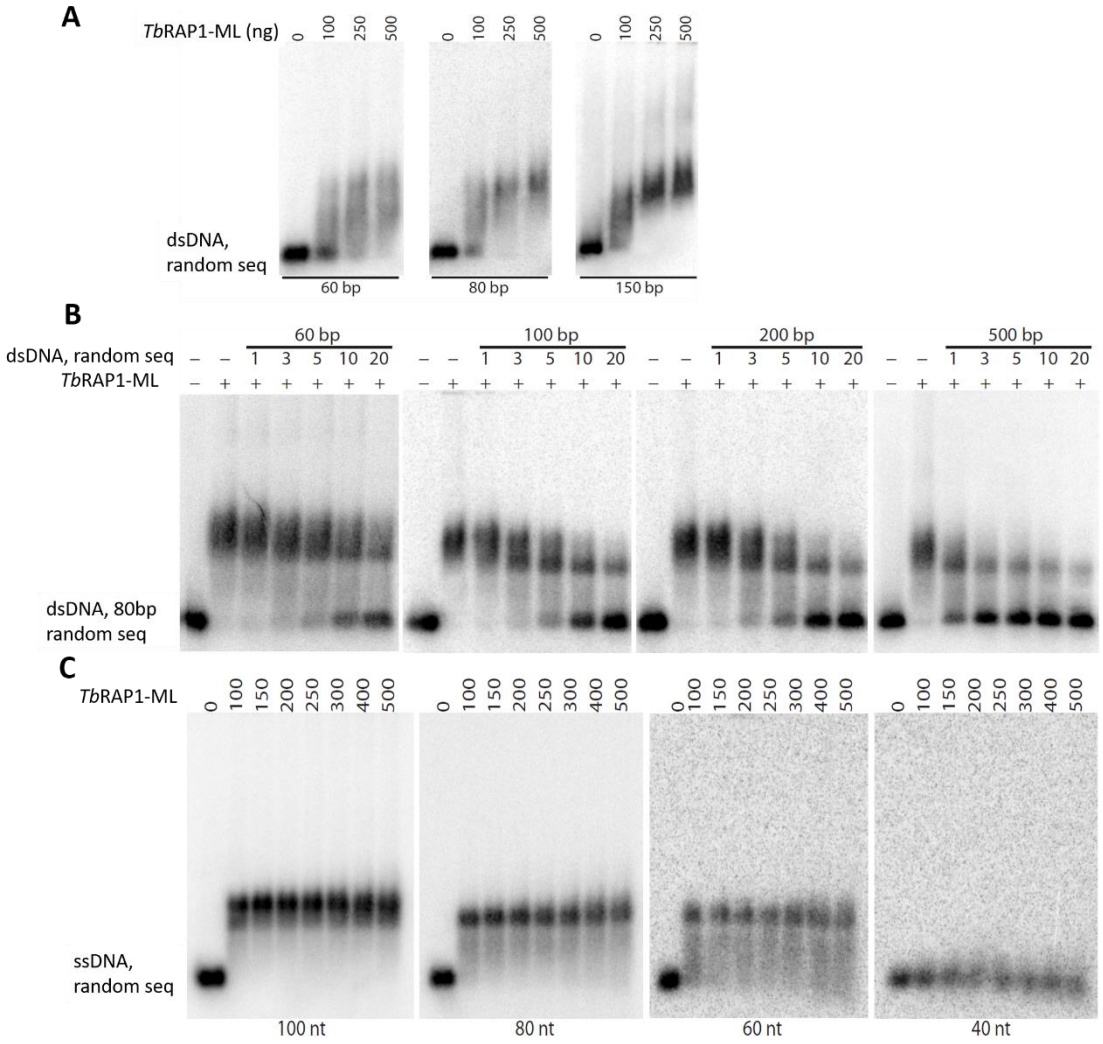


Figure 3.3 EMSA competitive experiments confirmed that *Tb*RAP1 has higher affinity for longer DNA substrates. (A) EMSA experiments were performed by using radiolabeled random dsDNA (of 60, 80, or 150 bp) oligos as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML. (B) EMSA using 250 ng of TrxA-His6-*Tb*RAP1-ML and a radiolabeled 80 bp random dsDNA as the substrate. Non-radiolabeled dsDNA with a random sequence of 60, 100, 200, or 500 bp were used as competitors. The amounts of competitors are indicated as molar folds of the probe. (C) EMSA experiments were performed by using radiolabeled 100, 80, 60, or 40 nt random

ssDNA oligo as the substrate to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML.

The same sequence length preference was also found in *Tb*RAP1's ssDNA binding activity. In the beginning, radiolabeled different length of random ssDNA (100, 80, 60, and 40 nt, respectively) were used as probes to detect EMSA shifts with increased addition of purified Trx-His6-tagged *Tb*RAP1-ML. EMSA result showed that 100 ng *Tb*RAP1-ML was sufficient to cause radiolabeled 100 nt random ssDNA because of RNA-protein complex formation (Figure 3.3 C). While 500 ng *Tb*RAP1-ML still failed to cause any EMSA shifts when using 40 nt random ssDNA as a probe. These data suggest that *Tb*RAP1-ML binds to shorter DNA oligos with weaker affinity (40 < 60 < 80 < 100 nt).

In summary, EMSA competitive experiments demonstrate that *Tb*RAP1-ML's DNA binding activities are not only sequence nonspecific, but also substrate length-dependent.

3.3 *Tb*RAP1 R/K patch is necessary and sufficient for DNA binding activity through electrostatic interaction

3.3.1 Construct design to investigate the key region that responsible for *Tb*RAP1's DNA binding activity

To map critical residues involved in *Tb*RAP1's DNA binding activities, we analyzed the residues with chemical shifts in HSQC titration result (Figure 3.2 B-D).

Interestingly, we found that the positively charged residues K738 and R741 within the $_{737}$RKRRR$_{741}$ patch (the R/K patch) underwent notable chemical shifts. Therefore, we hypothesized that the R/K patch within *Tb*RAP1-ML might play an important role in DNA binding activity.

To confirm the importance of the R/K patch in *Tb*RAP1's DNA binding activities, we designed two constructs within *Tb*RAP1-ML. Constructs *Tb*RAP1-ML-5A mutant with all five R and K residues replaced by A residues and *Tb*RAP1-734-761 with R/K patch only were generated (Figure 3.4).

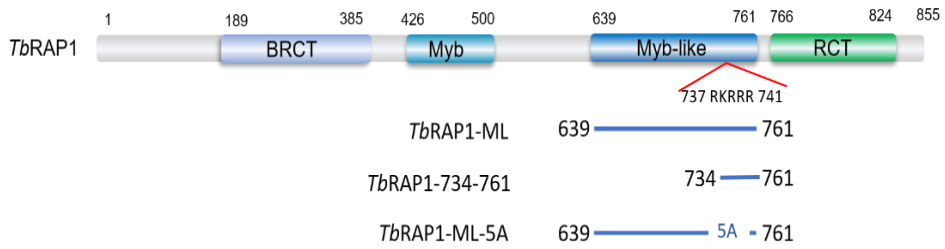


Figure 3.4 Constructs design within *Tb*RAP1-ML to investigate its DNA binding activity.

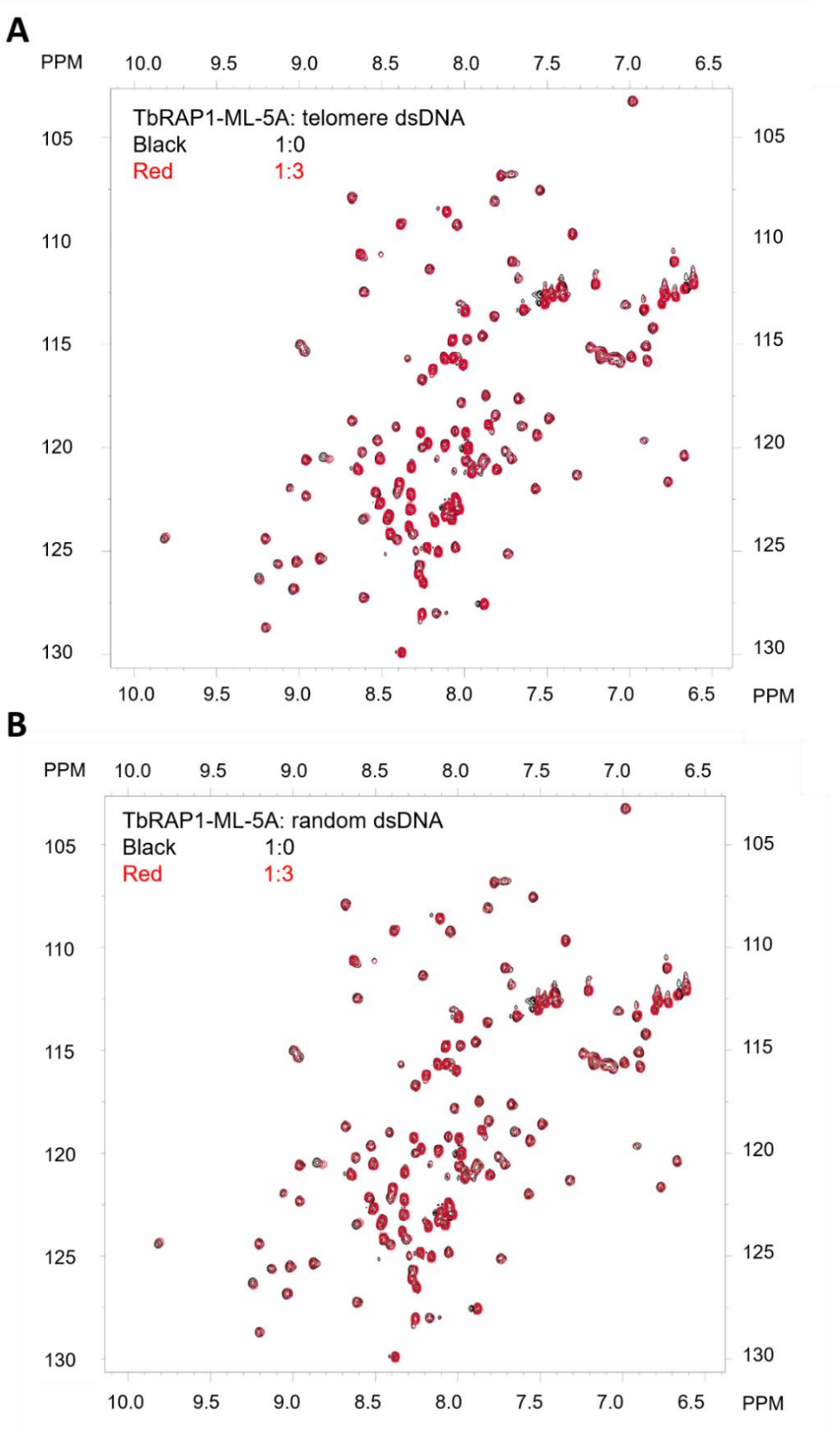
3.3.2 Deletion or mutation of R/K patch blocks *Tb*RAP1-ML's DNA binding activity *in vitro*

HSQC titration assay was performed to determine whether *Tb*RAP1-ML-5A retains DNA binding activity. ^{15}N-labeled recombinant protein *Tb*RAP1-ML-5A was obtained by culturing *E. coil* with M9 minimal media containing $^{15}NH_4Cl$. Before HSQC spectrum collection, purified ^{15}N-labeled *Tb*RAP1-ML-5A was dissolved in buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT)

containing 10% D_2O, with 0.1 mM final concentration.

And then, the HSQC NMR spectrum of ^{15}N-labeled *Tb*RAP1-ML-5A alone (black) was collected (Figure 3.5 A) to verify the folding of this new construct. The overall homogeneity of ^{15}N-labeled *Tb*RAP1-ML-5A was pretty good, indicating *Tb*RAP1-ML-5A was well folded under this condition.

Afterward, HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML-5A in the presence of telomeric dsDNA (TTAGGG*3, 18 bp) in 3× (red) molar excess were monitored (Figure 3.5 A). With analyzing by Sparky software, we found that no chemical shifts were observed. All peaks within these two spectra were well overlapped, suggesting that *Tb*RAP1-ML-5A fails to interact with telomeric dsDNA. A similar result was also found by using random dsDNA (Figure 3.5 B) or random ssDNA (Figure 3.5 C) as substrates titrated into *Tb*RAP1-ML-5A. Therefore, *Tb*RAP1-ML-5A HSQC titration result indicates that the $_{737}$RKRRR$_{741}$ patch is responsible for *Tb*RAP1's DNA binding activities.


A
PPM
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
TbRAP1-ML-5A: telomere dsDNA
Black 1:0
Red 1:3
105
110
115
120
125
130
PPM
B
PPM
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
TbRAP1-ML-5A: random dsDNA
Black 1:0
Red 1:3
105
110
115
120
125
130
PPM

C

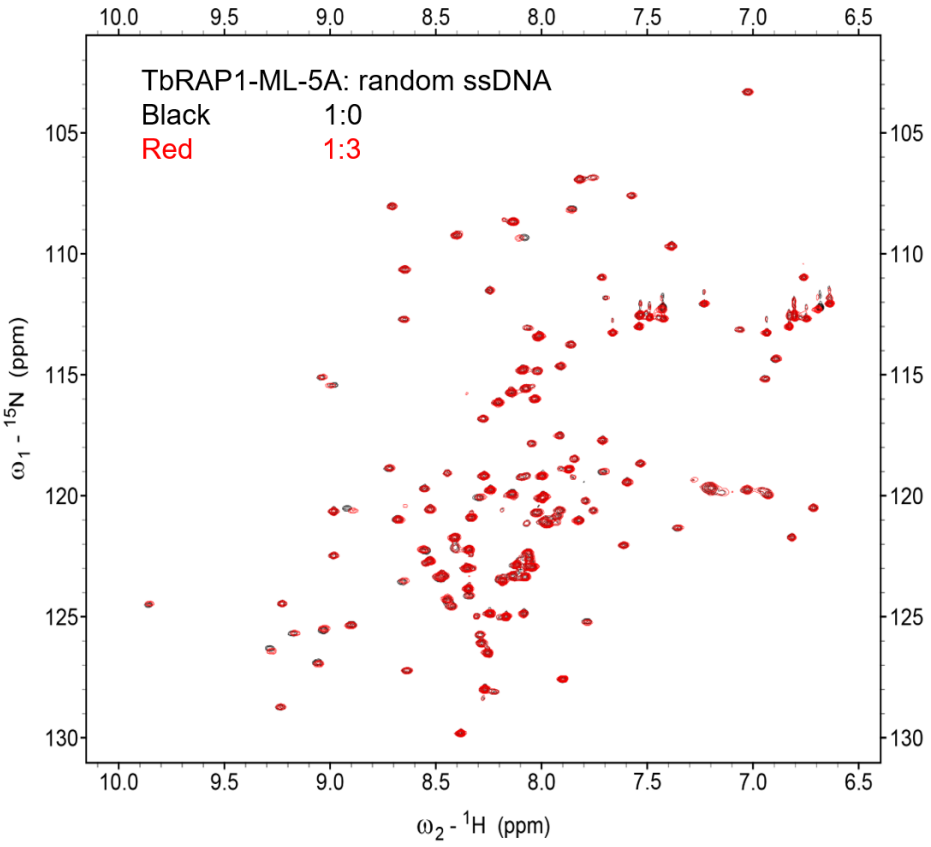


Figure 3.5 HSQC titration results between *Tb*RAP1-ML-5A and varied DNA oligos suggest that 5A mutant lost DNA binding activity. Two overlapped *Tb*RAP1-ML-5A ^{15}N HSQC NMR spectra before (black) and after (red) titrating with (A) telomeric dsDNA, (B) random dsDNA, and (C) random ssDNA, respectively, in protein-DNA molar ratio 1:3.

To further confirm the importance of $_{737}$RKRRR$_{741}$ patch in *Tb*RAP1-ML's DNA binding activity, construct *Tb*RAP1-ML-5A that mutated R/K patch and *Tb*RAP1-734-761 containing R/K patch only were used for EMSA assay.

Different amounts of protein *Tb*RAP1-ML-5A (0, 100, 250, and 500 ng, respectively) were mixed with radiolabeled telomeric dsDNA oligos (TTAGGG*12), respectively (Figure 3.6 A). After 30 min incubation, the mixture was loaded into a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer. Finally, gel shift was detected by a phosphor imager.

EMSA results showed that no gel shifts were observed in *Tb*RAP1-ML-5A because of no protein-DNA complex formation. It indicated that *Tb*RAP1-ML-5A did not bind to telomeric dsDNA anymore. *Tb*RAP1-ML-5A also failed to cause gel shift when using random dsDNA (Figure 3.6 B) and random ssDNA (Figure 3.6 C) as a probe. These data suggest that *Tb*RAP1-ML-5A loses its DNA binding activity.

Meanwhile, radiolabeled telomeric dsDNA (TTAGGG*12) was also used as probes to detect EMSA shifts with addition of purified Trx-His6-tagged *Tb*RAP1-734-761. Mixed different amounts of protein *Tb*RAP1-734-761 (0, 100, 250, and 500 ng, respectively) with 0.5 ng radiolabeled telomeric dsDNA (TTAGGG*12). After 30 min incubation, loading mixture into a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer and detecting gel shift under a phosphor imager. EMSA result showed that more obvious gel shifts were measured with the gradually increasing amount of *Tb*RAP1-734-761 addition (Figure 3.6 A). Similar shifts were observed when using random dsDNA (Figure 3.6 B) or random ssDNA (Figure 3.6 C) as substrate. These data suggest that *Tb*RAP1-734-761 still retains its DNA binding activity.

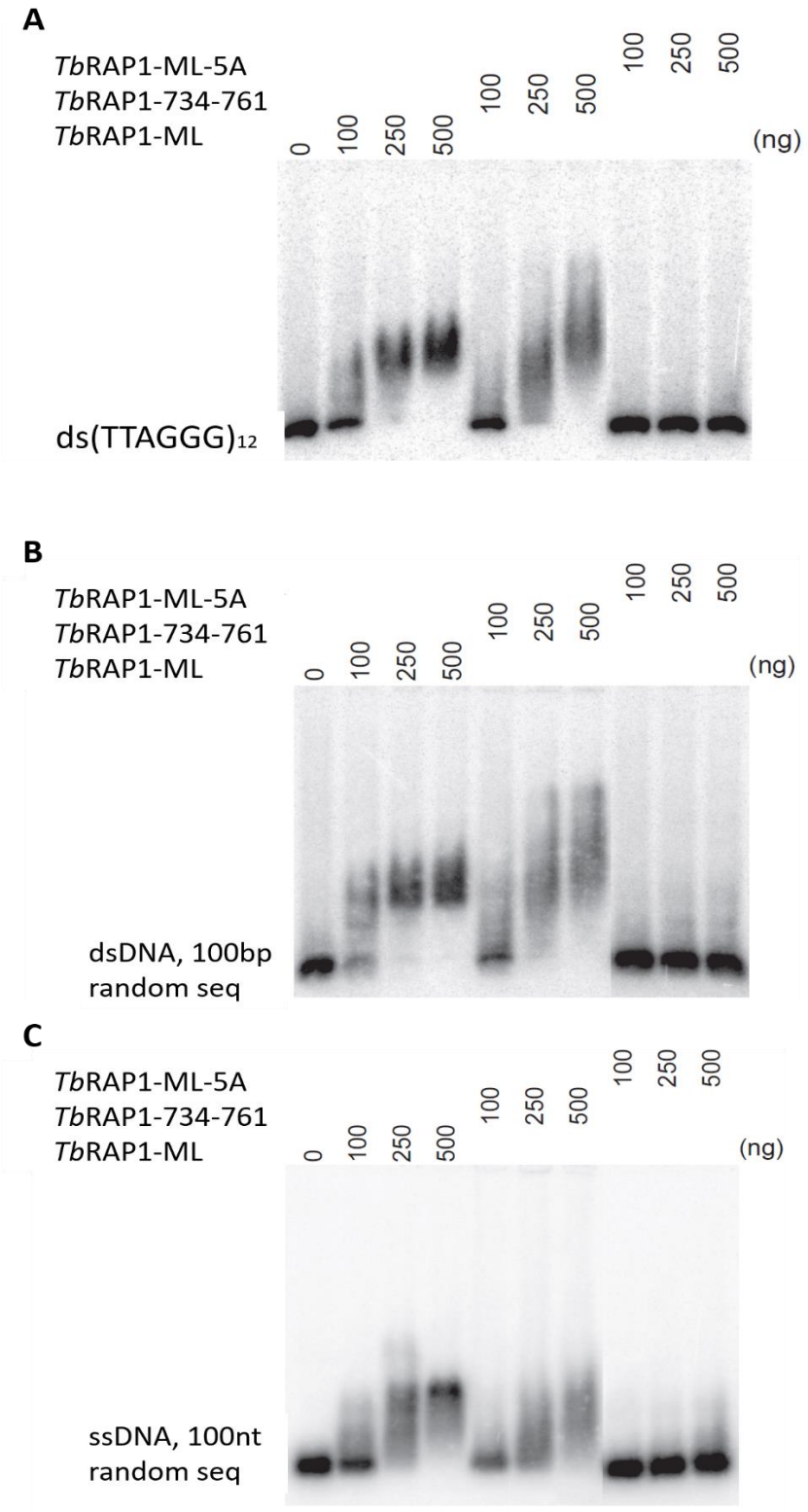


Figure 3.6 EMSA data show that the R/K patch within *Tb*RAP1-ML is important for DNA binding activities. (A) EMSA experiments were performed by using a

radiolabeled dsDNA containing $(TTAGGG)_{12}$ as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML, *Tb*RAP1-734-761, and *Tb*RAP1-ML-5A, respectively. (B) EMSA experiments were performed by using a radiolabeled random dsDNA as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML, *Tb*RAP1-734-761, and *Tb*RAP1-ML-5A, respectively. (C) EMSA experiments were performed by using a radiolabeled random ssDNA as a probe to detect shifts under treatment of Trx-His6-tagged *Tb*RAP1-ML, *Tb*RAP1-734-761, and *Tb*RAP1-ML-5A, respectively. The amount (ng) of protein used is indicated on top of each lane.

Combined these NMR titrations and EMSA data, we found that *Tb*RAP1-ML-5A lost its DNA binding activity while *Tb*RAP1-734-761 remained it. Therefore, we have concluded that *Tb*RAP1 has sequence-nonspecific dsDNA and ssDNA binding activities in the amino acid 734 to 761, which we named as the DNA binding (DB) region.

3.4 Telomere localization of *Tb*RAP1 requires the R/K patch *in vivo*

3.4.1 Establishing *Tb*RAP1 mutant strains used for DNA binding activity investigation

To further investigate the function of *Tb*RAP1's DNA binding activity *in vivo*, a series of cell-based experiments were carried out by our collaborator, Dr. Li Bibo from the Cleveland State University of USA.

Firstly, our collaborator has established a Cre-loxP-mediated conditional deletion system for *Tb*RAP1 function characterization (Figure 3.7). In *Tb*RAP1$^{F/+}$ strain, one *Tb*RAP1 allele was flanked by two loxP repeats (described as "F") which can be conditionally deleted after Cre treatment. While another WT *Tb*RAP1 allele containing (described as "+") can be replaced by various *Tb*RAP1 mutants. Specifically, to investigate which region within *Tb*RAP1 is necessary for *Tb*RAP1 telomere localization, we replaced the WT *Tb*RAP1 allele of *Tb*RAP1$^{F/+}$ with F2H-tagged mutants to generate *Tb*RAP1$^{F/mut}$ strains as described in the Table 3.1, including *Tb*RAP1$^{F/-}$, *Tb*RAP1$^{F/\Delta MybLike(\Delta ML)}$, *Tb*RAP1$^{F/\Delta DB}$, and *Tb*RAP1$^{F/5A}$.

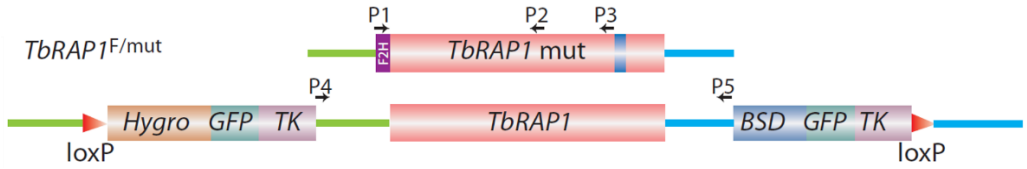


Figure 3.7 Establishing *Tb*RAP1 mutant strains. *Tb*RAP1$^{F/mut}$ strains contain two *Tb*RAP1 alleles including one floxed allele and one mutant allele, respectively. PCR primers used for validating the deletion of the floxed *Tb*RAP1 allele after Cre induction are marked.

Table 3.1 List of *Tb*RAP1 mutant strains used for its DNA binding activity investigation

Strains	Description
*Tb*RAP1$^{F/+}$	One floxed allele and one WT *Tb*RAP1
*Tb*RAP1$^{F/-}$	One floxed allele and one deleted allele
*Tb*RAP1$^{F/\Delta ML}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged mutant lacking the ML domain
*Tb*RAP1$^{F/\Delta DB}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged mutant lacking the DB domain
*Tb*RAP1$^{F/5A}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged $_{737}$RKRRR$_{741}$ to 5A mutant
*Tb*RAP1$^{F2H+/-}$	One F2H-tagged WT *TbRAP1* and one deleted allele

3.4.2 ChIP results indicate that the R/K patch is required for *Tb*RAP1 chromatin association

To examine whether F2H-tagged *Tb*RAP1 mutants associate with telomere chromatin, ChIP assay was performed in *Tb*RAP1$^{F/mut}$ cells (ΔML, ΔDB, and 5A) which lost DNA binding activity. Since *Tb*RAP1 occurs self-interaction through its BRCT domain, we removed the WT *Tb*RAP1 allele by Cre treatment to specifically examine *Tb*RAP1 mutants' behavior without the influence from the WT protein.

Firstly, western blot assay was performed to confirm the expression level of *Tb*RAP1 mutants. The whole-cell lysates were prepared with the *Tb*RAP1$^{F/mut}$ cells after Cre treatment for 30 hours. Western blot results suggested that *Tb*RAP1-ΔML were expressed at a subtly lower level than the *Tb*RAP1 WT (Figure 3.8 A). While *Tb*RAP1-ΔDB and 5A were expressed at the same level as *Tb*RAP1 WT.

Afterward, ChIP experiments were conducted in *Tb*RAP1$^{F/\Delta ML}$ (Figure 3.8 B), *Tb*RAP1$^{F/\Delta DB}$ (Figure 3.8 C), and *Tb*RAP1$^{F/5A}$ (Figure 3.8 D) after Cre treatment with 30 hours. *Tb*RAP1$^{F2H+/-}$ which contains one F2H-tagged *Tb*RAP1 WT was used as a positive control in ChIP assay. The HA antibody 12CA5 and a *Tb*TRF rabbit antibody were used in these *Tb*RAP1$^{F/mut}$ samples. ChIP results showed that both *Tb*RAP1 ΔML, ΔDB, and 5A mutants failed to associate with *Tb*RAP1 telomere chromatin, as no detectable telomere chromatin coprecipitation. *Tb*TRF, a telomeric dsDNA binding protein, was still associated with telomere chromatin in these mutants. Guided by these ChIP data, we have concluded that the R/K patch is required for *Tb*RAP1 chromatin association.

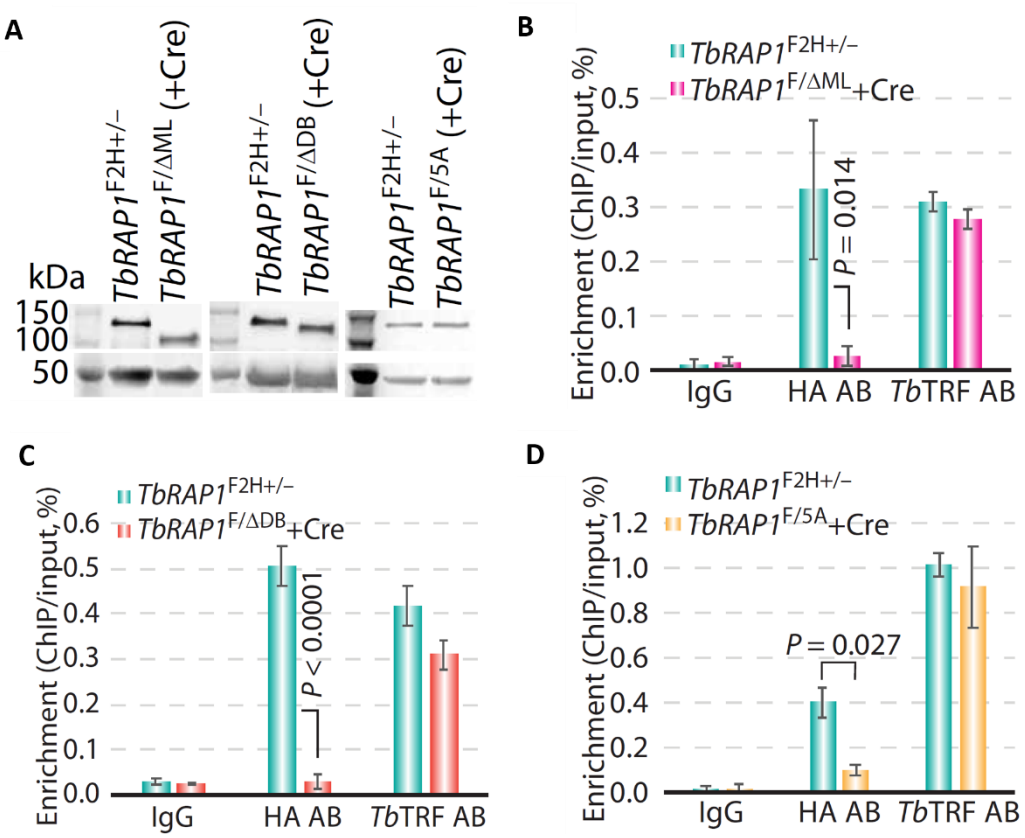


Figure 3.8 ChIP conducted from *Tb*RAP1$^{F/mut}$ cells indicated that the R/K patch is required for *Tb*RAP1 chromatin association. (A) Expression levels of F2H-tagged *Tb*RAP1 proteins were detected by western blot. The HA antibody 12CA5 was used to detect the expression level of *Tb*RAP1 ΔML, ΔDB, and 5A. Tubulin antibody TAT-1 was used for loading control. (B) ChIP experiments using the HA antibody 12CA5 and a *Tb*TRF rabbit antibody were done in *Tb*RAP1$^{F2H+/–}$ cells and Cre-induced (for 30 hours) *Tb*RAP1$^{F/ΔML}$ cells (B), *Tb*RAP1$^{F/ΔDB}$ cells (C), and *Tb*RAP1$^{F/5A}$ cells (D), respectively.

3.4.3 IF analyses suggest that the R/K patch is required for targeting *Tb*RAP1 to telomere chromatin

IF analyses were carried out to assess the subnuclear localization of *Tb*RAP1 in *Tb*RAP1$^{F2H+/-}$ and *Tb*RAP1$^{F/5A}$ cells after Cre treatment for 30 hours. In *Tb*RAP1$^{F2H+/-}$ cells, F2H-*Tb*RAP1 (red) partially colocalized with *Tb*TRF (green) in nucleus (Figure 3.9 A). 4′,6-diamidino-2-phenylindole (DAPI) was used to stain nucleus. However, F2H-*Tb*RAP1-5A (green) was no longer colocalized with *Tb*TRF (red), even though it was imported into the nucleus via the SV40 NLS (Figure 3.9 B), indicating F2H-*Tb*RAP1-5A did not localize at telomere anymore. Therefore, these IF data further confirmed that the R/K patch is required for targeting *Tb*RAP1 to telomere chromatin.

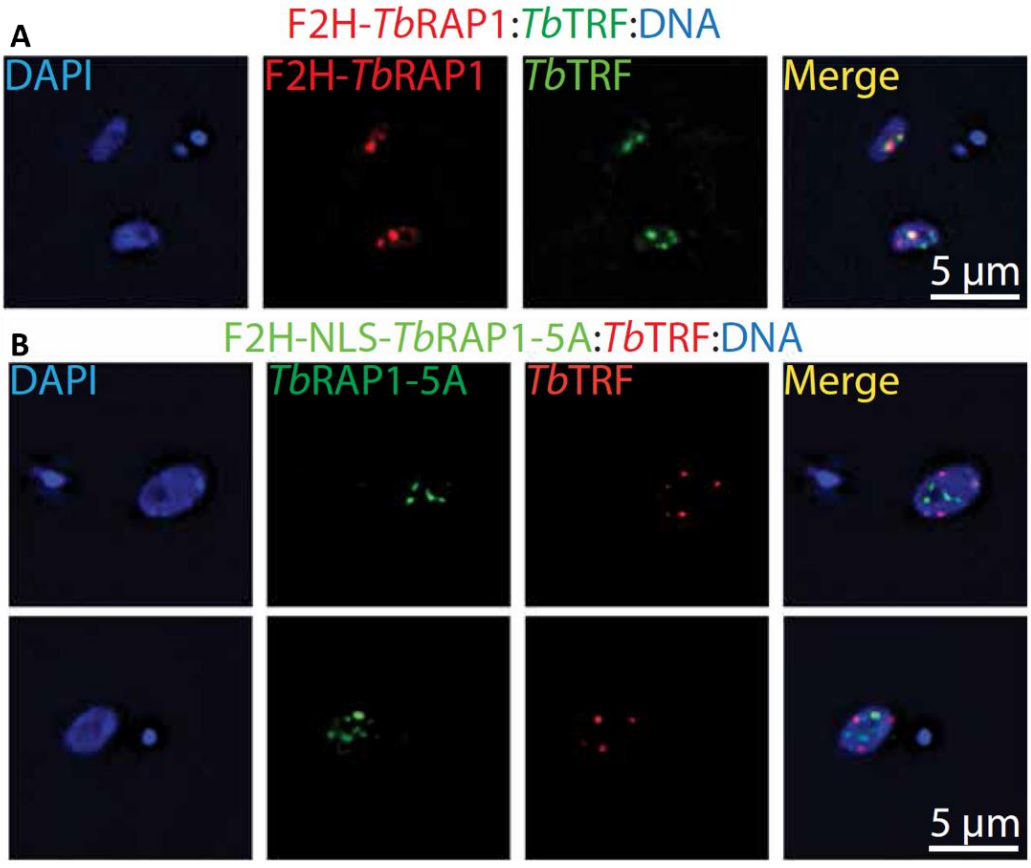


Figure 3.9 IF analyses suggested that the R/K patch is required for targeting *Tb*RAP1 to telomere chromatin. (A) IF analyses profile of *Tb*RAP1$^{F2H+/-}$ cells. (B) IF analyses profile of *Tb*RAP1$^{F/5A}$ cells. The 12CA5 antibody was applied to detect *Tb*RAP1 and a *Tb*TRF chicken antibody was used to detect *Tb*TRF. The size bars of all images are labeled in each panel.

3.5 Telomere localization of *Tb*RAP1 is essential for cell growth

To further investigate the functions of *Tb*RAP1's DNA binding activities, we detected cell growth in *Tb*RAP1$^{F/\Delta DB}$ cell which deleted DB region and *Tb*RAP1$^{F/5A}$ cell carrying R/K patch mutation after Cre induction. In the beginning, we confirmed F2H-NLS-tagged *Tb*RAP1 mutants protein expression level within these two *Tb*RAP1$^{F/mut}$ strains by western analysis. Afterward, *Tb*RAP1$^{F/mut}$ cell growth curves were detected under the condition with or without Cre induction, respectively. Three independent experiments were performed to obtain average values.

As shown in the western blot results (Figure 3.10 A), *Tb*RAP1$^{F/5A}$ cells before and after induction of Cre for 30 hours were harvested and analyzed. The 5A mutant was detected by the HA monoclonal antibody, while the endogenous *Tb*RAP1 was detected by a *Tb*RAP1 rabbit antibody. To detect both *Tb*RAP1 mutants and the endogenous *Tb*RAP1 within one gel, protein samples were run on a 7.5% Tris polyacrylamide gel and incubated by the *Tb*RAP1 rabbit antibody. Tubulin expression level was detected as a loading control. Western blot result showed that, after treatment with Cre for 30

hours, floxed-*Tb*RAP1 was disappeared while F2H-NLS-tagged *Tb*RAP1-5A was normally expressed in *Tb*RAP1$^{F/5A}$ cells.

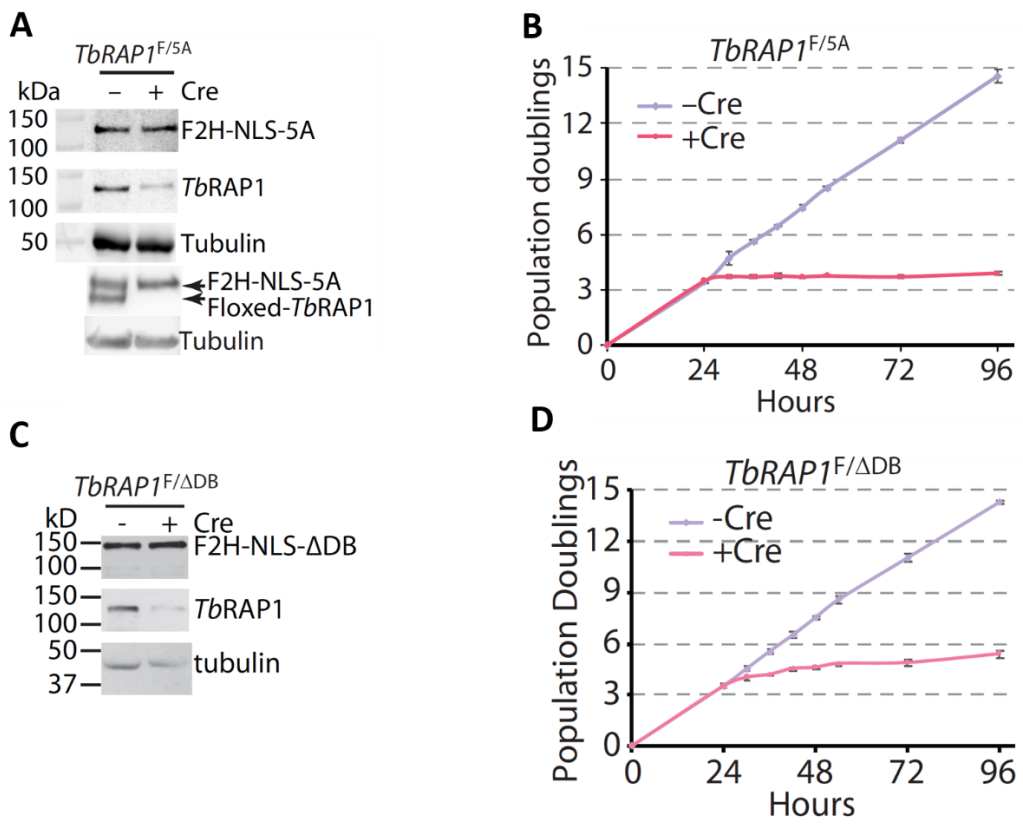


Figure 3.10 Telomere localization of *Tb*RAP1 is required for normal cell growth. (A) Western analyses profile in *Tb*RAP1 $^{F/5A}$ cells. (B) Cell growth curves with or without Cre treatment in *Tb*RAP1 $^{F/5A}$ cells. (C) Western analyses profile in *Tb*RAP1$^{F/\Delta DB}$ cells. (D) Cell growth curves with or without Cre treatment in *Tb*RAP1$^{F/\Delta DB}$ cells.

And then the cell growth curve over 96 hours was elaborated for under the condition with and without Cre treatment in *Tb*RAP1 $^{F/5A}$ cells (Figure 3.10 B). Three independent experiments were performed to obtain average values. The result showed

that cell population doublings did not increase after Cre treatment for 30 hours, indicating cell growth arrest in *Tb*RAP1$^{F/5A}$ strain.

We also performed western analyses and cell growth curve in *Tb*RAP1$^{F/\Delta DB}$ cells. As shown in the western blot result (Figure 3.10 C), the expression level of F2H-NLS-tagged *Tb*RAP1-ΔDB was normal and no *Tb*RAP1 WT was detected after Cre treatment for 30 hours. A similar cell growth curve was also obtained in *Tb*RAP1$^{F/\Delta DB}$ strain (Figure 3.10 D). Comparing to the control group without Cre treatment, cell population doublings showed no change after Cre treatment for 30 hours, indicating cell growth was arrested in *Tb*RAP1$^{F/\Delta DB}$ strain. In summary, these data suggest that the R/K patch-mediated DNA binding activity of *Tb*RAP1 is essential for cell viability.

3.6 *Tb*RAP1's DNA binding activity is essential for *VSG* silencing

Since *Tb*RAP1 is a critical suppressor for *VSG* switching, we examined whether R/K patch mediated *Tb*RAP1 DNA binding activity influences *VSG* silencing. We detected mRNA levels of several ES-linked *VSG*s in *Tb*RAP1$^{F/mut}$ strains (*Tb*RAP1$^{F/\Delta M}$ and *Tb*RAP1$^{F/\Delta DB}$) which lost *Tb*RAP1's DNA binding activity. The qRT-PCR technique was applied to measure mRNA level.

After Cre induction for 30 hours, multiple ES-linked silent *VSG*s, such as *VSG*8, *VSG*3, *VSG*6, *VSG*9, and *VSG*21, were derepressed several hundred-folds to thousand-folds in *Tb*RAP1$^{F/\Delta ML}$ (Figure 3.11 A) and *Tb*RAP1$^{F/\Delta DB}$ (Figure 3.11 B) cells. Consequently, the DB region, which is required for *Tb*RAP1's telomere localization, plays an importance role in *VSG* silencing.

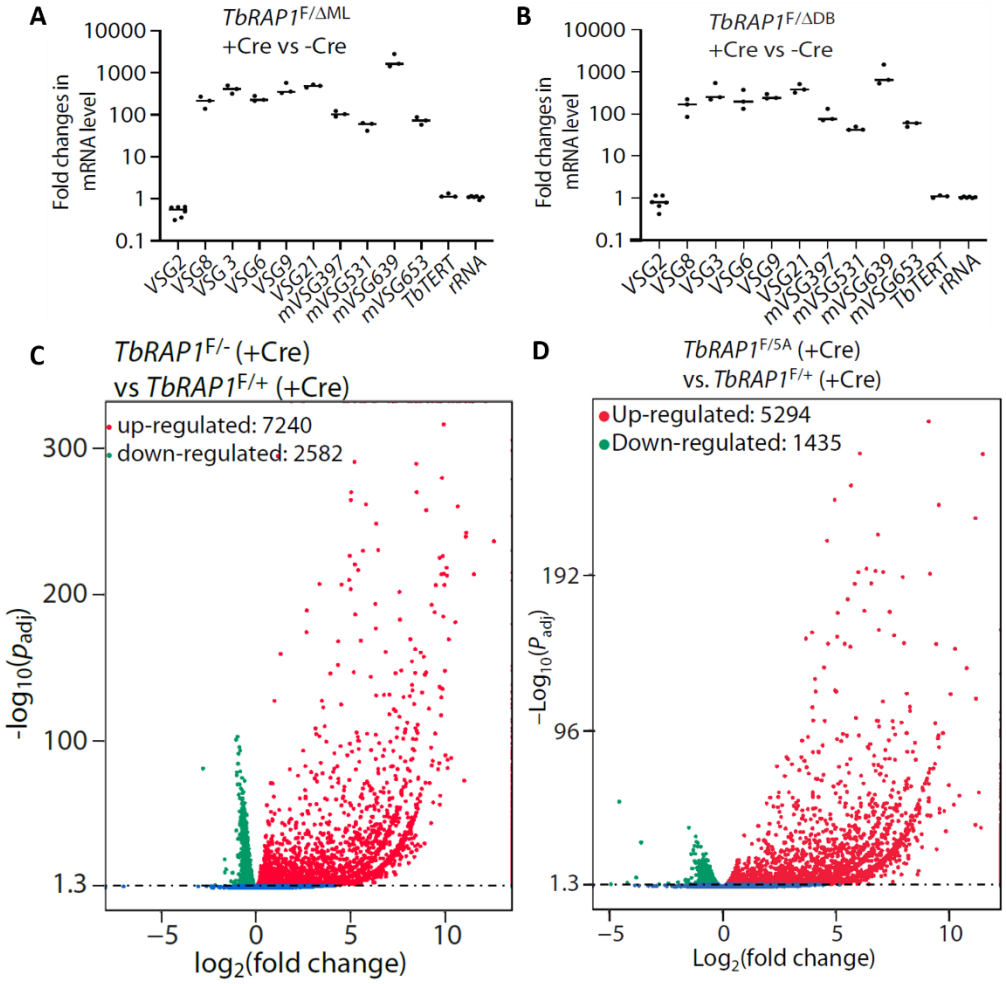


Figure 3.11 *Tb*RAP1's DNA binding activity is critical for *VSG* silencing. qRT-PCR of mRNA levels of the active *VSG*2, several silent ES-linked *VSG*s, and chromosome internal *Tb*TERT and rRNA in *Tb*RAP1$^{F/\Delta ML}$ (A) and *Tb*RAP1$^{F/\Delta DB}$ (B). A volcano plot of genes up-regulated and down-regulated in *Tb*RAP1$^{F/-}$ (C) and *Tb*RAP1$^{F/5A}$ (D) cells compared to *Tb*RAP1$^{F/+}$ cells 30 hours after Cre induction.

As mentioned in the introduction part, RAP1 homologs can regulate the expression of both subtelomeric and nontelomeric genes. Therefore, we hypothesized that *Tb*RAP1 might affect the expression of genes other than *VSG*s. To examine *Tb*RAP1's function in global gene expression, we performed RNA sequencing analysis in *Tb*RAP1$^{F/mut}$ that

lost DNA binding activity. As a control, we first compared the gene expression profile in *Tb*RAP1$^{F/-}$ with *Tb*RAP1 deletion and *Tb*RAP1$^{F/+}$ cells with *Tb*RAP1 WT after both were induced for Cre expression for 30 hours. Compared to *Tb*RAP1$^{F/+}$ cells, more than 7200 genes were up-regulated in *Tb*RAP1$^{F/-}$ cells, indicating *Tb*RAP1 suppresses global gene expression (Figure 3.11 C). At the same time, *Tb*RAP1$^{F/5A}$ cells showed more than 5200 gene up-regulated while only 1400 genes down-regulated (Figure 3.11 D), indicating *Tb*RAP1$^{F/5A}$ shares a similar transcriptome profile with *Tb*RAP1$^{F/-}$. Based on transcriptome profiles in *Tb*RAP1$^{F/5A}$, we suggest that the R/K patch within *Tb*RAP1 is also important for global gene expression regulation.

3.7 *Tb*RAP1's DNA binding activity is required for subtelomere/telomere region integrity

3.7.1 *Tb*RAP1's DNA binding activity protects telomere from DNA damage

In previous work related to *Tb*RAP1, we have known that *Tb*RAP1 suppresses *VSG* switching through maintaining telomere/subtelomere integrity. Telomeric/subtelomeric DNA damage, particularly that in the active ES, is a potent inducer of *VSG* switching (Glover*, et al.* 2013). To further investigate the function of *Tb*RAP1's DNA binding activity, we tested DNA damage in all *Tb*RAP1 mutants that lost DNA binding activity, including *Tb*RAP1$^{F/\Delta ML}$, *Tb*RAP1$^{F/\Delta DB}$, and *Tb*RAP1$^{F/5A}$, respectively. γH2A, a marker of DNA damage among eukaryotes, was used as a DNA damage indicator (Glover*, et al.* 2012).

The γH2A protein level in *Tb*RAP1 WT cells before and after phleomycin treatment was measured by western analysis as a positive control. While γH2A protein levels in *Tb*RAP1$^{F/mut}$ strains before and after Cre induction were analyzed by western blotting using γH2A rabbit antibody. Tubulin detected by antibody TAT-1 were used as a loading control.

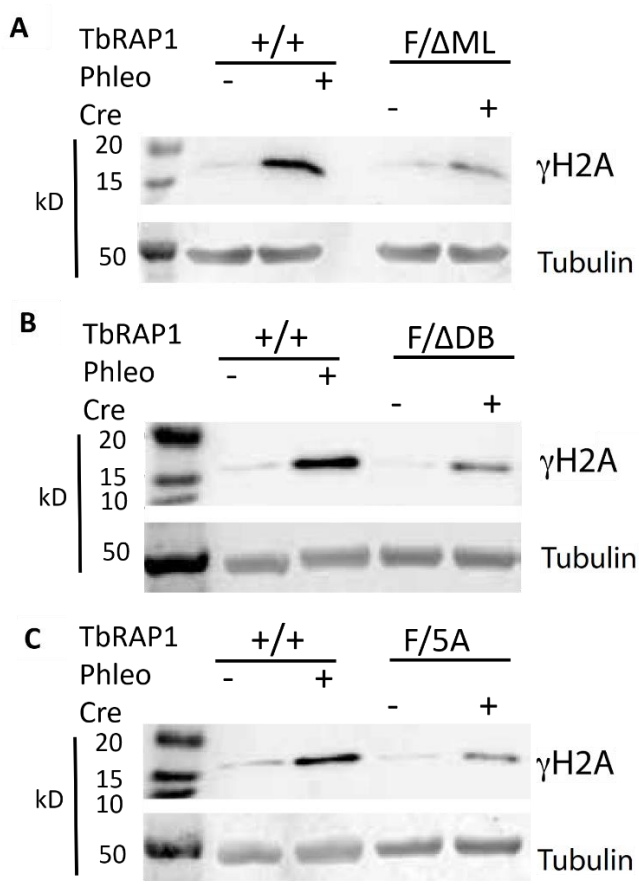


Figure 3.12 Western analysis shows that *Tb*RAP1's DNA binding activity protects telomere from DNA damages. Detect γH2A level in *Tb*RAP1$^{F/\Delta ML}$ (A), *Tb*RAP1$^{F/\Delta DB}$ (B), and *Tb*RAP1$^{F/5A}$ (C) cells before and after Cre treatment by western analysis, respectively. Tubulin level was detected as a loading control.

Western blotting showed that the γH2A level increased in WT cells after phleomycin treatment, a chemical that caused DNA damages. While γH2A accumulation also was detected in *Tb*RAP1$^{F/\Delta ML}$ (Figure 3.12 A), *Tb*RAP1$^{F/\Delta DB}$ (Figure 3.12 B), and *Tb*RAP1$^{F/5A}$ (Figure 3.12 C) cells after Cre induction, indicating that *Tb*RAP1's DNA binding activity protects telomere/subtelomere from DNA damages.

We also performed γH2A IF analyses in *Tb*RAP1$^{F/5A}$ cells before and after Cre induction. In IF analysis, *Tb*TRF, as a marker for the telomere, was detected by a *Tb*TRF rabbit antibody. While a γH2A rabbit antibody was used to detect γH2A in *Tb*RAP1$^{F/5A}$ cells before and after Cre induction. Nucleus was stained by DAPI. Afterward, the percentage of γH2A-positive nuclei in *Tb*RAP1$^{F/5A}$ cells before and after Cre induction was calculated.

The γH2A signal in IF analysis was faint (Figure 3.13 A), and there were only a few γH2A-positive nuclei (~7%) in *Tb*RAP1$^{F/5A}$ cells before Cre induction (Figure 3.13 B). In contrast, after Cre induction, the γH2A signal in IF analysis was very bright (Figure 3.13 A) and more than 90% of nuclei were γH2A positive (Figure 3.13 B). In addition, we detected *Tb*TRF in IF as a marker for the telomere and found that γH2A is partially colocalized with *Tb*TRF. These data suggest that some of the DNA damage occurs at the telomere vicinity when *Tb*RAP1 lost DNA binding activity.

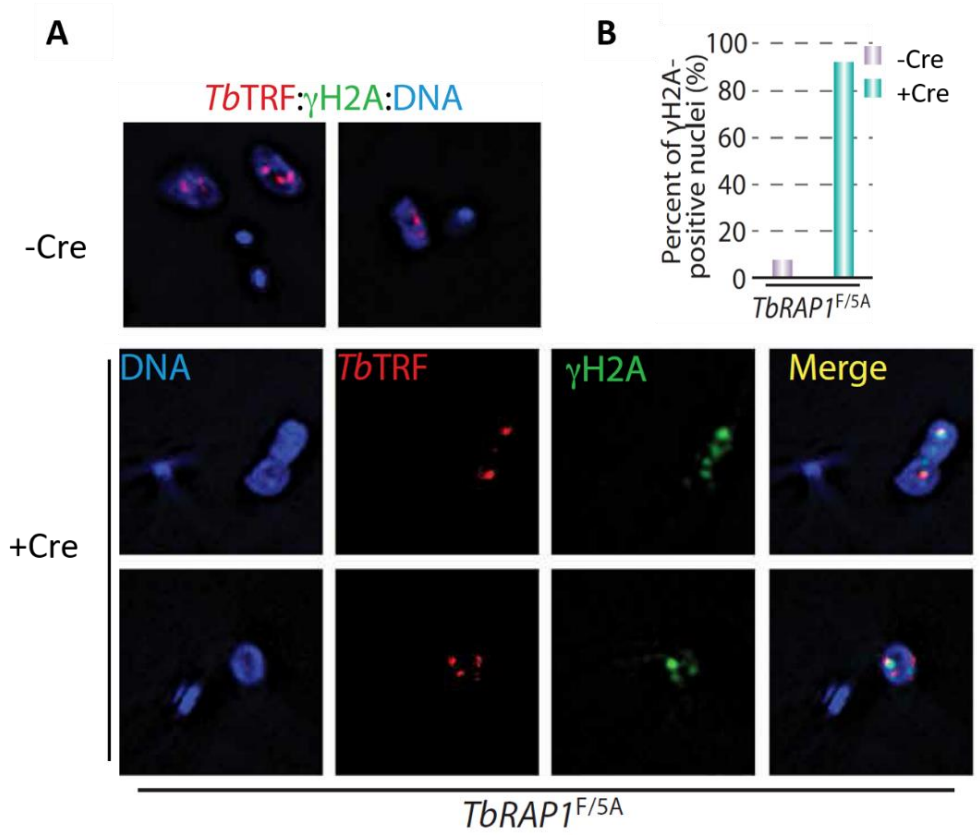


Figure 3.13 IF analysis indicates that DNA damage occurs at the telomere vicinity when *Tb*RAP1 lost DNA binding activity. (A) IF analysis in *Tb*RAP1$^{F/5A}$ cells. γH2A is labeled as green while *Tb*TRF is red. DAPI was used to stain nucleus. Merged images with signals from three channels are shown for -Cre cells (top). Images showing signals from individual and merged channels are shown in +Cre cells (bottom). (B**)** Profile of γH2A-positive nuclei percentage in *Tb*RAP1$^{F/5A}$ cells before and after Cre treatment.

3.7.2 *Tb*RAP1's DNA binding activity maintains subtelomeric region stability

To investigate the subtelomeric region stability when *Tb*RAP1 lost DNA binding active, ChIP assay between γH2A and subtelomeric DNA was performed in *Tb*RAP1$^{F/5A}$ cells. In ChIP assay, a γH2A rabbit antibody was used in *Tb*RAP1$^{F/5A}$

cells, followed by qPCR using primers specific to the indicated active and silent ES *loci.* SNAP50 is a chromosome internal gene.

ChIP results showed significantly more γH2A at the active ES, including *VSG*2 and 70-bp repeats loci after Cre induction in *Tb*RAP1$^{F/5A}$ cells (Figure 3.14). It indicates that there is an increased amount of DNA damage at the telomere and in the active ES located subtelomere when *Tb*RAP1 is no longer located at the telomere.

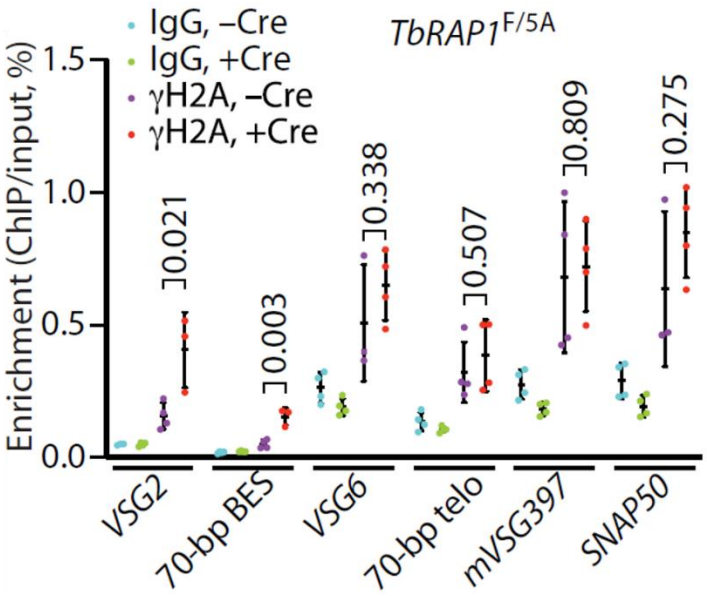


Figure 3.14 ChIP results indicate that subtelomeric DNA damage occurs when *Tb*RAP1 lost DNA binding activity. Average enrichment (ChIP/input) was calculated from three to six independent experiments.

Combined with these data, we have concluded that *Tb*RAP1's DNA binding activity is critical in maintaining subtelomeric/telomeric region stability. When *Tb*RAP1 lost DNA binding activity, more DNA damages were detected at subtelomeric/telomeric region.

4 Structural studies of *Tb*RAP1 Myb-like region

In previous studies, we have confirmed that *Tb*RAP1 Myblike domain has sequence non-specific DNA binding activity. This activity helps *Tb*RAP1 localize to telomere and plays a role in *VSG* switching suppression and subtelomeric /telomeric region stability. To further understand the function of *Tb*RAP1, we aim to determine the NMR structure of its MybLike domain (aa 639-761).

4.1 Expression and purification of *Tb*RAP1-ML

*Tb*RAP1-ML was cloned into pET32M vector with a thioredoxin (Trx) tag, a hexa-Histidine (His6) tag, and a human rhinovirus (HRV) 3C protease cleavage site. The recombinant gene was inserted into BamH1 and EcoRI restriction sites after the HRV 3C protease cleavage site, which enables the removal of the fusion tag by 3C protease digestion.

To express recombinant *Tb*RAP1-ML, the recombinant plasmid was transformed into *E. coli* BL21 (DE3). A single colony from transformation agar plate was taken into LB broth containing 100 μg/ml AMP, followed by overnight incubation at 250 rpm under 37 °C. Afterward, overnight culture was diluted into fresh LB broth containing 100 μg/ml AMP with a ratio of 1:100, and then incubated at 37 °C until OD_{600} reached 0.6-0.8. With 0.3 mM IPTG induction, bacteria were transferred to 30 °C and incubated for 6 hours. Finally, bacteria were harvest by centrifugation at 7000 rpm for 10 min.

During recombinant protein purification, the whole bacteria pellet was sonicated

in Tris buffer (50 mM Tris, pH 8.0, 150 mM NaCl) with 1 mM PMSF, 10 mM β-ME. The lysate after sonication was subjected to a high-speed centrifuge for 2 hours with 19, 000 rpm, 4 °C. Filter supernatant with 0.22um membrane followed by loading sample to His-column. Afterward, His-column was washed by His-binding buffer containing 25 mM imidazole to remove non-specific binding. The fusion protein with tag was on-column digested by HRV 3C protease. After digestion overnight, the fusion protein was divided into two parts including the Trx tag and target protein. After washing with Tris-buffer, the target protein was separated from this mixture.

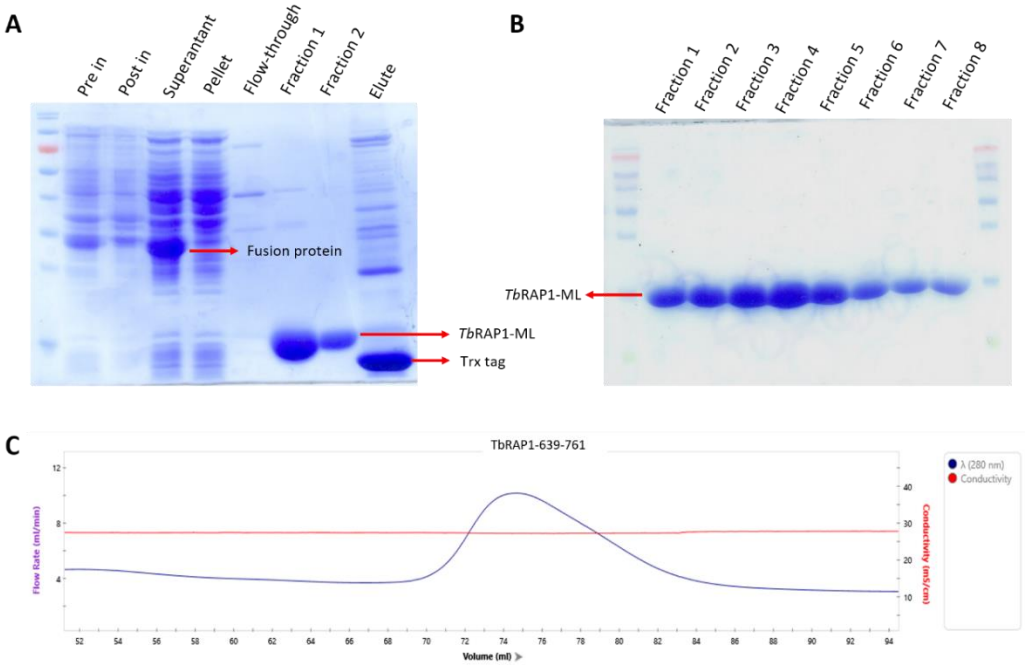


Figure 4.1 The purification of *Tb*RAP1-ML. (A) SDS-PAGE to characterize the expression level and purification of *Tb*RAP1-ML after affinity chromatography. Pre-in: the whole cell lysate before IPTG induction; Post-in: the whole cell lysate after IPTG induction; pellet: the pellet of cell lysate after centrifugation; supernatant: the supernatant of cell lysate after centrifugation; flow-through: the protein did not bind

with affinity column during loading sample. The target protein was eluted with Tris-buffer after on-column digestion with HRV 3C protease. While Trx-tag was washed out by Elution buffer with 500 mM imidazole. (B) SDS-PAGE of *Tb*RAP1-ML after gel filtration. (C) The profile of gel filtration.

As shown in SDS-PAGE analysis results (Figure 4.1 A), the fusion protein was well expressed after induced by 0.3 mM IPTG under 30°C. With sonication, the fusion protein was well dissolved in Tris buffer. After HRV 3C digestion, purified target protein *Tb*RAP1-ML was obtained. Afterward, the target protein without fusion tag was concentrated and then reloaded for later gel filtration purification. After gel filtration, *Tb*RAP1-ML was obtained finally (Figure 4.1 B-C).

4.2 Mass spectrometric analysis of *Tb*RAP1-ML

After purification, the molecular weight of *Tb*RAP1-ML was determined by MS. The mass spectrum result shows measured molecular weight is 14186.06Da (Figure 4.2), which is nearly equal to their theoretical values (14187.12Da), indicating the target proteins with high purify have been obtained.

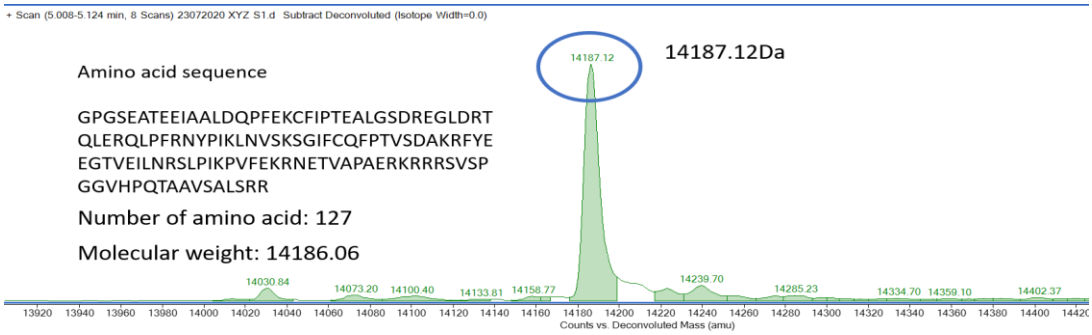


Figure 4.2 Mass spectrum of *Tb*RAP1-ML.

4.3 Structure analysis of *Tb*RAP1-ML by NMR

*Tb*RAP1-ML protein samples for NMR were isotopically labeled by expressing in M9 minimal medium with $^{15}NH_4Cl$ and ^{12}C- or ^{13}C-glucose. All NMR spectra recorded for the structure determination protocol were acquired at 10°C using 600 MHz Varian Inova spectrometers equipped with a cooled triple-resonance gradient probe. NMRPipe, NMRView, and Smartnotebook software were used respectively for data processing, data analysis, and assignment protocols. Backbone and aliphatic side chain assignments were obtained respectively from ^{1}H-^{15}N-HSQC, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HNHA, ^{1}H-^{13}C-HSQC experiments, and H(CCO)NH, (H)C(CO)NH, H(C)CH-TOCSY, (H)CCH-TOCSY experiments on the ^{15}N,^{13}C-labeled protein in 90% (v/v) H_2O/10% (v/v) D_2O. Resonances of the aromatic side chains were assigned through ^{1}H-^{13}C-HSQC and a 2D homo-nuclear NOESY experiment on a sample dissolved in 100 % (v/v) D_2O.

Based on a series of NMR data collection and analysis, we obtained a preliminary solution structure of *Tb*RAP1-ML. In total, more than 95% of backbone atoms and sidechain atoms were assigned in ^{1}H-^{15}N-HSQC spectrum (Figure 4.3).

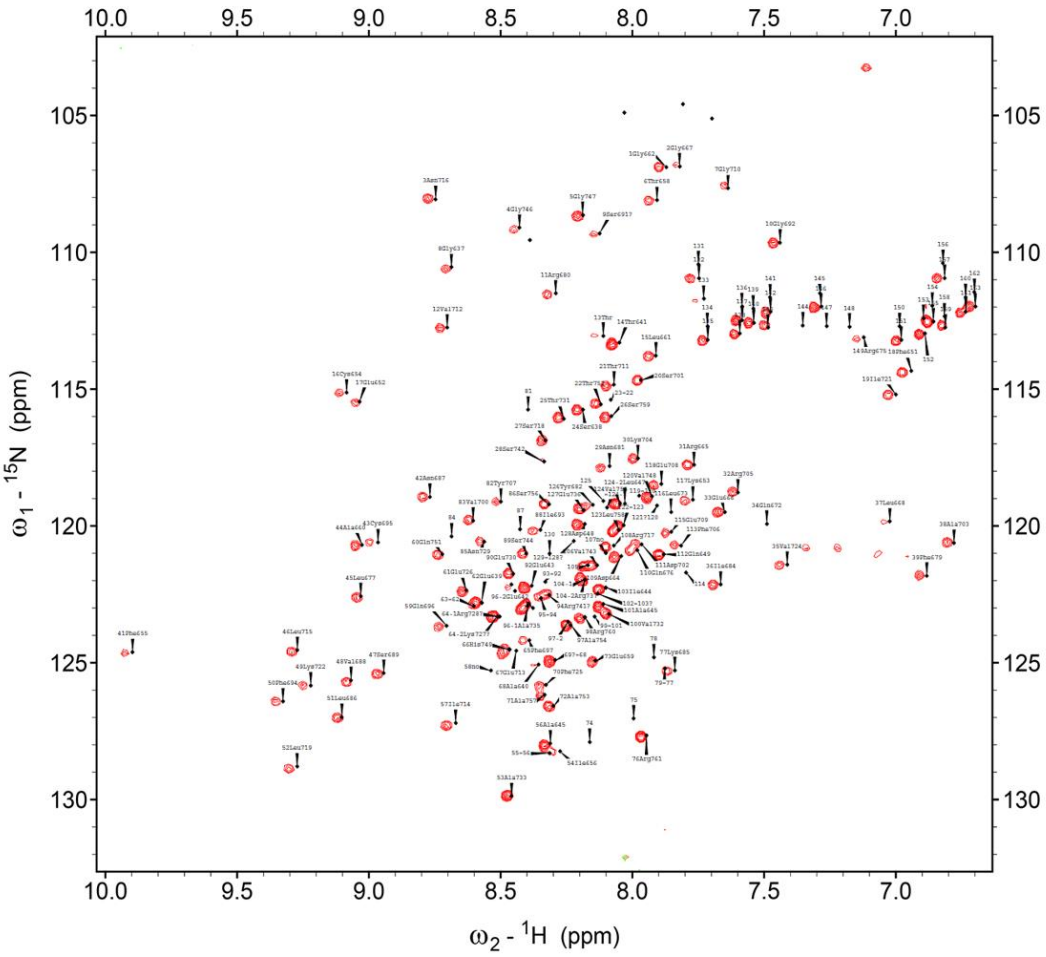


Figure 4.3 Assignment of *Tb*RAP1-ML in ^{1}H-^{15}N-HSQC spectrum.

4.4 The predicted Myb-like region of *Tb*RAP1 contains an unexpected RRM domain

According to NMR data analysis, secondary structure prediction of *Tb*RAP1-ML was carried out (Figure 4.4 A). Results show that it contains a globular domain (aa 639-733) followed by a long and flexible loop (aa 734-761). Specifically, the flexible loop represents the DB region identified in our previous study, which is required for targeting *Tb*RAP1 to the telomere region. Unexpectedly, the globular domain (aa 639-

733) is not the helix-turn-helix Myb domain folding as initially predicted. Instead, it shows a high similarity to the canonical RNA Recognition Motif (RRM) with the characteristic topology of a four-stranded anti-parallel β-sheet as the hydrophobic core plus two α-helices packed behind the β-sheet (Figure 4.4 B).

RRM, also known as the ribonucleoprotein (RNP) motif or RNA-binding domain (RBD), is one of the most abundant structural modules and has been identified in over 5,000 prokaryotic and eukaryotic proteins(Daubner, *et al.* 2013). The majority of RRM domains have RNA binding activity, some of them even can bind to both RNA and single-stranded DNA. With structural characterization of multiple RRM-RNA/ssDNA complexes, two RNP consensus sequences (RNP1 and RNP2 respectively) were thought to be involved in RNA binding activity. More specifically, RRM contains a conserved binding platform for RNA or ssDNA interaction. Nucleotides run across the surface of the four-stranded β-sheet within RRM and two of the nucleotides form direct stacking interactions with conserved aromatic residues on β1 and β3 strands.

Afterward, we superimposed *Tb*RAP1 RRM to the classic RRM domain of U1A spliceosome protein in complex with its cognate hairpin RNA (PDB ID: 1URN) by using structural analysis software WinCoot. *Tb*RAP1 RRM module superimposes well with U1A RRM domain with Root Mean Square Deviation (RMSD) of 3.54 Å for the main chain atoms (Figure 4.4 C). The core structural components such as the central β-sheet and two α-helices are highly similar between these two structures while the loops in between are highly flexible. Notably, F655 and F694 of *Tb*RAP1 match well to Y12

and F55 of U1A that are responsible for direct stacking interaction with the RNA substrate. These data further confirm that *Tb*RAP1 RRM is a canonical RRM module.

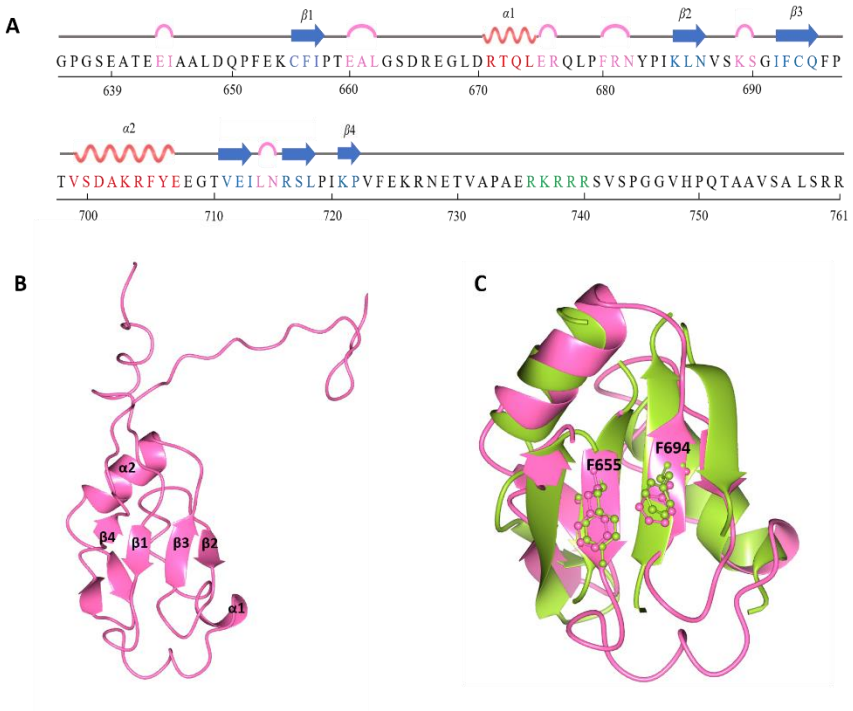


Figure 4.4 *Tb*RAP1-ML contains an RRM module followed by a long and flexible loop. (A) *Tb*RAP1-ML amino acid sequence information with secondary structure highlight. (B) Structure of *Tb*RAP1-ML. Secondary structure elements were labeled. (C) Superposition diagram between *Tb*RAP1 RRM and UIA (PDB: 1URN). *Tb*RAP1 RRM and its sidechain of two conserved aromatic residues (F655; F694) are highlighted in pink. U1A and its sidechain of two conserved aromatic residues (Y12; F55) are highlighted in green.

4.5 RRM of *Tb*RAP1 shows RNA/ssDNA binding potential

To explore whether the unexpected RRM module (*Tb*RAP1 639-733) have RNA or ssDNA binding potential, we also superimposed the *Tb*RAP1 RRM module to the

RRM1 domain of heterogeneous nuclear ribonucleoprotein (hnRNP) A1 in complex with either a short RNA oligo (PDB ID: 5MPG) (Figure 4.5 A) or a single-stranded telomeric DNA (PDB ID: 2UP1) (Figure 4.5 B). Similarly, the RMSD between the *Tb*RAP1 RRM module and the RRM1 domain of hnRNP A1 is 3.31 Å for the main chain atoms. Additionally, F655 and F694 of *Tb*RAP1 also superimpose exactly with F17 and F59 of hnRNP A1 that form stacking interaction with either the ssDNA or RNA substrates.

Furthermore, sequence alignment analysis reveals that the RRM domain of *Tb*RAP1 does contain the signature RNP1 and RNP2 motifs that are found in all RRM domains (Figure 4.5 C). Additionally, F655 and F694 of *Tb*RAP1 match exactly to the two most conserved aromatic residues within RNP1 and RNP2 that are critical for binding to ssDNA or RNA substrates. Taken together, our structural studies and sequence analysis suggest that the *Tb*RAP1 MybLike domain does contain a canonical RRM module that is structurally compatible with RNA or ssDNA binding.

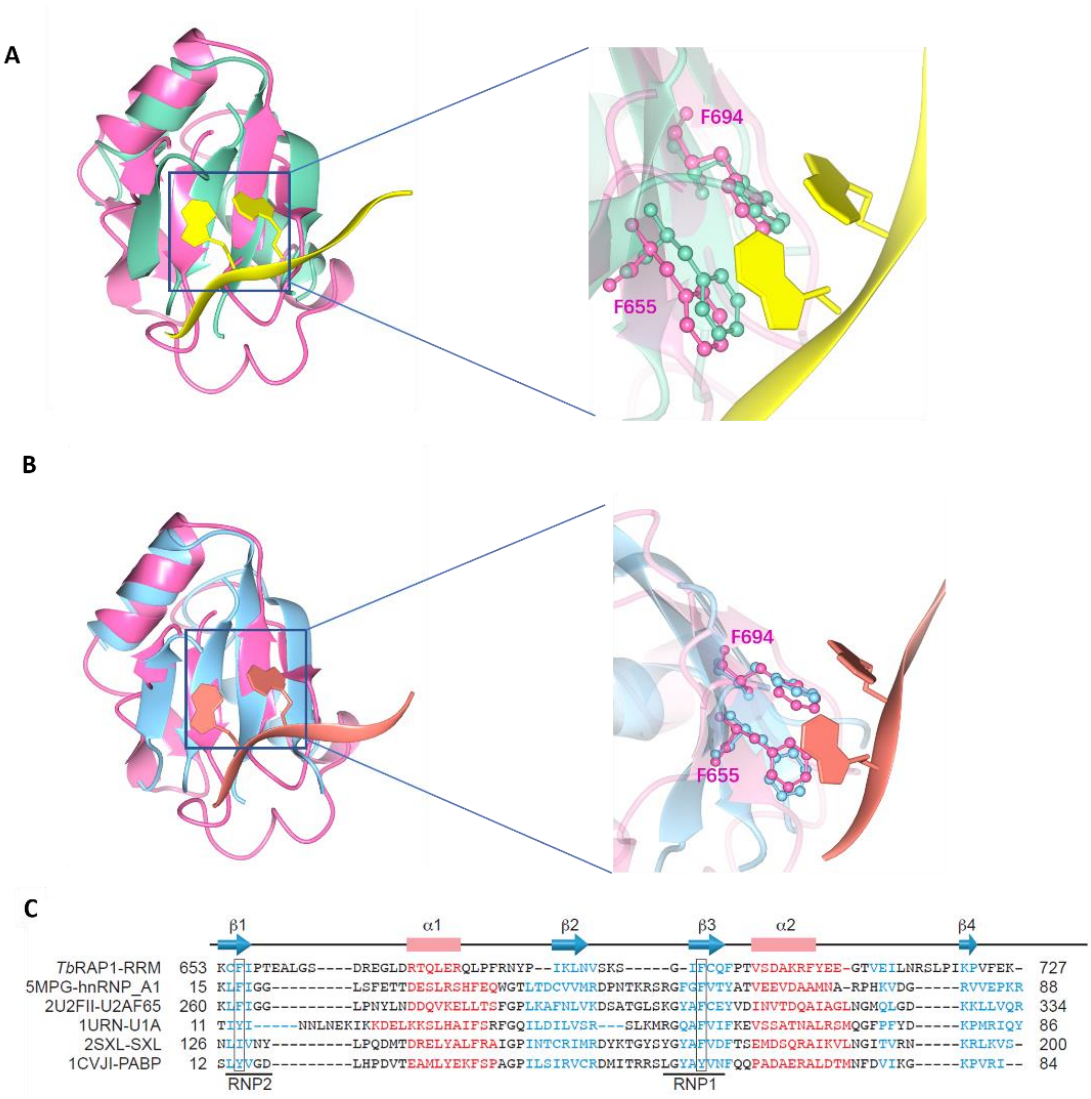


Figure 4.5 Structural superposition and sequence alignment results show that *Tb*RAP1 RRM has RNA/ssDNA binding potential. (A) *Tb*RAP1 RRM structural alignment with hnRNP A1 in complex containing RNA oligo (PDB: 5MPG). RNA oligo sequence: 5'-UUAGGG-3'. *Tb*RAP1 RRM and sidechain of two conserved aromatic residues (F655; F694) are highlighted in pink. hnRNP A1 RRM and sidechain of two conserved aromatic residues (F17; F59) are highlighted in medium aquamarine. RNA oligo sequence is highlighted in yellow. (B) *Tb*RAP1 RRM structural alignment with hnRNP A1 in complex containing DNA oligo (PDB: 2UP1). DNA oligo sequence: 5'-

TAGGGTTAGGG-3'. *Tb*RAP1 RRM and sidechain of two conserved aromatic residues (F655; F694) are highlighted in pink. hnRNP A1 RRM and sidechain of two conserved aromatic residues (F17; F59) are highlighted in light blue. DNA oligo sequence is highlighted in salmon. (C) Sequence alignment between *Tb*RAP1 RRM and varied RRMs.

5 Biochemical and functional characterization of potential RNA binding activity of *Tb*RAP1 and its role in *VSG* regulation

As described in chapter 4, we determined the NMR structure of *Tb*RAP1-ML containing an RRM module. With structural superposition and sequence alignment results, we predicted that *Tb*RAP1 RRM might exert its RNA binding activity through two conserved aromatic residues within the RRM module. To confirm our hypothesis, biochemical studies were conducted to investigate whether this RRM module binds to any single-stranded nucleic acids.

We first ruled out ssDNA as a potential substrate because our previous study (chapter 3) has demonstrated that the positively charged $_{737}$RKRRA$_{741}$ patch within *Tb*RAP1-ML is sufficient for binding both ssDNA and dsDNA. Consequently, we decided to focus our study on potential RNA-binding activity instead.

No RAP1 orthologs have been reported to contain an RRM module or have any RNA binding activity. We decided to test *Tb*RAP1's RNA binding activity by using the active *VSG* mRNA as substrate, since the active *VSG* is transcribed by RNA Pol I at a very high level. In particular, all *VSG*s have a common 14nt sequence in their 3'UTR regions(Cross et al., 2014; Ridewood et al., 2017). Due to *T. brucei* can express different *VSG*s at different times due to antigenic variation, it is very likely that *Tb*RAP1 can recognize all *VSG* RNAs if it binds to the active *VSG* mRNA. In addition, to detect *Tb*RAP-ML's RNA binding sequence preference, random RNA will also be applied in experiments.

5.1 *Tb*RAP1 RRM shows direct interaction with varied RNAs *in vitro*

5.1.1 EMSA data suggest that *Tb*RAP1-ML has RNA binding activity with RNAs

EMSA assay was carried out to detect *Tb*RAP1-ML's RNA binding activity. Firstly, radiolabeled full-length *VSG*13 and *VSG*16 mRNA were used as probes to detect EMSA shifts with addition of purified Trx-His6-tagged *Tb*RAP1-ML. Mixed different amounts of protein *Tb*RAP1-ML (0, 100, 250, and 500 ng, respectively) with 0.5 ng radiolabeled full-length *VSG*13 and *VSG*16 mRNA, respectively. After 30 min incubation, loading mixture into a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer and detecting gel shift under a phosphor imager. EMSA result showed that 100 ng *Tb*RAP1-ML was sufficient to cause full-length *VSG*13 mRNA shifts because of RNA-protein complex formation (Figure 5.1 A). Similar shifts were observed when mixed *Tb*RAP1-ML with full-length *VSG*16 mRNA (Figure 5.1 B). These data indicate that there is a direct interaction between *Tb*RAP1-ML and *VSG* mRNAs.

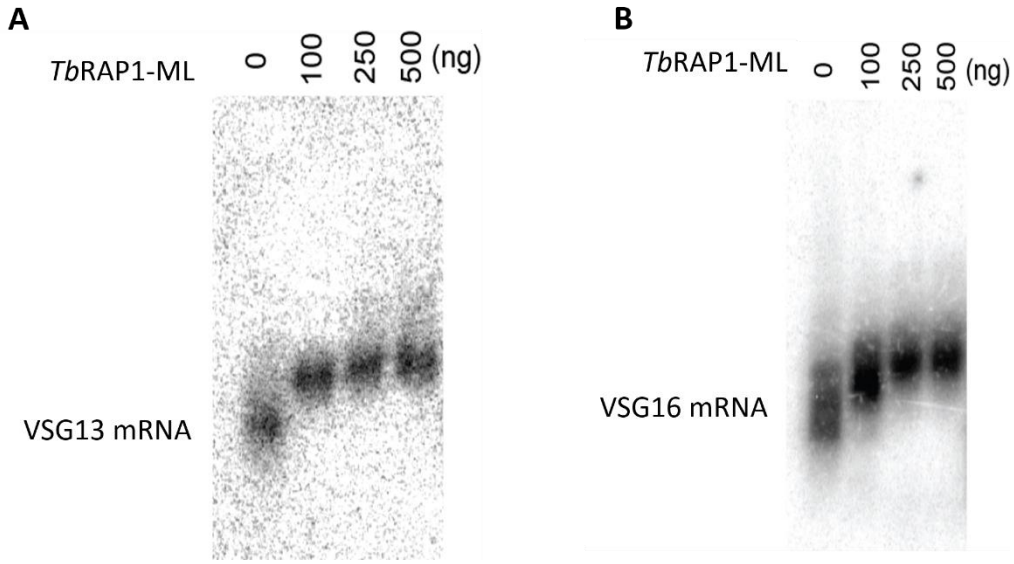


Figure 5.1 EMSA data show that *Tb*RAP1-ML binds to full-length *VSG* mRNAs.

(A) EMSA using Trx-His6-tagged *Tb*RAP1-ML and radiolabeled full-length *VSG*13 mRNA. (B) EMSA using Trx-His6-tagged *Tb*RAP1-ML and radiolabeled full-length *VSG*16 mRNA. The amount (ng) of protein used is indicated on top of each lane, respectively.

To investigate sequence preference of *Tb*RAP1-ML's RNA binding activity, *VSG*2 mRNA oligos and random sequence RNA oligos with the same length were used for EMSA assay. Briefly, mixed different amounts of protein *Tb*RAP1-ML (0, 5, 10, 20, 50, 100, 150, and 200 ng, respectively) with 0.5 ng radiolabeled 170 nt *VSG*2 mRNA oligos. After 30 min incubation, loading mixture into a 0.6% agarose gel in 0.5× Tris-borate EDTA running buffer. Afterward, gel shifts were detected under a phosphor imager. EMSA result showed that gel shifts became more obvious with the increasing amount of *Tb*RAP1-ML addition, indicating increased RNA-*Tb*RAP1-ML complex accumulation (Figure 5.2 A). When protein amount reached 50 ng, most of *VSG*2 mRNA oligos occurred shift. A similar interaction pattern also can be found when using random sequence RNA oligos as substrate (Figure 5.2 B). 50 ng of *Tb*RAP1-ML was sufficient to interact with all random RNA oligos.

Based on these EMSA data, we confirmed that *Tb*RAP1-ML, containing a conserved RRM domain, has sequence non-specific RNA binding activity, which can bind to both *VSG* mRNA and random sequence RNA.

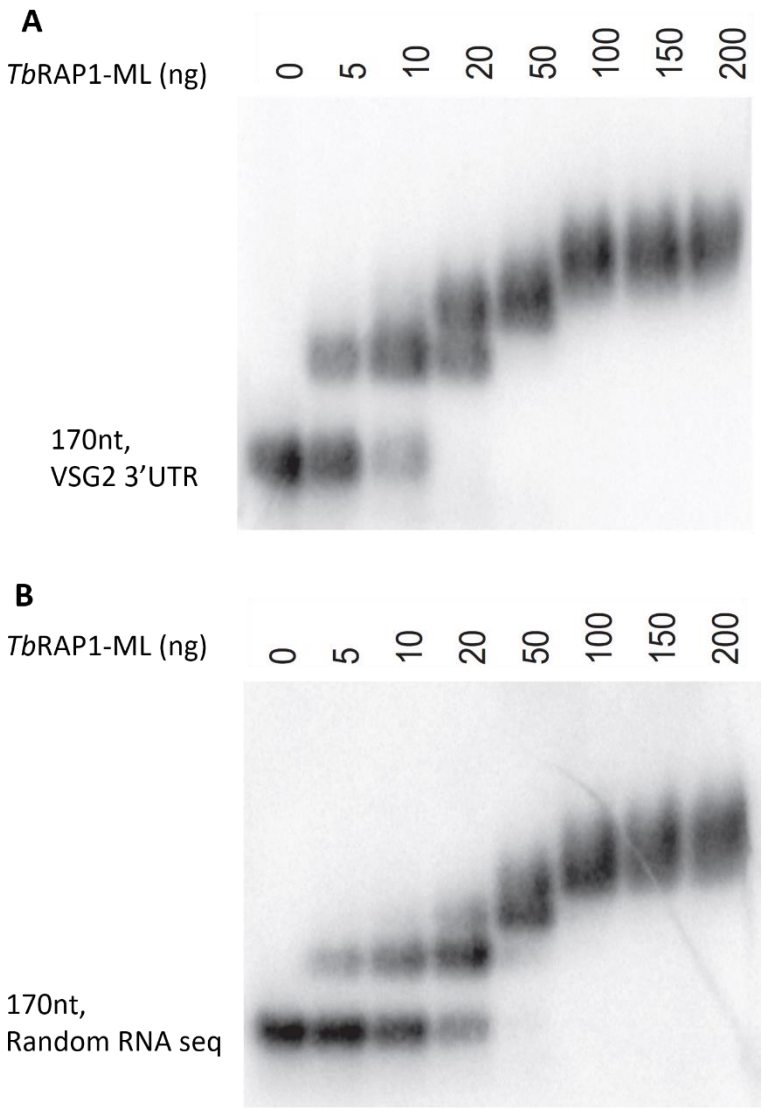


Figure 5.2 EMSA data demonstrate that *Tb*RAP1-ML binds to varied RNA oligos. (A) EMSA using Trx-His6-tagged *Tb*RAP1-ML and radiolabeled 170nt *VSG*2 3' UTR mRNA. (B) EMSA using Trx-His6-tagged *Tb*RAP1-ML and radiolabeled 170nt random RNA. The amount of protein used in each well is 0, 5, 10, 20, 50, 100, 150, and 200ng, respectively.

5.1.2 FP assay results confirm that *Tb*RAP1-ML binds to RNAs with moderate affinity

Fluorescent polarization assay was applied to further confirm the direct interaction between the *Tb*RAP1 RRM module and varied RNA oligos. We decided to test 34-nt *VSG* mRNA containing part of 3'UTR including the 16-mer consensus motif of all *VSG* 3'UTRs (termed "34-nt *VSG* mRNA"). Random sequence RNA oligo with the same length was also applied to FP assay as control.

RNA substrates were labeled with fluorescein 6-FAM at the 5' end and their fluorescence polarization was tracked after adding various amount of purified *Tb*RAP1-ML. Specifically, the concentration of protein ranging from 0 to 1200 μM mixed with 30 nM 5'6-FAM labeled RNAs. 34-nt *VSG* mRNA with the conserved 3'UTR sequence among *VSG*s and 34-nt random RNA were used, respectively. The buffer used in fluorescent polarization was described as below: 160 mM KCl, 40 mM sodium phosphate pH 7.4, 10% glycerol, 5 μM EDTA. After incubating the protein-RNA mixture for 30 min, fluorescent polarization value was observed by a microplate reader.

In the beginning, increased concentration of *Tb*RAP1-ML mixed with 34-nt *VSG* mRNA caused fluorescent polarization signal accumulation. When concentration of *Tb*RAP1-ML reached 450 mM, FP value became stable and was gradually saturated, with K_d estimated to be 282 μM (Figure 5.3 A). Similarity, *Tb*RAP1-ML showed direct interaction with 34-nt random RNA oligos with K_d 194 μM (Figure 5.3 B). These results

reveal that *Tb*RAP1-ML binds to both *VSG* mRNA and random RNA oligos with moderate affinity.

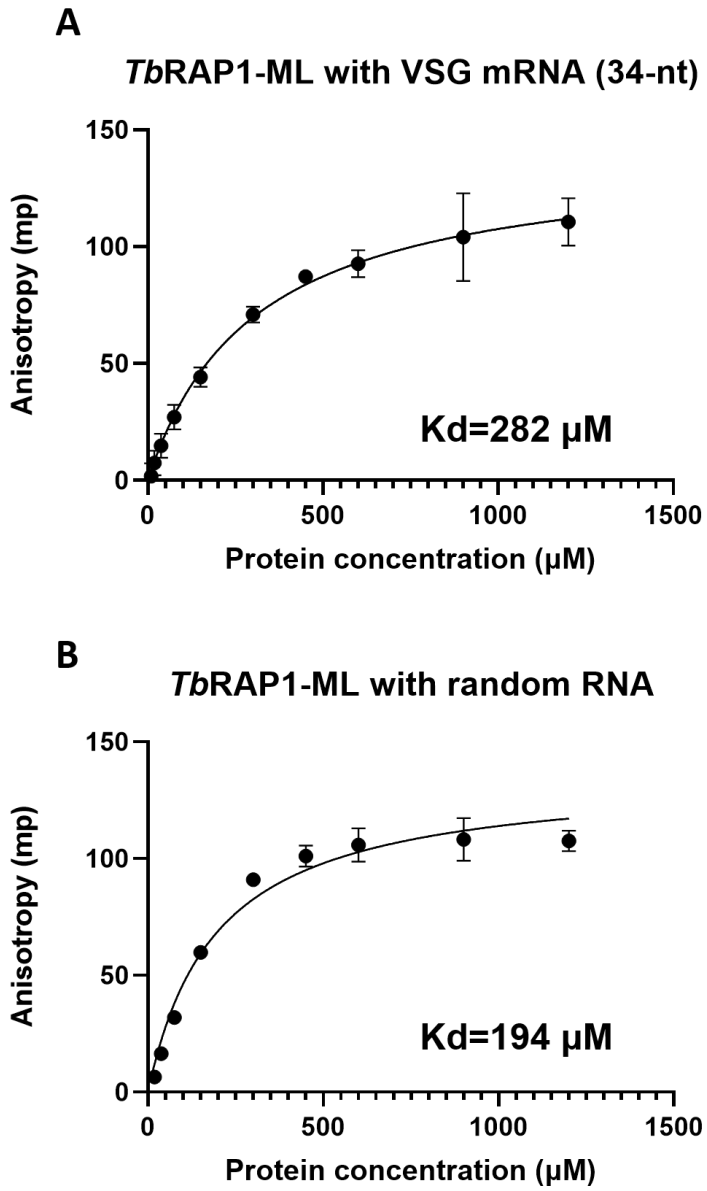


Figure 5.3 FP assay results demonstrated that *Tb*RAP1-ML interacts with varied RNAs. Three independent experiments were carried out for each assay and analyzed by Prism8 software. (A**)** FP profile between *Tb*RAP1-ML and 34-nt *VSG* mRNA**.** 34-nt *VSG* mRNA sequence: AAAACUUUUUGAUAUAUUUUAACACCAAAACCAG.

(B) FP profile between *Tb*RAP1-ML and 34-nt random RNA. 34-nt Random RNA sequence: ACAGACCACUACAAGAUACACAGUACAACCAACCA.

In summary, our binding assays confirmed that the *Tb*RAP1 RRM module does have RNA binding activity to both *VSG* mRNA and random RNA. However, the binding affinity is rather moderate compared to most RRM domains reported in the literature. Multiple factors may contribute to such moderate binding affinity, including but not limited to the short length of the RNA oligos used and limited RNA sequence tested especially for the *VSG* RNA molecule.

5.2 *Tb*RAP1-ML preferably binds to *VSG* mRNA through two conserved aromatic residues

Guided by previous EMSA and FP data, we have noticed that *Tb*RAP1-ML binds to both *VSG* mRNA and random RNAs with moderate binding affinity. However, biochemical characteristics of this protein-nucleic acid interaction are still unclear. Does *Tb*RAP1-ML's RNA binding activity have any sequence preference? Whether two conserved F residues within *Tb*RAP1 RRM module participate in this RNA binding activity? To answer these questions, we plan to further investigate the critical residues within *Tb*RAP1 that responsible for its RNA binding activity. To explore *Tb*RAP1-ML binding activity preference for RNA sequence, the binding patterns of *Tb*RAP1-ML to varied RNAs will be analyzed.

Compared with other *in vitro* assays, HSQC NMR titration assays not only can detect the interaction between protein and nucleic acid but also be able to directly

determine which amino acid may play critical roles within this interaction. After titrating nucleic acid substrate into ^{15}N-labeled protein, the greater chemical shifts caused by more obvious environment changes will be detected around amino acids which are closer to the binding region. Therefore, NMR titration assays were carried out to further investigate *Tb*RAP1-ML's RNA binding activity.

5.2.1 RRM module is involved in *Tb*RAP1-ML *VSG* mRNA binding activity

We performed NMR titration between *Tb*RAP1-ML and 34-nt *VSG* mRNA oligos. Firstly, the HSQC spectrum of ^{15}N-labeled *Tb*RAP1-ML was detected, and then 34-nt *VSG* mRNA oligo was gradually titrated into ^{15}N-labeled *Tb*RAP1-ML with protein-RNA molar ratio 1:1, 1:2, and 1:3, respectively. The HSQC spectra after each titration were monitored. After analyzing by Sparky software, we found that chemical shifts in ^{15}N-labeled *Tb*RAP1-ML became more obvious when more 34-nt *VSG* mRNA oligos titrated (Figure 5.4 A). These data were consistent with previous *in vitro* data and further confirmed that 34-nt *VSG* mRNA interacts with *Tb*RAP1-ML.

More importantly, concentration-dependent chemical shifts were observed for F655 and F694, the two conserved aromatic residues within *Tb*RAP1-ML that were predicted to be critical for RNA binding. Additionally, residues close to these two aromatic residues, including C654 and I656 (close to F655) as well as I693, C695, and Q696 (close to F694) also showed similar trend of chemical shifts (Figure 5.4 B). These data indicate that residues F655 and F694 participate in *Tb*RAP1-*VSG* mRNA binding activity.

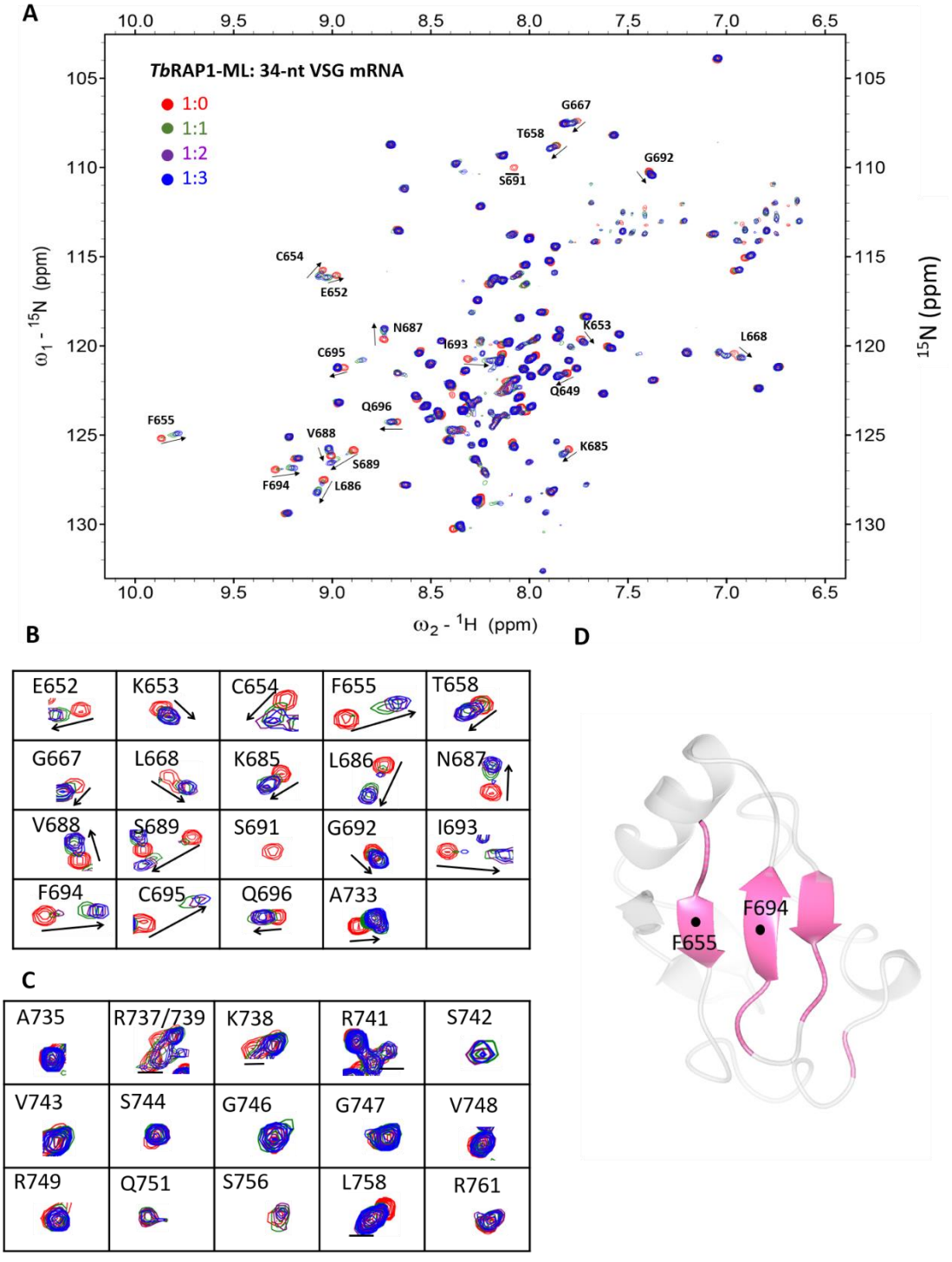


Figure 5.4 HSQC titration results show that *Tb*RAP1-ML binds to 34-nt *VSG* mRNA. (A) Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML in the absence (red) and presence of 34-nt *VSG* mRNA in 1× (green), 2× (purple), and 3×

(blue) molar excess. Chemical shift changed residues are labeled. Arrows indicate substantial chemical shifts after adding 34-nt *VSG* mRNA. (B) Chemical shifted residues within RRM module are labeled at box arrays. (C) ^{15}N-labeled *Tb*RAP1-ML residues within DB region showed no detectable chemical shifts with titrating of 34-nt *VSG* mRNA. (D) *Tb*RAP1-ML structure diagram highlighting all residues with chemical shifts. Two conserved aromatic residues (F655 and F655) are labeled as black cycles.

On the other hand, the residues within DB region, such as R737, K738, R739, and R741, showed no chemical shifts even under protein-RNA molar ratio 1:3 (Figure 5.4 C), indicating that DB region is not involved in *Tb*RAP1-*VSG* mRNA binding activity.

Highlighted all residues that caused chemical shift after titration in structure profile, we found that these residues are mostly located within the conserved RNP1 or RNP2 sequence motifs (Figure 5.4 D), two β strands that form the conserved RNA binding site in all RRM domains. These titration results suggest that the binding of *Tb*RAP1-ML to *VSG* mRNA depends on its RRM domain involving the conserved sequence motifs.

5.2.2 Two conserved F residues within RRM domain are indispensable for *VSG* mRNA binding activity

To confirm the importance of two conserved F residues for RNA binding activity, we generated 4 different mutants within *Tb*RAP1-ML, including 2FA, 2FS, 2FQ, and

2FL, respectively. Afterwards, we plan to obtain the HSQC spectrum with protein alone to check these mutants' folding in solution.

^{15}N-labeled recombinant *Tb*RAP1-ML mutants were cultured in M9 minimal media containing $^{15}NH_4Cl$. To collect HSQC spectrum, purified ^{15}N-labeled *Tb*RAP1-ML mutants were dissolved in buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10% D_2O, with 0.1 mM final concentration.

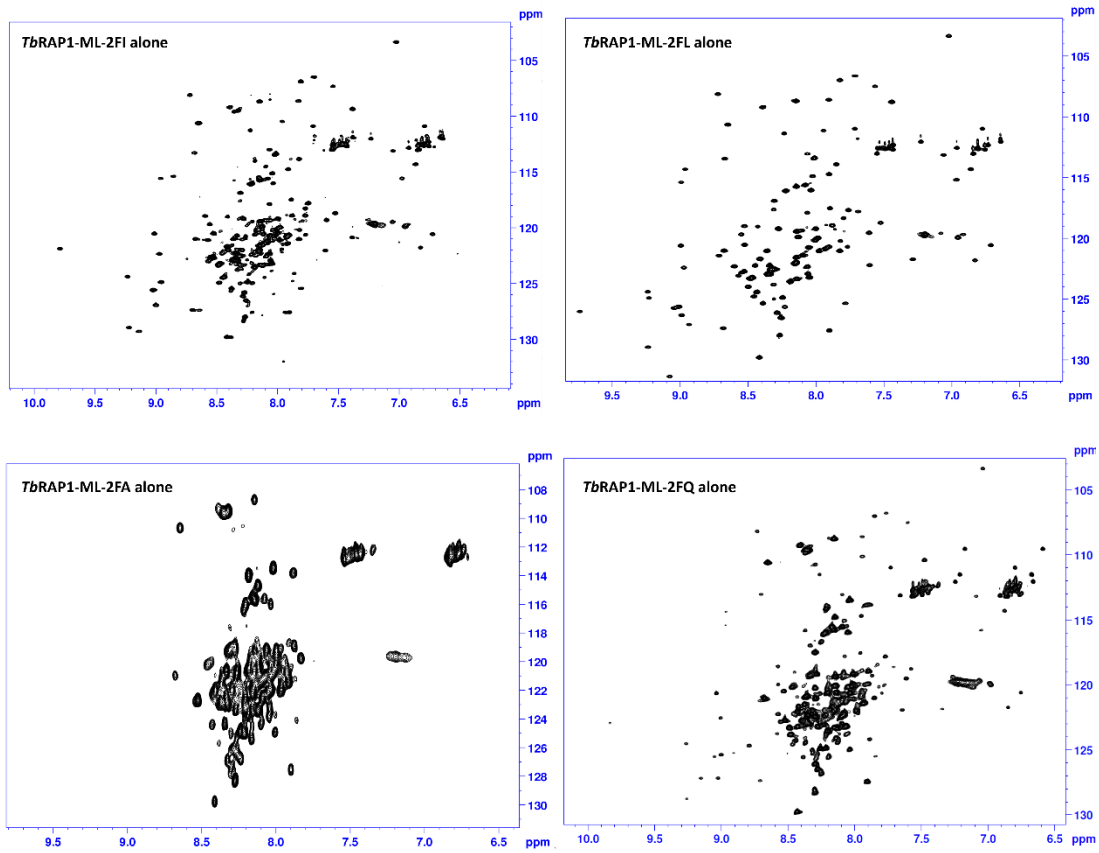


Figure 5.5 HSQC spectra of *Tb*RAP1-ML mutants showed that 2FL was well folded under this condition.

For 2FA and 2FQ mutants, the overall quality of the HSQC spectra significantly worsened with majority of the peaks showing poor dispersion. This change suggests

that the 2FA and 2FQ couldn't fold properly (Figure 5.5). In contrast, the HSQC spectrum of *Tb*RAP1-ML-2FL was largely identical to that for the wild-type, as all residues were clearly dispersed. It indicates that *Tb*RAP1-ML-2FL is well folded under this condition (Figure 5.5). Consequently, *Tb*RAP1-ML-2FL was used for HSQC titration assay to investigate the importance of 2F residues. Interaction between *Tb*RAP1-ML-2FL and 34-nt *VSG* mRNA were detected through NMR titration assay.

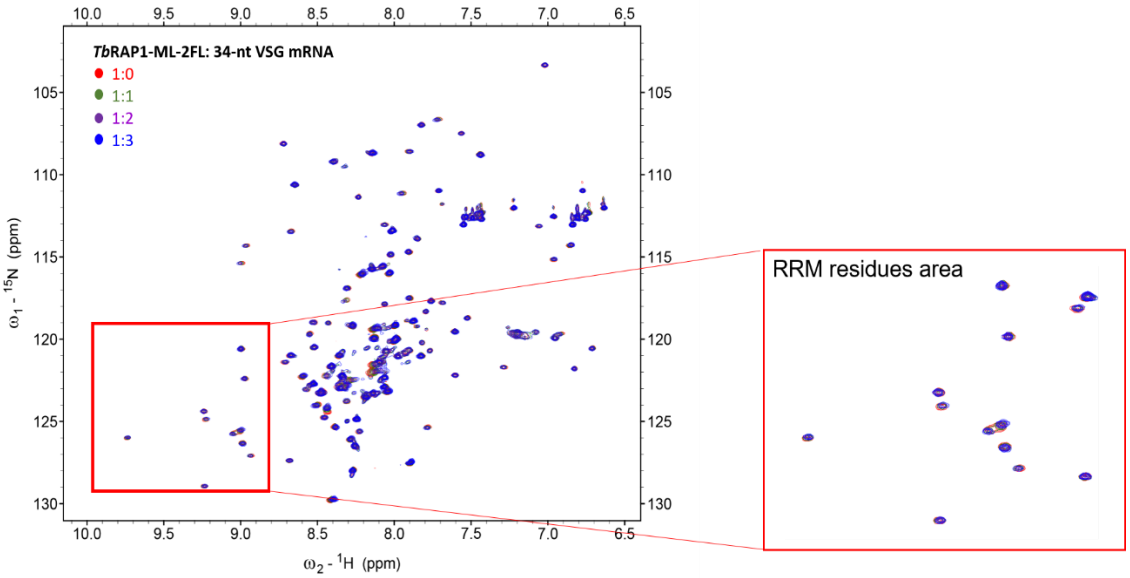


Figure 5.6 HSQC titration results show that *Tb*RAP1-ML-2FL fails to interact with 34-nt *VSG* mRNA. Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML-2FL in the absence (red) and presence of 34-nt *VSG* mRNA in 1× (green), 2× (purple), and 3× (blue) molar excess. Zoomed in spectra area that RRM residues located.

An increasing amount of 34-nt *VSG* mRNA oligo was titrated into ^{15}N-labeled *Tb*RAP1-ML-2FL with protein-RNA molar ratio 1:1, 1:2, and 1:3, respectively. The

HSQC spectra after each titration were collected and analyzed. HSQC titration data showed that no chemical shifts were observed, even the molar ratio of 2FL mutant to 34-nt *VSG* mRNA reached 1:3 (Figure 5.6). These data confirmed that the conserved F655 and F694 residues are critical for the *VSG* mRNA binding activity of *Tb*RAP1.

5.2.3 *Tb*RAP1-RRM specifically binds to *VSG* mRNA

The *Tb*RAP1-ML construct contains the RRM domain plus the C-terminal DB segment showed non-specific electrostatics-based binding to ssDNA and dsDNA as described in chapter 3. To assess whether the RRM domain is sufficient for *Tb*RAP1-ML's binding to *VSG* mRNA, we generated the *Tb*RAP1-RRM construct and assessed its binding to 34-nt *VSG* mRNA by NMR titration.

Before NMR titration assay, ^{15}N-labeled recombinant protein *Tb*RAP1-RRM was cultured in M9 minimal media containing $^{15}NH_4Cl$. High purity of ^{15}N-labeled *Tb*RAP1-RRM was diluted to 0.1 mM final concentration followed by dissolved in NMR buffer (20 mM sodium phosphate, pH 6.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) containing 10% D_2O.

In the beginning, the ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled *Tb*RAP1-RRM alone was collected to verify the folding of this new construct. The overall homogeneity of ^{15}N-labeled *Tb*RAP1-RRM was pretty good, and most of residues were visibly and able to distinguish from each other (Figure 5.7 A). It indicated that *Tb*RAP1-RRM was well folded under this condition. Consequently, NMR titration assay can be carried out to investigate *Tb*RAP1-RRM's RNA binding activities.

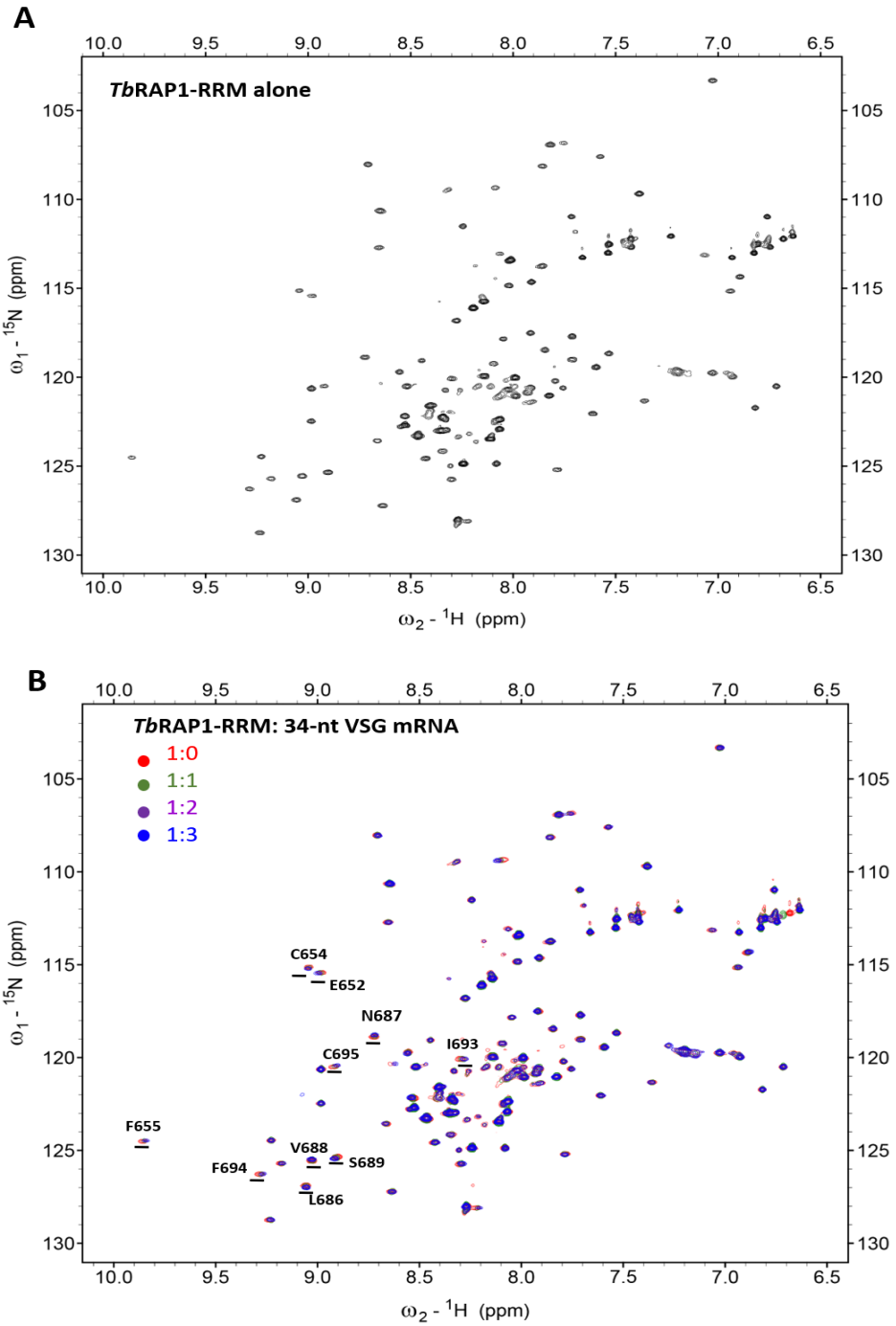

A
TbRAP1-RRM alone
B
TbRAP1-RRM: 34-nt VSG mRNA
1:0
1:1
1:2
1:3
C654
E652
N687
I693
C695
F655
V688
S689
F694
L686
ω2 - 1H (ppm)
ω1 - 15N (ppm)
C
Residues within β1
E652
C654
F655
Residues within β2
L686
N687
V688
S689
Residues within β3
I693
F694
C695

Figure 5.7 HSQC titration results display that *Tb*RAP1-RRM specifically binds to 34-nt *VSG* mRNA. (A) ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled *Tb*RAP1-RRM alone was collected. The protein concentration was 0.1 mM. (B) Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-RRM in the absence (red) and presence of 34-nt *VSG* mRNA in 1× (green), 2× (purple), and 3× (blue) molar excess. (C) Chemical shifted residues within RRM β*1*, β*2* and β*3* are labeled at box arrays, respectively. Arrows indicate substantial chemical shifts after adding the RNA substrate.

Afterward, 34-nt *VSG* mRNA was gradually titrated into ^{15}N-labeled *Tb*RAP1-RRM with protein-RNA molar ratio reached to 1:1, 1:2, and 1:3, respectively. Overlap the collected spectra, we found that there were detectable chemical shifts at residues F655 and F694 (Figure 5.7 B). Additionally, a few residues in these two F residues vicinity also occurred concentration-dependent chemical shifts, such as E652 and C654 (close to F655) as well as I693 and C695 (close to F694) (Figure 5.7 C). These chemical shifts are noticeably smaller than those observed in *Tb*RAP1-ML-34-nt *VSG* mRNA titration profile, suggesting that while the RRM domain binds to *VSG* mRNA through the conserved F655 and F694 residues, the DB domain may play a supporting role as well.

5.3 *Tb*RAP1-DB shows non-specific electrostatics-based binding to random RNA

Based on previous data, we have described that RRM module is critical for *Tb*RAP1's *VSG* mRNA binding. What is the function of DB region in *Tb*RAP1's RNA binding activity? The *Tb*RAP1-DB region (amino acid 734-761) has electrostatics-

based non-specific binding to ssDNA and dsDNA because of an R/K-rich patch ($_{737}$RKRRA$_{741}$). We hypothesized this region might bind to RNA in a manner similar to DNA.

5.3.1 *Tb*RAP1-ML binds to random RNA in a manner similar to DNA

To confirm our hypothesis, we firstly examined the interaction between *Tb*RAP1-ML and 35-nt random RNA through HSQC NMR titration. An increasing amount of random RNA oligo was titrated into ^{15}N-labeled *Tb*RAP1-ML with molar ratios 1:1, 2:1, and 3:1, respectively. The HSQC spectra after each titration were collected. Analyzing overlapped spectra by Sparky software, we found that many residues within *Tb*RAP1-ML occurred chemical shifts and disappeared after random RNA oligos titration (Figure 5.8 A). It indicates that *Tb*RAP1-ML also binds to random RNA.

Specifically, residues in DB region, such as K738, S742, G746, H749, Q751, L758, and R761, showed concentration-dependent chemical shifts after random RNA titration (Figure 5.8 B). While no such shifts were observed in the RRM domain, specifically in two conserved aromatic residues (F655 and F694). Upon protein to RNA molar ratio 1:1, some RRM residues possessed chemical shifts. However, when protein-RNA molar ratio reached 1:2 or 1:3, most of HSQC signals in RRM domain disappeared (Figure 5.8 C), including RRM β1 residues E652, K654 and C654 (close to F655), as well as RRM β3 residues I693 and C695 (close to F694). It is probably dues to structural unfolding induced by the excessive amount of RNA.

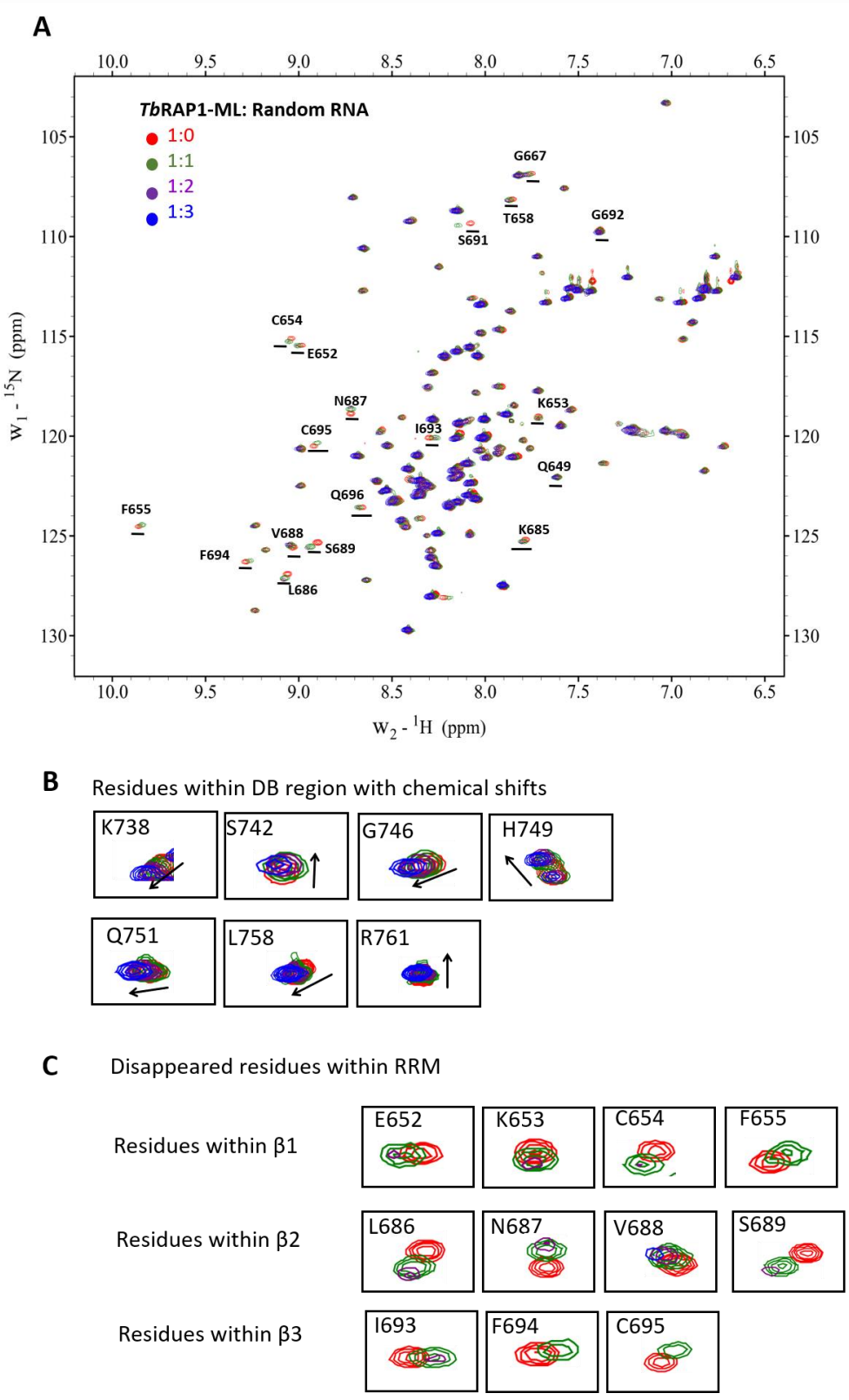


Figure 5.8 HSQC titration results show that *Tb*RAP1-ML binds to random RNA.

(A) Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML in the absence

(red) and presence of 35-nt random RNA in 1× (green), 2× (purple), and 3× (blue) molar excess. (B) Chemical shifted residues within DB region are labeled at box arrays. Arrows indicate substantial chemical shifts after adding the RNA substrate. (C) Disappeared residues within RRM β1, β2 and β3 module are labeled at box arrays. Random RNA sequence:ACAGACCACUACAAGAUACACAGUACAACCAACCA.

Interestingly, these chemical shifts observed in *Tb*RAP1-RNA NMR titration is largely identical to that observed in *Tb*RAP1-DNA titration. To compare the *Tb*RAP1's RNA titration profile with that of DNA, we overlapped the HSQC spectrum of titrating *Tb*RAP1-ML with 3 times ssDNA and 35-nt random RNA (Figure 5.9). After analyzing overlapped spectra by Sparky software, we observed *Tb*RAP1-ML binds to random RNA in a similar manner to random ssDNA.

Firstly, we found that part of residues in DB region occurred chemical shifts under both random RNA titration and ssDNA titration, such as residues A645, A733, S759, and R761 (labeled in yellow cycles). While the remaining DB region residues always showed no detectable chemical shifts no matter after ssDNA titration or random RNA titration, including E639, T641, and E751 (labeled in green cycles). Furthermore, most of RRM residues always disappeared after titrating with large amount of ssDNA or random RNA, including RRM β1 residues E652, K654, C654 and F655), as well as RRM β3 residues I693, C695, and F694. These data suggest that *Tb*RAP1 binds to random RNA through electrostatics-based non-specific binding, which is similar to DNA.

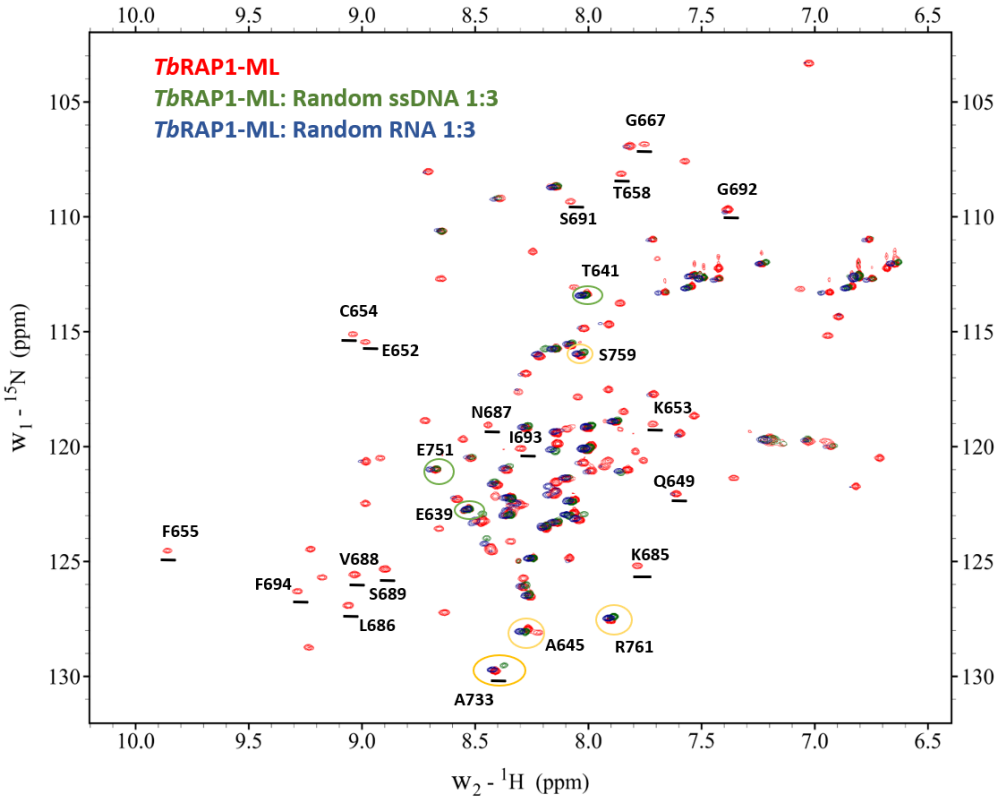


Figure 5.9 NMR titration analyses illustrate that *Tb*RAP1-ML binds to random RNA in a manner similar to DNA. NMR titration comparison between *Tb*RAP1-ML: random ssDNA (molar ratio 1:3, green) and *Tb*RAP1-ML: random RNA (molar ratio 1:3, blue). Spectrum of *Tb*RAP1-ML before titration is labeled as red. Residues always occurred chemical shifts in overlapped spectra are labeled by yellow cycle. Residues remain no chemical shifts in overlapped spectra are labeled by green cycle.

5.3.2 *Tb*RAP1-RRM fails to interact with random RNA

To further confirm the *Tb*RAP1 DB region binds to random RNA non-specifically through its positively R/K-rich patch, we performed HSQC NMR titration between *Tb*RAP1-RRM that without DB region and 35-nt random RNA.

We titrated a certain amount of 35-nt random RNA into *Tb*RAP1-RRM with protein-RNA molar ratio gradually increased from 1:1 to 1:3 (Figure 5.10). However, no chemical shifts were observed, and all residues were well overlapped even protein-RNA molar ratio reached 1:3. These data suggest that *Tb*RAP1-RRM loses its random RNA binding activity completely.

In summary, HSQC NMR titration studies revealed that random RNA failed to induce any chemical shifts in *Tb*RAP1-RRM suggesting that the R/K-rich path is critical for *Tb*RAP1's non-specific RNA binding activity.

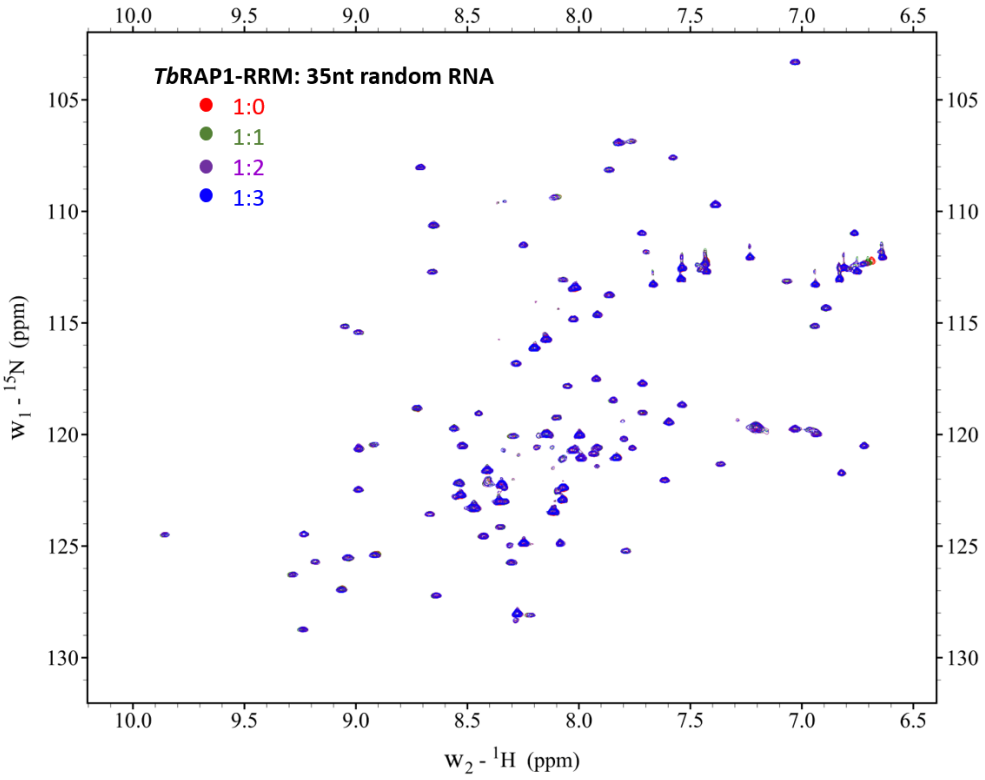


Figure 5.10 HSQC titration results indicate that *Tb*RAP1-RRM fails to interact with 35-nt random RNA. Four overlapped HSQC NMR spectra of ^{15}N-labeled

*Tb*RAP1-ML-5A in the absence (red) and presence of TERRA in 1× (green), 2× (purple), and 3× (blue) molar excess.

5.3.3 *Tb*RAP1-ML-2FL retains its non-specific RNA binding activity through DB region

In the previous section, the DB region was demonstrated be indispensable for *Tb*RAP1's non-specific RNA binding activity. We next investigated whether the non-specific RNA binding activity also relies on its RRM domain, the binding of *Tb*RAP1-ML-2FL to 35-nt random RNA oligo was tested.

Firstly, an increasing amount of 35-nt random RNA was titrated into ^{15}N-labeled *Tb*RAP1-ML-2FL with molar ratios 1:1, 2:1, and 3:1, respectively. The HSQC spectra after each titration were collected. After analyzing overlapped spectra by Sparky software, we found that some residues within DB region possessed detectable concentration-dependent chemical shifts, such as Q751 and H749 (Figure 5.11). It indicates that *Tb*RAP1-ML-2FL keeps its random RNA-binding activity.

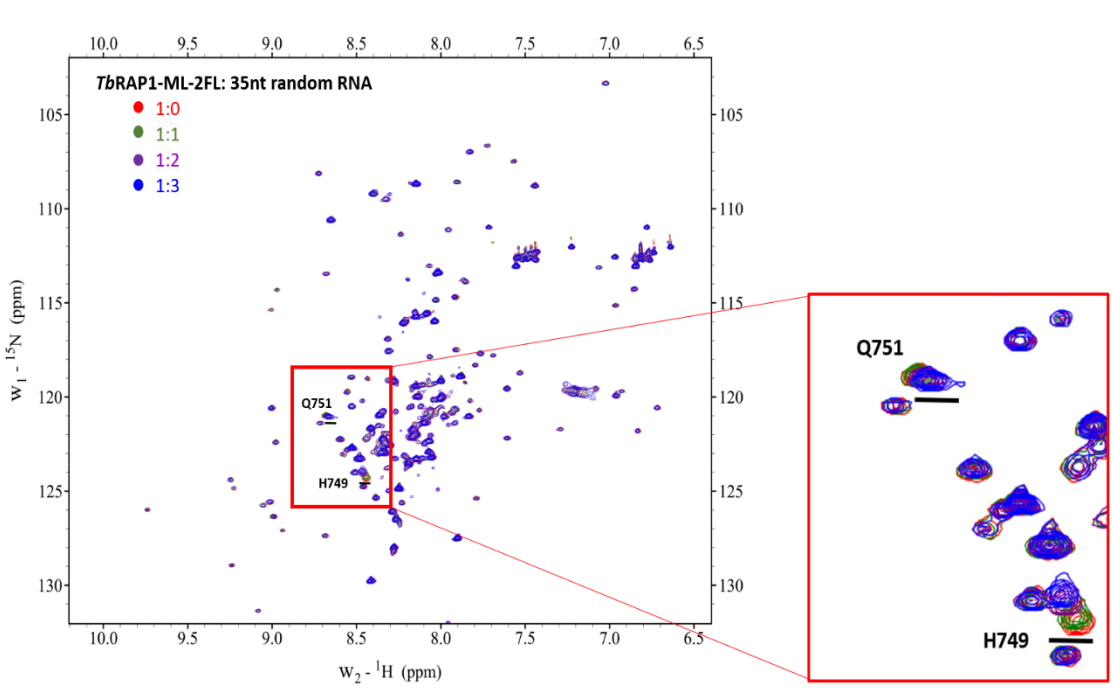


Figure 5.11 HSQC titration results demonstrate that *Tb*RAP1-ML-2FL retains its non-specific RNA binding activity through DB region. Four overlapped HSQC NMR spectra of ^{15}N-labeled *Tb*RAP1-ML-2FL in the absence (red) and presence of TERRA in 1× (green), 2× (purple), and 3× (blue) molar excess. Zoomed in spectra area which RRM residues located. Arrows indicate substantial chemical shifts after adding the RNA substrate.

In summary, NMR titration studies reveal that random RNA was able to induce measurable chemical shifts in the *Tb*RAP1 DB region. It was further confirmed that DB region is sufficient to cause weak and non-specific RNA binding. While this function does not rely on the RRM domain which preferably binds to *VSG* mRNA.

5.4 Both of RRM and DB region are required for *Tb*RAP1's active *VSG* mRNA association *in vivo*

5.4.1 Establishing *Tb*RAP1 mutant strains used for RNA binding activity investigation

In previous studies, our HSQC NMR titration experiments have confirmed that the *Tb*RAP1-ML possesses moderate RNA binding activity mediated by two structural regions. On the one hand, the RRM domain binds preferably to *VSG* RNA with the conserved F655 and F694 residues at the canonical RNA binding site playing critical roles. On the other hand, the DB region shows electrostatics-based non-specific RNA binding mediated by its R/K-rich patch in a manner similar to its binding with DNA. Guided by these data, our collaborator further examined whether *Tb*RAP1 has RNA binding activity *in vivo*.

To investigate the interaction between *Tb*RAP1 and active *VSG* mRNA, we established the *Tb*RAP1$^{F2H+/-}$ strain, in which one *Tb*RAP1 allele is deleted and the other has an N-terminal FLAG-HA-HA (F2H) tag.

To further determine whether *Tb*RAP1-ML mutants interact with the active *VSG* mRNA, Cre-loxP-mediated conditional deletion system was carried out. We replaced the WT *Tb*RAP1 allele of *Tb*RAP1$^{F/+}$ with F2H-tagged mutants to generate *Tb*RAP1$^{F/mut}$ strains as described in the below table, including *Tb*RAP1 $^{F/\Delta ML}$, *Tb*RAP1 $^{F/\Delta RRM}$, *Tb*RAP1 $^{F/2FQ}$, *Tb*RAP1 $^{F/\Delta DB}$, *Tb*RAP1 $^{F/5A,}$ and *Tb*RAP1 $^{F/2FA\&5A}$, respectively.

The genotypes of all *Tb*RAP1$^{F/mut}$ strains were confirmed by Southern and sequencing analyses.

Table List of *Tb*RAP1 mutant strains used for RNA binding activity investigation

Strains	Description
*Tb*RAP1$^{F/+}$	One floxed allele and one WT *Tb*RAP1
*Tb*RAP1$^{F/\Delta ML}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged mutant lacking the ML region containing
*Tb*RAP1$^{F/\Delta RRM}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged mutant lacking the RRM domain
*Tb*RAP1$^{F/2FQ}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged F655Q and F694Q mutant
*Tb*RAP1$^{F/\Delta DB}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged mutant lacking the DB region
*Tb*RAP1$^{F/5A}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged $_{737}$RKRRR$_{741}$ to 5A mutant
*Tb*RAP1$^{F/2FA\&5A}$	One floxed allele and one N-terminally F2H- and SV40 NLS-tagged F655A, F694A and $_{737}$RKRRR$_{741}$ to 5A mutant

5.4.2 *Tb*RAP1 interacts with active *VSG* mRNA *in vivo*

To detect *Tb*RAP1's interaction with active *VSG* mRNA, we did RNA IP assays with *Tb*RAP1$^{F2H+/-}$ and *Tb*RAP1$^{F/mut}$ strains. Firstly, western blot assays were performed to confirm expression level of *Tb*RAP1 mutants. The whole-cell lysates were prepared from the *Tb*RAP1$^{F/mut}$ strains. Results showed that F2H-*Tb*RAP1 ΔRRM and F2H-NLS-*Tb*RAP1 ΔML were expressed at a subtly lower level than the F2H-*Tb*RAP1 WT (Figure 5.12 A). While F2H-*Tb*RAP1-ΔDB, 2FQ (Figure 5.12 B), 5A (Figure 5.12 C), as well as 2FA&5A (Figure 5.12 D) were expressed at the same level as F2H-*Tb*RAP1 WT.

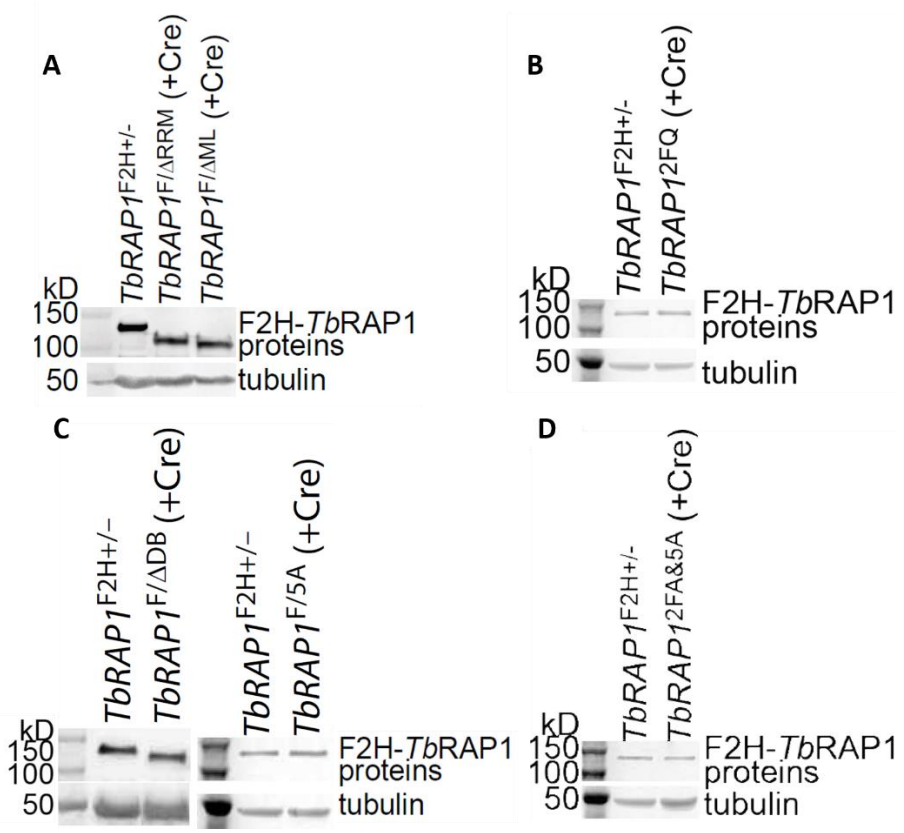


Figure 5.12 Expression levels of F2H-tagged *Tb*RAP1 proteins were detected by western blots. The HA antibody 12CA5 was used to detect the expression level of

*Tb*RAP1 ΔML and ΔRRM (A), 2FQ (B), ΔDB, and 5A (C), as well as 2FA&5A(D). Tubulin was detected by antibody TAT-1 used as a loading control.

Afterward, we did RNA IP with the HA monoclonal antibody 12CA5 (MSKCC monoclonal AB Core) in *Tb*RAP1$^{F2H+/-}$ strain. If any FLAG-HA-HA (F2H) tagged *Tb*RAP1 WT interacts with *VSG* mRNA, enriched *VSG* mRNA will be detected in the immunocomplexes through qRT-PCR analysis. Meanwhile, IgG was used as a negative control in the RNA IP assay. qRT-PCR analysis showed that a significant amount of *VSG*2 mRNA was immunoprecipitated by FLAG-HA-HA (F2H) tagged *Tb*RAP1 WT (Figure 5.13, lane of F2H+/-). Given *VSG*2 is the active *VSG* in these cells, the data confirmed that *Tb*RAP1 binds to the active *VSG* mRNA *in vivo*.

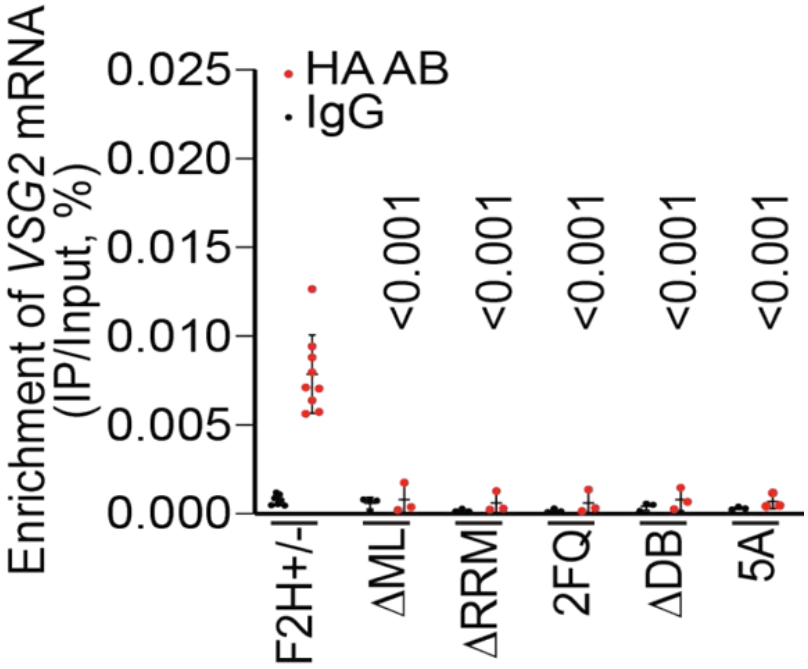


Figure 5.13 RNA-IP results among *Tb*RAP1$^{F2H+/-}$ and various *Tb*RAP1$^{F/mut}$ strains. Average enrichment was calculated from three to eleven independent experiments. *p*

values are indicated. Enrichment of *VSG* mRNA in *Tb*RAP1$^{F2H+/-}$ and various *Tb*RAP1$^{F/mut}$ strains, respectively.

5.4.3 Two conserved F residues within RRM is critical for active *VSG* mRNA association *in vivo*

Whether *Tb*RAP1 associate with *VSG* mRNA depends on ML domain? And whether two conserved F residues within RRM are critical for this association? To further investigate the role of two conserved F residues in *Tb*RAP1's association with *VSG* mRNA association, we also performed RNA IP assay in *Tb*RAP1 $^{F/mut}$ strains which depleted RRM module or mutated the two conserved F residues in RRM module, including *Tb*RAP1 $^{F/\Delta ML}$, *Tb*RAP1 $^{F/\Delta RRM,}$ and *Tb*RAP1 $^{F/2FQ}$, respectively.

Afterward, RNA IP experiments were conducted in *Tb*RAP1$^{F/\Delta ML}$, *Tb*RAP1$^{F/\Delta RRM}$, and *Tb*RAP1$^{F/2FQ}$, respectively. After treating with Cre for 30 hours, the floxed WT *Tb*RAP1 allele was removed in these three *Tb*RAP1 $^{F/mut}$ strains (ΔML, ΔRRM, and 2FQ mutants). Then, we can specifically examine *Tb*RAP1 mutants' behavior without the influence from the WT protein. The pull-down RNA products were analyzed by qRT-PCR using *VSG*2-specific primers. Results showed that deletion the RRM module or mutation of the two conserved F residues abolished the association of *Tb*RAP1 with active *VSG*2 mRNA (Figure 5.13).

5.4.4 *Tb*RAP1 DB region is necessary for targeting telomere to further associate with active *VSG* mRNA

5.4.4.1 *Tb*RAP1 with depletion or mutation of DB region fail to interact with active *VSG* mRNA

To further investigate the role of DB region in *Tb*RAP1's *VSG* mRNA association, we carried out RNA IP assay in *Tb*RAP1 $^{F/mut}$ strains which depleted DB region or mutated 5R/K to 5A in DB region, including *Tb*RAP1 $^{F/\Delta DB}$ and *Tb*RAP1 $^{F/5A}$.

After incubating with Cre for 30 hours, the whole cell lysates were collected for RNA IP assays. The IPed products were further analyzed by qRT-PCR using *VSG*2-specific primers. Interestingly, deletion of the DB domain or mutation of the $_{737}$RKRRR$_{741}$ patch (ΔDB and 5A mutants) also abolished *Tb*RAP1's interaction with active *VSG*2 mRNA *in vivo* (Figure 5.13). Therefore, the *Tb*RAP1 DB domain appears to be important for its random RNA binding activity.

5.4.4.2 DB region of *Tb*RAP1 is important for *Tb*RAP1 telomere localization

How to explain *VSG* mRNA's dissociation with *Tb*RAP1 mutants (*Tb*RAP1$^{F/\Delta DB}$ and *Tb*RAP1$^{F/\Delta 5A}$)? On the one hand, DB region has non-specific RNA binding activity through positive charged R/K residues *in vitro*. On the other hand, guided by the results of chapter 3, *Tb*RAP1 without DNA binding activity, a critical function of DB region, failed to localize to telomere. Therefore, we hypothesis that *Tb*RAP1 DB region is required for telomere localization to further associate with *VSG* mRNA.

To confirm our hypothesis, ChIP assay was performed in strains which mutated the two conserved F residues in RRM module (*Tb*RAP1$^{F/2FQ}$) and mutated both two conserved F residues and 5R/K residues (*Tb*RAP1$^{F/2FA\&5A}$). The WT *Tb*RAP1 allele was removed by Cre to specifically examine *Tb*RAP1 mutants' behavior without the influence from the WT protein. *Tb*RAP1$^{F2H+/-}$ which contains full-length *Tb*RAP1 WT was employed as a positive control for ChIP assay.

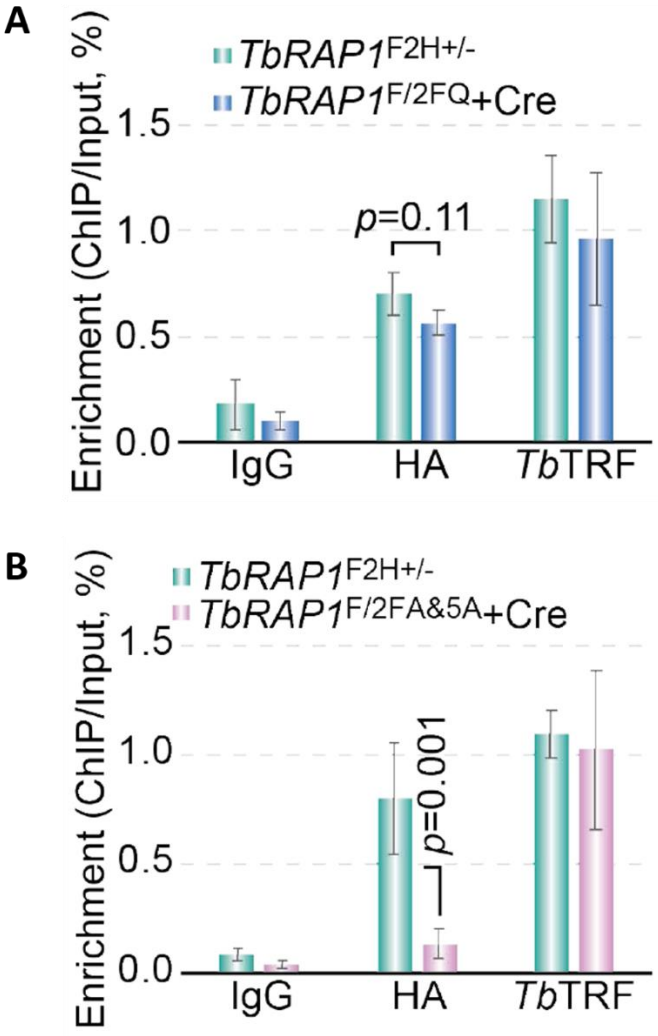


Figure 5.14 ChIP data display that *Tb*RAP1 DB region is important for *Tb*RAP1 telomere chromatin localization. (A) ChIP experiments results of Cre-induced (for 30

hours) *Tb*RAP1$^{F/2FQ}$. (B) ChIP experiments results of Cre-induced (for 30 hours) *Tb*RAP1$^{F/2FA\&5A}$. ChIP was conducted from *Tb*RAP1$^{F2H+/-}$ cells as a positive control. *p* values are indicated.

ChIP experiments were conducted in *Tb*RAP1$^{F/2FQ}$ and *Tb*RAP1$^{F/2FA\&5A}$ strains after treatment with Cre for 30 hours. The HA antibody 12CA5 and a *Tb*TRF rabbit antibody were used in both *Tb*RAP1$^{F/2FQ}$ and *Tb*RAP1$^{F/2FA\&5A}$ samples. Results showed that F2H-*Tb*RAP1-2FQ still associated with the telomere chromatin when analyzed, as a comparable amount of telomere chromatin was coprecipitated (Figure 5.14 A). In contrast, F2H-NLS-*Tb*RAP1-2FA&5A did not associate with the telomere chromatin, even though it was expressed at the same level as F2H-*Tb*RAP1 (Figure 5.14 B). It is because that this mutant lacks the R/K patch which is required for telomere localization. *Tb*TRF, telomeric dsDNA binding protein, as a positive control, was still associated with telomere chromatin in these mutants. These data suggest that DB region is important for *Tb*RAP1's telomere localization.

To further confirm the roles of DB region in *VSG* mRNA association, IF analyses were performed to examine the *Tb*RAP1 subnuclear localization in both *Tb*RAP1$^{F2H+/-}$ and *Tb*RAP1$^{F/2FQ}$ strains. DAPI was used to stain nucleus. In *Tb*RAP1$^{F2H+/-}$ cells, F2H-*Tb*RAP1 (red) partially colocalized with telomeric protein *Tb*TRF (green) in nucleus. Importantly, F2H-*Tb*RAP1-2FQ (green) remained partial colocalization with *Tb*TRF (red) (Figure 5.15), indicating F2H-*Tb*RAP1-2FQ still localized at telomere.

Consequently, IF analyses further showed that F2H-*Tb*RAP1-2FQ containing DB region, can localize at telomere like *Tb*RAP1 WT.

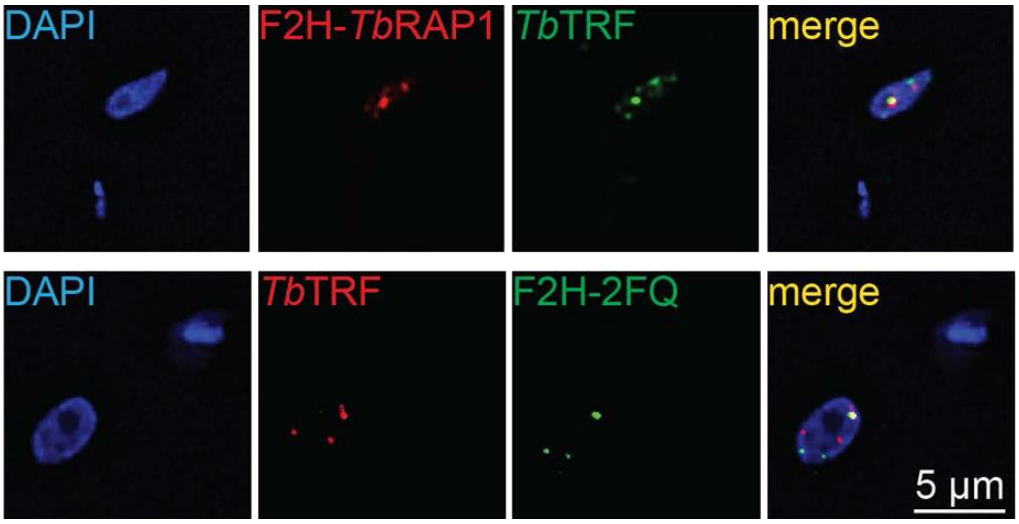


Figure 5.15 IF analyses results demonstrate that *Tb*RAP1 DB region is important for *Tb*RAP1 telomere chromatin localization. IF analyses were carried out in Cre-induced *Tb*RAP1$^{F/2FQ}$ cells. DAPI was used to stain nucleus. 12CA5 antibody was used to detect *Tb*RAP1 and a *Tb*TRF chicken antibody was used to detect *Tb*TRF. The size bars of all images are labeled in each panel.

In summary, we have confirmed that two F residues within RRM domain are indispensable for *Tb*RAP1's association with active *VSG* mRNA. While *Tb*RAP1's RNA binding activity is also dependent on the R/K-riched patch *in vivo*. Possibly because *Tb*RAP1 needs to be recruited to the telomere vicinity before it contacts with the active *VSG* mRNA.

5.5 RNA binding activity of *Tb*RAP1 is critical for normal cell proliferation

To determine the function of *Tb*RAP1's RNA binding activity, we examined various phenotypes in *Tb*RAP1$^{F/mut}$ cells that specifically disrupt RNA binding activity, including *Tb*RAP1$^{F/\Delta RRM}$ cells without RRM module, *Tb*RAP1$^{F/2FQ}$ cells carrying two conserved F residues mutation, and *Tb*RAP1$^{F/2FA\&5A}$ cells with RRM and DB region mutations, respectively.

Firstly, western analysis was used to investigate the continued expression of each F2H-*Tb*RAP1 mutants with RNA binding activity deficiency. These three *Tb*RAP1 $^{F/mut}$ cells before and after induction of Cre for 30 hours were harvested and analyzed by western blot. Specifically, *Tb*RAP1 mutants were detected by the HA monoclonal antibody and the endogenous *Tb*RAP1 were detected *Tb*RAP1 rabbit antibody. Tubulin expression level was detected as a loading control.

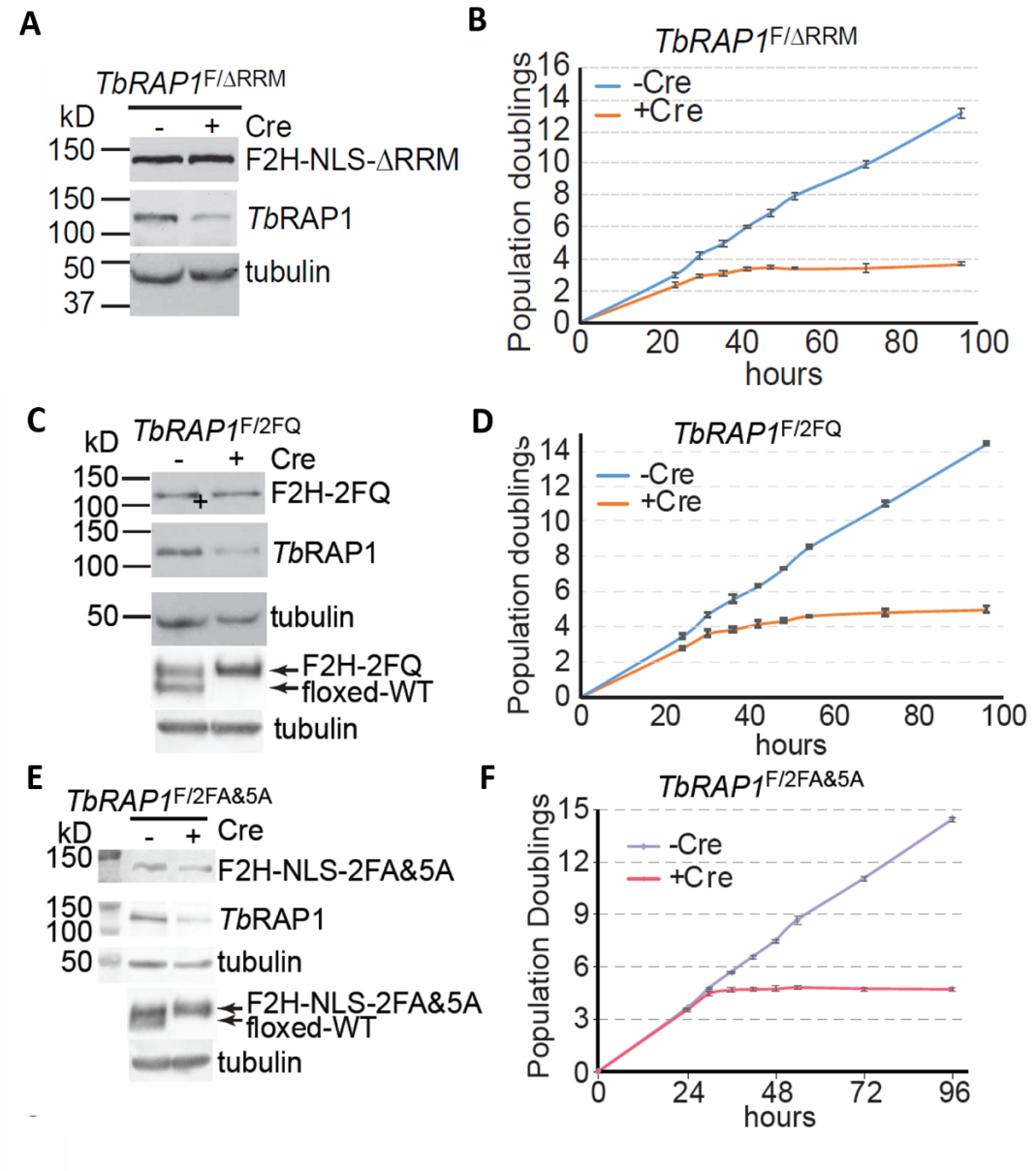


Figure 5.16 *Tb*RAP1's RNA binding activity is required for normal cell growth. (A) Western analyses profile of *Tb*RAP1 $^{F/\Delta RRM}$ cells. (B) Cell growth curves with or without Cre treatment in *Tb*RAP1 $^{F/\Delta RRM}$ cells. (C) Western analyses profile of *Tb*RAP1$^{F/2FQ}$ cells. (D) Cell growth curves with or without Cre treatment in *Tb*RAP1$^{F/2FQ}$ cells. (E) Western analyses profile of *Tb*RAP1$^{F/2FA\&5A}$ cells. (F) Cell growth curves with or without Cre treatment in *Tb*RAP1$^{F/2FA\&5A}$ cells.

Afterward, we calculated cell population for 96 hours under the condition with and without Cre treatment separately. Three independent experiments were performed to obtain average values. Subsequently, cell growth curves in each *Tb*RAP1$^{F/mut}$ were obtained, respectively.

In *Tb*RAP1$^{F/\Delta RRM}$ cells, western blot results showed that after treatment with Cre for 30 hours, F2H-NLS-tagged *Tb*RAP1-ΔRRM was normally expressed (Figure 5.16 A). Deletion of the floxed *Tb*RAP1 allele led to a growth arrest (Figure 5.16 B), indicating that the RRM domain is critical for *Tb*RAP1's functions in *VSG* regulation.

After inducing by Cre for 30 hours, floxed-*Tb*RAP1 was disappeared while F2H-NLS-tagged *Tb*RAP1-2FQ and 2FA&5A were normally expressed in *Tb*RAP1$^{F/2FQ}$ (Figure 5.16 C) and *Tb*RAP1$^{F/2FA\&5A}$ (Figure 5.16 E) cells, respectively. Like *Tb*RAP1$^{F/\Delta RRM}$ cells, we also observed cell proliferation inhibition in Cre-induced *Tb*RAP1$^{F/2FQ}$ (Figure 5.16 D) and *Tb*RAP1$^{F/2FA\&5A}$ (Figure 5.16 F) cells that lacking the RNA binding activity. Cell population doublings showed no accumulation after Cre treatment, indicating cell growth was arrested in both Cre-induced *TbRAP1*$^{F/FtoQ}$ and *TbRAP1*$^{F/2FA\&5A}$ cells. These results revealed that the RNA binding activity of *Tb*RAP1 is critical for normal cell proliferation.

5.6 *Tb*RAP1's RNA binding activity regulates genes expression

5.6.1 *Tb*RAP1's RNA binding activity suppresses *VSG* gene expression

To test the function of *Tb*RAP1's RNA binding activity in *VSG* gene expression, we detected mRNA levels of several ES-linked *VSG*s in *Tb*RAP1$^{F/mut}$ strains which lost

*Tb*RAP1's RNA binding activity, including *TbRAP1*$^{F/\Delta RRM}$ and *TbRAP1*$^{F/2FQ}$ cells. The qRT-PCR technique was applied to measure *VSG* mRNA level. With induction of Cre for 30 hours, multiple ES-linked silent *VSG*s, such as *VSG*8, *VSG*3, *VSG*6, *VSG*9, and *VSG*21, were derepressed several hundred-folds to thousand-folds in both *TbRAP1*$^{F/\Delta RRM}$ (Figure 5.17 A) and *TbRAP1*$^{F/2FQ}$ (Figure 5.17 B) cells. Based on these data, we conclude that the two conserved F residues within RRM module, which is required for *Tb*RAP1's RNA binding, play an important role in suppressing *VSG* gene switching.

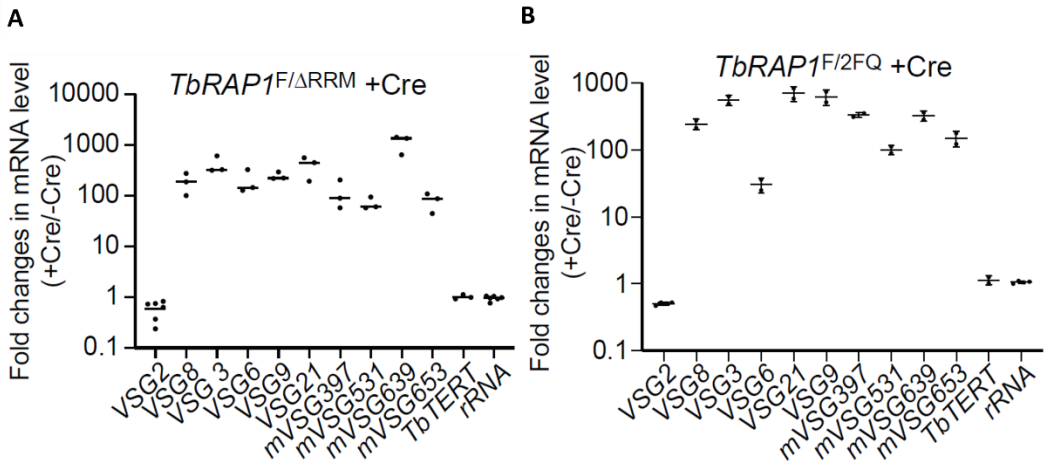


Figure 5.17 qRT-PCR results illustrate that *Tb*RAP1's RNA binding activity suppresses *VSG* gene expression. Varied *VSG* mRNA levels detected by qRT-PCR in Cre-induced *Tb*RAP1$^{F/\Delta RRM}$ cells (A) and *Tb*RAP1$^{F/\Delta RRM}$ cells (B), respectively.

5.6.2 *Tb*RAP1's RNA binding activity contributes to active *VSG* mRNA transcription

As described in chapter 3, *Tb*RAP1's DNA binding activity also functions on suppressing silent *VSG* gene expression. Is there any difference between the impacts of

*Tb*RAP1's DNA binding activity and that of RNA binding activity on active VSG expression? To test this hypothesis, we examined the mRNA level of active *VSG*2 and several ES-linked *VSG*s at different time points in Cre-induced *Tb*RAP1$^{F/2FQ}$ and *Tb*RAP1$^{F/5A}$ cells, respectively. After induced by Cre for 6, 12, 18, 24, and 30 hours, mRNA levels of active *VSG*2 and varied silent *VSG*s were detected by qRT-PCR.

Interestingly, we found that after induction of Cre, silent *VSG*s mRNA levels increased several hundred-fold in *TbRAP1*$^{F/2FQ}$ cells that lacking RNA binding activity. While the active *VSG2* mRNA level dropped to 50-60% of its original level (Figure 5.18 A). It indicates that, without RNA binding activity, active *VSG* failed to be fully transcribed.

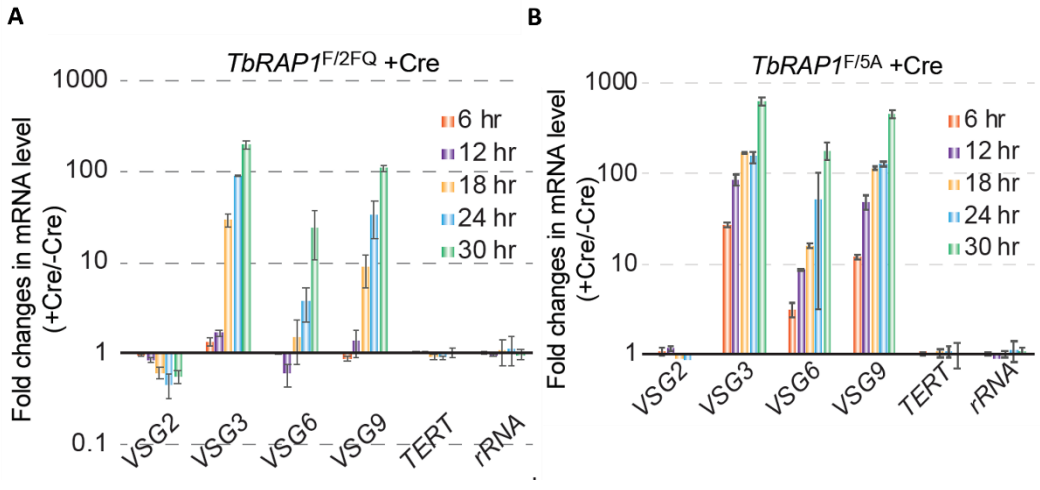


Figure 5.18 qRT-PCR results illustrate that *Tb*RAP1's RNA binding activity functions on active *VSG* mRNA transcription. Histogram of varied *VSG* mRNA level at different time points detected by qRT-PCR in Cre-induced *Tb*RAP1$^{F/2FQ}$ cells (A), and *Tb*RAP1$^{F/5A}$ cells (B), respectively.

On the contrary, after induction of Cre, the active *VSG* mRNA level did not decrease in *TbRAP1*$^{F/5A}$ cells that lost DNA binding activity, even though the mRNA levels of silent *VSGs* also increased several hundred-fold (Figure 5.18 B). These observations suggest that the RNA binding activity of *Tb*RAP1 may be important for keeping the active *VSG* fully transcribed.

5.6.3 *Tb*RAP1's RNA binding activity suppresses global genes expression

Since RAP1 homologs can regulate the expression of both subtelomeric and nontelomeric genes. To further investigate the function of *Tb*RAP1's RNA binding activity in global gene expression, we performed RNA sequencing analysis in *Tb*RAP1$^{F/mut}$ cells that lost RNA binding activity.

After 30 hours of Cre induction, ~5,000 genes were up-regulated in *TbRAP1*$^{F/2FQ}$ cells. As described in chapter 3, more than 7,000 genes were up-regulated in Cre-induced *TbRAP1*$^{F/-}$ cells with *Tb*RAP1 deletion. Fewer genes were up-regulated in the *TbRAP1*$^{F/2FQ}$ cells than the *TbRAP1*$^{F/-}$ cells, indicating 2FQ mutant appears to have a little milder derepression phenotype (Figure 5.19 A). Similarly, ~5,000 genes were up-regulated in the Cre-induced *TbRAP1*$^{F/2FA\&5A}$ cells (Figure 5.19 B). These data suggest that RNA binding activity is required for proper gene expression regulating by *Tb*RAP1. While the loss of the *VSG* mRNA and non-specific DNA binding activities does not have additive effects on gene derepression.

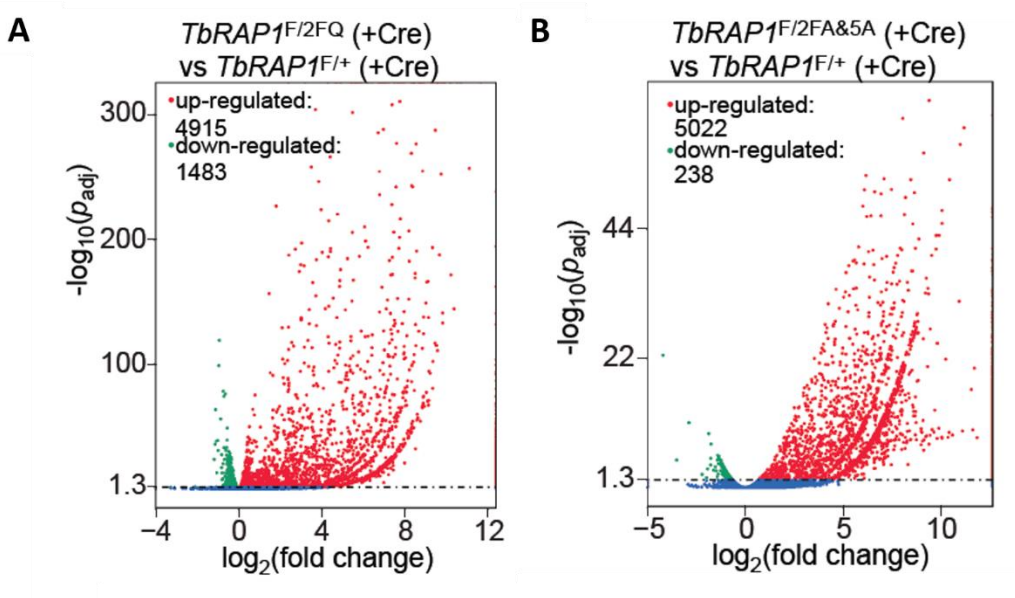


Figure 5.19 *Tb*RAP1's RNA binding activity is critical for suppressing global genes expression. A volcano plot of genes up-regulated and down-regulated in *Tb*RAP1^{F2FQ} **(A)** and *Tb*RAP1$^{F/2FA\&5A}$ **(B)** cells compared to *Tb*RAP1$^{F/+}$ cells 30 hours after Cre induction.

5.7 *Tb*RAP1's RNA binding activity maintains telomere/subtelomere integrity

We have noticed that *Tb*RAP1 suppresses *VSG* switching by maintaining telomere/subtelomere integrity. Telomeric/subtelomeric DNA damage, particularly that in the active ES, always leads to frequent *VSG* switching. As *Tb*RAP1 mutants that lose their RNA binding activity derepressed *VSG* switching, we hypothesis *Tb*RAP1's RNA binding activity may play a critical role in maintaining telomere/subtelomere integrity.

To confirm our hypothesis, we examined DNA damages in *Tb*RAP1 mutants that loss RNA binding activity, including *Tb*RAP1$^{F/2FQ}$ and *Tb*RAP1$^{F/2FA\&5A}$, respectively.

The level of γH2A, an indicator of the amount of DNA damages, was examined by western blots in *Tb*RAP1^{F2FQ} and *Tb*RAP1$^{F/2FA\&5A}$ cells.

Whole-cell lysates from WT cells before and after phleomycin treatment were prepared and analyzed by western blotting using a γH2A rabbit antibody as a positive control. At the same time, γH2A protein levels in *TbRAP1*$^{F/2FQ}$ and *TbRAP1*$^{F/2FA\&5A}$ cells before and after Cre induction were measured by western blot. While the tubulin detected by antibody TAT-1 were used as a loading control.

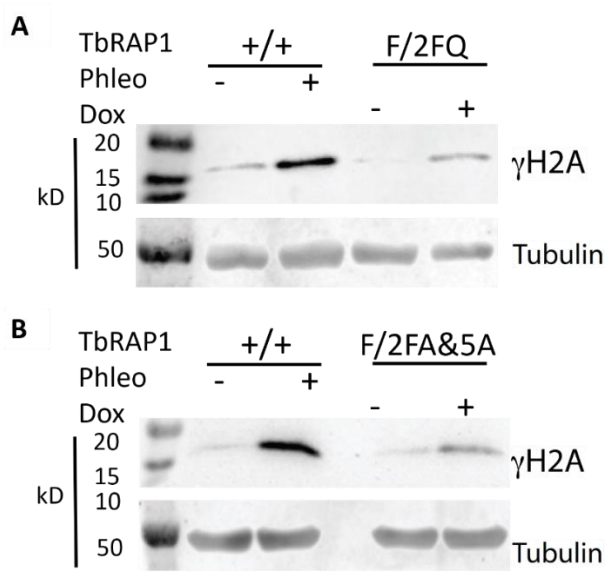


Figure 5.20 Western blotting results display that *Tb*RAP1's RNA binding activity protects telomere from DNA damage. Whole-cell lysates from WT cells before and after phleomycin treatment (as a positive control) and from *Tb*RAP1$^{F/2FQ}$ (A) and *Tb*RAP1$^{F/2FA\&5A}$ (B) cells before and after Cre induction were analyzed by western blotting using a γH2A rabbit antibody.

Western blotting showed that the γH2A level increased in WT cells after phleomycin treatment, a chemical that caused DNA damages. Similarly, an increased

level of γH2A was observed in Cre-induced *TbRAP1*$^{F/2FQ}$ (Figure 5.20 A) and *TbRAP1*$^{F/2FA\&5A}$ (Figure 5.20 B) cells, indicating that these cells had more DNA damages. In summary, we conclude that *Tb*RAP1's RNA binding activity protects telomere from DNA damages.

5.8 The binding of *Tb*RAP1 on RNA competes away its binding on DNA

Guided by previous data, we have noticed that *Tb*RAP1 has both DNA and RNA binding activities. The non-specific DNA binding activities depend on the $_{737}$RKRRR$_{741}$ patch in the *Tb*RAP1 DB domain, while the *VSG* mRNA binding activity relies on the RRM domain where F655 and F694 are essential. As the domains required for RNA binding and DNA binding are different, it is possible that *Tb*RAP1's binding activities of RNA and that of DNA are independent. To test this, we performed EMSA competition assays.

In the beginning, an increasing amount of non-radiolabeled 170nt *VSG*2 3'UTR RNA was used to compete with radiolabeled 100 bp duplex DNA containing (TTAGGG)$_{12}$ for *Tb*RAP1-ML binding. Surprisingly, the non-radiolabeled *VSG*2 3'UTR RNA fully competed *Tb*RAP1-ML away from radiolabeled dsDNA (TTAGGG)$_{12}$, when RNA: DNA molar ratio reached at 1.5 (Figure 5.21 A). Similarly, a non-radiolabeled RNA containing TERRA with (UUAGGG)$_{12}$ also competed away the *Tb*RAP1-ML binding on the duplex TTAGGG repeat containing DNA (Figure 5.21 B).

In contrast, the mixture of radiolabeled *VSG*2 3'UTR RNA and *Tb*RAP1-ML was used as the substrate, 170 times of 100 bp telomeric dsDNA could not outcompete *VSG*2 3'UTR RNA for binding with *Tb*RAP1-ML (Figure 5.21 C). Similarly, 170 bp dsDNA containing the *VSG*2 3'UTR sequence failed to outcompete with *Tb*RAP1-ML mixtures containing *VSG*2 3'UTR RNA, even when the DNA:RNA molar ratio reached 20 (Figure 5.21 D). These data suggest that RNAs are able to compete away *Tb*RAP1's binding on DNAs. Therefore, in the presence of the large amount of active *VSG*2 mRNA, *Tb*RAP1 is unlikely to bind the telomere DNA.

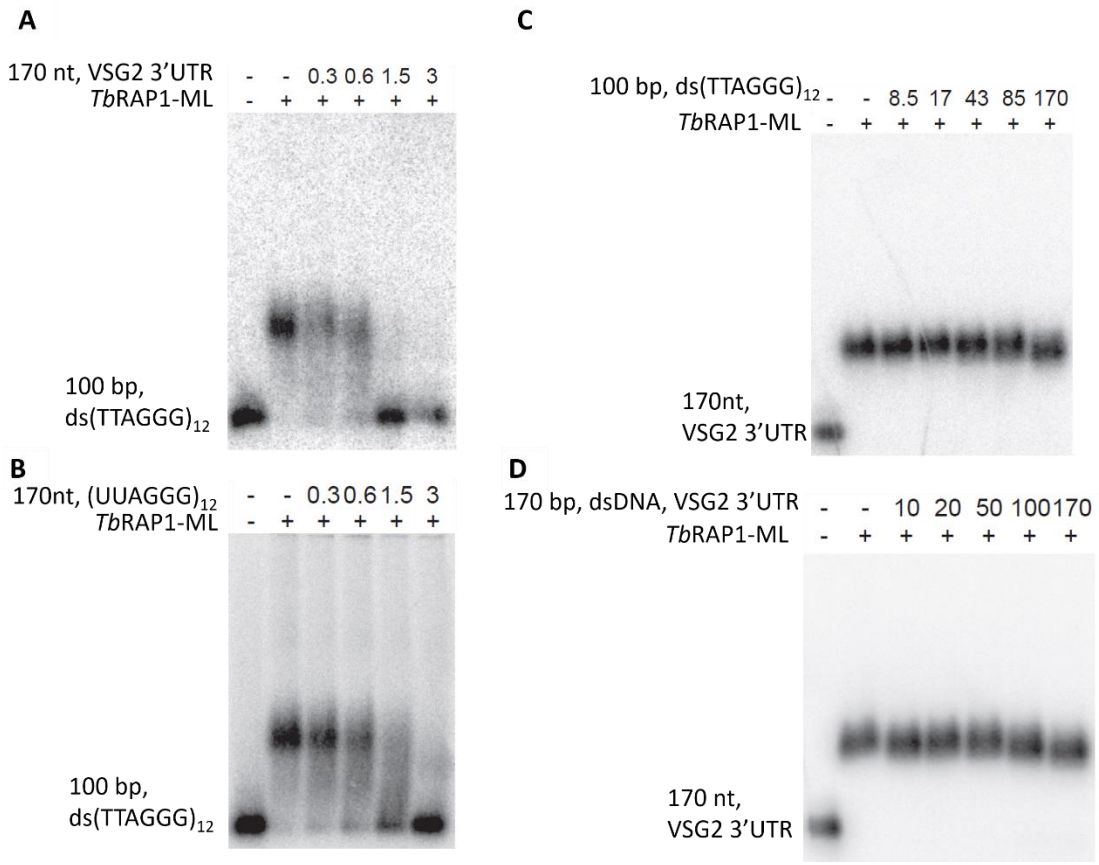


Figure 5.21 EMSA competitive experiments suggest that RNAs are able to compete away *Tb*RAP1's binding on DNAs. EMSA competition profile that non-radiolabeled 170 nt *VSG*2 3'UTR (A) and 170 nt $(UUAGGG)_{12}$ (B) compete *Tb*RAP1-ML away

from radiolabeled 100 bp ds(TTAGGG)$_{12}$, respectively. EMSA competition profile that non-radiolabeled 100 bp ds(TTAGGG)$_{12}$ (C) and 170 bp dsDNA with *VSG*2 3'UTR sequence (D) compete *Tb*RAP1-ML away from radiolabeled 170 nt *VSG*2 3'UTR, respectively. The molar ratios between competitors and substrates are indicated at the top line.

6 Conclusions

Trypanosoma brucei is a protozoan parasite that causes African sleeping sickness in humans and Nagana in cattle. The biggest challenge for this infectious disease treatment is that *T. brucei* can evade the host immune system response through regularly switching its major surface antigen *VSG*. Although the human innate immune system decreases the bloodstream form of parasites initially, the left pathogens with a new *VSG* continue to avoid attacks from the host immune system. More than 2,500 complete or truncated *VSG* genes are exclusively expressed from expression sites located at subtelomeric region which closes to telomere, in a monoallelic expression manner.

Telomeric proteins belonging to the Shelterin complex play important roles in *VSG* expression and switching. Specifically, our collaborator has identified *Tb*RAP1 as a component of the telomere Shelterin complex. Subsequent research works have suggested that *Tb*RAP1 is essential for *VSG* silencing, suppresses the telomeric transcription and telomeric R-loop levels, and suppresses DNA recombination-mediated *VSG* switching events.

However, the mechanism of *Tb*RAP1-mediated *VSG* silencing is still poorly understood. Here we have conducted a series of biochemical and structural studies of *Tb*RAP1 to elucidate the molecular mechanism of *Tb*RAP1 functions on silencing and *VSG* suppressing.

6.1 *Tb*RAP1 Myb-like region mediated DNA binding activity investigation and its role in *VSG* regulation

Firstly, we identified dsDNA and ssDNA binding activity of *Tb*RAP1 MybLike domain (aa 639-761) by EMSA assay and HSQC titration experiments *in vitro*. *Tb*RAP1's dsDNA and ssDNA binding activities are electrostatics-based and sequence nonspecific. *Tb*RAP1 mutant with $_{737}$RKRRR$_{741}$ patch to $_{737}$AAAAA$_{741}$ fails to interact with DNA anymore, indicating that these positively charged $_{737}$RKRRR$_{741}$ residues are required for *Tb*RAP1's DNA binding activity.

ChIP and IF analysis results illustrate that R/K patch mediated-*Tb*RAP1's DNA binding activity is necessary for *Tb*RAP1 telomere localization. Cell proliferation arrest in *Tb*RAP1 mutants that lost DNA binding activity, indicating DNA binding activity is critical for cell viability. Furthermore, this electrostatics-based DNA binding activity is also essential for *Tb*RAP1's *VSG* silencing and subtelomere/ telomere integrity.

6.2 Biochemical and functional characterization of potential RNA-binding activity of *Tb*RAP1 and its role in *VSG* regulation

We have determined that the NMR structure of *Tb*RAP1 MybLike domain (aa 639-761) contains an RRM module (aa 639-733) with the characteristic topology of a four-stranded anti-parallel β-sheet plus two α-helices, followed by a long and flexible loop (aa 734-761). Structural superimposition and sequence assignment results between *Tb*RAP1 RRM and other classic RRMs indicate that two conserved aromatic

residues (F655 and F694) of *Tb*RAP1 RRM may form stacking interaction with ssDNA or RNA substrates.

Furthermore, we have confirmed that *Tb*RAP1-ML interacts with RNA binding activity through EMSA, FP, and NMR titration experiments, respectively. More interestingly, two structural regions of *Tb*RAP1-ML contribute to this interaction in different mechanisms. The RRM domain binds preferably to *VSG* mRNA with the conserved F655 and F694 residues at the canonical RNA binding site playing critical roles. While the DB segment shows electrostatics-based non-specific RNA binding mediated by its R/K-rich patch. Most importantly, *Tb*RAP1 without RNA binding activity leads to *VSG* suppression and cell growth arrest, indicating that *Tb*RAP1's RNA binding activity is essential for *VSG* monoallelic expression and cell viability.

Different from DNA binding activity, *Tb*RAP1's RNA binding activity also contributes to active *VSG* mRNA fully transcript. Furthermore, EMSA competition assays suggest that RNAs can compete away *Tb*RAP1's binding on DNAs. Based on these findings, we suggest that *Tb*RAP1 executes its "repressor" role at silent ES while serves as an "activator" at active ES. Specifically, at the silent ES, *Tb*RAP1 localizes to telomeric DNA to repress silent *VSG* expression and maintains telomere/subtelomere stability (Figure 6.1 A). While at active ES, *Tb*RAP1 prefers to interact with *VSG* mRNA, rather than telomeric DNA, because the high level of active *VSG* mRNA is transcribed by RNA Pol I at this location (Figure 6.1 B). Therefore, *Tb*RAP1 remove from telomeric DNA and helps *VSG* expression at active ES.

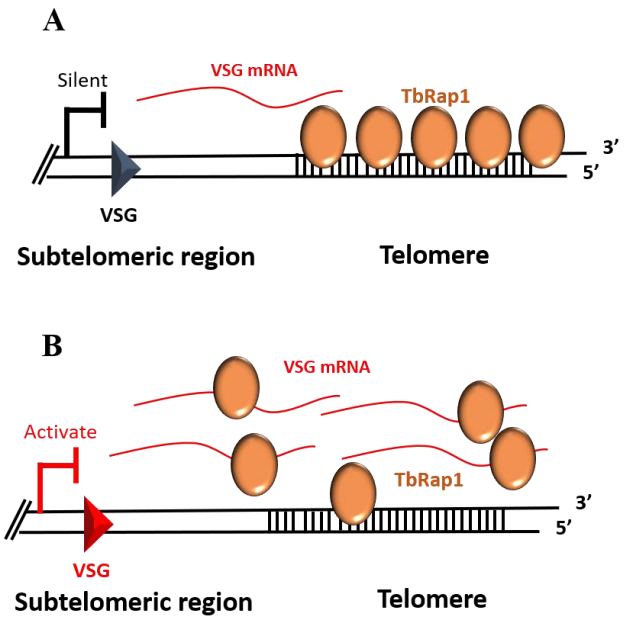


Figure 6.1 Proposed *Tb*RAP1-*VSG* mRNA binding model. (A) At the silent ES, *Tb*RAP1 localizes to telomeric DNA. (B) At active ES, *Tb*RAP1 prefers to interact with *VSG* mRNA.

6.3 Future studies

For DNA binding activity, we have confirmed that *Tb*RAP1's non-specific DNA binding activity is important for its telomere localization and *VSG* suppression. However, it is still unclear how *Tb*RAP1 targeting sequence-specific telomere through this sequence non-specific DNA binding activity.

We suppose that there may be additional partners work together with *Tb*RAP1 to execute its *VSG* suppression role. For example, *Tb*TRF, another telomeric protein in *T. brucei*, has been proved to interact with *Tb*RAP1 *in vivo*. In the future, we will investigate *Tb*RAP1's interaction with its potential binding partners to further illustrate this non-specific DNA binding activity.

For RNA binding activity, we have illustrated that *Tb*RAP1 contains an RRM domain with novel RNA binding activity that is critical for *VSG* silencing. However, it is still unclear about the sequence preference in *VSG* mRNA-mediated *Tb*RAP1's RNA binding activity. How does *Tb*RAP1 RRM exactly recognize *VSG* mRNA, such a large molecule? Whether *Tb*RAP1 will associate with RNAs that transcript from upstream ESs closed to *VSG* gene? To further investigate characteristics of *Tb*RAP1's *VSG* mRNA binding activity, varied lengths and different sequences of RNA oligos within *VSG* mRNA will be designed and analyzed.

In summary, my thesis work has provided biochemical and structural information to help understand the mechanism of *Tb*RAP1-mediated telomeric silencing and *VSG* silencing. The DNA-binding activity of *Tb*RAP1, caused by positively charged R/K patch, functions on its telomere localization. It is the first time that proves *Tb*RAP1 has nucleic acid binding activity. While *Tb*RAP1's *VSG* mRNA-binding activity, mediated by two F residues within RRM, is critical for active *VSG* fully transcript. This RNA-binding activity is completely novel among all known RAP1 homologs, indicating a more drastic evolution of telomere proteins than we have originally understood.

References

Marjia Afrin*, et al.*, *Trypanosoma brucei RAP1 Has Essential Functional Domains That Are Required for Different Protein Interactions.* Msphere, 2020. **5**(1).

Nancy S Bae, and Peter Baumann, *A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends.* Molecular cell, 2007. **26**(3): p. 323-334.

Matthew Berriman*, et al.*, *The genome of the African trypanosome Trypanosoma brucei.* science, 2005. **309**(5733): p. 416-422.

Elizabeth H Blackburn, Carol W Greider, and Jack W Szostak, *Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging.* Nature medicine, 2006. **12**(10): p. 1133-1138.

Yong Chen*, et al.*, *A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins.* Science, 2008. **319**(5866): p. 1092-1096.

G. A. Cross, L. E. Wirtz, and M. Navarro, *Regulation of VSG expression site transcription and switching in Trypanosoma brucei.* Mol Biochem Parasitol, 1998. **91**(1): p. 77-91.

George AM Cross, Hee-Sook Kim, and Bill Wickstead, *Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427.* Molecular and biochemical parasitology, 2014. **195**(1): p. 59-73.

Gerrit M Daubner, Antoine Cléry, and Frédéric HT Allain, *RRM–RNA recognition: NMR or crystallography⋯ and new findings.* Current opinion in structural biology, 2013. **23**(1): p. 100-108.

Titia De Lange, *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes & development, 2005. **19**(18): p. 2100-2110.

Marco De Vitis, Francesco Berardinelli, and Antonella Sgura, *Telomere length maintenance in cancer: at the crossroad between telomerase and alternative lengthening of telomeres (ALT).* International journal of molecular sciences, 2018. **19**(2): p. 606.

Eros Lazzerini Denchi, and Titia de Lange, *Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1.* Nature, 2007. **448**(7157): p. 1068-1071.

Ylli Doksani*, et al.*, *Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation.* Cell, 2013. **155**(2): p. 345-356.

Oliver Dreesen, Bibo Li, and George AM Cross, *Telomere structure and function in trypanosomes: a proposal.* Nature Reviews Microbiology, 2007. **5**(1): p. 70-75.

Louise Fairall*, et al.*, *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Molecular cell, 2001. **8**(2): p. 351-361.

Lucy Glover, Sam Alsford, and David Horn, *DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes.* PLoS Pathog, 2013. **9**(3): p. e1003260.

Lucy Glover, and David Horn, *Trypanosomal histone γH2A and the DNA damage response.* Molecular and biochemical parasitology, 2012. **183**(1): p. 78-83.

Jack D Griffith*, et al.*, *Mammalian telomeres end in a large duplex loop.* Cell, 1999. **97**(4): p. 503-514.

Shingo Hanaoka*, et al.*, *NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains.* Journal of molecular biology, 2001. **312**(1): p. 167-175.

Christiane Hertz-Fowler*, et al.*, *Telomeric expression sites are highly conserved in Trypanosoma brucei.* PloS one, 2008. **3**(10): p. e3527.
Dirk Hockemeyer*, et al.*, *POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end.* The EMBO journal, 2005. **24**(14): p. 2667-2678.
D. Horn, and G. A. Cross, *Analysis of Trypanosoma brucei VSG expression site switching in vitro.* Mol Biochem Parasitol, 1997. **84**(2): p. 189-201.
David Horn, and J David Barry, *The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes.* Chromosome research, 2005. **13**(5): p. 525-533.
Fuyuki Ishikawa, *Portrait of replication stress viewed from telomeres.* Cancer science, 2013. **104**(7): p. 790-794.
Mohammad A Jafri*, et al.*, *Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies.* Genome medicine, 2016. **8**(1): p. 69.
Sanaa E Jehi*, et al.*, *Suppression of subtelomeric VSG switching by Trypanosoma brucei TRF requires its TTAGGG repeat-binding activity.* Nucleic acids research, 2014. **42**(20): p. 12899-12911.
Sanaa E Jehi, Fan Wu, and Bibo Li, *Trypanosoma brucei TIF2 suppresses VSG switching by maintaining subtelomere integrity.* Cell research, 2014. **24**(7): p. 870-885.
Junko Kanoh, and Fuyuki Ishikawa, *spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast.* Current Biology, 2001. **11**(20): p. 1624-1630.
Jan Karlseder, Agata Smogorzewska, and Titia de Lange, *Senescence induced by altered telomere state, not telomere loss.* Science, 2002. **295**(5564): p. 2446-2449.
Sahn-ho Kim, Patrick Kaminker, and Judith Campisi, *TIN2, a new regulator of telomere length in human cells.* Nature genetics, 1999. **23**(4): p. 405-412.
Ming Lei, Elaine R Podell, and Thomas R Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection.* Nature structural & molecular biology, 2004. **11**(12): p. 1223-1229.
Charles Chung Yun Leung, and JN Mark Glover, *BRCT domains: easy as one, two, three.* Cell cycle, 2011. **10**(15): p. 2461-2470.
B. Li, *DNA Double-Strand Breaks and Telomeres Play Important Roles in Trypanosoma brucei Antigenic Variation.* Eukaryot Cell, 2015. **14**(3): p. 196-205.
Bibo Li, *DNA double-strand breaks and telomeres play important roles in Trypanosoma brucei antigenic variation.* Eukaryotic cell, 2015. **14**(3): p. 196-205.
Bibo Li, Amin Espinal, and George AM Cross, *Trypanosome telomeres are protected by a homologue of mammalian TRF2.* Molecular and cellular biology, 2005. **25**(12): p. 5011-5021.
Bibo Li, Stephanie Oestreich, and Titia De Lange, *Identification of human Rap1: implications for telomere evolution.* Cell, 2000. **101**(5): p. 471-483.
Weiqiang Lin*, et al.*, *Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity.* The EMBO journal, 2013. **32**(10): p. 1425-1439.
D Malvy, and François Chappuis, *Sleeping sickness.* Clinical Microbiology and Infection, 2011. **17**(7): p. 986-995.
Lucio Marcello, and J David Barry, *Analysis of the VSG gene silent archive in Trypanosoma brucei reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure.* Genome research, 2007. **17**(9): p. 1344-1352.
Richard McCulloch, Liam J Morrison, and James PJ Hall, *DNA recombination strategies during*

antigenic variation in the African trypanosome. Mobile DNA III, 2015: p. 409-435.
Nathan J Moerke, *Fluorescence polarization (FP) assays for monitoring peptide-protein or nucleic acid-protein binding.* Current protocols in chemical biology, 2009. **1**(1): p. 1-15.
Laura SM Müller*, et al.*, *Genome organization and DNA accessibility control antigenic variation in trypanosomes.* Nature, 2018. **563**(7729): p. 121-125.
Vishal Nanavaty*, et al.*, *Trypanosoma brucei RAP1 maintains telomere and subtelomere integrity by suppressing TERRA and telomeric RNA: DNA hybrids.* Nucleic acids research, 2017. **45**(10): p. 5785-5796.
Wilhelm Palm, and Titia de Lange, *How Shelterin protects mammalian telomeres.* Annual review of genetics, 2008. **42**: p. 301-334.
Gary N Parkinson, Michael PH Lee, and Stephen Neidle, *Crystal structure of parallel quadruplexes from human telomeric DNA.* Nature, 2002. **417**(6891): p. 876-880.
B Pina*, et al.*, *The different (sur) faces of Rap1p.* Molecular Genetics and Genomics, 2003. **268**(6): p. 791-798.
Rekha Rai*, et al.*, *TRF2-RAP1 is required to protect telomeres from engaging in homologous recombination-mediated deletions and fusions.* Nature communications, 2016. **7**(1): p. 1-13.
Sophie Redon*, et al.*, *Platinum cross-linking of adenines and guanines on the quadruplex structures of the AG 3 (T 2 AG 3) 3 and (T 2 AG 3) 4 human telomere sequences in Na+ and K+ solutions.* Nucleic acids research, 2003. **31**(6): p. 1605-1613.
Brice Rotureau, and Jan Van Den Abbeele, *Through the dark continent: African trypanosome development in the tsetse fly.* Frontiers in cellular and infection microbiology, 2013. **3**.
G. Rudenko, and L. H. Van der Ploeg, *Transcription of telomere repeats in protozoa.* EMBO J, 1989. **8**(9): p. 2633-8.
Christiane Schaffitzel*, et al.*, *In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei.* Proceedings of the National Academy of Sciences, 2001. **98**(15): p. 8572-8577.
Agnel Sfeir, and Titia De Lange, *Removal of Shelterin reveals the telomere end-protection problem.* Science, 2012. **336**(6081): p. 593-597.
Agnel Sfeir*, et al.*, *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication.* Cell, 2009. **138**(1): p. 90-103.
Jerry W Shay, and Woodring E Wright, *Hayflick, his limit, and cellular ageing.* Nature reviews Molecular cell biology, 2000. **1**(1): p. 72-76.
M Teresa Teixeira*, et al.*, *Telomere length homeostasis is achieved via a switch between telomerase-extendible and-nonextendible states.* Cell, 2004. **117**(3): p. 323-335.
Liza SM Wong*, et al.*, *Telomere biology in heart failure.* European journal of heart failure, 2008. **10**(11): p. 1049-1056.
Huawei Xin*, et al.*, *TPP1 is a homologue of ciliate TEBP-β and interacts with POT1 to recruit telomerase.* nature, 2007. **445**(7127): p. 559-562.
Yan Xu, *Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA.* Chemical Society Reviews, 2011. **40**(5): p. 2719-2740.
Xiaofeng Yang*, et al.*, *RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei.* Cell, 2009. **137**(1): p. 99-109.
J. C. Zomerdijk, R. Kieft, and P. Borst, *A ribosomal RNA gene promoter at the telomere of a mini-*

chromosome in Trypanosoma brucei. Nucleic Acids Res, 1992. **20**(11): p. 2725-34.