Structural Studies on RNA Helicase Associated with AU-Rich Element (RHAU) and DExD-box helicase 21 (DDX21)

by

Dhruba Ghosh

University of Manitoba

A thesis submitted to the Faculty of Graduate Studies

of the University of Manitoba

in partial fulfillment of the requirements of the degree of

Master of Science

Department of Chemistry

University of Manitoba

Winnipeg, Manitoba, Canada

ABSTRACT

G-quadruplexes (G4s) are physiologically significant tetra-stranded thermodynamically stable structures, characterized by stacked G- quartets, formed by guanine stretches on a single or multiple DNA/RNA strands, and implicated in various diseases. Helicase proteins capable of unfolding G4 structures include RNA helicase associated with AU-rich element (RHAU) and DExD-box helicase 21 (DDX21), that unwind both RNA and DNA G4 structures. In order to perform biophysical studies of the RHAU protein, full length RHAU ($RHAU_{FL}$) and a RHAU construct lacking its N-terminal 50 amino acid glycine-rich region ($RHAU_{51-1008}$) were tested in the optimization of the expression of the RHAU proteins in *E. coli*. In this thesis, we purified the $RHAU_{FL}$ protein using a comprehensive recombinant protein production approach involving gel electrophoresis, Western blot, cobalt affinity chromatography, size exclusion chromatography (SEC) and anion exchange chromatography. DDX21 binds G4 through its C-terminal domain. In order to have insight into its structure and G4 recognition, C-terminal DDX21 protein construct $DDX21_{C209}$ was expressed in *E. coli*, and purified using nickel affinity chromatography. Biophysical approaches including circular dichroism (CD) spectropolarimetry and isothermal titration calorimetry were used to check the quality of the $DDX21_{C209}$ protein, a thermal shift assay was used to investigate thermal stability of $DDX21_{C209}$, SEC and gel electrophoresis were performed following crystallization screening with and without a 22-nucleotide truncated human telomere DNA G4, hTel. In order to perform biophysical studies, full length DDX21 ($DDX21_{FL}$) protein was expressed in *E. coli* and its purification was attempted using nickel affinity chromatography. Crystallization of the G4 helicase proteins both in a free state and in complex with G4 structures would lead to a better understanding of G4 helicase structure and function.

ACKNOWLEDGEMENTS

I would like to thank my graduate programme supervisor, Dr. Jörg Stetefeld for his supervision, guidance and support. I would also like to extend my gratitude to the members of my advisory committee, Dr. Sean McKenna, Dr. Gregg Tomy and Dr. Francis Lin for their advice and guidance during this master's programme.

My humble appreciation goes to Dr. George Orriss who took his time to provide valuable research advice, insight into preparation of this thesis, and reviewed it with great care. I thank Dr. Aniel Moya Torres; without his continuous support, valuable advice and friendship, I would not have been at this platform. I thank Olga Francisco for her kind support in a foreign world. I appreciate research assistance of Dr. Markus Meier as I thank other members of our laboratory as well. I deeply appreciate advice and support of Dr. Evan Booy and Ewan McRae. I am eternally grateful to the former academic programme coordinator at the University of Manitoba, Mrs. Tuntun Sarkar for her selfless effort to direct me to the right path professionally and in life.

Finally, I would like to express my appreciation to Natural Sciences and Engineering Research Council of Canada (NSERC) that funded this research.

DEDICATION

I dedicate this thesis to my mother, Mallika Ghosh for she is the reason of the goodness in me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF COPYRIGHTED MATERIAL FOR WHICH PERMISSION WAS OBTAINED

Figure 1- Reproduced with permission from Nature Reviews Molecular Cell Biology.

https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=743ba255-6d17-4aba-8732-316e06e6ca51

Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).

Figure 1 appears on page 3.

Figure 2 - Reproduced with permission from Nucleic Acids Research.

https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=ac00648b-e400-405a-8da2-c139eed72510

Rhodes, D. & Lipps, H. J. Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).

Figure 2 appears on page 8.

Figure 3 - Reproduced with permission from The Journal of Biological Chemistry

http://www.jbc.org/site/misc/edpolicy.xhtml#copyright (Order License Id: 4547330146786)

Booy, E. P. *et al.* RNA helicase associated with AU-rich element (RHAU/DHX36) interacts with the 3′-tail of the long non-coding RNA BC200 (BCYRN1). *J. Biol. Chem.* **291**, 5355–5372 (2016).

Figure 3 appears on page 16.

Figure 4 - Reproduced with permission from Nature

https://s100.copyright.com/CustomerAdmin/PLF.jsp?ref=0742403d-82e6-4759-a390-d19f49f57efa

Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase

1. INTRODUCTION

1.1 G-QUADRUPLEXES

G-quadruplexes (G4s) are non-canonical nucleic acid secondary structures. These unusual structures are made up of four strands and have high thermodynamic stability. G4s can form in oligonucleotide sequences with consecutive stretches or tracts of multiple guanine (G) bases. Four guanine bases from separate tracts self-assemble into one planar G-quartet or -tetrad. The quartets are stabilized by hydrogen bonds (H-bonds) between the Watson-Crick and Hoogsteen faces of adjacent Gs. Multiple successive G-quartets stack upon one another by a spontaneous base stacking interaction. A monovalent or divalent cation, preferably potassium ion (K^+), positioned between the stacked quartets further stabilizes the G4 structure by co-ordinating with the O6 oxygen atoms of the guanines (Figure 1). As a result, G4 structures are highly thermostable, with reported melting temperature as high as 80º C [1–4]. Under physiological conditions monovalent K^+ and sodium ion (Na^+) confer the highest stability to G4 structures [5]. In contrast, smaller positive ions such as lithium ion (Li^+) destabilize G4 structures [6].

G4 quadruplexes can adapt different conformations based on several intrinsic and environmental conditions. Intrinsic factors comprise the nucleic acid sequence, the number of strands and strand orientation, glycosidic angles and the size of the loop connecting two adjacent G-rich strands. Environmental factors include the types of stabilizing cations, and while some cations stabilize G4 structures, heavy metal cations destabilize G-quadruplexes [1]. G4s can either be formed of either deoxyribonucleic acids (DNAs) or ribonucleic acids (RNAs). A number of high resolution crystal structures and NMR solution structures of structurally diverse DNA and RNA G-quadruplexes are now available [1,7]. For example the human telomeric d[AGGG(TTAGGG)$_3$] sequence is capable of forming DNA G4s as demonstrated by both NMR

and X-ray diffraction (XRD) studies [8]. The modified telomeric RNA sequence r(U_{Br}AGGGUUAGGGU) forms a G4 structure as reported by a crystal structure analysis which is further supported by the NMR structure for the unmodified r(UAGGGUUAGGGU) RNA sequence [9,10]. Multiple studies have shown the existence of G4-structures in the forms of both intermolecular and intramolecular quadruplexes [1] (Figure 1). Two or four G-rich strands can come together such that the quadruplex guanine bases are projected inside where they coordinate to the metal ion and fold into an intermolecular quadruplex. In contrast, unimolecular oligonucleotides form intramolecular quadruplexes due to the presence of G-tracts which are separated by a few non-guanine nucleotides [1]. Depending on the orientations of the quadruplex strands, G4-structures can be classified into the following types: parallel, antiparallel or mixed. Parallel G-quadruplexes are formed when the strands with the guanine stretches have the same polarity whereas in antiparallel G4 conformations the quadruplex strands run in opposite directions. In hybrid or mixed G4s three strands have the same polarity while the fourth strand runs in the opposite orientation [1]. The size of the loop between two G-tracts changes with the variation of G4 strand orientation. In case of unimolecular and bimolecular quadruplexes, the loop length ranges from a single to several hundred nucleotides. Parallel quadruplexes require longer loops to connect the top and the bottom quartets whereas antiparallel G4s can use much shorter loops to connect two strands on the same quartet [1]. G-tracts themselves are sometimes interrupted by non-G nucleotides which leads to the formation of a bulge (Figure 1); in consequence, the structural heterogeneity increases among separate G4s [1,11]. A 2017 Protein Database (PDB) search revealed that out of 231 PDB [12] entries only six G4 forming sequences were found to contain bulged nucleotides. An analysis of the backbone torsion angles of G4-structures identified a specific set of ζ-torsion angles caused by the

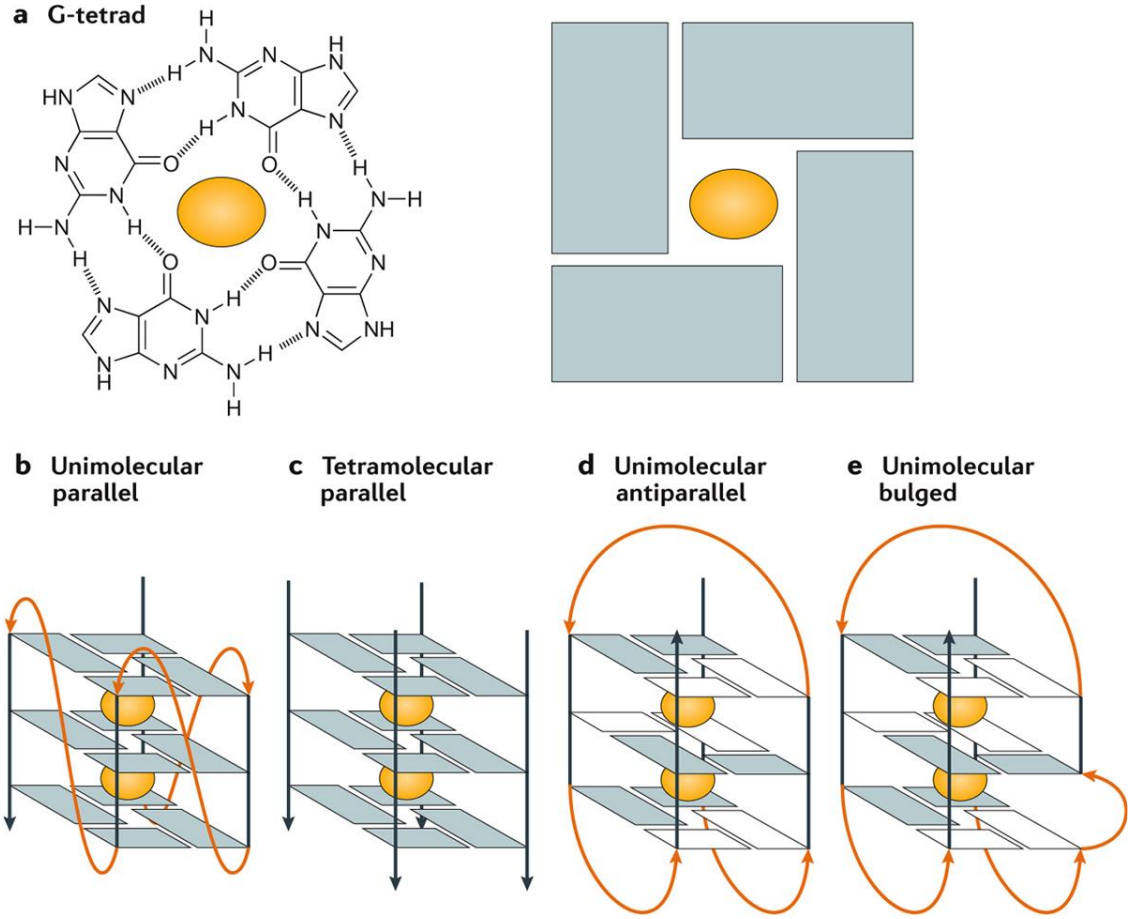


Figure 1. Schematic representation of G-quadruplex structures. (a) Left: G-quartet structure: G-quartets are the building blocks developed by the association of four guanine bases in a cyclic arrangement. Hoogsten hydrogen bonds (N1–N6 and N2–N7) primarily stabilize the quadruplex structure. (a) Right: Successive G-quartets stacked upon each other which build the quadruplex core. The core is further stabilized by coordination bonds between O_6 carbonyl groups and a monovalent alkali cation at the G4 center (orange) preferably K^+. Uni – and multi molecular G4s with different strand orientations: (b) unimolecular and parallel, (c) tetramolecular and parallel, (d) unimolecular and antiparallel and (e) unimolecular and antiparallel with a bulge caused by a G-tract discontinuity. Figure from "DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential" by Hertsch et al. [13].

loops and the bulges of G4s. Introduction of bulges causes an alteration of the DNA surface which results in a decrease in the stability of the G4 structure [14].

1.1.1 Stability of G-Quadruplexes

Regardless of variations in their conformation, G4-structures retain their high thermodynamic stability due to the stacking interactions of the G-quartets. A 2002 crystal structure [15] showed that a 22 nucleotide (d[AGGG(TTAGGG)$_3$]) DNA sequence from the human telomere could form a parallel intramolecular G4-quadruplex with a K^+ ion in the centre. In this structure, the guanine glycosidic bonds are arranged in the *anti*-conformation and π-π stacking interactions are effective between the three planar G-quartets. The bottom of one G4 strand is connected with the top of another strand with a propeller shaped loop without blocking the terminal G-quartets (Figure 1b). The central K^+ ions are located between the planar surfaces of the G-quartet such that they co-ordinate with the eight O_6-carbonyl groups of the same tetrad in a bi-pyramidal arrangement. Interestingly, when the K^+ is replaced by Na^+ the very same telomeric sequence folds mostly into an antiparallel conformation [15]. While the 2002 study revealed that quadruplexes mostly form antiparallel structures in the presence of excess Na^+, subsequent high resolution G4 structures have been reported where tetramolecular parallel quadruplexes can be formed in the presence of Na^+ ions [15–17]. In summary, G quadruplexes are mostly formed with Na^+ and K^+. Further studies revealed that Na^+ only has a strong stabilizing effect on DNA G4s, whereas K^+ significantly stabilizes both DNA and RNA G-quadruplexes [18]. Monovalent cations stabilize G4 structures from the highest to the lowest in the following order: $K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ [1]. Divalent cations can also impact G4 stability with the order as follows: $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$. Among the divalent cations, only Sr^{2+} has been shown to enhance RNA G4 stability, although Sr^{2+} ions have no known biological role. In the absence of other cations, Ca^{2+} and Mg^{2+}

ions do not impact G4 formation, whereas divalent heavy metal cations such as Pb^{2+}, Cd^{2+} and Hg^{2+} destabilize G4 structures. Cations stabilize G4s via two mechanisms, (a) cation-dipole interactions which are facilitated by the cation sitting between 2 G-tetrad layers such that it binds to 2x4 O6 carbonyl groups and (b) electrostatic interactions between the grooves and the phosphate backbone which is a common feature of most nucleic acid structures [19].

Due to variations in the stability among different G4 conformations, RNA G-quadruplex structures are less diverse than the DNA G4-structures. While DNA G-quadruplexes can adapt parallel, antiparallel and mixed conformations RNA G4s are usually limited to the parallel conformation. The 2'-hydroxyl (2'-OH) group of the ribose sugar has multiple impacts on G4 RNAs: (a) Intramolecular interactions are favored over intermolecular contacts for quadruplex stability. As a consequence, intramolecular structures of RNA G4s have enhanced stability compared to the DNA G4 structures. (b) 2'-OH attracts surrounding water (H_2O) molecules which leads to higher stability of the RNA G4s compared to the DNA counterparts. (c) Additional steric constraints are imposed on RNA G4 topology as 2'-OHs favor the anti-conformation over the syn-conformation of the ribose bases due to restraints on the glycosidic torsion angle and because ribose bases have a preference for C3'-endo sugar puckering which results in RNA G4s to preferentially adopt a parallel conformation [18].

1.1.2 *In Vivo* Distribution and Functions of G-Quadruplexes

Over three decades ago, Aaron Klug made a statement, "If G-quadruplexes form so readily *in vitro*, Nature will have found a way of using them *in vivo*." Since the 1960s it has been known that consecutive stretches of guanines show a tendency to self-assemble into G4 structures [20]. Initially, it was thought that these tetra-stranded secondary structures were aberrant due to the lack of any *in vivo* data. However, in the last few years the presence of both DNA and RNA G4-structures *in*

vivo has been reported. To determine the abundance of potential G-quadruplexes in the human genome *in silico* analyses were performed using the consensus sequence ($G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$). This analysis revealed that over 300 000 sequences are capable of forming G4-structures [21,22]. In order to exclude random G-rich sequences from the estimated number, a secondary analysis was performed where loop length between G- tracts was taken into account. The methodology to predict G4s requires further improvement for the complexity of the *in vivo* environment. However, it has been evident that a plethora of sequences are capable of forming G4 *in vivo,* and therefore G4s must be required to perform key biological roles. Therefore, performing additional experiments is crucial to obtain a more comprehensive understanding about the presence and functions of G4s in a biological context. Presently, it is established that G4s are found in varying abundance through different phases and states of cells, and are responsible in regulation of transcription, translation, mRNA splicing and therefore, G-quadruplexes are implicated in different diseases [23].

1.1.2.1 G-Quadruplexes at the Telomeric Regions

The first biologically relevant G4 discovered was a telomeric sequence found in the macronuclei of the ciliate, *Stylonychia lemnae*. This G4 sequence was detected using a single chain variable region fragment of an antibody specifically raised against the quadruplex [24]. In a later study, it has been shown that several G4-structures are present in telomeric strands in "beads-on-a-string" like arrangement where the quadruplexes are separated by flexible linker regions [25]. Telomeres are complexes of nucleoproteins, that are involved in capping the ends of chromosomes. This capping protects them from nuclease attacks and prevent the fusion between two nearby chromosomal ends during cell division [26]. Human chromosomes contain multiple repeats of the G4 forming TTAGGG sequence in high abundance at their 3'-terminus. Several kilobasepairs of

this repeat are followed by a few hundred base pair single stranded overhang that is required for loading the human telomerase reverse transcriptase (hTERT) [27]. hTERT is an enzyme that extends telomeres in germ cells. Telomeres regulate hTERT recruitment and protect the single-stranded G4 overhangs from degradation by the nucleases present in the cells [28].

1.1.2.2 G-Quadruplexes at the Promoters

In more than 40% of the human gene promoters, potential G4 forming sequences were identified [29]. G4s in the promoter regions regulate downstream gene expression by blocking transcription. For example, G4 motifs present in the promoter regions of both c-*MYC* and c-*KIT* oncogenes transcriptionally repress the respective gene expressions [3,30]. This is of significance given that c-MYC has been implicated in cell proliferation and various cancers such as breast, colon, cervical cancers and myeloid leukemia [3].

1.1.2.3 G-Quadruplexes at the Introns

G4-structures in the untranslated regions (UTRs) of RNA transcripts can also suppress the transcription of genes such as human proto-oncogene, *NRAS* and human zinc finger protein, Zic-1 [31,32]. Apart from its repressor function in transcription and translation, G-quadruplexes can interfere with mRNA splicing as well. *In vitro* experiments showed that stabilized G4-structures in intron 6 of the hTERT mRNA resulted in the generation of an inactive splice variant of the hTERT enzyme [33]. G4-structures on intron 3 involves in the production of a shorter splice variant of the mRNA for the tumor suppressor gene, p53 [34]. This in turn regulates the activity of the *TP53* gene which encodes the p53 protein. The early onset of cancers has been correlated with polymorphisms present in the G-quadruplex forming sequence of TP53 [35].

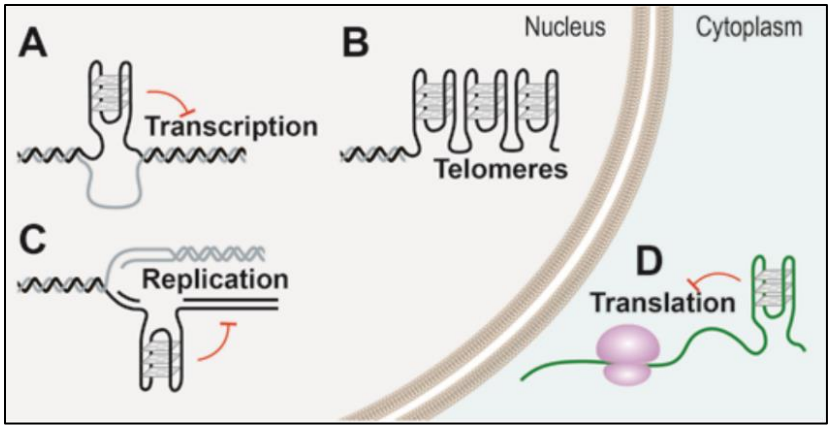


Figure 2. A schematic with possible locations of potential G-quadruplex forming regions in cells as revealed by genome wide studies. These potential G4-structures are distributed throughout the genome in a non-random fashion and particularly enriched in the genomic sequences, promoters and telomeres. Inside the nucleus, G4-structures can develop double stranded G-rich DNAs which transiently unfolds to its component single strands during the processes of replication (C) and transcription (A) and also at the single stranded G-rich extensions of telomeric DNAs (B). In the cytoplasm, G-rich regions of mRNAs can fold into G-quadruplexes and are further involved in translational control (D). The bent red T-bars indicate the hindrance to replication, transcription and translation caused by the presence of G-quadruplexes. Figure from "Survey and summary G-quadruplexes and their regulatory roles in biology" by Rhodes et al. [23]

1.1.2.4 G-Quadruplexes through the Cell Cycle

G4 numbers fluctuate throughout different phases of the cell cycle. In the G0/G1 phase when DNA replication is stopped, G-quadruplexes reach a minimum in terms of abundance. In contrast, a five-fold G4 number increase is observed when DNA replication takes place in the S phase. This in turn suggests that quadruplexes tend to form in G-rich sequences of single-stranded DNA [36]. The non-specific secondary structures of G-quadruplexes may cause the DNA and RNA polymerases to halt which in turn respectively inhibits replication and transcription that ultimately leads to the development of various diseases [37].

1.2 HELICASES AND THEIR IMPORTANCE

Helicases are proteins that catalyze the unwinding and / or rewinding (annealing) of structured DNA or RNA with or without the requirement of energy obtained from nucleoside triphosphate hydrolysis. In addition to the local structural rearrangement of secondary nucleic acid structures, helicases remodel complexes of protein and RNA or DNA by the displacement of bound proteins [38]. In higher organisms, helicases account for approximately 1% of the eukaryotic genome [39]. Depending on their substrate specificity these enzymes can be classified as helicases for DNA, RNA and both DNA – RNA [40]. The first discovered helicases were the DNA helicases of *Escherichia coli* [39]. Over the next decade, eukaryotic initiation factor4A (eIF4A) was discovered as the first RNA helicase [41]. Then followed the helicases that have both DNA and RNA helicase activities: SV40 virus large tumor antigen (T Antigen) is one such enzyme [42,43]. More recently, a further class of helicases which is relevant to this thesis was discovered, namely the G4 helicases which unwind G-quadruplexes. RHAU was the first studied human helicase in the context of its interaction with G-quadruplexes.

Helicases can be classified into six superfamilies (SF) based upon their structural organization [44]. Helicases which are incapable of forming ring structures are grouped into superfamilies 1 and 2 (SF1 and SF2) whereas the ring-forming helicases belong to superfamilies 3-6 (SF3-SF6) [45]. The best characterized helicases are Pif1 (SF1) and RecQ (SF2). In addition, on the basis of the substrate processivity, single-stranded DNA (ssDNA) helicases are grouped under α-helicases, whereas double-stranded DNA (dsDNA) helicases are recognized as the β-helicases [44].

1.2.1 G4 Helicases and Their Importance

G4 helicases can be classified into six superfamilies and follow a similar sub-classification as mentioned before [46]. Moreover, depending on the translocation polarity G4 helicases are categorized as type A (3′-5′ direction movement) and type B (5′-3′ direction movement) helicases [44]. G4 helicases are specialized enzymes that unfold thermodynamically stable G4s, which results in changes to the key regulatory processes of replication, transcription, duplex reannealing and splicing. As a result any abnormal G4 helicase activity has the potential to cause disease [46].

1.2.1.1 Superfamily 1 (SF1) Helicases

The superfamily can be subdivided into SF1A (3′-5′ translocation) and SF1B (5′-3′ translocation) depending upon the direction of translocation. One of the best studied SF1 helicases, Pif1 belongs to the SF1B subgroup, which contains the seven conserved motifs, I, Ia, II, III, IV, V, and VI - typical of SF1 enzymes [47]. Absence of Pif1 leads to a gross chromosomal rearrangement in the yeast genome [46].

1.2.1.2 Superfamily 2 (SF2) Helicases

SF2 is the largest G4 helicase superfamily, with the RecQ-like family helicases as the best studied SF2 subfamily [48]. It is a large family that consists of a C-terminal protein domain (RQC) and a DExH box domain which is comprised of the DEAD, DEAH and DExH subgroups. Biochemical studies revealed that SF2 helicases possess a catalytic core that shows the property of ATP-dependent directional movement at either the 3′ or 5′ end on single stranded or double stranded nucleic acid substrates [44]. G4 unwinding begins when the recognition domain, RecQ binds to the substrate nucleic acids. RecQ helicases can resolve both tetra-stranded and intramolecular G-quadruplexes [49,50].

The proteins relevant to this thesis are RHAU and DDX21 which are DEAH and DEAD box proteins, respectively. Both these proteins are highly abundant in human cells and are implicated in various diseases [46]. The structure and the mechanistic functions of the proteins and both DEAH and DEAD box families, are discussed later in their respective sections of the thesis. While a complete structure is now available for the RHAU protein [51] the crystal structures are still unavailable for DDX21. Due to the overexpression of DDX21 in multiple human cancers and a significant association with early death and a shorter disease-free survival in breast cancer patients, it comes out as a potential gene that could be exploited as a cancer drug target [52,53].

1.2.1.3 Superfamily 3-6 (SF3-6) Helicases

Structurally, the SF3-6 helicase families assemble into a toroid shaped hexameric architecture which radially arrange to form a bi-component NTP binding site. In case of SF3 and SF5 helicases, the hexameric components assemble into a flat closed-ring conformation to unwind nucleic acid by an NTP-coupled translocation-based mechanism. The direction of translocation is

determined by the order of NTP hydrolysis around the hexameric ring. In case of SF4 helicase, the hexameric ring structure unwinds nucleic acids also in an NTP-coupled mechanism but the helicase family binds its substrates in a notched lock washer conformation. The structurally dynamic SF6 family can adopt both the open flat-ring and notched lock washer conformations that again follows an NTP-coupled nucleic acid unwinding mechanism [54].

More particularly, SF3 helicases reportedly have enzymatic activities which include origin recognition and the unfolding of nucleic acid secondary structures [55]. SF3 helicases were first reported in the genomes of small DNA and RNA viruses [56]. Type A SF3 helicases comprise of four conserved motifs (motif A, B, B′, and C) and they maintain a 3′ -5′ directionality. The best characterized SF3 helicase is the papilloma virus helicase, E1 [44].

SF4 helicases were initially found and characterized in bacteria and bacteriophages where they function as replicative helicases. They comprise of five conserved sequence motifs (H1, H1a, H2, H3 and H4) [57]. All characterized SF4 members are type B helicases which translocate in a 5′-3′ direction. A well characterized SF4 helicase is the gp4 protein from bacteriophage T7 [44].

SF5 helicases share some common features with SF4 members, with Rho the best characterized SF5 helicase member. Rho is involved in the transcription termination in the bacterial system which binds to a consensus sequence at the transcription terminator pause site and translocates in the 5′-3′ direction along the transcript to unwind DNA/RNA duplexes. The three conserved motifs in the SF5 helicase family are motifs 1, 1a, and 2 [46].

An example of an SF6 helicase is the minichromosome maintenance protein (MCM), which acts as a replicative helicase in both archaea and eukaryotes [58,59]. In 2017, a cryo-electron microscopy (cryo-EM) study [60] used single-particle reconstruction to solve the structure of

Saccharomyces cerevisiae (*S. cerevisiae*) homolog of CMG bound to a forked DNA substrate. The structure revealed that CMG translocates in 5'-3' direction to reach the forked junction [60]. Based on the translocation polarity CMG acts as a Type B helicase. Other SF6 helicases such as RuvB act as dsDNA trnaslocases which process Holliday junctions (HJs) [61].

1.3 RNA HELICASE ASSOCIATED WITH AU-RICH ELEMENT (RHAU)

RNA **H**elicase Associated with **AU**-rich element (RHAU) is one of the most prominent G-quadruplex resolvases found in human cells. RHAU is also known as DEAH (Aspartic acid, Glumatic Acid, Alanine and Histidine) box protein 36 (DHX36), MLE-like protein 1 (MLEL1), G4 resolvase 1 (G4R1) and DNA resolvase 1. The full length human RHAU protein is comprised of 1008 amino acids and has a molecular weight of 114.8 kDa. A study in 2004 [62] detected two splice variants of RHAU in HeLa cell extracts which means that RHAU has two biologically relevant isoforms. Isoform I refers to the full length RHAU ($RHAU_{FL}$) protein which is expressed by the *Dhx36* gene and due to the presence of the nuclear localization signal (Δ517-530), it is localized in the nucleus. The isoform 2 or $RHAU^{\Delta 14}$ describes a 14 amino acid (Δ517-530) deletion form of the $RHAU_{FL}$ protein [62]. In the absence of the nuclear localization signal, $RHAU^{\Delta 14}$ is localized in the cytoplasm [62,63]. The concentration of RHAU is significantly higher in the nucleus compared to the cytoplasm [62]. The regulatory functions of RHAU in transcription, pre-mRNA processing, RNA folding and assembly, RNA localization and translation agree with its presence in both the cytoplasm and the nucleus, and its ability to interact with both the DNA and RNA G4s [64].

1.3.1 RHAU, a DEAH Family Helicase

RHAU is a DEAH-box (Asp-Glu-Ala-His) protein belonging to the SF2 helicase superfamily, which has the conserved Asp-Glu-Ala-His motif present at the ATP-binding site [46].

Structurally, the core helicase domain of DEAH/RHA helicase family consists of two conserved N-terminus RecA-like helicase domains (RecA1 and RecA2) followed by a winged helix (WH) domain and a C- terminal domain (CTD) [65]. The two RecA components together form a nucleotide 5′-triphosphate (NTP) binding cleft and a single-stranded nucleic acid binding surface. The N-terminal RecA-like domain, RecA1 contains the conserved motifs Q, I, Ia, Ib, II and III, whilst the C- terminal RecA-like domain, RecA2 consists of the conserved motifs IV, V, β-hairpin (HP), and VI in the CTD [65,66]. The HP motif acts as the unwinding element [67]. The C-terminal sub-domains of DEAH/RHA helicases are comprised of a degenerate winged-helix (WH), ratchet-like (RL), and OB-like fold (OB) sub-domains [68,69].

Members of the DEAH helicase family are capable of unfolding duplex DNA up to several dozen base pairs in the 3’ to 5’ direction following a translocation based mechanism [40]. Chen et al. [65] showed that DEAH/RHA helicases can process both DNA and RNA substrates. DEAH family helicases derive energy from ATP hydrolysis to drive the directional movement on a sequestered single strand as the complementary strand becomes displaced [70]. In addition to the remodelling of nucleic acid secondary structures, multifunctional DEAH helicases are also involved in the dissociation of nucleoproteins [46]. DEAH family helicases unwind tetramolecular DNA-quadruplexes following a mechanism which involves loading the helicase onto the ssDNA followed by its directional translocation. Steady state studies and single-turnover kinetics experiments showed that RHAU follows a mechanism typical of DEAH/RHA family members where RHAU loads onto a 3’-single stranded extension and translocates 3’-5’ to unwind DNA G4-structures [71].

1.3.2 RHAU Structure

In an attempt to better understand its function, a number of groups have attempted to obtain structural information about the RHAU protein [5,51,72]. The first crystal structure was that of the RHAU specific motif (RSM) in complex with hTR1-20 (G4 DNA) in which only the G4 DNA was present [5]. However, recent findings with bovine RHAU have produced X-ray structures for both a co-crystal RHAU•RSM construct (RHAU construct devoid of the N-terminal G-rich region) in complex with a DNA G-quadruplex containing a 3' single stranded overhang, and a RHAU•core structure (RHAU devoid of the N-terminal G-rich and RSM regions) [51]. The RHAU domain structure is typical of DEAH/RHA helicases. More specifically, the domains consist of an N terminus extension (NTE) which is comprised of two α-helices: α1 and α2. As found from the solution NMR study, the α1 helix which contains a glycine(G)-rich element is intrinsically disordered. Within the α1 helix of the NTE, the glycine rich element is followed by the DHX36 / RHAU specific motif (DSM / RSM). The NTE is connected with the RecA1 domain of the helicase core by a disordered linker. The core helicase of RHAU is comprised of two RecA-like domains (RecA1 and RecA2) in tandem. Finally, the core is followed by a C-terminus extension (CTE), typical of DEAH/RHA helicases, consisting of the subdomains degenerate-winged-helix (WH), ratchet-like (RL) and oligonucleotide and oligosaccharide-binding-fold-like (OB) [51] (Figure 4).

1.3.3 Substrates of RHAU

RHAU unfolds duplex RNA, RNA–DNA hybrids and other non-canonical DNA structures like triplexes [73]. Tran et al. [62] identified the protein as a high affinity binding partner for the AU-rich element (ARE) of urokinase plasminogen activator mRNA in HeLa cell lysates. Consequently, the protein was named RHAU (RNA helicase associated with AU-rich element). A pivotal 2005 study highlighted that the preferential function of RHAU was its G-quadruplex resolvase activity. The

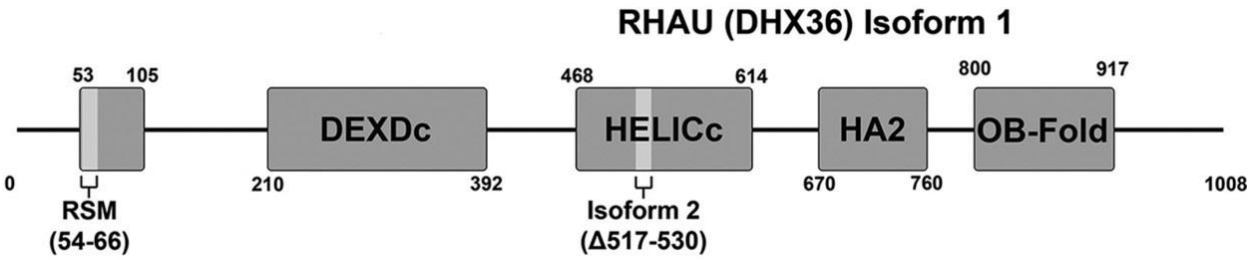


Figure 3. RHAU: A schematic of the respective domains. The domains are predicted by the NCBI conserved domain database. Figure adapted from "RNA helicase associated with AU-rich element (RHAU/DHX36) interacts with the 3′-tail of the long non-coding RNA BC200 (BCYRN1)" by Evan Booy [74].

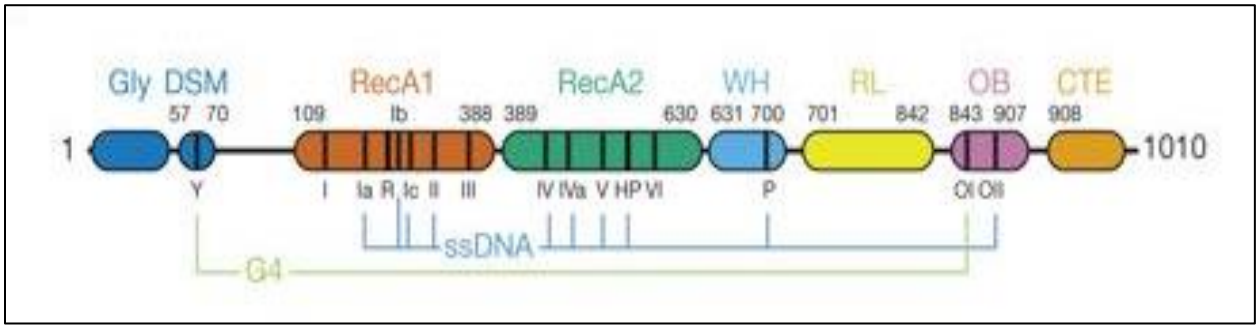


Figure 4. A schematic representation of the various domains of full-length bovine RHAU as recognized in a recent crystal structure [51] indicating the G4 and ssDNA-binding regions. Figure taken from "Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36" by Chen et al. [51].

key result of this study was the identification of the *in vitro* unwinding of tetramolecular DNA G-quadruplexes by RHAU. In a later study, RHAU was proposed to have a higher affinity for tetramolecular RNA quadruplexes [75]. A subsequent study showed that the affinity and the unwinding efficiency are the same for both DNA and RNA substrates [76]. At present, RHAU is one of the three extensively studied helicases that disrupt RNA G4s [71,77]. RHAU was identified as an ATP-dependent helicase capable of resolving intramolecular G-quadruplexes more efficiently than tetramolecular G-quadruplexes *in vitro* [76]. The G4 unwinding efficiency of RHAU depends on the nature of the G4 substrate. Less stable G-quadruplexes with fewer number of G-tetrads are more prone to unwinding by RHAU [78]. In addition, the length and sequence of the 3′-overhang of the G4 substrate also impact on the G4 unfolding. A 3'-overhang of at least 10 nucleotides (nt) is required for the optimal loading and G4 unwinding by RHAU whereas 3 or 5 nt overhang fails to support the loading of the helicase core. Sequence alteration of the 3'-extension influences G4 unwinding as this sequence is responsible for equilibrium binding of RHAU to G-quadruplexes. RHAU shows the highest activity on G4s with A-rich overhangs whereas it shows the least activity on T-rich overhangs. Interestingly, unlike many other DEAH-box proteins, the 5′ extension has no effect on G4 unfolding in the case of RHAU [71].

1.3.4 G4 Binding and Unwinding by RHAU

1.3.4.1 Mode of ATP-Independent G4 Binding and Preferential Substrate Selection

The co-crystal of RHAU•RSM and DNA G4 revealed that DNA binding induces the N-terminal DHX36 / RHAU-specific motif (DSM / RSM) to fold into an α-helix. Mutation of α1 Tyr69 facing the top tetrad of the G4 DNA drastically diminishes the helicase affinity for the substrate. Due to the altered structure of the mutated protein, the Y69A mutant of RHAU•RSM only weakly binds to the the top G4 quartet and so the mutated RHAU dissociates from the G-quadruplex readily.

Furthermore, the mutant protein lacks the repetitive unfolding activity [51]. The α1-helix of the NTE containing the RSM is projected outside of the helicase. A hydrophobic face formed by this α-helix ineracts with the 5′ end of the bound G4 and stacks on the nucleobases of the top tetrad [51]. This mode of binding is similar to G4-recognition by small planar molecules [79].

Together with the RSM and the C-terminal OB-fold-like subdomain, RHAU preferentially binds and unwinds parallel over anti-parallel G-quadruplexes [71] whereas it shows reduced or little activity against parallel – antiparallel mixed and fully anti-parallel G-quadruplexes, respectively [80–82]. The reduced activity is caused by the weakened interaction between the RSM and the G4 because of the steric clashes caused by specific G4 loops. If present, a 5′ G-tract with the opposite polarity also interferes with G4 binding to the OI loop of the OB subdomain [51].

1.3.4.2 Differential Affinity of RSM and $RHAU_{FL}$ for Substrate G4

While the RSM alone is capable of binding G4s the full length helicase binds G-quadruplexes with a much greater affinity. Previous studies reported that the dissociation constant of the G4 and the full length RHAU is 10 pM whereas that of G4 and an isolated RSM is 310 nM, respectively [76,83,84].

1.3.4.3 Mode of ATP-Independent G4 Unwinding

DNA binding to RHAU alone decreases G4 stability and unfolds quadruplexes one residue at a time [51]. This mode of G4 destabilization supports the previous studies where RHAU unwinding was reportedly highly sensitive to the G4 substrate stability [78,81]. Upon binding to RHAU the G4 loses one G-quartet at the top which is replaced by a non-canonical tetrad made up of Gs and other nucleotides. As shown in a FRET experiment both in the presence and absence of an ATP analog, binding of a three-tiered parallel G4 to the RHAU protein alone is capable of inducing

rearrangement of its helicase core. As revealed by superimposing the core RecA1 domains of the RHAU•core and DNA G4-bound RHAU•RSM X-ray structures, the mentioned structural rearrangement of the core region is executed by the rotation of the CTE and RecA2 domain. The crystal structure of the RHAU•RSM – DNA G4 complex also revealed that the core domains of RecA1 and RecA2 as well as the C-terminal domains arrange like a trefoil upon DNA binding. Several motifs and loops located in the core domains of RecA1 and RecA2 along with the C-terminal WH and OB domains associate to interact with ssDNA, such that the 3' face of the G4 contacts with the OB loop I. The core RecA1 domain, WH and OB loop II interact with the 3' single stranded phosphate backbone of the G4. Together the C-terminus of the RSM containing α1-helix and the OI loop of the OB domain form a composite surface. This surface extensively interacts with the sugar-phosphate backbone of 5' end single-stranded residues of the substrate DNA G4 by the formation of H-bonds. The overall cationic nature of the RSM helix provides a path for the negatively charged single-stranded DNA. Only the ssDNA can enter the positively charged groove between the RecA2 domain and the C-terminus [51] because the groove is too narrow to accommodate multi-stranded DNA [38,71,80,81]. This structural rearrangement of the core in turn pulls on the single-stranded DNA overhang resolving G4 by a single residue at a time. A 3′ single-stranded overhang is essential for the helicase activity of RHAU. The top G-quartet becomes eliminated as the helicase pulls out the bottom-most G of the bottom G-tetrad into the 3′ single-stranded extension. The presence of one Watson–Crick pair instead of a G•G Hoogsteen H-bond in the altered non-canonical top quartet drastically destabilizes the quadruplex structure, with ATP most likely assisting in the rapid release of DNA in the process [51].

1.3.5 Functions of RHAU

RHAU was highlighted for its function in mRNA deadenylation and decay. It has been

proposed that RHAU interacts with the exosome and polyadenylate-specific ribonucleases (PARN) in an RNA-independent manner whereas its NFAR1 and HuR binding activities are RNA-dependent [62]. Gradually, several other regulatory functions of RHAU have been suggested such as its roles in transcription and translation, mRNA destabilization and blood cell formation [62,64,85,86]. RHAU downregulates hematopoiesis in developing murine embryo causing hemolytic anemia in mice which results in anomalies of murine blood cell differentiation [64]. RHAU has also been reported to co-precipitate with the neuronal precursor-micro-RNA-134 (miRNA134) which is implicated in gene inhibition and also in the growth of dendritic spines that accept axonal signals in synaptic junctions [87].

1.3.5.1 RHAU in Transcription and Translation

As a regulator of transcription, RHAU induces tissue-nonspecific alkaline phosphatase (TNAP) transcription which in turn is involved in the formation and differentiation of bones [85]. As a translational regulator, RHAU is implicated in the regulation of transcriptional regulatory factor Yin Tang 1 (YY1) and the homeobox transcription factor Paired Like Homeodomain 1 (PITX1) [86,88]. RHAU resolves the G4s formed in the promoter regions located upstream of the YY1 gene and thus upregulates YY1 expression. Most of the studies involving YY1 suggest that YY1 has proliferative role in cancer development and progression. Therefore, in relation with the oncogenic YY1 gene, RHAU has a positive correlation with cancer [86,89]. In the case of PITX1, RHAU downregulates PITX1 expression. As a homeobox protein PITX1 plays an important role in the differentiation of the developing pituitary gland, and by acting as a tumour suppressor gene. Therefore, RHAU plays an important role in the development of cancer through the downregulation of PITX1 [86,88].

1.3.5.2 RHAU in Cellular Stress

The presence of RHAU has been reported in the nucleoplasm and nuclear speckles in HeLa cells. RHAU was found in association with RHA helicases such as p68 and p72 within nucleolar caps upon transcriptional arrest.[63] Nuclear speckles are the regions in the nucleus which are enriched with pre-mRNA splicing factors and nucleolar caps. Nuclear speckles consist of RNA binding proteins that are present in high abundance in the nucleolus during cellular stress or transcriptional arrest. RHAU targets and binds RNA in the stress granules. Stress granules are structures located in the cytoplasm that sequester RNAs and cellular proteins during cellular stress [90].

1.3.5.3 RHAU and Telomere Functions

Human RHAU binds and unwinds G-quadruplexes found in the RNA component of the human telomerase reverse transcriptase (hTERT) enzyme in the 5' to 3' direction. hTERT is a ribonucleoprotein which is highly active in proliferating germ cells and cancer cells, yet absent in somatic cells. The hTERT holoenzyme is comprised of a protein and an RNA counterpart. The catalytic protein subunit is TERT whereas its human telomeric RNA component is hTR that acts as a template for telomeric extension during reverse transcription [91]. hTERT becomes activated during unchecked replication in cancerous cells [92]. Two separate *in vitro* studies have shown that RHAU processes hTR G4s in a 5' to 3' direction [76,93] which was subsequently confirmed *in vivo* [94].

Human telomeric RNA is a 452 nucleotide long sequence which is transcribed by RNA polymerase II. In vertebrate telomeric RNA, there are eight highly conserved regions (CRs) in a 5' to 3' direction [95] and ten conserved paired helices (P) together forming four domains in the secondary structure. CR1 is a domain located 45 nucleotide downstream to the 5' end of hTR and

encodes the template sequence 5'-CUAACCCU-3'. During human telomeric reverse transcription, the correct nucleotide incorporation onto the template strand is mediated by the hTR-hTERT interaction and RNA-RNA base-pairing in the region upstream of the template sequence [96]. The P1 helix is formed upstream of the 5' template end and functions like an anchor which is tethered to the template through a short linker region. A mutation that disrupts the P1 helix formation results in the incorporation of additional nucleotides onto the template and generates telomeric repeats of incorrect length in the process [96]. The P1 helix is comprised of two helical components, P1a and P1b, which are separated by a loop region. It is P1b which is essentially needed for correct template incorporation. The nucleotides within the potential G4 forming regions of the hTR 5' terminus are required to form the P1 helix. During hTERT maturation, the G4-structures within the P1 sequence may function to protect the single stranded 5' terminal region of hTR from degradation. The protective G4s must be resolved to allow P1 helix formation and correct template demarcation [83,93]. The G4 unwinding event allows P1 to form its proper helical structure which is essentially required to ensure that the correct nucleotides are incorporated onto the template strand during telomeric reverse transcription [83,93,94]. RHAU interacts equally with 5'-terminus hTR G4 DNA and RNA sequences. Knockdown experiments revealed that the absence of endogenous RHAU in HEK293T cells results in a reduction of the average telomere length [83].

1.4 DExD-BOX HELICASE 21 (DDX21)

DExD-box helicase 21 (DDX21) is one of the most prominent RNA G-quadruplex resolvases found in human cells. DDX21 is a DEAD (Aspartic acid, Glumatic Acid, Alanine and Histidine) box helicase protein which is also known as RNA helicase 2 and Gu [97,98]. The full length human DDX21 ($DDX21_{FL}$) protein is comprised of 783 amino acids and has a molecular weight of 87.3 kDa. In 1993, the full length DDX21 was first purified from the nuclear extract of HeLa

cells [62]. DDX21 has two biologically existing isoforms, the full length DDX21 ($DDX21_{FL}$) is isoform 1 whereas the isoform 2 or $DDX21_{69-783}$ describes an N-terminal 68 amino acid deletion form of the $DDX21_{FL}$ protein.

1.4.1 DDX21, a DEAD-Box Family Helicase

DDX21 is classified as a DEAD-box (Asp-Glu-Ala-Asp) protein which has the conserved Asp-Glu-Ala-Asp motif present at the ATP binding site. The glutamine rich Q-motif which is an integral part of DEAD-box, is responsible for the helicase specificity towards ATP. DEAD-box proteins are comprised of a highly conserved helicase core which is responsible for the binding of ATP and RNA, and auxiliary domains that surround the core region (Figure 5). The two RecA-like domains of the helicase core form a cleft which contains an ATP-binding site. For a productive hydrolysis of ATP, the cleft must be closed. Structurally, the RNA-binding modules in both the core helicase domains of DEAD-box proteins are highly similar, which indicates their origin from the same anion-binding module [99]. DDX21 has a core helicase domain which is flanked by its N- and C-terminus extensions, and unwinds double stranded RNA [97,100].

DEAD box helicases use ATP to bind and unwind RNA duplexes in either the 3′ or 5′ end, where effective unwinding of duplexes is limited to less than two helical turns, and fewer than 10 or 12 base pairs. Mechanistically, the helicase family follows local strand separation where the helicase loads directly onto the duplex and unwinds the strands in an ATP-dependent fashion [99]. DDX21 unwinds dsRNA strictly in the 5'-3' direction following local strand separation [53].

1.4.2 DDX21 Expression and Localization

DDX21 is highly expressed in human cells ($\sim 4.2 \times 10^4$ molecules per cell) [97]. The helicase is present in both the nucleolus and nucleoplasm. c-JUN, a transcription factor which is involved

in apoptosis, interacts with DDX21 to promote its translocation from the nucleolus to the nucleoplasm [101–103]. WDR46, a nucleolar scaffold protein that contributes to the inner nuclear organization, interacts with DDX21 and localizes it in the nucleolus [104]. In addition, DDX21 colocalizes in the cytosol of myeloid dendritic cells (mDCs) in a complex together with another DEAD-box protein DDX1, RHAU and TIR domain-containing adapter molecule 1 (TICAM1), which is an adaptor molecule involved in the innate immune response. Upon infection with dsRNA virus, the multi-helicase-TICAM1 complex binds with the viral dsRNA and is translocated to the mitochondria [105]. DDX21 is translocated from the nucleus to cytoplasm in the early infection of Dengue virus in a mechanism which is yet unknown [106]. Finally, DDX21 is found in varying abundance in different cancer cells [107,108], suggesting roles of the protein in cancer progression.

1.4.3 DDX21 Structure

The DEAD-box protein DDX21 is comprised of a largely conserved helicase core that contains the DEAD sequence along with a flanking N-terminal domain and a C-terminal domain. These domains are highly divergent and are likely to interact with numerous proteins and RNAs within the cell [109]. The core helicase domain of DDX21 is comprised of the signature motifs (Q, I–VI) required for DEAD-helicase function (Figure 5) [110].

The C-terminal region of DDX21 has a unique motif containing three FRGQR repeats (Figure 5) and one PRGQR sequence. Each of these sequences are separated by five amino acids, with the FRGQR repeats contributing to the RNA binding activity of DDX21. For example, the PRGQR mutation to YEGIQ significantly affects the G4 binding affinity of the protein [77].

1.4.4 Substrates of DDX21

DDX21 is the first reported eukaryotic RNA helicase that unwinds dsRNA in the 5’-3’ direction[53].

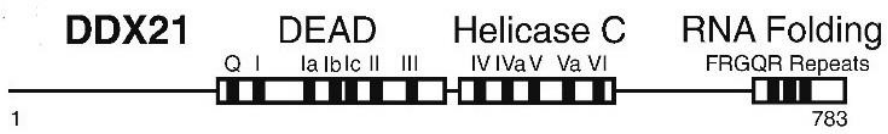


Figure 5. A schematic presentation of different domains of the DDX21 protein. As reviewed by Hammond et al., the core helicase domain of the protein is comprised of a DEAD domain and a helicase C domain. Figure adapted from "A Survey of DDX21 Activity During Rev/RRE Complex Formation" by Hammond et al. [110]

DDX21 is also capable of resolving hybrid DNA-RNA duplex provided the protein can access the 5' RNA end, and both RNA and DNA G4s [53,77]. McRae et al. [53] showed that DDX21 also binds to DNA G4 though with a weaker affinity than to RNA G4 substrates.

1.4.5 Mode of Function

A large excess of DDX21 is required in order to achieve complete unwinding of the dsRNA substrate. The initial study with the protein revealed its following two functions. Firstly, it has an ATP-dependent RNA duplex helicase activity and secondly an ATP-independent foldase activity to impart secondary structures into the single stranded RNA molecules [97]. Two different domains of DDX21 were considered to be responsible for these two respective functions. Firstly, the core helicase domain was found to unwind double stranded RNA and secondly, the C-terminus of the protein was thought to be responsible for the foldase function, respectively [97,100,111,112]. A subsequent study involving nuclease susceptibility assays provided further evidence that the G4 binding core domain of DDX21 was indeed responsible for the quadruplex unwinding [112].

The C-terminal 209 residues of the protein ($DDX21_{574-883}$/$DDX21_{C209}$) is sufficient for G4 binding although with a weaker affinity compared to the full-length protein. DDX21 has a unique motif in its C-terminal region that contains three FRGQR repeats and one PRGQR sequence. Each of these sequences are separated by five amino acids. PRGQR mutation to YEGIQ significantly reduces the affinity of the protein towards G-quadruplexes [112]. A thermal shift assay shows very little difference between the melting points of the native and the mutant proteins indicating the mutation has little or no effect on the protein secondary structure which in turn suggests that secondary structural change is not the cause behind the reduced affinity of the mutated protein [77]. Moreover, Valdez et al. [112] showed that the same mutation irreversibly destroys the foldase activity

of DDX21. However, in a later study it was found that DDX21 fails to fold single stranded RNAs which are incapable of forming G4s [77].

The C-terminus domain of DDX21 primarily binds directly to the phosphoribose backbone of the substrate G4. In the case of RNA G4 substrates, recognition and specificity of the helicase are mediated by the 2'-OH groups of the loop nucleotides. DDX21 does not interact with the ribose sugar of the terminal uracil (U) located in TERRA, an RNA G4 that is integral to the telomere which is involved in the telomeric heterochromatin formation and regulation of telomere length. The terminal G4 U could be providing structural stability to the loop of the TERRA G4 [53]. Unlike RHAU, DDX21 does not interact with the tetrad face of the G4 substrate, although both the helicases share affinity for the G4 loop sequence [53,113]. DDX21 binds to G4 RNAs with much higher affinity than that with DNA G4s. The higher affinity is attributed to the interactions between the C-terminus of DDX21 and the 2′ OH groups of the G4 loops. Isothermal titration calorimetry (ITC) suggests that the driving force for the binding between DDX21 and G4 is a coulombic interaction. Electrostatic force works between the polar arginine residues within the $DDX21_{C209}$ construct and the G4 phosphoribose backbone of the ribose sugars and the loop sequence, where the contacts are observed by 2D saturation transfer difference (2D-STD) NMR [53]. Although G4 recognition is usually mediated by the interactions between the protein and the G4 loop, DDX21 reportedly interacts with the entire ribose sugar backbone of guanine core [53,114]. The interaction between G4 substrates and DDX21 can be disrupted by using ligands such as Distamycin A derivatives that bind to the grooves of G4 quadruplexes [53,115].

1.4.5.1 Mode of ATP-Independent – ATP Dependent Helicase Machinery

A basal level of helicase activity was observed for DDX21 in the absence of ATP. Interestingly, the ATP-independent helicase activity of the DDX21 protein is absent in the core

$DDX21_{183–620}$ construct, whereas it retains almost 80% of the ATP-dependent helicase activity of the $DDX21_{FL}$. This result indicates a synergy between the activities of the core and the accessory domains of DDX21 during RNA remodeling [110]. It was hypothesized that the $DDX21_{C209}$ construct causes a distortion of the G4 structures in an ATP-independent mechanism such that the quadruplex secondary structure adapts a less stable conformation which allows the core helicase to readily unwind the G4 using the energy obtained from ATP hydrolysis [53].

1.4.6 Functions of DDX21

The several regulatory functions of DDX21 including its involvement in AIDS [110], cancer [107,116] are discussed in the following sections.

1.4.6.1 DDX21 in Human Immunodeficiency Virus (HIV) Infection

Immunoprecipitation and phenotypic characterization experiments showed that DDX21 is up-regulated in Human immunodeficiency virus (HIV) infection, thus DDX21 supports viral replication leading to an efficient HIV virus production [117,118]. HIV-1 transports unspliced and partially spliced viral RNA transcripts to the cytoplasm using a specialized nuclear export pathway. Rev is the central viral protein in this pathway, which binds to the Rev response element (RRE) regions in stem IIB on unspliced viral transcripts. This binding event eventually causes the viral RNAs to be transported from nucleus to the cytoplasm via the host cell chromosome region maintenance 1 (CRM-1) export pathway [119]. DEAD box proteins bind with Rev using their DEAD domains. DDX21 uses its DEAD domain to bind to Rev with high affinity and stimulates its own ATPase activity. The nuclear diffusion inhibitory signal (NIS) of Rev directly interacts with DDX21 [110]. A dose-dependent inhibition of the ATPase function and ATP-dependent helicase activity of DDX21 by Rev has already been reported. However, the *in vivo* Rev concentration is

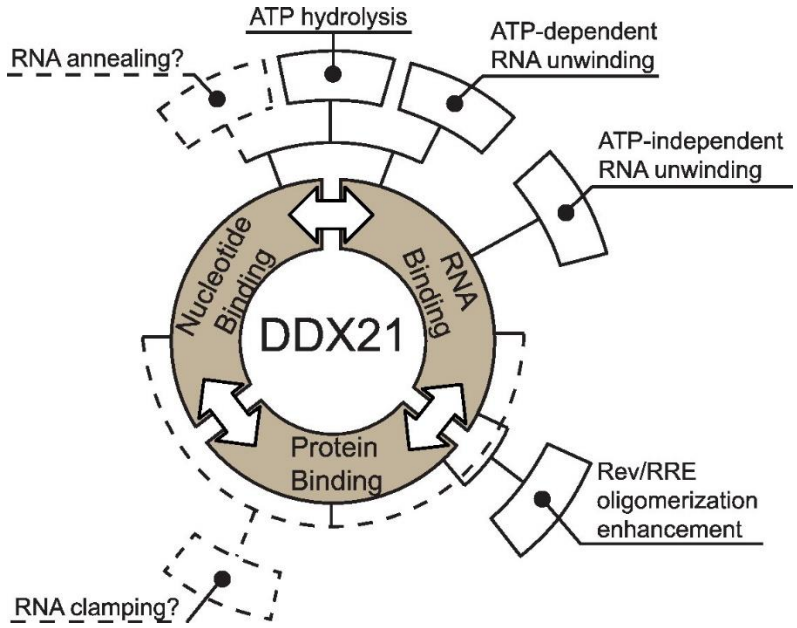


Figure 6. A schematic representing the binding and the enzymatic functions of DDX21 in relevance to HIV infection. The helicase protein has binding affinities for RNA, nucleotide, and protein (The boxes at the center). Often one binding event affects another binding phenomenon (indicated by double-edged arrows). More than one binding event can synergistically function to develop enzymatic effects (indicated by the boxes in the outer layer). Multiple binding events are often needed for a single enzymatic activity (solid and dashed lines from outer to inner layers). The solid black boxes in the outer tier are representatives of enzymatic or functional activities reviewed by Hammond et al. The dashed black boxes indicate DDX21 activities [110]. Figure adapted from Putnam et al. [120] and taken from Hammond et al. [110]

not high enough to inhibit the ATPase function of the total nuclear DDX21 population [121,122]. DDX21 enhances Rev binding to the RRE, which initiates an enhanced oligomerization pattern of Rev on RRE. Both DDX21 and another DEAD box protein DDX1 can accelerate Rev monomer binding to the RRE [110]. Therefore, a knockdown of either DDX1 or DDX21 has a severe impact on HIV release from the cells. The overlapping functions of these two helicase proteins indicate their direct interaction and their roles as an ensemble in pathways involving double-stranded RNA unwinding [105,118,123]. The specific roles of DEAD-box proteins in the HIV life cycle remains unclear, partly because of the vaguely understood mechanistic roles of these helicases in native cellular processes but also due to the lack of authentic helicase substrates for the HIV-specific functions [110].

1.4.6.2 DDX21 in Cancer

DDX21 is expressed in high abundance in a subset of breast cancer tissues and established cell lines compared to normal breast tissues [116]. As a result, the protein can be used as a biomarker for patients diagnosed with these cancers. An acute knockdown of DDX21 drastically reduces tumorigenicity both *in vitro* and *in vivo* [116]. Consequently, DDX21 might be considered as a therapeutic target for breast cancer which is the second most common cancer among women in Canada and the United States [116].

Interestingly, a 2018 study [107] reported DDX21 as a suppressor of breast cancer metastasis. Downregulation of miR-218-5p, a miRNA predicted to target DDX21, stabilizes DDX21 expression. A common indication of metastatic breast cancer is the impaired expression of miRNAs [124]. Both *in vivo* and *in vitro* studies revealed that DDX21 inhibits epithelial-mesenchymal transition (EMT), a crucial event in progression towards general cancer metastasis [107,125]. Snail, a zinc-finger transcription factor and the most studied master regulator of EMT and

various tumor metastasis, binds to the E-box sequences within the target gene promoters and triggers EMT during cancer progression [126–128]. Overexpression of Snail downregulates DDX21 expression and thus reverses inhibition of cell invasion mediated by the DEAD-box protein. Interestingly, DDX21 binds to the Snail promoter, functions independent of its helicase activity and suppresses Snail transcription [107].

1.5 OBJECTIVES

We investigated optimization of expression and purification of the RHAU protein to use it further in biophysical studies and finally to crystallize the protein. We used a comprehensive protein expression approach to optimize production of the full length RHAU ($RHAU_{FL}$) and RHAU lacking N-terminal 50 amino acid sequence ($RHAU_{51\text{-}1008}$) in *E. coli*. In addition, we attempted to optimize the purification of the $RHAU_{FL}$ protein. Although structures of free RHAU and RHAU in complex with nucleic acids have recently been published using non-human homologs of the protein [51,72], structures of the human RHAU protein both in free and nucleic acid bound forms could give a more comprehensive insight into the G4 recognition by helicases in humans.

We also produced the C-terminal construct of DDX21 ($DDX21_{C209}$) and the full length DDX21 ($DDX21_{FL}$) proteins in *E. coli* strain to further use the proteins in biophysical studies and explore their propensity to grow crystals. In addition, we performed quality control check for the folding and nucleic acid binding properties of the $DDX21_{C209}$ protein by pre-established methods [53] following its application in crystallization trials both in a free state and in complex with the quadruplex-forming 22 nucleotide DNA sequence from the human telomeric G4 (hTel). Although recent findings [53] provided information regarding G4 binding to the $DDX21_{C209}$ protein, a better understanding of the full length protein binding and mechanistic functions of the DDX21 protein in relation to its structure could be realized from the crystal structure. This is important for

investigation due to the significant association of the DDX21 protein with RNA G4s in physiological environment [77] and the involvement of DNA and RNA G4s in the various regulatory pathways [23] including correct functioning of hTERT [83,93].

2. MATERIALS AND METHODS

2.1 Cloning of Plasmids

The cloned construct of ecoli_DHX36_TEV_HIS8_pET-21a(+)/(His_9-$RHAU_{FL}$) plasmid was designed by Dr. Markus Meier of our laboratory, and the plasmid was ordered from a commercial facility. The *E. coli* codon optimized $RHAU_{FL}$ DNA sequence with a C416S point mutation was cloned into pET21a(+) plasmid using the *Nde* I and *Eco* RI restriction sites on the vector with a nine histidine (His_9) tag followed by a tobacco etch virus protease (TEV) cleavage site inserted 5' to the RHAU DNA sequence. To prepare a fresh stock of plasmid, the cloned construct was chemically transformed into *E.coli* DH5α cells (New England BioLabs). The transformation was plated out onto LB-agar/ampicillin (Fisher Scientific) plate with the plate incubated overnight at 37°C. A 5mL plasmid culture was grown overnight in LB/ampicillin medium at 30°C and the plasmid was purified using the GeneJet plasmid DNA mini-prep kit (Thermo-Fisher Scientific). Each of the putative clones were digested with *Nde* I and *Eco* RI to check that an insert of the correct size was obtained when analyzed on a 1% TAE agarose gel. Only those clones which gave a fragment of the correct size were sent for DNA sequencing (RIOH DNA Sequencing, Winnipeg) to confirm that the DNA sequence was correct.

The cloning of His_9-$RHAU_{51\text{-}1008}$ plasmid was performed by Dr. Evan Booy of the McKenna laboratory. His_9-tagged $RHAU_{51\text{-}1008}$ was cloned into the pET21a(+) vector using the *Nde* I and *Eco* RI restriction sites on the vector with a His_9 tag followed by a tobacco etch virus protease (TEV) cleavage site inserted 5' to the RHAU DNA sequence. The cloned construct was transformed into Z-competent *E. coli* DH5α cells, with the transformation plated out onto a LB-agar/ampicillin plate and incubated overnight at 37°C. Single colonies were inoculated into 5mL LB/ampicillin medium at 30 °C overnight, with the resultant plasmid cultures purified using the

GeneJET Plasmid Extraction Kit (Thermo-Scientific). The presence of the correct insert in the plasmid was confirmed by restriction enzyme digestion using *Nde* I and *Eco* RI, followed by agarose gel electrophoresis on a 1% TAE agarose gel. Those clones containing an insert of the correct size were subjected to DNA sequencing to confirm that the sequence was correct. All sequencing reactions were carried out by a commercial facility (RIOH DNA Sequencing, Winnipeg, Manitoba, Canada).

The cloning of both the $DDX21_{C209}$ and $DDX21_{FL}$ plasmids were performed by the McKenna laboratory. His_6-tagged $DDX21_{C209}$ was cloned into the pET-28b(+) vector using the *Nde* I and *Xho* I restriction sites on the vector with a six histidine (His_6) tag followed by a thrombin cleavage site inserted 5' to the $DDX21_{C209}$ sequence.

His_6-tagged $DDX21_{FL}$ was cloned into the pET-28b(+) vector using the *Nde* I and *Xho* I restriction sites on the vector with the sequential tags of IgG domain B1 of Protein G (GB1) domain, N-utilization substance (NusA), streptavidin binding peptide (Strep II) and an eight histidine (His_8) sequence followed by a Tobacco Etch Virus protease (TEV) cleavage site inserted 5' to the $DDX21_{FL}$ isoform I cDNA sequence.

2.2 Expression of Proteins

To express the His_9-$RHAU_{FL}$ protein, the His_9-$RHAU_{FL}$ plasmid was first transformed into $CaCl_2$ competent *E. coli* C41(DE3) and *E. coli* BL21(DE3) cells, with the transformations plated onto LB-agar/ampicillin plates and incubated overnight at 37°C. Expression of His_9-$RHAU_{FL}$ was initially performed as follows. A single colony was inoculated into 0.5mL LB broth supplemented with 100µg/mL of ampicillin, shaken overnight at 180rpm at 30°C. A 50mL LB/ampicillin culture was inoculated with the pre-culture and grown at 37°C with constant shaking at the same speed. Once the OD_{600} reached 0.5, the expression of the RHAU protein was induced by adding isopropyl

β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5mM. In order to determine the optimum expression duration after induction, cells were harvested at hourly intervals for eight hours, with a final time point taken the following morning. The 1mL cell samples were centrifuged at a speed of 14,800rpm for 1min in a microfuge, with the cell pellets resuspended in 70µL milliQ water, 5µL glycerol, 5µL β-mercaptoethanol (β-ME) and 20µL of 5x gel loading dye (0.5M Tris-HCl, pH6.8, 20mM TCEP, 10% SDS, 25% glycerol, 0.02% Xylene Cyanol FF and 0.02% Bromophenol Blue). Samples were analyzed using SDS-PAGE by loading onto an 8% polyacrylamide gel with the gel run at 7V for 90min before staining for 30min with Coomassie Blue (50% methanol, 8% acetic acid, 0.2% Coomassie Blue G-250) and destaining for 30min. Expression of RHAU was checked with Western blot (see section 2.4 for details) using a mouse monoclonal anti-RHAU primary antibody (12F33), raised against a peptide corresponding to the C terminus of RHAU, 991-1007 amino acids [129].

The overexpression of the His_9-$RHAU_{FL}$ protein was optimized by varying the expression strain, IPTG concentration, the OD_{600} at which the cells were induced and finally the temperature at which the experiment was performed after induction. In one set of experiments, using the expression strain *E. coli* C41(DE3), the IPTG concentration was kept constant at 0.5mM, the induction OD_{600} was left unaltered at 0.5 but the induction temperature was varied between 16°C, 20°C, 28°C and 37°C. In another set of experiments, following a modified Yoo et al. protocol [130], the selected expression strain was *E. coli* BL21(DE3), the IPTG concentration was 0.01mM, the induction OD_{600} was either 0.4 or 0.8, and the induction temperature was 16°C.

The standard condition used for the expression of the His_9-$RHAU_{FL}$ protein was as follows. A single colony of His_9-$RHAU_{FL}$ transformed into C41(DE3) was inoculated into 50mL LB supplemented with 100µg/mL ampicillin and grown overnight at 180rpm at 30°C. The following

morning 10mL of pre-culture was added to 1L LB/ampicillin medium and grown at 180rpm at 37°C until the OD_{600} reached 0.5. Expression was induced by adding 0.5mM IPTG and the culture was grown for a further four hours at 37°C. The cells were harvested by centrifugation at 6500rpm for 20min at 4°C, with the cell pellet subsequently stored at -80°C.

To express His_9-$RHAU_{51-1008}$ protein, the His_9-$RHAU_{51-1008}$ plasmid was transformed into $CaCl_2$ competent *E. coli* C41(DE3) and *E. coli* BL21(DE3) cells, with the transformations plated onto LB – Agar / Ampicillin plates and incubated overnight at 37°C. Expression of His_9-$RHAU_{51-1008}$ was initially performed as follows. A single colony was inoculated into 0.5mL LB broth supplemented with 100µg/mL of ampicillin, shaken overnight at 180rpm at 30°C. A 50mL LB / ampicillin culture was inoculated with the pre-culture and grown at 37°C with constant shaking at the same speed. In order to optimize the overexpression of the truncated RHAU protein, once the OD_{600} reached 0.5, the expression of RHAU was induced by addition of IPTG to varying final concentrations of 0.5mM and 2mM, with the induction temperature varied between 37°C and 28°C. In order to determine the optimum expression duration after induction, cells were harvested at hourly intervals for five hours, with a final time point taken the following morning. The 1mL cell samples were centrifuged at a speed of 14,800rpm for 1min in a microfuge, with the cell pellets resuspended in 70µL milliQ water, 5µL glycerol, 5µL β-mercaptoethanol (β-ME) and 20µL of 5x gel loading dye. Samples were analyzed using SDS-PAGE by loading onto an 8% polyacrylamide gel. Expression of RHAU was checked by performing a Western blot (see section 2.4 for details) using the mouse monoclonal anti-RHAU primary antibody (12F33) for detection of the RHAU protein.

The His_6-$DDX21_{C209}$ protein was expressed following a previously described protocol [53] in the *E. coli* strain BL21(DE3). However, the cells were harvested by centrifugation at 6500rpm

for 20min at 4°C, and stored at -80°C.

Expression of the His_8-$DDX21_{FL}$ protein was performed following the same protocol as described for the truncated protein, His_6-$DDX21_{C209}$.

2.3 Purification of Proteins

To better purify His_9-$RHAU_{FL}$ protein multiple purification protocols were employed. In a common step in the purification of the RHAU protein, the cell pellet stored in the freezer was thawed on ice. In the first method of purification the thawed cells were resuspended in 30mL of Lysis Buffer pH7.0 (20mM sodium phosphate, 2mM imidazole, 300mM NaCl) which was used in a previously described protocol [131], in the presence of the protease inhibitor cocktail (1mM AEBSF, 10µM pepstatin A and 10µM bestatin hydrochloride) supplemented with 0.4mg/mL DNase I and 1mM $MnCl_2$. The cell pellet was homogenized in the above buffer using Dounce homogenization and incubated with DNase I for 30min on ice before lysing the cells using a French Press. The protein was purified by affinity chromatography using 5mL HisPur™ cobalt resin (Thermo Scientific) that had been equilibrated in Lysis Buffer pH7.0. Following the loading of the cell lysate, the column was washed with Lysis Buffer pH7.0 until the absorbance at 280nm (A_{280nm}) of the flowthrough was 0.02. The protein was then eluted with a buffer containing 20mM sodium phosphate, 150mM imidazole, pH7.0 and 300mM NaCl. The eluate was concentrated using a 50kDa molecular cut-off (MWCO) concentrator up to 1mg/mL estimated spectrophotometrically at 280nm, and dialysed overnight into an imidazole free buffer containing 20mM sodium phosphate, pH7.0 and 300mM NaCl. The concentrated elute was further purified by size exclusion chromatography (SEC) using a Superdex 200 10/300 GL column (GE Healthcare) in 20mM sodium phosphate, pH7.0, 300mM NaCl run at 0.4mL/min. SEC fractions were analyzed by SDS-PAGE on an 8% polyacrylamide gel. Subsequent to the SDS-PAGE analysis the protein sample

was concentrated in a 100kDa MWCO Amicon Ultra concentrator in an attempt to remove lower molecular weight contaminants. To determine the possible disulfide linkages between the His_9-$RHAU_{FL}$ protein with the contaminants, the protein sample was treated with β-mercaptoethanol (β-ME). To find out if nucleic acid contaminants in the protein sample were responsible for the binding of the RHAU protein to contaminants, spectrophotometric readings were taken at 260nm and 280nm and their ratio was determined.

To overcome precipitation and aggregation problem with the purified His-tagged protein, a better buffer condition for the cobalt affinity purification needed to be developed. With that aim and to reduce the possible harmful effect of imidazole on the protein folding, the minimum imidazole concentration in the elution buffer was determined. In order to do that the elution of His_9-$RHAU_{FL}$ was performed with a gradient of imidazole concentration from 0 to 500mM with the main buffer components as follows. 20mM HEPES, 300mM NaCl, 1mM TCEP, pH7. An ÄKTA (GE Healthcare Life Sciences) FPLC was run at 0.4mL/min to perform the imidazole gradient purification. Purification fractions of varying imidazole concentration were analyzed by SDS-PAGE on an 8% polyacrylamide gel.

In order to avoid using imidazole during the purification of the His-tagged protein, the next purification protocol was based on lowering the pH to elute the protein from a histidine affinity cobalt resin column. After the lysis step, the protein was bound to a cobalt affinity column in a binding buffer containing 20mM Tris-HCl, pH8.5, 300mM NaCl, 2mM imidazole hydrochloride, 2mM β-mercaptoethanol and 10% glycerol. The resin-lysate slurry was washed with the wash buffer 20mM Tris-HCl, pH8.0, 300mM NaCl, 2mM imidazole hydrochloride, 2mM β-mercaptoethanol, 10% glycerol, before eluting the protein in a buffer containing 20mM acetic acid/acetate, pH5.0, 300mM NaCl, 2mM β-mercaptoethanol and 10% glycerol. The eluate was

first dialysed into the β-ME free buffer 20mM MES, pH6.0, 1mM TCEP, 300mM NaCl, 1mM EDTA, 10% glycerol before dialysing into a second EDTA-free buffer containing 20mM MES, pH6.0, 1mM TCEP, 300mM NaCl and 10% glycerol. The purity of the pH eluted protein was analysed by SDS-PAGE on an 8% polyacrylamide gel.

In the third His_9-$RHAU_{FL}$ purification protocol, which was also based on lowering the pH to elute the His-tagged protein from a cobalt resin column, the thawed cells were resuspended into lysis buffer containing 20mM Tris/Tris-HCl, pH8.5, 50mM NaCl, 2.5mM $MgCl_2$, 0.1mM $CaCl_2$ and 10% glycerol, in presence of the protease inhibitors (1mM AEBSF, 10μM pepstatin A and 10μM bestatin hydrochloride), and 2μg/mL RNase A. Following Dounce homogenization of the cell suspension, the cells were lysed using a pre-cooled French Press at 4°C. The lysed cells were treated with 0.4mg/mL DNase I at 4°C for 30min, with the ionic strength of the lysate increased to 500mM by addition of NaCl salt, and 1.5μL β-ME/mL lysate volume was used to reduce disulfide bonds to thiol groups. The cobalt resin was equilibrated with 5 column volumes of Equilibration Buffer pH8.5 (20mM Tris-HCl, 500mM NaCl, 2mM β-mercaptoethanol, 10% glycerol). Immediately after the passage of the lysate through the equilibrated resin beads, the resin was washed with 5 column volumes of Equilibration Buffer. The column resin was washed with wash buffer (100mM sodium phosphate, pH7.0, 1M NaCl, 2mM β-mercaptoethanol, 10% glycerol) until the A_{280} value of the flowthrough reached 0.02. The protein was eluted with a buffer containing 20mM acetic acid/acetate, pH5.0, 300mM NaCl, 2mM β-mercaptoethanol and 10% glycerol. The eluted protein was then dialysed into a buffer containing 20mM MES, pH6.0, 1mM TCEP, 300mM NaCl and 10% glycerol. The presence of His_9-$RHAU_{FL}$ in the various eluates was verified by SDS-PAGE on an 8% polyacrylamide gel.

The final purification method was based on anion exchange chromatography where two 1mL Hi-Trap Sepharose CaptoQ columns were used in tandem. The supernatant of the cell lysate was applied to the columns equilibrated in a buffer containing 20mM Bicine-NaOH, pH9.0. The resin bound proteins were eluted with a linear salt gradient of 0 - 500mM NaCl over 12.5 column volumes followed by a 0.5M - 1M NaCl step over 3.5 column volume, using an ÄKTA™ (GE Healthcare Life Sciences) FPLC system.

The His_6-$DDX21_{C209}$ (referred later as $DDX21_{C209}$) protein was purified following a previously described protocol [53] by affinity chromatography using HisPur™ Ni-NTA resin with the following modifications. The 4°C thawed cells taken out from -80°C storage was resuspended using Dounce homogenization in the previously described [53] 30mL ice cold Lysis Buffer pH7.5 (50mM Tris-HCl, 150mM NaCl, 500mM KCl, 1mM PMSF, 5mM DTT, containing 1mM PMSF). The cells were lysed by sonication (Sonic Dismembrator Model 500, Fisher Scientific) as previously described [53] but for a total run time of 12min. In this modified purification protocol, the lysed cells were centrifuged at 22000rpm for 50min at 10°C before applying to 2mL of the HisPur Ni-NTA resin and purifying as described. The eluate was immediately dialysed into 50mM Tris-HCl, pH7.5, 300 mM KCl (final buffer) using a 6-8 kDa molecular weight cut off (MWCO) membrane (Spectra/Por® - Spectrum Labs). Following the dialysis, the protein in the 300 mM KCl buffer was centrifuged at 14100rpm for 15min and the supernatant was stored at 4°C. The extinction coefficient ($\varepsilon_{280\ nm}$=19605$M^{-1}cm^{-1}$) for His_6-$DDX21_{C209}$ was calculated using the ExPASy ProtParam web-tool and the concentration of the protein was determined spectrophotometrically.

The purification of His_8-$DDX21_{FL}$ was attempted following a previously described protocol [53] for the His_6-$DDX21_{C209}$ protein purification.

2.4 Western Blotting

In order to determine the success of the RHAU protein expression in the induction study for the His_9-$RHAU_{FL}$ and His_9-$RHAU_{1-1008}$ constructs, and the presence of His_9-$RHAU_{FL}$ during its purification by affinity chromatography, size exclusion chromatography and ion-exchange chromatography, Western blots were performed using a mouse α-RHAU primary antibody (12F33), and a secondary α-mouse antibody as described. The protein samples were denatured by sodium dodecyl sulfate by heating at 95°C for 5mins prior to loading onto an 8% polyacrylamide gel which was run at 7A for 90mins. The protein bands were then transferred from the gel to a nitrocellulose membrane soaked in 10mM CAPS, pH11.0, 10% v/v methanol using a semi-dry electroblotting apparatus, with a current of 0.11A applied for a duration of 45mins. Following the transfer step, the membrane was blocked with Phosphate Buffered Saline supplemented with 0.1% Tween 20 (PBS-T) and 5 % (w/v) Coffee Whitener. After an hour of incubation at room temperature, the membrane was washed using PBS-T buffer for three consecutive 10mins intervals. Next, the nitrocellulose membrane was incubated with mouse α-RHAU primary antibody at a dilution of 1:5000 for an hour in blocking solution. The membrane was then washed for 3x10mins with PBS-T + 5% coffee whitener, before applying the horseradish peroxidase (HRP) conjugated anti-mouse secondary antibody in blocking solution. The membrane was washed for a final time with PBS-T, before incubating with 1mL Luminata Forte Western HRP substrate (EMD Millipore, USA) for a minute at room temperature and imaging using a fluorescent detector (FluorChem Q, Cell Biosciences).

In order to detect the presence of histidine tag in the RHAU protein, the above described method was followed except that a single anti-6x His tag primary antibody fused to HRP (ab18184, Abcam) was used.

The success of His_8-$DDX21_{FL}$ purification by affinity chromatography was checked with Western blot following the method described above for RHAU proteins except that using a mouse primary antibody raised against the C-terminus of human DDX21, and an HRP conjugated anti-mouse secondary antibody was used.

2.5 G-Quadruplex Preparation

The DNA equivalent of the human telomeric (hTel) sequence of 5'-Cy5-d(AGGGTTAGGGTTAGGGTTAGGG)-3' synthesized for the McKenna laboratory was used for the studies. The sequence was dissolved in a buffer containing 50mM Tris-HCl, pH7.5 and 300mM KCl at a concentration of 10mM, heated at 95°C for 3min followed by slow cooling to room temperature. The solution was nutated for 1min to obtain hTel in its quadruplex form. The molar extinction coefficient of the hTel dimer ($\varepsilon_{260\ nm}$=210600$M^{-1}cm^{-1}$) was used to calculate the concentration of the folded G4 hTel using IDT SciTools® (Primer-Quest Program, Integrated DNA Technologies).

The telomeric repeat containing RNA (TERRA) sequence of 5'-Cy5-r(AGGGUUAGGGUUAGGGUUA GGG)-3' was also obtained from the McKenna laboratory. In order to form G-quadruplexes the sequence was processed similarly as described before for hTel. The molar extinction coefficient of the TERRA dimer ($\varepsilon_{260\ nm}$=210000$M^{-1}cm^{-1}$) was used to calculate the concentration of the folded G4 TERRA.

2.6 Preparation of Protein and Protein-G4 Complexes Prior to Crystallization Trials

$DDX21_{C209}$ protein in a buffer containing 50mM Tris-HCl, pH7.5 and 300mM KCl was concentrated using a 3kDa molecular cut-off (MWCO) concentrator to 4.2mg/mL and 5.2mg/mL, and left at room temperature for two separate crystallization trials. $DDX21_{C209}$ protein was similarly concentrated to 1mg/mL and combined with human telomeric (hTEL) DNA G4 in a 2:1

molar ratio by very slow pipetting followed by nutating in a buffer containing 50mM Tris-HCl, pH7.5 and 300mM KCl. Following the mixing, the DNA-protein complex was further concentrated using the 3kDa concentrator to 7.9mg/mL, and left at room temperature for the crystallization trial. Before addition into the protein solution the concentration of the DNA was measured spectrophotometrically using the absorption coefficient ($\varepsilon_{260\ nm}$=210600$M^{-1}cm^{-1}$). The final concentration of the protein was measured using the *DC* Protein Assay (Bio-Rad). In a separate attempt to further purify the concentrated DNA-protein complex by SEC, the protein only sample and the protein-G4 complex with protein concentration of 1mg/mL was passed through a Superdex 200 10/300 GL column (GE Healthcare) in 50mM Tris-HCl, pH7.5, 300mM KCl run at 0.4mL/min. SEC fractions were analyzed by SDS-PAGE on an 8% polyacrylamide gel.

2.7 Circular Dichroism (CD) Spectropolarimetry

Circular dichroism spectra were collected by Ewan McRae on a J-810 spectropolarimeter (Jasco Inc., USA) using a 0.1cm quartz cell (Hellma) following a previously described protocol [53]. However, the sample concentration in the cell was 0.7mg/mL, the average spectrum of the buffer was subtracted from the average spectrum of the sample, and the buffer condition for the $DDX21_{C209}$ protein was 10mM phosphate, pH7.5, 150 mM NaF.

2.8 Isothermal Titration Calorimetry (ITC)

ITC for the $DDX21_{C209}$ protein - TERRA G4 complex was performed using a MicroCal ITC200 (GE Healthcare) system following a previously described protocol [53] by Ewan McRae. However, the sample was prepared only in 50mM Tris-HCl, pH 7.5, 100mM KCl where the protein and the G4 were combined in a 3:1 molar ratio and centrifuged for 10min at 14,100rpm prior to the experiment, the binding experiment was performed at room temperature. Control experiment was performed by TERRA G4 injection into the buffer to obtain the heat effects of RNA dilution.

2.9 Thermal Shift Assay (TSA)

The thermal shift assay (Thermo-Fisher Scientific Ottawa, ON, Canada) for the $DDX21_{C209}$ protein was performed following the manufacturer's recommended protocol (1× PTS buffer and 1× PTS dye) and a previously described protocol [53]. Fluorescence was measured using a StepOnePlus real-time PCR system (ThermoFisher Scientific Ottawa, ON, Canada) with the existing ROX spectral settings. Data with the slowest temperature increase of 1% was recorded for every 1°C interval. The StepOne-Plus software was used to calculate the first derivative of the fluorescence data over temperature, which was normalized to the largest slope value, and was further used to determine the melting temperature of the protein.

2.10 Crystallization Trials

Crystallization trials with $DDX21_{C209}$ and hTel DNA G4·$DDX21_{C209}$ complex were separately conducted using commercial crystallization kits from Hampton Research (CA, USA). The protein sample, G-quadruplex and the protein-G4 complex were prepared as described in the preceding sections of this thesis. Conditions for crystallization were tested by using the vapour diffusion technique. Sitting drops were prepared taking 0.6µL of protein/protein-G4 complex and adding to 0.6µL of reservoir buffer solution. The reservoir solution volume in each well was 50 µL. The plates were covered and sealed with clear protective film and incubated at 20°C. The PEG based JBScreen 1 – 5, ammonium sulfate based JBScreen 6, MPD based JBScreen 7 and MPD – alcohol based JBScreen 8, as well as Crystal Screen HT, JBScreen JCSG ++1 – ++4 and Index™ HR2 - 144 screening kits (Jena Bioscience Hampton Research, CA, USA) were used to test possible crystallization conditions for $DDX21_{C209}$ at 4.2mg/mL and 5.2mg/mL protein concentrations, and at 7.9mg/mL protein concentration for the hTel DNA G4·$DDX21_{C209}$ complex. Oligonucleotide concentration within the complex was spectrophotometrically determined

following the protein concentration determination by *DC* Protein Assay (Bio-Rad). The crystal plates were observed daily for the first week, every three days for the next week, weekly for the subsequent month and then monthly for the remainder of the time.

3. RESULTS

In an attempt to perform biophysical studies of the RHAU protein, expression of His_9-$RHAU_{FL}$ which is 114.8kDa was attempted in *E. coli* C41(DE3) and *E. coli* BL21(DE3), whereas expression of His_9-$RHAU_{51-1008}$ construct lacking the glycine-rich N-terminal 50 amino acids, which is 110kDa was performed only in *E. coli* BL21(DE3).

3.1 Confirmation of Successful Cloning of His_9-$RHAU_{FL}$ Plasmid

Potential clones of the His_9-$RHAU_{FL}$ were confirmed to contain an insert of the correct size by performing a double restriction digestion with *Nde* I and *Eco* RI followed by electrophoresis on a 1% agarose gel. pET21a(+) has a size of 5.3kb and the His_9-$RHAU_{FL}$ insert has an expected size of 3.1kb. Looking at Figure 7 it is clear that an insert of the correct size for His_9-$RHAU_{FL}$ is present in each of the three clones tested.

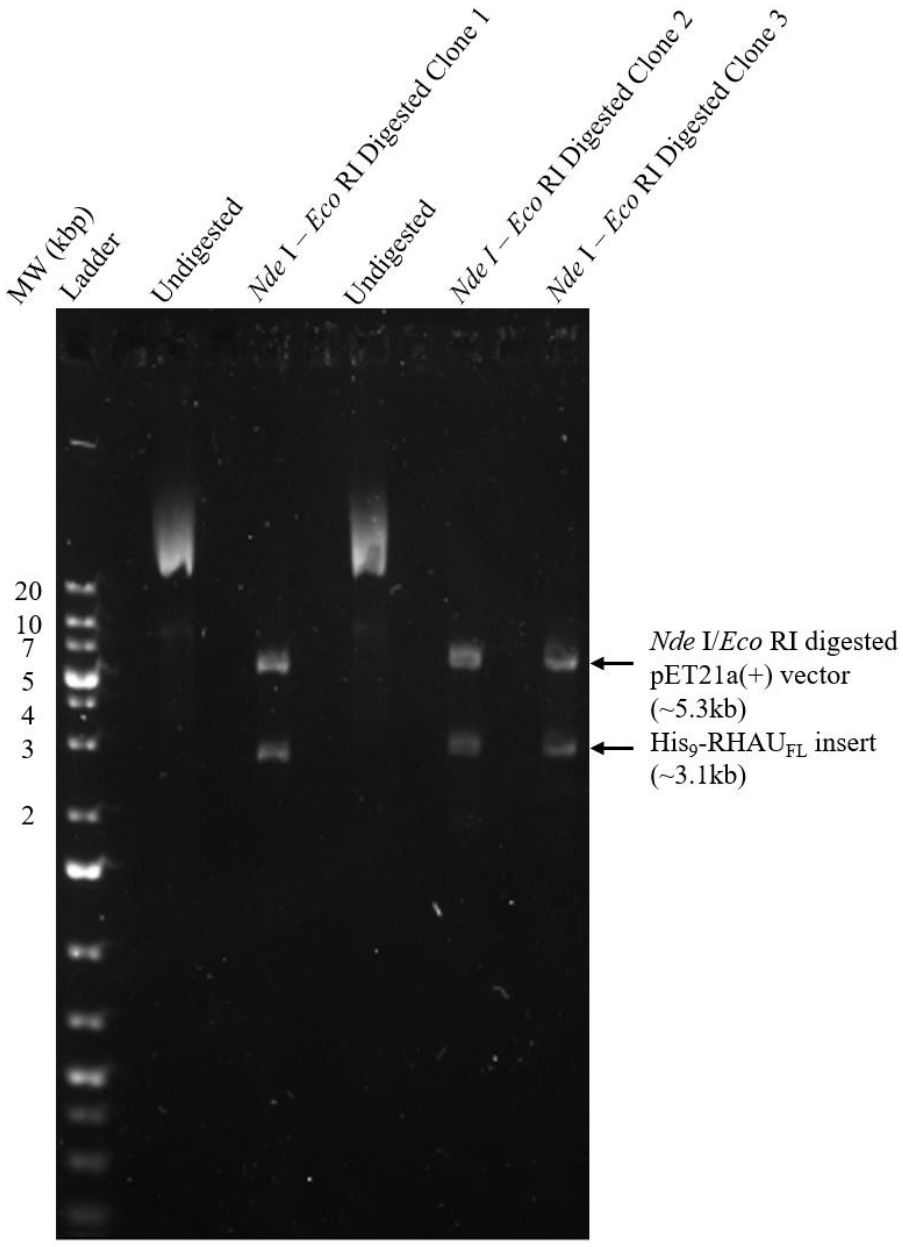


Figure 7. Agarose gel electrophoresis assay. Undigested and *Eco* RI - *Nde* I double digested samples of His_9-$RHAU_{FL}$ plasmid were run on 1% gel. The plasmid DNAs were stained with Hydra Green.

3.2 Optimization of His_9-$RHAU_{FL}$ and His_9-$RHAU_{51-1008}$ Protein Expression Conditions

The expression of His_9-$RHAU_{FL}$ was performed in the *E. coli* strain C41(DE3), with samples taken at 0, 1, 2, 3, 4 and 5hr after induction to monitor the expression level. Analysis of the samples by SDS-PAGE was inconclusive (Figure 8A), but a band running immediately above the 100 kDa marker was observed on the Western blot (Figure 8B) suggesting that the His_9-$RHAU_{FL}$ was produced as expected. The Western blot data of the same samples further revealed that His_9-$RHAU_{FL}$ expression increased continuously up to 4 hours after IPTG induction (Figure 8B), and so in all further experiments the His_9-$RHAU_{FL}$ protein was expressed for 4 hours after induction before harvesting.

In order to further optimize the expression protocol for His_9-$RHAU_{FL}$, the concentration of IPTG, the point of induction, and the temperature for the expression were also altered. For each experiment the results were assessed by SDS-PAGE and Western blot analysis (Figure 9A-9E). As a result of this experiment series we concluded that His_9-$RHAU_{FL}$ expression was optimal using the *E. coli* strain C41(DE3), with induction at a cell density OD_{600} of 0.5, 0.5mM IPTG and a growth temperature of 20°C (Figure 9B). It was observed that the RHAU protein was induced but slightly less expressed at an OD_{600} of 0.7, 0.5 mM IPTG and a growth temperature of 28°C (Figure 9C). In case of expression induction at an OD_{600} of 2, 0.5mM IPTG and a growth temperature of 20°C RHAU protein was not expressed (Figure 9A). The expression of His_9-$RHAU_{FL}$ in *E. coli* strain BL21 (DE3) was negligible (Figure 9D-E).

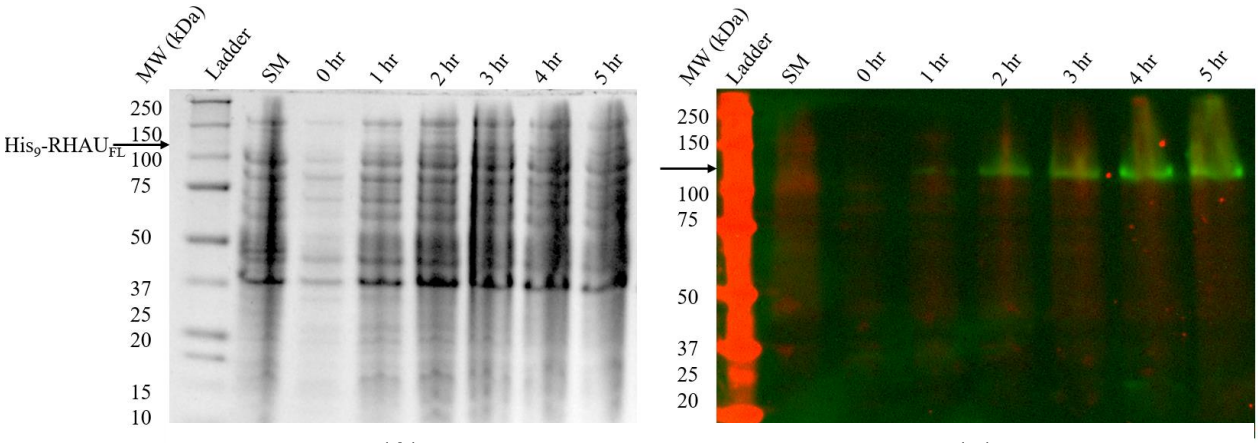
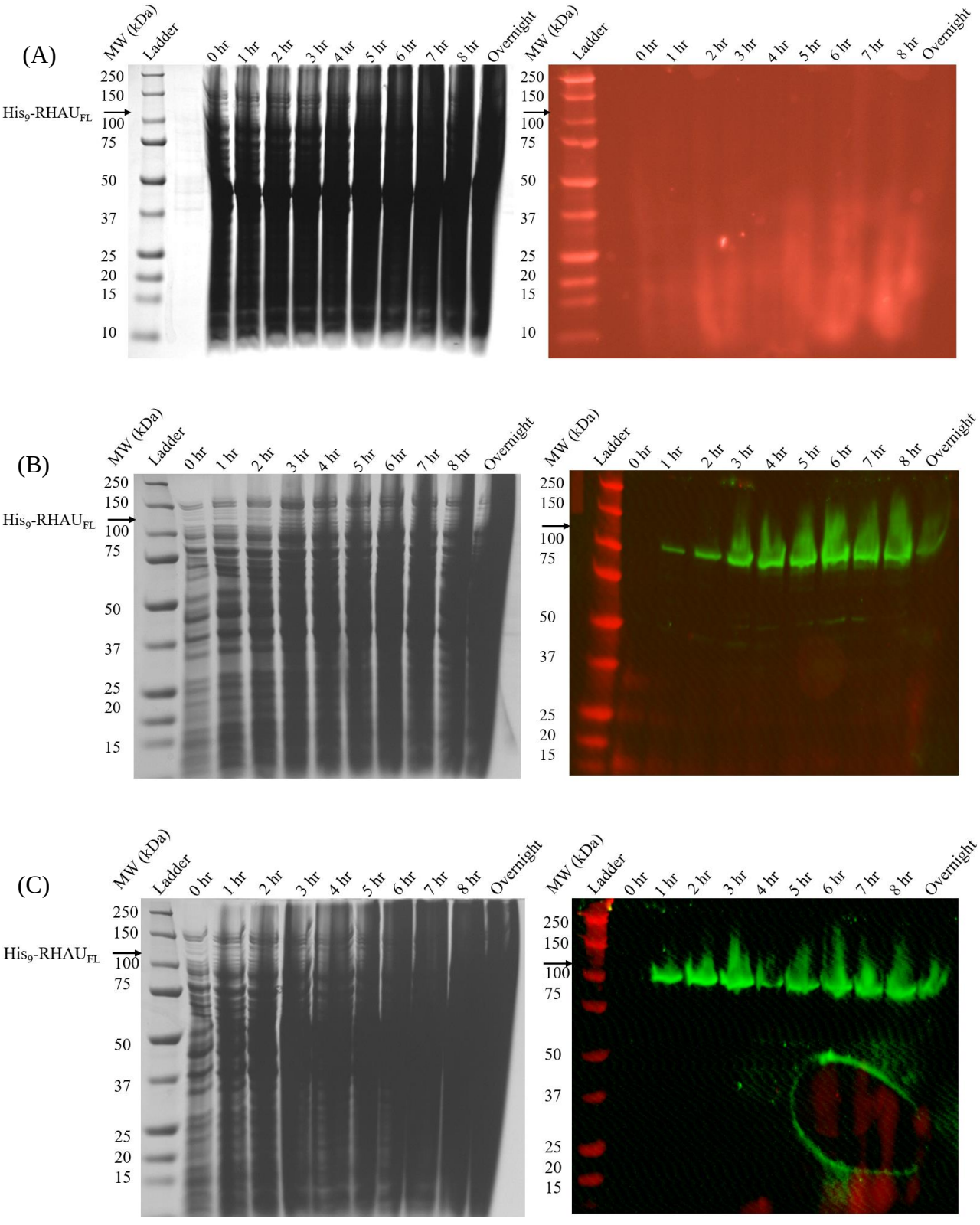


Figure 8. SDS-PAGE and Western blot analysis of the His_9-$RHAU_{FL}$ protein expression in *E. coli* C41(DE3) cells. Cell samples grown at 20°C were collected at regular intervals after 0.5mM IPTG induction at OD_{600} of 0.5, along with a pre-culture/starting material (SM) sample. The collected samples were processed and run on 8% polyacrylamide gel to perform (A) SDS-PAGE and (B) Western blot using an α-RHAU primary antibody for detection of the RHAU protein.


(A)
MW (kDa)
Ladder
0 hr
1 hr
2 hr
3 hr
4 hr
5 hr
6 hr
7 hr
8 hr
Overnight
250
150
100
75
50
37
25
20
15
10
His9-RHAUFL
(B)
(C)

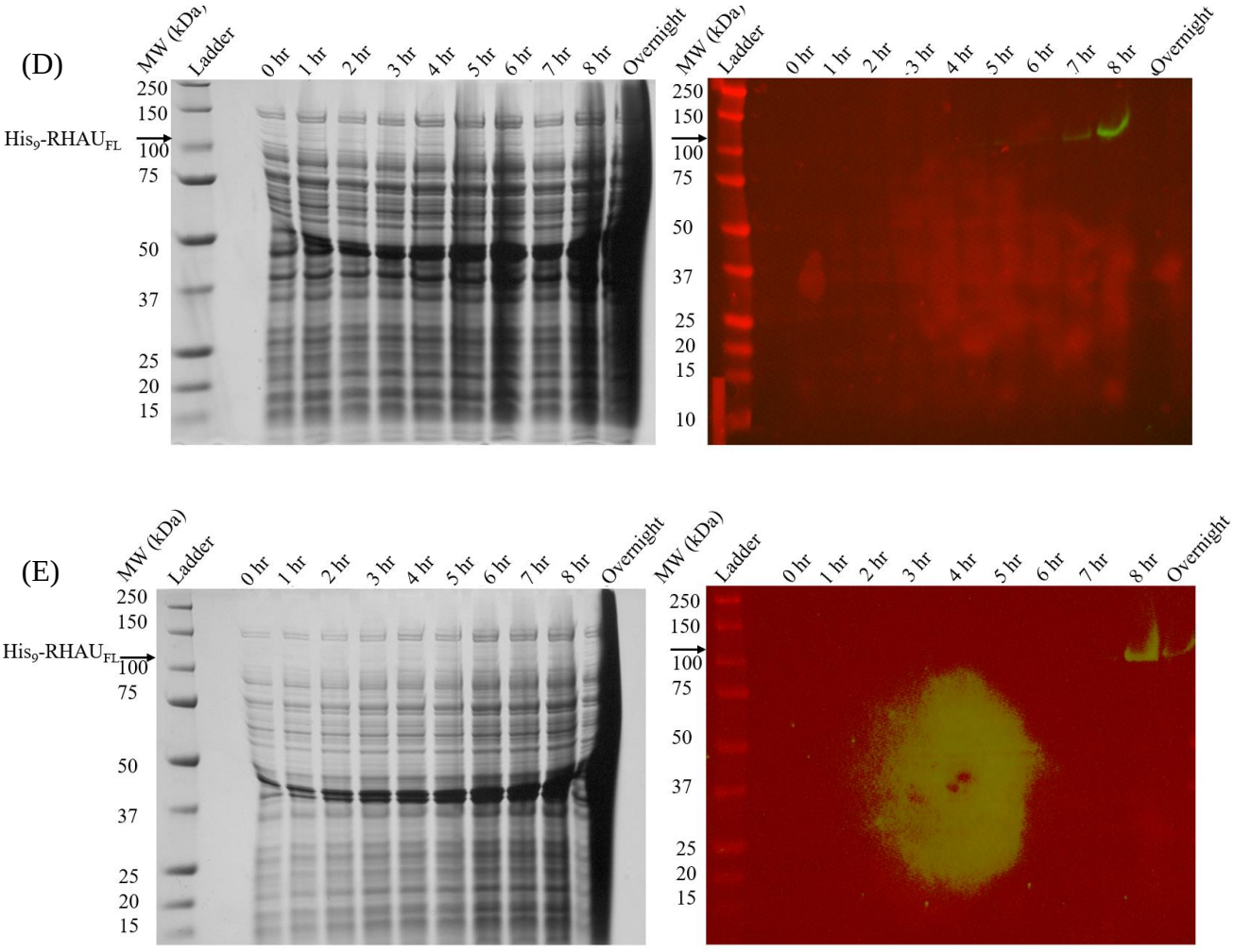
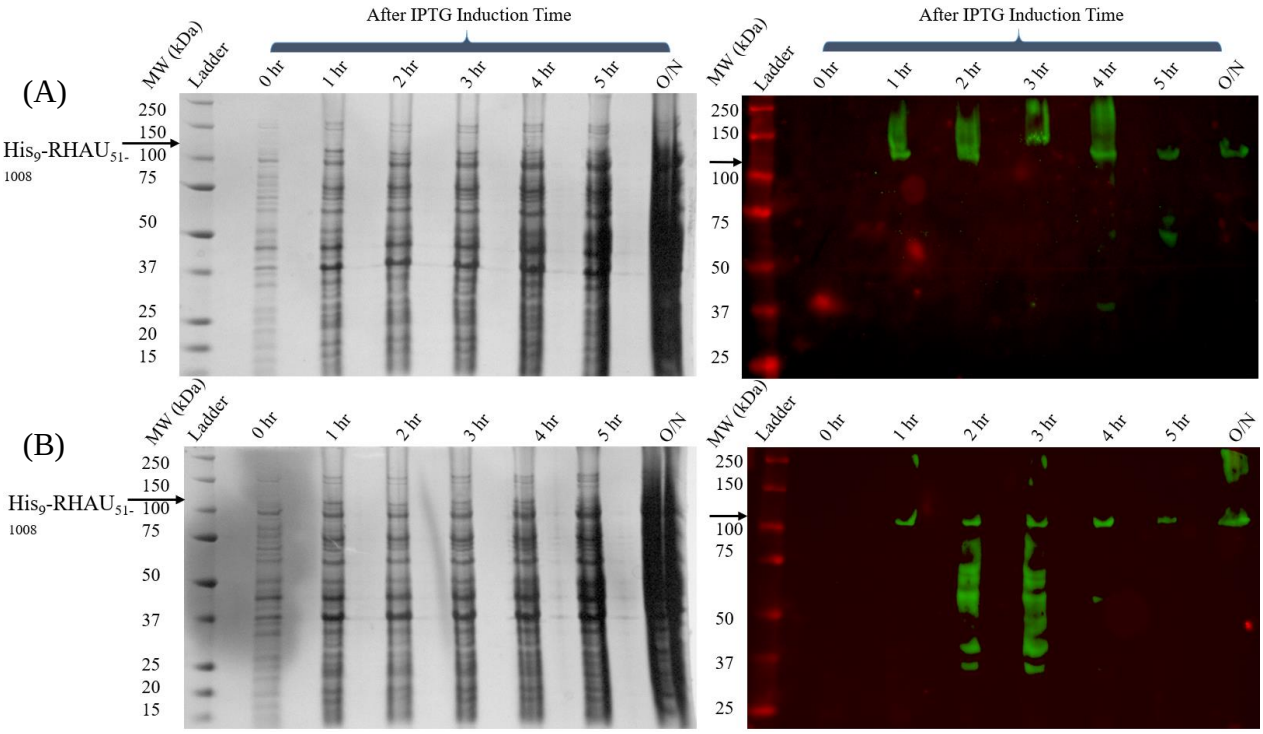


Figure 9. SDS-PAGE and Western blot analysis of the His_9-$RHAU_{FL}$ protein expression in *E. coli* strains. Expression of the protein in C41(DE3) cells were performed with (A) 0.5 mM IPTG induction at OD_{600} of 2.0 and 20°C growth temperature, (B) 0.5 mM IPTG induction at OD_{600} of 0.5 and 20°C growth temperature, (C) 0.5 mM IPTG induction at OD_{600} of 0.7 and 28°C growth temperature. Expression of the protein in BL21(DE3) cells with (D) 0.01mM IPTG induction at OD_{600} of 0.4 and 16°C growth temperature, and (E) 0.01mM IPTG induction at OD_{600} of 0.7 and 16°C growth temperature. Samples collected at regular interval were processed and run on 8% polyacrylamide gel to perform both SDS-PAGE (images on the left) and Western Blot (images on the right) using α-RHAU primary antibody.

The expression of His_9-$RHAU_{51-1008}$ was performed in the *E. coli* strain BL21(DE3), with samples taken at 0, 1, 2, 3, 4 and 5hr after induction to monitor the expression level. Analysis of the samples by SDS-PAGE was inconclusive (Figure 10) but a band running immediately above the 100kDa marker was observed in the Western blot (Figure 10) suggesting that the His_9-$RHAU_{51-1008}$ was produced as expected. In an attempt to optimize the His_9-$RHAU_{51-1008}$ expression protocol, the concentration of IPTG, the point of induction, and the temperature for the expression were also altered. For each experiment, the results were assessed by SDS-PAGE and Western blot analysis (Figure 10). As a result of the experiment series, we concluded that His_9-$RHAU_{51-1008}$ expression was optimal when cells were induced at an OD_{600} of 0.7, by 2mM IPTG, and a growth temperature of 28°C was maintained (Figure 10D). His_9-$RHAU_{51-1008}$ expression was comparatively less when cells were induced at an OD_{600} of 0.4 with 0.5mM and 2mM IPTG, and a growth temperature of 37°C was maintained (Figure 10A-B). The His_9-$RHAU_{51-1008}$ protein expression was the least when cells were induced at an OD_{600} of 0.7 with 0.5mM IPTG, and at a growth temperature of 28°C (Figure 10C).


(A)
After IPTG Induction Time
MW (kDa)
Ladder
0 hr
1 hr
2 hr
3 hr
4 hr
5 hr
O/N
250
150
100
75
50
37
25
20
15
His9-RHAU51-1008
(B)

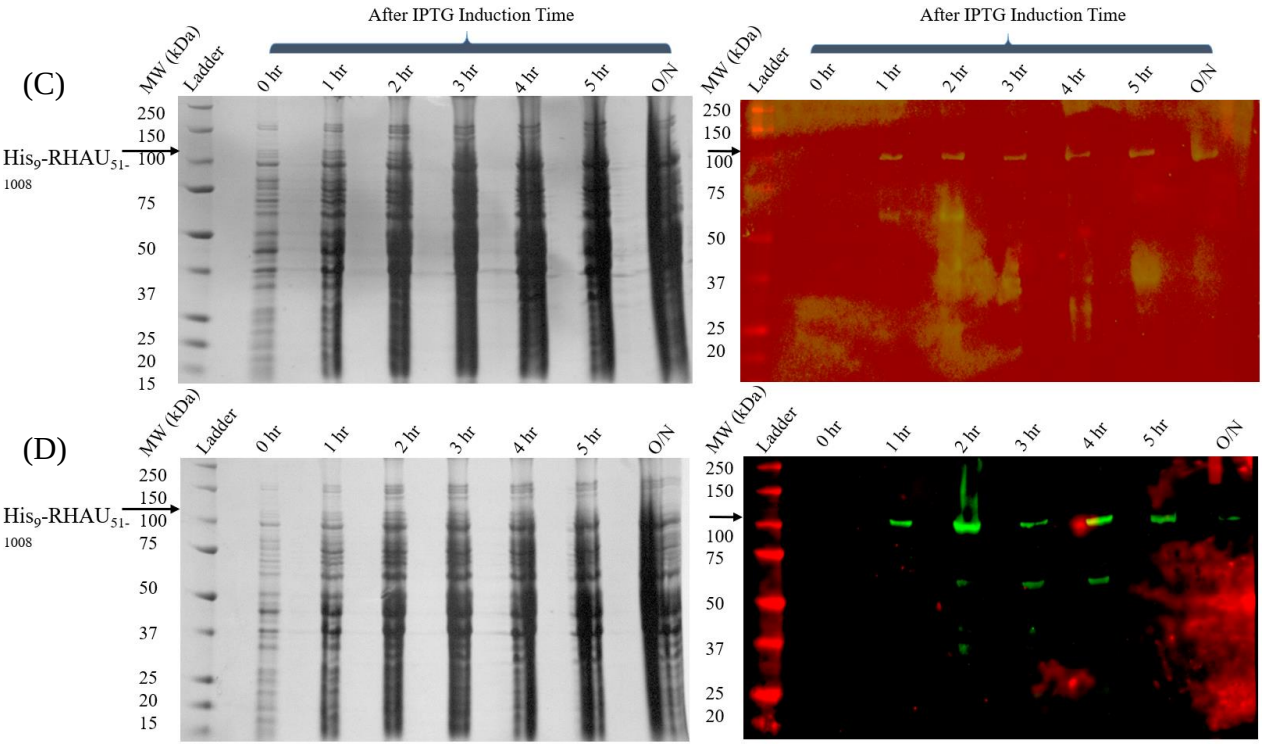
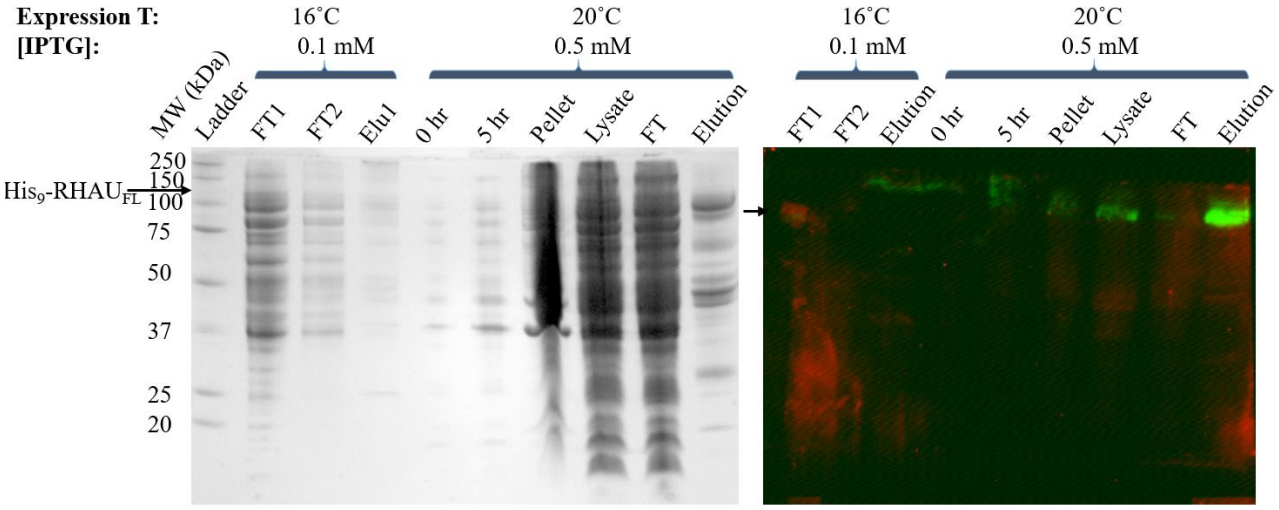


Figure 10. SDS-PAGE and Western blot analysis of the His_9-$RHAU_{51-1008}$ protein expression in *E. coli* strain BL21(DE3). Protein expression was performed with (A) 0.5mM IPTG induction, OD_{600} of 0.4 and 37°C expression temperature, (B) 2mM IPTG induction, OD_{600} of 0.4 and 37°C expression temperature, (C) 0.5mM IPTG induction, OD_{600} of 0.7 and 28°C expression temperature, and (D) 2mM IPTG induction, OD_{600} of 0.7 and 28°C expression temperature. Samples collected at regular interval were processed and run on 8% polyacrylamide gel to perform both SDS-PAGE (images on the left) and Western Blot (images on the right) using α-RHAU primary antibody.

3.3 Optimization of His_9-$RHAU_{FL}$ Protein Purification Conditions

The initial attempt at purification was made by affinity chromatography using cobalt resin beads following the first described purification protocol involving a buffer containing 20mM sodium phosphate, pH7.0. The first purification attempt was made from 16˚C grown cells, induced with 0.01mM IPTG, where the RHAU protein was absent in both the flowthrough, and almost undetectable in the imidazole buffer eluate (Figure 11). All subsequent attempts at RHAU purification was made from optimized 20˚C grown cells as described in the Material and Methods section. In the second purification attempt, RHAU protein was detected in both the cell lysate, and in the imidazole buffer eluate (Figure 11). Some RHAU protein was also detected in pellet fraction, suggesting that the cells might not have been fully lysed (Figure 11). As suggested by the SDS-PAGE data (Figure 11A), the purified eluate contained a number of polypeptide species that co-purified with the RHAU protein. Moreover, the RHAU eluate was precipitating at a concentration higher than 0.1mg/mL. In order to further purify the RHAU protein, the cobalt resin purified eluate was concentrated up to 1mg/mL as estimated spectrophotometrically at 280nm, and a 350µL sample was loaded onto a Superdex 200 10/300 GL SEC column. The major peak eluted in the void volume together with multiple smaller peaks (Figure 12A). SDS-PAGE analysis of the column fractions showed that bands corresponding to the His_9-$RHAU_{FL}$ protein were found mostly in the void volume fraction (Figure 12B). That His_9-$RHAU_{FL}$ was present in a void volume fraction of SEC suggests that the His_9-$RHAU_{FL}$ protein was aggregated under these conditions. The SDS-PAGE band pattern looked very clean for the void volume peak sample compared to the cobalt-affinity column eluate. In order to find out if the His_9-$RHAU_{FL}$ protein could be separated from the other polypeptide species present in the void volume sample, the following methods were employed. In an attempt to find out if the His_9-$RHAU_{FL}$ (114.8kDa) protein was loosely bound



(A) (B)

Figure 11. SDS-PAGE and Western blot analysis of the His_9-$RHAU_{FL}$ protein purification from *E. coli* strain *C41(DE3)* grown both at 16°C and 20°C after IPTG induction. Along with purification fractions, 0^{th} hr and 5^{th} hr cell samples for 20°C induced were also collected. The samples were processed and run on 8% polyacrylamide gel to perform both (A) SDS-PAGE and (B) Western Blot using α-RHAU primary antibody.

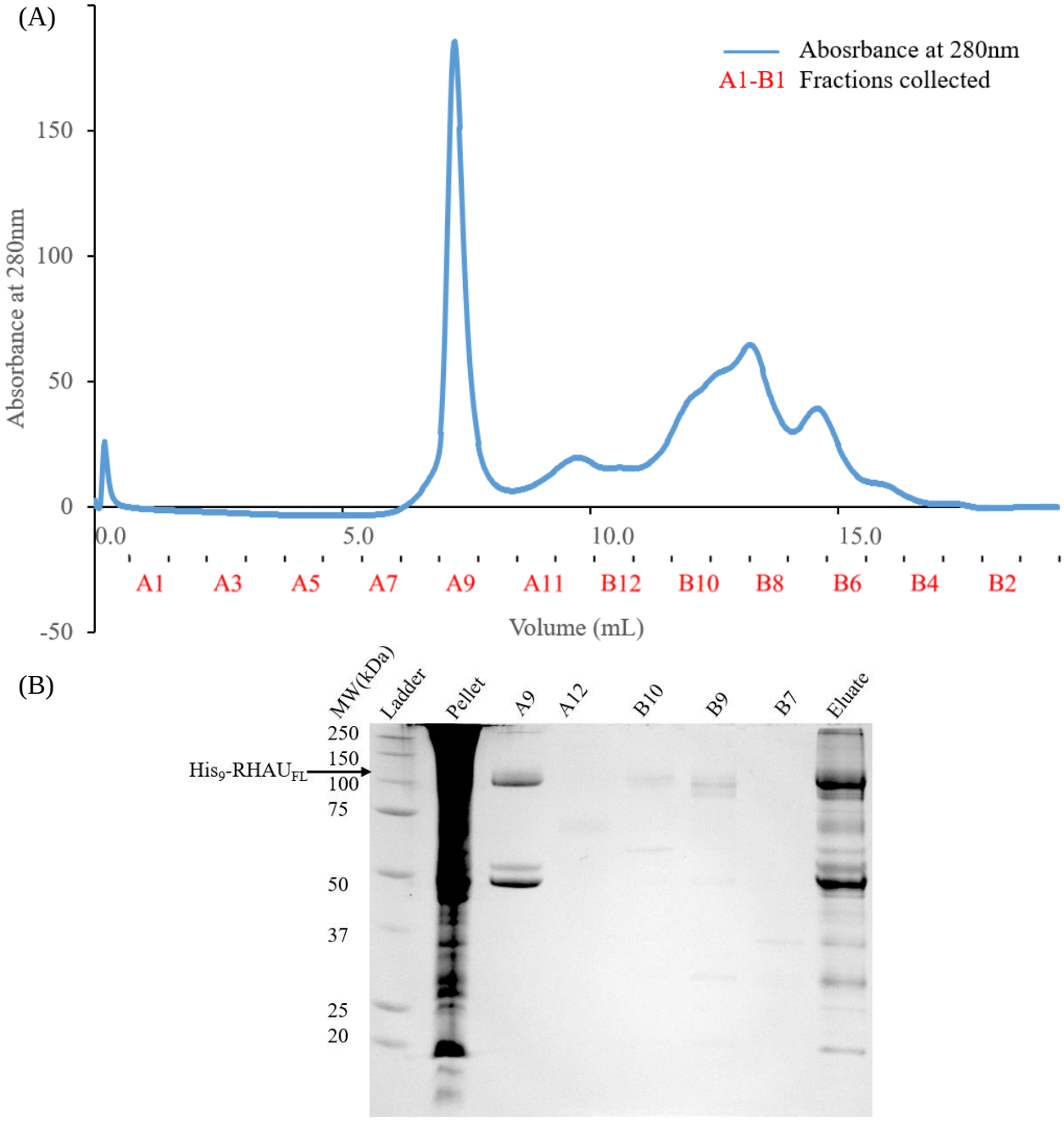
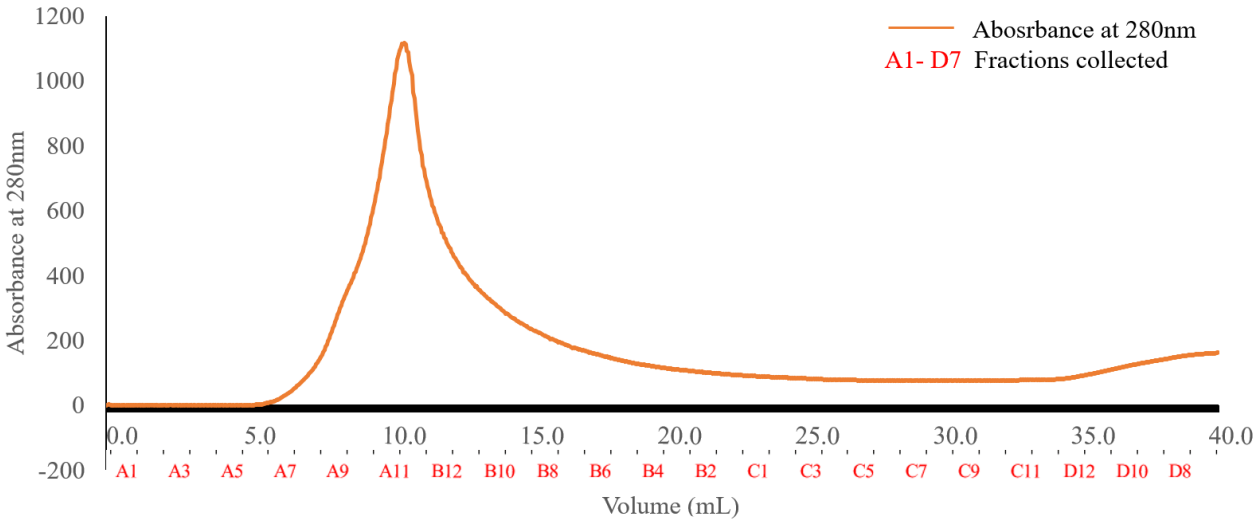


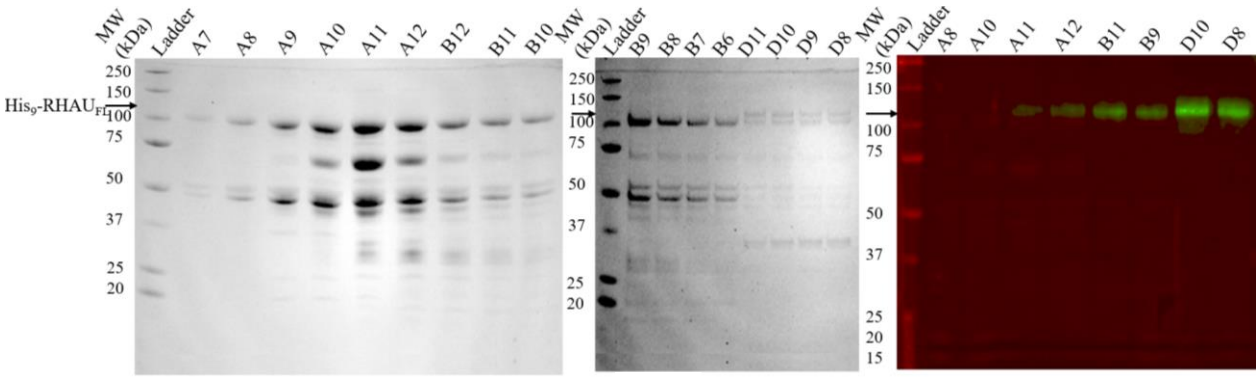
Figure 12. Size exclusion chromatography and SDS-PAGE analysis of cobalt affinity-SEC purified His_9-$RHAU_{FL}$ protein samples. (A) A SEC profile of the proteins eluted from the Superdex 200 10/300 GL column. (B) SDS-PAGE analysis on 8% polyacrylamide gel with the samples of SEC eluates (A9, A12, B10, B9 and B7), cobalt affinity eluate (Eluate), and the cell pellet (Pellet) obtained after cell lysis.

with the other polypeptide species present in the void volume sample using non-covalent forces, the eluate was passed through a 100kDa filter. No protein was found in the eluate filtrate and the concentrate revealed the same denaturing gel pattern as the void volume eluate, which indicates that non-covalent forces were not responsible for the interactions between the void volume polypeptide species. To check the possibility of disulfide linkages functioning as the binding force between His_9-$RHAU_{FL}$ (16 Cys residues present in the protein sequence) and other polypeptide species in the void volume, the eluate was treated with β-mercaptoethanol (β-ME). However, this also failed to separate His_9-$RHAU_{FL}$ from the other polypeptide species. To evaluate whether the binding interaction between the polypeptide species was because of the nucleic acid binding potential of the His_9-$RHAU_{FL}$ protein, the presence of DNA and/or RNA in the void volume sample was spectrophotometrically determined by measuring the ratio between absorbances at 260nm (A_{260}) and 280nm (A_{280}) UV light wavelength. However, A_{260}/A_{280} ratio of 0.61 indicated absence of nucleic acid contamination. The results of the above described experimental approaches suggested that the void volume proteins were aggregated.

So far, the purification of the tagged $RHAU_{FL}$ protein, was attempted from crude extracts of cell lysate using metal affinity chromatography, where imidazole competes for the 9-His tag of His_9-$RHAU_{FL}$. Imidazole is known to cause stability issues with certain proteins [132] and so in an attempt to minimize this potential effect on His_9-$RHAU_{FL}$ protein, the minimum required concentration of imidazole for the elution of His_9-$RHAU_{FL}$ from the cobalt affinity column was determined. The combined approach of SDS-PAGE and Western blot revealed that the His_9-$RHAU_{FL}$ protein was mostly purified in high imidazole concentration of 350mM (Figure 13B).



(A)



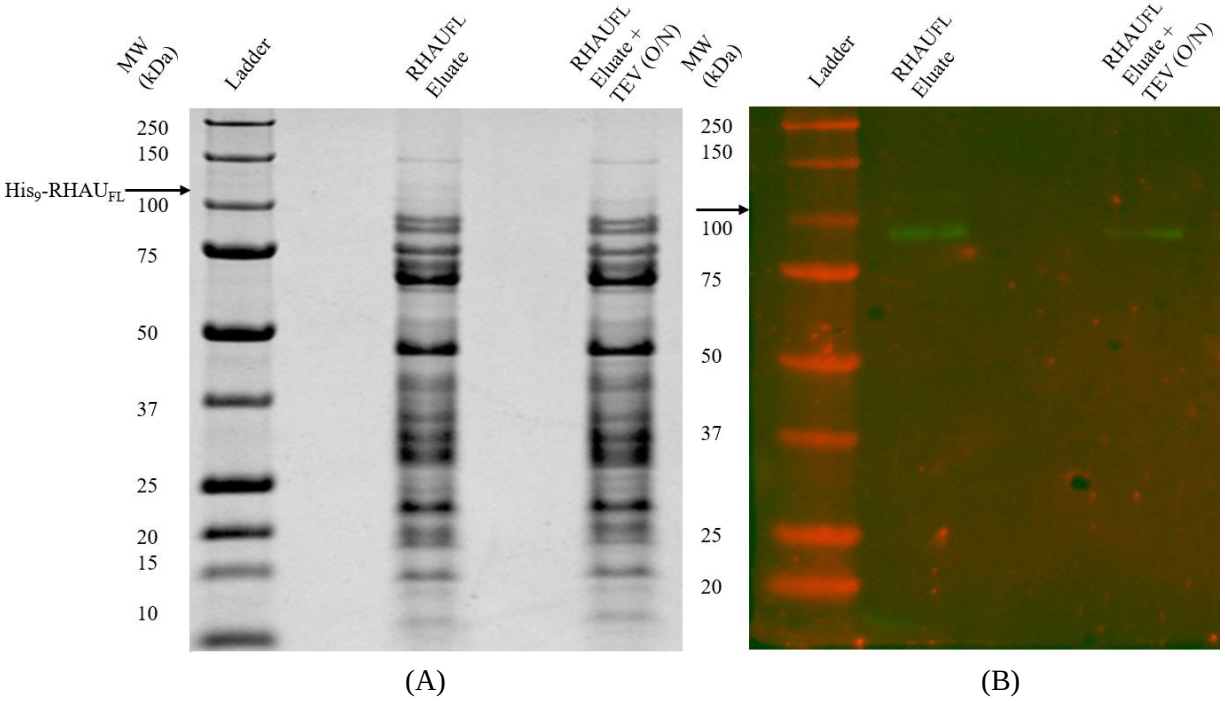
(B)

Figure 13. Imidazole gradient elution of the His_9-$RHAU_{FL}$ protein and SDS-PAGE – Western blot analysis of the eluted samples. (A) A protein concentration vs A_{280} profile of the proteins eluted from cobalt affinity column using increasing imidazole concentration of 0 – 500mM. (B) Eluted protein samples were run on 8% polyacrylamide gel to perform SDS-PAGE (images on the left) and Western blot (image on the right) using α-RHAU primary antibody.

In order to assess if it was the imidazole present in the elution buffer that was causing the aggregation of the His_9-$RHAU_{FL}$ protein, a new protocol with alternate buffers was explored to see if the His_9-$RHAU_{FL}$ could be purified more successfully. In the first experiment, immediately following the collection from the cobalt affinity resin beads, the eluate was exchanged into a buffer containing 20mM sodium phosphate, pH8.0, 300mM NaCl by dialysis. Although the pH8 phosphate buffer had a pH of 0.6 unit higher than the pI (7.42) value of the His_9-$RHAU_{FL}$ protein, the protein was highly precipitated. A second buffer change experiment in which the protein was exchanged into a buffer containing 20mM MES-NaOH, pH6.0 resulted in the complete precipitation of the protein. Interestingly, in the final buffer exchange experiment where the protein was exchanged into a buffer containing no salt but both DTT and EDTA (50mM Tris-HCl, pH8.5, 1mM DTT, 0.5mM EDTA), the precipitation of the eluted protein was significantly reduced.

In order to assess if the 9-His tag of the His_9-$RHAU_{FL}$ protein was affecting the protein stability, the removal of 9-His tag was attempted using tobacco etch virus (TEV) protease as His_9-$RHAU_{FL}$ and the 9-His affinity tag are separated by a TEV protease recognition sequence. In spite of being highly specific for its substrates TEV treatment was unsuccessful in cleaving the 9-His tag as confirmed on Western blot using α-His antibody (Figure 14B). The buffer used was 20mM HEPES, pH8.0, 300mM NaCl, 1mM TCEP which is compatible for TEV protease cleavage but the TEV cleavage did not work likely due to the aggregation of the protein rendering the TEV site inaccessible to the TEV enzyme.

In order to further minimize the precipitation of the eluates, and to obtain His_9-$RHAU_{FL}$ in the non-aggregated form in the cobalt affinity resin purification, a non-denaturing lysis and binding buffer was used with 20mM Tris-HCl, pH8.5, 300mM NaCl, 2mM imidazole hydrochloride, 2mM β-mercaptoethanol, 10% glycerol, where the buffer additive glycerol

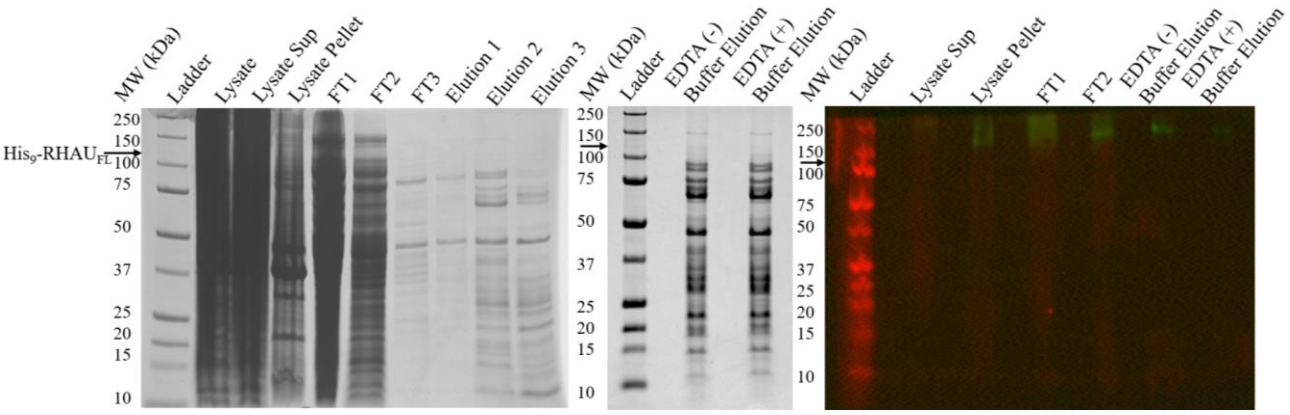


(A) (B)

Figure 14. SDS-PAGE and Western blot analysis of the TEV treated His_9-$RHAU_{FL}$ protein sample eluted from the cobalt affinity column. Protein samples before and after TEV treatment were run on 8% polyacrylamide gel to perform both (A) SDS-PAGE and (B) Western blot using α-His antibody.

enhances protein folding, a low concentration of 2mM imidazole helps remove non-specific proteins that could otherwise bind with the cobalt affinity resin. The wash buffer composition 20mM Tris-HCl, pH8.0, 300mM NaCl, 2mM imidazole hydrochloride, 2mM β-mercaptoethanol, 10% glycerol was used to remove unspecifically bound proteins from the cobalt resin. To avoid any issues with imidazole, elution of the His_9-$RHAU_{FL}$ protein was achieved by using a buffer containing 20mM acetic acid/acetate, pH5.0, 300mM NaCl, 2mM β-mercaptoethanol and 10% glycerol. His_9 - $RHAU_{FL}$ came off the column both in the wash buffer and the elution buffer as suggested by the SDS-PAGE and Western blot data (Figure 15). Surprisingly, the color of the eluate was changed from colorless to yellow sometime after performing the elution. This color change might have been caused by a small amount of cobalt leeching from the column into the sample which then reacted with β-mercaptoethanol. In an attempt to eliminate any possible cobalt contamination of the His_9-$RHAU_{FL}$ sample with cobalt, the pH eluted sample was immediately exchanged into a β-mercaptoethanol free, EDTA containing buffer of 20mM MES-NaOH, pH6.0, 1mM TCEP, 300mM NaCl, 1mM EDTA and 10% glycerol. In order to remove EDTA the protein was subsequently transferred to a buffer containing 20mM MES-NaOH, pH6.0, 1 mM TCEP, 300mM NaCl and 10% glycerol. The His_9-$RHAU_{FL}$ protein was found to be more stable as it precipitated less in the MES buffer of pH 6.0 than in the pH5 elution buffer where the protein precipitated highly even at 4 °C.

In order to purify higher yield of stable folded His_9-$RHAU_{FL}$, the next cobalt affinity purification protocol had the following changes: The cells were resuspended in a lysis buffer that was supplemented with RNase A to eliminate any RNA. The lysed cells were then treated with DNase I to remove any contaminating DNA. The ionic strength of the lysate was increased from 300mM to 500mM and β-mercaptoethanol was added as a reducing agent. During the cell lysis,

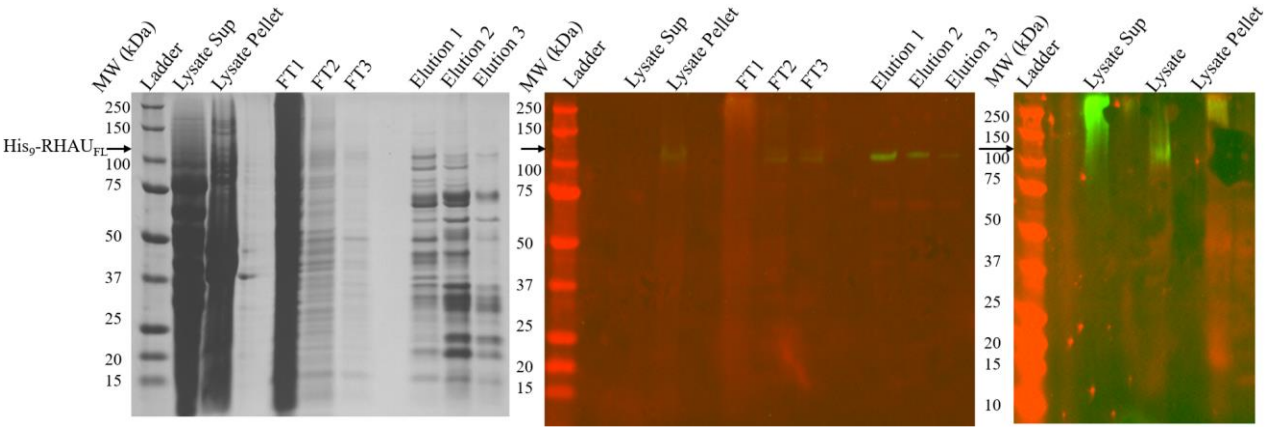


(A) (B) (C)

Figure 15. SDS-PAGE and Western blot analysis of cobalt affinity purification samples of the His_9-$RHAU_{FL}$ protein. Elution was performed by lowering pH. Purification samples were run on 8% polyacrylamide gel to perform both (A, B) SDS-PAGE and (C) Western blot using α-His antibody.

nucleic acid digestion and cobalt affinity resin binding steps, a non-denaturing buffer consisting of 20mM Tris-HCl, pH8.5, 500mM NaCl, 2.5mM $MgCl_2$, 0.1mM $CaCl_2$ and 10% glycerol was used. Furthermore, an absence of reducing agents and a low ionic strength were maintained to avoid inhibition of DNase I activity. Based upon the result of the earlier stability experiments a pH of 8.5 was chosen for this buffer in an attempt to lose less His_9-$RHAU_{FL}$ through precipitation of the sample. In order to remove non-specific proteins from the resin, the column was immediately washed with a buffer containing 20mM Tris-HCl, pH8.5, 500mM NaCl, 2mM β-mercaptoethanol and 10% glycerol. The resin was subsequently washed with wash buffer (100mM sodium phosphate, pH7.0, 1M NaCl, 2mM β-mercaptoethanol and 10% glycerol) to eliminate further non-specific binding of proteins to the cobalt resin. The elution was performed with 20mM acetic acid/acetate, pH5.0, 300mM NaCl, 2mM β-mercaptoethanol, 10% glycerol. Color change of the eluate was not observed by this method of purification, although His_9-$RHAU_{FL}$ was mostly precipitated as suggested by Western blot data (Figure 16). Therefore, the eluate was dialyzed into an EDTA-free buffer containing 20mM MES-NaOH, pH6.0, 1mM TCEP, 300 mM NaCl and 10% glycerol, where the protein precipitated less than before as observed visibly after centrifugation.

In order to remove unspecifically bound proteins on the basis of their net surface charge, anion exchange chromatography was employed where sepharose resin beads with the positively charged ligand, - $N(CH_3)^{3+}$ was used. His_9-$RHAU_{FL}$ and the other proteins in the cell lysate with a pI less than the pH9.0 of the sepharose resin-equilibration buffer (20 mM bicine-NaOH, pH 9.0), become negatively charged and so bind to the positively charged groups on the resin. The bound proteins were eluted with a NaCl salt gradient, with the eluate fractions analyzed by SDS-PAGE. This analysis suggested that there were multiple degradation products of His_9-$RHAU_{FL}$, as



(A) (B) (C)

Figure 16. SDS-PAGE and Western blot analysis of cobalt affinity purification samples of the His_9-$RHAU_{FL}$ protein. Elution was performed by lowering pH. Purification samples were run on 8% polyacrylamide gel to perform both (A) SDS-PAGE and (B, C) Western blot using α-His antibody.

suggested by the Western blot data (Figure 17). In the end of the purification experiment series, the conditions tested did not provide a pure sample of His_9-$RHAU_{FL}$. In order to focus on the DDX21 protein in the limited timeframe of this Master's thesis, further work on RHAU was discontinued.

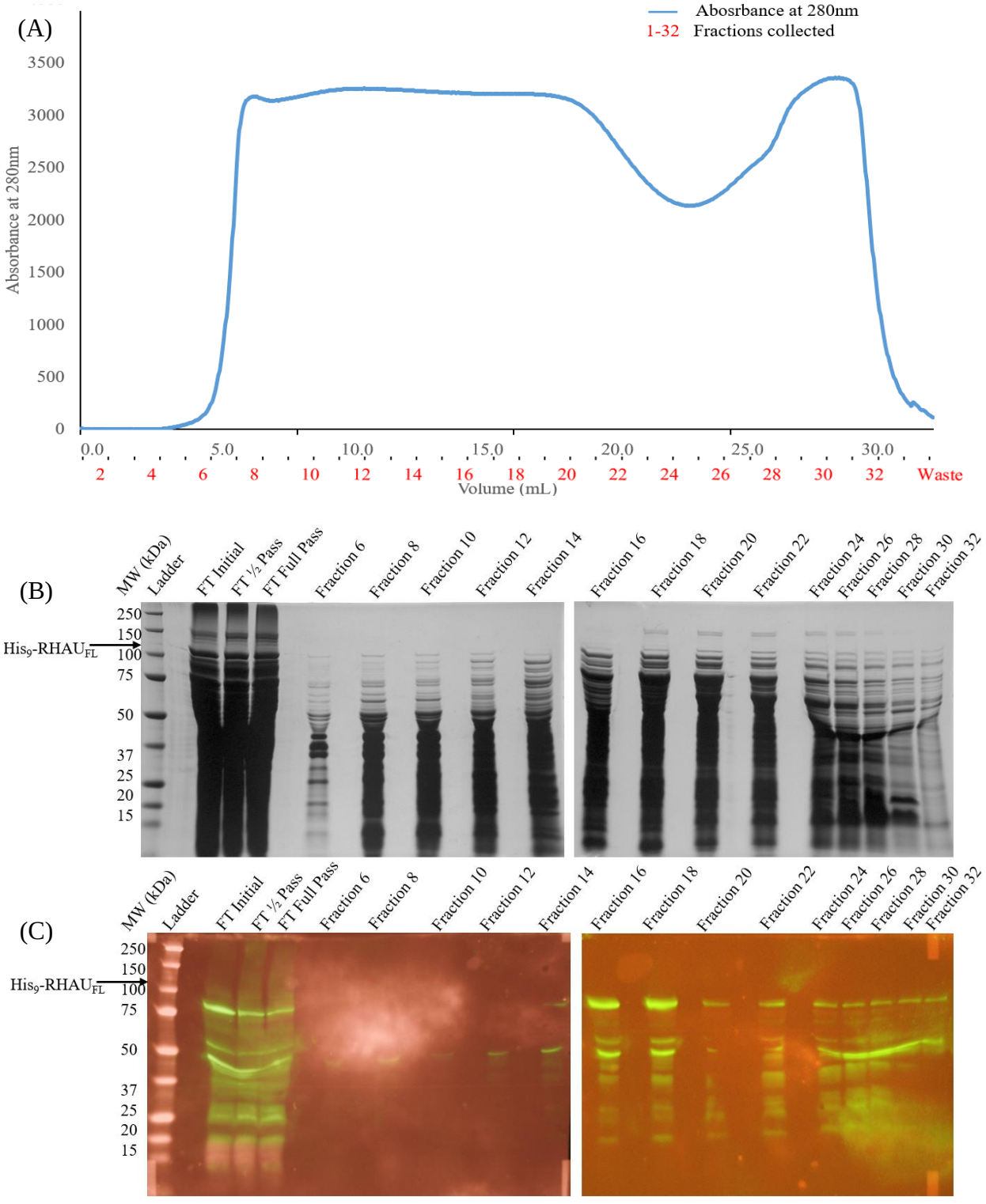


Figure 17. NaCl elution of the His_9-$RHAU_{FL}$ protein and SDS-PAGE – Western blot analysis of the eluted samples. (A) A protein concentration vs A_{280} profile of the eluted samples run on 8% polyacrylamide gel for (B) SDS-PAGE and (C) Western blot using α-RHAU primary antibody.

In order to perform biophysical analysis of the DDX21 protein, expression of His_6-$DDX21_{C209}$ (24.4kDa) which is sufficient for the interaction between full length DDX21 protein and the G4 DNA substrate [77], and the expression of the FLAG-tagged full length DDX21 ($DDX21_{FL}$), which is 152.6kDa in size, was attempted in *E. coli* BL21 (DE3).

3.4 Expression and Purification of DDX21

In order to check for the presence of His_6-$DDX21_{C209}$/$DDX21_{C209}$ protein in the purification fractions obtained from the His_6-$DDX21_{C209}$ expressed in BL21(DE3) cells, SDS-PAGE was performed. We could clearly see a band running between the 37kDa and the 25kDa markers (Figure 18), suggesting that the His_6-$DDX21_{C209}$ was produced in the strain BL21(DE3) and purified as expected. A smaller amount of His_6-$DDX21_{C209}$ protein was detected in the lysate pellet and in the first washing buffer fraction (flowthrough 1), suggesting incomplete lysis of the cells and that a portion of $DDX21_{C209}$ bound more loosely to the nickel resin than the main fraction of the protein (Figure 18A).

A very faint band running close to the 150kDa marker (Figure 19A) in the lysate pellet, flowthrough 1 and eluate from the $DDX21_{FL}$ purification suggests that the $DDX21_{FL}$ protein may have been expressed in the strain BL21(DE3) and purified to a very low extent as suggested by SDS-PAGE data. Further analysis of the SDS-PAGE data suggests presence of closely positioned multiple lower molecular weight bands (Figure 19A) in the eluate indicating that the $DDX21_{FL}$ protein might be degraded. A band running immediately below 75kDa marker (Figure 19B) in the TEV treated protein sample indicated that the $DDX21_{FL}$ protein was further degraded, which was confirmed by Western blotting (Figure 19C) using an α-DDX21 antibody. The combined results of SDS-PAGE (Figure 19B) and Western blotting (Figure 19C) suggest that in the TEV treated sample purification, the DDX21 degradation product was also present in the flowthrough along

with the eluate. Because of the degradation issues with the $DDX21_{FL}$ protein, we decided to focus solely on the $DDX21_{C209}$ protein in the limited timeframe of this Master's thesis, and so further work on $DDX21_{FL}$ was discontinued.

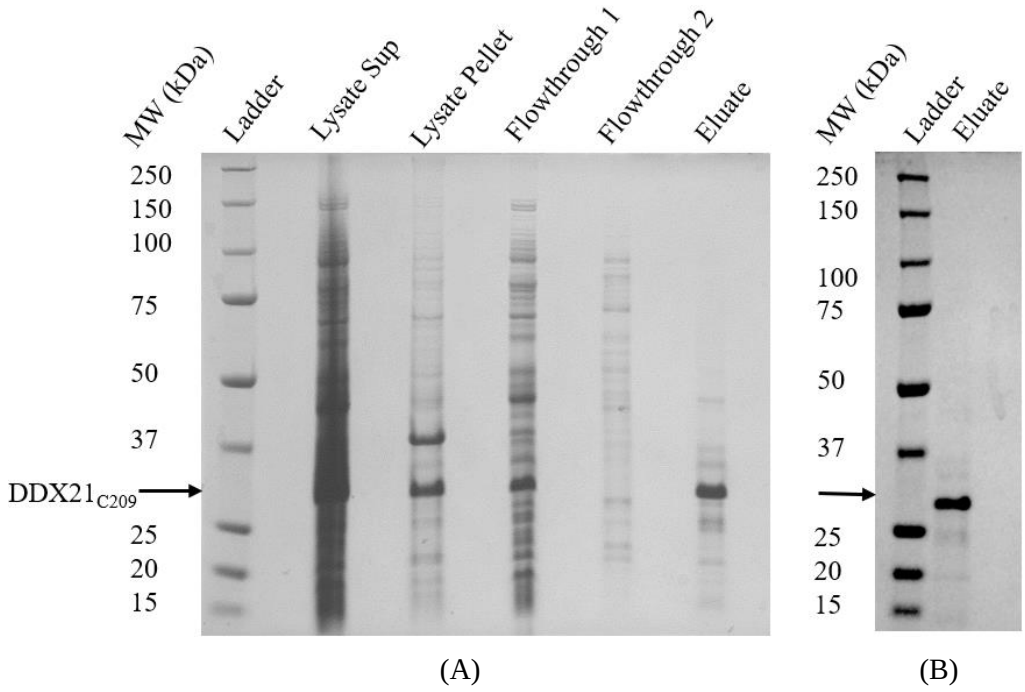


Figure 18. SDS-PAGE analysis of the $DDX21_{C209}$ protein purification. SDS-PAGE was performed on 8% polyacrylamide gel with (A) Different purification fraction samples and (B) The eluate obtained prior to crystallization trial.

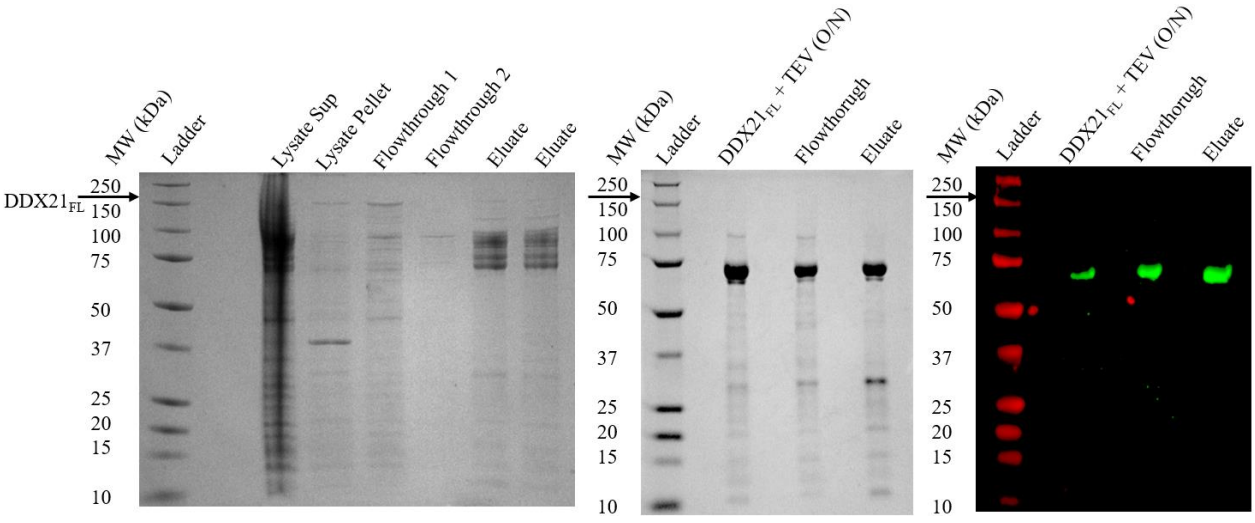


Figure 19. SDS-PAGE and Western blot analysis of nickel affinity purification samples of the $DDX21_{FL}$ protein. The samples were run on 8% polyacrylamide gel to perform both (A) SDS-PAGE. Samples from untagged $DDX21_{FL}$ protein purification attempt by TEV treatment were analyzed by both (B) SDS-PAGE and (C) Western blot using α-DDX21 antibody.

3.5 Preparation of Folded, Active, Stable DDX21$_{C209}$ Protein

CD spectroscopy was performed to check that the Ni-affinity purified DDX21$_{C209}$ protein was folded in a pre-established method by Ewan McRae. Duplicate experiments of UV CD spectroscopy show that the Ni-affinity purified DDX21$_{C209}$ protein was folded. In addition, negative bands at 207nm (near 208nm) and 220nm (near 222nm) and a positive band at 191nm (near 194nm) [133] in the spectra support the presence of α-helices in our purified DDX21$_{C209}$ protein (Figure 20).

The ITC experiment was performed to check that this sample of Ni-affinity purified DDX21$_{C209}$ protein was capable of binding to TERRA RNA G4 as previously established [53]. The thermodynamic parameters for DDX21$_{C209}$ binding to G4s obtained by ITC are as follows. N = 0.210 ± 0.01 sites, k_a = 3.53x10^6 ± 5.46x10^{5}M^{-1}, ΔH = - 27.58 ± 1.41kcal/mol and ΔS = -62.5cal/mol/deg which are highly similar to that previously reported [53]. We therefore concluded that the protein was functionally active and of sufficient quality to proceed to the next step (Figure 21).

In order to identify the best buffer composition for DDX21$_{C209}$ protein to minimize its tendency to precipitate, 8 salt concentrations ranging from 0 to 1M NaCl and 12 primary buffer conditions were tested to see which combination conferred the highest protein thermal stability. The thermal shift assay results indicated that the protein was most stable in a buffer containing 20mM sodium citrate, pH5.0 and 1M NaCl where the melting point (T_m) of the protein was 58.6°C. However, the amount of protein precipitation was unchanged in this buffer compared to the buffer containing 50mM Tris-HCl, pH7.5, 300mM KCl. The Tris buffer in the assay kit buffer screens, conferring the highest thermal stability to the protein contains 20mM Tris-HCl, pH8.0 and 1M NaCl (Figure 22) where the T_m of the protein was 51.1°C.

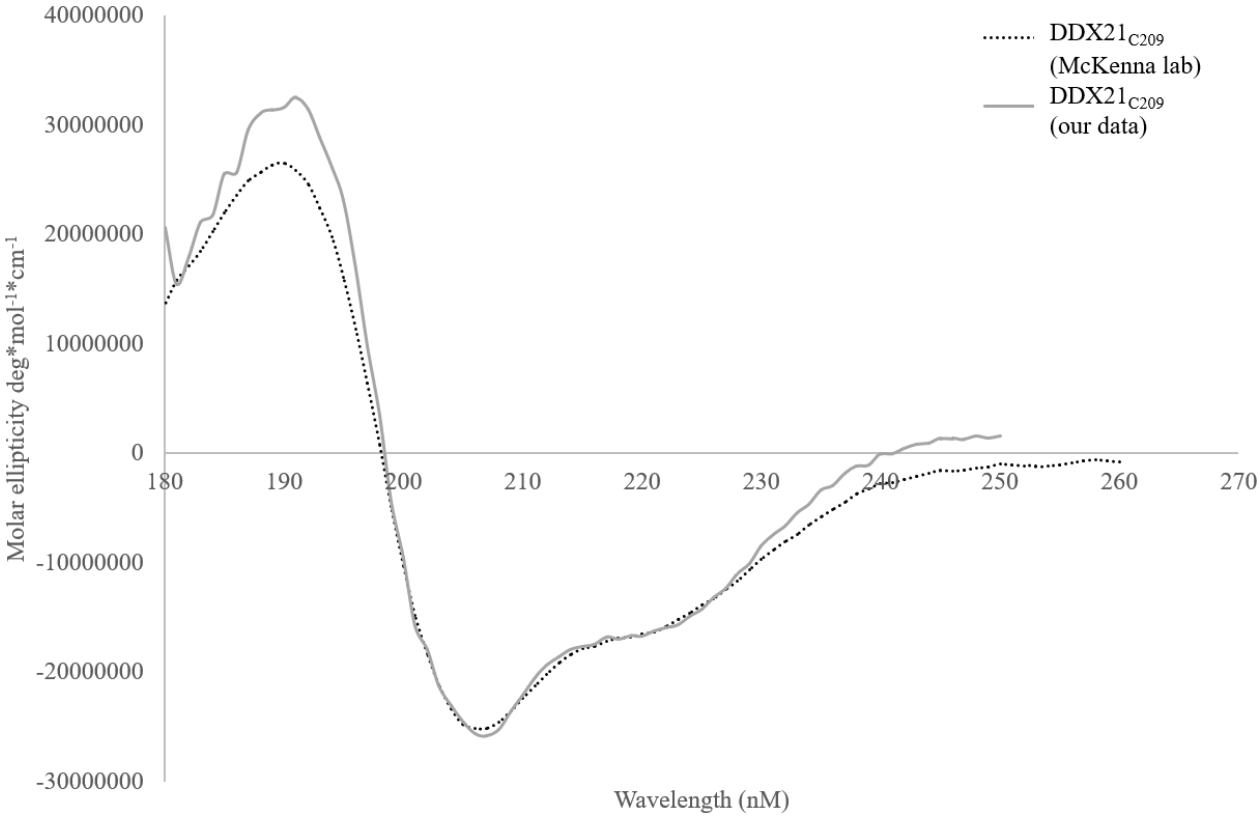


Figure 20. Far UV-visible CD spectra of the $DDX21_{C209}$ protein. The spectra were recorded at 20°C in a buffer containing 10 mM sodium phosphate, pH 7.5 and 150 mM NaF.

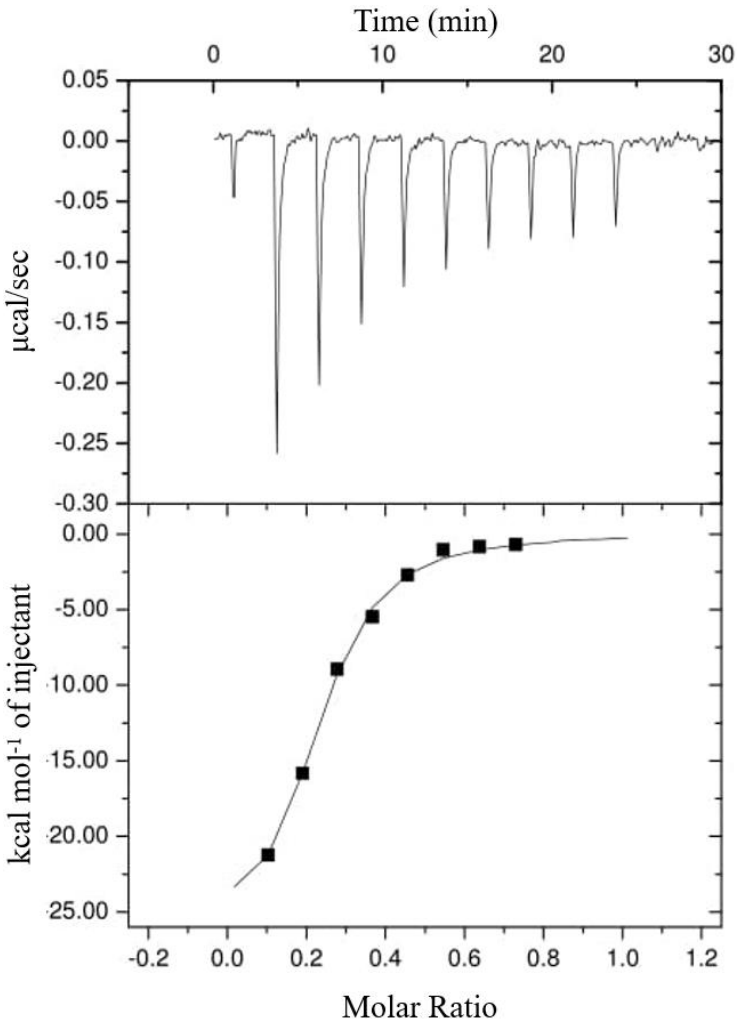


Figure 21. ITC titration of non-labelled TERRA RNA G4 into the $DDX21_{C209}$ protein. Binding interaction between the G4 RNA and the protein was measured at 20°C in a buffer containing 50mM Tris-HCl, pH 7.5 and 100mM KCl.

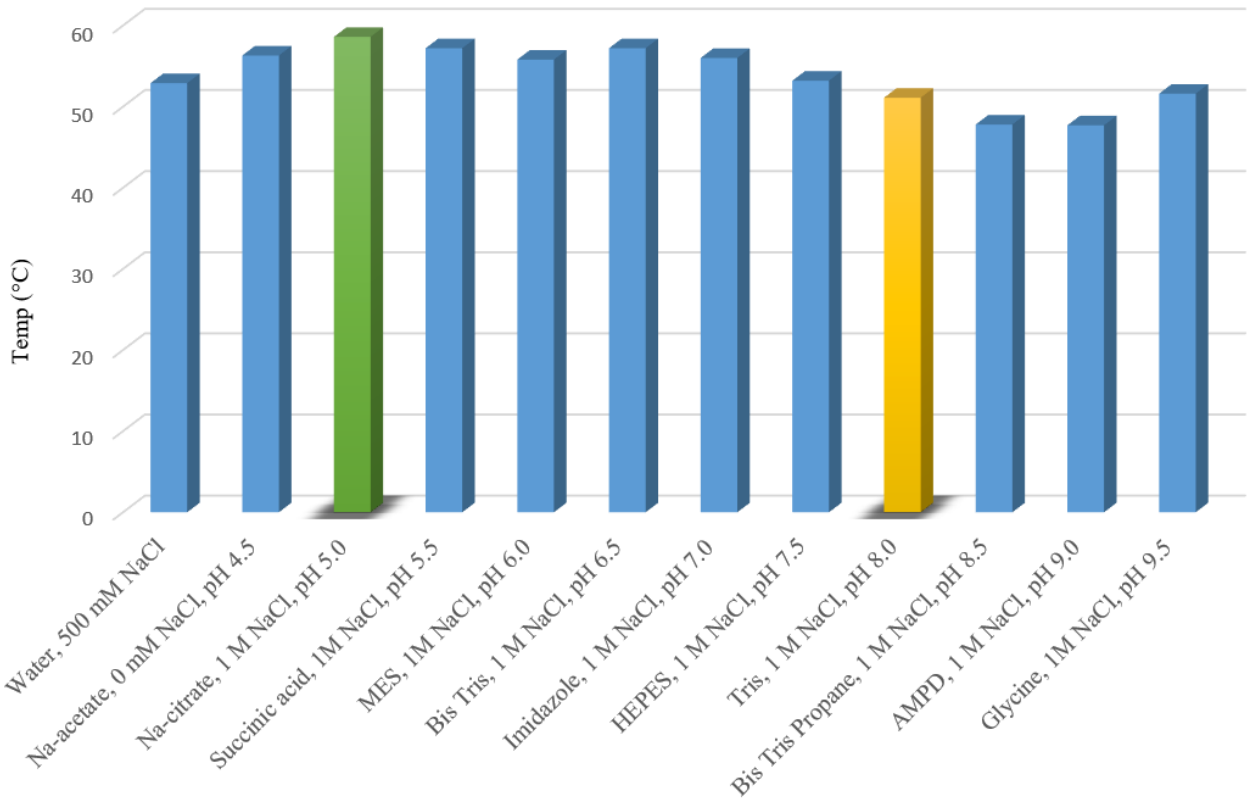


Figure 22. Overview of the thermal shift assay showing the best salt concentrations for the 12 primary buffer conditions for the DDX21$_{C209}$ protein. The bar graph indicates only those salt concentrations for each buffer type that confers maximum experimental thermal stability to the protein. Green, yellow and blue bars indicate the buffer conferring highest thermal stability, the Tris buffer conferring the maximum thermal stability, and NaCl concentrations for other buffer types conferring maximum thermal stability to the protein, respectively.

3.6 Crystallization Trials with DDX21$_{C209}$ Protein, and DDX21$_{C209}$·hTel Complex

In order to further purify the DDX21$_{C209}$ protein and the DDX21$_{C209}$·hTel complex (2:1 mole/mole) prior to the crystallization trials, samples of the protein and the complex each containing 1mg/mL DDX21$_{C209}$ protein estimated by the *DC* Protein Assay were loaded onto a Superdex 200 10/300 GL SEC column. SDS-PAGE analysis of the two eluted SEC peaks (Figure 23A) showed that only the final eluted peak sample resulted in protein bands. These bands were smear-like in pattern but correspond to the DDX21$_{C209}$ protein in size (Figure 23B), which indicated that the DDX21$_{C209}$ was perhaps degraded. Furthermore, *DC* Protein Assay analysis of the final eluted peak sample showed that only 5% of the protein came off the SEC column, which indicates that 95% of the protein was aggregated in the column. Spectrophotometric analysis at 260nm and 280nm provided an A_{260}/A_{280} ratio of 1.81, which indicated the presence of nucleic acids in the larger peak sample as expected.

In order to find suitable conditions for crystallization of the DDX21$_{C209}$ protein, and DDX21$_{C209}$·hTel complex, SEC purification step was omitted as the samples were mostly aggregating in the column. The protein only samples were prepared at 4.2mg/mL and 5.2mg/mL concentrations in two sets, and the DDX21$_{C209}$·hTel complex was similarly prepared at a protein concentration of 7.9mg/mL, and tested using the commercial crystallization screens JBScreen 1 – 5, JBScreen 6, JBScreen 7 JBScreen 8, Crystal Screen HT, JBScreen JCSG ++1 – ++4 and Index^{TM} HR2 - 144 (Hampton Research, CA, USA), but no crystals were obtained.

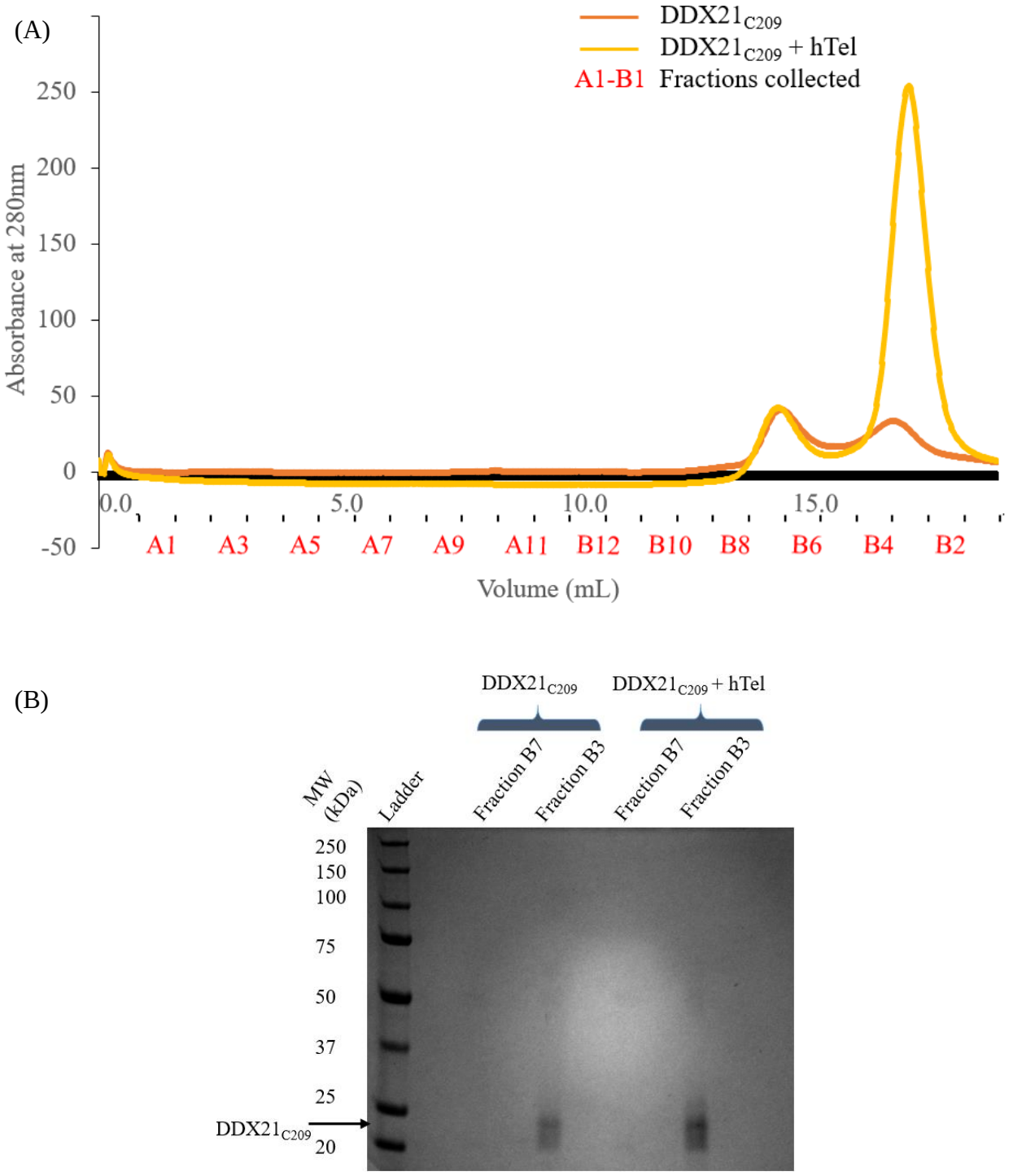


Figure 23. Size exclusion chromatography and SDS-PAGE analysis of the DDX21$_{C209}$ protein and the DDX21$_{C209}$·hTel complex. (A) SEC profiles of the samples eluted from the Superdex 200 10/300 GL column. (B) SDS-PAGE analysis on 8% polyacrylamide gel with the samples of SEC eluates (B7 and B3) of both the protein and the complex.

4. DISCUSSION

G-quadruplex structures, which are located *in vivo* in telomeric regions, promoters and in the first intron regions of various genes, have key regulatory roles in transcription, translation and mRNA splicing, and hence are implicated in different diseases [23]. The presence of a number of G4-interacting protein partners in the cellular environment further corroborates the functional importance of G-quadruplexes [46]. RHAU is one of the most well studied G4 helicases which interacts with G4s made of individually DNA or RNA, or DNA-RNA hybrid [46]. At the time when the RHAU project was started for this thesis, the mechanistic and structural details of the RHAU protein was largely unknown. Shortly afterwards, there was a burst of recent findings starting with its mechanistic highlights [71] and ending up with its crystal structures both in a free state and in complex with nucleic acids, although using the *Drosophila* [72] and bovine [51] homologs of the human RHAU protein which provided a much clearer insight into how the protein functioned.

Expression of $RHAU_{FL}$ was attempted in the *E. coli* strains C41(DE3) and BL21(DE3). Due to the recombinant nature of the $RHAU_{FL}$ protein, it was expected that expression of the protein in these strains would be plentiful [134]. In an attempt to increase the expression, protein induction conditions were modulated based on various troubleshooting protocols for recombinant protein production [134,135]. Despite using multiple strains of *E. coli*, varying the induction cell density (OD_{600}), IPTG concentration, induction time (for cell growth after induction), and growth temperature, the expression was not as high as expected. The reason behind low level protein expression could be that the heterologous human RHAU protein is harmful to the bacterial strains used. The harmful effect could be due to the differences in the post-translational modifications between prokaryotes and eukaryotes. Moreover, due to the large size of the human $RHAU_{FL}$ protein, it might be better expressed in eukaryotic cells than in the bacterial system [134]. Since,

expression of human $RHAU_{FL}$ had previously been reported in a bacterial system [130], we had not expected expression of $RHAU_{FL}$ to be highly challenging. One possible solution could have been to use a different codon optimization, as human codon usage is different to that found in bacteria [134]. Among all the attempted trials with different conditions, $RHAU_{FL}$ expression was the highest when the *E. coli* strain C41(DE3) was used and the cultures were induced at a cell density OD_{600} of 0.5 with 0.5mM IPTG, and the cells were grown at 20°C for 4hr. The slow growth of the cells due to the low growth temperature and C41(DE3) cells favored the proper folding of the RHAU protein [134]. After standardizing the expression parameters, only these conditions were used for $RHAU_{FL}$ protein expression, which was used in its purification. The truncated RHAU protein, $RHAU_{51\text{-}1008}$ was expressed in the *E. coli* strain BL21(DE3), where expression was the highest, when the cells were induced at a cell density OD_{600} of 0.7 with 2mM IPTG and the cells were grown at 28°C for 2hr.

A comprehensive purification approach was applied to overcome the challenges to abundantly purify the $RHAU_{FL}$ protein in its native form. First, purification of the RHAU protein was attempted by metal affinity chromatography exploiting the affinity of cobalt resin beads for the His_9-tag present in the recombinant His_9-$RHAU_{FL}$ protein. The imidazole eluted RHAU protein came off the column with a number of other polypeptide species which were together precipitating highly even at a low concentration (above 0.1mg/mL). Size exclusion chromatography (SEC) was applied to separate the RHAU protein from the other polypeptide species present in the cobalt affinity eluate based on their molecular size. In the SEC eluted sample, the RHAU protein was mostly found in a void volume fraction, where the proteins present are aggregated as corroborated by the fact that the RHAU protein was not separated from the few other polypeptide species present in the void volume eluted sample, despite eliminating the possibility that non-covalent force,

nucleic acids and disulfide bonds could be responsible for the RHAU protein to interact with the other polypeptide species. To minimize the possible harmful effect of imidazole [132] on the stability of the RHAU protein, minimum imidazole concentration required for maximum elution of the RHAU protein from the cobalt resin beads was determined to be 350mM which was much higher than the concentration (150mM) used in the first purification protocol. Yield of $RHAU_{FL}$ was extremely low upon elution of the protein using 350mM imidazole. Moreover, keeping the imidazole concentration at 350mM did not improve solubility issues with the RHAU protein. To eliminate the possible harmful effect of imidazole on the RHAU protein, a second approach of the RHAU protein purification was applied, where the pH was lowered to elute the protein from the histidine affinity cobalt resin column. The eluate contained the RHAU protein and many other polypeptide species, which still precipitated highly. During buffer exchange experiments, the eluate protein precipitation was minimized in the absence of salt (NaCl) in the buffer, however, this eluate could not be applied to the SEC column to separate the RHAU protein from the other polypeptide species because in the absence of salt proteins stick to the SEC column. To check whether the His-tag of the $RHAU_{FL}$ protein was affecting the protein stability, we attempted to cleave off the tag using TEV protease. However, this was only partially successful as we observed a partial removal of the tag. This was most likely because the proteins aggregated rendering the TEV site inaccessible to the TEV enzyme. The only time the RHAU protein precipitating less in a salt containing buffer was when the pH eluted protein was put into a buffer containing 20mM MES, pH6.0, 1 mM TCEP, 300mM NaCl and 10% glycerol. However, the RHAU protein was present at a very low level at this point. The third and the final approach of purification involved anion exchange chromatography, where negatively charged RHAU protein at pH9.0 bound to positively charged - $N(CH_3)^{3+}$ ligands of sepharose resin beads. However, when the RHAU protein

was eluted using NaCl gradient, the protein was found to be degraded. At this time, given the challenges with human RHAU expression and purification the decision was made to focus on another G4-binding protein, DDX21.

DDX21, like the RHAU protein, is one of the most important G4 helicases which interacts with both RNA and DNA G4s [53,77]. While the absence of RHAU has little or no impact on the global unwound state of RNA G4s [136], it has been implicated that DDX21 extensively interacts with RNA G-quadruplexes in a physiological context [77]. The C-terminal 209 residues of the protein ($DDX21_{C209}$) alone are sufficient for G4 binding although with a weaker affinity compared to the full-length protein [77]. The protein has implications in multiple diseases and pathogenesis which includes cancer [102,116]. Recent findings provided an overview of the binding interactions between the $DDX21_{C209}$ protein and G4-structures [53], but a crystallographic structure is yet to be available. In order to better understand the mechanistic details of RNA G4-helicases, DDX21 is one of the key proteins to study.

Expression of the $DDX21_{C209}$ protein, which is sufficient for interacting with G4 DNA [77], was performed in *E. coli* strain BL21(DE3), and purification of the protein by metal affinity chromatography was performed using Ni-NTA resin. Expression of the full length DDX21 protein ($DDX21_{FL}$) was also performed in *E. coli* strain BL21(DE3). However, purification of the protein by nickel affinity chromatography was unsuccessful, as the eluate contained degraded proteins. To check whether the degraded DDX21 protein had its tag attached TEV cleavage was performed, which was partially successful, perhaps because $DDX21_{FL}$ derivative had its TEV cleavage site buried in its structure. Further research could optimize the purification process; for example, lysing the cells using French Press instead of using sonication could be one way to avoid degradation of

the protein. To fully focus on the $DDX21_{C209}$ protein further experimental works on $DDX21_{FL}$ protein was discontinued.

Biophysical techniques were used to perform quality control of the purified $DDX21_{C209}$ protein. We used CD spectroscopy and ITC to show that our sample protein had the similar folding and the thermodynamic parameters of G4 binding, respectively as that prepared previously by the McKenna laboratory. The CD spectrum for our purified $DDX21_{C209}$ protein is characteristic of that obtained by the McKenna laboratory purified protein with characteristic negative bands near 208nm and 222nm, respectively and positive band near 193nm, which implicates the presence of α-helices in the protein. Thermodynamic analysis by ITC was performed to investigate the binding between $DDX21_{C209}$ protein and TERRA RNA G4 in a 2:1 molar ratio, which was selected as it is the RNA counterpart of the physiologically important 22-nucleotide truncation of the human telomere (hTel). ITC measured similar changes in positive enthalpy and negative entropy, and similar apparent dissociation constant (k_d) value as reported previously [53], which shows that the $DDX21_{C209}$ protein binds TERRA as expected, and hence, the protein is active. SEC was applied in an attempt to purify a homogeneous $DDX21_{C209}$·hTel complex. The SEC data indicated a sharp increase in the later eluted peak absorbance when $DDX21_{C209}$ protein was added onto the column in combination with hTel G4 in a 2:1 molar ratio, which indicates that the $DDX21_{C209}$ protein and hTel DNA G4 form a trimeric complex made of two molecules of the protein and one molecule of DNA G4. Spectrophotometric analysis at 260nm and 280nm showed the presence of nucleic acids in the eluted sample, which indicates that the protein and the DNA G4 forms complex in this molar ratio. For subsequent experiments involving this complex formation, SEC purification was excluded as the protein mostly stuck to the column and the SEC eluted proteins were degraded possibly due to its precipitation inside the column. To further investigate the binding between

DNA G4 and $DDX21_{C209}$, an attempt was made to grow crystals of $DDX21_{C209}$·hTel DNA G4 complex in a 2:1 molar ratio. To gain insight into the structural details of the $DDX21_{C209}$ protein at atomic level, attempts were made to grow crystals of the DDX21 protein in a 2:1 molar ratio at different protein concentration, however no crystals were obtained. Future attempt at growing crystals with higher concentration of the protein could provide success in crystallization. Optimization of the SEC purification could also yield success in growing crystals with the protein-nucleic acid complex. An anticipated high-resolution structure of $DDX21_{C209}$·hTel complex might provide a much clearer understanding of the interaction between the C-terminus of the DDX21 protein and the G-quadruplex. Further research on producing the full length DDX21 protein and its crystallization, both in a free state and in complex with nucleic acids, could provide valuable insight into the G4 helicase structure and function.

5. SUMMARY AND FUTURE DIRECTIONS

G-quadruplexes are thermodynamically stable secondary structures formed within one or more polynucleotide strands containing multiple stretches of guanine rich regions [23]. G4 structures have been found within telomeres and important regulatory regions of gene promoters, heavy chain switch regions of immunoglobulins and mutational hotspots [23,25,29]. At a cellular level G4s are found both inside nucleus and cytoplasm, and they vary in abundance throughout different phases of the cell cycle [23,36]. The *in vivo* interaction between G4s found in various regions and G4 helicases is of interest given their role in key biological processes [46]. RHAU shows major G4 unfolding activity in human cell lysates and is implicated in various diseases [23,75]. Therefore, we attempted to optimize expression and purification of the full length human RHAU protein to use biophysical studies and crystallization trials for the purpose of understanding its structure and the mechanisms underlying its G4 recognition activity. Due to concerns about the ability of the full-length protein

to crystallize, we also prepared a truncated $RHAU_{51-1008}$ version of the protein lacking its N-terminal glycine rich region.

Optimization of expression of the RHAU proteins was performed following a comprehensive recombinant protein production approach [134] highlighted in section 2. Gel electrophoresis and Western blot were used to conduct analysis of the protein expression. In section 3, based on the protein expression data, we showed optimal His_9-$RHAU_{FL}$ expression in the *E. coli* strain C41(DE3) (Figure 9B), and optimal His_9-$RHAU_{51-1008}$ expression in *E. coli* strain BL21(DE3) (Figure 10D). We followed a systematic approach to purify the expressed His_9-$RHAU_{FL}$ protein as highlighted in section 2. However, based on the protein purification data obtained, we were unable to purify the His_9-$RHAU_{FL}$ protein as highlighted in section 3. The use of a different plasmid construct varying either or both of vector plasmid and RHAU sequence codon optimization might result in better expression and successful purification of $RHAU_{FL}$ protein [134]. To fully focus on the DDX21 protein, His_9-$RHAU_{FL}$ purification was not attempted.

Like the RHAU protein, DDX21 is another physiologically important G4 helicase, highly abundant in cells and capable of recognizing both RNA and DNA G4s [53,77,97]. Unlike RHAU, DDX21 preferentially interacts with RNA G4s in physiological context [77,136]. DDX21 has been implicated in various diseases and so there is an interest in understanding its involvement in diseases. The C-terminal 209 residues of the DDX21 protein ($DDX21_{C209}$) is capable of binding to G4 although with a weaker affinity compared to the $DDX21_{FL}$ protein [77]. We expressed the $DDX21_{C209}$ and $DDX21_{FL}$ versions of the protein in the *E. coli* strain BL21(DE3), purified the $DDX21_{C209}$ protein and attempted to purify the $DDX21_{FL}$ protein for biophysical studies and crystallization so that we could try to better understand the biological role of DDX21.

Expression and purification of DDX21 proteins was performed in *E. coli* strain BL21(DE3) as highlighted in section 2. Gel electrophoresis was used to conduct analysis of the DDX21 protein expression. In section 3, based on the purification data, we showed that the DDX21 proteins were expressed (Figure 18,19). Nickel affinity purification was used to purify the proteins. In section 3, based on the protein purification data, we showed that the $DDX21_{C209}$ protein was purified (Figure 18). However, in the use of the $DDX21_{FL}$ protein, we found it degraded (Figure 19). To solely focus on the $DDX21_{C209}$ protein, work on $DDX21_{FL}$ was discontinued. As a quality control check of our purified $DDX21_{C209}$ protein, spectropolarimetry and ITC were performed as highlighted in section 2. In section 3, based on the CD spectra analysis, we showed that our purified $DDX21_{C209}$ protein was properly folded as expected (Figure 20), and based on the ITC result, we showed that our purified $DDX21_{C209}$ protein was capable of binding to TERRA G4 (Figure 21) as reported previously [53]. Thermal shift assay (TSA) was performed to find out the buffer condition conferring the highest thermal stability to the protein as highlighted in section 2. While based on the TSA result, the protein was most stable in a buffer containing 20mM sodium citrate, pH5.0 and 1M NaCl (Figure 22) the protein precipitation remained unchanged compared to the buffer containing 50mM Tris-HCl, pH7.5, 300mM KCl, used as final buffer in the purification. TSA result further showed that the Tris buffer conferring the highest thermal stability to the protein contains 20mM Tris-HCl, pH8.0 and 1M NaCl (Figure 22). SEC was used to purify a homogeneous $DDX21_{C209}$·hTel DNA G4 complex as highlighted in section 2. SEC result indicates that the $DDX21_{C209}$ protein forms a trimeric complex with hTel DNA G4 in a molar ratio of 2:1, however, the protein mostly stuck to the column and the eluted protein was degraded (Figure 23) as highlighted in section 3. $DDX21_{C209}$ protein and $DDX21_{C209}$·hTel DNA G4 complex samples were prepared excluding SEC purification for crystallization trials as highlighted in section 2.

Crystallization trials with the $DDX21_{C209}$ protein at a concentration of 4.2mg/mL and 5.2mg/mL, and $DDX21_{C209}$·hTel DNA G4 complex at 7.9mg/mL protein concentration did not yield crystals as highlighted in section 3. Use of higher concentration of the $DDX21_{C209}$ protein and/or optimization of SEC purification to purify a homogeneous protein and protein-G4 complex could yield crystals, which would provide further insight into the $DDX21_{C209}$ protein structure and mechanisms underlying $DDX21_{C209}$·G4 interaction. Production of the $DDX21_{FL}$ protein and analysis of the DDX21 proteins both in a free state and in complex with G4 would provide a better understanding of the protein structure and the mechanisms behind its G4 recognition and unwinding.

REFERENCES

1. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA : sequence , topology and structure. **34**, 5402–5415 (2006).

2. Bugaut, A. & Balasubramanian, S. 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res.* **40**, 4727–41 (2012).

3. Raiber, E.-A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **40**, 1499–508 (2012).

4. Martadinata, H. & Phan, A. T. Formation of a stacked dimeric G-quadruplex containing bulges by the 5'-terminal region of human telomerase RNA (hTERC). *Biochemistry* **53**, 1595–600 (2014).

5. Meier, M. *et al.* Binding of G-quadruplexes to the N-terminal Recognition Domain of the RNA Helicase Associated with AU-rich Element (RHAU) *. **288**, 35014–35027 (2013).

6. Bhattacharyya, D., Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, 1–14 (2016).

7. Malgowska, M., Czajczynska, K., Gudanis, D., Tworak, A. & Gdaniec, Z. Overview of the RNA G-quadruplex structures. *Acta Biochim. Pol.* **63**, 609–621 (2016).

8. Yuwei, Chen; Danzhou, Y. Sequence, Stability, Structure of G-quadruplexes and their drug interactions. *Curr Protoc Nucleic Acid Chem* **17**, 1–26 (2013).

9. Martadinata, H. & Phan, A. T. Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+solution. *J. Am. Chem.*

Soc. **131**, 2570–2579 (2009).

10. Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **38**, 5569–5580 (2010).

11. Mukundan, V. T. & Phan, A. T. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences.(Report). *J. Am. Chem. Soc.* **135**, 5017–5028 (2013).

12. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

13. Hänsel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: Detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **18**, 279–284 (2017).

14. Meier, M. *et al.* Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res.* **46**, 5319–5331 (2018).

15. Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. **417**, 3–7 (2002).

16. Laughlan, G. *et al.* The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* **265**, 520–4 (1994).

17. Phillips, K., Dauter, Z., Murchie, A. I. H., Lilley, D. M. J. & Luisi, B. The Crystal Structure of a Parallel-stranded Guanine Ê Resolution Tetraplex at 0 . 95 A. *J Mol Biol.* **273**, 171–182 (1997).

18. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *J. Mol. Biol.* **429**, 2127–2147 (2017).

19. Ariyo, E. O. *et al.* Biophysical characterization of G-quadruplex recognition in the PITX1 mRNA by the specificity domain of the helicase RHAU. **10**, 1–150 (2017).

20. Gellert, M., Lipsett, M. N. & Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **48**, 2013–2018 (1962).

21. Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).

22. Schiavone, D. *et al.* Determinants of G quadruplex-induced epigenetic instability in REV 1-deficient cells. *EMBO J.* **33**, 2507–2520 (2014).

23. Rhodes, D. & Lipps, H. J. Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).

24. Schaffitzel, C. *et al.* In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8572 (2001).

25. Yu, H.-Q., Miyoshi, D. & Sugimoto, N. Characterization of Structure and Stability of Long Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* **128**, 15461–15468 (2006).

26. Hackett, J. A., Feldser, D. M. & Greider, C. W. Telomere Dysfunction Increases Mutation Rate and Genomic Instability. *Cell* **106**, 275–286 (2001).

27. Blackburn, E. H. Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett.* **579**, 859–62 (2005).

28. Maizels, N. & Gray, L. T. The G4 genome. *PLoS Genet.* **9**, e1003468 (2013).

29. Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human

genome. **35**, 406–413 (2007).

30. Siddiqui-jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. **99**, 11593–11598 (2002).

31. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–21 (2007).

32. Arora, A., Dutkiewicz, M. & Scaria, V. Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. **14**, 1290–1296 (2008).

33. Gomez, D. *et al.* Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* **32**, 371–9 (2004).

34. Sagne, C. *et al.* Age at cancer onset in germline TP53 mutation carriers: association with polymorphisms in predicted G-quadruplex structures. *Carcinogenesis* **35**, 807–15 (2014).

35. Bonnal, S. *et al.* A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons. *J. Biol. Chem.* **278**, 39330–6 (2003).

36. Biffi, G., Tannahill, D., Mccafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182–186 (2013).

37. Sarkies, P. *et al.* FANCJ coordinates two pathways that maintain epigenetic stability at G-

quadruplex DNA. *Nucleic Acids Res.* **40**, 1485–98 (2012).

38. Vaughn, J. P. *et al.* The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **280**, 38117–20 (2005).

39. Wu, Y. Unwinding and rewinding: Double faces of helicase? *J. Nucleic Acids* **2012-12-01**, 6–19 (2012).

40. Pyle, A. M. Translocation and Unwinding Mechanisms of RNA and DNA Helicases. *Annu. Rev. Biophys.* **37**, 317–336 (2008).

41. Merrick, W. C. Found in translation - the discovery of the first RNA helicase, eIF4A. *RSC Biomol. Sci.* **19**, v–x (2010).

42. Stahl, H., Dröge, P. & Knippers, R. DNA helicase activity of SV40 large tumor antigen. *EMBO J.* **5**, 1939 (1986).

43. Scheffner, M., Knippers, R. & Stahl, H. RNA unwinding activity of SV40 large T antigen. *Cell* **57**, 955–963 (1989).

44. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23 (2007).

45. Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).

46. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).

47. Zhou, X. *et al.* Structural and Functional Insights into the Unwinding Mechanism of

Bacteroides sp Pif1. *Cell Rep.* **14**, 2030–2039 (2016).

48. Bennett, R. J. & Keck, J. L. Structure and Function of RecQ DNA Helicases. *Crit. Rev. Biochem. Mol. Biol. 2004, Vol.39(2), p.79-97* **39**, 79–97 (2004).

49. Liu, J.-Q., Chen, C.-Y., Xue, Y., Hao, Y.-H. & Tan, Z. G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *J. Am. Chem. Soc.* **132**, 10521 (2010).

50. Wang, Q. *et al.* G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res.* **39**, 6229 (2011).

51. Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465–483 (2018).

52. Robert, F. & Pelletier, J. Perturbations of RNA helicases in cancer. *Wiley Interdisciplinary Reviews: RNA* **4**, 333–349 (2013).

53. McRae, E. K. S., Davidson, D. E., Dupas, S. J. & McKenna, S. A. Insights into the RNA quadruplex binding specificity of DDX21. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 1973–1979 (2018).

54. Jackson, R. N., Lavin, M., Carter, J. & Wiedenheft, B. Fitting CRISPR-associated Cas3 into the Helicase Family Tree. *Curr. Opin. Struct. Biol.* **24**, 106–114 (2014).

55. Hickman, A. B. & Dyda, F. Binding and unwinding: SF3 viral helicases. *Curr. Opin. Struct. Biol.* **15**, 77–85 (2005).

56. Gorbalenya, A. E., Koonin, E. V & Wolf, Y. I. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Lett.* **262**, 145–148

(1990).

57. Ilyina, T., Gorbalenya, A. & Koonin, E. Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J. Mol. Evol.* **34**, 351–357 (1992).

58. Sakakibara, N., Kelman, L. M. & Kelman, Z. Unwinding the structure and function of the archaeal MCM helicase. *Molecular Microbiology* **72**, 286–296 (2009).

59. Boos, D., Frigola, J. & Diffley, J. F. Activation of the replicative DNA helicase: breaking up is hard to do. *Curr. Opin. Cell Biol.* **24**, 423–430 (2012).

60. Georgescu, R. *et al.* Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation. *Proc. Natl. Acad. Sci.* **114**, E697–E706 (2017).

61. Sharples, G. J., Ingleston, S. M. & Lloyd, R. G. Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. *J. Bacteriol.* **181**, 5543 (1999).

62. Tran, H., Schilling, M., Wirbelauer, C., Hess, D. & Nagamine, Y. Facilitation of mRNA Deadenylation and Decay by the Exosome-Bound, DExH Protein RHAU. *Mol. Cell* **13**, 101–111 (2004).

63. Iwamoto, F., Stadler, M., Chalupníková, K., Oakeley, E. & Nagamine, Y. Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp. Cell Res.* **314**, 1378–91 (2008).

64. Lai, J. C. *et al.* The DEAH-box helicase RHAU is an essential gene and critical for mouse hematopoiesis. *Blood* **119**, 4291–4300 (2012).

65. Chen, M. C. & Ferré-D'Amaré, A. R. Structural Basis of DEAH/RHA Helicase Activity. *Crystals* **7**, 1–13 (2017).

66. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. *Cell* **125**, 287–300 (2006).

67. Büttner, K., Nehring, S. & Hopfner, K.-P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–52 (2007).

68. Walbott, H. *et al.* Prp43p contains a processive helicase structural architecture with a specific regulatory domain. *EMBO J.* **29**, 2194–204 (2010).

69. He, Y., Andersen, G. R. & Nielsen, K. H. Structural basis for the function of DEAH helicases. *EMBO Rep.* **11**, 180–186 (2010).

70. Jarmoskaite, I. & Russell, R. RNA Helicase Proteins as Chaperones and Remodelers. *Annu. Rev. Biochem.* **83**, 697–725 (2014).

71. Yangyuoru, P. M., Bradburn, D. A., Liu, Z., Xiao, T. S. & Russell, R. The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism. *J. Biol. Chem.* **293(6)**, 1924–1932 (2017).

72. Chen, W. F. *et al.* Molecular Mechanistic Insights into Drosophila DHX36-Mediated G-Quadruplex Unfolding: A Structure-Based Model. *Structure* **26**, 403–415.e4 (2018).

73. Jain, A., Bacolla, A., Chakraborty, P., Grosse, F. & Vasquez, K. M. Human DHX9 helicase unwinds triple-helical DNA structures.(Report). *Biochemistry* **49**, 6992–6999 (2010).

74. Booy, E. P. *et al.* RNA helicase associated with AU-rich element (RHAU/DHX36) interacts with the 3′-tail of the long non-coding RNA BC200 (BCYRN1). *J. Biol. Chem.* **291**, 5355–

5372 (2016).

75. Creacy, S. D. *et al.* G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **283**, 34626–34 (2008).

76. Giri, B. *et al.* G4 Resolvase 1 tightly binds and unwinds unimolecular G4-DNA. *Nucleic Acids Res.* **39**, 7161–7178 (2011).

77. McRae, E. K. S. *et al.* Human DDX21 binds and unwinds RNA guanine quadruplexes. *Nucleic Acids Res.* **45**, 6656–6668 (2017).

78. Chen, M. C., Murat, P., Abecassis, K., Ferré-D'Amaré, A. R. & Balasubramanian, S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* **43**, 2223–2231 (2015).

79. Ohnmacht, S. A. & Neidle, S. Small-molecule quadruplex-targeted drug discovery. *Bioorg. Med. Chem. Lett.* **24**, 2602–2612 (2014).

80. Smaldino, P. *et al.* Mutational Dissection of Telomeric DNA Binding Requirements of G4 Resolvase 1 Shows that G4-Structure and Certain 3'-Tail Sequences Are Sufficient for Tight and Complete Binding: e0132668. *PLoS One* **10**, 1–21 (2015).

81. Tippana, R., Hwang, H., Opresko, P. L., Bohr, V. A. & Myong, S. Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8448–8453 (2016).

82. Heddi, B., Cheong, V. V., Martadinata, H. & Phan, A. T. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide–

quadruplex complex. *Proc. Natl. Acad. Sci.* **112**, 9608–9613 (2015).

83. Booy, E. P. *et al.* The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* **40**, 4110–4124 (2012).

84. Lattmann, S., Giri, B., Vaughn, J. P., Akman, S. A. & Nagamine, Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* **38**, 6219–6233 (2010).

85. Kim, H. *et al.* J BMR Histone Deacetylase Inhibitor MS-275 Stimulates Bone Transcription. **26**, 2161–2173 (2011).

86. Huang, W. *et al.* Yin Yang 1 contains G-quadruplex structures in its promoter and 5'-UTR and its expression is modulated by G4 resolvase 1. *Nucleic Acids Res.* **40**, 1033–49 (2012).

87. Bicker, S. *et al.* The DEAH-box helicase DHX36 mediates dendritic localization of the neuronal precursor-microRNA-134. *Genes Dev.* **27**, 991–6 (2013).

88. Booy, E. P. *et al.* The RNA helicase RHAU (DHX36) suppresses expression of the transcription factor PITX1. *Nucleic Acids Res.* **42**, 3346–3361 (2014).

89. Guangchao, S. *The Regulation of YY1 in Tumorigenesis and its Targeting Potential in Cancer Therapy*. *Molecular and Cellular Pharmacology* **1**, 157-176 (2009).

90. Lattmann, S., Selak, N., Iwamoto, F. & Fujiki, Y. Recruitment of the RNA Helicase RHAU to Stress Granules via a Unique RNA-binding Domain. **283**, 35186–35198 (2008).

91. Weinrich, S. L. *et al.* Recruitment of the RNA Helicase RHAU to Stress Granules via a Unique RNA-binding Domain. *Nat. Genet.* **17**, 498–502 (1997).

92. Kirkpatrick, K. L. & Mokbel, K. The significance of human telomerase reverse transcriptase (hTERT) in cancer. *Eur. J. Surg. Oncol.* **27**, 754–760 (2001).

93. Sexton, A. N. & Collins, K. The 5’ guanosine tracts of human telomerase RNA are recognized by the G-quadruplex binding domain of the RNA helicase DHX36 and function to increase RNA accumulation. *Mol. Cell. Biol.* **31**, 736–43 (2011).

94. Lattmann, S., Stadler, M. B., Vaughn, J. P., Akman, S. A. & Nagamine, Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA Gquadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.* **39**, 9390–9404 (2011).

95. Chen, J. L., Blasco, M. a & Greider, C. W. Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–14 (2000).

96. Chen, J.-L. & Greider, C. W. Template boundary definition in mammalian telomerase. *Genes Dev.* **17**, 2747–52 (2003).

97. Flores-rozas, H. & Hurwitzs, J. Characterization of a New RNA Helicase from Nuclear Extracts of HeLa Cells Which Translocates in the 5 ’ to 3 ’ Direction *. **268**, 21372–21383 (1993).

98. Valdez, B. C. *et al.* A nucleolar RNA helicase recognized by autoimmune antibodies from a patient with watermelon stomach disease. *Nucleic Acids Res.* **24**, 1220–1224 (1996).

99. Linder, P. & Jankowsky, E. From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.* **12**, 505–516 (2011).

100. Ou, Y., Fritzler, M. J., Valdez, B. C. & Rattner, J. B. Mapping and Characterization of the Functional Domains of the Nucleolar Protein RNA Helicase II / Gu. **398**, 389–398 (1999).

101. Eriksson, M., Saffrich, R., Ansorge, W. & Bohmann, D. Complex Functions of AP-1 Transcription Factors in Differentiation and Survival of PC12 Cells. **21**, 4369–4378 (2001).

102. Westermarck, J. *et al.* The DEXD/H-box RNA helicase RHII/Gu is a co-factor for c-Jun-activated transcription. *EMBO J.* **21**, 451–460 (2002).

103. Holmström, T. H. *et al.* c-Jun supports ribosomal RNA processing and nucleolar localization of RNA helicase DDX21. *J. Biol. Chem.* **283**, 7046–7053 (2008).

104. Hirai, Y. *et al.* Nucleolar scaffold protein, WDR46, determines the granular compartmental localization of nucleolin and DDX21. *Genes to Cells* **18**, 780–797 (2013).

105. Zhang, Z. *et al.* DDX1, DDX21, and DHX36 Helicases Form a Complex with the Adaptor Molecule TRIF to Sense dsRNA in Dendritic Cells. *Immunity* **34**, 866–878 (2011).

106. Dong, Y. *et al.* DDX21 translocates from nucleus to cytoplasm and stimulates the innate immune response due to dengue virus infection. *Biochem. Biophys. Res. Commun.* **473**, 648–653 (2016).

107. Zhang, H. *et al.* A double-negative feedback loop between DEAD-box protein DDX21 and Snail regulates epithelial-mesenchymal transition and metastasis in breast cancer. *Cancer Lett.* **437**, 67–78 (2018).

108. Cao, J. *et al.* DDX21 promotes gastric cancer proliferation by regulating cell cycle. *Biochem. Biophys. Res. Commun.* **505**, 1189–1194 (2018).

109. Fuller-Pace, F. V. DExD/H box RNA helicases: Multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.* **34**, 4206–4215 (2006).

110. Hammond, J. A. *et al.* A Survey of DDX21 Activity During Rev/RRE Complex Formation.

J. Mol. Biol. **430**, 537–553 (2018).

111. Valdez, B. C., Henning, D., Perumal, K. & Busch, H. RNA-unwinding and RNA-folding activities of RNA helicase IVGu Two activities in separate domains of the same protein. **807**, 800–807 (1997).

112. Valdez, B. C. Structural domains involved in the RNA folding activity of RNA helicase II / Gu protein. **6402**, 6395–6402 (2000).

113. Ariyo, E. O. *et al.* Impact of G-quadruplex loop conformation in the PITX1 mRNA on protein and small molecule interaction. *Biochem. Biophys. Res. Commun.* **487**, 274–280 (2017).

114. McRae, E. K. S., Booy, E. P., Padilla-Meier, G. P. & McKenna, S. A. On characterizing the interactions between proteins and guanine quadruplex structures of nucleic acids. *J. Nucleic Acids* **2017**, 1–11 (2017).

115. Martino, L. *et al.* Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]4. *J. Am. Chem. Soc.* **129**, 16048–16056 (2007).

116. Zhang, Y., Baysac, K. C., Yee, L. F., Saporita, A. J. & Weber, J. D. Elevated DDX21 regulates c-Jun activity and rRNA processing in human breast cancers. *Breast Cancer Res.* **16**, 1–18 (2014).

117. Krishnan, V. & Zeichner, S. L. Alterations in the expression of DEAD-box and other RNA binding proteins during HIV-1 replication. *Retrovirology* **1**, 1–5 (2004).

118. Naji, S. *et al.* Host Cell Interactome of HIV-1 Rev Includes RNA Helicases Involved in

Multiple Facets of Virus Production. *Mol. Cell. Proteomics* **11**, M111.015313 (2012).

119. Fornerod, M., Ohno, M., Yoshida, M. & Mattaj, I. W. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**, 1051–1060 (1997).

120. Putnam, A. A. & Jankowsky, E. DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 884–893 (2013).

121. Pomerantz, R. J., Seshamma, T. & Didier, A. Efficient Replication of Human Immunodeficiency Virus Type 1 Requires a Threshold Level of Rev: Potential Implications for Latency. *J. Virol.* **66**, 1809–1813 (1992).

122. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 1–8 (2011).

123. Edgcomb, S. P. *et al.* DDX1 is an RNA-dependent ATPase involved in HIV-1 Rev function and virus replication. *J. Mol. Biol.* **415**, 61–74 (2012).

124. Wang, L. & Wang, J. MicroRNA-mediated breast cancer metastasis: from primary site to distant organs. *Oncogene* **31**, 2499–2511 (2011).

125. Nieto, M. angela, Huang, R. yun-J., Jackson, R. a. & Thiery, J. paul. EMT: 2016. *Cell* **166**, 21–45 (2016).

126. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat. Cell Biol.* **16**, 488–494 (2014).

127. Batlle, E. *et al.* The transcription factor Snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89 (2000).

128. Nieto, M. A. The snail superfamily of zinc-finger transcription factors. *Nat. Rev. Mol. Cell*

Biol. **3**, 155–166 (2002).

129. Chalupníková, K. *et al.* Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. *J. Biol. Chem.* **283**, 35186–35198 (2008).

130. Yoo, J. S. *et al.* DHX36 Enhances RIG-I Signaling by Facilitating PKR-Mediated Antiviral Stress Granule Formation. *PLoS Pathog.* **10**, 1–18 (2014).

131. Marushchak, O. *Biophysical investigation of G-quadruplex recognition by the N-terminal construct of RNA helicase associated with AU-rich element (RHAU) / by Oksana Marushchak.* (Winnipeg, Manitoba : University of Manitoba, 2014).

132. Mezzasalma, T. M. *et al.* Enhancing Recombinant Protein Quality and Yield by Protein Stability Profiling. *J. Biomol. Screen.* 418–428 (2007). doi:10.1177/1087057106297984

133. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc.* **1**, 2876–2890 (2009).

134. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli : advances and challenges. **5**, 1–17 (2014).

135. Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Front. Microbiol.* **5**, 1–20 (2014).

136. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science.* **353**, aaf5371-1-aaf5371-8 (2016).

APPENDIX

List of Abbreviations

2D-STD	2-dimensional saturation transfer difference spectroscopy
AIDS	Acquired immune deficiency syndrome
AP-1	Activating protein-1
ARE	AU-rich element
β-ME	β-mercaptoethanol
CD	Circular dichroism
CR	Conserved region
CRM-1	Chromosome region maintenance 1
Cryo-EM	Cryo-electron microscopy
CTD	C- terminal domain
CTE	C-terminal extension
DDX21	DExD-box helicase 21
$DDX21_{C209}$	C-terminus 209 amino acid construct of DDX21
$DDX21_{FL}$	Full length DDX21
DHX36	DEAH-box protein 36 (DHX36)
DNA	Deoxyribonucleic acid

DSM	DHX36 / RHAU specific motif
E. coli	*Escherichia coli*
EGF	Epidermal growth factor
eIF4A	Eukaryotic initiation factor4A
EMT	Epithelial-mesenchymal transition
FPLC	Fast protein liquid chromatography
GB1	IgG domain B1 of Protein G
G4	G-quadruplex
G4R1	G4 resolvase 1
Glu	Glumatic Acid
His	Histidine
HIV	Human immunodeficiency virus
HJ	Holliday junction
hTel	Human telomeric DNA
hTERT	Human telomerase reverse transcriptase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
MCM	Minichromosome maintenance protein

mDC	Myeloid dendritic cell
miRNA134	micro-RNA-134
NIS	Nuclear diffusion inhibitory signal
MLEL1	MLE-like protein 1
NMR	Nuclear magnetic resonance
NTE	N terminus extension
NTP	Nucleotide 5′-triphosphate
NusA	N-utilization substance
OB	OB-like fold
PARN	Polyadenylate-specific ribonucleases
PDB	Protein database
P-helix	Paired helix
PITX1	Paired Like Homeodomain 1
RHAU	RNA Helicase Associated with AU-rich element
$RHAU_{51-1008}$	RHAU protein lacking N-terminal 50 amino acids
$RHAU_{FL}$	Full length RHAU
RL	Ratchet-like
RNA	Ribonucleic acid

RNP	Ribonucleoprotein
RRE	Rev response element
RSM	RHAU specific motif
S. cerevisiae	*Saccharomyces cerevisiae*
SEC	Size exclusion chromatography
SF	Superfamily
Strep II	Streptavidin binding peptide
TERRA	Telomeric non-coding ribonucleic acid
TEV	Tobacco etch virus protease
TICAM1	TIR domain-containing adapter molecule 1
TNAP	Tissue-nonspecific alkaline phosphatase
TSA	Thermal shift assay
UTR	Untranslated region
WH	Winged helix
XRD	X-ray diffraction
YY1	Yin Tang 1