University of Nottingham
UK | CHINA | MALAYSIA

Molecular Biology of Human DDX49 – Novel biochemical activities

A thesis submitted to the University of Nottingham for the degree of Master of Research

Sabesan Anandavijayan, B. Sc (Hons.)

September 2022

Abstract:

DExD-box helicases belong to superfamily II of RNA helicases and are characterised by a conserved Asp-Glu-X-Asp motif in their primary protein sequence. These helicases are often involved in pre-mRNA splicing, ribosome biogenesis and also cell cycle mediation; as such, they are implicated in the oncogenesis and proliferation of various cancer types. The overexpression of DEAD-box protein 49 (DDX49) has been associated with a favourable prognosis in cervical cancer patients, however, its structure and function has yet to be fully elucidated. We report that DDX49 is a dual function helicase and nuclease, and discuss how CRISPR-Cas9 mediated downregulation of DDX49 may affect the proliferation and viability of U2OS osteosarcoma cells.

Table of Contents

Chapter 1: Introduction

1.1 Helicases

Helicases are a diverse group of enzymes that utilise the energy released from ATP hydrolysis to unwind or remodel nucleic acid substrates; they may also be able to displace or otherwise interact with nucleic acid-protein complexes. Helicases are ubiquitous across all forms of cellular life and are involved in almost all biological processes that feature DNA or RNA, including DNA replication, transcription, and RNA maturation.[1]

Helicases are subdivided into six superfamilies (SFs) based on conserved sequence motifs in their ATP hydrolysis and helicase core domains. The catalytic cores of helicases are comprised of RecA-like folds (termed 'RecA-like' due to their similarity with the ATP-binding core of RecA recombination protein) that are responsible for DNA binding and adenosine triphosphate (ATP) hydrolysis.[2] SF1 & 2 contain two RecA-like fold domains connected by a flexible linker region to form a "dumbbell"-like structure (Figure 1A) while the core domains of SF3-6 are comprised of six to twelve RecA-like folds which form a hexameric, toroidal structure (Figure 1B). However, despite the extensive structural similarity, SF1 and SF2 helicases differ in the mechanism of unwinding and the range of "accessory domains".[1,3]

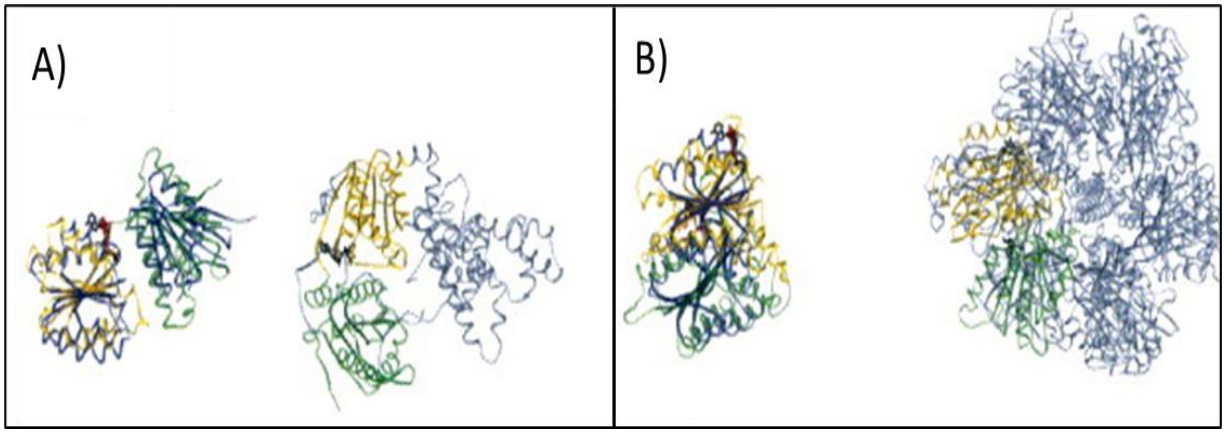

Figure 1: Side by side comparison of the RecA fold domains in a typical A) Superfamily 2 helicase and B) Superfamily 3 helicase*. Image modified from*[4]

Helicases can be further differentiated based on their translocation polarity (the direction in which helicases can move across polynucleotide strands), which is either in a 3' to 5' direction across the polynucleotide strand (type A), or in the opposite 5' to 3' direction (type B).[2] While often nucleic acid-protein interactions occur on the polynucleotide strand to which the protein has translocated, this may not always be the case.[1,3]

1.1.1 Superfamily 2 Helicases

Superfamily 2 is the largest and most diverse of the helicase superfamilies, they are primarily characterised by 9 highly conserved amino acid motifs (Q, Ia, Ib, Ic, II, III, IV, V, VI) located in between the cleft of the RecA-like domains; the conserved motifs coordinate the interactions with nucleic acids and with nucleotide triphosphates (NTPs).[5] SF2 can be further classified according to sequence conservation into families such as: RecQ-like, Rad3/XPD, DExD-box, DExH/RHA and also Swi/Snf.[1] The majority of SF2 helicases have type A translocation polarity, although there are exceptions such as members of the XPD subfamily.[6]

The residues within the conserved motifs of SF2 helicases primarily contact the phosphodiester backbones of the nucleic acids, with other interactions necessary for unwinding being carried out by other accessory domains, altering helicase function and specificity. Motif I ((G/A)xxxxGKT/S) is also known as the Walker A motif, it is responsible for binding the phosphate group of a nucleotide.[7] Motif II also makes contacts with the nucleotide; the aspartate residue can coordinate Mg^{2+} ions required for ATP hydrolysis.[1]

The focus of this research shall be on the DExD-box helicase DDX49, but first, the DExD-box sub-family will be discussed further.

1.1.2 DExD-Box Helicases

DExD-box helicases are a large family of ATP-dependent, SF2 helicases found across all three domains of life, they are characterised by a distinct Aspartate-Glutamine-X-Aspartate conserved motif in their primary structure (where X is any amino acid), this has the abbreviation: DExD (hence the family name). The DExD motif is contained within motif II of the helicase's amino acid sequence. The motif confers ATP hydrolysis and RNA unwinding ability, but many DExD-box helicases have also displayed a strong binding and unwinding affinity for DNA substrates, and some DExD-box helicases have additional roles conferred by their N and C-terminal domains. DExD-box helicases have often sparked interest for their multitude and diverse roles in RNA metabolism which include ribosome biogenesis, pre-mRNA splicing, mRNA export, and RNA decay. Due to this, they are often implicated in the oncogenesis and proliferation of cancer types.[4]

1.1.3 Mechanism of nucleic acid unwinding by DExD-box helicases

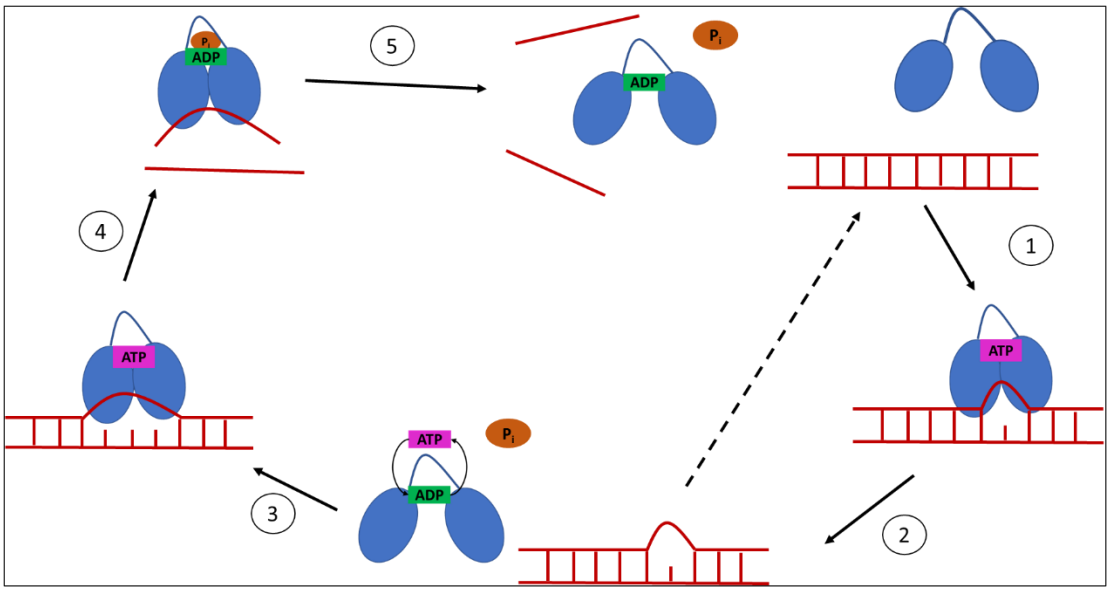

Figure 2: The mechanism of duplex substrate unwinding by DExD-box helicases. *The RecA core domains (blue) of DExD-box helicases unwind duplexes via local strand separation. Lines represent nucleic acid strands, and the two blue ovals represent the two RecA domains joined by a flexible linker; these are far apart in the absence of the substrate and ATP. Step 1) RecA domains adopt a closed conformation upon binding to ATP and duplexed nucleic acids. Step 2) ATPase activity opens the conformation of the helicase core domain, releasing the substrate, further duplex destabilisation requires ATP hydrolysis, as shown in Step 3. The partially opened substrate may reassociate back to a fully wound substrate (dotted arrow) in a non-productive cycle. Step 4) ADP-P_i bound core domains are still associated with the bent substrate strand. Step 5) Release of the phosphate ion opens the conformation of the core domains, allowing release of the strand. Original figure based on*[8]

In the absence of double-stranded substrate and ATP, the RecA-like domains at the core of a DExD-box helicase are splayed far apart, exhibiting an “open” conformation. Cooperative binding with a substrate and ATP is associated with the RecA-like domains adopting a closed conformation. The change in conformational state bends a strand of the double-stranded substrate to induce local destabilisation of the duplex. The unbent strand of the duplex dissociates from the helicase-substrate complex, while the bent strand remains bound to the helicase core domains.

The bound ATP is hydrolysed, and the inorganic phosphate ion is released, causing the RecA-like domains to open and release the second strand. The ADP is then released, allowing the conformation of the helicase to open, and resetting it for further catalytic cycles.[8] While each unwinding event requires ATP binding hydrolysis, not every cycle results in strand separation, therefore separation of a whole substrate may require multiple cycles.[8,9]

1.2 DDX49

DDX49 is a DEAD-box helicase expressed in *Homo sapiens*, its genetic locus is on chromosome 19. DDX49 may be involved in mRNA export, this was evidenced by a reduction in cytoplasmic mRNA levels when DDX49 expression was knocked down in HeLa cervical cells using small-interfering RNA (siRNA). The same study also noted that DDX49 was found to be recruited to the promoter regions of rDNA, and that siRNA-mediated depletion of DDX49 reduced the levels and stability of pre-ribosomal rRNA, this compounded with its regulatory role in mRNA export suggest that DDX49 could also affect global translation and protein synthesis.[10] While it has been stated that DDX49 is localised to the nucleolus, this contradicts immunofluorescence assays where it was localised within the mitochondria.[10,11] DDX49 is reported to be well conserved across the Animal kingdom, but no clear homologues have been found in other eukaryotic kingdoms, or even in the *Bacteria* or *Archaea* domains.[12]

Like many other DExD-box helicases, the level of expression of DDX49 has been implicated in cancer development: a high expression of DDX49 is associated with a favourable prognosis in cervical cancer patients (Figure 1)

but an unfavourable prognosis in both renal cancer patients and liver cancer patients. DDX49 was also found to be a biomarker for the metastases of lung cancer. It has been found that downregulating DDX49 in non-small cell lung cancer cells also decreased levels of AKT and beta-catenin proteins; this also reduced the proliferation of the cells. AKT and beta-catenin are involved in the AKT/beta-catenin signalling pathway, in which high levels of their expression results in activation of downstream targets involved in cellular growth and survival. Therefore, DDX49 may promote the tumorigenesis of non-small cell lung cancer by promoting the AKT/beta-catenin signalling pathway and its expression may be useful as a biomarker for lung cancer metastases, similar to DDX3.[13,14]

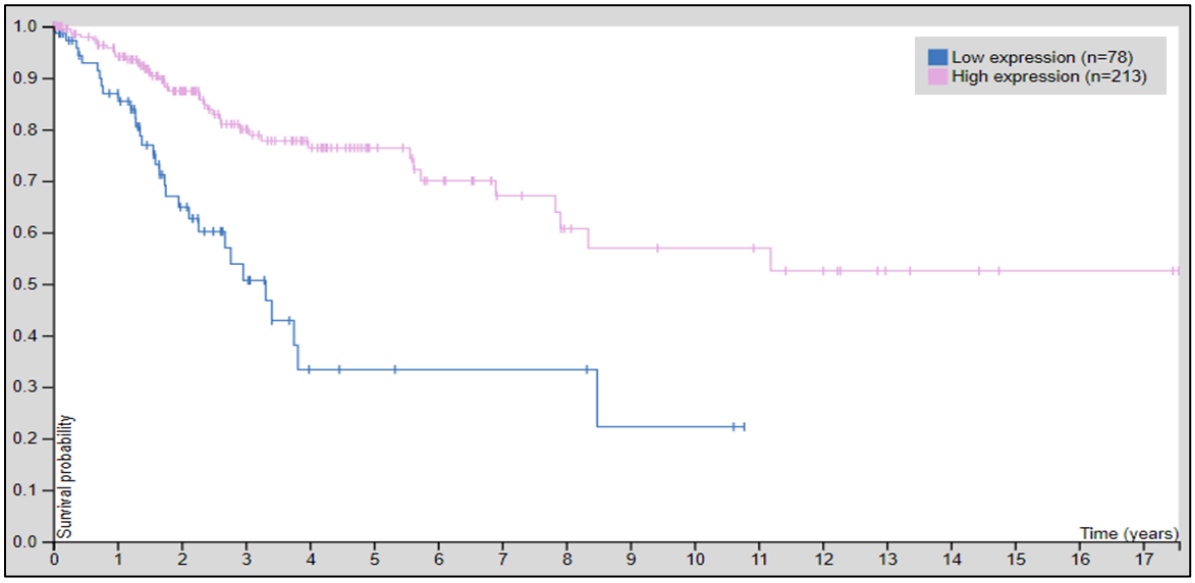

Figure 3: Cervical cancer patients with a high expression of DDX49 have a higher survival prognosis than patients with a low expression of DDX49*. A "high" or "low" expression is defined as the level of mRNA expression either above or below a FPKM (number of fragments per kilobase of exon per million reads) score of 16.85. Kaplan-Meier plots show whether the mRNA expression level of a gene is significantly associated (p<0.001) with the probability of patient survival, where 1.0 corresponds to 100%. Graph taken from* [11]

Although previous research has shown that DDX49 displays ATPase and RNA helicase activity that is typical of most DEAD-box proteins[10], its precise

biochemical nature has been poorly characterised. From DDX49's amino acid sequence, crystal structures have been predicted.

1.3 Structural characteristics of DDX49

As mentioned earlier, the precise biochemical properties of DDX49 have not yet been fully characterised, however its primary amino acid sequence has been resolved, and from this, two crystal structure have been predicted: from the AlphaFold protein structure database (shown in Figure 4A)[15], and from Protein Homology/Analogy Recognition Engine 2.0 (Phyre2.0), (shown in Figure 4B).[16] I will discuss various motifs and structural features that the predicted structures of DDX49 may resemble.

1.3.1 Overall structure description

DDX49 has a primary protein structure consisting of 483 amino acids, it has the typical features of a DExD-box helicase: a core helicase domain comprised of 2 RecA-like folds that are connected by a flexible linker, this is located towards the N-terminus. However, close to the C-terminus, DDX49 also possesses an accessory domain which precedes a region of high intrinsic disorder.

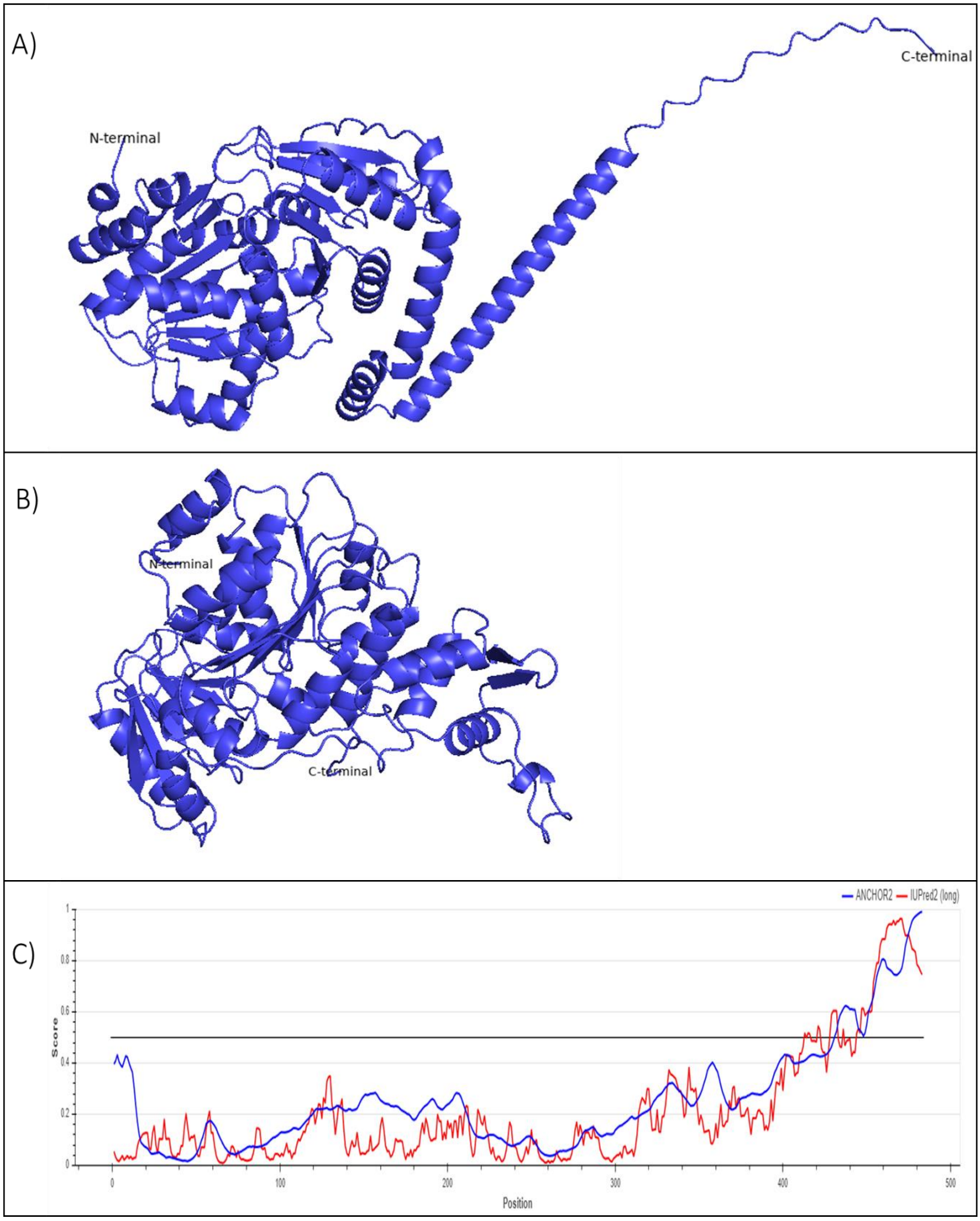

Figure 4: The full crystal structures of DDX49 predicted by using A)* the A*lphafold protein structure database*[15]*, and also with B) the Phyre2.0 protein fold recognition server.[16] *The crystal structures were modelled using the PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC.*[17]*C) The structural disorder within DDX49 predicted by IUPRED2, available at: https://iupred2a.elte.hu* [18]*. Scores returned by IUPRED2 reflect the probability of a given residue being part of a disordered region.*

1.3.2 Core domain motifs

The core domain of DDX49 is very typical of member of the DExD-box family. The Walker A and B motifs are well conserved across homologues of DDX49 (Figure 5).

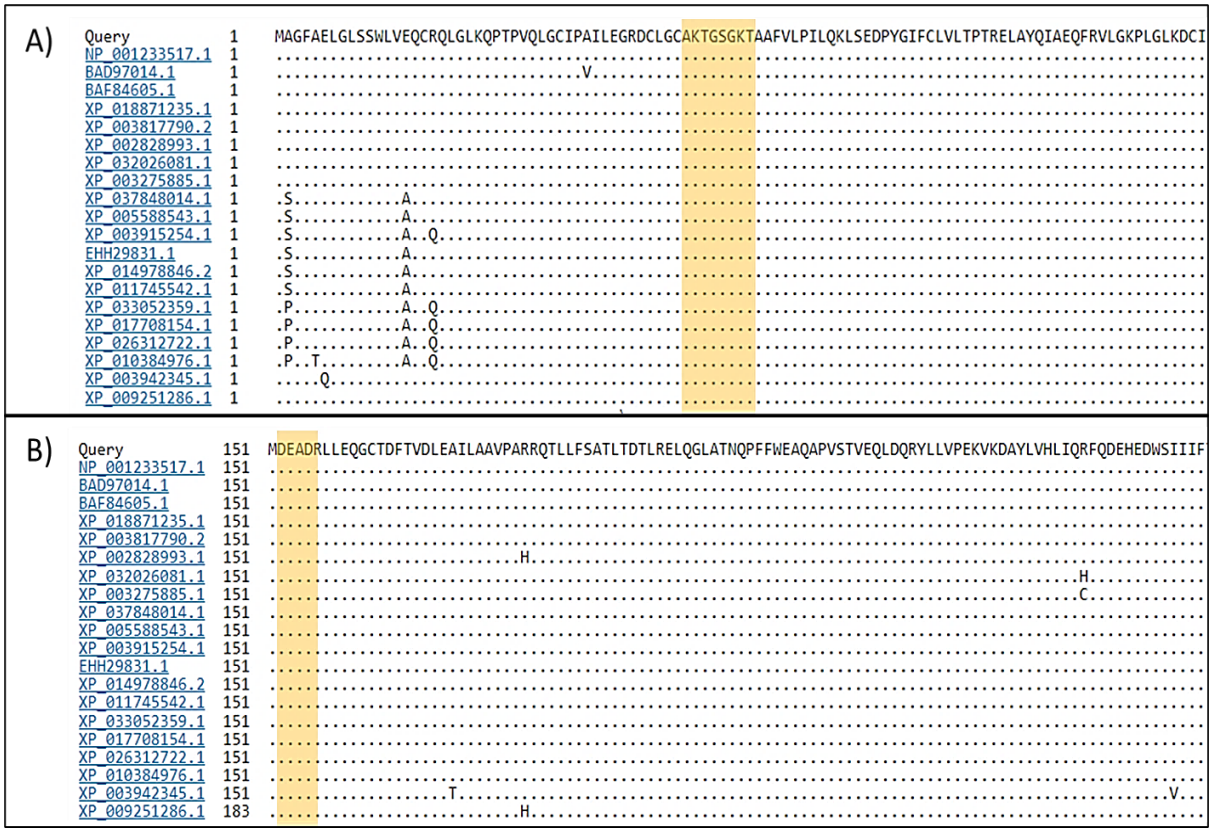

Figure 5: A) The Walker A motif (AKxxxxGKT) is well conserved across non-mammalian homologues of Hsp_DDX49, as is B) the Walker B motif (DExD). *These conserved sequences are highlighted in yellow. The amino acid sequence of Hsp_DDX49 was inputted into the BLASTp algorithm. Dots underneath the sequence indicate that the residue is conserved.*[12]

1.3.3 Accessory Domain motifs

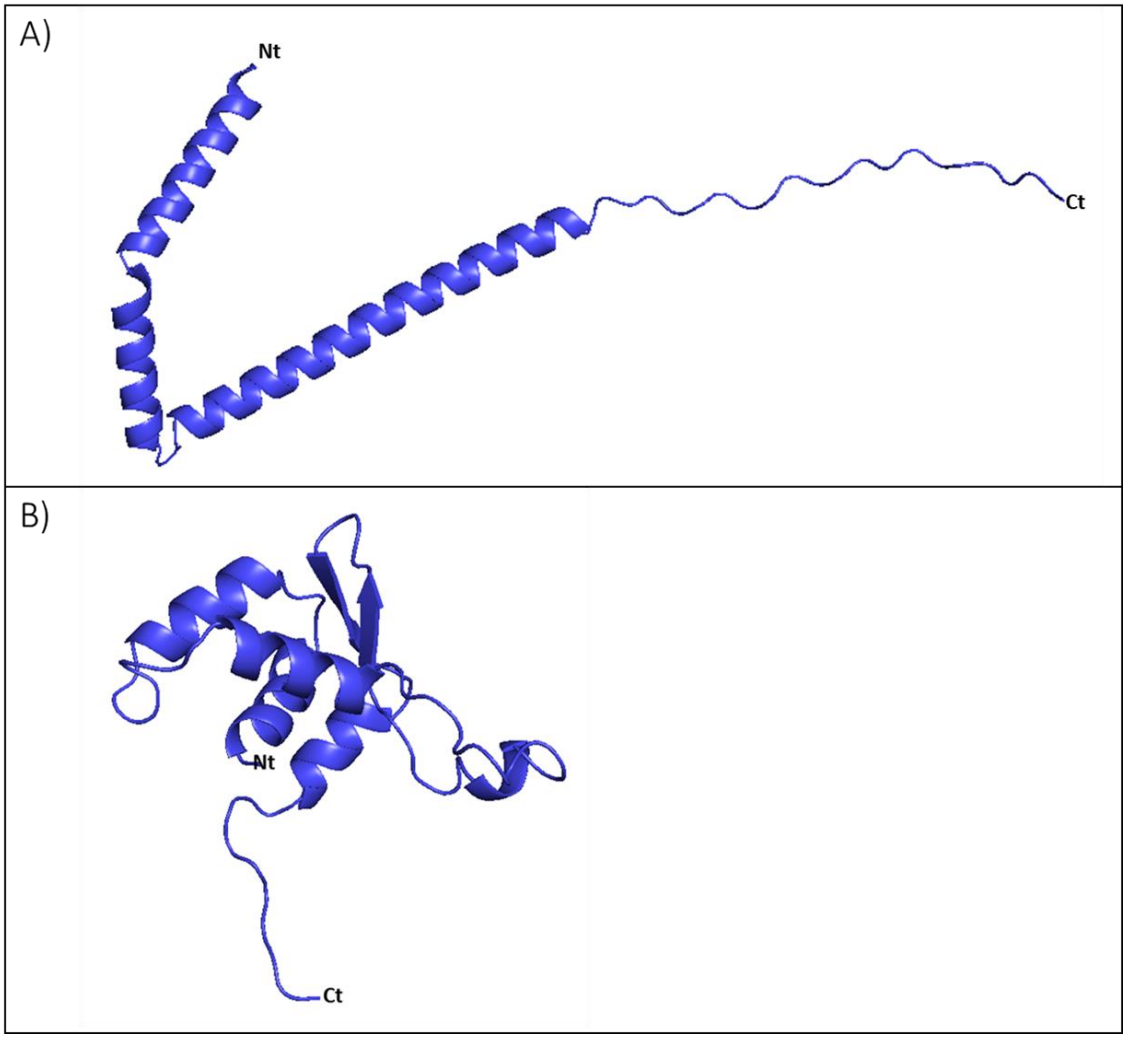

Figure 6: The Crystal structures of residues 383-483 of DDX49 were predicted by using A) the AlphaFold protein structure database[15], and also with B) the Phyre2.0 protein fold recognition server.[16] *The crystal structures were modelled using the PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC.*[17]

DDX49 is predicted to have an accessory domain towards its C-terminus between amino acid residues 383 to 483, the function of which has not been elucidated thus far. This region in the protein's structure is well conserved between non-mammalian homologues of DDX49, but not between other members of the DExD-box family. AlphaFold (Figure 6A) predicts that this accessory domain consists of two alpha helices connected by a linker region and followed by a chain of amino acids with intrinsic disorder. Whereas

Phyre2.0 (Figure 6B) predicts that the accessory domain consists of three alpha-helices, followed by two beta strands and a final, fourth alpha helix (α1 α2 α3 β1 β2 α4). There are several motifs that these predictions resemble, I shall discuss these motifs.

Possibly the most obvious structural motif that the AlphaFold-predicted accessory domain resembles is a di-helical helix-turn-helix (HTH) motif, these can bind DNA in the major groove. HTH motifs typically consist of alpha helices joined by a short chain of amino acids, the second helix (towards the C-terminus) is often associated with sequence recognition, while the N-terminal alpha helix is responsible for stabilising the interaction with DNA. Most HTH motifs in DNA-binding proteins are tri-helical, while DDX49's apparent HTH motif appears to be di-helical, however it could be argued that residues 383-400 make up a separate alpha-helix that is linked to the other helices by Histidine-401, thus fitting the tri-helical HTH model.[19,20]

The topology of the Phyre2.0-predicted accessory domain resembles a winged helix domain, which can bind DNA. A typical winged helix is around 100 amino acid residues long and consists of three alpha-helices at its core, a beta sheet, and flexible loops (or "wings").[21] The FOX (Forkhead box) family of proteins are Winged-Helix transcription factors, they have key roles in embryonic development and cellular differentiation.[22] The C-terminal of DEAH-box helicases are highly conserved and may contain winged-helix domains, the proteins are closely related to, and are involved in many processes that DEAD-box helicases play a part in, such as Dhr1 in ribosome biogenesis.[23] However, there are slight differences in topology between a

typical winged helix domain and the accessory domain of DDX49 described by Phyre2.0.

Within the primary structure of the accessory domain of Hsp_DDX49 is an Aspartate-Proline-Aspartate (DPD) motif that is conserved among non-mammalian homologues of the protein. An Aspartate-x-Aspartate motif (DxD) is also found to be conserved in topoisomerase-primase (TOPRIM) motifs and is involved with the coordination of Mg^{2+} ions.[24] TOPRIM motifs are found in many enzyme families such as: Type IA and type II topoisomerases, members of the RecR family as well as bacterial and archaeal OLD-family nucleases; reverse gyrase from *Thermatoga maritima* has a TOPRIM domain in addition to many of the domains found in SF2 helicases (albeit with variations of the consensus sequences of the motifs). TOPRIM motifs can catalyse the cleavage or formation of phosphodiester bonds, and this tends to be Mg^{2+}-dependent.[25,26] It could be hypothesised that the accessory domain has an Mg^{2+}-dependent function akin to TOPRIM motifs, and mutation of the DPD motif in DDX49 may shed light on this.

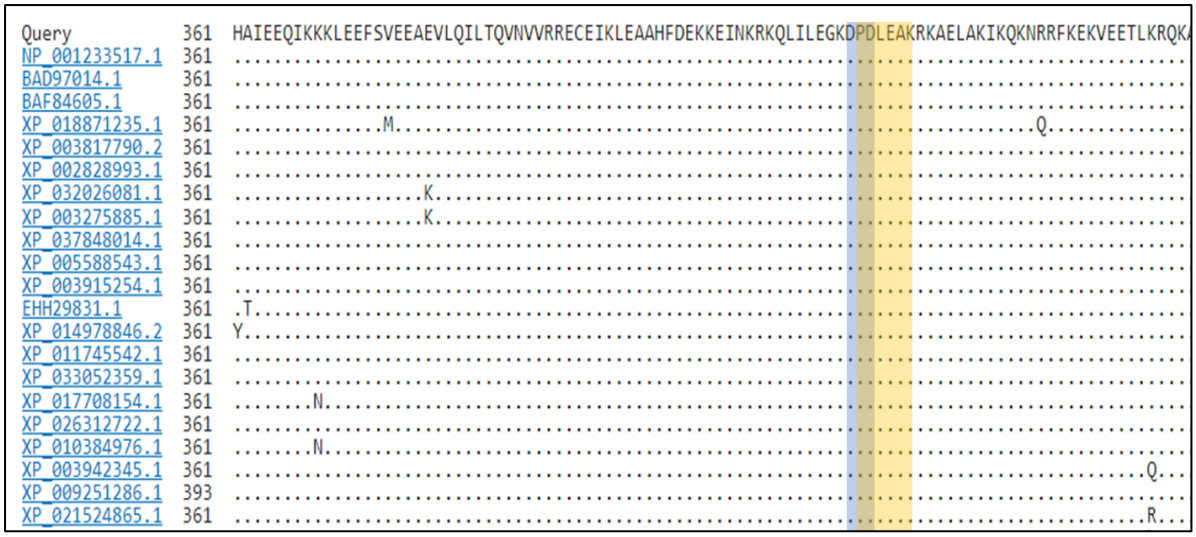

***Figure 7: The DxD motif (highlighted in blue) found within the TOPRIM domains is well conserved in non-mammalian homologues of Hsp_DDX49 as is the motif that resembles the PD-(D/E)xK domain (highlighted in yellow**). The amino acid sequence of Hsp_DDX49 was inputted into the BLASTp algorithm. Dots underneath the sequence indicates that the residue is conserved.*[12,27,28]

Another signature motif that the primary sequence of DDX49's accessory domain may be resemble is of the PD-(D/E)x_nK domain, which has been identified in several Mg^{2+} dependent nucleases, including type II restriction endonucleases; this may be seen in DDX49 as Proline-Aspartate - Leucine - Glutamic Acid – Alanine – Lysine (PDLEAK). Within the PD-(D/E)x_nK superfamily, this motif is weakly conserved but it lends further credence to the hypothesis that the accessory domain of DDX49 has an Mg^{2+} dependent nuclease function.[27]

Since many DExD-box helicases have a strong binding preference for RNA substrates, it is possible that DDX49's accessory domain is involved in RNA binding. The most abundant RNA binding secondary superstructure in higher vertebrates is the RNA recognition motif (RRM), this typically adopts a β1α1β2β3α2β4 topology, forming a four-strand beta sheet against two alpha helices. Since neither DDX49 accessory domain greatly resembles this, it can

be ruled out.[29] The differences in prediction and structural representation of DDX49 between AlphaFold and Phyre2.0 may be attributed to differences in the prediction algorithms; AlphaFold also uses a deep learning approach to predict protein structures whereas Phyre2.0 relies upon homology modelling.

1.4 DExD-box helicases in Ribosome Biogenesis

In *S. cerevisiae*, at least 19 DExD/H-box helicases are involved in ribosome biogenesis: seven are implicated in small subunit biogenesis, and ten are implicated in large subunit biogenesis and two others are involved in the assembly of both subunits.[30]

RNA helicases serve a purpose to dissociate snoRNAs (small nucleolar RNAs) from pre-rRNA. Many snoRNAs have sequence motifs complementary to pre-rRNA, these can guide associated proteins to make covalent modifications to the pre-rRNA. Dbp8 is a DEAD-box helicase that is essential to SSU maturation in Yeast, its depletion *in* vivo has been linked to a deficit in 40S ribosome subunit levels. Processing of cut sites in 35S pre-rRNA were also inhibited when Dbp8 expression was knocked down. Dbp8 has a co-factor, Esf2 that is known to associate with the U3 snoRNP at the 5-ETS region of the 35S pre-rRNA and it is thought Dbp8 is recruited to the pre-rRNA substrate when binds to Esf2's C-terminal domain.[31] Therefore, it can be assumed that Dbp8 plays a direct role in 18S rRNA formation and in turn, 40S subunit maturation.

The three DEAD-box helicases Dbp6, Dbp7 and Dbp9 are involved in 60S subunit maturation, and they seem to be functionally connected. Dbp6 is part

of the Npa1 complex, which chaperones rRNA folding and modification that is critical to pre-60S subunit assembly.[32]

DDX49 has been implicated in the regulation of ribosome biogenesis, it may do this by binding to the rDNA locus of 47S rRNA, thus regulating the synthesis of pre-ribosomal 47S rRNA. [10]

1.5 DExD-Box Helicases in Cancer

The aberrant function or expression of DExD-box helicases has been implicated in the development of cancers,[33] there are several examples of this.

The DEAD-box helicase DDX5 is overexpressed in many cancer types such as colorectal cancer and non-small cell lung cancer.[34] DDX5 can act as a transcriptional co-activator to Beta-Catenin which drives the transcription of several proto-oncogenes such as C-myc and Cyclin D1 through the Wnt-signalling pathway. In colorectal cancer cells, overexpressed DDX5 (along with Beta-catenin) occupies the promoter region for the AKT family of genes, increasing their transcription. Consequently, increased AKT protein expression downregulates the expression of genes targeted by FOXO3a, including p27kip1 which regulates cell cycle progression at phase G1 and therefore contributing to tumorigenesis.[34,35]

DDX3 is another DEAD-box helicase implicated in the development of cancers, but whether it is tumour-promoting or tumour-supressing is subject to debate and may depend on the cancer type.[36] DDX3 plays an oncogenic role in the tumorigenesis of breast cancers, its overexpression correlates with promoted cellular growth and proliferation in the breast cancer cell line MCF

10A.; DDX3 was also found to repress the expression of E-cadherin, a tumour-suppressing gene. Hypoxia is a common trait of solid tumours, whilst MCF 10A cells are in this state, HIF-1a (hypoxia inducible factor 1A)) induces the transcriptional activation of the DDX3 promoter, upregulating its expression and likely helping the cells to survive under hypoxia. Tumour metastasis was inhibited *in vivo* when DDX3 expression was knocked down in the MCF-10A cell line.[37] Contrastingly in Hepatocellular carcinoma, DDX3 has tumour supressing qualities and so it is downregulated in hepatocellular carcinoma cells. Studies in these cells have demonstrated that the declined expression of DDX3 also results in the downregulation of p21 expression, leading to the upregulation of Cyclin D1 and consequently early progression of the cell cycle into S-phase.

DDX49 is implicated in various cancer types, its overexpression in cervical cancer patients is associated with a favourable survival prognosis, and it is though that it may be involved in the tumorigenesis of non-small cell lung cancer.[11,13]

1.6 Project Aims

The aims of this project were to characterise the biochemical properties of DDX49 *in vitro,* as they have currently not yet been elucidated. This would be achieved by assaying purified DDX49 with different DNA and RNA substrates, highlighting a preference for either nucleic acid substrate. We have also made predictions identifying structural motifs in C-terminal accessory domain of DDX49, and we aim to conduct mutagenesis in key residues of DDX49 in hopes of gaining insight into the protein's function.

In addition, I aimed to knock out DDX49 expression *in cellulo* (within living cells) in U2OS osteosarcoma cells using CRISPR-Cas9 mediated gene editing techniques. This could generate a phenotype which could be compared against the wild-type to see if the depletion of DDX49 in U2OS cells has any major effect on their growth.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Chemicals

Chemicals used and listed within 2.2. Methods are supplied from either Fisher Scientific or Sigma-Aldrich unless otherwise stated.

2.1.2 Bacterial strains and cell lines

Table 2.1 List of cell strains and cell lines as mentioned in the methods and results sections.

***E. coli* strain**	Supplier	Genotype
DH5a	Invitrogen	F– φ80lacZΔ M15 Δ (*lacZYA-argF*) *U169 recA1 endA1 hsdR17* (rK– mK+) *phoA supE44* λ- *thi–1 gyrA96 relA1*
BL21AI	Invitrogen	F-*omp*T *hsd*S_B (r_B- m_B-) *gal dcm ara*B: T7RNAP-*tet*A

Table 2.2 List of cell strains and cell lines as mentioned in the methods and results sections.

Cell line	Information
U2OS	Human osteosarcoma cell line cultivated from a differentiated sarcoma of the tibia of a 15-year-old, Caucasian female patient.

2.1.3 Media and Supplements for E. coli culture

Table 2.3.1 Buffers, media used in culturing *E. coli* cells throughout this project.

Media Name	Composition
LB Broth	10g/L Tryptone 10g/L Sodium Chloride 5g/L Yeast extract pH adjusted to 7.0 using NaOH
LB Agar	10g/L Tryptone 10g/L Sodium Chloride 5g/L Yeast Extract 15g/L Agar pH adjusted to 7.0 using NaOH

Table 2.3.2 Supplements used for *E. coli* culture.

Supplement	Supplier	Working Concentration
Ampicillin	Sigma	50μg/mL
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	VWR	0.5mM
L-Arabinose	Sigma	0.2% (w/v)

2.1.4 Media and supplements used for Human cell culture

Table 2.4.1 Media used in Human cell culture.

Media Name	Composition
Dulbecco's modified Eagle's medium (DMEM)	Supplemented with 10% Fetal bovine serum, 2mM L-glutamine, 100 units/mL penicillin, 100μg/mL streptomycin

Table 2.4.2 Supplements used in Human cell culture

Supplement	Working Concentration
Fetal bovine serum (FBS)	10%
L-glutamine	2mM
Penicillin/streptomycin	100 units/mL penicillin, 100μg/mL streptomycin

2.1.5 Plasmids and nucleic acid substrates

Table 2.5.1 List of plasmids used for the characterisation of DDX49.

Cloned Plasmids	Description	Method
Pet100D.DDX49	Codon optimised His-tagged DDX49 cloned into Pet100D backbone	As delivered by manufacturer (GeneART, Thermofisher)
pAP6	pET100D.DDX49 mutated to express DDX49$^{D152A/155A}$	Site directed mutatgenesis of pET100D.DDX49 by Ashley Parkes
pFK1	pET100D.DDX49 mutated to express DDX49$^{D422A/424A}$	Site directed mutatgenesis of pET100D.DDX49 by Fiorela Kapllanaj

Table 2.5.2 List of duplex substrates, and their constituent oligonucleotides used in the biochemical analysis of DDX49.

Substrate name	Oligonucleotide name	Oligonucleotide sequence 5' to 3'
Fork 2A	MW12	TCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14 – 5' Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Fork 2B	MW12 – 5' Cy5	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Fork 3	MW14 – 5' Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA

	MW12	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	PM16	TGCCGAATTCTACCAGTGCCAGTGAT
Fork 4	MW14 – 5' Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	MW12	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	PM17	TAGCAATGTAATCGTCTATGACGTTG
Anisotropy Fork 2	HelQ1 – 5' FAM	TTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
FRET Fork 2	MW12-Cy3	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14-Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
FRET Fork 3	MW12-Cy3	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14-Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	PM16	TGCCGAATTCTACCAGTGCCAGTGAT
FRET Fork 4	MW12-Cy3	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW-14-Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	PM17	TAGCAATGTAATCGTCTATGACGTTG
RNA Fork 2	MW12	UCGAUCCUCUAGACAGCUCCAUGAUCACUGGCACUGGUAGAAU
	MW14 RNA-Cy5	UCAUAGACGAUUACAUUGCUACAUGGAGCUGUCUAGAGGAUCCGA
DNA: RNA Fork 1	5 prime DNA lead Cy5	AACGATTCTCGTTCTACAACTAGGCTCATGG
	5 prime RNA lag	AGAAAAUAUCUUGUAGAACGAGAAUCGUU
DNA: RNA Fork 2	5 prime DNA lag Cy5	GGTACTCGGATCAACATCTTGCTCTTAGCAA
	5 prime RNA lag	UUGCUAAGAGCAAGAUGUUCUAUAAAAGAGG

Table 2.5.3 List of primers and guide RNAs used in the biochemical analysis of DDX49.

Name	Details/Use	Oligonucleotide sequence 5' to 3'
49_exon_3_fw	Forward primer used for the amplification of exon 3 of the *Hsp_DDX49* gene	GAAGGAGGGATGTTCCAGGC
49_exon_3_rev	Reverse primer used for the amplification of exon 3 of the *Hsp_DDX49* gene	AACATTGCTAGGACTGGGCC
49_human_fw	Forward primer used for the pET100D.DDX49 insert	GCAAGCTTATGGCCGGCT
49_human_rev	Reverse primer used for the pET100D.DDX49 insert	GCGAATTCTCAGACGAGTCC
49_APA_FW	Forward primer used for the mutagenesis of pET100D.DDX49 to produce DDX49$^{D422A/424A}$	GGCACTGGAAGCAAAACGTAAAG
49_APA_REV	Reverse primer used for the mutagenesis of pET100D.DDX49 to produce DDX49$^{D422A/424A}$	GGTGCTTTACCTTCCAGGATCAG

49_exon_3_crRNA	CRISPR RNA used for guiding CRISPR-Cas9 mediated gene editing of exon 3 of the *Hsp_DDX49* gene	AAAGACTGCATCATCGTCGGTGG

2.1.6 General solutions and buffers composition

Table 2.6.1 Running Buffers

Stock Solution	Composition
10xTris-borate-EDTA (TBE) buffer	1M Tris 1M Boric acid 20mM EDTA
10xSDS PAGE running buffer	250mM Tris 1.92M Glycine 1% (v/v) SDS

Table 2.6.2 SDS PAGE

Solution Name	**Composition**
4x Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) loading buffer	50mM Tris pH 6.8 2% SDS 10mg Bromophenol blue 10% Glycerol
Coomassie Brilliant Blue Stain	50% Methanol 10% Glacial acetic acid 0.001% Coomassie Brilliant Blue R-250
Destain	20% Ethanol 10% Glacial acetic acid Made up to 1L with distilled water

Table 2.6.3 Substrate preparation

Solution Name	**Composition**
10x Annealing Buffer	100mM Tris pH 7.5 500mM NaCl
Substrate elution Buffer	50% Methanol 10% Glacial acetic acid 0.001% Coomassie Brilliant Blue R-250

Table 2.6.4 *In vitro experimental buffers*

Solution Name	**Composition**
5x Helicase Buffer	100mM Tris pH 7.5 500μg/mL BSA 30% glycerol
Stop Solution	2.5% SDS 200mM EDTA 2mg/mL Proteinase K

2.6.5 *Purification Buffers*

Solution Name	Composition
Ni-NTA Buffer A	25mM Tris pH 7.5 500mM NaCl 25mM Imidazole 10% Glycerol
Ni-NTA Buffer B	25mM Tris pH 7.5 500mM NaCl 400mM Imidazole 10% Glycerol
Heparin Buffer A	25mM Tris pH 7.5 100mM NaCl 10% Glycerol
Heparin Buffer B	25mM Tris pH 7.5 1M NaCl 10% Glycerol
DDX49 Storage Buffer	25mM Tris pH 7.5 200mM NaCl 35% Glycerol 5mM DTT

2.2 Methods

2.2.1 General Microbiology

2.2.1.1 Transformation of competent cells protocol

Competent cells were retrieved from the -80°C freezer and thawed on ice. 10ng of vector plasmid was added to 100μL of competent *Escherichia coli* DH5a/BL21 AI cells and gently mixed via pipetting in a 1.5mL Eppendorf microcentrifuge tube. The solutions were placed on ice for 5 minutes before being heat shocked for 90 seconds in a 42°C water bath (Grant) after which they were placed back on ice for 5 minutes. 900μL LB broth was added and gently mixed to the solution under aseptic conditions (roaring Bunsen burner flame) and incubated for 1 hour in a 37°C water bath (BFL). The cells were pelleted via centrifugation in a (Eppendorf 5254R) for 1 minute at 20,000g and the supernatant was discarded. The pellet was resuspended in 200μL of LB broth under aseptic conditions and plated evenly across a LB broth ampicillin agar plate (15g/L, 50μg/mL ampicillin) using a spreader. The plates were then incubated overnight at 37°C in a plate incubator (Carbolite).

2.2.1.2 Inoculation of overnight Bacterial cultures

Bacterial colonies of transformed *Escherichia coli* DH5a/BL21 AI cells were picked from a LB broth ampicillin agar plate using a 10μL pipette tip and placed inside a glass tube with 5mL LB broth and supplemented with 100mg/mL Ampicillin. The culture tubes were placed overnight in a rotating incubator at 37°C (New Brunswick TC-7 roller drum).

2.2.2 Human Cell Culture

2.2.2.1 Management of Human U2OS cells

U2OS osteosarcoma cells were used in the *in cellulo* analysis of *Hsp_DDX49.* The U2OS cell line was derived from the bone tissue of a 15-year-old Human female, they exhibit epithelial adherent morphology. The cell line was kindly gifted to the Bolt lab by Richard D Wood from the University of Texas.

The cells were kept in the vapour phase of liquid nitrogen, stored as aliquots of approximately $2x10^6$ cells suspended in 1mL DMEM + 10% DMSO. To seed a new flask of cells, an aliquot is thawed, and the cells were transferred to a fresh 50mL centrifuge tube where 19mL pre-warmed DMEM was slowly added; this was then centrifuged at 300g, ramp rate = 4 for 10 minutes. The media was discarded, and the pellet resuspended in 5mL DMEM and transferred to a $25cm^2$ cell culture flask which was kept in a 37°C incubator with 5% CO_2 (Sanyo). Growth of the cells was monitored until it had reached 70% confluency, which was determined via bright-field microscopy (Olympus). Then, the media was removed, and the adherent cells were suspended with 1xTrypsin-EDTA solution that was subsequently quenched with a 1:10 ratio of trypsin: growth media, and the resuspended cells were transferred to $75cm^2$ flask and stored in a 37°C incubator with 5% CO_2 (Sanyo).

Cell cultures stored within $75cm^2$ flasks were kept in the 37°C incubator until they reached 70% confluency, they were passaged by aspirating their media, washing twice with phosphate-buffered saline followed by trypsinisation with 1xTrypsin-EDTA solution. The flask was incubated at 37°C until detachment of the cells from the flask's surface was visible. The trypsin was quenched

with a 1:10 ratio of trypsin to growth media. The desired number of cells was then seeded into a new 75cm^2 flask and incubated at 37°C with 5% CO_2.

When freezing down cell lines for preservation, the cells were seeded onto 100mm cell culture dishes and grown to 70% confluency. The cells were then washed, trypsinised and quenched as described above but were transferred to a 50mL centrifuge tube. Cell number and viability was measured by mixing a 10μL sample with 0.4% Trypan blue, which was added to a haemocytometer for counting. Following this, the cells were resuspended in DMEM with 10% DMSO for a density of 2x10^6 cells per mL, and aliquoted into cryo-tubes. These tubes were first frozen slowly overnight at -80°C and then transferred into liquid nitrogen for long-term storage. All handling of human cells were conducted with aseptic technique in a safety cabinet (HeraSafe)

2.2.3 General DNA and RNA Manipulation

2.2.3.1 Polymerase Chain Reaction

For the amplification of DNA, Q5 polymerase (M0491, NEB) was used along with 2.5μM reverse/forward primers, this was supplemented with the addition of 1mM deoxynucleotide triphosphates (dNTPs) and Q5 reaction buffer (M0491, NEB). The appropriate annealing temperatures for successful amplification was calculated using an online tool (Tm calculator, NEB, https://www.tmcalculator.neb.com/). Generally, for a PCR reaction, the following conditions were used on a Verifi thermocycler: 98°C to denature DNA for 30 seconds, then 35 cycles of: 98°C for 10 seconds, 20 seconds at the calculated annealing temperature and extension at 72°C for 30 seconds per kB DNA, this would be followed by a final extension step at 72°C for 10

minutes. Successful PCR amplification was determined by agarose gel electrophoresis.

2.2.3.2 Site Directed Mutagenesis

Site directed mutagenesis to produce the DDX49$^{D420A/D422A}$ mutant was carried out by the undergraduate student Fiorela Kapllanaj but overseen by Ashley Parkes and me (Sabesan Anandavijayan). To conduct this, PCR was carried using mutant primers designed and ordered by Ashley Parkes and as described above. Following this, the amplified product was run on a 1-2% w/v agarose gel to verify successful amplification. Once verified, bands at the correct size were extracted via the protocol mentioned below. To remove methylated, template DNA, and to circularise the plasmid, a reaction mix containing: DpnI (R0176S, NEB), T4 polynucleotide kinase (PNK, M0201S, NEB) and T4 DNA ligase in 1x ligase reaction buffer was added to the amplified product and left to incubate for 1 hour at room temperature. It was then transformed into *Escherichia coli* DH5a for cloning and plated onto LB broth ampicillin plates and incubated overnight at 37°C.

2.2.3.3 Annealing of DNA and/or RNA oligomers into substrates

All DNA: DNA, DNA: RNA and RNA: RNA substrates used in assays were produced by denaturing 5μM of complementary ssDNA or ssRNA nucleotides at 95°C for 10 minutes and then allowed to cool overnight to room temperature. For their purification, DNA: DNA substrates were run on an 8% Native poly-acrylamide gel for 120 minutes and visualised by scanning with an Amersham Typhoon 5 biomolecular imager which detects their tagged - fluorescent signal. The appropriate sample bands were excised with a scalpel

and purified, by diffusion, over 48 hours in 250μL annealing buffer. DNA: RNA substrates and RNA: RNA substrates were kept on ice after annealing and used as immediately as possible.

2.2.4 Gel Electrophoresis

2.2.4.1 Agarose Gel Electrophoresis

DNA samples were mixed with 6x Purple Gel loading dye (NEB) and loaded onto a 1x TBE 1-2% w/v agarose gel stained with 0.2μg/mL ethidium bromide (EtBr). The gels were run at 140V for 60 minutes using a PowerPac Basic power supply (BioRad) to allow sufficient migration and resolution of the sample bands. Gels were imaged using a U: Genius Bio-imaging system.

2.2.4.2 Agarose Gel Extraction

DNA products were excised from the agarose gel in a dark room using a scalpel and a transilluminator 2040 EV (Stratagene). DNA was extracted using a gel extraction kit (Qiagen) and according to manufacturer's instruction. The DNA was eluted into nuclease-free water.

2.2.4.3 Sodium – dodecyl sulphate poly-acrylamide gel electrophoresis (SDS-PAGE)

Samples containing protein were mixed with 4xSDS loading buffer and 30mM dithiothreitol (DTT), these were then boiled for 10 minutes at 95°C using a heat-block. Samples were run on a 10% acrylamide for 60-90 minutes to allow sufficient migration and resolution of the sample bands. To visualise the protein bands, the gels were stained using Coomassie brilliant blue stain and de-stained using destain buffer. The gels may also be stained with SyproRuby

stain which has a higher sensitivity for detecting proteins, this was destained and imaged via U: Genius Bio-imaging system.

2.2.4.4 Native polyacrylamide gel electrophoresis (Native – PAGE).

Helicase and electromobility shift assays (EMSAs) can be visualised by running sample products on an 8 – 10% acrylamide TBE gel for 1 – 2 hours at 140V. The gel would then be scanned with an Amersham Typhoon 5 biomolecular imager which detects their tagged - fluorescent signal.

2.2.5 In Vitro Experimentation

2.2.5.1 Substrate unwinding gel-based assays (Helicase Assay)

To analyse DDX49's helicase activity, DNA: DNA, DNA: RNA and RNA: RNA unwinding assays involved reaction mixes containing 1xHB and 25nM 5' Cy5 - labelled substrate, and 250nM competitor DNA. Prior to assay optimisation, helicase reactions were supplemented with 5mM ATP, 5mM MgCl2 and 25mM DTT; this was changed to 5mM ATP, 10mM and 25mM DTT following assay optimisation. The desired concentration of DDX49 (or a selected mutant) was then spiked into the reaction and incubated at 37°C for 30 minutes in a water bath. The competitor DNA used was identical to the labelled oligomer in sequence but lacked the fluorescent tag; this prevented substrate re-annealing. Reaction samples were mixed with 6x orange-loading dye and loaded onto a 10% polyacrylamide TBE gel to migrate at 140V for 1 hour for Native-PAGE analysis. After this, the gel was imaged using an Amersham Typhoon 5 biomolecular imager. Images were processed with the software tool ImageJ. Helicase activity was qualitatively compared to a no-

protein substrate control lane, and a sample that had been boiled for 10 minutes at 95°C to represent fully unwound substrate.

2.2.5.2 Electromobility Shift Assay (EMSA)

Electromobility shift assays were used to observe the association of DDX49 (or selected mutants) with various nucleic acid substrates. Reactions were carried out in 1xHB, 25nM substrate and the desired protein concentration. This was then incubated for 10 minutes at 37°C before a sample was taken to be mixed with 6x orange-loading dye and loaded onto an 8% polyacrylamide TBE gel. The gel was run for 120 minutes at 120V and visualised using an Amersham Typhoon 5 biomolecular imager.

2.2.5.3 Fluorescent resonance energy transfer (FRET)

To quantitively observe the unwinding kinetics of DDX49 on DNA: DNA substrates, fluorescent resonance energy transfer was used. Experiments were carried in reaction mixes containing: 1xHB, 50nM DNA substrate, 250nM competitor DNA, 5mM ATP, 10mM MgCl2 and 5mM DTT. These were pipetted into the wells of a black 96xwell Packard microplate, which was inserted into a BMG FLUOstar Omega microplate reader. 500µg/mL Bovine serum albumin was added to the buffers to minimise non-specific binding of proteins. The gain value was then adjusted to the well with the highest signal and DDX49 was spiked into the reaction at the desired concentration. The microplate was incubated inside the FLUOStar at 37°C for up to 30 minutes, and the wells were scanned every minute for up to 30 minutes. The Cy3 probe on the DNA substrate was excited at a wavelength of 540nm, to emit a

FRET signal with a wavelength of 680nm Experiments were repeated in triplicate.

2.2.6 Purification of DDX49 and selected mutants

2.2.6.1 Protein overexpression of DDX49 in Escherichia coli

E. coli BL21 AI was transformed with a plasmid encoding DDX49 or a mutant, this was then plated onto a LB broth ampicillin plate. Single colonies were picked from the plate and used to inoculate 5mL LB broth with 50μg/mL ampicillin, this was then incubated overnight at 37°C. Overnight cultures were used to inoculate fresh LB broth in a 1:100 dilution, supplemented with fresh ampicillin and grown to OD_{600} of 0.6. DDX49 protein expression was induced upon the addition of 0.2% L-arabinose and 0.5mM isopropyl B-D-1-thiogalactopyranoside (IPTG) and cultured overnight at 18°C. Successful overexpression was verified via SDS-PAGE analysis.

This method was scaled up 36x for protein purification. Following overnight expression, 1mL samples were taken for analysis before and after inducing expression. The rest of the cultures were centrifuged in an Avanti J-26 XP centrifuge using a JLA 10.500 rotor for 10minutes at 4000g at 4°C, the biomass was pelleted and resuspended in Ni-NTA buffer A and stored at -80°C until further use.

2.2.6.2 Purification of DDX49

For purification, frozen biomass was thawed on ice and sonicated for 8 minutes, in 10 second pulses on ice. The lysed samples were then clarified by centrifugation (Avanti JA-25-50) at 35,000g for 35 minutes. Protein purification

was carried out using an AKTA Start system (GE Healthcare Life Sciences): following clarification, the supernatant was decanted, and the pellet was discarded. The supernatant was loaded onto a 5mL Ni-NTA column at a rate of 1mL per minute. Unbound material was collected for analysis and the column was washed with Ni-NTA buffer A until the chromatogram reached baseline. Fractions were eluted in Ni-NTA buffer B using an Imidazole gradient up to 100%.

Fractions positive for DDX49 were determined by SDS-PAGE, pooled, and dialysed overnight at 4°C in Heparin buffer A. The following day, the dialysed sample was clarified via centrifugation (Eppendorf centrifuge 5430 R) to remove any precipitates. The sample was then loaded onto a 1mL Heparin column. The unbound material was collected, while the sample was column washed in Heparin buffer A until the chromatogram reached baseline. Fractions were eluted in Heparin buffer B in a gradient of increasing ionic strength and were collected for analysis. DDX49-positve fractions were once again determined by SDS-PAGE and these were pooled together for overnight dialysis in DDX49 storage buffer. The purified DDX49 protein was then aliquoted and stored at -80°C. Protein concentrations in μM were determined by absorption readings on a DeNovix nanodrop spectrophotometer using Beer lambert's law and an extinction coefficient of 27350$M^{-1}cm^{-1}$.

2.2.7 CRISPR-Cas9 mediated DDX49 knockout in the U2OS cell line

2.2.7.1 Formation of CRISPR-guide RNA

For the formation of the guide RNA used to direct CRISPR-Cas9 mediated gene editing, both the CRISPR-RNA (crRNA) and the trans-activating CRISPR RNA (tracrRNA) were mixed together to a final concentration of 1μM and allowed to anneal at 95°C for 5 minutes before being cooled to room temperature, as per manufacturer's instructions. As a control to verify successful Cas9 RNP transfection into U2OS cells, ATTO dye-labelled tracrRNA was also annealed with crRNA in a separate reaction, as per manufacturer's instructions. If the Cas9 RNP with the ATTO-labelled tracrRNA is successfully incorporated into the cell, it is detectable by fluorescent microscopy.

2.2.7.2 Transfection of CRISPR-Cas9 RNP into U2OS cells

Prior to transfection, U2OS cells were grown to 80% confluency, washed, trypsinised and resuspended in DMEM; They were then centrifuged at 300g for 10mins. The supernatant was discarded, and the pellet was resuspended in an appropriate volume of DMEM so that the cells could then be seeded 40,000 per well of a 24-well microplate. After seeding, the cells were incubated overnight at 37°C, so they may settle and adhere to the well's surface.

Cas9 and the annealed guide RNA were mixed to a 1:1 ratio in serum free media (SFM), this was followed by the addition of 2.5μL of Ca9Plus reagent

(Thermofisher); the mix was incubated at room temperature for 5 minutes. Separately, the CRISPRMax reagent was diluted according to manufacturer's instructions and then the Cas9 RNP mix was added to this in a 1:1 ratio; it was then incubated for 20 minutes at room temperature. The transfection mix was added to the seeded cells in the plate so that the final RNP concentration in the well was 11nM. The 24-well plate was then incubated for 48 hours at 37°C.

2.2.7.3 Post transfection analysis

Following transfection, ATTO488 expression in control cells were visualised using an Axiovert S100. Once successful transfection was confirmed, the cells were grown to 70% confluency in the 24-well plate before being removed and seeded into a 96-well plate at 5 cells per well. The 96-well plate was incubated at 37°C to allow single cell colonies to develop.

Wells with single cell colonies were examined using an Axiovert S100, and once verified, allowed to be split into a 48-well plate. Wells with more than one colony were discarded. Cells seeded into the 48-well plate were incubated at 37°C, and their genomic DNA was extracted using a PureLink Genomic DNA Mini-kit according to manufacturer’s instructions. These were amplified via PCR and sent for Sanger sequencing (Deepseq, University of Nottingham) to verify successful gene editing. TIDE analysis was used on the sequencing results, a bioinformatics method used to evaluate gene editing efficiency by comparing the edited sequence to the original, unedited sequence.

Once sufficiently grown, cells from selected wells in the 48-well plate were split into a 25cm^2 flask and then a 75cm^2 flask. These cells were maintained and also preserved in liquid nitrogen as mentioned above.

2.2.8 In cellulo phenotyping

2.2.8.1 WST-1 Cellular Proliferation Assay

To determine the cellular proliferation of U2OS cells that have heterozygous expression of DDX49, WST-1 cell proliferation reagent (Sigma) was used. The assays were carried out to manufacturer's instructions: both edited and unedited ("wild-type") cells were seeded into seven 96-well plates at 40,000 cells per well (total well volume was 100μL), and then incubated at 37°C for up to 196 hours. Following the incubation periods, 10μL of WST-1 reagent was added to the cells and incubated for 4 hours at 37°C. The absorbance of light at 440nm was read using a BMG FluoStar Omega microplate reader. Background absorbance was corrected with a media-only blank.

2.2.8.2 Wound-healing assay

To investigate the cellular migration capability of the DDX49-heterozygous U2OS phenotype, wound healing assays were used. Wild-type and edited U2OS cells were split into separate 100mm cell culture dishes and incubated at 37°C until they reached 100% confluency. Upon reaching full confluency, the cells were washed with PBS and the media was replaced. Then, a sterile 20-200μL pipette tip was used to scratch a straight line on the surface of the dish, creating a gap between adherent cells. The cell culture dishes were incubated for up to 72 hours and imaged every 24 hours using an Axiovert

S100 to monitor and measure cellular migration across the gap. Experiments were repeated in triplicate.

Chapter 3: Overexpression, purification, and biochemical analysis of DDX49

3.1 Introduction

DDX49 has been reported to be a DExD-box RNA helicase, it can unwind double-stranded RNA substrates in the presence of ATP.[10] Following bioinformatic analysis, key motifs of interest within the structure of DDX49 had been identified, in particular, the DPD motif was in a region of high structure disorder. In addition, DDX49's interaction with DNA and RNA substrates has been poorly elucidated.

3.2 Aims

The aims of this work were to purify Hsp_DDX49 from the bacterial cell lysate of *E. coli* BL21 AI, and also to conduct mutational analysis of DDX49 by purifying select variants of DDX49 that have certain functional motifs inactivated. One mutant protein ($DDX49^{D422A/424A}$) was purified by an undergraduate research student Fiorela Kapllanaj under mine and PHD student Ashley Parkes' supervision.

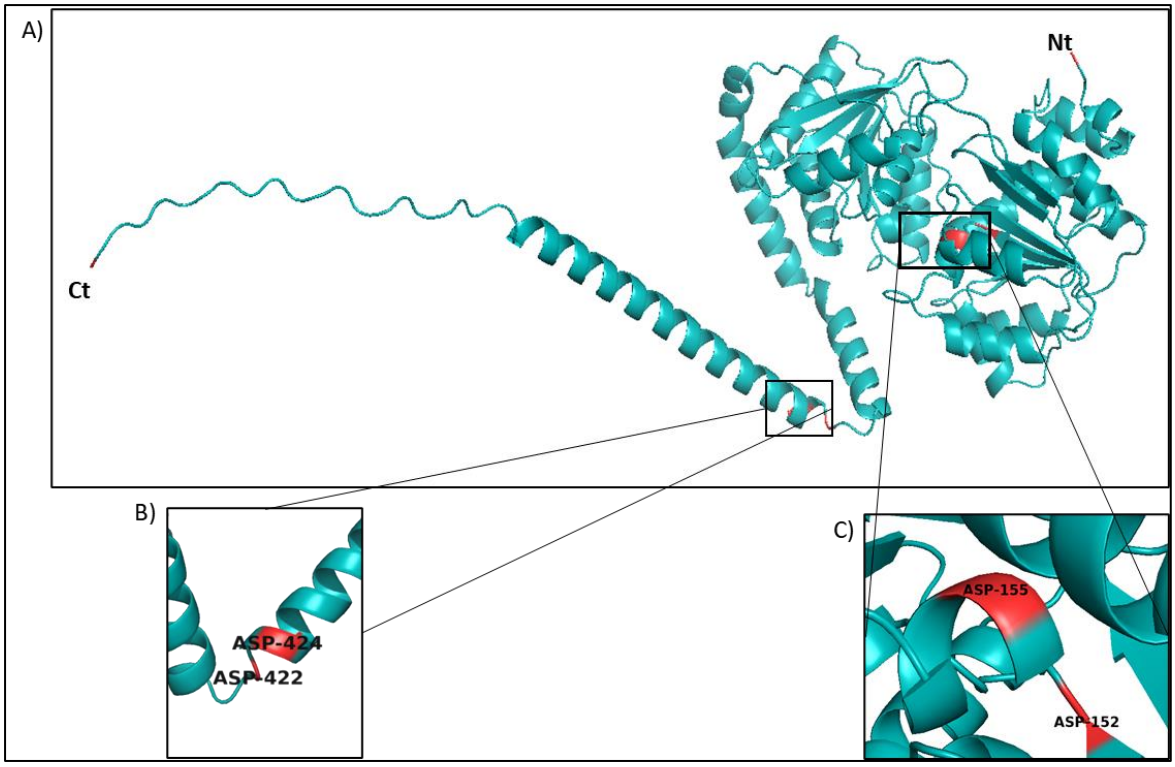

Figure 8: A) The crystal structure of DDX49 predicted by AlphaFold[15], with residues of interest expanded and highlighted in red for mutagenesis. B) D422 and D424 of the DPD region. C) D152 and D155 of the Walker B motif region. *Crystal structure imaged using the PyMol Molecular Graphics System, Version 2.0 Schrödinger, LLC*[17]

3.3 Overexpression and purification of DDX49 and select mutants

3.3.1 Overexpression of DDX49

To overexpress DDX49, the pET100D plasmid was transformed into *E. coli* BL21 AI, a strain derived from *E. coli* BL21 cells but modified to allow T7 promoter induced expression. Cultures of transformed *E. coli* BL21 AI were grown in 2L baffled flasks at 37°C until OD_{600} of 0.6, where IPTG and L-arabinose was added. The flasks were left to shake overnight at 18C prior to harvest. The addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), a molecular mimic of allolactose, binds to the *lac* repressor and in an allosteric manner, releases it from the *lac* operator. This allows access of the T7 promoter in pET100D by T7 RNA polymerase, therefore allowing high

expression of DDX49. SDS-PAGE confirmed successful overexpression (Figure 9). This method was also used to overexpress selected mutants of DDX49.

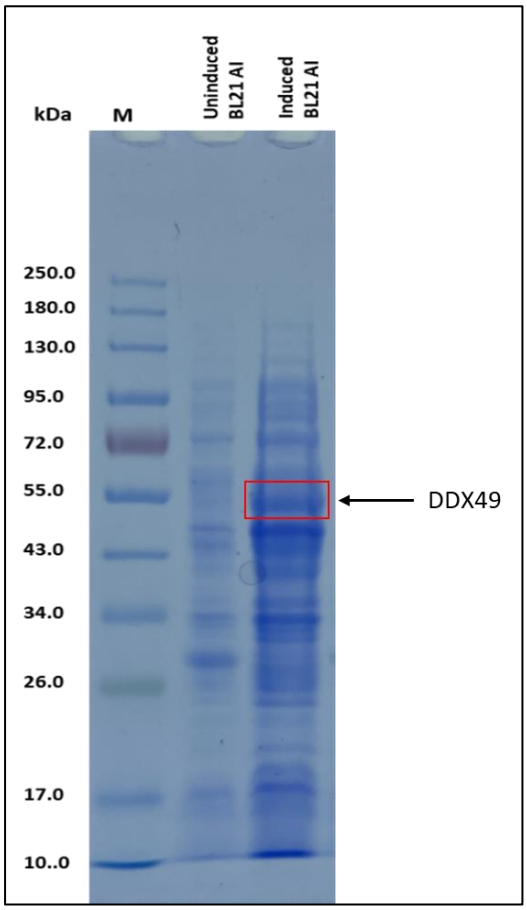

***Figure 9: Successful overexpression of "wild-type" DDX49 by BL21 AI E. coli**, as shown by SDS-PAGE on a Coomassie stained 10% Acrylamide gel.*

3.3.2 Purification of DDX49

To purify DDX49, harvested biomass were lysed via sonication and clarified via centrifugation; the supernatant was loaded onto a Ni-NTA column for purification by affinity chromatography. For ease of purification, the amino acid sequence of DDX49 and its mutants have been modified to include a poly-histidine tag (6xHis-tag) on their N-terminus, this has a strong affinity for Ni^{2+} ions thus allowing them to remain associated with the resin beads in the

NTA column while other material can flow through and be discarded. 6xHis-tagged proteins can be eluted from the column through the addition of imidazole; this is usually added gradually so that non-target proteins that may be weakly associated with the column resin can be eluted first. Imidazole bears a similar structure to histidine and can outcompete the affinity of histidine with Ni^{2+} ions.

DDX49 is expected to have a mass of approximately 55kDa, protein purification fractions were analysed through SDS-PAGE to check for DDX49's presence. Wild-type DDX49 bound well to the Ni-NTA column albeit with a few contaminating proteins (Figure 10A). DDX49-positive fractions were pooled and dialysed overnight in Heparin buffer, concentrating the eluted proteins which were then loaded onto a heparin column. Heparin columns contain highly sulphated polysaccharides which can bind to molecules via electrostatic interactions between its sulfate groups and negatively charged residues on the target molecule. A high concentration of NaCl can disrupt these interactions within the heparin column, therefore an increasing concentration of NaCl was applied to the column to elute DDX49. Chromatographic traces were used to aid in identifying potential DDX49-positive fractions from each column (Figure 10B, D).

The selected mutants of DDX49: $DDX49^{D152A/D155A}$ and $DDX49^{D422A/D424A}$ were purified in the same manner as the wild-type. $DDX49^{D422A/D424A}$ was purified by undergraduate research student Fiorela Kapllanaj for her final year research project, under mine and PHD student Ashley Parkes' supervision.

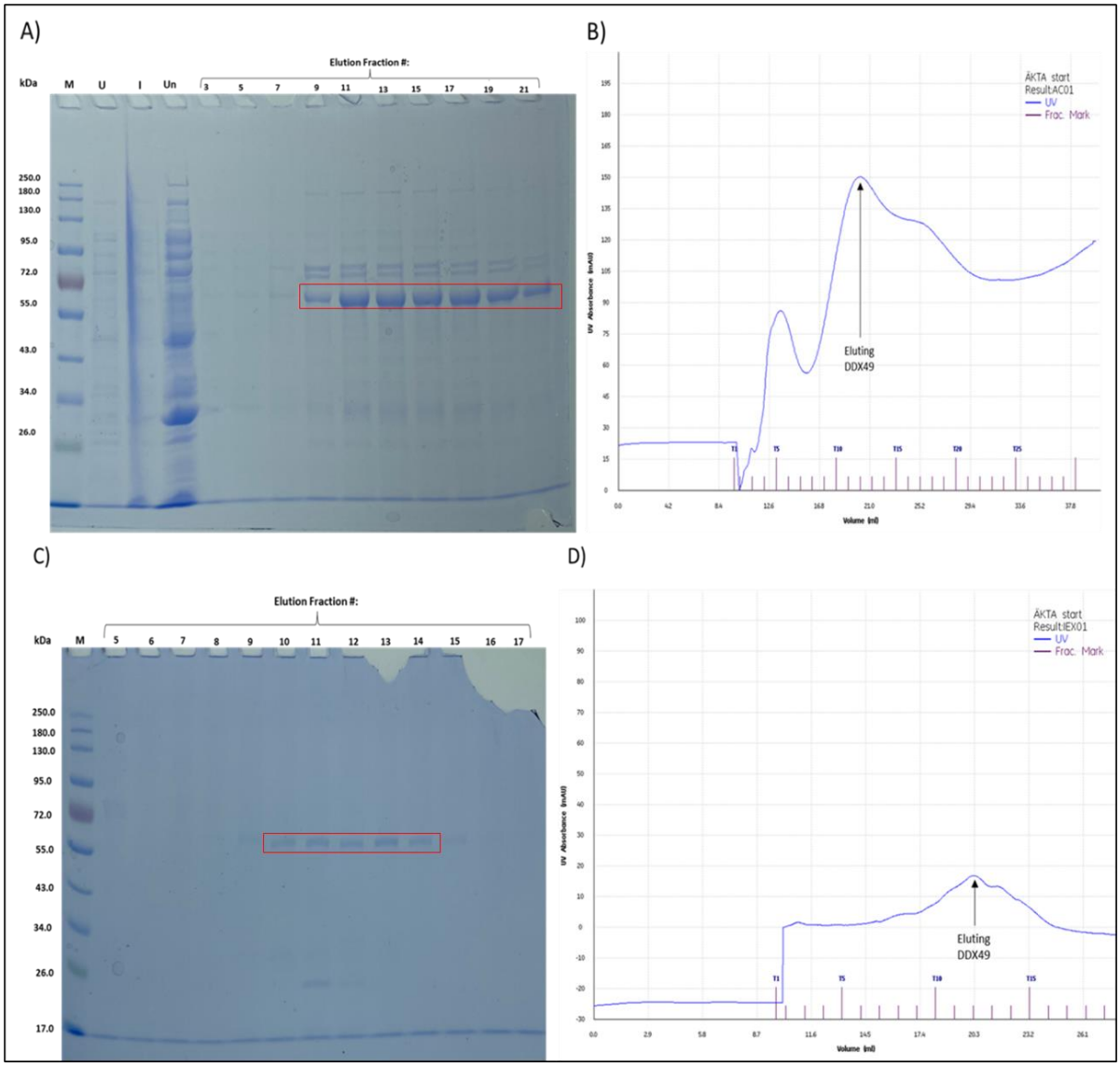

Figure 10: Purification of wild-type DDX49, *eluted fractions from Ni^{2+} - NTA affinity chromatography were loaded onto A) an 8% SDS-PAGE gel along with the samples from the uninduced (U) and induced (I) bacterial protein lysate and proteins not bound to the affinity column. Fraction samples that had been loaded onto the gel were determined by B) a chromatographic trace which recorded the absorbance of UV wavelengths at 280nm by eluting proteins. Selected DDX49-positive fractions (encircled in red) were pooled, dialysed, and purified further via C) Heparin affinity chromatography, of which eluted fractions to be loaded onto an 8% SDS-PAGE gel was determined by D) a chromatographic trace which recorded the absorbance of UV wavelengths at 280nm by eluting proteins.*

Purified wild-type and the mutated DDX49$^{D152A/155A}$ and DDX49$^{D422A/424A}$ were verified and compared against each other through SDS-PAGE (Figure 11). The purified wild-type DDX49 and DDX49$^{D152A/155A}$ mutant were both shown to contain a contaminant whereas DDX49$^{D422A/424A}$ did not. The presence of this contaminant was inconsistent across multiple, independent purifications of DDX49 and in SDS-PAGE gels, so it is possible the contaminant arose due to variations in the methods of purification or analysis.

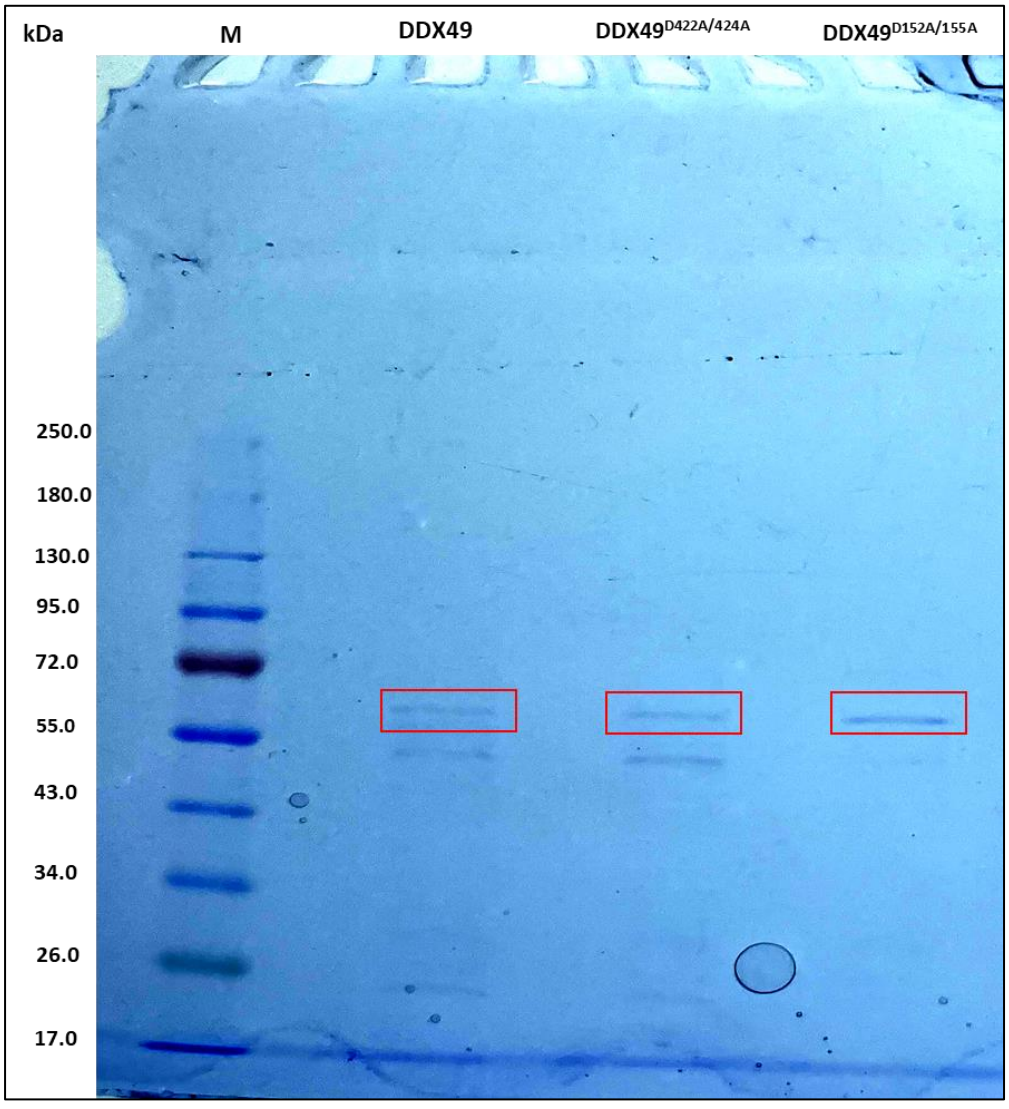

***Figure 11: Wild-type DDX49 and the selected mutants**: DDX49$^{D152A/D155A}$ and DDX49$^{D422A/D424A}$ on a Coomassie stained, 8% SDS-PAGE gel. A marker (M) was also loaded to determine the size of proteins on the gel. The position of DDX49 on the gel is encircled in red.*

3.4 Biochemical analysis of DDX49

3.4.1 Analysis of DDX49 nucleic acid substrate binding

Electromobility shift assays (EMSAs) were used to determine the binding affinity of DDX49 for DNA and RNA substrates. Reactions were loaded onto an 8% TBE gel and subjected to Native polyacrylamide gel electrophoresis to allow the separation of protein-bound substrates from unbound substrates. Substrates subjected to an electrical charge migrate down the TBE gel based upon their size, therefore protein-bound substrate migrate slower than unbound proteins. Nucleic acids were labelled with a Cy5-fluorescent probe on the 5' end of an oligonucleotide, allowing their identification when scanned with an Amersham Typhoon 5 biomolecular imager.

First, wild-type DDX49 was titrated into reactions containing either Fork 2A dsDNA or Fork 2 dsRNA. DExD-box proteins are typically ATP-dependent RNA helicases, but some members of the family, such as DDX43, display affinity for DNA substrates.[38] DDX49 displayed a higher preference for dsRNA substrates than dsDNA, however, it was also observed that the dsRNA substrate had also separated into its single-stranded constituents, raising the possibility that DDX49 had been binding to ssRNA molecules in the reaction mixes (Figure 12).

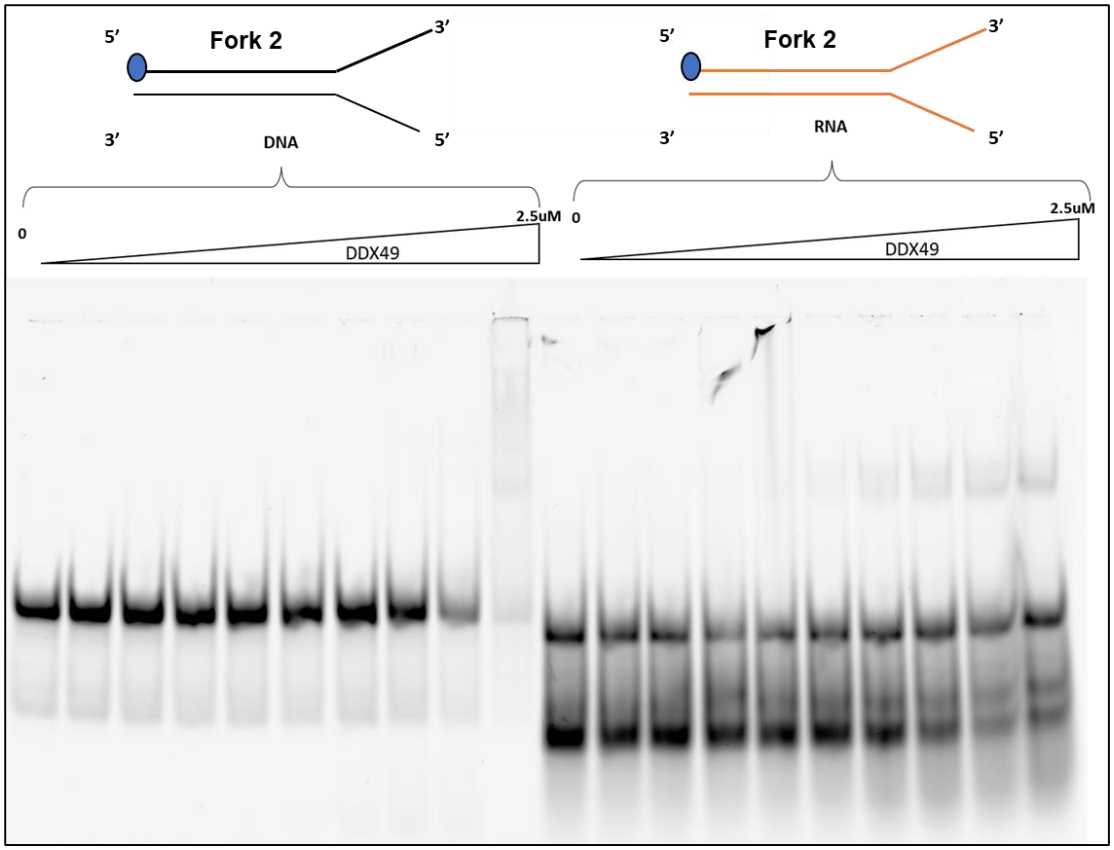

Figure 12: Electromobility shift assay (EMSA) shows that DDX49 may prefer forked, double - stranded RNA over forked, double-stranded DNA. *DDX49 was spiked into 20uL reactions pools containing: 1x HB, 25nM substrate and 5mM DTT. The reactions were incubated for 10 minutes at 37°C. 5' Cy5-labelled substrates that is not bound by DDX49 migrates further down the gel than bound substrate.*

Therefore, DDX49 was titrated into reactions containing either 5' Cy5-labelled MW14 ssDNA or 5' Cy5-labelled ssRNA. Here, it was shown that DDX49 could bind to either substrate but had a larger preference for ssRNA than ssDNA (Figure 13).

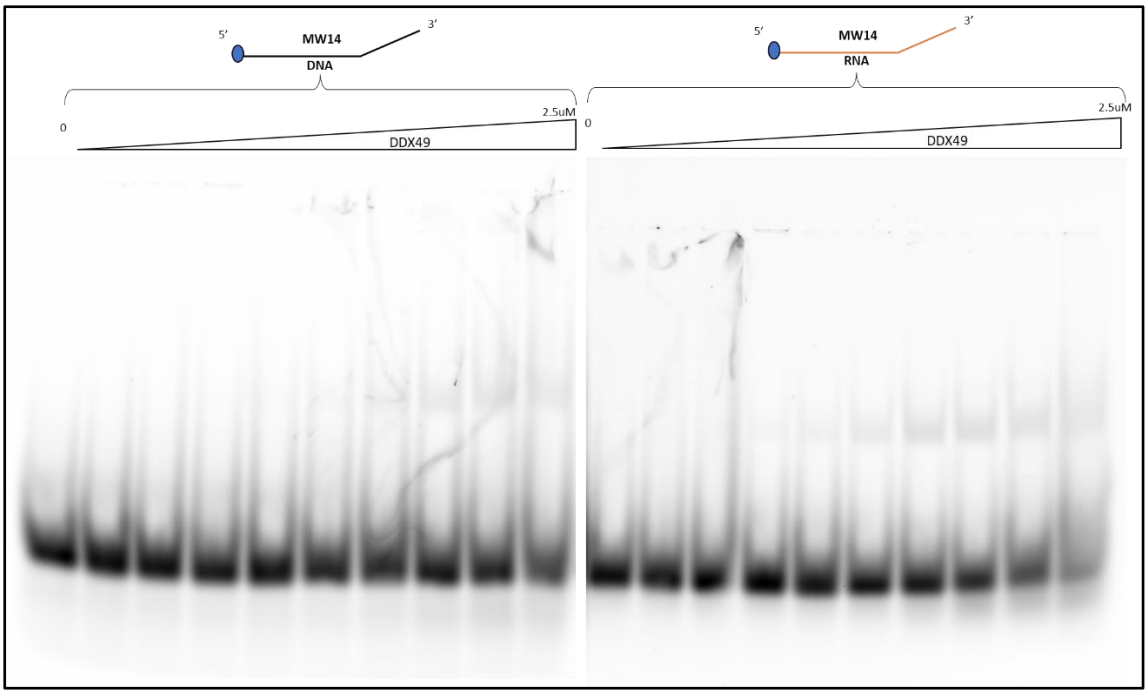

Figure 13: Electromobility shift assay (EMSA) shows that DDX49 prefers single stranded RNA over single-stranded DNA. *DDX49 was spiked into 20uL reactions pools containing: 1x HB, 25nM substrate and 5mM DTT. The reactions were incubated for 10 minutes at 37°C and then the reaction mixes were loaded onto an 8% acrylamide gel. 5' Cy5-labelled oligonucleotides that are not bound by DDX49 migrates further down the gel than bound oligonucleotides.*

To determine the binding characteristics of the mutant DDX49$^{D422A\&D424A,}$ this was titrated into reactions containing Fork 2 dsDNA and compared against the wild-type. The mutant was initially observed at first to not associate with the substrate, but only did so when 5mM ATP and 10mM $MgCl_2$ was added to the composition of the TBE gel and TBE buffer; the wild-type associated with the substrate in both cases (Figure 14). It could be theorised that the accessory domain of DDX49 can bind nucleic acid substrates independently of ATP, but association of the protein's core domain may require the cooperative binding of ATP.

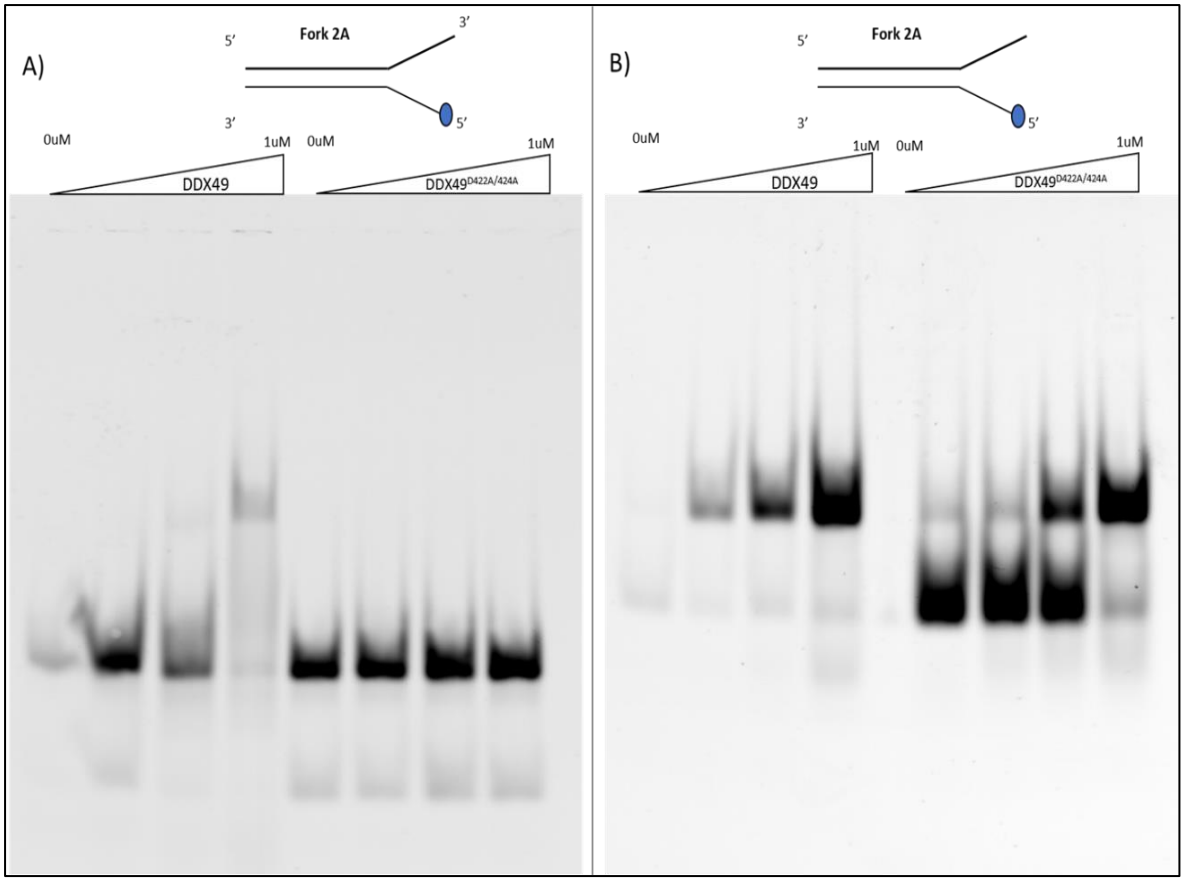

Figure 14: A) DDX49$^{D422A/D424A}$ does not bind to Fork 2A DNA unless B) in the presence of ATP and $MgCl_2$. *5' Cy5-labelled DNA that is not bound by DDX49 migrates further down the gel than bound DNA. After incubation for 30 minutes at 37°C, reaction mixes were loaded onto an 8% acrylamide gel.*

3.4.2 Optimisation of DDX49-mediated unwinding of DNA substrates

Originally, DDX49 unwinding assays were conducted with an equimolar ATP: $MgCl_2$ concentration and 25mM DTT. Nuclease activity alongside helicase activity by DDX49 was observed but was inconsistent across experimental repeats. Therefore, it was necessary to optimise the reaction conditions of DDX49 to achieve consistent results. In reaction pools of 1µM DDX49 and 25nM Fork 2A DNA substrate, either ATP (Figure 15A) or $MgCl_2$ (Figure 15B) was titrated in at increasing concentrations while all other reaction constituents remained constant. A 1:2 ratio of ATP:$MgCl_2$ was deemed the optimal conditions for obtaining consistent helicase and nuclease results on

Fork 2A DNA substrates by DDX49 and was used in all further unwinding assays. Then, DDX49 was titrated in at increasing concentration with the optimised helicase reaction conditions (Figure 15C).

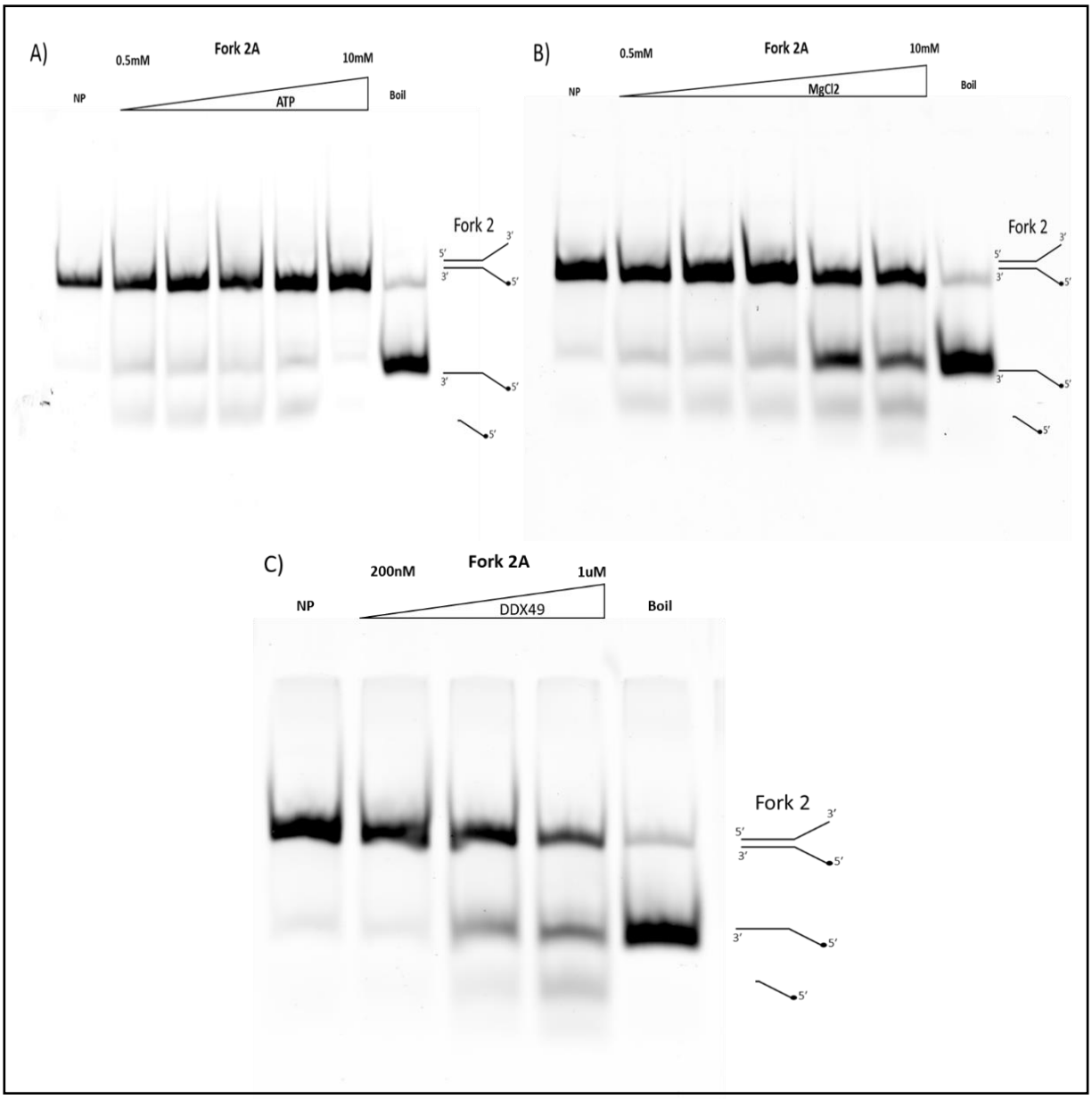

Figure 15: Optimisation of helicase assay conditions for DDX49 and 25nM 5' Cy5-labelled Fork 2A DNA*. A) Up to 10mM ATP was titrated into reaction mixes containing: 1μM DDX49, 1xHB, 2mM $MgCl_2$ and 25mM DTT. (B) Up to 10mM $MgCl_2$ was titrated into reaction mixes with 1μM DDX49, 5mM ATP and 25mM DTT. (C) Once optimised, up to 1μM DDX49 was titrated into reactions with 5mM ATP, 10mM $MgCl_2$, 25nM Cy5-Fork 2A DNA and 25mM DTT. After incubation for 30 minutes at 37°C, reaction mixes were loaded onto a 10% acrylamide gel.*

3.4.3 Analysis of dsDNA unwinding by DDX49 and selected mutants

Helicase assays were used to determine several biochemical characteristics of DDX49. First, the translocation polarity of DDX49 was determined with by adding the protein to reactions containing either Fork 3 dsDNA or Fork 4 dsDNA (Figure 16). Fork 3 is blocked on its 3' end with the PM16 oligonucleotide while Fork 4 is blocked on its 5' end with the PM17 oligonucleotide. DDX49 was shown to unwind Fork 4 DNA but not Fork 3 DNA, indicating that it translocates in a 3' to 5' direction, therefore it is a type A translocase.

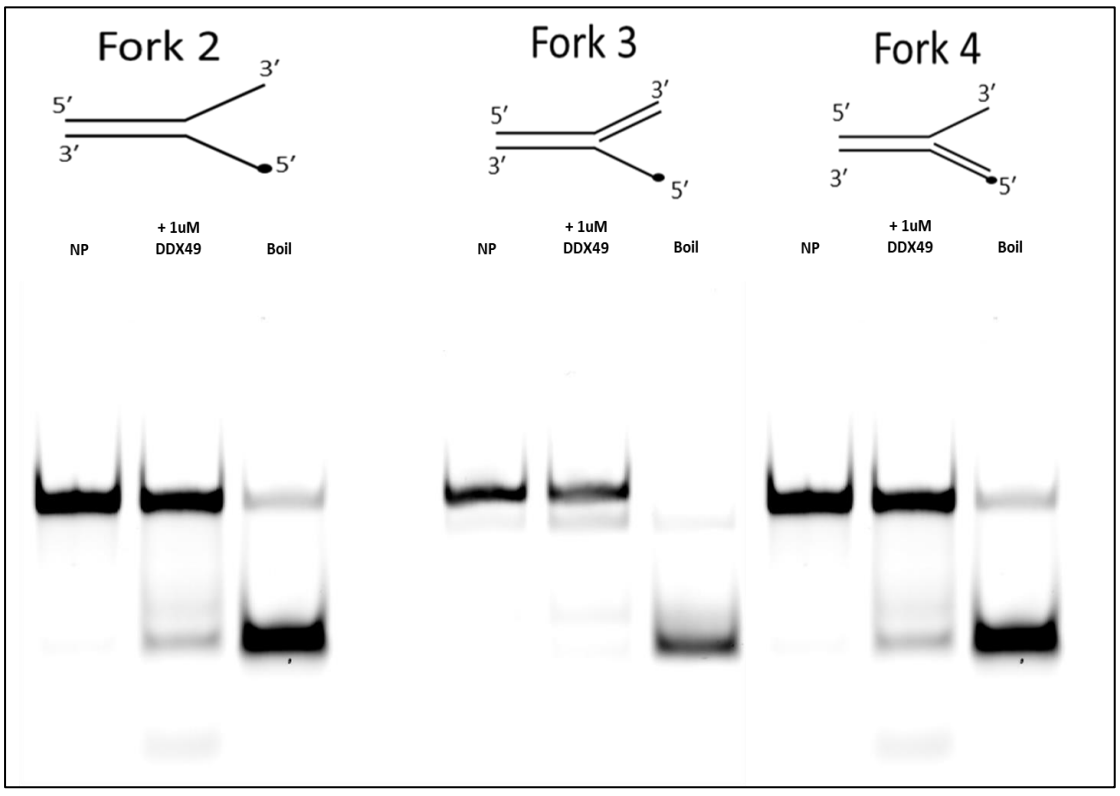

Figure 16: DDX49 translocates in a 3' to 5' direction. *1μM DDX49 was added to reaction pools with: 1xHB, 5mM ATP, 10mM $MgCl_2$, 25mM DTT, 250nM unlabelled MW14 ssDNA and 25nM of either 5' Cy5 – labelled Fork 2A, Fork 3 or Fork 4 DNA. After incubation for 30 minutes at 37°C, reaction mixes were loaded onto a 10% acrylamide gel.*

The unwinding capability of wild-type DDX49 was compared against the purified mutants. $DDX49^{D152A\&D155A}$ was mutated in the Walker B motif in the core domain of DDX49, therefore inhibiting its ATP hydrolysis ability. $DDX49^{D422A\&D424A}$ was mutated in the accessory domain, it was hypothesised that this may inhibit DDX49's novel nuclease activity that was seen in the wild-type. To test this, DDX49 or a mutant was added into reaction pools with 25nM Fork 2A DNA, 5mM ATP, 10mM $MgCl_2$, 25mM DTT and 250nM unlabelled MW14 ssDNA. The reactions were incubated for 30 minutes, and stopped with proteinase K. To determine if ATP hydrolysis was necessary for substrate unwinding, ATPγS was used instead of ATP in a separate reaction mix. Both mutants of DDX49 were shown to unwind Fork 2A substrates, even with the non-hydrolysable analogue ATPγS. Nuclease activity was observed in the wild-type and the Walker B mutant ($DDX49^{D152A\&D155A}$), but not in the accessory domain mutant. This confirms the hypothesis that the accessory domain is responsible for nuclease function. It can also be suggested that the nuclease ability of DDX49 is ATP-independent as it has been shown to cut the substrate in the absence of ATP or ATPγS. This assay was repeated with multiple, independent purification batches of DDX49 and the selected mutants to ensure that the observed activity is consistent and not due to a contaminant in a single batch.

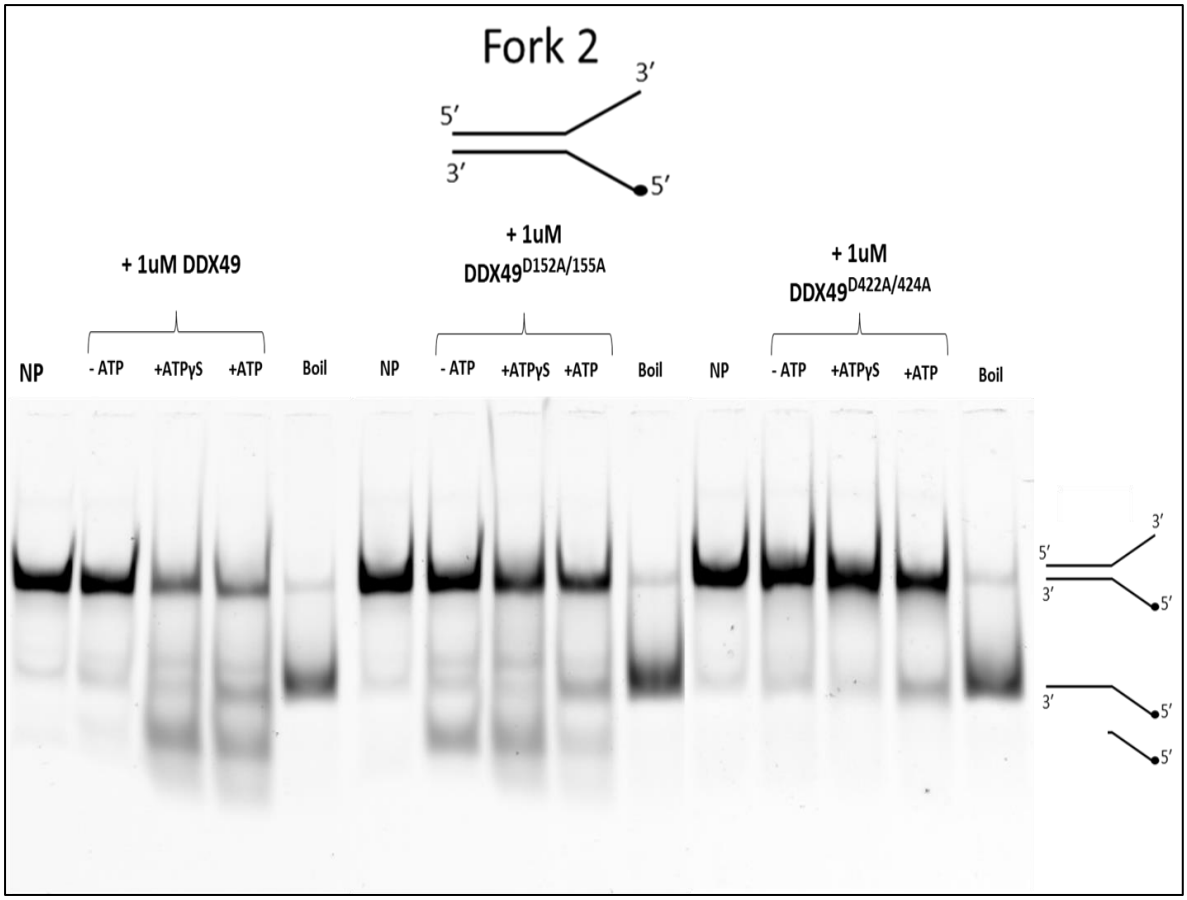

Figure 17: Mutation of DDX49 at residues Asp-422 and Asp-424 inhibits its nuclease function. *Either DDX49, DDX49$^{D152A/155A}$ or DDX49$^{D422A/424A}$ was added to reaction pools with: 1xHB, 25nM 5' Cy5-labelled Fork 2A DNA, 10mM $MgCl_2$, 250nM unlabelled MW14 ssDNA and 5mM of ATP, or its non-hydrolysable analogue ATPγS. Unwound Fork 2A DNA and cut DNA fragments travel faster down the gel.*

To ascertain the size of fragments cut from the Fork 2A substrate by DDX49 as a result of nuclease activity, the reactions were loaded onto a 15% acrylamide Urea denaturing gel and subjected to electrophoresis (Figure 18). A marker ladder containing Cy5-labelled DNA of varied sizes was also loaded onto the gel. Cut fragments migrated further down the gel than uncut fragments.

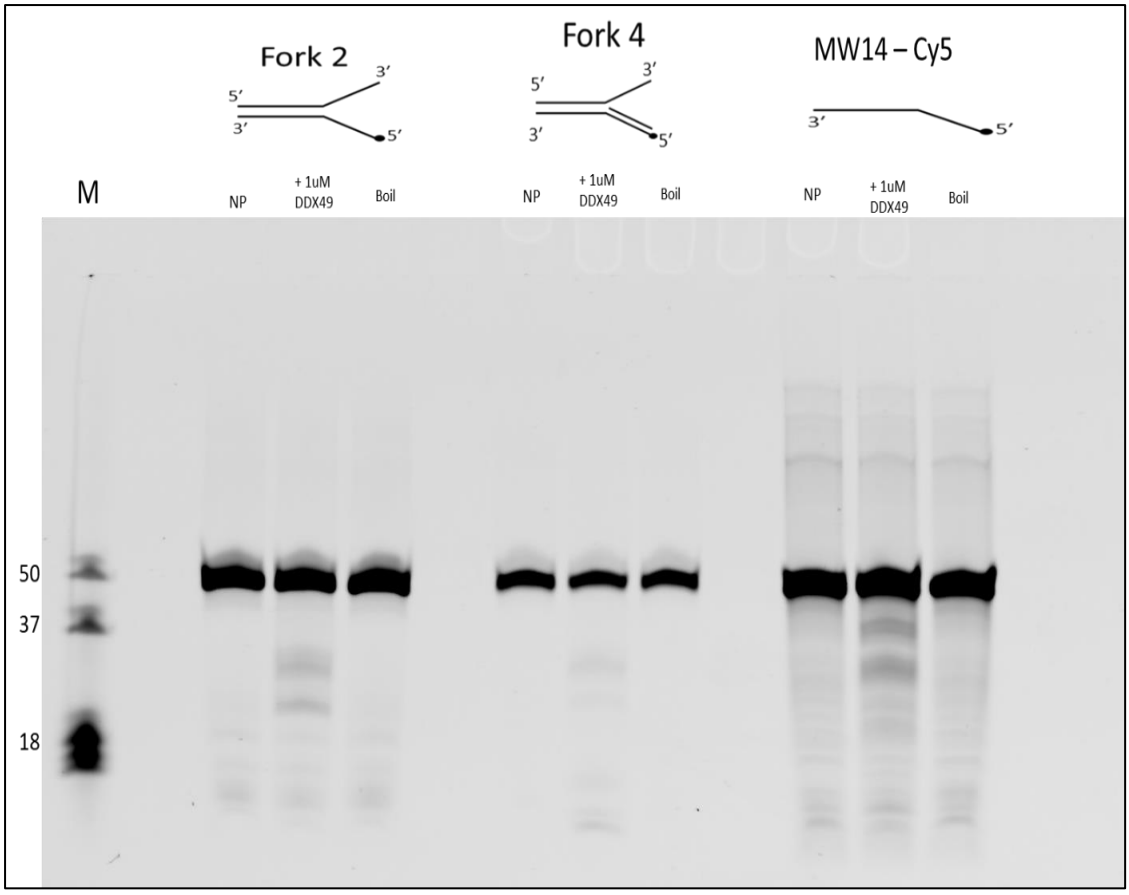

Figure 18: DDX49 cuts Fork 2A DNA to create 20bp sized DNA fragments*. DDX49 was added to reaction mixes containing: 1xHB, 10mM $MgCl_2$, 250nM unlabelled MW14 ssDNA, 5mM of ATP and 25nM 5' Cy5-labelled Fork 2A DNA or Fork 4 DNA or MW14 ssDNA. The reactions were incubated for 30 minutes at 37°C, before being boiled and loaded onto a Urea, 15% acrylamide denaturing gel.*

The unwinding ability of DDX49 was compared between Fork 2 dsDNA and Fork 2A dsRNA substrates. Wild-type DDX49 or the accessory domain mutant was added to reaction mixes containing 25nM Fork 2 dsDNA or Fork 2 dsRNA, 1xHB, 5mM ATP, 10mM $MgCl_2$, 25mM DTT and 250nM unlabelled MW14 ssDNA or MW14 ssRNA (Figure 19). There appears to be a preference for RNA substrates by DDX49 although Fork 2 dsRNA seems to separate even without the addition of the DDX49 helicase. Nuclease activity is also present in both reactions indicating it is not specific to either RNA or DNA substrates.

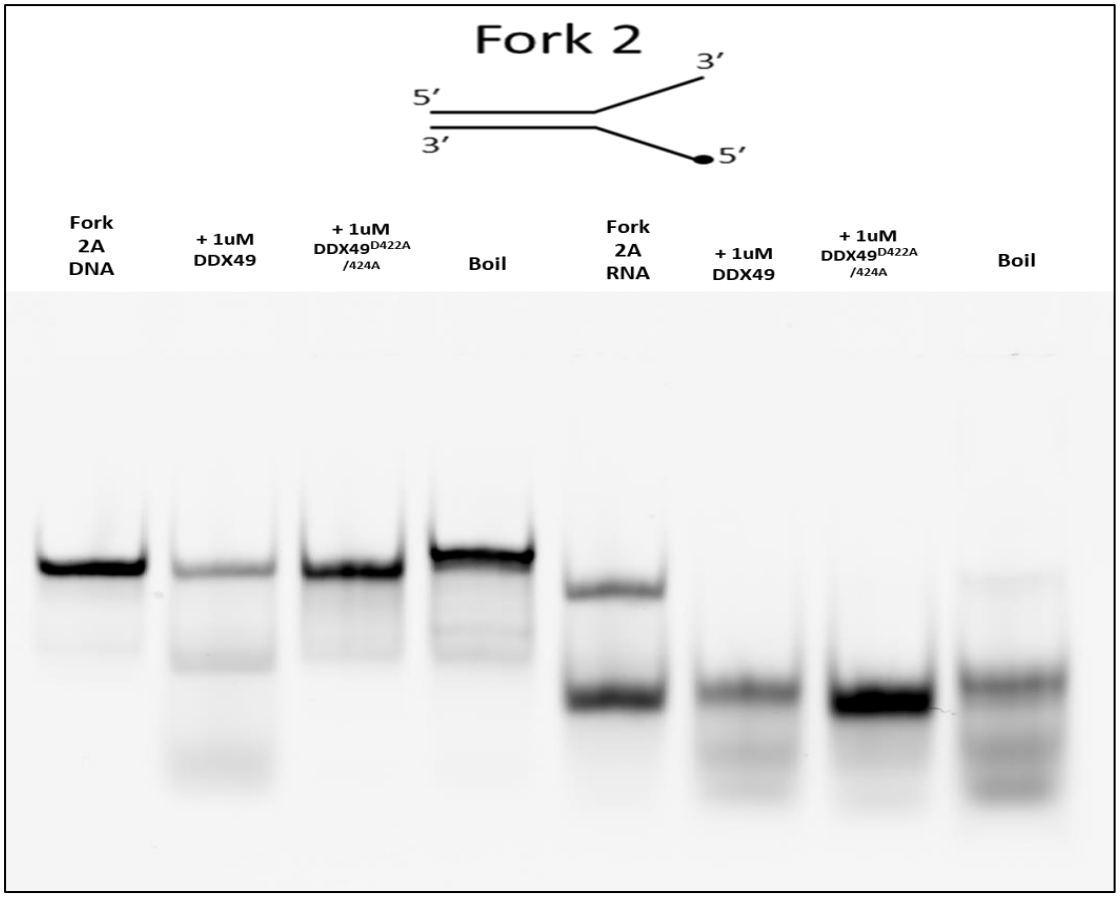

Figure 19: DDX49 unwinds both DNA and RNA substrates*. DDX49 was spiked into reactions mixes containing: 1xHB, 5mM ATP, 10mM $MgCl_2$, 25mM DTT and 25nM Fork 2A DNA or RNA substrate. The reactions were incubated for up to 30 minutes at 37°C before being loaded onto a 10% acrylamide gel for electrophoresis.*

To determine if DDX49 can unwind DNA: RNA hybrids, the wild-type protein or $DDX49^{D422A/424A}$ were added to reaction mixes containing 25nM Cy5-labelled DNA: RNA Fork 1, 1xHB, 5mM ATP, 10mM $MgCl_2$, 250nM competitor DNA and 25mM DTT (Figure 20). The DNA: RNA hybrid substrate was annealed using a thermocycler and used as immediately as possible. The reactions were incubated at 37°C and for 30 minutes. Both proteins were able to unwind the hybrid substrate, and the wild-type protein exhibited nuclease activity that had also been observed when DDX49 interacted with DNA: DNA and RNA: RNA substrates. However, some of the substrate had unwound without the addition of DDX49 or $DDX49^{D422A/424A}$.

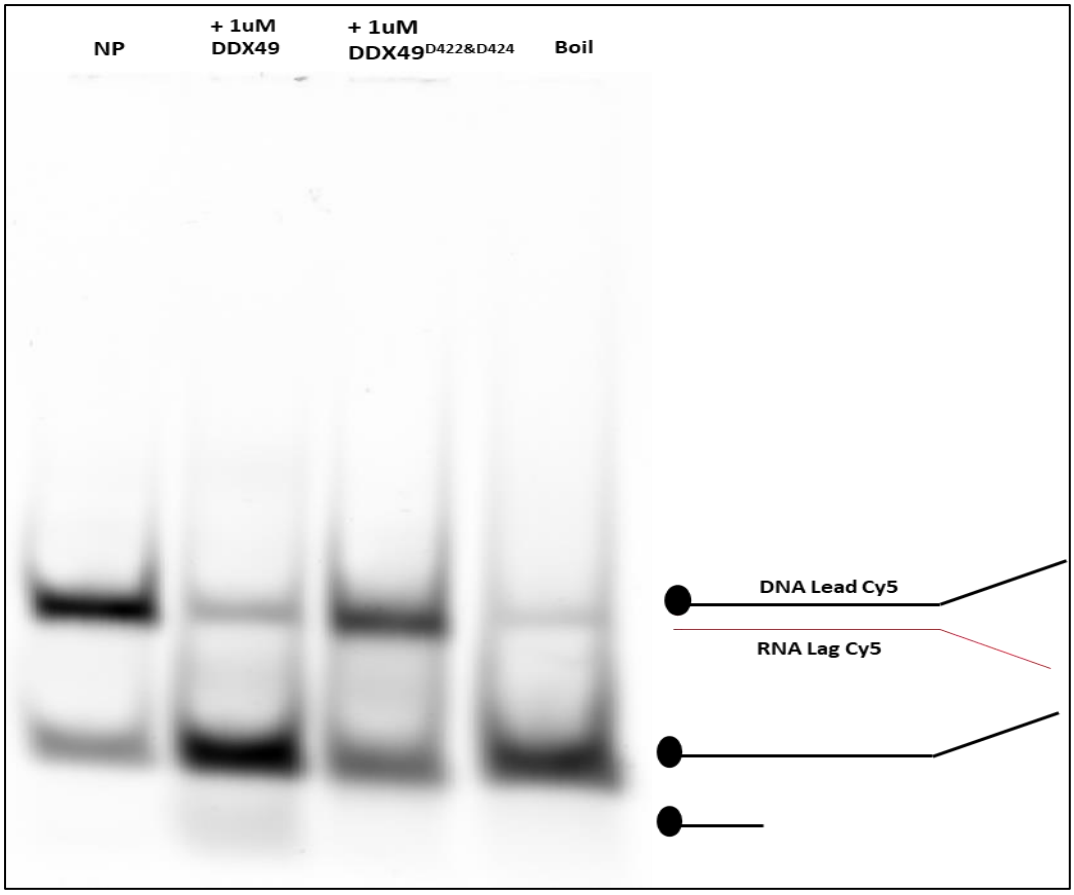

Figure 20: DDX49 can unwind DNA: RNA hybrid substrates*. DDX49 was spiked into reactions mixes containing: 1xHB, 5mM ATP, 10mM $MgCl_2$, 25mM DTT and 25nM DNA: RNA Fork 1 substrate. The reactions were incubated for up to 30 minutes at 37°C before being loaded onto a 10% acrylamide gel for electrophoresis.*

To qualitatively determine the unwinding activity of DDX49 over time, the wild-type protein was added to multiple reaction mixes containing Fork 2A DNA and incubated for up to 30 minutes. Reaction mixes were removed from the 37°C water bath and stopped at 2.5-minute intervals. These were then loaded onto a 10% acrylamide TBE gel for analysis (Figure 21). The amount of unwound substrate was shown to have increased over time and with it, the nuclease activity of DDX49. This suggests that DDX49 may first unwind duplex substrates before then cutting them.

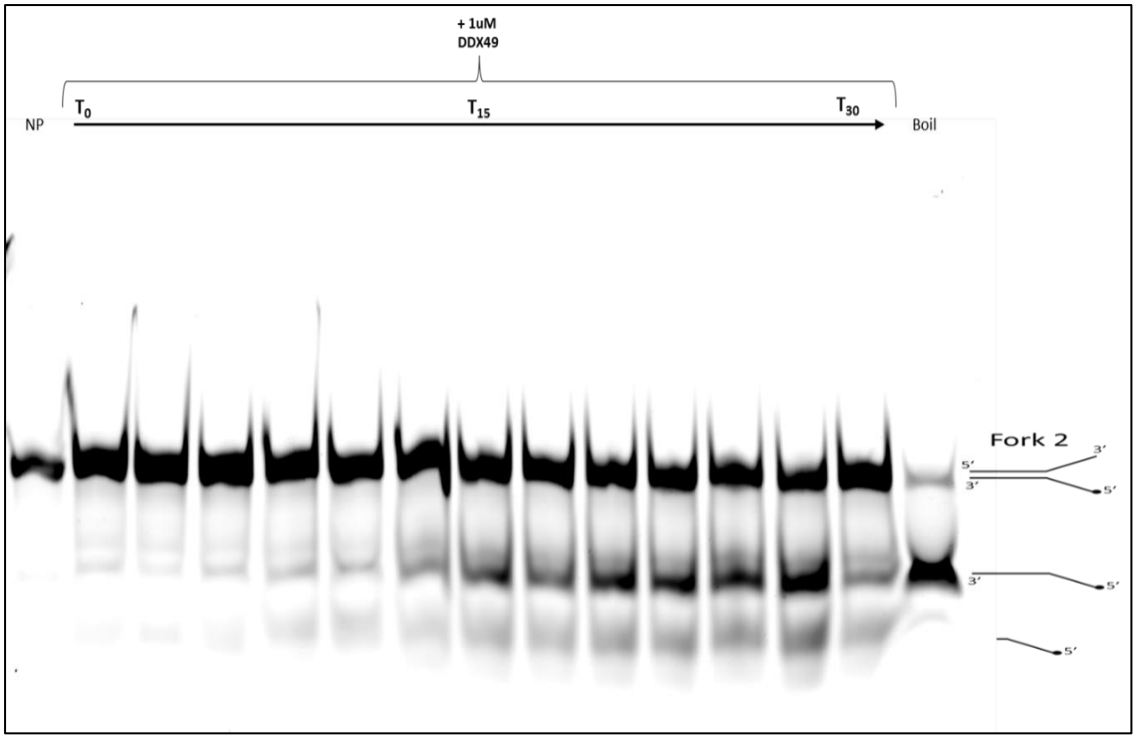

Figure 21: The helicase and nuclease activity of DDX49 increases over time. *DDX49 was added to reaction mixes containing: 1xHB, 5mM ATP, 10mM $MgCl_2$, 25mM DTT and 25nM Fork 2A DNA substrate. Reactions were incubated for up to 30 minutes and stopped every 2.5 minutes with Stop mix.*

Unwinding activity of different dsDNA substrates was also quantitatively measured using FRET assays. Duplex FRET substrates consist of a 5' Cy5-labelled strand, and a 3' Cy3- labelled strand. Excitation of the Cy3 labelled strand at 540nM when it is in close proximity of the Cy5 probe results in the emission of the FRET signal at 680nM. When a duplex FRET substrate is unwound, the Cy5 and Cy3-labelled strands move apart, and the FRET signal is lost. Therefore, one can assume that unwinding activity directly corresponds to the loss of FRET signal over time. DDX49 was loaded into reaction mixes containing either FRET Fork 2, FRET Fork 3 or FRET Fork 4. These reactions were incubated in the wells of a 96-well black Packard microplate at 37°C, and the FRET signals were measured for up to 30 minutes (Figure 22). The data obtained correlates with the gel-based helicase

assays shown above and confirms that DDX49 translocates and unwinds in the 3' to 5' direction.

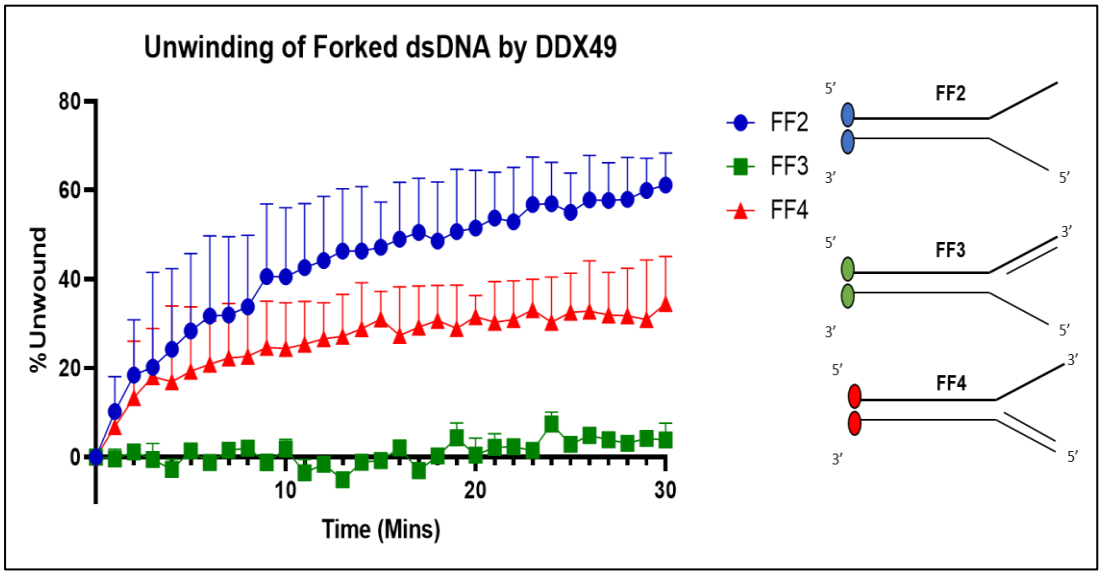

Figure 22: DDX49 can unwind FRET Fork 2 and FRET Fork 4 substrates not FRET Fork 3, showing it is a 3' to 5' translocase*. DDX49 was added to reaction mixes with 50nM 5' Cy5-labelled/3' Cy3-labelled forked DNA, 5mM ATP, 10mM $MgCl_2$ and 5mM DTT. This was incubated within the BMG FluoSTAR Omega microplate reader for 30 minutes at 37°C. Experiments were repeated in triplicate.*

The unwinding activity of DDX49 and its mutants was also compared quantitively using FRET assays. DDX49 or its mutant was added to reactions with FRET Fork 2 and incubated in the wells of a 96-well black Packard microplate at 37°C and the FRET signals were measured for up to 30 minutes (shown in Figure 23). Compared to the wild-type, unwinding activity by the Walker B mutant was only slightly inhibited and greatly inhibited in the accessory domain mutant.

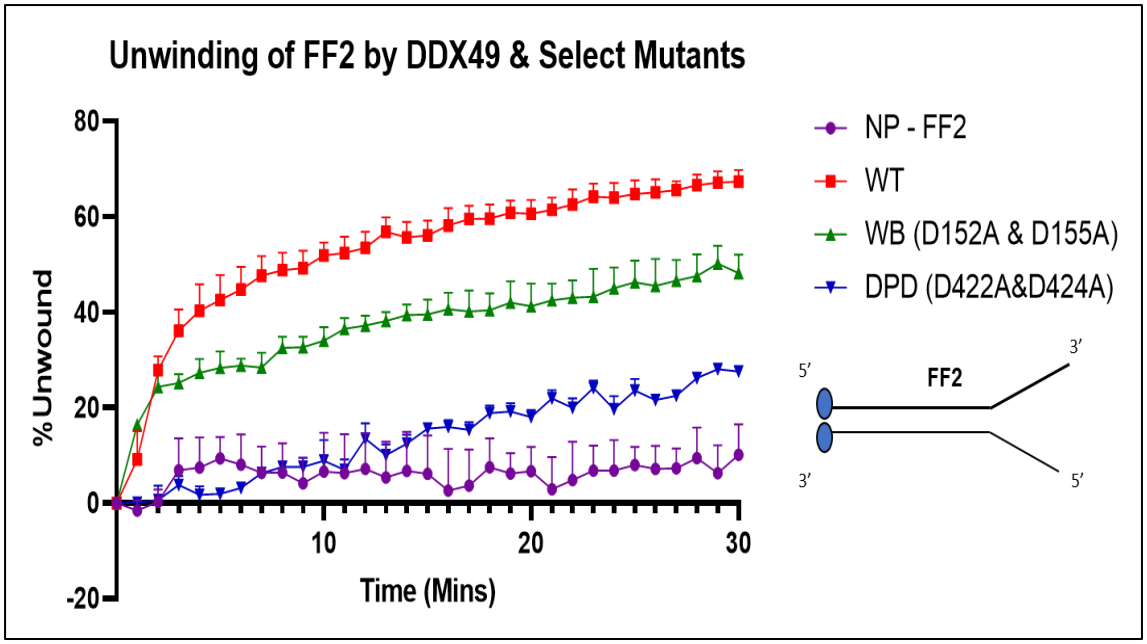

Figure 23: A quantitative comparison of helicase activity of DDX49 and selected mutants*. 1μM DDX49 or its mutant was added to reaction mixes with 50nM 5' Cy5-labelled/3' Cy3-labelled forked DNA, 5mM ATP, 10mM $MgCl_2$ and 5mM DTT. This was incubated within the BMG FluoSTAR Omega microplate reader for 30 minutes at 37°C. Experiments were repeated in triplicate.*

3.5 Summary

In summary, DDX49 is confirmed to be an ATP-dependent helicase with an apparent binding preference for RNA substrates over DNA substrates, and it translocates and unwinds duplex substrates in a 3' to 5' direction. In addition to this, DDX49 displays novel nuclease activity which may be ATP-independent and mutational analysis has shown that this may be conferred by DDX49's DPD motif in the accessory domain. Inactivation of the Walker B motif in the core domain only slightly inhibits the helicase ability of DDX49, and this may be explained by the general mechanism of unwinding by DExD-box helicases, where ATP hydrolysis primarily serves to dissociate the protein from the substrate and recycle it.

Chapter 4: *CRISPR-Cas9 mediated gene editing of DDX49 in U2OS cells*

4.1 Introduction

DDX49 is thought to play a regulatory role in protein synthesis by influencing mRNA export from the nucleus to the cytoplasm, and there is also evidence of its involvement in ribosome biogenesis.[10] DDX49 is also reported by the Cancer Dependency Map project to be a common essential gene[39], therefore its deletion in Human cells would result in cell death. However, the effect of knocking out or downregulating of DDX49 gene expression through CRISPR-Cas9 gene editing has not been reported.

4.2. Aims

The aims of this work were to establish a phenotype of U2OS osteosarcoma cells where DDX49 gene expression was knocked out by CRISPR-Cas9 mediated genetic editing.

4.3 Transfection of Cas9-guideRNA ribonucleoprotein into U2OS cells

40,000 U2OS cells were seeded in the wells of a 24-well plate, they were then transfected with Cas9 that was complexed with guide RNA that targeted exon 3 of the DDX49 gene, along with lipofectamine CrisprMAX reagent. To confirm successful transfection, ATTO-labelled guide RNA was also transfected into a separate well of U2OS cells. These cells were imaged using brightfield and GFP fluorescent microscopy (Figure 24).

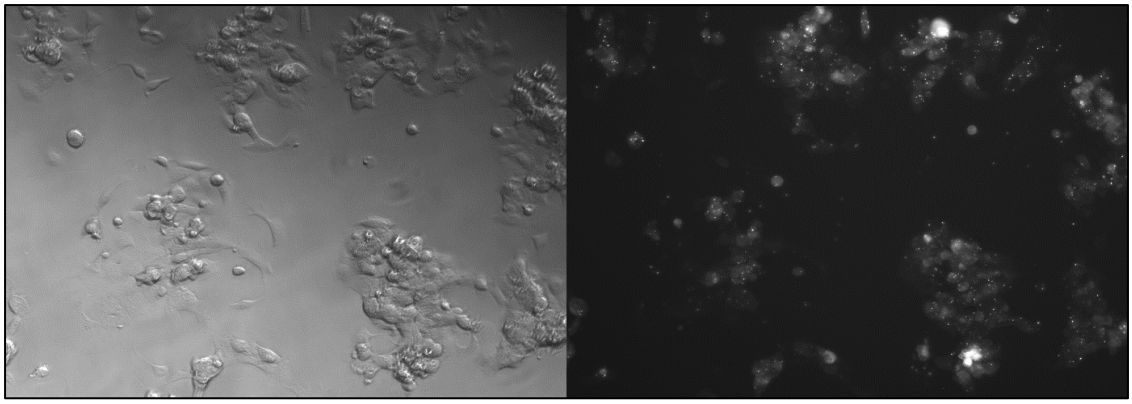

Figure 24: U2OS cells imaged 48 hours after transfection with the Cas9-guideRNA Ribonucleoprotein*. Left) Imaged with brightfield microscopy. Right) Imaged with fluorescent microscopy.*

Once successful transfection was confirmed, singe colonies were picked for clonal expansion and cultured in 75cm^2 flasks. The genomic DNA of the edited cells was extracted and sent for Sanger sequencing. By comparing the edited sequences against the wild-type, the editing efficiency by Cas9 was determined by TIDE analysis. This showed that only one allele of DDX49 was edited, so DDX49 gene expression by these cells was heterozygous (Figure 25).

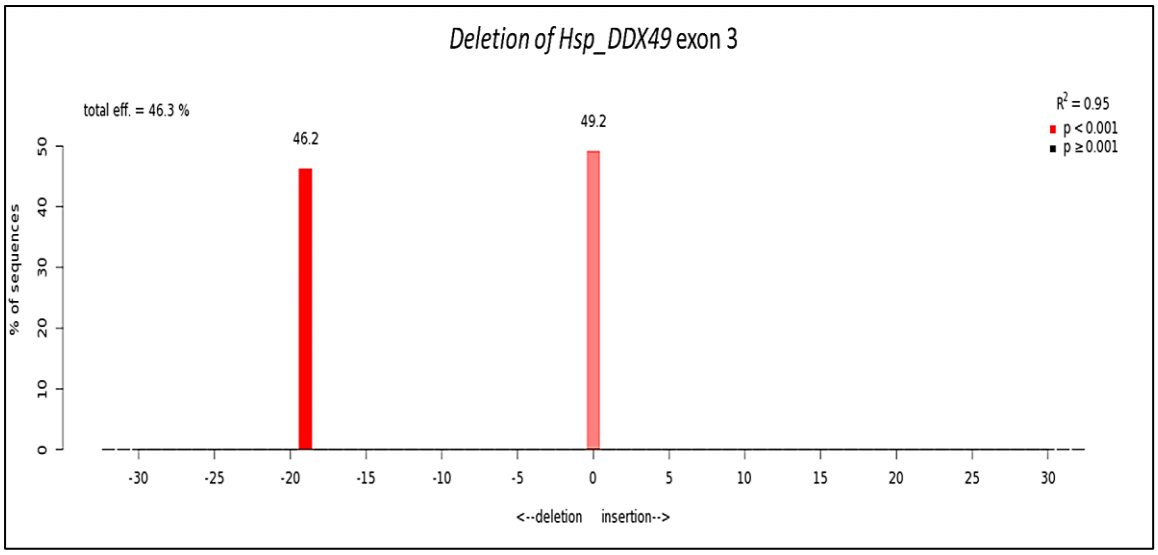

Figure 25: TIDE analysis shows that almost half of genome sequences in the colony of Cas9-transfected U2OS cells were edited successfully*, therefore they can be considered heterozygous. An insignificant percentage of sequences had off-target effects (p>0.002).*

4.4 Phenotyping of edited U2OS cells

DDX49 expression was downregulated in U2OS cells by CRISPR-Cas9 mediated gene editing, the cellular proliferation of these cells compared against the wild-type was measured using WST-1 assays. The heterozygous expression of DDX49 by U2OS cells was shown to inhibit their cellular proliferation (Figure 26).

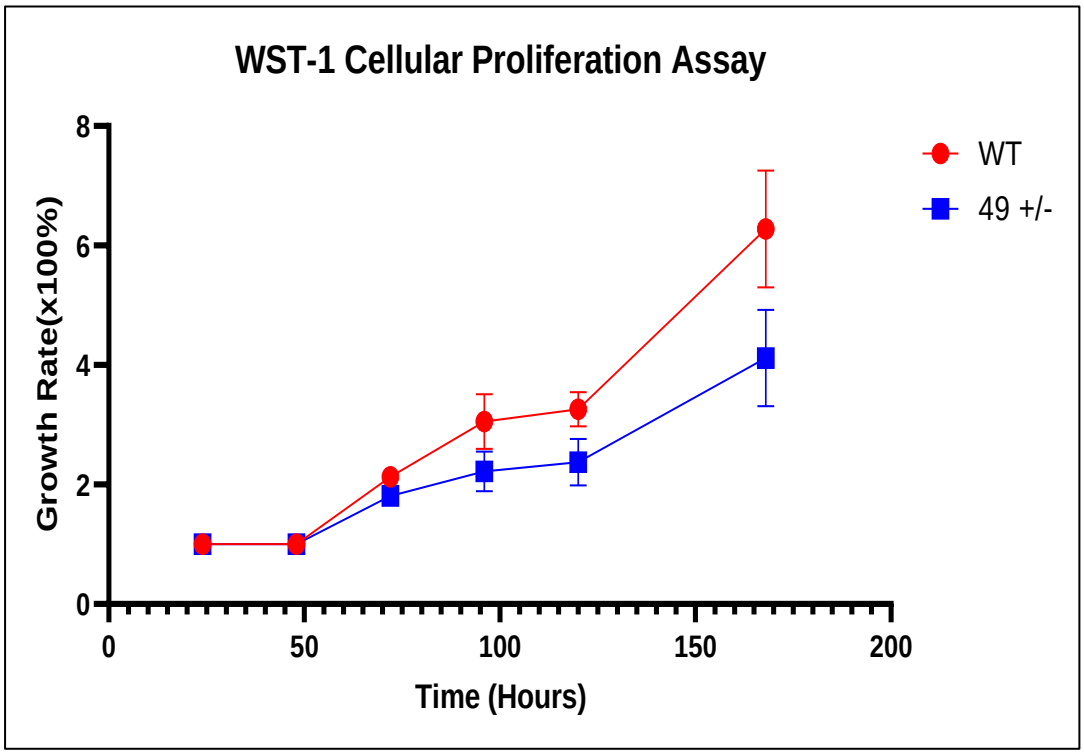

Figure 26: WST-1 assay shows that U2OS that have been heterozygous for DDX49 expression (blue), proliferate at a similar rate to wild-type homozygous U2OS cells (red) (n=3)

The cellular migration of the edited U2OS cells was compared against the wild-type was measured using wound-healing assays (Figure 27). Wild-type and edited cells were grown to full confluency in cell culture dishes, and a scratch was made in the monolayer. Migration across the scratch was monitored in both phenotypes for up to 72 hours. The heterozygous

expression of DDX49 by U2OS cells was shown to slow down their cellular migration.

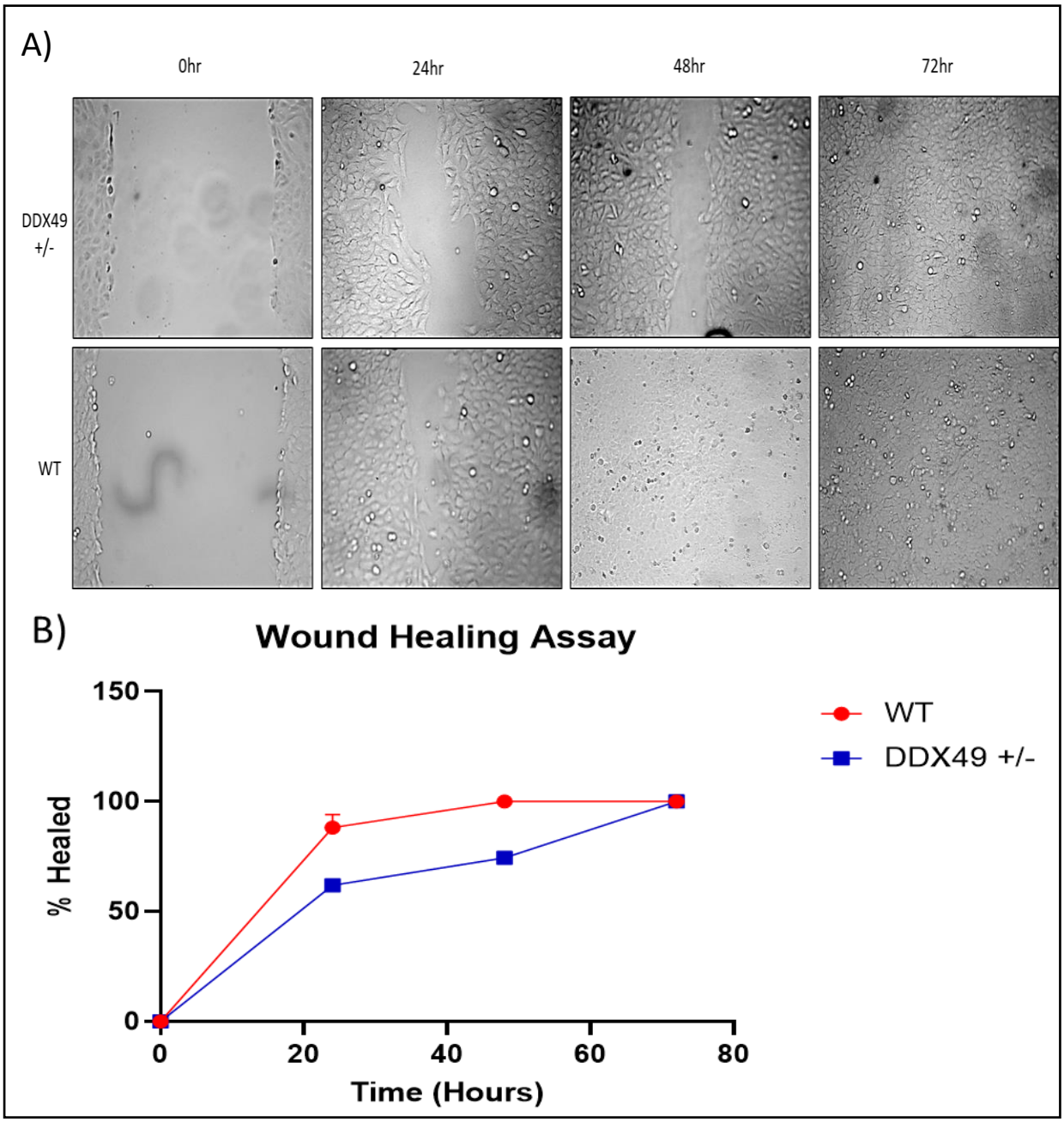

Figure 27: Wound healing assay using wild-type (WT) U2OS cells and U2OS cells that have heterogenous DDX49 expression (DDX49+/-). *A monolayer of cells was scratched and was monitored over 72 hours. A) Progressive cell migration across the scratch imaged using brightfield microscopy. B) The width of the scratch was recorded over 72 hours and plotted as a percentage. DDX49+/- cells healed at a slower rate than the WT cells (n=3)*

4.5 Summary

In summary, we aimed to knock out DDX49 gene expression within U2OS osteosarcoma cells using CRISPR-Cas9 mediated gene editing. However, only one allele of the DDX49 gene was successfully knocked out, so expression was downregulated instead. This created a mild sickness phenotype in the U2OS cells rather than cell death as expected from a full knockout.

Chapter 5: Discussion and Further Work

The aims of this research were to explore the biochemical characteristics of DDX49 through the mutational *in vitro* analysis of selected residues of interest, and also to create a phenotype in U2OS cells where DDX49 gene expression was knocked out through CRISPR-Cas9 mediated gene editing. Through biochemical characterisation, the precise cellular function of DDX49 could be realised. From the work presented, we have revealed novel properties of Human DDX49, and have produced a phenotype in U2OS cells that displays heterozygous DDX49 expression.

5.1 Analysis of DDX49's binding with nucleic acid substrates

Towards its C-terminus, DDX49 was predicted to possess an accessory domain, which resembles structural motifs associated with DNA binding. A conserved Aspartate-Proline-Aspartate (DPD) motif within the accessory domain was mutated into the non-polar Alanine-Proline-Alanine (APA) motif to see if any protein-nucleic acid interactions were affected.

Through the use of EMSAs, it was observed that wild-type DDX49 could bind to both duplex and single-stranded DNA and RNA substrates but had a clear preference for single stranded RNA substrates. Unlike wild-type DDX49, the accessory domain mutant was shown to not bind forked dsDNA substrates except when ATP and Mg^{2+} was added to the electrophoresis buffer and gel. The RecA-like domains in the core region of DExD-box helicases adopt a closed conformation around a substrate when it is cooperatively bound with

ATP. Therefore, the data may suggest that DDX49's accessory domain binds to DNA independently of ATP-binding. To validate this further, a mutant of DDX49 where both the Walker A motif (associated with ATP binding) and the DxD motif in the accessory domain are both inactivated, could be generated, and subjected to EMSAs in the presence of ATP and Mg^{2+}. This may show a complete inhibition of DDX49-substrate interaction.

In addition, the use of anisotropy work may be necessary to quantitively show DDX49's affinity for various types of DNA and RNA substrates, and how the mutation of selected residues in the primary structure of DDX49 can affect this. Fluorescence anisotropy is a technique that measures the rotational motion of fluorescently labelled molecules in solution. When a labelled molecule becomes associated with a protein, this causes an increase in anisotropy which can be recorded. By plotting this change in anisotropy against protein concentration, the equilibrium dissociation constant (K_d) of the protein can be recorded.[40] Mutants of DDX49 may have different K_d values compared to the wild-type and this may give insight into preferred substrates of DDX49.

5.2 Analysis of DDX49 unwinding data

Through the use of helicase assays, DDX49 was observed to show both helicase and nuclease activity. A Walker B-inactivated mutant was created though the mutation of both Aspartate residues in the Walker B motif, this inhibits ATP hydrolysis by the protein. Helicase activity by this mutant was only slightly inhibited, this may be because ATP hydrolysis was only required to release the DExD-box helicase from the substrate, recycling it. The

nuclease activity did not seem to be affected by inactivation of the Walker B motif, suggesting that this function is independent of ATP hydrolysis.

We have also shown that DDX49 can unwind duplex DNA and RNA substrates as well as DNA: RNA hybrids, in the presence of ATP, few DExD-box helicases can unwind all of these types of substrates, for example: DDX5 and DDX43[38]. Further exploration of DDX49's substrate unwinding preference would involve titrating DDX49 into helicase reactions with either duplex DNA or RNA. In addition, the ATP hydrolysis activity of DDX49 with either nucleic acid substrate could also be compared through the use of Malachite Green phosphate assay, which can quantify the amount of free phosphate ions yielded from ATP hydrolysis.

FRET assays have shown that wild-type DDX49 can unwind roughly 50% of Fork 2A DNA, which possesses a 24 base-pair long duplex DNA. In comparison, DDX43 can unwind roughly 40% of 19 base-pair long duplex DNA, so DDX49 seems to be more processive than DDX43.[38] The degree of processivity by DDX49 can be explored by using duplex substrates of varying length and composition.

The use of alternative cations to Mg^{2+} ions may alter the activity and specificity of DDX49. In some enzymes, this has been explored through biochemical characterisation and reported that Mn^{2+} ions can substitute for Mg^{2+} ions in the active site[41,42], as is the case for the Werner syndrome (WRN) helicase, which also displays both nuclease and helicase activity[43]; this could also be explored in DDX49, an experiment could be set up where Manganese ions are titrated into reaction mixes with DDX49, ATP and a

variety of substrates. A similar experiment has been conducted in the investigation of the mitochondrial helicase hSuv3, where it had been found that using Mn^{2+} ions in lieu of Mg^{2+} ions resulted in greater helicase activity.[44] This experiment could also be repeated with a variety of different divalent metal ions.

Many DExD-box helicases possess an accessory domain in their N-terminus or C-terminus, these can have a variety of functions such as target recognition or stabilising the protein-substrate interaction. The accessory domain mutant of DDX49 was also subjected to helicase assays, it was shown to unwind DNA substrates but failed to cut them like the Wild-type and Walker B mutant. It can be concluded that the accessory domain is responsible for nucleic acid substrate binding and nuclease activity. The accessory domain mutant also displayed reduced helicase activity as evidenced by FRET assays. It is possible that the accessory domain of DDX49 also stabilises substrate binding to the core domain. The dual helicase-nuclease nature of DDX49 is similar to the Cas3 protein first identified in *E. coli* K12 strain, which also possesses DExD/H – like domain and bears many structural motifs akin to SF2 helicases. Cas3 is involved mainly in the anti-viral immunity of bacteria, it is theorised that its nuclease function can cleave ssDNA R-loop structures that form when crRNA associate with protospacer DNA.[45] DDX49 has already been implicated in anti-viral activities, where its overexpression in Human BCBL-1 cells was associated with an inhibition of the effects of Kaposi's sarcoma-associated herpesvirus (KSHV). DDX49 was found to bind with KSHV transcripts but how it may interact with them to limit pathogenesis remains to be elucidated.[46]

It is likely that DDX49 does not act alone in its cellular functions, and it is possible that the disordered C-terminus of DDX49 may only fold correctly in the presence of an accessory protein. To identify other proteins that may interact with DDX49, BioID proximity labelling could be used. This involves fusing a biotin ligase to one of the termini of DDX49 such that when it is expressed in cells, it can be induced to biotinylate proteins that are close in proximity and may be interacting with DDX49. These biotinylated proteins are purified from the cell lysate and identified using mass spectrometry.[47] From this, a range of proteins that may only transiently interact with DDX49 could be screened and then evaluated further using other assays such as a two-hybrid screen. Anisotropy could also be used by fluorescently labelling DDX49 or the protein of interest, and measuring the change in anisotropy if they interact with each other.[48] Through these methods, an accessory protein could be identified and lend further insight into the function of DDX49.

5.3. Analysis of CRISPR-Cas9 mediated downregulation of DDX49 expression in U2OS cells

The initial aim was to completely knockout DDX49 gene expression in U2OS cells using CRISPR-Cas9 gene editing methods. However, TIDE analysis had shown that nearly 50% of genome sequences in the created phenotype had been edited successfully, therefore it can be assumed that the phenotype has heterozygous expression of DDX49. This was actually advantageous as DDX49 is an essential gene so its complete knockout would result in cell death, but with a heterozygous phenotype (DDX49+/-), the effects of downregulated DDX49 expression could now be monitored. The edited cells displayed a mild sickness phenotype compared to the wild-type, its capacity

for proliferation was reduced; this demonstrates that DDX49 is vital to proper cell growth. However, this work was only conducted in osteosarcoma cells and similar gene editing could be conducted in a cervical cancer cell line (such as HeLa) to gain insight as to how a high expression of DDX49 in result in a better survival prognosis for cervical cancer patients. HeLa is an HPV-positive cell line, HPV (Human papillomavirus) is a major factor in the oncogenesis of cervical cancer.[49] Since DDX49 has been implicated in anti-viral immunity, whether it interacts with HPV proteins or transcripts could be a potential topic of exploration. As well as this, DDX49 could be overexpressed *in cellulo* to see if this may alter cell-line growth.

The expression of DDX49 in the edited cells could be quantitated and compared against the wild-type, this may further confirm that DDX49 expression is reduced in a heterozygous phenotype. This could be achieved by using qPCR, the total RNA from the edited and wild-type cells would be extracted, reverse transcribed into cDNA and then amplified using primers designed to bind to the *DDX49* gene. A smaller accumulation of the amplified *DDX49* gene in the edited cells compared to the wild-type may show that DDX49 expression has been downregulated in the heterozygous phenotype. Similarly, western blotting with an antibody specific to DDX49 could be carried out on the protein lysate recovered from the edited and wild-type cells, a heterozygous phenotype may show less protein expression.

As mentioned earlier, DDX49 was demonstrated to have nuclease function *in vitro*, however the role of this either *in vivo* (within living organisms) or *in cellulo* (within living cells) is unknown. To investigate this, prime editing on Human primary cells could be used. Prime editing is as equally versatile and

programmable as CRISPR-Cas9 editing, however prime editing instead utilises a catalytically impaired Cas9 nickase that is fused to an engineered reverse transcriptase, and it is programmed with prime-editing guide RNA (pegRNA). The pegRNA contains an RNA template for synthesis of the edited gene sequence. Once the fusion protein and the associated pegRNA are introduced into the cell, the target DNA sequence is nicked by the fusion protein to expose a 3' hydroxyl group that will be used to initiate the reverse transcription of the RNA template directly into the target site. This creates heteroduplex DNA with one edited and one unedited strand. The cell can use DNA mismatch repair mechanisms to copy the edited strand onto the complementary strand, or vice-versa (Figure 28).[50] This technique can be used to provide genetic information or to mutate the residues of expressed proteins. In this context of this work, the *DDX49* gene could be mutated *in cellulo* so that the mutant protein DDX49$^{D422A/424A}$ is expressed instead. If this method yields a phenotype where cellular growth or protein synthesis is impaired, then it could highlight the significance of DDX49's nuclease function.

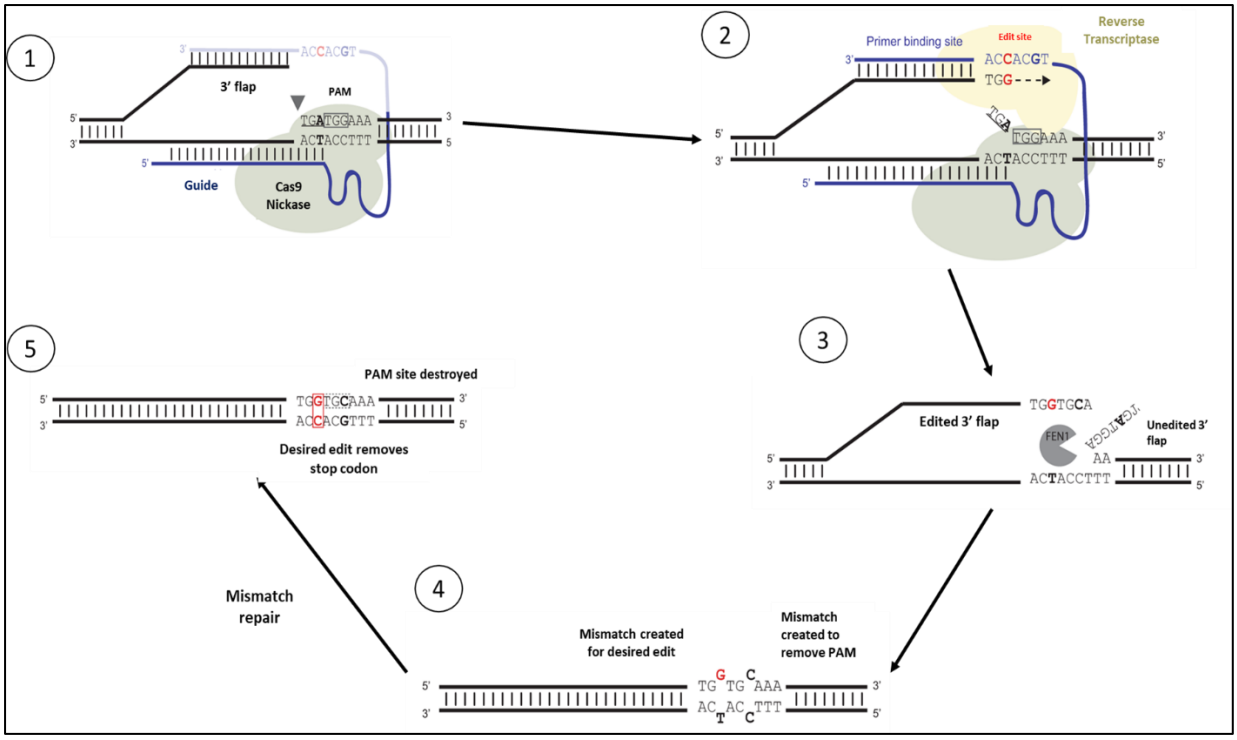

Figure 28: Prime editing involves the use of a Cas9 nickase fused to reverse-transcriptase, this is associated with a prime-editing guide RNA (pegRNA). Step 1) The Cas9-nickase/pegRNA complex binds to the target region and creates a nick upstream of the PAM site, producing a 3' flap. Step 2) The 3' flap interacts with the complementary primer-binding site on the 3' end of the pegRNA. The reverse transcriptase utilises this site to extend the 3' flap. Step 3) The edited 3' flap displaces the unedited 5' flap, which is removed by FEN1 nuclease from the cell. Step 4) A mismatch in base pairs is created from the edit, and this is resolved in Step 5) where mismatch repair results in correctly edited DNA or the original sequence. The Cas9 nickase is shown in green and the reverse transcriptase in yellow. Figure modified from [51]

5.4 Acknowledgments

Firstly, I would like to thank my supervisors: Dr. Edward Bolt, Tom Killelea and Ashley Parkes, whose training and expertise have been invaluable to this research. I wish to thank Fiorela Kapllanaj as well, whose hard work has been pivotal to this project. Additional thanks to my colleagues in the Bolt Lab who have been ever helpful with their knowledge and patience. Finally, I would also like to thank my parents and family, who have supported me throughout all my endeavours.

References

1. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 helicases: Family matters. *Current Opinion in Structural Biology* vol. 20 Preprint at https://doi.org/10.1016/j.sbi.2010.03.011 (2010).

2. Byrd, A. K. & Raney, K. D. Superfamily 2 helicases. *Frontiers in Bioscience* vol. 17 Preprint at https://doi.org/10.2741/4038 (2012).

3. Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry* vol. 76 Preprint at https://doi.org/10.1146/annurev.biochem.76.052305.115300 (2007).

4. Linder, P. & Jankowsky, E. From unwinding to clamping â€" the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology* vol. 12 Preprint at https://doi.org/10.1038/nrm3154 (2011).

5. Jankowsky, E. & Fairman, M. E. RNA helicases - one fold for many functions. *Current Opinion in Structural Biology* vol. 17 Preprint at https://doi.org/10.1016/j.sbi.2007.05.007 (2007).

6. Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. R. & White, M. F. The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains. *Mol Cell* **23**, (2006).

7. Deltoro, D. *et al.* Walker-A Motif Acts to Coordinate ATP Hydrolysis with Motor Output in Viral DNA Packaging. *J Mol Biol* **428**, (2016).

8. Xing, Z., Wang, S. & Tran, E. J. Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with S. cerevisiae ortholog Dbp2 in transcriptional control and glucose metabolism. *RNA* **23**, (2017).

9. Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc Natl Acad Sci U S A* **105**, (2008).

10. Awasthi, S. *et al.* DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Res* **46**, (2018).

11. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science (1979)* **357**, (2017).

12. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, (1990).

13. Lian, X. *et al.* DDX49 is a novel biomarker and therapeutic target for lung cancer metastases. *J Cell Mol Med* **24**, (2020).

14. Li, H. K. *et al.* DDX3 represses stemness by epigenetically modulating tumor-suppressive miRNAs in hepatocellular carcinoma. *Sci Rep* **6**, (2016).

15. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, (2021).

16. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, (2015).

17. Delano, W. L. The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography* **40**, (2002).

18. Dosztányi, Z. Prediction of protein disorder based on IUPred. *Protein Science* **27**, (2018).

19. Pellegrini-Calace, M. & Thornton, J. M. Detecting DNA-binding helix-turn-helix structural motifs using sequence and structure information. *Nucleic Acids Res* **33**, (2005).

20. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* vol. 29 Preprint at https://doi.org/10.1016/j.femsre.2004.12.008 (2005).

21. Littlefield, O. & Nelson, H. C. M. A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocrystal. *Nat Struct Biol* **6**, (1999).

22. Cirillo, L. A. & Zaret, K. S. Specific Interactions of the Wing Domains of FOXA1 Transcription Factor with DNA. *J Mol Biol* **366**, (2007).

23. Roychowdhury, A. *et al.* The DEAH-box RNA helicase Dhr1 contains a remarkable carboxyl terminal domain essential for small ribosomal subunit biogenesis. *Nucleic Acids Res* **47**, (2019).

24. Sorokin, E. P. *et al.* Inhibition of Mg2+ binding and DNA religation by bacterial topoisomerase I via introduction of an additional positive charge into the active site region. *Nucleic Acids Res* **36**, (2008).

25. Confalonieri, F. *et al.* Reverse gyrase: A helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc Natl Acad Sci U S A* **90**, (1993).

26. Aravind, L., Leipe, D. D. & Koonin, E. v. Toprim - A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res* **26**, (1998).

27. Kosinski, J., Feder, M. & Bujnicki, J. M. The PD-(D/E)XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics* **6**, (2005).

28. Cheng, B. *et al.* Asp-to-Asn Substitution at the First Position of the DxD TOPRIM Motif of Recombinant Bacterial Topoisomerase I Is Extremely Lethal to E. coli. *J Mol Biol* **385**, (2009).

29. Maris, C., Dominguez, C. & Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS Journal* vol. 272 Preprint at https://doi.org/10.1111/j.1742-4658.2005.04653.x (2005).

30. Martin, R., Straub, A. U., Doebele, C. & Bohnsack, M. T. DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biology* vol. 10 Preprint at https://doi.org/10.4161/rna.21879 (2013).

31. Granneman, S. *et al.* The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Res* **34**, (2006).

32. Kressler, D., de la Cruz, J., Rojo, M. & Linder, P. Dbp6p Is an Essential Putative ATP-Dependent RNA Helicase Required for 60S-Ribosomal-Subunit Assembly in Saccharomyces cerevisiae . *Mol Cell Biol* **18**, (1998).

33. Cai, W. *et al.* Wanted DEAD/H or alive: Helicases winding up in cancers. *Journal of the National Cancer Institute* vol. 109 Preprint at https://doi.org/10.1093/jnci/djw278 (2017).

34. Hashemi, V. *et al.* The role of DEAD-box RNA helicase p68 (DDX5) in the development and treatment of breast cancer. *Journal of Cellular Physiology* vol. 234 Preprint at https://doi.org/10.1002/jcp.26912 (2019).

35. Wang, Z. *et al.* DDX5 promotes proliferation and tumorigenesis of non-small-cell lung cancer cells by activating β-catenin signaling pathway. *Cancer Sci* **106**, (2015).

36. He, Y. *et al.* A double-edged function of DDX3, as an oncogene or tumor suppressor, in cancer progression (Review). *Oncology Reports* vol. 39 Preprint at https://doi.org/10.3892/or.2018.6203 (2018).

37. Botlagunta, M. *et al.* Expression of DDX3 is directly modulated by hypoxia inducible factor-1 alpha in breast epithelial cells. *PLoS One* **6**, (2011).

38. Talwar, T. *et al.* The DEAD-box protein DDX43 (HAGE) is a dual RNA-DNA helicase and has a K-homology domain required for full nucleic acid unwinding activity. *Journal of Biological Chemistry* **292**, (2017).

39. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).

40. Weinberg, R. L., Veprintsev, D. B. & Fersht, A. R. Cooperative Binding of Tetrameric p53 to DNA. *J Mol Biol* **341**, 1145–1159 (2004).

41. Hsu, M.-T. & Berg, P. Altering the specificity of restriction endonuclease: effect of replacing Mg^{2+} with Mn^{2+}. *Biochemistry* **17**, 131–138 (1978).

42. Sissi, C. & Palumbo, M. Effects of magnesium and related divalent metal ions in topoisomerase structure and function. *Nucleic Acids Res* **37**, 702–11 (2009).

43. Choudhary, S., Sommers, J. A. & Brosh, R. M. Biochemical and Kinetic Characterization of the DNA Helicase and Exonuclease Activities of Werner Syndrome Protein. *Journal of Biological Chemistry* **279**, 34603–34613 (2004).

44. Venø, S. T., Witt, M. B., Kulikowicz, T., Bohr, V. A. & Stevnsner, T. Regulation of the human Suv3 helicase on DNA by inorganic cofactors. *Biochimie* **108**, 160–168 (2015).

45. He, L., James, M. S. J., Radovcic, M., Ivancic-Bace, I. & Bolt, E. L. Cas3 protein—a review of a multi-tasking machine. *Genes* vol. 11 Preprint at https://doi.org/10.3390/genes11020208 (2020).

46. Serfecz, J. C. *et al.* DExD/H Box Helicases DDX24 and DDX49 Inhibit Reactivation of Kaposi's Sarcoma Associated Herpesvirus by Interacting with Viral mRNAs. *Viruses* **14**, 2083 (2022).

47. Sears, R. M., May, D. G. & Roux, K. J. BioID as a Tool for Protein-Proximity Labeling in Living Cells. *Methods Mol Biol* **2012**, 299–313 (2019).

48. Gijsbers, A., Nishigaki, T. & Sánchez-Puig, N. Fluorescence Anisotropy as a Tool to Study Protein-protein Interactions. *Journal of Visualized Experiments* (2016) doi:10.3791/54640.

49. Okunade, K. S. Human papillomavirus and cervical cancer. *J Obstet Gynaecol (Lahore)* **40**, 602–608 (2020).

50. Anzalone, A. v. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, (2019).

51. Scholefield, J. & Harrison, P. T. Prime editing – an update on the field. *Gene Ther* **28**, 396–401 (2021).

Chapter 6: Appendices

7.1 Appendix 1

7.1.1 Homo sapiens DDX49 protein sequence

Hsp_DDX49:

MAGFAELGLSSWLVEQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSG
KTAAFVLPILQKLSEDPYGIFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVG
GMDMVAQALELSRKPHVVIATPGRLADHLRSSNTFSIKKIRFLVMDEADRLL
EQGCTDFTVDLEAILAAVPARRQTLLFSATLTDTLRELQGLATNQPFFWEAQ
APVSTVEQLDQRYLLVPEKVKDAYLVHLIQRFQDEHEDWSIIIFTNTCKTCQIL
CMMLRKFSFPTVALHSMMKQKERFAALAKFKSSIYRILIATDVASRGLDIPTV
QVVINHNTPGLPKIYIHRVGRTARAGRQGQAITLVTQYDIHLVHAIEEQIKKKL
EEFSVEEAEVLQILTQVNVVRRECEIKLEAAHFDEKKEINKRKQLILEGKDPD
LEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRAGHKGRPPRTPSGSHSG
PVPSQGLV

Supplementary Figure 1: Hsp_DDX49 amino acid sequence as used