



University of
Nottingham

UK | CHINA | MALAYSIA

Molecular characterisation of DDX49

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

Fiorela Kapllanaj (B.Sc. Hons)

November 2023

ABSTRACT	4
CHAPTER 1: INTRODUCTION	5
1.1 DEAD BOX HELICASES	5
1.2 DEAD BOX HELICASE 49	12
1.3 DDX49 IN DISEASE	14
DDX49 STRUCTURAL MOTIFS	17
1.5 PROJECT AIMS	18
CHAPTER 2: METHODS	19
2.1 MATERIALS.....	19
2.1.1 Strains.....	19
2.1.2 Plasmids.....	20
2.1.3 Primers.....	21
2.1.4 Agar plate preparation.....	22
2.1.5 Substrates	23
2.1.6 General Buffers	24
2.1.7 Purification buffers.....	24
2.1.8 Acrylamide based gels.....	25
2.2 GENERAL PROTOCOL	26
2.2.1 Site directed mutagenesis.....	26
Exponential amplification via PCR.....	26
Analytical digest of PCR product	27
2.2.2 Agarose gel electrophoresis	28
2.2.3 Treatment and Enrichment of PCR product.....	28
2.2.4 Competent cells generation.....	29
2.2.5 Heat shock bacterial transformation.....	29
2.2.6 Overnight growth	30
2.2.7 Miniprep of overnight growth mixture.....	30
2.2.8 Protein overexpression	31
2.2.9 SDS-PAGE analysis	31
2.2.10 Protein purification.....	32
2.2.11 Substrate preparation	34
2.2.12 Helicase Assay.....	34

2.2.13 Electrophoretic mobility shift assay (EMSA).....	35
2.2.14 Fluorescence resonance energy transfer (FRET).....	35
2.2.15 CRISPR-Cas9 mediated DDX49 prime editing in U2OS cells	36
2.2.16 Maxiprep.....	37
2.2.17 Multiple sequence alignment	39
2.2.18 3D structure of DDX49.....	39
CHAPTER 3. ANALYSIS OF THE D-X-D MOTIF FUNCTION IN DDX49.....	40
3.1 INTRODUCTION.....	40
3.2 SEQUENCE ALIGNMENT OF DDX49 HOMOLOGUES	40
3.3 MUTAGENESIS OF DDX49	41
3.4 OVEREXPRESSION AND PURIFICATION OF DDX49 WILD TYPE AND SELECT MUTANTS	42
3.5 BIOCHEMICAL ANALYSIS OF DDX49, DDX49 ^{D422A/D424A} AND DDX49 ^{K421A} UNWINDING	45
3.6 BIOCHEMICAL ANALYSIS OF DDX49, DDX49 ^{D422A/D424A} AND DDX49 ^{K421A} DNA BINDING AFFINITY.....	56
3.7 CRISPR-CAS9 MEDIATED PRIME EDITING OF A DDX49 CONSERVED MOTIF IN U2OS CELLS	58
3.8 SUMMARY OF CHAPTER 3	62
CHAPTER 4: ANALYSIS OF IDR REMOVAL ON DDX49 CATALYTIC ACTIVITIES.....	63
4.1 INTRODUCTION.....	63
4.2 ANALYSIS OF DDX49 ^{AK407-V483} AND DDX49 ^{AE432-V483} HELICASE ACTIVITY.....	66
4.3 ANALYSIS OF DDX49 ^{AK407-V483} AND DDX49 ^{AE432-V483} DNA BINDING.....	67
4.4 SUMMARY OF CHAPTER 4	69
CHAPTER 5: DISCUSSION.....	70
5.1 DDX49 WILD-TYPE EXHIBITS DNA HELICASE AND NUCLEASE ACTIVITY	70
5.2 DDX49 D-X-D MOTIF IS ESSENTIAL FOR NUCLEASE ACTIVITY	72
5.3 DDX49 INTRINSICALLY DISORDERED C-TERMINUS IS ESSENTIAL FOR PROTEIN FUNCTION	78
<i>Acknowledgments</i>	80
CHAPTER 6. APPENDIX	81
6.1 <i>References</i>	86

ABSTRACT

DEAD box helicases play important biological roles in RNA metabolism, transcription and translation, thereby affecting gene expression and protein synthesis. Among these proteins, DEAD-box helicase 49 (DDX49) is ATP-dependent RNA helicase activity involved in regulation of 47S rRNA transcription, nuclear export and ribosome biogenesis via unexplored mechanisms. DDX49 has been shown to facilitate progression of a range of cancers, however due to the lack of functional and mechanistic insights, studies are needed to unravel how dysregulation DDX49 contributes to cancer development. In this project we aim to characterise DDX49 enzymatic activities via mutagenesis of key domains of DDX49, followed by *in-vitro* helicase and DNA binding assays with wild-type and mutant proteins. We provide evidence that DDX49 displays DNA binding, unwinding and nuclease activities and how these functions are related to key structural motifs of DDX49 *in-vitro* and in U2OS osteosarcoma cells.

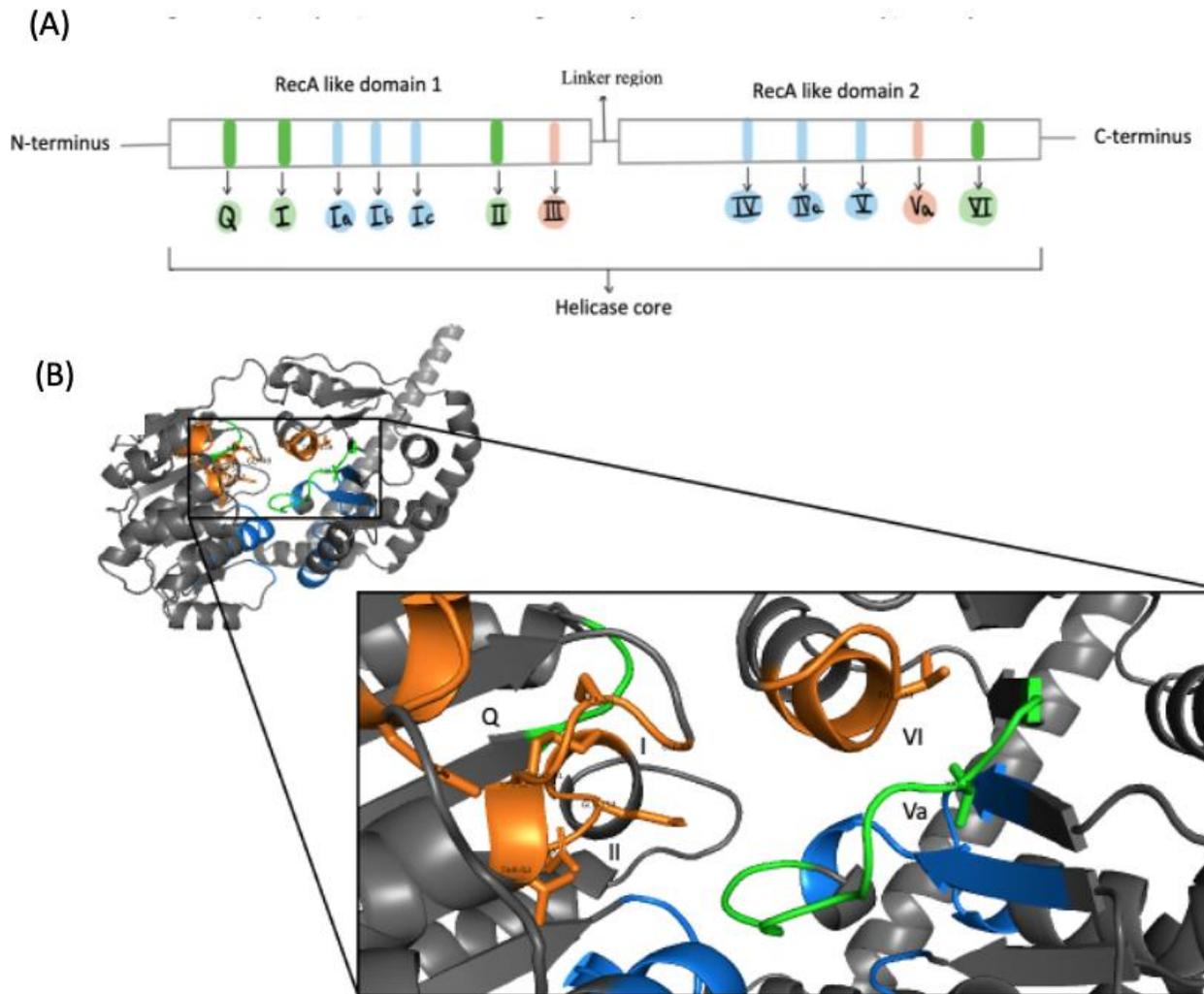
Chapter 1: Introduction

1.1 DEAD box helicases

DEAD box proteins are the largest family of RNA helicases, composed of 37 members and are involved in the central and essential physiological aspects of RNA metabolism.

They are characterised by a structurally highly conserved helicase core, composed of 2 RecA like domains connected via a flexible linker and flanked by the N and C terminus domains of the protein, as shown in figure 1.1 (Donsbach and Klostermeier, 2021).

The interplay of the 9 conserved motifs of the helicase core gives these family of proteins the ability to hydrolyse ATP, bind and unwind RNA duplexes. The characteristic Motif II, also known as the Asp-Glu-Ala-Asp (D-E-A-D) motif, together with motif Q, I and VI carry out ATP binding and hydrolysis, as shown in figure 1.1 (Linder and Jankowsky, 2011).



(A) The diagram shows the general structure of DEAD-box protein helicase core divided into two RecA-like domains connected via a linker sequence. Conserved domains labelled in green (Q, I, II and VI) hydrolyse ATP; domains labelled in blue (Ia/b/c, IV, IVa and V) bind to RNA and pink motifs (II and Va) motifs connect RNA binding and ATP hydrolysis. (Linder and Jankowsky, 2011) (B) 3D structure of the conserved DDX helicase core as modelled by PyMol (Schrödinger and DeLano, 2022).

DEAD-box helicases bind ATP at the interface of the two RecA domains and bind RNA on the surface of the RecA domains via the 2'-OH groups of the sugar backbone. This confers DEAD-box helicases non-specific RNA binding as shown in figure 1.1(B). (Donsbach and Klostermeier, 2021). However, some members of this family contain accessory domains which confer RNA substrate selectivity, such as DDX43 (Yadav et al, 2021) Unwinding of a duplex can be achieved via two mechanisms, either through translocation or local strand separation without translocation along the substrate.

Unlike other RNA/DNA helicases, DEAD-box helicases show no unwinding polarity, as they do not translocate on the RNA strand (Gilman, Tijerina and Russell, 2017).

Instead, DEAD-box proteins directly load onto the RNA and pry strands of typically 10-12 nucleotides in length apart. Their unwinding efficiency decreases with increasing substrate length, which is most likely adapted to their physiological RNA substrates, which are typically less than one helical turn in length. This mechanism of action is termed local strand separation (Linder and Jankowsky, 2011).

Binding of ATP and RNA molecule to the RecA domains, promotes a conformation change from the open (ON) state to the closed (OFF) state as shown in figure 1.2, which aligns the two RecA domains and ATP binding pocket, assembling the active site for ATP hydrolysis and formation of a kink in the RNA backbone. This causes the destabilisation of a few base pairs of the duplex, leading to the dissociation of the first RNA strand and local unwinding. The release of the phosphate group of the ATP molecule upon

hydrolysis, is coupled to the dissociation of the second RNA strand as shown in figure 1.2 (Sengoku et al., 2006).

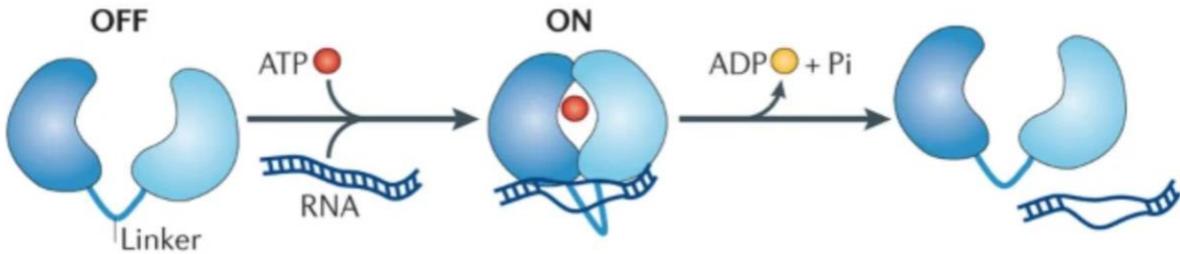


Figure 1.2. DEAD-box helicases are non-processive ATPases that perform local unwinding. A short linker region between the two RecA domains confers these proteins high flexibility. They can switch from a ligand-free OFF conformation to an ATP/RNA-bound ON conformation, resulting in the two RecA domains coming together. This conformational change unwinds RNA base pairs. Upon hydrolysis of ATP, the RNA substrate is released, and the enzyme is recycled (Bourgeois, et al., 2016).

Unwinding by the DEAD box helicases occurs when dynamic fluctuations of an RNA duplex arise, or when a duplex RNA contains extended 5' or 3' ends, which stimulates unwinding. RNA strand separation is dependent on ATP binding to the catalytic helicase core, whereas its hydrolysis drives the dissociation of the protein from the RNA substrate, as shown in figure 1.2 (Bohnsack et al., 2023).

Some DEAD- box helicases can bind ATP without hydrolysing it, subsequently promoting RNA binding for half-lives of hours, which forms the basis for how some enzymes function, such as the DEAD-box helicase eukaryotic initiation factor 4AIII (eIF4AIII). This clamping on RNA, allows eIF4AIII to act as an assembly platform and form a multiprotein complex as part of the exon-junction complex during mRNA translation at the ribosome. Thus, some of these proteins are more accurately described as ATP-dependent RNA binding, than RNA helicases (Bohnsack et al., 2023).

Despite the highly conserved structure of the helicase core, DEAD-box proteins' functions are further characterised by intrinsically disordered regions (IDR), as N and C-terminus domains flanking the conserved core. These range from a few amino acids to several hundreds and are usually intrinsically disordered regions, thus do not have a defined 3D structure. (Linder and Jankowsky, 2011). These intrinsically disordered regions confer additional physiological roles such as nucleotide binding, protein-protein interactions, self-assembly and liquid-liquid phase separation of these helicases into stress granules (Weis and Hondele, 2022).

One such example is DDX19, which contains a C-terminus disordered region, which has been shown to form an α - helix which is able to contact the N terminus containing the nucleotide binding sites of the core helicase domains of the protein. This α - helix displaces the amino acid Arginine 429 which is essential for the ATPase reaction to occur. The position of the helix in the core domains of DDX19, prevents formation of the ligand bound ON-conformation as shown in figure 1.2, thus autoregulating its ATP-dependent RNA

binding activity. Upon ssRNA binding, the N-terminus is displaced from the α - helix induced close state conformation, allowing the ATP to bind (Collins et al,2009).

Several proteins harbouring intrinsically disordered regions have shown to undergo phase separation under different conditions, one of them being DDX3X. Cancer associated mutations of DDX3X are linked to its helicase core and to its ability to form liquid-liquid phase separation (LLPS). Its N-terminus IDR is responsible for LLPS *in-vitro*, while its acetylation leads to inhibition of these liquid droplets forming, thereby suggesting an essential role of these IDR's in physiological functions and disease (Saito et al., 2018).

DEAD-box helicase involvement in essential cellular processes means that it is not surprising that when expression or function are dysregulated, they are implicated in the development and progression of cancer. Whether they play causative roles in oncogenic or are dysregulated as a consequence of cancer has yet to be elucidated.

Some well characterised DEAD-box proteins which overexpressed in cancer include, eIF4A which drives translation in cancer cells and DDX3X, which promotes tumorigenesis at the transcriptional level, as described in table 1.1. On the other hand, DDX5, appears to be differentially post translationally modified, facilitating cell proliferation and anti-apoptotic signals (Cai et al, 2017).

Table 1.1 Functions attributed to DEAD-box proteins DDX3, eIF4A and DDX5 and their importance in cancer.

DEAD-box helicase	Physiological role	Importance in cancer
DDX3	<p>Involved in translation initiation.</p> <p>Involved in transcription initiation of cancer related genes, such as activation of, Interferon-β and repression of E-cadherin</p>	<p>Promote translation initiation by cyclin E1, therefore facilitating G1/S transition and promoting cell growth.</p> <p>Overexpression of DDX3 represses E-Cadherin expression which increases motility of cell and facilitates EMT (epithelial to mesenchymal transition). (Fuller-Pace, 2013)</p>
eIF4A	<p>eIF4A1 initiates cap-dependent translation.</p> <p>eIF4A3 is a component of the exon-junction complex, thereby coordinating pre-mRNA splicing to non-sense mediated decay, mRNA localisation and translation.</p>	<p>eIF4A overexpression promotes proliferation, invasion, migration and epithelial to mesenchymal transition (Xue et al., 2021)</p>
DDX5	<p>Prevents excessive genomic instability by playing a role in DNA repair. Transcriptional co-activator of β-catenin/c-Myc pathway</p>	<p>Frequently dysregulated and mutated in cancers.</p> <p>Activates and increases expression of oncogene such as Wnt target genes. facilitates EMT and promotes</p>

		therapeutic drug resistance (Cai et al., 2017)
--	--	--

1.2 DEAD Box helicase 49

Many members of the DEAD-box family currently remain uncharacterised with only few of them being extensively studied; DEAD Box helicase 49 (DDX49) is one such uncharacterised protein. DDX49 is described as a probable ATP-dependent RNA helicase, as it is part of the SF2 family of helicases, which show preference for RNA binding and has been confirmed to unwind RNA substrates (Awasthi et al., 2018). The same study confirming its RNA helicase activity also showed it is co-localised with the nucleolar protein fibrillarin, thus confirming its nucleolar localisation. Furthermore, immunofluorescent staining of human cell line U2OS cells shows DDX49 mitochondrial localisation (Uhlen et al., 2017).

Upon DDX49 knockdown in HeLa cells via siRNA, there was a significant decrease in fluorescently labelled poly(A)⁺RNA in the cytoplasm compared to a siRNA control, where DDX49 was not depleted. RT-PCR also confirmed that mRNA export levels decreased in a splicing independent manner, due to single exon genes having similar reduction of spliced genes (Awasthi et al., 2018). This suggested that DDX49 could be involved in mRNA export from the nucleus. Its nucleolar localisation was then linked to effects on pre-ribosomal RNA levels in the nucleolus; DDX49 siRNA knockdown decreased the levels of 47S RNA by 60%, whereas its overexpression did not have any significant

effects. This suggests that DDX49 could be involved in the steady state regulation of 47S rRNA levels (Awasthi et al., 2018). Upon treatment of DDX49-knockdown cells with the transcriptional inhibitor Actinomycin D, mRNA levels measured using RT-PCR showed a substantially reduced stability of 47S rRNA. Furthermore, DDX49 depleted cells also showed decreased fluorescent staining of newly nascent RNA in the nucleoli of cells. DDX49 was also found to be localised at rDNA locus regulator regions and promoter (H0) and at the 3' end of the transcribed gene (H13). This suggested that DDX49 might be involved in transcription of pre-ribosomal RNA by binding to the DNA locus of 47S RNA regulatory regions.

Puromycin treated HEK293 cells depleted in DDX49, showed a decreased rate of protein synthesis compared to control scrambled siRNA cells. Furthermore, knockdown of DDX49 in HeLa cells significantly decreased cellular proliferation and colony forming ability, whereas DDX49 overexpression induced increased cellular proliferation coupled with only a subtle increase in mRNA export levels and pre-rRNA 47S RNA levels. In conclusion, these findings establish that DDX49 seems to be essential in protein synthesis and cellular growth via its involvement in ribosome biogenesis, mRNA export, regulation of ribosomal pre-mRNA transcription and stability via unknown mechanisms as shown in figure 1.2.1, (Awasthi et al., 2018).

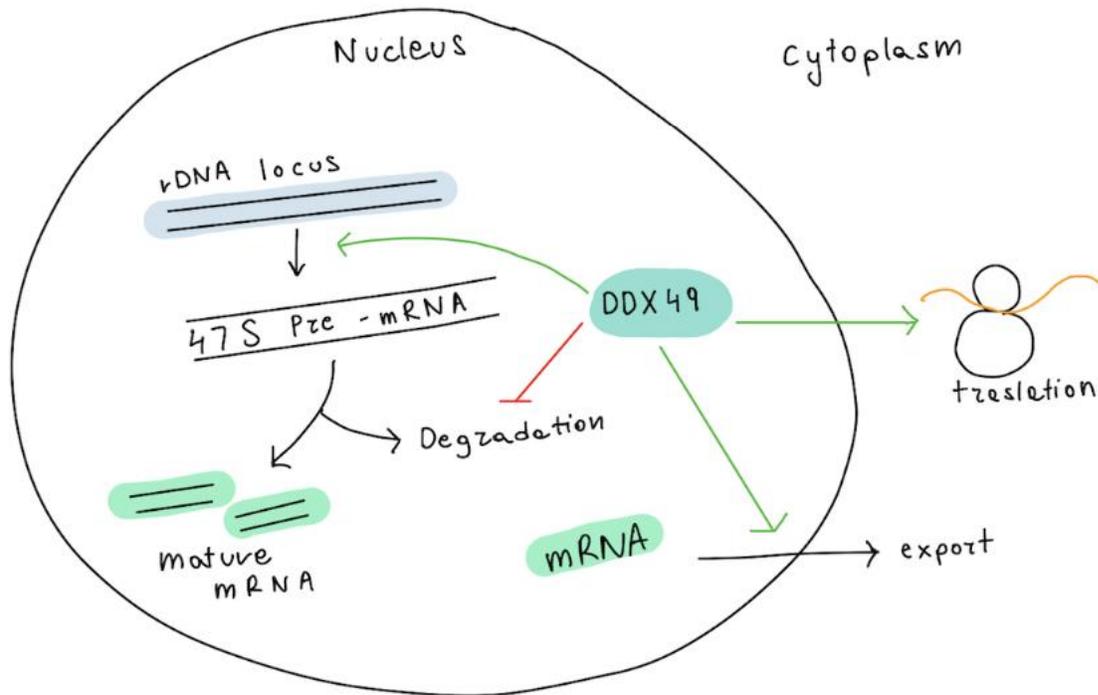


Figure 1.2.1. Diagram illustrating functional roles of DDX49. The diagram shows DDX49 localised in the nucleolus contributes to stability and transcription and maturation of 47S pre- RNA and inhibition of its degradation. DDX49 facilitates mRNA export into the cytosol and mRNA translation. (Awasthi et al., 2018)

1.3 DDX49 in disease

DDX49 appears to be mutated in cancers including colorectal, stomach, melanoma, uterine and prostate (Cai et al., 2017). A study by Tao et al., (2023), showed that DDX49 was significantly elevated in prostate cancer tissues in association with disease recurrence. Its downregulation was associated with suppressed proliferation, cell cycle arrest and facilitated cell death.

Analysis of RNA extracted from patient breast cancer stem cells revealed that DDX49 was aberrantly expressed, compared to control embryonic stem cells. Upon its knockdown, cancer involved stemness associated transcription factors OCT3/4 and SOX2 significantly decreased, with parallel increase in NANOG and CD34, with unaffected phenotype of cells (Apostolou et al., 2015).

DDX49 was found to be overexpressed in lung adenocarcinoma tissues and cells and its knockdown in lung adenocarcinoma cells inhibited cell proliferation, invasion and promoted apoptosis. (Wang et al., 2020) Furthermore, DDX49 showed to be differentially expressed in positive versus negative lymph node metastases in lung cancer patients. In addition, downregulation of DDX49 in non-small lung cancer cells (NSCLC) resulted in suppressed NSCLC growth and proliferation (Lian et al., 2020).

DDX49 overexpression has been correlated with poor survival prognosis in renal cancer and liver cancer patients (Uhlen et al., 2017). Hepatocellular carcinoma cells (HPCs) were shown to overexpress DDX49 compared to low expression seen in healthy livers.

Knockdown of DDX49 in HPCs significantly decreased cellular proliferation and metastasis *in-vitro* with subsequent recovery of invasion and proliferation following DDX49 overexpression (Dai et al., 2021).

On the other hand, DDX49 overexpression has been correlated with favourable prognosis in cervical cancer patients (Uhlen et al., 2017), however contrasting evidence shows its suppression in cervical cancer cells (CCLs) is associated with anti-tumour activity via

inactivation of PI3K/AKT and Wnt/ β -catenin pathways, which have extensively been shown to promote cancer cell survival. This was determined by reduced cell viability proliferation and invasion and increased apoptosis of CCLs upon DDX49 silencing, whilst overexpression facilitated cell cycle progression and suppressed apoptosis (Chen et al., 2023).

Computational analysis by Awasthi et al, shows DDX49 is an interactor of APP (Amyloid-beta A4 protein). Furthermore, DDX49 is upregulated in early-stage Alzheimer's disease mouse model with mutations in the APP gene. (Aladeokin et al., 2019) This suggests DDX49 could play a role in the early stages of Alzheimer's disease development.

DDX49 seems to be involved viral activities. It acts as a pattern recognition receptor (PRR) by negatively regulating Kaposi Sarcoma Associated Herpesvirus via direct interaction with viral mRNA IE and E (early gene transcripts) in the nucleus, thereby inhibiting latent viral reactivation (Serfecz et al., 2022).

Furthermore, DDX49 was shown to be involved in HIV viral replication by interacting with the HIV protein Gag (Jager et al., 2012).

DDX49 structural motifs

Upon sequence alignment of DDX49 homologues in a range of organisms (unpublished results Bolt et al.) conservation of the amino-acid sequence D-P-D (Aspartate-Proline-Aspartate) was highlighted in the C-terminus of the proteins. Inferred sequences from homology are not confirmed orthologs, thereby reducing the robustness of the investigation, due to paralogues having different functionalities.

The D-x-D motif has been shown to play a functional role in TOPRIM nucleases (Yang et al., 2010), by facilitating divalent metal coordination for nucleotide sugar binding. The conserved region of DDX49 could play a role in enzyme catalysis of DDX49 involved in functions that have been established or yet to be characterised.

S. fasciatus	LKEYPV--QEKEVLKILTQVNVTRRECEIKLESTDFDEKKEINKRKQLILEGKDPELEAK	428
C. elegans	LDELKV--SPKHVTKYVTQVLVAKKEAELKLENQKFGEKKEINRRKELLSGMDDEADR	517
D. melanogaster	LTEHPI--DQRMVERIFMQVNVTRRESEMQLDNNDFDERAQNYRRKTWIMEGKDPDQMEA	432
A. gambiae	LTEYSV--DDRLVQRIFMQVKVARAEAEINLDNKDFDERKHKYRRLRWIQEGLDPDEMEA	432
C. lupus	LEELSV--EEAKVLQILTQVNVVRRRECEIKLEAANFDEKKEINKRKQLILEGKDPDLEAK	424
M. musculus	LAELVV--EEAEVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEGKDPDLEAK	428
R. norvegicus	LAELAV--EEAQVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEGKDPDLEAK	428
H. sapiens	LEEFVS--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
P. troglodytes	LEEFVS--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
M. mulatta	LEEFVS--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
B. taurus	LEEFPV--EEAQVLQILTQVNVVRRRECEIKLEAANFDEKKEINKRKQMILEGKDPDLEAK	336
D. rerio	LKEFPI--EEKEVLKILTQVNVTRRQCEIKLESTDFDEKKEINKRKQMI LDGKDPDLEEK	428
M. unicolor	LKDFAV--EESVVLKILTQVNVTRRECEIKLESTDFDEKREKNRKQLILDGKDPDLEAK	428
S. punctatus	LQEFVS--EERDVLGILTQVNVVRRRECEIKLEATDFDEKKEINKRKQMILEGKDPDLEEK	428
P. muralis	LQEFVS--EEHSVNLILTQVNVVRRRECEIRLEATDFDEKKEINKRKQMILEGKDPDLEAQ	477

Figure 3.2.1. Multiple sequence alignment of DDX49 homologues using Clustal Omega⁰. The highlighted sequence D-P-D shows to be conserved across mammals and non-mammals. DDX49 orthologs were used for sequence alignment; labelled sequences (*) were inferred from homology. (McWilliam et al., 2013)

1.5 Project aims

The project aims will be to investigate DDX49 enzymatic activities *in-vitro* by conducting biochemical assays with DNA and RNA substrates, highlighting any differences in protein activities between substrates. Subsequently, site directed mutagenesis will be conducted to mutate key conserved amino acid sequences (D-P-D) and the C-terminus intrinsically disordered region of the protein, to investigate their function via substrate binding and unwinding assays. Finally, we will mutate the conserved D-P-D sequence using prime editing in U2OS cells to assess its impact in-cellulo.

Chapter 2: Methods

All chemicals are supplied by New England Biolabs™ (Ipswich, MA), Fisher Scientific (Waltham, Massachusetts, United States) and Sigma Aldrich (St. Louis, Missouri, United States), unless otherwise stated.

2.1 Materials

2.1.1 Strains

For bacterial transformation of encoding plasmids, *E. coli* DH5- α cells (Genotype: *F- ϕ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17(rK-, mK+) phoA supE44 λ -thi-1 gyrA96 relA1*) strain was used. (Thermo-Fisher™). In DH5- α cells, the *recA1* mutation disrupts homologous recombination and the *endA1* mutation disrupts endonuclease activity to increase DNA yield formation, by stabilising DNA during transformation (Durfee et al., 2008).

For protein overexpression, BL21AI (Genotype: *F-ompT hsdSB (rB- mB-) gal dcm araB: T7RNAP- tetA*) *E. coli* strain was used (Thermo-Fisher™). The BL21AI strain encodes for an arabinose inducible T7 RNA polymerase and is deficient in cytoplasmic protease Lon and outer membrane protease ompT. *Lon* and *ompT* mutations prevent recombinant protein degradation during expression (Rosano and Ceccarelli, 2014).

2.1.2 Plasmids

p100.DDX49 (DDX49 BL pET 100D, Thermofisher TM) is a codon optimised circular plasmid of 7213 base pairs. It encodes an ampicillin resistant gene (AmpR) for selection of bacterial colonies containing the plasmid, and a T7 promoter upstream to allow for high-level expression of the gene of interest (DDX49) upon IPTG induction. It also encodes multiple cloning sites and His-tagged DDX49.

BPK1520 (Addgene plasmid #65777; <https://www.addgene.org/65777/>) encodes for the *Streptococcus pyogenes* Cas9 gRNA backbone for human cell expression. It contains a BsmBI restriction site to insert the sgRNA spacer sequence (17-20 nt) of interest into the vector. It also contains an ampicillin resistance gene for successful colony selection and a U6 promoter for high expression of small RNAs in eukaryotic cells.

pU6-pegRNA-GG-acceptor (Addgene plasmid #132777; <https://www.addgene.org/132777/>) is used for prime editing in mammalian cells, which recognises a specific PAM sequence of the target DNA in mammalian cells. This contains an ampicillin resistance gene for bacterial colony selection and a U6 promoter for high-level expression of transcripts. It also contains an mCherry label which results in growth of red colonies. The mCherry sequence is disrupted upon successful golden gate assembly of the pEG-RNA in the plasmid, thereby indicating successful cloning upon loss of red fluorescence in bacterial colonies.

2.1.3 Primers

Codon optimised primers and oligonucleotides shown in table 2.1, 2.2 and 2.3, were lyophilized at 25 nM scale and resuspended in UltraPure DNase/RNase-Free Distilled Water following manufacture instructions (Thermo-Fisher TM).

Table 2.1 Codon optimised primer sequences used for Sanger sequencing (T7 terminator primer) and p100.DDX49 site directed mutagenesis of DDX49 to generate *DDX49*^{D422A/D424A}, *DDX49*^{K421A}, *DDX49*^{Δ408-483} and *DDX49*^{Δ432-483} mutant proteins.

Name of Primer	Sequence 5'-3'
T7 terminator primer	GCTAGTTATTGCTCAGCGG
<i>DDX49</i> ^{D422A/D424A} Forward primer	GGCACTGGAAGCAAACGTAAAG
<i>DDX49</i> ^{D422A/D424A} Reverse primer	GGTGCTTTACCTTCCAGGATCAG
<i>DDX49</i> ^{K421A} Forward primer	ATCCTGGAAGGT GCCGATCCGGACCTGGAA
<i>DDX49</i> ^{K421A} Reverse primer	CAGCTGCTTGCGTTTGTGATTTCT
<i>DDX49</i> ^{Δ408-483} / <i>DDX49</i> ^{Δ432-483} Forward primer	TAACTAGCATAACCCCTTG
<i>DDX49</i> ^{Δ408-483} Reverse primer	TTTCTCATCGAAGTGTGC
<i>DDX49</i> ^{Δ432-483} Reverse primer	GGCTTTACGTTTTGCTTC

Table 2.2. Codon optimised primer sequences used for golden gate cloning of sgRNA plasmid (p65777) and pEG-RNA acceptor plasmid (p132777).

Name of Primer	Sequence
49_APA_sgF	CACCGTGCGCTTGGCCTCCAGGTCAGTTTTAGA
49_APA_sgR	TAGCTCTAAACTGACCTGGAGGCCAAGCGCAC
49_APA_scaffF	GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGAGTCG
49_APA_scaffR	GCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTA GCCTTATTTTAACTTGCTATTTTC
49_APA_extensF	GTGCTGCCAGGCCCTGCCCTGGAGGCCA
49_APA_extensR	AAAATGGCCTCCAGGGCAGGGGCCTGGCA
49_APA_PE3_sgF	CACCGTGGCCAGGTTCCCTGGCGG
49_APA_PE3_sgR	AAACCCGCCAGGGAACCTGGCCAC

2.1.4 Agar plate preparation

Agar (3 g) was added to 200 ml of Lysogeny Broth (10 g/L Tryptone, 10 g/L NaCl and 5 g/L of Yeast Extract (pH 7.0) topped to 1 L with distilled water) followed by autoclaving. This is then microwaved for 10 minutes. Once cooled, the solution is supplied with ampicillin (1 µg/ml). 25 ml of the mixture is poured in each agar plate under sterile conditions and allowed to dry in an incubator at 42 °C.

2.1.5 Substrates

DNA substrates were generated using labelled and unlabelled oligonucleotide MW12, MW14, PM16 and PM17 and 1X annealing buffer. These oligos generated the following substrates:

- Fork2 (5 μ M 5' cyanine-MW14, 6 μ M MW12)
- Fork3 (5 μ M 5' cy5-MW14, 6 μ M MW12, PM16)
- Fork4, (5 μ M 5' cy5-MW14, 6 μ M MW12, PM17)
- FretFork2 (5 μ M 5' cyanine5-MW14, 5 μ M 3' Cyanine3-MW12)

Table 2.3 Oligos used to make DNA substrate Fork2, Fork3, Fork 4 and FretFork2

Oligo-nucleotide substrate	Label	Nucleotide sequence
MW12	-	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAAT TCGGC
MW14	-	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGA TCCGA
PM17	-	TAGCAATGTAATCGTCTATGACGTTG
PM16	-	TGCCGAATTCTACCAGTGCCAGTGAT
3' Cy3-MW12	Cy3	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAAT TCGGC
5' Cy5-MW14	Cy5	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGA TCCGA

2.1.6 General Buffers

Table 2.4 General buffer compositions.

Stock solution	Composition
10X Tris Borate EDTA (TBE) buffer	10.8 g of Tris, 5.5 g of Boric acid and 40 mM EDTA (Ethylenediaminetetraacetic acid))
10X SDS running buffer	250 mM Tris, 1.92 M Glycine and 1% (v/v) SDS
10X annealing buffer	10 mM Tris pH 7.5, 50 nM NaCl
10X helicase buffer	Tris pH 7.5 (10 mM), Bovine Serum Albumin (25 mg/ml) and v/v Glycerol (11.25%)
Stop buffer	2.5% v/v SDS buffer, 200 nM EDTA (Ethylenediaminetetraacetic acid) and 2 mg/ml Proteinase K

2.1.7 Purification buffers

Table 2.5 Buffers used in protein purification.

Stock solution	Composition
Nickel buffer A	25 mM Tris (pH 7.5), 500 Mm NaCl, 25 mM Imidazole and 10% v/v glycerol
Nickel buffer B	25 mM Tris (pH 7.5), 500 mM NaCl, 400 mM Imidazole and 10% v/v glycerol
Heparin buffer A	25 mM Tris (pH 7.5), 100 mM NaCl, 10% v/v glycerol

Heparin buffer B	25 mM Tris (pH 7.5), 1M NaCl, 10% glycerol
DDX49 storage buffer	25 mM Tris (pH 7.5), 200 mM NaCl, 35% glycerol and 5 mM DTT (Dithiothreitol)

2.1.8 Acrylamide based gels

Table 2.6. Acrylamide based gels used for electrophoresis.

Stock solution	Composition
10% native gel	8% w/v acrylamide, 0.5X TBE (10.8 g of Tris, 5.5 g of Boric acid and 40 mM EDTA), 0.135% v/v TEMED (Tetramethyl ethylenediamine), 0.054% v/v APS (ammonium persulfate) and 18.7 ml of dH ₂ O
5% native gel	5% w/v acrylamide v/v 1X TBE, 0.01X v/v APS, 0.05% v/v TEMED and distilled water topped up to a volume of 44.567 ml
SDS PAGE Separating gel	0.1X SDS, 10% w/v acrylamide, 0.4 M Tris (pH 8.8), 0.08% v/v APS, 9.5 µl TEMED and distilled water (6.18 ml)
SDS PAGE Stacking gel	0.01% SDS, 5% w/v acrylamide, 0.122 M Tris (pH 6.8), 0.01% v/v APS, TEMED (3 µl) and 1.75 ml of distilled water

2.2 General protocol

2.2.1 Site directed mutagenesis

Exponential amplification via PCR

PCR reaction mixture (50 μ l) contained 0.5X Q5 buffer, 200 μ M deoxynucleotide triphosphates (dNTPs), Q5 Hot Start High-Fidelity Polymerase (1 U), forward primer (500 nM), reverse primer (500 nM), 10 ng of DNA plasmid and nuclease-free water. Yield (ng/ μ l) of plasmid PCR product were measured using DeNovix DS-11 spectrophotometer (at 260 nm) blanked with nuclease-free water.

Optimal primer annealing temperature (67°C) for primers was calculated using NEB Tm Calculator TM (<https://tmcalsculator.neb.com/#!/main>). A gradient PCR was also conducted with annealing temperatures spanning 60°C, 62°C, 64°C, 66 °C, 68°C and 70°C, and a control sample at 70°C, as shown in table 2.2.1.

Table 2.2.1. PCR programme utilised for site directed mutagenesis.

<i>Step</i>	<i>Temperature (°C)</i>	<i>Time (seconds)</i>
<i>Initial denaturation</i>	98°C	30
<i>Denaturation</i>	98°C	10
<i>Annealing of primers</i>	60-70°C	30
<i>Elongation</i>	72°C	120
<i>Final extension</i>	72°C	120

Analytical digest of PCR product

Following exponential amplification of PCR, the sample is digested with restriction enzymes to measure the molecular weight of the digested product, which would confirm the plasmid has been correctly amplified. Plasmid DNA (100 ng) was added to rCutSmart Buffer (2 mM Tris-acetate, 1 mM magnesium acetate, 5 mM potassium acetate, 10 µg/ml Recombinant Albumin pH 7.9) to ensure optimal enzyme activity, HindIII (20 U) and 1 µl of NdeI (20 U) and nuclease-free water to make up 20 µl reaction. Samples were then incubated at 37°C for 60 minutes.

2.2.2 Agarose gel electrophoresis

The digested product is then subjected to gel electrophoresis to separate the excised plasmid fragments and measure their molecular weight. A 1% w/v agarose gel was prepared by adding 1 g of agarose to 100 ml of 1X TBE and microwaving for 150 seconds. Once the solution had cooled down, ethidium bromide was added to a final concentration of 50 mM. 15 µl of the of the digested PCR product was added to the agarose gel, following addition of purple loading dye (2X final concentration) (NEB#B7025) and run for 60 minutes at 140 V using a Bio-Rad power supply (Bio-rad.com). Gels were imaged under UV exposure using a U:Genius3 Bio-imaging system (Syngene).

2.2.3 Treatment and Enrichment of PCR product

Parental DNA is removed, and PCR product is circularised resulting in increased transformation efficiency. This was achieved by adding 100 ng of PCR product sample to T4 DNA ligase (20 U), T4 PNK (polynucleotide kinase, 10 U), Dnpl (1 U), T4 DNA ligase buffer (final concentration 5 mM Tris-HCl, 1 mM MgCl₂, 0.1 mM ATP, 1 mM DTT (pH 7.5)), topped up to 10 µl with dH₂O. The mixture was incubated at room temperature for 60 minutes, followed by transformation of the whole ligation mixture into competent *DH5-α*, to allow for subsequent DNA purification and sequencing.

2.2.4 Competent cells generation

Escherichia coli (*E. coli*) DH5 α /BL-21AI was inoculated with Mu broth in a ratio of 1:100 respectively. Cells were incubated in 37 °C water bath until optical density (OD) of 0.5 was reached, measured using a Spectronic 20+ (Thermo, Scientific). Cells were pelleted via centrifugation at 6800 g for 5 minutes, supernatant discarded. Pellets resuspended in chilled 0.1 M CaCl₂ and incubated on ice for 1 hour. The centrifugation and resuspension steps were repeated, followed by addition of 30% v/v glycerol. Aliquots of 100 μ l were snap frozen on dry ice and stored at -80°C.

2.2.5 Heat shock bacterial transformation

Plasmids were transformed in either competent *DH5- α* strain or *E. coli* for DNA purification and sequencing or into the BL-21AI strain of *E. coli* for protein overexpression. Both strains followed the same heat-shock transformation protocol.

Under sterile conditions, 10 ng of the plasmid of interest was added to 100 μ l of competent cells (*DH5- α* /BL-21AI), followed by incubation on ice for 5 minutes. The sample was then heat-shocked at 42 °C for 90 seconds, followed by ice incubation for 5 minutes. Lysogeny Broth was then added in a 10:1 ratio to the competent cells. This was followed by a 1 hour incubation in a 37 °C water bath, to allow for optimal bacterial cell growth. The mixture was then centrifuges at 15'000 g for 1 minute, the supernatant (900 μ l) was discarded, and the pellet resuspended in the remaining mixture. The resuspended mixture was spread on the agar plate containing ampicillin (1 μ g/ml), which was incubated overnight at 37 °C to allow for growth of colonies.

2.2.6 Overnight growth

Colonies grown from the incubated agar plates were picked individually and added to 5 ml of LB medium with ampicillin (1 µg/ml) in culture tubes under sterile conditions.

These are then incubated overnight at 37 °C, allowing for single bacterial colonies to expand.

2.2.7 Miniprep of overnight growth mixture

The culture tubes incubated overnight for bacterial growth were subjected to a miniprep experiment, using the QIAprep® Spin Miniprep Kit (supplied from QIAGEN) and following the provided Quick Start Protocol provided by the kit. For this procedure, all centrifugation steps were carried out at 15'000 g unless otherwise stated.

Quick Start Protocol procedure:

Bacterial cultures incubated overnight were centrifuged at 6800 g for 5 minutes and supernatant discarded. The pellet was resuspended with 250 µl of buffer P1, followed by addition of 250 µl of buffer P2. Buffer N3 (350 µl) containing chaotropic salts, was then added and the mixture thoroughly inverted 4-6 times, to allow DNA binding to the silica membrane of the column. This was followed by centrifugation for 10 minutes. The supernatant (800 µl) was then added into the QIAprep2.0 spin column by pipetting and centrifuged for 1 minute.

The column was then washed with buffer PB (500 µl) containing chaotropic salts and centrifuged for 1 minute. Subsequent column wash with 750 µl of wash buffer PE was followed by centrifugation for 1 minute. The DNA was finally eluted via addition of elution buffer EB (50 µl) and centrifugation for 1 minute.

2.2.8 Protein overexpression

Mutant plasmids generated via site directed mutagenesis were transformed into *E. coli* BL-21AI cells using the protocol described in (section 2.2.5), followed by overnight bacterial growth. LB with ampicillin (0.1 mg/l) was added to the bacterial culture in a ratio of 1:100 and inoculated in a 37 °C shaking incubator at 200 g for about 90 minutes. Once the optical density measurement (OD) reached 0.6, the cells were induced by addition of 0.5 mM IPTG (isopropyl β -D-1- thiogalactopyranoside) and 0.1% (v/v) Arabinose, to initiate protein overexpression. These flasks were incubated at 18 °C overnight to minimise bacterial cell replication. This was followed by centrifugation at 5000 g for 10 minutes at 4°C. The pellets were resuspended in buffer A (25 mM Tris (pH 7.5), 500 mM NaCl, 25 mM Imidazole and 10% (v/v) glycerol) and the protease inhibitor PMSF (phenylmethylsulfonyl fluoride) a ratio of 40:1, resulting in a 40 ml final mixture stored at - 80°C for subsequent protein purification.

2.2.9 SDS-PAGE analysis

SDS-page gel electrophoresis was used to measure the molecular weight of the components of the bacterial cellular contents and confirm the overexpression of the protein of interest. Protein containing samples (1 ml) was centrifuged and resuspended with 150 μ l of dH₂O (SDW), 1X SDS-loading dye and 0.1 mM DTT. The mixture was incubated at 95°C for 10 minutes and 15 μ l of the mixture was loaded onto a 10% SDS-page gel composed of separating gel and stacking gel.

Samples were electrophoresed for 60 minutes at 140 V, followed by staining with Coomassie Brilliant Blue (Thermo-Fisher TM) for 15 minutes and de-stained repeatedly in 10% (v/v) ethanol, 7.5% (v/v) acetic acid. Bands on gel were compared against molecular weight marker CPS (coloured standard protein, NEB #P7719S) to measure molecular weight.

2.2.10 Protein purification

Protein purification was carried out using a Nickel affinity column and a Heparin affinity column, using the AKTATM chromatography system. Sample was loaded and run through a Nickel affinity column (HisTrap HP, 5 ml), following column preparation steps described as followed: HisTrap HP was washed with 5 column volumes (CV) (2 ml/min) of wash Nickel Buffer A, followed by 5 CV (2 ml /min) of elution Nickel Buffer B and 5CV of Nickel Buffer A (2 ml/min).

Prior to loading the protein containing mixture into the column, the bacterial cell biomass made in section 2.2.9 and stored at - 80 °C, was thawed on ice, followed by sonication of 10 ml individual aliquots, in short bursts of 10 seconds for 1 minute. The mixture combined into 2 aliquots (20µl) was centrifuged at 35'000 g, using an Avanti J-26 XP centrifuge (Beckman Coulter, rotor: JA 25.5) for 35 minutes at 4°C. The protein containing supernatant was decanted and loaded into the HisTrap HP and flow-through collected. This was followed by column wash-through with Nickel Buffer A (5 CV, 2 ml/minute) and collection of the eluate containing weakly bound proteins, for subsequent analysis. The protein was then eluted via column wash-through with Nickel Buffer B (5 CV) into 25 fractions (1ml /fraction). The protein containing fractions were

analysed via SDS-PAGE analysis and pooled into a dialysis tube (Dialysis Tubing D104, BioDesign). This was subjected to overnight dialysis in Heparin Buffer A at 4°C, using a magnetic stirrer.

The dialysed protein containing solution was decanted from the dialysis tubing and centrifuged at centrifugation at 5000g for 5 minutes at 4°C, to remove any precipitate. The supernatant was then loaded onto a HiTrap Heparin HP 1 ml column, following column preparation steps described below. The column was washed with 5 CV (1ml/min) of Heparin Buffer A, followed by 5 CV (1ml/min) of Heparin Buffer B and 5 CV of Heparin Buffer A. Once the sample was loaded, flow-through was collected and the column was run with 5 CV (1 ml/min) Heparin Buffer A and wash-through collected. The protein was then eluted by running of Heparin Buffer B (1 ml/min), into 15 fractions (1ml/fraction). Once the protein containing fractions were determined via SDS-PAGE analysis, these were pooled into a dialysis tube and were dialysed overnight at 4 °C in DDX49 storage buffer, using a magnetic stirrer. The purified protein containing solution was then aliquoted, snap frozen using dry ice and stored at - 80°C. Protein concentration was measured using the Nano Drop 2000 spectrophotometer (Thermofisher Scientific™) and calculated using Beer Lambert's law and an extinction coefficient of $27350\text{M}^{-1}\text{cm}^{-1}$.

2.2.11 Substrate preparation

All substrates listed in section 2.1.5, were prepared by heating at 95 °C for 10 minutes and allowed to slowly cool overnight. 1X Orange G dye was added to the substrates prior to loading on a 10% native TBE gel and run at 120 V for 3 hours using a Bio-Rad power supply (Bio-rad.com). Gel bands corresponding to the correct side of the desired substrates were cut out using a sterile scalpel and submerged in 200µl of 10X elution buffer (20mM Tris (pH 8), 50mM NaCl), for at least 2 days at 4°C. Substrate was quantified using the Nano Drop 2000 spectrophotometer (Thermofisher Scientific™). Absorption at 260nm and extinction coefficients were applied to the Beer Lambert law to calculate substrate concentration (µM).

2.2.12 Helicase Assay

20µl reaction mixtures containing 1X helicase buffer 5mM ATP, 5 mM Mg²⁺, 25 mM DTT, 25 nM of substrate (Fork 2/3/4), 25 nM unlabelled MW14 (TRAP) and dH₂O were incubated with 2 µl of protein of interest (varying concentration) at 37°C for 30 minutes and subsequently treated with 2µl of stop buffer, to stop the reaction. 15µl samples were loaded in the 10% native gel, following addition of 5µl of Orange-G (80% (v/v) glycerol, orange G powder). Gel electrophoresis was performed for 60 minutes at 140 V in 1X TBE buffer using a Bio-Rad power supply (Bio-rad.com) and imaged using Amersham Typhoon™ image scanner.

2.2.13 Electrophoretic mobility shift assay (EMSA)

20 µl master mixes composed of 1X Helicase buffer, 5 mM DTT (Dithiothreitol), 25 nM of substrate, and DDX49 (0 nM, 200nM, 500 nM and 1µM) were incubated at 37 °C for 30 minutes. Following addition of 5µl Orange G, 20µl samples were loaded on 5% native gel, The gel was run in 1X TBE buffer for 90 minutes at 140 V using a Bio-Rad power supply (Bio-rad.com) and imaged using Amersham Typhoon™. The EMSA was then repeated as described above using 1X TB (10.8 g of Tris, 5.5 g of Boric acid, 0.1 mM ATP and 1 mM MgCl₂) Instead of 1X TBE buffer.

2.2.14 Fluorescence resonance energy transfer (FRET)

DNA unwinding activity was quantified using fluorescence resonance imaging transfer (FRET). Experiments were carried out in duplicate 50 µl reactions (1x HB, 5 mM DTT, 5 mM ATP, 5 mM Mg²⁺, 50 nM labelled substrate and 250 nM unlabelled complementary single stranded oligo (MW14/TRAP). 45 µl of master mix was aliquoted into a 96 well plate followed by spiking of 5 µl of DDX49 in varying concentrations (1 µM, 500 nM, 250 nM). Absorbance changes were measured using a pre-warmed (37 °C) BMG FLUOstar Omega microplate reader measuring sample emission at 590 nm at 1-minute intervals for 30 minutes. Gains were adjusted to the highest signal (Cy3 control) prior to addition of protein. Data points were analysed using Prism (GraphPad) software.

2.2.15 CRISPR-Cas9 mediated DDX49 prime editing in U2OS cells

GOLDEN GATE ASSEMBLY

SgRNA plasmid generation

To generate a phosphorylated oligo duplex, a 20 µl reaction mix containing 1X T4 ligase buffer, 1 µM of oligo inserts (sgF and sgR) and 0.5 µl of T4 PNK (T4 polynucleotide kinase) was incubated at 37 °C for 1 hour followed by heating at 95°C and cooling to 10°C at a rate of -5°C/minute, using a PCR machine.

The vector backbone p65777 (500 ng) was digested and dephosphorylated in 40 µl reaction containing 1X NEB3.1r buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂, 100 µg/ml BSA, pH 7.9), 2.5 Units of Alkaline Phosphatase Calf Intestinal (CIP) and 10 U of BsmBI-v2, at 55°C for 1 hour. The mixture was run through agarose gel electrophoresis of the sample and gel extraction of the plasmid. The vector backbone (10-40 ng) was then ligated with 500 nM of duplex DNA in a 10 µl reaction containing 1X T4 ligase buffer and 200 U of T4 ligase, at room temperature for 1 hour. This was followed by transformation into DH5α cells.

PegRNA plasmid generation

The vector backbone p132777 (1 µg) was digested in a 30 µl reaction containing 1X Cutsmart buffer and 1 U of Bsa-HF2, at 37 °C for 1 hour. The mixture was run through agarose gel electrophoresis and the correct band removed and gel extracted.

Oligonucleotides in the following pair were annealed separate in 25 µl reactions containing 1X annealing buffer: SgF with SgR, ScaffF with ScaffR and extensF with extensR. In a PCR machine, the mixtures were heated at 95°C and gradually cooled (0.1°C/s) to 22°C. Annealed oligo nucleotide (1 µM) formed of ScaffF and ScaffR was phosphorylated to generate the sgRNA scaffold in a 25 µl reaction containing 1X T4 ligase buffer and 5 U of T4 Polynucleotide Kinase (T4 PNK). PegRNA golden gate assembly reaction (10 µl) containing 100 µM of each annealing oligonucleotide pair, 1X T4 ligase buffer, 30ng of digested p132777 vector and 400 U of T4 ligase, was incubated at room temperature for 1 hour, followed by transformation into DH5α cells.

2.2.16 Maxiprep

PegRNA and sgRNA plasmids were amplified to reach desired concentration required for transfection in human cells using the ZymoPURE™ plasmid maxiprep kit, as described below. PegRNA and sgRNA DH5α colonies for the were picked individually and grown into 150 ml of LB broth, followed by centrifugation at 3400 g for 10 minutes, discarding the supernatant. The pellet was resuspended 14 ml of P1 buffer via pipetting, followed by addition of 14 ml of buffer P2 and immediately mixing by gentle inversion. This was followed by addition of 14 ml of buffer P3, gently mixing until the sample turned yellow and loading the lysate onto the ZymoPURE™ Syringe Filter X and waiting for 10 minutes for the precipitate to float on top. To the filtered solution, 14 ml of Binding Buffer was added, followed by inversion. The mixture (10 ml) was added to the column followed by centrifugation at 500 g for 2 minutes, repeating these steps until the whole mixture has

passed through the spin column. The column was then washed with 10 ml of wash buffer 1 via centrifugation at 500 g for 2 minutes. This was repeated twice with wash buffer 2. The spin column was then washed with no buffer at 16000 g for 1 minutes to remove any residual wash buffer. This was repeated with the addition of Elution buffer (400 μ l) and placing the spin column onto a clean centrifuge tube.

Transfection of U2OS cells

40,000 U2OS cells were seeded in a 24 well plate for 24 hours prior to transfection using the protocol (Lipofectamine 3000 transfection protocol, Thermofisher), described below: U2OS cells were grown to 80% confluency, washed using PBS (Phosphate Buffered Saline), trypsinised and resuspended in 10 ml of DMEM (Dulbecco's Modified Eagle Medium). 10 μ l of 0.4% (w/v) Trypan blue were added to 10 μ l of cells and viewed via bright-field microscopy (Olympus) and cells were counted. This was repeated 3 times to average the number of cells counted, followed by calculation of the total number of cells in 1 ml.

Resuspended cells were centrifuged (Biofuge Primo centrifuge) at 350 g for 10 minutes at room temperature, followed by resuspension with DMEM media in the appropriate ratio to achieve 40,000 cells/ml. Subsequently, 1 ml of cells were added into each well of a 24 well plate and incubated overnight at 37°C.

The lipofectamineTM 3000 reagent (1.5 μ l) was diluted with serum-free media (25 μ l).

This was then added to in a 1:1 ratio with the DNA master mix containing 2 μ l of p300 reagent, 750 ng of Cas9, 250 ng of PegRNA and 870 ng sgRNA plasmids, diluted in

serum free media to a final volume of 50 μ l. The mixture was finally added to the seeded cells (individual well) of the pre-incubated 24-well plate, followed by incubation at 37°C for 3 days.

2.2.17 Multiple sequence alignment

BLAST sequences of DDX49 homologues (See Appendix) obtained from Uniprot protein database (Bateman et al., 2022) were fed into Clustal Omega software program (McWilliam et al., 2013) for multiple protein alignment. Sequence alignment was carried out with orthologs of DDX49, and 3 sequences inferred from homology (Figure 3.2.1). Inferred homology suggests that the protein could have a paralogue nature and exhibit different functionalities from orthologs of DDX49, thereby affecting the robustness of the sequence alignment analysis. (Dandage et al., 2019)

2.2.18 3D structure of DDX49

Visualisation of DDX49 crystal structure was enabled via UCSF chimera (Pettersen et al, 2004), Alpha Fold software programs (Jumper et al., 2021) and Pymol (Schrödinger and DeLano, 2020).

Intrinsic disorder was assessed via IUPred3 software program (Erdős et al., 2021) which predicts the tendency of amino acids to be in a disordered region of a protein. The disorder score of each residue ranges from 0 (No disorder predicted) to 1.0 (Maximum score for disordered residue).

Chapter 3. Analysis of the D-x-D motif function in DDX49

3.1 Introduction

DDX49 has been reported to unwind RNA duplexes in an ATP dependent manner. To further investigate enzymatic activities, DDX49 sequence alignment was performed to identify key conserved motifs of interest. This was followed by site directed mutagenesis of amino acids of interest and overexpression and purification of the mutant protein. The wild-type and mutant proteins were investigated *in-vitro* via substrate binding and unwinding assays and the mutant DDX49^{D422A/D424A} was further investigated *in vivo* using CRISPR Cas-9 prime editing in U2OS cells.

3.2 Sequence alignment of DDX49 homologues

In order to gain a mechanistic insight into the relationship between the sequence of DDX49 and its functions, a sequence alignment of homologues of DDX49 was performed across species from the animal kingdom and yeast, as seen in figure 3.2.1 whereby we observed sequence conservation of Aspartate-Proline-Aspartate (D-P-D) in the mammals and non-mammals, suggesting potential function. The D-x-D motif was shown to be involved in coordinating divalent metal cations for nucleotide sugar binding (Yang et al., 2010), thereby suggesting that mutation of this motif by substituting the charged aspartic acid residues, could disrupt the nucleotide affinity and catalytic activities of wild-type DDX49.

C. elegans	LDELKV--SPKHVTKYVTQVLVAKKEAELKLENQKFG EKKEINRRKELLMSGMDEDEADR	517
D. melanogaster	LTEHP I--DQRMVERIFMQVNVTRRESEMQLDNNDFDERAQNYRRKTWIMEGKDPDQMEA	432
A. gambiae	LTEYSV--DDRLVQRIFMQVKVARAEAEINLDNKDFDERKHKYRRLRWIQEGLDPDEMEA	432
C. lupus	LEELSV--EEAKVLQILTQVNVVRRRECEIKLEAANFDEKKEINKRKQLILEGKDPDLEA	424
M. musculus	LAELVV--EEAEVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEGKDPDLEAK	428
R. norvegicus	LAELAV--EEAQVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEGKDPDLEAK	428
H. sapiens	LEEF SV--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
P. troglodytes	LEEF SV--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
M. mulatta	LEEF SV--EEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEGKDPDLEAK	428
B. taurus	LEEF PV--EEAQVLQILTQVNVVRRRECEIKLEAANFDEKKEINKRKQMILEGKDPDLEAK	336
D. rerio	LKEFPI--EEKEVLKILTQVNVTRRQCEIKLESTDFDEK KINKRKQMILDGKDPDLEEK	428
M. unicolor	LKDFAV--EESV LKILTQVNVTRRRECEIKLESTDFDEKREKNRKQLILDGKDPDLEAK	428
S. punctatus*	LQEF SM--EERDVLGILTQVNVVRRRECEIKLEATDFDEKKEINKRKQMILEGKDPDLEEK	428
P. muralis*	LQEF SV--EEHSVLNILTQVNVVRRRECEIRLEATDFDEKKEINKRKQMILEGKDPDLEAQ	477
S. fasciatus*	LKEYPV--QEKEVLKILTQVNVTRRRECEIKLESTDFDEKKEINKRKQLILEGKDPDLEAK	428

Figure 3.2.1. Multiple sequence alignment of DDX49 homologues using Clustal Omega^o. The highlighted sequence D-P-D shows to be conserved across mammals and non-mammals. DDX49 orthologs were used for sequence alignment; labelled sequences (*) were inferred from homology. (McWilliam et al., 2013)

3.3 Mutagenesis of DDX49

In order to investigate the functional role of the D-P-D sequence, plasmid p100.DDX49 was mutated using site directed mutagenesis techniques in *E. coli* BL21-AI cells. The Aspartate residues D422 and D424, were substituted with Alanine (A) residues, resulting in the sequence A-P-A. Alanine has a short and hydrophobic side chain which does not affect the main protein conformation (Betts and Russell, 2003), but disrupts any interaction that was present between the aspartate residues and charged side chains of the protein. Following analysis of DDX49^{D422A/D424A} using *in-vitro* functional experiments, we hypothesised that residue K421 would be involved in the mechanism of metal coordination for nucleic acid binding; therefore, site directed mutagenesis was performed

to achieve a K421A in collaboration by undergraduate student Louise Martin under my and PhD student Ashley Parkes. Successful site directed mutagenesis was verified sequencing via Sanger sequencing (Genewiz.com).

3.4 Overexpression and purification of DDX49 wild type and select mutants

Overexpression was initiated by transforming in *E. coli* BL21-AI cells with the desired plasmid (wild-type or mutant) previously miniprepmed from *E. coli* DH5- α cells. Bacterial cultured first grown overnight (50 ml) and subsequently upscaled to 2 L and induced with L-arabinose binds to the *AraBAD* promoter in the plasmid and allows efficient transcription of the T7 RNA polymerase, whereas IPTG binds to the LacI repressor releasing it from the *lac operator*, thereby allowing binding of the T7 RNA polymerase to the T7 promoter and expression of your gene of interest found downstream the T7 promoter.

The harvested biomass of DDX49 stored at -80°C was sonicated and centrifuged, and the protein containing supernatant loaded onto the Ni^{2+} -NTA column. The immobilised Ni^{2+} ions of the column resin interact with the Histidine residues (x6) attached to the N-terminus of DDX49, thereby capturing and separating DDX49 from the rest of the bacterial cell contents. The column is washed with Nickel buffer A containing a low concentration of Imidazole. Imidazole can compete for the Ni^{2+} binding site, thereby displacing proteins bound to the column. At a low concentration (25 mM), imidazole displaced proteins forming weak interactions with Ni-NTA resin, thereby discarding non-specific binding of co-contaminant proteins. Gradually increasing the concentration of Imidazole (400 mM) eluted the protein and fractions were analysed via SDS-PAGE gel electrophoresis to confirm the presence of the protein of interest, as seen in figure 3.3.1. The fraction

containing DDX49 were pooled followed by dialysis overnight in Heparin wash buffer. Heparin is immobilised in a porous bead and acts a specific affinity ligand due to its high composition of anionic sulphate groups and its linear and helical structure mimics nucleic acids (Bolten et al.,2018). Therefore, nucleic binding proteins like DDX49 bind to the heparin molecules, separating it from the cell lysate. Upon the addition of low concentration of NaCl (100 mM) in the Heparin wash buffer, weak interactions of co-contaminant proteins are displaced from the column. Gradually increasing NaCl concentration to 1 M, resulted in the elution of DDX49. Fractions containing DDX49 were confirmed via SDS-PAGE analysis, pooled and underwent overnight dialysis in DDX49 storage buffer.

The selected mutants DDX49^{D422A/D424A} and DDX49^{K421A} were purified using the same method described above. DDX49^{K421A} was purified in collaboration with undergraduate student Louise Martin and PhD student Ashley Parkes' supervision.

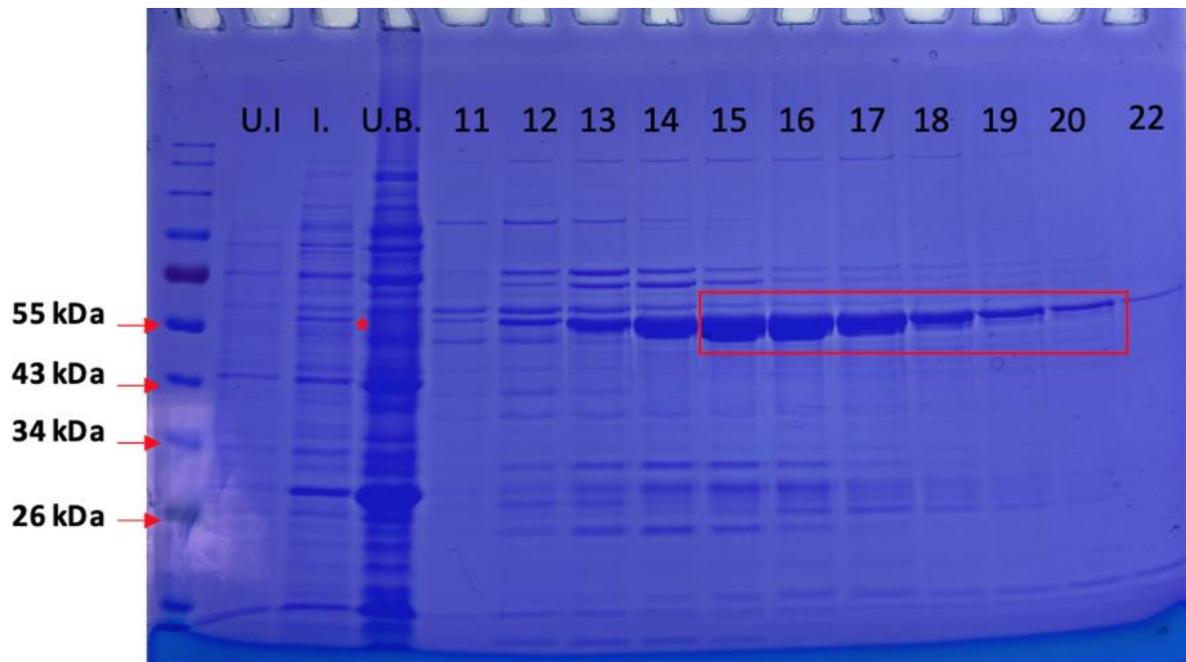


Figure 3.4.1 Coomassie stained 10% SDS page gel showing successful overexpression and elution of *DDX49^{D422A/D424A}* from Ni-NTA column. A band at around 55 kDa is present in lane I. (Induced) as depicted by the red star but is absent in the U.I (Uninduced) lane, confirming successful induction of DDX49 expression (MW=59 kDa). Lane U.B. (Unbound) shows unbound proteins eluted from Ni-NTA column. Lane 11-22 show fraction eluted from a Ni²⁺-NTA column, lanes framed in red represent enriched DDX49 and correspond to the fractions that were subsequently pooled and dialysed.

3.5 Biochemical analysis of DDX49, DDX49^{D422A/D424A} and DDX49^{K421A} unwinding

DDX49 has been established as ATP-dependent RNA helicase, however its behaviour with DNA has not been explored. We investigated DDX49 unwinding with different substrates, to determine potential unwinding activity and polarity.

Unwinding assays were performed using a 10% native gel and the protein was incubated with fluorescently Cy-5 dsDNA substrate Fork2, Fork3 and Fork4.

Substrate re-annealing following unwinding, is reduced via addition of an unlabelled oligonucleotide (TRAP) complementary to the non-labelled strand of the substrate as shown in figure 3.5.1, to visualise more accurate intensity of the single stranded Cy5 labelled oligonucleotide unwound by DDX49.

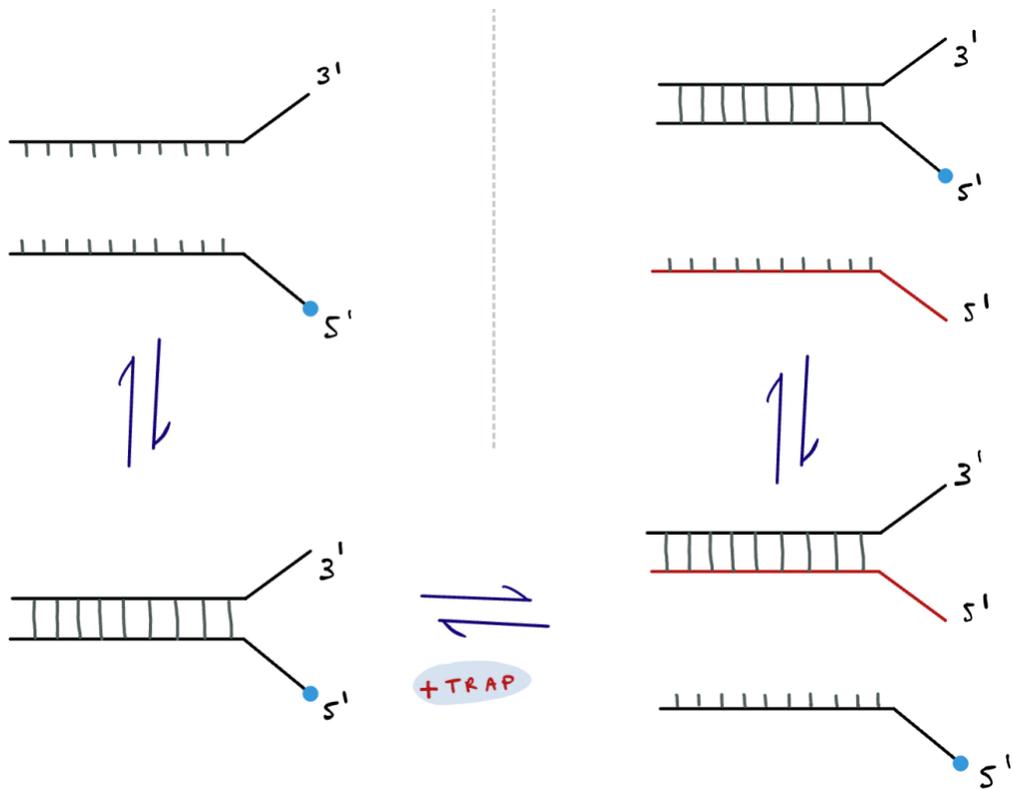


Figure 3.5.1. DDX49 shows unwinding of Fork2 in the presence and absence of TRAP. In the absence of TRAP (single stranded complementary unlabelled nucleotide), the unwound oligonucleotide MW12 and MW14-Cy5 naturally re-anneal, thereby seemingly reducing helicase activity detected. By adding unlabelled MW14, there will be an equilibrium between the labelled and unlabelled strand annealing to MW12, thereby decreasing the formation of Fork2 after it has been unwound by DDX49.

The unwinding assay showed that DDX49 can unwind dsDNA substrate Fork2 and Fork4 but not Fork3 as shown in figure 3.5.2, suggesting it has 3'-5' unwinding polarity. This is because the 3' ends on the non-complementary region of the substrates are single stranded, whereas it is double stranded of Fork3 via binding of the 3' end to PM16, thereby blocking DDX49 from unwinding.

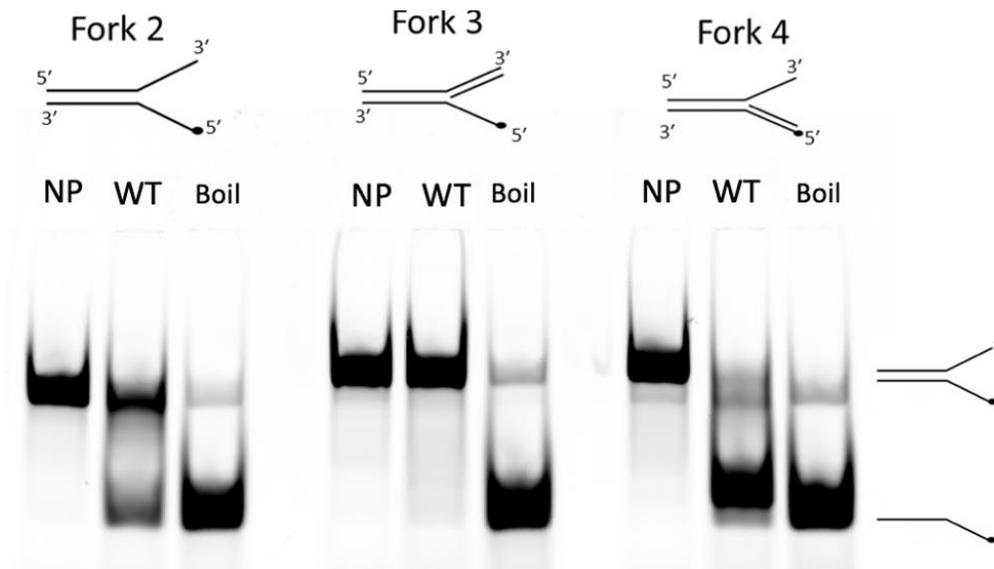


Figure 3.5.2. Unwinding assay of wild-type DDX49 with substrates Fork2, Fork3 and Fork4 in 5 mM ATP and 5 mM Mg²⁺. Figure 3.5.2 shows dark bands representing the Cy5 label on the substrates which travel at different distances based on molecular weight. Control samples without DDX49 include: no protein (NP) showing minimal unwinding and Boil, showing maximal unwinding following heat denaturation at 95°C for 10 minutes. Unwinding is present in the WT lane of Fork2 and Fork4 but is absent in WT Fork3. The 10% Native gel underwent gel electrophoresis for 60 minutes at 140V.

DDX49 unwinding activity on dsDNA was further investigated in the absence and presence of ATP and Mg^{2+} to observe whether its enzymatic activities are ATP-dependant or independent. Figure 3.5.3 shows a 10% native of an unwinding assay whereby DDX49 wild type was incubated in the presence of ATP/ Mg^{2+} , Mg^{2+} only and no ATP/ Mg^{2+} in reaction mixtures containing either Fork2 or Fork4. No unwinding seems to be present in the reactions mixtures containing either Fork2 or Fork4 in the absence of ATP/ Mg^{2+} . Unwinding of Fork2 by DDX49 takes place in the presence of ATP/ Mg^{2+} as well as in the absence of ATP. In both conditions, a smaller DNA fragment is present below the ssDNA, which suggests that DDX49 is behaving like a nuclease and excising the substrate. Fork4 is also unwound by DDX49 in the presence of ATP/ Mg^{2+} , however in the absence of ATP, the ssDNA seems to be fully digested by the nuclease activity of DDX49, as seen by the absence of the ssDNA band in figure 3.5.3.

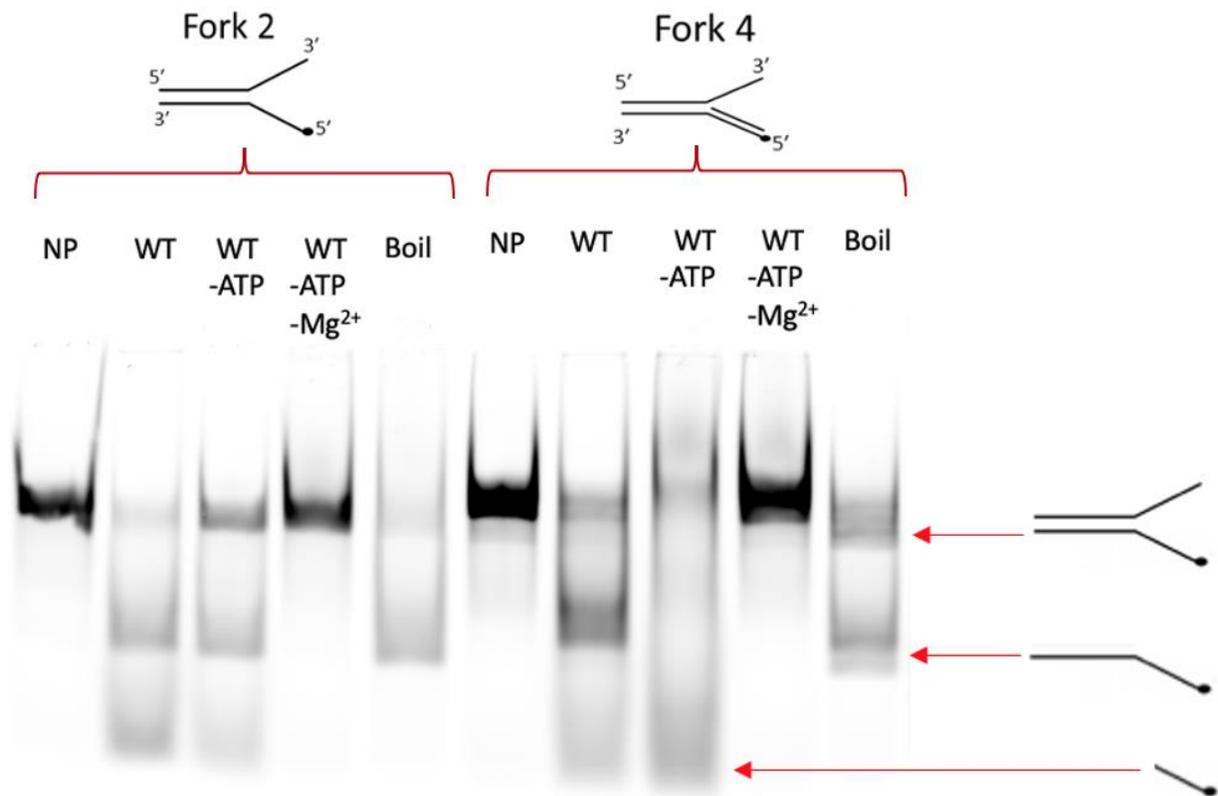


Figure 3.5.3. Unwinding assay of wild-type DDX49 with substrates Fork2, Fork4 in the presence or absence of 5 mM ATP and 5 mM Mg²⁺. Figure 3.5.3 shows dark bands representing the Cy5 label on the substrates which travel at different distances based on their molecular weight. Control samples without DDX49 include: no protein (NP) showing minimal unwinding and Boil, showing maximal unwinding following heat denaturation at 95°C for 10 minutes. Unwinding is present in the WT lane of Fork2 and Fork4 but is absent in the no ATP/Mg²⁺ condition. DDX49 is exhibiting nuclease and helicase activity with Fork2/4 in an ATP independent manner, with F4 being fully digested. The 10% Native gel underwent gel electrophoresis for 60 minutes at 140V.

Unwinding assays were repeated with DDX49 mutant DDX49^{D422A/D424A} to observe changes in unwinding or nuclease activity of DDX49 following mutations of the conserved D-x-D motif of DDX49. Figure 3.5.4 shows a 10% Native gel of DDX49 WT and DDX49^{D422A/D424A} incubated with Fork2 or Fork4. DDX49 WT displayed Fork2/4 unwinding as well as nuclease activity, whereas DDX49^{D422A/D424A} displayed reduced helicase activity and disrupted nuclease activity.

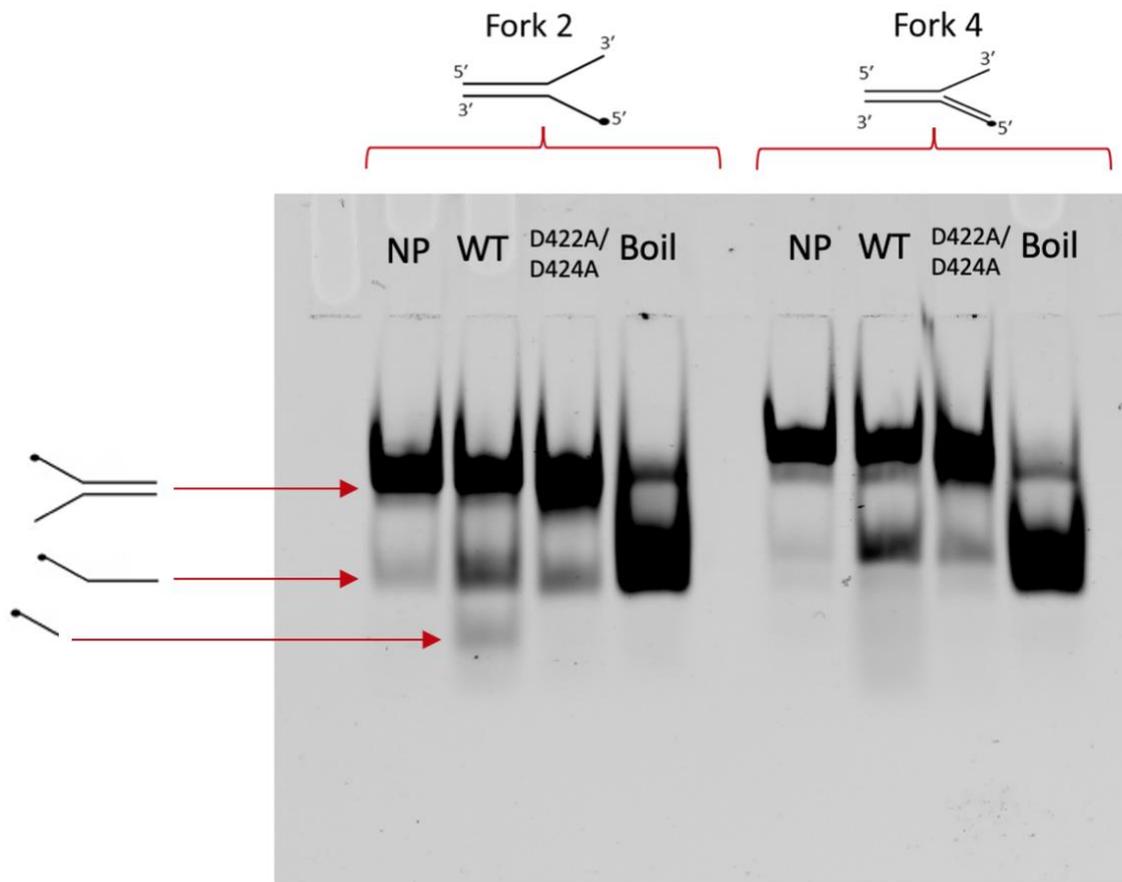


Figure 3.5.4. Unwinding assay of DDX49 and DDX49^{D422A/D424A} with substrates Fork2 and Fork4. Fluorescently labelled Fork2/4 are represented by dark bands. Control samples include No Protein (NP) showing minimal unwinding and Boil, showing maximal unwinding. WT lanes of Fork2 and Fork4 display a ssDNA band of darker intensity compared to the NP sample, indicating unwinding activity. Smaller DNA fragments are present in these lanes, indicating nuclease activity by DDX49, which is absent in the mutant lanes (D422A/D424A).

Site directed mutagenesis of the conserved D-P-D sequence resulted in disruption of the nuclease activity seen in DDX49 wildtype. We therefore hypothesised that the D-x-D motif could be directly involved in the nuclease activity of the protein, via a divalent metal ion coordination by the 2 aspartate residues of the sequence D-x-D, acting in parallel with a nucleophile able to de-protonate a water molecule. Lysine 421 in DDX49 sits just before the conserved sequence, therefore we proceeded to use site directed mutagenesis to achieve a K421A substitution, overexpression and purification of the mutant protein as described for WT DDX49 and DDX49^{D422A/D424A}.

Unwinding assays were performed using DDX49 wildtype and selected mutants to compare their unwinding and nuclease activity, as seen in figure 3.5.5. DDX49 WT and DDX49^{D422A/D424A} behave as previously described in figure 3.5.4. Figure 3.5.5 shows WT lane (DDX49 wild type) behaving as a helicase and nuclease, as seen by the single stranded band and DNA fragment below. DDX49^{D422A/D424A} unwinding activity seems to be reduced as seen by the band of lower intensity compared to WT. DDX49^{K421A} seems to have greater unwinding and nuclease activity compared to WT, as seen by the darker intensity of the band in lane K421A.

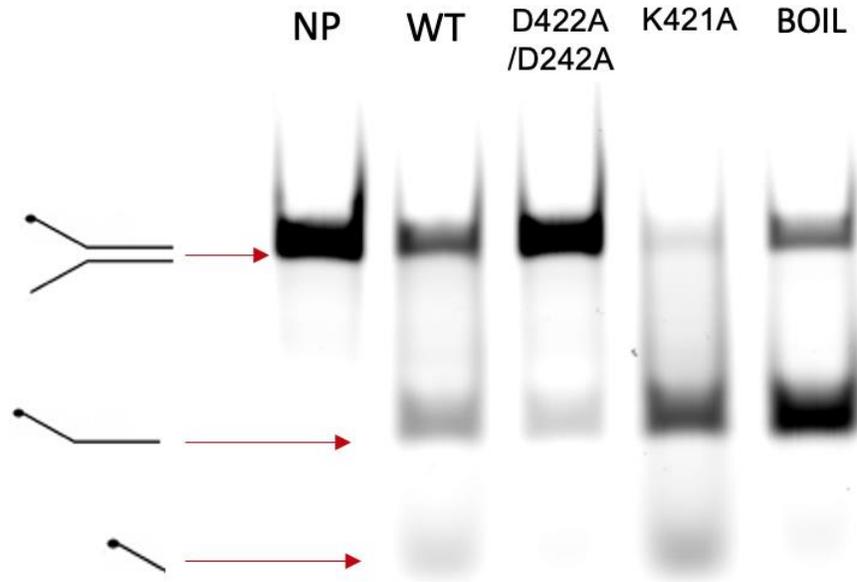


Figure 3.5.5. Unwinding assay (10% native gel) of DDX49 wild type (WT), DDX49^{D422A/D424A} (D422A/D424A) and DDX49^{K421A} (K421) in the presence of 5 mM ATP/Mg²⁺ and 25 nM Fork2. Dark bands represent fluorescently labelled DNA in duplex and single stranded form. Control lanes include No Protein (NP) showing no unwinding and Boil showing most of the substrate being unwound upon heat denaturation at 95 °C for 10 minutes. WT and K421 shows unwinding and excision of the substrate, whereas D422A/D424A shows reduced unwinding compared to WT.

To quantitatively measure the unwinding activity of DDX49 and mutants, a fluorescence resonance energy transfer (FRET) assay was performed. FRET is a distance dependent physical mechanism by which an excited fluorophore (donor) transfers the energy to another fluorophore (acceptor) when in close proximity.

These were carried out using Fret-fork2 duplex formed of Cy5-MW14 and Cy3-MW12.

The difference in emission detected between the substrate in duplex and unwound form allows for a quantitative measure of unwinding activity of a protein, as seen in figure 3.5.6.

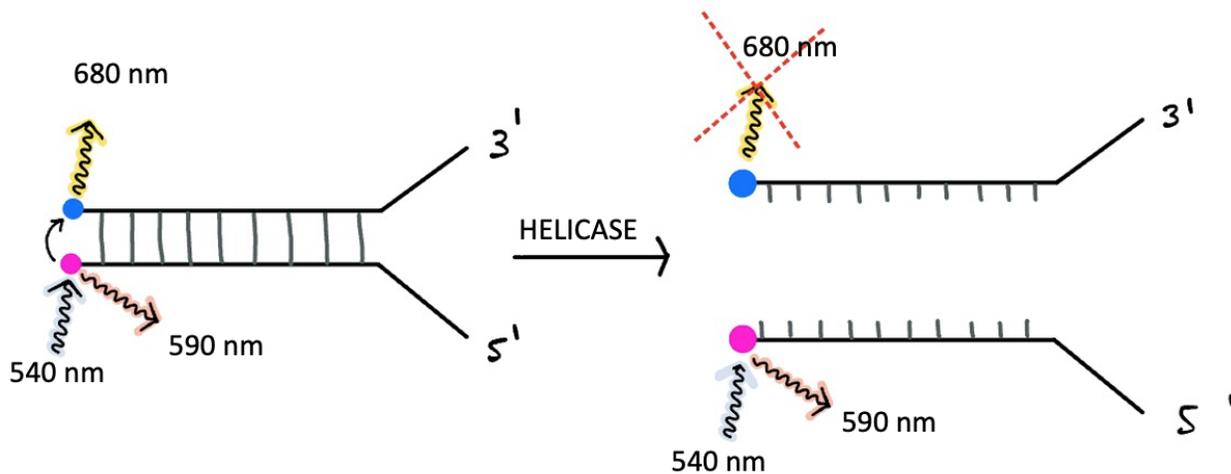


Figure 3.5.6 Fluorescence resonance energy transfer (FRET) diagram showing FretFork2 excitation and emission wavelengths with substrate in duplex and unwound form. The diagram shows FretFork2 in duplex being excited at 540 nm, resulting in emission wavelengths at 590 nm and 680 nm. Once the duplex is unwound, emission at 680 nm is lost. Unwinding is directly proportional to the signal at 680 nm loss, which can be detected to quantitatively measure helicase activity of a protein.

DDX49 WT, DDX49^{D422A/D424A} and DDX49^{K421A} were incubated with 50 nM FretFork2 substrate in reaction mixtures containing 5 mM ATP/Mg²⁺, for 30 minutes at 37°C. Readings were taken at 1 minutes intervals using the BMG FLUOstar Omega. DDX49 wildtype unwound around 50-60% of substrate, whereas DDX49^{D422A/D424A} unwound only 20% of substrate as shown in figure 3.5.7, in agreement with the previous gel-based results obtained (figure 3.5.5). On the other hand, DDX49^{K421A} unwound 75% of substrate at a faster rate than the wildtype, suggesting it is behaving as a hyperactive helicase as seen in figure 3.5.7.

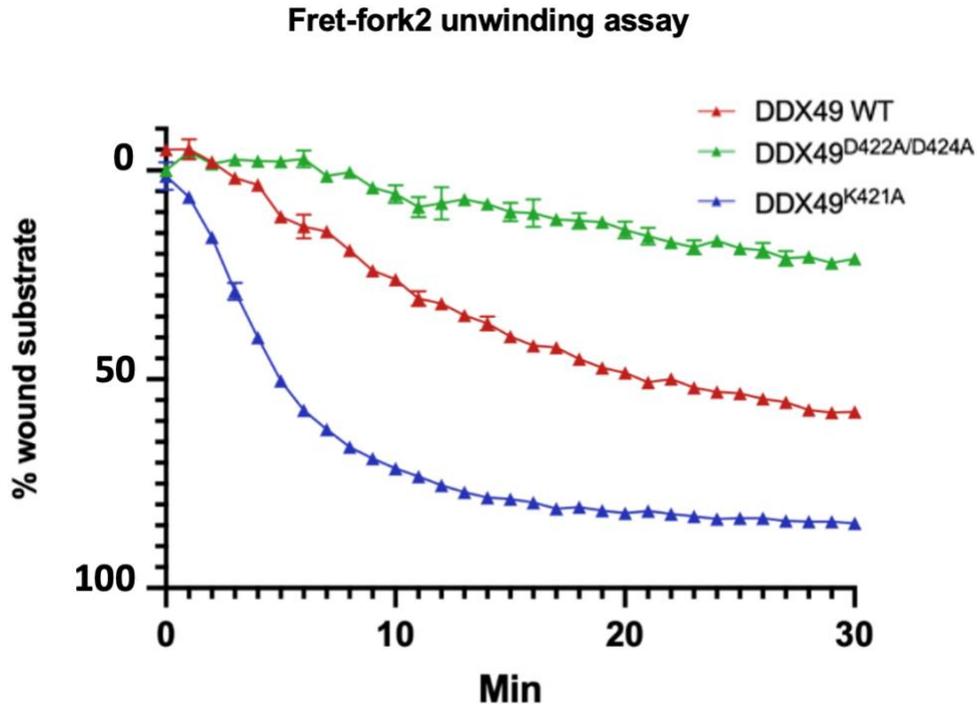


Figure 3.5.7. FRET assay of DDX49 WT, DDX49^{D422A/D424A} and DDX49^{K421A} with substrate Fret-Fork2 in 5 mM ATP and 5 mM Mg²⁺. WT shows around 50-60% unwinding of substrate, whereas DDX49^{D422A/D424A} shows impaired helicase activity, achieving only 20% unwinding. DDX49^{K421A} reaches 75% unwinding, suggesting it is acting as a hyperactive helicase.

3.6 Biochemical analysis of DDX49, DDX49^{D422A/D424A} and DDX49^{K421A} DNA binding affinity.

To investigate the ability of DDX49 and select mutants to bind to DNA substrate fork2, an electrophoretic mobility shift assay (EMSA) was carried out at different protein concentrations (0 nM, 250 nM, 500 nM and 1 μ M). Reactions (20 μ L) were loaded onto a 5% native gel and underwent native gel electrophoresis to separate the DNA/protein bound complex and the unbound DNA (Fork2). This occurs due to samples with higher molecular weight travelling slower through the gel compared to lower molecular weight samples.

DDX49 showed a concentration dependent binding to Fork2, with binding to substrate occurring at the lowest protein concentration used (250 nM). DDX49^{D422A/D424A} showed decreased binding affinity to Fork2 as seen in figure 3.6.1 (A) by the shift of the faint band occurring at 500 nM, with increased binding at 1 μ M. DDX49^{K421A} also displayed reduced binding affinity to Fork2 as seen by the weak shift in binding at 1 μ M in figure 3.6.1 (B). The DNA/protein complexes formed by DDX49 WT and DDX49^{K421A} showed in 3.6.1 (B) were running at different positions, suggesting that the two proteins might be binding to the substrate in different modalities, via polymers and monomers of DDX49 WT and DDX49^{K421A} respectively.

A)

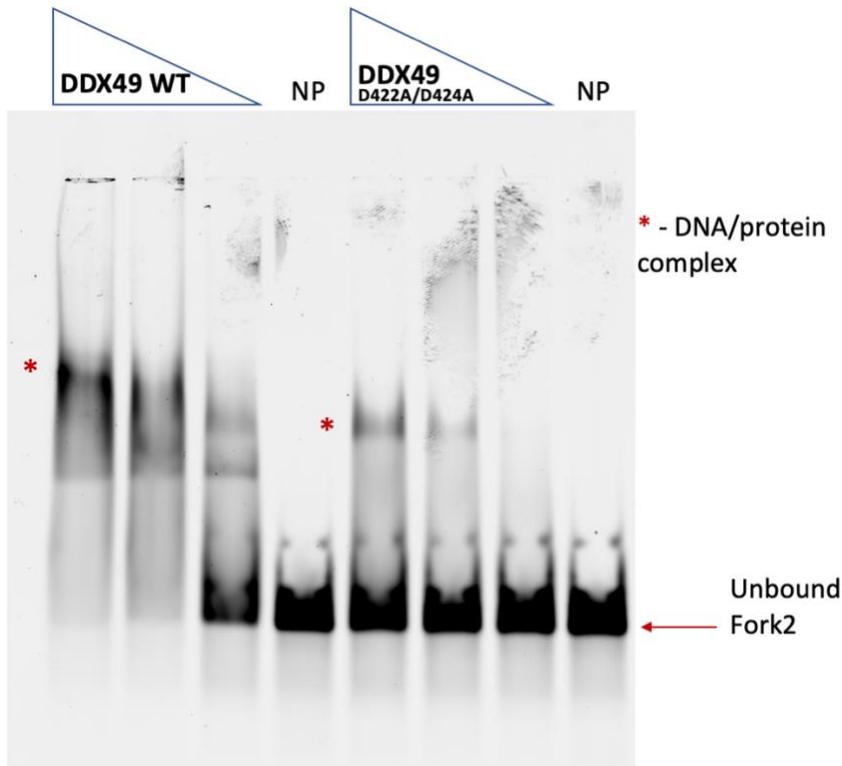
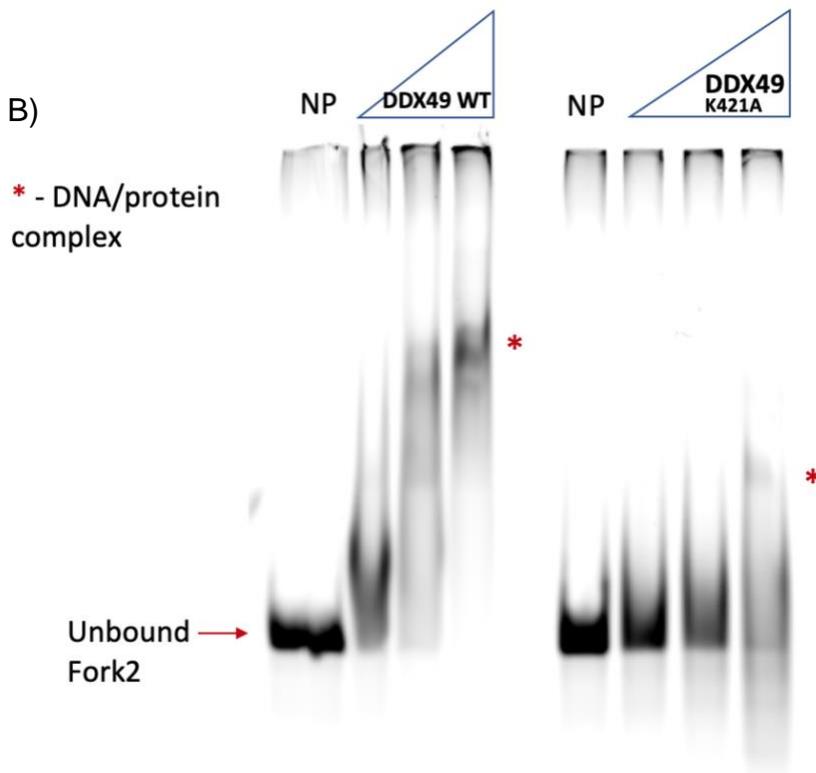


Figure 3.6.1.
Electrophoretic shift mobility assay (EMSA) of DDX49 WT, DDX49^{D422A/D424A} and DDX49^{K421A} (0-1 μ M) with Fork2.

(A) 5% Native gel of DDX49 and DDX49^{D422A/D424A} at different protein concentrations (1 μ M, 500 nM, 250 nM) and 0 nM (NP) show that DDX49 has a higher binding affinity for Fork2 compared to DDX49^{D422A/D424A}, which weakly binds to DNA at 500 nM.

B)



(B) 5% Native gel of DDX49 and DDX49^{K421A} at different protein concentrations (1 μ M, 500 nM, 250 nM) and 0 nM (NP) show that DDX49^{K421A} is weakly binding DNA at 1 μ M, compared to WT which binds to Fork2 at 250 nM.

3.7 CRISPR-Cas9 mediated prime editing of a DDX49 conserved motif in U2OS cells

Sequence alignment (figure 3.2.1) has highlighted the conservation of the sequence D-P-D; this motif has been shown to be present in other proteins (D-x-D) and to be responsible for coordination of a metal cation for nucleic acid binding. *In-vitro* biochemical assays described in section 3.5, show that mutating the D-P-D sequence causes loss of nuclease activity and reduced DNA substrate binding. In order to investigate the effects of mutating the DDX49^{D422A/D424A} mutant *in vivo*, CRISPR-Cas9 mediated prime editing of the D-P-D sequence was used in order to substitute D422 and D424 into alanine residues. This would allow generation of a phenotype in U2OS cells in which the CRISPR-Cas9 prime editing would result in genome editing of the conserved motif and measurement of changes in cell viability.

Prime editing is a genome editing tool which enables targeted insertion, deletion or substitution by fusing an RT to a prime editing guide RNA (pegRNA) and a Cas9-nickase, which is reprogrammed to only cleave one DNA strand thereby forming a 'nick' in the host genome.

The pegRNA contains a spacer sequence which can hybridise to a target site on the host DNA. The Cas9 nickase primes initiation of reverse transcription by generating a 3'-hydroxyl group, which allows RT to extend the pegRNA containing the desired mutation into the target site, as seen in figure 3.7.1.

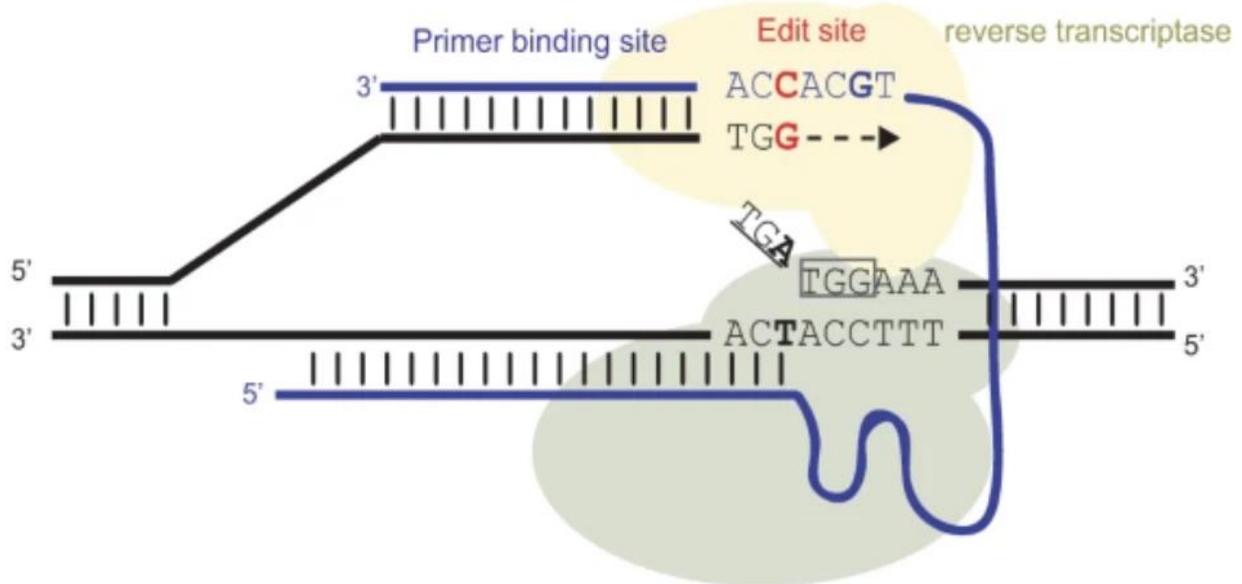


Figure 3.7.1. Prime editing overview. The nCas9 recognises the PAM sequence TGG and induces a nick in the dsDNA thus allowing the pegRNA to anneal to the host DNA forming a primer for new DNA synthesis by RT. The RT extends the 3' flap incorporating the desired mutation. FEN1 will displace the PegRNA allowing the newly edited 3' flap to anneal back on the DNA strand, where the mutated codon will cause formation of a bulge. (Scholefield and Harrison, 2021)

PegRNA and sgRNA were generated co-operatively with me, undergraduate student Louise Martin and PhD student Ashely Parkes. The seeding, transfection and imaging of U2OS was carried out by PhD student Ashley Parkes.

40,000 U2OS cells were seeded in a 24 well plate and transfected as described in method 2.2.15. These were then transfected with Cas9 plasmid (P169850), PegRNA (p32777) and sgRNA (56777) and along with lipofectamine 3000 reagent. To control for

contaminants in the plasmids to be transfected, cells were transfected with either Cas9 plasmid (P169850), PegRNA (p32777), sgRNA (56777), lipofectamine 3000 (0.75 μ l and 1.5 μ l) and grown in media only. The cells were imaged using brightfield microscopy (Olympus). U2OS cells grown in media-only maintained full confluency, whereas prime edited cells showed no survival, as see in figure 3.7.2. Control samples exhibited full survival, with some reduced confluency in control sample with 1.5 μ l of transfection reagent lipofectamine 3000. In order to quantitatively measure confluency, an automated cell counter could be used to accurately measure viability and upon gene editing (Stoddart, 2011).

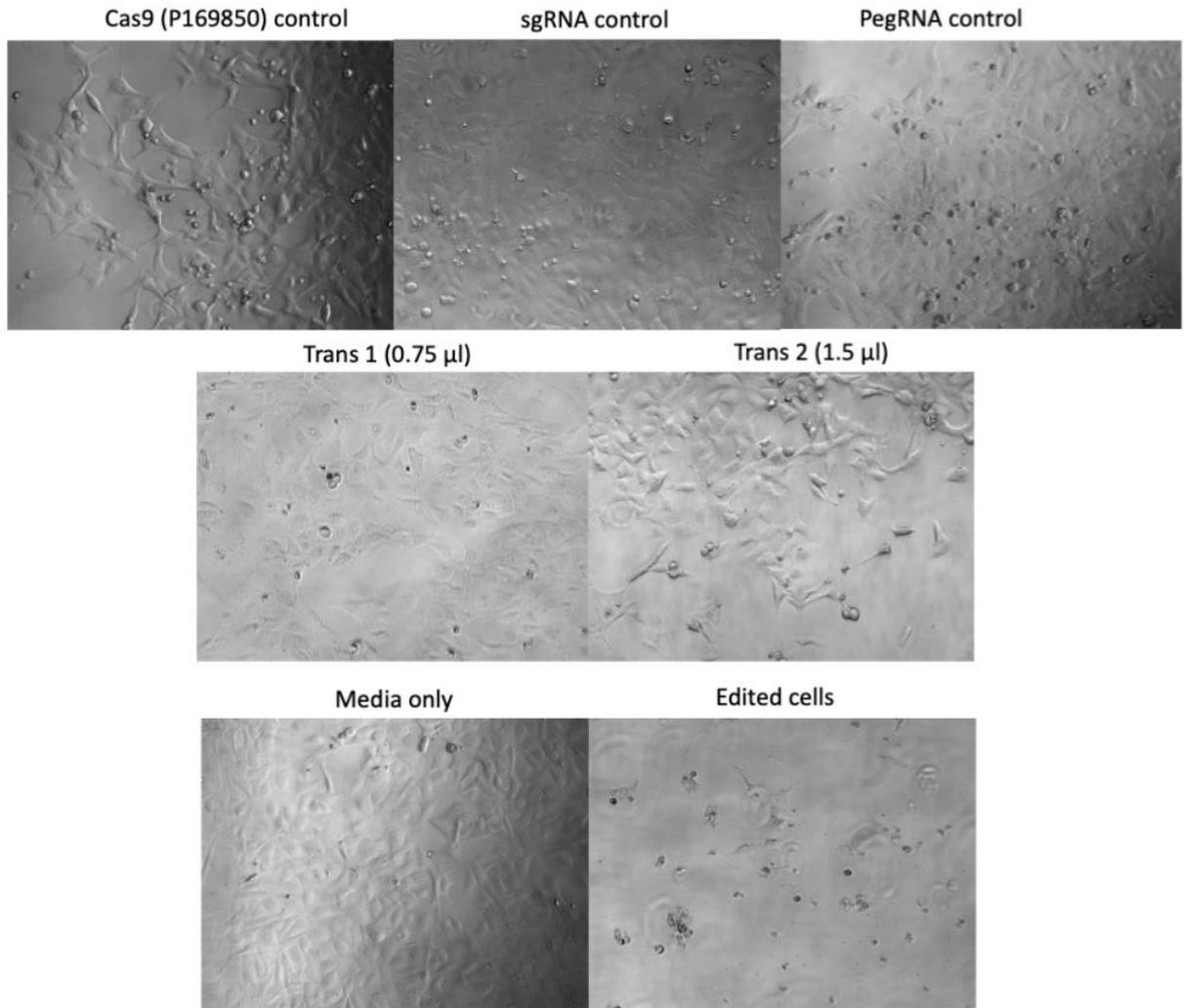


Figure 3.7.2. Brightfield microscopy images showing U2OS cells following CRISPR-Cas-9 prime editing of the conserved D-x-D motif of the DDX49 gene. Cells were imaged after three-day incubation at 37 °C, following transfection with Cas9 plasmid (P169850), PegRNA (p32777), sgRNA (56777) with lipofectamine 3000 reagent. Control cells grown in media only maintain full confluency, whereas prime edited cells exhibit no cell viability as shown by the absence of cells. Cells transfected with individual transfection plasmids, 0.75 μl and 1.5μl lipofectamine 3000 reagent only (Trans 1 and Trans 2) show cells unaffected, with reduced confluency in the 3000 reagent only.

3.8 Summary of Chapter 3

In summary, DDX49 wildtype was shown to have 3'-5' DNA helicase activity and DNA binding, with novel nuclease activity of DNA substrates Fork2 and Fork4. Following mutagenesis of the conserved region of DDX49 (D-x-D), the mutant DDX49^{D422A/D424A} showed disrupted of nuclease activity and reduced DNA helicase and binding activity. Mutagenesis of K421 adjacent to the conserved motif, resulted in mutant DDX49 hyper-active helicase activity, reduced DNA binding and unaffected nuclease activity. In vivo mutagenesis of the conserved motif (D-x-D) using CRISPR-Cas9 mediated prime editing of U2OS cells resulted in no cell viability.

Chapter 4: Analysis of IDR removal on DDX49 catalytic activities

4.1 Introduction

Despite the highly conserved structure of the helicase core, DEAD-box proteins exhibit a range of specialised activities, such as the ability to form liquid-liquid phase separation by DDX19. (Collins et al, 2009). These are conferred through non conserved structured such as intrinsically disordered regions (IDRs). In order to identify potential IDRs in DDX49, IDR software predictors fIDPnn and IUPred3 were used, as shown in figure 4.1.1 and 4.1.2.

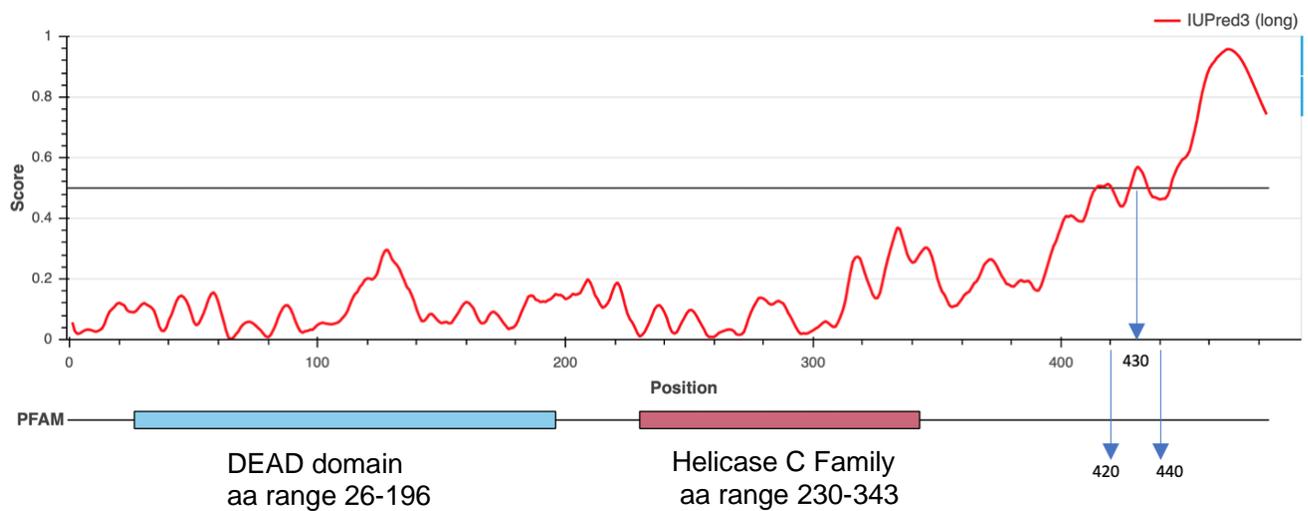


Figure 4.1.1. IUPred3 predictor of structural disorder of DDX49 residues. The IUPred3 predictor software predicts the probability of a given residue (aa) to be a part of an intrinsically disordered region of a protein. Here the scores are >0.5 between residue 420-440, more specifically around residue 430. IUPred3 webserver available at: <https://iupred.elte.hu> .(Erdős et al., 2021)

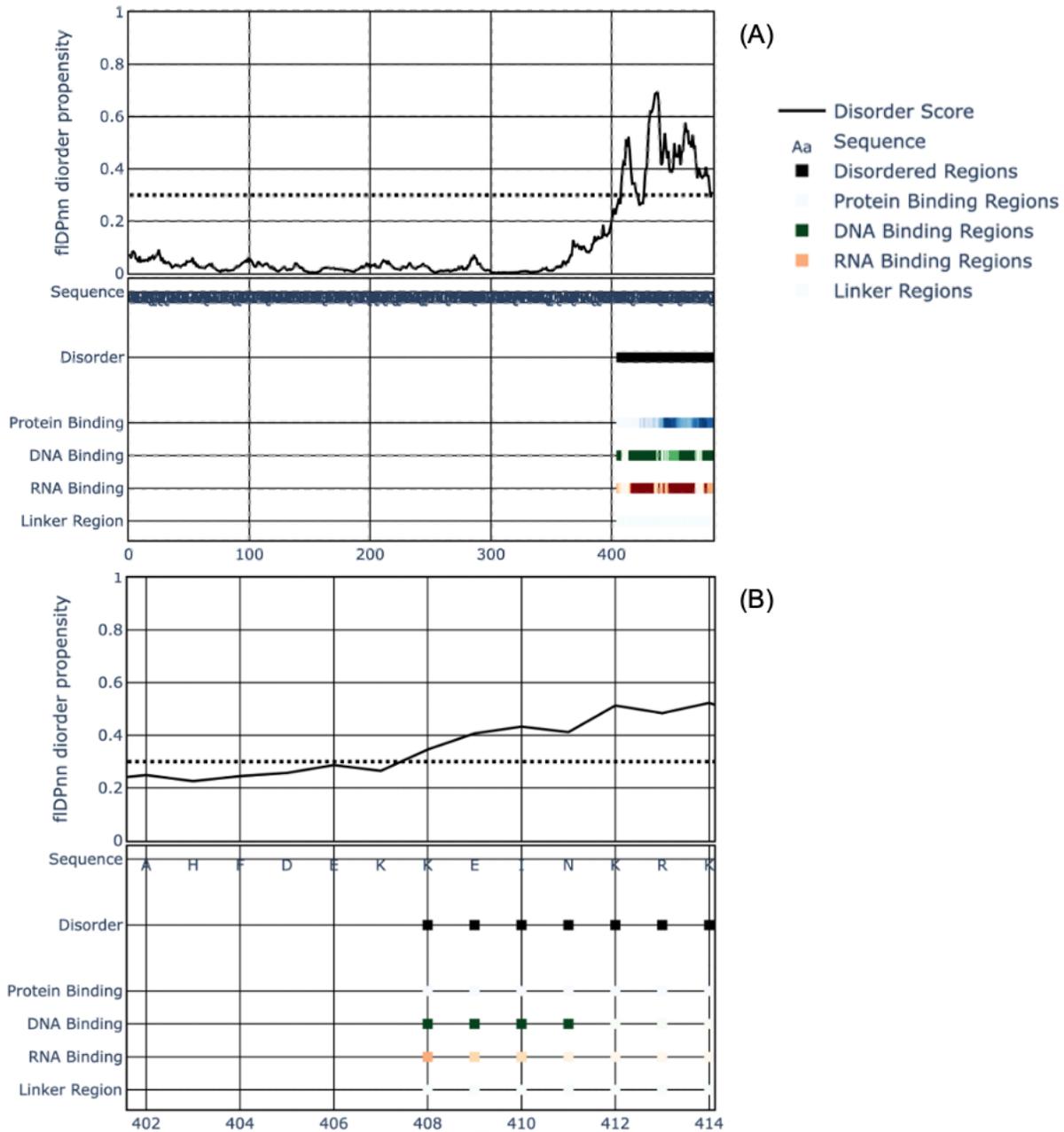


Figure 4.1.2. fIDPnn disorder propensity of residues of DDX49. (A) The disorder propensity is above the threshold of 0.3 in the C terminus of DDX49, which is also predicted to have protein, RNA and DNA binding regions, as shown by the dark coloured bars. (B) Disorder begins 408K as seen by its disorder propensity >0.3. fIDPnn available at: <http://biomine.cs.vcu.edu/servers/fIDPnn/>. (Hu et al., 2021)

Figures 4.1.1 and 4.1.2 show high probability of the C-terminus region of DDX49 to be an IDR. In figure 4.1.1, the IUPred3 predictor shows that around residue 430, the probability score is above the threshold of 0.5, suggesting a high probability that the IDR region of DDX49 begins around residue 430. On the other hand, the fIDPnn predictor in figure 4.1.2 shows residue 408 to score above the threshold of 0.3, suggesting that the IDR region of DDX49 begins at residue 408.

In order to investigate whether the C-terminus of DDX49 is essential for its catalytic activities, site directed mutagenesis of DDX49 was performed to generate C-terminus truncated mutants DDX49 Δ K408-V483 and DDX49 Δ E432-V483. These were tested using unwinding assays and DNA binding assays. Mutant proteins were generated using the same method of DDX49 wild type described in section 3.4 and visualised using a 10% SDS-page gel as shown in figure 4.1.3. DDX49 wildtype, DDX49^{D422A/D424A} and DDX49^{K421A} bands are seen slightly above the 55k Da mark, which corresponds to their molecular weight (MW) of 59 kDa. DDX49 Δ K408-V483 (MW = 45.8 kDa) and DDX49 Δ E432-V483 (MW = 48.6 kDa) bands show reduced molecular weight and consequently travelled more through the gel (fig 4.1.3), due to the truncation of the C-terminus of DDX49.

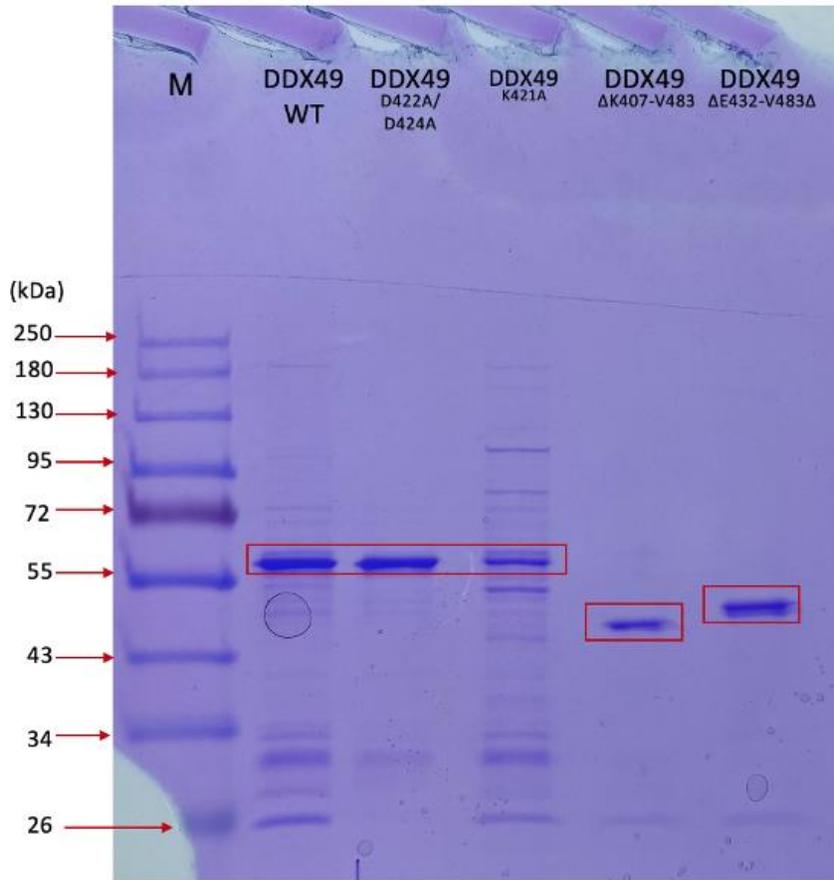


Figure 4.1.3. Coomassie stained 10% SDS PAGE-gel of purified DDX49 wild type (WT), $DDX49^{D422A/D424A}$, $DDX49^{K421A}$, $DDX49^{\Delta K408-V483}$ and $DDX49^{\Delta E432-V483}$. A molecular marker (M) is used to measure molecular weight of the proteins. $DDX49$ wild type and selected mutants are encircled in red.

4.2 Analysis of $DDX49^{\Delta K407-V483}$ and $DDX49^{\Delta E432-V483}$ helicase activity

Unwinding assays were performed using a 10% native gel and the protein was incubated with fluorescently Cy-5 dsDNA substrate Fork2. Reaction mixtures were incubated with 1 μ M protein for 30 minutes at 37 °C, followed by native gel electrophoresis for 60 minutes at 140 V. Figure 4.2.1 shows that $DDX49^{\Delta K408-V483}$ and $DDX49^{\Delta E432-V483}$ perform no unwinding with Fork2, unlike $DDX49$ wildtype, as show in figure 3.5.3. This suggests that the IDR's of $DDX49$ may be essential for DNA unwinding and nuclease activity.

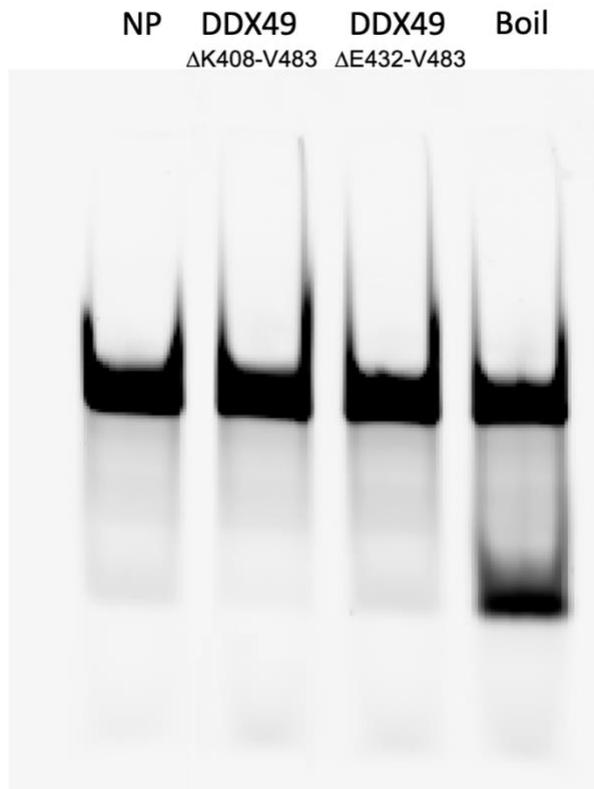


Figure 4.2.1. Unwinding assay of DDX49 Δ K408-V483 and DDX49 Δ E432-V483 with DNA substrate fork2. Dark bands represent the Cy5 label of Fork2. Control samples without DDX49 include no protein (NP) showing minimal unwinding and Boil, showing maximal unwinding following 10-minute heat denaturation at 95°C. No unwinding activity is seen with either mutant protein. The 10% Native gel underwent gel electrophoresis for 60 minutes at 140V.

4.3 Analysis of DDX49 Δ K407-V483 and DDX49 Δ E432-V483 DNA binding

To investigate the ability of DDX49 Δ K408-V483 and DDX49 Δ E432-V483 DNA binding activity to substrate fork2, an electrophoretic mobility shift assay (EMSA) was carried out at different protein concentrations (0 nM, 250 nM, 500 nM and 1 μ M). Reactions (20 μ L) were incubated for 30 minutes at 37°C, followed by loading on a 5% native gel. Samples underwent native gel electrophoresis to separate the DNA/protein bound complex and the unbound DNA (Fork2). No DNA binding was detected of DDX49 Δ K408-V483 and DDX49 Δ E432-V483 as shown in figure 4.3.1, suggesting that the IDR of DDX49 may be essential for DNA binding activity. A limitation of the assay is the lack of a positive control

in figure 4.2.1, which would allow for comparison wild-type DDX49 induced substrate unwinding.

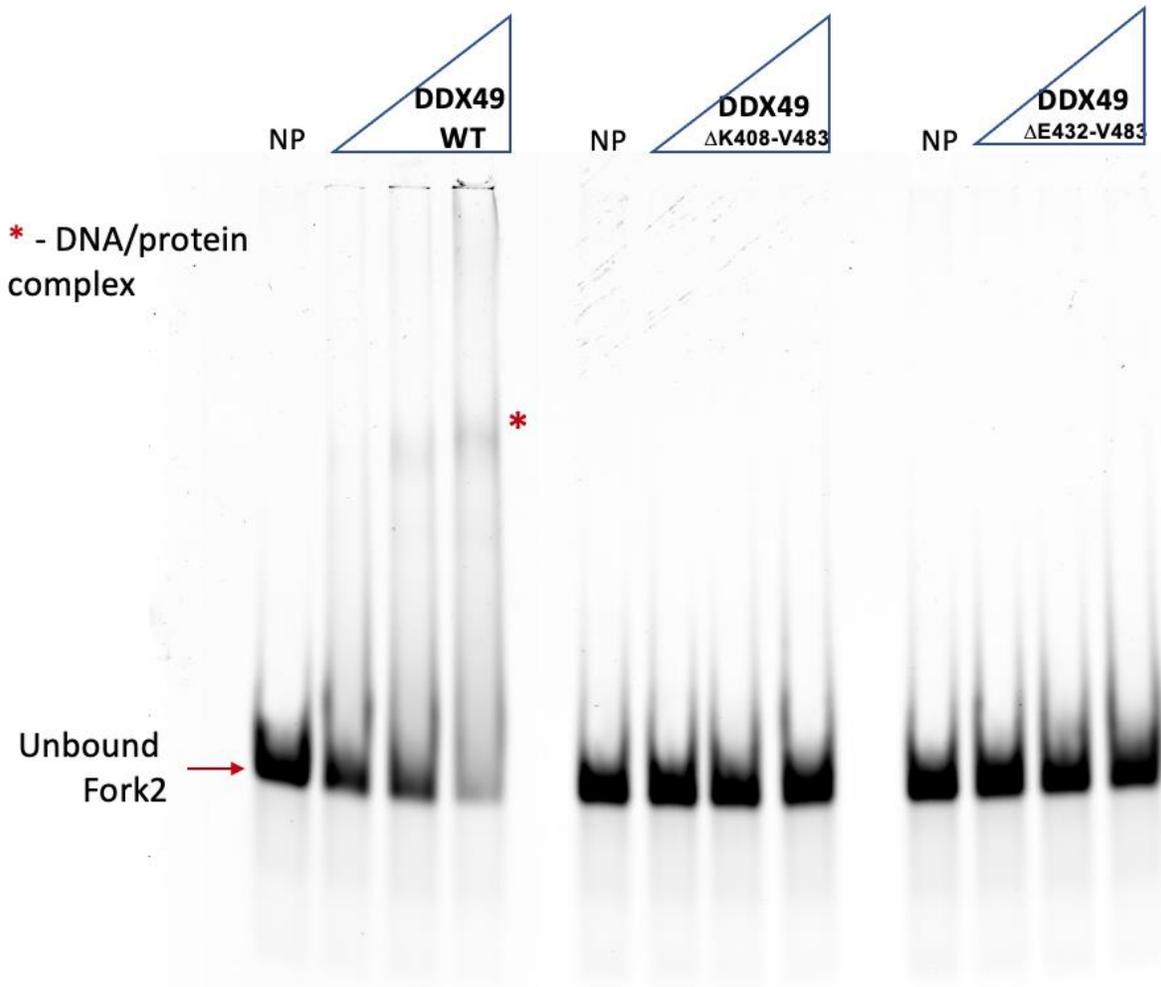


Figure 4.3.1. Electrophoretic mobility shift assay of DDX49 wild type (WT), $DDX49^{\Delta K408-V483}$ and $DDX49^{\Delta E432-V483}$ at varying concentrations (250 nM, 500 nM and 1 μ M) and 0nM (NP) with fork2. DDX49 Wild type shows concentration dependant binding to fork2, as seen by the band shift (*) representing the protein/fork2 bound complex. Unbound fork 2 travels more during gel electrophoresis due to its lower molecule weight. $DDX49^{\Delta K408-V483}$ and $DDX49^{\Delta E432-V483}$ show no DNA binding activity.

4.4 Summary of Chapter 4

In summary, software predictors were used to identify the C-terminus intrinsically disordered region (IDR) of DDX49, showing it begins at residues 408 or around 430. Following this analysis, we performed site directed mutagenesis to generate 2 mutant proteins lacking the IDR from position K408 and E432. The overexpressed and purified DDX49^{ΔK408-V483} and DDX49^{ΔE432-V483} show no unwinding activity or DNA binding activity with Fork2.

Chapter 5: Discussion

In this project we aimed to produce DDX49 mutant forms, targeting a conserved amino acid sequence via site directed mutagenesis *in-vitro* and prime editing in U2OS cells. Furthermore, we aimed to investigate the role of DDX49 intrinsically disordered region by removing its C-terminus via site directed mutagenesis at 2 different sites. These were followed by *in-vitro* functional analysis of DNA binding and unwinding data of wild-type and mutant proteins with DNA fork substrates.

5.1 DDX49 wild-type exhibits DNA helicase and nuclease activity

Investigational studies into the enzymatic activities of DDX49 (Awasthi et al., 2018) have established its ATP-dependent RNA helicase activity however, insights into its ability to interact with DNA are lacking. Thus, we investigated DDX49 wild type and mutant unwinding activity with DNA fork substrates.

By incubating DDX49 with different fork substrates we observed unwinding of DNA duplexes in a 3'-5' direction. This was shown by the effective unwinding of Fork2 and Fork4, and no unwinding with Fork3 seen in figure 3.5.2. DNA fork 3 has a double stranded 5' end, whereas Fork2 is a simple duplex and Fork4 has a double stranded 3' end, as seen in figure 4.1.1.

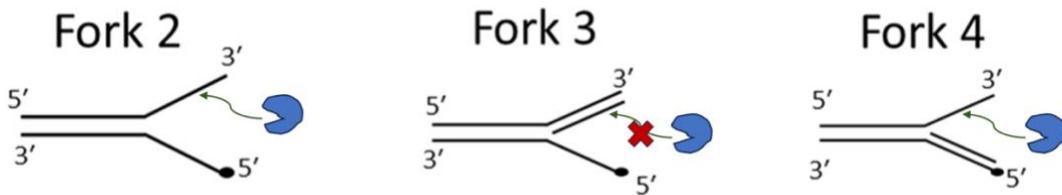


Figure 4.1.1. DNA fork substrate structure and coupling with DDX49. DDX49 as depicted in blue can bind to Fork2 and Fork4 as their 3' end is ssDNA, leading to strand separation. The 3' end of Fork3 is double stranded, preventing DDX49 binding and unwinding.

Unwinding assays also revealed that DDX49 was excising the DNA fork substrates, as seen by DNA fragments with smaller molecular weight than the ssDNA in figure 3.5.3.

A study Awasthi et al (2018) describes DDX49 as an ATP dependent RNA helicase, however no unwinding assays were carried out with DNA based substrates. Furthermore, the duplex used in the study was composed of a 10-mer labelled oligonucleotide and a 40-mer unlabelled nucleotide. Whereas, in the unwinding assays carried out in this project, labelled oligonucleotides used to generate duplex substrates are 50 nucleotides long and resulted in nuclease activity being detected. This difference in labelled oligonucleotide length used in this project and the Awasthi et al. (2018) study, suggests that DDX49 nuclease activity has not previously been detected, potentially due to the short length of the labelled oligonucleotide (10 nucleotides). Any nuclease activity taking place on the 40 mer oligonucleotide would have gone undetected due to its unlabelled nature. DDX49 also showed to unwind DNA Fork2 and Fork 4 in the absence of ATP as seen in figure 3.5.3. Previously, Awasthi et al (2018) had described DDX49 as an ATP-dependant RNA helicase, however due to DDX49 not being able to act as a nuclease on

short oligonucleotides, the unwinding seen in this project could be a side effect of the ATP-independent nuclease activity of DDX49. In order to investigate this further, an unwinding assay could be repeated in the presence of ATP and the absence of magnesium, to show whether limiting DDX49 nuclease activity (limited by the absence of Mg^{2+}), results in an ATP-dependent helicase activity of the protein.

5.2 DDX49 D-x-D motif is essential for nuclease activity

Following sequence alignment and structure prediction analysis we identified a putative functional motif in the auxiliary domain of DDX49, involving residue D422 and D424. These residues resemble the conserved D-x-D motif found in other proteins such as glycosyl transferases (Bush et al., 1998) and TOPRIM nucleases (Yang et al., 2010), whereby it is suggested to coordinate a divalent metal cation for nucleotide sugar binding. Thus, we hypothesised that D422 and D424 could coordinate the Mg^{2+} used in the unwinding assays, which in turn forms electrostatic interactions with the negatively charged sugar phosphate backbone of DNA. Nucleases utilize a variety of nucleophiles, the most common being water to cleave a scissile phosphate bond (Yang, 2010). The cleavage of a phosphodiester bond could be achieved via activation of a nucleophile such as water via a positively charged amino acid such as lysine (K), as shown in figure 4.1.2.

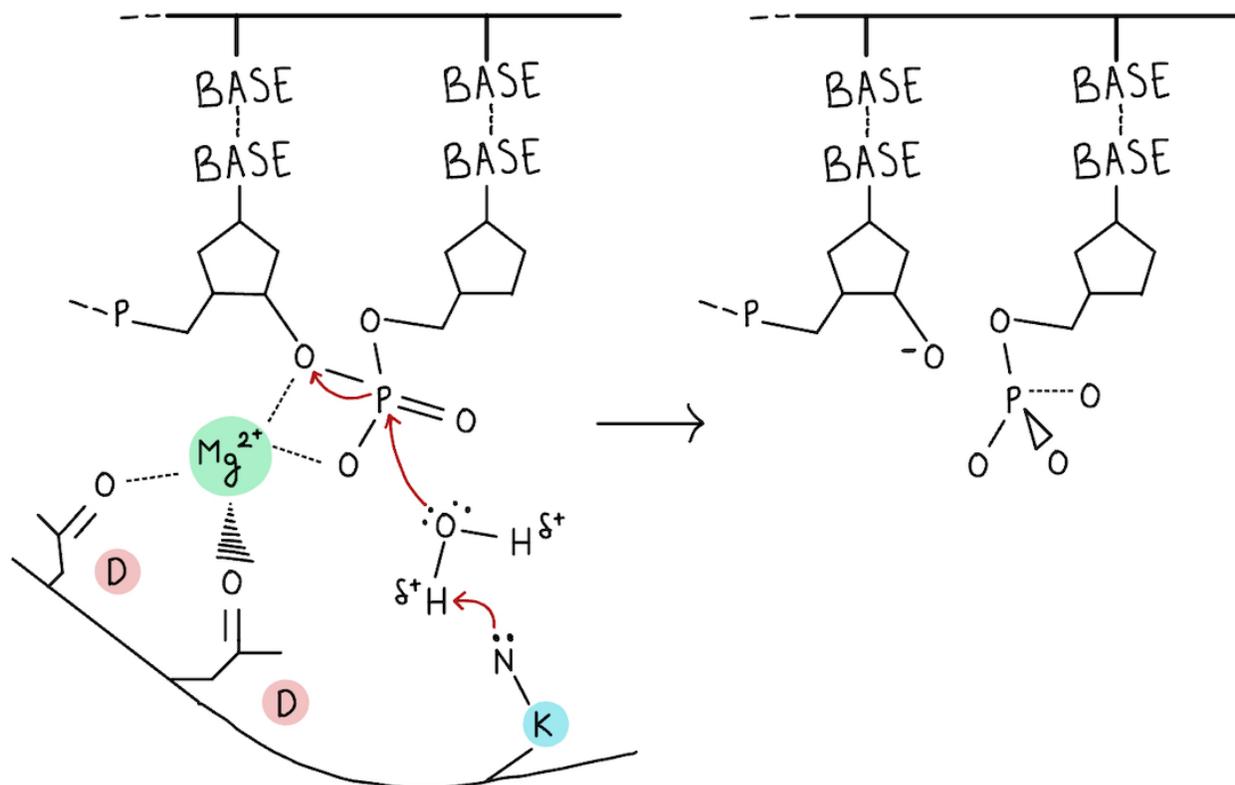


Figure 4.1.2. Potential metal ion dependent catalysis mechanism of DDX49. The diagram shows aspartic acid residues D422 and D424 coordinating a Mg^{2+} which forms electrostatic interaction with the negatively charged nucleic acid strand backbone. The lysine residue (K) deprotonates water, which in turn acts as a nucleophile as represented by the red arrows. This results in the excision of the phosphodiester bond as seen on the right-hand side of the diagram.

In order to investigate whether the D-x-D motif of DDX49 contributed to the helicase and/or nuclease activity observed, site directed mutagenesis was performed in order to substitute the aspartic acid residues into alanines, followed by overexpression and purification of the mutant protein DDX49^{D422A/D424A}. Unwinding assays of DDX49^{D422A/D424A}

showed significantly reduced unwinding, only reaching 20% substrate unwinding compared to the 50-60% unwinding of DDX49 wildtype as seen in figure 3.5.7.

DDX49^{D422A/D424A} mutant did not exhibit any nuclease activity and significantly reduced DNA binding affinity, as seen in figure 3.6.1. Due to DDX49^{D422A/D424A} mutant having significantly reduced catalytic activities, we hypothesised that DDX49 nuclease activity depicted in figure 4.1.2 could involve the conserved residues D422, D424 and K421. Thus, we proceeded to substitute K421 with an alanine residue, overexpress and purify the mutant protein to produce DDX49^{K421A}. This mutant exhibited nuclease activity and unexpected hyperactive helicase activity. This implied that K421 was not involved in the nuclease activity of DDX49, however it could act as a molecular hand break for DDX49 by regulating its helicase activity. Figure 4.1.3 shows a 3D model from residue K421-D424, whereby the K421 is facing a different direction compared to D422/D424. This would suggest that K421 due to its position might not be involved in the potential metal ion dependent catalysis mechanism of DDX49 depicted in figure 4.1.2. An alternative lysine proximal to the D-x-D motif, such as K428 could instead be involved in the mechanism, due to its positioning of the side chain, facing the same direction as D422 and D424, as shown in figure 4.1.4.

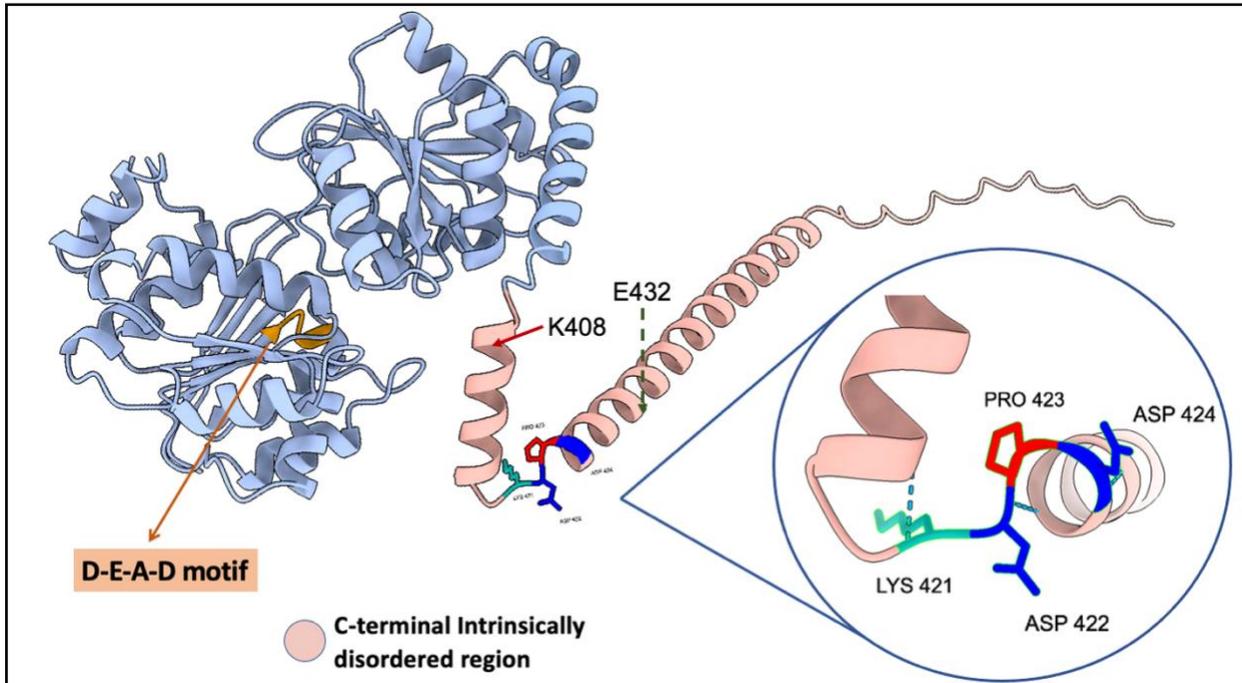


Figure 4.1.3. DDX49 and K421-D424 3D structure as modelled by UCSF chimera. The C-terminus intrinsically disordered region (IDR) is highlighted from residue D404, showing the IDR as helix-turn-helix motif with residues and K421-D424 position at the junction between the 2 helices. Residues K408 and E432 represent the position at which DDX49 was truncated. (Pettersen et al., 2004)

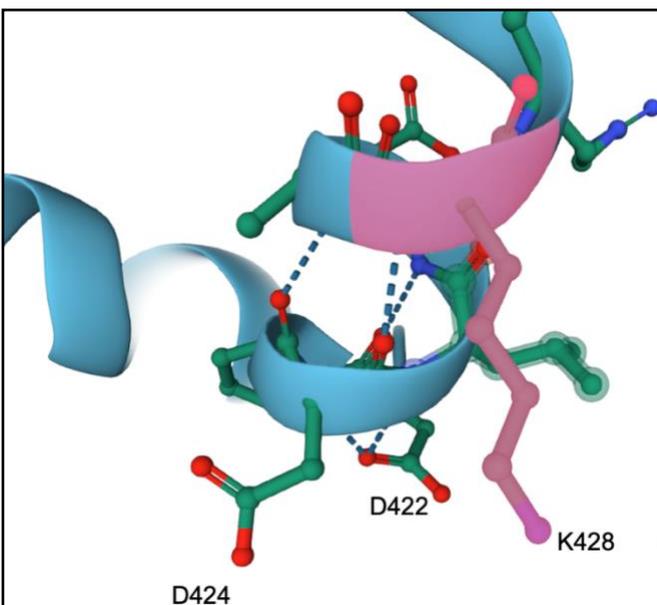


Figure 4.1.4. 3D structure of DDX49 as predicted by Alpha-Fold, highlighting residue K428. (Jumper et al., 2021)

FRET assays by Louise Martin using ATP- γ -s, identified that DDX49^{K421A} helicase activity was unaffected by the presence of non-hydrolysable ATP- γ -s, unlike wild-type DDX49. This implies that DDX49^{K421A} no longer requires ATP hydrolysis to dissociate from the substrate, as proposed by the model of local strand separation in figure 1.2. DDX49^{K421A} also exhibits reduced DNA binding as seen in figure 3.6.1, which signifies that structural confirmation changes in DDX49^{K421A} which decouples the from substrate binding and ATP hydrolysis drive substrate unwinding of DDX49^{K421A}, resulting in reduced DNA binding affinity.

DDX49^{K421A} hyperactive helicase activity could also be due to co-contaminant protein eluted during purification. These co-contaminants can be seen in figure 4.1.3. To investigate this, the protein could be purified using different chromatography columns, such as Phenyl Sepharose column which relies on the hydrophobic interaction between the protein and the Sepharose beads of the column. (Cytiva, 2020) This could then be followed by a Heparin column, to reduce elution of co-contaminants.

In order to investigate the biological significance of the D-x-D motif, in-cellulo gene editing of U2OS cells (osteosarcoma). CRISPR-Cas 9 mediated genome editing was used to substitute the D422/D424 into alanine residues by transfecting sgRNA, pegRNA and the CRISPR-cas9 encoding plasmids. Formation of the plasmids was achieved via a golden gate assembly reaction whereby multiple inserts are incorporated into a vector backbone. The editing of the U2OS cells resulted in no cell viability as seen in figure 3.7.2, whereas the control samples-maintained confluency, suggesting that their cell death observed in the edited cells, is unlikely due to co-contamination of the plasmids used for transfection.

In-vitro analysis of DDX49^{D422A/D424A} demonstrated the significant reduction in DDX49 unwinding, DNA binding and loss of nuclease activity; this loss of function could be reflected in the edited cells, and disrupt the fundamental biological functions proposed by Awasthi et al, such as DDX49 association with the 47S rDNA locus in the nucleus, mRNA export, regulation of ribosomal pre-mRNA transcription. The cumulative effect of negatively impacting these functions could have led to U2OS cell death, suggesting that the D-x-D motif plays an essential role in the catalytic activities of DDX49.

5.3 DDX49 intrinsically disordered C-terminus is essential for protein function

Following DDX49 sequence analysis using intrinsically disordered region (IDR) software predictors IUPred3 and IUPred3, we identified that the C-terminus of DDX49 is likely to be intrinsically disordered, due to the high probability scores seen in figure 4.1.1 and 4.1.2. The region where the IDR begins in the C terminus are however not in agreement between the 2 software predictors. Thus, we decided to truncate the C terminus of DDX49, starting at residue 432 and 408 until residue 483 (last protein amino acid), via site directed mutagenesis. This was followed by overexpression and purification as described in section 3.4, resulting in the generation of DDX49^{ΔK408-V483} and DDX49^{ΔE432-V483}. Both truncated proteins showed no Fork2 binding or unwinding and no nuclease activity, suggesting a total loss of DDX49 function.

Loss of function been seen in both mutants, suggests that the C-terminus intrinsically disordered region of DDX49 potentially begins at residue E432, as retaining the region between K408-E432 therefore including the presence of the D-x-D motif, does not lead to DDX49 regaining its catalytic activities.

These findings highlight the role that the C-terminus of DDX49 might play in order to participate in catalytic and cellular functions.

DDX49^{D422A/D424A} exhibits significant reduction in DNA binding affinity and unwinding activity, with loss of nuclease activity. The D-x-D mutagenesis therefore might only

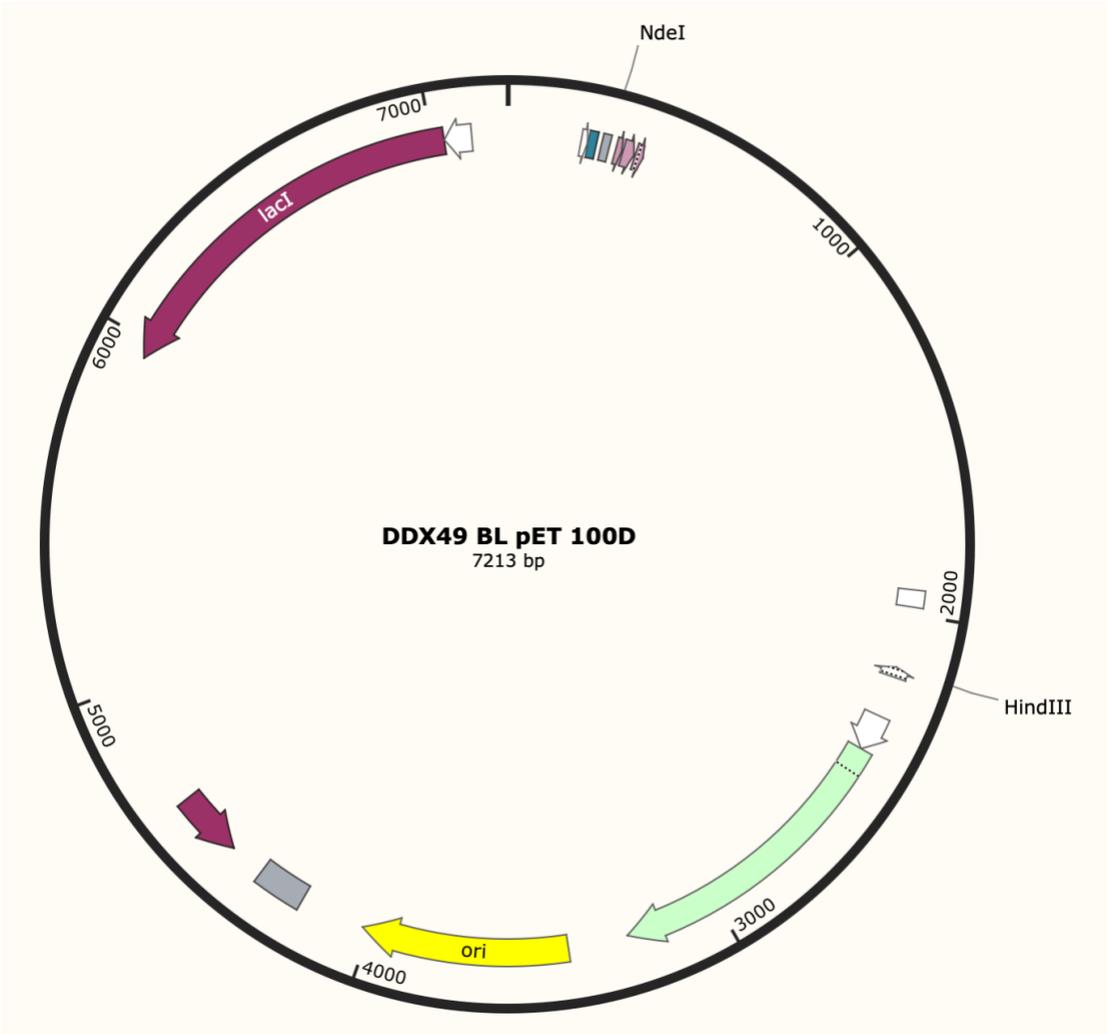
partially disrupt the function of a larger structure, comprising the C-terminus domain of DDX49, which acts a second nucleotide recognition or binding domain alongside the DEAD-box helicase core. Mutating the DEAD-box motif of the helicase core and analysing DNA binding and unwinding would allow investigation of the DNA binding and unwinding ability provided by the C terminus motif of the protein only. DDX49 binding to Fork2 was previously shown by master's student Sabesan Anandavijayan, which suggests that the DEAD-domain and the C-terminus of DDx49 coordinate to allow DNA binding and unwinding, with the nuclease activity being specifically linked to the D-x-D motif and intrinsically disordered C-terminus of DDX49.

Acknowledgments

I would like to thank my supervisor Dr Ed Bolt, for his unwavering support through my studies despite all obstacles encountered. I would also like to thank postdoctoral fellow Ashley Parkes for his guidance, contribution and training and master's students Sabesan Anandavijayan and Louise Martin for their key contributions to the data presented in this project.

Chapter 6. Appendix

pET.100DDX40 plasmid produced using SnapGene.



BLAST sequences used for sequence alignment.

Homo Sapiens

MAGFAELGLSSWLVEQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSDEPYGIFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALELSRK
PHVVIATPGRADHLRSSNTFSIKKIRFLVMDEADRLEQGCTDFTVDLEAILAAVPARR
QTLFSATLTDTLRELQGLATNQPFWEAQAPVSTVEQLDQRYLLVPEKVKDAYLVHLLIQ
RFQDEHEDWSIIIFTNTCKTCQILCMMLRKFSFPTVALHSMMKQKERFAALAKFKSSIYR
ILIATDVASRGLDIPTVQVVINHNTPLPKIYIHRVGR TARAGRQGGQAITLVTQYDIHLV
HAIEEQIKKKLEEFVVEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEG
KDPDLEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRAGHKGRPPRTPSGSHSGPVPSQ
GLV

Pan Troglodytes

MAGFAELGLSSWLVEQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSDEPYGIFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALELSRK
PHVVIATPGRADHLRSSNTFSIKKIRFLVMDEADRLEQGCTDFTVDLEAILAAVPARR
QTLFSATLTDTLRELQGLATNQPFWEAQAPVSTVEQLDQRYLLVPEKVKDAYLVHLLIQ
RFQDEHEDWSIIIFTNTCKTCQILCMMLRKFSFPTVALHSMMKQKERFAALAKFKSSIYR
ILIATDVASRGLDIPTVQVVINHNTPLPKIYIHRVGR TARAGRQGGQAITLVTQYDIHLV
HAIEEQIKKKLEEFVVEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEG
KDPDLEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRAGHKGRPPRTPSGSHSGPVPSQ
GLV

Macaca Mulatta

MSGFAELGLSSWLVAQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSDEPYGIFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALELSRK
PHVVIATPGRADHLRSSNTFSIKKIRFLVMDEADRLEQGCTDFTVDLEAILAAVPARR
QTLFSATLTDTLRELQGLATNQPFWEAQAPVSTVEQLDQRYLLVPEKVKDAYLVHLLIQ
RFQDEHEDWSIIIFTNTCKTCQILCMMLRKFSFPTVALHSMMKQKERFAALAKFKSSVYR
ILIATDVASRGLDIPTVQVVINHNTPLPKIYIHRVGR TARAGRQGGQAITLVTQYDIHLV
HAIEEQIKKKLEEFVVEAEVLQILTQVNVVRRRECEIKLEAAHFDEKKEINKRKQLILEG
KDPDLEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRVGHKGHPPRAPPGSHSGPVPSQ
GPA

Canis Lupus

MSSACAGKQSGHARPELAGRMAGFAELGLSSWLVEQCRQLGLKQPTPVQLGCIPAILEG
RDCLGCAKTGSGKTAAFVLPILQKLSDEPYGIFCLVLTPTRELAYQIAEQFRVLGKPLGL
KDCIIVGGMDMVAQALELSRKPHVVIATPGRADHLRSSNTFSIKKIRFLVMDEADRLE
QGCTDFTVDLEAILAAVPTRRQTLFSATLTDTLKELQGLATNQPFWEAQAPVRTVEQL
DQRYLLVPEKVKDAYLVHLLIQNFQDEHEDWSIIIFTNTCKTCQILCMMLRKFNPTVALH
SMMKQKERFAALAKFKSSIYRILIATDVASRGLDIPTVQVVINHNTPLPKIYIHRVGR
ARAEKQLEELVVEAKVLQILTQVNVVRRRECEIKLEAANFDEKKEINKRKQLILEGKDPD
LEAKRKAELAKIKQKNRRFKEKVEEQALQRQKASRAGHGGRPPGGRPPRAPPEAHSAPAPIHGQP

Bos Taurus

AGFAELGLSSWLVEQCRQMGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSSEDPYGFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMATNQPFWEAQA
PVRTVEQLDQRYLLVPEKVKDAYLVHLIQNFQDEHEDWSIIIFTNTCKTCQILCMMLRKF
NFPTVALHSMKQKERFAALAKFKSSYRILIATDVASRGLDIPTVQVVIHNTPLPKI
YIHRVGRTARAGRQQAITLVTQYDIHLVHAIEEQIKKKLEFPVEEAQVLQILTQVNVV
RRECEIKLEAANFDEKKEINKRKQMILEGKDPDLEAKRKAELAKIKQKNRRFKEKVEQTL
QRQKASRTDRRGRPPRARPEASLSLAPAQGA

Mus Musculus

MAGFAEIGLSSWLVEQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSSEDPYGFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALELSRK
PHVVIATPGRADHLRSSNTFNMKKIQFLVMDEADRLEQGCTDFTDLETILAAMPARR
QTLFSATLTDTLKELQGLATNEPFFWEAQATVRTVEQLDQRYLLVPEKVKDAYLVHLVQ
TFQDQLEDCSIIFTNTCKTCQILCMMLRKFNFPTVALHSMKQKERFAALAKFKSSYR
ILIATDVASRGLDIPTVQVVIHNTPLPKIYIHRVGRTARAGRQQAITLVTQYDIHL
HAIEEQIKQQLAELVVEAEVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEG
KDPDLEAKRKAELAKIKQKNRRFKEKVGQTLRRQKAGSTVRRSRPPRSRPQEPAAQEAQD

Rattus Norvegicus

MAGFAELRLSSWLVEQCRQLGLKQPTPVQLGCIPAILEGRDCLGCAKTGSGKTAAFVLP
LQKLSSEDPYGFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVTQALELSRK
PHVVIATPGRADHLRSSNTFNMKKIRFLVLEADRLEQGCTDFTADLETILSAVPARR
QTLFSATLTDTLKELQGLATNQPFWEAQATVRMVEQLDQRYLLVPEKVKDAYLVHLVQ
TFQDQLEDCSIIFTNTCKTCQILCMMLRKFNFPTVALHSMKQKERFAALARFKSSTYR
ILIATDVASRGLDIPTVQVVIHNTPLPKIYIHRVGRTARAGRQQAITLVTQYDIHL
HAIEEIKQQLAELAVEEAQVLQILTQVNVVRRRECEIKLEASHFDEKKEINKRKQMILEG
KDPDLEAKRKAELAKIKQKNRRFKEKVEQTLRQQKAGSTGRRSRPPRPRPQEPARAEQAN

Sphenodon Punctatus

MAGFRELGLAPWLVAQAEQLGLSRPTPVQAACIPPTLQGRDCMGCATGSGKTAAFVLP
LQKLSSEDPYGFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALDLSRK
PHVVIATPGRADHLRSSNTFSLKKIKFLVLEADRLEQGCTDFTKDLEVILAAMPAAAR
QTLFSATLTDTLNELKGLAMNKPFFWESQSEVRTVEQLDQRYLLVPERVKDAYLVHLIQ
TFQDQHEDWSIIIFTNTCKNCQILNMMMKRNFPSVALHSMLKQKQRFALAKFKSSVFK
ILIATDVASRGLDIPTVQVVIHNTPLPKIYIHRVGRTARAGRHGIAITLVTQYDIHLV
HAIEEQIKMKLQEFMEERDVLGILTQVNVVRRRECEIKLEATDFDEKKEINKRKQMILEG
KDPDLEEKRKTELARIKKNNFRARVQQTLQEKQELQMRRKLRKRVRQQKAAGKES

Danio Rerio

MATFESLGLSEWLIQQCKQMGISRPTAVQEKCIPAILDGRDCMGC AKTGSGKTA AFVLPV
LQKLS EDPYGVFCLVLTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVTQGLELSK
PHVVVATPGRLADHIRSSDTINLNRIQFLIMDEADRLL EQGCTDFTKDLEVILSAVPAKR
QTLLFSATLTDTLQQLQSIAMNRPFFWEHKSDVQTVEELDQRFILTPEKV KDAYLVH LIQ
TFQDEHDDWSIIIFTNTCKSCQILTMMLREFNFPTISLHSM MKQRQRFANLAKFKSNVFK
ILIATDVAARGLDIPTVQVVINHNTPLPKIYIHRVGRTARAGRNGVSITLVTQYDIHLI
NAIEEQNQTKLKEFPIEEKEVLKILTQVNVTRRQCEIKLESTDFDEKKKINKR KQMILDG
KDPDLEEKRKNELEKIRGKNKLLKGNPQSKKEKKKHEKSAKATEVGS

Drosophila Melanogaster

MQRKEANPFQILGLRPWLVKQLTKLGLKGATPIQQKCIPAILAGQDCIGAAKTGSGKTFA
FALPILERLSEEPVSHFALVLTPTHELAYQISEQFLVAGQAMGVRVCVVS GGTDQMVESQ
KLMQRPHIVVAMPGR LADHLTGCDTFSFDNLKYL VVDEADRMLNGDFDESLSIIERCLPK
TRQNLFFSATMKDFIKESSIFPIASDCFEWSQSDVATVETLDQRYLLCADYDRDMVLIE
ALRKYREENENANVMIFTNTKKYCQLLSMTLKNMEIDNVCLHGFMRQKERVAALS RFKSN
QIRTLIATDVAARGLDIPSVELVMNHMLPRTPK EYIHRVGRTARAGRKGMSISIFRFRD
LELLAAIEEEINTKLTEHPIDQRMVERIFMQVNVTRRESEMQLDNNDFD ERAQNYRRKTW
IMEGKDPDQMEALYRKKQKDKLREIRRKRLQHA EPAASEEGKALLQDERFKSVDSARFE
KKGKGRSRATQEDTPTKPLKRLNKEKPVAQKGRADVKKDKA

Podarcis Muralis

MLGVVLQTAGHCRGARRDAQHMGRRLSRCLVGLLPFLSLFPPQRRGAPAMSGFRALGLSP
WLVSQAEQLGLSRPAPVQEACIPPALQGRDCMGC AQTGSGKTA AFVLPILQKLS EDPYGI
FCLVLTPTRELAHQIAEQFRVLGKPLGLKDCIIVGGMDMVAQALELSRKPHVVIATPGRL
ADHLRSSNTFSLQKIKFLVLDEADRLL EQGCTDFTKDLEVILAAVPASRQTMLFSATLTD
TLNELKQIAMNKPFFWESKSEVRTVEQLDQRYLLVPEKV KDAYLVHLVQTFQDEHEDWSI
IIFTNTCKNCQILNMMLRRFNFP SVALHSM MKQKQRF AALAKFKSSVFKILIATDVAARG
LDIPTVQVVINHNTPLPKIYIHRVGRTARAGRNGIAITLVTQYDIHLVHAIEEQIKMKL
QEFVVEEHSVLNILTQVNVVRRECEIRLEATDFDEKKEINKR KQMILEGKDPDLEAQRKA
ELAKIRKKNVEFRSKVQQVLEDKKQVVLHRKLRKKRQRQRKQPAAEKITFKRIP

Microcaecilia Unicolor

MAGFAALGLSPWLREQCAQLGISRPTPVQESCIPVILNGQDCMGC AKTGSGKTA AFVLP
LQKLS EDPYGVIFCLVMTPTRELAYQIAEQFRVLGKPLGLKDCIIVGGMDMVTQALELSRK
PHIVVATPGRLADHIRSSDTFSIKKIKFLVLDEADRLL EQGCTDFTKDLEVILGAIPAQR
QTLLFSATLTDTLQQLKTIAMNKPFFWESSEVRTVDELDQRYILVPEKV KDAYLVH LIQ
KFQDEHEDWSIIIFTSTCKTCQILNMMLREFNFPSVALHSM MKQKQRFATLAKFKSSIFK
ILIATDVAARGLDIPTVQVVINHNTPLPKLYVHRVGRTARAGRHGIAITLVTQYDIHLV
HTIEQQIKIKLDFAVEESEVLKILTQVNVTRRECEIKLESTDFDEKREKNR KQLILDG
KDPDLEAKRKAEELEKIKRQKRKFKDQIQETLQRRQEVQLQRKREQKQQAILEKA

Anopheles Gambiae

MSLWTDKKFSDLGLTYWITRQTEKLGRRPTPIQVEIPRILQGQDCIGAAKTGSGKTF
FALPILQKLSEPTANFALVLTPTHELAHQIAEQFIVAGQPMNARVCVVTGGTDYLIESQ
QLEKRPHIIVAMPGRADHLTGCNTYSFAALQFLVVDEADRMLCGSFDEDLMVINRFLPA
KRQNLFFSATLKDFLKTSIVFPIADDDVFEWSEQSPVATVETLDQRYLLCADYDRDVTMVE
ALRKYKEDQEDASIMIFTNSKKDCQILSMLNSFGFSNVCLHGFLRQRERVAALNKFKSK
HVRIMIATDVASRGLDIHDVQLVVNHRLPKKPIEYIHRVGRTARAGRAGMAISILRFPRD
LEALGEIALINTKLTEYSVDDRLVQRIFMQVKVARAEAEINLDNKDFDERKHKYRRLRW
IQEGLDPDEMEAKWKEDMKAREQERRERLRQENEERRKRDQTIASPAVVNDARFQAAAS
DKKFKRKRKFIPTKLNELIEQRKTEAKGGKKARARGKGDKKIVKKKAKMAQ

Salarias Fasciatus

MGDFSSLGLSDWLINQCKQLGINKPTPVQENCMPPILEGRDCMGCAGTSGKTAAFVLPV
LQKLSEEPYGFCLVLTPTRSLAYQIAEQFRVLGKPLGLRDCIIVGGMDMVTQAMELSNQ
PHVVVATPGRLADHIRSSNTFSMSRIQFLILDEADRILLEQGCTDFTKDLEVILGILPAKR
QTLIFSATLTDTLQELKSAMNKPFFWESKSETRTVEELDQRYILTPETVKDAYLVHLIQ
KFTDEHDDWSIIIFTNTCKNCQVLTMMMLREFNFPTISLHSMKQKQRFANLAKFKASVYK
ILIATDVAARGLDIPTVQVVINQNTPLPKIYIHRVGRTARAGRNGVSITLVTQYDIHLV
HAIEEQTQTKLKEYPVQEKEVLKILTQVNVTRRECEIKLESTDFDEKKEINKRKQLILEG
KDPELEAKRKAELEKIRNQKKFKQRIQEDIERQKFGQKKKNFMKHHRSRSH

C.elegans

MSKVCDAILSDDDDLEMDDEEQEKPTSSKSVKNEEFEEFEDGEDVEDEEDEASDEDEE
SEGEEGDEFKSSDDTPKPIQISEDNMTTKKFSQLGVCSWITQQLQTMQIKTATPVQAACI
PKILEGSDILGCARTGTGKTLAFAIPILQKLSVDPYGIYALILPTRELAFAQIAEQFTAL
GKPITLKCSVIVGGRSLIHQARELSERPHVVVATPGRLADLIESDPDTIAKVFKKIQFFV
LDEADRMLLEGQYNDQLKPIFESISEKRQTLILLSATITNNINMLHRVSTRKPYFFEDKGD
DESTVDRLEQKYVVCVAVKDAYLVYVVKNYSEKNPKSSVMIFAQTCRECCALAYMFEGL
GFRVGLHSQIPQKQRLAALSFRSKTLQVICTDVASRGLDIPHVDLVVNHNVPCPKT
YIHRVGRSARAGRFGSALSFVTQYDVELLQAVEQVIGKKLDELKVPKHVTKYVTQVLVA
KKEAELKLENQKFGKKEINRRKELMSGMDEDEADRHLEEMRTRRMTNSKRKLEKISGQ
LDRDRFQKKLAAKKVKKEEN

6.1 References

Aladeokin, A.C., Akiyama, T., Kimura, A., Kimura, Y., Takahashi-Jitsuki, A., Nakamura, H., Makihara, H., Masukawa, D., Nakabayashi, J., Hirano, H., Nakamura, F., Saito, T., Saido, T. and Goshima, Y. (2019). Network-guided analysis of hippocampal proteome identifies novel proteins that colocalize with A β in a mice model of early-stage Alzheimer's disease. *Neurobiology of Disease*, 132, p.104603.

doi:<https://doi.org/10.1016/j.nbd.2019.104603>.

Apostolou, P., Toloudi, M. and Papisotiriou, I. (2015). Identification of genes involved in breast cancer and breast cancer stem cells. *Breast Cancer: Targets and Therapy*, p.183. doi:<https://doi.org/10.2147/bctt.s85202>.

Awasthi S., Verma M., Mahesh A., K. Khan M.I., Govindaraju G., Rajavelu A., Chavali P.L., Chavali S. and Dhayalan A. (2018). DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Research*, 46(12), pp.6304–6317. doi:<https://doi.org/10.1093/nar/gky231>.

M.J. Betts, R.B. Russell. Amino acid properties and consequences of substitutions. *Bioinformatics for Geneticists*, M.R. Barnes, I.C. Gray eds, Wiley, 2003. pp.289–316. doi:<https://doi.org/10.1002/0470867302.ch14>.

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L.J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A. and Insana, G. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*. doi:<https://doi.org/10.1093/nar/gkac1052>.

Bohnsack, K.E., Yi, S., Venus, S., Jankowsky, E. and Bohnsack, M.T. (2023). Cellular functions of eukaryotic RNA helicases and their links to human diseases. *Nature Reviews Molecular Cell Biology*, 24(10), pp.749–769. doi:<https://doi.org/10.1038/s41580-023-00628-5>.

Bolten, S.N., Rinas, U. and Scheper, T. (2018). Heparin: role in protein purification and substitution with animal-component free material. *Applied Microbiology and Biotechnology*.102(20), pp.8647–8660. doi:<https://doi.org/10.1007/s00253-018-9263-3>.

Bourgeois, C. F., Mortreux, F. & Auboeuf, D., 2016. The multiple functions of RNA helicases as drivers and regulators of gene expression. *National Review of Molecular Cellular Biology*, 17(7), pp. 426-438. doi:<https://doi.org/10.1038/nrm.2016.50>.

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D. and Klaus Aktories (1998). A Common Motif of Eukaryotic Glycosyltransferases Is Essential for the Enzyme Activity of Large Clostridial Cytotoxins. *Journal of Biological Chemistry*. 273(31), pp.19566–19572. doi:<https://doi.org/10.1074/jbc.273.31.19566>.

Cai, W., Chen, Z., Rane, G., Singh, S., Choo, Z., Wang, C., Yuan, Y., Tuan Zea Tan, Newsholme, P., Yap, C.T., Pongor, L.S., Yang, H., Lee, M., Boon Cher Goh, Sethi, G., Touati Benoukraf, Vinay Tergaonkar and Alan Prem Kumar (2017). Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. 109(6) doi:<https://doi.org/10.1093/jnci/djw278>.

Chen, Y., Ma, Q. and Sun, N. (2023). DDX49 Promotes Proliferation and Metastasis of Cervical Cancer by Regulation of AKT and Wnt/ β -Catenin Signalings. *Annals of Clinical and Laboratory Science*. 53(2), pp.271–277. doi:<https://pubmed.ncbi.nlm.nih.gov/37094865/>.

Collins, R., Karlberg, T., Lari Lehtiö, Schutz, P., Susanne, Dahlgren, L.G., Hammarström, M., Weigelt, J. and Herwig Schüler (2009). The DEXD/H-box RNA Helicase DDX19 Is Regulated by an α -Helical Switch. 284(16), pp.10296–10300. doi:<https://doi.org/10.1074/jbc.c900018200>.

Cytiva, 2020. *Phenyl Sepharose High Performance Hydrophobic Interaction Chromatography*. Available

at:<https://www.cytivalifesciences.com/en/us/shop/chromatography/resins/hydrophobic-interaction/phenyl-sepharose-high-performance-p-06293>

Dai, H., Feng, J., Nan, Z., Wei, L., Lin, F., Jin, R., Zhang, S., Wang, X. and Pan, L. (2021). Morphine may act via DDX49 to inhibit hepatocellular carcinoma cell growth. *Aging (Albany NY)*, 13(9), pp.12766–12779. doi:<https://doi.org/10.18632/aging.202946>.

Dandage, R. and Landry, C.R. (2019). Paralog dependency indirectly affects the robustness of human cells. *Molecular Systems Biology*, 15(9). doi:<https://doi.org/10.15252/msb.20198871>.

Donsbach, P. and Klostermeier, D. (2021) Regulation of RNA helicase activity: principles and examples. *Biological Chemistry*, Vol. 402 (Issue 5), pp. 529-559. <https://doi.org/10.1515/hsz-2020-0362>

Durfee, T., Nelson, R., Baldwin, S., Plunkett, G., Burland, V., Mau, B., Petrosino, J.F., Qin, X., Muzny, D.M., Ayele, M., Gibbs, R.A., Csorgo, B., Posfai, G., Weinstock, G.M. and Blattner, F.R. (2008). The Complete Genome Sequence of *Escherichia coli* DH10B: Insights into the Biology of a Laboratory Workhorse. *Journal of Bacteriology*. 190(7), pp.2597–2606. doi:<https://doi.org/10.1128/jb.01695-07>.

Erdős, G., Pajkos, M. and Dosztányi, Z. (2021). IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Research*, 49(W1), pp.W297–W303. doi:<https://doi.org/10.1093/nar/gkab408>.

Hu, G., Katuwawala, A., Wang, K., Wu, Z., Ghadermarzi, S., Gao, J. and Kurgan, L. (2021). fIDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nature Communications*,12(1), p.4438.
doi:<https://doi.org/10.1038/s41467-021-24773-7>.

Fuller-Pace, F.V. (2013). DEAD box RNA helicase functions in cancer. *RNA Biology*, 10(1), pp.121–132. doi:<https://doi.org/10.4161/rna.23312>.

Gilman B., Tijerina P. and Russell R. (2017). Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochemical Society Transactions*.45(6), pp.1313–1321.
doi:<https://doi.org/10.1042/bst20170095>.

Jumper, J. et al. “Highly accurate protein structure prediction with AlphaFold.” *Nature* (2021). (596), pp. 583–589 Doi:10.1038/s41586-021-03819-2.

Lian, X., Xiang, D., Peng, C., Chen, J., Liao, M., Sun, G. and Zhang, Z. (2020). DDX49 is a novel biomarker and therapeutic target for lung cancer metastases. *Journal of Cellular and Molecular Medicine*. 24(1), pp.1141–1145.
doi:<https://doi.org/10.1111/jcmm.14734>.

Linder, P. and Jankowsky, E. (2011). From unwinding to clamping — the DEAD box RNA helicase family. *Nature Reviews Molecular Cell Biology*, 12(8), pp.505–516.
doi:<https://doi.org/10.1038/nrm3154>.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P. and Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(W1), pp.W597–W600. doi:<https://doi.org/10.1093/nar/gkt376>.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin T.E. 92004. "UCSF Chimera - A Visualization System for Exploratory

Research and Analysis." *J. Comput. Chem* (2004). (13):1605-1612. Available at:
<https://www.rbvi.ucsf.edu/chimera/>

Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, 5(172).
<https://doi.org/10.3389/fmicb.2014.00172>.

Saito, M., Hess, D., Eglinger, J., Fritsch, A.W., Kreysing, M., Weinert, B.T., Choudhary, C. and Matthias, P. (2018). Acetylation of intrinsically disordered regions regulates phase separation. *Nature Chemical Biology*, 15(1), pp.51–61.
[doi:https://doi.org/10.1038/s41589-018-0180-7](https://doi.org/10.1038/s41589-018-0180-7).

Serfecz, J.C., Hong, Y., Gay, L.A., Shekhar, R., Turner, P.C. and Renne, R. (2022). DExD/H Box Helicases DDX24 and DDX49 Inhibit Reactivation of Kaposi's Sarcoma Associated Herpesvirus by Interacting with Viral mRNAs. *Viruses*. 4(10), p.2083.
[doi:https://doi.org/10.3390/v14102083](https://doi.org/10.3390/v14102083).

Scholefield J. and Harrison P.T. (2021). Prime editing – an update on the field. *Gene Therapy*. 28(7), pp.396–401. [doi:https://doi.org/10.1038/s41434-021-00263-9](https://doi.org/10.1038/s41434-021-00263-9).

Schrödinger, L. & DeLano, W., (2020). *PyMOL*, [doi:http://www.pymol.org/pymol](http://www.pymol.org/pymol).

Stoddart, M.J. (2011). Cell viability assays: introduction. *Methods in Molecular Biology (Clifton, N.J.)*, 740, pp.1–6. [doi:https://doi.org/10.1007/978-1-61779-108-6_1](https://doi.org/10.1007/978-1-61779-108-6_1).

Tao, J., Ge, Q., Meng, J., Liang, C., Hao, Z. and Zhou, J. (2023). Overexpression of DDX49 in prostate cancer is associated with poor prognosis. *BMC Urology*, 23(1).
[doi:https://doi.org/10.1186/s12894-023-01251-4](https://doi.org/10.1186/s12894-023-01251-4).

Toru Sengoku, Osamu Nureki, Nakamura, A., Kobayashi, S. and Yokoyama, S. (2006). Structural Basis for RNA Unwinding by the DEAD-Box Protein *Drosophila* Vasa. 125(2), pp.287–300. [doi:https://doi.org/10.1016/j.cell.2006.01.054](https://doi.org/10.1016/j.cell.2006.01.054).

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhor, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., Sanli, K., von Feilitzen, K., Oksvold, P., Lundberg, E., Hober, S., Nilsson, P., Mattsson, J., Schwenk, J.M., Brunnström, H. and Glimelius, B. (2017). A pathology atlas of the human cancer transcriptome. *Science*, 357(6352). doi:<https://doi.org/10.1126/science.aan2507>.

Wang, X., Gu, G., Zhu, H., Lu, S., Abuduwalli, K. and Liu, C. (2020). LncRNA SNHG20 promoted proliferation, invasion and inhibited cell apoptosis of lung adenocarcinoma via sponging miR-342 and upregulating DDX49. *Thoracic Cancer*, 11(12), pp.3510–3520. doi:<https://doi.org/10.1111/1759-7714.13693>.

Weis, K. and Hondele, M. (2022). The Role of DEAD-Box ATPases in Gene Expression and the Regulation of RNA–Protein Condensates. *Annual Review of Biochemistry*, 91(1), pp.197–219. doi:<https://doi.org/10.1146/annurev-biochem-032620-105429>.

Xue, C., Gu, X., Li, G., Bao, Z. and Li, L. (2021). Expression and Functional Roles of Eukaryotic Initiation Factor 4A Family Proteins in Human Cancers. *Frontiers in Cell and Developmental Biology*, 9. doi:<https://doi.org/10.3389/fcell.2021.711965>

Yadav, M., Singh, R.S., Hogan, D., Vidhyasagar, V., Yang, S., Chung, I.Y.W., Kusalik, A., Dmitriev, O.Y., Cygler, M. and Wu, Y. (2021). The KH domain facilitates the substrate specificity and unwinding processivity of DDX43 helicase. *The Journal of Biological Chemistry*, 296, p.100085. doi:<https://doi.org/10.1074/jbc.RA120.015824>.

Yang, W. (2010). Nucleases: diversity of structure, function and mechanism. *Quarterly Reviews of Biophysics*, 44(01), pp.1–93. doi:<https://doi.org/10.1017/s0033583510000181>