

Molecular biology of the DExD-box helicases DDX49 and DDX52

Ashley J. Parkes, BSc (Hons), MSc



**University of
Nottingham**
UK | CHINA | MALAYSIA

**Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy**

School of Life Sciences

University of Nottingham

September 2023

Acknowledgements

Initially, I would like to thank both the BBSRC and The University of Nottingham for funding and providing the opportunity to carry out this project. I would equally like to thank Professor Ed Bolt for his supervision and guidance throughout this research project, and for being understanding when things got a bit crazy in the middle there. Additional thanks to my second supervisor Professor Thorsten Allers, whose guidance on his rotation helped establish a working knowledge of genetics needed for my later projects.

I would also like to thank past and present members of the Bolt lab for their help and assistance throughout this project, especially for being so accommodating in the initial stages of returning from the COVID lockdown with split shift work. Particular thanks go to Dr Andrew Cubbon for his prior work in establishing human cell culture in the lab, to Dr Ryan Buckley and Dr Tom Killilea for their guidance on the cloning and microbiology side of the project (and the occasional Guinness after a long day!), and to the talented undergraduates and MRes students who I have had the opportunity to share DDX49 and DDX52 with: Sabesan Anandavijayan, Fiorela Kapllanaj, Louise Martin and Phillip Springer. Additional thanks go to Deborah Briggs and Alex Rathbone, who helped greatly with cell culture and RT-qPCR methods, respectively.

Finally, I would like to thank all my family, friends, colleagues, and the occasional therapist for being there alongside me through this journey. I'd always anticipated a PhD would be a difficult task, but I didn't realise the biggest challenge would come from everything that happened outside of the PhD! So thank you for being there through all the pandemics, the break-ups and the self-doubts. I would have crashed and burned long ago without all of you. With that, a slice of silly inspirational poetry to bring us home.

*In the oncoming of night, light ever dimming,
When there's no shore in sight, keep swimming.*

Table of Contents

Acknowledgements.....	2
List of abbreviations.....	11
Abstract.....	15
List of figures.....	16
List of tables.....	21
Chapter 1: Introduction.....	23
1.1 DNA and RNA.....	23
1.1.1 The Central Dogma of Molecular Biology.....	24
1.1.2. Functional RNAs.....	25
1.2 ATP-dependent nucleic acid processing helicases, translocases and remodelling enzymes.....	26
1.2.1 RNA helicases.....	26
1.2.2 Superfamily 2 helicases.....	27
1.2.3. DExD/DExH-box helicases.....	27
1.2.4. DExH-box helicases.....	28
1.2.5. DExD-box helicases.....	29
1.3 Ribosome Biogenesis.....	35
1.3.1. A brief overview of ribosome assembly.....	36
1.3.2. Ribosome biogenesis associations with animal disease.....	38
1.3.3. DExD-boxes in ribosome biogenesis.....	40
1.4. DNA replication.....	42
1.4.1. Initiation of DNA replication at origins of replication.....	42
1.4.2. Elongation of replication.....	43
1.4.3. Termination.....	44
1.5. DNA repair and recombination.....	46
1.5.1 Homology-Directed repair.....	46

Molecular biology of the DExD-box helicases DDX49 and DDX52

1.5.2. DNA recombination	53
1.5.3. Non-homologous end-joining	55
1.6. Asgard Archaea	57
1.6.1. MCP8718128.1	59
1.7. Gene editing	59
1.7.1. CRISPR-Cas9	60
1.7.2. Transposons, Base editors, and prime editors.....	62
1.8 Aims of the PhD project.....	63
Chapter 2: Materials and Methods.....	65
2.1. Materials	65
2.1. Chemicals and reagents.....	65
2.2. General methodologies concerning DNA and RNA.....	86
2.2.1. Materials, services and nucleic acid purification	86
2.2.2. Agarose gel electrophoresis.....	86
2.2.3. Polymerase Chain reaction	87
2.2.4. Site-directed mutagenesis.....	87
2.2.5. RT-qPCR.....	88
2.2.6. Annealing and purification of nucleic acids	89
2.3. General microbiology protocols.....	90
2.3.1. Growth and storage of <i>E. coli</i> strains	90
2.3.2. Preparation of chemically competent <i>E. coli</i>	90
2.3.3. Transformation of chemically competent <i>E. coli</i>	90
2.4. Cloning and protein purification	91
2.4.1. Cloning of AA.49, DDX49 and DDX52	91
2.4.2. General protocol for protein over-expression within <i>E. coli</i>	93
2.4.3. General protocol for protein purification	93
2.4.4. Streptococcus pyogenes recombinant NLS-Cas9	94

2.5. General assays.....	95
2.5.1. SDS-polyacrylamide-gel electrophoresis (SDS-PAGE)	95
2.5.2. Bradford Assay	95
2.5.3. Bicinchoninic acid assay	95
2.5.4. Mass Spectrometry analysis	96
2.6. Biochemical assays	96
2.6.1. Gel-based unwinding assays	96
2.6.2. FRET-based unwinding assays.....	96
2.6.3. Gel-based annealing assays	97
2.6.4. FRET-based annealing assays	97
2.6.5. Electromobility shift assays	98
2.6.6. ATPase assays	99
2.6.7. Nuclease protection assays	99
2.6.8. Flourescence polarisation	99
2.7. Human cell culture	100
2.7.1. Human cell lines, media and supplements.....	100
2.7.2. Routine growth and passaging of cells	101
2.7.3. Lysis of cultured cells	101
2.8. General cell culture assays	102
2.8.1. WST-1 viability assays	102
2.8.2. ATPlite assay	102
2.8.3. Western blot	103
2.8.4. Transfection protocols	104
2.8.5. CRISPR-Cas9 genetic editing of human cell lines	105
2.8.6. CRISPR-Cas9 Prime editing of human cell lines	106
2.8.6.2. Transfection.....	108
2.8.6.3. Analysis	108

2.9. Ribosome biogenesis methods	108
2.9.1. Northern blotting.....	108
Chapter 3: Nucleic acid interactions of Probable ATP-dependent helicase DDX52	110
3.1. Introduction.....	110
3.1.1. Pre-established roles of DDX52.....	110
3.1.2. A role for DDX52 in DNA repair and replication?	110
3.1.3. Aims and objectives	110
3.2. Results.....	111
3.2.1. Bioinformatics.....	111
3.2.2. Purification of recombinant wild type and mutant DDX52	115
3.2.3. Initial optimisation of recombinant DDX52	117
3.2.4. Interrogating DDX52 binding.....	120
3.2.5. The unwinding of DDX52	124
3.2.6. Investigating DDX52 N-terminus annealase activity	130
3.2.7. DDX52 binding protects DNA from nucleases	134
3.3. Discussion	137
3.3.1. DDX52 possesses distinct structural differences compared to Lhr and Rok1	137
3.3.2. Binding characteristics of DDX52.....	138
3.3.3. DDX52 acts as both a helicase and annealase	140
3.4. Conclusions	144
3.5. Future Research.....	145
Chapter 4: Investigating DDX52 within tumour cells.....	146
4.1. Introduction.....	146
4.1.1. DDX52 within the cell.....	146
4.1.2. Phenotypes of DDX52.....	146
4.1.3. Aims and objectives of research	147

4.2. Results.....	147
4.2.1. Development and optimisation of a gene-editing pipeline for human cells	147
4.2.2. CRISPR editing of <i>DDX52</i>	154
4.2.2.1 Testing of gRNAs	154
4.2.2.2. Full gene editing run and establishment of a monoclonal cell line	156
4.2.2.3. On-target testing.....	157
4.2.2.4. Off-target testing.....	158
4.2.3 Phenotyping <i>DDX52</i> heterozygotes	160
4.2.3.1. DDX52 concentration appears unaffected within western blots	160
4.2.3.2. <i>DDX52</i> ^{+/-} cells phenotypes include impeded cellular migration and proliferation.....	161
4.2.3.3. <i>DDX52</i> ^{+/-} cells show no clear phenotype in response to cytotoxic agents.....	163
4.2.4. Trialling workflows for investigations of ribosome biogenesis	164
4.2.4.1. Northern blotting.....	164
4.2.4.2. RT-qPCR.....	165
4.3. Discussion	167
4.3.1. CRISPR-Cas9 workflow development.....	167
4.3.2. Guide RNA testing.....	170
4.3.3. The successful generation of <i>DDX52</i> heterozygotes	170
4.3.4. <i>DDX52</i> ^{+/-} cells present impacted cell migration and proliferation .	173
4.3.5. <i>DDX52</i> ^{+/-} cell phenotypes to DNA damaging agents.....	174
4.3.6. <i>DDX52</i> ^{+/-} cells possess lower levels of 47S and 28S rRNA	174
4.4. Conclusion.....	175
4.5. Future work.....	175

Chapter 5: Biochemistry and cell models of probable ATP-dependent helicase DDX49 177

5.1 Introduction 177

5.1.1 The role of DDX49..... 177

5.1.2. Functions of DDX49 within RNA processing 177

5.1.3. DDX49 roles within viral immunity 177

5.1.4. DDX49 within disease 178

5.1.5. Aims and objectives 179

5.2. Results..... 179

5.2.1. Bioinformatics..... 179

5.2.2. Purification of recombinant hDDX49 183

5.2.3. Mass Spectrometry of DDX49 contaminant 185

5.2.4. Assessing activity of wild type and mutant DDX49 187

5.2.5. CRISPR-Cas9 mediated gene editing of *DDX49* 194

5.2.6. CRISPR-Cas9 prime editing of U2OS cell lines 200

5.3. Discussion 203

5.3.1. DDX49 has a characteristic DExD-box domain with a unique C-terminus extension 203

5.3.2. *E. coli* Crp was consistently copurified with DDX49 204

5.3.3. DDX49 C-terminus forms binding complexes with DNA 204

5.3.4. DDX49 is an ATP-dependent 3' to 5' helicase 205

5.3.5. Characterisation of a novel nuclease site..... 206

5.3.6. *DDX49*^{+/-} cells were successfully generated and showed a phenotype of inhibited migration 207

5.3.7. The abundance of several rRNA species is altered in *DDX49*^{+/-} cells 208

5.3.8. CRISPR-Cas9 prime editing of *DDX49*..... 209

5.4. Conclusions 209

5.5. Future work.....	210
Chapter 6: Investigating putative Asgard origins of probable ATP-dependent helicase DDX49	212
6.1 Introduction.....	212
6.1.1. MCP8718128.1 (AA.49)	212
6.1.2. Yeast Dbp8.....	213
6.1.3. Aims and objectives	213
6.2 Results.....	214
6.2.1. Bioinformatics.....	214
6.2.2. Purification of recombinant AA.49	218
6.2.3. Assessing the activity of wild type AA49	218
6.2.3.3. AA.49 is an ATP-independent helicase.....	222
6.3. Discussion	223
6.3.1. Blast searches of DDX49 identified a potential Archaea homolog	223
6.3.2. Bioinformatics suggests AA.49 shares closer similarities to <i>S. cerevisiae</i> Dbp8 than hDDX49	223
6.3.3. AA49 only partially shares the same biochemical activity as DDX49	223
6.4. Conclusion and alternative hypothesis	225
6.5. Future perspectives	226
Chapter 7: Conclusions and future perspectives	227
7.1. Summary of research	227
7.1.1. DDX52 is a 3' to 5' ATP-dependent helicase with novel annealase activity	228
7.1.2. Potential roles for DDX52 in cancer progression and DNA replication/repair.....	229
7.1.3. Abundance of rRNA in <i>DDX52</i> ^{+/-} cells suggests yeast predictions of DDX52 function are inappropriate.....	229

Molecular biology of the DExD-box helicases DDX49 and DDX52

7.1.3. DDX49 is a 3' to 5' ATP-dependent helicase with novel ATP-independent nuclease activity	230
7.1.4. Roles of DDX49 within cancer and ribosome biogenesis	231
7.1.5. Links between DDX49 and a putative Asgard homolog are redundant	232
7.2. Thesis summary	233
Reference list	234
PIP reflective statement	270
Outreach and Public Engagement Reflection Form.....	277

List of abbreviations

A	Adenine
ADP	Adenosine diphosphate
APS	Ammonium persulphate
ATM	Ataxia-telangiectasia mutated
ATP	Adenine Triphosphate
ATR	Ataxia telangiectasia and Rad3-related
BIR	Break induced replication
BLM	Bloom syndrome protein
bp	Base pair
BRCA	Breast cancer gene
BSA	Bovine Serum Albumin
C	Cytosine
Cas	CRISPR associate protein
C-DDX52	C-terminus fragment of DDX52
CDK	Cyclin dependent kinases
CMG	Cdc45-MCM-GINS
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats
CRISPRi	CRISPR-interference
crRNA	CRISPR RNA
DBCO	Dibenzocyclooctyne
Dbp	Dead box protein
DDX	DEAD-box helicase
DExD	Aspartic Acid – Glutamic Acid – (x) - Aspartic acid
DExH	Aspartic Acid – Glutamic Acid – (x) - Histidine
dHJ	Double Holliday Junction
D-loop	Displacement loop
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
dsDNA	Double-stranded DNA

dsRNA	Double-stranded RNA
DSB	Double strand break
DSBR	Double strand break repair
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
eIF4AIII	Eukaryotic initiation factor 4A-III
EMSA	Electrophoretic mobility shift assay
ESP	Eukaryotic signature proteins
ETS	External transcribed spacer
FBS	Foetal bovine serum
G	Guanine
GIN5	Go-ichi-ni-san protein
gRNA	Guide RNA
HDR	Homology-directed repair
HJ	Holliday junction
HR	Homologous recombination
ICL	Inter-strand crosslink
IDR	Intrinsically disordered region
ITS	Internal transcribed spacer
InDel	Insertion/Deletion
IPTG	Isopropyl b-D-1thiogalactopyranoside
KO	Knock out
LB	Luria Bertani
LSU	Large sub-unit
MCM	Mini-chromosome maintenance
MGE	Mobile genetic element
MMC	Mitomycin C
MMEJ	Microhomology mediated end joining
MRN	MRE11, Rad50 and NBS1
MW	Molecular weight
N-DDX52	N-terminus fragment of DDX52
NHEJ	Non-homologous end joining

Ni-NTA	Nitrilotriacetic acid
NTP	Nucleotide triphosphate
OD	Optical density
ORC	Origin recognition complex
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
pegRNA	Prime editing guide RNA
Pol	Polymerase
Pre-rRNA	Precursor ribosomal RNA
PTMs	Post translational modifications
R-loop	Ribonucleic acid loop
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
Rok1	Rescuer of KEM1
RPA	Replication Protein A
Rrp	Ribosomal RNA processing
rRNA	Ribosomal RNA
RT	Reverse transcriptase
S-phase	Synthesis phase
SDS	Sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing
SDW	Sterile distilled water
SEC	Size exclusion chromatography
SF1	Superfamily 1
SF2	Superfamily 2
sgRNA	Single guide RNA
siRNA	short interfering DNA
SNP	Single nucleotide polymorphism
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
SSA	Single strand annealing

SSC	Saline-sodium citrate
SSU	Small sub-unit
T	Thymine
TALEN	Transcription activator-like effector nuclease
TBE	Tris borate ethylenediaminetetraacetic acid
TEMED	Tetramethylethylenediamine
tracrRNA	trans-activating crispr RNA
Tris	Trisaminomethane
U	Uracil
UV	Ultraviolet
WRN	Werner syndrome ATP-dependent helicase
WT	Wild type
ZFN	Zinc-finger nuclease

Abstract

The DExD-box family of helicases represents the largest family of helicases within eukaryotes and have been primarily associated with all aspects of RNA biology, including processes of mRNA synthesis, pre-mRNA processing and ribosome biogenesis. However, despite their key roles and associations with cancer and other diseases, the biochemical activity and function of many of the human forms of these helicases remain uncharacterised. Additionally, despite becoming popularly synonymous as 'RNA helicases', it has become clear in recent years that in addition to their canonical roles within RNA processing, many of these proteins are multi-functional and play important roles within processes of DNA repair, transcriptional regulation and viral immunity, amongst others. In this study we examine and characterise two poorly studied helicases, Probable ATP-dependent RNA helicases DDX49 and DDX52, which are both associated with several cancers and have previously identified connections within viral immunity and DNA repair, respectively. We hypothesised that both genes processed DNA substrates in addition to RNA substrates and by testing recombinant proteins within *in vitro* assays with several DNA substrates were successful in confirming this hypothesis, as well as identifying novel nuclease and annealing activities within DDX49 and DDX52, respectively. We also developed and optimised a CRISPR-Cas9 gene knockout system for mammalian cells and successfully generated and performed preliminary phenotyping of heterozygous U2OS cell lines. Finally, we explore and performed a comparative study of DDX49 with a potential homolog from Asgard Archaea, providing unexpected but novel insight into the biochemistry of the yeast protein Dbp8.

List of figures

Figure 1.1: The chemical structure of the oxyribose and deoxyribose sugars composing the sugar backbone of RNA and DNA, respectively.

Figure 1.2: Multiple domains of DExD box helicases compared with DExH box helicases.

Figure 1.3: Binding models of DExD-box helicases.

Figure 1.4: Partial structural model of DDX52 in complex with ADP.

Figure 1.5: A brief overview of ribosome biogenesis.

Figure 1.6: Overview of a DNA replication fork within DNA replication.

Figure 1.7: A schematic diagram of the process of DNA resection within eukaryotic cells.

Figure 1.8: An overview of the double strand break repair and synthesis-dependent strand annealing models of homology-directed repair.

Figure 1.9: An overview of the non-homologous end joining pathway.

Figure 1.10: An illustration of CRISPR-Cas9 based gene editing.

Figure 2.1: An overview of the FRET assays carried out in this study, using Cy3 and Cy5 as an acceptor and donor pair, respectively.

Figure 3.1: A predicted structural model of human DDX52.

Figure 3.2: Output of IUPred2 and ANCHOR2 predictions for Human DDX52.

Figure 3.3: Structural and domain comparisons between DDX52 and Lhr.

Figure 3.4: Alphafold prediction model of DDX52 superimposed onto *S. cerevisiae* Rok1

Figure 3.5: Comparative sequence alignments of *S. cerevisiae* Rok1 and Human DDX52.

Figure 3.6: An overview of recombinant DDX52 purification.

Figure 3.7: SDS-PAGE of WT DDX52 and associated mutants used within this study.

Figure 3.8: Low concentrations of DDX52 are poorly proficient at unwinding a flayed DNA duplex.

Figure 3.9: Assessing the optimal concentrations of ATP (A) and MgCl₂ (B) for DDX52 unwinding of a flayed DNA duplex.

Figure 3.10: Electromobility shift assays on nucleic acid species show well aggregation as a function of wild type DDX52 concentration.

Figure 3.11: Electromobility shift assays on DDX52 mutants incubated with ssDNA and flayed duplexes reveal distinct binding profiles between C-terminus and full length DDX52 as a function of concentration.

Figure 3.12: Fluorescent polarisation assays as a function of protein concentration show DDX52 binds RNA with a higher affinity than DNA.

Figure 3.13: Unwinding assays confirm a 3' to 5' polarity for DDX52.

Figure 3.14: Summary helicase assays of all DDX52 mutants reveals inhibited unwinding within all mutants and apparent annealing within the N-terminus

Figure 3.15: DDX52 exhibited no selective preference for RNA in unwinding assays on DNA:RNA hybrids and D/R-loops.

Figure 3.16: Helicase assays show unwinding by DDX52 is dependent upon ATP hydrolysis.

Figure 3.17: ATPase assays suggested higher ATPase activity in the presence of ssDNA vs ssRNA.

Figure 3.18: Annealing reactions confirmed DDX52 activities as an ATP-independent annealase on DNA:DNA and DNA:RNA as a function of concentration.

Figure 3.19: FRET-based annealing reactions confirm DDX52 annealing activity is localised to the N-terminus.

Figure 3.20: Annealing assays confirm all mutants apart from F28_T18del mutant possess annealing activities.

Figure 3.21: Nuclease protection assays show DDX52 protects ssDNA.

Figure 3.22. Nuclease protection assays show that DDX52 protects ssRNA and ssDNA.

Figure 4.1: Purification of recombinant NLS-Cas9.

Figure 4.2: Transfection efficiency was improved using forward transfection technique.

Fig. 4.3. Plasmid condition and operator expertise was validated upon high levels of GFP plasmid transfection with optimised Lipofectamine 3000 reagent.

Figure 4.4. ATTO-labelled gRNA complexes confirm that optimised conditions are appropriate for RNP complexes.

Figure 4.5. Commercial Cas9 was more active than purified recombinant Cas9 in activity assays.

Figure 4.6. Exon 5A and 5B guide RNAs demonstrated successful editing in T7 endonuclease assays.

Figure 4.7: TIDE analysis of sequencing results reveals low levels of successful editing.

Figure 4.8: Preliminary TIDE analysis results from the full workflow indicate editing efficiency.

Figure 4.9: Sequencing analysis of samples taken from well D4 generation of a heterozygote.

Figure 4.10: Sequencing trace demonstrating decomposition within off-target site mm4_intron_TTTY21_chrY_9719917.

Figure 4.11: Western blots suggest no change in DDX52 expression levels within edited cell lines.

Figure 4.12: Cell migration and proliferation of wild type and *DDX52*^{+/-} U2OS cells.

Figure 4.13: *DDX52*^{+/-} response to cytotoxic agents.

Figure 4.14: Transfer of U2OS WT and *DDX52* +/- RNA to northern blot membrane was successful.

Figure 4.15: Cycle thresholds in qPCR show reduced levels of 47S, 28S and *DDX52* transcripts within *DDX52*^{+/-} cell lines

Figure 4.16. Plasmid map of pAC29 (pNLS-Cas9).

Figure. 5.1: Predicted structural model of human DDX49.

Figure 5.2: DDX49 possesses a unique C-terminus extension compared to its putative yeast homolog Dbp8.

Figure 5.3: Sequence alignments of DDX49 with its yeast homolog DBP8.

Figure 5.4: IDPR predictions reveal disordered C-termini in both DDX49 and Dbp8.

Figure 5.5: BLASTP data shows conservation of a DPD amino acid sequence when comparing DDX49 to non-mammalian species.

Figure 5.6: Purification of recombinant DDX49.

Figure 5.7: Summary gel of purified recombinant wild type DDX49 and the mutants used within this study.

Figure 5.8: The impure fraction of DDX49 as stained by RUBY staining and viewed under UV.

Figure 5.9: Electromobility shift assay for binding affinity of 'impure' (i-DDX49) and 'pure' (p-DDX49) DDX49 against nucleic acids.

Figure 5.10: Electromobility shift assay shows C-terminus mutants fail to bind DNA and RNA.

Figure 5.11: Examining DDX49s activity on DNA substrates in vitro.

Figure 5.12: DDX49 shows a 3' to 5' polarity of unwinding.

Figure 5.13: DDX49 cuts DNA but not RNA in an ATP-independent manner.

Figure 5.14: Summary of unwinding and nuclease activity of mutants.

Figure 5.15: gRNAs targeting exon 3 and exon 4 successfully edited genomic DNA.

Figure 5.16: *DDX49* editing shows high efficiencies.

Figure 5.17: A *DDX49* heterozygote cell line was generated with a 19 bp deletion.

Figure 5.18: *DDX49*^{+/-} cells show no difference in growth rate when compared with wild type cells.

Figure 5.19: *DDX49* heterozygotes show inhibited migration within wound healing assays

Figure 5.20: Cycle thresholds in qPCR shows the abundance of 47S, 28S and 5.8S rRNA species is impacted within *DDX49*^{+/-} cells.

Figure 5.21: Prime editing of *DDX49* D422A/D424A leaves cells unviable.

Figure 5.22: A comparison of substrates used to study DDX49

Figure 6.1: Sequence alignment of the ORF for MCP8718128.1 (AA.49) against *S. cerevisiae* Dpb8 in yeast and *H. sapiens* DDX49.

Figure 6.2: Comparison of AA.49 and DDX49 C-terminus shows low levels of conservation.

Figure 6.3: Purification of recombinant AA.49

Figure 6.4: Electromobility shift assay for binding affinity of AA49 against nucleic acids.

Figure 6.5: Examining the effect of temperature on AA49 unwinding of a flayed DNA duplex.

Figure 6.6: Examining the effect of salt concentration on AA49 unwinding of a flayed DNA duplex.

Figure 6.7: AA.49 unwinds independently of ATP.

List of tables

Table 1.1: Characteristics of DNA vs RNA.

Table 2.1: Summary, recipes and associated uses of common laboratory buffers.

Table 2.2: Summary and recipes of commonly used gels.

Table 2.3: A summary of the purification buffers used throughout this study and their composition.

Table 2.4: A summary of the plasmids used within this study.

Table 2.5: Oligonucleotides used within this study.

Table 2.6: An overview of DNA substrates used within this study.

Table 2.7: crRNA sequences used within this study.

Table 2.8: Oligonucleotides used within prime editing reactions.

Table 2.9: Strains of *E. coli* used within this study.

Table 2.10: Bacteria culture medium used within this study.

Table 2.11: Media supplements used within *E. coli* culture.

Table 2.12: Human cell lines used within this study

Table 2.13: A table of DMEM supplements used within human cell culture.

Table 2.14: A summary of the antibodies used within this study and their derived species and dilutions.

Table 3.1. DDX52 N-terminal truncation annealed DNA at a rate over two-fold that of comparative wild-type concentrations in FRET-based annealing reactions.

Table 4.1. *DDX52* crRNAs used within this study

Table 4.2: The top five off-target sites identified for *DDX52* gRNA 5B.

Table 4.3: GAPDH and UBE2 show superior performance as housekeeping genes for RT-qPCR assays.

Table 5.1. Proteins identified within mass spectrometry analysis with unique peptides >2.

Table 5.2. An overview of gRNAs tested for editing *DDX49* within this study.

Table 6.1: Potential Archaeon homologs of DDX49 as noted within NCBI BLASTp database.

Chapter 1: Introduction

1.1 DNA and RNA

Nucleic acids were first described in the mid-19th century, but it wasn't until the early twentieth century that DNA and RNA, initially termed thymus and yeast nucleic acids respectively, were noted to possess several important different structural and chemical characteristics (outlined in table 1.1) (1, 2). Key to differences in stability between the two is the single-stranded vs double-stranded nature of RNA vs DNA, believed to play a role in 'protecting' the bases and conferring an added level of stability. However whilst RNA is commonly single stranded, there are several notable exceptions to the rule: DNA:RNA hybrid molecules, such as R-loops and RNA G-quadruplexes, have been observed in vivo whilst RNA readily folds into stem-loop structures that give rise to RNA molecules such as ribozymes and riboswitches (3–5). In perhaps the most critical example, RNA plays a central role in the central dogma of biology by folding into tertiary and quaternary assemblies, giving rise to tRNA and ribosomal subunits, respectively.

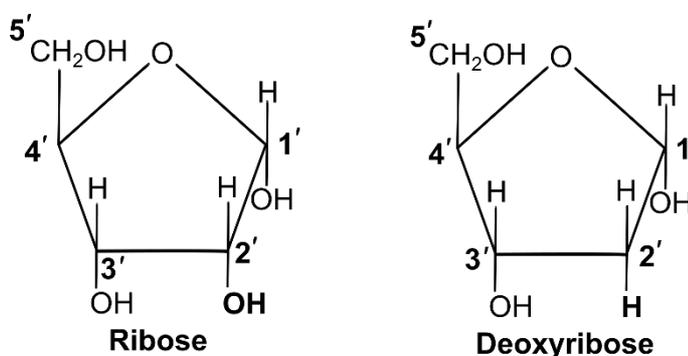


Figure 1.1: The chemical structure of the oxyribose and deoxyribose sugars composing the sugar backbone of RNA and DNA, respectively. The site of the oxygen absent in DNA is indicated in bold.

Table 1.1: Characteristics of DNA vs RNA. A comparison of common characteristics of DNA vs RNA.

	DNA	RNA
Composition	Commonly double-stranded helix	Commonly single stranded, with some exceptions
Functional role	Storage of hereditary information	Various, commonly the transcription and regulation of hereditary information and non-coding RNA (ncRNA)
Sugar-phosphate backbone	Deoxyribose	Ribose
Nucleotides	A, C, G, T	A, C, G, U
Stability	Stable	Unstable
Form	A-form, B-form or Z-form helix	A-form helix only
Associated polymerases	DNA polymerase	RNA polymerase and reverse transcriptase

1.1.1 The Central Dogma of Molecular Biology

Across all domains of life, the genetic information of an organism is encoded within DNA or RNA. In the basic framework of the central dogma, information is described as flowing unidirectionally: DNA is transcribed into the form of transportable cassettes of messenger RNA (mRNA) which transports this to the ribosomes of the cell, molecular factories that transcribe this genetic code into proteins. However, this basic framework is not always applicable. The discovery of reverse transcriptases demonstrated that RNA could be reverse transcribed into DNA, the functions of which are most commonly described in viruses, where they function to replicate the viruses genetic information. Within eukaryotes, reverse transcriptases most notably play a role in the formation of retrotransposons; self-replicating sequences within the genomes of eukaryotes, notably LINE-1 and Alu elements, that utilise reverse transcriptase to copy

themselves to another region of the genome (6). The central dogma has further been challenged in research on prion-based diseases, where it has been argued that information transfer occurs from protein-protein (7).

1.1.2. Functional RNAs

More recently, the classical dogma has been challenged by the changing view that RNAs are not only intermediary messenger molecules but serve additional functional purposes within the cell (8). Whilst the importance of ribosomal RNA (rRNA) (covered in section 1.3) in constituting the ribosome is well-recognised, increasing recognition attributed to non-coding RNAs (ncRNA) such as microRNAs (miRNA) and long noncoding RNAs (lncRNAs) is increasing our knowledge of RNAs additional roles (9). miRNAs are short RNA transcripts, approximately 18-24 nucleotides in length, that are implicated within processes of cellular proliferation, homeostasis and the DNA damage response (10–12). The majority of mammalian mRNAs are thought to be under regulation by miRNAs: in particular, miRNAs have been associated with mRNA repression through recruitment of the miRNA-induced silencing complex (13, 14). Studies suggest an estimated 2300 miRNAs are encoded within the genome; whilst many of these are considered dispensable on an individual scale, complete loss of the miRNA processing machinery leaves affected cells inviable (15, 16).

Generally defined as transcripts > 200 nucleotides in length, the human genome encodes over 60,000 lncRNAs (17). lncRNAs act in regulatory roles within a variety of processes such as modulating chromatin structure and function, RNA splicing and the DNA damage response (18–20). As a result of associations with disease, including as oncogenes, research into lncRNAs has seen a significant rise in interest as advances in understanding lncRNA biology informs new therapeutic designs (21). Notable examples include the lncRNAs NORAD and MALAT1: NORAD has been associated with suppressing metastasis through repression of the YAP pathway, with loss of NORAD upregulating chromosomal instability and aneuploidy, whilst overexpression of MALAT1 promotes proliferation and metastasis in a number of tumours (22–24). However, other studies show MALAT1 acting as a tumour suppressor and inhibiting breast and colorectal cancer cell metastasis, confirming further

research is necessary before lncRNAs can be used as prognostic markers or drug targets (25, 26).

1.2 ATP-dependent nucleic acid processing helicases, translocases and remodelling enzymes

Helicases are molecular motors that employ free energy released by ATP hydrolysis to catalyse binding and remodeling of DNA, RNA and nucleic acid-protein complexes: almost all biological processes that involve DNA or RNA utilize one or several helicases (27). These include essential processes of DNA replication and repair, chromatin remodeling and viral immunity (28, 29). Helicases can be classified into six superfamilies (SF1-SF6) dependent upon conserved sequence motifs within each superfamily, with each superfamily capable of further division dependent upon their polarity of unwinding; type A (3' to 5') or type B (5' to 3') helicases. There are, however, common structural and functional similarities across all superfamilies, notably the presence of a 'core domain' within each enzyme consisting of RecA-like folds that utilize the binding and hydrolysis of nucleotide triphosphates (NTPs) to induce conformational changes within the helicase. SF1 and SF2 helicases consist of two RecA-like folds in a single polypeptide chain, with other superfamilies forming hexameric rings consisting of 6 or 12 RecA folds. The focus of this study will be on Superfamily II helicases. The structure and mechanism of other superfamilies are well-described elsewhere (30).

1.2.1 RNA helicases

RNA helicases are molecular motors for the remodelling and processing of RNA and ribonucleoprotein (RNP) complexes. RNA helicases are found within superfamilies I and II (SF1 and SF2), with the majority found in the SF2 family of helicases (27). RNA helicases are involved in almost all cellular RNA processes and are often thought to primarily function within RNA unwinding, however this has been described as misleading with a number showing a wide variety of biochemical activities (31, 32). A number of RNA helicases, notably DExD-box helicases, unwind RNA in a non-processive fashion and only process secondary substructures whilst other RNA helicases, such as DDX5 and DDX21, have been shown to exhibit RNA annealing activity (33–35).

1.2.2 Superfamily 2 helicases

The superfamily 2 class of helicases represents the largest superfamily of helicases and are commonly composed of 9 motifs atypical of ATPases: Q, Ia, Ib, II through VI. This superfamily can be further divided into several groups, including RecQ-like, Ski2-like and DExD/DExH box helicases, amongst others (27). Motifs I, II and VI show the highest conservation across both superfamily 1 and 2 and coordinate binding and hydrolysis of triphosphates, with motifs I and II also referred to as the Walker A [GXXXXGKT/S] and Walker B [R/K-XXXGXXX-L-hhhD] motifs, respectively (27). The Walker A and B motifs are ubiquitous across all ATPases: motif I binds phosphate groups within triphosphates whilst motif II binds divalent metal cations that act to stabilize ATP hydrolysis intermediates, with NTP binding by motif VI proposed to induce closure of the cleft between domains 1 and 2 (36, 37). It has been proposed that motifs Ia, IV and V play important roles in nucleic acid binding, whilst motifs III and Va have been primarily associated with coordinating the NTP and nucleic acid binding sites (30). The Q motif regulates ATP binding and hydrolysis, whilst also associated with coordinating nucleotide binding (38, 39). The majority of studied superfamily II helicases, with the exception of XPD helicases, have been shown to exhibit 3' to 5' directions of unwinding (30). The focus of this research is on three members of the DExD box family of helicases: DDX49, DDX52 and a potential Asgard archaea homolog of DDX49 referred to herein as AA.49.

1.2.3. DExD/DExH-box helicases

DExD/DExH (DDX/DHX) proteins, often colloquially referred to as DEAD/DEAH-box helicases, are a class of putative RNA helicases defined by eight conserved motifs (presented within figure 1.2) within the family. Initially grouped together under the umbrella term DDX, the group has since been divided based upon differences within these motifs: most prominently, the presence of a histidine residue within the motif II of DExH proteins in substitution of the second aspartic acid residue within DExD-box proteins. In addition, there are conserved differences within the other motifs in addition to a complete absence of the Q-motif within DExH-box helicases: these differences are outlined in figure 1.2 (40, 41).

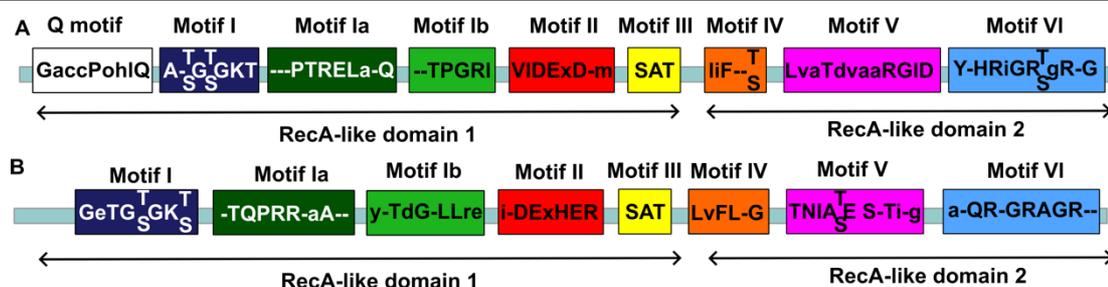


Figure 1.2: Multiple domains of DExD box helicases compared with DExH box helicases. The positions of conserved motifs within the RecA-like domains of DExD-box (A) and a DExH-box helicases (B). Whilst nomenclature of the families is based upon variation within motif II, variation within other motifs is also presented: notably the absence of a Q motif within DExH-box helicases. Motifs within members of each family have typically been defined by the order conserved amino acids appear respective to motifs I and II.

In contrast to the high conservation amongst the central motifs of either family, members of the family show a high level of diversity within their N and/or C-terminus domains that are believed to confer functional specificity through interactions with substrates and interacting factors and have also been implicated within the multifunctional roles of the proteins (40, 42). The C-terminus domain of DP103 (DDX20), for example, is associated with intrinsic transcriptional repression, whereas the N-terminus of RNA helicase A (DHX9) is implicated within transactivation and Pol II binding (43, 44).

As a result of their key functional roles within RNA processing and beyond, dysfunctions with DExD/H-box helicases have been associated with a large variety of diseases, most notably a group of diseases referred to as ribosomopathies. DExD-boxes have also been identified as biomarkers within cancer and are suggested to have therapeutic potential (45–47). However, there remain members of both families where binding and interaction dynamics remain poorly characterised and understood, with further exploration likely to inform future avenues of understanding disease and applying therapeutics (48).

1.2.4. DExH-box helicases

As with other superfamily II helicases, the core of DExH-box helicases feature two RecA-like folds, but additionally feature an N-Terminus extension and several conserved C-terminus sub-domains; including a degenerate winged

helix (WH), a ratchet-like domain and an oligosaccharide binding (OB) fold (49, 50). The N-terminus extension shows high levels of diversity and is believed to confer substrate specificity, shown experimentally in the specificity of DHX36 (RHAU) with G-quadruplexes (51–53). Notably the Q-motif, a domain within DExD-box and Ski2-like helicases that confers specificity for ATP is absent within DExH-box helicases, allowing binding and hydrolysis of other NTPs (54). In addition α -helix 8, conserved across DExD-box proteins and believed to play a role in cleft closure, is absent within DExH-box proteins (55). DExH-box helicases have been noted to be translocating helicases, distinct from the unwinding mechanisms of DExD-box helicases: studies and optimisations of these helicases as molecular motors, such as the helicase Hel308, has seen them applied within novel sequencing technologies (56, 57).

Functionally, DExH-box helicases have largely been associated with mRNA metabolism (31). However, DHX-box helicases have frequently also been shown to be multi-functional. DHX9, for instance, was initially shown to bind both nucleic acids with a twenty-fold higher affinity for ssRNA, but more recent studies have identified similar functional roles to the DNA repair helicases BLM and WRN; including processing of DNA replication forks and G4 quadruplexes, promoting the recruitment of BRCA1 in homologous recombination and D-loop synthesis (58–60). In addition, DHX9 cooperates with CREB-binding protein through its N-terminus to activate RNA polymerase II transcription (43, 61).

1.2.5. DExD-box helicases

DExD-box (DDX) helicases comprise the largest family of superfamily II helicases and are primarily associated with all aspects of RNA biology, including chaperoning mRNAs from synthesis to decay, remodelling of RNA and RNP complexes and ribosomal biogenesis (62, 63). Unlike other SF2 helicases, including DHX-box helicases, DDX proteins possess an alternative mechanism of unwinding RNA. In an early study, eIF4A was shown to unwind RNA duplexes of up to 15 bp with low processivity that decreased as substrate length and stability increased with increased initial rates of unwinding being concentration dependent (64). Further study revealed that DExD-box helicases unwind DNA in a non-processive mechanism of local strand separation, unlike other helicases which unwind via processive translocation along the duplex (65).

Current models of DExD-box binding and unwinding originate from studies of the *Saccharomyces cerevisiae* DExD-box helicase Mss116p (66). Within this model, ATP binds to the RecA1 domain whilst domain 2 functions as an RNA recognition domain: in some proteins, RNA recognition involves interactions across both domains but following complex formation RNA unwinding appears to function similarly across the family (67). Inter-domain contacts between conserved motifs create a functional ATPase site and closes the helicase core (68). This forces the RNA into a sharp bend, introducing a kink within one strand that is incompatible with a double-stranded conformation, resulting in local strand separation (69, 70). ATP hydrolysis releases the tightly bound nucleic acid, allowing the enzyme to be recycled: due to this mechanism of unwinding only several base pairs, DExD-box helicases are known to be subject to futile cycles whereby ATP is hydrolysed before RNA is fully dissociated (illustrated within figure 1.3) (71, 72). Interestingly, binding models noted that α -helix 8 appears to preclude the binding of RNA by blocking the binding site. It has been suggested that interactions between motif V and the C-terminus aspartic acid residue of motif II – the terminal residue of α -helix 8 – rotate α -helix 8 out of the binding site to facilitate binding (55).

The roles of DExD-box proteins within processes of pre-rRNA processing, pre-mRNA splicing and as mRNA chaperones have been well described in literature, with examples of the former described within section 1.3.3. (33, 73, 74). However similar to the DExH family, DExD-box proteins have been reported to play a variety of additional biological functions outside of their canonical roles (40). DDX1 and DDX3X have both been shown to colocalise to sites of DNA damage, with knockdowns of both resulting in DNA damage accumulation, whilst DDX1 is also associated with the removal of DNA:RNA hybrids (75, 76). DDX5 and DDX17 have both been shown to have active roles as both coactivators and corepressors within chromatin organization and transcriptional regulation, acting in concert with lncRNAs and ribonucleoprotein complexes (77, 78). Meanwhile, the viral RNA sensing properties of DExD-box RIG-1 (DDX58) have been commonly described and are essential for detection of a range of pathogenic viruses, including influenza A and flaviviruses (79).

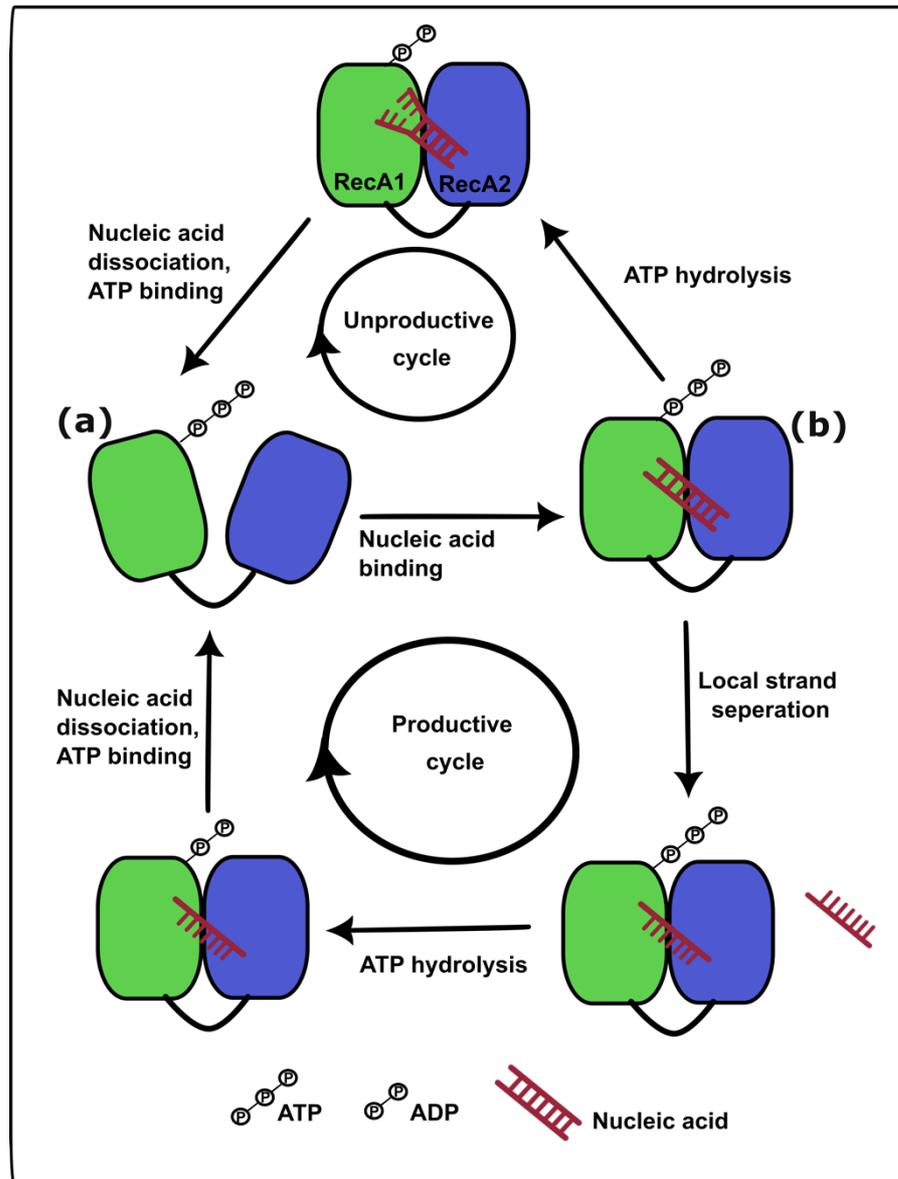


Figure 1.3: Model outlining production and unproductive unwinding cycles of DExD-box helicases. The RecA domain alternates between an open ATP-bound state (a) or closed ATP and RNA-bound (b) state, returning to the open conformation upon ATP hydrolysis. In unproductive cycles, ATP hydrolysis occurs but local strand separation is insufficient to fully separate the duplex prior to dissociation from the helicase, particularly in the case of longer duplexes. In productive cycles, ATP hydrolysis sufficiently dissociates the duplex into two single strands.

Counterwise, several DExD-box helicases are essential host factors for viral replication whereby viruses hijack their activities: DDX1, DDX5 and DDX6 have been associated with facilitating HIV and SARS-CoV2 replication (80, 81). As a

result of these and links of many DExD-box helicases with roles within cancer and other diseases, they are of interest to therapeutics and further research is necessary to characterise their biochemistry and identify additional functions of these proteins within the cell. Two candidate members of this family were identified as poorly studied and of relevance to research: DDX49 and DDX52.

1.2.5.1. Probable ATP-dependent RNA helicase DDX49

Probable ATP-dependent RNA helicase DDX49 is a human enzyme encoded by the *DDX49* gene. There have been limited studies on DDX49, with previous research showing that DDX49 presents robust activity as an RNA helicase, with a regulatory role in mRNA export (82). Awasthi et al (2018) also note that DDX49 is localized to the nucleolus: however, this conflicts with immunofluorescent studies suggesting it is localized to the mitochondria (83). High expression of DDX49 has been associated with highly favourable prognoses within cervical cancer but unfavourable prognoses in renal and liver cancers within data sets collated from comparisons of genome-wide RNA expression and patient survival (83). In addition, research has shown associations with lung cancer, with knockdowns of *DDX49* inhibiting proliferation and migration within PC-9, A549 and H460 lung cancer cell lines, whilst overexpression of the microRNA miR-342 showed that it targeted DDX49 to likewise inhibit lung adenocarcinoma cell proliferation and migration (84, 85). It has also been suggested that inhibitory effects of morphine on hepatocellular carcinoma cells are a result of down-regulation of DDX49 (86).

In addition to roles within cancer, DDX49 has been linked with the pathogenesis of Alzheimers disease, with computational models predicting an interaction between DDX49 and amyloid protein precursor (APP) with increased expression of DDX49 seen in App^{NL-F} mice (82, 87). DDX49 was shown to possess antiviral activity against gammaherpesvirus transcripts in vivo, reducing transcription of lytic viral genes, whilst pre-print data suggested associations with the SARS-COV-2 protein ORF9C (ORF14 in some literature) (88, 89). This potentially implicates DDX49 with the viral/anti-viral activities exhibited by other DDX helicases; however, no currently published literature has explored this relationship.

1.2.5.2. Putative yeast DDX49 homolog (Dbp8)

The purported yeast homolog of DDX49, Dbp8 has been primarily associated with roles within ribosome biogenesis. Previous research found depletion of Dbp8 in vivo results in defects in production of mature 18S rRNA due to impaired cleavage at sites A₀, A₁ and A₂, resulting in 40S ribosomal subunit deficiency (90). Similarly, overexpression of *Dbp8* dominant negative mutants resulted in accumulation of 22S pre-rRNA species, associated with delays in A1 and A2 processing (91). Dbp8 shares interactions with Hsp90 and Esf2, with Esf2 stimulating Dbp8 ATPase activity by acting as a cofactor with an additional suggested role in recruiting Dbp8 its targets (92, 93). Novel bioinformatic comparisons of Dbp8 between DDX49 and a related protein with Asgard Archaea can be found within chapters 5 and 6, respectively.

1.2.5.3. Probable ATP-dependent RNA helicase DDX52

Probable ATP-dependent RNA helicase DDX52 is a human enzyme encoded by the *DDX52* gene. As with the other DExD-box proteins included in this study, very little is known about the biological function of DDX52, with crystal structures showing that its RecA1-domain is highly conserved with other DExD-box proteins (Fig. 1.4) (55). In addition, it is predicted to be a homolog of the *S. cerevisiae* helicase Rok1, a protein with a functional role within ribosome biogenesis and catalytic activity of Rok1 necessary for release of snR30 from pre-ribosomal particles (94). Curiously, the study that identified DDX52 as a homolog of yeast Rok1 – originally referring to DDX52 as Hussy-19 – did so based on a 420 amino acid sequence which differs from the canonical 599 amino acid sequence of DDX52 (95).

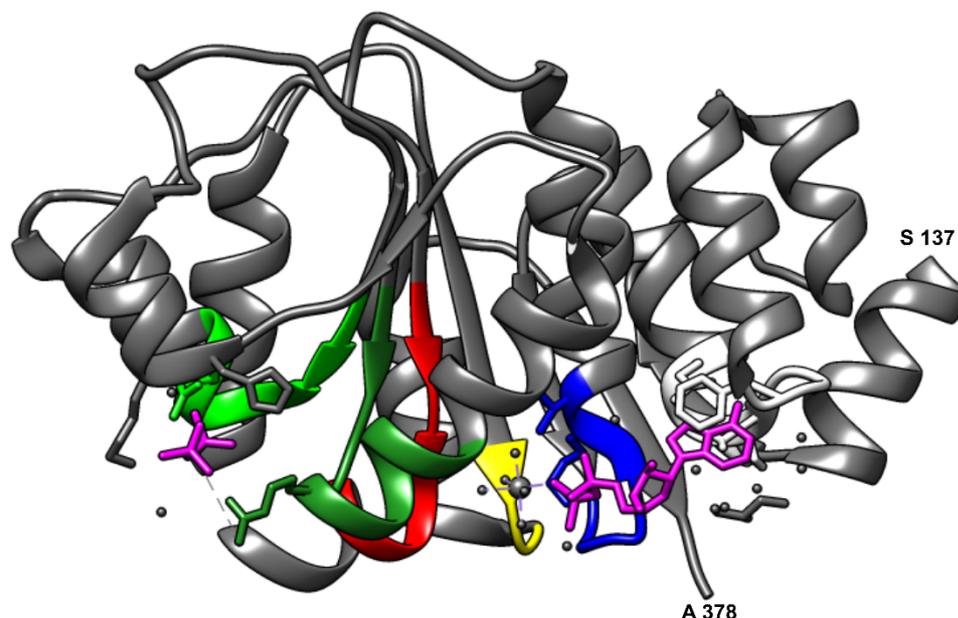


Figure 1.4: Partial structural model of DDX52 in complex with ADP.

DDX52 residues S137-A378 (indicated) are pictured in complex with ADP. Modelled by Schutz et al (2010) using X-ray diffraction (PDB: 3DKP) (96). RecA1 motifs are featured and coloured as in figure 1.2. Bound ADP in magenta.

DDX52 has been shown to upregulate expression of the transcription factor c-myc, with knockdowns of DDX52 inhibiting proliferation of melanoma cells in vitro and prostate cancer cells in vitro and in vivo (97, 98). In addition, high expression of DDX52 has been implicated as unfavourable and favourable prognostic markers in liver and colorectal cancers, respectively (83). Single nucleotide polymorphisms of *DDX52* have been associated with increases in bone mineral density in men, notably in the lumbar spine (99). Knockdowns of DDX52 increased replication of myxoma virus, suggesting an anti-viral role (100). Contrary, knockdowns of DDX52 decreased the production of infectious HIV-1 particles implicating it as a cofactor for HIV-1 replication (101).

Studies of mitotic phosphorylation have implicated DDX52 within DNA repair, with phosphorylation of residue S99 in response to DNA damage from ionizing radiation (102). Two additional residues (Y35 and S39) were found to be phosphorylated within the M-phase of mitosis in a separate study by Dephoure et al (2008) (102). It was noted by Buckley et al (2020) that DDX52 appeared to share structural similarities to the *Mycobacterium smegmatis* protein Lhr, a

protein suggested to play a role within DNA repair and replication based upon interactions with stalled and broken replication forks (104).

1.2.5.4. Putative yeast DDX52 homolog (Rok1)

Rok1 (rescuer of kem1) is a DExD-box helicase within yeast and a purported homolog of DDX52. Bioinformatic comparisons of Rok1 with DDX52 can be found within chapter 3. Previous biochemistry has shown Rok1s functions are ATP-dependent and that it is capable of unwinding duplexes in vitro, with increasing rates of displacement as 3' extension length increased (105, 106). Rok1 has also been implicated within cell cycle progression, with depletion and overexpression arresting the cell cycle in G1/S phase (107). Rok1 mutants have shown it is essential within yeast strains for viability, with depletion blocking 18S rRNA synthesis by inhibiting pre-rRNA processing at sites A₀, A₁ and A₂ potentially through interaction with snoRNPs (108, 109).

Rok1 has been closely linked with the yeast protein Kem1 (XRN1 in humans), with Kem1 reported to regulate ROK1 expression and expression of Rok1 on high copy number plasmids rescuing loss of kem1 functions (110, 111). Interestingly expression on low copy number plasmids was not enough to rescue null mutations suggesting that, if the proteins are functionally related, Rok1 possesses a weaker activity. Rok1s activity in ribosome biogenesis is closely linked with its cofactor Rrp5, with ATP-bound Rok1 stabilising binding of Rrp5 to 40S ribosomes whilst ATP-hydrolysis allows release of Rrp5 for its role within 60S subunit assembly (112). Likewise Rrp5 is implicated within association of Rok1 with 35S rRNA, with addition of Rrp5 also enhancing Rok1 specificity and catalysing duplex annealing (113, 114). Interestingly, despite this cooperativity, studies have demonstrated that overexpression of Rok1p is sufficient to rescue 18S rRNA defects upon Rrp5p depletion (115).

1.3 Ribosome Biogenesis

Ribosomes are molecular factories within cells that produce all cellular proteins during translation. As such the assembly of ribosomes (ribosome biogenesis) is a complex and heavily regulated process, involving co-ordination of all three RNA polymerases, approximately 200 accessory factors and over 80 ribosomal proteins (116). As such, even minor changes in the pathway give rise to a range

of diseases termed ribosomopathies (117). Within eukaryotes, the ribosome is constituted of two subunits: a small 40S subunit, responsible for recognising and binding mRNA and composed of 18S ribosomal RNA (rRNA) and 33 ribosomal proteins (r-proteins), and the large 60S subunit responsible for forming peptide bonds and composed of 25S, 5.8S and 5S rRNA and 47 r-proteins (118).

Evidence suggests that the basic mechanisms of biogenesis is conserved within eukaryotes, with most experimental evidence arising from genetic and biochemical techniques within the yeast *S. cerevisiae* (119). However due to the increased complexity of human ribosomes, including extended expansion segments, human ribosomal biogenesis requires many additional components (120). However, the human pathway of ribosomal biogenesis is poorly studied in comparison to the established model from yeast: as such this section will cover the established models with note to exceptions in human models.

1.3.1. A brief overview of ribosome assembly

Assembly of eukaryotic ribosomes predominantly takes places within the nucleolus, with later steps of maturation occurring in the nucleoplasm and cytoplasm. All three RNA polymerases are used within ribosome biogenesis, with the four rRNAs encoded within a tandem array of 9.1kb units that is repeated up to 200 times on chromosome 12. Pol I and Pol III synthesise the 35S rRNA precursor (47S in humans) and 5S rRNA precursor, respectively, whilst Pol II synthesises snoRNAs and mRNAs encoding the ribosomal proteins (121, 122). The 35S/47S rRNA precursor is flanked by two external transcribed spacers at each end (5' ETS and 3' ETS) and is composed of 28S, 18S and 5.8S rRNA molecules separated by two internal transcribed spacers (ITS1 and ITS2) (see figure 1.5). The nascent transcripts associate with a combination of ribosomal proteins, cis-acting elements and trans-acting factors to form a series of large ribonucleoprotein (RNPs) where pre-rRNA folding and modification take place. These ribosomal precursor particles have also been seen to contain an unknown number of non-r-proteins whose identity is uncertain (119, 123).

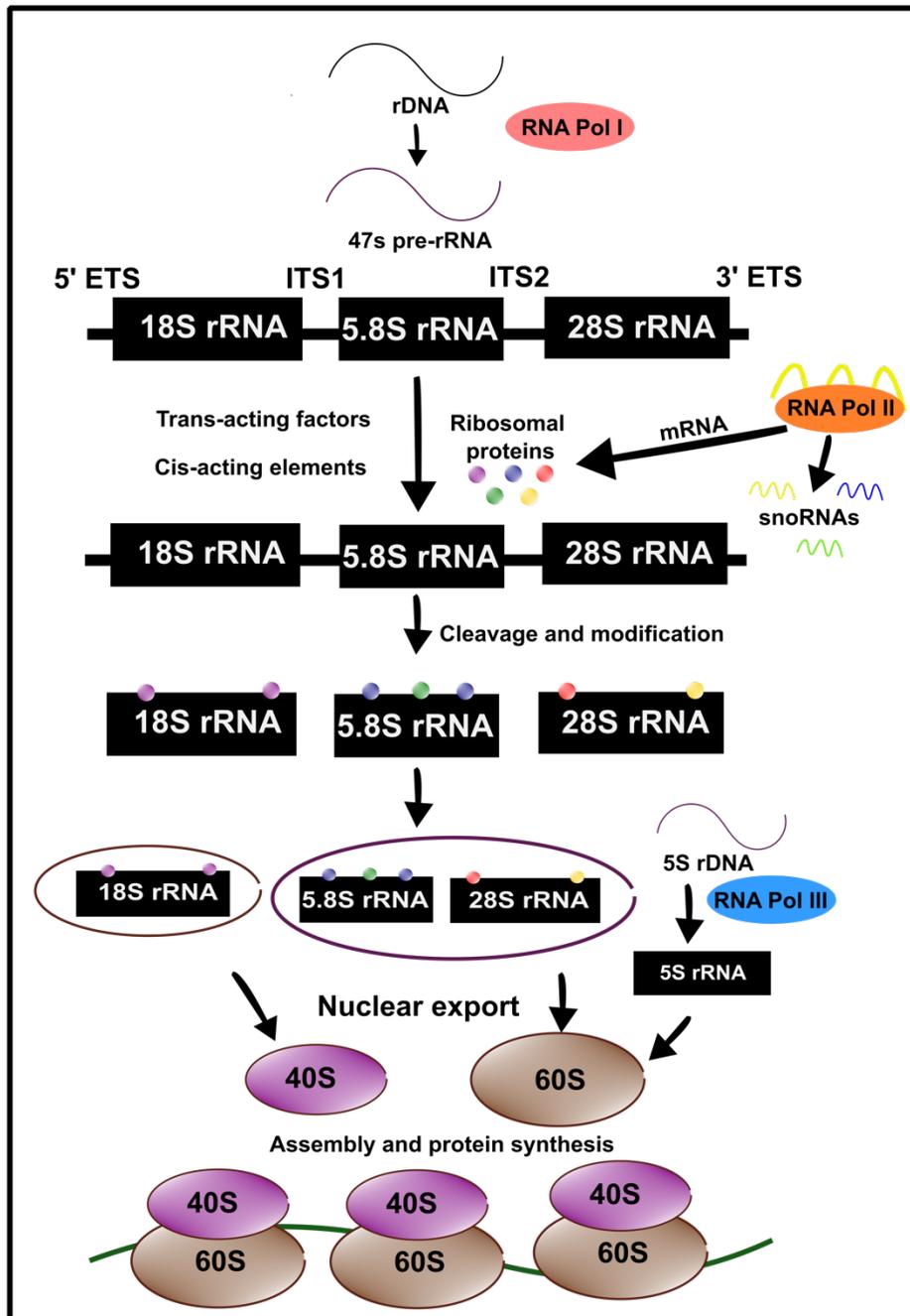


Figure 1.5: A brief overview of ribosome biogenesis. Transcription of rRNA proceeds initially through the action of RNA Polymerase I along with a subset of other factors to synthesise 47S pre-rRNA. This precursor rRNA transcript encodes three rRNAs flanked by two external transcribed spacers at each end (5' ETS and 3' ETS) and separated by two internal transcribed spacers (ITS1 and ITS2). Through a series of cleavage and modification steps, 47S pre-rRNA is processed into mature 18S, 5.8S and 28S rRNA. RNA Polymerase II synthesises snoRNAs and the mRNA of ribosomal proteins that participate

within ribosome biogenesis. Assembly of the 18S rRNA and ribosomal proteins gives rise to the small (40S) subunit. 5S rRNA is synthesised through the actions of RNA Pol III and assembly of 5S, 5.8S and 28S rRNA with ribosomal proteins forms the large (60S) subunit. Subunits are exported from the nucleus and join to form competent ribosomes for protein synthesis.

Maturation of pre-rRNA begins with the endonucleolytic cleave of the 5'-ETS and ITS1 sequences: within yeast, 18S rRNA is entirely generated by a series of endonucleolytic cleavages whilst in mammalian cells a series of endonucleolytic and exonucleolytic steps are necessary (123). Mature 5.8S is subsequently generated via two different pathways to produce a short form or a long form. Following endonucleolytic cleavage of ITS2, maturation of the 3' end of 5.8S and 5' end of 25S (28S) rRNA is facilitated through the action of 3' → 5' and 5' → 3' exonucleases, respectively. Maturation of the 3' end of both 18S rRNA and 5.8S rRNA is completed upon export to the cytoplasm by endonucleolytic and exonucleolytic steps, respectively.

As such, ribosome biogenesis is a complex process regulated by a vast number of proteins, cis and trans-acting factors. In addition to the DHX-box subfamily of proteins, these are beyond the scope of this study and can be found comprehensively reviewed elsewhere (119, 124, 125). The rest of this section will focus on discussion of ribosome biogenesis within disease and the role of DExD-boxes and/or their reported yeast homologs in ribosome biogenesis.

1.3.2. Ribosome biogenesis associations with animal disease

The ribosome has recently become re-evaluated as research identifies heterogenous ribosome composition, suggesting this gives rise to ribosome specialisation (126, 127). As such, due to their critical role as protein production factories within the cell, it has been found that non-lethal alterations within even individual parts of the ribosomal machinery can induce cellular dysfunction and even disease. Ribosomopathies are a series of disorders characterised by haploinsufficiency in genes that encode key proteins or accessory factors associated with ribosomal biogenesis. Often these mutations result in tissue-specific phenotypes associated with development dependent upon the individual protein affected. The most well-studied ribosomopathy is that of

Diamond-Blackfan anemia, with examples of other phenotypes including Treacher Collins syndrome, Shwachman Diamond syndrome and Dyskeratosis congenita (128). Likewise, ribosomopathies have been implicated within neurodevelopment diseases such as microcephaly (116, 129). As the network of proteins and accessory factors that result in different forms of ribosomopathies is extensive, comprehensive reviews can be found elsewhere (128, 130, 131).

Ribosomopathies have also been associated with playing a role within the processes of aging, cancer, neurological and cardiovascular disease and viral infections such as COVID-19. Aging has previously been associated with increases in nucleolous activity as a result of increased rDNA transcription, with decreased rates of translation associated with increases in an organisms lifespan (132). However this is a subject of debate as studies have also implied that decreases in nucleoli size and rRNA transcription are associated with premature aging phenotypes within both Bloom and Werner syndrome (116, 133). Through similar mechanisms, ribosomopathies are associated with neurodegenerative diseases, with observations in both Alzheimers and Parkinsons supporting impaired nucleoli activity (134, 135).

Central to protein synthesis, ribosomal dysfunction has been associated with driving tumorigenesis and cancer: a number of the ribosomopathies listed above result in a predisposition to cancer development (136). It has been suggested that hyperactivation of ribosome biogenesis might increase rates of total protein synthesis and aberrant ribosomes that escape degradation, resulting in a reduction in the translation fidelity of mRNAs and contributing to cancer initiation and progression (137). In addition, the extra ribosomal roles of some proteins has been implicated within oncogenesis: for example, a number of ribosomal proteins participate within a 'negative feedback loop' with the oncogenic factor c-myc, with heterozygosity of Rpl11 and Rpl24, resulting in upregulation and downregulation of c-myc, respectively (138, 139). Recent successes using RNA polymerase inhibitors highlights how further understanding of the composition of the mechanisms of these dysfunctional ribosomes can potentially be used to inform drug development, including the design of drugs that preferentially bind 'onco-ribosomes' (136).

1.3.3. DExD-boxes in ribosome biogenesis

1.3.3.1. Yeast as a model organism

Several DExD-box RNA helicases are associated with ribosome biogenesis in humans, but very few have been characterised in detail. Based upon expected conservation of the ribosome biogenesis pathway amongst eukaryotes, many are expected to share functions with their yeast homologues (140). This has been experimentally supported in some instances: depletions of DDX48 (eIF4AIII) and fal1p within human and yeast cells, respectively, resulted in decreased 18S rRNA levels, whilst complementation of DDX48 within fal1 Δ (null) yeast strains rescued 18S rRNA biogenesis (141, 142). Similarly like it's reported yeast homolog the helicase DDX47 has been associated with early pre-rRNA processing, potentially in 60S subunit assembly based upon affinity towards 26S and 5.8S pre-rRNAs (143).

However, there are several experimental cases with conflicting evidence between yeast and human homologs. DDX5 has been associated with the processing of 28S and 18S rRNA as well as the promotion of 47S rRNA transcription (144, 145). However DDX5s role in ribosome biogenesis shares functional redundancy with DDX17, with only silencing of both genes showing negative impacts on cell proliferation: notably, the reported yeast homolog of DDX5 (Dbp2p) does not show such redundancy (144, 146). In vivo depletion of Dbp8 in yeast blocked synthesis of 18S rRNA whilst Awasthi et al (82) described the human homolog, DDX49, as regulating transcription and synthesis of 47S rRNA (90).

In addition recent research has identified a role of DDX1, a young gene absent in yeast, within rRNA processing through knockout studies though its role is presently unknown, suggesting a role exclusive to higher eukaryotes (147). Similarly, no yeast homologs have been observed for DHX9 and DHX33, with DHX9 responsible for the processing of IGS-rRNA into pRNA necessary for heterochromatin formation at rRNA genes, whilst knockdowns of DHX33 decreased rRNA transcription through decreases in Pol I association (148, 149). As such characterisation in humans is vital to confirming current models of ribosome biogenesis.

1.3.3.2. Human

DDX5, DDX10, DDX17, DDX21 DDX24, DDX27, DDX51 and DDX54 have been linked with the production of 18S and 28S rRNAs (74, 150, 151). As mentioned previously, DDX5 shares a functionally redundant role in ribosome biogenesis with the very similar DDX17, associated with the processing of 28S and 18S rRNA as well as the promotion of 47S rRNA transcription (144, 145). DDX10 and DDX18 were identified within the SSU processome, with DDX10 localised within a complex involved within 18S maturation (152). Knockdowns of DDX18 reduced rRNA transcription due to increased occupancy of PRC2 and H3K27me3 on rDNA loci (153). A helicase deficient mutant of DDX21 with ATPase activity rescued 28S rRNA production but did not restore 18S rRNA levels, suggesting a role of DDX21 helicase activity in 18S rRNA processing (150). Interestingly overexpression of DDX21s paralogue, DDX50, inhibited rRNA production, suggesting an antagonistic relationship between the two proteins (150).

DDX24 was essential for processing of ITS1 site 2 and 28S rRNA synthesis, similar to its interactions between its yeast homolog Mak5p with 25S rRNA, but conflicts were noted between studies regarding the impact upon specific pre-rRNAs (154–156). DDX27 regulates formation of the 3' end form of 47S rRNA, though the mechanism of this is unclear: it has been suggested that it might play a similar role to DDX51, which plays a role in 28S rRNA maturation by promoting dissociation of U8 snoRNA from pre-ribosomal subunits (157, 158). Silencing of DDX51 reduced the levels of all precursors, whilst knockdowns of DDX54 was associated with an accumulation of 30S pre-rRNA (154, 159).

1.3.3.3. Perspectives on future research

It is thus evident that whilst similarities show consistency between models, a substantial number of differences across yeast and humans indicate that for characterising the mechanisms and design of effective therapeutics towards disease impacted by these, research should be guided by testing the human version of these proteins. Research is also needlessly complicated by the use of different nomenclature, with researchers using names interchangeably: for example, in a study by Yan et al (2003) on DDX20 the authors predominantly refer to it as DP103 whilst Bizen et al (2022) in a similar study primarily refer to

it as DDX20. This inconsistent practise complicates consolidation of research on these proteins, stemming from the assumption that both yeast and human proteins function the same.

1.4. DNA replication

Reported homologies of DDX52 with Lhr, previously shown to interact with destabilised replication forks, presents it as a candidate for a potential role within DNA replication (104). The accurate and complete replication of an organism's DNA is essential for cell growth and division across all domains of life. DNA replication is a semi-conservative process across all known organisms, though several key differences arise between bacteria and eukaryotes: whereas bacteria replicate their DNA continuously, even beginning a new round of replication before a prior round is complete, eukaryotic replication occurs primarily within the S phase. Additionally, bacteria typically only possess one single origin of replication per chromosome whilst the size and complexity of eukaryotic and some archaea genomes necessitates multiple origins of replication per chromosome (161). These origins of replication appear as 'replication bubbles', forming as the DNA double helix opens. As such, eukaryotic DNA replication is a tightly regulated process to ensure that DNA molecules on each chromosome are only copied once per cell division (162). This section will focus on the eukaryotic model of DNA replication.

1.4.1. Initiation of DNA replication at origins of replication

Initiation within Eukaryotes occurs via assembly of the pre-replicative complex (pre-RC) by successive binding of the origin recognition complex (ORC) and cell division cycle 6 (CDC6) to each potential origin of replication during the G1 phase of the cell cycle (163). Activation of ORC by CDC6 results in subsequent recruitment of chromatin licensing and DNA replication factor 1 (Cdt1), which acts as a chaperone for binding of the eukaryotic replicative helicase MCM2-7 complex, forming the pre-RC and a licensed origin of replication.

Within the pre-RC, MCM2-7 is inactive and encircle dsDNA, forming a tight dimer interface. In late G1/early S-phase, pre-RCs are activated through the action of CDK, DDK and other accessory factors, resulting in binding of CDC45 and the GINS complex to the MCM2-7 complex, forming active CDC45-MCM-

GIN5 (CMG) helicase. The mechanisms behind activation of the MCM complex for unwinding represents a long-standing question, with recent research suggesting Histone O-GlcNAcylation plays a role through orientating DDK recruitment to chromatin (164). Translocation of MCM along the leading strand in a 3' to 5' direction unwinds the origin, assembling two replication forks that replicate DNA in opposite directions through the action of DNA polymerases.

1.4.2. Elongation of replication

Elongation is the process by which the daughter DNA strand is synthesised through addition of free deoxyribonucleotide triphosphates (dNTPs) to the exposed 3-hydroxyl group of the last incorporated nucleotide in a series of base-pairing and chain formation reactions catalysed by DNA polymerases. Though the human proteome contains a number of polymerases, only four are involved within DNA synthesis; Pol α , δ , ϵ and γ , with the latter of which only involved within mitochondrial DNA replication (165). As all DNA polymerases synthesize in a 5' to 3' direction, co-ordination is required to synthesise both leading and lagging strands. A simplified diagram of the replication fork is presented in figure 1.6.

To begin elongation, RPA assists with the loading of the four-subunit complex pol α -primase. Primase subunits create short RNA primers of 7-10 nt in length, with pol α extending these primers to ~ 30 bases in length (166, 167). Once the initial primer is in place, an activity known as 'polymerase switching' occurs and, in the case of leading strand synthesis (5' – 3'), Pol α dissociates from the template to be replaced by Pol ϵ , which exhibits higher processivity and carries out continuous synthesis along the leading strand. Pol ϵ also possesses intrinsic 3' – 5' exonuclease activity allowing it to proofread its own errors. Pol δ acts alongside Pol ϵ throughout elongation, though there is still debate as to how the labour is divided. The preeminent model posits that Pol ϵ is restricted to the leading strand with Pol δ functioning on the lagging strand, supported by mutational bias on the leading and lagging strands of Pol ϵ and Pol δ defective cells, respectively (168, 169).

In either case, lagging strand synthesis proceeds discontinuously through repeated priming and synthesis of short fragments of DNA, known as Okazaki

fragments, by pol α -primase and Pol δ respectively. The Flap endonuclease Fen-1 processes the 5' end of Okazaki fragments whilst DNA ligase I converts the fragments into one continuous strand (170, 171). As primer synthesis is several magnitudes slower than the rate of leading strand synthesis, the mechanism by which synthesis of both strands is coordinated is an ongoing question. Previous models have suggested that priming pauses leading strand synthesis and that leading-strand synthesis proceeds during primer synthesis, forming a lagging strand ss-loop emerging from the helicase which infrequently pauses leading strand synthesis (172–174). As elongation proceeds, topoisomerases are recruited to relieve topological stress that accumulates in unreplicated DNA ahead of the polymerase whilst simultaneous production of new histone proteins reassociate with the newly synthesised duplex to re-establish the nucleosome.

1.4.3. Termination

Termination of DNA replication begins when converging replication forks intersect to a point that the formation of supercoils in replicated DNA becomes unfavourable. Typically occurring when 150 bp of parent DNA remains, due to the stiffness of DNA, relief of topological stress becomes dependent upon precatenanes (175). The CMG helicases bypass each other and translocate until reaching a downstream Okazaki fragment, at which point they pass the ssDNA-dsDNA junction and proceed on dsDNA – interactions between CMG and Pol ϵ in leading strand synthesis have been proposed to drag the polymerase with it, freeing up the 3' end for other replication machinery (176, 177). Evidence further suggests that resultant gaps on the leading strand and final Okazaki fragment is subsequently processed through the recruitment of Pol δ and Fen1 (175, 178).

Most termination events are not sequence specific, but at least two site-specific termination events occur. These include replication fork barriers (RFBs) that stall the replication fork, notably within rDNA loci. Each rDNA cluster contains a termination element bound by a terminator protein which stalls replication fork progression preventing collisions between replication and transcriptional machinery which can induce genome instability (179). The mechanisms by which termination at RFBs proceeds has not been examined, but it has been

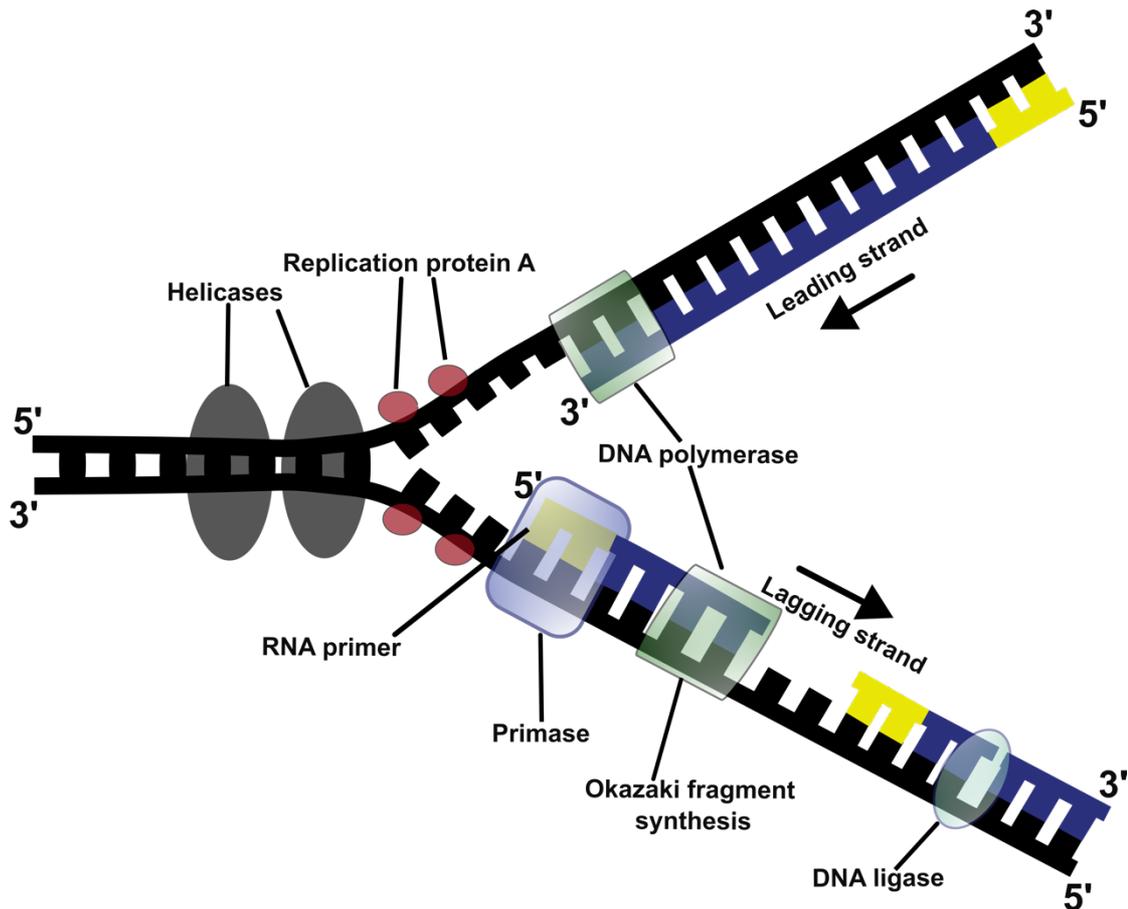


Figure 1.6: Overview of a replication fork within DNA replication. Leading and lagging strand synthesis proceeds in a 5' to 3' direction in a continuous and discontinuous manner, respectively. Primase generates short RNA primers (indicated in yellow) to initiate synthesis on the leading strand and each Okazaki fragment on the lagging strand. Okazaki fragments are later linked together through the actions of DNA ligase. For presentation purposes base pair lengths are not to scale.

suggested that replication forks readily restart and terminate normally upon displacement of the terminator protein (175). A second instance of site-specific termination occurs within telomeres. Telomeres consist of long stretches of short repeating DNA – TTAGGG in humans – and are essential as the ends of chromosomes resemble double-stranded breaks, predisposing them to fusion or degradation. It has been suggested that, due to association of pol α with CMG, termination occurs when pol α slides off the chromosome end with CMG, preventing further priming and resulting in telomere shortening with DNA lost

on the lagging strand roughly equivalent to the length of individual Okazaki fragments (180).

1.5. DNA repair and recombination

An organisms DNA is constantly exposed to damage through a combination of DNA damaging agents and errors within the replication machinery, with persisting damage resulting in stalling of the replication fork and transcription machinery, potentially resulting in mutagenic events (181). As such, robust and efficient repair pathways are essential for the cells response to damage and the maintenance of genome stability and perpetuation of the organism. Failures in repair giving rise to cell death or uncontrolled replication resulting in diseases such as cancer: as such, associations of several cancers within humans with both DDX49 and DDX52 presents a potential role for these helicases within pathways of DNA repair (84, 98, 182). This section will primarily focus on homology-directed repair and non-homologous end-joining double-strand break repair pathways: comprehensive reviews of single-strand break repair pathways can be found elsewhere (183, 184). Alternative end-joining mechanisms, notably the mutagenic pathways of single strand annealing (SSA) and microhomology-mediated end joining (MMEJ), are similarly covered elsewhere (185–188).

1.5.1 Homology-Directed repair

Homologous recombination is a process whereby genetic material is exchanged between two homologous sequences of nucleic acids, most commonly DNA. This has several functions, notably in the repair of double-strand breaks by homology-directed repair (HDR) and generating genetic diversity in meiosis. As HDR uses a homologous strand as a template for repair, repair by this pathway is largely accurate. Comparatively, however, it is also a slower process due to the large network of proteins involved (189). HDR can be subdivided into several models of repair: double-strand-break repair (DSBR), synthesis-dependent strand annealing (SDSA) and break-induced replication (BIR). Though DSBR and SDSA are distinct repair pathways, both are Rad51 recombinase-dependent and require prerequisite formation of a displacement loop (D-loop) to function. In contrast, whilst the canonical BIR pathway is similarly Rad51-dependent, studies within yeast have described a Rad51-

independent mechanism of BIR in strand invasions with shorter homologies (~33 bp) than those in canonical BIR (~ 100 bp) (190). It has been suggested that Rad51 filaments are inhibitory to this type of repair, but the mechanisms of this type of BIR are still largely unknown (191, 192). The mechanisms of BIR were considered outside of the scope of this study and can be found reviewed extensively elsewhere (215–217).

1.5.1.1. DNA resection

Within the initial step of short-range DNA resection, the MRN (MRE11-RAD50-NBS1) complex and phosphorylated CtIP are recruited to the site of a double-strand break. Endonuclease activity of MRE11, driven by RAD50 ATPase activity and cofactors NBS1 and CtIP, generates a nick (as indicated within figure 1.7) within the dsDNA. Limited 3' to 5' exonuclease activity of MRE11, accelerated by interactions with EXD2, subsequently creates a 3' ssDNA overhang of up to 300 nt (193, 194). This initial step is slow and limited to the DNA end, but has the capacity to process double-strand breaks with protein adducts and secondary structures, whilst cell-cycle-dependent phosphorylation of CtIP acts as a key limiting factor restricting HDR to S/G2 phases, when a template strand is available (195, 196).

Replication protein A (RPA) coats 3' ssDNA overhangs to protect them from nuclease activity and inhibit the formation of secondary structures (197). The 3' ssDNA overhangs generate an entrance point for the longer-range resection enzymes EXO1 and DNA2 which resect DNA in a 5' to 3' direction to produce a 3' ssDNA overhang of up to 7kb. DNA2 mediated end resection is dependent upon helicase activity, with DNA2 cooperating with the helicases BLM and WRN to cleave ssDNA generated through helicase unwinding (198, 199). The mechanisms that regulate extent of DNA resection remains an open question, with extensive resection favouring the mutagenic single-strand annealing process (187, 196). Both Neddylation and CtIP have been closely linked with modulating excess resection, specifically CtIP residues 550-600 and its interactions with BRCA1 (200–202).

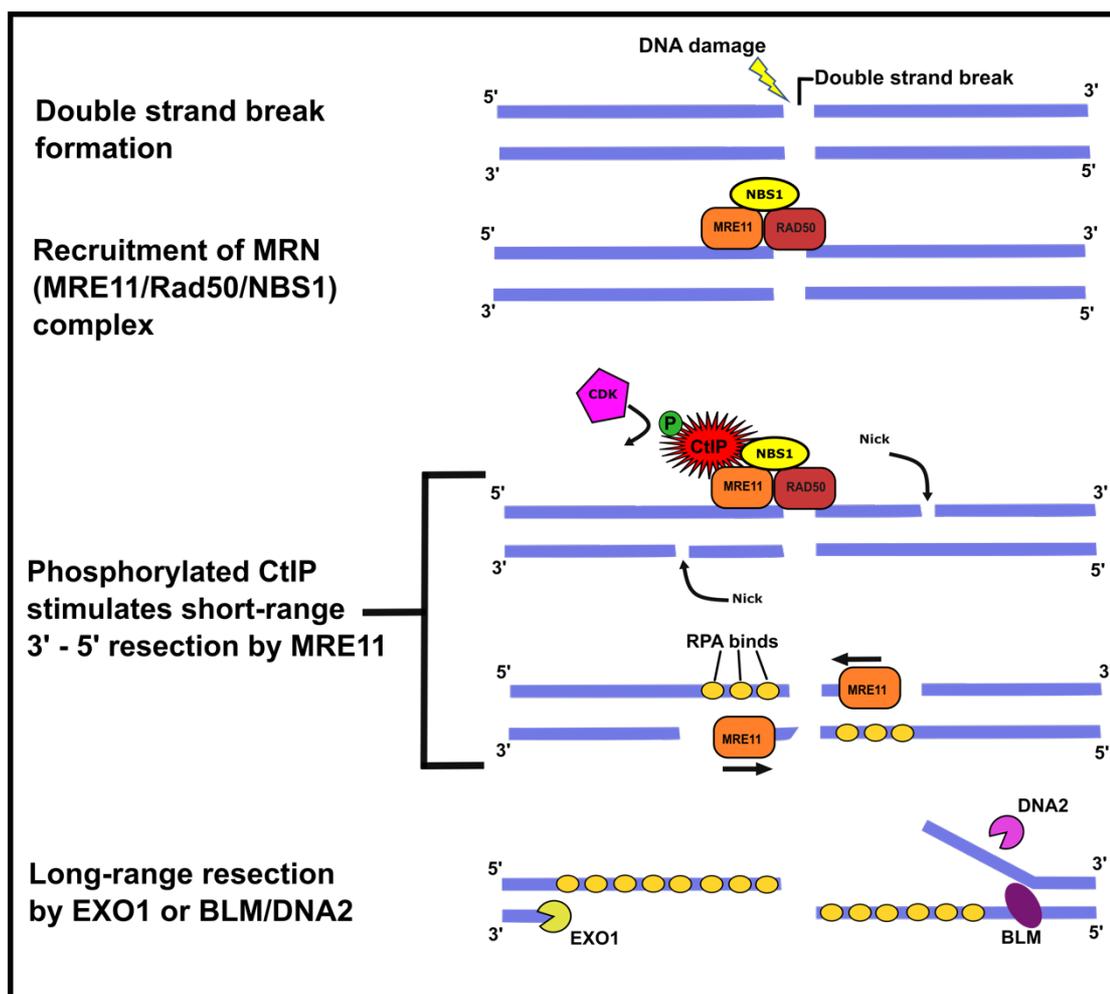


Figure 1.7: A schematic diagram of the process of DNA resection within eukaryotic cells. DNA damage triggers a double strand break. The MRN (MRE11/RAD50/NBS1) complex and phosphorylated CtIP are recruited to the site of the double-strand break. The function of CtIP is dependent upon its phosphorylation by CDK, as presented. Endonuclease activity of MRE11, with RAD50 and CtIP as cofactors, generates a nick within the dsDNA with subsequent limited exonuclease activity of MRE11 generating a short 3' ssDNA overhang. This facilitates recruitment and long-range DNA resection by the exonuclease EXO1 and endonuclease DNA2, with DNA2 acting in concert with Bloom helicase (BLM) mediated DNA unwinding to cleave ssDNA. Binding of RPA protects the ssDNA overhangs from nuclease activity.

RPA-ssDNA complexes trigger the ATR-ATRIP pathway, recruiting ATR and ATM which target protein kinases CHK1 and CHK2, reducing CDK activity and arresting the cell cycle (203). Exchange of RPA for RAD51 forms a RAD51-

ssDNA complex referred to as the presynaptic filament. High affinity of RPA for ssDNA precludes RAD51 binding, forming an inhibitory barrier that is overcome by recombination mediator proteins; notably Rad52 in yeast and BRCA2 in humans. BRCA2 inhibits ATP-hydrolysis and dsDNA binding activities of RAD51, promoting ssDNA binding, whilst the BRCA-associated protein DSS1 assists in displacing RPA by mimicking DNA and reducing RPA affinity for ssDNA (204). Whilst Rad52 plays a critical role in Rad51 loading in yeast, human RAD52 is believed to play a subtler role in recombination and the single strand annealing pathway; however it remains a topic of ongoing research with recent studies continuing to identify new interactions and functions, notably interactions with DSS1 within SSA and BIR (205).

1.5.1.2. Homology search and strand invasion by recombinase

The mechanisms behind homology search are still unclear. Proposed models include suggestions that the presynaptic filament probes the genome at random, making temporary contacts with various duplexes until finding a stable homologous sequence (206). It has also been suggested that one-dimensional sliding of the presynaptic filament along the DNA plays a key role (207). Research within yeast has shown Rad54 co-ordinates ATP-dependent translocation of a migrating DNA structure during homology search, also demonstrating a mechanism whereby Rad54 and RPA co-operate in bidirectional homology sampling (208). Whilst further research is required to understand homology search, it concludes with the presynaptic filament invading duplex DNA and displacing the original strand to form a displacement loop (D-loop) structure. Newly displaced ssDNA is stabilised by RPA whilst yeast studies have demonstrated Rad54 ATPase activity stimulates Rad51 displacement from the invading strand, promoting D-loop stability (209).

Following D-loop formation repair proceeds according to three pathways, with pathway choice influenced by the expansion of D-loops and the success of second end capture: stable D-loops successfully capture the second end and proceed by the double-strand-break repair (DSBR) pathway, whilst disruption of the D-loop channels repair towards the synthesis-dependent strand annealing (SDSA) (210). Metazoan RTEL1 and mammalian Rad54, FANCM and RECQ1 have been associated with D-loop disruption and thus SDSA

promotion, with RECQ1 preferentially displacing D-loops where invasion occurred at the 5' end: such intermediates are potentially toxic, posing a functional use for RECQ1 (211). Rad52 in yeast has been shown to encourage second end capture through promoting the annealing and flexibility of ssDNA complexed with RPA (212, 213). Stalling or failure of DSBR and/or SDSA results in repair directed by a further process known as break-induced replication (BIR).

1.5.1.3. Double-Strand Break Repair pathway

Following D-loop stabilisation, subsequent second end capture by the opposing resected end followed by further synthesis and ligation results in formation of a double Holliday junction (dHJ). Operating via Rad52 catalysed annealing in yeast, human RAD52 also mediates second end capture in vitro but mutant *RAD52* phenotypes are inconsistent with such a critical role HR making its role within second-end capture unclear and suggests potential redundancy (209, 214). The BRCA2 homolog Brh2 of *Ustilago maydis* has been shown to promote second-end capture, however human BRCA2 is incapable of annealing RPA-coated DNA (215, 216). Proposals have suggested an unidentified protein may fulfil this role or act as a co-factor enabling BRCA2 annealing: alternatively, it has been suggested that second end capture may occur via a second independent strand invasion step, though it has been tentatively referenced this mechanism might preclude SDSA in human cells (215).

Holliday junctions are processed by mechanisms of resolution or dissolution, regulated according to the type of molecule being processed and the needs of the cell (217). Resolution is described in section 1.5.2 and results in either crossover or noncrossover products: crossover products are undesirable within mitotic cells and resolution is suppressed whilst in meiotic cells crossover events are promoted, notably through Cdc5-mediated phosphorylation (217, 218). Dissolution is performed by a complex referred to as a dissolvasome, comprised of BLM, topoisomerase III α , RMI1 and RMI2. ATP-motor activity of BLM catalyses convergent migration of Holliday junctions towards each other, forming a hemicatenane, whilst TopoIII α relieves positive supercoiling ahead of the migration junction and nicks the hemicatenane to form two non-crossover products (219). Biochemical studies indicate the role of RMI1 is likely in later

steps of dissolution, specifically decatenation (219). RMI2s role in dissolution is believed to be minimal, recruiting the Fanconi anaemia protein FANCM to stalled replication forks (220).

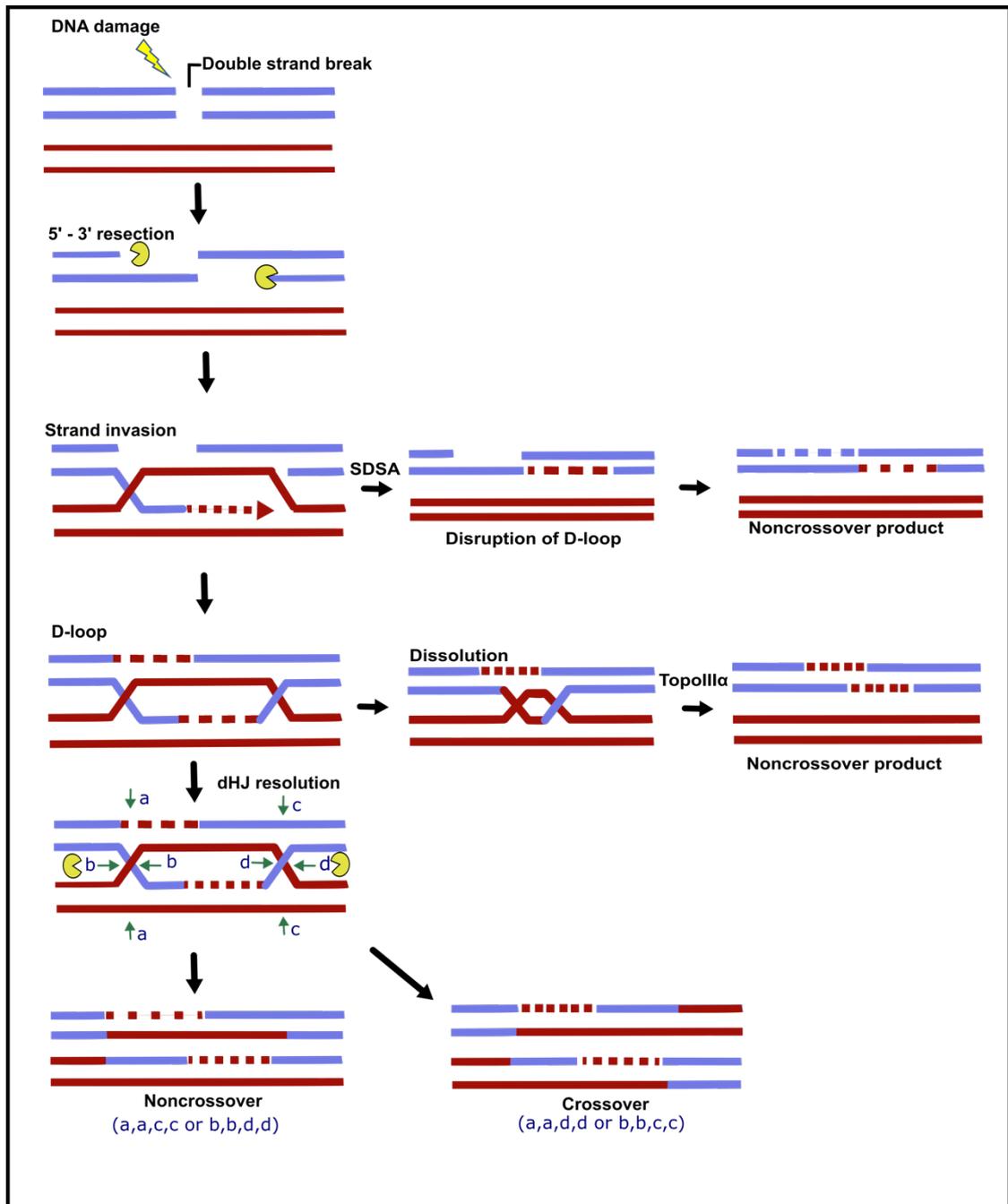


Figure 1.8: An overview of the double strand break repair and synthesis-dependent strand annealing models of homology-directed repair. DNA resection triggers the formation of 3' ssDNA overhangs. Strand invasion of DNA results in the formation of a D-loop. Disruption of D-loops promotes synthesis-dependent strand annealing, whilst successful second end capture favours

pathways of double-strand break repair. In dissolution, convergent migration of Holliday junctions forms a hemicatenane that is processed by topoisomerase III α to form a noncrossover product. Resolution of intermediates by Holliday junction resolvases results in either crossover products (cleavage combinations b and c or a and d) and non-crossover products (cleavage combinations b and d or a and c) resulting in either crossover or non-crossover products. Nucleases are depicted in yellow.

1.5.1.4. Synthesis-dependent strand annealing (SDSA)

Synthesis-dependent strand annealing is a recombinase-dependent repair pathway distinct from DSBR that results in exclusively non-crossover (NCO) products. SDSA has been well-studied and is considered the predominant mechanism of mitotic repair in budding yeast, but is poorly characterised within humans (221). In this model, limited DNA synthesis at the 3' termini of the invading strand occurs prior to displacement of the nascent DNA by collapse of the D-loop (209). Mph1/Fml1/FANCM have been associated with both releasing the invading strand to promote SDSA whilst also mediating the bypass of DNA lesions through regression of the replication fork (222). Nascent DNA anneals the opposite end of the DSB in the original strand, with DNA synthesis and ligation completing repair.

Signatures of SDSA include the presence and detection of gene conversion events, however an inability to distinguish between NCO arising from SDSA and other pathways of NCO DSB repair means little is known about mechanisms of annealing (223). Sequencing has noted that non-crossover events appeared to occur within clusters, and within spots that crossover events are frequent suggesting a shared mechanism (224). The SMARCAL1 ortholog Marcal1 in *Drosophila* has been posited as a candidate mediator of annealing, with mutants showing reduced levels of both SDSA and single strand annealing (SSA) mediated repair (225). Annealing in SDSA shares several similarities with RAD52-dependent SSA: however, the mild phenotype of human RAD52 mutants suggests that SDSA in human cells operates via a different mechanism (215). SDSA has been implicated within CRISPR/Cas9 gene editing, proposed as the primary repair mechanism for ssODNS: as a result further understanding of SDSA will help inform the use of CRISPR platforms (226, 227).

1.5.2. DNA recombination

Recombination of DNA is the process by which genetic information is exchanged between maternal and paternal chromosomes during meiosis. This is essential for correct chromosome segregation, with at least one crossover forming per chromosome within meiosis I (228). Meiotic recombination also promotes genetic diversity and adaptation by creating novel combinations of an organisms genetics; however it can also result in a loss of fitness by breaking up favourable allelic pairs – termed recombinational load (229, 230). Unlike in DNA repair pathways, meiotic recombination is initiated by programmed induction of double strand breaks within prophase 1 at many genetic loci by the topoisomerase-like Spo11 protein.

Spo11 remains bound to DNA and is subsequently released by local endonucleolytic cleavage, likely by Mre11, prior to resection as described within homology-directed repair (see section 1.5.1) (231). Unlike in mitosis, research within yeast implies Dna2 is not required for meiotic resection but may play a role in the removal of RPA-ssDNA filaments (232). Homology search, strand exchange and resolution of intermediates then proceed in a similar fashion as previously discussed within the homology-directed repair pathway however, in addition to Rad51, meiosis involves the meiosis-specific recombinase DMC1. DMC1 has previously been proposed to catalyse strand exchange between homologous chromosomes with RAD51 acting between sister chromatids and NCO recombination, however studies within Arabidopsis have demonstrated that in the absence of Rad51 no change in meiotic CO events were observed, suggesting that DMC1 is capable of catalysing the repair of all meiotic DSB into both CO and NCO events (233, 234).

As with homology-directed repair, the mechanisms by which DMC1 and RecA family members facilitate the genome-wide search for homology is still largely unclear but mapping of DMC1 and RAD51 binding on ssDNA in vivo has suggested that DMC1 is responsible for strand exchange within meiosis (235, 236). Recombination concludes through the formation and resolution of a double Holliday junction giving rise to cross-over products, or synthesis-dependent strand annealing (SDSA) which gives rise to non-crossover products. Resolvases, notably GEN1, EXO1, MUS81-EME1, SLX1-SLX4 and

MutLy (MLH1-MLH3 heterodimer), induce co-ordinated nicks within the junction, separating sister chromatids to produce crossover or non-crossover products (see figure 1.8). The action of MutLy (MLH1-MLH3 heterodimer) is believed to bias resolution towards crossover events, stimulated by EXO1 and PCNA, however the mechanisms of MutLy activity are still poorly understood (237, 238). Research within yeast has indicated a role for Exo1 within promoting crossovers through recruitment of Cdc5 kinase to MLH1-MLH3, with recent research suggesting Exo1 further promotes crossing over by protecting nicked DNA ends from ligation (239, 240). Full overviews of the known mechanisms of resolvases, including the formation of SMX trinuclease (SLX1-SLX4, MUS81-EME1 and XPF-ERCC1) complex, are available in detail elsewhere (241, 242).

Crossover recombination is essential for accurate chromosome segregation in meiosis. More thorough reviews of meiosis can be found elsewhere. In brief, crossovers arise prior to meiosis I at junctions between homologous chromosomes, physically represented in structures known as chiasmata, with the meiotic specific component Rec8 cohesin holding homologous chromosomes together: crossing over proximal to the centromere is rare, with frequent crossovers associated with missegregation (243). Upon onset of anaphase I, Rec8-cohesin complexes within the chromosome arms are phosphorylated prior to subsequent cleavage by separase to ensure separation of homologs. To ensure proper orientation and segregation of sister chromatids within meiosis II, cohesion proximal to the centromeres is protected from degradation by the Sgo1/PP2A protein complex: the phosphatase PP2A counteracts phosphorylation of Rec8, protecting cohesin from degradation (244, 245). This protection must subsequently be removed for chromatid segregation within anaphase II. It has been hypothesised tension-dependent relocation of Sgo1 plays a role in deprotecting cohesion, but the mechanisms of deprotection are not fully understood and additional factors are believed to be necessary (246, 247).

Only a limited number of double-strand breaks formed at the start of meiosis are repaired to form crossover events: in mice and most studied organisms, the number of crossovers is estimated to be at least ten-fold less than the number of DSB, with SDSA evidenced as a major pathway within meiotic NCO formation

(221, 248, 249). However, this can vary wildly between organisms, with *S. cerevisiae* exhibiting higher frequencies of crossing over: up to 90 crossover events for 150-170 programmed DSB events, with suggestions this may arise from weaker genetic interference (250, 251). The rate of crossing over appears to be tightly regulated through the principles of obligate crossover formation, whereby at least one crossover is formed between homologous chromosomes, whilst studies within a number of organisms, including yeast and mice, have observed evidence of phenomenon's of crossover interference and crossover homeostasis (252, 253). As crossover products are undesirable in mitotic DNA repair, resolvase activity is strictly limited to later stages in the cell cycle: SMX is primarily activated in prometaphase, whilst GEN1 is excluded from the nucleus and can only access DNA following nuclear envelope breakdown (254, 255).

1.5.3. Non-homologous end-joining

The second classically studied pathway of DSB repair, NHEJ is distinct from homology-directed repair as it does not require a homologous template to repair the DSB. Consequently the pathway is more error-prone, often resulting in insertion/deletion events at the lesion sites, and can occur at any stage within the cell cycle whilst HDR is limited to stages in the cell cycle where a template strand is available and when CDK1 (inactive within early G1 phase) is available, with studies suppressing CDK1 shown to inhibit DNA end resection (256, 257). CDKs role in DNA resection is briefly outlined in figure 1.7. Though NHEJ competes with HDR in the S and G2 phases of the cell cycle, suggesting a preference towards error-free repair whenever a sister chromatid is available: notably, the protein CYREN has been shown to act as an inhibitor of NHEJ within these phases by binding to Ku70/80 heterodimers to protect single stranded overhangs (258).

Mechanisms of NHEJ are largely determined by composition of the break, notably if the break is blunt ended or possesses 3'/5' overhangs: a representation of these mechanisms are outlined in figure 1.9. The initial step in all pathways is the binding of the Ku complex: a heterodimer consisting of Ku70 and Ku80 subunits with an exceptionally high affinity for DNA ends, which functionally acts to both protect DNA ends from nucleases and as a scaffold to

interact with other NHEJ factors (259, 260). In blunt-ended double-strand breaks, Ku promotes binding of XRCC4-DNA ligase IV to the site of the DSB, which joins the blunt ends together by direct ligation. XRCC4 promotes the activity and adenylation of DNA ligase IV, the latter of which is critical for ligation, and has been shown to interact with XLF to potentially bridge the broken DNA ends (261).

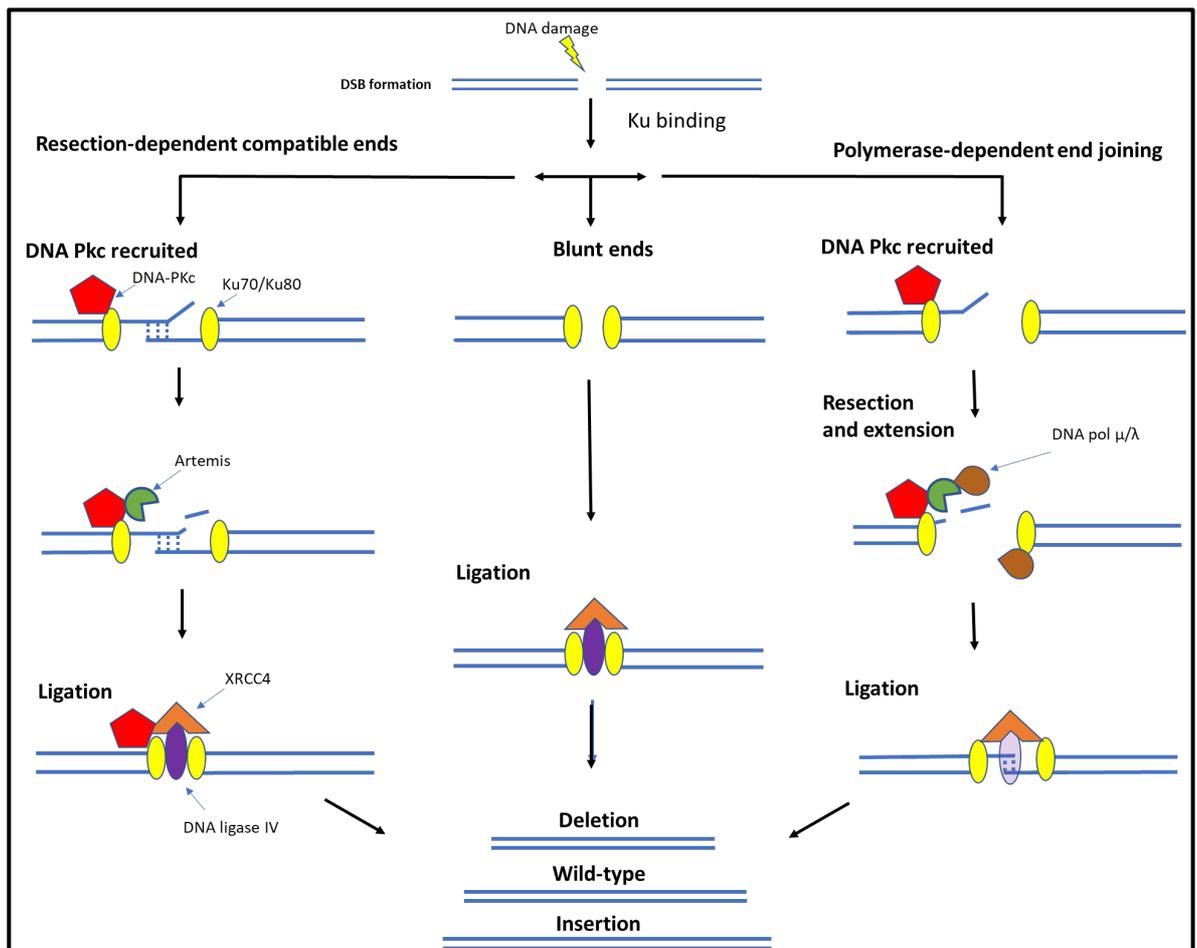


Figure 1.9: An overview of the non-homologous end joining pathway. End type dictates the repair process: blunt ends show a preference for no further end processing, resection-dependent ends with short regions of microhomology require resection by Artemis-DNA-PKcs whilst for some incompatible ends DNA polymerase μ and/or λ is recruited in combination with Artemis-DNA-PKc.

For incompatible ends the protein kinase DNA-PKc, which has a high affinity for Ku-DNA ends, is recruited in complex with the protein Artemis.

Autophosphorylation of DNA-Pkc upon binding activates Artemis endonuclease activity, enabling resection of incompatible 5' or 3' ends and facilitating end ligation by XRCC4-DNA ligase IV: it is thought phosphorylated DNA-PKc removes the self-inhibitory action of Artemis' C-terminus domain on the catalytic N-terminus domain (262). Dependent upon break complexity, data suggests that Polymerase μ and λ fills in gaps at junction sites and promote end annealing (259). In cases where end processing is not required, addition of DNA-PKcs, Artemis and Pol μ were not shown to promote end-joining, suggesting DNA prefers direct ligation to further processing (262).

1.6. Asgard Archaea

Living organisms have typically been classified according to the three domain model of life; consisting of bacteria, archaea and eukarya domains, with the latter two originally believed to possess a common ancestor as a result of similarities between archaea and eukaryotic enzymes, including those found within DNA replication/repair and ribosome biogenesis (263–266). Hypotheses of eukaryotic origins have recently shifted to the theory that eukaryotes arose from within the archaea clade rather than from a common ancestor; with the most promising candidate now commonly believed to have arisen from the recently identified Asgard lineage of Archaea (267, 268).

The Asgard superphylum of archaea were first formally recognised in 2015 with the identification of the *Lokiarchaeota*: the additional phylas of *Thor*-, *Odin*- and *Heimdallarchaeota* were identified soon thereafter (269, 270). Since then the group has been expanded to over 18 proposed phyla with high levels of diversity recorded between individual lineages (271, 272). Notably, differences have been recorded within metabolic pathways across the superphylum, with *Lokiarchaeota* and *Thorarchaeota* identified as likely to use organic compounds through the Wood-Ljungdahl pathway (WLP), whilst *Heimdallarchaeota* lack most of the enzymes necessary for WLP but later lineages encode respiratory oxidases and nitrate reductases not found within other Asgard groups (273).

Interest in Asgard archaea has grown as a basis of their evolutionary affinity with eukaryotes and identification of proteins originally thought to be signature to eukaryotes (274). As a result, Asgard archaea are under investigation as a

potential bridging point between prokaryotic and eukaryotic life, becoming key players within the two domain and eocyte hypotheses, which argues that eukaryotes originated from within the Asgard clade (268, 275). In particular, the *Heimdallarchaeota–Wukongarchaeota* group has previously been implicated as the closest bridging point between eukaryotes and the archaea based on protein sequence conservation, however it is possible that the eukaryotic ancestor resides in an archaea yet to be cultivated such as the recently discovered *Njordarchaeota* lineage (272, 276).

Strong evidence supporting this hypothesis originated upon phylogenomic analysis of *Lokiarchaeota* and the identification of complex membrane remodelling and trafficking systems similar to those found within eukaryotes (269, 277). These include the ESCRT subcomplex I and II, GTPase families and SNARE-like proteins; believed to be eukaryotic signature proteins (ESPs), these proteins were identified within Asgardarchaeota with archaeal SNARE-like proteins shown to interact with eukaryotic SNARE proteins (278–280). In addition, two supersized expansion segments — features thought to be unique to eukaryote ribosomes — were noted within *Lokiarchaeota* and *Heimdallarchaeota*, raising the possibility these arose before the last eukaryotic common ancestor (281).

Due to their extreme range of conditions, cultivation of strains of Archaea are notably difficult within the laboratory. It has been suggested that Asgard lineages are dependent upon symbiotic interactions for catabolism and anabolism, further complicating cultivation, thus much knowledge of Asgardarchaeota biology is based upon data from cultivation-independent metagenomic assemblies from a wide scope of sample locations (282). However, even high quality metagenome assemblies present low levels of contamination and studies have found that many proteins, including ribosomal proteins, fail to meet phylogenetic criteria, raising questions over the validity of ESPs (283, 284). As such cultivation of Asgard lineages or evidence that archaeal genes produce functional proteins operating similar to eukaryotes is still eagerly awaited to support the eocyte hypothesis (285).

1.6.1. MCP8718128.1

BLAST searches carried out on a member of this study, DDX49, against the Archaea domain identified a potential homolog of DDX49 within Asgardarchaeota. The identified sequence, MCP8718128.1, was a product of a metagenome assembled genome carried out within the yeast *Candida tropicalis*. At the time of identification, the study corresponding with this sequence had not yet been published. Jagadeeshwari et al (2023) observed that the average nucleotide identities of binned genomes within this metagenomic study appeared to correspond to an Asgard archaea within the *Heimdallarchaeota* branch.

As previously discussed, evidence that archaeal genes produce functional proteins functioning like eukaryotes is highly desirable for support of the eocyte hypothesis, indicating this still represents a novel area of research. As the scope of this study includes DDX49 and *Heimdallarchaeota* has been propositioned as the closest bridging point between archaea and eukaryotes, we included AA.49 identified within the scope of our study to investigate if it shared functional homology with recombinant *H.sapiens* DDX49.

1.7. Gene editing

A field that has flourished since the late 1900's, genome editing is the process by which biotechnological tools are used to manipulate the genetic material of an organism. Stemming from pioneering experiments using the I-SceI nuclease from *S. cerevisiae* within mice cells, key to modern gene editing is the ability to make targeted double-strand break within the DNA sequence in question, stimulating mutagenesis by exploiting the natural DNA repair pathway systems; notably those of non-homologous end joining and homology-directed repair (287, 288). Early gene editing systems took advantage of this and focused on targeted nucleases, such as zinc finger nucleases (ZFN) and TALEN systems. However, these technologies suffered several drawbacks making them inaccessible to many molecular laboratories: significantly, each target site required the custom design and engineering of an entire protein making them expensive and labour-intensive.

1.7.1. CRISPR-Cas9

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins comprise a system of adaptive immunity within prokaryotes whereby foreign mobile genetic elements (MGEs) are identified and memorised through three stages: adaptation, expression and interference. In brief, upon recognition of a short (2-4 bp) sequence known as the Protospacer-Adjacent Motif (PAM), Cas proteins cleave a portion of the DNA sequence (the protospacer) and integrate it between two repeats within the CRISPR locus in the form of a spacer. The array is later transcribed and matured into mature crRNA by Cas proteins which acts as a guide to recognise identical or similar sequences within other MGEs (289). The context of this study will focus on applications of CRISPR within gene editing, particularly Cas9-based editing: the molecular biology and other classes of CRISPR systems within prokaryotes can be found reviewed elsewhere (290–292).

Precision targeting of the gRNA-Cas9 complex was utilised in pioneering applications for genome engineering by the Doudna lab, who successfully applied a dual tracrRNA:crRNA system to induce targeted double-stranded breaks within DNA, and Zhang lab who adopted and successfully used this design in vitro within mammalian cell lines (293, 294). Studies since have predominantly applied CRISPR-induced double-strand breaks and the natural repair systems of NHEJ and HDR to facilitate gene editing. In NHEJ-based gene editing, a gRNA-CRISPR-Cas9 complex binds a genetic sequence complementary to the gRNA to induce a double-strand break and exploit the error prone nature of NHEJ (see section 1.5.3) to facilitate editing through disruption of key elements or frameshifts (295). NHEJ-based editing has been successfully applied within certain diseases, such as treatment of Duchennes muscular dystrophy in mice models and modelling of disease in cell models, however the random nature of generated indels means therapeutic applications are limited (296). By extension as indels generated are heterogenous in size, this can result in edits that remain in-frame making functional and careful characterisation of editing essential (297).

Homology-directed gene editing, often termed 'precise gene editing', is an approach whereby a 'donor' single stranded oligonucleotide (ssODN)

containing the desired sequence is introduced into the cell alongside the gRNA-CRISPR-Cas9 complex. This donor sequence possesses homology arms at both terminals that match the broken genomic DNA and can act as a template for the homology-directed repair pathway for integration into the broken DNA. Whilst this is naturally advantageous over NHEJ-based editing, HDR focused approaches are commonly met with relatively low efficacies, with notable issues include the restriction of HDR to the S and G2 phases of the cell and the two repair pathways acting in direct competition with each other (298).

Several strategies to improve the efficiencies of gene editing, particularly precise gene editing, have focused on exploiting regulatory factors to bias repair towards either pathway. These are reviewed extensively elsewhere and include examples of fusing Cas9 to other proteins, inhibition of key repair pathway proteins and cell cycle synchronisation (299). In addition to poor on-target efficiencies, editing is complicated by off-target effects: instances where the Cas9 cleaves in other areas of the genome than intended. In vitro this can result in inaccurate phenotypes and extensive delays to research, with outgrowth of monoclonal cell populations being time consuming, whilst the complications of off-target effects within a therapeutic setting could be severe. A variety of in silico tools have been developed to predict the frequency of off-target and on-target effects, including the web tool CRISPOR. CRISPOR provides an assessment of off-target effects using MIT specificity scores in addition to Doench and Moreno-Mateos scores to predict on-target efficiency (300). These scores are applied depending on the promotor used to express guides, such as the exogenous promotor U6 (Doench) and in vitro T7 promotors (Moreno-Mateos). Detailed outlines of each score can be found within their respective papers (301, 302) .

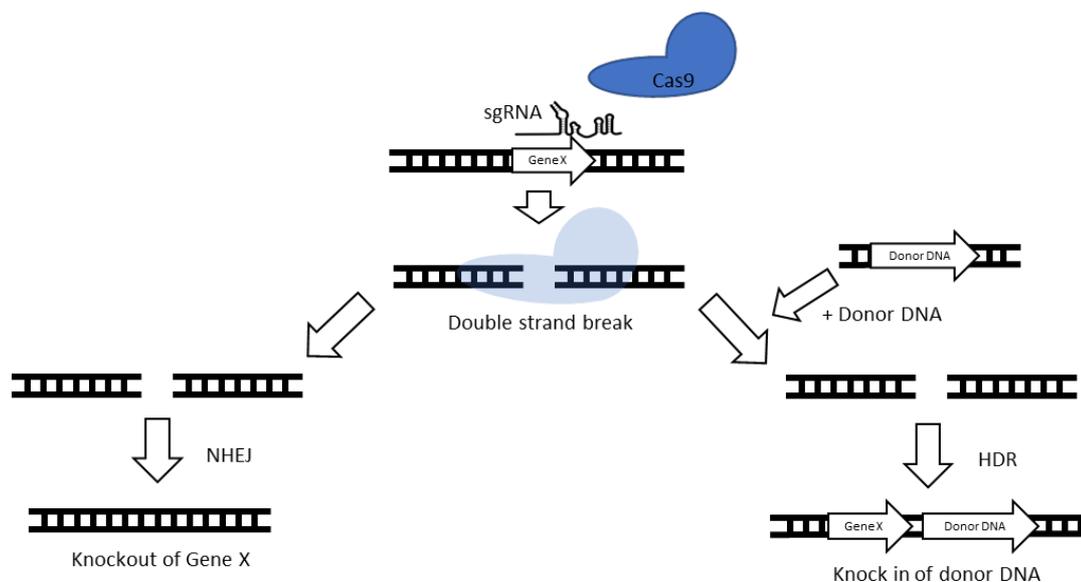


Figure 1.10: An illustration of CRISPR-Cas9 based gene editing. Cleavage by Cas9-gRNA complex induces a double-strand break within the DNA and the repair pathways of non-homologous end-joining and homology directed repair are manipulated to create a deletion or insert a piece of donor DNA, respectively. Figure adapted from Parkes et al (2021) (303).

1.7.2. Transposons, Base editors, and prime editors

Advances in CRISPR research resulted in development of further gene editing platforms, including Cas-fused transposases and base editors. In brief, Cas-fused transposases combine the actions of transposable elements with the precision targeting of gRNA-guided Cas enzymes to achieve targeted transposition. Base editors perform gene editing without inducing double-stranded breaks using a Cas9-nickase (Cas9n) fused with naturally occurring (CBE) and engineered (ABE) deaminase enzymes to deaminate the target nucleotide and create a single-stranded break within DNA. Deaminated nucleotides are recognised as noncanonical and are subsequently repaired through DNA repair mechanisms to several outcomes, including successful C/G – T/A (CBEs) and A/T – G/C (ABEs) transitions. A key limitation to base-editors has been their inability to generate precise base-edits outside of these transition mutations. Both transposases and base editors are reviewed in further detail elsewhere (304, 305).

To overcome these hurdles, a method referred to as prime editing was introduced by Anzalone et al (306). Prime editing operates similarly to HDR-based gene editing but includes an extension to the guide RNA (pegRNA) containing the desired template and utilises M-MLV reverse-transcriptase (RT) fused to Cas9n. This induces a nick, generating a 3' DNA flap which interacts with a 'primer binding site' at 3' terminal of pegRNA to act as a primer site for new DNA synthesis. M-MLV RT replicates the template and extends the 3' flap, displacing the unedited 5' flap which is cleaved by the nuclease FEN1. Mismatch repair proceeds and leaves either a precisely edited sequence or the original sequence, which Cas9n/pegRNA can rebind and reattempt editing.

Prime editing has already been successfully used in therapeutic models to treat disease, achieving 11.1% gene correction in treated mice with phenylketonuria and successfully restoring dystrophin expression with an efficiency of 52% in an iPSC cell model (307, 308). The technique has also seen several optimisations: Kim et al (2021) noted increased effectiveness using a 13-nt PBS, 12-nt RTT and a GC-rich PBS, though conflicting studies suggest optimum length of the PBS is sequence-dependent (310). Design of pegRNAs has been simplified with web tools such as multicrispr and pegfinder (311, 312). Despite the advantages of prime editing, it has been acknowledged that in cases with no bystander effects base editing invariably be more efficient than primer editing and other methods of precise editing (310).

1.8 Aims of the PhD project

As discussed within section 1.2.5, two poorly studied DExD-box helicases were identified of interest based upon homology with previously studied proteins and previously reported multi-functional roles of DExD-boxes outside of their candidate roles. Our research aims to investigate these helicases and identify novel roles outside of their predicted roles with ribosomal biogenesis. Candidates of interest include gaps in knowledge that exist within mechanisms of homology search and the emerging role of RNA species within processes of DNA repair. The primary objectives of this study therefore are:

- To biochemically characterise the activities of both DDX49 and DDX52 on DNA and (where possible) RNA species and identify any potential DNA replication and/or repair activities that it might function within.
- Create a CRISPR-Cas9 gene-editing pipeline for disruption of *DDX49* and *DDX52* genes in human cell lines and investigate phenotypes, implementing new assays where necessary.
- Investigate potential links between DDX49 and an Asgard archaea homolog discovered during bioinformatic studies.

Chapter 2: Materials and Methods

2.1. Materials

2.1. Chemicals and reagents

2.1.1. Reagents

Unless stated otherwise, the chemicals and reagents used within this study were purchased from a combination of the following companies: Merck/Sigma-Aldrich (Darmstadt, Germany) Thermo Fisher Scientific (Waltham, Massachusetts, US) and VWR international (Radnor, Pennsylvania, US). Nucleic acid preparation and extraction kits were purchased from Promega (Madison, Wisconsin, US). RNA extraction kits were purchased from QIAGEN (Hilden, Germany). Cell culture media was purchased from Lonza (Basel, Switzerland) whilst additional supplements and reagents were from Sigma-Aldrich unless otherwise stated.

2.1.2. Consumables

Unless stated otherwise all plasticware including serological pipettes, tissue culture flasks and plates used for tissue culture were purchased from the following companies: Corning (Corning, New York, US), Greiner (Kremsmünster, Austria), VWR international, Sarstedt (Nümbrecht, Germany) and Merck/Sigma-Aldrich.

2.1.3 Buffers and recipes

Table 2.1: Summary, recipes and associated uses of common laboratory buffers.

Buffer	Composition	Uses
10x TBE	890 mM Tris, 890 mM Boric acid, 20 mM EDTA	Preparation and running of agarose, native and denaturing PAGE gels
10x SDS running buffer	250 mM Tris, 1.92 M Glycine, 1 % SDS	Electrophoresis of SDS-PAGE gels

Chapter 2: Materials and Methods

4x SDS-PAGE loading dye	200 mM Tris-Cl (pH 6.8), 8% SDS, 0.4% bromophenol blue, 40% glycerol	Loading of samples for SDS-PAGE gels
Orange G loading dye	80% glycerol, Orange G powder	Loading of samples for native PAGE gels
Denaturing gel loading dye	75% formamide, 20 mM EDTA, 20% glycerol, Orange G powder	Loading of samples for denaturing PAGE gels
1x Coomassie blue stain	0.05% Brilliant Blue R-250, 10% acetic acid, 40% methanol	Protein staining of SDS-PAGE gels
Methylene blue stain	0.03% methylene blue, 0.3 M sodium acetate (pH 5.3)	Northern blotting
1x Destain buffer	10% acetic acid, 20% methanol	Destain of SDS-PAGE gels
5x Helicase buffer	100 mM tris pH 7.5 500 µg/mL 35% glycerol	EMSA and FRET + gel-based helicase assays
Proteinase K stop solution	5% SDS 200 mM EDTA 2 mg/mL proteinase K	Helicase and nuclease protection assays
10 x Annealing buffer	100 mM Tris-HCl pH 7.5 500 mM NaCl	Preparation of DNA and RNA substrates for assays
Elution buffer	4 mM Tris pH 8.0 10 mM NaCl	Elution of DNA and RNA substrates for assays
RIPA buffer	50 mM Tris-HCl pH 8 150 mM NaCl 1% NP-40	Lysis and extraction of proteins from cell culture

Chapter 2: Materials and Methods

	0.5% Sodium deoxycholate 0.1% SDS 1 tablet of protease cocktail inhibitor	
Cas9 working buffer	20 mM HEPES; 150 mM KCl, pH 7.5	Cas9 dilutions
Cas9 working reagent	20 mM HEPES 100 mM NaCl 5 mM MgCl ₂ 0.1 mM EDTA, pH 6.5	Cas9 activity assays
50 x HT buffer stock solution	1.5 M HEPES, 1.5 M triethanolamine	Northern blotting
RNA dye	2.1x HT buffer, 1 mM EDTA, 0.04% bromophenol blue	Northern blotting
20 x SSC	3 M NaCl 0.3 M sodium citrate	Northern blotting
50 x Denhardt's solution	1% (w/v) Ficoll, 1% (w/v) PVP, 1% BSA	Northern blotting
Hybridisation solution	5 x SSC, 5 x Denhardt's solution, 0.5% SDS	Northern blotting
Hybridisation wash solution	2 x SSC, 0.1% SDS	Northern blotting
Stripping solution	0.1 x SSC, 0.1% SDS	Northern blotting

Table 2.2: Summary and recipes of commonly used gels.

Gel type	Percentage	Composition
Native PAGE	10%	10 % acrylamide/bis-acrylamide (37.5:1) (v/v) 1 x TBE (v/v) 0.05 % APS (v/v) 0.125% TEMED (v/v)
	5%	5 % acrylamide/bis-acrylamide (37.5:1) (v/v) 1 x TBE (v/v) 0.05 % APS (v/v) 0.125% TEMED (v/v)
Denaturing PAGE	15%	15 % acrylamide/bisacrylamide (19:1) 1 x TBE (v/v) 7 M Urea (w/v) 5 % formamide (v/v)
SDS-PAGE separating	10%	10 % acrylamide/bis-acrylamide (37.5:1) (v/v) 0.42 M Tris pH 8.8 (v/v) 0.124 % SDS (v/v) 0.084 % APS (v/v) 0.1% TEMED (v/v)
SDS-PAGE stacking	5%	5 % acrylamide/bis-acrylamide (37.5:1) (v/v) 0.125 M Tris pH 6.8 (v/v) 0.1 % SDS (v/v) 0.1 % APS (v/v) 0.1 % TEMED (v/v)
Agarose	1%	1 % agarose (w/v) 1 x TBE buffer
	2%	2 % agarose (w/v) 1 x TBE buffer

Chapter 2: Materials and Methods

Denaturing agarose gel	1%	1 % agarose (w/v) 1 x HT buffer 1.3% formaldehyde (v/v)
------------------------	----	---

Table 2.3: A summary of the purification buffers used throughout this study and their composition.

	Composition	Used with
Ni-NTA buffer A	25 mM Tris pH 7.5, 25 mM imidazole, 500 mM NaCl, 10% (v/v) glycerol	DDX49, DDX52, AA.49
Ni-NTA buffer B	25 mM Tris pH 7.5, 400 mM imidazole, 500 mM, 10% (v/v) glycerol	DDX49, DDX52, AA.49
Ni-NTA buffer A (NLS-Cas9)	20 mM Tris pH 8, 20 mM imidazole, 250 mM NaCl, 10% (v/v) glycerol	NLS-Cas9
Ni-NTA buffer B (NLS-Cas9)	20 mM Tris pH 8, 400 mM imidazole, 250 mM NaCl, 10% (v/v) glycerol	NLS-Cas9
Heparin buffer A	25 mM Tris pH 7.5, 200 mM NaCl, 10% (v/v) glycerol	DDX49, DDX52, AA.49
Heparin buffer B	25 mM Tris pH 7.5, 1 M NaCl, 10% glycerol	DDX49, DDX52, AA.49
Heparin buffer A (NLS-Cas9)	20 mM HEPES-KOH pH 7.5, 150 mM KCl, 10% (v/v) glycerol	NLS-Cas9
Heparin buffer B (NLS-Cas9)	20 mM HEPES-KOH pH 7.5, 1 M KCl, 10% (v/v) glycerol	NLS-Cas9

Chapter 2: Materials and Methods

DDX49 storage buffer	25 mM Tris pH 7.5, 200 mM NaCl, 35% (v/v) glycerol, 2 mM DTT	DDX49, AA.49
DDX52 storage buffer	25 mM Tris pH 7.5, 200 mM NaCl, 35% (v/v) glycerol	DDX52
SEC buffer A	20 mM HEPES-KOH pH 7.5, 500 mM KCl, 10% glycerol (v/v) and 1 mM DTT	NLS-Cas9

2.1.4. Plasmids, oligonucleotides and gRNAs

Unless stated otherwise oligonucleotides used in cloning, PCR, sequencing and substrate preparation were ordered from and synthesised by Sigma-Aldrich. Oligonucleotides used in PCR and sequencing were ordered using desalt purification method whilst oligonucleotides used in substrate preparation and assays were purified using HPLC purification method. Backbone vectors were obtained from Addgene unless stated otherwise. pEGFP was a kind gift from Ronald Chalmers. Plasmids were cloned by the user stated and unless stated otherwise cloning protocols were designed by the author.

2.1.4.1. Plasmids

Table 2.4: A summary of the plasmids used within this study. Plasmids designated pAP were cloned by the author, pPS by student Philipp Springer, pKF by student Fiorela Kapllanaj, pLM by student Louise Martin and pAC by Dr. Andrew Cubbon.

Plasmid name	Description	Resistance
pETDuet-1	Sigma-Aldrich vector for protein co-expression containing two MCS with an incorporated His-tag and S-tag, respectively	Ampicillin
pUC19	Backbone vector for pUC19 used in Cas9 gRNA testing assays	Ampicillin

Chapter 2: Materials and Methods

pACYCDuet-1	Sigma-Aldrich vector for protein co-expression containing two MCS with an incorporated His-tag and S-tag, respectively	Chloramphenicol
pcDNA3.1-GFP	pcDNA3.1 with GFP incorporated for expression within human cells	Ampicillin, Neomycin
BPK1520	Addgene vector #65777 containing cassette for expression of SpCas9 sgRNA for prime editing	Ampicillin
pU6-pegRNA-GG-acceptor	Addgene vector #132777 containing cassette for expression of pegRNA for prime editing	Ampicillin
SpyCas9 PE2	Addgene vector #169850 for expression of SpyCas9 prime editor with optimised NLS	Ampicillin
pEGFP	GFP plasmid optimised for expression in human cells.	Ampicillin, Neomycin
pBad-mvenusN	pBad backbone with mvenus N-terminus inserted into NheI restriction sites	Ampicillin
pRSF1-mvenusC	pRSF1 backbone with mvenus C-terminus inserted into NcoI restriction sites	Kanamycin
pAP1	His- <i>DDX52</i> cloned into BamHI-HindIII pet-Duet His-tag reading frame	Ampicillin
pAP2	FLAG- <i>DDX49</i> cloned into BglII – KpnI pet-Duet MCS2 reading frame	Amp
pAP5	pAP1 with <i>DDX52</i> motif II sequence mutated to encode D318A, D321A	Amp
pAP6	<i>DDX49</i> .pET100D with <i>DDX49</i> sequence mutated to encode D152A/D155A	Amp

Chapter 2: Materials and Methods

pAP11	His-tagged AA.49–pET100 D, produced by Thermo Fisher GeneArt	Amp
pAP16	BPK1520 with nicking gRNA for <i>DDX49</i> D422A/D424A mutation	
pAP17	pAP1 with <i>DDX52</i> ORF mutated to encode F28_T128del (IDR)	
pPS2	N-terminus <i>DDX52</i> truncation in pET- Duet	Amp
pPS3	C-Terminus <i>DDX52</i> truncation in pET- Duet	Amp
<i>DDX49</i> .pET100D	<i>DDX49</i> in pET100D, GeneART construct	Amp
pAC29	Cas9 cloned into pACY-Duet with an inserted NLS	Chloramphenicol
pLM2	<i>DDX49</i> .pET100D with <i>DDX49</i> mutated to encode K421A	Amp
pKF1	<i>DDX49</i> .pET100D with <i>DDX49</i> mutated to encode D422/424A	Amp
pKF2	pU6-pegRNA-GG-acceptor with pegRNA for <i>DDX49</i> D422A/D424A	Amp

2.1.4.2. Oligonucleotides

Table 2.5: Oligonucleotides used within this study. Oligonucleotides designated RB were designed by Dr. Ryan Buckley and AP by the author.

ID	Alt name	Sequence	Function
RBX1	Hsa <i>DDX52</i> F (BamHI)	CAGTGCGGATCC GGACGTACATGA CC	Fw primer for PCR of <i>DDX52</i> gene for insertion into pET-Duet
RBX2	Hsa <i>DDX52</i> R (HindIII)	CAGTGCAAGCTTT TACGACTTATCCT CTAACGCG	Rev primer for PCR of <i>DDX52</i> gene for insertion into pET-Duet

Chapter 2: Materials and Methods

AP1	AP1	UAGCAAUGUAAU CGUCUAUGA	Construction of RNA fork 4
AP2	<i>DDX49</i> BamHI fw	GGATCCATGGCTA GCATGACTGGT	Cloning of <i>DDX49</i> into pAP4 BamHI/HindIII MCS
AP3	<i>DDX49</i> HindIII rev	GCAAGCTTTTATT GCTCAGCGGTGG C	Cloning of <i>DDX49</i> into pAP4 BamHI/HindIII MCS
AP4	52_AESA_F W	AGCGCGAAATTGT TCGAAGACGGTAA AAC	Fw primer for SDM of pAP1 with <i>DDX52</i> mutated to encode D318/321A
AP5	52_AESA_ REV	TTCCGCCACCACA AGCCACTCCAC	Rev primer for SDM of pAP1 with <i>DDX52</i> mutated to encode D318/321A
AP8	AP_49_FW	CAGTGCAGATCTG GATTATAAAGACG AT	Cloning of FLAG-tagged <i>DDX49</i> into pET-Duet BglIII/KpnI MCS
AP9	AP_49_RE V	CAGTGCAGTACC TTAAACTAAACC	Cloning of FLAG-tagged <i>DDX49</i> into pET-Duet BglIII/KpnI MCS
AP17	49_AEAA_F W	gcggcgCGTCTGCT GGAACAGGGT	Fw primer for SDM of pET100D-49 to encode D152/155A
AP18	49_AEAA_ REV	ttccgcCATAACCAG AAAGCGGATTTTT TTG	Rev primer for SDM of pET100D-49 to encode D152/155A
AP25	49_AEAA_F W2	GCAGCACGTCTG CTGGAACAGGGT	Fw primer for SDM of pET100D- <i>DDX49</i> to encode D152/155A
AP26	49_AEAA_ REV2	TTCTGCCATAACC AGAAAGCGGATTT TTTTG	Rev primer for SDM of pET100D- <i>DDX49</i> to encode D152/155A

Chapter 2: Materials and Methods

AP27	52_AESA_F W2	agcgcaAAATTGTT CGAAGACGGTAA AAC	Fw primer for SDM of pAP1 with <i>DDX52</i> mutated to encode D318A, D321A
AP28	52_AESA_ REV2	ttctgcCACCACAAG CCTACTCCAC	Rev primer for SDM of pAP1 with <i>DDX52</i> mutated to encode D318A, D321A
AP31	ORF9C_FW	GCGCCATATGTG GTCACATCC	Cloning of ORF9C GeneArt string into pET-Duet NdeI-KpnI MCS
AP32	ORF9C_RE V	CGCGGGTACCTT AATCGGTCAG	Cloning of ORF9C GeneArt string into pET-Duet NdeI-KpnI MCS
AP33	MW14_RNA _CY5	UCAUAGACGAUU ACAUUGCUCACAU GGAGCUGUCUAG AGGAUCCGA	Labelled for RNA binding assays and RNA/DNA hybrid assays
AP34	MW12_RNA	UCGGAUCCUCUA GACAGCUCCAUG AUCACUGGCACU GGUAGAAU	Complementary strand for RNA and DNA hybrid assays
AP37	49_APA_F W	ggcaCTGGAAGCA AAACGTAAAG	Fw primer for SDM of pET100D- <i>DDX49</i> to encode D422/424A
AP38	49_APA_R EV	ggtgcTTTACCTTCC AGGATCAG	Rev primer for SDM of pET100D- <i>DDX49</i> to encode D422/424A
AP41	52_DESD_ FW	gcagatAAATTGTTC GAAGACGGTAAAA C	Fw primer for SDM of pAP1 motif II
AP42	52_DESD_ REV	ttcatcCACCACAAG CCTACTCCAC	Rev primer for SDM of pAP1 motif II

Chapter 2: Materials and Methods

AP43	52_SEQ_D ESD	AATTA ACTCCCGT TTGTTAC	Sequencing primer for pAP1 motif II mutant
AP44	49_SEQ_D EAD	GTCAGCTGGGCC TGAAACAG	Sequencing primer for pET100D motif II mutant
AP51	52_EX4_F W	TGGAAGA ACTGCA GTGTGGG	Fw primer for amplification of <i>DDX52</i> exon 4 flanking region
AP52	52_EX4_RE V	ATGGGTTCAATGC CTGCCTT	Rev primer for amplification of <i>DDX52</i> exon 4 flanking region
AP53	52_EX5_F W	TGTCCAAGCAGG GCTATATT	Fw primer for amplification of <i>DDX52</i> exon 5 flanking region
AP54	52_EX5_RE V	ACCTTCCCTAGTG ATTGAACA	Rev primer for amplification of <i>DDX52</i> exon 5 flanking region
AP55	49_EX2_F W	CCCCAAATTCACG TGCTCCTGG	Fw primer for amplification of <i>DDX49</i> exon 2 flanking region
AP56	49_EX2_RE V	GTACCCTTTCTGC TGCCTGCC	Rev primer for amplification of <i>DDX49</i> exon 2 flanking region
AP57	49_EX4_fW	GAAGGAGGGATG TTCCAGGC	Fw primer for amplification of <i>DDX49</i> exon 4 flanking region
AP58	49_EX4_RE V	AACATTGCTAGGA CTGGGCC	Rev primer for amplification of <i>DDX49</i> exon 4 flanking region
AP59	49_EX9_F W	ATATCGCAGCTCA AGAGGCC	Fw primer for amplification of <i>DDX49</i> exon 9 flanking region

Chapter 2: Materials and Methods

AP60	49_EX9_RE V	AGGACAACTCCC CAACTTGC	Rev primer for amplification of <i>DDX49</i> exon 9 flanking region
AP62	45s rRNA PROBE	CGAGATGGAATG GATGG	Probe containing internal azide corresponding to 45s rRNA for Northern Blotting
AP63	5.8S rRNA PROBE	CGCAGCTACTTCT TGCG	Probe containing internal azide corresponding to 5.8s rRNA for Northern Blotting
AP66	<i>DDX52</i> - offtarget1fw	TCGTCGGCAGCG TCGCAGCAGCCT GGTTCTCGTGG	Fw primer for testing off- target sites of Cas9 KO
AP67	<i>DDX52</i> - offtarget1re v	GTCTCGTGGGCT CGGGCCGATGGC CTTCCCCACAC	Rev primer for testing off- target sites of Cas9 KO
AP68	<i>DDX52</i> - offtarget 2fw	TCGTCGGCAGCG TCTCCAATATCCG AGAAAGGAAACCT	Fw primer for testing off- target sites of Cas9 KO
AP69	<i>DDX52</i> - offtarget2re v	GTCTCGTGGGCT CGGTGCAACTGA ACACAATGGAAGT CA	Rev primer for testing off- target sites of Cas9 KO
AP70	<i>DDX52</i> - offtarget3fw	TCGTCGGCAGCG TCAGGCCACATAC CACCCAGCATCA	Fw primer for testing off- target sites of Cas9 KO
AP71	<i>DDX52</i> - offtarget3re v	GTCTCGTGGGCT CGGACCATGCAG CCTCAGTAGCTGC C	Rev primer for testing off- target sites of Cas9 KO
AP72	52_C_term_ fw	GGATCCACGGTC GAACAGGA	Cloning of C-terminus of <i>DDX52</i>

Chapter 2: Materials and Methods

AP73	52_C_term_ rev	GGAAGCTTTTACG ACTTATCCTC	Cloning of C-terminus of DDX52
AP76	52_N_TER M_FW	GGATCCGGACTG ACATG	Cloning of N-terminus of DDX52
AP77	52_N_TER M_REV	GGAAGCTTTTAGG CTTGCATTTGAA	Cloning of N-terminus of DDX52
AP92	N-term-fw	TAAGCAGGATCC GGACTGACATG	Fw primer for construction of N-terminus truncation of pAP1
AP93	N-term-rev	TAAGCAGGAAGCT TTTAGGCTTGCA TTGAA	Rev primer for construction of N-terminus truncation of pAP1
AP94	C-term-fw	AACAGGATCCAC GGTCGAACAGGA	Fw primer for construction of C-terminus truncation of pAP1
AP95	C-term-rev	AGCAGGAAGCTTT TACGACTTATCCT C	Rev primer for construction of C-terminus truncation of pAP1
AP96	52_DelbigDi p_fw	AGTGGTAAATTGG AGAATCTTCGCAA G	Fw primer for mutant with construction of F28_T128del pAP1 mutant
AP97	52_DelbigDi p_rev	AAAGCGCGCGGC ATC	Rev primer for mutant with construction of F28_T128del pAP1 mutant
AP98	52- Nterm_fw(2)	AAGCTTGCGGCC GCATAATGC	Fw primer for construction of N-terminus truncation of pAP1
AP99	52_N- term_rev(2)	TTAAGCGCTATTA CGCGCCCC	Rev primer for construction of N-terminus truncation of pAP1
AP106	49_PrimE_ DPD_fw	GTGGTGCGAAGA GAGTGTGA	Sequencing prime edits

Chapter 2: Materials and Methods

AP107	49_PrimeE_ DPD_rev	ACAACCTCCCGGT CATGTTC	Sequencing prime edits
AP122	AP33 DNA	TCATAGACGATTA CATTGCTACATGG AGCTGTCTAGAG GATCCGA	DNA version of AP33 for EMSA assays
AP137	FAM polyU(35)	UUUUUUUUUUUUUU UUUUUUUUUUUUUU UUUUUUUUUUUUUU	FAM-labelled RNA oligonucleotide for fluorescent polarisation assays
AP138	47s_fw	GCTGACACGCTG TCCTCTG	Oligo for RT-qPCR amplification of 47S rRNA
AP139	47s_rev	ACGCGCGAGAGA ACAGCAG	Oligo for RT-qPCR amplification of 47S rRNA
AP140	28s_fw	CCGCTGCGGTGA GCCTTGAA	Oligo for RT-qPCR amplification of 28S rRNA
AP141	28s_rev	TCTCCGGGATCG GTCGCGTT	Oligo for RT-qPCR amplification of 28S rRNA
AP142	18s_fw	CGGCGACGACCC ATTCGAAC	Oligo for RT-qPCR amplification of 18S rRNA
AP143	18s_rev	GAATCGAACCTG ATTCCCCGTC	Oligo for RT-qPCR amplification of 18S rRNA
AP144	beta- actin_fw	ACCACCATGTACC CTGGCATT	Oligo for RT-qPCR amplification of beta-actin
AP145	beta- actin_rev	CCACACGGAGTA CTTGCGCTCA	Oligo for RT-qPCR amplification of beta-actin
AP146	UBE2_Fw	TGCCTGAGATTGC TCGGATCT	Oligo for RT-qPCR amplification of UBE2
AP147	UBE2_rev	TCGCATACTTCTG AGTCCATTCC	Oligo for RT-qPCR amplification of UBE2
AP148	DDX52 qPCR_fw	GCTACATTTTCAGC AACTTGACCAG	Oligo for RT-qPCR amplification of DDX52

Chapter 2: Materials and Methods

AP149	DDX52 qPCR_rev	CTCGTGTGGTGA TATAATC	Oligo for RT-qPCR amplification of DDX52
AP150	5.8s rRNA_fw	CCTCGTACGACTC TTAGCGGT	Oligo for RT-qPCR amplification of 5.8S rRNA
AP151	5.8s rRNA rev	GCACGAGCCGAG TGATCC	Oligo for RT-qPCR amplification of 5.8S rRNA
sgpUC19	Guide RNA targeting pUC19	TAATACGACTCAC TATAGGGTGCTGC AAGGCGATTAAGT GTTTTAGAGCTAG AAATAGCAAGTTA AAATAAGGCTAGT CCGTTATCAACTT GAAAAAGTGGCA CCGAGTCGGTGC TT	Cas9 cleavage assays: gRNA sequence to target pUC19 plasmid

2.1.4.3. DNA substrates

Table 2.6: An overview of DNA substrates used within this study. All sequences are listed as 5' to 3'.

Name	Oligonucleotide name	Modification	5'-3' Sequence
Fork 2A	MW12	-	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTCGGC
	Cy5-MW14	5' Cy5	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA
Fork 2A (DNA:RNA)	MW12	-	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTCGGC

Chapter 2: Materials and Methods

	AP33	5'Cy5	UCAUAGACGAUUACA UU GCUACAUGGAGCUGUC UAGAGGAUCCGA
Fork 3	MW12	-	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTTCGGC
	Cy5-MW14	5' Cy5	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA
	PM16	-	TGCCGAATTCTACCAGT GCCAGTGAT
Fork 4	MW12	-	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTTCGGC
	Cy5-MW14	5' Cy5	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA
	PM17	-	TAGCAATGTAATCGTCT ATGACGTTG
FRET fork 2	MW12 Cy5	5' Cy5	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTTCGGC
	MW14 Cy3	3' Cy3	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA
FRET fork 3	MW12 Cy5	5' Cy5	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTTCGGC
	MW14 Cy3	3' Cy3	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA

Chapter 2: Materials and Methods

	PM16	-	TGCCGAATTCTACCGT GCCAGTGAT
FRET fork 4	MW12 Cy5	5' Cy5	TCGGATCCTCTAGACAG CTCCATGATCACTGGCA CTGGTAGAATTCCGGC
	MW14 Cy3	3' Cy3	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA
	PM17	-	TAGCAATGTAATCGTCT ATGACGTTG
DNA:DNA Annealing	ELB 40	-	GGAGCTCCCTAGGCAG GATCGTTCGCGACGATG GCCTTCGAAGAGCTCCA GTTACGGATACGGATCC TGC
	ELB 41	5'Cy5	GCAGGATCCGTATCCGT AACTGGAGCTCTTCGAA GGCCATCGTCGCGAAC GATCCTGCCTAGGGAG CTCC
DNA:RNA annealing	ELB 40R	-	GGAGCUCCCUAGGCAG GAUCGUUCGCGACGAU GGCCUUCGAAGAGCUC CAGUUACGGAUACGGA UCCUGC
	ELB 41	5'Cy5	GCAGGATCCGTATCCGT AACTGGAGCTCTTCGAA GGCCATCGTCGCGAAC GATCCTGCCTAGGGAG CTCC
FRET annealing	Cy5-ELB41	5'Cy5	GCAGGATCCGTATCCGT AACTGGAGCTCTTCGAA GGCCATCGTCGCGAAC

Chapter 2: Materials and Methods

			GATCCTGCCTAGGGAG CTCC
	Cy3-ELB40	3'Cy3	GGAGCTCCCTAGGCAG GATCGTTCGCGACGATG GCCTTCGAAGAGCTCCA GTTA CGGATACGGATCCTGC
D/R-loop (no invasion)	PM4	5'Cy5	GGGTGAACCTGCAGGT GGGCGGCTGCTCATCG TAGGTTAGTTGGTAGAA TTCGGCAGCGTC
	RGL19	-	GACGCTGCCGAATTCTA CCAGTGCCTTGCTAGGA CATCTTTGCCACCTGC AGGTTCACCC
R-loop (invasion, no flap)	PM4	-	GGGTGAACCTGCAGGT GGGCGGCTGCTCATCG TAGGTTAGTTGGTAGAA TTCGGCAGCGTC
	RGL19	-	GACGCTGCCGAATTCTA CCAGTGCCTTGCTAGGA CATCTTTGCCACCTGC AGGTTCACCC
	PM6 RNA	5' Cy5	AAAGAUGUCCUAGCAA GGCAC
R-loop (invasion, 3' flap)	PM4	-	GGGTGAACCTGCAGGT GGGCGGCTGCTCATCG TAGGTTAGTTGGTAGAA TTCGGCAGCGTC
	RGL19	-	GACGCTGCCGAATTCTA CCAGTGCCTTGCTAGGA CATCTTTGCCACCTGC AGGTTCACCC

	PM7 RNA	5'Cy5	AAAGAUGUCCUAGCAA GGCACGAUCGAGCGGA UAUCUAUGACCAU
D-loop (Invasion, no flap)	PM4	-	GGGTGAACCTGCAGGT GGGCGGCTGCTCATCG TAGGTTAGTTGGTAGAA TTCGGCAGCGTC
	RGL19	-	GACGCTGCCGAATTCTA CCAGTGCCTTGCTAGGA CATCTTTGCCACCTGC AGGTTACCCC
	PM6	5' Cy5	AAAGATGTCCTAGCAAG GCAC
D-loop (Invasion, 3' flap)	PM4	-	GGGTGAACCTGCAGGT GGGCGGCTGCTCATCG TAGGTTAGTTGGTAGAA TTCGGCAGCGTC
	RGL19	-	GACGCTGCCGAATTCTA CCAGTGCCTTGCTAGGA CATCTTTGCCACCTGC AGGTTACCCC
	PM7	5' Cy5	AAAGATGTCCTAGCAAG GCACGATCGAGCGGAT ATCTATGACCAT

2.1.5. CRISPR materials

2.1.5.1. Knockout guide design

crRNA and tracrRNA were synthesised by Integrated DNA Technologies (IDT). To increase the probability that edits would produce a greater impact upon the open reading frame, earlier exons were targeted. CRISPR guides were chosen using two online tools: IDTs Predesigned Alt-R™ CRISPR-Cas9 guide RNA (IDT, https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) and the CRISPR targets track on the human genome browser (313).

Oligonucleotides used for generation of amplicons of CRISPR on-target and off-target sites were designed using the CRISPOR tool (Concordet and Haeussler, 2018, <http://crispor.tefor.net/>).

Table 2.7: crRNA sequences used within this study. *DDX52* exons 5A and 5B are labelled as such due to close proximity.

Gene	Exon	Sequence	PAM
<i>DDX49</i>	2	AAGCTGTCTGAGGATCCCTA	TGG
<i>DDX49</i>	3	AAAGACTGCATCATCGTCGG	TGG
<i>DDX49</i>	4	GCTCTCTCGGAAACCACACG	TGG
<i>DDX52</i>	4	GCTTGCATTTGGATTGGCGT	AGG
<i>DDX52</i>	5A	TGGCTGGCAAGTTCTCGTGT	TGG
<i>DDX52</i>	5B	CTCGTGTTGGTGATATAATC	AGG

2.1.5.2 CRISPR prime editing guide design

Modification sites of interest were chosen based on motifs, notably motif II. For *DDX49*, an additional site of editing was chosen based upon a conserved site of interest noted within bioinformatic analysis in section 5.2.1. DNA regions flanking the edit site of interest were extracted from the human genome browser and oligonucleotides for the construction of prime editing plasmids designed using the online tool pegFinder (<http://pegfinder.sidichenlab.org>) with modifications inserted based upon the desired sequence (312, 313).

2.1.5.3. Prime editing oligonucleotides

Table 2.8: Oligonucleotides used within prime editing reactions.

Oligonucleotides designated SA were designed by student Sabesan Anandavijayan, AP by the author.

ID	Alt-name	Sequence	Purpose
SA5	49_APA_sgF	caccgTGCgCTTGGCCTCCAGG TCAgttttaga	Prime editing to mutate <i>DDX49</i> to
SA6	49_APA_sgR	tagctctaaaacTGACCTGGAGGCC AAGCGCAc	

Chapter 2: Materials and Methods

SA7	49_APA_scaff F	GCTAGAAATAGCAAGTTAAAT AAGGCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGTCG	encode D422A/D424A
SA8	49_APA_scaff R	GCACCGACTCGGTGCCACTTT TTCAAGTTGATAACGGACTAG CCTTATTTTAACTTGCTATTTC	
SA9	49_APA_exten sF	gtgcTGCCAGGCCCTGCCCTG GAGGCCA	
SA10	49_APA_exten sR	aaaaTGGCCTCCAGGGCAGGG GCCTGGCA	
SA11	49_APA_PE3_ sgF	caccGTGGCCAGGTTCCG	
SA12	49_APA_PE3_ sgR	aaacCCGCCAGGGAACCTGGC CAC	

2.1.6. Microbiology materials

2.1.6.1. Bacterial strains

Table 2.9: Strains of *E. coli* used within this study.

<i>E. coli</i> strain	Supplier	Genotype
DH5 α	Invitrogen	F- ϕ 80lacZ Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK- mK+) <i>phoA supE44</i> λ - <i>thi-1 gyrA96 relA1</i>
BL21-AI	Invitrogen	F- <i>ompT hsdS_B</i> (r _B ⁻ m _B ⁻) <i>gal dcm araB::T7RNAP-tetA</i>

2.1.6.2. Media and supplements

Table 2.10: Bacteria culture medium used within this study. All media was autoclaved before use and were used under aseptic technique.

Media	Recipe
Mu Broth	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, pH 7
Mu Broth with Agar	10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, 15 g/L Agar, pH 7

Table 2.11: Media supplements used within *E. coli* culture. All supplements were filtered using a 0.22 µm filter before use. L-arabinose was filtered and autoclaved.

Supplement	Stock concentration	Working concentration
Ampicillin	100 mg/mL	100 µg/mL
IPTG	1 M	0.5 mM
L-arabinose	20%	0.1%
Chloramphenicol	35 mg/mL	35 µg/mL
Kanamycin	50 mg/mL	50 µg/mL
Tetracycline	10 mg/mL	10 µg/mL

2.2. General methodologies concerning DNA and RNA

2.2.1. Materials, services and nucleic acid purification

Unless stated otherwise, all restriction enzymes, polymerases and DNA modification enzymes were obtained from New England Biolabs and reactions carried out according to manufacturer instructions (NEB, Massachusetts, USA). Sanger sequencing of plasmid DNA was carried out externally by Genewiz (Leipzig, Germany). Sanger sequencing of PCR amplicons within the CRISPR workflow were sequenced by the University of Nottingham Deepseq service (Nottingham, United Kingdom).

For subcloning work and preparation of high-quality plasmid stocks, plasmids were transformed into the DH5α strain of *E. coli* and extracted using Wizard® Plus SV Minipreps DNA Purification System (Promega) or ZymoPURE II Plasmid Maxiprep (Zymo Research, California, USA) according to manufacturer instructions. Gel extraction and PCR cleanups were carried out using Wizard® SV Gel and PCR Clean-Up System (Promega) according to manufacturer instructions.

2.2.2. Agarose gel electrophoresis

An appropriate percentage of agarose was dissolved in 1x TBE buffer through boiling and allowed to cool. Ethidium bromide (0.2 µg/mL) was added for nucleic

acid staining, solution mixed and gel poured evenly. Electrophoresis was carried out within 1x TBE buffer at 140 V for 60-90 minutes dependent on ideal band separation. DNA was visualised using a PCR and digests were imaged using UV detection with a U:Genius³ gel documentation system (Syngene, Bangalore, India); UV was substituted for Blue/White Light Transilluminator (Invitrogen) detection if gel extraction was required.

2.2.3. Polymerase Chain reaction

DNA was amplified using either Vent® or Q5® High-Fidelity polymerases (NEB): unless specifically stated otherwise either polymerase was used. The annealing temperature (x) of primer pairs was calculated using an online tool (NEB, Tm calculator, <https://tmcalculator.neb.com/#!/main>) and adjusted upon the polymerase used.

Vent PCR reactions were carried out in 1x Thermopol reaction buffer with 500 nM of each primer, 200 uM dNTPS, 50 ng DNA (unless stated otherwise) and 1 unit of polymerase. Unless stated otherwise, DNA was initially denatured at 95 °C for 5 minutes prior to 35 cycles of the following program: denaturation at 95 °C for 30 seconds, (x) annealing temperature for 30 seconds and an extension time of 1 minute per kilobase. A final extension was carried out at 72 °C for five minutes and sample stored at 4 °C until removal.

Q5 reactions were carried out in 1x Q5 reaction buffer with 500 nM of each primer, 200 uM dNTPs, 50 ng DNA (unless stated otherwise) and 1 unit of polymerase. Unless stated otherwise, DNA was initially denatured at 98 °C for 30 seconds prior to 35 cycles of the following program: denaturation at 98 °C for 10 seconds, (x) annealing temperature for 30 seconds and an extension time of 30 seconds per kilobases. A final extension was carried out at 72 °C for two minutes and sample stored at 4 °C until removal.

2.2.4. Site-directed mutagenesis

Primers for site-directed mutagenesis (SDM) were designed using an online tool (NEB, NEBaseChanger, <https://nebasechanger.neb.com>). PCR reactions were carried out as described in section 2.2.3 with the exception that 10 ng template DNA was used. Successful amplification was confirmed through visualisation of DNA bands in TBE-agarose gels stained with ethidium bromide.

Bands were purified using the Wizard® SV Gel and PCR Clean-Up System. To ensure template DNA was removed and ligate the SDM product, DNA was treated with 1 U of Dpn1 (R0176S, NEB), T4 polynucleotide kinase (M0201S, NEB) and T4 DNA ligase in 1 x ligase reaction buffer at 16 °C overnight. Reactions were transformed into the DH5 α strain of *E. coli* under appropriate antibiotic selection and plasmid DNA extracted using Wizard® Plus SV Minipreps DNA Purification System. Successful mutagenesis was confirmed by external sanger sequencing.

2.2.5. RT-qPCR

2.2.5.1. RNA extraction and cDNA synthesis

Total RNA was extracted from cell lines using RNeasy mini kit (Qiagen) according to manufacturer protocols. Synthesis of cDNA was performed using superscript IV (Thermo Fisher) and oligonucleotide (dT)20 (Thermo Fisher) according to manufacturer protocols in a GuardOne laminar flow cabinet (Starlab) where appropriate. Synthesised cDNA was subsequently incubated with 5 units of RNAase H (New England Biolabs) at 37 °C for twenty minutes to remove residual RNA from the sample.

2.2.5.2. qPCR

Sequences of the *Beta-actin*, *UBE2*, *GAPDH*, *DDX52* genes and the 47S, 28S, 18S and 5.8S rRNA sequences were retrieved and primers designed based on these sequences (see table 2.5). Beta-actin, UBE2 and GAPDH functioned as housekeeper controls. Amplifications were carried out in 20 μ L reaction solutions consisting of 1 x SYBR Green Master Mix, 500 nM of each primer and 5 ng of cDNA unless otherwise stated. PCR conditions were 95 °C for 40 seconds followed by 40 cycles of 95 °C for one second, 60 °C for 20 seconds and 95 °C for one second. Primer specificity was examined through melting curve analysis (95 °C for 1 s, 60 °C for 20 s and a rise in temperature to at a ramp rate of 0.15 °C/s). To confirm reproducibility, each assay was performed with technical triplicates for each of the three biological samples.

2.2.6. Annealing and purification of nucleic acids

2.2.6.1. Annealing of DNA:DNA substrates for gel-based and FRET assays

Unless otherwise stated, DNA:DNA substrates used within biochemical assays were prepared through incubation of 5 μM of labelled substrate with 6 μM of unlabelled oligonucleotide (s) in 1 x annealing buffer at 95 °C for ten minutes before cooling to room temperature overnight. Orange G loading dye was added to the reaction mix and DNA:DNA substrates were subsequently loaded onto a 10% native gel (1x TBE, 10% (v/v) polyacrylamide (37.5:1)) and migrated for three hours at 120 V. Gel band was excised and purified via diffusion in 250 μL elution buffer over 48 hours: if Cy5 was not visible via naked eye, then gel was imaged using Amersham Typhoon 5 biomolecular imager and gel slices extracted based upon extrapolation of migration distance. To maintain stability of substrate, no further concentration of the substrate was taken.

Concentration of substrate was determined through absorbance measurements at A260 using DeNovix DS-11 + spectrophotometer (DeNovix). Substrates A260 was measured and this value applied to the Beer-Lambert Law ($A=\epsilon cl$) whereby A corresponds to absorbance at A260, ϵ the substrates extinction coefficient, c is concentration and l is the optical path length. Respective extinction coefficients of substrates were calculated using the OligoAnalyzer™ online tool (Integrated DNA Technologies, <https://eu.idtdna.com/calc/analyzer>).

2.2.6.2. Annealing of substrates containing RNA

Unless otherwise stated, substrates containing RNA were prepared through incubation of 1 μM of labelled DNA/RNA with 2 μM of unlabelled DNA/RNA in RNase-free annealing buffer in a GuardOne laminar flow cabinet (Starlab) and heated at 95 °C for ten minutes before cooling for three hours. No further purification was performed. Substrates were freshly prepared each day, as needed.

2.2.6.3. Annealing of CRISPR substrates

CRISPR gRNAs used within knockout assays were prepared by creating a solution of 1 μM of Alt-R crRNA and Alt-R tracrRNA (IDT) in nuclease-free duplex buffer in a sterile cell-culture hood, prior to heating at 95 °C for five

minutes. Mixes were allowed to cool to room temperature for four hours prior to use.

2.3. General microbiology protocols

2.3.1. Growth and storage of *E. coli* strains

Solid cultures of *E. coli* were grown at either 37 °C overnight or for 72 hours at 30 °C and subsequently stored at 4 °C to minimise colony expansion. Liquid cultures were incubated overnight at 37 °C with shaking at 180 RPM. For small scale DNA purifications 5 mL of culture volume was inoculated and plasmids purified using a Wizard® Plus SV miniprep DNA purification system. For larger scale DNA purifications or where plasmids were required to be endotoxin-free, 150 mL or 300 mL of culture media was inoculated dependent upon the plasmids copy number and purified using a maxiprep kit according to manufacturer's instructions.

For long-term storage of cultures, cultures were supplemented with 20% glycerol (v/v) and flash frozen using dry ice. Stocks were subsequently stored at – 80 °C.

2.3.2. Preparation of chemically competent *E. coli*

Stored cell strains were streaked on to solid LB medium supplemented with appropriate antibiotics. Single colonies were picked and inoculated overnight in LB broth. *E. coli* culture was inoculated at a 1:100 dilution in fresh LB and cultured until OD₆₀₀ of 0.6. Culture was subsequently incubated on ice in ice-cold, filtered 0.1 M calcium chloride solution (CaCl₂) for thirty minutes prior to centrifugation at 4 °C for ten minutes at 4000 RPM. Pellet was carefully resuspended in ice-cold, filtered 0.1 M CaCl₂ solution and incubated for a further hour on ice, prior to centrifugation at 4 °C for 10 minutes at 4000 RPM. Pellet was resuspended in 85 mM ice-cold CaCl₂ supplemented with sterile glycerol added to a final concentration of 15% v/v. Aliquots were flash-frozen and stored at - 80 °C.

2.3.3. Transformation of chemically competent *E. coli*

Chemically competent *E. coli* was transformed using heat-shock protocol. In brief, 50-100 ng of plasmid DNA was added to 100 uL of competent cells and incubated on ice for five minutes. The cells were incubated at 42 °C for 90

seconds, prior to further incubation on ice for five minutes. LB was added to a final volume of 1 mL and cells were incubated for one hour at 37 °C for 60 minutes with shaking at 120 RPM.

2.4. Cloning and protein purification

All column chromatography was carried out using AKTA Start system (GE healthcare, Chicago, Illinois, USA) and columns were sourced from GE healthcare. All dialysis steps were performed using 8 kDa molecular weight cut-off (MWCO) dialysis tubing (Thermo Fisher) in at least 1:100 sample to buffer volume. Purified protein concentrations were determined using Bradford assay (see section 2.5.2) and via absorption readings at 280 nM using a DeNovix DS-11 + spectrophotometer (DeNovix). The protein A₂₈₀ was measured against a blank of storage buffer and this value applied to the Beer-Lambert Law ($A = \epsilon cl$) whereby A corresponds to absorbance at A₂₈₀, ϵ the proteins extinction coefficient, c is concentration and l is the optical path length. Respective extinction coefficients of proteins was determined using the ProtParam online tool (Swiss Institute of Bioinformatics, <https://web.expasy.org/protparam/>). It was assumed that all residues were reduced.

2.4.1. Cloning of AA.49, DDX49 and DDX52

2.4.1.1 Cloning of DDX52 and mutants for over-expression within *E. coli*

DNA sequences encoding the open reading frame of *H. sapiens* DDX52 were codon optimized for expression in *E. coli* and synthesized by GeneArt (Thermo Fisher) in string format with BamHI and HindIII sites incorporated upstream and downstream, respectively. DDX52 ORF was amplified via PCR, verified by agarose electrophoresis and successful PCRs purified using the Wizard® SV Gel and PCR Clean-Up System prior to digestion with BamHI-HF and HindIII-HF overnight at 37 °C. pETDuet-1 was digested likewise in parallel and the DDX52 ORF ligated overnight at 16 °C into the His-tag containing MCS1 of pETDuet-1 with 20 units of T4 DNA ligase. Ligation was subsequently transformed into the DH5 α strain of *E. coli* via heat-shock method (see section 2.3.3). Successful colonies were inoculated and plasmids miniprepmed using the Wizard® Plus SV Minipreps DNA Purification System, prior to testing via

analytical restriction digest using BamHI and HindIII and further verification via Sanger sequencing (Source Bioscience, Nottingham). Mutants of *DDX52* were Sanger sequenced by a separate company (Genewiz).

2.4.1.2. Cloning of FLAG tagged *DDX49* for over-expression within *E. coli*

DNA sequences encoding the open reading frame of *H. sapiens DDX49* were codon optimized for expression in *E. coli* and synthesized by GeneArt in string format with BglII and KpnI sites incorporated upstream and downstream, respectively. *DDX49* ORF was amplified via PCR, verified by agarose electrophoresis and bands purified using the Wizard® SV Gel and PCR Clean-Up System prior to digestion for one hour with KpnI-HF in Cutsmart® buffer at 37 °C. NaCl was spiked into reaction mix to a final concentration of 100 mM, BglII added and reaction incubated overnight at 37 °C. pETDuet-1 was digested likewise in parallel and the *DDX49* ORF ligated into MCS site 2 overnight at 16 °C. Ligation mixture was subsequently transformed into the DH5α strain of *E. coli*. Successful colonies were inoculated and plasmids purified using Wizard® Plus SV Minipreps DNA Purification System prior to verification analytical restriction digest using NdeI-HF and KpnI-HF and Sanger sequencing (Source Bioscience). Following practical difficulties in purification of FLAG-tagged *DDX49*, a readily prepared His-tagged construct in pET-100D was obtained from GeneART. Further methodologies associated with the FLAG-tagged purification will not be detailed further within this review.

2.4.1.3. Cloning of His-tagged *DDX49* and AA.49 for overexpression in *E. coli*

The DNA sequences for the ORF of *DDX49* and MCP8718128.1 (hereafter referred to as AA.49) were codon-optimized for expression in *E. coli* and were ordered from GeneArt readily incorporated into pET-100D plasmid with an incorporated His-tag. Respective plasmids were transformed into DH5α *E. coli* via heat-shock method (see section 2.3.3). Colonies were inoculated and plasmids purified using the Wizard® Plus SV Minipreps DNA Purification System prior to assessment via analytical restriction digest using NdeI-HF and EcoRI-HF and further validation by Sanger sequencing (Genewiz).

2.4.2. General protocol for protein over-expression within *E. coli*

Prior to overexpression, plasmids expressing the protein of interest were transformed into the BL21-AI strain of *E. coli* (see section 2.3.3) and plated onto LB-agar with appropriate antibiotic. Single colonies were picked and inoculated in 25 mL LB supplemented with appropriate antibiotic overnight at 37 °C. Overnight culture was subsequently inoculated 1:100 in 2 L of fresh LB supplemented with antibiotic and grown at 37 °C until reaching an OD₆₀₀ of 0.6. Culture was further supplemented with 0.5 mM IPTG and 0.1% (w/v) L-arabinose prior to incubation with shaking overnight at 16 °C. Biomass was subsequently clarified at 4500 RCF for 10 minutes at 4 °C in an Avanti J-26 XP centrifuge prior to resuspension in appropriate lysis buffer supplemented with 1 mM PMSF. If biomass was not to be used immediately, it was flash frozen and stored at - 80 °C.

2.4.3. General protocol for protein purification

For overexpression, see section 2.4.2. Clarified biomass was resuspended in Ni-NTA buffer A (25 mM Tris pH 7.5, 500 mM NaCl, 25 mM Imidazole, 10% glycerol) supplemented with 1 mM PMSF. CaCl₂ and DNase1 were added to final concentrations of 5 mM and 1 µg/mL, respectively. Sample was incubated on ice for thirty minutes prior to lysis with a Vibra Cell VC-50 sonicator (Sonics & Materials, Newtown, Connecticut, USA). Lysed samples were clarified by centrifugation using an Avanti J-26 XP centrifuge at 35,000 RCF for 35 minutes. Pellet was discarded and supernatant loaded onto a 5 ml HisTrap HP His-tag column (GE Healthcare) at 1 mL per min. Flowthrough was collected for analysis and column washed with Ni-NTA buffer A until UV signal baseline. Bound DDX52 was eluted in Ni-NTA buffer B (25 mM Tris pH 7.5, 500 mM NaCl, 400 mM Imidazole) using a gradient of increasing imidazole up to 100%. DDX52 positive fractions were determined by SDS-PAGE, pooled and dialysed overnight at 4 °C into heparin buffer A (25 mM Tris pH 7.5, 200 mM NaCl, 10% glycerol). Sample was clarified at 5000 RCF for ten minutes at 4 °C to remove precipitate prior to loading on a 1 ml HiTrap™ Heparin HP column (GE Healthcare). Flowthrough was collected and column washed in heparin buffer

A until UV signal baseline. Proteins were eluted in a gradient of increasing ionic strength of heparin buffer B (25 mM Tris pH 7.5, 1 M NaCl, 10% glycerol). Fractions containing proteins of interest were determined via SDS-PAGE, pooled and dialysed overnight into respective storage buffer. Purified protein was aliquoted and stored at -80 °C for future use.

2.4.4. Streptococcus pyogenes recombinant NLS-Cas9

2.4.4.1. Cloning

Plasmid pAC29 was previously prepared by Andrew Cubbon and Liam Chau. Plasmid was transformed into the DH5a strain of *E. coli* by heat-shock method and plated onto LB-agar plates supplemented with chloramphenicol (see section 2.3.3.). Colonies were inoculated and plasmids purified using the Wizard® Plus SV Minipreps DNA Purification System. Successful plasmid preparation was confirmed via analytical digest and subsequent Sanger sequencing (Source Bioscience).

2.4.4.2. Purification

For overexpression, see section 2.4.2. NLS-Cas9 protein was purified using a His-affinity tag and based upon a method published by Anders et al (314). Culture was resuspended in Ni-NTA buffer A (20mM Tris-HCl pH 8.0, 250 mM NaCl, 20 mM Imidazole, 10% (v/v) glycerol)) supplemented with a protease inhibitor cocktail tablet. CaCl₂ and DNase1 were added to final concentrations of 5 mM and 1 µg/mL, respectively, and lysate incubated on ice for 30 minutes. Suspension was subsequently lysed using a Vibra-Cell VC 50T sonicator and clarified by centrifugation using an Avanti J-26 XP centrifuge at 35,000 RCF, 4 °C for 35 minutes. Supernatant was loaded in Ni-NTA buffer A onto a HisTrap HP His-tag column (GE Healthcare) and was eluted in Ni-NTA buffer B (20 mM Tris-HCl pH 8.0, 250 mM NaCl, 400 mM Imidazole, 10% glycerol (v/v)) up to a gradient of 100%. Fractions containing absorbance peaks of note were analysed using SDS-PAGE and Cas9 positive fractions pooled and dialysed overnight at 4 °C in low salt containing Buffer A (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 10% (v/v) glycerol, 1 mM DTT). Dialysed sample was clarified to remove any precipitate prior to loading on a 1 ml HiTrap™ Heparin HP column (GE Healthcare Life Sciences). Protein was eluted in heparin buffer B (20mM HEPES-KOH pH 7.5, 1 M KCl, 10% (v/v) glycerol, 1mM Dithiothreitol (DTT)) up

to a gradient of 100%. Fractions of interest were analysed by SDS-PAGE before being pooled and loaded on to a HiPrep 16/60 Sephacryl S-300 HR (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES-KOH pH 7.5, 500 mM KCl, 10% glycerol (v/v), 1 mM DTT). His-NLS-Cas9 containing fractions were identified via SDS-PAGE, pooled and concentrated using a centrifugal concentrator (Pierce, 100 kDa MWCO). Concentrated NLS-Cas9 was aliquoted, flash frozen and stored at -80 °C for long-term storage.

2.5. General assays

2.5.1. SDS-polyacrylamide-gel electrophoresis (SDS-PAGE)

Protein samples (prepared in 4 x SDS-PAGE loading dye with 0.1 M DTT) were incubated at 95 °C for 10 min prior to loading and separation on a 10% SDS-PAGE gels (see table 2.2) at 140 V until dye front reached the end of the gel. Unless used in western blot or mass spectrometry analysis, gels were rinsed in SDW prior to staining in Coomassie blue stain with rocking for fifteen minutes. Coomassie stain was decanted, gel rinsed further with SDW and subsequently incubated in destain solution until background stain was removed. Protein sizes were determined by comparisons of protein bands of interest with a molecular weight standard (Color Prestained Protein Standard, Broad Range, NEB).

2.5.2. Bradford Assay

Purified protein concentrations were determined using Bradford reagent (B6916, Sigma-Aldrich) using the microplate procedure according to manufacturer instructions. In brief, protein standards at concentrations of 0.05 - 0.5 mg/mL were prepared using bovine serum albumin (BSA) in appropriate protein storage buffer. Samples were measured neat and at dilutions of 1:2 and 1:5. Protein standards and samples were mixed with Bradford Reagent as instructed and incubated for two minutes prior to reading at 595 nM on a FLUOstar Omega (BMG Labtech, Offenburg, Germany). Prism software was used to generate a standard curve and interpolate sample points. Standards and samples were measured in duplicate.

2.5.3. Bicinchoninic acid assay

Total protein from complex mixtures was determined using Pierce™ BCA Protein Assay Kit (Thermo Scientific) according to the microplate procedure and

manufacturer's instructions. In brief, protein standards were prepared at a range from 25 – 2000 µg/mL using kit provided albumin standards. Samples were diluted appropriately and samples and BCA working reagent was added and incubated for 30 minutes at 37 °C. Absorbance was measured at 562 nm using a FLUOstar Omega microplate reader (BMG labtech, Ortenberg, Germany). A standard curve was generated using Prism software and used to interpolate sample concentrations. Samples and standards were measured in duplicate.

2.5.4. Mass Spectrometry analysis

SDS-PAGE gels were run as described in section 2.5.1 and stained in SYPRO ruby stain (Bio-Rad) according to manufacturer instructions. In brief; gel was covered overnight in 50 mL of ruby stain with rocking, prior to rinsing in a 10% methanol:7% acetic acid solution for 60 minutes. Gel was washed in SDW, imaged using a Blue/White Light Transilluminator and bands of interest extracted carefully. Gel slice was analysed externally using LC/MS by the Cambridge Centre for Proteomics (University of Cambridge, Cambridge, United Kingdom).

2.6. Biochemical assays

2.6.1. Gel-based unwinding assays

Helicase unwinding assays were carried out using 25 nM Cy5 DNA substrate in 1 x helicase assay buffer (20 mM Tris pH 7.5, 7% glycerol, 100 µg/mL BSA) supplemented with 5 mM ATP, 5 mM MgCl₂ and 25 mM DTT unless otherwise stated. Helicases were added at the desired concentration and reactions incubated for thirty minutes at 37 °C prior to the addition of stop mix. Reactions were migrated on a 10% polyacrylamide gel for 60 minutes at 140 V prior to imaging on an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30).

2.6.2. FRET-based unwinding assays

FRET-based helicase assays were carried out using 50 nM Cy5 DNA substrate in 1 x helicase assay buffer (20 mM Tris pH 7.5, 7% glycerol, 100 µg/mL BSA) supplemented with 5 mM ATP, 5 mM MgCl₂ and 10 mM DTT unless otherwise stated. Protein was spiked in immediately prior to reading and reactions analysed at 37 °C for thirty minutes on FLOUstar Omega (BMG-Labtech) with

readings at wavelengths of 590 nM and 680 nM taken every minute. Data was analysed by dividing FRET signals by Cy3 signal and normalising against a no protein control. Threshold gains were set before the addition of protein through use of 50 nM fully annealed FRET substrate and Cy3-only control wells.

2.6.3. Gel-based annealing assays

Helicase annealing assays were carried out using 15 nM Cy5 DNA substrate in 1 x helicase assay buffer (20 mM Tris pH 7.5, 7% glycerol, 100 µg/mL BSA) supplemented with 25 mM DTT. Helicases were preincubated with the DNA/RNA oligonucleotide of interest for five minutes at 37 °C, prior to addition of 15 nM of a complementary oligonucleotide. Reactions were incubated for thirty minutes at 37 °C prior to the addition of stop mix, unless stated otherwise. Reactions were migrated on a 10% TBE for 60 minutes at 140 V prior to imaging on an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30).

2.6.4. FRET-based annealing assays

FRET-based annealing assays were carried out using 50 nM Cy5 DNA substrate in 1 x helicase assay buffer (20 mM Tris pH 7.5, 7% glycerol, 100 µg/mL BSA) supplemented with 1 mM ATP, 1 mM MgCl₂ and 5 mM DTT unless stated otherwise. Helicases were added to the wells and pre-incubated for two minutes prior to addition of 50 nM Cy3-labelled substrate. Reactions were then analysed at 37 °C for thirty minutes on FLOUstar Omega with readings at wavelengths of 590 nM and 680 nM taken every minute. Data was analysed by dividing FRET signals by Cy3 signal and normalising against a fully annealed substrate. Threshold gains were set before addition of Cy3-labelled substrate through use of 50 nM fully annealed FRET substrate and Cy3-only control wells.

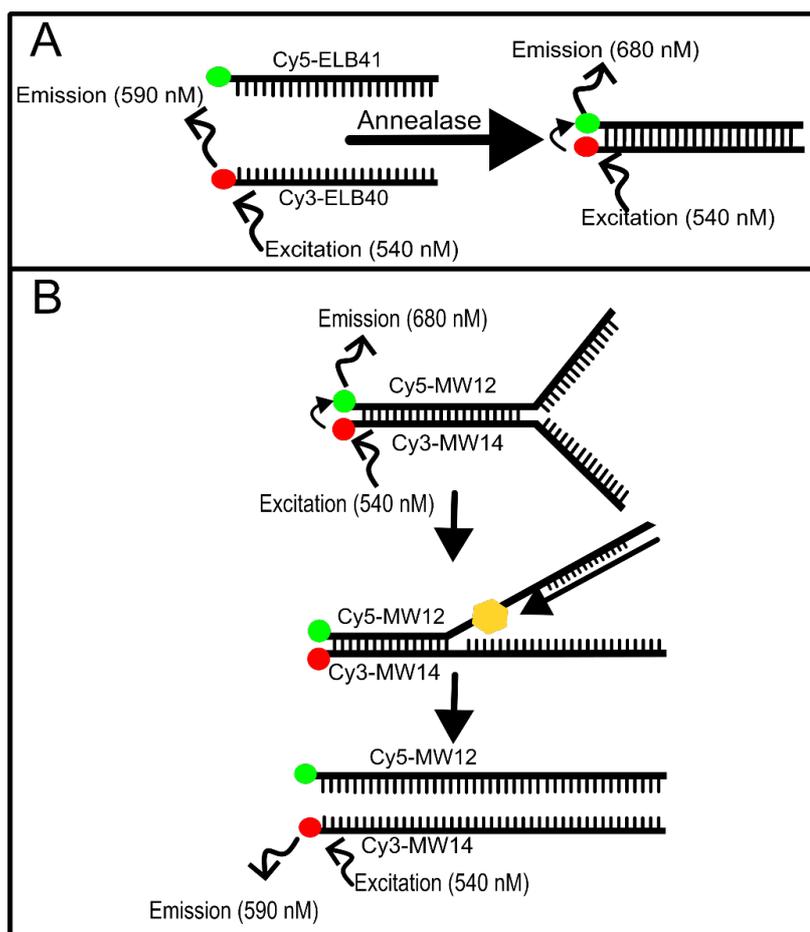


Figure 2.1: An overview of the FRET assays carried out in this study, using Cy3 and Cy5 as an acceptor and donor pair, respectively. A) Following annealing of oligonucleotides, Cy3 and Cy5 labelled oligonucleotides are brought into close proximity and upon excitation at 540 nm emit at a wavelength of 680 nm. B) Following unwinding of the forked DNA duplex, Cy3-labelled strand is liberated and a signal is emitted at 590 nm.

2.6.5. Electromobility shift assays

Interactions between proteins and DNA as part of a nucleoprotein complex was studied using electromobility shift assays (EMSAs). EMSAs were carried out using 25 nM Cy5-labelled substrate in 1 x helicase assay buffer supplemented with 25 mM DTT. Helicases were incubated with substrates of interest for thirty minutes, prior to addition of orange G loading dye. Samples were migrated on a 5% native polyacrylamide gel at 140 V for ninety minutes. In some cases, the gel was run at 4 °C for an extended time. Imaging was carried out using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30).

2.6.6. ATPase assays

ATPase activity was measured using BIOMOL® Green assay kit (Enzo life sciences) in accordance with manufacturers guidelines. In brief, phosphate standard concentrations were prepared ranging from 0.625 – 40 μ M. Samples were prepared and helicase unwinding reactions carried out as described in section 2.6.1. Reactions were terminated through addition of BIOMOL® Green Reagent and plates were incubated for thirty minutes at room temperature prior to absorbance measuring at OD_{620nm} in a FLUOstar Omega plate reader. A standard curve was generated using Prism software through which sample concentrations were interpolated. Assays were carried out in triplicate and appropriate controls used to detect possible phosphate contamination. All ATPase assays were performed using non-autoclaved materials as advised in manufacturer instructions.

2.6.7. Nuclease protection assays

Nuclease assays were optimised according to the nuclease used. Assays were carried out using 25 nM Cy5-labelled substrates in 1 x Helicase buffer supplemented with 25 mM DTT. Increasing concentrations of protein were pre-incubated with DNA for ten minutes at 37 °C, prior to the addition of 1 x nuclease buffer and twenty units of nuclease. Reactions were incubated at 37 °C for thirty minutes prior to addition of stop mix and denaturing loading dye. Reactions were migrated on a 15% denaturing polyacrylamide gel for 180 minutes at 100 W prior to imaging on an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30).

2.6.8. Fluorescence polarisation

Fluorescence polarisation (FP) assays were performed using 40 nM of fluorescein-labelled DNA/RNA in 1 x helicase assay buffer (20 mM Tris pH 7.5, 7% glycerol, 100 μ g/mL BSA) supplemented with 10 mM DTT. Concentrations of protein were spiked in, incubated at 37 °C for ten minutes and analysed on a FLUOstar Omega plate reader. The R-value was determined according to the following equation:

$$R = \frac{I_a - I_b}{I_a + 2I_b}$$

A no protein control was included and the average R for this was subtracted from those containing protein. Data was analysed using PRISM software and dissociation constant (Kd) values assessed using a curve fit based on a one site specific binding model.

2.7. Human cell culture

2.7.1. Human cell lines, media and supplements

Table 2.12: Human cell lines used within this study. U2OS-derived cell-lines were a kind gift from Richard Wood (MD Anderson Cancer Centre, University of Texas, Texas, US).

Cell line	Information
U2OS	Human osteosarcoma cell line reported to be hypertriploid from a moderately differentiated sarcoma of the tibia of a 15 year old female.
<i>DDX49 +/-</i>	U2OS cell line modified by CRISPR editing to produce a heterozygous gene edit of <i>DDX49</i>
<i>DDX52 +/-</i>	U2OS cell line modified by CRISPR editing to produce a heterozygous gene edit of <i>DDX52</i>

Table 2.13: A table of DMEM supplements used within human cell culture.

Supplements were aliquoted and stored at -20 °C and filtered prior to use.

Supplement	Working concentration
L-glutamine	2 mM
Penicillin-streptomycin	100 I.U/mL Penicillin, 0.1 mg/mL Streptomycin
Foetal bovine serum	10%

When new media solutions were constituted, 50 mL of stock medium (Serum-free media) was removed, aliquoted and stored at 4 °C prior to addition of supplements. L-glutamine was supplemented into media and supplemented

media replaced every six weeks to prevent formation of toxic by-products. Supplements were aliquoted and stored at -20 °C.

2.7.2. Routine growth and passaging of cells

All cell lines were cultured in DMEM media under conditions of 37 °C, 5% CO₂. For initial seeding, cryovials were retrieved from storage in vapour-phase liquid nitrogen and incubated briefly within a 37 °C water bath until approximately 50% of the solution was defrosted. Cells were transferred to a sterile 50 mL tube and 19 mL of appropriate media was gently added to dilute DMSO present within the freezing solution. Cells were pelleted via centrifugation at 350 x g for ten minutes. Supernatant was decanted and cells were resuspended in 5 mL media prior to being transferred to a T25 flask. Upon reaching a confluency of ~80%, cells were transferred to a T75 flask.

Cells were routinely passaged upon reaching 80% confluency. Media was aspirated from cells and the cell layer washed twice with phosphate buffered saline (PBS). Cells were incubated within 1 x Trypsin/EDTA for 5 minutes, shaken and incubated for a further one minute. Trypsin was quenched with addition of 10 x volume of media. For routine passages, cell split ratio was based on percentage confluency. Alternatively, cell number was assessed by mixing a small volume of cells at a 1:1 ratio with 0.4% Trypan blue (Thermo Fisher) and a cell count performed using a haemocytometer (Marienfeld, Lauda-Königshofen, Germany). Cell lines were passaged for a maximum of ten passages after thawing prior to discarding.

For cryopreservation of cell lines, cells were seeded into a 10 cm dish and cultured to ~80% confluency. Cells were then passaged and counted as noted above. Cells were spun down at 350 g for 10 minutes, resuspended in cryopreservation solution (80% complete medium, 10% DMSO) and aliquoted into cryotubes at a density of 1.5–2 x 10⁶ cells/mL. Tubes were frozen at -80 C for 24 hours prior to transfer to vapour-phase liquid nitrogen.

2.7.3. Lysis of cultured cells

As proteins of interest were localised in the nucleus, cells were lysed in RIPA buffer. Briefly, cells were trypsinised and resuspended as normal. A cell count was performed and cells spun down at 350 RCF for 10 minutes at 4 °C. Cells

were washed twice in 1 x PBS and pellet resuspended in mL ice-cold RIPA buffer at a ratio of 1 mL , prior to incubation on ice for twenty minutes with vortexing every four minutes. Lysed cell suspension was centrifuged at 12,000 RPM for 20 minutes at 4 °C, supernatant aliquoted and stored at -80 °C for downstream use. Total protein concentration was determined by BCA assay (see section 2.5.3).

2.8. General cell culture assays

2.8.1. WST-1 viability assays

Cell proliferation and cytotoxicity assays were performed using WST-1 cell proliferation reagent (Sigma-Aldrich) according to the manufacturer's guidelines. In brief, 4000 cells per well were seeded into 96 well plates and incubated for twenty-four hours at 37 °C and 5% CO₂. For cytotoxicity assays, media was aspirated and replaced after 24 hours with media supplemented with relevant additive. Following incubation, 10 µL of WST-1 reagent was added at relevant time-points and plates incubated at 37 °C and 5% CO₂ for four hours. Absorbance was recorded at 440 nM using a FLOUstar Omega microplate reader (BMG Labtech). Background absorbance was correcting using wells containing media-only and for cytotoxicity assays results were normalised against cells incubated in media with no supplements.

2.8.2. ATPlite assay

In toxicity assays where WST-1 reagent was inappropriate, the ATPlite luminescence assay system (PerkinElmer) was used according to the manufacturers guidelines. In brief, 4000 cells per well were seeded into 96 well plates and incubated for twenty-four hours at 37 °C and 5% CO₂. Media was aspirated and replaced after 24 hours with media supplemented with relevant additive. Following incubation, 50 µL of mammalian cell lysis solution was added to wells and plate was shaken at 700 rpm for five minutes. 50 µL of substrate solution was subsequently added to wells and plate was shaken for a further five minutes at 700 rpm. Plate was covered for ten minutes prior to reading luminescence on a FLOUstar Omega (BMG-Labtech).

2.8.3. Western blot

A summary of antibodies used can be found in table 2.14. 35 µg of cell lysate was subjected to 10% SDS-PAGE under reducing conditions and transferred to polyvinylidene fluoride (PVDF, 0.45 µm) membranes via semi-dry transfer at 12 V for 30 minutes. Membranes were subsequently blocked with 3% milk in 1 x TBS with 0.1% Tween 20 with gentle agitation for one hour at room temperature. Blocking buffer was replaced and supplemented with primary antibodies of interest before incubation overnight at 4 °C with gentle agitation. Membrane was washed five times with TBS-tween prior to incubation with secondary, HRP-labelled antibodies for sixty minutes at room temperature. TBS-tween wash steps were repeated, and detection was carried out using the enhanced chemiluminescent reagent (Promega). Image capture was performed by LAS-3000 imaging software (Fujifilm).

Table 2.14: A summary of the antibodies used within this study and their derived species and dilutions.

Antibody	Manufacturer	Derived species	Dilution
Anti-DDX52	Sigma-Aldrich	Mouse	1:1000
Anti-DDX49	Sigma-Aldrich	Mouse	1:1000
Anti-beta actin	Santa Cruz Biotechnology	Rabbit	1:1000
Anti-Rabbit IgG (H+L) Secondary Antibody, HRP	Invitrogen	Goat	1:2000
Anti-Mouse IgG Cross-Adsorbed Secondary Antibody, HRP	Invitrogen	Goat	1:2000
Anti-biotin, HRP-linked Antibody	Cell signalling technology	Goat	1:2000

2.8.4. Transfection protocols

Transfection experiments were carried out in accordance with manufacturer protocols: a brief overview is given in the sections below. The plasmid pEGFP (a kind gift from Professor Ronald Chalmers), optimised for human expression of GFP, was used as a control to study successful transfection using an Axiovert S-100 (Zeiss) in combination with a HBO 100 lamp. gRNA was prepared from Alt-R CRISPR-Cas9 crRNA (IDT) and Alt-R CRISPR-Cas9 tracrRNA (IDT) as described in section 2.2.6. For jetCRISPR, forward and reverse transfection protocols were trialled. In brief, in forward transfection protocols cells are seeded 24 hours prior to transfection, whereas in reverse transfection cells are seeded and transfected simultaneously. Forward transfection was preferred and was used for subsequent methods. Initial trials used purified recombinant Cas9, later experiments used ALT-R Cas9 (Integrated DNA Technologies).

2.8.4.1. jetCrispr

Methods were carried out according to a combination of jetCRISPR (Polyplus, Strasbourg, France) manufacturer protocols and those of the ALT-R™ Cas9 System (IDT). Cells were grown to 80% confluency and seeded at a density of 80,000 cells per well in a 24-well plate. Prior to assay, Cas9 protein was diluted to 1 μ M in Cas9 working buffer. To prepare RNP solution, Cas9 protein and gRNA were gently mixed at a 1:1 ratio in serum free medium (SFM) to form an RNP complex of 330 nM and incubated at room temperature for 10 minutes. jetCRISPR reagent was added to RNP solution and mixed gently by pipette prior to incubation at room temperature for a further 15 minutes. RNP transfection solution was spiked into culture media at a final concentration of 33 nM RNP complex per well. Cells were incubated at 37 °C (5% CO₂) for 48 hours prior to further analysis. pEGFP was used as a transfection positive control.

2.8.4.2. Lipofectamine 3000

Methods were carried out according to manufacturer protocols. Cells were grown to ~80% confluency and seeded at a density of 40,000 cells per well in a 24 well plate. 1.5 μ L of Lipofectamine 3000 (Thermo Fisher) reagent was diluted in serum-free media as per manufacturer protocols. Cas9 was diluted to 1 μ M in Cas9 working buffer. Cas9 and gRNA were mixed at a 1:1 ratio in serum free media to form an RNP complex of 330 nM prior to the addition of P3000

reagent. Diluted lipofectamine 3000 reagent and RNP complex were mixed 1:1 and incubated at room temperature for fifteen minutes. Transfection solution was then added to each well at a final concentration of 16 nM RNP complex per well. pEGFP was used as a transfection positive control.

2.8.4.3. CRISPRmax

Methods were carried out according to manufacturer protocols. Cells were grown to ~80% confluency and seeded at a density of 40,000 cells per well in a 24 well plate. Cas9 was diluted to 1 μ M in Cas9 working buffer. Cas9 and gRNA were mixed at a 1:1 ratio in serum free media and Cas9Plus reagent added to form a final RNP concentration of 240 nM and incubated at room temperature for five minutes. CRISPRMAX reagent was diluted in serum free media as per manufacturer instructions and the RNP solution added at a 1:1 ratio. Transfection solution was incubated for twenty minutes at room temperature. Transfection solution was then added to each well such that the final concentration of RNP complex was 11 nM per well. pEGFP was used as a transfection positive control.

2.8.5. CRISPR-Cas9 genetic editing of human cell lines

Cells were grown to ~80% confluency and seeded at a density of 40,000 cells per well in a 24 well plate. After 24 hours, transfection was carried out using Alt-R Cas9 and ALT-R gRNAs (IDT) and CRISPRMAX transfection reagent (Thermo Fisher) as described in section 2.8.4.3. After 48 hours, cells were trypsinised and cell count performed in triplicate. A series of dilution steps were performed to a final cell density of five cells per mL prior to seeding into two 96 well plates such that 0.5 cells were seeded per well. Remaining undiluted cells were retained and genomic DNA purified using PureLink™ genomic DNA mini kit (Invitrogen) according to manufacturer instructions for downstream assessments of editing efficiency.

2.8.5.1. PCR amplification of genomic DNA amplicons for evaluating CRISPR editing

PCR primers for 800 bp regions flanking both on-target and off-target sites were designed using the CrispOR online program (Tefor infrastructure, CRISPOR, <http://crispor.tefor.net>). Genomic DNA was extracted using the PureLink™

Genomic DNA Mini Kit according to the manufacturers protocol and DNA content recorded using a DeNovix Spectrophotometer using the following equation:

$$\text{dsDNA concentration} = 50 \mu\text{g/mL} \times \text{OD260} \times \text{dilution factor.}$$

A total of 100 ng of was PCR amplified using Q5 DNA polymerase. Successful amplification was visualised using TBE agarose gels stained with ethidium bromide and products extracted using Wizard® SV Gel and PCR Clean-Up System. Amplicons were stored at -20 °C for downstream analysis.

2.8.5.2. T7E1 genome editing assay

A reaction mix of 100 ng of amplicon DNA (from section 2.8.5.1) in 1 x T7 reaction buffer was heated at 95 °C for ten minutes and cooled to 25 °C at a rate of -0.3 °C per second to encourage heteroduplex DNA formation. T7 endonuclease (NEB) was added to the reaction mix and incubated at 37 °C for one hour. Finished reactions were run on a 2% agarose gel for 60 minutes at 140 volts prior to visualisation.

2.8.5.3. Tracking of Indels by Decomposition

Amplicons shown to be positive for heteroduplex formation in section 2.8.5.2 were sent for Sanger sequencing (Deepseq, The University of Nottingham). Edited amplicons were compared with wild type amplicons using TIDE analysis suite (<http://shinyapps.datacurators.nl/tide/>, Netherlands Cancer Institute, Amsterdam, Netherlands) with parameters set to identify Indels with a maximum size of 50 bp and a P-value threshold of 0.001 (315).

2.8.6. CRISPR-Cas9 Prime editing of human cell lines

2.8.6.1. Plasmid construction

Plasmids p169850, p 132777 and p65777 were ordered from Addgene as agar slabs and miniprepmed using a Wizard® Plus SV Minipreps DNA Purification System. All plasmids were maxi prepmed of endotoxins prior to transfection into human cells. Further information on plasmid construction can be found within supplementary information of Anzalone et al (2019) (306) .

2.8.6.1.1. Nicking gRNA

Plasmids were constructed using golden gate cloning according to methods described in Anzalone et al (2019). In brief, oligonucleotides listed in table 2.8 were phosphorylated and annealed at concentrations of 500 nM using 5 units of T4 PNK (NEB) in 1 x T4 ligase buffer. Reaction mix was incubated at 37 °C for 60 minutes, prior to heating to 95 °C for 5 minutes and subsequent cooling to 10 °C at -5 °C per minute. For control over cooling steps, reaction was carried out in a Miniamp thermocycler. Plasmid #65777 was digested using BsmBI restriction enzyme (NEB) at 55 °C for 60 minutes with simultaneous dephosphorylation using CIP (NEB). Reaction mix was visualised on a 1% TBE agarose gel and linear plasmid extracted by gel extraction. A reaction mix containing 40 ng of digested plasmid and 50 nM annealed oligonucleotides was incubated at 16 °C overnight with 200 units of T4 ligase. Ligation mix was subsequently transformed into DH5 α strain of *E. coli* using heat-shock method (see section 2.3.3). Due to small size of insert, successful cloning was verified through Sanger sequencing only (Genewiz).

2.8.6.1.2. pegRNA

Plasmids were constructed using golden gate cloning according to methods described in Anzalone et al (2019). In brief, plasmid #132777 was digested using 20 units of Bsa1-HF2 at 37 °C overnight prior to visualisation on a 1% TBE agarose gel and isolation of 2.2 kb fragment via gel extraction. Oligonucleotides listed in table 2.8 were annealed in 1 x annealing buffer through heating at 95 °C for 3 minutes prior to gradual cooling (0.1 °C/s) to 22 °C. For control over cooling steps, reaction was carried out in a Miniamp thermocycler. Components 1 and 3 were subsequently diluted to 1 μ M in SDW. Component 2 was phosphorylated at a concentration of 1 μ M using T4 PNK in 1 x T4 DNA ligase buffer at 37 °C for 60 minutes. A reaction mix of 30 ng digested vector and 100 nM of each component was ligated overnight at 16 °C using 400 units of T4 ligase. Ligation mix was transformed into DH5 α strain of *E. coli* using heat-shock method (see section 2.3.3). Unsuccessful clones were identifiable as red colonies due to presence of intact RFP cassette. Successful colonies were confirmed by Sanger sequencing (Genewiz).

2.8.6.2. Transfection

Cells were processed and counted as described in section 2.7 and seeded into a 24-well plate at a density of 40,000 cells per well. After 24 hours, transfection was carried out Lipofectamine 3000 transfection reagent according to manufacturer guidelines and methods described in Anzalone et al (2019). It is noted that Anzalone et al (2019) used Lipofectamine 2000. In brief, 1.5 uL of Lipofectamine™ 3000 Reagent was added to 25 uL of serum-free media and vortexed. Separately, a master mix of plasmid DNA, P3000™ Reagent and serum-free media was prepared in 50 uL total volume such that a final volume of 750 ng Cas9-PE3, 250 ng pEGRNA and 83 ng of nicking sgRNA would be transfected. 25 uL of master mix was added to solution, mixed via pipette, and incubated at room temperature for fifteen minutes prior to addition to the wells. Media was mixed gently via pipette and cells incubated for 72 hours prior to analysis.

2.8.6.3. Analysis

Cell health was assessed and imaged using an Axiovert S 100 microscope. Genomic DNA of edited and wild type cells was extracted using PureLink™ Genomic DNA Mini Kit according to manufacturer protocol, sites of interest were PCR amplified as described in section 2.8.5.1 and amplicons sent for Sanger sequencing (Genewiz). Sanger sequencing traces were analysed using TIDER analysis suite (316).

2.9. Ribosome biogenesis methods

2.9.1. Northern blotting

To trial radiation-free methods, a methods were based upon those described by Miller et al (2018) with several exceptions. Whereas Miller et al used probes labelled via click chemistry with IR800 dyes, our method trialled probes labelled via click chemistry with DBCO-Cy5. In addition, due to unavailability of a radiation-decommissioned hybridisation chamber, an improvised method was performed using a shaking incubator chamber. Total RNA was extracted from cells of interest using RNeasy mini kit (Qiagen) according to manufacturer protocols.

2.9.1.1. Click chemistry

Probes were constructed possessing internal azides (IDT) of which 2.5 μM was incubated with 50 μM DBCO-Cy5 in SDW and incubated at either 4 °C or 37 °C overnight. Conjugated DNA was subsequently purified using illustra™ MicroSpin™ G-25 columns and DNA concentration determined using a DeNovix spectrophotometer.

2.9.1.2. Northern blotting

RNA samples were mixed with 2 x formaldehyde loading solution at a 1:1 ratio in a GuardOne laminar flow cabinet and heated at 70 °C for five minutes to denature RNA prior to cooling to room temperature. Samples were electrophoresed on a 1% agarose gel supplemented with 1 x HT buffer and 1.3% formaldehyde for 10 minutes at 6 V/cm prior to reduction of voltage to 3.5 V/cm until dye front reached 3 cm from end of the gel. Gel was transferred to a Hybond N+ (Amersham) membrane via capillary blotting overnight. Membrane was stained with methylene blue to confirm RNA transfer prior to destaining in SDW. Membrane was UV crosslinked using a setting of 300,000 $\mu\text{J}/\text{cm}^2$ and stored in 0.1 x SSC buffer until hybridisation. Membrane was preincubated at 45 °C for one hour, prior to addition of 10 pmol of Cy5-probe and incubation overnight at 45 °C. Hybridisation solution was decanted, membrane rinsed once with wash solution and imaged using an Amersham Typhoon 5 biomolecular imager.

Chapter 3: Nucleic acid interactions of Probable ATP-dependent helicase DDX52

3.1. Introduction

3.1.1. Pre-established roles of DDX52

DDX52 is a putative RNA helicase predicted to play a role within rRNA processing and ribosome biogenesis. However, very little data is available showing the biochemistry or activity of DDX52 in vitro, with reported functions based upon its homology with the yeast protein Rok1. However, initial reports of this homology appear to be based on DDX52 peptide sequences that are different from the now canonically accepted amino acid sequence (95). Interactions have been demonstrated between DDX52 and UTP23, indicating that interactions seen at the expansion segment 6 (ES6) region of 18S rRNA are likely conserved across humans and yeast (318).

3.1.2. A role for DDX52 in DNA repair and replication?

Supplementary data from Buckley et al (2020) identified that the *Mycobacterium smegmatis* protein Lhr shared potential homology with DDX52. The same study suggested that Lhr acted on stalled or broken replication forks, with a role within DNA repair. Similarly, supplementary data from two separate studies by Matsuoka et al (2008) and Dephore et al (2008) have demonstrated the phosphorylation of DDX52 residues in response to DNA damage, implicating DDX52s potential interaction within DNA repair (102, 103).

3.1.3. Aims and objectives

Whilst it's predicted homology to the yeast protein Rok1 and thus role in RNA processing is widely reported, there is very little available evidence to underline the functions or biochemistry of DDX52. In addition, the association of DDX52 with cancer and the diverse roles of DExD-box proteins across a wide range of pathways highlights the importance of initial biochemical characterisation. This chapter focuses on examining and establishing the initial biochemistry of the proteins. The objectives were as such thus:

- 1) To examine the initial biochemistry of the protein and establish substrate preferences of recombinant DDX52
- 2) To establish any potential dual-role functions of DDX52; notably, in the processing of DNA replication forks, viral immunity or chromosomal regulation
- 3) To trial new methods and begin preliminary investigations into looking into DDX52's role within ribosome biogenesis.

3.2. Results

3.2.1. Bioinformatics

3.2.1.1. Predictions of DDX52 structure reveal an intrinsically disordered N-terminus loop

The only existing experimental structural model of DDX52 is of the commonly conserved RecA1 domain. To examine full-length DDX52 and identify sites of interest, structural predictions as modelled by the Alphafold software suite were studied (Fig. 3.1). N-terminus residues 1-127 were modelled with low confidence. The helicase core, featuring the RecA1 and RecA2 domains, shows high conservation with other members of the DExD-box family. DDX52 possesses a relatively novel Motif II sequence, consisting of a DESD-box than the canonical DEAD-box. Most strikingly, N terminal residues F28-T128 are predicted to form a large and novel intrinsically disordered region, resembling an extended loop. In initial identification this was informally termed BD motif (Big Dipper motif), reflected in oligonucleotide names in section 2.1.4, prior to later referral as IDR loop. The presence of this intrinsically disordered region was reinforced within predictions by the IUPred2A software (Fig. 3.2), with residues 57 – 101 reported to feature high levels of intrinsic disorder. Parallel ANCHOR2 scores, a representation of the likelihood of a partner protein to trigger a transition from disorder-to-order, suggest that this is unlikely to function as a disordered binding region.

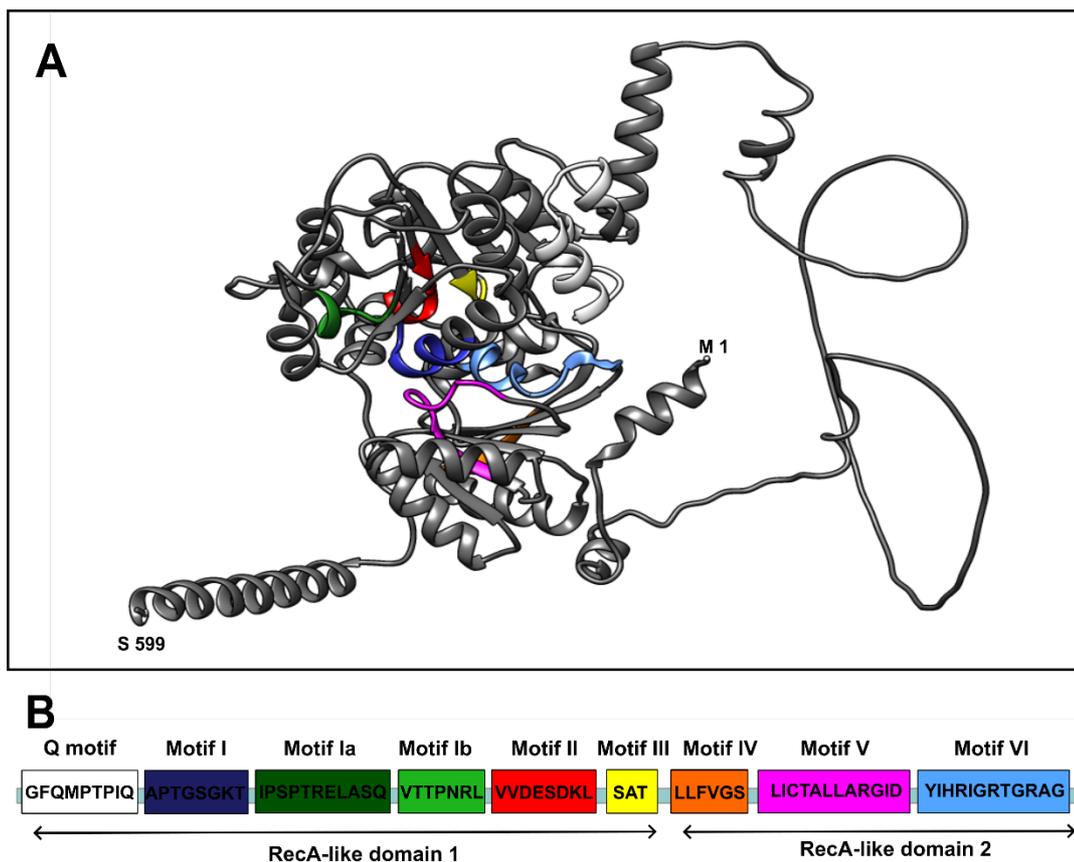


Figure 3.1: A predicted structural model of human DDX52. (A) The structure of human DDX52 as predicted by Alphafold software (319). Motifs are coloured as depicted in B and the N and C-terminus residues are indicated. (B) Motifs and domains of DDX52 with associated amino acid sequences.

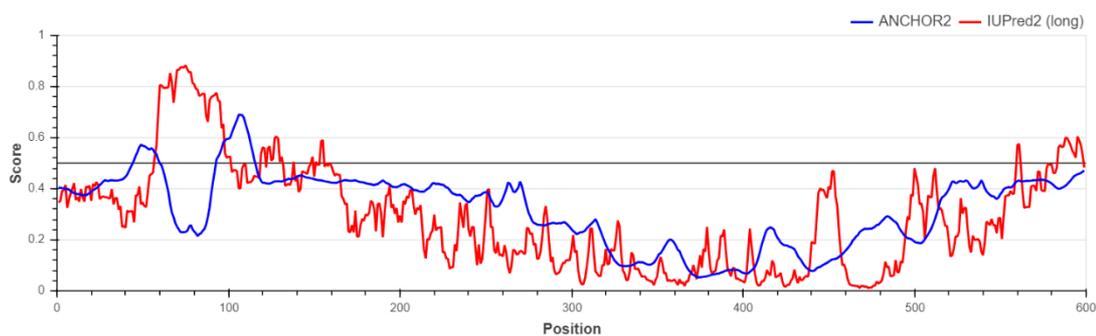


Figure 3.2: Output of IUPred2 and ANCHOR2 predictions for Human DDX52. IUPred2 and ANCHOR2 scores are shown in red and blue, respectively (320).

3.2.1.2. DDX52 comparisons with Lhr and Rok1

To examine previous comparisons in literature, the structure of DDX52 as predicted by Alphafold was superimposed onto the structure of *Mycobacterium smegmatis* Lhr as determined by X-ray diffraction (Fig. 3.3). The core helicase domains of DDX52 showed homology with the RecA-like domains of Lhr, but the N and C terminal of both proteins were distinctly dissimilar. Sequence comparisons between conserved motifs show further differences, with Lhr possessing a DExH-box and RecA-like domains characteristic of an Ski2-like helicase in comparisons to the DExD-box domains of DDX52.

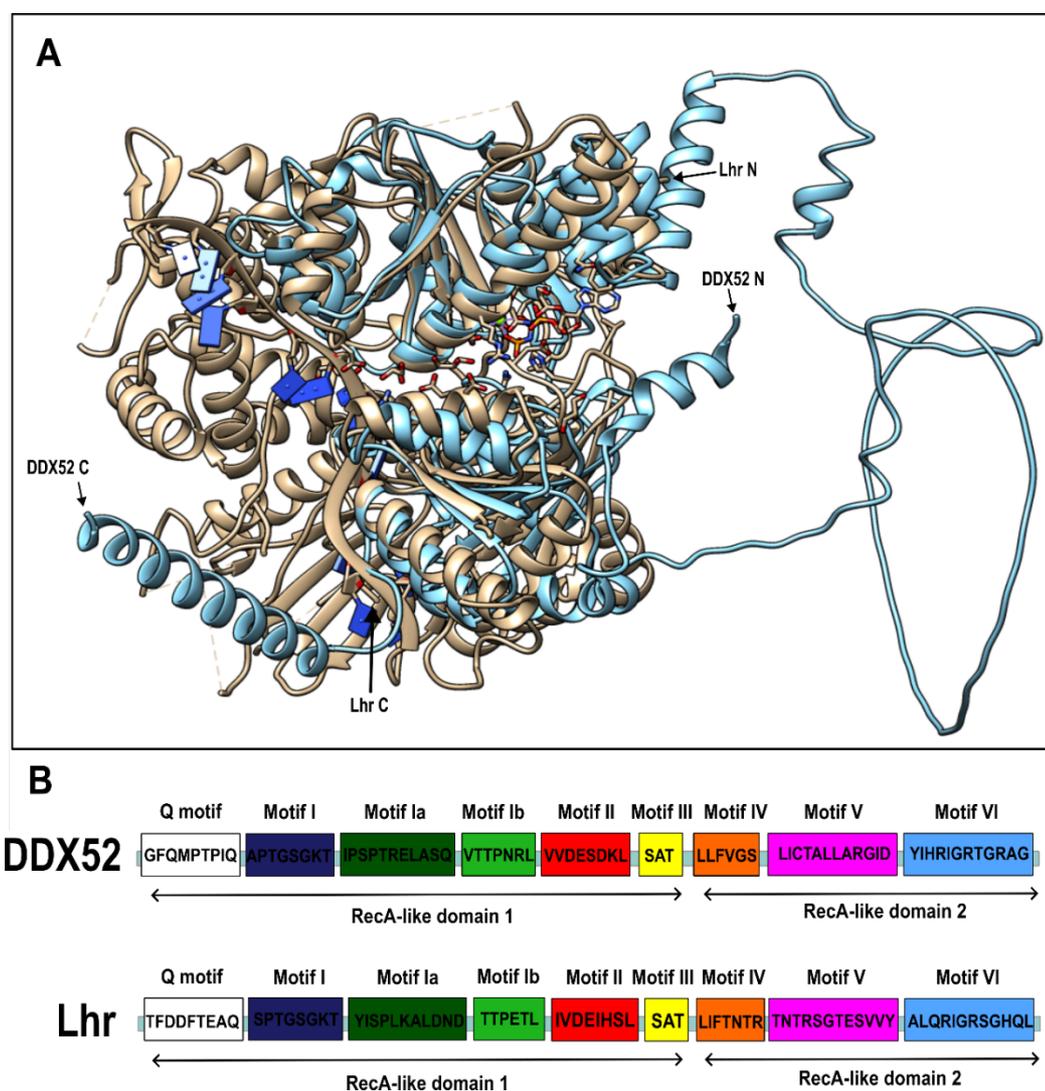


Figure 3.3: Structural and domain comparisons between DDX52 and Lhr.

(A) Alphafold prediction model of DDX52 superimposed onto *M. smegmatis* Lhr (PDB: 5V9X). DDX52 is shaded in light blue and Lhr shaded in cream. Respective N and C termini are indicated. Lhr structure is depicted bound to

ssDNA as modelled by Ejaz et al (2018) through X-ray diffraction. **(B)** Comparisons of domains of DDX52 and Lhr.

DDX52 has also commonly been described as the human homolog of the *S. cerevisiae* protein Rok1. Structural and sequence comparisons between DDX52 and Rok1 were carried out (Figs 3.4 and 3.5). The structure of Rok1 has not been resolved thus structures are based upon Alphafold predictions, with Rok1 N-terminus residues 1 - 87 modelled with low confidence. High degree of conservation was noted within the core helicase RecA1 and RecA2 domains, with homology also noted within the C-terminus of the protein. The C-terminus helical extension of DDX52 resembles that of Rok1 but the two protrude at different angles. Both proteins possess an intrinsically disordered 'loop' within their N-terminus, but the loop noted within DDX52 is substantially larger. Sequence alignments (Fig. 3.5) confirm these observations, with substantial variation within the N-termini of both proteins, notably irregular insertions of up to 27 amino acids within DDX52, whilst conserved residues are observed within the core and C-terminus.

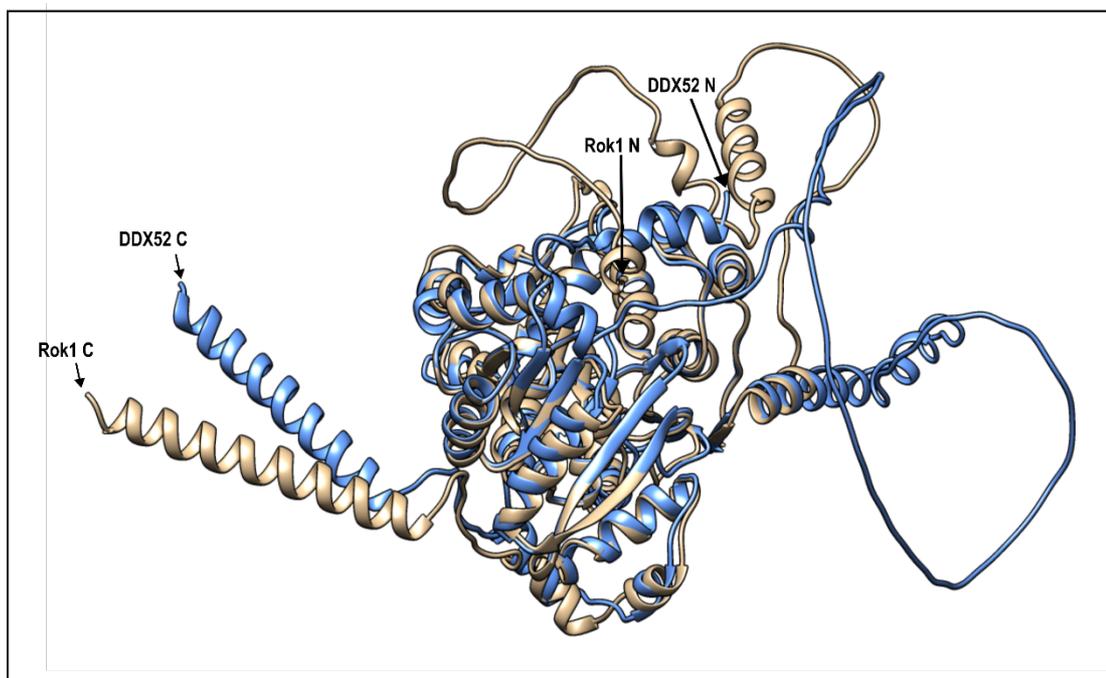


Figure 3.4: Alphafold prediction model of DDX52 superimposed onto *S. cerevisiae* Rok1. DDX52 is shaded in light blue and Rok1 shaded in cream. Respective N and C termini are indicated. Both protein structures are depicted as predicted by Alphafold software (319) .

Chapter 3: Nucleic acid interactions of DDX52

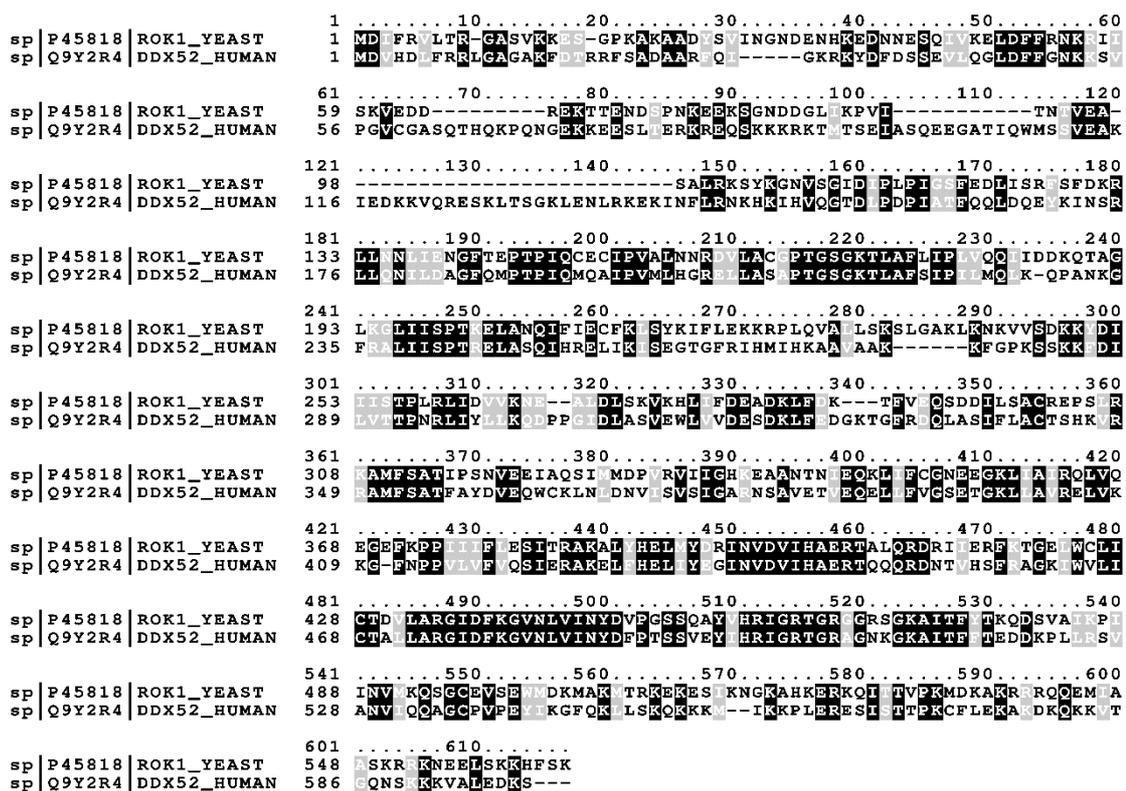


Figure 3.5: Comparative sequence alignments of *S. cerevisiae* Rok1 and Human DDX52. Sequence alignments were carried out using T-coffee and modelled using Boxshade software (322).

3.2.2. Purification of recombinant wild type and mutant DDX52

To purify DDX52 and associated mutants, the human ORF was codon-optimised for *E. coli* and relevant plasmids (see table 2.4) transformed and expressed within the BL21-AI strain of *E. coli*. Initial pilot overexpressions confirmed expression of recombinant DDX52 (Fig. 3.6A). Purification proceeded as outlined in section 2.4.3. DDX52 included an N-terminus his-tag incorporated from the vector plasmid and following lysis was purified using a HisTrap HP His. This was successful and was sufficient to get pure DDX52 with few contaminants (Fig. 3.6B), however previous trials and A260 values noted copurification of DNA with protein. Following a dialysis step to exchange the protein preparation into a low salt containing buffer and remove high levels of imidazole, further purification was carried out using a HiTrap™ Heparin HP column to separate protein from DNA. This was successful (Fig. 3.6C). DDX52 positive fractions were dialysed into a high glycerol containing storage buffer,

which had the effect of further concentrating the protein to highlight any contaminants (Fig. 3.6D).

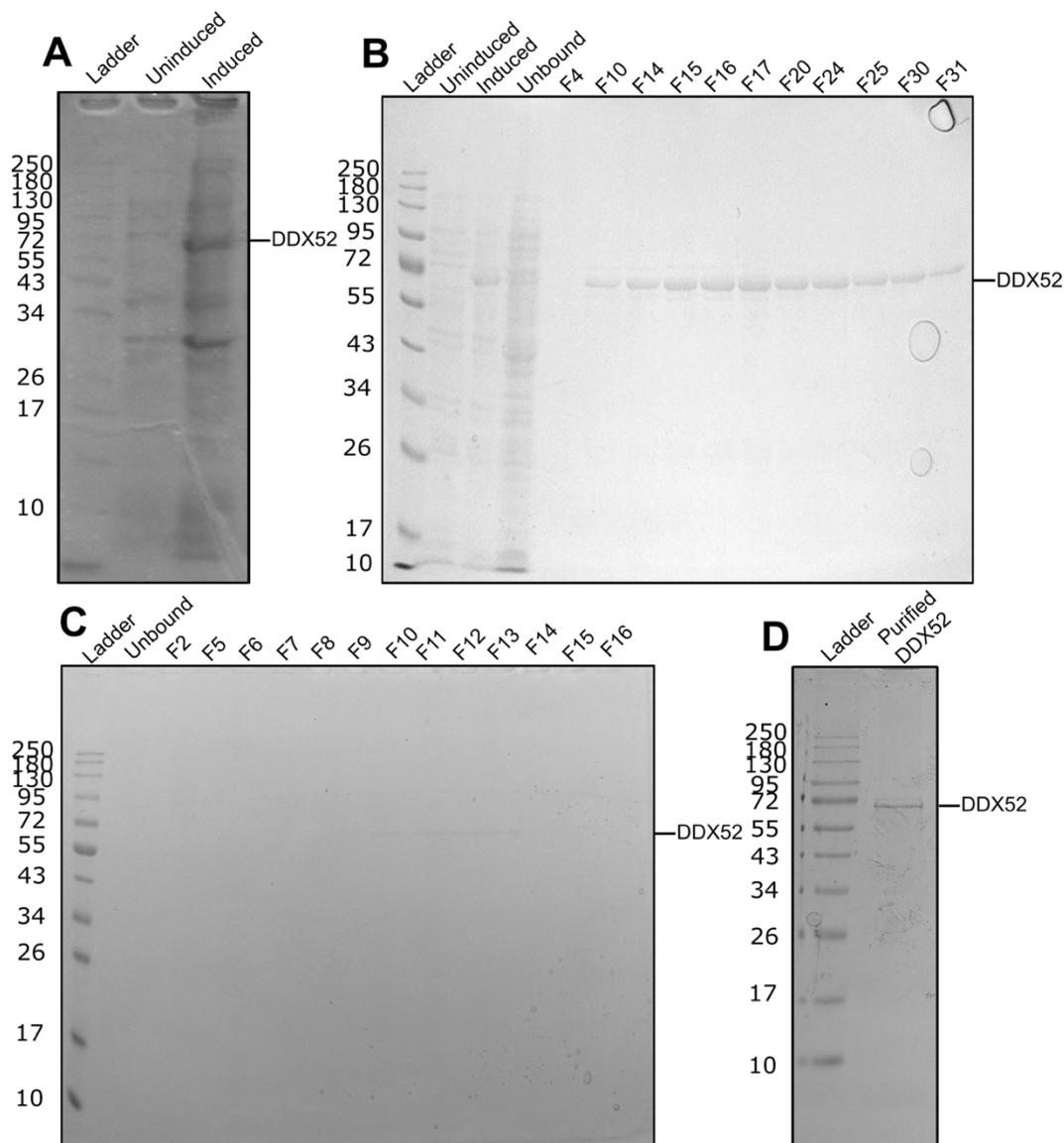


Figure 3.6: An overview of recombinant His-DDX52 purification. His-DDX52 was overexpressed in the BL21-AI strain of *E. coli* and methods carried out as described in section 2.4. Recombinant DDX52 with incorporated His-tag possessed an expected molecular weight of 68.2 kDa. **(A)** DDX52 overexpression was confirmed in a scaled-down induction assay, comparing lysate from transformed *E. coli* that were induced and uninduced. **(B)** Protein lysate was loaded onto a HisTrap HP His-tag column and DDX52 positive fractions identified by SDS-PAGE. **(C)** Sample was passed down a HiTrap™ Heparin HP column and DDX52 positive fractions identified by SDS-PAGE. **(D)** Purified recombinant DDX52.

To further probe the protein, a series of mutants were cloned and purified in addition to the wild type. The isoelectric point (pI) was determined for each mutant using ProtParam online tool and no deviations from the protocol in section 2.4.3 were necessary (323). To investigate if DDX52 acts in an ATP-dependent mechanism and confirm helicase activity was a product of DDX52 and not contaminants, a double mutant (D318A, D321A) within motif II was created based upon similar mutants generated by Garbelli et al (2011). N and C-terminus truncations were cloned in collaboration with student Philipp Springer according to the position of the RecA1 and RecA2 domains to dissect separation of function and activity between these parts of the protein. An additional F28_T128del mutant was purified based upon the N-terminus IDR loop identified within bioinformatic analysis. C-terminus mutant was purified in collaboration with student Philipp Springer, other mutants by the author. The purification of each desired mutant was successful (see Fig. 3.7 for summary gel) and the protein was stored at -80 °C.

3.2.3. Initial optimisation of recombinant DDX52

As no reported literature of DDX52 biochemistry was available at the time of purification, there was no previously known activity to validate the purified protein was active. As such, all biochemistry was novel and reaction conditions of protein required optimisation. DDX52 reaction conditions were optimised by examining unwinding activity of DDX52 on a flayed duplex substrate through methods described in chapter 2 section 2.6.1. Initially, DDX52s activity as a function of protein concentration was examined (Fig. 3.8) and negligible unwinding observed seen until concentrations of 500 and 1000 nM indicating poor proficiency as a helicase. A concentration of 500 nM DDX52 was used to optimise DDX52s activity as a function of ATP (Fig. 3.9A) and optimal activity seen at concentration of 2 mM ATP. This was used to subsequently optimise concentrations of MgCl₂ (Fig. 3.9B) and optimal activity was seen at conditions of 2 mM MgCl₂, indicating a 1:1 ATP:MgCl₂ ratio.

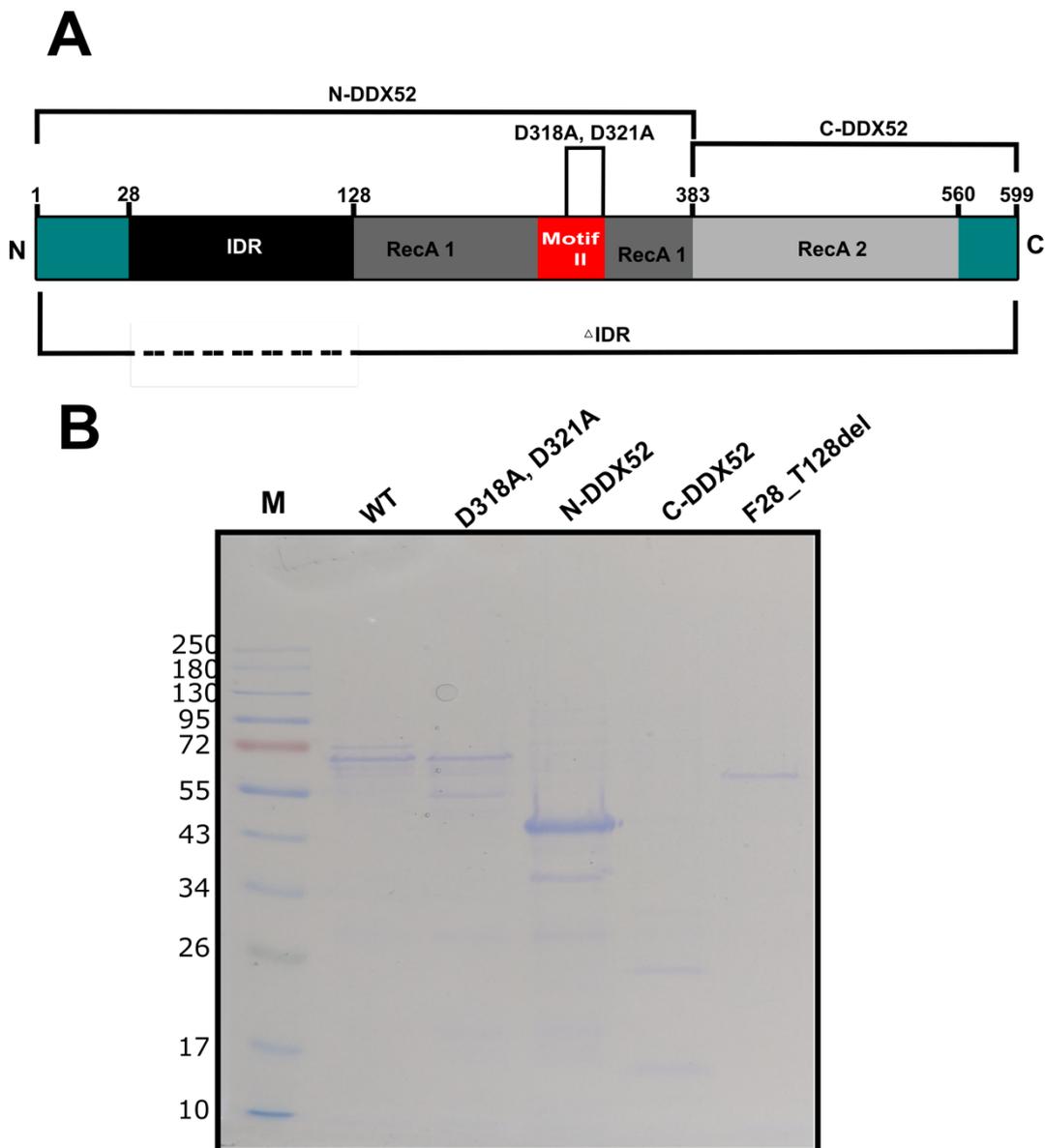


Figure 3.7: SDS-PAGE of WT DDX52 and associated mutants used within this study. A) Graphical representation of the mutants used within this study. B) An SDS-PAGE showing the purified mutants used for testing within this study. All mutants contained and were purified using incorporated His-tags. Expected molecular weights were as follows: WT - 69 kDa, D318A, D321A – 69 kDa, N-DDX52 – 44.5 kDa, C-DDX52 – 24.5 kDa, F28_T128del – 57.5 kDa. C-terminus mutant was purified by student Philipp Springer.

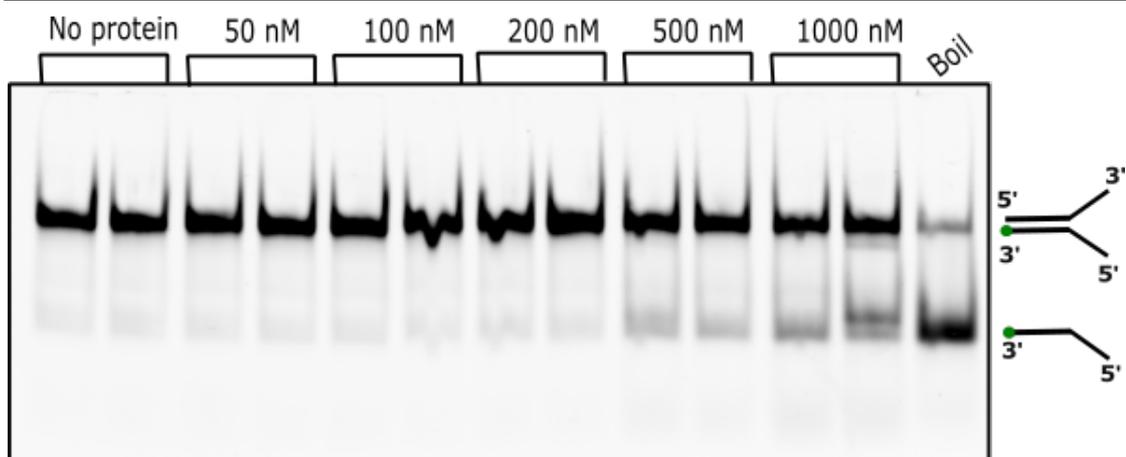


Figure 3.8: Low concentrations of DDX52 are poorly proficient at unwinding a flayed DNA duplex. Unwinding was analysed by testing five DDX52 concentrations in duplicate with 25 nM of a flayed duplex DNA substrate supplemented with 5 mM MgCl₂, 5 mM ATP and 25 mM DTT in 1 x helicase buffer. Included as controls were a substrate only reaction and a boiled substrate only reaction to indicate full dissociation.

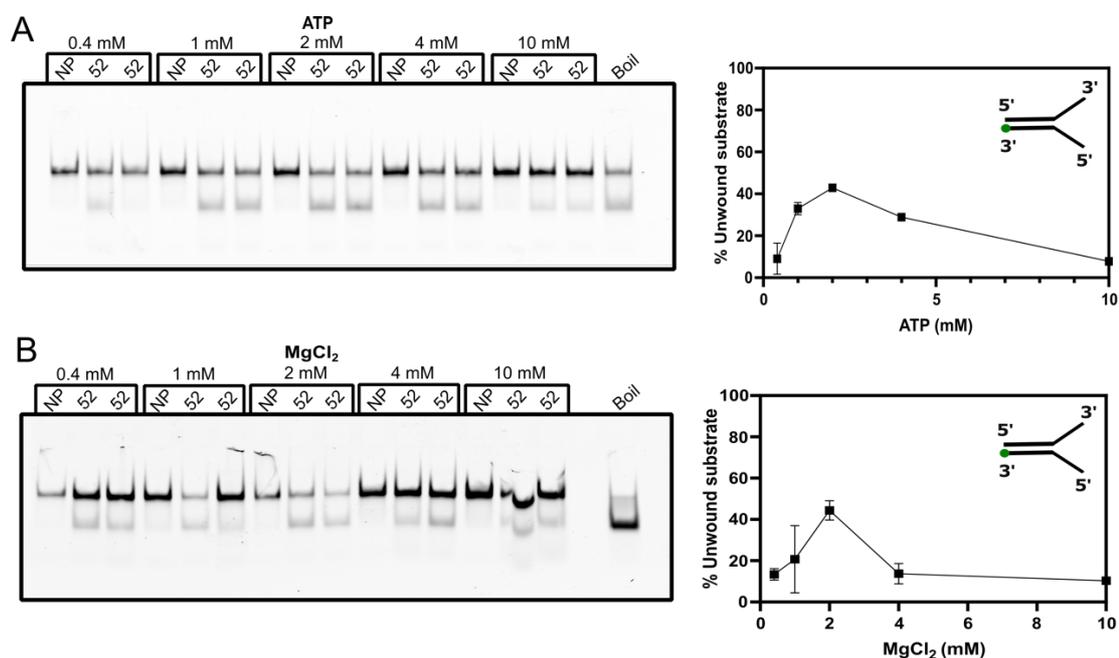


Figure 3.9: Assessing the optimal concentrations of ATP (A) and MgCl₂ (B) for DDX52 unwinding of a flayed DNA duplex. Unwinding was analysed using a DDX52 concentration of 500 nM with 25 nM of a flayed duplex DNA substrate supplemented with MgCl₂, ATP and 25 mM DTT in 1 x helicase buffer. Figures A and B used 2 mM MgCl₂ and 2 mM ATP, respectively. Included as

controls are substrate only and boiled substrate reactions to indicate stable substrate and full dissociation, respectively.

3.2.4. Interrogating DDX52 binding

3.2.4.1. The N and C terminal of DDX52 possess different nucleic acid binding characteristics within gel-based assays

Binding affinity of wild type DDX52 to several nucleic acid species — notably ssDNA, ssRNA and flayed duplex DNA — as a function of protein concentration was assessed by electrophoretic mobility shift assays (EMSA) as described in section 2.6.5. Studies used a 45-base pair Cy5 labelled ssDNA oligonucleotide (AP122), a 45-base pair Cy5 labelled ssRNA substrate (AP33) and a 50-base pair flayed duplex (fork 2B). DDX52 was unable to form a stable complex formation with tested substrates (Fig. 3.10A). Instead, a smearing pattern of the Cy5 signal and concentration-dependent formation of protein-DNA complexes within wells was observed. This was particularly apparent at 1000 nM DDX52 where 100% of free DNA was shifted. Trials with different conditions, including running the gel at 4 °C and without EDTA, did not improve banding patterns. An agarose based EMSA was trialled and aggregation within the wells replicated (Fig. 3.10B).

Purified DDX52 mutants were examined to identify if they impacted DDX52 nucleic acid binding (Fig. 3.11A). The F28_T128del mutant appeared inert: excess DNA was also noted within these lanes. There was little difference in the binding profiles of DDX52 and DDX52^{D318A,D321A}. Likewise N-DDX52 showed a similar profile, but formation of well aggregates was noticeably increased. C-DDX52, in contrast, formed semi-stable complexes with some evidence of additional banding appearing at higher concentrations. This difference in binding was further studied on a flayed duplex (Fig. 3.11B) and stability of the binding profile of the C-terminus dramatically increased whilst confirming higher bands above concentrations of 400 nM. Band shifts were also seen at concentrations eight-fold smaller than those seen for the wild type. This contrast in binding profile suggests that the C-terminus and N-terminus of DDX52 possess distinct mechanisms of binding nucleic acid species.

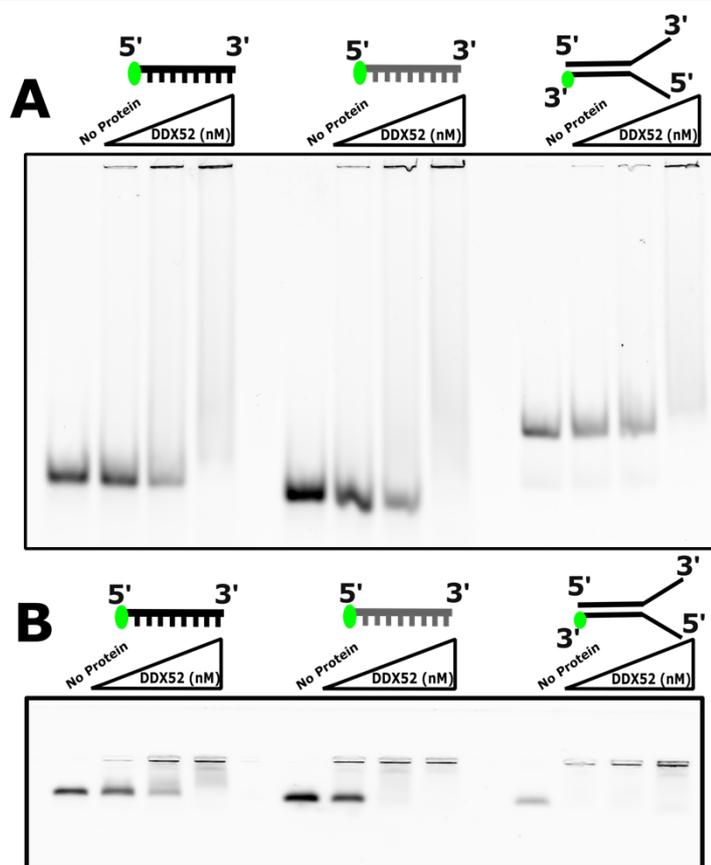


Figure 3.10: Electromobility shift assays on nucleic acid species show well aggregation as a function of wild type DDX52 concentration. EMSAs were performed as described in section 2.6.5. **(A)** Wild type DDX52 at concentrations of 0, 200 500 and 1000 nM was incubated with (from l-r) Cy5 labelled ssDNA, ssRNA and fork 2B (25 nM) for thirty minutes prior to electrophoresis on a 5% Native PAGE gel. **(B)** Wild type DDX52 at concentrations of 0, 200, 500 and 1000 nM were incubated with Cy5 labelled ssDNA, ssRNA and fork 2B (25 nM) for thirty minutes prior to electrophoresis on an Agarose gel.

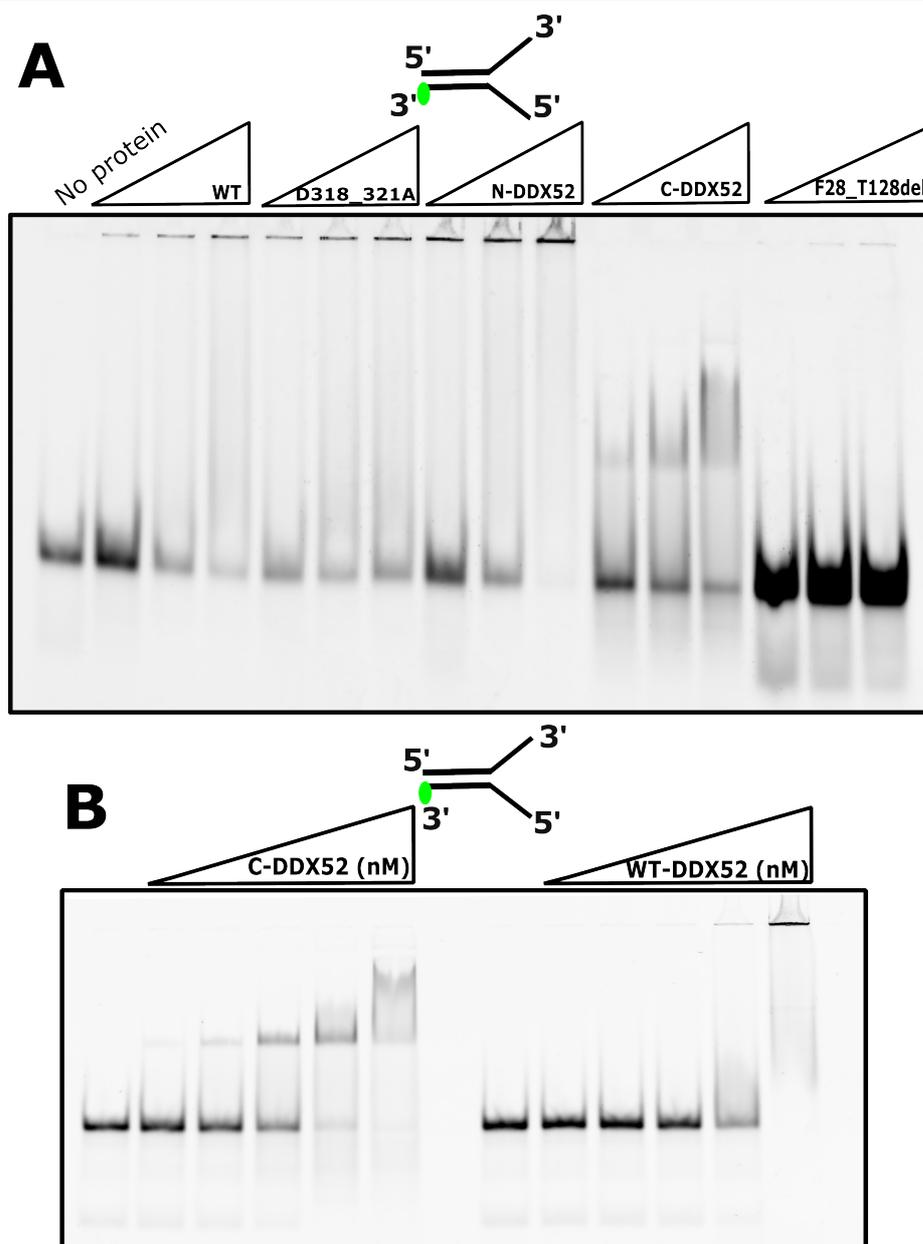


Figure 3.11: Electromobility shift assays on DDX52 mutants incubated with ssDNA and flayed duplexes reveal distinct binding profiles between C-terminus and full length DDX52 as a function of concentration. EMSAs were performed as described in section 2.6.5. **(A)** Wild type and mutant variants of DDX52 (as indicated) were incubated with Cy5 labelled ssDNA prior to running on a 5% native gel. Concentration range included 200, 500 and 1000 nM protein. **(B)** C-terminus truncation and wild type DDX52 were incubated with 25 nM Cy5 labelled fork 2B prior to running on a 5% native gel. Concentration range included 50, 100, 200, 400, 800 nM protein.

3.2.4.2. Fluorescence polarisation indicates DDX52 binds DNA with a higher affinity than RNA

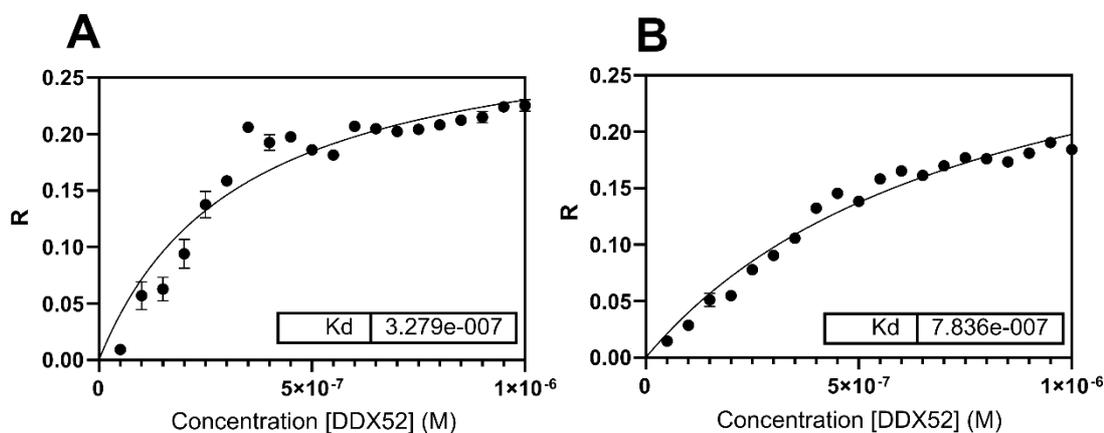


Figure 3.12: Fluorescent polarisation assays as a function of protein concentration show DDX52 binds RNA with a higher affinity than DNA.

Both data sets were carried out on a FLOUstar Omega as described in section 2.6.8. **(A)** Polarisation was examined on a series of 20 protein concentrations of 0-1000 nM incubated in duplicate with a 35-nucleotide poly-thymine 5' fluorescein labelled ssDNA. **(B)** Polarisation was examined on a series of 20 protein concentrations of 0-1000 nM incubated in duplicate with a 35-nucleotide poly-uracil 5' fluorescein labelled ssRNA. K_d s were calculated in PRISM software. No protein controls were run and subtracted to account for non-specific signals.

As EMSAs were suboptimal for studying wild type binding, a fluorescence polarisation method was employed to study the binding affinity of DDX52 to both ssDNA and ssRNA. Figure 3.12 shows that for both FAM-labelled ssDNA and ssRNA substrates, polarisation was observed as a function of DDX52 concentration confirming DDX52 binding. The upper range of DDX52 concentration was sufficient to plateau FP value and K_d s of 328 and 784 nM were recorded for DNA and RNA binding, respectively, suggesting DDX52 binds DNA with a higher affinity than RNA.

3.2.5. The unwinding of DDX52

3.2.5.1. DDX52 loads on and unwinds DNA strands in a 3' to 5' direction

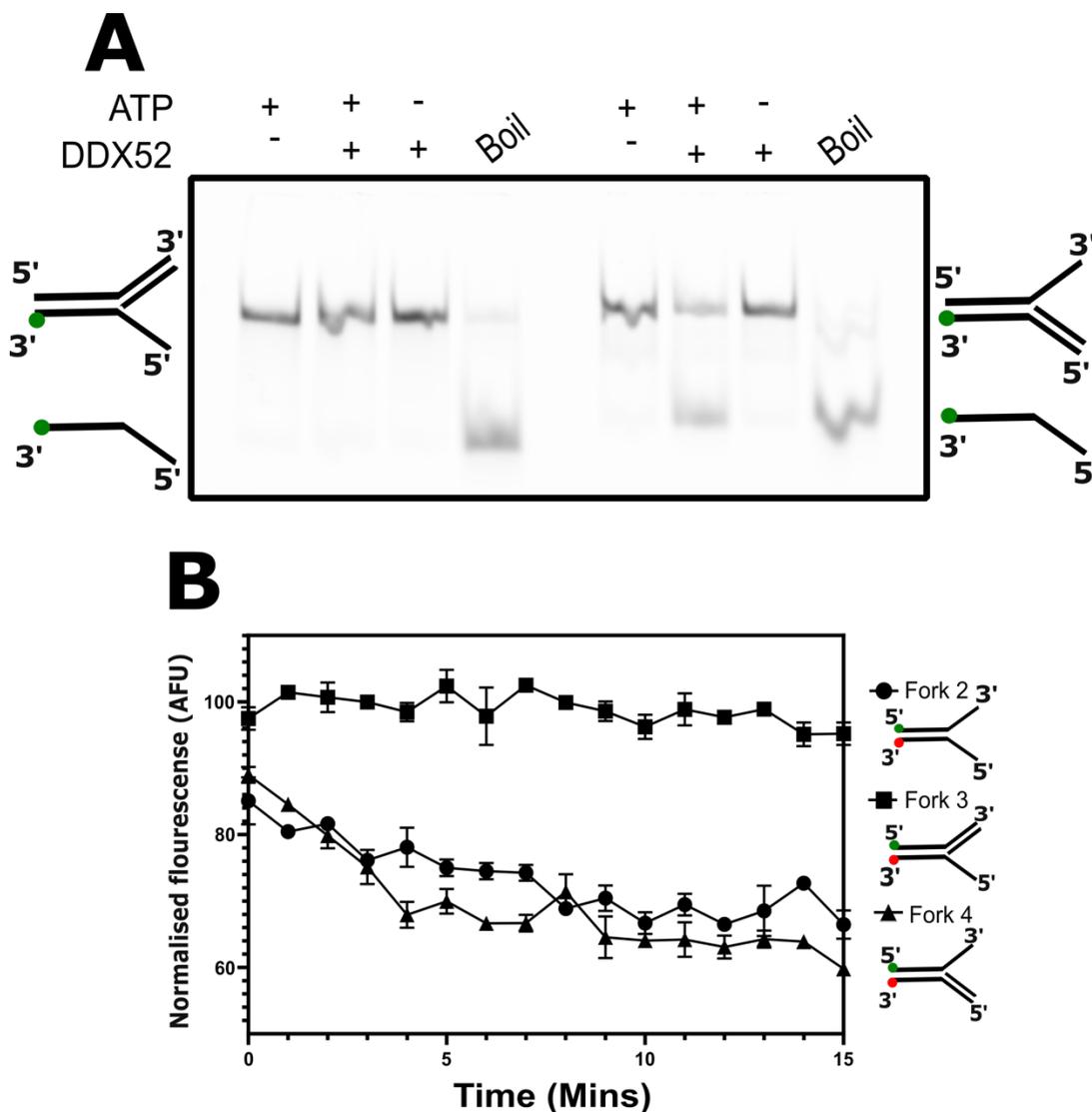


Figure 3.13: Unwinding assays confirm a 3' to 5' polarity for DDX52.

Helicase assays were performed as described in sections 2.6.1 and 2.6.2. **(A)** Gel-based helicase assay to examine DDX52 (500 nM) processivity on flayed duplexes (25 nM) with 3' and 5' flaps supplemented with 2 mM MgCl₂, 2 mM ATP, 25 mM DTT in 1 x helicase buffer. **(B)** FRET-based unwinding assay to examine DDX52 (500 nM) processivity on flayed duplexes (50 nM) with 3', 5' and open flaps supplemented with 2 mM MgCl₂, 2 mM ATP, 10 mM DTT in 1 x helicase buffer. Reactions were repeated in duplicate, normalised against a previously annealed control and the range of standard error is shown.

DDX52s direction of loading and unwinding was assessed (Fig. 3.13) by measuring the unwinding activity of DDX52 on a set of substrates through methods described in sections 2.6.1 and 2.6.2. No unwinding was seen on flayed duplexes with a 5' flap, whilst clear DNA strand separation was seen in duplexes with 3' flaps, with a maxima of ~ 40 percent unwound recorded in the FRET-based unwinding assay (Fig. 3.13B). This indicates a mechanism for 3'-5' loading and strand separation by DDX52 on flayed DNA duplexes. It is also noted in figure 3.13A that DDX52 activity is negligible in the absence of ATP, corroborating evidence of ATP dependency in section 3.2.5.3.

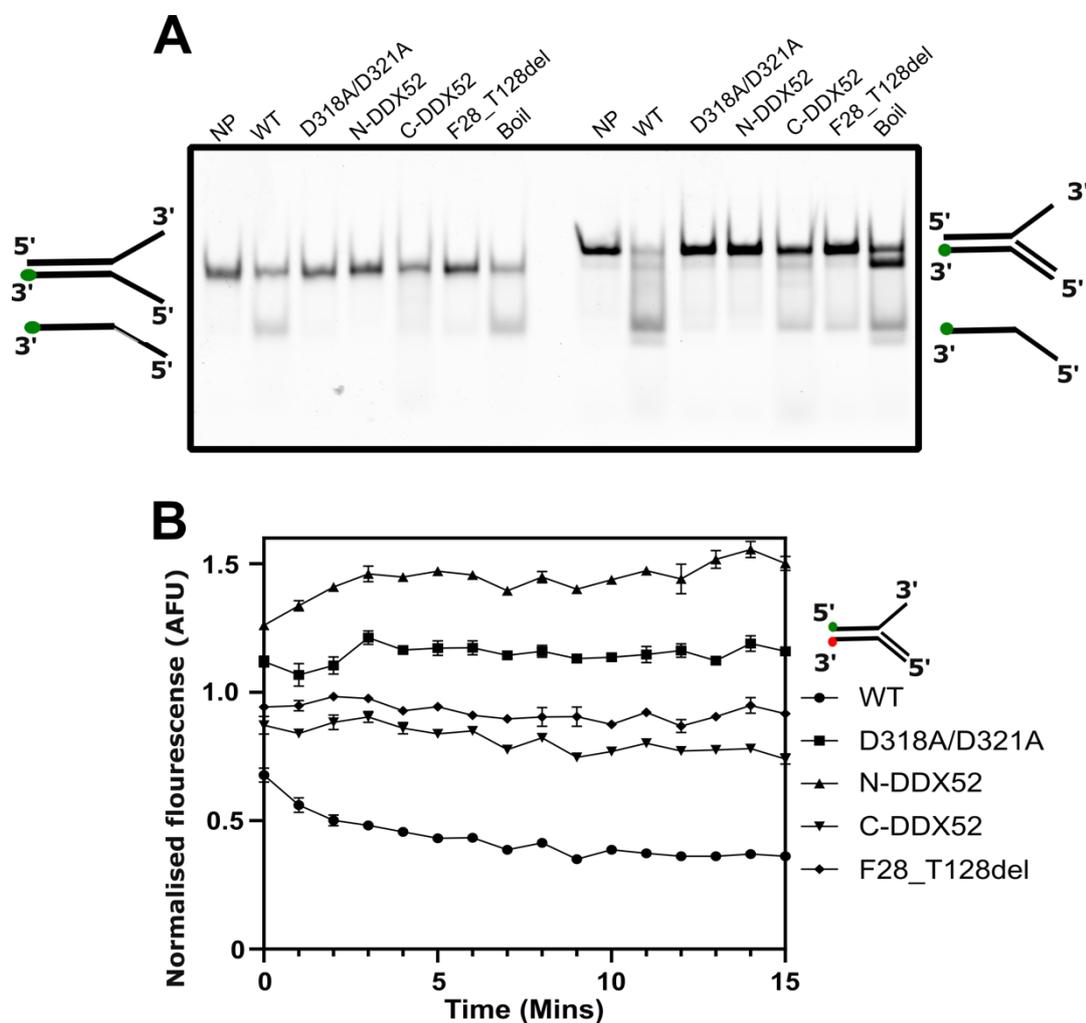


Figure 3.14: Summary helicase assays of all DDX52 mutants reveals inhibited unwinding within all mutants and apparent annealing within the N-terminus. Helicase unwinding assays were performed as described in section 2.6.1. **(A)** Processivity of DDX52 and associated mutants (500 nM) was compared on flayed duplexes (25 nM) with open and 3' flaps supplemented with

2 mM MgCl₂, 2 mM ATP, 25 mM DTT in 1 x helicase buffer in a gel-based assay. **(B)** Processivity of unwinding by DDX52 and associated mutants (500 nM) was compared on substrates (50 nM) with a 3' flap supplemented with 2 mM MgCl₂, 2 mM ATP, 10 mM DTT in 1 x helicase buffer in a FRET-based assay. Samples were run in duplicate, standardised against a no protein control and the standard error is shown.

The unwinding activity of purified DDX52 mutants was examined in both gel-based and FRET-based assays (Fig 3.14). It was noted in FRET-based assays that wild type DDX52 had already substantially unwound the duplex (~70%) by the time the first reading was taken. Both C-DDX52 and F28_T218del showed unwinding of substrates with a 3' flap in figure gel-based assays, however in FRET-assays this activity was negligible. Interestingly, both N-DDX52 and D318A, D321A mutants showed a higher FRET signal than the no protein control: in the case of N-DDX52 this was substantial and was increased to 125% that of the no protein by the time of initial reading. This indicates potential annealing whilst also suggesting that not all Cy5 and Cy3 strands are fully associated within the substrate used within the assay.

3.2.5.2. DDX52 processes D-loops but not R-loops

To further interrogate substrate preferences of DDX52, unwinding assays were conducted on different types of substrates. DDX52 showed no strong preference for unwinding DNA or RNA leading strands (Fig 3.15A), with quantification in ImageJ showing equal levels of activity on both DNA (47% unwound) and RNA (54% unwound) leading strands. No activity was seen in the absence of ATP. Processivity was also examined on a series of D-loops and R-loops (Fig. 3.15B). DDX52 was observed to process D-loops in the presence and absence of 3' flaps with no clear preference for either. Comparable activity was seen on R-loop with no flap, but activity on R-loops with a 3' flap was weaker than the corresponding D-loop. Presence of a lower band within R-loop substrates suggested potential unstable oligonucleotides or RNase contamination. A 5' flap substrate was not tested as oligonucleotides were unavailable and previous tests indicated DDX52 was not expected to load in this direction. This indicates similar activities on both nucleic acids with a preference for D-loops with 3' flaps over R-loops.

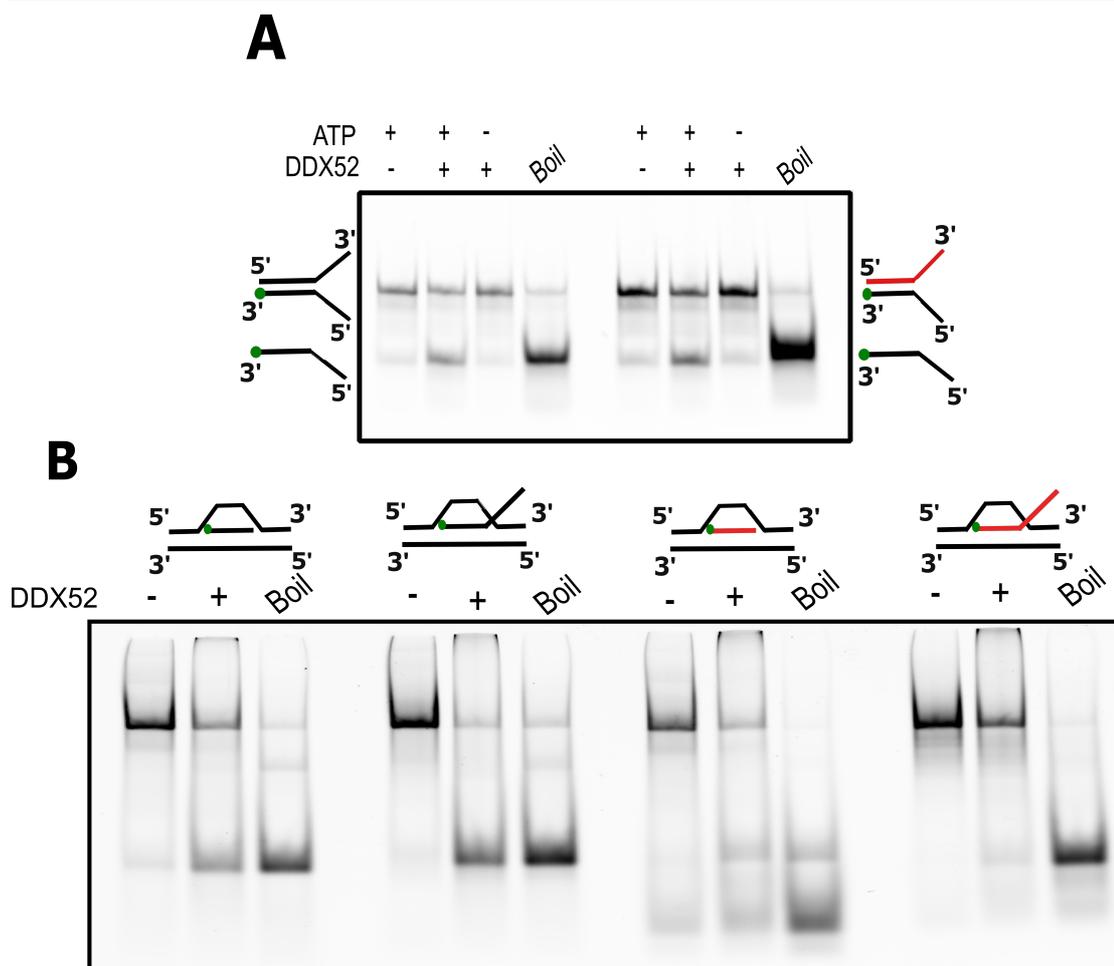


Figure 3.15: DDX52 exhibited no selective preference for RNA in unwinding assays on DNA:RNA hybrids and D/R-loops. Gel-based helicase unwinding assays were performed as described in section 2.6.1. RNA strands are indicated in red. **(A)** DDX52 (500 nM) processivity was compared on flayed duplexes consisting of DNA:DNA and DNA:RNA hybrids (25 nM) supplemented with 2 mM MgCl₂, 2 mM ATP, 25 mM DTT in 1 x helicase buffer in a gel-based assay **(B)** DDX52 (500 nM) processivity was compared on D-loops and R-loops with no flaps and 3' flaps (25 nM) supplemented with 2 mM MgCl₂, 2 mM ATP, 25 mM DTT in 1 x helicase buffer.

3.2.5.3. Unwinding by DDX52 is ATP-dependent

Having confirmed DDX52's polarity, the protein was examined to determine if unwinding was ATP-dependent. In initial assays this was carried out using ATPγs, a slow hydrolysing ATP analog. Figure 3.16A shows that in presence of ATPγs, unwinding by DDX52 on both substrates was abolished. A similar effect can also be seen in the ATP-negative lanes of Figure 3.13A. To confirm

unwinding is ATP-dependent and due to contamination within purification batches, site-directed mutagenesis was carried out on motif II (see section 2.2.4) to replace the aspartic acid residues of the DExD-box with alanine (D318A, D321A). An unwinding assay was performed comparing WT-DDX52 and DDX52^{D318A, D321A} (Fig. 3.16B). This assay showed unwinding was completely abolished on fork 4 whilst minor unwinding was observed on fork 2. These results suggest an ATP-dependent mechanism of unwinding that does not completely prohibit limited substrate unwinding within motif II mutants.

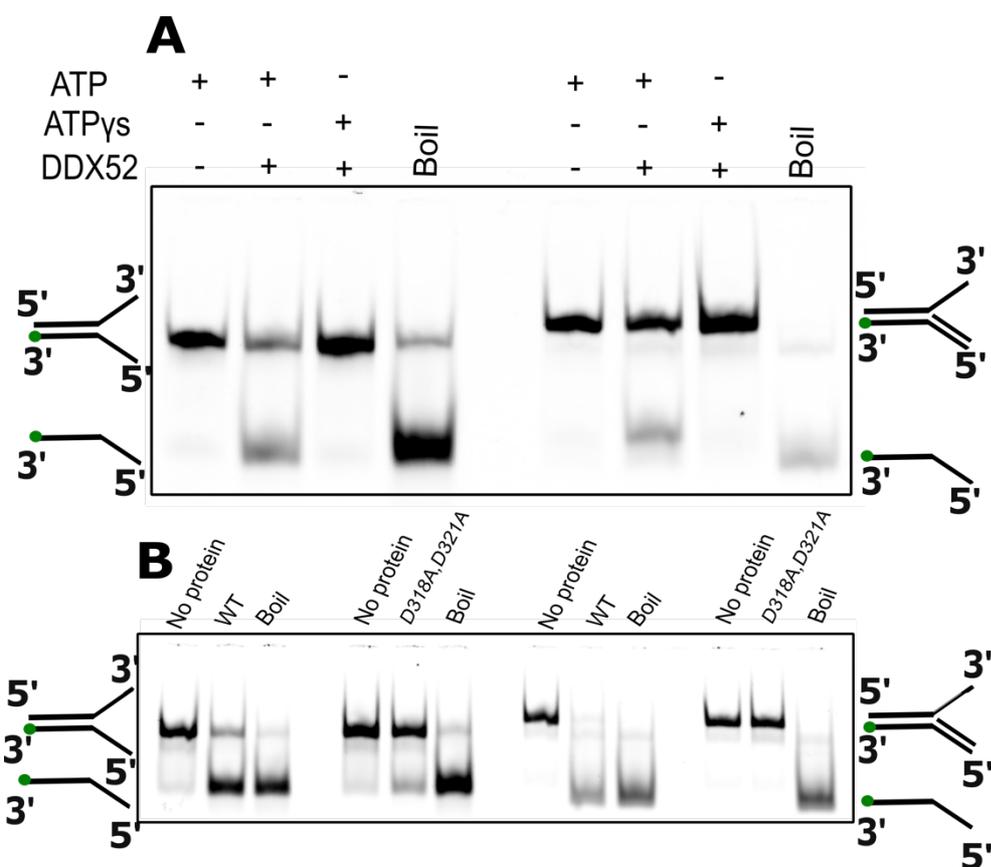


Figure 3.16: Helicase assays show unwinding by DDX52 is dependent upon ATP hydrolysis. All parts show results of DDX52 helicase reactions performed as described in section 2.6.1 within TBE 10% polyacrylamide gels. **(A)** DDX52 (500 nM) was tested on flayed duplexes supplemented with either 2 mM ATP or the slowly hydrolysable ATP analog ATP- γ -S, 2 mM MgCl₂ and 25 mM DTT in 1 x helicase buffer in a gel-based assay. **(B)** Wild type DDX52 and a motif II mutant (DDX52^{D318/321A}) (500 nM) were tested on flayed duplexes featuring an open fork and a 3' flap (25 nM) supplemented with 2 mM MgCl₂, 2 mM ATP, 25 mM DTT in 1 x helicase buffer in a gel-based assay.

3.2.5.4. DDX52 ATPase activity is more active on ssDNA than ssRNA

To examine ATPase activity of DDX52 in the presence of different nucleic acid species, a malachite green assay was carried out (Fig. 3.17). As stability of DNA:RNA complexes could not be verified these were not tested. Control wells were included to test for phosphate contamination within reagents and none was recorded. DDX52 showed a higher rate of ATP hydrolysis in the presence of ssDNA compared to ssRNA and forked DNA, with a mean phosphate liberated of ~1.5 fold that of ssRNA and ~2.5 fold higher than DDX52 in the absence of nucleic acids. It was noted that the data range for forked DNA showed high levels of standard error, particularly in comparison to the tight range seen in the absence of nucleic acids.

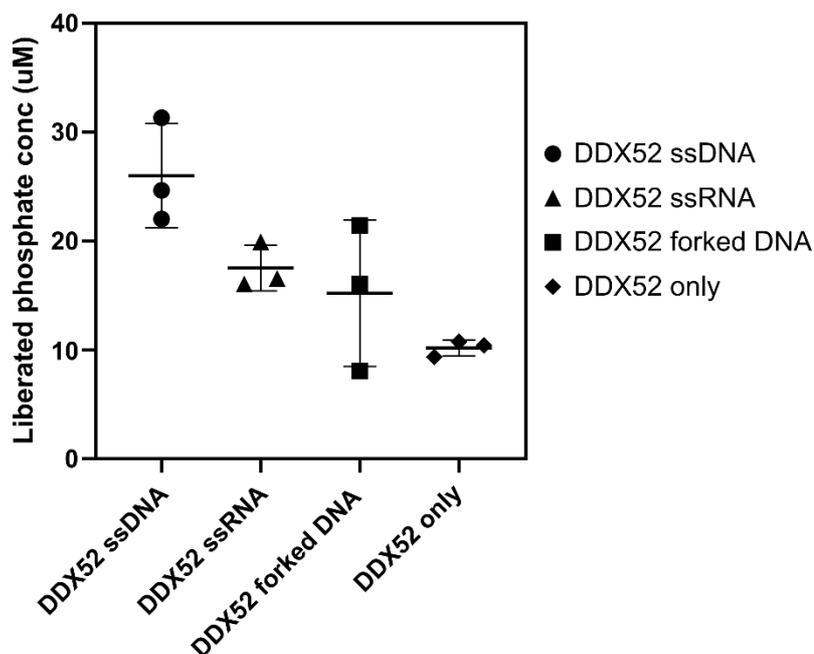


Figure 3.17: ATPase assays suggested higher ATPase activity in the presence of ssDNA vs ssRNA. Malachite green activity assays were performed as described in section 2.6.6 to measure ATPase activity of wild type DDX52 (200 nM) in the absence of nucleic acid substrates and 45-mer ssDNA, ssRNA and flayed duplex. A standard curve was generated, and concentration of liberated phosphate determined according to manufacturer instructions. Reagent only controls confirmed negligible presence of contaminating phosphatase activity.

3.2.6. Investigating DDX52 N-terminus annealase activity

3.2.6.1. DDX52 acts as an ATP-independent annealase

FRET experiments on N-DDX52 (Fig. 3.14B) appeared to show potential annealase activity. To confirm if DDX52 acted as an annealase, complementary oligonucleotides were incubated in the presence of a DDX52 concentration gradient in low ATP conditions as described in section 2.6.3. DDX52 was confirmed to act as an annealase on both DNA:DNA and DNA:RNA substrates (Fig. 3.18A). Interestingly, DDX52 appeared more effective at annealing DNA:DNA substrates than DNA:RNA substrates: achieving full annealing at concentrations of 100 nM and 200 nM, respectively. Testing in the absence of ATP and MgCl₂ (Fig. 3.18B) confirmed that annealing occurred independently of ATP. All assays showed spontaneous annealing within the no protein control, particularly within DNA:DNA samples in Fig. 3.18B. A fret-based annealing reaction (Fig. 3.18C) further confirmed DDX52 annealing activity, with a rate of annealing over double that the rate of spontaneous annealing.

3.2.6.2. DDX52 annealing activity is driven by the N-terminus

Purified DDX52 mutants were tested to determine their annealing activity in comparison to wild type protein. Comparisons between wild type, N-DDX52 and C-DDX52 demonstrated minor annealing activity within C-DDX52 with only trace levels of annealing visible at concentrations above 400 nM DDX52 (Fig. 3.19B). In contrast, the N-terminus truncation (Fig. 3.19A) appeared to show hyperactive annealing, achieving the same maximum levels of annealing of wild type DDX52 at half the concentration and reaching almost complete annealing of substrate at a concentration of 800 nM – over twice the level of annealing seen within wild type DDX52 at the same concentration. This suggests that annealing activity is localised to the N-terminus/RecA1 domain of DDX52.

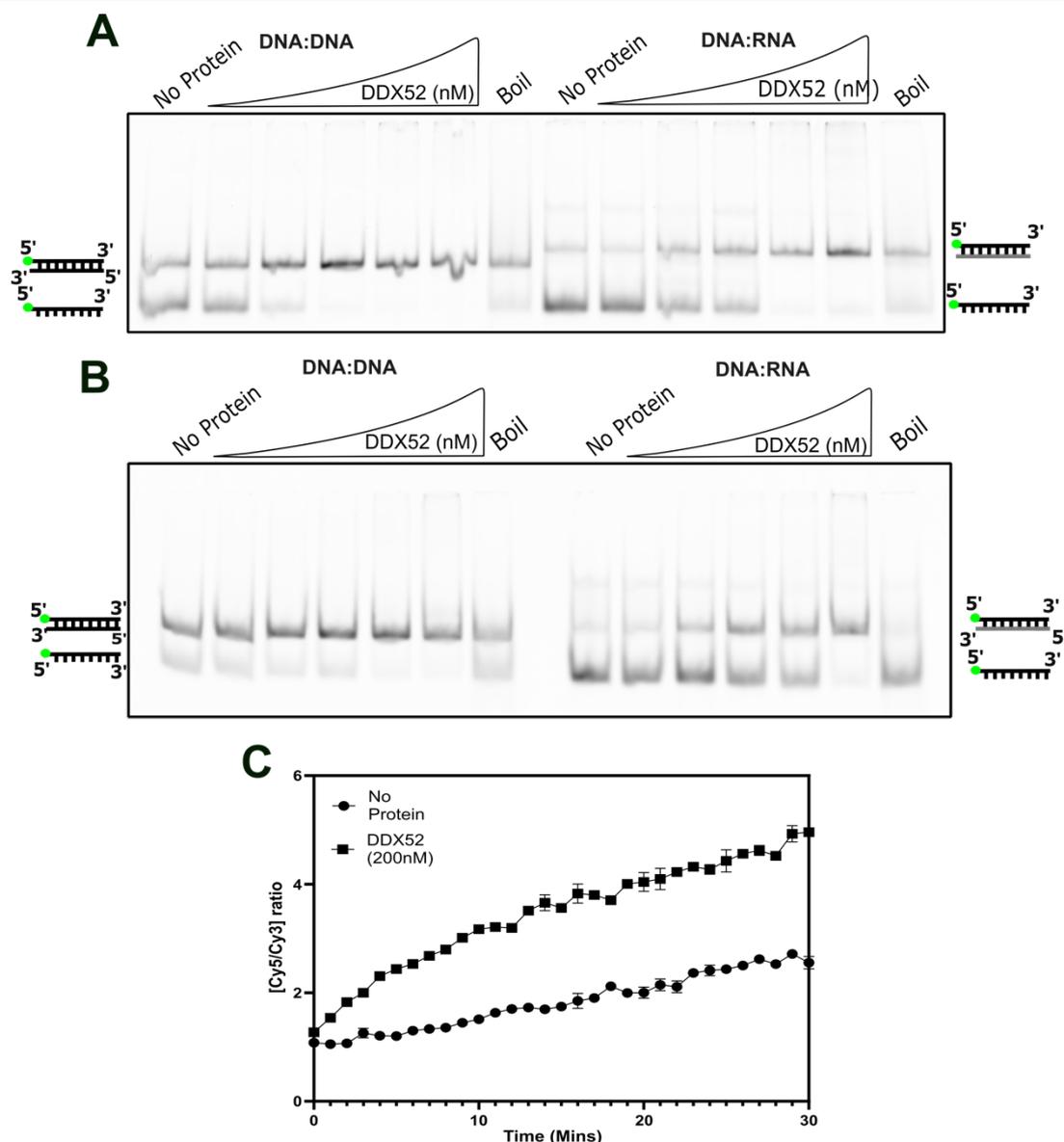


Figure 3.18: Annealing reactions confirmed DDX52 activities as an ATP-independent annealase on DNA:DNA and DNA:RNA as a function of concentration. A + B were performed as described in section 2.6.3 on 10% TBE polyacrylamide gels and used protein concentrations of 25, 50, 100, 200, 500 nM. **(A)** DNA:DNA and DNA:RNA annealing was measured in an assay containing 1 mM ATP, 1 mM MgCl₂ and 25 mM DTT with annealing increasing as a function of concentration. **(B)** DNA:DNA and DNA:RNA annealing was confirmed to be ATP-independent in an assay depleted of ATP and Mg²⁺. **(C)** Annealing was further confirmed in FRET-based analysis of DDX52 (200 nM) as described in section 2.6.4.

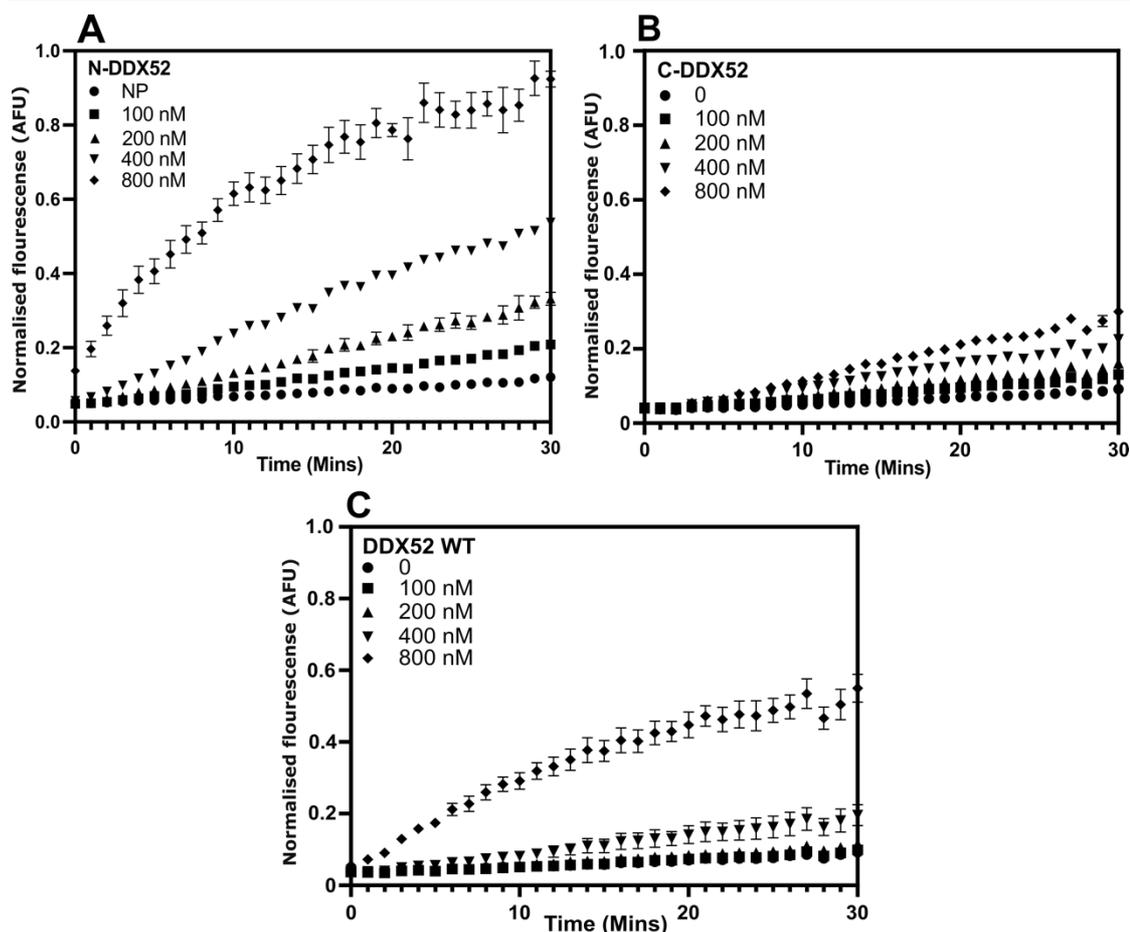


Figure 3.19: FRET-based annealing reactions confirm DDX52 annealing activity is localised to the N-terminus. A, B + C) Annealing of N-DDX52, C-DDX52 and WT-DDX52 using a FRET-based method. Reactions were carried out in duplicate using 50 nM of complementary ssDNA oligonucleotides supplemented with 1 mM MgCl₂, 1 mM ATP and 10 mM DTT in 1 x helicase buffer for thirty minutes at 37 °C. Data was normalised against a previously annealed control.

Protein	Ta50 (800 nM)	Ta50 (400 nM)
Wild-type	29 minutes	N/A
N-terminal	8 minutes	28 minutes
C-terminal	N/A	N/A

Table 3.1. DDX52 N-terminal truncation annealed DNA at a rate over two-fold that of comparative wild-type concentrations in FRET-based annealing reactions. Data is presented as the time taken to anneal 50% of DNA relative to previously annealed control. Table is complementary to data presented in Figure 3.19.

In annealing assays conducted in the absence of magnesium, additional bands were noted above the fully annealed substrate at concentrations of N-DDX52 above 100 nM (Fig. 3.20). Upon assessing the annealing activity on the other mutants, similar band formation was noted within the motif II mutants (D318A, D321A). This additional band was absent within other mutants and the wild type. IDR loop deletions show a lack of annealing (F28_T128del) implicating this as playing a key role within the annealing activity of DDX52. Interestingly in comparison to the FRET data in Fig. 3.19B, the C-terminus appears proficient at annealing.

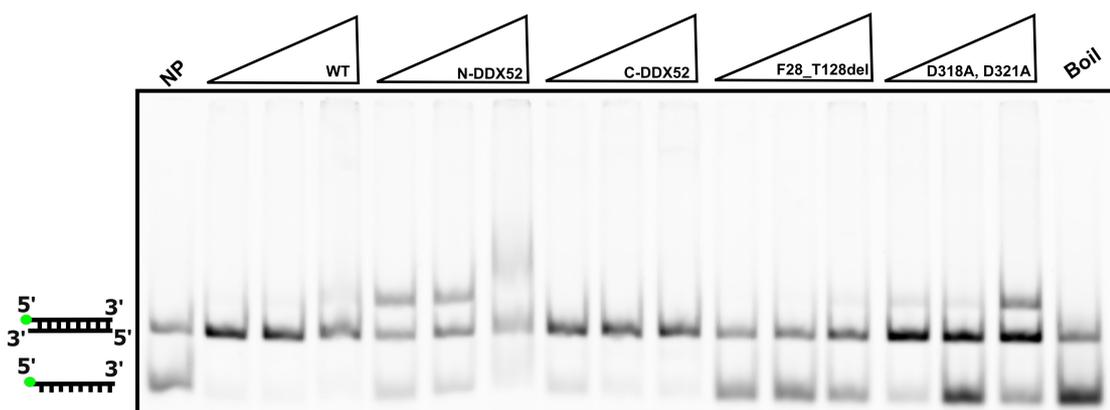


Figure 3.20: Annealing assays confirm all mutants apart from F28_T18del mutant possess annealing activities. Assays were performed using 15 nM complementary DNA oligonucleotides supplemented with 25 mM DTT in 1 x helicase buffer as described in section 2.6.3. Protein concentrations of 100, 200 and 500 nM were used.

3.2.7. DDX52 binding protects DNA from nucleases

Following on from data seen in section 3.2.4, it was hypothesised that DDX52 was forming filament like structures across nucleic acids that might function to block the activity of nucleases. To examine this, nuclease assays were conducted as described in section 2.6.7 on single-stranded DNA and RNA that had been preincubated with DDX52. As this was a novel assay, initially S1 nuclease activity on DNA was determined by running a series of samples to determine appropriate incubation time and units (Fig. 3.21A). Using this data, assay conditions were optimised to deliver a concentration and incubation time for S1 nuclease appropriate for optimum digestion of DNA (20 units, 30 minutes). It was noted from data in Fig. 3.21B and literature that S1 nuclease demonstrated weaker activity on ssRNA compared to ssDNA (325). To better examine if DDX52 showed a protective function on RNA, the P1 nuclease was likewise examined and optimised (Fig. 3.22A). Optimum reaction conditions for digestion were determined to be the same as S1 (20 units, 30 minutes).

Data indicates that addition of 20 units of S1 and P1 at an incubation time of 30 minutes resulted in almost complete digestion of the ssDNA or ssRNA in their respective assay. In the presence of increasing concentrations of DDX52, the band intensity of DNA and RNA was noticeably increased supporting the hypothesis that DDX52 binding protects DNA and RNA from both S1 and P1 degradation (Fig. 3.21A and 3.22B). It was noted, however, that even at the highest concentration of DDX52 a substantial decrease in intensity was noted in comparison to the no protein control suggesting that DDX52 does not fully protect nucleic acids or that a higher concentration is necessary to completely protect DNA.

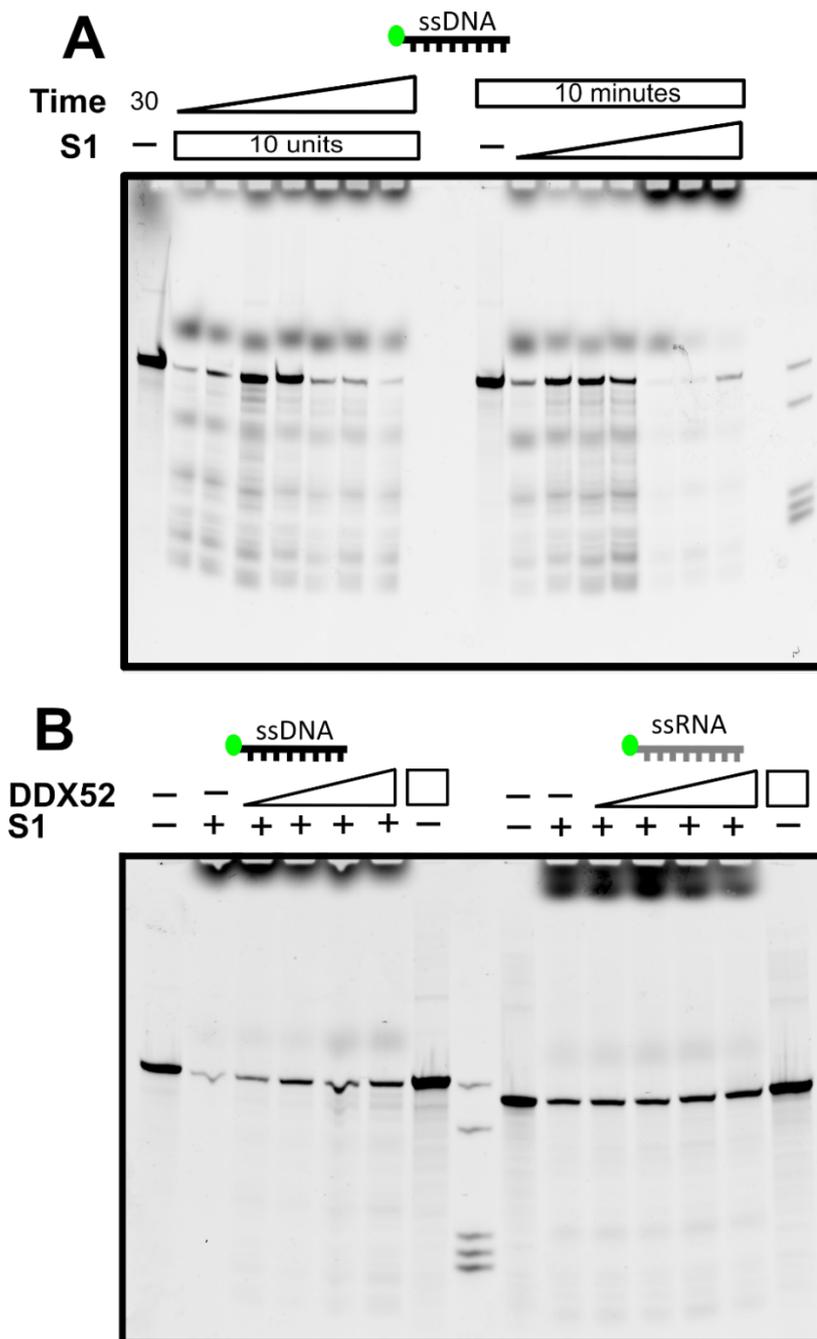


Figure 3.21: Nuclease protection assays show DDX52 protects ssDNA. (A) 5'-cy5-labelled ssDNA (50 bp) was incubated with S1 nuclease at increasing units or incubation time as indicated. Units of S1 were 1, 2, 5 10, 20, 50 and 100 units and time increments were 1, 2, 3, 5, 10, 20 and 30 minutes. Reactions were deproteinized and analysed by denaturing gel electrophoresis. **(B)** 5'-cy5-labelled ssDNA (50 bp) and ssRNA (45 bp) were preincubated with increasing concentrations of DDX52 prior to addition of 20 units S1 nuclease. After 30 minutes, reactions were deproteinized and analysed on a 15% denaturing gel. DDX52 protein concentrations were 100, 200, 500 and 1000 nM.

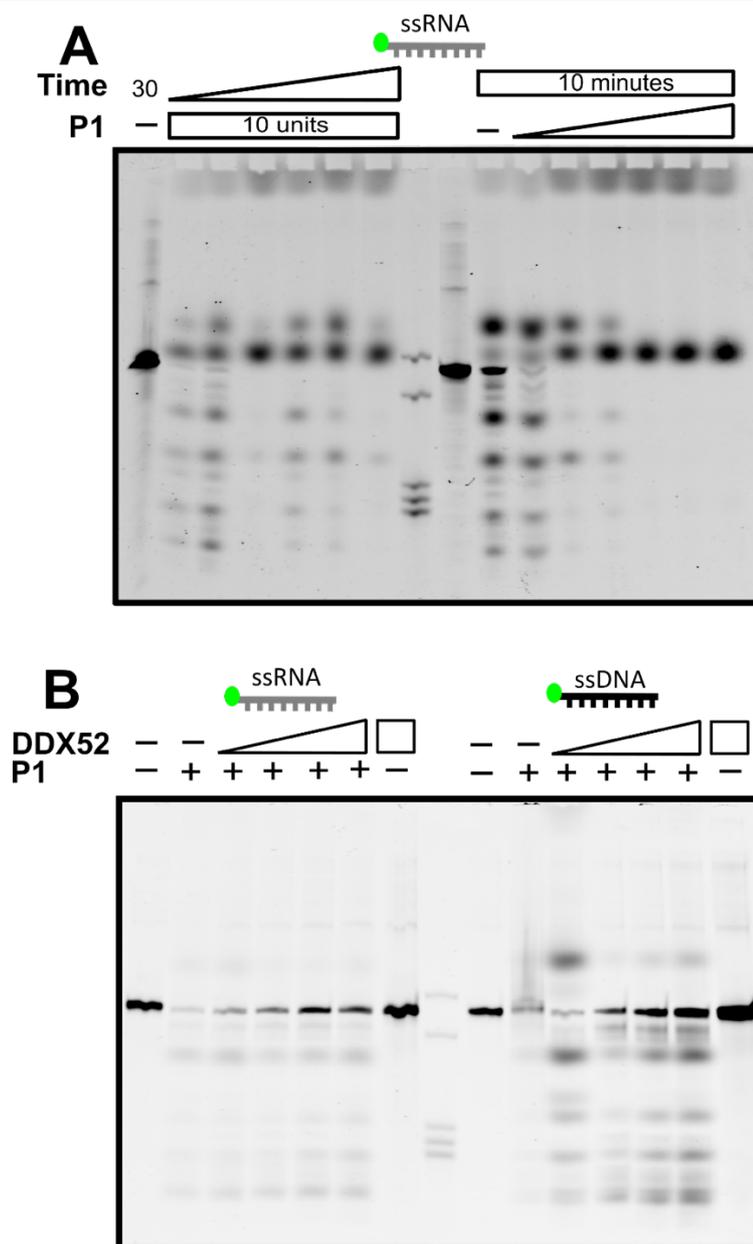


Figure 3.22. Nuclease protection assays show that DDX52 protects ssRNA and ssDNA. (A) 5'-cy5-labelled ssRNA (45 bp) was incubated with P1 nuclease at increasing units or incubation time as indicated. Units of S1 were 1, 2, 5, 10, 20, 50 and 100 units and time increments were 1, 2, 5, 10, 20 and 30 minutes. Reactions were deproteinized and analysed on a 15% denaturing gel. (B) 5'-cy5-labelled ssRNA (45 bp) and ssDNA (45 bp) were preincubated with increasing concentrations of DDX52 prior to the addition of 20 units P1 nuclease. After 30 minutes, reactions were deproteinized and analysed on a 15% denaturing gel. DDX52 protein concentrations were 100, 200, 500 and 1000 nM.

3.3. Discussion

As of the time of writing, no previous in vitro biochemistry data has been reported for DDX52. Like other DExD-box proteins, DDX52 is a putative RNA helicase thus an affinity, if not exclusivity, for RNA was anticipated. However, increasing research is showing duality of binding of several of these proteins to DNA species. Notably, for DDX52, it has been reported as being phosphorylated in response to DNA damage and posited to play a role in transcriptional regulation by way of reported interactions with C-myc, implying function beyond its predicted role within ribosome biogenesis (98, 103).

We thus hypothesised that DDX52 was a candidate for a multi-functional DExD-box helicase that might interact with DNA in addition to RNA. This was supported by previous work by Buckley et al (2020) that noted similarities between DDX52 and the Lhr, a helicase postulated to be active at stalled or broken replication forks. In this aim the work was largely successful, delivering insights into DDX52s interactions with DNA and dissecting this down to the sub-sections of the protein.

3.3.1. DDX52 possesses distinct structural differences compared to Lhr and Rok1

It was previously reported that human DDX52 shared structural similarities with *Mycobacterium smegmatis* Lhr (104) . This was explored (Fig. 3.3) and whilst structural similarities were noted within the core RecA-like domains, the N and C termini were distinctly different. This is relevant as it has been suggested that these extensions are what confer the functional and substrate specificities of DExD-boxes (40). It is considered that, due to the lack of a complete structural model for DDX52, inferences are likely to be imprecise. However, considering key differences between DDX52 and Lhr - notably being DExD-box and Ski2-like helicases, respectively - it is expected that they should demonstrate different mechanisms of unwinding DNA, with Lhr and DDX52 being processive and nonprocessive helicases, respectively. It is thus considered unlikely that they are related.

There is also reported homology between DDX52 and the yeast helicase Rok1. Conservation between the two proteins was noted within the core

RecA1/RecA2 domains and the C-terminus, but the N-terminus of DDX52 possesses an extended intrinsically disordered region (IDR) taking on the appearance of a loop of 100 amino acids that is absent within Rok1. As mentioned previously, these termini are likely the source of functional and substrate specificities within the family. In particular, IDRs are frequently associated with binding interaction partners, forming multivalent interactions and are enriched within transcription factors (326–328). ANCHOR2 scores (Fig. 3.2) suggest that this domain is not brought into order with an interaction partner. As a result of poor homology within the N-terminus, it is inconclusive if DDX52 and Rok1 are true homologs. It is also noted that the original study suggesting the two were homologous used a noncanonical sequence of DDX52 shorter than the now accepted sequence. If the two proteins do share functional homology, it is hypothesised that N-terminus extensions might deliver additional functionalities not seen within Rok1.

3.3.2. Binding characteristics of DDX52

3.3.2.1. DDX52 N and C-termini possess opposing binding mechanisms

Using electro-mobility shift assays (EMSAs) we performed an analysis of the binding activity of DDX52 on ssDNA, ssRNA and a flayed duplex. Results (section 3.2.4) clearly showed interaction with each nucleic acid species tested, with no apparent preference for any species. However, banding patterns were unsatisfactory for analysis with a prominent smearing effect noted, thought to be a result of poor association of the protein with the DNA substrates. Across all assays, it was observed that as protein concentration increased, Cy5 signals appeared within the wells of the gel, indicative of DNA ‘trapped’ in the well. Banding patterns similar to this are characteristic of the formation of multiple nucleoprotein species; a feature frequently seen within proteins involved within chromatin remodelling (329, 330).

DDX52 mutants were tested and clear separation of function between wild type and C-terminus truncations were observed. Stable banding patterns were noted within assays conducted on C-DDX52 incubated with flayed DNA duplexes when compared with wild type protein (Fig. 3.11B). Testing of the N-terminus truncation (Fig. 3.11A) replicated the aggregation seen within the well observed in the wild type, confirming this binding is localised to the N-terminus. In

addition, the N-terminus exhibited higher levels of Cy5 signal in the well than the wild type. It is hypothesised that the two binding mechanisms of both sides of the protein may act in competition with each other, notably within the roles of annealing and unwinding seen with assay results.

It was hypothesised that DDX52 was forming filaments along the nucleic acids, with the feature believed to contribute to this filamentation being the intrinsically disordered region (IDR); the 'loop' seen within Alphafold predictions of DDX52. IDR loops such as this have been implicated in forming multivalent interactions with other IDRs, potentially giving rise to multiple nucleoprotein species and the banding patterns seen here (328). In F28_T128del mutants where this was deleted, binding appeared almost completely abolished. It is considered that conflicting mechanisms of binding may be regulated by an as yet unidentified accessory factor; such as the interplay between Rad52 and BRCA2 to displace RPA, allowing formation of Rad51 filaments (331).

3.3.2.2. DDX52 exhibits a higher affinity for DNA than RNA

Data from EMSAs were less than satisfactory for direct comparisons of nucleic acid affinity. An anisotropy-based method was developed and successfully employed. As indicated by EMSA data, DDX52 was weakly associative with nucleic acid species with a high dissociation constant. However, it was noted that it appeared to have a higher affinity for DNA with a K_d of 328 nM: over half that of the K_d of 784 nM recorded for RNA. This challenges the preconception that processing RNA is the primary function of DExD-boxes and is relatively unique amongst DExD-boxes, with helicases studied previously showing expected affinities for RNA (332). This higher affinity for DNA supports a proposed role for DDX52 as a helicase with dual functionality, whilst also agreeing with results seen in later ATPase assays and assays comparing processivity of D-loops vs R-loops that appear to indicate DDX52 possessing more activity on DNA.

3.3.2.3. DDX52 protects nucleic acids from degradation

It was hypothesised based upon banding patterns seen within EMSAs (Figs. 3.10-3.11) that DDX52 may form filaments along ssDNA strands, protecting them from degradation. To examine this, a nuclease protection assay was optimised and employed to examine the effects of increasing concentration of DDX52 on the degradation of ssDNA and ssRNA. Clear increases in band concentration were noted as DDX52 concentrations increased, indicating DDX52 was inhibiting degradation of ssDNA. Initial tests showed no clear change in RNA levels compared to the S1 only control, suggesting DDX52 only has protective capacity for ssDNA. However, it was considered that S1 nuclease is 5x more active on ssDNA than ssRNA (325). A further assay using P1 nuclease was developed and optimised for the study of DDX52 interactions with RNA and confirmed that DDX52 binding also protected RNA from degradation.

Proteins previously shown to protect nucleic acids from nuclease activity have been linked with roles within DNA replication and repair, notably RPA and hSSB1 + 2 in humans and SSB in *E. coli* (333–335). Similarly, yeast Rad52 and Rad59 have been shown to protect DNA within nuclease protection assays similar to those performed here (336). This supports a potential role for DDX52 with DNA replication and repair. Alternatively, nuclease protection effects have previously been demonstrated as an effect of the RSC chromatin remodelling complex enveloping nucleosomes, with human SWI/SNF showing similar banding patterns in EMSAs as seen for wild type DDX52 within this study (337, 338).

3.3.3. DDX52 acts as both a helicase and annealase

3.3.3.1. DDX52 is a poorly proficient helicase

The reaction conditions of DDX52 were optimised and an optimum MgCl₂:ATP ratio of 1:1 was determined (Fig. 3.9). A concentration gradient of DDX52 was used to investigate unwinding as a function of concentration. DDX52 showed negligible unwinding activity up until 500 nM. This is poor proficiency, with the putative homolog Lhr showing 100 nM was sufficient to get unwinding (104). As reported in bioinformatic analysis, Lhr is a processive DExH-box helicase whilst DDX52 as a DExD-box helicase is anticipated to unwind on a basis of local

strand separation, unwinding a few nucleotides at a pair before using the energy of ATP to release and be recycled (63, 339). As a result poor processivity was anticipated, with previous work reporting that unwinding by DExD-box helicases requires a large excess of protein to substrate (41). As a result, the concentration of DDX52 is a limiting factor, with multiple protein:DNA complexes reported as necessary for full separation of longer duplexes (65).

Length of substrate is also considered a limited factor, with previous studies demonstrating effective unwinding using duplexes of than >12 NT. Studies with yeast DbpA have shown that longer duplexes results in futile ATP turnovers as a result of competitive reannealing of the duplex (340). However, for the purposes of this study full unwinding was considered unnecessary. An equilibrium existed between unwinding and spontaneous annealing of the substrate. Unwinding assays can be improved through use of unlabelled 'trap' DNA, however early into this study concerns arose that excess ssDNA might inhibit DDX52 unwinding so this was not used within assays.

3.3.3.2. DDX52 is a 3' to 5' ATP-dependent helicase

As a superfamily II helicase, it was hypothesised that DDX52 would possess a 3' to 5' polarity of unwinding. Both gel-based and FRET-based assays supported this hypothesis, with DDX52 only capable of acting on substrates possessing a 3' flap. Assays using ATP- γ -S, a poorly hydrolysable ATP analog, suggested that DDX52 was an ATP-dependent helicase. To confirm this, a motif II mutant (DDX52^{D318A, D321A}) was purified following double mutations previously used to study DDX3X by Garbelli et al (2011). This mutant also exhibited no unwinding and suggested that results were unlikely to the result of a contaminant. It was noted that the motif II mutant showed limited activity on substrates possessing open flaps. This is in line with expectations based upon other members of the DExD box family with ATP hydrolysis only necessary for the recycling of the helicase; as DDX52 is loaded at an excess in this reaction, partial unwinding is likely a concentration effect of DDX52 effecting local strand separation as noted by Rogers et al (1999) (71).

Malachite green assays noted that DDX52 exhibited more ATPase activity on ssDNA than ssRNA. As with the increased binding affinity for DNA as opposed

to RNA, this challenges the preconception that DDX52 is primarily an RNA helicase. In addition, the range of the data might suggest an inconsistent method of hydrolysis on forked DNA compared to single stranded substrates. A consideration is made that it has been noted in previous research that properties of RNA used can affect the biochemical characteristics of DExD-box ATPase activity (341).

3.3.3.3. DDX52 processes DNA:RNA hybrid substrates

Assays carried out showed no clear difference in unwinding was seen between DNA:DNA and DNA:RNA hybrids, with both substrates unwound to roughly 50%. Subsequent assays carried out demonstrated that DDX52 was able to proficiently unwind D-loops and R-loops, even in the absence of 3' flaps. A slight preference was noted for D-loops possessing 3' flaps over their R-loop counterpart, potentially supporting previous higher affinities for DNA seen in binding and ATPase assays. DEAD box activity on R-loops has been previously reported within DDX5, DDX17 and DDX21 (342, 343). DHX9 has been reported to act on D-loops, but possessed a preference for R-loops and DNA-based G-quadruplexes, suggesting a preference for D-loops is quite novel (344). However further evidence is necessary to confirm this preference as no in substrates possessing no flap. This data implies a role of DDX52 within the processing of D-loops and R-loop and thus a role within the maintenance of genome stability.

3.3.3.4. DDX52 possesses annealing activity and promotes the formation of DNA:DNA and DNA:RNA duplexes

Annealing assays were carried out (Fig. 3.18) and DDX52 annealed DNA:DNA and DNA:RNA in an ATP-independent manner. Annealing activity was concentration dependent and appeared biased towards DNA:DNA annealing, but this could be a consequence of reduced stability of RNA:DNA hybrid duplexes (345). It is noted that substantial levels of annealing in the no protein control indicates spontaneous annealing of duplexes irrespective of protein. This effect has been seen within other studies and it is considered acceptable to compare the spontaneous rate with reaction mixes containing protein (346, 347). Notably, DDX52s reported yeast homolog Rok1 also showed duplex annealing activities (114).

DDX52 strand annealing still proceeds with low levels of ATP present, interesting given the contradictory activities of annealing and unwinding - dual activities in unwinding and annealing have previously been reported within the helicases *hel308* and *Polθ*, with helicase activity of *Polθ* only recently identified as a result of annealing activity masking its unwinding (348, 349). Studies in *Ded1* suggest that the opposing activities are modulated by the ratio of ADP and ATP concentrations, with the ATP-dependent closure of the RecA domains prohibiting dsDNA/dsRNA accommodation (114, 350). Alternatively this could be regulated in part by a cofactor, with studies on *Rok1* showing that annealing was increased in the presence of the cofactor *Rrp5* (114). It is considered that these contradictory activities may elaborate on DDX52s poor proficiency as a helicase.

Annealing assays on DDX52 mutants demonstrated that N-terminus truncations of DDX52 were hyperactive as annealases, achieving almost 100% annealing in FRET-based assays and over twice that seen in wild type DDX52 (Fig. 3.19). Poor activity was seen within the C-terminus domain when examined in FRET assays, appearing to confirm annealing is localised to the N-terminus – however it is noted that gel-based assays showed C-DDX52 as being more proficient. It is suggested that negatively charged D/E residues within the N-terminus may act as nucleic acid mimics to facilitate this, particularly in residues 72-117 (21.7% E), with a similar effect reported within *S. cerevisiae* DExD-box *Dbp6* (351). Alternatively, multivalent interactions between DDX52 IDRs may act to bring complementary oligonucleotides together, with annealing shown to be abolished within F28_T128del mutants (328).

Interestingly, incubation of N-DDX52 and motif II with DNA in the absence of magnesium resulted in generation of a third, higher band not corresponding to either single-stranded or double-stranded substrate (Fig 3.19-3.20). The identity of this substrate is unknown and attempts to extract it from the gel were unsuccessful. Magnesium has been shown to possess an inhibitory role on the formation of Topo IV–DNA complexes when used as a cofactor in reactions, whilst motif II is involved in the co-ordination of magnesium ions (352, 353). Alternatively it is suggested that these mutants may prohibit DDX52 strand separation resulting in accumulation of this substrate, with motif II mutants

previously found to impact conformational changes associated with activity (354).

3.4. Conclusions

In this chapter, we have presented novel activities of the DExD-box helicase DDX52 including the first evidence of its biochemical properties. In addition to expected properties of superfamily II helicases, unwinding in an ATP-dependent 3' to 5' direction, we record interesting activities in processing replication forks and D-loops and R-loops, potentially implicating DDX52 within DNA replication and repair and the maintenance of genome stability. Most notably, we present the first evidence that the N-terminus of DDX52 acts as a DNA:DNA and DNA:RNA annealase. RNA annealing activity has previously been seen within other DExD-box helicases with Rok1, the putative DDX52 yeast homolog, shown to catalyse RNA strand annealing however the DNA annealing activities of these proteins is not present, not reported or has been overlooked (114). Hence our results for DDX52 represent a novel activity.

It is proposed that contradictory helicase and annealase activities of DDX52 identified within this study are likely to be regulated by accessory factors, with the annealing activity of Rok1 regulated by the cofactor Rrp5: it would be interesting to investigate potential interactions between DDX52 and the human homolog of Rrp5, PDCD11. Finally, regarding DDX52s potential as a dual-functionality helicase, evidence showing an apparent higher affinity for DNA suggests additional biological roles to its canonical predicted role within rRNA processing. It is hypothesised this may be within chromatin remodelling or transcriptional regulation, supported by banding patterns seen within EMSAs and results within nuclease protection assays. High throughput pulldown experiments have previously revealed DDX52 interactions with the histone associated cluster HIST1H4A, furthering this implication, whilst reported promotion of c-myc expression by DDX52 implies a potential role within regulating transcription (355, 356). However further work is needed to investigate these potential additional roles.

3.5. Future Research

Within this research, a novel IDR loop was identified within the N-terminus site that was associated with DDX52 activity as an annealase. Further research should further examine this region and identify any highly conserved residues for further investigation. The final suggests that DDX52 may play a role within chromatin remodelling and transcriptional regulation – further research should examine this by combining biochemical results with results gained from a cell culture model, notably using techniques of chromatin immunoprecipitation to examine DDX52 interactions with chromatin. Finally, further research should explore the impact of cofactors on the activity of DDX52 including PCD11, RPA and additional factors of interest such as HIST1H4A, c-myc and other recombinases; in particular if these proteins regulate the unwinding and annealing activities of DDX52.

Chapter 4: Investigating DDX52 within tumour cells

4.1. Introduction

4.1.1. DDX52 within the cell

Whilst the biochemistry of DDX52 has not been studied, effects of DDX52 in vivo have been documented to a limited extent. Studies within zebrafish have implicated it within a role within 47S rRNA synthesis and suspending growth, however the authors note the complexities of determining if this is due to its role within pre-rRNA synthesis or a result of interaction with specific mRNA transcripts (357). Studies within the Chinese mitten crab implied a role of DDX52 within spermatogenesis, whilst also participating within microtubule and P-body regulation (358). *DDX52* SNPs have also been associated with irregularities within the folate metabolism pathway (359).

4.1.2. Phenotypes of DDX52

DDX52 has been implicated within several subsets of cancers and disease. Previous research has shown that knockdown of DDX52 suppresses proliferation of both malignant melanoma and prostate cancer cells through regulation of c-myc, identifying it as a potential therapeutic target within these diseases (97, 98). Analysis of DDX52 expression as a function of mortality within cancer has associated overexpression of DDX52 as unfavourable and favourable prognostic markers within liver and colorectal cancers, respectively (83). SNPs within *DDX52* have been associated with bone mineral density within men (99). siRNA screens identified that DDX52 played inhibitory roles within myxoma virus replication and classed it as antiviral (100). Contrary, siRNA knockdowns of DDX52 decreased production of infectious HIV-1 particles implicating it as a cofactor for HIV-1 replication (101).

Very recent research identified *DDX52* as a salient gene within idiopathic pulmonary arterial hypertension (360). Notably, *DDX52* has commonly been noted as a gene deleted within 17q12 deletion syndrome: a copy number variant (CNV) disorder associated with neurodevelopmental disorders, diabetes and

renal cysts (361–363). Elucidating the role of *DDX52* within symptoms is difficult as a result of scarce characterisation but robust expression of *DDX52* in early embryonic development, along with four other genes affected by 17q12Del, has led to suggestions these genes may play a role within early expression and patterning events (362). This same group of genes was noted within a 17q12 case study that presented novel manifestations of Marfanoid symptoms and in cases presenting macrocephaly, lending support to this hypothesis (364, 365).

4.1.3. Aims and objectives of research

Whilst studies have associated *DDX52* with several subsets of disease, a number of these are from high throughput screens and a suitable gap in research was identified in exploring potential phenotypes within tumour cell lines. In addition, associations of *DDX52* with cancer and the diverse roles of DExD-box proteins across a wide range of pathways highlights the importance of exploring and validating phenotypes further. Having successfully tested the biochemistry of *DDX52* using recombinant purified protein, we were interested in characterising it within a human cell model. As no methods for gene editing or studying phenotypes were developed within our lab, this chapter will focus on the development and testing of workflows for these and establishing initial phenotypes where appropriate. The objectives were as such thus:

- 1) To test and implement an effective gene-editing pipeline for generation of CRISPR-Cas9 edited cell lines.
- 2) Test the on-target and off-target effects of CRISPR-Cas9 editing.
- 3) Establish phenotypes for the line; notably, impacts upon cellular proliferation in addition to susceptibility of edited cell lines to a range of drugs and DNA damaging agents.

4.2. Results

4.2.1. Development and optimisation of a gene-editing pipeline for human cells

As no method was incorporated within the lab for gene editing human cell lines, methods required optimization and testing. Guide RNAs used to target *DDX52* consisted of guides created using IDT's tool and through probing of the human

genome browser. Several transfection reagents, methods and gRNAs were tested for efficiency of gene editing.

4.2.1.1 Purification of recombinant Cas9 for gene editing

To purify Cas9 for use in downstream gene editing experiments, the plasmid pAC29 was transformed and expressed in *E. coli* BL21-AI. Following purification using a HisTrap HP His-tag column, Cas9 was successfully identified (Fig 4.1A) and Cas9 positive fractions dialysed to remove imidazole and exchange the protein preparation into a low-salt containing buffer. To separate protein from other DNA-binding proteins and/or DNA, further purification was carried out via ion-exchange chromatography using a HiTrap™ Heparin HP column and Cas9 successfully eluted (Fig 4.1B). To separate Cas9 from lower molecular weight proteins, it was successfully further purified through size exclusion chromatography using a HiPrep 16/60 Sephacryl S-300 HR column (Fig 4.1C). Cas9 positive fractions were then dialysed into storage buffer, concentrated and intact Cas9 successfully identified via SDS-PAGE (4.1D). Less concentrated contaminants were seen in a molecular weight range from 95 kDa -130 kDa.

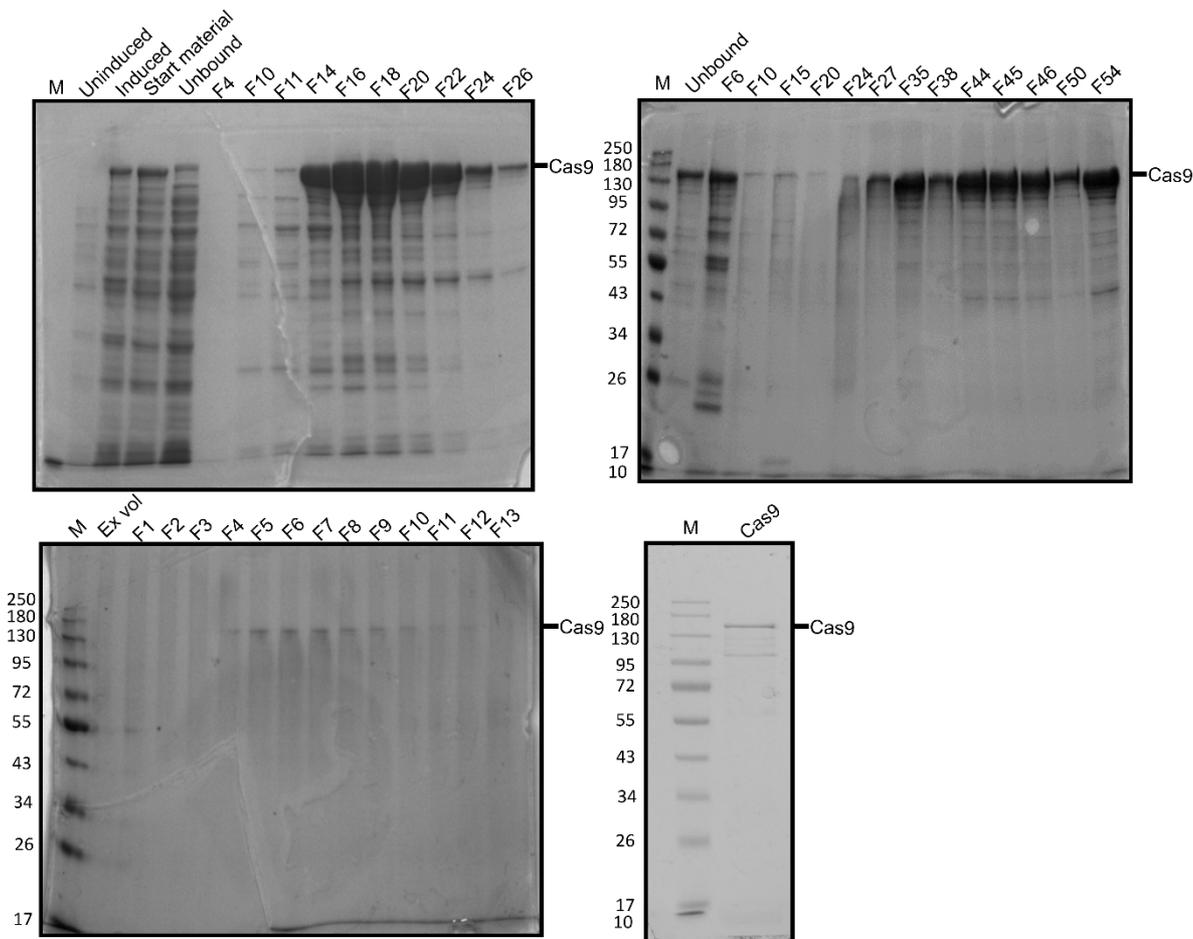


Figure 4.1. Purification of recombinant NLS-Cas9. NLS-Cas9 was overexpressed in the BL21-AI strain of *E. coli*. Expected molecular weight of NLS-Ca9 was 164.5 kDa. **(A)** Protein lysate was loaded onto a HisTrap HP His-tag column and purified using an imidazole gradient over 5 column volumes. NLS-Cas9 positive fractions were pooled and dialysed overnight at 4 °C. **(B)** Sample was passed down a HiTrap™ Heparin HP column and eluted over a gradient of 0.15 -1 M NaCl. NLS-Cas9 positive fractions were pooled and dialysed into storage buffer overnight at 4 °C. **(C)** Sample was further purified through size exclusion chromatography using a HiPrep 16/60 Sephacryl S-300 HR column. NLS-Cas9 positive fractions were pooled and dialysed into storage buffer overnight at 4 °C. **(D)** Purified protein was stored and intact NLS-Cas9 confirmed via SDS-PAGE.

4.2.2.2. Forward transfection and CRISPRMAX protocols showed the highest transfection efficiencies

For transfection trials, a plasmid optimized for human expression of GFP (pEGFP) was used to assess transfection using an Axiovert S100 microscope combined with HBO 100 Microscope Illuminating system. Two methods of transfection were trialed. In forward transfection cells were seeded twenty-four hours prior to transfection, whereas in reverse transfection both seeding and transfection were simultaneous. Cells that were reverse transfected (Fig. 4.2A) showed lower overall GFP expression (~50%) of those cells that were transfected using the forward transfection method (Fig 4.2B), indicating that the forward transfection method appeared more optimal. In addition, brightfield images indicate cell health in reverse transfected cells were poor.

To confirm that transfection efficiencies weren't affected by plasmid condition or operator error, cells were transfected with the optimized plasmid transfection reagent Lipofectamine™ 3000 and high levels of GFP expression were achieved (Fig. 4.3). Following a review of literature, the reagent CRISPRmax was trialed with forward transfection and similar expressions of GFP (Fig. 4.2C) were seen to comparative transfections using jetCRISPR. It is noted that cell population appears increased in the CRISPRmax trial, but poor quality of the brightfield image in 4.2C makes direct comparisons unreliable. To confirm results seen for pEGP were applicable for RNP complexes, an assembled ATTO labelled gRNA complex was transfected with Cas9 using CRISPRmax and forward transfection and high levels of transfection were achieved (Fig. 4.4).

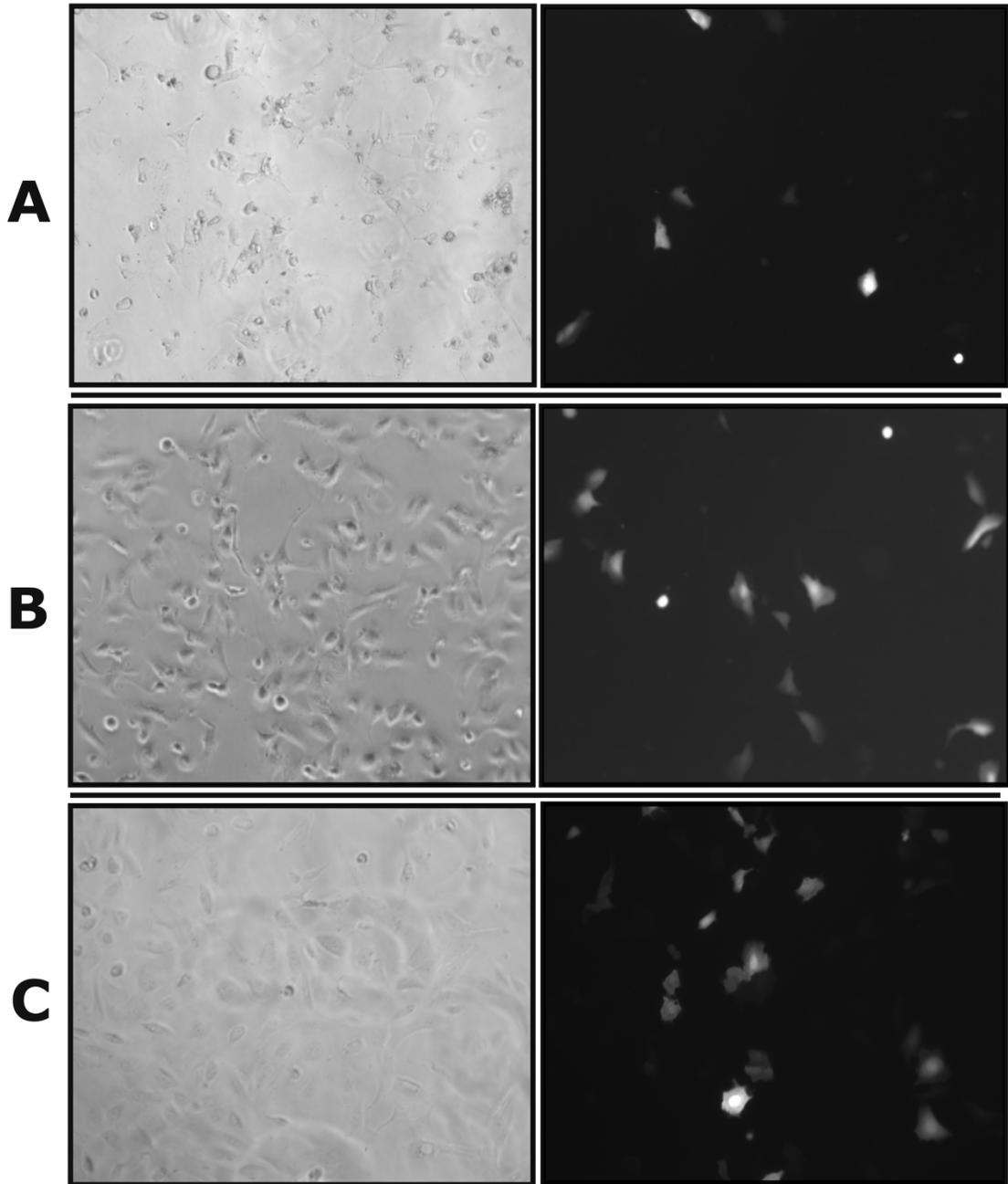


Figure 4.2. Transfection efficiency was improved using forward transfection technique. All transfections used the GFP human expression plasmid pEGFP. **(A)** Human U2OS cells were transfected with jetCRISPR using reverse transfection technique. **(B)** Human U2OS cells were transfected with jetCRISPR using forward transfection technique. **(C)** Human U2OS cells were forward transfected using CRISPRmax reagent.

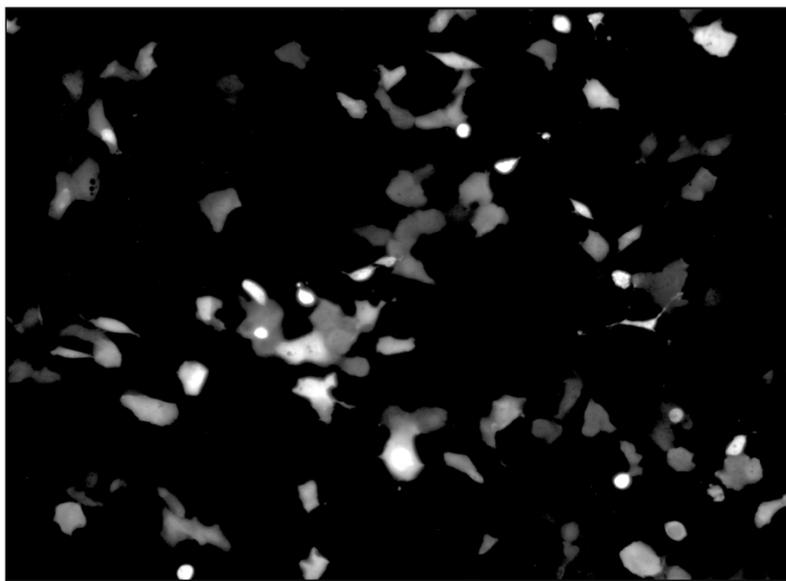


Figure 4.3. Plasmid condition and operator expertise was validated upon high levels of GFP plasmid transfection with optimised Lipofectamine 3000 reagent. Fluorescent image of cells transfected with Lipofectamine 3000 reagent and ATTO-labelled gRNA. A bright-field image was not available.

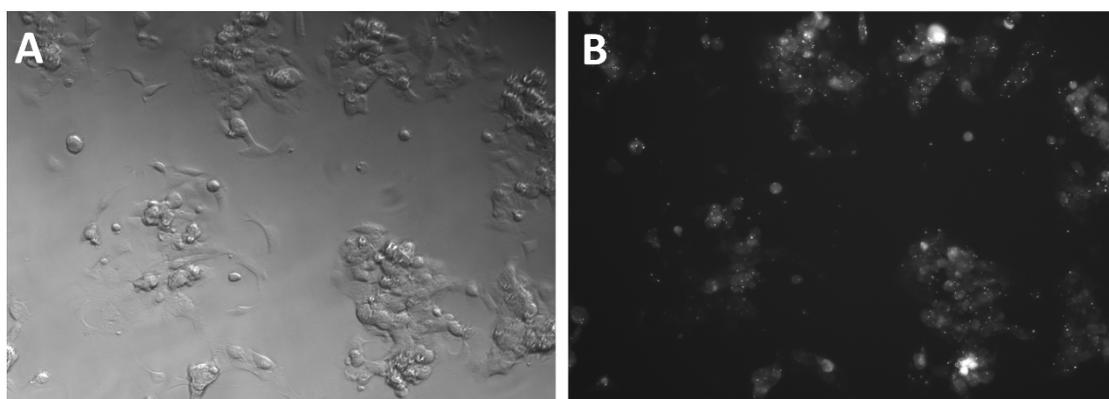


Figure 4.4. ATTO-labelled gRNA complexes confirm that optimised conditions are appropriate for RNP complexes. (A + B) Brightfield and fluorescent images, respectively, of cells transfected with CRISPRmax and ATTO-labelled gRNA. The tracrRNA component of the complex was ATTO-labelled.

4.2.2.3. Purified recombinant Cas9 was less active than commercial Cas9

The activity of recombinant Cas9 was examined using a series of cleavage assays. Initially, activity of recombinant Cas9 purified by the author was compared with that purified by a previous operator (Dr Andrew Cubbon), and

AC-Cas9 successfully linearised pUC19 plasmid when combined with a gRNA corresponding to pUC19 DNA (Fig. 4.5A). AP-Cas9 activity was ambiguous. Completely linearised plasmid was not observed in either case when compared with a BamHI comparison. Comparisons of cleavage activity between AP-Cas9 and commercial Cas9 (Integrated DNA Technologies) were carried out using gRNAs targeting exon 5B on a PCR amplicon of exon 5 from genomic DNA. Both sets of Cas9 showed cleavage activity (Fig. 4.5B). IDT Cas9 was notably more active than AP-Cas9 at 40 nM RNP complex, whilst comparisons at 80 nM RNP were inconclusive. Neither Cas9 achieved complete digestion of the amplicon, but assays indicated that gRNA did produce functional cleavage. As AP-Cas9 appeared weakly active, commercial Cas9 was used in further gene editing experiments.

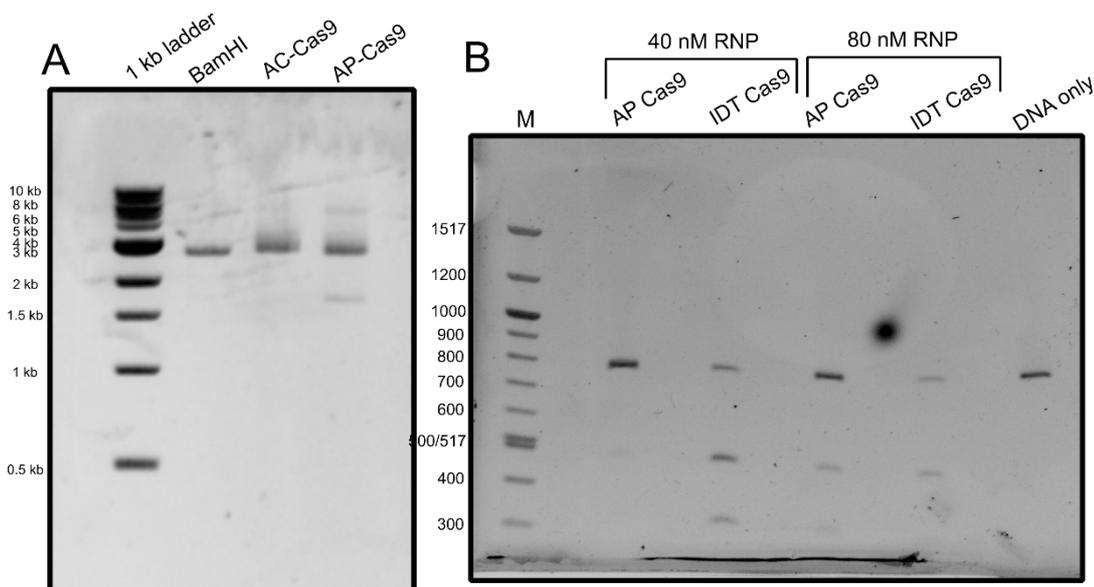


Figure 4.5. Commercial Cas9 was more active than purified recombinant Cas9 in activity assays. (A) Purified recombinant Cas9 from two operators was incubated with 200 ng of pUC19 plasmid with gRNA for sixty minutes at 37 °C in 1 x Cas9 working reagent prior to running on a 1% agarose gel. Linearised plasmid digested with BamHI was used as a positive control. (B) Wild type Exon 5 of *DDX52* was amplified via PCR and 200 ng incubated with both purified recombinant and commercial (IDT) Cas9 and Exon 5B gRNA in 1 x Cas9 working reagent at two different RNP concentrations for 60 minutes at 37 °C before running on a 2% agarose gel. Cleavage products of 500 and 300 bp were expected.

4.2.2. CRISPR editing of *DDX52*

4.2.2.1 Testing of gRNAs

Human U2OS cells were transfected as described in section 2.8.4.3. Genomic DNA was extracted following incubation and PCR amplification of target regions carried out. The PCR amplification for WT exon 5 initially failed: results were still investigated using T7E1 endonuclease (Fig. 4.6A). Two bands were seen for exon 5A edits, indicating heteroduplex formation and successful gene editing in gRNA corresponding to exon 5A, though nothing was noted for exon 5B gRNA. This matched industry predictions of on-target scores seen in table 4.1. The amplicon of the exon 4 edit was of insufficient quality to form a conclusion. To confirm results alongside a more representative negative control, a repeat PCR amplification was performed (Fig. 4.6B) and WT exon 5 successfully amplified. Heteroduplex formation in exon 5A was successfully replicated. Exon 4 amplifications were of better quality and no editing seen. Interestingly, repeat amplifications observed heteroduplex formation within gRNA targeting exon 5B not seen in the first amplification, however this still appeared weaker than cleavage seen in exon 5A.

Table 4.1. *DDX52* crRNAs used within this study. The on-target and off-target scores as predicted by manufacturer are presented.

Gene	Exon	Sequence	PAM	On-target score	Off-target score
<i>DDX52</i>	4	GCTTGCATTT GGATTGGCGT	AGG	50	80
<i>DDX52</i>	5A	TGGCTGGCAA GTTCTCGTGT	TGG	70	79
<i>DDX52</i>	5B	CTCGTGTTGG TGATATAATC	AGG	56	80

Both sets of exon 5A and 5B amplifications were sent for subsequent sequencing and edited sequences compared with wild type amplicons using TIDE (Tracking of Indels by Decomposition) analysis suite. TIDE analysis (Fig. 4.7) revealed low editing efficiencies across both guide RNAs used. In contrast

to results within T7E1 assays, higher editing efficiencies were noted within exon 5B than exon 5A; with means of 12% and 10%, respectively. This contradicts industry predictions in table 4.1. Predominant InDels were single base pair insertion and deletions in exon 5B, whereas exon 5A showed a higher range of deletions greater than 5 base pairs.

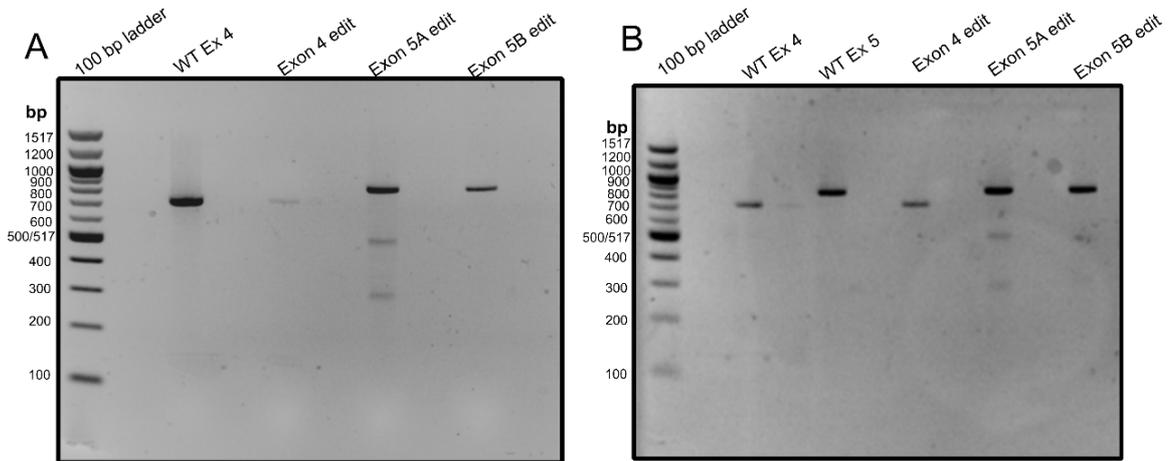


Figure 4.6. Exon 5A and 5B guide RNAs demonstrated successful editing in T7 endonuclease assays. (A) Genomic DNA was extracted from edited U2Os cells after 48 hours, target sites amplified and heteroduplex formation examined using T7E1 endonuclease assay. WT exons were used as positive controls: the PCR for WT exon 5 was unsuccessful and was not included. (B) To obtain a more appropriate negative control, amplification of target sites was repeated and heteroduplex formation examined using T7E1 endonuclease as in figure 4.5A. Cleavage products of 500 and 300 bp were expected in successful editing.

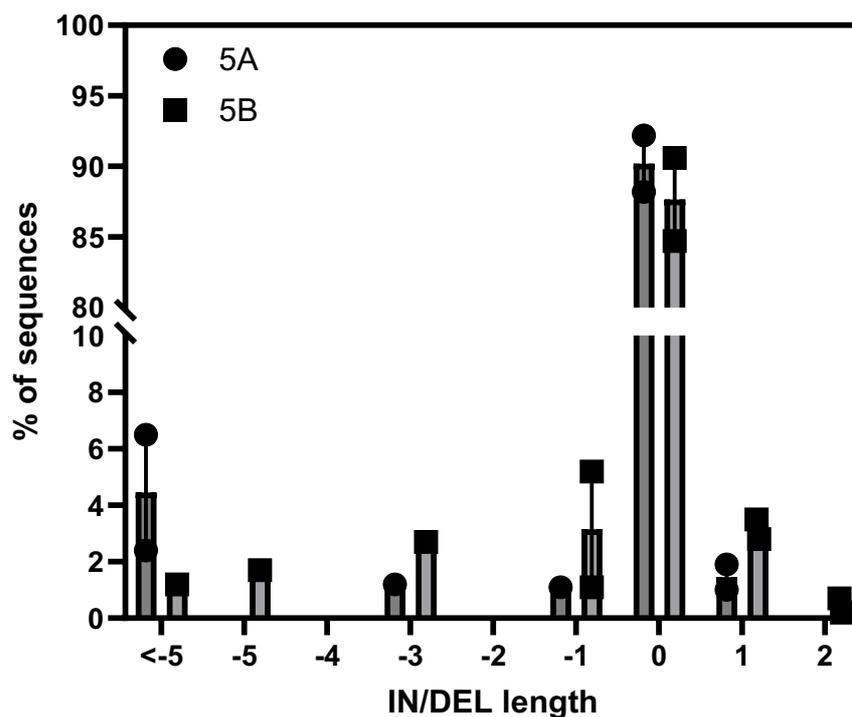


Figure 4.7: TIDE analysis of sequencing results reveals low levels of successful editing. Genomic DNA was extracted and amplicons of target sites sanger sequenced with sequencing data subsequently ran through TIDE analysis. Results presented are from two PCR repeats.

4.2.2.2. Full gene editing run and establishment of a monoclonal cell line

Following better performances noted within TIDE analysis (Fig. 4.7), Exon 5B gRNA was used to carry out a full workflow. Cells were transfected, incubated and detached prior to seeding at a density of 0.5 cells per mL in 96 well plates via a series of dilution steps. Genomic DNA of the remaining cells was extracted and tested via TIDE sequencing and editing efficiencies of 30.7% recorded (Fig. 4.8). After seven days, wells were analysed using microscopy and single cell forming colonies noted. These were tracked, cultivated and it was noted that a number of these colonies were not viable for the establishment of colonies: from an initial twenty-one wells noted to have cells present, only four possessing single cell colonies were able to be fully expanded into continuous cell lines and sequenced.

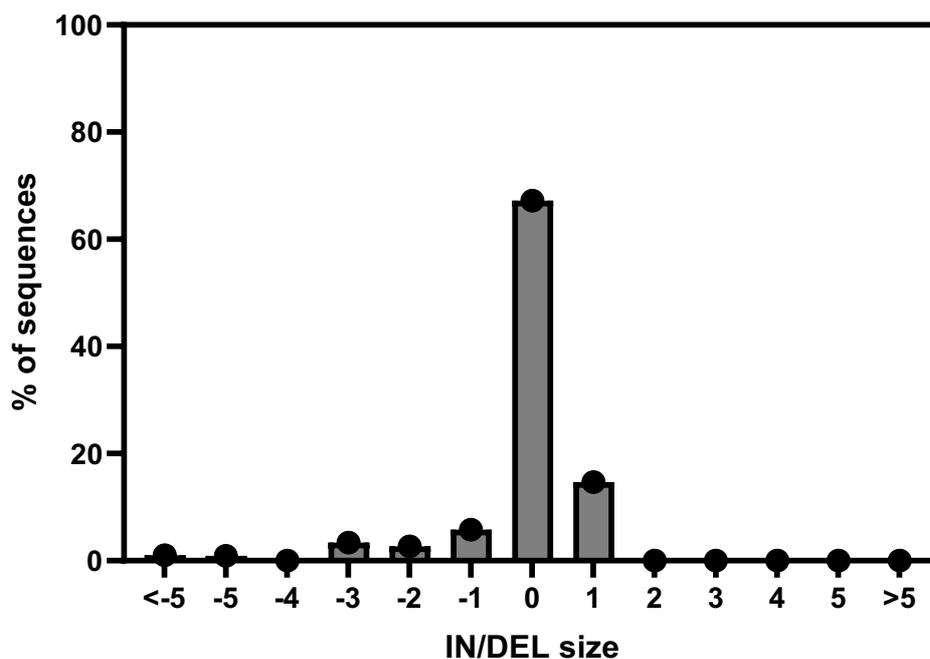


Figure 4.8: Preliminary TIDE analysis results from the full workflow indicate editing efficiency. Genomic DNA was extracted and amplicons of target site 5B was sanger sequenced with and compared with wild type DNA through TIDE analysis. $P > 0.05$.

4.2.2.3. On-target testing

On target editing was examined by comparing amplicon sites in edited cells vs wild type cells via sanger sequencing as described in section 2.8.5.1. Two lines (B4 + D2) were identical to the wild type cell line. Cells isolated from well B2 showed successful generation of a single base pair insertion (Thymine), but percentage ratios indicated a monoclonal progenitor was ambiguous. Well D4 showed successful generation of a single base pair deletion (Adenine) and a 1:1 ratio of wild type to deletion suggested presence of a heterozygous cell line. Well B2 also presented evidence of heterozygosity, but D4 was considered a better candidate.

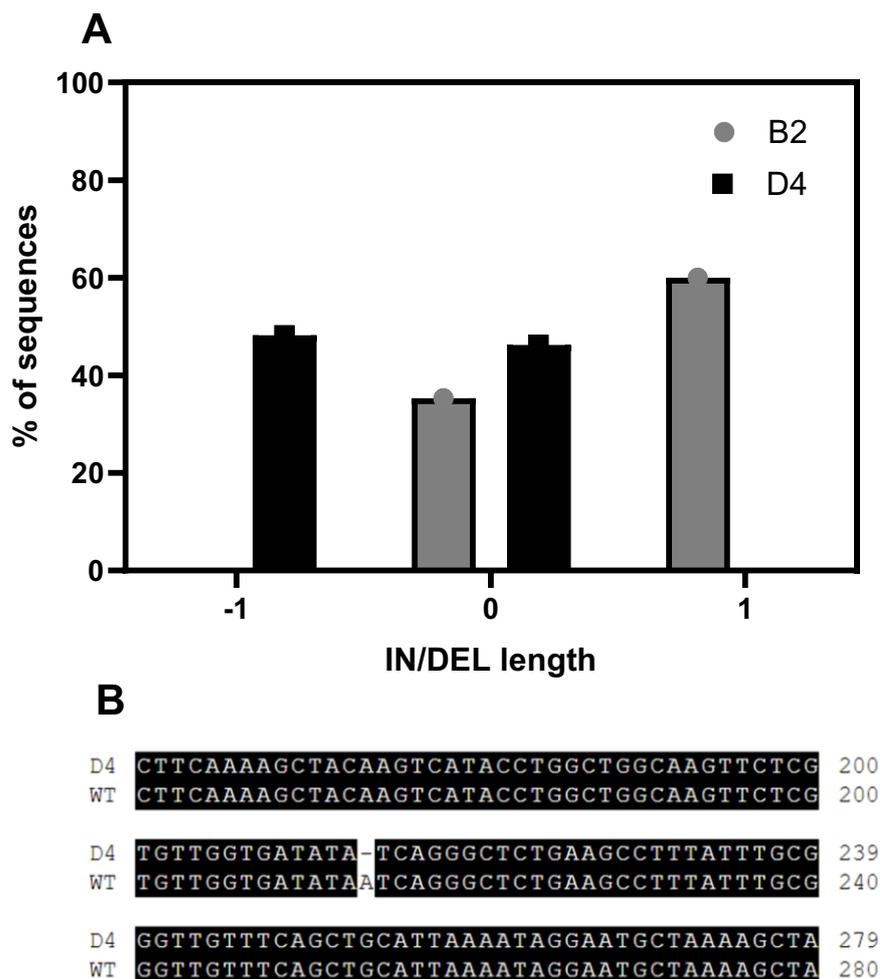


Figure 4.9: Sequencing analysis of samples taken from well D4 generation of a heterozygote. (A) In/Del lengths of wells confirmed successful editing in both wells B2 and D4 with approximately 50:50 sequence variation seen in D4. **(B)** Sequence comparisons confirmed a single adenine base pair deletion within week D4 in comparison to the wild type sequence.

4.2.2.4. Off-target testing

Potential off-target sites were identified using CRISPOR software. In total, 50 off-target sites were identified: 3 of which consisted of 3 mismatches whilst 47 consisted of 4 mismatches. The majority were intergenic or intronic regions with only two potential exon targets identified: RSR2 and ZNF717/MIR4273, with both possessing low-scoring CFD scores. The top 5 based on CFD score are shown in table 4.2. A number of these were unable to have primers designed using the software, noted to likely be due to the presence of repetitive

sequences within the flanking region. The presence of repeat sequences was verified using the human genome browser RepeatMasker track (366, 367). Off-target sites 3-5 were ordered for testing, though mm4_intron_TTTY21 was expected to be negative as the Y chromosome is absent within U2OS cells.

Table 4.2: The top five off-target sites identified for *DDX52* gRNA 5B.

Name	CFD score	Mismatches	Primers?	Nearby repeats
mm4_intergenic_DC AF5 Y_RNA_chr14_6 9137388	0.44	4	No	L1ME4a, MER44A, MER20
mm4_intron_EYS_ chr6_64597939	0.4	4	No	LINE L1PB4, Alu, L2A
mm4_intergenic_A C009505.2 NCK2_ chr2_105863591	0.34	4	Yes	MIR
mm3_intron_CAM KMT_chr2_443678 31	0.3	3	Yes	HAL1, Tigger4b, OldhAT1, X6B_LINE
mm4_intron_TTTY 21_chrY_9719917	0.29	4	Yes	(GT) _n , SST1

Off-target sites 3 and 4 were successfully amplified and no off-target effects were identified. However, chrY_9719917 also showed successful amplification, though it was anticipated that U2OS cells did not possess a Y chromosome. Upon comparing wild type and heterozygous cells, high levels of sequence decomposition were noted following a continuous sequence of 31 thymine bases (Fig. 4.10). The reported sequence was cross-referenced with the human genome browser and was found to be identical to a region on chromosome 17 with highly similar sequences also reported on chromosome Y.

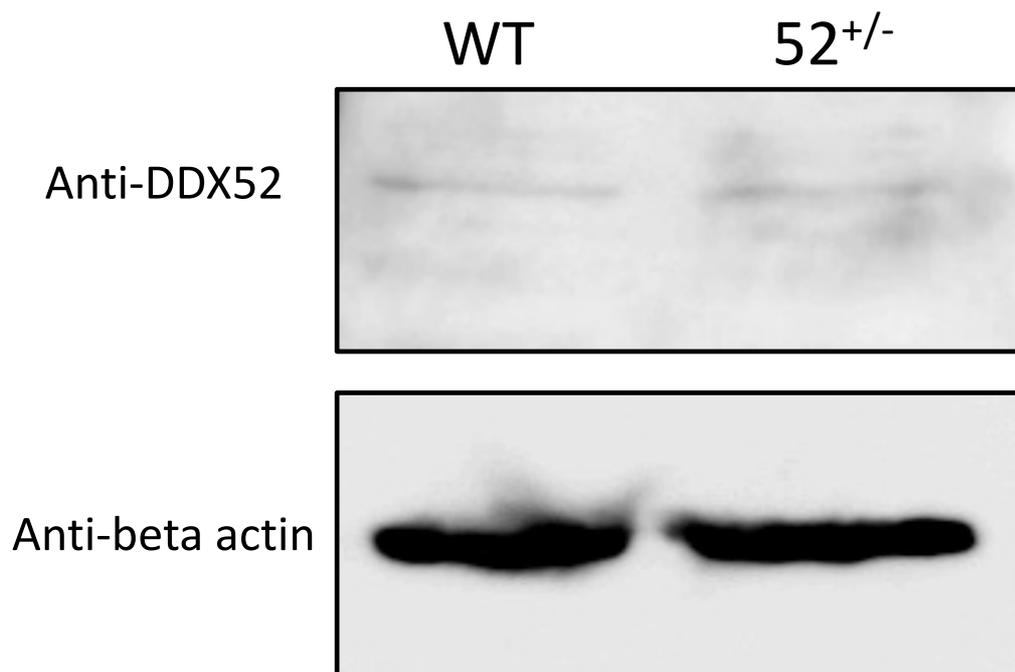


Figure 4.11: Western blots were inconclusive at distinguishing change within DDX52 expression levels in wild-type and DDX52^{+/-} cell lines. Cell lysate (35 ug) was loaded on to an SDS-PAGE gel and probed using anti-DDX52 and anti-beta-actin antibodies, prior to probing and analysis with HRP labelled secondary antibodies. Beta actin was measured as a loading control.

4.2.3.2. DDX52^{+/-} cells phenotypes include impeded cellular migration and proliferation

To phenotype the *DDX52^{+/-}* U2OS cell line in comparison to wild type cells, a scratch-based 'wound-healing' assay was employed to examine if cellular migration was impacted over time (Fig. 4.12A). Wild type cell lines were observed to have recovered to 100% confluency after a period of only 24 hours, whilst the heterozygous cell line still had a sizeable wound indicating reduced migration. In addition, *DDX52^{+/-}* cells demonstrated reduced adhesion and population density.

To quantify cell proliferation, the WST-1 assay was employed to measure changes in the number of viable cells over time. As it was not possible to generate a reading at 0 hours, the 24-hour time point was standardised as 1 and cell growth normalised as a ratio to the absorbance recorded at this time. No clear differences within cell proliferation were seen within the first 48 hours.

After 72 hours a significant difference within cell growth became apparent, becoming more distinct to several magnitudes of difference by 120 hours. This appears to indicate *DDX52* heterozygosity has a long-term impact on proliferation over time. It was also observed that prolonged passaging appeared to impact the viability of *DDX52* heterozygotes, notably cells began to lose viability after 8-10 passages. These effects were not able to be examined in detail.

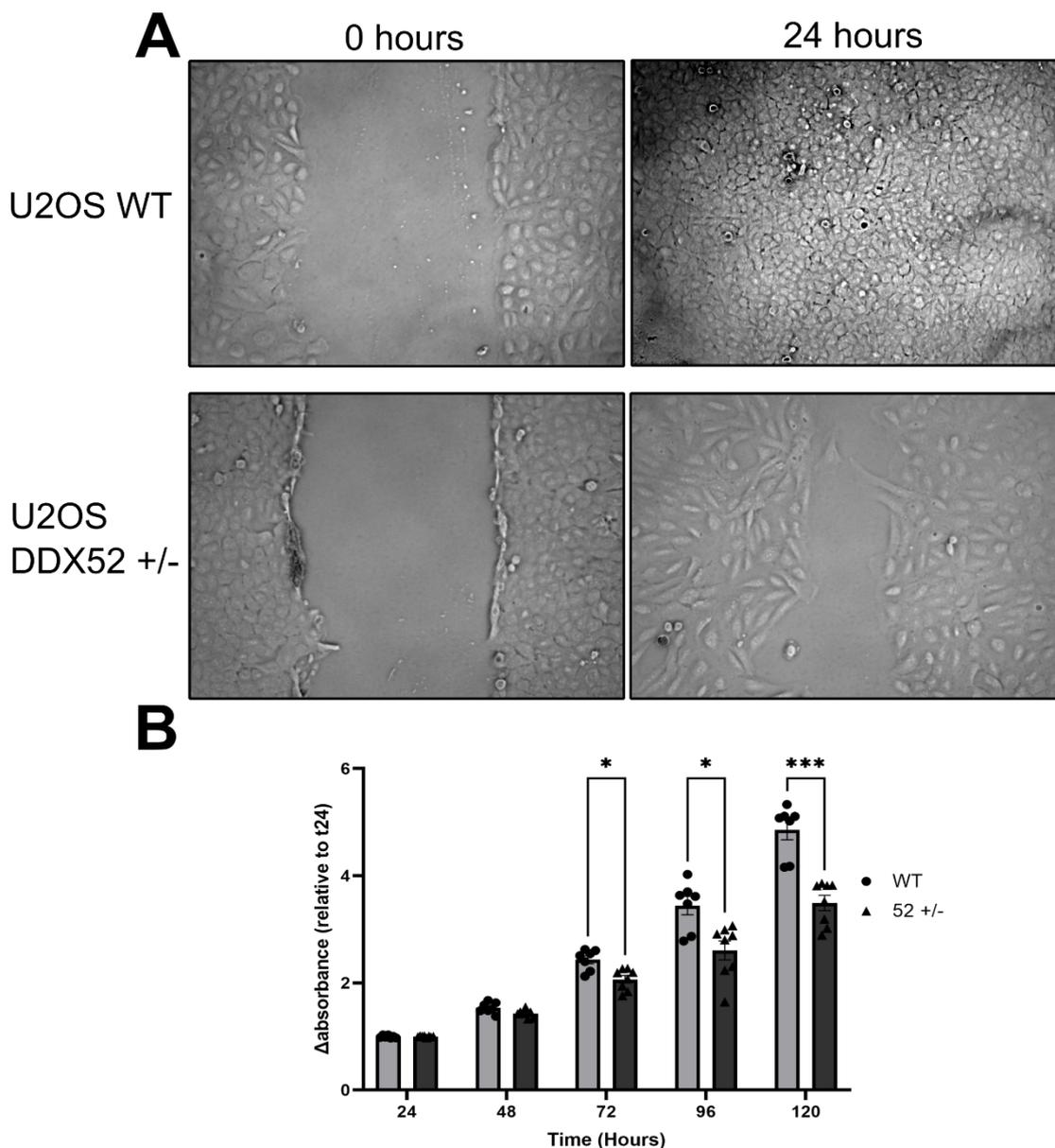


Figure 4.12: Cell migration and proliferation of wild type and *DDX52*^{+/-} U2OS cells. (A) Cell migration of U2OS WT and 52 +/- cells was analysed using scratch ‘wound healing assay’. Images were acquired at 0 and 24 hours in in

in vitro scratch assay. *DDX52*^{+/-} cells showed slower rate of migration when compared with WT cells. **(B)** Using the WST-1 assay, wild type and *DDX52*^{+/-} proliferation was assessed over five days. Graph shows that after 72 hours of seeding, proliferation of wild type cells was significantly increased compared to heterozygous cells. n = 7 wells per group, 3000 cells per well, Two-way ANOVA, *P<0.05, ***P<0.001

4.2.3.3. *DDX52*^{+/-} cells show no clear phenotype in response to cytotoxic agents

Having established migration and proliferation phenotypes associated with heterozygous cell lines, we next investigated the impact of genotoxic agents on the viability of *DDX52*^{+/-} U2OS cells in comparison to wild type cells. Experiments were performed as discussed in sections 2.8.1 (Mitomycin C) and 2.8.2 (Cisplatin) and results compared with a control group consisting of a media-only control. Heterozygous cells showed an apparent increased susceptibility to Mitomycin C at concentrations of 50 nM and above, however it was noted that more replicates were needed to validate this data. The crosslinking agent cisplatin was also tested with both sets of cells showing a similar dosage response to increasing concentrations of cisplatin with no conclusive difference between *DDX52*^{+/-} and wild type cells.

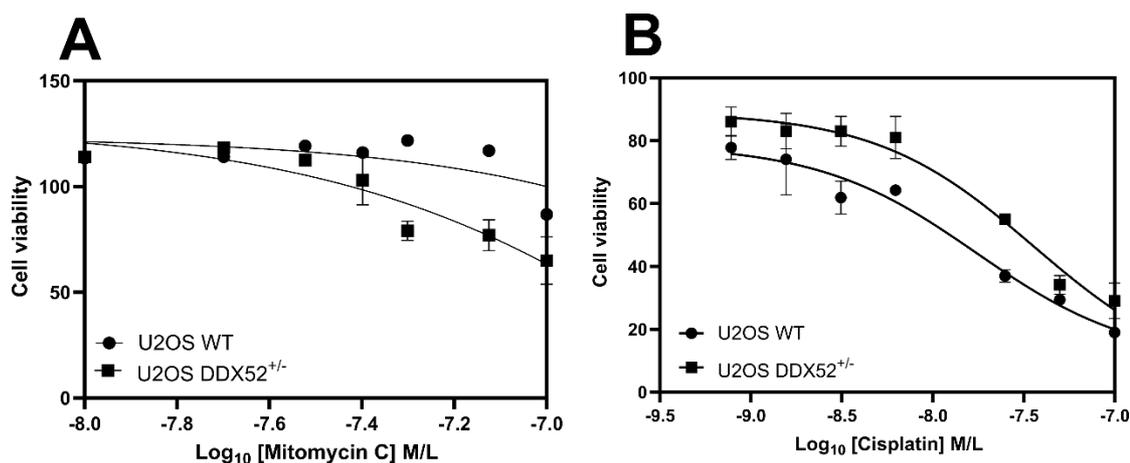


Figure 4.13: *DDX52*^{+/-} response to cytotoxic agents. (A) Response of *DDX52*^{+/-} cells to the cross-linking agent Mitomycin C was analysed using a series of concentrations and WST-1 assay kit. Heterozygote cells showed decreased levels of cell viability in response to higher concentrations of

mitomycin C compared to the wild type cells. N = 2. **(B)** Response of *DDX52*^{+/-} cells to the Cisplatin was analysed using a series of concentrations and the ATPlite assay kit. N=4.

4.2.4. Trialling workflows for investigations of ribosome biogenesis

No method was established for examining ribosome biogenesis. Methods of northern blotting and RT-qPCR were trialled. In an attempt to circumvent radiation use, a method was trialled using fluorescent based northern blot probes similar to those described by Miller et al (2018) with two major differences. Whereas Miller et al (2018) used internal azides labelled with near-IR dye by click chemistry in this a workflow internal azides labelled with Cy5 were used. A HRP-conjugated biotin labelled oligonucleotide was also trialled based on methods outlined by Huang et al (2014) (368). Finally, as no decommissioned hybridization ovens were available a shaking incubator was used for this step.

4.2.4.1. Northern blotting

Following copper-free click chemistry using a DBCO-Azide reaction, probes were purified using a spin column with A260 and A649 values of 0.649 and 0.212 were seen, respectively, indicating successful labelling of DNA. Total RNA was successfully extracted from wild type and *DDX52*^{+/-} cells and northern blotting carried out as described in section 2.9. Successful transfer of RNA to the membrane was confirmed using methylene blue stain (Fig. 4.14) post transfer, but practical difficulties within the loading process led to poor transfer of *DDX52* +/- RNA. It was noted downstream that methylene blue interfered with Cy5 signal thus this step was omitted in later attempts. Hybridisation was carried out using probes corresponding to 47S rRNA and assessed using an Amersham Typhoon but no fluorescence was observed. Repeats were attempted with alternative probes targeting other rRNA species and biotin labelled probes but were unsuccessful. Work was discontinued.



Figure 4.14: Transfer of U2OS WT and *DDX52* +/- RNA to northern blot membrane was successful. Total RNA was extracted and transferred by capillary transfer to northern blot membrane. RNA was imaged via staining with methylene blue.

4.2.4.2. RT-qPCR

4.2.4.2.1. Validation of housekeeping genes for normalisation of RNA expression

As methods of northern blotting were unsuccessful, methods of RT-qPCR were trialled to examine if they could be used to identify differences within rRNA levels between heterozygous and unedited cells. As a new workflow, an initial study was carried out to test the suitability of three housekeeping genes for the normalisation of RNA expression: Beta-actin, GAPDH and UBE2.

Table 4.3: GAPDH and UBE2 show superior performance as housekeeping genes for RT-qPCR assays. Cycle thresholds are presented and the standard deviation presented for each sample. Samples were run in triplicate but the top row of the plate failed. N=2.

Sample	Beta-actin average CT	GAPDH average CT	UBE2 average CT
WT	0.000	19.454+/-0.32	26.362+/-0.034
52	27.655+/- 5.85	16.759+/-0.034	24.979 +/- 0.17

Both GAPDH and UBE2 housekeeping probes were successful and showed high levels of consistency between samples. Beta-actin signals were poor and inconsistent, with all samples within the wild type sample failing, indicating poor suitability as a housekeeper gene. GAPDH and UBE2 were used going forward. It was observed that all tests on the top row of the plate failed – these wells were avoided in future assays and UPW added to wells surrounding test samples to limit evaporation.

4.2.4.2.2. qPCR suggests 47S, 28S and DDX52 transcripts are reduced in *DDX52*^{+/-} cells

GAPDH and UBE2 were consistent as housekeeping genes and expression levels of rRNA species normalised to these readings. Total RNA was extracted, reverse transcribed and a series of PCR probes used to amplify several rRNA species in addition to PCR probes targeting *DDX52* transcripts. Cells were not synchronised prior to RNA extraction. There was no substantial difference within 18S and 5.8S rRNA levels. Levels of 47S and 28S rRNA were notably reduced, however, with 28S rRNA levels over two-fold lower than those recorded within wild type cells. In addition, *DDX52* results were five-fold lower than within wild type cells. Amplification of *DDX52* within both wild type and *DDX52*^{+/-} cells had a high cycle threshold of detection, 37.308 and 39.184 respectively, suggesting that *DDX52* expression is low within the cell. This as reflected within repeat assays where *DDX52* transcript levels within *DDX52*^{+/-} cells were frequently unable to be amplified above the threshold within the cycle limit.

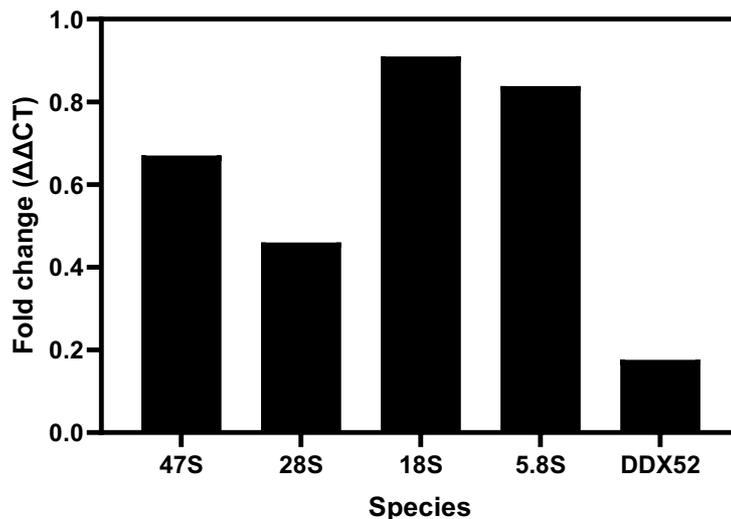


Figure 4.15: Cycle thresholds in qPCR show reduced levels of 47S, 28S and *DDX52* transcripts within *DDX52*^{+/-} cell lines. Fold differences were calculated through comparisons of $\Delta\Delta CT$ values with wild type cells, with wild type rRNA levels set as '1'. Sets of data for both housekeepers were consistent and samples normalised to GAPDH signals. N = 3.

4.3. Discussion

4.3.1. CRISPR-Cas9 workflow development

4.3.1.1. Purified recombinant NLS-Cas9 was unsuitable for gene editing

Purification of NLS-Cas9 was successful but figure 4.1D indicates that the purification batch possessed high molecular weight impurities. It was considered that this would likely mean the protein was unsuitable for transfection into human cells due to the potential for undesirable effects. Due to similarities in molecular weight to NLS-Cas9, a higher resolution size exclusion column would be necessary to remove these contaminants. In addition His-tags have been well reported to pull down contaminants thus purification using a more specific tag could be used to improve purity, such as a strep-tag. Alternatively, Cas9 could be delivered via plasmid expression, though this can result in undesired effects: such as random integration of vector into the chromosome (369).

Purified NLS-Cas9 was subsequently tested through a series of cleavage assays and complete digestion of amplicons was not observed, indicating lower

Chapter 4: Investigating DDX52 within tumour cells

than desired protein activity. Losses in protein activity can occur due to proteolysis and aggregation during the purification process. Alternatively, it has been suggested that histidine imidazolyl side chains present within His-tags can impact biological activity of the protein, including signal sequence processing (370). The plasmid used within this study, pAC29, possesses a His-tag immediately 5' of the NLS-sequence (Fig. 4.16). Whilst the impact of this on the cleavage activities of Cas9 is unknown, it does suggest this could potentially impact transport to the nucleus upon transfection and downstream gene editing. This could be alleviated using TEV-protease to cleave the His-tag, however as a TEV protease site is not included within the plasmid this was not possible. The commercial Alt-R Cas9 was trialed and increased cleavage activity was seen within assays. This was used within further trials.

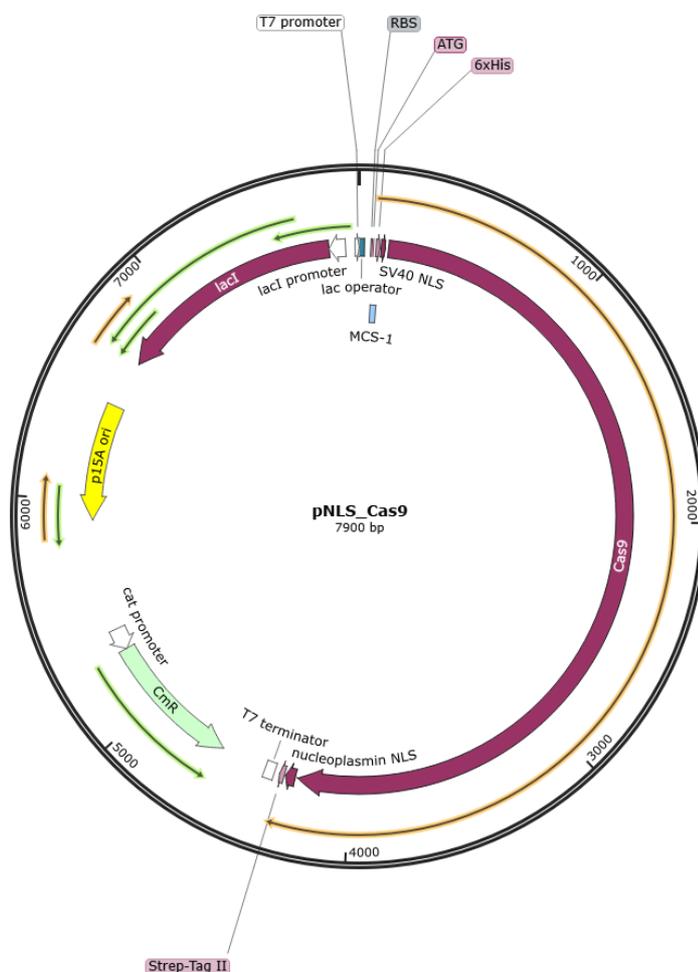


Figure 4.16. Plasmid map of pAC29 (pNLS-Cas9). The N-terminus located 6 x His-tag and nuclear localization signal (NLS) are indicated.

4.3.1.2. An optimised workflow was developed using forward transfection with CRISPRmax

Trials of reverse and forward transfections suggested that forward transfection performed better than reverse transfection. This is interesting as existing literature suggests reverse transfection is the superior method for transfection, likely due to increased surface availability within suspension cells (371). It is suggested that results seen here are influenced by the operator and forward transfection being a less intensive process, with cell condition being poor in brightfield images of reverse transfected cells. Alternatively, transfection reagents may have different toxicities with different transfection methods (372). Considerations are made that cell expansion within forward transfection protocols must be accounted for: initially, 40,000 cells per well were seeded but following 24 hours of incubation it was estimated that ~50,000-70,000 cells per well would be present by the time of transfection according to a division time of 29 hours for U2OS cells. This is still well within transfection limits.

High expression of GFP was observed following forward transfection with Lipofectamine 3000 reagent. However, whilst research has previously used Lipofectamine for RNP transfection, it is not optimised for the transfection of RNP complexes in comparison to other reagents. jetCRISPR reagent was initially trialled for transfection, but overall poor transfection efficiencies were seen. At the time of testing, recent literature had demonstrated that the editing efficiency of jetCRISPR was poor in comparison to CRISPRmax (373, 374). In addition, a search of available literature failed to identify studies successfully using jetCRISPR. Whilst forward transfection with CRISPRmax delivered comparable transfection efficiencies to those performed with jetCRISPR, CRISPRmax was chosen as an optimal transfection reagent according to evidence from previous literature.

The transfection reagent jetCRISPR has since been successfully used within gene editing and studies of DNA damage response (375, 376). However, it is noted that as of the time of writing the product is no longer available for purchase. It was considered that pEGFP was a poor model for studying RNP transfection: to confirm optimised transfection workflow was applicable for RNP complexes, ATTO-labelled gRNA complexes were transfected and high levels

of transfection were recorded. Future transfection optimisations should utilise these ATTO-labelled gRNAs in conjunction with non-essential genes.

4.3.2. Guide RNA testing

Guide RNAs targeting *DDX52* were tested for editing efficiency. Interestingly, in contradiction to manufacturer predictions, following TIDE analysis gRNA targeting exon 5B showed higher editing efficiency than exon 5A. This also contradicted results of the T7E1 assay where exon 5A delivered more distinct results than exon 5B. Common limitations within enzyme mismatch cleavage (EMC) assays include that commonly used endonucleases are poor at initiating cleavage at single-base loops, making them inadequate for the detection of single-base pair indels (377). This is reflected within the data presented here, with exon 5A consisting of larger In/Del events than exon 5B according to TIDE analysis, with exon 5B predominantly forming single base pair insertions/deletions. Testing all guide RNAs by sequencing is recommended for future experiments: following negative results in EMC assays, exon 4 DNA was not sequenced so the performance of this guide RNA is unknown. Whilst this matches manufacturer predictions for exon 4, the performance of exon 5B suggests this should have been tested.

Overall editing efficiencies were poor and different levels of efficiency were seen between PCR repeats of the same samples. This is likely a phenomenon referred to as 'jackpotting' resulting in PCR bias: due to the PCR stochasticity, templates amplified early are exposed to exponential amplification producing variations within results (378). The use of smaller amplification cycles could mitigate this effect. Later gene editing experiments on *DDX49* (section 5.2.6) would obtain superior efficiencies suggesting that this might be poor on-targeting efficiency of *DDX52* gRNAs. Alternatively, as higher editing efficiencies were seen within the full gene editing run of *DDX52*, the cause of low efficiencies could be operator inexperience.

4.3.3. The successful generation of *DDX52* heterozygotes

Following sequencing, a successful heterozygous gene edit was confirmed with a deleted adenine base within the site of interest. This was expected to introduce a frameshift that would eliminate protein activity. Throughout

outgrowth, a number of single cell colonies were unable to be expanded suggesting gene editing resulted in cells becoming unviable. This is supported by reported status of *DDX52* within CRISPR knockout screens as an essential gene in over 90% of tested cell lines (379). It is hypothesized that *DDX52* is essential as it plays a key role in ribosome biogenesis like its predicted yeast homolog, thus *DDX52* null cells lack the ability to produce functional ribosomes and hence the cells proteome.

The information heterozygotic cells may give on phenotypes and functional roles may be limited depending upon the haplosufficiency of the gene (380). No previous data on *DDX52* haplosufficiency is available. Heterozygous loss of ribosomal proteins has previously been suggested as producing qualitative alterations in ribosomes necessary for cancer progression (137). A method to consider in future would be using RNA interference as opposed to knockout methods. In screens using RNA interference (RNAi), it is reported that over 95% of cell lines are not *DDX52* dependent suggesting that knockdowns of *DDX52* are sub-lethal (381).

4.3.3.1. No off-target editing was seen within predicted sites

Fifty off-target sites of interest for testing were determined using the CRISPOR prediction tool based on the guide RNA used. Of these, only three were within three mismatches. It was decided to attempt to test the top five predicted sites - the extensive testing of all identified off-target sites was beyond the scope and budget of this study and it has been reported that many off-target sites identified by CRISPOR are likely false positives (300). It was subsequently found that primers targeting the top two off-target sites were unable to be designed – as noted within table 4.2, this is likely due to the surrounding region containing repetitive gene sequences.

No off-target editing was identified in sites tested. Interestingly, site chrY_9719917 was successfully amplified: as U2OS cells are derived from a female donor, amplification of this target was expected to return a negative result as the Y chromosome is absent. Following comparisons with the human genome browser the sequencing data was found to correspond to a region on chromosome 17 sharing high similarity with the Y chromosome. Similarities

could be the result of chromosomal heteromorphisms, but this is of little interest or consequence to this study (382). Interrogations of the sequencing trace revealed that sequencing decomposition occurred following a long stretch of thymine residues. It was concluded this was a result of polymerase slippage induced by the homopolymer thymine region (383). This further highlights the complexities in testing off-target sites.

As described, amplicon-based evaluation of off-target effects are limited and impeded by several complexities within design. Comparisons with next-generation sequencing approaches suggest that use of predicted sites fails to uncover rare off-target events (384). An ideal, more thorough evaluation would be using whole-genome sequencing. However this is largely impractical and unfeasible for many labs, including this study, and is rarely used with an assessment of CRISPR studies in plants reporting 72% used prediction methods similar to applied within this study (385). As such there is still no gold-standard technique, meaning that whilst genetic editing is now more accessible, thorough off-target assessment remains a barrier to many studies (384).

4.3.3.2. *DDX52*^{+/-} impacts on *DDX52* protein expression are inconclusive

Data from western blots (figure 4.11) suggests no differences within *DDX52* protein expression between heterozygous and wild type cell lines, but it is considered from faint banding that concentrations of *DDX52* are low within the sample so resolution between the two bands is unreliable. Resolution could be improved by enriching the nucleololus fraction through a series of subcellular fractionation steps, whilst the lysis buffer TRIzol may be more appropriate for extracting these proteins (386). Alternatively, immunoprecipitation has been proven effective within the enrichment of low abundance proteins though it may be inappropriate for comparisons of expression (387).

The expression of *DDX52* was further examined using methods of RT-qPCR to identify if there was any difference within levels of *DDX52* RNA transcripts and a five-fold reduction within *DDX52* $\Delta\Delta CT$ was noted in comparison to the wild type cells. This data appears to confirm that heterozygosity severely impacts *DDX52* expression levels. However, it was hypothesised that expression levels would only be two-fold lower within heterozygous cells - additional repeats are

recommended to confirm this data. It was also noted that DDX52 had a high cycle threshold within both cell lines, suggesting that it is canonically expressed at low abundance within the cell and supporting data seen within western blots.

It is considered that the U2OS cell line used within this study possesses a polyploid karyotype. A review was conducted and copy number profiles confirmed that *DDX52* possessed a normal copy number within this cell line (388). Previous literature has also reported that chromosome 17 within U2OS cells is unaffected by aneuploidy (389). The original scope of this study was intended to include the human near-haploid cell line HAP1, however due to logistical issues as a result of the COVID-19 pandemic this was not possible.

4.3.4. *DDX52*^{+/-} cells present impacted cell migration and proliferation

To investigate effects of *DDX52* heterozygosity on cell migration, ‘wound-healing’ assays were conducted and reduced levels of migration were observed within *DDX52*^{+/-} cells, with wild type cells recovering after 24 hours whilst *DDX52*^{+/-} failed to fully recover. A distinct loss of cell-cell adhesion was also observed, but it is unclear if similar effects were observed as wild type cells migrated – previous research indicates loss of cell-cell adhesion at the cell rear is common within wound-healing (390). It is considered that whilst this assay was effective, it is rudimentary and the manual pressure method applied here can result in variations in the size and area of the scratch. More accurate methods of creating scratches are available, such as lasers and electric currents (391, 392).

WST-1 assays were employed to quantitatively compare proliferation between the two cell lines and after a period of 48 hours reduced proliferation was seen. This became more significant over prolonged incubation. These support previous results seen within prostate and melanoma cancer cell lines, with knockdowns of *DDX52* shown to inhibit proliferation through regulatory interactions with c-Myc (97, 98). Reduced levels of migration seen within *DDX52*^{+/-} represents a novel result not previously seen within literature, though this has been previously reported in other DExD-box helicases (84, 393). Critically, tumour cell migration is essential within steps of cancer metastasis;

melanoma in particular is a highly metastatic cancer, with a survival prognosis of 6-12 months once cancer becomes metastatic (394, 395). Results here thus present DDX52 as a potential therapeutic target within metastatic cancers.

4.3.5. DDX52^{+/-} cell phenotypes to DNA damaging agents

Viability assays were conducted using the cross-linking agents mitomycin C and cisplatin, of interest to this study based upon previously reported bioinformatic links between DDX52 and the *Mycobacterium smegmatis* helicase Lhr, with deletions of Lhr linked with increased sensitivity to these cross-linking agents (104, 396). The results here show a clear increased susceptibility to mitomycin C within the heterozygous cell line at higher concentrations of mitomycin C, however as results here were only obtained in duplicate and based upon issues with the dataset further data is needed to confirm this phenotype. It is presumed that treated cells showing higher viability than untreated cells is a result of operator error, though the effect being reproduced across both tested cell lines is interesting. Treatments with Cisplatin showed similar dose responses across both cell lines but no clear increased susceptibility.

4.3.6. DDX52^{+/-} cells possess lower levels of 47S and 28S rRNA

Initial trials of housekeepers were successful and GAPDH and UBE2 presented ideal characteristics for assays. Beta actin levels were poor and inconsistent. Western blot data within section 4.2.3.1. suggests that beta actin is prevalent within cell extracts, so the cause of this is unknown. It is recommended additional primer pairs are trialled. A decreased abundance of 47S rRNA species was noted, supporting previous data from zebrafish models that suggest DDX52 is involved within the synthesis of 47S rRNA (357). No previous data has associated DDX52 with the processing of 28S rRNA and is hypothesised that inhibitions of 47S rRNA synthesis could present a downstream effect on 28S rRNA levels. Notably, 18S rRNA levels were unaffected within this study, with yeast models suggesting that the putative *S. cerevisiae* homolog, Rok1, processes 18S rRNA (108, 109). Importantly, Zebrafish DDX52 shares closer homology to *H. sapiens* DDX52 than *S. cerevisiae* Rok1, suggesting this may represent separation of function between yeast and higher organisms. Alternatively, transcription of rRNA genes fluctuates throughout the cell cycle with transcription maximal within S- and G2

phases (397). As cells were not synchronised prior to total RNA extraction, it is unknown if cell cycle distribution within each sample set impacts results seen here: it is recommended that future studies synchronise cells to further validate the results.

4.4. Conclusion

This chapter describes the successful development and implementation of a pipeline for the generation of gene edits within tumour cell lines as well as the initial phenotyping of a generated heterozygous cell line. Whilst further optimisation is possible, this workflow was successfully applied to other proteins within this study (see section 5.2.6) showing versatility. However, it is noted that whilst the likeliest sites of off-target activity were assessed, this is not extensive and the potential for off-target editing at other sites within the genome cannot be dismissed. Western blots suggested that DDX52 expression levels were unchanged within the cell, but RT-qPCR seemingly confirmed a five-fold reduction in DDX52 expression levels. This was more extensive than was hypothesised for a heterozygous cell line.

Efforts to study ribosome biogenesis via a radiation-free northern blotting method were unsuccessful whilst methods trialled with RT-qPCR suggest reductions in 47S within *DDX52^{+/-}* cells that are in line with previous research, as well as novel reductions in 28S rRNA levels not reported previously. Interestingly 18S rRNA levels were unaffected, conflicting with predicted phenotypes based upon the putative yeast homolog Rok1. Initial phenotyping with *DDX52^{+/-}* cells presents similar proliferation phenotypes as those seen previously within prostate and melanoma cancer cell lines, as well as novel results for cell migration that have not been reported before. This data presents DDX52 as a potential therapeutic target of interest, particularly within highly metastatic cancers.

4.5. Future work

At commencement of this work, no methods for studying ribosome biogenesis were readily implemented. A northern blotting trial was conducted using click chemistry labelled fluorescent probes similar to protocols outlined by Miller et al (2018). Click chemistry labelling of probes and transfer of total RNA to the

northern blotting membrane was successful but no signal was achieved following probing. It is unclear if this is an issue with probe hybridisation, if probe sensitivity is poor or another unrecognised issue. It is suggested that for future work a conventional radiation-based method is applied and a hybridisation oven is used.

Whilst optimisations were successful in implementing a gene editing workflow that was successful here and in further experiments on *DDX49* (Chapter 5), these were not extensive and further optimisation is possible. It is recommended a more appropriate strategy is utilised using RNP complexes instead of the GFP transfection methods used here. Alternative methods of transfection can also be trialled, including nucleofection (398). Future optimisations should utilise gRNAs targeting non-essential genes, similar to methods described by Chen et al (2018). Finally, further phenotyping *DDX52*^{+/-} cell lines can be carried out, including testing with further genotoxic agents and repeating RT-qPCR analyses following cell synchronisation.

Chapter 5: Biochemistry and cell models of probable ATP-dependent helicase DDX49

5.1 Introduction

5.1.1 The role of DDX49

DDX49 is a putative RNA helicase implicated within roles in RNA processing, viral infection and has been associated with several cancers (82, 84, 101). However, whilst data on the activity of DDX49 on small RNA duplexes has been well-described, little exists concerning its interactions with DNA. Reported to be a homolog of the *S. cerevisiae* protein Dpb8, DDX49 is distinguished from Dpb8 by a C-terminus extension of 52 amino acids. Both *Dbp8* and *DDX49* are classified as essential genes: *dbp8-Δ* mutants were unviable within *S. cerevisiae* whilst CRISPR knock-out studies report *DDX49* is essential within over 99% of tested human cell lines, with RNAi studies suggesting ~ 55% of tested cell lines show dependency following knockdowns of *DDX49* (90, 379, 381).

5.1.2. Functions of DDX49 within RNA processing

Whilst the biochemistry of DDX49 with DNA has not been previously reported, previous studies on RNA have identified roles within regulation of mRNA export and pre-rRNA levels within the cell (82). Specifically, multiple studies have linked DDX49 with the export of poly (A)⁺ RNA from the nucleus (82, 400). The Awasthi study also demonstrated that DDX49 was present in relatively low abundance within the cell, with DDX49 ATPase activity inhibited in the presence of both RNA oligonucleotides and total RNA by 17% and 45%, respectively, raising suggestions that RNA may act to modulate DDX49 activity.

5.1.3. DDX49 roles within viral immunity

One of the most common reported associations of DDX49 is its implicated role in interactions with several viruses. DDX49 has previously been reported to be a required host factor for replication of Human Immunodeficiency Virus (HIV-1), with particular interactions with the HIV-1 protein polyprotein Gag through affinity tagging and mass spectrometry (401, 402). Additionally, DDX49 was

upregulated in microarray analysis in HIV patients showing viral progression (403). Overexpression of DDX49 inhibited reactivation of gammaherpesvirus through inhibiting transcription of lytic viral genes (89). During the COVID-19 pandemic, it was also reported that DDX49 was reported to have associations with the SARS-COV2 protein ORF9C (also known as ORF14) (88). Though research later suggested OFC9C is not translated by SARs-COV2, more recent research suggesting detrimental impacts within cardiomyocytes means the nature of this protein remains ambiguous (404, 405) .

5.1.4. DDX49 within disease

Although DDX49 has previously been linked within the pathogenesis of Alzheimer disease, following observations it is upregulated within early stages of the disease, the protein has most prominently been linked with several cancers (87). The increased expression of DDX49 within prostate cancer tissues has been associated with poor prognosis within patients, whilst knockdowns suppressed PC-3 cell line proliferation and migration (406). DDX49 has been associated with lung cancer in multiple studies, with studies in vivo showing DDX49 promotes tumour cell proliferation through upregulation of the AKT/ β -catenin pathway, with knockdowns inhibiting proliferation and migration in PC-9 and H460 cell lines (84, 85). Similar interactions between DDX49 and the AKT/ β -catenin pathway were seen within cervical cancer cells (407).

Interestingly, data from the human protein atlas indicates that DDX49 overexpression is associated with highly positive prognoses within cervical cancer outcomes (83). DDX49 thus represents an attractive therapeutic target and diagnostic marker. DDX49 has been identified as a biomarker within lung cancer metastases and, in conjunction with EGFR and T-stage, could pose relevancy as a predictive model for diagnosis of lymph node metastasis (84, 400). Supporting DDX49s potential as a target for treatment, Morphine has previously been shown to inhibit growth in hepatocellular carcinoma by downregulating DDX49, though the mechanisms of this are unknown (86).

5.1.5. Aims and objectives

Whilst enzymatic activities of DDX49 with RNA have been well-reported by Awasthi et al (2018), little has been reported regarding interactions with DNA. With DExD-box proteins often reported as multifunctional and reported associations of DDX49 with cancer, this was identified as a gap within research. In addition, in the early stages of this research we identified an exceptionally well-conserved site within the C-terminus of DDX49 and a novel nuclease activity believed to be associated with this site. This chapter focuses on examining and establishing the initial enzymatic activity of DDX49 on DNA and investigating a novel activity. The objectives were as such thus:

- 1) To investigate biochemistry of recombinant DDX49, particularly in interactions with DNA-based substrates.
- 2) To investigate a potential link with a novel enzymatic activity and a well-conserved sequence within the proteins C-terminus.
- 3) Attempt to establish a CRISPR-edited monoclonal cell line of *DDX49* and phenotype where applicable.

5.2. Results

5.2.1. Bioinformatics

5.2.1.1. Comparisons of DDX49 and Dbp8 predicted models reveal an intrinsically disordered C-terminus extension

The structure of DDX49 was explored using prediction models taken from AlphaFold database with conserved motifs mapped using chimera software (Fig. 5.1). Amino acids 457-483 were modelled with low confidence by AlphaFold. Structural models reveal a conventional helicase core consisting of the two RecA-like domains, RecA1 and RecA2, with a short N-terminus domain and a canonical DEAD box motif. The most interesting feature is the extended C-terminus domain located from amino acids 383-483 and consisting of three alpha helices of 6, 4 and 11 turns, sequentially, with the latter two consisting of two linear helices separated by a sharp helix-turn-helix motif. This is followed by a disordered region from L462-V483, though this is modelled with low confidence.

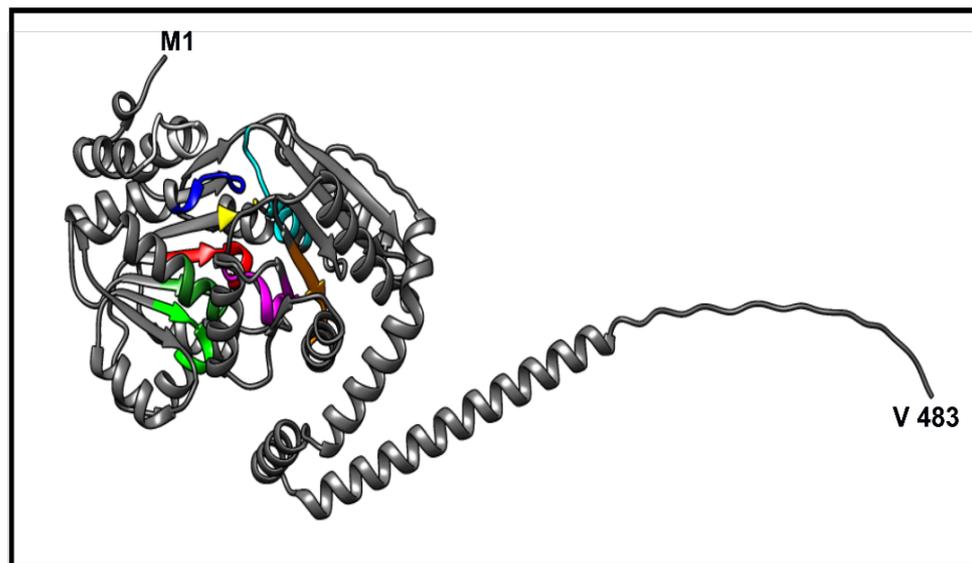
A**B**

Figure. 5.1: Predicted structural model of human DDX49. (A) The predicted model of human DDX49 as predicted by Alphafold software with motifs coloured as depicted in 5.1B (319). (B) The motifs and domains of DDX49 with associated amino acid sequences.

The Alphafold structure was compared with DDX49's purported yeast homolog Dbp8 (Fig. 5.2). Dbp8 predictions contained regions of low (residues 415-427) and very low (residues 428-431) accuracy. Both structural models showed structural conservation within both the helicase core and the first two C-terminus alpha helices, forming an almost perfect superimposition. However, Dbp8 possesses an abridged C-terminus; the helix-turn-helix, 10 turn linear helix and extensive disordered region are completely absent. Sequence alignments (Fig. 5.3) confirmed conservation within the N-terminus and conserved motifs of both proteins, also confirming that the C-terminus extension identified within figure 5.1 is comprised of 52 amino acids and is unique to DDX49.

To further examine the intrinsically disordered region (IDR), the amino acid sequences of both DDX49 and Dbp8 were analysed using IUPred2A server (Fig. 5.4) (320). In agreement with Alphafold predictions, the helicase core and N-terminus of both proteins were predicted to be ordered. Dbp8 and DDX49

were predicted to possess similar scores of intrinsic disorder from residues 404 – 431 and 444 – 483, respectively. Parallel ANCHOR2 scores within DDX49, a representation of the likelihood of a partner protein to trigger a transition from disorder-to-order, suggest that this IDR is likely to function as a disordered binding region. Interestingly, ANCHOR2 scores within Dbp8 do not suggest this is shared amongst both proteins.

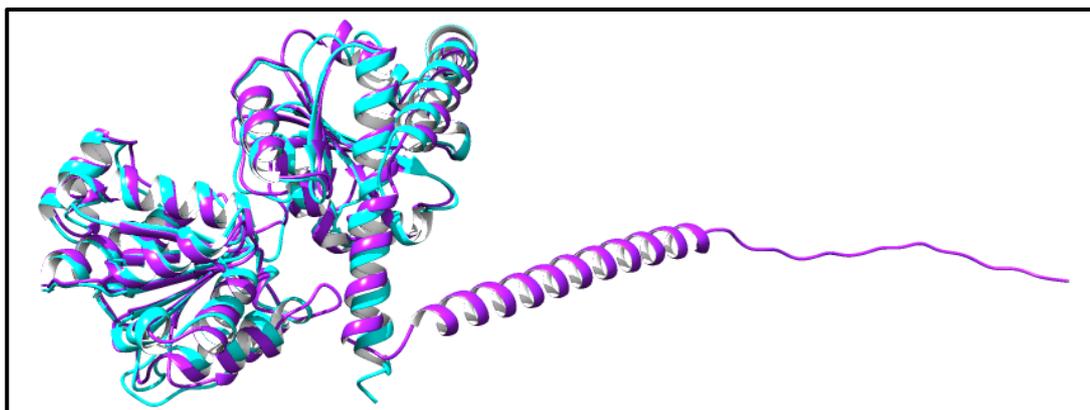


Figure 5.2: DDX49 possesses a unique C-terminus extension compared to its putative yeast homolog Dbp8. Dbp8 is presented in cyan, DDX49 in magenta. Models were taken from Alphafold database and superimposed using Chimera software.

sp P38719 DBP8_YEAST	110.....20.....30.....40.....50.....60
sp Q9Y6V7 DDX49_HUMAN	1	MADFKSLGLSKWLTWESLRANKTQPTAQLCKKICIPKILEGRDCGGAKTGSCKTAFAGCP
sp P38719 DBP8_YEAST	6170.....80.....90.....100.....110.....120
sp Q9Y6V7 DDX49_HUMAN	61	LTKWSEDPSCFGVLTPTRELAMQIAEQFTALGSSNIRVSVIVGGESVQQAALLQRK
sp P38719 DBP8_YEAST	121130.....140.....150.....160.....170.....180
sp Q9Y6V7 DDX49_HUMAN	121	PHFVATPGRLAHHEMSSGDDTVGGMAYLVDEADILLTS---TFADHLATCSAIP
sp P38719 DBP8_YEAST	181190.....200.....210.....220.....230.....240
sp Q9Y6V7 DDX49_HUMAN	178	PKDRQTLLEATTDQKSLQNAAPVKCKPPLPAYQVESVDNVAIPSTKIEYLVEPEH
sp P38719 DBP8_YEAST	241250.....260.....270.....280.....290.....300
sp Q9Y6V7 DDX49_HUMAN	230	VKAYLYQLT--CEYENKALIEFVNRMTAEILLRRTLQLEVRVASLHSONMPQOERTN
sp P38719 DBP8_YEAST	301310.....320.....330.....340.....350.....360
sp Q9Y6V7 DDX49_HUMAN	290	ALAFVSSIYRILLATDVASRGLDIPVQVFNHNTGLEKLYIHRVGRGTARAGRQQAII
sp P38719 DBP8_YEAST	361370.....380.....390.....400.....410.....420
sp Q9Y6V7 DDX49_HUMAN	350	FVTRQD SRQATE RINKK TETNKVHTAVRKALTKVTKA RESL A QKENFGE
sp P38719 DBP8_YEAST	421430.....440.....450.....460.....470.....480
sp Q9Y6V7 DDX49_HUMAN	408	KRQKAKON--GRDPDLEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRAGHKG PPR
sp P38719 DBP8_YEAST	481490.....
sp Q9Y6V7 DDX49_HUMAN	468	PSGSHSGPVPSSQGLV

Figure 5.3: Sequence alignments of DDX49 with its yeast homolog DBP8. Sequence alignments were carried out using T-Coffee server and shaded using Boxshade software (322).

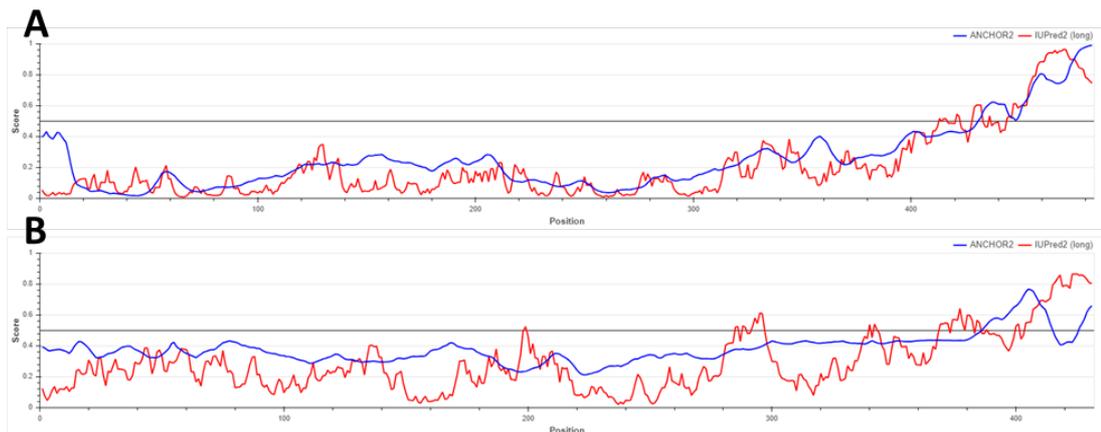


Figure 5.4: IDPR predictions reveal disordered C-termini in both DDX49 and Dbp8. IDPR scores are in red. Anchor2 scores are also presented (blue). Data was taken from IUPred2A server (320). **(A)** *H. sapiens* DDX49. **(B)** *S. cerevisiae* Dbp8.

5.2.1.2. DDX49 C-terminus contains a novel conserved site

Bioinformatic analysis of DDX49 identified a DPD sequence of amino acids near the C-terminus that is highly conserved across all species: highlighted in figure 5.5. In cases where substitutions were noted, this was always a conservative change [D>E]. Structural analysis of the protein found that this site sits within the final helix-turn-helix site located within the C-terminus of the protein.

Query	382	LQILTQVNVVRRCEIKLEAAHFDEKKEINKRQLILEGKDPDLEAKRKAELAKIKQKNRRFKEKVEETLKRQKAGRAGHKGRPPRTPSG	471
ADQ32774.1	382	471
XP_037741377.1	382	.G.....A.....TD.....M.....K..KQ.R.R.QQ..QEK.	455
TFK06454.1	382	.S.....A.....TD.....M.....R..KQ.R.R.QQ..QEK.	455
XP_005279030.1	382	.S.....A.....TD.....M.....M.....R..KQ.R.R.QQ..QEK.	455
XP_039371522.1	382	.S.....T.....TD.....M.....K..KQ.RAR.QQ..QEK.	455
XP_021234262.1	382	.D...H...T.....E..YMD.....M.....E.....K..NK.R...QQ..QEK.	455
XP_034278496.1	382	.K.....R...TD.....M.....MQ.....KQKLK.QSQ.QQ..QERQKQLQ..	462
XP_028569218.1	382	.N.....R...TD.....M.....Q.....RK..VE.RS..QQV.EDK.	455
XP_019393883.1	382	.D.....T.....N.....M.....E...V...K..KQ.R.S.QQA.Q	452
NP_001376236.1	382	.D...H...T.....E..YMD.....M.....E.....K..NK.R...QQ..QEK.	455
XP_038239681.1	412	.G.....A.....TD.....M.....K..KQ.R.R.QQ..QEK.	485
XP_039178237.1	382	.K.....R...TD.....M.....MQ.....K.KLE.QSR.QQ..QERQKQLQ.R	462
NP_989409.1	382	.K.....T.....CTD.....E...KQKKK...HIQ..MQKKREAQ	458
XP_010723344.1	382	.D..NH...T.....E..YMD.....M.....E.....K..NK.R...QQ..QEK.	455
XP_034974886.1	382	.N.....R...TD.....M.....Q.....RK..VE.RSR.QQV.EDK.	455
XP_006028902.1	382	.D.....R...E.....M.....E.....R..KQ.RDS.RQ..Q	452
XP_031459941.1	382	.D..NH...IT.....E..YMD.....MV.....E.....K..NK.R...QQ..QEK.	455
XP_020636636.1	382	.S.....GTD.....M.....EQ.....RK..VQ.RTQ.RQ..E	452
XP_015666211.1	382	.K.....R...TD.....M.....MQ.....K.KLE.QNR.QQ..QERQKQLQ.R	462
XP_030365574.1	382	.D.....T.....E..GMD.....M.....K...QCR...QQV.QKR.	455
XP_019369162.1	382	.D.....T.....N.....M.....E.....R..KQ.RDS.QQ..Q	452
XP_030073184.1	382	.K.....T.....STD...R.K...D.....E...RQK.K..DQIQ...Q.RQ	455
XP_014004839.1	382	.K.....A.....STD.....M.....V.....E..RSEK.K...RIQ..IQKKQH.	457
XP_035268953.1	382	.K.....T.....STD...R...M.....Q.....ER.RSEK.K...RIQ.SIQ.KQ	455
XP_015741597.1	382	.N..NH...T.....E..YMD.....M.....E.....N..K..NK.R...QQ..QER.	455

Figure 5.5: BLASTP data shows conservation of a DPD amino acid sequence when comparing DDX49 to non-mammalian species. Data was

taken from BLASTP and curated so that duplicate results for each species were removed (408).

5.2.2. Purification of recombinant hDDX49

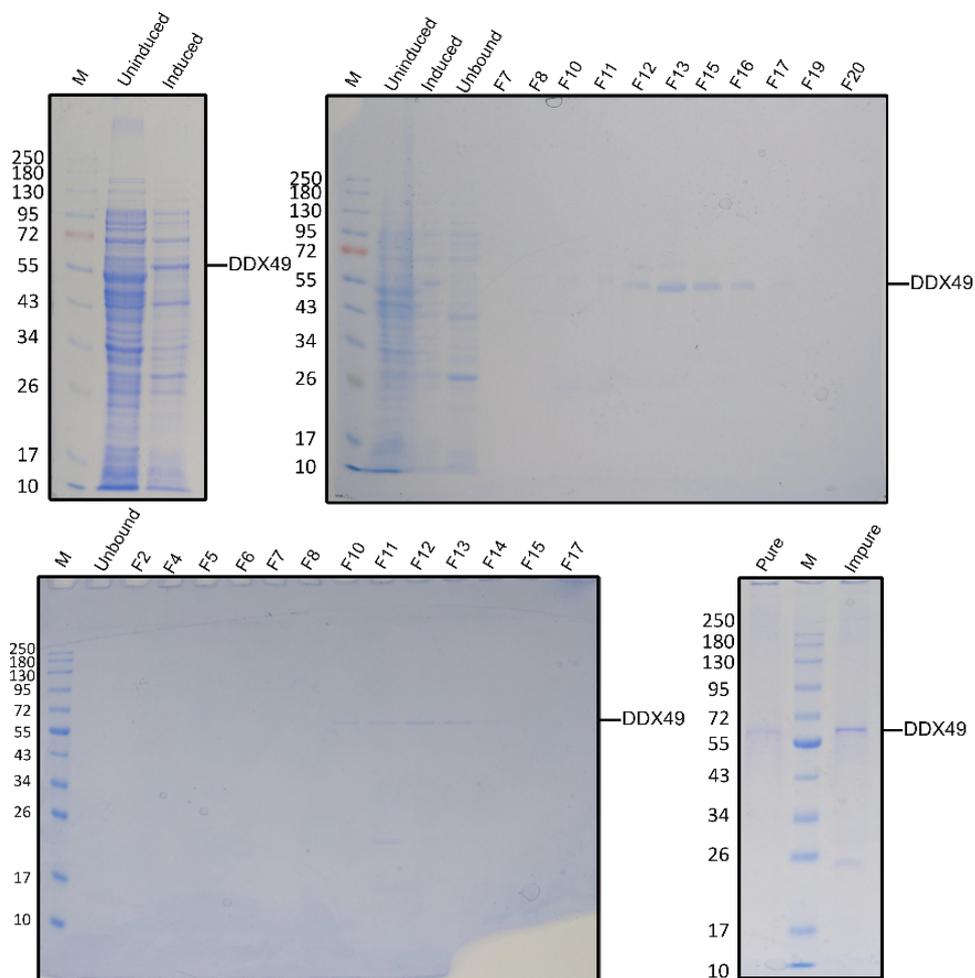


Figure 5.6: Purification of recombinant DDX49. DDX49 was overexpressed in the BL21-AI strain of *E. coli* and methods carried out as described in section 2.4. DDX49 had an expected mW of 60.5 kDa. A) DDX49 overexpression was confirmed in a scaled-down induction assay, comparing uninduced and induced *E. coli* BL21-AI. B) Protein lysate was loaded onto a HisTrap HP His-tag column and DDX49 positive fractions identified by SDS-PAGE prior to dialysis into low salt containing buffer. C) Sample was purified using a HiTrap™ Heparin HP column and DDX49 positive fractions were pooled separately: fractions 10-12 were dialysed as 'impure DDX49' and fractions 13-15 dialysed as 'pure DDX49' into storage buffer. D) 'Impure' and 'pure' protein fractions were stored and intact DDX49 was confirmed via SDS-PAGE.

DDX49 was cloned, expressed, and purified as described in section 2.4. Purification was successful and both 'impure' and 'pure' fractions of DDX49 collected for downstream testing (Fig. 5.6). To examine the ATP dependency of DDX49 activity and further validate helicase activity, a motif II mutant (D152A/D155A) was constructed based upon similar experiments by Garbelli et al (2011). Following interesting data collected in initial screens, two further mutants were purified to examine a well-conserved potential nuclease site within the C-terminus; a D422A/D424A and a K421A mutant. These were cloned and purified in collaboration with students Fiorela Kapplanaj and Louise Martin, respectively. It was noted that the preparation of K421A possessed substantially more contaminants than other preparations.

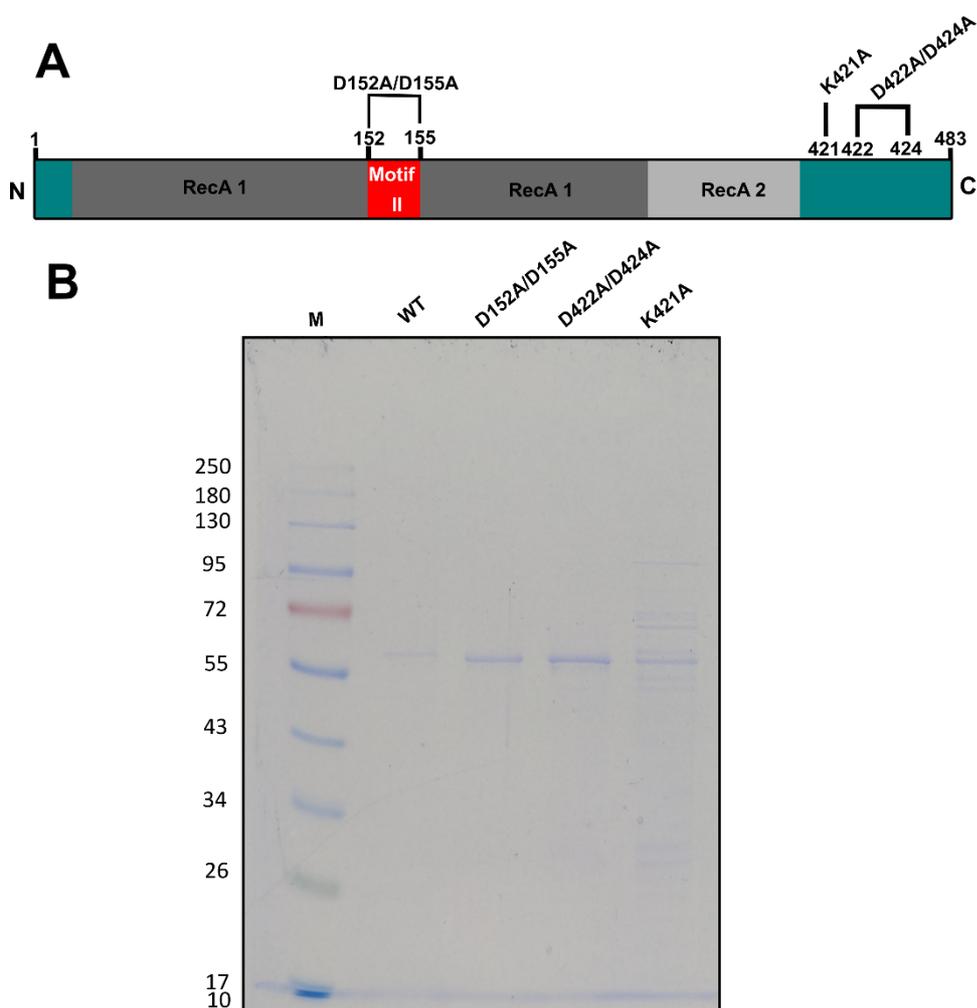


Figure 5.7: Summary gel of purified recombinant wild type DDX49 and the mutants used within this study. Expected molecular weight of all proteins was 60.5 kDa.

5.2.3. Mass Spectrometry of DDX49 contaminant

A notable contaminant of ~24 kDa consistently copurified with DDX49 across multiple batches (Fig. 5.6D). The impure sample was run on an SDS-PAGE gel, the band excised and sent for mass spectrometry analysis externally to identify it. To minimise interference in downstream mass spectrometry, gel was stained using ruby staining prior to band excision (Fig. 5.8).

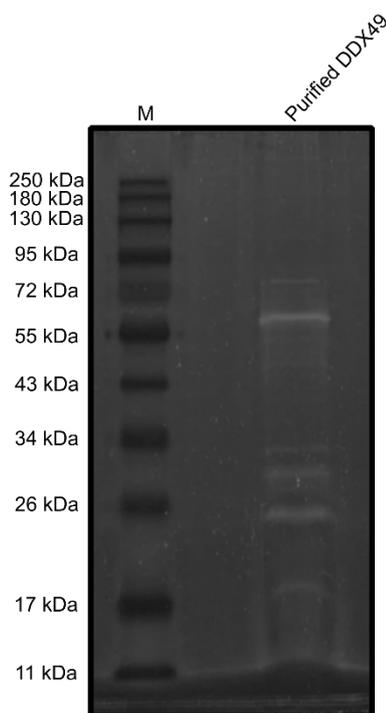


Figure 5.8: The impure fraction of DDX49 as stained by RUBY staining and viewed under UV. The ~20 kDa band was extracted and sent for external mass spectrometry analysis (see section 2.5.4).

Mass spectrometry analysis identified the contaminant as the *E. coli* protein Crp based upon % of total spectra and unique peptides identified (table 5.1). A number of human epidermal proteins were identified in spectra, e.g. keratin, and were presumed contaminants from sample handling and extraction so were excluded from analysis. Fragments of DDX49 were identified along with several additional *E. coli* proteins, but the overall % of the total spectra was low suggesting low abundance.

Table 5.1. Proteins identified within mass spectrometry analysis with unique peptides >2. Protein threshold was set at 95% with a minimum number

of peptides of 2. Excluding DDX49, human proteins present were assumed to be contaminants from skin and are omitted from this table.

Name	Unique peptides	% of total spectra	Protein mW
cAMP-activated global transcriptional regulator CRP	21	3.5	24
Probable ATP-dependent RNA helicase DDX49	10	0.19	54
50S ribosomal protein L3	9	0.22	22
GTP cyclohydrolase 1	8	0.12	25
30S ribosomal protein S4	6	0.09	23
RNA-binding protein Hfq	4	0.12	11
Transcriptional regulator Yqjl	5	0.09	23
60 kDa chaperonin	6	0.08	57
FKBP-type peptidyl-prolyl cis-trans isomerase SlyD	3	0.06	21
ATP synthase subunit beta	5	0.08	50
Isocitrate dehydrogenase [NADP]	3	0.04	46
NADH-quinone oxidoreductase subunit E	4	0.06	19
Ferric uptake regulation protein	4	0.05	17
DNA-directed RNA polymerase subunit beta	3	0.04	155
Aconitate hydratase A	3	0.05	98
Protein InaA	3	0.05	25
p-aminobenzoyl-glutamate hydrolase subunit A	3	0.04	47
Chaperone protein DnaK	3	0.04	69

5.2.4. Assessing activity of wild type and mutant DDX49

5.2.4.1. WT-DDX49 binds RNA with a slightly higher affinity than DNA

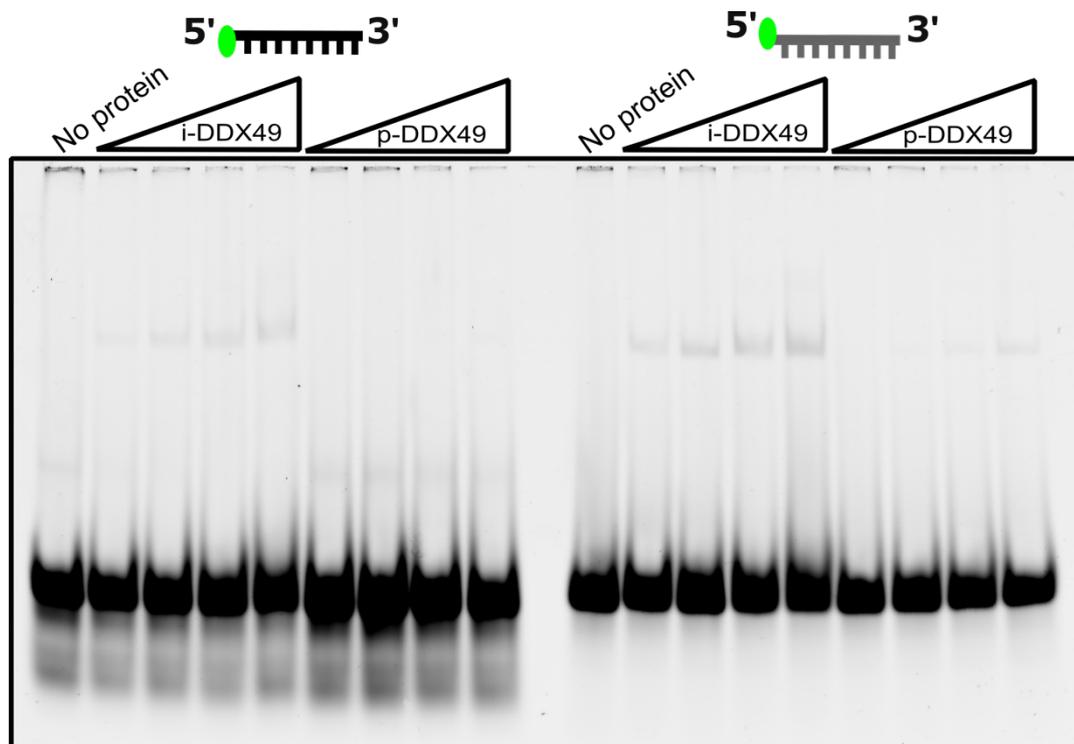


Figure 5.9: Electromobility shift assay for binding affinity of ‘impure’ (i-DDX49) and ‘pure’ (p-DDX49) DDX49 against nucleic acids. EMSA was carried out as described in section 2.6.5. Concentration ranges included 100, 200, 500 and 1000 nM protein. ssDNA substrate (represented in black) was 50-NT and ssRNA (represented in grey) substrate was 45-NT.

The binding affinity of DDX49 to several nucleic acid species — notably ssDNA, ssRNA and flayed duplex DNA — as a function of protein concentration was assessed by electrophoretic mobility shift assays (EMSA) as described in section 2.6.5. Initial studies (Fig. 5.9) used a 50-base pair Cy5 labelled ssDNA oligonucleotide (MW14) and a 45-base pair Cy5 labelled ssRNA substrate (AP33). Later studies (Fig. 5.10) used both 45-base paired Cy5 labelled ssDNA (AP122) and ssRNA oligonucleotides (AP33). Both fractions bound nucleic acids with poor affinity, reflected within the high exposures required to view band shifts (Fig. 5.9). Impure fraction pool formed stable nucleoprotein complexes with DNA and RNA with low affinity, whilst stable complexes were only noted within incubations with RNA for the pure fraction pool. Both sets of fractions showed a slightly increased affinity to the ssRNA substrate. Despite

increased binding seen within impure fractions, the 'pure' fraction was used for subsequent assays as it was considered more reliable as an indicator of DDX49 activity.

The binding affinity of DDX49 mutants was analysed on ssDNA and ssRNA (Fig. 5.10 A +B). As previously shown, very weak band shifts were seen when ssDNA was incubated with wild type DDX49. Both D422A/D424A and K421A mutants showed no binding to either nucleic acid, whilst motif II mutants showed smearing patterns and evidence of well aggregation at concentrations of 1000 nM of protein. Interestingly, incubations of RNA with both wild type and the K421A mutant exhibited apparent nuclease activity of the RNA. The cause of this is unknown, but it is noted that these assays do not contain the Mg^{2+} or ATP used in other assays to examine DDX49 nuclease activity.

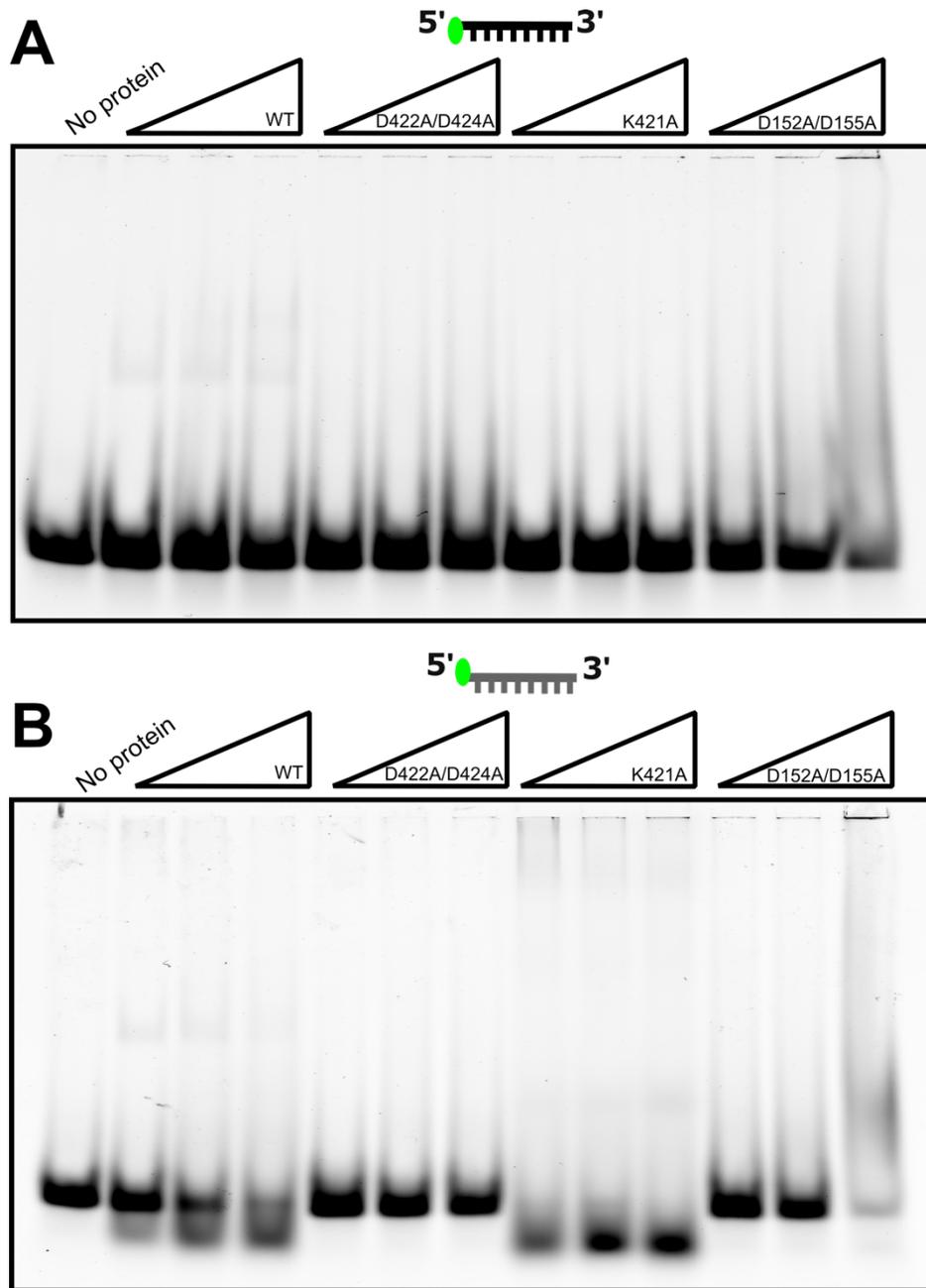


Figure 5.10: Electromobility shift assay shows C-terminus mutants fail to bind DNA and RNA. Protein concentrations are 200, 500 and 1000 nM, respectively. **(A)** EMSAs conducted on a 45-NT ssDNA oligonucleotide. **(B)** EMSAs performed on a 45-NT ssRNA oligonucleotide.

5.2.4.2. DDX49 has a 3' to 5' polarity of unwinding and possesses nuclease activity on the lagging strand

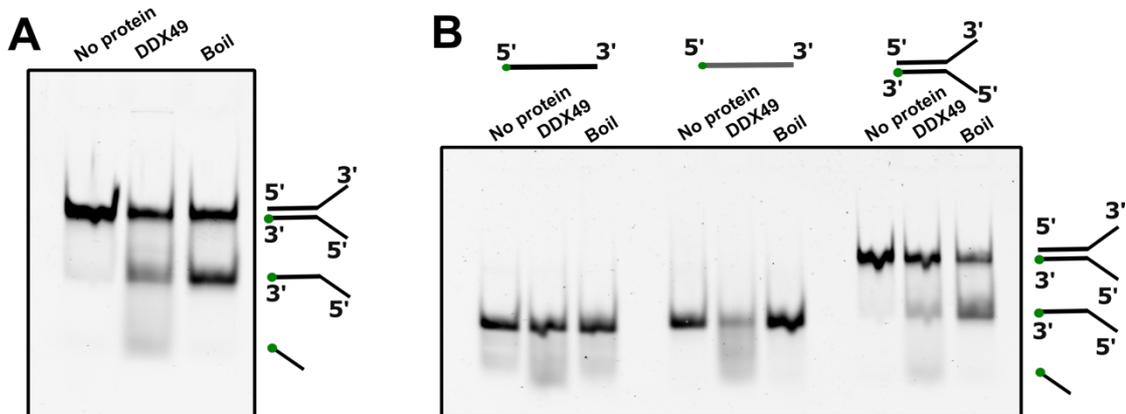


Figure 5.11: Examining DDX49s activity on DNA substrates in vitro. Unwinding was analysed through incubation of a DDX49 concentration of 500 nM with 25 nM of a flayed duplex DNA substrate as described in section 2.6.1. (A) DDX49 was tested against an open flayed duplex to assess initial activity. (B) DDX49 was tested on ssDNA, ssRNA and a flayed DNA duplex.

Unwinding activity of DDX49 was assessed in experiments using a flayed duplex (Fig. 5.11A) using methods described in section 2.6.1. Unwinding was successfully observed but a second and smaller band was noted. This activity was replicated in repeat experiments and across multiple batches of purified DDX49, including mutants of DDX49. As the third band migrates further than the boiled substrate, this activity must be on the labelled strand: in this case, the lagging strand. A subsequent assay confirmed that nuclease activity appeared to also be present on both ssDNA and ssRNA (Fig. 5.11B), however formation of a distinct band as seen within flayed duplexes was not observed.

mechanism for 3'-5' loading and unwinding of DDX49 on DNA strands, with apparent nuclease activity on the 5' flap. Polarity was confirmed using FRET-based analysis (Fig. 5.12B) which similarly indicated a lack of unwinding activity on substrates featuring 5' flaps. Substrates with an open fork and a 3' flap were both unwound, with the open fork showing the largest amount of unwinding.

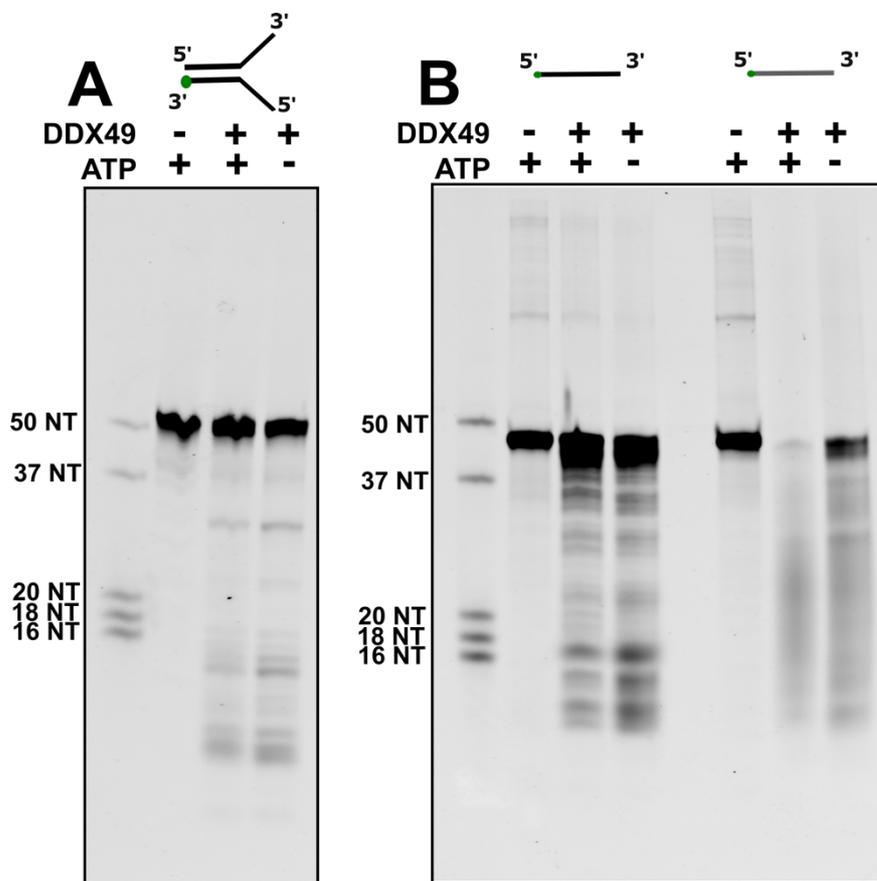


Figure 5.13: DDX49 cuts DNA but not RNA in an ATP-independent manner. Helicase assays were carried out as described in section 2.6.1 and subsequently run on a 15% denaturing gel. Both figures were run on the same gel. **(A)** Nuclease activity on a 50 NT flayed duplex substrate **(B)** Nuclease activity on (l-r) 45 NT ssDNA and ssRNA substrates.

To study nuclease activity in higher detail, a series of substrates were incubated with DDX49 in the presence and absence of ATP prior to running on a denaturing gel. Figure 5.13 A + B were run on the same gel but are separated for convenience. Nuclease activity was recorded for all substrates, with nuclease activity much higher on ssRNA samples in comparison with both ssDNA and flayed duplex, to the point of almost complete digestion. Nuclease

activity on ssRNA was demonstrably reduced in the absence of ATP implying a mechanism where ATP assists within the cleavage of RNA. ssDNA (45-NT) predominantly formed cleavage products of ~37-NT and ~16-NT. No substantial difference was seen between ATP +ve and -ve samples for other nucleic acid species.

5.2.4.3. DDX49 helicase activity is ATP-dependent and regulated by the C-terminus

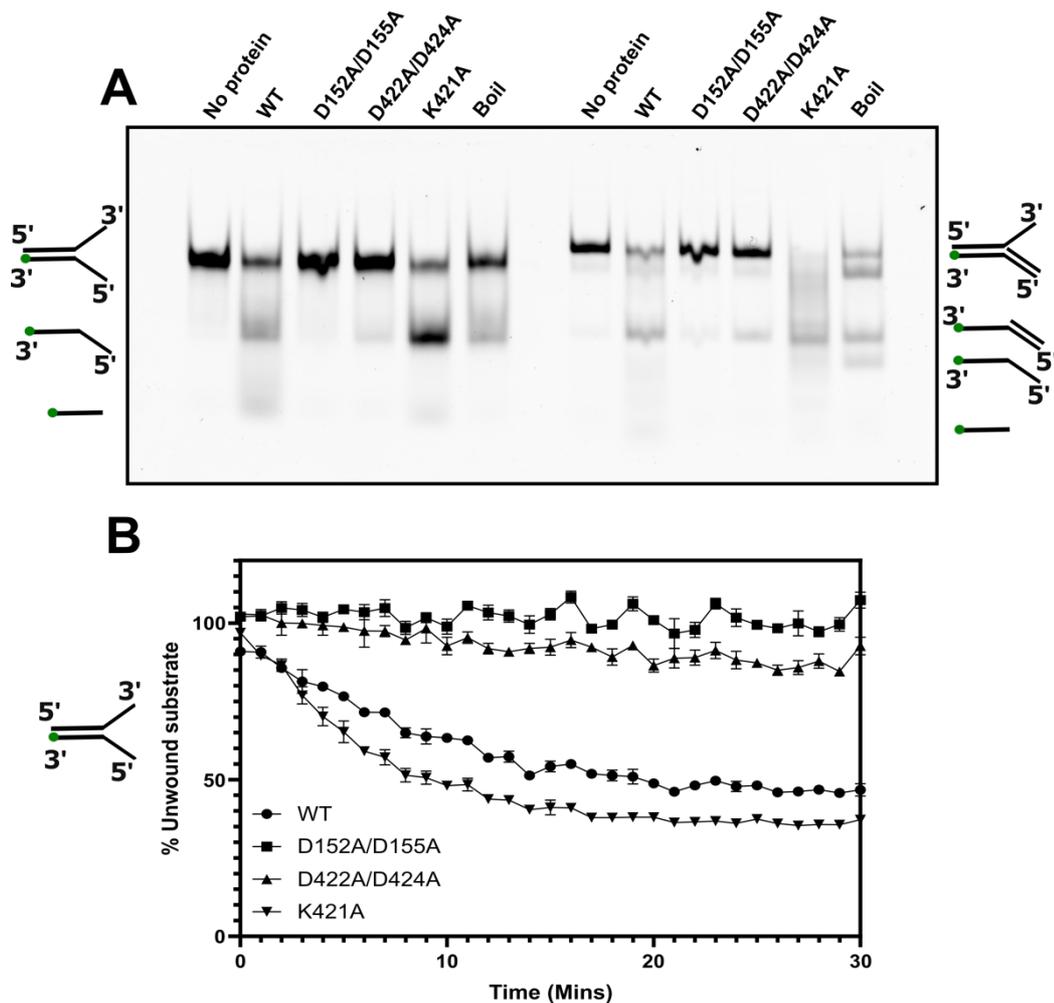


Figure 5.14: Summary of unwinding and nuclease activity of mutants. (A) DDX49 and associated mutants (500 nM) were tested as described in section 2.6.1. against flayed duplexes with open and 3' flaps (25 nM), respectively, to determine unwinding and nuclease activity. Controls include a substrate only reaction and a boiled reaction intended to represent full dissociation. **(B)** DDX49 (500 nM) was tested as described in section 2.6.2 against flayed duplex with

open flaps (50 nM). Samples were run in duplicate and normalised against reactions containing no protein. Standard error is shown.

The purified mutants described in section 5.2.2 were incubated with and without 3' flaps (Fig. 5.14A). No nuclease activity was seen within both motif II and D422A/D424A mutants, with greatly reduced nuclease activity seen within the K421A mutant. The motif II mutant possessed no helicase activity, indicating ATP hydrolysis is required for effective unwinding, whilst the D422A/D424A mutant showed weak helicase activity. Interestingly, the K421A mutant appears to possess hyperactive helicase activity in comparison to wild type protein, with almost 100% dissociation seen within the 3' flap substrate. This implies a model whereby aspartic residues 422 and 424 play a key role within nuclease activities of DDX49, with K421 acting as a regulator between nuclease and unwinding activities of the protein.

5.2.5. CRISPR-Cas9 mediated gene editing of *DDX49*

Alongside studying DDX49 biochemistry using recombinant DDX49, a human cell model was of interest. CRISPR-Cas9 gene editing was performed using methods outlined in section 2.8.5 and the optimised workflow developed and described within chapter 4. Initial guide RNA testing was performed by the author and the full knockout workflow carried out in collaboration with MRes student Sabesan Anandavijayan.

5.2.5.1. gRNA testing

Table 5.2. An overview of gRNAs tested for editing *DDX49* within this study. On-target and off-target scores as predicted by Integrated DNA Technologies online tool are presented.

Gene	Exon	Sequence	PAM	On-target score	Off-target score
<i>DDX49</i>	2	AAGCTGTCTG AGGATCCCTA	TGG	66	33
<i>DDX49</i>	3	AAAGACTGCA TCATCGTCGG	TGG	66	91

<i>DDX49</i>	4	GCTCTCTCGG AAACCACACG	TGG	74	79
--------------	---	--------------------------	-----	----	----

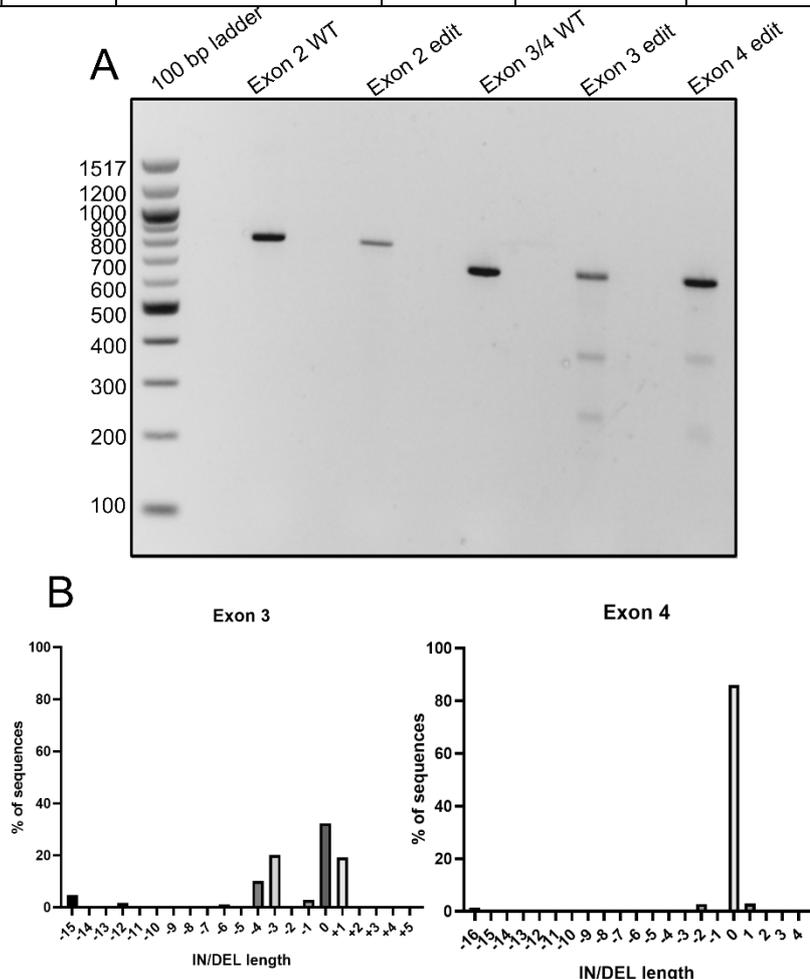


Figure 5.15: gRNAs targeting exon 3 and exon 4 successfully edited genomic DNA. (A) T7 endonuclease activity on PCR amplicons of the genomic DNA flanking the editing site. (B) Recorded IN/DEL lengths in exons 3 and 4 noted in TIDE (tracking of indels by decomposition) analysis of Sanger sequenced PCR amplicons of the genomic DNA flanking the respective editing site.

Human U2OS cells were edited using gRNAs as described in section 2.8.5, genomic DNA extracted and regions flanking the editing site amplified by PCR. PCR amplicons were digested with T7 endonuclease and within exon 3 and 4 edits, bands were noted at ~350 and ~250 bps confirming successful editing (Fig. 5.15A). Exon 3 cleavage products were more intense than those seen in exon 4. Amplicons were subsequently sent for Sanger sequencing and

analysed by TIDE analysis (Fig. 5.15B): as exon 2 returned a negative T7 assay result, this was not sent for sequencing. Consistent with the T7 assay results, high levels of editing were noted within exon 3 with deletions up to 15 BP and only 30% of DNA unchanged in comparison to the wild type. Interestingly, TIDE results for exon 4 were comparably low in comparison to those seen in the T7 assay in addition to less than expected given on-target efficiencies predicted in table 5.2. The most prominent edits were single base pair insertions and 3 base pair deletions within exon 3.

5.2.5.2. *DDX49* heterozygotes were successfully generated

Based upon gRNA tests, gRNA targeting exon 3 was used for the full editing run. A full gene editing run was carried out as described in section 2.8.5 and after 48 hours cells diluted to 0.5 cells per well across two 96 well plates. Remaining cells had their DNA extracted, the editing site amplified and were submitted to TIDE sequencing to evaluate editing efficiency (Fig. 5.16). High levels of editing were recorded with a total efficiency of 69.7%.

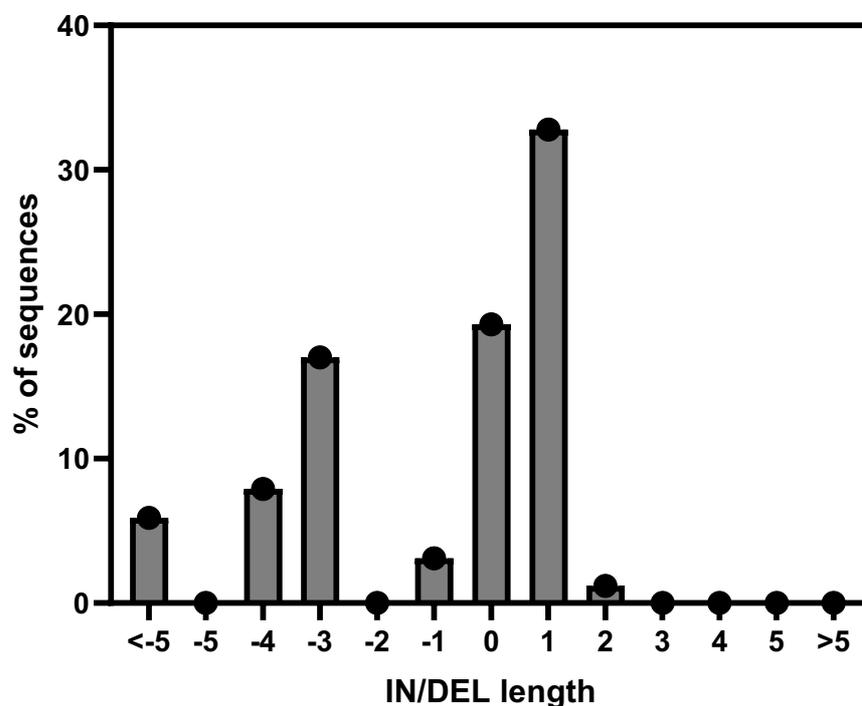


Figure 5.16: *DDX49* editing shows high efficiencies. Exon 3 of the edited cells was sequenced and compared with wild type cells by TIDE analysis. Only P-values of >0.05 were measured.

Several single cells studied during outgrowth were not viable as colony establishing units, replicating results seen within the workflow when carried out with *DDX52*. Wells showing single colony forming units were scaled up and following sequencing a successful heterozygote gene edit was identified within well A3 corresponding to a 19 bp deletion (Fig. 5.17).

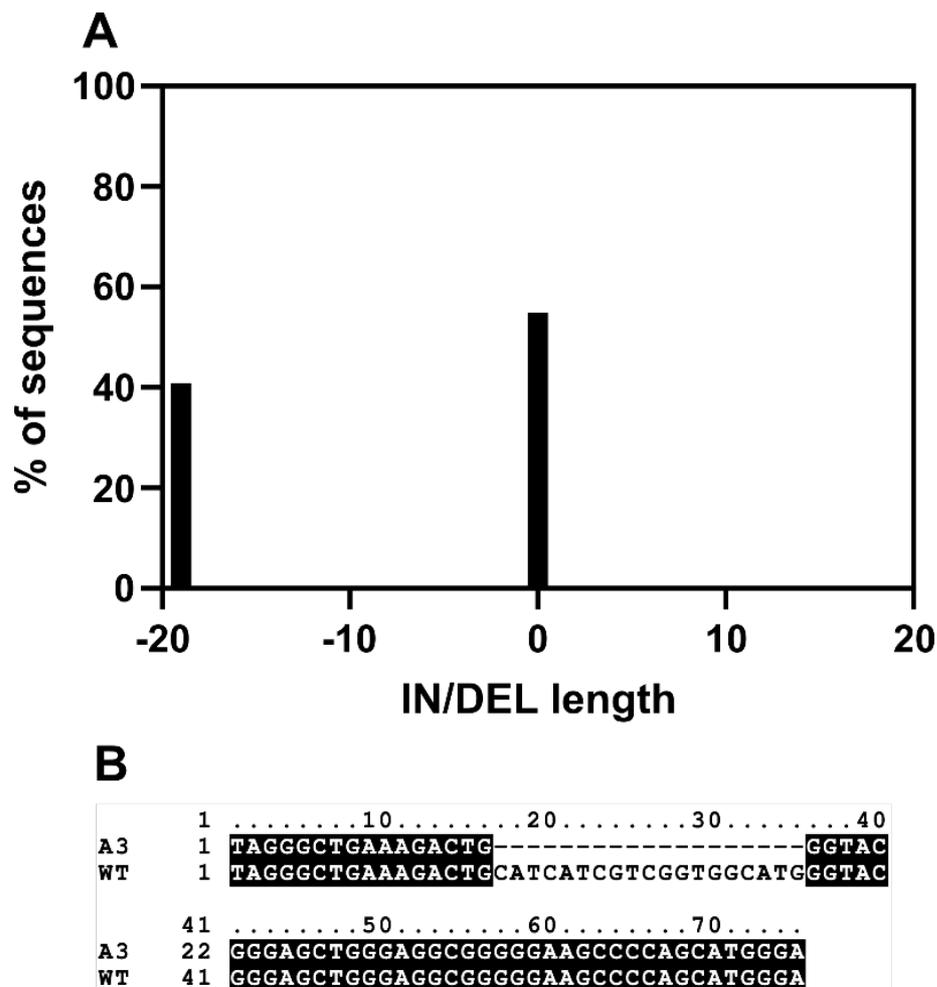


Figure 5.17: A *DDX49* heterozygote cell line was generated with a 19 bp deletion. (A) Cells were sequenced and compared with wild type cells by TIDE analysis. **(B)** Alignments indicate that a 19-base pair deletion was generated.

5.2.5.3. *DDX49* knockout cells demonstrate inhibited cell migration

Initial phenotyping was carried out by MRes student Sabesan Anandavijayan and their work is presented here using WST-1 assays and wound healing (scratch) assays. Having reviewed the students WST-1 assay data in the course of writing this report, it was concluded that due to perceived differences at time

0 data was inappropriate for normalisation. The rate of change of absorbance is presented instead (Fig. 5.20). Slopes were tested for significant differences and no significant difference in rate of growth was seen between the two.

To phenotype *DDX49*^{+/-} U2OS cell lines migration in comparison to wild type cells, a scratch-based 'wound-healing' assay was employed to examine if cellular migration was impacted over time (Fig. 5.21). Wild type cell lines were observed to have recovered to 100% confluency after a period of 48 hours, whilst heterozygous cell lines still had a sizeable wound indicating reduced migration. *DDX49*^{+/-} heterozygous cells recovered to full confluency after a period of 72 hours.

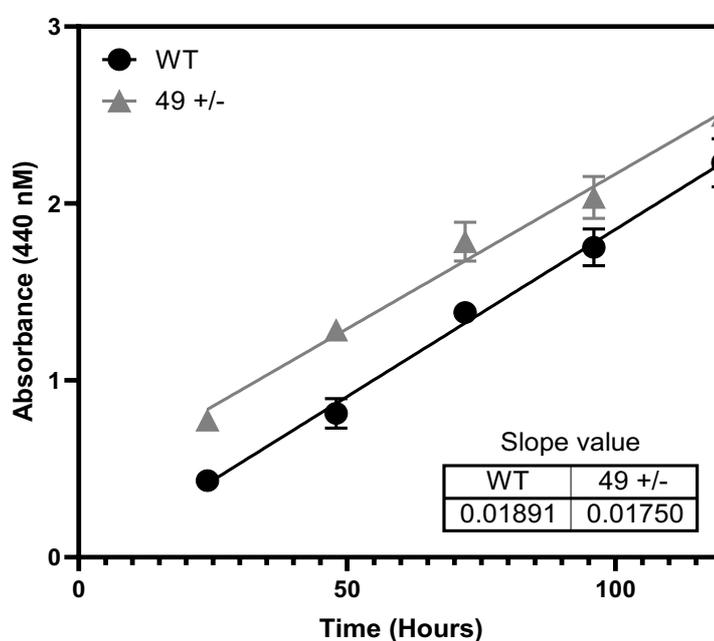


Figure 5.18: *DDX49*^{+/-} cells show no difference in growth rate when compared with wild type cells. Assay performed by student Sabesan Anandavijayan and results reinterpreted by the author. Statistical analysis was performed with PRISM software.

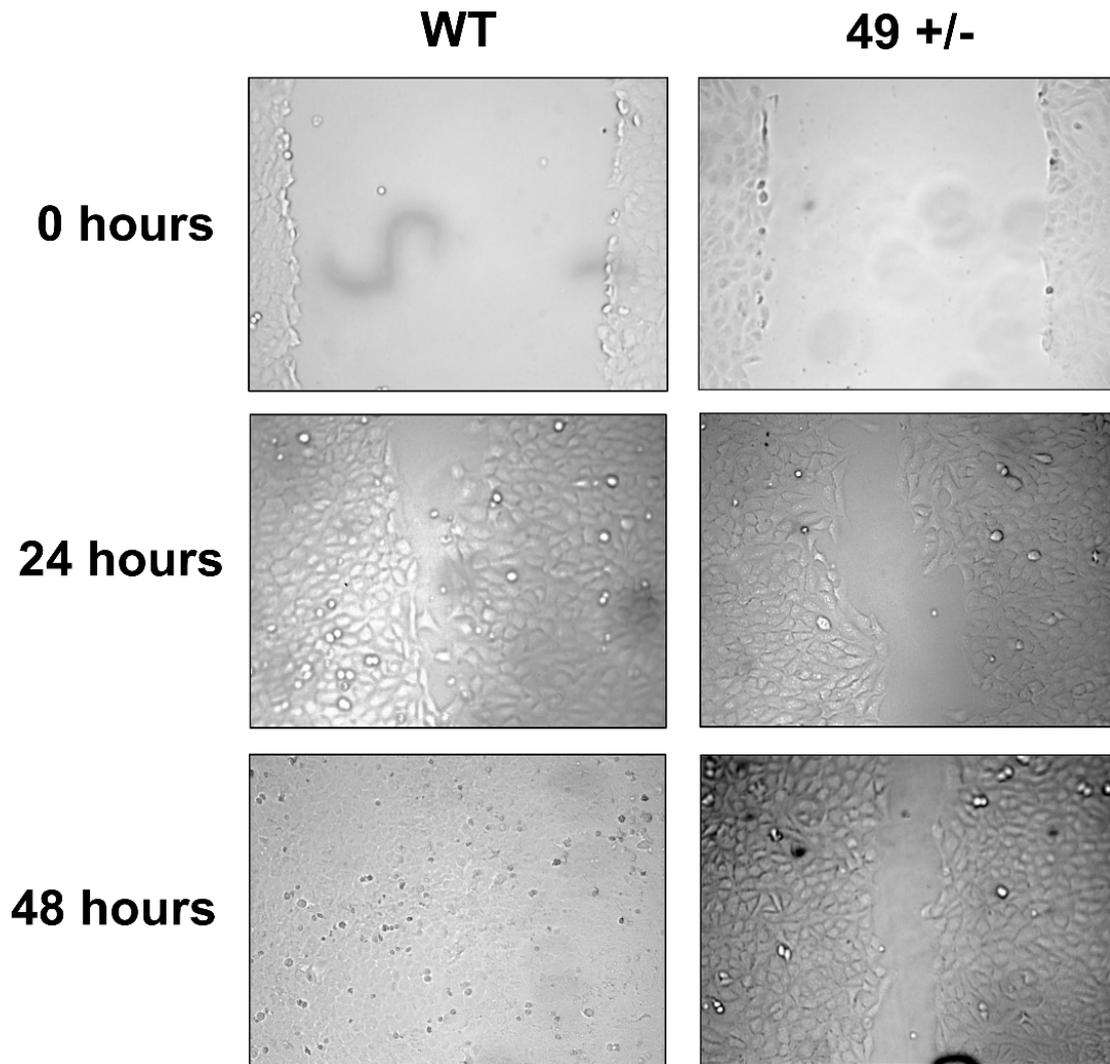


Figure 5.19: *DDX49* heterozygotes show inhibited migration within wound healing assays. Cells were incubated over a 48 hour period and imaged with an Axiovert S 100 microscope. Assay performed by MRes student Sabesan Anandavijayan

5.2.5.4. *DDX49*^{+/-} impacts abundance of 47S, 28S and 5.8S rRNA species

Using the RT-qPCR method described in section 2.2.5, the abundance of several rRNA species were compared with those seen within wild type cells. Housekeeping gene levels were of high quality and rRNA species were normalised using GAPDH. Abundance of 47S rRNA was increased by ~1.7 fold, whilst abundance of 28S and 5.8S rRNA species was around two-fold lower than levels within the wild type. Levels of 18S rRNA were decreased but not to a notable extent.

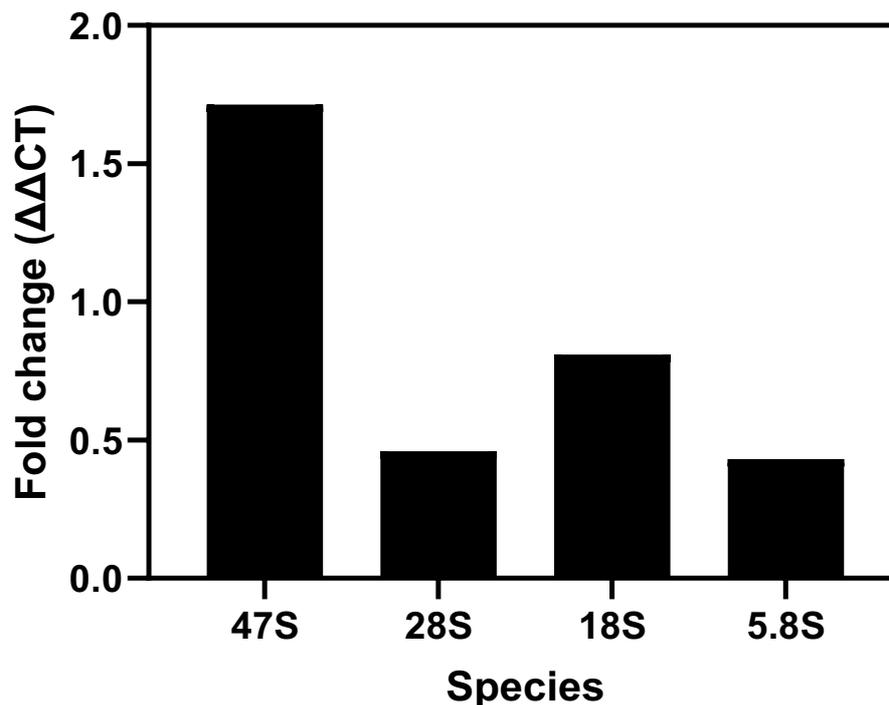


Figure 5.20: Cycle thresholds in qPCR shows the abundance of 47S, 28S and 5.8S rRNA species is impacted within *DDX49*^{+/-} cells. Fold differences were calculated through comparisons of $\Delta\Delta\text{CT}$ values with wild type cells, with wild type rRNA levels set as '1'. Sets of data for both housekeepers were consistent and samples normalised to GAPDH signals. N = 3.

5.2.6. CRISPR-Cas9 prime editing of U2OS cell lines

To further examine the impact of perturbations in vitro, a prime editing technique based upon work by Anzalone et al (2019) and described in section 2.8.6 was used to introduce specific mutations in lieu of a knockout. As this was a new technique, testing and optimisation was necessary. For testing, a D421A/D423A mutation was chosen to test the efficiency of editing. This was based upon the justification that it was hypothesised that, as *DDX49* is labelled an essential gene, site-based editing of the DExD motif thought to be essential could create a similar phenotype to a knockout. As the identified DPD motif was, at this point, considered to serve an ancillary function it was hypothesised that this could be edited to produce a less severe phenotype.

Components of the prime editing reaction were constructed in collaboration with MRes student Fiorela Kapllanaj and undergraduate student Louise Martin with transfections carried out by the author. To confirm any visible effects were not the product of contamination or endotoxin within the plasmid stocks, transfection controls were included consisting of each component plasmid separately as well as transfection reagent controls to determine reagent toxicity (Fig. 5.22). As can be seen, in comparison with media only controls the cells transfected with all components (hence, 'prime edited') were notably sick and unviable. Cells transfected with p169850 – Cas9 nickase (H840A) fused to M-MLV RT – showed potential signs of yeast contamination, though it is unclear if this is cell debris. In all controls tested, the cells were substantially healthier than transfection with all components, indicating that the prime editing reaction performed here renders cells unviable.

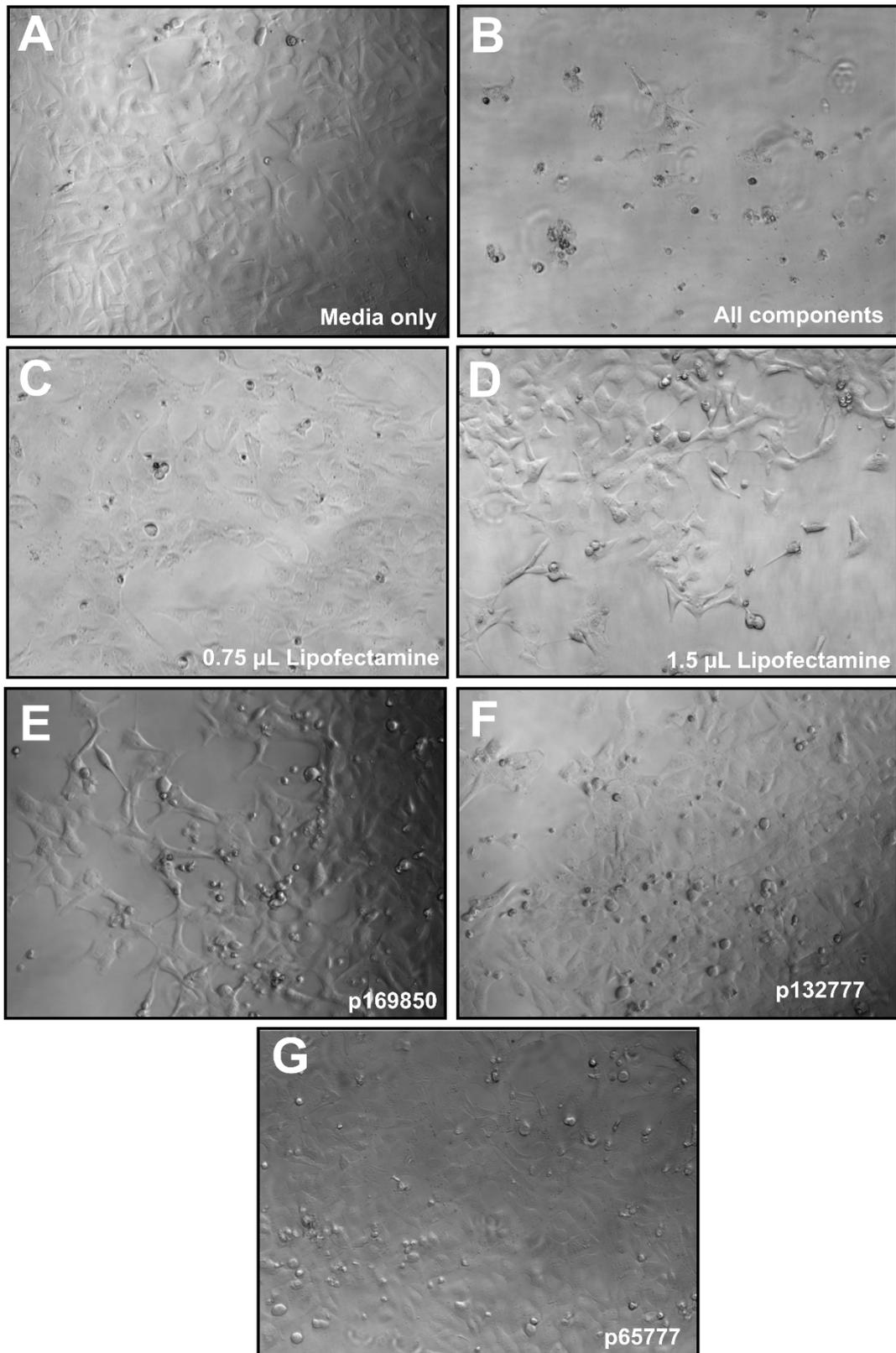


Figure 5.21: Prime editing of *DDX49* D422A/D424A leaves cells unviable. (A) A media only control. (B) U2OS cells transfected with all components of the prime editing reaction. (C+D) U2OS cells transfected with only 0.75 uL and 1.5 uL of lipofectamine transfection reagent. (E) U2OS cells transfected with

plasmid 169850 (M-MLV RT fused to Cas9 (H840A) nickase) (F) U2OS cells transfected with plasmid 132777 (pegRNA) (G) U2OS cells transfected with plasmid 65777 (nicking guide RNA).

5.3. Discussion

5.3.1. DDX49 has a characteristic DExD-box domain with a unique C-terminus extension

At the time of writing, as with other DExD-box proteins within this study, there is no readily available protein structure or bioinformatic analysis of DDX49. An initial examination was performed to examine the structure of the protein and compare it with its putative yeast homolog Dbp8. It was noted that both proteins possessed an intrinsically disordered C-terminus of varying lengths along with a conserved RecA-like helicase core domain, in accordance with expectations as unique features within the N/C-termini are well-defined within DExD-box proteins (40).

Comparisons between the two proteins revealed notable differences within the C-terminus domains, with DDX49 showing an extended C-terminus and a helix-turn-helix motif absent within Dbp8. The extent of structural conservation suggests that the two proteins are related, but the extent of the C-terminus extension represents a key difference within the two proteins. As discussed within chapter 3, human DDX52 shows similar extensions within the protein in comparison to its putative yeast counterpart. It is hypothesised that these extensions may arise in conjunction with the increasing complexity of human ribosome biogenesis in comparison to yeast (120). Notably, the yeast protein has not been reported to be involved with the same processes of mRNA export or poly(A)⁺ export as DDX49: these functions in yeast have been ascribed to the DExD-box protein Dbp5 instead (409).

Strikingly, the helix-turn-helix that marks the site where the proteins diverge is the proposed site of nuclease activity reported later within this study. No nuclease activity has been reported within studies on Dbp8. However, it is considered that the IUA2Pred predictions for Dbp8 predict an intrinsically disordered region within the C-terminus that appears to be clearly ordered within Alphafold predictions (Figs. 5.1 and 5.4). This region is modelled with low

confidence by AlphaFold and it has been reported to lack accuracy for intrinsically disordered regions, illustrating limitations of the software and the importance of employing multiple bioinformatic tools in assessing proteins (410). Despite uncertainty, the size of the C-terminus extension makes it a likely candidate for differences in activity between the two proteins.

5.3.2. *E. coli* Crp was consistently copurified with DDX49

Purification of recombinant DDX49 was successful, but throughout purification a major contaminant was consistently reproduced across batches. Following unexpected downstream results, the band was excised and sent for mass spectrometry analysis and was identified as the *Escherichia coli* protein Crp. In addition, DDX49 fragments were noted confirming successful expression. Due to the size difference between the two proteins, further purification of DDX49 using size-exclusion chromatography was considered: however this was avoided due to poor stability and yield in wild type DDX49 purifications.

Crp is an *E. coli* DNA binding protein associated with transcriptional regulation. It was noted that Crp features the sequence HxHxH, potentially sufficient for binding of the protein to the chelating column (411). However, it is noted that Crp was not noted to elute within purifications of other proteins in this report, such as DDX52 and its associated mutants. Alternatively, it is hypothesised that DDX49-Crp interactions result in Crp eluting with DDX49. Whilst of potential interest, based upon differences between *E. coli* and a lack of homologous proteins to Crp in humans, a potential interaction was not examined further.

5.3.3. DDX49 C-terminus forms binding complexes with DNA

Published research has not yet compared RNA and DNA binding affinities of DDX49. The results for the purer fraction on DNA and RNA (Fig. 5.9) appear to show a preference for RNA over DNA, with complexes seen in incubations with RNA at 200 nM and above, whilst no complexes were seen within DNA incubations. Overall band intensity was weak and high exposures were needed to see complexes, implying a low affinity of DDX49 for both nucleic acids. It is suggested that this low affinity might act as a self-regulatory mechanism to reduce background nuclease activity evidenced within this report. The impure fraction showed increased binding affinity to DNA compared to the pure fraction.

Combined with mass spectrometry data discussed within section 5.3.2, it is proposed that increases in affinity are contributed by the DNA binding contaminant Crp (412).

All mutants showed atypical binding to both nucleic acids. In the case of the D422A/D424A, nucleic acid binding was completely abolished. It is proposed that this is a key site within nucleic acid binding of DDX49, supported by its location within the helix-turn-helix structural motif: helix-turn-helix domains are well documented as sites of DNA binding (413). Motif II mutants further demonstrated this, forming smearing patterns and well aggregation similar those seen within studies of wild type DDX52 (section 3.2.4). Whilst nucleic acid recognition is typically associated with the RecA2 domain within DExD-boxes, involvement of both domains has previously been reported (67). Smearing could indicate weak association or that mutations within the motif II site facilitate the formation of multiple nucleoprotein species. Interestingly, nuclease activity was seen within incubations of RNA with wild type and K421A mutants, indicating nuclease activity occurs independently of ATP and MgCl₂ in the presence of RNA and implying a preference for RNA. This concurs with the canonical role of DExD-boxes as RNA helicases. It is unclear why the wild type did not exhibit this within initial assays (Fig. 5.9) but the possibility of RNase contamination is not excluded.

5.3.4. DDX49 is an ATP-dependent 3' to 5' helicase

As a member of the superfamily II family of helicases, it was anticipated that DDX49 would unwind DNA with a 3' to 5' polarity. This was confirmed within assays that compared helicase activity on flayed duplexes possessing 3' and 5' flaps (Fig. 5.12). As shown within helicase assays, DDX49 possesses poor proficiency as a helicase. This is not unexpected: unwinding by DExD-box helicases has previously been shown to require a large excess of protein to substrate, with multiple protein:DNA complexes needed for full separation of longer duplexes (41, 65). It is interesting to note that the study by Awasthi et al (2018) reported DDX49 as a robust helicase, however their study used smaller substrates than those used throughout this study (82) .

Motif II mutants displayed a lack of unwinding, indicating unwinding is ATP-dependent. However, it is noted that EMSAs (figure 5.10) show the binding characteristics of motif II mutants are altered which may also impact strand separation. Curiously, DDX49 K421A mutants displayed hyperactivity as helicases despite a lack of binding in EMSAs at concentrations two-fold higher than those used in helicase assays. It is hypothesised that this lysine residue may act to regulate the helicase and nuclease activities of DDX49, potentially through post translational modification. Lysine acetylation has previously been reported to play a regulatory role within a number of processes, including splicing, chromatin remodelling and DNA replication: acetylation of DDX3X, for example, is critical in regulating its function in stress response whilst acetylation within the helix-turn-helix motifs of bacterial PhoP inhibited DNA binding (414–416). It has also been suggested that TOPRIM adjacent lysines within topoisomerases influences magnesium binding within the catalytic site (417) .

5.3.5. Characterisation of a novel nuclease site

Nuclease activity was observed within helicase assays and was determined to be acting on the 5' flap of the lagging strand. Literature suggests the recurring contaminant Crp is not noted to have DNA cutting activity. It was initially hypothesised that DDX49 was acting to flap-trim the flayed duplex, however activity recorded on ssDNA and ssRNA suggests this interaction is not flap specific. The protein sequence was analysed and a well-conserved sequence was noted within residues 422-424 (DPD). It was hypothesized, based upon similarities to Toprim (topoisomerase-primase) motifs (DxD), that this was the catalytic site of the nuclease activity (419).

The nuclease activity was next examined on a denaturing gel containing urea and it was seen that the activity appeared to be hyperactive on RNA in the presence of ATP whilst acting independently of ATP on DNA species. Absence of ATP strongly inhibited nuclease activity on ssRNA. TOPRIM domains have previously been reported in DnaG-type primases and type 1A topoisomerases; activity of the latter operates independently of ATP with catalytic activity of the TOPRIM domain dependent upon magnesium ions (420). Additionally RNase M5, involved in maturation of 5.8S rRNA in gram-positive bacteria, possesses a TOPRIM domain critical for its function (421). It is thus hypothesised that the

putative TOPRIM domain of DDX49 may play an essential function within the processing of human rRNA precursors.

Larger bands within the no protein control of figure 5.13B indicates the presence of secondary structure formation: it is hypothesised that ATP might enhance nuclease activity within DDX49 using strand separation to process secondary structures. Alternatively, it is not excluded that this could be activity of an RNase. It is curious to note that nuclease activity was not seen in previous work by Awasthi et al (82). However, it was noted that the substrates used within the Awasthi study were of a shorter length than those used here (Fig. 5.23), with the labelled strand only 10nt in length and smaller than the prominent products seen within nuclease assays (Fig. 5.13). It was also noted that purified DDX49 within the Awasthi study appeared less pure than the protein here, raising a possibility that a contaminant may inhibit DDX49 cleavage activity. Further studies could investigate if the length of nucleic acid impacts nuclease activity.

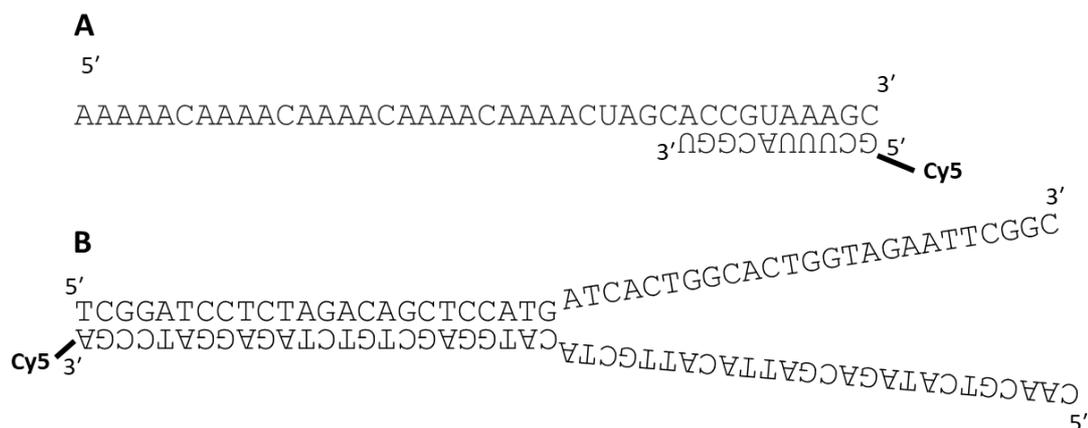


Figure 5.22: A comparison of substrates used to study DDX49. (A) The substrate used in the Awasthi study. **(B)** The substrates used within this study.

5.3.6. *DDX49*^{+/-} cells were successfully generated and showed a phenotype of inhibited migration

The inability to generate a stable homozygous deletion of *DDX49* within this study supports Depmap data for *DDX49*, suggesting that *DDX49* is an essential gene within this cell line amongst numerous others. As such, only a heterozygous cell line was able to be generated. A 19 bp deletion comprising 40% of sequences was recorded: an assumption was made that deviations from

50% were a result of PCR bias, as previously discussed in chapter 4. The polyploidy nature of U2Os cells was considered as a potential impacting factor, but previous research indicates that whilst chromosome 19 possesses structural abnormalities it is unaffected by polyploidy (389, 422). In addition, *DDX49* has been reported to possess a regular copy number within U2OS cells (388). Off-target testing was not conducted by the student and is recommended.

Phenotyping assays were performed by MRes student Sabesan Anandavijayan and no difference in the rate of cell growth was observed. Previous research has shown that *DDX49* knockdowns inhibit cell proliferation, implying heterozygous cells are able to resist his effect and are haplosufficient (84, 85, 406, 407). However, upon reviewing the data the student collected it was noted that an operator error had occurred, hence the use of rate of growth as an alternative. It is recommended that this assay is repeated. Wound healing assays to study cell migration revealed that migration was inhibited within *DDX49^{+/-}* cells. This supports previous research which has shown similar phenotypes of inhibited migration when *DDX49* is knocked down within lung, prostate and cervical cancer cell lines (84, 85, 406, 407). This presents *DDX49* as a potential therapeutic target within cancer metastases, as reported by Lian et al (2020)

5.3.7. The abundance of several rRNA species is altered in *DDX49^{+/-}* cells

DDX49 and its putative yeast counterpart have been implicated within processes of ribosome biogenesis, with Dbp8 associated with the synthesis of 18S rRNA. RT-qPCR analysis was performed to examine the abundance of rRNA species and decreased levels of 28S and 5.8S rRNA were observed within *DDX49^{+/-}* cells, with a 1.7-fold increase in abundance of 47S rRNA. No change was seen in the levels of 18S rRNA. Previous on 28S and 5.8S rRNA have not been reported whilst a lack of change in 18S rRNA suggests that *DDX49* and Dbp8 may be functionally distinct. It is noted that data directly contradicts previous data from Awasthi et al (2018), whereby 47S rRNA levels were reduced upon knockdown of *DDX49* (82). As the previous study used siRNA methods and likely achieved higher levels of knockdown than our heterozygous model, this does not prohibit a hypothesis whereby heterozygous

DDX49 is sufficient to stimulate 47S transcription but insufficient to process 47S rRNA. Deficiencies in processing of 47S rRNA could potentiate downstream decreases in 28S and 5.8S rRNA, seen within this data, though why 18S rRNA levels remain unaffected is unknown. It is recommended that repeat tests are performed to validate results seen here.

5.3.8. CRISPR-Cas9 prime editing of *DDX49*

To trial an alternative method of gene editing, a prime editing reaction was carried out to edit the DPD amino acid sequence within *DDX49*. This was chosen as it was suspected to play a functional role that was potentially non-essential, unlike mutations in conserved motifs. The full prime editing reaction was lethal to cells. This tentatively suggests that the DPD site may be essential: substitutions within the DxD TOPRIM motif of *E. coli* Topoisomerases have previously presented lethal phenotypes (423). It is considered that prime editing protocol is in its infancy and due to lethality an assessment of its editing efficiency could not be made. Controls implied that the effect was not induced by one specific plasmid or transfection condition. Testing the method on a non-essential gene is recommended to validate results seen here.

Further work could examine prime editing of motif II to examine if this also presents a lethal phenotype. Studies in other DExD-box helicases have identified separate phenotypes between unique sequences and the DExD-box domain: studies within *Xenopus* oocytes, for instance, demonstrated that downregulation of DDX21 reduced 28S and 18S rRNA production but the expression of a helicase deficient mutant was sufficient to rescue 28S rRNA production but not 18S rRNA (150). Plasmids to test this were constructed but were not completed in time for testing. Alternatively, an approach using other editing techniques could be trialled, notably base editing.

5.4. Conclusions

In this work we have presented novel biochemical activities of human recombinant DDX49, reporting novel nuclease activity that is hypothesised to be a highly conserved DPD motif within the C-terminus of the protein. Mutants of this motif support this hypothesis and we suggest the site may act functionally similar to TOPRIM domains seen within Topoisomerases and Primases. We

also report canonical helicase activity, including a 3' to 5' polarity of unwinding and ATP-dependent strand separation of flayed duplexes. Mutations (K421A) adjacent to the TOPRIM motif results in hyperactive helicase activity suggesting a mechanism whereby the helicase activity is regulated within the C-terminus.

In addition, we present the generation of a *DDX49*^{+/-} cell line and preliminary phenotyping of the heterozygous cell line using proliferation and migration assays. *DDX49*^{+/-} cells exhibited a phenotype of inhibited migration, agreeing with previous research and highlighting the gene as a potential therapeutic target within metastatic cancers. The abundance of rRNA species is also impacted, though these are contradictory to previous research and require further validation. It is noted that off-target testing and protein expression analysis was not conducted by the student so these results require further validation. Furthermore, issues were noted within how the student performed proliferation assays so these require repeating.

5.5. Future work

The possibility that nuclease activity seen within these assays is the product of a contaminating protein is not excluded, however the activity was recorded across multiple protein batches of wild type and mutant proteins and across multiple operators. In addition, upon assessment via SDS-PAGE the purified protein preparation here possessed less contaminants than those applied within other studies (82). To validate results, it is recommended that purification using a more specific tag is attempted. The initial framework of this study initially used a highly specific FLAG-tag, but practical limitations made purification of this mutant unfeasible. It is recommended that a Strep-tag purification is carried out to achieve higher specificity.

Further mutagenesis of the protein is recommended, particularly the generation of N and C-terminus truncations to examine if the two termini differ in functionality or binding characteristics, notably including truncations that exclude the extended C-terminus (including the conserved DPD site) that is absent within the purported yeast homolog. At the conclusion of the authors study early work had begun on this by MRes student Fiorela Kapllanaj but the generation of results was still in process.

Whilst preliminary phenotyping of the heterozygous cell line was carried out within this study, proliferation assays require repeats and further phenotyping is possible. Based upon potential similarities between the conserved DPD site with TOPRIM motifs in Topoisomerases, toxicity assays using the topoisomerase I inhibitor Camptothecin are of interest. Future work should also further investigate the preliminary results seen from qPCR results and examine if ribosome biogenesis is affected within these cells. As described in chapter 4, northern blot methods trialled were unsuccessful so were not attempted here, but upon development these will be key to investigating the role of DDX49 within ribosome biogenesis.

Chapter 6: Investigating putative Asgard origins of probable ATP-dependent helicase DDX49

6.1 Introduction

6.1.1. MCP8718128.1 (AA.49)

In the course of work conducted on DDX49, BLAST searches for potential homologs within Archaea species identified an open reading frame of interest within the Asgard superphylum. Interest within Asgard has grown since their identification due to their evolutionary affinity with eukaryotes, including identification of proteins previously thought to be eukaryotic signature proteins (ESPs); proteins that define eukaryotes from prokaryotes (274). These include ESCRT subcomplex I and II, GTPase families and SNARE-like proteins (278–280). As a result, the Asgard are considered strong candidates to act as the bridging point between prokaryotic and eukaryotic life: notably the *Heimdallarchaeota–Wukongarchaeota* group, on a basis of protein sequence conservation and the identification of supersized expansion segments within rRNA that were previously believed to be unique to eukaryotic ribosomes (281).

This potential homolog, MCP8718128.1, was identified from metagenome assembled genomes carried out within the yeast *Candida tropicalis* (286). For the purposes of this study, this identified protein of interest will be referred to hereon as AA.49. As a product of a metagenome study, no further data is available regarding the protein. The study by Jagadeeshwari et al (2023) observed that the average nucleotide identities of binned genomes corresponded to the *Heimdallarchaeota* branch of the Asgard superphylum. This was considered of interest as homology between the two proteins could assist within the characterisation of DDX49 through comparative studies, as well as providing further support towards an Asgardian origin of eukaryotic life, with the *Heimdallarchaeota–Wukongarchaeota* group previously implicated as the closest bridging point between prokaryotes and eukaryotes (281).

6.1.2. Yeast Dbp8

Following conclusion of research and publication of the work by Jagadeeshwari et al (2023) that deposited the database entry on this homolog, it was noted that the protein identified possessed a 100% sequence identity for the Dbp8 protein of the host organism *Candida tropicalis* used for their metagenome sequencing (286, 424). It has previously been reported that metagenomic studies are prone to low levels of contamination, with the study by Jagadeeshwari et al (2023) noting a contamination of <5% within the assembled genome: such contamination has previously raised concerns within metagenome studies following identification of anomalous proteins, with a number failing to meet phylogenetic criteria (283, 284, 286).

The purported yeast homolog of mammalian DDX49, Dbp8 has been primarily associated with roles within ribosome biogenesis. Previous research has noted that in vivo depletion of Dbp8 gives rise to 40S ribosomal subunit deficiency as a result of impaired cleavage at sites A₀, A₁ and A₂ in the process of mature 18S rRNA production (90). The overexpression of *Dbp8* dominant negative mutants resulted in accumulation of the 22S pre-rRNA species, similarly associated with defects within A1 and A2 processing (91). Dbp8 shares interactions with Hsp90 and Esf2, with Esf2 acting as a cofactor to stimulate Dbp8 ATPase activity and recruit Dbp8 to targets (92, 93). Novel bioinformatic comparisons of *S. cerevisiae* Dbp8 with DDX49 and AA.49 (putative *C. tropicalis* Dbp8) can be found within section 6.2.1.

6.1.3. Aims and objectives

Due to difficulties cultivating Asgard archaea, previous studies have heavily relied upon data from cultivation-independent metagenomic assemblies with evidence that archaeal genes produce proteins similar to eukaryotes eagerly awaited to support the eocyte hypothesis (282, 285). Following identification of a potential Asgard homolog of DDX49, AA.49, comparisons between the two proteins were of interest to both support this hypothesis and compliment research on DDX49. However, metagenomic assemblies are prone to low levels of contamination with studies finding that many identified proteins fail to meet phylogenetic criteria, raising questions over the validity of identified ESPs (283, 284). As discussed in section 6.1.2, following the conclusion of this study

the publication of the paper by Jagadeeshwari et al (2023) led to considerations that the gene identified may be a contaminant from the host organism: *Candida tropicalis* Dbp8. As of the time of writing the database entry has not been retracted, thus for the purposes of this study the protein will continue to be referred to AA.49 and the research will be presented with a focus on its original aims. However, it is noted that with both proteins sharing 100% identity the results presented here are also applicable to *Candida tropicalis* Dbp8. The aims of this study were as such thus:

- Investigate potential connections between DDX49 and AA.49 using bioinformatics.
- Purify recombinant AA.49 and perform initial characterisation of protein biochemistry.
- Compare the biochemistry of AA.49 and DDX49 to inform activities seen within the study on human DDX49 in chapter 5.

6.2 Results

6.2.1. Bioinformatics

6.2.1.1. DDX49 BLAST results reveal an Asgardian ORF

To identify potential archaea homologs of DDX49, a blastp search was carried out using the human amino acid sequence with criteria set to limit the results to those that were found within the clade Archaea (425). The top two results corresponded to a DEAD-box helicase ORF identified within Asgard group archaea, showing strong E-values and query covers to DDX49 (table 6.1). An RNA helicase from *Euryarchaeota* presented comparable scores, but study focused on the top Asgard result (MCP8718128.1). The accession was investigated and found to originate from metagenomic studies within the yeast *C. tropicalis*. No publication was available at the time of publication. Sequence alignments were carried out between AA.49, DDX49 and the putative DDX49 yeast homolog Dbp8 (Fig. 6.1). The N-terminus and core helicase domains of DDX49, Dbp8 and AA.49 show strong conservation, but poor conservation was observed within the C-termini of AA.49 and DDX49. It was noted that AA.49 and Dbp8 appeared more closely related than AA.49 and DDX49, however a lysine

rich sequence of amino acids was noted within the C-terminus of AA.49 that was absent within Dbp8.

6.2.1.2. Structural comparison between DDX49 and AA.49

To compare predicted structures of DDX49 and AA.49, the structure of AA.49 was predicted using ColabFold, an online version of AlphaFold script, and the two proteins compared using chimera (426). The structural prediction of DDX49 is discussed in section 5.2.1.1, but it is reiterated that amino acids 457-483 were modelled with low confidence by AlphaFold: the confidence levels of AA.49 were not provided by the script. In agreement with sequence alignments in figure 6.1, the modelling of DDX49 vs AA.49 shows high levels of conservation within the N-terminus and RecA1 and RecA2 domains (Fig. 6.2) however DDX49 possesses an extended C-terminus that is absent within AA.49. This matches results recorded within section 5.2.1.1. with this extended C-terminus also absent within Dbp8. Notably, this implies that AA.49 lacks the proposed DDX49 nuclease site.

Table 6.1: Potential Archaeon homologs of DDX49 as noted within NCBI BLASTp database. Criteria for search was set as the human DDX49 sequence and was limited to sequences within the clade 'Archaea'. Only the top ten results have been included (425).

Description	Scientific Name	Max Score	Query Cover	E value	Accession
DEAD/DEAH box helicase	Asgard group archaeon	368	85%	2.00E-121	MCP8718128.1
DEAD/DEAH box helicase	Asgard group archaeon	352	85%	1.00E-114	MCP8715898.1
RNA helicase	Euryarchaeota archaeon	346	85%	3.00E-113	MBI20465.1
DEAD/DEAH box helicase	Candidatus Nitrosocosmicus sp. SS	276	78%	1.00E-85	WP_149861599.1
DEAD/DEAH box helicase	Candidatus Nitrosocosmicus hydrocola	275	78%	3.00E-84	WP_148685637.1
TPA: DEAD/DEAH box helicase	Candidatus Altiarchaeales archaeon	267	77%	4.00E-83	HIE33951.1
RNA helicase	Candidatus Nitrosocosmicus sp.	271	78%	4.00E-83	NOJ30821.1
DEAD/DEAH box helicase	Candidatus Nitrosocosmicus franklandus	271	78%	4.00E-83	WP_134483650.1
DEAD/DEAH box helicase	Candidatus Thermoplasma tota archaeon	267	78%	1.00E-82	MCJ2514029.1
DEAD/DEAH box helicase	Candidatus Nitrosocosmicus oleophilus	267	78%	4.00E-82	WP_196817237.1

MCP8718128.1	110.....20.....30.....40.....50.....60
sp P38719 DBP8_YEAST	1	MS-FNELGAKWLEALDAMKIYKPTSFOSACIEQITLKGHDICGGAKTGSCKTIAPA PM
sp Q9Y6V7 DDX49_HUMAN	1	MADFKSLGLSKWLESLRAMKITQPTATQKACIEKILEGRDCIGGAKTGSCKTIAPA PM
MCP8718128.1	6170.....80.....90.....100.....110.....120
sp P38719 DBP8_YEAST	61	LTKWSEDPYGFGLVLTPTRELAQIABQFTALGSMNINIKVSVIVGGEDIVKQALELQ K
sp Q9Y6V7 DDX49_HUMAN	61	LCKLSEDPYGFCLVLTPTRELAQIABQFRVLGKPGKDCIVVGGEDIVKQALELSRK
MCP8718128.1	121130.....140.....150.....160.....170.....180
sp P38719 DBP8_YEAST	121	PHFVIATPGRLAHDHNSGSETSGLRKRYLVLDEADRLLSNS---FSDLKRCFDVLP
sp Q9Y6V7 DDX49_HUMAN	121	PHFVIATPGRLAHDHRSNTEFS---IKKIRFLVDEADRLLGCTDFTVDEAI AAP
MCP8718128.1	181190.....200.....210.....220.....230.....240
sp P38719 DBP8_YEAST	177	PSKRQTLFLFTATITDAVRALEKKEPQTEGKPPFMHEVEVDRVAIPSTISISYFVPSY
sp Q9Y6V7 DDX49_HUMAN	178	PKKRQTLFLFTATITDQVSLQNAFVKGKPPFAYQVEVDNVAIPSTIKTBYLVPEH
MCP8718128.1	241250.....260.....270.....280.....290.....300
sp P38719 DBP8_YEAST	237	VKEAYDNSL---LKKKDATAFVNRTTAE LRRTLKLLFRVASLSEMPQAEKRN
sp Q9Y6V7 DDX49_HUMAN	238	VKEAYLYQDLT---CEENKFAIFVNRTTAE LRRTLQLFRVASLSEMPQAEKRN
MCP8718128.1	301310.....320.....330.....340.....350.....360
sp P38719 DBP8_YEAST	295	SLHRFKACAAARLIATDVASRGLDIPFVQVVDIPDPDFIHRVGRPARAGRQDAI
sp Q9Y6V7 DDX49_HUMAN	296	SLHRFANAARLIATDVASRGLDIPFVQVVDIPSDPDFIHRSGRTPARAGRQDAI
MCP8718128.1	361370.....380.....390.....400.....410.....420
sp P38719 DBP8_YEAST	355	SIGEDIDRIQSIBERINKKMELEDVDDKVI DSNVVTAAKRESMEMKENFGER
sp Q9Y6V7 DDX49_HUMAN	356	SPVTD DSRIOAIE RINKKMTETNKVHDPAVIKATTKVTKAKRESMAKQENFGER
MCP8718128.1	421430.....440.....450.....460.....470.....480
sp P38719 DBP8_YEAST	415	KIN KKHGLLEKKNES-----K
sp Q9Y6V7 DDX49_HUMAN	416	KQKKQ-N GKS-----L
MCP8718128.1	481490.....
sp P38719 DBP8_YEAST	434	K--KQKKE-----
sp Q9Y6V7 DDX49_HUMAN	430	R-----S-----
	467	RTPSGSHSGPVPSQGLV

Figure 6.1: Sequence alignment of the ORF for MCP8718128.1 (AA.49) against *S. cerevisiae* Dpb8 in yeast and *H. sapiens* DDX49. Sequence alignments were carried out using T-coffee server and comparisons highlighted using boxshade software.

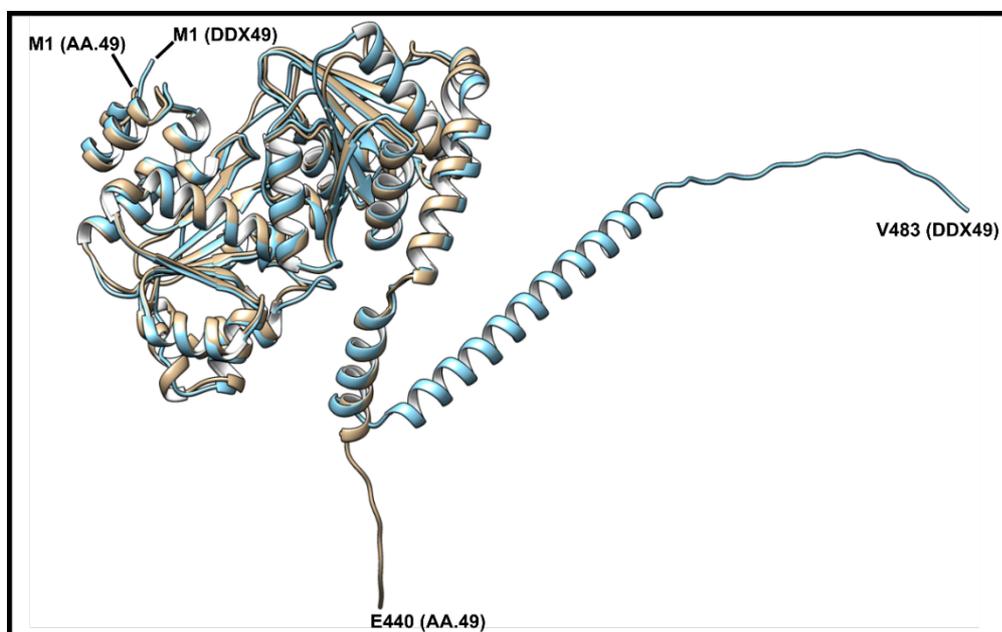


Figure 6.2: Comparisons of AA.49 and DDX49 C-terminus reveals low levels of conservation. AA.49 is represented in cream, DDX49 in light blue. DDX49 structure was taken from alphafold server, AA.49 structure determined using ColabFold. N and C-terminus residues are indicated (319, 426).

6.2.2. Purification of recombinant AA.49

To purify AA.49, an open reading frame encoding the amino acid sequence was designed and codon-optimised for *E. coli* expression using Thermofisher GeneArt and implemented into a pET100-D expression vector. The relevant plasmids (table 2.4) were transformed and expressed within *E. coli* BL21 AI cells. As this was a comparative study, no mutants were generated or purified. A small-scale pilot study confirmed successful overexpression of AA.49 (Fig. 6.3A). AA49 included an N-terminus his-tag and, following overexpression and cell lysis, was successfully purified using a HisTrap HP His-tag column (Fig. 6.3B). To further remove contaminants and separate DNA from pooled fractions, AA49 was dialysed into a low salt containing buffer (Heparin buffer A) and successfully eluted from a HiTrap™ Heparin HP column (Fig. 6.3C), prior to dialysis into a high glycerol containing storage buffer and storage at -80 °C. Intact purified protein was confirmed via additional SDS-PAGE (Fig. 6.4D)

6.2.3. Assessing the activity of wild type AA49

6.2.3.1. AA.49 and DDX49 do not share binding characteristics

The binding affinity of AA.49 to several nucleic acid species as a function of protein concentration was assessed through electrophoretic mobility shift assays (EMSAs) as described in section 2.6.5. Initial studies used a 45-base pair Cy5 labelled ssDNA oligonucleotide (AP122), a 45-base pair Cy5 labelled ssRNA substrate (AP33) and a 50 base pair fork duplex (fork 2B). Results demonstrated that AA49 was unable to form a stable complex with any nucleic acid species, showing a smearing pattern and concentration-dependent aggregation within the wells of the gel (Fig. 6.4). This contrasts with results observed with wild type DDX49 in section 5.2.4.1, where poor but stable binding patterns were noted. In addition, AA.49 appears to bind nucleic acid species with higher affinity than DDX49 with 1000 nM sufficient to completely supershift nucleic acid species.

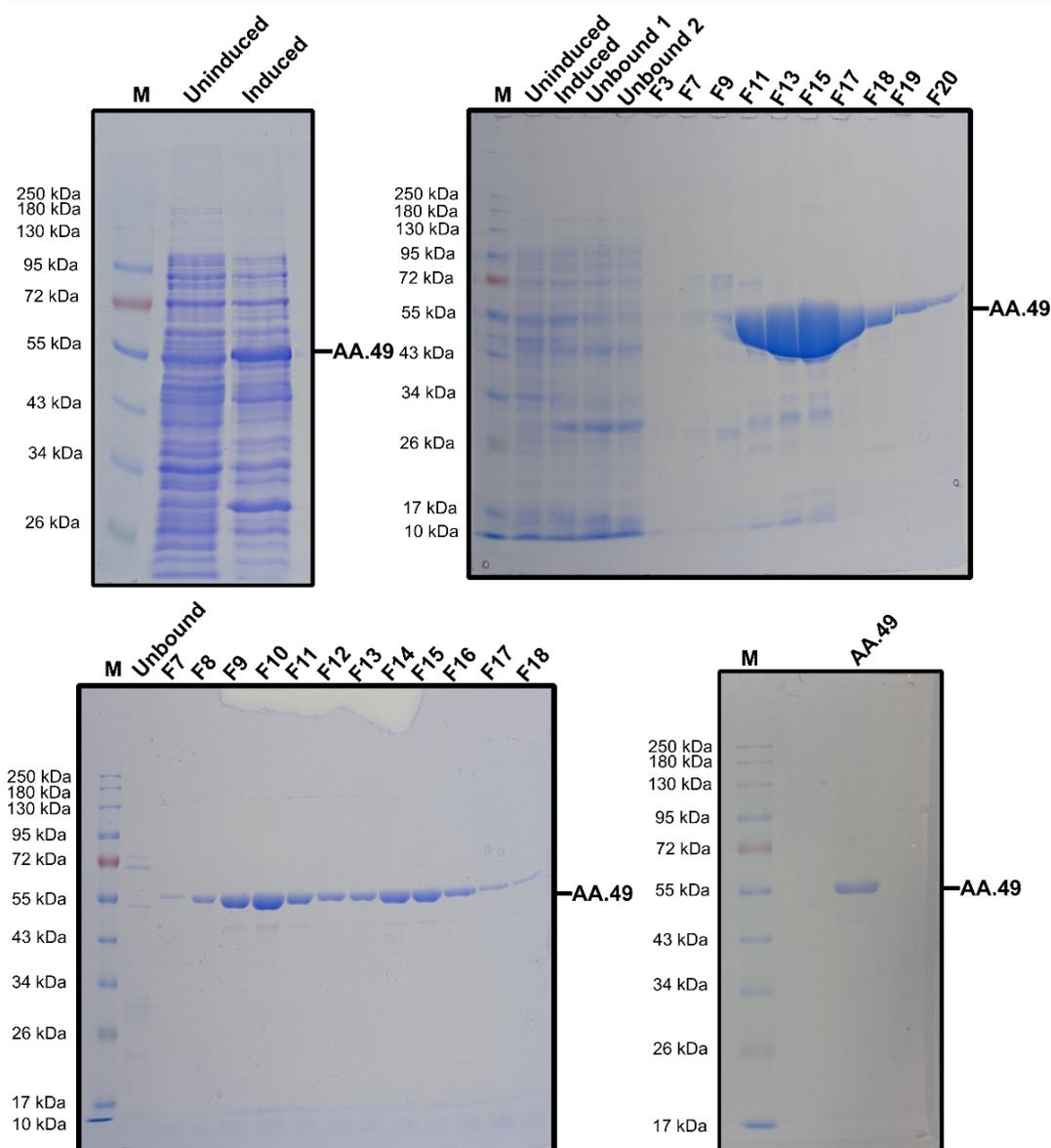


Figure 6.3: Recombinant AA.49 was successfully purified from *E. coli* AA.49. AA49 was overexpressed in *E. coli* strain BL21-AI and methods carried out as described in section 2.4. Molecular weight of AA.49 is 55 kDa. **(A)** Protein overexpression was confirmed in a scaled down pilot overexpression, comparing lysates of cells transformed with pAP11 that were induced and uninduced. **(B)** Upscaled lysate containing overexpressed AA49 was purified using a HisTrap HP His-tag column. Fractions positive for AA.49 (F9 – F20) were pooled and dialysed into a low-salt buffer overnight. **(C)** Dialysed sample was purified using a HiTrap™ Heparin HP column and protein containing fractions pooled and dialysed into protein storage buffer overnight. **(D)** Purified protein was run on a 10% SDS-PAGE gel to assess purity and stored at -80 °C for downstream use.

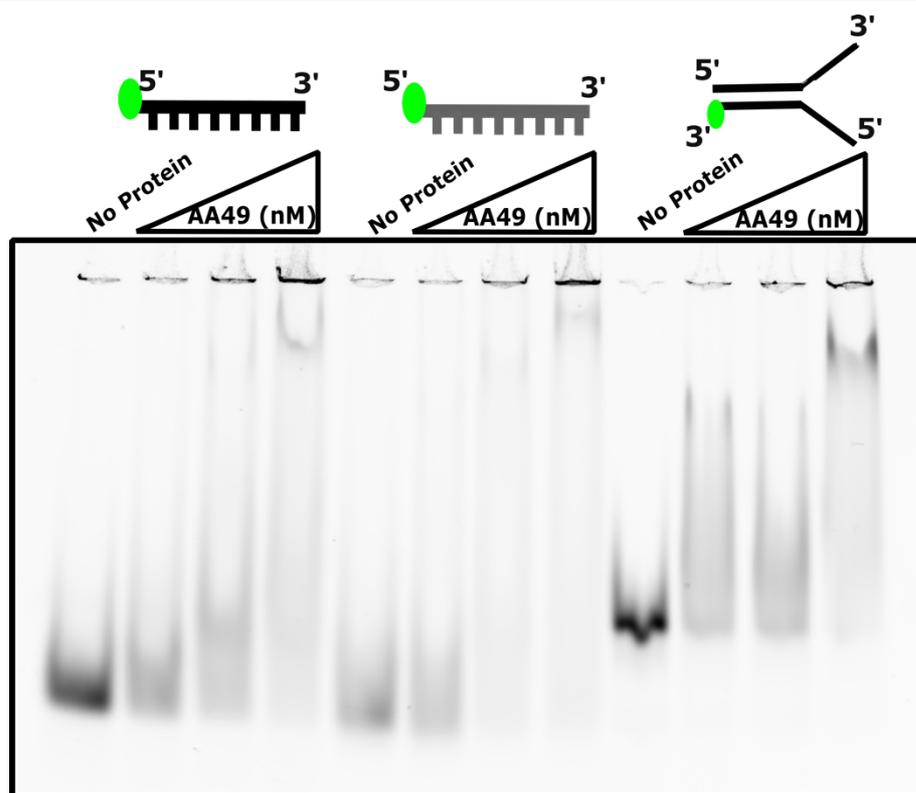


Figure 6.4: Electromobility shift assay for binding affinity of AA49 against nucleic acids. Protein was incubated with (from l-r) Cy5 labelled ssDNA, ssRNA and a flayed DNA duplex at 25 nM for thirty minutes in 25 mM DTT and 1 x Helicase buffer at prior to running on a 5% native gel. Concentration range included 200, 500 and 1000 nM protein.

6.2.3.2. AA.49 is thermostable but inhibited by high NaCl concentrations

AA.49 was initially trialled in conditions similar to those used in assays on DDX49 but poor activities were seen. As the activity at 37 °C was relatively poor and archaea species show a diverse range of native temperatures, an assay was performed to examine if incubation at other temperatures improved activity (Fig 6.5). It was observed that unwinding increased at higher temperatures, appearing to be most efficient at temperatures of 47 °C, reaching a percentage unwound substrate of ~80%. Following this unwinding appeared to plateau. It was not possible to test higher temperatures as no protein controls within earlier assays demonstrated that the flayed duplex DNA substrate became unstable at these temperatures.

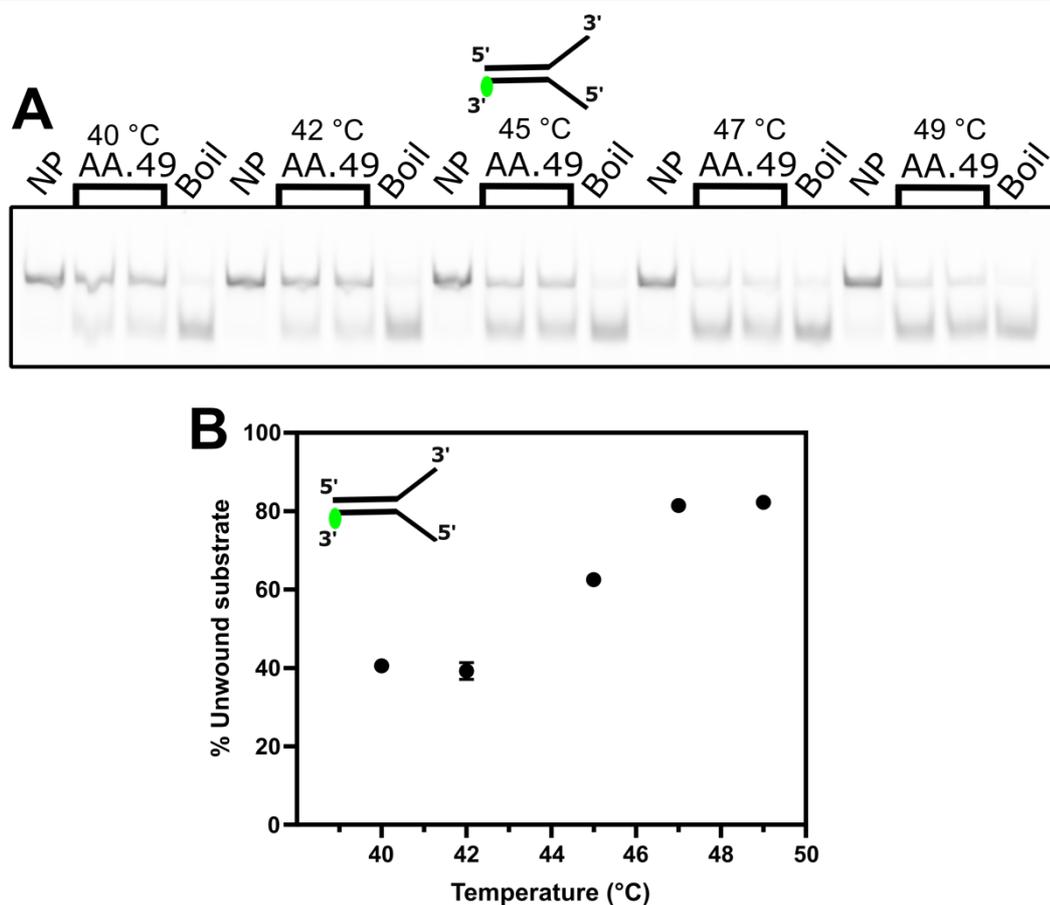


Figure 6.5: Examining the effect of temperature on AA49 unwinding of a flayed DNA duplex. (A) Unwinding was analysed by testing five different temperatures in duplicate with 25 nM of a flayed duplex DNA substrate supplemented with 5 mM ATP, 5 mM MgCl₂, 25 mM DTT and 250 nM ‘trap’ DNA in 1 x helicase buffer. Included as controls were a substrate only reaction and a boiled substrate only reaction to indicate full dissociation. (B) Graphical representation of the bands seen in figure 6.5 A. Bands were quantified using imageJ software and normalised based upon controls.

As some archaea species are halophilic, the effect of salt concentration on the unwinding activity of AA.49 was examined through use of a gradient of increasing sodium chloride concentration (Fig. 6.6). It was observed that unwinding was inhibited as the concentration of sodium chloride increased, with unwinding appearing completely abolished at 60 mM NaCl. The highest unwinding was seen at concentrations of 0 and 20 mM, though it was noted that the duplicates for 0 mM showed poor consistency.

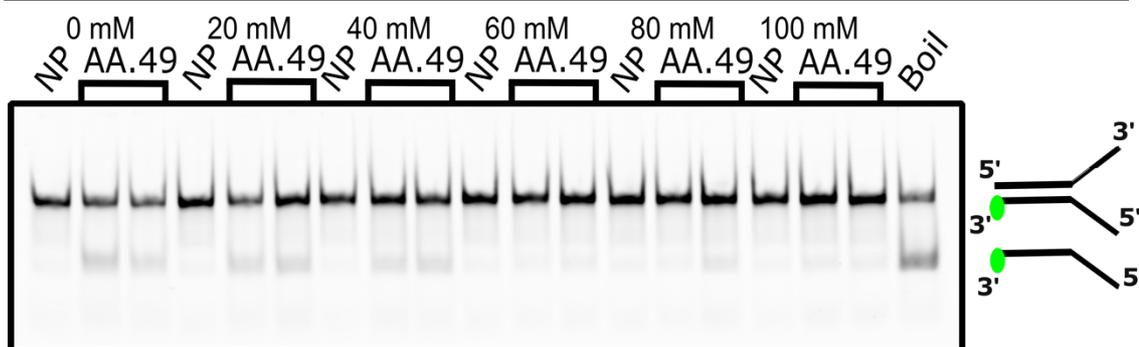


Figure 6.6: Examining the effect of salt concentration on AA.49 unwinding of a flayed DNA duplex. Unwinding was analysed by testing six different concentrations of sodium chloride in duplicate with 25 nM of a flayed duplex DNA substrate supplemented with 5 mM ATP, 5 mM MgCl₂ and 25 mM DTT in 1 x helicase buffer. Included as controls were substrate only reactions and a boiled substrate only reaction to indicate full dissociation.

6.2.3.3. AA.49 is an ATP-independent helicase

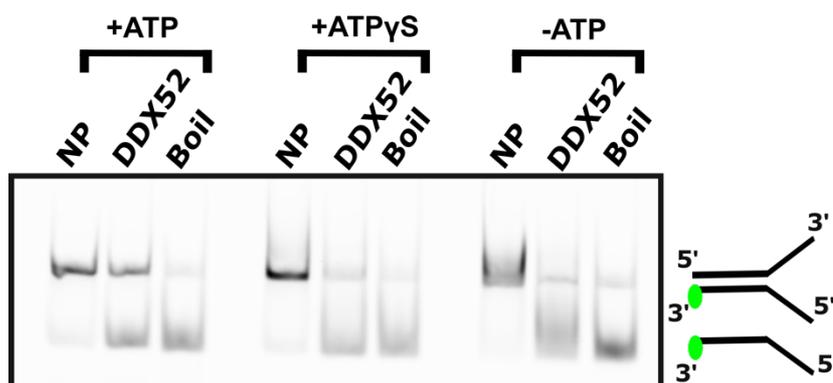


Figure 6.7: AA.49 unwinds independently of ATP. Unwinding was analysed by testing AA.49 (500 nM) in reaction pools containing 25 nM of a flayed duplex DNA substrate and ATP, ATPγS and no ATP supplemented with 5 mM MgCl₂ and 25 mM DTT in 1 x helicase buffer. Included as controls were substrate only reactions and a boiled substrate only reaction to indicate full dissociation.

To examine if AA.49 was an ATP-dependent helicase, a series of reactions were set up containing ATP, the slow hydrolysing ATP analog ATPγS and no ATP (Fig.6.7). Results indicated that AA.49 unwound DNA in all reactions, suggesting that it operates as an ATP-independent helicase. No clear differences within the proportion of unwound DNA was seen, however it was noted that in the absence of ATP a smearing pattern was observed.

6.3. Discussion

6.3.1. Blast searches of DDX49 identified a potential Archaea homolog

To explore potential archaea homologs, DDX49 was placed into a BLAST search with results filtered to Archaea. Several high-ranking results were noted (table 6.1) with the closest being MCP8718128.1 (AA.49), a protein ORF noted within a recent metagenomic study by Jagadeeshwari et al (2023) and reportedly corresponding to a protein with the Asgard clade of archaea (286). It was of interest that both Asgard results were closer in identity than the closest *Euryarchaeota* result: recent theories have posited that the eukaryotes arise from within the Asgard clade of archaea, with others suggesting that a *Euryarchaeota* origin (271, 427). If AA.49 and DDX49 are related, this supports an Asgard model of eukaryogenesis.

6.3.2. Bioinformatics suggests AA.49 shares closer similarities to *S. cerevisiae* Dbp8 than hDDX49

Bioinformatic analysis was carried out and it was noted that whilst the N-terminus and core helicase domains were well-conserved across AA.49 and DDX49, the extended C-terminus seen within DDX49 was absent with AA.49. This is similar to results seen in comparisons between DDX49 and the yeast protein Dpb8 (section 5.2.1.1). Peptide sequence alignments support similarities between Dbp8 and AA.49. It is hypothesised that AA.49 is more likely to be a homolog of Dbp8 but this hypothesis does not preclude that *H. sapiens* DDX49 arose from mutational events within the ORF of AA.49 throughout evolutionary history. Alternatively, following publication of the study associated with the metagenome study that identified AA.49 an alternative hypothesis was formed – this alternative hypothesis is discussed in section 6.4.

6.3.3. AA49 only partially shares the same biochemical activity as DDX49

The biochemistry of AA.49 was investigated and initial EMSAs (Fig. 6.4) suggest that DDX49 and AA.49 exhibited different binding activities, with a binding pattern more closely related to the smearing and aggregation seen

within DDX52 (see section 3.2.4) than stable binding patterns seen within wild type DDX49. As reported within that section, this binding pattern is consistent with the presence of multiple nucleoprotein species, often seen within proteins associated with chromatin remodelling (329, 330). Helicase assays further highlighted differences, with AA.49 showing less proficient activity at 37 °C than assays using DDX49. In addition, no nuclease activity was observed and unlike DDX49 the protein was able to proficiently unwind DNA in the absence of ATP.

It is hypothesised that increased thermostability of AA.49 is contributed by its compact nature, with the protein lacking the substantial C-terminus extension of human DDX49 (428). However, contributions of compactness to protein thermostability are debated (429). It is interesting to note that whilst AA.49 unwound at an optimum temperature of 47 °C, *S. cerevisiae* Dbp8 has previously been reported to have an optimum temperature of 30 °C (92). Salt gradients demonstrated that AA.49 activity was inhibited at higher salt concentrations, suggesting that the enzyme is unlikely to have originated from a halophilic organism (430). Similar to findings regarding temperature, this contrasts with yeast Dbp8 where optimum activity required 300 mM salt concentration: though it is noted that both temperature and salt assessments of activity examined ATPase activity opposed to the visible unwinding used here (92). Regarding the alternative hypothesis proposed in section 6.4, no data is available for Dbp8 within *C. tropicalis* or any species of *Candida*.

ATP-independent unwinding seen within AA.49 is interesting as this conflicts with common properties of DExD-box helicases, which canonically require the hydrolysis of ATP for recycling of protein and efficient unwinding. DExD-box proteins have commonly been reported to exhibit ATP-independent activities, but these are generally associated with ancillary functions (431). It would be interesting to perform a concentration gradient in samples with and without ATP to examine if this is induced by concentration saturation or another mechanism, though ATP-independent duplex unwinding has previously also been reported in the DExD-box proteins CYT-19 and Mss116p (432, 433).

6.4. Conclusion and alternative hypothesis

AA.49 was chosen for this short study after similarities with DDX49 were noted within BLAST analysis. As a result of similarities between AA.49 and *S. cerevisiae* Dbp8 noted within bioinformatics, and functional differences between AA.49 and DDX49, the data here would suggest that AA.49 is more likely an Asgard homolog of *S. cerevisiae* Dbp8 than *H. sapiens* DDX49. This is supported by differences in binding and unwinding characteristics, with a lack of nuclease activity in AA.49 corresponding with a truncated C-terminus in both AA.49 and Dbp8. This is notable as this means both proteins lack the putative nuclease site identified within our study of DDX49 (chapter 5). However, an alternative hypothesis is proposed by the author.

At the time of this study there was insufficient data to fully investigate deposited metagenome data and the study began and concluded prior to the publication of the research that identified the ORF. In the publication, the authors reported <5% contamination within the genome (286). It has been reported previously that metagenomic sequencing studies are prone to contamination from the host organism, raising concerns that some proteins identified within these studies are anomalous (284). Bioinformatics within this study primary focused on comparisons with *S. cerevisiae* and precluded investigations of other yeast strains. Following publication of the study the protein sequence of Dbp8 from the organism used to host the metagenome, *Candida tropicalis*, was compared with the sequence of AA.49 and both sequences were identical (Fig. 6.8). It is noted that *C. tropicalis* Dpb8 is still classed as a hypothetical protein and that the database entry for MCP8718128.1 (AA.49) has yet to be retracted.

CLUSTAL O(1.2.4) multiple sequence alignment		
AA. 49	MSFNELGVAKWLSSEALDAMKIYKPTSIQSACIPQILKGHDCIGGAKTGSCKTIAFAAPML	60
Candida	MSFNELGVAKWLSSEALDAMKIYKPTSIQSACIPQILKGHDCIGGAKTGSCKTIAFAAPML *****	60
AA. 49	TQWSEDPYGVFGLVLTPTRELALQIAEQFTALGSNMNIKVSIVGGEDIVKQALELQKKP	120
Candida	TQWSEDPYGVFGLVLTPTRELALQIAEQFTALGSNMNIKVSIVGGEDIVKQALELQKKP *****	120
AA. 49	HFVIATPGRADHILNSGEETISGLRRVKYLVLDEADRLLSNSFGSDLKRCFDVLPPEK	180
Candida	HFVIATPGRADHILNSGEETISGLRRVKYLVLDEADRLLSNSFGSDLKRCFDVLPPEK *****	180
AA. 49	RQTLFTATITDAVRALKEKPQTEGKPPVFMHEVETVDKVAIPSTLSISYVFVPSYVKEA	240
Candida	RQTLFTATITDAVRALKEKPQTEGKPPVFMHEVETVDKVAIPSTLSISYVFVPSYVKEA *****	240
AA. 49	YLNSILNLEKFKDATAVVFVNRTITAEVLRRLRKLDFRVALHSEMPQAERTNSLHRFK	300
Candida	YLNSILNLEKFKDATAVVFVNRTITAEVLRRLRKLDFRVALHSEMPQAERTNSLHRFK *****	300
AA. 49	AGAARILIATDVASRGLDIPTVELVNFIDIPADPDFIHRVGR TARAGRKGDAVSIIEK	360
Candida	AGAARILIATDVASRGLDIPTVELVNFIDIPADPDFIHRVGR TARAGRKGDAVSIIEK *****	360
AA. 49	DIDRIQSIEERINKMELLEDDVDDKVIKDSLNVTTAAKRESMMEMDKENFGERKKINRK	420
Candida	DIDRIQSIEERINKMELLEDDVDDKVIKDSLNVTTAAKRESMMEMDKENFGERKKINRK *****	420
AA. 49	KHGLELEKNKPSKKKQKKE	440
Candida	KHGLELEKNKPSKKKQKKE *****	440

Figure 6.8: Sequence alignment of AA.49 and the hypothetical protein Dbp from *Candida tropicalis*. Alignments were carried out using CLUSTAL Omega (322).

This suggests that the protein tested within this study is likely the ORF of a contaminating protein from the host organism as opposed to a true Asgard protein. However, previous literature does suggest that conflicts exist between Dbp8 and AA.49 – notably that ATP hydrolysis within Dbp8 is optimum at 30 °C and 300 mM KCl (92). As this study was not focused on Dbp8 this could not be verified in comparative assays. Furthermore, this study did not examine ATP hydrolysis rates and the data reported for Dbp8 is for *S. cerevisiae*, not *C. tropicalis*. It is concluded therefore that this study highlights one of the difficulties associated with investigating Asgard proteins and, whilst irrelevant to the initial question, as *C. tropicalis* Dbp8 has not previously been studied this represents novel research on this protein.

6.5. Future perspectives

Based upon the conclusions made here, it is not considered worthwhile pursuing any further work on the Asgard hypothesis unless the validity of the proteins archaeal origin is confirmed. However, the data presented and purified protein can be used further to investigate *C. tropicalis* Dbp8 and make further comparisons with both *S. cerevisiae* Dbp8 and hDDX49.

Chapter 7: Conclusions and future perspectives

7.1. Summary of research

DDX49 and DDX52 are two DExD-box helicases that remain poorly characterised regarding their functional activities within the cell. Previous research has closely linked them with their putative yeast homologs, Dbp8 and Rok1 respectively, implying roles with the processing and export of rRNA and mRNA molecules (90, 112). Studies have also implicated both within a number of diseases, predominantly in associations with several cancers (84, 97, 98, 406). Despite this, the biochemistry and function of both *H. sapiens* DDX49 and DDX52 remain overlooked, representing a gap in research, and their association with several subsets of cancer make them attractive as therapeutic biomarkers and targets. In addition, a range of DExD-box helicases participate in additional roles outside their canonical roles as RNA helicases (40). This work aimed to explore the biochemistry and function of both proteins, using a combination of biochemical assays and cell-line models, with a particular focus on their interactions with DNA species and potential associations with species seen within DNA replication and repair. In addition, following identification of a putative Asgard homolog of DDX49, the scope of this study was extended to investigate if this putative Asgard homolog and DDX49 were evolutionarily and functionally related.

Overall, the aims of this research project were:

- To biochemically characterise the activities of both DDX49 and DDX52 on DNA and (where possible) RNA species and identify any potential DNA replication and/or repair activities that it might function within.
- Construct a CRISPR-Cas9 gene-editing workflow for disruption of *DDX49* and *DDX52* genes in human cell lines and investigate phenotypes, implementing new assays where possible.
- Investigate potential links between DDX49 and an Asgard archaea homolog discovered during bioinformatic studies.

7.1.1. DDX52 is a 3' to 5' ATP-dependent helicase with novel annealase activity

Previous research posited similarities between the *M. smegmatis* helicase Lhr and *H. sapiens* DDX52 (104). Bioinformatic comparisons between the two concluded that whilst they share conserved core helicase domains, key differences mean that closer associations were dismissed. Further comparisons between DDX52 and its putative yeast homolog Rok1 identified strong conservation across both proteins, however DDX52 possessed a substantially extended IDR loop within the N-terminus in comparison to *S. cerevisiae* Rok1. Binding models of DDX52 demonstrated the apparent formation of multiple nucleoprotein complexes and banding patterns appeared similar to those previously seen within chromatin remodelling complexes (338). Fluorescent anisotropy assays revealed an apparent affinity for DNA, challenging the preconception that DDX52 is primarily an RNA helicase.

As a member of the DExD-box helicase family of proteins, it was hypothesised that DDX52 would show activity as an ATP-dependent helicase with a 3' to 5' polarity of unwinding. This was confirmed within assays on duplexes containing 3' and 5' flaps and previously unreported activity on DNA replication forks was revealed. Substitutions of wild type DDX52 for motif II mutants (D318A, D321A) and assays using a slow hydrolysing ATP analog, ATP γ S, further demonstrated that DDX52 exhibited an ATP dependent mechanism of unwinding. In assays using DNA:RNA hybrid duplexes, no clear difference in unwinding was seen further supporting the hypothesis that DDX52 does not preferentially unwind RNA.

Interestingly, this research demonstrated a novel annealing activity for DDX52. Annealing has previously been reported within the putative homolog Rok1, however this was in conjunction with cofactors (Rrp5) and only RNA:RNA annealing activity was reported (114). It was hypothesized that annealing activity was localised to the IDR loop located within the N-terminus of DDX52 and the generation of N-terminus and C-terminus truncations confirmed this hypothesis. Furthermore, deletions of the IDR loop abolished annealing though it was noted that this mutant was weakly active in general. It would be of interest to further investigate this region of the protein to identify strongly conserved

residues within homologs and perform substitutions within these residues, in addition to exploring cofactors that may stimulate either helicase or annealase activities of the protein including homologs of *S. cerevisiae* Rrp5.

7.1.2. Potential roles for DDX52 in cancer progression and DNA replication/repair

Previous research has associated DDX52 with several subsets of cancers, notably melanoma and prostate cancers, with studies demonstrating that knockdowns of *DDX52* inhibited cell proliferation (97, 406). A heterozygous cell line was successfully generated within this study and confirmed proliferation phenotypes seen within these studies. In addition, migration assays imply inhibited migratory phenotype within *DDX52* heterozygotes – consistent with *DDX52*'s role in metastatic melanoma and raising it as a potential biomarker and therapeutic target within metastatic cancers. Viability assays suggest increased susceptibility of *DDX52*^{+/-} to the cross-linking agent Mitomycin C, but dataset is insufficient to fully support this. Further studies should validate this data and examine further DNA damaging agents, such as hydrogen peroxide.

Biochemical assays on recombinant DDX52 revealed that DDX52 was able to dissociate replication forks possessing 3' flaps as well as both D-loops and R-loops, with an apparent preference for D-loops. In addition, nuclease protection assays demonstrated that DDX52 was able to act in a protective capacity on both DNA and RNA nucleic acid species, like the DNA replication and repair enzymes Replication protein A and Rad52 recombinase (333, 336). Fluorescent polarisation and ATPase assays suggesting a preference of DDX52 for DNA may transiently support this hypothesis. It would be interesting to compare R-loops and D-loops within the cell model here, such as detection using S9.6 antibody. Further work should also incorporate methods of chromatin immunoprecipitation to support functional hypotheses seen from binding assays.

7.1.3. Abundance of rRNA in *DDX52*^{+/-} cells suggests yeast predictions of DDX52 function are inappropriate

Attempts to incorporate a non-radiation fluorescent-based northern blotting protocol were unsuccessful. An RT-qPCR method was explored instead,

previously used effectively on other DExD boxes to study 47S rRNA levels (82). Reduced levels of 47S rRNA was seen within *DDX52*^{+/-}, supporting previous data from zebrafish models suggesting DDX52 is involved within the 47S synthesis (357). Decreased levels of 28S rRNA were also seen, potentially as a downstream effect of inhibited 47S synthesis. Interestingly no change within the abundance of 18S rRNA was seen with studies suggesting that the putative *S. cerevisiae* homolog, Rok1, processes 18S rRNA (108, 109). This potentially represents contradiction between human and yeast models, discussed in section 1.3.3.1. with Zebrafish DDX52 sharing closer homology, including a similar IDR loop, to *H. sapiens* DDX52 than *S. cerevisiae* Rok1. Further studies should attempt a radiation-based northern blotting protocol to further validate results as well as further RT-qPCR analysis combined with CRISPRi to disrupt DDX52 expression.

7.1.3. DDX49 is a 3' to 5' ATP-dependent helicase with novel ATP-independent nuclease activity

As a member of the DExD-box helicase family of proteins, it was predicted that DDX49 would show activity as an ATP-dependent helicase with a 3' to 5' polarity of unwinding. This was confirmed within assays on duplexes containing 3' and 5' flaps, with the helicase and nucleic acid binding assays confirming that DDX49 possessed a weak affinity for nucleic acid species tested within the framework of this study. However, during these assays a novel nuclease activity was identified and further investigation revealed that DDX49 appeared to be cutting DNA on the lagging strand.

Bioinformatic analysis hypothesised that this nuclease activity was localised to a highly conserved TOPRIM-like DPD amino acid sequence within the C-terminus of DDX49 (419). A D422A/D424A mutant was generated within this site and nuclease activity was abolished, supporting this hypothesis. This has previously been unreported within previous studies of DDX49, but it was considered that this may have arisen due to differences in substrates between the two studies (82). Denaturing gels revealed that cleavage of DNA and DNA duplexes occurred independently of ATP, but that RNA nuclease activity was strongly stimulated with the addition of ATP. ATP-independent cleavage by TOPRIM domains resembles activity seen within type 1A topoisomerase

enzymes (420). A mechanism whereby ATP-dependent helicase activity assists in processing of RNA secondary structures to allow cleavage is proposed but requires further study.

Interestingly, a purified K421A mutant showed hyperactive helicase activity in comparison to the poor activity of wild type DDX49, suggesting a mechanism whereby this lysine residue regulates helicase activity of the protein. It is hypothesised that this is potentially a result of lysine impacting magnesium binding within the catalytic site or post-translational modification similar to regulatory activities seen within other proteins (375–378). The purification preparation by the student carrying out this research was observed to possess several contaminants in comparison to other protein preparations and it is suggested that research is validated through the preparation of a purer batch.

7.1.4. Roles of DDX49 within cancer and ribosome biogenesis

Previous research has associated DDX49 with several subsets of cancers, notably lung cancers, with studies demonstrating that knockdowns of *DDX49* inhibited cell proliferation and migration (84, 85). A *DDX49* heterozygous cell line was successfully generated and migratory phenotypes in previous studies replicated results seen in previous research and highlighting it as a potential therapeutic marker within cancer. Proliferation studies were inconclusive due to issues noted within the data and require repeating. As a result of highly positive prognoses seen in cervical cancers where DDX49 was overexpressed, it is suggested that further research explore *DDX49* phenotypes within these cell lines: with DDX49 associated with viral immunity, comparisons between HPV-positive and HPV-negative cervical cell lines would be of great interest (83).

RT-qPCR analysis of *DDX49*^{+/-} cells revealed an increase in abundance of 47S rRNA and decreased abundance of 28S and 5.8S rRNA species. This conflicts with previous research whereby siRNA knockdowns of *DDX49* showed reductions in 47S rRNA (82). However, differences within methodologies suggest that previous studies achieved a more complete knockdown than our heterozygous model and it is considered that *DDX49*^{+/-} may be sufficient to promote transcription of 47S rRNA but still show deficiencies in rRNA processing. The abundance of 18S rRNA was unaffected, conflicting with

results observed within the putative yeast homolog Dbp8 – such differences could arise from areas of poor homology between the two proteins, with DDX49 possessing a substantially extended C-terminus in comparison to *S. cerevisiae* Dbp8. As a result of the contradictory results seen within comparisons to previous research and the infancy of our developed method it is strongly advised that results seen within this study are validated through additional testing, including the use of northern blot analysis to support the findings here.

7.1.5. Links between DDX49 and a putative Asgard homolog are redundant

During research on DDX49, a putative Asgard homolog was identified through BLAST searches. At the time of this small-scale study the research that identified this putative homolog had not been published and upon publication and review of this research, it was identified that the Asgard homolog in question was likely the result of misclassification of contaminating proteins from the host organism *Candida tropicalis* used within the metagenomic study (286). Biochemistry supported this, with 'AA.49' possessing little similarity to the biochemistry of *H. sapiens* DDX49.

Nevertheless, this study presents novel data on *C. tropicalis* Dbp8 and bioinformatic analysis reveals strong homologies between *C. tropicalis* and *S. cerevisiae* Dbp8. In addition, this study presents novel activity of yeast Dbp8 on flayed DNA duplexes as well as an ATP-independent mechanism of unwinding DNA duplexes, however this needs verification using lower concentrations of AA.49. It is interesting that temperature and salt concentrations differ from those seen in *S. cerevisiae* Dbp8, though it is unclear if these are related to experimental differences, with the activity of *S. cerevisiae* Dbp8 studied through analysis of ATPase activity and the salt KCl opposed to our studies NaCl measurement: KCl has previously been shown to be less inhibitory to both organisms than NaCl (434). Thus, whilst redundant to the original objective, this study has identified previously unreported activities of yeast Dbp8 on as well as an interesting role as an ATP-independent helicase.

7.2. Thesis summary

This thesis has explored the biochemistry and potential functions of two DExD-box helicases, DDX49 and DDX52, along with dismissing an incorrectly classified Asgard homolog of DDX49 but gaining novel insights into the biochemistry of *C. tropicalis* Dbp8. Through a two-pronged approach of examining these helicases we have successfully:

- Confirmed that DDX52 is a 3' to 5' ATP-dependent helicase with an apparent preference for DNA within in vitro biochemical assays.
- Identified a novel annealing activity of DDX52 and successfully demonstrated that this activity is localised within the N-terminus of the protein, with data supporting a role of the IDR loop region in this activity.
- Successfully generated a heterozygous cell line of *DDX52*, with phenotypes supporting previous research as well as identifying a novel inhibited migratory phenotype relevant to metastatic cancers.
- Confirmed that DDX49 is a 3' to 5' ATP-dependent helicase with a weak affinity for nucleic acid species tested within the frame of this study.
- Identified a novel nuclease activity within DDX49 and presented data supporting that nuclease activity is localised to a TOPRIM-like domain within the extended C-terminus of the protein.
- Successfully generated a heterozygous cell line of *DDX49*, with an inhibited migratory phenotype supporting previous research within as well as novel data on the impacts of heterozygosity on rRNA species.
- Dismissed a potential link between DDX49 and a misclassified putative Asgard archaea homolog, whilst gaining insight into the biochemistry of *C. tropicalis* Dbp8.

Reference list

1. Dahm,R. (2005) Friedrich Miescher and the discovery of DNA. *Dev. Biol.*, **278**, 274–288.
2. Levene,P.A. (1910) ON THE BIOCHEMISTRY OF NUCLEIC ACIDS. *J. Am. Chem. Soc.*, **32**, 231–240.
3. Ganser,L.R., Kelly,M.L., Herschlag,D. and Al-Hashimi,H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.*, **20**, 474–489.
4. García-Rubio,M., Barroso,S.I. and Aguilera,A. (2018) Detection of DNA-RNA Hybrids In Vivo. In Muzi-Falconi,M., Brown,G.W. (eds), *Genome Instability: Methods and Protocols*, Methods in Molecular Biology. Springer, New York, NY, pp. 347–361.
5. Varshney,D., Spiegel,J., Zyner,K., Tannahill,D. and Balasubramanian,S. (2020) The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.*, **21**, 459–474.
6. Deininger,P. (2011) Alu elements: know the SINEs. *Genome Biol.*, **12**, 236.
7. Daus,M.L. (2016) Disease Transmission by Misfolded Prion-Protein Isoforms, Prion-Like Amyloids, Functional Amyloids and the Central Dogma. *Biology*, **5**, 2.
8. Bandyra,K. and Luisi,B. (2015) Central dogma alchemy. *RNA*, **21**, 558–559.
9. Robinson,V.L. (2009) Rethinking the central dogma: Noncoding RNAs are biologically relevant. *Urol. Oncol. Semin. Orig. Investig.*, **27**, 304–306.
10. Frost,R.J.A. and Olson,E.N. (2011) Control of glucose homeostasis and insulin sensitivity by the Let-7 family of microRNAs. *Proc. Natl. Acad. Sci.*, **108**, 21075–21080.
11. Li,Y., Tong,Y., Liu,J. and Lou,J. (2022) The Role of MicroRNA in DNA Damage Response. *Front. Genet.*, **13**, 850038.
12. Wang,Y., Baskerville,S., Shenoy,A., Babiarz,J.E., Baehner,L. and Blelloch,R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat. Genet.*, **40**, 1478–1483.
13. Fabian,M.R. and Sonenberg,N. (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.*, **19**, 586–593.
14. Friedman,R.C., Farh,K.K.-H., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

References and additional information

15. Alles,J., Fehlmann,T., Fischer,U., Backes,C., Galata,V., Minet,M., Hart,M., Abu-Halima,M., Grässer,F.A., Lenhof,H.-P., *et al.* (2019) An estimate of the total number of true human miRNAs. *Nucleic Acids Res.*, **47**, 3353–3364.
16. Vidigal,J.A. and Ventura,A. (2015) The biological functions of miRNAs: lessons from in vivo studies. *Trends Cell Biol.*, **25**, 137–147.
17. Chang,H.Y. and Qi,L.S. (2023) Reversing the Central Dogma: RNA-guided control of DNA in epigenetics and genome editing. *Mol. Cell*, **83**, 442–451.
18. Huarte,M., Guttman,M., Feldser,D., Garber,M., Koziol,M.J., Kenzelmann-Broz,D., Khalil,A.M., Zuk,O., Amit,I., Rabani,M., *et al.* (2010) A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, **142**, 409–419.
19. Romero-Barrios,N., Legascue,M.F., Benhamed,M., Ariel,F. and Crespi,M. (2018) Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.*, **46**, 2169–2184.
20. Statello,L., Guo,C.-J., Chen,L.-L. and Huarte,M. (2021) Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.*, **22**, 96–118.
21. Blokhin,I., Khorkova,O., Hsiao,J. and Wahlestedt,C. (2018) Developments in lncRNA drug discovery: where are we heading? *Expert Opin. Drug Discov.*, **13**, 837–849.
22. Lee,S., Kopp,F., Chang,T.-C., Sataluri,A., Chen,B., Sivakumar,S., Yu,H., Xie,Y. and Mendell,J.T. (2016) Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. *Cell*, **164**, 69–80.
23. Ronchetti,D., Agnelli,L., Taiana,E., Galletti,S., Manzoni,M., Todoerti,K., Musto,P., Strozzi,F. and Neri,A. (2016) Distinct lncRNA transcriptional fingerprints characterize progressive stages of multiple myeloma. *Oncotarget*, **7**, 14814–14830.
24. Tan,B.-S., Yang,M.-C., Singh,S., Chou,Y.-C., Chen,H.-Y., Wang,M.-Y., Wang,Y.-C. and Chen,R.-H. (2019) lncRNA NORAD is repressed by the YAP pathway and suppresses lung and breast cancer metastasis by sequestering S100P. *Oncogene*, **38**, 5612–5626.
25. Chen,Q., Zhu,C. and Jin,Y. (2020) The Oncogenic and Tumor Suppressive Functions of the Long Noncoding RNA MALAT1: An Emerging Controversy. *Front. Genet.*, **11**.
26. Kim,J., Piao,H.-L., Kim,B.-J., Yao,F., Han,Z., Wang,Y., Xiao,Z., Siverly,A.N., Lawhon,S.E., Ton,B.N., *et al.* (2018) Long noncoding RNA MALAT1 suppresses breast cancer metastasis. *Nat. Genet.*, **50**, 1705–1715.

References and additional information

27. Fairman-Williams,M., Guenther,U. and Jankowsky,E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–24.
28. Baldaccini,M. and Pfeffer,S. (2021) Untangling the roles of RNA helicases in antiviral innate immunity. *PLOS Pathog.*, **17**, e1010072.
29. Racki,L.R. and Narlikar,G.J. (2008) ATP-dependent Chromatin Remodeling Enzymes: Two Heads are not Better, Just Different. *Curr. Opin. Genet. Dev.*, **18**, 137–144.
30. Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.*, **76**, 23–50.
31. Bourgeois,C.F., Mortreux,F. and Auboeuf,D. (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.*, **17**, 426–438.
32. Jankowsky,E. and Fairman,M.E. (2007) RNA helicases — one fold for many functions. *Curr. Opin. Struct. Biol.*, **17**, 316–324.
33. Jarmoskaite,I. and Russell,R. (2014) RNA helicase proteins as chaperones and remodelers. *Annu. Rev. Biochem.*, **83**, 697–725.
34. Rössler,O.G., Straka,A. and Stahl,H. (2001) Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res.*, **29**, 2088–2096.
35. Valdez,B.C., Henning,D., Perumal,K. and Busch,H. (1997) RNA-unwinding and RNA-folding Activities of RNA Helicase II/Gu — Two Activities in Separate Domains of the Same Protein. *Eur. J. Biochem.*, **250**, 800–807.
36. Bhattacharyya,B. and Keck,J.L. (2014) Grip it and rip it: Structural mechanisms of DNA helicase substrate binding and unwinding. *Protein Sci.*, **23**, 1498–1507.
37. Kim,J.L., Morgenstern,K.A., Griffith,J.P., Dwyer,M.D., Thomson,J.A., Murcko,M.A., Lin,C. and Caron,P.R. (1998) Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. *Structure*, **6**, 89–100.
38. Cordin,O., Tanner,N.K., Doère,M., Linder,P. and Banroques,J. (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.*, **23**, 2478–2487.
39. Ding,H., Guo,M., Vidhyasagar,V., Talwar,T. and Wu,Y. (2015) The Q Motif Is Involved in DNA Binding but Not ATP Binding in ChIR1 Helicase. *PLoS ONE*, **10**, e0140755.
40. Fuller-Pace,F.V. (2006) DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res.*, **34**, 4206–4215.

References and additional information

41. Tanner,N.K. and Linder,P. (2001) DExD/H Box RNA Helicases: From Generic Motors to Specific Dissociation Functions. *Mol. Cell*, **8**, 251–262.
42. Xie,J., Wen,M., Zhang,J., Wang,Z., Wang,M., Qiu,Y., Zhao,W., Zhu,F., Yao,M., Rong,Z., *et al.* (2022) The Roles of RNA Helicases in DNA Damage Repair and Tumorigenesis Reveal Precision Therapeutic Strategies. *Cancer Res.*, **82**, 872–884.
43. Aratani,S., Fujii,R., Oishi,T., Fujita,H., Amano,T., Ohshima,T., Hagiwara,M., Fukamizu,A. and Nakajima,T. (2001) Dual Roles of RNA Helicase A in CREB-Dependent Transcription. *Mol. Cell. Biol.*, **21**, 4460–4469.
44. Yan,X., Mouillet,J.-F., Ou,Q. and Sadovsky,Y. (2003) A novel domain within the DEAD-box protein DP103 is essential for transcriptional repression and helicase activity. *Mol. Cell. Biol.*, **23**, 414–423.
45. Arna,A.B., Patel,H., Singh,R.S., Vizeacoumar,F.S., Kusalik,A., Freywald,A., Vizeacoumar,F.J. and Wu,Y. (2023) Synthetic lethal interactions of DEAD/H-box helicases as targets for cancer therapy. *Front. Oncol.*, **12**, 1087989.
46. Cai,W., Xiong Chen,Z., Rane,G., Satendra Singh,S., Choo,Z., Wang,C., Yuan,Y., Zea Tan,T., Arfuso,F., Yap,C.T., *et al.* (2017) Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. *JNCI J. Natl. Cancer Inst.*, **109**, djw278.
47. Cui,Y., Hunt,A., Li,Z., Birkin,E., Lane,J., Ruge,F. and Jiang,W.G. (2021) Lead DEAD/H box helicase biomarkers with the therapeutic potential identified by integrated bioinformatic approaches in lung cancer. *Comput. Struct. Biotechnol. J.*, **19**, 261.
48. Strittmatter,L.M., Capitanchik,C., Newman,A.J., Hallegger,M., Norman,C.M., Fica,S.M., Oubridge,C., Luscombe,N.M., Ule,J. and Nagai,K. (2021) psiCLIP reveals dynamic RNA binding by DEAH-box helicases before and after exon ligation. *Nat. Commun.*, **12**, 1488.
49. He,Y., Andersen,G.R. and Nielsen,K.H. (2011) The function and architecture of DEAH/RHA helicases. *Biomol. Concepts*, **2**, 315–326.
50. He,Y., Andersen,G.R. and Nielsen,K.H. (2010) Structural basis for the function of DEAH helicases. *EMBO Rep.*, **11**, 180–186.
51. Chen,M.C. and Ferré-D’Amaré,A.R. (2017) Structural Basis of DEAH/RHA Helicase Activity. *Crystals*, **7**, 253.
52. Lattmann,S., Giri,B., Vaughn,J.P., Akman,S.A. and Nagamine,Y. (2010) Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.*, **38**, 6219–6233.
53. Meier,M., Patel,T.R., Booy,E.P., Marushchak,O., Okun,N., Deo,S., Howard,R., McEleney,K., Harding,S.E., Stetefeld,J., *et al.* (2013) Binding

References and additional information

- of G-quadruplexes to the N-terminal Recognition Domain of the RNA Helicase Associated with AU-rich Element (RHAU) *. *J. Biol. Chem.*, **288**, 35014–35027.
54. Robert-Paganin,J., Halladjian,M., Blaud,M., Lebaron,S., Delbos,L., Chardon,F., Capeyrou,R., Humbert,O., Henry,Y., Henras,A.K., *et al.* (2017) Functional link between DEAH/RHA helicase Prp43 activation and ATP base binding. *Nucleic Acids Res.*, **45**, 1539–1552.
55. Schutz,P., Karlberg,T., van den Berg,S., Collins,R., Lehtio,L., Hogbom,M., Holmberg-Schiavone,L., Tempel,W., Park,H.W., Hammarstrom,M., *et al.* (2010) Comparative Structural Analysis of Human DEAD-Box RNA Helicases. *Plos One*, **5**, 11.
56. Caldwell,C.C. and Spies,M. (2017) Helicase SPRNTing through the nanopore. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 11809–11811.
57. Pyle,A.M. (2008) Translocation and Unwinding Mechanisms of RNA and DNA Helicases. *Annu. Rev. Biophys.*, **37**, 317–336.
58. Chakraborty,P. and Hiom,K. (2021) DHX9-dependent recruitment of BRCA1 to RNA promotes DNA end resection in homologous recombination. *Nat. Commun.*, **12**, 4126.
59. Lee,C.G. and Hurwitz,J. (1992) A new RNA helicase isolated from HeLa cells that catalytically translocates in the 3' to 5' direction. *J. Biol. Chem.*, **267**, 4398–4407.
60. Wang,X., Zhang,S., Zhang,Z., Mazloum,N.A., Lee,E.Y.C. and Lee,M.Y.W. (2023) The DHX9 helicase interacts with human DNA Polymerase δ 4 and stimulates its activity in D-loop extension synthesis. *DNA Repair*, **128**, 103513.
61. Nakajima,T., Uchida,C., Anderson,S.F., Lee,C.G., Hurwitz,J., Parvin,J.D. and Montminy,M. (1997) RNA helicase A mediates association of CBP with RNA polymerase II. *Cell*, **90**, 1107–1112.
62. Linder,P. and Fuller-Pace,F. (2015) Happy birthday: 25 years of DEAD-box proteins. *Methods Mol. Biol. Clifton NJ*, **1259**, 17–33.
63. Linder,P. and Jankowsky,E. (2011) From unwinding to clamping — the DEAD box RNA helicase family. *Nat. Rev. Mol. Cell Biol.*, **12**, 505–516.
64. Rogers,G.W., Richter,N.J. and Merrick,W.C. (1999) Biochemical and Kinetic Characterization of the RNA Helicase Activity of Eukaryotic Initiation Factor 4A*. *J. Biol. Chem.*, **274**, 12236–12244.
65. Yang,Q., Del Campo,M., Lambowitz,A.M. and Jankowsky,E. (2007) DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Mol. Cell*, **28**, 253–263.

References and additional information

66. Mallam,A.L., Campo,M.D., Gilman,B., Sidote,D.J. and Lambowitz,A.M. (2012) Structural basis for RNA duplex recognition and unwinding by the DEAD-box helicase Mss116p. *Nature*, **490**, 121–125.
67. Samatanga,B. and Klostermeier,D. (2014) DEAD-box RNA helicase domains exhibit a continuum between complete functional independence and high thermodynamic coupling in nucleotide and RNA duplex recognition. *Nucleic Acids Res.*, **42**, 10644–10654.
68. Theissen,B., Karow,A.R., Köhler,J., Gubaev,A. and Klostermeier,D. (2008) Cooperative binding of ATP and RNA induces a closed conformation in a DEAD box RNA helicase. *Proc. Natl. Acad. Sci.*, **105**, 548–553.
69. Del Campo,M. and Lambowitz,A.M. (2009) Structure of the Yeast DEAD Box Protein Mss116p Reveals Two Wedges that Crimp RNA. *Mol. Cell*, **35**, 598–609.
70. Sengoku,T., Nureki,O., Nakamura,A., Kobayashi,S. and Yokoyama,S. (2006) Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell*, **125**, 287–300.
71. Liu,F., Putnam,A. and Jankowsky,E. (2008) ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 20209–20214.
72. Wurm,J.P. (2023) Structural basis for RNA-duplex unwinding by the DEAD-box helicase DbpA. *RNA*, **29**, 1339–1354.
73. Liu,Y.-C. and Cheng,S.-C. (2015) Functional roles of DExD/H-box RNA helicases in Pre-mRNA splicing. *J. Biomed. Sci.*, **22**, 54.
74. Martin,R., Straub,A.U., Doebele,C. and Bohnsack,M.T. (2013) DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol.*, **10**, 4–18.
75. Cargill,M.J., Morales,A., Ravishankar,S. and Warren,E.H. (2021) RNA helicase, DDX3X, is actively recruited to sites of DNA damage in live cells. *DNA Repair*, **103**, 103137.
76. Li,L., Germain,D.R., Poon,H.-Y., Hildebrandt,M.R., Monckton,E.A., McDonald,D., Hendzel,M.J. and Godbout,R. (2016) DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Mol. Cell. Biol.*, **36**, 2794–2810.
77. Fuller-Pace,F.V. (2013) The DEAD box proteins DDX5 (p68) and DDX17 (p72): Multi-tasking transcriptional regulators. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, **1829**, 756–763.
78. Giraud,G., Terrone,S. and Bourgeois,C.F. (2018) Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep.*, **51**, 613–622.

References and additional information

79. Bonaventure,B. and Goujon,C. (2022) DExH/D-box helicases at the frontline of intrinsic and innate immunity against viral infections. *J. Gen. Virol.*, **103**.
80. Ariumi,Y. (2022) Host Cellular RNA Helicases Regulate SARS-CoV-2 Infection. *J. Virol.*, **96**, e0000222.
81. Fullam,A. and Schröder,M. (2013) DExD/H-box RNA helicases as mediators of anti-viral innate immunity and essential host factors for viral replication. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, **1829**, 854–865.
82. Awasthi,S., Verma,M., Mahesh,A., Govindaraju,G., Bacterial, Parasite Disease Biology,I.,Rajiv Gandhi Center for Biotechnology, Trivandrum 695 014, Rajavelu,A., Bacterial, Parasite Disease Biology,I.,Rajiv Gandhi Center for Biotechnology, Trivandrum 695 014, Chavali,P.L., *et al.* (2018) DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Res.*, **46**, 6304–6317.
83. Pontén,F., Jirström,K. and Uhlen,M. (2008) The Human Protein Atlas—a tool for pathology. *J. Pathol.*, **216**, 387–393.
84. 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. *J. Cell. Mol. Med.*, **24**, 1141–1145.
85. Wang,X., Gu,G., Zhu,H., Lu,S., Abuduwaili,K. and Liu,C. (2020) LncRNA SNHG20 promoted proliferation, invasion and inhibited cell apoptosis of lung adenocarcinoma via sponging miR-342 and upregulating DDX49. *Thorac. Cancer*, **11**, 3510–3520.
86. 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*, **13**, 12766–12779.
87. Aladeokin,A.C., Akiyama,T., Kimura,A., Kimura,Y., Takahashi-Jitsuki,A., Nakamura,H., Makihara,H., Masukawa,D., Nakabayashi,J., Hirano,H., *et al.* (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. *Neurobiol. Dis.*, **132**, 104603.
88. Laurent,E.M.N., Sofianatos,Y., Komarova,A., Gimeno,J.-P., Tehrani,P.S., Kim,D.-K., Abdouni,H., Duhamel,M., Cassonnet,P., Knapp,J.J., *et al.* (2020) Global BiID-based SARS-CoV-2 proteins proximal interactome unveils novel ties between viral polypeptides and host factors involved in multiple COVID19-associated mechanisms. *bioRxiv*, 10.1101/2020.08.28.272955.
89. 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

References and additional information

- of Kaposi's Sarcoma Associated Herpesvirus by Interacting with Viral mRNAs. *Viruses*, **14**, 2083.
90. Daugeron,M.-C. and Linder,P. (2001) Characterization and mutational analysis of yeast Dbp8p, a putative RNA helicase involved in ribosome biogenesis. *Nucleic Acids Res.*, **29**, 1144–1155.
91. Granneman,S., Bernstein,K.A., Bleichert,F. and Baserga,S.J. (2006) Comprehensive Mutational Analysis of Yeast DEXD/H Box RNA Helicases Required for Small Ribosomal Subunit Synthesis. *Mol. Cell. Biol.*, **26**, 1183–1194.
92. Granneman,S., Lin,C., Champion,E.A., Nandineni,M.R., Zorca,C. and Baserga,S.J. (2006) The nucleolar protein Esf2 interacts directly with the DExD/H box RNA helicase, Dbp8, to stimulate ATP hydrolysis. *Nucleic Acids Res.*, **34**, 3189–3199.
93. Zhao,R., Davey,M., Hsu,Y.-C., Kaplanek,P., Tong,A., Parsons,A.B., Krogan,N., Cagney,G., Mai,D., Greenblatt,J., *et al.* (2005) Navigating the Chaperone Network: An Integrative Map of Physical and Genetic Interactions Mediated by the Hsp90 Chaperone. *Cell*, **120**, 715–727.
94. Bohnsack,M.T., Kos,M. and Tollervey,D. (2008) Quantitative analysis of snoRNA association with pre-ribosomes and release of snR30 by Rok1 helicase. *Embo Rep.*, **9**, 1230–1236.
95. Stanchi,F., Bertocco,E., Toppo,S., Dioguardi,R., Simionati,B., Cannata,N., Zimbello,R., Lanfranchi,G. and Valle,G. (2001) Characterization of 16 novel human genes showing high similarity to yeast sequences. *Yeast*, **18**, 69–80.
96. Schutz,P., Karlberg,T., van den Berg,S., Collins,R., Lehtio,L., Hogbom,M., Holmberg-Schiavone,L., Tempel,W., Park,H.W., Hammarstrom,M., *et al.* (2010) Comparative Structural Analysis of Human DEAD-Box RNA Helicases. *Plos One*, **5**, 11.
97. Wang,Q., Qian,L., Tao,M., Liu,J. and Qi,F. (2021) Knockdown of DEAD-box RNA helicase 52 (DDX52) suppresses the proliferation of melanoma cells in vitro and of nude mouse xenografts by targeting c-Myc. *Bioengineered*, **12**, 3539–3549.
98. Yu,W., Ma,H., Li,J., Ge,J., Wang,P., Zhou,Y., Zhang,J. and Shi,G. (2021) DDX52 knockdown inhibits the growth of prostate cancer cells by regulating c-Myc signaling. *Cancer Cell Int.*, **21**, 430.
99. Zeng,H., Ge,J., Xu,W., Ma,H., Chen,L., Xia,M., Pan,B., Lin,H., Wang,S. and Gao,X. (2023) Twelve Loci Associated With Bone Density in Middle-aged and Elderly Chinese: The Shanghai Changfeng Study. *J. Clin. Endocrinol. Metab.*, **108**, 295–305.
100. Rahman,M.M., Bagdassarian,E., Ali,M.A.M. and McFadden,G. (2017) Identification of host DEAD-box RNA helicases that regulate cellular

References and additional information

- tropism of oncolytic Myxoma virus in human cancer cells. *Sci. Rep.*, **7**, 15710.
101. Williams,C.A., Abbink,T.E.M., Jeang,K.-T. and Lever,A.M.L. (2015) Identification of RNA helicases in human immunodeficiency virus 1 (HIV-1) replication - a targeted small interfering RNA library screen using pseudotyped and WT HIV-1. *J. Gen. Virol.*, **96**, 1484–1489.
102. Matsuoka,S., Ballif,B.A., Smogorzewska,A., McDonald,E.R., Hurov,K.E., Luo,J., Bakalarski,C.E., Zhao,Z., Solimini,N., Lerenthal,Y., *et al.* (2007) ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science*, **316**, 1160–1166.
103. Dephoure,N., C,Z., Villén,J., Beausoleil,S., Bakalarski,C., Elledge,S. and Gygi,S. (2008) A quantitative atlas of mitotic phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 10762–10767.
104. Buckley,R., Kramm,K., Cooper,C., Grohmann,D. and Bolt,E. (2020) Mechanistic insights into Lhr helicase function in DNA repair. *Biochem. J.*, **477**, 2935–2947.
105. Garcia,I., Albring,M.J. and Uhlenbeck,O.C. (2012) Duplex Destabilization by Four Ribosomal DEAD-Box Proteins. *Biochemistry*, **51**, 10109–10118.
106. Oh,J.Y. and Kim,J. (1999) ATP hydrolysis activity of the DEAD box protein Rok1p is required for in vivo ROK1 function. *Nucleic Acids Res.*, **27**, 2753–2759.
107. Jeon,S. and Kim,J. (2010) Upstream open reading frames regulate the cell cycle-dependent expression of the RNA helicase Rok1 in *Saccharomyces cerevisiae*. *FEBS Lett.*, **584**, 4593–4598.
108. Song,Y., Kim,S. and Kim,J. (1995) ROK1, a high-copy-number plasmid suppressor of kem1, encodes a putative ATP-dependent RNA helicase in *Saccharomyces cerevisiae*. *Gene*, **166**, 151–154.
109. Venema,J., Bousquet-Antonelli,C., Gelugne,J.-P., Caizergues-Ferrer,M. and Tollervey,D. (1997) Rok1p Is a Putative RNA Helicase Required for rRNA Processing. *Mol. Cell. Biol.*, **17**, 3398–3407.
110. Kim,J., Jeon,S., Yang,Y.-S. and Kim,J. (2004) Posttranscriptional regulation of the karyogamy gene by Kem1p/Xrn1p exoribonuclease and Rok1p RNA helicase of *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.*, **321**, 1032–1039.
111. Kim,S. and Kim,J. (1992) Studies on KEM1 gene controlling mitotic cell division in yeast: molecular cloning of high copy suppressor (ROK1) of kem1. *Korean J. Microbiol.*, **30**, 37–41.
112. Khoshnevis,S., Askenasy,I., Johnson,M.C., Dattolo,M.D., Young-Erdos,C.L., Stroupe,M.E. and Karbstein,K. (2016) The DEAD-box

References and additional information

- Protein Rok1 Orchestrates 40S and 60S Ribosome Assembly by Promoting the Release of Rrp5 from Pre-40S Ribosomes to Allow for 60S Maturation. *PLoS Biol.*, **14**, e1002480.
113. Vos,H.R., Bax,R., Faber,A.W., Vos,J.C. and Raué,H.A. (2004) U3 snoRNP and Rrp5p associate independently with *Saccharomyces cerevisiae* 35S pre-rRNA, but Rrp5p is essential for association of Rok1p. *Nucleic Acids Res.*, **32**, 5827–5833.
114. Young,C.L., Khoshnevis,S. and Karbstein,K. (2013) Cofactor-dependent specificity of a DEAD-box protein. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, E2668–E2676.
115. Torchet,C., Jacq,C. and Hermann-Le Denmat,S. (1998) Two mutant forms of the S1/TPR-containing protein Rrp5p affect the 18S rRNA synthesis in *Saccharomyces cerevisiae*. *RNA*, **4**, 1636–1652.
116. Farley-Barnes,K.I., Ogawa,L.M. and Baserga,S.J. (2019) Ribosomopathies: old concepts, new controversies. *Trends Genet.*, **35**, 754–767.
117. Narla,A. and Ebert,B.L. (2010) Ribosomopathies: human disorders of ribosome dysfunction. *Blood*, **115**, 3196–3205.
118. Baßler,J. and Hurt,E. (2019) Eukaryotic Ribosome Assembly. *Annu. Rev. Biochem.*, **88**, 281–306.
119. Kressler,D., Linder,P. and de la Cruz,J. (1999) Protein trans-Acting Factors Involved in Ribosome Biogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **19**, 7897–7912.
120. Anger,A.M., Armache,J.-P., Berninghausen,O., Habeck,M., Subklewe,M., Wilson,D.N. and Beckmann,R. (2013) Structures of the human and *Drosophila* 80S ribosome. *Nature*, **497**, 80–85.
121. Paule,M.R. and White,R.J. (2000) Transcription by RNA polymerases I and III. *Nucleic Acids Res.*, **28**, 1283–1298.
122. Riggs,D.L. and Nomura,M. (1990) Specific transcription of *Saccharomyces cerevisiae* 35 S rDNA by RNA polymerase I in vitro. *J. Biol. Chem.*, **265**, 7596–7603.
123. Henras,A.K., Plisson-Chastang,C., O'Donohue,M.-F., Chakraborty,A. and Gleizes,P.-E. (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip. Rev. RNA*, **6**, 225–242.
124. Dörner,K., Ruggeri,C., Zemp,I. and Kutay,U. (2023) Ribosome biogenesis factors-from names to functions. *EMBO J.*, **42**, e112699.
125. Fromont-Racine,M., Senger,B., Saveanu,C. and Fasiolo,F. (2003) Ribosome assembly in eukaryotes. *Gene*, **313**, 17–42.

References and additional information

126. Chaillou, T. (2019) Ribosome specialization and its potential role in the control of protein translation and skeletal muscle size. *J. Appl. Physiol.*, **127**, 599–607.
127. Mauro, V.P. and Edelman, G.M. (2002) The ribosome filter hypothesis. *Proc. Natl. Acad. Sci.*, **99**, 12031–12036.
128. Danilova, N. and Gazda, H.T. (2015) Ribosomopathies: how a common root can cause a tree of pathologies. *Dis. Model. Mech.*, **8**, 1013–1026.
129. Karaca, E., Harel, T., Pehlivan, D., Jhangiani, S.N., Gambin, T., Coban Akdemir, Z., Gonzaga-Jauregui, C., Erdin, S., Bayram, Y., Campbell, I.M., *et al.* (2015) Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. *Neuron*, **88**, 499–513.
130. Aspesi, A. and Ellis, S.R. (2019) Rare ribosomopathies: insights into mechanisms of cancer. *Nat. Rev. Cancer*, **19**, 228–238.
131. Mills, E.W. and Green, R. (2017) Ribosomopathies: There's strength in numbers. *Science*, **358**, eaan2755.
132. Hansen, M., Taubert, S., Crawford, D., Libina, N., Lee, S.-J. and Kenyon, C. (2007) Lifespan extension by conditions that inhibit translation in *Caenorhabditis elegans*. *Aging Cell*, **6**, 95–110.
133. Hannan, K.M., Sanij, E., Rothblum, L., Pearson, R.B. and Hannan, R.D. (2013) Dysregulation of RNA polymerase I transcription during disease. *Biochim. Biophys. Acta*, **1829**, 342–360.
134. Hernández-Ortega, K., Garcia-Esparcia, P., Gil, L., Lucas, J.J. and Ferrer, I. (2016) Altered Machinery of Protein Synthesis in Alzheimer's: From the Nucleolus to the Ribosome. *Brain Pathol.*, **26**, 593–605.
135. Parlato, R. and Liss, B. (2014) How Parkinson's disease meets nucleolar stress. *Biochim. Biophys. Acta - Mol. Basis Dis.*, **1842**, 791–797.
136. Sulima, S.O., Hofman, I.J.F., De Keersmaecker, K. and Dinman, J.D. (2017) How Ribosomes Translate Cancer. *Cancer Discov.*, **7**, 1069–1087.
137. Pelletier, J., Thomas, G. and Volarević, S. (2018) Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat. Rev. Cancer*, **18**, 51–63.
138. Barna, M., Pusic, A., Zollo, O., Costa, M., Kondrashov, N., Rego, E., Rao, P.H. and Ruggero, D. (2008) Suppression of Myc oncogenic activity by ribosomal protein haploinsufficiency. *Nature*, **456**, 971–975.
139. Morgado-Palacin, L., Varetto, G., Llanos, S., Gómez-López, G., Martínez, D. and Serrano, M. (2015) Partial Loss of Rpl11 in Adult Mice Recapitulates Diamond-Blackfan Anemia and Promotes Lymphomagenesis. *Cell Rep.*, **13**, 712–722.

References and additional information

140. Abdelhaleem,M., Maltais,L. and Wain,H. (2003) The human DDX and DHX gene families of putative RNA helicases. *Genomics*, **81**, 618–622.
141. Alexandrov,A., Colognori,D. and Steitz,J.A. (2011) Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes Dev.*, **25**, 1078–1090.
142. Kressler,D., de la Cruz,J., Rojo,M. and Linder,P. (1997) Fal1p Is an Essential DEAD-Box Protein Involved in 40S-Ribosomal-Subunit Biogenesis in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **17**, 7283–7294.
143. Sekiguchi,T., Hayano,T., Yanagida,M., Takahashi,N. and Nishimoto,T. (2006) NOP132 is required for proper nucleolus localization of DEAD-box RNA helicase DDX47. *Nucleic Acids Res.*, **34**, 4593–4608.
144. Jalal,C., Uhlmann-Schiffler,H. and Stahl,H. (2007) Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res.*, **35**, 3590–3601.
145. Saporita,A.J., Chang,H.-C., Winkeler,C.L., Apicelli,A.J., Kladney,R.D., Wang,J., Townsend,R.R., Michel,L.S. and Weber,J.D. (2011) RNA Helicase DDX5 Is a p53-Independent Target of ARF That Participates in Ribosome Biogenesis. *Cancer Res.*, **71**, 6708–6717.
146. Bond,A.T., Mangus,D.A., He,F. and Jacobson,A. (2001) Absence of Dbp2p Alters Both Nonsense-Mediated mRNA Decay and rRNA Processing. *Mol. Cell. Biol.*, **21**, 7366–7379.
147. Suzuki,T., Katada,E., Mizuoka,Y., Takagi,S., Kazuki,Y., Oshimura,M., Shindo,M. and Hara,T. (2021) A novel all-in-one conditional knockout system uncovered an essential role of DDX1 in ribosomal RNA processing. *Nucleic Acids Res.*, **49**, e40.
148. Leone,S., Bär,D., Slabber,C.F., Dalcher,D. and Santoro,R. (2017) The RNA helicase DHX9 establishes nucleolar heterochromatin, and this activity is required for embryonic stem cell differentiation. *EMBO Rep.*, **18**, 1248–1262.
149. Zhang,Y., Forys,J.T., Miceli,A.P., Gwinn,A.S. and Weber,J.D. (2011) Identification of DHX33 as a mediator of rRNA synthesis and cell growth. *Mol. Cell. Biol.*, **31**, 4676–4691.
150. Henning,D., So,R.B., Jin,R., Lau,L.F. and Valdez,B.C. (2003) Silencing of RNA Helicase II/Gua Inhibits Mammalian Ribosomal RNA Production*. *J. Biol. Chem.*, **278**, 52307–52314.
151. Wu,T., Nance,J., Chu,F. and Fazio,T.G. (2021) Characterization of R-Loop-Interacting Proteins in Embryonic Stem Cells Reveals Roles in rRNA Processing and Gene Expression. *Mol. Cell. Proteomics*, **20**, 100142.

References and additional information

152. Turner,A.J., Knox,A.A., Prieto,J.-L., McStay,B. and Watkins,N.J. (2009) A Novel Small-Subunit Processome Assembly Intermediate That Contains the U3 snoRNP, Nucleolin, RRP5, and DBP4. *Mol. Cell. Biol.*, **29**, 3007–3017.
153. Zhang,H., Wu,Z., Lu,J.Y., Huang,B., Zhou,H., Xie,W., Wang,J. and Shen,X. (2020) DEAD-Box Helicase 18 Counteracts PRC2 to Safeguard Ribosomal DNA in Pluripotency Regulation. *Cell Rep.*, **30**, 81–97.
154. Tafforeau,L., Zorbas,C., Langhendries,J.-L., Mullineux,S.-T., Stamatopoulou,V., Mullier,R., Wacheul,L. and Lafontaine,D.L.J. (2013) The Complexity of Human Ribosome Biogenesis Revealed by Systematic Nucleolar Screening of Pre-rRNA Processing Factors. *Mol. Cell*, **51**, 539–551.
155. Yamauchi,T., Nishiyama,M., Moroishi,T., Yumimoto,K. and Nakayama,K.I. (2014) MDM2 Mediates Nonproteolytic Polyubiquitylation of the DEAD-Box RNA Helicase DDX24. *Mol. Cell. Biol.*, **34**, 3321–3340.
156. Zagulski,M., Kressler,D., Bécam,A.-M., Rytka,J. and Herbert,C.J. (2003) Mak5p, which is required for the maintenance of the M1 dsRNA virus, is encoded by the yeast ORF YBR142w and is involved in the biogenesis of the 60S subunit of the ribosome. *Mol. Genet. Genomics*, **270**, 216–224.
157. Kellner,M., Rohrmoser,M., Forné,I., Voss,K., Burger,K., Mühl,B., Gruber-Eber,A., Kremmer,E., Imhof,A. and Eick,D. (2015) DEAD-box helicase DDX27 regulates 3' end formation of ribosomal 47S RNA and stably associates with the PeBoW-complex. *Exp. Cell Res.*, **334**, 146–159.
158. Srivastava,L., Lapik,Y.R., Wang,M. and Pestov,D.G. (2010) Mammalian DEAD Box Protein Ddx51 Acts in 3' End Maturation of 28S rRNA by Promoting the Release of U8 snoRNA. *Mol. Cell. Biol.*, **30**, 2947–2956.
159. Milek,M., Imami,K., Mukherjee,N., Bortoli,F.D., Zinnall,U., Hazapis,O., Trahan,C., Oeffinger,M., Heyd,F., Ohler,U., *et al.* (2017) DDX54 regulates transcriptome dynamics during DNA damage response. *Genome Res.*, **27**, 1344–1359.
160. Bizen,N., Bepari,A.K., Zhou,L., Abe,M., Sakimura,K., Ono,K. and Takebayashi,H. (2022) Ddx20, an Olig2 binding factor, governs the survival of neural and oligodendrocyte progenitor cells via proper Mdm2 splicing and p53 suppression. *Cell Death Differ.*, **29**, 1028–1041.
161. Robinson,N.P. and Bell,S.D. (2005) Origins of DNA replication in the three domains of life. *FEBS J.*, **272**, 3757–3766.
162. Hu,Y. and Stillman,B. (2023) Origins of DNA replication in eukaryotes. *Mol. Cell*, **83**, 352–372.
163. Costa,A. and Diffley,J.F.X. (2022) The Initiation of Eukaryotic DNA Replication. *Annu. Rev. Biochem.*, **91**, 107–131.

References and additional information

164. Zou,Y., Pei,J., Long,H., Lan,L., Dong,K., Wang,T., Li,M., Zhao,Z., Zhu,L., Zhang,G., *et al.* (2023) H4S47 O-GlcNAcylation regulates the activation of mammalian replication origins. *Nat. Struct. Mol. Biol.*, **30**, 800–811.
165. Loeb,L.A. and Monnat,R.J. (2008) DNA polymerases and human disease. *Nat. Rev. Genet.*, **9**, 594–604.
166. Stillman,B. (2008) DNA polymerases at the replication fork in eukaryotes. *Mol. Cell*, **30**, 259–260.
167. Walter,J. and Newport,J. (2000) Initiation of Eukaryotic DNA Replication: Origin Unwinding and Sequential Chromatin Association of Cdc45, RPA, and DNA Polymerase α . *Mol. Cell*, **5**, 617–627.
168. Nick McElhinny,S.A., Gordenin,D.A., Stith,C.M., Burgers,P.M.J. and Kunkel,T.A. (2008) Division of Labor at the Eukaryotic Replication Fork. *Mol. Cell*, **30**, 137–144.
169. Shinbrot,E., Henninger,E.E., Weinhold,N., Covington,K.R., Göksenin,A.Y., Schultz,N., Chao,H., Doddapaneni,H., Muzny,D.M., Gibbs,R.A., *et al.* (2014) Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome Res.*, **24**, 1740–1750.
170. Goulian,M., Richards,S.H., Heard,C.J. and Bigsby,B.M. (1990) Discontinuous DNA synthesis by purified mammalian proteins. *J. Biol. Chem.*, **265**, 18461–18471.
171. Hosfield,D.J., Mol,C.D., Shen,B. and Tainer,J.A. (1998) Structure of the DNA Repair and Replication Endonuclease and Exonuclease FEN-1: Coupling DNA and PCNA Binding to FEN-1 Activity. *Cell*, **95**, 135–146.
172. Lee,J.-B., Hite,R.K., Hamdan,S.M., Sunney Xie,X., Richardson,C.C. and van Oijen,A.M. (2006) DNA primase acts as a molecular brake in DNA replication. *Nature*, **439**, 621–624.
173. Nelson,S.W., Kumar,R. and Benkovic,S.J. (2008) RNA Primer Handoff in Bacteriophage T4 DNA Replication. *J. Biol. Chem.*, **283**, 22838–22846.
174. Duderstadt,K.E., Geertsema,H.J., Stratmann,S.A., Punter,C.M., Kulczyk,A.W., Richardson,C.C. and van Oijen,A.M. (2016) Simultaneous Real-Time Imaging of Leading and Lagging Strand Synthesis Reveals the Coordination Dynamics of Single Replisomes. *Mol. Cell*, **64**, 1035–1047.
175. Dewar,J.M. and Walter,J.C. (2017) Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell Biol.*, **18**, 507–516.
176. Dewar,J.M., Low,E., Mann,M., Räschle,M. and Walter,J.C. (2017) CRL2Lrr1 promotes unloading of the vertebrate replisome from chromatin during replication termination. *Genes Dev.*, **31**, 275–290.

References and additional information

177. Langston,L.D., Zhang,D., Yurieva,O., Georgescu,R.E., Finkelstein,J., Yao,N.Y., Indiani,C. and O'Donnell,M.E. (2014) CMG helicase and DNA polymerase ϵ form a functional 15-subunit holoenzyme for eukaryotic leading-strand DNA replication. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 15390–15395.
178. Zhou,Z.-X., Lujan,S.A., Burkholder,A.B., Garbacz,M.A. and Kunkel,T.A. (2019) Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat. Commun.*, **10**, 3992.
179. Hamperl,S. and Cimprich,K.A. (2016) Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell*, **167**, 1455–1467.
180. Chow,T.T., Zhao,Y., Mak,S.S., Shay,J.W. and Wright,W.E. (2012) Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening. *Genes Dev.*, **26**, 1167–1178.
181. Tiwari,V. and Wilson,D.M. (2019) DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. *Am. J. Hum. Genet.*, **105**, 237–257.
182. Pontén,F., Jirström,K. and Uhlen,M. (2008) The Human Protein Atlas--a tool for pathology. *J. Pathol.*, **216**, 387–393.
183. Chatterjee,N. and Walker,G.C. (2017) Mechanisms of DNA damage, repair and mutagenesis. *Environ. Mol. Mutagen.*, **58**, 235–263.
184. Caldecott,K.W. (2022) DNA single-strand break repair and human genetic disease. *Trends Cell Biol.*, **32**, 733–745.
185. Seol,J.H., Shim,E.Y. and Lee,S.E. (2018) Microhomology-mediated end joining: Good, bad and ugly. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.*, **809**, 81–87.
186. Sinha,S., Villarreal,D., Shim,E.Y. and Lee,S.E. (2016) Risky business: Microhomology-mediated end joining. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.*, **788**, 17–24.
187. Bhargava,R., Onyango,D.O. and Stark,J.M. (2016) Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet.*, **32**, 566–575.
188. Vu,T.V., Das,S., Nguyen,C.C., Kim,J. and Kim,J.-Y. (2022) Single-strand annealing: Molecular mechanisms and potential applications in CRISPR-Cas-based precision genome editing. *Biotechnol. J.*, **17**, e2100413.
189. Mao,Z.Y., Bozzella,M., Seluanov,A. and Gorbunova,V. (2008) Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair*, **7**, 1765–1771.

References and additional information

190. Ira, G. and Haber, J.E. (2002) Characterization of RAD51-Independent Break-Induced Replication That Acts Preferentially with Short Homologous Sequences. *Mol. Cell. Biol.*, **22**, 6384–6392.
191. Haber, J.E. (2018) DNA repair: the search for homology. *BioEssays News Rev. Mol. Cell. Dev. Biol.*, **40**, e1700229.
192. Kramara, J., Osia, B. and Malkova, A. (2018) Break-Induced Replication: The Where, The Why, and The How. *Trends Genet.*, **34**, 518–531.
193. Lamarche, B.J., Orazio, N.I. and Weitzman, M.D. (2010) The MRN complex in double-strand break repair and telomere maintenance. *FEBS Lett.*, **584**, 3682–3695.
194. Broderick, R., Niemnuszcz, J., Baddock, H.T., Deshpande, R.A., Gileadi, O., Paull, T.T., McHugh, P.J. and Niedzwiedz, W. (2016) EXD2 promotes homologous recombination by facilitating DNA end resection. *Nat. Cell Biol.*, **18**, 271+.
195. Anand, R., Ranjha, L., Cannavo, E. and Cejka, P. (2016) Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol. Cell*, **64**, 940–950.
196. Reginato, G. and Cejka, P. (2020) The MRE11 complex: A versatile toolkit for the repair of broken DNA. *DNA Repair*, **91–92**, 102869.
197. Sugiyama, T., Zaitseva, E.M. and Kowalczykowski, S.C. (1997) A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.*, **272**, 7940–7945.
198. Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P. and Janscak, P. (2014) DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J. Biol. Chem.*, **289**, 27314–27326.
199. Zhao, F., Kim, W., Kloeber, J.A. and Lou, Z. (2020) DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells. *Exp. Mol. Med.*, **52**, 1705–1714.
200. Cruz-García, A., López-Saavedra, A. and Huertas Sánchez, P. (2014) BRCA1 Accelerates CtIP-Mediated DNA-End Resection. *Cell Rep.*, **9**, 451–459.
201. Howard, S.M., Ceppi, I., Anand, R., Geiger, R. and Cejka, P. (2020) The internal region of CtIP negatively regulates DNA end resection. *Nucleic Acids Res.*, **48**, 5485–5498.
202. Jimeno, S., Fernández-Ávila, M.J., Cruz-García, A., Cepeda-García, C., Gómez-Cabello, D. and Huertas, P. (2015) Neddylation inhibits CtIP-mediated resection and regulates DNA double strand break repair pathway choice. *Nucleic Acids Res.*, **43**, 987–99.

References and additional information

203. Jackson,S.P. and Bartek,J. (2009) The DNA-damage response in human biology and disease. *Nature*, **461**, 1071–1078.
204. Zhao,W.X., Vaithiyalingam,S., Filippo,J.S., Maranon,D.G., Jimenez-Sainz,J., Fontenay,G.V., Kwon,Y., Leung,S.G., Lu,L., Jensen,R.B., *et al.* (2015) Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry. *Mol. Cell*, **59**, 176–187.
205. Stefanovie,B., Hengel,S.R., Mlcouskova,J., Prochazkova,J., Spirek,M., Nikulenkov,F., Nemecek,D., Koch,B.G., Bain,F.E., Yu,L.P., *et al.* (2020) DSS1 interacts with and stimulates RAD52 to promote the repair of DSBs. *Nucleic Acids Res.*, **48**, 694–708.
206. Renkawitz,J., Lademann,C.A. and Jentsch,S. (2014) DNA damage Mechanisms and principles of homology search during recombination. *Nat. Rev. Mol. Cell Biol.*, **15**, 369–383.
207. Greene,E.C. (2016) DNA Sequence Alignment during Homologous Recombination. *J. Biol. Chem.*, **291**, 11572–11580.
208. Crickard,J., Moevus,C., Kwon,Y., Sung,P. and Greene,E. (2020) Rad54 Drives ATP Hydrolysis-Dependent DNA Sequence Alignment During Homologous Recombination. *Cell*, **181**, 1380–1394.
209. Ranjha,L., Howard,S.M. and Cejka,P. (2018) Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, **127**, 187–214.
210. Brandsma,I. and Gent,D.C. (2012) Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr.*, **3**, 9.
211. Heyer,W.D., Ehmsen,K.T. and Liu,J. (2010) Regulation of Homologous Recombination in Eukaryotes. *Annu. Rev. Genet. Vol 44*, **44**, 113–139.
212. Brouwer,I., Zhang,H., Candelli,A., Normanno,D., Peterman,E.J.G., Wuite,G.J.L. and Modesti,M. (2017) Human RAD52 Captures and Holds DNA Strands, Increases DNA Flexibility, and Prevents Melting of Duplex DNA: Implications for DNA Recombination. *Cell Rep.*, **18**, 2845–2853.
213. Nimonkar,A.V., Sica,R.A. and Kowalczykowski,S.C. (2009) Rad52 promotes second-end DNA capture in double-stranded break repair to form complement-stabilized joint molecules. *Proc. Natl. Acad. Sci.*, **106**, 3077–3082.
214. Nimonkar,A.V. and Kowalczykowski,S.C. (2009) Second-end DNA capture in double-strand break repair. *Cell Cycle*, **8**, 1816–1817.
215. Liu,J. and Heyer,W.D. (2011) Who’s who in human recombination: BRCA2 and RAD52. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 441–442.

References and additional information

216. Mazloum,N. and Holloman,W.K. (2009) Second-end capture in DNA double-strand break repair promoted by Brh2 protein of *Ustilago maydis*. *Mol. Cell*, **33**, 160–170.
217. Matos,J., Blanco,M.G., Maslen,S., Skehel,J.M. and West,S.C. (2011) Regulatory Control of the Resolution of DNA Recombination Intermediates during Meiosis and Mitosis. *Cell*, **147**, 158–172.
218. West,S.C., Blanco,M.G., Chan,Y.W., Matos,J., Sarbajna,S. and Wyatt,H.D.M. (2015) Resolution of Recombination Intermediates: Mechanisms and Regulation. *Cold Spring Harb. Symp. Quant. Biol.*, **80**, 103–109.
219. Bizard,A.H. and Hickson,I.D. (2014) The Dissolution of Double Holliday Junctions. *Cold Spring Harb. Perspect. Biol.*, **6**, 14.
220. Ling,C., Huang,J., Yan,Z.J., Li,Y.J., Ohzeki,M., Ishiai,M., Xu,D.Y., Takata,M., Seidman,M. and Wang,W.D. (2016) Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov.*, **2**, 18.
221. McMahonill,M.S., Sham,C.W. and Bishop,D.K. (2007) Synthesis-dependent strand annealing in meiosis. *Plos Biol.*, **5**, 2589–2601.
222. Meng,X. and Zhao,X. (2016) Replication fork regression and its regulation. *FEMS Yeast Res.*, **17**, fow110.
223. Zapotoczny,G. and Sekelsky,J. (2017) Human Cell Assays for Synthesis-Dependent Strand Annealing and Crossing over During Double-Strand Break Repair. *G3 GenesGenomesGenetics*, **7**, 1191–1199.
224. Williams,A.L., Genovese,G., Dyer,T., Altemose,N., Truax,K., Jun,G., Patterson,N., Myers,S.R., Curran,J.E., Duggirala,R., *et al.* Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *eLife*, **4**, e04637.
225. Holsclaw,J.K. and Sekelsky,J. (2017) Annealing of Complementary DNA Sequences During Double-Strand Break Repair in *Drosophila* Is Mediated by the Ortholog of SMARCAL1. *Genetics*, **206**, 467–480.
226. Byrne,S.M., Ortiz,L., Mali,P., Aach,J. and Church,G.M. (2015) Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res.*, **43**, 12.
227. Paix,A., Folkmann,A., Goldman,D.H., Kulaga,H., Grzelak,M.J., Rasoloson,D., Paidemarry,S., Green,R., Reed,R.R. and Seydoux,G. (2017) Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E10745–E10754.

References and additional information

228. Wang,S., Zickler,D., Kleckner,N. and Zhang,L. (2015) Meiotic crossover patterns: Obligatory crossover, interference and homeostasis in a single process. *Cell Cycle*, **14**, 305–314.
229. Charlesworth,B. and Charlesworth,D. (1975) An experimental on recombination load in *Drosophila melanogaster*. *Genet. Res.*, **25**, 267–274.
230. Felsenstein,J. (1974) The evolutionary advantage of recombination. *Genetics*, **78**, 737–756.
231. Furuse,M., Nagase,Y., Tsubouchi,H., Murakami-Murofushi,K., Shibata,T. and Ohta,K. (1998) Distinct roles of two separable in vitro activities of yeast Mre11 in mitotic and meiotic recombination. *EMBO J.*, **17**, 6412–6425.
232. Zhai,B., Zhang,S., Li,B., Zhang,J., Yang,X., Tan,Y., Wang,Y., Tan,T., Yang,X., Chen,B., *et al.* (2023) Dna2 removes toxic ssDNA-RPA filaments generated from meiotic recombination-associated DNA synthesis. *Nucleic Acids Res.*, **51**, 7914–7935.
233. Singh,G., Ines,O.D., Gallego,M.E. and White,C.I. (2017) Analysis of the impact of the absence of RAD51 strand exchange activity in Arabidopsis meiosis. *PLOS ONE*, **12**, e0183006.
234. Szurman-Zubrzycka,M., Baran,B., Stolarek-Januszkiewicz,M., Kwaśniewska,J., Szarejko,I. and Gruszka,D. (2019) The dmc1 Mutant Allows an Insight Into the DNA Double-Strand Break Repair During Meiosis in Barley (*Hordeum vulgare* L.). *Front. Plant Sci.*, **10**.
235. Sansam,C.L. and Pezza,R.J. (2015) Connecting by breaking and repairing: mechanisms of DNA strand exchange in meiotic recombination. *FEBS J.*, **282**, 2444–2457.
236. Hinch,A.G., Becker,P.W., Li,T., Moralli,D., Zhang,G., Bycroft,C., Green,C., Keeney,S., Shi,Q., Davies,B., *et al.* (2020) The Configuration of RPA, RAD51, and DMC1 Binding in Meiosis Reveals the Nature of Critical Recombination Intermediates. *Mol. Cell*, **79**, 689–701.
237. Hunter,N. (2015) Meiotic Recombination: The Essence of Heredity. *Cold Spring Harb. Perspect. Biol.*, **7**, a016618.
238. Kulkarni,D.S., Owens,S.N., Honda,M., Ito,M., Yang,Y., Corrigan,M.W., Chen,L., Quan,A.L. and Hunter,N. (2020) PCNA activates the MutLγ endonuclease to promote meiotic crossing over. *Nature*, **586**, 623–627.
239. Gioia,M., Payero,L., Salim,S., V,G.F., Farnaz,A.F., Pannafino,G., Chen,J.J., Ajith,V.P., Momoh,S., Scotland,M., *et al.* (2023) Exo1 protects DNA nicks from ligation to promote crossover formation during meiosis. *PLOS Biol.*, **21**, e3002085.

References and additional information

240. Sanchez,A., Adam,C., Rauh,F., Duroc,Y., Ranjha,L., Lombard,B., Mu,X., Wintrebert,M., Loew,D., Guarné,A., *et al.* (2020) Exo1 recruits Cdc5 polo kinase to MutLy to ensure efficient meiotic crossover formation. *Proc. Natl. Acad. Sci.*, **117**, 30577–30588.
241. Wyatt,H.D.M., Laister,R.C., Martin,S.R., Arrowsmith,C.H. and West,S.C. (2017) The SMX DNA Repair Tri-nuclease. *Mol. Cell*, **65**, 848-860.e11.
242. Wyatt,H.D.M. and West,S.C. (2014) Holliday Junction Resolvases. *Cold Spring Harb. Perspect. Biol.*, **6**, a023192.
243. Hughes,S.E. and Hawley,R.S. (2020) Meiosis: Location, Location, Location, How Crossovers Ensure Segregation. *Curr. Biol.*, **30**, R311–R313.
244. Katis,V.L., Galova,M., Rabitsch,K.P., Gregan,J. and Nasmyth,K. (2004) Maintenance of Cohesin at Centromeres after Meiosis I in Budding Yeast Requires a Kinetochore-Associated Protein Related to MEI-S332. *Curr. Biol.*, **14**, 560–572.
245. Vi,K., Jj,L., R,I., A,B., E,O., B,H., K,M., K,N. and W,Z. (2010) Rec8 phosphorylation by casein kinase 1 and Cdc7-Dbf4 kinase regulates cohesin cleavage by separase during meiosis. *Dev. Cell*, **18**.
246. Marston,A.L. (2015) Shugoshins: Tension-Sensitive Pericentromeric Adaptors Safeguarding Chromosome Segregation. *Mol. Cell. Biol.*, **35**, 634–648.
247. Nerusheva,O.O., Galander,S., Fernius,J., Kelly,D. and Marston,A.L. (2014) Tension-dependent removal of pericentromeric shugoshin is an indicator of sister chromosome biorientation. *Genes Dev.*, **28**, 1291–1309.
248. Saito,T.T. and Colaiácovo,M.P. (2017) Regulation of crossover frequency and distribution during meiotic recombination. *Cold Spring Harb. Symp. Quant. Biol.*, **82**, 223.
249. Gray,S. and Cohen,P.E. (2016) Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annu. Rev. Genet.*, **50**, 175–210.
250. Chakraborty,P., Pankajam,A.V., Lin,G., Dutta,A., Krishnaprasad,G.N., Tekkedil,M.M., Shinohara,A., Steinmetz,L.M. and Nishant,K.T. (2017) Modulating Crossover Frequency and Interference for Obligate Crossovers in *Saccharomyces cerevisiae* Meiosis. *G3 GenesGenomesGenetics*, **7**, 1511–1524.
251. Gray,S., Allison,R.M., Garcia,V., Goldman,A.S.H. and Neale,M.J. (2013) Positive regulation of meiotic DNA double-strand break formation by activation of the DNA damage checkpoint kinase Mec1(ATR). *Open Biol.*, **3**, 130019.

References and additional information

252. Cole,F., Kauppi,L., Lange,J., Roig,I., Wang,R., Keeney,S. and Jasin,M. (2012) Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nat. Cell Biol.*, **14**, 424–430.
253. Martini,E., Diaz,R.L., Hunter,N. and Keeney,S. (2006) Crossover homeostasis in yeast meiosis. *Cell*, **126**, 285–295.
254. Chan,Y.W. and West,S.C. (2014) Spatial control of the GEN1 Holliday junction resolvase ensures genome stability. *Nat. Commun.*, **5**, 11.
255. Wyatt,H.D.M., Sarbajna,S., Matos,J. and West,S.C. (2013) Coordinated Actions of SLX1-SLX4 and MUS81-EME1 for Holliday Junction Resolution in Human Cells. *Mol. Cell*, **52**, 234–247.
256. Lieber,M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.*, **79**, 181–211.
257. Shrivastav,M., De Haro,L.P. and Nickoloff,J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res.*, **18**, 134–147.
258. Arnoult,N., Correia,A., Ma,J., Merlo,A., Garcia-Gomez,S., Maric,M., Tognetti,M., Benner,C.W., Boulton,S.J., Saghatelian,A., *et al.* (2017) Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature*, **549**, 548–552.
259. Davis,A.J. and Chen,D.J. (2013) DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.*, **2**, 130–143.
260. Liu,X.Y., Shao,Z.P., Jiang,W.X., Lee,B.J. and Zha,S. (2017) PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice. *Nat. Commun.*, **8**, 13816.
261. Andres,S.N., Vergnes,A., Ristic,D., Wyman,C., Modesti,M. and Junop,M. (2012) A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res.*, **40**, 1868–1878.
262. Chang,H.H.Y., Pannunzio,N.R., Adachi,N. and Lieber,M.R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.*, **18**, 495–506.
263. Barry,E.R. and Bell,S.D. (2006) DNA Replication in the Archaea. *Microbiol. Mol. Biol. Rev.*, **70**, 876–887.
264. Jüttner,M. and Ferreira-Cerca,S. (2022) A Comparative Perspective on Ribosome Biogenesis: Unity and Diversity Across the Tree of Life. In *Ribosome Biogenesis: Methods and Protocols*, Methods in Molecular Biology. Springer US, New York, NY, pp. 3–22.
265. Woese,C.R., Kandler,O. and Wheelis,M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.*, **87**, 4576–4579.

References and additional information

266. Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.*, **74**, 5088–5090.
267. Martin, W.F., Garg, S. and Zimorski, V. (2015) Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. B Biol. Sci.*, **370**, 20140330.
268. Williams, T.A., Foster, P.G., Cox, C.J. and Embley, T.M. (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature*, **504**, 231–236.
269. Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T.J.G. (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, **521**, 173–179.
270. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., *et al.* (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, **541**, 353–358.
271. Da Cunha, V., Gaïa, M. and Forterre, P. (2022) The expanding Asgard archaea and their elusive relationships with Eukarya. *mLife*, **1**, 3–12.
272. Liu, Y., Makarova, K.S., Huang, W.-C., Wolf, Y.I., Nikolskaya, A.N., Zhang, X., Cai, M., Zhang, C.-J., Xu, W., Luo, Z., *et al.* (2021) Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature*, **593**, 553–557.
273. Spang, A., Stairs, C.W., Dombrowski, N., Eme, L., Lombard, J., Caceres, E.F., Greening, C., Baker, B.J. and Ettema, T.J.G. (2019) Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.*, **4**, 1138–1148.
274. Wu, F., Speth, D.R., Filosof, A., Crémière, A., Narayanan, A., Barco, R.A., Connon, S.A., Amend, J.P., Antoshechkin, I.A. and Orphan, V.J. (2022) Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat. Microbiol.*, **7**, 200–212.
275. Doolittle, W.F. (2020) Evolution: Two Domains of Life or Three? *Curr. Biol.*, **30**, R177–R179.
276. Xie, R., Wang, Y., Huang, D., Hou, J., Li, L., Hu, H., Zhao, X. and Wang, F. (2022) Expanding Asgard members in the domain of Archaea sheds new light on the origin of eukaryotes. *Sci. China Life Sci.*, **65**, 818–829.
277. Dacks, J.B. and Field, M.C. (2018) Evolutionary origins and specialisation of membrane transport. *Curr. Opin. Cell Biol.*, **53**, 70–76.
278. Hatano, T., Palani, S., Papatziadou, D., Salzer, R., Souza, D.P., Tamarit, D., Makwana, M., Potter, A., Haig, A., Xu, W., *et al.* (2022) Asgard archaea

References and additional information

- shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat. Commun.*, **13**, 3398.
279. Neveu,E., Khalifeh,D., Salamin,N. and Fasshauer,D. (2020) Prototypic SNARE Proteins Are Encoded in the Genomes of Heimdallarchaeota, Potentially Bridging the Gap between the Prokaryotes and Eukaryotes. *Curr. Biol.*, **30**, 2468–2480.
280. Surkont,J. and Pereira-Leal,J.B. (2016) Are There Rab GTPases in Archaea? *Mol. Biol. Evol.*, **33**, 1833–1842.
281. Penev,P.I., Fakhretaha-Aval,S., Patel,V.J., Cannone,J.J., Gutell,R.R., Petrov,A.S., Williams,L.D. and Glass,J.B. (2020) Supersized Ribosomal RNA Expansion Segments in Asgard Archaea. *Genome Biol. Evol.*, **12**, 1694–1710.
282. Imachi,H., Nobu,M.K., Nakahara,N., Morono,Y., Ogawara,M., Takaki,Y., Takano,Y., Uematsu,K., Ikuta,T., Ito,M., *et al.* (2020) Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature*, **577**, 519–525.
283. Da Cunha,V., Gaia,M., Gadelle,D., Nasir,A. and Forterre,P. (2017) Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.*, **13**, e1006810.
284. Garg,S.G., Kapust,N., Lin,W., Knopp,M., Tria,F.D.K., Nelson-Sathi,S., Gould,S.B., Fan,L., Zhu,R., Zhang,C., *et al.* (2021) Anomalous Phylogenetic Behavior of Ribosomal Proteins in Metagenome-Assembled Asgard Archaea. *Genome Biol. Evol.*, **13**, evaa238.
285. Makarova,K.S., Wolf,Y.I. and Koonin,E.V. (2019) Towards functional characterization of archaeal genomic dark matter. *Biochem. Soc. Trans.*, **47**, 389–398.
286. Jagadeeshwari,U., Sasikala,C., Rai,A., Indu,B., Ipsita,S. and Ramana,C.V. (2023) Characterization of metagenome-assembled genomes of two endo-archaea of *Candida tropicalis*. *Front. Microbiomes*, **1**.
287. Choulika,A., Perrin,A., Dujon,B. and Nicolas,J.-F. (1995) Induction of Homologous Recombination in Mammalian Chromosomes by Using the I-SceI System of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **15**, 1968–1973.
288. Rouet,P., Smih,F. and Jasin,M. (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, **14**, 8096–8106.
289. Killelea,T. and Bolt,E.L. (2017) CRISPR-Cas adaptive immunity and the three Rs. *Biosci. Rep.*, **37**, BSR20160297.
290. Makarova,K.S., Wolf,Y.I., Iranzo,J., Shmakov,S.A., Alkhnbashi,O.S., Brouns,S.J.J., Charpentier,E., Cheng,D., Haft,D.H., Horvath,P., *et al.*

References and additional information

- (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
291. Lee,H. and Sashital,D.G. (2022) Creating memories: molecular mechanisms of CRISPR adaptation. *Trends Biochem. Sci.*, **47**, 464–476.
292. Marraffini,L.A. and Sontheimer,E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181–190.
293. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**, 816–821.
294. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
295. Richardson,C.D., Ray,G.J., DeWitt,M.A., Curie,G.L. and Corn,J.E. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.*, **34**, 339–344.
296. Kupatt,C., Windisch,A., Moretti,A., Wolf,E., Wurst,W. and Walter,M.C. (2021) Genome editing for Duchenne muscular dystrophy: a glimpse of the future? *Gene Ther.*, **28**, 542–548.
297. Guo,T., Feng,Y.-L., Xiao,J.-J., Liu,Q., Sun,X.-N., Xiang,J.-F., Kong,N., Liu,S.-C., Chen,G.-Q., Wang,Y., *et al.* (2018) Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol.*, **19**, 170.
298. Saintigny,Y., Delacote,F., Boucher,D., Averbek,D. and Lopez,B.S. (2007) XRCC4 in G1 suppresses homologous recombination in S/G2, in G1 checkpoint-defective cells. *Oncogene*, **26**, 2769–2780.
299. Richardson,C., Kelsh,R.N. and J. Richardson,R. (2023) New advances in CRISPR/Cas-mediated precise gene-editing techniques. *Dis. Model. Mech.*, **16**, dmm049874.
300. Concordet,J.-P. and Haeussler,M. (2018) CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.*, **46**, W242–W245.
301. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R., *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
302. Moreno-Mateos,M.A., Vejnar,C.E., Beaudoin,J.-D., Fernandez,J.P., Mis,E.K., Khokha,M.K. and Giraldez,A.J. (2015) CRISPRscan:

References and additional information

- designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods*, **12**, 982–988.
303. Parkes,A., Kemm,F., He,L. and Killelea,T. (2021) CRISPR systems: what's new, where next? *The Biochemist*, **43**, 46–50.
304. Ma,W., Xu,Y.-S., Sun,X.-M. and Huang,H. (2021) Transposon-Associated CRISPR-Cas System: A Powerful DNA Insertion Tool. *Trends Microbiol.*, **29**, 565–568.
305. Porto,E.M., Komor,A.C., Slaymaker,I.M. and Yeo,G.W. (2020) Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.*, **19**, 839–859.
306. Anzalone,A.V., Randolph,P.B., Davis,J.R., Sousa,A.A., Koblan,L.W., Levy,J.M., Chen,P.J., Wilson,C., Newby,G.A., Raguram,A., *et al.* (2019) Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**, 149–157.
307. Böck,D., Rothgangl,T., Villiger,L., Schmidheini,L., Matsushita,M., Mathis,N., Ioannidi,E., Rimann,N., Grisch-Chan,H.M., Kreutzer,S., *et al.* (2022) In vivo prime editing of a metabolic liver disease in mice. *Sci. Transl. Med.*, **14**, eabl9238.
308. Chemello,F., Chai,A.C., Li,H., Rodriguez-Caycedo,C., Sanchez-Ortiz,E., Atmanli,A., Mireault,A.A., Liu,N., Bassel-Duby,R. and Olson,E.N. (2021) Precise correction of Duchenne muscular dystrophy exon deletion mutations by base and prime editing. *Sci. Adv.*, **7**, eabg4910.
309. Kim,H.K., Yu,G., Park,J., Min,S., Lee,S., Yoon,S. and Kim,H.H. (2021) Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.*, **39**, 198–206.
310. Scholefield,J. and Harrison,P.T. (2021) Prime editing – an update on the field. *Gene Ther.*, **28**, 396–401.
311. Bhagwat,A.M., Graumann,J., Wiegandt,R., Bentsen,M., Welker,J., Kuenne,C., Preussner,J., Braun,T. and Looso,M. (2020) multicrispr: gRNA design for prime editing and parallel targeting of thousands of targets. *Life Sci. Alliance*, **3**, e202000757.
312. Chow,R.D., Chen,J.S., Shen,J. and Chen,S. (2021) A web tool for the design of prime-editing guide RNAs. *Nat. Biomed. Eng.*, **5**, 190–194.
313. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler, and D. (2002) The Human Genome Browser at UCSC. *Genome Res.*, **12**, 996–1006.
314. Anders,C., Niewoehner,O., Duerst,A. and Jinek,M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, **513**, 569–573.

References and additional information

315. Brinkman,E.K., Chen,T., Amendola,M. and van Steensel,B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, **42**, e168.
316. Brinkman,E.K., Kousholt,A.N., Harmsen,T., Leemans,C., Chen,T., Jonkers,J. and van Steensel,B. (2018) Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res.*, **46**, e58.
317. Miller,B.R., Wei,T., Fields,C.J., Sheng,P. and Xie,M. (2018) Near-infrared fluorescent northern blot. *RNA*, **24**, 1871–1877.
318. Wells,G.R., Weichmann,F., Sloan,K.E., Colvin,D., Watkins,N.J. and Schneider,C. (2017) The ribosome biogenesis factor yUtp23/hUTP23 coordinates key interactions in the yeast and human pre-40S particle and hUTP23 contains an essential PIN domain. *Nucleic Acids Res.*, **45**, 4796–4809.
319. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
320. Mészáros,B., Erdős,G. and Dosztányi,Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.*, **46**, W329–W337.
321. Ejaz,A., Ordonez,H., Jacewicz,A., Ferrao,R. and Shuman,S. (2018) Structure of mycobacterial 3'-to-5' RNA:DNA helicase Lhr bound to a ssDNA tracking strand highlights distinctive features of a novel family of bacterial helicases. *Nucleic Acids Res.*, **46**, 442–455.
322. Madeira,F., Pearce,M., Tivey,A.R.N., Basutkar,P., Lee,J., Edbali,O., Madhusoodanan,N., Kolesnikov,A. and Lopez,R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.*, **50**, W276–W279.
323. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In Walker,J.M. (ed), *The Proteomics Protocols Handbook*, Springer Protocols Handbooks. Humana Press, Totowa, NJ, pp. 571–607.
324. Garbelli,A., Beermann,S., Cicco,G.D., Dietrich,U. and Maga,G. (2011) A Motif Unique to the Human Dead-Box Protein DDX3 Is Important for Nucleic Acid Binding, ATP Hydrolysis, RNA/DNA Unwinding and HIV-1 Replication. *PLOS ONE*, **6**, e19810.
325. Desai,N.A. and Shankar,V. (2003) Single-strand-specific nucleases. *FEMS Microbiol. Rev.*, **26**, 457–491.

References and additional information

326. Brodsky,S., Jana,T., Mittelman,K., Chapal,M., Kumar,D.K., Carmi,M. and Barkai,N. (2020) Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity. *Mol. Cell*, **79**, 459–471.
327. Babu,M.M. (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.*, **44**, 1185–1200.
328. 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. *Nat. Rev. Mol. Cell Biol.*, **24**, 749–769.
329. Georgel,P.T., Horowitz-Scherer,R.A., Adkins,N., Woodcock,C.L., Wade,P.A. and Hansen,J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.*, **278**, 32181–32188.
330. Georgel,P.T., Palacios DeBeer,M.A., Pietz,G., Fox,C.A. and Hansen,J.C. (2001) Sir3-dependent assembly of supramolecular chromatin structures in vitro. *Proc. Natl. Acad. Sci.*, **98**, 8584–8589.
331. Ma,C.J., Gibb,B., Kwon,Y., Sung,P. and Greene,E.C. (2017) Protein dynamics of human RPA and RAD51 on ssDNA during assembly and disassembly of the RAD51 filament. *Nucleic Acids Res.*, **45**, 749–761.
332. Sharma,D., Putnam,A.A. and Jankowsky,E. (2017) Biochemical differences and similarities between the DEAD-box helicase orthologs DDX3X and Ded1p. *J. Mol. Biol.*, **429**, 3730–3742.
333. Deng,S.K., Chen,H. and Symington,L.S. (2015) Replication Protein A prevents promiscuous annealing between short sequence homologies: Implications for genome integrity. *BioEssays News Rev. Mol. Cell. Dev. Biol.*, **37**, 305–313.
334. Lawson,T., El-Kamand,S., Kariawasam,R., Richard,D.J., Cubeddu,L. and Gamsjaeger,R. (2019) A Structural Perspective on the Regulation of Human Single-Stranded DNA Binding Protein 1 (hSSB1, OBFC2B) Function in DNA Repair. *Comput. Struct. Biotechnol. J.*, **17**, 441–446.
335. Meyer,R.R. and Laine,P.S. (1990) The single-stranded DNA-binding protein of Escherichia coli. *Microbiol. Rev.*, **54**, 342.
336. Wu,Y., Siino,J.S., Sugiyama,T. and Kowalczykowski,S.C. (2006) The DNA binding preference of RAD52 and RAD59 proteins: implications for RAD52 and RAD59 protein function in homologous recombination. *J. Biol. Chem.*, **281**, 40001–40009.
337. Asturias,F.J., Chung,W.-H., Kornberg,R.D. and Lorch,Y. (2002) Structural analysis of the RSC chromatin-remodeling complex. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 13477–13480.

References and additional information

338. Schnitzler,G., Sif,S. and Kingston,R.E. (1998) Human SWI/SNF interconverts a nucleosome between its base state and a stable remodeled state. *Cell*, **94**, 17–27.
339. Enders,M., Ficner,R. and Adio,S. (2022) Regulation of the DEAH/RHA helicase Prp43 by the G-patch factor Pfa1. *Proc. Natl. Acad. Sci.*, **119**, e2203567119.
340. Henn,A., Cao,W., Hackney,D.D. and De La Cruz,E.M. (2008) The ATPase Cycle Mechanism of the DEAD-box rRNA Helicase, DbpA. *J. Mol. Biol.*, **377**, 193–205.
341. Putnam,A.A. and Jankowsky,E. (2013) DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, **1829**, 884–893.
342. Song,C., Hotz-Wagenblatt,A., Voit,R. and Grummt,I. (2017) SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev.*, **31**, 1370–1381.
343. Polenkowski,M., Allister,A.B., Burbano de Lara,S., Pierce,A., Geary,B., El Bounkari,O., Wiehlmann,L., Hoffmann,A., Whetton,A.D., Tamura,T., *et al.* (2022) THOC5 complexes with DDX5, DDX17, and CDK12 to regulate R loop structures and transcription elongation rate. *iScience*, **26**, 105784.
344. Chakraborty,P. and Grosse,F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair*, **10**, 654–665.
345. Kankia,B.I. and Marky,L.A. (1999) DNA, RNA, and DNA/RNA Oligomer Duplexes: A Comparative Study of Their Stability, Heat, Hydration, and Mg²⁺ Binding Properties. *J. Phys. Chem. B*, **103**, 8759–8767.
346. Kantake,N., Madiraju,M.V.V.M., Sugiyama,T. and Kowalczykowski,S.C. (2002) Escherichia coli RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. *Proc. Natl. Acad. Sci.*, **99**, 15327–15332.
347. McGrath,S.L., Huang,S.H. and Kobryn,K. (2021) Single stranded DNA annealing is a conserved activity of telomere resolvases. *PLoS ONE*, **16**, e0246212.
348. Ozdemir,A.Y., Rusanov,T., Kent,T., Siddique,L.A. and Pomerantz,R.T. (2018) Polymerase θ -helicase efficiently unwinds DNA and RNA-DNA hybrids. *J. Biol. Chem.*, **293**, 5259–5269.
349. Lever,R.J., Simmons,E., Gamble-Milner,R., Buckley,R.J., Harrison,C., Parkes,A.J., Mitchell,L., Gausden,J.A., Škulj,S., Bertoša,B., *et al.* (2023) Archaeal Hel308 suppresses recombination through a catalytic switch that controls DNA annealing. *Nucleic Acids Res.*, 10.1093/nar/gkad572.

References and additional information

350. Yang,Q. and Jankowsky,E. (2005) ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry*, **44**, 13591–13601.
351. Khreiss,A., Capeyrou,R., Lebaron,S., Albert,B., Bohnsack,K.E., Bohnsack,M.T., Henry,Y., Henras,A.K. and Humbert,O. (2023) The DEAD-box protein Dbp6 is an ATPase and RNA annealase interacting with the peptidyl transferase center (PTC) of the ribosome. *Nucleic Acids Res.*, **51**, 744–764.
352. Lebbink,J.H.G., Fish,A., Reumer,A., Natrajan,G., Winterwerp,H.H.K. and Sixma,T.K. (2010) Magnesium Coordination Controls the Molecular Switch Function of DNA Mismatch Repair Protein MutS. *J. Biol. Chem.*, **285**, 13131–13141.
353. Hiasa,H. (2002) The Glu-84 of the ParC Subunit Plays Critical Roles in Both Topoisomerase IV–Quinolone and Topoisomerase IV–DNA Interactions. *Biochemistry*, **41**, 11779–11785.
354. Iaccarino,I., Marra,G., Dufner,P. and Jiricny,J. (2000) Mutation in the magnesium binding site of hMSH6 disables the hMutS α sliding clamp from translocating along DNA. *J. Biol. Chem.*, **275**, 2080–2086.
355. Gilmore,J.M., Sardi,M.E., Groppe,B.D., Thornton,J.L., Liu,X., Dayebgadh,G., Banks,C.A., Slaughter,B.D., Unruh,J.R., Workman,J.L., *et al.* (2016) WDR76 Co-Localizes with Heterochromatin Related Proteins and Rapidly Responds to DNA Damage. *PLOS ONE*, **11**, e0155492.
356. K,N., G,S., Jr,B., Bm,F., Nv,N., Te,B., Lc,C., Pa,F., S,L., T,F., *et al.* (2019) H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat Cell Biol*, **21**, 311–318.
357. Tseng,T.-L., Wang,Y.-T., Tsao,C.-Y., Ke,Y.-T., Lee,Y.-C., Hsu,H.-J., Poss,K.D. and Chen,C.-H. (2021) The RNA helicase Ddx52 functions as a growth switch in juvenile zebrafish. *Development*, **148**, dev199578.
358. Li,Q., Yang,H., He,L. and Wang,Q. (2018) Characterization of the Es-DDX52 involved in the spermatogonial mitosis and spermatid differentiation in Chinese mitten crab (*Eriocheir sinensis*). *Gene*, **646**, 106–119.
359. Duan,S., Huang,R.S., Zhang,W., Mi,S., Bleibel,W.K., Kistner,E.O., Cox,N.J. and Dolan,M.E. (2009) Expression and alternative splicing of folate pathway genes in HapMap lymphoblastoid cell lines. *Pharmacogenomics*, **10**, 549–563.
360. Kasavi,C. (2023) Idiopathic Pulmonary Arterial Hypertension: Network-Based Integration of Multi-Omics Data Reveals New Molecular Signatures and Candidate Drugs. *OMICS J. Integr. Biol.*, **27**, 315–326.

References and additional information

361. Guzmán,G.E., Madariaga,I., Vargas,C.J., Galeano,L.B., Guerra,M.A. and Nastasi,J.A. (2023) Identification of 17q12 microdeletion syndrome in a Latin American patient with maturity-onset diabetes of the young subtype 5: a case report. *J. Med. Case Reports*, **17**, 152.
362. Warren,E.B., Briano,J.A., Ellegood,J., DeYoung,T., Lerch,J.P. and Morrow,E.M. (2022) 17q12 deletion syndrome mouse model shows defects in craniofacial, brain and kidney development, and glucose homeostasis. *Dis. Model. Mech.*, **15**, dmm049752.
363. Xin,S. and Zhang,X. (2023) Case Report: Diabetes mellitus type MODY5 as a feature of 17q12 deletion syndrome with diabetic gastroparesis. *Front. Endocrinol.*, **14**, 1205431.
364. Roberts,J.L., Gandomi,S.K., Parra,M., Lu,I., Gau,C.-L., Dasouki,M. and Butler,M.G. (2014) Clinical report of a 17q12 microdeletion with additionally unreported clinical features. *Case Rep. Genet.*, **2014**, 264947.
365. Țuțulan-Cuniță,A., Pavel,A., Dimos,L., Nedelea,M., Ursuleanu,A., Neacșu,A., Budișteanu,M. and Stambouli,D. (2022) Phenotypic Variability of 17q12 Microdeletion Syndrome – Three Cases and Review of Literature. *Balk. J. Med. Genet. BJMG*, **24**, 71–82.
366. Jurka,J. (2000) Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**, 418–420.
367. Tarailo-Graovac,M. and Chen,N. (2009) Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Curr. Protoc. Bioinforma.*, **25**, 4.10.1-4.10.14.
368. Huang,Q., Mao,Z., Li,S., Hu,J. and Zhu,Y. (2014) A non-radioactive method for small RNA detection by northern blotting. *Rice*, **7**, 26.
369. Sioson,V.A., Kim,M. and Joo,J. (2021) Challenges in delivery systems for CRISPR-based genome editing and opportunities of nanomedicine. *Biomed. Eng. Lett.*, **11**, 217–233.
370. Hearn,M.T. and Acosta,D. (2001) Applications of novel affinity cassette methods: use of peptide fusion handles for the purification of recombinant proteins. *J. Mol. Recognit.*, **14**, 323–369.
371. Amarzguioui,M. (2004) Improved siRNA-mediated silencing in refractory adherent cell lines by detachment and transfection in suspension. *BioTechniques*, **36**, 766–768, 770.
372. Hook,B. and Schagat,T. Reverse Transfection Using FuGENE® 6 and FuGENE® HD. *Promega Corp.*
373. Kumar,R., Le,N., Tan,Z., Brown,M.E., Jiang,S. and Reineke,T.M. (2020) Efficient Polymer-Mediated Delivery of Gene-Editing Ribonucleoprotein

References and additional information

- Payloads through Combinatorial Design, Parallelized Experimentation, and Machine Learning. *ACS Nano*, **14**, 17626–17639.
374. Wan,T., Chen,Y., Pan,Q., Xu,X., Kang,Y., Gao,X., Huang,F., Wu,C. and Ping,Y. (2020) Genome editing of mutant KRAS through supramolecular polymer-mediated delivery of Cas9 ribonucleoprotein for colorectal cancer therapy. *J. Controlled Release*, **322**, 236–247.
375. Banerjee,D., Langberg,K., Abbas,S., Odermatt,E., Yerramothu,P., Volaric,M., Reidenbach,M.A., Krentz,K.J., Rubinstein,C.D., Brautigan,D.L., *et al.* (2021) A non-canonical, interferon-independent signaling activity of cGAMP triggers DNA damage response signaling. *Nat. Commun.*, **12**, 6207.
376. Hirata,M., Wittayarat,M., Namula,Z., Le,Q.A., Lin,Q., Takebayashi,K., Thongkittidilok,C., Mito,T., Tomonari,S., Tanihara,F., *et al.* (2021) Generation of mutant pigs by lipofection-mediated genome editing in embryos. *Sci. Rep.*, **11**, 23806.
377. Bennett,E.P., Petersen,B.L., Johansen,I.E., Niu,Y., Yang,Z., Chamberlain,C.A., Met,Ö., Wandall,H.H. and Frödin,M. (2020) INDEL detection, the ‘Achilles heel’ of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Res.*, **48**, 11958–11981.
378. Sentmanat,M.F., Peters,S.T., Florian,C.P., Connelly,J.P. and Pruett-Miller,S.M. (2018) A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci. Rep.*, **8**, 888.
379. Dempster,J.M., Boyle,I., Vazquez,F., Root,D., Boehm,J.S., Hahn,W.C., Tsherniak,A. and McFarland,J.M. (2021) Chronos: a CRISPR cell population dynamics model. *Genome Biol.*, **22**, 343.
380. Morrill,S.A. and Amon,A. (2019) Why haploinsufficiency persists. *Proc. Natl. Acad. Sci.*, **116**, 11866–11871.
381. Dempster,J.M., Rossen,J., Kazachkova,M., Pan,J., Kugener,G., Root,D.E. and Tsherniak,A. (2019) Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. 10.1101/720243.
382. Liehr,T. (2022) Chromosomal Heteromorphisms and Cancer Susceptibility Revisited. *Cells*, **11**, 3239.
383. Kieleczawa,J. (2006) Fundamentals of Sequencing of Difficult Templates—An Overview. *J. Biomol. Tech.*, **17**, 207–217.
384. Atkins,A., Chung,C.-H., Allen,A.G., Dampier,W., Gurrola,T.E., Sariyer,I.K., Nonnemacher,M.R. and Wigdahl,B. (2021) Off-Target Analysis in Gene Editing and Applications for Clinical Translation of CRISPR/Cas9 in HIV-1 Therapy. *Front. Genome Ed.*, **3**, 673022.

References and additional information

385. Sturme,M.H.J., van der Berg,J.P., Bouwman,L.M.S., De Schrijver,A., de Maagd,R.A., Kleter,G.A. and Battaglia-de Wilde,E. (2022) Occurrence and Nature of Off-Target Modifications by CRISPR-Cas Genome Editing in Plants. *ACS Agric. Sci. Technol.*, **2**, 192–201.
386. Bai,B. and Laiho,M. (2012) Efficient sequential recovery of nucleolar macromolecular components. *Proteomics*, **12**, 3044–3048.
387. Trieu,E.P. and Targoff,I.N. (2015) Immunoprecipitation: Western Blot for Proteins of Low Abundance. In Kurien,B.T., Scofield,R.H. (eds), *Western Blotting: Methods and Protocols*, Methods in Molecular Biology. Springer, New York, NY, pp. 327–342.
388. Rouillard,A.D., Gundersen,G.W., Fernandez,N.F., Wang,Z., Monteiro,C.D., McDermott,M.G. and Ma'ayan,A. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database*, **2016**, baw100.
389. Janssen,A. and Medema,R.H. (2013) Genetic instability: tipping the balance. *Oncogene*, **32**, 4459–4470.
390. Ponti,A., Machacek,M., Gupton,S.L., Waterman-Storer,C.M. and Danuser,G. (2004) Two Distinct Actin Networks Drive the Protrusion of Migrating Cells. *Science*, **305**, 1782–1786.
391. Keese,C.R., Wegener,J., Walker,S.R. and Giaever,I. (2004) Electrical wound-healing assay for cells in vitro. *Proc. Natl. Acad. Sci.*, **101**, 1554–1559.
392. Zordan,M.D., Mill,C.P., Riese II,D.J. and Leary,J.F. (2011) A high throughput, interactive imaging, bright-field wound healing assay. *Cytometry A*, **79A**, 227–232.
393. Wang,H., Yu,J., Wang,X. and Zhang,Y. (2019) The RNA helicase DHX33 is required for cancer cell proliferation in human glioblastoma and confers resistance to PI3K/mTOR inhibition. *Cell. Signal.*, **54**, 170–178.
394. Freitas,J.T., Jozic,I. and Bedogni,B. (2021) Wound Healing Assay for Melanoma Cell Migration. In Hargadon,K.M. (ed), *Melanoma: Methods and Protocols*, Methods in Molecular Biology. Springer US, New York, NY, pp. 65–71.
395. Schadendorf,D., Fisher,D.E., Garbe,C., Gershenwald,J.E., Grob,J.-J., Halpern,A., Herlyn,M., Marchetti,M.A., McArthur,G., Ribas,A., *et al.* (2015) Melanoma. *Nat. Rev. Dis. Primer*, **1**, 15003.
396. Warren,G.M., Ejaz,A., Fay,A., Glickman,M.S. and Shuman,S. (2023) Mycobacterial helicase Lhr abets resistance to DNA crosslinking agents mitomycin C and cisplatin. *Nucleic Acids Res.*, **51**, 218–235.

References and additional information

397. Iyer-Bierhoff,A. and Grummt,I. (2019) Stop-and-Go: Dynamics of Nucleolar Transcription During the Cell Cycle. *Epigenetics Insights*, **12**, 2516865719849090.
398. Seki,A. and Rutz,S. (2018) Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *J. Exp. Med.*, **215**, 985–997.
399. Chen,C.-H., Xiao,T., Xu,H., Jiang,P., Meyer,C.A., Li,W., Brown,M. and Liu,X.S. (2018) Improved design and analysis of CRISPR knockout screens. *Bioinformatics*, **34**, 4095–4101.
400. Zhang,Z., Lian,X., Yue,H., Xiang,D. and Niu,Z. (2023) Identification and validation of the model consisting of DDX49, EGFR, and T-stage as a possible risk factor for lymph node metastasis in patients with lung cancer. *Thorac. Cancer*, **14**, 1492–1499.
401. Jäger,S., Cimermancic,P., Gulbahce,N., Johnson,J.R., McGovern,K.E., Clarke,S.C., Shales,M., Mercenne,G., Pache,L., Li,K., *et al.* (2012) Global landscape of HIV–human protein complexes. *Nature*, **481**, 365–370.
402. Zhou,H., Xu,M., Huang,Q., Gates,A.T., Zhang,X.D., Castle,J.C., Stec,E., Ferrer,M., Strulovici,B., Hazuda,D.J., *et al.* (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe*, **4**, 495–504.
403. Salgado,M., López-Romero,P., Callejas,S., López,M., Labarga,P., Dopazo,A., Soriano,V. and Rodés,B. (2011) Characterization of host genetic expression patterns in HIV-infected individuals with divergent disease progression. *Virology*, **411**, 103–112.
404. Finkel,Y., Mizrahi,O., Nachshon,A., Weingarten-Gabbay,S., Morgenstern,D., Yahalom-Ronen,Y., Tamir,H., Achdout,H., Stein,D., Israeli,O., *et al.* (2021) The coding capacity of SARS-CoV-2. *Nature*, **589**, 125–130.
405. Liu,J., Zhang,Y., Han,L., Guo,S., Wu,S., Doud,E.H., Wang,C., Chen,H., Rubart-von der Lohe,M., Wan,J., *et al.* (2022) Genome-wide analyses reveal the detrimental impacts of SARS-CoV-2 viral gene Orf9c on human pluripotent stem cell-derived cardiomyocytes. *Stem Cell Rep.*, **17**, 522–537.
406. 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 Urol.*, **23**, 66.
407. Chen,Y., Ma,Q. and Sun,N. (2023) DDX49 Promotes Proliferation and Metastasis of Cervical Cancer by Regulation of AKT and Wnt/ β -Catenin Signaling. *Ann. Clin. Lab. Sci.*, **53**, 271–277.

References and additional information

408. NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D13.
409. Tseng,S.S., Weaver,P.L., Liu,Y., Hitomi,M., Tartakoff,A.M. and Chang,T.H. (1998) Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *EMBO J.*, **17**, 2651–2662.
410. Gomes,P.S.F.C., Gomes,D.E.B. and Bernardi,R.C. (2022) Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy. *Front. Bioinforma.*, **2**, 983306.
411. Spriestersbach,A., Kubicek,J., Schäfer,F., Block,H. and Maertens,B. (2015) Chapter One - Purification of His-Tagged Proteins. In Lorsch,J.R. (ed), *Methods in Enzymology*, Laboratory Methods in Enzymology: Protein Part D. Academic Press, Vol. 559, pp. 1–15.
412. Gent,M.E., Gärtner,S., Gronenborn,A.M., Sandulache,R. and Clore,G.M. (1987) Site-directed mutants of the cAMP receptor protein-DNA binding of five mutant proteins. *Protein Eng. Des. Sel.*, **1**, 201–203.
413. Harrison,S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715–719.
414. Saito,M., Iestamantavicius,V., Hess,D. and Matthias,P. (2021) Monitoring Acetylation of the RNA Helicase DDX3X, a Protein Critical for Formation of Stress Granules. *Methods Mol. Biol.*, **2209**, 217–234.
415. Choudhary,C., Kumar,C., Gnad,F., Nielsen,M.L., Rehman,M., Walther,T.C., Olsen,J.V. and Mann,M. (2009) Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science*, **325**, 834–840.
416. Ren,J., Sang,Y., Tan,Y., Tao,J., Ni,J., Liu,S., Fan,X., Zhao,W., Lu,J., Wu,W., *et al.* (2016) Acetylation of Lysine 201 Inhibits the DNA-Binding Ability of PhoP to Regulate Salmonella Virulence. *PLOS Pathog.*, **12**, e1005458.
417. Sorokin,E.P., Cheng,B., Rathi,S., Aedo,S.J., Abrenica,M.V. and Tse-Dinh,Y.-C. (2008) Inhibition of Mg²⁺ binding and DNA religation by bacterial topoisomerase I via introduction of an additional positive charge into the active site region. *Nucleic Acids Res.*, **36**, 4788–4796.
418. Mills,M., Tse-Dinh,Y.-C. and Neuman,K.C. (2018) Direct observation of topoisomerase IA gate dynamics. *Nat. Struct. Mol. Biol.*, **25**, 1111–1118.
419. Aravind,L., Leipe,D.D. and Koonin,E.V. (1998) Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.*, **26**, 4205–4213.
420. Spakman,D., Bakx,J.A.M., Biebricher,A.S., Peterman,E.J.G., Wuite,G.J.L. and King,G.A. (2021) Unravelling the mechanisms of Type 1A

References and additional information

- topoisomerases using single-molecule approaches. *Nucleic Acids Res.*, **49**, 5470–5492.
421. Allemand,F., Mathy,N., Brechemier-Baey,D. and Condon,C. (2005) The 5S rRNA maturase, ribonuclease M5, is a Toprim domain family member. *Nucleic Acids Res.*, **33**, 4368–4376.
422. Raftopoulou,C., Roumelioti,F.-M., Dragona,E., Gimelli,S., Sloan-Béna,F., Gorgoulis,V., Antonarakis,S.E. and Gagos,S. (2020) Karyotypic Flexibility of the Complex Cancer Genome and the Role of Polyploidization in Maintenance of Structural Integrity of Cancer Chromosomes. *Cancers*, **12**, 591.
423. Cheng,B., Annamalai,T., Sorokin,E., Abrenica,M., Aedo,S. and Tse-Dinh,Y.-C. (2009) Asp-to-Asn Substitution at the First Position of the DxD TOPRIM Motif of Recombinant Bacterial Topoisomerase I Is Extremely Lethal to *E. coli*. *J. Mol. Biol.*, **385**, 558–567.
424. Skrzypek,M.S., Binkley,J., Binkley,G., Miyasato,S.R., Simison,M. and Sherlock,G. (2017) The *Candida* Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.*, **45**, D592–D596.
425. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
426. Mirdita,M., Schütze,K., Moriwaki,Y., Heo,L., Ovchinnikov,S. and Steinegger,M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods*, **19**, 679–682.
427. Williams,T.A., Foster,P.G., Nye,T.M.W., Cox,C.J. and Embley,T.M. (2012) A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc. Biol. Sci.*, **279**, 4870–4879.
428. Russell,R.J., Gerike,U., Danson,M.J., Hough,D.W. and Taylor,G.L. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, **6**, 351–361.
429. Kumar,S., Ma,B., Tsai,C.J., Sinha,N. and Nussinov,R. (2000) Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci. Publ. Protein Soc.*, **9**, 10–19.
430. Oren,A. (2013) Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.*, **4**.
431. Zhao,X. and Jain,C. (2011) DEAD-Box Proteins from *Escherichia coli* Exhibit Multiple ATP-Independent Activities ν . *J. Bacteriol.*, **193**, 2236–2241.
432. Chen,Y., Potratz,J.P., Tijerina,P., Del Campo,M., Lambowitz,A.M. and Russell,R. (2008) DEAD-box proteins can completely separate an RNA

References and additional information

- duplex using a single ATP. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 20203–20208.
433. Campo, M.D., Tijerina, P., Bhaskaran, H., Mohr, S., Yang, Q., Jankowsky, E., Russell, R. and Lambowitz, A.M. (2007) Do DEAD-Box Proteins Promote Group II Intron Splicing Without Unwinding RNA? *Mol. Cell*, **28**, 159–166.
434. García, M.J., Ríos, G., Ali, R., Bellés, J.M. and Serrano, R. (1997) Comparative physiology of salt tolerance in *Candida tropicalis* and *Saccharomyces cerevisiae*. *Microbiol. Read. Engl.*, **143 (Pt 4)**, 1125–1131.

PIP reflective statement

PIP host: Sygnature Discovery, Nottingham, UK

I was fortunate to have previously worked within industry as a research scientist prior to commencing my PhD: however, my role within my previous company was largely focused on biomarker based diagnostics of disease and I had always been eager to obtain some experience within drug discovery. Having been interested in Sygnature discovery since attending a careers event hosted by the university during my masters, I applied for and was grateful to be offered the opportunity to contribute to their research.

The goals of my project were:

- The optimisation of the extraction and lysis of mouse brain samples.
- The development of an automated western blot analysis of biomarkers of neurodegeneration within brain and serum samples.
- Comparisons between total tau and phosphorylated tau levels.

By the end of my placement, working alongside fellow BBSRC DTP student Hannah Lockington and our Sygnature supervisor Tatiana Rosenstock, we had successfully tested several regions of mouse brain for both tTau and several pTau targets. This conferred and our results were later presented at the Brain and Neuroscience Advances conference in April 2023. Early on in my placement I was also fortunate to obtain some experience in mitochondrial modulators. I also gained the opportunity to bring in my own expertise when we were troubleshooting our serum assays - having worked in immunodiagnostics of human serum previously, I had a unique perspective.

The placement offered me a new and fresh insight into drug discovery and helped reignite my interest within a potential career in industry. I was also fascinated to gain some experience and knowledge in a new field, having not previously had any experience within the area of neuroscience. I was particularly impressed with the client-based nature of Sygnatures work, seeing clear examples of the opportunity to gain some variety and also lead projects. I also developed skills that were relevant to my , including in cell lysis, western blot sample preparation and toxicity assays that would later come in useful

References and additional information

during my cell culture work on the PhD, as well as the efficient and methodical upkeep of electronic lab-books.

Whilst I have currently accepted a contract with the University of Cardiff for a post-doctoral research position, I feel confident that the experience I gained during my placement at Sygnature will continue to be an asset going forward with my career and may one day result in a return to the company.

References and additional information

PIP Project Form

To the student

You must send this to the prospective PIP host once an in principle agreement has been made to host you on a placement there. The PIP host should return the form to you and you should agree the details. You must then return the form to _ so that contracts can be issued.

PLEASE RETURN THIS FORM TO

AT LEAST 6 WEEKS PRIOR TO YOUR INTENDED START DATE. FUNDING AND EXPENSES MUST BE APPROVED AT THE PRE-PIP MEETING. DO NOT PAY OUT ANY MONEY BEFORE IT HAS BEEN APPROVED.

To the host organisation

Thank you for your interest in hosting a Nottingham postgraduate student at your organisation. You will be sent a standard placements contract via email within the next two weeks.

Internship Dates (start and end)	04 July 22 – 30 th September 2022
Organisation name	Sygnature Discovery Ltd
Organisation address	Sygnature Discovery Ltd Laurus Building BioCity Pennyfoot Lane Nottingham NG1 1GF United Kingdom
Key contact name	Dr. Tatiana Rosenstock
Key contact email	T.Rosenstock@sygnaturediscovery.com
Key contact telephone number	0115 941 5401

References and additional information

Student name	Ashley Parkes
PIP project title	Neurite Outgrowth and the Effect of Neuroplastic-like Compounds

References and additional information

PIP project summary	<p>The visiting PIP student will work within the bioscience department at Sygnature for a period of 3 months (July – September 2022), and the research project area aligned to on-going internal neuroscience-based drug discovery research. This will focus primarily on the area of neuropsychiatry examining the effect of neuroplastic-inducing compounds on modulating neuron architecture.</p> <p>Neuroplasticity is a key measure of drug efficacy in neuroscience and relates to the ability of neuronal networks in the brain to change through growth and reorganisation (to support learning/memory, environmental effects, etc). Currently, there is much interest in identifying novel drugs that promote neuroplastic changes in the CNS for the treatment of both neurodegenerative and neuropsychiatry diseases. Traditionally, neuroplastic effects are measured by the use of complex, invasive imaging techniques <i>in vivo</i>. As part of Sygnature’s commitment to supporting CNS-drug discovery, we are interested in building a simplified <i>in vitro</i> cell model to detect and quantify neuroplastic structural changes in neurons, which will support the rapid testing of novel drug-like compounds, in early-stage drug discovery projects, that hopefully is representative and predictive of the effects observed in more complex <i>in vivo</i> studies.</p> <p>The student will work with a cellular model of neurite outgrowth, employing differentiated PC-12 cells exposed to nerve growth factor (a well-established promotor of neurite outgrowth). The objective is to evaluate multiple different biological parameters aligned to neurite outgrowth upon treatment of PC-12 cells with a selection of different compound classes (covering multiple known neuroplastic mechanisms of action), through the use and application of IncuCyte-based kinetic, live cell imaging instrumentation.</p>
---------------------	--

References and additional information

<p>PIP project outcomes</p>	<p>The student will obtain an excellent insight into laboratory-based research within a commercial drug discovery environment, together with the opportunity to be trained in new cutting-edge assay techniques and the application of these in the discovery of new medicines. At the end of the placement, the student will be invited to present a summary of the project (and placement) to the wider bioscience department.</p> <p>The overall theme of the PIP student project is aligned with on-going CNS drug discovery research currently being undertaken at Sygnature, and the scientific data generated will be used to further our understanding of neurodegenerative/neuropsychiatric disorders.</p>								
<p>PIP project line manager</p>	<p>Dr. Tatiana Rosenstock</p>								
<p>Out of pocket expenses covered by the PIP host:</p> <p>If the PIP host is based outside of Nottingham (20 mile radius), please state which of the students out pocket expenses incurred, will be covered by your PIP host.</p>	<table border="0"> <tr> <td><input type="checkbox"/> Accommodation</td> <td>Cost £.....</td> </tr> <tr> <td><input type="checkbox"/> Travel (International)</td> <td>Cost £.....</td> </tr> <tr> <td><input type="checkbox"/> Travel (National)</td> <td>Cost £.....</td> </tr> <tr> <td><input type="checkbox"/> Other costs</td> <td>Cost £.....</td> </tr> </table> <p>(Please provide details below)</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>	<input type="checkbox"/> Accommodation	Cost £.....	<input type="checkbox"/> Travel (International)	Cost £.....	<input type="checkbox"/> Travel (National)	Cost £.....	<input type="checkbox"/> Other costs	Cost £.....
<input type="checkbox"/> Accommodation	Cost £.....								
<input type="checkbox"/> Travel (International)	Cost £.....								
<input type="checkbox"/> Travel (National)	Cost £.....								
<input type="checkbox"/> Other costs	Cost £.....								

References and additional information

<p>Out of pocket expenses that you request support for from the University of Nottingham:</p> <p>If any out of pocket expenses will be incurred and not covered by the PIP host, please give details here.</p>	<input type="checkbox"/> Accommodation	Cost £.....
	<input type="checkbox"/> Travel (International)	Cost £.....
	<input checked="" type="checkbox"/> Travel (National)	Cost £153
	<input type="checkbox"/> Other costs	Cost £.....
	(Please provide details below)	
	
	
	
	
	
<p>FUNDING MUST BE APPROVED BY THE POSTGRADUATE FUNDING MANAGER PRIOR TO THE COMMENCEMENT OF THE PLACEMENT. Out of pocket estimated expenses must be supported with screenshots or a breakdown of the costs involved.</p>		

Outreach and Public Engagement Reflection Form

Student name:	Ashley Parkes
Public engagement/outreach activity:	Nottingham Festival of Science and Curiosity
Date of activity:	17 th February 2020 – 18 th February 2020
Summary of activity:	<p>Science busking to a large number of children (mostly aged 3-10 years old) with a variety of simple and fun experiments. This included explaining the science behind the experiments and also getting the parents involved.</p> <p>Guiding people around the festivals events and answering any questions they had.</p> <p>Helping out with media including recordings and also obtaining consent for pictures to be used in promotional material.</p>
Reflection on experience (250-400 words): <i>What did you enjoy? What did you find challenging? What would you have done differently?</i>	<p>One aspect I particularly enjoyed was seeing the childrens faces light up as they performed each experiment. It was also enjoyable to hear their interpretations of what they were experiencing, with some descriptions of the wire experiment ranging from 'soft' to 'hands feeling like ice cream'. It was also enjoyable trying to engage the more shy children into the experiments, particularly when successful. Over time I became more comfortable in engaging the kids and was much more confident and adept at performing each type of busking, elaborating on my explanations of the experiments with this.</p> <p>The second day at Rushcliffe arena was much more challenging as there were more events and a lot more people on this</p>

References and additional information

	<p>day: at one point, when the other volunteers were on lunch, I had a large crowd of people (20-30) around the busking table and had to quickly manage each experiment to keep them engaged. It was also a lot faster-paced and there were no breaks in the people attending the stall.</p> <p>One thing I would have liked to have done differently is to have volunteered over the weekend as well, particularly as my rotation lab had a workshop in the Victoria centre, but due to having made some plans this was not possible.</p> <p>Another thing I could have done differently is simplify things a bit better, as whilst most of the children understood how I described it, some of the younger ones had trouble understanding references to vibrations, etc. On occasion more complicated words slipped out too which is something that became less common as I gained more confidence.</p>			
<p>How many people did you engage with as part of your activity?</p>				<p>50+</p>
<p>What age ranges engaged with your activity?</p>	<p>0-11</p>			
<p>Have you been involved in any other outreach or public engagement activity over the last year? If so, please list.</p>	<p>Careers talk at Birmingham City University in March 2019 whilst working in industry. An additional talk to students from Aston university in May 2019.</p>			

Please attach any feedback received as part of your public engagement activity when you submit this form.