Analysis of catalytic and non-catalytic regions of the human DNA repair helicase HelQ

A thesis submitted to the University of Nottingham for the degree Doctor of Philosophy.

Tabitha R. Jenkins, B.Sc.

October 2020

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

BBSRC
bioscience for the future

Acknowledgements

I would not have been able to complete this project without the support of many people. Firstly, thank you to my supervisor Ed Bolt and second supervisor Panos Soultanas for the challenging project I have had over the last four years. Thank you for giving me many opportunities to develop as a scientist, learn new skills and push HelQ to the limit. I really appreciate the guidance, availability and support. I also would like to thank my funding body BBSRC that has supported my project and the University of Nottingham. Thank you to my collaborators, particularly Denis Ptchelkine for all his help with the structural biology and the lifesaving advice when it came to HelQ cloning. I would also like to thank Chris Cooper for his help with N-HelQ bioinformatics, Gemma Harris for her help with AUC and ITC methods, Vincenzo Taresco for help with SEC MALS and Michelle Hawkins with the replication assays. I learnt so many techniques and skills working with these amazing collaborators so thank you.

I really appreciate all the help from my lab friends. In particular to Ryan Buckley for helping me to push back at HelQ when things got tricky, for chatting through crazy lab ideas and teaching me at the start (and the crème eggs). I also want to thank Hannah Betts for continuing the mantel of HelQ, while it is a challenging protein it is fascinating, and I wouldn't leave it with anyone else. Shout out to the best fellow DTP student Amelia Townley, we got through this together and I wouldn't have had it any other way. Also, to Harry for getting me through lock down and He, Liam and Tom for all the coffee breaks and cake. I couldn't have asked for a better lab and I wish them luck in the future.

A massive thank you for all the support from my family Mark, Susan, Eloise and Ozzy; you have had to listen to me talk HelQ for years and COVID made writing that little (lot) more stressful, but I couldn't have done it without your help. The biggest thank you to James Grey for listening and understanding through every second of my PhD and reading through my thesis multiple times - I owe you!

Contents Page No.

List of Figures

List of Tables

List of Abbreviations

A	Adenine
ADP	Adenosine diphosphate
AGF	Analytical gel filtration
AP	Apurinic sites
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related
AUC	Analytical ultracentrifugation
BER	Base excision repair
BIR	Break induced replication
BLM	Bloom syndrome protein
BN-PAGE	Blue native polyacrylamide gel electrophoresis
BSA	Bovine serum albumin
Bp	Base pair
BRCA	Breast cancer associated gene
CDK	Cyclin dependent kinases
CRISPR	Clustered Regularly interspaced short palindromic repeats
Cas9	CRISPR associated protein 9
CryoEM	Cryogenic electron microscopy
C	Cytosine
C-HelQ	C-terminal domain of HelQ fragment
DBDs	DNA binding domains
DDR	DNA damage response
dHJ	double Holliday junction
D-loop	Deoxyribonucleic acid loop

DEAE	Diethylaminoethanol
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
dsDNA	Double stranded DNA
DSB	Double strand break
DSBR	Double strand break repair
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
FA	Fanconi anaemia
FBH1	F-box helicase 1
FBS	Foetal bovine serum
G	Guanine
G4	G quadruplex DNA
G1/2	Gap 1/2
GG-NER	global genome nucleotide excision repair
HGT	Horizontal gene transfer
HJ	Holliday junction
HR	Homologous recombination
ICL	Inter-strand crosslink
ITC	Isothermal titration calorimetry
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1thiogalactopyranoside
IR	Ionising radiation
J6 HJ	Holliday junction
KO	Knock out
LB	Luria Bertani

M	Mitosis
M	Molar (mmol/L)
MBP	Maltose-binding protein
MCM	Mini-chromosome maintenance
MeP	Methyl phosphonate
MMC	Mitomycin C
MMR	Mismatch repair
MMEJ	Microhomology mediated end joining
MST	Microscale thermophoresis
MW	Molecular weight
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
N-HelQ	N-terminal domain of HelQ fragment
NPF	Nucleoprotein filament
NTA	Nitrilotriacetic acid
OB fold	Oligonucleotide binding fold
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PHYRE2	Protein homology recognition engine version 2
PTMs	Post translational modifications
R-loop	Ribonucleic acid loop
ROS	Reactive oxygen species
RPA	Replication Protein A
RPM	Rotations per minute

RNA	Ribonucleic acid
RT	Room temperature
s	Seconds
S-1/2	Phosphorothioate-1/2
SAXS	Small Angle X-ray scattering
SCEs	Sister chromatid exchanges
SDS	Sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing
SDW	Sterile distilled water
SEC	Size exclusion chromatography
SEC MALS	SEC and multi angle light scattering
Sf cells	*Spodoptera frugiperda* cells
SF1	Superfamily 1
SF2	Superfamily 2
sgRNA	Single guide RNA
siRNA	short interfering DNA
SOC	Super optimal broth with catabolite repression
SUMO	Small ubiquitin-like modifier
SIM	SUMO-interacting motif
ssDNA	Single-stranded DNA
SSA	Single-strand annealing
SSB	Single-strand DNA binding protein
S-phase	Synthesis phase
T	Thymine
TBE	Tris-aminomethane borate ethylenediaminetetraacetic acid
TC-NER	Transcription coupled nucleotide excision repair
TCEP	Tris(2-carboxyethyl)phosphine

TEMED	Tetramethylethylenediamine
TLS	Translesion synthesis
TopBP1	Topoisomerase binding protein 1
Tris	Trisaminomethane
UHRF1	Ubiquitin-like, PHD and ring-finger containing 1 protein
UV	Ultraviolet
WHD	Winged helix domain
WRN	Werner syndrome ATP-dependent helicase

Abstract

Maintaining genome stability is essential to support DNA replication for all life to continue. Multiple systems are in place to ensure this occurs by scrutinising replication as it happens, detecting and repairing changes to DNA caused by damaging agents and removing physical blocks. Repair during replication is essential to ensure errors are not replicated, homologous recombination (HR) is one example; repairing DNA errors by using the homologous template from the sister chromosome. While these pathways exist to maintain genome integrity it is essential that they are tightly regulated to prevent unnecessary activation or downstream impacts on the genome. Therefore, a network of proteins are involved in controlling genome maintenance. Metazoan HelQ DNA helicase is a single-stranded DNA ATPase with 3′ to 5′ translocase activity that unwinds forked DNA structures. It is hypothesised that HelQ activity is crucial in promoting DNA replication and repair through the regulation of HR. However, the mechanism is unknown. While HelQ has shown to co-elute with essential repair proteins including RPA, the Rad51 paralogues and ATR and has been linked to cancer and repair-related diseases, little is known about how HelQ behaves in nature. Here, I was able to generate yields of HelQ and HelQ fragments for biochemical analysis of the recombinant proteins. We show that HelQ forms active dimers that unwind DNA fork substrates but no other intermediate DNA structures. Functional analysis of a catalytically active and non-catalytically active region of HelQ aids in the dissection of HelQ structure and function. I also report that HelQ interacts with the single-stranded DNA binding protein RPA *in vitro* and the N-terminal ORFan domain is able to displace RPA from DNA through an unknown mechanism. This leads us to hypothesis a model of HelQ activity and translocation on DNA.

Chapter 1: Background to the Research

1.1 Project context.

The subject of this thesis is the human helicase HelQ in the context of replication-coupled DNA repair. Maintaining genome integrity at replication forks is critical to ensure the efficacious transfer of genetic material through generations – successfully completed replication of a genome is a prerequisite for cell division in all domains of life. Genome maintenance is achieved by a complex network of proteins and pathways of DNA replication, transcription and repair. However, DNA damage can be missed or errors in replication can lead to limitations in genome maintenance and therefore tight control is required to prevent instability arising. DNA repair has evolved to occur alongside DNA replication, gene transcription and protein synthesis. It has been proposed that HelQ acts in replication-coupled DNA repair possibly by controlling the extent of how homologous recombination occurs.[1] To assess this in context I review processes of DNA replication, genome maintenance and relevant DNA damage repair. However, in order to understand the extensive networks of genome maintenance we must first understand the molecule it is in place to protect, DNA.

1.2 The structure of deoxyribonucleic acid (DNA).

DNA is the hereditary material in eukaryotes, bacteria, archaea and some viruses.[2] In mammalian cells, most DNA is located in the cell nucleus called nuclear DNA, but 1% of it is found in the mitochondria coding for 13 proteins required for its specific activity. Mitochondria, double-membrane bound organelles found in most eukaryotes, produce most of the cells supply of chemical energy in the form of adenosine triphosphate (ATP). The information in DNA is stored in four chemical bases: purines which are adenosine (A) and guanine (G) and pyrimidines which are thymidine (T) and cytosine (C) (figure 1.1).[2,3] The sequence of the DNA

bases dictates the order of amino acids for the synthesis of specific proteins, which carry out an organisms functions including catalysing metabolic reactions, replication, responding to stimuli and providing cellular structure.[4] The entirety of human DNA is known as the human genome. However, only 1% of the DNA is made up of protein-coding regions, known as genes. The majority is non-coding DNA commonly referred to as 'junk' DNA.[5] 'Junk' DNA acts as regulatory elements, including promoters, enhancers and silencers, controlling gene activity to determine when and where genes are turned 'on' and 'off'.[5] 'Junk' DNA also encodes binding elements for downstream DNA effects such as transcription activation. Every person has two copies of each protein-coding gene, each known as an allele, one inherited from each parent. Allele variations result in the traits observed across a population. In human DNA, more than 99% of the 3 billion bases that make up the DNA are conserved between people.[6] The vast amount of DNA and the encoded complexity in base sequence, highlights the importance for systems to maintain sequence fidelity. This also requires tight packaging to fit the information into a cell.[7,8]

DNA is made up of nucleotide molecules consisting of a deoxyribose sugar, a nitrogenous base and a phosphate group. These individual molecules of deoxyribonucleoside triphosphates (dNTPs) are assembled into DNA chains. Hydrogen bonds form base pairs between adenine and thymine or cytosine and guanine (figure 1.1).[2,9] The arrangement of the nucleotides into two long strands with central base pairs forms the spiral structure of the double helix. The orientation of base pairs across the helix results in variations in the DNA backbone. The DNA major groove is where the backbone spiral is further apart and the minor groove where they are close together (figure.1.1).[3]

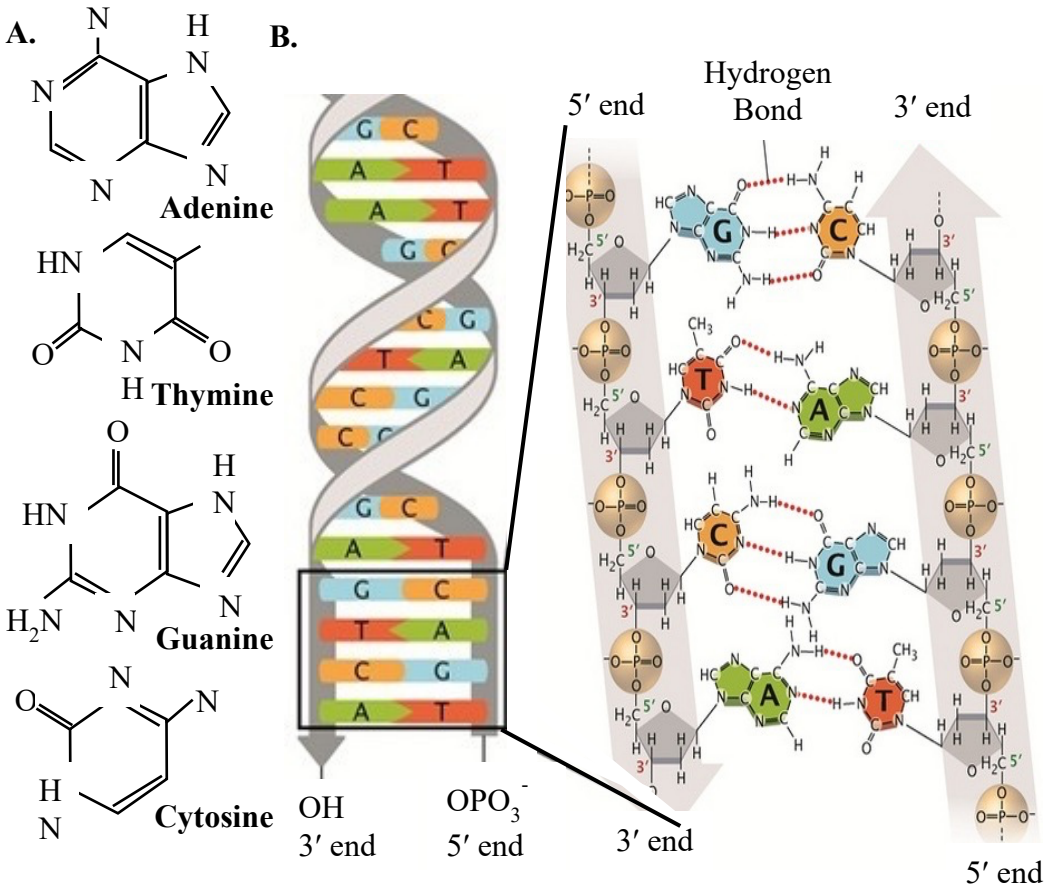


Figure 1.1 Chemical structure of DNA. A. DNA is made up of chains of nucleotides consisting of deoxyribose sugar, a nitrogenous base and a phosphate group. The four bases of DNA are adenine (A), guanine (G), cytosine (C) and thymine (T). **B.** The majority of DNA exists as a double-helical structure of B-form DNA. The schematic shows the phosphodiester backbone of DNA and the base pairs that run through the middle. Also shown are the structures of each base pair found in B-form DNA between A and T and between G and C. Hydrogen bonding between the bases is shown by dotted lines. The DNA strands run anti-parallel. Image adapted from *Nature Education*.[3]

DNA adopts three major structural forms, A, B and Z, which are determined by the base sequence and DNA environment.[10,11] The Watson-Crick B-form, the most commonly known structure, forms as the two strands of DNA, in a right-handed helix, wind around the same axis and are held together by hydrogen bonding between the bases in an anti-parallel conformation (figure 1.1).[3] A-form DNA, also right-handed, contains a deoxyribose molecule in a different conformation resulting in altered topology. In addition, the base pairs are centred over the helical axis in B-form DNA whereas in A-form they are displaced from the central axis and closer

to the major groove. Z-DNA forms when the DNA is in an alternating purine-pyrimidine sequence and the two strands coil in a left-handed helical structure.[10,12] The guanine residues cause distortions requiring a different conformation of the sugar in order for base pairs to form creating a zig-zag pattern in the phosphodiester backbone.[13] DNA can be forced to adopt different DNA conformations in the presence of DNA binding proteins.[10] Other DNA forms including C-DNA, D-DNA and E-DNA but these are uncommon in the human genome.[10]

Most bacterial DNA is contained within a circular molecule called a bacterial chromosome which folds to form an irregular structure known as a nucloid.[14,8] In addition, bacteria have smaller DNA plasmids obtained from neighbouring bacteria via conjugation or from the environment.[15] Plasmids contain genes that are usually unessential for day-to-day survival but help overcome stressful situations. As in bacteria, the absence of internal membranes means archaeal DNA also exists as a single circular strand. This review will focus on the machinery that maintains DNA in the eukaryotic system.

The amount of genetic material in a single eukaryotic cell varies between species. The human genome is 3.1 mega-base pairs which is required to fit into a range of cell sizes from 30 μm^3 (sperm cell) to 4,000,000 μm^3 (oocyte) with a median of ~3,000 μm^3.[16] The problem of fitting so much information into a small space is overcome by tight packaging around specialised proteins called histones. DNA strands are tightly coiled around histones forming nucleosomes which is further condensed around scaffold proteins into chromatin. Histone proteins also signal in DNA replication and transcription (figure 1.2).[9,11] The majority of cells are diploid meaning there is two copies of all genetic information organised as 23 pairs of chromosomes whereas haploid cells contain a single set (figure 1.2).[2,9,11] The paired chromosomes have a central constriction site, called the centromere,

which divides the chromosome into two arm-like structures, the short arm, ‘p’ and a long arm, ‘q’ (figure 1.2).

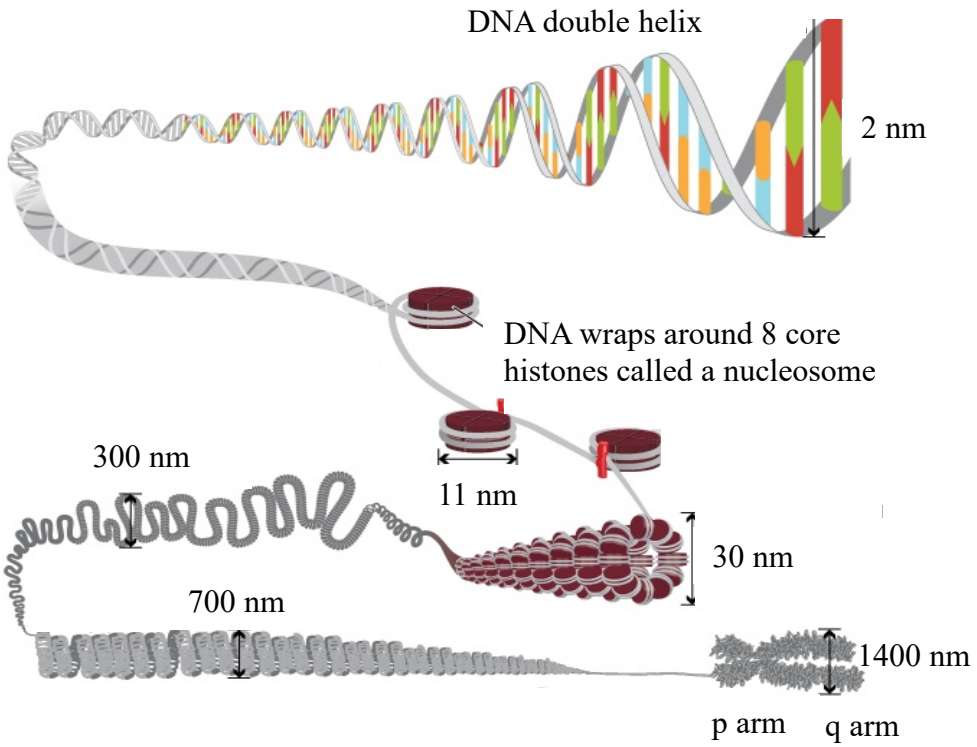


Figure 1.2 The packaging of eukaryotic DNA. Illustration taken and adapted from *Genetics - A conceptual approach*.[9,11] The DNA double helix coils around histone proteins forming nucleosome complexes. DNA is packaged into chromosome structures with a central constriction site.

In addition to the packaging of DNA, access is sometimes required that relies on the DNA being unpacked to allow for single-stranded DNA binding proteins (e.g. RPA see section 1.5) to bind and initiate downstream activities. Unpacking is achieved by chromatin remodeller proteins that unpack segments of DNA by sliding back and forth and replacing the histone proteins. Access is required for DNA replication (section 1.3), transcription and repair to occur and replies on organisation and tight regulation. This is discussed in more detail later in this chapter.

In addition to DNA, ribonucleic acid (RNA), is found in all domains of life.[17] Like DNA, RNA is made up of nucleotides but contains a ribose sugar with an

additional hydroxyl group and replaces thymine with uracil in base usage. Human RNA is found as a single strand folded onto itself and is involved in the catalysis of biological reactions, control of gene expression, sensing and signal transduction within the cell. Cellular organisms use messenger RNA (mRNA) to transfer the information contained in protein-coding genes for translation into specific proteins.[17] Transfer RNA (tRNA) transport amino acids to the ribosome for condensation into proteins.[17]

Bacterial RNA, small RNAs (sRNAs) of 50-500 nucleotides form stem loop structures and regulate cellular mechanisms.[18] sRNAs bind protein targets or mRNA to modify their function and respond to stress e.g. cold shock.[18] Maintaining DNA integrity is essential to prevent downstream impact on RNA.

Cell propagation is the increase of cellular material for multicellular organisms to grow and replace dead cells. This is essential for all living organisms and involves the cell cycle, a complex network, to ensure correct DNA duplication. The cell cycle is an ordered series of processes leading to cell division into two daughter cells but it relies on the ability of DNA to be replicated (see section 1.4).[19] Each strand of DNA can serve as a template for duplicating the base sequence to produce an identical copy for a new cell, this is discussed in the next section specific to eukaryotes. Understanding DNA replication and cellular division allows us to understand the complexity of the system that ensures correct DNA replication. It also allows us to predict failure in the system and the importance for DNA repair mechanisms to maintain DNA integrity.

1.3 Eukaryotic DNA replication.

1.3.1 Replication catalysed by the replisome at replication forks.

DNA replication is the process of nascent DNA strand synthesis from copies of parent DNA strands.[20] The replisome, a protein complex consisting of DNA replication related proteins, is diverse in structure and protein composition but analogous in function between eukaryotes, bacteria and archaea and this review will focus on eukaryotes.[20]

The replisome functions through the major catalytic activities of helicases, DNA polymerases, a primase, ligases and topoisomerases which make up the main enzymes of the replisome.[20,21] DNA helicases initiate replication by catalysing ATP-dependent DNA strand separation creating a fork-like structure known as a replication fork (figure 1.3).[22] See section 1.15 for helicase mechanism.

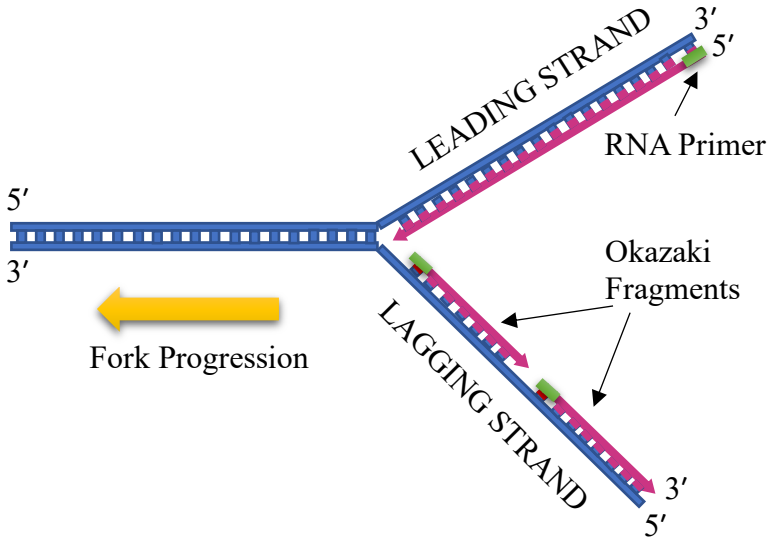


Figure 1.3 The DNA replication fork. The schematic shows the progression of the translocating helicase and DNA synthesis along the leading and lagging strands in a 5′ to 3′ direction. The lagging strand is synthesised in one continuous 5′ to 3′ strand whereas the lagging strand requires disrupted synthesis in the form of Okazaki fragments.

Replication is initiated at two stages, the replicative helicase is loaded at the replication origins, known as replicating licencing, which occurs in late M and G1 stages of the cell cycle (see section 1.4) with the help of the origin recognition

complex (ORC), the eukaryotic initiator and co-factors Cdc6 and Cdt1.[23] A pair of replicative helicases bind at each origin as an inert head-to-head double hexamer and remains bound until the end of G1 creating bi-directional replication complexes. In addition, Ddf4-dependent (DDK) and cyclin-dependent kinases (CDK) (see section 1.4) promote origin firing and helicase activation during S phase where they are assembled in the CMG helicase complex, discussed later.

The regions of exposed single-stranded DNA (ssDNA) in the replication fork are named the 'leading strand' and 'lagging strand' dependent on the anti-parallel nature of DNA duplexes and referring to the direction of replication fork migration in 3D space (figure 1.3).[22] The leading strand template, exposed 3′ to 5′, is used to synthesise continuously a nascent DNA strand in the 5′ to 3′ direction. The opposite lagging strand template is used to synthesise nascent DNA in the opposite direction in space, also 5′ to 3′. This results in the formation of discontinuous smaller 5′ to 3′ fragments known as Okazaki fragments that require ligation.[20]

The CMG complex is the eukaryotic replicative helicase consisting of Cdc45, go-ichi-ni-san (GINS) complex and hexameric Mcm2-7 proteins (figure 1.4).[22,24,25] Mcm2-7 belongs to the AAA^+ ATPase superfamily that use ATP hydrolysis as a motor force.[26] Mcm2-7 unwinds DNA by pulling the double-stranded DNA (dsDNA) through its central pore and excluding the lagging strand along an internal barrier in the hexameric complex.[27]

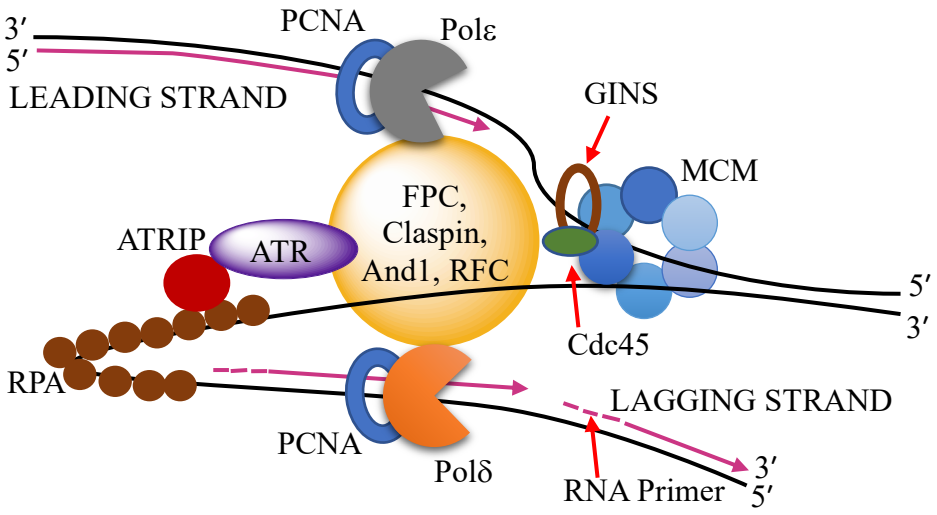


Figure 1.4 Components of the eukaryotic replicative helicase as part of the replisome complex. The replicative complex includes hexameric Mcm2-7 helicase and accessory GINS and Cdc45 factors. Mcm is recruited to replication origins during G_1. The polymerase enzymes for DNA replication are pol ε and pol δ. Other factors involved in replication include the clamp loader protein PCNA, Rcf ATPase enzyme, the ATR/ATRIP checkpoint kinase complex and RPA.

Cdc45 coordinates helicase activity ahead of the replication fork and creates a physical bridge between the helicase and polymerase.[28] Phosphorylation of Mcm4, by the S phase kinase Ddk, allows the stable formation of Cdc45-Mcm2-7 prior to replication and initiates the assembly of the replisome at the origins of replication.[22] The GINS complex (Sld5 (go), Psf1 (ichi), Psf2 (ni) and Psf3 (san)) physically assembles Mcm2-7 and Cdc45.[24,29]

Once DNA is exposed at the replication fork, topoisomase II relaxes the DNA supercoiling created by helicase unwinding by catalysing transient single and double strand breaks, crossing the strands through one another and resealing the gap.[20] Primase, a type of DNA-dependent RNA polymerase, synthesises short RNA sequences that are complementary to a single-stranded region at the replication start point of the template DNA. The RNA binds the DNA and acts as a starting position for DNA replication by the polymerase enzyme.[30] DNA polymerase, primed by an

RNA primer, polymerises a new DNA strand. DNA is synthesised through semi-conservative replication resulting in one parental DNA strand being retained duplexed with one nascent strand of DNA (figure 1.5).[31,32] The structure of DNA polymerases resemble a right hand where the template DNA runs along the palm domain and the thumb and fingers fold around the ssDNA. The polymerase finger domains adopt an open and closed conformation allowing the enzyme to attach deoxyribonucleotides (dNTPs) in the surrounding environment and test them with the corresponding DNA template.[33] The polymerase maintains the DNA in a 90 degree bent state within the enzyme.[34] Once correctly selected, the dNTP is docked and a hydrogen bond spontaneously forms between the base pairs because of the stable conformation created in the closed enzyme structure.[34] The polymerase catalyses the formation of a phosphodiester bond attaching the new dNTP in to the backbone of DNA. The hydroxyl group at the 3′ end of DNA attacks the α phosphate of the incoming dNTP which is condensed into a phosphodiester bond and a pyrophosphate molecule (section 1.3.2).[35] The energy for this process is obtained from the hydrolysis of the dNTP into dNMP and free pyrophosphate, driving the reaction forward. One mechanism proposed for the ability of DNA polymerase to migrate is the sliding of the newly synthesised dsDNA within the hand shaped structure of DNA polymerase instigated by conformational change.[34] The stacking of the newly formed base pairing into the double helix requires a conformational change that may move the DNA through the polymerase enzyme. Therefore, the ‘active’ site of the polymerase will be repositioned along the next available ssDNA region of the template strand.

Exoribonuclease enzymes remove the RNA primers from the newly synthesised DNA fragments and replace them with DNA nucleotides. The ends of

the DNA strands in eukaryotes, known as telomeres, require a specialised DNA polymerase, telomerase, which synthesises repetitive non-coding DNA to prevent the ends of each chromosome from fusing.[36] Bacteria do not have telomeres due to the circular nature of the chromosome. Replication is terminated by polymerase collisions with the next Okazaki fragment causing its dissociation from the template DNA or at the end of the telomeres.[20] The linear nature of chromosomes means polymerase is unable to replicate the whole chromosome in a single strand. This results in loss of DNA at the telomere after each round of replication.[36] The DNA fragments are then ligated together by DNA ligase I which forms a bond between the terminal 5′ phosphate of one strand and the 3′ deoxyribose group on the next (figure 1.4).[37,38]

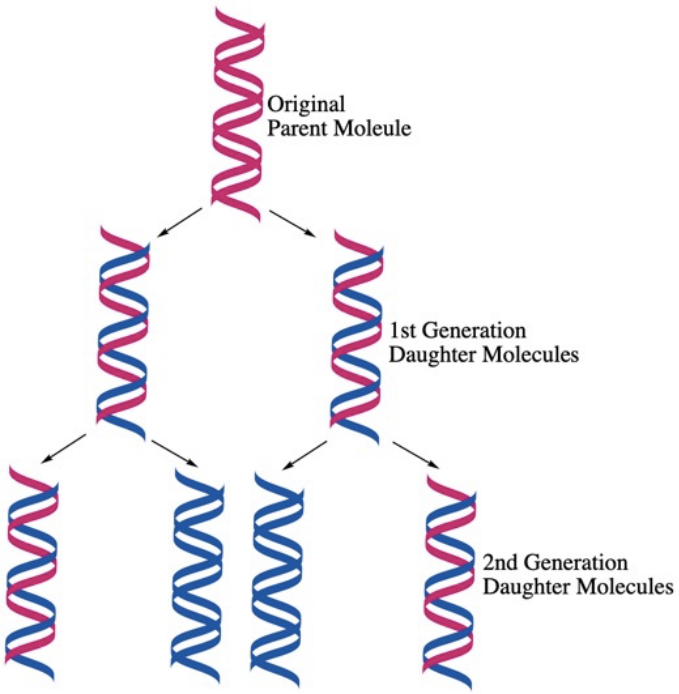


Figure 1.5 The mechanism of DNA replication is semi-conservative. An illustration adapted from Meselson and Stahl.[31] Semi-conservative replication is the process of new daughter DNA molecules being duplicated in cell production. Each new DNA molecule contains one original DNA strand and one nascent strand of DNA.

Replication termination can also happen when two DNA forks migrating in opposite directions collide, this is fork convergence and does not seem to be sequence specific.[39] As the DNA between converging forks becomes too short to super-coil, DNA can become stiff and therefore the release of tension in the DNA relies on the formation of pre-cantenanes. Pre-cantenanes occur when the entire fork rotates clockwise relative to the direction of fork migration. This counteracts the overwinding of the DNA and causes the replicated sister DNA to cross over each other, relieving the DNA stress.[39] As the replication forks encounter each other, the replisome dissociates from the DNA, likely through active disassembly. The removal of CMG is considered key to termination and may involve ubiquitination of Mcm7.[39] After replication, decatenation unlinks the pre-cantenanes behind the forks and the gaps are resolved through gap filling.

The next section will assess in more detail the proteins involved in regulating the replication process to ensure successful and controlled DNA duplication under normal cellular conditions. However, as would be expected in complex systems, replication can falter or become stalled due to damage (section 1.7) resulting in errors that physically impair the chemical composition of DNA and can be detrimental to the cell. Therefore, much like the evolution of DNA maintenance during normal DNA replication, a diverse network of DNA repair proteins have evolved to recognise damage. The restoration of DNA back to its original form will be assessed later in this review.

1.3.2 Polymerase enzymes.

There are three main eukaryotic replicative polymerases, α (1), ε (2) and δ (3) which are part of the B family of DNA polymerases.[20,40] In all DNA and RNA polymerases two magnesium ions (Mg^{2+}) are coordinated by the nucleic acid

substrate and catalytic residues of the polymerase to catalyse DNA polymerisation.[35] Metal ion A interacts with the 3′-hydroxyl on the primer which is hypothesised to lower the strength of the negative charge facilitating its nucleophilic attack on the α phosphate of the incoming dNTP. Metal ion B is proposed to facilitate the leaving of the pyrophosphate molecule.[35]

DNA is at risk of damage that can be caused by internal factors such as replication errors. One of the first lines of defence to DNA base changes is the proof-reading ability of DNA polymerase during replication. DNA base pairing requires specific complementary base insertion in the correct order by DNA polymerase, however, copying errors can occur. The DNA polymerases α (1) and δ (3) contain a 3′ to 5′ exonuclease to remove mismatched bases introduced into the newly synthesised DNA. With a mismatch, the polymerase pauses, transfers the growing strand to the exonuclease subunit to chew back the mismatched base before continuing.[41] While these proof-reading defences are in place to avoid permanent changes, some changes are missed by DNA polymerase and become embedded into the DNA sequence. Further pathways of repair of these changes are discussed in later in this review.

Interactions between DNA helicases and polymerases, the workhorse enzymes of replication, is essential for the successful duplication of DNA. This relationship is tightly regulated by accessory proteins of the replisome. If DNA unwinding occurs too prematurely relative to polymerisation, large regions of DNA become exposed which can lead to the activation of DNA damage signalling and induce DNA repair when not required.[22] While these repair pathways are essential for replication fidelity, premature or unnecessary activation can waste resources or even be detrimental to the cell.

1.3.3 Essential other replication proteins.

The switching of polymerases during replication is controlled by a DNA sliding clamp, *Proliferating cell nuclear antigen* (PCNA), which recruits regulatory factors, and a clamp loader protein, RFC (figure 1.6).[42] PCNA is a homotrimer that strengthens DNA binding of the DNA polymerase, aids incorporation of dNTPs at the active site and releases the newly forming dsDNA. PCNA is loaded at replication forks by a clamp loader complex, a complex of five AAA^+ ATPase enzymes, Rfc1, Rfc2, Rfc3, Rfc4 and Rfc518, collectively known as Replication Factor C (RFC).[43,44] RFC recognises junctions of RNA primer and DNA template and hydrolyses ATP allowing PCNA to load onto DNA.[42,45]

DNA replication is essential to cell proliferation to ensure the genetic material is copied for transfer into new cells. Once the DNA is replicated, the cell with two copies of the entire genome, undergoes division to create two new daughter cells regulated by the cell cycle.

1.4 The eukaryotic cell cycle.

DNA replication is carried out at specific points within the life cycle of a cell. The eukaryotic cell cycle regulates this to ensure appropriate and necessary proliferation.[19] Maintenance and control of this process reduces the likelihood of cellular defects, replication mistakes, DNA damage, injury and associated disease. This is important to maintain genome integrity.

Nascent DNA produced during successful replication is separated into two cells in a process called mitosis. The eukaryotic cell cycle consists of stages of DNA synthesis (S) and mitosis (M), separated by two 'gap' stages (G_1 and G_2) where cells prepare for the next stage (figure 1.6).[19] In addition, a quiescent stage where cells are not actively dividing, known as G_0, is where most cells are found. Cells

respond to growth factors, stimulating entry into the cell cycle from G_0. Cells are committed to DNA replication when they are no longer able to respond to growth factors, this is known as the G_1 restriction point.[46]

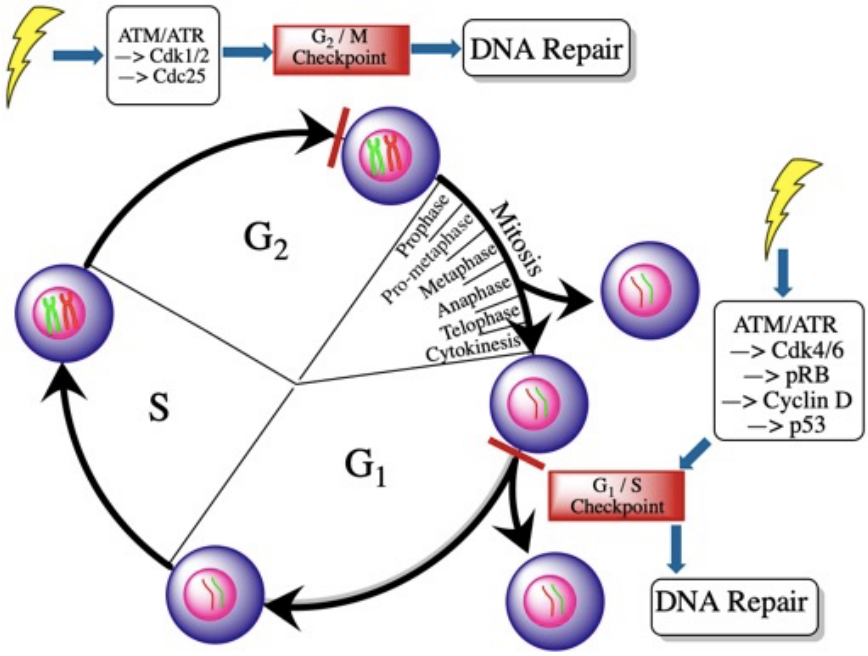


Figure 1.6 The stages of the eukaryotic cell cycle. The cell cycle is essential for cell division to be achieved. There are 4 stages to the cell cycle: Growth 1 (G_1), DNA synthesis (S), growth 2 (G_2) and mitosis (M). Cdks (cyclin dependent kinases) and cyclins regulate progression and control to prevent cellular defects. Checkpoints ensure progression only occurs when required and rely on activity of the ATM/ATR checkpoint kinases. These proteins prevent replication and diverge the cycle into DNA repair if required.

Cyclin dependent serine/threonine kinases (Cdks) and cyclin proteins regulate the progression of the cell through G_1, S, G_2 and M phase.[47] G_1, S and G_2 stages collectively make up interphase which direct the cell into mitosis, which is subdivided into prophase, pro-metaphase, anaphase and telophase.[19]

Cyclins are only expressed at certain times of the cell cycle to interact with Cdks and transport them to the nucleus.[48,49] Progression through the cell cycle is dictated by the assembly of Cdk and cyclin complexes which are important to ensure Cdks are only activated at appropriate stages during cell division. Cdk1 is

phosphorylated on entry into the nucleus by Wee1 to prevent premature exposure to active kinases and inhibit mitosis.[50] Cdc25 phosphatase activates Cdk1 when required. Cdk1 subsequently inactivates Wee1 and Cdc25 preventing further activity.[51] This allows for the controlled activation of elements within the cell cycle at the correct stage and highlights the elaborate network of proteins and phosphorylation events required for cell division.

Activated Cdks phosphorylate target proteins resulting in both activating and inhibitory affects to regulate cell cycle progression. Retinoblastoma protein (pRb) is targeted by Cdk1 for progression from G_1. Upon phosphorylation, pRB dissociates from the E2F family of transcription factors activating transcription of proteins for cell cycle progression.[52–54]

In addition to activating kinases that promote progression of the cell cycle, progression is also controlled by negative regulators to prevent progression, for example the Cdkn1a and Cdkn2a protein families.[19] The Cdkn2a locus, coding for the p16, p15, p18 and p19 proteins, bind Cdk4 and 6 preventing interaction with cyclin proteins for progression from G_1. This prevents the dissociation of pRb keeping E2F inhibited.[55] The Cdkn1a locus, coding for $p21^{Waf1}$, p27 and p57 proteins, inhibit most Cdks with the exception of Cdk1.[55–58]

As well as controlled phosphorylation events, the cell cycle has checkpoints as a second line of defence to ensure appropriate progression.[19] These check points create fail-safe measures against damage during replication (section 1.7).[46] For example, the DNA replication check point ensures replication is completed fully prior to chromosome segregation during mitosis. In addition, these checkpoints are used to ensure the cell is viable to continue and to prevent the unnecessary use of ATP and other cellular resources. Causes of cell unviability include intrinsically,

due to cell size, or extrinsically, due to cell nutrition or DNA damaging agents and cause cell arrest.[59]

The DNA damage checkpoints prevent cellular progression from G_1 into S phase or G_2 into M phase. The G_1/S phase check point is controlled by the DNA damage checkpoint protein p53.[60] p53 binds damaged DNA to activate the transcription of the p21 proteins resulting in Cdk inhibition. In the case of irreversible DNA damage, and therefore prolonged cell cycle arrest at G_1/S, p53 initiates controlled cell death, apoptosis. ATR and ATM kinases are also important at both DNA damage checkpoints (section 1.6).[60] Improper regulation of the cell cycle as a result of the absence of checkpoint response pathways would lead to abnormal tissue growth, neoplasm, and cancer.[61]

Protein networks described here are essential in the regulation of cell cycle progression and successful DNA replication to maintain genome integrity. The system is regulated by phosphorylation events, protein stoichiometry and negative feedback loops. As well as the fail-safe checkpoints to prevent or signal damage, further pathways of repair are required when DNA changes do arise from DNA damaging agents (section 1.7). In some instances, the tight control of the cell cycle and checkpoints are not sufficient to protect the cell and therefore mechanisms are required to restore the DNA and restart stalled replication (section 1.9). Replication protein A (RPA), discussed next, is a prominent mediator in these protein networks to maintain normal DNA replication and activate repair pathways when required.

1.5 Replication Protein A (RPA).

1.5.1 RPA in replication.

RPA is essential in maintaining genome integrity and plays a central role in DNA replication and cell cycle progression in eukaryotes.[62] RPA acts as a regulator

for DNA and telomere maintenance, checkpoint response, DNA repair and protein recruitment. In times of cellular stress, RPA has a role in repair (section 1.12). Under normal conditions, RPA interacts with DNA polymerase α (1), PCNA and Rfc complexes to coordinate replication.[62]

RPA maintains telomeres during DNA replication by destabilising G-quadruplex (G4) structures.[63,64] G-quadruplexes are four-stranded DNA helical structures assembled by guanine-rich oligonucleotides that protect telomeres because of their high level of stability.[65] These secondary structures inhibit the association of the Shelterin complex subunits Pot1 and Ccq1 which are important in distinguishing chromosome ends from sites of DNA damage. In the absence of Shelterin, telomeres are recognised as damage.[66] RPA binding prevents G-quadruplexes forming at the telomeres of lagging strands allowing the recruitment of the Shelterin complex for telomere protection.[67]

1.5.2 Eukaryotic RPA structure.

RPA is a heterotrimeric complex of RPA70, RPA32 and RPA14. RPA70, the largest subunit, consists of 4 DNA-binding domains (DBD) with multiple oligonucleotide-binding (OB) folds (figure 1.7 A).[68] The OB domain is a 120-residue β-barrel necessary for DNA binding.[69,70] RPA32 has one OB fold and a 60-residue winged-helix domain (WHD) and RPA14 has a single OB fold. These domains are essential for the different interactions RPA makes with DNA and result in different downstream effects.

1.5.3 The binding modes of eukaryotic RPA to DNA.

The binding of RPA to ssDNA determines the role RPA has in regulating DNA. RPA coated ssDNA stabilises the DNA, prevents nuclease attack and the formation

of secondary structures, and recruits secondary proteins.[68] The function of RPA is activated by specific interactions between RPA and DNA.[71]

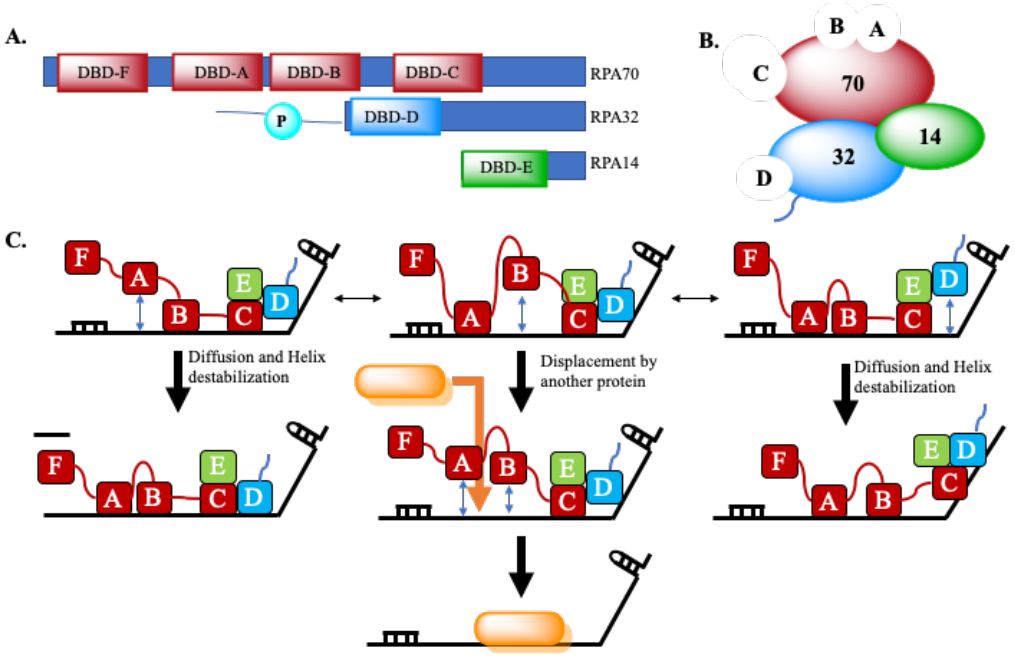


Figure 1.7 DNA binding domains of RPA. A. Trimeric RPA is made up of 70 kDa, 32 kDa and 14 kDa proteins with varying numbers of DNA OB-fold binding domains. **B.** Cartoon of the trimeric RPA and position of DNA interacting domains within each subunit (A-D). **C.** Schematic adapted from Chen *et al.*[67] of the DNA binding states of RPA and how they dictate the activities of RPA in downstream signaling.

RPA binds DNA in two states, an exposed unstable 8-nucleotide binding mode and a stable 30-nucleotide binding mode (figure 1.7 C).[67,72,73] The orientation of the DBD-A and B on DNA, which dictate the conformation of RPA, are different in the two binding modes. The 30-nucleotide mode has a kinked DNA conformation compared to the more extended 8-nucleotide mode. The crystal structure of RPA from the fungus *Ustilago maydis* highlighted the importance of the BC linker in the RPA structure in both the 30-nucleotide binding mode and the transition from 8- to 30-nucleotide binding (figure 1.8).[74–76] Some crystal structures for human trimeric RPA have been solved for the core regions (PB number 1L10)[77] with DNA and between RPA14 and RPA32 subunits (2PQA), the solved structure from *U. maydis*

is used here to show the solved structure of the trimer. The multiple binding modes of RPA on DNA, its flexibility and strong affinity for ssDNA allows rapid association and dissociation with ssDNA allowing for a range of quick responses that directly promote downstream pathways.[78] In addition, the crystal structure of RPA showed the formation of a compact quaternary structure of RPA70 and RPA32 with an extended RPA14 which may be responsible for recruiting and interacting with effector proteins (figure 1.8).[76]

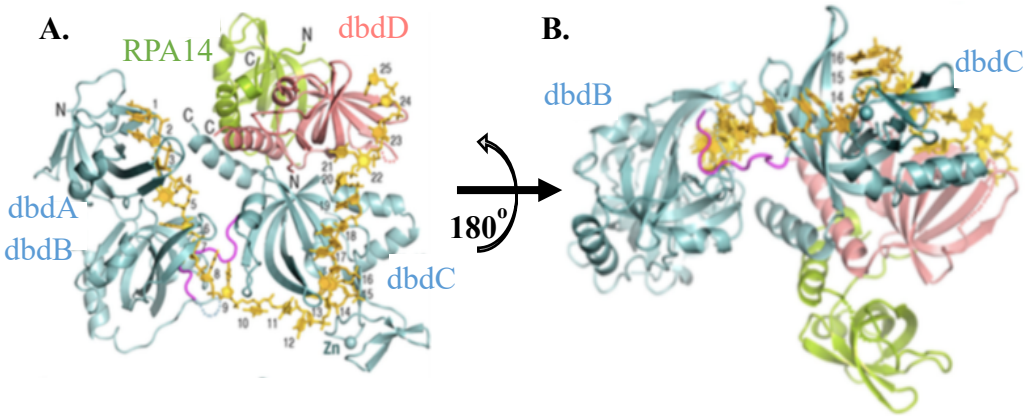


Figure 1.8 RPA domain structure. Figure adapted from Fan *et al*. of the crystal structure of trimeric RPA from the fungus *Ustilago maydis*.[75] Structure of heterotrimeric RPA, RPA70 (blue) with DBD-A, B and C and the 10-residue BC linker (purple). RPA32 (pink) with DBD-D and RPA14 (green) with a single OB fold. Model shows RPA interaction to ssDNA (yellow) and the coordination of zinc (sphere) within RPA70 DBD-C. **A.** Overall structure of the complex and **B.** View rotated 180º about the horizontal axis.

1.6 Signal and response proteins in maintaining the genome.

In addition to RPA, proteins in checkpoint control and genome maintenance ensure repair pathways are the last resort when replication is prohibited.

1.6.1 ATR and ATM kinases.

Cell cycle progression relies on maintenance and control by checkpoints and accessory proteins to ensure essential criteria are met for the next step (section 1.4). This is critical to maintain genome integrity, which if not maintained, would impact

the evolution of life and transmission of genetic material to future generations. These controls ensure appropriate cellular division and prevent instances of permanent changes to the DNA. While these systems are mostly successful, there are rare occasions where problems arise as a result of replication errors or due to damaging agents, both internal and external causing, that are missed and require immediate attention by downstream pathways (section 1.7). The eukaryotic system relies on multiple levels of control in addition to the cell cycle to ensure absolute certainty of appropriate replication and cellular division (section 1.9).

The *ataxia-telangiectasia mutated* (ATM) and *ATM and Rad3-related* (ATR) protein kinases regulate DNA damage response at the cell cycle damage checkpoints to prevent continued propagation in the presence of damaged DNA.[79] ATM and ATR signal for the control of cell cycle transitions, DNA replication, repair and apoptosis. These large kinases have similar structural and functional traits and are part of the *phosphoinositide 3-kinase (PI3K)-related protein kinase* (PIKK) family.[80] ATM and ATR target an overlapping subset of substrates involved in cell cycle arrest and the activation of DNA repair pathways.[81,82]

ATM is active mostly at DNA double-strand breaks (DSBs) and ATR is active throughout DNA synthesis.[83] ATR is localised at sites of DNA damage by RPA-ssDNA signals and regulates replication origin firing and prevent premature mitosis.[83] ATR requires activity of the *ATR-interaction protein* (ATRIP).[84,85] Direct interactions between ATRIP and RPA70 suggests it is ATR-ATRIP that is activated by recognition of ssDNA-RPA.[86,87]

1.6.2 ATR effectors involved in DNA damage check point responses.

ATR-ATRIP is essential in cellular response to DNA damage and activating downstream repair pathways. ATR is activated by *Topoisomerase-binding protein 1* (TopBP1), a *breast cancer-1 C-terminal* (BRCT) domain containing protein, part

of the 9-1-1 complex (Rad9-Hus1-Rad1) (figure 1.9).[88,89] The mechanism of how TopBP1 activates ATR is poorly defined, however, the 9-1-1 and TopBP1 complexes are recruited independently to sites of DNA damage and are both involved in activating ATR.[86,90–94] Co-localisation of ATR-ATRIP and the 9-1-1/TopBP1 complexes result in recruitment activity by ATR to promote activation of repair.[79]

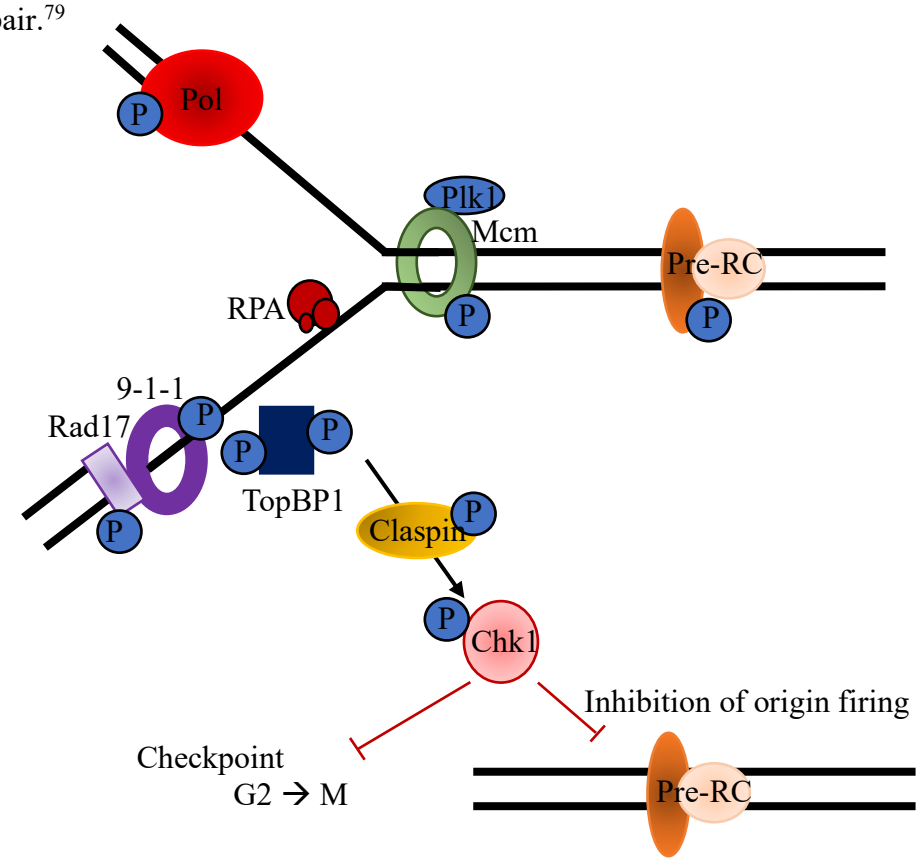


Figure 1.9 ATR signaling in the DNA damage response check point. ATR is activated at stalled replication forks. ATR targets proteins for phosphorylation (blue circle) to regulate replication and signal DNA damage. ATR phosphorylated Chk1 signals DNA damage to the rest of the nucleus and inhibits cell cycle proteins. Other ATR targets include Mcm2-7, RPA, Rad9, TopBP1 and Rad17 (9-1-1) to control replication and stabilise stalled forks. ATR phosphorylated of Mcm2-7 also results in the binding of Polo-like kinases-1 (Plk1) which may result in promoting DNA replication restart at stalled forks. legend= Pre-RC, pre-replicative complex; Pol, polymerase. Schematic adapted from Cimprich *et al.*[86]

ATR phosphorylates the checkpoint kinase Chk1 at damaged DNA stimulating it to signal DNA damage to the rest of the nucleus (figure 1.9).[87,95,96] Claspin, a mediator protein found at replication forks, binds ATR-phosphorylated Rad17 to

sustain Chk1 phosphorylation and amplify the damage signal.[87,97–99] Phosphorylated Chk1 is released from DNA to phosphorylate Cdc25 phosphatase and prevent cell cycle progression. The ATR pathway aims to prevent mitosis during DNA stress and activate downstream repair pathways.[100,101] The ATR-Chk1 signalling slows down or inhibits replication origin firing which is essential in reducing DNA synthesis under DNA-damaged conditions. The activation of repair at stalled replication forks results in the pathways occurring simultaneously. This can remove the need to displace the replisome from DNA saving time and resources during replication.

Furthermore, the ATR-Chk1 pathway is implicated in genome protection against replication forks blocked by R-loops.[102] R-loops are structures that form in the DNA during transcription where the newly synthesised RNA folds back and re-anneals with the template DNA. This results in the displacement of the homologous strand and formation of stable RNA-DNA hybrids.[103,104] R-loop structures require removing to restore the DNA double helix.

ATR stimulates activity of repair proteins at stalled replication forks initiating downstream repair pathways of recombination (section 1.10). Some examples include activation of the *Breast cancer early onset gene 1* (BRCA1), *Werner syndrome ATP-dependent helicase* (WRN) and *Bloom syndrome protein* (BLM) and recruitment of the Fanconi Anaemia (FA) protein, FancD2, to damage foci, initiating repair.[105]

After DNA damage signalling and recruitment of repair proteins, ATR promotes replication fork stability and recovery of replication. The role of ATR in fork stability is not well understood. However, ATR has been shown to interact with components of the replisome including polymerase ε (2) and PCNA which dissociate from the stalled forks in the absence of ATR signalling.[106–109] Therefore,

ATR may reassemble or maintain assembly of the replisome during DNA repair. This would allow for the rapid re-initiation of replication. Furthermore, ATR interacts with the Mcm2-7 complex at stalled forks (figure 1.9).[110–113] ATR phosphorylated Mcm2 binds to *Polo-like kinase-1* (Plk1) promoting DNA replication recovery (figure 1.9).[114] In these instances, ATR bridges replication and repair coordinating their activity.

ATR is essential in stabilising stalled forks, signalling damage and promoting replication restart after collapse. Therefore, it is important to understand the role of ATR in activating DNA repair during replication. The next section looks at the DNA damaging agents that result in changes to the DNA that require these elaborate protein networks to prevent impact the genome integrity.

1.7 DNA damage as a source of genome instability.

The integrity and stability of DNA is essential to ensure normal cellular function which relies on the highly specific sequence within DNA and therefore small changes to these can be detrimental to downstream protein production. Disruptions to the DNA can be fatal because it can result in errors in the genetic code, disruption to DNA replication or gene transcription.[41] Damage that is inaccurately repaired can lead to DNA mutations where changes occur to the specific base sequence which can be replicated and passed onto subsequent cell generations. DNA can become altered through chromosome translocation, where entire portions of a chromosome detach and reattach to other chromosomes. Furthermore, physical changes to DNA include replication slippage which occurs when the DNA strands are denatured, and base pairs reform displaced which results in miss-pairing and ssDNA structures protruding from the dsDNA. Damaging agents can also cause DNA breaks, chemical adducts, protein and DNA crosslinking. DNA damage that occurs in non-replicating cells (e.g. brain and

muscle cells) can cause aging. This review focuses on DNA damaging agents and changes to DNA that result in blocks to replication.

DNA damaging factors that infer change to the DNA can arise internally, for example through errors in replication or respiration in aerobic organisms, or externally, for example by mutagenic chemicals and radiation. If left unchecked, these damaging agents can lead to detrimental mutations disrupting the DNA that causes disease or cell death.[115,41]

There are two main groups of external DNA damaging factors; direct and indirect acting chemical carcinogens.[41] Direct-acting carcinogens are electrophiles that react with the negative charge of oxygen and nitrogen in DNA distorting the base pairing and resulting in nucleotide mismatch during replication. Indirect-acting carcinogens are unreactive, water-soluble compounds which are converted into carcinogens by enzymes found within the body for example detoxifying liver enzymes.[41] Indirect acting polycyclic aromatic hydrocarbons (PAHs) are found in atmospheric pollutants such as oil, coal, cigarette smoke and exhaust fumes.[116] PAHs are chemically modified to form DNA-reacting metabolites. The mutagenic metabolites of PAHs, diol epoxides, quinones and radical PAH cations, bind DNA at specific locations forming bulky DNA adducts.[117]

DNA strand breaks can be caused by both external damaging factors (e.g. UV radiation and chemicals (e.g. peroxides)) and by internal factors (e.g. DNases).[118] Breaks in the phosphodiester backbone can happen across both strands, double-strand breaks (DSBs), across one strand, single-strand breaks (SSBs) or staggered across two strands.[118] These types of changes to the DNA can result in replication stalling because the replisome is unable to migrate.

The simplest form of changes to the DNA structure caused by damaging agents include single base mutations. These are predominantly a result of exposure to X-

rays or UV radiation but can also arise due to internal replication errors. Base switching is either transitional, the like-for-like replacement of pyrimidines and purines or in trans-versional, the substitution of a pyrimidine for a purine or vice versa.[119] Other simple forms of damage include nucleotide insertions and deletions called point mutations. These changes to single nucleotides can lead to frameshifts in the translation of the nucleotide sequence and can disrupt the gene sequence creating complications in protein translation.[119] These DNA changes are predominantly caused by replication errors and detected by proof-reading of the polymerase enzymes (section 1.3.2).

Internal DNA damaging factors can arise as a by-product of normal cellular metabolism in aerobic organisms resulting in the formation of reactive oxygen species (ROS) and subsequent DNA oxidation. ROS produced under normal cellular conditions in the body require removal.[120] In addition, external factors can lead to the formation of ROS and subsequent imbalance of ROS. An imbalance of ROS and the inability to detoxify the ROS, can lead to DNA oxidative stress. DNA oxidation occurs at guanine residues which have high oxidation potential resulting in 8-hydroxyguanosine (8-OHdG).[121] PAHs form benzo(a)pyrene diol epoxide (BPDE), a highly reactive species, that leads to oxidative stress. BPDE covalently binds guanine in DNA producing BPDE adducts leading to tumour formation. DNA damage is considered the most significant consequence of oxidative stress.

DNA can become deaminated, the total removal of individual bases forming apurinic/apyrimidinic (AP) sites which, if left unrepaired, can prevent transcription. This type of hydrolytic damage arises due to ROS imbalances.[122]

DNA-protein crosslinks (DPCs) are another form of damage to DNA where proteins become covalently linked to DNA and create problems of steric blockades to transcription and replication.[123] These blockades can lead to mutations, genomic

instability and cell death. They can be caused both endogenously (e.g. enzymatic intermediates on DNA) or by external carcinogens or chemotherapeutic agents.[123]

Internal and external DNA damaging agents can result in DNA becoming chemically crosslinked. This occurs when two independent reactive groups in a single alkylating molecule react with two bases in the DNA.[124] Crosslinks commonly occur between position N7 of guanine and the N2 of guanine on the opposite DNA strand. DNA crosslinking can occur between nucleotides on the same strand (inter-strand crosslink) or on opposite strands (intra-strand crosslinks) (ICLs) (figure 1.10).[125] ICL inducing agents occur both naturally (e.g. psoralens, mitomycin C (MMC) and nitrous acids) and synthetically (e.g. alcohol and a high fat diet).

While both internal and external DNA damaging agents are unavoidable, the changes they bring to the DNA can be regulated or reversed, either by cell cycle control or through DNA repair pathways. These pathways within, and in addition to, cell cycle control act as barriers to prevent and reverse damage to the DNA. Furthermore, these pathways signal to ensure replication is stalled in order to prevent damaged DNA being replicated. The next section discusses the essential repair pathways that ensure restart and progression of successful DNA replication which have evolved to reverse persistent damage in order to maintain DNA integrity.

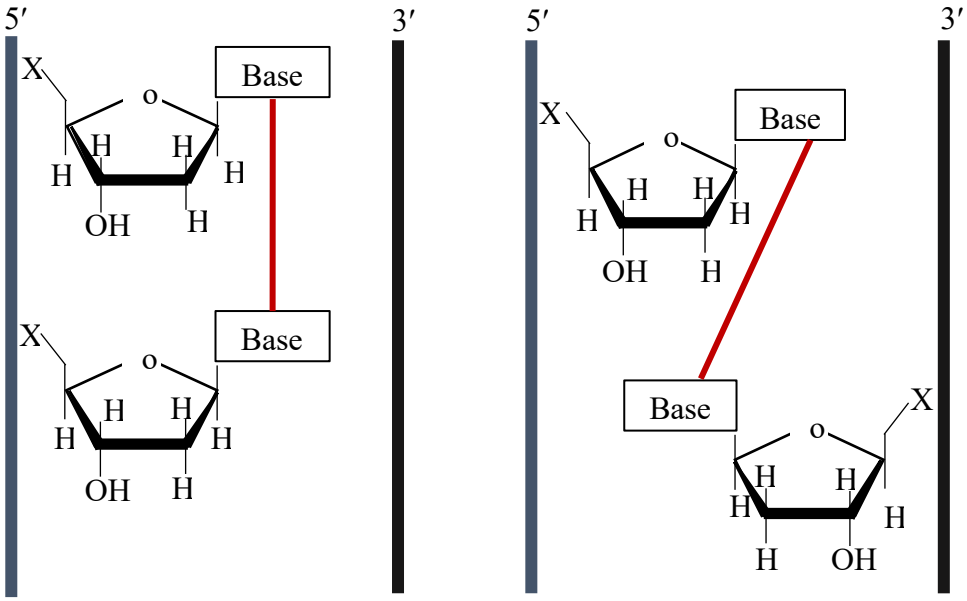


Figure 1.10 Inter- and intra-strand crosslinks (ICLs) in DNA. DNA crosslinking is caused by natural (e.g. DNases) and synthetic (e.g. MMC) crosslinking agents. ICLs are covalent bounds between two guanine bases on the same DNA strand (intra-strand crosslinks) or on opposite DNA strands (inter-strand crosslinks). Accumulation of ICL damage may lead to cell death and cancer.

1.8 Genome maintenance through mechanisms of repair in eukaryotes.

DNA repair is a network of processes where damage to DNA is identified and corrected. As discussed previously, this is essential because cells cannot function if damage disrupts the integrity and accessibility of essential information in the DNA. The type of change created in DNA dictates the type of repair required. For example, changes to ssDNA can be repaired by pathways of base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR). Changes to dsDNA, such as breaks, are repaired by pathways of non-homologous end-joining (NHEJ), the Fanconi anaemia pathway (FA) and homologous recombination (HR) (figure 1.11).[126]

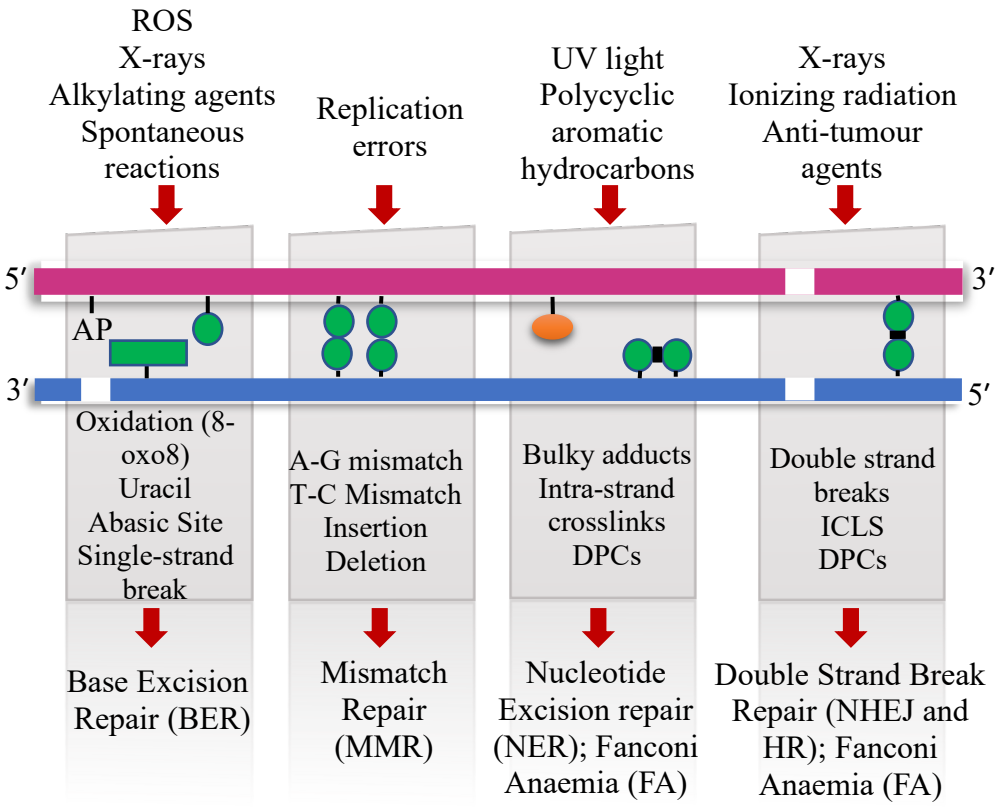


Figure 1.11 Major mechanisms of genome maintenance to restart stalled replication. DNA can be damaged by internal and external DNA damaging agents. These results in different types of damage that change the chemical make-up of DNA, some included here are DNA oxidation, base mismatch, inter-strand crosslinks (ICLs), DNA-protein crosslinking (DPC) and DNA breaks. Repair pathways have evolved to reverse the change to DNA caused by damaging agents and maintain the genome integrity. The major eukaryotic pathways discussed here involved in reversing damage that prevent replication progression include base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER), non-homologous end joining (NHEJ) and homologous recombination (HJ) and the Fanconi Anaemia pathway (FA). Other mechanisms of genome maintenance included the mutagenetic translesion synthesis (TLS) and direct reversal.

Furthermore, there are pathways that directly reverse the damage to DNA. For example, methylation damage where enzymes such as methyl guanine methyl transferase (MGMT) directly reverse the damage without the need for a template. Protein-DNA crosslinks are also repaired by a multitude of pathways. These include proteolytic cleavage of the protein moiety by specific proteases and a combination of NER, FA or HR to repair the gap.[123] In addition, other process that allow for tolerance of DNA damage includes the eukaryotic pathway translesion synthesis

(TLS); however, TLS is mutagenetic and can result in further problems. The mechanisms to maintain the genome are vast, however, this review will focus on those shown in figure 1.11 because of their involvement in repairing blocked replication.

Signals during the cell cycle result in activation of repair pathways. The activation of DNA repair mechanisms require tight regulation because in some instances repair can lead to undesired DNA changes. Furthermore, even with the extensive network of checkpoints and downstream repair pathways, DNA damage can persist in the DNA leading to disease and cell death.

BER repairs individual bases that do not distort the DNA helix structure.[127] The five enzymatic reactions of BER are highly conserved between bacteria and humans. DNA glycosylase recognises and removes the altered base leaving an abasic site. Apurinic endonuclease (APE2) cleaves this site leaving a 3′ OH and 5′ deoxyribose phosphate (5′dRP) termini removed by the lyase activity of DNA polymerase β. The missing base is replaced by DNA polymerase and the resulting nick is repaired by DNA ligase complexed to *X-ray repair cross-complementing protein 1* (XRCCI).[127]

MMR repairs DNA from damage that occurs during replication and is missed by proof reading activity of polymerase enzymes.[128] Differences in the newly synthesised DNA strand are recognised by MMR proteins, PCNA, RFC and RPA. MutSα (Msh2-Msh6 heterodimer) recognises and binds the mismatched base causing ATP-dependent conformational change of PCNA.[129,130] MutLα (Mlh1-Pms2 heterodimer) binds MutSα-DNA and nicks the daughter strand DNA. The mismatched site is excised by exonuclease enzymes and filled by DNA polymerase. MutLα contains a latent endonuclease activity that is activated by PCNA to nick

the DNA in a strand specific manner, excising only the strand with the initial nick.[128]

There are two pathways of NER repair; global genome NER (GG-NER) and transcription-coupled NER (TC-NER).[131] In GG-NER, the entire genome is scanned for helix distortions caused by disruptions in base pairing. GG-NER involves the *xeroderma pigmentosum complementation group* (XPA-G) proteins.[132] XPC, the main damage sensor, binds the lesion and recruits *transcription factor II H* (TFIIH) which verifies the site as effected. DNA helicases (XPB and XPD) with opposite polarities assisted by XPA extend the open DNA configuration which signals the lesion as real. The damage site is removed by structure-specific endonucleases XPF-ERCC1 and XPG and the exposed DNA triggers PCNA, DNA polymerase and ligase to fill the gap. TC-NER recognises lesions in template DNA when RNA polymerase ε (2) is stalled during transcription.[131]

The Fanconi Anaemia (FA) proteins work with other repair proteins to remove ICLs in DNA. FA is a rare genetic syndrome that initially presents with bone marrow failure and can be diagnosed from a hypersensitivity to ICL-inducing agents.[133] The DNA repair FA pathway is comprised of a network of 19 Fanc proteins and many associated proteins.[134] The pathway is activated when ICLs at stalled replication forks are detected during S phase by *ubiquitin-like, PHD and ring-finger containing 1 protein* (UHRF1) and the FancM-Mhf1-Mhf2 (FAAPs) complex. This recruits BRCA1 (see section 1.11) to displace the CMG replicative helicase and allows FancM to activate the ATR-dependent damage checkpoint response.[133] FAAPs also recruits the FancD2-A heterodimer and the FA core complex to execute the 'unhooking' step to produce a DSB for repair.[134] The FA pathway is able to detect and remove ICLs by the combined actions of NER and homologous recombination.

DSBs can be re-joined with little processing by non-homologous end joining (NHEJ) or repaired using homologous sequences by homologous recombination (HR).[126,135] Whether HR or NHEJ is used to repair DSBs is determined by the stage of the cell cycle. HR is used before the cell enters mitosis shortly after DNA replication when sister chromatids are readily available. NHEJ is usually active in DSB repair when the cell is in G_1 of the cell cycle, when the cell is growing but not ready to divide.[136] The packaging of DNA presents a barrier to HR repair and requires activity by ATP-dependent chromatin remodelling proteins to relax the chromatin and access the homologous template. The pathways of HR are discussed in section 1.9.

In addition to the pathways described here, eukaryotic DNA damage tolerance can also occur through translesion synthesis (TLS).[137] While elaborate systems are in place to restore DNA after it has become damaged, at times TLS is required to allow temporary tolerance to damage without mediating repair to ensure survival. TLS is the process where specialised DNA polymerases (Rev1, Polζ, Polκ, Polη, Polι) replicate through DNA lesions.[137] TLS is important in resistance to DNA damage by restarting stalled replication forks and filling in gaps in the DNA. However, this process is highly mutagenetic resulting in the introduction of mutations. The TLS polymerases use the damaged DNA as a template and insert nucleotides opposite the lesion despite the impact modified nucleotides may have on the DNA. This can result in the incorrect insertion of DNA bases in newly synthesised strands.

1.9 DNA homology dependent repair: Homologous recombination.

Stalled replication forks caused by DNA damage prevent the progression of replication. It is vital for stalled replication forks to be rescued to prevent fork

collapse and further damage. Therefore, repair is essential to reverse the damage and allow the replisome to continue. HR is the exchange of nucleotide sequences between identical, homologous, or near-identical DNA sequences.[138] HR is conserved across the three domains of life suggesting it is universal and essential. HR repairs DSBs caused by ionising radiation or DNA damaging chemicals, restarts stalled replication forks and results in horizontal gene transfer (HGT). HR also produces new combinations of DNA sequences during meiosis, the process of making gametes (sperm and egg cells) in mammals. The new DNA combinations provide genetic variation in future generation which enable populations to evolve.

The primary pathways of HR repair are double-strand break repair (DSBR) and synthesis-dependent strand annealing (SDSA) which initially start identically (figure 1.12). Other pathways of HR repair include break-induced replication (BIR) and single-strand annealing (SSA). BIR occurs when the helicase encounters DSBs at the replication fork during DNA replication and contributes to the repair of broken replication forks that have only one end.[139] However, eukaryotes are less dependent on BIR for completion of replication due to the presence of telomerase to solve the chromosome end problem.[139] SSA repairs DSBs between two repeat sequences and does not require a separate homologous sequence. The break is repaired by using the repeat sequence as the template.[140] SSA can result in the loss of the internal damage region.[135] These different repair pathways exist and can occur simultaneously because of the complexity of eukaryotes. With a large range of cell types that make up organisms and the different types of DNA damage that can occur, different pathways are required. Furthermore, these multiple pathways are in place to ensure repair only occurs when absolutely necessary.

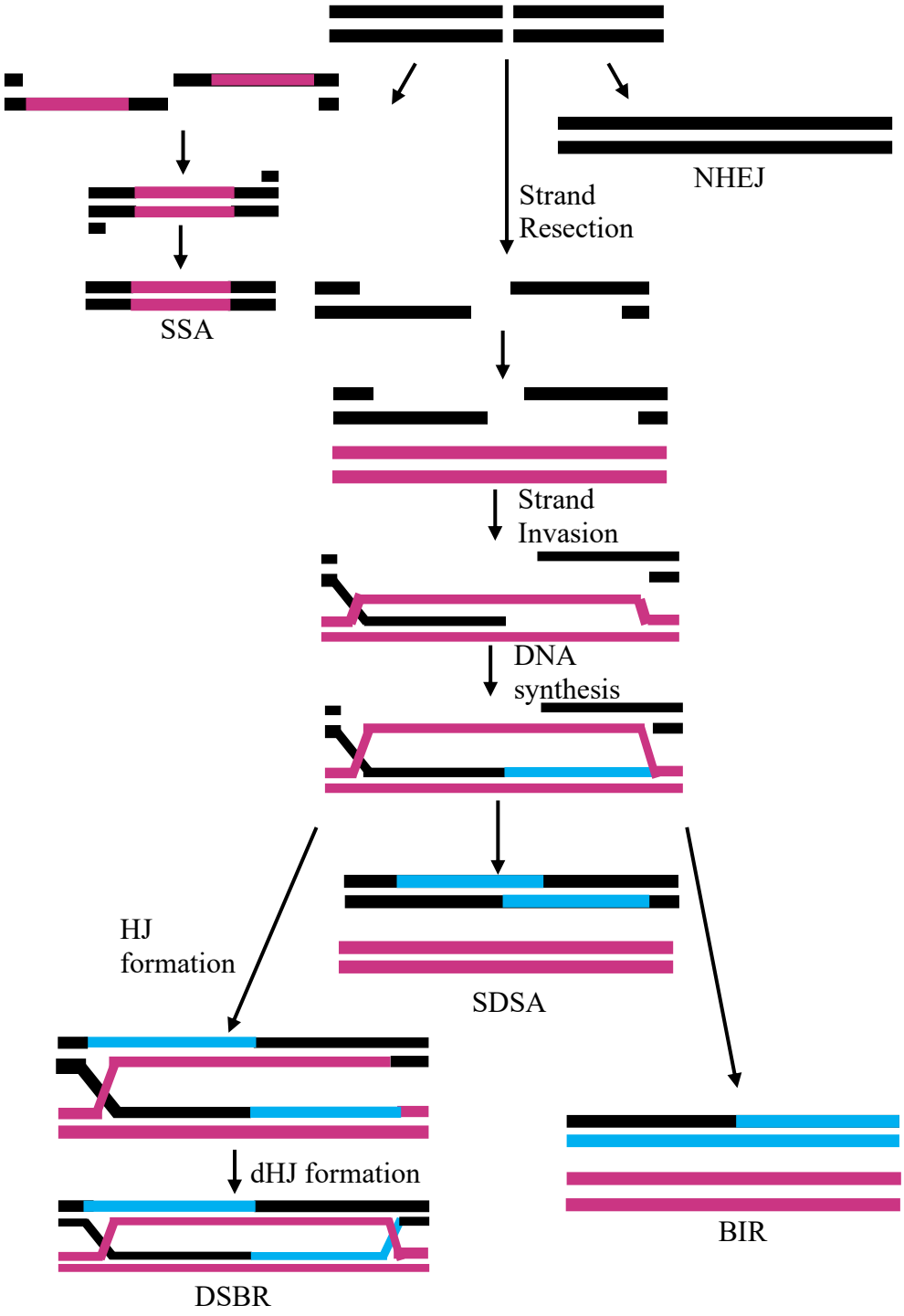


Figure 1.12 The pathways involved in repair of dsDNA breaks. Double strand breaks are repaired by a variety of pathways shown here. The DSB site resected to allow exposure of ssDNA to initiate DNA repair. These pathways include Single-strand annealing (SSA), Non-homologous end joining (NHEJ) and Homologous Recombination (HR) which is further divided into Break-induced repair (BIR), Synthesis-dependent strand annealing (SDSA) and Double-strand break repair (DSBR).

Initiation of HR requires ssDNA which is generated at DSBs by resection of the DNA ends. End resection is tightly regulated and commits the cell to repair by HR instead of NHEJ.[138] HR is made up of three stages: pre-synapsis (strand resection and strand invasion); synapsis (strand exchange and branch migration) and post synapsis (resolution of the Holliday junction (HJ)). DSBR and SDSA follow similar initiation pathways, however, DSBR can lead to crossover products and genetic diversity, focused on in this review (figure 1.13).[138,141]

Events that occur pre-synapsis are essential for the activation of HR. Exposed DNA, resected at sites of DSBs or at stalled replication forks, is bound by RPA that protect the DNA and recruit downstream repair proteins.[62,142] This includes the recruitment of the recombinase enzyme Rad51. The recombinase enzymes are part of the RecA family with the archaeal homologue, RadA and the bacterial homologue RecA. These proteins contain conserved Walker A and B motifs essential for ATPase activity.[142] The assembly of Rad51 on exposed DNA relies the displacement of RPA aided by accessory proteins discussed later.[143] The Rad51 recombinase nucleates ssDNA and polymerises onto the DNA forming a right handed nucleoprotein filament (NPF).[144] The NPF facilities homology searching prior to synapsis. The NPF assembles onto DNA in both orientations in an ATP-dependent manner with each monomer interacting with three nucleotides. ATP hydrolysis results in filament disassembly and therefore mediator proteins are required to stabilise the NPF and prevent premature ATP hydrolysis and filament dissociation. The ssDNA binding protein Rad52 physically interacts with Rad51 recruiting both monomeric Rad51 and ringed structures of Rad51 towards RPA-coated ssDNA to activate the assembly of NPFs.[145] Like Rad52, a Rad55-Rad57 heterodimer physically interacts with Rad51 by binding ssDNA.[146] Five eukaryotic proteins, related in structure to Rad55 and Rad57 known as the Rad51 paralogues,

form complexes to assemble and protect the Rad51 presynaptic filament (section 1.10).[1,147,148] Binding of Rad51 physically changes the chemical structure of DNA expanding its conformation to allow easier homology searching. Once assembled, the NPF invades the homologous DNA template initiating synapsis.

One strand of the homologous DNA template is displaced by the NPF upon strand invasion resulting in a displacement loop (D-loop) forming.[143] The D-loop is a central step for replication fork restart preserving DNA sequence, that may potentially be damaged, from the homologous template. The invading NPF base pairs with the complimentary sequence in the homologous template. Rad54 promotes the exchange of base pairs from the homologous DNA strand by displacing Rad51 to allow DNA synthesis.[149] The D-loop migrates along the template DNA in homology searching to find the correct template DNA in a process termed branch migration. DNA polymerase enzymes use the homologous DNA as a template to prime replication. The ssDNA damage site is extended through replication forming a Holliday junction (HJ) (figure 1.13).[149] The exposed DNA on the opposite side of the damage site base pairs with the displaced homologous template of the D-loop allowing for replication of new DNA on the opposite break strand. This is referred to as second-end capture.[150]

Post-synapsis involves the resolution of HR intermediates which is required to ensure rapid replication restart.[145,151] However, HJ resolution can lead to gene conversion events and therefore is only initiated when absolutely necessary. The newly synthesised original invading strand is ligated to its original strand forming a HJ. Likewise, the second-capture strand in ligated with its original strand forming a double Holliday junction structure (dHJ). HR is resolved separating the dHJ which results in the formation of both cross-over and non-crossover products (figure 1.13). Cross-over products arise when the HJ is cut on the crossing strand

and the other HJ is cut on the non-crossing strand. As mentioned, this genetic exchange is important during mitosis for population variation, however, can be detrimental during repair.

The introduction of crossover events during HR resolution requires mechanisms to be in place to block DSBR during pre-synapsis and synapsis prior to HJ formation and divert repair. SDSA is a favoured alternative pathway of HR during DNA repair because it produces non-crossover products. SDSA occurs prior to the second-end capture step of DSBR but after extension of the invading DNA strand. SDSA dissociates the newly formed duplex DNA preventing the formation of a HJ. The second damage strand is then synthesised using the newly synthesised repaired DNA as a template (figure 1.13). SDSA stems from HR strand invasion and therefore this is an essential check point to ensure extensive HR only occurs when necessary.[145,151]

Break induced repair (BIR) is an alternative pathway to DSBR and SDSA. BIR repairs breaks that have one end and is associated with repair of broken replication forks and replication at telomere ends. While eukaryotes are less reliant on BIR for the replication of telomeres due to the presence of telomerase, BIR is associated with DNA repair.[139] As with other pathways of HR, end resection and single-strand DNA end invasion into the homologous template occurs to form a D-loop. Conservative replication then occurs from the point of invasion giving a newly synthesised DNA strand. Like, DSBR, BIR can result in chromosomal rearrangements and therefore also requires factors to prevent long tract replication of DNA by BIR from occurring.

There are many proteins involved in the regulation of these HR pathways to prevent any pathways going too far to cause irreversible DNA damage and divert repair to a more favourable pathway.

The large number of proteins required for this is to ensure a high level of control is maintained and that the pathways can be regulated by a number of check points; the pathway will only be successful if all proteins involved are active and play their role. We hypothesise that HelQ (section 1.14) is involved in this activity of promoting SDSA or BIR after strand invasion in order to prevent cross-over events. With a broad range of proteins involved in this system, understanding how these proteins are involved in HR can build a picture of how helicases, in particular HelQ, may play a role in the process; discussed in the next section.

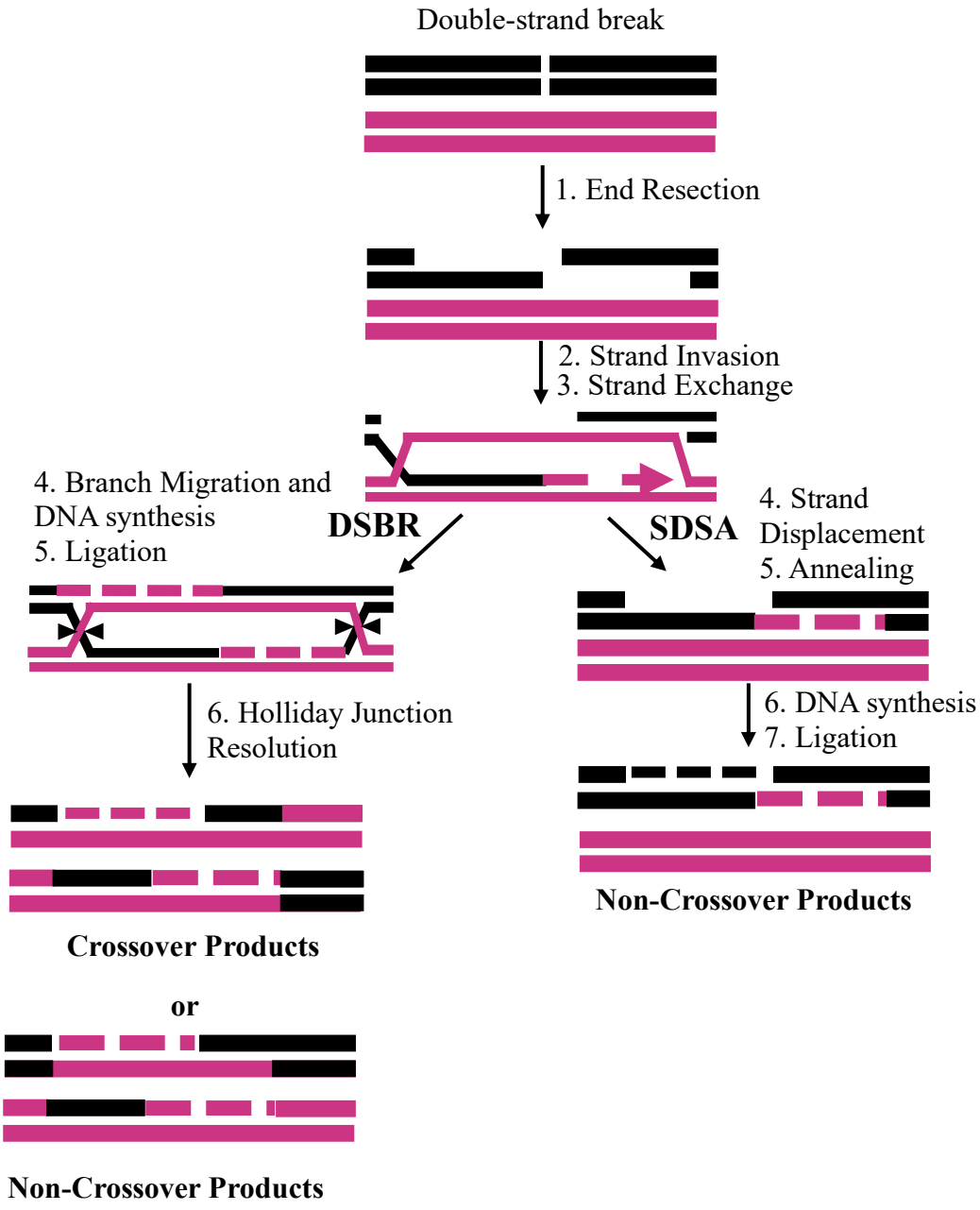


Figure 1.13 The process of homologous recombination (HR). DSBR and SDSA are initiated by end resection exposing ssDNA. ssDNA is coated with Rad51 recombinase forming the NPF that invades the homologous template DNA prior to strand exchange between the damage DNA and homologous template. DSBR is the extensive HR pathway that results in the formation of a double Holliday Junction (dHJ) which, when resolved, can lead to the formation of both non-crossover and crossover products. Mechanisms are in place to divert extensive HR to pathways that don't result in cross-over events such as SDSA. SDSA resolves HR without the formation of HJs and subsequence crossover products.

1.10 Recombination dependent repair proteins.

1.10.1 Rad51 recombinase in recombination repair.

Rad51 is a DNA-dependent ATPase that exists as a monomer and can bind ssDNA and dsDNA.[152] Rad51 assembles as long helical polymers on resected ssDNA to catalyse strand exchange as described previously.[152] The central mechanism of Rad51 in HR is highly conserved. Rad51 and the bacterial orthologue RecA polymerise faster onto ssDNA then dsDNA. DNA flexibility, which increases at areas of DNA damage as DNA is exposed, is essential for stimulating Rad51 activity.[153] Rad51 activity is regulated through phosphorylation events. However, its constitutive phosphorylation by tyrosine kinase ABL1 causes a higher frequency of HR events in repeat regions of DNA, resulting in chromosomal instability and cancer progression.[154] In addition, a single point mutation in Rad51 (Arg-150-Gln) gives a 3- to 4-fold decrease in ATPase activity reducing HR and increasing the risk of breast cancer.[155] Rad51 mutations are also observed in Fanconi Anaemia (section 1.9).[156] Rad51 mutations associated with FA increase sensitivity to DNA-damaging agents, chromosomal instability and defective DNA repair.[1,157]

Rad51 NPF formation is an essential turning point of HR in DNA repair. The destabilisation of the Rad51 NPFs is one way to prevent extensive HR from proceeding to prevent potential chromosomal rearrangements and deletions in favour of SDSA.[154]

1.10.2 Rad51 paralogue proteins in recombination.

Rad51 NPF assembly relies on the activity of the accessory Rad51 paralogue proteins that, alongside Rad52, recruit Rad51 to DNA break sites. While Rad52 directly interacts with Rad51 stimulating Rad51-catalysed strand invasion, physical interactions with the Rad51 paralogues is not as well defined.[146,158]

The paralogues, Rad51B, Rad51C, Rad51D, XRCC2 and XRCC3, share 20-30% sequence conservation to Rad51 and to each other.[159] These conserved sequences are located around the Walker A and B domains and function as DNA-binding motifs with weak ATPase activity.[159] The paralogues form two major complexes, BCDX2 (Rad51B-C-D-XRCC2) and CX3 (Rad51C-XRCC3), (figure 1.14) but other sub-complexes can assemble (e.g. Shu complex).[158,160]

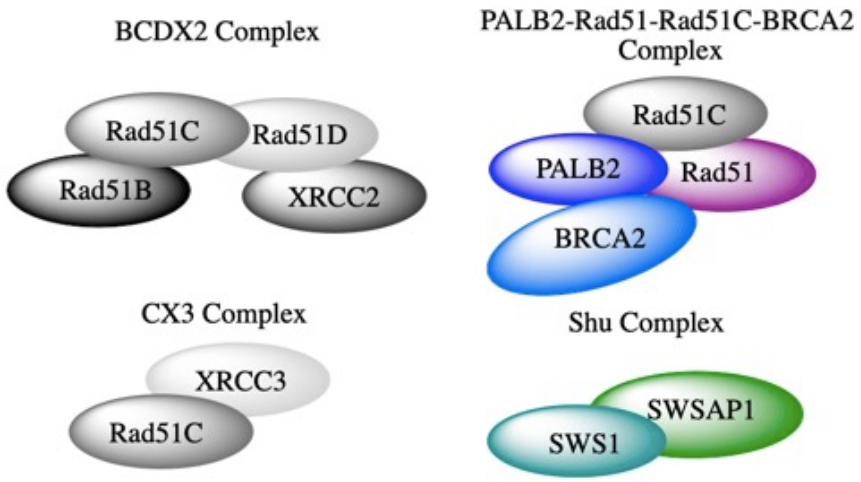


Figure 1.14 Complex formation of the Rad51 paralogue proteins. Rad51, the recombination protein essential in homologous recombination, has a set of paralogue proteins required for activation and aiding its activity. These proteins form complexes shown here. The two major complexes are the BCDX2 and CX3 complexes. The paralogues also interact with other proteins including BRCA2 and PLAB2 to form sub-complexes. Additionally, the Shu complex (SWS1 and SWSAP1) is a highly conserved regulator complex involved in HR.

Cells have increased chromosomal aberrations in the absence of the paralogue proteins highlighting their importance in mediating HR.[161,162] DNA damage caused by infrared radiation (IR) results in the assembly of Rad51 and Rad51C at nuclear foci co-localising with RPA.[163] Rad51C foci remained after the Rad51 disassembled suggesting a role of Rad51C in early and late stage HR.[163] In addition, Rad51C is involved in delaying cell cycle progression in response to DNA damage implicating Rad51C in multiple stages of DNA repair control.[163]

The paralogues are also associated with maintaining telomeric DNA and are important in telomere stability.[164] G-rich telomeric DNA protects the chromosome ends from end-to-end fusions, breakage and rearrangements. Rad51D and Rad54 promote HR reactions at telomeres and their deficiency results in telomere shortening and loss of the telomere cap during replication.[164]

HR is crucial for the repair of most lethal DNA damage and therefore loss of the paralogues can be detrimental.[165] Rad51C has been implicated as a tumour suppressor gene because heterozygous inactivation leads to tumour development.[166] Rad51C germline mutations have also been linked to the development of FA-like disorder, breast and ovarian cancers.[167,168]

1.10.3 BRCA1 and BRCA2 in recombination.

HR control in DNA repair is required to balance the need to restart replication against the undesirable by-product of genetic diversity. *Breast cancer early onset gene 1* (BRCA1) and BRCA2 proteins co-localise with Rad51 at nuclear foci and gene mutations leading to breast, ovarian, pancreatic and prostate cancers.[143,148,169]

BRCA1 consists of an *N-terminal Really Interesting New Gene* (RING) domain, a central large unstructured region and a coiled-coil domain with tandem *BRCA1 carboxy-terminal repeats* (BRCTs).[170] BRCA1 complexes with proteins in transcription, cell cycle checkpoint activation and HR and participates in end resection to promote HR.[169,171–175] It achieves this by inhibiting the end resection inhibitor protein, *p53-binding protein* (53BP1), and recruiting the *partner and localiser of BRCA2* protein (PALB2) to DSBs (figure 1.15).[175] It has also been shown to interact with the endonuclease CtIP, essential in DNA end resection, and the FancJ helicase involved in DSB repair.[176,177] In addition to end resection, BRCA1 aids in Rad51 NPF assembly.

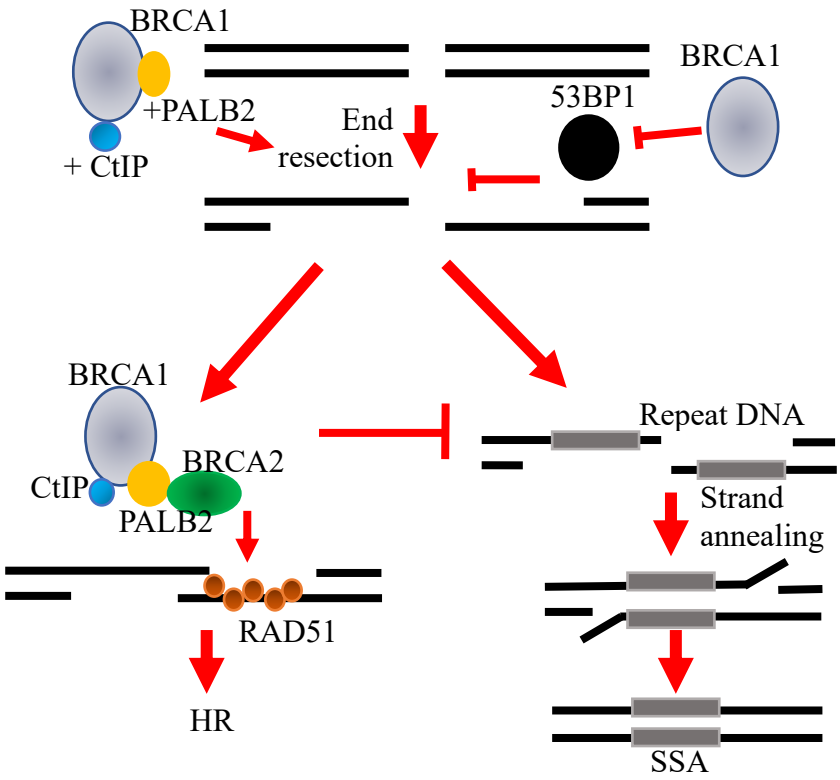


Figure 1.15 BRCA protein complexes involved in DSB repair and HR progression. BRCA1 promotes end resection and recruits PALB2 during early stage HR. BRCA1 inhibits the resection inhibitor protein 53BP1 driving HR end resection forward. BRCA1 also regulates resection by recruiting CtIP. BRCA2 promotes the assembly of Rad51 NPFs on ssDNA. In the absence of the Rad51 filament formation, the 3′ ssDNA is diverted into SSA repair.

BRCA2, also involved in stimulating HR, has eight conserved BRC repeats essential in activating Rad51 loading onto ssDNA and preventing Rad51 nucleation on dsDNA (figure 1.15).[178–182] Loss of BRCA2 results in fewer end resected DNA products being directed into HR.[183,184] The BRCA proteins are essential in promoting HR and highlights the extensive network of proteins required for successful and appropriate HR to occur. As mentioned previously, a central protein in activating DNA repair during stalled replication is RPA.

1.11 RPA as a first responder in replication coupled repair.

RPA signals stalled DNA replication to initiate protein recruitment for DNA repair, focused on in this review (figure 1.16).[185,186] RPA is essential in DNA maintenance in times of DNA damage evident by reduced levels of RPA70 eliciting

a number of severe phenotypes including slow S phase progression, replication arrest and reduced cell viability.[187–190]

RPA binds exposed ssDNA in a specific binding mode (see section 1.5.3) acting in the DNA damage response checkpoints to detect DNA lesions, specifically DSBs, and initiate repair (see figure 1.6).[190] At DSBs, RPA-ssDNA recruits Rad51 to bind DNA and catalyse recombination inhibiting alternative pathways (e.g. microhomology-mediated end joining (MMEJ)) in favour of HR.[191–193] While the mechanism of Rad51 recruitment by RPA is unknown, the interaction is essential for fork reversal to promote repair.[185]

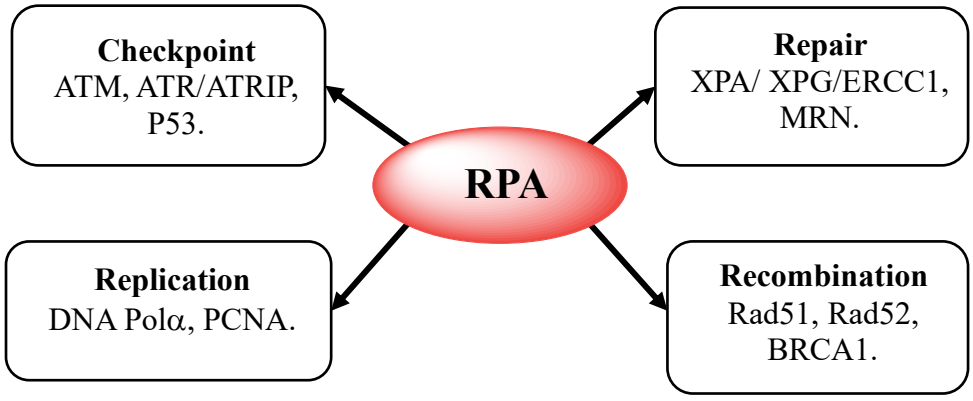


Figure 1.16 Summary of RPA interacting proteins. RPA is involved in a number of processes of DNA repair and replication. RPA recruits downstream proteins involved in damage response, repair and recombination to areas of damage. Some of the interacting partners of RPA are shown here.

RPA also stimulates activity of Rad52 and BRCA2 required for HR mediation. This implicates Rad52 in accelerating the Rad51-mediated displacement of RPA to form the pre-synaptic complex for HR.[194] RPA is also involved in HR dissociation to prevent extensive late stage HR by aiding NPFs disassembly by anti-recombinogenic mediators e.g. Srs2 (section 1.12.4).[195]

RPA-DNA damage response also relies on interactions with ATRIP, Rad9, Mre11, p53, XPA and ATM and ATR kinases to activate checkpoint response.[196] RPA transiently interacts with proteins allowing for the constant exchange of binding partners.[185] RPA-XPA interactions form a DNA damage recognition

complex which has a higher specificity for damage then XPA alone signalling DNA damage sites to promote repair.[197] This results in fewer non-specific DNA-RPA aggregates forming.[173]

RPA32 recruits the DNA annealing helicase *SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1*/*HepA-related protein* (Smarcal1/Harp) at stalled replication forks.[185] RPA stimulates Harp fork reversal on the leading strand to promote repair and inhibits Harp on the lagging strand. Excessive fork reversal is prevented by ATR phosphorylation of Harp.[185]

The inactivation of late stage HR repair is essential to prevent genome instability. Further to the roles of RPA mentioned here, an elaborate network of helicases, resolvases and dissolvases eliminate intermediate HR structures to block DSBR and promote alternative pathways making them essential in maintaining genome integrity. The next section focuses on the role played by helicases to protect the genome from the negative implications of HR but firstly we need to understand their function.[198]

1.12 DNA helicase enzymes.

Various helicase enzymes are needed to maintain genome stability in cells. Helicases couple translocation of DNA or RNA with separation of duplex strands.[199,200] DNA helicases disrupt hydrogen bonding between two DNA strands as an essential process in order to access the information stored in DNA, both in replication and repair.[200] The disruption of hydrogen bonds is coupled with ATP hydrolysis and is dependent on the presence of a DNA effector, usually ssDNA. The multi-functional activity of helicases mean they are prevalent in most biological processes.[201,202] Therefore, their prominent involvement in cellular function means helicase defects and expression downregulation are seen in

developmental diseases, neurodegenerative diseases and cancer.[203] The fundamental activities of helicases are reviewed here before applying them to HelQ specifically.

1.12.1 Mechanism of action by helicase proteins.

Helicases hydrolyse nucleoside triphosphate (NTP) and utilise the free energy to fuel translocation along DNA, unwinding the duplex in the process. Helicases are described as molecular motor proteins due to their ability to convert chemical energy into mechanical energy.

Helicase DNA strand separation can occur through 'passive' or 'active' mechanisms.[204] Both mechanisms require ATP hydrolysis but the difference refers to the stability of the duplex DNA.[205] Passive unwinding suggests the helicase is not involved directly in destabilisation of the duplex. As the helicase translocates along the DNA it traps the ssDNA at a thermally fraying fork and does not come into contact with dsDNA. Active mechanism of unwinding suggests the helicase interacts directly with dsDNA and actively destabilises the duplex. For example, the *E. coli* PcrA helicase actively unwinds duplex DNA through two separate, but coupled, processes of ssDNA translocation and duplex destabilisation driven by ATP hydrolysis.[204]

The mechanism of DNA translocation and duplex destabilisation differs between helicases. Most proposed active mechanisms for helicase catalysed DNA unwinding requires multiple DNA binding domains and therefore they tend to form oligomeric structures.[206] The mechanism of action that helicases can adopt are discussed here next, with specific examples.

An active mechanism requires helicases to be able to interact with both ssDNA and dsDNA and bind simultaneously during at least one stage of unwinding.

Helicases that adopt an active mechanism can be further categorised based on how they assemble and move on DNA, these include an the 'inchworm' model (e.g. *E. coli* monomeric PcrA) and the 'rolling' model (e.g. *E. coli* dimeric Rep and UvrD).[204,206] Furthermore, these mechanisms can be further sub-divided, for example the 'quantum inchworm' proposed for the RecBCD helicase. This suggests that different domains carry out translocation and unwinding activity as two separate and consecutive events.[207] In addition, the 'torsional' model does not require the helicase to interact directly with the duplex DNA but interacts at the ss/dsDNA junction and distorts the adjacent duplex DNA through ATP-induced conformational change.[206]

Helicases that adopt higher oligomeric state can also be categorised based on mechanism of translocation. For example, hexameric helicases such as the eukaryotic Mcm2-7 which actively unwinds dsDNA through the currently accepted model of steric exclusion.[208] These models suggest ssDNA is excluded around the outside of a ring shaped hexameric structure as the helicase translocates along the DNA.

More specific mechanisms of translocation and unwinding for individual helicases can be hypothesised based on crystal structures and helicase function. These can be derivatives of known models or new models. One example is the 'docking simulation' mechanism of the RecQ helicase BLM which is specific to BLM activity with Holliday junctions in branch migration.[209] This review does not assess these more specific mechanisms of translocation and activity. Mechanism of activity varies across helicases families, discussed here next.

1.12.2 Superfamily-1 (SF1) and superfamily-2 (SF2) helicases.

DNA and RNA helicases are predominantly split into two superfamilies; SF1 and SF2.[199,210] While structure, general unwinding activity, translocation and ATP

hydrolysis is conserved, the unwinding mechanisms vary considerably as discussed. Unwinding can occur as the helicase translocates along the DNA, translocation can occur without unwinding and unwinding can occur without translocation.

DNA helicases translocate in different directions because of the opposite chemical polarities in the DNA double helix. Helicases that require a 3′ ssDNA binding site exhibit a 3′-5′ polarity and vice versa.[200] The nine SF2 subfamilies and three SF1 subfamilies are categorised by translocation mechanism and unwinding polarity (figure 1.17).[203] Hel308 and HelQ, relevant to this work and discussed in more detail later, are Ski-like SF2 helicases.[210] Other SF2 helicases include bacterial RecG and the RecQ helicases.[211]

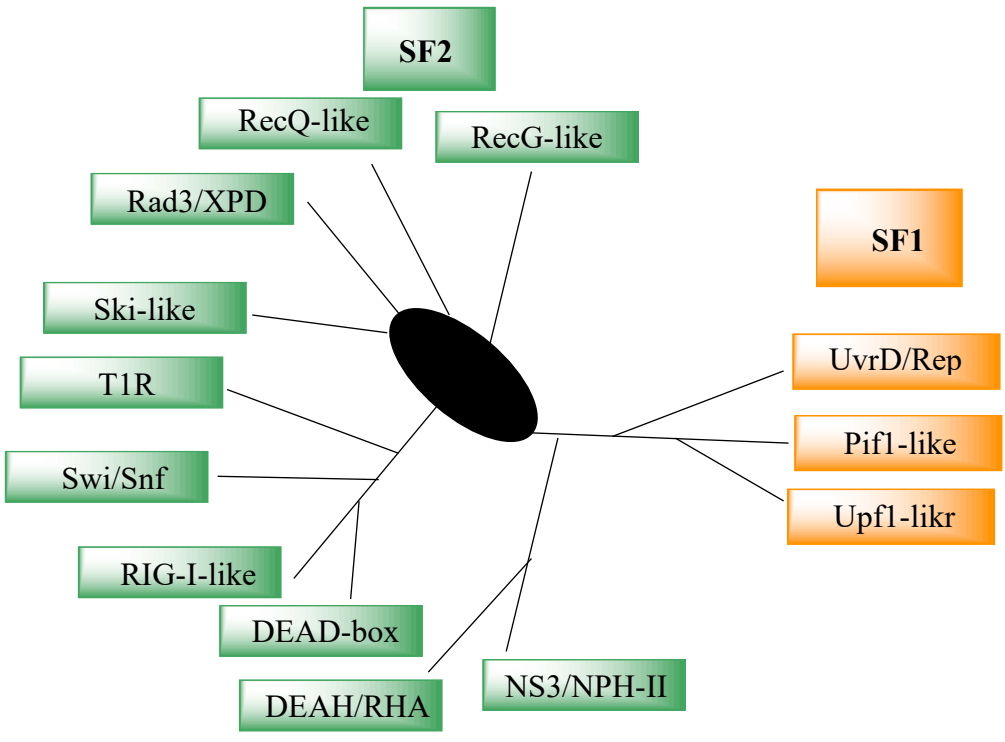


Figure 1.17 A summary of the helicase super-families' SF1 and SF2. DNA and RNA helicases are categorized on sequence homology. SF1 consists of three sub-families and SF2 of nine sub-families. The families are separated based on helicase translocation mechanism and unwinding polarity.

Helicases have a conserved helicase core consisting of two similar RecA-like domains with strong homology to the bacterial recombination protein RecA. The RecA-like domains are important in phosphodiester hydrolysis, evident from their presence in P-loop NTPases.[212] The P-loop, also known as the Walker loop or phosphate-binding loop, is the Walker A motif that binds to ATP. The P-loop NTPase fold is the most prevalent domain of nucleotide-binding protein folds which hydrolyses the phosphoanhydride bond between phosphate molecules of ATP.[212] Mutagenesis within the RecA-like domains impair the coupling of ATP hydrolysis and DNA binding inactivating the protein. In some helicases (e.g. PriA, UvrD and RecD2) ATP hydrolysis changes the 'state' of the protein, distorting the conformation of the ATP binding site during unwinding. This suggests an unwinding mechanism coordinated by conformational change.[203]

The DNA binding interface of helicases, found conserved within families, is located on the surface of the protein opposite to the ATP binding site.[203] Ski-like and DEAH/RHA SF2 families also have a conserved β-hairpin domain important in translocation mechanism. For example in archaeal Hel308, the homologue of HelQ, the β-hairpin acts as a 'pin' to separate the duplex strands at the ssDNA-dsDNA junction.[203] The β-hairpin, also found in the WRN and BLM helicases, wedges into the stacked bases at the duplex terminus, separating the strands by resulting in the loss of base-base stacking. Different residues have been shown to be essential in DNA binding within the different β-hairpin of the different helicases.[209] The C- and N-terminal domains of helicases contain accessory domains important for specific protein function.[213] These include regions of nuclease activity, DNA binding domains (zinc fingers and OB folds), protein-protein interacting regions and oligomerisation regions.

As well as unwinding duplex DNA, a small group of helicases can unwind and resolve DNA structures that arise during specific cellular pathways (e.g. HR).[203,214] For example, *E. coli* RuvAB complex, as part of the branch migration RuvABC complex, unwinds DNA to resolve HJs.[215] Helicases have also been associated with the unwinding of R-loops, stable RNA-DNA hybrids formed during transcription.[104] Helicases process flap DNA structures and DNA triplexes (e.g. WRN and BLM), G-quadruplexes (e.g. BLM, WRN and FancJ) and Z-DNA. Alternative roles of helicases include strand annealing and protein displacement, discussed next (figure 1.18).[186,200]

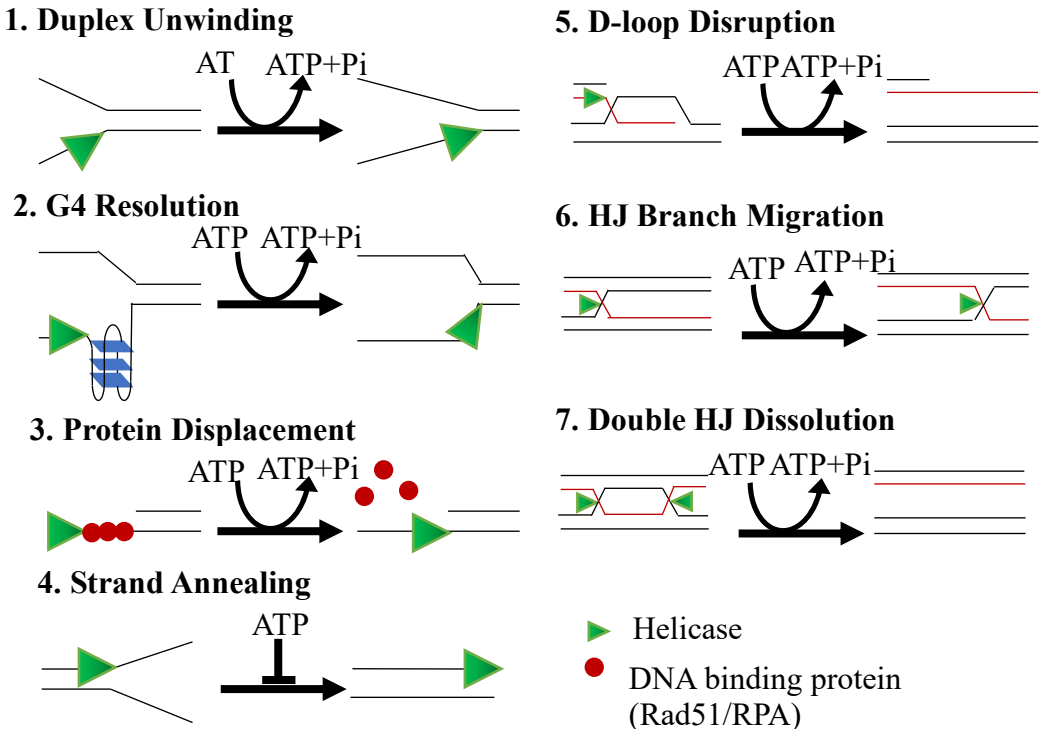


Figure 1.18 A summary of DNA processing activity by helicases. A schematic adapted from Awate *et al.*[186] Helicases are associated with a range of DNA processing activities including processing of intermediate DNA structures e.g. HJs, G4 quadruplexes, triplexes, R-loops and D-loops. Helicases are also important in strand annealing and protein displacement during DNA stress.

1.12.3 Accessory replicative helicases.

Failure to restart replication at stalled replication forks can result in genome rearrangements and cell death.[216] Replication stalling can be caused by copying

errors or damage to the DNA and can result in physical protein barriers that are inhibited from migrating along the DNA.[217] Not only does this stall replication, these barriers can prevent access to the DNA for repair proteins. These barriers can be caused by proteins in transcription, replication and other DNA-binding proteins.[218] Sometimes translocation of the replisome is sufficient to displace nucleoprotein complexes for replication to continue. However, barriers on the DNA can cause the replisome to dissociate preventing replication continuing. This not only requires the barrier to be removed, but also the replisome to be reloaded onto the DNA. Accessory replicative helicases are essential in fork clearance of protein blocks. This role of helicases arose from the finding that DNA unwinding by *E. coli* Rep was not inhibited by the lac repressor-operator complex when compared with the DnaB helicase.[219] Since then, many helicases, predominantly SF1, have been identified as protein displacement helicases (e.g. bacterial DinG and UvrD and eukaryotic Rrm3 and Pfh1). Physical and functional coupling of helicases to other proteins, such as RPA, can significantly improve unwinding in the presence of a barrier by increasing enzyme processivity.[217] These helicases promote DNA repair by displacing barriers and providing access to the DNA for repair proteins and also promote replication restart at stalled replication forks. Protein displacement activity shown by accessory helicases is also important in regulating late stage HR by dismantling NPFs, discussed next. This highlights the importance of helicases in times of DNA stress to maintain DNA integrity.

1.12.4 Helicases in homologous recombination.

Helicases in DNA repair are associated with blocking extensive HR and re-directing damage sites into more favourable pathways of SDSA.[1,220] The role of helicases in HR and similarities to HelQ (section 1.14) allow us to hypothesis that this is where HelQ acts.

Several SF1 helicases translocate along DNA displacing recombinase proteins and dismantling the pre-synaptic RecA NPFs physically (e.g. bacterial PcrA) or indirectly (e.g. bacterial UvrD).[1,221] PriA only disassembles NPFs after RecA has hydrolysed ATP.[222] UvrD, and the Srs2 yeast homologue, activates the recombinase ATPase activity to disassemble the NPF.[1] Srs2 contains a C-terminal Rad51-interacting domain for displacement of Rad51 as it translocates ssDNA.[223] The human UvrD homologue *F-box helicase 1* (FBH1) shows similar activity implicating the UvrD family as important in early HR control.[224]

Human RecQ helicase homologues (WRN, BLM, RecQ1, RecQ4 and RecQ5) are involved in HR synaptic D-loop checkpoints.[225,226] RecQ proteins unwind D-loops and HJs at stalled replication forks during HR synapsis in order to redirect HR prior to DSBR.[227] RecQ1 has strand-annealing activity which allows the re-direction of extensive late stage HR to SDSA.[228,229] Furthermore, BLM complexes with *DNA topoisomerase 3-α* (TopIIIα) loosening DNA supercoiling to disrupt HJ formation and prevent cross-over reactions.[230]

Srs2 DNA helicase acts to block late stage HR, evident by a hyper-recombination phenotype in Srs2 mutants.[231] Srs2, recruited by RPA, disassembles Rad51 NPFs allowing occupation by RPA to prevent re-annealing and promote SDSA.[232] It is proposed that Srs2-Rad51 interactions stimulate ATPase hydrolysis within the Rad51-ssDNA filament causing its dissociation.[233]

Helicases are also associated with promoting HR. The human SF2 FA helicase FancM, conserved in archaea (Hef), does not have helicase unwinding activity but translocates along DNA in an ATP-dependent manner to promote replication fork reversal and HJ migration.[234,235]

Human HelQ and archaeal Hel308 have been implicated in HR regulation through gene deletion phenotypes.[236,237] These proteins promote DNA repair through HR but negatively impact late stage HR, ensuring the process does not go beyond the formation of D-loops. HelQ is hypothesised to promote Rad51 NPF formation by removing pre-bound RPA from ssDNA.[237] *C. elegans* HelQ physically interacts with Rad51-dsDNA complexes but not Rad51-ssDNA which suggests a role in D-loop dissassembly.[236] Hel308 and HelQ are discussed in sections 1.13 and 1.14 respectively.

1.12.5 Helicases associated with disease.

The role of helicases in DNA processing and remodelling of nucleoprotein complexes makes them prominent in disease, particularly in age-related diseases and cancer. Bloom syndrome is a rare disease caused by chromosomal aberrations in the *BLM* helicase gene. BLM resolves 3′ overhang DNA, G-quadruplexes, D-loops and HJs.[229] Therefore, *BLM* mutations result in abnormalities in HR regulation resulting in increased sister chromatid exchanges (SCEs).[229] Mutations in the RecQ family helicase WRN leads to Werner syndrome, characteristic of premature aging caused by altered telomere maintenance.[229] RecQ5 helicase is linked to increased cancer susceptibility due to impaired HR regulation.[238] As with BLM, RecQ5 deficiency results in increased SCEs and HR events subsequently leading to chromosomal rearrangements upon DNA stress.[239]

Another well documented disease associated with helicases and HR is Fanconi Anaemia (section 1.9). FA is characterized by a hypersensitivity to ICL-inducing agents, tumours, congenital malformations and acute myeloid leukaemia.[134] Other helicase associated diseases include Xeroderma pigmentosum, Cockayne's syndrome, trichothiodystrophy and cerebro-oculo-facio-skeletal syndrome (COFS) as a result of mutations in the XPD helicase. Mutations in Mcm4, RecQ1, and HelQ

have also been associated with cancer.[240] The prevalence of disease highlights the importance of studying the role of helicases in maintaining genome stability.

1.13 Homologues of HelQ.

Studies carried out on the homologues of HelQ has allowed for predictions and models to be made for the role and activity of human HelQ and act as a starting place for work here.

1.13.1 Archaeal Hel308 helicase.

Archaea tend to lack helicase orthologues involved in replication restart at stalled forks. However, RecQ is found in a few archaea likely as a result of bacterial horizontal gene transfer.[241] Similarities in DNA replication and repair between eukaryotes and archaea means the absence of RecQ involved in fork restart is surprising. Therefore, it was proposed that archaea may contain analogous helicases to RecQ.[241] Archaeal Hel308 from *Methanothermobacter thermautitrophicus* was identified from a genetic screen in a fork-stalling *E. coli* strain as giving a phenotype that was similar to that obtained by reintroducing RecQ.[242] This implied a similar function in helicase-DNA interactions at stalled replication forks. This has since led to the biochemical analysis of recombinant Hel308 from *Methanothermobacter* to characterise the activity of Hel308 in replication fork stalling.[241,242]

Archaeal Hel308 is a SF2 helicase and an archaeal orthologue of the metazoan HelQ and PolQ and *Drosophila melanogaster* Mus308 helicases.[243] The Hel308 protein with a molecular weight of 78.5 kDa was identified because of conservation with the helicase domain of the *Drosophila* Mus308 protein.[242] As mentioned previously, Hel308 is implicated in replication fork restart and therefore may be involved in replication-coupled repair and genomic stability. Hel308 from *Pyrococcus furiosus* (Hjm) and *Methanothermobacter* have been used to

biochemically study archaeal Hel308.[211] The biochemical data from archaeal Hel308 can provide guidance towards explaining phenotypes observed for metazoan orthologues, such as Mus308, as well as the human homologue HelQ. Therefore, archaeal Hel308 has been used here to model predictions for HelQ activity and fork restart.

Biochemical analysis of archaeal Hel308 has identified 3′ to 5′ DNA polarity with limited ATPase activity. Helicase assays *in vitro* have shown that Hel308 binds and unwinds partial fork DNA substrates with exposed ssDNA on the leading strand template with high efficiency. Furthermore, Hel308 displaces the lagging strand from the duplex DNA.[242,243]

The crystal structure of Hel308/Hjm has been solved from *Saccharolobus solfataricus* (PBD ID 2VA8)[211], *Archaeoglobus fulgidus* with and without DNA (PBD ID 2P6R; 2P6U)[244] and in multiple orientations and complexes from *Pyrococcus furiosus* (PDB 2ZJA; 2ZJ8; 2ZJ5; 2ZJ2).[245,246] The domain organisation implicated by structures from different organisms were relatively similar, therefore, the solved Hel308 complexed with DNA structure from *Archaeoglobus fulgidus (Afu)* is described here.[244] A solved crystal structure of Hel308 with DNA allows us to visualise the organisation of the five protein domains (figure 1.19).[211] The core RecA-like ATPase motor domains 1 (residues 1-197) and 2 (200-416) are conserved within helicase proteins and act as the motor for driving activity. ATP hydrolysis within the RecA-like domains induces conformational change for translocation and unwinding activity. An exposed β-wing as part of the WHD, mentioned previously, (residues 426-501) and a 'recognition' ratchet α-helix, prominent in DNA binding proteins, were also

identified. WHD and RecA-like domain 1 are tightly associated, a common characteristic of Ski2-like helicases.[247]

Domain 4 (residues 502-644) is a seven-helical bundle specific to helicase unwinding activity. Domain 5 (residues 647-705) is a helix-hairpin-helix domain associated with ssDNA binding proteins.[211] Domain 5 contains a positively charged RAR (Arg-Ala-Arg) motif conserved across the Hel308 family that allows the domain 5 'cap' structure to point towards the central pore of the protein formed by domains 1 to 4. The organisation of Hel308 has allowed for predictions in translocation mechanism and DNA loading to be made.

Hel308 has been used as a model to study HelQ, however, with only 25% sequence similarity (figure 1.20), the use of Hel308 as a model for HelQ is a question assessed in work carried out here. The domain organisation is predicted to be similar but little else is known about similarities between the role and function of the proteins.

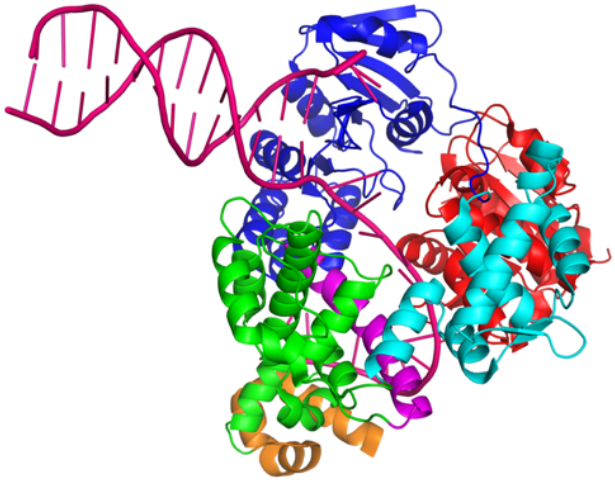


Figure 1.19 Crystal structure of *Afu* Hel308 protein. The atomic resolution structure of the HelQ archaeal homologue Hel308 from *Archaeoglobus fulgidus (Afu)* bound to a 3′ ssDNA tailed duplex (pink). Hel308 consists of RecA-like domains 1 and 2 (dark blue and red), the WHD (pale blue), domain 4 (green) with the ratchet domain (magenta) and the domain 5 cap (orange). PDB accession number 2P6R.

1.13.2 Eukaryotic PolQ helicase.

Human DNA helicase PolQ, also known as DNA polymerase theta, is a 290 kDa protein containing an N-terminal helicase-like domain and a C-terminal DNA polymerase domain with a central non-structured region.[248] PolQ is one of 15 DNA human polymerases and was first identified by its homology to the Drosophila Mus308 gene. The N-terminal helicase domain (PolQ-HLD), residues 67-894, has strong sequence homology to HelQ, Hel308 and other SF2 helicases (figure 1.20).

PolQ-HLD is made up of five protein domains that form a globular structure. Like HelQ and Hel308, PolQ contains two RecA-like domains essential for ATPase activity. The crystal structure of PolQ isolated an additional 32 residue region not seen in any other SF2 helicase structure.[249] The additional residues, located within the second RecA-like domain, form an extended loop and α-helix that packs along the first RecA-like domain and is involved in oligomerisation.[249] PolQ-HLD forms a tetrameric clover-shape with each monomer interacting with all three other monomers. These interactions rely on interfaces at the additional region of domain 4 and may explain why tetrameric structures are not adopted by other Hel308-like proteins.[249]

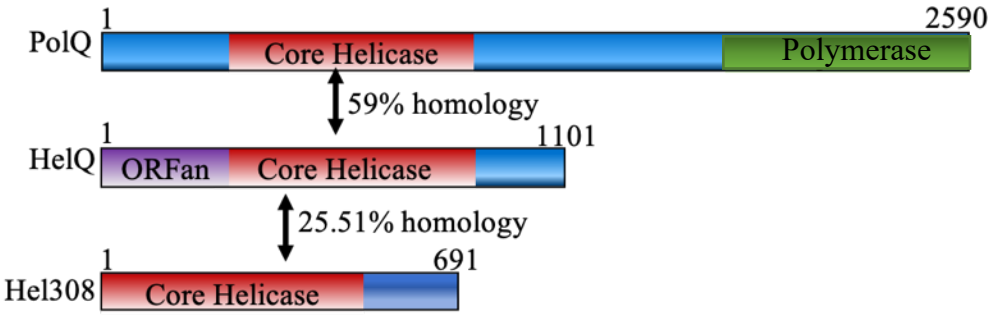


Figure 1.20 A schematic to show the homology between Hel308, HelQ and PolQ. Human PolQ and HelQ proteins and archaeal *Mth* Hel308 share sequence homology within the core helicase domains (red). In addition, PolQ has a C-terminal polymerase domain (green). The core helicase between HelQ and PolQ show strong amino acid sequence homology. Neither PolQ nor Hel308 contain the N-terminal domain found in HelQ known as the ORFan domain do to the lack of sequence and function homology (purple).

PolQ is an essential factor in promoting alternative NHEJ repair and acts in MMEJ and DSB repair.[250] Therefore, like other helicases mentioned, PolQ is a negative regulator of late stage HR important in maintaining genome integrity. In MMEJ, PolQ aligns microhomology sequences at DSBs promoting end joining.[250] PolQ acts as an annealing helicase ligating cut sites at DSBs preventing activation of downstream HR repair of DSBs. PolQ has been associated with a hyper-sensitive to DNA damaging agents and increased cancer predisposition; 70% of breast cancers show a 5-fold increase in PolQ levels.[250]

In addition to annealing and unwinding activity, PolQ is also associated with protein displacement as an accessory helicase. PolQ-HLD can displace RPA from resected DSBs, inhibiting RPA70, to allow annealing and end joining by alt-NHEJ and therefore block recombination repair. PolQ opposes the activity of RPA to balance repair carried out by HR and NHEJ.[251]

This background knowledge about the homologues of HelQ has allowed us to use them as models to make predictions of HelQ structure and activity.

1.14 Human HelQ helicase.

Work carried out here focused on the SF2 Ski-2 like helicase HelQ. HelQ is an ATP and ssDNA-dependent helicase that translocates DNA 3′ to 5′.[252] HelQ is found in metazoans including human, mouse, Drosophila and *C. elegans*. The archaeal homologue, Hel308, has been used as a model for human HelQ function. HelQ has been studied biochemically but little is known about its mechanism of action, structure, stable oligomeric state or role in replication coupled DNA repair.

HelQ is suggested to promote genome stability at stalled replication forks by acting in HR and ICL repair but the mechanism for this is unknown. In *C. elegans* a combined loss of HelQ and the Rad51 paralogue Rfs blocked DSB repair after homologous strand invasion preventing HR progression.[253] This suggests a failure to disassemble Rad51 from NPFs that is required during strand invasion for homology search to continue.[253] Additionally, in a DNA damage background, HelQ deficiencies lead to the accumulation of HR intermediates (HJs) suggesting extensive HR was able to occur.[142] These phenotypes imply HelQ is a mediator or negative regulator for the progression of extensive HR by promoting post-synaptic Rad51 filament disassembly after strand invasion. Interactions with RPA and the Rad51 paralogue complex (see section 1.14.1) may imply that HelQ is indirectly involved in Rad51 disassembly by impacts on Rad51 accessory proteins. Further work to establish direct interactions between HelQ and Rad51 are required.

HelQ has also been associated with ICL repair. A decreased survival after ICL treatment in *C. elegans* lacking HelQ implicates HelQ in ICL repair and potential links to the FA pathway.[254,255] Truncated HelQ proteins expressed in mice models gave rise to hypo-gonadism phenotypes, a trait associated with FA mouse models.[255] Furthermore, reduced cellular levels of HelQ resulted in mild MMC

sensitivity. These similar deletion phenotypes observed between HelQ and the FA pathway implies similarities in the mechanism of ICL repair.[255]

Previous studies have resulted in HelQ being associated with pathways of DNA repair with links to HR regulation and ICL damage. While the mechanism and exact involvement is yet to be determined, HelQ is highlighted as a protein of interesting when assessing DNA repair at stalled replication forks to maintain genome integrity.

1.14.1 Interactions of HelQ with other proteins.

HelQ has been shown to co-localise *in vivo* at ssDNA with the HR proteins Rad51 and the FA protein FancD2.[253,256] Furthermore, proteomic analysis revealed links between HelQ and a Rad51 paralogues complex (BCDX2) and DNA damage response kinase ATR.[236,253,256] These protein-protein links further implicate HelQ in the regulation of HR events at damaged replication forks and therefore in maintaining the genome (figure 1.21).[237] Only proteins in the BCDX2 complex and not the CX3 complex were isolated with HelQ. These proteins were still detected in pull-downs with HelQ even after treatment with MMC further supporting the notion that HelQ acts in cellular resistance to ICLs.[236]

Further interaction studies have also identified that HelQ co-localises with RPA at areas of cisplatin induced DNA damage *in vivo*.[256] This interaction is potentially unsurprising because of the extensive role of RPA in bridging replication stalling and DNA repair. RPA may recruit HelQ to regions of DNA exposed at stalled replication forks for downstream repair. However, the extent of the relationship is yet to be established and was a focal point during this PhD project.

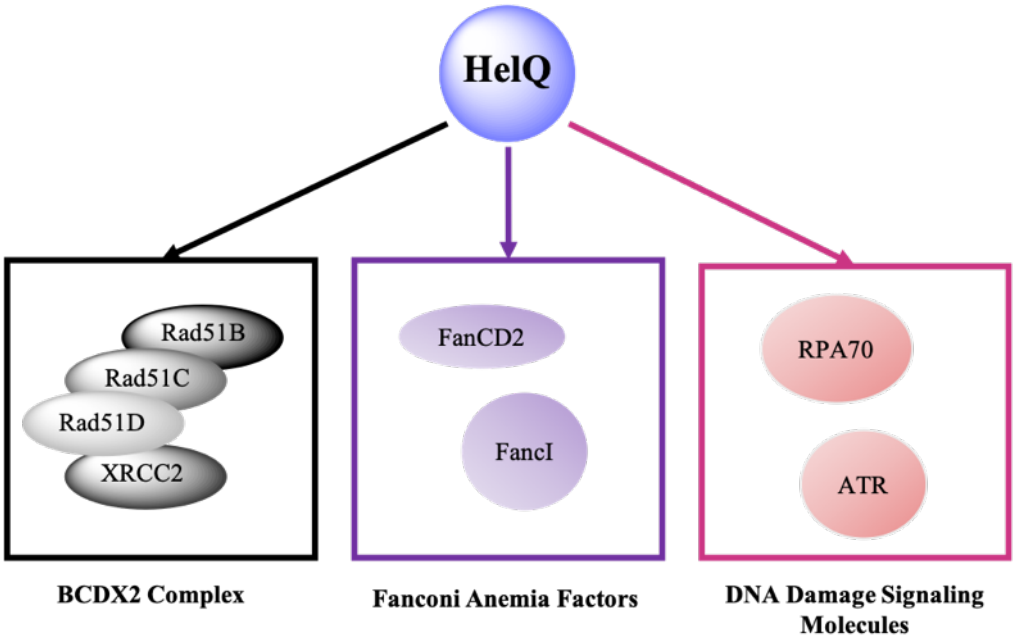


Figure 1.21 Proteins linked to HelQ. Schematic adapted from Adelman *et al*.[236] of the proteins found to associate with HelQ. The interaction network was built based on mass spectroscopy results and bioinformatic predictions carried out using BIOGRID, STRING and MINT.

At a genetic level, *HelQ* is epistatic with *FancD2* and *JMJD-5* genes.[254,257] This suggests that one gene determines whether or not the trait of a second gene is expressed and therefore the genes are antagonistic.[257,258] This genetic link has been shown in *C. elegans,* however, no physical protein predications have been observed.[257] JMJD-5 is a demethylase protein which regulated histone 3 Lys-36 methylation that is required for late stage HR.[259] The antagonistic relationship would suggest that in the absence of *JMJD-5* expression, late stage HR is not activated and therefore downregulation is not required and subsequently *HelQ* expression is not activated.[258] In the reverse, if expressed *JMJD-5* stimulates late stage HR, HelQ expression may be required to block HR progression. Likewise, due to its role in homology-directed DNA repair, the expression of *FancD2* is likely to impact *HelQ* expression depending on HR control required. These genetic links further support the hypothesis that HelQ is involved in downregulating extensive HR.

1.14.2 Association of HelQ with disease onset or resistance to treatments.

Understanding the role of HelQ in DNA repair and maintaining genome integrity is important because of its prevalence in cancer. Disruptions to HelQ is associated with breast and ovarian cancer and oesophageal squamous cell carcinomas.[260,261] A genetic screen of *HelQ* in 185 breast and ovarian cancer families revealed seven exon sequence changes and five missense variants of HelQ.[166,260] Furthermore, HelQ can be used as a novel indicator of chemotherapy resistance for epithelial ovarian cancer.[262] The over-expression of HelQ increased cellular resistance to the chemotherapy drug, cisplatin, in cancer cells. It was hypothesised that HelQ was initiating an increase in DNA repair response. Loss of HelQ resulted in loss of cisplatin resistance and reduced expression of the repair proteins.[262] This identified HelQ as a new diagnostic marker for ovarian cancer, a disease that is challenging to understand and diagnose.

Increasing evidence that HelQ is found associated with cancer and a potential link to chemo-resistance makes it an interesting and important target for research. This motivates us to understand the fundamental biochemistry of HelQ. In order to predict the role of HelQ in regulating HR and where it goes wrong, we first need to understand the fundamentals of HelQ structure and activity with DNA.

1.14.3 Predicted structural organisation of HelQ.

A crystal structure has not been solved for HelQ and therefore predicted models are based on sequence homology with other helicases (e.g. Hel308 and PolQ).[242] Highly conserved RecA-like ATPase domains consisting of the Walker A and B motifs in HelQ are important for ATP binding and hydrolysis and subsequent translocation activity. Other conserved domains to Hel308 include the helical ratchet domain and WHD.[263]

Previous work in the Bolt group identified specific mutations within the predicted WHD (Tyr-818-Ser and Lys-819-Asp) that removed the ability of HelQ to bind duplex DNA.[264] This suggests the WHD is essential in interacting with fork DNA structures close to the ssDNA-dsDNA junction. The WHD is hypothesised to be important in substrate selection of forks with an intact lagging strand and ssDNA gaps on the leading strand.[247]

The oligomeric state of HelQ is problematic to predict in the absence of a solved structure. Previous studies suggested HelQ adopts a hexameric structure eluting from gel filtration at 600 kDa.[265] However, this predicted structure is unlikely because it would suggest the presence of 12 RecA like domains, not seen in other classic hexameric helicases. The oligomeric state of a helicase can allow for models of translocation to be predicted. Hexameric helicases predominantly adopt a steric exclusion mechanism of translocation. This is where ssDNA is pulled through a central pore created by the ring structure of the six protein monomers while the non-translocating strand is excluded around the outside of the protein. Steric exclusion has been observed for hexameric Mcm2-7 and tetrameric PolQ. Work carried out here aimed to further assess the oligomeric state of HelQ and the potential mechanisms of translocation. Understanding how HelQ interacts with DNA builds a model for how HelQ may act at stalled replication forks which acted as a good starting point for work carried out here.

1.15 Summary.

Maintaining genome integrity is crucial for life and as described here is controlled by an elaborate network of mechanisms. The multi-level control during the cell cycle to prevent damage as well as repair pathways to resolve damage ensures changes to DNA rarely prevail. However, while the majority of damage is rectified in order for replication to continue, damage can become permanent and if

this occurs in prominent protein-coding genes can be detrimental. This review focused on replication-coupled DNA repair, the mechanism that processes damaged DNA in coordination with the replisome to ensure coupling of replication and repair while maintaining genome stability. One mechanism is through HR, and while this is beneficial for fork restart and DSB repair, extensive late stage HR (DSBR) comes with its own problems that need to be avoided. Proteins are essential to block extensive HR and prevent the potential troublesome cross-over events it can cause. This is where we propose HelQ acts implicating HelQ in maintaining genome integrity.

1.16 Project outlook.

Understanding the complex system of DNA repair pathways is paramount to predicting how these pathways can go wrong and lead to disease. Identifying the proteins involved, their structure, mechanism and the affects produced in their absence, can allow for advancement in disease prevention and correction. Determining the mechanism of helicases is important because of the essential role helicases play in maintaining genome stability.

Furthermore, HelQ is clinically relevant due to its association with cancer and an increased incidence of ovarian cancer and fertility defects. Understanding more about HelQ can help develop knowledge on the complex networks in controlling HR.

1.17 Project hypothesis.

We propose that human HelQ supports DNA replication forks by suppressing late stage HR (DSBR) and promoting alternative repair pathways (e.g. SDSA) and that this is fundamentally conserved through evolution.

1.18 Project aims and objectives.

The aim of this project was to gain a better understanding of how healthy cells utilise HelQ by developing a mechanism and model of HelQ activity. These aims were achieved by looking at a number of smaller objectives to understand the fundamental biochemistry of the helicase.

1. To determine the active oligomeric state of HelQ when bound to DNA and build a model for the assembly of HelQ oligomers and its role in translocation and unwinding.
2. To create a model for the binding, unwinding and ATPase activity of HelQ and different regions of HelQ with DNA.
3. To determine direct and indirect protein interactions between HelQ and RPA and isolate the impact of RPA on HelQ activity.
4. To characterise the catalytically active core helicase and catalytically inactive N-terminal ORFan regions of HelQ. To determine how the non-homologous region of HelQ is important to specific HelQ activity.

Chapter 2: Materials and Methodology

2.1 Chemicals

Chemicals listed within the methodology sections were supplied either by Fisher Scientific or Sigma unless stated otherwise.

2.2 Antibiotics

Table 2.1 Antibiotics and their used concentrations.

Antibiotic	Stock concentration	Used concentration
Ampicillin	100 mg/mL	50 µg/mL
Gentamicin	50 mg/mL	7 µg/mL
Chloramphenicol	35 mg/mL	35 µg/mL
Kanamycin	4 mg/mL	40 µg/mL
Tetracycline	1 mg/mL	10 µg/mL
S-gal	20 mg/mL	66.67 µg/mL

2.3 Bacterial strains and cell lines

Table 2.2 Cell strains and cell lines.

Name	Use	Company	Features
DH 5α	Work horse *E. coli* strain used for general cloning.	Invitrogen (#18265017)	Amp^R
NEB 5α	Replacement for DH 5α for more difficult cloning.	NEB (#C2987I)	Amp^R
Max-efficiency DH 5α	RPA cloning at RCaH.	Invitrogen (#18258012)	Amp^R
BL21 A1	Work horse *E. coli* strain used for general protein over-expression.	Invitrogen (#C6077003)	Tet^R
BLR (DE3)	Transformation of RPA vector for protein over-expression at RCaH.	Sigma-Aldrich (#69053)	Tet^R
BL21 (DE3) pLysS	Transformation of RPA vector for protein over-expression at RCaH.	Bioline (#Bio-85033)	$Chlm^R$
XL1-Blue	Transformation of site directed mutagenesis cloning of HelQ containing plasmids.	Agilent Technologies (#200236)	Tet^R
XL10-Gold	Transformation of hexaHis-SUMO-HelQ for plasmid maintenance.	Stratagene (#200314)	Tet^R and $Chlm^R$
One Shot TOP10/P3	Transformation to MBP/StrepII-HelQ containing vectors at RCaH.	Invitrogen (#C505003)	Tet^R and Amp^R
TURBO (DH10MultiBac)	HelQ bacmid recombination at RCaH.	Invitrogen (#10359-016)	Kan^R and Tet^R

BacterioMatch II (B2H) validation	BacterioMatch II (B2H) hybrid system.	Stratagene (#200192)	Kan^R

2.4 Oligonucleotides and primers

All oligonucleotides used for primers and the construction of DNA substrates used in this work were sourced from Sigma and were supplied lyophilized. All oligos were diluted in sterile-distilled water (SDW) to 100 μM.

Table 2.3 Primers used for cloning, plasmid verification and mutagenesis.

Gene Target	Use	5′-3′ sequence (1= Forward; 2 = Reverse)
FL-HelQ	To insert XhoI/XbaI restriction sites for cloning of hexaHis-SUMO-WT HelQ into pACEBacI vectors with StrepII- and MBP-tags.	1:GGCCGGGCACTCGAGATGGCACACCATCACCACCATCAC 2:GGCCGGGCGTCTAGATTATGCTTTATCGGTTGAGCTTGCAACTGC
FL-HelQ	To create a point mutation Phe-974-Ala point mutation in the ratchet domain in MBP-hexaHis-SUMO HelQ.	1:GCTAGCAGCTGTGTTCTGCATTTTTGTGAGGAACTGGAAGAGT 2:GCTTGCGGTGCCGGTCAGCAGATTCTGAATATAGC
FL-HelQ	To create a point mutation of Gln-965-Ala in the ratchet domain in MBP-His-SUMO HelQ.	1:GCGAATCTGCTGACCGGCACCGCAAGCTTTAGCA 2:AATATAGCCACGAGGCATATTGAATTTTTCGCTAACGGTCCAAATATTAGT
FL-HelQ	To create a Walker A motif Lys-365-Ala point mutation into StrepII-hexaHis-SUMO HelQ.	1:GCAACCCTGGTTGCCGAAATTCTGATGCTGCAAGAACTG 2:ACCACCGCTGGTCGGCAGTGAATAAATCAGGTTTTTAC
FL-HelQ	To create a Walker B Asp-463-Ala point mutation into StrepII-hexaHis-SUMO HelQ.	1:CTGAACTGCATATGATTGGTGAAGGTAGCCGTGGTGCAA 2:CAACAACAACCAGACCCAGGCTATCAATACGACCGG
FL-HelQ	To create an N-terminal truncated protein encoding DNA terminating at Ile-240 from FL-HelQ with introduced XhoI/XbaI restriction sites.	1(XhoI cut site): GGCCGGGCACTCGAGATGGCACACCATCACCACCATCAC 2(XbaI cute site): GGCCGGGCGTCTAGACAGATCATACGGGGTGGTCAGATAAATCAGATGCAGCAGGC
pACEBacI	Primers flanking insertion site on pACEBacI bacmid for PCR verification.	1:TGTAAAACGACGGCCAGT 2:CAGGAAACAGCTATGAC
FL-HelQ	Internal forward primers for sequencing.	F2:AGCGGTCTGAGCAGCTTTGG F4:CCCTGGCAGCCGGTGTTAAT F5:CGCAACCGATTTTGCAACCGA F6:ACCTGGTGCCGATCTGCTGT F7:AAGCAAGCGGTCAGGCGATT
FL-HelQ	Internal reverse primers for sequencing.	R1:CCGGACTATCGCTCAGCAGC

N-HelQ	Cloning of N-HelQ from pET14b into pRSF-1b. Protein terminates at Ile240.	1 (XhoI cut site): GGCCGGGCACTCGAGATGGCACACCATC ACCACCATCAC 2 (HindIII cut site): GGCCGGGCAAGCTTCTAGTTATGCGGCAG TTCTTCATTAACGGTATTATGGC
PWI mutant N-HelQ	To create a double point mutation Asp-142-Gly and Phe-143-Gly in N-HelQ PWI predicted regions.	1:ACACGCAACCGGTGGTGCAACCGAAA 2:TTTTTATGTTCCGGCAGCTGC
Truncated N-HelQ	To create a truncated mutant of N-HelQ. Terminates at Val-76.	1:CTCGAGCATATGGCTGCCGCGC 2:GTTGAAGTTCAGCCGCTGCTGCTGA
Phosphomimetic PWI N-HelQ	To create a phosphomimetic N-HelQ with point mutations Ser-158-Asp and Ser-178-Asp.	1:CAGTTCGGTCAGATTGCCAATGGTGGTA ATCAGCAGTTTGT 2:CAGACCGATAAACACACAGAAAATCAG CTAGGTTATGAAGGTGTTAC
HelQ C-terminal domain	To create a C-terminal fragment of HelQ (C-HelQ) with Sal1 and Not1 restriction sites from FL-HelQ.	1 (Sal1 cut site): GTTACGTCGACATGAATGCAAAAGCACA GACCCCGATTTTTAGC 2 (Not1 Cut site): GTTCAGCGGCCGCTTATGCTTTATCGGTT GAGCTTGCAACTGC
C-HelQ	To create a Walker B mutant of C-HelQ at position Asp-463 to make C-HelQ^{D463A}.	As for FL-HelQ.
C-HelQ	To create a Tyr-642-Ala point mutation in the predicted motif IVa in C-HelQ to produce C-HelQ^{Y642A}.	1:TTGGATGATCGCTTTTCGTCAGTTCAGT C 2:TCAGGATTACACTGGCTAAC
N-HelQ N1	To create N-HelQ protein 2 (N2) (PWI only). Protein terminates at Gln-245.	1 (NcoI cut site): GGCCGGGCACCATGGGATGAATGTGGTA GCCGTATTCGTCGTCGTGTTAGCC 2 (BamHI cut site): GGCCGGGCGGGATCCCTACAGCTGCATAT ATTTCTGTTCCAGATCATCAACCTGGGC
N-HelQ N2	To create N-HelQ protein 3 (N3) (PWI+ helicase 1). Protein terminates at Lys-540.	1 (NcoI cut site): GGCCGGGCACCATGGGATGAATGTGGTA GCCGTATTCGTCGTCGTGTTAGCC 2 (BamHI cut site): GGCCGGGCGGGATCCCTAACGACGTTTGG TAGGCGGAAAACGACC
N-HelQ N3	To create protein 4 for N-HelQ crystal trials (PWI + helicase 1 and 2). Protein terminates at Ile-720.	1 (NcoI cut site): GGCCGGGCACCATGGGATGAATGTGGTA GCCGTATTCGTCGTCGTGTTAGCC 2 (BamHI cut site): GGCCGGGCGGGATCCCTAGTGTTTCAGGT ACTCTTTGCTCAGGAATTTACAAATCATTT CGGC (65ºC)
N-HelQ	To clone N-HelQ DNA into a BioID BIR plasmid.	1 (KpnI cut site): GGCCGGGCAGGTACCGCATGGCACACCA TCACCACCATCACGGGA

RPA14	To clone RPA14 DNA into pTRG plasmid for B2H system.	1 (BamHI cut site): GGCCGGGCAGGATCCATGGTCGTCATCTACCTGCCTGGACAG 2 (XhoI cut site): GGCCGGGCGCTCGAGTTAATGCAGCTGGCACGACAGTTTCCC
RPA70	To clone RPA70 DNA into pTRG plasmid for B2H system.	1 (BamHI cut site): GGCCGGGCAGGATCCATGGTCGGCCAGCTGAGCGA 2 (XhoI cut site): GGCCGGGCGCTCGAGTCACATCAATGCACTTCTCCTGATGCTCATGACC

Table 2.4 Oligonucleotides to make DNA substrates for helicase assays and EMSAs.

DNA substrate	Composite strands	Extinction co-efficient L/(mole.cm)	Use	5′-3′ sequence
Fork 2a	MW12	461,700	Normal helicase activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	5′-Cy5-MW14	505,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Fork 2b	5′-Cy5-MW12	461,700	Normal helicase activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	495,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Methyl-Phosphonate	5′-Cy5-MePhos-1	496,400	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGCATCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14+T	505,100		CAACGTCATAGACG**A**TTACATTGCTACATGGATGCTGTCTAGAGGATCCGA
Abasic-1	Abasic-1	483,000	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGCTCCATGATCACTG*GCACTGGTAGAATTCGGC
	MW14	495,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Abasic-3	Abasic-3	483,000	Helicase activity in a chemically	GTCGGATCCTCTAGACAGCTC*CATGATCACTGGCACTGGTAGAATTCGGC

	MW14	495,600	modified background.	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Reverse Abasic-3	5′-Cy5-MW12	461,700	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14-Rv-Ab3	485,500		CAACGTCATAGACGATTACATTGCTACATG*AGCTGTCTAGAGGATCCGA
Reverse Methyl Phosphonate	5′-Cy5-MW12	496,400	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGC**A**TCCATGATCACTGGCACTGGTAGAATTCGGC
	MW13	453,100		TGCCGAATTCTACCAGTGCCAGTGATGGACATCTTTGCCCACGTTGACC
Phosphorothioate-1	Phosphoro-1	483,000	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	495,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
Phosphorothioate-2	Phosphoro-2	483,000	Helicase activity in a chemically modified background.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	495,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
BamHI^{EIIA} Roadblock	ELB303	492,800	Determining protein displace-ment activity.	ATCGACCTAGGGATCCGGTGCAATTCTTTTTTTTTTTTTTTTTTTTTTT
	5′-Cy5-ELB302	505,800		TTTTTTTTTTTTTTTTTTTTTTTTTTTTGAATTGGACCGGATCCCTAGGTCGAT
Extended Fork	ELB305	763,300	Normal helicase activity.	GCAAGCCTTCTACAGGTCGAATTGTGAGCGGATAACAATTCCGCAATACGGATTTTTTTTTTTTTTTTTTTTTTTTTT
	5′-Cy5-ELB304	770,300		TTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGTATTGCGGAATTGTTATCCGCTCACAATTCGACCTGTAGAAGGCTTGC
Biotin 3 Fork	5′-Cy5-MW12-Biotin 1	491,100	Determining protein displace-	GTCGGA**T**CCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC

	MW14 Biotin 3	485,500	ment activity.	CAACGTCATAGACGA TTACATTGCTACATG GAGCTGTCTAGAGAT CCGA
Bubble	5′-Cy5-BUB1	375,300	Determining RPA displace-ment activity.	CGAATCGATAGTCTC TTTTTTTTCGATCAG ATGCGATC
	BUB2	363,300		GATCGCATCTGATCG TTTTTTTTTAGACTA TCGATTCG
G4 Quadruplex	ADDx3 POLIA 3′	575,100	Helicase activity.	TCGCCACGTTTCGCC GTTTGCGGGGGTTTC TGCGAGGAACTTTGG AAAAAAAAAAAAA
J6 Holliday junction (HJ)	RGL-16	449,300	Helicase activity.	GACGCTGCCGAATTC TGGCTTGCTAGGACA TTCTTTGCCCACGTTG ACCC
	5′-Cy5-ELB-21	449,300		GACGCTGCCGAATTC TGGCTTGCTAGGACA TTCTTTGCCCACGTTG ACCC
	ELB-22	491,800		GGGTCAACGTGGGCA AAGAATGTCCTACGT CCGATACGGATAATC GCCAT
	ELB-23	492,800		ATGGCGATTATCCGT ATCGGACGTAGGACA TGCTGTCTAGAGACT ATCCA
Re-annealing substrates 1 5′ overhang	5′- Cy5-vDNA ts4	597,500	End joining activity.	GCCTTTAACTCCCTG AAGACCTGATTTTTG ATTTATGGTCATTCTC GTTTTTGCGTTTGACA AAC
	vDNA bs11	656,600		ATGAGTGTTGTTCCA GTTTGTCAAACGCAA AAACGAGAATGACCA TAAATCAAAAATCAG GTCTT
Re-annealing substrate 2 5′ overhang	vDNA ts5	617,600	End joining activity.	TGGAACAACACTCAT ATAATGTAACTCAAT AAGCCTCTCGCTATTT TGGTTTTTATCGTCGT CTG
	vDNA bs12	672,300		GAGACCTGGAGCAAA CAGACGACGATAAAA ACCAAAATAGCGAGA GGCTTATTGAGTTAC ATTAT
Re-annealing substrate 3 3′ overhang	5′-Cy5-vDNA ts4.2	617,500	End joining activity.	AAGACCTGATTTTTG ATTTATGGTCATTCTC GTTTTTGCGTTTGACA AACTGGAACAACACT CAT
	vDNA bs11.2	511,300		GTTTGTCAAACGCAA AAACGAGAATGACCA TAAATCAAAAATCAG GTCTT

Re-annealing substrate 4 3′ overhang	vDNA ts5.2	470,000	End joining activity.	ATAATGTAACTCAATAAGCCTCTCGCTATTTTGGTTTTTATCGTCGTCTG
	vDNA bs12.2	663,500		CAGACGACGATAAAAACCAAAATAGCGAGAGGCTTATTGAGTTACATTATATGAGTGTTGTTCCA
Fork 3	5′-Cy5-MW12	461,700	Super-shifting activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	MW14	495,600		CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	PM16	245,900		TGCCGAATTCTACCAGTGCCAGTGAT
Duplex DNA (MW14)	5′-Cy5-MW14	495,600	Helicase activity.	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	CH4	476,600		GCCGAATTCTACCAGTGCCAGTGATCATGGAGCTGTCTAGAGGATCCGA
Duplex DNA (MW12)	5′-Cy5-MW12	461,700	Helicase activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	HelQ5	484,100		TGCCGAATTCTACCAGTGCCAGTGATCATGGAGCTGTCTAGAGGATCCGA
Partial duplex 3′ end	5′-Cy5-MW12	461,700	Helicase activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	PM16	245,900		TGCCGAATTCTACCAGTGCCAGTGAT
Partial duplex 5′ end	5′-Cy5-MW12	461,700	Helicase activity.	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC
	PM11	239,500		AGCCTAGGAGATCTGTCGAGGTAC
Cas9 binding fork	5′-Cy5-MW14	505,600	Determining protein displace-ment activity.	CAACGTCATAGACGATTACATTGCTACATGGAGCTGTCTAGAGGATCCGA
	MW12-Cas9	473,700		CGCCTTAAGATGCTTGGCTCCATGTAGCAATGTAATCGTCTATGACGTTG
D-loop	5′-Cy5-RGL19	549,500	Helicase unwinding	GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCC

			control for RNA R-loop.	ACCTGCAGGTTCACCC
	PM5	213,000		AAAGATGTCCTAGCAAGGCA
	PM4	582,800		GGGTGAACCTGCAGGTGGGCGGCTGCTCATCGTAGGTTAGTTGGTAGAATTCGGCAGCGTC
R-loop	5′-Cy5-RGL19	549,500	Helicase unwinding through RNA.	GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC
	ELB1-RNA	209,100		AAAGAUGUCCUAGCAAGGCAC
	PM4	582,800		GGGTGAACCTGCAGGTGGGCGGCTGCTCATCGTAGGTTAGTTGGTAGAATTCGGCAGCGTC
RNA-DNA hybrid 1	crRNA1	488,100	Helicase unwinding through RNA with RNA flap.	CCAUACCACACAUUUGGAUUGUAGGUGGUUGAUGUUAAUUUCAAUCCCAUUUUG
	ELB104	365,900		TAACATCAACCACCTACAATCCAAATGTGTGGTATGG
DNA-RNA hybrid 2	crDNA1	515,400	Helicase unwinding through RNA with DNA flap.	CCATACCACACATTTGGATTGTAGGTGGTTGATGTTAATTTCAATCCCATTTTG
	ELB104-RNA	355,500		UAACAUCAACCACCUACAAUCCAAAUGUGUGGUAUGG
DNA-DNA cr fork	5′-Cy5-TJ2-DNA	519,000	Helicase unwinding control for RNA R-loop.	GTTTTACCCTAACTTTAATTGTAGTTGCACCTACAATCCAAATGTGTGGTATGG
	crDNA1	515,400		CCATACCACACATTTGGATTGTAGGTGGTTGATGTTAATTTCAATCCCATTTTG
DNA-RNA cr fork	5′-Cy5-TJ2-DNA	519,000	Helicase unwinding through RNA hybrid fork.	GTTTTACCCTAACTTTAATTGTAGTTGCACCTACAATCCAAATGTGTGGTATGG
	crRNA1	488,100		CCAUACCACACAUUUGGAUUGUAGGUGGUUGAUGUUAAUUUCAAUCCCAUUUUG
DNA fork with RNA lagging strand	5′-Cy5-MW12	461,700	Helicase unwinding of DNA fork with lagging	GTCGGATCCTCTAGACAGCTCCATGATCACTGGCACTGGTAGAATTCGGC

	MW14	495,600	strand flap with RNA.	CAACGTCATAGACGA TTACATTGCTACATG GAGCTGTCTAGAGGA TCCGA
	TJ3-RNA	246,100		UAGCAAUGUAAUCG UCUAUGACGUUG

2.5 Plasmids and vectors

Table 2.5 Vector backbones used for cloning.

Plasmid	Size	Resistance	Features
pACEBacI	2904bp	$Gent^R$	ColE1 origin, polh promoter, SV40 and HSVtk terminators, MCS, Tn7L, Tn7R transposition elements.
pET14b	4671bp	Amp^R	pBR322 origin, T7 promoter/terminator, hexaHis-tag coding sequence, MCS (NdeI-BamHI)
pRSF-1b	3669bp	Kan^R	RSF origin, T7 promoter/terminator, hexaHis-tag coding sequence, MCS (NcoI-XhoI)
pET28b	5368bp	Kan^R	pBR322 origin, T7 promoter/terminator, 3 possible ORF
pETM10	5528bp	Kan^R	pBR322 origin, T7 promoter/terminator, TEV cleavable N-terminal MBP tag, MCS (KpnI-XhoI)
pACYCDuet1	4008bp	$Chlm^R$	p15A origin, T7 promoter/terminator, hexaHis-tag coding sequence, MCS (NcoI-NotI)
pBadHisA	4103bp	Amp^R	pUC origin, araBad promoter/terminator, N-terminal hexaHis tag, MCS (XhoI-HindIII)

Table 2.6 Plasmids used and made for cloning.

Name	Description	Resistance	Origin
pSN52	N-terminal hexaHis-SUMO-HelQ.	Amp^R	Sarah Northall PhD
pTJ01	N-terminal MBP-hexaHis-SUMO- HelQ DNA cloned into pACEBacI background.	$Gent^R$	This work
pTJ03	N-terminal StrepII-hexaHis-SUMO-HelQ DNA cloned into pACEBacI background.	$Gent^R$	This work
pTJ04	N-terminal MBP-hexaHis-SUMO- HelQ DNA cloned into pACEBacI background with point mutation in domain 4 ratchet Phe-974-Ala.	$Gent^R$	This work
pTJ05	N-terminal MBP-hexaHis-SUMO- HelQ DNA cloned into pACEBacI background with point mutation in domain 4 ratchet Gln-965-Ala.	$Gent^R$	This work
pTJ06	N-terminal MBP-hexaHis-SUMO-HelQ DNA cloned into pACEBacI background with point mutation in Walker A domain Lys-479-Ala.	$Gent^R$	This work
pTJ07	N-terminal MBP-hexaHis-SUMO-HelQ DNA cloned into pACEBacI background with point mutation in Walker B domain Asp-463-Ala. FL-$HelQ^{D463A}$.	$Gent^R$	This work
pTJ08	RPA14 subunit in pBadHisA plasmid background.	Amp^R	This work

pTJ09	N-Terminal hexaHis-SUMO tagged N-terminal terminated at Ile-240 of HelQ in pET14b background.	Amp^R	This work
pTJ10	PWI mutant N-HelQ hexaHis-SUMO tagged (N-$HelQ^{D142A/F143A}$).	Amp^R	This work
pTJ11	Truncated N-HelQ hexaHis-SUMO tagged.	Amp^R	This work
pHB01	HexaHis tagged C-terminal HelQ in pACYCduet background.	$Chlm^R$	Hannah Betts Masters project
pTJ12	Point mutation in the Walker B motif of C-HelQ (derived from pHB01) C-$HelQ^{D463A}$.	$Chlm^R$	This work
pTJ13	Phosphomemetic mutant N-HelQ (N-$HelQ^{S157/177D}$).	Amp^R	This work
pTJ14	N-terminal hexaHis-SUMO tagged N-HelQ truncated at Ile-240 in pRSF-1b background.	Kan^R	This work
pTJ15	Y642A C-HelQ mutant in pACYCduet background to give C-$HelQ^{Y642A}$.	$Chlm^R$	This work
pTJ16	TEV cleavable StrepII- tagged N-HelQ PWI only DNA terminating at position 244 subcloned from pTJ9 into pET28b-SF (SF-N1).	Kan^R	This work
pTJ17	TEV cleavable StrepII- tagged N-HelQ PWI region + RecA-like domain 1 DNA terminating at position 539 subcloned from pTJ9 into pET28b-SF (SF-N2).	Kan^R	This work
pTJ18	TEV cleavable StrepII tagged N-HelQ PWI +RecA-like 1 and 2 DNA terminating at position 719 subcloned from pTJ9 into pET28b-SF (SF-N3).	Kan^R	This work
pTJ19	TEV cleavable MBP tagged N-HelQ PWI only DNA terminating at position 244 subcloned from pTJ9 into pETM40 (MBP-N1).	Kan^R	This work
pTJ20	TEV cleavable MBP tagged N-HelQ PWI + RecA-like domain 1 DNA terminating at position 539 subcloned from pTJ9 into pTEM40 (MBP-N2).	Kan^R	This work
pTJ21	N-HelQ DNA cloned into pBT vector background (B2H system).	$Chlm^R$	This work
pTJ22	RPA14 DNA cloned into pTRG vector background (B2H).	Tet^R	This work
pTJ23	RPA70 DNA cloned into pTRG vector background (B2H).	Tet^R	This work
pTJ24	N-HelQ with PWI-like mutation DNA cloned into pBT vector background (B2H).	$Chlm^R$	This work

2.6 Commercial and non-commercial enzymes

Table 2.7 Commercial enzymes used in this work.

Name	Supplier	Features
XhoI	NEB	Cuts sequence: 5′-C*T C G A G-3′ 3′-G A G C T*C-5′

XbaI	NEB	Cuts sequence: 5′- T*C T A G A- 3′ 3′- A G A T C*T- 5′
BamHI	NEB	Cuts sequence: 5′-G*G A T C C-3′ 3′-C C T A G*G-5′
BamHI^{EIIIA}	NEB	Nuclease inactive BamHI, binds: 5′-G G A T C C-3′ 3′-C C T A G G-5′
DpnI	NEB	Digests methylated DNA to remove genomic contaminates.
CIP	NEB	Calf intestinal alkaline phosphatase catalyses the dephosphorylation of 3′ and 5′ ends of DNA for ligation.
LacI	NEB	Binds 5′-AATTGTGAGCGGATAACAATT-3′
HindIII	NEB	Cuts sequence: 5′-A*A G C T T-3′ 3′-T T C G A*A-5′
Terminal transferase	NEB	To fix a Cy5 fluorescent moiety onto DNA oligos.
T4 DNA ligase	NEB	Ligate plasmids for transformation.
Vent polymerase	NEB	Standard PCR polymerase enzyme.
Q5 DNA polymerase	NEB	Mutagenesis PCR polymerase enzyme.

Table 2.8 Non-commercial enzymes.

Name	Supplier	Features/ Use
E. coli PriA	In lab purified recombinant protein	Control proteins in assays.
E. coli REP	In lab purified recombinant protein	
E. coli RecQ	In lab purified recombinant protein	
E. coli RecG	In lab purified recombinant protein	
E. coli SSB	In lab purified recombinant protein	
BamHI^{EIIIA}	NEB MTA Material	Protein roadblock.
Streptavidin	Sigma- Aldrich #189730	Protein roadblock.
Aro1-RPA	Kindly supplied by the lab of Robert M. Brosh, Jr; National Institutes of Health, Baltimore.	Two residue substitutions in the DNA binding domain of RPA70. Phe-238 + Trp-361.
Human RPA	EnzyMax, LLC #61	Purified recombinant trimeric protein.

Table 2.9 Primary and secondary antibodies used in protein immuno-blotting.

Antibody	Supplier	Description
1º Anti-hexaHis	ThermoFisher #MA1-21315-BTIN	Mouse monoclonal, Biotin conjugate
1º Anti-SUMO1	ABCAM #ab32058	Rabbit monoclonal
1º Anti-StrepII	Caltag Medsystems #M211-3B	Rabbit polyclonal
1º Anti-HelQ	Merckmillipore #MABC955	Mouse monoclonal
2º Anti-Biotin	Cell Signalling #7075P5	HRP-linked AB targeting Biotinylated ladder
2º Anti-Mouse	Invitrogen goat anti-Mouse IgA #62-6720	Secondary antibody
2º Anti-Rabbit	Invitrogen stabilised peroxidase conjugated goat anti- Rabbit #32460	Secondary antibody

2.7 Solution composition

Table 2.10 Composition of commonly used laboratory gels.

Gel Type	Percentage	Components
Native protein gel	5%	28.3 mL dH_2O 11.68 mL 30% acrylamide (SIGMA #A7168-100ML) 4.5 mL 10x TBE 0.45 mL 10% APS and 22 µL TEMED
	10%	20.9 mL dH_2O 14.6 mL 30% acrylamide 4.5 mL 10x TBE 0.45 mL 10% APS and 22 µL TEMED
SDS Separating gel	8%	6.18 mL dH_2O 3 mL 30% acrylamide 1.4 mL 3 M Tris pH 8.8 112 µL 10% SDS 84 µL 10% APS and 9.5 µL TEMED
	10%	5.43 mL dH_2O 3.75 mL 30% acrylamide 1.4 mL 3 M Tris pH 8.8 112 µL 10% SDS 84 µL 10% APS and 9.5 µL TEMED
	12%	4.48 mL dH_2O 4.7 mL 30% acrylamide 1.4 mL 3 M Tris pH 8.8 112 µL 10% SDS 84 µL 10% APS and 9.5 µL TEMED
SDS Stacking gel	5%	1.75 mL dH_2O 0.5 mL 30% acrylamide 0.75 mL 0.5 M Tris pH 6.8 30 µL 10% SDS 30 µL 10% APS and 3 µL TEMED
Native PAGE gel	8%	5.43 mL dH_2O 3.75 mL 30% acrylamide 1.4 mL 3 M Tris pH 8.8 84 µL 10% APS and 9.5 µL TEMED
Agarose gel	1%	1 g agarose in 100 mL 1x TBE

Table 2.11 Composition of commonly used laboratory buffers.

Buffer	Use	Components
50 x TAE	Agarose gels	2 M Tris 1 M acetic acid 50 mM EDTA
10 x TBE	Agarose gels	1 M Tris 1 M boric acid 20 mM EDTA
10 x LIS	Native PAGE EMSA gels	67 mL 1 M Tris HCL pH 8.0 66 mL 0.5 M sodium acetate 100 mL 0.2 M EDTA pH 8.0 To 1 L SDW
10 x SDS running buffer	SDS PAGE	250 mM Tris 1.92 M glycine 1% SDS (v/v)

6 x SDS PAGE loading dye	SDS PAGE samples	375 mM Tris-HCL pH 6.8 6% SDS 48% (v/v) glycerol 5 mM DTT 0.03% (w/v) bromophenol blue
1 x Coomassie blue stain buffer	SDS PAGE	1 g Brilliant blue 100 mL acetic acid 400 mL methanol To 1 L with SDW
1x Coomassie blue de-stain buffer	SDS PAGE	100 mL acetic acid 300 mL methanol To 1 L with SDW
2x Native PAGE running buffer	Native PAGE protein analysis	6 g Tris 28.8 g glycine To 1 L with SDW
5x Orange G gel loading dye	Native EMSA TBE gels	80% glycerol Orange G powder
5x Helicase buffer (HB)	Helicase assays	100 mM Tris HCL pH 8 0.5 mg/mL BSA 0.5 M NaCl
STOP solution	Helicase assays	2.5% SDS 200 μM EDTA 2 μg/μL proteinase K
Mu Broth	*E. coli* cell protein over-expression	10 g/L tryptone 10 g/L NaCl 5 g/L yeast extract 2 mM NaOH
SOC Broth	Bacmid purification	2% tryptone 0.5% yeast extract 10 mM NaCl 2.5 mM KCL 10 mM $MgCL_2$ 10 mM $MgSO_4$ 20 mM glucose
Agar	Bacterial transformation	3 g agar per 200 mL Mu broth
M9 Minimal Media (10x)	Transformations and grow ups	64 g $Na2HPO_4$-$7H_2O$ 015 g KH_2PO_4 2.5 g NaCl 5.0 g NH_4CL To 1 L SDW, autoclaved.
1x PBS	Gel filtration, biophysical methods.	8 g NaCl 0.2 g KCL 1.44 g Na_2HPO_4 0.24 g KH_2PO_4 To 1 L SDW pH 7.4
10 x TG Buffer	Component of Western blot transfer buffer	30.3 g Tris 144 g glycine To 1 L with SDW
Transfer Buffer	Protein immunoblot	100 mL 10x TG 200 mL methanol To 1 L with SDW
10 x TBS	Protein immunoblot	90 g NaCl 60 g Tris To 1 L with SDW
1 x TBST	Protein immunoblot	100 mL 10x TBS 200 μL Tween20, to 1 L SDW

Table 2.12 Composition of specialised buffers.

Method	Buffer	Components
Silver Staining	Fix buffer	40% methanol 10% acetic acid
	Sensitise buffer	0.2% w/v sodium thiosulphate 30% methanol 68 g/L sodium acetate
	Silver stain buffer	0.25% $AgNO_3$
	Developing buffer	25 g/L sodium carbonate 0.00027% w/v sodium thiosulphate 0.014% v/v formaldehyde
	Stop buffer	14.6 g/L EDTA
Malachite Green ATPase assays	Colour reagent	3:1 malachite green hydrochloride (0.045% aqueous solution) to ammonium molybdate (4.2% in 4M HCL)
	Developing reagent	34% aqueous sodium citrate

Table 2.13 Buffers and media used in BacterioMatch II system.

Media	Components
M9 Additives Solution I	10 mM 20% glucose (filter sterilised) 5 mL 20 mM adenine HCL (filter sterilised) 50 mL 10x His drop out amino acid supplement (BD/Clontech, #630415) (autoclaved)
M9 Media Additives Solution II	0.5 mL 1 M $MgSO_4$ 0.5 mL 1 M Thiamine HCL 0.5 mL 10 mM $ZnSO_4$ 0.5 mL 100 mM $CaCl_2$ 0.5 mL 50 mM IPTG; (all components individually filter sterilised).
M9 Media Additives	Prepare solution I and II separately and mix 1:1.
10 x M9 Salts (1 L)	67.8 g Na_2HPO_4 30 g KH_2PO_4 5 g NaCl 10 g NH_4Cl To 1 L with sterile H_20 and autoclaved.
M9+ His-dropout Broth (500 mL)	380 mL sterile H_2O 50 mL 10x M9 Salts 1x preparation of M9 media additives (67.5 mL) Stored at 4°C for up to one month.
Nonselective Screening Medium (500 mL)	380 mL sterile H_2O + 7.5 g agar (autoclaved and cooled to 50°C) Add: 50 mL 10x M9 Salts 1x preparation of M9 Medial Additives (67.5 mL) 0.5 mL 25 mg/mL chloramphenicol 0.5 mL 12.5 mg/mL tetracycline Plates poured and stored at 4°C for up to 1 month.
3-AT Stock Solution 1 M (10 mL)	840.8 mg 3-AT (sigma, #A-8056) in 10 mL DMSO Store at -20°C
Selective Screening Medium (5 mM 3-AT) (500 mL)	380 mL sterile H_2O + 7.5 g agar (autoclaved and cooled to 50°C) Add: 50 mL 10x M9 Salts 1x preparation of M9 Medial Additives (67.5 mL) 0.5 mL 25 mg/mL chloramphenicol 0.5 mL 12.5 mg/mL tetracycline 2.5 mL 1 M 3-AT Plates poured and stored at 4°C for up to 1 month.

Table 2.14 Composition of protein purification buffers.

Name	Use	Components
Nickel charge buffer	To charge Ni-NTA HiTrap column	0.2 M nickel (III) chloride
Ammonium sulphate resuspension buffer	Resuspension buffer for the pellet of HelQ after the ammonium sulphate cut of HelQ	50 mM NaCl 20 mM imidazole 50 mM Tris pH 8 10% glycerol 1 mM DTT 1x EDTA-free Protease Inhibitor Tablet (#11836153001)
Ni-NTA A	Ni-NTA HiTrap column wash buffer	20 mM Tris pH 8 10% glycerol 1 M NaCl 5 mM imidazole 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet
Ni-NTA B	Ni-NTA HiTrap column elution buffer	20 mM Tris pH 8 10% glycerol 1 M NaCl 1 M imidazole 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet
Dialysis buffer 1	Post Ni-NTA HiTrap column to remove high salt	50 mM Tris pH 8 10% glycerol 150 mM NaCl 25 mM DTT
Heparin A	Heparin Column/ Q-Sepharose wash buffer	50 mM Tris pH 8 10% glycerol 150 mM NaCl 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet
Heparin B	Heparin Column/ Q-Sepharose elution buffer	50 mM Tris pH 8 10% glycerol 1 M NaCl 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet
Dialysis storage buffer	To remove high salt; final buffer conditions of HelQ	50 mM Tris pH 8 30% glycerol 150 mM NaCl 25 mM DTT
AGF buffer/ Size exclusion	Superdex200 wash buffer	50 mM Tris pH 8 10% glycerol 150 mM NaCl
RCaH lysis buffer	Pellet resuspension MBP Trap/ StrepII-Trap	600 mM NaCl 50 mM Tris pH 7.5 10% glycerol 5 mM TCEP 1x EDTA-free Protease Inhibitor Tablet
RCaH binding buffer	MBP Trap/ StrepII-Trap wash buffer	600 mM NaCl 50 mM Tris pH 7.5 10% glycerol 5 mM TCEP
RCaH MBP elution buffer	MBP Trap/ StrepII-Trap elution buffer	600 mM NaCl 50 mM Tris pH 7.5 10% glycerol 5 mM TCEP 20 mM maltose

RCaH StrepII elution buffer	MBP Trap/ StrepII-Trap elution buffer	600 mM NaCl 50 mM Tris pH 7.5 10% glycerol 5 mM TCEP 2.5 mM d-Desthiobiotin (Iba)
RCaH AGF binding buffer	Sepharose 6 wash buffer	200 mM NaCl 10 mM Tris pH 7.5 1 mM TCEP
HI-Buffer	RPA purification base buffer	30 mM HEPES 10% glycerol 1 mM DTT 0.02% Tween20 0.25 mM EDTA
Butyl Sepharose buffer A	Hydrophobic column wash buffer	50 mM Tris pH 8 10% glycerol 150 mM $(NH_4)_2SO_4$ 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet
Butyl Sepharose buffer B	Hydrophobic column elution Buffer	50 mM Tris pH 8 10% glycerol 1.5 M $(NH_4)_2SO_4$ 10 mM DTT 1x EDTA-free Protease Inhibitor Tablet

Table 2.15 Columns used in protein purification.

Chromatography Column	Description	Company
HiTrap Chelating HP (5 mL)	Metal affinity, hexaHis tagged proteins.	GE Healthcare Life Sciences (#17040901)
HiTrap Heparin (5 mL)	DNA-binding protein purification.	GE Healthcare Life Sciences (#17040701)
HiTrap Q HP anion exchange (1 mL)	Q-sepharose for strong anion exchange	GE Healthcare Life Sciences (#29051325)
Superdex 200 Increase 10/300 GL (24 mL)	Gel filtration, MW range: M_r 10000 to 600000.	GE Healthcare Life Sciences (#28990944)
Superose 6 Increase 10/300 GL (18 mL)	Size exclusion with MW range M_r 5K to 5M.	GE Healthcare Life Sciences (#29091596)
HiTrap IMAC HP Column (1 mL)	Immobilised metal ion affinity for poly-histidine tagged proteins.	GE Healthcare Life Sciences (#17092003)
Affi Gel® blue resin	Crosslinked agarose bead with covalently attached Cibacron Blue dye for purification of proteins via ionic, hydrophobic, aromatic or sterically active binding sites.	Bio-Rad (#1537301)
Hydroxyapatite (HAP)	Dehydrated power for RPA purification	Sigma Aldrich (#289396)
MonoQ 5/50 GL	Strong anion exchange	GE Healthcare Life Sciences (#17516601)
Butyl Sepharose HiTrap Butyl/FF (5 mL)	Aliphatic hydrophobic interaction purification.	GE Healthcare Life Science (#17-1357-01)
Strep-Tactin® XT superflow column (5 mL)	One-step purification of recombinant StrepII-tag fusion proteins using biotin.	Iba Solutions for Life Science (#2-4026-001)
Gravity Flow Strep-Tactin® Sepharose mini-column	Mini column (0.2 mL) for anti-StrepII affinity tag pull-down assays.	Stratech (#2-1202-550-IBA)

2.8 Bacterial molecular biology

2.8.1 Agarose gel electrophoresis

Agarose was dissolved in 1x TBE buffer (10x stock: 1 M Trizma base, 1 M Boric acid, 0.02 M EDTA) by boiling and cooled to touch. Ethidium bromide (EtBr) (0.2 µg/mL) was added to visualise DNA and dispersed evenly. Electrophoresis was carried out using a BioRad Cell electrophoresis tank and BioRad PowerPac power supply. Gel was run in 1x TBE at 120 V, constant current for 90 minutes. DNA was visualised using a U:Genius Bio-imaging system (Syngene).

2.8.2 Preparation of bacterial agar plates

Sterile agar made up of 3 g agar powder per 200 mL LB liquid broth. LB agar was sterilised by autoclaving and left to cool to below 50°C. Appropriate antibiotics to the correct concentration were added (see table 2.1). Plates contained 25 mL of agar per petri dish using aseptic technique and dried at 60°C.

2.8.3 Preparation of competent cells

2.8.3.1 Chemically competent cells

Bacterial glycerol stocks stored at -80°C were used to streak out bacterial cell strains on LB agar to make competent cells for transformations carried out in this work. 5 mL cultures were inoculated from single colonies overnight at 37°C in LB containing appropriate antibiotic selection. These overnight cultures were used to inoculate larger 50 mL LB cultures using a 1 in 50 dilution. Cultures were incubated at 37°C with shaking until an optical density (OD_{600}) of 0.4 (range of 0.3-0.5) was reached. OD was measured using a Spectronic 20+ (ThermoScientific). Cells were harvested by centrifugation at 3381 x *g* for 20 minutes at 4°C to pellet the cells (Centrifuge 5430R from Eppendorf). Cells were resuspended in 6.25 mL ice-cold sterile calcium chloride (0.1 M $CaCl_2$) and incubated on ice for 4 hours, supernatant

was discarded. Cells were harvested again by centrifugation and re-suspended in fresh sterile calcium chloride. Glycerol was added to 30% v/v. Competent bacterial cells were flash frozen on dry ice for 5 minutes and stored at -80ºC. The transformation efficiency was tested prior to the use of these cells. Table 2.2 refers to antibiotic selection for bacterial competent cells.

2.8.3.2 Electro-competent cells

As for chemically competent cells, glycerol stocks were used to streak out bacterial cell lines onto LB agar to make competent cells. 5 mL LB broth was inoculated from single colonies overnight at 37ºC containing appropriate antibiotic selection. 50 mL LB broth containing appropriate antibiotic selection was inoculated with 1 in 50 dilution from overnight pre-cultures. Culture was incubated at 37ºC with shaking until an optical density (OD_{600}) of 0.4 (range of 0.3-0.5) was reached. Cells were chilled on ice and harvested using centrifugation at 4000 x *g* for 10 minutes at 4ºC. Supernatant was discarded and the cells were resuspended in 50 mL ice-cold SDW. The cells were re-spun and washed two more times in sterile 10% glycerol to pellet the cells. Cells were pelleted at maximum speed for 15 seconds, the supernatant discarded and resuspended in 800 µL ice-cold sterile 10% glycerol. Competent bacterial cells were flash frozen on dry ice for 5 minutes and stored at -80ºC.

2.8.4 Plasmid construction

2.8.4.1 Polymerase Chain Reaction (PCR)

Forward and reverse primers were designed using SNAPGENE to amplify the target gene. Transfer of genes into vector backbones required introduction of appropriate restriction sites using primers that flanked the gene of interest. Primers were designed with linker regions of six bases after the inserted restriction site.

Restriction sites used to insert fragments were chosen based on the multiple cloning site of the vector backbone of interest.

PCR reactions of 50 µL were assembled to contain 50 ng template DNA, 100 pM forward primer, 100 pM reverse primer, 2 mM dNTP mix, 1x reaction buffer and 1 µL appropriate polymerase. Vent DNA polymerase (NEB #M0254S) was used for standard PCR and Q5 NEB (E0554S) was used for mutagenesis.

Table 2.16 PCR thermocycling conditions.

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	2 minutes	x1
Denaturation	98°C	2 minutes	x 30
Annealing	68°C (dependent on primer Tm)	45 seconds	
Extension	72°C	30 seconds/ kb	
Final extension	72°C	7 minutes	x 1

Melting and annealing temperatures and times were extended based on primer specificity and PCR optimisation. Other optimisation steps included addition of DMSO (4-10%), $MgCl_2$ (5 mM) and GC enhancer buffer. PCR was carried put using a Verti 96-well thermocycler (Applied Biosystems®). Table 2.3 lists the primers used in this work for gene amplification.

The PCR product was confirmed at the correct band size by running on a 1% agarose gel. NEB 1 kb ladder was used as a marker (#N3232S). PCR product was purified using Qiagen QIAquick PCR clean up kit (#28106) and eluted in 50 µL Tris pH 8.6 warmed to 65°C. In the presence of non-specific PCR amplification, PCR DNA products were gel extracted using the QIAquick Gel Extraction Kit. The PCR DNA products were excised in a dark room using a transilluminator 2040 EV (Stratagene).

2.8.4.2 Enzyme digestion of DNA vector and insert

DNA vector backbone and purified PCR products were digested with appropriate restriction enzymes chosen during cloning design. Digestion reactions were assembled including 1.5 ng PCR product or 1 ng vector, 2.5 µL enzyme 1, 2.5

µL enzyme 2 and 1x CUTsmart buffer (NEB) to 100 µL with SDW. The digestion reaction was incubated at 37ºC for 3 hours and inactivated at 65ºC for 10 minutes. 100 µL digested insert and vector were run on a 1% agarose gel; the bands were excised, and gel purified as described previously.

2.8.4.3 DNA vector de-phosphorylation

Gel purified vector backbone was de-phosphorylated in 1x CUTSmart buffer (NEB) with 1 µL CIP enzyme (NEB) at 37ºC for 1 hour. The reaction was deactivated at 65ºC for 15 minutes.

2.8.4.4 Ligation of DNA vector and insert

Ligation reaction conditions were calculated using NEBioCalculator.[266] Reactions of 1:2 and 1:3 ratio of DNA vector to PCR insert were used with 50 ng vector DNA. Reactions consisted of 1x T4 ligase buffer (NEB), fresh 25 mM ATP (pH 8) and 1 µL T4 ligase (NEB). Ligations were incubated at 16ºC overnight prior to transformation into DH 5α competent cells. Control reactions were carried out with excluded insert DNA.

2.8.4.5 Ligation transformation into chemically competent cells

Plasmid DNA, from miniprepped purified DNA stocks or ligation products, was added to 100 µL DH 5α *E. coli* cells made chemically competent as described, or NEB 5α cells for more difficult cloning and incubated on ice for 30 minutes. Cells were heat shocked at 42ºC for 45 seconds and returned to ice for a further 2 minutes. Recovery LB medium was added (900 µL) and the cells recovered at 37ºC for 1 hour with shaking. Cells were harvested by centrifugation at 18407 x *g* for 1 minute and re-suspended in 200 µL LB broth. Transformed cells were plated onto

agar plates containing the appropriate antibiotics using ascetic techniques and incubated at 37ºC overnight. The absence of colonies from the control plate was used to indicate successful ligation. 2 to 4 colonies were picked used to inoculate individual 5 mL LB broth containing appropriate antibiotics and incubated at 37ºC overnight.

Transformations for protein over-expression and plasmid maintenance required less DNA (250 ng) for transformation into chemically competent cells.

2.8.4.6 DNA MiniPrep purification

Inoculated 5 mL cultures left overnight were pelleted by centrifugation at 4000 g for 5 minutes. The plasmid DNA was then purified from the pellet using the Qiagen MiniPrep kit following manufacturer's instructions (#27104).

2.8.4.7 Verification of plasmid construction

Restriction digestion was used to verify correct DNA insertion into vectors (table 2.7 details enzymes). Digests were confirmed by agarose gel analysis. Verification by sequencing carried out by Source Biosciences of plasmid insertion sites and mutagenesis sites.

2.8.5 Site-directed mutagenesis (SDM)

2.8.5.1 Primer design

Mutagenesis within this work was carried out on FL-HelQ, N-HelQ and C-HelQ to introduce site-specific point mutations. Primers were designed with a single mismatch base change in the 5′ end of the forward primer. The primers were ordered with HPLC purification. PCR mutagenesis involved the amplification of entire vectors and was run as described previously for PCR using Q5 polymerase and gel extracted.

2.8.5.2 DpnI treatment to remove genome DNA template

DpnI is a type II restriction enzyme that specifically cleaved DNA containing methylated adenine and therefore is used to cleave genomic DNA template. Purified PCR products were treated with DpnI enzyme (NEB) in 1x CUTsmart buffer at 37ºC for 3-4 hours. The reaction was deactivated at 74ºC for 15 minutes. DpnI treated vectors were gel purified from 1% agarose gels.

2.8.5.3 Bacterial transformation of mutagenesis product

Vectors were ligated as described previously prior to transformation into XL-1 Blue or NEB 5α competent cells. Successful mutagenesis was verified using sequencing across mutation site.

2.9 Baculovirus molecular biology

2.9.1 Preparation of S-Gal agar plates

LB agar was supplemented with 0.3 mg/mL S-Gal dissolved in DMF, 0.6 mg/mL iron ammonium citrate, 57 ng/mL IPTG (VWR), 30 µg/mL kanamycin, 7 µg/mL gentamicin and 10 µg/mL tetracycline. Plates were stored wrapped in aluminium foil at 4ºC. Plates were dried before spreading of transformed cells.

2.9.2 Transformation of DH10Bac *E. coli* cells with pACEBacI plasmid

40 ng pACEBacI plasmid containing HelQ encoding DNA was transformed into 100 µL electro-competent DH10Bac TURBO *E. coli* cells. Cells incubated on ice for 30 minutes prior to electroporation using a Gene Pulser (BioRad). Cells were recovered in SOC medium (450 µL) at 37ºC overnight with shaking. Recovered cells were diluted 1 in 20 in SOC medium and plated onto S-gal plates, covered and left at 37ºC for 48 hours.

2.9.3 Bacmid purification

Bacterial blue/white colony screening was used on S-gal agar plates to determine correct plasmid uptake into electro-competent cells. White colonies contained the recombinant bacmid plasmid were selected for isolation of bacmid DNA. Several isolated large white colonies were picked and re-streaked onto S-gal plates to verify true recombination. Confirmed colonies were inoculated into 5 mL LB broth containing 30 μg/mL kanamycin, 7 μg/mL gentamycin and 10 μg/mL tetracycline antibiotics and grown at 37°C for 24 hours with shaking.

Inoculated overnight bacterial cultures were pelleted using centrifugation at 800 g for 3 minutes at 4°C. Supernatant was discarded and the pellet resuspended by gentle vortexing into 0.3 mL bacmid solution I (15 mM Tris-HCL (pH 8), 10 mM EDTA, 100 μg/ml RNaseA; filter sterilised and stored at 4°C). 0.3 mL bacmid solution II (0.2 M NaOH, 1% SDS; filter sterilised and stored at room temperature) was added slowly and incubated for five minutes at room temperature until it went translucent. 0.3 mL 3 M potassium acetate (pH 5.5) was added slowed with gently mixing and left to incubate on ice for 5 to 10 minutes. Contaminant protein and *E. coli* genomic DNA formed a white precipitate. Precipitate was pelleted by centrifugation at 4000 g for 10 minutes at 4°C. The supernatant was gently transferred to pre-prepared tubes containing 0.8 mL absolute isopropanol. The mix was inverted gently and incubated on ice for 10 minutes. The DNA was pelleted at 14000 g for 15 minutes at room temperature. The supernatant was discarded, the tube could be re-spun to remove any excess supernatant. The pellet was left to air dry under sterile conditions and dissolved in 40 μL SDW without pipetting and stored at 4°C to be used within 1 week.

2.9.4 Bacmid verification

Bacmid recombination events were verified using PCR. PCR mix included 50 ng bacmid, 1x HF buffer, 0.2 mM dNTPs mix, 5mM forward and reverse primers, 0.5 µL Pfusion polymerase (ThermoFisher #F530S) to 50 µL with SDW. Cycling conditions were used as described previously. PCR was verified on a 1% agarose gel to determine presence and orientation of bacmid insertion. Primer combinations targeting internal and external insert regions were used to determine presence and orientation of bacmid insert.

2.10 General biochemistry techniques

2.10.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS PAGE)

SDS PAGE separates proteins according to molecular weight by applying an electrical current which causes the negative charge of a protein to migrate towards the positive electrode.[267] SDS PAGE was regularly used in this work to monitor the presence of proteins in purification and pull-down assays. A suitable percentage v/v concentration of acrylamide is required to make a polyacrylamide gel suitable to the protein molecular weight. High molecular weight proteins require a lower percentage gel. SDS PAGE polyacrylamide gels consist of a lower separating gel and an upper stacking gel containing wells for protein sample loading. Table 2.10 lists the components of polyacrylamide gels. The gels are set on the addition of TEMED and 10% APS.

Gels were loaded into an SDS PAGE running tank (Bio-Rad Mini tank) and filled with 1x SDS PAGE running buffer (table 2.11). The wells were washed out with running buffer prior to sample loading. Samples were prepared by addition of 1x SDS PAGE loading dye and 25 mM DTT and boiled at 95°C for 10 minutes.

Samples were run alongside blue pre-stained protein standard, broad range 11-250 kDa (NEB #P7718) and positive and negative protein controls at 120 volts for 90 minutes.

Proteins were visualised using Coomassie stain buffer (table 2.11) for 30 minutes with shaking followed by de-stain buffer refreshed until proteins were visible. Best results appeared when de-stain was left overnight.

2.10.2 Native PAGE analysis

Native gels were used for protein assay analysis using PROTEAN spacer plates (BioRad). Gels were made containing 1x TBE solution, appropriate volume of 30% acrylamide and distilled water to 40 mL and stirred until incorporated (table 2.10). TEMED and 10x APS were added and poured between the casting plates. Gels were left to set for 30 minutes. Gels used in this work included 5%, 8% and 10% acrylamide TBE gels dependent on protein size.

2.10.3 Protein immuno-blotting

Protein detection using protein immuno-blotting followed described previously methods for Horseradish Peroxidase (HRP)-conjugated secondary antibody reporters.[268,269] SDS PAGE was carried out alongside a biotinylated protein marker (Bio-Rad #1610376). Proteins were transferred from SDS PAGE to PVDF membrane (GE Healthcare) using Bio-Rad mini-tank electrophoresis at 60 volts for 2 hours. The tank was assembled in chilled 1 x transfer buffer (1x Tris-Glycine (TG) buffer (30.3 g Tris, 144 g glycine) with 200 mL methanol to 1 L with distilled water). An ice pack was used to maintain cold temperatures.

The membrane was transferred to 1x TBST (1x TBS with 0.1% v/v Tween) and washed 3 times for 15 minutes each. The membrane was blocked for 1 hour at

4°C in blocking buffer (1x TBST containing 5% Marvel milk powder) then washed a further 3x 15 minutes in 1x TBST. The membrane was left overnight in blocking buffer containing 1:1000 primary antibody (this dilution was used unless stated otherwise in table 2.9) at 4°C with gentle agitation. The type of primary antibody was dependent on the protein being analysed. The membrane was washed a further 3x 15 minutes in 1x TBST. The membrane was transferred to blocking buffer containing appropriate secondary antibody (1:3000) and anti-biotin antibody (1:1000) for 4 hours at 4°C (table 2.9). The membrane was washed and treated with Pierce ECL western blotting substrate kit (#32109). Proteins were visualised using a FUJIFILM Lass-3000 Mini with exposures of 5, 10 and 20 minutes detecting the chemiluminescent signal.

2.10.4 Protein detection by silver staining

Silver staining used disposable plasticware and SDW to reduce contamination following previously described methods.[270] SDS PAGE analysis was carried out on samples as described in section 2.10.1 using fresh SDS running buffer. The gel was transferred to a clean weigh boat and fixed in fixing solution for 2 times 15 minute with intermediate buffer exchange (table 2.12). All buffers were made up fresh and chilled at 4°C. The gel was transferred to a fresh weigh boat containing sensitive buffer for 30 minutes followed by 3 x 5 minute SDW washes. The gel was incubated in Silver containing buffer for 20 minutes followed by 2x 1 minute SDW wash steps. Gel development was variable from 1 - 20 minutes dependent on protein concentration. Development was stopped by addition of EDTA.

2.11 HelQ baculovirus production

2.11.1 Transfection of Sf21 cells with HelQ bacmid DNA

Bacmid transfection was carried out using Sf21 insect cells (#B821-01). Sf21 cells and Sf9 cells (#B825-01) are traditional cell lines used for baculovirus work and originated from the USDA Insect Pathology Laboratory.[271] Both cell lines originated from the IPLBSF-21 cell line, derived from the pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda.*[272,273] Sf21 insect cells have a doubling time of 24 hours.

Sf21 cells were grown to 50% confluency into 50 mL Erlenmeyer flasks in 10 mL supplemented Grace's medium (ThermoFisher). Cells were evenly spread into monolayers and left to adhere at 28°C for 1 hour. Supplemented media was aspirated off and cells washed in un-supplemented Grace's medium (Thermo Fisher). The bacmid prep was added to 300 µL un-supplemented Grace's medium carefully. Cellfectin® II reagent (ThermoFisher) was diluted 1 in 10 in un-supplemented Grace's medium and added to bacmid mix to a total of 1 mL. Reaction was incubated at room temperature for 1 hour. Un-supplemented media was aspirated off Sf21 monolayer prior to slow addition of bacmid mix. Flasks were rotated to get an even distribution and incubated at 28°C for 5 hours. 5 mL supplemented Grace's media was added to bacmid mix and left for a further 60 hours.

Post-60 hours cells appear 'spongy' which is used to indicate successful infection. The cell monolayer was dislodged by gentle tapping to the flasks to allow cells to come away easily. The infected Sf21 cells were transferred to 50 mL tubes and pelleted by centrifugation at 400 g for 5 minutes. The supernatant was transferred to a fresh tube and labelled P1 virus and stored at 4°C.

P1 to P2 viral amplification to improve the volume of insect virus and increase infection rate was carried out using Sf21 insect cells. Sf21 cells were split to 30-40% confluency in 200 mL static flasks. Cells are adhered in supplemented Grace's media for 1 hour at 28°C. Supplemented media was aspirated off the cells prior to

incubation with 0.5 mL P1 virus stock diluted with 0.5 mL supplemented media. The flask was rotated to allow for even distribution and incubated at 28ºC for 1 hour. The flask was made up to 20 mL with supplemented medium and left for 60 hours at 28ºC.

P2 virus was harvested by centrifugation at 400 g for 5 minutes, transferred to a clean tube and stored at 4ºC. The pellet was stored for analysis by SDS PAGE to determine successful protein expression.

2.11.2 Transfection of insect cells with P2 virus

HiFive insect cells (Invitrogen #B855-02) were used to over-express protein while working at the Research Complex in Harwell (RCaH). HiFive cells were developed by the Boyce Thompson Institute for Plant Research, and originated from ovarian cells of the cabbage looper, *Trichoplusia ni*.[274,275] HiFive cells have a doubling time of 18 hours and are used in virus transfection as they provide 5-10-fold higher secreted protein expression then Sf9 cells. HiFive cells in suspension were split to 1-2 x10^6 cells/mL in Express Five SFM supplemented with L-glutamine media (ThermoFisher). Cells were infected with P2 virus 1 in 100 and incubated at 28ºC for 48 hours with shaking.

Infected HiFive cells were harvested at 400 g for 5 minutes at 4ºC. Supernatant was discarded and the pellet was resuspended in sterile lysis buffer (600 mM NaCl, 50 mM Tris pH 7.5, 10% glycerol, 5 mM TCEP, 1x protease inhibitor tablet EDTA free). Resuspended pellets were stored at -80ºC prior to purification.

Protein over-expression using Sf9 insect cells was carried out for purifications carried out in Nottingham. The process of P2 infection followed the same protocol as described above however Sf9 cells were used. Viral titres with Sf9 cells were carried out to optimising viral volume, see later for details.

2.12 Insect cell culture

2.12.1 Media and supplements

Sf9 insect cells were regularly used throughout this work. Sf9 cells were cultured using LONZA Insect Xpress protein-free media with L-glutamine (VWR #733-1205) supplemented with Pluronic acid (Gibco Life Technologies #24040032), penicillin-streptomycin solution (Sigma #P4333) and amphotericin B solution (Sigma #A2942). Sf9 cells have a doubling time of 72 hours.

2.12.2 Starting up Sf9 insect cells

All insect cell culture was carried out in a category 2 hood under sterile conditions. Media was pre-warmed at 27°C and added to flasks to 20% occupancy. 1 mL frozen Sf9 cell stocks ($1x10^{12}$ cells/ mL) at -80°C were rapidly defrosted at 37°C for 30 seconds and directly added into prepared 25 mL media. Cells were left at 27°C with shaking for 5 days. The cells were counted and split accordingly.

2.12.3 Maintaining Sf9 cell stocks

Cells were counted using a haemocytometer and split between $8\text{-}10x10^{6}$ cells/mL. Cells were generally split every 3-4 days. Cells were split into pre-warmed supplemented medium to $1x10^{6}$ cells/mL. Flasks of 10 mL cells were maintained on a weekly basis for rapid access to Sf9 insect cells readily dividing, this was known as the maintenance stocks. Unused dense cell stocks were discarded in ethanol.

2.12.4 Freezing Sf9 insect cells for storage

Cell cultures of 50 mL at $8x10^{6}$ cells/mL were harvested at 100 g for 10 minutes. The supernatant was discarded, and the cells were resuspended in fresh medium containing 20% Foetal Bovine Serum (FBS) (WolfLabs #SLS1166) and

10% DMSO (Sigma #D2438) to a density of $12x10^6$ cells/mL. Cells were transferred to cryotubes in 1 mL aliquots and frozen slowly by initial incubation at 4°C for a few hours, to -20°C overnight before long-term storage at -80°C.

2.12.5 Massing up insect cells for protein over-expression

The volume of cells was expanded for protein over-expression by diluting the entire maintenance stock to $1x10^6$ cells/mL and supplementing with fresh pre-warmed culture. Cells were continually expanded over several weeks until the required volume of 4 L was reached.

2.12.6 Viral titre optimisation for protein expression

Viral titres were used to determine optimum virus volumes and incubation times for Sf9 cells with new virus stocks. Grids were assembled with combinations of incubation times (days) and viral volumes (µL) to measure for maximum protein over-expression. Sf9 insect cells were distributed at a density of 600,000 cells per well in 500 µL supplemented media into sterile 24 well culture plates.

Cells were added to each well and spread by gently shaking. The cells were left to adhere at 27°C for 30 minutes. P2 virus at specified volumes was added to each well. Volumes of 0.5 µL, 2 µL, 5 µL, 10 µL and 20 µL were tested alongside a negative control (0 µL). Each volume was incubated at 27°C for 24, 48, 36, 42 and 48 hours. At each time point samples were taken for SDS PAGE analysis. Media was aspirated off and cells resuspended in 100 µL 2x SDS PAGE loading dye. Cells were treated for SDS PAGE analysis as described previously and run on 8% acrylamide SDS PAGE gel. Coomassie staining and protein immuno-blotting was used to detect the presence of HelQ.

2.12.7 Insect cell viral infection for wild type HelQ

HelQ protein over-expression was optimised at 4.2 µL P2 hexaHis-SUMO-StrepII-HelQ virus per $1x10^6$ Sf9 cells for 46-48 hours at 27ºC.

2.12.8 Insect cell viral infection for N-HelQ fragment

The N-terminal region of HelQ was produced by over-expression of FL-HelQ P2 virus infected Sf9 cells for 72 hours at 27ºC. Extended incubation periods resulted in natural protein degradation into stable fragments.

2.13 Insect cell protein expression system

2.13.1 Protein purification of FL-HelQ

FL-HelQ and N-terminal HelQ (N-HelQ) were over-expressed and purified using the Bac-to-Bac baculovirus expression system in Sf9 cells as described previously. Affinity tags used included an N-terminal hexaHis-tag, SUMO- tag and StrepII-tag. HelQ protein over-expression was optimised for use in Nottingham to 4.6 µL virus per $1x10^6$ Sf9 cells for 48 hours at 27ºC. N-HelQ was produced by over-expression with an incubation of 72 hours at 27ºC. See appendix for SNAPGENE plasmid maps of vectors used in this work. Infected cells were pelleted under centrifugation using an Avanti J-26 XP centrifuge (Beckman Coulter) and a JLA 10.500 rotor for 20 minutes at 2400 g.

Cells were resuspended in lysis buffer (150 mM Tris pH 8, 150 mM NaCl, 20 mM imidazole, 10% glycerol) containing COMplete EDTA free protease inhibitor tablets and stored at -80ºC until purification. Freezing was carried out in smaller aliquots to allow for faster thawing. All steps were carried out on ice to reduce degradation. The composition of purification buffers can be found in table 2.14. Cells were thawed slowly on ice before gentle lysis by sonication using a Soniprep

150 ultrasonic disintegrator (MSE) at 80% pulsed for 1 min per 5 mL biomass. Sonication was carried out for longer at a lower frequency then carried out by previous students in the lab.[252] These modifications reduced protein exposing to changes in temperature. Insoluble proteins were removed by centrifugation at 38000 x *g* in a pre-chilled centrifuge for 1 hour using an Avanti J-26 XP centrifuge with a JA 25.50 rotor. Clarified supernatant was fractionated by a 0-50% ammonium sulphate cut. The salt was ground to powder form using a pestle and mortar and added slowly with gentle mixing at a rate of 12 g every 40 minutes.

The ammonium sulphate biomass was pelleted using centrifugation at 59000 g for 10 minutes at 4ºC to remove excess salt. The pellet was resuspended in ammonium sulphate resuspension buffer (table 2.14) prior to repeated centrifugation at 92400 g for 40 minutes.

All purifications described here were carried out on the AKTA start FPLC system unless otherwise stated (GE Healthcare #29022094). Supernatant was filtered through 0.22 μm syringe filters (GE Healthcare) prior to being loaded onto a 5 mL charged Ni-NTA hexaHis-Trap column pre-equilibrated in NTA-A buffer (table 2.15). Filtration was required to remove increased cellular debris associated with an insect expression system compared to a standard *E. coli* system. Bound protein was washed with filtered high-salt NTA-A buffer and developed over a long ascending imidazole gradient to 1 M (Buffer NTA-A with 1 M imidazole). HelQ at 137 kDa elutes between 10-20% imidazole. Fractions were analysed on an 8% acrylamide gel for SDS PAGE analysis. HelQ containing fractions were pooled and dialysed over-night at 4ºC to remove the high salt (150 mM NaCl, 10% glycerol, 50 mM Tris pH 8).

Protein was loaded onto a 5 mL heparin column. Bound protein was washed in HepA buffer and developed with a gradual ascending NaCl gradient with HepB

buffer (1 M NaCl). HelQ eluted between 10-40% NaCl. HelQ was visualised as before. HelQ was dialysed in HelQ dialysis buffer 1 to reduce the salt before loading onto a 1 mL anion exchange Q-sepharose column. Dialysis steps were carried out for 2.5 hours at 4°C with frequent buffer changes to hasten the purification process. Bound protein was washed in buffer HepA and eluted over a similar NaCl gradient. HelQ eluted in two pools, a low concentrated cleaner pool and a higher concentrated dirty pool. Clean HelQ eluted between 35-65% NaCl. The pooled protein was dialysed for 3 hours at 4°C into 30% glycerol, 5 mM DTT, 150 mM NaCl, 50 mM Tris pH 8. Protein concentration was calculated using Bradford protocol. HelQ was aliquoted to prevent repeated freeze/thaw and stored at -20°C.

The inconsistency of HelQ binding to the heparin column suggested this step did not remove contaminants but extended protein exposure. The addition of a DNaseI treatment step did not improve HelQ protein binding to the heparin column. Therefore, reduced binding affinity to heparin was likely due to column saturation with contaminants and not HelQ-DNA complex formation within the cell. Subsequently, the heparin column was removed in later purifications.

2.13.2 Protein purification of N-HelQ from insect cell over-expression

N-HelQ purification was carried out as described above without the heparin column step. N-HelQ was visualised at 47 kDa on 10% acrylamide gels for SDS PAGE analysis. The C-terminal HelQ fragment was obtained by isolating DNA-binding fractions after removal of N-HelQ by purification through a heparin column as described above.

2.14 Bacterial protein expression system

2.14.1 Protein purification of N-HelQ fragment from *E. coli*

2.14.1.1 Construction of N-HelQ DNA into pET14b plasmid

DNA encoding N-HelQ was cloned for *E. coli* expression following the described previously methods. N-HelQ consists of the 46.36 kDa N-terminal fragment of HelQ. The 1062 DNA bases, including the SUMO- and StrepII-tags, were cloned from the pSN52 HelQ plasmid (terminating at isoleucine 240). A STOP codon was introduced into the plasmid using the reverse primer (table 2.3). See appendix for DNA sequence and SNAPGENE plasmid map. DNA encoding N-HelQ was cloned into pET14b plasmid using the XhoI and HindIII restriction enzymes (NEB) (pTJ9, table 2.6). The plasmids were selected for ampicillin resistance as described previously.

2.14.1.2 N-HelQ protein over-expression

The N-HelQ containing plasmid (pTJ9) was transformed into BL21 DE3 *E. coli* cells on ampicillin containing agar plates. N-HelQ was over-expressed by induction at 0.5 OD_{600} with 0.02 % L-arabinose and 1 mM IPTG for 3 hours at 37°C. Cells were harvested by centrifugation at 1351 g for 30 minutes. Pelleted cells were resuspended in resuspension buffer (table 2.14) and stored at -80°C prior to purification.

2.14.1.3 Protein purification of N-HelQ from *E. coli*

Biomass containing N-HelQ protein was defrosted slowly on ice and lysed by sonication for 1 minute intermittent pulsing per 5 mL of biomass. Cells were clarified at 38000 g for 1 hour at 4°C. Subsequent affinity chromatography steps using the same buffers as for HelQ purification included a 5 mL Ni-NTA column

developed over an imidazole concentration gradient (table 2.14). N-HelQ eluted at 30% imidazole. This was followed by over-night dialysis into dialysis buffer 1 and a 1 mL Q-sepharose column. N-HelQ eluted as two peaks, one at 50% and one at 100% NaCl. All fractions were analysed by SDS PAGE. Fractions at 100% showed a higher purity with fewer contaminants in SDS PAGE analysis but a lower concentration compared to peak 1. N-HelQ containing fractions were pooled and dialysed at 4ºC into storage buffer (table 2.14) for 3 hours. Protein was aliquoted for storage at -20ºC.

2.14.2 Subcloning of N-HelQ encoding DNA

DNA encoding N-HelQ was subcloned into pETM40 (cleavable MBP tag) and pET28b-SF (cleavable double StrepII-tag) for use in crystal trials following described methods (table 2.6). Three variants of N-HelQ with variations in DNA length included N1, the PWI only domain, N2, the PWI and RecA-like domain 1 and N3 the PWI domain, RecA-like domain 1 and 2. N-HelQ was also subcloned into pRSF1b for Kan^R.

2.14.3 Mutagenesis of N-HelQ encoding DNA

DNA encoding N-HelQ underwent mutagenesis and purification as described previously (table 2.3). Mutants made included a phosphomimetic N-HelQ with two serine to aspartate changes within the PWI domain, serine 158 and serine 178; a truncated N-HelQ at position valine 76 producing a 30 kDa protein exclusive of the PWI-like domain (Trun-N-HelQ) and a loss of function double mutant within the PWI-like domain of N-HelQ DFA → GGA at position 142 (PWI-N-HelQ) (table 2.6 shows plasmids made in this work). Amino acid positions are given with respect to start codon of HelQ in the absence of tags (see chapter 3 for more detail on mutants).

2.14.4 Protein purification of Replication Protein A (RPA)

Replication Protein A (RPA) purification was optimised at RCaH following previously documented methods.[276] Human recombinant RPA DNA was cloned into a pET vector including the three subunits of RPA under the control of an inducible T7 promoter making p11d-tRPA. The design and construction of this plasmid is described by Henricksen *et al.* (1994).[277]

RPA expression is toxic to *E. coli,* so optimisation to minimise the number of generations between transformation and induction was required. No starter cultures were used in this protocol as they significantly reduced expression. p11d-tRPA (50 ng) was transformed into BL21 (DE3) cells (table 2.2). One litre LB broth with 100 µg/mL ampicillin were inoculated from single colonies of freshly transformed cells (1 to 2 days old). Cultures were incubated overnight at 37ºC without aeration or shaking. The OD was measured at 600 nm and cells returned to 37ºC until an OD between 0.5-0.8 was reached. Cells were induced with 0.3 mM IPTG final and incubated for 3 hours. Cells were harvested by centrifugation at 1350 g at 4ºC. The pellet was resuspended in HI buffer (30 mM HEPES pH 7.8, 0.25 M EDTA, 10% glycerol, 1 mM dithiothreitol (DTT) and 0.01% Tween20) containing 1 mM phenylmethylsulphonyl (PMSF) (table 2.14). Re-suspended cells were then frozen and stored at -80ºC prior to purification.

Protein containing biomass was thawed on ice and lysed using sonification as described previously. Cells were clarified at 38000 g for 35 minutes at 4ºC and the supernatant was kept for further processing. The supernatant was loaded onto an Affi-gel blue chromatography column (table 2.15). The beads were prepared following BioRad's manufacturer's instructions.[278] 1 mL Affi-gel blue resin was embedded into a column and equilibrated in HI buffer. All waste was collected and disposed of appropriately. Affi-gel blue resin is a crosslinked agarose gel with

Cibacron blue dye covalently attached and purifies nucleotide-binding proteins with high affinity.[279] The lysate was applied to the column using gravity flow, washed with four column volumes of HI buffer containing 80 nM KCL. Subsequent washes included HI buffer containing 800 mM KCL, 500 mM NaSCN and 1.5 M NaSCN. Washes were collected and analysed by SDS PAGE. The column was regenerated by washing with HI containing 2 M NaSCN. RPA eluted with 1.5 M NaSCN. RPA was present in other fractions; however, this fraction produced the purest form of RPA.

RPA was passed through a hydroxyapatite (HAP) chromatography column. The column was prepared by initially hydrating the powder in PBS prior to washing in HI-buffer. The beads were incubated with protein with gentle shaking for 30 minutes and then the resin was left to settle by gravity before buffer was removed. The beads were equilibrated in HI containing 30 mM KCl prior to addition of the RPA containing Affi-blue fraction. The beads were washed with 3 column volumes of HI buffer. Subsequent washes included HI buffer with 80 mM KPO_4 and 500 mM KPO_4. All fractions were analysed using SDS PAGE. RPA eluted with 80 mM KPO_4 but can elute in other fractions.

HAP purified protein containing fractions were concentrated using a 1 mL MonoQ column. The RPA fraction was diluted 1:3 in HI buffer before loading onto equilibrated MonoQ in HI with 50 mM KCl. Subsequent washes with 4 column volumes of HI buffer containing 50 mM KCl and 100 mM KCl prior to elution over 10 column volumes of an ascending linear salt gradient (200-400 mM KCl). RPA eluted as a single peak at 300 mM KCl and analysed using SDS PAGE. RPA was pooled as a pure fraction and as an un-pure fraction. Remaining RPA-containing fractions were cleaned up using superdex 200 gel filtration. RPA was aliquoted and stored at -80^{o}C. The Bradford Assay was used to determine protein concentration.

2.14.5 Protein purification of RPA14

DNA encoding RPA14, the smallest subunit of RPA at 23.5 kDa, was subcloned from p11d-tRPA using XhoI and HindIII flanking forward and reverse primers. RPA14 DNA was cloned into pBadHisA with an N-terminal hexaHis-tag. RPA14 was cloned individually to determine if this was the interacting domain to N-HelQ.

RPA14 was over-expressed in BL21 AI *E. coli* chemically competent cells and induced at OD_{600} 0.4-0.6 with 1 mM IPTG and 0.02 % L-arabinose for 3 hours at 37°C. Cells were pelleted and resuspended in 150 mM NaCl, 20 mM Tris pH 8, 10% glycerol, 5 mM imidazole and protease inhibitor. Biomass containing protein were stored at -80°C prior to purification.

RPA14 protein was purified from biomass using standard conditions described previously. Biomass containing RPA14 protein was sonicated for 1 minute intermittent pulsing per 5 mL of biomass and clarified by centrifugation at 38000 g. The supernatant was loaded onto a chilled 5 mL charged Ni-NTA Hi-Trap column and eluted over a linear ascending imidazole gradient over five column volumes in 1 M NaCl, (table 2.14). RPA14 eluted between 200-300 mM imidazole. RPA14 containing fractions were dialysed into low salt and loaded onto a 2 mL heparin column (table 2.15). RPA14 has one DNA binding domain. RPA14 was eluted over a linear ascending salt gradient (150 mM to 1 M NaCl) over 15 column volumes. RPA14 eluted between 500-600 mM NaCl. RPA14 fractions were collected and dialysed into storage buffer and stored at -20°C.

2.14.6 C-HelQ purification from *E. coli*

Protein over-expression and purification protocols to purify hexa-His-tagged C-HelQ were carried out as previously described for *E. coli* protein purification. C-

HelQ was over-expressed from pHB01 (table 2.6) for 3 hours at 37ºC post induction by 1 mM IPTG and 0.2% L-arabinose. Biomass containing C-HelQ was sonicated and clarified by centrifugation at 38000 g. The supernatant was loaded onto a chilled 5 mL charged Ni-NTA Hi-Trap column and eluted over a linear ascending imidazole gradient (table 2.14). C-HelQ containing fractions were resolved on a 1 mL Q-sepharose column using a linear gradient of 150 mM -1 M NaCl. C-HelQ was dialysed and stored at -20ºC.

2.15 Methods of protein determination

2.15.1 Bradford protein assay

Bradford protein assay was used to calculate protein concentrations. The Bradford reagent (Sigma #B6916-500ML) was brought to room temperature. A standard curve using a serial dilution of BSA protein concentrations was made using 1 mL Bradford reagent and 1 mg/mL BSA stock. Dilutions were made up of 0, 0.1 1, 2, 3, 5, 7 µg/mL of BSA. The reactions were left to incubate at room temperature for 30 minutes.

Dilutions were transferred to cuvettes and the absorbance measured at 595 nm using a Bradford reagent only blank. The absorbance was plotted against the known BSA concertation to produce a standard curve using PRISM 7 graphpad.[280] The concentration of the protein was calculated using two dilutions of protein in Bradford reagent and averaged. The protein dilution was left at room temperature for 30 minutes prior to measuring. The concentration was calculated using the standard curve, ensuring the dilution factor into Bradford reagent was considered. The average of the two samples was used to calculate the molar concentration by dividing by the protein extinction coefficient.[281,282]

2.15.2 NanoDrop spectrophotometer

A NanoDrop 2000 spectrophotometer (ThermoScientific) was used to confirm protein concentration but was not used when calculating the concentration for experimental use. Protein measurements were taken at A280 absorbance after blanking with appropriate storage buffer. An average of three readings was used to calculate the molar concentration using the protein extinction coefficient value.

2.15.3 Protein summary gel using SDS PAGE

SDS PAGE was used to test the purity of the protein and confirm the concentration visually. Samples were prepared with protein and 4x SDS PAGE loading dye, 25 mM DTT made up to 20 µL with SDW. The samples were boiled at 95°C for 10 minutes and spun by centrifugation before loading and analysis.

2.16 Biochemical assays

2.16.1 Preparation of fluorescently labelled DNA substrates

Dyes can be used as fluorescent labels on DNA because of their ability to absorb and then emit energy. The absorbance and emission wavelengths differ for different dyes used. For the purpose of this work, most DNA substrates were Cy5 end labelled. Cy5 fluorophores absorb at 550 nm and emit at 670 with an extinction coefficient of 250,000.[283] Oligos were ordered pre-labelled with Cy5 on the 5′ DNA end (Sigma) (table 2.4). Fluorescein labels were used for fluorescence anisotropy. Fluorescein fluorophores absorbs at 494 nm and emits at 512 nm.[283] DNA substrates were made using a 1.2:1 ratio of unlabelled to Cy5-labelled oligos. Reactions to anneal the DNA oligonucleotides into intermediate substrates were set up (50 µL) containing 5 µM labelled and 6 µM unlabelled oligo with 1x annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl and 1 mM EDTA). Substrates were heated to 95°C for 10 minutes and cooled to room temperature overnight. Oligos were gel

purified using 8% TBE gel run at 140 volts for 3 hours. Substrates were visualised by eye and cut out of the gel. DNA was eluted from the gel by diffusion at 4°C for 48 hours into 250 µL elution buffer (4 mM Tris pH 8, 10 mM NaCl). Substrates were spin concentrated using a SpeedVac (Eppendorf Concentrator 5301) to a volume of 50 µL. Concentration was determined by measuring the absorbance at 260 nm using a Nanodrop and conversion to molar concentration using Beer-Lambert law. The extinction coefficient was calculated using OligoAnalyzer 3.0.[284] DNA substrates were used at 25 nM for assays.

The G4 quadruplex 50 base oligonucleotide was 3′ labelled with Aminoallyl-UTP Cy5 using terminal deoxynucleotidyl transferase (TdT) following previously described methods.[285]

2.16.2 Helicase DNA unwinding assay

DNA unwinding activity of HelQ and HelQ fragments were analysed by detecting the dissociation of annealed DNA oligos and monitoring the Cy5 labelled DNA. DNA unwinding activity by HelQ proteins was assessed by the formation of ssDNA Cy5 oligonucleotides that migrate in TBE gels faster than the duplex Cy5 labelled substrates. Unwinding activity reactions (20 µL) were compiled of 1x helicase buffer (table 2.11), 5 mM $MgCl_2$, 5 mM ATP pH 8, 25 mM 1,4-dithiothreitol (DTT), 2.5 µM 'cold-trap' oligomer, 25 nM fluorescent substrate and diluted protein ranging between 0- 160 nM. Work carried out in the Bolt lab and work by Sarah Northall[252] on HelQ identified the need for an unlabelled strand 'cold-trap' of the same sequence as the labelled strand in order to detect unwinding. Helicase unwinding reactions were incubated at 37°C for 10 minutes prior to quenching by addition of STOP buffer (table 2.11). Reactions were loaded in Orange G loading dye (80% v/v glycerol, Orange G) on 10% w/v acrylamide TBE gels and run in 1x TBE buffer (table 2.11). Products were separated by

electrophoresis at 150 volts for 60 minutes. Gels were imaged using an Amersham Typhoon phosphor-imager detecting the Cy5 label florescence and analysed using GelAnalyzer 2010a and ImageJ software.[286,287] Graphs and statistical analysis shown in this work were generated using PRISM 7 (GraphPad).[280]

Helicase re-annealing assays were described as above under conditions described in the results chapters with complimentary oligos shown in table 2.4.

2.16.3 Electro-Mobility Shift Assay (EMSA)

Biochemical EMSAs are 'in-gel' techniques used to study protein-DNA interactions.[288] EMSA analysis detects stable protein-DNA bound complexes by monitoring the migration of fluorescently labelled DNA substrates. Complexes of a higher molecular weight migrate more slowly through the gel when compared to free fluorescently labelled DNA. Binding reactions (20 µL) were assembled with 1x HB, 25 mM DTT, 25 nM fluorescently labelled DNA substrate and varying protein concentrations ranging from 0- 500 nM. Protein dilutions that gave a final concentration were made as serial dilutions. Complexes were run in Orange G loading dye and resolved using 5% acrylamide TBE gels at 140 volts for 120 minutes. Gels were imaged and analysed as previously described.

2.16.4 Protein-protein super-shift assay

Super-shift EMSA assays were used to determine interactions between proteins on DNA. Specific interactions analysed were HelQ and RPA on fork 3 DNA. RPA was incubated with Cy5 labelled fork 3 DNA (25 nM) for 5 minutes at 37°C in 1x HB buffer and 25 mM DTT. HelQ was added to the reaction at stated concentrations (0-160 nM) to a final volume of 20 µL for a further 10 minutes. Reaction was resolved by electrophoresis on 5% acrylamide TBE gels at 120 volts for 3.5 hours.

Control assays included the use of Aro1-RPA and SSB (table 2.8). Gels were imaged and analysed as previously described.

2.16.5 ATPase activity assay

ATPase assays using a detergent-free malachite green colour reagent protocol as previously described.[289] Reactions were carried at as described for helicase unwinding assays in the presence of 25 nM DNA at 37ºC for 10 minutes and varying concentrations of HelQ to a final volume of 50 µL. The reaction was quenched by addition of 3:1 0.045% Malachite green with 4.2% ammonium molybdate (dissolved in 4 M HCl) and incubated for 2 minutes. 50 µL 34% aqueous sodium citrate was added for 30 minutes at room temperature. The absorbance was measured at 660 nm, blanked to a no DNA control. See table 2.12 for buffer components. A 1 in 10 serial dilution of potassium dihydrogen phosphate (0.1 M) was diluted in 1x HB and measured to create a standard curve of free phosphate.

ATPase assays were carried out with varying protein concentrations of FL-HelQ, N-HelQ and C-HelQ to determine differences in ATP hydrolysis activity. More extensive kinetics was carried out on FL-HelQ by varying DNA substrate concentrations and ATP concentrations to determine K_M and V_{max} values.

2.16.6 Native PAGE gel electrophoresis

Protein oligomeric state was tested using Blue Native PAGE (Invitrogen #BN1001BOX) Native PAGE 3-12% BIS-Tris gel kit as described in the manufacturer's instructions. Protein was pre-incubated in a range of conditions at 37ºC for 10 minutes. A boil control containing 10% SDS buffer was used to show monomeric protein.

A ‘home-made’ Native PAGE kit was used to test different percentage gels. Protein samples were assembled with addition of 1x HB, 25 mM DTT, 5 mM +/- ATP and $MgCl_2$, +/- 25 nM ssDNA to a final volume of 40 µL. Controls included boil control containing 10% SDS and single protein controls. Assays included N-HelQ and RPA combinations to determine interactions and a role in trimer dissociation. Protein samples were loaded with 1x Loading dye (50 mM Tris pH 6.8, 5% glycerol, 20 µg/mL Bro Blue) and run at 120 volts for 2 hours on 10% native gel (5 mL H_2O, 2.5 mL 40% acrylamide, 2.5 mL 3 M Tris pH 6.8, 50 µL 10% APS, 50 µL TEMED) using 1x running buffer (10x stock: 250 mM TRIZMA, 1.92 M glycine).

2.17 Protein roadblock assays

2.17.1 BamHI^{EIIIA} as a roadblock

A mutant variant of the BamHI restriction enzyme, BamHI^{EIIIA}(NEB), which was able to bind the recognition sequence but unable to cut it, was used as a protein barrier on forked DNA assays including the BamHI^{EIIIA} roadblock were conducted as described for helicase assays with a pre-incubation step of the BamHI^{EIIIA} with DNA. BamHI^{EIIIA} at varying concentrations 0-160 nM was incubated at 24^{o}C for 15 minutes with fluorescently labelled DNA substrate (25 nM). Reaction also included 1x HB, 25 mM DTT and 2.5 µM ‘cold-trap’ oligo. HelQ at varying concentrations 0-160 nM was added with 5 mM ATP and 5 mM $MaCl_2$ to the roadblock experiment final volume of 20 µL and incubated at 37^{o}C for 10 minutes before termination with STOP buffer. The reactions were resolved on 10% acrylamide TBE gels and analysed for unwinding.

Experiments including RPA were carried out with an additional second incubation step prior to addition of HelQ at described RPA concentrations at 37^{o}C

for 5 minutes. BamHI^{EIIIA} roadblock experiments were also carried out with RPA substitutions (SSB and Aro1-RPA) (table 2.8).

2.17.2 Restriction enzyme protection assay

Enzyme protection assays were used as a control to measure whether RPA alone could displace the BamHI^{EIIIA} roadblock from DNA in the absence of HelQ. The roadblock was assembled as described previously onto the labelled substrate (25 nM) in 1x HB and 25 mM DTT. RPA at concentrations stated in figure legends was added to the reaction and incubated at 37ºC for 10 minutes. Wild type BamHI (NEB) restriction enzyme (2U) was added and incubated for a further 20 minutes. The protein was deactivated at 65ºC for 10 minutes. The reaction was run on a 5% acrylamide TBE gel at 140 volts for 120 minutes. Controls included RPA only, wild type BamHI only and BamHI^{EIIIA} roadblock.

Gels were imaged as described and assessed for fluorescently labelled DNA degradation by wild type BamHI enzyme.

2.17.3 Streptavidin-biotin displacement assay

Streptavidin binding DNA contained two biotinylated thymidine residues within the duplex region of the fork substrate following documented protocols.[290] Streptavidin binding to biotin labelled Cy5 DNA fork substrates (25 nM) was optimised using EMSAs described previously. The concentration of free biotin (Sigma) to capture free streptavidin unbound to the biotinylated fork was optimised by addition into Streptavidin-DNA EMSA assays. Streptavidin concentration for DNA shifting was optimised to 1µM with 0.1 µM free biotin. Biotin labelled DNA substrates were pre-incubated with streptavidin on ice for 5 minutes with 1x HB and 25 mM DTT. Addition of HelQ (0-160 nM), 5 mM ATP, 5 mM $MgCl_2$ and free biotin to a total volume of 20 µL initiated the reaction. Unwinding activity was

terminated after 10 minutes with STOP buffer. The assay reactions were resolved on 10% acrylamide TBE gels at 150 volts for 90 minutes.

Experiments including RPA required an additional pre-incubation step at 37ºC for 10 minutes with streptavidin-DNA prior to addition of HelQ protein. Streptavidin displacement also included assays run with control proteins at stated concentrations (RecQ, RecG, PriA and Rep, table 2.8).

2.17.4 RNA polymerase displacement assay

Another barrier on DNA tested for HelQ displacement activity included RNA polymerase. The presence of the replisome on DNA prevents RNA polymerase activity, creating a DNA roadblock. Reaction was assembled with 2 nM plasmid DNA, 40 mM HEPES pH 8.0, 10 mM DTT,10 mM $Mg(CH_3COO)_2$, 2 mM ATP, 0.2 mM G/CUTP, 25 mM dNTPs, 20 mg/mL BSA. Protein replisome was set up including 160 nM DnaB, 160 nM DnaC, 1 μM SSB, 80 nM β, 30 nM HU, 200 nM DnaG, 50 nM DNA polymerase III, 25 nM Tau clamp loader. Assays were initiated by addition of assembled protein replisome to reaction mix. Various combinations of HelQ (0-100 nM) and RPA (20 nM) were added to a final volume of 20 μL. Assay was started with addition of 1 μL RNA pol and 0.45 μM DnaA and incubated at 37ºC for 5 minutes. Radioactive was incorporated by addition of 1 μL SmaI (NEB) enzyme and 2 μL α^{32}PP dCTP for a further 2 minutes. The reaction was terminated by 0.5 M EDTA. Rep (100 nM) was used as a positive control for replisome displacement. Replicated DNA was purified to remove excess α^{32}PP dCTP not incorporated during replication.

Samples were loaded onto a submarine alkaline 0.7% denaturing agarose gel (200 mL 0.7% agarose in 2 mM EDTA), and chilled. The gel was transferred to 30 mM NaOH and set for 1 hour. The gel was run at 100 volts for 20 minutes in running buffer A (2 mM EDTA, 30 mM NaOH) and reduced to 25 volts for 16 hours. The

gel was fixed in TCA and washed in cold 5% trichloroacetic and SDW to remove ^{32}PP. The gel was dried using a DE81 dryer for 1 hour at 80^{o}C and imaged using a phosphoimager as described previously.

2.18 Protein pull-down assays

2.18.1 HexaHis-tag protein pull-down assay

Protein pull-down assays between HelQ and RPA were carried out using SIGMA iminodiacetic acid (IDA) (#11139-85-8) resin to trap $(His)_6$-tagged HelQ. Resin was charged with nickel and washed in Buffer A (50 mM Tris pH 7.5, 10% glycerol, 25 mM DTT, 150 mM NaCl, 5 mM imidazole). HelQ and RPA were incubated 1:1 on ice for 30 minutes in the presence and absence of ssDNA (MW12) at 25 nM. Resin was added to reaction with gentle agitation for 1 hour at 4^{o}C. Protein bound resin was washed 3 times in buffer A and eluted three times in high imidazole (1 M). Fractions were analysed using SDS PAGE.

2.18.2 StrepII-tag protein pull-down assay

Protein pull-down assays between HelQ and fragments and RPA were carried out using Iba Solutions Gravity flow Streptactin Sepharose column (#2-1202-001) following manufacturer's instructions. The assay utilised the StrepII-tagged HelQ and N-HelQ protein variants. HelQ was pre-incubated on ice with RPA in a 1:1 ratio for 30 minutes in the presence and absence of ssDNA (MW12) at 25 nM. Reactions were loaded onto a StrepII 0.2 mL column and washed with 3 column volumes of buffer A (150 mM NaCl, 20 mM Tris pH 7.5, 25 mM DTT, 10% glycerol). Protein was eluted with 3 column volumes of buffer A with 2.5 mM D-desthiobiotin (Iba). Fractions were run on 8% acrylamide gels for SDS PAGE analysis and visualised using silver staining.

2.18.3 Anti-SUMO protein pull-down assay

Anti-SUMO pull-down assays targeted the SUMO-tag of FL-HelQ and N-HelQ proteins. Reactions were assembled with 10 µL 1/20 dilution Anti-SUMO rabbit monoclonal antibody (table 2.9) with protein in IP buffer (25 mM Tris pH 7.2, 150mM NaCl) to a final volume of 50 µL. Proteins were added at 2 µM included: N-HelQ only, RPA only, C-HelQ only, N-HelQ and RPA, N-HelQ and C-HelQ. Reactions were incubated shaking at 4°C for 1 hour. 100µL Protein A resin slurry (Thermo Scientific #20333) was added and incubated for a further 2 hours. Beads were pelleted by centrifugation for 5 minutes at 2500 g. Supernatant was removed and processed for SDS PAGE analysis. Protein bound beads were washed twice with 50 µL IP buffer. Protein was eluted by boiling in 2x SDS PAGE loading dye for 5 minutes at 95°C. All wash and elution samples were analysed by SDS PAGE.

2.18.4 Protein pelleting assay

N-HelQ and RPA were incubated in a 1:1 molar ratio in the presence and absence of 25 nM ssDNA (MW12) with 1x HB and 25 mM DTT. Variations included addition of 5 mM ATP and $MgCl_2$ to a final reaction volume of 40 µL. The reaction was incubated at 37°C for 10 minutes. 20 µL reaction volume was centrifuged at 18407 x *g* for 10 minutes. Supernatant was removed and pellet resuspended in 4x SDS PAGE loading dye. All samples including the remaining 20 µL reaction mix, were treated for SDS PAGE analysis.

2.18.5 Saturated ammonium sulphate pull-down assay

This method was used to determine interactions between untagged proteins particularly between RPA and N-HelQ. An over-saturated solution of ammonium

sulphate was made by dissolving 55-60 g in 100 mL SDW. The solution was heated to aid dissolving and left to cool to room temperature. The saturated solution formed ammonium sulphate crystals on cooling.

Individual proteins at 1 μM were incubated with varying concentrations of saturated ammonium sulphate to a final volume of 50 μL. Saturations of 10%, 20%, 30%, 50%, 60% and 70% were tested for each protein. Reactions were incubated for 30 minutes on ice. Salted protein was pelleted by centrifugation at 18407 x *g* for 2 minutes. Supernatant was discarded and the salt pellet resuspended in SDW. Samples were processed for SDS PAGE analysis and run on 15% acrylamide gels.

Gels were analysed to determine the percentage of saturated ammonium sulphate needed to precipitate each protein. A percentage of ammonium sulphate was chosen where only one protein precipitated. N-HelQ precipitated at 10% and 20% where RPA did not. The assay was repeated with RPA and N-HelQ at 10% and 20% ammonium sulphate. Interactions were determined by salting out of both proteins during analysis.

2.18.6 S1 nuclease protection assay

S1 nuclease protection assays were used to assess HelQ displacement of RPA-DNA complexes. RPA binding was carried out with 25 nM fluorescently labelled ssDNA (MW12) in 1x HB, 1x nuclease buffer (40 mM sodium acetate, 500 mM NaCl, 2 mM zinc sulphate) and 25 mM DTT. The reaction was incubated at 37°C for 10 minutes. HelQ was added with 5 mM ATP and $MgCl_2$ to a final volume of 20 μL to RPA-DNA reactions and incubated at 37°C for 10 minutes. S1 nuclease was added for a further 10 minutes. The reaction was quenched at 65°C and run on a 10% acrylamide TBE gels for analysis.

2.18.7 Chemical crosslinking

Chemical crosslinking was carried out to detect transient interactions between N-HelQ and RPA following the manufacturer's instructions (Thermo Fisher #21590). The crosslinking agent bis(sulfosuccinimidyl)-2,2,7,7-suberate-d_0 (BS^3-d_0) was used which reacts with primary amine groups on protein side chains at pH 7-9 to form stable amide bonds. Each protein at 1 µM was cross linked individually to assess for protein populations and in pools of RPA and N-HelQ together. The crosslinker was tested at 0.5, 1, 2 and 5 mM and optimised for use at 0.5 mM. The assay was carried out at room temperature for 30 minutes, 60 minutes and 90 minutes in the presence and absence of fork 2b DNA (25 nM) before the reaction was processed for SDS PAGE analysis as described previously. Acrylamide gels were analysed for additional protein bands to suggest complex formation between N-HelQ and RPA.

2.19 BacterioMatch II Hybrid system

BacterioMatch II hybrid system was carried out following manufacturer's instructions (Stratagene #200412).[291] The method was developed by Dove, Joung and Hochschild[292,293] and further refined by Joung and Pabo.[294] The system is used to identify transient and novel interactions between target and bait proteins. This method was used to discover any minute interactions between N-HelQ and RPA. Detection of protein-protein interactions is based on transcriptional activation of the *HIS3* reporter gene which allows growth in the presence of 3-amino-1,2,4-triazole (3-AT), a competitive inhibitor of HIS3 enzyme. Transcription of the *HIS3* gene only occurs if N-HelQ, fused to the Lambda-cl element and interacts with the lambda operator, interacts with RPA70, fused to the RNA polymerase α subunit.

2.19.1 Plasmid construction

N-HelQ was prepared for insertion into bait pBT plasmid ($Chlm^R$). N-HelQ DNA was cloned from pTJ9 by PCR as described previously into pBT (3.2 kb) using EcoRI and XhoI restriction enzyme cut sites. The target plasmid, pTRG (Tet^R) was prepared for insertion of RPA70 and RPA14 DNA independently. RPA70 and RPA14 encoding DNA were cloned from p11d-tRPA independently into pTRG (4.4 kb) using BamHI and XhoI enzyme cut sites as described previously. Only RPA70 was successfully inserted into pTRG, RPA14 was unsuccessful at ligation into the target plasmid.

pBT and pTRG vectors were toxic when grown and therefore optimisation of competent cell strains was tested at both 30°C and 37°C. This protocol varied to the manufacturer's instructions. The vectors were made using NEB 5α cells and grown at 30°C. pBT-N-HelQ and pTRG-RPA70 were made and verified by restriction digest and sequencing.

2.19.2 N-HelQ bait plasmid validation

Expression of N-HelQ from pBT was verified by transformation into BacterioMatch II validation reporter competent cells (#200192). Empty vector (pBT-Ev) was used as a control. Transformation was carried out following standard protocols with a longer recovery time of 2 hours at 37°C. Cells were plated on LB-chloramphenicol plates and incubated at 30°C for 24 hours. Individual colonies were grown in 2 mL M9+ His-dropout broth (table 2.13) with 25 µg/mL chloramphenicol at 37°C for 16 hours. Fresh cultures were inoculated with a 1:100 dilution into fresh 2 mL M9+ His-dropout broth with 25 µg/mL chloramphenicol and 10 µM IPTG to induce protein expression. Cells were grown at 37°C to 0.5-0.6

OD_{600}. SDS PAGE samples were collected and processed for immuno-blot analysis (anti-HexaHis) to determine N-HelQ expression when compared to the empty vector control.

2.19.3 Testing self-activation by recombinant pBT-N-HelQ

Determination of the ability of the bait fusion protein in activating the reporter cassette in the absence of an interaction partner (RPA70) was required to determine false positives. No activation is required in the absence of RPA. pBT-N-HelQ and pTRG empty vector (Ev) (50 ng) were co-transformed into validation reporter cells. A negative control co-transformation of pBT-Ev with pTRG-Gal11^{P} was used. Transformation was carried out as described previously except the cells were heat-pulsed at 42^{o}C for 35 seconds and recovered in SOC medium (table 2.11). The initial recovery phase was carried out at 37^{o}C for 90 minutes with shaking. The cells were pelted at 2000 g, washed and resuspended in M9+ His-dropout broth. Cells were further recovered for 2 hours at 37^{o}C with shaking. The cells were then plated on the appropriate media, shown in table 2.13.

For plating onto the selective screening medium (+5mM 3-AT), 200 µL cells were plated (table 2.13). For Nonselective screening medium (-3-AT), 100 µL was removed from the 1 mL culture and diluted 1:100, 20 µL and 200 µL were plated. Plates were incubated at 37^{o}C for 24 hours. A second incubation of 16 hours allowed growth of cells for weak interactors or toxic proteins.

The plate with a 'countable' number of colonies on nonselective medium was used to determine the number of co-transformants obtained (table 2.13). A significant number of colonies was expected. Colonies were counted and measured as colony forming units (cfu). The percentage of co-transformants able to grow on 5 mM 3-amino-1,2,4-triazole (3-AT) was calculated by determining the percentage

of colonies on the selective medium compared to the adjusted number of colonies on the nonselective medium (adjusted for dilutions). A suitability of pBT-N-HelQ was obtained with a result of <0.1% co-transformants, verifying the absence of bait self-activation. Self-activation was also tested for pTRG-RPA70.

Table 2.17 Co-transformation combinations carried out to test self-activation of bait and target plasmids in B2H.

Plasmids	Plate	Purpose	Expected results
pBT-N-HelQ + pTRG-Ev	Nonselective screening Medium (no 3-AT) at 1:100; 20 and 200 µL	Measure number of co-transformants	10^2-10^3 cfu (corrected for dilutions)
	Selective screening medium (+5 mM 3-AT) with 200 µL	To determine if bait protein can activate the reporter cassette alone	<0.1% of total transformants verifies absence of bait self-activation. 0.1-1% Requires troubleshooting. >1% Not suitable due to self-activation.
pBT-Ev + pTRG-Gal11^P	Nonselective screening medium (no 3-AT) at 1:100; 20 and 200 µL	Measure number of co-transformants	10^2-10^3 cfu (corrected for dilutions)
	Selective screening medium (+5 mM 3-AT) with 200 µL	Negative control (non-interacting pair)	<0.1% as expected, verifies 3-AT selection is working. >0.1% indicates failure of 3-AT selection. Requires troubleshooting.
pBT-Ev+pTRG-RPA70	Nonselective screening medium (no 3-AT) at 1:100; 20 and 200 µL	Measure number of co-transformants	10^2-10^3 cfu (corrected for dilutions)
	Selective screening medium (+5 mM 3-AT) with 200 µL	To determine if bait protein can activate the reporter cassette alone	<0.1% of total transformants verifies absence of bait self-activation. 0.1-1% Requires troubleshooting. >1% Not suitable due to self-activation.
pBT-LRF2+pTRG-Ev	Nonselective screen medium (no 3-AT) at 1:100; 20 and 200 µL	Measure number of co-transformants	10^2-10^3 cfu (corrected for dilutions)
	Selective screening medium (+5 mM 3-AT) with 200 µL	Negative control (non-interacting pair)	<0.1% as expected, verifies 3-AT selection is working. >0.1% indicates failure of 3-AT selection. Requires troubleshooting.

2.19.4 Co-transformation to test for interactions

N-HelQ and RPA70 interaction screen was carried as described in the previous section. Plasmid combinations (table 2.18) were co-transformed into the B2H validation reporter strain and plated on selective screening medium plates (+5 mM 3-AT) and nonselective screening medium.

Table 2.18 Co-transformations carried out for B2H screening.

No.	Bait Plasmid	Target Plasmid	Purpose
1	pBT-N-HelQ	pTRG-RPA70	Testing pair
2	pBT-N-HelQ	pTRG-Ev	Negative control
3	pBT-Ev	pTRG-Gal11^{P}	Screening test validation
4	pBT-Ev	pTRG-RPA70	Negative control
5	pBT-LGF2	pTRG-Ev	Screening test validation
6	pBT-LGF2	pTRG-Gal11^{P}	Positive control
7	pBT-LGF2	pTRG-RPA70	Negative control
8	pBT-N-HelQ	pTRG-Gal11^{P}	Negative control

2.19.5 Analysis of B2H co-transformations

Colony numbers of each plate was calculated. Dilution factors were corrected for nonselective plates. The percentage of successful co-transformants on selective medium (+5 mM 3-AT) was calculated from the total number of possible co-transformants on the nonselective medium.

2.20 Analytical gel filtration (AGF) using a superdex200 column

2.20.1 Column preparation and calibration

AGF is a form of size exclusion chromatography and was used in protein purification and determination of oligomeric state. A Superdex 200 column (table 2.15) was loaded onto an Amersham pharmacia Biotech AKTA FPLC System and washed with SDW and buffer A (150 mM NaCl, 10% glycerol, 1 mM DTT, 20 mM Tris pH 7.5) at 0.25 mL/min. The column was kept at 4^{o}C.

The column was calibrated with the GE Healthcare calibration high molecular weight (HMW) standards kit (#10196234). Column void volume was calculated by

the initial running of 500 µL Blue Dextran (2,000,000 Mr) at 0.45 mL/min. The void volume is the initial volume that elutes off the column which no protein will reliably elute. The following standards were then run through the column as two pools: Ovalbumin (43000 Da), Conalbumin (75000 Da), Aldolase (158000 Da), Ferritin (440000 Da) and Thyroglobulin (669000 Da). A standard curve was produced of known molecular weight against the elution volume using PRISM 7 GraphPad.

2.20.2 Analysis of protein populations to determine elution volumes

Buffer conditions varied according to the protein and was also varied to assess the impact on buffer components on oligomeric state. The column was equilibrated with two column volumes of buffer prior to sample run. Samples were prepared on ice and incubated in different conditions prior to loading as stated in results. Protein samples were made up to 500 mL with buffer and injected into a 0.5 mL loop. The loop was loaded onto the column and run at 0.25 mL/min for 1.2 column volumes. Fractions were treated with SDS and analysed by SDS PAGE to confirm the presence of protein and EMSA/ helicase assay analysis was also carried out. The column was cleaned by washing with 0.2 M NaOH followed by SDW at 0.25 mL/min and stored long term in 20% ethanol.

2.21 Biophysical methods to determine protein structure and interactions

2.21.1 Dynamic Light Scattering (DLS)

DLS was carried out using a Malvern Zetasizer Nano S machine with 40 µL cuvettes (Sigma, #Z637939). Samples were prepared by filtration through appropriately sized filters (GE Healthcare) to remove larger contaminants. Proteins were dialysed into 150 mM NaCl, 10% glycerol and 20 mM Tris pH 8. Measurements were adjusted for buffer viscosity which can impact particle size.

Guidelines for accurate DLS suggest protein concentrations between 0.5-2.0 mg/mL should be used, however, due to the HelQ and C-HelQ instability, this yield was not reached. Therefore, readings were taken at the highest concentration available.

Protein was loaded into the cuvette and incubated with activating components as stated in results, on ice. The cuvette was then placed into the DLS machine. All readings were taken at 37°C. A total of 10 readings per sample were collected and analysed using Zetasizer Ultra-Pro instrument Software provided by Malvern Panalytical. Results were given as percentage intensity vs. diameter (nm) and % volume vs. diameter (nm).

2.21.2 Circular Dichroism (CD)

CD was carried out under standard conditions.[295,296] A 24-hour purge directly prior to give the machine time for the high tension (HT) voltage to normalise.[297] An HT value less then 1.3 suggests appropriate purging and the machine is ready to use. The program was set up and run using the Chirascan software. Protein sample (350 µL at 1 mg/mL) dialysed into 5 mM Tris pH 7.5 was loaded into a standard Hellma® 1 mm absorption cuvette. All experiments were carried out at 25°C. The wavelength measurement was optimised for measurable HT values by adjustment for each protein tested. An HT voltage of above 600 suggested the machine is saturated and would result in unreliable spectrum oscillating and therefore requires repeating at lower concentrations.[297] Results analysed and plotted in PRISM 7 GraphPad as CD in $mdeg.cm^{-1}.\mu mol^{-1}$ against wavelength in nm. Cuvettes were rinsed with Hellmanex and SDS and left to dry completely before reuse.

2.21.3 Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC MALS)

Size exclusion of HelQ proteins was carried out using a Superdex200 column at 1 mL/min as described previously. Multi-angle light scattering was detected using WYATT technology DAWN HELEOS8 light detector over 30 minutes at 25°C and analysed using ASTRA. SEC MALS was carried out on HelQ under different conditions. Condition A (50 mM Tris pH 8, 150 mM NaCl, 10% glycerol) condition B (condition A with 0.2 mM ATP, 0.2 mM $MgCl_2$, HelQ incubated in 5 mM $MgCl_2$ and 5 mM ATP at 37°C for 10 minutes) and condition C (condition B with HelQ incubated in 5 mM $MgCl_2$ and 5 mM ATP with 25 nM Cy5 end-labelled ssDNA at 37°C for 10 minutes).

2.21.4 Analytical Ultra-Centrifugation (AUC)

AUC is the study of macromolecule behaviour in solution when under the influence of a strong gravitational force.[298] AUC was carried out on a Beckman ProteomeLab XL AUC system. This method is used to characterise proteins in their native form. Mass becomes redistributed in a gravitational field until the gravitational potential energy balances the chemical potential energy. AUC measures the rate at which molecules become redistributed, known as the sedimentation velocity, providing information on the size and shape. AUC uses absorbance and requires an absorbance of 0.2-1 OD_{600} to test a protein.

This work was carried out at the RCaH complex in Harwell with the help and guidance of Gemma Harrison. Three concentrations of N-HelQ were tested: 0.25, 0.5 and 1 mg/mL at room temperature. N-HelQ was initially dialysed into 20 mM Tris pH 8 to remove glycerol and salt using in the storage buffer.

Extensive cleaning to ensure no contaminating proteins was required prior to accurate and precise loading of protein into the AUC cells. Software created an

absorbance spectrum of the sedimentation, in which sedimentation velocity of each protein species was calculated. Exact protein molecular weight was calculated from this.

2.21.5 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a physical technique used to determine thermodynamic parameters of interactions in solution.[299] ITC was carried out using a MicroCal iTC200 system in collaboration with the RCaH. ITC measures heat directly released from a reaction under a constant pressure, known as the Q_T value. This is proportional to the molar enthalpy change associated with the process, ΔH. The calculation used is $Q_T = V_o [M_T] \Delta H$; where V_o is the protein volume within the calorimetric cell and $[M_T]$ is the total concentration of protein in the cell.[299] A change in temperature is monitored on addition of protein 2 to protein 1, which is used to indicate interactions.[299]

RPA and N-HelQ were gel filtrated through S200 to remove contaminants and buffer exchanged into PBS. Protein 2, N-HelQ, with the highest concentration at 200 μM, was loaded into the syringe to 60 μL. Protein 1, RPA, was loaded into the calorimetric cell at 5 μM to a total volume of 200 μL. Ideally, this would have been at a concentration of 20 μM to allow for accurate readings.

The protein filled syringe was fully inserted over 17 pumps allowing for readings to be taken after each pump. This allowed for measurement of the change in temperature throughout addition of N-HelQ into RPA-containing cell. The syringe was pumped at a rate of 180 seconds per pump at 750 rpm paddle speed in the cell. These conditions are standard procedure for initial ITC measurements.[299] Sigmoidal patterns monitoring the temperature of RPA were detected and monitored over the course of the experiment as N-HelQ was loaded.

2.21.6 Setting up plates for crystal trials

Crystal trial screen plates were set up with N-HelQ using a mosquito® crystal liquid handler (SPT LabTech). Three plates with variations in buffer composition in order to optimise conditions for the protein were set up for incubation at both room temperature and 4°C. Plates were prepared in at 4°C. N-HelQ was purified and gel filtrated into 100 mM NaCl, 20 mM Tris pH 8 and 1 mM DTT. The protein was spin concentrated to 5 mg/mL.

Three 96 well plates that were set up included standard screen plates: Index HT (HR2-134), Crystal Screen HT (HR2-130) and Wizard 122 supplied from Hampton Research. These screening matrices were selected as they are simple components suitable for finding initial crystallisation conditions. The plates were stored and monitored over months for crystal formation.

2.21.7 Fluorescence anisotropy

Protein-DNA interactions were detected using the PerkinElmer EnVision benchtop plate reader (#2105) following standard protocols. Fluorescein labelled DNA oligos were made following described previously methods. Software used was Wallac EnVision Manager using the FP Fluorescein Dual filters. The software converted fluorescence emission values into measured fluorescence polarisation (mP) values using the equation mP = 1000 x (S-GXP)/(S+GXP); Where S and P are the emission filters and G is a factor to correct for effect of the emission filter transmission variations. Reactions were set up with 1x HB, 25 mM DTT, labelled DNA (to optimised concentration) and loaded into a 384 Nunc black plate. HelQ protein serial dilutions (indicated in results section) were set up and added to the reaction plate to a final volume of 50 µL. Readings were taken at 0, 5, 10 and 15 mins at 30°C. The results were blanked to no protein and plotted as change in mP against protein concentration in PRISM 7 GraphPad.

2.21.8 Microscale Thermophoresis (MST)

MST is a biophysical technique that was used to measure protein-protein interactions between N-HelQ and RPA.[300] The technique is based on the detection of a temperature-induced change in fluorescence of a target protein when a non-fluorescently labelled substrate is added. These readings are used to determine binding affinities. Experiments were carried out using the Nanotemper technologies standard protocol and standard software (Nanotemper Technologies Monolith NT.115). N-HelQ was labelled using the Nanotemper Monolith hexaHis-Tag labelling kit RED Tris-NTA (#NT-L118). A stock reaction mix was assembled with 2x HB, 50 mM DTT and 100 nM labelled N-HelQ and aliquoted into 16 x 10 µL reactions. A two times concentrated serial dilution of RPA was made with starting concentration of 80 nM. 10 µL of the serial dilation was added to each reaction mix, giving tube 1 a final RPA concentration of 40 nM. The mix was fed into the Nanotemper capillaries and the run started. Runs were carried out at 20% and 40% MST power with 1% excitation using a K_d model following manufacturer's instructions. Experiments were repeated at 25ºC and 37ºC and tested at a range of protein concentrations in the presence and absence of 25 nM unlabelled ssDNA (MW12).

2.22 Bioinformatic analysis and molecular modelling

A range of bioinformatical tools was used throughout this work receive DNA and protein sequences, carry out homology search's, sequence alignments, structure prediction modelling and interaction predictions. Most databases used are described in results chapters. Below are more regularly used resources.

2.22.1 Uniprot and ExPASy: sequence mining

Protein amino acid sequences were extracted from the Universal Protein Resource (Uniprot) database.[301] Sequences were obtained in FASTA format of single letter amino acid code from protein accession numbers. This format was chosen as the compatible format for other bioinformatical tools. ExPASy[282] was used as an online tool to extract information about proteins based on FASTA sequence input. These included protein pI value, extinction coefficient, predicted molecular weight and amino acid composition.

2.22.2 BLAST and CLUSTAL: homology search

Protein alignments were generated from FASTA input sequences using multiple tools. These tools use different algorithms, including pairwise and multiple sequence alignment. CLUSTAL OMEGA[302] and BLAST[303] were used to identify sequence homology between protein sequences of different species to give alignment and percentage homology output.

2.22.3 PHYRE2, Raptor and Pymol: protein structure predictors

Protein amino acid sequences were inserted into PHYRE2[304] and Raptor[305] to produce predicted 3D models of amino acid composition in space. Extensive modes of prediction were used for both tools. PBD files were extracted and inserted into Pymol[306] modelling software to annotate and highlight protein regions. High quality images and videos were produced using Pymol.

2.22.4 COTH Zhanglab and Symmdock: protein interaction predictors

Protein-protein interactions between both monomers of the same protein and potential protein partners were predicted using online tools. PHYRE2 PDB files of two proteins were imported into the co-threading of protein-protein complex structures tool (COTH online) developed by the Zhanglab at the University of

Michigan[307] in which 10 composition predictions in space were produced in PDB format for interpretation using Pymol. FASTA amino acid code was imported into Symmdock[308] with oligomeric state estimations to produce predictions of oligomeric structure.

Chapter 3: Analysis of the organisation and oligomeric state of HelQ

3.1 Introduction.

HelQ is a human Ski2-like SF2 DNA helicase that is implicated in cell survival and genome stability. HelQ is a ssDNA-dependent ATPase that translocates DNA with 3′ to 5′ directionality separating DNA duplex strands within fork structures, thus acting as a helicase.[256,265] Due to the lack of atomic resolution structures available, HelQ has been analysed *in silico* using PHYRE2 and I-TASSER to produce predictive structures. Based on these studies, HelQ can be divided into a conserved C-terminal core helicase region with two RecA-like domains, a winged-helix domain (WHD) and a DNA binding ratchet and an N-terminal region with an unknown function which lacks obvious sequence homology.

HelQ has highest amino acid sequence homology to the Ski2-like DNA repair helicase Hel308 from archaea.[242] Hel308 has been characterised using biochemistry to build a high-resolution crystal structure from *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* as well as by using genetics where genetic screens for synthetic lethality identified similar functions to RecQ.[242,245,247,264,309,310] The biochemical mechanisms by which HelQ is able to support DNA repair through the control of homologous recombination are not known, therefore, Hel308 has been used to understand more about HelQ. However, it is unclear if the many studies on Hel308 are good models for HelQ.

This chapter describes the purification of full length human HelQ protein (FL-HelQ) for detailed biochemical analysis, and dissection of HelQ protein fragments from its N-terminal (N-HelQ) and C-terminal (C-HelQ) regions. The three main objectives were; (a) to purify HelQ and the HelQ fragments to a higher yield; (b) to determine the active oligomeric state of HelQ in DNA binding and translocation;

(c) to define the contributions made to FL-HelQ function by N-HelQ and C-HelQ components.

3.2 Purification of HelQ using baculovirus expression system.

DNA encoding human HelQ was cloned and over-expressed from baculovirus in insect cells. A former PhD student in the Bolt group, Sarah Northall, optimised the production of soluble HelQ in Sf9 insect cells and developed a reproducible method for purification.[252] Further work carried out here improved protein yield for biophysical analysis. This was important because structural protein characterisation requires milligram quantities of HelQ that have not been achieved previously.

Bioinformatics analysis of human HelQ sequence indicates that it has a significant degree of protein instability. The calculated stability index (II) value, determined using the ProParam ExPASy tool described in section 2.23.1[282], is 45.58 with an estimated half-life of 1.1 hours *in vitro* (figure 3.1). The stability index is the volume of protein occupied by aliphatic side chains (alanine, valine, isoleucine and leucine) and is used to measure the instability of a protein (more than 40 is unstable). HelQ was predicted to be highly unstable suggesting this would need to be considered during recombinant protein expression. This was confirmed during purification because prolonged exposure of HelQ during the multiple steps of purification resulted in extensive degradation of FL-HelQ being observed to produce smaller protein fragments. In addition, figure 3.1 shows the distribution of residues within HelQ sequence and further calculations including the extinction coefficient used to calculate protein concentration.

A.

Amino Acid	Number of residues	As a % of total residues	Amino Acid	Number of residues	As a % of total residues
Ala	57	5.2	Leu	136	12.4
Arg	46	4.2	Lys	87	7.9
Asn	51	4.6	Met	22	2.0
Asp	50	4.5	Phe	37	3.4
Cys	22	2.0	Pro	39	3.5
Gln	44	4.0	Ser	89	7.8
Gly	64	5.8	Thr	66	6.0
Glu	95	8.6	Try	6	0.5
His	23	2.1	Tyr	42	3.8
Ile	61	5.5	Val	66	6.0

B.

ExPASy Measure	Value
Number of residues	1100
Molecular mass	124044.66
Theoretical pI Value	6.12
Half life	1.1 hours
Instability Index	45.58
Extinction Coefficient	96955 $M^{-1}cm^{-1}$

Figure 3.1 ExPASy proteomics analysis of HelQ residue composition. A. ExPASy ProPram analysis of the composition and distribution of residues in human HelQ. HelQ is made up of a relatively equal number of each residue with the exception of a high percentage of leucine residues which contribute to the high instability index. **B.** HelQ, with a predicted molecular weight of 124 kDa, has a short half-life of 1.1 hours requiring rapid purification to reduce levels of protein break down. The protein does not exist long within the cell. The instability index, 45.58, classifies HelQ as unstable (>40 is unstable). The extinction coefficient was used to calculate molar concentration and the isoelectric point (pI) was used for purification optimum buffer composition.

Different affinity protein tagging strategies were investigated in attempts to maximise protein recovery during purification. HelQ encoding DNA was cloned into baculovirus expression plasmids containing pACEBacI to generate MBP-tagged and StrepII-tagged HelQ proteins (figure 3.2, see appendix for plasmid maps). Sf21 insect cells were used for viral transfection because they produce large quantities of virus. Sf21 are fast growing therefore rapid duplication resulted in

larger quantities of virus being produced in shorter periods of time. Additionally, the Sf21 cells are more resilient with a higher probability of surviving viral transfection when compared to other insect cell lines. Viral titres were used to optimise the over-expression of HelQ to 5 mL/L of virus in HiFive insect cells for 46-48 hours at 27^{o}C as described in section 2.12.6. HiFive cells are effective as a standard procedure for protein over-expression in insect cells and these achieved higher levels of over-expressed HelQ protein. The purification of HelQ was modified from previously published protocols to reduce the length of time used from 72 to 24 hours, which was found to improve HelQ yield. Therefore, we gained a higher protein yield but reduced purity. A full list of the modifications made to the protocol can be seen in section 2.13.1.

HelQ was resolved by binding to Hi-Trap Ni-NTA, DNA-binding heparin and Q-sepharose affinity chromatography columns (figure 3.2 A-C). These methods utilised the HelQ hexaHis-tag, and its DNA binding affinity (heparin) and charge at pH 8. Size exclusion chromatography using a Superdex200 column was used to isolate FL-HelQ from its smaller degraded protein fragments. However, maximum FL-HelQ purity was not achieved because the smaller degraded fragments could not be isolated from FL-HelQ. This is because FL-HelQ aggregated or oligomerised with degraded protein fragments of HelQ (figure 3.2 D). Protein mixes of FL-HelQ and fragments were aliquoted to reduce further protein degradation by freeze thaw. This resulted in a higher concetration of HelQ, however, a reduced purity then previously reached.

An ATPase inactive mutant HelQ with a single point mutation (Asp-463-Ala, HelQ^{D463A}) in the Walker B domain was used as a negative control (figure 3.3). The Walker B domain is important for ATP hydrolysis and is therefore essential for HelQ function. Mutation of this residue knocks out the activity of ATP hydrolysis.

Protein stability problems in this mutant, similar to those of the wild type, required equivalent purification alterations.

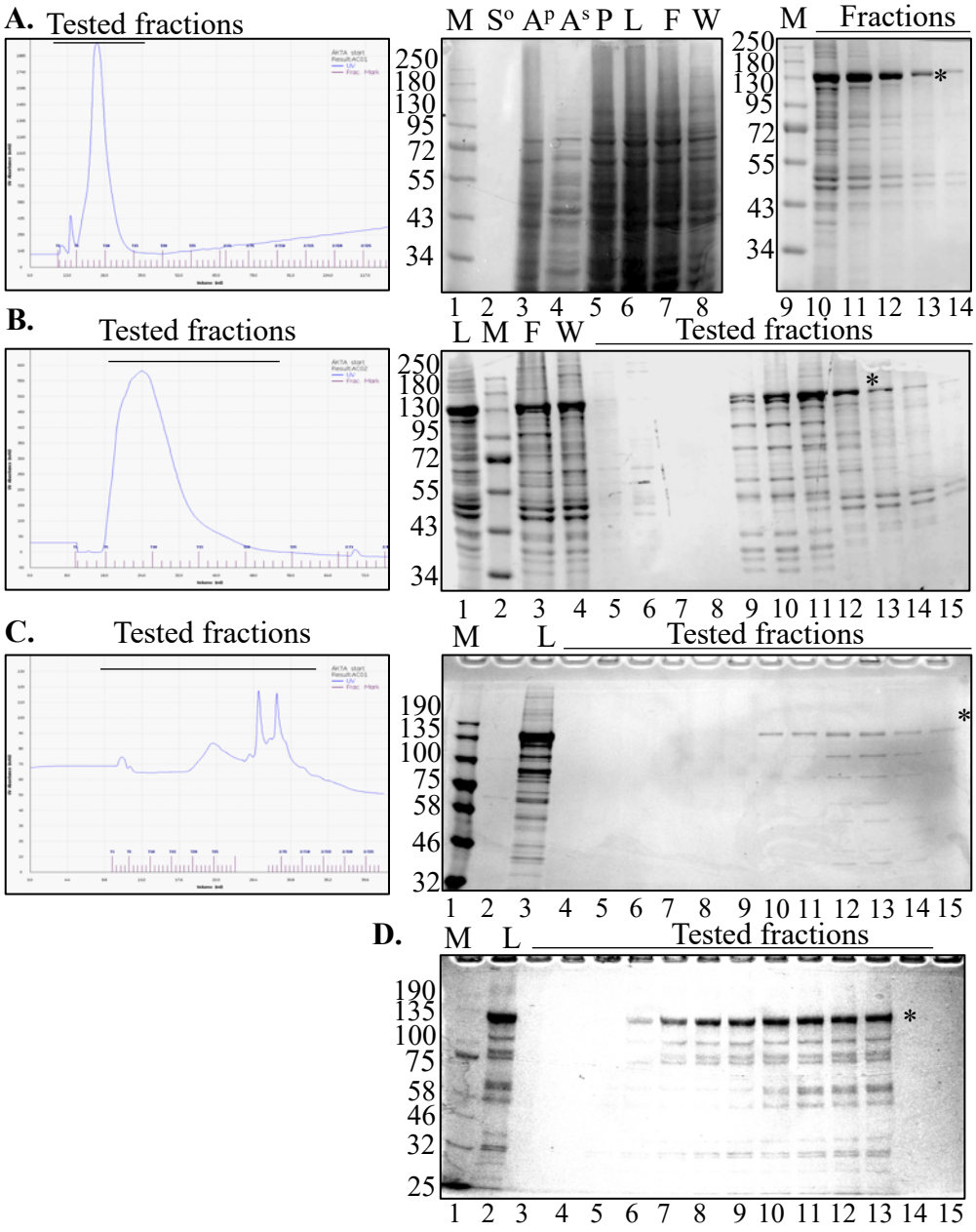


Figure 3.2 FL-HelQ protein purification. HelQ protein was over-expressed using baculovirus infected Sf9 insect cells. Biomass containing HelQ was sonicated and clarified using an ammonium sulphate precipitation prior to affinity purification. The panels show HelQ (141 kDa) in Coomassie stained 8% acrylamide SDS PAGE and UV trace analysis indicated by (*). **A.** Protein containing lysate was loaded onto a $NiCl_2$ charged HiTrap Ni-NTA column and eluted at 100-200 nM imidazole. Legend of samples during purification include marker (M), sonication (S^o), ammonium sulphate supernatant (A^S) and pellet (A^P), post hard spin pellet (P), loaded supernatant (S), loaded sample (L) flow-through (F) and wash-through (W). **B.** Pooled HelQ was dialysed and loaded on a 5 mL heparin column and eluted at 100-400 mM NaCl. **C**. HelQ was further isolated on a 1 mL Q-sepharose column and eluted at 350-600 mM NaCl. **D.** Size exclusion of HelQ did not isolate FL-HelQ from degraded HelQ protein fragments.

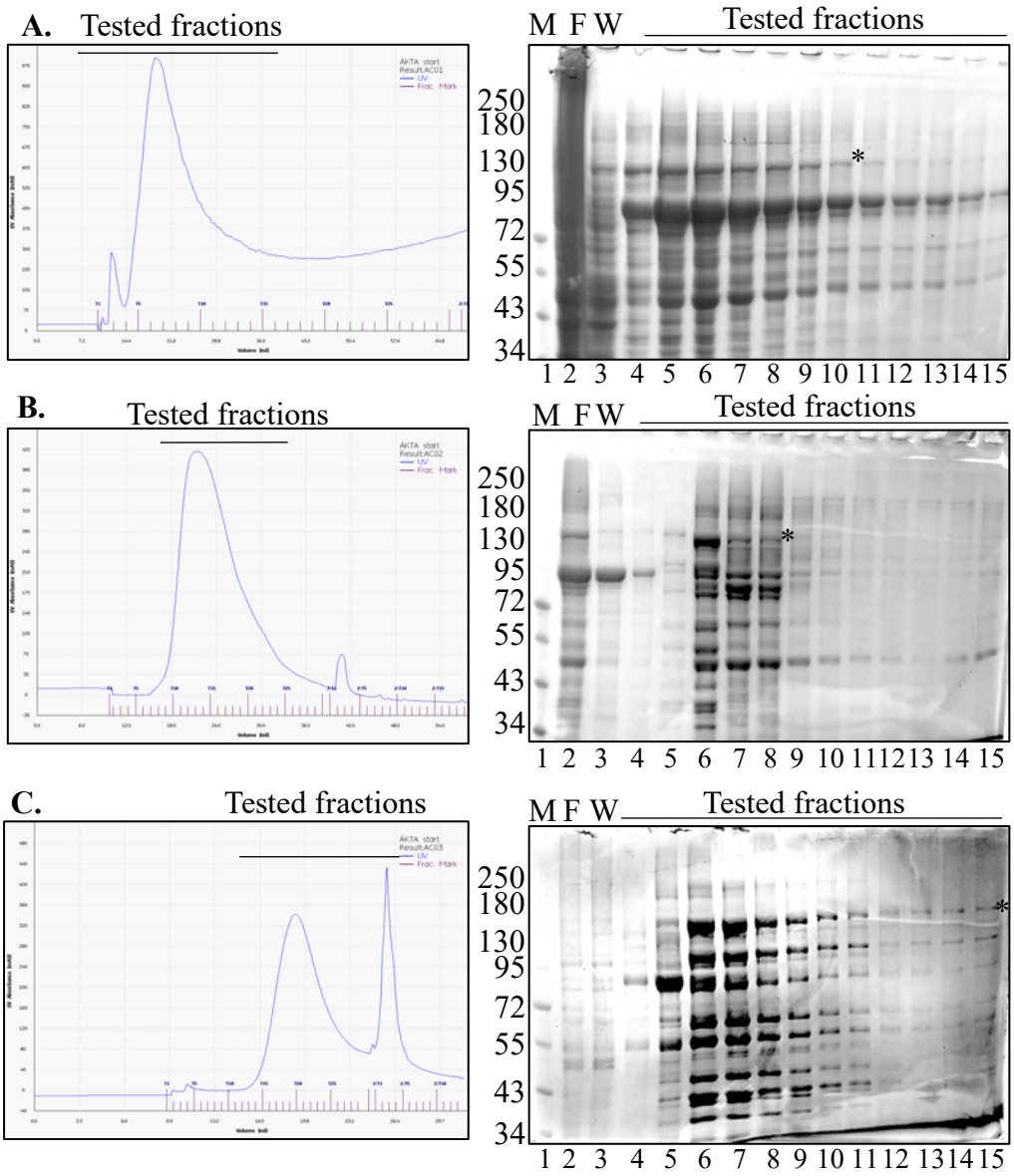


Figure 3.3 FL-HelQ^{D463A} protein purification. An ATPase inactive FL-HelQ mutant was made with a point mutation in the Walker B domain (Asp-463-Ala) and over expressed using baculovirus infected Sf9 insect cells. Biomass containing HelQ^{D463A} was sonicated and clarified using ammonium sulphate precipitation prior to affinity purification. The panels show FL-HelQ^{D463A} (141 kDa) in Coomassie stained 8% acrylamide SDS PAGE and UV trace analysis as indicated (*). **A.** HelQ containing biomass was loaded onto a $NiCl_2$ charged HiTrap Ni-NTA column and eluted at 100-200 nM imidazole. **B.** HelQ was pooled and dialysed before loading on a 5 mL heparin column and eluted at100-400 mM NaCl. **C**. FL-HelQ^{D463A} was further resolved using a 1 mL Q-sepharose column and eluted at 350-600 mM NaCl. A high level of degradation was observed throughout the purification. Legend = marker (M), loaded sample (L), wash-through (W), flow-through (F).

3.3 Prolonged expression of FL-HelQ resulted in natural protein degradation.

At the onset of this work, protein preparations of HelQ resulted in protein degradation into stable HelQ fragments. We noticed that HelQ protein over-expression between 48 and 72 hours resulted in the formation of a stable HelQ fragment that was identified by mass spectrometry as the 46.36 kDa tagged N-terminal HelQ region (N-HelQ) ending at Isoleucine 240 (see SNAPGENE files in appendix). This includes the predicted weight of N-HelQ at 34 kDa with the SUMO-hexaHis tag addition. The N-HelQ protein fragment does not contain any of the core helicase regions required for translocation and unwinding and is discussed in detail in the next sections. The fragment was considered stable enough for purification using an *E. coli* expression system as described in section 2.14.1. Differences in post translation modifications of N-HelQ *in E. coli* versus insect cells required consideration which is discussed in more detail later.

3.4 Purification of N-HelQ from *E. coli.*

The DNA encoding N-HelQ were sub-cloned from pSN52 into pET14b including the hexaHis- and SUMO- affinity tags, for expression of the 240 amino acid N-HelQ protein (figure 3.4 B-C). Purification of N-HelQ from *E. coli* showed similar purity and quantities as the protein isolated as a fragment from Sf9 over-expression of FL-HelQ (figure 3.4 A). Coomassie stained gels from SDS PAGE showed N-HelQ to have electrophoretic properties migrating at between 46-58 kDa against the predicted mass (46.36 kDa) but were consistent with the limitations of SDS PAGE for sizing proteins. N-HelQ remained stable throughout purification and was produced in large quantities, up to 5 mg/L, making it amenable for further biophysical studies where FL-HelQ was not.

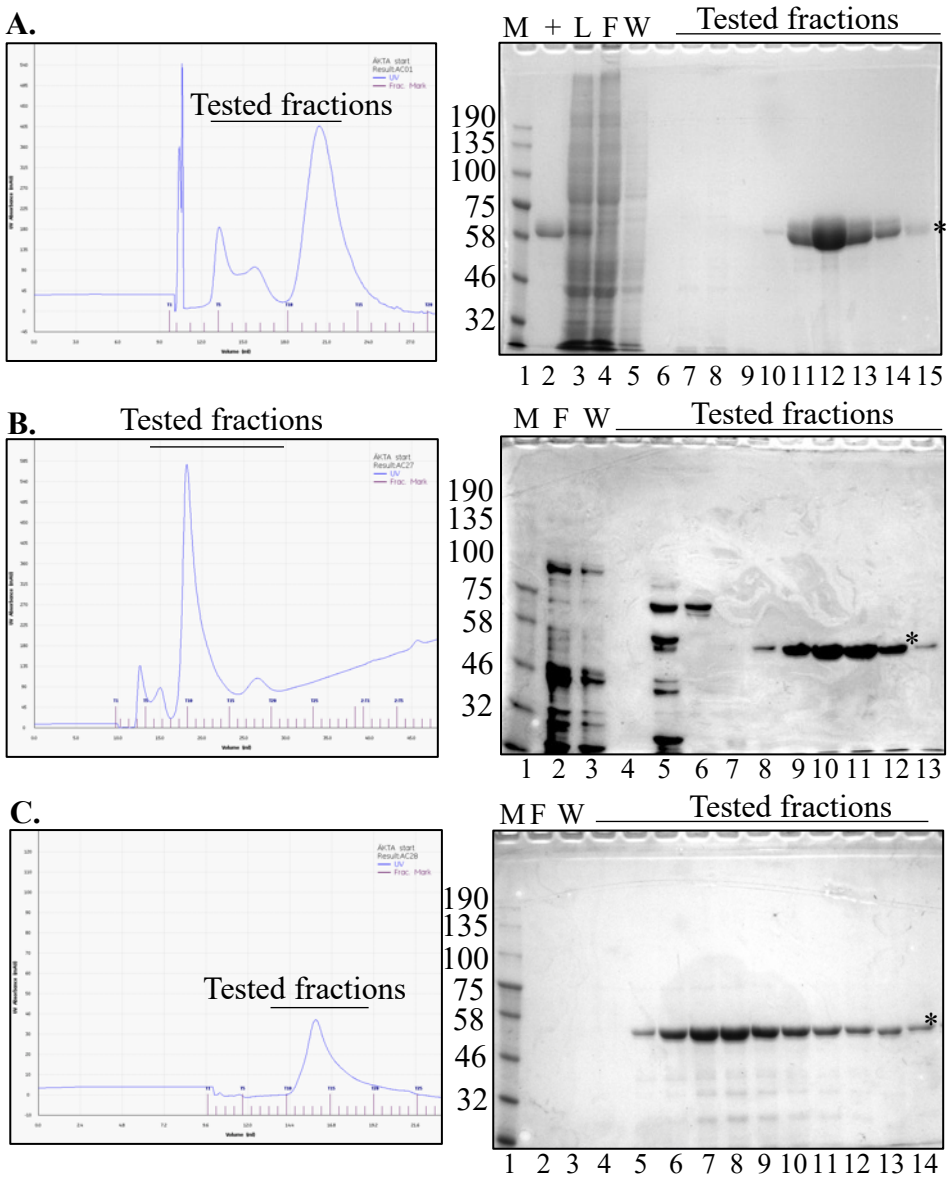


Figure 3.4 N-HelQ protein purification. A. Over-expression of FL-HelQ in Sf9 insect cells for 72 hours resulted in stable degradation to produce an N-terminal fragment. Biomass containing N-HelQ was sonicated and clarified prior to resolution by binding to a HiTrap Ni-NTA column. N-HelQ at 47 kDa (*) was pooled and dialysed for storage. **B.** DNA encoding SUMO-hexaHis-tagged N-HelQ (residues 1-240) was cloned into pET14b (pTJ9) for expression from *E. coli.* Biomass containing N-HelQ was clarified and resolved on a HiTrap Ni-NTA column (*). The panels show N-HelQ in Coomassie stained 8% SDS PAGE and UV trace analysis. **C.** N-HelQ was further resolved by a 1 mL Q-sepharose column. N-HelQ ran slightly higher than 47 kDa and reached high quantities. Legend = marker (M), wash-through (W), flow-through (F).

3.5 Characterisation of core C-HelQ.

We also purified the C-terminal predicted 'core-helicase' of HelQ (C-HelQ) as described in section 2.14.6. Comparisons between FL-HelQ with N-HelQ and C-HelQ fragments aided identification of the roles of each region. C-HelQ comprises predicted HelQ helicase domains: ATPase Walker A and B motifs within RecQ-like domains, a winged helix domain (WHD) and a domain IV that contains a predicted DNA ratchet. These domains are separate from the N-terminal region of unknown function at the onset of this work (figure 3.5 A).

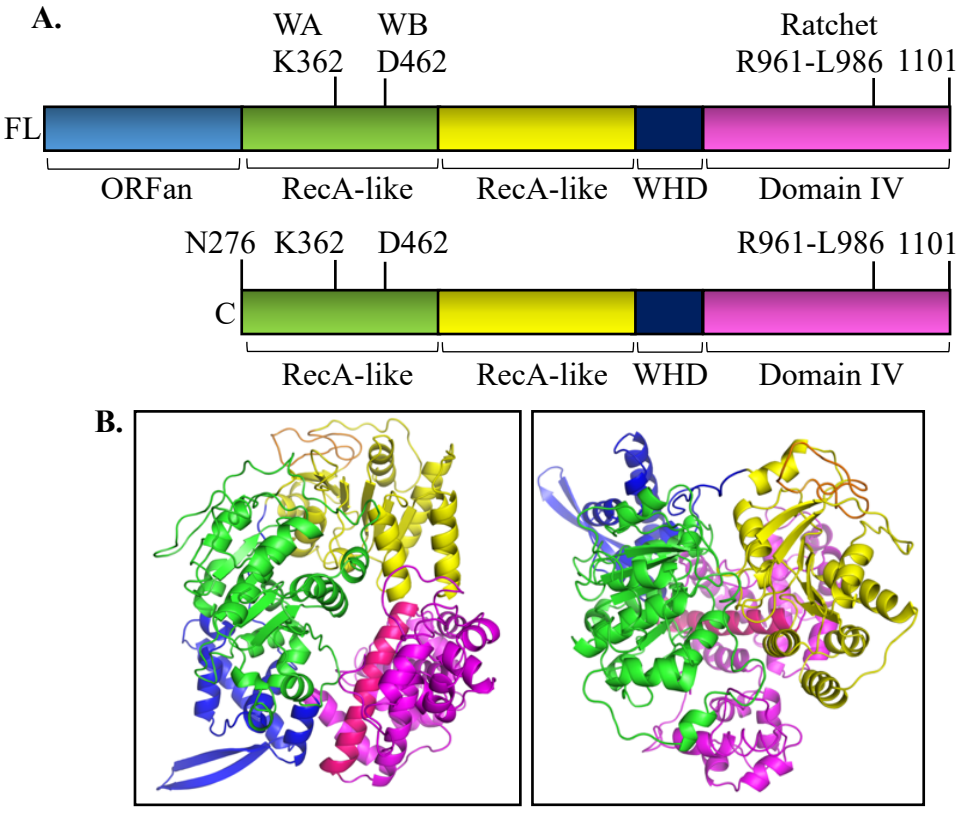


Figure 3.5 C-HelQ is predicted to consist of the 'core-helicase' of FL-HelQ. C-HelQ is the C-terminal 826 residue fragment of FL-HelQ. The DNA encoding C-HelQ was cloned into pACYUduet with an N-terminal hexaHis-tag for expression from *E. coli*. **A.** Cartoon representation of the organisation of FL-HelQ and C-HelQ including the RecA-like domains (green and yellow) with the Walker A (WA) and Walker B (WB) required for ATP hydrolysis, the winged-helix domain (WHD; navy) and domain IV (pink) with the DNA ratchet at residue positions 961-986. FL-HelQ also has the N-terminal region of unknown function (blue). **B.** Predictions based on known structures of Hel308 and PolQ were made using PHYRE2 to model the structure of C-HelQ shown in two orientations. Coloured regions refer to schematic in **A.**

C-HelQ is the work-horse of translocase and helicase activity (figure 3.5 B) and has strong sequence homology to the archaeal homologue Hel308 in comparison with FL-HelQ containing the N-terminal region (shown in more detail in figure 3.15). Therefore, C-HelQ was also used to assess the reliability of Hel308 as a model for HelQ.

3.6 Purification of C-HelQ from *E. coli.*

Work done by an undergraduate student Hannah Betts[311] led to successful cloning of DNA encoding C-HelQ into pACYCduet, for expression with a C-terminal hexaHis-Tag in *E. coli.* C-HelQ starts at position Asp-275 of FL-HelQ and has a calculated size of 96.3 kDa inclusive of the tag (figure 3.6). Cloning and protein over-expression of C-HelQ was problematic because of large amounts of protein degradation similar to FL-HelQ purification. Additionally, FL-HelQ is unable to be purified in *E. coli* due to toxicity and therefore C-HelQ cloning in *E. coli*, while possible, was also difficult. High levels of degradation suggested the core helicase region was responsible for HelQ instability.

The fractions containing C-HelQ from Ni-NTA columns showed brown colouration (figure 3.6 B). This facilitated subsequent rapid purification of C-HelQ without needing SDS PAGE analysis. C-HelQ was purified by subsequent Q-sepharose and size exclusion chromatography, for the latter eluting between 75-100 kDa (figure 3.6 C). Coomassie stained gels from SDS PAGE showed C-HelQ to have electrophoretic properties migrating at 85 kDa, smaller than the predicted size of 96 kDa. However, C-HelQ was confirmed by mass spectroscopy. This may be due to degradation resulting in a smaller more stable fragment or variations in protein structure impacting migration. Due to protein instability and degradation, recovery of C-HelQ resulted in low protein yield. Additionally, an ATPase inactive C-HelQ fragment (C-HelQ^{D463A}) was purified.

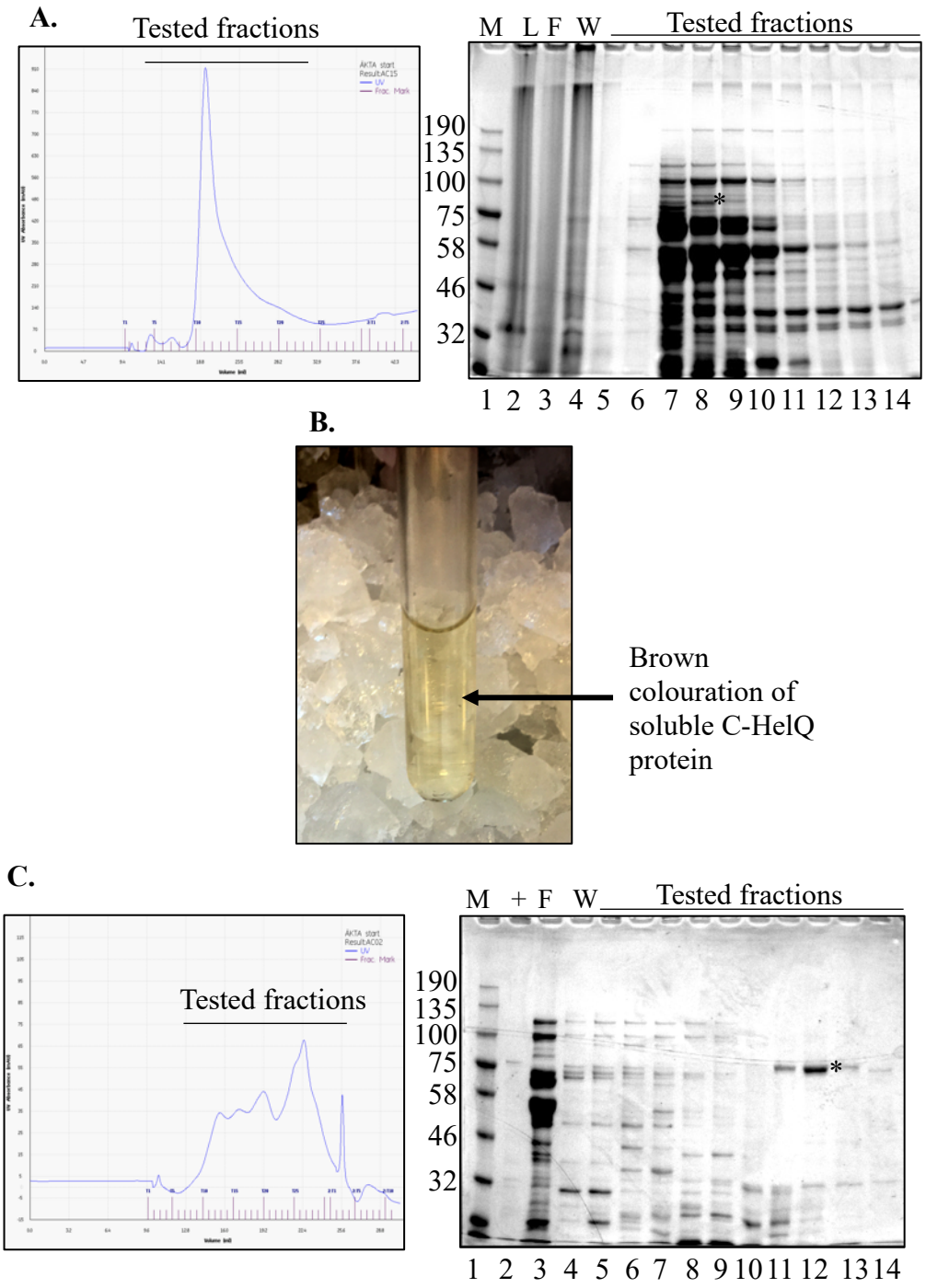


Figure 3.6 C-HelQ protein purification from *E. coli.* A. HexaHis-tagged C-HelQ (96 kDa) was sonicated, clarified and resolved by binding to a Hi-Trap Ni-NTA column. The panels show C-HelQ in UV trace and Coomassie stained 10% SDS PAGE analysis. C-HelQ eluted rapidly at 85 kDa from Ni-NTA (*). **B.** C-HelQ eluted with a brown colouration off the Hi-Trap Ni-NTA column allowing for rapid loading onto the Q-sepharose column. **C.** Pooled fractions were dialysed before loading and resolution on a 1 mL Q-sepharose column. C-HelQ containing fractions were pooled and dialysed for storage. Legend = marker (M), wash-through (W), flow-through (F).

The brown coloration may be due to the presence of an Iron-Sulphur cluster, discussed in more detail later in this chapter. The presence of an Fe-S cluster may explain the requirement of a reduced environment required for HelQ and incorrect folding of an Fe-S cluster in C-HelQ may result in high instability observed.

3.7 HelQ protein quantification and verification.

The concentration of purified protein was quantified by using the Bradford's assay and the protein extinction co-efficient calculated using the ExPASy ProtParam Bioinformatics Portal[282] tool (figure 3.1) allowing calculation of protein concentration (see section 2.15). Protein concentration was visually confirmed by SDS PAGE analysis by comparing band intensity of similar calculated concentrations of different proteins.

As expected from previous work,[265] purified HelQ was able to hydrolyse ATP (figure 3.7 A). See section 2.16.5 for described methods. However, N-HelQ, with no Walker A or B domain, was inactive as an ATPase. Interestingly, C-HelQ had reduced ATPase activity, suggesting C-HelQ was unstable in these reactions (figure 3.7 B). The incubation of N-HelQ with C-HelQ did not recover ATPase activity. This may be expected as the addition of N-HelQ and C-HelQ did not result in complete FL-HelQ, there are 35 residues missing that may be important. The fragments, in combination, are unable to mimic FL-HelQ, showing that maximum ATPase activity relies on the whole protein. These malachite green assays were not readily reproducable and therefore they were used to confirm ATPase activity not s realiable means of calculating ATPase activity.

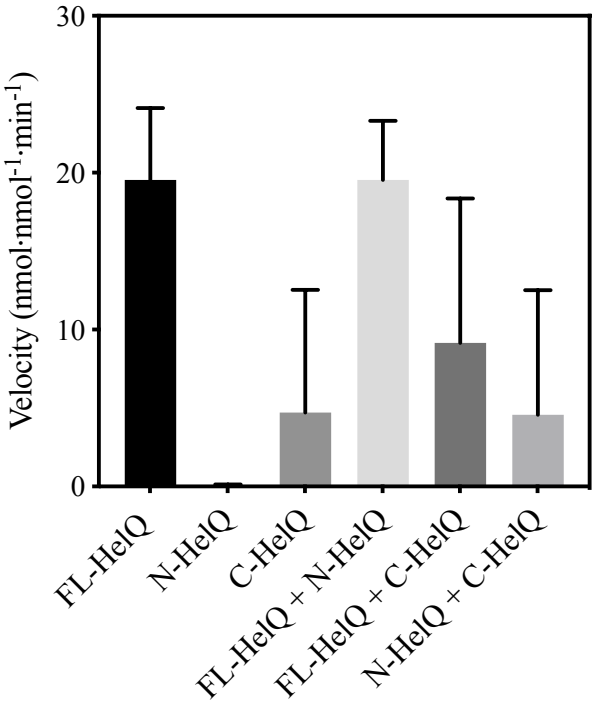


Figure 3.7 ATPase assays confirmed FL-HelQ activity. A. Malachite Green activity assays were used to calculate ATPase activity of FL-, C- and N-HelQ. A standard curve of known concentrations of free phosphate (μM) was calculated as y = 0.0091 x, where y is the OD and x is the concentration of phosphate (nM). **A.** The rate of reaction as nmol ATP hydrolysed per nmol HelQ per minute was calculated for each protein (used at 80 nM) with 25 nM fork 2b DNA in triplicate and shown as standard error from the mean. FL-HelQ showed the largest ATP hydrolysis activity as expected. Protein combinations were also tested including FL-HelQ + N-HelQ, FL-HelQ + C-HelQ and N-HelQ + C-HelQ.

3.8 Biochemical analysis of HelQ *in vitro* shows active HelQ dimers.

The superfamilies and clades of helicase classification includes proteins that are active in various oligomeric states.[200] Monomeric helicases include RecB, part of the *E. coli* RecBCD repair complex, and Pif1.[312,313] Many SF1 helicases are active as dimers, including bacterial Rep, UvrD and PcrA.[314,315] PolQ, a homologue of HelQ, encircles DNA as a tetramer.[248,250] Hexameric helicases, such as Mcm and *E. coli* Rho, form ring structures that unwind DNA by strand exclusion.[316–318] The oligomeric state of a helicase dictates is mechanism of translocation, for example, dimeric helicases can translocate and unwind DNA using a Rolling mechanism or Inchworm variant. The role of oligomerisation in translocation is discussed in

chapter 6. HelQ has been suggested previously to form hexamers eluting at 600 kDa by gel filtration.[265] However, this has not been confirmed using structural models. Work described in the next section assessed the oligomeric state of HelQ upon activation using a range of biochemical techniques in order to determine how HelQ may assemble onto DNA.

3.8.1 Native polyacrylamide gel electrophoresis (PAGE).

Native PAGE was firstly used to assess the likely HelQ oligomeric state in different conditions as described in section 2.16.6. HelQ activation was stimulated by incubation with ATP and magnesium. Analysis was carried out in the presence and absence of DTT because active HelQ requires a reduced environment for stability.[252] FL-HelQ migrated as a number of higher oligomers (figure 3.18); we expect to see monomeric tagged FL-HelQ at 137 kDa. The migrated protein bands were assigned based on the molecular weight markers, estimations were made based on the predicted size of HelQ as monomers and higher oligomers. Native PAGE sizing was used only as an estimate for higher oligomers because of the lack of absolute sizes. FL-HelQ appeared to be unaffected in Native PAGE by addition of activating components and migrated as monomers (146 kDa), dimers (260 kDa), tetramers (600 kDa), hexamers (900 kDa) and octamers (1236 kDa). Multiples of dimeric oligomers were prominent in FL-HelQ populations supporting the literature analysis and previous observations of hexameric HelQ.

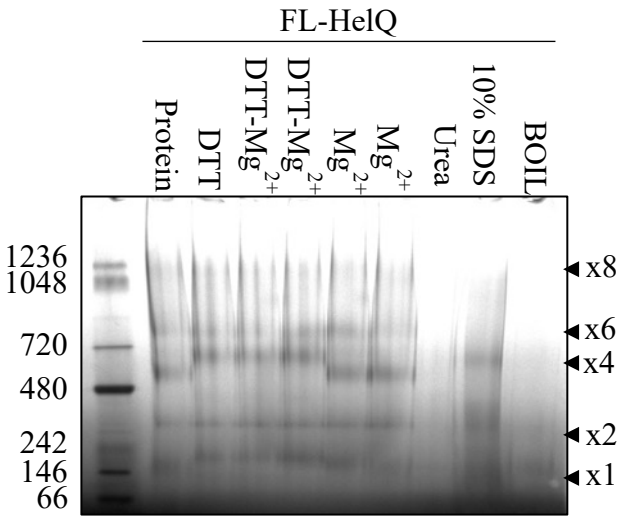


Figure 3.8 Native PAGE shows FL-HelQ forms higher oligomeric structures. Native PAGE of FL-HelQ shows formation of higher oligomers in varying conditions including the addition of 25 mM DTT, 5 mM $MgCl_2$, 5 mM ATP and 25 nM ssDNA (MW12). FL-HelQ formed multiple populations of dimers including dimeric (274 kDa), tetrameric (548 kDa), hexameric (822 kDa) and octameric (1098 kDa). Protein denaturing controls to reduce the protein to monomeric included boil, 10% SDS, proteinase K and urea.

3.8.2 Analytical gel filtration (AGF).

AGF was carried out on FL-HelQ in similar conditions discussed for Native PAGE (section 2.20). The molecular weight of the protein species was determined based on known protein standards. Figure 3.9 shows the standard proteins and their elution volumes used to calculate the molecular weight of each protein species. These standards were used as a guide and not as accurate molecular weight calculations. Elution volumes of unknown HelQ species were converted into molecular weight using the standard curve. Eluted HelQ was tested for DNA binding and unwinding activity to establish the oligomeric state of HelQ that was active.

A.

Protein	Size (Da)	LOG (MW)	UV (mAU)	Elution vol. (mL)	Kav
Ovalbumin	43000	4.633468	295.848	14.89	0.4160256
Conalbumin	75000	4.875061	364.672	14.16	0.3692308
Aldolase	158000	5.198657	322.529	12.62	0.2705128
Ferritin	440000	5.643453	327.616	10.42	0.1294872
Thyroglobulin	669000	5.825426	468.429	9.22	0.0525641

B.

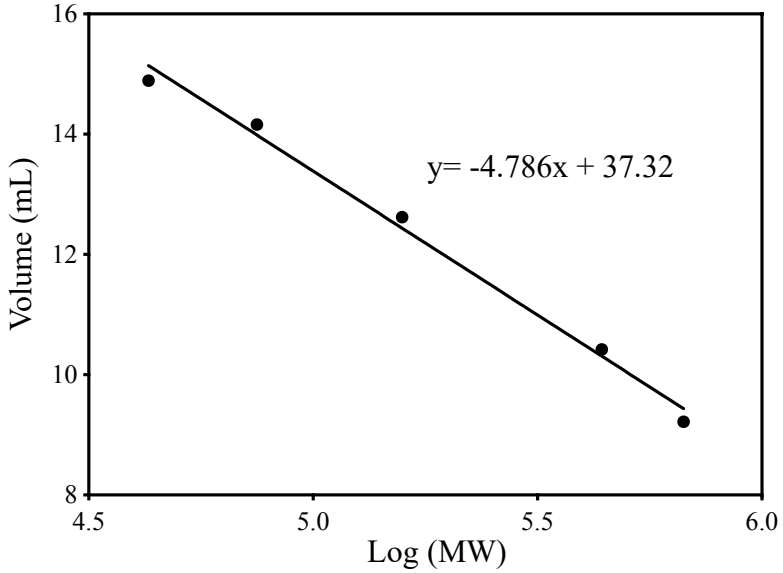


Figure 3.9 Analytical gel filtration (AGF) analysis of standard proteins. A. Table of GE Healthcare high molecular weight (HMW) standards run through the superdex200 gel filtration column in 150 mM NaCl, 20 mM Tris pH 7.5 and 10% glycerol. **B.** Elution volume of each protein was plotted against the log of the known molecular weight (MW) to create a line of best fit (y=-4.786x+37.32). This was used to calculate the molecular weight of proteins of unknown size under different conditions.

FL-HelQ formed predominantly aggregates (959 kDa) in the absence of activating components (figure 3.10 A; curve 1, peak 1). Smaller populations of dimeric (390 kDa; peak 2) and monomeric (37 kDa; peak 3) protein were also evident (predicted at 274 and 137 kDa respectively). Both aggregated (peak 1) and dimeric (peak 2) FL-HelQ bound to, and unwound, DNA in EMSAs (figure 3.10 B). However, monomeric FL-HelQ (peak 3) had no activity (figure 3.10 D). Therefore, this suggests that the oligomerisation of HelQ is essential for its activity on DNA. We assume that addition of ATP-Mg^{2+} required to test helicase activity to elution protein from peak 1 would shift the protein species to that represented in

peak 4; this suggests that aggregated protein is able to disassemble in order to form active oligomers.

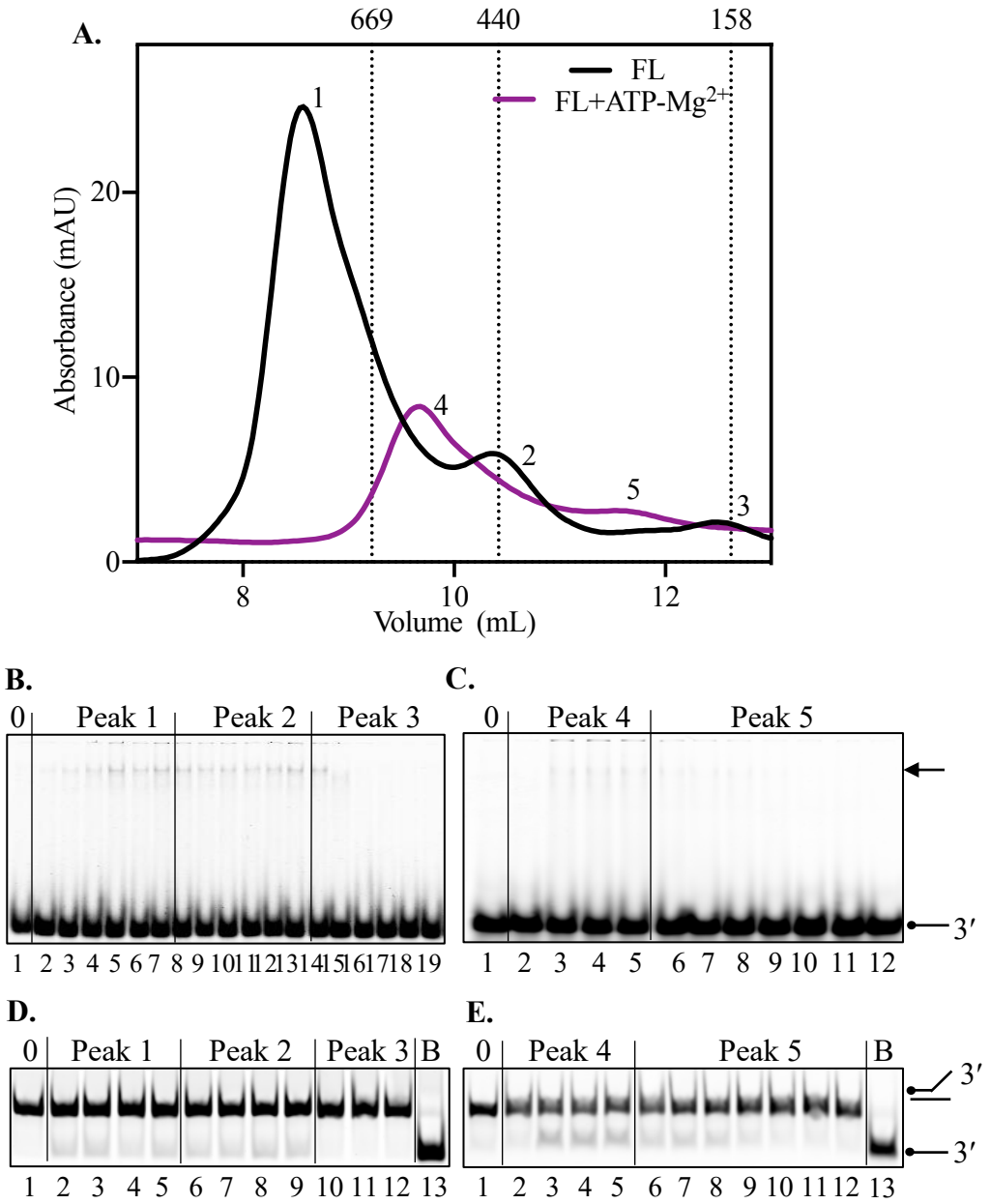


Figure 3.10 AGF analysis to show FL-HelQ forms dimers in the presence of ATP and $MgCl_2$. A. AGF UV trace data of HelQ-only (FL) at 959 kDa (black; curve 1) and HelQ with 5 mM ATP-Mg^{2+} at 570 kDa (purple; curve 2). HelQ formed smaller oligomers consistent with dimers when activated. Markers used are high molecular weight markers detailed in figure 3.9. Additionally, 5% acrylamide TBE EMSA analysis of **B.** HelQ-only (curve 1) and **C.** HelQ with ATP-$MgCl_2$ (curve 2) shows stable complex formation with Cy5-ssDNA. Filled dots indicate 5′ Cy5 labelled ends. 10% acrylamide TBE unwinding analysis of fork 2b (25 nM) by **D.** HelQ-only run (black) showed limited activity in peak 2 and **E.** HelQ with ATP-Mg^{2+} showed activity in peaks 4 and 5 (purple). Activated HelQ was able to unwind fork 2b.

The activation of FL-HelQ, by addition of ATP and magnesium, shifted FL-HelQ from aggregated to tetrameric protein at 570 kDa (curve 2, peak 4), predicted at 548 kDa. Activating components appear to impact HelQ oligomerisation by shifting it to more dimeric patterns previously discussed. A small population of monomeric HelQ is visible at 200 kDa (peak 5). Data presented here further supports dimeric forms of HelQ suggesting previously documented hexameric HelQ was aggregated protein in an inactive state.[265] Tetrameric FL-HelQ (peak 4) bound to, and unwound, DNA in EMSAs confirming it as an active oligomeric state (figure 3.10 C and E).

The addition of ssDNA (50-bp) to activated FL-HelQ further shifted the prominent species from 570 kDa (curve 2, peak 4) to 407 kDa (curve 3, peak 6) (figure 3.11 A). This is consistent with dimeric FL-HelQ (predicted at 274 kDa). The addition of ssDNA reduced protein aggregation, therefore, further stabilising the ATP-Mg^{2+} activated FL-HelQ into a dimeric form. EMSA gels run of the fractions eluted from the column confirmed complex formation between activated FL-HelQ and ssDNA (peak 6 and 7). The presence of smearing and aggregation in the wells (figure 3.11 B, peak 6) may be caused by complex instability or the presence of activating co-factors (ATP-Mg^{2+}). This would result in the iterative translocation of HelQ along ssDNA and subsequent complex disassembly. The presence of aggregated protein-DNA complexes is indicative of activated oligomeric HelQ.

The presence of ssDNA, in the absence of activating co-factors (ATP-Mg^{2+}), had no impact on the oligomerisation of FL-HelQ (figure 3.11 C, curve 1 and 4). This is apparent by similar peaks between curves 1 and 4. EMSA gels run of the fractions eluted off the column confirmed peak 8 (ssDNA + FL-HelQ) as aggregated protein, as seen for peak 2 previously (FL-HelQ only) (figure 3.11 D).

Additionally, monomeric and dimeric FL-HelQ (peaks 9 and 10) were observed to a lesser extent and formed complexes with ssDNA (predicted at 137 and 274 kDa respectively). Therefore, the activation and oligomerisation of FL-HelQ relies on the presence of ATP-Mg^{2+} and ssDNA alone is not sufficient.

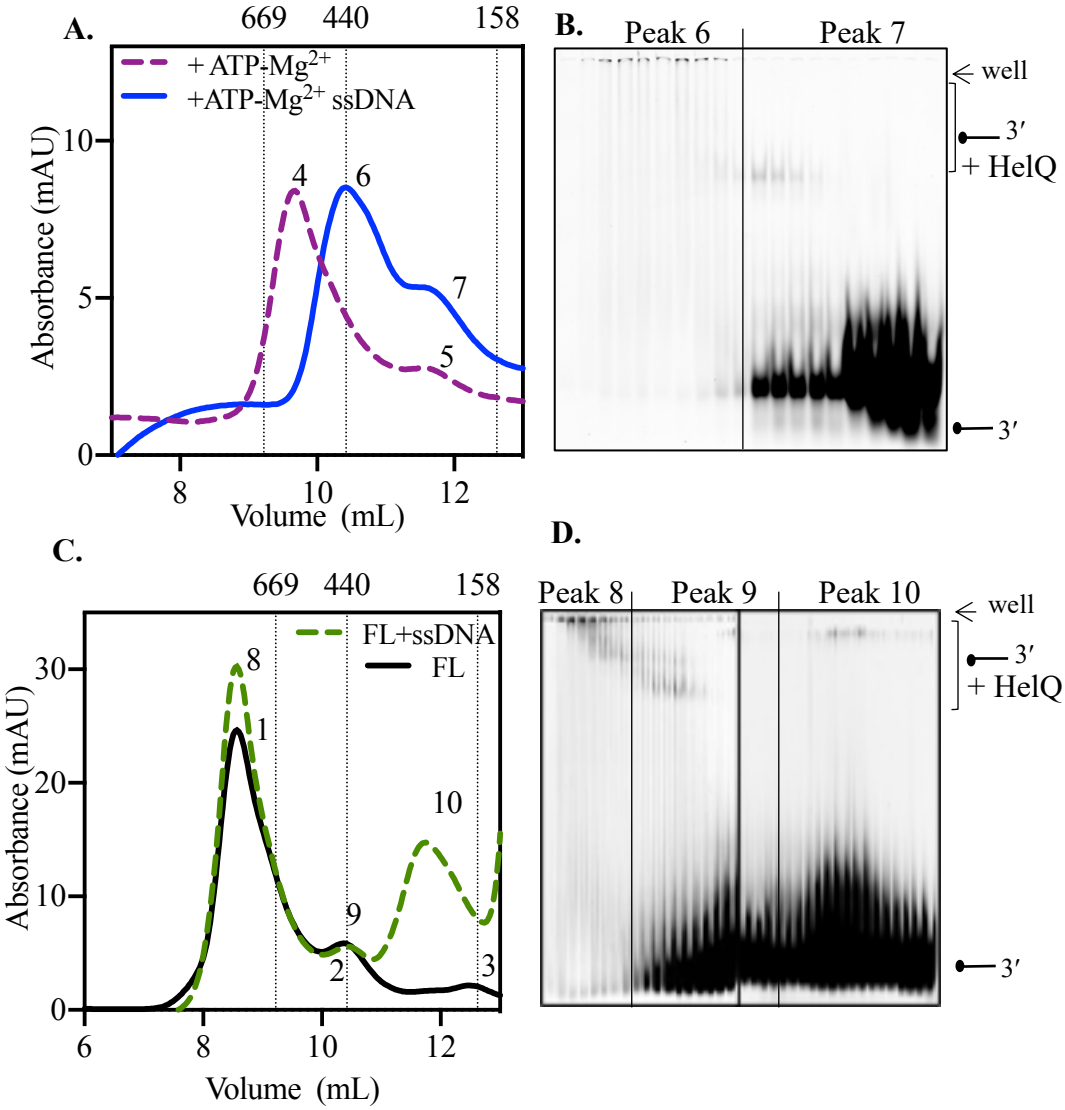


Figure 3.11 AGF analysis shows FL-HelQ forms active dimers when complexed with ssDNA. A. AGF UV trace data of FL-HelQ with 5 mM ATP-Mg^{2+} eluted at 570 kDa only (curve 2) and with 25 nM ssDNA at 407 kDa (curve 3). Markers used are high molecular weight markers (figure 3.9). **B.** 5% acrylamide TBE EMSAs of FL-HelQ with ATP- Mg^{2+} and ssDNA (curve 3) shows stable complex formation. **C.** AGF UV trace data of HelQ-only eluted at 570 kDa (curve 1) and HelQ with 25 nM ssDNA only eluted at 407 kDa (curve 4). **D.** 5% acrylamide TBE EMSAs of HelQ with ssDNA (curve 4) shows complex formation. Filled dots indicate 5′ Cy5 labelled ends.

AGF of ATPase inactive FL-HelQ^{D463A} with loss of the magnesium coordinating residue, was carried out as a control (figure 3.12). Addition of ATP-Mg^{2+} resulted in an oligomeric shift from aggregated (818 kDa, peak 1) to tetrameric protein (548 kDa, peak 2). The protein eluted over a large volume as a shallow peak suggesting the formation of unstable oligomers. This confirmed that ATPase activity is required for oligomerisation.

To summarise, HelQ oligomerises into an active state in the presence of ATP-Mg^{2+} to form predominantly dimers and tetramers (predicted 274 and 548 kDa respectively). The presence of ssDNA (50-bp) stabilises active HelQ potentially allowing for multiple HelQ dimers to bind.

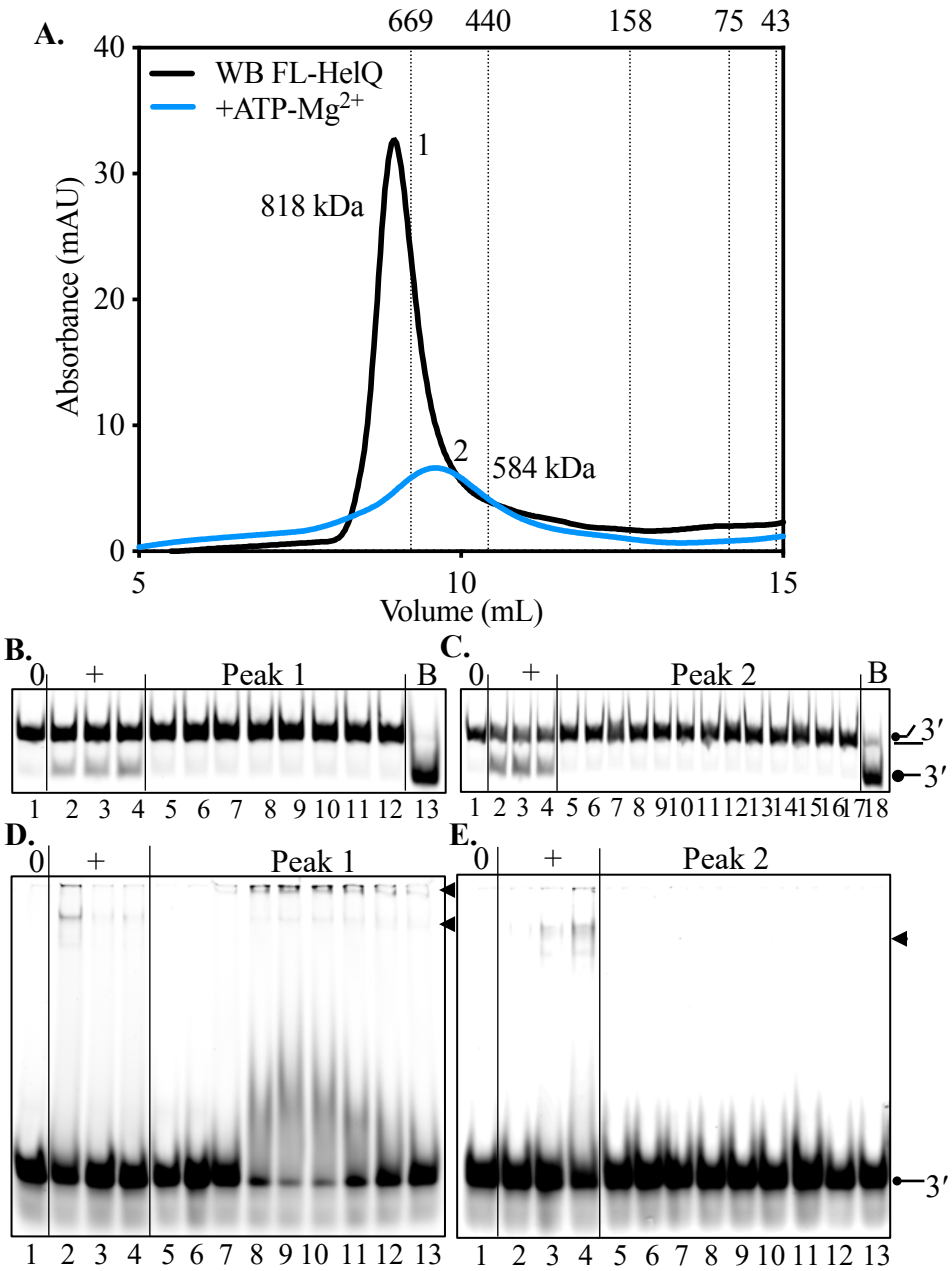


Figure 3.12 AGF analysis shows ATPase inactive FL-HelQ^{D463A} also forms dimers. A. AGF UV trace data of FL-HelQ^{D463A} only eluted at 818 kDa (peak 1) and with 5 mM ATP-Mg^{2+} at 584 kDa (peak 2). As seen for wild type HelQ, HelQ^{D463A} forms smaller oligomers when activated. Markers shown are high molecular weight standards (figure 3.9). Helicase unwinding activity of fork 2b (25 nM) was not observed by **B.** FL-HelQ^{D463A} or **C.** with ATP-Mg^{2+} in acrylamide TBE analysis. B represents a boil control to show dissociated substrate as indicated to right of gel. Filled dots indicate 5′ Cy5 labelled ends. **D.** EMSAs show protein-DNA complex formation (peak 1) but only weak complexes were detected when upon activation with ATP-Mg^{2+} (peak 2, **E.**).

3.8.4 Size exclusion chromatography with multi-angle light scattering (SEC MALS).

SEC MALS directly measures the mass of a sample independently of its elution time (section 2.21.3). SEC MALS removes the assumption that all molecules behave and elute similarly to the proteins commonly used as standards in AGF. The method uses precise alignment of the laser beam, laser polarization and detectors to give a measure of the particle size.[319] The appropriate refractive index increment (dn/dc) for SEC MALS was calculated by DLS (section 2.21.1). The dn/dc, a measure of the exact protein concentration, is unique to a sample and the solvent. No solvent interference was detected by DLS at the used protein concentration, therefore, the common dn/dc value of 0.185 was used.[320]

Experimental conditions were maintained for both SEC MALS and AGF. Addition of activating components (ATP-Mg^{2+}) shifted FL-HelQ from aggregated protein (598 kDa) to dimeric protein (265.2 kDa) (figure 3.13 A and B). Additionally, in the presence of ssDNA, FL-HelQ eluted predominantly dimeric (240.3 kDa) and monomeric (131.8 kDa) (figure 3.13 C). This further confirms FL-HelQ is active in dimeric form (predicted at 274 kDa).

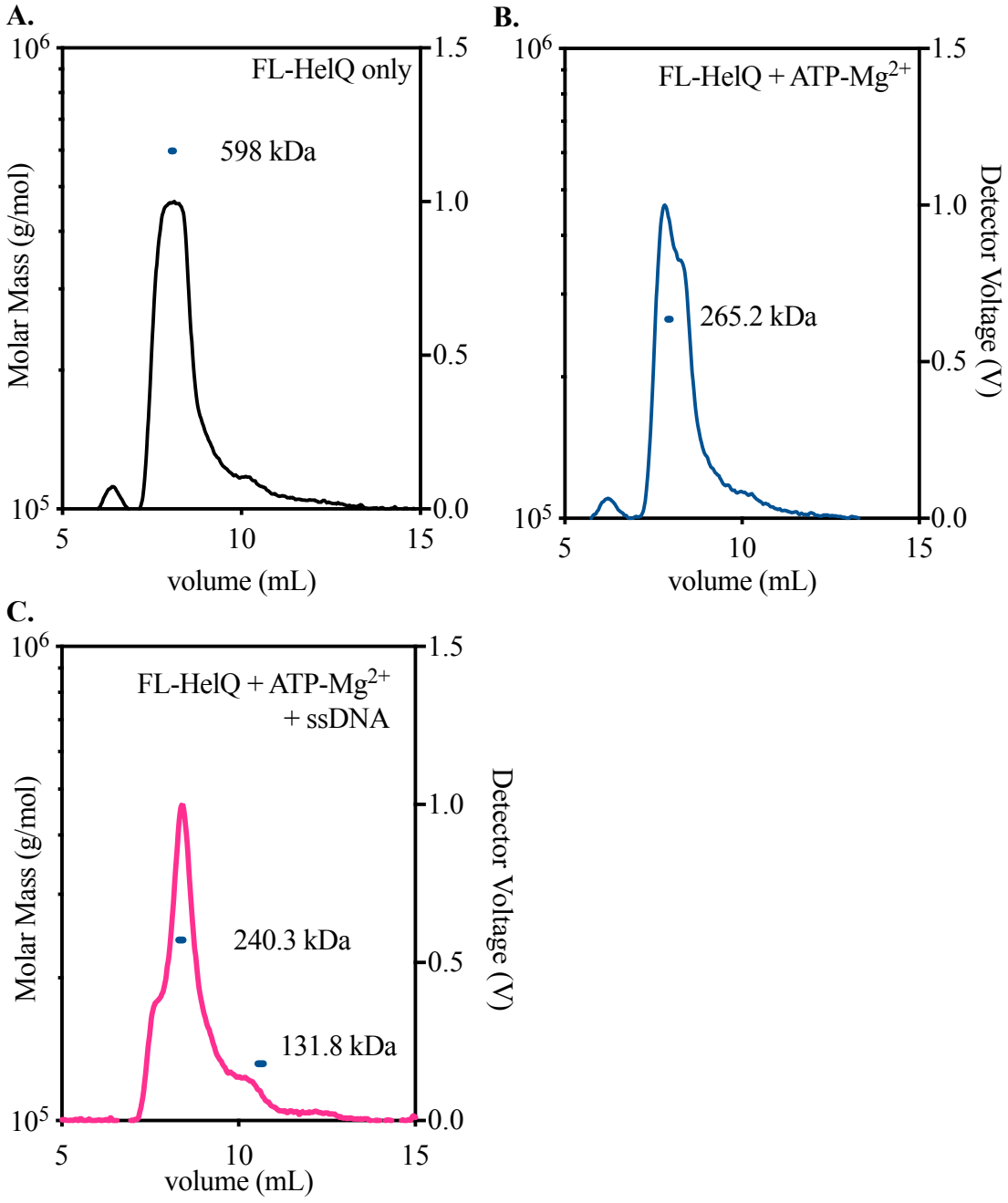


Figure 3.13 Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC MALS) of FL-HelQ. Size exclusion chromatography was carried out using a superdex 200 gel filtration column. Multi-angle light scattering was measured and analysed using ASTRA 6 and WYATT technology. Graphs show MALS detector voltage (V) plotted against elution volume of each protein species. **A.** FL-HelQ only eluted at 598 kDa, **B.** with 5 mM ATP-Mg^{2+} eluted at 265.2 kDa and **C.** with 5 mM ATP -Mg^{2+} and 25 nM ssDNA (MW12) at 240.3 kDa and 131.8 kDa indicating dimeric and monomeric protein.

3.9 FL-HelQ is predicted to form multiple dimeric oligomers *in silico.*

Structural modelling of proteins can provide an insight into their preferred oligomeric state and determine details about how proteins interact with DNA.

Structural models of HelQ rely on bioinformatic predictions because a crystal structure of FL-HelQ has not yet been solved. The algorithms used were those in the PHYRE2, Raptor and Pymol web-serves (section 2.22.3).[304,305,306] Protein modelling of FL-HelQ is difficult due to the presence of N-HelQ discussed in section 3.10.

Predictive modelling was used to assess and visualise the experimental findings observed of oligomeric HelQ. Dimeric HelQ was considered *in silico* because HelQ was experimentally shown to preferentially form dimers and further dimeric patterns. Additionally, homology to PolQ which forms a tetrameric structure discussed in chapter 1, implicates a tetrameric form of HelQ (figure 3.14 A-D). PolQ monomers each interact with the other three monomers in the tetrameric structure, while these residues were not found conserved in HelQ, other similarities made it an interesting avenue to pursue. This, along with the documented hexameric form of HelQ implicated higher oligomeric structures.[265] While this was not supported using SEC MALS and AGF presented here, the dimeric pattern of hexamers was considered important for HelQ modelling *in silico.*

Modelling of HelQ using the Symmdock[308] tool for PHYRE2[304] predictions of HelQ structure showed a strong probability of HelQ in dimeric and tetrameric forms (figure 3.14 C and D). This supported the results observed experimentally. The Symmdock tool predicts oligomerisation based on predicted structure, assessing interacting surfaces of monomers and rotational symmetry. The role of N-HelQ in FL-HelQ function, specifically oligomerisation, was next assessed in this work.

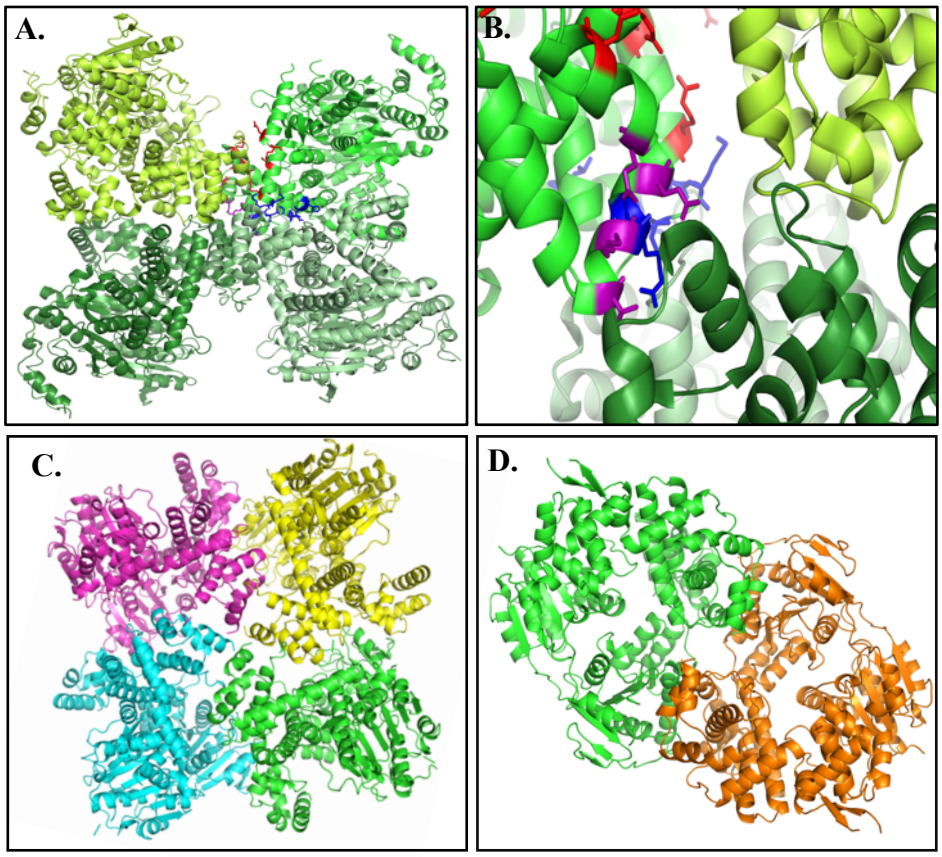


Figure 3.14 Symmdock predictions of the higher oligomeric state of FL-HelQ. A. PHYRE2 models of the solved crystal structure of the tetrameric conformation of PolQ helicase. Each monomer, in a different shade of green, interacts with the other three monomers. **B.** Residues involved in monomer binding and potentially oligomerization is highlighted in red, blue and purple and are magnified in upper right panel. Sequence homology between PolQ and HelQ suggests a similar mechanism of action may be adopted by HelQ. PHYRE2 predicted protein structures were used to model for higher oligomerisation of HelQ using the Symmdock tool. A high probability for oligomerisation of HelQ was identified as **C.** tetramer and **D.** dimer.

3.10 Bioinformatic analysis cannot structurally define N-HelQ.

Investigation of N-HelQ in isolation was useful for determining its function. N-HelQ is a protein that has no previously reported sequence or structural homology to any other protein, a so-called 'ORFan' protein[321] until studied here (figure 3.15 A). Protein sequence alignments between HelQ and archaeal Hel308 identified that similarities start at the first RecA-like domain (figure 3.15 B). This suggests N-HelQ has a specific role in metazoan HelQ orthologues.

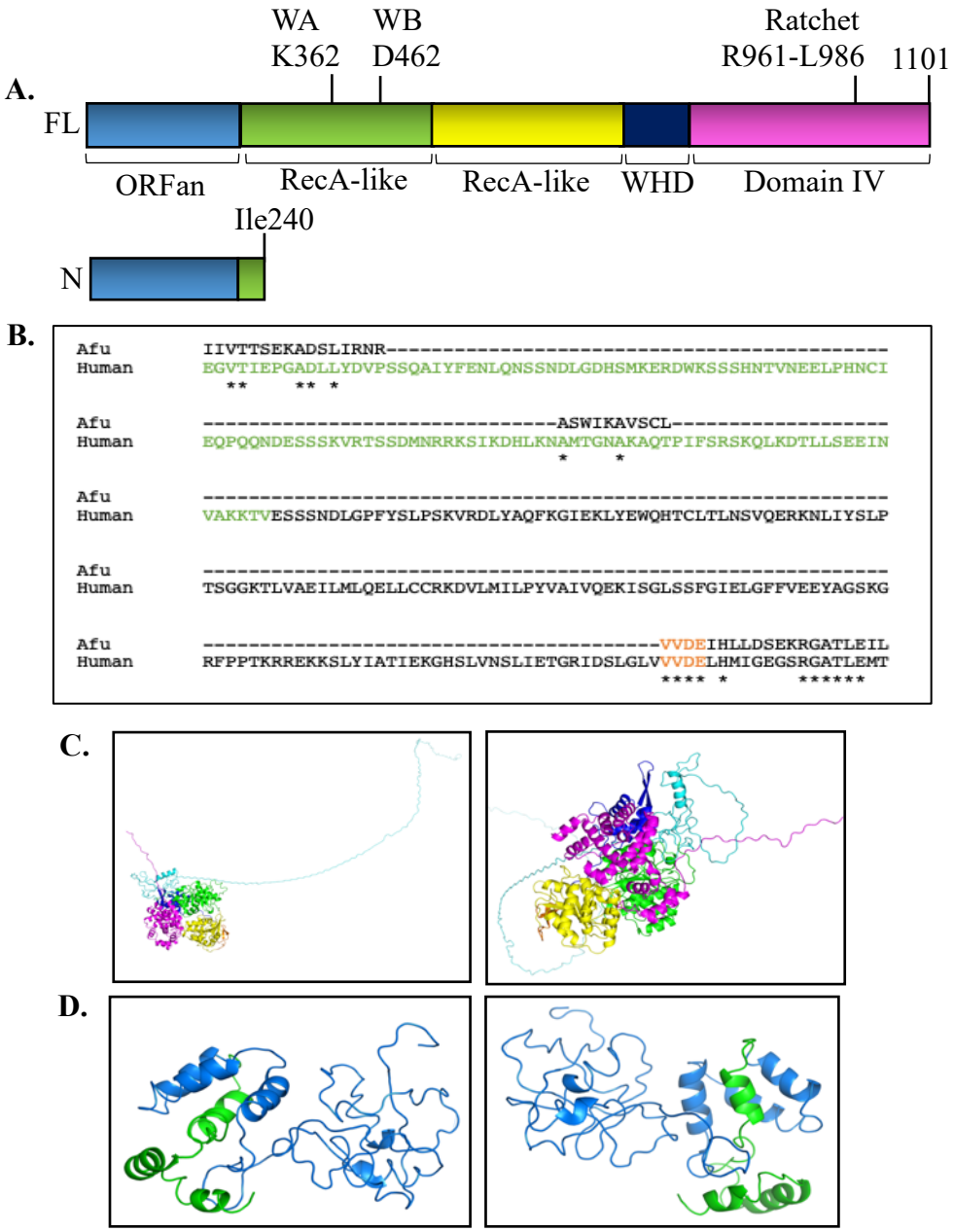


Figure 3.15 N-HelQ is a region of unknown function and no predicted structure can be defined. A. Cartoon representation of FL-HelQ and N-HelQ. A 240 residue N-terminal 'ORFan' region (blue), RecA-like domains (green and yellow) with Walker A (WA) and Walker B (WB) domains indicated, a winged helix domain (WHD; navy), and a helicase ratchet within Domain IV (pink). N-HelQ has a slight overlap into the RecA-like domain 1. **B.** The N-terminal region of human HelQ (green) is not found in archaeal Hel308 (Afu) in CLUSTAL sequence alignments. The sequence similarity starts at the Walker domains (orange). **B.** PHYRE2 structure prediction models of **C.** FL-HelQ and **D.** N-HelQ in two orientations. Structure predictions of N-HelQ were not able to be assigned when considered in the context of FL-HelQ suggesting disorder. N-HelQ sequence alone showed some helical conformation and some unstructured un-defined sequence. The colour reflects schematic in **A.**

Bioinformatic analysis of N-HelQ gave inconsistent and unreliable models when using the PHYRE2[304] protein structure prediction tool. Modelling of FL-HelQ in the same way was also unable to predict a meaningful structure for N-HelQ (figure 3.15 C-D). Therefore, this suggests a lack of structure and order within N-HelQ. Another bioinformatic tool, IUPred, was used to predict intrinsically disordered protein regions (IDPRs) within N-HelQ. Two regions within N-HelQ (residues 20-30 and 50-90) were predicted to be IDPRs, with a high probability (between 0.8-1.0) (figure 3.16). The probability was based on the distribution of residues that promote disorder (Arg, Pro, Glu, Gly, Ser, Ala and Lys).[322]

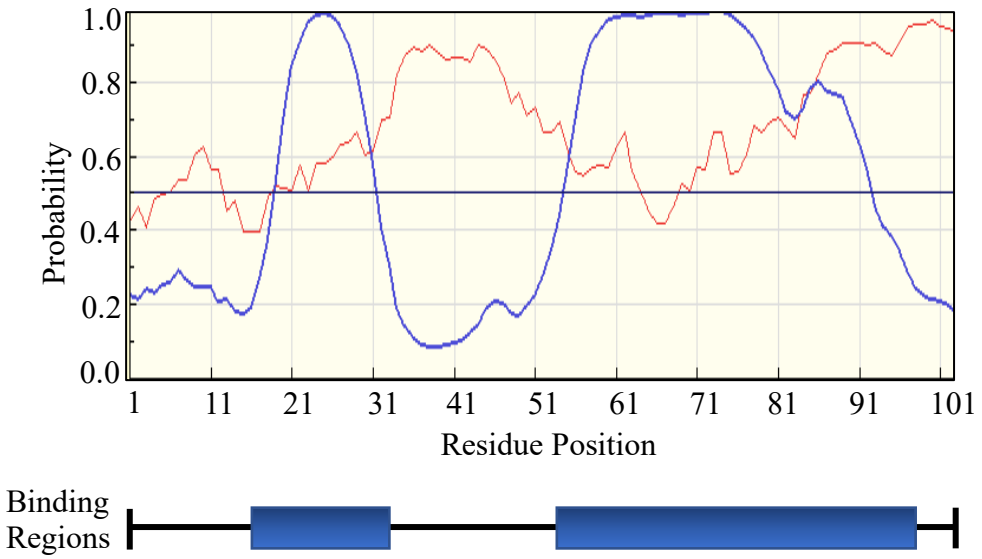


Figure 3.16 IUPred predicts regions of protein disorder in N-HelQ. IUPred calculator measured the probability of disorder of the residues to predict regions of protein disorder in N-HelQ (red line). Disorder was predicted with a probability between 0.8-1 between residues 31-55 and 75-101. We can therefore predict regions outside of these domains have elements of structure. ANCHOR also predicted protein-protein interacting domains within N-HelQ (upper panel; blue line). Two regions of strong probability are predicted outside of the disorder regions (lower panel). Protein interactions are considered important for the function of intrinsically disordered proteins.

IDPRs are characterised by extensive structural plasticity and spatio-temporal heterogeneity.[322] This means that regions of intrinsic disorder within proteins have ready conformational flexibility in response to their aqueous environment and surrounding proteins. These proteins, or protein regions, are found across all domains of life. A literature study in 2019 identified approximately 1,150 non-redundant proteins within the validated intrinsic disorder proteins (IDPs) database.[322] IDPs are more commonly encountered in eukaryotic compared to their prokaryotic homologues. Consequently, increased protein disorder correlates with more complex organisms and protein function.[323] These characterisations make them ideal for a range of unique biological processes.

The majority of IDPs are not completely disordered. They contain regions of structure, known as a 'pre-structured motifs'.[322,324] These motifs act as targets for substrate interactions including protein or DNA. IDPs also behave as the 'molecular glue' in complex assembly. These types of IDPs are known as 'morphers', adopting different structures upon interaction. The bioinformatic tool, ANCHOR, was used to predict regions of protein-protein interactions within the IDPRs of N-HelQ (figure 3.16). There were regions of predicted disorder and protein-protein interaction interfaces in N-HelQ sequence, suggesting a role for N-HelQ as a 'morpher'.

3.11 Experimental analysis confirmed N-HelQ as an IDP.

3.11.1 Circular Dichroism (CD) and Nuclear Magnetic Resonance (NMR).

CD, an absorption spectroscopy technique, uses circular polarised light to determine the chirality of molecules (section 2.21.2). The shape of the UV trace of known protein structures is used to produce a secondary structure read out for unknown samples. A change in protein conformation, for example during protein

activation, can change the outcome of the CD spectrum. CD can provide detail on how protein structure changes under different conditions.[295,325]

CD spectrum analysis confirmed protein disorder of hexaHis-SUMO-tagged N-HelQ (figure 3.17). Standard spectra from Greenfield *et al.* (2006) was used to determine the secondary structure, and lack of, in N-HelQ.[296] Structured regions detected at a wavelength of 200 nm in figure 3.17 may be consistent with the helical structure of the SUMO-tag. A SUMO-tag has a globular structure made up of an α-helix and a β-sheet.[326] Alternatively, as shown from IUPred predictions (figure 3.16), regions of order within the disorder are still present within N-HelQ which may impact this. Further CD of individual SUMO-tags would be required to confirm this.

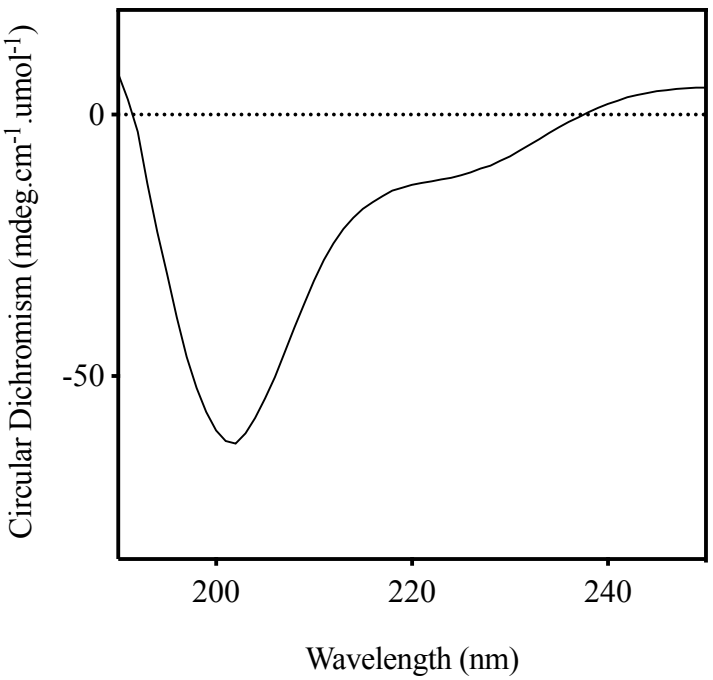


Figure 3.17 Circular dichroism (CD) confirmed intrinsic disorder within N-HelQ. CD absorption spectrum of hexaHis-SUMO-tagged N-HelQ at 1 mg/mL in 5 mM Tris pH 8. Results were normalized to buffer only control. Trace was compared to known protein structures and suggests protein disorder. Structure detected at 220 nm may be consistent with the SUMO-tag. SUMO has a globular structure consisting of an α-helix and a β-sheet.[325] Comparison to the structure analysis from Greenfield *et al.*[295] of CD spectra of known structures, this suggests a partial helical structure that may be due to the SUMO-tag. Alternatively, it may indicate regions of order that are predicted in N-HelQ and discussed later in this work. The presence of disorder confirms bioinformatic analysis observed.

NMR is a spectroscopy technique that observes local magnetic fields around atomic nuclei. This can be interpreted to give a molecular structure or confirm a lack of structure. Due to the dynamic nature of IDPs, NMR is a suitable technique to determine structural dynamics of these proteins. IDPs have a lack of dispersion of proton resonances required for spectra, however, the flexibility of these proteins cause slower relaxation rates allowing spectra to be acquired of even larger proteins.[327] Therefore, we used NMR to confirm the absence of order by looking for characteristic spectra patterns of highly flexible proteins. Preliminary NMR confirmed a lack of structure in N-HelQ. Clusters of residues plotted in a spectrum as shown in figure 3.18 dictate the type of structure a protein adopts. The distribution reported here is suggestive of regions of no structure (see insert in figure 3.18). Additionally, structure is still detected within N-HelQ which is expected as only two regions of disorder were predicted in IUPred (figure 3.16). This work requires exploration in order to further pinpoint the region of disorder.

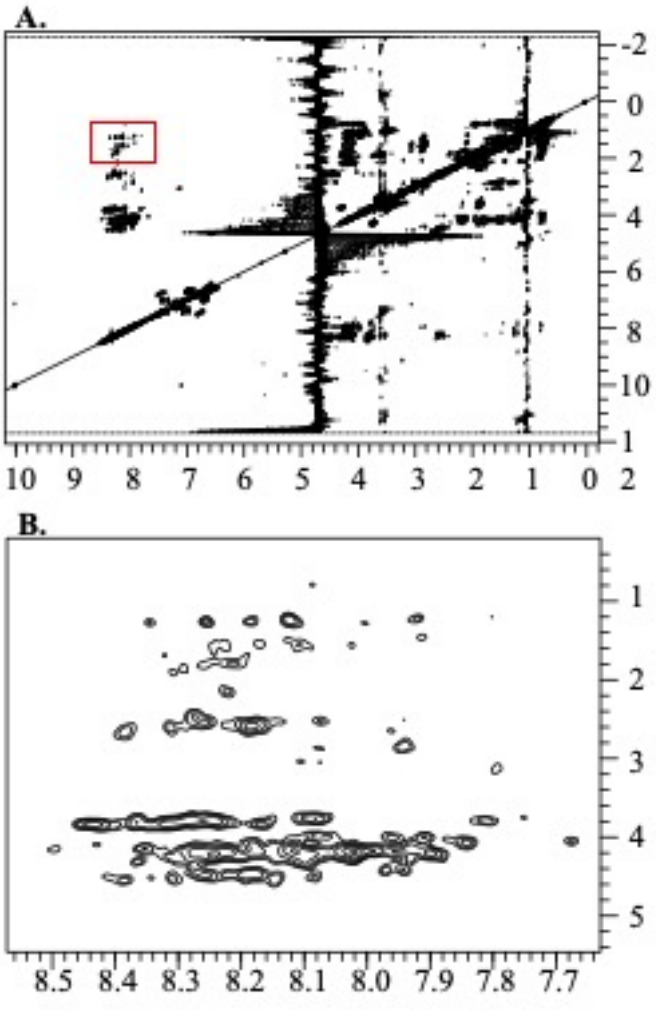


Figure 3.18 Preliminary Nuclear Magnetic Resonance (NMR) spectra for N-HelQ indicates disorder. Toscy NMR spectra for hexaHis-SUMO-N-HelQ. **A.** Full spectra produced at 1 mg/mL HelQ run for 24 hours. Region in red identifies areas of protein disorder. **B.** Magnified red box to identify region of disorder within N-HelQ. Structure observed from the spectra is likely from the SUMO-tag on the N-terminal of HelQ. Initial findings support protein disorder.

3.11.2 Preliminary crystallography of N-HelQ.

Crystal trials of N-HelQ were run in collaboration with Denis Ptchelkine at RCaH (section 2.21.6). N-HelQ did not crystallise in routinely tested crystallisation conditions. This may add to confirm the presence of intrinsic disorder within N-HelQ which would interfere with protein crystallisation. The presence of an activating component or protein partner may be required (chapter 5). Biophysical

analysis of N-HelQ confirmed the bioinformatic predictions of intrinsic disorder implicating N-HelQ in a range of biological roles (chapter 5).

Variants of untagged N-HelQ were made that included truncations purified for crystallography and biochemical analysis as described in section 2.14.2 (figure 3.19 A-D). It was determined that protein variations with different regions of structure may be required to stimulate protein crystallisation. These protein variants included N-HelQ 1 (N1), a 33.5 kDa protein terminating prior to the core helicase, N-HelQ 2 (N2), a 63 kDa protein terminating after the first RecA-like domain and N-HelQ 3 (N3) containing both RecA-like domains (80.5 kDa). The DNA encoding the protein fragments were cloned into pETM40 with a TEV-cleavable MBP-tag and pET28b-SF with a TEV cleavable StrepII-tag. PHRYE2 was used to predict the structure of these fragments (figure 3.19 E-G). Biochemical analysis of these proteins is described in later chapters.

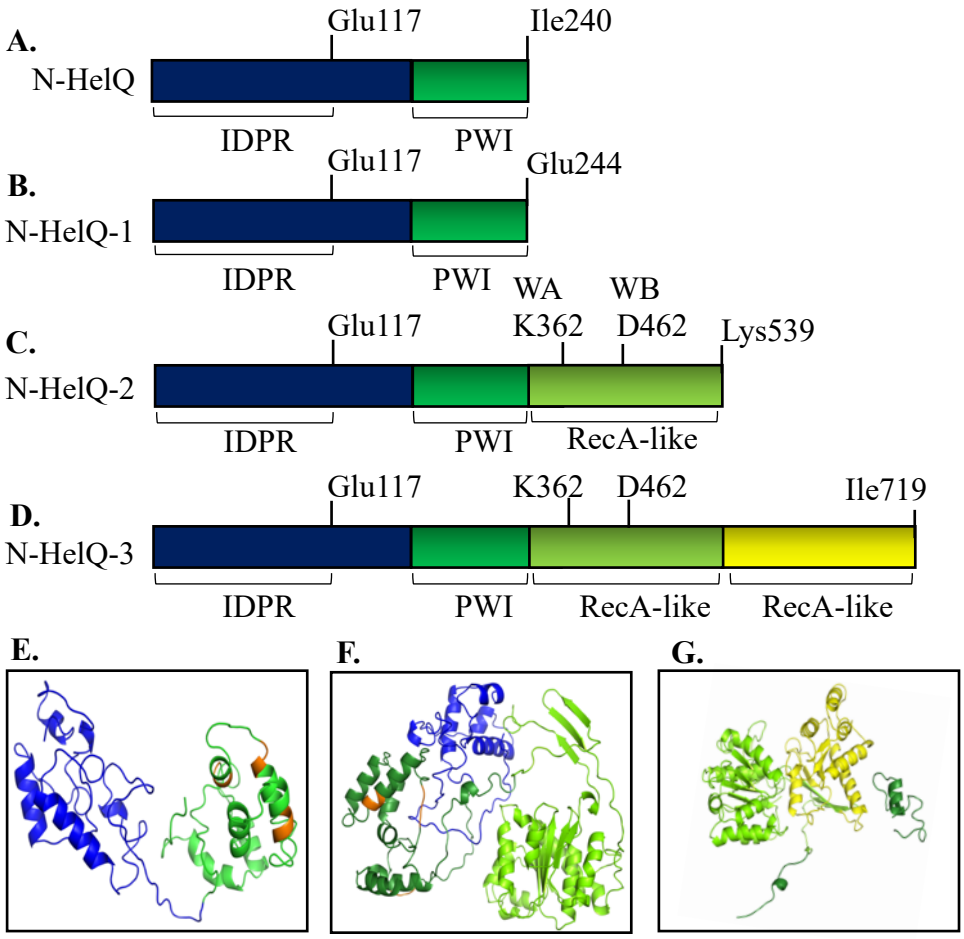


Figure 3.19 N-HelQ protein variations used in this work. DNA encoding variants in N-HelQ sequence length were cloned with TEV-cleavable StrepII- and MBP-tags into pET28b and pETM40. Cartoons are shown without protein affinity tags. **A.** WT-N-HelQ protein (240 amino acids) with region of intrinsic disorder terminating at Glu-117 (navy) and PWI-like domain (green). **B.** TEV-cleavable PWI-like domain only N-HelQ (N-HelQ-1) terminating at residue 244. **C.** PWI and RecA like domain 1 N-HelQ (N-HelQ-2) terminated at residue 539 and **D.** PWI and RecA like domains 1 and 2 N-HelQ (N3) terminated at residue 719. Predicted PHYRE2 structures of **E.** N1, **F.** N2 and **G.** N3. Pymol models are coloured corresponding to schematics. N3 PHYRE2 prediction was unable to assign the IDPR region (navy) within the protein.

3.12 Symmdock predicts higher oligomers of N-HelQ.

N-HelQ is predicted to contain IDPRs, potentially involving protein-protein interacting domains. IDPs typically rely on activation by a binding partner. It is hypothesised here that N-HelQ may require such a partner, potentially itself, which would result in oligomerisation. FL-HelQ is shown to form higher oligomers, therefore a role of N-HelQ in this was explored. Bioinformatic predictions of N-

HelQ oligomerisation, as for FL-HelQ, were made using Symmdock (section 2.22.4).[308,328] As described previously, HelQ was considered in multiple dimeric forms and was both predicted and shown to oligomerise in a dimeric pattern (figure 3.14). However, modelling of N-HelQ did not predict similar instances of higher oligomers and did not easily form dimers or tetramers (figure 3.20 A-B). A lower probability calculated may be because predicted structures produced in Symmdock cannot account for the 'morphing' phenotype that N-HelQ may display as an IDPR. Whilst bioinformatics does not support N-HelQ as having an oligomerisation role, this cannot be excluded. Modelling of N-HelQ has been shown previously to be problematic (figure 3.15) and therefore it is unsurprising other predictive tools would be equally unreliable. Additionally, literature suggests IDPRs likely rely on alternative binding partners for activation, or oligomerisation, which may be the case here. Therefore, oligomerisation of N-HelQ was explored experimentally prior to analysis of alternative roles.

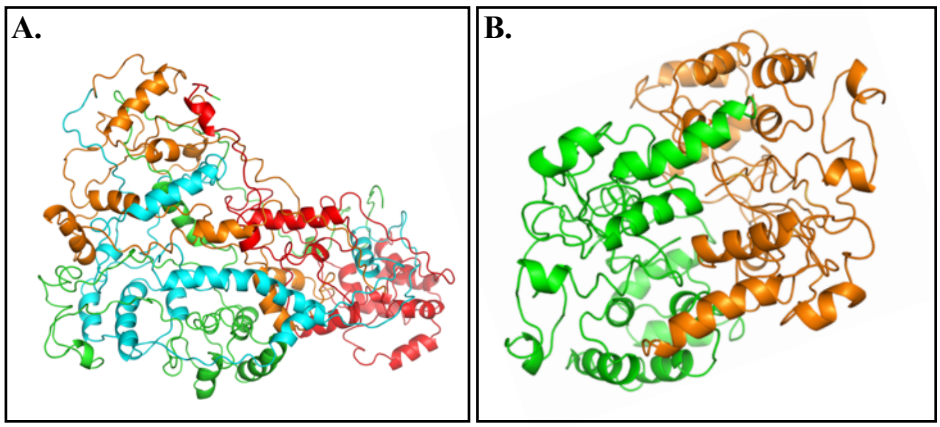
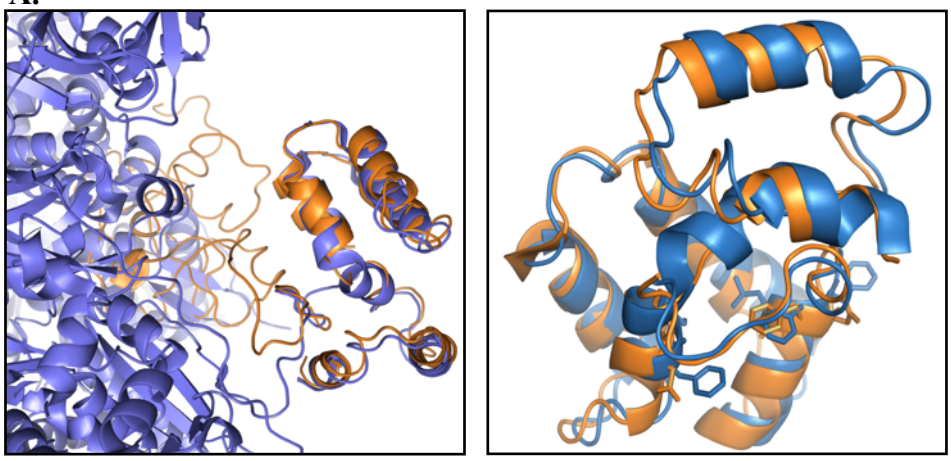


Figure 3.20 Symmdock modelling to predict oligomerization of N-HelQ. The Symmdock predictive tool was used to assess the probability of a role of N-HelQ in the oligomerisation of FL-HelQ as both **A.** tetramers and **B.** dimers. There were fewer possible iterations of N-HelQ forming either higher oligomer when compared to FL-HelQ. The best results are shown here. Predictive modelling may be limited because of the intrinsic disorder of N-HelQ.

Further bioinformatic analysis on N-HelQ was carried out by Chris Cooper at Huddersfield University. Based on predicted structure, a portion of N-HelQ superimposes onto a PWI-like domain of the RNA splicing helicase Brr2 (figure

3.21 A). A PWI domain has a canonical structure with a four α-helical bundle and conserved proline, tryptophan and isoleucine residues required for this structure to fold. N-HelQ, like Brr2, has a PWI-like domain with the same α-helical bundle but without the conserved residues. Therefore, the PWI-like domain of N-HelQ superimposes onto the helical structure in Brr2 with very little sequence homology (figure 3.21 B). These structured regions may account for the order observed among the disorder in CD and NMR analysis.

A.



B.

```
N-HelQ      -------VQPLLLSDSPECLVLGGGDTNPDLLRHMPTD-RG-VGDQPNDSEVDMFGDYDS        112
Brr2        YEQILQWVTEVLGNDIPHDLIIGTADIFIRQLKENEENEDGNIEERKEKIQHELGINIDS        179
                  *  :* .* *. *::* .*     *:.   :  * : :: :. : ::  : **

N-HelQ      FTENSFIAQVD-------------------------------------------------        123
Brr2        LKFNELVKLMKNITDYETHPDNSNKQAVAILADDEKSDEEEVTEMSNNANVLGGEINDNE        239
            :. *.::  :.

N-HelQ      --DLEQKYMQLPEHKKHATDFATENLCSESIK------NKLSITTIGNLTELQTDKHTEN        175
Brr2        DDDEEYDYNDVEVNSKKKNKRALPNIENDIIKLSDSKTSNIESVPIYSIDEFFLQRKLRS        299
              * * .* ::  :.*: .. *  *: .: **      .::. . * .: *:  ::: ..

N-HelQ      QSGYEGVTIEP-------------------------------------------GADL---        190
Brr2        ELGYKDTSVIQDLSEKILNDIETLEHNPVALEQKLVDLLKFENISLAEFILKNRSTIFWG        359
            :  **:..::                                                : :
```

Figure 3.21 Structural homology between N-HelQ and the Brr2 PWI-like domain. Bioinformatic analysis carried out by Chris Cooper (Huddersfield University) identified structural homology between the C-terminal region of N-HelQ and the PWI-like domain of the Brr2 slicing RNA helicase. The PWI domain of Brr2 is a small external domain involved in protein-protein interactions. **A**. PHYRE2 model overlay of predicted N-HelQ (orange) and Brr2 (purple) whole protein. Panel 2, PWI-like domain only (blue) structurally super-imposes onto N-HelQ (orange). **B.** The PWI-like domain shares structural but not sequence homology to the canonical PWI domain and therefore there is little homology between N-HelQ and Brr2. Residues of interest important for Brr2 PWI-like activity are highlighted in red with the HelQ alignment.

The PWI-like domain is conserved between yeast and humans. The PWI-like domain, found in Brr2, cannot bind DNA like the canonical PWI domain.[329] However, it transiently interacts with proteins of the spliceosome, including the Prp19 complex. This suggests that the N-terminal region of Brr2 is important for protein-protein interactions to assemble the spliceosome which is reinforced by the PWI-like domain.[329] To summarise, evidence of a PWI-like domain further supports a role for N-HelQ in protein-protein interactions to initiate function. A role of N-HelQ in protein-protein interactions was therefore tested in this work, both self-interacting (see oligomerisation in next section) and with other protein partners (see chapter 5.)

Despite a lack of sequence conservation between the PWI-like domains of N-HelQ and Brr2, a conserved aspartate residue (Asp-142), within the first helical turn, is necessary for Brr2 PWI-like function. The aspartate orientates away from a hydrophobic core that centres around a phenylalanine residue (Phe-143). These residues are important for function and therefore were targeted for mutagenesis in N-HelQ (section 2.14.3); the mutant hitherto is referred to as N-HelQ$^{D142A/F143A}$ (figure 3.22 A-B). This mutant variant was used to assess the role of the PWI-like domain in protein-protein interactions.

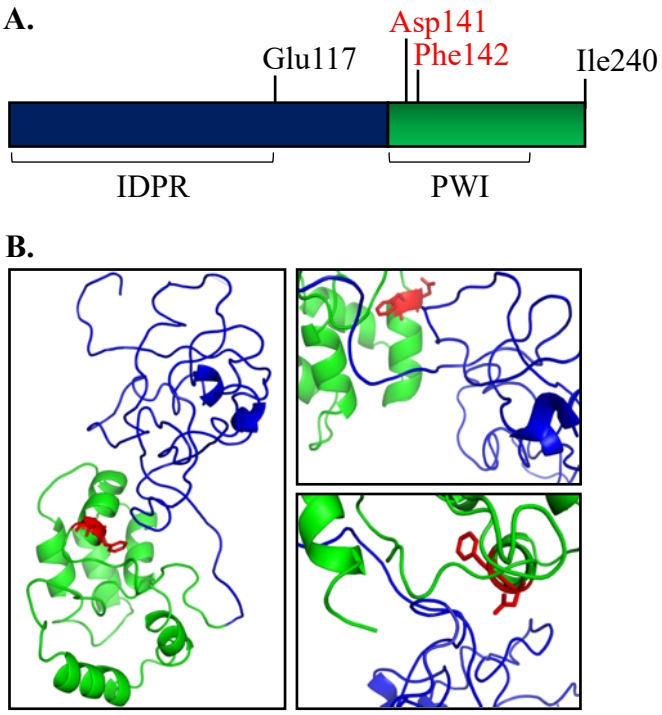


Figure 3.22 The PWI-like domain in N-HelQ as a target for SDM. A. A schematic of N-HelQ with the PWI-like domain (green). Residues Asp-141 and Phe-142 indicated were isolated in alignments with Brr2 as important for function and were targeted for SDM to produce HelQ $^{D142A/F143A}$. **B.** PHYRE2 predicted structural models of N-HelQ. The helical structure conformation of PWI-like domain (green) is located at the C-terminal region to the less structured region of disorder (blue). SDM targets identified in red. Right panels show the targeted residues magnified and their position in N-HelQ.

3.13 Post-translational modifications of N-HelQ are predicted *in silico.*

3.13.1 Phosphorylation of N-HelQ.

IDPs like other proteins can gain function by dynamic post-translational modifications (PTMs). Additionally, pull-down assays that suggested links between HelQ and the kinase ATR further implicates the potential phosphorylation of HelQ.[237] We first used the phosphorylation prediction calculator, NetPhos 3.1,[330] to interrogate the HelQ primary sequence for predicted interactions with 15 protein kinases, including ATR, cdc2, GSK3, PKC, ATM, PKA, CKII and DNAPK. NetPhos 3.1 produced a score for each potentially phosphorylated residue. There were 27 residues in N-HelQ with a phosphorylation potential above the threshold

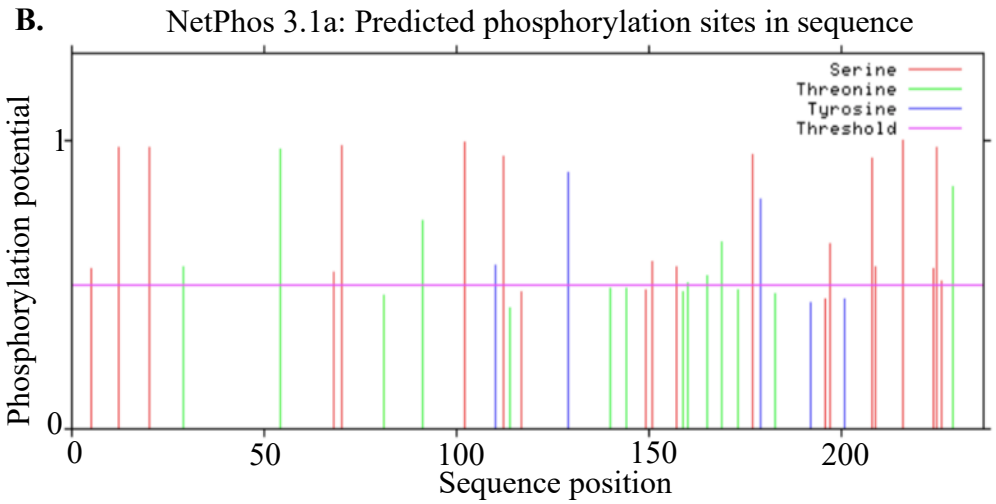
of 0.5 (50%), which were considered interesting as potential sites for phosphorylation (figure 3.23 A-B). Two candidate serine residues of the 27 hits were located in the predicted PWI-like domain of N-HelQ, conserved in Brr2, and subsequently were taken forward for experimental analysis (figure 3.23 C-D). While there were many serine residues to choose from, a potential role of the PWI-like domain made this a potentially interesting place to start.

The two predicted phosphorylated serine residues within the PWI-like domain were targeted for mutagenesis (section 2.14.3). Mutagenesis of Ser-157 and Ser-177 in N-HelQ to aspartate aimed to mimic phosphorylation; the mutant hitherto is referred to as N-HelQ$^{S157/177D}$ (figure 3.23 C-D).[324] The phosphorylation state of HelQ when purified in *E. coli* is different to HelQ from insect cells. Therefore, this mutant was important to mimic specific sites for phosphorylation.

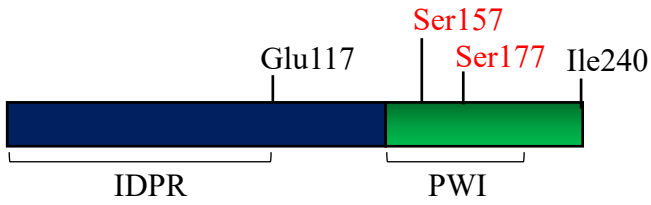
A.

```
DECGSRIRRRVSLPKRNRPSLGCIFGAPTAAELEPGDEGKEEEEMVAENR   #     50
RRKTAGVLPVEVQPLLLSDSPECLVLGGGDTNPDLLRHMPTDRGVGDQPN   #    100
DSEVDMFGDYDSFTENSFIAQVDDLEQKYMQLPEHKKHATDFATENLCSE   #    150
SIKNKLSITTIGNLTELQTDKHTENQSGYEGVTIEPGADLLYDVPSSQAI   #    200
YFENLQNSSNDLGDHSMKERDWKSSSHNTVNEELPHN                #    250
....S......S.......S........T.....................   #     50
...T.............S.S.....................T.........   #    100
.S.......Y.S.................Y.....................   #    150
S.....S..T....T...T.......S.Y.................S...   #    200
.......SS......S.......SSS..T........
```

B.



C.



D.

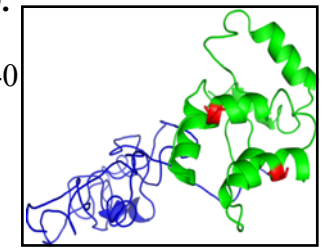


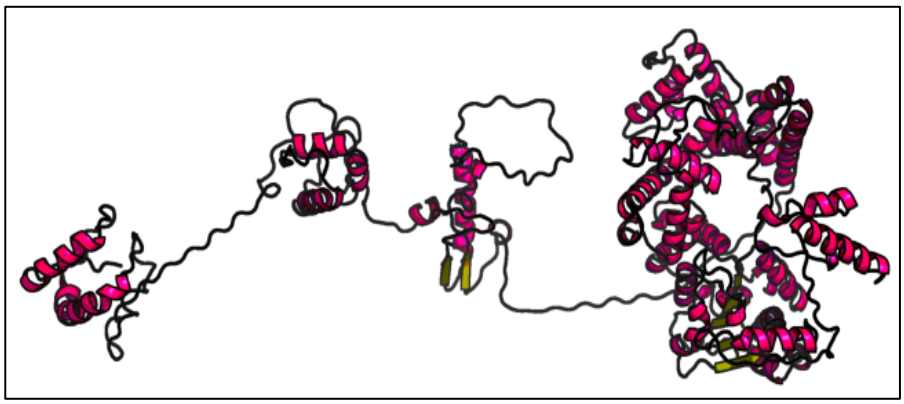
Figure 3.23 Predicted phosphorylation sites and SDM targets in N-HelQ. **A.** Potential targets for phosphorylation in N-HelQ were isolated using the bioinformatic tool NetPhos 3.1a. **A.** Serine (S), tyrosine (Y) and threonine (T) residues were isolated in N-HelQ protein sequence as potential targets. Residues within the conserved PWI-like sequence are highlighted in red. **B.** Residues were identified with a threshold of more the 0.5 as a candidate for phosphorylation. **C.** Candidates that overlapped with the PWI-like domain, shown in the cartoon schematic, were targeted for SDM to mimic phosphorylation activation. These included residues Ser-157 and Ser-177 which were mutated to aspartate. **D.** PHYRE2 predicted model of N-HelQ, the PWI-like domain (green) and targeted residues (red).

3.13.2 SUMO-Interacting Motifs in N-HelQ.

Other PTMs that were investigated using bioinformatics included ubiquitination and SUMOylation, in particular SUMO-Interacting Motifs (SIMs).[331] Small ubiquitin-like modifiers (SUMO) are associated with regulating gene transcription, the cell cycle, DNA repair and protein localisation. SUMOylation is the covalent attachment of SUMO to a target protein through an isopeptide bond. SIM interactions form electrostatic bonds between SUMO and the target protein. A β-strand on the SIM-containing protein binds either parallel or anti-parallel with a β-strand on SUMO1 or SUMO2. The specific binding to SUMO proteins occurs at S-X-S residues surrounded by hydrophobic and acidic regions or the presence of phosphorylated serine residues within the SIM.[331]

SUMO paralogues mediate specific protein-protein interactions when bound to a SIM. For example, in *S. cerevisiae,* the SIM-containing Srs2 helicase interacts with PCNA and Rad51 to abolish HR intermediates.[233] A SIM motif predicted in N-HelQ and sequence homology with Srs2 (28.36%) suggests a SIM domain may have a role in mediating protein interactions (figure 3.24).

A.



B.

Position	Peptide Area	Score (%)	Type
63-37	AGVLPVEVQPLL LSDSPEC	63.226	SUMO interaction
371-375	GKTLVAEILMLQ ELLCCRK	62.871	SUMO interaction
720-724	IDTIGESILILQ EKDKQQV	72.309	SUMO interaction
731-735	QEKDKQQVLELI TKPLENC	61.233	SUMO interaction
1080-1084	AEALQEEVEELL RLPSDFP	59.694	SUMO interaction

Figure 3.24 SUMOylation Interacting Motif (SIM) predictions in HelQ. Protein SIMs interact with SUMO molecules via non-covalent interactions. SIMs activate proteins targeting them for protein-protein interactions. **A.** RAPTOR prediction model of Srs2, a *S. cerevisiae* helicase, involved in abolishing undesired HR intermediates. Srs2 has sequence homology with HelQ with an undefined N-terminal region and also contains a SIM motif. Potential over-lapping structure/function with HelQ suggested HelQ may also have a SIM motif. **B.** GPS-SUMO tool predictions of SIMs in HelQ sequence. Consensus sequence ⋔-K-X-E-⋔s (⋔ = a hydrophobic amino acid) and the favored surrounding environment were identified in HelQ. There were 5 potential hits with a score greater than 50% for SIM interactions.

3.14 Biochemical analysis shows monomeric N-HelQ.

3.14.1 Native polyacrylamide gel electrophoresis (PAGE).

After unsuccessfully determining a role of oligomerisation of N-HelQ *in silico,* N-HelQ was biochemically assessed similarly to FL-HelQ (section 2.10.2). In Native PAGE, N-HelQ species were unaffected by the presence of ATP and magnesium consistent with the inability to hydrolyse ATP. The dominant N-HelQ population migrated between 146-242 kDa, consistent with a tetramer. Denaturing

controls dissociated tetrameric N-HelQ suggesting the formation of higher oligomers was stable (figure 3.25). Tetrameric N-HelQ is consistent with the pattern of multiple dimeric oligomers seen with FL-HelQ implicating N-HelQ as an active component in oligomerisation or in protein stabilisation. Protein aggregation was also detected at 480 kDa. N-HelQ may aggregate in the absence of activation by a binding partner.

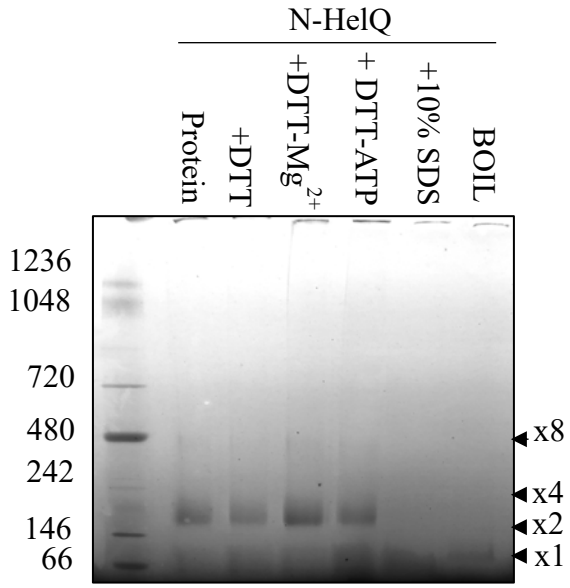


Figure 3.25 Native PAGE shows N-HelQ forms higher oligomeric structures. Native PAGE analysis showed N-HelQ formed higher oligomers under varying conditions including the addition of 25 mM DTT, 5 mM $MgCl_2$ and 5 mM ATP. N-HelQ formed stable tetramers (146-242 kDa) and monomers. This suggests N-HelQ is highly stable and does not aggregate. Protein denaturing controls to reduce the protein to monomeric included boil at 95°C and addition of 10% SDS.

3.14.2 Analytical Gel Filtration of N-HelQ.

AGF analysis of N-HelQ was carried out as previously described for FL-HelQ (section 2.20). While no DNA-binding motifs are predicted within N-HelQ we also carried out AGF in the presence of Cy5-ssDNA in case the presence of DNA impacted the environment to a more favourable one for N-HelQ. Multiple oligomeric states of N-HelQ were observed in the presence and absence of ssDNA (figure 3.26 A). The addition of ssDNA shifted the prominent N-HelQ species from

78 kDa (peak 2) to 67.7 kDa (peak 5), consistent with monomers/dimers (predicted at 47 and 94 kDa respectively). Furthermore, Cy5-ssDNA shifted N-HelQ into a second species at 91 kDa (peak 4), not observed with N-HelQ alone. This was consistent with dimeric N-HelQ (predicted at 94 kDa). Unbound Cy5-DNA also eluted off the column after 15 mL. EMSA gels run of the protein containing fractions suggested this was N-HelQ-DNA complexes (figure 3.26 B). The majority of N-HelQ was found in peaks 2 and 4, evident from SDS PAGE (figure 3.26 C). Therefore, N-HelQ may be able to weakly interact with ssDNA in dimeric form. Minimal protein aggregation was also evident in the presence of ssDNA (470 kDa, peak 3). Weak DNA binding is indicative of a role for N-HelQ that is activated upon oligomerisation. However, this may be caused by over-crowding or specific assay conditions. More extensive analysis is required to confirm this activity. Additionally, more prominent DNA binding by N-HelQ may require activation, for example through PTMs, to detect improved levels of binding. Therefore, AGF was carried out on N-HelQ$^{S158/178D}$.

AGF analysis of phosphomimetic N-HelQ$^{S158/178D}$ was carried out in the presence and absence of ssDNA (figure 3.27). There was no difference in peak distribution between wild type N-HelQ and phosphomimetic N-HelQ$^{S158/178D}$. The prominent peak of phosphomimetic N-HelQ$^{S158/178D}$ was observed at 72.4 kDa (peak 2) similar to 78 kDa for N-HelQ. However, addition of ssDNA resulted in aggregated phosphomimetic N-HelQ$^{S158/178D}$ (311.2 kDa, peak 3) as well as dimers (111 kDa, peak 4) and monomers (peak 5). EMSAs confirmed phosphomimetic N-HelQ$^{S158/178D}$ did not interact with ssDNA. Therefore, phosphorylation, at these sites, did not activate DNA binding or oligomerisation of N-HelQ.

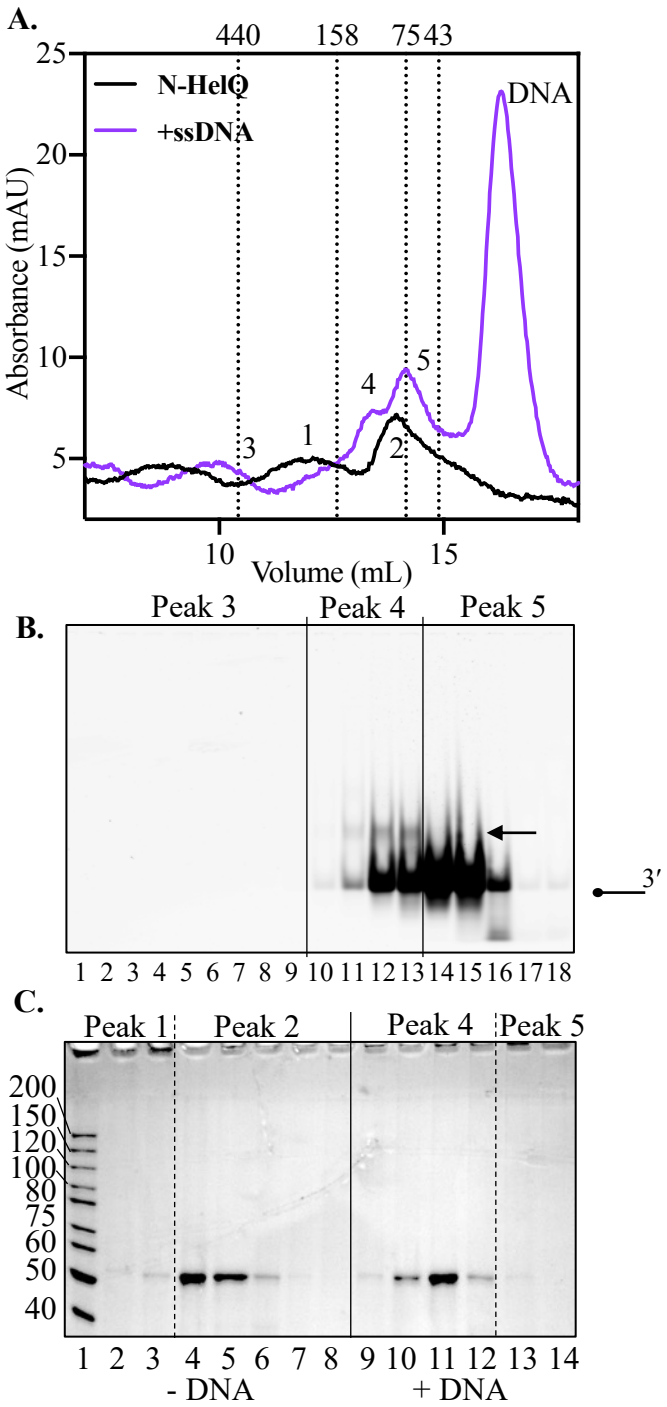


Figure 3.26 AGF shows N-HelQ forms predominantly monomeric and dimeric oligomers. A. AGF UV trace data of N-HelQ only (black solid) and N-HelQ with 25 nM 5′ Cy5-ssDNA (blue dotted). Markers used are high molecular weight markers (figure 3.9). There was very little change in oligomeric state with the addition of activating components, N-HelQ formed predominantly monomeric and dimeric in both conditions. Unbound Cy5-ssDNA can be seen eluting as indicated. **B.** 5% acrylamide TBE gel analysis of N-HelQ with ssDNA shows some week binding of dimeric N-HelQ. **C.** 8% acrylamide SDS PAGE analysis confirmed the presence of N-HelQ across the peaks.

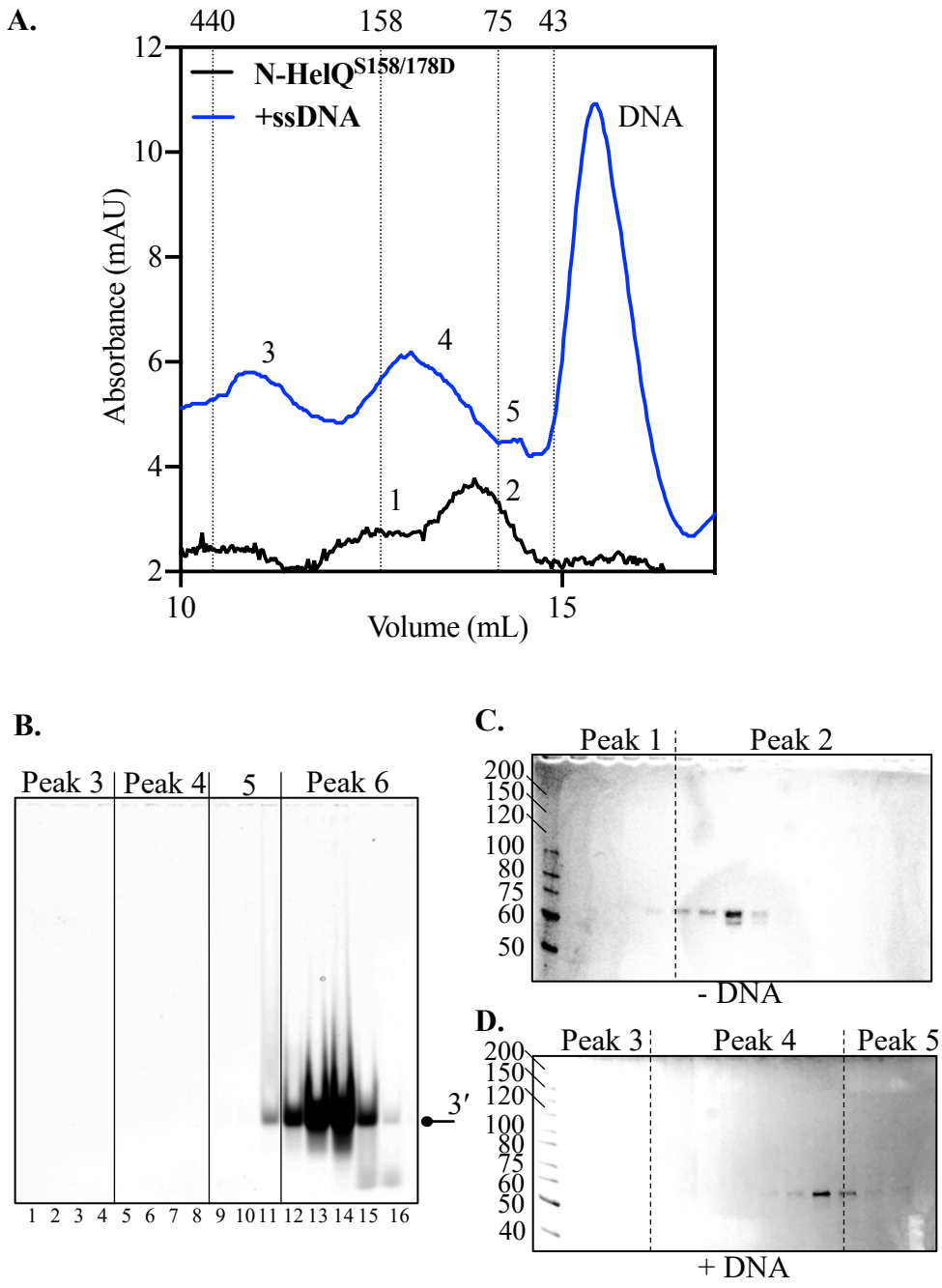


Figure 3.27 AGF shows phosphorylation of N-HelQ has no impact on oligomerisation. A. AGF UV trace data of phosphomimetic N-HelQ$^{S158/178D}$ only (black) and N-HelQ$^{S158/178D}$ with ssDNA (blue dotted). N-HelQ$^{S158/178D}$ forms a similar oligomeric pattern to wild type N-HelQ forming dimeric forms, however, more aggregation is observed. Markers shown are high molecular weight standards (figure 3.9). **B.** Acrylamide TBE EMSAs of N-HelQ$^{S158/178D}$ with ssDNA (25 nM) shows smearing complex formation of monomeric protein in peak 5 (lanes 13-14). The presence of N-HelQ$^{S158/178D}$ was confirmed in SDS PAGE analysis of **C.** protein only run in peak 2 (dimeric/monomeric) and **D.** protein with ssDNA in peaks 4 and 5 (dimeric/monomeric).

3.14.3 SEC MALS predicted N-HelQ as monomeric.

In SEC MALS, N-HelQ eluted as predominantly monomeric in the absence of activating agents at 47 kDa (figure 3.28). Protein at 165 kDa was consistent with aggregated or tetrameric protein (tetramer predicted at 188 kDa). This aligns with AGF data where N-HelQ eluted between monomeric and dimeric in the absence of DNA. Combined, these provide evidence that N-HelQ is not involved in oligomerisation but its existence as a monomer confirms it as highly stable (monomer at 47 kDa). Discrepancies between AGF and SEC MALS for calculated molecular weight were observed. These occurred because of differences in techniques and protein organisation and will be discussed later.

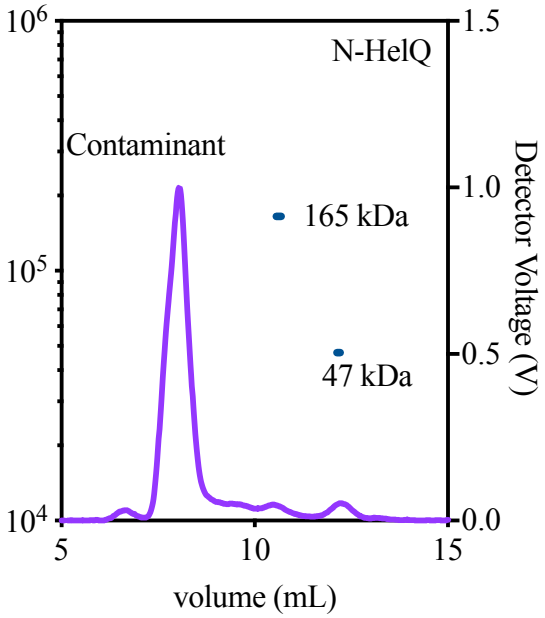


Figure 3.28 Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC MALS) of N-HelQ. Size exclusion chromatography was carries out using a superdex 200 gel filtration column. Multi-angle light scattering was measured and analysed using ASTRA 6 and WYATT technology. Graph shows MALS detector voltage (V) plotted against elution volume of protein species. N-HelQ run alone was detected at 165 kDa and 47 kDa consistent with predominantly monomeric.

3.15 Further biochemical analysis of the oligomeric state of N-HelQ.

3.15.1 Analytical ultra-centrifugation (AUC).

Results presented so far suggested N-HelQ exists in a stable state with minimal aggregation. Therefore, implicating N-HelQ in stabilising active HelQ. Native

PAGE and AGF suggested some higher oligomers forming and SEC MALS suggested monomers. The presence of activating agents did not impact the state of N-HelQ suggesting oligomerisation is not a function of N-HelQ. The lack of N-HelQ aggregation suggests a role of N-HelQ in stability. This may occur either during the assembly of higher oligomers, or once assembled. However, due to inconsistences between methods in defining the oligomeric state of N-HelQ, AUC was used as a definitive highly precise method (section 2.21.4).

AUC, carried out at RCaH, analyses the velocity of protein sedimentation in solution by assessing particle movement in high centrifugal fields. The sedimentation pattern over time, measured as absorbance, dictates the size and shape of molecules.[298] AUC was carried out on N-HelQ at 0.25 mg/mL, 0.5 mg/mL and 1 mg/mL (figure 3.29). AUC of FL-HelQ and C-HelQ was not achieved because of low protein yield. N-HelQ sedimented predominantly as monomeric (45.7 kDa, peak 1) with some dimeric N-HelQ (107 kDa, peak 2). Additionally, trace sedimentation peaks, at 238 kDa and 398 kDa, suggested some protein aggregation. AUC data was consistent with SEC MALS, confirming N-HelQ does not form higher oligomers. N-HelQ is unlikely to be responsible for FL-HelQ dimerisation. However, a lack of aggregation further implicates a role in stabilisation.

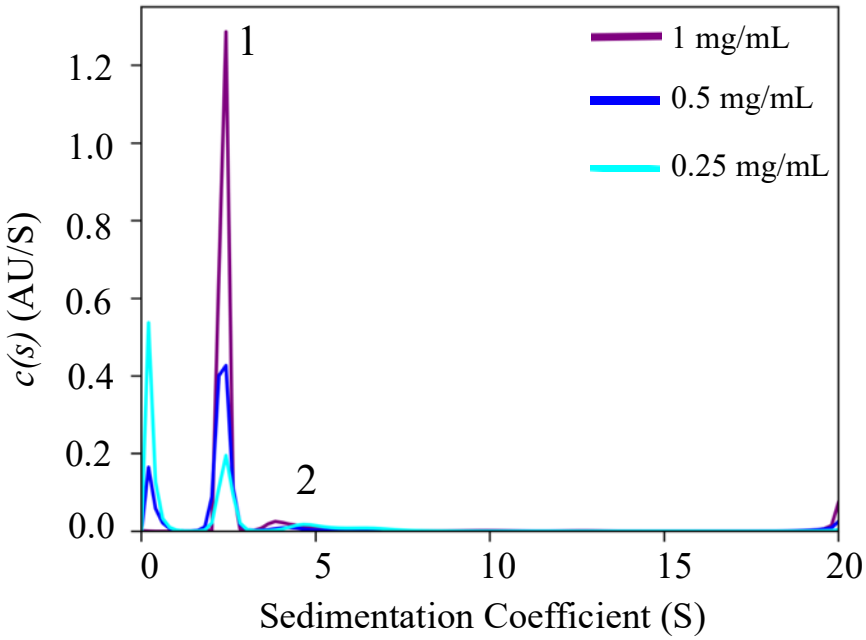


Figure 3.29 Analytical ultra-centrifugation (AUC) confirms monomeric N-HelQ. AUC analysis of N-HelQ at 0.25, 0.5 and 1 mg/mL as a highly precise method to assess protein oligomeric state. Absorbance over time of protein sedimentation was converted and plotted as the sedimentation coefficient (S). N-HelQ sedimentation pattern identified two populations at 45.7 kDa and 107 kDa, consistent with monomeric and dimeric N-HelQ. The peak at 0 sedimentation was background and not recorded as protein species.

3.16 Biochemical analysis shows C-HelQ as an unstable dimer.

3.16.1 Native polyacrylamide gel electrophoresis (PAGE).

In the absence of N-HelQ forming higher oligomeric species, C-HelQ was assessed for being responsible for protein oligomerisation. C-HelQ did not produce reliable results using native PAGE because of low protein yield (figure 3.30). A weak band between 146-242 kDa was consistent with a dimer (166 kDa). Protein aggregation was also detected at 720 kDa. C-HelQ is highly unstable in comparison to N-HelQ, this resulted in less defined populations being detected. Therefore, N-HelQ may stabilise higher oligomeric forms of HelQ which originates from C-HelQ. A similar pattern of dimeric oligomers of C-HelQ to FL-HelQ confirms this is the preferred form of the protein and is likely to require elements of the entire protein.

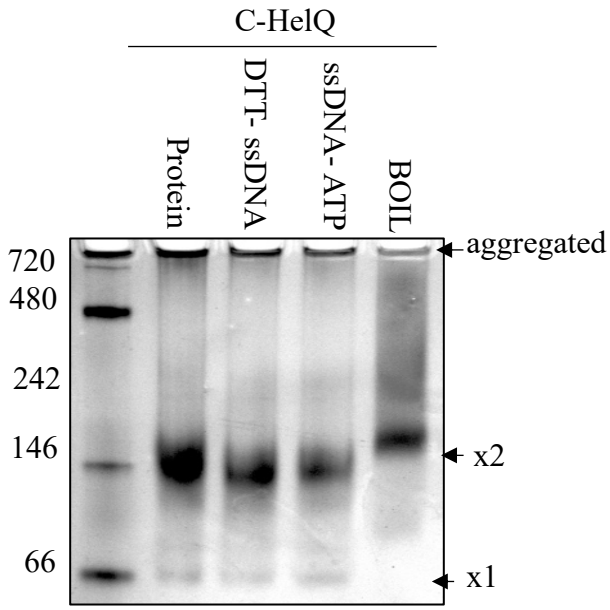


Figure 3.30 Native PAGE shows C-HelQ forms higher oligomeric structures. Native PAGE analysis shows C-HelQ forms higher oligomers under varying conditions including the addition of 25 mM DTT, 5 mM ATP and 25 nM ssDNA (MW12). C-HelQ formed mostly dimeric (166 kDa) and aggregated proteins in the presence of ssDNA and ATP. However, the boil control failed to dissociate the higher oligomeric state of C-HelQ. This provides further evidence to the instability of C-HelQ.

3.16.2 Analytical gel filtration of C-HelQ.

Consistent with observations for FL-HelQ, AGF of C-HelQ resulted in multiple species being detected. In the absence of activating co-factors, a single peak indicated monomeric C-HelQ at 65 kDa (figure 3.31 A; curve 1, peak1); predicted at 83 kDa. However, the reduced UV detection suggested very little protein eluted off the column. Therefore, this suggests C-HelQ remained aggregated on the top of the column, confirming its instability. Addition of activating co-factors (ATP-Mg^{2+}) resulted in a higher proportion of C-HelQ forming monomers at 64 kDa (curve 2, peak 3). Similar to FL-HelQ, protein activation may reduce the aggregation of C-HelQ, shifting it to an active state. Activated C-HelQ did not form higher oligomers to the same extent as activated

FL-HelQ. However, a small population, consistent with dimeric C-HelQ (218 kDa, peak 2), was observed (predicted at 166 kDa).

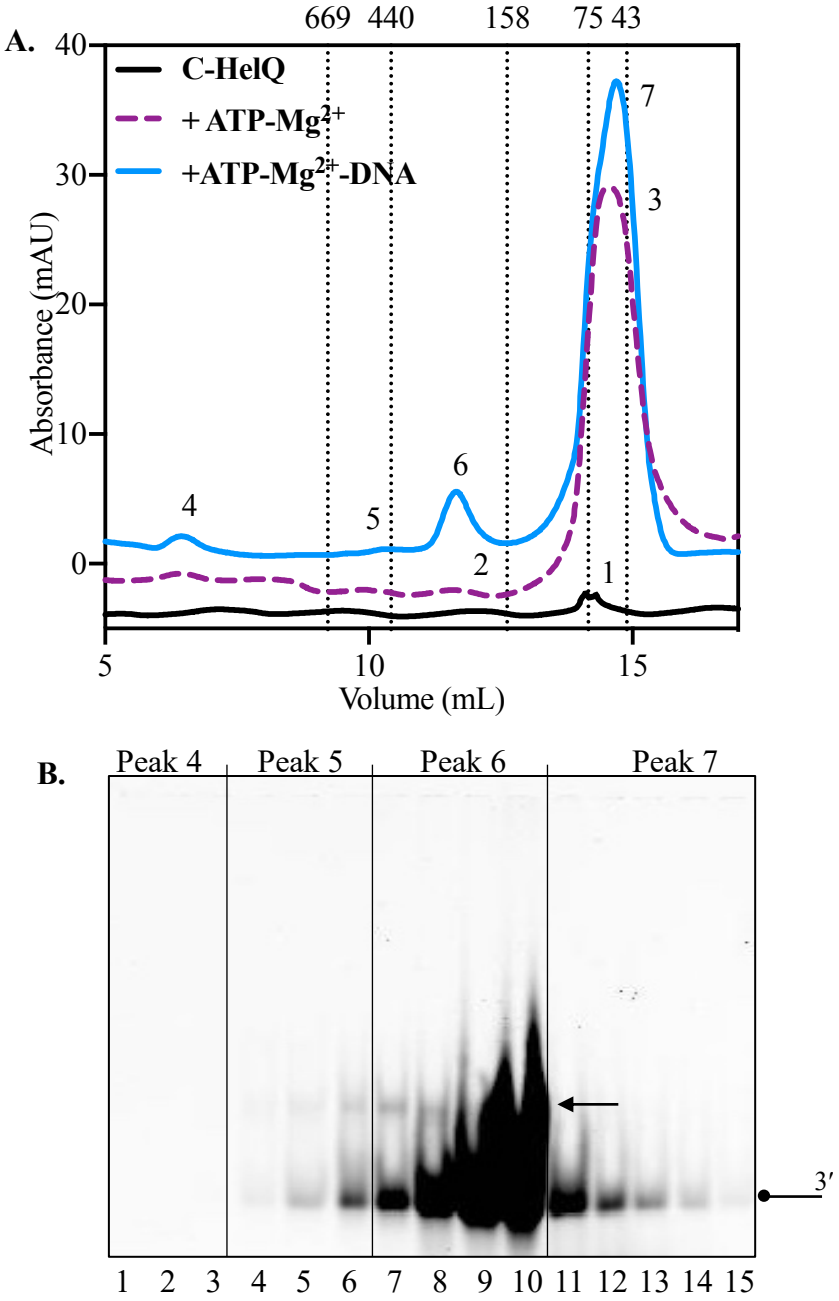


Figure 3.31 AGF shows C-HelQ forms unstable dimers. A. AGF UV trace analysis of C-HelQ protein only (black), C-HelQ with 5 mM ATP-Mg^{2+} (black dotted) and C-HelQ with 25 nM ssDNA (MW12) and 5 mM ATP-Mg^{2+}. Curve 1 (black), C-HelQ only, shows various peaks with one prominent peak of interest (peak 1). A lack of protein implicates the protein aggregated on the top of the column. In the presence of 5 mM ATP-Mg^{2+} (purple) similar monomeric C-HelQ was observed however less aggregation was evident. Further addition of ssDNA (blue) resulted in the formation of dimeric C-HelQ at 229 kDa (peak 6). **B.** Acrylamide TBE EMSA of curve 3 with ATP-Mg^{2+} and ssDNA shows formation of dimeric C-HelQ. Filled dots indicate 5′ Cy5 label.

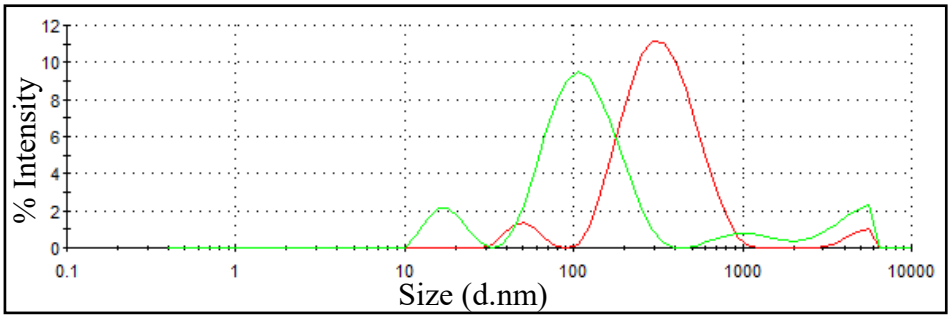
The presence of ssDNA (curve 3) further increased the proportion of dimeric C-HelQ (229 kDa, peak 6) and these complex formations were confirmed in EMSAs (figure 3.31 B). Additionally, monomeric C-HelQ (60 kDa, peak 7), the most abundant species, tetrameric (407 kDa, peak 5) and aggregated (peak 4) protein were observed. C-HelQ is activated and oligomerises in a similar dimeric pattern to FL-HelQ. However, the abundance of monomeric C-HelQ suggests it is unable to maintain stable complexes (predicted at 83 kDa). This may imply that oligomerisation occurs within the C-HelQ region but requires stabilisation from N-HelQ. AGF analysis of C-HelQ and N-HelQ suggests both fractions of the protein are required for HelQ to form active and stable higher oligomers.

3.17 Dynamic light scattering (DLS) as a tool to analyse HelQ stability.

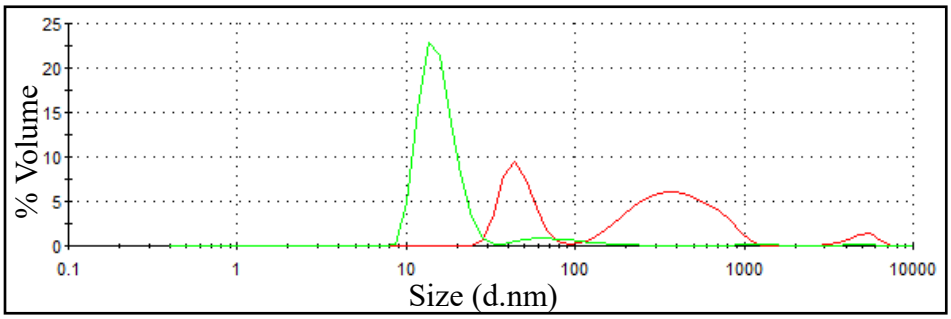
Analysis of the oligomeric state of HelQ implicates an active dimeric form. However, this analysis highlighted the level of instability within the HelQ protein. Dynamic light scattering was therefore used to assess protein stability and further assess protein aggregation (section 2.21.1). Initially, DLS was used to determine particle size of FL-, C- and N-HelQ proteins as a measure of protein aggregation and oligomeric state. A proteins size distribution, the spread of sizes within a population of a protein, can be used to determine how globular a protein is. For example, a population that has a small spread would be considered globular and a population that has a more varied range it likely to adopt a more linear conformation. The ratio of the different sizes within a single population can also be used as an indicator of protein oligomerisation. For example, a population may have a size of X and 4X, suggesting monomer and tetramer species. The Z-average, a measure of the average diameter of the protein size, was 261.75 nm for FL-HelQ and 93.15 nm N-HelQ (figure 3.32). This suggests N-HelQ does not form aggregates larger than FL-HelQ supporting the idea that N-HelQ is relatively stable.

The difference in diameter size between FL-HelQ and N-HelQ is similar to the difference in sequence length, suggesting similar oligomeric states. N-HelQ makes up 21% of the FL-HelQ sequence and 35% of the diameter of FL-HelQ. The poly-dispersity index (PdI), a measure of the distribution of molecular mass, was 0.2845 for FL-HelQ which was categorised based on known samples, as moderately mono-dispersed. Conversely, N-HelQ, at 0.5715, was categorised as mid-range dispersed. Therefore, N-HelQ had a more varied population then FL-HelQ, contradicting AGF and Native PAGE results.[332] This may be due to the intrinsic disorder of N-HelQ resulting in a larger variation of conformations of N-HelQ instead of N-HelQ existing in multiple higher oligomers. However, PdI is calculated from the absorbance intensity, a measure of the strength of the signal of each species which can be impacted by particle size. For example, a small population of a large particle will give a higher intensity to a large population of a small particle, therefore it does not represent the distribution within a sample. Analysis as percentage volume, a measure of the percentage of each species within the entire population, suggests the majority of N-HelQ exists as a single species by comparison with three for FL-HelQ (figure 3.32 B). This is consistent with AUC and SEC MALS data.

A.



B.



C.

	Z-Average (diameter nm)			PdI		
	1	2	Average	1	2	Average
FL-HelQ	270.0	253.5	261.75	0.288	0.281	0.2845
N-HelQ	93.83	92.47	93.15	0.587	0.556	0.5715

Figure 3.32 Dynamic light scattering (DLS) confirmed the stability of N-HelQ. A. DLS spectra of protein-only identified one prominent species for FL-HelQ (red) and N-HelQ (green) when looking at the intensity of different sized species within the population. **B.** FL-HelQ had two prominent species when analysing the proportion of a species in the total volume and N-HelQ remains a single prominent species. The data was normalized to storage buffer (150 mM NaCl and 10% glycerol). **C.** The average diameter calculated suggested similar oligomeric states of FL-HelQ and N-HelQ. The dispersity index (PdI), a measure of the distribution of molecular mass in each sample, suggests a moderately mono-dispersed FL-HelQ population compared to a mid-ranged poly-dispersed population of N-HelQ based on analysis from graph **A**.

More extensive DLS, using a Zetersizer Nano model, assessed protein stability by measuring protein aggregation in different conditions.[333] Changes in particle diameter was indicative of protein conformational changes and associated stability. Results are presented as percentage volume of each protein species, as an indication of the favourable state of the protein within each condition tested. Percentage volume indicates the largest species within the population, not the largest particle

size. N-HelQ favoured a single population consistent with previous DLS data (figure 3.33 A). Addition of ATP destabilised N-HelQ into two populations. Further addition of ssDNA recovered stability back to a single population.

C-HelQ was confirmed as unstable using DLS with a calculated PdI of 0.65 (figure 3.33 B). In the presence of ATP and ssDNA the PdI of C-HelQ shifted to 0.58 suggesting highly poly-dispersed populations. The change in PdI was insignificant implying activation did not result in a stable population being detected. Inconsistencies in replicas suggested dynamic aggregation and dissociation of C-HelQ particles. This provides further support to the evidence that C-HelQ, which may be able to dimerise, is unable to form oligomers that are stable enough to remain long enough for consistent detection. While a mono-dispersed population was not evident, addition of activating co-factors did result in structural changes of C-HelQ, shown in AGF analysis. These results, combined with previous observations, confirms the instability of the core helicase which is expected because of bioinformatics carried out on FL-HelQ. To expand on work carried out here, DLS could be used to assess different buffer conditions and additives to determine if the stability of C-HelQ could be improved.

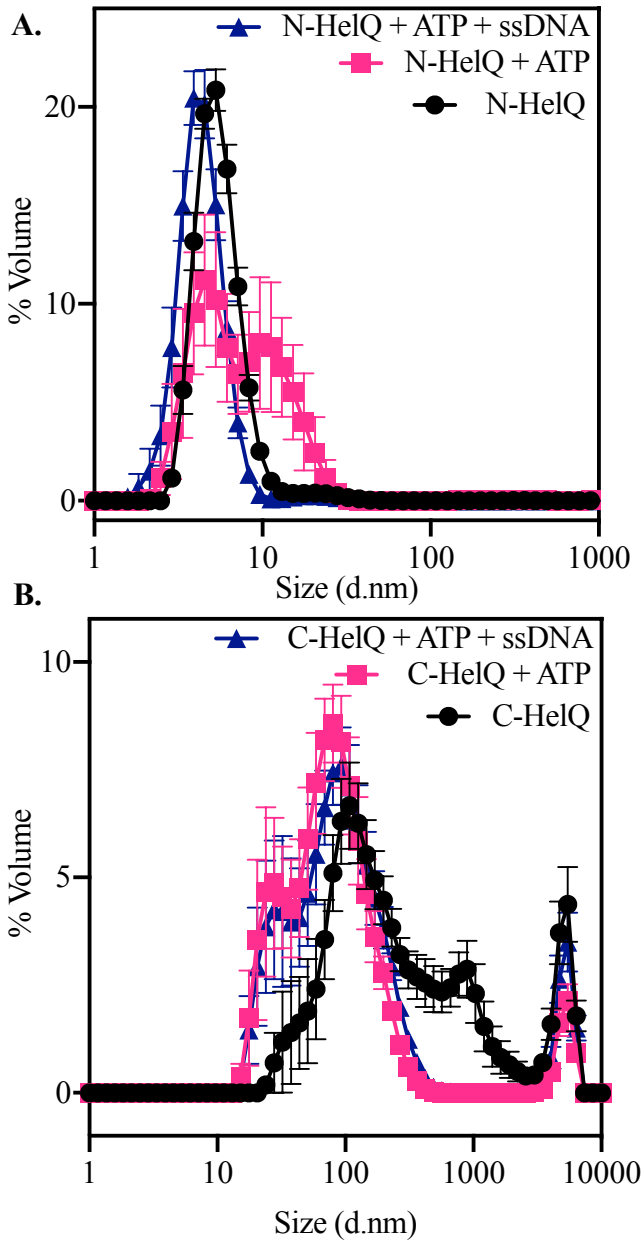


Figure 3.33 DLS showed the instability detected in HelQ fragments. DLS was carried out using the Malvern Zetasizer Nano S to measure population distribution dependent on particle size. Results are based on 10 replicas plotted as percentage volume against particle diameter (nm). This provides an indication of the dominant sized population within the sample in the different conditions tested. Protein was tested alone (black), with 5 mM ATP (pink) and with 5 mM ATP and 25 nM ssDNA (MW12) (blue) for **A.** N-HelQ and **B.** C-HelQ. N-HelQ remained relatively stable under different conditions. The presence of ATP slightly stabilized C-HelQ, however, the population remained poly-dispersed.

3.18 Summary and conclusions.

3.18.1 HelQ purification.

Optimisation of HelQ purification methods resulted in improved yields of HelQ protein. This was due to higher levels of over-expressed HelQ at the onset of

the purification process as well as rapid protocols that reduced time for degradation. HelQ protein yield was increased from 0.19 μM at the start of this work to up to 4.5 μM after alterations were made. While these concentrations were still low and high levels of degradation were observed, this allowed for further biochemical analysis and early structural studies to be carried out. Future work could further optimise purification including large scale insect cell over-expression which could aid in producing even larger quantities of HelQ per round of purification.

3.18.2 Modelling HelQ as a dimer.

FL-HelQ actively assembled onto ssDNA as dimers or multiple dimers. It can be concluded that HelQ is active as a dimer because of results obtained using SEC MALS. HelQ oligomerisation is activated by ATP-Mg^{2+} and to a lesser extent, ssDNA. Additionally, protein aggregation confirmed bioinformatic analysis of HelQ instability.

Human RecQ1, required for genome stability, exists in a similar dimeric form. RecQ1 also interacts with RPA, which is explored in chapter 5.[334] We hypothesise that dimeric HelQ binds to ssDNA close to the ssDNA-dsDNA junction of a duplex fork (see chapter 6 for proposed mechanism of action).

Variations detected between methods of oligomeric state analysis suggested HelQ oligomerisation is sensitive to surrounding conditions. Multiple protein species were detected in-gel analysis (Native PAGE) in comparison with fewer species using in-solution analysis (SEC MALS and AGF). Even variations between AGF and SEC MALS data using the same Superdex200 gel filtration system were observed. Therefore, the precise oligomeric state is hard to confirm in the absence of a solved protein structure. Additionally, differences in the calculated molecular weight between AGF and SEC MALS were observed. SEC MALS is considered a versatile and reliable means for characterising proteins in solution.[335] SEC MALS

does not estimate mass based on standards, instead, measurements are taken for each elution volume to distinguish a fixed molecular weight. AGF makes assumptions that SEC MALS does not, including: (a) HelQ shares the same conformation and specific volume to the standards and (b) like the standards, HelQ only sterically interacts to the column (charge or hydrophobic interactions do not occur). Furthermore, regions of disorder within globular proteins impact the techniques differently. Intrinsic disorder in N-HelQ impacts protein elution because of large Stokes radii in unstructured regions and non-linear oligomeric assemblies.[322,335,336] Therefore, for absolute molecular weight, SEC MALS was considered accurate and AGF was used in support. Therefore, in conclusion the molecular weight of HelQ in the presence of ATP, $MgCl_2$ and ssDNA is 240 kDa, a dimer.

3.18.3 HelQ fragments in oligomerisation.

Purification of N-HelQ from *E. coli* allowed for rapid production and large protein quantities. Data presented here suggests N-HelQ exists predominantly as stable monomers and DLS confirmed this stability. The stability of N-HelQ as a monomer implicates it in the stabilisation of active FL-HelQ. The inability of active C-HelQ to maintain a stable dimeric state further confirms a role for N-HelQ.

Core C-HelQ forms dimers similarly to FL-HelQ. While a mechanism is unknown and the data from AGF is limited, we hypothesis that C-HelQ contains the region responsible for dimerisation. AGF and DLS implicated C-HelQ as highly unstable and high levels of aggregated and monomeric C-HelQ implied the constant association and dissociation of C-HelQ monomers. Therefore, C-HelQ oligomerises but is unable to remain in this state. Consistent with FL-HelQ, the addition of activating co-factors (ssDNA and ATP), resulted in a slight shift in stability. In summary, we propose that elements within C-HelQ causes dimerisation

of active FL-HelQ which is stabilised by N-HelQ, therefore active dimers rely on the entirety of the protein. The instability of HelQ suggests rapid protein turnover in the cell. HelQ may be active at precise times during DNA repair and rapidly degraded to prevent further activity.

3.18.4 Iron-sulphur clusters in C-HelQ.

C-HelQ eluted with a brown colouration which may suggest the presence of iron-sulphur (Fe-S) clusters. Fe-S clusters are found within proteins to coordinate the assembly of iron and sulphide.[337] They are highly conserved and involved in many biological roles including DNA repair. Fe-S clusters have been reported in many DNA binding proteins and are essential components of DNA processing proteins including helicases. Clusters within SF2 helicases, including Chlr1, RTEL, FancJ and DinG have been located close to the helicase core.[338] The potential presence of Fe-S clusters, or the ability to coordinate iron, may be important for HelQ function. Further work could first determine the spectroscopic property of these C-HelQ fractions and if verified as a possible Fe-S cluster would then assess the role of Fe-S clusters in HelQ activity by mutagenesis. Fe-S clusters are characterised by cystine and histidine residues in close proximity and while there are no conserved regions of these residues across different HelQ proteins, the Ratchet (residues Arg961-Leu986) in domain IV contains a number of residues that could be targeted for mutagenesis.

3.18.5 Identification of intrinsic disorder in N-HelQ.

N-HelQ was predicted to contain regions of disorder which were confirmed experimentally by CD. Further CD analysis is required to assess N-HelQ without the SUMO-tag. Analysis of known IDPs could be used as a comparison. IDPs undergo transitions in order to activate function and subsequently the function of

N-HelQ may remain elusive until a partner is identified. Regulated folding or unfolding of IDPs can also be initiated by environment factors, including temperature, pH, redox potential or light. Identified IDPs suggest that protein function commonly occurs during the disorder-to-order transition.[324] A role of protein partners and environmental changes in activating N-HelQ requires assessment to establish the impacts on N-HelQ. C-HelQ, in the absence of N-HelQ, remains active in oligomerisation, translocation and unwinding. This suggests an additional, unknown role for N-HelQ.

Structural homology between the PWI-like domain of Brr2 and N-HelQ implicates a role of this region in protein-protein interactions. The Brr2 Ski2-like RNA helicase is required for remodelling activity during spliceosome activation.[329] It has been implicated in both the catalysis and disassembly of splicing. Brr2 is organised in a tandem of helicase units each containing dual RecA-like domains and a 400 amino acid N-terminal helicase region of unknown function.[329] Given the similarities to N-HelQ, a role of the PWI-like domain in protein-protein coordination is likely. Further analysis of the PWI-like domain in N-HelQ, with close attention to essential residues and links to the region of disorder could be further analysed. A recent near-atomic resolution crystal structure of the PWI-like domain within the N-terminal region of Brr2 providing a good place to start in N-HelQ.[329]

The potential phosphorylation of predicted sites within the PWI-like domain were assessed for a function in activating oligomerisation. However, phosphorylation did not activate oligomerisation.

The archaeal homologue Hel308 lacks the N-terminal region of human HelQ. Therefore, the function of N-HelQ will be in addition to any function of Hel308. Further analysis of the role of N-HelQ is explored in chapters 4 and 5.

3.19 Outlook and future work.

Future work to confirm the dimeric conformation is needed which could include assessing additional variants of FL-HelQ, including overlapping regions of N- and C-HelQ. Furthermore, mutagenesis of conserved residues of other SF2 helicases involved in dimerisation could isolate individual residues important for oligomerisation. Additional AUC analysis to calculate precise molecular weight of C- and FL-HelQ would also confirm oligomeric state in different conditions, however, larger quantities of protein would be required.

NMR analysis of N-HelQ started during this work will aim to continue to assess the structure and intrinsic disorder of N-HelQ in different conditions. This would be complimented by comparative CD analysis alluding to N-HelQ conformational changes in different contexts.

Elaboration of the impact of PTMs on FL-HelQ and N-HelQ is required to assess if PTMs affect function. The scope of this work was limited to the potential phosphorylation of two sites in N-HelQ. However, bioinformatic analysis identified potential SIM motifs that require further analysis.

Understanding oligomerisation enables modelling of HelQ translocation and unwinding mechanisms and inform us of the fundamental activity of HelQ. The assembly of dimeric FL-HelQ on DNA, for both translocation and unwinding, is addressed in chapter 4.

Chapter 4: Analysis of DNA processing activity by HelQ

4.1 Introduction.

Helicases are found across all domains of life.[200] They are important in nucleic acid metabolism for unwinding of duplex DNA or RNA powered by ATP-binding and hydrolysis that induces protein conformational change. Helicases are involved in maintaining genome stability through activities in DNA replication, repair, recombination and transcription.[229] Understanding how helicases interact with DNA is important because defects in translocation has been linked to diseases of premature aging and cancers. For example, mutations in the RecQ proteins (WRN, BLM and RecQ4) are associated with genomic instability and cancer.[339]

The mechanism of a DNA helicase is defined by their direction of translocation and preferred DNA substrate.[340,341] Determining these parameters can give clues about the role of helicases in the cell. HelQ binds to ssDNA and translocates along in a 3′ to 5′ direction. HelQ has also been shown to preferentially interact with forked substrates.[256] This has led to predictions for the mechanism of HelQ translocation and combined with dimeric HelQ identified in chapter 3, experiments described here have provided further insight into HelQ mechanism.

This chapter describes how HelQ interacts with DNA and RNA. C-HelQ and N-HelQ are also analysed for their roles in DNA interactions. The objectives of the work were: (a) to analyse interactions of HelQ with DNA and RNA, (b) assess HelQ helicase activity in the context of different structures and barriers, and (c) to delineate the roles of N-HelQ and C-HelQ, if any, in DNA binding and unwinding.

4.2 FL-HelQ has a higher binding affinity for DNA then HelQ fragments.

Electrophoretic mobility shift assays (EMSAs) were used to assess DNA binding by FL-, C- and N-HelQ proteins as described in section 2.16.3. End point EMSAs measured DNA binding as a function of protein concentration, using a 50-

base pair duplex fork with a 5′ Cy5 florescent label referred to as fork 2b (refer to table 2.4 for DNA sequence and figure 3.5 for fork schematic). This fork was used throughout this work as a positive control for functional HelQ DNA binding and helicase unwinding activity. EMSAs were used to assess for the stable complex formation between HelQ and DNA (figure 4.1 A-B). FL-HelQ predominantly formed a relatively stable complex with fork 2b, apparent from almost 100% free DNA being shifted into protein-DNA complex and a loss of the free Cy5 DNA signal (figure 4.1 B panel 1). A stable complex in EMSAs was suggested by the observation of a single species being detected and the absence of smearing or aggregation. Some protein was also unstable, visible as smearing of the DNA and well aggregation implying formation of less stable and aggregated protein-DNA complexes. This may be supported by the bioinformatics analysis on HelQ instability in chapter 3 (figure 3.1) indicating considerable instability of the purified protein. That may result in transient protein-DNA complexes alleviated by aggregation of the protein.

N-HelQ gave no evidence for stable binding to fork 2b DNA (figure 4.1 B). Some smearing of the Cy5 DNA signal observed at 500 nM N-HelQ (lane 7) is likely caused by non-specific macromolecular crowding of high protein concentrations around DNA. The EMSA data for N-HelQ indicates it is unlikely to have a prominent role in DNA binding. The C-HelQ fragment was also shown to form a stable complex with fork 2b, apparent from DNA being shifted into a protein-DNA complex (lane 11). However, less DNA was shifted into a DNA-protein complex with C-HelQ in comparison with FL-HelQ. Some smearing of the Cy5 DNA signal was observed at high concentrations of C-HelQ. This supports analysis of C-HelQ in chapter 3 indicating the ‘core-helicase’ is highly unstable. Analysis presented in chapter 3 (figure 3.13) indicated active FL-HelQ assembles as HelQ dimers and therefore is likely to interact with DNA as dimers. However,

C-HelQ was shown to form dimers but is unable to maintain stable DNA-protein complexes (figure 3.31). This may support a role in stabilising for N-HelQ to maintain the dimeric protein-DNA complexes suggesting both regions of HelQ are required to fulfil activity.

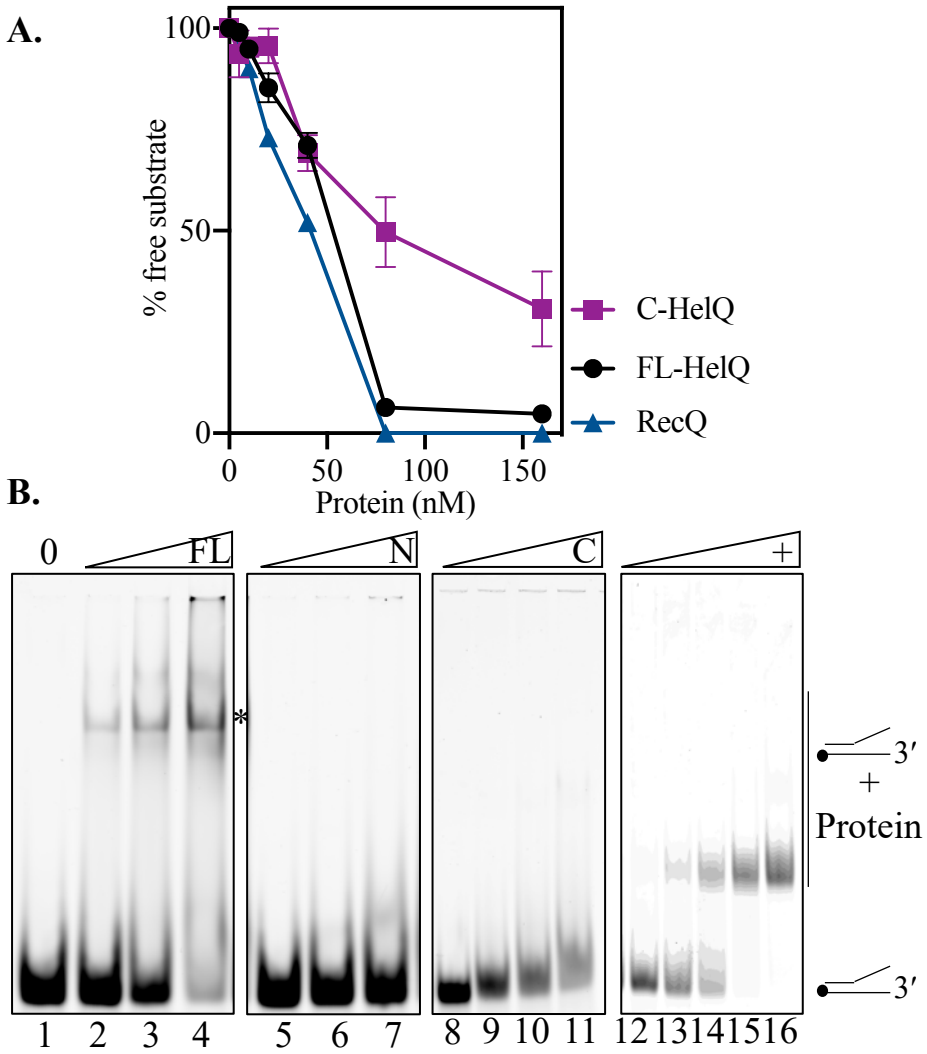


Figure 4.1 EMSAs show FL-HelQ and C-HelQ form complexes on forked DNA. A. FL-HelQ (black), C-HelQ (grey) and RecQ (blue) binding to fork 2b (25 nM). **B.** Also shown are representative TBE gels that were analysed in triplicate to calculate standard error from the mean shown here; FL-HelQ (FL), C-HelQ (C), N-HelQ (N) and RecQ (+) are as indicated. Substrate schematics shown with filled dots indicating 5′ Cy5 labelled ends. Proteins were used at 0, 5, 10, 20, 40, 80 and 160 nM. FL-HelQ and C-HelQ load onto DNA whereas N-HelQ does not.

4.3 FL-HelQ and C-HelQ unwind duplex DNA in a forked substrate.

HelQ activity was assessed by measuring helicase unwinding of DNA duplex in a forked structure (fork 2b) as described in section 2.16.2. Helicase unwinding

by HelQ of fork 2b as a function of protein concentration was assessed for FL-, C- and N-HelQ proteins (figure 4.2). Additionally, unwinding by HelQ of fork 2b as a function of time was assessed to determine if unwinding efficiency improved over longer periods of time (figure 4.3). Helicase unwinding of fork 2b by *E. coli* RecQ, a helicase in genome maintenance, was used as a positive control in this work due to its accessibility within the lab. RecQ had strong DNA unwinding ability apparent from 100% formation of Cy5 ssDNA from forked substrate at 50 nM.

FL-HelQ was unable to unwind fork 2b to 100% in these assay conditions (figure 4.2 A) and was appreciable less efficient as RecQ at fork 2b unwinding. Analysis of unwinding by FL-HelQ over time showed HelQ was unable to unwind fork 2b more then 60% of the total substrate. This may be supported by analysis on HelQ in chapter 3 indicating considerable instability of purified protein (figure 3.33). Therefore, the inability of HelQ to reach 100% unwinding of fork 2b after 30 minutes may be because of high levels of protein degradation. Additionally, the conditions required for optimal HelQ unwinding of fork 2b may not have been reached. Reduced fork 2b unwinding by HelQ may also indicate alternative roles in in DNA processing of HelQ additional to unwinding duplex DNA.

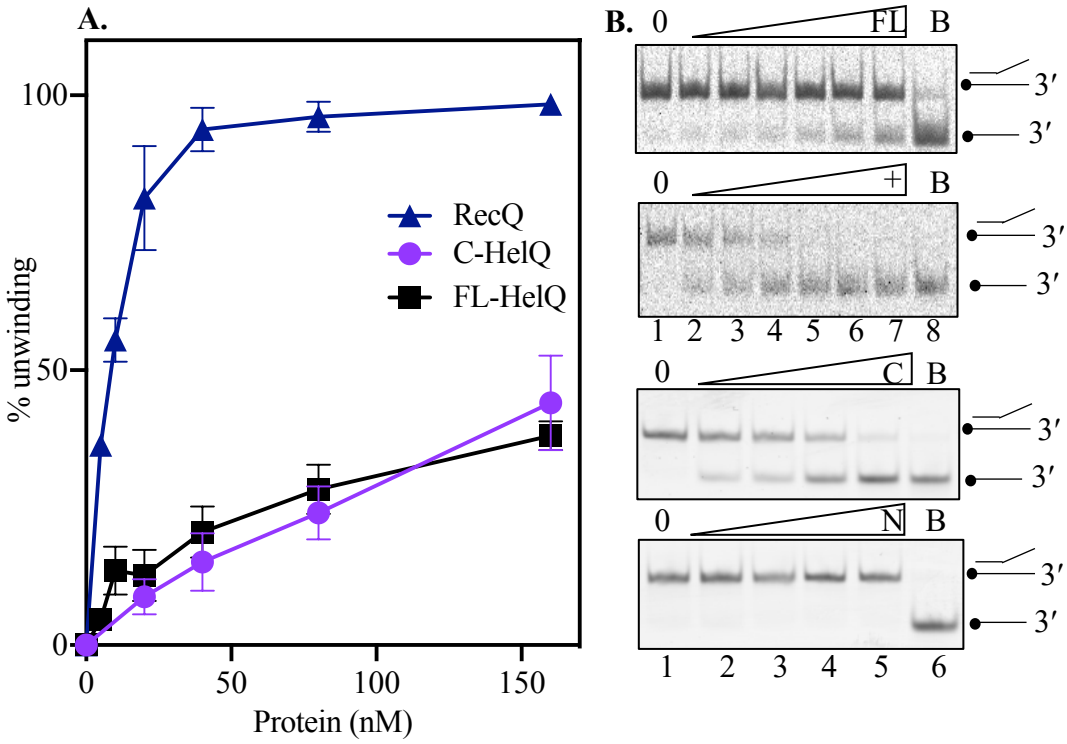


Figure 4.2 Helicase assays as a function of protein concentration show FL-HelQ and C-HelQ are active as helicases. A. A measurement of FL-HelQ (black), C-HelQ (purple) and *E. coli* RecQ (blue) unwinding activity of fork 2b (25 nM). **B.** Also shown are representative acrylamide TBE gels that were used to generate the data, reactions carried out in triplicate and shown as standard error from the mean. FL-HelQ (FL) and RecQ (+) were used at 0, 5, 10, 20, 40, 80 and 160 nM and C-HelQ (C) and N-HelQ (N) were used at 0, 20, 40, 80 and 160 nM. B represents a boiled reaction to fully dissociate the fork. Substrate schematics with filled dots indicating 5′ Cy5 label. FL-HelQ and C-HelQ unwound the DNA whereas N-HelQ did not.

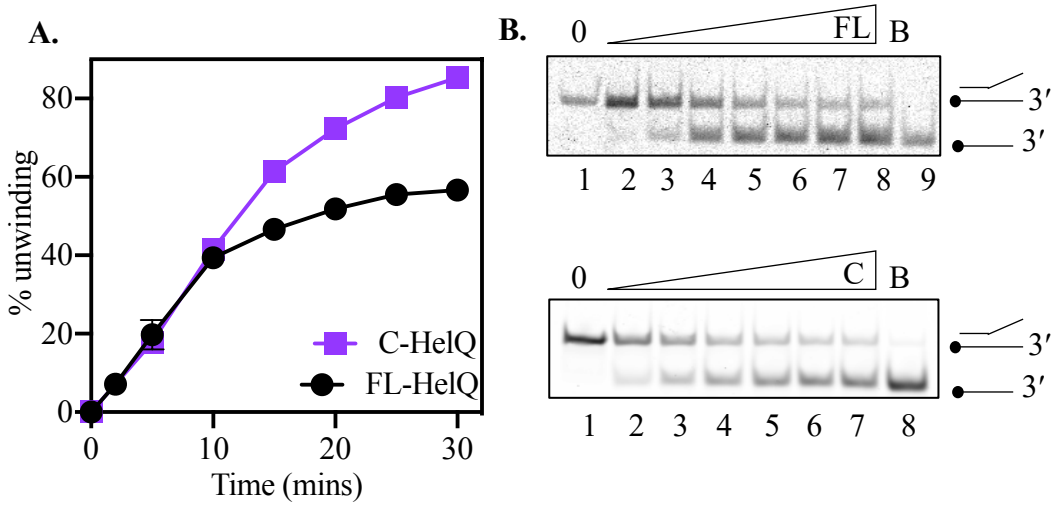


Figure 4.3 Helicase assays as a function of time show FL-HelQ and C-HelQ are active as DNA helicases. A. A measurement of FL-HelQ and C-HelQ (80 nM) unwinding activity of fork 2b (25 nM) over time. **B.** Also shown are representative TBE gels that were used to generate the data in triplicate and shown here as standard error from the mean. Substrate schematics are shown with filled dots indicating 5′ Cy5 label. Reactions were terminated at 0, 2, 5, 10, 15, 20, 25 and 30 minutes. B represents a boiled reaction to fully dissociate the fork. FL-HelQ reached 60% unwinding after 30 minutes.

N-HelQ does not contain any evidence of protein motifs required for ATP hydrolysis and as expected was unable to unwind the DNA. C-HelQ unwound fork 2b similarly to FL-HelQ (figure 4.2 A), in agreement with it being the 'core-helicase' element of HelQ. This also implies that N-HelQ is not required for the catalytic aspects of HelQ helicase activity. This is analysed further later in this chapter. Mutagenesis of the HelQ Walker B motif to inactivate the ATPase activity of FL-HelQ and C-HelQ resulted in a loss of unwinding activity, as expected, confirming unwinding observed was due to HelQ mechanism (figure 4.4).

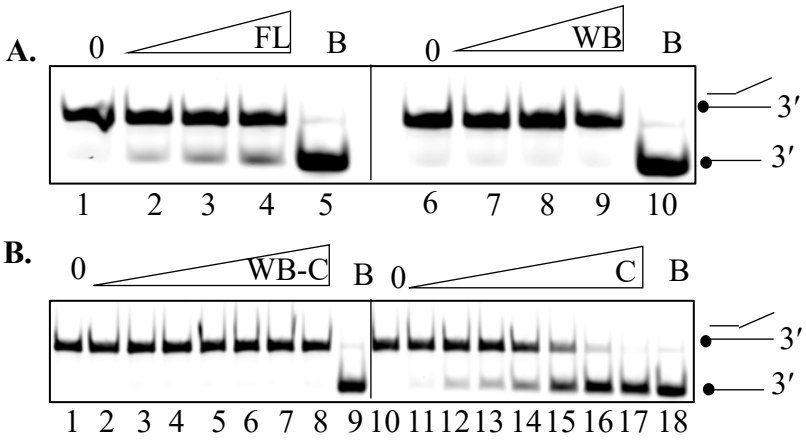


Figure 4.4 Helicase assays show ATPase inactive mutants of FL-HelQ and C-HelQ cannot unwind DNA. Representative acrylamide TBE gel to show **A.** FL-HelQ (FL) and HelQ^{D463A} Walker B mutant (WB) and **B.** C-HelQ (C) and C-HelQ^{D463A} Walker B mutant (WB-C) fork 2b (25 nM) unwinding. Proteins were used at 0, 10, 20, 40, 80 and 160 nM. B represents a boil control of dissociated substrate as indicated. Substrate schematics shown with filled dots indicating 5′ Cy5 label. Walker B mutation successfully removes ATPase activity.

4.4 Insights into HelQ DNA translocation using chemically modified DNA.

Defined chemical modifications were introduced at single sites into the unmodified fork 2b to assess their impact on HelQ activity on DNA (table 2.4). Biochemical analysis of DNA-protein interactions between HelQ and these modified substrates aimed to determine information about HelQ translocation mechanism. DNA modifications included an abasic site, phosphorothioate and methyl phosphonate substitutions (figure 4.5 A). Modifications at a single site were made within the duplex region on the translocating and non-translocating strand and in the ssDNA region of the translocating strand. The addition of abasic sites assessed the importance of internal DNA structure and base stacking on helicase activity (figure 4.5 B-C). Methyl phosphonate and phosphorothioate sites assessed the importance of the external backbone structure and were used to assess for electrostatic interactions by disrupting the DNA charge (figure 4.5 B). A methyl phosphonate substitution removed the negative charge and phosphorothioate substituted the negative charge.

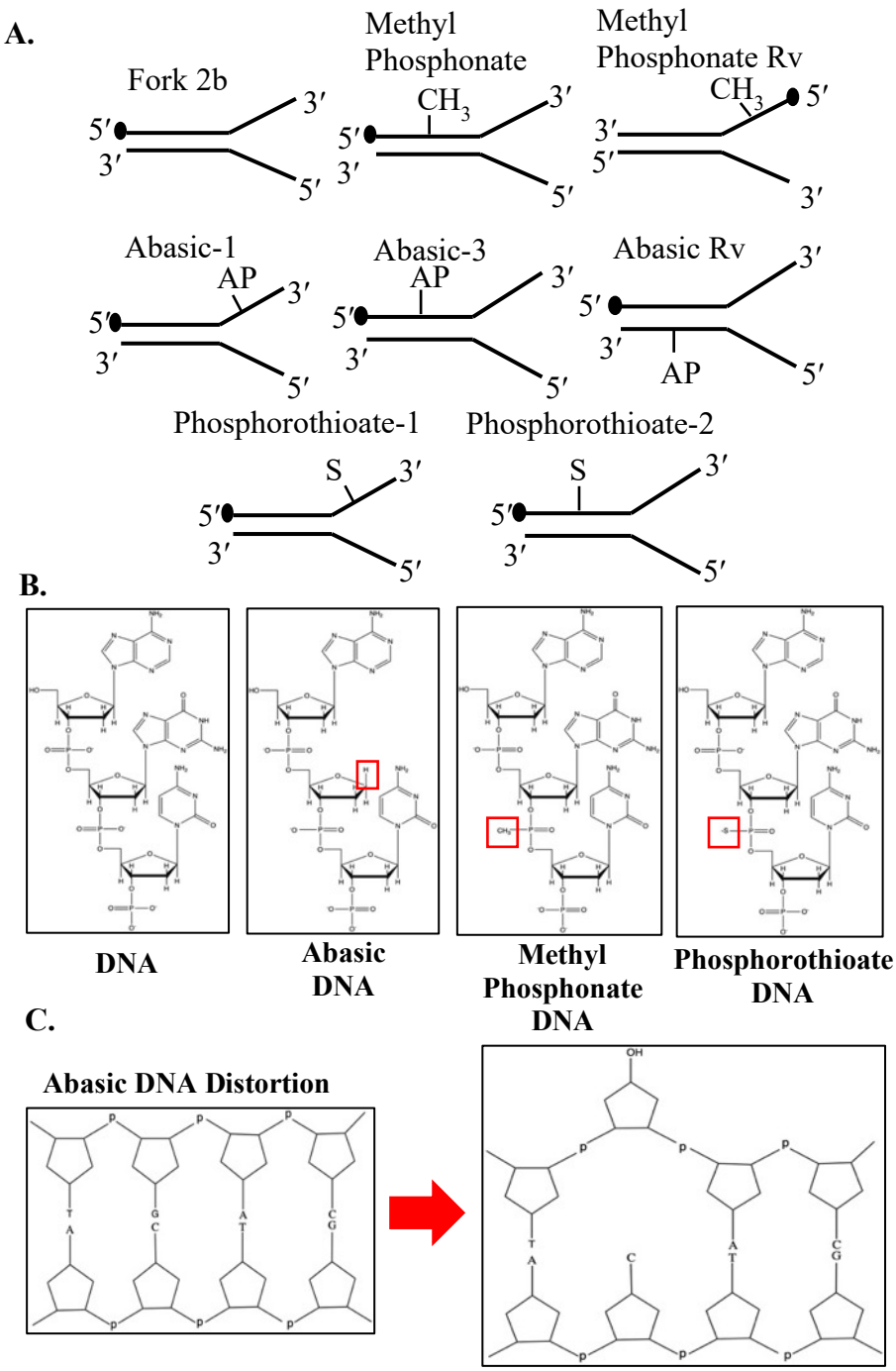


Figure 4.5 DNA schematics of the defined chemically modified substrates. **A.** Chemically modified DNA substrates were made based on the unmodified fork 2b sequence with a 5′ Cy5 label (indicated by filled dot). Chemical modifications were introduced at a single site within the duplex region and 3′ flap region of the DNA fork. Modifications included methyl phosphonate (CH_3), abasic sites (AP) and phosphorothioate (S) substitutions. **B.** Chemical representation of physical change to the DNA (indicated by red box). Left to right: normal nucleotide, abasic, methyl phosphonate and phosphorothioate substitution. **C.** Schematic of the change in base stacking caused by abasic sites.

We first tested substrates with modifications at single positions on the translocating strand. Helicase unwinding assays of HelQ at 80 nM as a function of time were carried out to determine the ability of HelQ to unwind DNA through a chemical block. All tested modification at a single position within the duplex region inhibited HelQ unwinding activity when compared with the unmodified fork 2b (figure 4.6 A-B). HelQ unwinding activity was reduced, evident from 40% less Cy5 ssDNA being observed for fork 2b with a methyl phosphonate and phosphorothioate substitution. An abasic site within the duplex region on the translocating strand had the greatest impact on HelQ unwinding activity with a reduction of 50% Cy5 ssDNA liberated when compared to unwinding of fork 2b. These results imply HelQ interacts with both the external backbone and the internal bases of the translocating strand of DNA with the presence of DNA bases being essential.

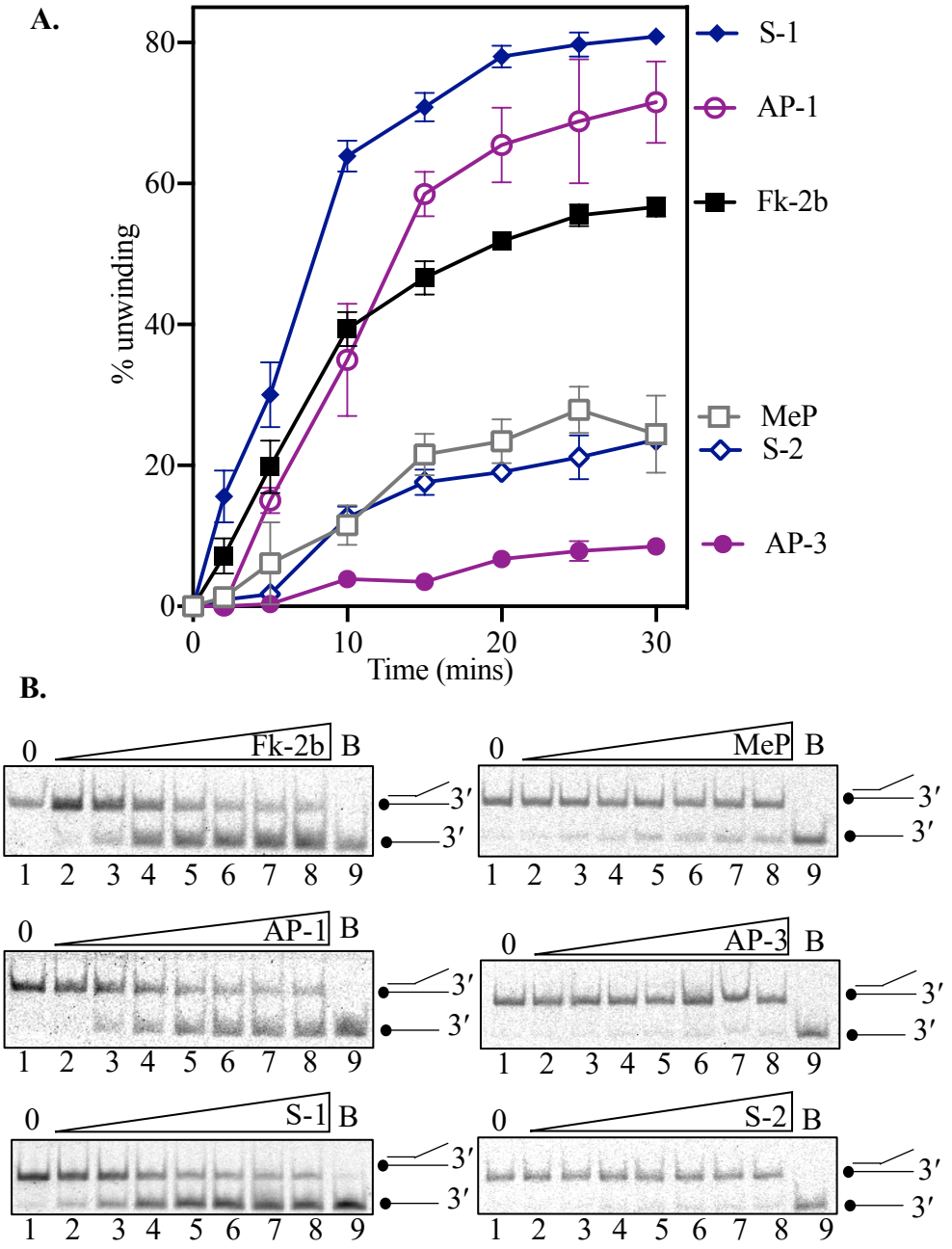


Figure 4.6 Helicase assays as a function of time show FL-HelQ does not unwind forked DNA with a chemical modification within the duplex region. A. A measurement of FL-HelQ (160 nM) unwinding of chemically modified fork 2b DNA (25 nM) over time. **B.** Also shown are representative TBE gels that were carried out in triplicate and used to generate the data as standard error from the mean as presented here. Modifications to the DNA included methyl phosphonate (MeP), phosphorothioate (S) and abasic sites (AP) in the duplex region and ssDNA region of the translocating strand. Reactions were terminated at 0, 2, 5, 10, 15, 20, 25 and 30 minutes. B represents a boil control of dissociated substrate as indicated. HelQ only unwound DNA with modifications in the ssDNA region (AP-1 and S-1).

We next tested DNA substrates containing the same modifications in the 3′-ended ssDNA region of the translocating strand. HelQ was able to unwind these chemically modified substrates similarly, if not better, to the unmodified fork 2b (figure 4.6), indicating that HelQ was unaffected by base or backbone modifications at these positions. These modifications were relatively close to the ssDNA-dsDNA junction at 7 base-pairs from the junction. This suggests that HelQ loads onto the forked DNA close to the ssDNA-dsDNA branchpoint and does not translocate across the modification site because if it did, we would expect to observe at least partial inhibition of fork unwinding. This also suggests that HelQ has a binding footprint of 7 bp or less. A partial improvement in unwinding by HelQ was observed for these chemically modified fork substrates compared to unmodified fork apparent by 10% more Cy5 ssDNA being detected (figure 4.6). A shorter region available for HelQ to load due to the presence of the modification may result in more successful HelQ loading and subsequent fewer 'false starts' ultimately resulting in more unwinding. However, this improvement is not considered significant enough to suggest the modifications improve HelQ unwinding of the forked DNA. Additionally, end point helicase assays as a function of HelQ protein concentration supported these findings for all chemical modifications of forked DNA (figure 4.7).

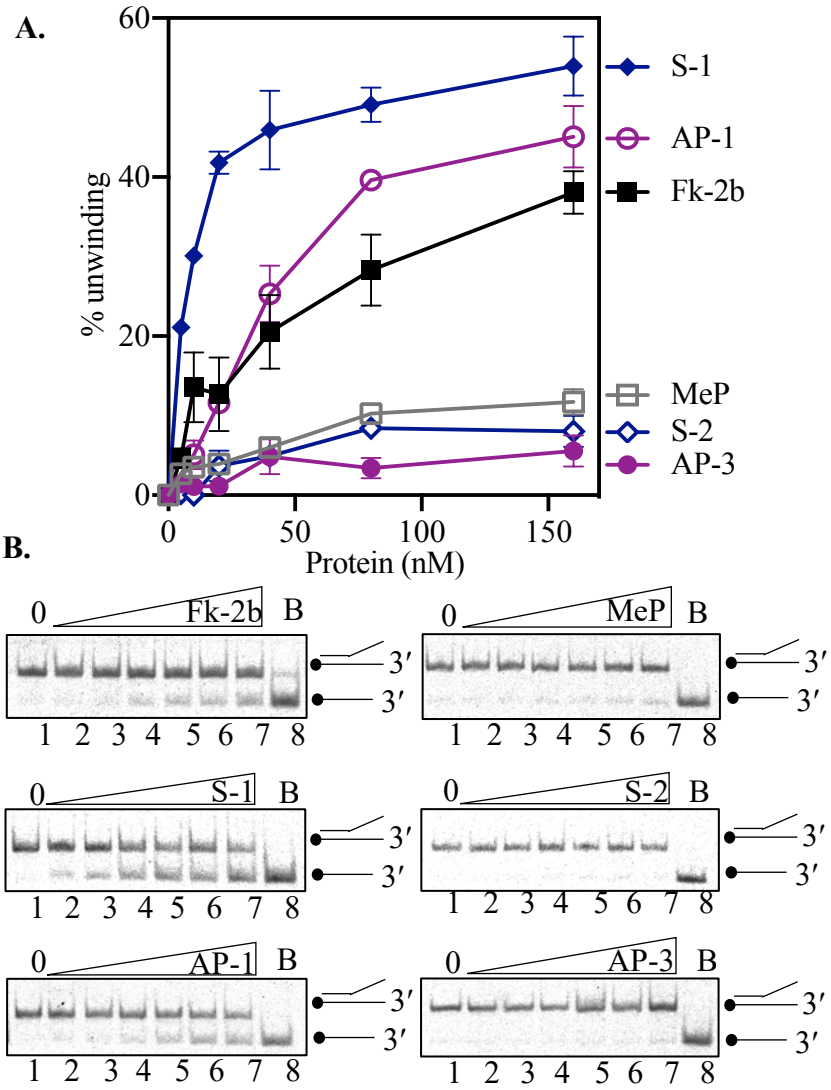


Figure 4.7 Helicase assays as a function of protein concentration show FL-HelQ does not unwind forked DNA with a chemical modification within the duplex region. A. A measurement of unwinding by FL-HelQ of chemically modified fork 2b DNA (25 nM). **B.** Also shown are representative TBE gels that were carried out in triplicate and used to generate the data as standard error from the mean. Modifications to the DNA included methyl phosphonate (MeP), phosphorothioate (S) and abasic sites (AP) in the duplex region and ssDNA region of the translocating strand. HelQ was used at 0, 5, 10, 20, 40, 80 and 160 nM. B represents a boil control to show dissociated substrates as indicated. HelQ unwound DNA with modifications in the ssDNA region (AP-1 and S-1) but not in the duplex region (AP-3, S-2 and MeP).

A similar chemically modified DNA forked substrate with an abasic site within the duplex region of the non-translocating strand was shown to not impact HelQ unwinding activity (reverse abasic, figure 4.5). This substrate was used to assess the location on the DNA that HelQ is likely to load by determining if the abasic site

impacts HelQ activity at this site. The ability of HelQ to unwind this modified fork suggests HelQ interacts predominantly with the translocating strand (figure 4.8 A-B). A reverse methyl phosphonate substrate was made with the methyl phosphonate substitution in the ssDNA region of the non-translocating strand. HelQ was unable to unwind through the chemical modified fork (figure 4.8 A-B). HelQ is likely to load onto both ssDNA regions of the forked structure, however, translocating towards the branch point results in observed unwinding. This implies HelQ does not rely on interactions with the non-translating strand to unwind forked DNA. Therefore, an inability to unwind through the reverse methyl phosphonate substrate may be as a result of a change in the conformation of the DNA caused by the modification which may prevent HelQ interacting efficiently with the translocating strand.

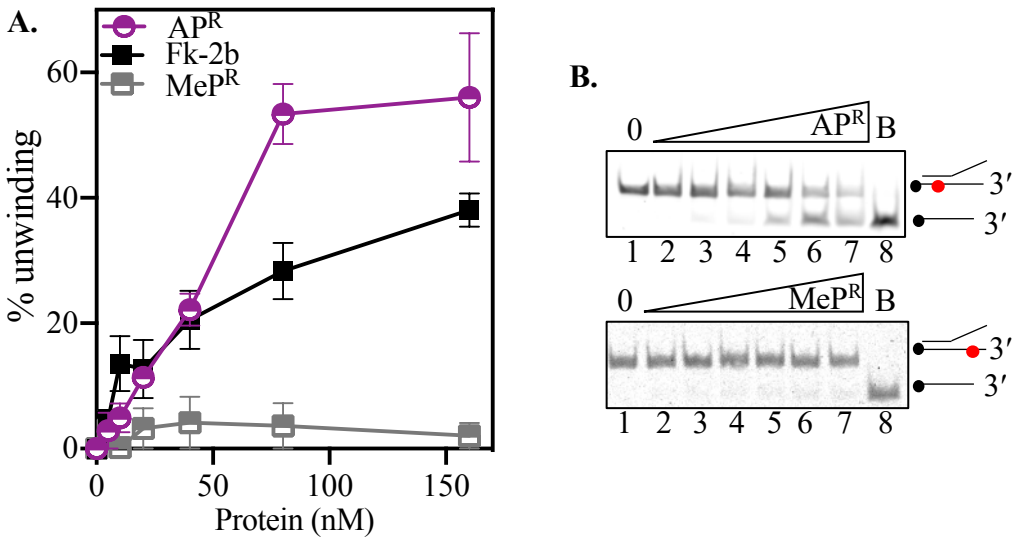


Figure 4.8 Helicase assays as a function of protein concentration show FL-HelQ is able to unwind DNA with an abasic site in the duplex region of the non-translocating strand. A. A measurement of HelQ, used at 0, 5, 10, 20, 40, 80 and 160 nM, unwinding of chemically modified fork 2b DNA (25 nM). **B.** Also shown are representative TBE gels carried out in triplicate used to generate the data as standard error from the mean for each modified fork. DNA modifications included methyl phosphonate (MeP^R) and abasic (AP^R) on the non-translocating strand of the fork (red dot). B represents a boil control to show dissociated substrates as indicated. HelQ unwound DNA with an abasic site on the non-translocating strand but not a methyl phosphonate substitution.

In summary, HelQ requires the internal and external DNA elements on the translocating strand of DNA to load and unwind the duplex. DNA bases in the non-translocating strand are not essential for helicase activity.

4.5 FL-HelQ forms stable complexes with chemically modified DNA.

The DNA binding activity of HelQ with chemical modified DNA substrates was measured using EMSAs to assess the impact of DNA modifications on HelQ DNA loading. The chemically modified DNA substrates did not impact the ability of HelQ to bind with the DNA when compared to binding to the unmodified fork 2b (figure 4.9 A). EMSAs showed complete shifting of the DNA substrate in the presence of HelQ to form HelQ-DNA complexes (figure 4.9 B). Therefore, the impact of modifications in the DNA on HelQ unwinding activity is not due to HelQ loading but likely because of translocation ability. This was expected because HelQ would be able to bind other regions of ssDNA. HelQ was also able to load onto the ssDNA oligos with the chemical modifications (figure 4.10 A-B).

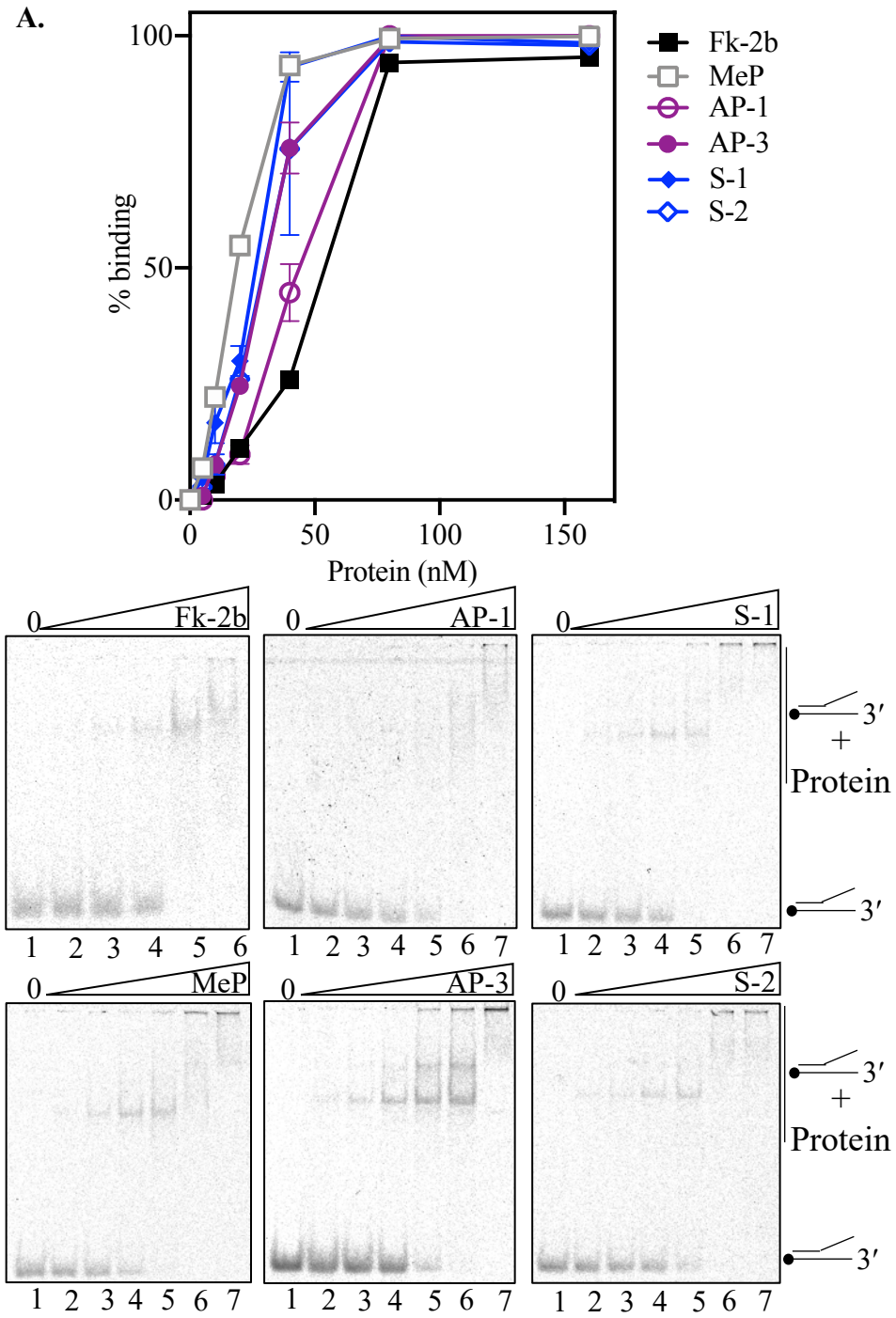


Figure 4.9 EMSAs show FL-HelQ is able to load onto chemically modified fork DNA. A. A measurement of FL-HelQ binding to forked DNA with chemical modifications (25 nM). **B.** Also shown are TBE EMSAs representations carried out in triplicate out as a function of HelQ concentration used at 0, 5, 10, 20, 40, 80 and 160 nM. DNA modifications included methyl phosphonate (MeP), phosphorothioate (S) and abasic sites (AP) in the duplex region and ssDNA region of the translocating strand.

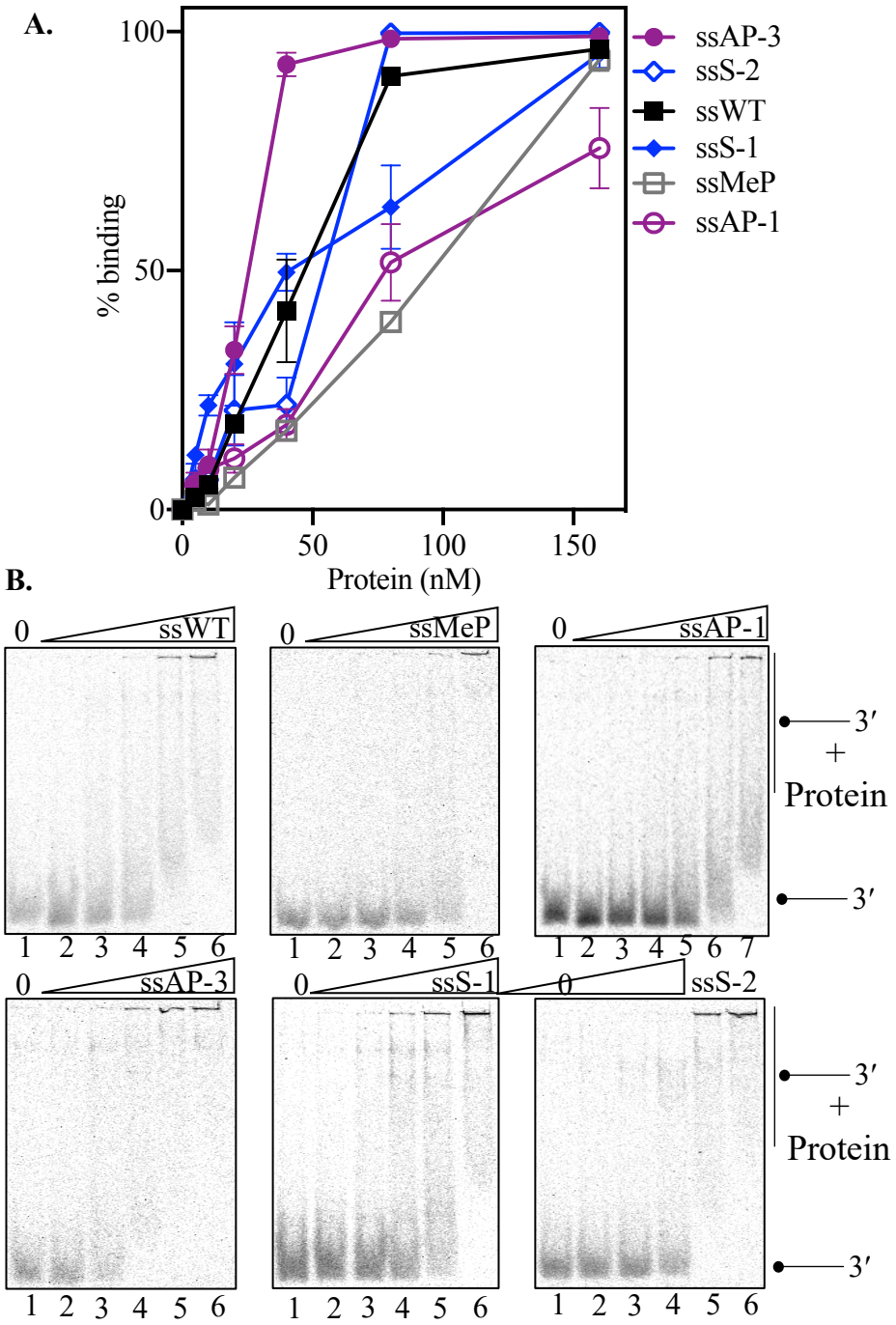


Figure 4.10 EMSAs show FL-HelQ is able to load onto chemically modified ssDNA. A. A measurement of FL-HelQ binding to ssDNA with chemical modifications (25 nM). **B.** Also shown are representative TBE EMSA gels carried out in triplicate carried out as a function of HelQ concentration used at 0, 5, 10, 20, 40, 80 and 160 nM. DNA modifications included methyl phosphonate (ssMeP), phosphorothioate (ssS) and abasic sites (ssAP) in the duplex region and ssDNA region of the translocating strand. HelQ binding was not impaired by the presence of DNA modifications.

4.6 Analysis of the ATPase activity of FL-HelQ.

4.5.1 Chemically modified DNA did not impact ATPase activity of HelQ.

Malachite green reporter assays were used to measure the ATPase activity of HelQ when confronted with chemically modified DNA in comparison with unmodified fork 2b as described in section 2.16.5. ATPase activity was measured by converting absorbance (OD_{620}) into nM ATP hydrolysed per nM HelQ per minute using a standard curve of free phosphate. HelQ unwinding of fork 2b resulted in 20 nM ATP/ nM HelQ/ minute ATPase activity (figure 4.11). The ATPase activity of HelQ remained similar when unwinding the chemically modified forked substrates ranging between 20 and 40 nM ATP/ nM HelQ/ minute (figure 4.11). The presence of an abasic site, methyl phosphonate or phosphorothioate substitution did not impact the ability of HelQ to hydrolyse ATP. Therefore, DNA modifications do not appear to prevent HelQ from loading and attempting to unwind the substrate. The inability of HelQ to unwind the substrate is due to disruptions in the interactions between DNA and HelQ during translocation.

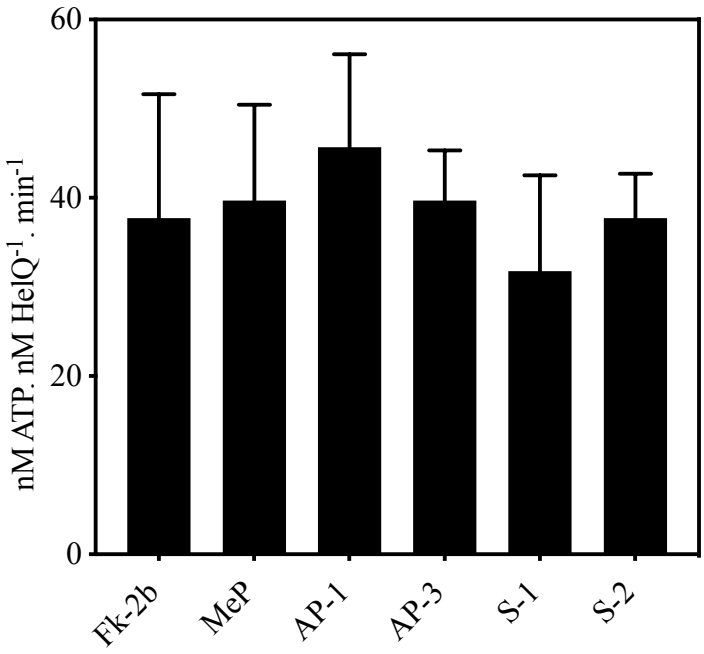


Figure 4.11 ATPase assays show FL-HelQ remains active when bound to chemically modified forked DNA. A measurement of the ATPase activity of FL-HelQ (80 nM) with chemically modified fork 2b (25 nM) using malachite green reporter assays. ATPase activity measured in triplicate and reported here as nM ATP hydrolysed per nM FL-HelQ per minute using a standard curve of free phosphate (y = 0.157x). HelQ activity was tested with unmodified fork 2b (Fk-2b), methyl phosphonate (MeP), abasic 1 (AP-1), abasic 3 (AP-3), phosphorothioate 1 (S-1) and phosphorothioate 2 (S-2). ATPase activity remained constant for all substrates.

These results confirm that unwinding observed in described assays is due to active HelQ and not substrate dissociation upon HelQ binding.

4.7 Chemical crosslinking negatively impacts FL-HelQ activity.

We hypothesis that HelQ forms stable, active dimers (section 3.8.4). HelQ protein was crosslinked by incubation with 0.2% glutaraldehyde in an attempt to stabilise the protein (section 2.18.7). FL-HelQ has proved to be highly unstable making it a difficult protein to assess structurally. We first tested the impact of crosslinking HelQ on the unwinding activity. It is likely that crosslinking would produce an assortment of HelQ species which would not have been limited to the proposed dimeric conformation. Crosslinking was aimed to stabilise HelQ in these

various conformations and complexes to ease structural analysis. However, crosslinked HelQ was unable to unwind fork 2b DNA and loading onto the fork was limited in comparison with unmodified HelQ (figure 4.12 A-B). While crosslinking may have stabilised HelQ oligomers, a loss of unwinding activity implied crosslinking removed protein flexibility required for activity. Therefore, this crosslinked protein was not taken further in structural studies because it did not produce a representative active HelQ protein. However, this data suggests that crosslinking increased the rigidity of HelQ preventing minute conformational changes that would allow protein 'morphing' around the DNA. Further implying flexibility is important for HelQ activity. Furthermore, crosslinking may have resulted in protein aggregation or degradation accounting for this reduced activity. A necessity for flexibility in HelQ is unsurprising because of the nature of intrinsically disordered proteins which require fluidity for activity.

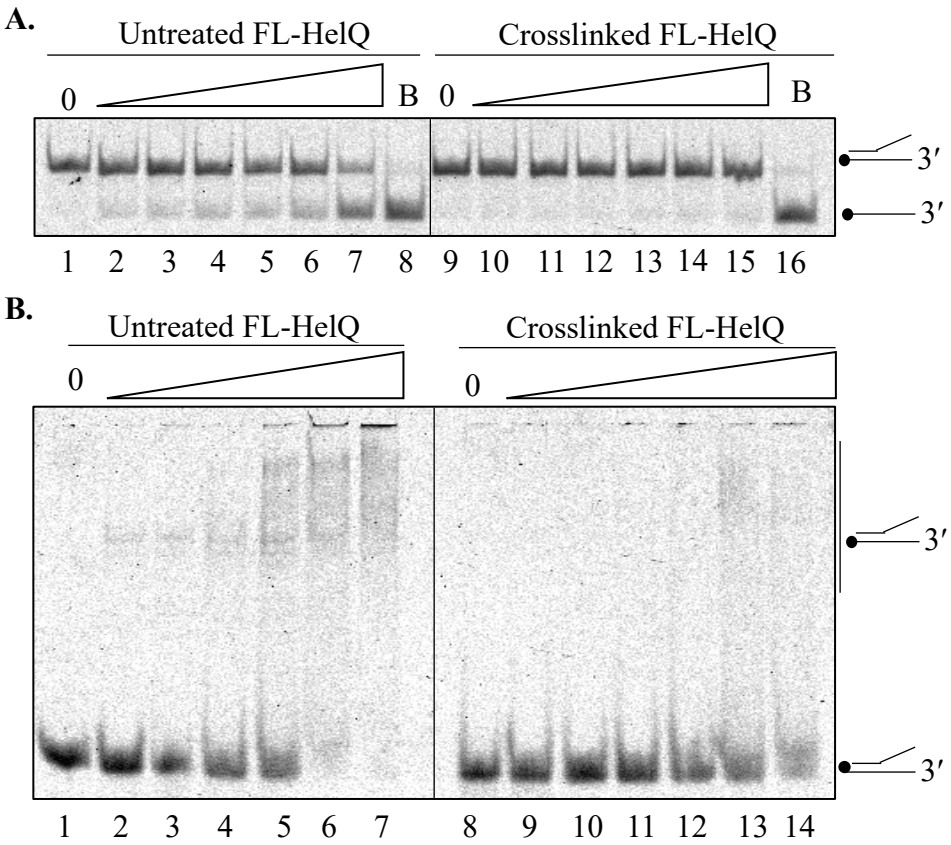


Figure 4.12 Crosslinked HelQ is inactive for DNA binding and unwinding. **A.** Representative TBE gels showing unwinding of fork 2b (25 nM) by native FL-HelQ compared with glutaraldehyde crosslinked HelQ. Protein was used at 0, 5, 10, 20, 40, 80 and 160 nM. B represents a boil control showing dissociated substrate as indicated by schematics (filled dot shows Cy5 end label). **B.** Representative TBE EMSA gels showing native FL-HelQ and crosslinked HelQ loading onto fork 2b (25 nM). Crosslinking inhibits unwinding and impairs DNA binding.

4.8 HelQ is unable to unwind through different DNA intermediate structures.

HelQ unwinding and binding activity was assessed with different DNA structures (table 2.4). This was used to measure the preference of HelQ for different substrates and determine what structures HelQ was able to interact with and unwind. Forked substrates are found physiologically during replication and DNA damage when DNA becomes exposed or resected.[157,216] We next tested other physiologically relevant intermediate DNA structures associated with repair and

replication *in vitro*. Analysis of HelQ with these structures provides evidence for specific roles of HelQ in DNA processing.

G-quadruplex (G4) structures occur in nature when interactions between four guanine bases, stabilised by a central cation, form a compact square planar arrangement (figure 4.13 A).[343] There are 375,000 known DNA sequences that form G4 structures in the human genome and are involved in activating and repressing a range of DNA activities, some examples are given here next.[65] It is hypothesised that G4 substrates sequester the 3′ end of telomeres to prevent telomerase extension making them essential for telomere protection.[343] They are also implicated in activating transcription initiation, translation and splicing. Therefore, in order for these processes to occur, efficient and controlled formation of these structures, as well as the unwinding of them, is essential to maintain stability of DNA processing activities and replication progression. Furthermore, failure to do so can lead to disease. Examples of such helicases that regulate G4 structures include SF1 Pif1 and SF2 RecQ, FancJ, BLM and WRN.[344]

HelQ was unable to unwind the synthetic G4 structure evident from no dissociation of the oligos on addition of HelQ (figure 4.13 B, lanes 8-14). RecQ, used as a positive control, was able to unwind the substrate completely (lanes 1-7). These results suggest HelQ is unlikely to be involved in G4 maintenance.

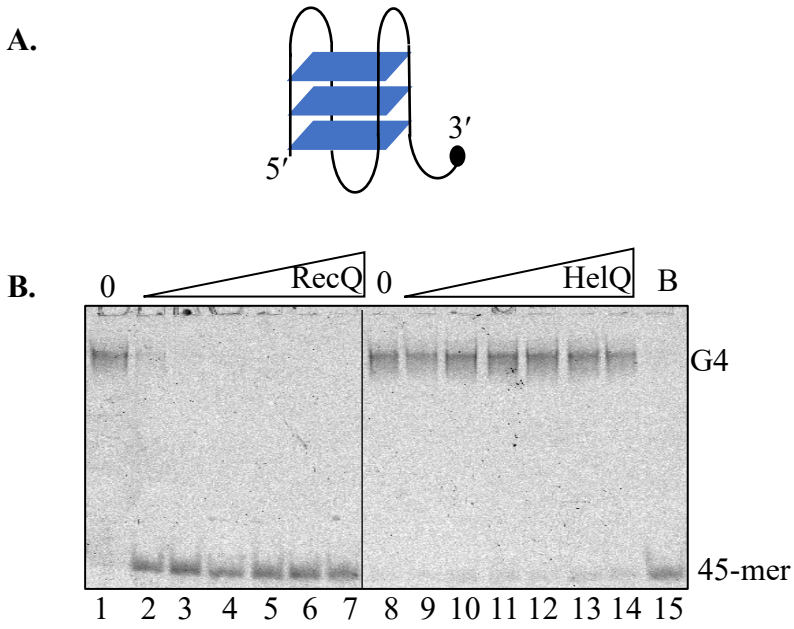


Figure 4.13 Helicase assays show HelQ does not unwind through a G4 quadruplex DNA structure. A. Cartoon schematic of the assembled 45 base oligonucleotide in the synthetic G4 quadruplex structure. Filled dot indicates 5′ Cy5 label. **B.** Representative TBE gel showing HelQ and *E. coli* RecQ unwinding of G4 DNA (25 nM). RecQ acts as a positive control unwinding the DNA substrate. Proteins were used at 0, 5,10, 20, 40, 80 and 160 nM and B represents boil control showing fully dissociated DNA strands (45-mer).

A Holliday junction (HJ), another physiological DNA structure, is a four-stranded DNA intermediate that forms during late stages of homologous recombination (figure 4.14 A). The spatial arrangement of two dsDNA molecules in the HJ allows DNA crossing over that is central to several physiological processes, including DNA repair or strand breaks and exchange of genetic material during meiosis.[138,198] In these contexts, resolution of HJs is required to ensure completion of DNA repair for resumption of DNA replication, or for chromosome disjunction during mitosis. HJ resolvases include Gen1, an XPG/Rad2 family protein of structure-selective endonucleases.[345] Gen1 cleaves the HJ recombination intermediates in order to separate the cross shaped structure. The bacterial analogue is RuvC.[345] RuvC is part of the RuvABC complex that mediates branch migration by resolving Holiday junctions crucial to bacterial DNA repair. The combined activity of the all three proteins: DNA binding RuvA, ATPase RuvB and resolvase

RuvC allows for the directional resolution of a HJ at specific sequences.[346,347] Similarly, helicases have also been implicated in HJ resolution. The *E. coli* helicase RecG has been implicated in the catalysis of nascent strand regression at stalled replication forks forming HJs and it is proposed helicases with similar activities may also be necessary for this activity.[348,349] Therefore, the association of helicases in joint molecules, as well as previous publications implicating HelQ in HR through interactions with Rad51 and the paralogues, (see chapter 1)[236] suggests HelQ as a potential candidate for HJ resolution.

HelQ was unable to unwind a synthetic HJ (figure 4.14 A-B) but was able to bind a HJ (figure 4.14 C). Therefore, the lack of unwinding by HelQ was not caused by reduced loading of HelQ onto the DNA. This suggests HelQ is unlikely to be a HJ resolution helicase involved in late-stage HR.

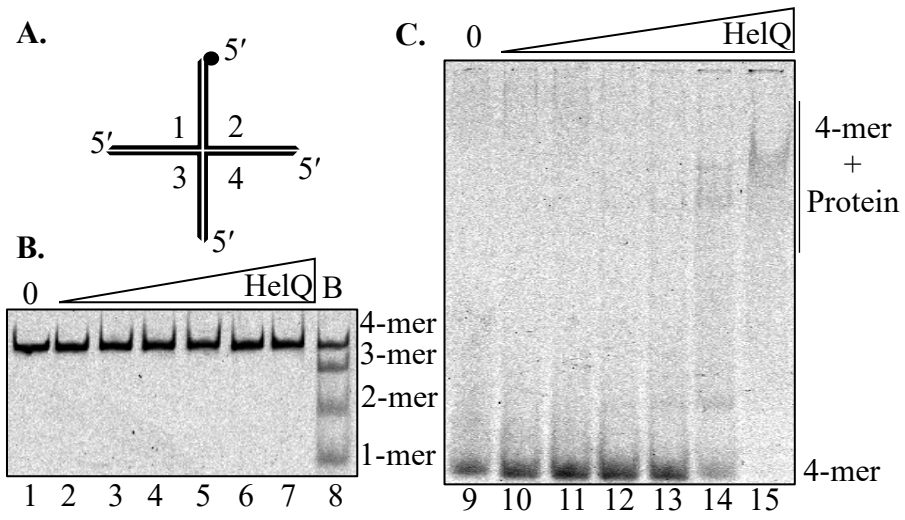


Figure 4.14 Helicase assays show FL-HelQ is unable to unwind a synthetic Holliday junction (HJ). A. Cartoon schematic of the four oligonucleotides (1-4) assembled into a J6 HJ structure. Filled dot indicates 5′ Cy5 label. **B.** Representative TBE gels showing HelQ is unable to unwind a J6 HJ (25 nM). HelQ was used at 0, 5, 10, 20, 40, 80 and 160 nM. B represents a boil control indicating strand separation to 3 strands (3-mer), 2 strands (2-mer) and complete dissociation (1-mer). **C.** Representative TBE EMSA gel showing HelQ binding to J6 (25 nM).

Other DNA intermediate structures that we tested for HelQ unwinding activity included an extended fork substrate (figure 4.15 A) and dsDNA (figure 4.15 B). The extended fork consisted of 80 base pairs (bp) with a longer duplex region of 52 bp compared to fork 2b with a 25 bp duplex region and ssDNA region. HelQ was unable to unwind the extended fork substrate evident from a lack of free Cy5 ssDNA being observed with addition of HelQ (figure 4.15 C). HelQ was also able to load onto the substrate suggesting HelQ attempted to unwind the fork (figure 4.15 E) Unwinding by HelQ of the DNA may occur, however, is not completed which would result in strand separation being observed. This suggests that HelQ is falling off the DNA substrate prior to reaching the duplex end which implies HelQ is not a very processive enzyme and targets small regions of duplex DNA for unwinding. Reduced processivity would be confirmed by testing additional substrates of varying lengths.

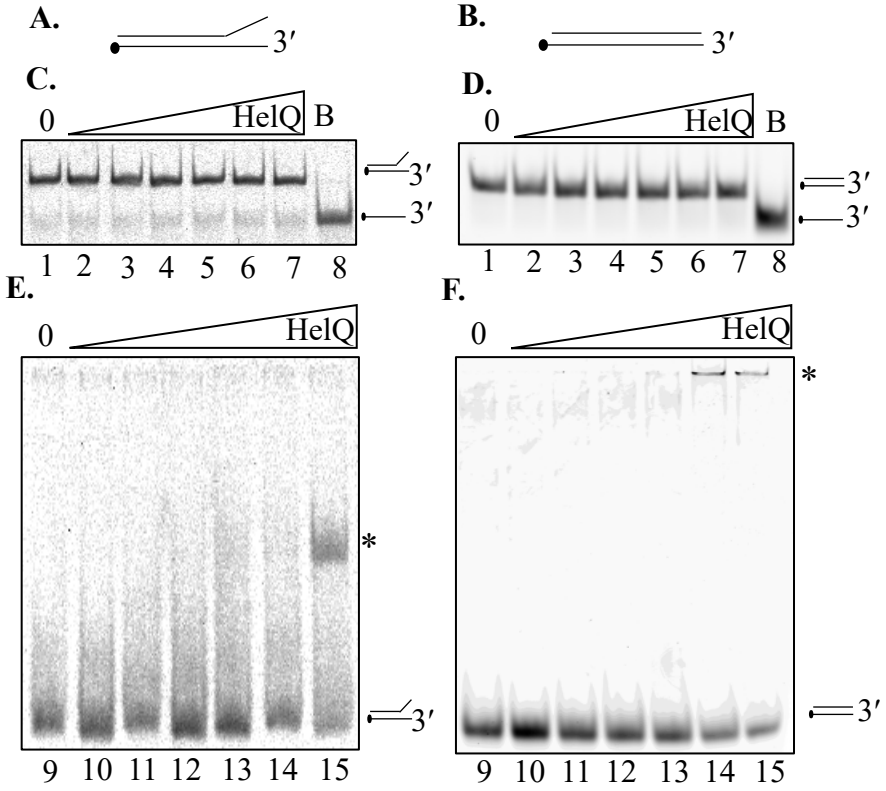


Figure 4.15 FL-HelQ does not unwind an extended fork structure or dsDNA. Cartoon schematic of **A.** an extended fork duplex and **B**. dsDNA tested in HelQ unwinding assays. Filled dots indicate 5′ Cy5 label. TBE representation of FL-HelQ unwinding of **C.** extended fork (25 nM) and **D.** dsDNA (25 nM) as a function of protein concentration. FL-HelQ was used at 0, 5, 10, 20, 40, 80 and 160 nM. B represents a boil control to show fully dissociated substrates. TBE EMSA representation of HelQ binding to the **E.** extended fork structure **F.** and dsDNA.

HelQ was also unable to unwind the dsDNA substrate consisting of 50 base pairs (figure 4.15 D). This was expected because HelQ requires a free 3′ ssDNA region to load onto and translocate along. EMSA analysis showed HelQ was also unable to bind dsDNA as expected (figure 4.15 E-F). Some protein-DNA aggregation was observed at high protein concentrations. Therefore, the inability of HelQ to unwind the substrate was likely due to the inability of HelQ to load onto a free 3′ ssDNA region.

HelQ was unable to unwind the synthetic DNA intermediate structures tested here (G4, HJ, extended fork and dsDNA). This implies that HelQ likely favours DNA structures with a ssDNA-dsDNA junction, as in fork 2b. HelQ requires loading to a free 3′ ssDNA of which G4, HJ and dsDNA did not have and therefore a lack of HelQ activity is likely because HelQ was unable to efficiently load onto the DNA. This is also supported by evidence previously published.[256] The inability of HelQ to process the extended fork structure may suggest limited processivity or the conditions were not favourable.

4.9 HelQ does not function as a re-annealing helicase.

HelQ was unable to bypass chemical modifications or synthetic intermediate DNA structures. An alternative role assessed for HelQ was as an end-joining helicase. PolQ, which includes a C-terminal DNA polymerase and an N-terminal SF2 HelQ-type DNA helicase, assists in alternative end-joining to suppress HR (chapter 1). This end-joining activity is mediated by the N-terminal helicase domain to join DNA ends.[250,350] Sequence conservation with HelQ implicates HelQ as an end-joining helicase which was assessed here as described in section 2.16.2. In addition, the eukaryotic RecQL4 helicase is involved in end-joining which is recruited to laser-induced dsDNA breaks and involved in non-homologous end

joining.[351,352] Further links between HelQ and homologous DNA repair helicases implicates HelQ as a potential player in end-joining.

A role in DNA end-joining by HelQ was tested by assessing if base pairing between two complementary ssDNA oligonucleotides into dsDNA could be promoted by HelQ (figure 4.16 A). Additionally, the ability of FL-HelQ to anneal DNA substrates with complimentary over-hangs was tested. Two pairs of DNA oligonucleotides were designed, one with 5′ complimentary base over-hangs and a second with 3′ overhangs (figure 4.16 B and C). One of each pair was Cy5 end labelled to detect oligo extension. Assay variations were used in an attempt to optimise re-annealing activity, these included the addition of magnesium, EDTA and ATP at room temperature, 30^{o}C and 37^{o}C (figure 4.16).

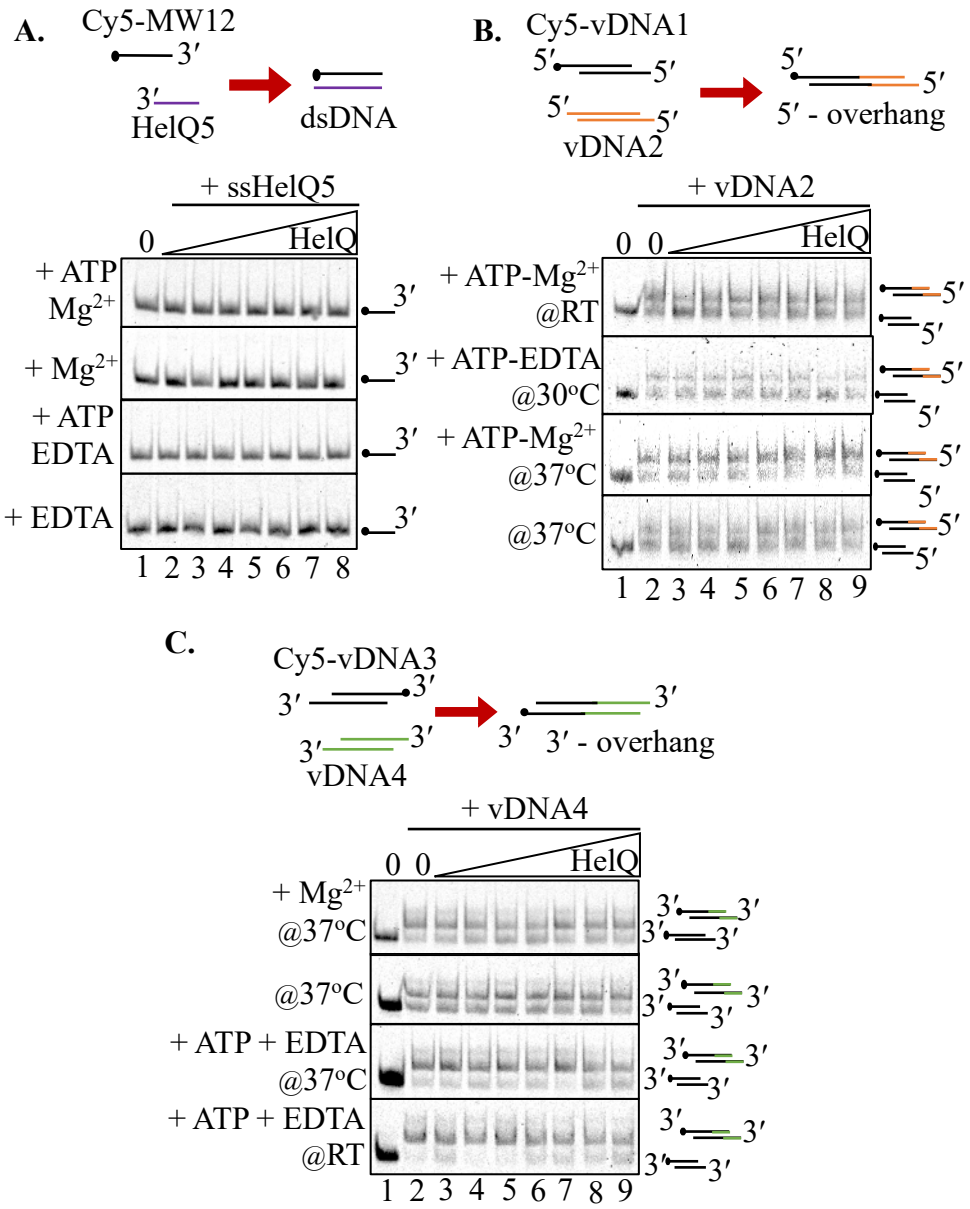


Figure 4.16 FL-HelQ does not re-anneal ssDNA or DNA with complimentary overhangs. Representative TBE gels to show HelQ is not active in the annealing of substrate combinations. **A.** Complimentary ssDNA oligonucleotides (cy5-MW14 and HelQ5), **B.** dsDNA with complimentary 5′-overhangs (Cy5-vDNA1 and vDNA2) and **C.** dsDNA with complimentary 3′-overhangs (Cy5-vDNA3 and vDNA4). Schematics shown above and to the right of gel panels indicate substrate formation upon annealing. HelQ was used at 0, 5, 10, 20, 40, 80 and 160 nM. Varying conditions tested as indicated included addition of 5 mM ATP, $MgCl_2$ and EDTA at room temperature (RT), 30°C and 37°C. Spontaneous self-annealing was observed for over-hang DNA.

FL-HelQ was unable to anneal any of the complimentary substrates tested. The formation of extended substrates observed was due to spontaneous annealing which

is evident from annealing occurring in the absence of HelQ. This data suggests HelQ is unlikely to be involved in end-joining to prevent HR.

4.10 HelQ cannot displace protein barriers during translocation.

Helicases are also found to be involved in protein displacement activity at stalled replication forks.[217,353] Therefore, HelQ was assessed for its ability to bypass a physical roadblock on DNA (section 2.17). Replication does not occur uninterrupted throughout the genome but encounters multiple types of barriers including both DNA sequences that from secondary structures and bound protein complexes e.g. transcribing DNA polmyerase.[354] Protein barriers also arise at sites of DNA damage, for example nicks in the DNA or bulky adducts which physically block protein translocation. As the replication fork encounters the nucleoprotein barrier it can induce barrier dissociation with no impact on replication or the barrier can stall the replisome.[217] If barriers remain, replication is prevented from continuing and can induce genome instability and fork collapse.[354–356] Fork stalling results in check point activation. This can occur by the replicative helicase Mcm2-7 continuing to unwind DNA exposing larger regions of ssDNA.[357,358] This stimulates RPA binding to signal damage response for recruitment of repair proteins to displace the protein block. These include helicases which are involved in fork clearance by unwinding through the barrier site and displacing the protein roadblock to provide access for repair proteins.[357] In these circumstances accessory helicases are involved in displacing the barrier (see chapter 1).[217] The ability of HelQ to unwind through different protein barriers on the DNA was considered as a role for HelQ in fork clearance.

4.10.1 BamHI^{EIIIA} as a protein roadblock.

The type II restriction endonuclease from *Bacillus amyloliquefaciens,* BamHI, binds and cuts a specific duplex recognition site (5′-GGATCC).[359] BamHI^{EIIIA} has a point mutation that prevents the endonuclease from cutting the DNA making it ideal as a physical barrier bound on DNA (figure 4.17 A). BamHI^{EIIIA} was able to bind to a DNA fork substrate containing the recognition site within the duplex region (figure 4.17 C). Therefore, this was used as a suitable model assay system to determine if DNA unwinding by HelQ alone was sufficient to remove the bound BamHI protein.

HelQ was able to unwind 50% of the fork substrate in the absence BamHI^{EIIIA} (figure 4.17 B). The ability of HelQ to unwind the fork substrate in end point assays in the presence of varying concentrations of BamHI^{EIIIA} was then analysed (figure 4.28 D). HelQ was unable to unwind the fork DNA in the presence of BamHI^{EIIIA} at the higher concentrations, 80 and 160 nM (figure 4.17 E lower panel), that correspond to conditions in which BamHI^{EIIIA} bound to100% of the DNA substrate (figure 4.17 C). Percentage unwinding by HelQ was observed to be reduced to less than 10% in the presence of 160 nM BamHI^{EIIIA} roadblock. HelQ was able to unwind the fork substrate in the presence of lower concentrations of BamHI^{EIIIA}, 10 and 20 nM (figure 4.17 E upper panel), where full shifting was not observed. Therefore, evident from EMSAs, unwinding observed is likely of naked fork DNA (figure 4.17 C).

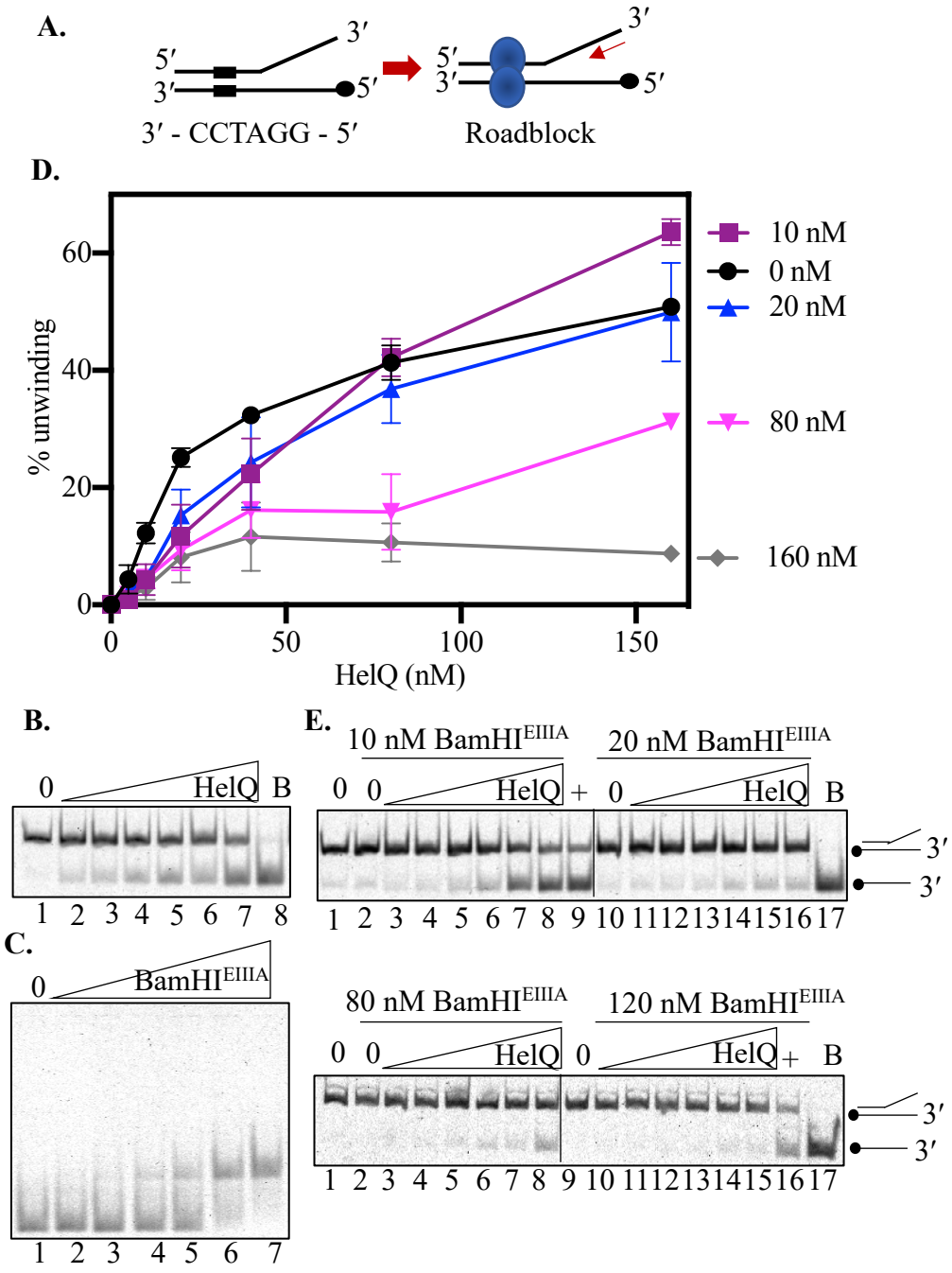


Figure 4.17 Protein displacement assays show FL-HelQ does not unwind DNA through a BamHI^{EIIIA} non-specific protein barrier. A. Cartoon schematic of the assembly of the BamHI^{EIIIA} protein roadblock bound to fork DNA at the cognate hexanucleotide site. Representative TBE gels showing the following: **B.** HelQ unwinding the naked DNA fork (25 nM) and **C.** BamHI^{EIIIA} binding the fork (25 nM). Proteins were used at 0, 5, 10, 20, 40, 80 and 160 nM. B represents a boil control to show dissociated substrates as indicated in schematics with filled dots indicating 5′ Cy5 label. **D.** A measurement to show HelQ is unable to unwind DNA (25 nM) in the presence of BamHI^{EIIIA} barrier. **E.** Also shown are representative TBE gels carried out in triplicate used to produce the graph as standard error from the mean.

4.10.2 Streptavidin-biotin as a protein roadblock.

Additional roadblocks were tested to further explore the notion of whether or not HelQ can displace DNA-bound barriers. Streptavidin, at 55 kDa, binds to biotin on the DNA forming a bulkier adduct when compared with the BamHI^{EIIIA} barrier at only 24.5 kDa.[360,361] The sizes of the adducts are different with streptavidin averaging at 54x58x48 Å compared to the smaller 45x35x25 Å of BamHI.[359,361] The binding affinity of streptavidin to biotin is very high with a K_d value of 10^{-15} M compared to 5.6^{-10} M for BamHI.[361,362] Streptavidin was therefore used to test HelQ displacement activity of a bulkier, larger and more strongly bound roadblock. DNA unwinding using a streptavidin-bound oligonucleotide was assessed using previous described methods.[363] A DNA fork substrate with biotinylated bases within the duplex region on the translocating and non-translocating strand was designed to create a scaffold for the streptavidin barrier (figure 4.18 A). EMSAs showed the formation of stable streptavidin-biotinylated fork DNA complexes (figure 4.18 C) and the concentration of free biotin required to bind displaced streptavidin (figure 4.18 D). However, HelQ was unable to unwind the biotinylated fork DNA in the absence of the roadblock (figure 4.18 B) suggesting the presence of biotin alone in the DNA inhibited HelQ unwinding activity.

While HelQ was unable to unwind the fork in the absence the roadblock (figure 4.18 E), HelQ was incubated with the DNA-roadblock complex to determine if HelQ can displace the streptavidin. Streptavidin was not displaced from the forked DNA on addition of HelQ apparent from no DNA shifting from a streptavidin bound to unbound state (figure 4.18 E). The streptavidin interaction with biotinylated DNA appeared to be very strong and remained even after treatment with proteinase K. This may suggest that the interaction was too strong in order for HelQ to have any impact.

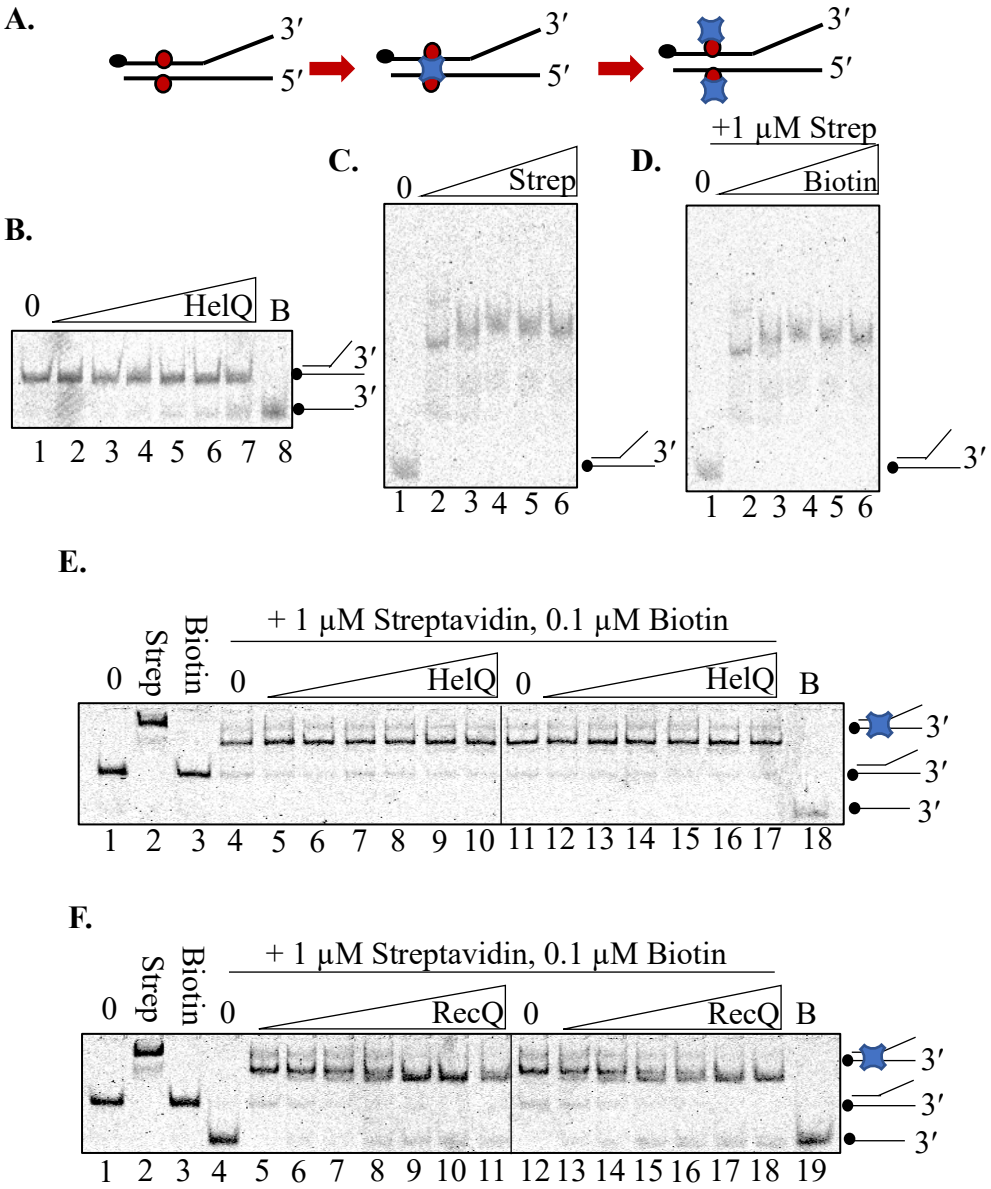


Figure 4.18 Protein displacement assay shows FL-HelQ does not unwind DNA through a streptavidin-biotin roadblock. A. Cartoon schematic of streptavidin (blue) assembly onto biotinylated DNA fork (red dots) as multiple complexes. Representative TBE gels showing the following: **B.** HelQ cannot unwind the biotinylated fork (25 nM) in the absence of streptavidin; **C.** streptavidin, used at 0, 0.1, 1, 2 and 3 µM, forms complexes on biotinylated-DNA and **D.** free biotin was optimised to 0.1 µM to saturate unbound streptavidin. **E.** Representative TBE gels showing HelQ, used at 0, 5, 10, 20, 40, 80 and 160 nM, cannot unwind the fork substrate (25 nM) in the presence of streptavidin (1 µM). B represents a boil control as indicated by schematics with filled dots indicating 5′ Cy5 labels. Streptavidin remained bound to the fork. **F.** Similarly, RecQ was unable to displace streptavidin and unwind the substrate.

A range of additional helicase proteins were tested for displacement activity of streptavidin from the biotinylated-DNA, including RecQ (figure 4.18 F), RecG,

PriA and Rep, however, displacement was not observed for any. These results suggest the streptavidin-biotin roadblock was too strong to be used in these assays and therefore may not be a reliable measure for the ability of HelQ to translocate through a protein barrier.

4.10.4 RNA polymerase as a protein roadblock to bacterial replisome DNA replication.

The ability of HelQ to translocate through the *E. coli* RNA polymerase to recover DNA replication by the replisome was tested. The *E. coli* helicase, Rep, was used as a positive control that is able to displace RNA polymerase from circular DNA to allow the replisome to continue replication.[364] The replisome (DNA polymerase III, Tau, DnaB, DnaC, SSB, β, HU, DnaB, DnaA) was assembled onto circular plasmid DNA and replication was initiated by addition of NTPs, ATP and magnesium (figure 4.19 lane 1). RNAP was added to the reaction to create a roadblock which stalled DNA replication. Stalled replication was apparent from reduced replication products being observed in denatured gels (lane 2). Replication was not recovered with addition of HelQ, DNA fragments suggesting stalled and incomplete replication were evident (lanes 4-6). Therefore, HelQ did not displace RNA polymerase from circular DNA. Rep was able to recover replication due to its ability to displace RNAP (lane 3).

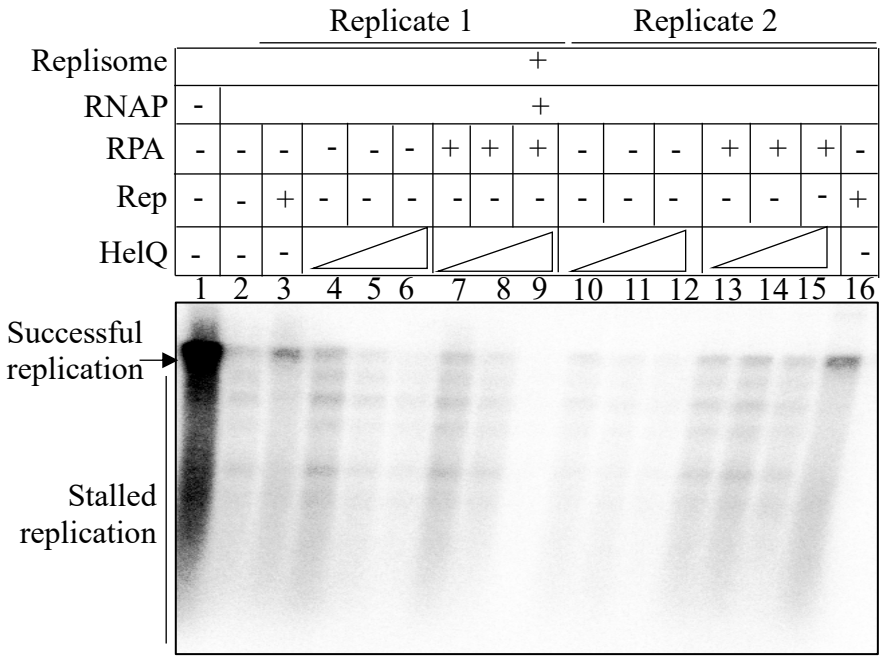


Figure 4.19 FL-HelQ is unable to displace RNA polymerase to stimulate replication restart. Representative denaturing agarose gel showing the assembly of the *E. coli* replisome and RNA polymerase (RNAP) on circular DNA (2 nM). The replisome successfully replicated the DNA (lane 1). Addition of RNAP halted replication by the replisome resulting in DNA fragmentation (lane 2). The positive control Rep (100 nM) displaced RNAP recovering replication (lane 3). Replication was not recovered with addition of HelQ at 25, 50 and 100 nM (lanes 4-6). Addition of RPA (20 nM, lanes 7-9) did not aid HelQ to displace RNAP (see chapter 5). Assay is in duplicate (Lanes 10-16).

4.10.5 Cas9 (CRISPR associated protein 9) as a protein roadblock.

DNA binding by BamHI^{EIIIA}, streptavidin, RNA polymerase and the replisome inhibited DNA unwinding by HelQ. Next, the CRISPR editing enzyme Cas9 was tested as a potential barrier on DNA because of its interactions with proteins in the FA repair pathway and subsequent link to HelQ (see below). Cas9, a CRISPR protein involved in the immunological defence in bacteria against DNA viruses, is an effector of adaptive immunity.[365] Cas9 recognises and loads onto specific DNA sequences, known as PAM sites, and then cleaves specific sequences based on DNA complimentary to guide RNA (gRNA) used to target the nuclease activity of Cas9. This gRNA directs the insertion or deletion of residues in the process of RNA

editing.[365] The Cas9 nuclease has been previously shown to be dislodged by translocating RNA polymerases.[366]

CRISPR-Cas genome editing is used to create targeted dsDNA breaks which are repaired by cellular pathways, specifically single-strand template repair (SSTR).[367] Cas9-induced SSTR requires the FA pathway, important in ICL repair, to direct repair from NHEJ to SSTR. The FancD2 helicase localises to Cas9-induced DSBs implicating FancD2 in regulating genome editing. HelQ is an interacting factor of the FA repair pathway that plays a role in SSTR as part of a FA subcomplex required for genome stability. In the same study, knockdown of HelQ markedly reduced SSTR creating a potential link between HelQ and Cas9-induced genome editing worth exploring.[367] We initially wanted to determine if HelQ was able to process the Cas9 R-loop interference complex by displacing Cas9.

A nuclease inactivated Cas9 protein, dCas9, was used to assess the protein as an R-loop complex barrier on DNA because it was able to load onto specific sequences but did not cut the DNA. This created a physical barrier to HelQ activity without impairing the DNA. dCas9 was inactivated by a double residue substitution Asp-10-Ala and His-840-Ala and purified from *Streptococcus pyogenes* by a student in the Bolt group, Chun Hang Lau.

A forked DNA substrate was designed for use in these assays that contained a Cas9 binding site, a PAM recognition site and complimentary DNA sequence to the gRNA (figure 4.20 A). HelQ was able to unwind the DNA fork in the absence of dCas9 in both HelQ optimised buffer (helicase buffer) and Cas9 optimised binding buffer (Cas9 binding buffer) (figure 4.21 B). However, dCas9 was only able to load and assemble the R-loop in Cas9 binding buffer (figure 4.20 C-D). Therefore, experiments were carried out in Cas9 binding buffer to ensure the formation of R-loops to test for HelQ displacement activity.

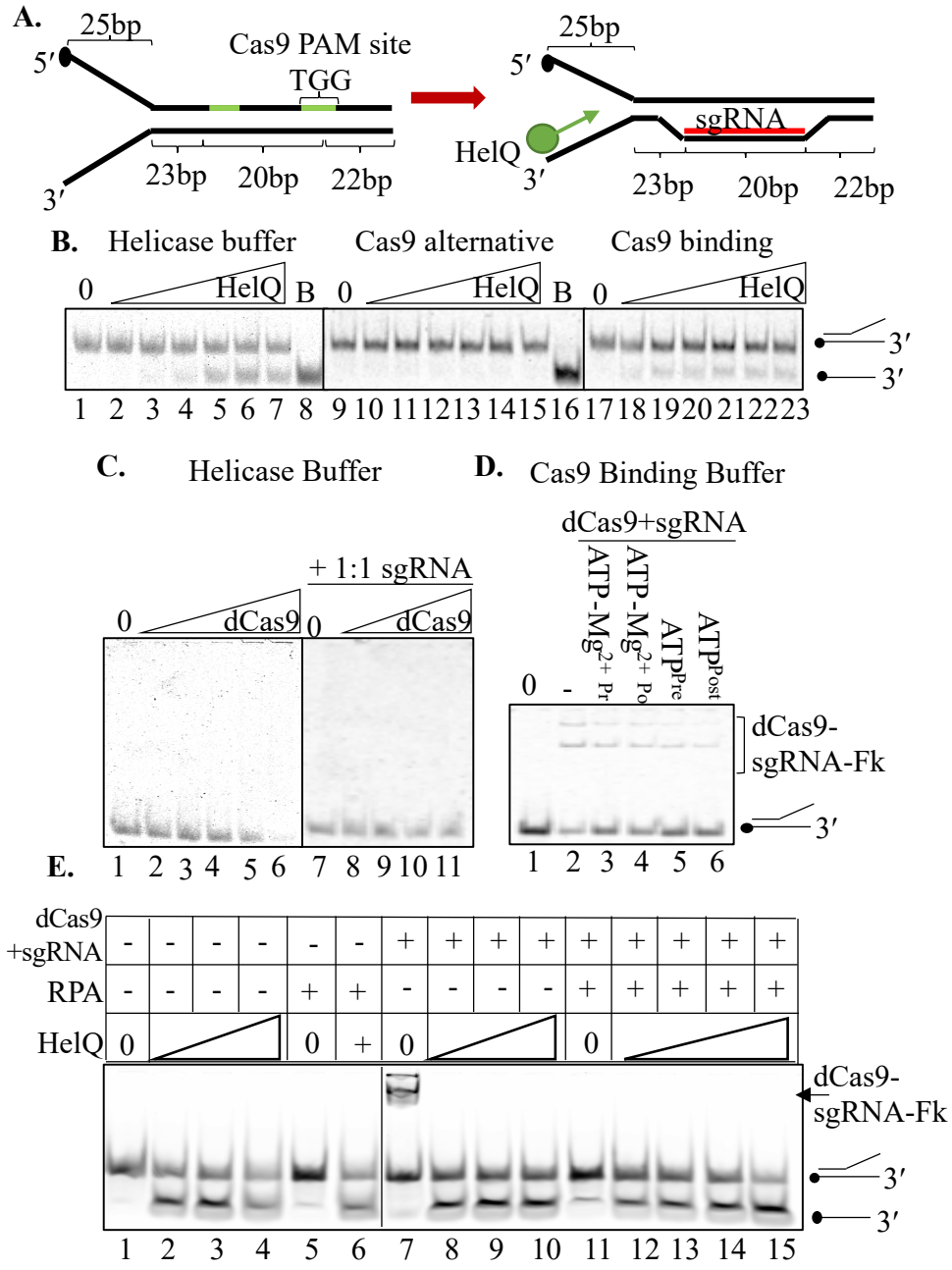


Figure 4.20 FL-HelQ does not unwind DNA in the presence of a Cas9 R-loop roadblock. A. Cartoon schematic showing the assembly of dCas9-sgRNA on the PAM-site containing fork into an R-loop. **B.** Representative TBE gel showing HelQ, used at 0, 10, 20, 40, 80 and 160 nM, can unwind the fork (25 nM) in helicase buffer (lanes 1-8) and Cas9 binding buffer (lanes 17-23). B shows a boil control showing substrate disassembly as indicated. Representative TBE EMSAs showing dCas9, used at 0, 10, 20, 40, 80 and 160 nM, binding to DNA in **C.** helicase buffer and **D.** Cas9 binding buffer. Addition of ATP-Mg^{2+} is also shown to have no impact on R-loop formation when added before (Pr) and after (Po) dCas9. **E.** HelQ unwinding analysis of DNA in the presence of 1:1 dCas9 (100 nM): sgRNA roadblock in Cas9 binding buffer. All reactions were terminated by addition of proteinase K with the exception of lane 7 showing successful formation of dCas9-sgRNA-DNA complex. HelQ could not translocate through the roadblock. Addition of RPA did not aid HelQ activity (see chapter 5).

dCas9 was incubated 1:2 with fork DNA and sgRNA to create a stable dCas9 R-loop complex evident in the absence of proteinase K in lane 7 (figure 4.20 E). However, free fork DNA unbound by dCas9 was also observed. All other reactions were terminated with proteinase K to terminate HelQ unwinding activity which resulted in R-loop complex dissociation. HelQ unwinding was observed in the presence of the dCas9 R-loop. However, this may be unwinding of the free unbound fork substrate (figure 4.20 lanes 8-10). Arguable, there is less unwinding observed in the presence of the R-loop (lanes 8-10) then in the absence of the R-loop (lanes 2-4) suggesting HelQ cannot displace dCas9 from DNA. These results are inconclusive for determining if HelQ can displace dCas9 in order to unwind the forked DNA, however, if displacement does occur it is likely to be minimal.

Data presented above is inconclusive on assessing if HelQ can displace a dCas9 R-loop complex from forked DNA. We next explored whether HelQ was able to displace dCas9 from dsDNA in the absence of unwinding activity (figure 4.21). Instead, we assessed for displacement by a mechanism of protein-protein interactions or binding to DNA. The ability of HelQ to displace dCas9 was measured by a reduction in the dsDNA-sgRNA-dCas9 complex band shifting evident in EMSAs. dCas9 formed stable complexes with DNA (figure 4.21 lane 8) but addition of HelQ did not impact the assembly of these complexes suggesting HelQ could not displace the complex (compare lanes 7 and 9). Furthermore, RPA incubated with HelQ prior to addition in these assays had no impact on dCas9 complex formation (lanes 11-13). RPA was explored as a potential protein partner of HelQ to aid in protein displacement from DNA, discussed in chapter 5. Evidence suggests HelQ is unlikely to be involved in processing R-loops. Studies elaborating on this were continued by Andrew Cubbon in the Bolt group.

In summary, evidence presented here implies HelQ does not translocate through a protein barrier and is unlikely to be involved in protein displacement *in*

vitro. However, it is important to recognise that HelQ may act on a specific protein barrier on DNA not assessed in this work because of the use of bacterial examples. Furthermore, HelQ may require the activity of an accessory protein, such as RPA, in order to displace barriers on the DNA.

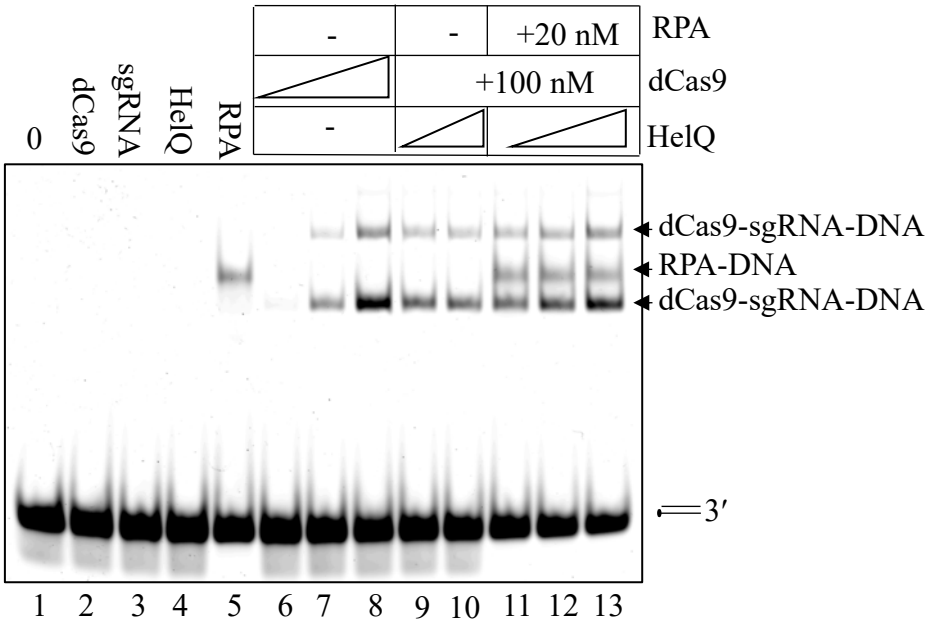


Figure 4.21 FL-HelQ does not displace dCas9 from dsDNA. Representative TBE EMSA gels showing HelQ activity upon loading onto a dCas9-sgRNA-dsDNA complex. 1:1 dCas9: sgRNA (100 nM) was assembled onto a PAM recognition site containing dsDNA substrate (25 nM) prior to addition of HelQ. HelQ was used at 0, 40, 80 and 160 nM. Addition of HelQ did not dissociate the dCas9-sgRNA-DNA complex as indicated to right of gel. RPA (20 nM) did not aid in HelQ activity (see chapter 5).

4.11 Analysis of C-HelQ in DNA binding and unwinding.

4.11.1 C-HelQ is fully active in the absence of the N-terminal of HelQ.

C-HelQ (85 kDa) consists of the RecA-like domains, WHD and domain IV of HelQ which retains the canonical ATPase hydrolysis activity. C-HelQ, without the N-terminal ORFan domain, was purified from *E. coli* and analysed independently to determine the role C-HelQ plays in active FL-HelQ. By doing this, assumptions could also be drawn to hypothesise the role of N-HelQ if C-HelQ behaved differently in its absence. C-HelQ was active as a helicase to a similar extent as

observed for FL-HelQ as shown in figure 4.2. Therefore, we concluded that C-HelQ, unsurprisingly due to the predicted domain organisation, is the 'core' active helicase and N-HelQ is not an essential component for this activity to occur.

Initially, we reasoned that the helicase activity of C-HelQ may be different in some respects from FL-HelQ, if N-HelQ had a role in the helicase function. However, C-HelQ could also not unwind dsDNA, a synthetic HJ or synthetic G4 structure as observed with FL-HelQ (figure 4.22 A-C). C-HelQ was able to partially unwind an extended DNA fork that FL-HelQ could not (figure 4.22 D). Unwinding of this substrate suggests C-HelQ was able to remain associated with the DNA for longer and therefore may be more processive than FL-HelQ. This may be as a result of the smaller protein size enabling C-HelQ to remain bound to the DNA for longer or less easily altered due to the smaller size and lack of intrinsically disordered N-HelQ. C-HelQ appeared to behave similarly to FL-HelQ in terms of substrate preference and unwinding activity. These results suggest N-HelQ is not required for DNA substrate recognition or translocation along the DNA.

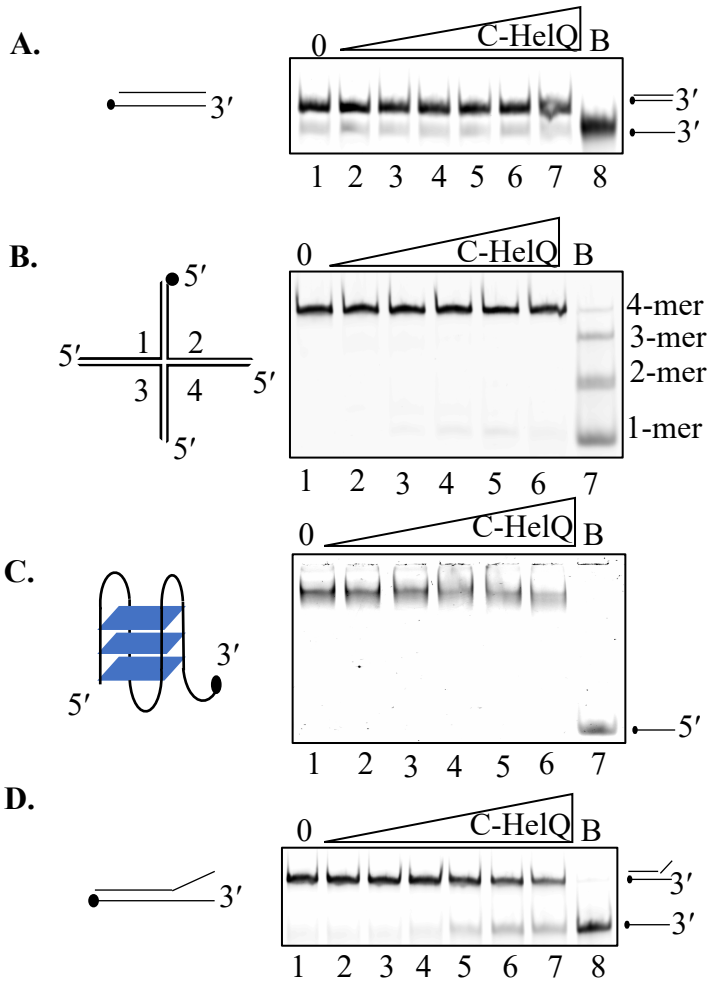


Figure 4.22 Helicase assays show C-HelQ is active as the 'core-helicase' and is inhibited by intermediate DNA structures. Representative TBE gels showing C-HelQ unwinding of **A.** dsDNA (25 nM), **B.** J6 HJ (25 nM), **C.** G4 quadruplex and **D.** an extended fork substrate (25 nM) as indicated in schematics to left of gel. Filled dots indicate 5′ Cy5 labels. C-HelQ was used at 0, 10, 20, 40, 80, 160 and 300 nM. B represents a boil control to show fully dissociated substrates. C-HelQ, similarly to FL-HelQ, did not unwind dsDNA or a HJ and showed weak unwinding of the extended fork substrate.

We next carried out fluorescence anisotropy on C-HelQ to test the binding ability to other DNA substrates (section 2.21.7) (figure 4.23). Fluorescence anisotropy was used to detect protein-DNA interactions in-solution. EMSAs trap DNA-protein complexes in-gel and therefore may not detect transient or fleeting interactions that in-solution methods can. Comparisons between the methods is discussed in section 4.15.

C-HelQ loaded onto a 3′ ssDNA DNA flap substrate (fork 3) with the highest binding affinity and a K_d value of 17.67 nM. C-HelQ was observed to load onto fork 2b with a K_d value of 54.83 nM and then ssDNA with a value of 148.9 nM. We can conclude that C-HelQ interacts with DNA and confirm C-HelQ, as observed with FL-HelQ, is able to load onto ssDNA and preferentially with a dsDNA-ssDNA junction.

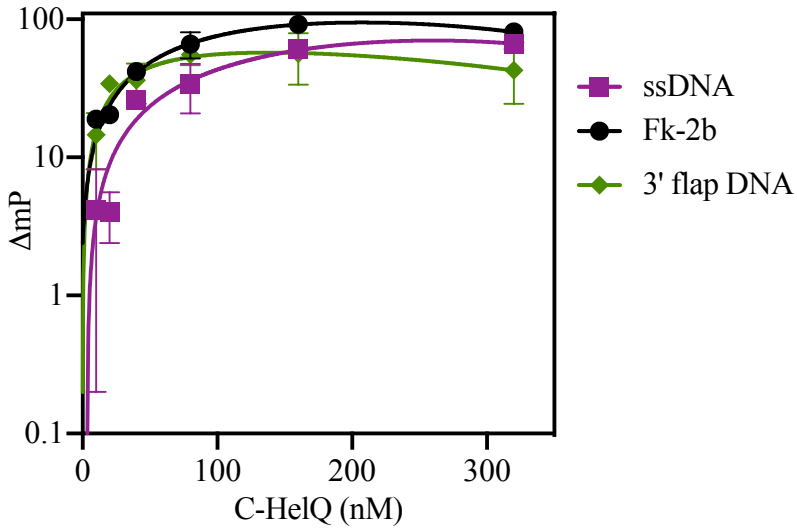


Figure 4.23 Fluorescence anisotropy shows C-HelQ loads onto ssDNA in fork structures with a strong binding affinity. A measurement of C-HelQ DNA binding using fluorescent anisotropy. Results based on triplicate reactions and shown here as standard error from the mean. C-HelQ was used at 0, 10, 20, 40, 80, 160 and 300 nM. DNA was optimised to 4 nM and was labelled with 5′ fluorescein. C-HelQ was able to bind to ssDNA, fork 2b and dsDNA with a free 3′ ssDNA flap.

4.11.2 C-HelQ helicase activity is inhibited by chemically modified DNA in the same way as FL-HelQ.

As for FL-HelQ, we next tested the unwinding activity of C-HelQ with chemically modified DNA substrates. We would expect that due to the presence of the core helicase properties of HelQ, C-HelQ would show similar behaviours to FL-HelQ when confronted with chemical modifications. This was indeed the case in experiments designed to allow for comparison between FL- and C-HelQ (figure 4.24 A-B).

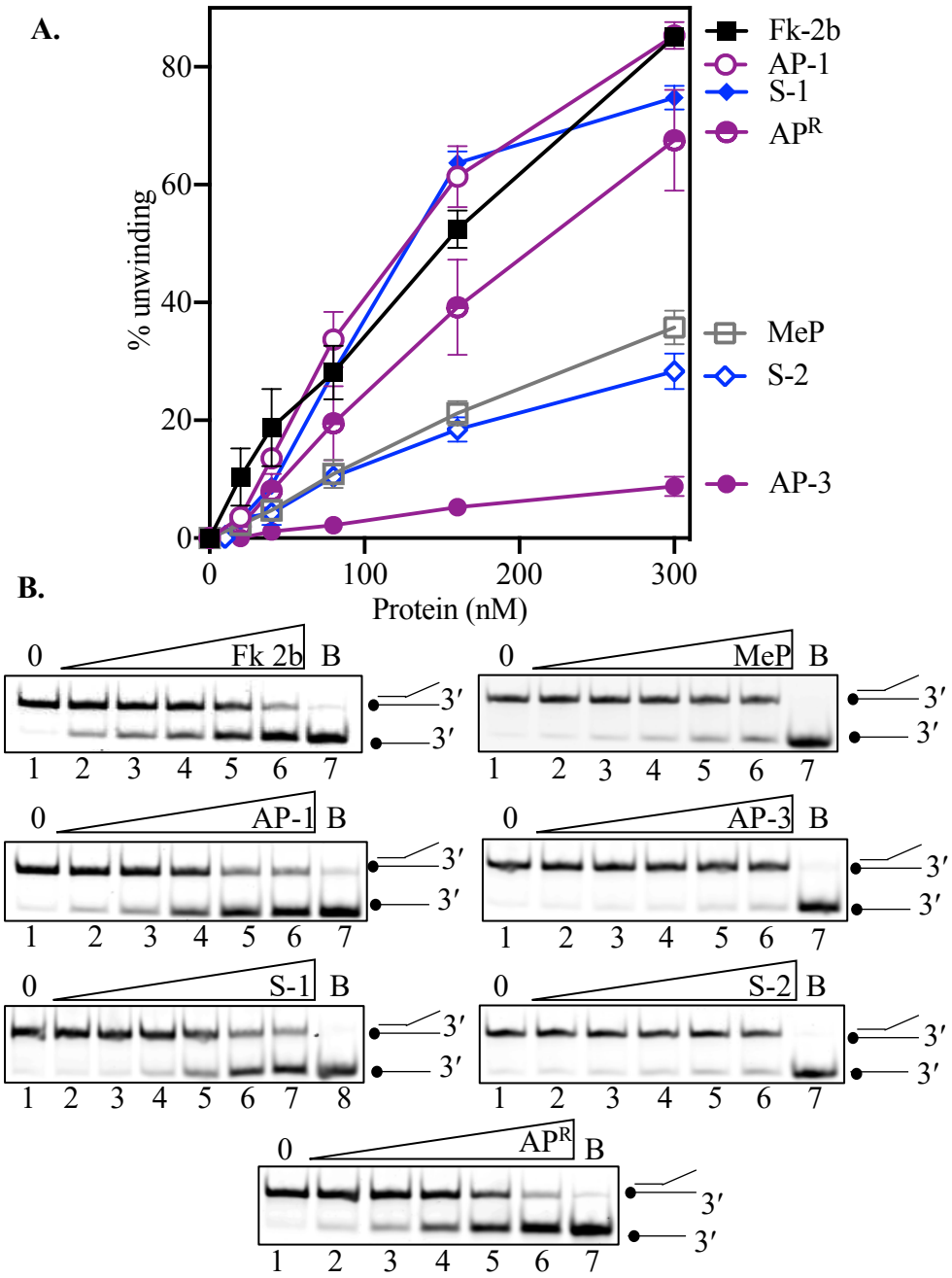


Figure 4.24 Helicase assays show C-HelQ 'core-helicase' activity is inhibited by internal chemical modifications. A. A measurement of C-HelQ unwinding of chemically modified fork 2b DNA (25 nM). **B.** Also shown are representative TBE gels carried out in triplicate to generate the graphs shown as standard error from the mean. DNA modifications included methyl phosphonate (MeP), phosphorothioate (S) and abasic sites (AP) in the duplex region and ssDNA region of the translocating strand and an abasic site on the non-translocating strand (AP^R). C-HelQ was used at 10, 20, 40, 80, 160 and 300 nM. B represents a boil control to show fully dissociated substrates. C-HelQ unwound DNA with modifications in the ssDNA region (AP-1 and S-1) but not with modifications in the duplex region (AP-3 and S-2).

An abasic site within the duplex region of the fork had the strongest inhibitory effect on C-HelQ unwinding, followed by the internal methyl phosphonate and phosphorothioate substitutions. C-HelQ was not impacted by modifications on the 3′ ssDNA region or the reverse abasic substrate. We therefore concluded enzyme activity of C-HelQ was identical to FL-HelQ confirming C-HelQ as the 'core-helicase' and that N-HelQ is not required for enzyme unwinding activity.

4.12 Analysis of C-HelQ activity with RNA containing substrates.

After assessment of HelQ activity with DNA, the ability of HelQ to process RNA was next analysed as for DNA substrates. Undesired DNA secondary structures can form during transcription and replication when DNA is exposed. Helicases and topoisomerase enzymes manage the formation of these structures, including R-loops.[103] R-loops are DNA-RNA hybrids that base pair across GC rich sequences. R-loops form when newly synthesised RNA threads back to hybridise with the template DNA as RNA polymerase moves along the DNA. RNA-DNA hybrids can result in the non-template DNA becoming displaced because the hybrids are more stable than dsDNA. Therefore, R-loops require resolving in order to restore the dsDNA. While R-loops are important in regulating gene expression, they are associated with genetic stress, particularly when replication forks collide.[103]

The ability of HelQ to process R-loops and RNA containing substrates was used to assess HelQ as an R-loop resolution helicase. Due to unforeseen issues with HelQ baculovirus production, there was limited sources of FL-HelQ protein available for these assays. Therefore, the core-helicase, C-HelQ, was predominantly assessed in R-loop resolution assays. Assays up to this point suggest C-HelQ behaves similarly to FL-HelQ and therefore could be used to predict FL-HelQ activity with RNA.

4.12.1 HelQ does not dissociate model R-loops *in vitro* but can dissociate RNA-DNA hybrids.

FL-HelQ and C-HelQ binding and unwinding activity to a synthetic R-loop was measured as previously described with DNA. The DNA 'bubble' shell in the absence of RNA as well as a DNA D-loop were used as controls. FL-HelQ was unable to unwind the R-loop or D-loop (figure 4.25 A). RecG, an *E. coli* branch migration protein that can dissociate R-loops, was used as a positive control (+) with R-loop dissociation being observed.[368] C-HelQ partially unwound the D-loop but was also unable to unwind the R-loop (figure 4.25).

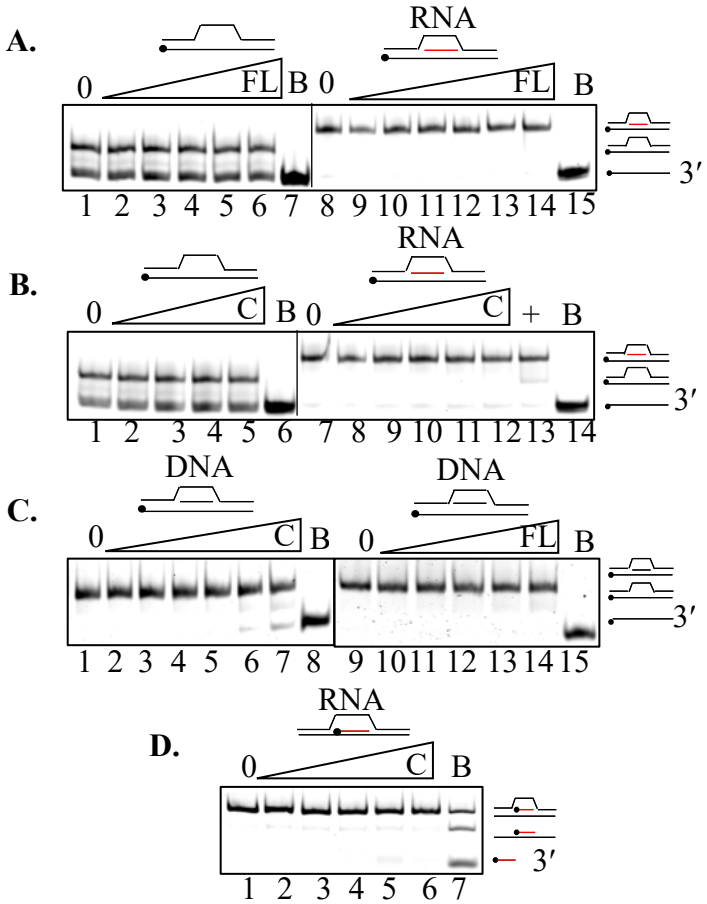


Figure 4.25 Helicase assays show FL-HelQ and C-HelQ could not displace RNA from a synthetic R-loop. Representative TBE gels showing **A.** FL-HelQ and **B.** C-HelQ unwinding activity of an R-loop with the Cy5 label on the DNA strand (DNA in black, RNA in red). **C.** FL-HelQ and C-HelQ were also shown to be unable to process the equivalent D-loop structure. **D.** C-HelQ could also not process an R-loop structure with the Cy5 label on the RNA strand. FL-HelQ (FL) was used at 0, 10, 20, 40, 80 and 160 nM and C-HelQ (C) at 0, 12, 37, 75, 150, 300 nM. B represents a boil control to show fully dissociated substrates as indicated in schematics to left of gel. RecG (100 nM) was used as a positive control for R-loop resolution.

The ability of C-HelQ to unwind the D-loop may be because C-HelQ can remain bound on the DNA for longer, which was observed with the extended fork substrate (figure 4.23). In addition, an alternative R-loop structure with the Cy5 label on the RNA strand was tested to follow migration of the RNA. Similarly, C-HelQ was unable to displace the RNA from the DNA shell (figure 4.26 D). From this data we concluded that C-HelQ, and likely FL-HelQ, is unlikely to be involved in R-loop resolution *in vitro*. This may be due to the absence of a sufficient length of ssDNA available for HelQ to load.

To determine if the inability of HelQ to displace RNA from an R-loop structure was because of the presence of RNA, we next assessed HelQ with RNA containing fork structures. Unwinding of DNA-RNA hybrid fork structures, similar to fork 2b with a 3′ loading RNA strand, was assessed alongside the equivalent DNA-DNA fork as a control. FL-HelQ could not unwind the RNA-DNA fork. However, C-HelQ was able to 100% dissociate the RNA from the fork structure evident from increased Cy5 ssDNA being observed (figure 4.26 A). These results imply C-HelQ is able to translocate along RNA. As previously discussed, C-HelQ may be more processive then FL-HelQ. However, the activity of the FL-HelQ protein may be questionable and poor due to the limitations in recombinant protein production. This may have created the variations in DNA unwinding between FL-HelQ and C-HelQ observed that may not be representative of new protein preparations. These results may also imply that the presence of N-HelQ is essential in regulating HelQ activity and directing the protein to DNA substrates only, this requires further exploration. Both FL-HelQ and C-HelQ were able to unwind the DNA-DNA fork equivalent (figure 4.26 B). Therefore, the lack of unwinding activity observed by FL-HelQ of the RNA-DNA fork is not due to fork length as seen with an extended

DNA fork substrate. These experiments require repeating using fresh HelQ protein preparations.

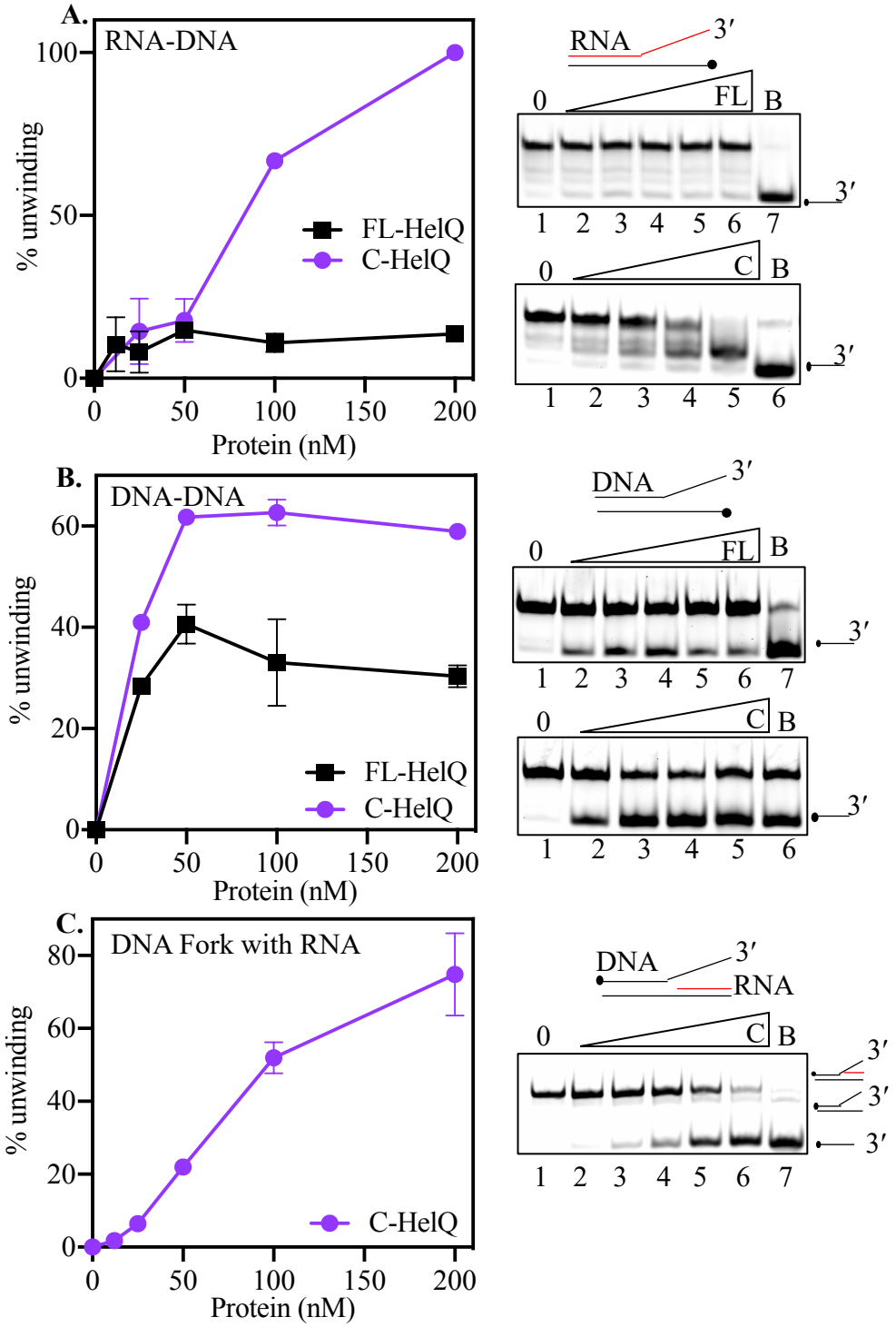


Figure 4.26 C-HelQ translocates along RNA to unwind RNA-DNA hybrids. **A.** A measurement of FL-HelQ and C-HelQ unwinding of a DNA-RNA fork hybrid (25 nM). Also shown are representative TBE gels carried out in duplicate to generate data presented as standard error from the mean. FL-HelQ (FL) was used at 0, 10, 20, 40, 80 and 160 nM and C-HelQ (C) at 0, 12, 37, 75, 150, 300 nM. B represents a boil control as indicated in schematics. **D.** C-HelQ and FL-HelQ were able to unwind the equivalent DNA-DNA fork and **C.** C-HelQ unwound a DNA fork with a free 3′ ssDNA flap created with RNA.

The ability of C-HelQ to process a DNA fork substrate with a complimentary RNA strand creating a 3′ ssDNA flap structure was tested (figure 4.26 C). C-HelQ was able to unwind this substrate to up to 80% confirming the presence of RNA does not impact C-HelQ unwinding ability.

These results suggest that C-HelQ is able to translocate along DNA in order to displace RNA as well as translocate along RNA to displace DNA. However, HelQ does not appear to be involved in the resolution of R-loop structures and branch migration.

4.13 Analysis of the role of a conserved residue motif IVa in C-HelQ.

4.13.1 Purification of the motif IVa mutant protein C-HelQ^{Y642A}.

A study carried out in the Bolt group on Hel308, the archaeal homologue of HelQ, identified a hyper-active mutation of a phenylalanine residue at position 295 (F-295-A) within the Hel308 protein motif IVa.[369,263] Motif IVa (GIA**F**HHAGL) is highly conserved between RecQ and Hel308 proteins; it is found in human proteins PolQ, BLM, HelQ, *E. coli* RecQ and other superfamily 2 DEAD-Box helicases. While the sequence is highly conserved, there is little known about the function of the domain. In solved crystal structures of SF2 helicases, this motif has been shown to form part of the β-sheet core of the RecA-like domain 2, implicating a role in enzymatic function. Furthermore, residues following the phenylalanine are part of a loop-helix region that are essential in RNA interactions of the DEAD-Box helicases.[370] Sequence alignment identified the region of interest in human HelQ studied in Hel308 (figure 4.27 A).

The F-295-A mutation within motif IVa Hel308 resulted in hyper-activity with an increase of 60% when compared to wild type Hel308. The mutant could unwind an extended fork substrate that wild type Hel308 was unable to unwind. Therefore,

the mutation improved activity and processivity which suggested that motif IVa has an auto-regulatory function by an unknown mechanism. Additional mutations within the domain also resulted in enhanced unwinding (data not published).[369]

The equivalent mutation was mapped to human C-HelQ at tyrosine position 642 and was used to make the mutant protein C-HelQ^{Y642A} (section 2.8.5). C-HelQ, rather than FL-HelQ, was used because of its ability to be cloned and purified in *E. coli.* The mutant protein was modelled using Pymol into C-HelQ predicted structures (figure 4.27 B-C).

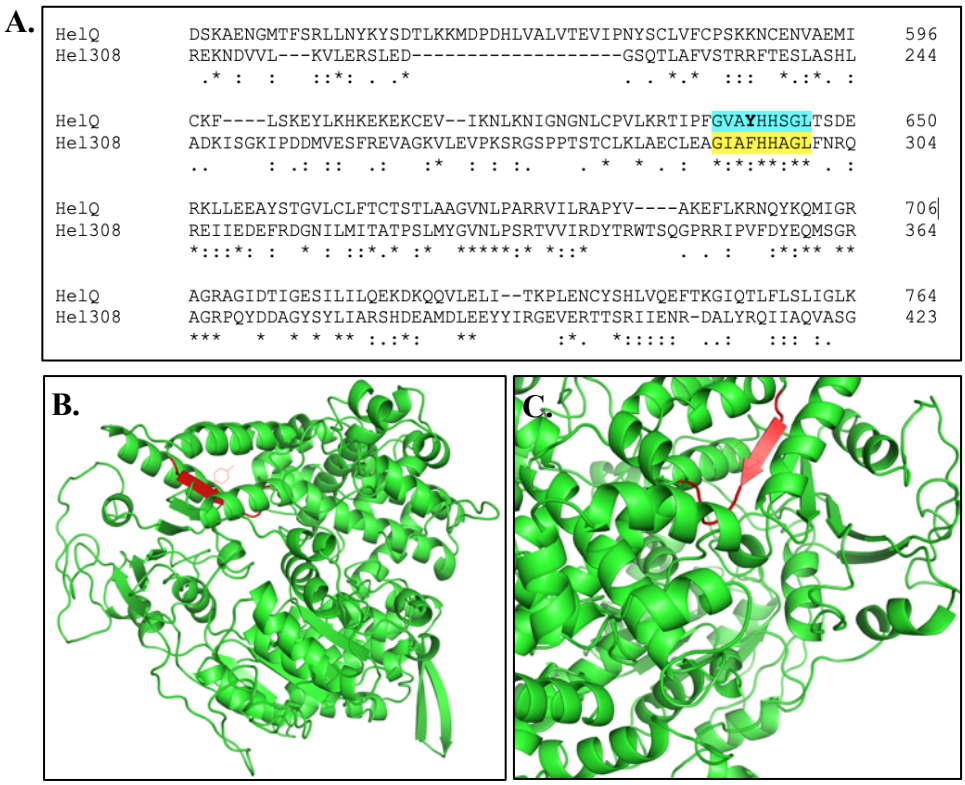


Figure 4.27 Conserved motif IVa found in Hel308 and HelQ. A. CLUSTAL sequence alignment of archaeal *Afu* Hel308 and human HelQ protein sequence. Motif IVa, conserved in SF2 helicases, was found conserved between Hel308 (yellow) and HelQ (blue). The HelQ tyrosine (Y) residue at position 642, which aligns with phenylalanine (F) in Hel308, was targeted for site-directed mutagenesis to produce the mutant protein C-HelQ^{Y642A}. **B.** The conserved motif IVa (red) was mapped to predicted models of C-HelQ in PHYRE 2 (green). **C.** Magnification of predicted model showing the target tyrosine residue.

C-HelQ^{Y642A} was purified from *E. coli* following the method described for wild type C-HelQ protein (figure 4.28). C-HelQ^{Y642A} showed similar signs of instability

and high levels of degradation to wild type C-HelQ. This was expected due to the instability seen for core C-HelQ and the predicted short half-life of FL-HelQ.

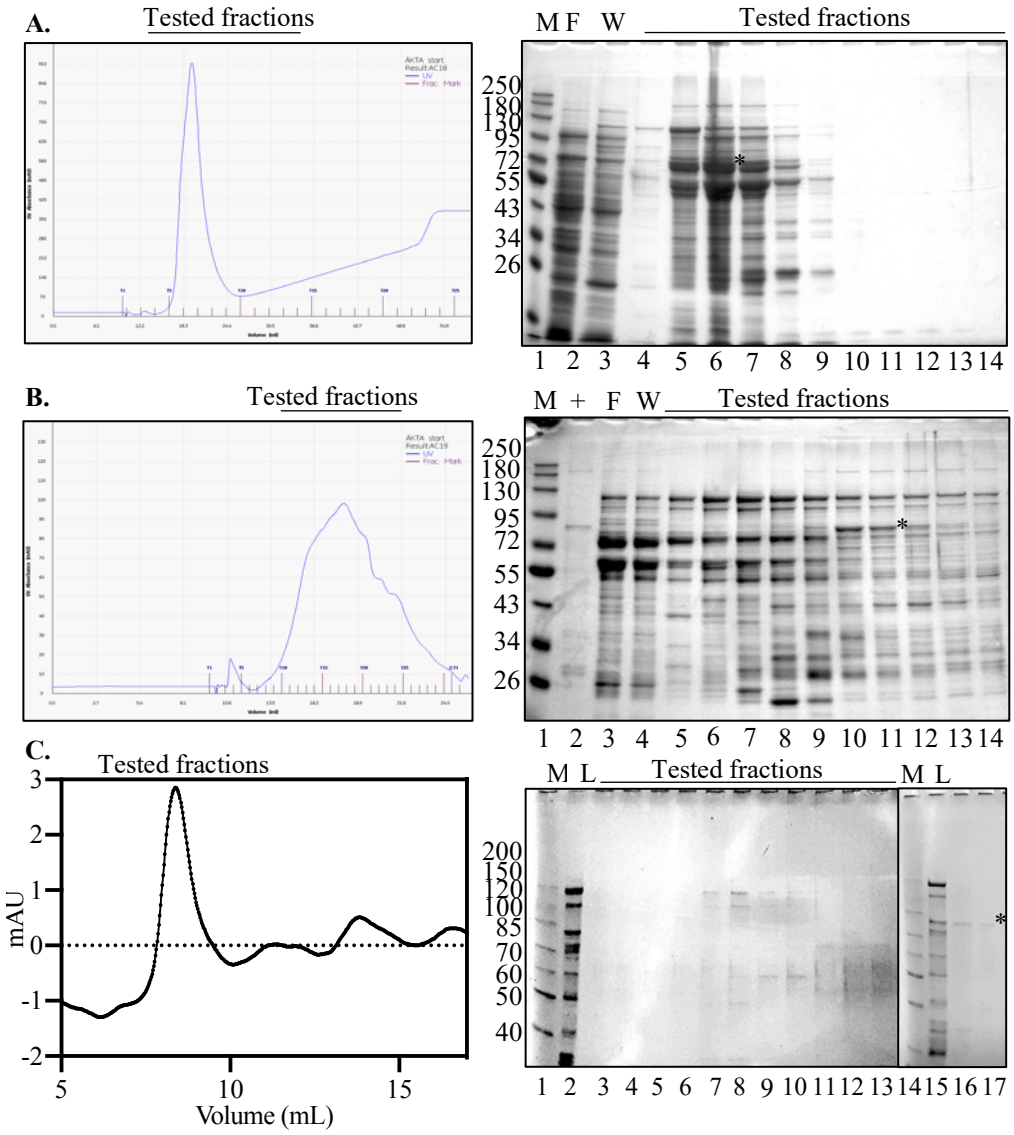


Figure 4.28 C-HelQ^{Y642A} protein purification from *E. coli*. A. HexaHis-tagged C-HelQ^{Y642A} (96 kDa) was sonicated, clarified and resolved on a $NiCl_2$ charged HiTrap Ni-NTA column. The panels show UV trace and Coomassie stained 8% SDS PAGE analysis of C-HelQ^{Y642A} eluting rapidly at low imidazole. C-HelQ^{Y642A} eluted at 85 kDa as indicated (*). **B.** C-HelQ^{Y642A} was further resolved by binding to a 1 mL Q-sepharose column. **C.** C-HelQ^{Y642A} containing fractions were pooled and dialysed prior to loading onto a superdex200 gel filtration column. A low yield of C-HelQ^{Y642A} was recovered. Legend = marker (M), wash-through (W), flow-through (F).

4.13.2 C-HelQ^{Y642A} does not behave as a hyper-helicase.

We initially assessed if C-HelQ^{Y642A} behaved as a hyper-helicase when compared to wild type C-HelQ. Comparisons between C-HelQ and C-HelQ^{Y642A} unwinding of fork 2b were made as a function of protein concentration (figure 4.29 A-B). ATPase inactive C-HelQ^{D463A} was used as a control for no unwinding activity. C-HelQ^{Y642A} was only able to unwind fork 2b to a maximum of 10% apparent from small amounts of Cy5 ssDNA being detected and showed no signs of hyper-activity. Therefore, the role of the highly conserved motif IVa is different in HelQ then to Hel308. The absence of the N-terminal ORFan domain in Hel308 implies it is not the lack of N-HelQ that resulted in the difference in phenotypes. It is proposed that motif IVa has a regulatory role in Hel308, a role that N-HelQ may be involved in FL-HelQ.

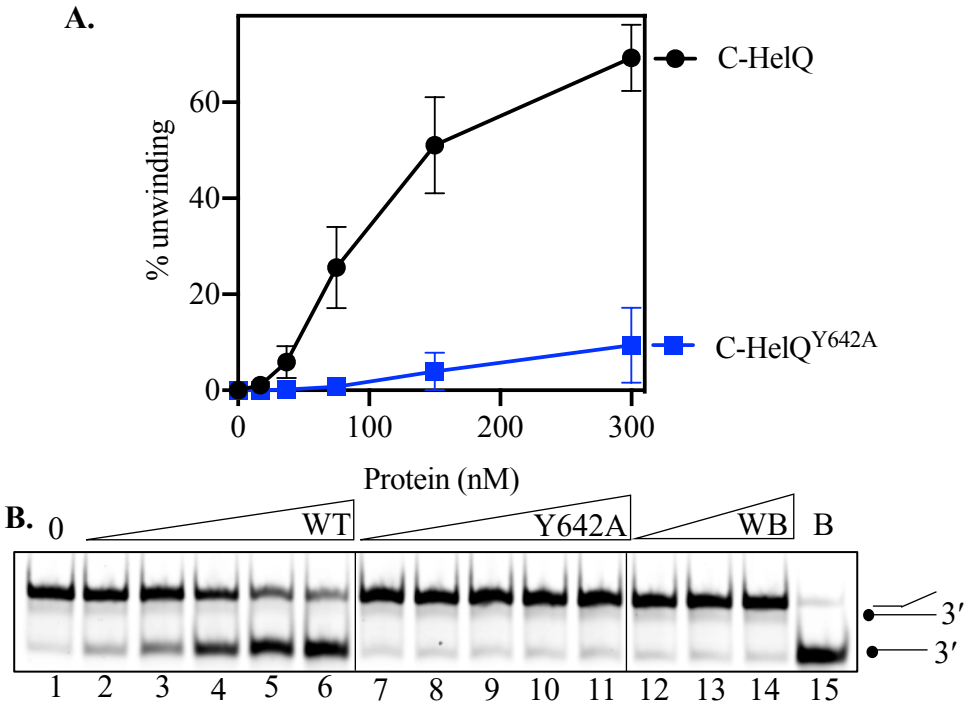


Figure 4.29 A mutation in motif IVa of C-HelQ (Y642A) removes the ability of C-HelQ to unwind forked DNA. A. A measurement of C-HelQ (black) and C-HelQ^{Y642A} (blue) unwinding of fork 2b (25 nM) in end-point assays. **B.** Also shown are representative TBE gels carried out in triplicate to generate the data and shown here as standard error from the mean. Proteins were used at 0, 10, 20, 40, 80, 160 and 300 nM. ATPase inactive C-HelQ^{D463A} (WB) was used as a negative control. B represents a boil control to show dissociated substrates as indicated. Wild type C-HelQ at 300 nM unwound 75% of the forked substrate in comparison to C-HelQ^{Y642A} with only 10% unwinding.

4.13.3 C-HelQ^{Y642A} assembles differently on DNA.

While C-HelQ^{Y642A} was unable to unwind forked DNA, it loaded onto both ssDNA and fork 2b in EMSAs. Additionally, C-HelQ^{Y642A} migrated faster on DNA then wild type C-HelQ on DNA (figure 4.30). Both C-HelQ^{Y642A} and C-HelQ formed single species on DNA. The difference in migration patterns suggest a role of rotational flexibility of the tyrosine residue. This would cause the mutant to interact and assemble differently onto the DNA. A faster migration suggests a more compact form of C-HelQ which may subsequently prevent its helicase activity. Further analysis of the role of motif IVa in conformation and flexibility is required. However, we conclude that in HelQ, the predicted motif IVa does not have a role in autoregulation, as previously proposed. Previous findings implicating the motif in RNA binding in the DEAD-Box helicases may suggest that this domain is important in substrate recognition and therefore may explain why differences are observed between the mutant and wild-type protein with DNA. This requires further exploration.

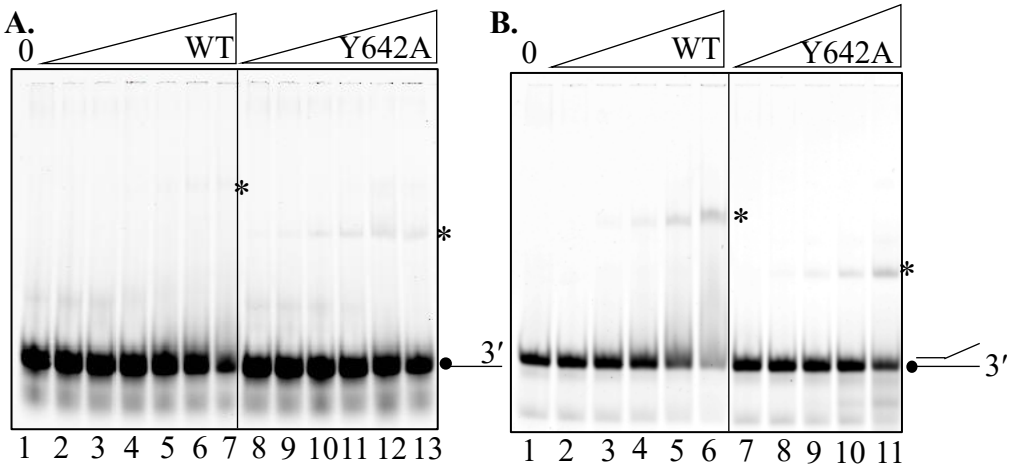


Figure 4.30 EMSAs show C-HelQ^{Y642A} loads onto DNA forming different complexes when compared with wild type C-HelQ. TBE EMSA gels showing C-HelQ (WT) and C-HelQ^{Y642A} (Y642A) binding with **A.** ssDNA (25 nM) and **B.** fork 2b (25 nM). Schematics show substrates with filled dots indicating 5′ Cy5 labels. Protein-DNA complex formation indicated (*). Proteins were used at 0, 20, 40, 80, 160 and 300 nM. C-HelQ^{Y642A} migrated faster on both substrates suggesting different assembly on DNA.

4.14 N-HelQ does not bind to DNA *in vitro*.

Work carried out here implies that C-HelQ is the core helicase domain necessary for unwinding activity. However, EMSAs suggested that C-HelQ is inadequate at forming stable DNA-protein complexes to a similar extent observed with FL-HelQ. Therefore, we propose that N-HelQ may have a role in stabilising the formation of the DNA-protein complex. Bioinformatic analysis described in chapter 3 predicted that N-HelQ has no conserved DNA binding motifs, confirmed with EMSAs and helicase activity assays presented in figure 4.1. In order to exhaust the notion that N-HelQ cannot load onto DNA independently, additional EMSAs were carried out with different DNA substrates. EMSAs were carried out using wild type N-HelQ, PWI-N-HelQ$^{D142A/F143A}$, phosphomimetic N-HelQ$^{S158/178D}$ and the N-HelQ truncated proteins (N1-3) introduced in chapter 3 (section 3.11.2).

N-HelQ was analysed for binding with forked fork 2b, ssDNA, synthetic HJ and dsDNA. EMSA in-gel analysis showed no evidence for protein-DNA interactions with wild type N-HelQ evident from no DNA shifting (figure 4.31 A). Additionally, PWI-N-HelQ$^{D142A/F142A}$ did not load onto the DNA (figure 4.31 B) confirming the PWI-like domain does not have a role in DNA binding. Weak interactions between phosphomimetic N-HelQ$^{S158/178D}$ and ssDNA were observed (figure 4.31 C). The activated phosphorylation mimic had a strong affinity for ssDNA when compared with wild type N-HelQ (figure 4.31 D). This implies that N-HelQ may require activation through protein modifications to activate DNA binding activity. However, the DNA binding observed was not significantly different enough between phosphomimetic N-HelQ and wild type N-HelQ to be convincing from the EMSA data alone. To confirm this as true DNA binding and not as a result of protein over-crowding a more reliable method of fluorescence anisotropy was used, shown later.

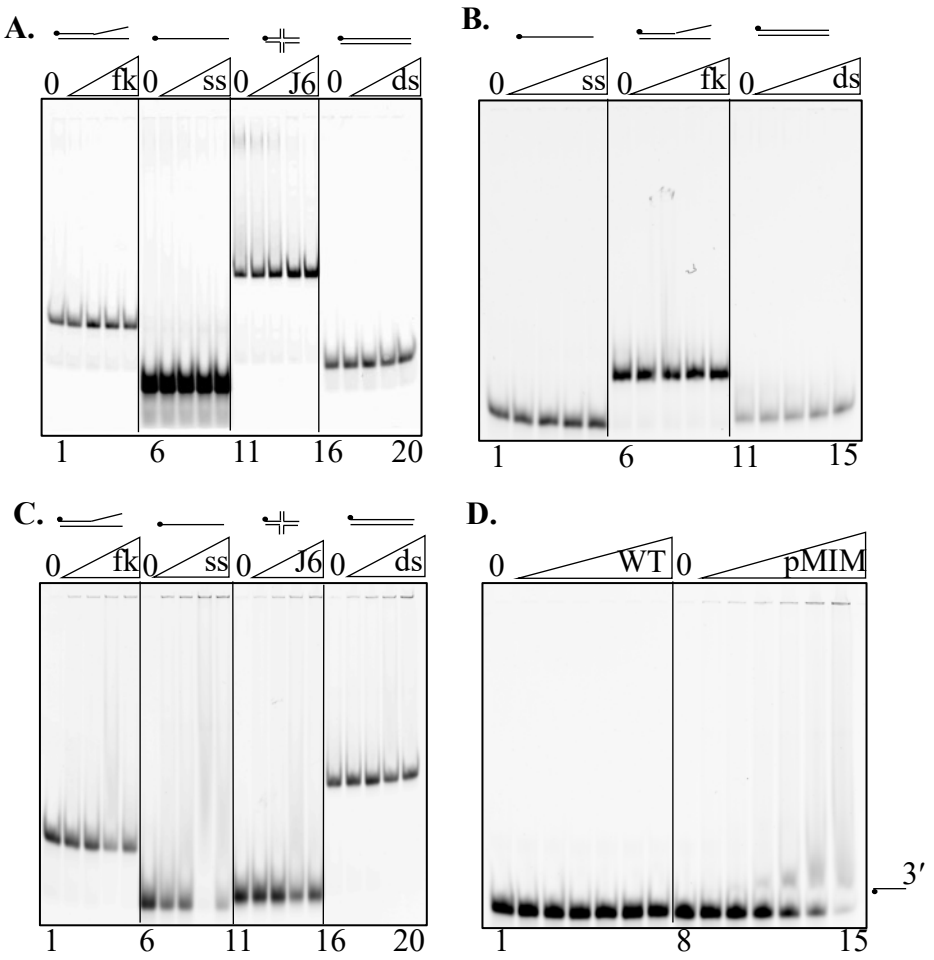


Figure 4.31 EMSAs show N-HelQ does not load onto intermediate DNA substrates. TBE EMSA gels showing the inability of N-HelQ, phosphomimetic N-HelQ$^{S158/178D}$ (pMIM) and N-HelQ$^{D142A/F143A}$ (PWI) to load onto DNA substrates (25 nM). Proteins were used at 0, 50, 100, 200 and 500 nM. Assays shown include **A.** N-HelQ, **B.** PWI and **C.** pMIM with fork 2b (fk), ssDNA (ss), Holliday Junction (J6) and dsDNA (ds). Substrates indicated above gel and filled dots represent 5′ Cy5 label. **D.** Also shown are EMSAs of pMIM-N-HelQ binding to ssDNA (25 nM) compared with N-HelQ. N-HelQ and PWI-N-HelQ$^{D142A/F143A}$ showed no DNA binding. However, weak binding by phosphomimetic N-HelQ$^{S158/178D}$ to ssDNA was observed.

Analysis of the N-HelQ -N1, -N2 and -N3 proteins for DNA binding were tested to determine the location of essential DNA binding regions in HelQ. N-HelQ-N1 did not bind to any DNA substrates as predicated because of its sequence similarity to wild type N-HelQ (figure 4.32 A). N-HelQ-N2, containing the RecA-like domain 1, loaded onto ssDNA and fork 2b (figure 4.32 B). However, formed complexes were unstable evident from observed binding being mostly smearing of

Cy5 DNA with no defined DNA-protein species. N-HelQ-N3 also loaded onto DNA forming weak but more defined DNA-protein complexes (figure 4.32 C). This observation was expected because N-HelQ-N3 contains the largest proportion of HelQ but still lacks the ratchet DNA binding domain considered important in successful DNA-protein complex formation. Reduced binding further supported the prediction that domain IV and the rachet helix are essential for DNA interactions. In addition, helicase assays showed that N-HelQ-N2 and -N3 were unable to unwind the DNA as expected (Figure 4.32 D) due to the lack of all the canonical helicase domains.

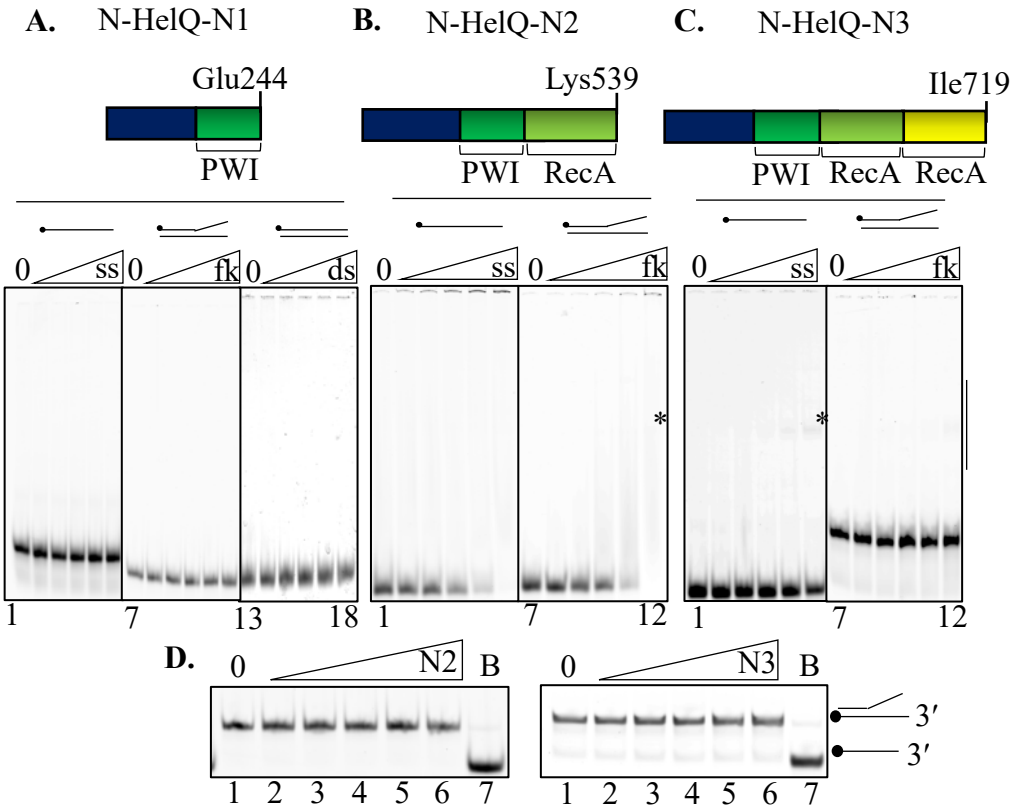


Figure 4.32 Extended N-HelQ proteins form weak DNA-protein complexes but have no unwinding activity. Representative TBE EMSA gels showing **A.** N-HelQ-N1, **B.** N-HelQ-N2 and **C.** N-HelQ-N3 binding with DNA substrates (25 nM). Substrates tested included ssDNA (ss), fork 2b (fk) and dsDNA (ds) as indicated above gels, filled dots represent Cy5 label. **D.** Also shown are unwinding activity assays of N-HelQ-N2 (left) and N-HelQ0N3 (right) with fork 2b (25 nM). Proteins were used at 0, 25, 50, 100, 200 and 500 nM. B shows a boil control of dissociated DNA. N2 and N3 showed weak DNA binding but no unwinding activity.

4.14.1 Fluorescence anisotropy showed N-HelQ does not load onto DNA.

EMSAs traps DNA-protein complexes in-gel and therefore may not detect transient interactions that in-solution methods can. Therefore, fluorescence anisotropy was used as a more sensitive method to detect N-HelQ DNA interactions (section 2.21.7). Phosphomimetic N-HelQ$^{S158/178D}$, with weak binding to DNA observed in EMSAs, was also tested to assess whether this effect observed was real. The fluorescence polarisation (mP) was measured for N-HelQ with 5 nM fluorescein-labelled ssDNA and fork 2b (figure 4.33 A). There was no observed increase in polarisation with increased concentration of wild type N-HelQ which confirmed the absence of protein-DNA interactions. The decrease in florescence across the replicas may suggest the protein was interfering with the intensity readings.

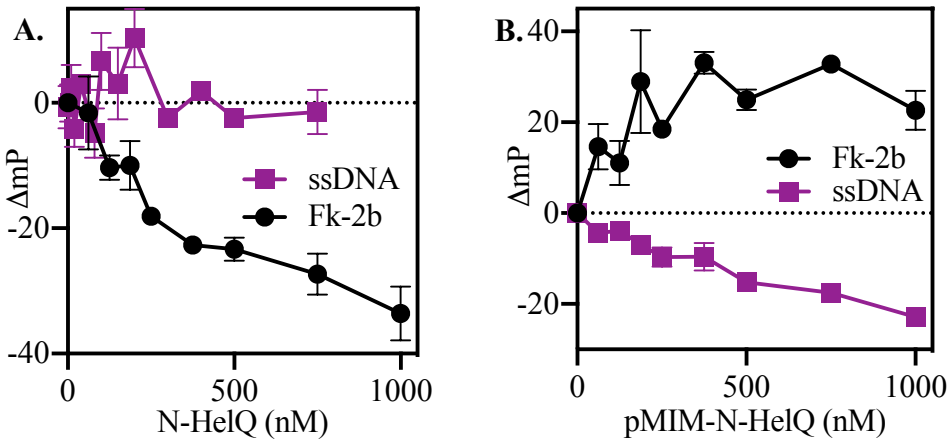


Figure 4.33 Fluorescence anisotropy shows N-HelQ does not load onto DNA. A measurement of DNA binding to ssDNA and fork 2b by **A.** N-HelQ and **B.** phosphomimetic N-HelQ$^{S158/178D}$. Reactions were in triplicate, and standard error from the mean is shown. Proteins were used at 0, 62, 125, 187, 250, 375, 500, 750 and 1000 nM. DNA was optimised to 4 nM and was labelled with fluorescein. N-HelQ did not interact with ssDNA of fork 2b. Phosphomimetic N-HelQ$^{S158/178D}$ showed weak binding to fork 2b.

Analysis of phosphomimetic N-HelQ$^{S158/178D}$ showed no interactions with ssDNA which contradicted results seen in EMSAs, however, did interact with forked DNA (figure 4.33 B). Phosphomimetic N-HelQ$^{S158/178D}$ had a K_d value of

1195 nM with ssDNA showing no binding and a K_d value of 81.35 nM with fork 2b. This may implicate phosphorylation in the activation of N-HelQ. N-HelQ may interact with DNA when in close proximity or when FL-HelQ assembles onto the DNA. However, these interactions were only observed when large quantities of protein were used which suggests the observation may be caused by protein over-crowding resulting in DNA proximity. Therefore, additional factors e.g. binding partners or additional PTMs may be required in conjunction with phosphorylation to see a greater effect. You would expect that if these phosphorylation events alone activated N-HelQ function, a more significant increase in this function would be detected.

There is little evidence to support a role of phosphorylation at these sites in the activation of N-HelQ to load onto DNA *in vitro*. In comparison with C-HelQ, a positive control for DNA binding (figure 4.24), which required 10 times less protein to give 10 times more activity then phosphomimetic N-HelQ$^{S158/178D}$. This concluded that N-HelQ is unlikely to be involved in the loading of HelQ protein onto DNA implicating an alternative role of N-HelQ that is explored in chapter 5.

4.15 Summary and conclusions.

This chapter assessed the activity of HelQ with DNA to be able to hypothesis how HelQ loads and assembles onto DNA and its mechanism of translocation. In addition, a role of HelQ in alternative helicase activities was also assessed. These are discussed here next.

4.15.1 HelQ is inhibited by both physical and chemical DNA roadblocks.

The assessment of HelQ activity with chemically modified fork substrates provided detail on the translocation mechanism of HelQ. In addition, these substrates mimicked a DNA damage environment. DNA base residues are

susceptible to chemical modifications from endogenous cellular proteins or the environment.[371] The hydrolysis of the sugar-phosphate bond gives rise to an abasic site which can result in localised distortion of the DNA double helix producing base stacking disruptions and twisting of the helical axis.[371] Therefore, the assessment of HelQ with abasic DNA provides detail on whether interactions rely on the structured hydrogen bonding and the intra- and inter-molecular base stacking interactions between aromatic residue side chains and HelQ.[201] Analysis of HelQ with methyl phosphonate and phosphorothioate groups provide insight into the importance of electrostatic interactions between HelQ and DNA.[201] These modifications are solely synthetic non-natural types of linkages used to determine translocation mechanism.[371] The phosphorothioate substrate also provided evidence for whether the size of the oxygen atom or length of the phosphate-oxygen bond is essential for HelQ interactions with DNA.[372]

Analysis of HelQ unwinding of chemical modified DNA *in vitro* identified essential interactions with the internal bases and external backbone of DNA. The requirement of DNA bases suggests HelQ requires the base stacking conformation of DNA to pull the DNA through the enzyme as it unwinds. Electrostatic interactions with the DNA backbone, perhaps in a supportive role to stabilise HelQ, have also been determined. These interactions are important in the duplex region but not the single-stranded region of the translocating DNA. Therefore, HelQ binds close to the branch point of the fork substrate. Analysis of reverse substrates suggested interactions with the base on the non-translocating strand are not essential for unwinding. In summary, HelQ translocates and unwinds DNA utilising interactions on the strand that it initially binds. ATPase activity and DNA binding were not impacted by the presence of the chemical modifications suggesting HelQ attempts to unwind the substrate but is halted at the modification.

HelQ was impacted by DNA chemical modifications, similarly to other helicases. For example, the FancJ helicase is inhibited by a modification on the phosphodiester backbone in the 5′ ssDNA tail but not the 3′ ssDNA tail, close to the fork junction point.[371] Another example, the WRN helicase, is stalled by a polyaromatic hydrocarbon DNA adduct on the non-translocating DNA strand. While this is different to an abasic site, it suggests essential interactions with the internal base of DNA.[371] The use of modified DNA can be used to determine translocation of DNA. The mechanism of translocation of hexameric Mcm2-7 has been identified through crystal structure analysis. Mcm2-7 relies solely on DNA interactions with the phosphodiester backbone of the translocating strand through a spiralled positively charged region and does not interact with the bases.[373]

In order to extend this work further, experiments looking at the replacement of bases with bulky adducts in DNA and assess HelQ limitations in processing these internal modifications. The abasic site changes the stacking of the duplex and therefore assessment of substrates with more distorted stacking could provide further enlightenment.

Unwinding activity was also prohibited by the presence of protein barriers. HelQ was unable to displace a number of different sized protein roadblocks with varying levels of DNA binding affinity of the roadblocks tested. However, it cannot be ruled out that HelQ has a role in protein displacement. Displacement activity may be specific to the type of protein roadblock and substrate. Helicases have been shown to act in protein displacement including eukaryotic Mcm2-7. Hexameric Mcm2-7 encircles DNA and unwinds via steric exclusion allowing for protein displacement.[371] DNA clearance is achieved through combined motor activity and the inability of the protein to pass through the central pore.[371,357] Dimeric HelQ would not act through steric exclusion that would allow displacement.

Helicases have been shown to displace proteins through the aid of a protein partner. FancJ and RecQ1, when recruited and complexed to RPA, are able to displace a protein roadblock from the DNA.[374] HelQ may require stimulation by a protein partner in order to initiate displacement activity, explored in chapter 5.

4.15.2 HelQ does not unwind intermediate DNA structures.

HelQ did not unwind a number of synthetic DNA substrates including a HJ, G4 quadruplex, dsDNA or an extended fork. The resolution of G4 quadruplexes, carried out by WRN and BLM helicases, is important for the maintenance of telomeres.[375] A conserved RQC sequence domain has been shown to bind the G-quadruplex and is essential in this activity. HelQ does not appear to fall into this category of helicases. HelQ was unable to process dsDNA as expected because of its strong affinity for ssDNA. This was confirmed by its inability to bind dsDNA. The inability of HelQ to process an extending fork substrate suggested HelQ is likely to be important in processing smaller regions of DNA during repair. To assess HelQ processivity, DNA substrates of varying lengths could be tested. Additionally, HelQ does not act in HJ resolution, a role shown by eukaryotic RecQ helicases.[376] Differences in DNA processing by RecQ and HelQ observed throughout this chapter means the inability for HelQ to resolve HJs is unsurprising. Five eukaryotic RecQ helicases have also been shown to exhibit strand-annealing activity.[214] Therefore, similarly, the inability of HelQ to re-anneal DNA in end-joining is unsurprising.

Work carried out here did not establish any additional roles in DNA processing of HelQ. HelQ is not a helicase involved in re-annealing, branch migration, R-loop or G4 quadruplex resolution *in vitro*. Data presented here, combined with high levels of protein instability, implies HelQ has a specific and short-lived role.

4.15.3 A comment on DNA binding in-gel vs. in-solution.

N-HelQ-DNA interactions were assessed using in-gel EMSA and in-solution anisotropy analysis with no interactions being detected. However, differences in phosphomimetic N-HelQ$^{S158/178D}$ binding were observed. In-gel methods 'trap' proteins and may not be representative of transient interactions of passing molecules. Therefore, for the purposes of N-HelQ, the anisotropy data is more representative of transient interactions between N-HelQ variants and DNA.

4.15.4 C-HelQ, but not N-HelQ, forms multiple interactions with DNA.

N-HelQ is not essential for DNA interactions evident from EMSAs and fluorescence anisotropy. Phosphorylation within the PWI-like domain did not appear to fully activate a function in DNA binding of N-HelQ, however, DNA binding was observed. Therefore, phosphorylation is unlikely to be a required modification of N-HelQ when considered alone or at only these two serine residues. Further analysis could look at mutating further serine residues or determine the effect of the phosphomimetic mutations in addition to other factors. Weak binding of wild type N-HelQ observed in EMSAs confirms a supportive role of N-HelQ once the core helicase assembles onto DNA. However, these interactions are not essential and are a result of close proximity. This is confirmed by reduced DNA binding activity of C-HelQ. The core helicase may require stabilising by N-HelQ. Further fluorescence anisotropy of the variants of N-HelQ with increasing presence of C-HelQ could be used to determine where DNA binding affinity increases.

Despite reduced binding activity, the core helicase C-HelQ was able to unwind DNA substrates similarly to FL-HelQ. Therefore, the role of N-HelQ is additional to DNA processing and is likely inclusive of protein stabilisation. The slightly improved helicase processivity of C-HelQ, in comparison to FL-HelQ, may be as a result of a less bulky protein maintaining its hold on the DNA for longer.

C-HelQ was unable to unwind substrates that FL-HelQ could not already unwind. Therefore, C-HelQ behaves similarly to FL-HelQ, adopting the same mechanism of translocation. This confirms N-HelQ is not involved in the mechanism of translocation and unwinding.

HelQ did not appear to be involved in the resolution of R-loops or D-loops. C-HelQ was able to process RNA containing fork substrates and translocate along RNA. Discrepancies between FL-HelQ and C-HelQ may be as a result of higher processivity of the smaller C-HelQ protein remaining bound to DNA for longer. Alternatively, the presence of N-HelQ may be important in differentiating between substrates. N-HelQ may prevent FL-HelQ associating with RNA containing substrates and may have a role on substrate specificity. This requires further exploration. A FL-HelQ protein with N-HelQ mutations to attempt to knock out this DNA/RNA identifying ability could be used to further assess this.

Conservation between HelQ and archaeal Hel308 motif IVa identified a potential tyrosine residue of interest (Tyr-642) in HelQ. This residue has been shown to be essential for RNA substrate binding in DEAD-Box helicases located in the RecA-like domain 2, therefore this motif is likely essential for enzymatic activity. Mutations in Hel308 resulted in hyper-activity implicating a role in auto-regulation that is knocked out. C-HelQ^{Y642A} did not give the same hyper-active phenotype observed with Hel308 suggesting the role of the motif is different. However, differences in DNA binding between C-HelQ and C-HelQ^{Y642A} were observed. C-HelQ^{Y642A} migrated at a faster rate on ssDNA compared to C-HelQ suggesting differences in rotational flexibility. The Y642A mutation may interfere with the DNA binding conformation of C-HelQ. C-HelQ potentially forms dimers or dimer of dimers when assembling on DNA which is required for helicase activity. The mutation may interfere with the assembly of dimers on the DNA and

therefore result in different migration patterns. The inability of the mutant to assemble as a dimer may account for the absence of unwinding activity. This implicates motif IVa in dimerisation and flexibility of C-HelQ or protein assembly onto the DNA and should be explored.

Assessment of hyper-active phenotypes can be used to assess the role of domains in protein regulation and is a potentially interesting avenue to pursue with HelQ. Another helicase that has been shown to gain hyper-helicase activity *in vitro* is UvrD.[377] Mutagenesis of two residues (Asp 403/4→Ala) in the 2B subdomain of UvrD resulted in improved unwinding ability. This mutation is suggested to weaken the interaction with the 1B subdomain allowing the helicase to adopt a more open conformation and unwind DNA using a different mechanism of 'strand displacement'.[377] Similar mutations within C-HelQ could provide evidence on how the protein assembles on to DNA.

4.16 Hypothesised model for HelQ binding and unwinding of DNA.

The precise mechanism of translocation and unwinding remains unidentified. However, further evidence suggests that HelQ predominantly interacts with the translocating strand allowing for hypotheses to be made. This favours an unwinding mechanism of strand displacement by a protein WHD wedge forced into the branch point as HelQ translocates along the DNA. HelQ can be ruled out as a replicative hexameric helicase that exhibits steric exclusion due to evidence described here and in chapter 3 (section 3.8). Evidence supports a fluid interaction between HelQ and DNA that relies on protein flexibility and multiple DNA interactions, for example from both monomers of the assembled oligomer. This may implicate a dimeric inchworm mechanism like many other SF2 dimeric helicases including Hel308. Evidence presented with HelQ on chemically modified DNA substrates suggests HelQ moves along DNA and is therefore unlikely to follow a reeling mechanism,

however, this cannot be ruled out and requires further analysis.[314] Translocation of HelQ along the DNA actively pushes the wedge into the junction separating the strands. A hypothesised model for the interactions between HelQ and DNA suggests DNA binding and unwinding is orchestrated by the core helicase C-HelQ. N-HelQ has an additional role which may include the stabilisation of C-HelQ and require activation in order for its role be become apparent. The model below (chapter 6) summaries this hypothesis which is explored in chapter 6.

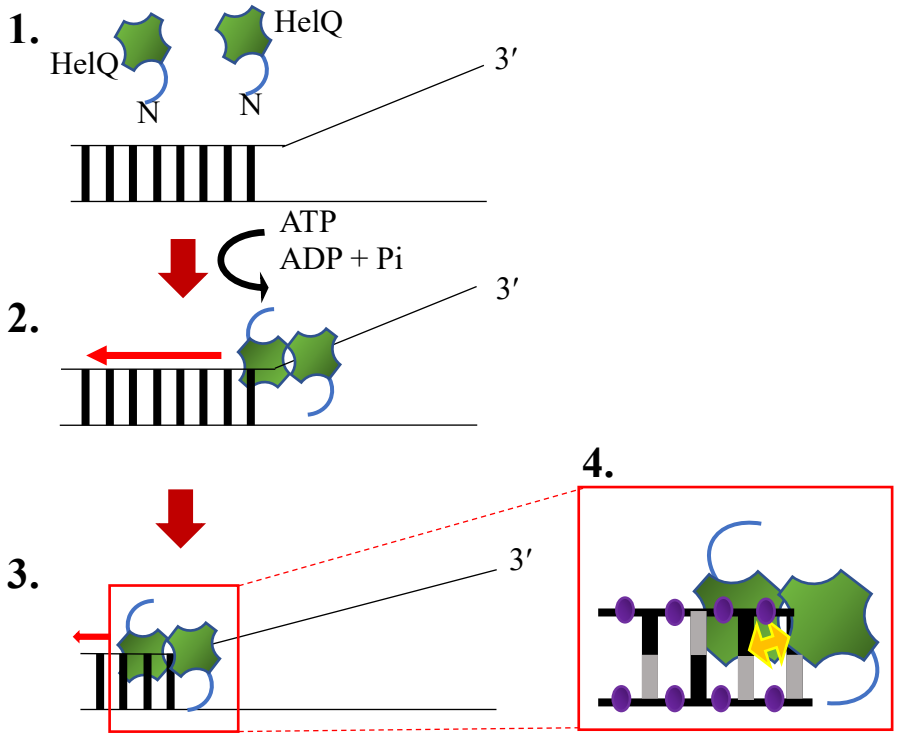


Figure 4.35 Proposed model of FL-HelQ binding to DNA. HelQ interacts and assembles onto DNA as an active dimer (1). HelQ binds to ssDNA region of the fork substrate and translocates along in a 3′ to 5′ direction using ATP hydrolysis (2). HelQ dimers interact with the internal base of the translocating strand as well as the external backbone, likely though electrostatic interactions (3-4). HelQ displaces the non-translocating strand via strand exclusion. HelQ does not appear to interact with the non-translocating strand.

4.17 Outlook and future work.

FL-HelQ baculovirus production and protein purification became a limiting factor during work carried out here. Stopped flow kinetics using higher protein concentrations to calculate V_{max} and k_M data for HelQ would add to the reliability

of the data presented here. Kinetics data could be collected for HelQ to assess DNA nucleotide step size that would confirm a dimeric rolling mechanism proposed here. Further chemical modifications in DNA, including the use of alternative bases such as 8-oxoguanine and thymine glycol substitutions, could be used to assess how HelQ binds and translocates along synthetic damaged DNA. Furthermore, structural analysis, including CryoEM, would be beneficial in confirming the mechanism of translocation proposed here to visualise DNA-protein interactions.

Activity of HelQ with RNA requires further exploration. A role of N-HelQ in substrate identification could be assessed by analysing N-HelQ mutations in the context of FL-HelQ. This would ascertain if N-HelQ activity is essential in interaction with RNA containing substrates. Furthermore, analysis of C-HelQ with different RNA containing substrates and the comparisons of C-HelQ ATPase activity between DNA and RNA could provide further evidence of a role of N-HelQ in substrate determination.

HelQ was not shown to displace the protein roadblocks tested here. However, there is opportunity to explore the role of partner proteins in HelQ activity. Chapter 5 discusses in detail a specific role for RPA in HelQ activation both in protein displacement and unwinding ability.

The phenotype observed for C-HelQ^{Y642A} suggested a different purpose for the human motif IVA then to Hel308. Analysis of C-HelQ^{Y642A} in protein assembly could be carried out using AGF and SEC MALS. Furthermore, the equivalent mutation in FL-HelQ would also be needed to assess for a role in stable dimerisation.

Understanding how helicases interact with DNA provide insights into specific helicase activities. Additionally, identifying the stage within the cell cycle a helicase may be detected can aid in understanding the complex system.

Chapter 5: Analysis of interactions between RPA and HelQ

5.1 Introduction.

Helicase activities in the cell are part of broader networks of nucleic acid processing pathways, in which the helicases interact functionally and physically with other proteins.[200] I showed evidence in chapter 4 that indicates against involvement of HelQ in displacement of DNA-bound proteins, dissociation of G-quadruplex or Holliday junctions, or DNA strand annealing (sections 4.8-4.10). However, it is likely that for HelQ to function in cells it may require partner proteins. HelQ has been shown to co-localise at exposed ssDNA with Rad51 and FancD2, two factors involved in homologous recombination repair of damaged forks.[253,256] Furthermore, proteomic analysis revealed links between HelQ and a Rad51 paralogues complex (BCDX2) and DNA damage response kinase ATR.[236] Association of HelQ with proteins of DNA repair, and the observation that genetic disruption of *HelQ* enhances cellular sensitivity to ICL-inducing agents, indicate a role for HelQ in HR and genome stability, potentially a pre-synaptic role of loading or removing Rad51.[236] Consistent with this, HelQ and the DNA damage responder RPA co-localise at cisplatin induced DNA damage *in vivo*.[256] RPA, an essential regulator of DNA replication and repair, has a large protein interaction pool but it's role with HelQ is unknown.[190] We hypothesise, based on previous work, that HelQ is recruited to exposed DNA by RPA, and sought to test this whilst establishing more knowledge about the interaction.

This chapter describes attempts made to analyse RPA recruitment of HelQ using purified recombinant proteins. The aim of the work was to (a) establish physical interactions between HelQ and RPA under different conditions *in vitro*; (b) assess the region of HelQ involved in these interactions and (c) determine the impact of RPA, in addition to recruitment, on HelQ. Work presented here provides

further evidence supporting a physical interaction between HelQ and RPA and a novel role of N-HelQ in RPA displacement.

5.2 Purification of RPA from *E. coli.*

Human RPA is a heterotrimer, which was purified from over-expression of the p11d-tRPA plasmid in. *E. coli*.[276,277] The expression of RPA protein from p11d-tRPA during bacterial transformation is toxic to *E. coli* cells and therefore measures were taken to overcome this, described in section 2.14.4.

RPA has a strong binding affinity for ssDNA making it ideal for its multiple roles in replication and repair.[62] RPA binds to DNA in different modes (chapter 1), a strong 30 nucleotide binding mode with a binding affinity (K_d) of 0.05 nM and a weaker 8 nucleotide binding mode with a K_d of 50 nM.[76] Using fluorescence anisotropy we detected similar DNA binding by RPA, evident from increased polarisation with increased RPA concentrations and a K_d value of 44 nM (figure 5.2 A). In EMSAs RPA loaded effectively onto the model fork 2b DNA substrate, evident from shifting of the Cy5 end-labelled substrate with increased RPA concentrations (figure 5.2 B). To a lesser extent the same was seen with RPA loading onto dsDNA (figure 5.2 B). Similar RPA-DNA binding patterns were observed as previously documented.[76] In addition, a strong affinity for ssDNA containing substrates confirmed that heterotrimeric RPA purified in this work was active and behaved as expected *in vitro*.

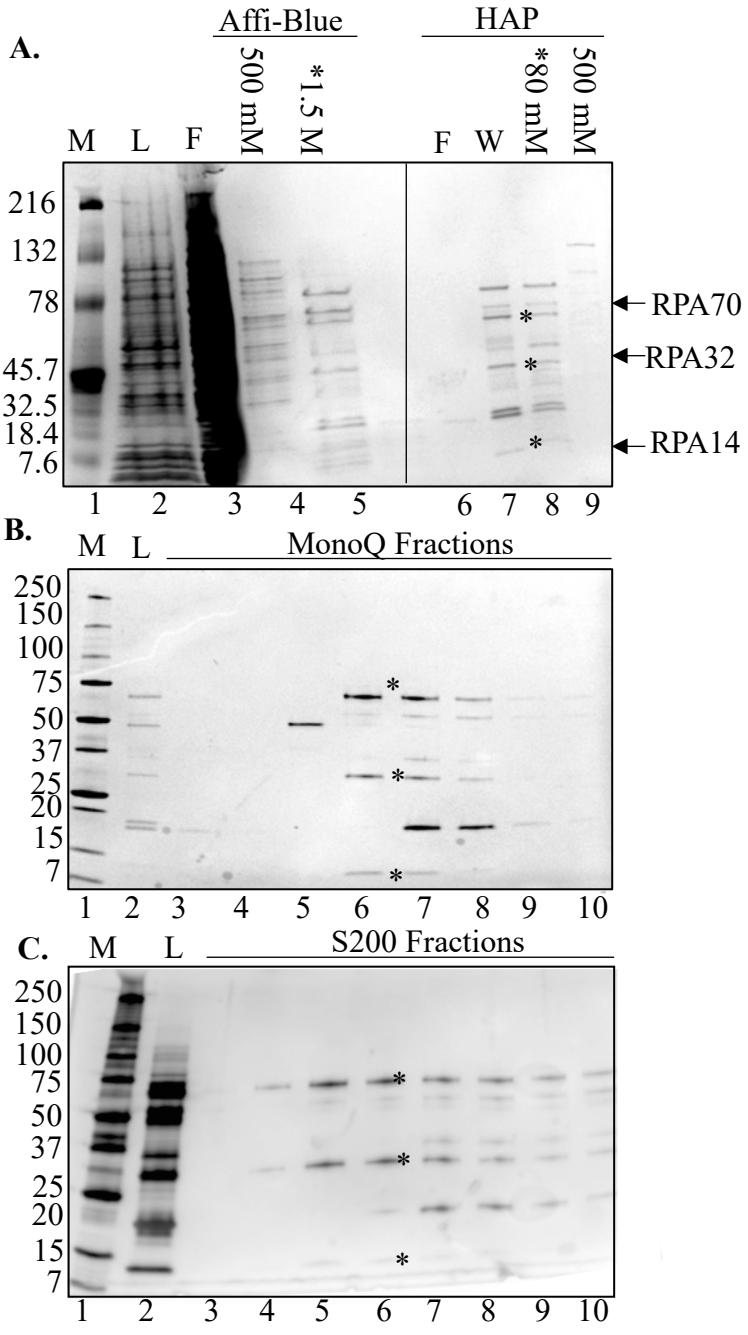


Figure 5.1 RPA protein purification. RPA was over-expressed from p11d-tRNA in *E. coli.*[2,3] Biomass containing RPA protein was sonicated and clarified prior to resolving RPA protein using affinity chromatography. The panels show Coomassie stained acrylamide gel for SDS PAGE analysis of RPA70, RPA32 and RPA14 as indicated (*). **A.** Affi-Blue gel chromatography showing the elution of RPA with 500 nM and 1.5 M sodium thiocyanate (NaSCN). The 1.5 M NaSCN fraction was further resolved by binding to hydroxyapatite (HAP) resin and eluted with potassium phosphate (KPO_4). **B.** Protein containing fractions were next resolved using MonoQ chromatography. Pure RPA protein was pooled and stored at -80°C. **C.** Impure RPA fractions were further separated using a superdex 200 column (S200). Legend= marker (M), clarified lysate (L), flow-through (F), wash-through (W).

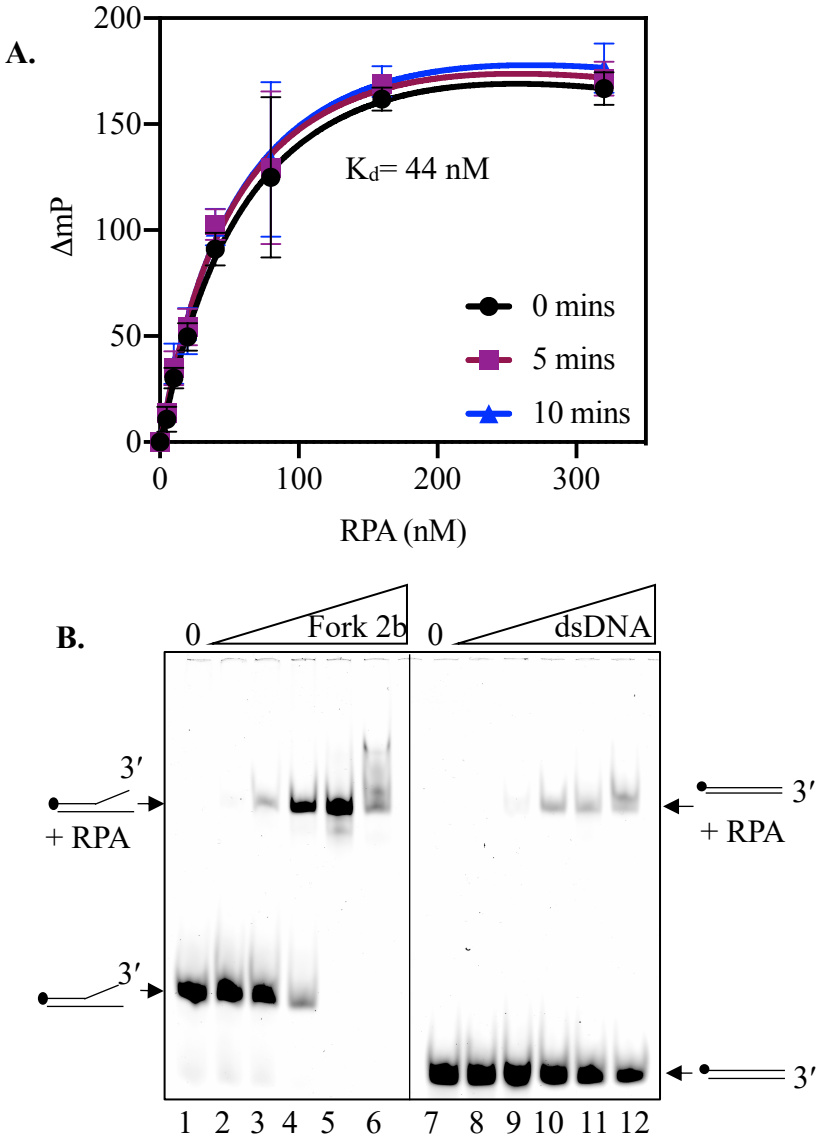


Figure 5.2 Fluorescence anisotropy measurements of ssDNA binding by purified RPA heterotrimer. RPA was used at 0, 5, 10, 20, 40, 80, 160 and 320 nM with 4 nM fluorescein labelled ssDNA. The change in polarization (mP) was measured at 0 (black), 5 (purple) and 10 minutes (blue). Reactions were carried out in triplicate and shown as standard error from the mean. **B.** Also shown are TBE gel EMSAs of RPA binding with fork 2b (25 nM) and dsDNA (25 nM). Schematics are shown with filled dots indicating 5′ Cy5 label. RPA loads onto fork 2b but has limited binding to dsDNA.

5.3 Recombinant FL-HelQ and RPA physically interact *in vitro.*

Co-localisation of HelQ with RPA at sites of damage induced exposed DNA *in vivo* suggests there may be a physical association at sites of accumulating ssDNA caused by replication blocks.[255] Furthermore, proteomic analysis has identified

associations between HelQ and RPA activated repair proteins, the Rad51 paralogue repair proteins (BCDX2 complex) and ATR.[236] We sought to understand more about the interaction between HelQ and RPA. Physical interactions between RPA and HelQ was analysed *in vitro* and the impact of activating components on the potential interactions. As part of this we wanted to dissect the HelQ fragments (chapter 3) and test them for interactions with RPA. This also generated new insights into the molecular mechanism of HelQ helicase function.

5.3.1 FL-HelQ and RPA interact in protein pull-down assays.

In vitro protein pull-down assays in this work utilized affinity tags on purified HelQ; hexa-histidine (hexaHis), SUMO and StrepII tags, see also section 2.18. The proteins interacted if untagged RPA co-eluted with tagged HelQ when the resin was washed in elution buffer. No interactions were isolated using the hexaHis-tag in protein pull-down assays because untagged RPA immobilised onto the resin when incubated in the absence of tagged HelQ. Therefore, elution of RPA with HelQ could not be differentiated from non-specific RPA-resin interactions. Non-specific RPA interactions may be as a result of patches of histidine residues causing RPA to bind the resin. Protein pull-down assays utilising the StrepII-tag identified physical interactions between tagged HelQ and untagged RPA. RPA and FL-HelQ eluted prematurely, prior to addition of D-desthiobiotin to displace StrepII-tagged HelQ, and were both detected in the wash fraction (figure 5.3). These results imply the formation of HelQ-RPA complexes shielded the 1 kDa StrepII-tag, located at the N-terminal of HelQ, preventing its immoblisation on the column. In addition, these results suggest RPA may interact locally to the StrepII-tag at the N-terminal region of HelQ.

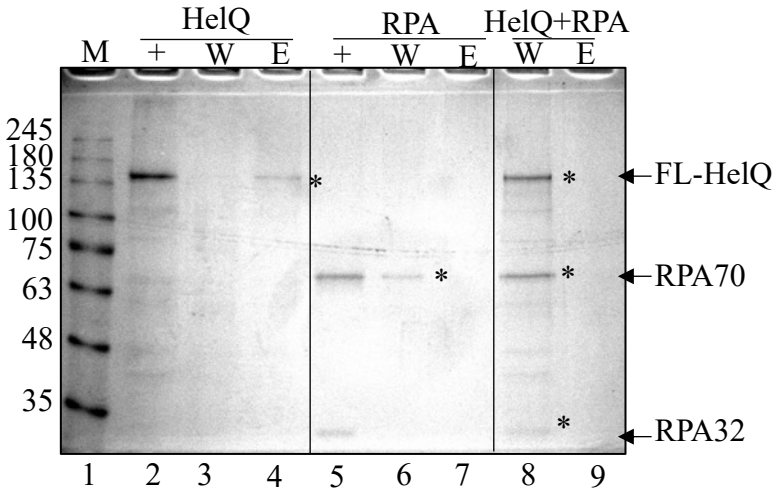


Figure 5.3 Protein pull-down assays show interactions between StrepII-tagged FL-HelQ and heterotrimeric RPA. Coomassie stained acrylamide gel SDS PAGE analysis showing HelQ immobilized onto anti-StrepII 0.2 mL mini column via interactions with its N-terminal StrepII-tag and eluted with 2.5 mM D-desthiobiotin (lanes 2-4). Untagged RPA alone did not interact with the column (lanes 5-7). Physical interactions between FL-HelQ and RPA are implied by HelQ being eluted from the resin by addition of RPA in the wash fraction (lanes 8-9). Legend= loaded protein (+), protein (*), wash-through (W), elution (E).

5.3.2 RPA and FL-HelQ interact on DNA in EMSAs.

The binding of RPA to DNA is important in determining the role RPA plays in recruiting downstream repair proteins. The binding of RPA to ssDNA dictates which proteins are recruited by RPA and the subsequent downstream repair pathway activated due to protein conformational changes.[190] Therefore, we assessed the role of DNA in RPA-HelQ interactions because RPA-DNA binding may be required for different downstream effects on HelQ to be observed. EMSAs were used to determine if HelQ interacted with RPA when in a DNA bound state (section 2.16.4). This would implicate RPA-DNA activation in HelQ recruitment to exposed ssDNA.

RPA was assembled onto a Cy5 labelled DNA fork substrate with a 3′ ssDNA flap (fork 3). Protein interactions were confirmed by the formation of super-shifted complexes with addition of FL-, C- and N-HelQ to RPA-DNA. The HelQ protein

fragments described in chapters 3 and 4 were added to RPA-DNA in these assays. FL-HelQ super-shifted with RPA-DNA evident from the formation of an additional fluorescently labelled complex not seen in RPA-DNA or FL-HelQ-DNA alone (figure 5.4 A lane 6). The additional complex observed suggests FL-HelQ and RPA interact on DNA.

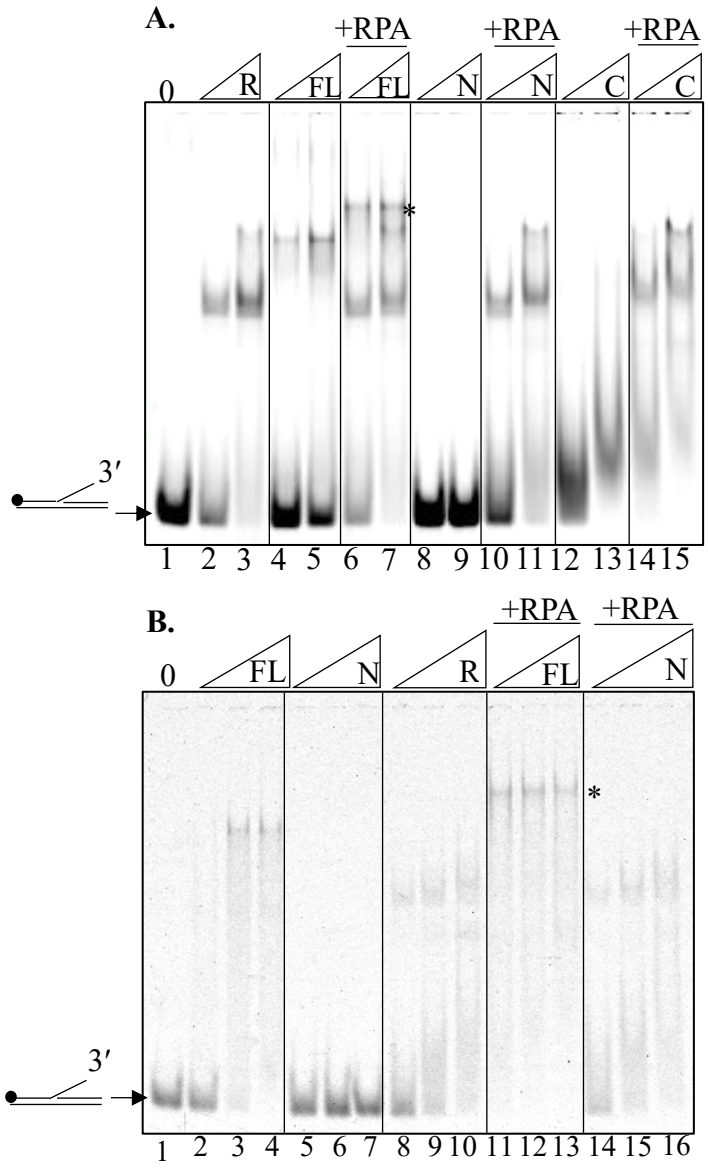


Figure 5.4 EMSAs show FL-HelQ super-shifts with heterotrimeric RPA bound to DNA. Acrylamide TBE gels to show that FL-HelQ, but not C-HelQ or N-HelQ, super-shifts with RPA (R; 120 nM) bound to fork 3 DNA (25 nM). HelQ proteins were used at 0, 80 and 160 nM. **A.** RPA was loaded onto DNA prior to addition of HelQ and **B.** RPA and HelQ were incubated prior to addition of DNA to give the same effect. Super-shifting observed is shown by * and a schematic of the fork substrate with filled dot indicating 5′ Cy5 label.

N-HelQ and RPA-DNA did not super-shift in EMSAs evident from no additional bands being detected (figure 5.4 A lane 11). The protein-DNA complexes observed in lanes including both RPA and N-HelQ were consistent with RPA-DNA only. Therefore, the addition of N-HelQ had no effect. N-HelQ does not load onto DNA (section 4.14) and therefore the lack of super-shifting implies N-HelQ cannot interact with RPA in a DNA-bound state. Similarly, C-HelQ did not super-shift with RPA-DNA as seen for FL-HelQ (figure 5.4 A lane 15). These results suggest that neither N-HelQ or C-HelQ independently can interact with RPA-DNA complexes *in vitro*.

Similar super-shifting of FL-HelQ and RPA bound to DNA was observed in EMSAs carried out where RPA and FL-HelQ were incubated prior to addition of DNA (figure 5.4 B). This may suggest that the presence of HelQ does not prevent RPA-DNA binding. EMSAs carried out with dsDNA, a less favourable substrate of both RPA and FL-HelQ, also showed similar super-shifting patterns (figure 5.5). I showed in chapter 4 (figure 4.16) that HelQ is unable to bind dsDNA and in figure 5.2 RPA weakly binds dsDNA. Therefore, super-shifting observed in figure 5.5 suggests HelQ is interacting with RPA when bound to DNA and HelQ is not interacting with DNA independent of RPA. This was used as a control to confirm that the proteins were not binding independently to the DNA but forming a complex on DNA. FL-HelQ and RPA also super-shifted in EMSAs carried out with the abasic 3 fork containing an internal abasic site (figure 5.5).

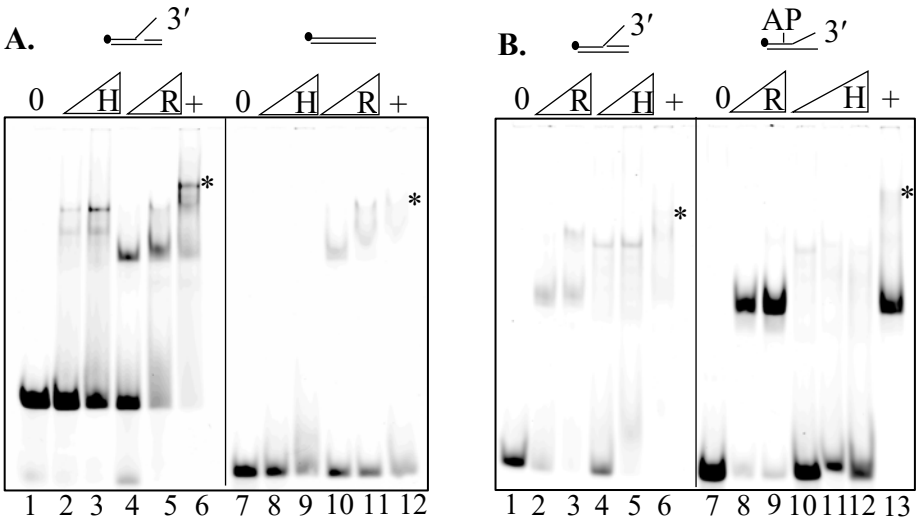


Figure 5.5 EMSAs show FL-HelQ super-shifts with RPA bound to dsDNA and chemically modified DNA. Representative EMSA acrylamide TBE gels to show that HelQ (H), used at 0, 80 and 160 nM, super-shifts with trimeric RPA (R; 120 nM) on **A.** dsDNA (25 nM) and **B.** abasic 3 fork (25 nM). Super-shifting on fork 3 DNA was used as a control. Schematics show the substrates tested with filled dots indicating Cy5 labels.

RPA was substituted with the *E. coli* ssDNA binding protein (SSB) as a control. No super-shifting was observed for FL-HelQ with SSB-DNA evident from no additional bands being detected (figure 5.6 A). This provides evidence that HelQ-RPA-DNA complexes are due to HelQ-RPA interactions and not HelQ-DNA interactions.

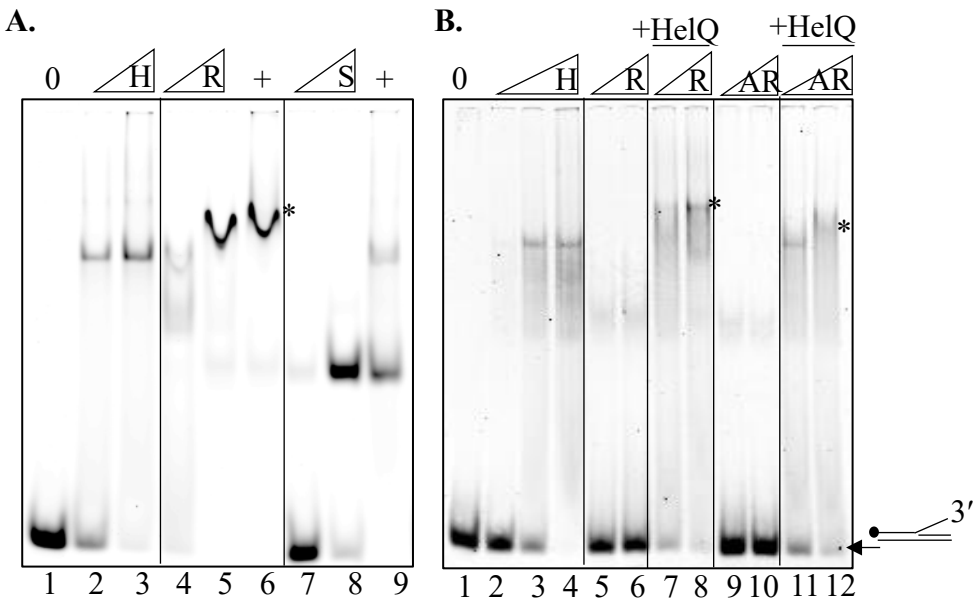


Figure 5.6 EMSAs show no super-shifting of FL-HelQ with *E. coli* SSB bound to DNA. Control EMSA acrylamide TBE gel to show the absence of super-shifting between FL-HelQ (H), used at 80 and 160 nM, and **A.** SSB (S), used at 0, 40 and 80 nM, bound to fork 3 (25 nM). **B.** Also shown are super-shifting EMSAs carried out with Aro1-RPA (AR), a DNA binding mutant of RPA used at 0, 60 and 120 nM. Aro1-RPA and FL-HelQ (H) formed an unstable super-shifted complex that ran differently to the additional complex with wild type RPA-DNA and HelQ evident from DNA-protein smearing.

A DNA binding mutant of RPA, Aro1-RPA, was used as an additional control in super-shifting EMSAs (see table 2.8).[374] Aro1-RPA has two residue substitutions, Phe-238-Ala and Trp-361-Ala, reducing its DNA binding ability (see table 2.6).[374] In these assays, Aro1-RPA had reduced DNA binding affinity as expected (figure 5.6 lanes 9-10). However, the positive control wild type RPA as shown in this gel did not bind to DNA as observed previously (figure 5.6 lanes 5-6 compared to figure 5.5 lanes 3-4). Wild type RPA still resulted in super-shifting with HelQ (figure 5.6 B lanes 7-8). However, Aro1-RPA with DNA and HelQ did not show the formation of a stable complex compared to wild type RPA with HelQ and band-shifting is consistent with HelQ-DNA only (lanes 3-4 compared to 11). Unstable complex formation between Aro1-RPA and HelQ on DNA is evident from Cy5 DNA smearing different to the HelQ-DNA complexes (lanes 12). This

suggests that RPA-DNA interactions are important for the stable effect observed and it is not a result of RPA interacting with HelQ bound to DNA. From these results we hypothesised that the impact, potentially recruitment, of RPA on HelQ requires initial RPA-DNA binding. In the absence of super-shifting for N-HelQ and C-HelQ with RPA-DNA in EMSAs (figure 5.4) we were unable to determine the region of HelQ responsible for RPA interactions. Therefore, we aimed to look for physical interactions between RPA and HelQ fragments in pull-down assays.

5.3.3 N-HelQ and C-HelQ, when isolated, do not interact with RPA in pull-down assays.

Protein pull-down assays as described for FL-HelQ in section 5.3.1 were carried out between RPA with N-HelQ and C-HelQ to assess for physical interactions. We predicted that N-HelQ may be involved in protein-protein interactions due to its lack of DNA processing activity shown in chapter 4. However, no interactions between untagged RPA and tagged N-HelQ were detected using either hexaHis-tag (figure 5.7 A) or StrepII-tag (figure 5.7 B) protein pull-down assays. Similar problems with RPA non-specific binding to the hexaHis-tag resin were observed. This is evident from RPA eluting from the beads in the elution fraction (figure 5.7 A). A range of conditions were tested to test for an interaction including the presence of magnesium and DNA, this had no impact on elution patterns of N-HelQ and RPA. In StrepII-tag pull-down assays RPA eluted in the wash fraction and N-HelQ remained bound to the resin via its StrepII tag until eluted with D-desthiobiotin. This suggests that RPA was not interacting with N-HelQ. Addition of ssDNA had no impact on protein-protein interactions. Furthermore, hexaHis-tagged pull-down assays did not show a physical interaction between RPA and phosphomimetic N-HelQ$^{S158/178D}$ (figure 5.8). While RPA non-

specific binding impacted the result, the elution of RPA32 in the wash fraction suggested no protein interactions between N-HelQ$^{S158/178D}$ and trimeric RPA. These results suggest N-HelQ does not interact with RPA *in vitro.*

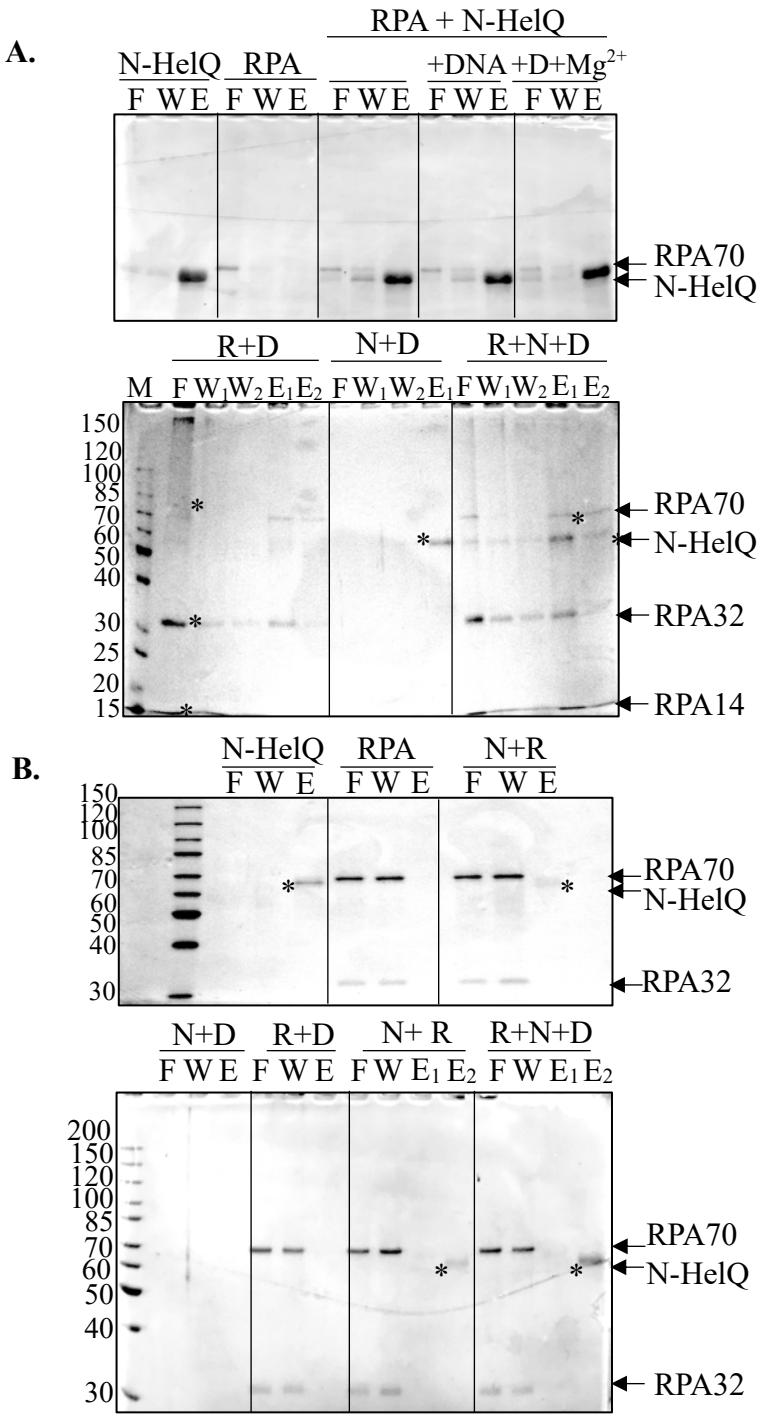


Figure 5.7 Protein pull-down assays show N-HelQ does not interact with heterotrimeric RPA. Coomassie stained 8% acrylamide gels showing SDS PAGE analysis of **A.** hexaHis-tag N-HelQ and RPA protein incubated and eluted with imidazole and **B.** StrepII-tag N-HelQ with untagged RPA incubated and eluted with D-desthiobiotin protein pull-down assays. Assays were carried out with tagged N-HelQ (N) alone, trimeric RPA (R) alone and in combination. Assays were also run with addition of 25 nM ssDNA (D) and 5 mM $MgCl_2$ (Mg^{2+}) as indicated. Legend= flow-through (F), wash-through (W), elution (E), protein (*).

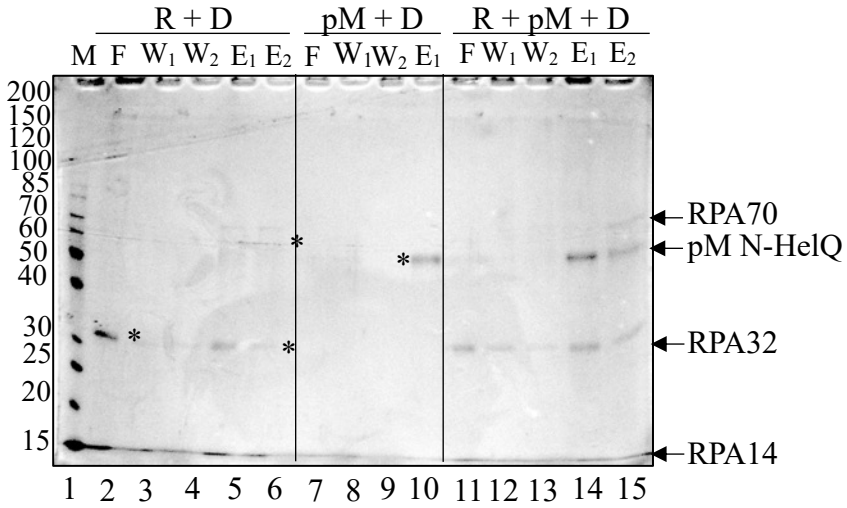


Figure 5.8 An attempt to show that RPA does not interact with phosphomimetic N-HelQ in hexaHis-tag protein pull-downs. Coomassie stained 8% acrylamide gels showing SDS PAGE analysis of hexaHis tagged phosphomimetic N-HelQ $^{S158/178D}$ (pM) protein pull-down with untagged heterotrimeric RPA (R). Assays were also carried out in the presence of 25 nM ssDNA (D). RPA70 non-specific binding to the anti-hexaHis column resulted in elution of RPA with imidazole. However, detection of RPA32 in the flow through suggests phosphomimetic N-HelQ does not interact with trimeric RPA. Legend= flow-through (F), wash-through (W), elution (E), protein (*).

Super-shifting EMSAs were carried out between RPA and phosphomimetic N-HelQ$^{S158/178D}$ to confirm observations reported in chapter 4 (figure 4.35) and figure 5.8 that N-HelQ is not activated when phosphorylated at these sites. Super-shifting was not observed between phosphomimetic N-HelQ$^{S158/178D}$ and RPA-DNA evident from no additional complexes being detected (figure 5.9). Combined, these results confirm that phosphorylation at these sites does not stimulate protein-protein interactions between N-HelQ and RPA.

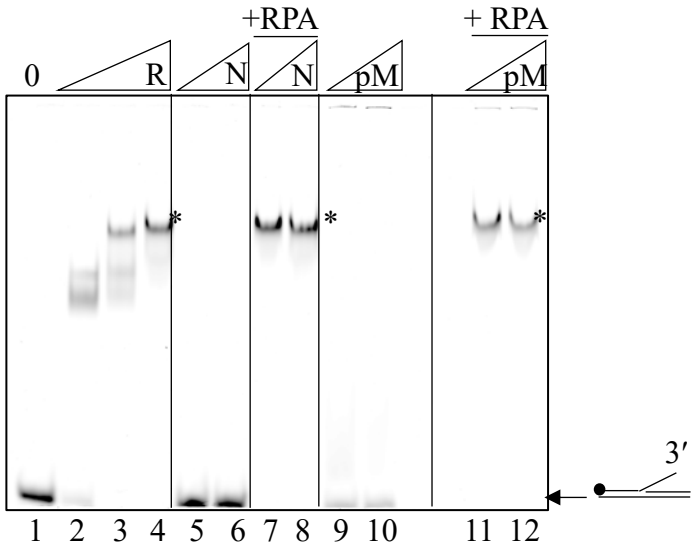


Figure 5.9 EMSAs show phosphomimetic N-HelQ does not super-shift with RPA bound to DNA. EMSA acrylamide TBE gel to show phosphomimetic N-HelQ$^{S158/178D}$ (pM), used at 0, 80 and 160 nM, does not super-shift with heterotrimeric RPA (R; 120 nM) bound to fork 3 DNA (25 nM). Also shown is N-HelQ with RPA-DNA used as a control for no super-shifting. Schematic shows DNA substrate with filled dot indicating Cy5 label.

HexaHis-tag protein pull-down assays were also carried out to assess for interactions between C-HelQ and RPA. However, as observed for pull-down assays with FL-HelQ, RPA non-specific binding to the hexaHis resin limited the detection of physical interactions. RPA co-eluted with C-HelQ after addition of imidazole in the presence and absence of ssDNA (figure 5.10). This result may be caused by RPA non-specific adherence to the column and not due to physical interactions between C-HelQ and RPA. However, evidence presented in super-shifting EMSAs (figure 5.4 A lanes 12-15) implied C-HelQ, in the truncated state used here, does not interact with RPA bound to DNA. However, we cannot rule out that C-HelQ is responsible for interactions with RPA. C-HelQ throughout the work carried out here has been shown to be highly unstable. Therefore, C-HelQ when assessed alone and in the absence of N-HelQ, may be too unstable to detect any interactions with RPA. Further analysis discussed later in this chapter is required to determine the presence of interactions specific to C-HelQ. This suggests C-HelQ does not interact with RPA and therefore elements in both N- and C-HelQ are required to stably interact with RPA *in vitro*.

Observations from protein pull-down assays and super-shifting EMSAs confirmed FL-HelQ and RPA physically interact with and without DNA. This interaction was not isolated to a specific region of HelQ in work carried out here, but it is unlikely to rely on interactions solely within N-HelQ.

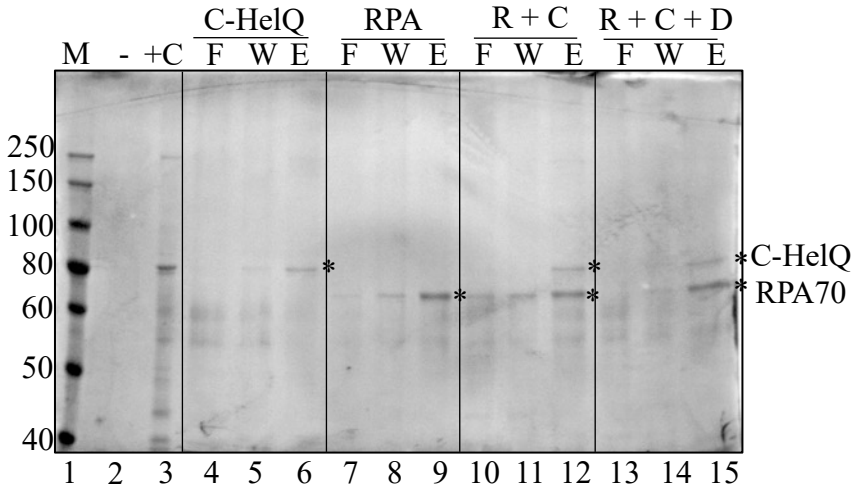


Figure 5.10 An attempt to assess interactions between C-HelQ and heterotrimeric RPA in hexaHis protein pull-down assays. Coomassie stained acrylamide gels in an attempt to show SDS PAGE analysis of protein interactions between C-HelQ (C) and trimeric RPA (R). Assay was also carried out in the presence of ssDNA (D). RPA showed non-specific binding to the hexaHis column and eluted with imidazole. RPA70 and C-HelQ co-eluted implying that interactions were caused by RPA non-specific binding. Legend = Marker (M), flow-through (F), wash-through (W), elution (E) and protein (*).

5.4 Analysis of the role of RPA in the HelQ-RPA complex.

We hypothesise that RPA recruits HelQ to exposed DNA that is known to accumulate at compromised replication forks. However, any downstream impact RPA has on HelQ activity after recruitment is unknown. RPA has a large pool of effector proteins and therefore it would be unsurprising for RPA to impact HelQ activity. For example, RPA may assemble HelQ into complexes for DNA repair, activate unwinding or act as a protein partner to the IDPR of N-HelQ (see chapter 3 section 3.10). The next section assesses the potential impacts RPA may have on HelQ activity.

5.4.1 RPA can stimulate HelQ to unwind through a BamHI barrier on DNA.

Data presented in chapter 4 (section 4.10) showed that FL-HelQ could not translocate through a protein barrier bound to fork DNA. As discussed in chapter 1, RPA is involved in recruitment of downstream proteins to activate repair pathways and has been shown to activate the helicases FancJ and RecQ1 to displace a protein barrier and unwind forked DNA.[62,374] This suggests that HelQ may also require activation by RPA to unwind these protein-bound DNA fork structures. A role of RPA with HelQ was assessed in the presence of the roadblocks BamHI^{EIIIA}, dCas9, streptavidin-biotin and RNAP roadblocks as discussed in chapter 4. FL-HelQ was able to unwind through the BamHI^{EIIIA} bound fork DNA in the presence of RPA, evident from the formation of Cy5 ssDNA (figure 5.11 A-B). These results initially implied RPA activated FL-HelQ to displace BamHI^{EIIIA} and be able to unwind the fork. The presence of RPA recovered HelQ unwinding activity to the rate observed in the absence of BamHI^{EIIIA} (figure 5.11 A). *E. coli* SSB was used as a control to confirm the effect observed was specific to RPA (figure 5.11 B). No unwinding activity was observed by FL-HelQ in the presence of SSB suggesting BamHI^{EIIIA} was not displaced. Unwinding was also not observed with Aro1-RPA which confirmed the effect was specific to RPA and required RPA binding to DNA (figure 5.11 B). N-HelQ was used in the assay to determine if observed unwinding was specific to HelQ and not a result of protein overcrowding leading to spontaneous substrate dissociation. No unwinding was observed in the presence of N-HelQ and RPA with the BamHI^{EIIIA} bound DNA. These controls suggested that RPA activated FL-HelQ to displace the BamHI^{EIIIA} and unwind the fork substrate.

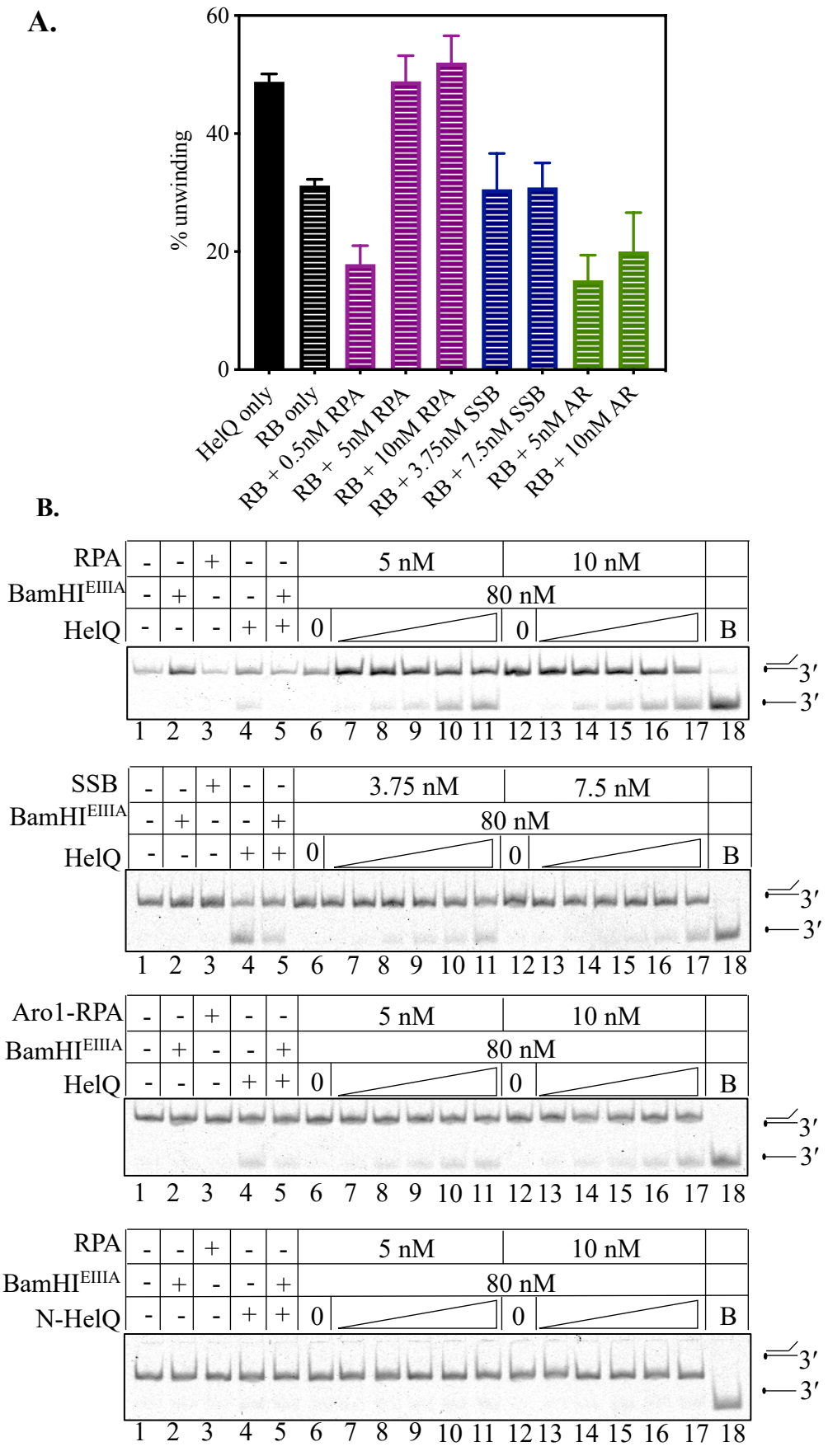

A.
% unwinding
60
40
20
0
HelQ only
RB only
RB + 0.5nM RPA
RB + 5nM RPA
RB + 10nM RPA
RB + 3.75nM SSB
RB + 7.5nM SSB
RB + 5nM AR
RB + 10nM AR
B.
RPA
BamHI^EIIIA
HelQ
5 nM
10 nM
80 nM
0
B
3′
3′
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
SSB
BamHI^EIIIA
HelQ
3.75 nM
7.5 nM
80 nM
Aro1-RPA
BamHI^EIIIA
HelQ
RPA
BamHI^EIIIA
N-HelQ

Figure 5.11 Protein displacement assays show RPA activates HelQ to bypass BamHI^{EIIIA} and unwind forked DNA. A. A measurement of HelQ unwinding activity of a BamHI^{EIIIA} bound fork substrate (25 nM) in the presence of RPA and control proteins. **B.** Also shown are representative TBE gels run in triplicate to present data shown here as standard error from the mean. Endpoint HelQ (160 nM) unwinding assays with 80 nM BamHI^{EIIIA} (RB) bound DNA (25 nM) were carried out with RPA (pink), SSB (blue) and Aro1-RPA (AR; green). RPA was used at 0.5, 5 and 10 nM; SSB was used at 3.7 and 7.5 nM and Aro1-RPA was used at 5 and 10 nM. B represents a boil control to show dissociated substrates with filled dots indicating 5′ Cy5 label.

Enzyme protection assays were used to assess if RPA was able to displace BamHI^{EIIIA} from DNA in the absence of HelQ (section 2.17.2). A lack of wild type BamHI endonuclease activity was used to confirm the presence of BamHI^{EIIIA} bound to DNA. There was less BamHI endonuclease activity in the presence of BamHI^{EIIIA} evident from intact fork (figure 5.11 lane 5). Endonuclease activity was not recovered when RPA was added (figure 5.11 lane 7). This suggests RPA did not displace BamHI^{EIIIA} to allow access for BamHI cutting and therefore confirmed that HelQ is necessary for BamHI^{EIIIA} displacement from DNA. This implies that RPA activates HelQ protein displacement activity and that it is not RPA alone displacing the barrier.

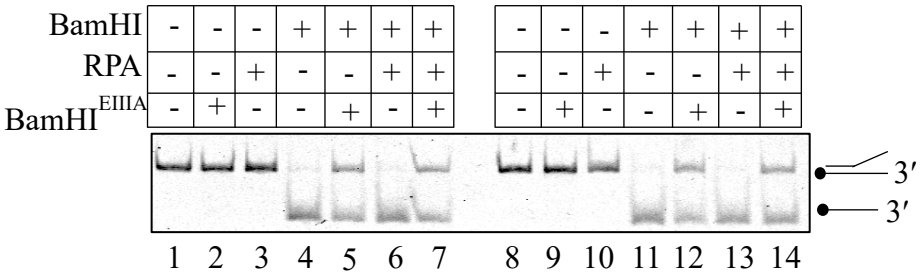


Figure 5.12 Restriction enzyme protection assays show RPA cannot displace BamHI^{EIIIA} from DNA in the absence of HelQ. TBE gel representation showing wild type BamHI (2U) endonuclease activity of fork DNA (25 nM) in the presence of the BamHI^{EIIIA} (180 nM) where digestion is reduced and in the presence of BamHI^{EIIIA} and RPA (20 nM). RPA does not displace BamHI^{EIIIA} to allow for BamHI digestion of exposed DNA. Schematics to shows intake DNA fork and digested Cy5 end labelled ssDNA (filled dot).

Displacement activity by HelQ in the presence of RPA was not observed with dCas9 and RNAP protein DNA barriers. These assays were carried out alongside analysis of HelQ displacement alone and can be seen in chapter 4. RPA was preincubated with the RNA polymerase bound DNA (figure 4.20) and dCas9 bound DNA (figure 4.21) prior to addition of HelQ. Unwinding of the DNA substrate by HelQ which would suggest displacement of the barrier was not observed in either assay. Furthermore, RPA did not aid HelQ to displace streptavidin from biotinylated DNA. The strong binding affinity (K_d= 10^{-15} M) of streptavidin compared to BamHI^{EIIIA} (K_d= 5.6^{-10} M) meant that streptavidin was not displaced by any of the trialled helicases and therefore no displacement with addition of RPA was unsurprising.[361,362]

Protein displacement from DNA by the combined activity of HelQ and RPA observed was specific to the BamHI^{EIIIA} barrier. These results imply that RPA does not have a prominent role in activating HelQ activity through protein bound DNA. We hypothesised that the effect observed in figure 5.11 was an anomaly caused by conformational changes in the DNA upon RPA binding weakening BamHI^{EIIIA}-DNA interactions allowing for bypass and unwinding by HelQ. The displacement effect observed was difficult to confirm as real because BamHI^{EIIIA} is not a physiologically relevant barrier that could be considered an interesting exception. To exhaust the idea that RPA activates HelQ to displace protein barriers on the DNA, more biologically relevant examples are required.

5.4.2 RPA does not inhibit, but does not improve, FL-HelQ unwinding activity.

Data presented in chapter 4 showed that HelQ is unable to unwind dsDNA, HJs or an extended fork substrate (figures 4.15 and 4.16). Unwinding assays were carried out in the presence of RPA to assess if RPA improved the ability of HelQ to process these intermediate DNA substrates. The assembly of RPA on dsDNA

and an extended fork substrate had no impact on HelQ unwinding evident from no increase in Cy5 ssDNA (figure 5.13 A-B). Assembly of RPA onto a synthetic HJ resulted in increased substrate dissociation in comparison with no RPA (figure 5.13 C). However, HJ dissociation was also observed by RPA in the absence of FL-HelQ. This suggests that RPA, and not HelQ unwinding activity, resulted in HJ dissociation. Therefore, RPA does not stimulate HelQ to process unfavoured intermediate DNA substrates *in vitro*.

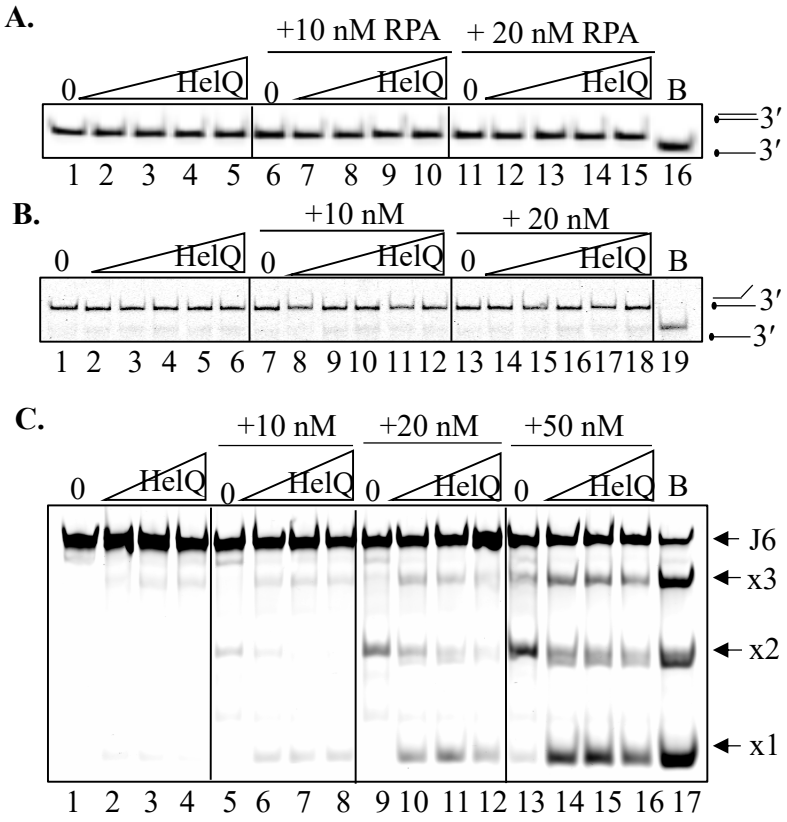


Figure 5.13 HelQ does not unwind intermediate DNA structures in the presence of RPA. TBE gel representation to show HelQ unwinding activity of intermediate DNA structures (25 nM) in the presence of RPA used at 10, 20 and 50 nM. FL-HelQ, used at 0, 40, 80 and 160 nM, could not unwind **A.** dsDNA, **B.** extended fork substrate and **C.** synthetic HJ DNA. B represents a boil control to show dissociated substrates as shown by schematics. Filled dot indicate Cy5.

5.4.3 RPA does not improve the rate of FL-HelQ unwinding.

The effect of RPA on HelQ unwinding of DNA fork structures was assessed in a similar way as described in section 5.4.2. The presence of RPA did not improve

the amount of fork unwound to form Cy5 ssDNA by HelQ in comparison with HelQ unwinding alone (figure 5.14 A). Similarly, RPA did not improve HelQ unwinding through the abasic 3 fork substrate with an internal abasic site (figure 5.14 B). Unwinding assays as a function of time were carried out to assess the effect of RPA on the rate of unwinding by HelQ of forked DNA. RPA did not affect the rate of HelQ unwinding over 30 minutes when compared to unwinding with no RPA (figure 5.14 C). Results presented here suggests RPA does not increase HelQ unwinding efficacy, however, the presence of RPA does not prohibit HelQ unwinding activity of these substrates, it does not act as a roadblock despite its higher binding affinity for DNA.

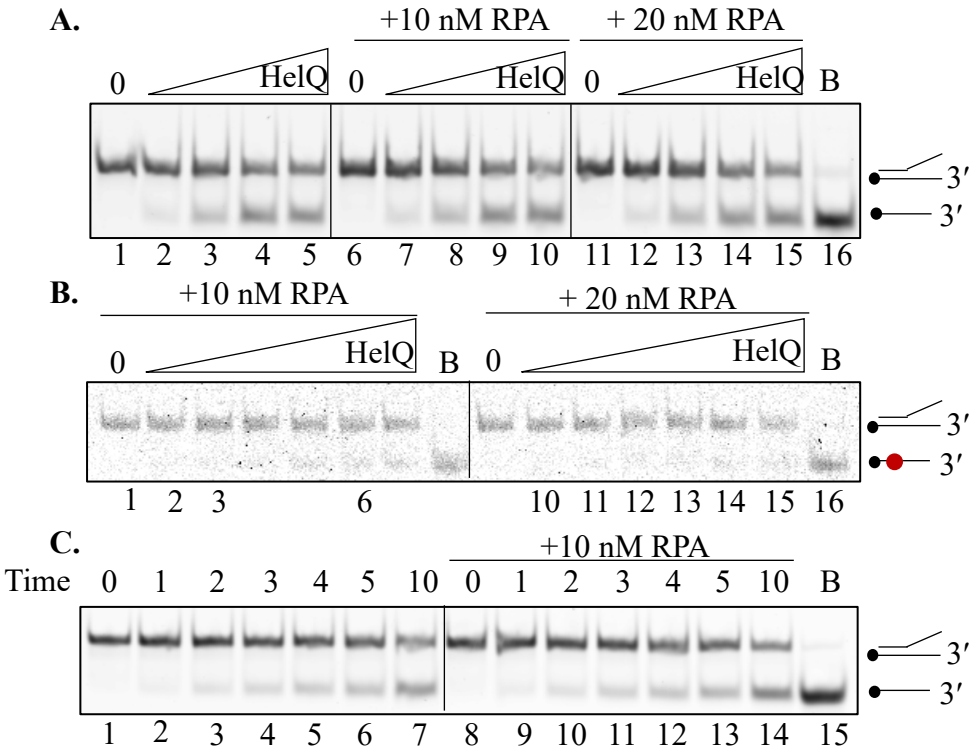


Figure 5.14 Helicase assays show RPA does not inhibit HelQ unwinding activity over time. Acrylamide TBE gels to show unwinding by HelQ, used at 0, 10, 20, 40, 80 and 160 nM of **A.** fork 2b DNA and **B.** abasic 3 fork with an internal AP site (25 nM) in the presence of RPA. **C.** Also shown are HelQ (80 nM) helicase assays of fork 2b (25 nM) as a function of time which are unaffected in the presence of RPA (100 nM). Reactions were terminated at 0, 1, 2, 3, 4, 5 and 10 minutes. B represents a boil control to show dissociated substrates illustrated by schematics.

5.4.4 Nuclease protection assays show FL-HelQ displaces RPA from DNA.

Data presented here implies that RPA does not stimulate the rate of DNA unwinding by HelQ or enzyme processivity. Therefore, we next assessed the effect of HelQ on RPA activity. We used S1 nuclease activity assays to assess if addition of HelQ to RPA-ssDNA complexes exposed ssDNA for nuclease digestion (see section 2.18.6). S1 nuclease digests ssDNA evident from no detectable Cy5 labelled ssDNA (figure 5.15 lane 4). RPA formed a complex with ssDNA (lanes 2-3) and with addition of S1 nuclease RPA-DNA complexes were still evident (lanes 5-6). This demonstrates protection of the DNA by RPA. Further addition of activated HelQ (+ATP-Mg^{2+}) resulted in partial recovery of S1 nuclease digestion of ssDNA (lanes 15-16). This suggests HelQ is able to displace the RPA from ssDNA exposing the ssDNA to digestion. The DNA would not be protected from S1 nuclease digestion by activated HelQ (+ATP-Mg^{2+}) due to the nature of activated HelQ continually loading, translocating and dissociating from the DNA (lanes 7-8). Displacement was not observed in the absence of ATP-Mg^{2+} suggesting the activity relies on the ability for HelQ to translocate on the DNA or be in an active state. This allowed us to hypothesise that activated HelQ may be involved in the displacement of RPA from DNA.

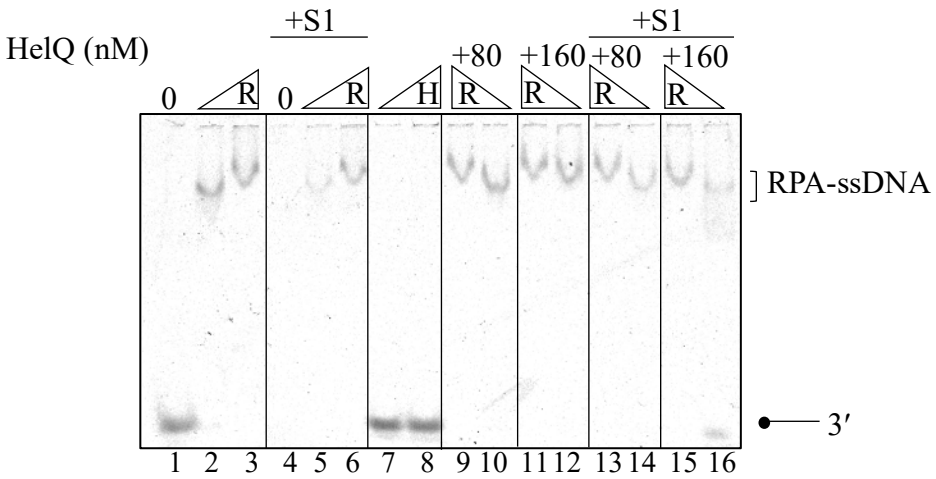


Figure 5.15 S1 nuclease protection assays to show weak displacement of RPA from DNA by HelQ. Representative TBE EMSA gels show S1 nuclease digestion of ssDNA (25 nM) in the presence of RPA (R) and HelQ (H). RPA was used at 12.5 and 25 nM and HelQ was used at 80 and 160 nM. RPA bound to the ssDNA protecting the ssDNA from S1 nuclease digestion. Addition of HelQ and 5 mM ATP and $MgCl_2$ resulted in partial recovery of S1 nuclease activity implying the displacement of RPA. Schematic shows ssDNA and filled dot indicates Cy5 label.

5.5 Analysis of N-HelQ with RPA.

5.5.1 RPA displacement is localised to N-HelQ.

Results presented in figures 5.15 and 5.3 suggest that HelQ may be involved in displacement of RPA from DNA and that HelQ may interact with RPA locally to the N-terminal StrepII-tag. N-HelQ was highlighted as a potential region of interest for activity with RPA because of the predicted regions of disorder and requirement of activating partners (section 3.10). This required further analysis and therefore we next looked at the relationship between RPA and N-HelQ. EMSAs were carried out with N-HelQ and RPA-DNA despite no observed physical interactions in pull-down assays.

N-HelQ displaced RPA from ssDNA evident from the increase in free Cy5 fork DNA being observed in EMSAs (figure 5.16 A-B). High concentrations of N-HelQ

(500 nM) were required for this effect to be observed which suggests further activation of N-HelQ may be required for this activity to be more efficient, potentially through post translational modifications. N-HelQ $^{D142A/F143A}$ was unable to displace RPA from DNA suggesting the predicted PWI-like domain is essential for displacement activity (figure 5.16 C). Furthermore, a truncated N-HelQ variant without the PWI-like domain and phosphomimetic N-HelQ$^{S158/178D}$, were unable to displace RPA from DNA (figure 5.16 D-E). This further confirmed the PWI-like domain is central to the displacement activity observed and implicates the PWI-like domain in protein-protein interactions. As a control, N-HelQ was unable to displace *E. coli* SSB bound to DNA confirming the displacement was specific to RPA-DNA (figure 5.16 F). In addition, RPA was not displaced from DNA by BSA confirming displacement was not a result of protein over-crowding (figure 5.16 G).

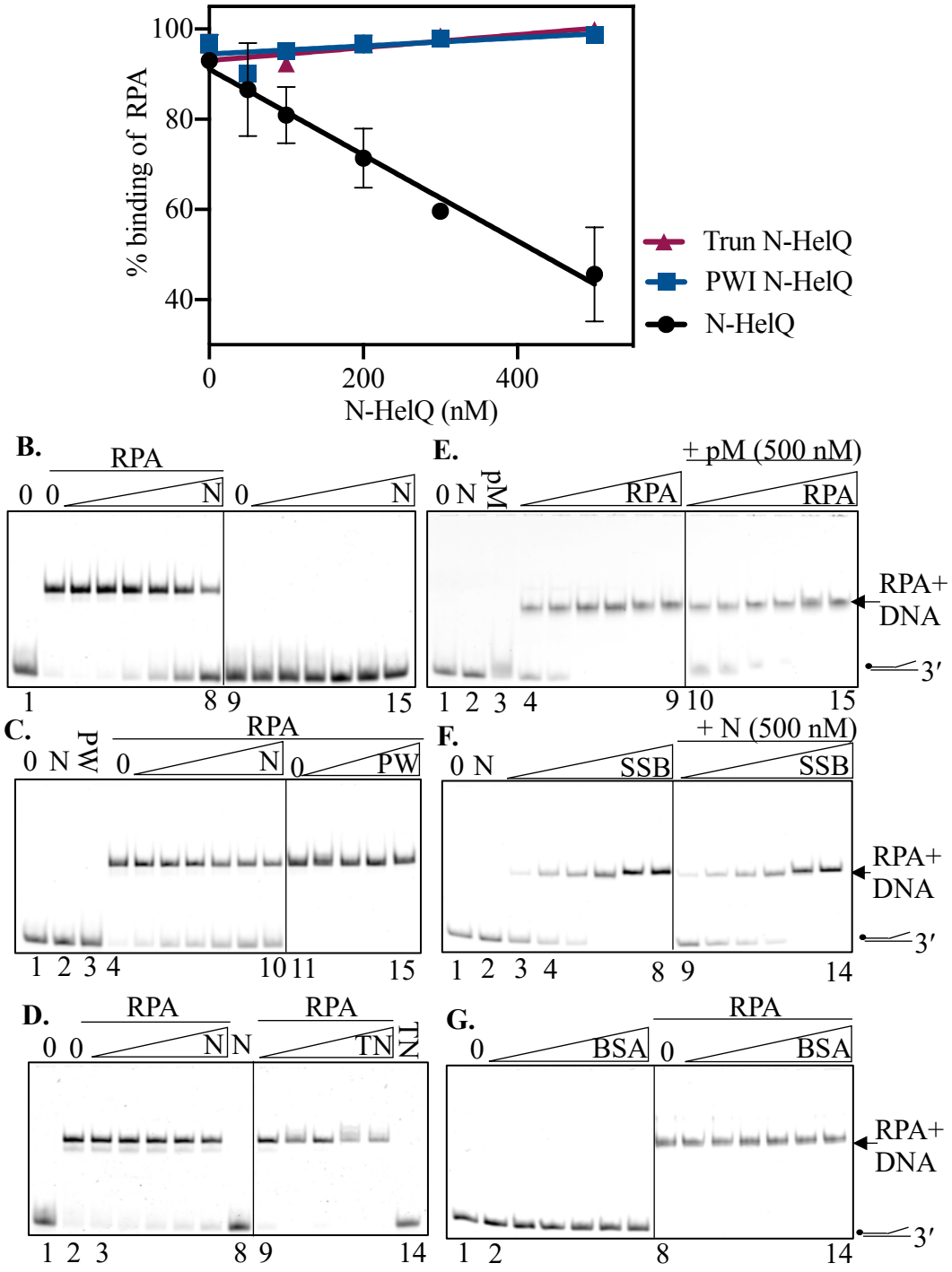


Figure 5.16 Protein displacement EMSAs show N-HelQ displaces RPA from ssDNA. A. A measurement of the impact of N-HelQ on RPA (12.5 nM) bound to fork 2b (25 nM) DNA. Also shown are EMSA TBE gels carried out in triplicate to generate the data shown as standard error from the mean. RPA displacement was observed by **B.** N-HelQ (N) but not **C.** PWI-like mutant N-HelQ$^{D142A/F143A}$ (PW), **D.** truncated N-HelQ (Val-76) (TN) or **E.** phosphomimetic N-HelQ $^{S158/178D}$ (pM). Protein was used at 0, 25, 50, 100, 200, 400 and 500 nM. Schematics show DNA substrates with filled dots indicating 5′ Cy5 label. **F.** N-HelQ did not displace *E. coli* SSB (0-300 nM) and **G.** BSA (0-500 nM) did not displace RPA from fork 2b.

Displacement EMSAs as a function of RPA concentration were carried out to assess whether N-HelQ was still able to displace large quantities of RPA from the DNA. RPA formed two distinct complexes on the DNA when used at higher concentrations (figure 5.17 B lanes 2-7). This suggests that the slower migrating band has more molecules of RPA that are interacting with each DNA molecule. Addition of N-HelQ resulted in more RPA-DNA complexes running at the faster migrating complex (lanes 9-12). This suggests that N-HelQ is able to displace RPA from the DNA at high concentrations (figure 5.17 B) as well as at low concentrations (figure 5.17 A). Therefore, N-HelQ displacement activity is not affected by the concentration of RPA.

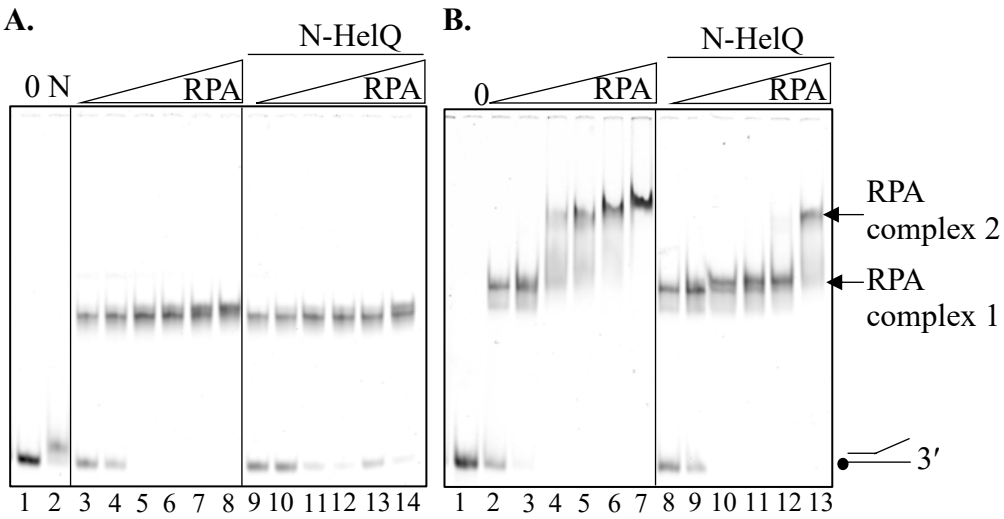


Figure 5.17 EMSAs show that N-HelQ can displace high concentrations of RPA from DNA. Representative EMSA TBE gels to show N-HelQ (500 nM) can displace RPA from fork 2b (25 nM) used at **A.** 0, 1, 2, 5, 10 and 20 nM and **B.** at 0, 7, 15, 30, 60 and 120 nM. RPA formed two complexes on DNA as indicated to the right of the gel. Schematics show DNA substrates with filled dot indicating 5′ Cy5 label.

5.5.2 Fluorescence anisotropy confirmed N-HelQ displaces RPA from DNA.

Fluorescence anisotropy was carried out to measure the binding affinity of RPA to DNA in the presence and absence of N-HelQ (see section 2.21.7). Addition of N-HelQ resulted in reduced binding of RPA for DNA indicated by reduced

change in fluorescence polarisation compared to RPA alone (figure 5.18 A). This suggests that N-HelQ displaced RPA from the DNA. N-HelQ is unable to interact with DNA and therefore the decrease in polarisation is indicative of RPA interactions with the fluorescently labelled DNA. The presence of N-HelQ resulted in reduced fluorescence polarisation (mP) of RPA bound to DNA being detected when compared to without N-HelQ (figure 5.18 A). BSA was used as a negative control and did not result in a change in fluorescence polarisation being detected for RPA and DNA (figure 5.18 B). This suggests the displacement effect observed is specific to N-HelQ and not as a result of protein over-crowding. Kd values were not calculated for these results due to the absence of saturation between RPA and DNA being reached. However, a change in fluorescence polarisation indicates that N-HelQ impacts RPA-DNA binding.

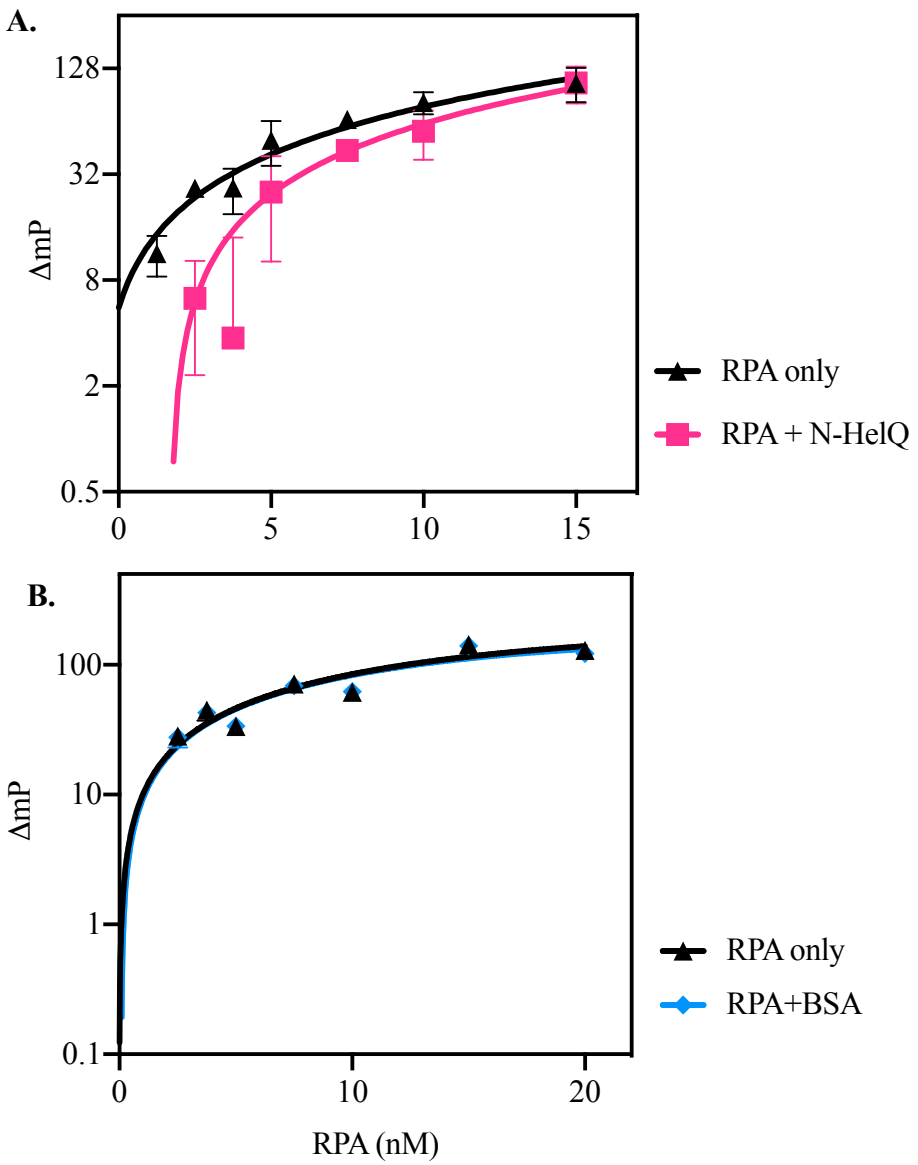


Figure 5.18 Fluorescence anisotropy shows RPA-DNA binding affinity is reduced in the presence of N-HelQ. A measurement of RPA binding to fluorescein labelled fork 2b DNA (5 nM) in the presence of **A.** N-HelQ (500 nM) and **B.** BSA (500 nM). RPA was used at 0, 2.5, 3.75, 5, 7.5, 10, 15 and 20 nM. RPA-DNA binding (black) was only impacted by the presence of N-HelQ (pink) and not BSA (blue) suggesting displacement activity.

5.5.3 N-HelQ displacement of RPA from DNA relies on the PWI-like domain.

Data shown here suggests that the predicted PWI-like domain is central to the ability of N-HelQ to displace RPA bound to ssDNA. The PWI-like domain, as discussed in chapter 3 (figure 3.21), was identified in N-HelQ due to potential structural homologies with the four-helix bundle of a PWI-like fold in the yeast

Ski2 family helicase Brr2 (figure 5.19). The PWI-like fold interacts with DNA and proteins; however, was not required for N-HelQ interactions with DNA (section 4.14). Therefore, due to similarities with Brr2 and the role in RPA displacement (figure 5.16) we predict the PWI-like fold has a role in protein-protein interactions.

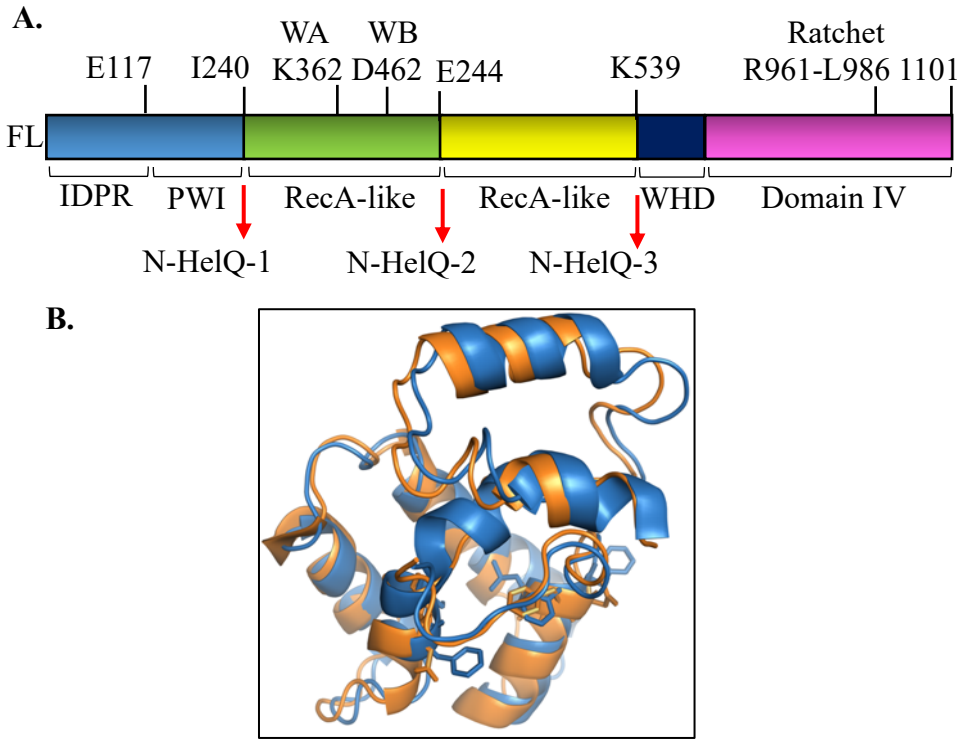


Figure 5.19 Predicted PWI-like fold in N-HelQ proteins. A. Representative schematic of full length HelQ with domains regions: N-terminal ORFan domain (blue), RecA-like domain 1 (green), Rec-A like domain 2 (yellow), wing-helix domain (navy) and domain IV (pink). The N-terminal region used in this work includes a predicted region of disorder (IDPR) and PWI-like domain. N-HelQ protein variants were purified in this work, N-HelQ -1 (I240), N-HelQ-2 (E244) and N-HelQ-3 (K539). **B.** The predicted PWI-like domain was identified through structural homology to the PWI-like fold of the Ski2 helicase Brr2. Shown is Brr2 (orange) superimposed onto N-HelQ (blue).

To further isolate the RPA displacement activity, assays were carried out with the additional N-HelQ proteins described in chapter 3 (section 3.11.2). This was in an attempt to confirm that the PWI-like fold is required for displacement activity to be observed. N-HelQ-N1 was also used to ensure the affinity tags did not impact the binding ability because this variant was made without the large SUMO tag. N-HelQ-N1 (PWI-like only; 36 kDa), which was unable to bind DNA, showed RPA-

displacement activity evident from an increase in free Cy5 fork 2b with increased N-HelQ (figure 5.20 A). N-HelQ-N2 (PWI-like and RecA-like 1; 60 kDa) and N-HelQ-N3 (PWI-like and RecA-like domains 1 and 2; 83 kDa), which were able to bind DNA, showed some displacement activity of RPA from DNA (figure 5.20 B-C). Observations with N2 and N3 proteins were hard to interpret because free DNA was shifted with N-HelQ once RPA had been displaced (figure 5.20 B-C). These results confirmed the displacement activity of N-HelQ and provided further evidence for a role of the PWI-like domain.

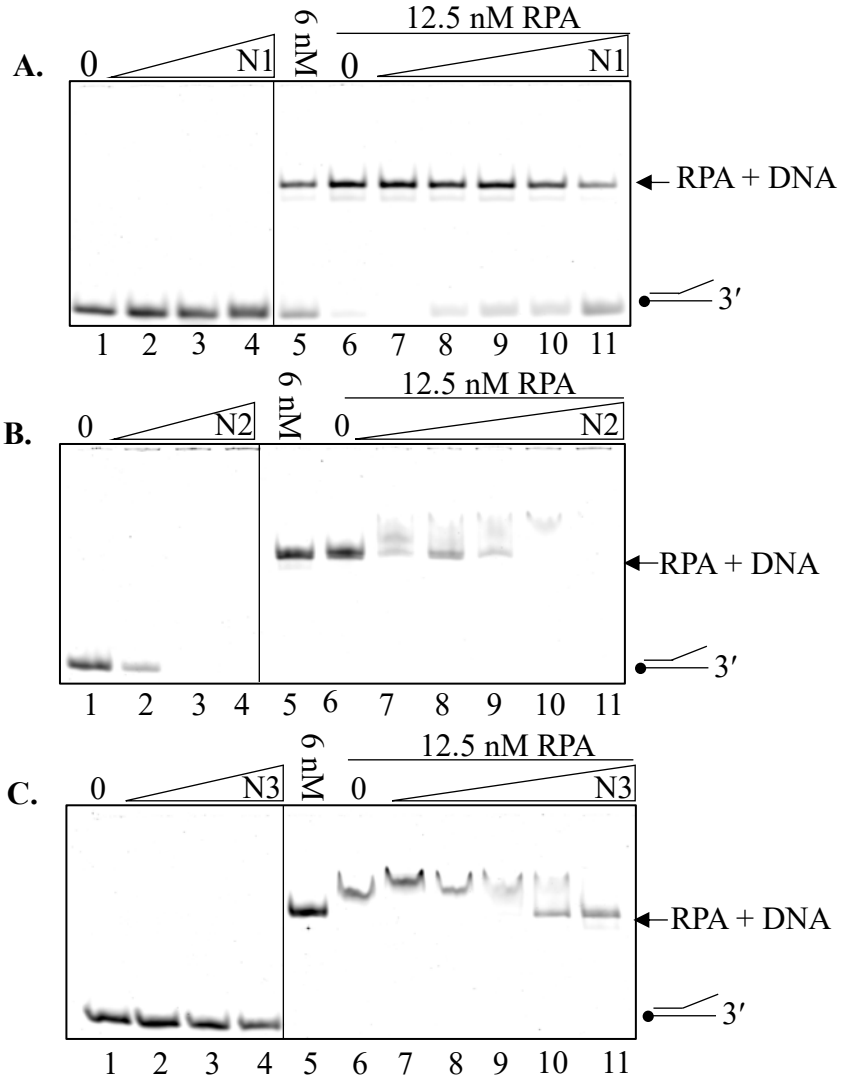


Figure 5.20 Displacement of RPA from DNA was shown with N-HelQ proteins of various lengths. Representative EMSA TBE gels to show addition of StrepII-tagged N-HelQ variants, used at 0, 25, 50, 100, 250 and 500 nM, results in the displacement of RPA, at 6 and 12.5 nM, from fork 2b (25 nM). Assays were carried out with **A.** N-N-HelQ-N1 (PWI-like only), **B.** N-HelQ-N2 (PWI-like and RecA-like 1) and **C.** N-HelQ (PWI-like and RecA-like 1 and 2). Schematics show forked substrate with filled dots indicating 5′ Cy5 label.

5.5.4 C-HelQ does not show displacement activity of RPA from DNA.

C-HelQ was assessed for its ability to displace RPA from forked DNA as described for N-HelQ. C-HelQ was unable to displace RPA from DNA evident from the absence of liberated Cy5 fork 2b being detected with addition of C-HelQ (figure 5.21). The ability of C-HelQ to bind DNA resulted in the fork 2b not bound to RPA, shifting with C-HelQ. However, the intensity of RPA-DNA complexes did not change implying C-HelQ was not displacing RPA and binding to the newly liberated substrate. This suggests that the displacement activity is specific to N-terminal region of HelQ.

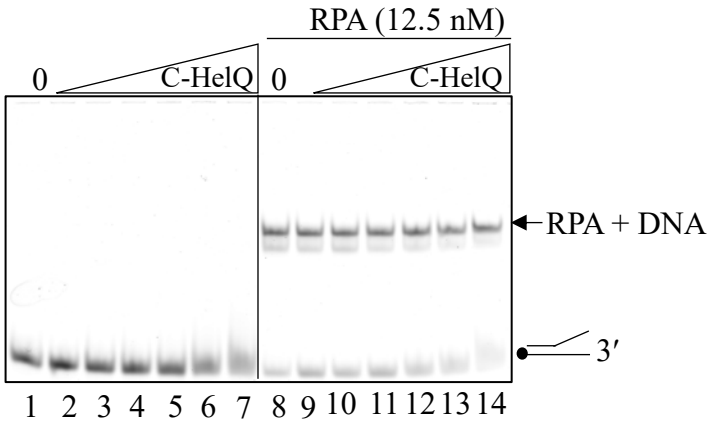


Figure 5.21 Protein displacement assays to show C-HelQ does not displace RPA from DNA. Representative EMSA TBE gels to show C-HelQ, used at 0, 6, 12, 25, 50, 100 and 200 nM, does not displace RPA (12.5 nM) from fork 2b DNA (25 nM). Schematic shows DNA substrate with filled dot indicating 5′ Cy5 label.

5.5.5 N-HelQ, but not FL-HelQ, can displace RPA from DNA substrates.

Additional protein displacement assays with FL-HelQ were carried out to assess if RPA displacement by N-HelQ was observed when part of the entire protein. While the S1 nuclease assays presented earlier imply some displacement activity by FL-HelQ, this observation was weak and therefore required further experimental evidence. This activity also required ATPase active HelQ while N-HelQ does not require ATP activity to displace RPA. This suggests that

independent activities of HelQ are working to see this affect. The formation of HelQ-DNA complexes on liberated DNA resulted in difficulties in isolating RPA displacement activity of FL-HelQ. Therefore, a DNA bubble substrate with internal ssDNA binding sites was used to assess displacement activity by FL-HelQ. EMSAs were carried out to show that RPA, but not FL-HelQ, could load onto the bubble DNA (figure 5.22 A). The bubble substrate represents 'breathing' DNA caused when a helicase unwinds a small section of the duplex DNA. Addition of FL-HelQ to RPA-bubble complexes resulted in HelQ binding to RPA evident from DNA aggregation in the gel wells (figure 5.22 A). This suggests FL-HelQ was not displacing RPA from DNA but forming protein-protein complexes with RPA bound to DNA. FL-HelQ binding to fork 2b was used as a control in these assays. N-HelQ was able to displace RPA from bubble DNA suggesting activity by N-HelQ is not substrate specific (figure 5.22 B). We are unable to conclude if FL-HelQ was able to displace RPA from DNA because of RPA-HelQ complex formation and therefore additional components may be required to see this affect for FL-HelQ. In addition, the lack of displacement observed may be as a result of insufficient concentrations of FL-HelQ being reached equal to those used with N-HelQ (500 nM). We hypothesise that the activities of translocation, unwinding, protein interaction and displacement of different regions of HelQ are catalysed simultaneously which is hard to capture in assays presented here. Therefore, the next question that we aimed to address was how the N-terminal domain of HelQ was able to displace RPA in the absence of DNA or RPA binding capabilities.

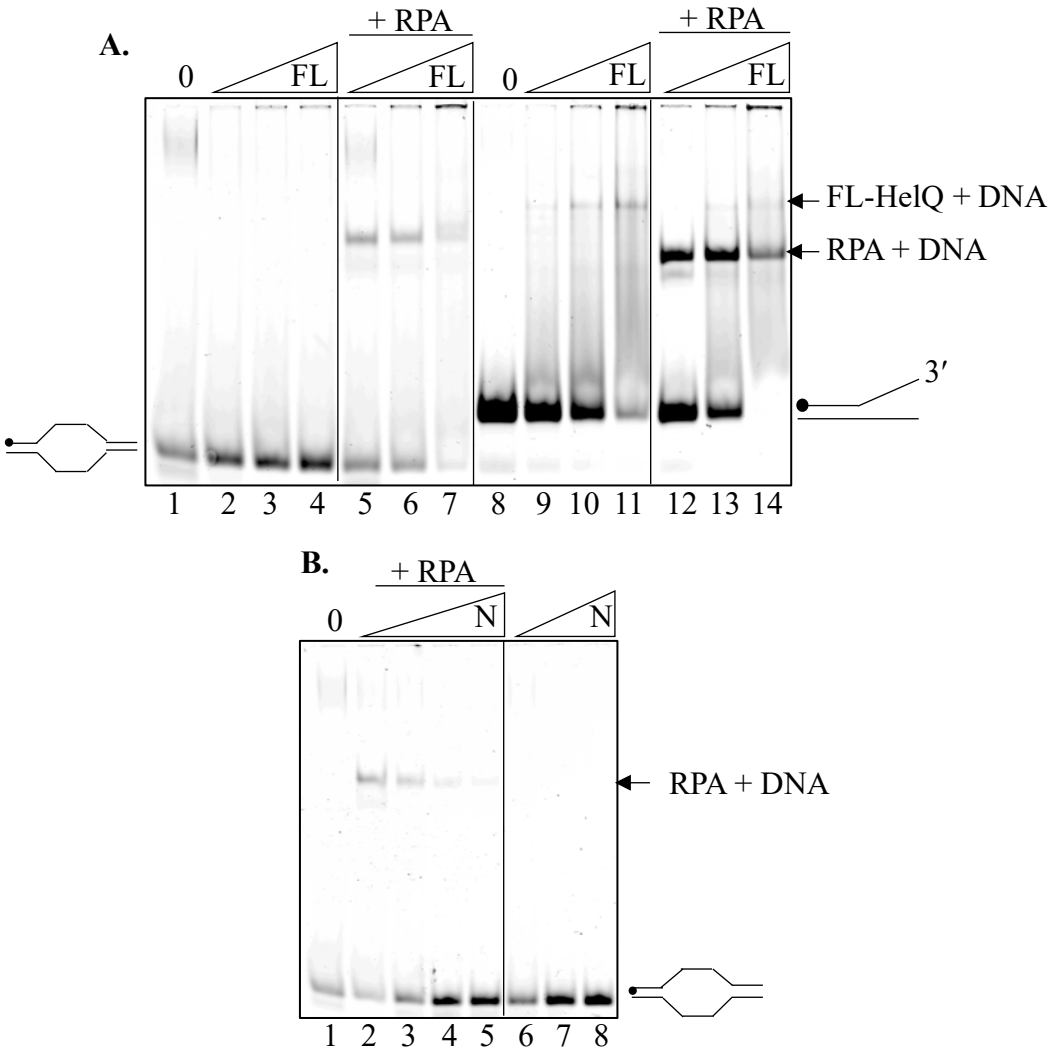


Figure 5.22 N-HelQ, but not FL-HelQ, displaces RPA from a bubble DNA substrate. A. Representative EMSA TBE gels to show HelQ (FL), used at 0, 80 and 160 nM, does not displace RPA (12.5 nM) from bubble DNA substrate or fork 2b (25 nM). HelQ is unable to load onto the bubble DNA. **B.** Also shown is RPA displacement activity by N-HelQ, used at 0, 100, 200 and 500 nM, from bubble DNA (25 nM). Schematics show DNA substrate with filled dots indicating Cy5 end label.

5.6 N-HelQ and RPA do not transiently interact.

Data presented here provides evidence for a role of N-HelQ in RPA displacement *in vitro*. However, N-HelQ and RPA are not shown to interact as shown in figure 5.7. We hypothesised that a physical interaction required interacting interfaces in both C- and N-terminal regions of HelQ. It is unclear how N-HelQ is able to displace RPA from DNA in the absence of N-HelQ interacting with either DNA or RPA. Therefore, we next assessed in more detail the presence

of transient interactions, common in IDPRs, between N-HelQ and RPA.[322] Furthermore, the predicated PWI-like domain in N-HelQ (section 3.12), important in the displacement activity (figure 5.16), has a role in protein-protein interactions in Brr2 further implying N-HelQ may transiently interact with RPA.[329]

5.6.1 Bioinformatic predictions of RPA-HelQ complex formation.

Predicted models of RPA with C-HelQ and N-HelQ were made using the co-threading of protein-protein complex structures tool (COTH online) (section 2.22.4).[307] Ten iterations for predicted interactions between RPA70 with C-HelQ and N-HelQ were produced and one of each is shown here (figure 5.23). RPA70 was used in these models because it is the main interacting subunit of heterotrimeric RPA.

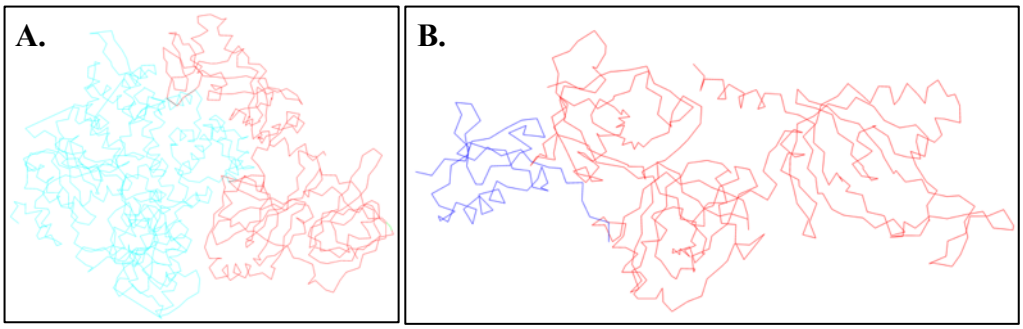


Figure 5.23 COTH modeling of potential complex formation for RPA70 with C-HelQ and N-HelQ. Ten iterations of predicted interactions between RPA70 (red) with, **A.** C-HelQ (cyan) and **B.** N-HelQ (dark blue) were produced. One iteration with a predicted physical contact using the COTH ZhangLab tool for each interacting partner is shown here using PBD files generated in PHYRE2. 40% of the predicted models for C-HelQ resulted in physical contacts and only 10% of the N-HelQ models resulted in physical contacts with RPA70.

Physical interactions between C-HelQ and RPA were predicted in 40% of the models produced implying these domains are unlikely to interact independently to form the stable complexes observed. Furthermore, only 10% of the models produced resulted in physical interactions between N-HelQ and RPA70. However, these predicted models do not account for the protein flexibility of disordered N-

HelQ. Interactions between N-HelQ and RPA may result in disorder-to-order transition within N-HelQ that could not be modelled here. More extensive biochemical methods to detect protein-protein interactions were carried out despite no reliable predicted models between N-HelQ and RPA70 being observed.

5.6.2 Analytical gel filtration (AGF) did not detect N-HelQ RPA complexes.

AGF was used to detect protein complex formation between N-HelQ and RPA with ssDNA in-solution (section 2.20). N-HelQ run in the presence of ssDNA eluted as two protein peaks at 1192 kDa (peak 1) and 279 kDa (peak 2) (figure 5.24 A). SDS PAGE analysis confirmed peak 1 as aggregated protein and peak 2 as higher oligomers of N-HelQ (figure 5.24 B). Heterotrimeric RPA run in the presence of ssDNA eluted at 251 kDa (peak 3) (figure 5.24 A). Due to the limitations of AGF described in chapter 3 (section 3.18.2) RPA peak at 251 kDa was considered as monomeric RPA bound to ssDNA. When pooled, N-HelQ and RPA with ssDNA peak analysis implied that N-HelQ eluted alone (peak 4 and 6), RPA eluted alone (peak 6) and an additional protein peak eluted at 640 kDa (peak 5). SDS PAGE analysis was inconclusive in confirming the components of peak 5. Weak protein smearing was observed at sizes consistent with N-HelQ and RPA70 (figure 5.24 C). These results did not confirm the presence of RPA-N-HelQ complexes, however, may imply that N-HelQ disrupts the heterotrimeric RPA resulting in the additional peak being detected. Alternatively, the additional peak may be due to particle proximity resulting in a change to the local environment leading to altered complex formation of individual proteins and subsequent changes to migration. These results were inconclusive because of the inability to confirm the components of peak 5, the additional peak when the proteins were incubated together. This required additional techniques to be explored to assess for protein interactions.

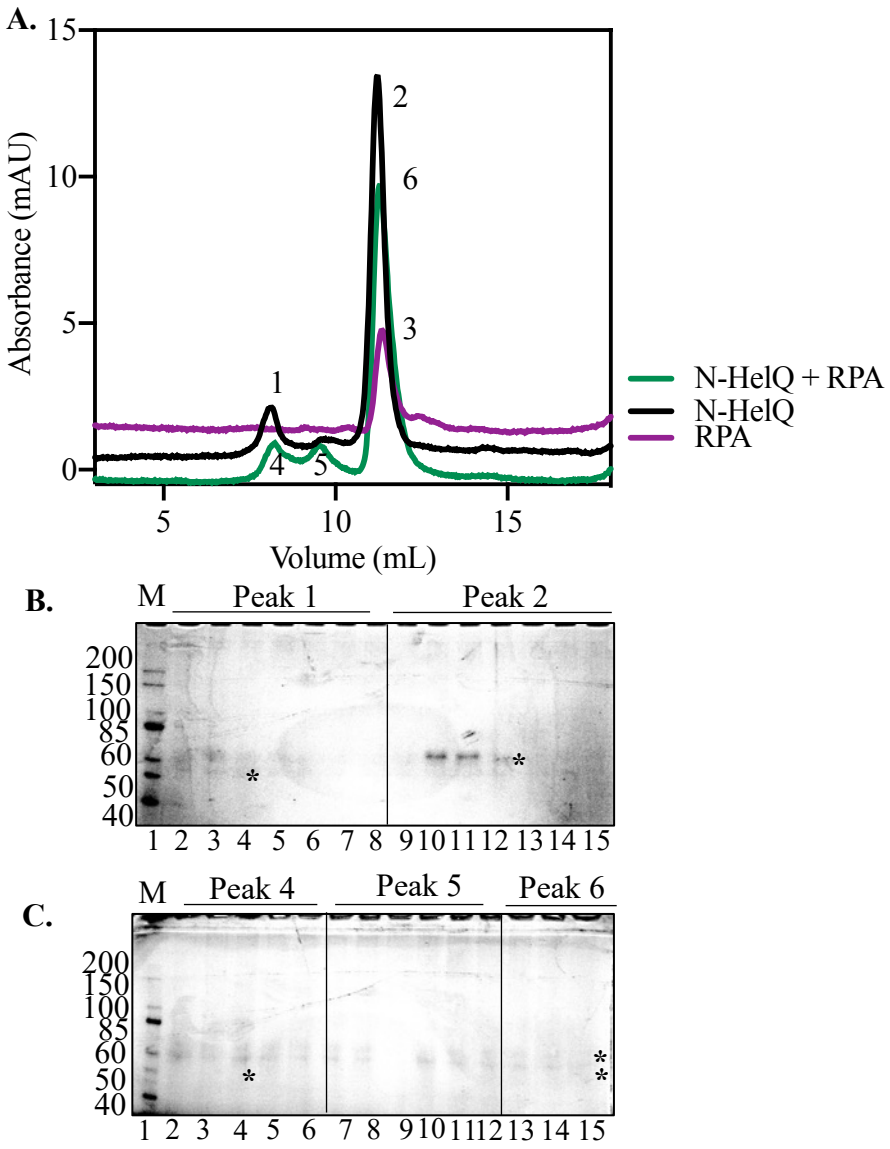


Figure 5.24 AGF identified an additional species in runs with N-HelQ and RPA. A measurement of N-HelQ and RPA populations with ssDNA (25 nM) using size exclusion chromatography Superdex 200 column. **A.** AGF traces for N-HelQ alone (black), RPA alone (purple) and N-HelQ with RPA (green) in the presence of ssDNA. Also shown are acrylamide gels for SDS PAGE analysis to show the presence of **B.** N-HelQ in peaks 1 and 2; and **C.** potential RPA70 and N-HelQ across peaks 4, 5 and 6.

5.6.3 Crosslinking and protein pelleting assays did not isolate interactions between N-HelQ and RPA.

Pools of N-HelQ and RPA were treated with the crosslinking agent bis(sulfosuccinimidyl)-2,2,7,7-suberate-d_0 (BS^3-d_0) to assess transient protein-protein interactions (section 2.18.7). RPA treated with BS^3-d_0 migrated as

heterotrimeric at 116 kDa, monomeric RPA70 and further combinations of RPA32 and RPA14 (figure 5.25 lane 2). N-HelQ treated with BS^3-d_0 migrated predominantly as monomers and dimers (lane 3). N-HelQ and RPA incubated together, in the presence and absence of the BS^3-d_0, resulted in a reduction in heterotrimeric RPA at 116 kDa being detected (lanes 5 and 6). Furthermore, bands consistent with crosslinked N-HelQ shifted from 60 kDa to 75 kDa (lanes 5 and 6) in the presence of crosslinker. This suggests that crosslinking of N-HelQ and RPA pools resulted in the loss of trimeric RPA and potential complexes between N-HelQ and RPA14 or RPA32 forming. The interactions are likely to be transient because the shift was only detected in the presence of crosslinker.

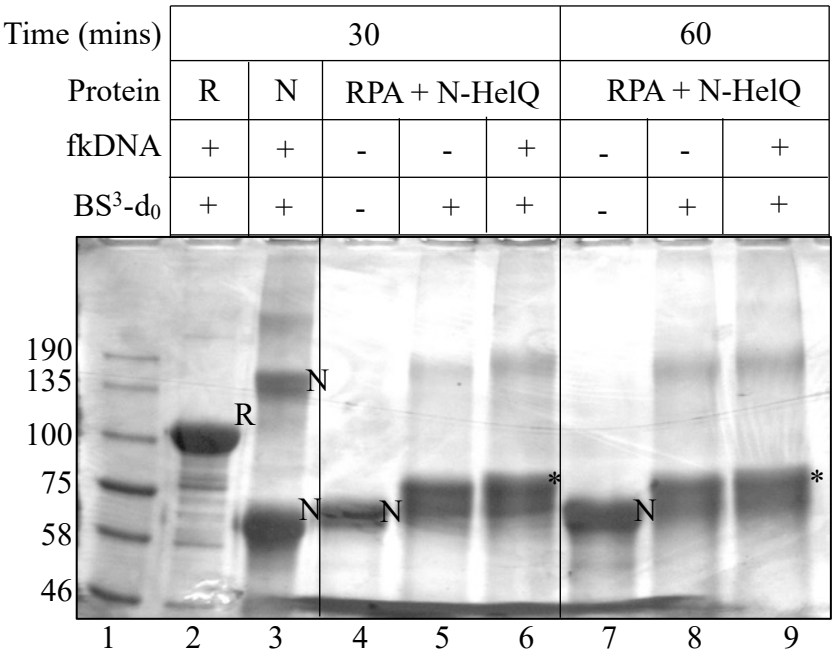


Figure 5.25 Interactions between N-HelQ and RPA were not detected using crosslinking. Representative 10% acrylamide gel to show SDS PAGE analysis of N-HelQ (N) and RPA (R) crosslinked by bis(sulfosuccinimidyl)-2,2,7,7-suberate-d_0 (BS^3-d_0) in the presence and absence of fork 2b DNA. Crosslinker was optimised to 0.5 mM and incubated with 1 μM protein at room temperature for 30 and 60 minutes. RPA crosslinked to form predominantly heterotrimeric at 116 kDa (lane 2) and N-HelQ formed predominantly monomeric (lane 3). No additional bands at a higher molecular weight that would indicate complex formation were detected when N-HelQ and RPA where incubated together (lanes 4-6). However, the heterotrimeric RPA band at 116 kDa was lost in the presence of N-HelQ for 30 minutes (lanes 4-6) and for 60 minutes (lanes 7-9). Slight shifting of the N-HelQ band at 58 kDa was observed in the presence of RPA (lanes 5-6) and crosslinker implying N-HelQ may be crosslinked with RPA14 or RPA32 (*).

The time incubated with the crosslinker and the presence of DNA had no impact on protein species detected. From this we can hypothesise that potentially a transient interaction between N-HelQ and RPA14 or RPA32 causes heterotrimeric RPA to destabilise and the trimer to dissociate.

5.6.4 Microscale thermophoresis (MST) did not show physical interactions between N-HelQ and RPA.

MST measures changes in fluorescence of a target protein when a non-fluorescently labelled substrate is added, detailed in section 2.21.8.[300] N-HelQ was labelled with Cy5 by addition to the hexaHis tag of the protein (see 2.21.8). There was no detectable change observed as Cy5 fluorescently labelled N-HelQ was incubated with RPA-DNA complexes (figure 5.26). Only one incidence of a sigmoidal pattern indicating interactions was detected in a number of assay conditions tested (figure 5.26 A). However, this was only observed at one time point and could not be replicated. No change in fluorescence signal was detected in all other runs suggesting no interactions between N-HelQ and RPA (figure 5.26 B).

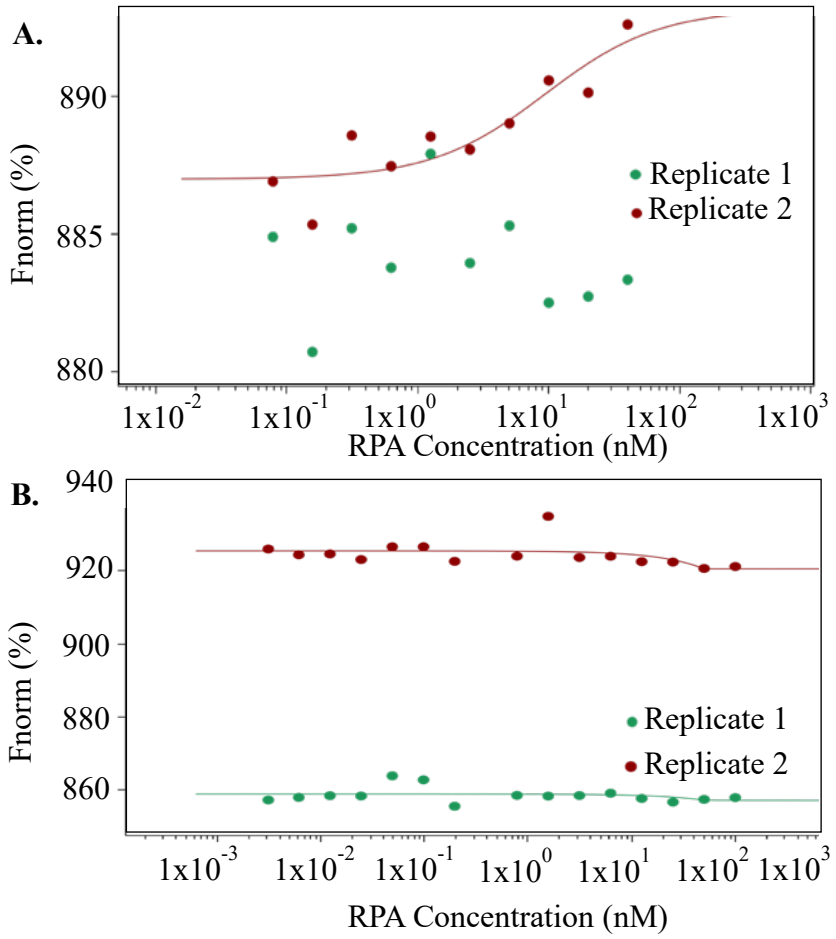


Figure 5.26 Microscale thermophoresis (MST) did not show interactions between N-HelQ and RPA. A measurement of protein-protein interactions using the Nanotemper Monolith NT.115. Cy5 fluorescently labelled N-HelQ was used at 50 nM and unlabeled RPA was used as a serial dilution starting at 40 nM. Experiments were run at both 20% and 40% MST power with a 1% excitation and plotted following a K_d model. Reactions were carried out in duplicate at 25°C and 37°C in the **A.** presence of ssDNA (25 nM) and **B.** absence of ssDNA. No sigmoidal pattern was for observed for duplicate data at a single time point at 37°C (green and red).

5.6.5 Isothermal titration calorimetry (ITC) did not show interactions between N-HelQ and RPA.

ITC, which detects minute changes in temperature, was used to measure weak interactions between N-HelQ and RPA (section 2.21.5).[299] Addition of N-HelQ to RPA resulted in no detectable heat signature evident from no sigmoidal pattern with addition of N-HelQ being observed (figure 5.27). This data suggests that N-HelQ

and RPA do not interact. However, these assays were not carried out in the presence of DNA which may be required to initiate N-HelQ binding to RPA.

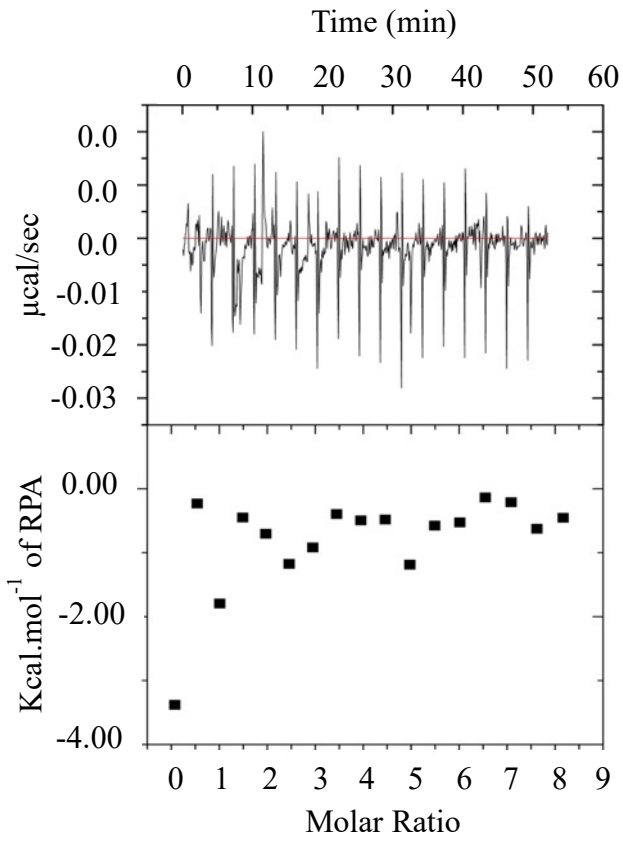


Figure 5.27 Isothermal titration calorimetry (ITC) did not show interactions between N-HelQ and RPA. A measurement of changes in heat when N-HelQ (200 µM) is pumped into an RPA containing calorimetric cell (5 µM). The upper panel shows temperature changes over time. Spicks in temperature observed coincide with the effect of unspecific heat associated with N-HelQ injection points. No large heat changes were observed implicating no protein interactions. The lower panel shows the calculated molar enthalpy change (Kcal.mol^{-1}) of RPA with addition of N-HelQ over time. The absence of a sigmoidal pattern indicates no relative temperature change suggesting no interactions. Molar ratio refers to the change in ratio of RPA to N-HelQ as N-HelQ is titrated into the reaction. A molar ratio <0.5 reflects unspecific background injection heat associated with the absence of N-HelQ.

5.7 Analysis of the ability of N-HelQ to dissociate heterotrimeric RPA.

5.7.1 Native PAGE does not show trimeric RPA dissociation by N-HelQ.

In the absence of detecting transient interactions between N-HelQ and heterotrimeric RPA, we hypothesised that N-HelQ may only interact with one

subunit of RPA. We predict that N-HelQ destabilises the heterotrimeric RPA and therefore any interaction that does occur is likely to be highly transient and with a single RPA subunit. Displacement activity observed by N-HelQ (figure 5.20) shows that RPA does not reload onto the DNA immediately after being displaced by N-HelQ. This suggests that RPA undergoes a structural or conformational change when displaced preventing reloading. This may be caused by trimeric dissociation which would slow down the ability of RPA to reanneal the DNA until the heterotrimer was reassembled. Therefore, we next carried out Native PAGE to assess if N-HelQ caused trimeric RPA dissociation (section 2.16.6). When run independently, RPA and N-HelQ migrated as multiple higher oligomers (figure 5.28 lanes 2 and 3). When run in the presence of ssDNA, RPA also migrated as an additional complex around 242 kDa (lane 4). Similar complexes were observed when RPA and N-HelQ were pooled (lanes 5 and 6). This suggests that the addition of N-HelQ did not result in RPA dissociation evident from the lack of bands at 70, 32 or 14 kDa.

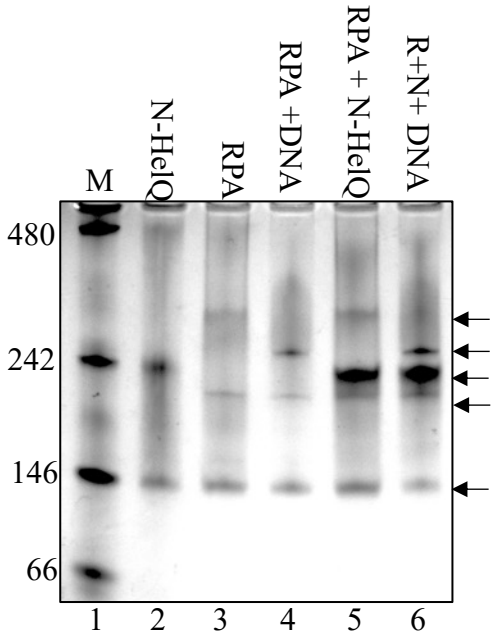


Figure 5.28 Native PAGE did not show heterotrimeric RPA dissociation in the presence of N-HelQ. Acrylamide Native PAGE gel to show heterotrimeric RPA (R; 116 kDa) runs at 146 kDa, 200 kDa and 300 kDa (lane 3). In the presence of ssDNA (25 nM) RPA also runs at 250 kDa (lane 4). N-HelQ runs at 146 kDa and 230 kDa (lane 3). The complexes that RPA form remain intact in the presence of N-HelQ (N; lane 5) and N-HelQ and ssDNA (lane 6). This suggests N-HelQ does not cause dissociation of the heterotrimer.

While RPA was not dissociated by N-HelQ in Native PAGE, a mechanism of RPA distortion by N-HelQ to displace RPA from DNA cannot be ruled out. We aimed to next assess if N-HelQ only interacts with an individual subunit of heterotrimeric RPA. We initially tested for interactions between N-HelQ and the individual RPA14 subunit to determine if heterotrimeric dissociation was caused by N-HelQ interacting with RPA14. RPA14 has been previously shown to be important in the stability of the RPA heterotrimer. We therefore wanted to assess if N-HelQ de-stabilised the heterotrimer RPA by targeting the RPA14 subunit.[77] Furthermore, RPA14 adopts an extended structure when assembled into the heterotrimer creating a platform for proteins to readily interact, this may allow N-HelQ to readily associate with it. Therefore, we assessed RPA14 with N-HelQ first.

5.7.2 Purification of RPA14 from *E. coli*.

We next assessed if N-HelQ only interacted with the smallest subunit of RPA, RPA14. DNA encoding RPA14 was cloned from p11d-tRPA into pBadHisA with an N-terminal hexaHis-tag and purified from *E. coli* as described in section 2.14.5 (figure 5.29). RPA14 was problematic to purify in isolation. This was likely because RPA14 is unstable unless complexed in heterotrimeric RPA or with RPA32.[77]

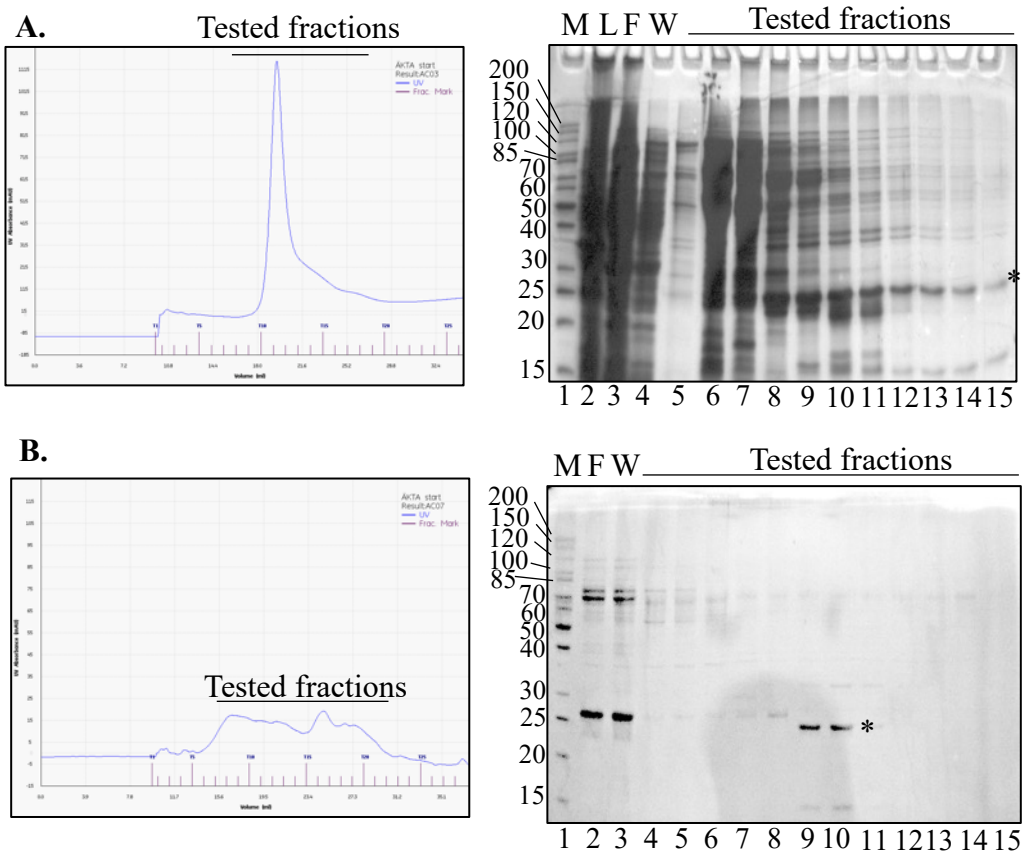


Figure 5.29 RPA14 protein purification from *E. coli*. A. DNA encoding RPA14 was cloned into pBadHisA for expression from *E. coli* with an N-terminal HexaHis-tag. Biomass containing RPA14 (23.5 kDa) was sonicated, clarified and resolved by binding to a $NiCl_2$ charged HiTrap Ni-NTA column. The panels show UV trace analysis and Coomassie stained acrylamide gels of eluted RPA14 are shown. **B.** RPA14 was further resolved using heparin affinity chromatography. Legend = marker (M), clarified lysate (L), flow-through (F), wash-through (W), eluted protein (*).

5.7.3 Interactions were not detected between RPA14 and N-HelQ.

Physical interactions between RPA14 and N-HelQ were measured as described previously with heterotrimeric RPA (figure 5.30). HexaHis-tag protein pull-down assays could not be carried out because both proteins had a hexaHis-tag. Furthermore, protein interactions were not determined used super-shifting in EMSAs (figure 5.30 A). RPA14 was unable to interact with fork DNA independently evident from no band shifting with Cy5 labelled DNA (lane 2-5). In the absence of N-HelQ-DNA interactions, no complexes were observed when RPA14 and N-HelQ were pooled with DNA (lane 8-9). This implies RPA14 and

N-HelQ do not interact on DNA when considered as separated from the full protein. Similarly, phosphomimetic N-HelQ did not super-shift with RPA14 (lanes 12-13). The smearing observed was consistent with the weak DNA binding of phosphomimetic N-HelQ (lanes 10-11).

Native PAGE was used to assess protein-protein interactions between N-HelQ and RPA14 by observing the formation of higher molecular weight complexes (figure 5.30 B). RPA14 migrated at 146 kDa and N-HelQ migrated around the 200 kDa marker when run alone (lanes 2-4). Similar migration patterns were observed when RPA14 and N-HelQ were incubated together (lanes 5-6). Likewise, the presence of DNA did not impact protein migration patterns (lanes 7-8). These results suggest no additional complex was formed between N-HelQ and RPA14 and therefore the two proteins do not physically interact.

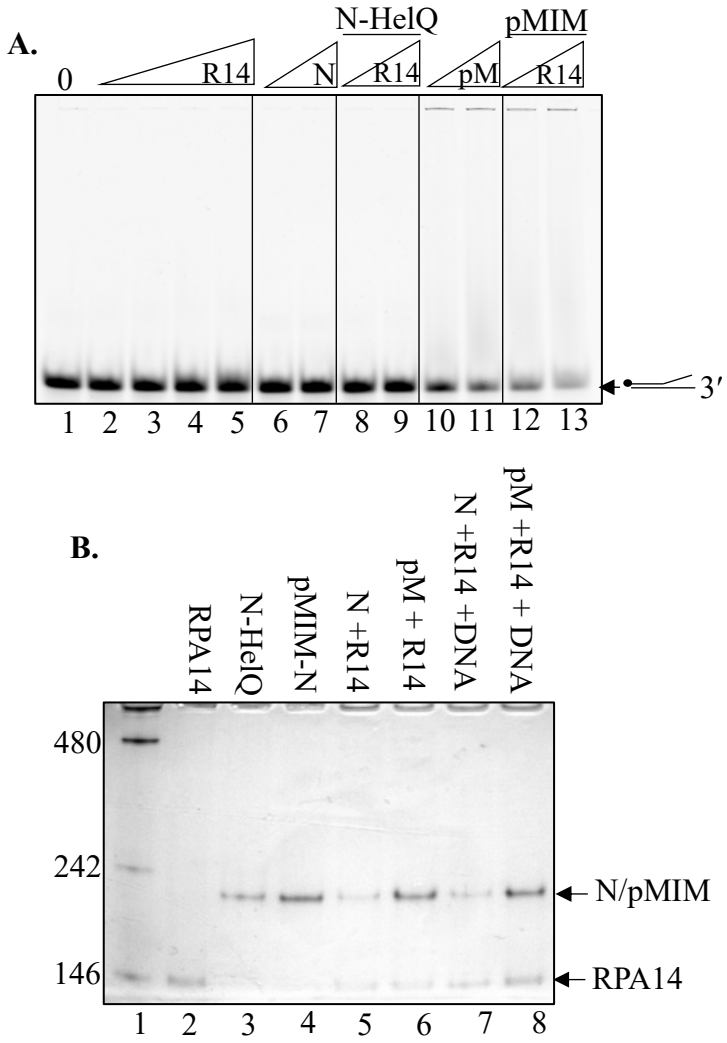


Figure 5.30 Native PAGE showed no physical interactions between RPA14 and N-HelQ. A. TBE EMSA gel to show RPA14 (R) was unable to bind to fork 2b DNA (25 nM) alone or in the presence of N-HelQ (N) and phosphomimetic N-HelQ$^{S158/178D}$ (pM) (500 nM). RPA14 was used at 0, 15, 30, 60 and 120 nM. **B.** Also shown are acrylamide gels for Native PAGE analysis of RPA14 in the presence and absence of N-HelQ and phosphomimetic N-HelQ$^{S158/178D}$. N-HelQ had no impact on the migration patterns of RPA14. Addition of ssDNA (25 nM) also had no impact on migration patterns.

These results give no evidence for an interaction between N-HelQ and RPA14 when RPA14 is considered as a separate entity to the heterotrimeric protein. However, analysis of N-HelQ with RPA14 as an independent subunit from the heterotrimer may impact the conditions that are required for the interaction to occur. As described in chapter 1, the assembly of heterotrimeric RPA together and onto DNA is essential for its activity. Therefore, N-HelQ may interact with RPA14 while in a specific conformation and upon interacting with RPA14 it results in

conformational change disrupting the heterotrimeric RPA causing its dissociation from DNA. Further work is required to assess RPA14 interactions with N-HelQ when part of heterotrimeric RPA complex.

5.8 Assessment of potential electrostatic interactions within a conserved motif.

The presence of electrostatic interactions between N-HelQ and RPA was measured to further understand the interaction observed (figure 5.31). The archaeal homologue Hel308 has essential electrostatic interaction domains that are found conserved in HelQ. Electrostatic interactions form when there are differences in the electrostatic charge of interacting proteins between positively charged nuclei and negatively charged electrons.[378] The region predicted to be involved in electrostatic interactions in Hel308 is located on the solvent exposed flexible linker in the winged helix domain (figure 5.31). The electrostatic motif is located in close proximity to key residues essential in the activity of the WHD including conserved tyrosine and phenylalanine residues within the solvent exposed α helix 20.[379]

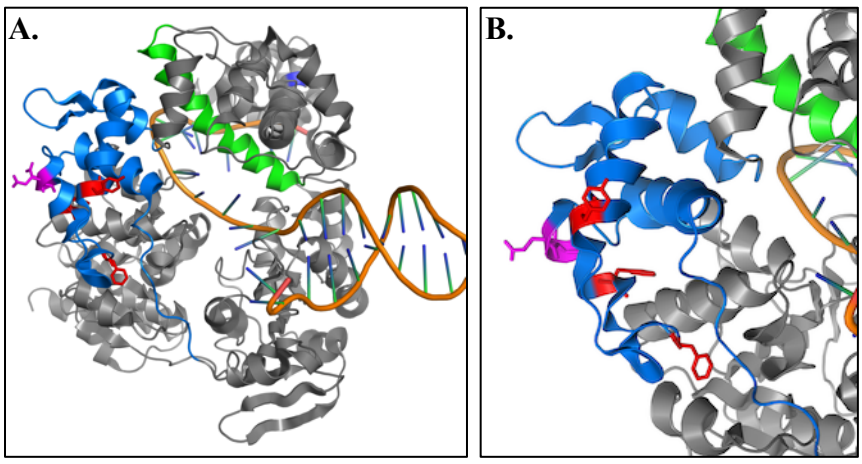


Figure 5.31 PHYRE2 models to show conserved residues in interacting interface of Hel308. PHYRE2 model of the archaeal *Afu* solved crystal structure of Hel308. **A.** Shown are the conserved tyrosine and phenylalanine residues (red) essential in the activity of the winged helix domain (blue) in DNA binding. Also shown is the DNA binding ratchet in domain IV (green). The conserved electrostatic residues located on the solvent exposed flexible linker of the WHD are shown in pink. These residues are conserved in HelQ. **B.** Also shown is a close up of the electrostatic interface in close proximity to the essential residues in the WHD.

This region packs against the RecA-like domain 1 and is proposed to be a recognition helix for DNA binding in branched substrates.[379] Furthermore, the WHD region containing the electrostatic interface is located closely to the DNA recognition ratchet in domain IV (figure 5.31). Therefore, this region is implicated as an interaction interface that may expand to protein-protein interactions. This motif was therefore considering important in the HelQ homologue.

Electrostatic potential maps were used to illustrate the charge distribution of FL-HelQ and predict the conserved regions involved in electrostatic interactions (figure 5.32 A). The conserved string of amino acids between HelQ and Hel308 potentially involved in electrostatic interactions are also shown (figure 5.32 B). A conserved positively charged double aspartate at position 771 in HelQ central to the conserved region (double glutamate in Hel308) was considered important for function (figure 5.32 C). Therefore, a HelQ peptide (49 residues), known as $HelQ^{PEP}$, with the central conserved region was synthesised (Severn Biotech Ltd.) to test for electrostatic interactions with RPA (figure 5.32 D).

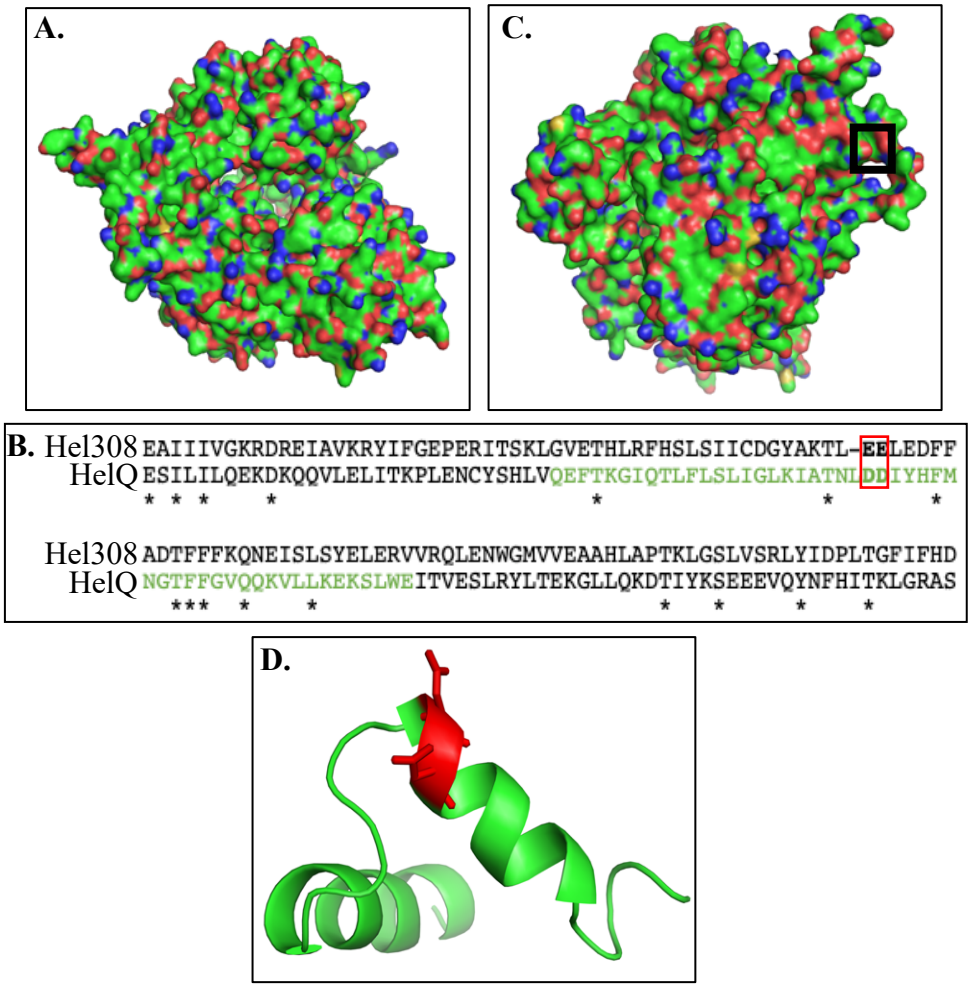


Figure 5.32 Surface mapping of FL-HelQ to show the electrostatic potential. A. Pymol generated electrostatic surface map of FL-HelQ with positively charged (red), negatively charged (blue) and neutral residues (green) shown. Electrostatic protein-protein interactions form from opposite charge attractions. **B.** Conserved region of electrostatic interacting domain in Hel308 found in HelQ containing central double positive aspartate (red box) using CLUSTAL sequence alignment. **C.** Pymol identifying the location of the conserved region of electrostatic interactions (black box). **D.** PHYRE2 HelQ predicated model of the synthesized peptide with central aspartate, highlighted in green in part B.

Similar super-shift EMSAs and pull-down assays were used to assess for interactions between RPA and HelQ^{PEP}. However, the small size of the peptide (5 kDa) meant that any shifting with RPA-DNA was impossible to detect even with extended gel running times (figure 5.33 A). Saturated ammonium sulphate pull-down assays were used to detect physical protein interactions (section 2.18.5). The concentration of ammonium sulphate required to precipitate HelQ^{PEP} was 10-40%

salt and for RPA was 40-60% (figure 5.33 B). Ammonium sulphate was used at 10% and 20% targeting precipitation of HelQ^{PEP} when the assay was carried out with pools of HelQ^{PEP} and RPA (figure 5.34 C). However, neither HelQ^{PEP} nor RPA precipitated when pooled (figure 5.34 C, lane 6). This may have been as a result of RPA and HelQ^{PEP} interacting resulting in precipitation patterns to vary. The reverse assay targeting precipitation of RPA was not possible because HelQ^{PEP} also precipitated at these salt concentrations (figure 5.33 B).

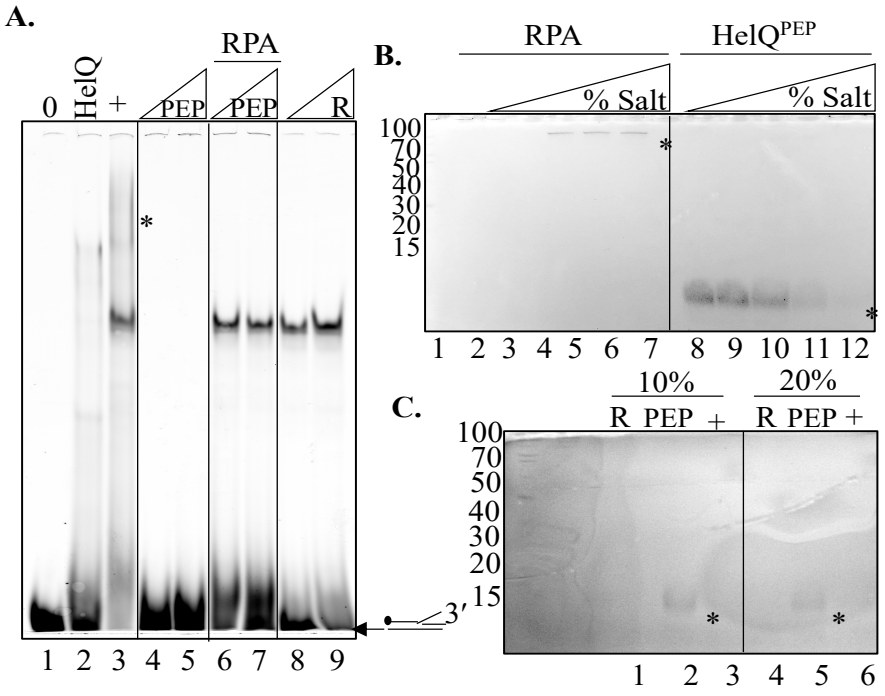


Figure 5.33 EMSAs and pull-down assays are inconclusive to determine electrostatic interactions between HelQ^{PEP} and RPA. A. TBE gel EMSAs to show RPA (R; 120 nM) bound to fork 3 (25 nM) in the presence of the conserved electrostatic peptide HelQ^{PEP} (PEP) used at 0, 80 and 160 nM. FL-HelQ (+) was used as a positive control to show super-shifting with RPA-DNA. **B.** Also shown is acrylamide gels of SDS PAGE analysis of saturated ammonium sulphate precipitation assays to detect interactions of untagged proteins. Ammonium sulphate precipitation of HelQ^{PEP} and RPA was tested individually at 10%, 20%, 40%, 50% and 60%. **C.** Ammonium sulphate used at 10% and 20% to target precipitation of HelQ^{PEP} resulted in no protein precipitation from pools of HelQ^{PEP} and RPA (+).

These results were inconclusive to determine if electrostatic interactions are central to HelQ-RPA interactions. Further work would be required to assess the conserved region in the context of FL-HelQ with RPA. Electrostatic interactions

may be important in HelQ-RPA interactions which would provide further evidence that both regions of FL-HelQ are required to form stable and detectable complexes with RPA.

5.9 Summary and conclusions.

5.9.1 A physical interaction between RPA-HelQ was detected *in vitro*.

FL-HelQ and RPA were shown to physically interact *in vitro* in the presence and absence of DNA. Protein-pull down assays showed StrepII-tagged HelQ was eluted from the resin with addition of heterotrimeric RPA. This implicated that RPA impacts the association of FL-HelQ with the resin, changing its conformation and preventing its continual annealing to the beads. Furthermore, EMSAs showed super-shifting of HelQ and RPA on a 3′ ssDNA flap fork, dsDNA and a modified abasic 3 fork. Control assays using *E. coli* SSB and a DNA binding mutant Aro1-RPA confirmed the specificity of the interaction. Assays with the Aro1-RPA binding mutant also suggested that the super-shifting detected was due to RPA-DNA complexes interacting with HelQ and not HelQ-DNA complexes interacting with RPA. This implies that RPA binds first and that recruitment of HelQ occurs after RPA is loaded onto DNA. The physical interaction detected here supports the co-localisation of RPA and HelQ observed *in vivo* previously.[62,380] However, the physical interaction was not isolated to either the C-terminal or N-terminal regions of HelQ when the fragments were taken in isolation.

5.9.2 RPA is displaced from DNA by N-HelQ.

Work presented here also showed that N-HelQ was able to displace RPA bound to ssDNA. EMSA displacement assays and fluorescence anisotropy showed that N-HelQ reduced RPA-DNA complexes and was isolated to the predicted PWI-like domain of N-HelQ using mutagenesis of conserved residues to the Brr2 helicase.

The PWI-like domain of the Brr2 helicase is involved in protein-protein interactions. Therefore, we predict that due to structural homology the displacement detected may be as a result of transient interactions between RPA and the PWI-like domain. S1 nuclease protection assays showed this effect could be seen by the FL-HelQ to a lesser extent which suggests that the role of N-HelQ was not as a result of N-HelQ being isolated from FL-HelQ.

Phosphorylation of the selected residues in the PWI-like domain is unlikely to be an activating PTM for protein-protein interactions. Unlike in chapter 4 when looking at interactions with DNA, no changes were observed when incubated with RPA. Phosphomimetic N-HelQ$^{S158/178D}$ did not impact displacement of RPA. This does not rule out a role of PTMs or phosphorylation in HelQ activity.

One hypothesis for the observed displacement is if the RPA heterotrimer becomes distorted upon N-HelQ binding. The protein may become distorted by losing its structure required for function, this may be as a result of dissociating the heterotrimer or shielding part of the protein to prevent activity. This may prevent downstream interactions of RPA with other proteins if RPA can no longer be recognised. Distortion of heterotrimeric RPA may impact RPA-DNA binding resulting in dissociation from ssDNA allowing for FL-HelQ, or other proteins, to bind. The strong binding affinity of RPA for DNA suggests that RPA would be able to readily re-anneal to DNA once displaced. Therefore, RPA structure must be impacted due to the absence of rapid re-annealing being reported in assays. The lack of inhibition by RPA-DNA to HelQ activity also suggests that HelQ is physically distorting RPA and preventing its re-annealing. We suggest that N-HelQ both displaces RPA, by an unknown mechanism, and causes RPA to remain unbound. This leads us to our hypothesised model of action.

5.9.3 Hypothesised model for RPA and HelQ interactions on DNA.

Interactions observed between HelQ and RPA likely rely on physical contacts with regions in both C-HelQ and N-HelQ. RPA displacement activity by FL-HelQ was hard to differentiate because of FL-HelQ-RPA complexes potentially remaining intact after displacement. HelQ may displace RPA to create space to bind DNA as well as prevent RPA activation of downstream proteins. HelQ may remain bound to RPA, or one subunit of RPA, to prevent its re-assembly to activate downstream proteins in alternative repair pathways.

In our model proposed here we have considered the activity of individual regions of HelQ in the context of FL-HelQ. N-HelQ is implicated in signalling, recruitment and stability of FL-HelQ while C-HelQ is essential to helicase activity. RPA recruits FL-HelQ to exposed DNA, N-HelQ, through the PWI-like domain, causes the dissociation of trimeric RPA from DNA and allowing for HelQ to bind. RPA maintains in close proximity to HelQ for inhibition of downstream signalling. HelQ translocates and unwinds the duplex.

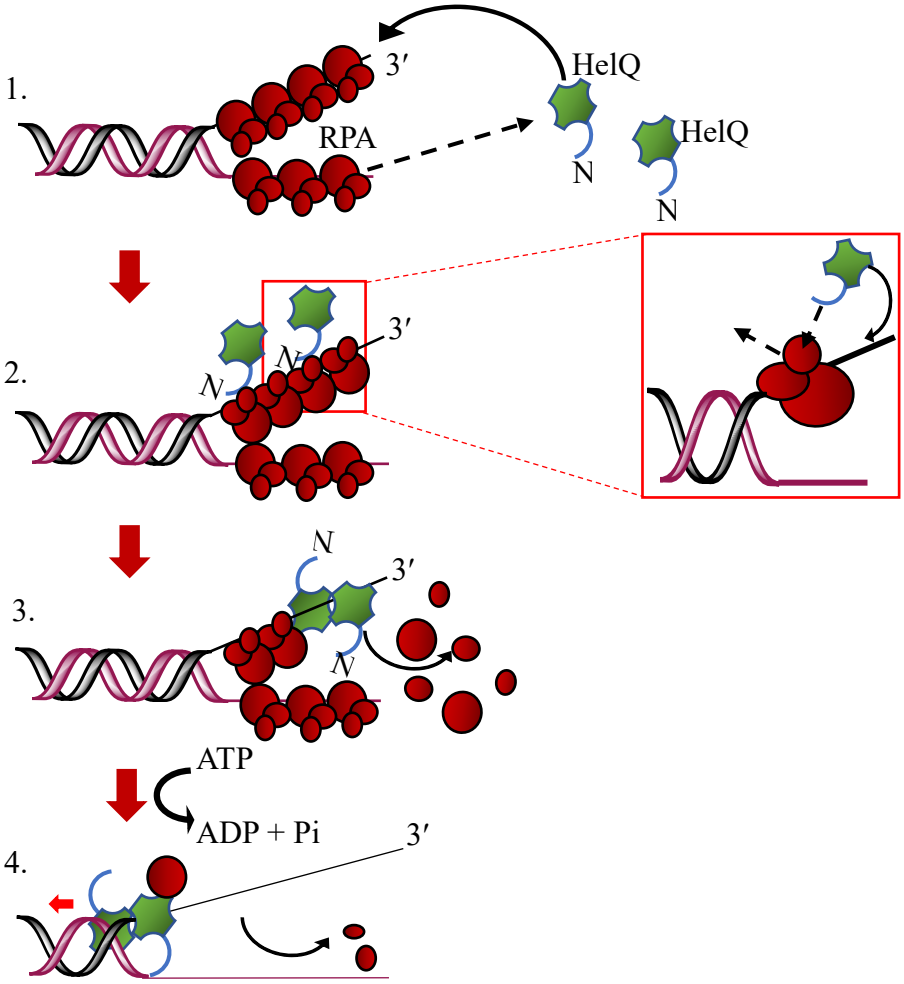


Figure 5.34 Hypothesised model of HelQ and RPA interactions on DNA. We hypothesise that FL-HelQ is recruited to exposed ssDNA at stalled fork structures by RPA-ssDNA complex formation. (1). When HelQ gets close to the RNA-bound ssDNA, it is unable to interact with the ssDNA due to the presence of RPA. N-HelQ transiently interacts with the trimeric RPA, potentially via the RPA14 subunit. N-HelQ interactions cause dissociation of the RPA heterotrimer. This distorts the heterotrimer causing it to fall off the DNA. Dissociation of the trimer from DNA results in interactions between N-HelQ and RPA to also dissociate (2). As RPA dissociates, it allows for HelQ, via the core helicase domain, to bind the newly exposed ssDNA (3). ATPase activity of HelQ results in translocation along the DNA. Subsequent interactions between HelQ and RPA displace the RPA roadblock as HelQ dimers move along DNA (4).

5.9.4 The impacts of RPA on HelQ activity.

Further to recruitment, biochemical analysis shown here did not imply any additional activation or inhibition by RPA on HelQ activity. RPA did not stimulate HelQ to unwind DNA substrates it could not unwind in the absence of RPA.

Furthermore, RPA did not improve HelQ processivity or improve the rate of HelQ unwinding. Interestingly, the presence of RPA with these substrates did also not inhibit HelQ activity. This suggests that RPA does not act as a roadblock to HelQ and provides further evidence that HelQ is able to displace RPA from the DNA in order to gain access.

RPA did appear to impact HelQ displacement activity to unwind through a BamHI^{EIIIA} protein barrier on fork DNA (figure 5.11). However, this result was not significant when considered alone due to a lack of physiological relevance because no further tested roadblocks were displaced by HelQ in the presence of RPA. Observations with the BamHI^{EIIIA} barrier were likely as a result of conformational change upon RPA binding reducing DNA binding affinity of BamHI^{EIIIA} allowing for displacement by HelQ. While helicases aid in fork clearance with and without RPA to allow access for repair proteins, we cannot conclude if this is a role for HelQ in the absence of a more specific roadblock.[177,217] We cannot rule out a role of HelQ in fork clearance, HelQ may be involved in the displacement of a specific protein from DNA. The Rad51 recombinase was been previously isolated in protein pull-down assays with HelQ potentially implicating HelQ in the assembly or disassembly of Rad51 NPFs.[236,237]

5.9.5 Isolating the region of HelQ that interacts with RPA.

The region of HelQ responsible for interacting with RPA was not isolated in work carried out here. In protein pull-down assays neither the N-HelQ nor C-HelQ fragment stably interacted with RPA. This may have been as a result of protein instability of C-HelQ (chapter 3), unoptimized binding conditions or that elements of both N-HelQ and C-HelQ are required for an interaction to form (figure 5.10). Furthermore, bioinformatic models did not predict significant physical interactions between RPA70 and N-HelQ and C-HelQ independently (figure 5.23). However, a

region of C-HelQ may still be responsible for interactions with RPA but are unable to be detected in these assays due to the instability of C-HelQ. Evidence presented in this work suggests that C-HelQ is highly unstable and may rely on stabilisation by the presence of N-HelQ. Therefore, in further experiments to assess for RPA interactions, C-HelQ activity could be analysed in the context of FL-HelQ by mutating the activity of N-HelQ and therefore stabilising the protein. The PWI-like mutations in N-HelQ in the context of FL-HelQ would knock out N-HelQ activity to focus on C-HelQ.

The predicted region of electrostatic potential located within C-HelQ may be involved in RPA binding. Experiments carried out here of the pocket of positively charged residues found conserved in Hel308 failed to conclude the presence of electrostatic interactions with RPA. However, if elements of C-HelQ and N-HelQ are required to form stable complexes with RPA, we would not expect to have observed interactions between $HelQ^{PEP}$ and RPA in the absence of N-HelQ (figure 5.34). The conservation of this motif in Hel308, located within essential DNA binding residues, implicates a role as a protein interaction interface. However, analysis of the 5 kDa peptide was limited due to the small size restricting the ability to detect protein shifts in EMSAs and pull-down assays. This region may therefore be essential for RPA interactions.

The displacement activity observed by N-HelQ, reliant on the PWI-like domain, suggested physical interactions between N-HelQ and RPA. Furthermore, regions of protein disorder are hypothesised to require protein binding partners to stimulate activity.[324] These interactions can be transient because of protein flexibility and therefore are hard to detect. However, extensive analysis did not identify any physical interactions between N-HelQ and RPA. Biophysical techniques using ITC, AGF and MST, and biochemical techniques, using protein crosslinking, pull-downs and super-shifting EMSAs, failed to establish a stable or

otherwise physical interaction. Debatably, weak protein interactions were detected through crosslinking, AGF and MST. Crosslinking resulted in the loss of the trimeric RPA at 116 kDa in the presence of N-HelQ which may provide evidence for protein dissociation. However, higher oligomeric species that would suggest complex formation were not evident. An additional population was detected in AGF; however, the presence of both N-HelQ and RPA was inconclusive. The additional species did imply that pooling the proteins resulted in changes to the local environment impacting the elution of the proteins. The detection of thermal changes in MST at two time points may implicate transient interactions. Alternatively, these results may be as a result of non-specific protein interactions due to particle proximity.

We hypothesise, based on results presented here, that RPA and N-HelQ do not physically interact independently. C-HelQ is more likely to be responsible for interactions with RPA but this was also not isolated and requires further work. C-HelQ requires stabilisation by N-HelQ to assembly with RPA. During the writing of this work, NMR of N-HelQ and RPA is underway to further confirm the absence of physical interactions or as an extensive method to assess transient interactions.

The inability of the displaced RPA to re-anneal to the DNA suggests N-HelQ may dissociate the heterotrimer. This may be caused by interactions between N-HelQ and one subunit of RPA resulting in conformational change within RPA and its subsequent dissociation from DNA. Native PAGE did not detect RPA dissociation; however, this effect may require specific conditions that were not met here. Furthermore, protein interactions were not detected between N-HelQ and RPA14 but N-HelQ-RPA14 interactions may rely on the conformation adopted by RPA14 when assembled into the heterotrimer. N-HelQ is likely to target RPA when in a specific conformation bound to DNA in order to deactivate RPA at very specific time points to prevent downstream effects. Therefore, it would be

unsurprising if N-HelQ only interacted with heterotrimeric RPA in a very specific conformation. Trimeric dissociation cannot be ruled out as the mechanism for N-HelQ displacement activity and may only be detected when N-HelQ and RPA14 are assembled on DNA as full protein entities.

5.10 Outlook and future work.

Further biochemical analysis is needed to expand on data presented in this chapter and confirm the hypothesised model. Further assessment of N-HelQ interactions with RPA70, RPA32 and RPA14 independently to identify the specific domain N-HelQ impacts to cause protein dissociation would be beneficial. Mutagenesis of heterotrimeric RPA could be used to isolate the interacting region with FL-HelQ and determine if there are any specific interactions to N-HelQ. Furthermore, assessment of C-HelQ in the presence of mutated N-HelQ could be assessed for RPA interactions. N-HelQ displacement via a mechanism of trimer dissociation suggests N-HelQ interactions with at least one RPA module are present, even if transient. Repeated analysis using ITC and NMR of individual RPA subunits with N-HelQ and C-HelQ could confirm the presence of transient interactions under different conditions. Furthermore, RPA mutagenesis impacting the different DNA binding modes of RPA to assess the impact of N-HelQ displacement activity and determine if these conformations are important for N-HelQ activated displacement to take place are required.

Determining the mechanism of action of how RPA impacts HelQ and the activity of HelQ on RPA can help establish a role of HelQ in DNA repair and regulating HR. Biochemical understanding of how proteins work in systems such as DNA repair helps progression in research of disease models and therapeutics. We need to understand how systems work normally before we can assess how they faulter. The RPA displacement activity of HelQ may implicate a negative role of

HelQ in preventing downstream HR or in aiding the loading of the Rad51 paralogues and Rad51 stimulating pathways of HR by removing the RPA roadblock. This implicates HelQ in genome stability which is explore further in chapter 6.

Chapter 6: Summary and discussion

6.1 Overview.

The original aims of this project were to investigate the human DNA helicase HelQ by (a) determining the active oligomeric state when bound to DNA; (b) creating a model of unwinding activity; (c) to assess protein interactions with RPA and (d) characterise the role of the N-terminal ORFan region. Work presented here identified HelQ as an active dimer for translocation on ssDNA that is especially reliant on interactions with DNA bases of the translocating strand (chapters 3 and 4). We present evidence for the site of crucial interaction between HelQ and RPA that is bound to DNA, suggesting a model in which HelQ can be recruited to DNA that is exposed during replication stress (chapter 5). We saw that RPA is displaced from DNA by the N-terminal domain of HelQ (N-HelQ), reliant on amino acid residues within a PWI-like domain. This is a significant step forward that is worthy of future research into the mechanism of RPA mediated HelQ loading in the context of the multiple events occurring at stressed replication forks in human cells. While progress was made in investigating HelQ, there are still aspects that require further analysis - these along with the conclusions that can be taken from these findings will be discussed in more detail here.

Archaeal Hel308 has previously been used as a model for human HelQ because it shares high overall amino acid sequence similarity (25.5%; figure 1.20). However, differences observed between the two helicases during this work has led to the reassessment of archaeal Hel308 as a model for HelQ. The crystal structures of Hel308 (section 1.13.1) revealed a monomeric five-domain structure with a central pore lined with DNA binding residues.[309] This monomeric arrangement is different to the dimeric model we propose for HelQ and therefore the helicases may use different mechanisms of action. Additionally, Hel308 can displace a

streptavidin molecule from biotinylated DNA but our data show that HelQ is ineffective at removing DNA-bound barriers.[309] Hel308 is proposed to be involved in the removal of bound proteins at stalled replication forks and recombination intermediates using motive force generated when translocating DNA. Whereas HelQ only appears to displace RPA from DNA by an ATP-independent process. Furthermore, the N-terminal region of HelQ that mediates, at least in part, interactions with RPA is not conserved with archaeal Hel308 (figure 3.9). This suggests a unique role of N-HelQ that is additional to core helicase activity. Finally, mutagenesis of a conserved phenylalanine/tyrosine residue within motif IVa of HelQ did not generate a hyper-active helicase as reported for Hel308.[381]

6.2 Determining the oligomeric state of HelQ when bound to DNA.

The first aim of this project was to determine the oligomeric state of HelQ when activated as an ATPase by binding to DNA. The oligomeric state of a helicase can be important for translocation and unwinding activity as it determines how a helicase interacts with the DNA (see section 1.12.2). Furthermore, oligomerisation is a level of regulation used to control helicase activity.[204]

Native PAGE, AGF and SEC MALS (figures 3.18-3.20) implicated that HelQ is active as a dimer in the presence of ATP-Mg^{2+} and ssDNA. The shift in apparent HelQ mass caused by activating the apoenzyme form adds new insight to HelQ function- previously it had been suggested HelQ was active when hexameric, but we believe this is an aggregated inactive form of HelQ. We suggest here that dimeric HelQ binds ssDNA and in an ATP-dependent manner translocates and unwinds the fork (figure 3.20) evident from AGF fractions isolated as dimers being active in both binding and unwinding DNA.

Bioinformatic analysis to predict the oligomeric state of HelQ, based on sequence homology to PolQ, hypothesised a tetrameric structure for the apoenzyme (figure 3.8).[248,250] Tetrameric apoenzyme PolQ relies on an additional extended loop region in domain IV not seen in other SF2 helicases, that therefore may explain why HelQ does not show this similar tetramer structure when translocating on DNA. However, while HelQ did not form tetramers *in vitro*, we cannot rule out that a dimeric dimer pattern of assembly may be important for certain activities of HelQ. This led us to question whether the active oligomeric state of HelQ changes based on its role at specific times. The helicases Rep, UvrD and RecQ4 are noted for varying oligomeric states modulating helicase function. These helicases can have varying levels of enzyme processivity dependent on oligomeric state.[382] Therefore, this may be the case for HelQ, whereby different oligomeric states give rise to alternative roles. For example, Ward *et al.* showed an ATPase inactive HelQ (Lys-197-Arg) rapidly disassembled Rad51-dsDNA filaments whereas wild type HelQ did not.[253] This indicates that HelQ may be involved in other DNA repair roles aside from its helicase activity and may depend on protein state. Cryogenic electron microscopy (CryoEM) could be used to explore these, as well as dimeric HelQ further dimerising, by solving the structure of HelQ in the presence and absence of DNA, activating components and protein partners to confirm a shift in oligomeric assembly.

By determination of the active form of HelQ and analysis of its unwinding activity using chemically modified substrates we now have a clear picture for the first time of the most fundamental aspect of HelQ function- its ability to unwind DNA. Results presented in chapter 4 suggest that HelQ relies on interactions with the translocating strand and internal bases within the duplex region and to a lesser

extent the DNA backbone (figure 4.6). Interactions with the DNA bases implicates the winged-helix domain (WHD) in DNA unwinding by acting as a wedge interacting with the base closest to the DNA branch point. Archaeal Hel308 has a β-hairpin within the WHD which acts as a pin to separate DNA at the ssDNA-dsDNA junction (section 1.12.3).[203] This suggests HelQ would translocate along the DNA, or could potentially reel in the DNA, and the WHD would force a gap into the ssDNA-dsDNA branch point separating the strands.[204,206 383] DNA reeling is not suggested by data including chemical modifications, however, we cannot rule this out as a mechanism.

The presence of modifications on the non-translocating strand had no impact on HelQ and therefore HelQ is likely to interact significantly only with the translocating strand (figures 4.7 and 4.8). HelQ does not appear to load onto dsDNA or require interactions with the non-translocating strand. Therefore, the insertion of a wedge into the junction point from HelQ loaded onto the translocating strand suggests HelQ directly destabilises the duplex and implies an active mechanism of unwinding (chapter 4).[204,205] The unidirectional movement of HelQ along the DNA would result in stabilisation of the ssDNA and subsequent dissociation and re-association of the duplex DNA at the junction, also known as DNA 'breathing'. The WHD wedge of HelQ would then be able to insert into the gap created during intermittent dissociating base pairs which would result in DNA separation allowing helicase translocation forward upon ATP-hydrolysis.[384] The specific directional translocation of HelQ may be necessary to ensure the WHD is orientated in the correct direction facing the ds-ssDNA junction. Analysis of other helicases with a WHD, including Hel308, BLM and WRN, could be used to predict residues within the WHD in HelQ important in DNA binding.[209]

Multiple models of mechanism action can be hypothesised for HelQ. The interaction of dimeric HelQ with one DNA strand and an active WHD wedge action suggests that HelQ may follow an inch-worm variation, described for many dimeric SF2 helicases.[384] Both monomeric and dimeric UvrD and Rep helicases follow a type of inch-worm mechanism which relies on the 'rolling' of a multi-subunit helicase along DNA.[382] The cooperative inch-worm mechanism results in increased unwinding with additional monomers as well as protein displacement and the ability to by-pass DNA lesions. This suggests that the oligomeric state and the subsequent loading of the helicase onto DNA may result in different activities. This may be important in HelQ activity and therefore further single molecule analysis of HelQ could be used to differentiate between these potential different activities.

6.3 Assessment of interactions between HelQ and DNA.

The second aim of the work carried out here was to assess DNA binding, unwinding and ATPase activities of HelQ, and the C- and N-HelQ fragments. HelQ was assessed with chemically modified DNA substrates as discussed in the section above which provided insight into the mechanism of translocation. The chemically modified substrates did not impact the ATPase activity of HelQ suggesting HelQ embarks on futile cycles of ATP hydrolysis stimulated by binding and re-binding to DNA. Stopped-flow analysis is necessary to confirm that the ATPase activity remains normal even in the absence of translocation. Furthermore, C-HelQ behaved similarly to HelQ implicating the 'core' helicase as solely responsible for translocation and unwinding activity. N-HelQ had no DNA binding or unwinding activity evident from EMSAs and fluorescence anisotropy (see later).

Additional roles of HelQ with DNA were assessed (figures 4.14-4.17). Helicases have been associated with branch migration, strand re-annealing, protein

displacement from DNA and the resolution of alternative DNA substrates.[186] *In vitro* analysis did not implicate HelQ in the resolution of Holliday junctions or G4 structures, in re-annealing of oligos or displacement of protein barriers on the DNA. HelQ was unable to process an extended fork substrate or D-loop suggesting the primary role of HelQ in repair is not in the processing of these specific substrates (figure 4.35).

A highly conserved phenylalanine residue within motif IV of SF2 DEAD-box helicases is thought to be important for activity.[370] This conserved phenylalanine, found in archaeal Hel308, gave a hyperactive unwinding phenotype when mutated.[381] The phenylalanine appears to be an anchor for maintaining the rigidity of the RecA-like domain. While the phenylalanine was not conserved in HelQ, a motif IV domain was identified with tyrosine. Mutagenesis of the tyrosine residue in C-HelQ (C-HelQ^{Y642A}) did not give the hyperactive phenotype but reduced unwinding ability (figures 4.21-4.24). Interestingly, EMSAs of the mutant protein resulted in different migration patterns to the wild type protein implicating the residue in protein assembly. Potentially mutagenesis removed the ability of the protein to dimerise which would explain why protein activity was lost. Mutagenesis of the residue in bacterial SF2 helicases results in sensitivity to temperature implicating it in conformational stability.[370] While temperature changes in eukaryotic cells is not physiologically relevant, it may implicate the motif in conformational stability of HelQ and requires further exploration.

C-HelQ were able to unwind RNA-containing fork substrates, by directional translocation of RNA instead of DNA. The absence of unwinding by FL-HelQ may be due to the poor quality of the protein or may suggest a role of N-HelQ in nucleic acid identification. N-HelQ may be important in differentiating between RNA and DNA and therefore be responsible for directing FL-HelQ to DNA only substrates.

This would require further analysis to assess FL-HelQ with RNA containing substrates with a loss-of-activity mutant of N-HelQ. This hypothesis requires further exploration to establish the extent of how C-HelQ behaves with RNA.

Furthermore, bioinformatic analysis using STRING 9 resulted in several hits between HelQ and RNA splicing proteins (see chapter 1). This included Dhx8, the *DEAH-box polypeptide 8 protein* that facilitates with nuclear exportation of spliced mRNA by releasing the RNA from the spliceosome and Cwc22, *the spliceosome-associated protein*.[385] HelQ was also predicted to interact with cell cycle proteins Polymerase α1 and Xab1 (*XPA binding protein 2*) involved in transcription-coupled repair important in pre-mRNA splicing. These interactions and a role in processing RNA may implicate HelQ in splicing. Furthermore, structural homology to the PWI-like domain found in the Brr2 splicing helicase further links HelQ to RNA splicing or an association with splicing repair (figure 3.13).[329] There are eight SF2 RNA helicases required for pre-mRNA splicing in eukaryotes of which the Ski2-like Brr2 helicase is one.[353]

Helicases also remove RNA at sites of DSBs which prevent efficient repair, e.g. the RNA-unwinding protein *DEAD box 1* (DDX1).[385] HelQ may have a role in removing RNA at DSBs to aid replication restart (figure 4.35-4.36).

6.4 Assessment of protein interactions between HelQ and RPA.

The next aim of this work was to learn more about the interaction of HelQ and RPA both physically and functionally. As seen in chapter 5, we were able to reconstitute *in vitro* the physical interaction of HelQ with RPA that had been reported in human cells (figure 5.3).[252] Coupled with the co-localisation of HelQ with Rad51 and ATR previously observed *in vivo*, this implicates a role of HelQ in HR damage response.[253] No other impact of RPA on HelQ activity was observed other than the potentially recruitment to exposed DNA. RPA has been shown to

impact the activity of the WRN helicase after recruitment.[386,387] We did not identify any further impacts of RPA on activating downstream activity of HelQ, however, this may rely on the presence of additional factors; future work could assess additional proteins such as the Rad51 paralogues, or specific structures, such as R-loops. However, the presence of RPA on DNA did not inhibit HelQ activity either. RPA did not act as a roadblock to HelQ suggesting that HelQ was able to translocate through or displace RPA bound to DNA. This implicates HelQ in protein clearing of RPA which may be important in regulating HR. HelQ may displace RPA to aid in the assembly of Rad51 and the paralog proteins or to displace RPA to prevent downstream DSBR activation. Further assessment of HelQ with other proteins in HR would be beneficial in determining where HelQ acts within this pathway.

We next aimed to isolate the region responsible for the interaction, however, the interaction was not isolated to either the N- or C-terminal of HelQ implicating either multiple points of contact or the protein fragments were too unstable to isolate the region of interaction (figures 5.3-5.6). A study carried out on WRN and BLM identified the region responsible for interactions to RPA and the subsequent impact on recruitment by RPA.[386] WRN has two RPA binding sites, a high affinity N-terminal site and a lower affinity C-terminal site and interactions with RPA elicit different responses.[386] Dual, or more, RPA binding sites may provide one explanation for our inability to determine a physical interaction of RPA with either N-HelQ or C-HelQ. Another explanation for the lack of detected interactions is the instability of C-HelQ, evident during purification and experiments presented in chapter 3. C-HelQ, when assessed alone, may be too unstable to form complexes with RPA and therefore requires the presence of N-HelQ for stability and to establish interactions. Further analysis looking at FL-HelQ with mutant N-HelQ could be used to confirm the presence of these interactions.

Isolation of the domain of RPA that interacts with HelQ may also aid in isolating additional effects on HelQ activity. RPA has a large influence on downstream proteins dependent on the specific connections made. For example studies have identified that the WRN, BLM and RecQL1 SF2 helicases interact with RPA70.[386,387] Analysis with the DNA binding mutant Aro1-RPA could be used to assess the importance of DNA activated complexes of RPA on HelQ recruitment.[374]

ATR pulls down with RPA and HelQ individually and therefore the specific relationship between HelQ and ATR with RPA could also be researched.[236,253,256] ATR is a checkpoint damage response kinase prominent in activation of downstream repair (chapter 1) and therefore interactions with HelQ would not be surprising. Potential interactions between HelQ and the BCDX2 paralogue complex identified by mass spectroscopy could also be expanded.[237] The BCDX2 paralogue complex, involved in early stage HR, may also implicate HelQ in early HR. Biochemical analysis of the human paralogue recombinant proteins has proved difficult, however, an emerging model organism, *Trypanosomes brucei,* with predicted homologues may aid in this analysis.[388] We speculate that these interactions may be important for the negative regulation of HR by HelQ. Additionally, analysis of HelQ with the CX3 complex may implicate HelQ in late stage HR, yet to be determined, or establish its role solely in early HR.

6.5 Characterisation of the active core and catalytically inactive ORFan domain of HelQ.

The final aim of this work arose during purification of full length HelQ protein, providing opportunity to assess individually the helicase active core (C-HelQ) and catalytically inactive N-terminal ORFan domain (N-HelQ). As discussed in previous sections, C-HelQ is active as a helicase showing similar DNA processing

to FL-HelQ and N-HelQ did not. We concluded that RPA recruitment is likely to be essential in activation of HelQ, however, we could not isolate this interaction. We did show that N-HelQ activity may reduce RPA-DNA binding (figure 5.19). N-HelQ may therefore be important in the negative regulation of HR to prevent activation of further repair proteins by RPA as with the FancJ and RecQ1 helicases, or in promoting assembly of downstream proteins by displacing RPA and exposing ssDNA.[374]

The ability to displace RPA was isolated to the PWI-like domain (figure 5.19-5.22). One way in which this may occur is that the PWI-like domain transiently interacts with RPA resulting in its dissociation from DNA (figure 5.24).[329] However, no such interactions have been detected. The lack of observed protein-protein interactions may highlight how transient the interaction is during displacement (figures 5.26-5.30). Further interaction analysis specific to the PWI-like domain of N-HelQ is required. Ultra-weak interactions (K_d > 100 μM) between proteins have become an increasing focus in research with NMR being a major player detecting interactions as weak as 3 mM.[389] NMR is underway to assess for any interactions between N-HelQ and RPA.[389]

N-HelQ may disrupt the heterotrimeric structure of RPA preventing the easy re-annealing of RPA to DNA. Native PAGE did not show dissociation of the trimeric RPA with addition of N-HelQ; however, more extensive biochemical analysis is required (figure 5.25). The interaction may be isolated to one region of RPA and therefore was not observed in work carried out here. As discussed previously, for a stable complex to form elements of both N- and C-HelQ may be required and therefore the absence of stable complex formation results in the absence of observed interactions.

Regions of intrinsic disorder (IDPRs) in helicases are associated with multiple roles. The intrinsically disordered N-terminal of the human RecQL4 is important for DNA binding to ssDNA, dsDNA and forked structures and in the resolution of G4 quadruplexes.[390] Disorder has also been shown to be active in protein coupling and binding. The disorder-function paradigm is still an emerging field with activation requirements still mostly unknown.[322,324] A lack of structure prevents prediction of useful functional annotation making characterisation difficult. The N-terminal of HelQ, while shown to be unessential in binding and unwinding, is likely required for a unique activity further to the RPA displacement activity that relies on the structured PWI-like domain (figures 4.25-4.27). This work did not identify a specific function of the IDPR. Further work could assess modifications or protein partners of N-HelQ IDPR. PTMs may be required to activate its role such as SUMO-interactions via SIMs discussed in chapter 3 (figure 3.15). Alternatively, the IDPR may interact with other proteins not tested here e.g. Rad51 or the BCDX2 paralogue complex. If interactions with the Rad51 proteins were detected, this could suggest HelQ involvement in Rad51 disassembly.[324]

In addition to analysis of N-HelQ independently, C-HelQ was also assessed for isolated activity. As discussed, C-HelQ was similar to FL-HelQ in terms of protein interactions and DNA processing. This acted as a useful control to isolate N-HelQ function. However, purification of C-HelQ resulted in a brown colouration in the fractions which may implicate the presence of an iron-sulphur cluster (figure 3.6). Members of the SF2 Fe-S cluster family include the DDX11, RTEL1 and FancJ helicases that are essential for DNA repair and genome stability.[385] Similarities to the FancJ helicase are unsurprising due to HelQ association with the FA pathway and RPA recruitment activities.[186] While preliminary bioinformatic analysis did not predict Fe-S with a high probability, this cannot be discounted. Alternatively, C-HelQ may interact with a Fe-S protein eluting with it during purification. Mass

spectroscopy could be used to determine if Fe-S containing proteins are present in these fractions. Additionally, analysis of the regions with low probability Fe-S predicted clusters could be used to assess the presence of a Fe-S cluster.

6.6 Supporting the hypothesis that HelQ is active in HR.

Work carried out in this project and summarised in this chapter added to the fundamental biochemical understanding of HelQ function. We also aimed to determine how these findings aid in the understanding of HelQ physiologically. Previous evidence indicated HelQ supporting active DNA replication by suppressing DSBR in favour of alternative pathways of HR. At the onset of this work, we hypothesised that HelQ is involved specifically in the negative regulation of DSBR in order to prevent cross-over events that can lead to genetic instability. Since the onset of this work, we considered a role for HelQ in the regulation of BIR to prevent cross-over events. RPA recruitment of HelQ to exposed DNA and displacement of RPA by HelQ may dampen the signal of RPA in recruiting downstream repair proteins e.g. Rad51, the paralogues, BRCA and FA proteins and subsequently preventing NPF formation. This supports the hypothesis that HelQ may act as a negative regulator of HR to promote replication restart and aid in the prevention of detrimental late stage HR introducing genetic variation.[138] RPA displacement may alternatively promote Rad51 and the paralogues binding in favour of HR activation. Further analysis is required to assess the role of HelQ in negative regulation or promoting HR.

HelQ has been linked to both RPA and BCDX2 involved in early stage HR and therefore we can predict HelQ is also likely to be involved in these early stages. Furthermore, the absence of D-loop disassembly by HelQ *in vitro* implies early regulation due to the presence of D-loops found during late-stage HR. Another hypothesise is that HelQ may be involved in regulating BIR. BIR occurs where only

one broken end can invade the homologous pathway and, as discussed in chapter 1, can result in conservative replication and cross-over events. Therefore, like DSBR, there are proteins involved in preventing BIR continuing past DNA repair which may lead to genome instability. HelQ may be involved in unwinding newly synthesised DNA during BIR to prevent replication continuing in this manner. Elaboration on the work carried out here is required to pinpoint the exact location that HelQ acts in HR and to build on this hypothesis, however, evidence presented here strongly supports a role of HelQ in regulating HR.

Bibliography

1. Northall, S. J., Ivancic-Bace, I., Soultanas, P. & Bolt, E. L. Remodeling and control of homologous recombination by DNA helicases and translocases that target recombinases and synapsis. *Genes (Basel).* **7**, 1–12 (2016).
2. Crick, F. & Watson, J. Molecular Strucutre of Nucleic Acids. *Nature* **171**, 737–738 (1953).
3. Pray, L. A. Discorvery of DNA Strucutre and Function: Watson and Crick. *Nat. Educ.* **1**, 100 (2008).
4. Berg, J. M., Tymoczko, J. L. & Stryer, L. Protein Structure and Function. in *Biochemistry* (W H Freeman, 2002).
5. Palazzo, A. F. & Gregory, T. R. The Case for Junk DNA. *PLoS Genet.* **10**, (2014).
6. Genome News Network. Genome Variations. *J. Craig Venter Institute* (2004). Available at: http://www.genomenewsnetwork.org/resources/whats_a_genome/Chp4_1.shtml. (Accessed: 1st September 2020)
7. Macalpine, D. M. & Almouzni, G. Chromatin and DNA Replication. *Cold Spring Harb. Perspect. Biol.* **5**, 1–22 (2013).
8. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
9. Pierce, B. *Genetics- A conceptual Approach*. (2006).
10. Zimmerman, S. B. The Three-Dimersional Structure of DNA. *Annu. Rev. Biochem.* **51**, 395–427 (1982).
11. Annunziato, A. T. DNA Packaging: Nucleosomes and Chromatin. *Nat. Educ.* **1**, 26 (2008).
12. Ussery, D. W. DNA Structure: A-, B- and Z-DNA Helix Families. *Encycl. Life Sci.* 1–7 (2002).
13. Ravichandran, S., Subramani, V. K. & Kim, K. K. Z-DNA in the genome: from structure to disease. *Biophys. Rev.* **11**, 383–387 (2019).
14. Donnell, M. O., Langston, L. & Stillman, B. Principles and Concepts of DNA Replication in Bacteria, Archaea, and Eukarya. *Cold Spring Harb. Perspect. Biol.* **3**, 1–13 (2013).
15. del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R. Replication and Control of Circular Bacterial Plasmids. *Microbiol. Mol. Biol. Rev.* **62**, 434–464 (1998).
16. Milo, R. & Phillips, R. *Cell Biology by the Numbers*. (Garland Science, 2015).
17. Lodish, H., Berk, A. & Zipursky, S. L. The Three Roles of RNA in Protein Synthesis. in *Molecular Cell Biology. 4th edition.* (ed. Freeman, W. H.) (2000).
18. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).
19. Schafer, K. A. The Cell Cycle: A Review. *Vet. Pathol.* **35**, (1998).
20. Burgers, P. M. J. & Kunkel, T. A. Eukaryotic DNA Replication Fork. *Annu. Rev. Biochem.* 1–22 (2017).
21. Sun, J. *et al.* The architecture of a eukaryotic replisome. *Nat. Struct. Mol. Biol.* **22**, 976–982 (2015).
22. Leman, A. R. & Noguchi, E. The Replication Fork: Understanding the Eukaryotic Replication Machinery and the Challenges to Genome Duplication. *Genes (Basel).* **4**, 1–32 (2013).
23. Bleichert, F. Mechanisms of replication origin licensing: a structural perspective. *Curr. Opin. Struct. Biol.* **59**, 195–204 (2019).
24. Takayama, Y. *et al.* GINS, a novel multi-protein complex required for chromosomal DNA replication in budding yeast. *Genes Dev.* **17**, 1153–1165 (2003).
25. Pacek, M., Tutter, A. V., Kubota, Y., Takisawa, H. & Walter, J. C. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol. Biol. Cell* **21**, 581–587 (2006).
26. Bochman, M. L. & Schwacha, A. The Mcm2-7 Complex Has In Vitro Helicase Activity. *Mol. Cell* **31**, 287–293 (2008).
27. Yuan, Z. *et al.* Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation. *Nat. Publ. Gr.* **23**, 217–224 (2016).
28. Masai, H. *et al.* Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. *J. Biol. Chem.* **281**, 39249–39261 (2006).
29. Gambus, A. *et al.* GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Chem Biol* **8**, 358–366 (2006).
30. Frick, D. N. & Richardson, C. C. DNA Primases. *Annu. Reivew Biophys.* **70**, 39–80

(2001).
31. Hanawalt, P. C. & Holmes, F. L. Density matters : The semiconservative replication of DNA. *PNAS* **101**, 17889–17894 (2004).
32. Meselson, M. & Stahl, F. W. The Replication of DNA in E. coli. *PNAS* **44**, 671–682 (1958).
33. Wigley, D. B. Structure and Mechanism of DNA Gyrase. *Nucleic Acids Mol. Biol.* **9**, 165–176 (1995).
34. Arias-Gonzalez, J. R. A DNA-centered explanation of the DNA polymerase translocation mechanism. *Sci. Rep.* **7**, 1–8 (2017).
35. Yang, W., Lee, J. Y. & Nowotny, M. Making and Breaking Nucleic Acids: Two-Mg2+-Ion Catalysis and Substrate Specificity. *Mol. Cell* **22**, 5–13 (2006).
36. Pfeiffer, V. & Lingner, J. Replication of telomeres and the regulation of telomerase. *Cold Spring Harb. Perspect. Biol.* **5**, 1–22 (2013).
37. Waga, S., Bauer, G. & Stillman, B. Reconstitution of complete SV40 DNA replication with purified replication factors. *J. Biol. Chem.* 10923–10934 (1994).
38. Lindahl, T. & Barnes, D. E. Mammalian DNA ligases. *Annu. Reivew Biochem.* **61**, 251–281 (1992).
39. Dewar, J. M. & Walter, J. C. Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell Biol.* **18**, 507–516 (2017).
40. Steitz, T. A. DNA polymerases: Structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).
41. Lodish, H., Berk, A. & Zipursky, S. L. DNA Damage and Repair and Their Role in Carcinogenesis. *Mol. Cell Biol. 4th Ed.* **15**, (2000).
42. Gonzalez-Magana, A. & Blanco, F. J. Human PCNA Structure, Function and Interactions. *Biomolecules* **10**, 1–19 (2020).
43. Cai, J. *et al.* ATP hydrolysis catalyzed by human replication factor C requires participation of multiple subunits. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11607–11612 (1998).
44. Ansbach, A. B. *et al.* RFC Ctf18 and the Swi1-Swi3 complex function in separate and redundant pathways required for the stabilization of replication forks to facilitate sister chromatid cohesion in Schizosaccharomyces pombe. *Mol. Biol. Cell* **19**, 595–607 (2008).
45. Skibbens, R. V., Corson, L. B., Koshland, D. & Hieter, P. Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev.* **13**, 307–319 (1999).
46. Zetterberg, A., Larsson, O. & Wiman, K. G. What is the restriction point? *Curr. Opin. Cell Biol.* **7**, 835–842 (1995).
47. Malumbres, M. & Barbacid, M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer* **9**, 153–166 (2009).
48. Darzynkiewicz, Z., Gong, J., Juan, Gl., Ardelt, B. & Traganos, F. Cytometry of cyclin proteins. *J. Quant. Cell Sci.* **25**, 1–13 (1996).
49. Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267–271 (1996).
50. Reinhardt, H.C. and Yaffe, M. . Kinases that Control the Cell Cycle in Response to DNA Damage: Chk1, Chk2 and MK2. *Curr. Opin. Cell Biol.* **21**, 245–255 (2009).
51. Lee, J., Kumagai, A. & Dunphy, W. G. Positive regulation of Wee1 by Chk1 and 14-3-3 proteins. *Mol. Biol. Cell* **12**, 551–563 (2001).
52. Kato, J., Matsushime, H., Hiebert, S. W., Ewen, M. K. & Sherr, C. Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes Dev.* **7**, 331–342 (1993).
53. Arroyo, M. & Raychaudhuri, P. Retinoblastoma-repression of E2F-dependent transcription depends on the ability of the retinoblastoma protein to interact with E2F and is abrogated by the adenovirus E1A oncoprotein. *Nucleic Acid Res.* **20**, 5947–5954 (1992).
54. Mudrak, I., Ogris, E., Rotheneder, H. & Wintersberger, E. Coordinated trans activation of DNA synthesis- and precursor-producing enzymes by polyomavirus large T antigen through interaction with the retinoblastoma protein. *Mol. Cell. Biol.* **14**, 1886–1892 (1994).
55. Hall, M., Bates, S. & Peters, G. Evidence for different modes of action of cyclin-dependent kinase inhibitors: p15 and p16 bind to kinases, p21 and p27 bind to cyclins. *Oncogene* **11**, 1581–1588 (1995).
56. Harper, J. W. *et al.* Inhibition of cyclin-dependent kinases by p21. *Mol. Biol. Cell* **6**, 357–484 (1995).

57. Chen, J., Saha, P., Kornbluth, S., Dynlacht, B. D. & Dutta, A. Cyclin-binding motifs are essential for the function of p21CIP1. *Mol. Cell. Biol.* **16**, 4673–4682 (1996).
58. LaBaer, J. *et al.* New functional activities for the p21 family of CDK inhibitors. *Genes Dev.* **11**, 847–862 (1997).
59. Maity, A., Gillies McKenna, W. & Muschel, R. J. The molecular basis for cell cycle delays following ionizing radiation: a review. *Radiother. Oncol.* **31**, 1–13 (1994).
60. Ingaramo, M. C., Sánchez, J. A. & Dekanty, A. Regulation and function of p53: A perspective from Drosophila studies. *Mech. Dev.* **154**, 82–90 (2018).
61. Greenblatt, Ma. S., Bennett, W. P., Hollstein, M. & Harris, C. C. Mutations in the p53 Tumor Suppressor Gene: Clues to Cancer Etiology and Molecular. *Pathogenesist* **54**, 4855–4878 (2006).
62. Shell, S. M., Zou, Y., Liu Y. and Wu, X. Functions of Human Replication Protein A (RPA): From DNA Replication to DNA Damage and Stress Responses. *J. Cell. Physiol.* **208**, 267–273 (2011).
63. Kibe, T., Ono, Y., Sato, K. & Ueno, M. Fission yeast Taz1 and RPA are synergistically required to prevent rapid telomere loss. *Mol. Biol. Cell* **18**, 2378–2387 (2007).
64. Safa, L. *et al.* Binding polarity of RPA to telomeric sequences and influence of G-quadruplex stability. *Biochimie* **103**, 80–88 (2014).
65. Murat, P. & Balasubramanian, S. Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.* **25**, 22–29 (2014).
66. De Lange, T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).
67. Audry, J. *et al.* RPA prevents G-rich structure formation at lagging-strand telomeres to allow maintenance of chromosome ends. *EMBO J.* **34**, 1342–1958 (2015).
68. Chen, R. & Wold, M. S. Replication Protein A: Single-stranded DNA's first responder : Dynamic DNA-interactions allow Replication Protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays* **36**, 1156–1161 (2014).
69. Flynn, R. L. & Zou, L. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Crit Rev Biochem Mol Biol* **45**, 266–275 (2010).
70. Murzin, A. G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861–867 (1993).
71. Gibb, B. *et al.* Concentration-dependent exchange of replication protein A on single-stranded DNA revealed by single-molecule imaging. *PLoS One* **9**, 1–13 (2014).
72. Kim, C., Paulus, B. F. & Wold, M. S. Interactions of human replication protein A with oligonucleotides. *Biochemistry* **33**, 14197–14206 (1994).
73. Blackwell, L. J. & Borowiec, J. A. Human replication protein A binds single-stranded DNA in two distinct complexes. *Mol. Cell. Biol.* **14**, 3993–4001 (1994).
74. Gomes, X. V. & Wold, M. S. Functional Domains of the 70-kilodalton subunit of the Replication Protein A. *Biochemistry* **35**, 10558–10568 (1996).
75. Fanning, E., Klimovich, V. & Nager, A. R. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acid Res.* **34**, 4126–4137 (2006).
76. Fan, J. & Pavletich, N. P. Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.* **16**, 2337–2347 (2012).
77. Bochkareva, E., Korolev, S., Lees-Miller, S. P. & Bochkarev, A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863 (2002).
78. Pokhrel, N. *et al.* Dynamics and selective remodeling of the DNA-binding domains of RPA. *Nat. Struct. Mol. Biol.* **26**, 129–136 (2019).
79. Awasthi, P., Foiani, M. & Kumar, A. ATM and ATR signaling at a glance. *J. Cell Sci.* **128**, 4255–4262 (2015).
80. Templeton, G. W. & Moorhead, G. B. G. The phosphoinositide-3-OH-kinase-related kinases of Arabidopsis thaliana. *EMBO Rep.* **6**, 723–728 (2005).
81. Brown, E. J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation and early embryonic lethality. *Genes Dev.* **14**, 397–402 (2000).
82. de Klein, A. *et al.* Targeted disruption of the cell-cycle checkpoint gene ATR leads to early embryonic lethality in mice. *Curr. Biol.* **10**, 479–482 (2000).
83. Shechter, D., Costanzo, V. & Gautier, J. Regulation of DNA replication by ATR: signaling in response to DNA intermediates. *DNA Repair (Amst).* **3**, 901–908 (2004).
84. Cortez, D., Guntuku, S., Qin, J. & Elledge, S. J. ATR and ATRIP: partners in checkpoint

signaling. *Science (80-.).* 1713–1716 (2001).
85. Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA–ssDNA complexes. *Science (80-.).* **300**, 1542–1548 (2003).
86. Ball, H. L. *et al.* Function of a conserved checkpoint recruitment domain in ATRIP proteins. *Mol. Cell. Biol.* **27**, 3367–3377 (2007).
87. Cimprich, K. A. & Cortez, D. ATR: An Essential Regulator of Genome Integrity. *Nat. Rev. Mol. Cell Biol.* **9**, 616–627 (2008).
88. Delacroix, S., Wagner, J. M., Kobayashi, M., Yamamoto, K. & Karnitz, L. M. The Rad9–Hus1–Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. *Genes Dev.* **21**, 1472–1477 (2007).
89. Marechal, A. & Zou, L. DNA Damage Sensing by the ATM and ATR Kinases. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
90. Lee, J., Kumagai, A. & Dunphy, W. G. The Rad9Hus1–Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. *J. Biol. Chem.* 28036–28044 (2007).
91. Kumagai, A., Lee, J., Yoo, H. Y. & Dunphy, W. G. TopBP1 activates the ATR–ATRIP complex. *Cell* 943–955 (2006).
92. Melo, J. A., Cohen, J. & Toczyski, D. P. Two checkpoint complexes are independently recruited to sites of DNA damage in vivo. *Genes Dev.* **16**, 198–208 (2001).
93. Zou, L., Cortez, D. & Elledge, S. J. Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev.* **2**, 198–208 (2002).
94. Choi, J. H., Lindsey-Boltz, L. A. & Sancar, A. Reconstitution of a human ATR-mediated checkpoint response to damaged DNA. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13301–13306 (2007).
95. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science (80-.).* **316**, 1160–1166 (2007).
96. Smolka, M. B., Albuquerque, C. P., Chen, S. H. & Zhou, H. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc. Natl. Acad. Sci. U. S. A.* 10364–10369 (2007).
97. Mu, J. J. *et. al.* A proteomic analysis of ataxia telangiectasia-mutated (ATM)/ATM-Rad3-related (ATR) substrates identifies the ubiquitin-proteasome system as a regulator for DNA damage checkpoints. *J. Biol. Chem.* 17330–17334 (2007).
98. Kumagai, A. & Dunphy, W. G. Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in Xenopus egg extracts. *Mol. Cell* **4**, 839–849 (2000).
99. Wang, X. Rad17 phosphorylation is required for claspin recruitment and Chk1 activation in response to replication stress. *Mol. Cell* **3**, 331–341 (2006).
100. Furnari, B., Rhind, N. & Russell, P. Cdc25 mitotic inducer targeted by chk1 DNA damage checkpoint kinase. *Science (80-.).* **5331**, 1495–1497 (1997).
101. Sanchez, Y. *et al.* Conservation of the Chk1 checkpoint pathway in mammals: linkage of DNA damage to Cdk regulation through Cdc25. *Science (80-.).* **277**, 1497–7501 (1997).
102. Matos, D. A. *et al.* ATR Protects the Genome against R Loops through a MUS81-Triggered Feedback Loop. *Mol. Cell* **77**, 514–527 (2020).
103. Allison, D. F. & Wang, G. G. R-loops: formation, function, and relevance to cell stress. *Cell Stress* **3**, 38–46 (2019).
104. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **3**, 398–411 (2019).
105. Andreassen, P. R., D'Andrea, A. D. & Taniguchi, T. ATR couples FANCD2 monoubiquitination to the DNA damage response. *Genes Dev.* **16**, 1958–1963 (2004).
106. Heffernan, T. P. *et al.* An ATR- and Chk1-dependent S checkpoint inhibits replicon initiation following UVC-induced DNA damage. *Mol. Cell. Biol.* **22**, 8552–8561 (2002).
107. Feijoo, C. *et al.* Activation of mammalian Chk1 during DNA replication arrest: a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. *J. Cell Biol.* **154**, 913–923 (2001).
108. Trenz, K., Smith, E., Smith, S. & Costanzo, V. ATM and ATR promote Mre11 dependent restart of collapsed replication forks and prevent accumulation of DNA breaks. *EMBO J.* **8**, 1764–1774 (2006).
109. Dimitrova, D. S. & Gilbert, D. M. Temporally coordinated assembly and disassembly of replication factories in the absence of DNA synthesis. *Nat. Rev. Mol. Cell Biol.* **10**, 686–694 (2000).
110. Bao, S. et al. ATR/ATM-mediated phosphorylation of human Rad17 is required for

genotoxic stress responses. *Nature* 969–974 (2001).
111. Brush, G. S., Morrow, D. M., Hieter, P. & Kelly, T. J. The ATM homologue MEC1 is required for phosphorylation of replication protein A in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 15075–15080 (1996).
112. Oakley, G. G. *et al.* UV-induced hyperphosphorylation of replication protein a depends on DNA replication and expression of ATM protein. *Mol. Biol. Cell* **12**, 1199–1213 (2001).
113. Ge, X. Q., Jackson, D. A. & Blow, J. J. Dormant origins licensed by excess Mcm2–7 are required for human cells to survive replicative stress. *Genes Dev.* **24**, 3331–3341 (2007).
114. Trenz, K., Errico, A. & Costanzo, V. Plx1 is required for chromosomal DNA replication under stressful conditions. *EMBO J.* **6**, 876–885 (2008).
115. Clancy, S. Mechanisms for Maintaining DNA Integrity. *Nat. Educ.* **1**, 103 (2008).
116. Mastrangelo, G., Fadda, E. & Marzia, V. Polycyclic Aromatic Hydrocarbons and Cancer in Man. *Environ. Health Perspect.* **104**, 1166–1170 (1996).
117. Common types of DNA damage Different types of repair fix different types of damage. *Berkeley Education* (2007). Available at: https://mcb.berkeley.edu/courses/mcb110spring/2007L9to10.pdf.
118. Cannan, W. J. & Pederson, D. S. Mechanisms and Consequences of Double-strand DNA Break Formation in Chromatin Wendy. *J. Cell. Physiol.* **231**, 3–14 (2016).
119. Lengar, P. An analysis of substitution, deletion and insertion mutations in cancer genes. *Nucleic Acid Res.* **40**, 6401–6413 (2012).
120. Shah, R. DNA: Damage Types and Repair Mechanisms. (2016). Available at: http://www.biologydiscussion.com/dna/dna-damage-types-and-repair-mechanisms-with-diagram/16332. (Accessed: 17th July 2019)
121. Lunec, J., Cooke, M. S., Evans, M. D. & Dizdaroglu, M. Oxidative DNA Damage: mechanisms, mutation, and disease. *FASEB J.* **17**, 1195–1214 (2003).
122. Shah, R. DNA : Damage Types and Repair Mechanisms (With Diagram). *Biology Discussion* Available at: http://www.biologydiscussion.com/dna/dna-damage-types-and-repair-mechanisms-with-diagram/16332.
123. Klages-Mundt, N. L. & Li, L. Formation and repair of DNA-protein crosslink damage. *Sci. China Life Sci.* **60**, 1065–1076 (2017).
124. Hunang, Y. & Li, L. DNA Crosslinking damage and cancer- a tale of friend and foe. *Transl. Cancer Res.* **2**, 144–154 (2013).
125. Hunang, Y. & Li, L. DNA Crosslinking damage and cancer- a tale of friend and foe. *Transl. Cancer Res.* **2**, 144–154 (2013).
126. Dexheimer, T. S. DNA Repair Pathways and Mechanisms. in *DNA Repair of Cancer Stem Cells* 19–32 (Springer, 2013).
127. Wallace, S. S., Murphy, D. L. & Sweasy, J. B. Base excision repair and cancer. *Cancer Lett.* **327**, 73–89 (2012).
128. Hsieh, P. & Yamane, K. DNA mismatch repair: Molecular mechanism, cancer and ageging. *Mech. Ageing Dev.* **129**, 391–407 (2009).
129. Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* **2010**, 1–16 (2010).
130. Kunkel, T. A. & Erie, D. A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.* **49**, 291–313 (2015).
131. Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
132. Patrick, S. M. & Turchi, J. J. Xeroderma Pigmentosum Complementation Group A Protein (XPA) Modulates RPA-DNA Interactions via Enhanced Complex Stability and Inhibition of Strand Separation Activity. *J. Biol. Chem.* 16096–16101 (2002).
133. Niraj, J., Färkkilä, A. & D'Andrea, A. D. The Fanconi Anemia Pathway in Cancer. *Annu. Rev. Cancer Biol.* **3**, 457–478 (2019).
134. Gille, J. J. P. *et al.* Diagnosis of fanconi anemia: Mutation analysis by multiplex ligation-dependent probe amplification and PCR-based sanger sequencing. *Anemia* **2012**, (2012).
135. Cristina Negritto, M. Repairing Double-Strand DNA Breaks. *Nat. Educ.* **3**, 26 (2010).
136. Lieber, M. R. The mechanism of human nonhomologous DNA End joining. *J. Biol. Chem.* **283**, 1–5 (2008).
137. Waters, L. S. *et al.* Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiol. Mol. Biol. Rev.* **73**, 134–154 (2009).
138. Renkawitz, J., Lademann, C. a & Jentsch, S. Mechanisms and principles of homology

search during recombination. *Nat. Rev. Mol. Cell Biol.* **15**, 369–83 (2014).
139. Malkova, A. & Ira, G. Break-induced replication: functions and molecular mechanism. *Curr. Opin. Genet. Dev.* **23**, 271–279 (2013).
140. Bhargava, R., Onyango, D. O. & Stark, J. M. Regulation of Single Strand Annealing and its role in genome maintenance Chromosomal break repair by the Single Strand Annealing (SSA) pathway. *Trends Genet* **32**, 566–575 (2016).
141. White, M. F. Homologous recombination in the archaea: the means justify the ends. *Biochem. Soc. Trans.* **39**, 15–9 (2011).
142. Heyer, W.-D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–39 (2010).
143. Sung, P. & Klein, H. Mechanism of homologous recombination : mediators and helicases take on regulatory functions. **7**, 739–750 (2006).
144. Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* **11**, 791–796 (2004).
145. Li, X. & Heyer, W.-D. Homologous recombination in DNA repair and DNA tolernace. *Cell Res.* **18**, 99–113 (2008).
146. Benson, F. E., Baumann, P. & West, S. C. Synergistic actions of Rad51 and Rad52 in genetic recombination and DNA repair. *Nature* 401–404 (1998).
147. Jeggo, P. A., Pearl, L. H. & Carr, A. M. DNA repair, genome stablity and cancer: a historical perspective. *CANCER Nat. Rev.* **16**, 35–42 (2016).
148. Bai, Y. & Symington, L. S. A Rad52 homologue is required for RAD51-independent mitotic recombination in Saccharomyces cerevisiae. *Genes Dev.* **10**, 2025–2037 (1996).
149. Suwaki, N. et al. RAD51 paralogs: Roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin. Cell Dev. Biol.* 898–905 (2011).
150. Amitabh, V. N. & Kowalczykowski, S. C. Second-end DNA capture in double-strand break repair: How to catch a DNA by its tail. *Cell Cycle* **8**, 1816–1817 (2009).
151. Andre Kuzminov. Homologous Recombination-Experimental Systems, Analysis and Significance. *EcoSal Plus* **4**, 1124–1129 (2011).
152. Ristic, D., Modesti, M., Heijden, T. Van Der & Noort, J. Van. Human Rad51 filaments on double- and single-stranded DNA : correlating regular and irregular forms with recombination function. **33**, 3292–3302 (2018).
153. Paoletti, F. *et al.* Molecular flexibility of DNA as a key determinant of RAD 51 recruitment. *EMBO J.* **39**, 1–15 (2020).
154. Son, M. Y. & Hasty, P. Homologous recombination defects and how they affect replication fork maintenance. *AIMS Genet.* **5**, 192–211 (2019).
155. Jiang, S., Lin, T., Xie, Q. & Wang, L. Network Analysis of RAD51 Proteins in Metazoa and the Evolutionary Relationships With Their Archaeal Homologs. *Front. Genet.* **9**, 1–9 (2018).
156. Ravera, S., Dufour, C., Degan, P. & Cappelli, E. Fanconi anemia: from DNA repair to metabolism. *Eur. J. Hum. Genet.* 475–476 (2018).
157. Marians, K. J. Lesion Bypass and the Reactivation of Stalled Replication Forks. 1–22 (2018).
158. Masson, J. *et al.* Identification and purification of two distinct complexes containing the five RAD51 paralogs. 3296–3307 (2001). doi:10.1101/gad.947001.and
159. Masson, J.-Y. *et al.* Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev.* **15**, 3296–3307 (2001).
160. Thacker, J. The RAD51 gene family, genetic instability and cancer. *Cancer Lett.* 125–135 (2005).
161. Yonetani, Y. *et al.* Differential and collaborative actions of Rad51 paralog proteins in cellular response to DNA damage. *Nucleic Acids Res.* **33**, 4544–4552 (2005).
162. Schild, D., Lio, Y., Collins, D. W., Tsomondo, T. & Chen, D. J. Evidence for Simultaneous Protein Interactions between Human Rad51 Paralogs. *J. Biol. Chem.* **275**, 16443–16449 (2000).
163. Badie, S. *et al.* RAD51C facilitates checkpoint signalling by promoting CHK2 phosphorylation. *J. Cell Biol.* 587–600 (2009).
164. Tarsounas, M. *et al.* Telomere maintenance requires the RAD51D recombination/repair protein. *Cell* **3**, 337–347 (2004).
165. Schoenmakers, E. F. P. M., Huysmans, C. & Van de Ven, W. J. M. Allelic knockout of novel splice variants of human recombination repair gene RAD51B in t(12;14) uterine leiomyomas. *Cancer Res.* **59**, 19–23 (1999).

166. Palttari, L. M. *et al.* RAD51C is a susceptibility gene for ovarian cancer. *Hum. Mol. Genet.* **20**, 3278–3288 (2011).
167. Vaz, F. *et al.* Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat. Genet.* **42**, 406–409 (2010).
168. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* **42**, 410–414 (2010).
169. Scully, R. *et al.* Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell* **88**, 265–275 (1997).
170. Prakash, R., Zhang, Y., Feng, W. & Jasin, M. Homologous Recombination and Human Health: The Roles of BRCA1, BRCA2, and Associated Proteins. *Cold Spring Harb. Perspect. Biol.* **7**, 1–27 (2015).
171. Morris, J. R. & Solomon, E. BRCA1: BARD1 induces the formation of conjugated ubiquitin structures, dependent on K6 of ubiquitin, in cells during DNA replication and repair. *Hum. Mol. Genet.* 807–817 (2004).
172. Suszynska, M. *et al.* Bard1 is a low/moderate breast cancer risk gene: Evidence based on an association study of the central european p.q564x recurrent mutation. *Cancers (Basel).* **11**, 1–18 (2019).
173. Vasquez, K. M., Christensen, J., Li, L., Finch, R. A. & GLazer, P. M. Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. *PNAS* **99**, 5848–5853 (2002).
174. Christensen, D. E., Brzovic, P. S. & Klevit, R. E. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat. Struct. Mol. Biol.* 941–948 (2007).
175. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer* **12**, 68–78 (2012).
176. Aparicio, T. & Gautier, J. BRCA1-CtIP interaction in the repair of DNA double-strand breaks. *Mol. Cell. Oncol.* **3**, 1–3 (2016).
177. Nath, S. & Nagaraju, G. *FANCJ helicase promotes DNA end resection by facilitating CtIP recruitment to DNA double-strand breaks*. *PLoS Genetics* **16**, (2020).
178. Carreira, A. *et al.* The BRC repeats of BRCA2 modulate the DNA-binding selectivity of RAD51. *Cell* **6**, 1032–1043 (2009).
179. Liu, J., Doty, T., Gibson, B. & Heyer, W. D. Human BRCA2 protein promotes RAD51 filament formation on RPA covered single-stranded DNA. *Nat. Struct. Mol. Biol.* **17**, 1260–1262 (2012).
180. Thorslund, T. *et al.* The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. *Nat. Struct. Mol. Biol.* **10**, 1263–1265 (2010).
181. Yu, D. S. *et al.* Dynamic control of Rad51 recombinase by self-association and interaction with BRCA2. *Mol. Cell* **4**, 1029–1041 (2003).
182. Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science (80-.).* 1837–1848 (2002).
183. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & M., J. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol. Cell Biol. 4th Ed.* 9305–9316 (2004).
184. Feng, Z. *et al.* Rad52 inactivation is synthetically lethal with BRCA2 deficiency. *PNAS* **108**, 686–691 (2010).
185. Bhat, K. P. & Cortez, D. RPA and RAD51: fork reversal, fork protection, and genome stability. *Nat. Struct. Mol. Biol.* **25**, 446–453 (2018).
186. Awate, S. & Brosh Jr., R. M. Interactive Roles of DNA Helicases and Translocases with the Single-Stranded DNA Binding Protein RPA in Nucleic Acid Metabolism. *Int. J. Mol. Sci.* **18**, 1–25 (2017).
187. Dodson, G. E., Shi, Y. & Tibbetts, R. S. DNA Replication Defects, Spontaneous DNA Damage, and ATM-dependent Checkpoint Activation in Replication Protein A-deficient Cells. *J. Biol. Chem.* 34010–34014 (2004).
188. Araya, R., Hirai, I., Meyerkord, C. L. & Wang, H.-G. Loss of RPA1 induces Chk2 phosphorylation through a caffeine-sensitive pathway. *FEBS Lett.* **579**, 157–161 (2005).
189. Recolin, B., Van der Laan, S. & Maiorano, D. Role of replication protein A as sensor in activation of the S-phase checkpoint in Xenopus egg extracts. *Nucleic Acid Res.* **40**, 3431–3442 (2012).
190. Haring, S. J., Mason, A. C., Binz, S. K. & Wold, M. S. Cellular Functions of Human

RPA1. *J. Biol. Chem.* **283**, 19095–19111 (2008).
191. Dou, H., Huang, C., Singh, M., Carpenter, P. B. & Yeh, E. T. Regulation of DNA repair through deSUMOylation and SUMOylation of replication protein A complex. *Mol. Cell* **39**, 333–345 (2010).
192. Dhingra, N., Wei, L. & Zhao, X. Replication protein A (RPA) sumoylation positively influences the DNA damage checkpoint response in yeast. *J. Biol. Chem.* **294**, 2690–2699 (2019).
193. Deng, S. K., Gibb, B., de Almeida, M. J., Greene, E. C. & Synington, L. S. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat Struct Mol Biol* 405–412 (2014).
194. Sugiyama, T. & Kowalczykowski, S. C. Rad52 Protein Associates with Replication Protein A (RPA)-Single-stranded DNA to Accelerate Rad51-mediated Displacement of RPA and Presynaptic Complex Formation. *J. Biol. Chem.* 31663–31672 (2002).
195. Pokhrel, N. *et al.* Monitoring Replication Protein A (RPA) dynamics in homologous recombination through site-specific incorporation of non-canonical amino acids. *Nucleic Acid Res.* **45**, 9413–9426 (2017).
196. Oakley, G. G. *et al.* Physical interaction between replication protein A (RPA) and MRN: involvement of RPA2 phosphorylation and the N-terminus of RPA1. *Biochemistry* **48**, 7477481 (2009).
197. Wu, X., Shell, S. M., Liu, Y. & Zou, Y. ATR-dependent checkpoint modulates XPA nuclear import in response to UV irradiation. *Oncogene* **5**, 757–764 (2007).
198. San-Segundo, P. A. & Clemente-Blanco, A. Resolvases, Dissolvases, and Helicases in Homologous Recombination: Clearing the Road for Chromosome Segregation. *Genes (Basel).* **11**, 1–22 (2020).
199. Mackintosh, S. G. & Raney, K. D. DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic Acid Res.* **34**, 4154–4159 (2006).
200. Brosh, R. M. J. & Matson, S. W. History of DNA Helicases. *Genes (Basel).* **22**, 1–45 (2020).
201. Raney, K. D. Chemical modifications of DNA for study of helicase mechanisms. *Bioorganic Med. Chem.* **22**, 4399–4406 (2014).
202. Matson, S. W. & George, J. W. DNA helicase II of Escherichia coli. Characterisation of the single-stranded DNA- dependent NTPase and helicase activities. *J. Biol. Chem.* **5**, 2033–2079 (1987).
203. Fairman-Williams, M. E., Guenther, U.-P. & Jankowsky, E. SF1 and SF2 helicases:family matters. *Curr. Opin. Struct. Biol.* **20**, 313–324 (2010).
204. Soultanas, P. & Wigley, D. B. Unwinding the 'Gordian knot' of helicase action. *Trends Biochem. Sci.* **26**, 47–54 (2001).
205. Manosas, M., Xi, X. G., Bensimon, D. & Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Res.* **38**, 5518–5526 (2010).
206. Lohman, T. M. Helicase-catalyzed DNA unwinding. *J. Biol. Chem.* **268**, 2269–2272 (1993).
207. Tuteja, N. & Tuteja, R. Unraveling DNA helicases Motif, structure, mechanism and function. *Eur. J. Biochem.* **271**, 1849–1863 (2004).
208. Chaban, Y. *et al.* Structural basis for DNA strand separation by a hexameric replicative helicase. *Nucleic Acids Res.* **43**, 8551–8563 (2015).
209. Kitano, K. Structural mechanisms of human RecQ helicases WRN and BLM. *Front. Genet.* **5**, 1–11 (2014).
210. Byrd, A. K. & Raney, K. D. Superfamily2 helicases. *Front. Biosci. (Landmark Ed.* **17**, 2070–2088 (2013).
211. Richards, J. *et al.* Strucutre of the DNA Repair Helicase Hel308 reveals DNA binding and autoinhibitory domains. *J. Biol. Chem.* **282**, 5118–5126 (2008).
212. Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **1**, 41–72 (2002).
213. Bernstein, D. A. & Keck, J. L. Domain mapping of Escherichia coli RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family. *Nucleic Acid Res.* 2778–2785 (2003).
214. Wu, Y. Unwinding and rewinding: Double faces of helicase? *J. Nucleic Acids* **2012**, (2012).
215. Lloyd, R. G., Benson, F. E. & Shurvinton, C. E. ffect of ruv mutations on recombination and DNA repair in Escherichia coli K12. *Mol. Genet. Genomics* 303–309 (1984).

216. Cox, M. M. *et al.* The importance of repairing stalled replication forks. *Nature* **404**, 37–41 (2000).
217. Brüning, J. G., Howard, J. L. & McGlynn, P. Accessory replicative helicases and the replication of protein-bound DNA. *J. Mol. Biol.* **426**, 3917–3928 (2015).
218. McGlynn, P., Savery, N. J. & Dillingham, M. S. The conflict between DNA replication and transcription. *Mol. Microbiol.* **85**, 12–20 (2012).
219. Yancey-wrona, J. E. & Matson, S. W. Bound lac repressor protein differentially inhibits the unwinding reactions catalyzed by DNA helicases. *Nucleic Acids Res.* **20**, 6713–6721 (1992).
220. Wang, S. W., Goodwin, A., Hickson, I. D. & Norbury, C. J. Involvement of Schizosaccharomyces pombe Srs2 in cellular responses to DNA damage. *Nucleic Acid Res.* 2963–2972 (2001).
221. Park, J. *et al.* PcrA helicase dismantles RecA filaments by reeling in DNA in uniform steps. *Cell* **142**, 544–555 (2010).
222. Fagerburg, M. V. *et al.* PcrA-mediated disruption of RecA nucleoprotein filaments—essential role of the ATPase activity of RecA. *Nucleic Acid Res.* 8416–8424 (2012).
223. Crickard, J. B. & Greene, E. C. Helicase Mechanisms During Homologous Recombination in Saccharomyces cerevisiae. *Annu. Reivew Biophys.* 255–275 (2019).
224. Tsutsui, Y., Kurokawa, Y., Ito, K., Siddique, S. P. & Kawano, Y. Multiple Regulation of Rad51-Mediated Homologous Recombination by Fission Yeast Fbh1. *PLOS Genet.* **10**, 1–15 (2014).
225. Croteau, D. L., Popuri, V., Opresko, P. L. & Bohr, V. A. Human RecQ helicases in DNA repair, recombination, and replication. *Annu. Rev. Biochem.* 519–552 (2014).
226. Larsen, N. B. & Hickson, I. D. RecQ helciases: Conserved guardians of genomic integrity. *Adv. Exp. Med. Biol.* 161–184 (2013).
227. Wu, L., Davies, S. L. & Hickson, I. D. Roles of RecQ family helicases in the maintenance of genome stability. *Cold Spring Harb. Symp. Quant. Biol.* 573–581 (2000).
228. Wu, Y. & Brosh, R. M. J. Distinct roles of RECQ1 in the maintenance of genomic stability. *DNA Repair (Amst).* 315–324 (2010).
229. Daley, J. M., Niu, H. & Sung, P. Roles of DNA Helicases in the Mediation and Regulation of Homologous Recombination. *Adv. Exp. Med. Biol.* 185–202 (2014).
230. Uniprot.org. Q13472 (TOP3A_HUMAN). Available at: http://www.uniprot.org/uniprot/Q13472. (Accessed: 30th June 2017)
231. Nguyen, M. M. & Livingston, D. M. The effect of a suppressed rad52 mutation on the suppression of rad6 by srs2. *Yeast* 1059–1064 (1997).
232. Krejci, L. *et al.* DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature* 305–309 (2003).
233. Antony, E. *et al.* Srs2 disassembles Rad51 filaments by a protein-protein interaction triggering ATP turnover and dissociation of Rad51 from DNA. *Mol. Cell* 105–115 (2009).
234. Gari, K., Decaillet, C., Delannoy, M., Wu, L. & Constantinou, A. Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc. Natl. Acad. Sci. U. S. A.* 16107–16112 (2008).
235. Gari, K., Decaillet, C., Stasiak, A. Z., Stasiak, A. & A., C. The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol. Cell* 141–148 (2008).
236. Takata, K., Reh, S., Tomida, J., Person, M. D. & Wood, R. D. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat Commun* **4**, 2338 (2013).
237. Adelman, C. A. *et al.* HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis. *Nature* **502**, 381–384 (2013).
238. Hu, Y. *et al.* RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev.* 3073–3084 (2007).
239. Andrs, M., Hasanova, Z., Oravetzova, A., Dobrovolna, J. & Janscak, P. RECQ5 : A Mysterious Helicase at the Interface of DNA Replication and Transcription. *Genes (Basel).* **11**, (2020).
240. Uchiumi, F., Seki, M. & Furuichi, Y. Helicases and human diseases. *Front. Genet.* **6**, (2015).
241. Bolt, E. L. Helicases that interact with replication forks: New candidates from archaea. *Biochem. Soc. Trans.* **33**, 1471–1473 (2005).

242. Guy, C. P. & Bolt, E. L. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res.* **33**, 3678–3690 (2005).
243. Marini, F. & Wood, R. D. A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. *J. Biol. Chem.* **277**, 8716–8723 (2002).
244. Büttner, K., Nehring, S. & Hopfner, K. P. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652 (2007).
245. Oyama, T. *et al.* Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm. *BMC Struct. Biol.* **9**, 1–12 (2009).
246. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acid Res.* **28**, 235–242 (2020).
247. Northall, S. J. *et al.* DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst).* **57**, 125–132 (2017).
248. Wood, R. D. & Doublie, S. DNA polymerase theta (POLQ), double-strand break repair, and cancer. *DNA Repair (Amst).* **44**, 22–32 (2016).
249. Newman, J. A., Cooper, C. D. O., Aitkenhead, H. & Gileadi, O. Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Strucutre* **23**, 2319–2330 (2015).
250. Wang, Z. *et al.* DNA Polymerase theta (POLQ) is important for repair of DNA double-strand breaks caused by fork collapse. *J. Biol. Chem.* **294**, (2019).
251. Mateos-Gomez, P. A. *et al.* The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nat. Struct. Mol. Biol.* **24**, 1116–1123 (2017).
252. Northall, S. J. Molecular Analysis of Human and Archaeal DNA Repair Helicases HelQ and Hel308 A thesis submitted to the University of Nottingham for the degree Doctor of Philosophy. (2017).
253. Ward, J. D. *et al.* Overlapping Mechanisms Promote Postsynaptic RAD-51 Filament Disassembly during Meiotic Double-Strand Break Repair. *Mol. Cell* **37**, 259–272 (2010).
254. Muzzini, D. M., Plevani, P., Boulton, S. J., Cassata, G. & Marini, F. Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. *DNA Repair (Amst).* **7**, 941–950 (2008).
255. Luebben, S. W. *et al.* Helq acts in parallel to Fancc to suppress replication-associated genome instability. *Nucleic Acids Res.* **41**, 10283–10297 (2013).
256. Tafel, A. A., Wu, L. & McHugh, P. J. Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *J. Biol. Chem.* **286**, 15832–15840 (2011).
257. Amendola, P. G. *et al.* JMJD-5/KDM8 regulates H3K36me2 and is required for late steps of homologous recombination and genome integrity. *PLoS Genet.* **13**, 1–24 (2017).
258. Cordell, H. J. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**, 2463–2468 (2002).
259. Wilkins, S. E. *et al.* JMJD5 is a human arginyl C-3 hydroxylase. *Nat. Commun.* **9**, (2018).
260. Pelttari, L. M., Kinnunen, L., Kiiski, J. I. & Khan, S. Screening of HELQ in breast and ovarian cancer families. *Fam. Cancer* **15**, 19–23 (2016).
261. Li, W. *et al.* Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. *Carcinogenesis* **34**, 1536–1542 (2013).
262. Long, J. *et al.* Gynecologic Oncology Helicase POLQ-like (HELQ) as a novel indicator of platinum-based chemoresistance for epithelial ovarian cancer. *Gynecol. Oncol.* **149**, 341–349 (2018).
263. Guy, C. P. & Bolt, E. L. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res.* **33**, 3678–3690 (2005).
264. Northall, S. J., Jones, N., Soultanas, P. & Bolt, E. L. DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst).* **57**, 125–132 (2017).
265. Marini, F. & Wood, R. D. A Human DNA Helicase Homologous to the DNA Cross-link Sensitivity Protein Mus308 *. *J. Biol. Chem.* **277**, 8716–8723 (2002).
266. NEW ENGLAND BioLAbs Inc. NEBioCalculator. *Version 1.11.0* (2020). Available at: https://nebiocalculator.neb.com/#!/ligation. (Accessed: 27th August 2020)
267. Manns, J. M. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins. *Curr. Protoc. Microbiol.* **22**, (2011).
268. Stott, D. I. Immunoblotting and dot blotting. *J. Immunol. Methods* **119**, 153–187 (1989).
269. Krainer, F. W. & Glieder, A. An updated view on horseradish peroxidases: recombinant production and biotechnological applications. *Appl. Microbiol. Biotechnol.* **99**, 1611–1625 (2015).

270. Chevallet, M., Luche, S. & Rabilloud, T. Silver staining of proteins in polyacrylamide gels. *Nat. Protoc.* **1**, 1852–1858 (2006).
271. Chen, Y. P., Gundersen-Rindal, D. E. & Lynn, D. E. Baculovirus-based expression of an insect viral protein in 12 different insect cell lines. *Vitr. Cell. Dev. Biol. - Anim.* **41**, 43–49 (2005).
272. Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. The establishment of two cell lines from the insectspodoptera frugiperda (lepidoptera; noctuidae). *In Vitro* **13**, 213–217 (1977).
273. O'Reilly, D. R., Brown, M. R. & Miller, L. K. Alteration of ecdysteroid metabolism due to baculovirus infection of the fall armyworm Spodoptera frugiperda: Host ecdysteroids are conjugated with galactose. *Insect Biochem. Mol. Biol.* **22**, 313–320 (1992).
274. Granados, R. R., Guoxun, L., Derksen, A. C. G. & McKenna, K. A. A new insect cell line from Trichoplusia ni (BTI-Tn-5B1-4) susceptible to Trichoplusia ni single enveloped nuclear polyhedrosis virus. *J. Invertabrate Pathol.* **64**, 260–266 (1994).
275. Davis, T. R. *et al.* Comparative recombinant protein production of eight insect cell lines. *Vitr. Cell Dev Biol-Animal* **29**, 388–390 (1993).
276. Binz, S. K., Dickson, A. M., Haring, S. J. & Wold, M. S. Functional Assays for Replication Protein A (RPA). *Methods Enzymol.* **409**, 11–38 (2006).
277. Henricksens, L. A., Umbrichtn, C. B. & Wold, S. M. Recombinant Replication Protein A Expression, Complex Formation and Functional Characterization. *J. Biol. Chem.* **269**, 11121–11132 (1994).
278. Bio-Rad. Affi-Gel Blue Gel Instruction Manual Catalog Numbers 153-7302 153-7301. *Bio-Rad Lab.* 18 (2000).
279. Wilson, J. E. Applications of blue dextran and cibacron blue F3GA in purification and structural studies of nucleotide-requiring enzymes. *Biochem. Biophys. Res. Commun.* **72**, 816–823 (1976).
280. GraphPad Software, I. PRISM GraphPad. *La Jolla California USA* Available at: www.graphpad.com.
281. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server; in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005).
282. Gasteiger, E. *et al.* ExPASy: the proteomics server for indepth protein knowledge and analysis. *Nucleic Acid Res.* **31**, 3784–3788 (2003).
283. Abcam. Fluorophore table. 325 (2014).
284. Idtdna. OligoAnalyzer Tool. (2020). Available at: https://www.idtdna.com/calc/analyzer. (Accessed: 23rd April 2020)
285. Guerra, C. E. Analysis of oligonucleotide microarrays by 3 ′ end labeling using fluorescent nucleotides and terminal transferase Short Technical Reports. *Boitechniques* **41**, 53–56 (2006).
286. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *BMC Bioinformatics* **529**, (2017).
287. Lazer, I. J. & Horvath-Lazar, E. GelAnalyzer 2010a. (2010). Available at: http://www.gelanalyzer.com/index.html.
288. Hellman, L. M. & Fried, M. G. Electrophoretic Mobility Shift Assay (EMSA) for Detecting Protein- Nucleic Acid Interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).
289. Nadeau, K., Das, A. & Walsh, C. T. Hsp90 chaperonins possess ATPase activity and bind heat shock transcription factors and peptidyl prolyl isomerases. *J. Biol. Chem.* **268**, 1479–1487 (1993).
290. Bruning, J. G., Howard, J. A. L. & McGlynn, P. Use of streptavidin bound to biotinylated DNA structures as model substrates for analysis of nucleoprotein complex disruption by helicases. *ELSEVIER Methods* **108**, 48–55 (2016).
291. StrataGene. BacterioMatch ® II Two-Hybrid System Library Construction Kit. *StrataGene*
292. Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein- protein contacts. *Nature* **386**, 627–630 (1997).
293. Dove, S. L. & Hochschild, A. Conversion of the ω subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes Dev.* **12**, 745–754 (1998).
294. Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7382–7387 (2000).

295. Greenfield, N. J. Analysis of Circular Dichroism Data. *Methods Enzymol.* **383**, 282–317 (2004).
296. Greenfield, N. J. Using circular dichroism spectra to estmate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2006).
297. Gore, M. G. Chapter 4: Introduction to Circular Dichroism by Rogers and Ismail. in *Spectrophotometry and Spectrofluorimetry: A Practiccal Approach* (ed. Gore, M. G.) 99–139 (Oxford University Press, 2000).
298. Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **84**, 143–179 (2008).
299. Saponaro, A. *et al.* Isothermal Titration Calorimetry: A Biophysical Method to Characterize the Interaction between Label-free Biomolecules in Solution. *Bio-protocol* **8**, 2570–2580 (2018).
300. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).
301. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acid Res.* **47**, D506–D515 (2019).
302. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol* **7**, (2011).
303. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
304. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
305. Källberg, M. *et al.* Template-based protein structure modelling using the RaptorX web server. *Nat. Protoc.* **7**, 1511–1522 (2012).
306. Schrödinger, L. The PyMOL Molecular Graphics System, Version 2.0.
307. Mukherjee, S. & Zhang, Y. Protein-protein complex structure prediction by multimeric threading and template recombination. *Structure* **19**, 955–966 (2011).
308. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, 363–367 (2005).
309. Richards, J. D. *et al.* Structure of the DNA Repair Helicase Hel308 Reveals DNA Binding and Autoinhibitory Domains *. **283**, 5118–5126 (2008).
310. Buttner, K., Nehring, S. & Hopfner, K. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652 (2007).
311. Betts, H. A protein dissection of the human helicase HELQ. 50 (2019).
312. Ramanagoudr-bhojappa, R. *et al.* Yeast Pif1 Helicase Exhibits a One-base-pair Stepping Mechanism for Unwinding Duplex DNA. *J. Biol. Chem.* **288**, 16185–16195 (2013).
313. Taylor, A. F. & Smith, G. R. Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.* **270**, 24451–24458 (1995).
314. Yang, Y. *et al.* Evidence for a functional dimeric form of the PcrA helicase in DNA unwinding. *Nucleic Acids Res.* **36**, 1976–1989 (2008).
315. Maluf, N. K., Fischer, C. J. & Lohman, T. M. A dimer of Escherichia coli UvrD is the active form of the helicase in vitro. *J. Mol. Biol.* **325**, 913–935 (2003).
316. Walmacq, C., Rahmouni, A. R. & Boudvillain, M. Testing the steric exclusion model for hexameric helicases: Substrate features that alter RNA-DNA unwinding by the transcription termination factor rho. *Biochemistry* **45**, 5885–5895 (2006).
317. Graham, B. W., Schauer, G. D., Leuba, S. H. & Trakselis, M. A. Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acid Res.* **39**, 6585–6595 (2018).
318. Kaplan, D. L., Davey, M. J. & Donnell, M. O. Mcm4, 6, 7 uses a “Pump in Ring” Mechanism to Unwind DNA by Steric Exclusion and Actively Translocate along a Duplex. *J. Biol. Chem.* **278**, 49171–49182 (2003).
319. Amartely, H., Avraham, O., Friedler, A., Livnah, O. & Lebendiker, M. Coupling Multi Angle Light Scattering to Ion Exchange chromatography (IEX-MALS) for protein characterization. *Sci. Rep.* **8**, 1–9 (2018).
320. Theisen, A., Johann, C., Deacon, M. P. & Harding, S. E. *Refractive increment data-book for polymer and biomoleclar scientists*. (Nottingham: Nottingham University Press, 2000).
321. Ekman, D., Björklund, Å. K., Frey-Skött, J. & Elofsson, A. Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions. *J. Mol. Biol.* **348**,

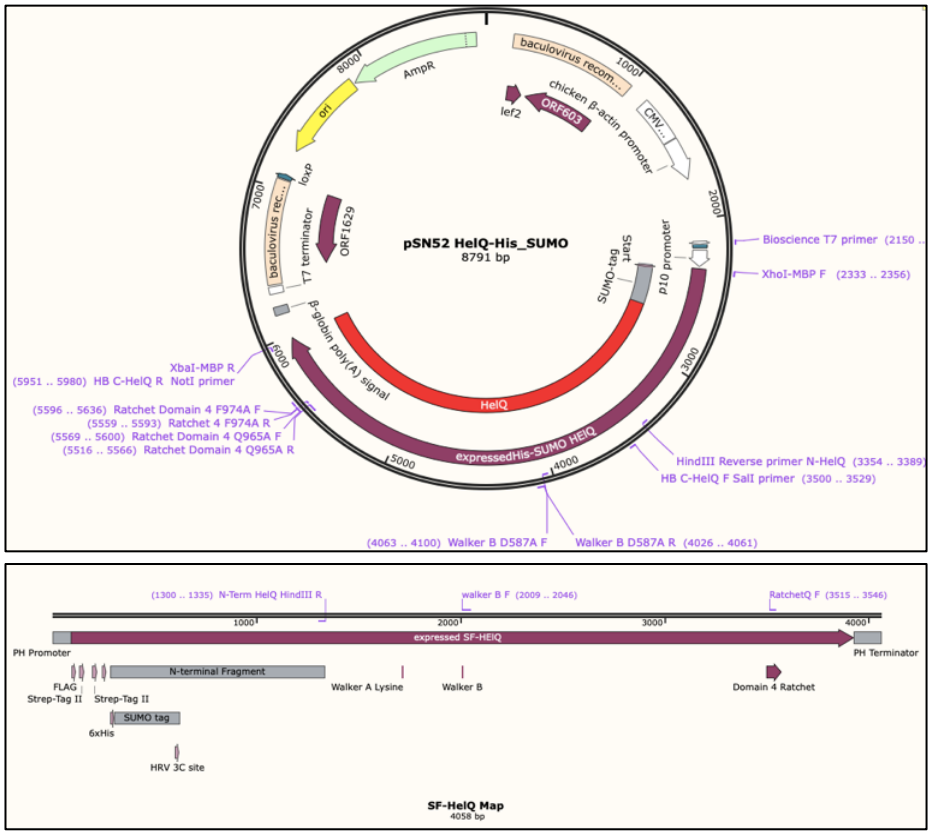
231–243 (2005).
322. Uversky, V. N. Intrinsically disordered proteins and their ‘Mysterious’ (meta)physics. *Front. Phys.* **7**, 8–23 (2019).
323. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
324. Darling, A. L. & Uversky, V. N. Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Front. Genet.* **9**, 1–18 (2018).
325. Sreerama, N. & Woody, R. W. Computation and Analysis of Protein Circular Dichroism Spectra. *Methods Enzymol.* **383**, 318–351 (2004).
326. Bayer, P. *et al.* Structure determination of the small ubiquitin-related modifier SUMO-1. *J. Mol. Biol.* **280**, 275–286 (1998).
327. Kosol, S., Contreras-Martos, S., Cedeño, C. & Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802–10828 (2013).
328. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. Geometry based flexible and symmetric protein docking. *Proteins* **60**, 224–231 (2005).
329. Absmeier, E. *et al.* A noncanonical PWI domain in the N-terminal helicase-associated region of the spliceosomal Brr2 protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 762–771 (2015).
330. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).
331. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281**, 16117–16127 (2006).
332. Instruments, M. Inform White Paper Dynamic Light Scattering. *Malvern Guid.* 1–6 (2011).
333. Arzenšek, D., Podgornik, R. & Kuzman, D. Dynamic light scattering and application to proteins in solutions. *Seminar* 1–18 (2010).
334. Cui, S. *et al.* Analysis of the unwinding activity of the dimeric RECQ1 helicase in the presence of human replication protein A. *Nucleic Acids Res.* **32**, 2158–2170 (2004).
335. Some, D., Amartely, H., Tsadok, A. & Lebendiker, M. Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (Sec-mals). *J. Vis. Exp.* **2019**, 1–9 (2019).
336. Uversky, V. N. Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. in *Intrinsically Disordered Protein Analysis. Methods in Molecular Biology (Methods and Protocols)* (ed. Dunker, A. (eds.) (Springer, New York, NY, 2012).
337. Lu, Y. Assembly and transfer of iron–sulfur clusters in the plastid. *Front. Plant Sci.* **9**, 1–17 (2018).
338. White, M. F. & Dillingham, M. S. Iron-sulphur clusters in nucleic acid processing enzymes. *Curr. Opin. Struct. Biol.* **22**, 94–100 (2012).
339. Opresko, P. L., Cheng, W. H. & Bohr, V. A. Junction of RecQ Helicase Biochemistry and Human Disease. *J. Biol. Chem.* **279**, 18099–18102 (2004).
340. Patel, S. S. & Donmez, I. Mechanisms of helicases. *J. Biol. Chem.* **281**, 18265–18268 (2006).
341. Manthei, K. A., Hill, M. C., Burke, J. E., Butcher, S. E. & Keck, J. L. Structural mechanisms of DNA binding and unwinding in bacterial RecQ helicases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4292–4297 (2015).
342. Harmon, F. G. & Kowalczykowski, S. C. Biochemical Characterization of the DNA Helicase Activity of the Escherichia coli RecQ Helicase. *J. Biol. Chem.* **276**, 232–243 (2001).
343. Moye, A. L. *et al.* Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **6**, (2015).
344. Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
345. Punatar, R. S., Martin, M. J., Wyatt, H. D. M., Chan, Y. W. & West, S. C. Resolution of single and double Holliday junction recombination intermediates by GEN 1. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 443–450 (2017).
346. Van Gool, A. J., Hajibagheri, N. M. A., Stasiak, A. & West, S. C. Assembly of the Escherichia coli RuvABC resolvasome directs the orientation of Holliday junction resolution. *Genes Dev.* **13**, 1861–1870 (1999).
347. Górecka, K. M. *et al.* RuvC uses dynamic probing of the Holliday junction to achieve

sequence specificity and efficient resolution. *Nat. Commun.* **10**, (2019).
348. McGlynn, P., Lloyd, R. G. & Marians, K. J. Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA is negatively supercoiled. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8235–8240 (2001).
349. Wardrope, L., Okely, E. & Leach, D. Resolution of joint molecules by RuvABC and RecG following cleavage of the Escherichia coli chromosome by EcoKI. *PLoS One* **4**, (2009).
350. Black, S. J., Kashkina, E., Kent, T. & Pomerantz, R. T. DNA polymerase θ: A unique multifunctional end-joining machine. *Genes (Basel).* **7**, (2016).
351. Shamanna, R. A. *et al.* RECQ helicase RECQL4 participates in non-homologous end joining and interacts with the Ku complex. *Carcinogenesis* **35**, 2415–2424 (2014).
352. Muzzolini, L. *et al.* Different quaternary structures of human RECQ1 are associated with its dual enzymatic activity. *PLoS Biol.* **5**, 0157–0168 (2007).
353. Cordin, O., Beggs, J. D., Cordin, O. & Beggs, J. D. RNA helicases in splicing RNA helicases in splicing. **6286**, 83–95 (2016).
354. Anand, R. P. *et al.* Overcoming natural replication barriers: Differential helicase requirements. *Nucleic Acids Res.* **40**, 1091–1105 (2012).
355. Mirkin, E. V. & Mirkin, S. M. Replication Fork Stalling at Natural Impediments. *Microbiol. Mol. Biol. Rev.* **71**, 13–35 (2007).
356. Meng, X. & Zhao, X. Replication fork regression and its regulation. *FEMS Yeast Res.* **17**, 2017 (2017).
357. Yardimci, H. *et al.* Bypass of a protein roadblock by a replicative DNA helicase. *Nature* **492**, 205–209 (2013).
358. Fu, Y. V. *et al.* Selective bypass of a lagging strand roadblock by the eukaryotic replicative DNA helicase. *Cell* **146**, 931–941 (2011).
359. Newman, M., Strzelecka, T., Dorner, L., Schildkraut, I. & Aggarwal, A. Structure of restriction endonuclease BamHI phased at 1.95 å resolution by MAD analysis. *Structure* **2**, 439–452 (1994).
360. Pingoud, A. & Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727 (2001).
361. Hendrickson, W. A. *et al.* Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2190–2194 (1989).
362. Lynch, T. W. & Sligar, S. G. Macromolecular hydration changes associated with BamHI binding and catalysis. *J. Biol. Chem.* **275**, 30561–30565 (2000).
363. Shin, J. & Kelman, Z. DNA unwinding assay using streptavidin-bound oligonucleotides. *BMC Mol. Biol.* **7**, 1–8 (2006).
364. Hawkins, M. *et al.* Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. *Nucleic Acids Res.* **47**, 5100–5113 (2019).
365. Hafiz Ishfaq Ahmad, al *et al.* CRISPR-Based Genome Editing. 47–68 (2018). doi:10.21775/cimb.028.047
366. Clarke, R. *et al.* Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. *Mol. Cell* **71**, 42-55.e8 (2018).
367. Richardson, C. D. *et al.* CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50**, 1132–1139 (2018).
368. Vincent, S. D., Mahdi, A. A. & Lloyd, R. G. The RecG Branch Migration Protein of E.coli Dissociates R-loops. *J. Mol. Biol.* **264**, 713–721 (1996).
369. Simmons, E. Genetic and BIochemical Analysis of the Archaeal Helicase Hel308. *University of Nottingham* (2018).
370. Banroques, J., Cordin, O., Doère, M., Linder, P. & Tanner, N. K. A Conserved Phenylalanine of Motif IV in Superfamily 2 Helicases Is Required for Cooperative, ATP-Dependent Binding of RNA Substrates in DEAD-Box Proteins. *Mol. Cell. Biol.* **28**, 3359–3371 (2008).
371. Khan, I., Sommers, J. A. & Brosh, R. M. Close encounters for the first time: Helicase interactions with DNA damage. *DNA Repair (Amst).* **33**, 43–59 (2015).
372. Tackett, A. J., Wei, L., Cameron, C. E. & Raney, K. D. Unwinding of nucleic acids by HCV NS3 helicase is sensitive to the structure of the duplex. *Nucleic Acids Res.* **29**, 565–572 (2001).
373. Meagher, M., Epling, L. B. & Enemark, E. J. DNA translocation mechanism of the MCM

complex and implications for replication initiation. *Nat. Commun.* **10**, (2019).
374. Sommers, J. A. *et al.* Novel Function of the Fanconi Anemia Group J or RECQ1 Helicase to Disrupt Protein-DNA Complexes in a Replication Protein A-stimulated Manner *. *J. Biol. Chem.* **289**, 19928–19941 (2014).
375. Rhodes, D. & Lipps, H. J. Survey and summary G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).
376. Pike, A. C. W. *et al.* Human RECQ1 helicase-driven DNA unwinding, annealing, and branch migration: Insights from DNA complex structures. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4286–4291 (2015).
377. Meiners, M. J., Tahmaseb, K. & Matson, S. W. The UvrD303 hyper-helicase exhibits increased processivity. *J. Biol. Chem.* **289**, 17100–17110 (2014).
378. Zhou, H.-X. & Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **118**, 1691–1741 (2018).
379. Northall, S. J. *et al.* DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst).* **57**, 125–132 (2017).
380. Fanning, E., Kilmovich, V. & Nager, A. . A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acid Res.* **34**, 4126–4137 (2006).
381. Lever, R. *et al.* In preparation: Hyper-active Hel308 helicase activity causes a hyper-recombination phenotype — insights into the cellular function of Hel308 in recombination and DNA repair. *Prep.* (2020).
382. Gwynn, E. J. *et al.* The Conserved C-Terminus of the PcrA/UvrD Helicase Interacts Directly with RNA Polymerase. *PLoS One* **8**, 1–11 (2013).
383. Woodman, I. L. & Bolt, E. L. Winged helix domains with unknown function in Hel308 and related helicases. *Biochem. Soc. Trans.* **39**, 140–144 (2011).
384. Lohman, T. M., Tomko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: Mechanisms and regulation. *Nat. Rev. Mol. Cell Biol.* **9**, 391–401 (2008).
385. Li, L. *et al.* DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Mol. Cell. Biol.* **36**, 2794–2810 (2016).
386. Doherty, K. M. *et al.* Physical and functional mapping of the replication protein A interaction domain of the Werner and Bloom syndrome helicases. *J. Biol. Chem.* **280**, 29494–29505 (2005).
387. Lee, M. *et al.* Multiple RPAs make WRN syndrome protein a superhelicase. *Nucleic Acids Res.* **46**, 4689–4698 (2018).
388. Gibson, W. The origins of the trypanosome genome strains Trypanosoma brucei brucei TREU 927, T. b. gambiense DAL 972, T. vivax Y486 and T. congolense IL3000. *Parasites and Vectors* **5**, 71 (2012).
389. Rowe, A. J. Ultra-weak reversible protein-protein interactions. *Methods* **54**, 157–166 (2011).
390. Keller, H. *et al.* The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. *Nucleic Acids Res.* **42**, 12614–12627 (2014).

Appendix 7.1 StrepII-hexaHis-SUMO-HelQ amino acid FASTA sequence and plasmid map.

*MDYKDDDDKGSAASWSHPQFEKGSAGSAAGSAAGSGGAGWSHPQFEKSDYDIPTTENLYFQG AGTLEEAHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEA FAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQISSGLEVLFQGP*DECGSRIR RRVSLPKRNRPSLGCIFGAPTAAELEPGDEGKEEEEMVAENRRRKTAGVLPVEVQPLLLS DSPECLVLGGGDTNPDLLRHMPTDRGVGDQPNDSEVDMFGDYDSFTENSFIAQVDDLEQ KYMQLPEHKKHATDFATENLCSESIKNKLSITTIGNLTELQTDKHTENQSGYEGVTIEPGA DLLYDVPSSQAIYFENLQNSSNDLGDHSMKERDWKSSSHNTVNEELPHNCIEQPQQNDES SSKVRTSSDMNRRKSIKDHLKNAMTGNAKAQTPIFSRSKQLKDTLLSEEINVAKKTIESSS NDLGPFYSLPSKVRDLYAQFKGIEKLYEWQHTCLTLNSVQERKNLIYSLPTSGGKTLVAEI LMLQELLCCRKDVLMILPYVAIVQEKISGLSSFGIELGFFVEEYAGSKGRFPPTKRREKKSL YIATIEKGHSLVNSLIETGRIDSLGLVVVDELHMIGEGSRGATLEMTLAKILYTSKTTQIIG MSATLNNVEDLQKFLQAEYYTSQFRPVELKEYLKINDTIYEVDSKAENGMTFSRLLNYKY SDTLKKMDPDHLVALVTEVIPNYSCLVFCPSKKNCENVAEMICKFLSKEYLKHKEKEKCE VIKNLKNIGNGNLCPVLKRTIPFGVAYHHSGLTSDERKLLEEAYSTGVLCLFTCTSTLAAG VNLPARRVILRAPYVAKEFLKRNQYKQMIGRAGRAGIDTIGESILILQEKDKQQVLELITKP LENCYSHLVQEFTKGIQTLFLSLIGLKIATNLDDIYHFMNGTFFGVQQKVLLKEKSLWEIT VESLRYLTEKGLLQKDTIYKSEEEVQYNFHITKLGRASFKGTIDLAYCDILYRDLKKGLEG LVLESLLHLIYLTTPYDLVSQCNPDWMIYFRQFSQLSPAEQNVAAILGVSESFIGKKASGQ AIGKKVDKNVVNRLYLSFVLYTLLKETNIWTVSEKFNMPRGYIQNLLTGTASFSSCVLHF CEELEEFWVYRALLVELTKKLTYCVKAELIPLMEVTGVLEGRAKQLYSAGYKSLMHLAN ANPEVLVRTIDHLSRRQAKQIVSSAKMLLHEKAEALQEEVEELLRLPSDFPGAVASSTDK A



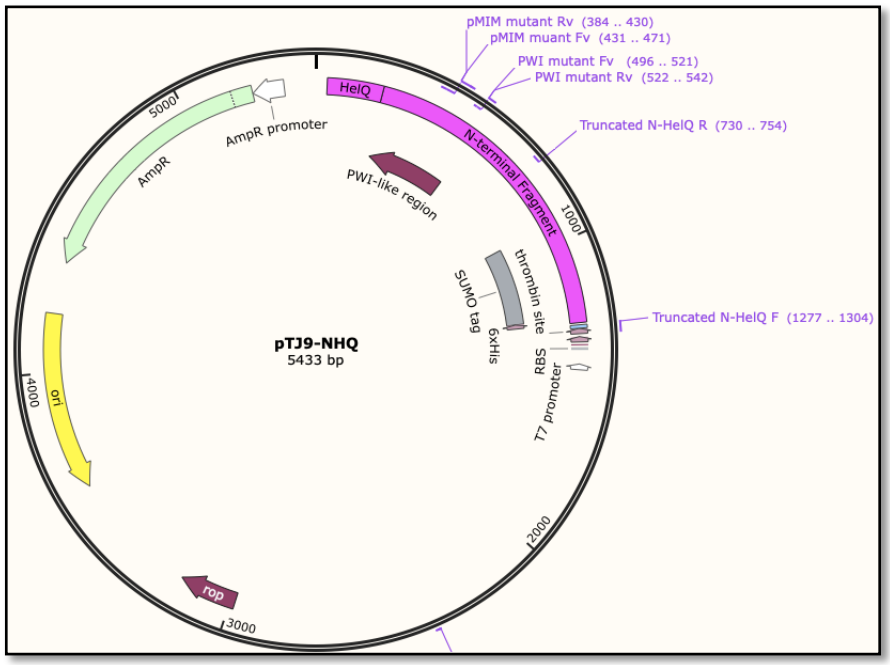
Plasmid maps for FL-HelQ. HelQ was cloned and expressed using baculovirus expression in insect cells. pSN52 FL-HelQ plasmid with an N-terminal hexaHis-SUMO tag. This vector was used for subcloning of all subsequent plasmids within this work. Primers used for HelQ mutagenesis and subcloning are identified. FL-HelQ was subcloned into the pACEBac1 vector with N-terminal Strep-II tags for purification.

Appendix 7.2 N-HelQ amino acid FASTA sequence and plasmid map with mass spectroscopy results.

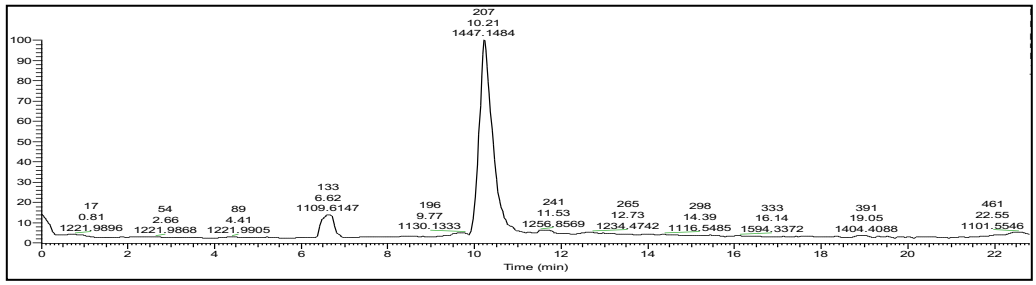
A.

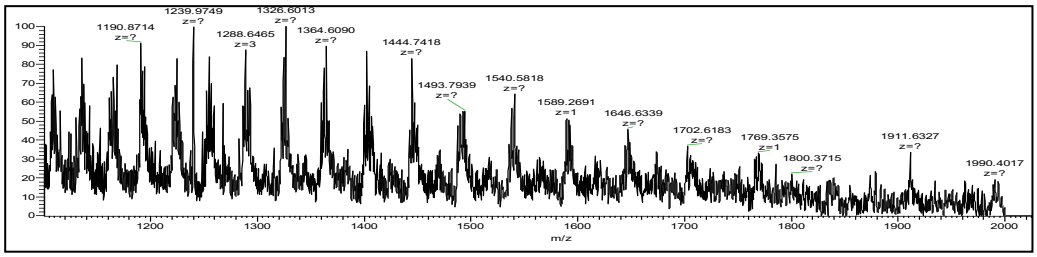
*MAHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKR QGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQISSGLEVLFQGPDECGSRIRRRVS LPKRNRPSLGCIFGAPTAAELEPGDEGKEEEEMVAENRRRKTAGVLPVEVQPLLLSDSPECLVL GGGDTNPDLLRHMPTDRGVGDQP*NDSEVDMFGDYDSFTENSFIAQVDDLEQKYMQLPEH KKHATDFATENLCSESIKNKLSITTIGNLTELQTDKHTENQSGYEGVTIEPGADLLYDVPSS QAIYFENLQNSSNDLGDHSMKERDWKSSSHNTVNEELPHNCI

B.

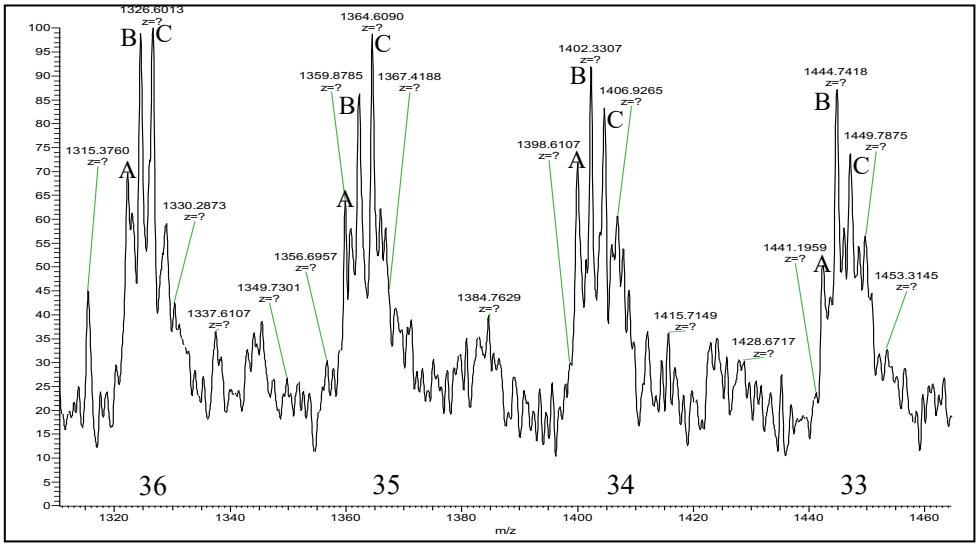


C.





D.



E.

Peak	m/z Spectra	Charge (+)	MW (Da)	Deconvoluted MW (Da)	Error
A	1322.3	36	47566.514	47564.11	2.281
	1259.9	35	47561.222		
	1399.9	34	47562.330		
	1442.4	33	47565.938		
B	1324.5	36	47645.714	47644.21	1.541
	1362.2	35	47641.722		
	1402.3	34	47643.930		
	1444.8	33	47645.138		
C	1326.6	36	47721.314	47722.930	2.004
	1364.5	35	47722.222		
	1404.7	34	47725.530		
	1492.4	33	47721.038		

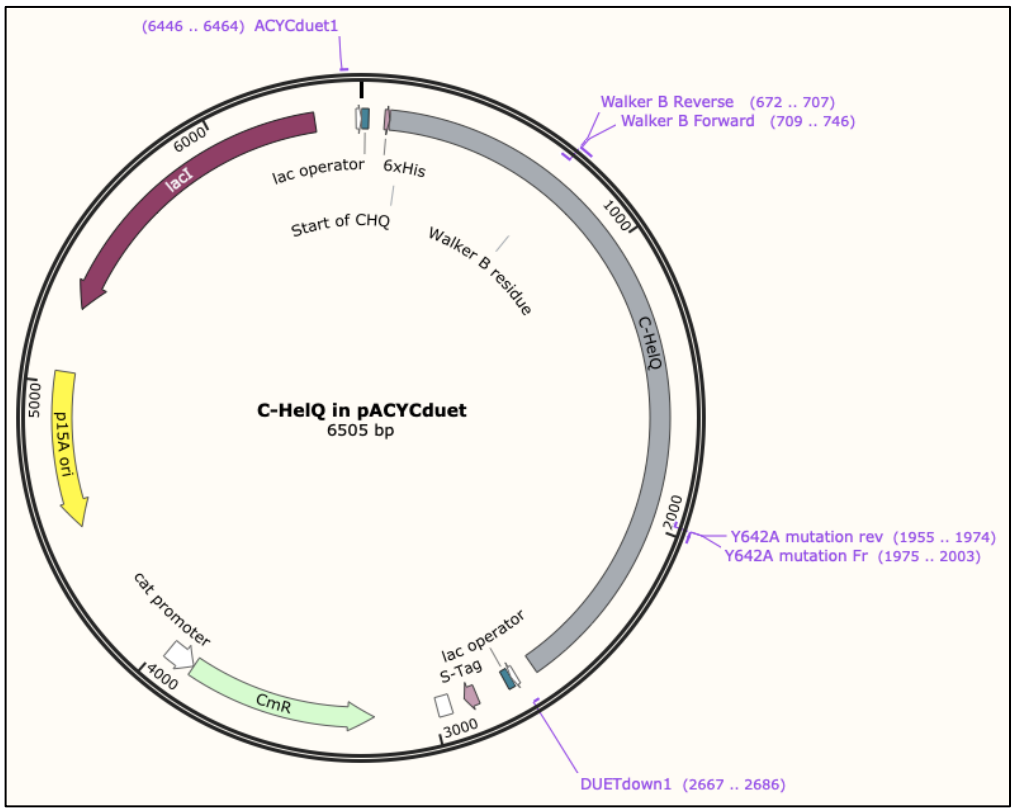
N-HelQ fragment identified from HelQ purification using mass spectroscopy. A. FASTA protein sequence of tagged N-HelQ**. B.** SNAP GENE file of pTJ9 N-HelQ *E. coli* expression plasmid. Extended over-expression for 72 hours of FL-HelQ in Sf9 insect cells resulted in the degradation of HelQ and the production of a stable N-terminal HelQ fragment. The fragment was isolated and analysed by mass spectroscopy at the University of Leicester Proteomics Facility. Mass spectrometry was carried out by trypsin digestion and LC-MS/MS analysis. **C**. Raw spectra (full spectra). **D.** Peak assignment and isolation of spectra in A. **E.** Peak analysis and calculations of individual and mean molecular weight data to give the deconvoluted molecular weight (MW) of 47.7 kDa.

Appendix 7.3 C-HelQ amino acid FASTA sequence and plasmid map.

A.

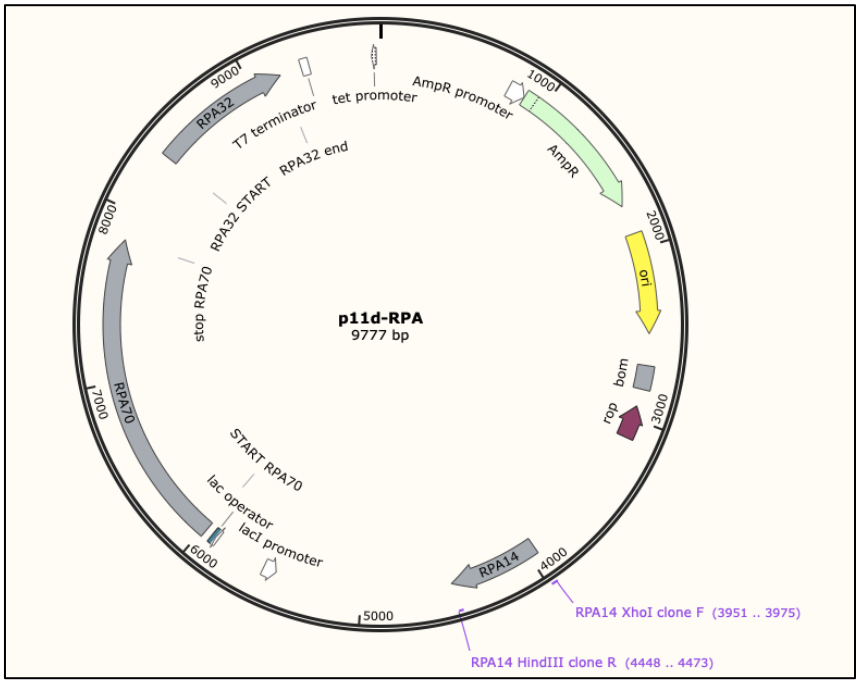
*MHHHHHH*NAKAQTPIFSRSKQLKDTLLSEEINVAKKTIESSSNDLGPFYSLPSKVRDLYAQ
FKGIEKLYEWQHTCLTLNSVQERKNLIYSLPTSGGKTLVAEILMLQELLCCRKDVLMILPY
VAIVQEKISGLSSFGIELGFFVEEYAGSKGRFPPTKRREKKSLYIATIEKGHSLVNSLIETGRI
DSLGLVVVDELHMIGEGSRGATLEMTLAKILYTSKTTQIIGMSATLNNVEDLQKFLQAEY
YTSQFRPVELKEYLKINDTIYEVDSKAENGMTFSRLLNYKYSDTLKKMDPDHLVALVTEV
IPNYSCLVFCPSKKNCENVAEMICKFLSKEYLKHKEKEKCEVIKNLKNIGNGNLCPVLKRT
IPFGVAYHHSGLTSDERKLLEEAYSTGVLCLFTCTSTLAAGVNLPARRVILRAPYVAKEFL
KRNQYKQMIGRAGRAGIDTIGESILILQEKDKQQVLELITKPLENCYSHLVQEFTKGIQTLF
LSLIGLKIATNLDDIYHFMNGTFFGVQQKVLLKEKSLWEITVESLRYLTEKGLLQKDTIYK
SEEEVQYNFHITKLGRASFKGTIDLAYCDILYRDLKKGLEGLVLESLLHLIYLTTPYDLVSQ
CNPDWMIYFRQFSQLSPAEQNVAAILGVSESFIGKKASGQAIGKKVDKNVVNRLYLSFVL
YTLLKETNIWTVSEKFNMPRGYIQNLLTGTASFSSCVLHFCEELEEFWVYRALLVELTKKL
TYCVKAELIPLMEVTGVLEGRAKQLYSAGYKSLMHLANANPEVLVRTIDHLSRRQAKQI
VSSAKMLLHEKAEALQEEVEELLRLPSDFPGAVASSTDKA

B.



Plasmid maps for C-HelQ. DNA encoding C-HelQ starting at position Asp-275 was cloned into pACYC-duet for expression from *E. coli* with an N-terminal hexa-His tag. A. FAST amino acid sequence inclusive of the tag resulting in a protein predicted at 96.3 kDa. B. SNAPGENE plasmid file of C-HelQ in pACYC-duet.

Appendix 7.4 p11d-tRPA plasmid map.



Plasmid map of p11d-tRPA to express trimeric RPA in *E. coli.* The map made from addgene (plasmid #102613) identifies the three protein subunits, RPA70, RPA32 and RPA14 that make up trimeric RPA. The plasmid was first described by Henricksen *et al.* 1994. This plasmid was used for protein over-expression of untagged RPA in *E. coli.* Due to its toxic nature it required tight regulation and control during over-expression and purification as described in chapter 2.