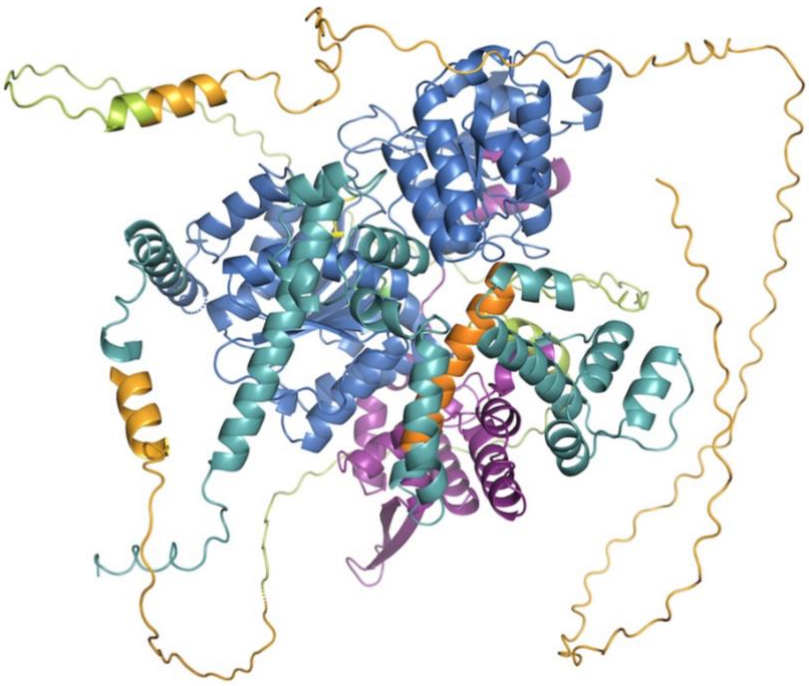
The Molecular Annealing Mechanisms of *H. sapiens* HelQ



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

Anna A. Lou-Hing, BSc (Hons)

September 2023

University of Nottingham
UK | CHINA | MALAYSIA

Abstract

The human genome is constantly under the threat of DNA damage as replicating cells are chemically exposed, resulting in double-strand breaks (DSBs). Humans evolved unique mechanisms for controlling these mutagenic effects. Homologous recombination (HR) repairs otherwise lethal DNA double-strand breaks. In mammals, multiple forms of HR have evolved for different contexts, including to underpin DNA replication so that genome duplication is completed before cell division. HR in this context is mutagenic, because it relies on unstable DNA synthesis by 'Break-Induced Replication' (BIR) within specialized HR DNA structures called D-loops. The extent of D-loop DNA synthesis can be restrained to the DNA break region to limit mutagenesis, but by mechanisms unknown in human cells. Genome instability is a hallmark of cancer, and if left untreated can be detrimental. Therefore, characterizing the role helicase proteins, namely HelQ, play in DNA repair and replication is of great importance.

Recent publications describe physical and functional interactions of HelQ and suggest possible mechanisms in which HelQ functions within HR-mediated processes. We currently lack mechanistic insight about the HelQ annealing reaction, and here begin reporting that this requires a 'core' catalytic domain (C-HelQ), and identify a HelQ mutation that triggers hyper-annealing. Intriguingly, this mutation also hyper-activated DNA annealing by the prokaryotic homologue of HelQ (Hel308), indicating an ancient evolutionarily conserved mechanism.

Acknowledgements

I would firstly like to acknowledge my appreciation for my supervisor Prof. Ed Bolt for giving me the opportunity to undertake a project in his lab. Thank you for giving me the space and support to develop the skills, techniques, and resilience to become a scientist. I feel I have learnt so much this year, and have been pushed further than what I would have ever expected I could do in a year.

I also want to say a massive thank you to the members of D58 (past and present) for their wisdom and guidance, and for creating such an enjoyable work environment throughout the year – especially to Tom Killelea, who has been so incredibly supportive and to whom I owe a great deal to; and to Ryan Buckley, who has always kindly listened and given advice to my never-ending questions.

Table of Contents

Chapter 1 Introduction

1.1 DNA Replication

The successful replication of genomes is a prerequisite to cell division across all domains of life. This involves a complex network of events within the cell cycle required for successful duplication of DNA, in which a DNA is replicated during cell division resulting in two daughter cells identical to its parental cell [1]. It is important to understand a cells journey through the cell cycle to understand the complexity of the systems involved in maintaining correct DNA replication, and more so, to understand the repair systems in place if DNA replication goes wrong.

DNA replication is the biological process, conserved across all domains of life, in which DNA sequences are maintained and replicated with high fidelity [2, 3]. In this tightly regulated process, nascent DNA is synthesized from a parental DNA strand, governed by a diverse group of functionally conserved proteins which form a protein complex known as the replisome [4]. While the replisome may be diverse in structure and composition, it is functionally analogous across eukaryotes, bacteria and archaea; with all organisms within these domains of life carrying out semi-conservative replication, originally hypothesized by Watson and Crick [5] [6], resulting in two daughter duplexes containing a parental strand and a newly synthesized strand. To allow for semi-conservative replication, the DNA duplex must first be separated, exposing the nitrogenous bases of the contiguous parental backbone to free DNA bases for complementary binding. However, in doing so this also exposes the single

stranded DNA to other unwanted DNA binding proteins and damaging agents which can limit DNA replication and be a source of DNA damage.

DNA replication can be distinctly categorized into three stages: initiation, elongation and termination, and take place at a branched DNA structures known as replication forks [7]. Tight regulation of DNA replication at each of these stages works to prevent the incorporation of DNA mutations into the genome by maintaining genome stability, achieved through a network of signaling and repair pathways, and therefore prevents the onset of disease cause by genomic instability such as cancer [8]. As such, defining these pathways and understanding their molecular mechanisms is crucial.

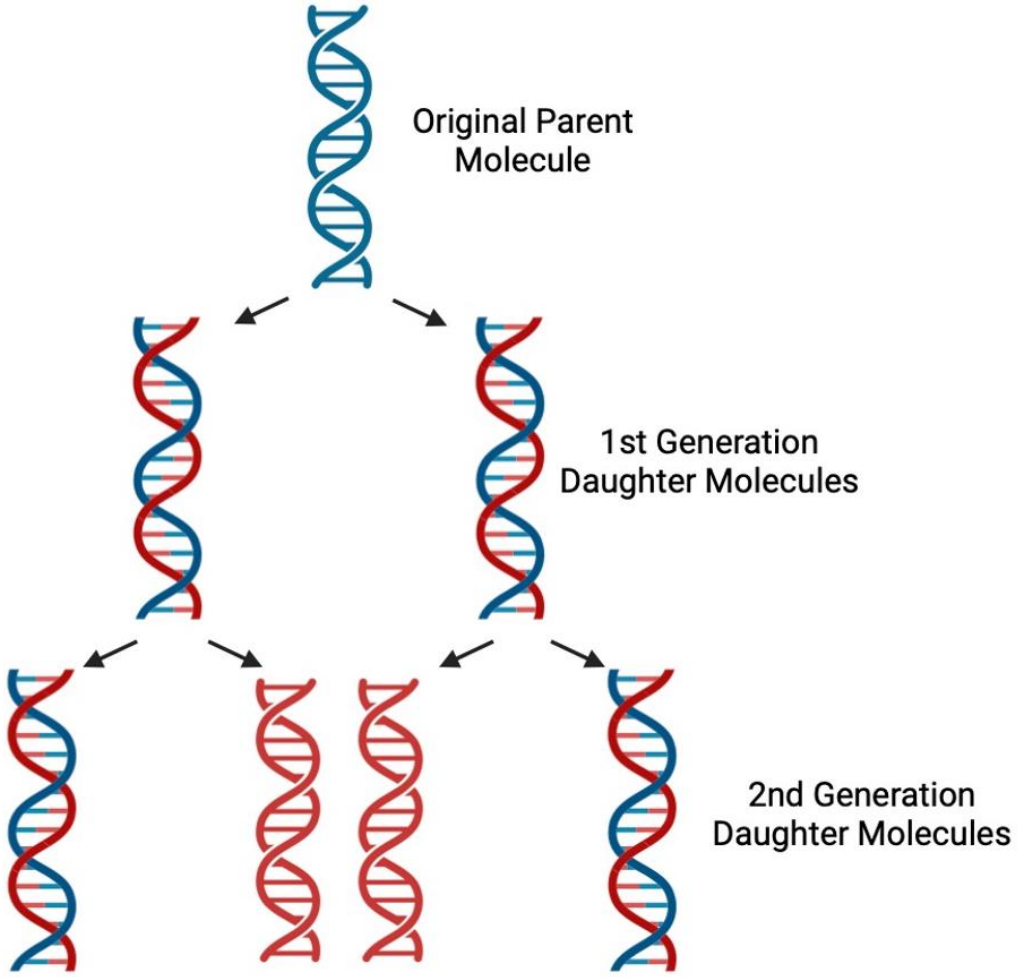


Figure 1.1 Semi-conservative DNA replication. DNA replication is semi-conservative, whereby the replicative process retains each parental strand (blue) in the process of producing two daughter molecules. Each new DNA molecule contains one original parental strand and one newly synthesized DNA strand (red). (Illustration made on BioRender, adapted from Meselson and Stahl [9]).

1.1.1 The Replisome

DNA replication is carried out by the replisome. This large, multi-protein complex functions through the catalytic activities of the proteins it comprises,

and has six core components which are functionally conserved across these three domains of life [10]. As with many life processes shared throughout the domains of life - bacteria, archaea and eukaryota – evolving from the last universal common ancestor (LUCA), the replisome apparatus appears to also have evolved from a common ancestor in archaea and eukaryotes; however, it has been noted the architectures of bacterial replisome machinery, and similar components of their replisomes being non-homologous in sequence when compared to archaea and eukaryota, may suggest the replisome evolved independently for bacteria [11]. Furthermore, the bacterial replisome between species, for example the *Bacillus subtilis* replisome seems to be more "eukaryotic-like" in architecture than the *Escherichia coli (E. coli)* replisome [12].

Assembly of the replisome occurs upon initial loading of its DNA helicase onto double stranded (ds)DNA where it unwinds it, exposing the single stranded (ss)DNA bases [13]. These unwound strands get protected from nucleases by a single-strand binding protein (Single-Strand Binding (SSB) protein in bacteria, and Replication Protein A (RPA) in archaea and eukaryotes), reducing ssDNA secondary structures [6] [11]. The helicases arrange themselves in a hexametric ring, surrounding DNA; in bacteria, this replicative helicase (DnaB in *Escherichia coli)* translocates ssDNA in a 5'-3' direction, on the lagging strand, whereas in archaea and eukaryotes, the replicative helicase MCM (Mini Chromosomal Maintenance protein) translocated in a 3'-5' direction, on the leading strand to separate the duplex [11]. As DNA polymerases cannot initiate synthesis *de novo*, they require a primed site to initiate nascent strand synthesis. The DNA polymerase synthesizes nascent DNA, in an ATP-

dependent manner, from free bases. The polymerases on each daughter strands are different: Polε synthesizes the leading strand while Polδ synthesizes the lagging strand [14]. The lagging strand is synthesized step-wise in 100-200 base pieces in eukaryotes [11] called Okazaki fragments [7, 15]. This requires a primase which primes the lagging strand for Okazaki fragment synthesis. A sliding DNA clamp and clamp loader (the beta subunit homodimer in bacteria, and PCNA homotrimer in archaea and eukaryotes) allows tethering of duplex DNA to the DNA polymerase [11]. The clamp loader maintains the association between then DNA polymerase and DNA helicase. Additionally, the eukaryotic replisome contains other components, Cdc45 and GINS, which interact with the MCM helicase [14].

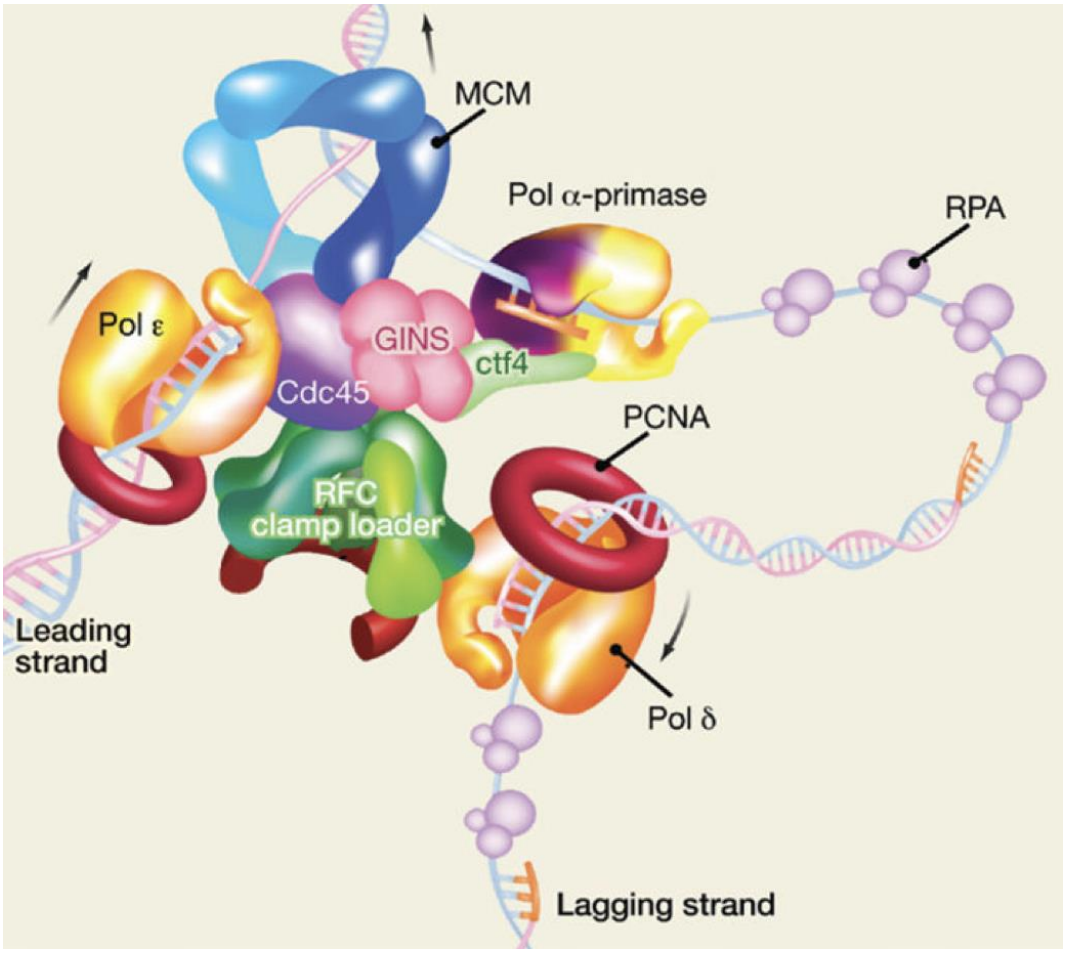


Figure 1.2 The eukaryotic replisome. Diagram of the eukaryotic replisome, highlighting the key functional components required for the replisome to carry out DNA replication. (Figure taken from Yao and O'Donnell [14]).

1.1.2 Origin of Replication and The Cell Cycle

DNA replication begins at positions along the chromosome, in which the DNA helix is opened, known as origins of replication. Typically, simple organisms, such as bacteria and yeast, have one origin of replication which is specified by well-defined DNA sequences at which initiator proteins bind to the DNA. In other eukaryotes, origins are less well-defined and possess multiple sites of origin across its genome – this is due to these higher organisms having a much larger genome compared to bacteria which have a smaller, circular chromosome [6].

Initiation at eukaryotic origins of replication give rise to the assembly of two divergent replication forks which move in opposing directions, beginning the process of DNA replication.

The eukaryotic cell cycle progresses through four phases: G1 (growth 1), S (synthesis), G2 (growth 2) and M (mitosis). The progression of a cell through the cell cycle is determined by various factors including extracellular growth factors, cell health such as cell size and progression through check points such as 'R' within G1 – all of which determine whether the cell can continue through the cell cycle or whether the cell enters G0 (quiescence), a metabolically active, reduced growth state [6, 2]. Eukaryotic replication occurs during the S phase of the cell cycle [10].

All three domains of life use origin-binding proteins to begin the process of activating an origin for replication [6]. These proteins are composed as AAA^+ family subunit(s); the AAA+ superfamily of proteins are a large and functionally diverse family of ATPases characterized by a highly conserved ATPase domain and are involved in a number of cellular functions involving the energy release from ATP hydrolysis [16]. In bacteria, DnaA forms a helical filament which binds to the origin of replication and ATP to unwind A/T-rich regions of bases, which results in a ssDNA 'bubble' where DNA helicases can load onto [17]. In this process, DnaA binds 9-bp stretches of DNA called DnaA boxes, also known as R-sites, and in the presence of ATP, can bind ATP-DnaA boxes, known as I-sites [18]. In eukaryotes, the ring-shaped, hexametric origin recognition complex (ORC), comprised of five ORC subunits (Orc1-5) related to AAA^+

proteins and a sixth AAA+ protein called Cdc6, binds DNA, but unlike in bacteria, this origin binding protein does not unwind DNA in the regions to which it binds. In archaea, AAA^+ proteins related to Orc1 and Cdc6 form a complex, with the number of subunits varying between archaeal species [6, 2]. The purpose of origin binding proteins in all domains of life is to initiate the loading of helicases onto DNA, allowing the formation of replication forks which move in opposite directions to the origin, allowing for DNA to be unwound.

1.1.3 Initiation, Elongation and Termination

The first of the three stages in DNA replication is initiation, where the replisome assembles onto DNA allowing its catalytic activities to take place. For this to occur, first dsDNA must be opened at origins of replication to allow loading of the replisome onto a parental DNA strand.

In bacteria, the melting mechanism of DnaA can be active or passive. The predefined origins of replication are bound by DnaA during origin recognition, at DnaA boxes or ATP-DnaA boxes [18]. Cooperative binding forms DnaA filaments which impose a positive tordial wrap onto bound DNA, causing melting of a flanking AT-rich DNA-unwinding element (DUE) [19]. The replicative helicase complex DnaB:DnaC can then load onto the unwound ssDNA, initiating replication. Similar events occur in archaea, where origin recognition boxes (ORB) flaked by an AT-rich DUE are bound by Orc proteins [20]. These proteins comprise a AAA^+ ATPase domain and a winged-helix domain (WHD) in which DNA helicases bind. In eukaryotes, initiation begins during late M / early G1 phase of the cell cycle. Orc proteins, forming a

preassembled hexameric complex, bind origins of replication, recruiting a AAA^+ ATPase protein Cdc6 to the complex upon DNA binding. This in turn recruits the heterohexameric replicative helicase MCM and the Cdt1 chaperone protein promoting DNA loading. As the cell progresses into S phase of the cell cycle, replisome assembly is completed, along with Cdc45, GINS and the MCM to form the active CMG helicase [21, 22].

dsDNA separation by the DNA helicase into a fork-like structure allows for elongation to occur. This replication fork structure allows other components of the replisome to access the exposed ssDNA. Once the ssDNA has been primed – by DnaG in bacteria and two primase subunits of DNA polymerase α (Pol α) in eukaryotes – the DNA polymerase can synthesize new dNTPs. The leading strand is continually synthesized by DNA Polε in eukaryotes (or DNA Pol III, in bacteria) and the lagging strand is synthesized in Okazaki fragments by DNA Polδ [23]. As mentioned previously, the DNA helicases and polymerases are bridged by proteins to provide structural stability and allows tethering of DNA to the polymerase. In eukaryotes and archaea, this protein is PCNA [24]. In bacteria, this protein is known as the β-clamp.

1.2 Replication Stress

Replisome processivity is constantly threatened by many inhibitory events during DNA replication. This may inhibit the functions of the DNA polymerase and/or DNA helicase and contribute to replicative stress, slowing the replisome and possibly resulting in stalled replication forks [25]. Sources of replicative stress include direct replication fork barriers, such as DNA lesions, and the formation of secondary DNA structures, such as R-loops and hairpins [26]. DNA lesions have particularly been found to be detrimental to fork progression, which have been found to arise from endogenous and exogenous sources which include UV radiation, reactive oxygen species (ROS) and chemical mutagens [27]. Additionally, replicative stress can also result from an imbalance in the dNTP pool, stalling the replicative polymerase. Stalled replication forks are a major source of genome instability and collapsed forks can lead to lethal double strand breaks within DNA. This can result in mutations in the genes encoding the essential proteins involved in replication and repair. With genome instability being a known hallmark of cancer and a driving factor in tumour progression, unresolved problems resulting from these events can be fatal. Tandem duplications that spontaneously accumulate in genomes, genome rearrangements and deleterious events are characteristic of many cancers with overall complex karyotypes defined by structural and numerical changes [28]. Understanding the molecular mechanisms which lead to such events are therefore of great importance to identify pathways that can be exploited in order to prevent these events from occurring [29].

Prevention of replication fork stalling is aided by accessory helicases that associate and translocate with the replisome, disrupting non-covalent nucleoprotein complexes ahead of the replication fork [30]. Examples of such accessory helicases include the *E. coli* Rep and UvrD helicase [30, 31]. This ensures translocation of the replisome is not impeded and limits the possibility of genome instability caused by stalled or collapsed replication forks. These interfering cellular stresses that pose a threat to genome stability have allowed cells to evolve a network of responses to address these obstacles caused by replication fork stalling and collapse and there are many response mechanisms in place to allow the replisome to bypass DNA damage or roadblocks that could lead to the stalling or collapse of replication forks – collectively these responses are known as DNA damage tolerance (DDT); in which pathways include (but are not limited to) translesion synthesis (TLS), template switching (TS), break-induced repair (BIR) and homologous recombination (HR) [32].

Yet despite these systems allowing progression of replication forks, DNA damage often persists, and so additional repair pathways exist to repair post-replicative DNA damage.

1.3 DNA Damage and Repair

Cellular DNA is continually exposed to mutagenic agents which result in chemical alterations to the structure of DNA. Ultimately this can result in aberrant replisome activity, leading to replication fork stalling and/or the incorporation of incorrectly paired bases. Several unique mechanisms for controlling mutagenic effects have been evolved. These specialized repair pathways exist to act at sites of DNA damage aiding the fidelity of DNA replication and other associated replicative processes. The mechanism of repair employed can vary, depending on the type of DNA lesion and source of DNA damage (Figure 1.3).

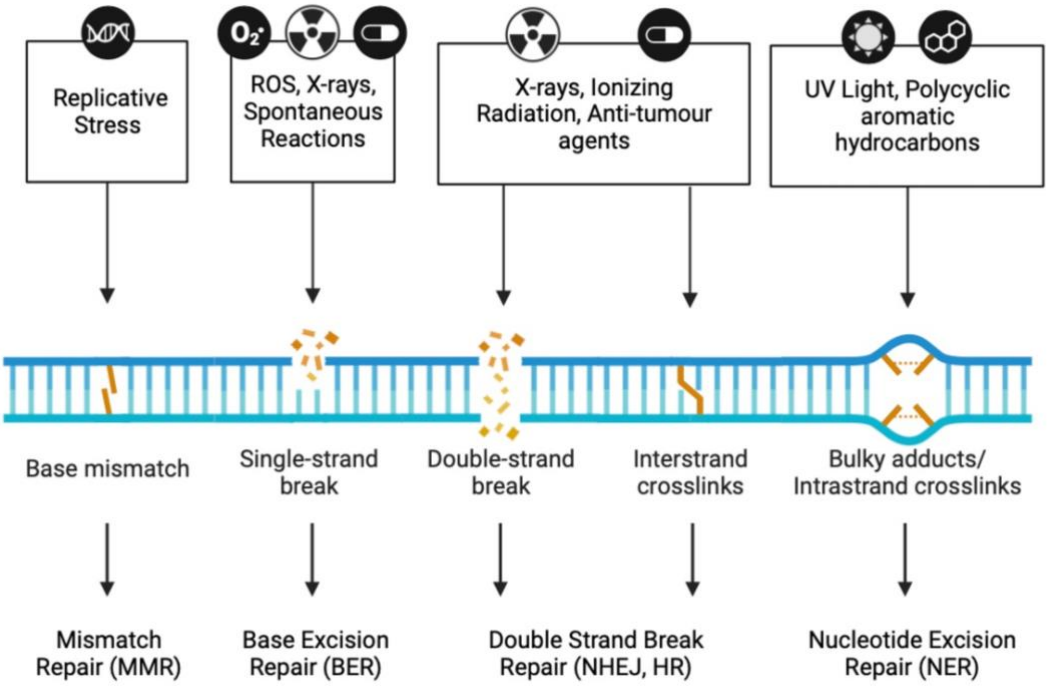


Figure 1.3 DNA Damage and Repair Pathways. DNA is exposed to exogenous and endogenous agents constant, resulting in DNA damage. There are many potential sources of DNA damage, such as X-rays, UV light and replicative stress. This results in a variety of DNA

lesions. Repair pathways have evolved to correct the chemical damage to DNA and maintain genome integrity. The repair mechanism required depends on the type of damage caused to DNA. Major eukaryotic repair pathways include (but are not limited to) mismatch repair (MMR), base excision repair (BER), homologous recombination (HR), non-homologous end joining (NHEJ) and nucleotide excision repair (NER) [33, 34, 35]. (Figure made on BioRender).

DNA replication is arguably the most essential biological process that occurs; however, it is also when the cell is at its most vulnerable state. Due to this vulnerability giving rise to DNA damage by the form of breaks and mutations within the structure of DNA, repair pathways have co-evolved with DNA replication to repair any damage caused in the process [33, 34, 35].

1.3.1 Mechanisms of DNA repair

Evidenced by the number of proteins cells invest into carrying out DNA repair or aiding its function, these repair pathways and mechanisms are highly important. Many repair proteins were first identified, in bacteria, through genetic studies indicating that when the proteins are mutated leading to a functional change, there was increased rate of mutation and sensitivity to DNA damaging agents [2]. It has since been shown that many human diseases can be linked to dysfunction of DNA repair proteins – possibly the most infamous being cancers. For example, mutations in the *Brca1* and *Brca2* genes, known to encode proteins involved in homologous recombination, are a known cause of hereditary breast and ovarian cancer [36].

Translesion synthesis (TLS) is one pathway of DNA repair that can be exploited to bypass leading strand damage, continuing downstream DNA replication. TLS

recruits alternative polymerases, which lack proofreading functions to bypass abnormal DNA, replacing the replicative polymerases. While these polymerases may be lower in processivity, it allows the progression of DNA synthesis [32].

Excision repair is another pathway that can process DNA lesions with chemical alterations, often caused by exogenous agents. These pathways involve DNA glycosylases that recognize specific altered DNA bases, catalyzing their hydrolytic removal. The base excision repair (BER) pathway targets small lesions, while nucleotide excision repair (NEB) target “bulky” lesions which lead to large changed in the DNA structure, often causing helix distortion [2, 31, 37].

The mismatch repair (MMR) pathway is a post-replicative repair system that finds incorrectly paired bases and insertion/deletion loops formed during DNA replication. A MutS homodimer recognizes any mismatched bases which then recruits other proteins such as MutL, MutH and UvrD to nick, remove and repair the DNA. MMR has been found to be a critical pathway in maintaining genome stability [38].

It is also important to note DNA end resection has a key role in DSB repair, owing to its role in error-free repair essential during homologous recombination. CtBP-interacting protein (CtIP) functions with the Mre11-Rad50-NBS1 (MRN) complex to generate 3’-ssDNA overhangs at the DSB ends. This is then reliant on downstream nucleases and helicases, for example exonuclease 1 (EXO1) or Bloom syndrome protein (BLM), for HR-mediated repair [39].

1.3.2 Homologous recombination

Homologous recombination (HR) is a ubiquitous process and critical DNA repair pathway for maintaining genomic stability. This process has co-evolved in cells to repair DNA damage that occurs during DNA replication, owing to its ability to repair DNA double strand breaks, and is crucial to recovering stalled or collapsed replication forks [40, 41]. Homologous recombination can also repair lesion-mediated breaks such as interstrand cross-links that can prevent transcription and replication. HR relies on the presence of a repair template to synthesize nascent DNA; in eukaryotic cells this is therefore limited to the synthesis (S) and growth 2 (G2) phases of the cell cycle, where there is a newly synthesized repair template. These HR events are found to occur in many circumstances during DNA replication, DNA repair, and telomere maintenance. HR competes with other repair pathways, such as the DDT mechanisms previously described, and non-homologous end joining (NHEJ) when repairing DSBs [32]. As HR is less mutagenic, it is more favorable than NHEJ.

HR encompasses a group of repair sub-pathways. These include the classical double-strand break repair (DSBR), synthesis-dependent strand annealing (SDSA) and break-induced repair (BIR). Whether it is resulting from ionizing radiation, an intermediate of other repair pathways or arising from replication through a single stranded break, double strand breaks can be incredibly genotoxic and leading to disastrous effects if not efficiently repaired [42].

HR-mediated repair can be divided into three overarching steps: pre-synapsis (end resection and strand invasion), synapsis (strand exchange and branch migration) and post-synapsis (Holiday junction (HJ) resolution). Prior to recombination, the broken ends of DNA must first be resected. This allows for the loading of DNA repair proteins (RPA in eukaryotes) onto the newly exposed ssDNA. In turn this amplifies the DNA damage repair signal, leading to cell cycle arrest. Recombinase proteins then replace RPA, which oligomerize onto the exposed 3' ssDNA to form a nucleoprotein filament, subsequently catalyzing homology search and strand invasion by Rad51 (RecA in bacteria and RadA in archaea) [43, 37]. The filament invades homologous regions of DNA, forming a displacement loop (D-loop), and as it contains a 3' end, it can promote extension by DNA polymerases to form a HJ [37]. Resolution or dissolution of the HJ completes the recombination process, ending with strand separation into two species of duplex DNA – resulting in crossover or non-crossover products [40].

Employing the functions of DNA repair helicases can be used for extensive remodeling to recover stalled replication forks. Fork reversal may result in a "chicken foot" intermediate, giving access to the lesion for repair [40]. Recombination proteins can then load onto the intermediate leading to D-loop formation, and the replisome can then be loaded onto the invading strand, restarting DNA replication.

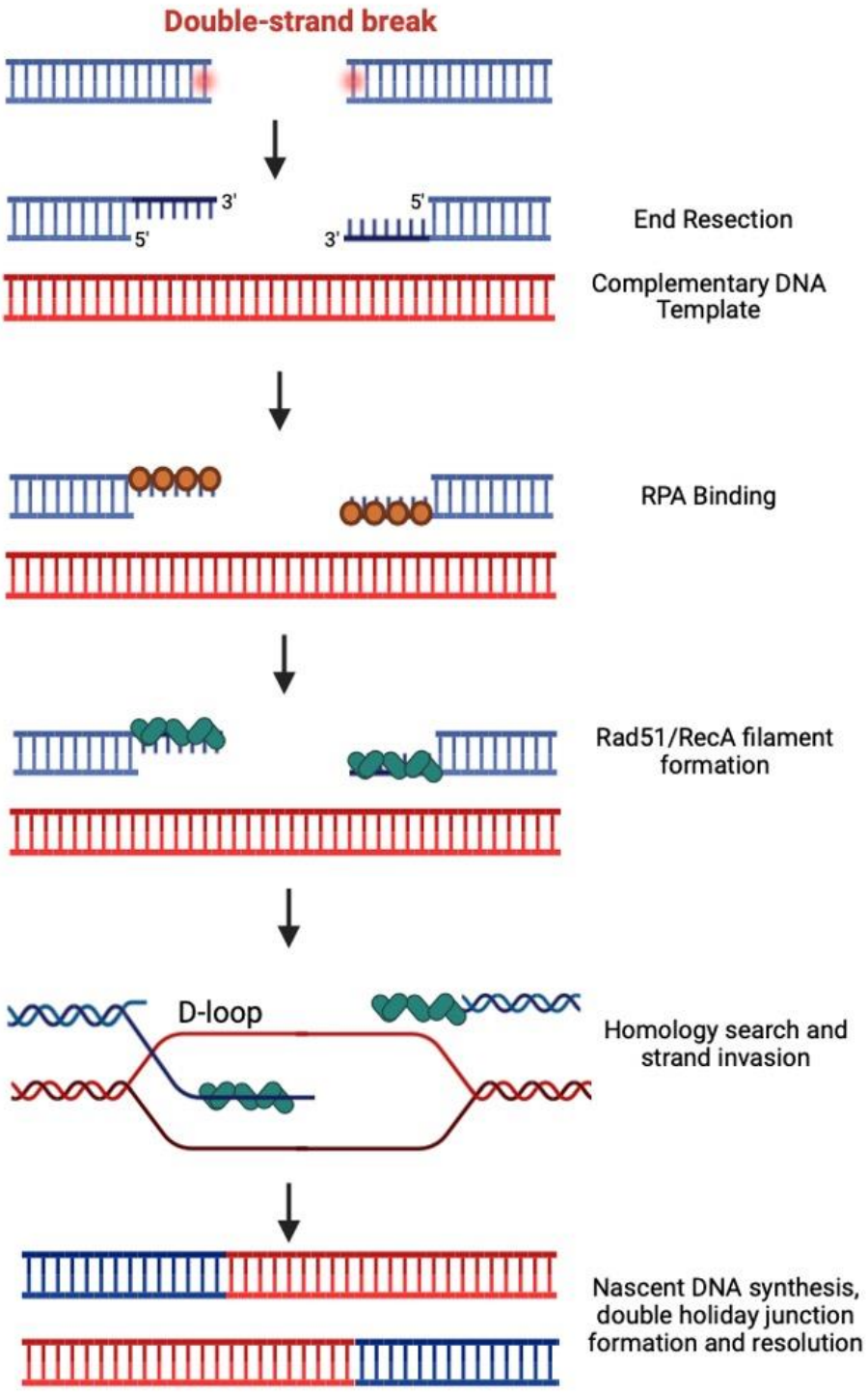


Figure 1.4 Overview of double-strand break repair via homologous recombination. Simplified overview of double strand break repair by homologous recombination proteins in eukaryotes, with key components

highlighted. Crossover and non-crossover products are possible. (Figure made on BioRender).

While HR has evolved as a form of protection against DNA double strand breaks and to aid the maintenance of genome stability, studies have shown it can be quite error -prone and mutagenic; especially DNA synthesis that occurs during the extension of the D-loop intermediate [44]. This continues to contribute to the hallmarks observed in cancers, owing to genome instability.

In eukaryotes, Ski-2 like DNA helicases have been identified to play important roles in alleviating replicative stress, promoting genome stability and restarting the replication fork.

1.4 DNA Helicases

DNA helicases play a central role in nucleic acid metabolism. They are a subset of a larger group of nucleic acid dependent ATPase's known as the AAA+ superfamily of molecular motor proteins. DNA helicases function with directional polarity (3'-to-5' or 5'-to-3') to translocate DNA through coupling the energy release from adenosine triphosphate (ATP) hydrolysis to a conformational change in its protein structure, resulting in nucleic acid unwinding or remodeling via breaking the hydrogen bonds between paired bases [45] [46]. They also play additional roles in displacing nucleoprotein complexes that inhibit DNA replication [47]. Helicase classification is dictated by the conserved domain structure and organization of the core ATPase and helicase domains – and other functional motifs – allowing further classification into superfamilies [48].

1.4.1 Superfamily-1 and Superfamily-2 Helicases

Helicases, which are ubiquitous across all domains of life, can be subdivided into six superfamilies (Superfamily-1 to -6 (SF1-6)), categorized by sequence similarity, structural organization and conservation of motifs within in core domains, such as the Walker A and Walker B motifs responsible for ATP binding and ATP hydrolysis [45]. Helicase subdivision can be further extended based on toroidal ring formation. SF3-6 are toroidal enzymes and SF1/2 are non-ring forming enzymes. Subfamilies within these groups are characterized by polarity, DNA/RNA substrate preference and by sequence, structural and mechanistic features [48].

Most helicase enzymes fall within the SF1 and SF2 families. These proteins show considerable structural similarities with near identical folds and extensive signature motif conservation within the dual RecA core-helicase domains [49]. Such features give rise to their common structure-function relationship, and importantly residues that co-ordinate ATP binding and hydrolysis. Despite this, the different helicases within these family's function within a wide range of processes and on a diverse range of substrates; and while they may be classified as helicases, duplex unwinding may not always their primary function [48].

SF2 helicases, the focus of this study, is the largest and most diverse family having 10 subgroups within it. They are collectively referred to as DExH/D helicases due to the conservation of the DExH/D box motif [48]. These helicases play roles within DNA transcription and repair, RNA metabolism and chromosome rearrangement [45]. There are nine primary signature motifs found to be conserved within SF2 helicases which are oriented within the two RecA-like domains, forming the helicase core [50]. These motifs are responsible for forming contacts with DNA allowing the helicase core to act as a motor domain when coupled with ATP hydrolysis [50].

1.4.2 Ski2-like and Hel308 Family Helicases

One relatively small family of helicases within the SF2 superfamily are the Ski2-like family that have functions in RNA degradation, processing and splicing [51]. These RNA helicases have distinct features that distinguishes them from other SF2 proteins. Ski2-like helicases are large in size and possess 12 conserved

sequence motifs within the two RecA-like domains. Structural features include a β-hairpin for duplex strand separation, a helical 'ratchet' domain involved in DNA unwinding and a winged helix domain (WHD). Along with the RecA domains, the helical ratchet and WHD form a ring-like four domain assembly known as the DExH/D box core – this common architecture is sometimes also referred to simply as the helicase core [51].

Hel308 proteins are a unique sub-group of Ski2-like family proteins that have a preference for DNA, and are implicated in DNA repair, recombination and genome stability. The first member of the family identified was the *Drosophila melanogaster* Mus308 [52]. This was identified in a genetic screen where mutation of the *mus308* gene caused hypersensitivity to inter- and intra-strand DNA crosslinks upon exposure to nitrogen mustard [53, 54]. The structure of mus308 showed an N-terminal SF2 helicase fused to a C-terminal DNA polymerase. This in turn lead to the identification of the human ortholog, PolQ, as it displayed the same structural arrangement. A subsequent discovery found another metazoan ortholog of this helicase-polymerase protein which only constituted the helicase – this is Hel308.

Hel308 family proteins are found in metazoans and archaea, but not in bacteria and fungi [55]. They have been shown to possess RecQ-like properties, and deficiencies cause hypersensitivity to ICL-inducing agents and function in DSB repair and recombination. In addition to the conserved motifs shared between Ski2-like proteins, Hel308 also contains an additional auto-inhibitory helix-

hairpin-helix (HhH) domain which limits helicase activity through acting as a molecular break [51].

As mentioned above, PolQ is a helicase-polymerase protein within the Hel308 family proteins, and the human ortholog of *D. melanogaster* Mus308. PolQ primarily functions as a DNA polymerase and has a role in TLS [56]. The C-terminal third of the protein is a family A DNA polymerase and the N-terminal third contains the helicase domain – a unique and unusual structure for Ski2-like family proteins [57]. It has been suggested that PolQ functions in DSB repair pathways, specifically in mechanisms involving end-joining repair and HR, alternative from the canonical NHEJ. These pathways can involve alternative end joining (altEJ), microhomology-mediated end joining (MMEJ), synthesis-dependent end joining (SD-MMEJ) and theta-mediated end joining (TMEJ) [57].

1.5 HelQ

HelQ is an SF2 ATP-dependent DNA helicase, within the Ski2-like subfamily, that primarily functions in DNA repair. While its oligomeric structure remains unresolved, its archaeal homologue Hel308 and other related helicases such as PolQ have been used as a model to predict its structural organization, mechanistic function and its potential roles and interactions within DNA repair and replication.

1.5.1 Discovery

The discovery of HelQ began in 1976 when mutations in the Drosophila gene mus308 were first identified in a mutagen sensitivity screen to cause hypersensitivity to ICL-inducing agents, as outlined earlier [52]. Two decades later in 1996, the Drosophila mus308 gene was successfully cloned and was reported to shown to share significant homology to DNA polymerase and helicase motifs [15]. A few years later, the human DNA helicase homologous to mus308 was isolated from cDNA expression libraries and named Hel308 [53].

Human Hel308 was identified to be located on chromosome 4q21 and encode an 1101 amino acid polypeptide, showing to function as a ssDNA-dependent ATPase and DNA helicase; and through stimulation by RPA, was found to displace duplex substrates. Following this, in 2003, a mammalian ortholog of mus308 has been identified, cloned and characterized, and named PolQ – it was suggested to function in DNA repair through helicase and polymerase activities [58].

In 2005, archaeal hel308 was identified and characterized. Hel308 is a Ski2-like helicase that is functionally conserved in archaea and eukaryotes. First identified in a genetic analysis using a *E. coli* strain that conditionally accumulates stalled forks, *Methanothermobacter thermautotrophicus* Hel308 expression caused synthetic lethality, with further analysis suggesting it functions at stalled replication forks independent of HR and the RuvABC pathway [59]. Characterization of the protein has since suggested Hel308 promotes genome stability through functioning in DNA repair at replication forks. It has been shown Hel308 functions as a ssDNA ATPase, conferring a 3'-to-5' polarity and has a preference to unwind forked DNA structures with lagging strands [59, 60]. Duplex DNA separation is facilitated by the joint activity in an α-helical ratchet in domain IV and β-hairpin in domain II. This stimulates a "inchworm" translocase mechanism along ssDNA backbone in which the two RecA domains undergo conformational changes [55]. Helicase activity is also stimulated by RPA.

In 2008, two distinct pathways of ICL repair were described in *Caenorhabditis elegans* involving PolQ and Hel308. PolQ was suggested to be involved in a repair pathway with BRC-1 (the nematode BRCA1 homologue) and Hel308 was suggested to be involved with FCD-2 (the nematode functional homologue of FANCD2) – suggesting its role in the FA pathway [61, 62].

1.5.2 HelQ protein structure and domain organisation

With HelQ being the human homologue of Hel308, it inevitably shares similarities in its structure and associated function. Human HelQ possesses the

highly conserved helicase core composed of the two RecA-like ATPase domains found in Ski2-like helicases, and as seen in many other SF2 family of proteins, giving HelQ its 3' to 5' ATP-coupled translocase activity and dsDNA unwinding ability. As with proteins within this family, HelQ also contains the nine conserved motifs, including the Walker A and Walker B motifs, essential for ATP binding and hydrolysis, and so also its subsequent translocation activity. HelQ also contains the winged-helix domain (WHD) and helicase ratchet which participate in DNA unwinding and are essential for interacting with forked DNA structures. HelQ however lacks the HhH auto-inhibitory domain found in archaeal Hel308. The N-terminal region of HelQ has been shown to be intrinsically disordered, yet despite this, it contains a PWI-like domain, responsible for the displacement of RPA, within this region [63].

1.5.3 Characterising the structure-function relationship of HelQ

Most recent literature on HelQ has focused on characterizing the mechanisms by which HelQ functions in DSB repair and the identification of potential protein-protein interactions that may suggest the role of HelQ in DNA repair and replication.

Novel data identified a conserved fold in human HelQ that triggers the displacement of RPA from ssDNA, suggesting a mechanism of how HelQ loads onto ssDNA at stalled replication forks and revealing why RPA stimulates HelQ helicase activity [64]. The functional role of the non-catalytic N-terminal region was first described, identifying its PWI-like domain to mediate the interaction

with RPA, allowing the loading of the core helicase domain onto ssDNA, triggering translocation as a dimer [64]. This knowledge is important for giving insight into the mechanisms of HelQ function relevant to avoiding genome instability.

Through CRISPR-induced double-strand breaks, it was found that HelQ is required for all double-strand break repair mechanisms that are guided by the annealing of complementary bases at end breaks; notoriously known to be mutagenic and therefore may be a primary cause of double-strand break induced genomic alterations [65]. It was found that loss of HelQ led to increased genome instability, causing patchwork insertions to be found at deleted junctions due to abortive rounds of polymerase theta activity, and tandem duplications found to spontaneously accumulate in the genomes of HelQ mutant animals.

HelQ's function as a double-strand break repair enzyme was also shown to be modulated by RPA and Rad51 in a co-factor-dependent mechanism [28]. Its helicase activity and an annealing function were shown to be differentially regulated by RPA and Rad51. This established the translocation of HelQ during DNA unwinding is stimulated by Rad51, and contrastingly, RPA inhibits DNA unwinding by HelQ, instead strongly stimulating DNA strand annealing [28].

Additionally, DNA polymerase delta (PolD) has been explored in an attempt to identify possible interactions with HelQ and mechanisms to control DNA synthesis [60]. PolD is involved in DNA break repair through several

homologous recombination pathways where it synthesizes new DNA but can trigger mutagenesis too. PolD is comprised of four subunits; the catalytic subunit PolD1, and three regulatory subunits, PolD2 which acts as a scaffold, PolD3 which is an intrinsically disordered protein and PolD4. PolD forms a complex with PCNA and RPA to allow for DNA synthesis. PCNA first binds to D-loops [66, 67] that are formed during homologous recombination, which then recruits PolD in the process. While mechanisms which prevent PolD beginning to replicate have been defined, the mechanisms for stopping PolD once it begins replication and repair are not known.

Mechanisms that limit or prevent mutagenesis arising during DNA synthesis via homologous recombination have been found to involve the deployment of helicase enzymes. They prevent or dissociate D-loop structures, instead priming new DNA synthesis from a strand break. This is achieved via ATP-dependent translocation of DNA at D-loops therefore disrupting base-pairing between the replication priming DNA strand and its template.

Recent studies revealed a potential mechanism to control DNA synthesis during homologous recombination through promoting DNA strand annealing through studying the interaction between PolD and HelQ [60]. It was shown that DNA synthesis can be halted by HelQ via PolD and a holoenzyme of PolD in complex with RPA and PCNA *in vitro.* This instead stimulated DNA single-strand annealing. Both functional and physical interactions between HelQ and PolD were explored; which revealed the interaction between the N-terminus of HelQ and PolD3 was significant for inhibiting DNA unwinding. This overall suggests

a mechanism by which HelQ limits the extent of mutagenic DNA synthesis during DNA repair – and therefore limiting carcinogenesis.

Most recently, single amino acid substitution mutations in *M. thermautotrophicus* Hel308 have shown to result in hyperactive DNA processing activity: DNA binding, unwinding and annealing functions [68]. Data revealed archaeal Hel308 suppressed recombination while promoting DNA repair, and that a motif within the dual RecA domain can act as a catalytic switch to modulate recombination and repair functions. As Hel308 is the archaeal homologue of human HelQ, it is of interest to learn more about the effects of these same mutations in HelQ.

1.6 Project Aims

The initial aims of this project were to study the biochemistry of the human DNA repair protein HelQ to develop our understanding and characterize a newly identified annealing mechanism and develop on any protein-protein interactions of interest. Experimental procedures set out to achieve this included assessing HelQ's annealing activity by gel- and FRET-based assays. Additionally, the hypothesis that HelQ requires longer regions of DNA strand complementarity than PolQ would also be assessed; informing why both proteins co-exist (predictively one for HR and the other for MMEJ). We currently lack insight about the HelQ annealing mechanism, and here begin reporting that this requires a core catalytic domain (C-HelQ) and identify mutations that trigger hyper-annealing phenotype.

Intriguingly, studies have revealed this mutation leads to hyper-activated DNA annealing by the archaeal homologue of HelQ, Hel308. This project was therefore further developed into assessing whether there is evidence for an ancient evolutionary conserved mechanism of DNA annealing, from archaea to humans.

Chapter 2 Materials and Methods

2.1 Chemicals

Chemicals listed within this methodology section were supplied either by ThermoFisher Scientific or Sigma, unless stated otherwise.

2.2 Antibiotics

Table 2.1 Antibiotics and their used concentrations.

Antibiotic	**Stock concentration**	**Working concentration**
Ampicillin	100 mg/mL	50 µg/mL
Chloramphenicol	35 mg/mL	35 µg/mL
Kanamycin	50 mg/mL	50 µg/mL

2.3 Bacterial cell strains

Table 2.2 List of *E. coli* cell strains used in this work.

Name	**Supplier**	**Use**	**Feature**
DH5α	Invitrogen	*E. coli* strain used for cloning.	Amp^R
BL21 AI	Invitrogen	*E. coli* strain used for protein over-expression.	Tet^R

2.4 Plasmids, oligonucleotides and DNA substrates

Oligonucleotides used for primers and the construction of DNA substrates were sources from Sigma, supplied lyophilized, and diluted in sterile-distilled water (SDW) to 100 μM.

Table 2.3 List of vector backbones used for cloning in this work.

Plasmid	Size	Resistance	Features
pET14b	4671bp	Amp^R	pBR322 origin, T7 promotor/terminator, hexaHis-tag coding sequence, MCS (NcoI-XhoI)
pACYC-duet	4008bp	$Chlm^R$	p15A origin, T7 promotor/terminator, hexaHis-tag coding sequence, MCS (NcoI-NotI)

Table 2.4 List of plasmids used in this work.

Plasmid	Description	Resistance	Origin
pTJ009	N-terminal hexaHis-SUMO tagged N-terminal terminated at Ile-240 of HelQ in pET14b background	Amp^R	Tabitha Jenkins
pHB001	HexaHis tagged C-terminal HelQ in	$Chlm^R$	Hannah Betts

	pACYC-duet background.		
pTJ015	Y642A C-HelQ mutant in pACYC-duet background to give C-HelQ^{Y642A}	Chlm^{R}	Tabitha Jenkins
pALH009	P960A-Y963A C-HelQ mutant in pACYC-duet background to give C-HelQP$^{960A-Y963A}$	Chlm^{R}	This work
pBAD-HisA-NmVenus	1-154 amino acids of mVenus cloned into pBAD-HisA via NheI site	Amp^{R}	Liu He
pRSF-1b-CmVenus	155-238 amino acids of mVenus cloned into pRSF-1b via NcoI site	Kana^{R}	Liu He
Plasmid #65777	Mamalian expression plasmid for prime editing sgRNA	Amp^{R}	Addgene
Plasmid #132777	Mamalian expression plasmid for prime editing pegRNA	Amp^{R}	Addgene

pTJ011	HexaHis SUMO tagged truncated N-HelQ.	Amp^R	Tabitha Jenkins
pTJ010	HexaHis SUMO tagged PWI N-HelQ mutant (N-$HelQ^{D142A/F143A}$).	Amp^R	Tabitha Jenkins
pRSF-CmVenus-NHelQ	N-HelQ (from pTJ009) cloned into pRSF-CmVenus.	$Kana^R$	Liu He
pALH001	Truncated N-HelQ cloned into pRSF-CmVenus (pRSF-CmVenus-$NHelQ^{Trun}$)	$Kana^R$	This work
pALH002	D142AF143A N-HelQ mutant cloned into pRSF-CmVenus. (pRSF-CmVenus-$NHelQ^{D142AF143A}$)	$Kana^R$	This work
pALH003	N-HelQ RG3 mutant cloned into pRSF-CmVenus. (pRSF-CmVenus-$NHelQ^{RG3}$)	$Kana^R$	This work
pALH004	N-HelQ RG3KG mutant cloned into	$Kana^R$	This work

	pRSF-CmVenus. (pRSF-CmVenus-NHelQ^{RG3KG})		
pALH005	C24A N-HelQ mutant cloned into pRSF-CmVenus. (pRSF-CmVenus-NHelQ^{C24A})	Kana^{R}	This work
pALH006	C74A N-HelQ mutant cloned into pRSF-CmVenus. (pRSF-CmVenus-NHelQ^{C74A})	Kana^{R}	This work
pALH007	C239A N-HelQ mutant cloned into pRSF-CmVenus. (pRSF-CmVenus-NHelQ^{C239A})	Kana^{R}	This work
pBAD-NmVenus-PolD1	PolD1 cloned into pBAD-NmVenus.	Amp^{R}	Liu He
pBAD-NmVenus-PolD2	PolD2 cloned into pBAD-NmVenus.	Amp^{R}	Liu He
pBAD-NmVenus-PolD4	PolD4 cloned into pBAD-NmVenus.	Amp^{R}	Liu He
pBAD-NmVenus-PolD3	PolD3 cloned into pBAD-NmVenus.	Amp^{R}	Liu He

pALH008	PolD3 KKRRR mutant cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3^{Quincunx})	Amp^{R}	This work
pALH010	Residues 1-106 of PolD3 cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3$^{aa1-106}$)	Amp^{R}	This work
pALH011	Residues 107-466 of PolD3 cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3$^{aa107-466}$)	Amp^{R}	This work
pOLD001	PolD3 IDPR1 cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3$^{169-188}$)	Amp^{R}	Olivia Downs
pOLD002	PolD3 IDPR2 cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3$^{199-232}$)	Amp^{R}	Olivia Downs
pOLD003	PolD3 IDPR3 cloned into pBAD-NmVenus.	Amp^{R}	Olivia Downs

	(pBAD-NmVenus-PolD3$^{274-393}$)		
pOLD004	PolD3 IDPR4 cloned into pBAD-NmVenus. (pBAD-NmVenus-PolD3$^{406-466}$)	Amp^{R}	Olivia Downs
pBAD-FLmVenus	Full length mVenus fluorescent protein cloned into pBAD-HisA.	Amp^{R}	Liu He

Table 2.5 List of primers used for cloning, mutagenesis and plasmid verification in this work.

Target Gene	**Use**	**5'-3' Sequence (F = Forward, R = Reverse)**
C-HelQ	To create a Pro-960-Ala, Tyr-963-Ala double point mutation in C-HelQ to produce C-HelQ$^{P960A-Y963A}$.	F:ATTCAATATGgcgCGTGGC gcgATTCAGAATCTGC R:TTTTCGCTAACGGTCCAAAT
Duet Down 1	Internal reverse primer for sequencing C-HelQ$^{P960A/Y963A}$.	5'GATTATGCGGCCGTGTACAA
spCas9 gRNA	To clone "sg" insert into "65777" vector	F: caccgTATAGCTCAAGT TGACGACC

	backbone via golden gate cloning for prime editing of F142D143 mutant.	R: aaacGGTCGTCAACTT GAGCTATAc
	To clone “sg” insert into “132777” vector backbone via golden gate cloning for prime editing of F142D143 mutant.	F: caccgTTCCGAGCAA AGATTTTCAGgttttaga R: tagctctaaaacCTGAA AATCTTTGCTCGGAAc
	To clone “scaff” insert into “132777” vector backbone via golden gate cloning for prime editing of F142D143 mutant.	F: GCTAGAAATAGCAAG TTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGT GGCACCGAGTCG R: GCACCGACTCGGTGCCA CTTTTTCAAGTTGATAACGGA CTAGCCTTATTTTAACTTGC TATTTC
	To clone “extens” insert into “132777” vector backbone via golden gate cloning for prime editing of F142D143 mutant.	F: gtgcACAGCCGCCGCCA CTGAAAATCTTTGCTCGGA R: aaaaTCCGAGCAAAGATT TTCAGTGGCGGCGGCTGT

spCas9 gRNA	To clone “sg” insert into “65777” vector backbone via golden gate cloning for prime editing of Y642A mutant.	F: caccgAACTTGAAGAATA TTGGCAA R: aaacTTGCCAATATTCTT CAAGTTc
	To clone “sg” insert into “132777” vector backbone via golden gate cloning for prime editing of Y642A mutant.	F: caccgTGTTAAGCCACTGT GGTGATgttttaga R: tagctctaaaacATCACCACAG TGGCTTAACAc
	To clone “scaff” insert into “132777” vector backbone via golden gate cloning for prime editing of Y642A mutant.	F: GCTAGAAATAGCAA GTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAA AGTGGCACCGAGTCG R: GCACCGACTCGGTGCCA CTTTTTCAAGTTGATA ACGGACTAGCCTTATTTTA ACTTGCTATTTC
	To clone “extens” insert into “132777” vector backbone via golden gate cloning	F: gtgcTGGAGTTGC CGCCCACCACAGTGGCTTAA R: aaaaTTAAGCCACTG TGGTGGGCGGCAACTCCA

	for prime editing of Y642A mutant.	

Table 2.6 List of Oligonucleotides used for biochemical assays.

Substrate name	**Oligonucleotide name**	**Oligonucleotide Sequence 5' to 3'**	**Use**
Fork2	MW12	GTCGGATCCT CTAGACAGCT CCATGATCAC TGGCACTGGT AGAATTCGGC	Gel-based Helicase Assay Substrate
	MW14 5' Cy5	CAACGTCATA GACGATTACA TTGCTACATG GAGCTGTCTA GAGGATCCGA	
FRET Fork2	MW12 5' Cy3	GTCGGATCCT CTAGACAGCT CCATGATCAC TGGCACTGGT AGAATTCGGC	FRET Helicase Assay Substrate
	MW14 5' Cy5	CAACGTCATA GACGATTACA TTGCTACATG	

		GAGCTGTCTA GAGGATCCGA	
FRET Anneal	ELB41 5- Cy5	GCAGGATCCG TATCCGTAAC TGGAGCTCTT CGAAGGCCAT CGTCGCGAAC GATCCTGCCT AGGGAGCTCC	Gel-based Helicase Assay Substrate / Anneal control for gel- and FRET-Annealing Assay
	ELB40 5'-Cy3	GGAGCTCCCT AGGCAGGATC GTTCGCGACG ATGGCCTTCG AAGAGCTCCA GTTACGGATA CGGATCCTGC	
ssDNA	ELB41 / ELB41 5-Cy5	GCAGGATCCG TATCCGTAAC TGGAGCTCTT CGAAGGCCAT CGTCGCGAAC GATCCTGCCT AGGGAGCTCC	Annealing Assay / EMSAs
ssDNA	ELB40 / ELB40 5-Cy3	GGAGCTCCCT AGGCAGGATC	Annealing Assay

		GTTCGCGACG ATGGCCTTCG AAGAGCTCCA GTTACGGATA CGGATCCTGC	
ssDNA	ELB40-F	CGGATCCTGC AAAAAAACAG TTTAAACGTA CGATTATTGC CGTCAATGTC GCAACTGATC CGTTTAAGTT	Annealing assays; 10-nt homology with 5’ of ELB41
ssDNA	ELB40-G	GTTACGGATA CGGATCCTCG AAAAAAACAG TTTAAACGTA CGATTATTGC CGTCAATGTC GCAACTGATC	Annealing assays; 18-nt homology with 5’ of ELB41
ssDNA	ELB40-H	AAGAGCTCCA GTTACGGATA CGGATCCTGC AAAAAAACAG TTTAAACGTA	Annealing assays; 30-nt homology with 5’ of ELB41

		CGATTATTGC CGTCAATGTC	
Cold-Trap	MW14	CAACGTCATA GACGATTACA TTGCTACATG GAGCTGTCTA GAGGATCCGA	Helicase assay
Anisotropy DNA	FAM-labelled poly(T)35	TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT TTTTT	Anisotropy ssDNA substrate
Anisotropy RNA	FAM-labelled poly(U)35	UUUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUU	Anisotropy ssRNA substrate

Table 2.7 List of commercially available 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’
HindIII	NEB	Cuts sequence: 5’-A*A G C T T-3’ 3’-T T C G A*A-5’
NdeI	NEB	Cuts sequence:

		5’-C A*T A T G-3’ 3’-G T A T*A C-5’
SalI	NEB	Cuts sequence: 5’-G*T C G A C-3’ 3’-C A G C T*G-5’
NotI	NEB	Cuts sequence: 5’-G C*G G C C G C-3’ 3’-C G C C G G*C G-5’
KpnI	NEB	Cuts sequence: 5’-G G T A C*C-3’ 3’-C*C A T G G-3”
BamHI	NEB	Cuts sequence: 5’-G*G A T C C-3’ 3’-C C T A G*G-3’
CIP	NEB	Calf intestinal alkaline phosphatase catalyzes the dephosphorylation of 3’ to 5’ ends of DNA for ligation.
T4 DNA Ligase	NEB	Used to ligate plasmids for transformation.
Q5 DNA Polymerase	NEB	Used in mutagenesis PCR polymerase enzyme.
T4 PNK	NEB	Used in KLD reaction step during site directed mutagenesis (SDM).

Dpn1	NEB	Digests methylated DNA to remove genomic contaminants.
S1 Nuclease	ThermoFisher	Degrades single stranded nucleic acids.

Table 2.8 List of primary and secondary antibodies used in this work.

Antibody	**Description**	**Use**
1º Anti-hexaHis	Mouse monoclonal, Biotin conjugate	Primary antibody for detecting hexaHis-tagged C-HelQ
1º Anti-N-HelQ	Rabbit monoclonal	Primary antibody for detecting N-HelQ
2º Anti-Biotin	HRP-linked antibody targeting Biotinylated ladder	For detecting the biotinylated ladder used as a marker
2º Anti-Mouse	Goat anti-Mouse IgA	Secondary antibody
2º Anti-Rabbit	Goat anti-Rabbit IgA	Secondary antibody

2.5 Solution Composition

Table 2.9 List of recipes of commonly used laboratory gels.

Gel Type	**Percentage**	**Composition**
Agarose Gel	1%	1 g agarose 100 mL 1X TBE

TBE Native Polyacrylamide Protein Gel	5%	29.3 mL dH_2O 6.7 mL 30% (w/v) acrylamide 4 mL 10X TBE 200 μL 10% (v/v) APS 50 μL TEMED
	10%	22.7 mL dH_2O 13.3 mL 30% (w/v) acrylamide 14 mL 10X TBE 200 μL 10% (v/v) APS 50 μL TEMED
Denaturing Urea Gel	15%	4.5 mL dH_2O 15 mL 40% (w/v) acrylamide 4 mL 10X TBE 16.8 g Urea 2 mL Formamide 100 μL 10% (v/v) APS 50 μL TEMED
SDS Separating Gel	10%	5.43 mL dH_2O 3.75 mL 30% (w/v) acrylamide 1.4 mL 3M TRIS pH 8.8 112 μL 10% (w/v) SDS 84 μL 10% (v/v) APS 9.6 μL TEMED
	12%	4.48 mL dH_2O

		4.7 mL 30% (w/v) acrylamide 1.4 mL 3M TRIS pH 8.8 112 μL 10% (w/v) SDS 84 μL 10% (v/v) APS 9.5 μL TEMED
SDS Stacking Gel	5%	1.75 mL dH_2O 3.5 mL 30% (w/v) acrylamide 0.75 mL 0.5M TRIS pH 6.8 30 μL 10% (w/v) SDS 30 μL 10% (v/v) APS 3 μL TEMED

Table 2.10 List of recipes of commonly used laboratory buffers.

Name	**Composition**
10X Tris-borate-EDTA (TBE) buffer	1 M TRIS 1 M Boric Acid 20 mM EDTA
10X SDS PAGE running buffer	250 mM TRIS 1.92 M Glycine 1% (w/v) SDS
4X SDS PAGE loading buffer	200 mM Tris-HCl pH6. 8% (w/v) SDS 0.4% (w/v) Bromophenol blue 40% (v/v) Glycerol

1X Coomassie Brilliant Blue Stain Buffer	40% (v/v) Methanol 10% (v/v) Glacial acetic acid 0.05% (w/v) Coomassie Brilliant Blue R-250
1X Coomassie Blue Destain Buffer	20% (v/v) Ethanol 10% (v/v) Acetic acid Made up to 1 L with distilled water
10X Annealing Buffer	100 mM Tris-HCl pH7.5 500 mM NaCl 10 mM EDTA
Elution Buffer	20 mM Tris pH 8.0 50 mM NaCl
5X Orange G Gel Dye	80% (v/v) Glycerol Orange G dye
5X Denaturing Gel Loading Dye	75% (v/v) Formamide 20 mM EDTA 20% (v/v) Glycerol Orange G Dye
5X Helicase Buffer (HB)	20 mM Tris-HCl pH7.5 10% (v/v) Glycerol 100µg/mL BSA
STOP Solution	2 mg/mL Proteinase K (Invitrogen) 2.5% (w/v) SDS 200 mM EDTA

Mu Broth	10 g/L Tryptone (BD) 10 g/L NaCl 5 g/L Yeast Extract (BD) adjusted to pH 7 using NaOH
Agar	3 g agar per 200 mL Mu broth
Anode Buffer	20 mL 5X Tris/CAPS 15 mL MeOH 65 mL dH_2O
Cathode Buffer	20 mL 5X Tris/CAPS 1 mL 10% (w/v) SDS 79 mL dH_2O
5X Tris/CAPS	36.34 g Tris base 44.26 g CAPS To 1 L dH_2O
10X TBS	90 g NaCl 60 g Tris To 1L with dH_2O
1X TBST	100 mL 10X TBS 200 µL Tween20 To 1 L with dH_2O

Table 2.11 List of recipes of purification buffers used in this work.

Buffer Name	**Use**	**Composition**
Nickel Charge Buffer	To charge HiTrap Chelating HP column	0.2 M Nickel (III) chloride
Ni-NTA Buffer A	HiTrap Chelating HP column wash buffer	20 mM Tris pH 8.0 150 mM NaCl 20 mM Imidazole 10% (v/v) Glycerol
Ni-NTA Buffer B	HiTrap Chelating HP column elution buffer	20 mM Tris pH 8.0 150 mM NaCl 1 M Imidazole 10% (v/v) Glycerol
Q-Sepharose Buffer A	HiTrap Q HP wash buffer	20 mM Tris pH 8.0 150 mM NaCl 10% (v/v) Glycerol
Q-Sepharose Buffer B	HiTrap Q HP elution buffer	20 mM Tris pH 8.0 1 M NaCl 10% (v/v) Glycerol
Gel Filtration Buffer	Superdex200 wash buffer	20 mM Tris pH 8.0 150 mM NaCl 20% (v/v) Glycerol
Dialysis Storage Buffer	To remove high salt; final buffer	20 mM Tris pH 8.0 150 mM NaCl 30% (v/v) Glycerol

	conditions of C-HelQ	

Table 2.12 List of columns used in purification in this work.

Chromatography Column	**Description**	**Supplier**
HiTrap Chelating HP (5mL)	Metal affinity, hexaHis tagged proteins	Cytiva
HiTrap Q HP anion exchange (1mL)	Q-sepharose for strong anion exchange	Cytiva
HiLoad 16/600 Superdex 200 pg	Gel filtration, MW range: M_r 10,000 – 600,000.	Cytiva
HiTrap Chealating HP (1mL)	Metal affinity, hexaHis tagged proteins	Cytiva

2.6 General Microbiology

2.6.1 Competent cells protocol

Escherichia coli DH5α/BL21 AI overnight cultures were used to inoculate Mu Broth in a 1:100 dilution. Cultures were grown in a Fisher Scientific shaking water bath in conical flasks at 37°C until OD_{600} 0.5 was reached. Cells were pelleted by centrifugation for 5 minutes at 4000 rpm using a centrifuge 5430R (Eppendorf) in falcon conical tubes, and the

supernatant discarded. The cell pellets were then resuspended in 1/8th volume chilled 0.1M $CaCl_2$ and left on ice for 1 hour before being pelleted again. Cells were resuspended in 1/8th volume chilled 0.1M $CaCl_2$ with glycerol added to 30% (v/v). 100μL aliquots were flash frozen with dry ice and stored at -80°C.

2.6.2 Competent cell transformation protocol

Competent cells were thawed on ice prior to 5 ng of vector plasmid being added per 100 μL of competent cells and mixed by gentle pipetting. Solutions were kept on ice for 30 minutes before being heat shocked at 42°C for 2 minutes and placed back in ice. Mu broth was added under sterile conditions and incubated in a shaking water bath at 37°C for 1 hour. Cells were pelleted by centrifugation for 1 minute at max speed and concentrated by resuspension in 1/2 volume Mu broth. The solution was then evenly distributed onto Mu broth ampicillin agar plates (50 μg/mL ampicillin) using spread plate technique. Plates were incubated overnight at 37°C.

2.6.3 Site Directed Mutagenesis

Conserved residues of interest located on C-HelQ were targeted for mutagenesis. Primers were designed using NEBaseChanger. Mutagenesis was performed using a makeshift Q5 site directed mutagenesis kit. In summary, pHB01 containing WT C-HelQ was subject to PCR with Q5 polymerase. PCR samples were assessed on a 1% (w/v) agarose gel in 1X TBE buffer to determine successful amplification. PCR

sample was subject to a KLD reaction with T4 PNK, T4 DNA Ligase and Dpn1 for 1 hour at room temperature. The resulting reaction was transformed in competent *E. coli* DH5α cells overnight using the transformation protocol previously described. Colonies were picked and plasmids were purified using the above method before being sent for sequencing by GENEWIZ from Azenta Life Science.

2.6.4 Plasmid cloning.

2.6.4.1 PCR

Forward and reverse primers for were designed to amplify a target gene of interest.

PCR reactions of 25 µL were assembled containing 5X reaction buffer, 10 mM dNTPs, 10 µM forward primer, 10 µM reverse primer, <1000 ng DNA template, 0.25 µL Q5 DNA Polymerase, 5X GC Enhancer and dH_2O.

Table 2.13 PCR Thermocycling Conditions

Cycle Step	Temperature	Time	Number of Cycles
Initial denaturing	98°C	30 seconds	x 1
Denaturing	98°C	10 seconds	x 30

Annealing	Variable (dependent on primer Tm)	30 seconds	
Extension	72°C	30 seconds/kb	
Final Extension	72°C	2 minutes	x 1

PCR products were confirmed by running a sample of the product on a 1% (w/v) agarose gel with NEB 1 kb ladder used as a marker. PCR products were purified using Promega PCR Clean-Up Kit.

2.6.4.2 DNA vector and insert ligation

The DNA vector backbone and purified PCR product were digested with appropriate restriction enzymes in 25 µL reactions containing 1X rCutSmart Buffer (NEB), 10 µL vector backbone, 0.5 µL each restriction enzyme, and additionally 1 µL CIP included in for the digestion of the vector backbone for de-phosphorylation. Digests were incubated at 37°C for.1 hour and following this ran on a 1% (w/v) agarose gel, bands excised in a dark room using a transilluminator and gels purified with a Promega Gel Extraction Kit.

Ligation reactions were calculated using NEBioCalculator with reactions having a vector:insert ratio of 1:3; with 40 ng vector used. Reactions additionally contained 1X T4 DNA Ligase buffer (NEB), 1 µL T4 DNA Ligase (NEB) and dH_2O. Reactions were ligated at room temperature for 1 hour, transformed in DH5α competent cells using the standard protocol

described earlier and plated onto LB plates with the appropriate antibiotic. Colonies were picked, miniprepped and digested to be run on a 1% (w/v) agarose gel to confirm band size. Successful ligations were verified with sequencing at Source Biosciences.

2.7 Gel Electrophoresis

2.7.1 Agarose gel electrophoresis

DNA samples, with 6X Purple DNA Loading Dye (NEB), were loaded onto a 1% (w/v) agarose gel stained with 0.2 µg/mL ethidium bromide (EtBr) to visualize. Electrophoresis was carried out using a BioRad Cell Electrophoresis tank and PowerPack, ran in 1X TBE, at 120 V, for 1 hour. DNA was visualized using Bio-imaging system (Syngene) with UV exposure.

2.7.2 SDS PAGE analysis

SDS PAGE was used in this work to identify the presence on proteins in over-expressions and purifications. Samples were added to 1X SDS Loading Buffer and 30 mM dithiothreitol (DTT) and boiled for 10 minutes at 95°C on a heat block. The SDS-PAGE gel comprised of a lower separating gel of suitable polyacrylamide percentage (based on the proteins molecular weight – high molecular weight required a lower percentage gel) and an upper stacking gel were used to run the samples on. 8 µL samples were run typically on a 10- or 12% (w/v) SDS gel, alongside blue pre-stained protein standard, at 140 V for 60 minutes to

allow full migration. Proteins were visualized using Coomassie Brilliant Blue stain for 30 minutes and de-stained with ‘Destain Buffer’.

2.7.3 Western blotting

Protein detection using protein immuno-blotting followed SDS-PAGE analysis when overexpression of proteins could not be detected with Coomassie blue staining. SDS-PAGE was carried out with the addition of a biotinylated ladder marker. Proteins were transferred from the SDS-PAGE gel to a PVDF membrane (GE Healthcare); first the PVDF membrane was soaked in 100% MeOH and the equilibrated in anode buffer for 30 minutes. Meanwhile, the SDS-PAGE gel was equilibrated in cathode buffer. A piece of blotting paper was soaked in each of the buffers too. The dry transfer western blotting system was assembled; platinum anode, blotting paper soaked in anode buffer, PVDF membrane, SDS gel, blotting paper soaked in cathode buffer and cathode assembly (bottom to top).

Electrophoresis took place at 120 mA (for a 8X10 cm gel) for 1 hour. The membrane was then blocked with 20 mL blocking buffer (TBST and 3% (w/v) milk powder) with gentle agitation for 1 hour at room temperature. Blocking buffer was then changed and supplemented with appropriate antibody, at an appropriate concentration; and left overnight with gentle agitation at 4°C. Five 4-minute washes with fresh ice-cold TBST were caried out at room temperature before being left at room temperature for 1 hour with 20 mL fresh blocking buffer and appropriate secondary

antibody. The blot was then washed with five 4-minute washes with ice-cold TBST, at room temperature. ECL Western Blotting Substrate (Promega) was prepared: 3 mL of Luminol Enhancer solution was mixed with 3 mL Peroxide solution and immediately pipetted onto of the blot. After 1-minute, excess solution was blotted off and the PVDF membrane visualised using FUJIFILM Lass-3000 Mini with exposures of 5, 10 and 30 minutes detecting the chemiluminescent signal.

2.8 Protein Over-Expression and Purification

2.8.1 Obtaining C-HelQ and Mutants

Wild type *Homo sapiens* C-HelQ (pHB001) was previously amplified using PCR and cloned into pACYC-duet background using SalI and NotI restriction sites by Hannah Betts. Mutant pHB001 plasmids were made by Q5 SDM using the standard protocol described above; giving pTJ015 (Tyr-642-Ala C-HelQ point mutant, work by Tabitha Jenkins) and pALH009 (Pro-960-Ala, Tyr-963-Ala C-HelQ point mutants, this work). Plasmids was transformed into competent *E. coli* BL21 AI cells, using the method previously described, for overexpression.

2.8.2 Overexpression of H. sapiens C-HelQ and Mutants

Transformed colonies were subject to a pilot protein overexpression. Single colonies were grown overnight in 5 mL Mu broth with 35 µg/mL chloramphenicol at 37°C in a test tube rotator. Mu broth containing 35 µg/mL chloramphenicol was inoculated to a 1:100 dilution using the overnight culture. Cells were grown to OD_{600} 0.6 37°C before being

induced with 0.2% (w/v) L-arabinose and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cell growth was continued for a further 3 hours at 37°C. Cells from 1 mL samples were pelleted through centrifugation and resuspended in 150 µL SDW. Samples were added to 4X SDS Loading dye and 30 mM DTT, heated at 95°C for 10 minutes and ran on a 10% (w/v) SDS PAGE gel to confirm expression or proteins and size of protein. Once successful overexpression was confirmed via SDS PAGE, an up-scale was performed to obtain sufficient cell biomass for protein purification. After overexpression, the culture was centrifuged using an Avanti J-26 XP centrifuge (Beckman Coulter) using a JLA 10.500 rotor for 15 minutes at 4000 rpm. Cell pellets were resuspended in 15 mL of biomass storage buffer and flash frozen for storage at -80°C.

2.8.3 Purification of C-HelQ and Mutants

To purify, frozen biomass was first thawed on ice, sonicated 5 mL at a time in 30 second intervals for a total of 1 minute 30 seconds, and clarified by centrifugation using an Avanti J-26 XP centrifuge (Beckman Coulter) with a JA 25.50 rotor at 16,500 rpm for 30 minutes. The supernatant was decanted and kept on ice to be loaded onto a 5 mL HiTrap Chelating HP column (GE Healthcare Life Sciences) which has been pre-equilibrated with ‘Ni-NTA Buffer A’. The flow through (eluate from loading the supernatant) and wash through (eluate containing weakly bound protein, washed with ‘Ni-NTA Buffer A’ were collected for analysis. C-HelQ was eluted over a gradient of increasing ionic strength, achieved by increasing the ratio of Ni^{2+}-NTA Buffer B:Ni^{2+}-NTA Buffer A.

Fractions eluted which contained C-HelQ, determined by UV absorption on the chromatographic trace graph and SDS-PAGE analysis, were pooled together to be loaded on to a 1 mL HiTrap Q HP anion exchange column pre-equilibrated with 'Q Sepharose Buffer A'. The flow through/wash through were collected as before and C-HelQ eluted over a gradient of increasing ionic strength, achieved by increasing the ratio of Q-Sepharose Buffer B:Q-Sepharose BufferA over the fractions eluted. As before, C-HelQ containing fractions confirmed by SDS PAGE analysis were pooled together to be loaded onto a HiLoad 16/600 Superdex 200 pg Gel Filtration column. C-HelQ was eluted over 2 hours with 'Gel Filtration Buffer', samples believed to contain pure C-HelQ were analyzed on a SDS PAGE gel and pooled together as before. The protein was then concentrated on a 1 mL HiTrap Chelating HP column before C-HelQ containing fractions being pooled together and dialyzed with dialysis storage buffer. Dialyzed purified protein was then aliquoted, and flash frozen on dry ice for storage at -80°C. A sample of purified protein was tested using a NanoDrop 2000 spectrophotometer to give a 280 nm absorption value. This value and the extinction coefficient of C-HelQ (83,630) were applied to the Beer-Lambert Law ($A = \varepsilon cl$) to calculate the protein concentration. The same purification method was used to purify both mutants of C-HelQ.

2.8.4 Analytical Gel Filtration

Analytical gel filtration (AGF) is a form of size exclusion chromatography used during protein purification to determine the oligomeric state of the

protein. A Superdex200 column was loaded onto a Cytiva ÄKTA Start, washed with dH_2O and Gel Filtration Buffer, and calibrated with BioRad's gel filtration protein standard comprised of the following: Thyroglobulin (bovine), MW. 670,000 Da; γ-globulin (bovine), MW. 158,000 Da; Ovalbumin (chicken), MW. 44,000 Da; Myoglobin (horse), MW. 17,000 Da; and Vitamin B12, MW. 1,350 Da. A standard curve of known molecular weights against elution volume was produced using PRISM 7 (GraphPad).

2.9 *In vitro* Experimentation

2.9.1 Preparation of DNA substrates

Double stranded DNA substrates listed in Table 2.6 were prepared using a 1.2:1 ratio of unlabeled to Cy-5 labelled oligos. 50 μL reactions were set up with 5 μM labelled and 6 μM unlabeled oligo with 1X Annealing Buffer. Substrates were heated to 95°C on a heat block for 10 minutes and cooled to room temperature overnight. 5X OG dye was added to each sample, which was then loaded onto a 10% (w/v) polyacrylamide TBE gel and run at 140 volts for 3 hours. Substrates were visualized by eye, cut out with a scalpel and DNA was eluted from the gel by diffusion at 4°C over 48 hours in 250 μL Elution Buffer. DNA was quantified using a NanoDrop 2000 spectrophotometer to take 260nm readings to apply to the Beer-Lambert Law, with the corresponding substrate extinction coefficient to calculate μM of DNA.

2.9.2 DNA Annealing Assays

DNA annealing activity of Hel308 family proteins were analyzed by detecting the association of a Cy-5 ssDNA oligonucleotide and an unlabeled ssDNA oligonucleotide; ssDNA oligos migrate further than dsDNA substrates in polyacrylamide TBE gels. 20 μL reactions were set up with 15 nM Cy5-labeled ssDNA substrate, 1X Helicase Buffer, 25 mM DTT and dH_2O, pre-incubated for 2 minutes on ice with diluted protein in the range of 25-800 nM, made as serial dilutions, and 15 nM complementary ssDNA substrate to initiate the reaction. Annealing reactions were incubated at 37°C for 5 minutes prior to quenching with the addition of STOP buffer. A pre-annealed dsDNA substrate was used as a control in addition to a no protein control. In some instances (particularly with substrates with smaller micro-homologies) spontaneous annealing was observed in the no protein control. An additional 'Pool" sample containing only the reaction pool (no protein and no complementary ssDNA) was included as an additional control. Reactions were loaded onto a 10% (w/v) polyacrylamide TBE gel with Orange G loading dye and run in 1X TBE buffer at 140 volts for 1 hour. Gels were imaged using an Amersham Typhoon phosphor-imager.

2.9.3 DNA Helicase Assays

DNA unwinding activity of Hel308 family proteins were analyzed by detecting the dissociation of a Cy5-labeled ssDNA substrate from a pre-annealed substrate; ssDNA oligos migrate further than dsDNA substrates in TBE gels. 20 μL reactions were set up with 1X HB, 5 mM

$MgCl_2$, 5 mM ATP pH 8.0, 25 mM DTT, 2.5 M cold-trap unlabeled substrate and 25 nM Cy5-fluorescently labelled dsDNA substrate, and the addition of diluted protein in the range of 25-800 nM, made as serial dilutions, which initiated the reaction. The need for a cold-trap, an unlabeled strand of the same sequence as the Cy5-labelled strand, was identified to be required to detect unwinding. Helicase reactions were incubated at 37°C for 10 minutes prior to quenching with addition of STOP buffer. Reactions were loaded onto a 10% (w/v) polyacrylamide TBE gel with Orange G loading dye and run in 1X TBE buffer at 150 volts for 45 minutes. Gels were imaged as previously described.

2.9.4 Electromobility Shift Assay (EMSA)

EMSAs are a biochemical method of detecting and studying protein-DNA interactions. This method detected stable protein-DNA bound complexes via monitoring the migration of fluorescently labelled DNA substrates. 20 μL reactions binding reactions are set up with 1X HB, 25 mM DTT, 25 nM fluorescently labelled DNA substrate and diluted protein within the range of 25-800 nM, made as serial dilutions. Samples were run a 5% (w/v) polyacrylamide TBE gel with Orange G loading dye at 140 volts for 120 minutes. Gels were imaged as previously described.

2.9.5 S1 Nuclease Protection Assay

S1 nuclease protection assays were used to determine the effect of C-HelQ in preventing ssDNA degradation by S1 nuclease; and to see the effect, if any, of the C-HelQ^{Y642A} mutation on exposure of ssDNA for S1

nuclease digestion. 20 μL reactions were set up containing 25 nM Cy5-labeled ssDNA, 25 mM DTT and 1X Helicase Buffer. HelQ/C-HelQ protein was added and incubated on ice for 5 minutes. 10 units of S1 Nuclease (NEB) was added, with S1 Nuclease Buffer (NEB), and incubated at 37°C for 20 minutes – time and nuclease concentration determined through optimisation of the assay. The reaction was quenched and run on a 15% (w/v) denaturing urea gel at 100 W for 180 minutes, with denaturing gel loading dye, to visualise.

2.9.6 Biophysical Methods

2.9.6.1 Fluorescence Resonance Energy Transfer (FRET) Assays

DNA annealing activity of Hel308 family proteins were analyzed using FRET techniques, detecting the association of a Cy5-labeled ssDNA substrate and a complementary Cy3-labeled ssDNA substrate using the BMG FLUOstar Omega Benchtop Plate Reader. 50 μL reactions were set up with 1X HB, 5 mM DTT and 50 nM Cy5-labeled ssDNA substrate. Diluted protein in the range of 25-800 nM, made as serial dilutions, were added to each pool sample aliquoted in a 96-well plate and incubated for 2 minutes at 37°C. The addition of 50 nM Cy3-labeled complementary ssDNA initiated the reaction. Reactions were incubated in the plate reader at 37°C and a 30-minute time-course assay was conducted, taking an emission reading at 670 nm and 590 nm (corresponding to the emission spectrum of the Cy3 and Cy5 dyes) every minute.

2.9.6.2 Anisotropy

Anisotropy was used to determine protein binding affinities on fluorescein (FAM) labelled ssDNA or ssRNA substrates. 70 μL reactions were set up with 1X Helicase Buffer, 5 mM DTT and 50 nM single stranded substrate. Reaction pool was aliquoted into a 96-well plate and upon the addition of protein serial dilutions, the binding reaction was initiated. Protein interactions were determined using the BMG FLUOstar Omega Benchtop Plate Reader. Anisotropy values were taken with excitation at 480 nm and emission at 560 nm, corresponding to the absorption and emission spectrum of the dye. Three readings were taken at 0 minutes (initial binding), 5 minutes and 10 minutes, incubated at 37°C. FLUOstar Omega software converted emission values into R-values, blanked with a no protein reaction, and were then plotted and analysed using PRISM GraphPad.

2.10 Bioinformatic Analysis and Molecular Modelling

A range of open source bioinformatic resources and tools were used throughout this work to retrieve DNA and protein sequences, carry out homology searches and sequence alignments, and predict the protein structure of unresolved proteins.

2.10.1 Sequence Mining

The Universal Protein Resource (UniProt) database (https://www.uniprot.org) was used to extract amino acid sequences in FASTA format, for compatibility with other bioinformatic tools. The online

resource ExPASy ProtPram (https://web.expasy.org/protparam/) was used to generate physico-chemical properties based on FASTA sequences, including extinction coefficients and molecular weights.

2.10.2 Homology Search

Protein sequence alignments were generated from FASTA sequences on ClustalOmega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo); identifying sequence homology of protein sequences between different proteins and different species.

2.10.3 Structure Prediction

Predicted 3D models of proteins were achieved by inserting the amino acid sequence of the protein into Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2). Files generated, and PBD files were extracted and inserted into the modelling software PyMOL (https://pymol.org/2/) to annotate the protein, producing high quality images.

Chapter 3 Biochemical analysis of *H. sapiens* C-HelQ and mutants

3.1 Introduction

The C-terminus of HelQ (C-HelQ) is the predicted 'core-helicase' of the protein, and reliable for the proteins' translocase and helicase activity; and is defined by the residue boundaries G275 to A1101. Having strong sequence homology to its archaeal homologue Hel308, C-HelQ comprises the functionally conserved helicase domains containing the Walker A and Walker B motifs, a winged-helix domain (WHD) and domain IV with a DNA ratchet. This therefore separates the predicted intrinsically disordered N-terminal region of full-length HelQ; with C-HelQ lacking the 274 most N-terminal residues of full-length HelQ – this missing region having thought to be predominantly an IDPR (Figure 3.1).

Recently, specific single-amino acid substitutions in archaeal Hel308 have shown to modulate its DNA processing activity, which will be described and shown within Chapter 4 [68]. Sequence alignment between Hel308 and *H. sapiens* C-HelQ have identified corresponding residues of interest in the human protein that produce a modulated phenotype in the archaeal protein. These residues are yet to be explored to assess the extent of a functional change, if any, of mutagenized C-HelQ protein.

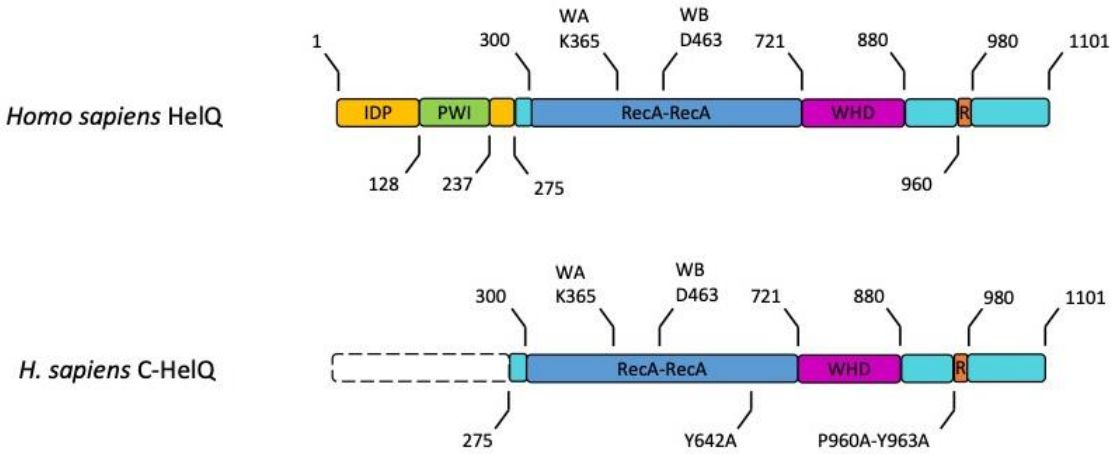
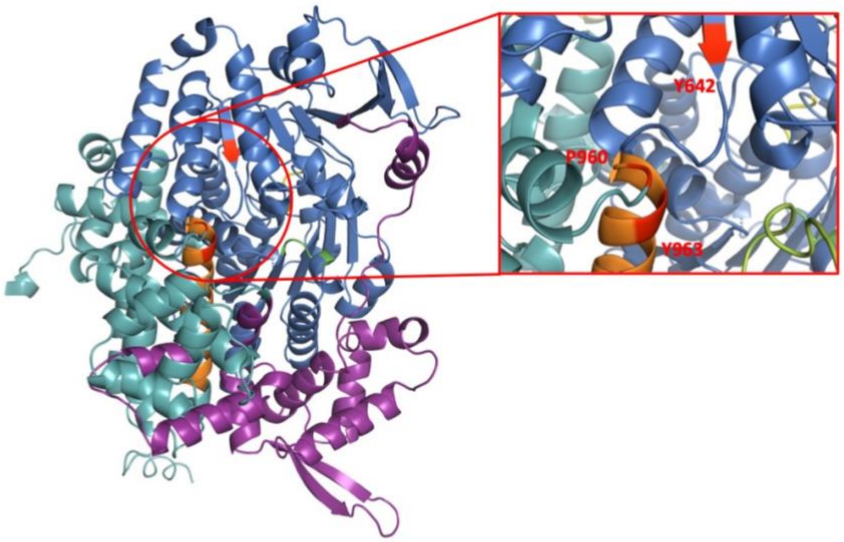


Figure 3.1 Domain map comparison between HelQ and C-HelQ. C-HelQ an 826 amino acid protein that lacks the first 274 amino acids of full-length HelQ. This missing region is thought to predominantly be an intrinsically disordered protein region (IDPR), despite containing a PWI-like fold which was found to have a single strong sequence match to a PWI-fold in the crystal structure of the yeast Ski-2 family helicase Brr2 [64]. Residues Y642 (work by Tabitha Jenkins) located within the dual RecA-domain, and P960 and Y963 (this work), located within the DNA ratchet, were targeted for mutagenesis and subsequent biochemical analysis.

The Y642 residue on C-HelQ, located within the RecA-RecA domain, was found to align with the F295 residue on *M. thermautotrophicus* Hel308; which when mutated to an alanine was found to result in a hyper-annealing phenotype. It was therefore of interest to see if a likewise mutation in C-HelQ would result in a similar change in activity. For this, a C-HelQ^{Y642A} mutant was cloned (work by Tabitha Jenkins). Similarly, the P960 residue on C-HelQ, located within the DNA ratchet of Domain IV, was found to align with the Y586 on *M. thermautotrophicus* Hel308; which when mutagenized to an alanine was found to result in a hyperactive DNA unwinding. It was however also noted that a tyrosine residue to be located three residues downstream in C-HelQ, Y963. Thus, it was decided for a double mutant, C-HelQ$^{P660A-Y963A}$ to be cloned (this work).

A



B

HsaHelQ	638-FGVAYHHSGLT-648	955-EKFNMPRGYIQ-965
MthHel308	291-AGIAFHHAGLF-301	581-NAFHVYAASTR-591

Figure 3.2 C-HelQ residues of interest for mutagenesis. C-HelQ predicted structure (A) shows residues mutated in this work. Sequence alignments (B) identified the aligned residues of interest from archaeal Hel308, found to modulate DNA binding, unwinding and annealing, in *H. sapiens* HelQ. Y642, located within the dual Rec-A domain, P960 and Y963, located within the DNA ratchet, were targeted for a single amino acid substitution for alanine, creating mutants C-HelQ^{Y642A} and C-HelQ$^{P960A\text{-}Y963A}$.

3.2 Overexpression and purification of WT *H. sapiens* C-HelQ and mutants

3.2.1 Cloning

Wild-type C-HelQ was cloned into a pACYC-duet backbone for expression with a C-terminal hexaHis-Tag in *E. coli,* pHB001 (work by Hannah Betts). Mutagenesis was carried out on C-HelQ to introduce site-specific point mutations. pHB001 has been used as the backbone for mutagenesis for C-HelQ mutants, such as C-HelQ^{Y642A} which has been successfully cloned prior to this project (work by Tabitha Jenkins).

Following a sequence alignment between *H. sapiens* C-HelQ and *M. thermautotrophicus* Hel308, residues P960 and Y963 on C-HelQ were identified to be of interest. Primers were designed on NEB Base Changer/NEB Builder to mutagenize both residues each to an alanine, resulting in a C-HelQ$^{P960A\text{-}Y960A}$ double mutant (see Table 2.5). Successful site-directed mutagenesis (SDM) of these residues produced the mutant: in short, the entire plasmid template was amplified via PCR amplification using Q5 polymerase, introducing the mutations. This was then followed with a KLD reaction for 1 hour at room temperature involving treatment with T4 PNK, T4 DNA Ligase and Dpn1 allowing for phosphorylation, ligation and template removal. The vector was then transformed into *E. coli* DH5α competent cells and successful mutagenesis was confirmed with a double restriction digest with SalI and NotI, and via sanger sequencing.

3.2.2 Overexpression

Vector constructs containing C-HelQ WT (pHB001), C-HelQ^{Y642A} (pTJ015) and C-HelQ$^{P960A\text{-}Y963A}$ (pALH009) were transformed into *E. coli* BL21AI competent cells (which are derived from *E. coli* BL21 cells) for overexpression. The addition of L-arabinose allowed controlled T7 RNA polymerase expression through the *ara*Bad promoter, and addition of IPTG results in removal of the *lac* repressor from the T7 promoter, allowing access by the T7 RNA polymerase and mitigating the repressors inhibitory effect. Cell biomass was obtained as described in 2.8.2: in summary, cultures were grown in baffled flasks to OD_{600} 0.6 at 37°C, induced with 0.2% (w/v) L-arabinose and 1 mM IPTG, and left to shake for an additional 3 hours at 37°C.

Protein overexpression of C-HelQ constructs were found to be problematic due to high levels of protein degradation, suggested to result from the core helicase region of HelQ being responsible for HelQ instability.

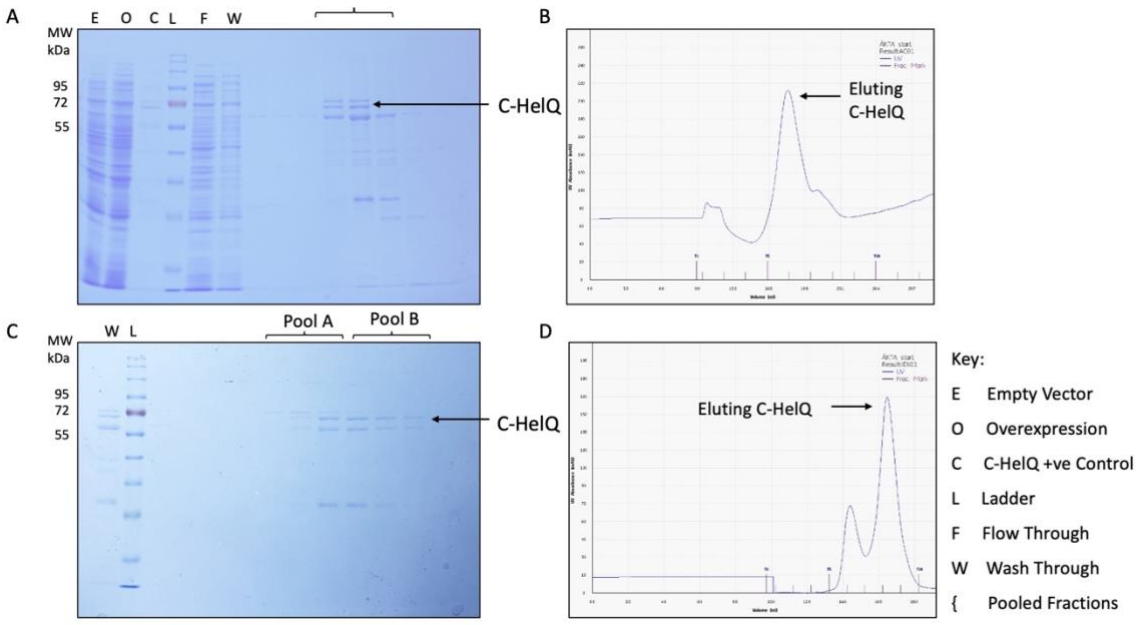


Figure 3.3 Purification of WT C-HelQ using Ni^{2+}-NTA affinity column and Q-Sepharose ion exchange column. Coomassie stained 10% (w/v) acrylamide SDS-PAGE analysis of fractions eluted off the (A) Ni^{2+}-NTA affinity column and (C) Q-Sepharose column. C-HelQ at ~72 kDa indicated with arrow. Specified fractions eluted from the Ni^{2+}-NTA were pooled together for purification by Q-Sepharose ion exchange column. Following subsequent purification, two pools were made and dialysed for activity analysis. Corresponding chromatographic trace from AKTA Start shows elution of protein by change in UV absorption. Increasing absorbance (peaks) indicates proteins from (B) Ni^{2+}-NTA column and (D) Q-Sepharose columns being eluted off, and potential fractions of interest.

Chromatography traces, reliant on the detection of UV absorption of aromatic residues, allowed the identification of C-HelQ eluting off each column. Fractions believed to contain C-HelQ were confirmed by SDS-PAGE analysis and Coomassie blue staining using a previously purified positive control C-HelQ sample. C-HelQ cloned into the plasmid pHB01 starts at position Asn-276 and has a molecular weight of approximately 96 kDa, inclusive of its tag. Coomassie stained gels from SDS-PAGE showed C-HelQ to migrate at ~72 kDa, a size

smaller than that it is predicted to be – owing to C-HelQ having properties that influence migration of the protein in a gel. This has been previously reported [69] in and where it was confirmed to be C-HelQ by mass spectrometry.

As the positive control (C-HelQ previously purified in the lab) contained a couple bands close to the expected C-HelQ band size (approximately 72 kDa), and as C-HelQ is known to be unstable, possibly resulting in degraded versions of the protein through the purification process, two pools of suspected C-HelQ containing fractions that came off the Q-Sepharose column were made, one primarily containing the topmost band and the other containing the second band. Each pool was subsequently dialysed, flash frozen and stored in at -80°C. C-HelQ^{Y642A} was also initially purified following this same method.

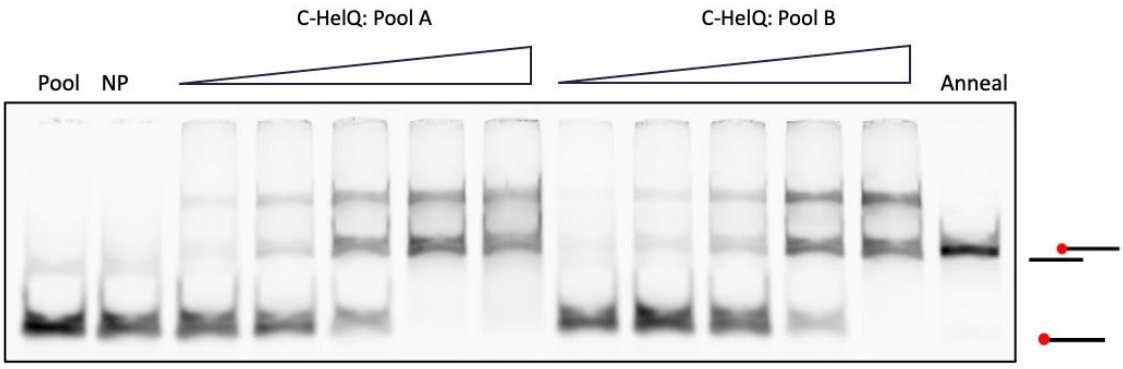


Figure 3.4 C-HelQ annealing activity test. Initial activity test, ran on 10% (w/v) polyacrylamide TBE gel for 60 minutes at 140 V, of both WT C-HelQ pools to determine annealing activity on ssDNA substrates with 18-nt micro-homology. Concentration titrations, 25, 50, 100, 200 and 400 nM protein, were performed at 37°C for 5 minutes, in buffer independent of ATP.

After dialysis both wild-type C-HelQ pools were tested for annealing activity prior to the study. Both Pool A and Pool B of C-HelQ were shown to anneal ssDNA with 18 base pairs of complementarities at their 5'-ends. Pool A demonstrated greater annealing activity than Pool B, fully annealing all DNA at 200 nM of C-HelQ compared to 400 nM with Pool B. Interestingly, the fully annealed DNA control identified both proteins were also annealing a secondary product of greater molecular weight – this was theorized to be a result of either a contaminant protein and/or the substrate forming another structure.

Following this first attempt at purification, it was decided additional steps would be taken in subsequent purifications to remove any contaminants and degraded forms of C-HelQ in an attempt to negate their effect on protein activity.

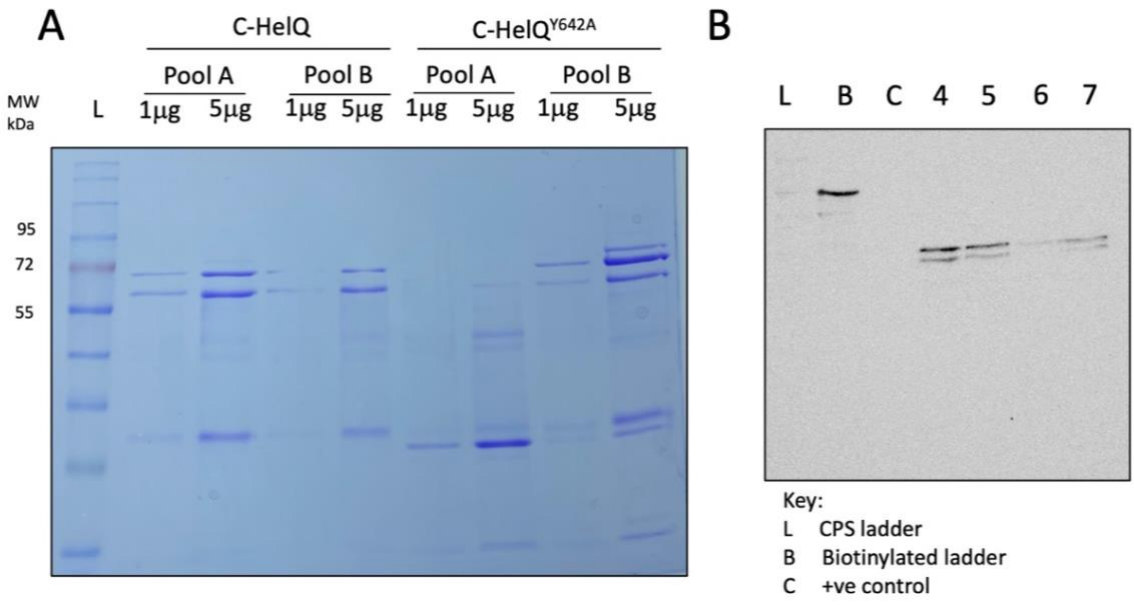


Figure 3.5 SDS-Page and Western Blot. (A) Coomassie stained 10% (w/v) acrylamide SDS-PAGE analysis of C-HelQ and C-HelQ^{Y642A} pooled proteins, each at 1 µg and 5 µg. (B) Western blot analysis of proteins. Lane 1: colour protein standard ladder; Lane 2: biotinylated ladder;

Lane 3: Control C-HelQ; Lane 4: C-HelQ Pool A; Lane 5: C-HelQ Pool B; Lane 6: C-HelQ^{Y642A} Pool A; Lane 7: C-HelQ^{Y642A} Pool B – 5 μg of each protein loaded.

SDS-PAGE and western blot analysis identified two proteins migrating to a size of ~60 kDa and ~70 kDa to be His-tagged; these predicted to be the second and third topmost bands on the SDS-PAGE gel. While the size of these bands could not be confirmed on the western blot due to recurrent lab issues with the biotinylated ladder and an issues with the C-HelQ positive protein control aliquot (previously purified in the lab) on the gel, the size difference between the third band, and a protein product of lower molecular weight shown on the SDS-PAGE gel suggest they are two ends of a cropped C-HelQ which would result in a protein product of the molecular weight of the second band. It can also be deduced that the topmost band of the SDS-PAGE gel to be a contaminant as the anti-His antibody did not detect it. To confirm this, proteins were separated by size-exclusion chromatography and pooled separately, in order to isolate the top two bands, to determine their activity.

Following elution off the 1 mL Q-Sepharose column, all fractions containing C-HelQ were pooled together and directly loaded onto a 120 mL S200 Gel Filtration column with the aim to separate the two bands that appear to be ~72 kDa in size on the SDS-PAGE gel, suspected to be a protein contaminant and C-HelQ. Fractions containing pure, isolated protein were then pooled together (with aid from the UV absorption chromatographic trace graphs as protein SDS-PAGE does not appear very well due to small amounts of protein being eluted off over multiple fractions), and loaded onto a 1 mL Ni^{2+}-NTA column for final concentration before being dialysed and stored at -80°C. The approximate

concentration was determined using the Beer-Lambert law with A260 absorption reading value obtained NanoDrop and C-HelQ's extinction coefficient value (83755).

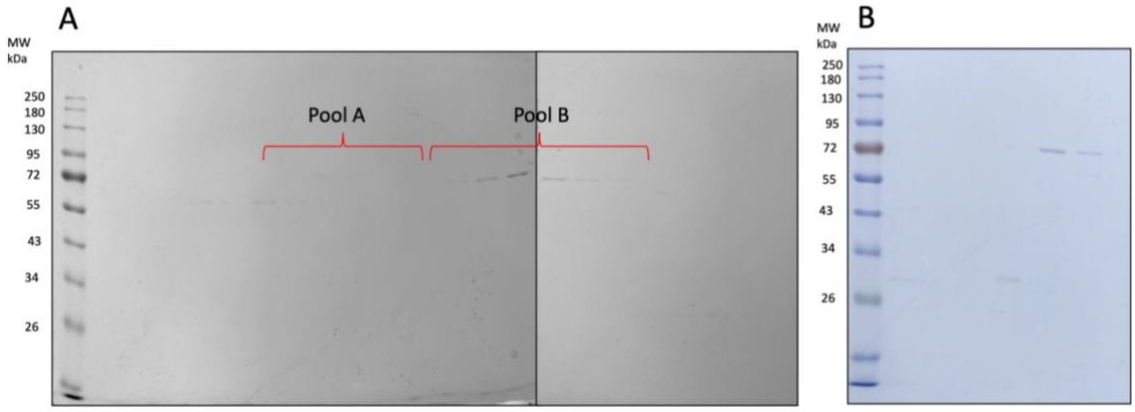


Figure 3.6 Purification of WT C-HelQ using gel filtration column and Ni^{2+}-NTA column. Coomassie stained 10% (w/v) acrylamide SDS-PAGE analysis of fractions eluted off the (A) S200 gel filtration chromatography column, eluting proteins off by size exclusion. Two pools were made, isolating the top and second band from the previous SDS-PAGE gel; predicted to be a contaminant and C-HelQ. Both pools were then run individually through 1 mL Ni^{2+}-NTA affinity column to concentrate the protein. (B) shows Pool B.

Following isolation of the ~72 kDa protein (topmost band) from the middle band (the protein suspected to be "full length" C-HelQ) by size-exclusion chromatography, and remaining purification procedure, the two pools were tested for unwinding activity prior to the study.

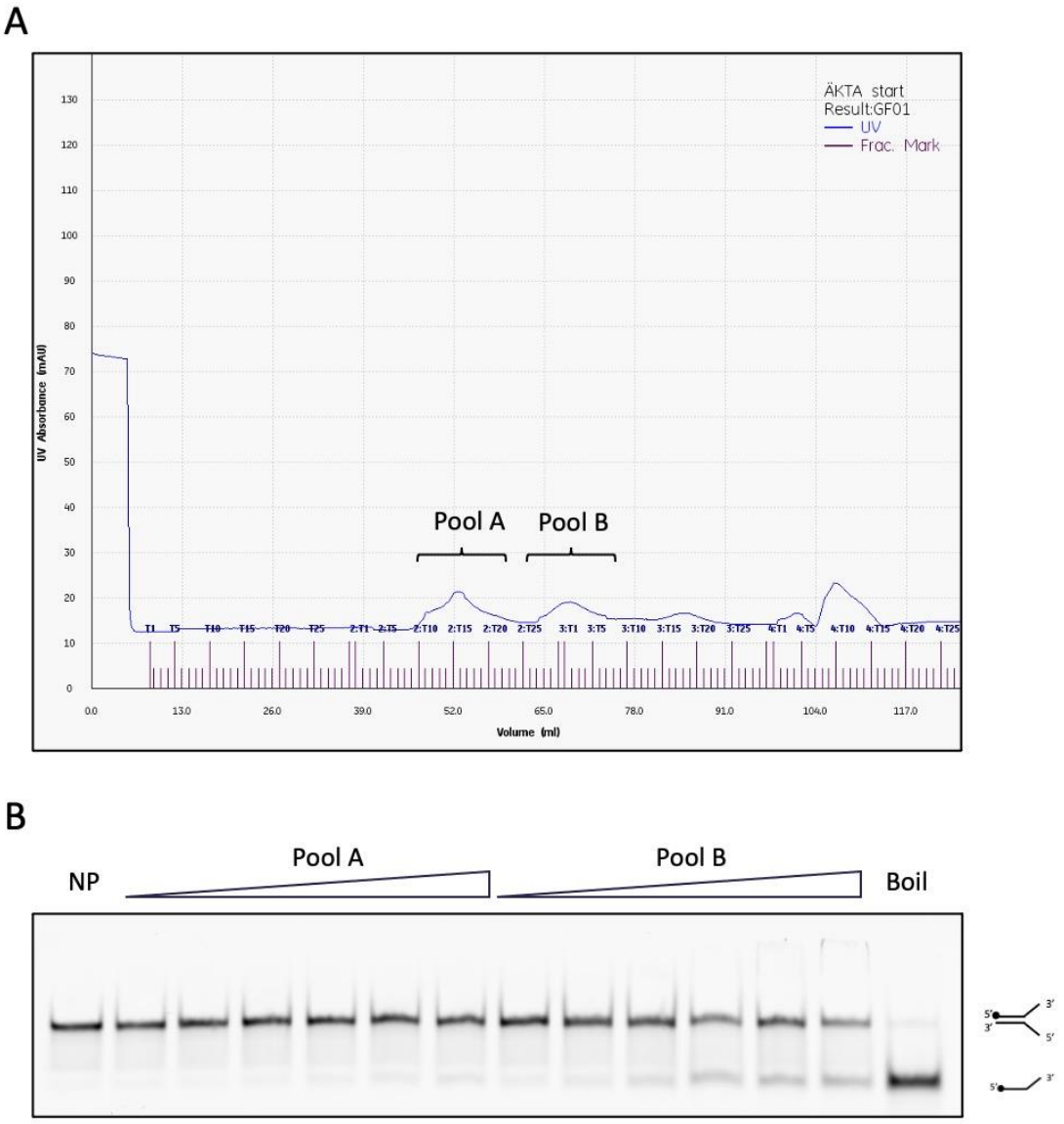


Figure 3.7 C-HelQ unwinding activity test. Initial activity of both C-HelQ pools, containing a different protein band as seen on previous SDS-PAGE gels, the top band and second band. Concentration titration reactions were performed at 37°C for 10 minutes. Pool A, the top band, shows no unwinding whereas Pool B shows increased unwinding with increased concentration of C-HelQ. This confirms Pool A is not C-HelQ and can be assumed to be a contaminant protein based on its size being greater than the second band, Pool B, which can be deduced to be C-HelQ.

As seen in Figure 3.7 by the chromatographic traces and unwinding assay, the first protein to be eluted off the size-exclusion chromatography column has no unwinding activity whereas the second protein to be eluted off does (these correlates to the top and second band seen on previous SDS-PAGE gels). This confirmed that the ~72 kDa protein (top band) is a contaminant and not C-HelQ, confirming the results of the western blot seen in Figure 3.5. Like full-length HelQ, C-HelQ is known, and expected, to unwind forked DNA substrates.

To summarise the purification process, C-HelQ is purified with first a 5 mL Ni^{2+}-NTA affinity column allowing the elution of his-tagged proteins, then with a 1mL Q-Sepharose ion exchange column allowing concentration of proteins. Following this, size exclusion chromatography with the S200 gel filtration column allows the separation of the contaminant protein and other degraded forms of the protein from “full-length” C-HelQ. This is followed with concentration on a 1 mL Ni^{2+}-NTA affinity column and dialysis. C-HelQ mutant proteins; ‘Y642A’ and ‘P960A-Y963A’ were also purified using the same method detailed above.

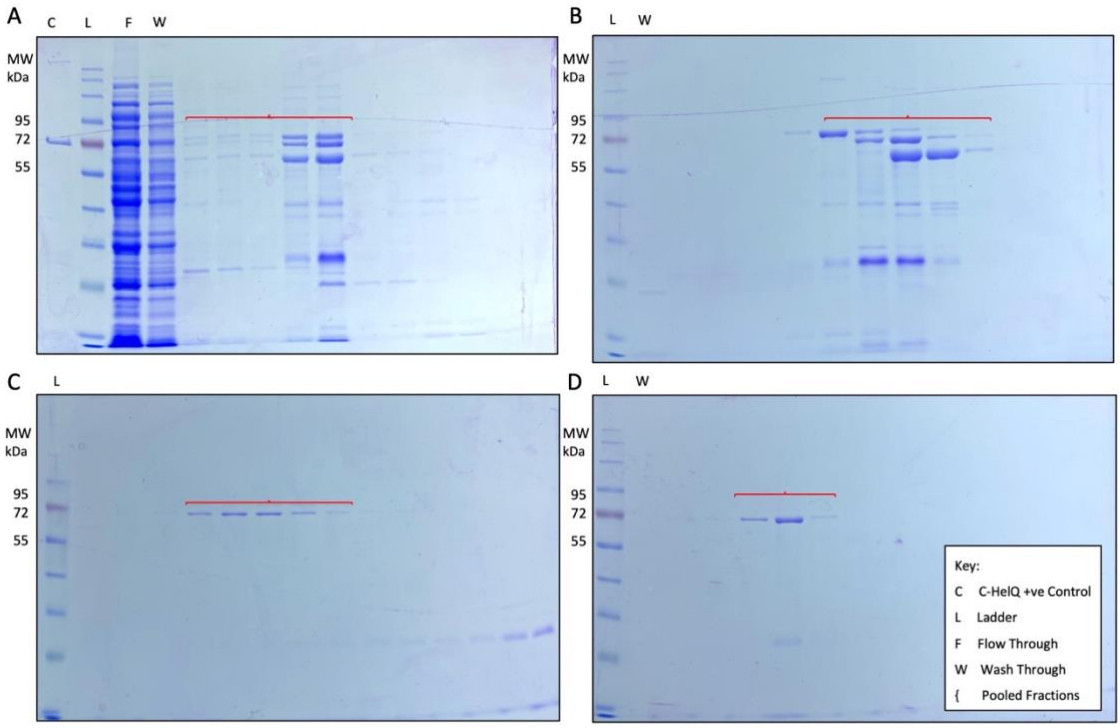


Figure 3.8 Purification of C-HelQ^{Y642A} protein. Coomassie stained 10% (w/v) SDS-PAGE analysis of fractions eluted of each of the four columns in the purification process: (A) 5 mL Ni^{2+}-NTA affinity column, (B) 1 mL Q-Sepharose ion exchange column, (C) 120 mL S200 size exclusion column and (D) 1 mL Ni^{2+}-NTA affinity column. Fractions pooled to be loaded onto the next column (or for dialysis in the final step) indicated.

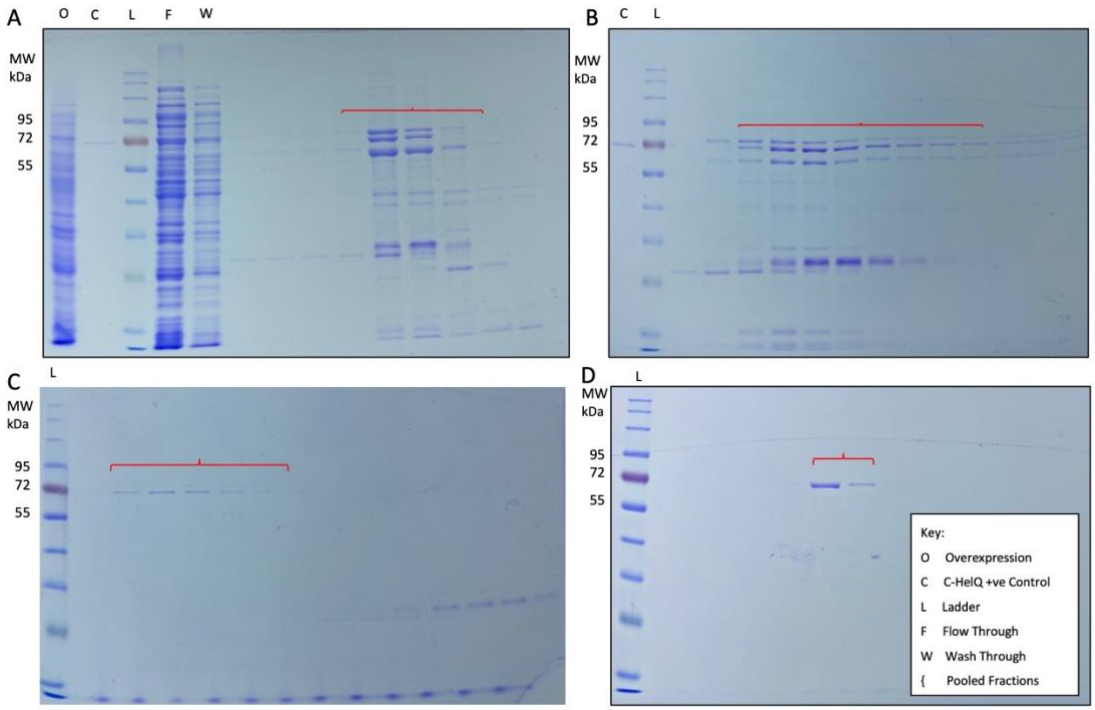


Figure 3.9 Purification of C-HelQ$^{P960A-Y963A}$ protein. Coomassie stained 10% (w/v) SDS-PAGE analysis of fractions eluted of each of the four columns in the purification process: (A) 5 mL Ni^{2+}-NTA affinity column, (B) 1 mL Q-Sepharose ion exchange column, (C) 120 mL S200 size exclusion column and (D) 1 mL Ni^{2+}-NTA affinity column. Fractions pooled to be loaded onto the next column (or for dialysis in the final step) indicated.

Comparative analysis of all three proteins, visualised through SDS-PAGE gel analysis, confirms approximate concentration of each protein, as determined using a NanoDrop, is accurate.

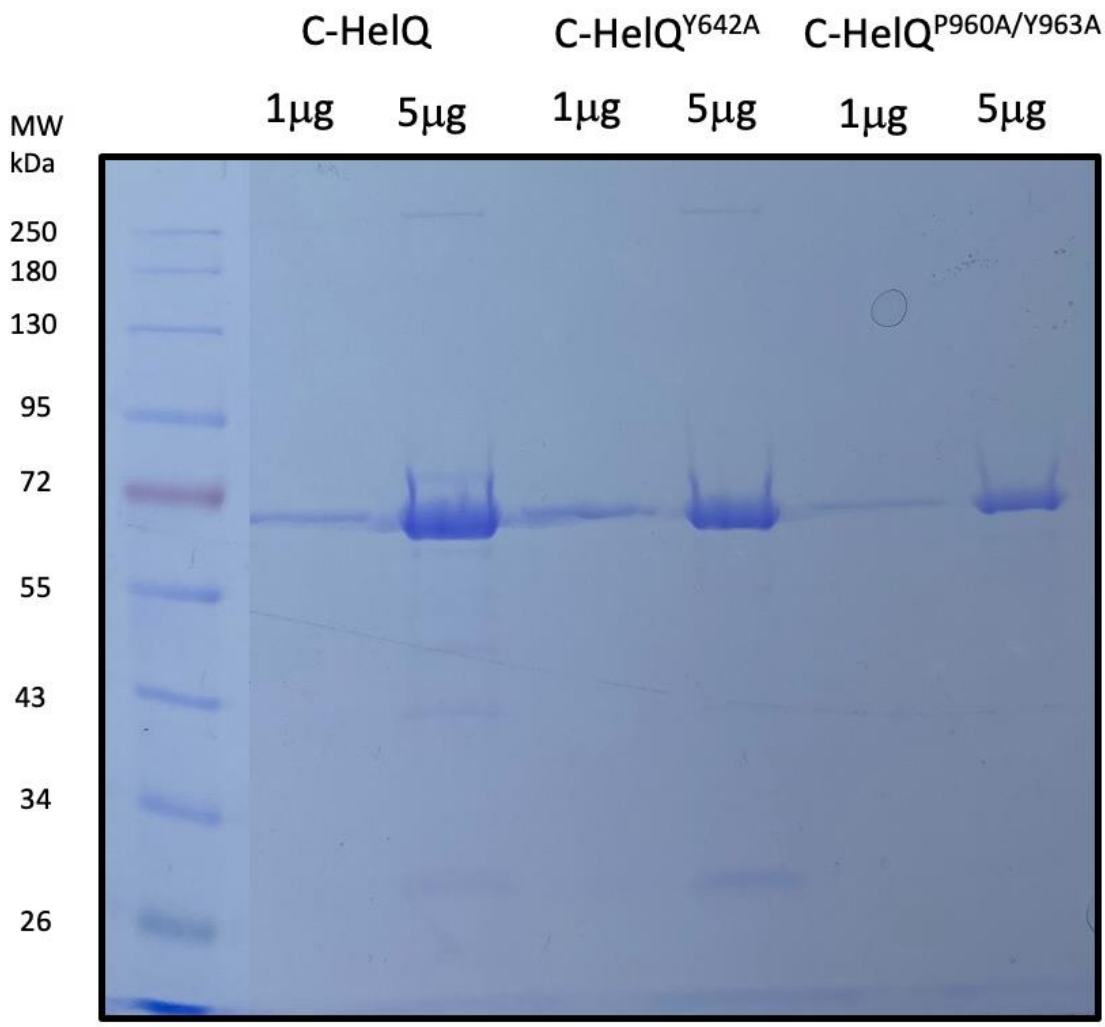


Figure 3.10 SDS-PAGE gel with 1 ug and 5 ug C-HelQ proteins. Coomassie stained 10% (w/v) SDS-PAGE analysis of 1 µg and 5 µg C-HelQ, C-HelQ^{Y642A} and C-HelQ$^{P960A-Y963A}$.

Analytical gel filtration with BioRad's Gel Filtration Standard was conducted alongside size-exclusion chromatography to determine the oligomeric state in which C-HelQ elutes off the column. This lyophilised mixture of five molecular weight markers ranging from 1,350 to 670,000 Da acts as a calibration standard to determine the molecular weight of the C-HelQ eluted. Figure 3.11 shows the standard proteins and their elution volumes used to calculate the molecular weight of C-HelQ. Each elution volume was plotted against the log of known

molecular weight (MW) to create a line of best fit (y=-14.93x+140). This was then used to calculate the molecular weight of C-HelQ based on its elution volume.

A

Component	Molecular Weight (Da)	LOG (MW)	UV (mAU)	Elution vol. (mL)
Thyroglobulin (bovine)	670,000	5.826075	97	52
γ-globulin (bovine)	158,000	5.198657	130	58
Ovalbumin (chicken)	44,000	4.643453	60	76
Myoglobin (horse)	17,000	4.230449	70	80
Vitamin B12	1,350	3.130334	62	90

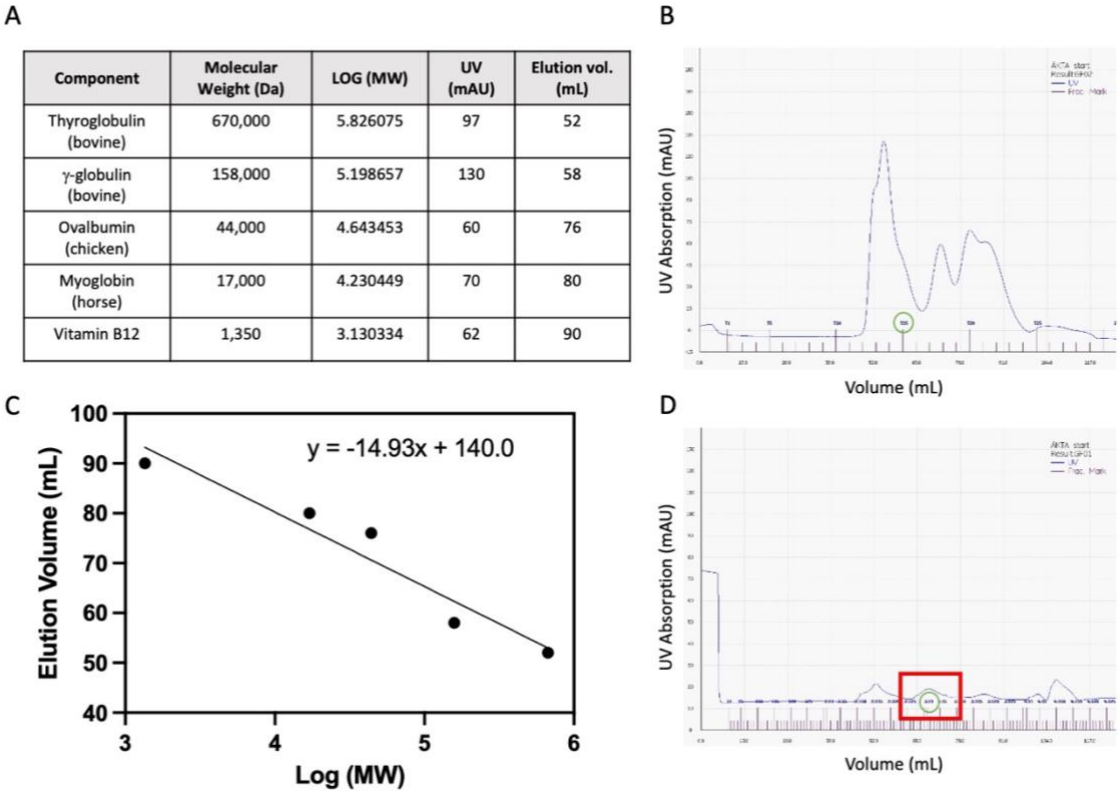


Figure 3.11 Analytical gel filtration of protein standards to determine the oligomeric state of C-HelQ.(A) Table list of the BioRad gel filtration protein standard run through the Superdex200 gel filtration column; (B) corresponding mAU graph. (C) Elution volume was plotted against log (MW), creating a line of best fit in which proteins of unknown size can be determined from. (D) mAU graph from gel filtration of C-HelQ: boxed in red is where C-HelQ is eluted. C-HelQ elutes off at 66 mL.

C-HelQ eluted off the gel filtration column as a ~90 kDa, as calculated from the line of best fit. With C-HelQ having a known size of 96 kDa, this indicates the protein has eluted off as a monomer. This is of particular interest as it has been

shown that full-length HelQ is purified as a dimer [69]; suggesting the N-terminal region may be involved in dimerization of HelQ.

3.3 Biochemical analysis of C-HelQ

3.3.1 Analysis of C-HelQ DNA binding

DNA-Protein interaction assays were performed to assess the DNA binding ability of C-HelQ and mutant protein *in vitro* through biochemical 'in-gel' electro-mobility shift assays (EMSAs) and biophysically through fluorescence anisotropy.

EMSA analysis, assessed on 5% (w/v) polyacrylamide TBE gels, allows the detection of stable protein-DNA bound complexes through visualising the migration of fluorescently labelled DNA substrates. Complexes of higher molecular weight migrate through the gel more slowly in comparison to unbound DNA, producing a visible band shift as protein concentration increases, therefore also identifying the concentration of protein at which all DNA forms a complex. EMSA analysis shows C-HelQ binds forked DNA substrates equally to both mutant C-HelQ proteins (Figure 3.12).

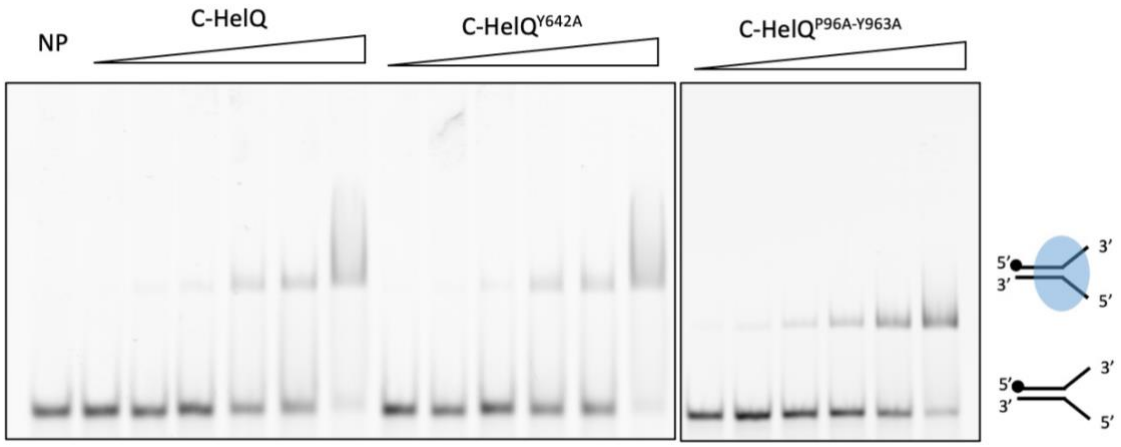


Figure 3.12 EMSA showing C-HelQ WT and mutant binding forked DNA. 5% (w/v) native TBE gels showing binding of wild-type and mutant C-HelQ on 25 nM 5’ Cy5-labelled forked DNA substrate. Protein concentration increases from 25 to 800 nM.

Fluorescence anisotropy, also known as fluorescence polarization, is the measurement of change of orientation of molecules in space. To overview the principles of fluorescence anisotropy, a fluorophore, which labels the substrate, emits light along difference axes of polarization. The rate at which the light emitted is fully depolarized reflects the anisotropy value. Binding of protein to DNA increases the size of the molecule, decreasing its mobility and so depolarization is slower – causing an increase in anisotropy [70]. This allows equilibrium dissociation constants (K_d values) to be determined.

Fluorescence anisotropy assays were performed to assess and compare the DNA binding ability of wild-type and mutant C-HelQ. Protein was titrated into a reaction pool containing a 5’ labelled 6-carboxyfluorescein (FAM)-labelled ssDNA substrate. Anisotropy readings were taken on a FLUOstar Omega plate reader. A no protein control was included, and its average R-value subtracted

from the readings of samples with protein. The R-value was determined according to the equation below:

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

Data was analysed using PRISM GraphPad software and K_d values were calculated based on the curve fit based on the one-site specific binding model.

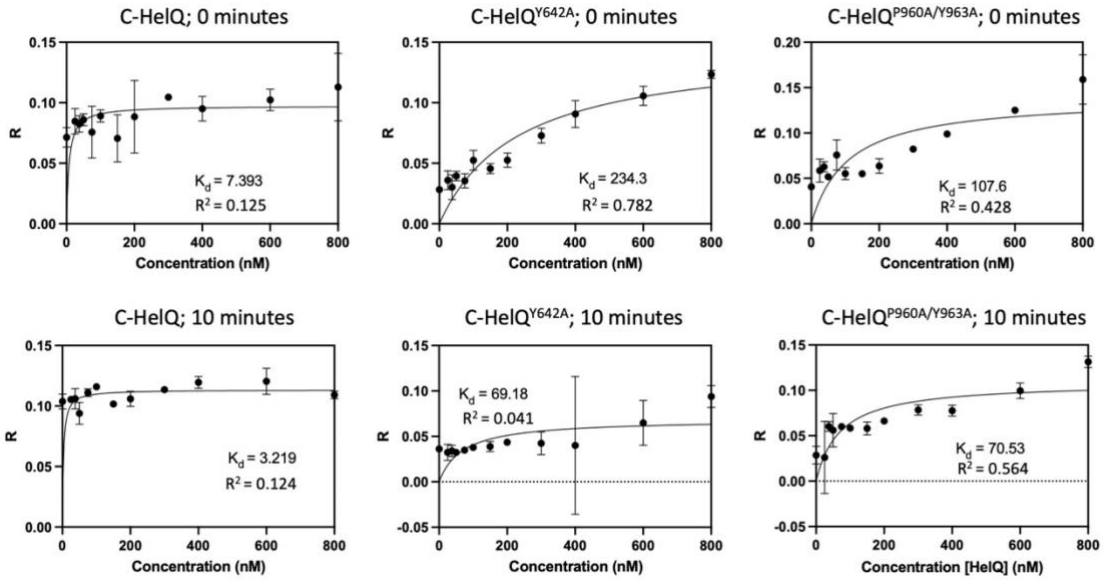


Figure 3.13 Binding affinities of C-HelQ proteins on ssDNA. Curves were generated through use of fluorescent anisotropy assays, showing the binding affinity of C-HelQ proteins for ssDNA initially and after incubation at 37°C for 10 minutes. Protein is titrated (at 0, 25, 37.5, 50, 75, 100, 150, 200, 300, 400, 600 and 800 nM) into the reaction pool containing 50 nM of single stranded substrate with a 5' labelled 6-carboxyfluorescein (FAM) reporter. Data plotted reflects means of duplicate data, with error bars indicating the standard deviation from the mean, calculated using Prism (GraphPad) software. K_d and R^2 values indicated on each graph, assessing the binding affinity and goodness of fit, respectively.

Interestingly, these preliminary anisotropy binding results assessed with this method do not seem to be comparable to the results seen on the EMSAs in regard to initial binding of mutant protein compared to wild-type protein binding. Y642A and P960A/Y963A mutations have a lower binding affinity than C-HelQ for ssDNA, demonstrated by higher K_d values. We also see all C-HelQ proteins binds the ssDNA substrate with a higher affinity after incubation at 37°C for 10 minutes. All three C-HelQ species K_d decreases after incubation, indicating more binding. After 10 minutes of incubation, the “Y642A” and “P960A-Y963A” mutants have very similar K_d values: 69.18 and 70.53, respectively. This indicates a direct contrast with what we have seen with full-length HelQ, outlined in Chapter 4, however more data points at lower concentrations to obtain a better curve fit would be required to be certain of this conclusion. It should be noted the goodness of fit of these models differs, indicated by R^2 values. Some models, such as C-HelQ^{Y642A} at time point 0 minutes, gives an R^2 value of 0.782, indicating a relatively strong positive linear relationship. On the other hand, some models, such as C-HelQ^{Y642A} at time point 10 minutes, gives an R^2 value of 0.041, indicating a weak relationship.

3.3.2 Analysis of C-HelQ unwinding and annealing activity

C-HelQ contains the catalytic motifs responsible for HelQ’s unwinding and annealing activity despite it lacking the N-terminal domain. The helicase and annealing activity of C-HelQ were assessed through end-point gel-based assays and time-course fluorescence resonance energy transfer (FRET) assays. Here, the comparison of assay results aims to reveal the differences in the unwinding/annealing activity of wild-type C-HelQ and mutant C-HelQ.

HelQ is thought to be involved in genome stability through acting at stalled replication forks. To investigate this, the helicase and annealing activity of C-HelQ were assessed using forked DNA substrates resembling stalled replication forks. Cy5 modifications (and additionally Cy3 modifications for FRET-based assays) were situated on the 5'-end of the leading DNA strand. Successful unwinding or annealing was identified and visualised by the accumulation of Cy5-fluorescently labelled DNA, pickup upon gel scanning with the Typhoon.

The unwinding activity was assessed by an end-point helicase assay, visualised on a 10% (w/v) polyacrylamide TBE gel, allowing the separation of Cy5-fluorescently labelled DNA species of different molecular weights by electrophoresis. Successful unwinding was identified and visualised by the accumulation of Cy5-fluorescently labelled DNA as imaged on the Typhoon phosphor-imager (Amersham). To minimise false negative results due to re-annealing of DNA, unlabelled ssDNA with the same sequence as the Cy5-labelled strand of the dsDNA substrate was added to the reaction pool in excess to anneal the dissociated product. Additionally, a no protein sample, "boiled" on a heat block for 10 minutes at 95°C was also used to identify the size of the unwound, fully dissociated product.

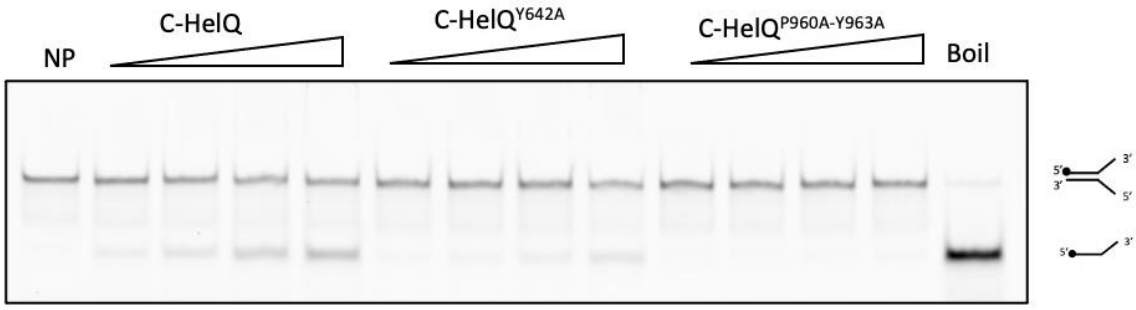


Figure 3.14 Unwinding of forked DNA by C-HelQ is modulated in mutant protein. Wild-type and mutant C-HelQ were assessed in an end-point assay to determine their unwinding activity for a forked DNA substrate (25 nM). Concentration titration reactions, with 50, 100, 200 and 400 nM of protein, were performed at 37°C for 10 minutes. A no protein control shows migration of forked DNA and a no protein boiled control "boil" shows ability for the substrate to be unwound and migration of the unwound Cy5-labelled substrate through the gel.

Wild-type C-HelQ and C-HelQ^{Y642A} show unwinding of the forked DNA substrate, but C-HelQ$^{P960A-Y963A}$ did not. These equivalent mutants in archaeal Hel308 resulted in a hyperactive unwinding phenotype (outlined in Chapter 4), yet interestingly in C-HelQ these mutations do not seem to result in the same hyperactive phenotype. Supressed unwinding by C-HelQ^{Y642A} as seen in Figure 3.14 aligns with already published data by Jenkins *et al* (2021) [64], who noted C-HelQ was inactivated by the Y642A amino acid mutation. This is however the first time we see data describing the effect of the P960A/Y963A mutation on DNA unwinding.

DNA annealing activity was also assessed in an end-point assay, visualised on a 10% (w/v) polyacrylamide TBE gel; showing the annealing of two ssDNA, one Cy5-labelled and one unlabelled. A pre-annealed DNA substrate of the two ssDNA strands used in the reaction pool was used as a control to allow size-identification of the fully annealed dsDNA product.

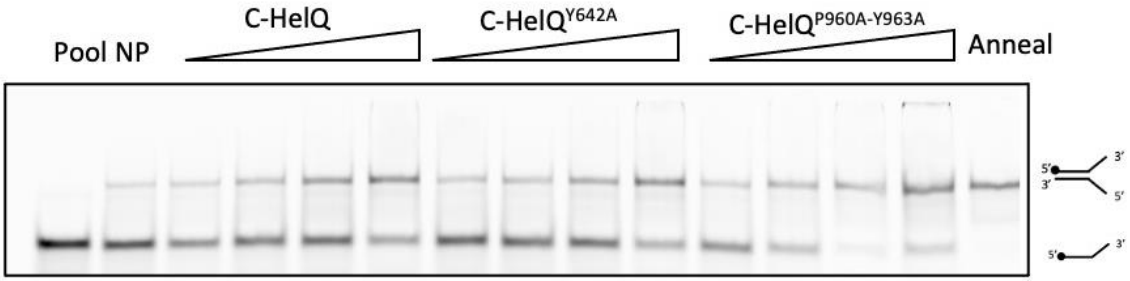


Figure 3.16 Annealing of complementary ssDNA by C-HelQ is modulated in mutant protein. Wild-type and mutant C-HelQ were assessed in an end-point assay to determine their annealing activity of two ssDNA substrates (15 nM). Concentration titration reactions, with 50, 100, 200 and 400 nM of protein, were performed at 37°C for 10 minutes. A no protein, no complementary DNA control "pool" shows migration of Cy5-labelled ssDNA. This was used in addition to a no protein "NP" control as spontaneous DNA annealing was observed. Additionally, a pre-annealed control "anneal" shows migration of fully annealed DNA substrate through the gel.

Annealing, independent of ATP, by the Y642A mutant here seems to be equally active to the wild-type protein, whereas the P960A/Y963A mutant shows increased annealing. Fluorescence resonance energy transfer (FRET) techniques were then employed to assess C-HelQ's annealing activity, giving more precise, real-time results of annealing activity over a 30-minute time period, incubated at 37°C.

FRET assays have been extensively used to study DNA-protein interactions in a distance-dependant interaction in which energy is transferred from the excitation of a donor molecular fluorophore to an acceptor molecular fluorophore. In this way, the changes in molecular dynamics between DNA and a protein can be studied; the energy transfer efficiency (E) can be quantified

and correspond to the closeness in proximity of the fluorophores. The greater the E value, the closer in proximity the two fluorophores are to each other.

FRET assays show annealing of two complementary 70-nt ssDNA by C-HelQ and C-HelQ^{Y642A}. In addition to the standard no protein control and a fully annealed substrate control used, a Cy3 control was also implemented. This was to account for any background Cy3 fluorescence, which was considered when calculating true FRET readings. FRET “Anneal values” were determined by first correcting FRET readings against the background Cy3 emission, then calculated based on a control reaction with fully annealed substrate.

FRET anneal values for C-HelQ^{Y642A} were approximately double compared to wild-type C-HelQ (Figure 3.17) implying the Y642A mutation results in a hyper-annealing phenotype. The results show for both proteins, annealing increases over the 30-minute time-course, with greater protein concentration used resulting in more annealing. The graphs show annealing by C-HelQ^{Y642A} plateaus implying complete annealing of all ssDNA, particularly in reactions with 400 nM protein. However, the reaction with wild-type C-HelQ does not implying maximum annealing is yet to be reached.

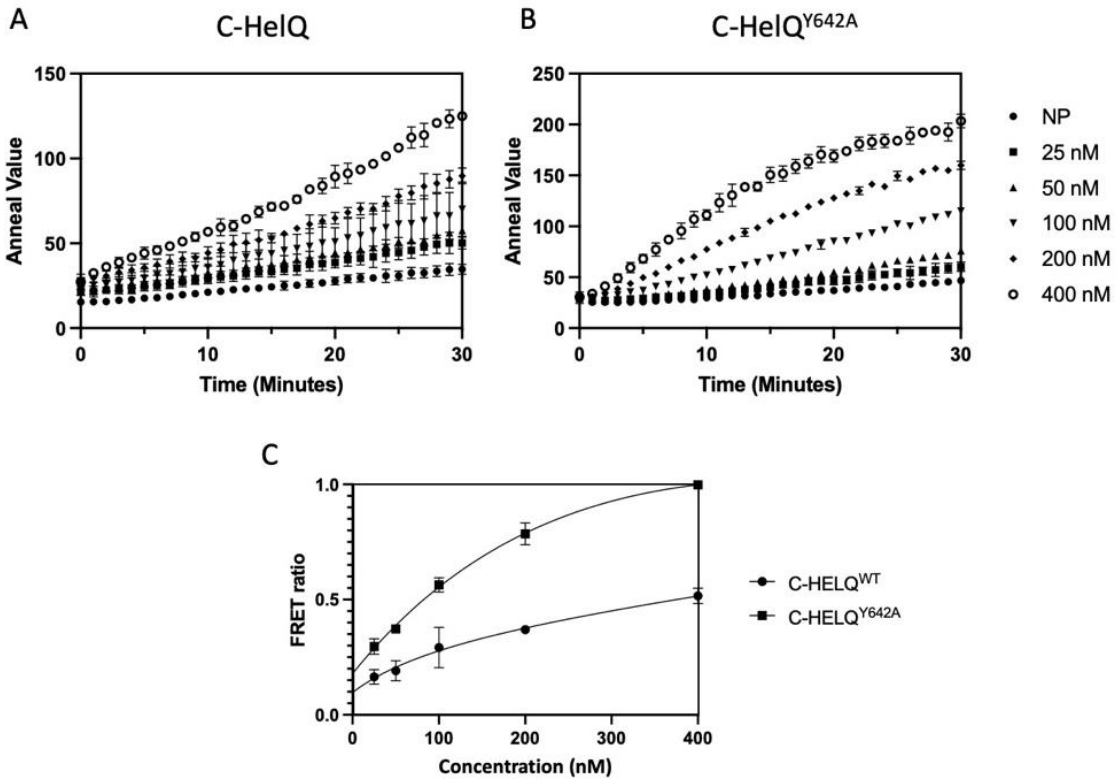


Figure 3.17 Annealing of complementary 70-nt ssDNA substrates by C-HelQ is modulated in mutant protein in FRET analysis. Wild-type C-HelQ (A) and mutant C-HelQ^{Y642A} (B) time-course FRET-assays were carried out at 37°C over 30-minutes, independent of ATP. Concentration titration reactions, with 25, 50, 100, 200, and 400 nM of protein, were performed to anneal two complementary ssDNA species 70-nt in length. “Anneal values” were determined by first correcting FRET readings against the background Cy3 emission, then calculated based on a control reaction with fully annealed substrate. (C) End-point anneal values of each concentration for each protein were plotted for comparison, with values given as a ratio against maximum annealing.

3.4 Further C-HelQ experimentation

3.4.1 Nuclease protection assays

Nuclease protection assays, with S1 nuclease, were used to determine the effect of C-HelQ in preventing ssDNA degradation by S1 nuclease. This was employed to see the effect, if any, of the C-HelQ^{Y642A} mutation on exposure of ssDNA for S1 nuclease digestion.

Firstly, the assay was optimised to identify the minimum concentration of S1 nuclease and incubation time at which all DNA is degraded, Figure 4.18. As incubation time increases, the more degradation of ssDNA by the S1 nuclease is seen. Likewise, the greater the unit concentration of S1 nuclease used, the more DNA degradation is seen. 10 units of S1 nuclease and an incubation time of 20 minutes at 37°C was chosen based on the optimisation assay.

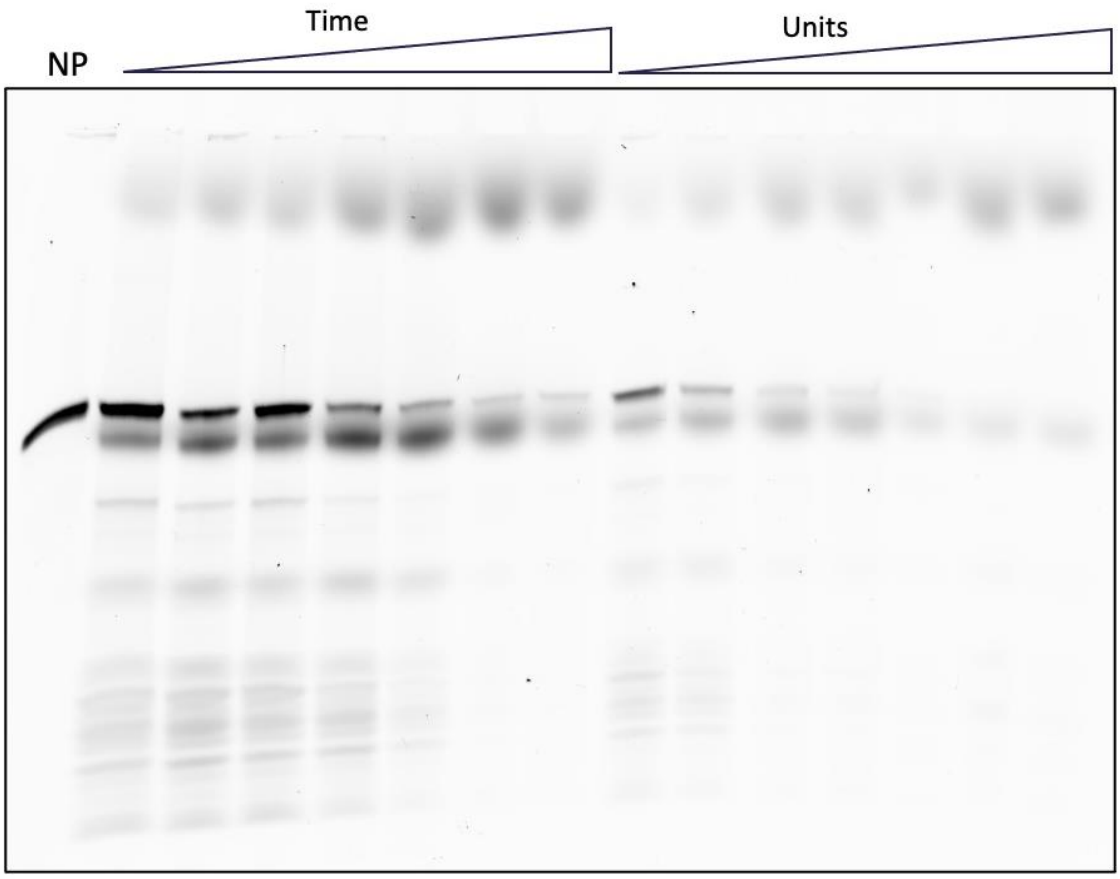


Figure 3.18 Nuclease Protection Assay Optimisation. Lane 1: NP; Lane 2-8: Time iterations, 1 minute, 2 minutes, 3 minutes, 5 minutes, 10 minutes, 20 minutes, 30 minutes (10 units); Lane 9-15: Unit iterations: 1 unit, 2 units, 5 units, 10 units, 20 units, 50 units, 100 units (10 minutes incubation). Run on 15% (w/v) denaturing gel for 180 minutes at 10 watts.

Preliminary data collected (Figure 4.19) shows S1 nuclease protection by HelQ, C-HelQ and C-HelQ^{Y642A}. Full-length HelQ was used as a control to compare C-HelQ nuclease protection to. The gel images suggest HelQ protects ssDNA against S1 nuclease digestion. This is noticeable at 100 to 400 nM of protein. C-HelQ is showing less protection from the S1 nuclease compared to HelQ, but more protection compared to the Y642A mutant, which shows digestion by the S1 nuclease.

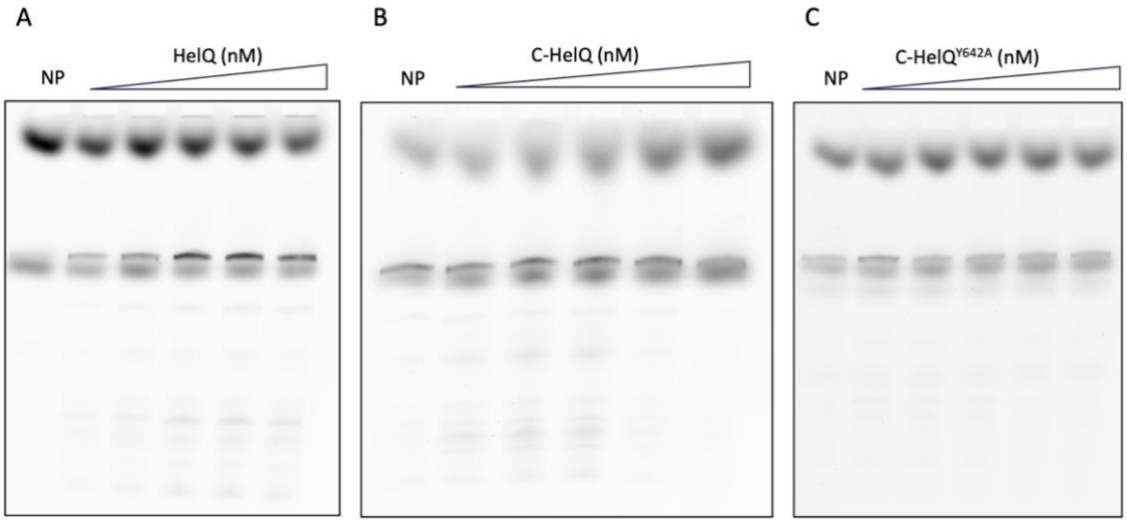


Figure 3.19 Nuclease Protection Assay. S1 Nuclease Protection Assay with (A) HelQ, (B) C-HelQ and (C) C-HelQ^{Y642A}. Protein titrated at 25, 50, 100, 200, 400 nM. Run on 15% (w/v) denaturing gel for 180 minutes at 5 watts. Assay was carried out with 10 units of S1 nuclease, incubated at 37C for 20 minutes.

3.4.2 Alternative annealing assay with full-length HelQ

Further assessment of C-HelQ annealing was tested with an alternative annealing assay using T4 DNA ligase and restriction enzymes, NdeI and BamHI (Figure 3.20). These two enzymes cut sites were chosen as they have 2- and 4-bp overhangs, respectively, so it could be determined whether this had any effect on annealing. In the absence of HelQ and T4 DNA ligase, pET14b cut with NdeI or BamHI show supercoiling compared to the uncut plasmid; supercoiling is indicated by DNA in these reactions migrating to a greater kb and indicates good quality of plasmid minipreps. HelQ and/or T4 DNA ligase is added to the reaction mix to determine their effect on annealing. Data as seen in Figure 4.20 indicate this method for testing DNA annealing does not give any conclusive results. Uncut pET14b treated with HelQ, irrespective of presence of T4 DNA ligase, show a second DNA fragment. There is no effect of HelQ or

T4 DNA ligase on pET14b cut with NdeI. pET14B cut with BamHI show treatment with T4 DNA ligase, irrespective of HelQ, also show multiple DNA fragments.

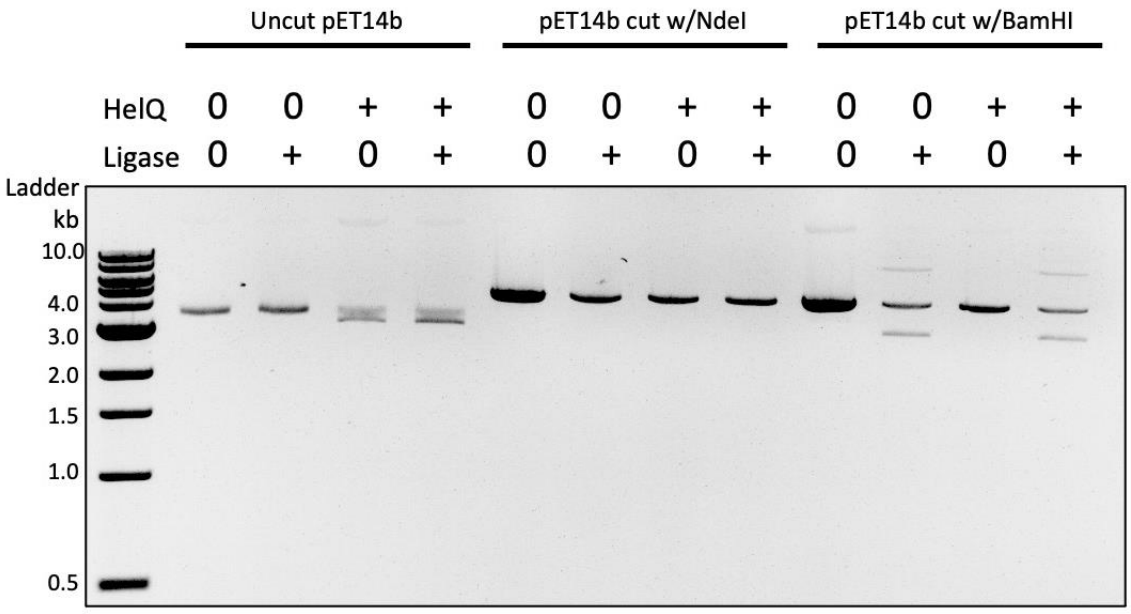


Figure 3.20 Alternative annealing assay with restriction enzymes.1% (w/v) agarose gel, ran for 1 hour at 140 V, shows migration of DNA products after treatment with HelQ and/or T4 DNA Ligase.

3.4.3 Prime editing

Assessing the proteins and mutations highlighted in this chapter *in vivo* would be of interest to determine any cellular effects. One method of doing so would be to exploit prime editing techniques. Prime editing eliminates several limiting factors of traditional gene editing strategies, such as CRISPR. Traditional methods of genome editing rely on creating a DSB to introduce a desired mutation through recruitment of error-prone DNA repair machinery. Instead, prime editing allows for site-specific nicks to be created in DNA through use of

a Cas9 nickase fused to a M-MLV reverse transcriptase, and an RNA template, avoiding any double-strand breaks.

While no *in vivo* experimentation has been conducted for this research, preliminary steps required for any future *in vivo* experimentation have been started; namely construction of prime editing plasmids for HelQ Y642A and HelQ D142A/F143A – the former of which (Y642A) this work details the background and biochemistry of, the latter is a double mutation in the PWI-like motif situated within the intrinsically disordered N-terminal of HelQ that has been previously studied in the Bolt lab by Tabitha Jenkins.

Prime editing oligos were designed using pegFinder [http://pegfinder.sidichenlab.org] in conjunction with human genome sequences extracted from the online source, the Human Genome Browser.

Golden gate assembly protocol was followed to assemble gRNA into the acceptor plasmid (Addgene, Plasmid #65777) and assemble pegRNA into the acceptor plasmid (Addgene, Plasmid #132777). Cloning of each sgRNA plasmid construct involved generating a phosphorylated oligo duplex with two inserts, digesting and dephosphorylating the vector backbone and ligating the duplex and plasmid together. Cloning of each pegRNA plasmid construct involved the annealing of three pairs of insert oligos (sg, scaffold and extension primers), generating a phosphorylated sgRNA scaffold and ligating the inserts into the digested vector backbone.

Ligations were transformed into DH5α cells using the standard heat-shock protocol, miniprepped and an analytical digest conducted. Plasmids were then sent for sequencing, confirming the presence of all appropriate inserts into each plasmid.

3.5 Summary

Biochemical analysis of C-HelQ show the wild-type protein functions in DNA processing activities through binding, unwinding and annealing DNA. DNA processing by C-HelQ is modulated by amino acid substitution mutations Y642A and P960A/Y963A.

The Y642A mutant – a mutation within the RecA domain of the protein – results in decreased binding and unwinding activity by the helicase enzyme, but demonstrates a hyper-annealing phenotype, seen in FRET-assay analysis. There is preliminary evidence to suggest C-HelQ^{Y642A} may also allow for DNA to be more susceptible to digestion by S1 nuclease, as determined with nuclease protection assays. The P960A/Y963A mutant – a mutation within the DNA ratchet motif – also results in decreased binding and unwinding activity, and, also, indicates an increased annealing function, as seen on gel-based assays. This would need to be further assessed using FRET-techniques. Additional nuclease protection would be of interest to assess. This data may provide insight into a structure-function relationship the mutant proteins adopt.

In addition, C-HelQ is eluted as a monomer, as determined by analytical gel filtration. This is in contrast to HelQ which has been previously shown to predominantly elute as a dimer [69]. This perhaps suggests a role for N-HelQ in dimerization. Further experimentation would be required to assess this difference when purifying proteins and establish a possible cause or mechanism for this difference.

This novel data begins to suggest that introducing these mutations may have caused a change in the structure-function relationship of the C-HelQ protein – possibly through a conformational change in the proteins structure from the substitution of amino acids, resulting in modulation to the protein biochemistry excreted by the protein on DNA processing activities. Discussed further in Chapter 6, this interpretation can be supported by recent literature (Lever *et al,* 2023) where molecular dynamics simulations of comparable mutants in the archaeal homologue of HelQ, Hel308, show changes to the structural conformation of the protein, resulting in a more open configuration [68].

Chapter 4 Analysis of Hel308 family proteins

4.1 Introduction

Hel308 family proteins, belonging to the SF2 family of DNA helicases, are thought to promote genome stability through interacting with stalled replication forks during DNA repair and replication via interactions with RPA and other homologous recombination proteins. Archaeal Hel308 and the metazoan homologue HelQ are able to bind DNA substrates, with a preference for 3' ssDNA ends, allowing loading of the protein onto DNA through residue interactions within the dual RecA domain, for ATP dependent translocation and DNA unwinding. Hel308's helicase mechanism is well characterized, yet it remains unclear how these activities contribute to genome stability. Recent studies reveal possible insight into mechanisms in which Hel308 modulates DNA recombination and repair.

4.2 Bioinformatic analysis of Hel308 family proteins

4.2.1 Homology search on Hel308 family proteins

Hel308 helicases promote genome stability in archaea and are found to be conserved in metazoans. Metazoan Hel308's name was based on the *Drosophila mus308. Mus308* was originally predicted to encode a helicase-polymerase but has since been shown to encode PolQ. Hel308 was therefore renamed to HelQ to align with this. Extensive regions of *H. sapiens* HelQ have strong sequence homology with archaeal Hel308 and other helicases, such as PolQ. This has facilitated the structural modelling of HelQ, as there is no resolved crystal structure of the protein. Despite this, there are significant non-homologous regions between these proteins suggesting these proteins also have distinct differences in their mechanisms.

Figure 4.1 Sequence alignment of Hel308 family proteins identifies highly conserved residues. Multiple protein sequence alignment of *H. sapiens* HelQ and four archaeal species of Hel308 (*Archaeoglobus fulgidus, Pyrococcus furiosus, Sulfolobus solfataricus* and *Methanothermobacter thermaltotrophicus*) by Clustal Omega. The Walker A (blue) and Walker B (green), also known as a DExH box, motifs highlighted show an example of sequence conservation of functionally important residues.

Sequence alignment of Hel308 from multiple archaeal species and *H. sapiens* HelQ show strong sequence homology, as indicated through example of the conservation of the Walker A and Walker B motifs in (Figure 4.1). Several

archaeal Hel308 species have had their crystal structures resolved which have provided detail in understanding its DNA helicase and translocase mechanisms. Along with the known structures of other SF2 family helicases, this has allowed HelQ structure and domain organization to be predicted through sequence alignment and the use of online structural modelling tools such as Phyre2. Hel308 family proteins share a core catalytically active domain responsible for its translocase and helicase activities, as seen in Figure 4.2. Here, the conservation of the Walker A and Walker B motif within the RecA domains can be seen.

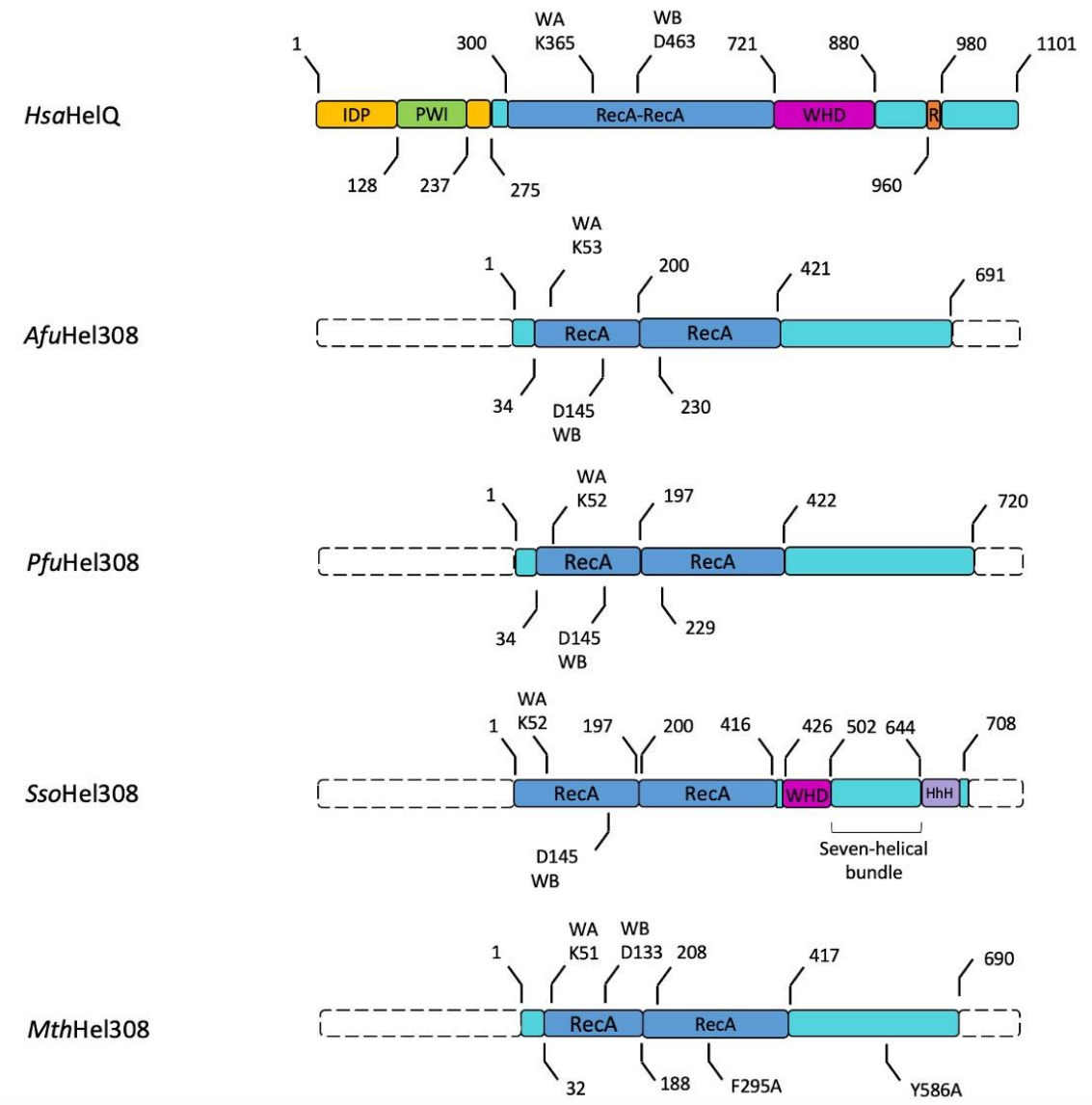


Figure 4.2 Domain map comparison of Hel308 family proteins. Gene profile of Hel308 family proteins showing conserved functional domains. Protein and domain lengths are proportional.

4.2.2 Structural prediction of HelQ

Structural models of HelQ are reliant on bioinformatic predictions as the crystal structure has not been fully resolved. Structural models can help predict its oligomeric state and determine how HelQ may interact with DNA and other proteins. Phyre2, a free web-based service for protein structural predication, was used for predictive modelling. The amino acid sequence of HelQ

(Supplementary Figure 1 – HelQ Sequence) was pasted into Phyre2, which generated a FASTA file which could be inserted into structural modelling software, such as PyMOL.

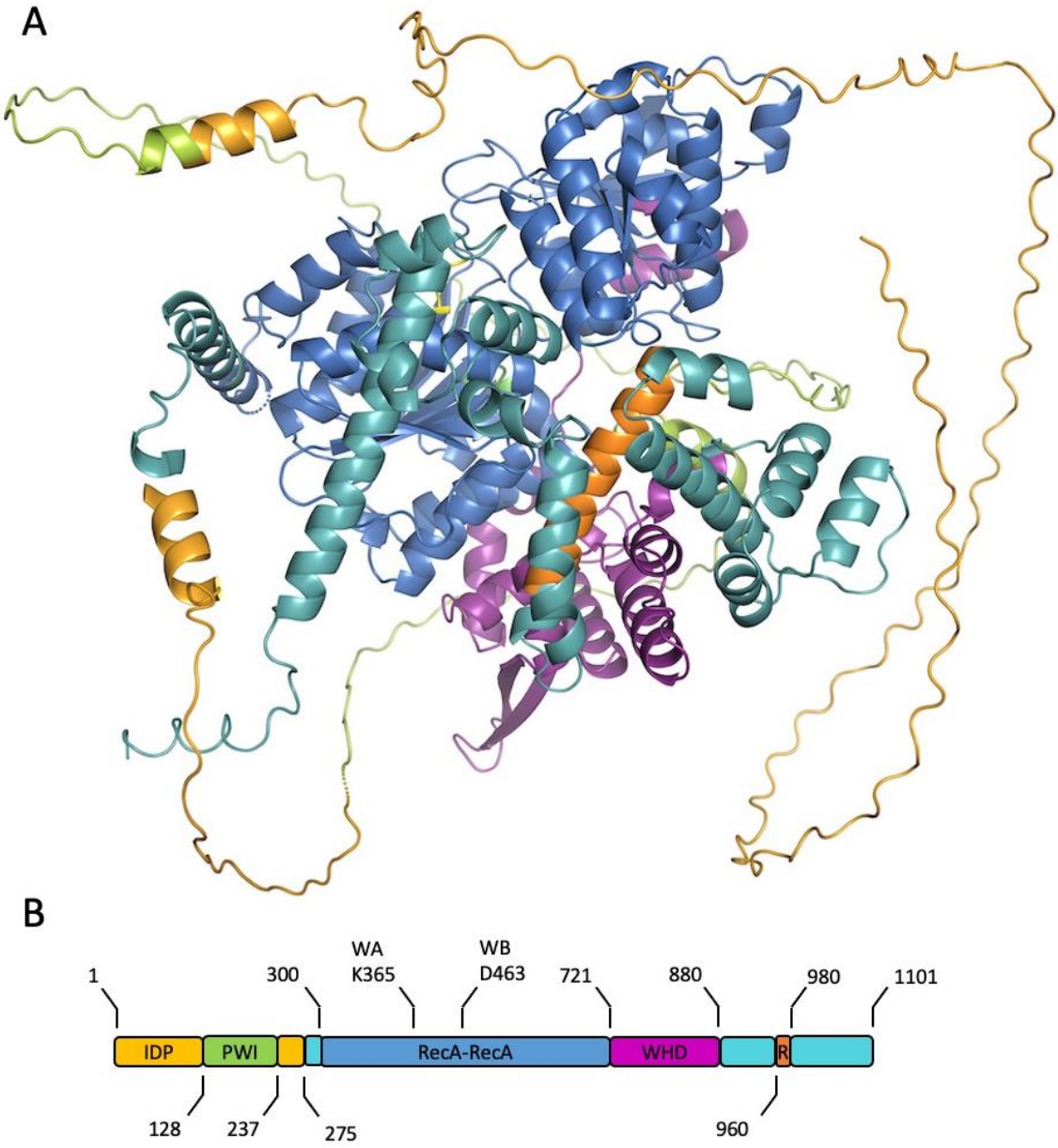


Figure 4.3 Predicted PyMOL model of HelQ with domain map. (A) Phyre2 predicted structural model of full-length *H. sapiens* HelQ. Yellow: N-terminal domain (N-HelQ); Green: PWI-like domain; Blue (with Dark Blue, Magenta and Orange): C-terminal domain (C-HelQ); Dark Blue: dual RecA-like domain; Magenta: winged-helix domain; Orange: DNA ratchet: Neon Yellow: Walker A motif; Neon Green: Walker B motif. (B) Domain map of full length HelQ for

comparison, colours correspond to predicted structural model colour coded structure. Models visualized and annotated using PyMOL. Colour labelling consistent with structural model with A.

Full length HelQ can be difficult to predict due to its N-terminal region being an intrinsically disordered protein (IDP). Throughout this research I have referred to full-length HelQ having been split where this IDPR ends, giving the intrinsically disordered N-HelQ and the catalytically functional C-HelQ. While the majority of HelQ's functional activity comes from within C-HelQ, N-HelQ does contain a PWI-like domain which has been shown to have a role in DNA loading through interaction with RPA [64]. Regions of intrinsic disorder, of >30 residues, have found to be present in 44% of the proteins in the human genome [71]. It is predicted that upon protein binding, IDPs undergo a disorder-to-order transition, forming secondary structural features and thus allowing the protein to participate in functional processes.

4.3 Biochemical analysis of *M. thermautotrophicus* Hel308

Protein overexpression and purification of *Methanothermobacter thermautotrophicus* Hel308 was previously achieved resulting in functional wild-type and mutant Hel308 (work caried out by Ryan Buckley and Rebecca Lever). Biochemical analysis of wild-type Hel308 and mutant Hel308 have been reported in literature [68]. The two Hel308 mutants used in this work are “F295A”, a single amino acid mutation within the RecA2 domain discovered to cause a hyper-recombination phenotype caused by hyper-active DNA annealing; and “Y586A”, a single amino acid mutation in the ‘ratchet’ motif shown to lead to hyper-active DNA unwinding. Reported here are the results of assays assessing the DNA processing abilities of these proteins, exploring how these important residues modulate DNA processing in Hel308.

4.3.1 Analysis of Hel308 DNA binding

Protein-DNA interactions were studied to assess the DNA binding ability of Hel308 and mutant protein. EMSA analysis, assessed on 5% (w/v) polyacrylamide TBE gels showed the formation of stable protein-DNA bound complexes through visualising the migration of fluorescently labelled forked DNA substrates.

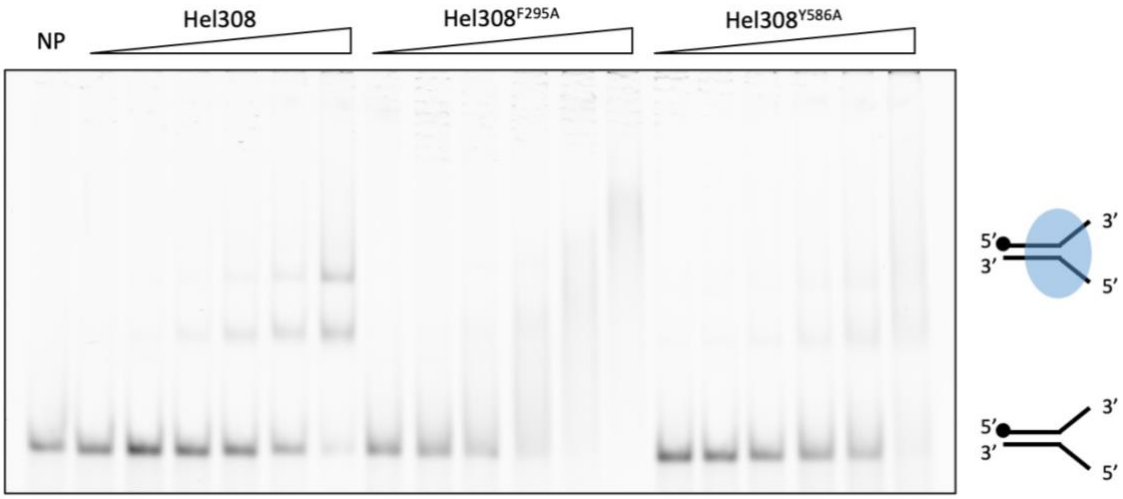


Figure 4.4 EMSA showing Hel308 WT and mutant binding. 5% (w/v) native TBE gels showing binding of wild-type and mutant Hel308 on 25 nM 5' Cy5-labelled forked DNA substrate. Protein concentration increases from 25, 50, 100, 200, 400, 800 nM. WT Hel308 and mutants bind forked DNA.

In EMSAs, all three Hel308 proteins formed stable complexes with the forked DNA substrate (Figure 4.4). Hel308^{F295A} demonstrated increased protein-DNA binding compared to wild-type Hel308 and Hel308^{Y586A}. This conclusion can be justified as there was a DNA-Hel308^{F295A} complexes form more readily than wild-type and Hel308^{Y586A} proteins, showing full binding of all DNA substrate. This can be seen especially when comparing the availability of free, unbound DNA substrate with of all proteins at the same concentration. This aligns with results described in current literature [68].

4.3.2 Analysis of Hel308 unwinding and annealing activity

Biochemical activity of wild-type and mutant Hel308 was also assessed to reproduce published data. The helicase activity was assessed by an end-point assay, visualised on a 10% (w/v) polyacrylamide TBE gel, allowing the

separation of Cy5-fluorescently labelled DNA species of different molecular weights by electrophoresis. As before, successful unwinding was identified and visualised by the accumulation of Cy5-fluorescently labelled DNA as imaged on the Typhoon phosphor-imager (Amersham). To minimise false negative results due to re-annealing of DNA, unlabelled ssDNA with the same sequence as the Cy5-labelled strand of the dsDNA substrate was added to the reaction pool in excess to anneal the dissociated product. Additionally, a no protein sample, “boiled” on a heat block for 10 minutes at 95°C was also used to identify the size of the unwound, fully dissociated product.

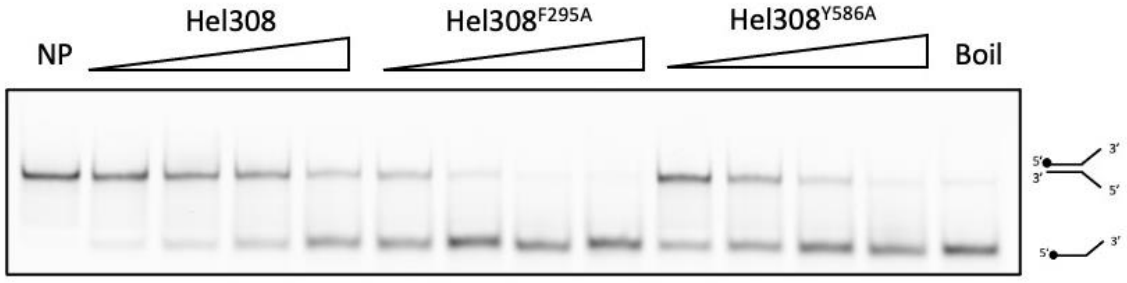


Figure 4.5 Unwinding of forked DNA by Hel308 is modulated in mutant protein. Wild-type and mutant Hel308 were assessed in an end-point assay to determine their unwinding activity for a forked DNA substrate (25 nM). Concentration titration reactions, with 10, 20, 40 and 80 nM of protein, were performed at 37°C for 10 minutes. A no protein control shows migration of forked DNA and a no protein boiled control “boil” shows ability for the substrate to be unwound, and migration of the unwound Cy5-labelled substrate through the gel.

Hel308 unwinding activity is modulated with single amino acid point mutations in the RecA2 or ratchet motif of Hel308. Both mutant proteins, Hel308^{F295A} and Hel308^{Y586A}, can be seen to unwind forked DNA hyperactively compared to wild-type Hel308 (Figure 4.5); with Hel308^{F295A} showing to modulate this activity

the greatest. This indicates to the significance of the domain IV – the dual RecA domain – in facilitating the DNA unwinding function of this helicase.

Hel308's DNA annealing activity was also assessed in an end-point assay, visualised on a 10% (w/v) polyacrylamide TBE gel; showing the annealing of two ssDNA, one Cy5-labelled and one unlabelled. A pre-annealed DNA substrate of the two ssDNA strands used in the reaction pool was used as a control to allow size-identification of the fully annealed dsDNA product.

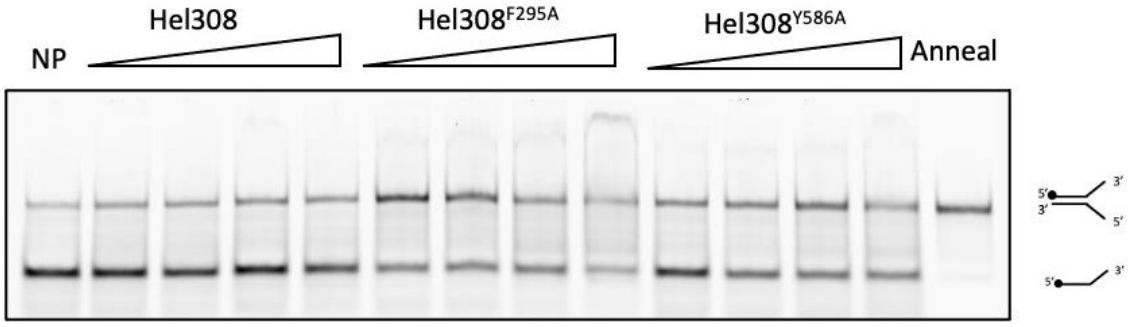


Figure 4.6 Annealing of complementary ssDNA by Hel308 is modulated in mutant protein. Wild-type and mutant Hel308 were assessed in an end-point assay to determine their annealing activity of two ssDNA substrates (15 nM). Concentration titration reactions, with 10, 20, 40 and 80 nM of protein, were performed at 37°C for 10 minutes. A no protein control shows migration of Cy5-labelled ssDNA and a pre-annealed control "anneal" shows migration of fully annealed DNA substrate through the gel.

Hel308 has the ability to anneal complementary single stranded DNA independently of ATP. Similar activity can be seen visually (Figure 4.6) in Hel308^{Y586A} when compared to wild-type Hel308. However, this activity is significantly increased with Hel308^{F295A} – further to this, the highest concentration of protein used caused protein aggregation in the well, further

emphasising the proteins modulated activity. This therefore identifies Hel308^{F295A} to be hyperactive in all three forms of DNA processing: binding, unwinding and annealing.

4.4 Biochemical analysis of *H. sapiens* HelQ

Hel308 has been used biochemically as a model for HelQ activity. It was therefore of interest to also assess the DNA processing activity of HelQ biochemically. HelQ used in this work has been commercially bought.

4.4.1 Analysis of HelQ binding, unwinding and annealing activity

DNA binding of HelQ was initially assessed, *in vitro,* biochemically though EMSAs. This was to show like Hel308, HelQ too can bind DNA substrates – this was assessed using the same forked DNA substrate as in the Hel308 assays, and additionally with ssDNA. Figure 4.7 details EMSA analysis showing HelQ forms stable complexes with forked DNA and a 5' Cy5-labelled 70bp ssDNA substrate.

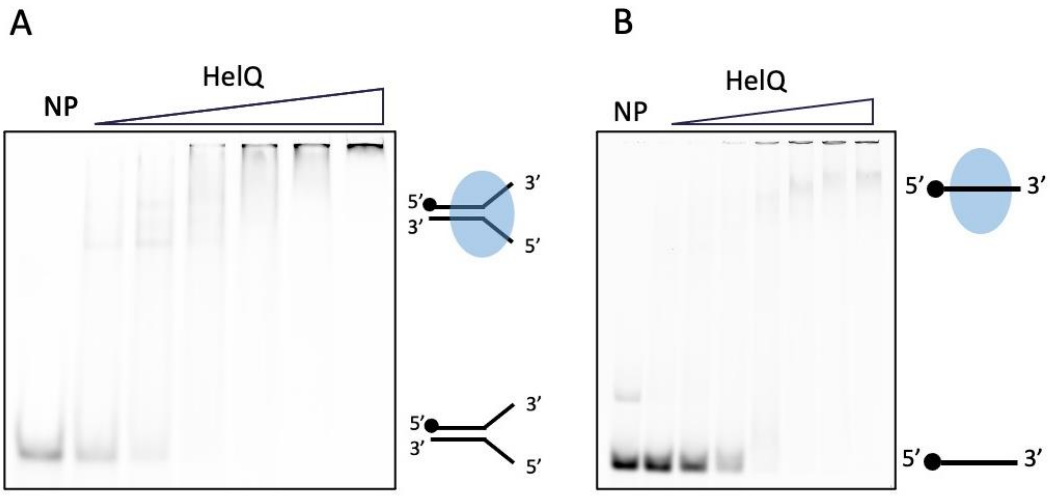


Figure 4.7 HelQ binds DNA substrates, forming stable protein-DNA complexes. (A) EMSA of HelQ, titrated at 25, 50, 100, 200, 400 and 800 nM, shows binding of forked DNA (25 nM), forming a stable complex. (B) EMSA of HelQ, titrated at 12.5, 25, 50, 100, 200, 400 and 800 nM, shows binding of ssDNA (25 nM), forming a stable complex. Assessed on a 5% (w/v) polyacrylamide TBE gel.

In addition to EMSA analysis of HelQ binding, fluorescence anisotropy, was also used to obtain real-time, precise data readings for binding, and to calculate equilibrium dissociation constants (K_d values) through plotting the change in anisotropy values against protein concentration.

Fluorescence anisotropy assays were performed to assess and compare the DNA and RNA binding ability of HelQ. Protein was titrated into a reaction pool containing a 5' labelled 6-carboxyfluorescein (FAM)-labelled ssDNA or a ssRNA substrate. As before, anisotropy readings were taken on a FLUOstar Omega plate reader and a no protein control was included, with its average R-value subtracted from the readings of samples with protein.

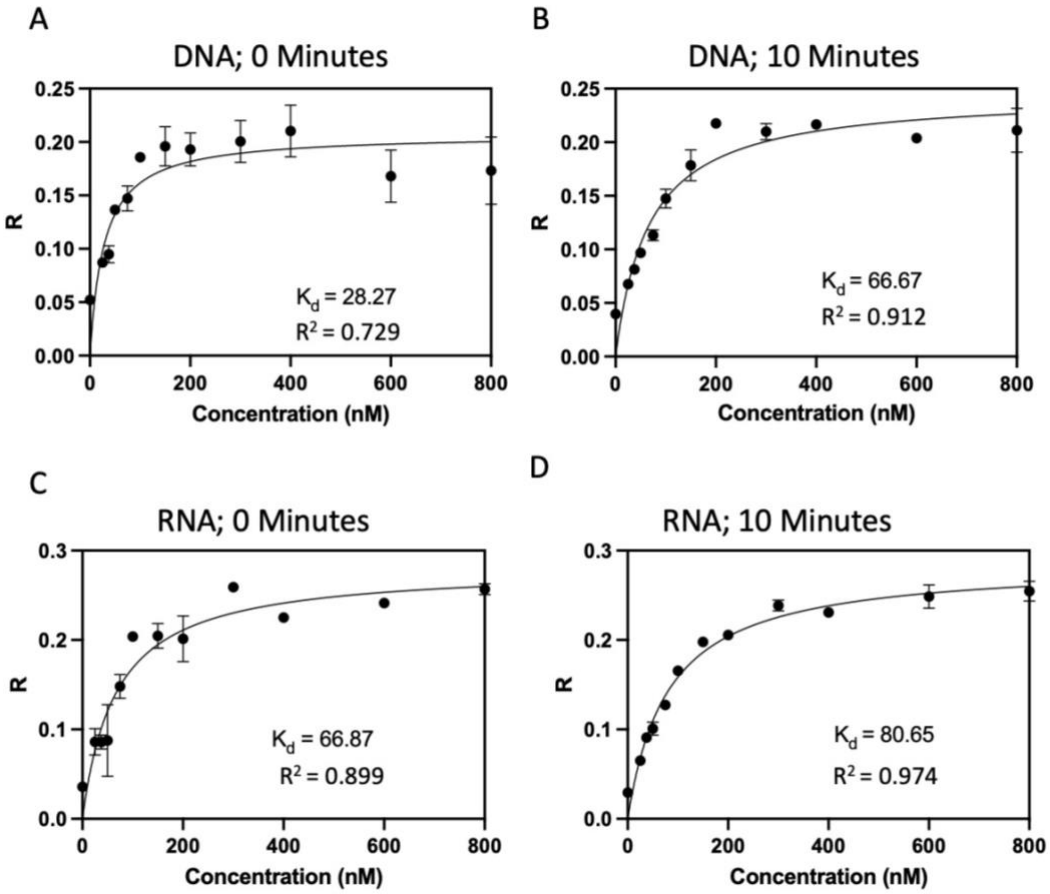


Figure 4.8 HelQ DNA binding vs HelQ RNA binding. Binding curves were generated using a fluorescence anisotropy assay, showing the binding affinity of HelQ for a ssDNA substrate and ssRNA substrate initially and after incubation at 37°C for 10 minutes. Protein is titrated (at 0, 25, 37.5, 50, 75, 100, 150, 200, 300, 400, 600 and 800 nM) into the reaction pool containing 50nM of single stranded substrate with a 5' labelled 6-carboxyfluorescein (FAM) reporter. Anisotropy is increased as protein binding increases allowing K_d values to be calculated. Data plotted reflects means of duplicate data, with error bars indicating the standard deviation from the mean, calculated using Prism (GraphPad) software. K_d and R^2 values indicated on each graph, assessing the binding affinity and goodness of fit, respectively.

HelQ demonstrates a preference for DNA substrates over RNA substrates, indicated by having a lower K_d value for the ssDNA substrate at each comparative time point than ssRNA (Figure 4.8). Upon initial binding, HelQ produced a K_d of 28 nM for ssDNA which was approximately 2.5 times lower

than the K_d for the ssRNA species (which had a a K_d of 66 nM). For both the DNA and RNA species, HelQ binds better initially, and after incubation at 37°C for 10 minutes shows less binding. High R^2 values indicate positive correlation and great goodness of fit for the model. This confirms what we already know about Ski2-like family helicases showing a preference for DNA than RNA, unlike other SF2 family proteins.

Following the same method as the assays assessing Hel308's unwinding activity, an end-point assay visualised on a 10% (w/v) polyacrylamide TBE gel was first used to show HelQ's ability to unwinding forked DNA, shown in (Figure 4.9).

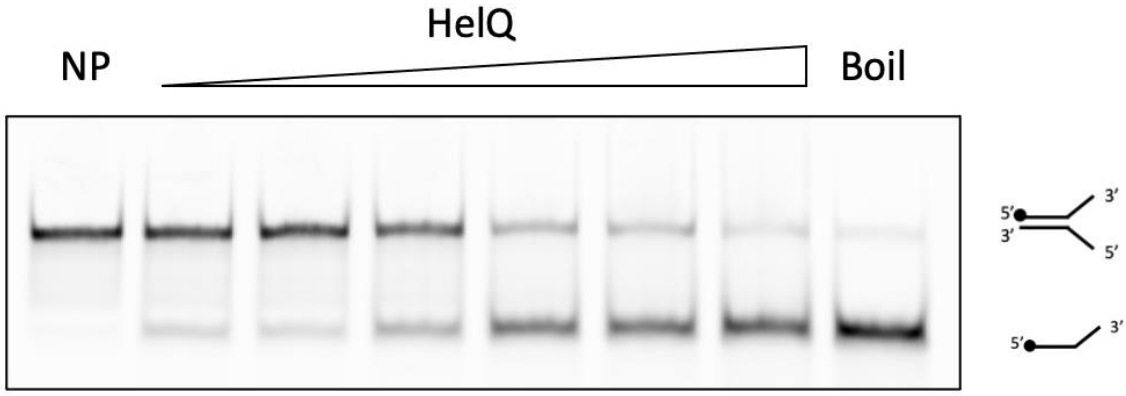


Figure 4.9 Unwinding of forked DNA by HelQ. Wild-type HelQ was assessed in an end-point assay to determine its unwinding activity for forked DNA (25 nM). Concentration titration reactions, with 25, 50, 100, 200, 400 and 800 nM of protein, were performed at 37°C for 10 minutes. A no protein control shows migration of forked DNA and a no protein boiled control "boil" shows ability for the substrate to be unwound and migration of the unwound Cy5-labelled substrate through the gel.

The DNA annealing activity of HelQ was assessed by an end-point, gel-based method and additionally in a time-course assay with FRET techniques.

As before, annealing end-point assays were assessed and visualised on 10% (w/v) polyacrylamide TBE gels. This time however, HelQ's ability to anneal ssDNA with different lengths of complementarity was assessed. This was decided in order to assess how HelQ anneals ssDNA substrates with different lengths of micro-homology comparatively to PolQ; PolQ is a known DNA polymerase that participates in double strand break repair pathways. It has been suggested that while PolQ may function in MMEJ, another protein may be substituted for the helicase function in extended-MMEJ (eMMEJ); that protein being HelQ [65]. We therefore wanted to assess the differences in abilities of HelQ and PolQ in annealing ssDNA with different microhomologies. The same leading 5'Cy5-labelled DNA strand was used in each assay, with a different unlabelled lagging strand having either 10-, 18-, 30- or the full 70-nt length of homology.

Protein overexpression and purification of the *Homo sapiens* PolQ helicase domain was previously achieved in the lab, resulting in functional PolQ (work caried out by He Liu).

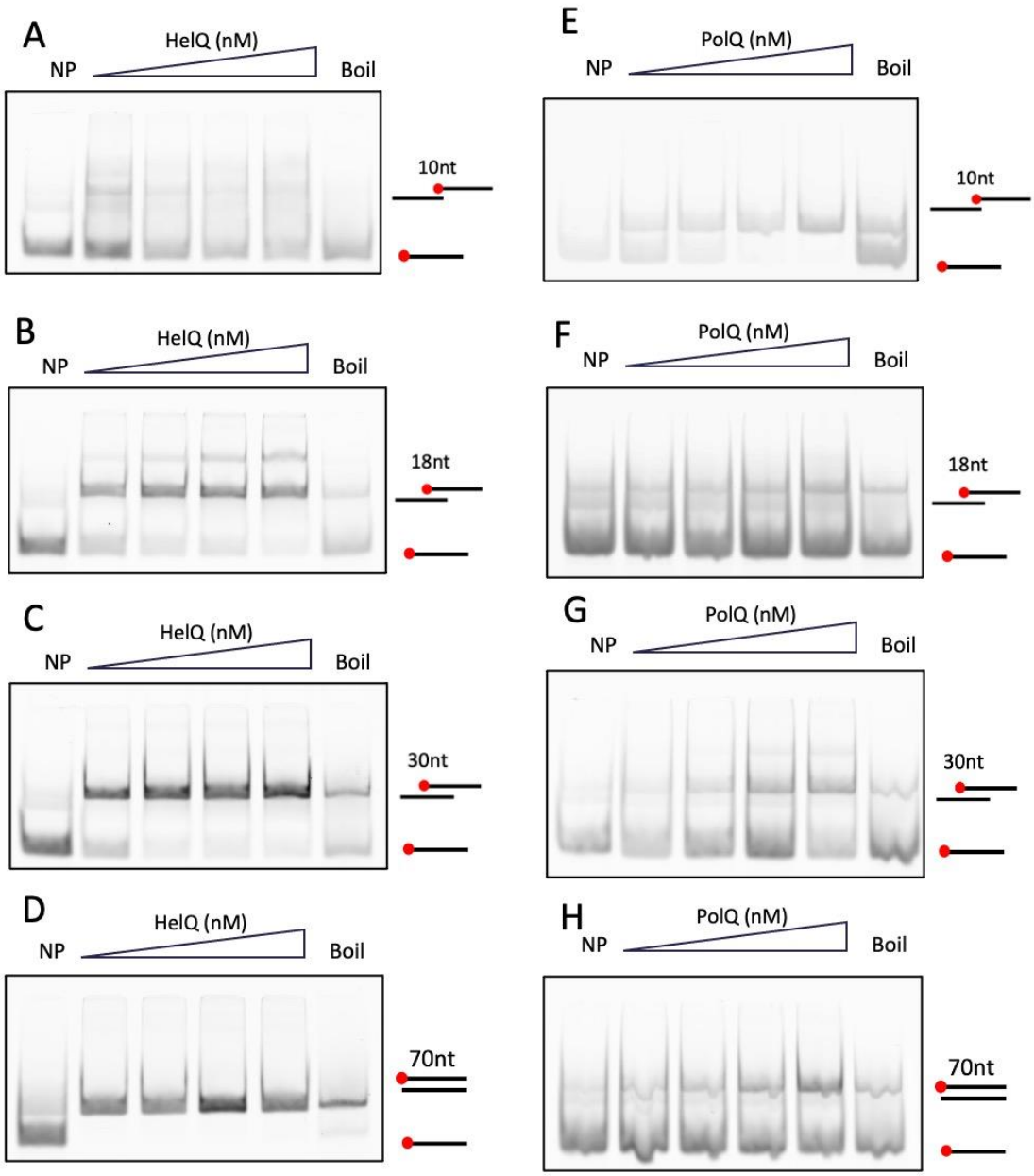


Figure 4.10 HelQ anneals homologies of greater length better than PolQ. (A-D) Wild-type HelQ was assessed in an end-point assay to determine its annealing activity for ssDNA substrates with different lengths of microhomology; 10-, 18-, 30- and the full 70-nt. Concentration titration reactions, with 100, 200, 400 and 800 nM of protein, were performed at 37°C for 5 minutes. A no protein control shows migration of Cy5-labelled DNA and a no protein boiled control “boil” was used to show migration of the wound Cy5-labelled substrate through the gel. (E-H) The same assays were carried out with PolQ at the same protein titration concentrations.

Originally a no protein reaction was “boiled” on a 95°C heat block for 10 minutes to anneal the complementary DNA strands, and used to show where the annealed substrate migrates to on the polyacrylamide TBE gel. However, the annealing of ssDNA was not consistent through application of this method, and in some cases did not result in an annealed product at all in the absence of protein. Following this, substrate preparation to produce a pre-annealed substrate were carried out and this method of making an annealed product was used for a control for all subsequent annealing assays.

Shown in Figure 4.10, HelQ anneals homologies of greater length more effectively than PolQ helicase domain. On the whole, for both proteins, as protein concentration increases, so does annealing activity. In addition, for both proteins, the greater the sequence homology of substrates, the greater the annealing activity displayed. However, it is seen that HelQ had a greater affinity for annealing the substrates of 18-, 30- and the full 70-nt homology than PolQ; suggesting a separate role for HelQ in HR-mediated end-joining processes than PolQ; possibly through eMMEJ.

FRET techniques were then employed to assess HelQ’s annealing activity, giving more precise, real-time results of unwinding activity over a 30-minute time period, incubated at 37°C.

Figure 4.11 confirms HelQ and PolQ are both able to anneal two 70nt homologous ssDNA. FRET anneal values for HelQ were approximately double that value seen for PolQ – highlighting HelQ anneals the 70-nt homology

substrate pair better than PolQ, confirming what we see in gel-based results. This perhaps further emphasises separate roles of HelQ and PolQ in MMEJ.

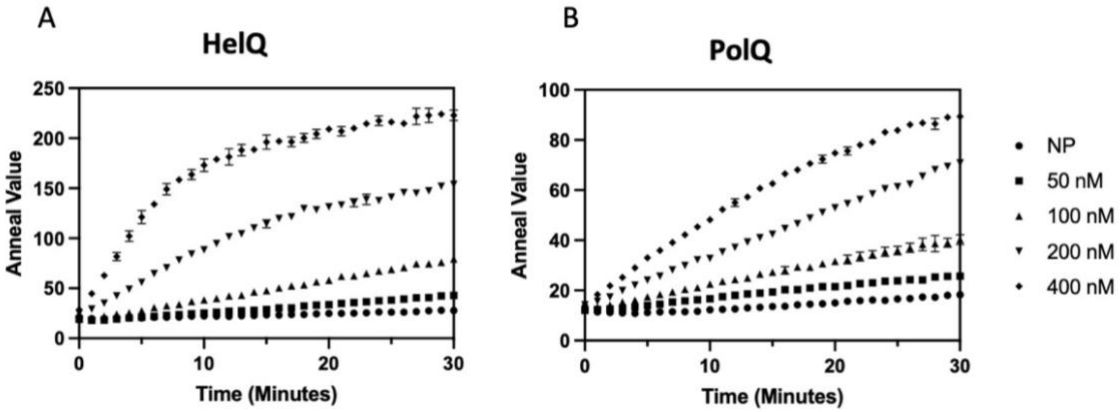


Figure 4.11 HelQ and PolQ anneals complementary 70-nt ssDNA substrates. HelQ (A) and PolQ (B) time-course FRET-assays were carried out at 37°C over 30-minutes, independent of ATP. Concentration titration reactions, with 50, 100, 200, and 400 nM of protein, were performed to anneal two complementary ssDNA species 70-nt in length. “Anneal values” were determined by first correcting FRET readings against the background Cy3 emission, then calculated based on a control reaction with fully annealed substrate.

4.5 Summary

To summarise, Hel308 family proteins have shown to share significant sequence and (predicted) structural homology, giving rise to comparable functions in DNA processing. We see both *M. thermautotrophicus* Hel308 and *H. sapiens* HelQ bind, unwind and anneal DNA substrates. This data has prompted the suggestion of an ancient evolutionary conserved molecular mechanism from archaea to humans in DNA repair and replication. Mutations in Hel308's core helicase domain show modulated DNA processing; with Hel308^{F295A} and Hel308^{Y586A} having the ability to bind and unwind forked DNA hyperactively compared to wild-type Hel308, and additionally Hel308^{F295A} showing hyperactive annealing compared to wild-type.

In addition, we have reported that HelQ has a greater affinity for annealing DNA substrates with greater lengths of microhomology compared to PolQ. With both enzymes being associated with DNA repair and replication, specifically MMEJ, this suggests a role for both enzymes in DSB repair dependant on the length of ssDNA.

Chapter 5 Preliminary analysis of the N-HelQ-PolD interaction

5.1 Introduction

To better understand the DNA strand annealing mechanism of HelQ, we turn to characterising its interaction with other proteins involved in HR and DNA break repair.

Recently, an interaction between N-HelQ and DNA Polymerase Delta (PolD) has been shown to halt DNA strand synthesis and promote DNA single-strand annealing [60]. We know that DNA break repair relies on the synthesis of new DNA by polymerase enzymes in HR and MMEJ processes through extending D-loops [72, 73]. Of which, PolD participates in DNA break repair, recovering cells from stalled or collapsed replication; however, in doing so may also contribute to genome instability through triggering mutagenesis via genetic rearrangements and tandem duplications [74, 75]. Helicase enzymes have been shown to be involved in mechanisms that limit the synthesis of mutagenic DNA during HR [76]. Here they aid the dissociation of D-loops, stimulating the priming of DNA synthesis. He and Lever *et al.* show HelQ halts DNA synthesis by PolD, which instead stimulated DNA annealing of homologous ssDNA by HelQ [60]. They further characterise an interaction between one of the subunits of the PolD complex, PolD3 (Figure 5.1), and N-HelQ. The HelQ interaction with the HR-regulatory subunit of PolD, PolD3, triggers HelQ to anneal DNA strands, a pre-requisite for ending HR so that normal genome duplication can resume.

N-HelQ is the N-terminal 274 amino acid region of HelQ, identified as an IDPR. Previously, it has been shown that N-HelQ is unable to bind DNA and utilises a PWI-like motif to modulate RPA activity [64]. Furthermore, N-HelQ selectively inhibits DNA synthesis by PolD [60].

Important residues within these two proteins involved in facilitating this physical interaction are yet to be explored. A five amino acid sequence within PolD3 was identified to be of interest: 319-KKRRR-323, which we have since named “Quincunx”. This sequence can be found within the largest IDPR of PolD3, IDPR3. This cluster of amino acids was determined to be of interest as clusters of 4-8 basic amino acids containing 4 or more positively charged residues, arginine (R) or lysine (K), are putatively accepted as being classical nuclear localisation signals (cNLS) [77]; therefore indicating this tract of amino acids to be a nuclear localisation signal. It was speculated that mutation of this region would likely impact PolD localisation, and whether this mutation would have any impact on PolD-protein interactions was of interest.

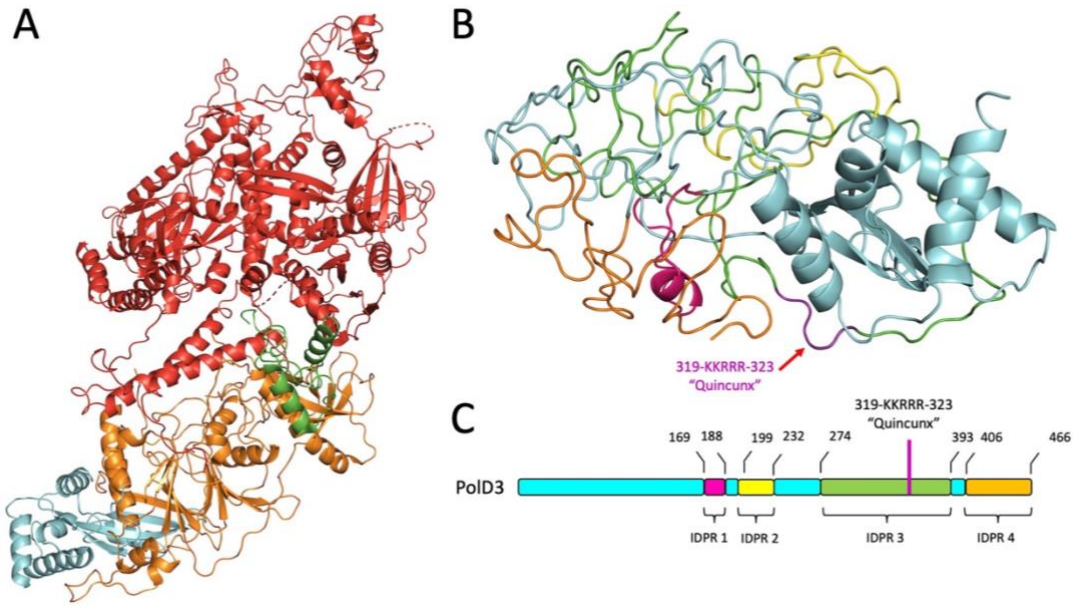


Figure 5.1 PolD Structure Graphic. (A) Structure of PolD complex. Red = PolD1, orange = PolD2, cyan = PolD3, green = PolD4. (B) PyMOL predicted structure of the PolD3 subunit of PolD complex. Intrinsically disordered protein regions (IDPRs) individually highlighted; pink = IDPR 1, yellow = IDPR 2, green = IDPR 3, orange = IDPR 4. KKRRR "Quincunx" motif indicated in purple. (C) Domain organisation of PolD3 subunit. Colour labelling consistent with B.

One assay that has been optimised to detect potential protein-protein interactions are bimolecular fluorescence complementation (BiFC) assays. In summary, this assay works through fusing two proteins to the C-terminal and N-terminal of a split fluorescent protein. Upon interaction of the two proteins, the proximity of the NmVenus and CmVenus being close together causes the full-length mVenus fluorescent reporter protein to be re-assembled, producing a fluorescence signal.

5.2 Cloning and overexpression of mVenus plasmids

5.2.1 Cloning of mVenus plasmids

Previous work in the lab (by Liu He) generated the vector backbones required for cloning work. In this system, mVenus fluorescent protein is split at A154/D155 into two fragment proteins, names NmVenus (155 aa) and CmVenus (84 aa). This site was chosen as it was shown to be effective in a previous study in *E. coli* [78]. The genes encoding the N- and C-terminal portions of the mVenus fluorescent protein (NmVenus = 1-154aa, CmVenus = 155-238aa) were amplified from p63RhoGED619-mVenus-N1 (Addgene, Plasmid #84339) and cloned into pBADHisA via the NheI restriction site and pRSF-1b via the NcoI restriction site, respectively. Constructs were confirmed via sequencing at SourceBioScience, UK. The gene encoding PolD3 was amplified from pPolD3 (GeneArt, ThermoFisher Scientific) and cloned into pBAD-NmVenus via XhoI and HindIII restriction sites, resulting in the pBAD-nMVenus-PolD3 plasmid. The gene encoding N-HelQ was amplified from pSN52 and cloned into pRSF-CmVenus via KpnI and XhoI sites, resulting in the pRSF-CmVenus-NHelQ plasmid.

Here following the same method, I cloned other versions of the PolD and N-HelQ proteins into these vector backbones. The genes encoding various N-HelQ mutants (N-HelQ^{Trun}, N-HelQ^{D142F143}, N-HelQ^{RG3} and N-HelQ^{RG3KG}) were

amplified from their respective plasmids (pTJ11, pTJ10, pTJ09 R51-52G and pTJ09 R51-52G-K42G, respectively) and cloned into pRSF-CmVenus via KpnI and XhoI restriction sites. Additionally, N-HelQ was subject to SDM to generate three other mutants (N-HelQ^{C4A}, N-HelQ^{C24A} and N-HelQ^{C74A}), using pTJ009 as the template, and cloned into pRSF-CmVenus via KpnI and XhoI restriction sites.

The genes encoding the three other subunits of PolD (PolD1, PolD2 and PolD4) were amplified from their respective plasmids (pPolD1, pPolD2 and pPolD4, respectively) and cloned into pBAD-NmVenus via XhoI and HindIII restriction sites. Additionally, a five-residue motif of interest was identified in PolD3: 319-KKRRR-323. SDM of PolD3 to mutagenize these residues to glycine residues was carried out, using pPolD3 as the template, and cloned into pBAD-NmVenus via XhoI and HindIII restriction sites. This mutant was named the “quincunx” mutant.

Analytical digests of all mVenus plasmids confirmed successful cloning of each plasmid (Figure 5.2), which was followed by sequencing for conformation.

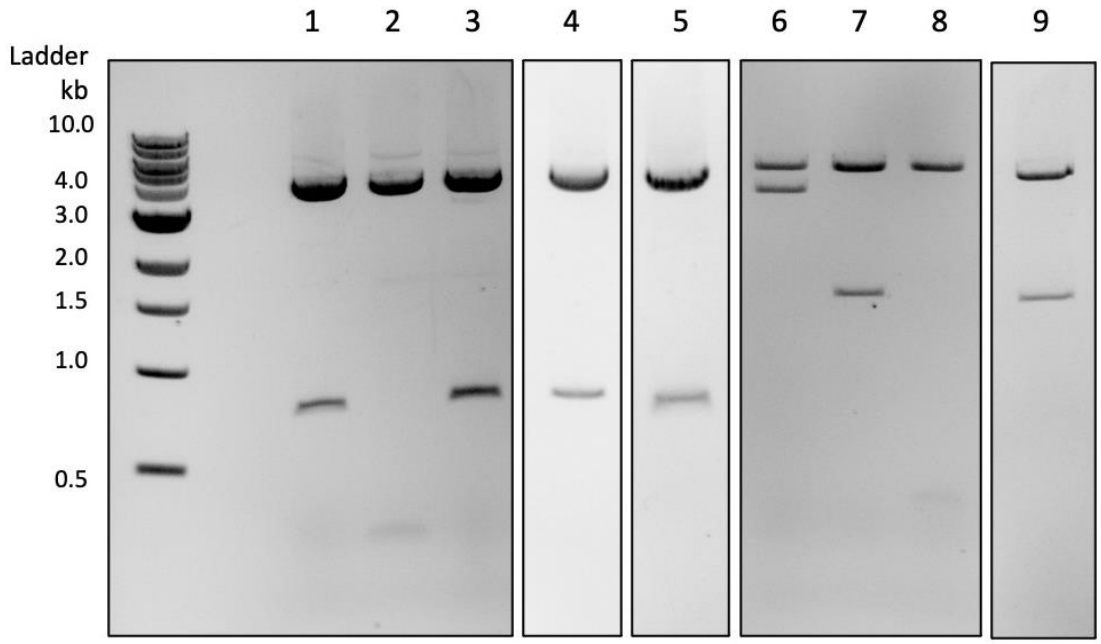


Figure 5.2 Analytical digest of mVenus plasmids confirms insert of correct size in each plasmid. Double digests of mVenus plasmids on 1% (w/v) agarose gel indicates inserts of the correct size were cloned into each vector. Lanes 1-9: Wild-type N-HelQ, N-HelQ^{Trun}, N-HelW^{D142F143}, N-HelQ^{RG3}, N-HelQ^{RG3KG}, PolD1, PolD2, PolD4, PolD3 “quincunx”.

N-HelQ-containing plasmid digests were expected to have a band at 3829 bp (the size of pRSF-CmVenus) and a second band at 732 bp, (or 240 bp for N-HelQ^{Trun}). For PolD plasmids, plasmid digests were expected to have a band at 4532 bp (the size of pBAD-NmVenus) and a second band the size of each PolD subunit: PolD1 = 3333 bp, PolD2 = 1419 bp, PolD4 =333 bp and PolD3 “quincunx” = 1432 bp. Bands of the correct size can be seen for each plasmid (Figure 5.2). Bands of smaller size (i.e. N-HelQ^{Trun} and PolD4) are less obviously seen on the agarose gel, but nevertheless are seen.

5.2.2 Protein overexpression of mVenus plasmids

Protein expression of each individual plasmid construct was necessary before assessing any interactions between proteins. For pilot protein overexpression, each vector construct was transformed into *E. coli* BL21AI competent cells, overnight cultures were inoculated into fresh LB with respective antibiotic in baffled flasks, incubated at 37°C with shaking until OD600 = 0.5, induced with 0.2% (w/v) L-arabinose and 1mM IPTG and left to shake for an additional 3 hours at 37°C.

Figure 5.3 of Coomassie stained SDS-PAGE analysis shows all PolD proteins were overexpressed. However, the N-HelQ protein over-expressions were not visible via Coomassie staining. N-HelQ protein expression was followed with western blot analysis for visualisation of protein expression.

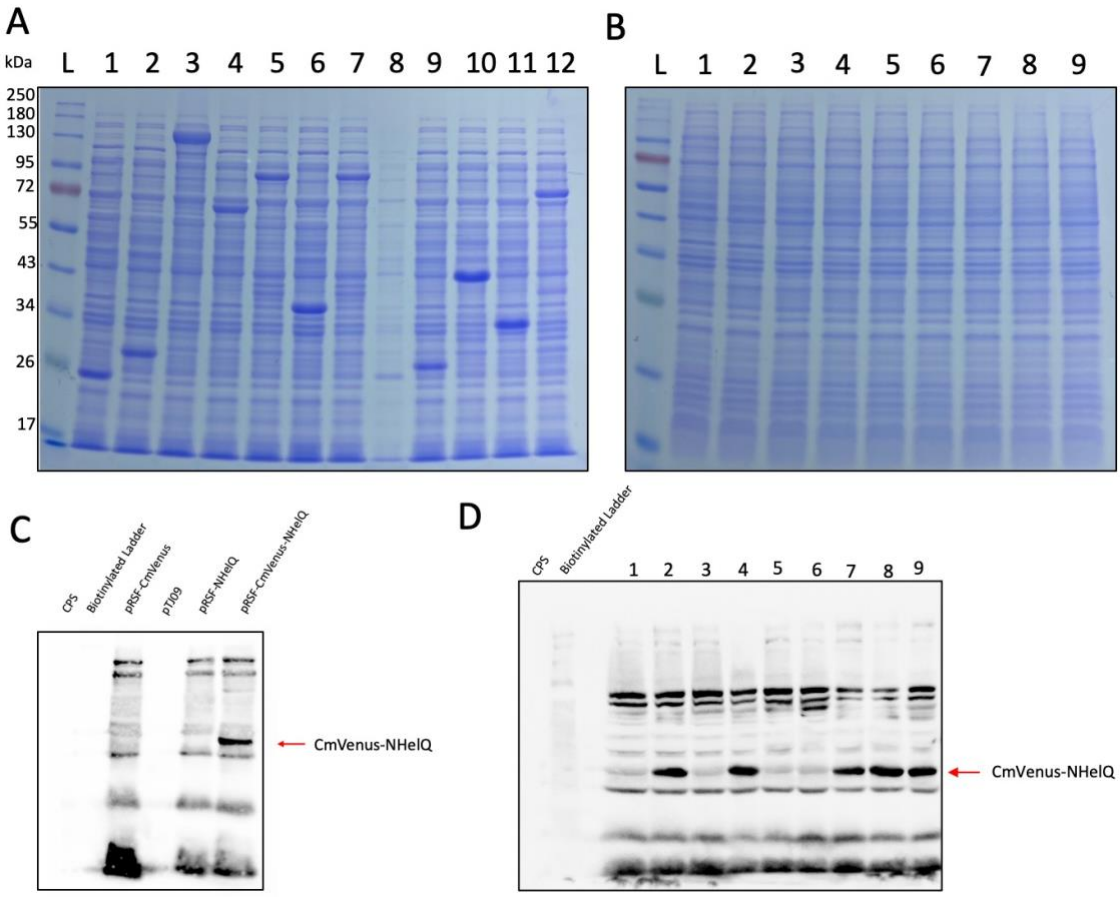


Figure 5.3 Pilot protein overexpression of mVenus plasmids. SDS-PAGE analysis shows successful overexpression of (A) PolD mVenus plasmids (10% (w/v) polyacrylamide SDS-PAGE gel). Lane 1-12: pBAD-NmVenus-PolD3aa1-106, -PolD3aa107-466, -PolD1, -PolD2, -PolD3, -PolD4, -PolD3 ”quincunx”, -PolD3_IDPR1, -PolD3_IDPR2, -PolD3_IDPR3, -PolD3_IDPR4 and pBAD-FLmVenus. Expression of N-HelQ mVenus plasmids could not be detected with Coomassie staining on12% polyacrylamide SDS-PAGE gel (B). Western blot, using primary anti-N-HelQ (rabbit) and secondary goat anti-rabbit and anti-biotin, was conducted on N-HelQ mVenus plasmids. A preliminary western blot (C) to detected presence of pRSF-CmVenus-NHelQ, against necessary controls. Expression of remaining N-HelQ mVenus plasmids (D) shows expression of six of the eight constructs. Lane1-9: pRSF-CmVenus (empty vector control), pRSF-CmVenus-NHelQ, -NHelQ^{Trun}, -NHelQ^{D142F143}, - N-HelQ^{RG3}, -N-HelQ^{RG3KG}, -NHelQ^{C4A}, -NHelQ^{C24A} and -NHelQ^{C72A}.

5.3 Bimolecular fluorescence complementation (BiFC) assay

Following successful cloning and overexpression of mVenus plasmids, interactions between protein pairs could be assessed. A testing pairs of plasmids were co-transformed into *E. coli* BL21 AI competent cells, protein expression was induced as before, and a 100 μL sample was out at regular time points across 3-4 hours to assess the interaction, if any, between the two proteins. Samples were spun-down to remove LB broth and resuspended in 150 μL ice-cold PBS before being transferred into a flat-bottomed 96-well plate for imaging at 515 nm with the Amersham Typhoon laser-scanner platform with Cy2 filter.

The physical interaction between wild-type PolD3 / PolD3^{Quincunx} mutant and N-HelQ / N-HelQ cysteine mutants was assessed to determine the effect of each mutation on protein-protein binding. Inducible co-expression of CmVenus fused to N-HelQ and NmVenus fused to wild-type PolD3 triggered fluorescence at 210 minutes – confirming the previously described N-HelQ-PolD3 interaction (Figure 5.4). NHelQ^{C4A}, NHelQ^{C24A} and NHelQ^{C72A} mutations did not have any effect on this interaction. Fluorescence was not observed in the CmVenus and NmVenus independent controls, or in co-expression with one protein of the testing pair missing. The N-HelQ-PolD3 interaction was found to be negated by the PolD3 Quincunx mutant – implicating the 319-KKRRR-323 motif to be important for the physical interaction with N-HelQ (Figure 5.4).

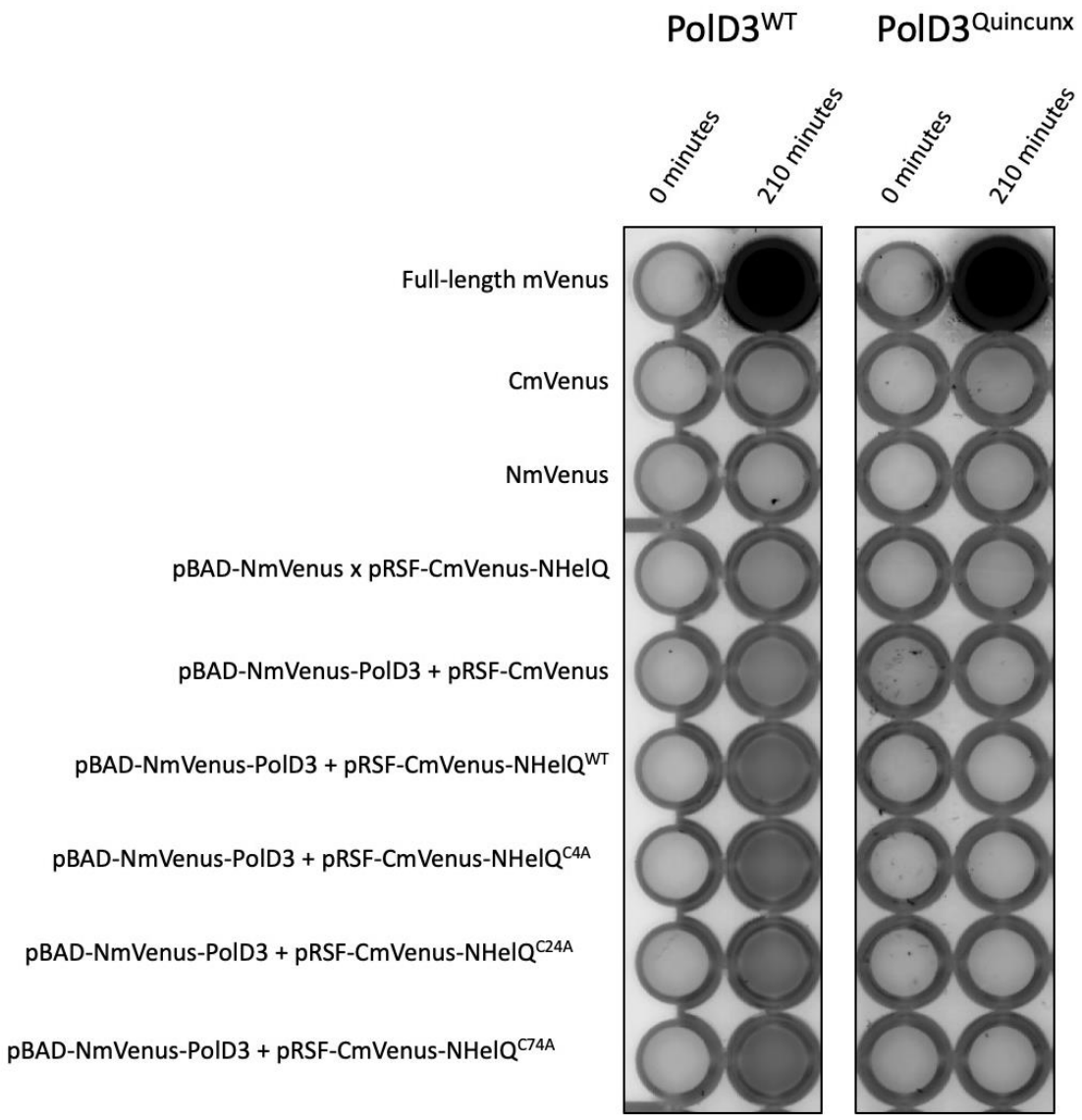


Figure 5.4 The physical interaction between PolD3 and N-HelQ is negated by PolD3^{Quincunx} mutation. BiFC assay, assessing the physical interaction between N-HelQ and PolD3 and mutants in *E. coli,* detectable in 96-well plates samples co-expressing NmVenus-PolD3 and CmVenus-NHelQ fusion proteins, indicate the mutation of the 319-KKRRR-323 motif in PolD3 to 319-GGGGG-323 inhibits the physical interaction between the two proteins. Other co-expression controls, partial mVenus protein controls and full-length mVenus control are indicated: CmVenus and nMvenus alone, CmVenus with NmVenus-PolD3 and NmVenus with CmVenus-NHelQ, and full mVenus as a positive control for continuous fluorescence.

5.4 Summary

Here we show preliminary data into characterising the interaction between N-HelQ and PolD, more specifically the third domain, PolD3. Mutagenesis of the 319-KKRRR-323 (Quincunx) motif of PolD3, which lies within its third intrinsically disordered region (IDPR3), inhibits its interaction with N-HelQ. This eludes to the KKRRR motif of PolD3 being significant for this interaction.

Further analysis of protein interactions between the other mVenus plasmids which were cloned will need to be explored further to establish a molecular mechanism for protein interaction between N-HelQ and PolD3.

Chapter 6 Discussion and Future Work

The aims of this study were to assess the DNA processing activity of wild-type and mutant C-HelQ and propose whether there is evidence to suggest an evolutionary conserved molecular mechanism from archaea to humans of the Hel308 family proteins role in DNA replication and repair. This initial data could be used to aid the proposition a molecular mechanism for how HelQ anneals DNA, and identify functional important residues involved in the process. Additionally, it was of interest to identify potential protein-protein interactions of HelQ; we here turn to build upon recently published literature describing the interaction between the N-terminus of HelQ and PolD.

6.1 DNA annealing by C-HelQ

Helicase unwinding assays in Chapter 3 indicate C-HelQ's unwinding activity is negatively modulate by Y642A and P960A/Y963A mutations, making it less effective as a helicase. However, the opposite result is seen for its annealing activity, with both mutations individually increasing DNA annealing activity of C-HelQ. When the binding affinity of DNA by C-HelQ is assessed, EMSA analysis shows binding of forked DNA by C-HelQ and the two mutants is near identical. This is however not the same when we assess the C-HelQ binding affinity of ssDNA substrates in anisotropy assays. Rather, this indicated Y642A and P960A/Y963A mutations inhibit DNA binding activity of C-HelQ.

Sense of this data could be made by proposing the mutations in C-HelQ cause a conformational change to its structure, giving it a more open conformation.

Molecular dynamics simulations and analyses of Hel308 have shown the Hel308^{F295A} mutant, the archaeal homolog of C-HelQ^{Y642A}, has a more open conformation compared to wild-type Hel308 [68]. Specifically, the distance between the F295 and Y586 residues (in Domain II and the ratchet in Domain IV, respectively) was nearly doubled with the F295A mutation. This mutation also resulted in a hyper annealing phenotype. C-HelQ^{Y642A} having a more open conformation could be used to explain the hyper-annealing activities demonstrated. An open conformation would make the protein more suited to DNA annealing, having more space to navigate two ssDNA species for annealing reactions. This open conformation may also result is less stable DNA-protein complexes, explaining why anisotropy analyses identify mutant protein to bind less stably than wild-type C-HelQ and therefore also exhibit decreased DNA unwinding activity.

It is possible this proposal may be supported by the preliminary results seen from nuclease protection assays. Assays assessing C-HelQ's ability to protect ssDNA from S1 nuclease digestion hint at the idea that the Y642A mutation has decreased protection from the nuclease compared to wild-type C-HelQ. This could suggest the mutant has a more open conformation and so the nuclease can access the ssDNA for digestion more easily. This nuclease protection data us not very robust – it is presented as a single assay, without replicates and has not been quantified. Replicates and quantification would be necessary to conduct statistical analysis. However, this preliminary data suggests there are grounds for future work pursuing this apparent lesser protection of ssDNA by C-HelQ^{Y642A} compared to wild-type C-HelQ, although this is not clear as of yet.

More experimentation is required to facilitate this prediction. ATPase assays could provide insight into the ATPase activity of mutant C-HelQ compared to wild-type and functionally characterise the proteins and could be used to confirm whether the Y642A mutation to C-HelQ causes a structural change, resulting in a more open conformation. Also, FRET data showing the modulated annealing activity the P960A/Y963A mutation has on C-HelQ would be beneficial. Additionally, *in vivo* experimentation via prime editing would provide insight into the cellular effects of these mutations – which is already in development.

It should be noted that EMSA and anisotropy analyses show differences in comparative binding affinities of C-HelQ proteins. While EMSAs show wild-type C-HelQ and mutant C-HelQ have approximately the same binding affinity, anisotropy does show the mutants have decreased binding affinity. DNA substrates used in these two analyses were different; EMSAs used Cy5-labeled forked DNA, resembling stalled replication forks, whereas anisotropy used FAM-labelled Poly(T)ssDNA. EMSA and anisotropy assays displaying different results isn't necessarily unusual, especially considering different DNA substrates were used. The differences between their method of assessment – with anisotropy results are read on a 96-well plate instantly compared to EMSAs in which stable protein-DNA complexes having to survive being loaded and ran through a gel – could be the cause of the varied results. Alternatively, this could also hint at the idea of mutant C-HelQ having a more open conformation – it is possible the mutant requires the binding of two strands of DNA to form a stable

complex, which it then pulls together to anneal. Identifying possible reasons for this difference would need to be illuminated through further experimentation – but at this point we are unable to confidently say why. I would propose anisotropy using a FAM-labelled forked DNA substrate would allow a more representative comparison with how C-HelQ binds forked DNA compared to ssDNA, and therefore also assessing how the two mutations may modulate this.

6.2 Comparing Hel308 and C-HelQ mutant activity

It is interesting to note the difference in DNA processing activities between comparable mutants in Hel308 and C-HelQ. In C-HelQ, we have demonstrated the Y642A mutation in the RecA domain results in decreased DNA binding and unwinding activity and increase DNA annealing activity. However, the same mutation in *M. thermautotrophicus* Hel308, F295A, display increased DNA binding, unwinding and annealing activity. Likewise, the P960A/Y963A mutation in the DNA ratchet of domain IV in C-HelQ also results in decreased DNA binding and unwinding activity and increase DNA annealing activity. Yet in Hel308, the comparative Y586A mutation shows increased DNA unwinding. While these two proteins share significant sequence and predicted structural homology (as demonstrated), it is possible the non-homologous regions could be responsible for this difference in activity, making mutant C-HelQ better at DNA annealing but worse at DNA unwinding. The most obvious sequential difference between the two proteins would be the presence of an additional C-terminal Domain V in C-HelQ. Further experimentation to assess the activity of this domain would be required to deduce if it is of any importance to the DNA binding, unwinding and annealing mechanism. This data presented does

however elude to a, at least, partial evolutionary conserved mechanism for DNA annealing by Hel308 family proteins as likewise mutations result in a comparable DNA hyper annealing phenotype.

6.3 Suggested role of HelQ and PolQ in MMEJ

To aid the assessment of HelQ's DNA annealing mechanism and its suggested role in DNA break repair pathways, understanding its activity in relation to PolQ, another known DSB repair enzyme, was of interest. Here, we show data suggesting HelQ has a higher affinity for annealing micro-homologies of greater length than PolQ – demonstrated through a comparative analysis of HelQ and PolQ annealing 10-, 18-, 30- and 70-nt microhomologies in gel-based assays, and additionally with assessing the two proteins ability to anneal two fully complementary 70nt ssDNA substrates using FRET techniques. FRET anneal values show greater annealing of the 70nt ssDNA substrates compared to PolQ. This could indicate separate roles of HelQ and PolQ in HR-mediated break repair, and explain the necessary existence of both proteins within homologous recombination – with PolQ functioning in MMEJ and HelQ required for eMMEJ. Further assessment of the ability of both proteins to anneal very short microhomologies (for example optimising assays for annealing 2-, 4-, 6-, 8-nt complementary regions of ssDNA) would provide further detail in supporting this theory.

6.4 Characterising the N-HelQ-PolD3 interaction

Following recently published research identifying a physical interaction between the N-terminal intrinsically disordered region of HelQ and the largest

subunit of the PolD complex, PolD3, we here describe preliminary data characterising the physical interaction between N-HelQ and PolD3 to be located within the 319-KKRRR-323 motif within IDPR3 of PolD3 to be significant – SDM of this region negates the interaction between the two proteins. This cluster of amino acids was determined to be a classical nuclear localisation signal which would likely impact PolD localisation when mutagenized. This could be used to annotate the molecular mechanism by which these proteins interact within the context of homologous recombination and DNA repair. Data presented would be required to be quantified and replicates conducted to conduct statistical analysis on the physical interaction.

The use of BiFC assays to continue to detect physical protein-protein interactions would need to be explored further, using other interaction pairs of mutants of N-HelQ and PolD3 to locate additional important residues for this interaction; and its mechanism within the context of homologous recombination to be addressed further.

Chapter 7 Bibliography

[1] K. A. Schafer, “The cell cycle: a review.,” *Vet Pathol.,* Vols. 35(6):461-78., 1998..

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, “The Structure and Function of DNA.,” (2002)..

[3] L. L. S. B. O'Donnell M, “Principles and concepts of DNA replication in bacteria, archaea, and eukarya.,” *Cold Spring Harb Perspect Biol.,* vol. 5(7):a010108. , 2013..

[4] J. Sun, Y. Shi, R. E. Georgescu, Z. Yuan, B. T. Chait, H. Li and M. E. O'Donnell, “The architecture of a eukaryotic replisome.,” Vols. 22(12):976-82., 2015..

[5] J. D. Watson and F. H. Crick, “Genetical implications of the structure of deoxyribonucleic acid.,” Vols. 171(4361):964-7., 1953.

[6] M. O'Donnell, L. Langston and B. Stillman, “Principles and concepts of DNA replication in bacteria, archaea, and eukarya.,” vol. 5(7):a010108., 2013.

[7] P. M. Burgers and T. A. Kunkel, “Eukaryotic DNA Replication Fork.,” Vols. 86:417-438., 20 June 2017.

[8] M. R. Kelley, D. Logsdon and M. L. Fishel, “Targeting DNA repair pathways for cancer treatment: what's new?,” Vols. 10(7):1215-37, 2014.

[9] M. Meselson and F. W. Stahl, “The replication of DNA in Escherichia coli.,” *Proc. Natl. Acad. Sci.,* Vols. 44, 671–682, 1958.

[10] A. R. Leman and E. Noguchi, “The Replication Fork: Understanding the Eukaryotic Replication Machinery and the Challenges to Genome Duplication,” *Genes (Basel).,* Vols. 4, 1–32., 2013..

[11] N. Yao and M. O'Donnell, “Bacterial and Eukaryotic Replisome Machines.,” vol. 3(1):1013., 2016.

[12] G. S. Briggs, W. K. Smits and P. Soultanas, “Chromosomal replication initiation machinery of low-G+C-content Firmicutes.,” *J Bacteriol.,* vol. 194, no. 19, pp. 5162-70., 2012.

[13] K. J. Marians, “Understanding how the replisome works.,” *Nat. Struct. Mol. Biol.,* vol. 15; 125–127., 2008..

[14] N. Y. Yao and M. O'Donnell, “SnapShot: The replisome.,” *Cell.,* Vols. 141(6):1088, 1088.e1., 2010.

[15] P. V. Harris, O. M. Mazina, E. A. Leonhardt, R. B. Case, J. B. Boyd and K. C. Burtis, “Molecular cloning of Drosophila mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes.,” *Mol Cell Biol,* Vols. 16(10):5764-71., 1996.

[16] J. Snider, G. Thibault and W. A. Houry, “The AAA+ superfamily of functionally diverse proteins.,” *Genome Biol.,* vol. 9, no. 4, p. 216, 2008.

[17] W. Messer, “The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication.,” *FEMS Microbiol Rev.,* Vols. 26(4):355-74., 2002..

[18] M. P. Martinez, J. M. Jones, I. Bruck and D. L. Kaplan, “Origin DNA melting-An essential process with divergent mechanisms.,” *Genes (Basel).,* Vols. 8, 1–13., 2017..

[19] W. Messer, “The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication.,” *FEMS Microbiol. Rev.,* Vols. 26, 355–374., 2002..

[20] D. Ausiannikava and T. Allers, “Diversity of DNA replication in the archaea.,” *Genes 8,* 2017.

[21] F. Bleichert, M. R. Botchan and J. M. Berger, “Crystal structure of the eukaryotic origin recognition complex.,” *Nature.,* Vols. 519, 321–6., 2015..

[22] Z. Yuan, L. Bai, J. Sun, R. Georgescu, J. Liu, M. E. O'Donnell and H. Li, “Structure of the eukaryotic replicative CMG helicase suggests a pumpjack motion for translocation.,” *Nat. Struct. Mol. Biol.,* vol. 23., 2016..

[23] L. Balakrishnan and R. A. Bambara, “Okazaki fragment metabolism. Cold Spring Harb.,” *Perspect. Biol.,* Vols. 5, a010173., 2013..

[24] K. Lee and K. Myung, “PCNA modifications for regulation of post-replication repair pathways.,” *Mol. Cells.,* Vols. 26, 5–11 ., 2008..

[25] K. Labib and B. Hodgson, “Replication fork barriers: pausing for a break or stalling for time?,” *EMBO Rep.,* Vols. 8, 346–353, 2007.

[26] S. Saxena and L. Zou, “Hallmarks of DNA replication stress.,” Vols. 82(12):2298-2314., 16 June 2022.

[27] V. Pagès and R. P. Fuchs, “How DNA lesions are turned into mutations within cells?,” *Oncogene.,* Vols. 21, 8957–8966, 2002..

[28] R. Anand, E. Buechelmaier, O. Belan, M. Newton, A. Vancevska, A. Kaczmarczyk, T. Takaki, D. S. Rueda, S. N. Powell and S. Boulton,

“HELQ is a dual-function DSB repair enzyme modulated by RPA and RAD51.,” *Nature.,* Vols. 601(7892):268-273., 2022.

[29] J. Hoeijmakers, “Genome maintenance mechanisms for preventing cancer.,” *Nature.,* Vols. 411, 366–374., 2001..

[30] J. G. Brüning, J. L. Howard and P. McGlynn, “Accessory replicative helicases and the replication of protein-bound DNA.,” *J Mol Biol.,* Vols. 426(24):3917-3928., 12 December 2014.

[31] M. J. Florés, N. Sanchez and B. Michel, “A fork-clearing role for UvrD.,” *Mol. Microbiol.,* Vols. 57, 1664–1675., 2005..

[32] G. Ghosal and J. Chen, “DNA damage tolerance: a double-edged sword guarding the genome.,” *Transl Cancer Res.,* Vols. 2(3):107-129., 2013.

[33] M. S. Cooke, M. D. Evans, M. Dizdaroglu and J. Lunec, “Oxidative DNA damage: mechanisms, mutation, and disease.,” *FASEB J.,* vol. 17, no. 10, pp. 1195-214., 2003.

[34] N. Klages-Mundt and L. Li, “Formation and repair of DNA-protein crosslink damage.,” *Sci. China Life Sci.,* vol. 60, no. 1065–1076, 2017.

[35] R. Shah, “DNA : Damage Types and Repair Mechanisms (With Diagram). Biology Discussion.,” [Online]. Available: http://www.biologydiscussion.com/dna/dna-damage-types-and-repair-mechanisms-with-diagram/16332..

[36] R. Roy, J. Chun and S. N. Powell, “BRCA1 and BRCA2: different roles in a common pathway of genome protection.,” *Nat Rev Cancer.,* Vols. 12(1):68-78., 2011.

[37] J. S. Sung and B. Demple, “Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA.,” *FEBS Journal.,* Vols. 273, 1620–1629, 2006..

[38] H. Wang and J. B. Hays, “Human DNA mismatch repair: Coupling of mismatch recognition to strand-specific excision.,” *Nucleic Acids Res.,* Vols. 35, 6727–6739., 2007..

[39] F. Zhao, W. Kim, J. A. Kloeber and Z. Lou, “DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells.,” *Exp Mol Med.,* vol. 52, no. 10, pp. 1705-1714., 2020.

[40] X. Li and W. D. Heyer, “Homologous recombination in DNA repair and DNA damage tolerance.,” *Cell Res.,* Vols. 18, 99–113., 2008..

[41] T. Nakano, S. Morishita, A. Katafuchi, M. Matsubara, Y. Horikawa, H. Terato, A. M. Salem , S. Izumi, S. P. Pack, K. Makino and H. Ide, “Nucleotide Excision Repair and Homologous Recombination Systems Commit Differentially to the Repair of DNA-Protein Crosslinks.,” *Mol. Cell.,* Vols. 28, 147–158., 2007..

[42] J. Vignard, G. Mirey and B. Salles, “Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up.,” *Radiotherapy and Oncology.,* Vols. 108, 362– 369., 2013..

[43] M. Spies, M. S. Dillingham and S. C. Kowakzykowski, “Translocation by the RecB motor is an absolute requirement for χ-recognition and RecA protein loading by RecBCD enzyme.,” *J. Biol. Chem.,* Vols. 280, 37078–37087., 2005..

[44] K. Rodgers and M. McVey, “Error-Prone Repair of DNA Double-Strand Breaks.,” *J Cell Physiol.,* Vols. 231(1):15-24., 2016.

[45] A. K. Byrd and K. D. Raney, “Superfamily 2 helicases.,” Vols. 17(6):2070-88., 1 June 2012.

[46] J. Snider, G. Thibault and W. A. Houry, “The AAA+ superfamily of functionally diverse proteins.,” vol. 9(4):216., 30 April 2008.

[47] A. K. Byrd and K. D. Raney, “Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA.,” *Nat Struct Mol Biol.,* Vols. 11, 531–538., 2004.

[48] M. E. Fairman-Williams, U. P. Guenther and E. Jankowsky, “SF1 and SF2 : family matters.,” Vols. 20, 313–324., 2010..

[49] M. R. Singleton, M. S. Dillingham and D. B. Wigley, “Structure and mechanism of helicases and nucleic acid translocases.,” *Ann Rev Biochem.,* vol. 76:23–50., 2007..

[50] J. Ye, A. R. Osborne, M. Groll and T. A. Rapoport, “RecA-like motor ATPases--lessons from structures.,” *Biochim Biophys Acta.,* Vols. 1659(1):1-18., 4 November 2004..

[51] S. J. Johnson and R. N. Jackson, “Ski2-like RNA helicase structures Common themes and complex assemblies.,” *RNA Bilogy,* Vols. 10, 33–43., 2013.

[52] J. B. Boyd, M. D. Golino, T. D. Nguyen and M. M. Green, “Isolation and characterization of X-linked mutants of Drosophila melanogaster which are sensitive to mutagens.,” *Genetics.,* Vols. 84(3):485-506., November 1976.

[53] F. Marini and R. D. Wood, “A human DNA helicase homologous to the DNA cross- link sensitivity protein Mus308.,” *J. Biol. Chem.,* Vols. 277, 8716–8723., 2002..

[54] J. B. Boyd, K. Sakaguchi and P. V. Harris, “mus308 Mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease.,” *Genetics.,* Vols. 125, 813–819., 1990..

[55] J. D. Richards, K. A. Johnson, H. Liu, A. M. McRobbie, S. McMahon, M. Oke, L. Carter, J. H. Naismith and M. F. White, “Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains.,” *J Biol Chem.,* Vols. 283(8):5118-26., Feb 2008.

[56] J. H. Yoon, J. Roy Choudhury, J. Park, S. Prakash and L. Prakash, “A role for DNA polymerase θ in promoting replication through oxidative DNA lesion, thymine glycol, in human cells.,” *J Biol Chem.,* Vols. 289(19):13177-85., 9 May 2014..

[57] R. D. Wood and S. Doublié, “DNA polymerase θ (POLQ), double-strand break repair, and cancer.,” *DNA Repair (Amst).,* Vols. 44:22-32, August 2016..

[58] F. Marini, N. Kim, A. Schuffert and R. D. Wood, “POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308.,” *J Biol Chem.,* Vols. 278(34):32014-9., 22 August 2003.

[59] C. P. Guy and E. L. Bolt, “Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands.,” *Nucleic Acids Res.,* Vols. 33(11):3678-90., 30 June 2005.

[60] L. He, R. Lever, A. Cubbon, M. Tehseen, T. Jenkins, A. O. Nottingham, A. Horton, H. Betts, M. Fisher, S. M. Hamdan, P. Soultanas and E. L. Bolt, “Interaction of human HelQ with DNA polymerase delta halts DNA synthesis and stimulates DNA single-strand annealing.,” *Nucleic Acids Res.,* Vols. 51(4):1740-1749., 28 February 2023.

[61] D. M. Muzzini, P. Plevani, S. J. Boulton, G. Cassata and F. Marini, “Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways.,” *DNA Repair (Amst).,* Vols. 7(6):941-50., 1 June 2008.

[62] G. L. Moldovan and A. D. D’Andrea, “How the fanconi anemia pathway guards the genome.,” *Annu. Rev. Genet.,* Vols. 43, 223–49., 2009..

[63] S. J. Northall, “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.

[64] T. Jenkins, S. J. Northall, D. Ptchelkine, R. Lever, A. Cubbon, H. Betts, V. Taresco, C. D. Cooper, P. J. McHugh, P. Soultanas and E. L. Bolt, “The HelQ human DNA repair helicase utilizes a PWI-like domain for DNA loading through interaction with RPA, triggering DNA unwinding by the HelQ helicase core.,” *NAR Cancer.,* vol. 3(1):zcaa043., 12 January 2021.

[65] J. A. Kamp, B. Lemmens, R. J. Romeijn, S. C. Changoer, R. van Schendel and M. Tilsterman, “Helicase Q promotes homology-driven DNA double-strand break repair and prevents tandem duplications.,” *Nat Commun.,* vol. 12(1):7126., 8 December 2021.

[66] C. Madru, G. Henneke , P. Raia, I. Hugonneau-Beaufet, G. Pehau-Arnaudet, P. England, E. Lindahl, M. Delarue, M. Carroni and L. Sauguet, “Structural basis for the increased processivity of D-family DNA polymerases in complex with PCNA.,” *Nat Commun.,* vol. 11, no. 1, p. 1591, 2020.

[67] M. Sebesta, P. Burkovics, S. Juhasz, S. Zhang, J. E. Szabo, M. Y. Lee, L. Haracska and L. Krejci, “Role of PCNA and TLS polymerases in D-loop extension during homologous recombination in humans.,” *DNA Repair (Amst).,* vol. 12, no. 9, pp. 691-8., 2013.

[68] R. J. Lever, E. Simmons, R. Gamble-Milner, R. J. Buckley, C. Harrison, A. J. Parkes, L. Mitchell, J. A. Gausden, S. Skulj, B. Bertosa, E. L. Bolt and T. Allers, “Archaeal Hel308 suppresses recombination through a catalytic switch that controls DNA annealing.,” *Nucleic Acids Res.,* Vols. 51(16):8563-8574., 2023.

[69] T. Jenkins, “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.,* 2020.

[70] R. Favicchio, A. I. Dragan, G. G. Kneale and C. M. Read, “Fluorescence spectroscopy and anisotropy in the analysis of DNA-protein interactions.,” *Methods Mol Biol.,* Vols. 543:589-611., 2009.

[71] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughbrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwcki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E.

Wright and M. M. Babu, “Classification of intrinsically disordered regions and proteins.,” *Chem Rev.,* Vols. 114(13):6589-631., 2014..

[72] M. McVey, V. Y. Khodaverdian, D. Meyer, P. G. Cerqueira and W. D. Heyer, “Eukaryotic DNA polymerases in homologous recombination.,” *Annu. Rev. Genet.,* Vols. 50, 393–421., 2016.

[73] R. A. Donnianni, Z. X. Zhou, S. S. Lujan, A. Al-Zain, V. Garcia, E. Glancy, A. B. Burkholder, T. A. Kunkel and L. S. Symington, “DNA polymerase delta synthesizes both strands during break-induced replication.,” *Mol. Cell.,* Vols. 76, 371–381., 2019..

[74] C. J. Sakofsky, S. Ayyar, A. K. Deem, W. H. Chung, G. Ira and A. Malkova, “Translesion polymerases drive microhomology-mediated break-induced replication leading to complex chromosomal rearrangements.,” *Mol. Cell.,* Vols. 60, 860–872., 2015.

[75] C. E. Smith, A. F. Lam and L. S. Symington, “Aberrant double-strand break repair resulting in half crossovers in mutants defective for Rad51 or the DNA polymerase delta complex.,” *Mol. Cell. Biol.,* Vols. 29, 1432–1441., 2009.

[76] E. Huselid and S. F. Bunting, “The regulation of homologous recombination by helicases.,” *Genes (Basel).,* Vols. 11, 498–505., 2020..

[77] J. Lu, T. Wu, B. Zhang, S. Liu, W. Song, J. Qiao and H. Ruan, “Types of nuclear localization signals and mechanisms of protein import into the nucleus.,” *Cell Commun Signal.,* vol. 19, no. 1, p. 60, 2021.

[78] K. E. Dillard, M. W. Brown, N. V. Johnson, Y. Xiao, A. Dolan, E. Hernandez, S. D. Dahlhauser, Y. Kim, L. R. Myler, E. V. Anslyn, A. Ke

and I. J. Finkelstein, “Assembly and translocation of a CRISPR-Cas primed acquisition complex.,” *Cell.,* Vols. 175, 934–946., 2018.

[79] A. F. Palazzo and T. R. Gregory, “The case for junk DNA.,” *PLoS Genet.,* vol. 10(5):e1004351., 2014..

Chapter 8 Supplementary Data

MAHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIF
FKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDL
DMEDNDIIEAHREQISSGLEVLFQGPDECGSRIRRRVSLPKRNR
PSLGCIFGAPTAAELEPGDEGKEEEEMVAENRRRKTAGVLPVEV
QPLLLSDSPECLVLGGGDTNPDLLRHMPTDRGVGDQPNDSEVD
MEGDYDSETENSFIAQVDDLEQKYMQLPEHKKHATDFATENLCS
ESIKNKLSITTIGNLTELQTDKHTENQSGYEGVTIEPGADLLYDVP
SSQAIYFENLQNSSNDLGDHSMKERDWKSSSHNTVNEELPHNCI
EQPQQNDESSSKVRTSSDMNRRKSIKDHLKNAMTGNAKAQTPIF
SRSKQLKDTLLSEEINVAKKTIESSSNDLGPFYSLPSKVRDLYAQ
FKGIEKLYEWQHTCLTLNSVQERKNLIYSLPTSGGKTLVAEILML
QELLCCRKDVLMILPYVAIVQEKISGLSSFGIELGFFVEEYAGSKG
RFPPTKRREKKSLYIATIEKGHSLVNSLIETGRIDSLGLVVVDELH
MIGEGSRGATLEMTLAKILYTSKTTQIIGMSATLNNVEDLQKFLQA
EYYTSQFRPVELKEYLKINDTIYEVDSKAENGMTFSRLLNYKYSD
TLKKMDPDHLVALVTEVIPNYSCLVFCPSKKNCENVAEMICKFLS
KEYLKHKEKEKCEVIKNLKNIGNGNLCPVLKRTIPFGVAYHHSGL
TSDERKLLEEAYSTGVLCLFTCTSTLAAGVNLPARRVILRAPYVAK
EFLKRNQYKQMIGRAGRAGIDTIGESILILQEKDKQQVLELITKPL
ENCYSHLVQEFTKGIQTLFLSLIGLKIATNLDDIYHFMNGTFFGVQ
QKVLLKEKSLWEITVESLRYLTEKGLLQKDTIYKSEEEVQYNFHIT
KLGRASFKGTIDLAYCDILYRDLKKGLEGLVLESLLHLIYLTTPYDL
VSQCNPDWMIYFRQFSQLSPAEQNVAAILGVSESFIGKKASGQAI
GKKVDKNVVNRLYLSFVLYTLLKETNIWTVSEKFNMPRGYIQNLL
TGTASFSSCVLHFCEELEEFWVYRALLVELTKKLTYCVKAELIPLM
EVTGVLEGRAKQLYSAGYKSLMHLANANPEVLVRTIDHLSRRQAK
QIVSSAKMLLHEKAEALQEEVEELLRLPSDFPGAVASSTDKA

Supplementary Data 1. HelQ Sequence. Anotated HelQ sequence. Colour coding aligns with PyMOL predicted striuctral model in Figure 4.3.