Neuroscience Area – PhD course in

Functional and Structural Genomics

The role of the accessory N- and C-terminal domains in modulating the helicase activity of human RecQ4

Candidate: Advisor:

Ilenia Bagnano   Silvia Onesti, PhD

Academic Year 2018-19

The role of the accessory
N- and C-terminal domains in
modulating the helicase activity
of human RecQ4

Ilenia Bagnano

A thesis submitted for the degree of
Doctor of Philosophy
in Functional and Structural Genomics
October 2019

ABSTRACT

RecQ helicases are ubiquitous DNA unwinding enzymes, essential in the maintenance of genome stability by acting at the interface between DNA replication, recombination and repair. Humans have five different paralogues of RecQ helicases namely RecQ1, BLM, WRN, RecQ4 and RecQ5. Germ-line mutations in the *recq4* gene give rise to three distinct human genetic disorders (Rothmund-Thomson, RAPADILINO and Baller-Gerold syndromes), characterized by genetic instability, growth deficiency and predisposition to cancer. Moreover, RecQ4 sporadic mutations are implicated in many types of cancer, including osteosarcoma and lymphoma. At the same time, overexpression of RecQ4 in tumour development and progression has been observed in breast, cervical and prostate cancer.

In addition to the central helicase core, RecQ4 has a unique N-terminal domain, which is essential for viability and has homology to the yeast Sld2 replication initiation factor, followed by a Zn knuckle, and an uncharacterized C-terminal domain. The role of these accessory regions in the function of the RecQ4 helicase was assessed by a combination of biochemical and biophysical experiments.

Many RecQ helicases have a C-terminal domain which folds into a helical bundle known as the Helicase and RnaseD-like C-terminal domain (HRDC). Despite the lack of sequence homology, a detailed bioinformatics analysis led us to predict that the RecQ4 C-terminal domain may fold as a HRDC domain. We expressed and characterized the C-terminal domain of human RecQ4. CD spectra analysis of the recombinant fragment does suggest it folds into a helical bundle, consistent with our prediction. We show that the domain does not bind nucleic acid substrates. To further investigate its possible role in assisting the unwinding, we tested the ability of the helicase domain, alone and in trans with the C-terminal domain, to unwind a variety of DNA and RNA structures and found that the presence of the C-terminal domain enhances *in vitro* unwinding for a variety of substrates, especially for R-loops, the substrate most efficiently unwound by RecQ4, as our colleagues have previously shown (unpublished experiments). We have also produced a number of C-terminal site-directed mutants, either based on sequence conservation or present in Rothmund-Thomson patients, and tested the effect of the mutations on the biochemical activity. Our experiments indicate that the mutation of all selected residues (except Arg1162) seem to affect the R-loops resolving activity of the helicase core of RecQ4, and especially the RTS patient mutations, which cause a dramatic reduction in the activity of the protein.

We also investigated the effect of an N-terminal region encompassing the end of the Sld2 homologous region and the Zn knuckle. This region does bind all the substrates with a preference for R-loops, Holliday Junctions, D-loops and hybrid fork-RNA/DNA and its presence, in trans with the helicase core domain of the protein, strongly enhances RecQ4 binding to nucleic acids. Moreover, our experiments show that its presence significantly increases the unwinding activity of the helicase core towards all the substrates, especially for R-loops.

To enhance our understanding of the activity of the N-terminal domain towards R-loops unwinding and resolution, a number of mutants within the zinc knuckle, and predicted

and/or shown to be functionally important in nucleic acid interaction, were produced and used in R-loops unwinding assays. Two mutants (Asn406Cys and in particular Phe404Ala/Trp412Ala) showed a significant decrease in R-loops resolving activity of the protein, suggesting a role for these residues in substrate interaction.

This complex scenario would suggest a role of N-terminal domain as a major nucleic acid interacting region of the protein, increasing the unwinding affinity towards all the substrates, while the C-terminal domain is an essential player in imparting substrate specificity to the protein, promoting R-loops resolution.

These results enrich our knowledge about RecQ4 structure, function and suggests novel roles of RecQ4 in DNA replication and genome stability.

Moreover, since misregulation of R-loops might cause neurological diseases and cancer, a detailed biochemical and structural analysis of human RecQ4 and its role in R-loops metabolism may also shed light on these devastating diseases.

Table of contents

CHAPTER 1

INTRODUCTION

1.1. Nucleic acid helicases: general properties

DNA and RNA helicases are a large class of ubiquitous and evolutionary conserved enzymes, encoded by all forms of life (Pike et al., 2015). They act as motor proteins that move directionally along a nucleic acid phosphodiester backbone, separating two annealed nucleic acid strands (i.e., DNA, RNA, DNA:RNA or RNA:DNA hybrids, depending on the nature of strand oriented in 5'-3' direction and on the process in which the intermediates are formed), using energy derived from nucleoside triphosphate (NTP) hydrolysis (Singleton, 2007; Spies, 2014). Approximately 1% of eukaryotic genes code for helicases (Wu, 2012), which are considered very abundant proteins, and the human genome codes for around 95 non-redundant helicases: 64 RNA helicases and 31 DNA helicases (Umate et al., 2011).

They are essential enzymes involved in all aspects of nucleic acid metabolism including DNA replication, DNA repair and DNA recombination, as well as RNA transcription, maturation and translation, splicing, ribosome synthesis and nuclear export processes.

These enzymes are characterized by peculiar biochemical properties: **rate**, **directionality**, **processivity**, **step size** and **mechanism of unwinding**, which can be **active** or **passive**. **Translocation rate**, which is often determined via the ATPase rate, can vary from a few to several thousand base pairs per second. Some helicases are less active until they bind a protein cofactor that highly stimulates their activity. The **directionality** is defined with respect to the fact that DNA is a bipolar molecule with two strands running in opposite direction and, once loaded on the strand, helicases show a directional bias and translocate either 5'–3' or 3'–5'. These enzymes travel along just one of the two strands even if they bind a double stranded DNA. The **processivity** reflects the ability of these enzymes to catalyze multiple cycles of reaction before releasing the product, as opposed to a distributive enzyme. The **step size** is the number of base pairs unwound per molecule of ATP hydrolyzed. Ultimately, **active helicases** are able to destabilize the nucleic acids, participating in the active unwinding of the duplex ahead of the fork, while **passive helicases** wait for thermal fraying of the strands at the fork and then are able to trap the strands in the unwound state (Lohman and Bjornson, 1996; Singleton et al., 2007).

1.1.1. Classification of nucleic acids helicases

Sequence analysis performed by Gorbalenya and Koonin (1993) and more recent comparative structural and functional analysis carried out by Wigley and co-workers

(Singleton et al., 2007) have shown that helicases can be classified in distinct superfamilies (SFs).

The first classification (Gorbalenya and Koonin, 1993), which was done on the basis of primary structure analysis (in particular the existence of short conserved aminoacidic motifs identified within each group and shared by more than one group) divided the helicases into two major groups, SF1 and SF2, with three additional smaller families. The two largest superfamilies share similar patterns of seven conserved sequence motifs. Helicase motifs appear to be organized in a **core domain** which provides the catalytic function, whereas optional additional inserts and amino- and carboxy-terminal sequences may comprise distinct domains with different accessory roles.

An updated classification, comprising six superfamilies, was later proposed on the basis of the identification of novel motifs, characteristic of a single family (Singleton et al., 2007). The six superfamilies and their associate motifs are shown in Figure 1.1.
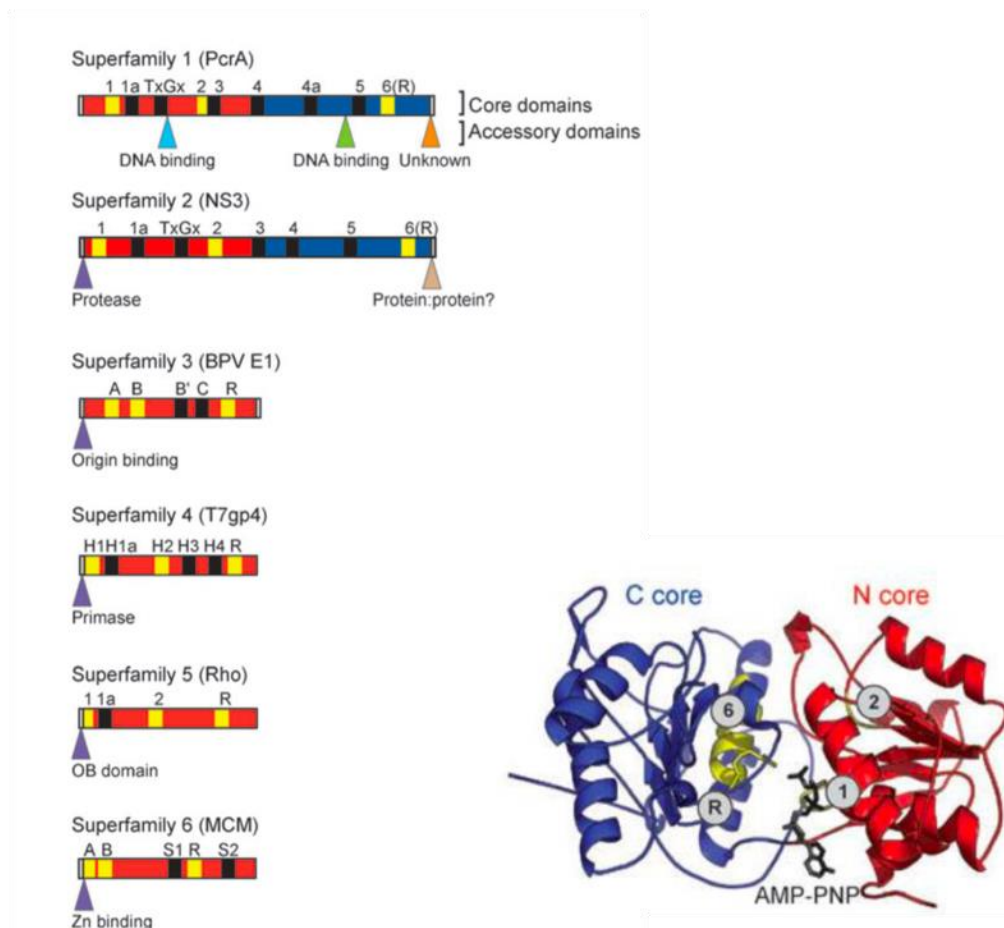
Figure 1.1. Helicase family classification. *Left*: Helicases family is composed of 6 superfamilies according to their primary sequence. Universal conserved motifs are colored in yellow. *Right*: The SF1 and SF2 core helicase is composed by two RecA-like domains, colored in blue and red. In a cleft between the two domains binds the ATP analog AMP-PNP (in black). Motif 1 and 2 are present in the N-core domain and are related to ATPases/ATP synthase Walker A and B motifs, while motif 6 in the C-core domain contains the arginine finger. The region involved in the conformational change in the

structure of the enzyme due to ATP binding and hydrolysis is highlighted in yellow (Pictures by Singleton et al., 2007).

The characteristic catalytic "**core domain**" is present in all the six superfamilies and folds either as a RecA-like or AAA+-like ATPase core with a nucleotide-binding site at the interface of the monomers. This structural unit is also a functional unit, able to convert chemical to mechanical energy by coupling NTP binding and hydrolysis to protein conformational changes. Common features of the core domains include conserved residues, involved in the NTP binding and hydrolysis, equivalent to the Walker A and Walker B boxes of many ATPases (Walker et al., 1982), and an "arginine finger" (R) that plays a key role in energy coupling (Scheffzek et al., 1997).

Walker A is a common motif in proteins that are associated with phosphate binding and it is present in many ATP or GTP utilizing proteins, as it is essential for ATP binding: the conserved lysine residue makes contacts with the β and γ phosphate of the ATP molecule, while serine/threonine residue plays a role in coordination of divalent metal ion Mg^{2+}, which is important for hydrolysis of ATP (Walker et al., 1982; Morozov et al., 1997).

Walker B is situated downstream of the A-motif and is involved in ATP hydrolysis. There is considerable variability in the sequence of this motif, with the only invariant features being a negatively charged residue following a stretch of bulky, hydrophobic amino acids (Pause & Sonenberg, 1992; Gorbalenya & Koonin, 1993; Koonin, 1993; Morozov et al., 1997; Bernstein & Keck, 2003; Pike et al., 2009).

The "arginine finger" (R) is a highly conserved and essential residue in many GTPase and AAA+ ATPase enzymes. It forms contacts with the γ-phosphate of the nucleotide, enabling structural organization of the active site for efficient catalysis via its nucleotide coordination (Nagy et al., 2016).

Helicases can be distinguished into two distinct types: those forming toroidal, predominantly hexameric structures, and those that do not. SF1 and SF2 enzymes contain two RecA-like domains and have mainly been considered to be monomeric or dimeric, with the catalytic site at the interface between the two RecA domains. They include the large majority of helicases. The toroidal helicases comprise SF3 to SF6 and comprise all of the cellular replicative helicases. They are generally hexameric (or double-hexameric) rings composed of 6 (or 12) individual RecA or AAA+ folds; the ATP (and thus the catalytic centre) lies at the interface between monomers.

Helicases and translocases can finally be classified on the basis of mechanistic differences. Enzymes with specificity for DNA, RNA, or DNA:RNA and RNA:DNA hybrids have all been identified and, according to their direction of translocation, they can be distinguished as type A helicases if they move on a substrate along the 3' to 5' direction or type B if they move along the 5' to 3' direction. Moreover, a further division is based on the type of substrate on which these enzymes translocate. They are named α or β helicases if the substrate is single or double stranded, respectively (Singleton et al., 2007).

1.1.2. The Superfamily 2 (SF2)

Superfamily 2 is the largest and most diversified of the helicase superfamilies. The SF2 helicases are implicated in all the aspects of RNA metabolism and many steps in DNA metabolism. It has been further divided, based on sequence homology, into distinct families including **RecQ-like**, **RecG-like**, **Rad3/XPD**, **Ski2-like**, **type I restriction enzyme**, **RIG-I-like**, **NS3/NPH-II**, **DEAH/RHA**, **DEAD-box** and **Swi/Snf** families. It also includes smaller groups, such as **type III restriction enzymes** and **Suv3** (Jankowsky et al., 2006; Jankowsky et al., 2007; Fairman-Williams et al., 2010).
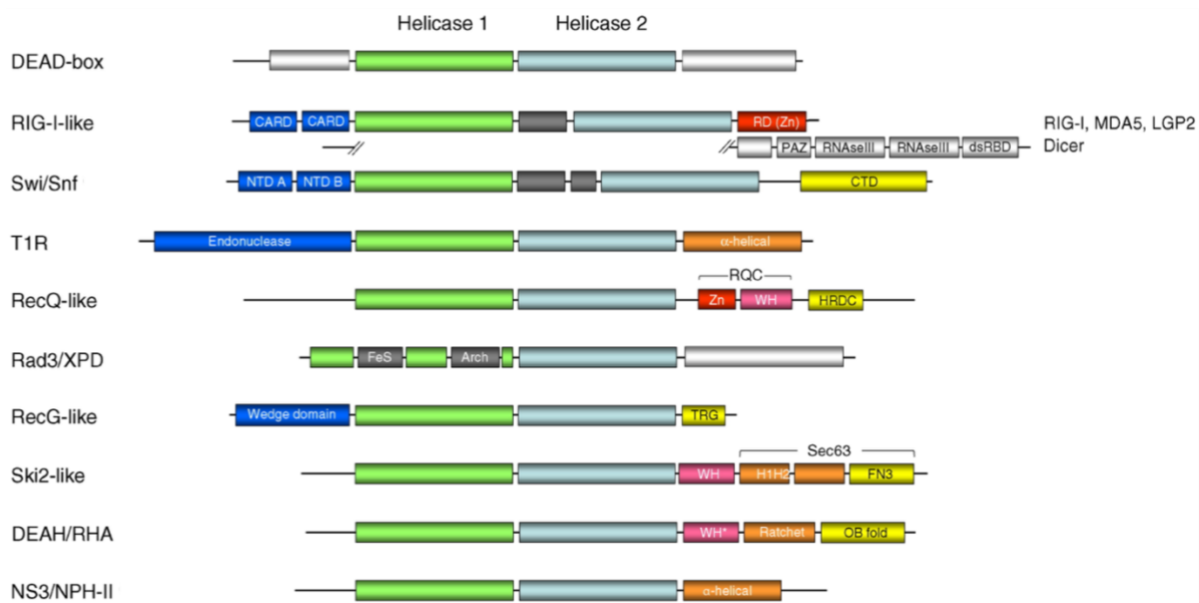
Figure 1.2. SF2 helicases family classification. Figure adapted from Fairman-Williams et al., 2010.

SF2 motor core contains 11 conserved motifs (Q, I, Ia, Ib, II, III, IV, IVa, V, Va, and VI). SF2 DNA translocases and processive helicases possess a **Q motif**, which coordinates the adenine base and is less conserved among those helicase families which do not show specificity for ATP (this motif is absent in the DEAH/RHA and viral DExH proteins, which are not specific for adenosine triphosphates) (Tanner et al., 2003). Helicase **motifs I and II**, containing the Walker A and B boxes, are among the most conserved motifs across the superfamily (Fairman-Williams et al., 2010). Located at the interface between the RecA-like folds, motifs I and II, along with motifs Q, III, Va, and VI, compose a pocket formed when the helicase domains are brought into close proximity upon ATP binding (Lohman & Bjornson, 1996). The subsequent hydrolysis of ATP to ADP and inorganic phosphate collapses this pocket and allows the separation of the helicase domains. **Motifs III and Va** are both involved in DNA-binding and NTP hydrolysis and are believed to play an essential role in transmitting the energy of ATP hydrolysis into motor function (Lohman & Bjornson, 1996; Zhang & Wigley, 2008). **Motif VI** is also involved in the coordination and hydrolysis of NTPs (Lohman & Bjornson, 1996; Zhang

& Wigley, 2008). **Motifs Ia, Ib, IV, IVa, and V** make extensive contacts with the phosphodiester backbone of the DNA (Bochkarev et al., 1999; Singleton et al., 2001).

The more studied subfamilies include the RecQ-like family, the DEAD-box RNA helicases and the Snf2-like chromatin remodeling enzymes and the best studied examples are the HEL308 (Buttner et al., 2007), flavivirus NS3 helicase (Luo et al., 2008) and the human RNA helicase DDX19 (both containing the DExH box) (Collins et al., 2009).

The crystal structure of SF2 HEL308 (Buttner et al., 2007), in complex with a dsDNA carrying a 3' single stranded tail, has revealed important features regarding the initiation of duplex unwinding reaction, such as the presence of a prominent hairpin loop involved in strand separation. Other structures of SF2 members are available: the flavivirus NS3 helicase (Luo et al., 2008) and the human RNA helicase DDX19 (both containing the DExH box) (Collins et al., 2009), the DEAD-box Vasa from Drosophila (Sengoku et al., 2006) and the DEAD-box ATPase eukaryotic initiation factor 4AIII (Andersen et al., 2006). These structures were crystallized either in their apo-forms or in complex with different nucleotides or nucleic substrates.

An example of complete structural cycle with many snapshots capturing a SF2 enzyme during ATP binding and hydrolysis coupled to translocation along nucleic acid is the NS3 helicase from Hepatitis C virus (HCV) (Gu and Rice, 2010). Three structures have been crystallized for NS3 HCV helicase: the binary complex enzyme-ssDNA, the tertiary complex enzyme-ssDNA-$ADP \cdot BeF_3$, mimicking the ATP bound state, and the tertiary complex enzyme-ssDNA- $ADP \cdot AlF_4^-$, mimicking the transition state of ATP hydrolysis.

The binding of the ATP analogues induces large conformational rearrangements in the enzyme and closes the cleft in the core domain, similarly to other ATPases composed of RecA-like domains. Motif II has a key role in changes upon nucleotide binding and hydrolysis. The nucleotide driven conformational changes are reflected also on the bound DNA, which bends and rotates altering the interactions between the helicase and its substrate. A cyclic loss and recovery of these interactions in the different nucleotide bound states, allows the enzyme to move along DNA with a 'ratchet' translocation mechanism. NS3 helicase is therefore able to maintain a unidirectional movement defining a step size of one base pair per ATP hydrolysis cycle.

1.2 The RecQ helicase family: structure and biochemistry

RecQ helicases belong to superfamily SF2 and are important enzymes in maintaining genome stability by acting at the interface between DNA replication, recombination and repair, with some also having a role in transcription. This is the reason why RecQ helicases are known as "Guardians of Genomic Integrity". The family name was chosen after the discoveries related to the RecQ gene of *E. coli*, which is the only family member present in this organism and was identified 20 years ago as a new mutation in a study focused on the isolation and characterization of a thymineless death-resistant mutant of *E. coli* (Nakayama et al., 1984).

They are conserved from bacteria to humans and several biochemical studies have demonstrated that they are able to unwind the DNA with a 3′ to 5′ direction in an ATP dependent manner (Larsen & Hickson, 2013). One of their main characteristics is that, beside the canonical forked duplexes, they can unwind a variety of DNA structures including displacement loops (D-loops; an intermediate in homologous recombination reactions), B-form DNA duplexes, DNA triple helices and G-quadruplexes. They can also promote annealing of complementary single-stranded DNA and branch migration of Holiday junctions (Bohr, 2008; Larsen & Hickson, 2013). Despite these findings and observations, their exact role in cellular processes are not fully elucidated.

Whereas most bacteria have only a RecQ homologue, eukaryotes tend to have multiple paralogues. A single RecQ homolog (named Sgs1) had been identified in yeasts, until a bioinformatics analysis suggested a putative ortholog in both plants and fungi, called Hrq1, which was suggested to be equivalent to RecQ4 (Gangloff et al., 1994; Stewart et al., 1997). Higher eukaryotes typically encode multiple RecQ orthologs, and five members have been identified in humans. The human RecQ helicases include: **RecQ1**, **BLM**, **WRN**, **RecQ4**, and **RecQ5** (Figure 1.3).

While RecQ1 and RecQ5 have not been linked to human diseases, defects in BLM and WRN are responsible for distinct genetic disorders: Bloom's syndrome (BS) and Werner's syndrome (WS) respectively (Ellis et al., 1995; Yu et al., 1996). RecQ4 mutations are linked to three autosomal recessive diseases: Rothmund-Thomson syndrome (RTS) type II, RAPADILINO syndrome and Baller-Gerold syndrome (BGS). The RecQ4-associated syndromes share common clinical features including skeletal abnormalities and growth retardation. In addition, RTS type II individuals are characterized by skin abnormalities, symptoms of premature aging and a high risk of developing osteosarcoma. The RAPADILINO phenotype typically lacks the unique RTS characteristics and these patients show an elevated risk for developing both osteosarcoma and lymphoma (Siitonen et al., 2009).

The fact that defects in human RecQ helicases causes similar but not-overlapping symptoms suggests that the 5 proteins have similar but not overlapping functions within the cells. However, the exact roles of each helicase, both in term of biochemistry and cellular role, are not yet fully understood.
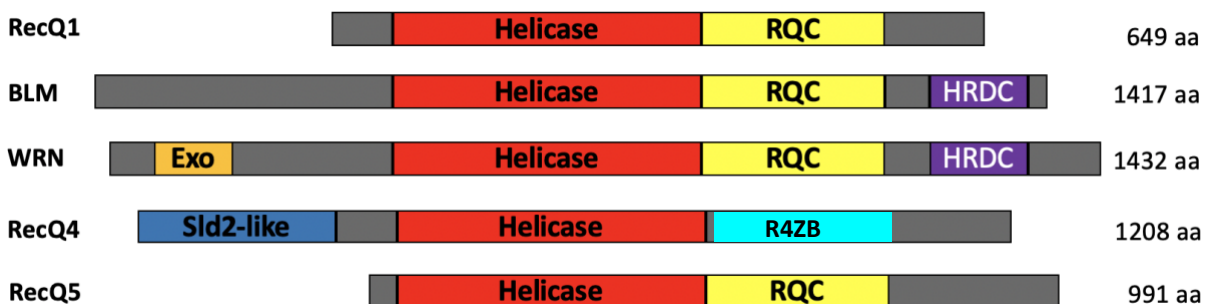
Figure 1.3. Schematic representation of five RecQ family members from humans: conserved and specific domains are indicated.

1.2.1. Structural features of RecQ helicases: an overview

In addition to a canonical SF2 domain, most RecQ helicases differ in the length of the N- and C-terminal domains that are flanking the catalytic core domain and feature some additional domains which assist helicase function, interactions with other proteins, interaction with a variety of DNA-metabolism intermediate structures and are also involved in regulation of protein subcellular localization, promotion of enzyme oligomerization and in providing additional enzymatic activities, such as the exonuclease activity in the N-terminal domain of WRN (Shen et al., 1998). The **RecQ carboxy-terminal (RQC) domain** is unique to RecQ helicases, while the **Helicase and RNase D C-terminal (HRDC) domain** is also found in other DNA binding proteins (Larsen & Hickson, 2013).

There is no obvious consensus on the stoichiometry of RecQ helicases, with studies proposing that the proteins act as monomers (Vindigni & Hickson 2009; Garcia et al. 2004), dimers (Vindigni & Hickson, 2009; Suzuki et al., 2009) and also trimers and hexamers (Karow et al., 1999; Perry et al., 2006). The crystal structure of RecQ1 with the DNA strongly suggests that at least RecQ1 (the most "basic" and less structurally complex member of the human RecQ family) is a functional dimer although is unusual among RecQ helicases in its quaternary structure, which can be dimeric or tetrameric. In fact, the RecQ1 tetramer seems to possess activities (Holliday Junctions branch migration and DNA strand annealing) that are not shared with the homodimer, which has only 3'–5' fork-unwinding activity (Pike et al., 2015).

However, the crystal structures of a number of RecQ helicases or RecQ fragments have been determined and, together with biochemical studies, have contributed to the understanding of the role of each single domain. Brief summaries of the structural information on each domain are reported below.

1.2.2. The core helicase domain

Crystallographic studies related to the core fragment of the RecQ helicase showed that the core helicase motif, like all the SF2 enzymes, is made up of the two RecA-like domains, named Helicase domain 1 and Helicase domain 2 (HD1 and HD2), with an ATPase catalytic site at the interface between the domains. This functional and structural unit was identified as the "molecular motor" of the helicase itself (Bernstein et al., 2003; Pike et al., 2009).
Motif I, which coincides with the **Walker A** motif (or P-loop, or phosphate-binding loop), is typically defined by a Gx4GK(S/T) consensus. It is a common motif in proteins that are associated with phosphate binding and it is present in many ATP or GTP utilizing proteins, as

it is essential for ATP binding: the conserved lysine residue makes contacts with the β and γ phosphate of the ATP molecule. The S/T residue plays a role in coordination of divalent metal ion Mg^{2+}, which is important for hydrolysis of ATP (Walker et al., 1982; Morozov et al., 1997). This motif differs slightly in RecQ helicases (TGxGKS is the RecQ helicases motif I) but retains the canonical function. The highly conserved aspartate residue in Motif II (or **Walker B**) with DExHC sequence coordinates the Mg^{2+} ion, while the glutamate acts as a catalytic base. The Walker A motif is thus involved in ATP binding, while the Walker B in ATP hydrolysis.

Besides the seven conserved motifs, the core helicase domain of RecQ helicases is characterized by an additional sequence element, termed 'motif 0', which is located N-terminally to motif I (Bernstein and Keck, 2003). This motif is well conserved in all RecQ enzymes from different organisms and is composed of four invariant and two conserved amino acids spaced by eight non conserved residues: 'LX3(F/Y/W)GX3F(R/K)X2Q' (Bernstein & Keck, 2003). This motif has been first identified in the *E. coli* RecQ helicase. The crystal structure of the nucleotide bound form of *E. coli* RecQ shows that the adenine moiety of ATPγS is located between the conserved aromatic residue (Tyr23) and Arg27 and is hydrogen-bonded to the conserved glutamine (Gln30). Mutation of a Gln residue to Arg residue in motif '0' of the human BLM gene is enough to cause Bloom's syndrome, suggesting motif '0' involvement in nucleotide binding and an important function in the core helicase domain (Ellis et al., 1995).

Helicase activity is crucial for RecQ helicase function *in vivo*, in fact point-mutations in this domain lead to a mutant phenotype in *S. cerevisiae* (Mullen et al., 2000), mice (Bahr et al., 1998) and humans (Rong et al., 2000).
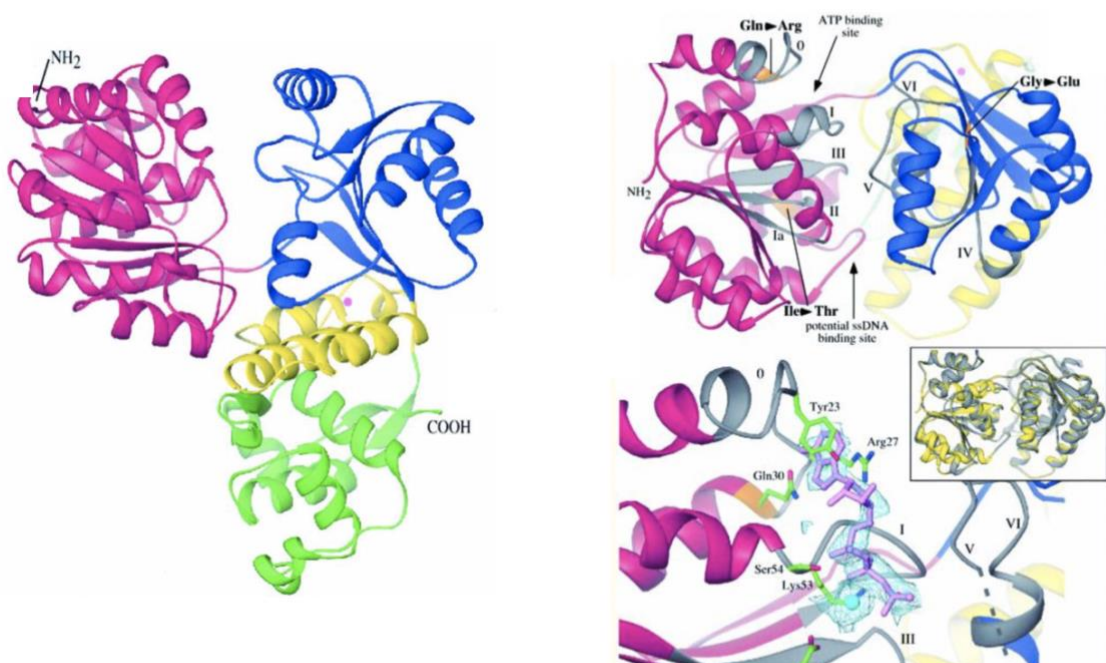
Figure 1.4. *E.coli* RecQ helicase. *Left:* Structure of the *E. coli* RecQ catalytic core [PDB code 1OYY]. The two distinct HD1 and HD2 are shown in magenta and blue, respectively. A bound Zn2+ ion is shown

as a magenta sphere. *Right (upper):* View into the cleft formed by the two helicase subdomains. Sites where nucleotide and ssDNA have been observed to bind in other helicase structures are indicated. *Right (lower):* Structure of ATPγS/Mn^{2+}-bound RecQ catalytic core. The ATPγS (adenosine 5'-O-(thiotriphosphate) adenine moiety is sandwiched between Tyr23 and Arg27, and hydrogen bonds are formed between the N6 and N7 atoms of the adenine and the side chain of Gln30. The triphosphate is bound by interactions with Lys53 and backbone amides from motif I. A Mn^{2+} ion (cyan) is bound by Ser54 from motif I and Asp146 from motif II. Picture adapted from Bernstein et al., 2003.

1.2.3. The RecQ C-terminal (RQC) domain

The RQC domain is unique to the RecQ helicases family. Less conserved than the core helicase domain, it is located downstream of the helicase domain and together they constitute the functional core unit for helicase activity. This region folds into a winged-helix motif, a subset of the helix-turn-helix superfamily (Hu et al., 2005; Kitano et al., 2010; Kim et al., 2013; Swan et al., 2014). Helix-turn-helix motifs, including the winged-helix, are known as major double-stranded (ds) DNA-binding domains and are found in many nuclear proteins (Gajiwala and Burley, 2000; Harami et al., 2013).

The RQC domain specifically consists of a Zn^{2+}-binding domain and Winged-Helix (WH) domain (Figure 1.5).

Berg and Shi suggested that Zn-binding domains in proteins can be crucial for several functions, such as critical in DNA binding and acting as surfaces that mediate protein-protein interactions (Berg & Shi, 1996). The structures of the Zn^{2+}-binding modules are highly similar between the bacterial and human enzymes: the RQC Zn-binding region binds Zn^{2+} using four conserved cysteines, located on two anti-parallel α-helices, which are brought together to form the metal binding site (Bernstein et al., 2003).

The Zn^{2+}-binding site is conserved among RecQ proteins and, consistently, several functional analyses have shown this region to be important for the enzyme function.

Indeed, two disease-causing BLM missense mutations mapping to highly conserved cysteine residues (Cys1036 and Cys1055) have been identified in the Zn-binding domain (Ellis et al., 1995). Additionally, when equivalent missense mutations found in BLM syndrome patients were introduced in the yeast BLM homologue, Sgs1, the mutants could not suppress DNA-damage sensitivity and showed increased frequency of sister chromatid exchange, believed to be caused by a high level of somatic inter-chromosomal recombination (Onoda et al., 2000).

The structures of RQC domains have been solved for *E. coli* RecQ, human RecQ1, WRN and BLM helicases (Bernstein et al., 2003; Pike et al., 2009; Kitano et al., 2010; Kim et al., 2013; Kitano, 2014).

The crystal structure of the truncated catalytic core of the human RecQ1$^{49–616}$ shows a β-hairpin, with an aromatic residue (Y564) at the tip, located in the C-terminal winged-helix domain, essential for DNA unwinding and Holliday junction (HJ) resolution activity of full-

length RecQ1, and required for dimer formation in RecQ1$^{49–616}$ and tetramer formation in full-length RecQ1 (Pike et al., 2009; Lucic et al., 2011). Mutagenesis studies on RecQ1 have shown that this prominent β-hairpin is essential for DNA unwinding and the substitution of the Tyr residue at the tip of the loop was sufficient to abolish the unwinding activity of the helicase (Pike et al., 2009; Lucic et al., 2011).
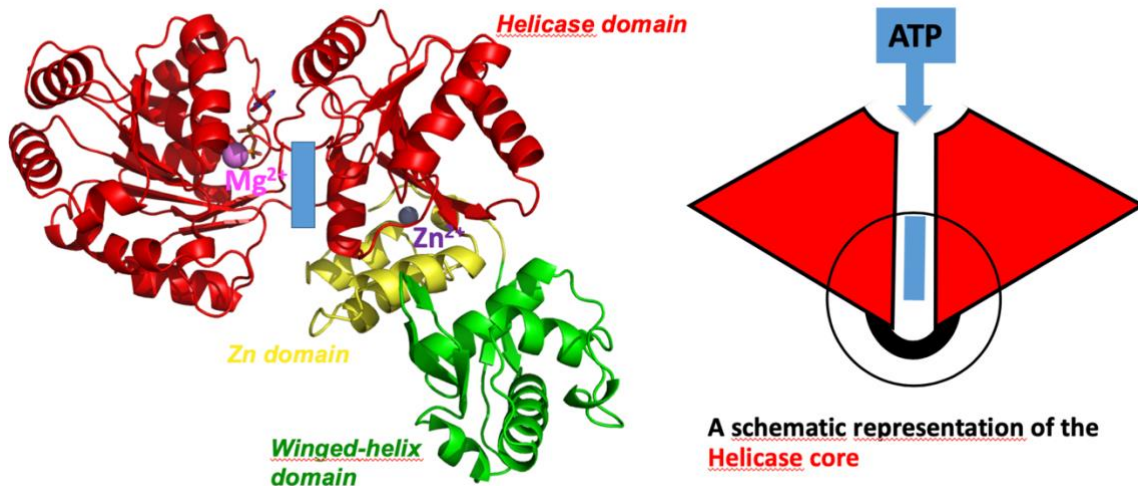
Figure 1.5. Crystal structure of human RecQ1 and a schematic representation of the catalytic core of RecQ helicases. *Left*: The crystal structure of the catalytic core of human RecQ1 [PDB code 2V1X], including the Helicase domain and the RQC domain. The domains are colored to reflect the different domains and subdomains; helicase RecA core is colored in red while, within the RQC, the Zn-binding subdomain is highlighted in yellow and the WH subdomain in colored in green. The bound Zn^{2+} ion is shown as a violet sphere. *Right*: A schematic representation of the helicase core of RecQ helicases, showing the two RecA-like domains in red and the ATP binding site between them in blue.

The corresponding β-hairpin in WRN is similarly capped by a phenylalanine at the same position, supporting a model in which the WH domain splits the DNA duplex using the β-wing as a wedge (Kitano et al., 2010).

The co-crystal structure of the WRN RQC domain bound to a DNA duplex has been deposited in Protein Data Bank (PDB) [PDB code 3AAF] and it represented the first example of a RecQ-DNA complex. Moreover, this structure successfully captured a DNA-unwinding event by the RQC domain. In particular, WRN RQC domain was found to bind duplex DNA in a novel DNA-interaction mode: the recognition helix, a principal component of helix-turn-helix motifs that are usually embedded within DNA grooves, was unprecedentedly excluded from the interaction. In this work, Kitano and colleagues showed that RQC domain specifically interacted with a blunt end of the DNA duplex and, in the absence of any other domain, unpaired a Watson-Crick base pair using the prominent hairpin structure β2–β3, which corresponds to the so-called β-wing of the winged-helix fold (Kitano et al., 2010).

Consistent with these results, mutagenesis of conserved Arg987 and Arg993 in the WH of WRN inactivates DNA binding and helicase activity (Kitano et al., 2010; Tadokoro et al., 2012). Other DNA helicases such as bacterial UvrD (Lee & Yang, 2006) and archaeal Hel308 (Buttner

et al., 2007) also possess a conserved β-hairpin to act as an unwinding element; however, these hairpins are located directly within the ATPase domains (i.e., helicase domains) and display various orientations toward DNA (Kitano et al., 2010).

The Helicase domain and the RQC domain together form the ~ 59 kDa functional core unit of *E. coli* RecQ and this fragment alone is sufficient to catalyze single stranded DNA (ssDNA) binding and 3' to 5' in vitro helicase activity with the same specific activity as the one of the full-length *E. coli* RecQ (Bernstein & Keck, 2003). In contrast with the human RecQ1 and WRN proteins, the *E. coli* RecQ β-hairpin is significantly shorter and lacks corresponding aromatic residues at the tip, indicating that the β-wing is not as important for DNA unwinding in the *E. coli* RecQ enzyme (Pike et al., 2009; Kitano et al., 2010; Lucic et al., 2010).

In 2013, the crystal structure of the BLM RQC domain bound to a phosphate ion was determined [PDB code 3WE2] (Kim et al., 2013). Subsequently, the co-crystal structure of a BLM large fragment (640 –1291 residues) in complex with a 3'-overhang DNA duplex [PDB code 4O3M] (Swan et al., 2014) was deposited.

This structure revealed unique structural features: the aromatic and non-polar residues at the tip of the β-wing, key elements that WRN uses for DNA strand separation, are each replaced by polar and acidic residues in BLM, and the presence of an asparagine instead of the aromatic residue at the tip of β-hairpin in RQC domain of BLM helicase was shown (Swan et al., 2014).

Moreover, an insertion between the N-terminal helices exhibiting a looping-out structure that extends at right angles to the β-hairpin was detected.

These BLM RQC features could be associated with the preferential activity of BLM toward HJs and, consistently with these results, deletion mutagenesis of this insertion interfered with binding to Holliday junction (Kim et al. 2013). The purified WH domain from BLM helicase shows strong G-quadruplex DNA binding activity (Huber et al., 2006).

Interestingly, the C-terminal residues of BLM RQC adopt a novel extended structure (referred to as the C-term extended loop) by being tightly packed against the domain core (Kim et al., 2013).
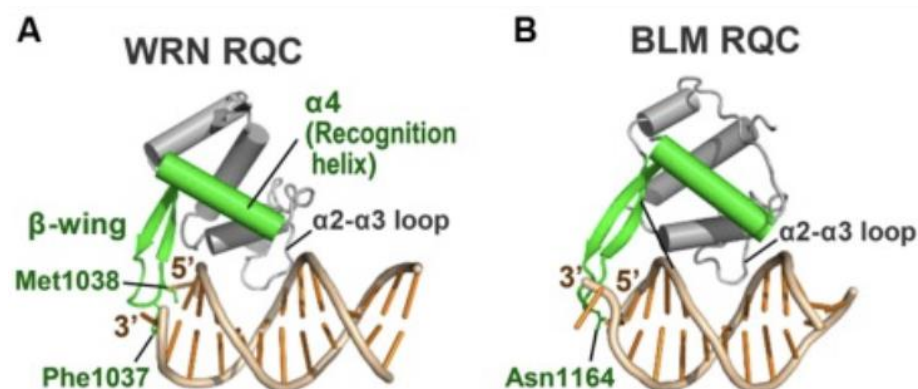
Figure 1.6. Crystal structure of human RQC winged-helix domains of WRN and BLM helicases. (A) WRN RQC domain [PDB code 3AAF; Kitano et al., 2010]. The RQC domain binds the DNA duplex terminus via the α2–α3 loop and the β-wing (in green), while the helix α4 (recognition helix, in green) is located apart from the DNA. (B) BLM RQC domain [PDB code 4O3M; Swan et al., 2014]. Picture adapted from Kitano et al., 2014.

This complex scenario shows that, whereas the relative orientation of the Zn-binding domain is stable, with respect to the helicase core, the WH domain can assume different orientations: these differences reflect distinct DNA recognition, translocation and unwinding dynamics as well as helicases substrates preferences, so their role in different genome maintenance pathways.
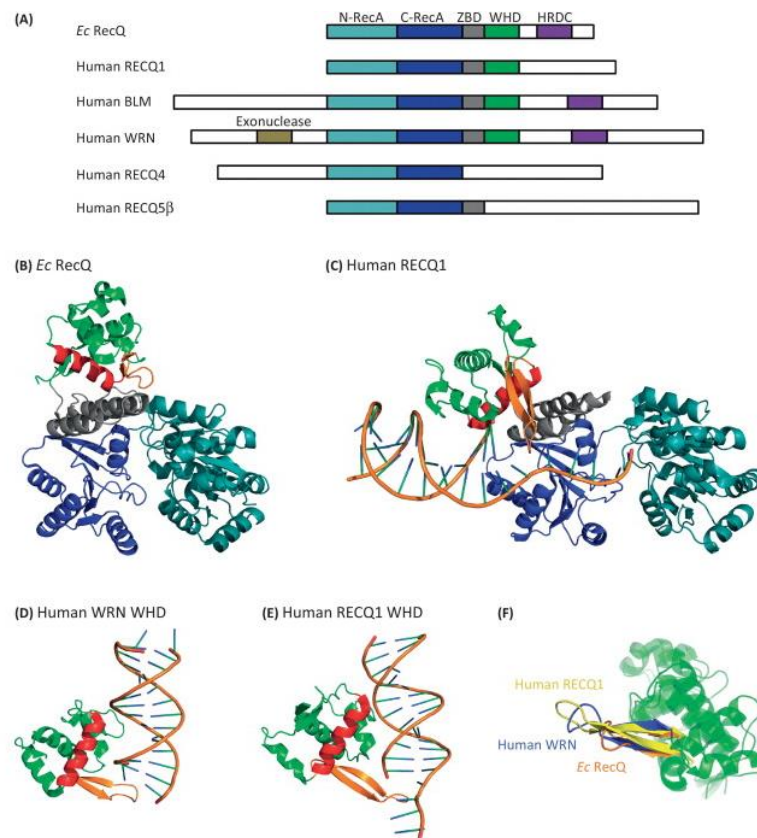
Figure 1.7. Winged helix domain (WHD) in RecQ helicases. (A) Schematic representation of the domain structure of *E. coli* and human RecQ helicases. (B) Crystal structures of truncated *E. coli* RecQ in the absence of DNA [PDB code 1OYW]. (C) Truncated human RecQ1 bound to 3'-tailed double-stranded dsDNA [PDB ID: 2WWY]. (D) Crystal structures of the WH domain of human WRN bound to blunt-ended dsDNA [PDB code 3AAF]. (E) Human RecQ1 bound to 3'-tailed dsDNA [PDB code 2WWY]. (F) Superimposition of WH domain structures of *E. coli* RecQ [PDB code 1OYW], with the loop W1 hairpin colored orange), human RecQ1 [PDB code 2WWY], W1 hairpin in yellow), and human WRN [PDB code 3AAF], W1 hairpin in blue). All the domains are colored as in (A). In B–E, the recognition helix is colored red, and W1 with adjacent β-strands is colored orange.

1.2.4. The Helicase-and-RNaseD-like-C-terminal (HRDC) domain

The third conserved region of RecQ helicases is the HRDC domain and it is found at the C-terminus of many RecQ helicases and RNAses. It exists in at least three genes linked to human diseases. These genes encode the RecQ helicases WRN and BLM, and the human PM-Scl autoantigen, an RNaseD homolog found in patients affected by polymyositis and scleroderma (Liu et al., 1999).

The HRDC region is the most variable of the conserved RecQ domains: for some eukaryotic RecQ proteins, such as RecQ1, RecQ4 and RecQ5, the presence of a HRDC domain is not obvious from the sequence, whereas several bacterial RecQ proteins have multiple HRDC domains. Although all HRDC domains have a similar helical bundle structure, they have different surface charge distributions and DNA-binding affinities (Vindigni et al., 2010).

Liu and colleagues determined the first three-dimensional structure of an HRDC domain from *S. cerevisiae* Sgs1 by NMR spectroscopy. This study revealed a positively charged region on the surface of the Sgs1p HRDC domain, that was proposed to interact with DNA, similar to the 1B module of SF1 helicases (Liu et al., 1999). They found that this basic patch on the surface of the Sgs1p HRDC domain can interact via electrostatic interactions with the phosphate backbone of the DNA. This basic patch is also present in the primary sequence of the HRDC module of *E. coli* RecQ, but not in the sequences of HRDC domain of WRN and BLM, suggesting that other interactions may be involved in the recognition of different substrates or protein cofactors (Vindigni et al. 2010). A structure of the HRDC domain from *E. coli* has a scaffold similar to HRDC from Sgs1. The isolated Sgs1 HRDC domain binds to ssDNA and 3'-overhanging duplex structures, whereas the *E. coli* RecQ HRDC binds only to ssDNA (Bernstein & Keck, 2005; Kim & Choi, 2010).

In BLM helicase, this domain is a crucial determinant for the efficient binding and unwinding of double Holliday Junctions (dHJ) (Tang et al., 2005). A study revealed unique features of BLM HRDC and, in particular, Sato and colleagues showed that BLM HRDC domain contains a patch of acidic residues which makes the domain surface extensively electronegative, suggesting an additional role of this motif apart from DNA binding (Sato et al., 2010). Kim and co-workers have proposed an electrostatic repulsion model in which upon BLM oligomerization, the HRDC domain from each BLM monomer uses electrical repulsion to separate the junction sites of dHJ DNA. In their model, the helicase core domain would then bind and unwind the double stranded region of the dHJ DNA because of its higher binding affinity for dsDNA. Subsequently, the dHJ DNA is divided into non-crossover products. The catalytic core and the HRDC domain showed then synergic cooperation (Kim & Choi, 2010).

A study by Newman and colleagues (2015) indicates a robust intramolecular association of the HRDC domain with the catalytic core, in a position to directly impact on the ATPase cycle. The distance of the HRDC domain from the DNA in their structure seems to suggest that a direct DNA binding role is unlikely, although it remains a possibility for DNA substrates with higher complexity. Further evidence for a role of the HRDC domain in the catalytic function of BLM comes from a structural and biochemical study in which an enzymatic

characterization of BLM mutants lacking the entire HRDC domain was performed. BLM mutants lacking the HRDC domain showed a decreased DNA unwinding activity and a generally less efficient coupling of ATP hydrolysis to DNA unwinding (Swan et al., 2014). This study also included a structure of BLM in complex with DNA that is very similar to the Newman and colleague's DNA complex, although a detailed comparison indicates that it may correspond to a slightly different state of the crystallized enzyme. They showed crystal structures of BLM-nanobody and BLM-DNA complexes which are virtually identical but the relative orientation of the WH domains is markedly different, with a rotation of 90° required to place the domains in equivalent positions. The WH domain in the DNA complex structure is in a similar position to that observed for human RecQ1 (Pike et al., 2009) and is poised to play a similar role in DNA strand separation. Newman and co-workers assume, therefore, that the position of the WH domain in the nanobody complex is different due to the absence of the DNA and, possibly, its stabilization by the nanobody (Newman et al., 2015).

All these findings, coupled with a detailed SAXS analysis, suggested a dynamic model for the BLM helicase mechanism and associated conformational changes that occur during the BLM catalytic cycle. In particular, the HRDC domain seemed to associate and dissociate from its binding cleft between the HD1 and HD2 domains together with nucleotide binding and release (Newman et al., 2015).

On the other hand, the HRDC domain of WRN helicase does not appear to interact with DNA *in vitro* (Liu et al., 1999; von Kobbe et al., 2003; Kitano et al., 2007). However, a WRN fragment containing the HRDC domain and additional residues at the C-terminus binds forked-duplex DNA and Holliday Junctions with high affinity (von Kobbe et al., 2003). The crystallized HRDC domain from WRN revealed that this protein possesses an additional N- and C-terminal extension to the standard helical bundle, which is missing in other RecQ helicases. The authors have suggested that WRN HRDC may be adapted to play a distinct function that involves protein-protein interactions, rather than protein-DNA interactions (Kitano et al., 2007). Consistent with these differences, the heterogeneity in the HRDC domains may correlate with functional differences between the various RecQ helicases.

1.2.5. Biochemical properties of RecQ helicases

The dynamic and diversified scenario related to RecQ helicase biochemical properties may provide a framework to better understand the physiological role of these helicases in processing different biological substrates: cycles of unwinding and reannealing may enable or antagonize specific DNA metabolic processes as well as provide chances to regulate RecQ helicases function and the DNA metabolic pathways in which they can act (Monnat & Sidorova, 2014).

RecQ helicases unwind DNA duplexes with moderate processivity by translocating along one of the two strands in the 3' to 5' direction while hydrolyzing ATP, and all exhibit DNA strand

annealing activity. The processivity of RecQ helicases in DNA unwinding is relatively low, in particular for WRN protein, which is unable to unwind duplex regions longer than 40 bp (base pairs). The presence of accessory factors, such as ssDNA binding proteins, can significantly enhance RecQ helicase processivity (Umezu & Nakayama, 1993).

Several studies indicate that many RecQ proteins exhibit preferential activity on wide range of substrates. Preferred substrates are branched DNA structures, including forked structures that mimic replication forks, and synthetic 4-way junctions that mimic Holliday Junctions (HJs). In fact, human BLM and WRN proteins promote branch migration of HJs. HJs resolution and dissolution are important in homologous recombination (HR) and DNA replication (Constantinou et al., 2000; Harmon & Kowalczykowski, 1998; Karow et al., 2000). RecQ helicases are also active in unwinding a number of atypical DNA structures, including gapped DNA, displacement loops (D-loops; an intermediate in homologous recombination reactions), DNA:RNA hybrids as well as DNA triplex (Bennett et al., 1998; van Brabant et al., 2000; Brosh et al., 2001; Mohaghegh et al., 2001 Machwe et al., 2002; Orren et al., 2002). The ability to process recombination intermediates formed during DNA replication is proposed to be a key function of the RecQ helicases (Hickson, 2003).

Sgs1, BLM, and WRN proteins are also efficient at unwinding G-quadruplex (G4) DNA, G-rich highly stable structures for which different cellular functions have been postulated (e.g. in telomere maintenance, DNA replication, transcription, and translation). The unwinding of G4 DNA by RecQ helicases requires a free 3' single-stranded tail (Sun et al., 1998; Mohaghegh et al., 2001; Huber et al. 2006, Sauer & Paeschke, 2017).

Recently, Stirling and his group have shown a role for Sgs1/BLM in R-loop (DNA:RNA hybrid and an associated non-template single-stranded DNA) suppression and in supporting DNA repair and replication fork stabilizing proteins as modulators of R-loop-mediated genome instability (Chang et al., 2017). Sgs1 loss increases R-loops and γ-H2A at sites of Sgs1 binding, replication pausing regions, and long genes, sensitizing cells to transcription-replication collisions. Moreover, analysis of BLM in Bloom's syndrome fibroblasts or by depletion of BLM from human cancer cells confirmed a role for Sgs1/BLM in suppressing R-loop-associated genome instability across species. In support of this potential direct effect, BLM is found physically proximal to DNA:RNA hybrids in human cells and can efficiently unwind R-loops *in vitro* (Chang et al., 2017).

RecQ enzymes are also able to promote the annealing of complementary single-stranded DNA molecules in ATP independent manner (Machwe et al., 2005; Sharma et al. 2005; Macris et al., 2006;). Khadka and collaborators (2016) compared double strand DNA annealing activity among the human RecQ proteins. Interestingly, RecQ1 and BLM had poor strand annealing in the presence of ATP, while WRN and RecQ4 showed intermediate incapacitation after addition of ATP. In contrast, RecQ5 exhibited strong annealing activity even in the presence of ATP. The initial rate of strand annealing activity of RecQ5 is also decreased in the presence of ATP as compared to the one in the absence of ATP. This shows that ATP might

have a general inhibitory effect on the strand annealing activity of RecQ proteins (Khadka et al., 2016).

In Figure 1.8 different RecQ substrates are shown.
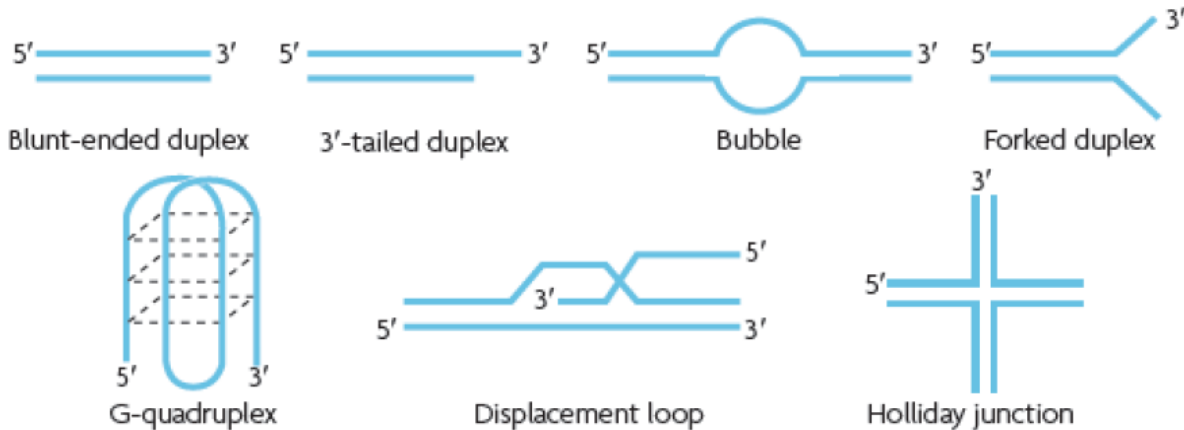
Figure 1.8. Some different DNA substrates for RecQ helicases. *In vitro*, RecQ can unwind a multitude of nucleic acid substrates, requiring neither an ssDNA tail nor a dsDNA end from which to initiate unwinding (Harmon & Kowalczykowski, 1998, 2000, 2001; Umezu et al., 1990). Translocation occurs along ssDNA and this translocation is tightly coupled to DNA unwinding (Manosas et al., 2010; Rad & Kowalczykowski, 2012a; Sarlós et al., 2012). Gap unwinding and several branched DNA substrates are preferred substrates for unwinding, including forked structures that mimic replication forks; D-loops and Holliday junctions, which resemble intermediates of the homologous recombination process, and G-quadruplex, formed at telomers and many other spots along the genome, are also common substrates for RecQ helicases (Bennet & Keck, 2004). Picture adapted from Chu & Hickson, 2009.

1.3. Cellular and physiological roles of RecQ helicases

1.3.1. RecQ helicases and DNA replication

DNA replication occurs in all living organisms and is the biological process of producing two identical replicas of DNA from one original DNA molecule. Since accurate duplication of the genome is critical to successful cell division, the process is carefully regulated and widely conserved. The DNA replication occurs in the S phase of the cell cycle and in eukaryotes it starts from many replication origins distributed along the genome, and therefore requires a complex regulation to ensure that each region of the DNA is replicated once and only once (Masai et al., 2010).

The regulation of DNA replication initiation in eukaryotes is the most complex and highly regulated. It starts from the assembly of a pre-replication complex (pre-RC), composed of six ORC proteins (ORC1-6), Cdc6, Cdt1, and a heterohexamer of the six MCM proteins (MCM2-7). The six subunit origin recognition complex (ORC, ORC1-6) associates with replication origins and the hexameric MCM2-7 complex, which is the catalytic core DNA helicase, and is

loaded onto the ORC-bound origins with the aid of Cdc6 and Cdt1 to form the pre-replicative complex (pre-RC, Figure 1.9). Two enzymes, known as the S-phase Cyclin-Dependent Kinase (S-CDK) and Dbf4-Dependent Kinase (DDK), play a major role in modulating this process and trigger the recruitment of Cdc45, MCM10, Sld2 and Sld3, Dpb11 and GINS, in yeast (Figure 1.9). The yeast replication factors Sld2 and Sld3 are the two essential targets of S-CDK and their phosphorylation in S phase leads them to interact with Dpb11, through interactions with its BRCT repeats (Tanaka et al., 2007b; Zegerman & Diffley, 2007). Whereas Dpb11 is clearly conserved in higher eukaryotes, no obvious Sld3 homologue has been identified with certainty in metazoan, whereas it has been suggested that the N-terminal region of the helicase RecQ4 may be the counterpart of Sld2 in eukaryotes (Sanghriti et al, Matsuno et al).

Then, in a series of steps which have not been elucidated in atomic details, the assembly of the replication fork occurs, involving origin melting and opening up of the DNA double helix, activation of the MCM2-7 helicase activity, recruitment of various factors involved in the tight control of the replication fork progression, including Cdc45 and GINS, which together with the MCM2-7 hexamer form the replicative helicase opening up the DNA double helix ahead of the fork. DNA replication can then start: Polα associates tightly with primase and extends a short DNA strand after the primer RNA synthesized by the primase. Subsequently, the five-subunit clamp loader, replication factor C (RFC), binds to the 3'-OH end of newly synthesized DNA and loads Polδ or Polε with a proliferating cellular nuclear antigen (PCNA, homotrimer sliding clamp that increases Polδ and Polε processivity). Polδ and Polε synthesize mainly the lagging and leading strands, respectively, at the replication forks (Kunkel and Burgers, 2008).

Several studies have demonstrated the involvement of human RecQ helicases in different aspects of DNA replication.
Evidences suggest that RecQ4 is among the essential DNA replication initiation factors. Studies performed with *Xenopus* egg extract system showed that the N-terminus of RecQ4 shares a weak but significant homology to the yeast Sld2 factor that assists in the recruitment of GINS to replication origins during S phase in a CDK activity dependent manner (Sangrithi et al., 2005; Matsuno et al., 2006; Thangavel et al., 2010). Indeed in *Xenopus* RecQ4 physically interacts with Cut5 (the *X. laevis* Dpb11 homologue). The N-terminal region of RecQ4 was shown to be essential for cell viability (Ichikawa et al., 2002; Matsuno et al., 2006; Abe et al., 2011) and it associates with several proteins involved in replication initiation like the MCM complex, MCM10, GINS, Cdc45. As a matter of fact, the absence of RecQ4 helicase was shown to significantly affect the formation of CMG complex (Xu et al., 2009; Im et al., 2009), However, its replication role is arguably preserved in RTS patients, since all the known RecQ4 mutants found in patients maintain the Sld2-like domain at the N-terminus of the protein.

Xu and colleagues demonstrated that RecQ4 is recruited to replication origins at late G1, after ORC and MCM complex assembly, while RecQ1 and additional RecQ4 are loaded at origins at the onset of S phase when licensed origins begin firing. Both proteins are lost from origins after DNA replication initiation, indicating either disassembly or tracking with the newly

formed replisome. Moreover, the group showed that RecQ4 is associated with the replicative MCM2-7 helicase and other factors of replication initiation complex through an interaction with MCM10, forming a complex with CTF4 in a process dependent on the CDK and DDK kinase activities (Xu et al., 2009; Im et al., 2009; Thangavel et al., 2010). This interaction is crucial for an efficient replication origin firing in human cells, although it does not seem required for cell viability. Xu and co-workers in fact reported that RecQ4 downregulation does not affect chromatin binding of MCM and Cdc6: these findings support the idea that RecQ4 is loaded on the origin downstream of these pre-RC complex factors but does affect chromatin recruitment of the GINS complex (Xu et al., 2009).

Moreover, other replication factors including RPA, PCNA and particularly DNA polymerase α display reduced binding to chromatin in the absence of RecQ4 (Sangrithi et al., 2005; Thangavel et al., 2010). RecQ4 appears then to function early, in replication initiation, when prereplication complex assembly takes place and active replisomes are assembled. RecQ1 helicase, which is also an integral component of the prereplication complex required for Kaposi's sarcoma-associated herpesvirus replication, is required for efficient replication initiation and may play an additional role during replication elongation (Wang et al., 2008).

Kanagaraj and co-workers suggested also a role for RecQ5β in DNA replication and fork repair by promoting the regression of stalled replication forks. The authors demonstrated that RecQ5β promotes strand exchange between arms of synthetic forked DNA structures, resembling a stalled replication fork in a reaction dependent on ATP hydrolysis. BLM and WRN can also promote strand exchange on these structures (Kanagaraj et al., 2006).

In particular, RecQ5β facilitates template-switching to catalyze the lagging strand unwinding and strand exchange on RPA-coated forked structures. Kanagaraj's group identified a short region located at the C-terminus of RecQ5β to be important for the processing of forked structures. They also found that RecQ5β interacts with the polymerase processivity factor, PCNA, as well as that this helicase localizes into DNA replication factories in S-phase nuclei (Kanagaraj et al., 2006). RecQ5 was also found to stabilize the replication fork allowing replication to overcome the effects of thymidine and complete the cell cycle (Blundred et al., 2010). Its overexpression releases cells from the cell cycle arrest upon thymidine treatment as these cells display fewer RPA foci and less γH2AX activation, compared to the cells with endogenous levels of RecQ5 (Blundred et al., 2010). RecQ5 appears also to be crucial in preventing genomic instability resulting from replication-transcription collisions, which is consistent with its role as a tumor suppressor (Hu et al., 2007).

In figure 1.9 a schematic representation of the assembly of DNA replication complexes is shown.
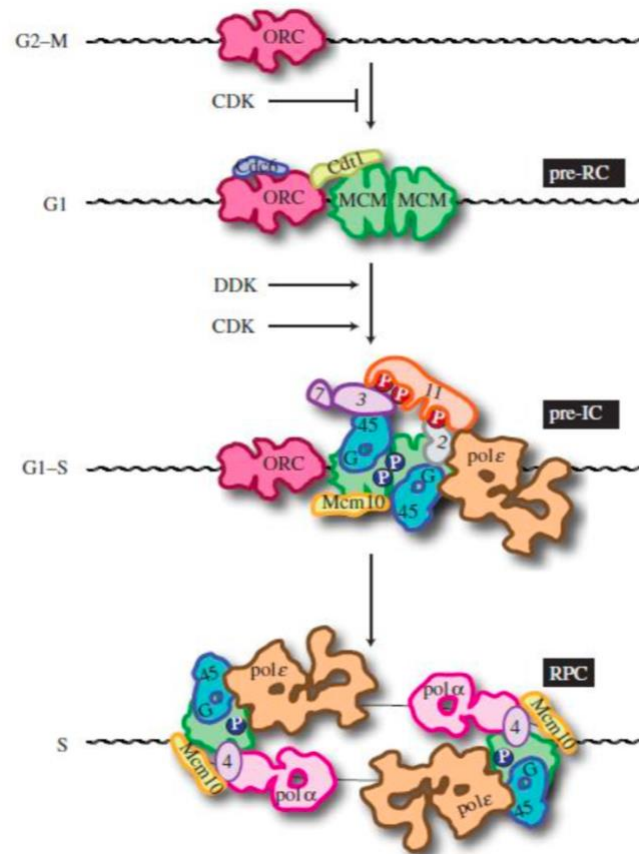
Figure 1.9. Assembly of DNA replication complexes in yeast. pre-RC: pre-replication complex; pre-IC: pre-initiation complex; RPC: replisome progression complex. Cell cycle phases permissive for the individual steps are shown on the left. Some of the protein names have been abbreviated: 11: Dpb11; 3: Sld3; 7: Sld7; 2: Sld2; G: GINS; 45: Cdc45; 4: Ctf4. See text for explanations (Diffley, 2011).

1.3.2. RecQ helicases and transcription

RecQ helicases play also a part in transcription. By several biochemical and cellular experiments, Balajee and co-workers indicated that WRN acts as a transcriptional activator modulating RNA polymerase II transcription, which could be explained considering the differences between gene-expression patterns in normal and Werner's syndrome lymphoblastoid cells (Balajee et al., 1999). Turaga and colleagues showed that WRN helicase plays also a part in regulating chromatin structure via interactions with chromatin accessibility factor 1. In fact, WRN-dependent changes in chromatin structure can reduce the incidence of DNA-strand breaks. WRN helicase would be then crucial in ensuring genome stability, by acting in concert with Topol to prevent DNA breaks, following alterations in chromatin topology (Turaga et al., 2007).

Among RecQ helicases, RecQ5 seems also to be able in promoting genome stabilization by inhibiting the transcription at the initiation to reduce transcription-associated replication impairment and recombination (Islam et al., 2010), although it can bind both the initiation

(Pol IIa) and elongation (Pol IIo) forms of the RNA polymerase II through two distinct and independent domains (Islam et al., 2010).

1.3.3. RecQ helicases and DNA repair

The complex DNA repair machinery is a collection of processes by which a cell identifies and corrects damage to the DNA molecules that encode its genome.

RecQ helicases take part in various DNA repair pathways and are able to interact with proteins involved in DNA repair. Consistently, they prefer to specifically bind DNA substrates that resemble DNA repair intermediates.

There are several types of damage to DNA due to endogenous cellular processes and exogenous agents, and the four main DNA repair pathways in the cells include:

- **Base Excision Repair** (**BER**), which repairs oxidative DNA base modifications such as 8- oxoguanine (8-oxoG), alkylation base damage and ssDNA breaks (SSBs);
- **Nucleotide Excision Repair** (**NER**), which repairs bulky helix-distorting DNA lesions and **mismatch repair** (**MMR**), which repairs single nucleotide mismatches and small insertion–deletion mispairs;
- **Double Strand Break Repair** (**DSBR**), which repairs double strand breaks (DSBs) either using the **Homologous Recombination** (**HR**) or the **Non-homologous End Joining** (**NHEJ**) pathway.

Base Excision Repair

BER removes oxidative, abasic and other monofunctional base modifications from DNA. These base lesions are recognized by a glycosylase that removes the damaged nitrogenous base while leaving the sugar–phosphate backbone intact, creating an apurinic/apyrimidinic (AP) site. Mammalian cells express multiple DNA glycosylases with overlapping DNA lesion specificity. The enzyme AP endonuclease 1 (APE1) cleaves AP sites, producing a single strand break intermediate, which can be utilized by DNA polymerase β (Pol β) for DNA repair synthesis, which is followed by ligation of nicked DNA by a DNA ligase (Wilson & Bohr, 2007).

WRN strongly stimulates NEIL1 glycosylase which incises formamidopyrimidines, 5-hydroxyuracil and, to a lesser extent, 8-oxoG. APE1 is inhibited by WRN (Ahn et al., 2004; Das et al., 2007), but stimulated by RecQ4 (Schurman et al., 2009) *in vitro*. Moreover, APE1 is upregulated in RTS cells and is overexpressed in some sarcomas, types of cancers that are common to RTS and WS (Wang et al., 2004; Schurman et al., 2009).

Shurman's group showed that primary RTS and RecQ4 siRNA knockdown human fibroblasts accumulate more H_2O_2-induced DNA strand breaks than control cells, suggesting that RecQ4

may promote repair of H_2O_2-induced DNA damage by activating BER, so it is capable to regulate both directly and indirectly base excision repair capacity. RTS primary fibroblasts display an upregulation of BER pathway genes and fail to respond like normal cells to oxidative stress accumulate. The cells accumulated more XRCC1 foci than control cells in response to endogenous or induced oxidative stress and have a high basal level of endogenous formamidopyrimidines, suggesting that RecQ4 deficiency is associated with defective repair of oxidative DNA damage, and may exhibit a hyper-oxidation phenotype. In cells treated with H_2O_2, RecQ4 co-localizes in the nucleus of cells after oxidative DNA damage with APE1 and FEN1, key participants in base excision repair, and biochemical experiments indicate that RecQ4 specifically stimulates the enzymatic functions of three BER proteins: the apurinic endonuclease activity of APE1, the incision of a 1- or 10-nucleotide flap DNA substrate by Flap Endonuclease I (FEN1) and also the DNA strand displacement activity of Pol β *in vitro* (Schurman et al., 2009).

Pol β is significantly stimulated by helicase-proficient, but not by helicase-defective, WRN (Harrigan et al., 2003). DNA Pol β lacks 3' proofreading exonuclease activity, so when DNA Pol β and WRN work in a coordinated manner, WRN exonuclease activity can play a role in correcting DNA synthesis errors, contribute to post-replication repair, and increase DNA synthesis fidelity during BER (Harrigan et al., 2006).

WRN, RecQ1 and RecQ4 interact with and are modulated by poly(ADP-ribose) polymerase 1 (PARP1), which plays essential roles in the cell, including DNA repair, translation, transcription, telomere maintenance and chromatin remodeling, adding moieties to chromatin-binding proteins, thereby modulating chromatin structure and function (Sousa et al., 2012; Thomas et al., 2013).

WRN-deficient cells fail to activate PARP1 in response to oxidative and alkylation DNA damage, indicating that the WRN–PARP1 interaction is biologically relevant and significant (von Kobbe et al., 2003). However, PARP1 is hyperactivated in RecQ1-depleted cells exposed to oxidative stress and in BLM-, WRN-, and RecQ5-depleted cells that are not exposed to exogenous stress (Gottipati et al., 2010; Sharma et al., 2012; Tadokoro et al., 2012).

Woo and colleagues showed, for the first time, that RecQ4 localizes to the nucleolus in response to oxidative stress (Woo et al., 2006), suggesting that RecQ4 helicase may have a role in one of several processes that are associated with the nucleolus, such as rRNA synthesis, protein folding, or cell cycle regulation through protein sequestration.

It is also reported that the C-terminal region of RecQ4, interestingly, is a PARP1 *in vitro* substrate, and PARP1 inhibitors were seen to alter the cellular localization of RecQ4 (Woo et al., 2006). Most mutations of the RECQL4 gene in patients have been shown to result in the truncation or loss of the entire helicase and C-terminal domain of RecQ4. Because the deletion mutants that lacked the C-terminal domain localized differently than the full- length protein, the authors speculated that these truncated protein products, in addition to potentially lacking critical enzymatic activity, may mislocalize in the cell and contribute to some of the abnormalities associated with Rothmund-Thomson syndrome (Woo et al., 2006).

Nucleotide Excision Repair

NER removes bulky lesions from DNA, such as UV-induced pyrimidine dimers and carcinogen adducts. It has been reported that WRN is stimulated by XPG (Xeroderma pigmentosum complementation group G) and RecQ4 interacts with XPA (Trego et al., 2011; Fan & Luo, 2008), two proteins involved in NER. However, RecQ helicase-deficient cells are not hypersensitive to UV irradiation and so there is no direct evidence that the RecQ helicases alter NER functionally *in vivo*.

DNA Double-Strand Break Repair (DNA DSBs)

DNA DSBs are common events in eukaryotic cells and can cause transient or permanent cell cycle arrest, mutagenesis, gross chromosomal rearrangements, cell death and tumorigenesis. The diverse causes of DSBs result in a diverse chemistry of DNA ends that must be repaired. There are three major pathways in human cells for repairing DSBs: **non-homologous end joining** (NHEJ), alternative **non-homologous end joining** (Alt-NHEJ) and **HR** (Ciccia et al., 2010; Boboila et al., 2012). HR-dependent DSBs repair is based on a sister chromatid template, is error free and occurs only during the late S and G2 phases of the cell cycle (Takata et al., 1998). NHEJ and Alt-NHEJ occur throughout the cell cycle and are error prone (Boboila et al., 2012). Alt-NHEJ is mainly active when NHEJ is impaired (Boboila et al., 2012). The five human RecQ helicases play putative roles in one or more subpathways of DSBs repair and interact with several key components of the DSBs repair machinery.

Non-homologous End Joining (NHEJ) and Alternative Non-homologous End Joining (Alt-NHEJ)

Non-homologous end joining (NHEJ) is a pathway that repairs double-strand breaks in DNA. Its name is due to the fact that the break ends are directly ligated without the need for a homologous template, in contrast to homology directed repair, which requires a homologous sequence to guide repair (Moore & Haber, 1996).

In nonproliferating human cells, DSBs are repaired by NHEJ and Alt-NHEJ. Several proteins are involved in NHEJ machinery: 53BP1, the Ku70/Ku80 heterodimer (Ku), DNA-PKcs and XLF/XRCC4/LIG4.

RecQ1 has been reported to interact with Ku and RecQ1-deficient cells show a reduced Ku-DNA binding activity, thus indicating a role of RecQ1 in NHEJ (Parvathaneni et al., 2013).

Shamanna and colleagues reported that RecQ4 interacts with Ku and DNA-PKcs via its N-terminal domain. RecQ4 also stimulates DNA binding of Ku to a blunt end DNA substrate, thus implicating its role in NHEJ (Shamanna et al., 2014).

It has been also found that WRN interacts with Ku and is a substrate of DNA- PKcs kinase. WRN exonuclease is stimulated by Ku and XRCC4/LIG4; in contrast DNA-PKcs stimulates WRN helicase but not WRN exonuclease (Rooney et al., 2004; Fattah et al., 2010; Bothmer et al., 2010; Bunting et al., 2010; Cooper et al., 2000; Li & Comai, 2000; Karmakar et al., 2002; Kusumoto et al., 2008; Kusumoto-Matsuo et al., 2010; Yannone et al., 2001).

The precise mechanism of Alt-NHEJ is poorly understood; however, there is often microhomology between the DSB break point and the DNA repair template (Boboila et al., 2012). During Alt-NHEJ, PARP1 may act as a DNA damage recognition protein, followed by end resection by MRE11, CtIP and EXO1 and ligation by LigIII/XRCC1 (Audebert et al., 2004; Wang et al., 2006; Dinkelmann et al., 2009; Xie et al., 2009; Wang et al., 2005). WRN, BLM and RecQ1 are shown to interact with EXO1 and MRE11/RAD50/NBS1 (MRN) and an interaction of WRN-LigIIIα has been also reported (Aggarwal et al., 2010; Gravel et al., 2008; Nimonkar et al., 2008; Doherty et al., 2005; Cheng et al., 2004; Cheng et al., 2005; Nimonkar et al., 2011; Sallmyr et al., 2008). These findings suggest a role of RecQ helicases in Alt-NHEJ.

Homologous recombination (HR)

HR is an important and conserved cellular pathway for DNA repair during mitosis and also for chromosomal paring and exchange during meiosis. It is activated during the late S and G2 phases of the cell cycle. It repairs DNA DSBs when a homologous template is available, for example at stalled replication forks during S phase, at DNA lesions during mitosis and at chromosomal pairing during meiosis (Alberts et al., 2002). HR can cause chromosome instability, therefore it is extremely important for the cells to strictly regulate this process. CtIP, MRN, RPA, DNA2, EXO1 and RAD51 are some of the proteins involved in this process (Ciccia and Elledge, 2010). CtIP and MRN bind to the DSB and consequently CtIP stimulates MRN to recruit two additional nucleases, EXO1 and DNA2. RPA plays a role in stimulating BLM DNA unwinding and enforcing DNA2 for 5' to 3' resection. Several studies report that MRN recruits and enhances the processivity of EXO1 and BLM increases the affinity of EXO1 for DNA ends. *In vitro* studies suggest that only BLM stimulates DNA end resection of DNA2, while RecQ1, BLM and WRN stimulate EXO1 (Aggarwal et al., 2010; Nimonkar et al., 2008; Doherty et al., 2005; Nimonkar et al., 2011). It has also been reported that BLM and RecQ5 show an anti-recombination activity by disrupting the RAD51 nucleoprotein filament, essential for homology search and D-loops formation. WRN, RecQ1 and RecQ4 also colocalize with RAD51; however, WRN and RecQ1 cannot disrupt RAD51 filaments (Hu et al., 2007; Wu et al., 2001; Bugreev et al., 2007; Schwendener et al., 2010). WRN has also been reported to interact with several other proteins involved in HR, including MRN, BRCA1, RAD52 and RAD54 (Cheng et al., 2004; Cheng et al., 2006; Baynton et al., 2003; Otterlei et al., 2006). Interestingly, RecQ4 is recruited to laser-induced DSBs, and RecQ4-deficient cells are sensitive to ionizing radiation. RecQ4 interacts with RAD51 in cells exposed to etoposide (Singh et al., 2010; Kohzaki et al., 2012; Petkovic et al., 2005). WRN and BLM helicases in vitro

preferentially bind to DNA substrates that resemble recombination intermediates such as Holliday junctions (HJs). Moreover, BLM in cooperation with topoisomerase IIIa and RMI1-RMI2 complex can process double HJs to generate non-crossover products (a process named dissolution of HJs), consistently with the fact that BLM defective cells are deficient in SCE (Wu and Hickson, 2003).

1.3.4. RecQ helicases and telomere maintenance

Telomeres are protein-DNA structures at chromosome ends, rich in repetitive nucleotide sequences (tandem-repeated short G-rich sequences - TTAGGG in humans), which protect the end of the chromosome from deterioration or from fusion with neighbouring chromosomes. They have a crucial role in preserving genome stability, survival and proliferation at the cellular level and in preventing degenerative diseases and cancer at the organism level. The Shelterin complex (also called telosome) is a protein complex known to protect telomeres in many eukaryotes from DNA repair mechanisms, as well as to regulate telomerase activity. This complex includes different subunits including TRF1 and TRF2 proteins, which are able to bind the double-stranded telomeric DNA, and POT1, a protein that binds to single-stranded telomeric DNA overhang. The absence of Shelterin causes telomere uncapping and thereby activates damage-signaling pathways that may lead to non-homologous end joining (NHEJ), homology directed repair (HDR), end-to-end fusions, genomic instability, senescence, or apoptosis (O'Sullivan and Karlseder, 2010). Much of the final double-stranded portion of the telomere forms a T-loop (Telomere-loop) that is invaded by the 3' (G-strand) overhang to form a small D-loop (Displacement-loop). The T-loop configuration represents a protective cap that protects the end of the chromosomes from the DNA damage response (DDR) machinery (Griffith et al., 1999; O'Sullivan & Karlseder, 2010; Calado and Young, 2012).

Several studies showed the interaction of WRN, BLM and RecQ4 with Shelterin proteins. WRN and BLM both interact with telomere proteins TRF1 and TRF2 (Opresko et al., 2002; Opresko et al., 2005) and WRN is enriched at the telomeres only during S-phase of the cell cycle, as revealed by live-cell imaging and direct chromatin immunoprecipitation of synchronized cells (Opresko et al., 2004). Moreover, *in vitro* both BLM and WRN unwind artificial D-loops substrate in coordination with TRF1 and TRF2, and tetrameric G-quadruplexes are an excellent substrate for WRN and BLM helicases (Huber at al., 2006). TRF1 and TRF2 bind with high affinity to telomere repeat sequences and one of their function is to recruit proteins that are essential for telomere maintenance and repair. Oxidative lesions in telomere sequences disrupt the binding of TRF1 and TRF2 to telomere repeat sequences *in vitro* (Opresko et al., 2005); these lesions will, therefore, need to be rapidly repaired to maintain genome stability. POT1 (protection of telomeres 1) is a ssDNA-binding protein that binds with high specificity to telomere repeats and strongly stimulates WRN and BLM helicases (Opresko et al., 2005). In this context, WRN can help to resolve aberrant DNA structures that tend to form as the replication fork progresses through telomeric repeats.

This model is consistent with the observation that the frequency of chromosome fusions is higher in WS fibroblasts than in normal fibroblasts and that overexpression of telomerase reduces the frequency of chromosomal aberrations in cells lacking WRN (Crabbe et al., 2007).

Similarly to WRN, BLM helicase localizes to a subset of telomeres in either late S or G2/M phase (Barefield and Karlseder, 2012). BLM facilitates telomere replication by resolving G4 structures formed during copying of the G-rich strand by leading strand synthesis. In addition, deficiency of BLM, or another G4-unwinding helicase, resulted in increased G4 structures in cells and, importantly, its deficiency led to greater increases in G4 DNA detected in the telomere, compared with G4 seen genome-wide (Drosopoulos et al., 2015). In a study of 2013, it was identified a role of RecQ4 in the repair of thymine glycol lesions to promote efficient telomeric maintenance (Ferrarelli et al., 2013). RecQ4- and BLM-deficient cells exhibit telomere fragility but, while WRN and BLM unwind G4s, preliminary data indicates that RecQ4 does not (Mohaghegh et al., 2001; Rossi et al., 2010).

1.3.5. Post translational modifications in RecQ helicases

Post-translational modifications (PTMs) refer to covalent enzymatic modification of proteins, following protein biosynthesis, to form a mature protein product. PTMs can occur on the amino acid side chains or at the protein's C- or N- terminus and are important components in cell signaling, regulating the catalytic activities, cellular localization and protein-protein interactions of RecQ helicases. WRN can be phosphorylated by ATM, ATR, DNA-PKcs and c-Abl tyrosine kinase (Ammazzalorso et al., 2010; Patro et al., 2011). Suppression of phosphorylation mediated by ATR alters the localization of WRN to nuclear foci and its colocalization with RPA *in vivo*, leading to breakage of stalled replication forks (Ammazzalorso et al., 2010). Serine/threonine phosphorylation of WRN by DNA-PKcs and tyrosine phosphorylation by c-Abl kinase inhibits exonuclease and helicase activity of WRN (Opresko et al., 2003). WRN is also acetylated, *in vivo* and *in vitro*, by the lysine acetyltransferase p300, which stimulates the helicase and exonuclease activity (Muftuoglu et al., 2008). WRN can also undergoes sumoylation but the functional consequence of this modification remains unclear (Kawabe et al., 2000). p300 acetylation of RecQ4 is required for nuclear localization and disrupts its nuclear import, leading to a cytoplasmic accumulation of RecQ4 (Dietschy et al., 2009). Phosphorylation of RecQ4 Sld2-like domain by CDK is crucial for DNA replication as it modulates its interaction with MCM10 (Xu et al., 2009). Phosphorylation of BLM helicase causes its dissociation from the nuclear matrix, facilitating BLM accumulation in nucleoplasmic fraction and, consequently, its arrival to the sites of DNA damage (Beamish et al., 2002; Dutertre et al., 2000). Moreover, BLM undergoes both sumoylation and ubiquitylation, essential in DSBs repair pathway (Eladad et al., 2005; Ouyang et al., 2009) and in nuclear partitioning (Tikoo et al., 2013), respectively.

1.3.6. Interactions between RecQ helicases

All the human RecQ helicases interact physically and functionally with other DNA metabolic proteins (Croteau et al., 2014), including other RecQ proteins. Several studies have documented a number of significant functional interactions between different RecQ helicases. BLM inhibits the WRN exonuclease, although WRN and BLM share similar substrate specificity and colocalize to nuclear foci (von Kobbe et al., 2002). BLM also interacts with the N-terminus of RecQ4, which in turn specifically stimulates BLM. BLM also promotes retention of RecQ4 at DSBs *in vivo* and cells defective in BLM and RecQ4 proliferate slowly with elevated SCEs (Singh et al., 2012). Cells that are defective in WRN and RecQ5 demonstrate synthetic lethality. *In vivo* RecQL5 dissociates slowly from DSBs in WS and BS fibroblasts. However, it specifically stimulates WRN but not BLM, and RecQ5 and WRN cooperate during repair and restart of synthetic stalled replication fork-like structures (Popuri et al., 2013). These results suggest that RecQ5 and WRN play cooperative and complementary roles.

1.4. RecQ helicases and diseases

1.4.1. Werner syndrome (WS)

Werner syndrome (WS) is a rare autosomal recessive disorder, caused by biallelic mutations of WRN, which encodes the multifunctional nuclear protein with helicase and exonuclease activities. It was reported for the first time in 1094 by the German scientist Otto Werner. WS patients typically exhibit premature aging phenotype, including short stature, early graying and loss of hair, bilateral cataracts and scleroderma-like skin changes (Epstein et al., 1966; Goto, 1997; Werner, 1985; Yu et al., 1996). The most common initial symptom, which is often recognized retrospectively, is the lack of a growth spurt during one's teens. Their symptoms include the greying and loss of hair, cataracts, skin atrophy, diabetes mellitus, atherosclerosis and malignancies (Oshima et al., 2017). Patient are also cancer-prone: in particular, they display an elevated incidence of sarcomas (Futami et al., 2008a). Median age of diagnosis is around 37 years of age (Oshima et al., 2017). Mutations in WRN gene lead to loss of the WRN protein and its catalytic activity (Moser et al., 2000).

Werner helicase may modulate gene expression in human cells by unwinding G-quadruplexes at transcription start sites (Tang et al., 2016) and is also involved in various pathways including double strand break (DSB) repair, replication, base excision repair, transcription and telomere maintenance. Consequently, in absence of WRN helicase, the pathways of DNA repair and telomere maintenance fail to suppress cancer and the aging symptoms seen in patients with WS. In addition to rapid telomere shortening and dysfunction, over-expression of oncogenes and oxidation induce a low response to overall cellular stress in Werner syndrome cells. High stress causes a synergistic effect, where WS cells become even more sensitive to agents that increase cell stress and agents that damage DNA. As a result, WS cells enter into a stage of aging prematurely (Multani & Chang, 2006). Interesting, emerging

evidences show that WRN plays a role in mitochondrial metabolism. Nuclear DNA damage leads to mitochondrial dysfunction in progeroid syndromes, particularly in people with neurodegeneration. WRN may be involved in discarding the damaged mitochondria through the mitophagy process (Shamanna et al., 2017).

1.4.2. Bloom syndrome (BS)

Bloom syndrome (BS) is a rare autosomal recessive disorder, characterized by pre- and post-natal growth deficiency, skin photosensitivity and a high predisposition to cancer development and is caused by defects in the BLM gene (Ellis et al., 1995; German et al., 2007). BLM deficient cells are characterized by increased sister chromatid exchange (SCE) and has been shown to function in replication fork stabilization, fork rescue and HR repair. Previous studies showed that ATM is required for BLM recruitment at the DSB sites, while MRE11 is required for both recruitment and retention (Tripathi et al., 2018). BLM protein is activated through the sumoylation by a SUMO E3 ligase, NSMCE2. While BLM deficient cells expressing SUMO-mutant BLM protein increased focal RPA and decreased focal RAD51, NSMCE2 deficient cells show an increased focal RAD51 and exhibit a defect in the generation of DSBs following the hydroxyurea-induced replication fork collapse. This raises the possibility that an important source of SCEs in BLM-deficient cells may be derived from the repair of DSBs, during collapsed-fork rescue (Ouyang et al., 2013). BLM helicase localizes to the nucleolus and PML bodies, moving in and out of the nucleolus in response to DNA damage and other stresses. Recently, BLM protein was reported to interact with topoisomerase IIα to localize to PML bodies and also interact with topoisomerase I to localize to nucleoli. A mouse model with the DD mutant Blm protein excluded from the nucleolus was generated. The DD mutation is located in the interaction domain in BLM with topoisomerase I (Behnfeldt et al., 2018). This mouse model displayed signs of accelerated aging and alterations of rDNA copy number. These data suggest the significance of ribosomal genome instability in the pathogenesis of BS (Tangeman et al., 2016).

1.4.3. Rothmund-Thomson syndrome (RTS), RAPADILINO syndrome and Baller-Gerold syndrome (BGS)

Rothmund-Thomson syndrome (RTS) is a rare autosomal recessive disorder linked to a mutation in the RecQ4 gene, as are the allelic disorders RAPADILINO syndrome and Baller-Gerold syndrome (BGS). RTS was described for the first time by Rothmund as an unusual skin change together with bilateral juvenile cataracts (Rothmund, 1868); subsequent cases were reported by Thomson in 1936 (Thomson, 1936). RTS is characterized by poikiloderma, small stature, and skeletal abnormalities (radial ray defects, ulnar defects, absent or hypoplastic patella, and osteopenia), juvenile cataracts and an increased risk of developing cancer, especially osteosarcoma and skin cancer. RAPADILINO syndrome was described as occurring

in the Finnish population and is characterized by irregular pigmentation with macules (but no poikiloderma), small stature, skeletal abnormalities and gastrointestinal abnormalities. Osteosarcomas and lymphomas are also reported. BGS is characterized by craniosynostosis, radial ray defects, skeletal dysplasia, short stature and poikiloderma. A midline NK/T cell lymphoma has been also reported in BGS. Approximately two third of RTS patients carry at least one truncation mutation of RecQ4, associated with the increased risk of osteosarcoma (Wang et al., 2001), while most common RAPADILINO mutation is the 44 amino acid in-frame deletion which abolishes helicase and ATPase activities of RecQ4 (Croteau et al., 2012).

As RecQ4 helicase plays an important role in DNA replication and repair, the cellular pathways in which is involved are dramatically altered. Mutations in RecQ4 result in increased genomic instability caused by DNA replication stress and/or defective DNA damage repair, and telomere attrition, which could affect mitochondrial function, leading to aging phenotypes (Lu et al., 2014).

RTS cells display a high frequency of chromosomal abnormalities such as translocations and trisomes (Vennos et al., 1992) and genomic instability may induce premature cellular senescence cell senescence (Lu et al., 2014). RecQ4 has a crucial role in DNA end resection, the initial step required for homologous recombination (HR)-dependent double-strand break repair (Lu et al., 2016). When RecQ4 is depleted, HR-mediated repair and 5' end resection are severely reduced *in vivo* and the regulation of pathway choice between homologous recombination (HR) and non-homologous end joining (NHEJ) during DSB repair is impaired (Lu et al., 2016; Lu et al., 2017).

1.4.4. RecQ helicases and cancer

RecQ helicases do not seem to directly regulate tumorigenesis. They function to prevent genomic instability and their loss or inactivation result in the accumulation of structural changes in oncogenes or tumor-suppressor genes, leading to cancer. RecQ helicases are able to suppress neoplastic transformation through control of chromosomal stability and many other similar caretakers are functionally linked to the RecQ helicases, indicating a common molecular basis for tumorigenesis in several apparently distinct cancer predisposition disorders (Levitt & Hickson, 2002; Hickson, 2003).

Since human RecQ helicases are involved in key steps of DNA replication and repair, all the human RecQ pathologies and loss or impairment in RecQ helicases structure and function are associated with a high risk of developing different types of cancer, with distinct characteristic phenotypic profiles.

Clinical features of Bloom syndrome may include non-Hodgkin's lymphoma, leukaemias and carcinomas of the breast, gut and skin, while Werner syndrome patients often suffer from tumors of mesenchymal origin, soft-tissue and osteogenic sarcomas and, less frequently, melanoma and thyroid cancers. Interestingly, MYC oncoprotein, a transcription factor that coordinates cell growth and division, directly promotes transcription of the human Werner syndrome gene. MYC overexpression causes genomic instability and sensitizes cells to

apoptosis and its overexpression in WRN syndrome fibroblasts or after WRN depletion from control fibroblasts led to rapid cellular senescence (Grandori et al., 2003).

Among the RecQ4 associated tumors, RTS patients are instead primarily suffering from non-melanoma skin cancers and osteogenic sarcoma, while patients with RAPADILINO syndrome are mainly at risk for lymphomas as well as osteogenic sarcoma (Wang et al., 2001; Hickson, 2003; Siitonen et al., 2009).

All five human RecQ helicases are also upregulated in various cancers, suggesting their requirement in rapidly dividing cells to repair replicative lesions or elicit an appropriate response in cell cycle checkpoint or gene expression. The functional importance of key structural elements within the helicase core as well as additional domains inspired the design of small molecules which can target distinct and specific domains (Brosh, 2013).

Some small molecule helicase inhibitors were identified from *in vitro* helicase assays using purified recombinant RecQ helicase proteins, oligonucleotide-based DNA substrates and small molecule libraries. The targeting of tumors with RecQ helicase inhibitors can provide a tumor-specific therapeutic advantage if combined with conventional chemotherapy. As RecQ helicases may be upregulated to maintain genomic stability in cells that are transformed or actively proliferating, innovative approaches to selectively introducing DNA damaging chemotherapy drugs or radiation to tumor tissues without adversely affecting normal cells is an active area of investigation (Brosh, 2013). Among those tested in cell-based assays, small molecule inhibitors of DNA unwinding catalyzed by WRN (Aggarwal et al., 2011, 2013b) and BLM (Nguyen et al., 2013) were found all negatively affect proliferation of cancer cells and induce DNA damage and chromosomal instability.

Small molecule helicase inhibitors against WRN operate in a manner that is dependent on the genetic background of the tumor. WRN, like other human RecQ helicases, is typically upregulated in its expression in various cancer cell lines; moreover, their downregulation by RNA interference has been shown to cause decreased proliferation and compromise the dividing activity of cancer cells (Brosh, 2013).

Recent structural and biochemical studies showed that the C-terminus of RecQ4 contains a unique Zn^{2+} binding domain (R4ZBD) and a region sharing homology to two winged helices, distinct from the RQC domain seen in other RecQs (Kaiser et al., 2017). The unique identity of RecQ4 C-terminal domain, which was found to be crucial for DNA unwinding (Kaiser et al., 2017), could be a potential site for molecular docking of small molecules in the upper or lower half of the R4ZBD-WH to modulate the helicase catalytic function.

1.5. RecQ4 helicase

1.5.1. Structural features

The human *recq4* gene consists of 21 exons, lies on chromosome 8q24.3 and encodes 1208 amino acid (aa) polypeptide, resulting in a 133 kDa protein with a highly conserved ATPase

domain located in its center including the two helicase subdomains HD1 and HD2, which adopt the typical RecA-like fold (Kaiser et al., 2017).

The N-terminal domain includes a region of weak sequence homology to *Saccharomyces cerevisiae* Sld2 (1–400 residues) (Sangrithi et al., 2005; Marino et al., 2013), a yeast essential initiation factor of DNA replication (Kammimura et al., 1998). The presence of a Sld2-like domain is a unique feature within the RecQs family as RecQ4 has an important role in the initiation of eukaryotic DNA replication, playing a role in assembling the replicative complex (Sangrithi et al., 2005; Thangavel et al., 2010). Experimental evidences showed that, in *Xenopus* oocyte extract, the N-terminus of RecQ4, including the Sld2 homology region, is essential for replication initiation and chromatin binding by the DNA polymerase α (Matsuno et al., 2006) and the deletion of RecQ4 N-terminal region is lethal in mice (Ichikawa et al., 2002).

An *in silico* analysis previously done in our laboratory (Marino et al., 2013) foresaw the presence of new structural and functional features. A second region of homology with Sld2, in addition to the N-terminal 150 amino-acid residues (Sangrithi et al., 2005), was identified. This region is highly conserved among species and characterized by numerous positively charged and aromatic residues, possibly involved in DNA binding. A cysteine-rich region, which constitutes the Zn knuckle, was also identified. It is located between the Sld2 homology region and the helicase core domain and, in both human and frog proteins, the presence of this region strongly enhances binding to nucleic acids. It was also hypothesized the presence of a putative RQC domain at the C-terminus of the helicase domain (Marino et al., 2013). Later, both human and frog RecQ4 Zn knuckles have been structurally and biochemically characterized by using CD and NMR spectroscopy and EMSA techniques respectively (Marino et al., 2016).

NMR studies of the first 54 aa revealed a homeodomain-like fold, that can function as a DNA-interaction motif (Ohlenschlager et al., 2012). A short Zn^{2+}-binding motif upstream of the ATPase domain, the so-called "Zn-knuckle" (aa 610–634) was then characterized by NMR in *Xenopus laevis* RecQ4 (Marino et al., 2016). Marino and collaborators showed that the *Xenopus* fragment assumes the canonical Zn knuckle fold, whereas the human sequence remains unstructured, consistently with a mutation of one of the Zn ligands.

Zn knuckles are short cysteine-rich sequences (with a typical CCHC pattern), wrapping around a Zn^{2+} ion, often present in multiple copies in nucleocapsid proteins of RNA retroviruses and in eukaryotic gene regulators. In fact, this module has also been found in transcriptional activators, binding to promoter DNA (Fields et al., 2008) and cellular nucleic-acid binding proteins involved in ssDNA binding (Armas et al., 2008). Within the RecQ4 paralogues, the Zn knuckle motif is well conserved, although could present some variants of the canonical zinc site: for example, the human sequence has the second cysteine substituted by an asparagine (CNHC), while the *Xenopus laevis* region contains the canonical Zn^{2+} ligands: CCHC. Asparagine and glutamine are occasionally found as Zn^{2+} ligands (Cleasby et al., 1996; Koutmos et al., 2008). This domain is located in a region that appears essential for cell viability and is likely to share, together with the Sld2-homology domain, the essential function of RecQ4 in DNA replication. Recently, Kaiser and collaborators published the crystal structure

of human RecQ4, containing the conserved helicase core followed by a novel domain that takes the place of the RQC domain. The new domain, named RecQ4-Zn^{2+}-binding domain (R4ZBD), features a zinc-binding site and two distinct types of winged-helix domains. Mutational and functional analysis suggest that the zinc-binding domain in RecQ4 is not directly involved in dsDNA separation but could play a role in specific DNA interaction. A possible function for the R4ZBD-winged-helix domains in RecQ4 could be to serve as a platform to interact with other proteins (Kaiser et al., 2017).
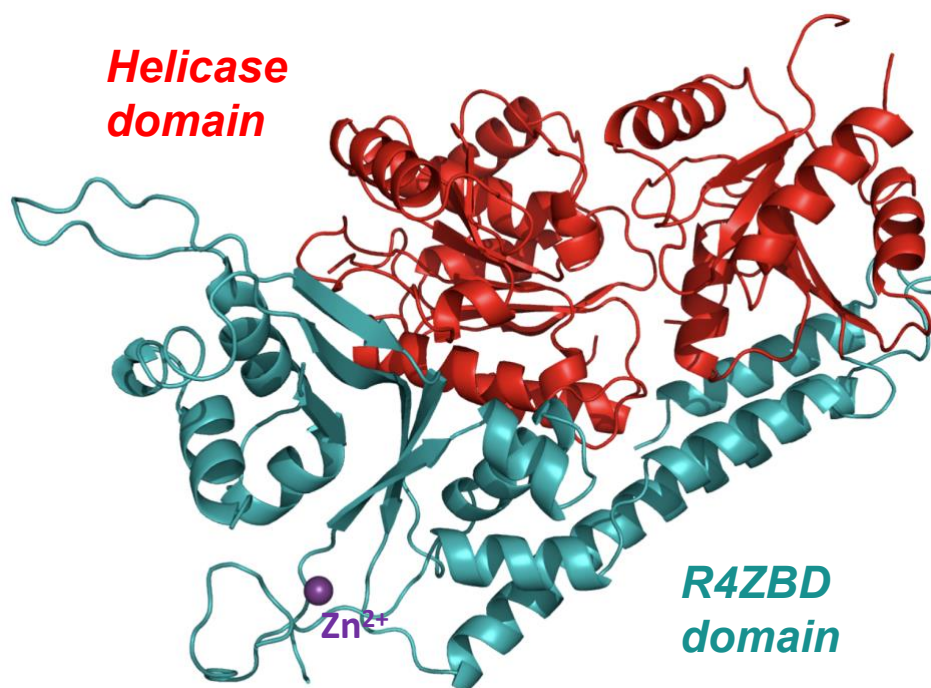
Figure 1.10. Structure of human RecQ4 helicase (aa 449–1111). The ATPase helicase domain, comprising HD1 and HD2, highly conserved among all RecQ proteins, is shown in red. In place of a RQC domain, RecQ4 features a structurally unique domain, termed RecQ4-Zn^{2+}-binding domain (R4ZBD), shown in cyan. The R4ZBD coordinates a Zn^{2+}-ion (purple sphere). PDB code: 5LST.

1.5.2. Biochemical features

In 2006, Macris and co-workers showed that RecQ4 has an ATPase function that is activated by DNA, with ssDNA being much more effective than dsDNA, but at that time they did not detect any helicase activity. They also discovered RecQ4 possessing a single-strand DNA annealing activity (Macris et al., 2006). RecQ4 DNA helicase activity was only later detected in reactions containing an excess of ssDNA to prevent its reannealing (Xu & Liu, 2009).
Suzuki and collaborators suggested that the previous failures in detecting RecQ4 helicase activity could be due to the low concentration of ATP used in the biochemical assays (Suzuki et al., 2009). RecQ4 has a 3'-5' directionality (Kitao et al., 1998; Kitao et al., 1999) and its helicase activity was shown to be inactivated by mutations in the conserved helicase core domain, responsible for the observed unwinding activity (Rossi et al., 2010).

A variety of Zn knuckles have been shown to bind both ssDNA and ssRNA, most having a preference for ssRNA (Buckman et al., 2003; Loughlin et al., 2011). Marino and collaborators conducted a detailed biochemical analysis by which they discovered that both human and frog Zn knuckles bind to a variety of nucleic acid substrates, with a mild preference for RNA (Marino et al., 2016).

For its part, the C-terminal domain (the last 92 aa residues) was found to be crucial for DNA unwinding. The RecQ4$^{427-1116}$ variant is in fact able to separate a 3'-overhang (3'-OH) DNA-substrate upon ATP addition but, surprisingly, helicase activity is more than five-fold increased for the RecQ4$^{427–1208}$ variant. RecQ4$^{427–1116}$ features all basic structural requirements for dsDNA separation, however the last 92 aa of RecQ4$^{427–1208}$ seem to contain additional elements, which are important for RecQ4s helicase activity (Kaiser et al., 2017).

The authors suggested a new functional model based on these features: the ssDNA could enter the RecQ4 helicase via HD2 and be directed towards HD1, positioning the dsDNA entity in close proximity to the RecQ4 C-terminal domain (Kaiser et al., 2017).

1.5.3. Cellular role

RecQ4 is the only RecQ helicase known to be present in both nucleus and mitochondria. In the nucleus it is predominantly localized to the nucleoplasm with a fraction of the protein colocalizing to telomeres and the nucleolus (Yin et al., 2004; Petkovic et al., 2005; Woo et al., 2006). In fact, RecQ4 has been shown to function in telomere maintenance and localizes to telomeres, associates with Shelterin proteins TRF1 and TRF2 and is able to resolve telomeric D-loop structures with the help of TRF1, TRF2 and POT1 specialized telomere DNA binding proteins (Ghosh et al., 2012). It was also found in the cytoplasm and more recently inside mitochondria: its intracellular localization is dynamic and appears to be cell type-specific (Petkovic et al., 2005; Croteau et al., 2012; De et al., 2012). There are evidences that RecQ4 loss alters mitochondrial integrity. The Mitochondrial Localization Signal (MLS), identified within the helicase N-terminus, is responsible for the localization of the RecQ4-p53 complex to the mitochondria. RecQ4-p53 interaction is disrupted after stress, allowing p53 translocation to the nucleus (De et al., 2012). RecQ4 was also found to be critical for skeletal development by modulating p53 activity *in vivo* (Lu et al., 2015). Like Sld2 in *S. cerevisiae*, RecQ4 is suggested to be involved in the assembling of DNA replication initiation complex (Thangavel et al., 2010), interacting with MCM and GINS complexes, and its N-terminus is essential for replication initiation in frog oocyte extract and chromatin binding by DNA polymerase α (Matsuno et al., 2006).

Several findings also indicate the involvement of RecQ4 in DNA repair. It colocalizes and functionally interacts with APE1, FEN1 and polymerase β, indicating a role of RecQ4 in base excision repair (Schurman et al., 2009).

RecQ4 has a crucial role in the first step of HR, in which is directly involved as an important participant. Its depletion severely reduces HR-mediated repair and 5' end resection *in vivo*. Moreover, inactivation of its helicase activity impairs DNA end processing and HR-dependent

DSBR, although without affecting its interaction with MRE11 and CtIP, suggesting an important role for RecQ4's unwinding activity in the process (Lu et al., 2016).

RecQ4 also interact with poly(ADP-ribose) polymerase 1 (PARP-1), which is implicated in transcriptional regulation, DNA repair and recombination (Woo et al., 2006). RecQ4 foci coincide with foci formed by human RAD51 and regions of ssDNA after induction of DSBs. (Petkovic et al., 2005). These findings further support a role for RecQ4 in the repair of DSBs by HR. Evidences also showed that UV treatment of human cells resulted in the colocalization of the nuclear foci formed with RecQ4 and XPA. RecQ4 could then directly interact with XPA, and this interaction could be stimulated by UV irradiation (Fan & Luo, 2008).

Interestingly, human mutant cells lacking RecQ4 were defective in UV-induced S-phase arrest, whereas cells defective in BLM exhibited a normal S-phase arrest following UV irradiation. A targeted inhibition of RecQ4 expression in human 293 cells showed a defect in inducing S-phase (replication) arrest following UV treatment and hydroxyurea treatment. These results suggested that RecQ4 may then have a unique role in replication fork arrest (Park et al., 2006).

1.6. R-loops metabolism and their involvement in human diseases

The term "R-loops" was given to reflect the similarity of these structures to D-loops: the "R" represents the involvement of an RNA strand, as they are three-stranded nucleic acid structures formed when the RNA hybridizes to a complementary DNA strand and a displaced single-stranded DNA. They were described for the first time in 1976 by Richard J. Roberts and Phillip A. Sharp, when their formation *in vitro*, in the presence of 70% formamide, was visualized by electron microscopy (Berget et al., 1977; Chow et al., 1977). R-loops have been detected in various organisms from bacteria to mammals, playing crucial roles in regulating gene expression, DNA and histone modifications, immunoglobulin class switch recombination (a process that allows activated B cells to modulate antibody production), DNA replication and genome stability, accumulating at preferred regions all over the genome, such as pericentromeric DNA, telomeres, ribosomal DNA or transcription termination regions (Groh & Gromak, 2014).

R-loops can repress transcription and promote transcriptional termination (Huertas & Aguilera, 2003; Tous & Aguilera, 2007; Skourti-Stathaki et al., 2011). Furthermore, R-loops are clearly associated with epigenetic mechanisms governing transcription, post-translational histone modifications and in protecting some active promoters from methylation (Ginno et al., 2012).

R-loops can be a dangerous source of DNA damage. They can sensitize DNA to damaging agents (Santos-Pereira et al., 2013), induce transcription-associated recombination (Huertas & Aguilera, 2003), double-strand breaks (DSBs), chromosome breaks, fragile site instability, and cause chromosome loss (Li & Manley, 2005; Sordet et al., 2009; Tuduri et al., 2009; Stirling et al., 2012). Therefore, cells need to tightly regulate the levels of R-loops as their

unbalance and dysregulation can impair R-loop-regulated processes, cause genome instability and lead to several human diseases.

Many proteins, required for efficient transcriptional elongation, termination, polyadenylation, RNA splicing, packaging and export, can regulate cellular R-loops levels either directly or indirectly, mostly by preventing RNA from hybridizing to DNA, thus reducing excessive R-loops accumulation (Huertas & Aguilera, 2003; Li & Manley, 2005; Skourti-Stathaki et al., 2011; Wahba et al., 2011; Stirling et al., 2012; Santos-Pereira et al., 2013). DNA topology can also influence and determine hybridization of RNA to DNA and, in this context, topoisomerases can play important roles in modulating R-loops levels. TOP3B, for example, is able to relax negative supercoiled DNA and reduces transcription-generated R-loops *in vitro*, promoting transcription, protecting DNA from damage and reducing the frequency of chromosomal translocations (Tuduri et al., 2009; Yang et al., 2014).

Cells also possess specific enzymes, including the members of the RNase H family, that specifically degrade the RNA in R-loops (Cerritelli & Crouch, 2009), and helicases that can unwind RNA/DNA hybrids (Misho et al., 2011; Skourti-Stathaki et al., 2011).

Mutations in proteins implicated in R-loops resolution can cause several devastating human diseases, often related to neurodegeneration. The RNA/DNA helicase SETX is required to resolve R-loops at termination elements, releasing RNA for degradation by the 5'–3' "torpedo" exonuclease Xrn2 prior to termination (Skourti-Stathaki et al., 2011) and its mutations cause amyotrophic lateral sclerosis type 4 (ALS4) and a recessive form of ataxia oculomotor apraxia type 2 (AOA2). These disorders are characterized by a progressive degeneration of motor neurons in brain and spinal cord, muscle weakness and atrophy (Chen et al., 2004; Moreira et al., 2004).

Endonucleases such as RNase H enzymes are able to cleave the RNA of RNA/DNA hybrids in a sequence-independent manner, promoting and maintaining genome stability by resolving R-loops that form during transcription and by removing misincorporated ribonucleotides from the DNA (Cerritelli & Crouch, 2009).

Mutations in any of the three subunits of RNase H2 cause Aicardi-Goutières syndrome (AGS), a neurological inflammatory disease, which resembles a congenital viral infection and is associated with accumulation of ribonucleotides in the DNA (Crow & Rehwinkel, 2009).

TREX1, ADAR1 and SAMHD1 are enzymes able to process retroelement-derived nucleic acids and help to suppress retroelements expansion in the host genome and their recognition by the immune system (Lee-Kirsch et al., 2014). Genome-wide studies have demonstrated that RNA/DNA hybrids are particularly abundant at retrotransposon elements in yeast cells (Chan et al., 2014), and the expansion of retroelements, due to mutations in TREX1, ADAR1 or SAMHD1, may lead to increased RNA/DNA hybrid levels, contributing to autoimmunity in AGS. Indeed, it has been demonstrated that RNA/DNA hybrids can be sensed by toll-like receptor 9 (TLR9) to induce pro-inflammatory cytokine and antiviral interferon production in dendritic cells (Rigby et al., 2014).

R-loops have been also proposed to play a role in triplet expansion pathologies. In fact, they can form over-expanded GAA and CGG repeats in cells from Friedreich's Ataxia (FRDA) and

Fragile X syndrome (FXS) patients, respectively (Groh et al., 2014). The abundance of these stable R-loops correlates with expansion size and they colocalize with the repressive chromatin marks characteristic of these diseases (Groh et al., 2014).

R-loops can corrupt genome integrity, increase DNA sensitivity to damaging agents, promote the formation of DSBs, chromosome breaks, fragile site instability, chromosome loss and recombination events (Hamperl & Cimprich, 2014), all of them important hallmarks of cancer. Moreover, mutations in proteins controlling R-loops levels have been identified in tumors and the link between R-loops and tumoral onset and progression has been validated by several findings. Kaposi's sarcoma-associated herpesvirus (KSHV), which causes multiple AIDS-related cancers, encodes the ORF57 protein, which can sequester the host hTREX complex, important for mRNA processing and export (Jackson et al., 2014). Sequestration of hTREX leads to KSHV-induced accumulation of R-loops and causes damage to the host DNA, contributing to tumorigenesis (Jackson et al., 2014).

Efficient transcription of the oncogene c-MYC requires that R-loops levels are kept low by the activity of DNA topoisomerase IIIB, which is recruited to arginine-methylated histones by the tudor domain containing 3 (TDRD3) protein (Yang et al., 2014). This R-loops-mediated mechanism of c-MYC gene regulation is relevant to tumor progression in breast cancer, which frequently shows overexpression of both c-MYC and TDRD3 (Yang et al., 2014). Yang and co-workers speculated that TDRD3 levels suppress R-loops in c-MYC, thereby allowing its enhanced expression, which correlates with poor cancer prognosis (Hynes & Stoelzle, 2009). Interestingly, tumor suppressor BRCA2, which is mutated in breast and ovarian cancer, is required to prevent R-loops accumulation and genome instability (Bhatia et al., 2014). However, it still remains unclear if R-loops play a specific role in transcription dysregulation in cancer.

R-loops have been also found implicated in cell senescence and play a complex and dynamic role in telomere length maintenance and cellular proliferative potential. In particular, the telomeric noncoding RNA TERRA forms R-loops which are induced when R-loops suppressors, such as RNase H or Thp2, are lost (Balk et al., 2013; Pfeiffer et al., 2013). Telomeric R-loops, in the absence of telomerase, can promote recombination-mediated telomere elongation via Rad52, and this delays the onset of cellular senescence (Balk et al., 2013). In contrast, in Rad52-deficient cells, R-loops accumulation leads to telomere shortening and premature senescence (Balk et al., 2013). Interestingly, cells from AOA2 patients with senataxin mutations contain shorter telomeres, suggesting a possible involvement of SETX in telomere stability (De Amicis et al., 2011).

Since their involvement in gene regulation and maintenance of genome integrity, R-loops and their interactors also represent a potential therapeutic target and have yet to be fully exploited in drug design, although various compounds that modulate DNA supercoiling and inhibit DNA topoisomerases as well as small-molecule inhibitors for RNase H2 were recently identified (Sordet et al., 2009; Powell et al., 2013; White et al., 2013).

1.7. RecQ helicases and R-loops

From unpublished studies previously carried out in our laboratory, RecQ4 was identified as a helicase able to efficiently resolve R-loops both *in vitro* and *in vivo*. A detailed biochemical analysis of the catalytic core of RecQ4, using a different DNA and RNA substrates, was performed and showed that, although the protein binds a variety of nucleic acids, it has a preference for R-loops; moreover R-loops are also the substrate most efficiently unwound by RecQ4.

When compared with all other human RecQ helicases, RecQ4 exhibited the highest R-loops resolving activity. WRN displayed a low level of R-loops unwinding activity, consistent with previous observation that WRN has limited helicase activity towards RNA:DNA hybrids such as Okazaki fragments (Chakraborty et al., 2010). Although RecQ5 has been shown to catalyse annealing of RNA to DNA substrates (Khadka et al., 2016), and has been suggested to be involved in R-loop metabolism (Kanagaraj et al., 2010), our colleagues observed only a weak unwinding activity towards R-loops *in vitro*.

These biochemical results were confirmed by a further *in vivo* analysis. Fibroblasts originating from a RTS patient with a RecQ4 truncated at the beginning of the helicase domain (maintaining an intact N-terminal domain, which is essential for DNA replication and cell viability) displayed a high frequency of R-loops, whereas the enforced expression of a full length RecQ4 in those same cells caused a marked decrease in the R-loops count.

Furthermore, shRNA-mediated downregulation of RecQ4 also caused accumulation of R-loops in a colon carcinoma cell line.

These results provided evidence for a novel functional role of RecQ4, during DNA replication, in resolution of R-loops at stalled replication forks (Mojumdar & Kenig, unpublished data).

Recently, loss of Sgs1, a yeast orthologue of human helicase BLM, was found having an important role in DNA replication and repair as its depletion increases R-loops accumulation and sensitizes cells to transcription-replication collisions. Moreover, analysis of BLM fibroblasts or by depletion of BLM from human cancer cells confirms a role for Sgs1/BLM in suppressing R-loops-associated genome instability across species. In support of a potential direct effect, BLM is found physically proximal to DNA:RNA hybrids in human cells, and can efficiently unwind R-loops *in vitro*. These evidences confirm that Sgs1/BLM have a crucial role in R-loops suppression and support the ideal involvement of RecQ helicases in stabilizing replication forks and as modulators of R-loop–mediated genome instability (Chang et al., 2017).

1.8. Present contribution

RecQ helicases are ubiquitous DNA unwinding enzymes, essential in the maintenance of genome integrity by acting at the interface between DNA replication, recombination and repair. Among RecQ helicases, RecQ4 is involved in DNA replication and maintenance of genome stability. Germ-line mutations in the RECQ4 gene give rise to three distinct human genetic disorders (Rothmund-Thomson, RAPADILINO and Baller-Gerold syndromes), characterized by genetic instability, growth deficiency and predisposition to cancer.

RecQ4 is one of the less biochemically characterized RecQ helicases, despite its role in genetic diseases and carcinogenesis. From studies previously carried out in our laboratory, we know that the catalytic domain of human RecQ4 binds and unwinds R-loops with a higher efficiency than other DNA or RNA substrates. Moreover, when compared to the other human RecQ helicases, RecQ4 shows the highest level of R-loops unwinding activity.

My PhD project focused on studying the contribution of additional domains (N- and C-terminal domain) of RecQ4 in binding different DNA and RNA substrates and in enhancing the catalytic ability of the helicase core in unwinding different nucleic acids substrates, by expressing and purifying them in large amounts suitable for biochemical and structural studies. The cloning and part of the expression and purification strategies were based on previous bioinformatics (Marino et al., 2013) and biochemical studies (Marino et al., 2016; Mojumdar et al., 2016) and were meant to provide a functional validation to newly identified and independent protein domains.

These findings are described into RESULTS AND DISCUSSION chapter.

CHAPTER 2

MATERIALS AND METHODS

2.1. Constructs and cell strain used in this study

Commercially available bacterial expression vectors and one bacterial expression vector kindly supplied by EMBL were used to produce the DNA constructs for the expression in *E. coli* of the recombinant proteins used in this study, as already described in literature (Marino et al., 2016; Mojumdar et al., 2017). The main features of these vectors are summarized in Table 2.1.

Vector	Promoter	Antibiotic resistance	Tag	Protease	Size (bp)	Cloning method	Supplier
pETM-30	T7	Kanamycin	6His-GST, N-term	-	6346	Restriction enzymes	EMBL
pET SUMO/CAT	T7	Kanamycin	6His-Sumo, N-term	Sumo protease	6307	RF	Invitrogen
pET-28b	T7	Kanamycin	6His- C-term	-	5368	Restriction enzymes	Novagene

Table 2.1. **List of vectors used for the cloning of RecQ4.** The vectors used and their characteristics: promoter, antibiotic resistance, fusion tag and protease eventually used to cleave the tag, size, supplier and the cloning strategy used. RF = restriction free cloning (van den Ent and Lowe, 2006).

The vectors containing RecQ4 DNA fragments were transformed into Rosetta 2 (DE3) *E. coli* strain. Rosetta 2 (DE3) host strain is BL21 derivative and is designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* but common in eukaryotic genes. This strain provides tRNAs for 7 rare codons (AGA, AGG, AUA, CUA, GGA, CCC and CGG) in a plasmid (pRARE) containing a chloramphenicol resistance. The tRNA genes are driven by their native promoters.

BL21 indicates the basic *E. coli* strain used for protein expression; it lacks both *lon* protease and *ompT* protease that can degrade the expressed proteins. DE3 indicates that the host is a lysogen of bacteriophage λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. Such strains are suitable for production of proteins from target genes cloned in pET vectors and IPTG is required to maximally induce expression of the T7 RNA polymerase in order to express recombinant genes cloned downstream of a T7 promoter.

2.2. Site directed mutagenesis

In order to understand the role of some conserved functional residues and two patient mutations in the C-terminal domain, as well as residues previously predicted to be functionally important in nucleic acid interaction in the N-terminal domain (Marino et al., 2016), site-directed mutagenesis was used to mutate critical residues in the pETM-30 and pET-28b vectors containing RecQ4 fragment 335-427 and 1124-1208, respectively. Mutants generated are listed in Table 2.2.

Mutant	Template DNA plasmid	Primer forward (5'-> 3')	Primer reverse (5'-> 3')
C403A/N406A	pETM-30 (RecQ4 335- 427)	CCAAGGAGTCTGCGTTCCT GGCGGAGCAGTTCGATCAC TGGGCAGCCC	GGGCTGCCCAGTGATCGAA CTGCTCCGCCAGGAACGCA GACTCCTTG
H411A/C416A	pETM-30 (RecQ4 335- 427)	GAGCAGTTCGATGCGTGG GCAGCCCAGGCGCCCCGGC CAGCAAGTGAG	CTCACTTGCTGGCCGGGGC GCCTGGGCTGCCCACGCAT CGAACTGCTC
N406C	pETM-30 (RecQ4 335- 427)	GAGTCTTGTTTCCTGTGTG AGCAGTTCGATCAC	GTGATCGAACTGCTCACAC AGGAAACAAGACTC
F404A/W412A	pETM-30 (RecQ4 335- 427)	GGAGTCTTGTGCGCTGAAC GAGCAGTTCGATCACGCGG CAGCCCAG	CTGGGCTGCCGCGTGATCG AACTGCTCGTTCAGCGCAC AAGACTCC
R1158A	pET-28b (RecQ4 1124-1208)	GAGAAGTTCTCCAGCGCAG CTGTGGCCCGCATC	GATGCGGGCCACAGCTGC GCTGGAGAACTTCTC
R1162A	pET-28b (RecQ4 1124-1208)	CAGCAGGGCTGTGGCCGC AATCTTCCACGGCATC	GATGCCGTGGAAGATTGCG GCCACAGCCCTGCTG
P1170A	pET-28b (RecQ4 1124-1208)	CACGGCATCGGAAGCGCCT GCTACCCGGCCCAG	CTGGGCCGGGTAGCAGGC GCTTCCGATGCCGTG
R1185A	pET-28b (RecQ4 1124-1208)	GACCGACGCTTCTGGGCAA AATACCTGCACCTG	CAGGTGCAGGTATTTTGCC CAGAAGCGTCGGTC
K1186A	pET-28b (RecQ4 1124-1208)	CGACGCTTCTGGAGAGCAT ACCTGCACCTGAGC	GCTCAGGTGCAGGTATGCT CTCCAGAAGCGTCG
T1200R	pET-28b (RecQ4 1124-1208)	CTGGTGGGCCTGGCCAGG GAAGAGCTCCTGCAG	CTGCAGGAGCTCTTCCCTG GCCAGGCCCACCAG

Table 2.2. List of all mutants and primers used for site directed mutagenesis. The introduced mutations are showed in red.

All reactions were performed using double-stranded plasmid DNA as a template and a pair of complementary oligonucleotide primers, incorporating the desired mutation.

Mutated vectors were generated in a 25 µl reaction containing 5-50 ng of template DNA, 125 ng forward primer, 125 ng reverse primer, 10 µl Pfu Turbo polymerase (2.5U/µl) and its 10X buffer and dNTPs mix using the following PCR cycles:

- 30 sec initial denaturation at 95°C;

  Then 18 cycles:

- 30 sec denaturation 95°C;
- 1 min annealing 55°C;
- 14 min extension 68°C (2 min/kb plasmid);
- ∞ at 4°C.

The reactions were then digested with 1 µl of the DpnI (stock 20,000U/ml) (NEB) restriction enzyme at 37°C, for 1.5 hr, to digest the methylated parental DNA. A 4 µl aliquot from the reaction mixture was then used to transform 45 µl DH5α competent cells (Stratagene) and the cell mixture was streaked on LB agar plates, containing kanamycin at a final concentration of 50 µg/µl. After overnight incubation at 37°C, preparation of plasmid DNA from isolated colonies proceeded.

2.2.1. Preparation of plasmid minipreps

Single colonies from agar plates were inoculated into 5 ml LB broth medium supplemented with the appropriate antibiotics and the cultures were grown for 16 hours at 37°C with shaking at 200 rpm. The cells were then pelleted by centrifugation at 3,500 g for 15 minutes at 4°C.

Plasmid DNA was purified from cell pellets using the QIAprep Spin Miniprep kit (Qiagen), according to the manufacturer's protocol, and stored at -20°C.

The amplified PCR products were analysed by agarose gel electrophoresis in 1% agarose in TBE Buffer. The typical yield of plasmid DNA was ~50ng/µl, checked spectrophotometrically measuring the absorbance at a wavelength of 260 nm using a Nanodrop (Thermo Scientific).

2.3. Protein expression and purification

2.3.1. Transformation

The expression vectors with the codon sequence for different portions of the human RecQ4 helicase were transformed into *E. coli* Rosetta 2(DE3) competent cells.

For the transformation, each plasmid was added to a 50 µl aliquot of Rosetta 2 (DE3) competent cells. The mixtures were incubated on ice for 30 min, heat-shocked at 42°C for 45 seconds and incubated on ice for 2 min. LB (Luria Bertani) medium, preheated to 42°C, was added to the mixture tubes to a final volume of 500 µL. The transformation reactions were then incubated for 1 hour at 37°C, shaking at 650 rpm in Eppendorf Thermomixer. After the transformation, the transformed cells were inoculated in LB broth with proper antibiotics (small inoculum): kanamycin at a final concentration of 50 µg/µl and chloramphenicol at a final concentration of 34 µg/µl. The small inoculum was then over-night incubated at 37°C.

2.3.2. Expression conditions

Using T7 systems, protein expression can be induced either with the chemical inducer isopropyl-β-d-thiogalactoside (IPTG) or by manipulating the carbon sources during *E. coli* growth (auto-induction).

Like allolactose, IPTG binds to the lac repressor and releases the tetrameric repressor from the lac operator in an allosteric manner, thereby allowing the transcription of genes in the lac operon, such as the gene coding for beta-galactosidase, a hydrolase enzyme that catalyzes the hydrolysis of β-galactosides into monosaccharides. Anyway, unlike allolactose, the sulfur (S) atom creates a chemical bond which is non-hydrolyzable by the cell, preventing the cell from metabolizing or degrading the inducer. The concentration of IPTG therefore remains constant and the expression of lac p/o-controlled genes would not be inhibited during the experiment (https://www.goldbio.com).

Protein production using autoinduction is based on diauxic growth of *E. coli* under the dynamic control of lac operon regulatory elements in a medium with mixtures of glucose and lactose. It omits the step of adding inducer to start protein production. For the experiment, an autoinduction medium containing glucose, glycerol, and lactose as carbon substrates and NH^{4+} as sole nitrogen source without addition of amino acids and vitamins is developed. As the *E. coli* cultures grow, they consume the glucose first. As the glucose runs out, they are forced to use the lactose which drives expression of the T7 promoter. Glycerol helps support growth without inhibiting T7 protein expression (Studier, 2005).

Different induction methods were used for distinct constructs involved in this study. Induction by IPTG was used for N-terminal domain (0.2 mM IPTG) and autoinduction was

used for the catalytic core of human RecQ4 as already described in literature (Marino et al., 2016; Mojumdar et al., 2017).

For C-terminal domain, different IPTG concentrations (0.1 mM to 1 mM) were tested before scaling up and 0.1 mM IPTG concentration was finally detected as the best one in term of both expression and solubility of the domain.

2.3.3. Purification techniques used in this study

Immobilized Metal Ion Affinity Chromatography (IMAC)

IMAC is a specialized variant of affinity chromatography where the proteins or peptides are separated according to their affinity for metal ions that have been immobilized by chelation to an insoluble matrix. At pH values around neutral, the amino acids histidine, tryptophan and cysteine form complexes with the chelated metal ions (e.g., Zn^{2+}, Cu^{2+}, Cd^{2+}, Hg^{2+}, Co^{2+}, Ni^{2+}, and Fe^{2+}). They can then be eluted by reducing the pH, increasing the mobile phase ionic strength or adding ethylenediaminetetraacetic acid (0.05 M) to the mobile phase. This technique is especially suited for membrane proteins, protein aggregates where detergents or high-ionic-strength buffers are required and, as in this case, for purifying recombinant proteins as poly-histidine fusion proteins. Proteins fused to a poly-histidine tag can be purified using a Nickel-NTA Resin, that uses nitrilotriacetic acid (NTA), a tetradentate chelating ligand, in a highly cross-linked 6% agarose matrix. Nickel generally provides good binding efficiency to His-tagged proteins but tends to bind nonspecifically to endogenous proteins that contain exposed histidine or histidine clusters.

For this reason, a low concentration of imidazole, which competes with the his-tag for binding to the metal-charged resin and thus is used for elution of the protein from an IMAC column, is typically added to both binding and wash buffers to interfere with the weak binding of other proteins and to elute any proteins that weakly bind. His-tagged protein is then eluted with a higher concentration of imidazole in Nickel Elution Buffer B.

Heparin Affinity Chromatography

Heparins are negatively charged polydispersed linear polysaccharides which have the ability to bind a wide range of biomolecules. The structure and negative charge of heparin enable it to mimic DNA in its overall binding properties. Heparin is used as a chromatography resin, acting as both an affinity ligand and an ion exchanger, as binding to it involves both charge and ligand specificity. Its polyanionic structure can mimic nucleic acids like DNA and RNA, making it useful for purification of nucleic acid-binding proteins including DNA and RNA polymerases and transcription factors. Also, since heparin is not susceptible to bacterial RNases and DNases, it is an ideal ligand to use for purification of DNA binding proteins from

bacterial lysates. Sepharose High Performance is the base matrix for HiTrap Heparin HP 5 ml, the column used in this study.

Size Exclusion Chromatography (SEC)

SEC is a method that allows to separate the proteins according to their molecular weight, by filtration through a gel matrix. The gel consists of spherical beads containing pores of a specific size distribution. Separation occurs when molecules of different sizes are included or excluded from the pores within the matrix. Small molecules diffuse into the pores and their flow through the column is retarded according to their size, while large molecules do not enter the pores and are eluted in the column's void volume.

SEC is also useful for an estimate of the protein size. A standard calibration curve was plotted for the determination of the size and the molecular weight of each protein of this study: for each calibration standard, the logarithm of its known molecular weight, log (MW), was plotted against its normalized elution volume (Ve/Vo). Using the least squares method, the following linear equation was calculated: $log(MW) = m(Ve/Vo) + b$, where m is the slope and b is the intercept.

The equation was then used for the determination of the log (MW) of the proteins of interest based on their elution volume (Ve) obtained by size-exclusion chromatography with the Superdex-200 or Superdex-75 column, respectively.

2.3.4. Large scale expression and purification

The expression and purification protocol of both **N-terminal domain** and **catalytic core** of RecQ4 was already described in literature (Marino et al., 2016; Mojumdar et al., 2017). Briefly:

N-terminal domain

GST-tagged wild type and mutant proteins were expressed in Rosetta 2(DE3) cells. Small inoculum was added in Terrific Broth (TB, Sigma) in ratio 1:40 (112.5 ml of small inoculum were added in 4.5 L of TB) at 18°C overnight, following induction with 0.2 mM IPTG in the presence of 0.1 mM ZnSO4. The cells were harvested at 5000 rpm for 30 min at 4°C and pellets were frozen at -80°C. The cell pellets from 1 L of culture were thawed and resuspended in 40 mL of affinity binding buffer (Nickel Binding Buffer A: 50 mM Hepes pH 7.5, 0.5 M NaCl, 1 mM TCEP, 5% glycerol, 10 mM Imidazole) supplemented with one tablet of Complete Protease Inhibitor Cocktail (Roche), 5 mM MgCl$_2$ and 0.1 mg/ml DnaseI (Sigma). The resuspended solution was incubated on ice for 30 min and stirred with a magnetic bar, in order to allow the lysis. The suspension was then sonicated on ice 15 min with short pulses of 15 seconds each, followed by 45 seconds pause, at 40% amplitude (Soniprep 150). The cell

lysates were clarified by centrifugation for 45 min at 26,000 g and the supernatant filtered through a 0.22 mm cut-off filter before applying to affinity columns.

Proteins were purified by Nickel-NTA fastflow resin and washed with high salt concentration to eliminate DNA contamination and aspecific binding. After the elution with 5 times CV of Nickel Elution Buffer B (50 mM Hepes pH 7.5, 0.5 M NaCl, 1 mM TCEP, 5% glycerol, 500 mM Imidazole), the proteins were run on a size exclusion chromatography (Superdex-200) in 250 mM NaCl, 20 mM Tris-HCl pH 7.5, 5% glycerol and 5 mM β-mercaptoethanol.

Catalytic core

The wild type protein was expressed in Rosetta 2(DE3) cells using auto-induction method (Studier, 2005) at 17°C for 48 hours and harvested by centrifugation. The cells were harvested at 5000 rpm for 30 min at 4°C and pellets were frozen at -80°C. The cell pellets from 1 L of culture were thawed and resuspended in 40 mL (for cells grown in TB) of affinity binding buffer (Nickel Buffer A: 50 mM Hepes pH 8.0, 0.5 M NaCl, 1 mM TCEP, 5% glycerol, 10 mM Imidazole), supplemented with 2 mM of the protease inhibitor AEBSF (4-(2-aminoethyl) benzenesulfonyl fluoride, Sigma) and/or one tablet of Complete Protease Inhibitor Cocktail (Roche), 5 mM $MgCl_2$ and 0.1 mg/ml DnaseI (Sigma).

The resuspended solution was incubated on ice for 30 min and stirred with a magnetic bar, in order to allow the lysis. The suspension was then sonicated on ice 15 min with short pulses of 15 seconds each, followed by 45 seconds pause, at 40% amplitude (Soniprep 150). The cell lysates were clarified by centrifugation for 45 min at 26,000 g and the supernatant filtered through a 0.22 mm cut-off filter before applying to affinity columns. The clarified lysate was loaded onto HisTrap FF 5 ml column (GE Healthcare) pre-equilibrated in Nickel Buffer A, using ÄKTApurifier instrument (GE Healthcare). The column was washed with 10 column volumes (CV) of Nickel Buffer A with 25 mM Imidazole and 2 M NaCl (to remove DNA contaminations), and the protein was eluted with a linear gradient over 6 CV to reach a concentration of 250 mM Imidazole. The peak fractions containing the protein of interest were pooled and the concentration of imidazole was brought back to 10 mM by diluting the sample with Buffer A.

Proteins expressed from pET-SUMO/CAT expression vectors were subjected to removal of SUMO tag by SUMO Protease. SUMO Protease is a highly active cysteinyl protease which is a recombinant fragment of Ulp1 (Ubl-specific protease 1) from *Saccharomyces cerevisiae* (Mossessova and Lima, 2000). SUMO Protease recognizes the tertiary structure of the ubiquitin-like (UBL) protein, SUMO rather than an amino acid sequence, cleaving in a highly specific manner. The optimal temperature for cleavage is 30°C; however, the enzyme is active over wide ranges of temperature and pH (pH 7.0-9.0).

Protein samples were kept under overnight cleavage with SUMO protease at a w/w ratio of 1:500 at 4°C, in the following buffer: 50 mM Tris-HCl (pH 8.0), β-mercaptoethanol, 250 mM NaCl. Cleavage was checked by SDS-PAGE gel.

Following digestion, SUMO Protease and His- SUMO tag were removed from the cleavage reaction by Nickel affinity chromatography using the polyhistidine tag at the N-terminus of the protease.

Cleaved reaction was passed through a HisTrap FF 5 ml column (GE Healthcare) pre-equilibrated in Nickel Buffer A.

The flow-through was then collected and further diluted 10 times by Heparin Buffer A (50 mM Hepes pH 7.5, 1 mM TCEP and 5% glycerol) to reduce the NaCl concentration to 50 mM. This sample was loaded onto HiTrap Heparin 5 ml column (GE Healthcare) equilibrated in Heparin Buffer A with 50 mM NaCl. The column was washed with 10 column volumes (CV) of Heparin Buffer A with 50 mM NaCl, and the protein was eluted with a linear gradient over 6 CV to reach a concentration of 1 M NaCl. The peak fractions containing the protein of interest were pooled and concentrated using Amicon® Ultra-4 centrifugal filter unit with 30 kDa cutoff (Millipore) to further purify it using size exclusion chromatography (Superdex-200 10/300 GL-GE Healthcare) in 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 5% glycerol and 5 mM β-mercaptoethanol.

C-terminal domain

The wild type protein and its mutants were expressed as 6His-tagged fusion proteins in Rosetta 2 (DE3) cells. Small inoculum was added in Terrific Broth (TB, Sigma) in ratio 1:40 (112.5 ml of small inoculum were added in 4.5 L of TB) at 18°C overnight, following induction with 0.1 mM IPTG. The cells were harvested at 5000 rpm for 30 min at 4°C and pellets were frozen at -80°C. The cell pellets from 1 L of culture were thawed and resuspended in 40 mL of affinity binding buffer (Nickel Binding Buffer A: 50 mM Hepes pH 8.0, 0.5 M NaCl, 1 mM TCEP, 5% glycerol v/v, 10 mM Imidazole) supplemented with one tablet of Complete Protease Inhibitor Cocktail (Roche), 5 mM $MgCl_2$ and 0.1 mg/ml DnaseI (Sigma). The resuspended solution was incubated on ice for 30 min, stirring with a magnetic bar, in order to allow the lysis. The suspension was then sonicated on ice 15 min with short pulses of 15 seconds each, followed by 45 seconds pause, at 40% amplitude (Soniprep 150). The cell lysates were clarified by centrifugation for 45 min at 26,000 g and the supernatant filtered through a 0.22 mm cut-off filter before applying to affinity columns. Proteins were purified by Nickel-NTA fastflow (Qiagen) resin and washed with high salt concentration to eliminate DNA contamination and aspecific binding. After the elution with 5 times CV of Nickel Elution Buffer B (50 mM Hepes pH 7.5, 0.5 M NaCl, 1 mM TCEP, 5% glycerol, 500 mM Imidazole), the proteins were run on a size exclusion chromatography (Superdex-75 10/300 GL-GE Helthcare) in 50 mM Hepes pH 8.0, 250 mM NaCl, 5% glycerol v/v and 5 mM β-mercaptoethanol.

The concentration of any protein was determined by using Nanodrop Instrument (Thermo Scientific) and protein purity was analyzed as described below.

2.4. Biochemical characterisation of protein

2.4.1. Determination of protein concentration

Protein concentrations were determined spectrophotometrically by measuring the absorbance at 280 nm and applying the Lambert-Beer equation using the theoretical extinction coefficient of each protein. Theoretical values of molecular weight (MW), isoelectric point (pI) and extinction coefficient at 280 nm of all purified proteins were calculated from the amino acid sequences using ProtParam tool available at the Expasy server (http://ca.expasy.org/tools/protparam.html).

2.4.2. Sodium Dodecyl Sulphate – PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples were analysed by electrophoresis under denaturing, reducing conditions using discontinuous SDS-polyacrylamide gels (4% for the stacking gel and 10% or 12% for the separating gel). All the samples were prepared by addition of 4x SDS sample buffer and heated at 95°C for 5 min. Electrophoresis was carried out at a constant voltage (200V) for 45-55 min in 1x Tris-glycine buffer. Gels were stained with Instant Blue, then stored in water.

2.4.3. Oligonucleotide preparation for helicase and nucleic acid binding assays

All the oligonucleotides used for EMSA and helicase assay were chemically synthesized and purified by reverse-phase high pressure liquid chromatography (Sigma-Aldrich, Suffolk, UK). Each oligo was resuspended in Tris-EDTA buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Oligonucleotide sequences used in this work are reported in Table 3.3.

Fork-DNA (D1:D2 and D1:D3), fork-RNA (R1:R2 and R1:R3), hybrid fork-DNA/RNA (D1:R2 and D1:R3) and hybrid fork-RNA/DNA (R1:D2 and R1:D3) substrates were prepared by annealing the complementary fluorescent labelled and unlabelled oligonucleotides at a 1:2 M ratio in annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA) by heating at 95°C for 5 min and slowly cooling at room temperature. The fluorescent labelled DNA and RNA oligonucleotides (6FAM label at 5' end and BHQ1 quencher at 3' end) are shown in Table 3.3. For Holliday Junctions the substrates D4:D5:D7:D8 and D4:D6:D7:D8 were prepared by annealing the oligos D4, D5, D6, D7 and D8 in a molar ratio of 1:1.5:1.5:1.5 in 50mM Tris-HCl pH 8.0 and 10mM $MgCl_2$; following denaturation at 100°C for 5 min, the DNA was allowed to anneal by overnight cooling to room temperature.

For D-loop (D4:D9:D11 and D4:D10:D11) and R-loop (R4:D9:D11 and R4:D10:D11) substrates the molar ratio used were 1:1.25:2.5 in 6mM Tris-HCl pH 7.5, 7mM $MgCl_2$, 50mM NaCl and 1mM DTT.

The annealing method used for D-loops and R-loops was heating at 99°C for 5 min followed by incubations at 67°C for 1 hour, at 37°C for 30 min and at 25°C for 3-4 hours or overnight. All the products were analyzed by electrophoretic mobility shift assay (EMSA) on 6% (w/v) polyacrylamide gels, where labelled unannealed strands were used as markers.

Strand name	Sequence (5'-3')
D1	[6FAM]CTACTACCCCCACCCTCACAACCTTTTTTTTTTTTTTT
D2	TTTTTTTTTTTTTTTGGTTGTGAGGGTGGGGGTAGTAG [BHQ1]
D3	TTTTTTTTTTTTTTTGGTTGTGAGGGTGGGGGTAGTAG
D4	[6FAM]GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATC
D5	ATCCTCTCTAGAGTCGACCTGCAGGCATGCAAGCTAAGCCATCTACTTCGT [BHQ1]
D6	ATCCTCTCTAGAGTCGACCTGCAGGCATGCAAGCTAAGCCATCTACTTCGT
D7	GATGAGTTGAGGCTGGGAAAAGTTACTGTAGCC
D8	GGCTACAGTAACTTTTCTCGACTCTAGAGAGGAT
D9	GTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAAC [BHQ1]
D10	GTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAAC
D11	GTTGTAAAACGACGGCCAGTGCCTTTTCCCAGCCTCAATCTCATCACTCTAGAGGATCCCCGGGTAC
R1	[6FAM]cuacuaccccccacccucacaaccuuuuuuuuuuuuuu
R2	uuuuuuuuuuuuuuugguugugagggtggggguaguag[BHQ1]
R3	uuuuuuuuuuuuuuugguugugagggtggggguaguag
R4	[6FAM]gcuugcaugccugcaggccagccucaaucacauc
Cap1	CTACTACCCCCACCCTCACAACC
Cap2	GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATC

Table 3.3. Oligonucleotides used in this study. 6FAM (6-Carboxyfluorescein) label at 5'end and BHQ1 (Black Hole Quencher 1) label at 3'end. DNA sequences are in capital letters and RNA sequences in small letters.

2.4.4. Electrophoretic Mobility Shift Assay

Binding of the single protein concentration (5 μM for the C-terminal domain) or binding of increasing amounts of protein (0, 0.1, and 0.2 μM for the helicase domain alone and together with the C-terminal domain; 0, 0.05, and 0.1 μM and for the helicase domain alone and together with the N-terminal domain) on the single strand labelled substrates (at final concentration of 20 nM) was detected by a gel mobility shift assay.

The nucleic acid binding assay was performed by mixing the purified proteins and each substrate in 20 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 8 mM DTT, 0.1 mg/ml BSA and 5% Glycerol in 20 μl of reaction volume, incubated at room temperature for 30 min. The reaction mixture was then loaded on a 6% non-denaturing polyacrylamide gel and run at 4°C in TBE buffer. Fluorescent labelled substrates were detected by fluorescent scanner (ImageQuant, GE Healthcare) and quantification of protein bound nucleic acid was performed with ImageQuant image analysis software (GE Healthcare). One-site total binding

model was used to fit the data, from three independent experiments, using GraphPad-Prism software.

2.4.5. Helicase Assay

The helicase activity was measured by using fluorescence resonance energy transfer (FRET) in which the fluorophore emits when the double stranded substrate is unwound by the helicase (Tani et al. 2010). The assay was performed in 20 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 50 mM KCl, 8 mM DTT, 0.1 mg/ml BSA and 5% Glycerol with 10 nM substrate, 3 mM ATP and 125 nM Capture strand to prevent reannealing (Cap1 for fork-DNA and fork-RNA; Cap2 for Holliday Junctions, D-loops and R-loops, Table 3.3) in 25 μl of reaction volume. The unwinding reaction was started by incubating increasing concentrations (0 to 160 nM) of the purified protein in the reaction mixture. The reaction mixture was incubated at 37°C for 30 min. The fluorescence intensity was recorded using Infinite F200 PRO TECAN instrument. To have a measurement of 100% unwinding the reaction was incubated at 95°C and measured. The assay was done in triplicate. The percentage of unwinding was calculated and plotted using GraphPad-Prism software.

2.5. Structural analysis of the C-terminal domain: CD spectroscopy

Circular dichroism (CD) spectra of the purified wild type C-terminal domain was recorded at 25°C on a Jasco J-810 Spectropolarimeter at wavelength 185–260 nm, in 10 mm quartz cells, band width 1 nm, response 1 sec, data pitch 0.1 nm and scanning speed 20 nm/min. Each peptide was dissolved at 50 μM in 50 mM phosphate buffer pH 8.0 and 50 mM NaF. Three scans were done for each sample and averaged and every spectrum was corrected for the buffer.

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Bioinformatics analysis

A bioinformatics analysis previously carried out on RecQ4 had identified the presence of a cysteine-rich region, predicted to fold into a Zn knuckle, immediately upstream of the helicase domain, and a revised prediction of the Sld2-homology region, extending throughout the whole of the yeast Sld2 replication factor (Marino et al., 2013). Both the human and *Xenopus* Zn knuckles of RecQ4 were found to bind a variety of nucleic acid substrates, with a mild preference for RNA. Besides, the segment located upstream the Zn knuckle, that is highly conserved and rich in positively charged and aromatic residues, strongly enhances binding to nucleic acids in both the human and *Xenopus* proteins (Marino et al., 2016).

We have further examined the last 100 amino-acid residues of RecQ4. This region does not show any similarity to known protein sequence, beside the RecQ4 orthologues; however, within this family, the level of conservation is significant, pointing to an essential role for this domain in the cellular function of RecQ4 proteins.

We run various threading/fold recognition algorithms, such as I-tasser, FFAS03, HHPRED, but none of these were able to predict the presence of a specific fold. However, secondary structure prediction algorithms consistently show the presence of 4 or 5 α-helices (Figure 3.1). This consideration opens up the intriguing possibility that the C-terminus may fold into a HRDC domain, as seen in a number of other RecQ helicases. If that is true, the lack of sequence homology would not be surprising, since the HRDC domains are very poorly conserved at the sequence levels.

As some positively charged residues of the C-terminal domain are conserved (such as Arg1158, Arg1162, Arg1185 and Lysine 1186), we predicted they may be functionally important, possibly being involved in nucleic acid interaction. To verify our prediction, these residues were mutated to alanine. Additionally, residues reported to be mutated in patients, Pro1170 and Thr1200, were also mutated and analyzed. On the basis of an earlier publication by Kääriäinen and colleagues, it is known that the mutation Thr1200ArgfsX26, found in a RAPADALINO patient, causes an early stop codon or a frameshift, both leading to truncated polypeptide (Kääriäinen et al., 1989). On the other hand, the mutation Pro1170Leu was found in concert with other two amino acid substitutions (Arg522Cys and Val799Met), but none of them was yet proven to be a pathogenic change (Siitonen et al., 2009).

Figure 3.1. Sequence analysis of the C-terminal domain of RecQ4. Sequence alignment (ClustalW2, Chenna et al., 2003) of the C-terminal domain of RecQ4 from various organisms. Identical or very similar residues are highlighted in cyan. Arg1158, Arg1162, Arg1185 and Lys1186, predicted to be functionally relevant, are highlighted by a green star. Pro1170 and Thr1200, mutated in RTS patients, are highlighted by a red star. The secondary structure prediction (Psipred, Heringa, 2000; McGuffin et al., 2000) shows the presence of five helices, shown above as purple rectangles.

3.2. Expression and purification of proteins used in this study

3.2.1. Expression and purification of the C-terminal domain (Ct) of hRecQ4 and its mutants

In order to test the above prediction and to better understand the role of hRecQ4 C-terminal domain (comprising amino acid residues 1124-1208) in the nucleic acid binding and unwinding activities of the catalytic core of the protein, the Ct region was cloned in a pET-28b vector to produce a 6His-tagged protein. This construct was then expressed in *E. coli* in TB at 18°C overnight, following induction with 0.1 mM IPTG, and purified by Nickel affinity chromatography, then injected on a gel filtration column Superdex 75 10/300 (Figure 3.2). As the region is rather small (84 amino-acid residues), we choose not to cleave the tag.

The protein expressed well (Figure 3.2 A), ran on size exclusion chromatography as a monomer with a molecular weight of about 10 kDa (Figure 3.2 B), and could be purified to a reasonable amount.

Conserved positively charged residues predicted to be functionally important, such as Arg1158, Arg1162, Arg1185 and Lys1186, were mutated to alanine. Additionally, residues reported to be mutated in RTS patients such as Pro1170 and Thr1200 were also mutated, to Ala and Arg, respectively. These mutant proteins were expressed and purified following the same protocol as for the wild type protein and the amounts and level of purity obtained appeared to be similar to the wild type protein.

The purity of the proteins was estimated to be > 90% by densitometry of Coomassie brilliant blue stained SDS-PAGE gel. The purified samples were kept at -80°C for further structural and biochemical characterization.
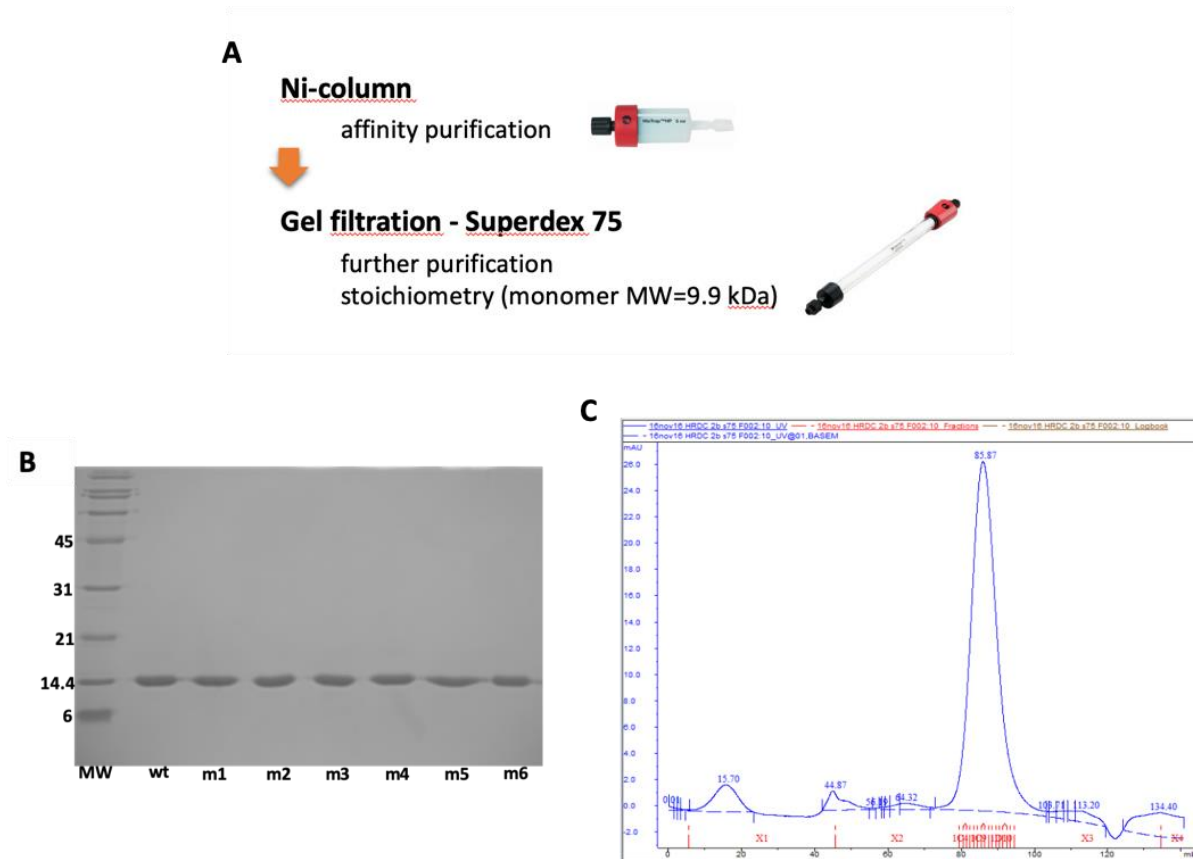
Figure 3.2. hRecQ4-Ct and its mutants. (A): Purification strategy of the Ct domain, comprising Nickel affinity purification, followed by size-exclusion chromatography. **(B)**: SDS-PAGE analysis of the expression and purification of Ct domain. Lane MW: molecular marker; wt: Wild type C-terminal domain; m1: Ct-Arg1158Ala; m2: Ct-Arg1162Ala; m3: Ct-Pro1170Ala; m4: Ct-Arg1185Ala; m5: Ct-Lys1186Ala; m6: Ct-Thr1200Arg (15% SDS-PAGE gel). **(C)**: Size exclusion chromatography indicates that the recombinant hRecQ4-Ct is a monomer of approximately 10 kDa in solution.

3.2.2. Expression and purification of the helicase domain (Hel) and the N-terminal domain (UpZnk) of hRecQ4 and its mutants

The catalytic domain of hRecQ4, comprising amino acid residues 445-1112 (hRecQ4-Hel), and the N-terminal region encompassing the last region of homology with the yeast Sld2 factor and the Zn knuckle, comprising amino acid residues 335-427 (hRecQ4-UpZnK), as well as their mutants, were expressed and purified as previously described in literature and in Chapter 2 (Marino et al., 2016; Mojumdar et al., 2017).
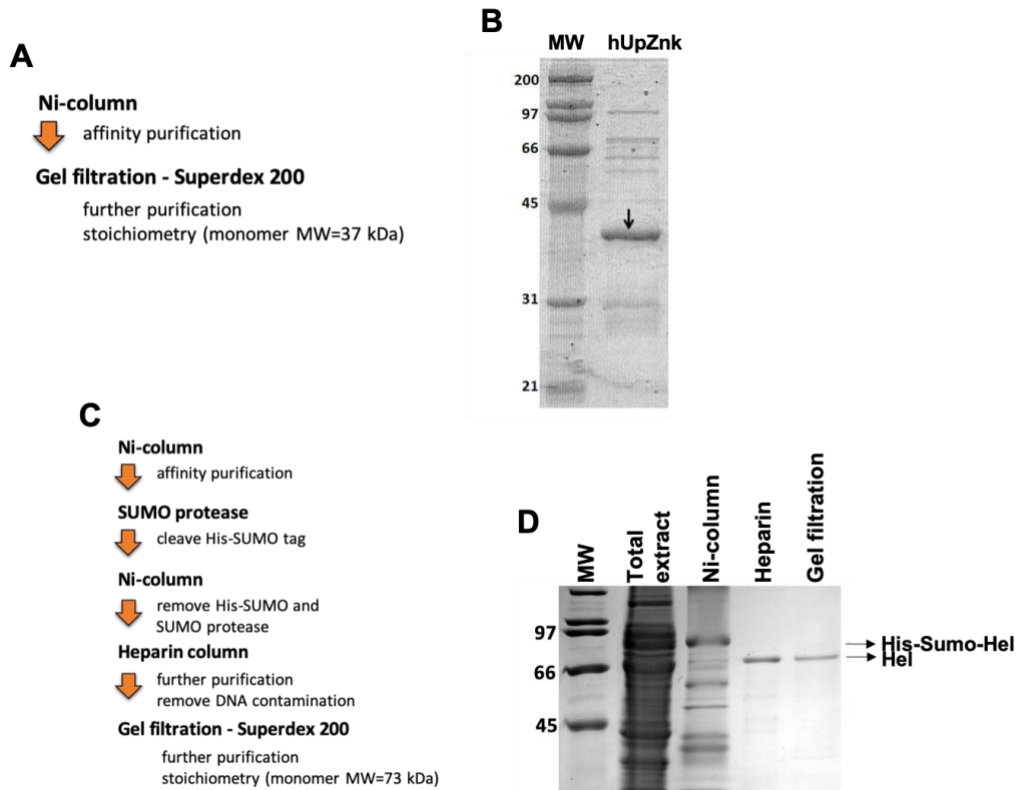
Figure 3.3. hRecQ4-UpZnk and hRecQ4-Hel. (A): Purification process of the N-terminal domain. After Nickel affinity purification, the 6His-GST tag was not cleaved to further stabilize the protein. **(B)**: SDS-PAGE analysis of the expression and purification of hRecQ4-UpZnk**. (C)**: Purification process of the catalytic domain of hRecQ4. After Nickel affinity purification, the 6His-SUMO tag was cleaved. **(D):** SDS-PAGE analysis of the expression and purification of hRecQ4-Hel.

3.3. Preparation of the nucleic acid substrates used in this study

We choose to test hRecQ4 unwinding activity of different DNA and RNA substrates which are able to mimic important physiological helicase substrates. Partially duplex substrates, such as fork-DNA and fork-RNA, hybrid fork-DNA/RNA and hybrid fork-RNA/DNA are interesting substrates because it is known that binding to ssDNA stimulates ATP hydrolysis for most helicases, whereas dsDNA usually does not provide the same level of activation. In fact, most helicases require a ssDNA "tail" adjacent to a dsDNA region in order to initiate unwinding of the duplex.

Several studies indicate that many RecQ proteins exhibit preferential activity on a wide range of substrates. Preferred substrates are branched DNA structures, including forked structures that mimic replication forks, and synthetic 4-way junctions that mimic Holliday junctions (HJs). In fact, human BLM and WRN proteins promote branch migration of HJs, crucial in homologous recombination (HR) and DNA replication (Harmon & Kowalczykowski, 1998; Constantinou et al., 2000; Karow et al., 2000).

RecQ helicases are also known to be active in unwinding a number of unusual DNA and RNA structures, including displacement loops (D-loops; an intermediate in homologous recombination reactions), DNA-RNA/RNA-DNA hybrids and R-loops, which form naturally during essential cellular functions such as transcription and replication and also be an important source of genome instability, a hallmark of cancer and genetic diseases (Bennett et al., 1998; van Brabant et al., 2000; Brosh et al., 2001; Machwe et al., 2002; Mohaghegh et al., 2001; Orren et al., 2002).

Binding to nucleic acid substrates was carried out using Electrophoretic Mobility Shift Assay (EMSA), based on the observation that protein–DNA complexes migrate more slowly than free linear DNA or RNA fragments when subjected to non-denaturing polyacrylamide gel electrophoresis.

The hRecQ4 capability in unwinding different nucleic acid substrates was instead tested by using a fluorescent assay which is based on the principle of fluorescent resonance energy transfer (FRET), in which the photon released from an excited fluorophore, known as the "donor," is absorbed by an "acceptor" molecule, which can be a fluorophore or a nonfluorescent molecule, also known as quencher. If the acceptor is a fluorophore then the absorbed energy will be emitted as a fluorescence characteristic of the acceptor; if it is a nonfluorescent (quencher) molecule, then the absorbed energy will be lost as heat (Mojumdar & Deka, 2019).

Radioactivity has been the predominant method of DNA labeling in EMSAs. Traditionally, DNA probes (labeled before performing the experiment to allow its specific detection after electrophoresis) have been radiolabeled with ^{32}P by incorporating an $[\gamma\text{-}^{32}P]dNTP$ during a 3' fill-in reaction using Klenow fragment or by 5' end labeling using $[\gamma\text{-}^{32}P]ATP$ and T4 polynucleotide kinase. Following electrophoresis, the gel is exposed to X-ray film to visualize the results.

Recent advances in fluorescent dyes and scanning methods have prompted the use of fluorescent tagging of DNA as an alternative to radioactivity for the advantages of easy handling and saving time. Indeed, numerous nonradioactive methods for performing EMSAs binding and unwinding assays are today available. Another advantage in using fluorescent probes unwinding is that it can be detected in real time and their use also lacks radiation hazards and the consequent problems of waste handling and disposal, thus improving safety. On the other hand, fluorescence is not necessarily easier or more convenient to use because it requires specialized equipment of its own and because quenching makes absolute and/or reproducible quantification sometimes difficult.

Moreover, there are two disadvantages of this method. The first is that radioactivity-based assays are far more sensitive, and allow to follow much smaller amounts of nucleic acids. The second disadvantage is a significant biological problem: chemically tagging a molecule with a fluorescent dye can radically change the structure of the molecule, which in turn changes the way that molecule interacts with other molecules. In contrast, intrinsic radiolabeling of a molecule can be done without altering its structure in any way (for example, substituting a

[32]P for a phosphorus atom or ^{14}C for a carbon atom does not change conformation, structure or any other property of the molecule). Thus, an intrinsically radiolabeled molecule is identical to its unlabeled counterpart.

However, we carried out EMSAs binding and FRET unwinding experiments using fluorescent probes due to regulatory concerns associated with the use of radioactivity.

In particular, all the DNA and RNA oligonucleotides used in this study for nucleic acid binding and unwinding experiments were prepared by annealing complementary fluorescent 5'-end-labeled with [6FAM] (6-Carboxyfluorescein) fluorophore and unlabeled oligonucleotides (EMSAs binding assays) or BHQ1 (Black Hole Quencher 1) labeled at 3'-end strands (FRET unwinding assays).

All the oligonucleotides used for EMSA binding assays (picture 3.4 A) and helicase assays (picture 3.4 B) were chemically synthesized and purified by reverse-phase high pressure liquid chromatography (Sigma-Aldrich). Fork-DNA, fork-RNA, hybrid fork-DNA/RNA and hybrid fork-RNA/DNA substrates were prepared by annealing in tube, by using a PCR thermal cycler, the complementary fluorescent labelled and unlabeled oligonucleotides at a 1:2 M ratio in a proper annealing buffer.

Holliday Junctions substrates were prepared by annealing the 4 oligos in a molar ratio of 1:1.5:1.5:1.5 in a proper annealing buffer, at 100°C for 5 min. The DNA was then allowed to anneal by overnight cooling to room temperature.

For D-loop and R-loop substrates were prepared by annealing the 3 oligos in a molar ratio of 1:1.25:2.5, in the adapted annealing buffer, as already carefully described in Chapter 2. All the products were analyzed by electrophoretic mobility shift assay (EMSA) on 6% (w/v) polyacrylamide gels, where a labelled unannealed strand was used as markers, and detected in-gel with the help of an appropriate imaging system (see Materials and Methods chapter for details).

It is important to stress that, in contrast to DNA, since RNA is very susceptible to degradation by enzyme-catalyzed hydrolysis, working with RNA substrates is more demanding due to both the chemical instability of RNA and because of the ubiquitous presence of RNases. Unlike DNases, which require metal ions for activity, RNases are found everywhere and have no requirement for metal ion cofactors. They can also maintain activity even after prolonged boiling or autoclaving. Special precautions must be taken when working with RNA substrates and all reagents and equipment must be specially treated to inactivate RNases prior to use.

All the oligonucleotides used in this study are showed in the picture 3.4 A, B.
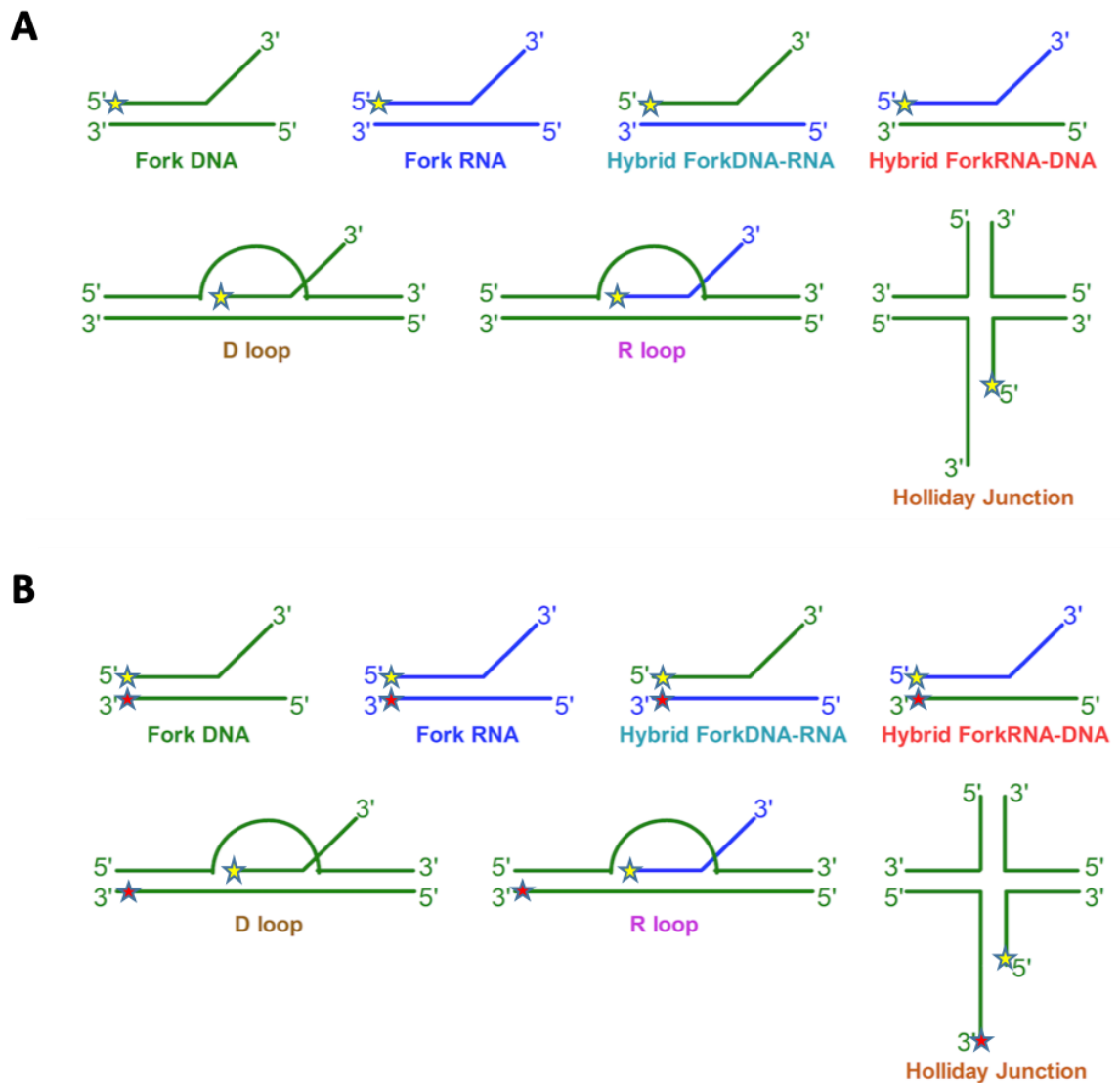
Figure 3.4. DNA and RNA substrates used in this study. (A): EMSA binding experiments substrates. [6FAM] fluorophore is represented by a yellow star. **(B):** FRET unwinding experiments substrates. [6FAM] fluorophore is represented by a yellow star; BHQ1 quencher is represented by a red star. All the substrates were prepared following the protocols described in Chakraborty & Grosse (2011).

3.4. Nucleic acid binding preference for the RecQ4 N-terminal region (hRecQ4-UpZnK, 335-427)

In order to evaluate the role of N-terminal region of hRecQ4, containing the additional Sld2 homologous region and the zinc knuckle (hRecQ4-UpZnK, residues 335-427), EMSA for detecting protein-nucleic acid interaction was performed with 10 nM of fluorescently labelled substrates and increasing concentration (0-400 nM) of hRecQ4-UpZnK.

Fluorescent labelled substrates were detected by a fluorescent scanner and quantification of protein bound to a nucleic acid was performed with a densitometric image analysis software (ImageQuant, GE Healthcare). The fraction of bound DNA was determined from the background-subtracted signal intensities by using the expression: bound/(bound + unbound). The resulting ratio is multiplied by 100 to get the percentage of nucleic acid binding.

One-site total binding function of GraphPad-Prism software was used to fit the data obtained from three independent experiments. This model does not measure nonspecific binding directly. Instead, it fits only total binding by assuming that the amount of nonspecific binding is proportional to the concentration of labeled ligand.

The N-terminal region shows affinity towards almost all the substrates with a preference for R-loops, Holliday Junctions, D-loops and hybrid fork-RNA/DNA (Figure 3.5 A, B). From experiments previously carried out in our laboratory, it is known that hRecQ4-Hel protein is able to bind several DNA and RNA substrates, with a preference for R-loops and fork-RNA, which seem to be the preferred substrates, followed by D-loops and Holliday Junctions. It is possible to speculate that these two distinct domains could then synergistically contribute to the substrate recruiting and selectivity.

To further determine the impact of the N-terminal region on nucleic acid binding activity of the catalytic core of hRecQ4, EMSA was then performed with 10 nM of the four substrates showing the highest affinity (R-loops, Holliday Junction, D-loops and hybrid fork-RNA/DNA) and increasing concentration (0-0.1 µM) of hRecQ4-Hel protein, together with or without an equimolar ratio of hRecQ4-UpZnk (0.1 µM). The presence of the hRecQ4-UpZnK seems to increase the binding affinity of the helicase core towards the best substrates. Surprisingly, the major contribution to bind the substrates seems to be due to the N-terminal region itself; interestingly, the EMSA band corresponding to the N-terminal region:nucleic acid complex shows a shift at a molecular weight higher than the one corresponding to the helicase core:nucleic acid complex. It is possible to speculate that this effect could be due to an unexpected stoichiometry of the complex itself (Figure 3.5 C, D).
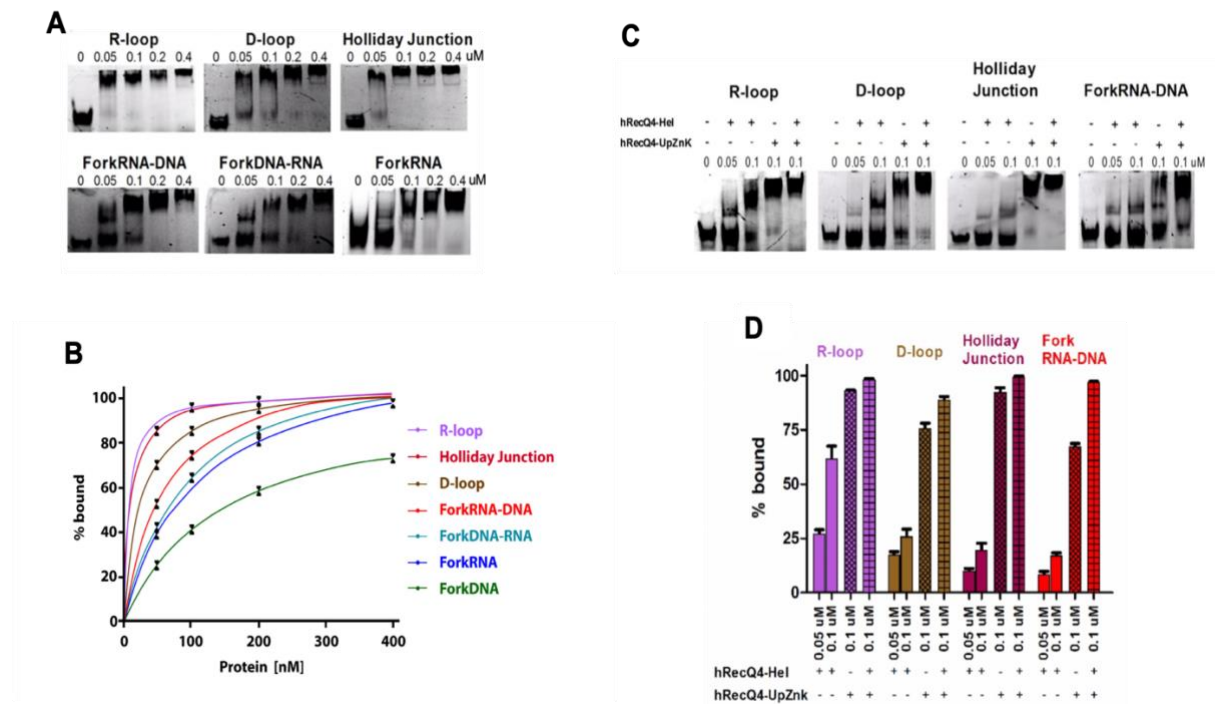
Figure 3.5. hRecQ4-UpZnK nucleic acid binding preferences. (A): EMSA gels showing nucleic acid binding at increasing concentration (0-0.4 μM) of hRecQ4-UpZnK with 10 nM substrates: R-loop (R4:D10:D11), D-loop (D4:D10:D11), Holliday Junction (D4:D6:D7:D8), hybrid fork-RNA/DNA (R1:D3), hybrid fork-DNA/RNA (D1:R3), fork-RNA (R1:R3) and fork-DNA (D1:D3), (see Figure 3.4). **(B)** Comparison of nucleic acid binding activity of hRecQ4-UpZnk towards various substrates. One-site total function of Graphpad-Prism was used to fit the data points.

(C): EMSA gels showing the nucleic acid binding activity of two concentrations of hRecQ4-Hel (0.05-0.1 μM), without/with the presence of an equimolar amount of hRecQ4-UpZnK, towards R-loop, D-loop, Holliday Junction and hybrid fork-RNA/DNA; a lane corresponds to the hRecQ4-UpZnK alone, as a comparison. **(D)** Bar graph showing the quantification of the EMSA shown above. Each experiment was done in triplicate.

3.5. The C-terminal domain of hRecQ4: structural characterization by circular dichroism

Circular Dichroism (CD) is an absorption spectroscopy method which uses circularly polarized light to investigate structural aspects of optically active molecules, based on the differential absorption of left and right circularly polarized light. UV CD is used to determine aspects and characteristics of protein secondary structure (Johnson W.C., 1990; Woody R., 1995).

There are characteristic UV CD spectra for all-α-helix, all-β-sheet and a random coil protein (Figure 3.6). The spectrum for an all- α-helical protein has two negative bands at 222 and 208 nm, and a positive band at ~ 190 nm. The spectrum for an all-β-sheet protein has, in general, a negative band between 210 - 220 nm and a positive band between 195 - 200 nm. Spectra for β -sheet proteins are more diverse than those for α-helical proteins because they may be

present at parallel, anti-parallel, or mixed conformations, and can be twisted in many ways. The spectrum for a disorderly random coil protein has a negative band at around 200 nm (Corrêa & Ramos, 2008).
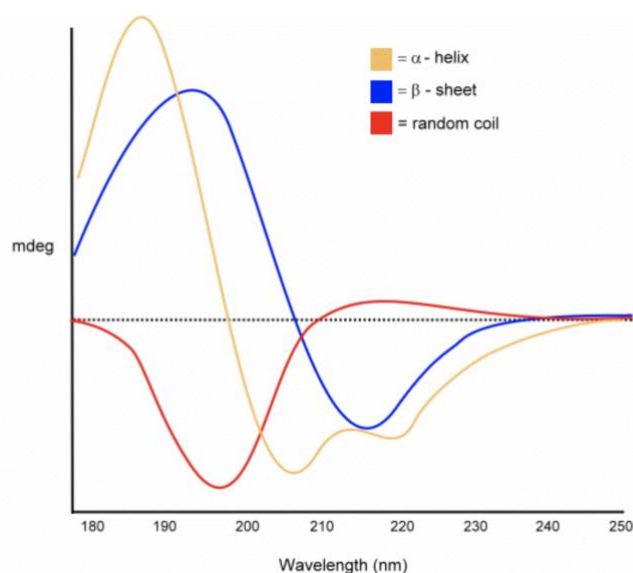

Figure 3.6. Typical CD-spectra for each secondary structure element. The three basic secondary structures of a polypeptide chain (helix, sheet, coil) show a characteristic CD spectrum. A protein consisting of these elements should therefore display a spectrum that can be deconvoluted into the three individual contributions.

The Circular Dichroism spectrum of hRecQ4-Ct shows two negative peaks of similar magnitude at 222 and 208 nm and a positive band at ~ 190 nm, suggesting that the protein is folded and is mainly composed of alpha-helical structures (Figure 3.7 A, B). DICHROWEB, an online server for protein secondary structure analysis from circular dichroism spectroscopic data, was used to analyze the spectrum and to obtain more relevant information about the secondary structure of the C-terminal domain. The primary function of the DICHROWEB server is to provide a user-friendly interface and calculation platform for a range of popular secondary structure calculation algorithms and reference databases, thereby facilitating the analysis of CD spectroscopic data. It currently supports five popular and freely available analysis algorithms: SELCON3, CONTINLL, CDSSTR, VARSLC and K2d (Whitmore & Wallace, 2004). According to the secondary structure prediction, the spectrum analysis showed the presence of five helices in the region of interest (Figure 3.7 A, B).
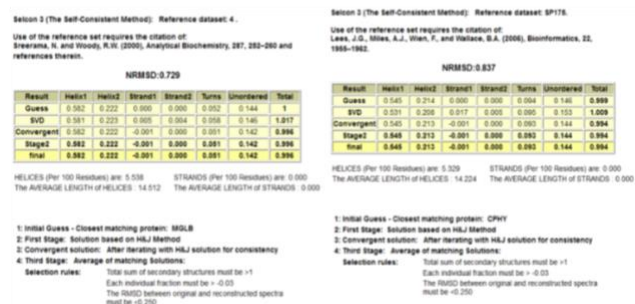
The CD analysis of the recombinant hRecQ4-Ct domain thus confirms the secondary structure prediction and is consistent with our hypothesis that this domain may fold as an HRDC domain, as in other RecQ helicases.
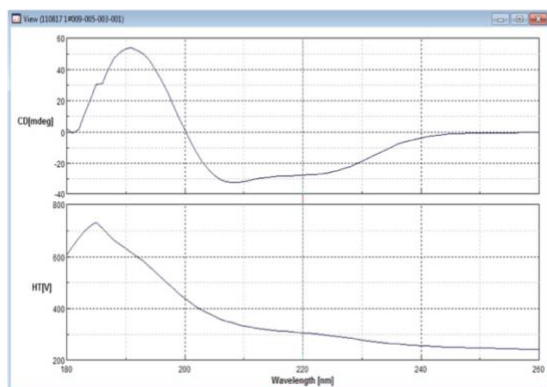
Figure 3.7. Circular dichroism of Ct domain of hRecQ4. (A): Circular Dichroism spectrum of 5 µM Ct domain shows a typical alpha-helical spectrum, suggesting a well folded protein. **(B):** Circular Dichroism spectrum analysis (Dichroweb) shows the presence of a helical bundle (approximate 5 helices of 14 residues length).

3.6. The C-terminal domain of hRecQ4: biochemical characterization

In order to verify the possible role of the Ct domain in binding nucleic acid, EMSA was performed with 10 nM of fluorescently labelled DNA and RNA substrates (Figure 3.4 A) and an excess of hRecQ4-Ct, but the protein showed no affinity for nucleic acid even at very high protein concentration (5 µM), as shown in Figure 3.8 A below.

Moreover, to assess whether the C-terminal domain may have a synergistic effect when added in trans to the catalytic domain, EMSA were then carried out with increasing concentrations (0-0.2 µM) of hRecQ4-Hel with and without an equimolar amount of hRecQ4-Ct. However, no significant effect was observed in the presence of the C-terminal domain (Figure 3.8).

In our hands thus the C-terminal region does not play a noticeable role in nucleic acid binding. However, it has to be stressed that in our experiment the C-terminal domain was added in trans; it is possible that a different effect can be obtained when the whole protein is used. Indeed, it is reported in literature that the fragment hRecQ4[427–1208] (which includes the C-terminal) binds tighter to ssDNA, compared to the shorter hRecQ4[427–1116] (comprising only the catalytic core), indicating that the last 92 aa of RecQ4 contribute significantly to DNA binding (Kaiser et al., 2017).
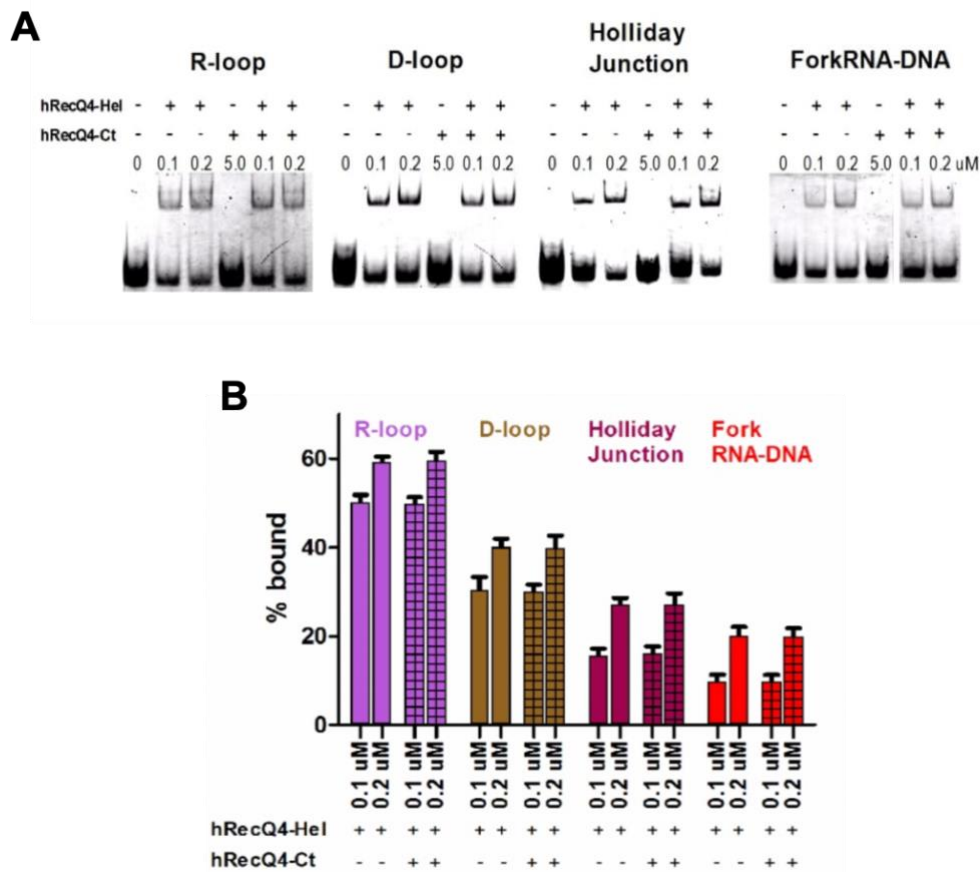
Figure 3.8. The C-terminal region does not exhibit any nucleic acid binding when added in trans to the catalytic core. (A): EMSA gels showing the nucleic acid binding activity of increasing concentration (0-0.2 μM) of hRecQ4-Hel without/with the presence of an equimolar amount of hRecQ4-Ct towards R-loop (R4:D10:D11), D-loop (D4:D10:D11), Holliday Junction (D4:D6:D7:D8) and hybrid fork-RNA/DNA (R1:D3). **(B):** Bar graph showing the quantification of the EMSA described above. The C-terminal region does not show any visible nucleic acid binding activity. Each experiment was done in triplicate.

3.7. The role of the N- and C-terminal region of RecQ4 in R-loops unwinding and resolution

To further dissect the role of the N-terminal and C-terminal domains of RecQ4, a FRET based helicase assay was performed using increasing concentration (0-160 nM) of all recombinant proteins (hRecQ4-Hel, hRecQ4-UpZnk and hRecQ4-Ct).

Two main constraints to be satisfied for the correct execution of this biochemical assay are: the distance between the donor and acceptor, which should be between 10 A° to 100 A°; the absorption range of the acceptor must overlap the emission range of the donor. Some fluorophores, such as TET, HEX, and FAM, with an emission range between 500 nm to 550 nm, are quenched by quenchers, such as Black hole quencher 1 (BHQ1) and dabcyl, with an absorption range of 450 nm to 550 nm (Mojumdar & Deka, 2019).

As already mentioned, the double-stranded nucleic acid substrate is composed of two complementary strands, labeled with a fluorophore and a quencher respectively, and hybridized to each other to form a duplex substrate. This substrate is then used for an unwinding experiment: as the fluorophore-labeled strand is separated from the quencher, an increase in the fluorescent emission intensity is observed. This increased fluorescence is measured using a spectrophotometer and the helicase activity of the enzyme is estimated in terms of percentage of unwinding, using a total unwound substrate as control.
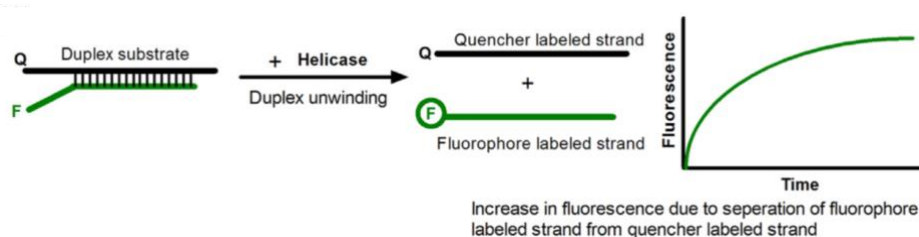
Figure 3.9. Schematic description of FRET method. The duplex substrate is prepared by hybridizing a fluorophore(F)-labeled strand (green) with quencher(Q)-labeled strand (black). Where the fluorescence from the fluorophore molecule is emitted, it is then absorbed by the quencher. As the duplex is unwound, the fluorophore-labeled strand is released, and the fluorescence signal increases with time. This increase in fluorescence signal is measured and the activity of the helicase is estimated (Figure adapted by Mojumdar & Deka, 2019).

The effect of the N-terminal region in enhancing the catalytic activity of the helicase domain was determined by adding an equimolar amount of hRecQ4-UpZnk to the helicase reaction containing hRecQ4-Hel. The addition of the N-terminal domain shows a significant increase in the unwinding activity of the helicase core towards all the substrates, especially for R-loops, as it shown in the picture below (Figure 3.10 A).
In a previous study on the N-terminal region of RecQ4 carried out in our laboratory, various residues within the zinc knuckle were mutated and then predicted and/or shown to be functionally important in nucleic acid interaction. To enhance our understanding of the activity of the N-terminal domain towards R-loops unwinding and resolution, these mutants were used in R-loops unwinding assays.
These mutants were produced as already described and were added in equimolar amount to the helicase reaction in trans with the RecQ4 helicase core: UpZnk-m1: Cys403Ala/Asn406Ala; UpZnk-m2: His411Ala/Cys416Ala; UpZnk-m3: Asn406Cys and UpZnk-m4: Phe404Ala/Trp412Ala (Marino et al., 2016).
The two mutants UpZnk-m1 and UpZnk-m2 did not show any significant change in R-loops unwinding, while UpZnk-m3 and in particular UpZnk-m4 showed a significant decrease in R-loops resolving activity of the protein, suggesting that phenylalanine and tryptophan can play a role in substrate interaction (Figure 3.10 A).

The biochemical characterization of the N-terminal mutants shows that phenylalanine (Phe404) and tryptophan (Trp412) present in the Zn knuckle are important for R-loops resolution. The two aromatic residues might interact with the nucleic acid bases by establishing π-π stacking interactions, as predicted from the NMR structure (Marino et al., 2016).

It is important to stress that we decided to target the Zn ligands (Cys403, Asn 406, His411 and Cys 416). In the human sequence one of the canonical Zn ligands (Cys) is substituted by an Asn. Mutation of these residues to alanine did not significantly affect R-loops unwinding; an attempt of reconstituting a canonical Zn knuckle, by changing Asn406 to Cys surprisingly showed a decrease in the R-loops resolving activity of the protein. Based on the NMR structure, two aromatic residues (Phe404 and Trp412) were predicted to be involved in nucleic acid binding. Indeed, their mutation to alanine strongly affected the unwinding, confirming that these residues play a role in substrate interaction (Marino et al., 2016).

Similarly, the impact of the C-terminal on unwinding activity of the helicase core was determined performing a similar unwinding assay, by adding an equimolar amount of hRecQ4-Ct to the helicase reaction containing hRecQ4-Hel. The C-terminal domain seems to significantly enhance unwinding activity of the protein towards specific substrates, with a strong preference towards R-loops, as it shown in the picture below (Figure 3.10 B).

Some conserved residues were predicted to be functionally important such as Arg1158, Arg1162, Arg1185 and Lysine 1186 and were mutated to alanine. Additionally, residues reported to be mutated in RTS patients, such as Pro1170 and Thr1200, were also mutated (Figure 3.1 A). These mutants were added to the helicase reaction with the helicase core of RecQ4. All the residues, except Arg1162, seem to affect the R-loops resolving activity of the protein, especially the RTS patient mutations, which showed a significant reduction in the activity of the protein (Figure 3.10 B).
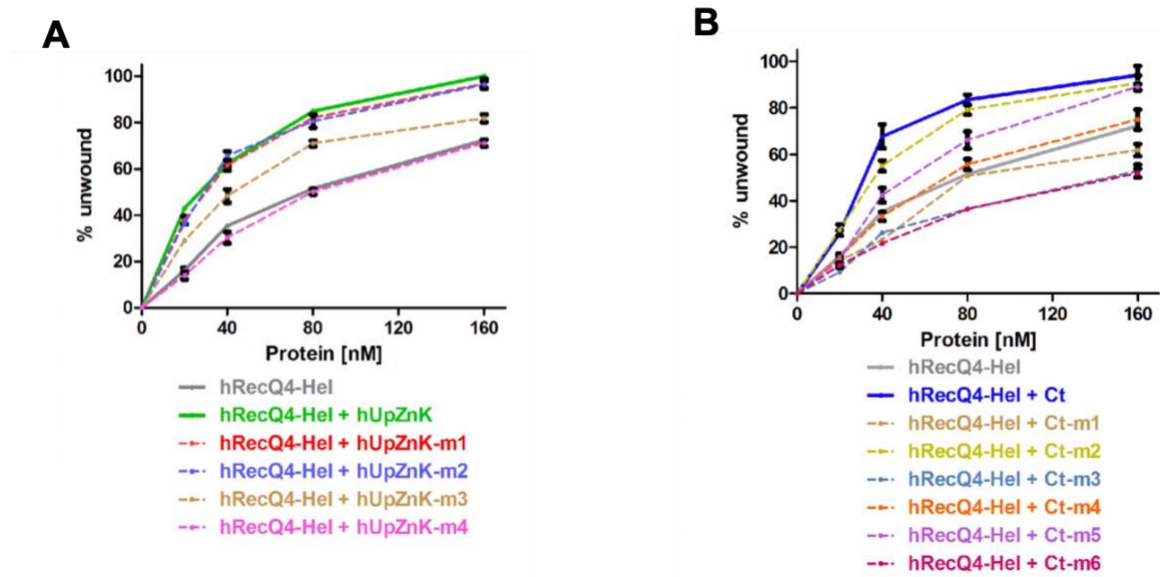
Figure 3.10. The effect of the N-terminal and C-terminal wild type and mutants on the helicase activity of the hRecQ4-Hel. (A): The effect of the N-terminal wild type and mutants: UpZnk-m1: Cys403Ala/Asn406Ala; UpZnk-m2: His411Ala/Cys416Ala; UpZnk-m3: Asn406Cys; UpZnk-m4: Phe404Ala/Trp412Ala. UpZnk-m3 and UpZnk-m4 showed a significant decrease in R-loops resolving activity of the helicase core of hRecQ4

(B): The effect of the C-terminal wild type and mutants. All the residues, except Arg1162, seem to affect the R-loops resolving activity of the helicase core of hRecQ4, especially the RTS patient mutations (Ct-m1: Arg1158Ala; Ct-m2: Arg1162Ala; Ct-m3: Pro1170Ala; Ct-m4: Arg1185Ala; Ct-m5: Lys1186Ala; Ct-m6: Thr1200Arg).

3.8. A synergic role for the N- and C-terminal region of RecQ4 in nucleic acid unwinding

Finally, an unwinding assay was carried out with all the three recombinant domains of RecQ4 (Hel, UpZnk and Ct) together in equimolar amounts (Figure 3.11). While the N-terminal domain seems to enhance unwinding of all substrates, the C-terminal domain only enhances unwinding of selected substrates (R-loops, D-loops). In a few cases there is a synergic effect between N- and C-terminal domains. When all the domains are present the best substrates are R-loops.

All these observations suggest that the N-terminal domain plays a role in improving the affinity of the protein towards all the substrates, whereas the C-terminal domain seems to play a role improving the affinity towards specific substrate, so playing an important role in improving substrate specificity to RecQ4 for R-loops structures.
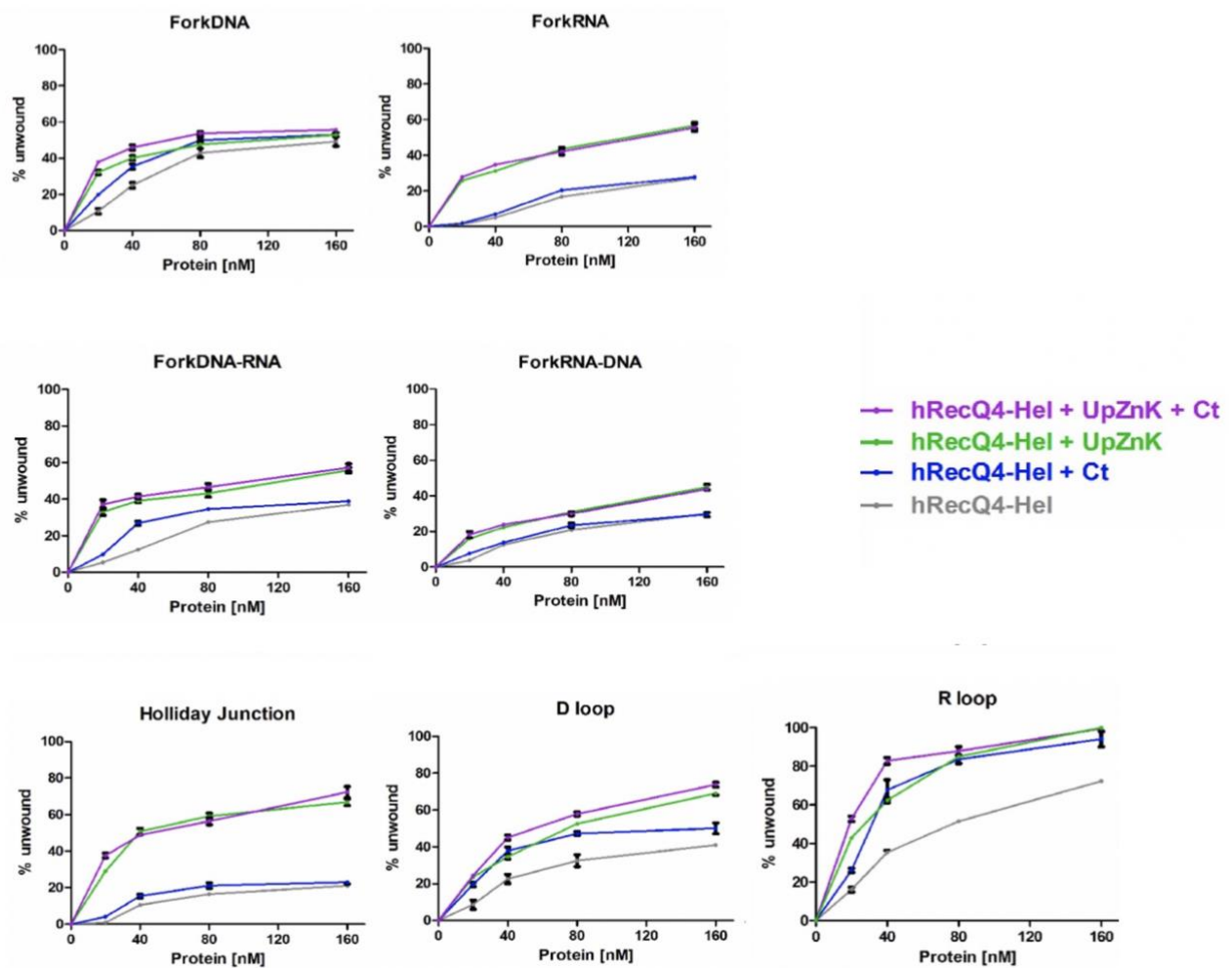
Figure 3.11. The effect of N- and C-terminal domains in nucleic acid unwinding activity of the helicase core. Comparison of helicase activity of increasing amounts (0-160 nM) of hRecQ4-Hel (catalytic core, 445-1112) without and with equimolar amount of hRecQ4-UpZnk (335-427) and/or hRecQ4-Ct (1124-1208) with various substrates (10nM) shows that the N-terminal region enhances substrate unwinding activity towards all the substrates and the C-terminal region gives substrate specificity to the protein, preferring R-loops. In a few cases there is a synergic effect between N- and C-terminal domains. When all the domains are present the best substrates are R-loops. From these data, it seems that the major players in nucleic acid binding and unwinding are these three regions, encompassing amino acids from 335 to 1208.

3.9. Possible *in vivo* significance of the RecQ4 activity towards R-loops

The RecQ4 helicase is known to play an essential role in DNA replication, as well as in DNA repair. It has been found involved in several cellular pathways such as telomere maintenance (Ghosh, 2012), double-strand break repair (Lu et al., 2016), base excision repair (Schurman et al., 2009), maintenance of mitochondrial integrity (Croteau et al., 2012), but no unifying scenario on its exact role in the cells has yet emerged.

Here, a detailed biochemical analysis of the catalytic core of human RecQ4 was carried out, with or without the other additional protein domains, using a variety of substrates (including fork-DNA, fork-RNA, hybrid DNA/RNA forks, Holliday junctions, D-loops and R-loops), to understand the role of the N- and C-terminal domains in nucleic acid interaction and in enhancing the catalytic activity of the helicase domain.
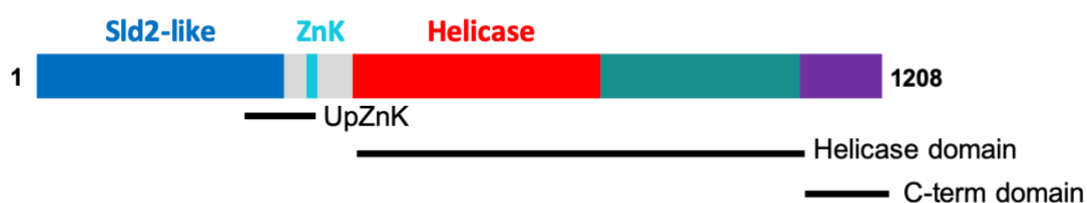
Figure 3.12. Schematic diagram of the domain organization of human RecQ4. In blue, the Sld2-like domain. In light blue, the Zn-knuckle and, highlighted in black, the region of interest (UpZnK). In red, the helicase domain. In green, the new recently characterized R4ZNBD. In purple, the C-terminal domain.

Fork-DNA does not have a high affinity for RecQ4, but it is a good substrate for its helicase activity. On the contrary, although RecQ4 has a high affinity towards fork-RNA substrates, it does not unwind them efficiently. Hybrid forks are also bound with a reasonable affinity but not unwound efficiently. D-loops, and especially R-loops seem indeed to be the preferred substrates for unwinding.

These findings, together with the fact that fork-DNA is unwound more efficiently than fork-RNA, seem to suggest that RNA is important for substrate recognition, but the protein has a preference for translocating along a DNA strand, in a 3' to 5' direction. The N-terminal domain includes a Zn knuckle that has been shown to bind both DNA and RNA, with a mild preference for the latter (Marino et al., 2016), reinforcing the idea that RecQ4 has a preference for RNA containing substrates. This suggests that RNA-containing nucleic acids such as R-loops could be important physiological substrates of RecQ4. Results from our lab also suggest that RecQ4 has the highest R-loops resolving activity among all human RecQ helicases, followed by BLM.

R-loops metabolism is associated to neurodegenerative diseases and cancer development (Sollier & Cimprich, 2015; Groh & Gromak, 2014) and may also be related to the RTS underlying pathology. Moreover, R-loops play an important role in various cellular processes, such as efficient transcription termination, epigenetic modifications and immunoglobulin class switching (Aguilera & García-Muse, 2012; Hamperl & Cimprich, 2014). Although various nucleases (RNaseH1 and RNaseH2) and helicases (Pif1, DHX9 and Sen1) have already been found implicated in R-loop removal and resolution, additional proteins and DNA repair factors may help in removing R-loops and preventing transcription-associated DNA damage (Stirling & Hieter, 2016).

R-loops have been further found implicated in the initiation of DNA replication process in bacterial cells and mitochondria and recent findings seem to propose a role of R-loops also in the initiation of DNA replication in eukaryotic cells, suggesting that they may be responsible for the initiation of origin-independent replication events in ribosomal DNA (Stuckey et al., 2015) or can contribute to replication origin specification (Lombraña et al., 2015). It is also possible that RecQ4 may play a role in R-loop resolution in the context of DNA replication initiation, both in the nucleus and in mitochondria. Considering this complex scenario, we may suggest a role for RecQ4 in the resolution of R-loops at stalled replication forks.

In this context, characterizing the metabolic role of RecQ4 in R-loops processing and resolution could have crucial implications to better understand RTS and other RecQ4 associated disorders, as well as its role in carcinogenesis and response to cancer therapies.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

The human RecQ4 helicase has a well conserved helicase domain, which conforms to the features of the RecQ family of helicases. For all the RecQ helicases, however, the catalytic core include an additional domain that is known as the RQC (RecQ C-terminal domain), comprises a Zn-binding domain and a Winged-Helix (WH) domain, and is essential for the activity (Pike et al., 2009; Lucic et al., 2011), which appeared to be absent in RecQ4. A more detailed sequence analysis (Marino et al. 2013) revealed the presence of a putative Zn-binding region and a WH domain, suggesting that RecQ4 could have a divergent RQC domain; indeed site-directed mutagenesis of key residues in this region did confirm the importance of the region for the biochemical activity (Mojumdar et al., 2017). However, the crystal structure of human RecQ4 revealed that the region following the helicase core did contain a Zn-binding and two WH domains, but their arrangement was different from a canonical RQC, and unique to RecQ4 sequences (Kaiser et al., 2017).

In addition to the catalytic core, RecQ4 possesses a N-terminal and a C-terminal extension. The first 150 amino-acid residues of RecQ4 were shown to be homologous to the N-terminus of the yeast replication factor Sld2 (Sangrithi et al., 2005). The NMR structure of the first 54 amino acids at the N-terminus of human RecQ4 is known and it forms a helical bundle resembling a homeodomain (Ohlenschläger et al., 2012). A detailed bioinformatic study, previously done in our laboratory (Marino et al., 2013), predicted a second region of homology with the C-terminal region of Sld2, encompassing amino acids 343-368 in RecQ4. Moreover, the residues between the Sld2-like domain and the helicase core comprises a cysteine-rich region classified as "retrovirus Zn finger like" or "Zn knuckle". This region was found to indeed fold as a Zn knuckle, as predicted, and to bind a variety of nucleic acid substrates, with a mild preference for RNA (Marino et al., 2016). Besides, the segment located upstream the Zn knuckle, that is highly conserved and rich in positively charged and aromatic residues, strongly enhances binding to nucleic acids in both the human and *Xenopus* proteins (Marino et al., 2016).

Although a good amount of structural information is known, not much is clear about RecQ4 biochemical aspects and, in particular, its functional role in human cells. Our group has tried to identify, by a detailed biochemical analysis, the best human RecQ4 substrate in terms of both binding and catalytic unwinding.

A detailed biochemical analysis of the catalytic core of RecQ4, using different DNA and RNA substrates, identified RecQ4 as a helicase able to efficiently resolve R-loops *in vitro*. In fact, although the protein binds a variety of nucleic acids, it has a preference for R-loops. Our collaborators have also found that R-loops are the substrate most efficiently unwound by

RecQ4 and, when compared with all other human RecQ helicases, RecQ4 exhibits the highest R-loops resolving activity. These biochemical results were then confirmed by a further analysis *in vivo*. Fibroblasts originating from a RTS patient with a RecQ4 truncated at the beginning of the helicase domain (maintaining an intact N-terminal domain, which is essential for DNA replication and cell viability) displayed a high frequency of R-loops, whereas the enforced expression of a full length RecQ4 in those same cells caused a marked decrease in the R-loops count. Furthermore, shRNA-mediated downregulation of RecQ4 also caused accumulation of R-loops in a colon carcinoma cell line. These results provided evidence for a novel functional role of RecQ4, during DNA replication, in resolution of R-loops at stalled replication forks (Mojumdar & Kenig, unpublished data).

In this work we have further analyzed the structural and functional role of the independent accessory domains of the catalytic core of the protein to understand if/how they could be involved in substrate recognition and in enhancing both binding and unwinding of different DNA and RNA substrates. We have also mutated a number of aminoacidic residues conserved throughout the RecQ family and predicted to be functionally important, or based on mutations found in patients, and tested their effect on nucleic acid interaction and processing.

4.1.1. The N-terminal domain

The analysis of the N-terminal region of human RecQ4 was focused on biochemical characterization of the region encompassing the last region of homology with the yeast Sld2 factor and the Zn knuckle (amino acid residues 335-427), with the aim to define its role in assisting the catalytic core in nucleic acid binding and unwinding, and in particular in enhancing R-loops affinity and resolution by RecQ4.

The summary of this work is as follows:

- The N-terminal region shows binding affinity towards almost all the substrates with a preference for R-loops, Holliday Junctions, D-loops and hybrid fork-RNA/DNA.

- The in trans addition of the N-terminal domain to the unwinding reaction with the catalytic core of RecQ4 shows a significant increase in the unwinding activity of the helicase core towards all the substrates, especially for R-loops.
-
- N-terminal region residues within the zinc knuckle, predicted and/or shown to be functionally important in nucleic acid interaction, were mutated and added in equimolar amount to the helicase reaction with the helicase core of RecQ4. In particular we targeted the Zn ligands (Cys403, Asn 406, His411 and Cys 416); it has to be stressed that in the human sequence one of the canonical Zn ligands (Cys) is

substituted by an Asn. Mutation of these residues to alanine did not significantly affect R-loops unwinding; an attempt of reconstituting a canonical Zn knuckle, by changing Asn406 to Cys surprisingly showed a decrease in the R-loops resolving activity of the protein. Based on the NMR structure, two aromatic residues (Phe404 and Trp412) were predicted to be involved in nucleic acid binding. Indeed, their mutation to alanine strongly affected the unwinding, confirming that these residues play a role in substrate interaction.

4.1.2. The C-terminal domain

The analysis of the C-terminal region of human RecQ4 was focused on a structural and biochemical characterization of the last 100 amino-acid residues and on defining its role in enhancing R-loops binding and unwinding by the catalytic core of the protein. An analysis of the sequence reveals the presence of 4 or 5 helices. Although no structure prediction/threading provides strong evidence for the likely 3D structure, the presence of a likely helix bundle may suggest a fold that resemble the HRDC domains that are present in many RecQ helicases. No sequence homology is detectable with other HRDC domains, but this is a fold with no clear structural determinants or high level of conservation.

The structural and biochemical studies of this fragment done in this work are summarized below:

- An expression and purification protocol of the C-terminal domain of hRecQ4, comprising the last 100 aminoacidic residues, and its mutants, was developed and the proteins were expressed and purified in a reasonable amount for structural and biochemical studies.

- Circular dichroism analysis of C-terminal domain of hRecQ4 showed a typical alpha-helical spectrum, suggesting a well folded protein. Circular Dichroism spectrum analysis by Dichroweb also showed the presence of a helical bundle (approximate 5 helices of an average length of 14 residues). These results confirmed the secondary structure predictions and are consistent with the hypothesis that this domain may fold as an HRDC domain, as in other RecQ helicases.

- The purified 6His-tagged C-terminal domain of human RecQ4 does not exhibit any nucleic acid binding activity but it seems to significantly enhance unwinding activity of the protein towards specific substrates, with a strong preference towards R-loops, when added in trans to the binding reaction with the catalytic core of RecQ4.

- Some Ct domain conserved residues, predicted to be functionally important such as Arg1158, Arg1162, Arg1185 and Lysine 1186, were mutated to alanine. Residues

reported to be mutated in RTS patients, such as Pro1170 and Thr1200, were also mutated. These mutants were added to the helicase reaction with the helicase core of RecQ4. All the residues, except Arg1162, seem to affect the R-loops resolving activity of the protein. Intriguingly, whereas the mutations of the positively charged residues simply reduce or abolish the enhancement of unwinding caused by the presence of the C-terminal domain, the presence of the two mutations derived from the patients seem to have an effect on the activity that is worse than the absence of the C-terminal domain itself.

- An unwinding assay was carried out with all the three recombinant domains of RecQ4 (Hel, UpZnk and Ct) together in equimolar amounts. While the N-terminal domain seem to enhance unwinding of all substrates, the C-terminal domain enhances unwinding of selected substrates (R-loops, D-loops). In a few cases there is a synergic effect between N- and C-terminal domains. When all the domains are present the best substrates are R-loops.

In summary, during my PhD we have further characterized accessory domains of human RecQ4 and reported the effect of their presence on binding and unwiding activity of the protein. These results enhance the knowledge about human RecQ4 and provide a useful base of a detailed study to dissect the mechanism of action of the protein in cells metabolism and R-loops resolution.

4.2. Future work

These results provided us a useful base to obtain a detailed structural and biochemical characterization of human RecQ4 and for further cellular studies to better understand RecQ4 functional role in cellular metabolism. We therefore want to continue our studies performing the following experiments:

- Using a combination of structure determination techniques such as X-ray crystallography, NMR and SAXS to further structurally characterize the protein. In particular we are aiming to visualize the whole protein using Cryo-EM, and to determine the structure of either the full length or the catalytic core (possibly together with regions of the N/C-terminus) in the presence of suitable substrates (nucleic acids such as single strands, forks, hybrid forks, G-quadruplexes, Holliday junctions; and/or nucleotides such as ADP, ATP, non-hydrolysable ATP analogues, analogues of the transition states).

- Investigating the role of each protein domain in affecting R-loops resolution *in vivo*, by cloning different protein constructs in a mammalian vector and trasfecting mammalian cells, to confirm our biochemical findings.

- Carrying out a detailed biochemical study of the annealing activity of RecQ4, to dissect the role of RecQ4 and its domains in annealing various substrates.

- Identifying the potential interaction partners of the protein. The laboratory can provide a large number of proteins involved in human DNA replication, including MCM2-7, MCM89, MCM10, Cdc45, GINS, AND-1, Cdt1. As for substrates involved in DNA repair, we will rely on collaborations with colleagues working in the field.

REFERENCES

Abe T, Yoshimura A, Hosono Y, Tada S, Seki M, Enomoto T. (2011) The N-terminal region of RECQL4 lacking the helicase domain is both essential and sufficient for the viability of vertebrate cells. Role of the N-terminal region of RECQL4 in cells. *Biochim. Biophys. Acta*, 1813, 473-479.

Aggarwal M, Sommers JA, Morris C, Brosh RM Jr. (2010) Delineation of WRN helicase function with EXO1 in the replicational stress response. *DNA Repair,* 9, 765–776.

Aggarwal M, Sommers JA, Shoemaker RH, Brosh RM., Jr. (2011). Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc. Natl. Acad. Sci. U.S.A*., 108, 1525–1530.

Aggarwal M, Banerjee T, Sommers JA, Iannascoli C, Pichierri P, Shoemaker RH, et al. (2013b). Werner syndrome helicase has a critical role in DNA damage responses in the absence of a functional fanconi anemia pathway. *Cancer Res*., 73, 5497–5507.

Aguilera A, García-Muse T. (2012) R loops: from transcription byproducts to threats to genome stability. *Mol Cell*., 46, 115-124.

Ahn B, Harrigan JA, Indig FE, Wilson DM 3rd, Bohr VA. (2004) Regulation of WRN helicase activity in human base excision repair. *J Biol Chem*., 279, 53465-53474.

Alberts B, Johnson A, Lewis J, et al. (2002) Molecular Biology of the Cell. 4th edition. New York: Garland Science.

Ammazzalorso F, Pirzio LM, Bignami M, Franchitto A, Pichierri P. (2010) ATR and ATM differently regulate WRN to prevent DSBs at stalled replication forks and promote replication fork recovery. *EMBO J.*, 29, 3156–3169.

Armas P, Aguero TH, Borgognone M, Aybar MJ, Calcaterra NB. (2008) Dissecting CNBP, a zinc-finger protein required for neural crest development, in its structural and functional domains. *Journal of Molecular Biology,* 382, 1043–1056.

Audebert M, Salles B, Calsou P. (2004) Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *J. Biol. Chem*., 279, 55117–55126.

Ausiannikava D, Allers T. (2017) Diversity of DNA Replication in the Archaea. *Genes (Basel)*, 8, 56.

Bahr A, De Graeve F, Kedinger C, Chatton B. (1998) Point mutations causing Bloom's syndrome abolish ATPase and DNA helicase activities of the BLM protein. *Oncogene,* 17, 2565-2571.

Balajee AS, Machwe A, May A, Gray MD, Oshima J, Martin GM, Nehlin JO, Brosh R, Orren DK, Bohr VA. (1999) The Werner syndrome protein is involved in RNA polymerase II transcription. *Mol Biol Cell,* 10, 2655-2668.

Balk B, Maicher A, Dees M, Klermund J, Luke-Glaser S, Bender K, Luke B (2013) Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol.,* 20, 1199–1205.

Barefield C, Karlseder J. (2012) The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures. *Nucleic Acids Res*., 40, 7358–7367.

Baynton K, Otterlei M, Bjørås M, von Kobbe C, Bohr VA, Seeberg E. (2003) WRN interacts physically and functionally with the recombination mediator protein RAD52. *J Biol Chem*., 278, 36476-36486.

Beamish H, Kedar P, Kaneko H, Chen P, Fukao T, Peng C, Beresten S, Gueven N, Purdie D, Lees-Miller S, et al. (2002) Functional link between BLM defective in Bloom's syndrome and the ataxia-telangiectasia-mutated protein, ATM. *The Journal of boil. chem*., 277, 30515-30523.

Behnfeldt JH, Acharya S, Tangeman L, Gocha AS, Keirsey J, Groden J. (2018). A tri-serine cluster within the topoisomerase IIalpha-interaction domain of the BLM helicase is required for regulating chromosome breakage in human cells. Hum. *Mol. Genet.,* 27, 1241–1251.

Bennett RJ, Sharp JA, Wang JC. (1998) Purification and characterization of the Sgs1 DNA helicase activity of Saccharomyces cerevisiae. *J Biol Chem.,* 273, 9644-9650.

Bennett RJ, Keck JL. (2004) Structure and function of RecQ DNA helicases. *Crit Rev Biochem Mol Biol*., 39, 79-97.

Berg JM, Shi Y. (1996) The galvanization of biology: a growing appreciation for the roles of zinc. *Science,* 271, 1081-1085.

Berget SM, Moore C, Sharp PA. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74, 3171–3175.

Bernstein DA, Keck JL. (2003) Domain mapping of Escherichia coli RecQ defines the roles of conserved N- and C-terminal regions in the RecQ family. *Nucleic Acids Res*., 31, 2778-2785.

Bernstein DA, Keck JL. (2005) Conferring substrate specificity to DNA helicases: role of the RecQ HRDC domain. *Structure*, 13, 1173-1182.

Beyer DC, Ghoneim MK, Spies M. (2012) Structure and Mechanisms of SF2 DNA Helicases. *DNA Helicases and DNA Motor Proteins*, 47–73.

Bhatia V, Barroso SI, García-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. (2014) BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, 511, 362-365.

Blundred R, Myers K, Helleday T, Goldman AS, Bryant HE. (2010) Human RECQL5 overcomes thymidine-induced replication stress. *DNA Repair (Amst)*, 9, 964-975.

Boboila C, Alt FW, Schwer B. (2012) Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. *Adv Immunol*., 116, 1-49.

Bochkarev A, Bochkareva E, Frappier L, Edwards AM. (1999) The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J*., 18, 4498–4504.

Bohr VA. (2008) Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. *Trends Biochem Sci*., 33, 609-620.

Bothmer A, Robbiani DF, Feldhahn N, Gazumyan A, Nussenzweig A, Nussenzweig MC. (2010) 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med*., 207, 855–865.

Brosh RM Jr, Majumdar A, Desai S, Hickson ID, Bohr VA, Seidman MM. (2001) Unwinding of a DNA triple helix by the Werner and Bloom syndrome helicases. *J Biol Chem*., 276, 3024-3030.

Brosh RM Jr. (2013) DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer*, 13, 542-558.

Buckman JS, Bosche WJ, Gorelick RJ. (2003) Human immunodeficiency virus type 1 nucleocapsid zn(2+) fingers are required for efficient reverse transcription, initial integration processes, and protection of newly synthesized viral DNA. *Journal of Virology,* 77, 1469–1480.

Bugreev DV, Yu X, Egelman EH, Mazin AV. (2007) Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev*., 21, 3085–3094.

Bunting SF, Callen E, Wong N, Chen HT, Polato F, et al. (2010) 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell,* 141, 243–254.

Buttner K, Nehring S, Hopfner KP. (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol*., 14, 647-652.

Calado R, Young N. (2012) Telomeres in disease. *F1000 Med. Rep*., 4, 8.

Capp C, Wu J, Hsieh TS. (2009) Drosophila RecQ4 has a 3'-5' DNA helicase activity that is essential for viability. *J Biol Chem*., 284, 30845-30852.

Cerritelli SM, Crouch RJ. (2009) Ribonuclease H: the enzymes in eukaryotes. *FEBS J*., 276, 1494-1505.

Chakraborty P, Grosse F. (2010) WRN helicase unwinds Okazaki fragment-like hybrids in a reaction stimulated by the human DHX9 helicase. *Nucleic Acids Res*., 38, 4722-4730.

Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. (2014) Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet*., 10, e1004288.

Chang EY, Novoa CA, Aristizabal MJ, Coulombe Y, Segovia R, Chaturvedi R, Shen Y, Keong C, Tam AS, Jones SJM, Masson JY, Kobor MS, Stirling PC. (2017) RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. *J Cell Biol*., 216, 3991-4005.

Chen S, Sayana P, Zhang X, Le W. (2013) Genetics of amyotrophic lateral sclerosis: an update. *Mol. Neurodegener*., 8, 28.

Cheng WH, von Kobbe C, Opresko PL, Arthur LM, Komatsu K, Seidman MM, Carney JP, Bohr VA. (2004) Linkage between Werner syndrome protein and the Mre11 complex via Nbs1. *J Biol Chem*., 279, 21169-21176.

Cheng WH, Sakamoto S, Fox JT, Komatsu K, Carney J, Bohr VA. (2005) Werner syndrome protein associates with γH2AX in a manner that depends upon Nbs1. *FEBS Lett*., 579, 1350–1356.

Cheng WH, Kusumoto R, Opresko PL, Sui X, Huang S, Nicolette ML, Paull TT, Campisi J, Seidman M, Bohr VA (2006) Collaboration of Werner syndrome protein and BRCA1 in cellular responses to DNA interstrand cross-links. *Nucleic Acids Res*., 34, 2751-2760.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.*, 31, 3497-3500.

Chow LT, Gelinas RE, Broker TR, Roberts RJ. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12, 1–8.

Chu WK, Hickson ID. (2009) RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer*., 9, 644-654.

Ciccia A, Elledge SJ. (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell.,* 40, 179–204.

Cleasby A, Wonacott A, Skarzynski T, Hubbard RE, Davies GJ, Proudfoot AE, Bernard AR, Payton MA, Wells TN. (1996) The X-ray crystal structure of phosphomannose isomerase from Candida albicans at 1.7 angstrom resolution. *Nature Structural Biology,* 3, 470–479.

Collins R, Karlberg T, Lehtiö L, Schütz P, van den Berg S, Dahlgren LG, Hammarström M, Weigelt J, Schüler H. (2009) The DEXD/H-box RNA helicase DDX19 is regulated by an {alpha}-helical switch. *J Biol Chem*., 284, 10296-10300.

Constantinou A, Tarsounas M, Karow J.K, Brosh R.M, Bohr V.A., Hickson I.D., West S.C. (2000) Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest, *EMBO* Rep, 1, 80-84.

Cooper MP, Machwe A, Orren DK, Brosh RM, Ramsden D, Bohr VA. (2000) Ku complex interacts with and stimulates the Werner protein. *Genes Dev*., 14, 907–912.

Crabbe L, Jauch A, Naeger CM, Holtgreve-Grez H, Karlseder J.(2007) Telomere dysfunction as a cause of genomic instability in Werner syndrome. *Proc Natl Acad Sci U S A*., 104, 2205-2210.

Croteau DL, Rossi ML, Canugovi C, Tian J, Sykora P, Ramamoorthy M, Wang Z, Singh DK, Akbari M, Kasiviswanathan R, Copeland WC, Bohr VA. (2012) RECQL4 localizes to mitochondria and preserves mitochondrial DNA integrity. *Aging Cell*., 11, 456-466.

Croteau DL, Popuri, V, Opresko, PL, Bohr VA. (2014) Human RecQ Helicases in DNA Repair, Recombination, and Replication. *Annu. Rev. Biochem*., 83, 519-552.

Crow YJ, Rehwinkel J. (2009) Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum Mol Genet*., 18, 130-136.

Das A, Boldogh I, Lee JW, Harrigan JA, Hegde ML, Piotrowski J, de Souza Pinto N, Ramos W, Greenberg MM, Hazra TK, Mitra S, Bohr VA. (2007) The human Werner syndrome protein stimulates repair of oxidative DNA base damage by the DNA glycosylase NEIL1. *J Biol Chem*., 282, 26591-26602.

De S, Kumari J, Mudgal R, Modi P, Gupta S, Futami K, Goto H, Lindor NM, Furuichi Y, Mohanty D, Sengupta S. (2012) RECQL4 is essential for the transport of p53 to mitochondria in normal human cells in the absence of exogenous stress. *J Cell Sci*., 125, 2509-2522.

De Amicis A, Piane M, Ferrari F, Fanciulli M, Delia D, Chessa L. (2011) Role of senataxin in DNA damage and telomeric stability. *DNA Repair (Amst).*, 10, 199-209.

Dietschy T, Shevelev I, Pena-Diaz J, Huhn D, Kuenzle S, Mak R, Miah MF, Hess D, Fey M, Hottiger MO, et al. (2009) p300-mediated acetylation of the Rothmund-Thomson- syndrome gene product RECQL4 regulates its subcellular localization. *Journal of cell science*. 122, 1258-1267.

Diffley JFX. (2011) Quality control in the initiation of eukaryotic DNA replication. *Philos Trans R Soc Lond B Biol Sci*., 366, 3545-3553.

Dinkelmann M, Spehalski E, Stoneham T, Buis J, Wu Y, et al. (2009) Multiple functions of MRN in end-joining pathways during isotype class switching. *Nat. Struct. Mol. Biol*., 16, 808–813.

Doherty KM, Sharma S, Uzdilla LA, Wilson TM, Cui S, et al. (2005) RECQ1 helicase interacts with human mismatch repair factors that regulate genetic recombination. *J. Biol. Chem*., 280, 28085–28094.

Drosopoulos WC, Kosiyatrakul ST, Schildkraut CL. (2015) BLM helicase facilitates telomere replication during leading strand synthesis of telomeres. *J Cell Biol*., 210, 191-208.

Dutertre S, Ababou M, Onclercq R, Delic J, Chatton B, Jaulin C, Amor-Gueret M. (2000) Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase. *Oncogene,* 19, 2731-2738.

Eladad S, Ye TZ, Hu P, Leversha M, Beresten S, Matunis MJ, Ellis NA. (2005) Intra- nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. *Human molecular genetics,* 14, 1351-1365.

Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, Proytcheva M., German J. (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell,* 83, 655-666.

Epstein CJ, Martin GM, et al. (1966) Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. *Medicine* (Baltimore), 45, 177-221.

Fairman-Williams ME, Guenther UP, Jankowsky E. (2010) SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol.,* 20, 313-324.

Fan W, Luo J. (2008) RecQ4 facilitates UV light-induced DNA damage repair through interaction with nucleotide excision repair factor xeroderma pigmentosum group A (XPA). *J Biol Chem*., 283, 29037-29044.

Fattah F, Lee EH, Weisensel N, Wang Y, Lichter N, Hendrickson EA. (2010) Ku regulates the nonhomologous end joining pathway choice of DNA double-strand break repair in human somatic cells. *PLoS Genet*. 6, e1000855.

Ferrarelli LK, Popuri V, Ghosh AK, Tadokoro T, Canugovi C, Hsu JK, Croteau DL, Bohr VA. (2013) The RECQL4 protein, deficient in Rothmund-Thomson syndrome is active on telomeric D-loops containing DNA metabolism blocking lesions. *DNA Repair (Amst)*., 12, 518-528.

Fields S, Ternyak K, Gao H, Ostraat R, Akerlund J, Hagman J. (2008) The 'zinc knuckle' motif of Early B cell factor is required for transcriptional activation of B cell-specific genes. *Molecular Immunology*, 45, 3786–3796.

Foster JW, Slonczewski JL. (2008) Microbiology: an Evolving Science. 1st ed.

Futami K, Ishikawa Y, Goto M, Furuichi Y, Sugimoto M. (2008a) Role of Werner syndrome gene product helicase in carcinogenesis and in resistance to genotoxins by cancer cells. *Cancer Sci,* 99, 843-848.

Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. (1994) The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol.,* 14, 8391-8398.

Garcia PL, Liu Y, Jiricny J, West SC, Janscak P. (2004) Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide, *EMBO J.*, 21, 2882-2891.

German J, et al. (2007) Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Hum Mutat.,* 28, 743-753.

Ghosh AK, Rossi ML, Singh DK, Dunn C, Ramamoorthy M, Croteau DL, Liu Y, Bohr VA. (2012) RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions in telomere maintenance. *J Biol Chem*., 287, 196-209.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*., 45, 814-825.

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res*., 17, 4713-4730.

Gorbalenya AE, Koonin EV. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol.,* 3, 419-429.

Goto M. (1997) Hierarchical deterioration of body systems in Werner's syndrome: implications for normal ageing. *Mech Ageing Dev.,* 98, 239-254.

Gottipati P, Vischioni B, Schultz N, Solomons J, Bryant HE, Djureinovic T, Issaeva N, Sleeth K, Sharma RA, Helleday T. (2010) Poly(ADP-ribose) polymerase is hyperactivated in homologous recombination-defective cells. *Cancer Res.,* 70, 5389-5398.

Grandori C, Wu KJ, Fernandez P, Ngouenet C, Grim J, Clurman BE, Moser MJ, Oshima J, Russell DW, Swisshelm K, Frank S, Amati B, Dalla-Favera R, Monnat RJ Jr. (2003) Werner syndrome protein limits MYC-induced cellular senescence. *Genes Dev*., 17, 1569-1574.

Gravel S, Chapman JR, Magill C, Jackson SP. (2008) DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev*., 22, 2767–2772.

Graves-Woodward KL, Weller SK. (1996) Replacement of gly815 in helicase motif V alters the single-stranded DNA-dependent ATPase activity of the herpes simplex virus type 1 helicase-primase. J Biol Chem., 271, 13629-13635.

Griffith JD, Comeau L, et al. (1999) Mammalian telomeres end in a large duplex loop. *Cell,* 97, 503-514.

Groh M, Gromak N. (2014) Out of balance: R-loops in human disease. *PLoS Genet*., 10, e1004630.

88

Groh M, Lufino MM, Wade-Martins R, Gromak N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet*., 10, e1004318.

Hall MC, Matson SW. (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol.*, 34, 867-877.

Hamperl S, Cimprich KA. (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)*., 19, 84-94.

Harmon FG, Kowalczykowski SC. (1998) RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination. *Genes Dev.,* 12, 1134-1144.

Harmon FG, Kowalczykowski SC. (2000) Coupling of DNA helicase function to DNA strand exchange activity. *Methods Mol Biol*., 152, 75-89.

Harmon FG, Kowalczykowski SC. (2001) Biochemical characterization of the DNA helicase activity of the escherichia coli RecQ helicase. *J Biol Chem*., 276, 232-243.

Harrigan JA, Opresko PL, von Kobbe C, Kedar PS, Prasad R, Wilson SH, Bohr VA. (2003) The Werner syndrome protein stimulates DNA polymerase beta strand displacement synthesis via its helicase activity. *J Biol Chem*., 278, 22686-22695.

Harrigan JA, Wilson DM 3rd, Prasad R, Opresko PL, Beck G, May A, Wilson SH, Bohr VA. (2006) The Werner syndrome protein operates in base excision repair and cooperates with DNA polymerase beta. *Nucleic Acids Res*., 34, 745-754.

Heringa J. (2000) Computational methods for protein secondary structure prediction using multiple sequence alignments. *Curr Protein Pept Sci*., 1, 273-301.

Hickson ID. (2003) RecQ helicases: caretakers of the genome. *Nat Rev Cancer*., 3, 169-178.

Hu Y, Raynard S, Sehorn MG, Lu X, Bussen W, Zheng L, Stark JM, Barnes EL, Chi P, Janscak P, Jasin M, Vogel H, Sung P, Luo G. (2007) RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev*., 21, 3073-3084.

Huber MD, Duquette ML, Shiels JC, Maizels N. (2006) A conserved G4 DNA binding domain in RecQ family helicases. *J Mol Biol.,* 358, 1071-1080.

Huertas P, Aguilera A. (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell*., 12, 711-721.

Hynes NE, Stoelzle T. (2009) Key signalling nodes in mammary gland development and cancer: Myc. *Breast Cancer Res*., 11, 210.

Ichikawa K, Noda T, Furuichi, Y. (2002) Preparation of the gene targeted knockout mice for human premature aging diseases, Werner syndrome, and Rothmund-Thomson syndrome caused by the mutation of DNA helicases. *Nippon yakurigaku zasshi*, 119, 219-226.

Im JS, Ki SH, Farina A, Jung DS, Hurwitz J, Lee JK. (2009) Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. *Proc Natl Acad Sci U S A*., 106, 15628-15632.

Islam MN, Fox D 3rd, Guo R, Enomoto T, Wang W. (2010) RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. *Mol Cell Biol*., 30, 2460-2472.

Jackson BR, Noerenberg M, Whitehouse A. (2014) A novel mechanism inducing genome instability in Kaposi's sarcoma-associated herpesvirus infected cells. *PLoS Pathog*., 10, e1004098.

Jankowsky E, Bowers H. (2006) Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. *Nucleic Acids Res*., 34, 4181-4188.

Jankowsky E, Fairman ME. (2007) RNA helicases--one fold for many functions. *Curr Opin Struct Biol.,* 17, 316-324.

Kääriäinen H, Ryöppy S, Norio R. (1989) RAPADILINO syndrome with radial and patellar aplasia/hypoplasia as main manifestations. *Am J Med Genet*., 33, 346-351.

Kaiser S, Sauer F, Kisker C. (2017) The structural and functional characterization of human RecQ4 reveals insights into its helicase mechanism. *Nat Commun*., 27, 15907.

Kamimura Y, Masumoto H, Sugino A, Araki H (1998) Sld2, which interacts with Dpb11 in Saccharomyces cerevisiae, is required for chromosomal DNA replication. *Mol Cell Biol.,* 10, 6102–6109.

Kanagaraj R, Saydam N, Garcia PL, Zheng L, Janscak P. (2006) Human RECQ5beta helicase promotes strand exchange on synthetic DNA structures resembling a stalled replication fork. *Nucleic Acids Res*., 34, 5217-5231.

Kanagaraj R, Huehn D, MacKellar A, Menigatti M, Zheng L, Urban V, Shevelev I, Greenleaf AL, Janscak P. (2010) RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. *Nucleic Acids Res.*, 38, 8131-8140.

Karmakar P, Snowden CM, Ramsden DA, Bohr VA. (2002) Ku heterodimer binds to both ends of the Werner protein and functional interaction occurs at the Werner Nterminus. *Nucleic Acids Res.*, 30, 3583-3591.

Karow JK, Newman RH, Freemont PS, Hickson ID. (1999) Oligomeric ring structure of the Bloom's syndrome helicase. *Curr Biol.;* 9, 597-600.

Karow JK, Constantinou A, Li JL, West SC, Hickson ID. (2000) The Bloom's syndrome gene product promotes branch migration of holliday junctions, *Proc Natl Acad Sci U S A.*, 97, 6504-6508.

Kawabe Y, Seki M, Seki T, Wang WS, Imamura O, Furuichi Y, Saitoh H, Enomoto T. (2000) Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1.
J Biol Chem., 275, 20963-20966.

Khadka P, Croteau DL, Bohr VA. (2016) RECQL5 has unique strand annealing properties relative to the other human RecQ helicase proteins. *DNA Repair,* 37, 53-66.

Kim YM, Choi BS. (2010) Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. *Nucleic Acids Res.*, 38, 7764-7777.

Kitano K, Yoshihara N, Hakoshima T. (2007) Crystal structure of the HRDC domain of human Werner syndrome protein, WRN. *J Biol Chem.*, 282, 2717-2728.

Kitano K, Kim SY, Hakoshima T. (2010) Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure,* 18, 177-187.

Kitano K. (2014) Structural mechanisms of human RecQ helicases WRN and BLM. *Front Genet*. 5, 366.

Kitao S, Ohsugi I, Ichikawa K, Goto M, Furuichi Y, Shimamoto A. (1998) Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes. *Genomics.*, 54, 443-452.

Kohzaki M, Chiourea M, Versini G, Adachi N, Takeda S, et al. (2012) The helicase domain and Cterminus of human RecQL4 facilitate replication elongation on DNA templates damaged by ionizing radiation. *Carcinogenesis*, 33, 1203–1210.

Koonin EV (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.*, 21, 2541–2547.

Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G. (1997) Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell.*, 90, 635-647.

Koutmos M, Pejchal R, Bomer TM, Matthews RG, Smith JL, Ludwig ML. (2008) Metal active site elasticity linked to activation of homocysteine in methionine synthases. *Proceedings of the National Academy of Sciences of the United States of America,* 105, 3286–3291.

Kunkel TA, Burgers PM. (2008) Dividing the workload at a eukaryotic replication fork. *Trends Cell Biol.*, 18, 521-527.

Kusumoto R, Dawut L, Marchetti C, Wan Lee J, Vindigni A, et al. (2008) Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry*, 47, 7548–7556.

Kusumoto-Matsuo R, Opresko PL, Ramsden D, Tahara H, Bohr VA. (2010) Cooperation of DNA-PKcs and WRN helicase in the maintenance of telomeric D-loops. *Aging*, 2, 274–284.

Larizza L, Roversi G, Volpi L. (2010) Rothmund-Thomson syndrome. *Orphanet J Rare Dis.*, 29; 2.

Larsen NB, Hickson ID. (2013) RecQ Helicases: Conserved Guardians of Genomic Integrity. *Adv Exp Med Biol.*, 767, 161-84.

Lee JY, Yang W. (2006) UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell*, 127, 1349-1360.

Lee-Kirsch MA, Wolf C, Günther C. (2014) Aicardi-Goutières syndrome: a model disease for systemic autoimmunity. *Clin Exp Immunol.*, 175, 17-24.

Levitt NC, Hickson ID. (2002) Caretaker tumour suppressor genes that defend genome integrity. *Trends Mol Med.*, 8, 179-186.

Li B, Comai L. (2000) Functional interaction between Ku and the Werner syndrome protein in DNA end processing. *J. Biol. Chem*., 275, 28349–28352.

Li X, Manley JL. (2005) Inactivation of the SR Protein Splicing Factor ASF/SF2 Results in Genomic Instability. *Cell.*, 122, 365–378.

Liu Z, Macias MJ, Bottomley MJ, Stier G, Linge JP, Nilges M, Bork P, Sattler M. (1999) The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins. *Structure,* 7, 1557-1566.

Lohman TM, Bjornson KP. (1996) Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.,* 165, 169–214.

Lohman TM, Tomko EJ, Wu CG. (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol*., 9, 391-401.

Lombraña R, Almeida R, Álvarez A, Gómez M1. (2015) R-loops and initiation of DNA replication in human cells: a missing link? *Front Genet*., 28, 158.

Loughlin FE, Gebert LF, Towbin H, Brunschweiger A, Hall J, Allain FH. (2011) Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nature Structural and Molecular Biology,* 19, 84–89.

Lu H, Fang EF, Sykora P, Kulikowicz T, Zhang Y, Becker KG, Croteau DL, Bohr VA. (2014) Senescence induced by RECQL4 dysfunction contributes to Rothmund-Thomson syndrome features in mice. *Cell Death*., 5, e1226.

Lu H, Shamanna RA, Keijzers G, Anand R, Rasmussen LJ, Cejka P, Croteau DL, Bohr VA. (2016) RECQL4 Promotes DNA End Resection in Repair of DNA Double-Strand Breaks. *Cell Rep*., 16, 161-173.

Lu L, Harutyunyan K, Jin W, Wu J, Yang T, Chen Y, Joeng KS, Bae Y, Tao J, Dawson BC, Jiang MM, Lee B, Wang LL. (2015) RECQL4 Regulates p53 Function in vivo During Skeletogenesis. *J Bone Miner Res*., 30, 1077-1089.

Lu L, Jin W, Wang LL. (2017) Aging in Rothmund-Thomson syndrome and related RECQL4 genetic disorders. *Ageing Res Rev*., 33, 30-35.

Lucic B, Zhang Y, King O, Mendoza-Maldonado R, Berti M, Niesen FH, Burgess-Brown NA, Pike AC, Cooper CD, Gileadi O, Vindigni A. (2011) A prominent β-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation. *Nucleic Acids Res*., 39, 1703-1717.

Luo D, Xu T, Watson RP, Scherer-Becker D, Sampath A, Jahnke W, Yeong SS, Wang CH, Lim SP, Strongin A, Vasudevan SG, Lescar J. (2008) Insights into RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J.*, 27, 3209-3219.

Machwe A, Xiao L, Theodore S, Orren DK. (2002) DNase I footprinting and enhanced exonuclease function of the bipartite Werner syndrome protein (WRN) bound to partially melted duplex DNA. *J Biol Chem.*, 277, 4492-4504.

Machwe A, Xiao L, Groden J, Matson SW, Orren DK. (2005) RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange. *J Biol Chem.*, 280, 23397-23407.

Macris MA, Krejci L, Bussen W, Shimamoto A, Sung P. (2006) Biochemical characterization of the RECQ4 protein, mutated in Rothmund-Thomson syndrome. *DNA Repair (Amst).*, 5, 172-180.

Manosas M, Xi XG, Bensimon D, Croquette V. (2010) Active and passive mechanisms of helicases. *Nucleic Acids Res.*, 38, 5518-5526.

Marino F, Vindigni A, Onesti S. (2013) Bioinformatic analysis of RecQ4 helicases reveals the presence of a RQC domain and a Zn knuckle. *Biophys Chem.*, 177-178, 34-39.

Marino F, Mojumdar A, Zucchelli C, Bhardwaj A, Buratti E, Vindigni A, Musco G, Onesti S. (2016) Structural and biochemical characterization of an RNA/DNA binding motif in the N-terminal domain of RecQ4 helicases. *Sci Rep.,* 18, 21501.

Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem.*, 79, 89-130.

Matson SW, Bean DW, George JW. (1994) DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays.*, 16, 13-22.

Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa, H. (2006) The Nterminal noncatalytic region of Xenopus RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Molecular and cellular biology*, 26, 4843-4852.

McGuffin LJ, Bryson K, Jones DT. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, 16, 404-405.

Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell.*, 41, 21-32.

Mohaghegh P, Karow JK, Brosh RM Jr, Bohr VA, Hickson ID. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res.,* 29, 2843-2849.

Mojumdar A, De March M, Marino F, Onesti S. (2017) The Human RecQ4 Helicase Contains a Functional RecQ C-terminal Region (RQC) That Is Essential for Activity. *J Biol Chem.*, 292, 4176-4184.

Monnat RJJ, Sidorova J. (2014) Human RECQ helicases: roles in cancer, aging, and inherited disease. *Advances in Genomics and Genetics,* 5, 19-33.

Moore JK, Haber JE. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Mol Cell Biol.*, 16, 2164-2173.

Moreira MC, Klur S, Watanabe M, Németh AH, et al. (2004) Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet.,* 36, 225-227.

Morozov V, Mushegian AR, Koonin EV, Bork P. (1997) A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases. *Trends Biochem Sci.*, 22, 417-418.

Moser MJ, Kamath-Loeb AS, et al. (2000) WRN helicase expression in Werner syndrome cell lines. *Nucleic Acids Res.*, 28, 648-654.

Moyer SE, Lewis PW, Botchan MR. (2006) Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci U S A.*, 103, 10236-10241.

Muftuoglu M, Kusumoto R, Speina E, Beck G, Cheng WH, Bohr VA. (2008) Acetylation regulates WRN catalytic activities and affects base excision DNA repair. *PLoS One*, 3, e1918.

Mullen JR, Kaliraman V, Brill SJ. (2000) Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae. *Genetics,* 154, 1101-1114.

Multani AS, Chang S. (2007) WRN at telomeres: implications for aging and cancer. *J Cell Sci.*, 120, 713-721. Review.

Nagy GN, Suardíaz R, Lopata A, Ozohanics O, Vékey K, Brooks BR, Leveles I, Tóth J, Vértessy BG1, Rosta E. (2016) Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *J Am Chem Soc*., 138, 15035-15045.

Nakayama H, Nakayama K, Nakayama R, Irino N, Nakayama Y. and Hanawalt PC. (1984) Isolation and genetic characterization of a thymineless death-resistant mutant of Escherichia coli K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway. *Mol Gen Genet.,* 195, 474-480.

Newman JA, Savitsky P, Allerston CK, Bizard AH, Özer Ö, Sarlós K, Liu Y, Pardon E, Steyaert J, Hickson ID, Gileadi O. (2015) Crystal structure of the Bloom's syndrome helicase indicates a role for the HRDC domain in conformational changes. *Nucleic Acids Res*., 43, 5221-5235.

Nguyen GH, Dexheimer TS, Rosenthal AS, Chu WK, Singh DK, Mosedale G, et al. (2013). A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem. Biol*., 20, 55-62.

Nimonkar AV, Ozsoy AZ, Genschel J, Modrich P, Kowalczykowski SC. (2008) Human exonuclease 1 and BLM helicase interact to resect DNA and initiate DNA repair. *Proc. Natl. Acad. Sci. USA*., 105, 16906-16911.

Nimonkar AV, Genschel J, Kinoshita E, Polaczek P, Campbell JL, et al. (2011) BLM-DNA2- RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev*., 25, 350–362.

O'Sullivan RJ, Karlseder J. (2010) Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol*., 11, 171-181.

Ohlenschläger O, Kuhnert A, Schneider A, Haumann S, Bellstedt P, Keller H, Saluz HP, Hortschansky P, Hänel F, Grosse F, Görlach M, Pospiech H. (2012) The N-terminus of the human RecQL4 helicase is a homeodomain-like DNA interaction motif. *Nucleic Acids Res*. 40, 8309-8324.

Onoda F, Seki M, Miyajima A, Enomoto T. (2000) Elevation of sister chromatid exchange in Saccharomyces cerevisiae sgs1 disruptants and the relevance of the disruptants as a system to evaluate mutations in Bloom's syndrome gene. *Elsevier,* 459, 203-209.

Opresko PL, von Kobbe C, et al. (2002) Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. *J Biol Chem*., 277, 41110-41119.

Opresko PL, Cheng WH, von Kobbe C, Harrigan JA, Bohr VA. (2003) Werner syndrome and the function of the Werner protein; what they can teach us about the molecular aging process.
Carcinogenesis, 24, 791-802.

Opresko PL, Otterlei M, et al. (2004) The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. *Mol Cell*., 14, 763-774.

Opresko PL, Mason PA, Podell ER, Lei M, Hickson ID, Cech TR, Bohr VA. (2005) POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol Chem*., 280, 32069-32080.

Orren DK, Theodore S, Machwe A. (2002) The Werner syndrome helicase/exonuclease (WRN) disrupts and degrades D-loops in vitro. Biochemistry, 41, 13483-13488.

Oshima J, Sidorova JM, Monnat RJ Jr. (2017) Werner syndrome: Clinical features, pathogenesis and potential therapeutic interventions. *Ageing Res Rev*., 33, 105-114.

Otterlei M, Bruheim P, Ahn B, Bussen W, Karmakar P, Baynton K, Bohr VA. (2006) Werner syndrome protein participates in a complex with RAD51, RAD54, RAD54B and ATR in response to ICL-induced replication arrest. *J Cell Sci*., 119, 5137-5146.

Ouyang KJ, Woo LL, Zhu J, Huo D, Matunis MJ, Ellis NA. (2009) SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Biol*., 7, e1000252.

Ouyang KJ, Yagle MK, Matunis MJ, Ellis NA. (2013) BLM SUMOylation regulates ssDNA accumulation at stalled replication forks. *Front Genet*., 4, 167.

Park SJ, Lee YJ, Beck BD, Lee SH. (2006) A positive involvement of RecQL4 in UV induced S-phase arrest. *DNA and cell biology*, 25, 696-703.

Parvathaneni S, Stortchevoi A, Sommers JA, Brosh RM Jr, Sharma S. (2013) Human RECQ1 interacts with Ku70/80 and modulates DNA end-joining of double-strand breaks. *PLoS ONE*, 8, e62481.

Patro BS, Frøhlich R, Bohr VA, Stevnsner T. (2011) WRN helicase regulates the ATR-CHK1-induced S-phase checkpoint pathway in response to topoisomerase-I-DNA covalent complexes. *J Cell Sci*., 124, 3967-3979.

Pause A, Sonenberg N. (1992) Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J*., 11, 2643-2654.

Perry JJ, Yannone SM, Holden LG, Hitomi C, Asaithamby A, Han S, Cooper PK, Chen DJ, Tainer JA. (2006) WRN exonuclease structure and molecular mechanism imply an editing role in DNA end processing. *Nat Struct Mol Biol.,* 13, 414-422.

Petkovic M, Dietschy T, Freire R, Jiao R, Stagljar I. (2005) The human Rothmund- Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. *Journal of cell science*, 118, 4261-4269.

Pfeiffer V, Crittin J, Grolimund L, Lingner J (2013) The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *EMBO J,* 32, 2861-2871.

Pike AC, Shrestha B, Popuri V, Burgess-Brown N, Muzzolini L, Costantini S, Vindigni A, Gileadi O. (2009) Structure of the human RECQ1 helicase reveals a putative strand-separation pin. *Proc Natl Acad Sci U S A.*, 106, 1039-1044.

Pike AC, Gomathinayagam S, Swuec P, Berti M, Zhang Y, Schnecke C, Marino F, von Delft F, Renault L, Costa A, Gileadi O, Vindigni A. (2015) Human RECQ1 helicase-driven DNA unwinding, annealing, and branch migration: insights from DNA complex structures. *Proc Natl Acad Sci USA,* 112, 4286-4291.

Popuri V, Huang J, Ramamoorthy M, Tadokoro T, Croteau DL, Bohr VA. (2013) RECQL5 plays co-operative and complementary roles with WRN syndrome helicase. *Nucleic Acids Res*., 41, 881-899.

Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH, Chédin F, LaSalle JM. (2013) R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation*. Proc Natl Acad Sci U S A.*, 110, 13938-13943.

Rad B, Kowalczykowski SC. (2012) Translocation of E. coli RecQ helicase on single-stranded DNA. *Biochemistry*, 51, 2921-2929.

Rigby RE, Webb LM, Mackenzie KJ, Li Y, Leitch A, Reijns MA, Lundie RJ, Revuelta A, Davidson DJ, Diebold S, Modis Y, MacDonald AS, Jackson AP. (2014) RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J.*, 33, 542-558.

Rooney S, Chaudhuri J, Alt FW. (2004) The role of the non-homologous end-joining pathway in lymphocyte development. *Immunol.,* 200, 115-131.

Rong SB, Väliaho J, Vihinen M. (2000) Structural basis of Bloom Syndrome (BS) causing mutations in the BLM helicase domain. *Molecular Medicine,* 6, 155-164.

Rossi ML, Ghosh AK, Kulikowicz T, Croteau DL, Bohr VA (2010) Conserved helicase domain of human RecQ4 is required for strand annealing-independent DNA unwinding. *DNA Repair (Amst).*, 9, 796-804.

Rothmund A. (1868) Über Cataracte in Verbindung mit einer eigenthümlichen Hautdegeneration. *Albrecht von Graefes Arch Klin Ophthal,* 14, 159.

Saikrishnan K, Powell B, Cook NJ, Webb MR, Wigley DB. (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell*, 137, 849-859.

Sallmyr A, Tomkinson AE, Rassool FV. (2008) Up-regulation of WRN and DNA ligase IIIαin chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks. *Blood,* 112, 1413-1423.

Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, Venkitaraman AR. (2005) Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell*, 121, 887-898.

Santos-Pereira JM, Herrero AB, García-Rubio ML, Marín A, Moreno S, Aguilera A. (2013) The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev.*, 27, 2445-2458.

Sarlós K, Gyimesi M, Kovács M. (2012) RecQ helicase translocates along single-stranded DNA with a moderate processivity and tight mechanochemical coupling. *Proc Natl Acad Sci U S A*, 109, 9804-9809.

Sato A, Mishima M, Nagai A, Kim SY, Ito Y, Hakoshima T, Jee JG, Kitano K. (2010) Solution structure of the HRDC domain of human Bloom syndrome protein BLM. *J Biochem.*, 148, 517-525.

Sauer M, Paeschke K. (2017) G-quadruplex unwinding helicases and their function in vivo. *Biochem Soc Trans.*, 45, 1173-1182.

Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, et al. (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science*, 277, 333-338.

Schurman SH, Hedayati M, Wang Z, Singh DK, Speina E, Zhang Y, Becker K, Macris M, Sung P, Wilson DM 3rd, Croteau DL, Bohr VA. (2009) Direct and indirect roles of RECQL4 in modulating base excision repair capacity. *Hum Mol Genet.*, 18, 3470-3483.

Schwendener S, Raynard S, Paliwal S, Cheng A, Kanagaraj R, et al. (2010) Physical interaction of RECQ5 helicase with RAD51 facilitates its anti-recombinase activity. *J. Biol. Chem*., 285, 15739-15745.

Sedlackova H, Cechova B, Mlcouskova J, Krejci L. (2015) RECQ4 selectively recognizes Holliday junctions. *DNA Repair (Amst)*., 30, 9.

Shamanna RA, Singh DK, Lu H, Mirey G, Keijzers G, Salles B, Croteau DL, Bohr VA. (2014) RECQ helicase RECQL4 participates in non-homologous end joining and interacts with the Ku complex. *Carcinogenesis*, 35, 2415-2424.

Shamanna RA, Croteau DL, Lee JH, Bohr VA. (2017) Recent Advances in Understanding Werner Syndrome. *F1000Res*. 28, 1779.

Sharma S, Sommers JA, Choudhary S, Faulkner JK, Cui S, Andreoli L, Muzzolini L, Vindigni A, Brosh RM Jr, (2005) Biochemical analysis of the DNA unwinding and strand annealing activities catalyzed by human RECQ1. *J Biol Chem,* 280, 28072-28084.

Sharma S, Phatak P, Stortchevoi A, Jasin M, Larocque JR. (2012) RECQ1 plays a distinct role in cellular response to oxidative DNA damage. *DNA Repair (Amst).,* 11, 537-549.

Shen JC, Gray MD, Oshima J, Kamath-Loeb AS, Fry M and Loeb LA. (1998) Werner syndrome protein. I. DNA helicase and dna exonuclease reside on the same polypeptide, *J Biol Chem,* 273, 34139-34144.

Siitonen HA, Sotkasiira J, Biervliet M, Benmansour A, Capri Y, Cormier-Daire V, Crandall B, Hannula-Jouppi K, Hennekam R, Herzog D, Keymolen K, Lipsanen-Nyman M, Miny P, Plon SE, Riedl S, Sarkar A, Vargas FR, Verloes A, Wang LL, Kääriäinen H, Kestilä M. (2009) The mutation spectrum in RECQL4 diseases. *Eur J Hum Genet.,* 17, 151-158.

Singh DK, Karmakar P, Aamann M, Schurman SH, May A, Croteau DL, Burks L, Plon SE, Bohr VA. (2010) The involvement of human RECQL4 in DNA double-strand break repair. *Aging Cell*., 9, 358-371.

Singh DK, Popuri V, Kulikowicz T, Shevelev I, Ghosh AK, et al. (2012) The human RecQ helicases BLM and RECQL4 cooperate to preserve genome stability. *Nucleic Acids Res*., 40, 6632-6648.

Singleton MR, Scaife S, Wigley DB. (2001) Structural analysis of DNA replication fork reversal by RecG. *Cell*, 107, 79-89.

Singleton MR, Wigley DB. (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol*., 184, 1819-1826.

Singleton MR, Dillingham MS, Wigley DB. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*., 76, 23-50.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*., 42, 794-805.

Sollier J, Cimprich KA. (2015) Breaking bad: R-loops and genome integrity. *Trends Cell Biol*., 25, 514-522.

Sordet O, Redon CE, Guirouilh-Barbat JEE, Smith S, Solier SEP, Douarre CEL, Conti C, Nakamura AJ, Das BB, Nicolas E, et al. (2009) Ataxia telangiectasia mutated activated by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Reports*, 10, 887-893.

Sousa FG, Matuo R, Soares DG, Escargueil AE, Henriques JA, Larsen AK, Saffi J. (2012) PARPs and the DNA damage response. *Carcinogenesis*, 33, 1433-1440.

Spies M. (2014) Two steps forward, one step back: determining XPD helicase mechanism by single-molecule fluorescence and high-resolution optical tweezers. *DNA Repair (Amst),* 20, 58-70.

Stewart E, Chapman CR, Al-Khodairy F, Carr AM, Enoch T. (1997) rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. *EMBO J.,* 16, 2682-2692.

Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, Hieter P. (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & Development*, 26, 163-175.

Stirling PC, Hieter P. (2017) Canonical DNA Repair Pathways Influence R-Loop-Driven Genome Instability. *J Mol Biol*., 429, 3132-3138.

Stuckey R, García-Rodríguez N, Aguilera A, Wellinger RE. (2015) Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proc Natl Acad Sci U S A*, 112, 5779-5784.

Sun H, Karow JK, Hickson ID, Maizels N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem*., 273, 27587-27592.

Suzuki T, Kohno T, Ishimi Y. (2009) DNA helicase activity in purified human RECQL4 protein. *J Biochem*., 146, 327-335.

Swan MK, Legris V, Tanner A, Reaper PM, Vial S, Bordas R, Pollard JR, Charlton PA, Golec JM, Bertrand JA. (2014) Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA. *Acta Crystallogr D Biol Crystallogr*., 70, 1465-1475.

Tadokoro T, Kulikowicz T, Dawut L, Croteau DL, Bohr VA. (2012) DNA binding residues in the RQC domain of Werner protein are critical for its catalytic activities. *Aging,* 4, 417-429.

Tanaka S, Umemori T, Hirai K, Muramatsu S, Kamimura Y, Araki H. (2007) CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. *Nature,* 445, 328-332.

Tanaka T, Umemori T, Endo S, Muramatsu S, Kanemaki M, Kamimura Y, Obuse C, Araki H. (2011) Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J*., 30, 2019-2030.

Tang S, Wu MKY, Zhang R, Hunter N. (2015) Pervasive and essential roles of the Top3-Rmi1 decatenase orchestrate recombination and facilitate chromosome segregation in meiosis. *Mol Cell*., 57, 607-621.

Tang W, Robles AI, Beyer RP, Gray LT, Nguyen GH, Oshima J, Maizels N, Harris CC, Monnat RJ Jr. (2016) The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum Mol Genet*., 25, 2060-2069.

Tangeman L, McIlhatton MA, Grierson P, Groden J, Acharya S. (2016) Regulation ofBLM nucleolar localization. *Genes,* (Basel), 7.

Tanner NK, Linder P. (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell.,* 8, 251-262.

Tanner NK, Cordin O, Banroques J, Doère M, Linder P. (2003) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell*., 11, 127-138.

Thangavel S, Mendoza-Maldonado R, Tissino E, Sidorova JM, Yin J, Wang W, Monnat RJ Jr, Falaschi A, Vindigni A. (2010) Human RECQ1 and RECQ4 helicases play distinct roles in DNA replication initiation. *Mol Cell Biol*., 30, 1382-1396.

Thomas C, Tulin AV. (2013) Poly-ADP-ribose polymerase: machinery for nuclear processes. *Mol Aspects Med*., 34, 1124-1137.

Thomson MS (1936) Poikiloderma congenitale. *Br J Dermatol,* 48, 221.

Tikoo S, Madhavan V, Hussain M, et al. (2013) Ubiquitin-dependent recruitment of the Bloom Syndrome helicase upon replication stress is required to suppress homologous recombination. *EMBO J.*, 32, 1778–1792.

Tous C, Aguilera A. (2007) Impairment of transcription elongation by R-loops in vitro. *Biochem Biophys Res Commun.*, 360, 428-432.

Trego KS, Chernikova SB, Davalos AR, Perry JJ, Finger LD, Ng C, Tsai MS, Yannone SM, Tainer JA, Campisi J, Cooper PK. (2011) The DNA repair endonuclease XPG interacts directly and functionally with the WRN helicase defective in Werner syndrome. *Cell Cycle*, 10, 1998-2007.

Tripathi V, Agarwal H, Priya S, Batra H, Modi P, Pandey M, Saha D, Raghavan SC, Sengupta S. (2018) MRN complex-dependent recruitment of ubiquitylated BLM helicase to DSBs negatively regulates DNA repair pathways. *Nat. Commun.,* 9, 1016.

Tuduri S, Crabbé L, Conti C, Tourrière H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, et al. (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell. Biol.,* 11, 1315-1324.

Turaga RV, Massip L, Chavez A, Johnson FB, Lebel M. (2007) Werner syndrome protein prevents DNA breaks upon chromatin structure alteration. *Aging Cell.*, 6, 471-481.

Umate P, Tuteja N, Tuteja R. (2011) Genome-wide comprehensive analysis of human helicases. *Commun Integr Biol.,* 4, 118-137.

Umezu K, Nakayama K, Nakayama H. (1990) Escherichia coli RecQ protein is a DNA helicase. *Proc Natl Acad Sci U S A*. 1990 Jul;87(14):5363-7. Erratum in: *Proc Natl Acad Sci U S A,* 87, 9072.

Umezu K, Nakayama H. (1993) RecQ DNA helicase of Escherichia coli. Characterization of the helix-unwinding activity with emphasis on the effect of single-stranded DNA-binding protein. *J Mol Biol.*, 230, 1145-1150.

van Brabant AJ, Stan R, Ellis NA. (2000) DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet.*, 1, 409-459.

Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, 97, 75-84.

Vennos EM, Collins M, James WD. (1992) Rothmund-Thomson syndrome: review of the world literature. *J Am Acad Dermatol*., 27, 750-762.

Vindigni A, Hickson ID. (2009) RecQ helicases: multiple structures for multiple functions? *HFSP J.,* 3, 153-164.

Vindigni A, Marino F, Gileadi O. (2010) Probing the structural basis of RecQ helicase function. *Biophys Chem.,* 149, 67-77.

von Hippel PH. (2004) Helicases become mechanistically simpler and functionally more complex. *Nat Struct Mol Biol*., 11, 494-496.

von Kobbe C, Karmakar P, Dawut L, Opresko P, Zeng X, Brosh RM Jr, Hickson ID, Bohr VA. (2002) Colocalization, physical, and functional interaction between Werner and Bloom syndrome proteins. *J Biol Chem*., 277, 22035-22044.

von Kobbe C, Harrigan JA, May A, Opresko PL, Dawut L, Cheng WH, Bohr VA. (2003) Central role for the Werner syndrome protein/poly(ADP-ribose) polymerase 1 complex in the poly(ADP- ribosyl)ation pathway after DNA damage. *Mol Cell Biol*., 23, 8601-8613.

Wahba L, Amon JD, Koshland D, Vuica-Ross M. (2011) RNase H and Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability. *Mol. Cell*., 44, 978-988.

Walker JE, Saraste M, Runswick MJ, Gay NJ. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.,* 1, 945-951.

Wang LL, Levy ML, Lewis RA, Chintagumpala MM, Lev D, Rogers M, Plon SE. (2001) Clinical manifestations in a cohort of 41 Rothmund-Thomson syndrome patients. *Am J Med Genet*., 102, 11-17.

Wang D, Luo M, Kelley MR. (2004) Human apurinic endonuclease 1 (APE1) expression and prognostic significance in osteosarcoma: enhanced sensitivity of osteosarcoma to DNA damaging agents using silencing RNA APE1 expression inhibition. *Mol Cancer Ther*., 3, 679-686.

Wang H, Rosidi B, Perrault R, Wang M, Zhang L, et al. (2005) DNA ligase III as a candidate component of backup pathways of nonhomologous end joining. *Cancer Res*., 65, 4020-4030.

Wang M, Wu W, Wu W, Rosidi B, Zhang L, et al. (2006) PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res*., 34, 6170-6182.

Wang Y, Li H, Tang Q, Maul GG, Yuan Y. (2008) Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: involvement of host cellular factors. *J Virol*., 82, 2867-2882.

Werner O. (1985) On cataract in conjunction with scleroderma. Otto Werner, doctoral dissertation, 1904, Royal Ophthalmology Clinic, Royal Christian Albrecht University of Kiel. *Adv Exp Med Biol*., 190, 1-14.

White R, Saxty B, Large J, Kettleborough CA, Jackson AP. (2013) Identification of small-molecule inhibitors of the ribonuclease H2 enzyme. *J Biomol Screen*., 18, 610-620.

Wilson DM III, Bohr VA. (2007) The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair*, 6, 544–559.

Woo LL, Futami K, Shimamoto A, Furuichi Y, Frank KM. (2006) The Rothmund- Thomson gene product RECQL4 localizes to the nucleolus in response to oxidative stress. *Exp Cell Res*., 312, 3443-3457.

Wu X, Maizels N. (2001) Substrate-specific inhibition of RecQ helicase. *Nucleic Acids Res*., 29, 1765-1771.

Wu Y. (2012) Unwinding and rewinding: double faces of helicase? *J Nucleic Acids,* 2012, 140601.

Wu L, Hickson ID. (2003) The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426, 870-874.

Xie A, Kwok A, Scully R. (2009) Role of mammalian Mre11 in classical and alternative nonhomologous end joining. *Nat. Struct. Mol. Biol*., 16, 814-818.

Xu X, Liu Y. (2009) Dual DNA unwinding activities of the Rothmund-Thomson syndrome protein, RECQ4. *EMBO J*., 28, 568-577.

Xu X, Rochette PJ, Feyissa, EA, Su TV, Liu Y. (2009) MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication. *EMBO J*., 28, 3005-3014.

Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. (2014) Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell*., 53, 484-497.

Yannone SM, Roy S, Chan DW, Murphy MB, Huang S, et al. (2001) Werner syndrome protein is regulated and phosphorylated by DNA-dependent protein kinase. *J. Biol. Chem*., 276, 38242-38248.

Yin J, Kwon YT, Varshavsky A, Wang W. (2004) RECQL4, mutated in the Rothmund-Thomson and RAPADILINO syndromes, interacts with ubiquitin ligases UBR1 and UBR2 of the N-end rule pathway. *Hum. Mol. Genet*., 13, 2421-2430.

Yu CE, Oshima J, et al. (1996) A YAC, P1, and cosmid contig and 17 new polymorphic markers for the Werner syndrome region at 8p12-p21. *Genomics*, 35, 431-440.

Yuan Z, Riera A, Bai L, Sun J, Nandi S, Spanos C, Chen ZA, Barbon M, Rappsilber J, Stillman B, Speck C, Li H. (2017) Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1. *Nat Struct Mol Biol*., 24, 316-324.

Zegerman P, Diffley JF. (2007) Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. *Nature*, 445, 281-285.

Zhang X, Wigley DB. (2008) The "glutamate switch" provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat Struct Mol Biol*., 15, 1223-1227.