



**UNIVERSITÀ
DEGLI STUDI
DI TRIESTE**

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVII CICLO DEL DOTTORATO DI RICERCA IN CHIMICA

Supported by an AIRC fellowship for Italy

Borsa di Studio AIRC Guglielmina Lucatello e Gino Mazzega FIRC - ID 26979-2021

Distinct and overlapping roles of human RecQ helicases in the recognition and unwinding of different G-quadruplexes

Settore scientifico-disciplinare: **BIO/10**

**DOTTORANDA
ELISA DE PIANTE**

**COORDINATORE
PROF. ENZO ALESSIO**

**SUPERVISORE DI TESI
PROF. SILVIA ONESTI**

**CO-SUPERVISORE DI TESI
LUISA MR NAPOLITANO**

ANNO ACCADEMICO 2023/2024

Abstract

G-quadruplexes (G4) are non-canonical nucleic acid structures, which are widely distributed across the genome, but are particularly enriched within human oncogenic promoters and at telomeres. They have been implicated in many cellular pathways such as transcription, replication, and telomere homeostasis. Their formation and regulated unwinding are critical processes for genome stability, and their dysregulation has been linked to cancer development.

Several helicases, including the RecQ family helicases and the Iron-Sulfur (FeS) helicase family, have been shown to target and regulate G4 structures, and hence play a key role in G4 metabolism.

RecQ helicases are highly conserved in evolution and are implicated in many aspects of nucleic acid metabolism and genome stability. In human cells there are five RecQ (RecQ1, Bloom, Werner, RecQ4 and RecQ5), involved in cancer development and progression. Beside opening DNA duplexes, they are specialized in unwinding non-canonical DNA structures, including D-loops, R-loops, triple helices and G4s. Although there are various reports on the interactions between RecQ helicases and G4s, there is the need of a more systematic analysis to clarify their specificity towards different types of G4 structures, and to better understand their distinct and/or overlapping roles in G4 metabolism.

We expressed and purified the catalytic domains of RecQ1, BLM, WRN, RecQ4 and RecQ5 and first evaluated their activity on a canonical DNA Fork substrate, as well as D-loops and R-loops. All the proteins displayed the ability to unwind these substrates, although with different efficacies, demonstrating to be active fragments. We then explored their interaction with a panel of eleven physiologically relevant unimolecular G4s, from different genomic regions (promoter or telomeric) and with different topologies (parallel, antiparallel or hybrid). We analyzed both the binding affinity of each RecQ with the diverse G4s, as well as their ability to unwind them. The results showed both specificity and selectivity: each helicase showed a different range of affinity and unwinding capability towards G4s, and at the same time we observed significance G4 selectivity within each helicase.

BLM and WRN are able to bind with high affinity and unwind many G4s, exhibiting a different pattern of G4 recognition. As previous reports suggest that the helicase activity can be influenced by G4 thermal stability in response to ionic conditions, CD spectroscopy assessed the stability of each G4 in various buffers, and confirmed that the preferences were not simply due to the relative stability. Interestingly, BLM ability to unwind G4s remained unaffected by increasing potassium concentrations, whereas WRN activity decreased in high KCl environments. In contrast, RecQ1, RecQ4 and RecQ5 show very low or no binding and unwinding activity.

The more stable helicase-G4 complexes were used in a large number of crystallization experiments testing diverse crystallization conditions, in the hope to obtain crystals suitable for X-ray diffraction, providing information on the determinants of the interaction. However, no well-diffracting crystal was obtained.

The crystal structure of a bacterial RecQ bound to a resolved G4 revealed the presence of a putative guanine-binding pocket that was suggested to be common to RecQ helicases (Voter et al., 2018). We carried out site-directed mutagenesis experiments on BLM and WRN to disrupt the pocket and experimentally test this hypothesis: our preliminary data confirm its role in G4 interaction. Further experiments are still in progress.

Additionally, we investigated the *Escherichia coli* helicase DinG, a DNA damage-inducible FeS helicase implicated in DNA repair. It unwinds DNA duplex with a 5' to 3' polarity and shows specificity for several unusual DNA structures, being active on D-loops, R-loops and G-quadruplexes structures. With the aim of better understanding the role of DinG in G4 metabolism, the *E. coli* DinG protein was expressed and purified, and its interactions with different unimolecular G4 topologies were systematically analyzed. In collaboration with the University of Naples, we elucidated the thermodynamic and kinetic aspects of G4-helicase interactions and tested its unwinding activity towards these substrates. The G4/DinG interactions have also been investigated in the presence of well-known G4 ligands, which can significantly interfere with several biological processes involving G4s. Our results show that DinG can bind and unwind various unimolecular DNA G4 structures at physiological concentrations, with a preference to unwind G4s in a parallel conformation. Moreover, when G4s were stabilized by selective ligands, DinG activity decreased or was abolished, although G4 recognition remained unaffected.

Overall, our findings highlight the specificity of helicase interactions with G4s and highlight the importance of investigating multiple G4 structures to fully understand their role in G4 metabolism.

Acknowledgments

Alla fine di questo elaborato desidero ringraziare tutti coloro che mi hanno sostenuto, supportato e sopportato, rendendo possibile questo percorso e il suo compimento.

Innanzitutto, vorrei ringraziare Silvia Onesti per avermi dato l'opportunità di iniziare questa esperienza e aver creduto in me nel corso di tutti questi anni. Le sue conoscenze e la sua esperienza mi hanno permesso di imparare molto e, unite alla sua disponibilità, guida e gentilezza, mi hanno fatto ottenere molti risultati. Non da meno ringrazio Luisa Napolitano, che fin da subito mi ha preso sotto la sua ala insegnandomi tutto ciò che ho imparato finora. Senza le sue indicazioni, sarebbe stato certamente più difficile districarmi tra le insidie del mio progetto di ricerca. Oltre che mentore è stata anche un'ottima compagna di viaggio in questi anni trascorsi in laboratorio.

Ringrazio Elettra Sincrotrone Trieste e la Fondazione AIRC per la Ricerca sul Cancro per avermi dato la possibilità e gli strumenti per fare tutto ciò, oltre che per aver finanziato la mia attività di ricerca.

Ringrazio i collaboratori, in particolare Concetta Giancola, dell'Università degli Studi di Napoli, che mi ha fatto conoscere e mi ha introdotto nel mondo dei G-quadruplexes. È stata una delle prime a supportare questo mio percorso. Ringrazio anche Jussara Amato e Federica D'Aria, che si sono sempre dimostrate collaborative e disponibili.

Un ringraziamento va anche a Joseph Newman, Opher Gileadi e Alessandro Vindigni per avermi fornito i costrutti su cui lavorare e per essersi dimostrati disponibili nel rispondere a dubbi ed incertezze, favorendo un dialogo scientifico costruttivo.

Ringrazio anche tutti i membri del laboratorio, che fin da subito mi hanno accolto e messo a proprio agio, rendendo l'ambiente di lavoro molto più piacevole.

Grazie a tutti voi e ai vostri consigli ho accresciuto le mie conoscenze e le mie competenze.

Elisa De Piante was supported by an AIRC fellowship for Italy "Guglielmina Lucatello e Gino Mazzega FIRC – ID 26979-2021"

Contents

Abstract	I
Acknowledgments.....	III
Contents.....	V
List of Tables	IX
List of Figures	XI
Introduction	1
1. DNA secondary structures.....	1
2. G-quadruplex substrates	2
2.1. G-quadruplexes biological functions.....	5
2.1.1. G-quadruplexes in DNA replication.....	5
2.1.2. G-quadruplexes in transcription.....	6
2.1.3. G-quadruplexes at telomeres.....	7
2.2. Displacement loops and the interplay between G-quadruplexes and R-loops	8
2.3. G-quadruplexes as therapeutic target and tool	9
2.3.1. G-quadruplex ligands	10
2.3.2. G-quadruplexes binding proteins	10
3. Helicases.....	12
3.1. Superfamily 1 and Superfamily 2 helicases (SF1 and SF2)	12
3.1.1. Helicase architecture	12
3.1.2. Translocation mechanism.....	14
3.2. G-quadruplex unwinding helicases	14
3.3. Crystal structures of helicases bound to G4	16
3.3.1. Co-crystal structure of DXH36 bound to cMYC G4	17
3.3.2. Co-crystal structure of CsRecQ bound to a resolved G4 DNA.....	17
3.3.3. Co-crystal structure of ToPif1 bound to G4	19
3.3.4. Co-crystal structure of ScPif1 bound to G4	19
4. Human RecQ helicases	21
4.1. RecQ domain organization	21
4.2. Biochemical activities of human RecQ helicases.....	24
4.3. Functional role of human RecQ helicases	25
4.3.1. RecQ helicases in DNA repair	26

4.3.2. RecQ helicases in DNA replication.....	27
4.3.3. RecQ helicases in telomere maintenance	28
4.4. RecQ-associated syndromes	29
4.5. RecQ helicases as therapeutic targets for cancer.....	31
4.6. Human RecQ helicases in G-quadruplex metabolism.....	32
4.6.1. BLM in G4 metabolism	32
4.6.2. WRN in G4 metabolism.....	35
4.6.3. RecQ1 in G4 metabolism	36
4.6.4. RecQ4 in G4 metabolism	37
4.6.5. RecQ5 in G4 metabolism	38
5. Iron-Sulphur Cluster helicases	39
5.1. <i>E. coli</i> DinG helicase	40
6. Project aim.....	42
Materials and Methods	44
7. Construct design and cloning.....	44
7.1. Cloning strategy used: Ligation Independent Cloning.....	44
7.2. Expression vectors used for cloning human RecQ fragments and <i>E. coli</i> DinG.....	44
7.3. Insert preparation.....	45
7.4. Vector preparation.....	46
7.5. Ligation and transformation	47
7.6. Preparation of plasmid minipreps.....	48
7.7. Bioinformatic analysis for guanine-specific binding pocket identification.....	48
7.8. Site directed mutagenesis of BLM and WRN	49
8. Protein expression in bacterial <i>E. coli</i> cells.....	50
8.1. Test expression and solubility assay	50
8.2. Small scale expression test and purification.....	52
9. Purification technique used for Protein Large-scale purification	52
9.1. Ni-affinity Chromatography or Immobilized Metal Ion Affinity Chromatography (IMAC).....	52
9.2. Heparin affinity chromatography	53
9.3. Tag removal and dialysis	53
9.4. Cation Exchange chromatography.....	53
9.5. Size Exclusion Chromatography (SEC)	54
10. Large-scale expression and protein purification	55
10.1. RecQ1 purification	56
10.2. BLM purification.....	56

10.3. BLM S897A and BLM S965A purifications	57
10.4. WRN purification	57
10.5. RecQ4 fragments purifications.....	58
10.5.1. RecQ4 fragment 371-1116	58
10.5.2. RecQ4 fragment 1-55	59
10.5.3. RecQ4 fragment 1-133	59
10.5.4. RecQ4 fragment 1-427	60
10.6. RecQ5 purification	61
10.7. <i>E. coli</i> DinG purification	61
11. Biochemical and biophysical studies	62
11.1. Determination of protein concentration.....	62
11.2. Sodium Dodecyl Sulphate – PolyAcrylamide Gel Electrophoresis (SDS-PAGE).....	62
11.3. Preparation of oligonucleotides	62
11.4. Binding assay	65
11.4.1. Electrophoretic Mobility Shift Assays (EMSA).....	65
11.4.2. Fluorescence Anisotropy	66
11.5. Helicase assay	67
11.5.1. Gel-based helicase assay	67
11.5.2. FRET-based helicase assay.....	68
11.5.3. Two-step fluorescent-based helicase assay of G4	69
11.5.4. Two-step fluorescent-based helicase assay of G4 in presence of G4 ligands	71
11.6. Circular Dichroism (CD) of G-quadruplex.....	71
12. Structural studies: protein crystallization	72
Results and discussion.....	74
13. Expression and purification of human RecQ helicases	74
13.1. An overview.....	74
13.2. Purification of human RecQ1 (49-619).....	76
13.3. Purification of human BLM (636-1298)	77
13.4. Purification of human WRN (517-1093).....	78
13.5. Purification of human RecQ4 (371-1116)	80
13.6. Purification of human RecQ5 (11-582).....	81
13.7. Purification of RecQ4 N-terminal fragments	83
13.7.1. Small scale expression test.....	83
13.7.2. Large scale expression	85
14. Assembly and characterization of G-quadruplex structures	89

14.1. CD spectra of G4s.....	91
14.2. Variations of G4s melting temperature in different buffers.....	94
15. RecQ helicases: G4s interactions	96
15.1. The catalytic domain of human RecQ helicases are active helicases able to unwind forked duplex substrates.	96
15.2. Human RecQ helicases unwind non-canonical substrates	97
15.3. BLM and WRN interact with and unfold G-quadruplex structures	100
15.4. RecQ1, RecQ4 and RecQ5 do not show great interaction with G-quadruplexes.....	105
15.5. The N-terminal domains of RecQ4 display selective affinity for G-quadruplex structures.....	108
15.6. RecQ4 full-length binds with high affinity G-quadruplex structures.....	111
16. A putative G4-binding pocket in human RecQ helicases	113
16.1. Structural comparison.....	114
16.2. The putative residues are conserved among organisms.....	115
16.3. Production of BLM and WRN mutants.....	119
16.3.1. Test expression of BLM S897A and BLM S965A mutants	119
16.3.2. Purification of BLM S897A and BLM S965A mutants	120
16.4. Single point mutations decreased BLM unwinding activity towards G-quadruplexes	122
17. Structural studies	124
18. AlphaFold predictions.....	125
19. The interaction between <i>E. coli</i> DinG and G-quadruplexes	127
19.1. Expression and purification of DinG.....	127
19.2. Biochemical and biophysical characterization of DinG helicase	128
19.2.1. DinG binds and unwinds G-quadruplex structures with different efficacy.....	130
19.2.2. DinG unwinding activity is compromised by G4 ligands.....	132
Conclusions	134
Future work	138
References	140

List of Tables

Table 1. List of some helicases and selected features, which have been described to regulate G4 structure in vivo. (Teng et al., 2021)	11
Table 2. Clinical features of RecQ disorders. (Mojumdar, 2020; Abu-Libdeh et al., 2022)	30
Table 3. The vectors used and their main features	45
Table 4. List of primers used for the cloning of different N-terminal fragments of RecQ4	46
Table 5. List of mutants and primers used for site-direct mutagenesis	49
Table 6. Details of construct and final expression protocol	55
Table 7. Oligonucleotides used to make EMSA and helicase assay substrates	63
Table 8. G4 oligonucleotides used to make binding assay substrates	64
Table 9. G4 oligonucleotides used to make RecQs helicase assay substrates	65
Table 10. G4 oligonucleotides used to make <i>E. coli</i> DinG helicase assay substrates	65
Table 11. List of G-quadruplexes used in the present study	90
Table 12. List of G-quadruplexes used in the present study	100
Table 13. List of G-quadruplexes used in the present study	129

List of Figures

Figure 1. Structure of Watson–Crick pairing and Hoogsteen pairing scheme in the DNA base pair.....	2
Figure 2. G-quadruplex polymorphism.	4
Figure 3. The putative biological function of G-quadruplex in DNA transcription (A, B, C, D, E), replication (G, H) and telomere maintenance (F) (Adapted from Teng et al. 2021, Bettin et al, 2019 and Robinson et al., 2021).	5
Figure 4. SF2 family helicase.	13
Figure 5. Scheme illustration of inchworm translocation mechanism. (From Singleton et al., 2007).	14
Figure 6. Involvement of G-quadruplex helicases in cellular processes in vivo.	16
Figure 7. Overall structure of the DHX36–G-quadruplex DNA complex.	17
Figure 8. The guanine-specific binding pocket of the CsRecQ helicase.....	18
Figure 9. Structure of ToPif1 complexed with G4 DNA and ADP·AlF₄ in cartoon mode.	19
Figure 10. Cartoon representation of the crystal structure of ScPif1 bound to G4.....	20
Figure 11. Domain organization of select bacterial and human RecQ helicases.....	22
Figure 12. Crystal structures of the five human RecQ family members.	23
Figure 13. Proposed models for BLM–G4 unfolding mechanisms.....	33
Figure 14. FeS cluster helicases.....	39
Figure 15. <i>E. coli</i> DinG.	41
Figure 16. 6% non-denaturing polyacrylamide gel EMSA representation.	66
Figure 17. Schematic representation of fluorescence anisotropy.....	67
Figure 18. SDS-polyacrylamide gel representative of a resolved helicase assays.....	68
Figure 19. Schematic representation of fluorescence helicase assay based on fluorescence resonance energy transfer (FRET).....	69
Figure 20. A schematic diagram of the two-step fluorescent-based G4 helicase assay.....	70
Figure 21. Schematic representation of RecQ helicases.	75
Figure 22. Purification steps for RecQ1.....	77
Figure 23. Purification steps of BLM.....	78
Figure 24. Purification steps of WRN.	79
Figure 25. Final purification steps of RecQ4 371-1116.....	81
Figure 26. Purification steps of RecQ5.....	82
Figure 27. Schematic representation of N-terminal domains of RecQ4 helicase.	83

Figure 28. An example of small-scale expression and purification test for two different N-terminal domain fragments 1-55 and 1-133 of RecQ4.	84
Figure 29. Examples of small-scale expression and purification tests for the N-terminal domain fragments 1-427 of RecQ4.	85
Figure 30. Purification steps of 1-55 RecQ4 fragment.	86
Figure 31. Purification steps of 1-133 RecQ4 fragment.	87
Figure 32. Purification steps of 1-427 RecQ4 fragment.	88
Figure 33. Schematic representation of CD experiment with representative CD spectra of G4 structures with different topologies.	91
Figure 34. CD characterization of G4s folding.	93
Figure 35. Thermal stability of G4 structures in different buffers.	95
Figure 36. Comparison of RecQ helicases activity on the equivalent DNA Fork substrate.	97
Figure 37. Full-length (FL) RecQ4 binding and unwinding activity on loop substrates.	98
Figure 38. Comparison of RecQ helicases activity on the equivalent A. R-loops and B. D-loops substrates.	99
Figure 39. BLM and WRN interacting with various G-quadruplexes.	101
Figure 40. BLM and WRN unwinding activity towards various G-quadruplexes.	103
Figure 41. BLM and WRN unwinding activity towards various G-quadruplexes in two different annealing buffers.	104
Figure 42. RecQ1, RecQ4 and RecQ5 binding to various G-quadruplexes.	106
Figure 43. RecQ1, RecQ4 and RecQ5 unwinding activity towards various G-quadruplexes.	108
Figure 44. RecQ4 1-427 binding to various G-quadruplexes.	109
Figure 45. Comparison of A. RecQ4 1-55 and B. RecQ4 1-133 binding affinity towards a variety of G4 substrates.	111
Figure 46. RecQ4 full-length binds to various G-quadruplexes.	112
Figure 47. The guanine specific pocket of the CsRecQ helicase (PDB ID: 6CRM).	113
Figure 48. Superimposition of bacterial RecQ and human RecQ helicases structures.	115
Figure 49. Sequence analysis of RecQ helicases. Sequence alignment of RecQ helicases from various organism.	118
Figure 50. Medium-scale expression and purification test of BLM S897A and BLM S965A mutants.	120
Figure 51. Purification steps of BLM S897A.	121
Figure 52. Purification steps of BLM S965A.	121

Figure 53. BLM WT, BLM S897A and BLM S965A unwinding activity towards various G-quadruplexes.....	123
Figure 54. AlphaFold3 models of the interaction between RecQ and G4s.	126
Figure 55. Large scale purification of <i>E. coli</i> DinG.	127
Figure 56. CD spectra	129
Figure 57. Time evolution SPR sensorgrams.....	130
Figure 58. DinG helicase activity towards G4s.....	131
Figure 59. DinG helicase activity in the presence of G4 ligands.....	132

Introduction

Non-canonical DNA structures, especially G-quadruplexes (G4s), have gained a lot of attention in the last two decades. Even though at first they were seen as structural curiosities, they are now recognized to be involved in several key genome functions. Due to the very large amount of literature on the structure, physical-chemical properties and biological role of G-quadruplexes, it is impossible to provide a comprehensive review of the field. This chapter is therefore not meant to be an in-depth review of G4 structure and functions, but simply to briefly introduce the subject.

1. DNA secondary structures

The most widely recognized structure of DNA is the B-form double helix, a right-handed spiral of antiparallel strands held together by complementary base pairing, first described by Watson and Crick in 1953 (Watson & Crick, 1953). This structure is based on specific hydrogen bonding between nucleotide bases, a mechanism now known as Watson-Crick base pairing, where adenine (A) pairs with thymine (T) via two hydrogen bonds, and guanine (G) pairs with cytosine (C) via three hydrogen bonds (Figure 1). This complementary pairing not only ensures the accurate replication of genetic information but also maintains the structural stability of the DNA helix. B-DNA is the most energetically favorable and predominant form under physiological conditions, facilitating interaction with various proteins involved in transcription, replication, and repair.

Despite the centrality of the Watson-Crick double helix, it has become clear that DNA is capable of adopting a variety of non-canonical secondary structures under specific conditions. These structures include displacement loops (D-loops and R-loops), triple helices, hairpin structures, Holliday junctions, i-motifs, and G-quadruplexes. Such non-canonical structures often arise in specific genomic contexts and play important roles in regulating DNA metabolism and gene expression.

While Watson-Crick base pairing remains the standard mechanism for double-stranded DNA formation, many of these alternative structures are stabilized through different modes of interaction, of which the most common is Hoogsteen base pairing. Hoogsteen bonding shows an unusual base pair modality in which the purine base is rotated 180° around its glycosidic bond, adopting a different orientation and allowing the formation of structures with unique geometries (Figure 1). Unlike Watson-Crick base pairs, which maintain a uniform helical distance, Hoogsteen base pairing creates a more compact conformation with a shorter C1'-C1' distance between bases (Hoogsteen, 1963). Hoogsteen base-pairing is not limited to A-T and C-G, but can involve other pairs, such as G-G; Hoogsteen base-pairing involves a different “face” of the base with respect to Watson-Crick base

pairing, and it is therefore compatible with it, for example in the formation of triple helices. The interplay between Watson-Crick and Hoogsteen base pairing highlights the structural versatility of DNA and its ability to form specialized and non-canonical structures that influence gene regulation and genome stability (Conroy et al., 2022; Duardo et al., 2023).

This thesis focuses on DNA G-quadruplex structures, exploring their interaction with helicases and their implications for genomic stability and cancer.

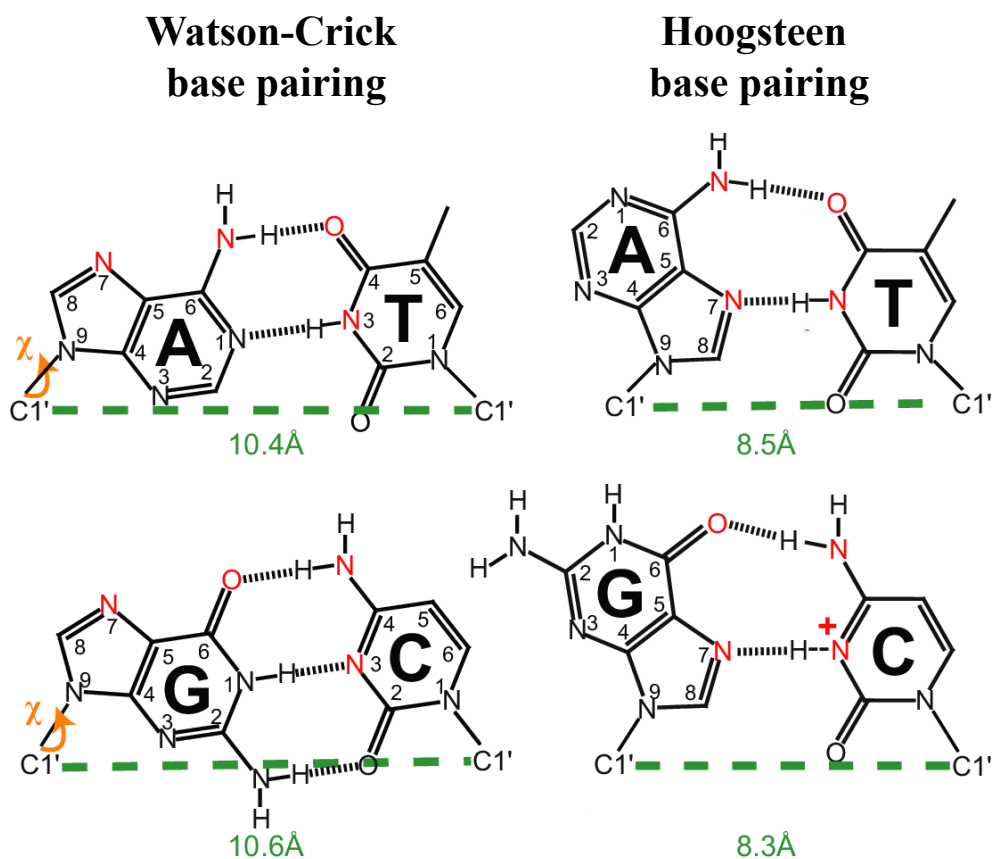


Figure 1. Structure of Watson–Crick pairing and Hoogsteen pairing scheme in the DNA base pair with highlighted key geometrical differences. Heavy atoms involved in HG hydrogen bonds (in red), syn angle (in orange) and constricted C1'–C1' distances (in green). Average C1'–C1' distances from the survey are shown for each base-pair type. (From Zhou et al., 2015)

2. G-quadruplex substrates

G-quadruplexes are non-canonical DNA or RNA structures formed by guanine-rich sequences that are able to build up in stacked planar G-tetrads, each consisting of four guanines bases held together by Hoogsteen hydrogen bonds. The hydrogen bond network within a single quartet is not enough to guarantee a stable structure, therefore two or more G-quartets are linked and primarily stabilized

through the overlap of π orbitals of stacked guanines (π - π stacking). The whole structure is additionally stabilized by monovalent cations (in the order of preference $K^+ > Na^+ > Li^+$) which are positioned along the central axis. (Figure 2A).

Adjacent G-tracts are connected by stretches of nucleotides, called loops, which can differ in length and base composition. Length, sequence and conformation of the loops influence not only the G4 topology but also its stability. In fact, G4s with longer loops have a lower thermodynamic stability, and single nucleotide substitutions in the loops sequence appear to strongly affect G4s stability. In particular pyrimidines (T, C) are preferred over adenine, both as single base and longer loops (Guedin et al., 2008).

According to the number of DNA strands that forms the G4 structure, they can be classified as intermolecular G4 (bimolecular when formed by two distinct DNA strands or tetramolecular by four DNA strands, Figure 2C) or intramolecular G4 (or unimolecular G4, formed by one single DNA strand, Figure 2D). As it is harder to conceive a situation where two or four distinct DNA strands may come together to form an intermolecular G4 structure in vivo, we expect that unimolecular G4 structures are much more physiologically relevant within the cell. However, there is limited evidence for the occurrence of bimolecular G4, for example being physiological components of cellular nucleoli and contributing to certain phenotypes in ageing-derived diseases. In particular, the accumulation of unresolved intermolecular G4s might contribute to the formation of the pathological aggregates that are typical hallmarks of neurological disorders, acting as nucleation points for condensation (Liano et al., 2022).

Intramolecular G4 can adopt various topologies depending on the sequence composition, number of G-tetrads, loop length, ionic environment and temperature. All these factors can affect also its stability. Guanine bases in each quartet can adopt syn or anti glycosidic bond angles. (Figure 2B). According to the orientation of the G-strands one respect to the other and to the loop nature, G4 can be divided into parallel, antiparallel or hybrid structures. Depending on the conformation, grooves between G-tracts are of variable size. Moreover, loops can adopt lateral (edge-wise loop), diagonal or propeller (double-chain-reversal) orientation (Figure 2D).

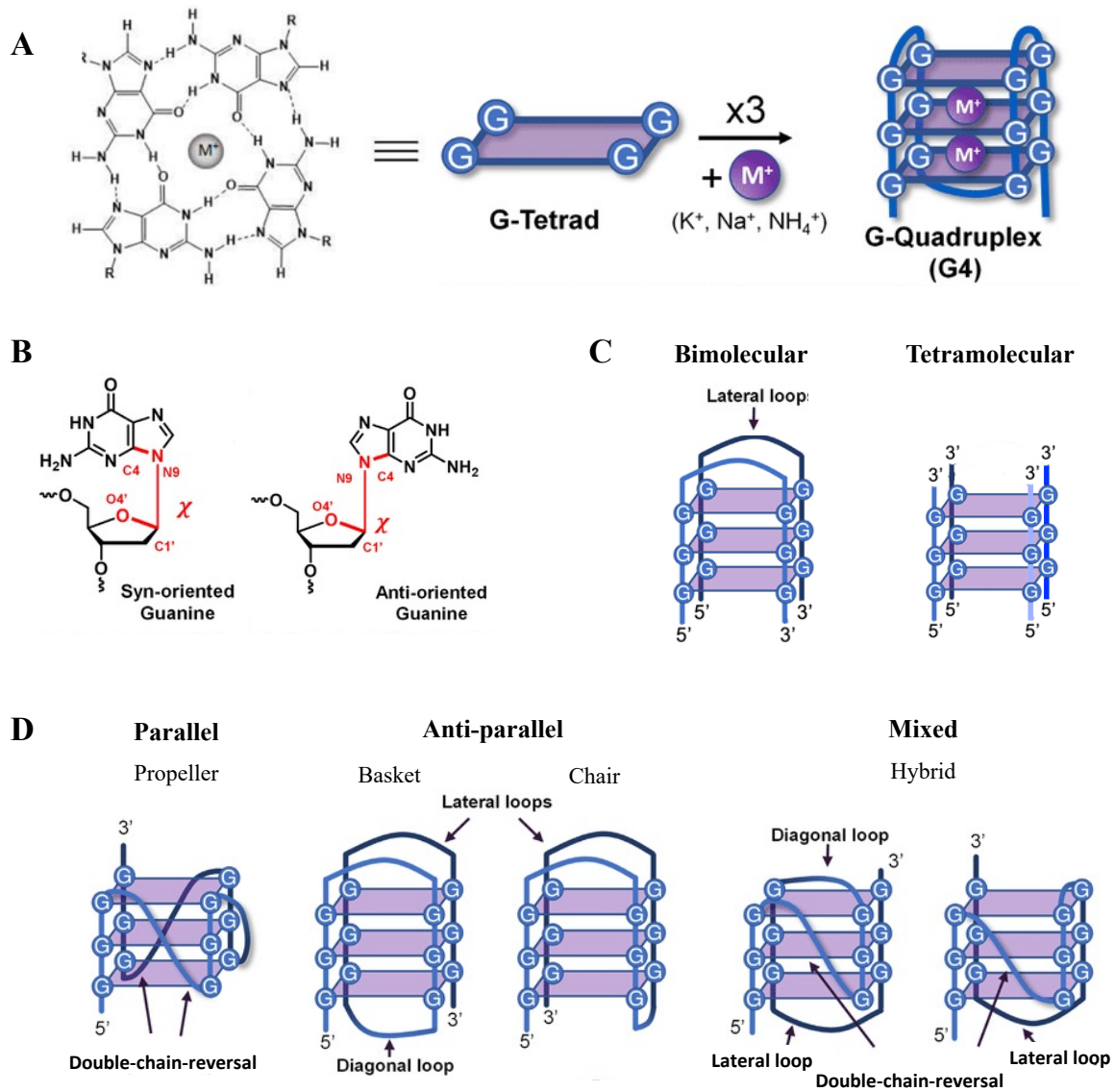


Figure 2. G-quadruplex polymorphism. **A.** Guanines arrangement into a G-tetrad structure, stabilized by a monovalent cation (M^+), and G-quadruplex formation by stacking of three tetrads. Hoogsteen bonds are represented as dotted lines. **B.** Illustration of syn- and anti-oriented guanines, the dihedral angle is highlighted in red. **C.** Intermolecular G4 topology representations. **D.** Intramolecular G4 topology representations with G4 loop types. (Adapted from Cadoni et al., 2021).

Many bioinformatic tools and computational algorithms have been developed to predict the potential formation of quadruplexes in the genome of several organism directly from DNA or RNA sequences, following a consensus motif: $G_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}N_xG_{\geq 3}$ where G are guanines and N represent any residue (between 1 and 7). These studies predicted over 376000 sequences with the potential to form G4 structures in the human genome. (Puig Lombardi & Londoño-Vallejo, 2020) A combination of biophysical methods and different approaches are needed to address whether the sequence is forming

a G-quadruplex in vitro and to properly characterize it (Luo et al., 2023). Moreover, functional biology assay can assess the role of G4s in many fundamental functions for cell maintenance.

2.1. G-quadruplexes biological functions

G-quadruplexes can be found in key regions that regulate genome function, such as DNA replication origins, promoters, gene transcriptional regulatory regions and telomere, and there is growing evidence that they play an important role in a range of biological processes (Figure 3).

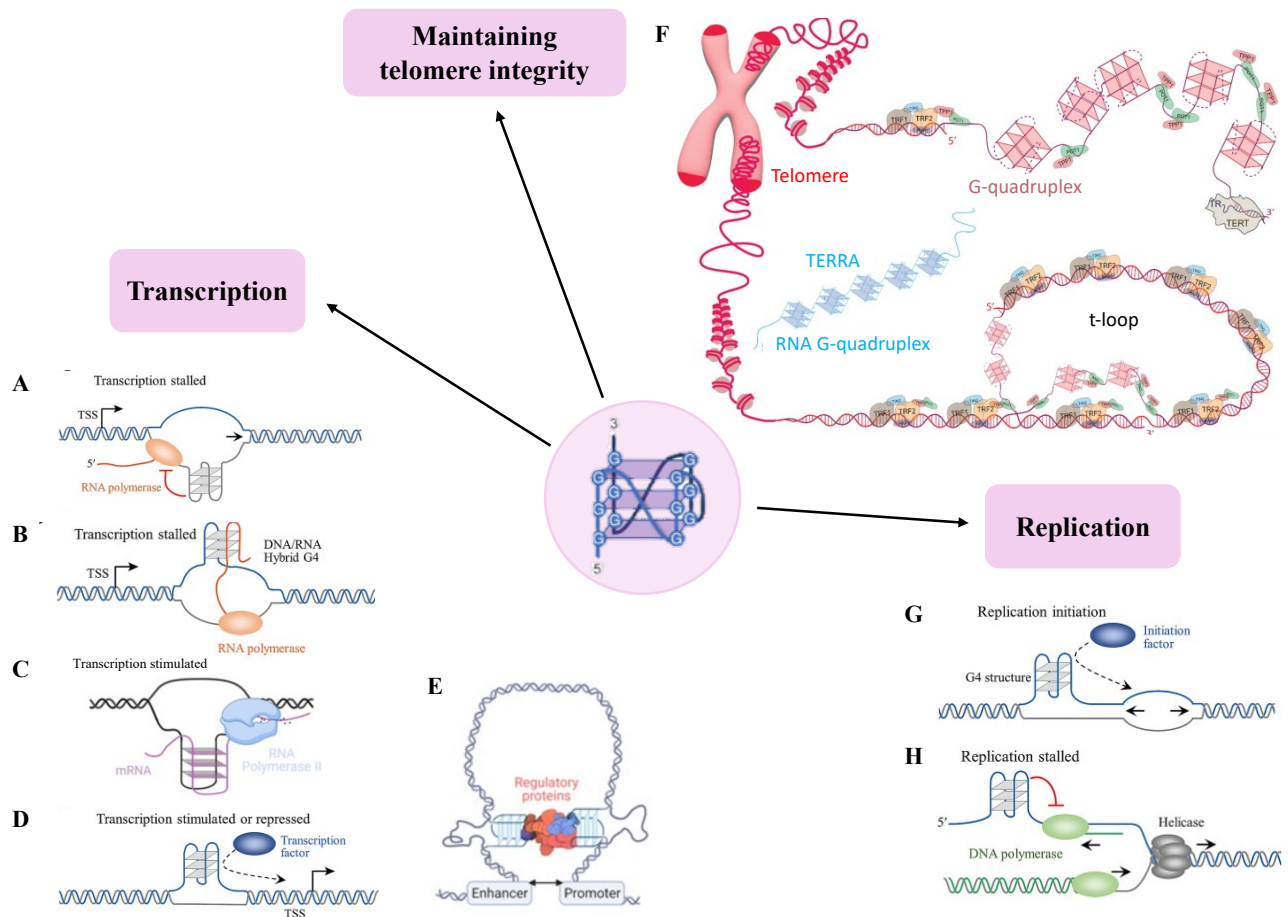


Figure 3. The putative biological function of G-quadruplex in DNA transcription (A, B, C, D, E), replication (G, H) and telomere maintenance (F) (Adapted from Teng et al. 2021, Bettin et al, 2019 and Robinson et al., 2021). TSS = transcription start sites.

2.1.1. G-quadruplexes in DNA replication

Many studies have demonstrated that the formation of G4 structures is stimulated in human cells during the S-phase (Murat & Balasubramanian, 2014). G4 structures may have a dual role in DNA

replication, acting as both a challenge to be resolved for fork progression and a regulatory element in the replication process (Figure 3G,H; Teng et al., 2021).

Some studies have led to the hypothesis that G4 DNA structures may somehow be involved in the mechanism of replication initiation. The human replication origin has been predicted to contain G4-motif (Besnard et al., 2012), which can be specifically bound by replication initiation factors, thus helping in recruiting them to the replication origin sites (Figure 3G). For example, the human origin recognition complex has been shown to bind G4-forming sequences in vitro (Hoshina et al., 2013). On the other hand, stable G4 structures can pose an impediment for DNA polymerase progression, leading to replication stalling, DNA-damage and genomic and epigenetic instability, particularly in the absence of specialized helicases (Castillo Bosch et al., 2014) or other factors (Schiavone et al., 2014) that resolve G4 structures in DNA (Hänsel-Hertsch et al., 2017) (Figure 3H).

2.1.2. G-quadruplexes in transcription

As transcription starts, transcription bubble generates transiently exposed single-strand segments with the potential to form G-quadruplexes. G4-forming sequences are indeed enriched in the promoters and both upstream and downstream of other regulatory regions of actively transcriptional start sites compared with the rest of the genome. (Huppert & Balasubramanian, 2007). Their abundance in these key regions in diverse vertebrates highlights their likely conserved role in transcriptional regulation. G4 structures could act on multiple levels during transcription, acting as transcriptional roadblocks and repressors, but also as transcriptional activators (Figure 3A,B,C;D,E; Agarwal et al., 2014; Robinson et al., 2021). These structures are thought to influence transcription by modulating the accessibility of the RNA polymerase complex. During DNA transcription, the presence of G4 on the DNA template strand blocks the progression of RNA polymerase (Fig. 3A), causing a transcriptional arrest or termination, directly decreasing the expression of the target genes (Smestad & Maher, 2015). G4 has been reported to form on the DNA non-template strand forming a stable DNA:RNA hybrid intermolecular G4 with the nascent RNA, which can block the transcription process (Figure 3B; Belotserkovskii et al., 2017). This intermolecular DNA:RNA hybrid G4s can also significantly form in the DNA template strand in actively transcribed genes (Figure 3C), where it may stabilize R-loops formed on the template strand, allowing for enhanced transcription (Robinson et al., 2021).

The formation of G4 upstream of the transcriptional start sites may both stimulate or repress transcription (Figure 3D). Interestingly, G4s are abundant in the promoters of proto-oncogenes such as *c-MYC*, *KRAS*, and *BCL2*, where their formation can block the binding of transcriptional activators, resulting in transcriptional repression (Balasubramanian et al., 2011). Conversely, G4 structures may also promote transcription by stabilizing open promoter conformations (Huppert & Balasubramanian,

2007). G4 structures may not only be involved in proximal transcriptional control at promoters, but its formation can promote long-range DNA interactions between complexes of gene regulation sites (Figure 3E; Robinson et al., 2021). Moreover, G4 formation may regulate gene transcription promoting transcriptional activation through the modulation of chromatin structure. G4 may engage key regulators of chromatin architecture, leading to global changes in chromatin structure. Further studies are needed to assess this direct correlation (Robinson et al., 2021).

2.1.3. G-quadruplexes at telomeres

Telomeres are repetitive DNA sequences made up of multiple G-rich repeats found at the ends of the chromosome. Telomeric DNA in humans usually contains a double-stranded region of several kilobases in length, formed by the TTAGGG sequence repeated from 300 to 8000 times, as well as a 3'-tail single-stranded repeat region with 10–50 TTAGGG repeats. Telomeres and their single-stranded overhangs gradually shorten with successive cell divisions, as part of the natural aging process, but can be elongated by telomerase, a reverse transcriptase that extends the telomeric repeat using its integral telomerase RNA as a template. This enzyme is activated in the majority of cancers, where is essential in preventing replication-dependent telomere loss in highly proliferative and cancer cells (Gruber et al., 2021). A telomerase-independent mechanism, the Alternative Lengthening of Telomeres (ALT) pathway, is used by some cells to maintain their telomere length, enabling them to evade the typical limits on cell division, contributing to cellular immortality in some cancers. It relies on homologous recombination and DNA repair mechanisms, in which telomeres from different chromosomes exchange genetic material or use DNA from other telomeric regions as a template to elongate their own (Zhang & Zou, 2020).

The repetitive sequence, rich in Gs, is particularly prone to forming G-quadruplex structures, which can regulate telomere metabolism (Rhodes & Lipps, 2015), and can also facilitate T-loop formation as it provides a large number of potential target sites that can be invaded by the telomeric overhang (Griffith et al., 1999). T-loops are structures in which the 3' overhang invades an upstream double-stranded region of the telomere, forming a loop of several kilobases of DNA, thereby protecting the otherwise vulnerable G-rich single-stranded telomere (Figure 3F). T-loops would need to be resolved during S-phase for telomere replication to occur; it is possible that during this part of the cell cycle, G-quadruplexes represent an alternative telomere-capping solution to the T-loop, preventing the degradation of telomeric ends by nucleases (Bryan, 2020). (Figure 3F). Telomere Repeat-containing RNA, or TERRA, also contributes to end-capping structure and function via regulation of telomere length and maintenance of genome stability. It is thought to accumulate

preferentially at short telomeres and can associate with the G-rich telomere strand displaced and forming G4 DNA structures (Figure 3F; Bettin et al., 2019).

The persistent presence of G4s at telomeres during DNA replication is problematic. Depletion of many proteins known to interact with telomeric G4 oligonucleotides results in telomere shortening, altered telomere replication rate or formation of fragile telomeres (Varshney et al., 2020). Moreover, G4 formation can control access to telomeres of telomerase based on its topology: antiparallel and hybrid G4 can inhibit telomerase elongation limiting its access to the 3' end of telomere, whilst parallel G4 can be extended due to partial G4 resolution by telomerase in vitro (Paudel et al., 2020).

2.2. Displacement loops and the interplay between G-quadruplexes and R-loops

Displacement loops (D-loops) and R-loops are additional key intermediate structures in DNA metabolism, including replication, repair, recombination, and transcription.

A D-loop forms when a ssDNA invades a homologous double-stranded DNA region, displacing one of the original strands and pairing with the complementary sequence. This structure is essential in homologous recombination (HR), a process that repairs double-strand breaks (DSBs) in DNA, by enabling accurate genetic information exchange between homologous sequences. D-loops serve as a platform for DNA synthesis, allowing the invading strand to use the undamaged homologous DNA as a template for repair, ensuring the faithful repair of DSBs (Piazza et al., 2019). The proper formation and resolution of D-loops are essential for maintaining genomic stability, as aberrant D-loop structures can lead to mutations or chromosomal rearrangements (San Filippo et al., 2008).

R-loops are three-stranded nucleic acid structures that are formed co-transcriptionally when an RNA strand invades a DNA duplex by displacing one DNA strand and hybridizing with the complementary one. They are abundant (5% of the human genome) and play a role in cellular metabolism including chromosome segregation (Kabeche et al., 2018), gene expression control (Nierhs and Luke, 2020), transcription termination (Skourti-Stathaki et al., 2011) and mitochondrial replication (Pohjoismaki et al., 2010). R-loops presence needs to be tightly controlled as they can be harmful to genome integrity, by causing the formation of both single and double-strand breaks, as well as stalling or collapsing of replication forks (Sollier & Cimprich, 2015; Allison & Wang, 2019). Deregulation of R-loops increases cell sensitivity to DNA damaging agents, formation of double-strand breaks, chromosome loss and recombination events, thus enhancing genome instability, a hallmark of cancer. Indeed, the overabundance of R-loops has been reported in a number of syndromes, human neurological disorders, autoimmune diseases and cancer (García-Muse & Aguilera, 2019).

An intriguing connection between R-loops and G-quadruplexes has gained significant attention due to their overlapping roles in DNA metabolism and their collective contribution to genomic instability (De Magis et al., 2019; Miglietta et al., 2020; Wulfridge & Sarma, 2024). G4 structures can form on the displaced single strand of R-loops, forming a G-loop, depending on the high transcription rate and the negative supercoiling of the DNA template, in both physiological and pathological contexts (Duquette et al., 2004). Interestingly, the requirements for R-loop generation are also compatible with G4-formation, as RNA:DNA hybrids form most stably with C-rich regions on the template strand of DNA (Huppert, 2008). Thus, the creation of R-loops often displaces a single-stranded G-rich sequence which may fold into a G-quadruplex.

G4 structures have been shown to modulate the formation of R-loops, particularly in regions of high transcriptional activity, and stabilize them. The stabilization of R-loops can be achieved with the specific action of G4 ligands, which may stabilize preexisting G4 structure in the displaced strand of R-loops through direct binding, in turn stabilizing DNA:RNA hybrids and increasing its length. The spreading of R-loops at transcribed genes results in the accumulation of DNA DSBs, that can activate molecular pathways leading to either cell killing or genome instability (De Magis et al., 2019). During replication-transcription conflicts, there is a high propensity to form R-loops at stalled replication forks of common fragile sites, where many G4-forming sequences reside. G4-associated R-loop formation is a major driver of chromosome fragility within actively transcribed common fragile sites. Immunofluorescence microscopy showed that RTEL1 helicase can prevent the accumulation of G4-associated R-loops, which interfere with DNA replication fork progression, thus maintaining genome stability (Wu et al., 2020). Some helicases, such as BLM and DDX9, are known to actively unwind both G-quadruplexes and R-loops, helping to prevent their accumulation induced by reactive oxygen species (Tan et al., 2020) and contribute to unhindered transcription (Chakraborty & Grosse, 2011). A recent study demonstrated a positive effect of G4s on R-loops, allowing for enhanced transcription: G4s formed specifically on the non-template strand enhance R-loops formation resulting in significantly increased transcription, RNA polymerase initiation and elongation (Lee et al., 2020).

Understanding the cooperative effects of R-loops and G4s in DNA metabolism is essential for developing therapeutic strategies, especially in cancer, where these structures are frequently dysregulated.

2.3. G-quadruplexes as therapeutic target and tool

Improper folding, stabilization or unwinding of G4s can lead to severe dysfunction of fundamental cellular processes, resulting in pathological outcomes. Due to the significant impact of

G-quadruplexes on human diseases, especially in cancer, and their elevated polymorphism, significant efforts have been made to use G4s as therapeutic targets.

2.3.1. G-quadruplex ligands

The unique features of the G4 structure permits the design of structure-selective small molecules that can manipulate folded G4s. More than thousands of G4 ligands have been reported in the G-quadruplex Ligands Database (<http://www.g4ldb.com>, Davis et al., 2024) providing many opportunities for the developing of novel anticancer therapies.

The first efforts were focused on targeting G4-telomeric sequences with the aim to inhibit telomere extension by telomerase in cancer cells, blocking tumor growth and metastasis. BRACO-19 ligand is able to stabilize G4 at the ssDNA overhang of telomeres, preventing telomerases activity and leading to telomere shortening and end-to-end chromosomal fusions in cancer cells (Zhou et al., 2016). Another approach was to target G4-containing genes involved in cancer progression, suppressing directly the oncogenes expression, using G4-binding molecules able to stabilize G4 structures in the promoters (Awadasseid et al., 2021). In this context, the most successful target is *c-MYC* G4: the GQC-05 compound has been demonstrated to be a potent and selective stabilizer of the *c-MYC* G-quadruplex, inducing cytotoxicity and suppress its transcription in Burkitt's lymphoma (Brown et al., 2011).

G4 ligands have been proposed to have therapeutic effects in cancer patients when used at low dosages. In particular, they can be exploited to stimulate antitumor immunity and optimize immunotherapy efficacy in unresponsive tumors, acting on a subset of immune-related and autophagy-related pathways. The findings support the novel use of G4 ligands as enhancers of immune gene expression by harnessing G4-mediated replication and transcription stress in cancer cells at non-cytotoxic dosage, rather than for their cell killing activity as chemotherapeutic agents (Miglietta et al., 2022).

However, absolute specificity for a unique G4 is a hard challenge, due to the common core that all G4s share.

2.3.2. G-quadruplexes binding proteins

A number of proteins have evolved to bind and/or manipulate G4 structures (Mishra et al, 2016). Since G-quadruplex structures can act as rope knots in DNA replication but also interferes with the transcriptional and translational machinery, they need to be unwound to maintain proper genome and cellular functions. Among the proteins able to regulate these structures, particularly important are helicases. Helicases able to unwind G4 have been the focus of multiple studies, due to their connection

with many human diseases (Table 1). They are able to recognize and disrupt G4 structures in a regulated manner, differing in strand and target specificity, and sometimes show the ability to displace G4-stabilizing ligands. For examples, DHX36, BLM and WRN helicases show different G4 conformation specificity and their resolving activity is enough to displace a stably bound G4-ligand (Tippiana et al., 2016). Helicases can affect genome instability on many levels, targeting G4 structures during different biological processes. Further details on G4-interacting helicases will be provided later.

Name	Family	Polarity	Organism	Functions	Genetic diseases
Pif1	SF1 Pif1-like	5'-3'	<i>S. cerevisiae</i>	Telomere maintenance, DNA replication, mitochondrial genome maintenance, DS repair, epigenetic regulation	
Dna2	SF1 Upf1-like	5'-3'	<i>S. cerevisiae</i>	Telomere maintenance, DNA replication	
XPD	SF2 Rad3/XPD	5'-3'	<i>H. sapiens</i>	Nucleotide excision repair, transcription regulation	Cockayne syndrome, COFS syndrome, Xeroderma pigmentosum
FANCD1	SF2 Rad3/XPD	5'-3'	<i>H. sapiens</i>	DNA replication, DNA crosslink repair, epigenetic regulation	Fanconi anemia, breast cancer
DinG	SF2 Rad3/XPD	5'-3'	<i>E. coli</i>	DNA repair, replication fork restart after DNA damage	
DDX11	SF2 DEAH	5'-3'	<i>H. sapiens</i>	DNA replication, sister chromatid cohesion, post-replicative repair	Warsaw Breakage Syndrome
RTEL1	SF2 DEAH	5'-3'	<i>H. sapiens</i>	Telomere maintenance, DNA replication, DNA transcription	Dyskeratosis congenita
RecQ	SF2 RecQ-like	3'-5'	<i>E. coli</i>	DNA repair, DNA replication, homologous recombination	
RecQ1	SF2 RecQ-like	3'-5'	<i>H. sapiens</i>	Telomere maintenance, DNA replication, DNA repair, homologous recombination,	RECQ1 One syndrome
BLM	SF2 RecQ-like	3'-5'	<i>H. sapiens</i>	Telomere maintenance, DNA replication, DNA transcription, DNA repair, homologous recombination	Bloom syndrome
WRN	SF2 RecQ-like	3'-5'	<i>H. sapiens</i>	Telomere maintenance, DNA replication, DNA transcription, DNA repair, homologous recombination	Werner syndrome
RecQ4	SF2 RecQ-like	3'-5'	<i>H. sapiens</i>	Telomere maintenance, DNA replication, DNA repair, homologous recombination	Rothmund-Thomson syndrome, RAPADILINO syndrome, Baller-Gerold syndrome
RecQ5	SF2 RecQ-like	3'-5'	<i>H. sapiens</i>	DNA replication, homologous recombination	
DHX36	SF5 DEAH	5'-3'	<i>H. sapiens</i>	Telomere maintenance, transcription regulation, translation regulation	

Table 1. List of some helicases and selected features, which have been described to regulate G4 structure in vivo. (Teng et al., 2021). The present work focuses on different G4-specific helicases, including human RecQ helicases and *E. coli* DinG helicases.

3. Helicases

Helicases are a class of enzymes vital to all organisms. The number of helicases expressed in higher organisms is strikingly high, with the human genome encoding for 95 distinct helicase proteins. They are involved in virtually all aspects of nucleic acid metabolism, including DNA replication, repair, recombination, transcription, translation, RNA splicing, mRNA transport, chromosome segregation and telomere maintenance.

Helicases are molecular motors that use the energy of ATP hydrolysis undergoing conformational changes: they bind to a single-stranded nucleic acid substrate (ssDNA or ssRNA), move to an ss/ds junction, and then separate the duplex (Patel & Donmez, 2006; Singleton et al., 2007). Whereas some enzymes unwind canonical nucleic acid structures (such as dsDNA), other are specialised to deal with non-canonical structures such as G-quadruplexes, D-loops, R-loops, Holliday junctions and triple helices. Some helicases also have the ability to catalyze the opposite reaction, reannealing complementary DNA strands (Wu, 2012).

Mutations in genes encoding DNA helicases have been observed in several inherited human diseases, linked with cancer and ageing, emphasizing the essentiality of DNA helicases in cellular metabolism (Suhasini & Brosh, 2013; Brosh, 2013).

Helicases can have a preference for DNA or RNA substrates (Pyle, 2008) and may differ in strand directionality, moving on the single-strand in the 3'-5' or 5'-3' direction. They can have very different translocation speed on nucleic acids strands and processivity, which is a measure of the average number of base pairs unwound by the helicase in a single binding event.

According to sequence similarity and the presence of common motifs in the catalytic core, helicases have been classified into six superfamilies (SF1-6). Superfamily 1 (SF1) and Superfamily 2 (SF2) helicases are predominant, generally monomeric or dimeric, whilst the other helicase superfamilies (from SF3 to SF6) form hexameric rings. They all share three motifs within the catalytic core, namely the Walker A, Walker B and Arginine finger motifs (Singleton et al., 2007).

3.1. Superfamily 1 and Superfamily 2 helicases (SF1 and SF2)

3.1.1. Helicase architecture

The core of the SF1/SF2 monomeric helicases is formed by two RecA-like domains (Figure 4A): in SF1, these domains are called 1A and 2A, while in SF2, they are named helicase domain 1 and 2 (HD1 and HD2). The ATP binding site placed at the interface of the two domains (Figure 4B). This helicase core is fundamental for enzyme activity, using the energy of ATP binding and hydrolysis to perform mechanical work. The ATPase cycle allows conformational changes with the two RecA-like

domains moving relative to one another, providing energy for the translocation and nucleic acid unwinding activities. Beside the core catalytic RecA domains, most helicases feature additional N- and C-terminal domains with specific roles. These domains influence or even define the function of a helicase, being critical for their specificity in both substrate binding selectivity, as well as protein-protein interactions.

The RecA-like domains include several characteristic sequence motifs (for example Q, I, Ia, Ib, II, III, IVa, V, Va and VI in SF2; Figure 4C) (Fairman-William et al., 2010). Motif I (Walker A motif) and Motif II (Walker B motif) are particularly important for binding and hydrolysis of the triphosphate; the subsequent hydrolysis of ATP to ADP and inorganic phosphate underpins the relative movement of the helicase domains. Motif III and V are implicated in coordinating ATP hydrolysis and nucleic acid binding, playing an essential role in transmitting the energy of ATP hydrolysis into motor function. Motif VI, also called the “arginine finger” due to an essential arginine residue, extends from HD2 into the HD1 core, enabling communication between the domains (Loham & Bjornson, 1996; Zhang & Wigley, 2008).

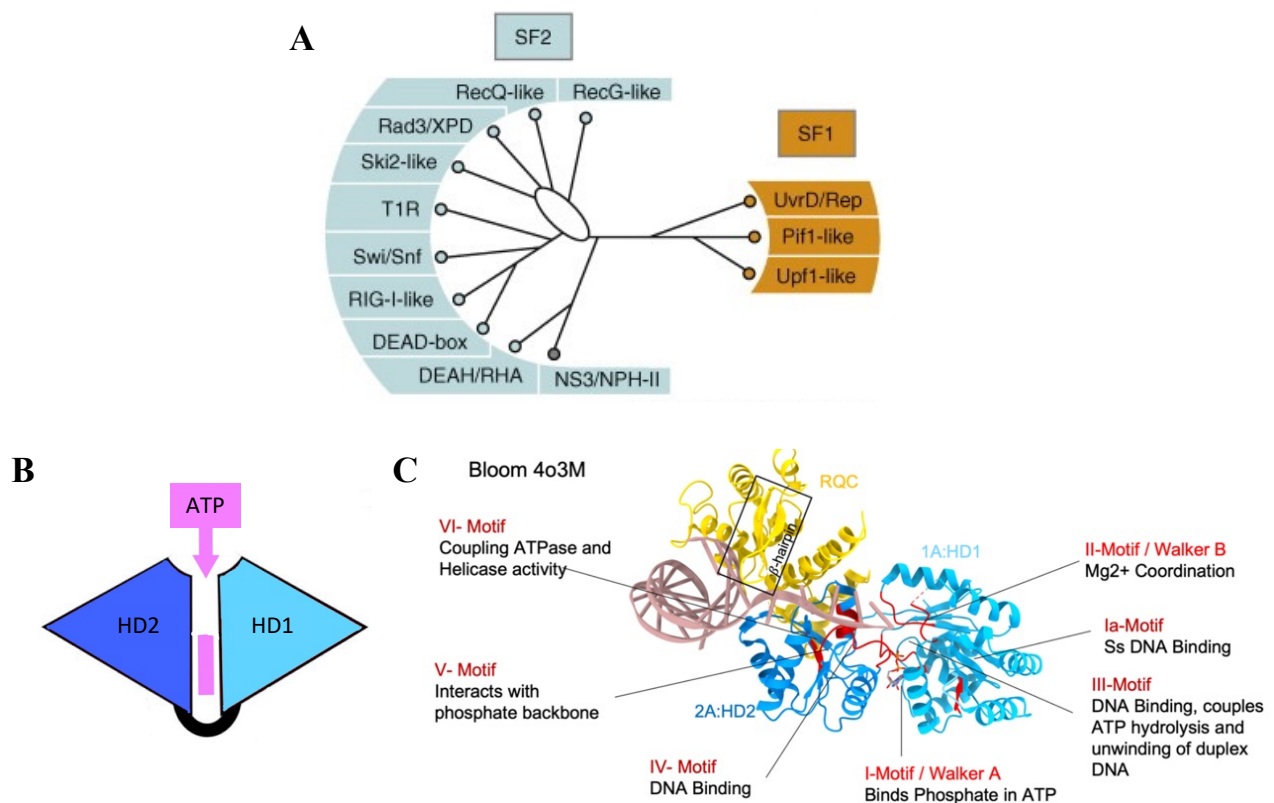


Figure 4. SF2 family helicase. **A.** Schematic representation of helicase families belonging to superfamily 1 (orange) and superfamily 2 (sky blue). (Adapted from Fairman-Williams et al., 2010). **B.** Schematic representation of the helicase catalytic core, showing the two RecA domains in two shades of blue with an ATP binding pocket in between (shown in magenta). **C.** Structural overview of BLM helicase (PDB:4O3M) as

representative of the SF1/SF2 architecture. Domains are color coded as follow: the two RecA domains as in B; the additional RQC domain which is unique to the RecQ family is in yellow; the canonical helicase motifs are highlighted in red, and the ATP analogue is show as ball-and-stick representation. The DNA forked substrate is shown in beige.

The helicases investigated in the present work all belong to Superfamily 2, which is the largest and most diversified of the helicase superfamilies, and includes a number of distinct families, based on sequence homology (e.g. RecQ-like, RecG-like, Rad3/XPD, Ski2-like, DEAH/RHA, DEAD-box families; Figure 4A).

3.1.2. Translocation mechanism

The translocation mechanism along a nucleic acid used by SF1 and SF2 helicases was first proposed for bacterial PcrA helicase, which belongs to SF1. Velankar et al. (1999) described an inchworm mechanism in which the binding of ATP initiates a relative movement of RecA-like domains 1A and 2A (which are equivalent to HD1 and HD2 in the SF2 family) and at the same time regulates the relative affinity of the two domains for the single-stranded DNA substrate. In the ATP-bound state, domain 1A releases the ssDNA while domain 2A binds to it. ATP hydrolysis leads to movement of the two RecA domains, coordinating binding and releasing of nucleic acid by each RecA-like domain, resulting in translocation of PcrA along the nucleic acid (Figure 5) (Velankar et al., 1999).



Figure 5. Scheme illustration of inchworm translocation mechanism. (From Singleton et al., 2007).

3.2. G-quadruplex unwinding helicases

To date, several helicases from different organism have been described to regulate non canonical DNA structure, in particular G4 formation and unwinding (Table 1). Although G4 structures can be beneficial in some contexts, it is known that their formation can be detrimental to certain cellular processes (Mendoza et al., 2016). In this section I will provide some examples of helicases acting of G4 structures, and their *in vivo* implications (Figure 6).

Among the helicases shown to affect G4 unfolding at telomeres, important examples are the human RecQ helicase family members Bloom (BLM) and Werner (WRN), the RTEL1 (*Regulator of telomere length 1*) helicase, a Fe-S helicase, and DHX36, an RNA helicase belonging to the DEAH box family. In WRN-deficient cells, the (TTAGGG)₆ sequence is not properly duplicated resulting in large deletions during the replication of human telomeric sequence (Damerla et al., 2012). *In vivo*, the loss of BLM activity leads to a significant increase in G4 structures at telomeres, accompanied by changes in replication speed (Drosopoulos et al., 2015). In the absence of RTEL1, studies revealed that telomeres are short and become fragile (Le Guen et al., 2013). In DHX36-knockdown HEK293T cells, telomeres are not elongated due to a downregulation of telomerase activity, suggesting that DHX36 unwinds the G4s within telomerase RNA regulating its function (Booy et al., 2015).

A few helicases have a role in modelling G4 during DNA replication and genome stability, such as *Fanconi anemia group J helicase* (FANCI), Pif1 helicase and the RecQ helicases. Loss of FANCI function in human patient cell line EUFA0030, where the protein is mutated, correlates to large genomic deletions near to G4 motifs (London et al., 2008). Consistently, defects in the *C. elegans* FANCI ortholog dog-1, can cause deletion of G4 DNA and genome instability (Kruisselbrink et al., 2008). In the absence of the yeast Pif1 helicase (ScPif1), G-rich sequences are genetically unstable in replicating yeast cells, leading to a significant increase in gross chromosomal rearrangements (Ribeyre et al., 2009). Another study highlighted the formation of DNA breakage at the G4 sites, in Pif1-deficient mutants, due to the stalling of the DNA replication fork (Paeschke et al., 2011).

Helicases may influence transcription by unfolding G4 structures. XPD, a member of the Rad3/XPD family helicases and a component of the TFIIH transcription factor, is enriched near the transcription start site of numerous highly transcribed genes, in particular where G4 motif are present. By unwinding local G4 structures, it may play a role in transcribing genes linked to specific signaling pathways and regulatory pathways associated with specific cancers (Gray et al., 2014).

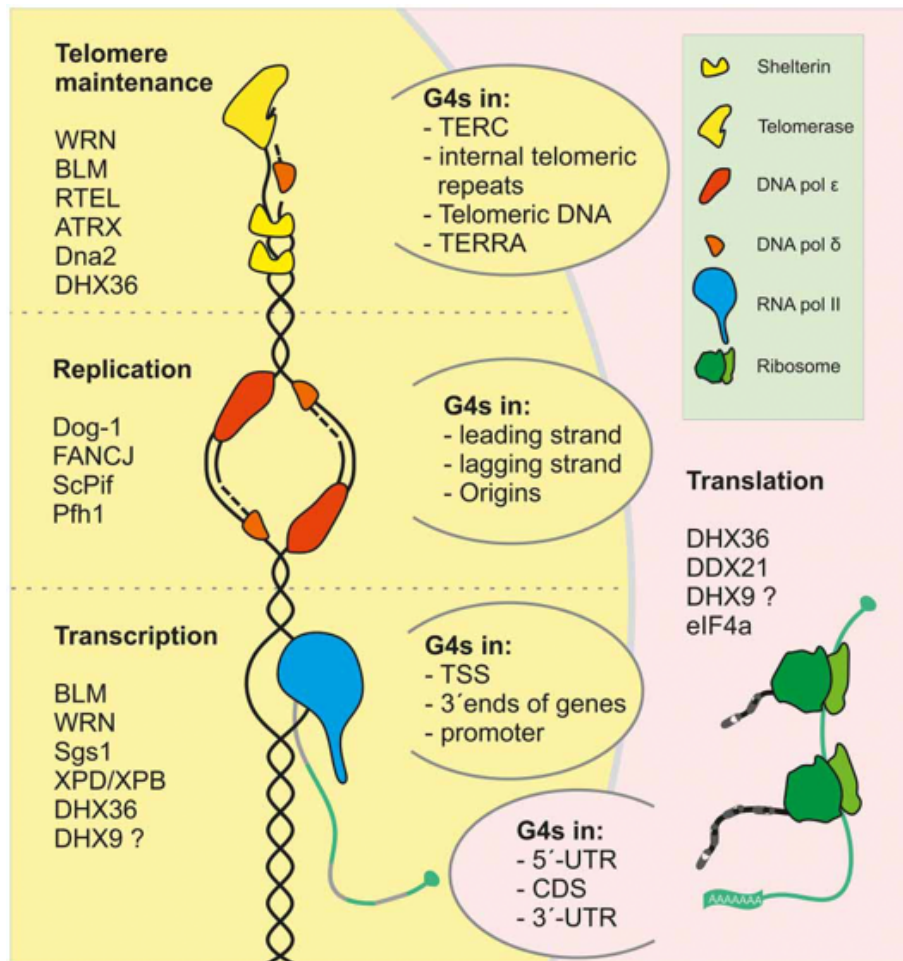


Figure 6. Involvement of G-quadruplex helicases in cellular processes in vivo. G4 related-cellular processes localized in the nucleus (yellow background) or in the cytosol (red background). G4 structure-resolving helicases are listed according to their contribution to each process. Locations of potential G4-forming events are listed in corresponding bubbles. (From Sauer & Paeschke, 2017)

3.3. Crystal structures of helicases bound to G4

Structural information on the interaction between G4 and protein scarce. In particular, data on helicases and G4s are limited to the structures of the *Bos taurus* DHX36 helicase bound to cMYC G4 (PDB ID: 5VHE, Chen et al., 2018), the *Cronobacter sakazakii* RecQ (CsRecQ) bound to a resolved G4 (PDB ID: 6CRM, Voter et al., 2018), *Thermus oshimai* Pif1 (ToPif1) complexed with a G4 with both 5' and 3' tails (PDB ID: 7OAR, Dai et al., 2022) and to the most recent *Saccharomyces cerevisiae* Pif1 (ScPif1) bound to a G4 DNA with a 5' single-stranded DNA (ssDNA) segment (PDB ID: 8XAK, Hong et al., 2024).

3.3.1. Co-crystal structure of DHX36 bound to cMYC G4

The co-crystal structure of DHX36 bound to the cMYC G4 with a 3'-ssDNA segment has for many years provided the only structural information on the interaction between a helicase and a G4 substrate. The structure shows the folded G4 structure recognized by an N-terminal DHX36-specific motif (DSM, blue helix in Figure 6) that fold into an α -helix after the binding with the G4, creating a nonplanar hydrophobic surface that stacks on the top of the 5' face of the G4. Moreover, the C-terminal OB-fold domain (in violet in Figure 7) create contacts with both the G4 structure and the adjacent ssDNA, participating to the unfolding of the G4. The ssDNA 3' tail, which fits in a single-stranded binding channel formed by the RecA1, RecA2 and the C-terminal domain, is pulled by the helicase causing the G4 unwinding one residue at a time (Chen et al., 2018).

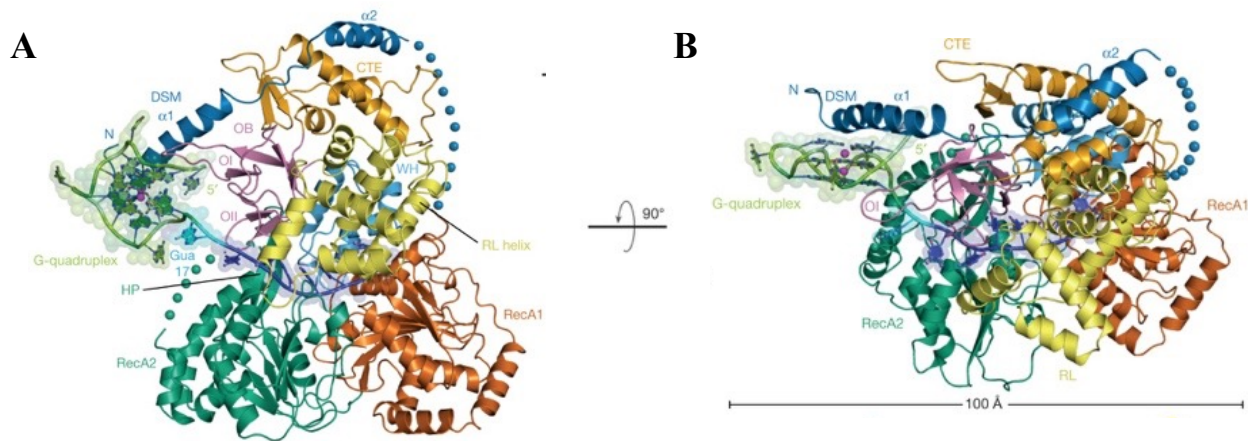


Figure 7. Overall structure of the DHX36–G-quadruplex DNA complex. **A.** Cartoon representation of the co-crystal structure of DHX36 bound to *cMYC* G4; colour-coded: helicase domains RecA1 and RecA2 in orange and green respectively, WH domain in light blue, DSM domain in blue, OB-fold domain in violet, RL domain in yellow; G4 structure in light green. **B.** Cartoon representation as in A, rotated by 90°. (From Chen et al., 2018).

3.3.2. Co-crystal structure of CsRecQ bound to a resolved G4 DNA

When the bacterial CsRecQ was co-crystallised in the presence of a G4 DNA, surprisingly the crystal structure showed the presence of a resolved G4, despite the absence of ATP (Voter et al., 2018), providing some insights into the possible mechanism with which other G4-resolving helicases, and in particular RecQ helicases, may unwind G4 structures.

Voter et al. (2018) thus suggested a guanine-flipping and sequestration mechanism for G4 unwinding by the helicase, based on the presence of a guanine specific binding pocket (GSP) on the surface of RecQ. The 3'-most guanine base of the unfolded G4 (G21) is sequestered within this pocket thanks

to the formation of hydrogen bonds with the RecQ residues Ser254 and Asp312 (Figure 8A). Other two residues of GSP, Lys248 and Trp347, capped the pocket at the 5' end and 3' end, respectively (Figure 8B). Comparisons with the DNA-free and duplex DNA-bound bacterial RecQ structures (Bernstein et al., 2003; Manthei et al., 2015), revealed that access to the GSP is occluded by interactions between Lys248 with Asp312, making GSP and all the hydrogen contacts uniquely selective for guanines. Alterations in the GSP blocked the unwinding of G4 DNA, maintaining the ability to bind to and unwind duplex DNA, highlighting its essential role in G4 resolution. In the model proposed, the ssDNA of the G4 is firstly recognized by the helicase, placing the G4 structure near to the GSP. The 3'-most guanine of the G4 flips from the folded structure and occupies the pocket, destabilizing the G4 structure. Lastly, the hydrolysis of ATP would permit the translocation of the base along the helicase domain allowing G4 unwinding (Voter et al., 2018).

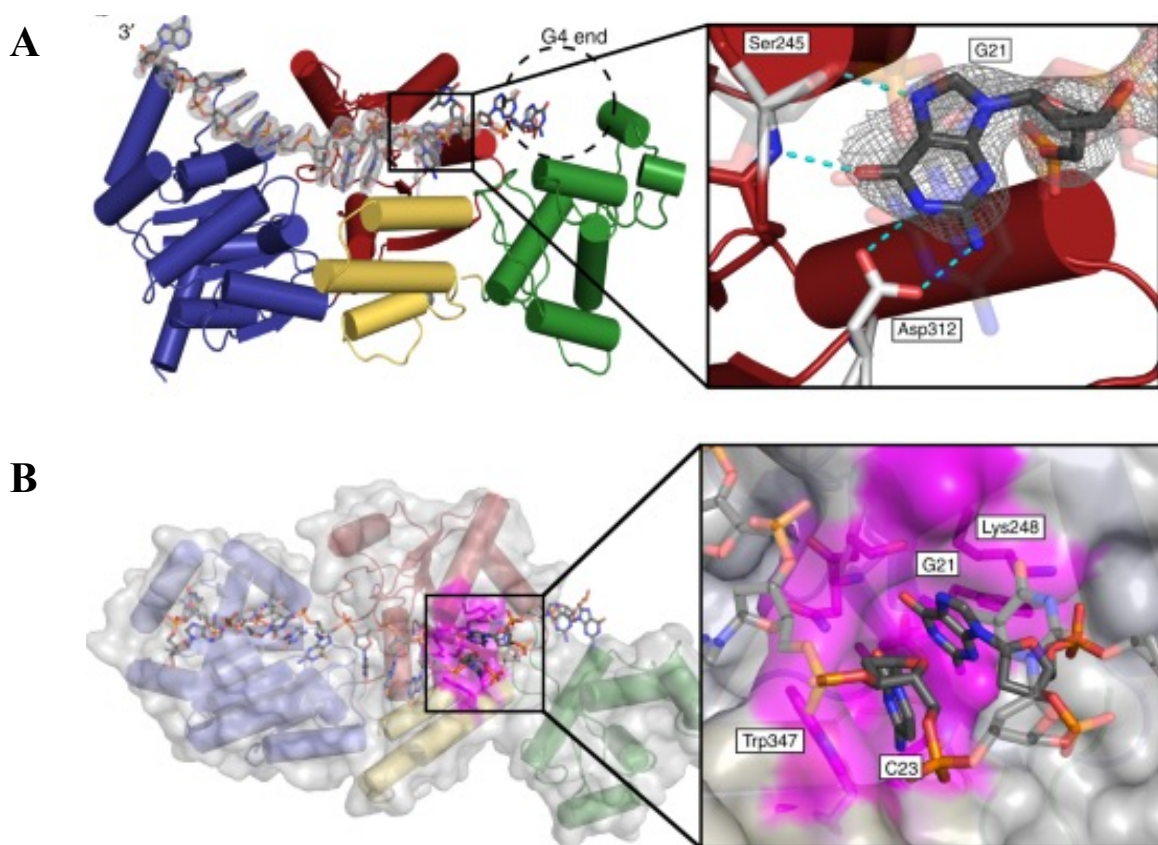


Figure 8. The guanine-specific binding pocket of the CsRecQ helicase. **A.** Crystal structure of CsRecQ bound to resolved G4 DNA. The helicase domains are colored in blue and red, the C-terminal domain (RQC) in green and the Zn²⁺-binding domain in yellow. The expected location of the G4 is highlighted. (Insert) Hydrogen bonds are represented by dashed lines. **B.** Surface representation of the CsRecQ bound to the resolved G4 with the GSP colored in magenta. (Insert). (From Voter et al., 2018).

3.3.3. Co-crystal structure of ToPif1 bound to G4

The crystal structure of ToPif1 bound to a G4 is the first G4-helicase structure among SF1 helicases (Dai et al., 2022). However, the structure shows one G4 sandwiched between two ToPif1 molecules (molecule a and molecule b in Figure 9), with each ToPif1 molecule mostly interacting with the flanking ssDNA regions (Figure 9). This structure therefore does not provide many insights on the structural determinant of helicase:G4 recognition. Indeed, when this structure is compared to the most recent structure of yeast Pif1 bound to a G4 (where the tetrads make significant interactions with the protein – see next section) no similarity to the mode of G4 binding can be detected.

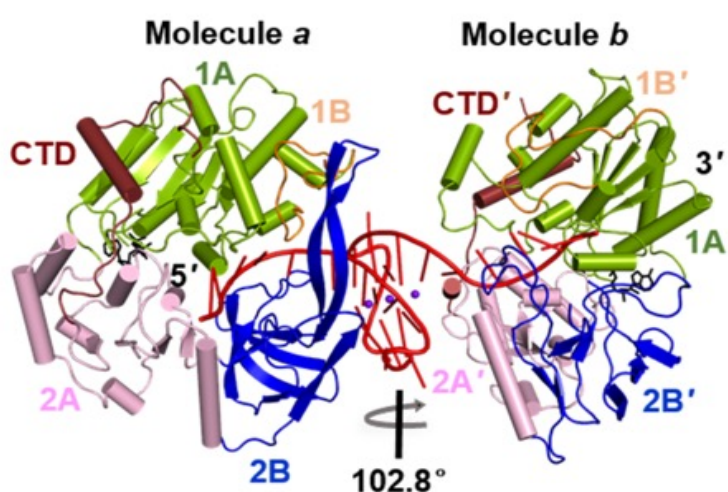


Figure 9. Structure of ToPif1 complexed with G4 DNA and ADP·AlF4 in cartoon mode. Color code: conserved domains 1A in green, 1B in orange, 2A in pink, 2B in blue and C-terminal (CTD) in brown; G4 DNA in red and ADP·AlF4 as black stick. ToPif1 molecules binding at the 5'- and 3'-ends of G4 DNA are labeled as molecule a and molecule b, respectively. (From Dai et al., 2022)

3.3.4. Co-crystal structure of ScPif1 bound to G4

The co-crystal structure of ScPif1 bound to a G4 DNA with a 5'-ssDNA segment revealed a different mode of G4 recognition compared to the previous resolved crystal structure of ToPif1 helicase (Dai et al., 2022). ScPif1 interact with and unfold in a sequential manner the G4 structure mainly through the conserved wedge region in the 1A domain, helped by additional but weak interactions of 2B domain with the peripheral region of the G4. Residue R324 present in the wedge region play a key role in connecting ScPif1 to the G4 structure, interacting with G12 in the 5'-most G-tetrad (Figure 10). As a demonstration, mutations in this residue significantly reduces G4 unfolding activity. Hong et al. described G4 unwinding mechanism, in which the binding of the G4 DNA causes a rotational

conformational change of 2B domain. At this point, R324, anchoring the G12, pushes the first guanine of the 5'-most tetrad into the ssDNA binding region thanks to ATP hydrolysis-dependent translocation (Hong et al., 2024).

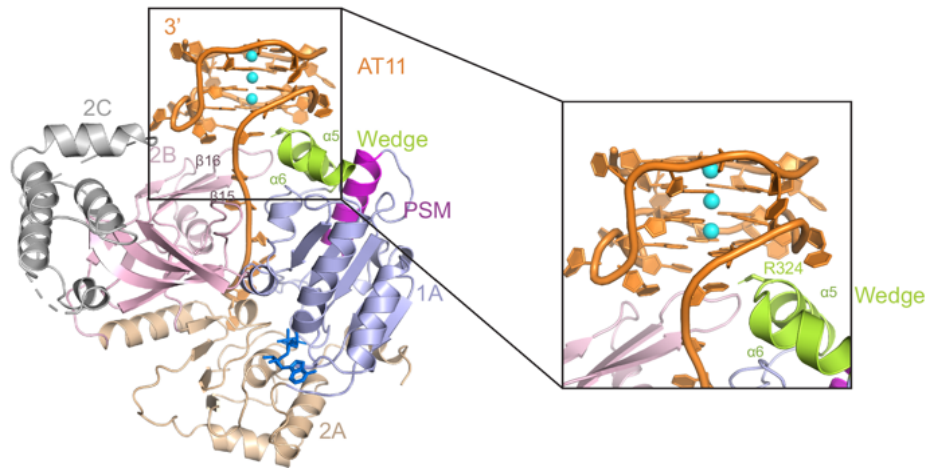


Figure 10. Cartoon representation of the crystal structure of ScPif1 bound to G4. Color code of the conserved domains and motifs: 1A in light blue, 2A in wheat, 2B in light pink, 2C in gray, wedge in limon, and Pif1 signature motif (PSM) in magenta. The AT11 G4 DNA is colored in orange. The key residue R324 involved in the interaction is shown as sticks. (From Hong et al., 2024).

Although the characteristics of the catalytic domain are almost shared among helicases, they unfold G-quadruplex with a specific directionality, they show a selectivity for specific G4 topologies and they act on different pathways. Thus, there is the need of a more in-depth analysis to better elucidate the variety of mechanisms existing among helicases that are used to resolve G4s.

4. Human RecQ helicases

RecQ helicases belong to a ubiquitous family of DNA unwinding enzymes and have emerged as essential to guarantee proper genome stability, by performing a broad variety of functions at the interface of DNA replication, recombination and repair as well as telomere stability. The family of RecQ helicases is named after the *Escherichia coli* recQ gene, which plays a role in recombination and recovering from replication fork stalling at sites of UV-induced DNA damage (Nakayama et al., 1984).

RecQ proteins are DNA helicases belonging to superfamily SF2, and unwind a variety of DNA structures in the 3'–5' direction, including duplex DNA, D-Loop, R-Loop, DNA junctions and G-quadruplexes.

RecQ proteins are highly conserved from bacteria to man. Unicellular organisms typically express a single RecQ-like DNA helicase, such as *E. coli* RecQ in bacteria and Sgs1 in budding yeast. In contrast, human cells contain five homologs: RecQ1, Bloom (BLM), Werner (WRN), RecQ4 and RecQ5, four of which are associated with rare genetic diseases. While RecQ5 is not linked to any human hereditary syndromes, defects in human RecQ1, BLM, WRN and RecQ4 cause rare recessive autosomal diseases: RECON (RECql One) syndrome (RecQ1), Bloom syndrome (BLM), Werner syndrome (WRN), and Rothmund–Thomson/RAPADILINO/Baller–Gerold syndromes (RecQ4). They share common characteristics of genetic instability, skin related problems and cellular senescence, however each syndrome has its own clinical symptoms and developmental patterns. All human RecQ helicases alterations lead in the end to cancer predisposition, even if their cancer profiles are different (Mojumdar, 2020). This suggests that human RecQ helicases have distinct, but partially overlapping, specialized biological functions in the cells (Croteau et al., 2014).

4.1. RecQ domain organization

The average fold of RecQ helicases generally includes three conserved domains: the core helicase domain, the RecQ C-terminal (RQC) domain and the helicase and RNase D-like C-terminal (HRDC) domain. However, a number of paralogs miss one or more of these domains. Additional C- and N-terminal extension regions confer distinct functions or characteristics to each homolog, such as specific protein-protein interactions, distinct nucleic acid binding properties, regulation of nuclear or mitochondrial localization, enzyme oligomerization, or additional enzymatic activities (Figure 11).

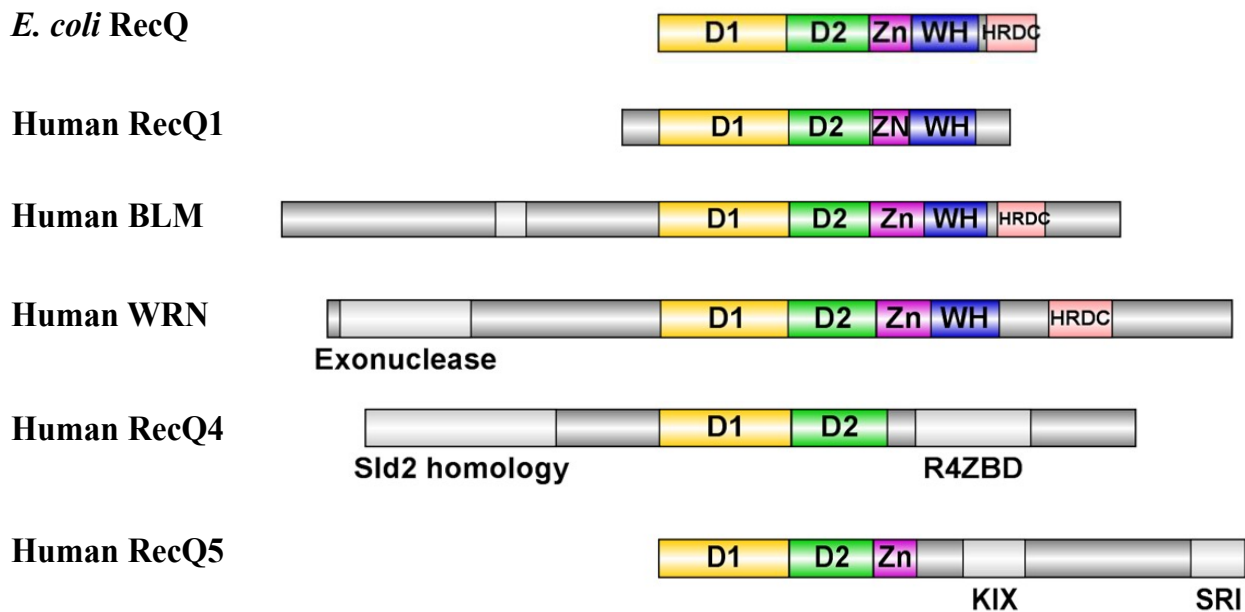


Figure 11. Domain organization of select bacterial and human RecQ helicases. Domains color-coded: helicase core domain D1 (HD1) and D2 (HD2) in yellow and green respectively; the RQC domain, formed by the Zn²⁺-binding domain in violet and the WH domain in purple; the HRDC domain, which is not present in all RecQ family helicases, in pink; addition domains in shadows of grey. (From Newman & Gileadi, 2020)

The structural understanding of the human RecQ helicase family has increased massively thanks to the crystal structures of the helicase core of all human paralogs (Figure 12). In addition, structures of RecQ1 (Pike et al., 2009; Pike et al., 2015), BLM (Swan et al., 2014; Newman et al., 2015) as well as the isolated WRN WH domain (Kitano et al., 2010; Newman et al., 2020) in complex with DNA substrates have been determined, revealing a common helicase mechanism. The structural information on RecQ helicases has been reviewed by Newman & Gileadi (2020).

The helicase core domain features the common fold of the catalytic core of SF2 helicases (Figure 4C), formed by two adjacent RecA-like domains (HD1 and HD2, in yellow and green in Figure 11 and Figure 12, respectively), including all the canonical SF2 motifs. In addition, the HD1 of all RecQ helicases features a conserved aromatic-rich sequence motif, the aromatic-rich loop (ARL), located between helicase motifs II and III, that act as a ssDNA sensor element, coupling DNA-binding to structural rearrangements in HD1.

The RQC domain is the second most conserved domain and is unique to the RecQ family helicases. It is located downstream of the helicase domain and together they constitute the functional core unit for helicase activity. It consists of a helical Zn²⁺-binding domain and a winged-helix (WH) domain (in violet and purple, respectively, in Figure 11). The zinc-binding domain, where four cysteine

residues coordinate a single zinc ion, is found in all human RecQ helicases and is important for maintaining the structural integrity of the protein. It may be involved in nucleic acid interactions and acting as surfaces that mediate protein-protein interactions (Guo et al., 2005; Ren et al., 2008). The WH domain shows a helix-turn-helix fold and is missing in RecQ4 and RecQ5. It is characterized by the presence of a β -hairpin motif with an aromatic residue at the tip, which is important for DNA unwinding at least in eukaryotic RecQ helicases, since it is used as a wedge to split DNA duplex. It plays a role in protein-DNA interaction making extensive contacts with the junction between double and single-stranded regions, as well as protein-protein interactions (Lucic et al., 2011). Moreover, RQC domain has been suggested to have a role in binding to non-canonical substrates, such as G-quadruplex structures in BLM and Holliday junctions in WRN (von Kobbe et al., 2003; Huber et al., 2006). RecQ5 possess a single α -helix that occupies a similar position to the missing WH domain and appears to be essential for helicase activity but not ATPase activity (Newman et al., 2017). Instead of the RQC, RecQ4 features a unique domain, the RecQ4 Zn²⁺-binding (R4ZBD) domain (in grey in Figure 11D), that consists of a zinc-binding region and two distinct folds that share some similarities with the WH domain (Kaiser et al., 2017).

The HRDC domain is the third conserved region among RecQ helicase localized at the C-terminus, but among the human RecQ helicases it is only found in WRN and BLM, suggesting that it plays a specialized role (in pink in Figure 12). It has a helical bundle structure, with different surface charges distribution among RecQ helicases, that gives specific DNA-binding affinities (Vindigni et al., 2010). For example, in BLM helicase this domain is a crucial determinant for the efficient binding and unwinding of double Holliday junctions (Wu et al., 2005).

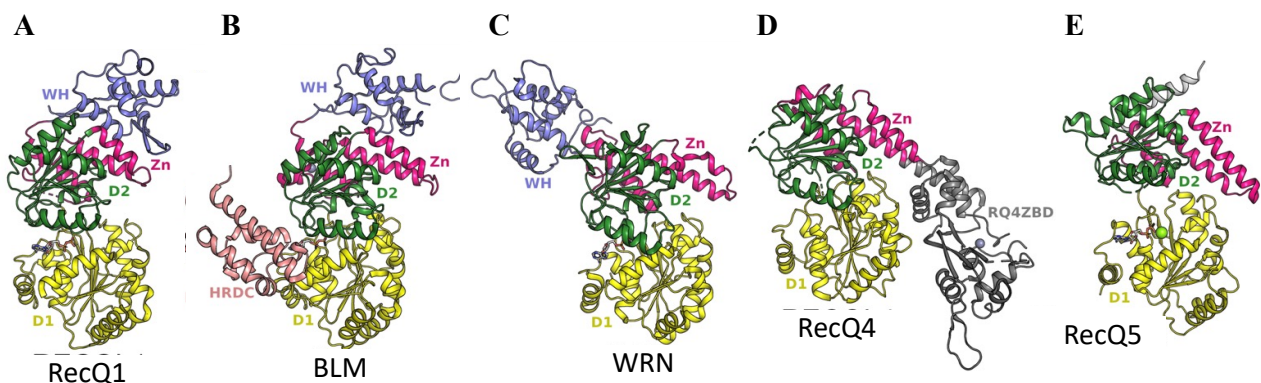


Figure 12. Crystal structures of the five human RecQ family members viewed from the same orientation. Color code as in Figure 11. **A.** RecQ1 (PDB ID: 2V1X; Pike et al., 2009). **B.** BLM (PDB ID: 4CGZ; Newman et al., 2015). **C.** WRN (PDB ID: 6YHR; Newman et al., 2020). **D.** RecQ4 (PDB ID: 5LST; Kaiser et al., 2017). **E.** RecQ5 (PDB ID: 5LB8; Newman et al., 2017). (Adapted from Newman & Gileadi, 2020)

Among the additional domains present in RecQ helicases, particularly interesting is the N-terminal domain of RecQ4 which is homologous to yeast Sld2. The Sld2-like N-terminal contains small, structured regions connected by long disordered segments. Among the structured regions the first 54 amino acids have been visualised by NMR and shows a helical bundle resembling a homeodomain-like DNA-interaction motif (Ohlenschläger et al., 2012), includes positively charged and aromatic residues possibly involved in DNA binding and has been shown to bind Holiday-Junctions and G-quadruplex DNA (Keller et al., 2014). Another small folded region just upstream of the helicase domain is a Zn knuckle that was solved by NMR and was shown to bind nucleic acid, with a preference for RNA (Marino et al., 2016).

Characterizations of RecQ helicases demonstrated that, while the helicase domains tend to be monomeric, they can adopt also higher-order oligomeric and dimeric forms, which can modulate the protein activity. There are various reports of RecQ1, BLM and WRN observed as tetrameric or hexameric species in electron microscopy studies (Muzzolini et al., 2007; Karow et al., 1999; Compton et al., 2008). However, the distribution of these oligomeric states in cells remain unknown.

4.2. Biochemical activities of human RecQ helicases

As already mentioned above, human RecQ helicases are ATP-dependent proteins able to unwind nucleic acid structures with a 3' – 5' polarity. Product amounts, reaction kinetics and protein affinities for the substrate in side-by-side reactions were used to determine the substrate preferences. Consistent with their multiple roles in DNA metabolism, they prefer substrates that are DNA metabolic intermediates. They are able to unwind the usual forked DNA duplex and hybrid substrates (DNA:RNA hybrid duplex), both being involved in replication and repair. They act also on a wide range of non-canonical substrates associated with repair, transcription and recombination, such as displacement loops (D-loops and R-loops), gapped DNA substrates, triple helices, Holliday junctions and G-quadruplex.

While some substrates, such as the forked DNA, Holliday junctions and D-loops, are processed by all the RecQ helicases, others, like R-loops, bubbles and G-quadruplex are more specific. A summary of RecQ substrate specificity is reported below (Sharma et al. 2006). Since human RecQ helicases have a 3' to 5' directionality, they cannot unwind duplexes DNA without a 3' ssDNA tail. D-loops are processed by BLM, WRN, and RecQ1, with BLM showing a preference for mobile D-loops. WRN, in addition to its helicase activity, also uses its exonuclease function to degrade D-loops with 3' ssDNA tails. Holliday junctions can be unwound by various RecQ helicases, including RecQ1 and RecQ5, although the binding of the bacterial RuvA protein to the HJ core inhibits this process. Triplex DNA helices, formed by the interaction of a third strand with the major groove of a duplex, are

processed by BLM and WRN, especially in the presence of a 3' ssDNA tail. Additionally, WRN, RECQ1, and BLM helicases act on 5' DNA flaps, formed during DNA replication and repair.

Although it is obvious that RecQ helicases have both distinct and overlapping roles in DNA metabolism, the exact substrate specificity of each human RecQ helicase is still not yet well established.

The activity of DNA unwinding by human RecQ helicases can be considerably enhanced by accessory factors, such as ssDNA-binding proteins. For example, a key nuclear protein partner of DNA helicases is Replication Protein A (RPA), which binds and coats the ssDNA with a very high affinity protecting it from damage. RPA interact physically and functionally with RecQ helicases stimulating their unwinding activity (Awate & Brosh, 2017). The interaction of RPA with WRN dramatically enhanced long duplex DNA (up to 850 base pairs in length) unwinding by WRN *in vitro* in a specific manner (Brosh et al., 1999). Also RecQ1, BLM and RecQ5 physically interact with RPA, and their DNA unwinding activity is strongly and specifically stimulated by this ssDNA interacting protein (Brosh et al., 2000; Cui et al., 2003; Garcia et al., 2004). Conversely, the forked duplex DNA unwinding catalyzed by purified recombinant RecQ4 was only modestly stimulated by RPA (Rossi et al., 2010).

Besides their helicase activity, RecQ helicases can also catalyze the opposite reaction of annealing of complementary strands in an ATP-independent fashion, and this activity is strongly inhibited by nucleotide binding (Sharma et al., 2005; Cheok et al., 2005). Annealing activity is potentially required to promote branch migration of DNA junctions (such as Holliday junction) or to promote the regression of stalled replication forks, which are characteristic activities of several RecQ helicases. Khadka and co-workers compared the DNA double-strand annealing activity between human RecQ proteins, showing that the relative strand annealing activities of the five RecQ proteins are not simply inversely related to their helicase activities. Notably, RecQ5 possesses relatively strong ATP-independent annealing activity on duplex substrates compared to the other RecQs (Khadka et al., 2016). The biological function of the annealing activity of RecQ helicases *in vivo* needs to be determined yet.

4.3. Functional role of human RecQ helicases

Human RecQ helicases interact with a wide variety of nucleic acid substrates and protein partners, playing multiple important roles in DNA metabolism, including DNA repair, recombination, replication, transcription and telomere maintenance.

4.3.1. RecQ helicases in DNA repair

In vitro and *in vivo* studies demonstrate that RecQ helicases have a preference for substrates that are DNA repair intermediates and are also able to interact with and modulate key DNA repair players. This suggests these helicases play a role in nearly all DNA repair pathways, in particular Base Excision Repair (BER) and Double-strand Break (DSB) repair. Although human RecQ helicases are recruited to DSBs with similar kinetics, there is a difference in their retention times, suggesting diverse and differential roles both in DNA damage signalling and repair processes (Lu & Davis, 2021).

BER corrects small DNA lesions, such as oxidated and alkylated bases, by using a glycosylase that removes the damaged base, leaving an apurinic/apyrimidinic (AP) site. AP endonuclease 1 (APE1) cleaves AP sites, producing a single-strand break intermediate, used by DNA polymerase β (Pol β) for DNA repair synthesis. Some RecQ helicases can regulate the transcription of BER-related genes, such as RecQ1, RecQ4 and RecQ5 (Sharma et al., 2012; Schurman et al., 2009; Tadokoro et al., 2012), or unwind several single-strand break BER intermediate, like WRN (Ahn et al., 2004), thus modulating BER indirectly.

DSBs are potentially lethal lesions that can cause transient or permanent cell-cycle arrest, mutagenesis, chromosomal rearrangements and ultimately cell death or tumorigenesis. Human cells can repair DSBs by nonhomologous end joining (NHEJ) or homologous recombination (HR). HR repairs DSBs when a homologous template is available, for example at stalled replication forks during the S phase, at DNA lesions during mitosis, and at chromosomal pairing during meiosis. NHEJ occurs throughout the cell cycle and is error prone.

RecQ1 is involved in both NHEJ and HR: knockdown experiments in human cells demonstrated that a reduced expression of RecQ1 leads to increased cell death following ionizing radiation as well as increased sister chromatid exchange in HeLa cells (Sharma & Brosh, 2007).

BLM and WRN are indispensable in both the early stages of HR, where they participate in DNA end resection (e.g. interacting with DNA2), as well as in the late stages of HR for the dissolution of Holliday junctions (Nimonkar et al., 2011; Sturzenegger et al., 2014; Constantinou et al., 2000). Moreover, they both process HR intermediates, such as D-loops, through different mechanisms. BLM drives proper HR by displacing the invading strand from the D-Loop formed by RAD51 during strand invasion and sister chromatid exchange, disrupting RAD51-ssDNA filaments (Bugreev et al., 2009). WRN helicase acts on the invading strand of the D-loop, completely displacing it or degrading it with its exonuclease activity (Opresko et al., 2009). WRN physically and functionally interacts with multiple members of the NHEJ machinery, including the Ku heterodimer, DNA-PK, and the

XRCC4-LIG4 complex, which stimulates its exonuclease activity (Cooper et al., 2000; Kusumoto et al., 2014; Kusumoto et al., 2008).

RecQ4 is recruited immediately to DSBs and, differently to BLM and WRN, it plays a crucial role only in the initial step of HR inducing DNA 5' end resection through its helicase activity (Lu et al., 2016a). RecQ4 also promotes NHEJ, interacting with the Ku complex and stimulating its binding to a blunt-ended dsDNA substrate. Depletion of RecQ4 reduced the end-joining activity on DNA substrates both *in vitro* and *in vivo* (Shamanna et al., 2014).

RecQ5 is recruited quickly to laser-induced DSBs, remaining for a shorter time compared to BLM and WRN, but longer than RecQ4 (Popuri et al., 2012a). Likewise BLM, RecQ5 act on RAD51 and prevent the formation of aberrant RAD51 filaments on the extended invading strand, inhibiting D-Loop formation and displacing it from ssDNA, avoiding inappropriate HR events (Paliwal et al., 2014).

4.3.2. RecQ helicases in DNA replication

The genomes of eukaryotic cells are replicated during the S phase of the cell cycle, starting from many replication origins distributed along the genome. Some RecQ helicases were shown to either participate in the initiation of the replication fork or remodel aberrant replication fork structures, especially under conditions of replication stress, and recovering stalled forks.

RecQ1 and RecQ4 are integral components of the replication complex and play distinct roles in DNA replication initiation and replication fork progression *in vivo*. RecQ4 is recruited to replication origins at the late G1 phase, after the pre-replicative complex is assembled, and interacts with MCM10 and DNA polymerase α , to promote DNA replication initiation (Thangavel et al., 2010; Xu et al., 2009). The N-terminal region of RecQ4 shows significant homology to yeast replication initiation factor Sld2, which in *S. cerevisiae* plays a key role in assembling the replicative complex, interacting with several proteins involved in replication initiation. This fusion event between the entire Sld2 and human RecQ4 helicase strongly argues in favour of its unique role in the early processes of eukaryotic DNA replication initiation (Matsuno et al., 2006). RecQ1 is recruited in early S phase, playing a role in replication initiation. (Thangavel et al., 2010). It has an additional role in replication fork progression, acting at damaged or stalled replication forks, helping in replication fork restart. RecQ1 promotes strand exchange and actively unwinds the leading strand of the fork, processing stalled replication forks (Popuri et al., 2012b). RecQ1 has been demonstrated to have a unique role among RecQ helicases in promoting the restart of reversed replication forks, caused by the inhibition of topoisomerases, through branch migration or DNA unwinding (Berti et al., 2013).

Conversely, BLM, WRN and RecQ5 do not have a role in DNA replication initiation, but function at the origins in the S phase in response to replication stress to promote fork recovery or repair (Bachrati & Hickson, 2008). BLM promotes resolution of aberrant DNA structures at stalled replication forks. It assembles into a complex (with topoisomerase III α and RMI1) and is able to dissolve Holliday junctions which can arise upon replication fork stalling, by rewinding the parental DNA strands and annealing the two nascent strands (Wu & Hickson, 2003). This activity is also important at the sites of DNA replication termination, in order to efficiently complete DNA replication (Wu et al., 2005). Similarly to BLM, WRN promotes the restart of stalled replication forks. The unwinding activity of WRN is coupled to the nuclease activity of DNA2 to degrade the nascent DNA promoting replication restart, demonstrated by impaired recovery of stalled forks and accumulation of reversed forks in WRN-deficient cells (Thangavel et al., 2015). Furthermore, WRN exonuclease activity is useful in protecting the nascent DNA strand against degradations after replication stress (Iannascoli et al., 2015).

RecQ5 is the only member of the human RecQ family to directly interact with the RNA polymerase II (RNAPII), promoting the resolution of replication-transcription collisions and thus being involved in both transcription and transcription-associated replication processes. In particular, RecQ5 counteracts replication fork stalling by preventing RNAPII and RNAPI from pausing or arresting at the transcribed regions of chromatin, where they form a barrier for replication fork progression (Urban et al., 2016).

4.3.3. RecQ helicases in telomere maintenance

Human RecQ helicases play significant roles in the proper maintenance of telomeres, acting on their replication, recombination and repair. They are able to resolve obstacles during telomeric DNA replication that can stall or block replication fork, such as the G-quadruplexes which can form in the G-rich sequence. They can avoid telomere shortening by interacting with other telomere-binding proteins involved in the ALT pathway, a recombination mechanism used by the cells to try to restore telomeric length when the telomeres become critically short.

BLM, WRN and RecQ1 were shown to maintain the stability of telomeres playing a role in the dissociation of alternate/complex DNA intermediates arising during replication of telomeric DNA, which can potentially block replication and cause telomere fragility. BLM and WRN are able to resolve long telomeric forked duplexes and D-Loop/T-Loop structures, stimulated by the telomeric single-strand DNA-binding protein POT1 (Ghosh et al., 2009; Opresko et al., 2005). RecQ1 resolves telomeric D-loops and Holliday junction substrates by interacting with telomere sheltering proteins.

Loss of RecQ1 results in dysfunctional telomeres, telomere loss and telomere shortening, elevation of telomere sister-chromatid exchanges and increased telomere fragility (Popuri et al., 2014). BLM and WRN are also able to modulate G-quadruplex structures, promoting the restart of stalled forks thus facilitating proper elongation of telomeric ends (Drosopoulos et al., 2015; Ghosh et al., 2009). Moreover, they are implicated in the ALT pathway helping to resolve recombination intermediates (Stavropoulos et al., 2002; Mendez-Bermudez et al., 2012). WRN has a critical function in telomere maintenance, as suggested by the characteristic premature senescence features and accelerated telomeric shortening of Werner syndrome patients. Its association to telomere-specific sheltering proteins that are essential for telomere capping has been shown to suppress defects in telomere lagging strand synthesis (Crabbe et al., 2004).

RecQ4 also associates with telomeres and telomeric factors: RTS patients show elevated levels of fragile telomeric ends and RecQ4-depleted human cells accumulate telomere replication defects, including fragile sites, sister chromatid exchanges and double-strand breaks at telomeric sites (Ghosh et al., 2012). In association with sheltering proteins or working synergically with WRN or the telomeric CST (CTC1-Stn1-Ten1) complex, RecQ4 has been shown to resolve telomeric D-loop structures (Ghosh et al., 2012; Li et al., 2023).

No evidence for a role of RecQ5 in telomeric maintenance is available in the literature.

4.4. RecQ-associated syndromes

Mutations in four of the five human RecQ helicases lead to distinct genetic diseases: BLM helicase is associated with Bloom syndrome, WRN helicase with Werner syndrome, RecQ4 with three diseases (Rothmund-Thomson syndrome, RAPADILINO and Baller-Gerold syndrome) and RecQ1 with RECON syndrome. These genetic disorders exhibit a unique set of clinical and cellular features (listed in Table 2), underlining that these helicases, even if are structurally and functional related, perform different tasks within the cell. However, all these disorders present cancer predisposition, consistent with a role of these proteins in the key steps of DNA replication and repair, but their cancer profiles are different (Table 2).

RecQ5 has not been associated with heritable diseases, but it has been associated with increased cancer rates.

Syndrome (abbr) (gene)	Main Clinical Features	Cancer Predisposition
Blooms syndrome (BS) (BLM)	Dwarfism, beaked nose, narrow face, pigmentation, redness, and dilated blood vessels in skin, mental retardation, type-II diabetes, immunodeficiency, lung problems, low or no fertility	Early onset; colon, breast and lung cancers; leukaemia and lymphomas; sarcomas
Werner syndrome (WS) (WRN)	Premature aging phenotypes (bilateral cataracts, skin alterations, premature grey and loss of hair, short stature), osteoporosis, hypogonadism, diabetes, soft tissue calcification, premature atherosclerosis, myocardial infarction and stroke, reduced fertility.	Early onset of primary sarcomas and mesenchymal tumors
Rothmund-Thomson syndrome (RTS) (RecQ4)	Poikiloderma, juvenile cataracts, growth retardation, skeletal and dental abnormalities (dysplasia, osteopenia), sparse scalp hair, hypogonadism, skin disorders, light sensitivity, age prematurely, congenital short stature.	Early onset of osteosarcomas
RAPADILINO syndrome (RAPA) (RecQ4)	Short stature, abnormal pigmentation, growth retardation, skeletal abnormalities (limb malformation, radial defects such as hypoplasia and aplasia of thumbs and radius), gastrointestinal malfunction.	Lymphoma and osteosarcoma
Baller-Gerold syndrome (BGS) (RecQ4)	Craniosynostosis, radial aplasia and hypoplasia, skeletal dysplasia, poikiloderma, growth retardation, short stature	Osteosarcoma
RECql One syndrome (RECON) (RecQ1)	Striking progeroid facial features, tiny and pinched nose, prominent premaxilla, skin photosensitivity, xeroderma, slender and elongated thumbs, postnatal grow retardation and muscle wasting.	Breast cancer

Table 2. Clinical features of RecQ disorders. (Mojumdar, 2020; Abu-Libdeh et al., 2022)

The unique features of each pathology underline the role of each helicase in DNA metabolism. BS cells show increased frequency of sister chromatid exchanges and HR, are sensitive to UV radiation and have a high number of DNA gaps, breaks and rearranged chromosomes. This highlights the role of BLM in repair pathways and in restarting of replication forks (Cunniff et al., 2017). WS cells display replicative senescence and severe shortening of telomere, demonstrating WRN role in telomere maintenance and its preference for telomeric G4 structures. Similarly to BLM, WS cells exhibits increased incorrect recombination and high frequency of chromosomal deletions and DNA breaks, in line with WRN function in DNA repair (Oshima et al., 2017). RTS cells are genetically unstable and display high frequency of chromosomal abnormalities, consistent with the main role of RecQ4 in DNA replication. Interestingly, the RTS or cancer-causing mutations are spread over the helicase region, except the N-terminal region, which is essential for embryonic survival (Martins et al., 2023). The cells from the recently identified RECON disorder show chromosomal instability and,

in particular, a replication defect after inhibition of topoisomerases but not PARP1. This result confirms the prominent role played by RecQ1 in restarting reversed forks (Abu-Libdeh et al., 2022).

4.5. RecQ helicases as therapeutic targets for cancer

The role of these helicases in cancer is thus complex: whereas their inactivation predisposes to cancer, their expression is often up-regulated in cancer cells, probably reflecting the need to counteract the increased level of DNA damage and replicative stress that is produced in rapidly proliferating cancerous cells. Thus, RecQ helicases can be important factors for the survival and proliferation of tumors and for the resistance to DNA-damaging agents used in chemotherapy.

As RecQ helicases are dysregulated in cancer, they may be novel and direct therapeutic targets for the treatment of cancer or other diseases (Brosh, 2013). Many efforts are therefore employed to identify candidate small molecule inhibitors specifically targeting RecQ helicases.

Small molecule inhibitors of the WRN helicase activity have already been identified by in vitro screenings and found to inhibit the proliferation of human endocervical carcinoma cells (Aggarwal et al., 2013). A potent inhibitor of BLM helicase has also been discovered through high-throughput screening of a chemical compound library and medicinal chemistry optimization (Nguyen et al., 2013). A small molecule that inhibits the catalytic activity of BLM more than other RecQ family helicases has been recently characterized and shown to induce cell cycle arrest and increased apoptosis in human multiple myeloma cell lines due to accumulation of DNA damage (Ovejero et al., 2022). More recently, a novel WRN inhibitor with potent and selective anti-tumor activity in cancers with microsatellite instability (MSI) has been identified and is currently in clinical trials. This inhibitor binds allosterically at the interface of the HD1 and HD2 helicase domains, thereby locking WRN in an inactive conformation (Ferretti et al., 2024).

Evidence for a role of RecQ helicases in G4 metabolism opens new possibilities of targeting these structures with small G4 ligands, which can interfere with protein activity thus modulating various DNA processes. A plethora of small molecules have been designed to bind and stabilise G4s, thus potentially affecting the growth of cancer cells at multiple levels, by modulating oncogene transcription, triggering apoptosis by stabilizing telomeric G4s, and possibly affecting DNA replication via the putative G4s present at origins of replication (Criscuolo et al., 2022). However, most of these ligands have shown to be not efficient or selective enough to be used as therapeutic agents.

Further elucidating the molecular determinants of the G4s:helicases interactions will increase our understanding of the system and provide solid bases for identifying and developing novel anticancer strategies.

4.6. Human RecQ helicases in G-quadruplex metabolism

The RecQ helicases are the most likely players in G4 metabolism. It has been suggested that the RQC domain, which is peculiar to some RecQ family helicases may be a determinant for binding DNA substrates and especially G-quadruplexes (Huber et al., 2006). However, no structural data is available to support or disprove this hypothesis.

4.6.1. BLM in G4 metabolism

BLM was one of the first identified helicases able to bind and unwind both intermolecular and intramolecular G4 DNA in an ATP-dependent manner. Similarly to duplex DNA, it requires a 3'-single-stranded region for loading and resolving the G4 structure, not being able to unwind tailless G4 or G4 with a 5' tail (Sun et al., 1998). BLM shows different unwinding efficacy based on G4 topology: the protein unwinds intermolecular G4 better than duplex DNA, whilst intramolecular G4s are unwound with lower efficiency (Budhathoki et al., 2014). The protein failed to unfold a 4-layered G4 at various ATP concentrations in presence of K⁺ ions, whilst in Na⁺ buffers, where the G4 is less stable, it unwinds this G4 structure efficiently (Wu et al., 2015). These results indicate that the efficiency of BLM G4 unwinding varies among G4 structures. This also suggests that G4 structures can differently affect the processing of genomic DNA, possibly hindering the translocation of BLM on DNA. (Liu et al., 2010).

BLM unfolds G4 through different mechanisms depending on the molecular and structural environments of the G-quadruplex.

Biophysical studies based on single molecule FRET suggested that BLM-catalyzed G4 unfolding may be ATP-independent at high concentrations and required a ssDNA overhang region in the vicinity of the G4 motif. Unwinding occurs thanks to the binding of BLM to the ssDNA near to the G4 and through the cooperation between the RQC and the HRDC domains, the latter binding to the ssDNA and stabilizing the RQC domain interaction with the G4 structure, thus leading to the destabilization and unfolding of the G4 in the absence of ATP (Figure 13A). It is possible that this ATP-independent G4 unwinding activity can be due to thermodynamic fluctuations after BLM addition and trap of the unfolded G4 by BLM, supported by the fact that the efficiency of BLM-mediated G4 unfolding correlates with the binding stability of BLM to ssDNA overhang, as modulated by the nucleotide

state, ionic conditions, overhang length and overhang directionality (Chatterjee et al., 2014; Budhathoki et al., 2014).

Two different ATP-dependent mechanisms for unwinding G4 structures, depending on different G4 structures associated with distinct biological contexts, were suggested using single-molecule FRET (Wu et al., 2015). When in presence of a DNA structure-containing G4, representative of an intermediate formed in DSBs or at telomere, BLM unfolds the three-layered G4 (with a 3' ssDNA tail and a 5' dsDNA) in a stepwise manner (Figure 13B). BLM does this by transitioning through four distinct conformational states, two of which probably correspond to G-hairpin and G-triplex structures. After fully unfolding the G4 structure, BLM repetitively unwinds the duplex DNA until it eventually dissociates. However, when the G4 structure resembles a structure formed during DNA replication or RNA transcription (G4 with a 3' dsDNA followed by a 12-nucleotide ssDNA segment adjacent to the G4 motif), the unfolding mechanism changes. In this case, BLM binds at the junction between the dsDNA and the ssDNA. It then stretches and pulls in the ssDNA, creating a loop. BLM translocates along the ssDNA, rather than the G4 structure itself, to unfold the G4 (Figure 13C). (Wu et al., 2015).

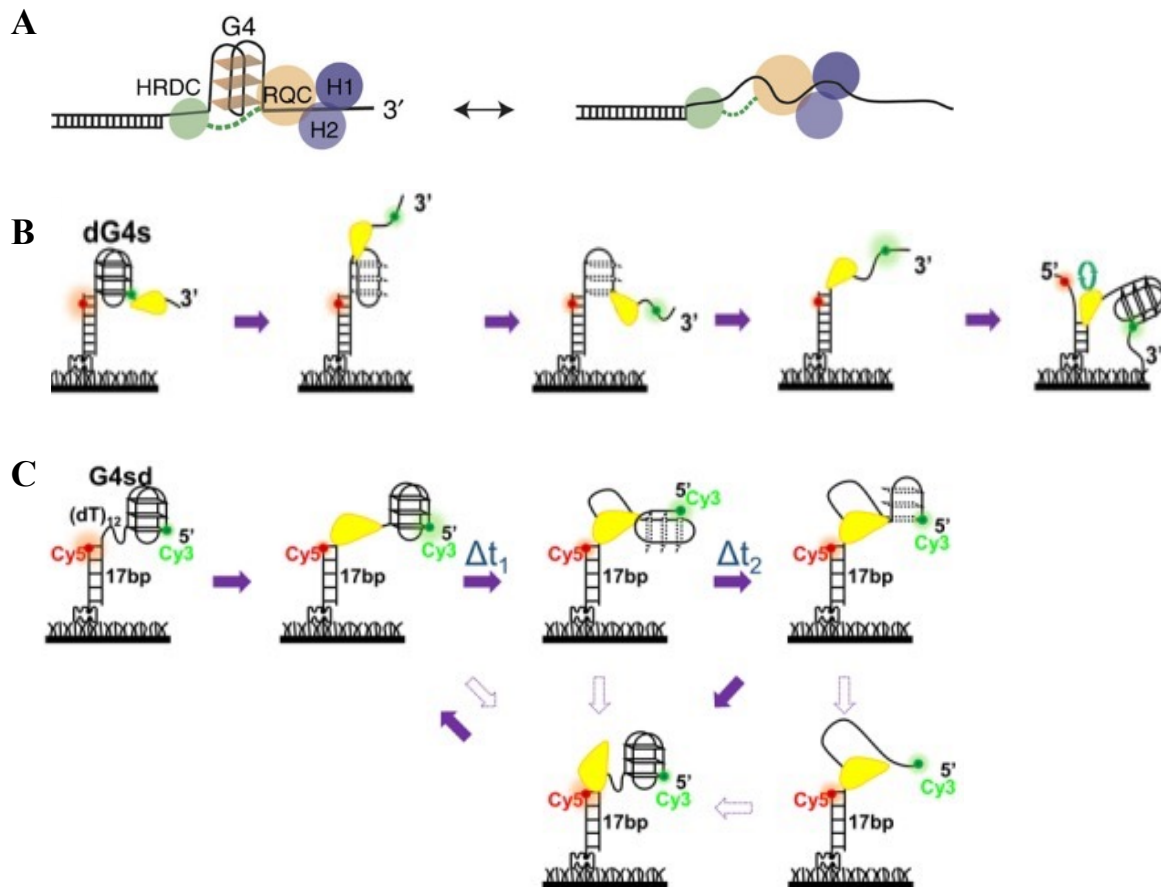


Figure 13. Proposed models for BLM–G4 unfolding mechanisms. A. ATP-independent unfolding of a G4 motif and regulation of BLM activity via binding to the ssDNA and cooperativity of RQC–HRDC domains

(From Chatterjee et al., 2014). **B.** Schematic representation of the proposed ATP-dependent model showing BLM unfolding G4s formed from a double-strand break or at telomere. DNA constructs (black) composed of a G4 strand labelled with Cy3 (green) and a complementary stem strand labelled with Cy5 (red) is immobilized on a PEG-coated coverslip. BLM protein is represented by a yellow blob. **C.** Schematic diagram of the proposed ATP-dependent model to explain how BLM unfolds G4s during DNA replication and RNA transcription. Elements colour-coded as in B. (From Wu et al., 2015).

The specialized role of BLM in G4 processing could be crucial for many G4-associated biological processes such as telomere maintenance, replication, recombination and transcription. These functions help in maintaining genome stability and might explain the characteristic features of Bloom syndrome (Cunniff et al., 2017).

BLM plays a critical role in maintaining telomere stability promoting efficient replication by unwinding G4 structures, which form during leading strand synthesis at telomeres. In BLM-deficient cells there is a reduction in the rate of fork progression through telomere repeats and this is due to the presence of unprocessed G4 structures that impeded the copy of the leading strand (Drosopoulos et al., 2015). BLM regulates replication also in transcribed genes and suppresses recombination at G4 motif by resolving G4 structures in the leading strand (van Wietmarschen et al., 2018). In these regions, the G4 structures are known to be barriers for DNA synthesis, leading to stalled replication forks, anomalous recombination and genome instability. In BS cells and in BLM-depleted cells, sister chromatid exchange are enriched in transcribed genes containing G4 compared to non-transcribed genes and also compared to wild-type cells (van Wietmarschen et al., 2018).

BLM also regulates gene expression by directly binding and unwinding G4 structures that are enriched at transcription start sites, thus contributing to the pathogenesis of Bloom syndrome. Genes that were found to be downregulated in BS patients had a significant enrichment of G4 motifs upstream of the transcription start site, suggesting the formation of G4 structures that impede the transcription. (Nguyen et al., 2014).

BLM can work synergically with other proteins in resolving G4 structures. Deficiency of either BLM and WRN resulted in greater increases in telomere dysfunction with a higher level of G4 detected in the telomere compared with G4 seen genome-wide, in addition to a considerable reduction in telomere length (Drosopoulos et al., 2015). BLM may use multiple mechanisms that are not mutually exclusive to maintain telomere stability, also collaborating with FANCD1 on removing G4 structures in telomeric leading strand from opposite directions. Additionally, FANCD1 was demonstrated to act in concert with both BLM and WRN to maintain epigenetic stability by guaranteeing continuous replication at G4-forming DNA sequences. FANCD1 moves in a 5' to 3' direction, thus the opposite approaching of

the three helicases on the DNA could facilitate G4 disruption and consequently the processive DNA replication through the quadruplex (Sarkies et al., 2012)

4.6.2. WRN in G4 metabolism

WRN helicase, together with BLM, is one of the well-known helicases able to unwind G4 structures both *in vitro* and *in vivo*. It has robust ATP-dependent G4 DNA unwinding activity and sometimes cooperates with other helicases modulating G4s, having both unique and overlapping functions. WRN requires a 3' tail to bind and unwind G4 structures, not being active on blunt G4s (Fry & Loeb, 1999). Similar to BLM, WRN helicase is predicted to interact with G4 structures through the RQC domain. G4 unwinding activity was shown against two G4 structures: the G4 found in the promoter of cMYC G4 and a telomeric G4 sequence. Based on an NMR study of the WH domain, a few amino acids in the RQC domain were shown to be involved in G4 binding and unwinding: Thr-1024, Thr-1086 and Leu-1093 (Ketkar et al., 2017). Interestingly, these residues are located close to the β -wing in a flexible C-terminal region of the RQC domain, suggesting that this portion could adopt multiple conformations to easily accommodate the dynamic nature of the G4 structure (Ketkar et al., 2017). WRN shows an apparent low processivity on G4 structures, but also on dsDNA. Its repetitive activity of moving backward and forward across the G4 sequence, keeps the G4 constantly in an open conformation, helping other enzymes in their functions during transcription, replication, DNA repair and telomere maintenance. Some G4 structures derived from DNA intermediates from different cellular context may also hinder WRN translocation along the DNA, regulating its activity in different structural environments (Wu et al., 2017).

WRN helicase has a role in G4-mediated replication, both at telomeres and in genetic regions. Differently from BLM, WRN has been proposed to exert its helicase function on the lagging strand of replication. The lack of WRN helicase activity leads to impaired telomere G-rich lagging strand synthesis; the exposed G-rich ssDNA is degraded by nucleases, leading to end-to-end fusions of telomeres and telomere loss from single sister chromatids (Crabbe et al., 2004). WRN role in telomere explains the premature aging and chromosomal instability characteristic of the Werner syndrome. Studies on late-generation telomerase and WRN-deficient mice show the panel typically observed in WS patients: dysfunctional telomeres with accelerated telomere shortening and elevated recombination rates between the telomeres of sister chromatids (restored by the expression of active WRN), replicative senescence, chromosomal instability and premature aging (Laud et al., 2005).

Similarly to BLM, WRN modulates gene transcription by targeting G4 DNA, promoting their unwinding. It has been shown that many of the upregulated genes in WS (and also BS) cells, that

contribute to Werner syndrome pathogenesis or acquired disease risk, have significant associations with regions characterized by a high presence of G4 motifs (Johnson et al., 2010). Genes that are down-regulated in both WS cells and WRN-depleted cells show significant enrichment of G4s at the transcription start site, contrary to genes exhibiting increased expression (Tang et al., 2016). Consistent with short stature being a classical sign in WS patients, WRN has been shown to regulate chondrogenesis through the transcriptional regulation of SHOX (short-stature homeobox), promoting its expression by unwinding G4 structures (Tian et al., 2022). Despite a similar mechanism as observed in BLM syndrome, miRNAs analysis revealed a different gene expression signature in WS cells compared to BS cells, in line with the fact that WS and BS are distinct diseases. This suggests that WRN and BLM are helicases able to specifically interact with diverse G4 structures in order to differentially modulate human gene expression, and that G4-containing genes appear specifically resolved by the individual helicase (Tang et al., 2016).

4.6.3. RecQ1 in G4 metabolism

RecQ1 was reported to bind G4 substrates with an affinity similar to the one observed for BLM and there are evidence showing that RecQ1 binds to the G4 motifs in the promoters of downregulated genes having a role in transcription regulation (Lu et al., 2016b). However, RecQ1 was reported to inefficiently unfold both intermolecular and intramolecular G4s *in vitro*, even in the presence of saturating concentration of human RPA (Popuri et al., 2008).

Recent studies show new evidence about the ability of human RecQ1 to unwind G4 structures. The homolog *Bos taurus* RecQL (BtRecQL), which shares 90,5% identity with human RecQ1, was shown to be able, in its monomeric form, to unfold intramolecular G4s in an ATP-dependent manner and with an efficiency comparable to BLM protein. Moreover, in contrast with the previous studies, the monomeric form displays higher activity compared to the tetrameric and the dimeric forms of the protein (Liu et al., 2021). Another study, performed by Dabral et al., demonstrated the role of RecQ1 in facilitating Kaposi Sarcoma associated Herpesvirus (KSHV) lytic DNA replication through its helicase activity. RecQ1, recruited by a component of the prereplication complex indispensable for the activation of promoters of viral genes, thus for lytic reactivation (RTA protein), binds selectively to the G4 forming sites in the origin of lytic DNA replication (oriLyt) region of KSHV. In particular, the inhibition of RecQ1 helicase activity at G4 sites using NMM, a strong G4 stabilizer, led to a reduction in actively replicated DNA, genome persistence and copies, virion concentration and late gene expression (Dabral et al., 2023).

Additional biochemical studies need to be performed to better understand the mechanism of interaction and regulation between RecQ1 and G4 structures.

4.6.4. RecQ4 in G4 metabolism

The N-terminal domain of RecQ4, in particular the first 400 amino acids, binds to G4 structures with higher affinity compared to ssDNA and dsDNA, whilst a C-terminal fragment of RecQ4, including the helicase core and the C-terminal domain, is not able to bind to these non-canonical structures (Keller et al., 2014). A more recent analysis suggested the presence of a flexible positively charged RecQ4-specific motif (RSM, between residues 348-388), which is conserved among various RecQ4 homologues, and is able to bind with significant affinity to G4 structures as well as to RPA protein. These results suggest the N-terminus of RecQ4 is necessary and sufficient for binding to G4 structures. Interestingly, the binding of RSM to the G4 is strong enough to inhibit the helicase activity of various helicases able to unwind G4 structures, such as Pif1 and FANCI, and may thus represent a barrier for processing enzymes, implying a possible signal function of RecQ4 during temporal activation of replication origins (Papageorgiou et al., 2023).

Contrary to previous reports, a recent paper suggests that RecQ4 alone, in the presence of ATP, is able to unwind intramolecular G4 structures, even if this unwinding activity may not be robust as that of BLM and WRN (Li et al., 2023). Additionally, RecQ4 interacts both *in vitro* and *in vivo* with human CST (CTC1-STN1-TEN1), a ssDNA-binding complex that cooperates with the replisome to aid in stalled fork rescue, and this interaction significantly enhanced the helicase activity of RecQ4 in resolving G4 structures. CST is required for the engagement of RecQ4 to the G4-rich sequences in the genome and to enhance RecQ4 activity to unwind G4 structures, suggesting a collaborative nature of these two proteins in resolving G4s and preserving genome and telomere stability. In particular, RecQ4 is recruited by and interacts with CST at telomeres, resolving telomeric G4 structures in order to maintain normal telomere DNA replication and telomere integrity. Similarly, RecQ4 collaborates with CST to restart the replication forks, specifically stalled by G4 structures (Li et al., 2023). However, the results reported in this paper may not be entirely reliable due to several key concerns with the experimental conditions used in the helicase assays. First, the large excess of DNA (1M) relative to the protein concentration (maximum of 100 nM) likely creates an imbalanced system that does not accurately reflect physiological conditions, potentially leading to non-specific effects. Additionally, the use of 150 mM NaCl instead of the more common KCl is problematic, as sodium ions are known to stabilize G4 structures less than potassium ions, possibly altering the results by artificially facilitating G4 unwinding. Furthermore, the reported ability of RecQ4 helicase-dead mutant as well as the wild-type protein in the absence of ATP to resolve G4 structures raises additional questions.

4.6.5. RecQ5 in G4 metabolism

The G4 unwinding activity of RecQ5 is relatively weak in vitro when compared to BLM and WRN, suggesting it is probably not primarily involved in G4 destabilization. The G4 unfolding activity of RecQ5 in presence of two different G4 structures (the human telomeric G4 sequence and the less stable non-telomeric G4 formed by thrombin binding aptamer TBA) were characterized under different salt conditions. Results suggest that the ATP-dependent G4 unfolding by RecQ5 is strongly correlated with G4 stability, but it is conceivable that this weak activity may help to compensate for WRN and BLM when their activity is impaired (Budhathoki et al., 2016). However, there is still no evidence that RecQ5 might modulate G4 structures in vivo to regulate cellular processes in cells.

5. Iron-Sulphur Cluster helicases

Helicases characterized by the presence of an Iron-Sulphur (FeS) cluster are a family of helicases well conserved from bacteria to humans. Structural information on the FeS helicase family mostly comes from the crystal structures of the archaeal XPD enzymes (Fan et al., 2008; Liu et al., 2008) and the bacterial homologue DinG (Chen & Wigley, 2018). They feature two unique domains, whose functions have not been fully defined. The FeS domain (in red in Figure 14), which contains the Iron-Sulphur cluster, is a four-helix assembly stabilized by FeS cluster via four conserved cysteines, and has been shown to be essential for the helicase activity. In its normal, unoxidized state, the cluster is likely to exist as a 4Fe-4S cluster. The ARCH domain (in blue in Figure 14) is made up of four α helices and a four-stranded β sheet. The ARCH and the FeS domain together form a pore, in which ssDNA binds. Mutations of the cysteines to serines result in the loss of cluster. The loss of cluster affects the helicase activity but not the ATPase activity stimulated by ssDNA (Rudolf et al., 2006).

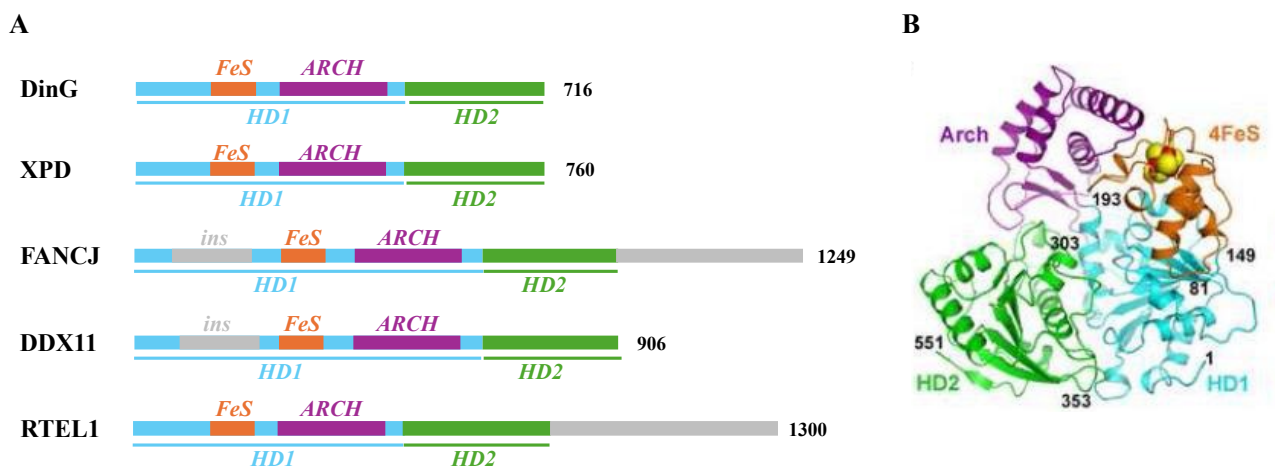


Figure 14. FeS cluster helicases. **A.** Schematic primary sequence representation of FeS helicases. Different domains are highlighted with different colors; in addition to the two RecA-like domains (HD1 and HD2, in cyan and green respectively), the helicase core contains the characteristic FeS domain (in orange) as well as a unique domain called ARCH (in purple). **B.** The crystal structure of the XPD helicase (PDB ID: 3CRV. Fan et al., 2008) as example of the characteristic domains present in FeS helicases (color coded as in A).

In archaea and eukaryotes, FeS helicases are ubiquitous, with human cells including four paralogues (XPD, FancJ, DDX11 and RTEL1, Figure 14A), which have important roles in DNA replication, DNA repair, telomere metabolism, and sister chromatid cohesion, and are involved in genetic diseases and cancer metabolism (Estep & Brosh, 2018). In bacteria the situation is more confused: most, but not all, bacteria contain at least one protein sharing similarity to the archaeal and eukaryotic FeS

helicases, but in a subset of bacteria the helicase is fused to an N-terminal exonuclease-like domain, and in some cases (as in *Staphylococcus aureus*) the helicase domain has lost the FeS cluster and thus the ability to unwind DNA (McRobbie et al., 2012).

Most members of the FeS helicase family, including FANCI (Brosh & Wu, 2021), RTEL1 (Kotsantis et al., 2020) and DDX11 (Pisani et al., 2018), are able to bind and unwind non-canonical DNA structures, in particular G4s.

5.1. *E. coli* DinG helicase

DinG (DNA damage-inducible gene G) from *Escherichia coli* is the prototype of a bacterial functional FeS helicase, and has been suggested to have a role in DNA repair and in replication fork restart pathways after DNA damage. (Rudolf et al., 2006; Gupta et al., 2005; Uringa et al., 2011). DinG comprises the four characteristic domains of FeS helicases (Figure 14A and Figure 15A). The canonical HD1 and HD2 domains have the expected fold seen in other structures, with an ATP-binding site located between them (Singleton et al., 2007). The FeS domain, which is inserted within HD1, folds in an α -helical structure and contains an Iron-Sulphur cluster at its center. The FeS cluster is essential for its activity, in fact a reduction of the cluster reversibly switched off the DNA helicase activity (Ren et al., 2009). The ARCH domain sits above the other domains and contacts the FeS domain above the ssDNA binding surface.

Both overexpression and deletion of *dinG* gene do not have a significantly effect on cell survival after UV irradiation, indicating that the protein is not essential for repairing DNA photolesions or restoration of replication arrest after UV-damage (Voloshin et al., 2003). The crystal structure of *E. coli* DinG bound to a ssDNA, both in the presence and absence of an ATP analogue (ADP-BeF₃; Figure 15B), revealed further information on the DNA translocation mechanism of SF2 5'-3' helicases, and in particular of XPD class of enzymes (Cheng & Wigley, 2018). The ssDNA fits in a positive-charged groove across the ATPase domains and it translocates first through the HD2 domain and then across the HD1 domain. The two domains undergo a relative conformational change caused by ATP binding and after ATP hydrolysis they relaxed to the apo form.

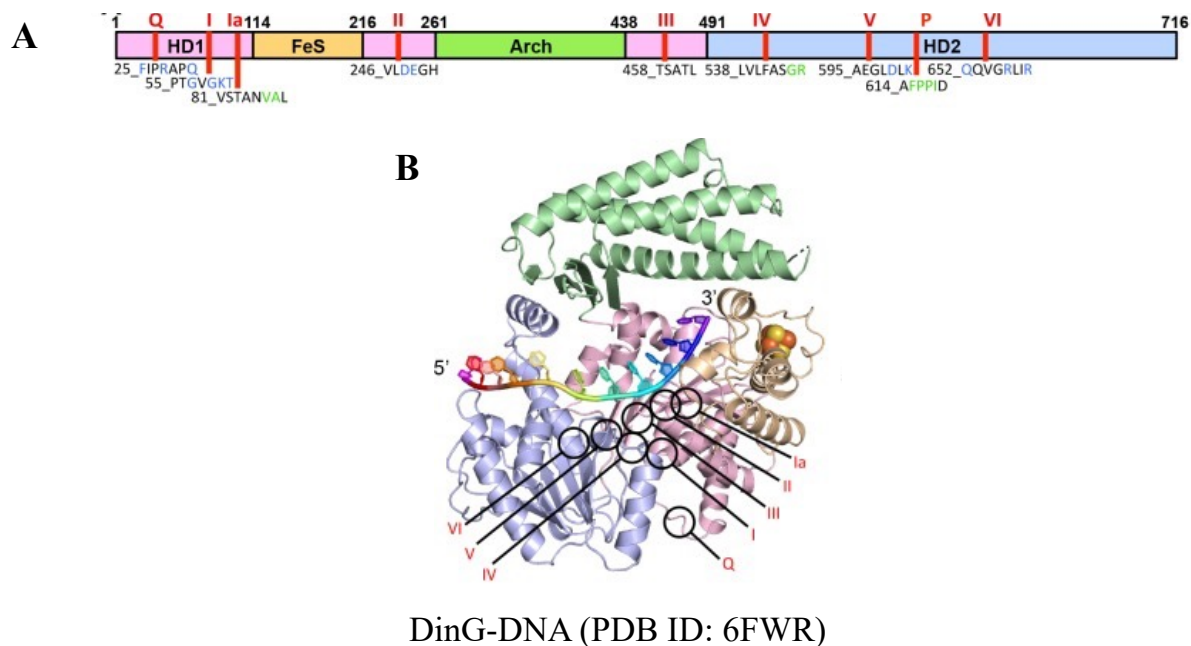


Figure 15. *E. coli* DinG. **A.** Domain organization of *E. coli* DinG. The HD1, HD2, FeS and Arch domains are colored in pink, lilac, wheat and green, respectively. Conserved motifs are labelled and listed. **B.** Structure cartoon view of DinG bound to ssDNA colored as in A. (From Cheng & Wigley, 2018).

DinG is a helicase able to unwind duplex DNA with a 5'-3' directionality and it requires a 11-15 nucleotide 5'-tail for efficient DNA duplex unwinding. It unwinds not only DNA-forked structures, but also DNA-RNA hybrid duplexes and bifurcated substrates, such as flap substrates, which can be considered a model for a replication fork. This suggests the role of DinG in unwinding duplex ahead of a stalled replication fork, translocating along the lagging DNA strand. The protein is also able to bind and unwind non-canonical substrates, such as D-loops and R-loops, suggesting a possible role in modulating inducible stable DNA replication and constitutive stable DNA replication (Voloshin & Camerini-Otero, 2007).

DinG may play a role also in G-quadruplex stability: both *E. coli* and *M. tuberculosis* DinG were shown to be able to resolve G4 DNA structures. *M. tuberculosis* DinG was shown to be active on bimolecular and tetramolecular G4s, unwinding G4 structures with both 5' and 3' tails and with parallel and antiparallel conformations (Thakur et al., 2014). Although Barthi et al. (2013) demonstrated that *E. coli* DinG is able to unwind bimolecular and tetramolecular G4s but not unimolecular G4s, recent evidence suggest that the helicase may remove unimolecular G4 structures formed in either the leading or laggings strand during replication (Parekh et al., 2023). Biochemical and biophysical evidence of DinG ability to unwind the biologically relevant unimolecular G4 structures were recently published by our laboratory (De Piante et al., 2023) and are presented in this thesis.

6. Project aim

G-quadruplexes and RecQ helicases have emerged as novel and direct targets for anticancer therapies. However, most ligands that show potential in targeting these structures lack the necessary selectivity to be used as therapeutic agents, limiting their success so far. Understanding the molecular determinants of the G4s:helicases interactions will increase our understanding of the system and provide a solid groundwork for identifying and developing novel anticancer strategies.

RecQ helicases have been shown to be involved in multiple processes in DNA metabolisms and appear to have specific but partially overlapping functions. Although there are multiple reports on the interactions between RecQ helicases and G4s, there is the need for a more systematic analysis to clarify their specificity towards different G4 topologies, and to better understand their distinct and/or overlapping roles in G4 metabolism. Moreover, structural information on G4 and helicase complexes is scarce, being limited to few crystal structures, preventing a full understanding of the mechanisms of interaction.

The aim of this project is to perform a comprehensive analysis of the interaction between human RecQ helicases and various G4s, in order to elucidate their specificity and selectivity towards these non-canonical structures and to understand their roles in G4 metabolism.

For this purpose, we planned to express and purify the catalytic domains of each human RecQ helicases, and test their interactions with a variety of G4 structures. We choose well characterized G4s with different origins and topologies. A wide range of biochemical and biophysical techniques were applied to dissect the interactions and identify the highest-affinity helicase-G4 partners. A complete biochemical characterization required the expression and the production of significant amounts of pure, stable, well-folded and active protein, a process that was challenging and time-consuming.

Since the crystal structure of a bacterial RecQ bound to a resolved G4 revealed the presence of a putative guanine-binding pocket within RecQ helicases, we also decided to assess the presence of this motif in human RecQ helicases to better understand the molecular determinants of the interaction. For this purpose, a bioinformatic analysis comparing helicases structures was first performed, followed by site-directed mutagenesis on BLM and WRN to disrupt the pocket. An additional biochemical characterization confirmed its role in G4 interaction.

The highest-affinity helicase-G4 complexes were then subjected to extensive crystallization experiments for high-resolution X-ray diffraction studies. Structural data on RecQs-G4 complexes

would be very useful for elucidating not only the function of this specific family of helicases in G4 metabolism, but also for shedding light on the role of the whole eukaryotic SF2 family.

Since the ability to bind and unwind G-quadruplex is also shared across most members of the FeS helicase family, and given the evidence of *E. coli* DinG activity on G4 structures, we performed a detailed study of the interactions between the *E. coli* DinG and a variety of G4s. This side project helped us to set up the experimental conditions used in our biochemical assays. A series of G4-rich sequences from different genomic locations (promoter and telomeric regions), and different organisms (human and bacteria), able to form unimolecular G4 structures with diverse topologies, were analyzed. *E. coli* DinG represents an attractive model system for the function of the more challenging and medically relevant eukaryotic homologs, besides being itself a useful target, due to its presence in a number of widespread bacterial pathogens.

Therefore, my project was focused on the following milestones:

- Produce large amounts of pure recombinant proteins; these included the catalytic domain of human RecQ helicases (RecQ1, BLM, WRN, RecQ4, and RecQ5) encompassing the catalytic core, fragments of the N-terminal domain of RecQ4 and the full-length *E. coli* DinG.
- Perform site-direct mutagenesis of BLM and WRN helicases. Express and purify these mutants.
- Prepare and characterize various G4 structures, from different genomic origins and with diverse topologies.
- Carry out a systematic biochemical analysis testing the nucleic acid binding preferences and helicase activities towards several canonical and, mainly, non-canonical substrates.
- Perform structural analysis, using X-ray crystallography, to obtain medium or high-resolution data of G4:helicase complexes, enhancing our understanding of the molecular details of their interactions.

The proposed work should provide critical insights into the roles of both RecQ helicases and G-quadruplexes in cancer onset, progression and therapy resistance. A deeper understanding of the interplay between G4s and helicases may provide novel therapeutic possibilities and could aid in the development of novel anticancer drugs.

Materials and Methods

7. Construct design and cloning

7.1. Cloning strategy used: Ligation Independent Cloning

Ligation Independent Cloning (LIC) was the method used for cloning the RecQ4 fragments into the selected vectors. This technique, developed at the Structural Genomic consortium at Oxford, obviates the need for the time-consuming ligation step of traditional cloning methods. Indeed, in traditional cloning, base-pairing in the short overlapping regions (usually 4 bp) does not provide enough stability to hold the plasmid together through the transformation/replication process. LIC employs long overhangs to form a stable association between fragments, allowing for transformation without ligation. It uses the 3'→5' exonuclease activity of T4 DNA Polymerase to create overhangs of varying length (typically 10-12 bp) based on a specific sequence, with complementarity between the vector and insert (Gileadi et al., 2008). Complementary overhangs in the gene of interest are generated during PCR by incorporating appropriate extensions into the primers and subsequently treating them with T4 DNA polymerase. The annealed but nicked vector product is then repaired during the replication cycle thanks to the incorporation of dGTP, which limits in the reaction the exonuclease processing to the first complementary C residue. Circularization can occur between vector molecules and PCR fragments through the 12 nucleotide cohesive ends, but not in mixtures lacking insert fragments.

7.2. Expression vectors used for cloning human RecQ fragments and *E. coli* DinG

The bacterial expression vector pNH-TrxT, kindly provided by Opher Gileadi-Structural Consortium, Oxford, was used to obtain the DNA constructs for the expression of recombinant proteins in *E. coli*. The RecQ fragments I worked with (RecQ1, BLM, WRN, RecQ4 and RecQ5) and *E. coli* DinG had been previously cloned into different vectors: the pNIC-CTHF (provided by O. Gileadi or Pavel Savitsky) for RecQ1, BLM, WRN and RecQ5 fragments, the pNH-TrxT for RecQ4 fragments, and the pNIC28-Bsa4 for DinG. The main features of all the vectors used are summarized in Table 3.

Vector	Promoter	Antibiotic Resistance	Tag	Protease	Size (bp)
pNH-TrxT	LacI	Kanamycin	6His-TrxT	TEV	7602
pNIC-CTHF	T7	Kanamycin	C-term. 6His-FLAG	TEV	7260
pNIC28-Bsa4	T7	Kanamycin	6His	TEV	7284

Table 3. The vectors used and their main features. The vectors used and their characteristics: promoter, antibiotic resistance, fusion tag, protease used to cleave the tag and size. LIC = Ligation-Independent Cloning (Gileadi et al., 2008).

7.3. Insert preparation

DNA sequences were amplified by Polymerase Chain Reaction (PCR), using the Q5 high fidelity polymerase (NEB). The 50 μ L of PCR reaction mixture contained: 5-10 ng of template DNA, 500 ng of forward and reverse primers, 200 μ M of dNTPs mix, 0.3 μ L Q5 high fidelity polymerase (0.01U/ μ L) (NEB) and 1X Q5 Reaction Buffer (NEB). The PCR program consisted of cycles of denaturation, annealing, and extension, followed by a final extension and cooling steps.

The PCR program used was the following:

- 30 s initial denaturation 98°C

Then 35 cycles:

- 30 s denaturation 98°C
- 30 s annealing 55°C
- 60 s extension 72°C

At the end of the cycles:

- 120 s final extension 72°C
- Cooling at 10°C

The primer sequences used to amplify N-terminal fragment of RecQ4 are listed in Table 4.

N-terminal RecQ4 Fragment	Primer	Sequence (5'-3')
1-55	RecQ4_1-55_Fp	TACTTCCAATCCATGGAGCGGCTCCGGGAC
	RecQ4_1-55_Rv	TATCCACCTTTACTGTCTCAGGCCTGGCCCGTGGTACGCTT
1-133	RecQ4_1-133_Fp	TACTTCCAATCCATGGAGCGGCTCCGGGAC
	RecQ4_1-133_Rv	TATCCACCTTTACTGTCTCAGGCCTCTTCCTAGAGGCCA
1-427	RecQ4_1-427_Fp	TACTTCCAATCCATGGAGCGGCTCCGGGAC
	RecQ4_1-427_Rv	TATCCACCTTTACTGTCTCAGGCCTCTTCCTAGAGGCCA

Table 4. List of primers used for the cloning of different N-terminal fragments of RecQ4. Fp: Forward primer, Rv: Reverse primer.

The amplified PCR products were analysed by agarose gel electrophoresis. The agarose gel was prepared dissolving 1% w/v agarose in Tris-Acetate-EDTA (TAE) Buffer (40 mM Tris-acetate, 1 mM Ethylenediaminetetraacetic acid pH 8.0), run in TAE running buffer at 90 V for 40 minutes. The DNA insert was purified using the PCR purification kit (Macherey-Nagel) according to the manufacturer's instructions. Their concentration was determined by measuring the UV absorption at a wavelength of 260nm with a Nanodrop 2000 spectrophotometer (Thermo Scientific).

Purified PCR products were treated with T4 DNA polymerase in the presence of dCTP or dGTP.

The procedure for treating PCR products with T4 was the following (20 µL reaction):

- 0.2 pmol PCR product
- 1x NEB Buffer 2
- 0.1 µL of BSA (1 mg/ml)
- 2.5 mM dCTP/dGTP
- 0.5 mM DTT (100mM)
- 0.15 U/µL of T4 DNA polymerase
- H₂O to reach final volume of 20 µL

The reaction is incubated at 22°C for 30 minutes, followed by inactivation at 75°C for 20 minutes.

7.4. Vector preparation

The SGC-made LIC vectors (Gileadi et al., 2008) all contain the SacB gene flanked by two BsaI sites. These restriction sites are used to linearize the vector and to remove the SacB gene, allowing the negative selection on 5% sucrose. 5 µg of each vector were digested with BsaI enzyme according to the manufacturer's protocol (NEB) and then purified using the PCR purification kit

(Macherey-Nagel) and eluted in Tris EDTA (TE) buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0). The digested and purified vector was further treated with T4 DNA polymerase (NEB) in the presence of dGTP. Because of the 3'-5' exonuclease activity of the polymerase, the bases are removed from both 3'-ends until the first guanine (G), when the polymerase activity ensures the guanine incorporation, avoiding the further digestion of the vector.

According to the protocol, 600 ng of digested LIC vector is treated with T4 DNA polymerase to produce enough vector for approximately 20 annealing reactions.

Reaction components are:

- 600 ng of digested vector
- 1x NEB Buffer 2
- 0.1 µL of BSA (1 mg/ml)
- 2.5 mM dGTP
- 0.5 mM DTT (100 mM)
- 0.15 U/µL of T4 DNA polymerase
- H₂O to reach final volume of 20 µL

The reaction is incubated at 22°C for 30 minutes, followed by inactivation at 75°C for 20 minutes.

7.5. Ligation and transformation

To facilitate the annealing of the T4 polymerase-treated vector and insert, we combined the vector and insert in different ratios (1:1, 1:2, and 1:3): 1 µL of treated vector was mixed with 1-3 µL of treated insert. Annealing reactions were incubated 20 minutes at room temperature (RT) then transferred to ice. Subsequently, the ligation mixture was used for the transformation with 50 µL of competent *E. coli* DH5α cells using the heat shock method: the mixture was incubated on ice 30 minutes, then up to 42°C for 42 seconds and at the end in ice for 3 minutes. Transformed cells were plated onto small Petri dishes of LB containing Kanamycin 50 µg/mL and 5% sucrose. The presence of sucrose is essential to positively select the cells containing the cloned vectors. Only empty vectors, without the insert, contain an intact *sacB* gene, which encodes for an enzyme that converts the sucrose present in the plates into levan, a toxic compound that accumulates in periplasm of *E. coli* cells, causing cell death (Gay et al., 1983). After overnight (ON) incubation at 37°C, the only visible colonies on the plates should correspond to cells containing cloned vectors.

7.6. Preparation of plasmid minipreps

Single colonies from agar plates were inoculated into 10 ml LB broth medium supplemented with the appropriate antibiotics and the cultures were grown for 16 hours at 37°C, at 200 rpm. The cells were pelleted at 3,500 x g for 15 minutes at 4°C. Plasmid DNA was purified from cell pellets using the E.Z.N.A. Plasmid DNA Mini Kit (OMEGA Bio-Tek), according to the manufacturer's protocol, and stored at -20°C. To verify the presence of the insert, vectors were checked by PCR amplification of the DNA construct using the appropriate primers. The PCR program was the same as the one described above, used to amplify the DNA constructs (paragraph 1.3). PCR products were analysed by agarose gel electrophoresis, prepared as previously described. The average yield of plasmid DNA was around 50 ng/μL, checked spectrophotometrically measuring the absorbance at a wavelength of 260 nm using a Nanodrop 2000 (Thermo Scientific).

7.7. Bioinformatic analysis for guanine-specific binding pocket identification

To identify the presence of a guanine-specific binding pocket (GSP; Voter et al., 2018) in human RecQ helicases and the residues which could be involved in G-quadruplex interaction, a structural superimposition was performed using PyMOL software. The resolved crystal structures of the bacterial RecQ (PDB ID: 6CRM, Voter et al., 2018) was aligned with the crystal structure of each human RecQ helicase (RecQ1, PDB ID: 2WWY, Pike et al., 2015; BLM, PDB ID: 4CGZ, Newman et al., 2015; WRN, PDB ID: 6YHR, Newman et al., 2021; RecQ4, PDB ID: 5LST, Kaiser et al., 2017; RecQ5, PDB ID: 5LB8, Newman et al., 2017).

Conservation analysis is one of the most widely used methods for predicting functionally important residues in protein sequences. In order to look for sequence homology among homologues, a first database search was carried out by means of BLAST algorithm (<https://blast.ncbi.nlm.nih.gov>), using the human RecQ helicase fragments sequences as search model and default parameters. Multiple sequence alignments were generated with Clustal Omega (EMBL-EBI, <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) program, comparing the protein sequences of human RecQ1, BLM, WRN, RecQ4 or RecQ5 with the relative sequences of similar length (in FASTA format) from different organism, including *Drosophila melanogaster*, *Danio rerio* (zebrafish), *Caenorhabditis elegans*, *Xenopus laevis*, *Aplysia californica*, and *Hydra vulgaris*. From the output of the sequence alignment, it was possible to identify the degree of conservation of the newly identified residues through different RecQ species.

7.8. Site directed mutagenesis of BLM and WRN

In order to understand the role of some conserved residues previously predicted to be functionally important in G-quadruplex interaction, site-direct mutagenesis was used to mutate critical residues in the pNIC-CTHF vectors containing BLM fragment 636-1298 and WRN fragment 517-1093. Mutants generated are listed in Table 5.

Mutant	Primer	Sequence (5'-3')
BLM S897A	BLM_Ser897Ala_Fp	TTACTGCCTC GCG AGGCGAGAAT
	BLM_Ser897Ala_Rv	ATTATCCCTGAATCATATGGG
BLM S965A	BLM_Ser965Ala_Fp	GATTCATGCA GCG CTCCCTAAATC
	BLM_Ser965Ala_Rv	ACAAATCGCACGTCC
WRN S773A	WRN_Ser773Ala_Fp	CTACTGTCCT GCG AGAAAAATGAC
	WRN_Ser773Ala_Rv	ATGATTGTTGGACCTTC
WRN D872A	WRN_Asp872Ala_Fp	GGCTCCTGCA GCG ATTA ACTTAAATAG
	WRN_Asp872Ala_Rv	CAGAGGACGTGACAAG

Table 5. List of mutants and primers used for site-direct mutagenesis. The introduced mutations are shown in red. Fp: Forward primer, Rv: Reverse primer.

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 μ L PCR reaction mixture containing 100 ng of template DNA, 500 ng of forward and reverse primers, 10 μ M of dNTPs mix, 0.3 μ L Q5 high fidelity polymerase (0.01 U/ μ L) (NEB) and 1X Q5 Reaction Buffer (NEB). The PCR program used was the following:

- 30 s initial denaturation 98°C

Then 25 cycles:

- 10 s denaturation 98°C
- 30 s annealing 61°C
- 300 s extension 72°C

At the end of the cycles:

- 120 s final extension 72°C
- Cooling at 10°C

PCR product was checked on EtBr-stained agarose gel (0.7% w/v agarose). The methylated parental plasmid was digested using DpnI enzyme for 2 hrs at 37 °C, treated with T4 Polynucleotide Kinase (New England Biolabs) and T4 ligase (New England Biolabs) for 20 min at RT, and transformed into

DH5 α cells using the heat-shock method. Transformed cells were plated onto Petri dishes of LB agar containing Kanamycin 50 $\mu\text{g/mL}$ and, after overnight incubation at 37°C, colonies on the plates should correspond to cells containing mutant vectors. All clones were confirmed by sequencing.

For plasmid miniprep preparation, single colonies from agar plates were inoculated into 10 ml LB broth medium supplemented with the appropriate antibiotics and the cultures were grown for 16 hours at 37°C, at 200 rpm. The cells were pelleted at 3,500 x g for 15 minutes at 4°C. Plasmid DNA was purified from cell pellets using the E.Z.N.A. Plasmid DNA Mini Kit (OMEGA Bio-Tek), according to the manufacturer's protocol, and stored at -20°C. The average yield of plasmid DNA was around 150 ng/ μL , checked spectrophotometrically measuring the absorbance at a wavelength of 260 nm using a Nanodrop 2000 (Thermo Scientific).

8. Protein expression in bacterial *E. coli* cells

Before attempting protein large-scale expression, the optimization of growth conditions and protein purification were carried out using small scale batch purification, as described below. Various conditions were tested, including different *E. coli* strains, induction methods and growth temperatures.

8.1. Test expression and solubility assay

E. coli expression strains used:

- Rosetta 2 (DE3) (*Novagen*): derived from BL21 strain, they are designed to enhance the IPTG induced expression of eukaryotic proteins that contain codons rarely used in *E. coli*. This strain also provides the tRNAs for seven rare codons (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) in a plasmid (pRARE) containing the chloramphenicol resistance.
- Rosetta 2 (DE3) pLysS (*Novagen*): variant containing an additional plasmid providing rare codons, limiting the basal expression of target genes.
- BL21 (DE3) (*Stratagene*): basic strain containing the λ DE3 lysogen which carries the gene for T7 RNA polymerase under control of the lacUV5 promoter allowing high-level expression of T7-regulated genes induced by IPTG induction. It lacks both lon protease and ompT protease that can degrade the expressed proteins.
- BL21 (DE3) Star (*Thermo Fisher Scientific*): BL21 cells, containing a mutated rne gene (rne131) which encodes a truncated RNase E enzyme that lacks the C-terminal 477 amino acids of the

enzyme required for mRNA degradation. Thus, the strain lacks the ability to degrade mRNA, resulting in an increase in mRNA stability.

- BL21 (DE3) Star pRARE2: phage-resistant derivative of BL21 (DE3), which has been transformed with plasmid pRARE2, which carries seven rare codon tRNA genes.
- B834 (*Stratagene*): they are the parental strain of BL21 (DE3) and are a methionine auxotroph, usually used to label proteins with selenomethionine. They also lack lon and ompT proteases.
- LOBSTR (low background strain) (*Thermo Fisher Scientific*): derived from the BL21 (DE3) strain, it carries genomically modified copies of *arnA* and *slyD*, whose protein products exhibit reduced affinities to Ni and Co resins, resulting in a much higher purity of the target protein. This strain has been optimized for one-step downstream polyhistidine-tag affinity purification and is ideal for poorly expressing proteins.

Vector containing RecQ4 DNA fragments 1-55 and 1-133 were transformed into BL21 Star strain cell. Vector containing RecQ4 DNA fragment 1-427 was transformed into BL21, BL21 Star, BL21 Star pRARE2 and Rosetta 2 strain cells. Vector containing BLM mutants were transformed into BL21 Star pRARE2 strain cell, while vector containing WRN mutants were transformed into LOBSTR strain cell.

For small scale expression, classical induction was performed by inoculating 10-50 mL of Luria Broth (LB) or Terrific broth (TB), supplemented with proper antibiotics, with single colonies of *E. coli* cells harboring the plasmid tested or with plasmid minipreps. Protein expression was induced with IPTG at an OD₆₀₀ of 0.5-1.2 for LB media and 1.6-2.2 for TB media. IPTG concentrations of 0.5 mM and 1 mM were used; induction temperatures of 18 °C was tested.

E. coli cells harboring the plasmid with RecQ4 fragments 1-55 or 1-133, and with BLM or WRN mutants, were used to inoculate 50 mL of TB in 250 ml flasks, to allow sufficient aeration during growth. Instead, *E. coli* cells harboring the plasmid with RecQ4 fragment 1-427 was used to inoculate 50 mL of LB in the same conditions. For all the RecQ4 fragments, protein expression was induced by addition of 0.5 mM IPTG, at 18°C ON. BLM and WRN mutants protein expression were induced by addition of 0.2 mM and 0.1 mM IPTG, respectively, at 18°C overnight.

For large-scale growth, the optimized conditions were scaled to volumes ranging from 1L to 6L, depending on the expression levels.

8.2. Small scale expression test and purification

After protein expression, the pellet of *E. coli* cells was resuspended in 1-10 mL of Lysis buffer (50 mM Tris-HCl pH 7.5, 0.5-1 M NaCl, 10 mM Imidazole, 10% glycerol, 0.5 mg/mL DNase (Life Technologies), 1 mM PMSF (4-(2-aminoethyl)benzenesulfonyl fluoride, Sigma), 2 µg/ml Aprotinin (Sigma), 4 µg/ml Leupeptin (Sigma), 0.7 µg/ml Pepstatin A (Sigma), 1 mM TCEP).

The resuspended pellet was agitated on wheel for 15 minutes and then cells were disrupted using a sonicator (Bandelin® SonoPlus ultrasonic cell disruptor – KE 76) (10” sonication, 20” pause, for two times), obtaining the Total Extract. The Total extract was ultra-centrifuged at 16000 x g for 30 min and the soluble fraction was then incubated with 50 µL of Ni-NTA agarose resin (Qiagen) to perform a His-tagged protein purification in batch. Before use, the resin was prepared according to the manufacturer’s instructions: pre-washed twice in water and equilibrated once in Lysis buffer. After 1 hour of incubation at 4°C, the suspension was centrifuged at 500 x g for 15 minutes at 4°C to sediment the resin. The supernatant was removed (unbound sample) and the resin was washed three times with 10 CV of Lysis buffer, followed by centrifugation at 500 x g for 5 min at 4°C. The resin was then incubated with 1.5 CV (column volume) Elution buffer (50 mM Tris pH 7.0, 0.5-1 M NaCl, 300 mM imidazole, 10% glycerol, 1 mM TCEP) at 4° for 10 min, and then spun down for 10 mins at 500 x g at 4°C as above. The elution fractions were then collected and analysed.

9. Purification technique used for Protein Large-scale purification

9.1. Ni-affinity Chromatography or Immobilized Metal Ion Affinity Chromatography (IMAC)

IMAC is based on the interaction of certain protein residues (histidines, cysteines, and to some extent tryptophans) with cations of transition metals. Proteins fused to a 6 histidine (6His) tag can be purified using a Ni-affinity resin (Qiagen), that uses nitrilotriacetic acid (NTA), a tetradenate chelating ligand, in a highly cross-linked 6% agarose matrix. Nickel generally provides good binding efficiency to His-tagged proteins but also binds non-specifically to endogenous proteins that may contain histidine clusters. For this reason, a small amount of imidazole is added in the binding/wash buffer since it helps to control off-target binding before the elution of proteins.

9.2. Heparin affinity chromatography

Heparin is a negatively charged polydispersed linear polysaccharide which has the ability to bind a wide range of biomolecules and can work as a high-capacity cation exchanger due to its anionic sulphate group. It can therefore be used for affinity chromatography, being particularly useful for isolating DNA binding proteins, where the interaction often involves charged residues. HiTrap Heparin column HP (Cytiva) is prepacked with Heparin Sepharose® High Performance for purification of proteins with affinity for heparin (e.g., DNA binding proteins, coagulation factors, etc.).

9.3. Tag removal and dialysis

Proteins expressed from pNH-TrxT and pNIC-CTHF expression vectors were subjected to removal of Trx-6His-tag and of C6His-MYC-tag, respectively, by TEV (Tobacco Etch Virus) protease, which recognizes a cleavable site on the fusion tag. The optimal temperature for cleavage is 30°C; however, the enzyme is active over wide ranges of temperature (4°C - 30°C) and pH (7.0 - 9.0). The cleavage was performed by dialyzing overnight the protein solution mixed to 6His-TEV protease against the TEV buffer (20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% Glycerol, and 1 mM DTT) at 1:20 ratio between protein and TEV. This also helps to reduce the concentration of imidazole contained in the Ni-NTA-eluted sample, since the imidazole is an inhibitor of TEV protease. The TEV Buffer was 200 times the volume of samples (i.e. 10 ml of samples and 2 L of dialysis buffer). The membrane (SnakeSkin™ Dialysis Tubing, Thermo Scientific) used allows the diffusion of molecules up to 10 kDa. Following digestion, TEV Protease, uncleaved protein and tags were removed from the cleavage reaction by Ni-NTA fastflow resin (Qiagen) using the polyhistidine tag at the N-terminus of the tag. Cleavage was checked by SDS-PAGE gel.

9.4. Cation Exchange chromatography

Ion Exchange (IEX) chromatography can separate molecules or groups of molecules that have only slight differences in charge. Separation is based on the reversible interaction between a charged molecule and an oppositely charged chromatography medium. An IEX medium comprises a matrix of spherical particles substituted with charged groups covalently bound that are negatively (strong sulfopropyl [SP]) or positively (strong quaternary ammonium [Q]) charged: in the case of cationic exchange the matrix is negative in order to bind proteins with a surface positive charge. The matrix is usually porous to provide a high internal surface area. The medium is packed into a column to form

a packed bed. The bed is then equilibrated with buffer which fills the pores of the matrix and the space in between the particles. Typically, conditions are selected to ensure that the molecules of interest bind to the medium as they are loaded onto the column. Protein's net surface charge is determined by the relationship between the pH of the solution and the pI of the protein and will change gradually as the pH of the environment changes. Each protein has its own unique net charge versus pH relationship which can be visualized as a titration curve. When the pH value is below the pI value, a major part of amino acids is protonated; on the other hand, when the pH value is above the pI value, the net charge of the protein is negative. However, beside the overall charge, the interactions with the matrix are determined by local clusters of charge, which critically depends on each protein structure. The stronger is the net charge of the proteins, the higher is the salt concentration needed to elute them. Elution was performed with salt gradient in both cases.

9.5. Size Exclusion Chromatography (SEC)

Size Exclusion chromatography is a powerful purification technique that allows to separate proteins according to their molecular weight, by filtration through a gel matrix (assuming a globular conformation). The Increased Superdex™ 200 10/300 GL column (© GE Healthcare) used for protein purification is made of specific matrix, composed by spherical beads of dextran and/or agarose, 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, generating an elution volume inversely proportional to protein molecular weight. Thus, SEC can also be used for estimating the protein size. For the determination of the molecular weight of the protein, a standard calibration curve was plotted as follows: for each calibration standard, the logarithm of its known molecular weight [$\log(MW)$] was plotted against its normalized elution volume (V_e/V_0), generating the following linear equation

$$\log(MW) = m \frac{V_e}{V_0} + b$$

where m is the slope and b is the intercept. The equation was then used for the determination of the $\log(MW)$ of the protein of interest based on its elution volume (V_e).

10. Large-scale expression and protein purification

The expression protocols used for all the proteins studied in the present work are described in the following Table 6.

Construct	Cell line	medium	OD ₆₀₀	Induction conditions
RecQ1	Rosetta 2	TB	1.5-2	0.2 mM IPTG ON 18°C
BLM	BL21 Star pRARE2	TB	2.5-3	0.2 mM IPTG ON 18°C
BLM S897A	Rosetta 2	TB	2.5-3	0.2 mM IPTG ON 18°C
BLM S965A	BL21 Star pRARE2	TB	2.5-3	0.2 mM IPTG ON 18°C
WRN	LOBSTR	TB	2-3	0.1 mM IPTG ON 18°C
RecQ4 371-1116	BL21 Star pRARE2	TB	2-3	0.2 mM IPTG ON 18°C
RecQ4 1-55	BL21 Star	TB	1.5	0.5 mM IPTG ON 18°C
RecQ4 1-133	BL21 Star	TB	1.5	0.5 mM IPTG ON 18°C
RecQ4 1-427	BL21	LB	0.8	0.5 mM IPTG ON 18°C
RecQ5	BL21 Star pRARE2	TB	2-3	0.2 mM IPTG ON 18°C
<i>E. coli</i> DinG	B834	LB (+10 μ M FeCl ₃)	0.8	0.15 mM IPTG ON 18°C

Table 6. Details of construct and final expression protocol.

The recombinant RecQ1, BLM, WRN and RecQ5 were expressed and purified as described previously (Pike et al., 2009, Newman et al., 2015, Newman et al., 2020, Newman et al., 2017), with some variations.

10.1. RecQ1 purification

For purification of RecQ1 protein, a 2 L pellet was resuspended in 100 ml Lysis buffer (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 10 mM imidazole, 5 % glycerol), supplemented with 1 mM TCEP, 1 mM PMSF, 2 µg/ml Aprotinin, 4 µg/ml Leupeptin, 0.7 µg/ml Pepstatin A and 0.1 mg/ml Lysozyme. The cells were disrupted with a sonicator (4 round 40% amplitude, 5' on and 1 min pause) and then incubated 20 mins at RT with 0.1 mg/ml DNase I supplemented with 5 mM MgCl₂. The lysate was clarified by centrifugation for 1 hour at 14,000 rpm, and the supernatant, representing the soluble fraction, was incubated with 1 mL of Ni-affinity resin (Qiagen). After 2 hours of incubation at 4°C, the resin was collected by gravity using a bench column and then washed with 20 resin volumes of Lysis buffer, 5 resin volume of 2M NaCl to eliminate the DNA in excess and 20 resin volumes of washing buffer (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 30 mM imidazole, 5 % glycerol). The protein was then eluted with a volume of Elution Buffer (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 500 mM imidazole, 5 % glycerol) equal to 2 times the volume of the resin used. The incubation time with the elution buffer was 10 minutes, and the elution fraction was collected. The protein elution was repeated four times. Following Ni-NTA elution, the sample was concentrated using Amicon Ultra 15 mL centrifugal filters (Millipore) with a 30 kDa molecular weight cut-off. SEC was carried out on the analytic ÄKTA system using an Increase Superdex™ 200 10/300 GL column (© GE healthcare). The SEC buffer comprises 50 mM Hepes-NaOH pH 7.5, 250 mM NaCl, 5% glycerol, 1 mM TCEP.

10.2. BLM purification

2 L culture pellet of BLM protein was resuspended in 100 mL Lysis Buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol), supplemented with 1 mM PMSF, 4 µg/ml Leupeptin, 2 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A, 1 mM TCEP. After the pellet was stirred and completely resuspended for about 30 min at 4°C, cells were lysed using the sonicator (5 rounds 40% amplitude, 5 sec on 1 min off). After sonication, 20 units/ml of Benzonase was added to the pellet and kept in agitation for 15 min at RT. The final lysate was divided up into 50 mL beakers and clarified by centrifugation for 30 min at 30,000 g. The supernatant, representing the soluble fraction, was loaded on a HisTrap FF crude 5 mL column (Cytiva) using the analytic ÄKTA system, pre-equilibrated in Nickel buffer A (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 30 mM imidazole, 1 mM TCEP). Elution was done with 5 CV of an imidazole gradient created by mixing Nickel buffer A and Nickel buffer B (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 250 mM imidazole, 1 mM TCEP). The eluted fractions from IMAC were diluted to reduce salt concentration to 250 mM NaCl and loaded onto a HiTrap Heparin column HP 5 mL (Cytiva) within an ÄKTA purifier system (GE Healthcare) for

Heparin chromatography. In this case elution was done with a salt gradient created by mixing two buffers: Heparin buffer A (50 mM Hepes-NaOH pH 7.5, 5% glycerol, 1 mM TCEP) and Buffer B (50 mM Hepes-NaOH pH 7.5, 5% glycerol, 1 M NaCl, 1 mM TCEP). The eluted fractions from Heparin were further analysed through SDS-PAGE. Pure protein eluted at $\approx$ 500 mM salt concentrations. The C-terminal fusion tag (His-FLAG) was then removed from BLM protein using the TEV protease. The cleavage was performed by dialyzing overnight the protein solution mixed to 6His-TEV protease against the TEV buffer (20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT) at 1:20 ratio between protein and TEV. The Ni-affinity resin was used to retain both the tag and the TEV protease, while the protein was collected in the flow-through. It is crucial at this point to add 30 mM imidazole to protein solution before Ni-NTA incubation. The flow-through and the eluted fractions were further analysed through SDS-PAGE. As a final polishing step, protein was concentrated and injected into a Increased Superdex™ 200 26/60 GL column (© GE healthcare) for Size Exclusion chromatography (SEC buffer: 10 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM TCEP).

10.3. BLM S897A and BLM S965A purifications

500 mL culture pellet of BLM S897A and BLM S965A proteins was resuspended in 30 mL Lysis Buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol), supplemented with 1 mM PMSF, 4 μ g/ml Leupeptin, 2 μ g/ml Aprotinin, 0.7 μ g/ml Pepstatin A, 1 mM TCEP. The purification protocol followed exactly the one used for BLM WT protein (described in the previous section 10.2.).

10.4. WRN purification

4 L culture pellet of WRN protein was resuspended in 100 mL Lysis Buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol, 10 mM imidazole), supplemented with 1 mM PMSF, 4 μ g/ml Leupeptin, 2 μ g/ml Aprotinin, 0.7 μ g/ml Pepstatin A, 1 mM TCEP, 0.1 mg/ml Lysozyme. After the pellet was stirred and completely resuspended for about 30 min at 4°C, cells were lysed using the sonicator (5 rounds 45% amplitude, 5 sec on 1 min off). After sonication, 20 units/ml of Benzonase was added to the pellet and kept in agitation for 15 min at RT. The final lysate was clarified by centrifugation for 30 min at 13,000 g. The supernatant, representing the soluble fraction, was incubated with 800 μ L of Ni-affinity resin (Qiagen). After 1h 30 of incubation at 4°C, the resin was collected by gravity using a bench column and then washed with 15 resin volumes of Lysis buffer, 5 resin volume of 2M NaCl to eliminate the DNA in excess and 10 resin volumes of washing buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol, 45 mM imidazole). The protein was then eluted with 1.5 resin volume of

Elution Buffer (50 mM Hepes pH 7.5, 500 mM NaCl, 5% glycerol, 300 mM imidazole), repeated four times. The incubation time with the elution buffer was 10 minutes, and the elution fraction was collected. The protein elution was repeated four times. The eluted fractions were diluted to reduce salt concentration to 250 mM NaCl and loaded onto a HiTrap Heparin column HP 1 mL (Cytiva) within an ÄKTA purifier system (GE Healthcare). Elution was done with a salt gradient created by mixing two buffers: Heparin buffer A (25 mM Hepes-NaOH pH 7.5, 250 mM NaCl, 2.5% glycerol, 1 mM TCEP) and Buffer B (25 mM Hepes-NaOH pH 7.5, 2.5% glycerol, 1 M NaCl, 1 mM TCEP). The eluted fractions from Heparin were further analysed through SDS-PAGE. Pure protein eluted at $\approx$ 500 mM salt concentrations. The C-terminal fusion tag (His-FLAG) was then removed from WRN protein using the TEV protease. The cleavage was performed by dialyzing overnight the protein solution mixed to 6His-TEV protease against the TEV buffer (20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM DTT) at 1:20 ratio between protein and TEV. The Ni-affinity resin was used to retain both the tag and the TEV protease, while the protein was collected in the flow-through. Similarly to BLM, it is crucial at this point to add 30 mM imidazole to protein solution before Ni-NTA incubation. As a final polishing step, protein was concentrated and injected into a Increased Superdex™ 200 10/300 GL column (© GE healthcare) for Size Exclusion chromatography (SEC buffer: 20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP).

10.5. RecQ4 fragments purifications

10.5.1. RecQ4 fragment 371-1116

RecQ4 371-1116 fragment, encompassing the catalytic core of the helicase and part of the N-terminal domain (with an extra helix discovered from bioinformatic studies) was expressed and purified with the help of another PhD student from the laboratory. 3 L cell pellet was resuspended in 200 mL of Lysis buffer (50 mM Hepes-NaOH pH 8.0, 300 mM NaCl, 10% Glycerol, 1 mM MgCl₂, 15 mM imidazole), complemented with 1 mM PMSF, 2 µg/ml Aprotinin, 4 µg/ml Leupeptin, 0.7 µg/ml Pepstatin A, 1 mM TCEP, 0.1 mg/ml Lysozyme. After cell disruption with sonicator (30 rounds 45% amplitude, 10 sec on and 50 sec off) the crude extract was incubated with 0.1 mg/ml DNase I at RT for 10 mins and then centrifuged at 4 °C for 1 h at 30,000 g.

The supernatant was incubated ON with 1.5 mL Ni-affinity resin (Qiagen) pre-equilibrated with Lysis Buffer. The column was washed with 20 CV of Lysis buffer, then with 5 CV of Lysis buffer supplemented with 2M NaCl and again with 20 CV Lysis buffer to decrease salt concentration. The protein was eluted with 1.5 CV (repeated five times) of Lysis buffer supplemented with 300 mM imidazole. The eluate was collected, and the salt concentration was reduced to 50 mM NaCl. The

solution was applied to a Capto HiRes S 5/50 1 mL column, pre-equilibrated with MonoS buffer (20 mM HEPES pH 7.2, 50 mM NaCl, 10% Glycerol, 1 mM MgCl₂, 1 mM TCEP). Protein was eluted using a linear gradient over 40 CV up to a concentration of 300 mM NaCl.

Appropriate fractions were pooled, and the Trx-6xHis-tag was removed by batch incubation with TEV protease in an overnight reaction at 4°C. After tag removal, as a final polishing step, the solution was concentrated and applied to an analytical Superdex 200 Increase 10/300GL Size Exclusion column, using protein-storage buffer (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% Glycerol, 1 mM MgCl₂, 1 mM TCEP). Appropriate fractions were pooled and concentrated. Protein was flash frozen in liquid nitrogen and stored at -80 °C.

10.5.2. RecQ4 fragment 1-55

500 mL pellet obtained from the expression of RecQ4 fragment 1-55 was resuspended in 20 mL Lysis buffer (20 mM Hepes-NaOH pH 8.0, 300 mM NaCl, 15 mM imidazole, 20% Glycerol, 2.5 mM MgCl₂), supplemented with 1 mM TCEP, 1 mM PMSF, 4 µg/ml Leupeptin, 2 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A and 1 mg/ml Lysozyme, and cells were lysed by sonication (4 round 45% amplitude, 5' on 30' pause). After sonication, 0.1 mg/ml DNase I was added to the pellet and kept in agitation for 15 min at RT. The lysate was clarified by centrifugation 30 mins at 13,000 xg and the supernatant was incubated with 1 mL Ni-affinity resin (Qiagen). After 1h of incubation at 4°C, the resin was collected by gravity using a bench column and then washed with 10 resin volumes of Lysis buffer, 5 resin volume of 2M NaCl to eliminate the DNA in excess and 10 resin volumes of PBS 1x. The protein was then eluted with 1.5 resin volume of Elution Buffer (20 mM Hepes pH 8.0, 300 mM NaCl, 300 mM imidazole, 20% Glycerol, 2.5 mM MgCl₂), repeated for three times. The eluted fraction was then diluted to a final concentration of 50 mM NaCl and loaded onto Hi Trap SP FF 5 mL column (Cytiva) to perform a cation exchange chromatography. Elution was performed with a linear gradient mixing IEX buffer A (20 mM Hepes-NaOH pH 7.2, 10% Glycerol, 2.5 mM MgCl₂, 1 mM TCEP) and IEX buffer B (20 mM Hepes-NaOH pH 7.2, 10% Glycerol, 2.5 mM MgCl₂, 1 M NaCl, 1 mM TCEP). The eluted fractions from cation exchange chromatography were further analysed through SDS-PAGE. The appropriate fractions were pooled and concentrated to be injected into a Superdex™ 75 10/300 GL column (© GE healthcare) for Size Exclusion chromatography (SEC buffer: 20 mM Hepes-NaOH pH 7.5, 300 mM NaCl, 10% glycerol, 2.5 mM MgCl₂, 1 mM TCEP).

10.5.3. RecQ4 fragment 1-133

2 L culture pellet of RecQ4 fragment 1-133 was resuspended in 100 mL Lysis buffer (50 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% Glycerol), supplemented with 1 mM TCEP, 1

mM PMSF, 4 µg/ml Leupeptin, 2 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A and 1 mg/ml Lysozyme, and cells were lysed by sonication (4 round 45% amplitude, 5' on 30' pause). After sonication, 0.1 mg/ml DNase I was added to the pellet and kept in agitation for 15 min at RT. The lysate was clarified by centrifugation 30 mins at 12,000 xg and the supernatant was incubated with 1 mL Ni-affinity resin (Qiagen). After 1h of incubation at 4°C, the resin was collected by gravity using a bench column and then washed with 20 resin volumes of Lysis buffer and 5 resin volume of 2M NaCl to eliminate the DNA in excess. The protein was then eluted with 1.5 resin volume of Elution Buffer (50 mM Hepes pH 8.0, 150 mM NaCl, 300 mM imidazole, 5% glycerol), repeated for three times. The eluted fraction was then diluted to a final concentration of 50 mM NaCl and loaded onto Hi Trap SP FF 5 mL column (Cytiva) to perform a cation exchange chromatography, following the same protocol used for RecQ4 fragment 1-55 (section 10.5.2). The eluted fractions from cation exchange chromatography were further analysed through SDS-PAGE and the appropriate fractions were pooled and concentrated.

10.5.4. RecQ4 fragment 1-427

2L culture pellet of RecQ4 fragment 1-427 was resuspended in 100 mL Lysis buffer (20 mM Tris pH 8.0, 1 M NaCl, 5 mM imidazole), supplemented with 1 mM TCEP, 1 mM PMSF, 4 µg/ml Leupeptin, 2 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A and 1 mg/ml Lysozyme, and cells were lysed by sonication (4 round 45% amplitude, 5' on 30' pause). After sonication, 0.1 mg/ml DNase I was added to the pellet and kept in agitation for 15 min at RT. The lysate was clarified by centrifugation 30 mins at 12,000 g and the supernatant was incubated with 1 mL Ni-affinity resin (Qiagen). After 1h of incubation at 4°C, the resin was collected by gravity using a bench column and then washed with 20 resin volumes of Lysis buffer and 5 resin volume of 2M NaCl to eliminate the DNA in excess. The protein was then eluted with 1.5 resin volume of Elution Buffer (50 mM Hepes pH 8.0, 300 mM NaCl, 300 mM imidazole, 10% Glycerol, 2.5 mM MgCl₂), repeated for three times. The eluted fractions were diluted to reduce salt concentration to 50 mM NaCl and loaded onto a HiTrap Heparin column HP 1 mL (Cytiva) within an ÄKTA purifier system (GE Healthcare). Elution was done with a salt gradient created by mixing two buffers: Heparin buffer A (20 mM Tris pH 8.0, 20% Glycerol, 1 mM TCEP) and Buffer B (20 mM Tris pH 8.0, 20% Glycerol, 1 M NaCl, 1 mM TCEP). The eluted fractions from Heparin chromatography were further analysed through SDS-PAGE and the appropriate fractions were pooled and concentrated.

10.6. RecQ5 purification

RecQ5 purification protocol follows the same steps used for WRN purification (section 10.4). Briefly, a pellet from 2 L culture expressing RecQ5 protein was lysed, and the clarified lysate was purified by Ni-affinity resin with the same protocol used for WRN protein. Following Ni-NTA elution, the sample was diluted to reduce salt concentration to 250 mM NaCl and loaded onto a HiTrap Heparin column HP 1 mL (Cytiva). Elution was done with a salt gradient created by mixing Heparin buffer A and Heparin buffer B, as for WRN. This step was followed by overnight TEV cleavage to remove the C-terminal tag (His-FLAG), at a protease:protein ratio of 1:50 in TEV buffer (20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM DTT). Protein was recovered using reverse Ni-NTA chromatography, concentrated and injected into a Superdex™ 200 10/300 GL column (© GE healthcare) for Size Exclusion chromatography (SEC buffer: 20 mM Hepes-NaOH pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP).

10.7. *E. coli* DinG purification

2 L cell pellets were resuspended in 50 mL Lysis buffer (20 mM Tris-HCl pH 8.0, 5 mM imidazole, 1 M NaCl), complemented with 0.5 mM TCEP, 1 mM PMSF, 4 µg/ml Leupeptin, 2 µg/ml Aprotinin, 0.7 µg/ml Pepstatin A and 1 mg/ml Lysozyme and incubated for 15 min at 4°C. Cells were disrupted with sonicator (8 round 45% amplitude, 5' on 30' pause). 0.1 mg/ml DNase I was added and, after an incubation at RT for 15 min, the crude extract was clarified by centrifugation at 4 °C for 50 min at 18,000 g. The cell extract was incubated with 1 mL Ni-affinity resin (Qiagen) pre-equilibrated with Lysis Buffer. After 1 hr at 4°C, the resin was washed with 20 resin volume of Lysis Buffer and eluted in 10 resin volume with Elution buffer (20 mM Tris-HCl pH 8.0, 300 mM imidazole, 500 mM NaCl, 20% Glycerol). After elution, the salt concentration was reduced to 150 mM NaCl and the solution was applied to a 1 mL Heparin FF column (Cytiva), pre-equilibrated with Heparin buffer A (20 mM Tris pH 8.0, 20% Glycerol, 0.5 mM TCEP). The protein was eluted with a linear salt gradient over 40 CV, by mixing Heparin buffer A and Heparin buffer B (20 mM Tris pH 8.0, 20% Glycerol, 1M NaCl, 0.5 mM TCEP) up to a concentration of 700 mM NaCl. The eluted fractions from Heparin were further analysed through SDS-PAGE and the appropriate fractions were pooled and concentrated.

11. Biochemical and biophysical studies

11.1. 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 for each protein. Theoretical values of Molecular Weight (MW), isoelectric point (pI) and extinction coefficient at 280 nm of all purified proteins, either individually or as a complex, were calculated from the amino acid sequences using the Protparam tool on the Expasy server (<http://www.expasy.ch/tools/protparam.html>).

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

Protein samples were analysed by electrophoresis under denaturing, reducing conditions using either discontinuous SDS-polyacrylamide gels (4% for the stacking gel and 10%, 12% or 15% for the separating gel) or 4-20% gradient gels. Samples were prepared by addition of 4x SDS sample buffer (200 mM Tris-HCl pH 6.8, 8% SDS, 0.08% w/v Bromophenol Blue, 40% glycerol, 400 mM DTT) in a ratio of 1:3 and heating at 95°C for 5 min. Electrophoresis was carried out at a constant voltage (200V) for 45-55 min in 1x Tris-glycine-SDS buffer. Gels were stained with Instant Blue (Giotto Biotech), then stored in water.

11.3. Preparation of oligonucleotides

All oligonucleotides, purchased from Biomers (<https://www.biomers.net>), were chemically synthesized and purified by high pressure liquid chromatography (HPLC grade).

Lyophilized oligos were resuspended to a concentration of 100 μ M and stored at -20°C. Oligonucleotides were resuspended in Tris-EDTA (TE) buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, pH 8.0), supplemented with RNase inhibitor (New England Biolabs) in case of RNA substrates. G4 forming sequences oligonucleotides were resuspended in Phosphate buffer (10 mM KH_2PO_4 , 50 mM KCl, pH 7.5). Oligonucleotide sequences used in this study are listed in Table 7, 8, 9 and 10.

For fluorescent EMSA and helicase assays, the single oligonucleotide forming the ssDNA was 5'-end-labelled with 6FAM. Fork DNA substrate was prepared by annealing the 6FAM-labelled oligonucleotide to a 3-fold or 2.6-fold excess (gel-based helicase assay and FRET-based helicase

assay, respectively) of the unlabelled complementary strand or of the complementary quencher strand (gel-based helicase assay and FRET-based helicase assay, respectively) in Annealing buffer (20 mM Tris-HCl pH 7.5, 50 mM NaCl) by heating at 95°C for 5 min and then cooling through thermocycler, setting a temperature gradient of one degree each minute until 15°C. For D-loops, R-loops and bubbles the protocol was optimized, by annealing the 6FAM-labelled oligonucleotide to a 1.5-fold excess of the unlabelled complementary strand and to 2.5-fold excess of the invading strand in 6 mM Tris-HCl pH 7.5, 7 mM MgCl₂, 50 mM NaCl and 1 mM DTT. The reaction was heated at 95°C for 5 minutes, then incubated at 67°C for 1 hour, at 37°C for 30 minutes, and finally at 22°C for 2 hours and 30 minutes. The reaction was then run on a 6% native polyacrylamide gel in TBE buffer, purifying from gel the band corresponding to the annealed loop structure and then D/R-loops were eluted from gel slices by overnight dialysis in TE buffer.

G4 oligonucleotides used for binding assays were prepared resuspending the 5'-end-labelled-6FAM strands in 10 mM Tris-HCl pH 7.5, 150 mM KCl and 0.5 mM EDTA. The reaction was heated at 95°C for 5 minutes, then cooled by decreasing the temperature of 1°C every minute. The reaction was then incubated overnight at 4°C. G4 DNA systems used in the helicase assay were annealed at 1 µM concentration by preparing a mixture of 1 µM S-G4 (dabcyl-labelled oligonucleotide) and 0.85 µM Oligo FAM (FAM-labelled oligonucleotide) in two different buffers: Buffer K/Na (20 mM Tris-HCl pH 7.2, 5 mM MgCl₂, 1 mM KCl and 99 mM NaCl) and Buffer K (20 mM Tris-HCl pH 7.2, 5 mM MgCl₂, 100 mM KCl). The 1.2-fold excess of dabcyl-labelled strand was used in order to assure the total hybridisation of the FAM-labelled strand and therefore to achieve the maximum quenching of the fluorescent signal. Solutions were denatured at 95°C for 5 min and then slowly cooled to room temperature and stored at 4°C overnight before use.

Name	Sequence (5'-3')
D1L	*CTACCCCCACCCTCACAACCTTTTTTTTTTTTTTTTTTTTTT
D3L	TTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTGTGAGGGTGGGGGTAG
D2Q	TTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTGTGAGGGTGGGGGTAG*
D9	*GTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGCACTGGCC GTCGTTTACAAC
D11	GTTGTAAAACGACGGCCAGTGCCTTTTCCCAGCCTCAATCTCATCACTCTAGAGG ATCCCGGGTAC
D4L	GCTTGCATGCCTGCAGGCCAGCCTCAATCTCATCTTTTT
R4L	gcuugcaugccugcaggccagccucaaucucaucuuuuu
D4 5'	CTACTCTAACTCCGATTAAGCTTGCATGCCTGCAGG
R4 5'	cuacucuaacuccgauuaagcuugcaugccugcagg
D4B	GCTTGCATGCCTGCAGG
R4B	gcuugcaugccugcagg
Cap1	CTACTACCCCCACCCTCACAACC

Table 7. Oligonucleotides used to make EMSA and helicase assay substrates. D1L, D3L and D2Q were

used to prepare DNA fork. D9, D11, D4L, R4L, D4 5', R4 5', D4B and R4B were used to prepare D/R-loops and bubble. D1L is 5'-end-labelled with 6FAM (*); D2Q is 3'-end-labelled with BHQ-1 (*). Cap1 is the capture strand used to prevent the refolding of the fork. Lower case letters refer to RNA oligonucleotides.

Name	Sequence (5'-3')
cMYC	*TGAGGGTGGGTAGGGTGGGTAA
cKIT1	*TAAGGGAGGGCGCTGGGAGGAGGG
KRAS	*AGGGCGGTGTGGGAATAGGGAA
Mcm2	*GAAGGGACACGGAGGGGCGGGCCAGAGGGTCC
BU1A (+3.5)	*TAAGGGCTGGGTGGGTGCTGTGAAGGGCTGGG
ρ-globin	*TAAGGGGAGTAAAAGGGAGCGGGGTGCTGGG
Bcl2	*GGGCGCGGGAGGAATGGGCGGG
Bcl2 Mid	*ATAGGGCGCGGGAGGAAGGGGGCGGG
Tel ₂₃	*TAGGGTTAGGGTTAGGGTTAGGG
Tel ₂₆	*TTAGGGTTAGGGTTAGGGTTAGGGTT
2KM3	*TAAGGGCTAGGGCTAGGGCTAGGG
G4#2	*TAATGGGTTTGGGTTTGGGTTTGGGT
G4#4	*TTTTGGGTGGGTGGGTGGGTTTT
Zic1	*GGTGGGGGGCGGGGAGGCCGGG
EC-6	*GGTGGGGAGGGGTAAGGGG
EC-7	*GGGGGAGGAGGACGGGGG
EC-9	*GGGGCGGGGTGGGTTGG

Table 8. G4 oligonucleotides used to make binding assay substrates. All oligonucleotides are 5'-end-labelled with 6FAM (*). EC-6, EC-7 and EC-9 are G4 sequences from *E. coli*, used for DinG assays.

system	Sequence (5'-3')
Oligo FAM	AAGGCAACTCGTCTC*
S-cMYC	*GAGACGAGTTGCCTTATTGAGGGTGGGTAGGGTGGGTAA-(A) ₁₁
S-cKIT1	*GAGACGAGTTGCCTTATTAAGGGAGGGCGCTGGGAGGAGGG-(A) ₁₁
S-KRAS	*GAGACGAGTTGCCTTATAAGGGATAAGGGTGTGGCGGGA-(A) ₁₁
S-Mcm2	*GAGACGAGTTGCCTTATGAAGGGACACGGAGGGGCGGGCCAGAGGGTCC-(A) ₁₁
S-BU1A (+3.5)	*GAGACGAGTTGCCTTATTAAGGGCTGGGTGGGTGCTGTGAAGGGCTGGG-(A) ₁₁
S-BCL2 Mid	*GAGACGAGTTGCCTTATTAGGGCGCGGGAGGAAGGGGGCGGG-(A) ₁₁
S-Tel ₂₆	*GAGACGAGTTGCCTTATTAGGGTTAGGGTTAGGGTTAGGGTT-(A) ₁₁
S-2KM3	*GAGACGAGTTGCCTTATTAAGGGCTAGGGCTAGGGCTAGGG-(A) ₁₁
S-G4#2	*GAGACGAGTTGCCTTATTAATGGGTTTGGGTTTGGGTTTGGGT-(A) ₁₁
S-G4#4	*GAGACGAGTTGCCTTATTGGGTGGGTGGGTGGG-(A) ₁₁
Trap	GAGACGAGTTGCCTT
C-cMYC	(T) ₁₁ -ACTCCACCCATCCACCCATTATAAGGCAACTCGTCTC
C-cKIT1	(T) ₁₁ -ATTCCCTCCCGCGGACCCCTCCCTCCTCCCATAGGCAACTCGTCTC
C-KRAS	(T) ₁₁ -TCCCGCCACACCCTTATCCCTTATAAGGCAACTCGTCTC
C-Mcm2	(T) ₁₁ -CTTCCCTGTGCCTCCCCGCCCGGTCTCCAGGATAAGGCAACTCGTCTC
C-BU1A (+3.5)	(T) ₁₁ -CCCAGCCCTTGACAGCACCCAGCCCTTAATAAGGCAACTCGTCTC
C-BCL2 Mid	(T) ₁₁ -TATCCCFCCCTCCTTCCCCCGCCCATAGGCAACTCGTCTC
C-Tel ₂₆	(T) ₁₁ -AATCCCAATCCCAATCCCAATCCCAATAAGGCAACTCGTCTC
C-2KM3	(T) ₁₁ -ATTCCCGATCCCGATCCCGATCCCGATCCCATAGGCAACTCGTCTC

C-G4#2	(T) ₁₁ -ATTACCCAAACCCAAACCCAAACCCAATAAGGCAACTCGTCTC
C-G4#4	(T) ₁₁ -CCCACCCACCCACCCAATAAGGCAACTCGTCTC

Table 9. G4 oligonucleotides used to make RecQs helicase assay substrates. G4 DNA system formed by annealing Oligo FAM with 1.2-fold excess of S-G4. Oligo FAM is 3'-end-labelled with 6FAM (*); S-G4s are 5'-end-labelled with Dabcyl (*). C-G4 is the complementary strand able to open all the unreacted system. Trap G4 is the capture strand used to prevent the refolding of the G4 system.

system	Sequence (5'-3')
Oligo FAM	*CTCTGCTCAACGGAA
S-cMYC	(A) ₁₁ -AGGGTGGGTAGGGTGGGTATTCCGTTGAGCAGAG*
S-cKIT1	(A) ₁₁ -AGGGAGGGCGCGCTGGGAGGAGGGTATTCCGTTGAGCAGAG*
S-T30695	(A) ₁₁ -GGGTGGGTGGGTGGGTATTCCGTTGAGCAGAG*
S-Tel ₂₃	(A) ₁₁ -TAGGGTTAGGGTTAGGGTTAGGGTATTCCGTTGAGCAGAG*
S-KRAS	(A) ₁₁ -AGGGCGGTGTGGGAATAGGGAATATTCCGTTGAGCAGAG*
S-BCL2	(A) ₁₁ -GGGCGCGGGAGGAATGGGCGGGTATTCCGTTGAGCAGAG*
S-Zic1	(A) ₁₁ -GGTGGGGGGGCGGGGGAGGCCGGGTATTCCGTTGAGCAGAG*
S-EC6	(A) ₁₁ -GGTGGGGAGGGGTAAGGGGTATTCCGTTGAGCAGAG*
S-EC7	(A) ₁₁ -GGGGGAGGAGGACGGGGGTATTCCGTTGAGCAGAG*
S-EC9	(A) ₁₁ -GGGGCGGGGTGGGTATTCCGTTGAGCAGAG*
Trap	TTCCGTTGAGCAGAG
C-cMYC	CTCTGCTCAACGGAATACCCACCCTACCCACCC- (T) ₁₁
C-cKIT1	CTCTGCTCAACGGAATACCCTCCTCCCAGCGCCCTCCC- (T) ₁₁
C-T30695	TCTTGCTCAACGGAATACCCACCCACCCACCC- (T) ₁₁
C-Tel ₂₃	CTCTGCTCAACGGAATACCCTAACCCCTAACCCCTAACCCCTA- (T) ₁₁
C-KRAS	CTCTGCTCAACGGAATACCCCTCTTCCCTCTTCCCACACCGCCC-(T) ₁₁
C-BCL2	CTCTGCTCAACGGAATACCCGCCCATTCCTCCCGCGCCC-(T) ₁₁
C-Zic1	CTCTGCTCAACGGAATACCCGGCCTCCCCCGCCCCCCCACC(T) ₁₁
C-EC6	CTCTGCTCAACGGAATACCCCTTACCCCTCCCCACC-(T) ₁₁
C-EC7	CTCTGCTCAACGGAATACCCCGTCCTCCTCCCCC(T) ₁₁
C-EC9	CTCTGCTCAACGGAATACCAACCCACCCGCCCC(T) ₁₁

Table 10. G4 oligonucleotides used to make *E. coli* DinG helicase assay substrates. G4 DNA system formed by annealing Oligo FAM with 1.2-fold excess of S-G4. Oligo FAM is 5'-end-labelled with 6FAM (*); S-G4s are 3'-end-labelled with Dabcyl (*). C-G4 is the complementary strand able to open all the unreacted system. Trap G4 is the capture strand used to prevent the refolding of the G4 system.

11.4. Binding assay

11.4.1. Electrophoretic Mobility Shift Assays (EMSA)

The Electrophoretic Mobility Shift Assay (EMSA) is used to identify interactions between proteins and nucleic acids. In this assay, the electrophoretic mobility of a protein-nucleic acid complex is compared to that of free nucleic acid. The binding of a nucleic acid to one or more proteins changes

its mobility through a non-denaturing gel matrix, usually reducing it (Figure 16). In addition to the specificity of the interaction between the protein and nucleic acid, this type of assay can be used to investigate the affinity and the kinetics of the interaction (Hellman & Fried, 2007).

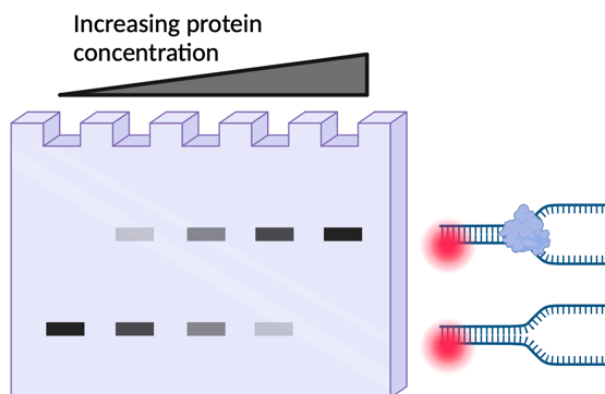


Figure 16. 6% non-denaturing polyacrylamide gel EMSA representation. The electrophoretic mobility of a protein-nucleic acid complex is lower than that of free nucleic acid. Created with BioRender.com.

The experiments were performed by incubating increasing concentration of purified protein fragments with 6FAM labelled substrates (Table 7 and 8) at a final concentration of 10 nM in a 20 μ l reaction mixture. The binding assays for RecQ helicases fragments (RecQ1, BLM, WRN, RecQ4 FL, RecQ4 1-427 and RecQ5) have been performed in 20 mM Tris-HCl pH 7.5, 10 mM KCl, 5% glycerol, 0.1 m/ml BSA and 0.5 mM TCEP. After incubation for 30 min at RT, the reaction products were separated on a 6% non-denaturing polyacrylamide gel run in 1X TBE buffer (0,089 M Tris Base, 0,089 M Boric Acid, 0,002 M EDTA). The assays have been repeated three times. Labelled nucleic acid fragments were detected with fluorescence (ImageQuant, LAS4000, GE Healthcare). The quantification of protein-nucleic acid complexes and the percentage of binding was calculated and plotted using GraphPad Prism.

11.4.2. Fluorescence Anisotropy

Fluorescence Anisotropy can be used as a sensitive non-destructive technique for the characterisation of protein-DNA interactions, even in a high throughput format. It is based on the change in fluorescence polarization upon protein binding to a labelled nucleic acid, in our case with the 6FAM fluorophore. It can be used to obtain a DNA-binding curve if the molecular size of the protein-DNA complex is sufficiently different from the free fluorescing component, thus providing information about binding affinity (Figure 17). The technique does not provide positional information.

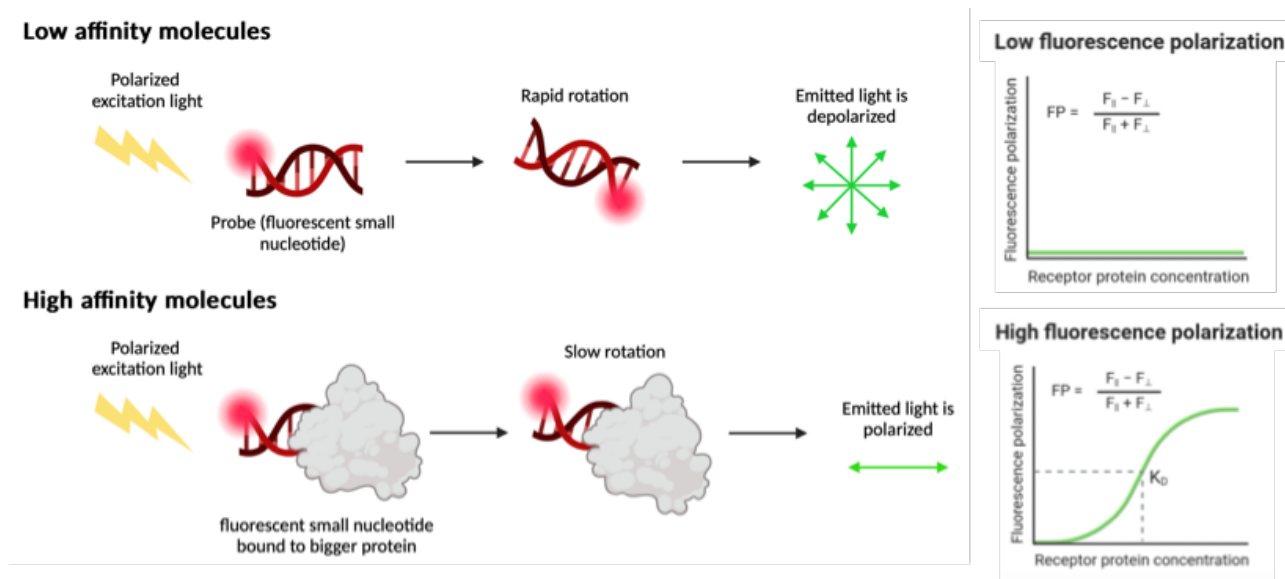


Figure 17. Schematic representation of Fluorescence Anisotropy. Rapidly rotating small molecule fluorophore gives low FP signal (low mP). The association of a relatively large molecule, such as a protein, with the small molecule fluorophore slows down the motion of the fluorophore, leading to increased FP signal. Created with BioRender.com.

The oligonucleotides for the assay were prepared as described in the paragraph 5.3. The assay is performed incubating increasing protein concentrations with 5 nM labelled nucleic acid for 30 mins at 25°C. The reaction buffer used for RecQ4 fragments 1-55 and 1-133 was 20 mM Tris-HCl pH 7.5, 10 mM KCl, 5% glycerol, 0.1 m/ml BSA and 0.5 mM TCEP. The reaction buffer used to perform the binding assays with G4 substrates was 20 mM Tris-HCl pH 7.5, 1 mM MgCl₂, 100 mM KCl and 0.5 mM TCEP. The mixtures are then loaded in 384 well plates (Corning Assay plate, no lid, low volume, flat bottom, non-binding surface, black polystyrene) and fluorescence polarization was measured with a Tecan Infinite F200 Pro plate reader equipped with filters for excitation and emission wavelengths at 485 and 535 nm, respectively. The instrument automatically subtracts the background fluorescence value from each sample measurement and subtracted values are output in milliunits of polarization (mP). Data were analysed using GraphPad Prism. The averaged polarization of the samples for each protein dilution was plotted against the corresponding protein concentration.

11.5. Helicase assay

11.5.1. Gel-based helicase assay

The standard gel-based helicase assay monitors nucleic acid unwinding using a modified DNA duplex where one strand is labelled with fluorophore (in our case with the 6FAM). The DNA is incubated

with a helicase in the presence of ATP and specific reaction components (Mg^{2+} , buffer, etc.). If the protein is active as a helicase, the duplex DNA is resolved into two single strands. To prevent duplex re-annealing during the reaction, a strand complementary to the unlabelled strand (capture strand) is added to the reaction. When the reaction is over, the protein is degraded by adding Protease K. The positive control of the assay is the DNA duplex resolved by heat-shocking (incubated for 5 mins at 95°C and then immediately switched to 4°C), while the negative control is a reaction mix without the protein. The reaction mixtures and the controls are loaded onto an 8% SDS-polyacrylamide gel and run at RT in TBE + 0.1% SDS. The bands corresponding to DNA duplex and the separated single strands can be visualised on a gel by a fluorescence reader (Figure 18). The stronger is the helicase activity, the higher is the amount of duplex resolved in two single strands.

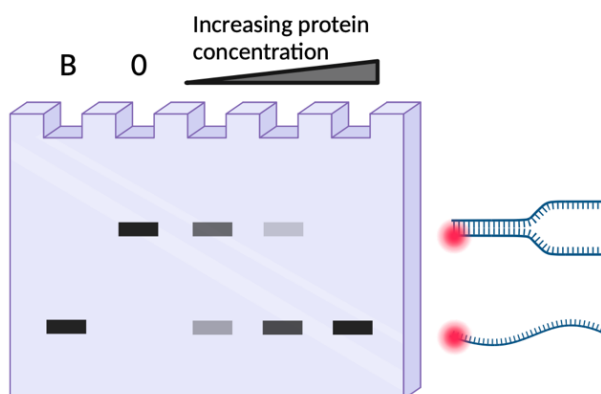


Figure 18. SDS-polyacrylamide gel representative of a resolved helicase assays. The electrophoretic mobility of a DNA duplex is lower than that of ssDNA. Created with BioRender.com.

The helicase assay was performed incubating increasing concentrations of protein with 40 nM of DNA fork, 1 μM capture strand (Cap1, Table 7) and 5 mM ATP in 20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM MgCl_2 , 5% glycerol, 0.1 mg/ml BSA and 0.5 mM TCEP. The reaction was incubated at 37°C for 30 minutes and terminated using a stop buffer (8 mM EDTA, 0.5% SDS, 0.1 mg/ml proteinase K and 4% glycerol) to denature the protein. As control for 100% unwinding the substrate was boiled to fully separate the two strands. The unwinding reaction was monitored by loading the samples onto a 8% native PAGE gel. Experiments were done in triplicate.

11.5.2. FRET-based helicase assay

An alternative way to perform a helicase activity assay is using the fluorescence resonance energy transfer (FRET) technique, in which the fluorophore emits when the double stranded substrate is unwound by the helicase. The fluorophore (6FAM) is covalently attached to the 5'-end in the

fluorescent strand and quenched by the quencher (BHQ-1) that is covalently attached to the 3'-end in the quencher strand via FRET. When the double-stranded substrate (dsDNA in Figure 19) is unwound by the helicase, the fluorophore (red dot) emits bright light upon its release from the quencher (green dot). The capture strand, which is complementary to the quencher strand, prevents the reannealing of the unwound duplex (Figure 19). This technique allows to collect data continuously throughout the reaction in real time.

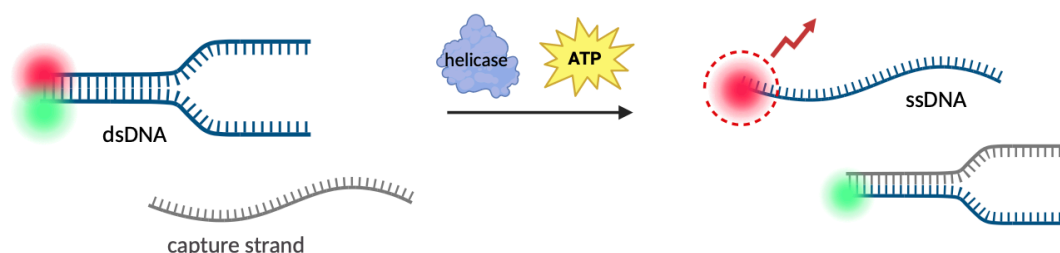


Figure 19. Schematic representation of fluorescence helicase assay based on fluorescence resonance energy transfer (FRET). When the fluorophore and the quencher are close, the fluorescence emission is blocked; when the duplex is resolved, the 6FAM and BHQ-1 are no longer in close proximity and the fluorescence emission from the 6-FAM is detected. Created with BioRender.com.

The assay was performed in 5 mM ATP in 20 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM MgCl₂, 5% glycerol, 0.1 mg/ml BSA and 0.5 mM TCEP with 15 nM substrate and 430 nM Capture strand (Cap1, Table 7) in 30 µl of reaction volume. The unwinding reaction was started by incubating a single concentration of the purified protein in the reaction mixture containing 5 mM ATP. The reaction mixture was loaded in 384 well plates (Corning Assay plate, no lid, low volume, flat bottom, non-binding surface, black polystyrene) and the fluorescence intensity was recorded at 37°C continuously in real time for 30min or 1h using a photosensor (Tecan® Infinite F200Pro instrument). To have a measurement of 100% unwinding the reaction was boiled and measured. The assay was done in triplicate. The percentage of unwinding was calculated and plotted using GraphPad Prism.

11.5.3. Two-step fluorescent-based helicase assay of G4

To study the unwinding activity of the proteins towards G4 substrates, an alternative helicase assay technique has been used. The two-step fluorescent-based helicase assay is a high throughput assay requiring both low sample concentrations and reaction volumes, which allows testing the kinetic behaviour of the enzyme in real time in a 96-well plate format (Mendoza et al., 2015; De Piante et al., 2023). It uses a simple DNA system consisting of a G4-forming sequence, in which the

oligonucleotides were modified with the 6FAM fluorophore and the Dabcyl quencher, to allow monitoring the unfolding of the G4 substrate by fluorescence spectroscopy. The system is composed of three oligonucleotides (Table 9 and 10): the 3'-FAM-labelled single strand (Oligo FAM), the G4 motif forming strand labelled at the 5'-end with dabcyl (S-G4) and the trap oligonucleotide (Trap), annealed as previously described in paragraph 5.3. Since the helicases tested in the present work are not able to unwind blunt-end dsDNA, the increasing fluorescence, detected consequently to the unwinding of the dsDNA region at the 5'-end of the system, is due to the unfolding of the G4 sequence prior to the double strand region.

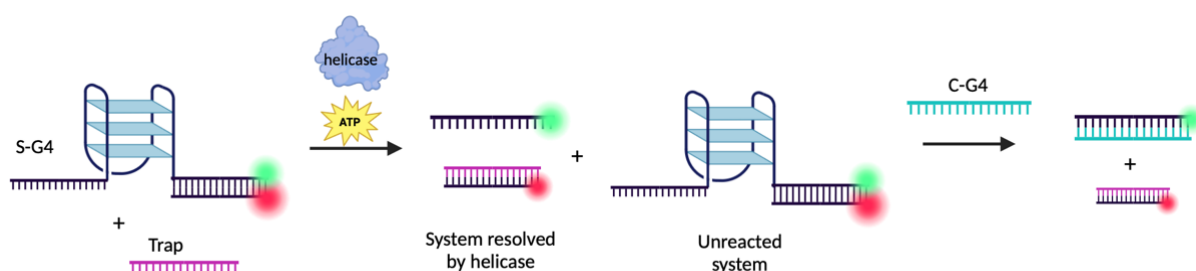


Figure 20. A schematic diagram of the two-step fluorescent-based G4 helicase assay. In the first step, upon addition of ATP, the helicase unwinds the G4/duplex system releasing the labelled single-stranded probe. In the second step, the dabcyl-complementary strand C-G4 was added, displacing any remaining duplex leading to the maximum emission signal possible. Created with BioRender.com.

Helicase reactions were carried out in triplicate in 384-well plates (Corning Assay plate, no lid, low volume, flat bottom, non-binding surface, black polystyrene) at 25°C and fluorescence monitored in a microplate reader (Tecan Infinite F200 PRO). Every replicate contained a 30 µl solution of 40 nM FAM-dabcyl system (S-G4 + Oligo FAM) previously annealed, 6 nM or 20 nM of RecQ helicases or *E. coli* DinG, respectively, and 200 nM of Trap oligonucleotide (unlabelled, complementary to the FAM-labelled strand). Next, 5 mM of ATP solution was added to every well and the fluorescence emission was recorded every 30 s (the excitation wavelength was set at 485 nm and the emission wavelength at 535 nm). Once the maximum emission was reached and the signal was stable (30 min or 1h, depending on the helicase tested), to measure the maximum level of fluorescence emission of the unwound substrate, 2 µM of a strand complementary to the dabcyl-labelled sequence (C-G4) was added to every well in a second step and emission was again monitored every 30 s for 30 min/1h. This C-G4 DNA strand displaces the Oligo FAM sequence in any unreacted G4-duplex system, leading to a highly stable duplex DNA in an irreversible process (Figure 20). Data were analysed taking in consideration the emission reached after 30mins/1h related to the one obtained at the end of

the second step of the reaction, and the percentage of unwinding was plotted using GraphPad Prism. Statistical analysis was performed on GraphPad Prism using an Ordinary One-Way ANOVA test or a Two-Way ANOVA test, as indicated in figure captions. P values less than 0.05 were considered statistically significant. Only most appropriate statistical significance is shown.

11.5.4. Two-step fluorescent-based helicase assay of G4 in presence of G4 ligands

The two-step fluorescent-based helicase assay can be also used to screen a series of G4 structures together with G4 ligands, to see the behaviour of the helicase activity and to assess the possible inhibitory effect of pharmacological G4 binders on enzyme activity.

I tested *E. coli* DinG helicase G4 unwinding activity in presence of G4 ligands, which are known to stabilize the G4 structure. The same procedure as used for G4 substrate screening was applied for the G4-ligand screening. A 30- μ l solution of 40 nM G4 system, 20 nM *E. coli* DinG, 200 nM Trap single-strand and 1 μ M of selected ligand was prepared in 20 mM Tris-HCl buffer (pH 7.2, 10 mM MgCl₂, 1 mM KCl and 99 mM NaCl). ATP and the single-stranded complementary sequence were added following the same procedure as described above. The emission was monitored every 30 s. Experiments were done in triplicate and data were plotted using GraphPad Prism.

11.6. Circular Dichroism (CD) of G-quadruplex

For assessing the folding of tested DNA oligonucleotides into G4 structures and determine their conformation and melting temperature, CD spectroscopy was performed. Desalted oligonucleotides were chemically synthesised and purchased from Eurofins (Munich, Germany) (Table 8, without fluorophore). They were first resuspended in water at 500 μ M concentration and kept at 4°C for 24h, and then diluted to a final concentration of 10 μ M in different buffers, with different concentrations of KCl, NaCl and LiCl, as indicated in the text. Solutions were denatured at 95°C for 5 min and then slowly cooled to room temperature and stored at 4°C overnight before use.

CD spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a PFD-425S/15 Peltier temperature controller using a quartz cell of 1 mm optical path length (Hellman Analytics) after 5 mins at 25°C to allow the sample to equilibrate. The temperature was set at 25°C and the spectra were recorded over a wavelength range of 230-320 nm and averaged over three scans. The following parameters were used: standard sensitivity, 200 nm min⁻¹ scan rate, 1 s response time and 1 nm bandwidth. Acquired spectra were baseline corrected for signal contribution due to the buffer. For T_m determination, spectra were recorded over a temperature range of 20-100°C, with temperature increase of 2°C/min. The same parameters as described above were used.

CD melting data were converted to fraction folded (θ) plot using the following equation:

$$\theta = \frac{CD_{\gamma} - CD_{\gamma}^{min}}{CD_{\gamma}^{max} - CD_{\gamma}^{min}}$$

where CD_{γ} is the ellipticity of the monitored wavelength at each temperature, CD_{min} is the lowest ellipticity and CD_{max} the highest ellipticity. Data were plotted θ vs T in GraphPad prism and then fitted into a sigmoidal Boltzmann function. x_0 value corresponds to T_m .

12. Structural studies: protein crystallization

In order to obtain crystals of a molecule, it is necessary to bring the molecule solution to a supersaturation state in which the solute, in our case the protein, can start packing in ordered 3D arrays, thus promoting nucleation and, ultimately, crystal growth. To this end, different combination of buffers, pH values, temperature, additives (often organic solvents as 2-propanol, ethanol or dimethylsulfoxide) and precipitants (salts as ammonium sulfate, manganese (II) chloride or polymers as polyethylenglicole or jeffamine) can be used. Because of the large number of parameters involved, it is impossible to pre-determine the exact conditions in which a protein will eventually crystallize. A number of crystallization screening kits are available on the market for the initial phase of the crystallization trials with a new protein, and many had been used in the present work, including Morpheus Molecular Dimensions, MIDAS plusTM Molecular Dimension, BCS Molecular Dimension, JCSG plusTM Molecular Dimension. After the first phase, the range of variations around those combinations of chemicals can be reduced since we can focus on the optimization of those conditions that yielded the most promising preliminary results in terms of nucleation and small crystal formation. Through several rounds of optimization, crystal of sufficient quality and dimensions may eventually be obtained.

In order to check a broad range of crystallization conditions, 96-well plates were used. The vapor diffusion method, in particular the sitting drop technique, was used. It is based on a drop lying on a flat surface, usually constituted by a mix of the protein solution and a crystallization solution, left to equilibrate with a given amount of the crystallization solution in a closed environment. During equilibration the water will diffuse out of the drop, thus moving the system from the stable through the supersaturated state, possibly leading to nucleation and crystal formation. A drop of 0.1 μ L of the concentrated protein solution (BLM 5.4-20 mg/ml) was mixed in 1:1 ratio of reservoir to protein. A reservoir solution of 70 or 30 μ L, depending on the multi-well plate, containing a different composition for each well of one or more of the components above mentioned, was used. Each

reservoir solution contains a precipitant agent, which is responsible for the “salt out” process of the protein from the solution, by subtracting the available water molecules, thus facilitating reaching the supersaturation state and the crystallization. The process aimed to reach the so-called meta-stable state, where the protein is too concentrated to be fully soluble but too less concentrated to become unstable and aggregate. The MOSQUITO © TTP Labtech automatically mixed the drop, and the plates were stored at two different temperature conditions (20°C and 4°C) since the temperature at which equilibration takes place is also important, influencing the diffusion rate.

Results and discussion

Although the main focus of the project was to investigate the interaction between human RecQ helicases and G-quadruplex structures, I initially worked on a side project, on the bacterial helicase DinG, in collaboration with C. Giancola and J. Amato from the University of Naples.

The rationale of starting with DinG as a model system for G4 interaction was to facilitate the establishment of experimental procedures and provide a preliminary protocol to refine our approach. The choice of DinG was based on the following two reasons: 1. bacterial helicases, like DinG, are simpler to express and purify in large amounts compared to human helicases; 2. the assay selected for assessing G4 unwinding (Mendoza et al., 2015) was originally designed for Pif1, a helicase with a 5'-3' directionality. Testing the assay first with DinG, another 5'-3' helicase, allowed us to use the original substrates developed by the authors, helping minimize unknown variables while we were getting acquainted with the methodology. The results obtained from the analysis have been published (De Piante et al., 2023) and are briefly presented in the last section (section 18).

13. Expression and purification of human RecQ helicases

13.1. An overview

As described above, RecQ are among the helicases mostly implicated in G-quadruplex metabolism. The main aim of this project was to perform a comprehensive and systematic analysis of the interaction between various G4 structures and the human RecQ helicases, evaluating their ability to bind and unwind these substrates. To achieve this, we used a variety of unimolecular G4s derived from promoter and telomeric regions, representing different topologies (parallel, antiparallel and hybrid), which had been well-characterized in the literature.

A significant challenge in carrying out this extensive analysis was the need for large quantities of highly pure protein. Expressing and purifying the full-length protein is particularly challenging, due to large size of these helicases and to the presence of multiple unstructured and disordered regions. Moreover, some RecQ helicases contain extra domains with distinct functions, such as the exonuclease domain in WRN and the RNA polymerase binding domain in RecQ5, which could interfere with the detection of the helicase activity. To address these difficulties, we focused on truncated versions of RecQ proteins that include the core helicase region, containing the catalytic domains essential for the unwinding activity. The laboratory routinely expressed a truncated form of the human RecQ1 (49-616), the full-length human RecQ4 and a large number of RecQ4 fragments containing the catalytic RecQ-like domain and/or the N-terminal unique domain. Thanks to the

kindness of Opher Gileadi and Joseph Newmann (Oxford Structural Genomic Consortium), that provided plasmids for the expression of the other proteins, we produced recombinant, truncated forms of BLM (residues 636-1298), which encodes the catalytic and HRDC domains; WRN (residues 517-1093), which includes the helicase core but lacks the exonuclease and HRDC domains; and RecQ5 (residues 11-582), which encodes the helicase core, Zn domain, and large part of the KIX domain. Working with truncated versions of RecQ helicases was particularly valuable for studying the role of the helicase core in G4 metabolism. This approach allowed for a more focused comparison of the core functionalities across different helicases, providing clearer insights into their mechanisms without the potential interference of additional domains. On the other hand, the lack of some functional domains could hide the role played by the helicase in the interaction with G4s. A schematic representation of RecQ helicases used in the present work is shown in Figure 21.

One of the aim of the project was also to obtain structural information on helicases:G4 interactions by attempting co-crystallization of RecQ helicases with their high-affinity G4 substrates, an experiment that also required large amounts of purified protein. Despite multiple attempts, I was unable to obtain diffracting crystals of RecQ helicases in complex with G4 structures.

Furthermore, I conducted a bioinformatic analysis on human RecQ helicases and the bacterial CsRecQ, focusing also on sequence and structural alignment across RecQ helicases orthologues, to identify the presence of a potential guanine-specific binding pocket which has been characterized in bacterial RecQ (Voter et al., 2018). Mutants of BLM and WRN were subsequently cloned, expressed, and purified to assess the role of key residues possibly implicated in forming the binding pocket.

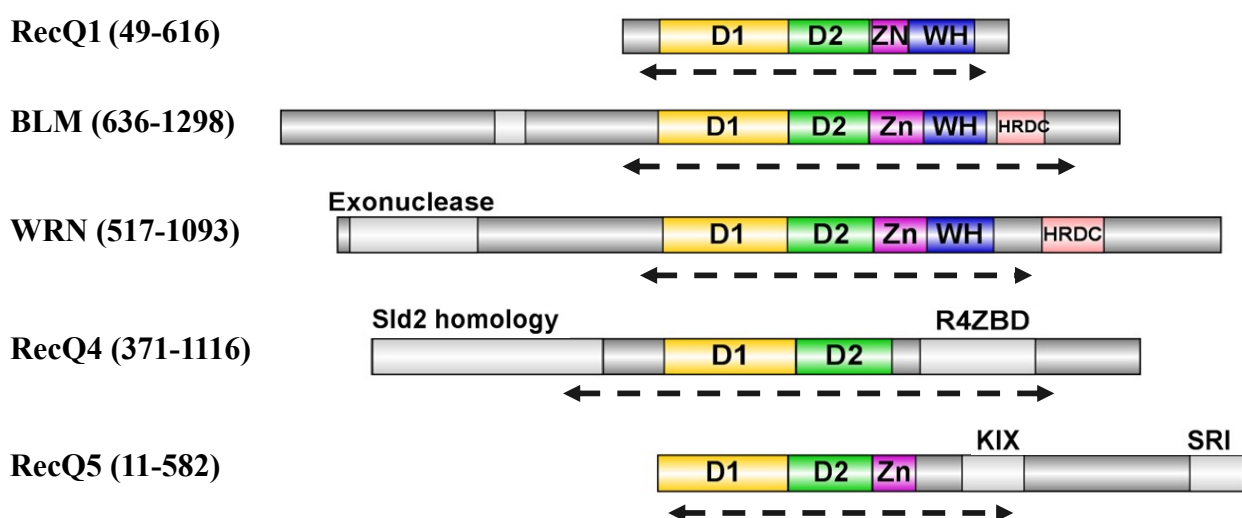


Figure 21. Schematic representation of RecQ helicases. The two RecA domains forming the helicase core are shown in yellow and green (HD1 and HD2 domains respectively), the WH domain in blue, the Zn-binding

domain in violet, and the HRDC domain in pink. Additional domains which are not part of the RecQ framework are colored in different shades of grey. Black arrows underline proteins boundaries.

The large-scale expression and purification of RecQ helicases thus constituted a significant part of my PhD project. Although some proteins were routinely produced in our lab and the plasmids of others were generously provided by a collaborating research group, optimization of the purification protocols was necessary for several reasons.

First, the pure amount of protein required for the experiments was substantial, which meant I needed to refine the protocols to obtain larger yields from the same or even smaller volumes of starting cell culture, thereby reducing the need for repeated, time-consuming culture growth. Moreover, replicating purification protocols from other laboratories often proves challenging, particularly when not all procedural details are fully documented. Even small variations in buffer composition, reagent concentrations, timing, or skipping minor steps (such as specific washes or incubation times) can have a significant impact on the overall success of the process. Thus, fine-tuning these conditions was essential to achieve consistent and efficient protein production.

All RecQ helicases were expressed in bacterial cell lines, using LB or TB media. The proteins possess a fusion tag at the N-terminal (in the case of RecQ1 and RecQ4 fragments) or C-terminal (for the other RecQs), which has been removed in almost all cases as it may interfere with the activity of the purified protein and/or crystallization.

13.2. Purification of human RecQ1 (49-619)

According to the literature, the full-length RecQ1 protein does not express well in bacterial cells, gives small yields and is purified as a mixture of different oligomeric forms, each of which possesses a different activity (Muzzolini et al., 2007). A truncated version of the protein, with modest deletions of both the N-terminal and C-terminal regions of the protein, encompassing amino acids 49-619 (Pike et al., 2009) has long been produced in the laboratory, as part of a collaboration with Alessandro Vindigni (Washington University). We further optimized the large-scale expression and purification protocols. The construct was expressed in *E. coli* cells in TB, following induction at 18°C with 0.2 mM IPTG. Only a few steps were necessary to obtain a large amount of pure RecQ1 protein, described in detail in section 10.1 in Materials and Methods. The Ni-affinity step was very successful, being able to remove the majority of contaminants (Figure 22A). Thus, the protein was directly injected in a gel filtration S200 column for Size Exclusion chromatography to eliminate the remaining impurities, eluting in a single monomeric peak, consistent with its molecular weight of ~70 KDa

(Figure 22B). The protein final yield was about 0.8 mg of pure RecQ1 per liter of cell culture, adequate for biochemical experiments.

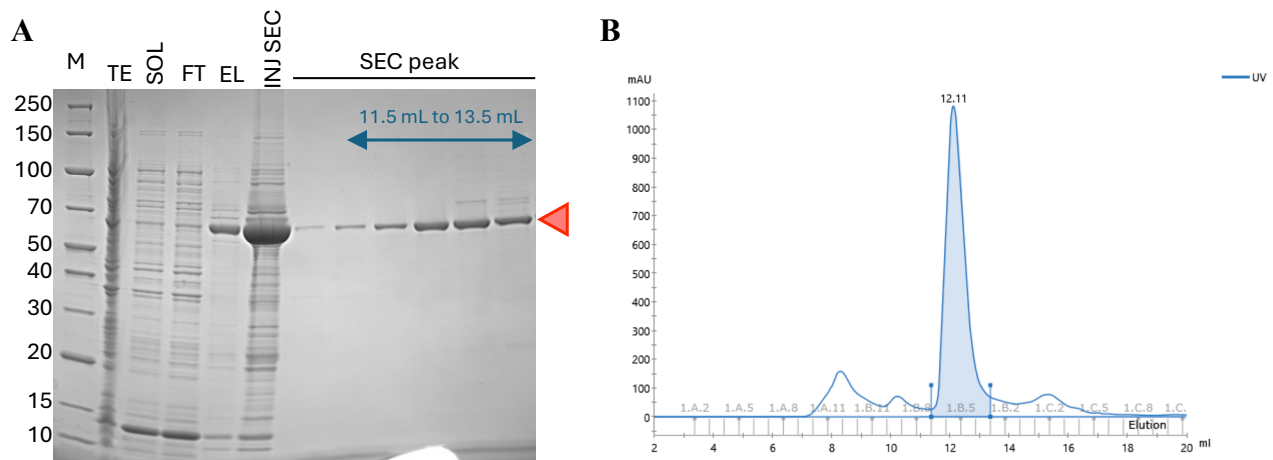


Figure 22. Purification steps for RecQ1. **A.** SDS-page analysis of RecQ1 purification (Ni-affinity and Size Exclusion chromatography); the band corresponding to the target protein is shown with a red triangle. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin. INJ SEC: Sample injected into SEC; SEC peak: Eluted fractions from the main SEC peak. **B.** Elution profile from a Superdex 200 column; the fraction visualised in the gel are shaded in blue.

13.3. Purification of human BLM (636-1298)

BLM protein encompassing amino acids from 636 to 1298, spanning the helicase, RQC and HRDC domains, was expressed as a fusion with C-terminal 6His-FLAG tag and purified as previously described in literature (Newman et al., 2015) and in section 10.2 in Material and Methods. The protocol consists of a Ni-affinity chromatography column (Figure 23A), followed by Heparin chromatography. Although the heparin step helped to remove the excess DNA and some non-DNA binding proteins, it was later abandoned as it did not provide further purification. After the Ni-affinity step, the protein was then directly placed in a dialysis membrane together with the TEV protease and dialyzed overnight to reduce the concentration of imidazole, which interferes with the TEV activity. The difference between the tagged and untagged protein is hardly visible due to the small size of the tag (Figure 23A). An additional Ni-affinity incubation in batch is needed to remove the uncleaved protein, the tag, as well as the His-tagged TEV protease. A key factor influencing the right unbinding of the untagged protein from the Ni-NTA resin, which is not mentioned in the protocol described in literature, is to add little amount of imidazole (30 mM) before the Ni-NTA resin incubation. This step

was added in the optimized protocol since the omission of this addition resulted in a protein stuck to the resin and eluted together with the tag and the protease. As a final polishing step, a Superdex™ 200 26/60 GL was applied. Some proteolytic fragments were still present, but they could be visualized when overloading the gel, suggesting that a very large proportion of the protein was intact. Overall, the protein was judged pure enough for biochemical assays and structural studies (Figure 23B). BLM eluted as a monomer with a molecular weight of about 76 KDa (Figure 23C).

Among all RecQ proteins, BLM is the only one from which we could obtain milligrams yields (about 11 mg) from 2 liters of cell culture, being able to concentrate it until 25 mg/ml. This makes BLM the ideal protein for performing all tests and serving as a control in assays when necessary.

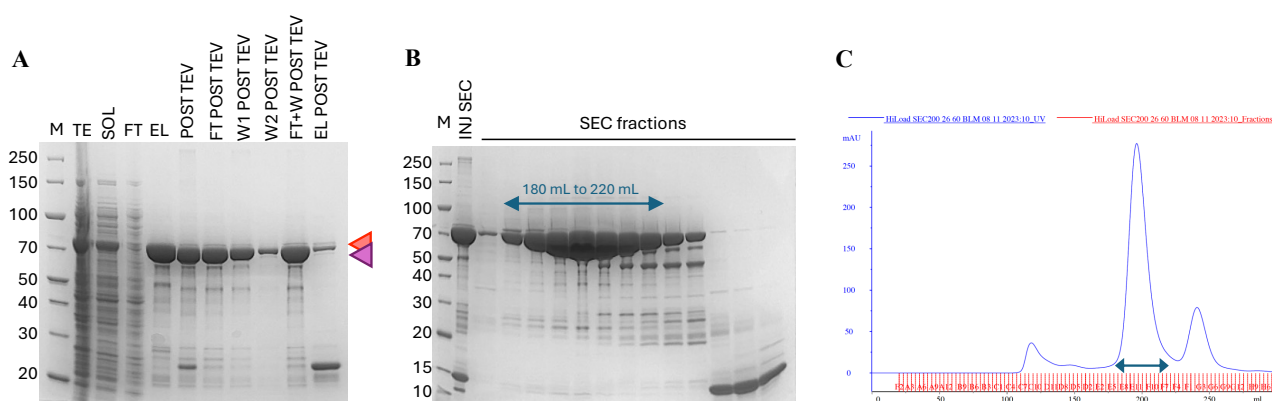


Figure 23. Purification steps of BLM. **A.** SDS-page analysis of the sample from Ni-affinity purification and protein before and after TEV protease cleavage. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-affinity; EL: Eluted fraction from Ni-affinity; POST TEV: protein after tag-cleavage; FT POST TEV: Flow Through from Ni-NTA resin post tag-cleavage; W1-W2 POST TEV: wash fraction from Ni-NTA resin post tag-cleavage; EL POST TEV: eluted fraction from Ni-NTA resin post tag-cleavage. **B.** SDS-page analysis of Size Exclusion chromatography purification. Pooled fractions are highlighted in blue. Labels as follow: M: Molecular Weight standards (in KDa); INJ SEC: Injected sample; SEC fractions: eluted fractions from the main peak. **C.** Elution profile from a Superdex™ 200 26/60 GL column; the fraction collected are highlighted by blue arrow.

13.4. Purification of human WRN (517-1093)

In our study we used a WRN fragment encompassing amino acids 517-1093, whose structure was one of the most recently resolved by Newman et al, in 2020. The protein fused to a C-terminal 6His-FLAG tag was expressed in *E. coli* cells in TB and purified as previously described in literature (Newman et al., 2020) and in section 10.4 of Material and Methods.

Since we faced some difficulties in obtaining the protein, due to the presence of significant proteolysis resulting in protein loss, medium-scale expression and purification tests were carried out in order to further optimize the purification protocol to obtain larger amount of pure protein. Following Ni-affinity purification (Figure 24A), the sample was subjected to Heparin chromatography. This step was not very successful as the protein eluted across a wide salt range and other prominent bands, possibly proteolytic fragments, appeared during the run (Figure 24B and C). A small amount of pure protein was subjected to tag cleavage by TEV protease in an overnight reaction, followed by Ni-affinity. Despite the successful cleavage, protein stuck to the resin and eluted together with the tag and the protease (Figure 24D). We decided to concentrate the eluted fraction and inject it onto a Superdex 200 column for Size Exclusion chromatography to remove the proteolysis products (Figure 24D). The chromatogram revealed two main peaks corresponding to the monomer form of WRN protein (highlighted by a blue area in Figure 24E) and the other one to the proteolyzed fragments. The protein final yield was about 8 μ g per liter of cell culture. Whereas we have managed to obtain enough WRN for biochemical studies, we struggled to purify enough protein for structural analysis, as in our hand the protein aggregated and was prone to proteolysis. Personal correspondence with J. Newman confirmed that indeed it was a struggle to obtain enough protein for crystallization.

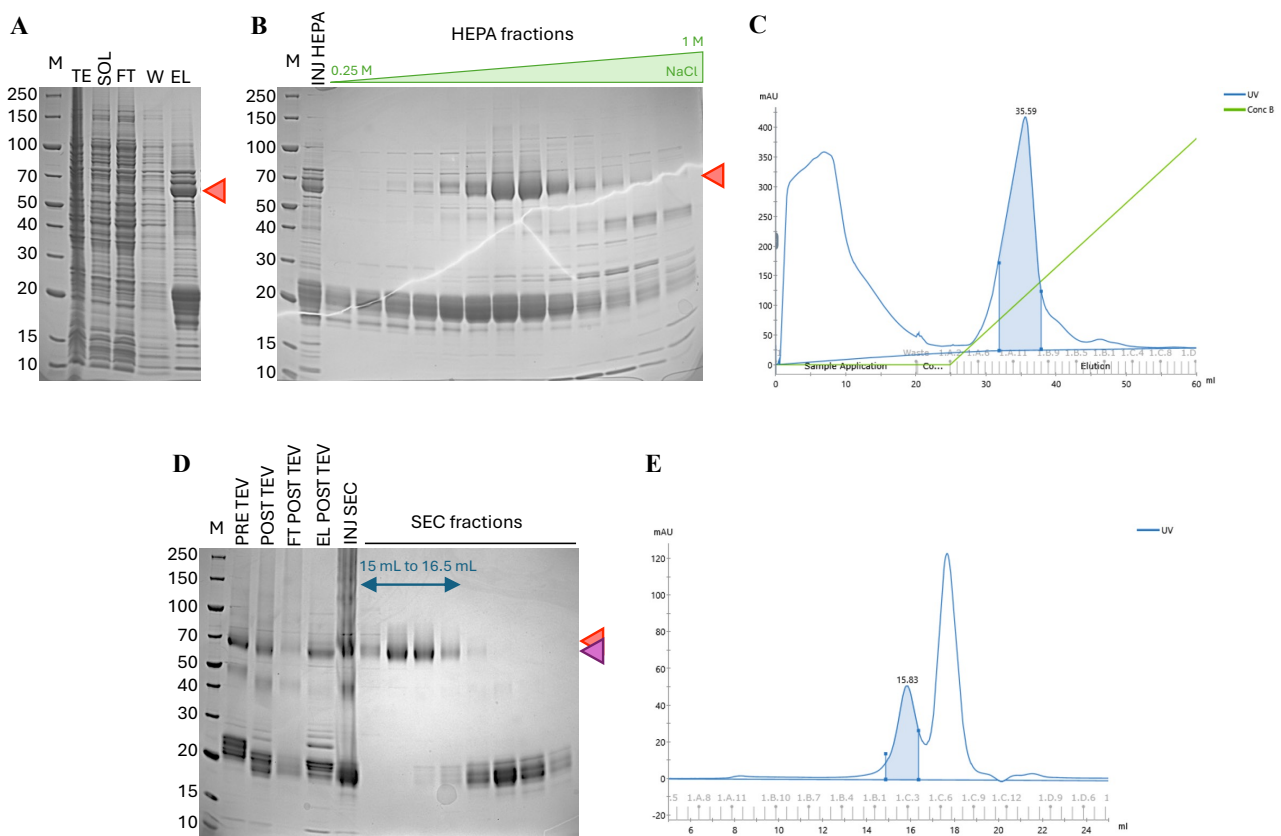


Figure 24. Purification steps of WRN. A. SDS-page analysis of the sample from Ni-affinity purification; the

band corresponding to the target protein is shown with a triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin. **B.** Heparin-affinity purification. Labels as follow: M: Molecular Weight standards (in KDa); INJ HEPA: Injected sample; HEPA fractions: eluted fractions from the main peak. **C.** Chromatogram of the heparin chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the proteins bound to heparin molecules. Pooled fractions are shaded in blue **D.** SDS-PAGE analysis of the protein samples before and after TEV protease cleavage, and collected fractions from Size Exclusion Chromatography purification. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); PRE TEV: before TEV cleavage; POST TEV: protein after tag-cleavage; FT POST TEV: Flow Through from Ni-NTA resin; EL POST TEV: eluted fraction from Ni-NTA resin; INJ SEC: Injected sample; SEC fractions: Eluted fractions from the main peak. **E.** Size Exclusion chromatography indicates that recombinant WRN is a monomer of approximately 66 KDa; the fraction are shaded in blue.

13.5. Purification of human RecQ4 (371-1116)

Many RecQ4 fragments encompassing the core helicase domain have been cloned and expressed, including a longer construct that comprises the C-terminal domain, which has been shown to play a role in the protein's activity (Mojumdar et al., 2017), as well as various regions of the N-terminal domain, comprising the last part of homology with the yeast Sld2 factor and the Zn-knuckle (aa 391-427). Among these constructs, we focused on the 371-1116 fragment, as it proved to be the most stable one, able to be purified in larger quantities, allowing us to perform an extensive biochemical analysis.

RecQ4 fragment 371-1116 was cloned as a 6His-thioredoxin tag fusion protein, and purified in collaboration with Gianluca Centrone, a PhD student of our laboratory. The detailed purification steps are described in section 10.5.1 of Materials and Methods, and comprised a Ni-affinity chromatography, followed by an anion exchange chromatography; the tag was then removed, and the cleaved protein was directly injected in a Superdex 200 column. Size Exclusion chromatography was able to separate the cleaved protein from the TEV protease. A fraction of the protein eluted in the void volume, suggesting the presence of aggregates, but a considerable amount eluted as a monomer with a molecular weight of about 82 KDa (Figure 25A and B). Although the yields were not very high (about 20 μ g per liter of cell culture), this protein was indeed challenging, and we managed to obtain enough protein to carry out biochemical analysis.

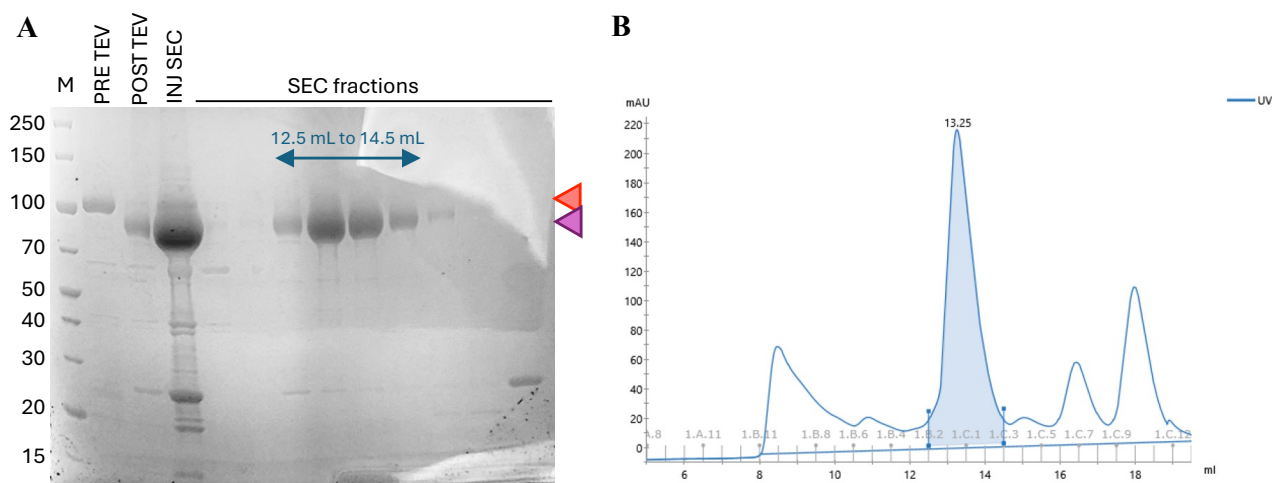


Figure 25. Final purification steps of RecQ4 371-1116. **A.** SDS-PAGE analysis of the protein samples before and after TEV protease cleavage, and collected fractions from Size Exclusion chromatography purification. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Pooled fractions are highlighted by a blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); PRE TEV: before TEV cleavage; POST TEV: protein after tag-cleavage; INJ SEC: Injected sample; SEC fractions: Eluted fractions from the main peak. **B.** Elution profile from a Superdex 200 column; the fraction are shaded in blue.

13.6. Purification of human RecQ5 (11-582)

RecQ5 fragment comprising amino acids 11-582 includes the conserved Zn^{2+} binding subdomain, which was found to be essential for efficient helicase activity, and part of the C-terminal region, in which some regions responsible for the binding to various cellular partners were deleted. Expression and purification protocols of the catalytic domain of RecQ5 tagged to a C-terminal 6His-FLAG tag were previously described in literature (Newman et al., 2016) and in section 10.6 in Material and Methods, and purification followed the same steps also used for WRN purification. The construct was expressed in TB and induced with 0.2 mM IPTG at 18°C overnight. It was purified by Ni-affinity chromatography, followed by Heparin chromatography (Figure 26A and B). The overnight TEV cleavage performed in a dialysis membrane had 100% efficiency. Follow another Ni-affinity chromatography to separate the cleaved protein (found in the flow through) from the tag and the TEV protease. Protein was applied to a SEC column (Superdex 200), which resulted in a significant portion eluting as a single peak, corresponding to a monomer (Figure 26C and D).

The final yields of 0.3 mg of pure RecQ5 per liter of cell culture, obtained after the purification process, were appropriate for the subsequent biochemical analysis.

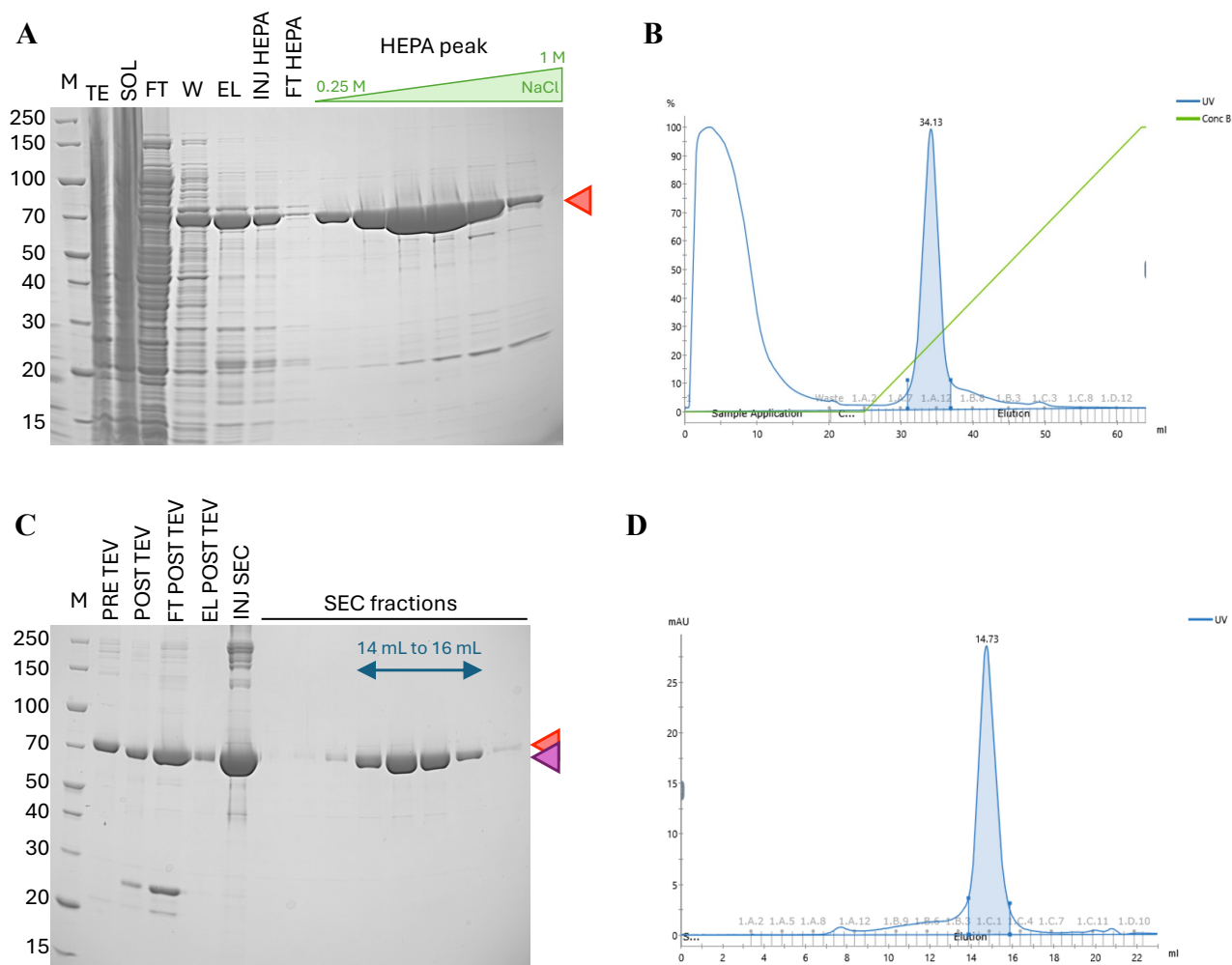


Figure 26. Purification steps of RecQ5. **A.** SDS-page analysis of the sample from Ni-affinity and Heparin-affinity purifications; the band corresponding to the target protein is shown with a triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin; INJ HEPA: Injected sample; FT HEPA: Flow Through sample; HEPA peak: Eluted fractions from the main peak. **B.** Chromatogram of the Heparin chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the proteins bound to heparin molecules. **C.** SDS-PAGE analysis of the protein samples before and after TEV protease cleavage, and collected fractions from Size Exclusion chromatography purification. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); PRE TEV: before TEV cleavage; POST TEV: protein after tag-cleavage; FT POST TEV: Flow Through from Ni-NTA resin; EL POST TEV: eluted fraction from Ni-NTA resin; INJ SEC: Injected sample; SEC fractions: Eluted fractions from the main peak. **D.** Elution profile from a Superdex 200 column; the fraction are shaded in blue.

13.7. Purification of RecQ4 N-terminal fragments

It has been reported that the N-terminal domain of RecQ4 has a role in interacting with G-quadruplex structures (Papageorgiou et al., 2023). In order to confirm these data and to better understand the role of the Sld2-like N-terminal region, we cloned multiple fragments of the first 427 amino acids of RecQ4 to perform a systematic analysis of the interaction with G4s. It has to be stressed that the RecQ4 N-terminal region contains long regions that are highly disordered: expressing and purifying intrinsically disordered proteins presents a challenge, as the proteins tend to aggregate or are prone to significant amount of proteolysis, and often one needs to compromise purity against speed of purification, in order to obtain a reasonable amount of sample.

Fragments encompassing 1-55, 1-133, 1-198, 1-251, 1-368, 1-391 and 1-427 amino-acid residues were cloned into a pNH-TrxT vector to produce a 6His-thioredoxin-tagged protein (Figure 27).

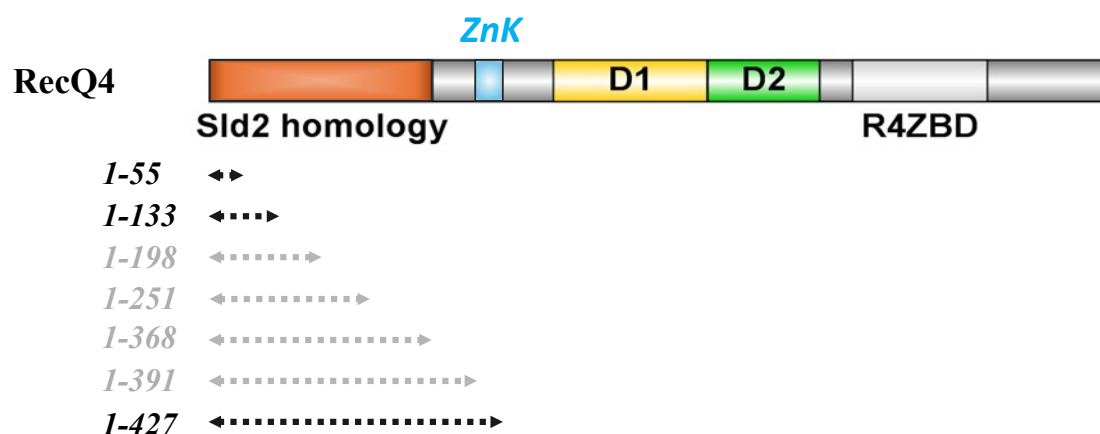


Figure 27. Schematic representation of N-terminal domains of RecQ4 helicase. The catalytic domain is shown in yellow and green (HD1 and HD2 domains respectively), the Zn-knuckle domain in cyan, the Sld2 homology domain in orange and the R4ZBD in light grey. Black arrows underline proteins boundaries. In black the proteins expressed and purified; in grey the proteins whose expressions need to be optimized.

13.7.1. Small scale expression test

For each construct, we typically tested a large number of bacterial strains (BL21 DE3, BL21 Star, BL21 Star pRare2, Rosetta2, etc.), various expression temperatures (18°C and 27°C), and other expression parameters (IPTG concentration, expression time, etc.).

Preliminary results show that it is possible to produce a good amount of RecQ4 fragments 1-55, 1-133 and 1-427 in bacterial cells. As the most encouraging result was with the 1-55 and 1-133 constructs

in BL21 Star strain (Figure 28A and B) and with the 1-427 fragment in BL21 strain (Figure 29C), further purification attempts were performed focusing on these three proteins.

Few attempts were carried out for the other fragments (1-198, 1-251, 1-368, 1-391), with no promising results (data not shown). More time and effort are needed to explore other expression conditions.

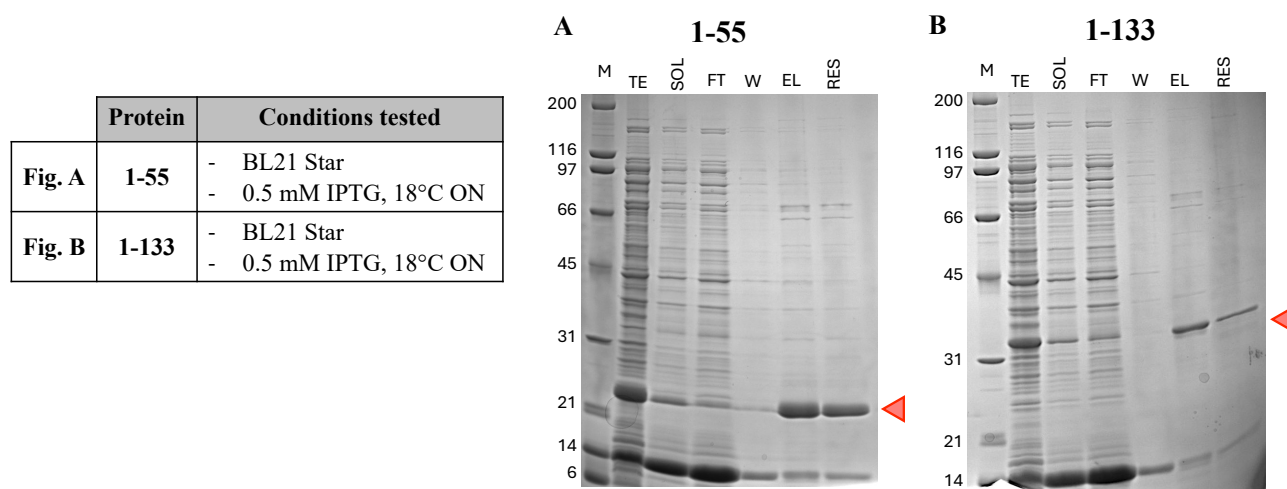


Figure 28. An example of small-scale expression and purification test for two different N-terminal domain fragments 1-55 and 1-133 of RecQ4. SDS-PAGE analysis of the fractions from the small-scale expression test, for two different constructs: **A.** 1-55 and **B.** 1-133, in BL21 Star cells. Samples from the different stages of the purification protocol are examined. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin; RES: sample of Ni-NTA resin. The expected position of the band corresponding to the fusion protein is shown with a triangle.

	Conditions tested
Fig. A	- BL21 Star pRare2 - 1 mM IPTG, 18°C ON
Fig. B	- Rosetta 2 - 1 mM IPTG, 18°C ON
Fig. C	- BL21 - 1 mM IPTG, 18°C ON
Fig. D	- BL21 - 1 mM IPTG, 27°C ON
Fig. E	- BL21 Star - 0.5 mM IPTG, 18°C ON
Fig. F	- BL21 Star - 0.5 mM IPTG, 18°C ON

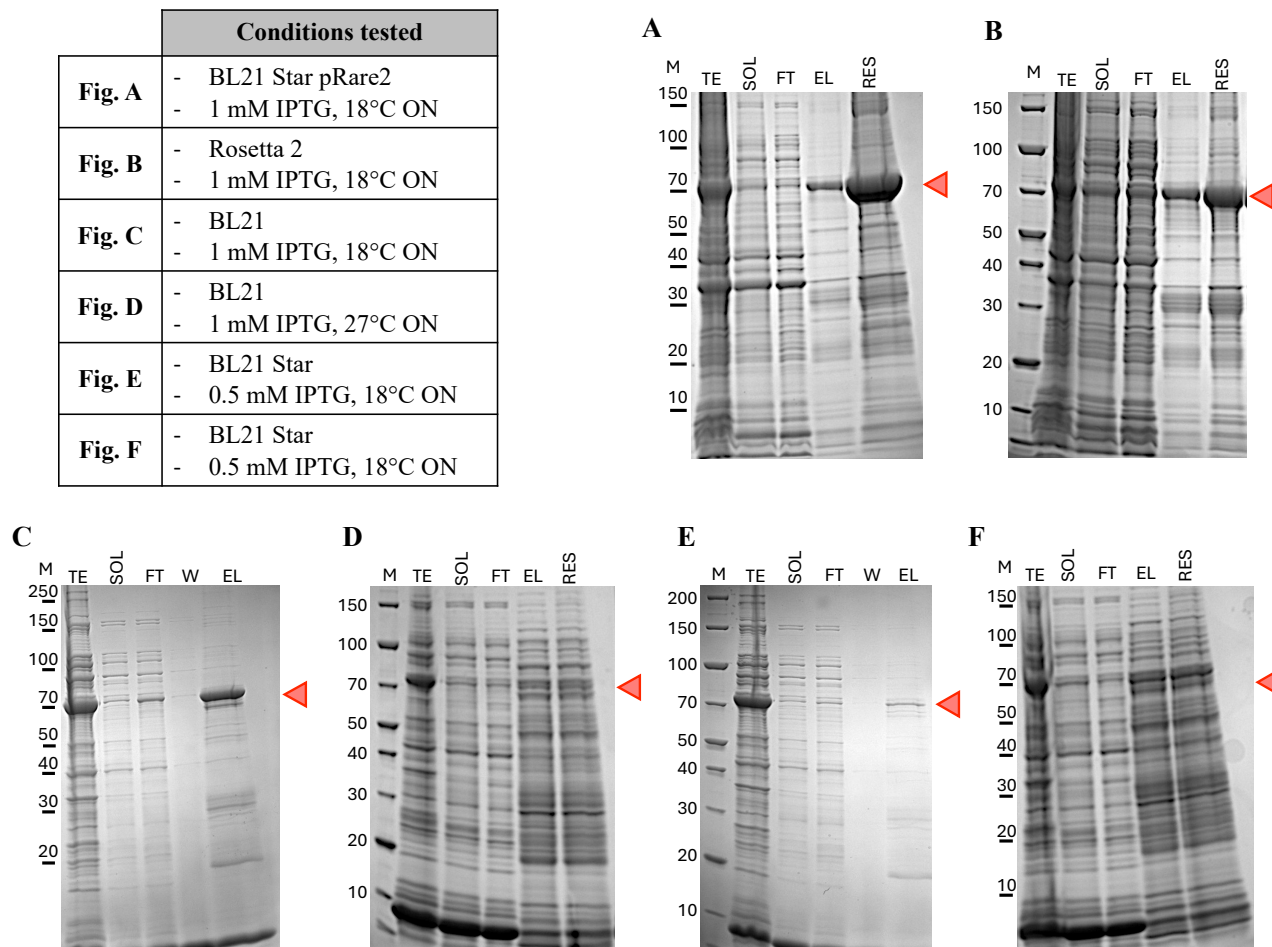


Figure 29. Examples of small-scale expression and purification tests for the N-terminal domain fragments 1-427 of RecQ4. SDS-PAGE analysis of the fractions from the small-scale expression test for 1-427 construct in various cell strains and conditions. Samples from the different stages of the purification protocol are examined. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin; RES: sample of Ni-NTA resin. The expected position of the band corresponding to the fusion protein is shown with a triangle.

13.7.2. Large scale expression

1-55 and 1-133 RecQ4 fragments were expressed in TB at 18°C overnight, following induction with 0.5 mM IPTG and purified by Ni-affinity chromatography and cation exchange chromatography, which successfully removed a significant portion of impurities (Figures 30A, B, C for 1-55; Figure 31 for 1-133). As the fragments are rather small, the tag was not cleaved to avoid destabilizing the proteins. After these two steps, fragment 1-55 was injected on a gel filtration column Superdex™ 75 10/300 GL (Figure 30D and E). The protein ran on Size Exclusion chromatography as a monomer with a molecular weight of about 20 KDa (consistent with the presence of a His-thioredoxin tag of

about 14 kDa), and could be purified in large amounts (final yields of about 3 mg per liter of cell culture). For the fragment 1-133, the fractions from ion exchange chromatography step were pulled and directly concentrated (final yields of about 100 µg per liter of cell culture). No further purification step using a gel filtration column was performed, due to time constraints and since we judged the protein to be pure enough for the biochemical characterization (Figure 31B).

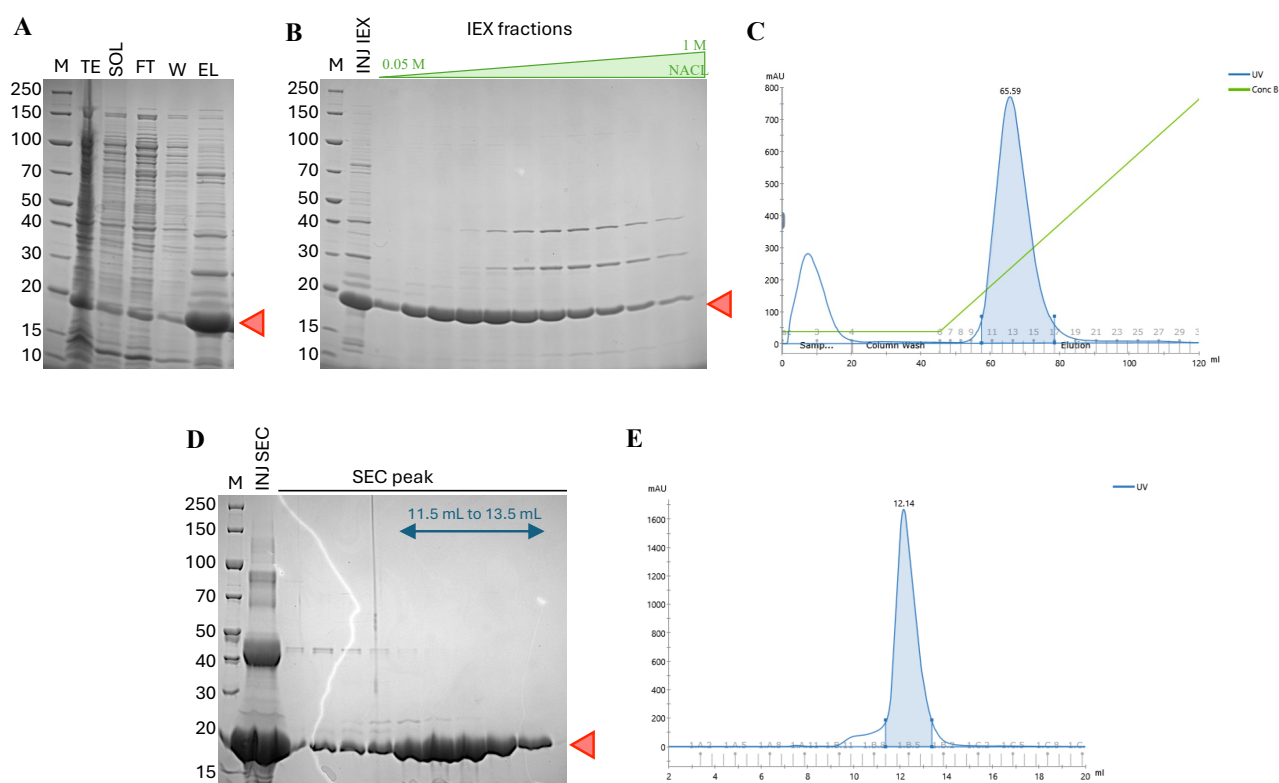


Figure 30. Purification steps of 1-55 RecQ4 fragment. **A.** SDS-page analysis of the sample from Ni-affinity purification; the band corresponding to the target protein is shown with a triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin. **B.** SDS-page analysis of cation exchange chromatography purification. Labels as follow: M: Molecular Weight standards (in KDa); INJ IEX: Injected sample; IEX fractions: Eluted fractions from the main peak. **C.** Chromatogram of the cation exchange chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the protein. **D** SDS-page analysis of Size Exclusion chromatography purification. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); INJ SEC: Injected sample; SEC peak: Eluted fractions from the main peak. **E.** Elution profile from a Superdex 75 column; the fraction are shaded in blue.

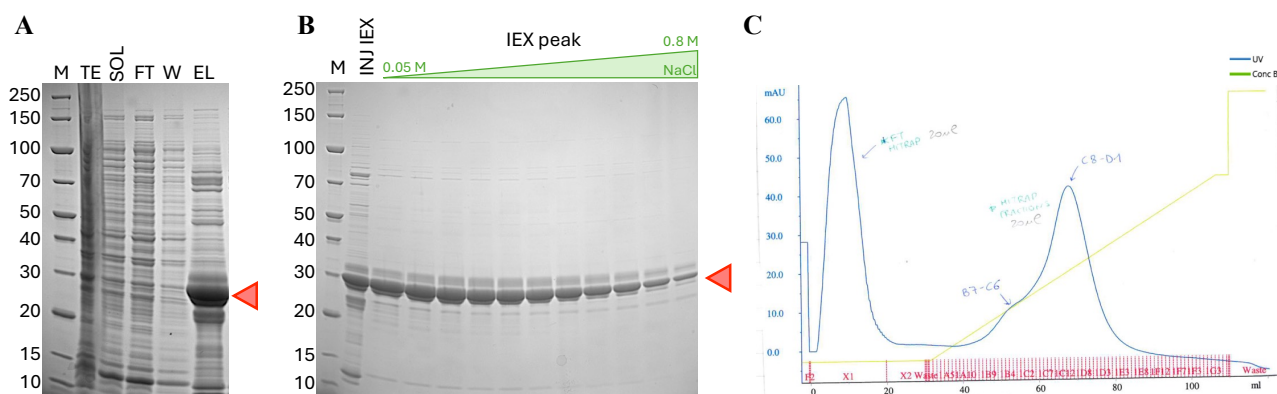


Figure 31. Purification steps of 1-133 RecQ4 fragment. **A.** SDS-page analysis of the sample from Ni-affinity purification; the band corresponding to the target protein is shown with a triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin. **B.** SDS-page analysis of cation exchange chromatography purification. Labels as follow: M: Molecular Weight standards (in KDa); INJ IEX: Injected sample; IEX peak: eluted fractions from the main peak. **C.** Chromatogram of the cation exchange chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the protein.

The fragment including amino acids 1-427 RecQ4 was expressed in LB at 18°C overnight, following induction with 1 mM IPTG and purified by Ni-affinity chromatography and Heparin chromatography as a final step (Figure 32). Although the small-scale Nickel purification from the construct expressing 1-427 RecQ4 fragment showed a reasonable amount of the target protein, other prominent bands appeared during the large-scale purification steps (Figure 32B), which are likely to represent proteolytic fragments. Due to the large proteolysis, we decided not to proceed with further purification steps. Previously, we carried out some attempts injecting the protein in the gel filtration Superdex 75 column, but the increasing amount of time needed for the last step resulted in amplified proteolysis and thus in protein lost. This protein was used for a preliminary biochemical characterization, although a better optimization of the purification protocol is needed.

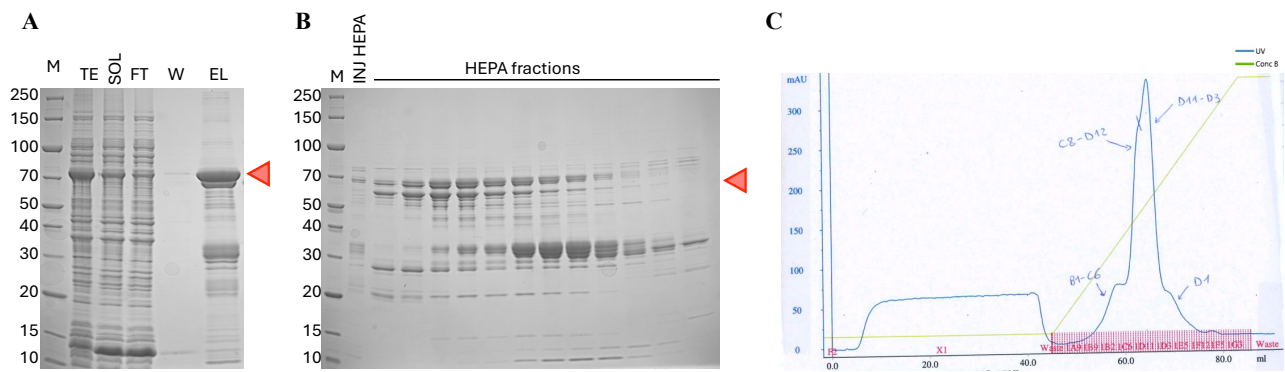


Figure 32. Purification steps of 1-427 RecQ4 fragment. **A.** SDS-page analysis of the sample from Ni-affinity purification; the band corresponding to the target protein is shown with a triangle. Labels as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin; **B.** Heparin chromatography purification. Labels as follow: M: Molecular Weight standards (in KDa); INJ HEPA: Injected sample; HEPA fractions: Eluted fractions from the main peak. **C.** Chromatogram of the Heparin chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the proteins bound to heparin molecules.

14. Assembly and characterization of G-quadruplex structures

G-quadruplexes are highly polymorphic, adopting many diverse topologies, conformations and arrangements, depending on the exact nucleotide sequence, the loop length and the presence of flanking nucleotides; moreover, the exact conformations they assume and their relative stabilities are affected by several environmental factors, such as the temperature, the ionic strength, the presence of specific ions and buffer conditions. For the present study we choose G4s that are well-characterized in literature and whose main features are listed in Table 11.

The majority of the G4 structures fold in a parallel conformation and comprises the G4s from the promoters of cMYC (PDB ID: 1XAV), cKIT1 (PDB ID: 2O3M), KRAS (PDB ID: 5I2V) and Mcm2, and the G4s found in the intronic regions of ρ -globin (ρ -globin G4) and BU1A (BU1A (+3.5) G4). The three G-tetrads of the cMYC G4 (PDB ID: 1XAV) are linked by three double-chain-reversal loops, two formed by single nucleotide (T) and one by a double nucleotide (TA), which arrangement results in a very stable G-quadruplex conformation (Ambrus et al., 2005). cKIT1 G4 (PDB ID: 2O3M) is peculiar since its structure features an isolated non-G-tract guanine that participates in the formation of the G-tetrad core, despite the presence of four G-tracts containing three consecutive guanines. Another significant difference with cMYC G4 is the loop composition. cKIT1 has four loops: two single nucleotide loops forming two double-chain-reversal loops; a two nucleotides loop that connects two adjacent corners of the G-tetrad core; and a five-residue segment that allows the terminal Gs to be inserted back to the core (Phan et al., 2007). KRAS G4 (PDB ID: 5I2V) forms a parallel structure with three G-quartets connected by a four--nucleotide loop and two one-residue double-chain reversal loops. Interestingly, the presence of a thymine bulge between the G-quartets can be potentially used as a target for small ligands (Kerkour et al., 2017). Less is known about Mcm2 G4, which has only been characterized by CD spectroscopy and UV melting experiments, but it is expected to have three G-tetrads and short loops connecting the strands (Huppert et al., 2007). Both ρ -globin and BU1A (+3.5) G4s have been implicated in epigenetic instability, but structural studies on these specific G4s are limited. The “+ 3.5” refers to the identified G4 motif localized on the coding strand, 3.5 kb downstream to the transcription start site of the *BU-1* locus (Sarkies et al., 2012). They are characterized by three stacked G-tetrads with short loops (Sarkies et al., 2010, Sarkies et al., 2012).

Bcl2 Mid G4 (PDB ID: 2F8U), G4#2 [(GGGTTT)₄] and the human telomeric G4 (Tel₂₆, PDB ID: 2JPZ) sequences formed hybrid [3 + 1] G4 structures containing three parallel and one antiparallel G-strands. Loop configuration differs from one G4 to another. Bcl2 Mid G4 has two lateral loops of different lengths and a third single-nucleotide double-chain reversal side loop, which interact with the core G-tetrad in a specific way, defining and stabilizing the unique G4 structure (Dai et al., 2006).

Tel₂₆ G4 adopts a Hybrid-2 type (which differs from Hybrid-1 type in its loop arrangement and strand orientation) with the three TTA loops arranging in lateral-lateral-side conformation. Adenine residues, together with the 3'-flanking segment (AA), are involved in the formation of a capping structure that stabilizes the hybrid-2 folding (Dai et al., 2007). G4#2 G4 is a simplified sequence of the human telomeric one, which is found in telomeres of ciliates of the genus *Paramecium*, that can be used as a starting point to evaluate the impact of loop composition to protein binding (Lago et al., 2017). Interestingly, the sequence variant (CTAGGG)_n (PDB ID: 2KM3) of the human telomeric sequence adopts a novel antiparallel conformation, involving two G-tetrads sandwiched between a G•C base pair and a G•C•G•C tetrad (Lim et al., 2009). The G-rich sequence G4#4 named *T30695* (or *T30923*) (PDB ID: 2LE6) forms a dimeric all-parallel-stranded G-quadruplex structure (Do et al., 2011).

G4	PDB	Sequence (5'-3')	Topology	Reference
cMYC	1XAV	TGAGGGTGGGTAGGGTGGGTAA	promoter parallel	Ambrus et al., 2005
cKIT1	2O3M	TAAGGGAGGGCGCTGGGAGGAGGG	promoter parallel	Phan et al., 2007
KRAS	5I2V	AGGGCGGTGTGGGAATAGGGAA	promoter parallel	Kerkour et al., 2017
Mcm2		GAAGGGACACGAGGGGCGGCCAGAGGGTCC	promoter parallel	Huppert et al., 2007
BU1A (+3.5)		TAAGGGCTGGGTGGGTGCTGTCAAGGGCTGGG	intronic region, parallel	Sarkies et al., 2012
ρ-globin		TAAGGGGAGTAAAAGGGAGCGGGGTGCTGGG	intronic region, parallel	Sarkies et al., 2010
Bcl2 Mid	2F8U	ATAGGGCGCGGGAGGAAGGGGGCGGG	promoter hybrid	Dai et al., 2006
Tel₂₆	2JPZ	TTAGGGTTAGGGTTAGGGTTAGGGTT	telomere hybrid	Dai et al., 2007
(CTAGGG)_n	2KM3	TAAGGGCTAGGGCTAGGGCTAGGG	telomere antiparallel	Lim et al., 2009
G4#2		TAATGGGTTTGGGTTTGGGTTTGGGT	synthetic hybrid	Lago et al., 2017
G4#4 (T30695)	2LE6	TTTTGGGTGGGTGGGTGGGTTTT	dimeric parallel	Do et al., 2011

Table 11. List of G-quadruplexes used in the present study. Sequence and characteristics of each G4 sequence. Whenever a structure is available, the PDB ID is provided. Gs that participate in G4 formation are in bold.

Since one of the main factors influencing G4 topology and stability is the nature and the concentration of metal ions, which either lay between G-tetrads or stay within the G-tetrad plane, it is crucial to characterize G4 structures based on the buffer composition used in biochemical assays. In particular, it is known that among the monovalent cations, K⁺ ions stabilize G4s more than Na⁺ ions, which stabilize more than Li⁺ ions. Also changing the relative concentrations of ions can affect the stability.

14.1. CD spectra of G4s

Circular dichroism (CD) spectroscopy is a primary tool for the characterization of G-quadruplex structures, as this technique not only provides information on the nature of the structure (i.e. whether the nucleotide sequence folds as a G4 or not) but also on the topology of the quadruplex. Different G4 structures, arising from differences in G-quartet stacking, strand segment orientation, and loop arrangement, display unique CD spectra signatures. Generally, parallel G4s adopt a positive ellipticity at 260 nm and a negative one at 240 nm; hybrid G4s exhibit a negative peak at around 245 nm and two positive bands at around 260 nm and 295 nm, the latter being predominant; the CD pattern of antiparallel G4s shows a negative band at 260 nm and two positive peaks at 240 nm and 295 nm (Figure 33).

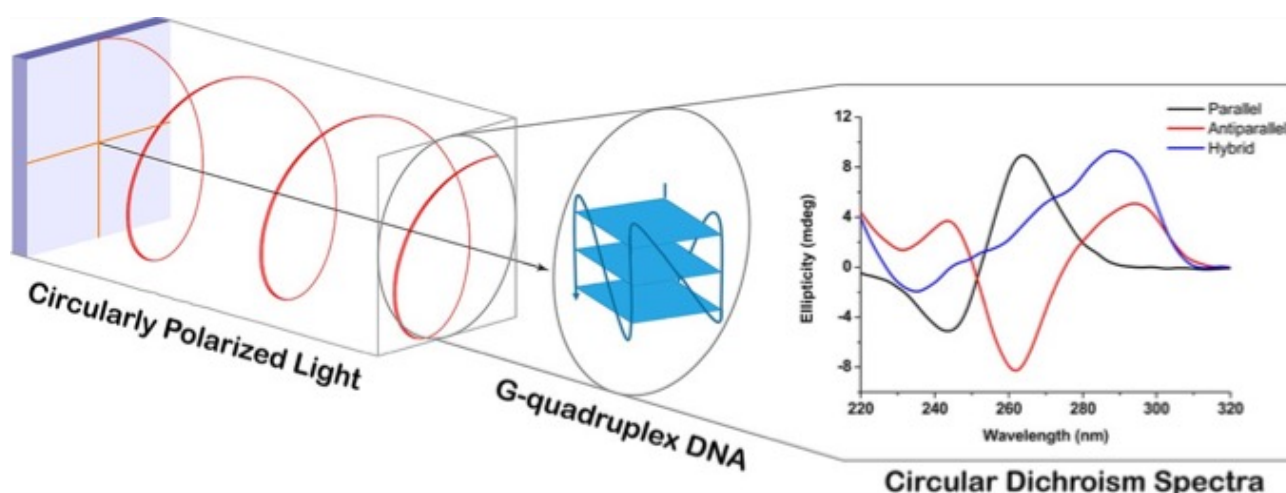


Figure 33. Schematic representation of CD experiment with representative CD spectra of G4 structures with different topologies: parallel G4 in black, antiparallel G4 in red and hybrid G4 in blue). From Carvalho et al. (2017).

To clarify whether the G4s used in the present study are mostly folded or unfolded under the selected conditions and under different buffer compositions used for biochemical studies, we performed CD spectroscopy to confirm G4 formation and characterize the general topology. It is important to highlight that at first sight, the CD spectra of a parallel G4 can be very similar to the one of dsDNA duplex. For this reason, the use of Li^+ ions can help since differences in ellipticities between potassium or lithium-only conditions may be interpreted as evidence of G4 formation.

CD spectroscopy was applied to analyze all the G4 sequences in the presence of 100 mM KCl, 150 mM KCl, 1 mM KCl and 99 mM KCl, or 100 mM LiCl. CD spectra were acquired as described in section 11.6 in Materials and Methods. Each of the recorded spectra corresponded to a typical G4

topology, revealing conformational changes in response to varying ion concentrations. In most cases, the presence of Li^+ ions disrupted G4 folding, resulting in nearly flat or atypical CD spectra, further supporting robust G4 formation under other conditions tested.

cMYC, cKIT1, BU1A (+3.5) and G4#4 G4s adopt a parallel conformation in all the conditions tested, showing very little differences in the maximum ellipticity signal at 260 nm (Figure 34A, B, E, F). Different is the situation of KRAS and Mcm2 G4s, which also adopt a parallel conformation in the presence of high K^+ ions concentrations (green and yellow spectra in Figure 34C and D): in the presence of only 1 mM KCl and 99 mM NaCl, the CD spectra overlap with the ones acquired in presence of 100 mM LiCl, implying the fact that the G4s is not well-formed in the case of KRAS (Figure 34C), or it assumes an hybrid-like conformation as seen for Mcm2 (Figure 34D). ρ -globin G4 is peculiar since it adopts a mainly parallel topology, but the CD spectrum shows some amount of peak at 290 nm suggesting partial formation of antiparallel structure (Figure 34G). In contrast with literature that reports Bcl2 Mid G4 folding as a hybrid-type G4, in our hand the same G4 sequence adopts a parallel conformation (Figure 34H).

The antiparallel conformation is the rarest folding assumed by known G4s. (CTAGGG)_n G4 structure adopts an antiparallel conformation in all the conditions tested, however showing lower ellipticity when in the presence of low K^+ concentrations (red spectra in Figure 34M).

G4#2 G4 adopts a hybrid conformation with two positive peaks around 260 nm and 290 nm and a negative peak at 240 nm (Figure 34I). As already reported in literature (Dai et al., 2007), the most polymorphic G4 is the telomeric G4 (Tel₂₆). The hybrid-type structure appears to be the major conformation formed in human telomeric sequences in K^+ solution, confirmed by the CD spectra acquired in the presence of high concentrations of potassium. Notably, in the presence of low K^+ ions (as in the case of 1 mM KCl and 99 mM NaCl, red spectra in Figure 34L), the antiparallel conformation prevails, showing two positive peaks at 290 nm and 240 nm and a negative peak at 260 nm (Figure 34L).

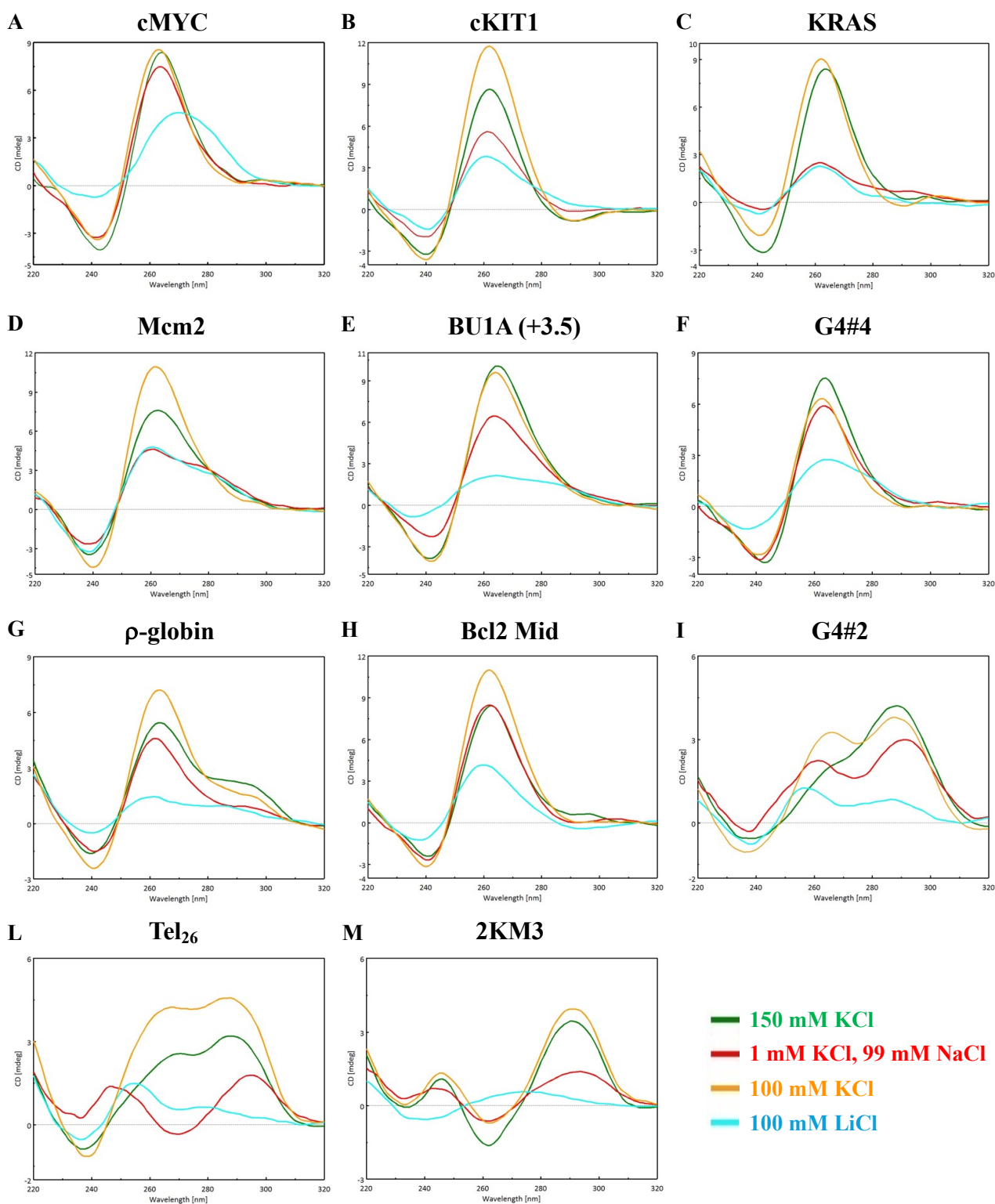


Figure 34. CD characterization of G4s folding. A. cMYC, B. cKIT1, C. KRAS, D. Mcm2, E. BU1A (+3.5), F. G4#4, G. ρ -globin, H. Bcl2 Mid, I. G4#2, L. Tel₂₆, M. 2KM3 G4s in the presence of 150 mM KCl (green spectra), 100 mM KCl (yellow spectra), 1 mM KCl and 99 mM NaCl (red spectra) and 100 mM LiCl (cyan spectra).

14.2. Variations of G4s melting temperature in different buffers

CD spectra can be useful to monitor changes in the secondary structures of G4s during thermal denaturation experiments. We also performed this type of assay to assess the stabilities of the G4 structures used in the study, in the presence of different buffer compositions. All the G4s spectra show an increasing stability in the presence of higher K^+ concentration. The high stability ($T_m > 70^\circ\text{C}$) of many G4 structures at physiological K^+ concentration (100 mM or 150 mM KCl, yellow and green spectra respectively in Figure 35) is an indication of their potential ability to fold also in a cellular environment and to play a biological role in the regulation of gene expression. In some cases, the high potassium concentration induces a G4 stability higher than 90°C , which is the upper limit of the instrument sensitivity. Moreover, the presence of lithium significantly lowers the stability of G4 structures with some exceptions in which the G4 structures seem to be very stable even in non-folding conditions (e.g. Mcm2 G4 and telomeric G4).

These results confirm that G4 stability is influenced by the potassium concentration used in the assay, raising the question of whether helicase activity also varies in response to G4 stability under these specific conditions. Specifically, in the presence of 1 mM KCl and 99 mM NaCl, most of the G4 structures exhibit reduced stability, supported by differences in the melting temperatures (ΔT shown in Figure 35), which is sometimes similar to that observed under lithium-only conditions. Therefore, helicase activity should be assessed under both low and high potassium concentrations, the latter being more physiologically relevant.

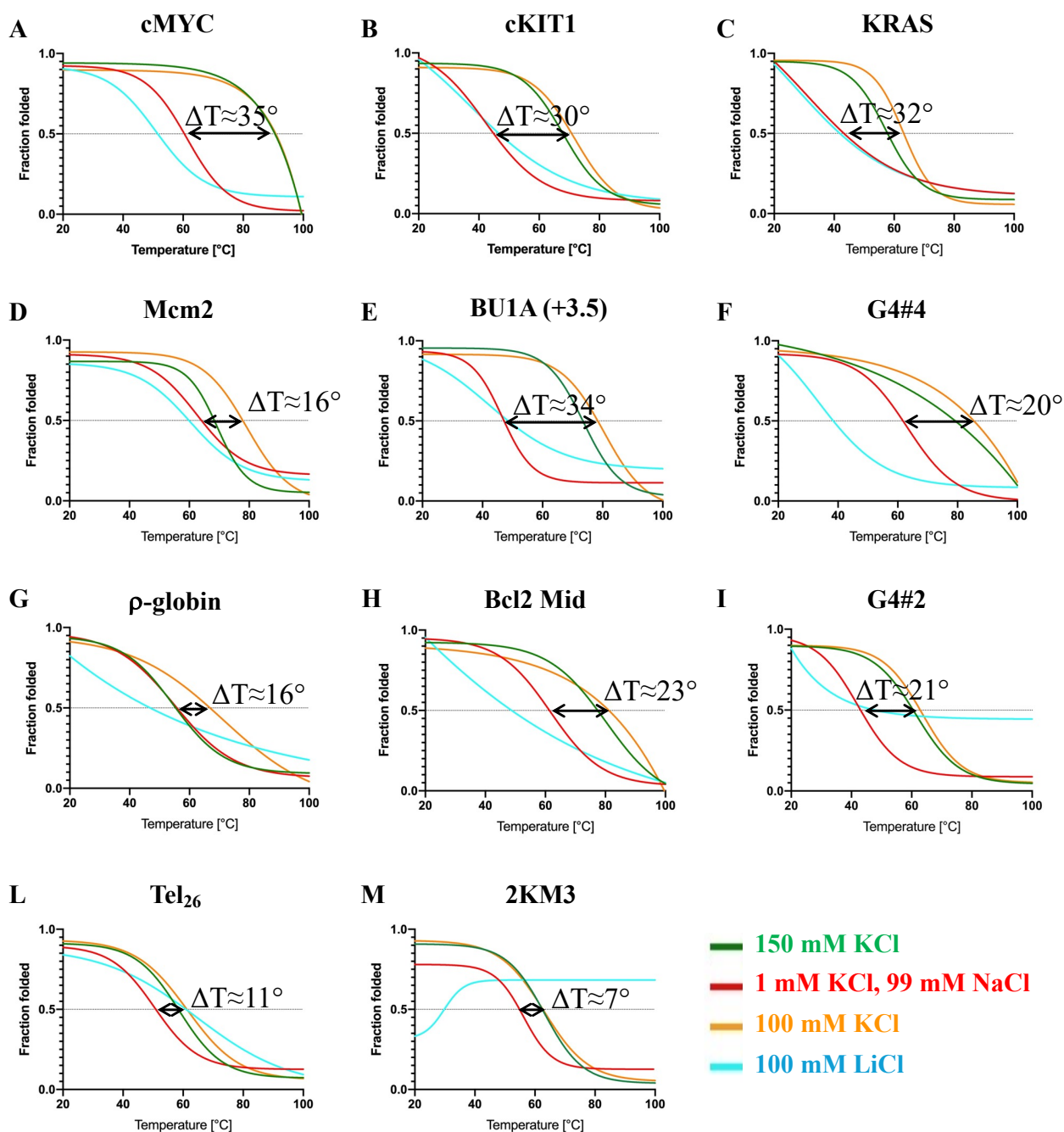


Figure 35. Thermal stability of G4 structures in different buffers. A. cMYC, B. cKIT1, C. KRAS, D. Mcm2, E. BU1A (+3.5), F. G4#4, G. ρ-globin, H. Bcl2 Mid, I. G4#2, L. Tel₂₆, M. 2KM3 G4s denaturation curves in the presence of 150 mM KCl (green spectra), 100 mM KCl (yellow spectra), 1 mM KCl and 99 mM NaCl (red spectra) and 100 mM LiCl (cyan spectra). ΔT is the difference of melting temperature between 100 mM KCl and 1 mM KCl/99 mM NaCl conditions.

15. RecQ helicases: G4s interactions

Many reports have shown that RecQ helicases can bind and unwind a variety of DNA and RNA substrates (Lu & Davis, 2021). The goal of this study was to further investigate their interaction with more non-canonical structures, in particular G-quadruplexes. Before focusing on G4 interaction, we first examined the activity of RecQ fragments on forked duplex DNA to confirm their ability to unwind DNA substrates. This served as an internal control to ensure that the helicases we were working with were active, allowing us to confidently conclude that any observed lack of G4 unwinding activity was due to the helicase inability to target these structures, rather than a lack of activity of the proteins.

15.1. The catalytic domain of human RecQ helicases are active helicases able to unwind forked duplex substrates.

Experiments to evaluate RecQ helicases activity were carried out using a DNA Fork substrate with a 20-nucleotide base-paired region and a 26-nucleotide 3' tail. For the helicase assay, a DNA Fork with one strand labeled with 6-FAM was incubated with increasing concentrations (0-100 nM) of RecQ1, BLM, WRN, RecQ4, and RecQ5 in a buffer containing a capture strand (to prevent reannealing) and Mg^{2+} -ATP, as described in the Materials and Methods section 11.5.1. The generation of single-stranded DNA was monitored via gel electrophoresis.

The results demonstrated that RecQ helicases efficiently unwind the DNA fork substrate, supporting that RecQ fragments used in this study are active, even at low protein concentrations (100 nM) (Figure 36). Notably, BLM and WRN exhibited significant activity, fully unwinding the substrate at just 10 nM. RecQ1 also showed near-complete unwinding at this concentration. RecQ4 displayed a 100% unwinding at 100 nM. Although RecQ5 did not achieve a high percentage of unwinding at 100 nM (approximately 50%), the experiments confirmed that it is an active helicase. Our results are consistent with what reported in the literature: RecQ5 is known to be a weak helicase that shows very low unwind activity towards forked duplex DNA. The protein requires the ssDNA-binding protein RPA in order to mediate the efficient unwinding of oligonucleotide-based substrates. In particular, RecQ5 fragment 1-475, which contains the entire helicase catalytic core but lacks the C-terminal region, exhibits a significantly greater DNA helicase activity than the full-length enzyme, suggesting that we are working with the most-active RecQ5 fragment (Garcia et al., 2004).

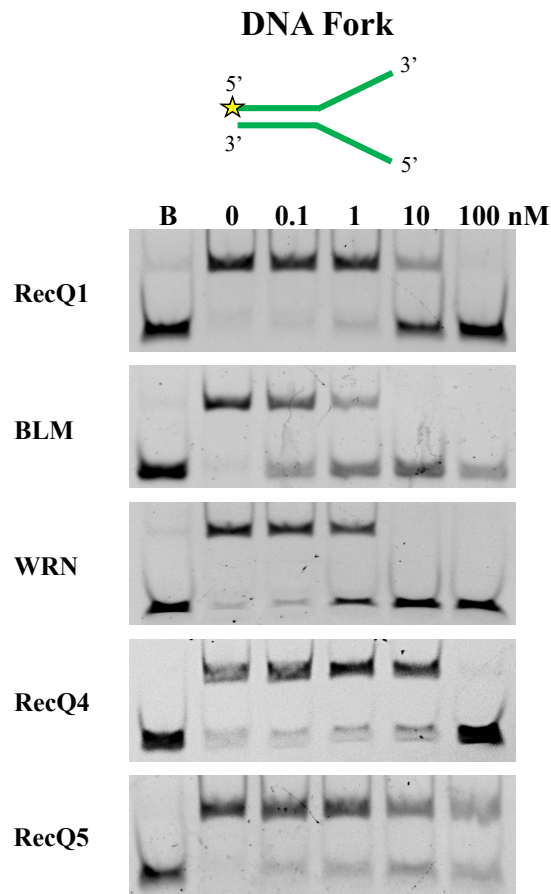


Figure 36. Comparison of RecQ helicases activity on the equivalent DNA Fork substrate. Gel-based helicase assays were carried out with 10 nM DNA Fork at increasing concentrations of protein (from 0.1 to 100 nM). The substrates were resolved on 8% polyacrylamide gels supplemented with SDS. The substrate was denatured in the absence of helicase ("Boiled") to serve as a control.

15.2. Human RecQ helicases unwind non-canonical substrates

Since RecQ helicases have been shown to bind and unwind various non-canonical nucleic acid structures, a parallel analysis was performed also on displacement loops, including D-loops and R-loops. These substrates are characterized by an invading DNA or RNA strand that anneals with a DNA bubble scaffold, and they represent pivotal intermediates of homologous recombination.

In our lab, we carried out an additional study focused on a detailed biochemical characterization of the full-length (FL) protein of RecQ4. Together with Luisa MR Napolitano, I tested the binding and unwinding activity of full-length RecQ4 on DNA Fork, bubble, D-loops, and R-loops structures, including also loops with different tail orientations. Both binding and unwinding were analyzed using gel-based assays, incubating increasing concentrations of full-length RecQ4 (0-200 nM for binding and 0-240 nM for unwinding) with various loop substrates.

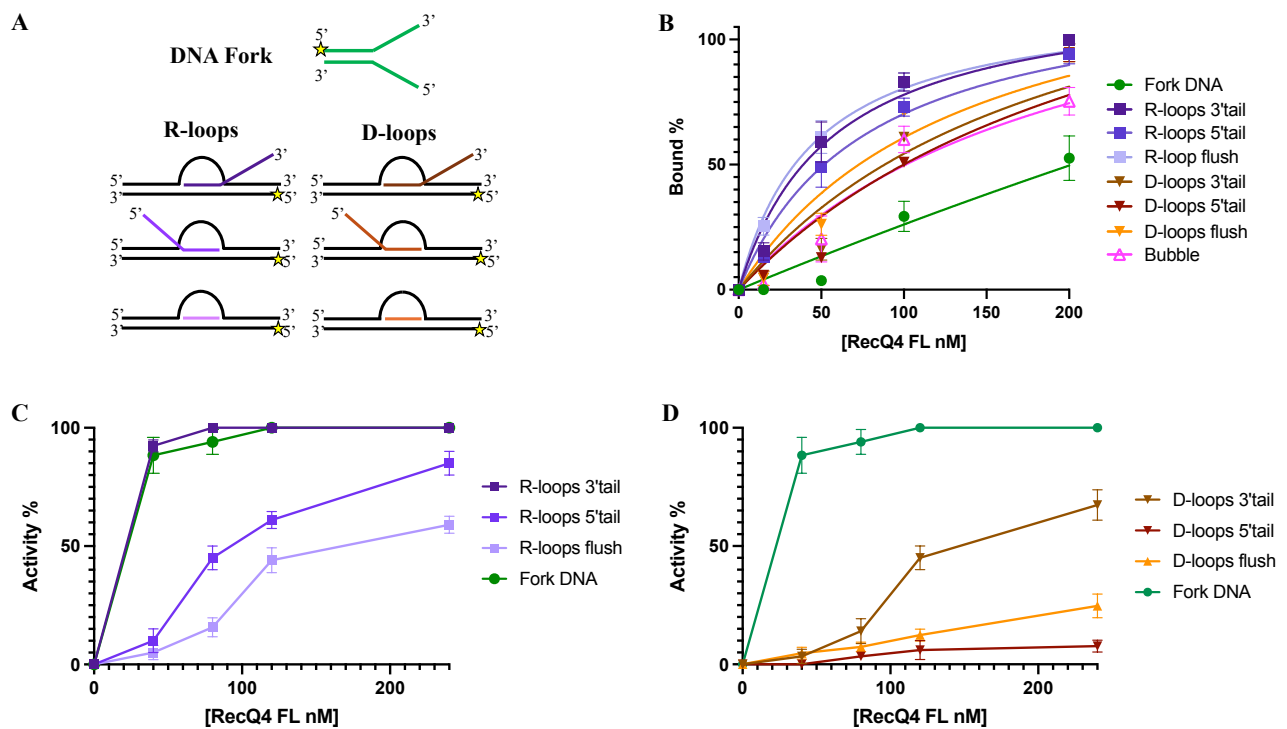


Figure 37. Full-length (FL) RecQ4 binding and unwinding activity on loop substrates. **A.** A schematic diagram of the substrates used. **B.** Comparison of FL RecQ4 binding activity towards, DNA Fork, R-loops, D-loops and bubble. **C.** Comparison of helicase activity of FL RecQ4 on different R-loops substrates. **D.** Comparison of helicase activity of FL RecQ4 on different D-loops substrates.

Our data showed that full-length RecQ4 binds R-loops with higher affinity than D-loops and bubble structures (Figure 37B). Among them, for both the R-loops and D-loops, the 3' tail and flush structures appeared to be bound more efficiently than the 5' tail ones. FL RecQ4 also most efficiently unwound the R-loops substrates over the DNA Fork and the D-loop (Figure 37C and D) with a clear preference for the 3' tail structure, consistent with its 3' to 5' directionality. Considering all these results, R-loops seemed to be the preferred substrate followed by the D-loops.

In order to have a broader understanding of loop substrates specificities among human RecQ helicases, we then compared the R-loops and D-loops resolution activity of the catalytic domain of RecQ1, BLM, WRN, and RecQ5.

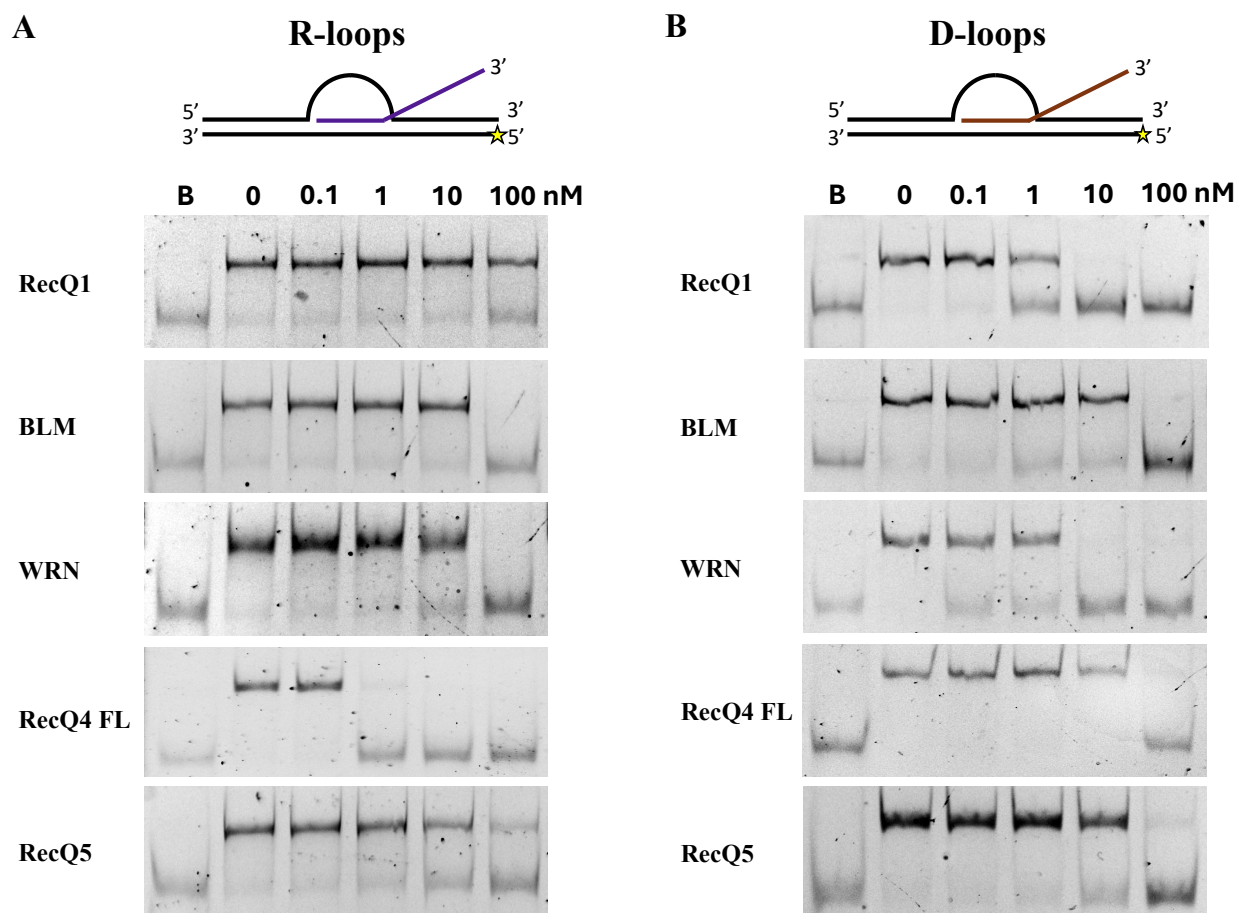


Figure 38. Comparison of RecQ helicases activity on the equivalent A. R-loops and B. D-loops substrates. Gel-based helicase assays were carried out with 10 nM R-loops or D-loops at increasing concentrations of protein (from 0.1 to 100 nM). The substrates were resolved on 8% polyacrylamide gels supplemented with SDS, as described in Material and Methods. The substrate was denatured in the absence of helicase ("Boiled") to serve as a control.

Using the same concentration range used for the DNA Fork assay (0-100 nM), all tested human RecQ helicases are able to unwind both R-loops and D-loops structures. RecQ1 and WRN both exhibited a net preference for the DNA Fork (Figure 37) and D-loops (Figure 38B), even if for the R-loops a slight activity has been observed in RecQ1 and a good resolution ability has been observed in WRN (Figure 38A). BLM was highly proficient in unwinding both loop substrates, albeit it displayed a preference for the DNA Fork (Figure 38). As previously noted, RecQ5 showed the lowest activity overall, although its ability to unwind these non-canonical substrates was greater than its activity toward the DNA Fork (Figure 38). Taken together, these findings indicate that, while both DNA Forks and D-loops were efficiently unwound by several RecQ helicases, RecQ4 has a clear preference for R-loops, making it the most efficient human RecQ helicase for resolving this substrate.

It has to be stressed that here we are comparing the full-length RecQ4 with the catalytic domains of RecQ1, BLM, WRN, RecQ5; this is justified by the fact that in most of these cases it has been shown that there isn't a significant difference in the unwinding properties of the full-length protein vs the catalytic domain, whereas in the case of RecQ4 the presence of the N-terminal and C-terminal domains significantly affects the helicase activity. However, a similar analysis performed by another PhD student is currently ongoing, and it focuses on the characterization of the interactions between displacement loop structures and the RecQ4 fragment 371-1116. As described above this is the fragment that was chosen for the analysis of the interaction with G4s, as it includes the catalytic domain and is thus more directly comparable to the catalytic domains of the other RecQ helicases.

15.3. BLM and WRN interact with and unfold G-quadruplex structures

Many studies reported the efficient activity of BLM and WRN helicases in resolving G-quadruplexes. However, these analyses mainly report the activity towards one or two different G4s, rather than carrying out a systematic evaluation of different G4 structures. For this reason, we performed a comprehensive study on multiple G4 structures with different topologies and from different regions of the genome, particularly focusing on unimolecular G4s, to assess the ability of the helicase to unwind more or less efficiently a given G4 substrate based on its nature. The complete list of G-quadruplexes used in the assays has been previously reported in Table 11, and for convenience is summarized in Table 12 below.

G4	PDB	Topology
cMYC	1XAV	promoter parallel
cKIT1	2O3M	promoter parallel
KRAS	5I2V	promoter parallel
Mcm2		promoter parallel
BU1A (+3.5)		intronic region, parallel
ρ -globin		intronic region, parallel
Bcl2 Mid	2F8U	promoter parallel
Tel ₂₆	2JPZ	telomere hybrid
(CTAGGG) _n	2KM3	telomere antiparallel
G4#2		synthetic hybrid
G4#4 (T30695)	2LE6	dimeric parallel

Table 12. List of G-quadruplexes used in the present study. Summary of G4 characteristics. Whenever a structure is available, the PDB ID is provided.

The RecQ helicases binding activity towards the selected G4 has been tested using Fluorescence Anisotropy, a technique widely used to study protein-protein and protein-nucleic acids interactions. In the case of protein-nucleic acids interactions, Fluorescence Anisotropy relies on the change in fluorescence polarization, due to the different mobility that fluorophore molecules attached to nucleic acids have in solution, if the molecular size of the protein-DNA complex is sufficiently different from the unbound nucleic acid. The measured polarization directly correlates with the fraction of bound ligand in solution, providing an easy way to monitor protein-ligand interactions. A range of RecQ helicases concentrations were used to test the binding to the G4 sequences (Table 12).

The results showed that both BLM and WRN bind a plethora of G4 structures with variable bind strength, with some G4 showing high affinities and others lower affinities, confirming that the interactions between helicases and G4 do depend on the G4 structure and topology, and that a panel of G4s need to be tested when assessing the interaction between a protein and quadruplexes (Figure 39). Interestingly, the two helicases exhibited similar substrate preferences, with the cMYC G4 and Tel₂₆ G4 showing higher affinity, and KRAS, ρ -globin, G4#2 and G4#4 being among the less tightly bound. There seem to be no clear preference for either promoter or telomeric G4 topologies, despite the expectation that WRN may have a more specific role in telomer maintenance.

Analyzing and comparing the structures of the G4 showing high affinity, or low affinity, we could not find any consistent correlation between G4 structural or topological peculiarities and helicase binding, highlighting the complexity of G4 recognition.

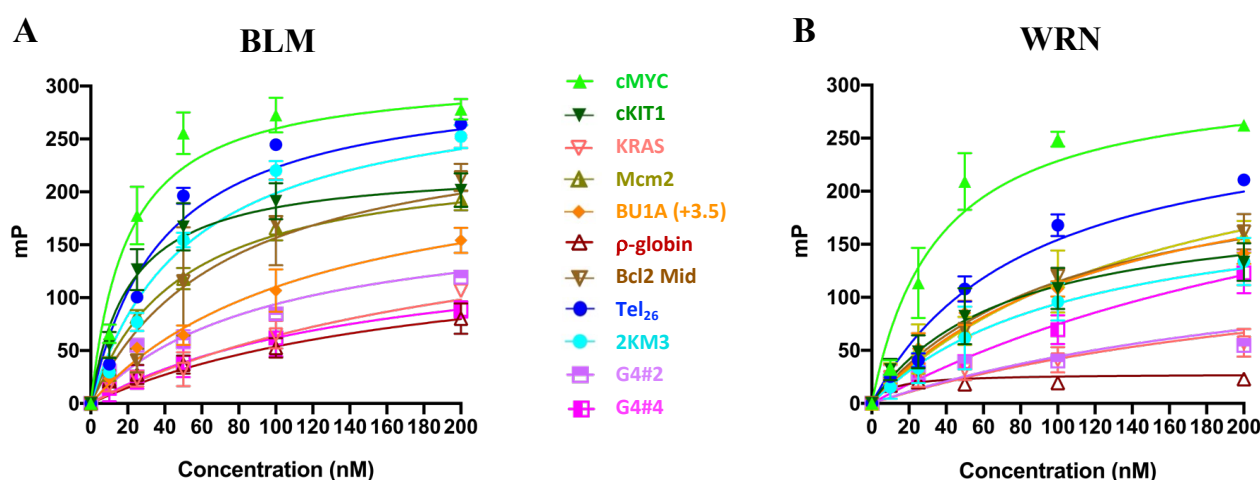


Figure 39. BLM and WRN interacting with various G-quadruplexes. Comparison of **A.** BLM and **B.** WRN binding affinity towards different G-quadruplexes, using Fluorescence polarization technique. One-site total binding function of GraphPad-Prism was used to fit the data points. Error bars indicate mean deviation between three independent experiments.

Although one would expect a helicase to recognize the preferred substrate with high affinity, the correlation between affinity and unwinding is not always linear. In order to characterize RecQ helicases activity towards G-quadruplex structures, we adapted the previously used 96-well plate fluorescent G4-based helicase assay (De Piante et al., 2023), based on the assay proposed by the laboratory of J-L. Mergny (Mendoza et al., 2015).

The assay used a DNA system consisting of a G4-forming sequence, modified with the FAM and dabcy1 quenchers to allow monitoring the unfolding of the G4 substrate by fluorescence spectroscopy. Compared to more conventional techniques based on gel electrophoresis, this technique can follow the kinetic behavior of the enzyme in real time.

Since RecQ helicases have the opposite directionality (from 3' to 5') compared to Pif1 and DinG helicase (from 5' to 3') and since they required a tail to load on the substrate, we had to adjust the G4-system used, reversing the position of the 11-nt tail from the 5'-end to the 3'-end.

Moreover, the fluorescent assay is a sensitive technique, allowing the use of low sample concentrations and reaction volumes. However, many factors can interfere with the helicase ability to open the G4 structure. Thus, I optimized the assay conditions for each RecQ helicase, such as the G4 annealing buffer (20 mM Tris pH 7.2, 1 mM KCl, 99 mM NaCl), the reaction buffer (20 mM Tris pH 7.5, 5 mM MgCl₂, 50 mM KCl, 0,1 mg/ml BSA, 1 mM TCEP), the timing of reaction (30 min for BLM and 1h for the other RecQ proteins) and the optimal protein concentration. The maximum emission observed, and the kinetics of the reaction were directly dependent on the enzyme concentration. We evaluated a range of concentrations of helicase and were able to identify a concentration in which many G4 structures were unfolded by RecQ proteins. We were able to detect a maximum enzyme activity reached with most of the substrates at BLM and WRN concentrations of 6 nM. The arbitrary 1 mM KCl/99 mM NaCl annealing buffer solution was selected since this K⁺/Na²⁺ mixture was found to be a good compromise to allow optimal activity of the protein while the quadruplex structure was properly folded.

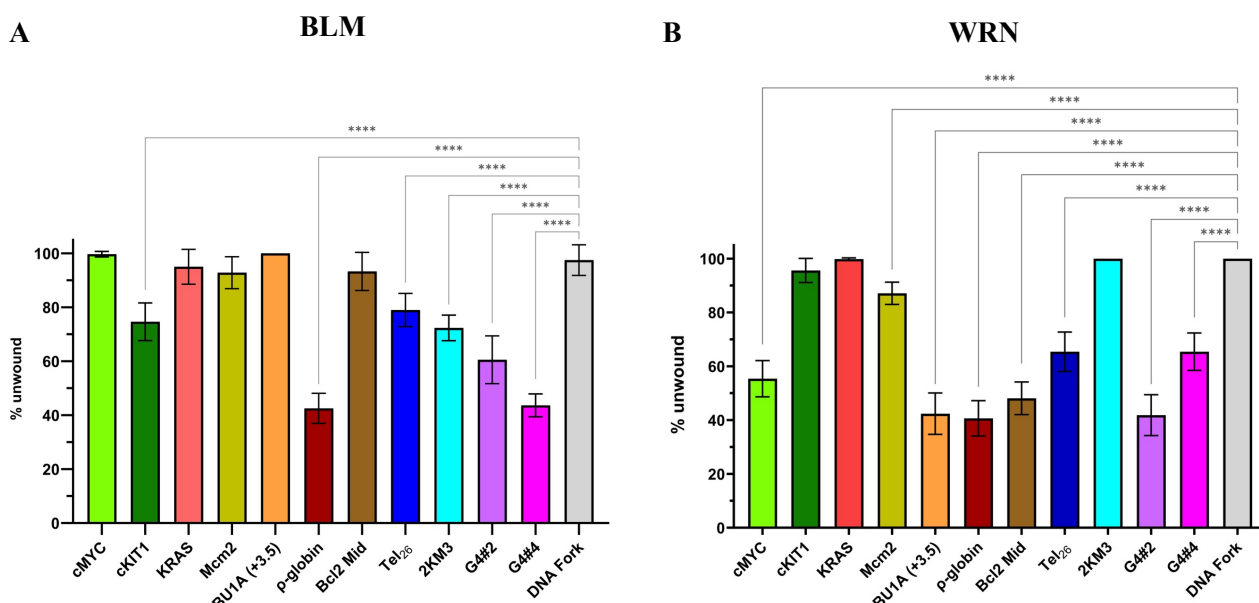


Figure 40. BLM and WRN unwinding activity towards various G-quadruplexes. Quantification of **A.** BLM and **B.** WRN helicase activity against selected G4 DNA systems; error bars indicate standard deviation of three independent experiments. DNA Fork has been used as comparison. Most appropriate statistical significance has been reported (**** represents statistical significance at $P < 0.0001$; data analyzed with GraphPad Prism, using Ordinary One-Way ANOVA test, comparing the mean of each column with the mean of the DNA Fork column, chosen as control). Helicase assays were performed in annealing buffer containing 1 mM KCl and 99 mM NaCl, and in reaction buffer with 50 mM KCl.

By analyzing multiple G4 sequences, we observed distinct selectivity and varying unwinding efficacies across different substrates. BLM emerged as the more versatile and proficient helicase, successfully unwinding the majority of the G4 substrates within 30 minutes, achieving an unwinding efficiency of over 70%. (Figure 40A). Notably, the G4 structures that were least efficiently unwound also exhibited the weakest binding (ρ -globin, G4#2 and G4#4, Figure 39A and Figure 40A), suggesting a potential correlation between helicase recognition and subsequently resolving ability. This broad activity contrasts with WRN, which, while still displaying significant G4 helicase activity, efficiently unwound a narrower range of G4 structures compared to BLM (Figure 40B). WRN required a longer incubation time (1 hour) to reach a stable unwinding plateau, indicating a slower, less versatile mechanism. Remarkably, binding affinity did not always correlate with helicase activity: for example, cMYC G4 is the preferred substrate in terms of binding but it is unwound only partially compared to other G4s such as KRAS and 2KM3, which are, on the contrary, less bound (Figure 39B and Figure 40B). This could suggest that additional mechanisms beyond simple sequence recognition may influence the unwinding process.

Moreover, taking into consideration the stability of each G-quadruplex, studied with thermal denaturation experiments (results are shown in the previous section 14.2), we wondered if differences in the unwinding rates of each protein for different G4 structures are correlated to G4 stability. Interestingly, the results indicate no clear correlation between the thermal stability of G4s and the helicase ability to unwind them, suggesting that the specificity of BLM or WRN towards certain G4s is not simply due to their relative stability. This is reinforced by the fact that the two helicases show different efficacies and substrate specificities. Moreover, if the differences in activity were simply due to differences in G4 stability rather than in protein selectivity, we should have seen for both helicases the same trend.

It is well-established that cations, and in particular potassium, play a pivotal role in G4 stability. Studies demonstrated that conditions aimed to increase the thermal stability of a G4 substrate such as a higher potassium concentration, decrease the unwinding ability of the enzyme (Mendoza et al., 2015). Similarly to what was performed in the above-cited study, we wanted to observe if an enrichment with KCl in the annealing buffer would influence the general level of unwinding of G4s by BLM and WRN helicases. To this aim, fluorescent-based helicase assay was a useful tool since it allowed us to monitor at the same time the protein activity towards several G4s, testing various conditions. We modulated the salt concentration in the G4 annealing buffer used by increasing the potassium from 1 mM KCl with 99 mM NaCl to 100 mM KCl, leaving unchanged all the other parameters.

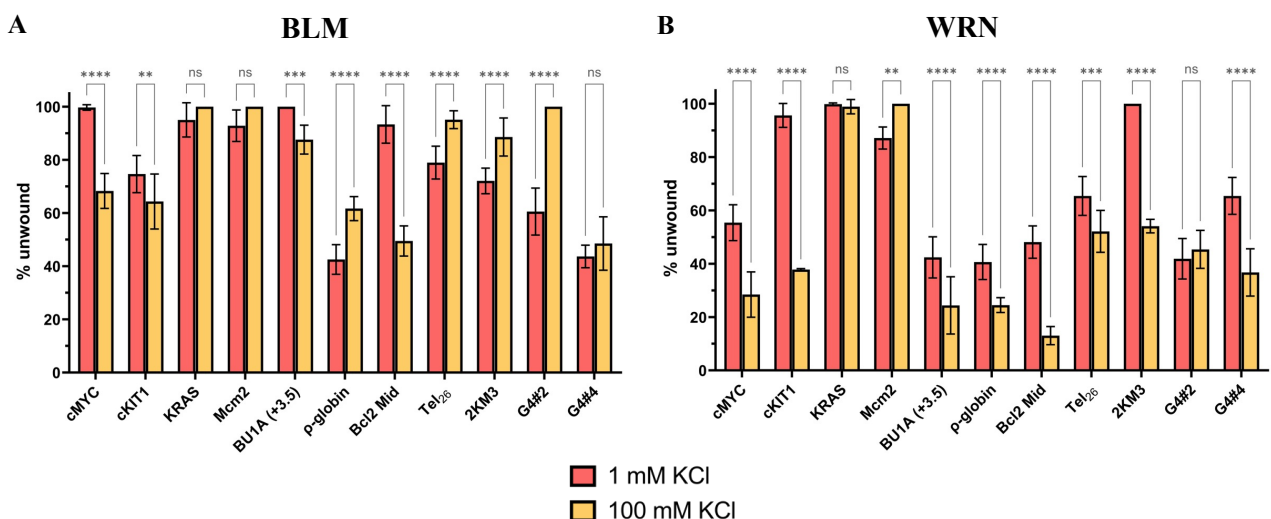


Figure 41. BLM and WRN unwinding activity towards various G-quadruplexes in two different annealing buffers. Comparison of **A. BLM** and **B. WRN** unwinding activity in annealing buffer containing 1 mM KCl with 99 mM NaCl (red bars) or 100 mM KCl (yellow bars). The reaction buffer contains 50 mM KCl. Error bars indicate the standard deviation between at least three independent experiments. Statistical

significance has been reported (****, *** and ** represent statistical significance at $P < 0.0001$, $P < 0.001$ and $P < 0.01$ respectively. ns represents statistically non-significant difference ($P > 0.05$); data analyzed with GraphPad Prism, using Two-Way ANOVA, comparing each cell mean with the other cell mean in that row).

We assessed BLM and WRN helicase activity towards all the G4s tested previously by increasing potassium concentration to 100 mM in the annealing buffer. Both BLM and WRN are active in the presence of either Na^+ or K^+ ; the optimal concentration of monovalent cations is 50 mM (which is the concentration used in the reaction buffer), but they are still highly active at higher ionic strength (100 mM KCl). In particular, our results show that the amount of G4 unwound by BLM is generally not or just slightly impacted whatever the saline concentration, and the amount of system unwound in solution is almost similar in both conditions for each system (Figure 41A). On the contrary when the K^+ concentration is increased the resolution of G4s by WRN is significantly decreased (typically about 50%). It is interesting to notice that this is not happening with the KRAS and Mcm2 G4s (Figure 41B).

These results suggest that BLM unwinding activity towards G4 DNA is mostly independent of G4 thermal stability and the nature of the monovalent cation at this ionic strength, whilst in most cases differences in WRN activity were observed as a result of the difference in stability of the G4 substrate to be unfolded.

15.4. RecQ1, RecQ4 and RecQ5 do not show great interaction with G-quadruplexes

To date, there is limited evidence supporting interactions between human RecQ1, RecQ4, or RecQ5 helicases and G-quadruplexes. However, once again, most of these studies used one or two G4s as an example. Therefore, we conducted a systematic analysis of various G4 topologies and sequences, to verify whether we could identify any G-quadruplex that could serve as a substrate for these RecQ helicases.

An extensive binding analysis carried out using Fluorescence Anisotropy confirmed that despite the large number of G4s tested, all three RecQ helicases show very low affinity towards all the substrates, suggesting G-quadruplexes are not their preferred targets.

RecQ1 is not able to bind to any G4 (Figure 42A), which could be due to its inability to interact with G4 structures or to the absence of a 3' tail. RecQ1 has been shown to efficiently binds tetramolecular G4s with either a 7- or 20-nt 3'-tail with an affinity similar to that measured for BLM protein (Popuri et al., 2008). However, they never tested the affinity on unimolecular G4s. Further studies analyzing

the interaction of RecQ1 with unimolecular G4s containing a 3'-tail could better clarify the reason for the lack of consistency between our results and those previously obtained.

RecQ4 and RecQ5 display a low but measurable binding affinity (Figure 42B and Figure 42C). There is not a particular preference for one topology over the other, probably suggesting that G-quadruplexes are not their target substrates.

It has to be stressed that the low binding capacity of RecQ4 towards G4 structures can be due to the use of RecQ4 fragment 371-1116, which misses most of the N-terminal domain and, in particular, the entire RecQ4-specific motif (RSM, aa 348-388) which has been suggested to have an important role in G4 interaction (Papageorgiou et al., 2023). Preliminary analyses on the interaction between the N-terminal domain of RecQ4 and G4s have been performed, and initial results are reported in the next section (section 15.5).

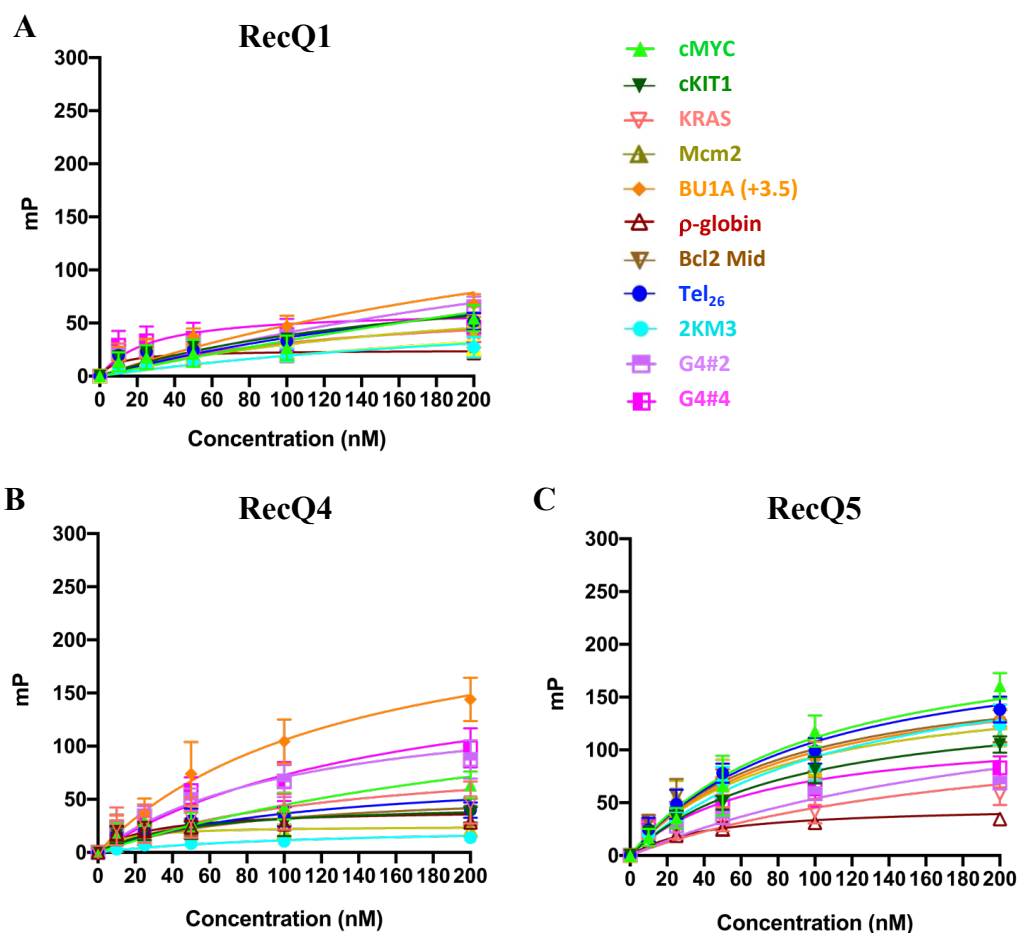


Figure 42. RecQ1, RecQ4 and RecQ5 binding to various G-quadruplexes. Comparison of A. RecQ1, B. RecQ4 and C. RecQ5 binding affinity towards different G-quadruplexes, using Fluorescence polarization technique. One-site total binding function of GraphPad-Prism was used to fit the data points. Error bars indicate mean deviation between three independent experiments.

Since there is not always a direct correlation between the binding affinity to G-quadruplexes and the ability of the helicase to unwind the same substrates, we studied the helicase activity of RecQ1, RecQ4, and RecQ5 towards the panel of G4s tested before.

Our results showed that, while BLM and WRN efficiently unwind many G-quadruplexes, with different efficiencies towards different G4s, in the same conditions and protein concentration, RecQ1 has very low helicase activity while RecQ4 and RecQ5 display no detectable unwinding activity.

Notably, although RecQ1 is not able to bind to unimolecular G4s, it shows a slight helicase activity, even though negligible. These results are in line with previous reports indicating that the helicase is unable to resolve any G4 structure, either with different 3'-tail lengths or different sequences and topologies (Popuri et al., 2008).

Although the control experiments shown in section 15.1 have demonstrated that RecQ4 and RecQ5 are able to unwind DNA Forks (Figure 36), these helicases are not able to resolve G4 structures (Figure 43B and C). RecQ4 is able to perturb the chemical equilibrium of G-rich regions, thus opening the G4 structure, even in the absence of ATP (Li et al., 2023). The experimental conditions used in that work raised some doubts about the results obtained: the incubation time of 2h at 4°C and the use of 150 mM NaCl in the reaction buffer could facilitate the resolving activity of the protein, being the G4 structures less stable in these conditions. Our results are mostly consistent with other studies.

The fact that the protein binds but does not unwind G4 suggests the possibility of RecQ4 having a role in protecting G4 structure from other helicases rather than in resolving it, or acting as a recruiter of other important factors to the G-rich regions.

Our findings on the inability of RecQ5 to unfold G4 structures reflect previous reports showing that the helicase is proficient in unwinding at low protein concentration (50 nM) the human telomeric G4 only in the presence of 50 mM NaCl, or in unfolding the TBA G4, which is significant less stable, even at higher KCl concentrations. In fact, according to Budhatoki et al. (2016), RecQ5 unfolding activity resulted to be much weaker compared to the one of BLM and WRN, and showed a strong dependence on G4 stability, suggesting that this might not be its primary function. This could explain why, in the experimental conditions used in our study (6 nM protein concentration; reaction buffer with 50 mM KCl), RecQ5 did not show any unwinding activity (Figure 43C). Higher enzyme concentration could be needed to detect some helicase activity. However, we decided to set and use the same protein concentration for all the RecQ helicases to better compare the G4 unwinding efficacy among them.

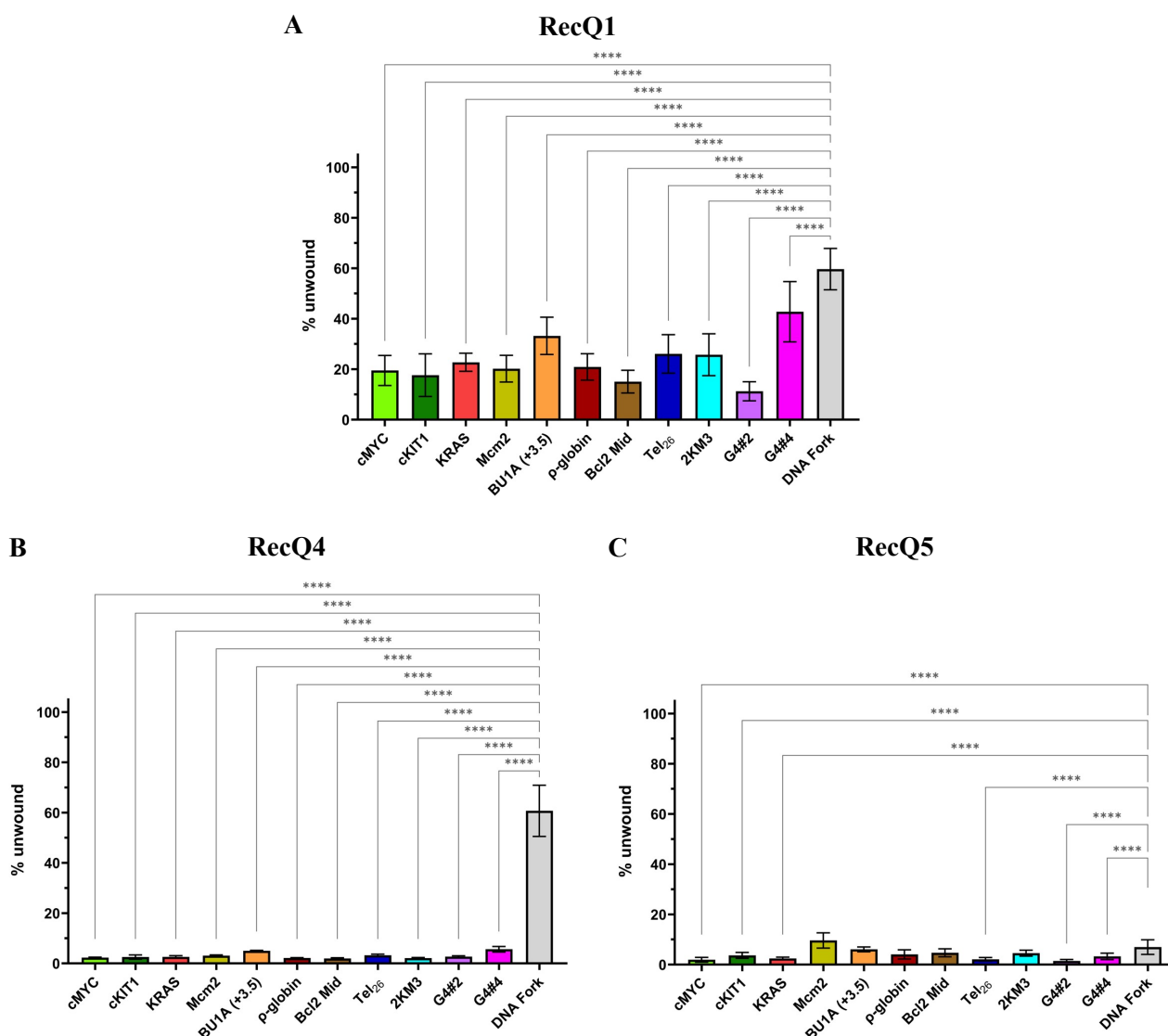


Figure 43. RecQ1, RecQ4 and RecQ5 unwinding activity towards various G-quadruplexes. Quantification of A. RecQ1, B. RecQ4 and C. RecQ5 helicase activity against selected G4 DNA systems; error bars indicate standard deviation of three independent experiments. DNA Fork has been used as comparison. Most appropriate statistical significance has been reported (**** represents statistical significance at $P < 0.0001$; data analyzed with GraphPad Prism, using Ordinary One-Way ANOVA test, comparing the mean of each column with the mean of the DNA Fork column, chosen as control). Helicase assays were performed in annealing buffer containing 1 mM KCl and 99 mM NaCl, and in reaction buffer with 50 mM KCl.

15.5. The N-terminal domains of RecQ4 display selective affinity for G-quadruplex structures

Our results showed that the catalytic domain of RecQ4 (fragment 371-1116) alone is not able to interact with G4 structures, in line with the reported literature. A couple of studies revealed the

importance of the N-terminal region of RecQ4, which is homologous to the yeast replication initiation factor Sld2, in binding G-quadruplexes (Keller et al., 2014, Papageorgiou et al., 2023).

For this reason and due to the absence of a systematic analysis of the N-terminal region with multiple G4 sequences, we decided to produce various N-terminal domains in order to dissect their G4 specificities.

Due to the limited time and the time-consuming process of expressing and purifying lots of fragments, we focused our attention only on RecQ4 fragments 1-55, 1-133, and 1-427 (Figure 27), which showed better levels of expression in small-scale expression tests. We carried out an extensive biochemical characterization, studying the binding affinity of these three fragments towards the same panel of G4s used for RecQ helicases analysis. To assess the binding we used gel-based electrophoretic mobility shift assay (EMSA) that permit to directly visualize not only binding, but the behaviour of the protein:substrate complex. For example the protein:nucleic acid complex can remain in the well and not being able to migrate through the gel matrix due to protein aggregation; or it is possible to see multiple band shifts on the gel suggesting the presence of complexes with multiple stoichiometries, providing a clear advantage with respect to Fluorescence Anisotropy, especially for partially disordered proteins that may be more prone to aggregate. However, disadvantages of gel-based assays are that they are generally more difficult to quantify and tend to underestimate binding at low concentrations.

We started the analysis on 1-427 fragment (encompassing the whole N-terminal domain) using EMSAs.

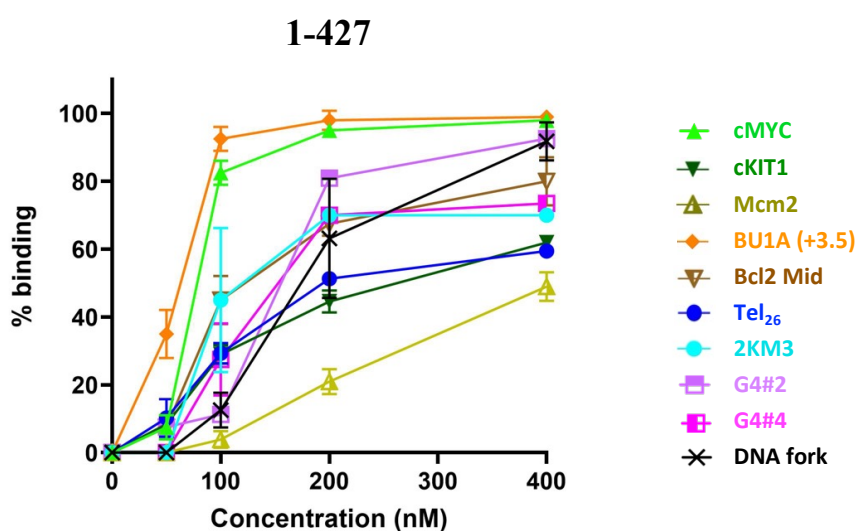


Figure 44. RecQ4 1-427 binding to various G-quadruplexes. Comparison of RecQ4 1-427 binding affinity towards different of G4 substrates, using fluorescence polarization technique. DNA Fork have been used as

comparison. GraphPad-Prism was used to connect the data points. Error bars indicate mean deviation between three independent experiments.

The RecQ4 1-427 fragment showed a high affinity for many G4 structures, in some cases even higher compared with the DNA Fork (Figure 44). Promoter G4s (cMYC, BU1A (+3.5) and Bcl2 Mid G4) are generally bound better than telomeric G4s (Tel₂₆, and (CTAGGG)_n G4). Focusing on lower concentrations (< 100 nM), we observed a significant affinity to all the G4s, higher compared to the control DNA Fork, except for Mcm2. These results confirm the preference of the N-terminal region of RecQ4 for G4 structures, in line with the reported literature (Keller et al., 2014), but extend the previous results by analyzing a large panel of G4s and detecting a degree of specificity.

It is not possible to perform a direct comparison of binding affinities between 1-427 and 371-1116 fragments due to the use of two different techniques. In any case, the high affinity demonstrated by this N-terminal domain demonstrate that this region is indeed crucial for RecQ4 to interact with G4 structures. Moreover, the ability of the N-terminal domain to bind to G4 DNA suggests that it may be involved in recruiting RecQ4 to replication origin sites or in regulating the structure and function of G4 DNA at these sites.

We then started dissecting the G4-binding determinants of the N-terminal region of RecQ4, by analyzing the G4 interaction proprieties of 1-55 and 1-133 fragments. It was not possible to complete the analysis with all the substrates due to time constrain, in particular for 1-133, but the preliminary results provided some significant information. In this case, we used Fluorescence Anisotropy technique to evaluate the binding affinity of both fragments with 5 nM of various G4 substrates. We first tried to perform the analysis using the gel-based assay, but we faced some troubles with the EMSA gels. It was not always possible to calculate a percentage of binding since in many cases the protein remained stuck in the well or no shift was evident, while at other times a smear was visible throughout the protein concentrations tested.

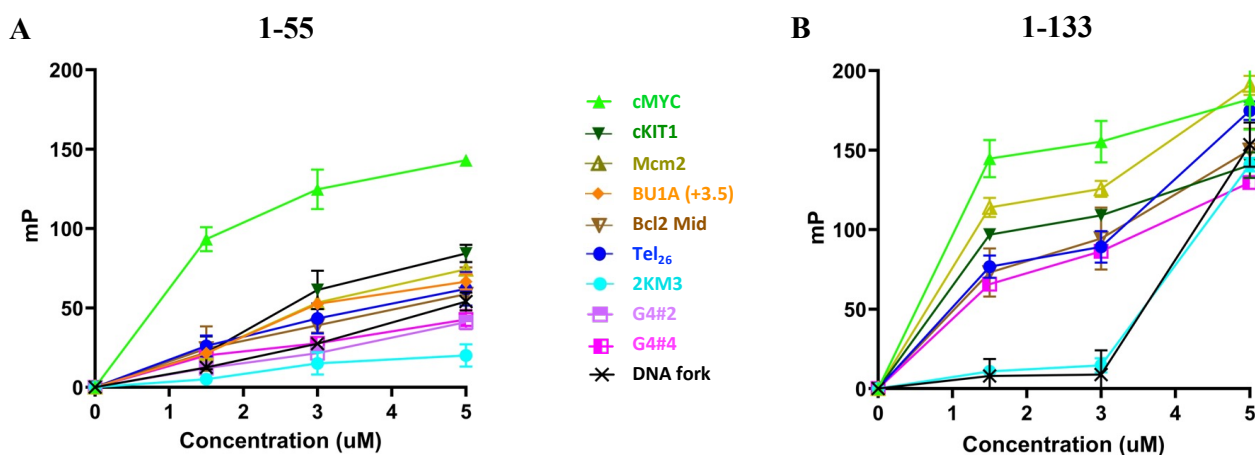


Figure 45. Comparison of A. RecQ4 1-55 and B. RecQ4 1-133 binding affinity towards a variety of G4 substrates. DNA Fork have been used as comparison. GraphPad-Prism was used to connect the data points.

Both 1-55 and 1-133 RecQ4 constructs displayed some G4 binding affinity, although a 100-fold higher concentration compared to 1-427 was required to saturate binding to the quadruplex. Interestingly, 1-55 fragment displayed a net preference for one single G4 structure, the cMYC G4 (Figure 45A), which stands out from the other G4s, indicating a possible preference of this fragment for this specific structure or, more in general, for promoter G4s. Similarly, 1-133 retained the preference for the same G4 and particularly for promoter G4s, as indicated by the high fluorescence values obtained with cMYC, Mcm2, cKIT1 and Bcl2 Mid G4s (Figure 45B).

Furthermore, both fragments showed considerably higher affinity for G4s compared to the control DNA Fork (Figure 45), demonstrating their preference for these non-canonical substrates. The elevated protein concentration needed in the assay to detect some binding signal indicated that these first 133 amino acids contribute to the overall RecQ4 G4 binding but are not important key residues. In fact, according to previous studies, another N-terminal region of the protein (the RSM motif, aa 348-388) is decisive and sufficient for G4 DNA binding properties (Papageorgiou et al., 2023).

15.6. RecQ4 full-length binds with high affinity G-quadruplex structures

Our results showed that the N-terminal domain 1-427 of RecQ4 binds with high affinity many G-quadruplexes, while the RecQ4 fragment 371-1116, lacking most of the Sld2-like domain, displays very low binding. We performed a preliminary biochemical analysis on the full-length RecQ4, which we routinely express and purify in our laboratory, to test its binding affinity towards different G4 sequences. For this purpose, Fluorescence Anisotropy technique was used, and increasing concentrations of RecQ4 full-length (0-200 nM) were incubated with various G4 structures.

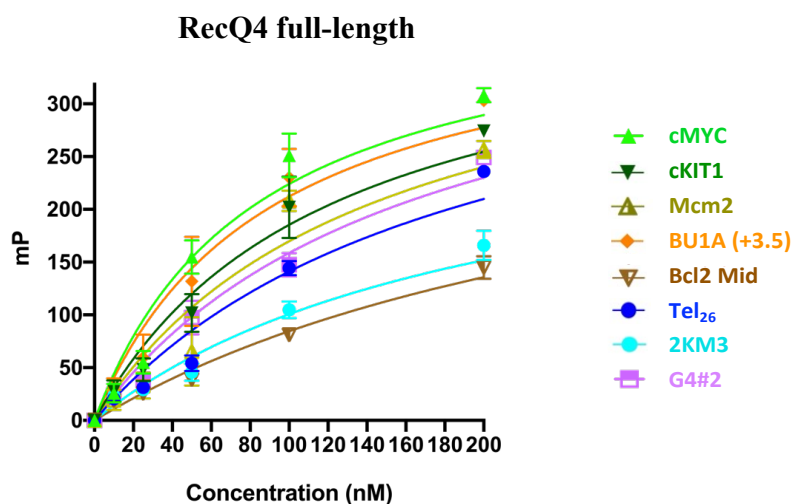


Figure 46. RecQ4 full-length binds to various G-quadruplexes. Comparison of RecQ4 full-length binding affinity towards different G-quadruplexes, using Fluorescence polarization technique. One-site total binding function of GraphPad-Prism was used to fit the data points. Error bars indicate mean deviation between three independent experiments.

The results clearly showed that RecQ4 full-length is able to bind with high affinity many G-quadruplexes, suggesting that the addition of the N-terminal domain to the catalytic core plays a key role in the interaction with G4 structures. Analyzing substrate preference, RecQ4 full-length displays a higher affinity for G4 sequences from promoter regions (cMYC, cKIT1, BU1A (+3.5), Mcm2) compared to telomeric G4s (Tel₂₆ in blue and 2KM3 in cyan) (Figure 46). These findings corroborate the idea that RecQ4 might play a role in the initiation of DNA replication interacting with G4 structures (Papageorgiou et al., 2023). A complete biochemical analysis is still in progress, studying the unwinding activity of the full-length RecQ4 towards these non-canonical substrates.

Anyway, considering the high binding affinity of the N-terminal region of RecQ4 and the full-length RecQ4 protein to G-quadruplexes, in particular promoter G4s, along with the inability of RecQ4 371-1116 fragment to resolve G4 structures; considering the evolutionary fusion event of RecQ4 protein with the yeast Sld2-like domain, which play an essential role in the initiation of DNA replication, along with its localization to replication origins and its inhibition of G4 unwinding (Papageorgiou et al., 2023); these findings suggest that RecQ4 may play a signaling role in regulating G4 accessibility to other helicases and coordinating DNA replication initiation by interacting with these structures. Further studies are needed to elucidate the in vivo mechanisms underlying this process.

16. A putative G4-binding pocket in human RecQ helicases

The crystal structure of a bacterial RecQ helicase shows how these enzymes may bind to a resolved G4 structure, revealing how the guanine can make specific interactions (Voter et al., 2013). The identification of a guanine-specific binding pocket (GSP) in bacterial RecQ, which is uniquely selective for G-quadruplex structures, raised the question of the possible presence of the same pocket also in human RecQ helicases.

The main residues of bacterial RecQ involved in G4 interactions are Ser245 and Asp312, which create uniquely selective contacts with the guanine base of the DNA chain (Figure 47). Hydrogen bonds are formed by the sidechain hydroxyl and backbone amide of the Serine and the sidechain carboxyl group from the Aspartate. As previously discussed in the introduction (section 3.3.2), this pocket was suggested to play an essential role in G4 resolution. Voter et al. (2018) demonstrated that when compromising the GSP by mutating one of the two residues from Serine/Aspartate to Alanine (Ser245Ala or Asp312Ala), the protein failed to unwind G4 DNA but maintained the ability to unwind duplex DNA, as well as the binding to both substrates.

Thanks to the availability of crystal structures of the catalytic core of human RecQ helicases, we carried out a structural comparison between the latter and the bacterial RecQ. This analysis aimed to reveal the putative presence of a G4-binding pocket, identifying the residues involved in the interaction between G4 and human RecQs.

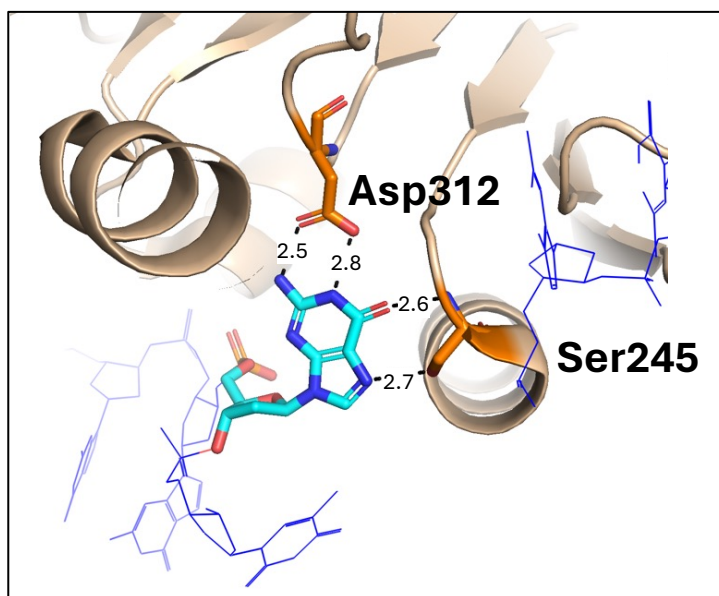


Figure 47. The guanine specific pocket of the CsRecQ helicase (PDB ID: 6CRM). The helicase backbone is shown in beige, with the DNA shown in blue. Residues Ser245 and Asp312 (with their side chains highlighted in orange) are involved in the interaction with the guanine (cyan) of G4 DNA. Hydrogen bonds are represented by dashed lines and bond distances in Å are shown.

16.1. Structural comparison

To investigate the potential presence of a guanine-binding pocket in human RecQ helicases, we conducted a structural analysis to identify residues potentially involved in interactions with the guanine base of G4 DNA. Using PyMOL, we superimposed the bacterial and human RecQ helicase structures. The overlap revealed a high degree of structural conservation in the region surrounding the pocket. Starting from the bacterial residues forming the pocket, we identified homologous or functionally similar residues in human RecQ helicases, positioned similarly or oriented in a suitable way to form hydrogen bonds with the guanine base of the G4 DNA.

Analysis of the [BLM helicase](#) structure revealed a potential GSP formed by Ser897 and Ser965, which overlap perfectly with bacterial Ser245 and Asp312, respectively (Figure 48).

Similarly, in [WRN helicase](#), Ser773 and Asp872 were identified as potential GSP residues, with Ser773 occupying the same position as bacterial Ser245, while Asp872 comes from an adjacent secondary structure element. In the WRN crystal structure Asp872 is currently not in a position that can engage with the guanine, but it is plausible that the presence of a G-rich nucleotide sequence could induce a small conformational shift, bringing Asp872 closer to the guanine base and allowing hydrogen bond formation (Figure 48).

In [RecQ5](#), a similar GSP can be found, involving Thr268, which aligns with bacterial Ser245, along with Asn335 and Asp367. While Asn335 mirrors the orientation of bacterial Asp312, Asp367 comes from an adjacent helix, with a similar position to the aspartate found in WRN (Figure 48).

In [RecQ1](#), Ser323 and Ser390 occupy the same positions and orientations as bacterial Ser245 and Asp312, respectively, suggesting they may have a role in the formation of the putative GSP (Figure 48).

On the contrary, the structure of [RecQ4](#) did not reveal a clear GSP. Only one residue, Arg709, aligned with bacterial Ser245, suggesting its possible interaction with the G4 DNA. Although no other residues evidently form a guanine-binding pocket, we hypothesize that conformational changes might bring residues such as Asn708, Glu823, and/or Asp824 into a favorable position to contribute to the pocket by forming contacts with the G4 structure (Figure 48).

The important role of a residue in contributing to the structure and function of a protein can usually be inferred from how conserved it appears in a multiple-sequence alignment of that protein and its homologs. The conservation of these key residues among species can support their potential role in forming the GSP, reinforcing the hypothesis that they might play a crucial role in the helicase function (as further discussed in section 16.4). Nonetheless, a co-crystal structure of human RecQ helicases bound to G4 DNA is necessary to confirm these predictions.

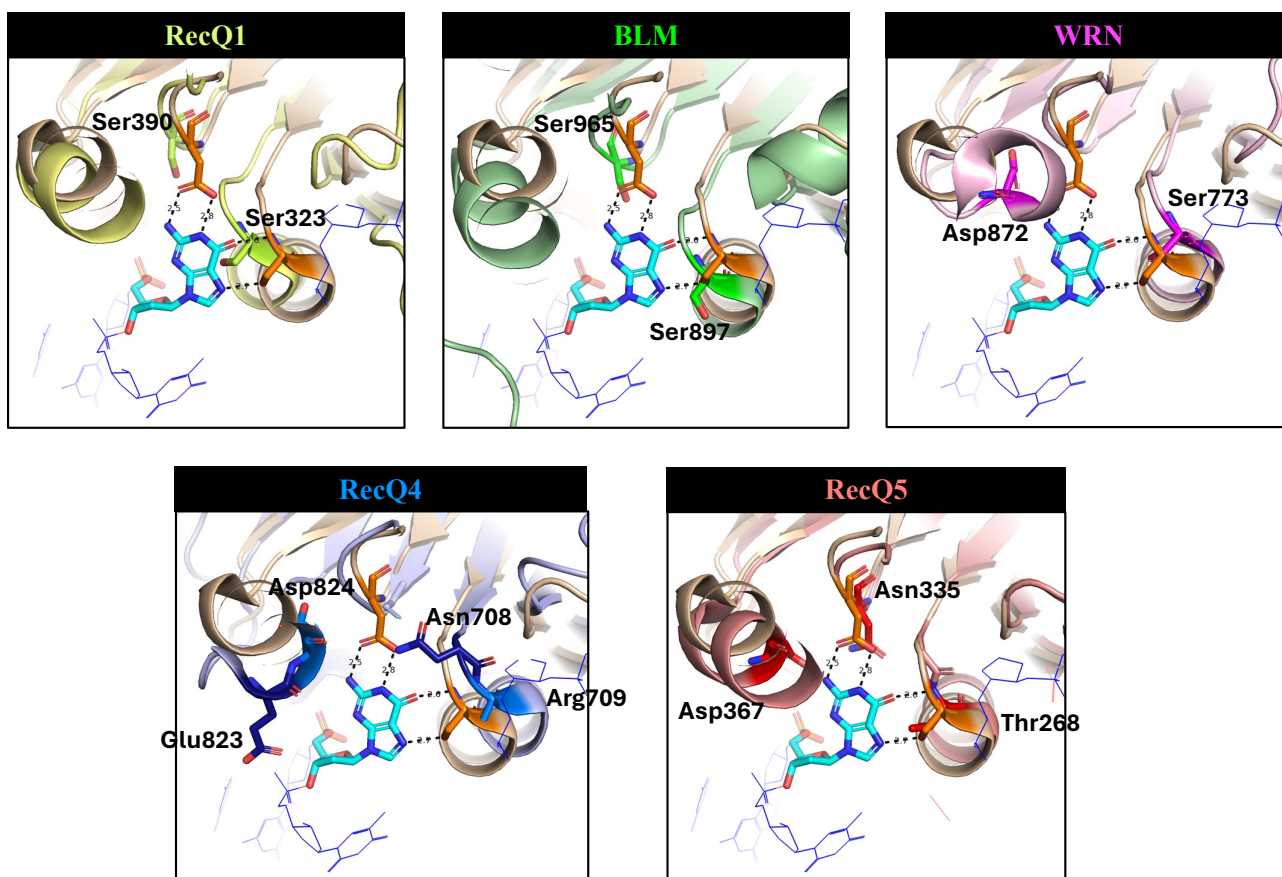


Figure 48. Superimposition of bacterial RecQ and human RecQ helicases structures. The bacterial RecQ (PDB ID: 6CRM) is shown in beige with the G-binding residues in orange, as in Figure 47. The analysis revealed the presence of putative key residues potentially involved in the interaction with the guanine base of the G4 DNA. Bacterial RecQ is in wheat and residues (Ser245 and Asp312) in orange; hydrogen bonds between CsRecQ and G4 are represented by dashed lines; RecQ1 (PDB ID: 2V1X) is in pale yellow and residues (Ser390 and Ser323) in lime; BLM (PDB ID: 4CGZ) is in pale green and residues (Ser965 and Ser897) in green; WRN (PDB ID: 6YHR) is in light pink and residues (Ser773 and Asp872) in magenta; RecQ4 (PDB ID: 5LST) is in light blue and residues (Arg709 and Asp824, Asn708 and Glu823) in two different shades of blue respectively; RecQ5 (PDB ID: 5LB8) is in salmon and residues (Thr268, Asn335 and Asp367) in red.

16.2. The putative residues are conserved among organisms

Sequence alignments were performed on RecQ orthologs (RecQ1, BLM, WRN, RecQ4 and RecQ5) from multiple organisms, to investigate the conservation of predicted residues across different species. Using the protein sequences of human RecQ helicases as input, and in particular the truncated sequences corresponding to the RecQ fragments used in this study, we compared them to the sequences of selected model organisms: *Drosophila melanogaster*, *Danio rerio* (zebrafish), *Caenorhabditis elegans*, *Xenopus laevis*, *Aplysia californica*, and *Hydra vulgaris*.

Initially, the BLAST tool (NCBI) (Altschul et al., 1997) was used to identify the presence of RecQ helicases in these organisms by finding regions of local similarity between protein sequences and generating significant alignments. The coverage percentage between human and model organism sequences was consistently >90%, indicating that a large portion of the organism RecQ sequence aligned with the human version. In many cases, the percentage identity between the human and selected organism RecQ sequences was also high, further demonstrating strong conservation across species. Interestingly, it was not possible to find some RecQ orthologues in some species: RecQ1 and WRN are missing in *Drosophila*, RecQ4 is absent in *C. elegans* and RecQ5 can't be found in *X. laevis*.

After obtaining the protein sequences of RecQ helicases from these organisms, we conducted a multiple sequence alignment using Clustal Omega (Sievers et al., 2012) to examine the conservation of specific residues. Our focus was on residues previously identified as potentially involved in forming the G-quadruplex binding site (GSP) in the five human RecQ helicases. The analysis revealed a high degree of conservation for most of the residues among RecQ orthologs, suggesting their potential importance in GSP formation and G4 interaction.

The identified residues in **RecQ1** (Ser390 and Ser323), **BLM** (Ser965 and Ser897), and **WRN** (Ser773 and Asp872) are highly conserved across species, remaining identical in nearly all organisms. In **BLM**, the only exception is *Hydra vulgaris*, where the two serines are replaced by alanine and threonine (Figure 49B). In **WRN**, Ser773 shows conserved substitutions: while *Danio rerio* retains the serine, other organisms feature threonine (Figure 49C). Since both serine and threonine are polar, amino acids having a hydroxyl as functional group, these substitutions are unlikely to disrupt the interaction with G4 DNA.

For human **RecQ4**, multiple residues were identified as potentially involved in GSP formation, including Arg709, Asp824, Asn708, and Glu823. Of these, Arg709 and Asp824 are identical in all aligned sequences, suggesting they could be key residues for guanine interaction. Asn708 shows a conserved substitution to threonine in all other species considered, which may still support its involvement in pocket formation. Conversely, Glu823 is not conserved in any species and can likely be excluded from the analysis (Figure 49D).

In human **RecQ5**, Thr268 and Asp367 exhibit conserved substitutions: threonine changes to serine in *C. elegans*, and aspartate to glutamate in *Aplysia californica*, reinforcing their potential role in GSP formation. Asn335, however, is not conserved across species indicating this is not an important residue (Figure 49E).

Overall the structural and sequence analysis suggested the possibility that the mechanism of G4 unwinding proposed for the *Cronobacter sakazakii* RecQ (Voter et al., 2028) may be applicable to all

RecQ helicases; the putative guanine specific pocket could potentially be present (with some variation) in all the five human helicases, and the residues involved (or a subset of the possible candidates) show a degree of conservation among species.

However, the conservation of the putative pocket does not appear to discriminate between the RecQ helicases that seem to be more or less active towards G4 unwinding. It is difficult to speculate why these residues are actually conserved in RecQ1, RecQ4 and RecQ5. The initial recognition of a folded G-quadruplex is likely to involve another region of the helicase and is a pre-requisite for the following unwinding. As in human cells the RecQ helicases have diverged to assume specialized function, it is likely that only some of these have maintained the full G4-unwinding function, whereas the others have primarily lost the initial site of interaction. The residues of the putative guanine-binding pockets are located at the interface between the helicase core and the RQC domain and they may be therefore also involved in the folding and stability of the helicase. Nonetheless, it has to be stressed that we are working with truncated forms of these helicases, and we cannot completely rule out the possibility that in the context of the full-length proteins, or in the presence of specific molecular partners, some G4 unwinding activity could be observed, explaining the conservation of a guanine-binding pocket.

A. RecQ1

RecQ1_HUMAN	YEVVRQKPSNTEDFIEDIVKLINGRYKGQSGIYCFSSQKDSEQVTVSLQNLGIHAGAYHAN	348
RecQ1_Danio rerio	YEVRFKD--NEDCTDQIASLIRGRYKNQSGIYVVFSSQKDAEVVATELQKRDIVAQPYHAN	352
RecQ1_C. elegans	YKVVQKPGSEDECTEEIAKTIKRFAGQTGIYICLSRNDCEKVAKALKSHGIAKAKHYHAY	343
RecQ1_Xenopus laevis	YEVRLKPSSSEDFIADITKLINSRYKGQSGIYCFSSQKDSEQVTMSLQKLGIKAGAYHAN	350
RecQ1_Aplysia californica	YQVLKPKSSQKETMEKILLISTFRFGMSGIYVCFSSRKDSEEVTSDLMSHGIAKAGCYHGD	338
RecQ1_Hydra vulgaris	YEVKHTLLSPSELVNEIATCISRFKEESGIYVCFSSRKDSEEVSTLLNRHGINSHCYHAD	344
	: . .: .* * : :***:*:*:*:* * . * : **	
RecQ1_HUMAN	LEPEDKTTVHRKWSANEIQVVVATVAFGMGIDKPDVRFVIHHSMSKSMENYYQESGRAGR	408
RecQ1_Danio rerio	MEPSHKSLVHQRWSSKKIQVVVATVAFGMGIDKADVRFVIHHSISKSIENYYQESGRAGR	412
RecQ1_C. elegans	MEPVDRSGAHQGWISGKIQVIVATVAFGMGIDKPNVRFVIHHSLPKSIENYYQESGRAGR	403
RecQ1_Xenopus laevis	MEPRDKSKVHTKWTANEIQIVVATVAFGMGIDKPDVRFVIHHSMSKSMENYYQESGRAGR	410
RecQ1_Aplysia californica	IDAKRSRSVRHMLWLSGEIQVVVATVAFGMGIDKPDVRFVIHHSISKSMENLYQESGRAGR	398
RecQ1_Hydra vulgaris	ISSDVKTQVHQLWIKGDIQVIVATVAFGMGIDKPNVRFVIHHSLSKSIENYYQESGRAGR	404
	:. :. :* * .**::***:***** :*****:: **:* *****	

B. BLM

BLM_HUMAN	KKPKKVAFDLCLEWIRKHHPPYDSGIYICLSRRECDTMDTLQRDGLAALAYHAGLSDSARD	928
BLM_Drosophila	KKGVSTLDDISRYIRSKPHQFSGIYICLSRKECDETSKMKCKDGVRAVSYHAGLTDTDRE	999
BLM_Danio rerio	KKPKKVDEECIQWIKKYYPRDSGIVYICLSRNDCTDLADSLQAGIAALAYHAGLSDSDRE	951
BLM_C. elegans	KAA-RSLINVVEKMKQLYPGKSGIYICLSRKECETVQMLTKAGLSAEVYHAGLNDNLRV	510
BLM_Xenopus laevis	KKPKRVALDCVEWIKKHHPNDSGIYICLSRHECDTMDTLQKEGLAALAYHAGLADSNRD	885
BLM_Aplysia californica	KKCSNATTDVIQMIQSRFRNDCGIVYCFSSRKECDNVAMDLRKAGLLANSYHAGLDDNKRA	939
BLM_Hydra vulgaris	KS-KNTLNDMVNKKIKTEFDNKGSIYICLSRRDCDYVANYLISSGITAAAYHAGLNDNERS	944
	* : . :. :**:*:*:*:*: : * : * ***** *	
BLM_HUMAN	EVQQKWINQDGCQVICATIAFGMGIDKPDVRFVIHHSLPKSVEGYQESGRAGRDEISH	988
BLM_Drosophila	SRQKDWLTG-KMRVICATVAFGMGIDKPDVRFVLIHSLPKSIEGYQESGRAGRDEIS	1058
BLM_Danio rerio	YVQNKWINQDGCQVMCATIAFGMGIDKPDVRFVIHSLPKSVEGYQESGRAGRDEISH	1010
BLM_C. elegans	SVQRSWIAN-KFDVICATIAFGMGIDKPDVRFVIHSLPKSIEGYQETGRAGRDMPSY	569
BLM_Xenopus laevis	YVQHKWINQDGCQVICATIAFGMGIDKPDVRFVIHSLPKSVEGYQESGRAGRDEIS	945
BLM_Aplysia californica	TVQERWLNEDGCKICATIAFGMGIDKADVRFVIHSLPKSMEGYQESGRAGRDMIAH	999
BLM_Hydra vulgaris	NVHESWLKN-KFKVVCATIAFGMGIDKADVRFVFIHSLPKSVEGYQESGRAGRDLKSV	1003
	:. *: :***:***** **:*:* :***:***:***** :	

C. WRN

WRN_HUMAN	QPFLVK-T--SSHWEFEGPTIIYCP S RKMTQQVTGELRKLNLSGTYHAGMSFSTRKDIH	808
WRN_Danio rerio	KRFLVK-KK-GGGYEFEGSAIVYCP S KKEAERVTTALFKLDIPCGVYHAGLSIKQRRETQ	304
WRN_C. elegans	GLFMKTDE--VKGRHFGGPTIIYCP T KQMVDVNCVLRRIQVRSAHYHAGLTKNQREKAH	490
WRN_Xenopus laevis	PQFLIK-KQQGSGWEFEGATIVYCP T RTKTSQVTAELIKLGIACGTYHAGMIKQRRREVH	751
WRN_Aplysia californica	NDFRGELEKKGLRYDFSGPTLIYCP T KKATGEVASVLQAMGVSCLSYHAAMSLSTRKEAH	512
WRN_Hydra vulgaris	SLMIEENVNGVRKYRFCGTTIVYCP T KKKVEEIANTLIGFGVTCEPYHAGLTPQRKKTH	487
	: * * : : * : : : * : : . * * : : * . : :	
WRN_HUMAN	HRFVRDEIQCVIATIAFGMGINKADIRQVIHYGAPKDMESYYQEIGRAGRDLQSSCHVL	868
WRN_Danio rerio	HQFMRDEIQCVVATVAFMGINKSDIRKVIHYGAPKEMESYYQEIGRAGRDLPSACHVL	364
WRN_C. elegans	TDFMRDKITIVATVAFMGIDKDPVRNVIHYGCPNNIESYYQEIGRAGRDLSPSICRVF	550
WRN_Xenopus laevis	HRFMRDEIQCVVATVAFMGINKADIRKVIHYGAPKEMESYYQEIGRAGRDLPSCHAL	811
WRN_Aplysia californica	HKFLNDRVQAVVATIAFGMGIDKDPVRKVIHYGAPKDIESYYQEIGRGRDQGPSFCTVL	572
WRN_Hydra vulgaris	NKFIRDELDCIVATVAFMGIDKDPDIRMVIHYGAPRDIESYYQEIGRAGRDLQSPFCYAF	547
	* . : . : : * : * * * : * * : * * * * * : * * : :	
WRN_HUMAN	WAPADINLNRHLLTEIRNE-KFRLYKLKMAKMEKYLHSSRCRRQIILSHFEDKQVQKAS	927
WRN_Danio rerio	WMPGDMLNKFILNQSKSE-RFRSYKIDMMAKMEKYLNSTKCRRLILCHFEDKRLRKVT	423
WRN_C. elegans	WAPKDLNTIKFKLRNSQQKEEVNLTMLRLQELVLTTVGCRRYQLLKHFPDPSYAK---	607
WRN_Xenopus laevis	WTQADMFNFRHMLGEIPNK-GFREYKLKMLTKMEKYLNSSTCRRKIILSHFEDKQLRKAS	870
WRN_Aplysia californica	YSKADFNISRHRHISTENE-KFRDHKMKMFAKMEQYLSSTQCRRRIILSYFETGNLE---	628
WRN_Hydra vulgaris	YNCADFAINRYFLAEVDDP-KFKEYKNEMITKMEQYLMTSSCRRDAILAHFDP-SIK---	602
	: * : : : . . : * : : * : * * : * : :	

D. RecQ4

RecQ4_HUMAN	DTDQALLTLLQGKRFQNLDSIIYCN R REDTERIAALLRCTLHAAWVPGSGG-----	735
RecQ4_Drosophila	NRDAALLQLNSERFEPQSIIYCN R REDECERIAAGFIRTCVQDRREPTQD-----	1129
RecQ4_Danio rerio	DKDQALVSLKGERFGALDSVIVYCN R REETSRIAALLRCTLQGVMLRESSKHTALQEDD	937
RecQ4_Xenopus laevis	NRDQALVTLKGERFGCLDSVIVYCN R REETQRISALLRSLQGVTVSANS-HLNSQEQE	994
RecQ4_Aplysia californica	-----	0
RecQ4_Hydra vulgaris	NRDEALTELLQMKPFLSCASIIYCN R RTDADRLASMLRNLPAKHTDENA-----QNQK	710
	: * : :	
RecQ4_HUMAN	-----RAPKTTAEAYHAGMCSRRERRRVQRAFMQGLRVVATVAFGMG	778
RecQ4_Drosophila	-----QTKKRKRNVNQAEPYHAGMPASRRRTVQKAFMSNELRIVVATIAFGMG	1117
RecQ4_Danio rerio	DNPVGKKKKALARKIRKPLKWIAEAYHAGLSASERRRVQNNFMCGELRVVATVAFGMG	997
RecQ4_Xenopus laevis	DTTVSDRKAQALKKIRPLKWIAEAYHAGMSAAERRRVQNNFMCGQLRLVATVAFGMG	1054
RecQ4_Aplysia californica	-----MSGRLRVVATVAFGMG	17
RecQ4_Hydra vulgaris	DATILKG----KRKNAKKFFWWNIESYHAGLKPSQRRQVQINFMGSELRIVVATVAFGMG	766
	* . * : * * : * * *	
RecQ4_HUMAN	LDKSDVRAVLHLGLPSPFESYVQAVGRAGRDGQPAHCHLFLQPQ G EDLRELRRHVHADST	838
RecQ4_Drosophila	INKPDIRAVIHYNMPNPFESYVQEIAGRAGRDGLPSHCHLFLDAKG G QDQSELRRHYVNSI	1237
RecQ4_Danio rerio	LDKSDVRGIIHYNMPKSFESYVQEIAGRAGRDGHPAHCHLFLDPEG G DLHELRRHIYADTV	1057
RecQ4_Xenopus laevis	LDKSDVRGIIHYNMPKSFESYVQEIAGRAGRDGKDAQCHLFLDPEG G DMDELRRHIYTDV	1114
RecQ4_Aplysia californica	LDKADVRAVIHYSLARSFESYVQEIAGRAGRDGSTSHCHVFLHPQ N DACELRRHIYNSV	77
RecQ4_Hydra vulgaris	LDKSDVRAVIHYNMPSPFENYVQEIAGRAGDGKPSYVHVFLDDE G NDLYELRRHIYGNVT	826
	: : * : : * : : . * . * * : * * * * : * : * : . * * * * : : :	

E. RecQ5

RecQ5_HUMAN	ALGQEADKG-----LSGCGIVYCR R REACEQLAIELSCRGVNKAYH	291
RecQ5_Drosophila	CLGNPKFK-----DTPKPQRCGIVYCR R DQVERMAIGVTKQGIGAVAYH	288
RecQ5_Danio rerio	ALGGN--TA-----EKGCGIVYCR R ESCEEVAHKLTKLGIAPKPYH	276
RecQ5_C. elegans	CLTVDAKTSSSNLTKEKAERSQNKKTFTGSAIVYCR S RNECGQVAKMLEIAGIPAMAYH	458
RecQ5_Aplysia californica	CLGGELTDG-----ETWNRDRCGIVYCR R RDGCEEIACQLTRKGLPTQAYH	288
RecQ5_Hydra vulgaris	CLKLK--D-----NIPLKERGCGIYCR R RECHNIAEKITKKGLLAKAYH	285
	. * * . * : * * : * : : * : : * :	
RecQ5_HUMAN	AGLKASERTLVQNDWEEKVPVIVATISFGMGVDKANVRVFAH N IAKSMAGYYQESGRA	351
RecQ5_Drosophila	AGLKTGERTEVQEAWMRGDPPIICATNSFGMGVDKPSVRFVH W VPQNVAAYYQESGRA	348
RecQ5_Danio rerio	AGLKSGDRSENQADWMADKVPVIVATISFGMGVDKANVRVFAH N LAKSLASYQESGRA	336
RecQ5_C. elegans	AGLGKDRNEVQEKWMNNEIPVVAATVAFMGIDKDPVRAVH S PSQNLAGYYQESGRA	518
RecQ5_Aplysia californica	AGLKTQVREQTQTDWMEGRVPVIAATISFGMGVDKANVRVFAH T MPKSMAGYYQESGRA	348
RecQ5_Hydra vulgaris	SALTKDKKLEVQNMWDGFPVPIICATISFGMGVDKHSVRCVVH N IPQTLEGYYQESGRA	345
	: . * : : * * * : : * : * : * * : * * : : : . * * : * * :	
RecQ5_HUMAN	GRDGKPSWCRLYYSR N DRDQVSFLIRKEVAKLQEKRGNK-AS-----DKATIMAFDAL	403
RecQ5_Drosophila	GRDGLQSYCRLYYGR D VRSIRFLQNDAAHRA-GRGDKELL-----TERAIKQFEKI	400
RecQ5_Danio rerio	GRDGLPSSCRIYYSPKRDQ L NFLIRKEISRQEKRGSEKEQ-----DKAPILDFEAM	389
RecQ5_C. elegans	GRDGKRSYCRYYSK D KNALNFLVSGELAKLREKAKKNNAEKEAEMQIKSIQTGLAKM	578
RecQ5_Aplysia californica	GRDNCPSWCRLYYSH R KETVAFLIATENNPKKNLQLAKLQ-----AKAEDSFSA	401
RecQ5_Hydra vulgaris	GRDKQSFRCRIYFD K DRDQIAYLISKEVERKEEK-----RENEMLNFEFM	394
	*** * * : * . : : : : : : : : : : :	

Figure 49. Sequence analysis of RecQ helicases. Sequence alignment of RecQ helicases from various organism. A. RecQ1, B. BLM, C. WRN, D. RecQ4, and E. RecQ5. Identical residues are highlighted in

yellow, green, magenta, cyan and red for RecQ1, BLM, WRN, RecQ4 and RecQ5, respectively. Very similar residues are red highlighted in grey. Residues when not conserved are in a different color.

16.3. Production of BLM and WRN mutants

In order to test the above prediction and to better understand the role of the selected residues in G-quadruplex unwinding activity of RecQ helicases, BLM and WRN proteins were mutated in the putative guanine binding site to verify whether this generated helicases inactive towards G4 structures. We initially focused our attention only on these two RecQs since, in our hand, they are the only ones able to bind with high affinity to and unwind G4s.

Serines and aspartate residues were mutated to alanine for many reasons. Alanine is a small, non-polar amino acid with a simple side chain that provides minimal structural impact, minimizing the chance of large structural changes or destabilization of the protein upon mutagenesis. Moreover, this mutation removes the side-chain group which is the main responsible for the formation of hydrogen bonds with the DNA.

Four different single mutants were designed and cloned: BLM Ser897Ala, BLM Ser965Ala, WRN Ser773Ala and WRN Asp872Ala. Protein sequences were then analyzed to verify the successful incorporation of the desired mutations. Unfortunately, the limited time available allowed me to only test the expression and purification of BLM variants.

16.3.1. Test expression of BLM S897A and BLM S965A mutants

For each BLM mutant, we tested two different bacterial strains: BL21 Star pRare2, the same used for expressing the WT protein of BLM, and Rosetta 2. The expression temperature and IPTG concentration were kept constant at 18°C and 0.2 mM, respectively.

Preliminary results show that it is possible to produce a reasonable amount of both BLM mutants S897A and S965A in either bacterial cell strain, starting from 500 mL cell culture (Figure 50). Given these encouraging and promising results, we choose to purify BLM S897A from Rosetta 2 (Figure 50B) and BLM S965A from BL21 Star pRare2 (Figure 50C), due to slightly higher yields.

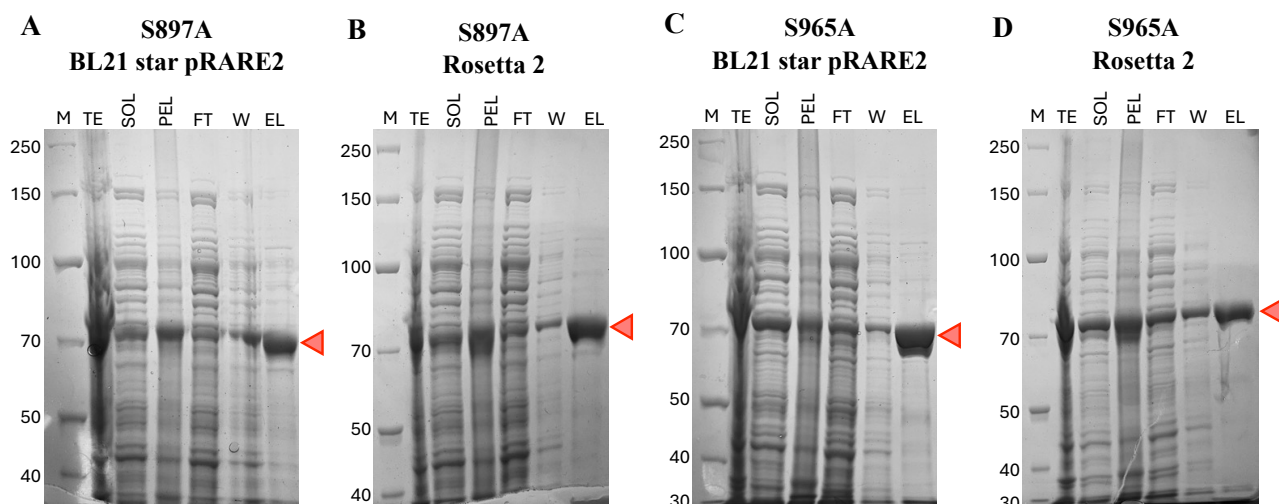


Figure 50. Medium-scale expression and purification test of BLM S897A and BLM S965A mutants. SDS-PAGE analysis of the fractions from the medium-scale expression test, for two different constructs (**A-B**. S897A and **C-D**. S965A) in (**A** and **C**) BL21 Star pRARE2 cells and in (**B** and **D**) Rosetta 2 cells. Samples from the different stages of the purification protocol are examined and labelled as follow: M: Molecular Weight standards (in KDa); TE: Total Extract; SOL: Soluble fraction; PEL: pellet representing the insoluble fraction; FT: Flow Through from Ni-NTA resin; W: wash fraction from Ni-NTA resin; EL: Eluted fraction from Ni-NTA affinity resin. A red triangle indicates the expected position of the band corresponding to the mutated protein.

16.3.2. Purification of BLM S897A and BLM S965A mutants

BLM single mutants S897A and S965A were expressed and purified as described in section 10.3 in Materials and Methods. The purification protocol was easy to set up for both proteins and it followed exactly the one used for the wild-type BLM protein. Samples from each purification step were analyzed by SDS-PAGE in Figure 51 for BLM S897A and in Figure 52 for BLM S965A.

Both BLM mutants eluted as a monomer with a molecular weight of about 76 KDa (Figure 51C and Figure 52C). From just 500 mL of cell culture, we were able to obtain a large amount of pure protein (yield of about 0.2 mg for BLM S897A and 0.5 mg for BLM S965A), which was used to perform some preliminary biochemical assays with G-quadruplexes.

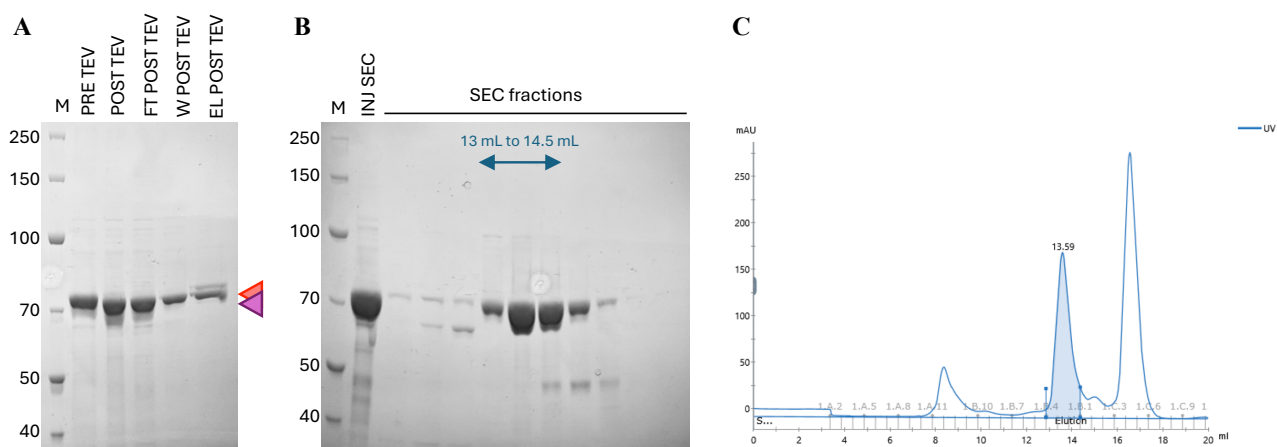


Figure 51. Purification steps of BLM S897A. **A.** SDS-page analysis of the protein sample and protein before and after TEV protease cleavage, followed by inverse Ni-affinity purification. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Labels as follow: M: Molecular Weight standards (in KDa); PRE TEV: Eluted fraction from Ni-NTA resin before tag-cleavage; POST TEV: protein after tag-cleavage; FT POST TEV: Flow Through from Ni-NTA resin post tag-cleavage; W POST TEV: wash fraction from Ni-NTA resin post tag-cleavage; EL POST TEV: eluted fraction from Ni-NTA resin post tag-cleavage. **B.** SDS-page analysis of Size Exclusion chromatography purification. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); INJ SEC: Injected sample; SEC fractions: eluted fractions from the main peak. **C.** Elution profile from a Superdex 200 column; the fractions are shaded in blue.

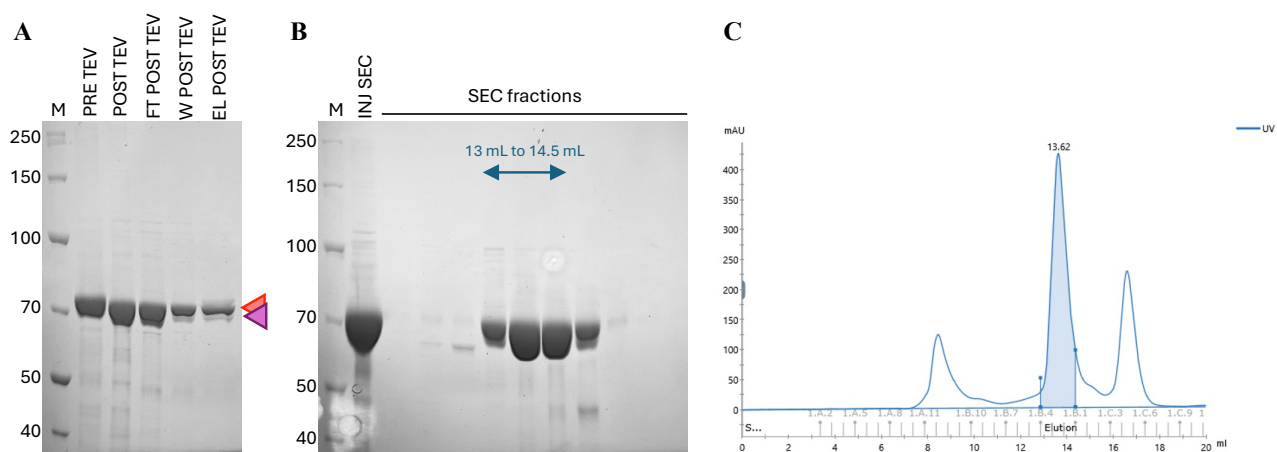


Figure 52. Purification steps of BLM S965A. **A.** SDS-page analysis of the protein sample and protein before and after TEV protease cleavage, followed by inverse Ni-affinity purification. The full-length target protein is marked with a red triangle, while the tag-cleaved protein is indicated by a violet triangle. Labels as follow: M: Molecular Weight standards (in KDa); PRE TEV: Eluted fraction from Ni-NTA resin before tag-cleavage; POST TEV: protein after tag-cleavage; FT POST TEV: Flow Through from Ni-NTA resin post tag-cleavage; W POST TEV: wash fraction from Ni-NTA resin post tag-cleavage; EL POST TEV: eluted fraction from

Ni-NTA resin post tag-cleavage. **B.** SDS-page analysis of Size Exclusion chromatography purification. Pooled fractions are highlighted by blue arrow. Labels as follow: M: Molecular Weight standards (in KDa); INJ SEC: Injected sample; SEC fractions: eluted fractions from the main peak. **c** Elution profile from a Superdex 200 column; the fractions are shaded in blue.

16.4. Single point mutations decreased BLM unwinding activity towards G-quadruplexes

A two-step fluorescence-based helicase assay was performed to determine the impact of putative GSP sequence changes on BLM-mediated G4 unwinding.

Prior to test G4 unwinding, we verified that both the wild-type BLM protein and the two single mutants were able to fully unwind the DNA Fork after 30 minutes, suggesting that the GSP variants retained duplex helicase activity with an efficiency very similar to that observed with the WT protein. These findings are consistent with the published data on the bacterial RecQ helicase showing that sequence alterations in the GSP impair G4 unwinding but not duplex DNA helicase activity (Voter et al., 2018).

Although both mutants displayed some helicase activity on G4 substrates (Figure 53), G4 unwinding was significantly decreased. The BLM S965A variant exhibited a significantly reduced activity towards cKIT1, KRAS, Tel₂₆, and 2KM3 G4 structures (by 60%, 60%, 60%, and 50% of wild-type levels, respectively, at 6 nM protein), while its activity towards cMYC G4 was reduced by 40%. The BLM S897A variant showed a more modest reduction in G4 unwinding activity, with a 35-40% decrease compared to the wild-type for most G4s, except for cMYC G4, where the reduction was about 50%.

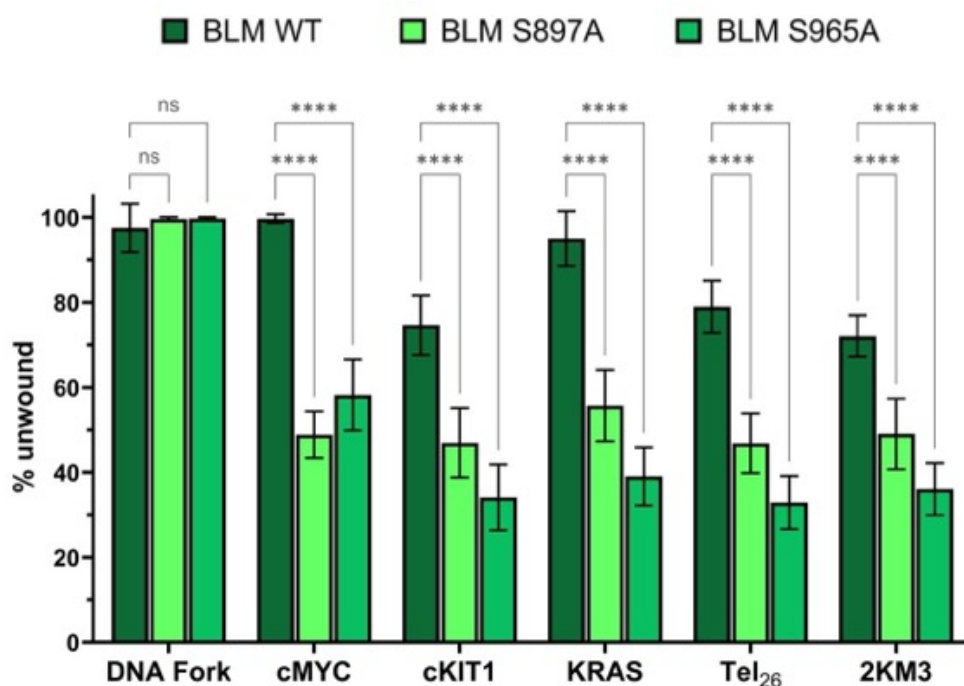


Figure 53. BLM WT, BLM S897A and BLM S965A unwinding activity towards various G-quadruplexes.

Preliminary analysis of BLM WT, BLM S897A and BLM S965A helicase activity against selected G4 DNA systems; error bars indicate standard deviation of three independent experiments. DNA Fork has been used as comparison. Statistical significance has been reported (**** represents statistical significance at $P < 0.0001$. ns represents statistically non-significant difference ($P > 0.05$); data analyzed with GraphPad Prism, using Two-Way ANOVA, comparing each cell mean with BLM WT chosen as control). Helicase assays were performed in annealing buffer containing 1 mM KCl and 99 mM NaCl, and in reaction buffer with 50 mM KCl.

These preliminary results highlight the importance of Ser897 and Ser965 residues in G4 unwinding. Although we expected that the single mutants would be unable to resolve G-quadruplex structures, we still observed some helicase activity, even if it significantly decreased. This suggests that mutating a single residue may not fully abrogate the helicase activity, and both residues may need to be altered for complete inhibition. A BLM variant with both residues mutated is planned for future studies. Further experiments are necessary to assess the affinity of each variant for G-quadruplexes, to determine whether these GSP mutations are critical for G4 binding. However, we expect that the GSP binding pocket binds a resolved rather than a folded G4 structure, and we therefore think that it is unlikely that the mutants would significantly affect the binding of a fully folded G4.

17. Structural studies

Important insight into the interactions between proteins and G4 structures would be provided by structural studies, in order to gain insight into the molecular determinants of interaction. Structural data could also provide a framework for the design or fine-tuning of small molecule ligands.

Over the past few decades, X-ray crystallography has been the primary technique used for solving protein structures and continues to be a key reference in the field. However, the crystallization step frequently represents a bottleneck in X-ray crystallography, particularly when working with helicases, as their inherent flexibility and polymorphism make it challenging to form the ordered lattice structures required for analysis. The rise of CryoEM and its ability to tackle smaller and smaller macromolecules could provide a possible tool to study flexible helicases; however, the size of all the RecQ helicases makes them very challenging samples for a microscopy experiments.

It is important to highlight that little structural data is available on how G4 structures interact with proteins in general and helicases in particular (see section 3.3 in Introduction), with only two or three structures providing significant insights.

I therefore attempted the characterization of the tridimensional atomic structure of proteins by single crystal X-ray diffraction. I performed some preliminary crystallization trials on the helicase:G4 complexes that showed the highest affinities, using the automated crystallization facility available in the laboratory in Elettra.

I first started with BLM protein, which showed a high affinity for many G4s and can be expressed and purified at high concentrations. The optimized purification protocol allowed me to obtain a large amount of very pure protein. Pure BLM protein solution was first concentrated to 5.4 mg/mL and 12.6 mg/mL for the first round of crystallization trials, then (based on the results of the first rounds, where many drops showed no or little precipitate) the protein concentration was increased up to 25 mg/mL for further screenings, to get closer to the protein solubility limit. My initial experiments focused on the co-crystallization with a couple of G4s having different topologies and showing the greatest affinity, e.g. cMYC G4 and Tel₂₆ G4. A large number of crystallization conditions need to be screened in order to find the best combination of buffers, pH values, temperature, additives and precipitants. I explored many different conditions using several 96-well crystallization screening kits that are commercially available: MD1-46 Morpheus, covering a range of pH, precipitants, PEGs and salt additives; MD MIDASplusTM, a crystallization screen ideal for soluble protein, protein/protein complexes, protein-nucleic acid complexes and sensitive macromolecular complexes; MD BCS Screen, that has been formulated to include PEG smears, enabling a broader search of chemical space than just using individual PEGs; and JCSG-plusTM Screen, an optimized sparse matrix screen based on previously successful crystallization conditions.

The first round of screening plates has been prepared using BLM protein in the presence of two different G4 structures in a ratio of 1:1.1 and 1 mM Mg-ADP. Plates have been incubated both at 4°C and 20°C, regularly checking for crystalline formations. In a few conditions very small crystals appeared; their diffraction properties were immediately evaluated on one of the two Elettra beamlines (XRD2) supporting macromolecular crystallography, but in all cases they turned out to be salt crystals. Additional crystallization trials have been performed, screening again the same crystallization kits, but this time increasing the protein concentration up to 25 mg/mL. Moreover, many other factors have been explored, changing the ratio between protein and G4 (1:1.25), using different G4 structures with an 11nt-single-strand tail at the 3'-end and adding various ATP analogs (such as ATP γ S, AMP-PNP). Despite all the attempts, no crystals were detected after more than one or two months of monitoring activity.

18. AlphaFold predictions

While this work was in progress, the introduction of Artificial Intelligence based structure predictions has sparked a revolution in structural biology. Although structure predictions remain predictions, and need to be validated by experimental data, undoubtedly AlphaFold (Senior et al., 2020) and RosettaFold (Baek et al., 2021) provide very useful framework to interpret biochemical and cellular information.

The first AI-based prediction servers were reasonably good at modelling compact domains and structures, but did not provide any information on multimeric proteins or protein-nucleic acid interaction. A significant breakthrough in 2024 (AlphaFold3, Abramson et al., 2024) opened up the possibility of modelling biomolecular interactions. However, it has to be stressed that, in our limited experience, the prediction of protein-protein interactions shows a higher degree of accuracy than the prediction of protein-nucleic acid interactions. For example, in the case of helicases, we have sometimes observed that the ssDNA does bind to the correct face of the protein, but shows the wrong polarity.

We tried also to use AlphaFold3 to obtain structural models of RecQ-G4 interactions, with very mixed results. The modelling depends critically on the G4 chosen. When using G4s without tail the result is always an unfolded polynucleotide chain bound to the canonical ssDNA binding groove of the helicase. We also obtain different results when using the full-length protein vs the catalytic domain; whether this can make sense in the case of RecQ4, where the N-terminal domain has been shown to

have a significant role in G4 binding, it is less easy to explain for other RecQ, where the evidence is that the catalytic domain is sufficient for binding and unwinding.

We are currently carrying out a systematic analysis, using a number of G4 with/without tails, and different RecQ fragments. Here I only briefly present the results for two complexes (Figure 54). The model for the catalytic domain of BLM helicase bound to a cMYC G4 with a polyT 3' tail shows the folded G4 and the tail binding in the canonical DNA binding groove; the protein does not make extensive interaction with the G4, and no specific elements involved in binding could be identified. More interestingly, the model for full-length RecQ4 bound to a telomeric G4 with a polyT tail confirms the prediction that the RSM motif in the N-terminal domain is indeed involved in G4 recognition (Papageorgiou et al., 2023).

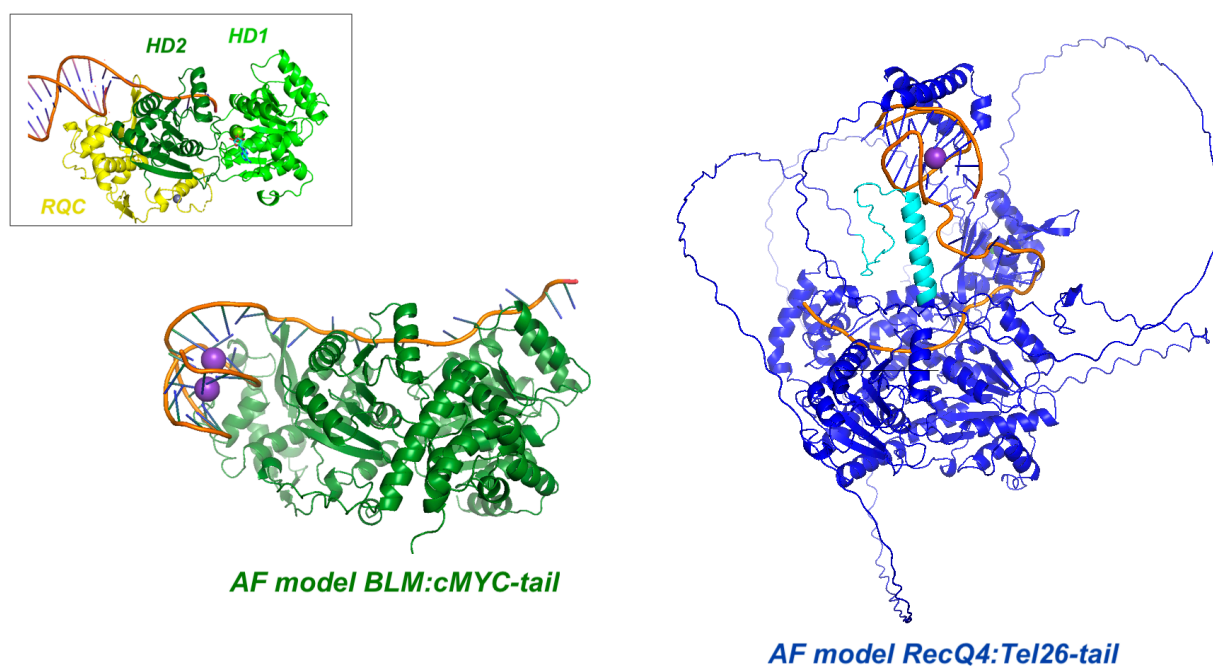


Figure 54. AlphaFold3 models of the interaction between RecQ and G4s. On the left panel the model for the catalytic domain of BLM in a complex with cMYC with a 3'tail; on the right panel the model for full-length RecQ4 in complex with a telomeric tailed G4. In the inset the crystal structure of RecQ1 bound to a DNA Fork is shown in the same orientation, with the HD1 domain in light green, the HD2 in dark green and the RCQ domain in yellow.

19. The interaction between *E. coli* DinG and G-quadruplexes

As reported at the beginning of the Results and Discussion section, our first inroad into studying helicase-G4 interactions used the bacterial DinG as a model system, in collaboration with C. Giancola and J. Amato at the University of Naples. DinG was easier to express and purify in large amounts with respect to human helicases and allowed us to become familiar with the unwinding assay used.

We thus performed a biophysical and biochemical analysis to study the specificity and selectivity of DinG in the binding and unwinding of unimolecular G4s in the presence and absence of well-known G4 ligands (De Piante et al., 2023). This side project enabled us to establish optimal experimental conditions for biochemical assays, particularly for the two-step fluorescence-based helicase assay, which was used in our lab for the first time to assess G4 unwinding. It was instrumental in determining the ideal reaction buffer with the correct ion concentration, the reaction timing, and the optimal protein concentration to use. These experiments were the starting point to perform the biochemical analysis of the human RecQ helicases towards G4s in a more reliable and fast way.

19.1. Expression and purification of DinG

The DinG helicase was cloned and expressed in *E. coli* cells, supplemented with FeCl₃. The protein was purified to homogeneity by Ni-affinity, followed by Heparin chromatography, which was able to remove all the impurities (Figure 55). Thus, no further polishing steps were needed since the optimized protocol allowed us to obtain large amount of pure protein, sufficient to perform the biochemical analysis. Concentrated protein solutions maintained a yellow color throughout the whole purification protocol, consistent with the retention of the FeS cluster, necessary for the activity.

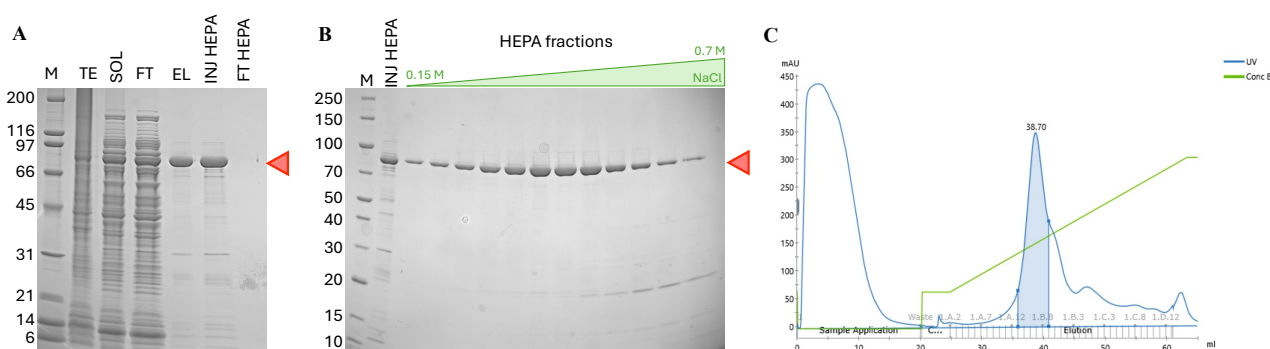


Figure 55. Large scale purification of *E. coli* DinG. A-B. SDS-PAGE analysis of purification steps from Ni-affinity to Heparin chromatography; the band corresponding to the target protein is shown with a triangle. Labels as follows: M: Molecular Weight standards (in kDa); TE: Total Extract; SOL: Soluble fraction; FT: Flow Through from Ni-NTA resin; EL: Eluted fraction from Ni-NTA resin; INJ HEPA: Injected sample; FT

HEPA: Flow Through from Heparin chromatography; HEPA fractions: Eluted fractions from the main peak. C. Chromatogram of the Heparin chromatography step. The blue curve is the absorbance measured at $\lambda = 280$ nm; the green curve shows the increasing NaCl concentration used to elute the proteins bound to heparin molecules.

19.2. Biochemical and biophysical characterization of DinG helicase

A systematic study of the interaction between recombinant DinG and a variety of G4s was carried out, focusing on both binding affinity and helicase activity. Although previous studies on *E. coli* DinG showed that the protein unwinds bimolecular but not unimolecular G4s, we were able to demonstrate for the first time that DinG not only binds with high affinity to unimolecular G4s but also resolves them. We selected a panel of well-characterized G4, including the 23-mer truncation of the human telomeric G4-forming sequence (Tel₂₃), KRAS, c-MYC, c-KIT1, BCL2 and T30695 (or T30923) (Table 13). Since our findings differ from earlier reports, which showed that DinG could not unwind unimolecular G4 (Bharti et al., 2013), we also tested the G4 sequence from the non-template strand of the transcribed regions of the *Zic1* gene used in that study. The topology of all the G4 structures was confirmed by CD spectroscopy (Figure 56 A, B, C, D, E, F, G). Furthermore, since DinG is an *E. coli* enzyme, we also explored its interaction with three different unimolecular parallel G4-forming sequences from the *E. coli* genome (Table 13) (CD spectra in Figure 56 H, I, L) (Cebrián et al., 2022). Lastly, two control sequences were included: the DNA Fork, selected as canonical helicase substrate, and the *G-scramble*, a G-rich oligonucleotide not able to assemble into a G4 structure.

G4	Sequence (5'-3')	Topology
cMYC	AGGGTGGGTAGGGTGGGT	promoter parallel
cKIT1	AGGGAGGGCGCTGGGAGGAGGG	promoter parallel
KRAS	AGGGCGGTGTGGGAATAGGGAA	promoter parallel
Bcl2 Mid	GGGCGCGGGAGGAATGGGCGGG	promoter parallel
Tel₂₃	TAGGGTTAGGGTTAGGGTTAGGG	telomere hybrid
T30695 (T30923)	GGGTGGGTGGGTGGGT	dimeric parallel
Zic1	GGTGGGGGGGCGGGGGAGGCCGGG	non-template strand of transcribed gene, parallel
EC-6	GGTGGGGAGGGGTAAGGGG	<i>E. coli</i> parallel
EC-7	GGGGGAGGAGGACGGGGG	<i>E. coli</i> parallel
EC-9	GGGGCGGGGTGGGTGGG	<i>E. coli</i> parallel
<i>G-scrambled</i>	GTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	

Table 13. List of G-quadruplexes used in the present study. Summary of G4 characteristics and sequences. Gs that participate in G4 formation are in bold.

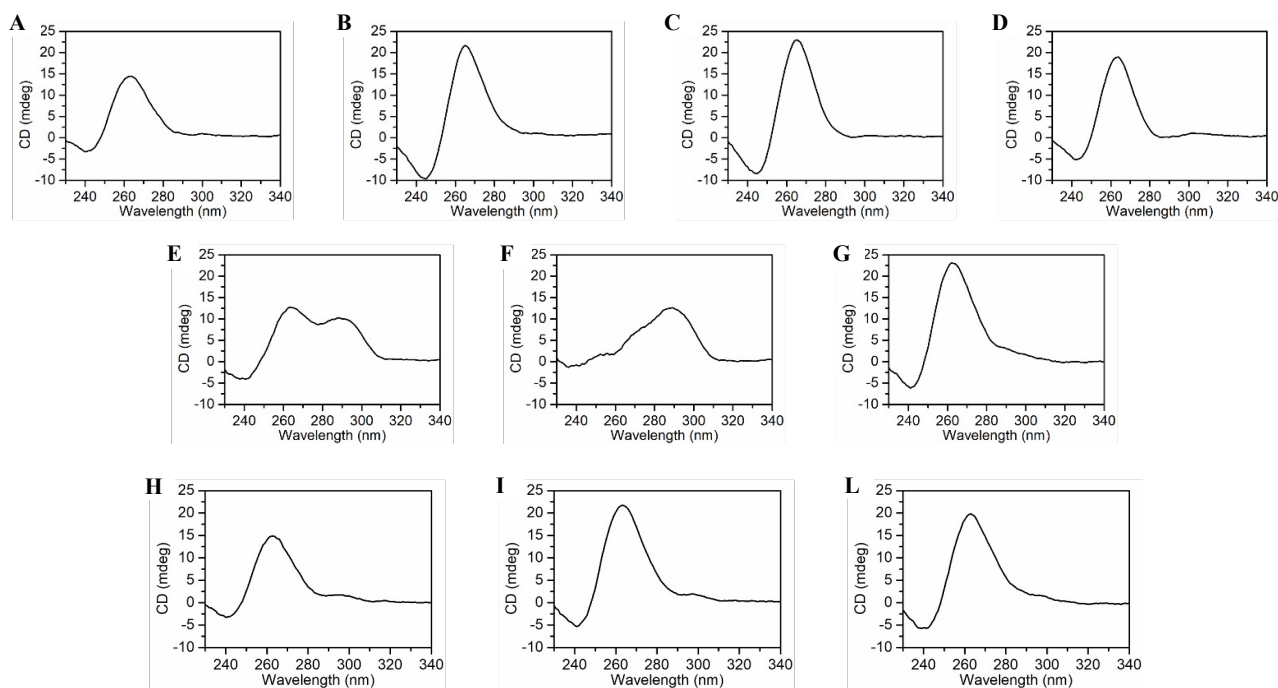


Figure 56. CD spectra of A. c-KIT1, B. c-MYC, C. KRAS, D. T30695, E. BCL2, F. Tel₂₃, G. Zic1, H. EC-6, I. EC-7, L. EC-9. Experiments conducted by our collaborators at the University of Naples.

19.2.1. DinG binds and unwinds G-quadruplex structures with different efficacy

To explore the binding interaction between DinG and G4s, our collaborators at the University of Naples conducted a physicochemical analysis using Surface Plasmon Resonance (SPR) technique. Compared to other methods for studying DNA–protein interactions, SPR provides valuable information about the affinity of interactions and the kinetic parameters, by measuring the association (k_{on}) and dissociation (k_{off}) rate constants, from which the equilibrium dissociation constant (K_D) can be estimated. Figure 56 shows SPR sensorgrams obtained by injecting increasing concentrations of G4s (analyte) over the DinG-immobilized sensor-chip.

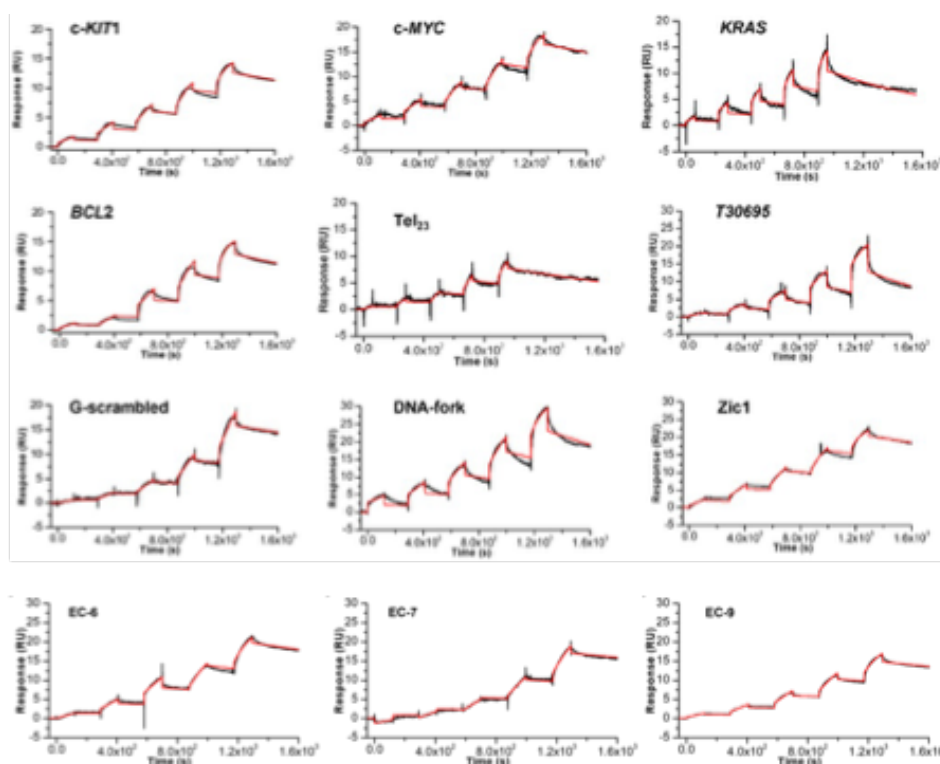


Figure 57. Time evolution SPR sensorgrams obtained at 25 °C by injection of increasing concentrations of each investigated G4 (from 0.062 to 1 μM) on the chip-immobilized DinG helicase. Experiments conducted by our collaborators at the University of Naples.

The results of the SPR analysis indicate that DinG binds to the investigated G4s with affinities similar or better to those observed for the DNA Fork substrate. The highest affinity value is seen for the interaction with the Tel₂₃, which forms a hybrid G4 structure as major conformation, followed by the BCL2, which folds in a hybrid G4 structure as well. The lowest affinities are observed for DinG binding to the T30695 sequence, which forms a dimeric all-parallel stranded G4 structure with single-nucleotide containing loops, as well as with the *G-scramble* sequence, which does not form

any G4 or other secondary structures. These results suggest that DinG recognition of DNA is dependent on the specific G4 sequence and conformation. The affinities towards G-rich sequences from *E. coli* DNA are similar to those observed for parallel human G4s.

We evaluated the ability of DinG to unwind the same panel of G4s, using the two-step fluorescence-based helicase assay described above (Mendoza et al., 2015). Originally developed for the Pif1 helicase, this assay was easily adapted for DinG, which shares the same unwinding polarity. Our results show that DinG was able to almost completely unwind c-KIT1 G4 (around 85%), with the same efficiency observed for the DNA Fork substrate (Figure 58B). KRAS and Tel₂₃ G4s were unwound with approximately 70–75% efficiency, while most of the other G4 structures (c-MYC, BCL2, EC6, EC7 and EC9) were unwound with an efficiency of about 50%. In contrast to previous studies, we found that DinG was able to unwind Zic1 G4, although at a lower level than seen for the best G4 substrates (c-KIT1 and KRAS). This discrepancy may be due to the relatively lower efficiency of DinG towards this particular substrate, and the use of a different assay. Consistently with the SPR data, the lowest unwinding activity was observed for the T30695 G4 structure (Figure 58), which also exhibited the weakest binding affinity.

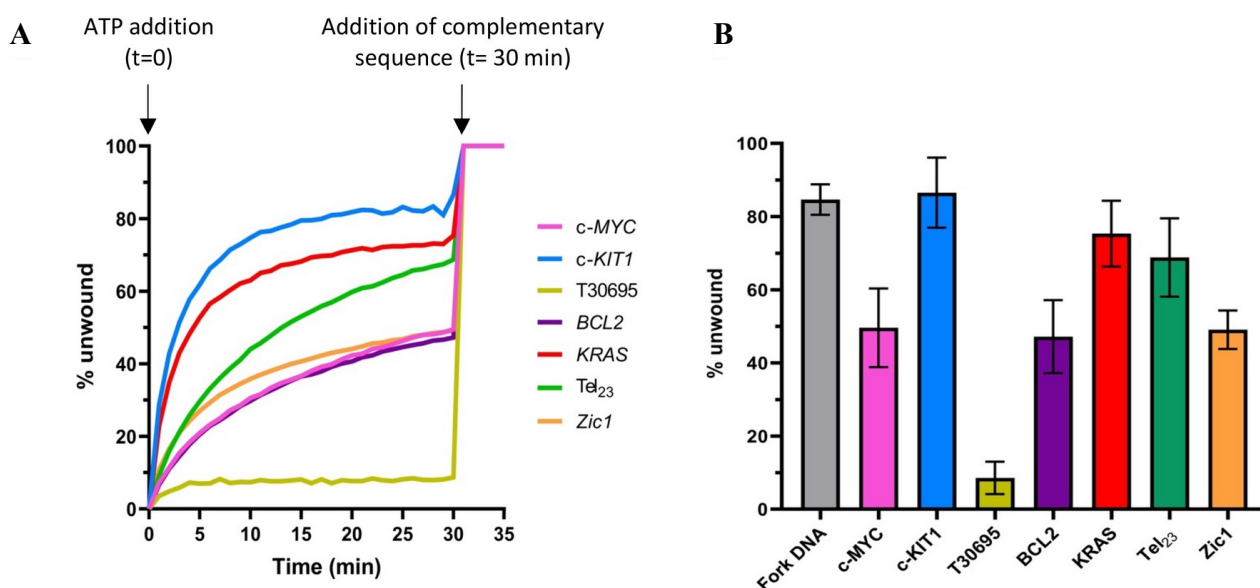


Figure 58. DinG helicase activity towards G4s. **A.** Representative plots of fluorescence emission vs time for unwinding of selected G4s (c-MYC, c-KIT1, T30695, Tel₂₃, KRAS, BCL2 and Zic1). ATP was added to begin the reaction (t=0 min); the complementary strands (C-c-MYC, C-c-KIT1, C-T30695, C-Tel₂₃, C-KRAS, C-BCL2 and C-Zic1 respectively) were added once the reactions reached a plateau (t=30 min) as described in section 10.5.3 in Materials and methods. **B.** Quantification of the helicase activity of *E. coli* DinG against selected G4s; error bars indicate standard deviation of three independent experiments.

19.2.2. DinG unwinding activity is compromised by G4 ligands

In the present work we also studied the impact of four well-characterized G4 ligands, namely Pyridostatin (PDS) (Müller et al., 2012), PhenDC3 (De Cian et al., 2007), BRACO-19 (Burger et al., 2005) and Netropsin (Pagano et al., 2010), in the binding and unwinding of DinG towards the G4s studied. These ligands are known to bind and stabilize G4 structures: while PDS, PhenDC3 and BRACO-19 show an end stacking mode of binding, Netropsin binds to the G4 grooves. These peculiarities allowed us to investigate whether the mode of ligand binding influenced DinG ability to bind and/or unwind G4s. Not surprisingly, when the G4 structure was stabilized by the selected G4 ligands, in most cases the activity was abolished, particularly with PhenDC3 and PDS. However, some residual helicase activity was observed in the presence of BRACO-19 and Netropsin, suggesting that these ligands have a lower impact on the helicase's G4 unwinding (Figure 59).

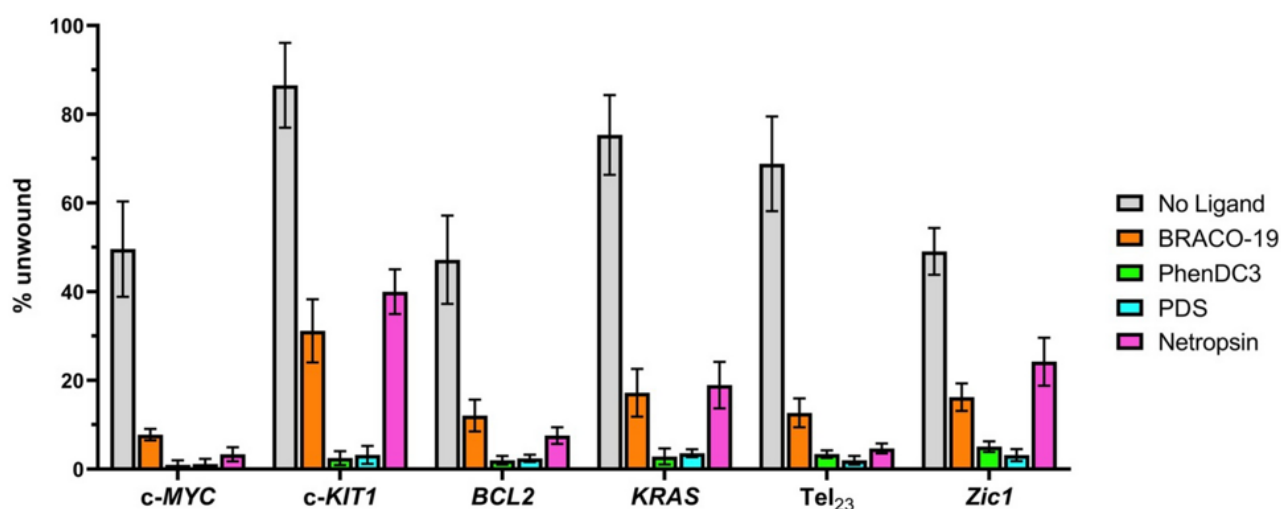


Figure 59. DinG helicase activity in the presence of G4 ligands. *E. coli* DinG helicase activity against c-MYC, c-KIT1, BCL2, KRAS, Tel₂₃ and Zic1 G4s in the absence or presence of selected G4 ligands (BRACO-19, PhenDC3, PDS and Netropsin). Error bars indicate standard deviation of three independent experiments.

To clarify whether the decreased activity of DinG is correlated to its lower capability to recognize and bind to the G4 when complexed to the ligands, SPR measurements were again carried out by our collaborators on the c-KIT1 G4, as the best substrate for unwinding (data not shown). The results show that the presence of BRACO-19, PhenDC3 and Pyridostatin affected but not completely hindered the binding affinity of DinG for the c-KIT1. In contrast, Netropsin had no significant impact on the G4 recognition by the helicase, likely due to its different binding mode compared to the other investigated G4 ligands. Notably, changes in binding affinity did not always correlate directly with

reduced unwinding activity, but it could be due to the different thermal stabilization properties induced by the ligands on the c-KIT1 G4 thermal stability. Thus, PhenDC3 and PDS significantly increased the melting temperature of c-KIT1 G4, while BRACO-19 and Netropsin had little or no effect on G4 stability respectively.

As DinG is present in several pathogenic bacteria (including *Mycobacterium tuberculosis*, *Neisseria meningitidis* and *Neisseria gonorrhoeae*) a better understanding of the mechanism of action of the FeS helicase DinG can have a medical impact in terms of design of novel drugs. Moreover, as a model system for the more challenging eukaryotic FeS helicases, such as FancJ, DDX11 and RTEL1, which have all being implicated in G4 metabolism, the study of the interaction between DinG and G4s in the absence and presence of G4 ligands may provide a framework to better harness the potential of G4 stabilization in genome stability and cancer onset and development.

Conclusions

This PhD thesis provided a comprehensive in vitro analysis of the interaction between human RecQ helicases and G-quadruplex structures, offering insights into the specificity and selectivity of different helicases in the metabolism of a variety of G4s.

G-quadruplexes are involved in many cellular pathways, such as transcription regulation, progression of the replication fork, DNA stability and repair, and telomere maintenance. Not surprisingly, given their widespread role in biology, they are implicated in genome instability and can have an impact on malignant transformation and cancer development. In particular, whereas G4 structures are distributed throughout the genome, they are enriched in certain regions, and often form very stable structures within human oncogenic promoters and at telomeres. For this reason, G4s are currently tested as a therapeutic target and many chemical molecules are being designed to target these structures, with the aim to downregulate transcription or to block telomere elongation in cancer cells. Among the helicases that evolved to unravel or modulate G4 structures, particularly important are human RecQ helicases, as various members of the family were shown to be able to bind and unwind G4 structures. RecQ helicases are involved in all aspects of nucleic acid metabolism, and their presence is essential to maintain genome integrity, replication, repair, and recombination. Moreover, they show a preference for non-canonical nucleic acid structures, including G-quadruplexes.

Through a series of biochemical assays, structural characterization attempts, and bioinformatic predictions, this work tried to elucidate the distinct and overlapping roles of the five human RecQ helicases (RecQ1, BLM, WRN, RecQ4, and RecQ5) in recognizing and processing G4s, advancing our understanding of their functional diversity.

For these purposes, we produced the catalytic domain of the five human RecQ helicases, and we cloned additional RecQ4 variants including its N-terminal domain, homologous to the yeast Sld2 factor, that had been implicated in G4 recognition. Expression and purification protocols were optimized, and successful expression and purification of all the proteins enabled detailed functional studies with G4s, measuring both the binding affinity and the helicase activity. Biochemical analysis firstly confirmed that all the catalytic RecQ helicase fragments exhibited helicase activity, with RecQ1, BLM, WRN, and RecQ4 showing complete resolution of the DNA Fork, consistent with literature reports about the roles of RecQ helicases in DNA replication and repair pathways. A systematic analysis of the interactions between each human RecQ with a panel of eleven diverse G4 was then carried out. In particular, BLM and WRN demonstrated high binding affinity and significant helicase activity towards many G4 structures, without clearly discriminating between promoter or telomeric G4s. RecQ1, RecQ4, and RecQ5 displayed medium to low or no affinity for unimolecular

G4s, with RecQ4 and RecQ5 lacking G4 unwinding activity altogether, whilst RecQ1 exhibited only minimal G4 unwinding capability. These findings align with previous reports indicating that RecQ1, RecQ4, and RecQ5 are less specialized for G4 resolution compared to BLM and WRN, supporting the idea that specific RecQ family members have evolved distinct roles in genome maintenance and in G4 metabolism, with some showing limited interaction with non-canonical DNA structures such as G4s.

G-quadruplexes exhibit a high degree of polymorphism that depends on the characteristics of local environment, sequence composition, loop length and composition, as well as tetrad geometry, which can have an impact on their structural stability and folding. All these features contribute to the unique three-dimensional conformations observed across G4 structures. As an helicase that binds G4 would need to engage with this unique 3D conformation, it was hypothesized that RecQ helicases might exhibit specificity for certain G4 structures, possibly recognizing unique topological or sequence-dependent characteristics. Although we did observe a degree of selectivity, with a clear preference for certain G4, upon detailed analysis of the structural determinants, we have not been able to detect any specific feature that discriminate between the G4 that are bound and unwind efficiently, from those that are less interacting. We demonstrated that the N-terminal domain of RecQ4 shows high binding affinity for several G4 structures, supporting the potential G4-interacting capabilities of the protein. Similar results were obtained also with the preliminary analysis done on the full-length protein of RecQ4, which binds with high affinity to many G4 structures. Its unwinding activity on G4s needs to be more extensively characterized, but preliminary data suggests that RecQ4 is less active towards G4 than the most efficient paralogs. These initial findings suggest that RecQ4 may have a regulatory role in G4 recognition but may lack the unwinding capacity of BLM and WRN, in line with previous reports that suggested RecQ4 may protect the G4 structure or recruit other factors.

Different monovalent cations (e.g. K^+ , Na^{2+} and Li^+) and their relative concentrations could stabilize particular G4 topologies, influencing also the G4 thermal stability. To make sure that the differences we saw were due to an intrinsic helicase preference and not simply to G4 stability, we carried out CD spectroscopy in various buffer conditions. Analyzing BLM and WRN ability to unwind G4 structures in response to different ionic conditions, we observed different behavior for the two most active helicases: BLM activity remained unaffected by increasing potassium concentrations, whereas WRN activity decreased in high KCl environments.

Based on the report of a presence of a guanine-specific binding pocket interacting with a resolved G4 in the crystal structure of the bacterial *Cronobacter sakazakii* RecQ helicase, we carried out a

structural bioinformatic analysis on all the human RecQ helicases, to verify whether this pocket is conserved, and could represent a general feature of RecQ:G4 interactions. This analysis revealed some residues that might be implicated in the pocket formation and thus in G4 recognition and processing in all human RecQ helicases. These residues resulted to be well conserved among RecQ of different species, confirming their possible role in G4 interaction. Site-direct mutagenesis was performed on BLM and WRN helicases (BLM Ser897Ala, Ser965Ala, WRN Ser773Ala, Asp872Ala), mutating the identified residues to check whether they would be critical for G4 unwinding. So far, preliminary results showed that BLM mutants exhibited reduced G4 unwinding but retained normal DNA Fork activity, suggesting these residues do play a key role in G4-specific resolution. Although additional studies are needed, however, this provides a novel contribution to the understanding of RecQ:G4 interactions, strongly suggesting that the mechanism proposed for CsRecQ may be extended to all RecQ.

In parallel, crystallization experiments were carried out on the most stable BLM:G4 complexes to try to shed light on the molecular details of the interactions. Although many enzymes have been effectively analyzed through macromolecular crystallography, human helicases are challenging to crystallize due to their high flexibility and conformational changes that can occur during their activity, making them extremely polymorphic. A large number of crystallization conditions were explored but we were not able to obtain any diffracting crystal.

A number of reports suggest a significant interplay between G-quadruplexes and R-loops. Together with Luisa MR Napolitano, we also performed a parallel analysis on the full-length protein of RecQ4 in order to study its role in R-loops resolution, which may occur at stalled replication forks or, more intriguingly, in the context of the initiation of the replication. Our results showed that, although the protein binds a variety of nucleic acids, it has a preference for R-loops which are unwound with the highest efficiency. We further measured R-loops and D-loop resolution among the five human RecQ helicases and showed that RecQ4 does indeed unwind R-loops with a much higher efficiency compared with the other RecQ family members. These data provide evidence for a novel functional role of RecQ4 protein in R-loops resolution, contributing to the understanding of the molecular basis of Rothmund Thomson syndrome and other RecQ4-associated disorders, and to the understanding of its role in carcinogenesis and response to cancer therapy. The difference in displacement loops unwinding among the five human RecQs further argues for a degree of diversification and specialization among the paralogs.

Starting with DinG as a model system for G4 interaction indeed facilitated the establishment of experimental procedures and of the protocol to use to study RecQ helicases activity. In collaboration

with the University of Naples, we carried out a detailed study of the interaction between *E. coli* DinG and a variety of G4s, by employing physiochemical and biochemical methodologies. Results showed that DinG binds to most of the investigated G4s with little discrimination, while it exhibits a clear degree of unwinding specificity towards different G4 topologies. Whereas the previous literature only reports *E. coli* DinG unwinding bi- or tetra-molecular G4s, we see for the first time the protein acting on the more physiological unimolecular G4s. Besides helping to clarify the precise role of the DinG helicase in bacterial G4 DNA metabolism and genomic processes, DinG can provide a working model for the more complex eukaryotic FeS helicases, which include medically important enzymes such as XPD, DDX11, RTEL1 and FANCI. The results obtained have been published (De Pianta et al., 2023).

Future work

Future work should focus on extending and completing the biochemical characterization of the interaction between human RecQ helicases and G-quadruplexes, further efforts in understanding the structural/molecular determinants of G4 specificity, and performing additional structural studies to provide deeper insights into the mechanistic basis for G4 interaction.

- We will complete the analysis of the interaction between the N-terminal domain of RecQ4 with G4 structures, finalizing the study of the binding affinity of the already-expressed proteins, but also trying to express and characterize the other cloned fragments.
- While the full-length RecQ4 displayed high binding affinity to G4 structures, its unwinding activity on these substrates remains unclear. We will further characterize its ability to unwind G4s and its potential regulatory role in G4 metabolism. This will clarify whether RecQ4 might play a direct role in G4 resolution or might function as a stabilizing factor recruiting other proteins for G4 unwinding.
- Bioinformatic predictions have identified potential G4-specific binding pockets in human RecQ helicases. Future work will focus on finalizing the analysis on BLM variants, trying also to produce the double mutant (with both residues mutated). WRN mutants should also be expressed and tested against G4 structures. Mutate the predicted guanine-binding pocket of the other RecQ helicases (RecQ1, RecQ4 and RecQ5) could be useful to get deeper insight into the function of this region. The validation of these pockets could be done also through structural studies, trying co-crystallizing the mutant together with a G4 structure, to further clarify the role of these residues in G4-specific interactions.
- We plan to expand the binding analysis using G4 sequences modified with the addition of 3'/5' tails of different lengths to investigate changes in the specificity depending on the presence of a ssDNA. This library would be then extended to bimolecular, testing both binding and helicase activity.
- Whether the cellular roles and implications of DNA G4s are now well established, RNA G4s are still relatively less explored and the study of their role in cellular processing and diseases is a rapidly developing field. Future analysis would include the exploration of a variety of RNA G4s. The ability of some RecQ in unwinding R-loops and DNA/RNA hybrids does suggest the possibility that these enzymes might indeed engage RNA.

- Given the involvement of G4 structures in oncogenic promoters and telomeric regions, and the crucial role of RecQ helicases in G4 metabolism, future work should explore the potential of targeting G4 structures for cancer therapy. Screening for G4-stabilizing ligands that could modulate the activity of BLM, WRN, or other RecQ helicases on G4s, could provide new therapeutic strategies for targeting cancer cells with high level of G4 expression.
- Despite numerous attempts performed and still ongoing, the crystallization of G4:RecQ complexes has proven challenging. CryoEM could provide an alternative, as it can handle the flexibility and conformational changes inherent to helicases, but the size of RecQs is too small for the current limits of the technique. However, there is evidence for substrate-triggered dimerization for some RecQ (eg. RecQ1 in a complex with a DNA fork). As Trieste is currently installing a high-hand microscope, we will soon have the possibility of exploring this possibility. High-resolution structures of these complexes would provide crucial information on the helicase-G4 interaction interface and insights into the design of inhibitors that specifically disrupt this interaction in pathological contexts.
- Although this work primarily focused on in vitro biochemical assays, future studies should extend these findings to in vivo models, such as knockdown or knockout of individual RecQ helicases in human cell lines or model organisms, or cells of patients deficient for RecQ helicases. These studies would provide insights into the physiological relevance of G4 resolution by different RecQ helicases, particularly concerning RecQ genetic diseases and RecQ-related cancer development, genome instability, and response to replication stress.

These experiments will provide a more comprehensive understanding of the molecular mechanisms of G4 recognition and processing by RecQ helicases, providing insight into both basic research and potential therapeutic applications in cancer and genome stability disorders.

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., ... Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016), 493–500. <https://doi.org/10.1038/s41586-024-07487-w>
- Abu-Libdeh, B., Jhujh, S. S., Dhar, S., Sommers, J. A., Datta, A., Longo, G. M., Grange, L. J., Reynolds, J. J., Cooke, S. L., McNee, G. S., Hollingworth, R., Woodward, B. L., Ganesh, A. N., Smerdon, S. J., Nicolae, C. M., Durlacher-Betzer, K., Molho-Pessach, V., Abu-Libdeh, A., Meiner, V., Moldovan, G. L., ... Stewart, G. S. (2022). RECON syndrome is a genome instability disorder caused by mutations in the DNA helicase RECQL1. *The Journal of clinical investigation*, 132(5), e147301. <https://doi.org/10.1172/JCI147301>
- Aggarwal, M., Banerjee, T., Sommers, J. A., & Brosh, R. M., Jr (2013). Targeting an Achilles' heel of cancer with a WRN helicase inhibitor. *Cell cycle (Georgetown, Tex.)*, 12(20), 3329–3335. <https://doi.org/10.4161/cc.26320>
- Agarwal, T., Roy, S., Kumar, S., Chakraborty, T. K., & Maiti, S. (2014). In the sense of transcription regulation by G-quadruplexes: asymmetric effects in sense and antisense strands. *Biochemistry*, 53(23), 3711–3718. <https://doi.org/10.1021/bi401451q>
- Ahn, B., Harrigan, J. A., Indig, F. E., Wilson, D. M., 3rd, & Bohr, V. A. (2004). Regulation of WRN helicase activity in human base excision repair. *The Journal of biological chemistry*, 279(51), 53465–53474. <https://doi.org/10.1074/jbc.M409624200>
- Allison, D. F., & Wang, G. G. (2019). R-loops: formation, function, and relevance to cell stress. *Cell stress*, 3(2), 38–46. <https://doi.org/10.15698/cst2019.02.175>
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Ambrus, A., Chen, D., Dai, J., Jones, R. A., & Yang, D. (2005). Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, 44(6), 2048–2058. <https://doi.org/10.1021/bi048242p>
- Awadasseid, A., Ma, X., Wu, Y., & Zhang, W. (2021). G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 139, 111550. <https://doi.org/10.1016/j.biopha.2021.111550>
- Awate, S., & Brosh, R. M., Jr (2017). Interactive Roles of DNA Helicases and Translocases with the Single-Stranded DNA Binding Protein RPA in Nucleic Acid Metabolism. *International journal of molecular sciences*, 18(6), 1233. <https://doi.org/10.3390/ijms18061233>
- Bachrati, C. Z., & Hickson, I. D. (2008). RecQ helicases: guardian angels of the DNA replication fork. *Chromosoma*, 117(3), 219–233. <https://doi.org/10.1007/s00412-007-0142-4>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., ... Baker,

- D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* (New York, N.Y.), 373(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- Balasubramanian, S., Hurley, L. H., & Neidle, S. (2011). Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?. *Nature reviews. Drug discovery*, 10(4), 261–275. <https://doi.org/10.1038/nrd3428>
- Belotserkovskii, B. P., Soo Shin, J. H., & Hanawalt, P. C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic acids research*, 45(11), 6589–6599. <https://doi.org/10.1093/nar/gkx403>
- Bernstein, D. A., Zittel, M. C., & Keck, J. L. (2003). High-resolution structure of the E.coli RecQ helicase catalytic core. *The EMBO journal*, 22(19), 4910–4921. <https://doi.org/10.1093/emboj/cdg500>
- Berti, M., Ray Chaudhuri, A., Thangavel, S., Gomathinayagam, S., Kenig, S., Vujanovic, M., Odreman, F., Glatter, T., Graziano, S., Mendoza-Maldonado, R., Marino, F., Lucic, B., Biasin, V., Gstaiger, M., Aebersold, R., Sidorova, J. M., Monnat, R. J., Jr, Lopes, M., & Vindigni, A. (2013). Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nature structural & molecular biology*, 20(3), 347–354. <https://doi.org/10.1038/nsmb.2501>
- Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J. M., & Lemaitre, J. M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology*, 19(8), 837–844. <https://doi.org/10.1038/nsmb.2339>
- Bettin, N., Oss Pegorar, C., & Cusanelli, E. (2019). The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. *Cells*, 8(3), 246. <https://doi.org/10.3390/cells8030246>
- Bharti, S. K., Sommers, J. A., George, F., Kuper, J., Hamon, F., Shin-ya, K., Teulade-Fichou, M. P., Kisker, C., & Brosh, R. M., Jr (2013). Specialization among iron-sulfur cluster helicases to resolve G-quadruplex DNA structures that threaten genomic stability. *The Journal of biological chemistry*, 288(39), 28217–28229. <https://doi.org/10.1074/jbc.M113.496463>
- Booy, E. P., McRae, E. K., & McKenna, S. A. (2015). Biochemical characterization of G4 quadruplex telomerase RNA unwinding by the RNA helicase RHAU. *Methods in molecular biology (Clifton, N.J.)*, 1259, 125–135. https://doi.org/10.1007/978-1-4939-2214-7_9
- Brosh R. M., Jr (2013). DNA helicases involved in DNA repair and their roles in cancer. *Nature reviews. Cancer*, 13(8), 542–558. <https://doi.org/10.1038/nrc3560>
- Brosh, R. M., Jr, & Wu, Y. (2021). An emerging picture of FANCDJ's role in G4 resolution to facilitate DNA replication. *NAR cancer*, 3(3), zcab034. <https://doi.org/10.1093/narcan/zcab034>
- Brosh, R. M., Jr, Orren, D. K., Nehlin, J. O., Ravn, P. H., Kenny, M. K., Machwe, A., & Bohr, V. A. (1999). Functional and physical interaction between WRN helicase and human replication protein A. *The Journal of biological chemistry*, 274(26), 18341–18350. <https://doi.org/10.1074/jbc.274.26.18341>

- Brosh, R. M., Jr, Li, J. L., Kenny, M. K., Karow, J. K., Cooper, M. P., Kureekattil, R. P., Hickson, I. D., & Bohr, V. A. (2000). Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *The Journal of biological chemistry*, 275(31), 23500–23508. <https://doi.org/10.1074/jbc.M001557200>
- Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H., & Brooks, T. A. (2011). Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. *The Journal of biological chemistry*, 286(47), 41018–41027. <https://doi.org/10.1074/jbc.M111.274720>
- Bryan T. M. (2020). G-Quadruplexes at Telomeres: Friend or Foe?. *Molecules* (Basel, Switzerland), 25(16), 3686. <https://doi.org/10.3390/molecules25163686>
- Budhathoki, J. B., Ray, S., Urban, V., Janscak, P., Yodh, J. G., & Balci, H. (2014). RecQ-core of BLM unfolds telomeric G-quadruplex in the absence of ATP. *Nucleic acids research*, 42(18), 11528–11545. <https://doi.org/10.1093/nar/gku856>
- Budhathoki, J. B., Maleki, P., Roy, W. A., Janscak, P., Yodh, J. G., & Balci, H. (2016). A Comparative Study of G-Quadruplex Unfolding and DNA Reeling Activities of Human RECQ5 Helicase. *Biophysical journal*, 110(12), 2585–2596. <https://doi.org/10.1016/j.bpj.2016.05.016>
- Bugreev, D. V., Mazina, O. M., & Mazin, A. V. (2009). Bloom syndrome helicase stimulates RAD51 DNA strand exchange activity through a novel mechanism. *The Journal of biological chemistry*, 284(39), 26349–26359. <https://doi.org/10.1074/jbc.M109.029371>
- Burger, A. M., Dai, F., Schultes, C. M., Reszka, A. P., Moore, M. J., Double, J. A., & Neidle, S. (2005). The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer research*, 65(4), 1489–1496. <https://doi.org/10.1158/0008-5472.CAN-04-2910>
- Cadoni, E., De Paepe, L., Manicardi, A., & Madder, A. (2021). Beyond small molecules: targeting G-quadruplex structures with oligonucleotides and their analogues. *Nucleic acids research*, 49(12), 6638–6659. <https://doi.org/10.1093/nar/gkab334>
- Carvalho, J., Queiroz, J.A. & Cruz, C. (2017). Circular Dichroism of G-Quadruplex: a Laboratory Experiment for the Study of Topology and Ligand Binding. *Journal of Chemical Education* 94 (10), 1547-1551. <https://doi.org/10.1021/acs.jchemed.7b00160>
- Castillo Bosch, P., Segura-Bayona, S., Koole, W., van Heteren, J. T., Dewar, J. M., Tijsterman, M., & Knipscheer, P. (2014). FANCD1 promotes DNA synthesis through G-quadruplex structures. *The EMBO journal*, 33(21), 2521–2533. <https://doi.org/10.15252/embj.201488663>
- Cebrián, R., Belmonte-Reche, E., Pirota, V., de Jong, A., Morales, J. C., Freccero, M., Doria, F., & Kuipers, O. P. (2022). G-Quadruplex DNA as a Target in Pathogenic Bacteria: Efficacy of an Extended Naphthalene Diimide Ligand and Its Mode of Action. *Journal of medicinal chemistry*, 65(6), 4752–4766. <https://doi.org/10.1021/acs.jmedchem.1c01905>
- Chakraborty, P., & Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair*, 10(6), 654–665. <https://doi.org/10.1016/j.dnarep.2011.04.013>
- Chatterjee, S., Zagelbaum, J., Savitsky, Sturzenegger, A., Huttner, D., Janscak, P., Hickson, I. D., Gileadi O., & Rothenberg E. (2014). Mechanistic insight into the interaction of BLM helicase

with intra-strand G-quadruplex structures. *Nat Commun* 5, 5556. <https://doi.org/10.1038/ncomms6556>

- Chen, M. C., Tippana, R., Demeshkina, N. A., Murat, P., Balasubramanian, S., Myong, S., & Ferré-D'Amaré, A. R. (2018). Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature*, 558(7710), 465–469. <https://doi.org/10.1038/s41586-018-0209-9>
- Cheng, K., & Wigley, D. B. (2018). DNA translocation mechanism of an XPD family helicase. *eLife*, 7, e42400. <https://doi.org/10.7554/eLife.42400>
- Cheok, C. F., Wu, L., Garcia, P. L., Janscak, P., & Hickson, I. D. (2005). The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic acids research*, 33(12), 3932–3941. <https://doi.org/10.1093/nar/gki712>
- Compton, S. A., Tolun, G., Kamath-Loeb, A. S., Loeb, L. A., & Griffith, J. D. (2008). The Werner syndrome protein binds replication fork and holliday junction DNAs as an oligomer. *The Journal of biological chemistry*, 283(36), 24478–24483. <https://doi.org/10.1074/jbc.M803370200>
- Conroy, D. W., Xu, Y., Shi, H., Gonzalez Salguero, N., Purusottam, R. N., Shannon, M. D., Al-Hashimi, H. M., & Jaroniec, C. P. (2022). Probing Watson-Crick and Hoogsteen base pairing in duplex DNA using dynamic nuclear polarization solid-state NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 119(30), e2200681119. <https://doi.org/10.1073/pnas.2200681119>
- 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 reports*, 1(1), 80–84. <https://doi.org/10.1093/embo-reports/kvd004>
- Cooper, M. P., Machwe, A., Orren, D. K., Brosh, R. M., Ramsden, D., & Bohr, V. A. (2000). Ku complex interacts with and stimulates the Werner protein. *Genes & development*, 14(8), 907–912.
- Crabbe, L., Verdun, R. E., Haggblom, C. I., & Karlseder, J. (2004). Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science (New York, N.Y.)*, 306(5703), 1951–1953. <https://doi.org/10.1126/science.1103619>
- Criscuolo, A., Napolitano, E., Riccardi, C., Musumeci, D., Platella, C., & Montesarchio, D. (2022). Insights into the Small Molecule Targeting of Biologically Relevant G-Quadruplexes: An Overview of NMR and Crystal Structures. *Pharmaceutics*, 14(11), 2361. <https://doi.org/10.3390/pharmaceutics14112361>
- Croteau, D. L., Popuri, V., Opresko, P. L., & Bohr, V. A. (2014). Human RecQ helicases in DNA repair, recombination, and replication. *Annual review of biochemistry*, 83, 519–552. <https://doi.org/10.1146/annurev-biochem-060713-035428>
- Cui, S., Klima, R., Ochem, A., Arosio, D., Falaschi, A., & Vindigni, A. (2003). Characterization of the DNA-unwinding activity of human RECQ1, a helicase specifically stimulated by human replication protein A. *The Journal of biological chemistry*, 278(3), 1424–1432. <https://doi.org/10.1074/jbc.M209407200>

- Cunniff, C., Bassetti, J. A., & Ellis, N. A. (2017). Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Molecular syndromology*, 8(1), 4–23. <https://doi.org/10.1159/000452082>
- Dabral, P., Uppal, T., & Verma, S. C. (2023). G-quadruplexes of KSHV oriLyt play important roles in promoting lytic DNA replication. *Microbiology spectrum*, 11(6), e0531622. <https://doi.org/10.1128/spectrum.05316-22>
- Dai, J., Chen, D., Jones, R. A., Hurley, L. H., & Yang, D. (2006). NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic acids research*, 34(18), 5133–5144. <https://doi.org/10.1093/nar/gkl610>
- Dai, J., Carver, M., Punchihewa, C., Jones, R. A., & Yang, D. (2007). Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K⁺ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic acids research*, 35(15), 4927–4940. <https://doi.org/10.1093/nar/gkm522>
- Dai, Y. X., Guo, H. L., Liu, N. N., Chen, W. F., Ai, X., Li, H. H., Sun, B., Hou, X. M., Rety, S., & Xi, X. G. (2022). Structural mechanism underpinning Thermus oshimai Pif1-mediated G-quadruplex unfolding. *EMBO reports*, 23(7), e53874. <https://doi.org/10.15252/embr.202153874>
- Damerla, R. R., Knickelbein, K. E., Strutt, S., Liu, F. J., Wang, H., & Opresko, P. L. (2012). Werner syndrome protein suppresses the formation of large deletions during the replication of human telomeric sequences. *Cell cycle (Georgetown, Tex.)*, 11(16), 3036–3044. <https://doi.org/10.4161/cc.21399>
- Davis, A. P., Wieggers, T. C., Sciaky, D., Barkalow, F., Strong, M., Wyatt, B., Wieggers, J., McMorran, R., Abrar, S., & Mattingly, C. J. (2024). Correction to 'Comparative Toxicogenomics Database's 20th anniversary: update 2025'. (2024). *Nucleic acids research*, gkae1185. Advance online publication. <https://doi.org/10.1093/nar/gkae1185>
- De Cian, A., Delemos, E., Mergny, J. L., Teulade-Fichou, M. P., & Monchaud, D. (2007). Highly efficient G-quadruplex recognition by bisquinolinium compounds. *Journal of the American Chemical Society*, 129(7), 1856–1857. <https://doi.org/10.1021/ja067352b>
- De Magis A, Manzo SG, Russo M, Marinello J, Morigi R, Sordet O, Capranico G. (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc Natl Acad Sci USA*. 116(3):816-825. <https://doi.org/10.1073/pnas.1810409116>
- De Piante, E., D'Aria, F., Napolitano, L.M.R., Amato, J., Pirrello, S., Onesti, S., & Giancola, C., (2023). Exploring the G-quadruplex binding and unwinding activity of the bacterial FeS helicase DinG. *Sci Rep* 13, 12610. <https://doi.org/10.1038/s41598-023-39675-5>.
- Do, N. Q., Lim, K. W., Teo, M. H., Heddi, B., & Phan, A. T. (2011). Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic acids research*, 39(21), 9448–9457. <https://doi.org/10.1093/nar/gkr539>
- Drosopoulos, W. C., Kosiyatrakul, S. T., & Schildkraut, C. L. (2015). BLM helicase facilitates telomere replication during leading strand synthesis of telomeres. *The Journal of cell biology*, 210(2), 191–208. <https://doi.org/10.1083/jcb.201410061>

- Duardo, R. C., Guerra, F., Pepe, S., & Capranico, G. (2023). Non-B DNA structures as a booster of genome instability. *Biochimie*, 214(Pt A), 176–192. <https://doi.org/10.1016/j.biochi.2023.07.002>
- Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & development*, 18(13), 1618–1629. <https://doi.org/10.1101/gad.1200804>
- Estep, K. N., & Brosh, R. M., Jr (2018). RecQ and Fe-S helicases have unique roles in DNA metabolism dictated by their unwinding directionality, substrate specificity, and protein interactions. *Biochemical Society transactions*, 46(1), 77–95. <https://doi.org/10.1042/BST20170044>
- Fairman-Williams, M. E., Guenther, U. P., & Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Current opinion in structural biology*, 20(3), 313–324. <https://doi.org/10.1016/j.sbi.2010.03.011>
- Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K., & Tainer, J. A. (2008). XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations. *Cell*, 133(5), 789–800. <https://doi.org/10.1016/j.cell.2008.04.030>
- Ferretti, S., Hamon, J., de Kanter, R., Scheufler, C., Andraos-Rey, R., Barbe, S., Bechter, E., Blank, J., Bordas, V., Dammassa, E., Decker, A., Di Nanni, N., Dourdoigne, M., Gavioli, E., Hattenberger, M., Heuser, A., Hemmerlin, C., Hinrichs, J., Kerr, G., Laborde, L., Jaco, I., Jiménez Núñez, E., Martus, H., Quadt, C., Cortés-Cros, M., Discovery of WRN inhibitor HRO761 with synthetic lethality in MSI cancers. *Nature* **629**, 443–449 (2024). <https://doi.org/10.1038/s41586-024-07350-y>
- Fry, M., & Loeb, L. A. (1999). Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *The Journal of biological chemistry*, 274(18), 12797–12802. <https://doi.org/10.1074/jbc.274.18.12797>
- García-Muse, T., & Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. *Cell*, 179(3), 604–618. <https://doi.org/10.1016/j.cell.2019.08.055>
- Garcia, P. L., Liu, Y., Jiricny, J., West, S. C., & Janscak, P. (2004). Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *The EMBO journal*, 23(14), 2882–2891. <https://doi.org/10.1038/sj.emboj.7600301>
- Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E., & Hoch, J. A. (1983). Cloning structural gene sacB, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli*. *Journal of bacteriology*, 153(3), 1424–1431. <https://doi.org/10.1128/jb.153.3.1424-1431.1983>
- Ghosh, A., Rossi, M. L., Aulds, J., Croteau, D., & Bohr, V. A. (2009). Telomeric D-loops containing 8-oxo-2'-deoxyguanosine are preferred substrates for Werner and Bloom syndrome helicases and are bound by POT1. *The Journal of biological chemistry*, 284(45), 31074–31084. <https://doi.org/10.1074/jbc.M109.027532>
- Ghosh, A. K., Rossi, M. L., Singh, D. K., Dunn, C., Ramamoorthy, M., Croteau, D. L., Liu, Y., & Bohr, V. A. (2012). RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions

- in telomere maintenance. *The Journal of biological chemistry*, 287(1), 196–209. <https://doi.org/10.1074/jbc.M111.295063>
- Gileadi, O., Burgess-Brown, N. A., Colebrook, S. M., Berridge, G., Savitsky, P., Smee, C. E., Loppnau, P., Johansson, C., Salah, E., & Pantic, N. H. (2008). High throughput production of recombinant human proteins for crystallography. *Methods in molecular biology (Clifton, N.J.)*, 426, 221–246. https://doi.org/10.1007/978-1-60327-058-8_14
- Gray, L. T., Vallur, A. C., Eddy, J., & Maizels, N. (2014). G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nature chemical biology*, 10(4), 313–318. <https://doi.org/10.1038/nchembio.1475>
- Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell*, 97(4), 503–514. [https://doi.org/10.1016/s0092-8674\(00\)80760-6](https://doi.org/10.1016/s0092-8674(00)80760-6)
- Gruber, H. J., Semeraro, M. D., Renner, W., & Herrmann, M. (2021). Telomeres and Age-Related Diseases. *Biomedicines*, 9(10), 1335. <https://doi.org/10.3390/biomedicines9101335>
- Guédin, A., De Cian, A., Gros, J., Lacroix, L., & Mergny, J. L. (2008). Sequence effects in single-base loops for quadruplexes. *Biochimie*, 90(5), 686–696. <https://doi.org/10.1016/j.biochi.2008.01.009>
- Guo, R. B., Rigolet, P., Zargarian, L., Femandjian, S., & Xi, X. G. (2005). Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase. *Nucleic acids research*, 33(10), 3109–3124. <https://doi.org/10.1093/nar/gki619>
- Gupta, R., Sharma, S., Sommers, J. A., Jin, Z., Cantor, S. B., & Brosh, R. M., Jr (2005). Analysis of the DNA substrate specificity of the human BACH1 helicase associated with breast cancer. *The Journal of biological chemistry*, 280(27), 25450–25460. <https://doi.org/10.1074/jbc.M501995200>
- Hänsel-Hertsch, R., Di Antonio, M., & Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature reviews. Molecular cell biology*, 18(5), 279–284. <https://doi.org/10.1038/nrm.2017.3>
- Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols*, 2(8), 1849–1861. <https://doi.org/10.1038/nprot.2007.249>
- Hong, Z., Byrd, A. K., Gao, J., Gao, J., Das, P., Tan, V. Q., Malone, E. G., Osei, B., Marecki, J. C., Protacio, R. U., Wahls, W. P., Raney, K. D., & Song, W. (2024) Eukaryotic Pif1 helicase unwinds G-quadruplex and dsDNA using a conserved wedge. *Nat Commun* 15, 6104. <https://doi.org/10.1038/s41467-024-50575-8>
- Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16, 907-916. <https://doi.org/10.1107/S0365110X63002437>
- Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., & Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. *The Journal of biological chemistry*, 288(42), 30161–30171. <https://doi.org/10.1074/jbc.M113.492504>

- Huber, M. D., Duquette, M. L., Shiels, J. C., & Maizels, N. (2006). A conserved G4 DNA binding domain in RecQ family helicases. *Journal of molecular biology*, 358(4), 1071–1080. <https://doi.org/10.1016/j.jmb.2006.01.077>
- Huppert J. L. (2008). Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. *Molecular bioSystems*, 4(6), 686–691. <https://doi.org/10.1039/b800354h>
- Huppert, J. L., & Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic acids research*, 35(2), 406–413. <https://doi.org/10.1093/nar/gkl1057>
- Iannascoli, C., Palermo, V., Murfuni, I., Franchitto, A., & Pichierri, P. (2015). The WRN exonuclease domain protects nascent strands from pathological MRE11/EXO1-dependent degradation. *Nucleic acids research*, 43(20), 9788–9803. <https://doi.org/10.1093/nar/gkv836>
- Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S., & Johnson, F. B. (2010). Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic acids research*, 38(4), 1114–1122. <https://doi.org/10.1093/nar/gkp1103>
- Kabeche, L., Nguyen, H. D., Buisson, R., & Zou, L. (2018). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science (New York, N.Y.)*, 359(6371), 108–114. <https://doi.org/10.1126/science.aan6490>
- Kaiser, S., Sauer, F., & Kisker, C. (2017). The structural and functional characterization of human RecQ4 reveals insights into its helicase mechanism. *Nature communications*, 8, 15907. <https://doi.org/10.1038/ncomms15907>
- Karow, J. K., Newman, R. H., Freemont, P. S., & Hickson, I. D. (1999). Oligomeric ring structure of the Bloom's syndrome helicase. *Current biology : CB*, 9(11), 597–600. [https://doi.org/10.1016/s0960-9822\(99\)80264-4](https://doi.org/10.1016/s0960-9822(99)80264-4)
- Keller, H., Kiosze, K., Sachsenweger, J., Haumann, S., Ohlenschläger, O., Nuutinen, T., Syväoja, J. E., Görlach, M., Grosse, F., & Pospiech, H. (2014). The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. *Nucleic acids research*, 42(20), 12614–12627. <https://doi.org/10.1093/nar/gku993>
- Kerkour, A., Marquevielle, J., Ivashchenko, S., Yatsunyk, L. A., Mergny, J. L., & Salgado, G. F. (2017). High-resolution three-dimensional NMR structure of the KRAS proto-oncogene promoter reveals key features of a G-quadruplex involved in transcriptional regulation. *The Journal of biological chemistry*, 292(19), 8082–8091. <https://doi.org/10.1074/jbc.M117.781906>
- Ketkar, A., Voehler, M., Mukiza, T., & Eoff, R. L. (2017). Residues in the RecQ C-terminal Domain of the Human Werner Syndrome Helicase Are Involved in Unwinding G-quadruplex DNA. *The Journal of biological chemistry*, 292(8), 3154–3163. <https://doi.org/10.1074/jbc.M116.767699>
- Khadka, P., Croteau, D. L., & Bohr, V. A. (2016). RECQL5 has unique strand annealing properties relative to the other human RecQ helicase proteins. *DNA repair*, 37, 53–66. <https://doi.org/10.1016/j.dnarep.2015.11.005>
- Kitano, K., Kim, S. Y., & Hakoshima, T. (2010). Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure (London, England: 1993)*, 18(2), 177–187. <https://doi.org/10.1016/j.str.2009.12.011>

- Kotsantis, P., Segura-Bayona, S., Margalef, P., Marzec, P., Ruis, P., Hewitt, G., Bellelli, R., Patel, H., Goldstone, R., Poetsch, A. R., & Boulton, S. J. (2020). RTEL1 Regulates G4/R-Loops to Avert Replication-Transcription Collisions. *Cell reports*, 33(12), 108546. <https://doi.org/10.1016/j.celrep.2020.108546>
- Kruisselbrink, E., Guryev, V., Brouwer, K., Pontier, D. B., Cuppen, E., & Tijsterman, M. (2008). Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C. elegans*. *Current biology : CB*, 18(12), 900–905. <https://doi.org/10.1016/j.cub.2008.05.013>
- Kusumoto, R., Dawut, L., Marchetti, C., Wan Lee, J., Vindigni, A., Ramsden, D., & Bohr, V. A. (2008). Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry*, 47(28), 7548–7556. <https://doi.org/10.1021/bi702325t>
- Kusumoto, R., Ghosh, D., Karmakar, P., May, A., Ramsden, D., & Bohr, V. A. (2014). Serines 440 and 467 in the Werner syndrome protein are phosphorylated by DNA-PK and affects its dynamics in response to DNA double strand breaks. *Aging*, 6(1), 70–81. <https://doi.org/10.18632/aging.100629>
- Lago, S., Tosoni, E., Nadai, M., Palumbo, M., & Richter, S. N. (2017). The cellular protein nucleolin preferentially binds long-looped G-quadruplex nucleic acids. *Biochimica et biophysica acta. General subjects*, 1861(5 Pt B), 1371–1381. <https://doi.org/10.1016/j.bbagen.2016.11.036>
- Laud, P. R., Multani, A. S., Bailey, S. M., Wu, L., Ma, J., Kingsley, C., Lebel, M., Pathak, S., DePinho, R. A., & Chang, S. (2005). Elevated telomere-telomere recombination in WRN-deficient, telomere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway. *Genes & development*, 19(21), 2560–2570. <https://doi.org/10.1101/gad.1321305>
- Le Guen, T., Jullien, L., Touzot, F., Schertzer, M., Gaillard, L., Perderiset, M., Carpentier, W., Nitschke, P., Picard, C., Couillaud, G., Soulier, J., Fischer, A., Callebaut, I., Jabado, N., Londono-Vallejo, A., de Villartay, J. P., & Revy, P. (2013). Human RTEL1 deficiency causes Hoyer-Hreidarsson syndrome with short telomeres and genome instability. *Human molecular genetics*, 22(16), 3239–3249. <https://doi.org/10.1093/hmg/ddt178>
- Lee, C. Y., McNerney, C., Ma, K., Zhao, W., Wang, A., & Myong, S. (2020). R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nature communications*, 11(1), 3392. <https://doi.org/10.1038/s41467-020-17176-7>
- Li, T., Zhang, M., Li, Y., Han, X., Tang, L., Ma, T., Zhao, X., Zhao, R., Wang, Y., Bai, X., Zhang, K., Geng, X., Sui, L., Feng, X., Zhang, Q., Zhao, Y., Liu, Y., Stewart, J. A., & Wang, F. (2023). Cooperative interaction of CST and RECQ4 resolves G-quadruplexes and maintains telomere stability. *EMBO reports*, 24(9), e55494. <https://doi.org/10.15252/embr.202255494>
- Liano, D., Monti, L., Chowdhury, S., Raguseo, F., & Di Antonio, M. (2022). Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chemical communications (Cambridge, England)*, 58(92), 12753–12762. <https://doi.org/10.1039/d2cc04872h>
- Lim, K. W., Alberti, P., Guédin, A., Lacroix, L., Riou, J. F., Royle, N. J., Mergny, J. L., & Phan, A. T. (2009). Sequence variant (CTAGGG)_n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. *Nucleic acids research*, 37(18), 6239–6248. <https://doi.org/10.1093/nar/gkp630>

- Liu, H., Rudolf, J., Johnson, K. A., McMahon, S. A., Oke, M., Carter, L., McRobbie, A. M., Brown, S. E., Naismith, J. H., & White, M. F. (2008). Structure of the DNA repair helicase XPD. *Cell*, 133(5), 801–812. <https://doi.org/10.1016/j.cell.2008.04.029>
- Liu, J. Q., Chen, C. Y., Xue, Y., Hao, Y. H., & Tan, Z. (2010). G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. *Journal of the American Chemical Society*, 132(30), 10521–10527. <https://doi.org/10.1021/ja1038165>
- Liu, N. N., Song, Z. Y., Guo, H. L., Yin, H., Chen, W. F., Dai, Y. X., Xin, B. G., Ai, X., Ji, L., Wang, Q. M., Hou, X. M., Dou, S. X., Rety, S., & Xi, X. G. (2021). Endogenous Bos taurus RECQL is predominantly monomeric and more active than oligomers. *Cell reports*, 36(10), 109688. <https://doi.org/10.1016/j.celrep.2021.109688>
- Lohman, T. M., & Bjornson, K. P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annual review of biochemistry*, 65, 169–214. <https://doi.org/10.1146/annurev.bi.65.070196.001125>
- London, T. B., Barber, L. J., Mosedale, G., Kelly, G. P., Balasubramanian, S., Hickson, I. D., Boulton, S. J., & Hiom, K. (2008). FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *The Journal of biological chemistry*, 283(52), 36132–36139. <https://doi.org/10.1074/jbc.M808152200>
- Lu, H., & Davis, A. J. (2021). Human RecQ Helicases in DNA Double-Strand Break Repair. *Frontiers in cell and developmental biology*, 9, 640755. <https://doi.org/10.3389/fcell.2021.640755>
- Lu, H., Shamanna, R. A., Keijzers, G., Anand, R., Rasmussen, L. J., Cejka, P., Croteau, D. L., & Bohr, V. A. (2016a). RECQL4 Promotes DNA End Resection in Repair of DNA Double-Strand Breaks. *Cell reports*, 16(1), 161–173. <https://doi.org/10.1016/j.celrep.2016.05.079>
- Lu, X., Parvathaneni, S., Li, X. L., Lal, A., & Sharma, S. (2016b). Transcriptome guided identification of novel functions of RECQ1 helicase. *Methods (San Diego, Calif.)*, 108, 111–117. <https://doi.org/10.1016/j.ymeth.2016.04.018>
- Lucic, B., Zhang, Y., King, O., Mendoza-Maldonado, R., Berti, M., Niesen, F. H., Burgess-Brown, N. A., Pike, A. C., Cooper, C. D., 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 research*, 39(5), 1703–1717. <https://doi.org/10.1093/nar/gkq1031>
- Luo, Y., Granzhan, A., Marquevielle, J., Cucchiari, A., Lacroix, L., Amrane, S., Verga, D., & Mergny, J. L. (2023). Guidelines for G-quadruplexes: I. In vitro characterization. *Biochimie*, 214(Pt A), 5–23. <https://doi.org/10.1016/j.biochi.2022.12.019>
- Manthei, K. A., Hill, M. C., Burke, J. E., Butcher, S. E., & Keck, J. L. (2015). Structural mechanisms of DNA binding and unwinding in bacterial RecQ helicases. *Proceedings of the National Academy of Sciences of the United States of America*, 112(14), 4292–4297. <https://doi.org/10.1073/pnas.1416746112>
- 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. *Scientific reports*, 6, 21501. <https://doi.org/10.1038/srep21501>

- Martins, D. J., Di Lazzaro Filho, R., Bertola, D. R., & Hoch, N. C. (2023). Rothmund-Thomson syndrome, a disorder far from solved. *Frontiers in aging*, 4, 1296409. <https://doi.org/10.3389/fragi.2023.1296409>
- Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y., & Takisawa, H. (2006). The N-terminal 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(13), 4843–4852. <https://doi.org/10.1128/MCB.02267-05>
- McRobbie, A. M., Meyer, B., Rouillon, C., Petrovic-Stojanovska, B., Liu, H., & White, M. F. (2012). *Staphylococcus aureus* DinG, a helicase that has evolved into a nuclease. *The Biochemical journal*, 442(1), 77–84. <https://doi.org/10.1042/BJ20111903>
- Mendez-Bermudez, A., Hidalgo-Bravo, A., Cotton, V. E., Gravani, A., Jeyapalan, J. N., & Royle, N. J. (2012). The roles of WRN and BLM RecQ helicases in the Alternative Lengthening of Telomeres. *Nucleic acids research*, 40(21), 10809–10820. <https://doi.org/10.1093/nar/gks862>
- Mendoza, O., Gueddouda, N. M., Boulé, J. B., Bourdoncle, A., & Mergny, J. L. (2015). A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands. *Nucleic acids research*, 43(11), e71. <https://doi.org/10.1093/nar/gkv193>
- Mendoza, O., Bourdoncle, A., Boulé, J. B., Brosh, R. M., Jr, & Mergny, J. L. (2016). G-quadruplexes and helicases. *Nucleic acids research*, 44(5), 1989–2006. <https://doi.org/10.1093/nar/gkw079>
- Miglietta, G., Russo, M., & Capranico, G. (2020). G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic acids research*, 48(21), 11942–11957. <https://doi.org/10.1093/nar/gkaa944>
- Miglietta, G., Marinello, J., Russo, M., & Capranico, G. (2022). Ligands stimulating antitumour immunity as the next G-quadruplex challenge. *Molecular cancer*, 21(1), 180. <https://doi.org/10.1186/s12943-022-01649-y>
- Mishra, S. K., Tawani, A., Mishra, A., & Kumar, A. (2016). G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Scientific reports*, 6, 38144. <https://doi.org/10.1038/srep38144>
- Mojumdar A. (2020). Mutations in conserved functional domains of human RecQ helicases are associated with diseases and cancer: A review. *Biophysical chemistry*, 265, 106433. <https://doi.org/10.1016/j.bpc.2020.106433>
- Müller, S., Sanders, D. A., Di Antonio, M., Matsis, S., Riou, J. F., Rodriguez, R., & Balasubramanian, S. (2012). Pyridostatin analogues promote telomere dysfunction and long-term growth inhibition in human cancer cells. *Organic & biomolecular chemistry*, 10(32), 6537–6546. <https://doi.org/10.1039/c2ob25830g>
- Murat, P. and Balasubramanian, S. (2014) Existence and consequences of G-quadruplex structures in DNA. *Curr. Opin. Genet. Dev.* 25, 22–29 doi:10.1016/j.gde.2013.10.012
- Muzzolini, L., Beuron, F., Patwardhan, A., Popuri, V., Cui, S., Niccolini, B., Rappas, M., Freemont, P. S., & Vindigni, A. (2007). Different quaternary structures of human RECQ1 are associated with its dual enzymatic activity. *PLoS biology*, 5(2), e20. <https://doi.org/10.1371/journal.pbio.0050020>

- Nakayama, H., Nakayama, K., Nakayama, R., Irino, N., Nakayama, Y., & Hanawalt, P. C. (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. *Molecular & general genetics*, *MGG*, *195*(3), 474–480. <https://doi.org/10.1007/BF00341449>
- Newman, J. A., & Gileadi, O. (2020). RecQ helicases in DNA repair and cancer targets. *Essays in biochemistry*, *64*(5), 819–830. <https://doi.org/10.1042/EBC20200012>
- Newman, J. A., Savitsky, P., Allerston, C. K., Bizard, A. H., Özer, Ö., Sarlós, K., Liu, Y., Pardon, E., Steyaert, J., Hickson, I. D., & Gileadi, O. (2015). Crystal structure of the Bloom's syndrome helicase indicates a role for the HRDC domain in conformational changes. *Nucleic acids research*, *43*(10), 5221–5235. <https://doi.org/10.1093/nar/gkv373>
- Newman, J. A., Aitkenhead, H., Savitsky, P., & Gileadi, O. (2017). Insights into the RecQ helicase mechanism revealed by the structure of the helicase domain of human RECQL5. *Nucleic acids research*, *45*(7), 4231–4243. <https://doi.org/10.1093/nar/gkw1362>
- Newman, J. A., Gavard, A. E., Lieb, S., Ravichandran, M. C., Hauer, K., Werni, P., Geist, L., Böttcher, J., Engen, J. R., Rumpel, K., Samwer, M., Petronczki, M., & Gileadi, O. (2020). Structure of the helicase core of Werner helicase, a key target in microsatellite instability cancers. *Life science alliance*, *4*(1), e202000795. <https://doi.org/10.26508/lsa.202000795>
- Nguyen, G. H., Dexheimer, T. S., Rosenthal, A. S., Chu, W. K., Singh, D. K., Mosedale, G., Bachrati, C. Z., Schultz, L., Sakurai, M., Savitsky, P., Abu, M., McHugh, P. J., Bohr, V. A., Harris, C. C., Jadhav, A., Gileadi, O., Maloney, D. J., Simeonov, A., & Hickson, I. D. (2013). A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chemistry & biology*, *20*(1), 55–62. <https://doi.org/10.1016/j.chembiol.2012.10.016>
- Nguyen, G. H., Tang, W., Robles, A. I., Beyer, R. P., Gray, L. T., Welsh, J. A., Schetter, A. J., Kumamoto, K., Wang, X. W., Hickson, I. D., Maizels, N., Monnat, R. J., Jr, & Harris, C. C. (2014). Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(27), 9905–9910. <https://doi.org/10.1073/pnas.1404807111>
- Niehhs, C., & Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. *Nature reviews. Molecular cell biology*, *21*(3), 167–178. <https://doi.org/10.1038/s41580-019-0206-3>
- Nimonkar, A. V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., Modrich, P., & Kowalczykowski, S. C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & development*, *25*(4), 350–362. <https://doi.org/10.1101/gad.2003811>
- Ohlenschläger, O., Kuhnert, A., Schneider, A., Haumann, S., Bellstedt, P., Keller, H., Saluz, H. P., 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 research*, *40*(17), 8309–8324. <https://doi.org/10.1093/nar/gks591>
- Opresko, P. L., Mason, P. A., Podell, E. R., Lei, M., Hickson, I. D., Cech, T. R., & Bohr, V. A. (2005). POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *The Journal of biological chemistry*, *280*(37), 32069–32080. <https://doi.org/10.1074/jbc.M505211200>

- Opresko, P. L., Sowd, G., & Wang, H. (2009). The Werner syndrome helicase/exonuclease processes mobile D-loops through branch migration and degradation. *PloS one*, 4(3), e4825. <https://doi.org/10.1371/journal.pone.0004825>
- Oshima, J., Sidorova, J. M., & Monnat, R. J., Jr (2017). Werner syndrome: Clinical features, pathogenesis and potential therapeutic interventions. *Ageing research reviews*, 33, 105–114. <https://doi.org/10.1016/j.arr.2016.03.002>
- Ovejero, S., Viziteu, E., Dutrieux, L., Devin, J., Lin, Y. L., Alaterre, E., Jourdan, M., Basbous, J., Requirand, G., Robert, N., de Boussac, H., Seckinger, A., Hose, D., Vincent, L., Herbaux, C., Constantinou, A., Pasero, P., & Moreaux, J. (2022). The BLM helicase is a new therapeutic target in multiple myeloma involved in replication stress survival and drug resistance. *Frontiers in immunology*, 13, 983181. <https://doi.org/10.3389/fimmu.2022.983181>
- Paeschke, K., Capra, J. A., & Zakian, V. A. (2011). DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell*, 145(5), 678–691. <https://doi.org/10.1016/j.cell.2011.04.015>
- Pagano, B., Fotticchia, I., De Tito, S., Mattia, C. A., Mayol, L., Novellino, E., Randazzo, A., & Giancola, C. (2010). Selective Binding of Distamycin A Derivative to G-Quadruplex Structure [d(TGGGGT)](4). *Journal of nucleic acids*, 2010, 247137. <https://doi.org/10.4061/2010/247137>
- Paliwal, S., Kanagaraj, R., Sturzenegger, A., Burdova, K., & Janscak, P. (2014). Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing. *Nucleic acids research*, 42(4), 2380–2390. <https://doi.org/10.1093/nar/gkt1263>
- Papageorgiou, A.C., Pospisilova, M., Cibulka, J., Ashraf, R., Waudby, C. A., Kadeřávek, P., Maroz, V., Kubicek, K., Prokop, Z., Krejci L., & Tripsianes, K., (2023). Recognition and coacervation of G-quadruplexes by a multifunctional disordered region in RECQ4 helicase. *Nat Commun* 14, 6751. <https://doi.org/10.1038/s41467-023-42503-z>
- Parekh, V.J., Węgrzyn, G., Arluison, V., Sinden, R.R. (2023). Genomic Instability of G-Quadruplex Sequences in *Escherichia coli*: Roles of DinG, RecG, and RecQ Helicases. *Genes*, 14, 1720. <https://doi.org/10.3390/genes14091720>
- Patel, S. S., & Donmez, I. (2006). Mechanisms of helicases. *The Journal of biological chemistry*, 281(27), 18265–18268. <https://doi.org/10.1074/jbc.R600008200>
- Paudel, B. P., Moye, A. L., Abou Assi, H., El-Khoury, R., Cohen, S. B., Holien, J. K., Birrento, M. L., Samosorn, S., Intharapichai, K., Tomlinson, C. G., Teulade-Fichou, M. P., González, C., Beck, J. L., Damha, M. J., van Oijen, A. M., & Bryan, T. M. (2020). A mechanism for the extension and unfolding of parallel telomeric G-quadruplexes by human telomerase at single-molecule resolution. *eLife*, 9, e56428. <https://doi.org/10.7554/eLife.56428>
- Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S., & Patel, D. J. (2007). Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *Journal of the American Chemical Society*, 129(14), 4386–4392. <https://doi.org/10.1021/ja068739h>
- Piazza, A., Shah, S. S., Wright, W. D., Gore, S. K., Koszul, R., & Heyer, W. D. (2019). Dynamic Processing of Displacement Loops during Recombinational DNA Repair. *Molecular cell*, 73(6), 1255–1266.e4. <https://doi.org/10.1016/j.molcel.2019.01.005>

- Pike, A. C., 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. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4), 1039–1044. <https://doi.org/10.1073/pnas.0806908106>
- Pike, A. C., 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. *Proceedings of the National Academy of Sciences of the United States of America*, 112(14), 4286–4291. <https://doi.org/10.1073/pnas.1417594112>
- Pisani, F. M., Napolitano, E., Napolitano, L. M. R., & Onesti, S. (2018). Molecular and Cellular Functions of the Warsaw Breakage Syndrome DNA Helicase DDX11. *Genes*, 9(11), 564. <https://doi.org/10.3390/genes9110564>
- Pohjoismäki, J. L., Holmes, J. B., Wood, S. R., Yang, M. Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., Rigby, R. E., Jackson, A. P., Spelbrink, J. N., Griffith, J. D., Crouch, R. J., Jacobs, H. T., & Holt, I. J. (2010). Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *Journal of molecular biology*, 397(5), 1144–1155. <https://doi.org/10.1016/j.jmb.2010.02.029>
- Popuri, V., Bachrati, C. Z., Muzzolini, L., Mosedale, G., Costantini, S., Giacomini, E., Hickson, I. D., & Vindigni, A. (2008). The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *The Journal of biological chemistry*, 283(26), 17766–17776. <https://doi.org/10.1074/jbc.M709749200>
- Popuri, V., Ramamoorthy, M., Tadokoro, T., Singh, D. K., Karmakar, P., Croteau, D. L., & Bohr, V. A. (2012a). Recruitment and retention dynamics of RECQL5 at DNA double strand break sites. *DNA repair*, 11(7), 624–635. <https://doi.org/10.1016/j.dnarep.2012.05.001>
- Popuri, V., Croteau, D. L., Brosh, R. M., Jr, & Bohr, V. A. (2012b). RECQ1 is required for cellular resistance to replication stress and catalyzes strand exchange on stalled replication fork structures. *Cell cycle (Georgetown, Tex.)*, 11(22), 4252–4265. <https://doi.org/10.4161/cc.22581>
- Popuri, V., Hsu, J., Khadka, P., Horvath, K., Liu, Y., Croteau, D. L., & Bohr, V. A. (2014). Human RECQL1 participates in telomere maintenance. *Nucleic acids research*, 42(9), 5671–5688. <https://doi.org/10.1093/nar/gku200>
- Puig Lombardi, E., & Londoño-Vallejo, A. (2020). A guide to computational methods for G-quadruplex prediction. *Nucleic acids research*, 48(1), 1–15. <https://doi.org/10.1093/nar/gkz1097>
- Pyle A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annual review of biophysics*, 37, 317–336. <https://doi.org/10.1146/annurev.biophys.37.032807.125908>
- Ren, H., Dou, S. X., Zhang, X. D., Wang, P. Y., Kanagaraj, R., Liu, J. L., Janscak, P., Hu, J. S., & Xi, X. G. (2008). The zinc-binding motif of human RECQ5beta suppresses the intrinsic strand-annealing activity of its DExH helicase domain and is essential for the helicase activity of the enzyme. *The Biochemical journal*, 412(3), 425–433. <https://doi.org/10.1042/BJ20071150>
- Ren, B., Duan, X., & Ding, H. (2009). Redox control of the DNA damage-inducible protein DinG helicase activity via its iron-sulfur cluster. *The Journal of biological chemistry*, 284(8), 4829–4835. <https://doi.org/10.1074/jbc.M807943200>

- Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, 43(18), 8627–8637. <https://doi.org/10.1093/nar/gkv862>
- Ribeyre, C., Lopes, J., Boulé, J. B., Piazza, A., Guédin, A., Zakian, V. A., Mergny, J. L., & Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. *PLoS genetics*, 5(5), e1000475. <https://doi.org/10.1371/journal.pgen.1000475>
- Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D., & Di Antonio, M. (2021). DNA G-quadruplex structures: more than simple roadblocks to transcription?. *Nucleic acids research*, 49(15), 8419–8431. <https://doi.org/10.1093/nar/gkab609>
- Rossi, M. L., Ghosh, A. K., Kulikowicz, T., Croteau, D. L., & Bohr, V. A. (2010). Conserved helicase domain of human RecQ4 is required for strand annealing-independent DNA unwinding. *DNA repair*, 9(7), 796–804. <https://doi.org/10.1016/j.dnarep.2010.04.003>
- Rudolf, J., Makrantonis, V., Ingledew, W. J., Stark, M. J., & White, M. F. (2006). The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Molecular cell*, 23(6), 801–808. <https://doi.org/10.1016/j.molcel.2006.07.019>
- San Filippo, J., Sung, P., & Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry*, 77, 229–257. <https://doi.org/10.1146/annurev.biochem.77.061306.125255>
- Sarkies, P., Reams, C., Simpson, L. J., & Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. *Molecular cell*, 40(5), 703–713. <https://doi.org/10.1016/j.molcel.2010.11.009>
- Sarkies, P., Murat, P., Phillips, L. G., Patel, K. J., Balasubramanian, S., & Sale, J. E. (2012). FANCI coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic acids research*, 40(4), 1485–1498. <https://doi.org/10.1093/nar/gkr868>
- Sauer, M., & Paeschke, K. (2017). G-quadruplex unwinding helicases and their function in vivo. *Biochemical Society transactions*, 45(5), 1173–1182. <https://doi.org/10.1042/BST20170097>
- Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M. N., Balasubramanian, S., & Sale, J. E. (2014). Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *The EMBO journal*, 33(21), 2507–2520. <https://doi.org/10.15252/emj.201488398>
- Schurman, S. H., Hedayati, M., Wang, Z., Singh, D. K., Speina, E., Zhang, Y., Becker, K., Macris, M., Sung, P., Wilson, D. M., 3rd, Croteau, D. L., & Bohr, V. A. (2009). Direct and indirect roles of RECQL4 in modulating base excision repair capacity. *Human molecular genetics*, 18(18), 3470–3483. <https://doi.org/10.1093/hmg/ddp291>
- Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>

- Shamanna, R. A., Singh, D. K., Lu, H., Mirey, G., Keijzers, G., Salles, B., Croteau, D. L., & Bohr, V. A. (2014). RECQ helicase RECQL4 participates in non-homologous end joining and interacts with the Ku complex. *Carcinogenesis*, 35(11), 2415–2424. <https://doi.org/10.1093/carcin/bgu137>
- Sharma, S., & Brosh, R. M., Jr (2007). Human RECQ1 is a DNA damage responsive protein required for genotoxic stress resistance and suppression of sister chromatid exchanges. *PloS one*, 2(12), e1297. <https://doi.org/10.1371/journal.pone.0001297>
- Sharma, S., Sommers, J. A., Choudhary, S., Faulkner, J. K., Cui, S., Andreoli, L., Muzzolini, L., Vindigni, A., & Brosh, R. M., Jr (2005). Biochemical analysis of the DNA unwinding and strand annealing activities catalyzed by human RECQ1. *The Journal of biological chemistry*, 280(30), 28072–28084. <https://doi.org/10.1074/jbc.M500264200>
- Sharma, S., Doherty, K. M., & Brosh, R. M., Jr (2006). Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. *The Biochemical journal*, 398(3), 319–337. <https://doi.org/10.1042/BJ20060450>
- Sharma, S., Phatak, P., Stortchevoi, A., Jasin, M., & Larocque, J. R. (2012). RECQ1 plays a distinct role in cellular response to oxidative DNA damage. *DNA repair*, 11(6), 537–549. <https://doi.org/10.1016/j.dnarep.2012.04.003>
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7, 539. <https://doi.org/10.1038/msb.2011.75>
- Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry*, 76, 23–50. <https://doi.org/10.1146/annurev.biochem.76.052305.115300>
- Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*, 42(6), 794–805. <https://doi.org/10.1016/j.molcel.2011.04.026>
- Smestad, J. A., & Maher, L. J., 3rd (2015). Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. *BMC medical genetics*, 16, 91. <https://doi.org/10.1186/s12881-015-0236-4>
- Sollier, J., & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. *Trends in cell biology*, 25(9), 514–522. <https://doi.org/10.1016/j.tcb.2015.05.003>
- Stavropoulos, D. J., Bradshaw, P. S., Li, X., Pasic, I., Truong, K., Ikura, M., Ungrin, M., & Meyn, M. S. (2002). The Bloom syndrome helicase BLM interacts with TRF2 in ALT cells and promotes telomeric DNA synthesis. *Human molecular genetics*, 11(25), 3135–3144. <https://doi.org/10.1093/hmg/11.25.3135>
- Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P., & Janscak, P. (2014). DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *The Journal of biological chemistry*, 289(39), 27314–27326. <https://doi.org/10.1074/jbc.M114.578823>

- Suhasini, A. N., & Brosh, R. M., Jr (2013). Disease-causing missense mutations in human DNA helicase disorders. *Mutation research*, 752(2), 138–152. <https://doi.org/10.1016/j.mrrev.2012.12.004>
- Swan, M. K., Legris, V., Tanner, A., Reaper, P. M., Vial, S., Bordas, R., Pollard, J. R., Charlton, P. A., Golec, J. M., & Bertrand, J. A. (2014). Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA. *Acta crystallographica. Section D, Biological crystallography*, 70(Pt 5), 1465–1475. <https://doi.org/10.1107/S139900471400501X>
- Tadokoro, T., Ramamoorthy, M., Popuri, V., May, A., Tian, J., Sykora, P., Rybanska, I., Wilson, D. M., 3rd, Croteau, D. L., & Bohr, V. A. (2012). Human RECQL5 participates in the removal of endogenous DNA damage. *Molecular biology of the cell*, 23(21), 4273–4285. <https://doi.org/10.1091/mbc.E12-02-0110>
- Tan, J., Wang, X., Phoon, L., Yang, H., & Lan, L. (2020). Resolution of ROS-induced G-quadruplexes and R-loops at transcriptionally active sites is dependent on BLM helicase. *FEBS letters*, 594(9), 1359–1367. <https://doi.org/10.1002/1873-3468.13738>
- Tang, W., Robles, A. I., Beyer, R. P., Gray, L. T., Nguyen, G. H., Oshima, J., Maizels, N., Harris, C. C., & Monnat, R. J., Jr (2016). The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Human molecular genetics*, 25(10), 2060–2069. <https://doi.org/10.1093/hmg/ddw079>
- Teng, F. Y., Jiang, Z. Z., Guo, M., Tan, X. Z., Chen, F., Xi, X. G., & Xu, Y. (2021). G-quadruplex DNA: a novel target for drug design. *Cellular and molecular life sciences: CMLS*, 78(19-20), 6557–6583. <https://doi.org/10.1007/s00018-021-03921-8>
- Thakur, R. S., Desingu, A., Basavaraju, S., Subramanya, S., Rao, D. N., & Nagaraju, G. (2014). Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA: implications for targeting G4 DNA as a novel therapeutic approach. *The Journal of biological chemistry*, 289(36), 25112–25136. <https://doi.org/10.1074/jbc.M114.563569>
- Thangavel, S., Mendoza-Maldonado, R., Tissino, E., Sidorova, J. M., Yin, J., Wang, W., Monnat, R. J., Jr, Falaschi, A., & Vindigni, A. (2010). Human RECQ1 and RECQ4 helicases play distinct roles in DNA replication initiation. *Molecular and cellular biology*, 30(6), 1382–1396. <https://doi.org/10.1128/MCB.01290-09>
- Thangavel, S., Berti, M., Levikova, M., Pinto, C., Gomathinayagam, S., Vujanovic, M., Zellweger, R., Moore, H., Lee, E. H., Hendrickson, E. A., Cejka, P., Stewart, S., Lopes, M., & Vindigni, A. (2015). DNA2 drives processing and restart of reversed replication forks in human cells. *The Journal of cell biology*, 208(5), 545–562. <https://doi.org/10.1083/jcb.201406100>
- Tian, Y., Wang, W., Lautrup, S., Zhao, H., Li, X., Law, P. W. N., Dinh, N. D., Fang, E. F., Cheung, H. H., & Chan, W. Y. (2022). WRN promotes bone development and growth by unwinding SHOX-G-quadruplexes via its helicase activity in Werner Syndrome. *Nature communications*, 13(1), 5456. <https://doi.org/10.1038/s41467-022-33012-6>
- Tippana, R., Hwang, H.J., Opresko, P.L., Bohr, V.A., & Myong, S. (2016). Single-molecule imaging reveals a common mechanism shared by G-quadruplex-resolving helicases. *Proceedings of the National Academy of Sciences*, 113, 8448 - 8453. <https://doi.org/10.1073/pnas.1603724113>

- Urban, V., Dobrovolna, J., Hühn, D., Fryzelkova, J., Bartek, J., & Janscak, P. (2016). RECQ5 helicase promotes resolution of conflicts between replication and transcription in human cells. *The Journal of cell biology*, 214(4), 401–415. <https://doi.org/10.1083/jcb.201507099>
- Uringa, E. J., Youds, J. L., Lisaingo, K., Lansdorp, P. M., & Boulton, S. J. (2011). RTEL1: an essential helicase for telomere maintenance and the regulation of homologous recombination. *Nucleic acids research*, 39(5), 1647–1655. <https://doi.org/10.1093/nar/gkq1045>
- van Wietmarschen, N., Merzouk, S., Halsema, N., Spierings, D. C. J., Guryev, V., & Lansdorp, P. M. (2018). BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nature communications*, 9(1), 271. <https://doi.org/10.1038/s41467-017-02760-1>
- Varshney, D., Spiegel, J., Zyner, K., Tannahill, D., & Balasubramanian, S. (2020). The regulation and functions of DNA and RNA G-quadruplexes. *Nature reviews. Molecular cell biology*, 21(8), 459–474. <https://doi.org/10.1038/s41580-020-0236-x>
- Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., & Wigley, D. B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell*, 97(1), 75–84. [https://doi.org/10.1016/s0092-8674\(00\)80716-3](https://doi.org/10.1016/s0092-8674(00)80716-3)
- Vindigni, A., Marino, F., & Gileadi, O. (2010). Probing the structural basis of RecQ helicase function. *Biophysical chemistry*, 149(3), 67–77. <https://doi.org/10.1016/j.bpc.2010.03.012>
- Voloshin, O. N., & Camerini-Otero, R. D. (2007). The DinG protein from Escherichia coli is a structure-specific helicase. *The Journal of biological chemistry*, 282(25), 18437–18447. <https://doi.org/10.1074/jbc.M700376200>
- Voloshin, O. N., Vanevski, F., Khil, P. P., & Camerini-Otero, R. D. (2003). Characterization of the DNA damage-inducible helicase DinG from Escherichia coli. *The Journal of biological chemistry*, 278(30), 28284–28293. <https://doi.org/10.1074/jbc.M301188200>
- von Kobbe, C., Thomä, N. H., Czyzewski, B. K., Pavletich, N. P., & Bohr, V. A. (2003). Werner syndrome protein contains three structure-specific DNA binding domains. *The Journal of biological chemistry*, 278(52), 52997–53006. <https://doi.org/10.1074/jbc.M308338200>
- Voter, A. F., Qiu, Y., Tippana, R., Myong, S., & Keck, J. L. (2018). A guanine-flipping and sequestration mechanism for G-quadruplex unwinding by RecQ helicases. *Nature communications*, 9(1), 4201. <https://doi.org/10.1038/s41467-018-06751-8>
- Watson, J. D., & Crick F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. <https://doi.org/10.1038/171737a0>
- Wu Y. (2012). Unwinding and rewinding: double faces of helicase?. *Journal of nucleic acids*, 2012, 140601. <https://doi.org/10.1155/2012/140601>
- Wu, L., & Hickson, I. D. (2003). The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature*, 426(6968), 870–874. <https://doi.org/10.1038/nature02253>
- Wu, L., Chan, K. L., Ralf, C., Bernstein, D. A., Garcia, P. L., Bohr, V. A., Vindigni, A., Janscak, P., Keck, J. L., & Hickson, I. D. (2005). The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *The EMBO journal*, 24(14), 2679–2687. <https://doi.org/10.1038/sj.emboj.7600740>

- Wu, W. Q., Hou, X. M., Li, M., Dou, S. X., & Xi, X. G. (2015). BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic acids research*, 43(9), 4614–4626. <https://doi.org/10.1093/nar/gkv361>
- Wu, W. Q., Hou, X. M., Zhang, B., Fossé, P., René, B., Mauffret, O., Li, M., Dou, S. X., & Xi, X. G. (2017). Single-molecule studies reveal reciprocating of WRN helicase core along ssDNA during DNA unwinding. *Scientific reports*, 7, 43954. <https://doi.org/10.1038/srep43954>
- Wu, W., Bhowmick, R., Vogel, I., Özer, Ö., Ghisays, F., Thakur, R. S., Sanchez de Leon, E., Richter, P. H., Ren, L., Petrini, J. H., Hickson, I. D., & Liu, Y. (2020). RTEL1 suppresses G-quadruplex-associated R-loops at difficult-to-replicate loci in the human genome. *Nature structural & molecular biology*, 27(5), 424–437. <https://doi.org/10.1038/s41594-020-0408-6>
- Wulfridge, P., & Sarma, K. (2024). Intertwining roles of R-loops and G-quadruplexes in DNA repair, transcription and genome organization. *Nature cell biology*, 26(7), 1025–1036. <https://doi.org/10.1038/s41556-024-01437-4>
- Xu, X., Rochette, P. J., Feyissa, E. A., Su, T. V., & Liu, Y. (2009). MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication. *The EMBO journal*, 28(19), 3005–3014. <https://doi.org/10.1038/emboj.2009.235>
- Zhang, X., & Wigley, D. B. (2008). The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nature structural & molecular biology*, 15(11), 1223–1227. <https://doi.org/10.1038/nsmb.1501>
- Zhang, JM., & Zou, L. (2020). Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. *Cell Biosci* 10, 30. <https://doi.org/10.1186/s13578-020-00391-6>
- Zhou, H., Hintze, B. J., Kimsey, I. J., Sathyamoorthy, B., Yang, S., Richardson, J. S., & Al-Hashimi, H. M. (2015). New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic acids research*, 43(7), 3420–3433. <https://doi.org/10.1093/nar/gkv241>
- Zhou, G., Liu, X., Li, Y., Xu, S., Ma, C., Wu, X., Cheng, Y., Yu, Z., Zhao, G., & Chen, Y. (2016). Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget*, 7(12), 14925–14939. <https://doi.org/10.18632/oncotarget.7483>