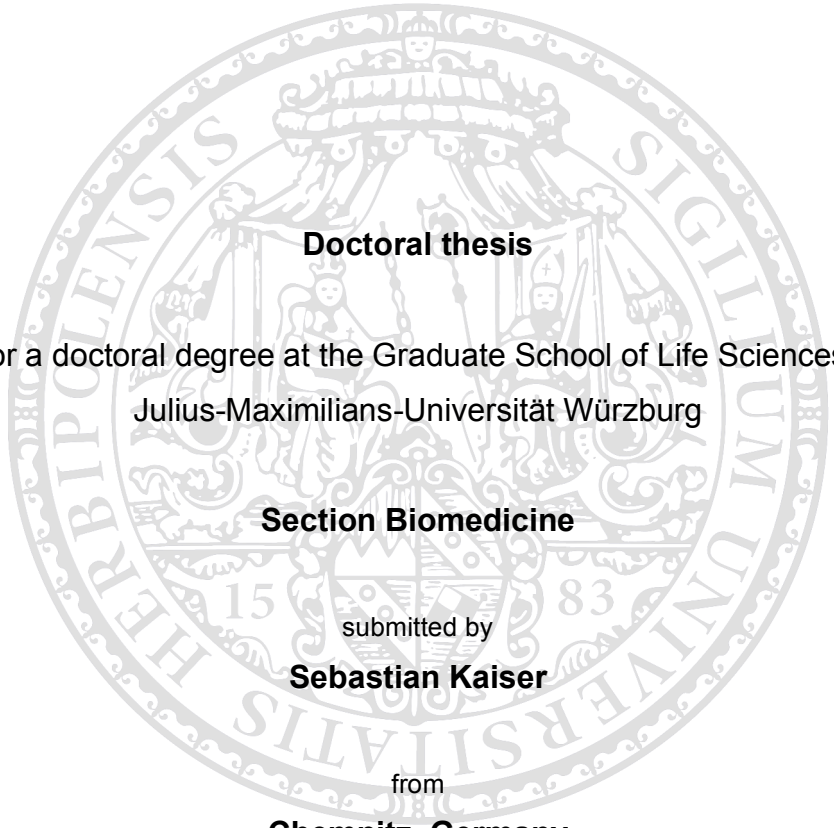
A RecQ helicase in disguise: Characterization of the unconventional Structure and Function of the human Genome Caretaker RecQ4

Die unkonventionelle RecQ Helikase RecQ4: Charakterisierung der ungewöhnlichen Struktur und Funktion eines essentiellen Beschützers des menschlichen Genoms



Doctoral thesis

for a doctoral degree at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg

Section Biomedicine

submitted by

Sebastian Kaiser

from

Chemnitz, Germany

Würzburg 2017

CC BY SA

Submitted on: ……

(office stamp)

Members of the *Promotionskomitee:*

Chairperson:	Prof. Manfred Gessler
Primary Supervisor:	Prof. Caroline Kisker
Supervisor (Second):	Prof. Thomas Müller
Supervisor (Third):	Prof. Martin Eilers
Supervisor (Fourth):	Prof. Volker Dötsch

Date of Public Defence: ……

Date of Receipt of Certificates: ……...

It always seems impossible

until it’s done

(A quotation generally attributed to Nelson Mandela or rocket pioneer Robert H Goddard)

Abstract

From the simplest single-cellular organism to the most complex multicellular life forms, genetic information in form of DNA represents the universal basis for all biological processes and thus for life itself. Maintaining the structural and functional integrity of the genome is therefore of paramount importance for every single cell. DNA itself, as an active and complex macromolecular structure, is both substrate and product of many of these biochemical processes. A cornerstone of DNA maintenance is thus established by the tight regulation of the multitude of reactions in DNA metabolism, repressing adverse side reactions and ensuring the integrity of DNA in sequence and function. The family of RecQ helicases has emerged as a vital class of enzymes that facilitate genomic integrity by operating in a versatile spectrum of nucleic acid metabolism processes, such as DNA replication, repair, recombination, transcription and telomere stability. RecQ helicases are ubiquitously expressed and conserved in all kingdoms of life. Human cells express five different RecQ enzymes, RecQ1, BLM, WRN, RecQ4 and RecQ5, which all exhibit individual as well as overlapping functions in the maintenance of genomic integrity. Dysfunction of three human RecQ helicases, BLM, WRN and RecQ4, causes different heritable cancer susceptibility syndromes, supporting the theory that genomic instability is a molecular driving force for cancer development. However, based on their inherent DNA protective nature, RecQ helicases represent a double-edged sword in the maintenance of genomic integrity. While their activity in normal cells is essential to prevent cancerogenesis and cellular aging, cancer cells may exploit this DNA protective function by the overexpression of many RecQ helicases, aiding to overcome the disadvantageous results of unchecked DNA replication and simultaneously gaining resistance against chemotherapeutic drugs. Therefore, detailed knowledge how RecQ helicases warrant genomic integrity is required to understand their implication in cancerogenesis and aging, thus setting the stage to develop new strategies towards the treatment of cancer.

The current study presents and discusses the first high-resolution X-ray structure of the human RecQ4 helicase. The structure encompasses the conserved RecQ4 helicase core, including a large fraction of its unique C-terminus. Our structural analysis of the RecQ4 model highlights distinctive differences and unexpected similarities to other, structurally conserved, RecQ helicases and permits to draw conclusions about the functional implications of the unique domains within the RecQ4 C-terminus. The biochemical characterization of various RecQ4 variants provides functional insights into the RecQ4 helicase mechanism, suggesting that RecQ4 might utilize an alternative DNA strand separation technique, compared to other human RecQ family members. Finally, the RecQ4 model permits for the first time the analysis of multiple documented RecQ4 patient mutations at the atomic level and thus provides the possibility for an advanced interpretation of particular structure-function relationships in RecQ4 pathogenesis.

Zusammenfassung

Vom simpelsten einzelligen Organismus bis hin zu hoch komplexen Lebensformen, genetische Information in Form von DNA repräsentiert die universelle Grundlage aller biologischer Prozesse, und damit die des Lebens selbst. Die Aufrechterhaltung der intakten Struktur und Funktion des Genoms ist daher von höchster Priorität für jede einzelne Zelle. Die DNA selbst, als aktives und komplexes Makromolekül, ist sowohl Substrat als auch Produkt einer Vielzahl dieser biochemischen Prozesse. Ein wesentlicher Aspekt für die Aufrechterhaltung genomischer Integrität besteht daher in der gezielten Regulation aller Prozesse des DNA Metabolismus, um die Konservierung der DNA in Sequenz und Funktion zu gewährleisten und unerwünschte Nebenreaktionen zu verhindern. Die Familie der RecQ Helikasen hat sich als eine essentielle Gruppe von Enzymen etabliert, die diese genomische Integrität gewährleisten, indem sie eine Vielzahl von DNA basierten Prozessen kontrollieren. Dies umfasst die Replikation, Reparatur, Rekombination und Transkription von DNA, sowie Prozesse, die der Stabilisierung der Telomere dienen. RecQ Helikasen werden von allen Zellen exprimiert und können in allen Domänen des Lebens – Bakterien, Archaeen und Eukaryoten nachgewiesen werden. Humane Zellen enthalten fünf verschiedene RecQ Helikasen, RecQ1, BLM, WRN, RecQ4 und RecQ5, welche sowohl individuelle als auch überlappende Funktionen in der Aufrechterhaltung genomischer Integrität innehaben. Eine Beeinträchtigung der Funktion der humanen RecQ Helikasen BLM, WRN und RecQ4 führt zu Krankheiten die durch eine erhöhte Wahrscheinlichkeit für die Entstehung von Krebs gekennzeichnet sind. Dies unterstützt die Theorie, dass die genomische Instabilität eine molekulare Grundlage für die Entstehung von Krebs darstellt. Allerdings repräsentiert die den RecQ Helikasen innewohnende Funktion der Aufrechterhaltung genomischer Integrität ein zweischneidiges Schwert. Während ihre Aktivitäten auf der einen Seite für normale Zellen essentiell sind, um Krankheiten und zelluläre Alterungserscheinungen zu verhindern, wird ihre DNA protektive Funktion

von Krebszellen genutzt, indem sie verschiedenste RecQ Helikasen überexprimieren und damit den nachteiligen Effekten der unkontrollierten DNA Replikation entgegenwirken. Zudem erlangen Tumorzellen durch die erhöhte Präsenz der RecQ Helikasen Resistenz gegenüber einer Vielzahl von Chemotherapeutika. Es ist daher von größter Bedeutung zu verstehen, wie genau die einzelnen RecQ Helikasen in der Entstehung von Krebs und dem Alterungsprozess involviert sind, um neue Ansätze in der Krebstherapie zu entwickeln.

Die vorliegende Arbeit präsentiert und diskutiert die erste detaillierte Röntgen-Kristallographische Struktur der humanen RecQ4 Helikase. Die vorgestellte Struktur umfasst den konservierten Kern der RecQ4 Helikase, einschließlich eines großen Teils ihres einzigartigen C-terminus. Eine Analyse des RecQ4 Modells weist sowohl eindeutige Unterschiede als auch unerwartete Gemeinsamkeiten im Vergleich mit anderen, untereinander strukturell und funktional ähnlichen, humanen RecQ Helikasen auf und erlaubt zudem Rückschlüsse auf die Funktion der einzigartigen C-terminalen RecQ4 Domäne. Die biochemische Charakterisierung verschiedener RecQ4 Varianten liefert funktionelle Einblicke in den Mechanismus der DNA Doppelstrangtrennung durch RecQ4 und deutet darauf hin, dass sich dieser in weiten Teilen vom Mechanismus der anderen humanen RecQ Helikasen unterscheidet. Letztlich repräsentiert das hier vorgestellte Modell der RecQ4 Helikase die Grundlage für die Analyse verschiedenster dokumentierter RecQ4 Patientenmutationen und erlaubt damit eine erste Abschätzung von Struktur-und-Funktions-Beziehungen bezüglich der bekannten RecQ4-assoziierten Krankheitsbilder.

Table of Contents

List of Figures

List of Tables

Abbreviations

2xYT	Double yeast and tryptone (LB variant)
3C	PreScission protease
5-FU	5-Fluorouracil
aa	Amino acid
ADP	Adenosine diphosphate
ALT	Alternative lengthening of Telomeres
AMP-PNP	Adenylyl-imidodiphosphate
ARL	Aromatic rich loop
ATP	Adenosine triphosphate
ATPγS	Adenosine 5'-O-(3-thio)triphosphate
BER	Base excision repair
BGS	Baller-Gerold Syndrome
BLM	Bloom Helicase
BRCv	Breast cancer susceptibility type 2 variant
CD	Circular dichroism
CDDP	*cis*-platinum derivate
cDNA	Complementary DNA
CEX	Cation exchange chromatography
CO	Crossover product
CPT	Topoisomerase inhibitor
CTD	C-terminal domain
CV	Column volume
Cy3	Cyanine 3
D-Loop	Displacement loop
dHJ	Double Holliday-Junction
DNA	Desoxyribonucleic acid
DSB	Double strand break
EBV	Epstein Barr Virus
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electro mobility shift assay
ESRF	European synchrotron radiation facility
FF	Fast flow
FL	Full length
FP	Fluorescence polarization

FT	Flow-through
G4	G-quadruplex DNA
HD1/2	Helicase domain 1 / 2
HEPES	4-(2-hydroxymethyl)-1-piperazineethanesulfonic acid
HGPS	Hutchinson-Gilford progeria syndrome
HJ	Holliday-Junction
HR	Homologous recombination
HRDC	Helicase-and-RNase-D C-terminal
HTH	Helix-turn-Helix
I-T	I-TASSER
IDA	Iminodiacetic acid
IMAC	Immobilized metal ion affinity chromatography
iPSCs	Induced pluripotent stem cells
IPTG	Isopropyl-β-D-thiogalactoside
IR	Ionizing radiation
ITC	Isothermal titration calorimetry
KIX	Kinase-inducible
LB	Lysogeny broth
LDH	Lactate dehydrogenase
MALS	Multi-angle light scattering
MMR	Mismatch repair
mRNA	Messenger RNA
mtDNA	Mitochondrial DNA
NADH	Nicotinamide adenine dinucleotide
NCO	Non-crossover
NDP	Nucleotide diphosphate
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NTA	Nitrilotriacetic acid
NTP	Nucleotide triphosphate
NTS	Nuclear targeting signal
OD	Optical density
OH	Overhang
PAR	Poly ADP-ribose
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
Ph2	Phyre2

Pi	Inorganic phosphate
PK	Pyruvate kinase
PML	Promyelocytic Leukemia
PTM	Posttranslational modification
R4ZBD	RecQ4-zinc-binding domain
RAPA	RAPADILINO syndrome
rDNA	Ribosomal DNA
RFU	Relative fluorescence unit
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RQC	RecQ-C-terminal
RTS	Rothmund-Thomson-Syndrome
SAD	Single wavelength anomalous dispersion
SCE	Sister chromatid exchange
SDS-PAGE	Sodium-dodecyl polyacrylamide gel electrophoresis
SDSA	Synthesis dependent strand annealing
SEC	Size exclusion chromatography
SF	Superfamily
SLIC	Sequence and ligation independent cloning
SRI	Set2-Rpb1-interacting
ssDNA	Single-stranded DNA
T-Loop	Telomeric loop
TB	Terrific broth
TED	Tris(carboxymethyl)ethylene diamine
TEV	Tobacco Etch Virus
TLS	Translesion synthesis
Tris	Tris(hydroxymethyl)-aminomethane
Trx	ThioredoxinA
UV	Ultraviolet light
WH	Winged helix
WRN	Werner helicase
WS	Werner syndrome
WT	Wild type
XP	Xeroderma pigmentosum
Zn	Zinc binding domain

This page has been left blank intentionally

1. Introduction

1.1. Genome integrity, aging and cancer

Maintenance of genomic integrity is a prime objective for every living organism in order to sustain viability and ensure the faithful passage of genetic material to the following generation. Like any other macromolecular structure in an aqueous solution, DNA is continuously subjected to a multitude of unintended chemical alterations, leading to what is broadly referred to as DNA damages. The result of such chemical reactions include bulky adducts within the DNA structure (e.g. 6,4-photoproducts), chemically modified bases (e.g. through alkylation and oxidation), deamination, single-stranded (ss) and double-stranded (ds) DNA breaks, interstrand crosslinks and intercalation events. In contrast to RNA, proteins and lipids, DNA is at the highest level of the cellular informational hierarchy and cannot simply be recreated once it is damaged. Consequently, DNA integrity is exceptionally dependent on repair and maintenance. DNA damages arise both endogenously and through environmental sources. Famous examples of the latter include ultraviolet (UV) light, ionizing radiation (IR) or chemical mutagens like platinum-based compounds and polycyclic aromatic hydrocarbons[1]. However, the vast majority of DNA damages originate from the byproducts of the cells metabolic pathways, non-enzymatic hydrolysis of glycosyl bonds and deamination[2,3]. The precise quantification of such events is ambitious as not all endogenous modifications can be determined experimentally and the methodology to measure such damages is prone to the artificial introduction of additional DNA modifications. Hence, estimates with respect to the number of endogenous DNA damages in the human genome cover a wide range from a minimum of 10,000 incidences[4] to a magnitude of 50,000-200,000 lesions[5] per cell and day. The bulk of these endogenous alterations are oxidative reactions, which are caused by reactive oxygen species (ROS), by-products of the oxidative metabolism, most notably the oxidative phosphorylation within the inner mitochondrial

membrane[3]. Imbalance between ROS-production and -protection is cumulatively described as oxidative stress, which is profoundly associated with human pathology and the process of aging. DNA damages are primarily obstacles for genuine DNA-based metabolism like transcription and replication. However, if unrepaired, such damages may result in their permanent manifestation in form of mutations, which in contrast to the former cannot be reversed once introduced. Thus, in order to cope with the constant load of damage to the DNA, nature has evolved complex and efficient networks of repair and maintenance programs, which ensure functionality of the DNA and the faithful preservation of the embedded information. Among the most prevalent pathways are the classical DNA repair mechanisms, including nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), interstrand crosslink repair or direct damage reversal. However, DNA maintenance is not only restricted to DNA repair, as many DNA lesions, like ssDNA and dsDNA breaks, also require complex repair mechanisms like homologous recombination (HR) or non-homologous end-joining (NHEJ). Generally, such mechanisms take action at multiple points in DNA metabolism and are occasionally used in a varied fashion within different cellular contexts, which constitutes the requirement to precisely recognize each such event and thoroughly regulate the mechanism in order to achieve a preferable result. Thus, all of these regulating enzymes contribute to DNA maintenance and genomic stability. Therefore, mere chemical alterations to the DNA are not the only threat to genome integrity. DNA metabolism as an entity has to be tightly guided and regulated in order to avoid illegitimate side reactions, which can lead to dramatic consequences for the cell. Inherently, DNA serves as the template for replication and transcription and is especially vulnerable during these stages compared to when it is densely packed in the form of heterochromatin[6]. Failure to safeguard such reactions will not only lead to local incidences like DNA breaks, adverse recombination events or transcription- and replication-fork collapse; on a more global scale, large chromosomal aberrations, rearrangements and aneuploidy are just as likely the outcome of a

compromised DNA metabolism. Such alterations will inevitably undermine cell homeostasis and viability through pathway imbalance based on the loss or gain of large quantities of genetic information after mitosis. Thus, genome integrity does not solely rely on the fidelity and functionality of DNA damage repair, but rather on the functionality of the entire arsenal of proteins involved in DNA metabolism, including the cell division process, nuclear integrity and telomere maintenance.

DNA damage is a steady state phenomenon, as chemical alterations are constantly introduced while they are continuously repaired at the same time[7]. Thus, the entire genome is never completely free of DNA errors and, as a consequence, mutations eventually accumulate within the DNA over time. In turn, these mutations modulate the capacity of the cell to sustain cell homeostasis and viability. Moreover, DNA polymerases, representing indispensible components of any DNA repair mechanism and the replication machinery, are not 100% accurate when synthesizing new DNA and thus mutations are most frequently generated during DNA repair and replication[8]. In addition, a cell might just as well bypass a lesion and accept its biological consequences if time is of the essence and DNA replication needs to proceed, a process known as translesion DNA synthesis (TLS)[9]. The accumulation of DNA damage and mutations in the genome leads to what is cumulatively described as genomic instability: the tendency to a decline in DNA maintenance as it becomes more and more likely that the load of unrepaired damage and/or subsequent mutations affect essential components of DNA metabolism networks. In addition to DNA damage and mutations, epimutations are a third source of genomic instability, which are defined as heritable changes of epigenetic markers that control gene expression, yet in contrast to mutations, are not caused by changes in the actual information content of the DNA[10]. Genome instability is therefore recognized as the prime and fundamental cause for both, aging and carcinogenesis and accordingly, most modern theories, which attempt to rationalize the aging process on the molecular level, acknowledge DNA damages and their biological consequences as the underlying cause of

aging[11]. In an attempt to systematically categorize cellular and molecular processes, which govern the aging process, nine candidate hallmarks of aging have been proposed, each supported by a large body of scientific evidence[12]. In addition to the aforementioned genomic instability and epigenetic alterations, these hallmarks further comprise telomere attrition, loss of proteostasis, deregulated nutrition sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion and altered intercellular communication[12]. All of these processes are inherently connected with each other through their biological context and thus the following picture emerges: Primary DNA-based hallmarks of aging like genomic instability, telomere attrition, epigenetic changes and loss of proteostasis affect secondary processes, such as nutritional sensing, mitochondrial function and cellular senescence. These secondary systems are considered to mediate in principal a beneficial compensation of the damages caused by the primary class of aging hallmarks. Yet, if the damage persists over an extended period of time, these secondary processes ultimately provoke effects like stem-cell exhaustion and altered intercellular communication. Thus, this functional hierarchy emphasizes the tight relationship between genome integrity and the aging process, which in turn highlights the importance of DNA repair and maintenance systems for health and longevity. Human progeroid syndromes represent prime examples for the direct visualization of this relationship by featuring a broad spectrum of prematurely appearing aging characteristics in young individuals, which can be ascribed to the malfunction of one particular gene that, in most cases, is associated with genomic maintenance programs[13,14]. One of the leading examples for such a genetic disorder is Werner Syndrome (WS), which most accurately resembles the phenotype of normal aging in adolescent individuals[15,16]. The aging characteristics of WS include short statue, scleroderma-like skin, graying hair, cataracts, hearing loss and the premature appearance of aging-related diseases like atherosclerosis, osteoporosis, type 2 diabetes and cancer. WS is causes by mutations within the *wrn* gene, which encodes a RecQ-like helicase, implicated in DNA repair, replication and telomere maintenance. A second

famous example is the Hutchinson-Gilford progeria syndrome (HGPS), which comprises a similar broad range of premature aging characteristics, yet in contrast to WS, develops already during infancy and consequently, affected children seldom exceed an age of 20 years[17]. The underlying cause of HGPS is a sporadic mutation within the *lmnA* gene, encoding the prelamin-A protein. As a result of the mutation, prelamin-A cannot be correctly processed to the mature Lamin-A form, leading to morphological and functional impairment of the nuclear lamina. This indirectly causes a deregulation of DNA repair, particularly DNA double-strand break repair, by affecting the processes of HR and NHEJ[18], which may directly contribute to the development of HGPS-related premature aging features. In contrast to full progeroid syndromes like WS or HGPS, so-called segmental progerias feature only a subset of premature aging characteristics. Examples for such are xeroderma pigmentodsum (XP) and Cockayne syndrome, both caused by defects in the nucleotide excision repair (NER) pathways[19], or the Rothmund-Thomson-Syndrome (RTS), which is caused by mutations within the RecQ-like helicase RecQ4, an important *Genome Caretaker* protein.

Many progeroid syndromes are accompanied by an early onset of carcinogenesis, which is generally ascribed to the increased genomic instability in these genetic disorders. As with aging, it is widely accepted that the underlying cause of cancer is genomic instability and, as a result thereof, an increased rate and accumulation of somatic mutations, affecting tumor suppressor genes, oncogenes and the tightly regulated networks that manage proliferation, apoptosis and cellular senescence. Analogous to the above-mentioned functional core characteristics that govern the aging process, carcinogenesis features a set of characteristic functional networks that are typically compromised in cancers. These hallmark characteristics encompass proliferative signaling, evasion of growth suppression, resistance to apoptosis, gain of replicative immortality, sustained angiogenesis and the activation of invasion and metastasis[20,21]. Deregulation of genes that manage and control these programs will ultimately lead to the development of one or more of these hallmark capabilities, eventually driving carcinogenesis. In

addition, recent studies highlight the initially underestimated proportion of epigenetic alterations that influence cancer development and progression, for instance by silencing DNA repair proteins, which will in turn increase mutation frequencies. Intriguingly, overexpression of certain DNA repair proteins may also have an advantageous effect on malignant cells as the presence of these proteins could pose as a signal to block apoptotic pathways by mimicking a “DNA repair in progress” cellular state. A cornerstone of this cell-centered somatic mutation theory of carcinogenesis is the principle of clonal evolution and expansion, which states that pre-malignant as well as malignant cells are both subjected to the natural selection for mutations that lead to an increase in fitness or proliferative advantages. Such “beneficial” mutations will thus expand and amplify throughout the entire pre-malignant or tumor tissue. Generally, multiple mutations in one genome are required for a pre-malignant cell to become cancerous. Sustained genomic instability thus increases the chances to acquire one or more of these driver mutations alongside with a broad variety of passenger mutations, which initially neither grant direct selective advantages nor exhibit extensive adverse affects under normal physiological conditions. A collection of pre-malignant cells might therefore feature one or two universal driver mutations, while the mutations, which eventually induce malignancy in combination with the ensemble of passenger mutations, will be variable throughout the tissue. Since tumors are typically detected at an advanced stage, this genetic mosaicism constitutes a central problem for the therapeutic treatment of cancer, as the genetic diversity provides a means for a minority of tumor cells to evade anti-cancer treatment and renders the regrowing tumor resistant to further treatment by clonal expansion of the resistance genotype. The principle anti-cancer strategy today aims to inflict cytotoxic effects on cancer cells in order to induce apoptosis or cell death[22]. However, as cancer cells originate from normal host cells, they share many common molecular features, resulting in severe side effects due to limited selectivity of many cytotoxic drugs for malignant cells. In most cases, the increased proliferation rate of cancer cells represents the only, albeit fairly

inadequate, selective marker compared to the non-proliferating somatic host cells. However, normal non-malignant tissue with elevated proliferation, like the bone marrow or the gastrointestinal system is affected as well during treatment. Thus, modern developments in chemotherapy aim to increase the selectivity of anti cancer drugs either by administration of specific antibodies against tumor epitopes, antibody-coupled small-molecule inhibitors or small-molecule inhibitors against proteins and molecular networks, which are particularly indispensible for cancer cells[23]. This strategy of targeted anti-cancer therapy relies excessively on (1) the detailed knowledge of the underlying mechanisms, which in case of failure and deregulation give rise to cancer, (2) how different types of cancer develop on the molecular level and which combination of deregulated mechanisms lead to the cancer phenotype and (3) the ability to precisely detect these deregulated proteins/mechanisms via certain biological markers in order to develop a tailored therapeutic therapy. DNA maintenance and repair processes constitute essential networks that, when compromised, not only drive such deregulating alterations through increased genomic instability, but also sustain cancerous cells by facilitating an elevated DNA damage repair capacity. Therefore, genomic maintenance mechanisms resemble fundamental areas of cancer research. Especially structural studies on key proteins involved in genome stability provide the groundwork for a bottom-up structure-based drug-design approach, which represents a powerful tool to develop potent inhibitors against target proteins and networks that are indispensible for cancer development and progression.

1.2. Structure and Function of Helicases

Nucleic acid metabolism is largely orchestrated by the protein family of helicases, which are central in the maintenance of genomic stability and constitute one of the largest groups of enzymes in living organisms[24]. Many DNA metabolism processes like transcription, replication and DNA repair require the accessibility of single stranded nucleic acids. Helicases facilitate this necessity via the separation of double-stranded DNA, the segregation of annealed RNA molecules or DNA-RNA hybrids and the remodeling of nucleoprotein complexes. They achieve these functions by directed translocation along single- or double-stranded nucleic acids utilizing the energy of nucleotide triphosphate (NTP) hydrolysis. Given their vital role in DNA- and RNA metabolism, helicases are ubiquitously expressed in all kingdoms of life and are exceptionally conserved in both, protein sequence and structure. One hallmark of helicases is a set of highly conserved sequence motifs (Fig. 1-1), which enable the common helicase-typical characteristics, e.g. NTP hydrolysis and the interaction with nucleic acid substrates[25]. The most conserved sequence elements are motifs I and II, which mediate interactions with the NTP and facilitate hydrolysis. Motif I, also referred to as the Walker A motif, features an essential lysine residue that binds to the NTP β-phosphate group, thereby stabilizing the transition state during hydrolysis. Motif II, the so-called Walker B motif, features a conserved DExx sequence, where the aspartate coordinates a catalytic magnesium ion required for transition state stabilization and activation of a nearby water nucleophile for phosphoanhydride hydrolysis[26,27].

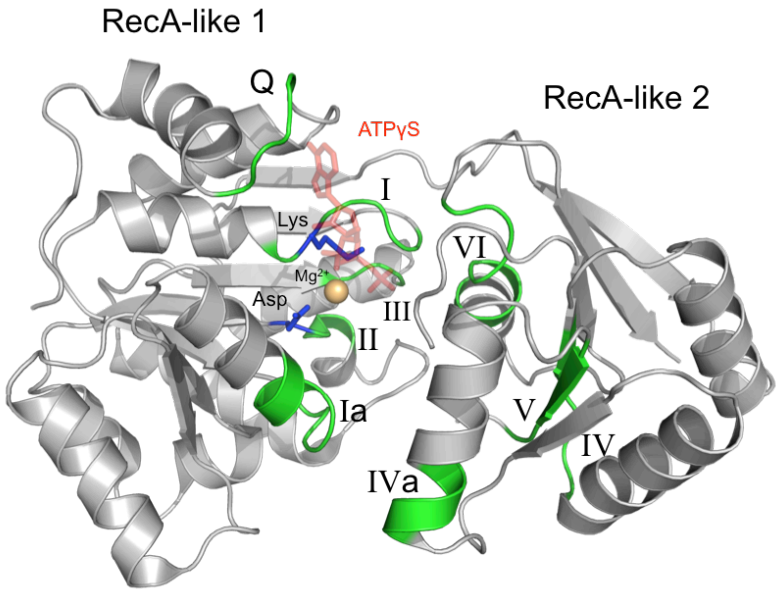


Figure 1-1: Conserved helicase motifs. The conserved helicase motifs are highlighted in green within the helicase core of *Ec*RecQ (PDB 1oyy). Motifs I and II, also known as Walker A and B motif, respectively, feature the catalytically important residues Lys and Asp (blue), which are illustrated in stick representation. The Asp-coordinated magnesium ion is depicted as orange sphere. A co-crystallized non-hydrolysable ATP analogue, ATPγS, is depicted in red, coordinated between the two RecA-like domains.

However, with the exception of the highly conserved sequence elements such as the Walker A and B sequences, not every helicase motif is present in each helicase and the degree of conservation among these motifs varies between different enzymes. These subtle differences, in fact, determine the classification of helicases into six superfamilies (SF1-6), some of which contain further sub-families and sub-groups[28]. This classification is constantly improved and refined as new helicase features are discovered and sequence information of new species become available. Although the main determinant for the affiliation with a particular SF is given by the presence and conservation of the helicase motifs, further classification is based on functional characteristics, like the preferred translocation polarity (e.g. either 3'-to-5' or 5'-to-3'), whether they act on DNA or RNA and use a single- or double-stranded nucleic acid substrate and their ability to hydrolyze either any NTP or only ATP. However, all of these functional characteristics are not

determinative for the integration into any particular SF. The basic functional unit of every helicase is the so-called helicase core, which features two protein domains, each resembling the structural fold of the recombination protein RecA. Within the two largest SFs, SF1 and SF2, these RecA-like domains are arranged in a tandem fashion within the same polypeptide chain and thus members of these SFs operate predominantly in a monomeric manner. In contrast, helicases from SF3-6 feature either six or twelve individual RecA-like copies with one RecA domain in each monomer, forming a hexameric or double-hexameric ring system, respectively[29]. The conserved helicase motifs, which are responsible for NTP hydrolysis, are typically located in the cleft between two adjacent RecA-like domains. NTP binding, -hydrolysis and NDP+P_i release enable the two neighboring domains to undergo a series of conformational changes relative to each other, which results in varying affinities of each RecA-like domain towards the nucleic acid substrate. Thus, repeated cycles of NTP hydrolysis are directly translated into a translocation motion along the nucleic acid, with a certain step-size per hydrolyzed NTP. This so called “inchworm mechanism” of translocation has been proposed based on the analysis of numerous helicase structures, which have been either crystallized in their apo form or with different nucleic acid substrates and/or NTP analogues, thus representing snapshots of each individual step of the inchworm reaction cycle[30]. Furthermore, a large body of functional data supports the theory of the inchworm mechanism. While the structural and functional analysis of the helicase core describes the conserved mechanism of translocation along nucleic acids, individual accessory domains of a helicase define its particular function and mode of action within its respective nucleic acid metabolism pathway. Such additional domains may be conserved within the respective sub-family or sub-group if they facilitate a common function. However, more often such additional domains are helicase specific and either grant additional functions like nuclease activity, warrant individual attributes like the recognition of a specific DNA- or RNA substrate or enable particular protein-protein interactions.

1.3. The family of RecQ helicases in genome maintenance

Among the different groups of DNA helicases, the RecQ family has emerged as a vital class of enzymes, whose members are involved in virtually all processes of genome maintenance and thus are often referred to as *Genome Caretaker* proteins[31]. RecQ proteins belong to the SF2 of helicases and exclusively exhibit a 3'-5' directionality by utilizing DNA-dependent ATP hydrolysis for translocation-based helicase activity. Impaired function of RecQ helicases is associated with deficiency in DNA damage repair, increased sensitivity to replicative and transcriptional stress as well as chromosomal aberrations and rearrangements, which ultimately promote carcinogenesis and accelerated aging in higher organisms[31-33]. RecQ proteins are universally conserved (Fig. 1-2), mostly featuring a single homolog in bacteria and other single-cellular organisms, while higher eukaryotes employ multiple paralogues. The founding member of the RecQ family was discovered in *Escherichia Coli* (*E. coli*, *Ec*RecQ)[34] where it acts in the HR pathway and DNA repair[35], the stabilization and repair of stalled replication forks at sites of DNA damage[36] and as suppressor of illegitimate recombination events[37]. Subsequently, a single RecQ homologue was identified in *Saccharomyces cerevisiae* (*S. cerevisiae*, Sgs1)[38], whereas two homologues have been discovered in *Schizosaccharomyces pombe* (*S. pombe*, Rqh1 and Hrq1)[39,40]. Based on the homology to the helicase core domain of *Ec*RecQ, five RecQ helicases have been identified in humans: RecQ1, BLM, WRN, RecQ4 and RecQ5.

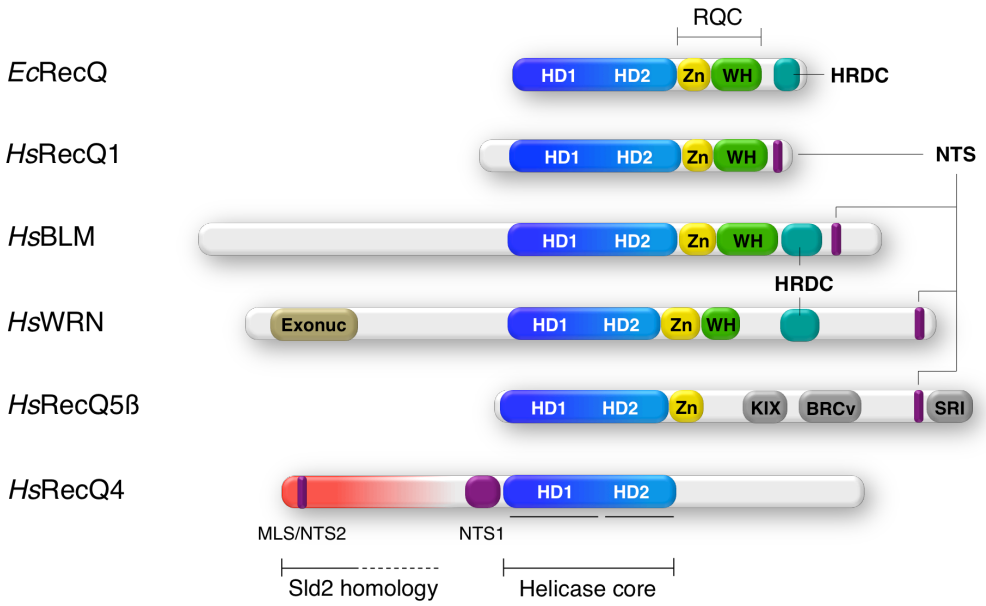


Figure 1-2: Domain architecture of RecQ helicases. RecQ helicases feature a set of characteristic elements, such as the conserved helicase core (HD1 and HD2, blue) and the less-conserved RecQ-C-terminal (RQC) domain, composed of a zinc-binding domain (Zn, yellow) and winged-helix (WH) domain (green). The helicase-and-RNaseD-C-terminal (HRDC) domain (turquoise) is only present in bacterial and yeast homologues, as well as in human BLM and WRN. Abbreviations: NTS – Nuclear targeting signal. MLS - Mitochondrial localization sequence. Exonuc – Exonuclease. KIX – Kinase inducible. BRCv – Breast cancer type 2 susceptibility variant. SRI – Set2-RPB1 interaction.

All human RecQ helicases exhibit both, cooperative and individual functions in genomic maintenance through their combined activities in DNA replication, recombination and repair as well as telomere stability[33,41]. From a general perspective, human RecQ helicases are multifunctional anti-recombination factors and tumor suppressor proteins. They maintain genomic integrity by recognizing specific events in the DNA metabolism by means of binding to an exceptionally broad spectrum of different nucleic acid intermediate structures and subsequently guide an appropriate response by recruiting additional proteins, enabling access to the DNA for further processing or stimulate the activity of other binding partners, thereby suppressing illegitimate side reactions. They mainly act during critical events in DNA metabolism, such as DNA replication, transcription and DNA repair where they counteract replicative and transcriptional stress in form of stalled or collapsed replication forks and transcription bubbles, especially in the context of unrepaired DNA

damages and so-called difficult-to-replicate GC-rich sites, which naturally challenge these processes. In addition to their helicase activity, which is required for many of their genome protective functions, all human RecQ helicases feature a potent nucleic acid strand annealing activity, a function that is allegedly equally important to ensure genomic stability, yet considerably less well studied[42]. Most intriguingly, human RecQ helicases exhibit a double-edged nature due to their function in maintaining genomic integrity. While their prime purpose is to prevent mutations and chromosomal aberrations, which contribute to aging and carcinogenesis, overexpression of RecQ helicases is a hallmark of many cancers, due to their inherit genome caretaker functions that protect cancerous cells from increased replicative stress and its associated DNA damages, which naturally occur in frequently dividing tissues, thus preventing cancerous cells to enter mitotic arrest or undergo apoptosis[43,44]. Therefore, human RecQ family members may serve as potential biomarkers for the identification of certain cancer types and have considerable potential as targets for the design of novel anti-cancer drugs in chemotherapy, as their inhibition generally potentiates the cytotoxic effects of conventional small-molecule inhibitors in cancer therapy and counteracts resistance development[45-47].

1.4. RecQ family Structure and Function

Most RecQ helicases feature a set of functionally and structurally conserved domains, which enable the translocation along DNA, the specific interaction with a variety of DNA substrates and, ultimately, genuine helicase activity. These RecQ-conserved domains comprise the helicase core, the RecQ-C-terminal (RQC) domain and the helicase-and-RNaseD-C-terminal (HRDC) domain (Fig. 1-2). The most conserved element is the helicase core, consisting of the two tandem RecA-like domains, termed helicase domain 1 and 2 (HD1 and HD2). The HDs harbor the conserved helicase sequence motifs, which determine the affiliation of all RecQ helicases to the group of SF2 helicases[29] (Fig. 1-1). The second most conserved element in RecQ helicases is the RQC domain, which is subdivided into a zinc-binding domain (Zn) followed by a winged-helix (WH) domain. The WH domain mediates dsDNA binding and harbors the structural elements for dsDNA strand separation[48-51]. The least conserved element in RecQ helicases is the HRDC domain, which is only present in BLM and WRN as well as bacterial and yeast homologs. It remains unclear whether the HRDC domain confers a common function. *In vitro* studies demonstrate that the HRDC domain is dispensable for conventional helicase activity in WRN and BLM and neither of the isolated HRDC domains exhibits DNA binding affinity[52,53]. However, crystal structures depict the HRDC domain in close proximity to the helicase core, suggesting a possible role in the modification of ATPase- and/or helicase activity[54,55] (Fig. 1-3 A). Moreover, it has been demonstrated that the HRDC domain of BLM is of particular importance for the BLM-mediated double Holliday Junction (dHJ) dissolution activity which can not be substituted by the HRDC domain of the WRN protein, suggesting an evolutionary differentiation of the individual HRDC domains[56]. This is further substantiated by the different surface charge properties of the BLM- and WRN HRDC domains which could account for individual protein-protein interactions, fine-tuning the distinct cellular functions of the two RecQ helicases[52,57]. Apart from the conserved elements present in most RecQ

helicases, many RecQ homologs feature long and unique N- and C-terminal extensions, which either grant additional functions or resemble protein-protein interaction modules, further broadening the complex functional network of these versatile enzymes. Examples include the N-terminally located exonuclease domain of WRN, the N-terminally located Sld2-like domain of RecQ4 or the C-terminally located BRCv- KIX- and SRI domains of RecQ5 (Fig. 1-2).

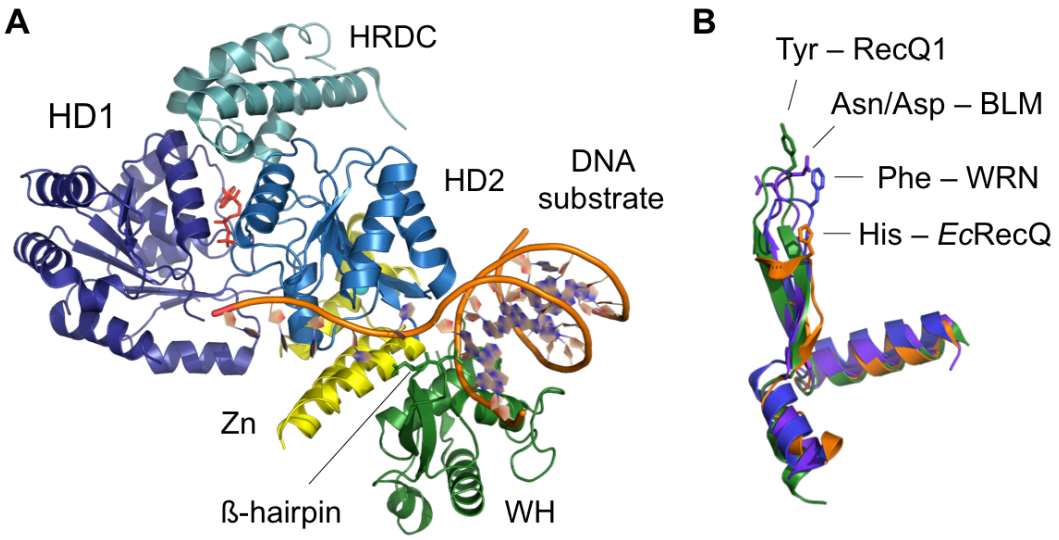


Figure 1-3: Prototypical structure of RecQ helicases. (A) The structure of human BLM (PDB 4cgz) illustrates the conserved RecA-like helicase domains HD1 (blue) and HD2 (light blue), which coordinate an ADP molecule between them (red, stick representation). The RQC domain is composed of the Zn domain (yellow) and the WH domain (green). The Zn coordinates a zinc ion (cyan sphere), whereas the WH domain mediates contacts with the DNA substrate and features the ß-hairpin element, which destabilizes the base pairing at the ss/dsDNA junction. The HRDC (turquoise) domain is located in close proximity to both HDs, where it might influence ATP-binding and/or –hydrolysis. **(B)** Superposition of the ß-hairpins of human RecQ1, BLM, WRN as well as *E. coli* RecQ, with their respective tip residues depicted in stick representation. By comparison, the bacterial ß-hairpin is shorter than the human versions of RecQ1, BLM and WRN.

Generally, the minimal functional unit of each RecQ helicase consists of the helicase core and the RQC domain. ATP-binding, -hydrolysis and ADP-release takes place in the cleft between HD1 and HD2 and is subsequently translated into a cycle of conformational changes among the two domains, which leads to a directed 3' to 5' translocation along the ssDNA strand[58]. The

WH fold within the RQC domain harbors a specific structural element, termed the ß-hairpin, which in the human RecQ helicases interacts with the paired bases at the ssDNA-dsDNA junction and destabilizes Watson-Crick base pairing (Fig. 1-3 A). Therefore, the ß-hairpin resembles a wedge element, which is pushed between paired bases upon DNA translocation, enabling DNA-double-strand separation in the process[48-50,59].

Surprisingly, the tip of the ß-hairpin, presenting amino acid residues responsible for the destabilization of paired bases, is not uniformly conserved in the RecQ family. Human RecQ1 features a tyrosine residue, whereas WRN utilizes a phenylalanine and a methionine for base-paring distortion. The BLM helicase features a T-shaped ß-hairpin tip, with an asparagine facing the base at the ssDNA/dsDNA junction and an aspartate pointing in the opposite direction. It has been proposed that this particular arrangement of the ß-hairpin tip in BLM might have functional importance particularly for its unique double Holliday Junction (dHJ) dissolution activity[60]. Bacterial RecQ homologs typically feature a histidine at the tip of the ß-hairpin, however, the entire pin is much shorter in bacterial RecQ homologs compared to the ß-hairpins of its human family members and does not interact with the bases at the ssDNA/dsDNA junction (Fig. 1-3 B). *In vitro* studies confirm that, in contrast to human RecQ helicases, the entire ß-hairpin is dispensable for helicase activity in bacterial RecQ homologs[51,61]. These observations led to the hypothesis that bacterial RecQs might utilize a different DNA unwinding mechanism than their mammalian homologs, despite the high structural conservation of the RQC domain in the RecQ family. One hypothesis proposes a bend-angle-driven helicase mechanism, as bacterial RecQ/DNA complex structures exhibit a larger bend-angle at the ssDNA/dsDNA junction (~90°) compared to the bend-angles observed in the DNA-bound structures of human BLM and RecQ1 (20-60°)[51]. However, due to the limited availability of DNA-bound RecQ structures, this hypothesis awaits further structural and experimental confirmation. Concerning the high structural conservation of the RQC domain and its functional importance for the helicase mechanism, it is intriguing that the entire RQC domain in RecQ4

as well as the RQC-WH domain in RecQ5 are not preserved in these important proteins (Fig. 1-2). A recent structural and functional analysis of human RecQ5 demonstrated the functional importance of a single α-helix, which directly follows the RecQ5 Zn domain. The experimental data suggest that this α-helix might adopt a wedge-like function comparable to the observed ß-hairpin element in RecQ1, BLM and WRN, thus enabling dsDNA strand separation in RecQ5[62]. These observations indicate that RecQ5, along with RecQ4, may utilize a very different dsDNA separation mechanism compared to the other human RecQ family members. However, due to lack of structural information regarding the C-termini of RecQ4 and RecQ5, this hypothesis remains to be validated.

An additional structural element that is conserved within the HD1 of all RecQ family members is the so-called aromatic-rich loop (ARL), which enables ATP hydrolysis upon binding of ssDNA[63]. The ARL by itself is a conserved structural feature found in many SF1 enzymes as well as the SF2 RecQ helicases, in which it is positioned in a surface exposed fashion within the cleft between the two RecA-like helicase domains[64,65]. However, in SF1 helicase members, the ARL constitutes the C-terminal end of helicase motif III, whereas the SF2 RecQ helicases feature the ARL between helicase motifs II and III[63]. Structurally, this distinctive positioning bears minor functional differences, yet a detailed mutational analysis of the *Ec*RecQ ARL revealed that the ssDNA sensor function is generally preserved[51,63]. The conserved *Ec*RecQ ARL residues Trp154, Phe158 and Arg159 were identified to detect the ssDNA-binding event and subsequently transmit the signal by initiating subtle structural rearrangements in the nearby helicase motifs, which ultimately favors ATP hydrolysis. In line with these functional data, a DNA bound structure of the bacterial RecQ homologue from *Cronobacter sakazakii* depicts the conserved ARL-phenylalanine side-chain in a stacked position between two DNA bases and the ARL-arginine residue in direct contact with the phosphodiester backbone, mediating an electrostatic interaction[51] (Fig. 1-4). Thus, the ARL in RecQ helicases serves

as an important and conserved ssDNA sensor element that couples ATP hydrolysis to the binding of the ssDNA cofactor.

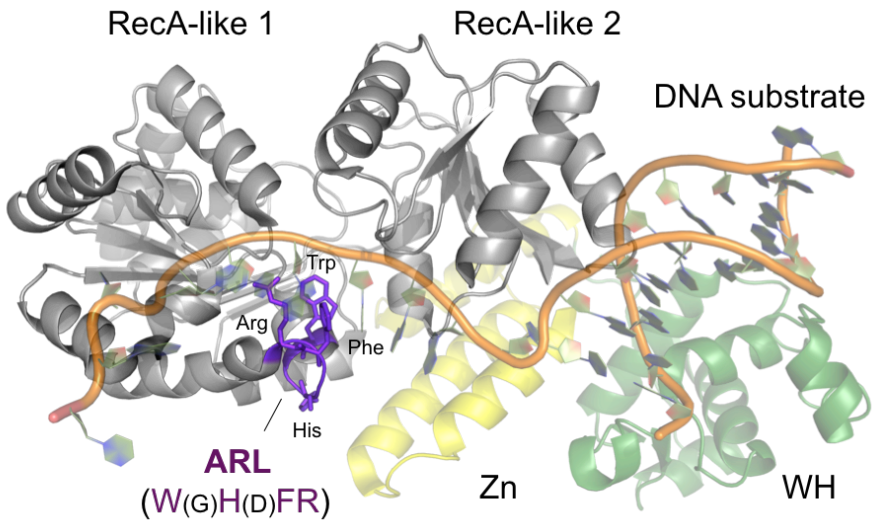


Figure 1-4: The RecQ conserved aromatic-rich loop (ARL). The DNA-bound structure of the RecQ helicase from *Cronobacter sakazakii* (PDB 4tmu) illustrates the direct interaction between the ARL and the ssDNA entity of the DNA substrate. The ARL phenylalanine mediates a stacking interaction with two bases, whereas the ARL arginine mediates electrostatic interactions with the backbone of the DNA (both residues depicted in stick representation). Ich würde die Reste nicht in rot darstellen, weil man sie dann praktisch nicht sehen kann.

1.5. RecQ cellular functions in genome maintenance

1.5.1. RecQ1, safeguarding DNA replication

The first *Ec*RecQ homologue identified in humans was RecQ1, which is the smallest of the human RecQs with a length of 649 amino acids (aa) and a molecular weight of 73 kDa, exhibiting the strongest homology to the *Ec*RecQ helicase among all human RecQ paralogues[66] (Fig. 1-5).

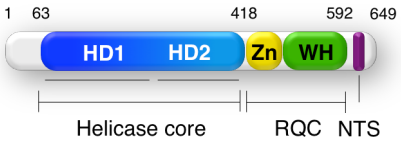


Figure 1-5: RecQ1. Domain architecture and boundaries of the human RecQ1 helicase.

RecQ1 is ubiquitously expressed in all tissues in a cell-cycle independent manner[43] and is ranked among the most abundant proteins of the cell[67]. In normal unperturbed cells, RecQ1 is an integral component of the replication complex, promoting fork stability, progression and repair in case of DNA damage induced fork stalling or collapse[68]. During replication stress, RecQ1 catalyzes fork regression and restart of stalled replication forks in conjunction with the DNA damage sensor poly (ADP-ribose) polymerase 1 (PARP-1)[68], which covalently attaches polymers of ADP-ribose (PAR) to itself and a variety of nuclear target proteins upon DNA-damage detection[69]. The PARP-1 mediated PARylation of RecQ1 prevents premature fork regression in order to permit sufficient time for the repair of the DNA damage[68]. In addition to replication fork DNA structures, RecQ1 binds and unwinds several other structural intermediates of DNA repair, such as displacement loops (D-loops) and Holliday junctions (HJs)[70]. However, similar to all other human RecQ helicases, RecQ1 requires a 3'-ssDNA loading site in order to resolve these structures via its ATP-dependent helicase/translocation activity[71]. In contrast to BLM, RecQ1 is not able to displace ssDNA-bound Rad51 directly via its helicase activity and thus cannot prevent D-loop formation, which is one of the first steps in HR-mediated repair of DNA double strand breaks (DSBs)[72]. The Rad51-bound nucleoprotein filament initiates the invasion of the complementary DNA double strand on the sister chromatid and searches for homologous regions in order to initiate HR-mediated repair[73]. The illegitimate initiation of this process represents unnecessary exchange of genetic information between sister chromatids (so-called sister chromatid exchanges, SCEs), which enhances the likelihood of chromosomal rearrangements, breaks and subsequent mutations, thereby driving carcinogenesis. Although unable to remove Rad51 from ssDNA via its helicase activity, RecQ1 mediates Rad51 displacement via branch migration

of the HJ, a structure that is established subsequently to the D-loop during HR[74]. Moreover, RecQ1 promotes unidirectional branch migration of the HJ[75], which may be required to remove toxic or non-productive attempts of HR-mediated replication fork restart. Interestingly, the N-terminus is required for the HJ resolution activity of RecQ1 as well as for the annealing of complementary DNA strands[70,76]. The latter was demonstrated by the crystallized RecQ1$^{49-616}$ variant, which is able to separate double stranded DNA but lacks annealing activity[49]. Furthermore, the N-terminus facilitates oligomerization of RecQ1, suggesting that strand annealing and HJ processing require an oligomeric form of the RecQ1 helicase[76], thus implying that DNA strand annealing is not simply the reverse form of DNA strand separation. Underlining its role in DNA damage repair, RecQ1 interacts with a variety of DNA repair proteins that participate in different repair pathways. In addition to PARP-1[77] and Rad51[78], both being involved in HR-mediated DSB repair, further binding partners encompass Exo1, MSH2/6 and the MRN complex, known to be involved in MMR[79] as well as Ku70/80, which facilitates the repair of DSBs via the error-prone NHEJ pathway[80].

Considering the important functions in DNA maintenance outlined above, it is surprising that RecQ1 knockout mice lack any characteristic phenotype under normal growth conditions, suggesting that redundant systems exist and that RecQ1 is not implicated in developmental processes[81]. Furthermore, RecQ1 dysfunction is not implicated in human diseases, however, a polymorphism within the 3' untranslated region (A159C) is associated with faster tumor growth and poor survival, which might be based on decreased mRNA stability or an altered protein expression level[82]. Notwithstanding, loss of RecQ1 increases genomic instability, incidences of SCEs, chromosomal breakage and sensitivity to ionizing radiation, which causes DSBs[81]. Moreover, RecQ1 has been identified as a breast cancer susceptibility factor as it modulates the expression of key genes involved in cell migration, invasion and metastasis[83,84]. Accordingly, silencing of RecQ1 results in enrichment of G-quadruplex (G4) DNA structures within RecQ1 promoter targets, suggesting that RecQ1 could regulate the stability of G4 motifs *in*

vivo[85]. Gene regulatory functions have also been identified for other RecQ helicases like BLM and WRN as well as other SF2 helicases like XPB and XPD[86,87]. However, similar to XPB, RecQ1 only binds to G4 motifs and is unable to unwind these DNA structures, in contrast to BLM and WRN, suggesting that either additional factors are required, which assist the G4 resolution activity of RecQ1, or that RecQ1 may only unwind rather specific and intrinsically unstable G4 motifs *in vivo*[85]. Like all other human RecQ helicases with the exception of RecQ5, RecQ1 is upregulated in many different cancer variants[43]. This overexpression specifically in malignant cells implies the requirement for increased DNA damage repair capacity and DNA maintenance in frequently dividing tumor tissues, which consequently confers resistance to anti-cancer drugs like *cis*-platinum compounds[88], doxorubicin[89] and DNA alkylating agents[90]. Accordingly, silencing of RecQ1 in tumor cells reduces proliferation and increases the sensitivity for cytotoxic agents[91,92]. Overexpression of RecQ1 (and other RecQ helicases) might also compensate for inactivated cell-cycle checkpoints, which can be compromised in cancer cells in favor of unchecked proliferation[45]. Consistently, depletion of RecQ1 will kill cancer cells due to mitotic catastrophe events based on entering mitosis with a load of unrepaired DNA damages, while normal proliferating cells remain unaffected due to their intact cell-cycle checkpoint system[93]. Among the cancer types, which exhibit characteristic upregulation of RecQ1 expression, are glioblastomas, hepatocellular carcinomas, ovarian cancers, melanomas and head-and-neck cancers[46,47,88,92,94]. Intriguingly, other cancer types display downregulated RecQ1 expression levels, whereas a significant inverse correlated with the expression levels of RecQ4 is observed, but not with those of WRN, BLM or RecQ5[95]. Thus, RecQ1 and RecQ4 have a considerable potential in anti-cancer therapy via their targeted inactivation and may, in addition, serve as a biological marker for the identification of certain types of cancer, guiding individual therapy.

1.5.2. BLM, the Bloom's Syndrome and cancer

Given its substantial implication in carcinogenesis, BLM is one of the most intensively studied RecQ family members and, as it is typical for all RecQs, has multiple implications in DNA replication, repair and telomere maintenance. The 1417 aa protein features the conserved helicase core domains, which are characteristic for all RecQ proteins and confer canonical ATP- and ssDNA dependent 3' to 5' helicase activity[96,97] (Fig. 1-6).

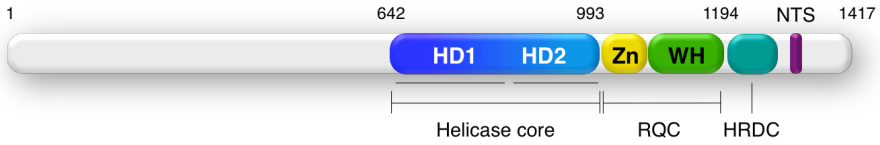


Figure 1-6: BLM. Domain architecture and boundaries of the human RecQ helicase BLM.

In addition, BLM exhibits large extensions at the N- and C-terminus, which grant discrete functions and feature regulatory elements, that are not present in other RecQ helicases[58]. While the C-terminal extension is fairly structured and harbors the HRDC fold, an auxiliary domain that is shared among the human RecQs only with WRN, the N-terminal region features only sparse secondary structure elements and is divergent even among BLM orthologs from different species[98]. However, a few conserved features of the N-terminus encompass regions that facilitate oligomerization and post-translational modifications (PTMs), which are required for certain cellular functions as well the localization of the protein to subnuclear structures such as the nucleolus or PML nuclear bodies[98,99]. In contrast to RecQ1, BLM is only expressed in actively dividing cells and peaks during S- and G2/M-phases of the cell cycle[100], consistent with the implication of BLMs role in DNA replication and recombination. In normal unperturbed cells, BLM is localized at replication forks, where it interacts with the flap endonuclease Fen1 and may thus be implicated in Okazaki fragment maturation[101].

Following replication stress, BLM is phosphorylated by the checkpoint kinases ATM and ATR and is actively recruited to stalled replication forks where it regulates both, fork regression via ATP-dependent HJ branch migration in order to stabilize the stalled fork (creating a so-called "chicken foot" structure) as well as the reversal of regressed forks to promote replication restart[102]. Among the best characterized functions of BLM are those in HR, which represents a versatile pathway in DNA metabolism that is utilized for the repair of DSBs, collapsed replication forks and eroded telomeres[103]. During HR, BLM is implicated in almost every step, which is best exemplified by the HR mediated repair of DSBs (Fig. 1-7). Here, BLM initially stimulates Dna2- and Exo1-mediated 5'-3' end resection on both DNA strands[104,105]. BLM further modulates the subsequent D-loop formation by removing ADP-bound Rad51 from the ssDNA and stimulating ATP-Rad51-ssDNA mediated strand invasion and homology search[72,106]. The D-loop structure is subsequently resolved via two different sub-pathways: (1) Synthesis dependent strand annealing (SDSA), which exclusively results in non-crossover products and (2) by the formation and processing of a double Holliday Junction (dHJ) structure. In the latter, BLM promotes branch migration of the two HJs towards each other, converting them into a hemicatenane structure, which is subsequently untangled by the dissolvasome complex, consisting of BLM, Topoisomerase III alpha (Top3a) and the RecQ-mediated genome instability proteins 1 and 2 (RMI1/2)[107-109]. Both activities, the dHJ dissolution and the decatenation of hemicatenane structures are unique functions of BLM that cannot be accomplished by other RecQ helicases[56]. In the absence of BLM, the two HJs are resolved by structure specific endonucleases, leading to equal amounts of non-crossover and crossover products. In contrast, the dissolution of the dHJ structure via BLM results, like the SDSA pathway, exclusively in non-crossover end products[107].

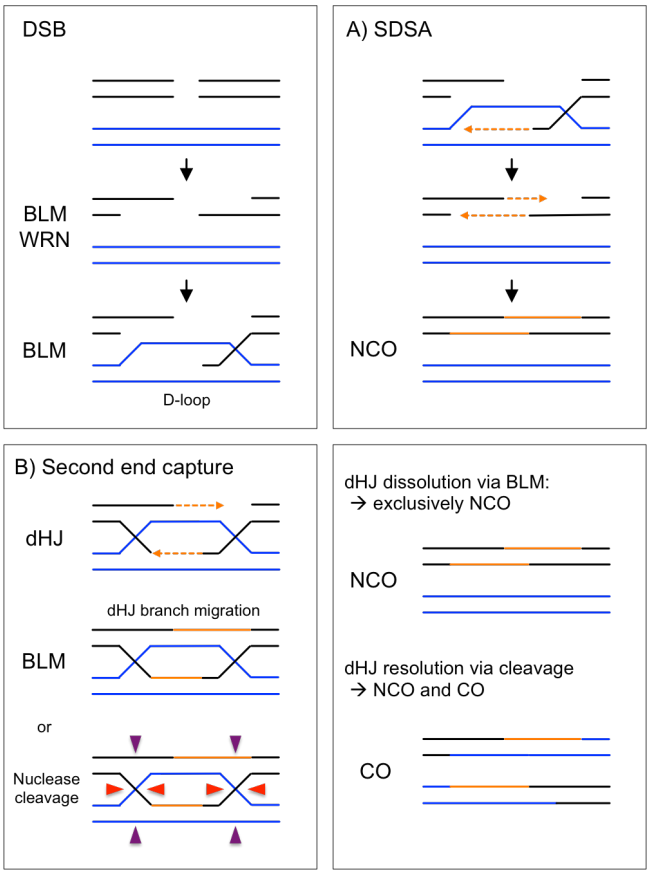


Figure 1-7: BLM and WRN in HR-mediated repair of DSBs. Both, BLM and WRN stimulate 5'-3' end resection at the DSB, while only BLM facilitates the Rad51-mediated homology search and subsequent strand invasion at the homologous sister chromatid, leading to D-loop formation. Subsequently, the DSB is repaired via (A) synthesis dependent strand annealing (SDSA), leading exclusively to non-crossover (NCO) products, or (B) the second end capture pathway, resulting in a double Holliday (dHJ) structure. The dHJ is either branch migrated by BLM (dHJ dissolution), again resulting in NCO products, or cleaved by structure-specific nucleases (dHJ resolution), resulting in NCO and crossover (CO) products, the latter being the foundation for sister-chromatid exchanges (SCEs).

Thus, BLM does not only regulate the initiation of HR at various DNA metabolism processes but also channels the outcome to non-crossover products in order to prevent excessive recombination of genetic material between sister chromatids. Consequently, the most characteristic feature of BLM defective cells is excessive HR, which leads to highly increased rates of SCEs, chromatid gaps and quadriradial chromosomes, representing unresolved HR intermediates[110].

BLM is also involved in telomere maintenance where it interacts with the shelterin complex proteins TRF1/2 and Pot1[111,112], a nucleoprotein structure that caps the ends of telomeres by the formation of a telomeric-loop (T-loop), thereby sequestering these structures and preventing the unintended initiation of side reactions. However, BLM defective cells do not exhibit accelerated erosion of telomeres, yet rather display a characteristic increase in telomeric SCEs (T-SCEs) and telomere associations as a result of

deregulated HR at telomeres[113]. Through its activities in HR, BLM is further involved in the alternative lengthening of telomeres (ALT) pathway[114], which is mediated by HR and exploited by various cancers in order to achieve telomerase-independent immortality[115]. Lastly, BLM is able to actively resolve G4 DNA structures[116], which are enriched in GC-dominated DNA areas such as telomeres, ribosomal DNA (rDNA) loci, promoter regions and immunoglobulin heavy chain switch regions[117-119]. Collectively, BLM aids in the replication and maintenance of these regions but is also actively involved in rDNA transcription[119], transcriptional regulation[87] and is required for proper immune system development[120].

Biallelic somatic mutations in BLM are causative for Bloom's syndrome (BS), a rare autosomal recessive disease that is characterized by genomic instability, in particular by increased incidences of SCEs, slow progression of replication and the accumulation of aberrant replication intermediate structures[97,121-124]. More than 60 mutations have been identified that inactivate BLM and subsequently cause BS with 90% of all missense mutations affecting the ATPase activity and thus render BLM a functional null allele[125]. BS patients exhibit a characteristic pre- and post-natal growth deficiency, leading to a very short but proportional statue, yet no other signs of major anatomic defects[121,126]. Affected individuals suffer from a mild immunodeficiency, increased susceptibility of type II diabetes as well as sunlight sensitivity and rash development at sun-exposed skin areas[127,128]. Although no abnormal sensitivity towards UV induced DNA damages has been reported[129,130], BS patients have an increased risk for basal and squamous cell carcinomas, but not for melanoma[131]. The most striking phenotype of BS is the substantially increased risk for a very broad spectrum of cancers. BS patients are on average 99 times more likely to be diagnosed with any cancer compared to the general population, with the mean age of diagnosis at 23 years and cancer being the leading cause of death[132]. The main reasons for this broad range of cancer susceptibility are most likely on the one hand the increased mutation rate, which is based on the excessive amount of unchecked HR and the subsequent accumulation of unresolved

HR intermediates, leading to chromosomal aberrations and mutations. Secondly, the increased rate of SCEs, which is about 20-50 times the frequency of normal cells[132], permits the vast distribution of a critical first-hit mutation via mitotic crossover events and consequently accelerates carcinogenesis. As of 2016, the Bloom's Syndrome registry listed 212 malignant neoplasms from 136 BS individuals, with the most common cancer types being leukemia and lymphoma (75/212) followed by colorectal cancer (31/212)[131]. Furthermore, BS patients exhibit an increased risk for developing particularly rare cancers, such as Wilms' tumor. Similar to RecQ1, BLM expression is upregulated in cancer cells as well as in Epstein-Barr virus (EBV) transformed cells[43] and, consequently, silencing of BLM was reported to suppress the growth of osteosarcoma U2OS cells and increases the sensitivity for cytotoxic drugs such as captothecin (CPT), the *cis*-platinum derivate CDDP and 5-Fluorouracil (5-FU)[133]. More recently, a high-throughput screening approach identified a lead compound, ML216, which preferentially inhibits BLM by decreasing its affinity to DNA substrates and leads to increased incidences of SCEs upon incubation in cell culture. However, although the effects of ML216 were only moderate on other related human RecQ helicases, such as RecQ1 and RecQ5, the compound also showed selectivity for WRN inhibition[134].

1.5.3. WRN, the Werner Syndrome and premature aging

The RecQ helicase WRN was first cloned in 1996 and has, at first glance, many features in common with BLM[135]. WRN is the largest of all human RecQ helicases with a length of 1432 aa and a molecular weight of 160 kDa (Fig. 1-8). Like BLM, it features vast N- and C-terminal extensions, which grant individual functions, in addition to the centrally located RecQ conserved domains. One of these unique features is the N-terminally located 3'-5' exonculease domain, whose precise cellular function still remains unclear[136,137].

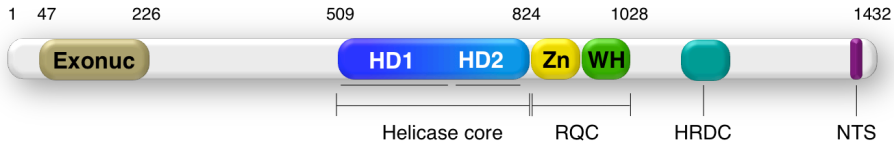


Figure 1-8: WRN. Domain architecture and boundaries of the human RecQ helicase WRN.

In contrast to BLM, WRN is not only expressed in dividing but also in quiescent and terminally differentiated cells, with increased expression in the pancreas, testis and ovaries[135]. Its preferred DNA substrates are, similarly to BLM, replication fork structures, DNA bubbles, G4 DNA and HJs and both proteins share activities in the HR pathway, through which WRN and BLM operate in the repair and stabilization of stalled replication forks, HR-mediated DSB repair and the maintenance of telomeres[103]. In HR, both proteins stimulate the Dna2 endonuclease activity, required for the 5'-3' resection of blunt DSB ends and for the reversal of the chicken foot structure at stalled and regressed replication forks[138,139] (Fig. 1-9). Like BLM, WRN is able to branch migrate a HJ *in vitro*, however, WRN fails to substitute for BLM in the dissolution of dHJ structures[56,140]. Furthermore, WRN interacts with the NHEJ proteins Ku70/80, DNA dependent protein kinase catalytic subunit (DNA-PKcs) and DNA ligase IV/XRCC4[141,142], which repair DSBs during G1 phase when no complementary sister chromatid is present for HR-mediated DSB repair, a function that is not shared with BLM but with RecQ1. Interestingly, Ku70/80 and DNA ligase IV/XRCC4 stimulate the 3'-5' exonuclease- but not the helicase activity of WRN[143], suggesting that the WRN exonuclease activity could facilitate the resection of non-paired 3'-overhang nucleotides during the end-joining process. Moreover, a double inactivation of BLM and WRN has more severe consequences compared to the inactivation of the single proteins, highlighting their non-redundant functions in replication fork recovery and DSB repair[132]. In addition to DSB repair, WRN might also be implicated in BER via the interaction with APE1 and Pol-β[144], in NER via its interaction with XPG[145] as well as in the TLS

pathway through the interaction with and stimulation of the TLS polymerases Polη, Polκ and Poli[146].

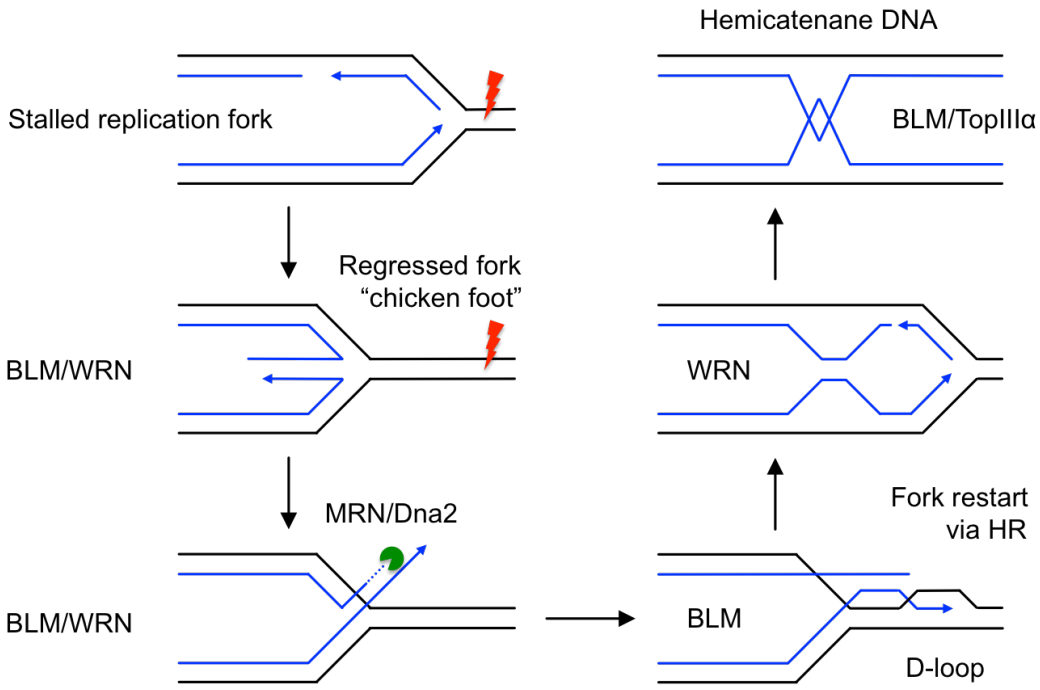


Figure 1-9: BLM and WRN at stalled replication forks. At stalled replication forks, BLM and WRN facilitate fork regression in order to stabilize the fork and prevent fork collapse, creating a chicken foot structure. To restart the fork, BLM and WRN stimulate 5’-3’ end resection, required for Rad51-mediated D-loop formation. The resolution of the D-loop structure leads to fork restart, while a possible entanglement of DNA strands, called hemicatenane DNA structures, are resolved by the dissolvasome complex, containing BLM and Topoisomerase IIIa.

Certainly, the most important function of WRN lies in the maintenance of telomeres, where it interacts with the shelterin proteins TRF1/2 and POT1, counteracts G4 motif formation at the G-rich telomeric strand, thus facilitating efficient telomere replication, and functions in the HR-mediated ALT pathway in order to prevent T-SCEs[147-149]. Although many of those functions are again shared with BLM, dysfunction of WRN results in a different telomeric phenotype compared to BLM inactivation, suggesting that WRN might be able to compensate for some telomeric BLM activities but not vice versa. Only WRN deficient cells exhibit a characteristic delay of telomere replication and a loss of telomeric TTAGGG repeats due to the formation of G4 motifs

that are not resolved during the replication process, subsequently initiating DNA damage signaling and telomere fusion events[148,149]. Furthermore, WRN defective cells enter senescence prematurely due to critically shortened telomeres, a feature that is not observed in BLM deficient cells[150]. The distinct functions of BLM and WRN are further exemplified by the different phenotypes of their respective human diseases, as outlined below.

Werner Syndrome (WS), a rare autosomal recessive disease caused by mutations in the WRN gene, was first described by Otto Werner in 1904[151]. So far, more than 70 different mutations have been described that cause WS, with point mutations in both, the helicase core and exonuclease domain[152]. Cells from WS patients exhibit slow replication, an increased mutation rate and genomic instability[153,154]. However, the type of genome instability is different from that displayed by BS patients. WS cells feature a so-called “variegated translocation mosaicism”, which is defined by a pseudodiploid, heterologous karyotype with a characteristic set of chromosomal aberrations[155]. In contrast to BS, WS cells neither show increased SCEs nor excessive HR, yet rather a defect in the HR pathway[156]. Like in BS, WS patients display a predisposition to develop cancer, however, while the cancer spectrum in BS is very broad, only certain cancer types are observed in patients with WS, with soft-tissue carcinoma, osteosarcoma or thyroid cancer being the most common cancer types[157,158]. As with all RecQ helicases, silencing of WRN expression is a gateway to increase genomic instability and drive carcinogenesis, yet many cancer types also display upregulation of WRN expression. For instance, in colorectal cancer WRN expression is suppressed via CpG-island promoter hypermethylation whereas c-Myc activates WRN expression in chronic myeloid leukemia cells[159,160]. In accordance with the genomic caretaker function in cancer cells, siRNA mediated knockdown of WRN was shown to increase the sensitivity for the topoisomerase inhibitors CPT and etoposide as well as for bleomycin[161,162]. In 2011, a small-molecule inhibitor was described, NSC 19620, which selectively inhibits WRN and acts synergistically with other DNA damage inducing agents to reduce proliferation and induce apoptosis,

opening the possibility for a future application in targeting c-Myc positive tumors[163,164]. In addition to the susceptibility for certain cancers, the most prominent characteristics of WS are the associated progeroid features: the premature appearance of a broad range of aging symptoms starting in adolescence. Such features include a lack of growth spurt, leading to a very short stature, skin atrophy, graying and loss of hair, bilateral cataracts (yet not the old age type) and the early onset of age-related diseases such as atherosclerosis, arteriosclerosis, osteoporosis and type II diabetes[165]. The main causes of death are cardiovascular disease and cancer at a mean age of 57 years[165]. Most interestingly, the progeroid features in WS may be directly related to WRN dysfunction in telomere maintenance. $WRN^{-/-}$ mice, which constitutively express murine telomerase, are healthy and do not exhibit any aging related symptoms. However, when combined with a knockout of m*Terc*, the murine telomerase RNA component, late generation mice feature premature aging characteristics, reminiscent to those in WS patients[166]. Moreover, induced pluripotent stem cells (iPSCs), generated from skin fibroblasts of a WS patient, lost the cellular phenotype associated with aging and subsequent re-differentiation led to the reappearance of the senescence phenotype. Of note, telomerase re-expression is part of the transcriptional reprogramming of iPSCs[167]. These studies highlight the tight association of telomere erosion and premature aging as well as the importance of WRN in telomere maintenance, which, in case of WRN dysfunction, cannot be compensated by any other RecQ helicase.

1.5.4. RecQ4, RTS and replication initiation

In 1998, RecQ4 and RecQ5 have been the last two human RecQ helicases identified and cloned so far. The human *recq4* gene encodes a 1208 aa protein of 133 kDa, featuring the conserved helicase domains at its center (Fig. 1-10). Both, the N- and C-terminus of RecQ4, exhibit vast extensions that are not homologous to any other RecQ helicase. Most striking, the RecQ4 C-terminus lacks the RecQ conserved RQC domain, which was

reported to be indispensable for helicase activity in RecQ1, WRN and BLM[48-50,59].

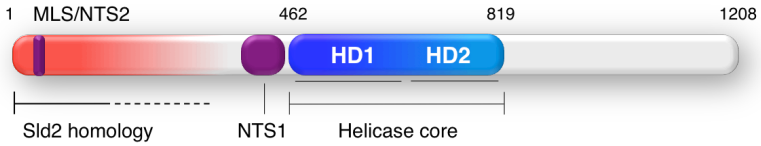


Figure 1-10: RecQ4. Domain architecture and boundaries of the human RecQ4 helicase.

RecQ4 and RecQ1 are the only two human RecQ helicases, which are directly implicated in DNA replication initiation. The N-terminal 370 aa of RecQ4 exhibit homology to the DNA replication initiation factors Sld2 from *S. cerevisiae* and DRC1 from *S. pombe*[168,169]. Accordingly, RecQ4 performs an essential function in the initiation of eukaryotic DNA replication via its N-terminal Sld2-like domain[168,170]. It was found that RecQ4 localizes to replication origins in G1/S-phase, where it is required for origin loading of the DNA replication polymerase α and further interacts with the replisome factors MCM10, the MCM2-7 helicase, Cdc45, the GINS complex and PCNA[170-172]. In line with the function in initiating DNA replication, RecQ4 expression is cell-cycle specific and peaks during S-phase, similar to BLM[173]. Furthermore, RecQ4 mRNA levels show tissue specific differences, with high expression levels in thymus, testis, placenta and the developing bone, whereas moderate expression has been reported in the heart, brain, colon and the small intestine[173]. Sequence analysis and functional studies of the N-terminus further revealed a mitochondrial localization signal (MLS) sequence within the first 84 aa. Thus, RecQ4 is the only RecQ helicase featuring mitochondrial localization in addition to its presence in the nucleus and the cytoplasm[174]. Accordingly, it has been shown that RecQ4 is important for mitochondrial DNA (mtDNA) synthesis and maintenance[174,175]. While the MLS coincides with a minor NTS, the major NTS of RecQ4 is located directly upstream of the helicase core[176]. This location too is

uncommon, as the NTS of other RecQ helicases are exclusively localized within their respective C-termini. The distribution of mitochondrial, cytoplasmic and nuclear RecQ4 content is highly dynamic and cell type specific, which suggests a complex subcellular regulation via posttranslational modifications of the two NTS and the MLS[177,178]. Although the N-terminus of RecQ4 appears to be largely unstructured, specific binding regions for RNA, Holliday-Junctions and G4 DNA have been proposed throughout the entire N-terminal 430 aa[179-181]. However, the structural basis for the recognition of these DNA substrates as well as their functional relevance remains elusive.

While the N-terminal domain of RecQ4 constitutes a vital component in embryogenesis and development[182], the DNA protective character of RecQ4 has been assigned to the helicase domain and the unique C-terminus, as mutations, which lead to RecQ4 related diseases, congregate within these regions[183]. Homozygous or compound heterozygous mutations in the *recq4* gene are associated with three different heritable syndromes: Rothmund-Thomson-Syndrome (RTS) type II, Baller-Gerold-Syndrome (BGS) and RAPADILINO (RAPA) syndrome[184-186]. RTS is the most prevalent syndrome associated with RecQ4 dysfunction and is characterized by a pre- and post-natal growth deficiency, skin abnormalities in form of hypo- and hyperpigmentation, atrophy and co-called spider veins (collectively described as poikiloderma), skeletal deformations, gastrointestinal disturbances and sparse or absent scalp hair and eye lashes[187]. Most importantly, RTS type II individuals exhibit an increased risk for developing osteosarcoma during childhood and squamous and basal cell carcinoma later in life[188]. Interestingly, RTS type I individuals, which do not exhibit mutations within the *recq4* gene and encompass about one third of all RTS patients, lack the prevalence for cancer development, thus linking cancer susceptibility in RTS type II directly to RecQ4 dysfunction[189]. Of note, the clinical features regarding skin abnormalities, sparse scalp hair, eyebrows and eyelashes are generally regarded as premature aging symptoms, which is unique to RTS among the RecQ4-associated syndromes. However, as the molecular origins

of these features are not yet well understood, it is unclear whether this phenotype truly reflects premature aging or if these symptoms are just interpreted as aging characteristics, yet are in fact of a different molecular origin. Therefore, it is still under debate whether RTS is a true premature aging disorder. RAPADILINO syndrome patients exhibit many of the RTS characteristics with the exception of poikiloderma and the premature aging phenotype. However, in addition to osteosarcoma, RAPADILINO individuals exhibit an increased risk for developing lymphoma. BGS is the scarcest syndrome among the RecQ4-asociated diseases and is characterized mainly by radial bone malformations and craniosynostosis. In contrast to RTS type II and RAPADILINO, BGS is genetically heterogeneous with mutations identified within the genes *recq4*, *fgfr2* and *twist*, all leading to a BGS phenotype[190,191]. Intriguingly, mutations in the *recq4* gene cannot be segregated in accordance to any of the RecQ4-associated diseases, rendering the analysis of disease mutations and their functional consequences rather difficult. The mutation spectrum within RecQ4 encompasses missense, nonsense as well as frameshift mutations[183]. The latter two mutations types typically lead to a premature termination of the polypeptide chain either within the helicase core or the unique C-terminus, thereby disrupting helicase activity and/or potentially deleting modules for protein-protein interactions. In addition, RecQ4 pre-mRNA is especially prone to splice site mutations due to atypically short intron sequences, which in some cases become too short to be efficiently spliced during mRNA maturation. One such case causes the deletion of exon 7, which harbors the main NTS of RecQ4. As a result, the deletion prevents the retention of RecQ4 within the nucleus, leading to the accumulation within the cytosol. This particular deletion mutant was found to be a founder-mutation associated with the RAPADILINO syndrome, as multiple homozygous patients are documented[185].

The molecular groundwork for RecQ4s function in genomic maintenance has been fueled by studies including *in vivo* localization, cell-based experiments investigating the effects of various DNA-damage causing agents, the

analysis of tissue derived from patients with RecQ4-associated diseases and the identification of RecQ4 protein interaction partners[192-196]. Furthermore, *in vitro* characterization of the RecQ4 helicase has highlighted preferred DNA substrates, thereby assisting to define RecQ4 functions in the context of DNA replication, recombination and repair. Here, it has been demonstrated that RecQ4 exhibits a more restrictive substrate specificity compared to other RecQ family members. *In vitro* helicase activity could only be verified for 3'-overhang- and forked duplexes as well as D-loop DNA substrates, yet not for HJs or G4 DNA[197]. One hotspot of RecQ4-activity is in the maintenance of telomeres, where it interacts with the telomeric TRF2 protein[192]. Accordingly, RecQ4 has been shown to resolve T-loops *in vitro* and cells derived from RTS patients exhibit increased telomere fragility, indicating RecQ4s importance for telomere replication and preservation. Furthermore, RecQ4 seems to play a role in chromosome segregation as primary RTS cells exhibit chromosomal abnormalities, most frequently trisomies of chromosomes 7 and 8[198,199]. Moreover, several ties connect RecQ4 to a variety of different DNA repair pathways. RecQ4 stimulates the activities of the BER proteins APE1 and polymerase ß, demonstrating a functional role for RecQ4 in the removal of oxidative DNA damages[195]. Another study reports the physical and functional interaction between RecQ4 and the NER protein XPA[200]. However, the importance for RecQ4 in NER is still under debate as some contradictory results regarding the cellular response of RecQ4 upon UV treatment have been reported[178,200]. Recent studies have further established an important role for RecQ4 in HR[201] and NHEJ[202], the two major mechanisms for the repair of DSBs. Consistently, RecQ4 is recruited to laser-induced DSBs and RTS cells as well as RecQ4-depleted cells exhibit sensitivity to ionizing radiation, which causes DSBs[194,203]. Furthermore, RecQ4 interacts and co-localizes with the HR initiation protein Rad51[193], a key protein in the HR mediated repair of DSBs. A central determinant for sensing and initiating the repair of DNA-breaks and oxidative DNA damages is PARP-1[69]. The C-terminal domain of RecQ4 has been shown to interact with PARP-1, both *in vitro* and *in vivo*, and was

demonstrated to serve as a substrate for PARP-1 mediated PARylation, suggesting a tight functional collaboration between the two proteins in the context of DNA repair[178]. Importantly, RecQ4 depletion has been associated with increased PARP-1 mediated apoptosis in prostate cancer cells[204]. Thus, the RecQ4/PARP-1 interaction represents an interesting target for cancer therapy. In the context of mitochondrial genome integrity, the interaction between RecQ4 and the tumor suppressor p53 constitutes an active area of study. In unstressed cells, p53 co-localizes with RecQ4 at sites of newly synthesized mtDNA and RecQ4 as well as p53 have been shown to exhibit a collaborative function in both, mtDNA replication and maintenance[174,175,205]. As mentioned above, RecQ4 is the only RecQ helicase, which has been found to localize within mitochondria, utilizing its MLS within the N-terminal 84 amino acids[174]. Since p53 lacks its own MLS, it has been proposed that p53 exploits the interaction with RecQ4 to translocate into the mitochondrial nucleoids via formation of the RecQ4/p53 protein complex. The interaction sites of both proteins were determined to conceal their respective NTS, thereby preventing the nuclear retention of p53 and RecQ4 simultaneously in favor of mitochondrial translocation via the MLS of RecQ4. Features of premature aging and senescence as well as susceptibility to cancer have been correlated with increased mutation rates in mtDNA[175] and are key characteristics in RTS patients, strengthening the notion that a dysfunctional interplay between RecQ4 and p53 could be an explanation for the development of these phenotypes at the molecular level. In addition to their collaborative effort in mtDNA maintenance, RecQ4 has been shown to modulate p53-function during skeletogenesis[206], which is underscored by the prominent skeletal abnormalities of RTS, RAPADILINO and BGS patients. Collectively, almost two decades of RecQ4 research have established a comprehensive and multifunctional role for RecQ4 in the maintenance of genomic integrity, whereas the precise contribution of RecQ4 in the molecular mechanisms that warrant genome integrity is still far from being fully understood. In line with most other RecQ helicases, RecQ4 upregulation has a tumor promoting function, specifically in prostate-, cervical- and breast

cancers and, consequently, depletion of RecQ4 in these cells resulted in reduced growth and survival of the cancer cells[44,204,207]. Moreover, the *recq4* gene location on chromosome 8q24 is a site that is frequently amplified in sporadic breast cancers[208,209], thus highlighting the oncogenic potential of RecQ4 as well as the opportunity to develop a targeted inhibitor against RecQ4 that may counteract breast-, cervical- and prostate cancers.

1.5.5. RecQ5, the transcriptional stress RecQ helicase

Together with RecQ4, RecQ5 was first cloned in 1998[173]. It is expressed in three alternative splice forms, termed RecQ5α, RecQ5β and RecQ5γ[210]. The α- and γ-isoforms exhibit C-terminal truncations and solely harbor the two RecA-like domains[211]. Consequently, both isoforms are located within the cytoplasm and do not show dsDNA strand separation but annealing activity *in vitro*[210,211]. RecQ5β (from here on referred to as RecQ5) is the main isoform, with a significantly higher expression level compared to the α- and γ-isoforms. It is localized in the nucleus and features the helicase core, the RQC-Zn and the entire RecQ5 C-terminus, which is required for helicase activity and further RecQ5-specific cellular functions[212] (Fig. 1-11).

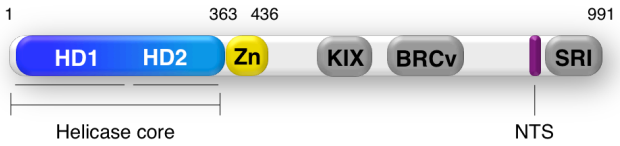


Figure 1-11: RecQ5. Domain architecture and boundaries of the human RecQ5 helicase.

Similar to RecQ1, RecQ5 is expressed ubiquitously and cell cycle independent[43] and functions in DNA replication, recombination, chromosome segregation and DNA repair. Furthermore, as with RecQ1, RecQ5 dysfunction has no implications in human pathogenesis. Functionally, RecQ5 is unique as it is the only RecQ helicase to interact with RNA polymerase II, thus operating at the interface of DNA replication and transcription[213,214]. A

central component for its function in transcription is the unique C-terminus, which features the kinase-inducible- (KIX) and the Set2-Rpb1-interacting (SRI) domains, responsible for the interaction with RNA Pol II[213,215]. The KIX domain interacts with the hypo-phosphorylated RNA Pol IIa, regulating its association with chromatin and thereby suppresses the initiation of transcription. In contrast, the SRI domain interacts with the hyper-phosphorylated form, RNA Pol IIo, inhibiting transcription elongation[214]. Consequently, knockdown of RecQ5 increases the fraction of chromatin bound RNA Pol II as well as the transcription of several genes[216], whereas overexpression of RecQ5 reduces chromatin-bound RNA Pol II[217]. As the RecQ5/RNA Pol II interaction is especially enhanced during S-phase, it has been suggested that RecQ5 regulates transcription events during DNA replication and facilitates DNA integrity by preventing head-on collisions between transcription bubbles and replication forks[217,218]. In contrast to other RecQ helicases, RecQ5 only unwinds forked duplexes, yet is incapable to resolve D-loops, G4 DNA and HJs[219,220]. In DNA replication, RecQ5 interacts with Fen1 and PCNA, promotes strand exchange and the regression of stalled replication forks[221,222]. The interaction between RecQ5 and Topoisomerase IIa has been shown to be important for the segregation of chromosomes after replication, as RecQ5 knockdown cells exhibit undercondensed and entangled chromosomes[223]. Furthermore, RecQ5 acts in HR, disrupting the Rad51 nucleoprotein complex formation and inhibiting MRE11 exonuclease activity, which highlights RecQ5s role as an anti-recombination factor[224,225]. Thus, similar to BLM knockdown cells, RecQ5 deficient cells exhibit increased SCEs, yet a double depletion is more severe than a deficiency of either BLM or RecQ5 alone, suggesting non-redundant functions of both proteins in HR[224]. Furthermore, RecQ5 may be involved in DSB repair, as RecQ5 depleted cells accumulate DSBs and are sensitive to gamma irradiation[226]. RecQ5 is actively recruited to DSBs, where it is associated with the MRN complex, a major sensor of DSBs, which promotes DSB repair and ATM mediated DNA damage signaling[226,227]. Other than in DSB repair, RecQ5 might also function in ssDNA break repair and BER, two

DNA repair pathways that share several key proteins. Indicative for the involvement of RecQ5 in these processes are the protein interactions with Fen1, PCNA and PARP-1[221,228]. Moreover, RecQ5 depleted human cells exhibit increased PAR levels and accumulation of endogenous DNA damages, which are mainly BER substrates[229]. As an exception from the rule, RecQ5 is the only RecQ helicase, which is not upregulated in cancer cells. However, RecQ5 confers resistance to CPT induced replication stress and maintains genome integrity as shown by RecQ5 knockout mice, which exhibit increased susceptibility to develop cancer[43,225]. In addition to preventing replicative and transcriptional stress, it has been proposed that RecQ5 may provide backup capabilities for some of the functions of other RecQ helicases. The shared anti-recombination function with BLM is one such example, where the depletion of either protein has similar consequences, e.g. the manifestation of increased SCEs and multiradial chromosomes[106,225]. Similarly, RecQ5 depletion is synthetically lethal with loss of WRN[230]. Both proteins physically and functionally interact with each other at stalled replication forks, where RecQ5 stimulates the helicase activity of WRN. Furthermore, both proteins mutually interact with Fen1, the MRN complex and PCNA, augmenting a functional overlap[222,231]. Lastly, RecQ5 shows a unique behavior regarding its annealing activity. Like RecQ4, RecQ5 is a rather weak helicase, yet displays a very potent annealing activity[42]. However, in contrast to all other human RecQs, DNA strand annealing by RecQ5 is not inhibited by ATP or RPA[42]. In general, it has been suggested for RecQ helicases that oligomerization could be the major switch between annealing- and helicase activity and that ATP binding typically favors monomer/dimer formation of RecQ helicases, thus promoting their DNA strand separation activity[232,233]. RecQ5, however, was found to be mainly monomeric, regardless of the ADP/ATP status. Moreover, only RecQ5 facilitates annealing of DNA/RNA hybrids, which may be the functional foundation for RecQ5-mediated RNA Pol IIo stalling during transcription elongation[42].

2. General scope of the thesis project

Structural and functional studies on RecQ helicases continue to provide essential mechanistic insights into this important class of genome maintenance proteins. Especially the characterization of the human RecQ family members contributes to the understanding of the molecular details in cancer development and the aging process. While some of the human RecQ helicases have been intensively studied in the past, others remain less well investigated. One of the rather poorly characterized RecQ members is the human RecQ4 protein, in part because of its challenging behavior in recombinant expression and purification. However, RecQ4 in particular represents an interesting target of study, as it seems to possess a unique structural composition compared to other RecQ family members, which might hold the key to understanding its distinctive individual functions within the genome maintenance systems.

The chief aim of this thesis project thus constituted a detailed structural and functional characterization of the human RecQ4 helicase. Structural information was obtained via protein X-ray crystallography, which necessitated the prior establishment of suitable protocols for recombinant protein expression and purification in order to meet the quantitative and qualitative requirements of protein crystallization. The functional analysis was performed *in vitro* via biochemical methods, targeting the analysis of helicase- and ATPase activities as well as the binding behavior to different DNA substrates. The combination of functional and structural studies made it possible to obtain valuable mechanistic insights into RecQ4s helicase mechanism and permitted a more thorough analysis of patient mutations that are causative for RecQ4 associated diseases.

This page has been left blank intentionally

3. Material and Methods

3.1. Materials

3.1.1. Chemicals and reagents

Table 3-1: List of chemical and reagents

Substance	Supplier
2'-Deoxyadenosine 5'-triphosphate (dATP), sodium salt solution	New England Biolabs
2'-Deoxycytidine 5'-triphosphate (dCTP), sodium salt solution	New England Biolabs
2'-Deoxyguanosine 5'-triphosphate (dGTP), sodium salt solution	New England Biolabs
2'-Deoxythymidine 5'-triphosphate (dTTP), sodium salt solution	New England Biolabs
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Carl Roth
Acetic acid	Carl Roth
Adenosine 5'-diphosphate sodium salt	Sigma-Aldrich
Adenosine 5'-(β,γ-imido)-triphosphate lithium salt hydrate	Sigma-Aldrich
Adenosine 5'-triphosphate disodium salt hydrate	Sigma-Aldrich
Agarose NEEO ultra quality	Carl Roth
Ammonium persulfate (APS)	Carl Roth
Ampicillin sodium salt	Carl Roth
Acrylamide/Bis-acrylamide (37.3:1)	Carl Roth
Bromphenol blue	Carl Roth
Chloramphenicol	Carl Roth
Coomassie Brilliant Blue G-250	Carl Roth
Coomassie Brilliant Blue R-250	Carl Roth
Dithiothreitol (DTT)	Carl Roth
Ethanol	Carl Roth
Ethylenediaminetetraacetic acid (EDTA)	Carl Roth
Glycerol	Carl Roth
Glycine	Carl Roth
Hydrochloric acid (HCl)	Carl Roth
Imidazole	Carl Roth
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Carl Roth
Kanamycin sulfate	Carl Roth
Magnesium chloride hexahydrate	Carl Roth
Methanol	Carl Roth
Nicotinamide adenine dinucleotide (NAD^+/H)	Carl Roth

Nickel(II) sulfate hexahydrate	Carl Roth
Orange G	Sigma-Aldrich
Phospho(enol)pyruvic acid trisodium salt hydrate	Sigma-Aldrich
Sodium chloride (NaCl)	Carl Roth
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxide (NaOH)	Carl Roth
SYPRO® Orange	Sigma-Aldrich
Tetramethylethylenediamin (TEMED)	Carl Roth
Tris-(2-carboxyethyl)-phosphine (TCEP)	Carl Roth
Tris-(hydroxymethyl)-aminoethane (Tris)	Carl Roth
Urea	Carl Roth
Xylene cyanol	Sigma-Aldrich

3.1.2. Consumables

Special consumables that were used in this study are listed below, excluding common equipment such as glass/plastic ware, reaction containers, etc.

Table 3-2: Selected laboratory consumables

Type	Model	Supplier
24-well hanging-drop crystallization plates	Crystalgen SuperClear™ Plate	Jena Bioscience
96-well sitting-drop crystallization plates	Crystalquick™ 1 square well, flat bottom, low profile	Greiner Bio-One
96-well thin-wall PCR plates	Microplate 96 well	Greiner Bio-One
384-well Microplate (ATPase)	µCLEAR®, High binding	Greiner Bio-One
384-well Microplate (Helicase, FP)	FLUOTRAC, High binding	Greiner Bio-One
Centrifugal Device	Amicon® Ultra-2 mL	Merck Millipore
Centrifugal Device	Macrosep® Advance	PALL
Cuvettes	Rotilabo®-single-use	Carl Roth
Cover slides (22 mm)	Siliconised	Jena Bioscience
Dialysis membranes	Spectra/Por®	Spectrum Laboratories
Optical quality sealing foil	VIEWseal™	Greiner Bio-One

3.1.3. Manufactured compounds, kits and enzymes

Table 3-3: Manufactured compounds, kits and enzymes

Designation	Supplier
2xYT (double yeast and tryptone) medium (bacterial growth)	Carl Roth
Bayer silicon grease medium viscosity	Jena Bioscience
Bovine Serum Albumin (BSA)	New England Biolabs
cOmplete EDTA-free protease inhibitor tablets	Roche
DNaseI	Invitrogen
Dpn1	New England Biolabs
GC buffer (PCR)	New England Biolabs
GeneRuler™ 1 kb DNA Ladder	Thermo Fisher Scientific
HF buffer (PCR)	New England Biolabs
Lactate dehydrogenase (LDH) / Pyruvate kinase (PK) mix	Carl Roth
Lysogeny broth (LB) medium (bacterial growth)	Carl Roth
Lysozyme	Carl Roth
Midori green Advance DNA stain	Biozym Scientific
NEBuffer™ 2 (Cloning)	New England Biolabs
NEBuffer™ 4 (Cloning)	New England Biolabs
Nucleospin Gel and PRC cleanup kit	Macherey-Nagel
Nucleospin Plasmid kit	Macherey-Nagel
PageRuler™ Prestained Protein Ladder	Thermo Fisher Scientific
Phusion high fidelity DNA polymerase	Thermo Fisher Scientific
PreScission (3C) protease	In-house production
T4 DNA Ligase Reaction Buffer	New England Biolabs
T4 Ligation buffer	New England Biolabs
Tobacco Etch Virus (TEV) protease	In-house production
Taq DNA Polymerase	New England Biolabs
Terrific Broth (TB) medium (bacterial growth)	Carl Roth

3.1.4. Scientific Equipment

Table 3-4: Scientific equipment

Type	Model	Supplier
Agarose gel electrophoresis system	Mini-Sub® Cell GT System	Bio-Rad Laboratories
Cell disruption system	M-110P	Microfluidics
Crystallography: cryo-loop	CryoLoop	Hampton Research
Crystallography: sample holder	CrystalCap™ Magnetic	Hampton Research
Crystallography: sample vial	CryoVial	Hampton Research
Crystallography: handling tool	CrystalWand™ Magnetic	Hampton Research
Crystallography: storage pucks	SPINE Puck	Jena Bioscience
FPLC system (Protein purification)	ÄKTA pure 25	GE Healthcare
FPLC system (MALS)	ÄKTApurifier	GE Healthcare
Liquid handling robot	Honeybee 963	Digilab
Microscope	SteREO Discovery.V12	ZEISS
Microscope camera	AxioCam MRc	ZEISS
Microplate reader	CLARIOstar®	BMG LABTECH
MALS detector	DAWN® 8 + HELEOS® II	Wyatt Technology
Molecular Imager	PharosFX™ Plus	Bio-Rad Laboratories
Polyacrylamide gel electrophoresis	Mini-PROTEAN system	Bio-Rad Laboratories
PCR-cycler	Mastercycler® EPgradient S	Eppendorf
Real-time (RT) PCR cycler	Mx3005P	Agilent Technologies
Refractometer	Optilab T-rEX	Wyatt Technology
Robotic sealing unit for microplates	RoboSeal	HJ-BIOANALYTIC
Rotor (6x 250 mL)	JLA 16.250	Beckman Coulter
Rotor (8x 50 mL)	JA-25.50	Beckman Coulter
Rotor (4 x 2250 ml)	JS-5.0	Beckman Coulter
Spectrophotometer	BioPhotometer	Eppendorf
Spectrophotometer	NanoDrop ND 1000	Peqlab
UV imaging system	Gel Doc™ XR System	Bio-Rad Laboratories
X-ray cryosystem	X-Stream™ 2000	Rigaku
X-ray detector	R-AXIS HTC	Rigaku
X-ray generator	MicroMax™-007 HF	Rigaku
X-ray optics	VariMax™	Rigaku

3.1.5. Bacterial strains and vectors

Table 3-5: Bacterial strains

Organism	Strain	Usage	Supplier
E. coli	Dh5α	Cloning, plasmid amplification	Invitrogen
E. coli	BL21star (DE3)	Protein expression	Life Technologies
E. coli	Rosetta2 (DE3)	Protein expression	Novagen
E. coli	ArcticExpress (DE3) RIL	Protein expression	Agilent Technologies
E. coli	ArcticExpress (DE3) RP	Protein expression	Agilent Technologies

Table 3-6: Vectors used for protein expression

Vector	Host	Structure/Comments	Resistance	Supplier
pGEX-6p1	Bacterial	GST-[3C]-POI cDNA source for human RecQ4	Ampicillin	Gift from Vilhelm A. Bohr, NIA, Baltimore, MD, USA
pETM-11	Bacterial	His_{6x}-[TEV]-POI, alternatively: POI-His_{6x}	Kanamycin	EMBL, Hamburg
pETM-14	Bacterial	His_{6x}-SUMO/smt3-[3C]-POI	Kanamycin	EMBL, Hamburg
pETM-22	Bacterial	TrxA-His_{6x}-[3C]-POI	Streptomycin	EMBL, Hamburg

3.1.6. Primer

All primer for molecular cloning, site-directed mutagenesis and sequencing were purchased from Sigma-Aldrich. SLIC cloning of RecQ4 variants into the pETM-14 or pETM-22 vectors permitted the use of identical primer pairs, as in both cases the RecQ4 inserts were placed between the sequence of the 3C protease cleavage site and the BamH1 restriction sequence.

Table 3-7: Primer list

Construct		Primer sequence (5'-3')
pETM-11	F	GGCGCCCTGAAAATAAAG
vector lin.	R	CGAATTCGAGCTCCGTC
pETM-11	F	AATCTTTATTTTCAGGGCGCCATGGAGCGGCTGCGGGACGTGCGG
FL_RecQ4	R	TGTCGACGGAGCTCGAATTCGTCAGCGGGCCACCTGCAGGAGCTC
pETM-11	F	TTTCAGGGCGCCATGGGGCAGTTGGCAGAG
Core	R	TCACCTTCTGGGTCCTGGTGGCCTAGGCTTAAG
pETM-11	F	TTTCAGGGCGCCATGGGGCAGTTGGCAGAG
Helc	R	TCAGCGGGCCACCTGCTGGCCTAGGCTTAAG
C-His pETM-11	F	AATCTTTATTTTCAGGGCGAGGAGCACCACCACCAC
vector lin.	R	GGTATATCTCCTTCTTAAAGTTAAATCAAAATTATTTCTAGAGGG
pETM-11	F	ACTTTAAGAAGGAGATATACCATGGAGCGGCTGC
FL_CHis	R	CTCGCCCTGAAAATAAAGATTCTCAGTAGTGGGGCGGGCCACCTG
pETM-11	F	ACTTTAAGAAGGAGATATACCATGGGGCAGTTGGCAGAG
Core_CHis	R	CTCGCCCTGAAAATAAAGATTCTCAGTAGTGGGCCTTCTGGGTCCTGGG
pETM-11	F	ACTTTAAGAAGGAGATATACCATGGGGCAGTTGGCAGAG
HelC_CHis	R	CTCGCCCTGAAAATAAAGATTCTCAGTAGTGGGGCGGGCCACCTG
pETM-14/22	F	GGATCCGAATTCGAGCTC
vector lin.	R	GGGCCCCTGGAAC
pETM-14/22	F	GAAGTTCTGTTCCAGGGGCCCATGGAGCGGCTGC
FL_RecQ4	R	CGACGGAGCTCGAATTCGGATCCTCAGCGGGCCAC
pETM-14/22	F	GAAGTTCTGTTCCAGGGGCCCATGGGGCAGTTGGC
HelC	R	CGACGGAGCTCGAATTCGGATCCTCAGCGGGCCAC
pETM-14/22	F	GAAGTTCTGTTCCAGGGGCCCATGGGGCAGTTGGC
Core	R	CGACGGAGCTCGAATTCGGATCCTCACCTTCTGGGTCCTG
pETM-14/22	F	GTTCTGTTCCAGGGGCCCATGGCTGTTGGGCCTGAG
$RecQ4^{427-997}$	R	CTCGAATTCGGATCCTCAGTCCTCAGGCAGCTG
pETM-14/22	F	GTTCTGTTCCAGGGGCCCATGGTGCTGCCACTCTACTC
$RecQ4^{450-997}$	R	CTCGAATTCGGATCCTCAGTCCTCAGGCAGCTG
pETM-14/22	F	GTTCTGTTCCAGGGGCCCATGGCTGTTGGGCCTGAG
$RecQ4^{427-1116}$	R	CTCGAATTCGGATCCTCACGGCTCCTGCCC
pETM-14/22	F	GTTCTGTTCCAGGGGCCCATGGTGCTGCCACTCTACTC
$RecQ4^{450-1116}$	R	CTCGAATTCGGATCCTCACGGCTCCTGCCC
K508A	F	CCTACAGGTGCCGGCGCGTCCCTGTGCTACCAG
Walker A	R	CTGGTAGCACAGGGACGCGCCGGCACCTGTAGG
D6105A	F	GCTTTTGCCTGCATTGCTGAGGCCCACTGCCTC
Walker B	R	GAGGCAGTGGGCCTCAGCAATGCAGGCAAAAGC
pETM-14/22	F	GTTCTGTTCCAGGGGCCCATGGCTGTTGGGCCTGAG
$RecQ4^{427-1208}$	R	CTCGAATTCGGATCCTCAGCGGGCCACCTG

R1021W	F	CAGGACAGGTGTGTGGCGTGGGACAG
	R	CTGTCCCACGCCACACACCTGTCCTG
W613L	F	CTGCCTCTCCCAGCTGTCCCACAACTTC
	R	GAAGTTGTGGGACAGCTGGGAGAGGCAG
H615L	F	CCCAGTGGTCCCTGAACTTCCGGCCC
	R	GGGCCGGAAGTTCAGGGACCACTGGG
F617L	F	GTCCCACAACTTACGGCCCTGCTAC
	R	GTAGCAGGGCCGTAAGTTGTGGGAC
R618L	F	CCCACAACTTCCTGCCCTGCTACCTG
	R	CAGGTAGCAGGGCAGGAAGTTGTGGG
RecQ4$^{427\text{-}1090}$	F	CGGGCCCTGCCTGGAGTGAGGATCCGAATTCGAGCTCCGTCGACAAGC
	R	CTCCAGGCAGGGCCCGCAG
del944-1044	F	ACCACCTATACCGGAGGTGAGCTGGCCTTCCACCTTCGCA
	R	AGGCCAGCTCACCTCCGGTATAGGTGGTCGCCAGCAGCTC
R894A/R895A	F	CAAGCAGCCCCAGGACCCGCGGCGGTCTGCATGGGCCATGAG
	R	TCATGGCCCATGCAGACCGCCGCGGGTCCTGGGGCTGCTTGG
K1048A	F	CCGCTGAGGAGGCGGACCAGATATG
	R	ATATCTGGTCCGCCTCCTCAGCGGT
R844A	F	TGGCTGTGAAGGCGCTGGTACAGCG
	R	GCTGTACCAGCGCCTTCACAGCCAG
R848A*	F	GTGAAGGCGCTGGTACAGGCGGTGTTCCCAGCCTGC
	R	GCAGGCTGGGAACACCGCCTGTACCAGCGCCTTCAC
K843A**	F	CGGACTTCCTGGCTGTGGCGGCGCTGGTACAGG
	R	CCTGTACCAGCGCCGCCACAGCCAGGAAGTCCG
R844E***	F	CTGGCTGTGGCGGAGCTGGTACAGGCGG
	R	CCGCCTGTACCAGCTCCGCCACAGCCAG
SeqT7	F	TAATACGACTCACTATAGGG (Sequencing)
SeqATG	F	ATGGAGCGGCTGCG (Sequencing)
Seq7	F	CCTAGATCCTGGCTGGTTACAGC (Sequencing)
Seq14	F	GGTGGTGCCACAGTCACAACC (Sequencing)
Seq21	F	GCTTTTGCCTGCATTGATGAGGCC (Sequencing)
Seq28	F	GAGAGCTACGTGCAGGCCGTG (Sequencing)
Seq35	F	ACTCCATGGGCTGGGAGCTGG (Sequencing)

* introduced in the R844A template

** introduced in the R843A/R848A template

*** introduced in the K843A/R844A/R848A template

3.1.7. Protein purification equipment

Table 3-8: Equipment used for protein expression

Designation	*Type*	*Supplier*
Econo-Column®	Column body, batch-purification	Bio-Rad Laboratories
HiLoad™ 16/60 Superdex™ 200	Preparative SEC FPLC column	GE Healthcare
HiLoad™ 26/40 Column Housing	Preparative Heparin FF Sepharose column	GE Healthcare
HiPrep™ Heparin FF 16/10 prepacked	Prepacked Heparin Sepharose 6 FF affinity chromatography column	GE Healthcare
HiTrap™ Heparin HP 5mL	Heparin affinity chromatography	GE Healthcare
Heparin Sepharose 6 FF	Heparin affinity chromatography medium	GE Healthcare
HisTrap™ FF crude 5 mL	Immobilized metal-ion affinity chromatography FPLC column	GE Healthcare
MonoQ™ 5/50 GL	High-resolution AIEX FPLC column	GE Healthcare
Protino® Ni-IDA	Immobilized metal-ion affinity chromatography resin	MACHEREY-NAGEL
Protino® Ni-NTA	Immobilized metal-ion affinity chromatography resin	MACHEREY-NAGEL
Protino® Ni-TED	Immobilized metal-ion affinity chromatography resin	MACHEREY-NAGEL
Superdex™ 200 Increase 10/300 GL	Analytical SEC FPLC column	GE Healthcare

3.1.8. Crystallization screens

Table 3-9: Commercially available crystallization screens

Name	Supplier
Additive Screen™	Hampton Research
Crystal Screen, Crystal Screen 2	Hampton Research
Index	Hampton Research
JCSG+	Molecular Dimensions
Optimix™ 3	Fluidigm
Optimix™ PEG	Fluidigm
PEGs Suite, PEGs II Suite	Qiagen
pH Clear Suite, pH Clear II Suite	Qiagen
Protein Complex Suite	Qiagen
Silver Bullets Bio™	Hampton Research
Wizard 1+2, Wizard 3+4	Emerald BioSystems

3.1.9. Software and databases

Table 3-10: Software and databases

Program	Description	Supplier/Reference
AIMLESS	Scaling and merging of diffraction data	Evans & Murshudov, 2013
APBS	Electrostatic surface potential calculation	Baker et al., 2001
Astra VI	MALS control and data analysis	Wyatt
AxioVision	Microscopy image recording software	ZEISS
Buccaneer	Automated crystallographic model building	Cowtan, 2006
CCP4	Software suite for macromolecular X-ray structure determination	Winn et al., 2004
CCP4i	Graphical interface to *CCP4*	Potterton et al., 2003
Clustal Omega	Multiple sequence alignment tool	Sievers & Higgins, 2014
Coot	Model-building software, X-ray crystallography	Emsley et al., 2010
CrystalClear	X-ray data collection and basic processing	Rigaku
Crank2	SAD phasing pipeline, including AFRO, CRUNCH2 and REFMAC	Pannu et al., 2011; de Graaff et al., 2001; Murshudov et al. 2011
DaliLite v. 3	Structural homology search	Holm L, Rosenström P, 2010
DSSP	Secondary structure assessment calculated from coordinate files	Touw et al., 2015

ExPASy ProtParam tool	Computation of physical and chemical properties of proteins	Artimo et al., 2012
ExPASy Translate tool	Translation tool of nucleotide sequences to protein sequences	Artimo et al., 2012
MARS	Clariostar data analysis software	BMG LABTECH
Microsoft Excel	Spreadsheet software	Microsoft Corporation
MolProbity	Structure validation for macromolecular crystallography	Chen et al., 2010
MXCube2	X-ray data collection GUI	Gabadinho et al., 2010
ODYSSEY	Infrared imaging software	LI-COR
Origin Pro	Graphics and data analysis software, V 8.6	OriginLab
PARROT	Automated electron density modification	Cowtan, 2010
PDB2PQR	PQR file generation (for APBS)	Dolinsky et al., 2004
PDQuest	2D analysis software (EMSA)	Bio-Rad Laboratories
PHASER	Phasing software	McCoy et al., 2007
Phenix	Software suite for macromolecular X-ray structure determination	Adams et al., 2010
Phyre2	Biosequence analysis; protein 3D-structure prediction	Kelley et al., 2015
PROPKA	Estimation of protonation states (for APBS)	Li et al., 2005
PyMOL	3-dimensional visualisation and graphical illustration software	Open Source, V1.7.6
Quantity One®	UV imaging system control; UV image recording and analysis	BioRad
UNICORN	FPLC instrument control; recording, analysis and management of chromatograms	GE Healthcare
XDS	Indexing, and integration of diffraction images	Kabsch, 2010

3.2. Methods

Several methods that were used in this study have been originally described in the thesis-associated publication by Kaiser *et al.* (2017), which presents and discusses the primary results of the thesis project at hand. Therefore, the methods described below represent an original excerpt from the published manuscript, extended or adapted to account for the additional methods that have been exclusively presented here.

3.2.1. Molecular Biology methods

3.2.1.1. Cloning and mutagenesis

First-generation RecQ4 variants, encompassing the human cDNA RecQ4 sequences (UniProtKB O94761) corresponding to full-length RecQ4, aa 460-895 and aa 460-1208, were cloned into the pETM-11 vector (EMBL Heidelberg, Germany) between the sequence of the Tobacco Etch Virus (TEV) cleavage site and the BamH1 restriction site using the sequence-and-ligation-independent-cloning (SLIC) method. Shortly, the SLIC cloning method is based on the separate preparation of the linearized plasmid and RecQ4 insert fragments. The respective RecQ4 insert and the linearized vector feature overlapping DNA sequences at the 3' and 5' ends. The dsDNA-overlap segments are transformed into 5' ssDNA overhangs via treatment with T4 polymerase, utilizing its 3'-5' exonuclease activity. Insert and vector are then annealed based on their overlapping single-stranded sequences and thus the RecQ4 insert is incorporated into the vector. The annealed construct is transformed into the *E. coli* Dh5α strain, which subsequently repairs DNA gaps and ligates the two fragments via its host DNA repair machinery. The resulting pETM-11 RecQ4 expression variants featured an N-terminal hexahistidine (His_{6x}) purification-tag, followed by the TEV cleavage site and the particular RecQ variant. The generation of the C-terminal His_{6x}-tag RecQ4 expression variants was based on the same

principle with the difference that RecQ4 insert sequences featured an additional C-terminal TEV sequence, which was subsequently positioned between the RecQ4 C-terminus and the C-terminal His_{6x}-tag. The sub-cloning of those first-generation RecQ4 variants into the pETM-14 vector followed the general SLIC approach as described for pETM-11. pETM-14 RecQ4 expression variants featured an N-terminal His_{6x}-tag, followed by a SUMO/smt3 solubility tag and a PreScission (3C) protease cleavage site. The C-terminal deletion variants $RecQ4^{460-1047}$ and $RecQ4^{460-1115}$ were created based on the pETM-14 $RecQ4^{460-1208}$ template, using a one-step SLIC reaction. In those cases, vector linearization was performed with forward primers that already featured a respective RecQ4 sequence overlap, corresponding to the particular C-terminal end of the RecQ4 variant (e.g. ending at aa 1047 or 1115, respectively), thus creating respective complementary overhangs during the ($RecQ4^{427-1208}$-containing) vector linearization. ssDNA overhang generation and subsequent annealing of the linearized vector thus resulted directly in the pETM-14 $RecQ4^{460-1047}$ and $RecQ4^{460-1115}$ variants without the generation of corresponding insert sequences. The final third-generation RecQ4 variants, encompassing the RecQ4 cDNA sequences of FL_RecQ4, aa 450-997, aa 427-997, aa 450-1116, aa 427-1116, aa 427-1208 and aa 427-1090 were cloned into the pETM-22 vector (EMBL, Heidelberg, Germany) between the 3C sequence and BamH1 restriction site, using the standard SLIC method, e.g. by separate insert amplification and vector linearization. All resulting constructs featured an N-terminal thioredoxinA (Trx) solubility tag, followed by a His_{6x}-tag and the 3C protease cleavage site. The pETM-22 $RecQ4^{427-1208}$ delH944-S1032 variant (upper WH domain deletion) was created based on the one-step SLIC principle as described above, using a vector-linearization primer that incorporated a double glycine-linker sequence between aa 943 and aa 1033. Single amino acid substitutions were generated according to the standard QuickChange site-directed mutagenesis protocol (Stratagene). The success of each PCR reaction (insert amplification and vector linearization) was standardly analyzed via 1.5 % agarose gel-electrophoresis.

The correct nucleotide sequence of all RecQ4 expression variants was verified by DNA sequencing (Eurofins Genomics, Ebersberg, Germany).

Table 3-11: RecQ4 variants for protein expression in *E. coli*

No.	Variant name	Vector	Insert	Comment
1	FL_RecQ4	pETM-11	1-1208	Full-length RecQ4
2	RecQ4_Core	pETM-11	460-895	Helicase core
3	RecQ4_HelC	pETM-11	460-1208	Helicase core + C-terminus
4	FL_CHis	pETM-11	1-1208	Full-length RecQ4
5	Core_CHis	pETM-11	460-895	Helicase core
6	HelC_CHis	pETM-11	460-1208	Helicase core + C-terminus
7	FL_RecQ4	pETM-14	1-1208	Full-length RecQ4
8	RecQ4_Core	pETM-14	460-895	Helicase core
9	RecQ4_HelC	pETM-14	460-1208	Helicase core + C-terminus
10	RecQ4_1047	pETM-14	460-1047	-
11	RecQ4_1115	pETM-14	460-1115	-
12	FL_RecQ4	pETM-22	1-1208	Full-length RecQ4
13	RecQ4$^{427\text{-}997}$	pETM-22	427-997	-
14	RecQ4$^{450\text{-}997}$	pETM-22	450-997	-
15	RecQ4$^{427\text{-}1116}$	pETM-22	427-1116	Crystallized RecQ4 variant
16	RecQ4$^{450\text{-}1116}$	pETM-22	450-1116	-
17	RecQ4$^{427\text{-}1208}$	pETM-22	427-1208	-
18	K508A	pETM-22	427-1208	Walker A mutant
19	D605A	pETM-22	427-1208	Walker B mutant
20	Del_H944-S1032	pETM-22	427-1208	Upper WH domain deletion
21	R894A/R895A	pETM-22	427-1208	Lower WH domain substitutions
22	K1048A	pETM-22	427-1208	Lower WH domain substitution
23	K843A/R844A/R848A	pETM-22	427-1208	Lower WH domain substitutions
24	K843A/R844E/R848A	pETM-22	427-1208	Lower WH domain substitutions
25	W613L	pETM-22	427-1208	ARL mutation
26	H615L	pETM-22	427-1208	ARL mutation
27	F617L	pETM-22	427-1208	ARL mutation
28	R618L	pETM-22	427-1208	ARL mutation
29	RecQ4$^{427\text{-}1090}$	pETM-22	427-1090	RAPADILINO mutation
30	R1021W	pETM-22	427-1208	BGS mutation

3.2.1.2. Protein expression and purification

Any plasmid, coding for a particular RecQ4 variant, was transformed into an *E. coli* expression strain via the standard heat-shock transformation method. These cells were then incubated for 1 hour at 37°C with 700 µL 2xYT (two times yeast/tryptone) sterile bacterial growth medium, and plated out on an LB-agar plate, containing the appropriate antibiotic(s). After a 16-18 h incubation period at 37°C, a single colony was used to inoculate a 2xYT over-night starter culture, containing the appropriate antibiotic(s).

Test expressions

The typical volume of a starter culture for test expressions was 10 mL. Expression tests were performed in a 200 mL format, using 2xYT bacterial growth medium. 2 mL of the over-night starter culture were used to inoculate the main culture, supplemented with the appropriate antibiotic(s). After a growth period at 37°C, recombinant protein expression was initiated as soon as the bacterial culture had reached an optical density (OD) of approximately 0.6 by addition of IPTG to the desired final concentration (typically 0.1 – 1.0 mM). Subsequently, the culture flasks were transferred to the desired expression temperature. Test expressions at 37°C typically lasted for 6-8 hours, whereas expression at lower temperatures were usually performed over-night (15-20 hours). In case of the ArcticExpress expression strains, induction as well as recombinant protein expression was performed at a temperature of 10-13°C for a duration of 24 hours. After recombinant protein expression, the entire cell pellet was lysed and batch-purified via immobilized-metal-ion-chromatography (IMAC) and size-exclusion-chromatography (SEC), if applicable, to analyze the expression and solubility of the desired RecQ4 variant.

Large-scale protein production

After completion of the testing phase, large-scale expression was typically performed in a 2 L culture volume, initially using 2xYT medium during the

early stages of the thesis project, which was eventually replaced by Terrific Broth (TB). A single-colony over-night culture was used to generate the starter culture, which typically comprised a volume of 100 mL 2xYT medium, supplemented with the appropriate antibiotic(s). 20 mL of this starter culture were used to inoculate the main culture, containing the appropriate antibiotic(s). Recombinant protein expression from bacteria grown in 2xYT medium was typically induced at an OD of 0.6, whereas the expression from bacteria grown in TB medium was induced at an OD of 2.5 by addition of the desired IPTG concentration. After the final optimization of the recombinant protein expression and purification methods, all RecQ4 variants were standardly expressed in TB medium, utilizing the BL21Star (DE3) *E. coli* strain (Novagen), which was supplemented with the pRARE2 plasmid from the Rosetta2 *E. coli* strain (Life Technologies). Cells were grown at 37°C to an OD of 2.5 (OD of 0.6 in case of 2xYT medium), transferred to 18°C and recombinant protein expression was induced by the addition of 0.1 mM IPTG. After 18 h of incubation, cells were harvested by centrifugation at 4000 x g for 20 minutes and either used immediately for cell lysis and protein purification or aliquoted in 50 mL tubes (containing 40-45 g cell pellet per tube) and stored at minus 80°C.

Cell lysis

For each purification batch, 40-45 g cell pellet was resuspended in 10 volumes of lysis buffer, complemented with two tablets cOmplete EDTA-free protease inhibitor (Roche), 1 mg mL^{-1} lysozyme and 300 U DNAseI and stirred for 1 hour at 4°C. All following steps were carried out at 4°C or on ice unless stated otherwise. The cell-suspension was lysed using the M-110P microfluidizer system (Microfluidics) according to the manufacturers instructions, which disrupts bacterial membranes on the basis of the Frech pressure cell press system. The resulting crude extract was centrifuged for 1 h at 38,000 x g.

Protein purification

After centrifugation, the supernatant was loaded on a pre-equilibrated 100 mL Heparin FF sepharose column (Heparin buffer A). The column was washed with 1.1 CV (column volume) Heparin buffer A, supplemented with 30 mM NaCl, and subsequently the protein was eluted with 3.5 CV Heparin buffer B. The eluate was collected and directly applied to a 5 mL HisTrap FF crude column, which was pre-equilibrated with IMAC buffer A. The HisTrap column was washed with 3 CV of IMAC buffer A, supplemented with 25 mM imidazole, and the protein was subsequently eluted using a linear gradient over 5 CV up to a concentration of 200 mM imidazole. Appropriate fractions were pooled and the purification-tag (His_{6x}-SUMO or Trx-His_{6x}) was removed by incubation with 3C protease for at least 3 hours. After purification-tag removal, the protein suspension was concentrated to 2 mL and applied to a Superdex 200 16/600 size exclusion column using the SEC-200 buffer. Peak fractions from the size exclusion column were analyzed via SDS-PAGE, pooled and the salt concentration was reduced to 35 mM NaCl using Zero-NaCl buffer. The protein solution was then applied to a MonoS 5/50 GL column, which was pre-equilibrated with MonoS buffer A. The column was washed with 5 CV of MonoS buffer A and the protein was eluted utilizing a linear gradient over 40 CV to a concentration of 200 mM NaCl. Appropriate fractions were analyzed via SDS-PAGE, pooled and concentrated to 500 µL. As a final polishing step, all RecQ variants were applied to an analytical Superdex 200 Increase 10/300GL size exclusion column, using SEC-100 buffer (e.g. storage buffer). The peak fractions were analyzed via SDS-PAGE, pooled and concentrated. The protein was flash frozen in liquid nitrogen and stored at minus 80°C.

All buffers were prepared with ultrapure H_2O from a TKA GenPure water system, using at least analytical grade chemicals. All buffers were filtered and, except for the cell lysis buffers, degassed prior to usage. Buffers were stored at 4°C for no longer than 4 weeks.

Table 3-12: Composition of buffers for protein expression

	Buffer A	Buffer B
Cell lysis	Low-salt (standard) 50 mM HEPES pH 8.0 50 mM NaCl 20% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	Medium-salt (Expression tests) 50 mM HEPES pH 8.0 300 mM NaCl 20% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP
Heparin affinity	20 mM HEPES pH 8.0 10 mM NaCl 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	20 mM HEPES pH 8.0 1 M NaCl 10% (w/v) Glycerol 15 mM Imidazole 2.5 mM $MgCl_2$ 1 mM TCEP
IMAC	20 mM HEPES pH 8.0 300 mM NaCl 10% (w/v) Glycerol 15 mM Imidazole 2.5 mM $MgCl_2$ 1 mM TCEP	20 mM HEPES pH 8.0 300 mM NaCl 10% (w/v) Glycerol 500 mM Imidazole 2.5 mM $MgCl_2$ 1 mM TCEP
SEC-200	20 mM HEPES pH 8.0 200 mM NaCl 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	
Cation affinity (MonoS)	20 mM HEPES pH 7.2 10 mM NaCl 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	20 mM HEPES pH 7.2 1 M NaCl 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP
SEC-100 (storage buffer)	10 mM HEPES pH 8.0 100 mM NaCl 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	
Zero NaCl	20 mM HEPES pH 7.0 10% (w/v) Glycerol 2.5 mM $MgCl_2$ 1 mM TCEP	

3.2.2. Crystallization and structure determination

3.2.2.1. Protein preparation

RecQ4 variants used for crystallization were diluted to the desired protein concentration using SEC-100 buffer and centrifuged for 30 min at 25,000 x g at 4°C. In case of co-crystallization attempts with a DNA substrate, protein and DNA were mixed in a molar ratio of 1:1.1, diluted to the desired protein concentration and incubated on ice for 1 hour before centrifugation.

3.2.2.2. Screening for crystallization conditions

Protein crystallization was performed via the vapor diffusion method in a sealed airtight container, using either the hanging- or sitting-drop setup. In both cases, the protein solution was mixed with a precipitant solution and placed in form of a crystallization drop either on the container lid or on an elevated pedestal over a large volume of precipitant solution (e.g. the mother liquor). The concentration difference of the precipitant between the crystallization drop and the mother liquor leads to an exchange of solvent via evaporation until an equilibration of vapor pressure is reached. The slow evaporation of solvent from the crystallization drop thus results in a steadily increasing concentration of protein and precipitant, which may result in the formation of protein crystals. Initial screening for suitable crystallization conditions was performed in a 96-well sitting drop format utilizing the commercially available screens, which are listed in Table 3-9. The crystallization setups were performed in a high-throughput low-volume format utilizing the HoneyBee 963 crystallization robot, which mixes (per well) 300 nL of protein solution with 300 nL precipitant solution to generate the crystallization drop and adds 40 µL of the precipitant solution to the bottom of each well to generate the mother liquor reservoir. The resulting 96-well crystallization plate was sealed with adhesive film and stored at the desired temperature. After an initial crystallization condition was identified, manual

fine screening was performed in the 24-well hanging drop format by varying the concentration of precipitant and/or protein, the pH or the protein/precipitant ratio. A detailed description of the fine screening approach can be obtained from the experimental results section.

3.2.2.3. Structure determination and model building

Protein crystals used for structure determination were obtained with the hanging-drop vapor diffusion method from crystallization setups containing 2.8-3.0 M sodium formate and 0.1 M imidazole pH 9.0 as precipitant and a protein concentration of 3 mg mL^{-1} with a protein-to-precipitant ratio of 2:1 in the crystallization drop. Crystallization was carried out at 20°C. Crystals were flash-frozen in mother liquor supplemented with 25% (v/v) glycerol. Diffraction data were collected at a temperature of 100 K at the beamlines ID29 (ESRF, Grenoble, France) and ID14.1 (BESSY, Berlin, Germany). Data were integrated and scaled with XDS[234] and Aimless[235]. Trials to solve the structure by molecular replacement were unsuccessful. As the diffraction data indicated the presence of an anomalous scatterer, an X-ray fluorescence scan was performed. This indicated the presence of Zn in the crystal. The Crank2 pipeline[236] was used to solve the structure in space group $P3_121$ by the single wavelength anomalous dispersion (SAD) method with data collected at the Zn-peak wavelength (Table 4-1). A single Zn ion was located with AFRO[237] / CRUNCH2[238]. Several cycles of phasing, density modification and chain-tracing with REFMAC[239], PARROT[240] and Buccaneer[241] resulted in an initial model consisting of 581 residues in 17 fragments. This model was manually completed with COOT[242]. Restrained refinement of coordinates, B-factors and TLS-parameters was performed with Phenix[243] against data extending to a resolution of 2.75 Å. The final model was refined to an R_{work} of 18.3% and an R_{free} of 23.1% with 97% of the residues in the most favored regions of the Ramachandran plot and none within disallowed regions.

3.2.2.4. Structure analysis and visualization

Analysis of secondary structure elements within the RecQ4$^{427-1116}$ model was performed using the DSSP program. A structure homology search was performed by DaliLite v3, using manually generated PDB files that contained the respective structure coordinates of the RecQ4$^{427-1116}$ domains. Electron density maps were created with the phenix.maps tool. Model images were created in PyMOL (Open source, V1.7.6). Electrostatic surface potentials were calculated with APBS utilizing the AMBER forcefield and visualized using the APBS plugin in PyMOL. Estimation of protonation states for an ionic strength of 150 mM NaCl at pH 7.0 was performed by PROPKA. PQR files were generated via PDB2PQR.

3.3. Biochemical and biophysical methods

3.3.1. Measuring protein concentrations

Estimation of protein concentrations in solution was performed by measuring the absorption of a protein sample at 280 nm, using a Nanodrop ND 1000 spectrophotometer. All protein samples were measured in their unfolded and reduced state. Therefore, 2 µL of the protein sample was diluted with 6 µL of an 8 M urea solution, resulting in the complete denaturation of the protein. Using the molar extinction coefficient for the particular RecQ4 variant (calculated by the ExPASy ProtParam tool), the protein concentration based on the absorption at 280 nm can be determined by application of the Lambert-Beer law:

$$c = A / \varepsilon \times d,$$

where c is the protein concentration [mol x L^{-1}], A is the absorption at 280 nm, ε the molar extinction coefficient [L x mol^{-1} x cm^{-1}] and d [cm] the path length of the light through the sample.

3.3.2. Agarose gel electrophoresis

Analysis of PCR reactions was performed by agarose gel-electrophoresis, using a 1.5% agarose gel. 0.9 g agarose were dissolved in 60 mL TAE buffer, supplemented with 5 µL MidoriGreen DNA stain and poured into an appropriate gel-cast container. After the gel solidified, it was transferred to the electrophoresis chamber and topped with TAE running buffer. 6 µL of each PCR reaction was mixed with 1 µL DNA loading buffer and loaded onto the agarose-gel. Size estimation of the PCR-produced DNA fragments was achieved by applying the separation of a DNA standard in parallel to the PCR samples. Electrophoretic separation was performed at 100 V for 30-40 minutes at room temperature. The separated DNA fragments were visualized by exposing the agarose gel to UV light.

Table 3-13: Agarose gel buffers

	Components
TAE (running) buffer	40 mM Tris/HCl, pH 8.0, 20 mM acetic acid, 1 mM EDTA
DNA loading buffer	10 mM Tris/HCl, pH 7.6, 0.03 % (w/v) bromphenol blue, 0.03 % (w/v) xylene cyanol, 60 % (v/v) glycerol, 60 mM EDTA

3.3.3. SDS-PAGE

Analysis of protein samples was routinely performed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Typically, a 16% separation gel was prepared and topped with a 4% stacking gel. The solidified gel was transferred to the electrophoresis chamber, which was filled with SDS-PAGE running buffer. 16 µL of a protein sample was mixed with 4 µL of SDS-PAGE loading buffer and heated at 95°C for 5 min. Subsequently, 10 µL per sample was loaded onto the SDS-gel. Proteins separated according to their size at 230 V for 50 min at room temperature. Size estimation was achieved by running a protein standard in parallel to the samples. After electrophoretic separation, the gels were washed twice with

water and stained with a Coomassie brilliant blue staining solution for 30-45 min, followed by washing and destaining, using destaining-solution.

Table 3-14: SDS-PAGE gel buffers and components

	Components
SDS-PAGE separation gel (480 mL stock)*	126.2 mL ddH_2O, 224 mL acrylamide/bis-acrylamide mix (37.5:1), 120 mL 1.5 M Tris/HCl pH 8.8, 4.8 mL 10% SDS
SDS-PAGE stacking gel (250 mL stock)*	132 mL ddH_2O, 50 mL acrylamide/bis-acrylamide mix (37.5:1), 62.5 mL 0.5 M Tris/HCl pH 6.8, 2.5 mL 10% SDS
SDS-PAGE loading buffer (5x)	250 mM Tris/HCl pH 6.8, 0.5 M DTT, 0.5% (w/v) bromphenole blue, 10% (w/v) SDS, 50% (v/v) glycerol
SDS-PAGE running buffer (10x stock)	30 g/L Tris/base, 144 g/L Glycine10 g/L SDS
Coomassie staining solution	50% (v/v) methanol, 10% (v/v) acetic acid, 0.1% (w/v) Coomassie R-250, 0.1% Coomassie G-250
Destaining solution	10% (v/v) ethanol, 5% (v/v) acetic acid

* Stock solutions were stored at 4°C. Gel polymerization was initiated by adding 1 vol-% freshly prepared 10% (w/v) APS solution and TEMED (2 µL per gel).

3.3.4. SEC-coupled multi-angle light scattering (SEC-MALS)

Molecular weight determination of the RecQ4$^{427\text{-}1116}$ variant in solution was performed using a size exclusion chromatography coupled multi-angle light scattering setup (SEC-MALS). 2 mg of the RecQ4$^{427\text{-}1116}$ protein (100 µL) were loaded onto a Superdex200 10/300 GL column, which was attached to an ÄKTApurifier FPLC system, equipped with a MALS detector and a refractive index (RI) monitor. Size exclusion chromatography was performed at room temperature in degassed SEC-200 buffer. Analysis of the molecular weight, based on the light scattering and RI signal, was performed using the supplied manufacturer software (Wyatt).

3.3.5. DNA Substrate preparation

Synthesized DNA oligonucleotides (biomers.net) were resuspended in sterile ultrapure water to achieve a concentration of 200 µM and stored at minus 20°C. Oligonucleotide sequences used for DNA substrate preparation are listed in Table 3-15. DNA annealing was performed in 20 mM HEPES pH 8.0 and 100 mM NaCl. 100 µL reactions including 10 µM of each, the Top- and Bottom-strand, were heated to 95°C for 5 min and gradually cooled to 15°C at a rate of 1°C min^{-1} using a PCR-thermal-cycler (Eppendorf). Annealed DNA substrates were stored at 4°C in a light protected container.

Table 3-15: Oligonucleotides for functional studies

Name	Label	Sequence (5' to 3')
T2-Cy3	5' Cy3	CCA TTC CAC CCT CTA
T3-Cy3	5' Cy3	CCA TTC CAC CCT CTA TTT TTT TTT TTT TTT
T7-Cy3	5' Cy3	CCA TTC CAC CCT CTA CCA TTT TTT TTT TTT TTT
T8-Cy3	5' Cy3	CAT CCA TTC CAC CCT CTA CCA TTT TTT TTT TTT TTT
T9-Cy3	5' Cy3	CAT TAC CCA TTC CAC CCT CTA CCA TTT TTT TTT TTT TTT
T11-Cy3	5' Cy3	CAT TAC CTT CCA TTC CAC CCT CTA CCA TCC
T12-Cy3	5' Cy3	CCA TTC CAC CCT CTA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
B1-Dab	3' Dabcyl	TAG AGG GTG GAA TGG
B1	-	TAG AGG GTG GAA TGG
B4-Dab	3' Dabcyl	TTT TTT TTT TTT TTT TAG AGG GTG GAA TGG
B4	-	TTT TTT TTT TTT TTT TAG AGG GTG GAA TGG
B8-Dab	3' Dabcyl	TTT TTT TTT TTT TTT TGG TAG AGG GTG GAA TGG
B9-Dab	3' Dabcyl	TTT TTT TTT TTT TTT TGG TAG AGG GTG GAA TGG ATG
B10-Dab	3' Dabcyl	TTT TTT TTT TTT TTT TGG TAG AGG GTG GAA TGG GTA ATG
B12-Dab	3' Dabcyl	GGA TGG TAG AGG GTG GAA TGG AAG GTA ATG
PolyT30	-	TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT

3.3.6. Electro mobility shift assays (EMSA)

DNA binding analysis via the electro mobility shift assay (EMSA) was performed using the native polyacrylamide gel electrophoresis (native PAGE) system. The binding reaction was performed in a 10 µL volume in the standardized assay buffer (20 mM HEPES pH 8.0, 10 mM NaCl, 5% glycerol, 1 mM $MgCl_2$, 0.5 mM TCEP). Increasing concentrations (0 - 1.0 µM) of a particular RecQ4 variant were incubated at room temperature with 50 nM of the DNA substrate for 15 minutes. Subsequently, the 10 µL binding reaction was supplemented with 2 µL DNA loading buffer and 6 µL per reaction were loaded onto the 4% native PAGE gel. The gel chamber was filled with pre-chilled (4°C) 1x Tris-glycine (TG) buffer and transferred to the cold room. Separation of protein-bound and unbound DNA species was performed at a constant current of 12 mA and a maximum of 150 V until the dye front had reached the lower edge of the gel (approximately 2-3 hours). The EMSA gels were analyzed by fluorescence scanning using the PharosFX™ molecular imaging system (Bio-Rad Laboratories).

Table 3-16: EMSA buffer compositions

	Components
10x Tris-glycine (TG) buffer	200 mM Tris/HCl pH 7.8, 1.9 M glycine
Native-PAGE 4% gel (4 gels)	2 mL 10x TG buffer, 2.7 mL acrylamide/bis-acrylamide mix (37.5:1), 15.1 mL ddH_2O, 200 µL 10% APS, 10 µL TEMED
DNA loading buffer (5x)	50 mM Tris/HCl pH 7.8, 0.15% Orange G, 60% (v/v) glycerol

Table 3-17: EMSA DNA substrates

EMSA Substrate	Top strand	Bottom DNA strand
Y-15-15	T3-Cy3	B4
15nt3'OH	T3-Cy3	B1
15nt5'OH	T2-Cy3	B4
15nt blunt	T2Cy3	B1

3.3.7. Fluorescence polarization assay

Fluorescence polarization (FP) measurements were performed in a 100 µL reaction volume containing assay buffer (20 mM HEPES pH 8.0, 10 mM NaCl, 5% glycerol, 1 mM $MgCl_2$, 0.5 mM TCEP) and 0.5 nM Cy3-labeled DNA substrate. Each reaction was supplemented with the RecQ4 protein, utilizing a serial dilution from 2 µM to 3.8 pM. FP values for each RecQ4 variant concentration were determined by measuring the parallel and perpendicular fluorescence of the Cy3-labeled DNA (Excitation 530 nm / Emission 580 nm) using the Clariostar microplate reader (BMG LABTECH). Maximum and minimum polarization values indicate fully bound and unbound ssDNA, respectively, and were used to normalize the ssDNA binding curves (if applicable). Normalized values were plotted to a logarithmical scale of the protein concentration and fitted using the logistic fitting algorithm in OriginPro (OriginLab). The protein concentration at half-maximal binding represents the apparent dissociation constant (K_d) for the particular RecQ4 variant. Unless stated otherwise, the presented FP data was obtained from one single experiment.

Table 3-18: Fluorescence polarization DNA substrates

FP Substrate	Top strand	Bottom DNA strand
30nt ssDNA	T3-Cy3	-
15nt ssDNA	T2-Cy3	-
15nt3'OH	T3-Cy3	B1
15nt blunt	T2Cy3	B1

3.3.8. *In vitro* helicase assay

Helicase assays were performed in a Clariostar microplate reader (BMG LABTECH) at 25°C in 50 µL reactions containing 1 µM protein (unless stated otherwise) and 50 nM of a DNA substrate in assay buffer (20 mM HEPES pH 8.0, 10 mM NaCl, 5% Glycerol, 1 mM $MgCl_2$, 0.5 mM TCEP). After recording the baseline fluorescence (60-90 s), the reaction was initiated by adding ATP to a final concentration of 1.25 mM (unless stated otherwise). The helicase reaction, visualized by increasing Cy3-fluorescence as the quencher-labeled bottom-DNA strand is separated from the Cy3-labeled top-DNA strand, was recorded by monitoring the fluorescence intensity (Excitation 530 nm / Emission 580 nm). If indicated, measurements using H_2O in place of ATP as well as reactions with buffer instead of protein served as blank and were subtracted from the helicase-data. Helicase rates were calculated based on the linear increase in fluorescence for the first 10 s after ATP addition. Unless stated otherwise, all fluorescence-based helicase assays represent data obtained from one single experiment.

Table 3-19: Helicase assay DNA substrates

Substrate	Top strand	Bottom DNA strand
15nt3'OH	T3-Cy3	B1-Dab
Y-15-15	T3-Cy3	B4-Dab
Y-18-15	T7-Cy3	B8-Dab
Y-21-15	T8-Cy3	B9-Dab
Y-24-15	T9-Cy3	B10-Dab
30nt3'OH	T12-Cy3	B1-Dab
30nt blunt	T11-Cy3	B12-Dab

3.3.9. *In vitro* ATPase assay

ATPase turnover rates were determined using an NADH-coupled ATP consumption assay, where ATP hydrolysis, performed by the helicase, is coupled to the proportional decrease of NADH concentration through an ATP regeneration system based on phosphoenolpyruvate (PEP), pyruvate kinase (PK) and lactate dehydrogenase (LDH). NADH absorbance at 340 nm was recorded over 30 minutes and converted into ATP concentration using the molar extinction coefficient of NADH. All reactions were carried out at 30°C using the Clariostar microplate reader (BMG LABTECH). Two types of ATPase assays were performed: (1) ATP-concentration dependent ATPase rates were measured in 50 µl reactions containing 0.3 mM NADH, 2.7 U LDH, 1.8 U PK, 2 mM PEP and 1 µM polyT30-ssDNA substrate in assay buffer (20 mM HEPES pH 8.0, 10 mM NaCl, 5% Glycerol, 1 mM $MgCl_2$, 0.5 mM TCEP). The reactions were supplemented with 0.3 µM RecQ4 helicase variant and initiated by addition of ATP with concentrations ranging from 0 mM to 10 mM. The ATP-concentration dependent ATPase data were fitted to the Michaelis-Menten-equation to calculate the maximal ATPase rates (V_{max}). (2) ssDNA-dependent ATPase rates were conducted in the same way as in (1) with the exception that the reactions varied in their polyT30-ssDNA concentration, representing a serial dilution from 1.024 µM to 0.125 nM, and a standard ATP concentration of 1.25 mM was used to initiate the reaction. ssDNA-dependent ATPase data were plotted to a logarithmic ssDNA-concentration scale and fitted to a logistic fitting function in OriginPro (OriginLab). K_{DNA} values represent the ssDNA concentration required for half-maximal activity. Unless stated otherwise, all experiments were performed in triplicates with protein from at least two different purification batches.

3.3.10. Theromfluor assay

In order to confirm intact protein folding, Thermofluor analysis was performed on selected RecQ4 variants, in which a slow and controlled heat increase leads to the thermal denaturation of the tested RecQ4 proteins. This unfolding reaction is visualized by an increase in the fluorescence signal, arising from a dye (SYPRO Orange), which binds to hydrophobic amino acids that become accessible to the dye upon unfolding of the protein, resulting in an increase in fluorescence. Proteins with a similar structural fold/stability therefore show overlapping thermal denaturing curves. For each RecQ4 variant, 5 µL of protein (1.74 mg mL^{-1}) were mixed with 1 µL of a 2.5% SYPRO Orange solution and 19 µL storage buffer (10 mM HEPES pH 8.0, 100 mM NaCl, 10% glycerol, 2.5 mM $MgCl_2$, 1 mM TCEP) and heated gradually to 95°C with a step size of 1°C min^{-1} using a qPCR machine (Stratagene Mx3005P), which permits simultaneous SYPRO Orange fluorescence monitoring (Excitation 492 nm / Emission 610 nm). Data were analyzed using Excel (Microsoft) and OriginPro (OriginLab).

4. Experimental Results

4.1. RecQ4 variants, their expression and purification

4.1.1. *In silico* secondary Structure prediction

The design of a suitable protein construct is crucial for the success of its expression, purification and crystallization. In this respect, *in silico* prediction tools provide insightful guidance by identifying secondary structure elements and potential domain boundaries of a structurally uncharacterized protein on the basis of primary sequence information. For the human RecQ4 helicase, secondary structure prediction was performed via two freely available online services: Phyre2 (Ph2) and I-TASSER (IT) (Fig. 4-1).

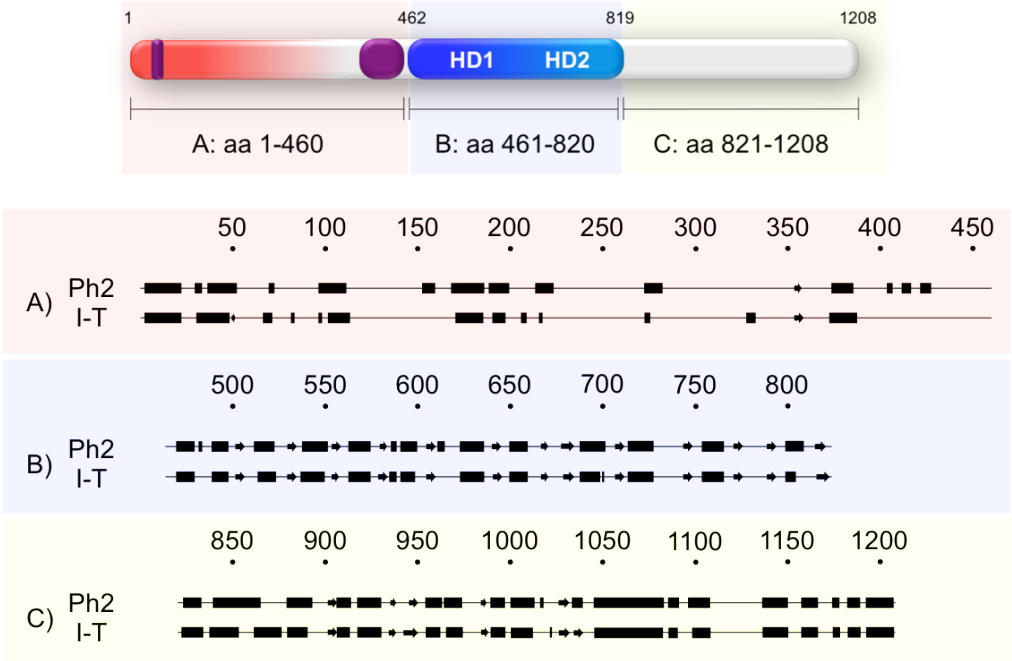


Figure 4-1: Prediction of secondary structure elements within human RecQ4. Panels A, B and C illustrate the predicted secondary structure elements for the RecQ4 sections of aa 1-460, 461-820 and 821-1208, respectively. While the N-terminus of RecQ4 (red box, A) exhibits only sparse structural elements with some deviations between the predictions by the Phyre2 (Ph2) and I-TASSER (I-T) algorithms, the helicase core (blue box, B) and the C-terminus (yellow box, C) are predicted to be densely structured with a high degree of congruence regarding the individual secondary structure elements. A-helices and β-sheets are depicted as boxes and arrows, respectively.

The results of both prediction tools suggested that the N-terminal domain of RecQ4 (Fig. 4-1 lane A) sparsely exhibits secondary structure elements, while the centrally located helicase core (Fig. 4-1 lane B) and the unique C-terminus (Fig. 4-1 lane C) are both highly structured. Moreover, with only minor exceptions, both algorithms predicted a highly congruent sequence of secondary structure elements for both, the helicase core and the C-terminus. In contrast, the N-terminus exhibited a somewhat lower degree of congruence, which is most pronounced within the first 50 aa and in smaller sections around aa 100, aa 170-200 and aa 375. Flexible and unstructured regions within a protein are unfavorable for both, protein expression and crystallization. Accordingly, the construct design for truncated RecQ4 variants that are suitable for protein expression and crystallization focused on the highly structured helicase core domains and the unique C-terminus.

4.1.2. Advancing RecQ4 expression in *E. coli*

This section provides an overview about the development of expression- and purification strategies of various RecQ4 variants, thereby summarizing the rationale behind adjustments of protein purification approaches and the adaptation of RecQ4 expression-construct design. An example of the final purification procedure is given in the following segment and a complete list of RecQ4 protein constructs is provided in the materials and methods section.
A prime objective of this thesis project required establishing the recombinant expression of RecQ4 variants in *E. coli*. To this end, three initial RecQ4 variants were cloned into the pETM-11 bacterial protein expression vector, featuring an N-terminal hexahistidine ($His_{(6x)}$) purification tag and a TEV cleavage site. In addition to the full-length RecQ4 sequence (RecQ4_FL), two truncated RecQ4 variants were designed based on the above described secondary structure prediction analysis: $RecQ4^{460\text{-}895}$, starting right before the helicase core and including a small fraction of the RecQ4 C-terminal domain, and $RecQ4^{460\text{-}1208}$, featuring the helicase core and the entire C-terminus (Fig. 4-2).

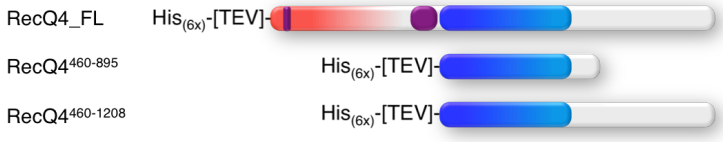


Figure 4-2: Overview of the initial pETM-11 RecQ4 expression variants. Full-length RecQ4 (RecQ4_FL) and the two RecQ4 truncation variants, RecQ4$^{460-895}$ and RecQ4$^{460-1208}$, feature an N-terminal hexahistidine tag (His_{6x}) and a Tobacco Etch Virus (TEV) protease cleavage sequence.

Conventional systematic expression tests, e.g. via detecting the overexpression of the desired protein by SDS-PAGE analysis of pellet- and supernatant samples, were not applicable due to RecQ4s poor soluble expression behavior. Thus, medium-scale 200 mL expression tests with target-protein enrichment via Ni^{2+}-NTA sepharose in batch mode were performed. Initial testing of the two RecQ4 truncation variants encompassed two different *E. coli* expression strains, BL21 star (DE3) and Rosetta (DE3), grown in 2xYT LB medium in combination with low (0.1 mM) and high (1.0 mM) Isopropyl-β-D-thiogalactoside (IPTG) induction and expression at 18°C and 37°C. The results indicated soluble production of both RecQ4 truncations at 18°C in both *E. coli* strains, with moderately higher yields at 1.0 mM IPTG induction. Similar expression tests for RecQ4_FL were unsuccessful. A follow-up large-scale (2 L) expression, however, revealed a small signal in the SDS-PAGE analysis when expressed for eight hours at 37°C and no signal at 18°C. Unfortunately, subsequent concentration of the RecQ4_FL IMAC elution fractions for analysis by size exclusion chromatography (SEC) led to protein precipitation, an occurrence that could not be averted by expressing RecQ4_FL with N-terminal solubility tags or by using optimized buffer conditions. Similarly, precipitation during protein concentration was encountered for the shorter RecQ4$^{460-895}$ truncation variant in the following large-scale expressions. The C-terminally extended RecQ4$^{460-1208}$ variant, however, endured the concentration procedure to a small extent, yet the remaining protein eluted in the void volume during SEC, indicating protein aggregation, with predominantly co-purified *E. coli* protein

contaminations in the soluble phase. In order to improve yield and purity, further optimization trials included RecQ4$^{460-1208}$ expression in Arctic Express (DE3) RP and RIL bacterial strains, the analysis of the binding behavior to different IMAC media (e.g. Ni^{2+}-TED, Ni^{2+}-IDA or Ni^{2+}-NTA) and the testing of C-terminal $His_{(6x)}$-tagged RecQ4 versions, however, with no improvements compared to previous purifications. Thus, all three RecQ4 variants were sub-cloned into the pETM-14 expression vector, featuring an N-terminal $His_{(6x)}$ coupled to an smt3/SUMO solubility tag and a PreScission protease (3C) cleavage site. The SUMO-RecQ4$^{460-1208}$ variant showed significantly increased yields after the IMAC purification step, whereas no improvement was achieved for RecQ4$^{460-895}$ or RecQ4_FL. Characteristic for the IMAC purification of SUMO-RecQ4$^{460-1208}$ was an elution at high imidazole concentrations (300-500 mM), suggesting that the protein might either interact with the sepharose medium or form aggregates/oligomers that would increase the number of $His_{(6x)}$-tags per aggregate/oligomer and would therefore require higher imidazole concentrations to be eluted from the column. In agreement with the aggregation hypothesis, the protein continued to form high-molecular weight aggregates as shown by analytical SEC. To address the aggregation problem, C-terminally shortened RecQ4$^{460-1208}$ variants were created in order to exclude the possibility that a partially unfolded C-terminus could be responsible for the aggregation behavior: RecQ4$^{460-1115}$ and RecQ4$^{460-1047}$. Both variants were expressed in *E. coli* and purified in batch mode via IMAC. To test whether high imidazole concentrations could have an influence on protein stability, elution was performed with a three-step imidazole gradient (100 mM, 250 mM and 500 mM) and all fractions were analyzed via analytical SEC (Fig. 4-3). Indeed, the RecQ4$^{460-1047}$ variant showed an improved 280/254 nm absorption ratio for the 100 mM imidazole elution, which gradually declined with increasing imidazole concentration, however, RecQ4$^{460-1047}$ still seemed to form aggregates as shown by the SEC elution profile. RecQ4$^{460-1115}$, on the other hand, was not affected by different imidazole concentrations and similarly eluted in the aggregate peak on the SEC column.

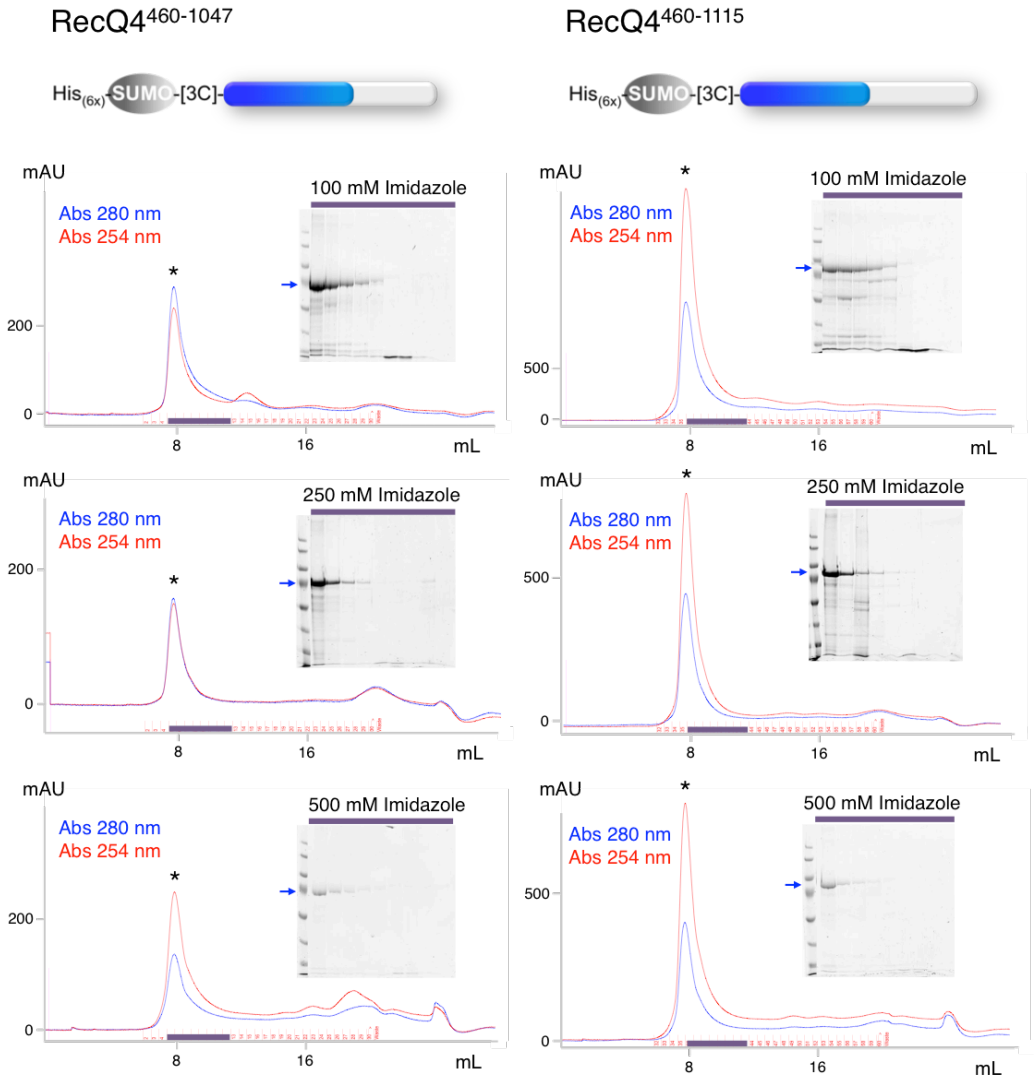


Figure 4-3: Analysis of the aggregation problem: Influence of the RecQ4 C-terminus and the imidazole concentration. C-terminal truncation variants of RecQ4$^{460-1208}$, RecQ4$^{460-1047}$ (left) and RecQ4$^{460-1115}$ (right), were purified via IMAC by elution with increasing imidazole concentration steps, which were subsequently concentrated and analyzed via analytical SEC. The SEC chromatograms for the 100 mM (top), 250 mM (middle) and 500 mM (bottom) imidazole steps demonstrate that both RecQ4 truncation variants elute in the aggregate peak (*) around 8 mL, suggesting that both variants form soluble aggregates irrespective of their truncation and the imidazole concentration which they were eluted with during the IMAC purification. While the absorption ratio between 280 nm (blue lines) and 254 nm (red lines) indicate partially unfolded protein and/or DNA contaminations for all the RecQ4$^{460-1115}$ chromatograms, this is not the case for the 100 mM and 250 mM imidazole elution SEC chromatograms of RecQ4$^{460-1047}$, suggesting a more stable protein fold or fewer DNA contaminations for this RecQ4 variant. Inserts within each chromatogram illustrate SDS-PAGE analyses of the corresponding SEC fractions (purple bar). A blue arrow marks the signal for the RecQ4 variants in the SDS-PAGE gels.

It was thus concluded that an unfolded C-terminus might not be the main reason for protein aggregation. Accordingly, further optimization trials to improve protein stability included adapted protein expression strategies, e.g. by the use of alternative expression strains and auto-induction medium (ZYP5052), as well as variations of the IMAC method by using cobalt- and zinc-complexd IMAC resins in combination with different non-imidazole elution procedures, however, with no improvement in terms of aggregation behavior. Eventually, further adaptation within the RecQ4 expression constructs were made, leading to four new variants, each with variable N- and C-terminal extensions: RecQ4$^{450\text{-}997}$, RecQ4$^{450\text{-}1116}$, RecQ4$^{427\text{-}997}$ and RecQ4$^{427\text{-}1116}$ (Fig. 4-4 A). A comparative purification of RecQ4$^{427\text{-}997}$ and RecQ4$^{450\text{-}997}$ still exhibited late imidazole elution, yet revealed for the first time an indication of monomeric protein on the analytical size exclusion column for both RecQ4 variants. However, much of the protein remained in the aggregate fraction and the 280/254 nm absorption ratio of the monomer peak still indicated a partially unfolded state. In addition to unfolded protein, the unfavorable 280/254 nm absorption ratio could also be indicative for DNA contaminations, which however was considered unlikely due to the standard use of DNaseI prior to protein purification and the non-responsive implementation of high-salt wash steps during previous IMAC purifications. Nonetheless, the cleared supernatant of a RecQ4$^{427\text{-}1116}$ expression batch was directly applied to a Heparin HP sepharose column and subsequent SDS-PAGE analysis of the elution fractions indicated the presence of RecQ4$^{427\text{-}1116}$ among many co-purified *E. coli* contaminants. The following IMAC purification exhibited an elution of RecQ4$^{427\text{-}1116}$ at significantly lower imidazole concentrations and, accordingly, the SEC chromatogram showed monomeric protein with an ideal 280/254 nm absorption ratio of approximately 2:1, indicating correctly folded and DNA-free protein (Fig. 4-4 B). A similar purification procedure for the RecQ4$^{450\text{-}997}$ variant showed successful binding to the heparin column, yet again high-imidazole elution and primarily monomeric, albeit partially unfolded protein on the size exclusion column. In summary, it was concluded that the C-terminal part, at

least until aa 1116, is indeed essential for proper protein folding and the heparin purification step is in addition required to detach RecQ4-bound DNA fragments, which must have been resistant to DNaseI treatment. As the RecQ4$^{427\text{-}1116}$ fractions still contained major *E. coli* contaminants after size exclusion chromatography, a subsequent cation exchange chromatography (CEX) purification step via a MonoS column was established, which significantly increased the purity of the protein.

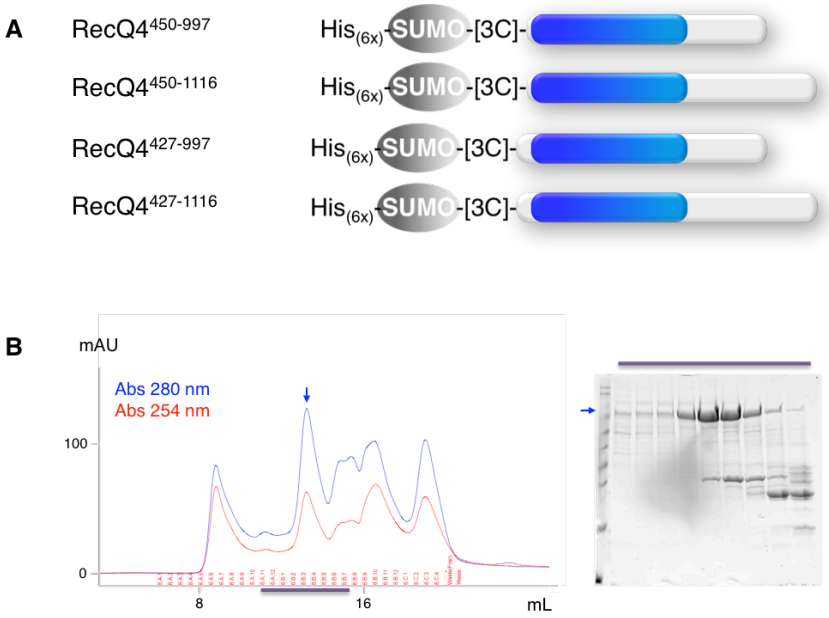


Figure 4-4: Solving the aggregation problem. (A) Revised pETM-14 SUMO-RecQ4 expression constructs, featuring new N- and C-terminal extensions. **(B)** Left: Analytical SEC chromatogram of SUMO-RecQ4$^{427\text{-}1116}$ after Heparin-IMAC purification. The RecQ4 variant elutes at 13.3 mL, representing monomeric protein. The absorption ratio between 280 nm (blue line) and 254 nm (red line) of approximately 2:1 indicates properly folded protein. SEC fractions (purple bar) were analyzed by SDS-PAGE. Right: SDS-PAGE gel of selected SEC fractions. The blue arrow indicates the signal for SUMO-RecQ4$^{427\text{-}1116}$ (89 kDa).

Follow-up purifications revealed that a fully C-terminal RecQ4$^{427\text{-}1208}$ variant could be equally well purified via this four-step purification procedure. In order to increase protein yields, it was further addressed whether the Heparin- and the IMAC purification steps are interchangeable, however, reversing the order of the heparin- and IMAC purification led to a change in

the contamination profile, which could not be rescued by the MonoS CEX and the final purification scheme was therefore kept in the previously established order of Heparin-IMAC-SEC-CEX. Over the course of many large-scale purifications, minor adjustments were made which all led to further advantages in expression and purification performance. Foremost, all protein expressions were performed in Terrific Broth (TB) medium, which primarily increases *E. coli* bacterial mass due to high-OD induction of recombinant protein expression. In contrast to the initial expression tests, low IPTG induction (0.1 mM) proved to be more beneficial than induction through a high IPTG concentration. Finally, all RecQ4 constructs were sub-cloned into the pETM-22 expression vector, which featured an N-terminal Thioredoxin (Trx) solubility tag, followed by $His_{(6x)}$ and a 3C cleavage site. The Trx-RecQ4 variants had a minor purity advantage compared to their SUMO-counterparts while maintaining a comparable yield. Altogether, the complete expression and purification process, although providing reliable performance, remains time- and material intensive. The low binding affinity and capacity of the heparin medium represents the major purification bottleneck, resulting in a final protein yield of approximately 1 mg per 50 g *E. coli* cell pellet. However, through the use of an Äkta Pure chromatography system, the Heparin- and IMAC steps can be systematically combined and the purification of up to three batches can be automatically scheduled one after another, which increases purification speed and permits the parallel purification of up to three RecQ4 protein variants.

4.1.3. Final RecQ4 purification scheme

The following purification protocol was developed for all RecQ4 variants that feature the conserved helicase core and the C-terminal domain, including $RecQ4^{427\text{-}1116}$ and $RecQ4^{427\text{-}1208}$ as well all corresponding mutational- and deletion variants that were used in this study. The protocol comprises five chromatography steps, which were optimized over the course of multiple large-scale purifications: Heparin affinity chromatography, coupled to an

IMAC purification step, followed by size exclusion chromatography (SEC), cation exchange chromatography (CEX) and a final analytical SEC step for protein polishing and the transfer of all RecQ4 proteins into a standardized storage buffer.

For each purification, 40-50 g of *E. coli* cell pellet was used, resulting in 200-250 mL supernatant after cell lysis and centrifugation. Further increase of the batch size was not applicable as the binding capacity of the heparin column was already at its maximum. The cleared supernatant was applied to a manually packed 100 mL Heparin FF sepharose 26/40 column, followed by a wash step with a low-salt buffer and protein elution, which was performed by a one-step high-salt procedure (Fig. 4-5 A). The Heparin-affinity step primarily serves to detach bound DNA fragments from the RecQ4 proteins and is not suitable to selectively separate the target protein from other heparin-bound contaminants via an elution gradient. The heparin column eluent was collected and directly applied to a 5 mL HisTrap FF crude IMAC column, from which RecQ4 was eluted with an imidazole gradient (Fig. 4-5 B). The entire IMAC gradient eluent was collected irrespective of additional contaminants, as these were removed during subsequent purification steps. In case of a parallel three-batch purification of the same protein variant, it was worthy to collect the heparin column flow-through of all three batches and apply them combined to an IMAC column. Due to the low binding affinity of the RecQ4 variants to the heparin matrix, a minor fraction of the RecQ4 protein was routinely retained in the Heparin flow-through, which in case of three or more consecutive purifications added up to be a sufficient amount of protein for a successive purification. As the transient interaction with Heparin sepharose has already detached the RecQ4-bound DNA, it was sufficient to start the so-called flow-through RecQ4 (FT_RecQ4) purification by loading the combined Heparin flow-through directly on a HisTrap column. The FT_RecQ4 was initially purified separately and its activity was analyzed *in vitro*, displaying no difference to the standard heparin-IMAC purified RecQ4 protein. Therefore, the FT_RecQ4 was routinely combined with the main RecQ4 pool in later purifications, prior to preparative SEC.

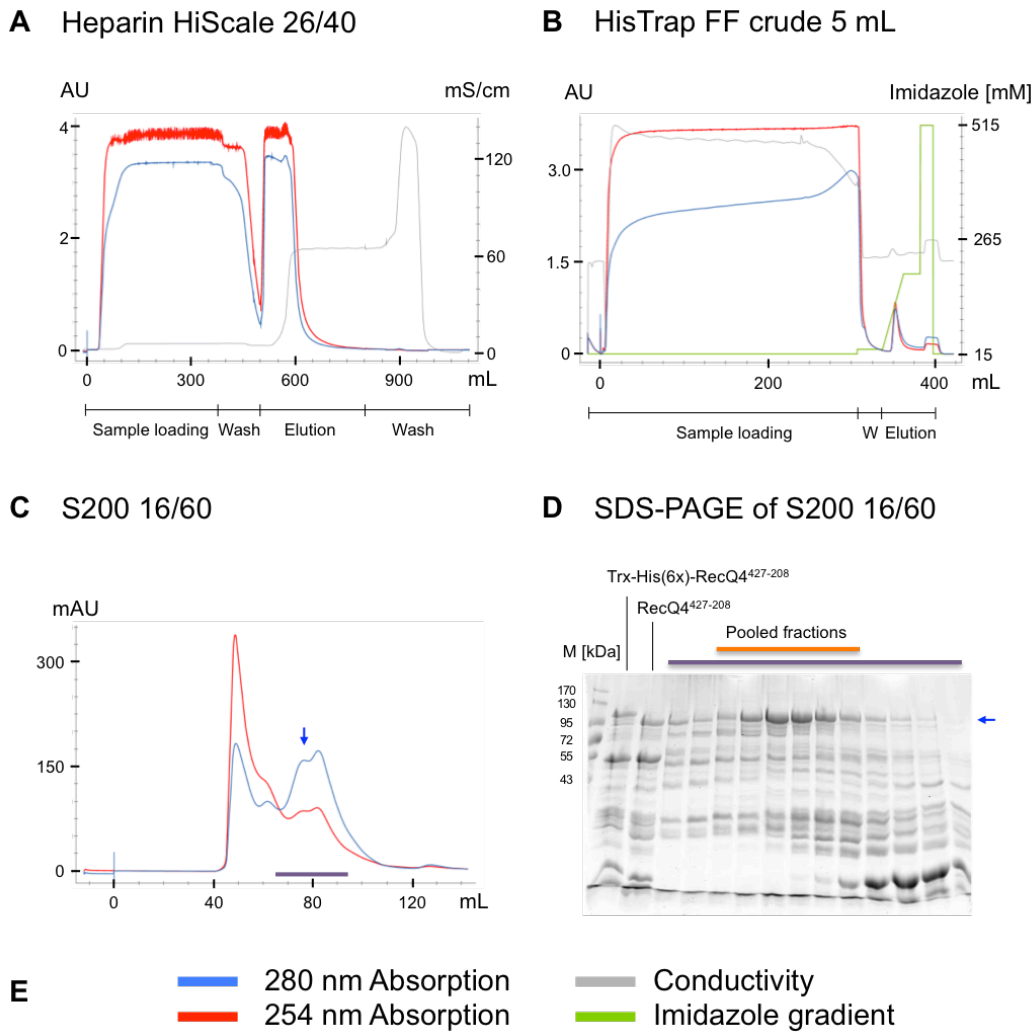


Figure 4-5: Standard purification of RecQ4$^{427-1208}$ – Heparin-IMAC-SEC. (A) Heparin 26/40 chromatogram, illustrating sample loading, the wash step and elution of bound proteins via a 1 M NaCl buffer (approx. 68 mS/cm). **(B)** IMAC chromatogram using a HisTrap FF crude 5 mL column. The entire heparin eluent was loaded, washed with 25 mM Imidazole (W) and eluted with a gradient up to 190 mM imidazole. Bound proteins elute at around 100 mM imidazole. **(C)** Superdex 200 16/60 chromatogram of the purification-tag cleaved RecQ4$^{427-1208}$ variant. RecQ4 elutes at 76 mL as indicated by the blue arrow. SEC fractions (purple bar) were analyzed via SDS-PAGE. **(D)** SDS-PAGE gel of the SEC fractions, including a sample before and after purification-tag cleavage. The blue arrow indicates the signal for RecQ4$^{427-1208}$ and the orange bar highlights the SEC fractions, which were collected for further purification steps. **(E)** Chromatogram legend.

The RecQ4 containing IMAC fractions were combined, incubated with 3C protease to cleave off the Trx-His$_{(6x)}$ purification tag and subsequently concentrated for preparative SEC via a Superdex 200 16/60 gel filtration

column (Fig. 4-5 C). Samples taken prior and after 3C protease incubation routinely served to confirm the presence of the target protein as the respective signals in the SDS-PAGE exhibit a characteristic shift to lower molecular weight for the RecQ4 protein without the Trx-$His_{(6x)}$ tag (Fig. 4-5 D). Concentrating the protein for preparative SEC was often prone to aggregation when using V-shaped centrifugation devices. To avoid very high concentrations at the tip of the V-filter-unit, centrifugal devices from PALL were used instead, which exhibit a different filter design, thereby preventing regional over-concentration and precipitation. Fractions of the SEC step were analyzed by SDS-PAGE (Fig. 4-5 D), appropriately combined and adjusted to a salt concentration of approximately 35 mM NaCl. The reduction of the salt concentration was achieved by drop-wise addition of Zero-NaCl buffer while the solution was continuously stirred in a cooled water container, thereby minimizing the amount of protein precipitation during the dilution process. The low-salt protein solution was then applied to a 1 mL MonoS 5/50 GL CEX column from which the bound RecQ4 protein was eluted with a shallow salt gradient (Fig 4-6 A). Depending on the RecQ4 variant, in particular on the length of the C-terminus, RecQ4 typically eluted between 90-130 mM NaCl, usually sufficiently separated from contaminants, which eluted at slightly lower and higher salt concentrations, respectively. Samples from the CEX purification step were analyzed by SDS-PAGE (Fig. 4-6 B), RecQ4-containing fractions were combined and subsequently concentrated for a final analytical SEC in order to remove remaining minor contaminations and adjust the purified RecQ4 variant to the common storage buffer (Fig. 4-6 C). Appropriate fractions from the analytical SEC were combined and concentrated after SDS-PAGE analysis (Fig. 4-6 D). Batches assigned for crystallization were either used directly or flash frozen in liquid nitrogen and stored at -80°C. Protein batches assigned for *in vitro* assays were diluted with storage buffer to a final concentration of 20 µM and aliquoted into 20 µL samples before freezing.

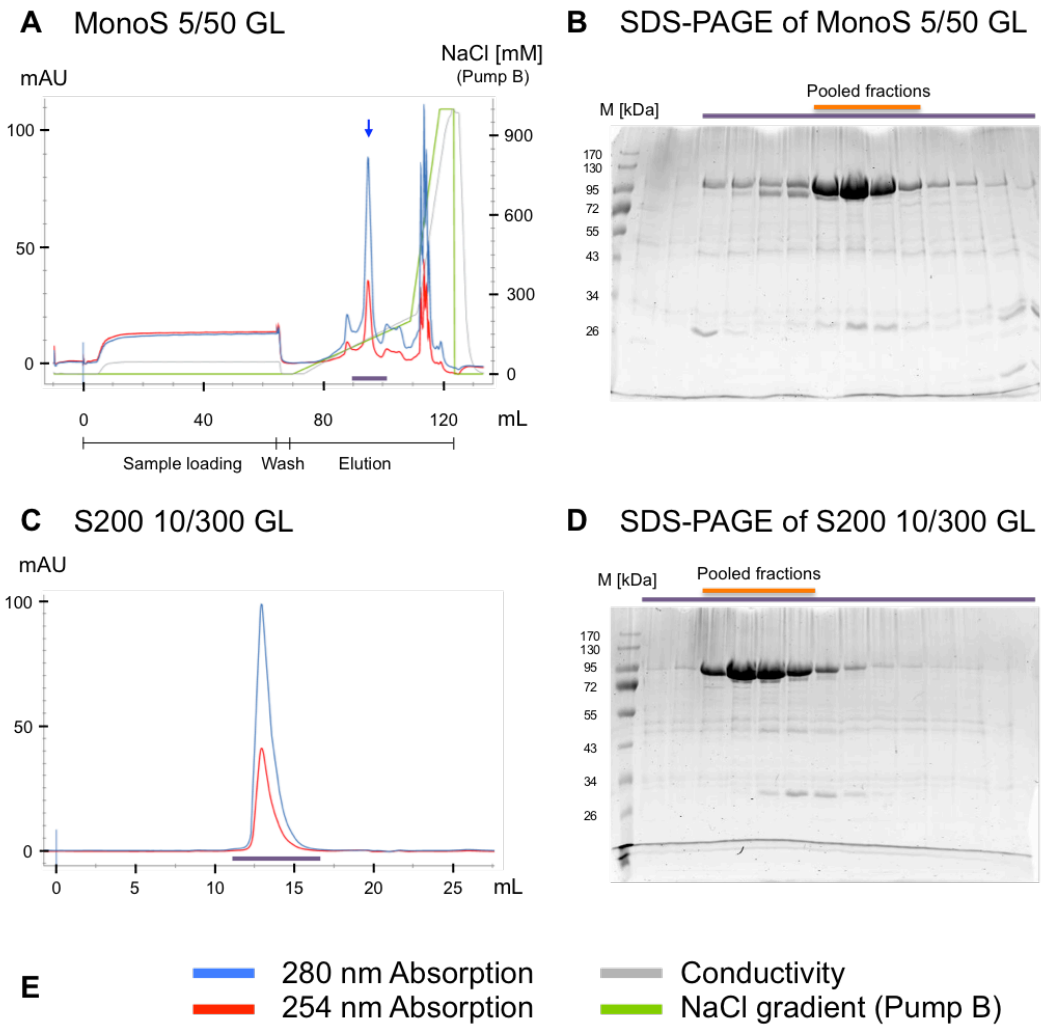


Figure 4-6: Standard purification of RecQ4$^{427\text{-}1208}$ – MonoS and analytical SEC. **(A)** MonoS 5/50 GL chromatogram, illustrating sample loading, wash and elution of RecQ4$^{427\text{-}1208}$. The RecQ4 variant elutes in a sharp peak, indicated by a blue arrow, at approximately 130 mM NaCl. **(B)** SDS-PAGE analysis of selected MonoS fractions (purple bar). RecQ4$^{427\text{-}1208}$ exhibits a molecular weight of 87 kDa. The orange bar highlights selected fractions for further processing. **(C)** Analytical SEC chromatogram of RecQ4$^{427\text{-}1208}$ using a Superdex 200 Increase 10/300 GL column. The protein elutes at 12.9 mL. Selected fractions (purple bar) were analyzed via SDS-PAGE. **(D)** SDS-PAGE gel of the analytical SEC fractions, representing the purified RecQ4$^{427\text{-}1208}$ variant (87 kDa). The orange bar indicates the SEC fractions, which were collected for subsequent assays or crystallization trials. **(E)** Chromatogram legend.

4.2. Structural and Functional characterization of RecQ4

4.2.1. Crystallization of RecQ4 variants

4.2.1.1. RecQ4$^{427-1116}$

Initial screening of RecQ4$^{427-1116}$ at room temperature revealed a suitable crystallization solution, comprising 0.1 M imidazole pH 8.0 and 3 M sodium formate (Fig. 4-7 A). Additional crystallization conditions, however, could not be identified. Systematic manual screening of pH and precipitant concentration in the 24-well hanging drop setup was pursued to further optimize the crystallization process, resulting in the formation of larger and qualitatively improved protein crystals (Fig. 4-7 B-D). In general, the diffraction quality of the optimized protein crystals appeared to be rather heterogenic, as for example two suitable protein crystals from the same crystallization experiment (e.g. the same crystallization drop) could exhibit a significant difference in diffraction quality. Thus, whenever possible, several protein crystals from one particular crystallization drop were harvested and analyzed. Moreover, the ideal crystallization condition appeared to vary between protein batches, most likely due to minor differences in protein batch quality and/or small variations in crystallization solution preparation. Therefore, crystallization of the RecQ4$^{427-1116}$ variant was always carried out via a screening setup, covering a sodium formate concentration between 2.6 M and 3.1 M, an imidazole pH range between 7.0 and 9.0 as well as two different protein concentrations and protein-to-precipitant ratios. The best-diffracting protein crystals exhibited a resolution maximum of about 3.0 Å when tested at the X-ray home source and were harvested from crystallization conditions comprising 2.7 M or 2.8 M sodium formate and 0.1 M imidazole pH 9.0, using a protein concentration of 3 mg/mL and a protein to precipitant ratio of 2:1. However, well-diffracting protein crystals were also grown at pH 7.0 and pH 8.0, or at a protein concentration of 2 mg/mL and at a 1:1 protein to precipitant ratio. Protein crystals usually appeared after 12 – 24 hours, often nucleating on tiny air bubbles or at the inner surface of the

crystallization drop, which typically developed a translucent and elastic skin on its surface (Fig. 4-7 B, E, F). The RecQ4$^{427\text{-}1116}$ crystals could be gently detached from the skin-like layer by softly poking at the crystal base. Protein crystals were typically harvested within three to four days, supplemented with 25% glycerol (v/v) as cryogenic protectant and flash frozen in liquid nitrogen.

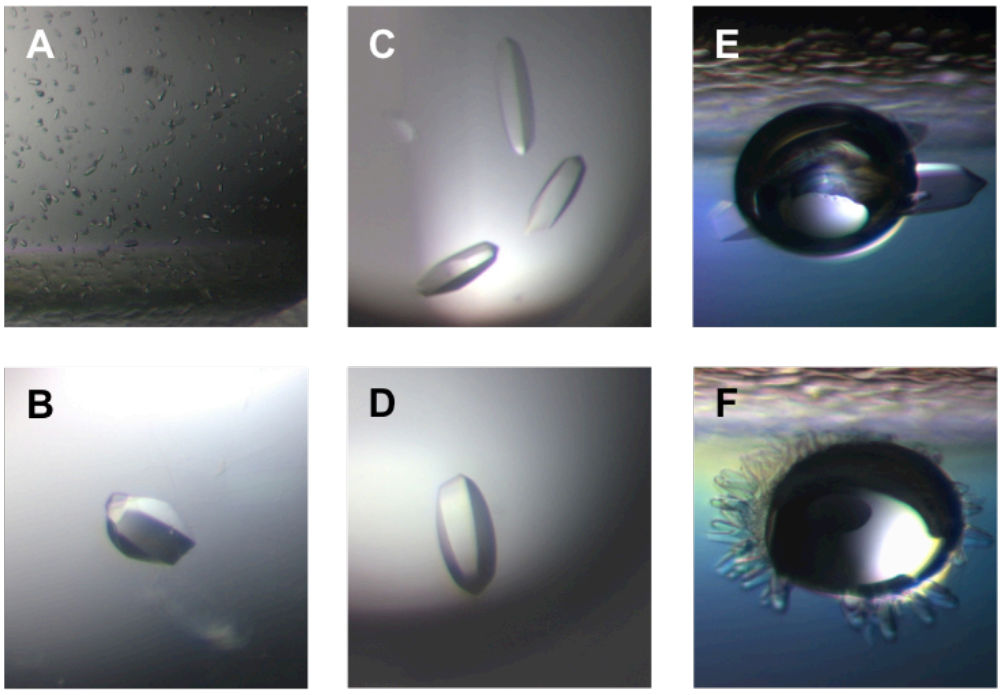


Figure 4-7: Protein crystals of RecQ4$^{427\text{-}1116}$. (A) Crystals from the initial screening condition (0.1 M imidazole pH 8.0, 3 M sodium formate). **(B)** Optimized fine screen crystal (0.1 M imidazole pH 9.0, 3 M sodium formate, 3 mg/mL RecQ4 protein, protein to precipitant ratio 2:1). The crystal base is attached to the skin-like layer that formed at the drop surface. **(C)** Optimized protein crystals grown in 0.1 M imidazole pH 8.0, 2.6 M sodium formate using 2 mg/mL protein and a protein to precipitant ratio of 1:1. **(D)** RecQ4$^{427\text{-}1116}$ protein crystal grown in 0.1 M imidazole pH 9.0, 2.8 M sodium formate using a protein concentration of 3 mg/mL and a protein to precipitant ratio of 2:1. **(E)** Protein crystals, which nucleated at an air bubble in the additive screen (D08, 0.01 M urea). **(F)** Ring of RecQ4$^{427\text{-}1116}$ protein crystals around an air bubble, grown in the additive screen (F04, 0.2 M NDSP-201).

4.2.1.2. Other RecQ4 variants and DNA co-crystallization trials

Several attempts have been undertaken to improve the diffraction quality of the native RecQ4$^{427\text{-}1116}$ protein crystals. Initial optimization trials included the replacement of either the original precipitant with potassium formate or the substitution of the original buffer component with Tris(hydroxymethyl)-aminomethane (Tris) or 4-(2-hydroxymethyl)-1-piperazineethanesulfonic acid (HEPES). Although RecQ4$^{427\text{-}1116}$ successfully crystallized in the adapted conditions, protein crystals generally appeared to be not as large and well-diffracting as the crystals from the original condition. Additional attempts aimed at incorporating either ADP or the non-hydrolysable ATP analogue AMP-PNP into the helicase core of RecQ4$^{427\text{-}1116}$ via co-crystallization or soaking approaches, as this might stabilize the two RecA-like domains and thereby facilitate a well-ordered packing of the protein crystal. However, both nucleotide additives did not increase the diffraction quality of the RecQ4$^{427\text{-}1116}$ protein crystals. Furthermore, a commercially available 96-well additive screen (Hampton Research) was applied to search for reagents that could enhance crystal quality. This additive screen was performed via a pipetting robot that supplemented 96 wells, each featuring the same original crystallization condition, with 96 different crystallization additives. Although many conditions produced protein crystals, only a few grew large enough to be harvested and tested due to the typical fluctuations regarding the ideal crystallization parameters of RecQ4$^{427\text{-}1116}$ (Conditions E06, E12, F02, H10). However, all tested protein crystals that grew with additive supplementation exhibited either a comparable or inferior diffraction quality compared to the originally improved crystals. Further adaptations to increase diffraction quality constituted crystal growth at a lower temperature. The standardized manual 24-well crystallization setup was performed, stored and evaluated at 4°C, which in conclusion led to a decrease in the size of the final protein crystals and slowed down the crystallization process, yet did not improve the diffraction quality. The final attempt to optimize protein crystallization for the RecQ4$^{427\text{-}1116}$ variant included the search for an alternative crystallization

condition. Therefore, initial screening of the apo-protein was repeated using an increased protein concentration of 6 mg/mL and 9 mg/mL, however, without success. This process was combined with the search for suitable crystallization conditions to produce DNA-bound RecQ4$^{427\text{-}1116}$ protein complex crystals, for which the RecQ4$^{427\text{-}1116}$ protein was incubated with either ssDNA (T3-Cy3) or a Y-fork DNA substrate (T3-Cy3:B4) in a molar protein:DNA ratio of 1:1.1 and screened at a protein concentration of 3 mg/mL. Again, no suitable crystallizations condition could be identified in these setups.

In addition to the RecQ4$^{427\text{-}1116}$ variant, the C-terminally extended RecQ4$^{427\text{-}1208}$ has been utilized to screen for suitable crystallization conditions. This search was performed for the apo-protein at a concentration of 3 mg/mL and 6 mg/mL at room temperature and 4°C, but so far no protein crystals could be observed. Attempts to crystallize the RecQ4$^{427\text{-}1208}$ variant in complex with a DNA substrate included supplementation with a ssDNA fragment (T3-Cy3) and the 15nt-polyT 3'OH DNA substrate (T3-Cy3:B1), which were incubated with RecQ4$^{427\text{-}1208}$ in a molar protein:DNA ratio of 1:1.1 and screened at a protein concentration of 3 mg/mL, again, without success.

4.2.2. Structure and function of the human RecQ4 helicase

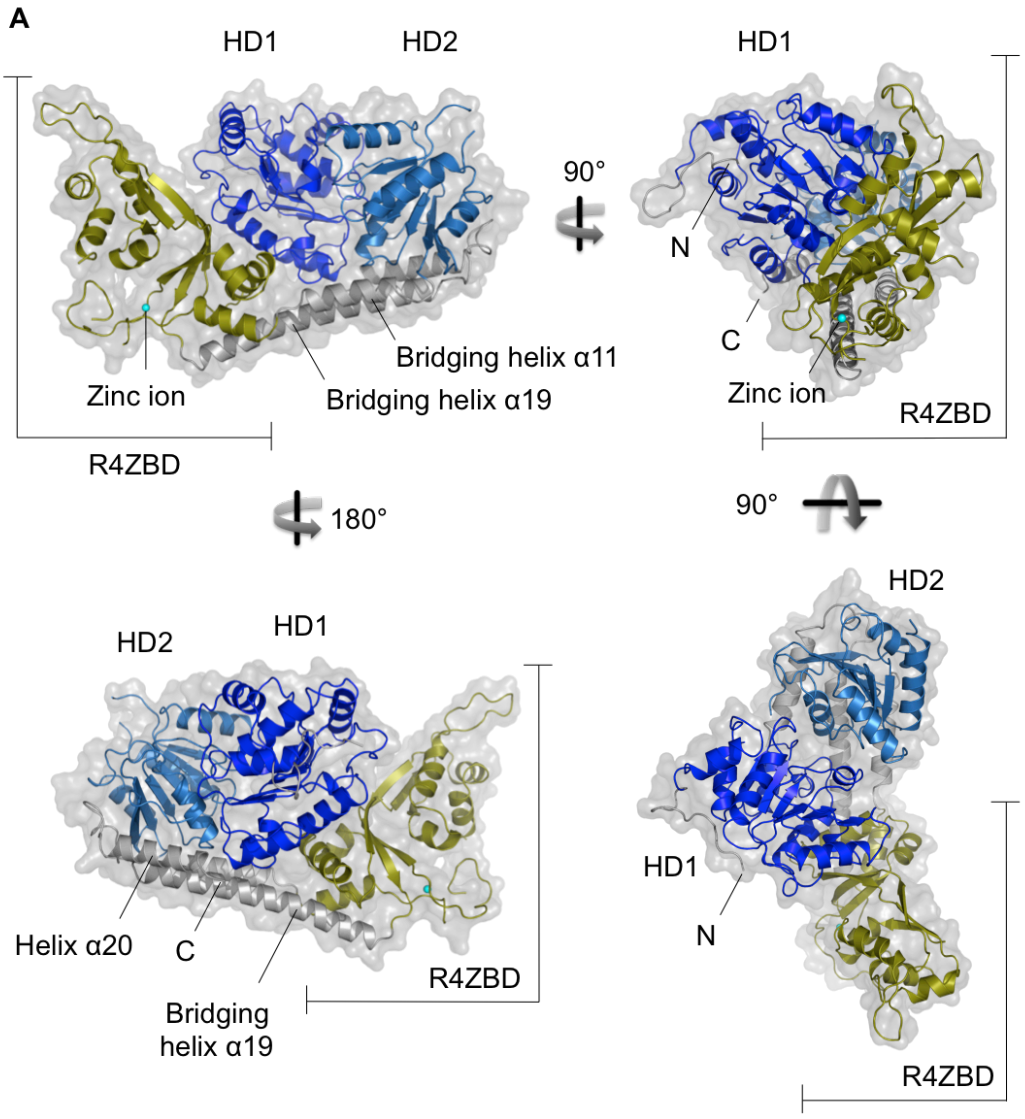
Central elements of the results outlined below have been originally presented and discussed in the publication by Kaiser *et al.* (2017). In parts, textual descriptions and figures have been adapted from the published manuscript.

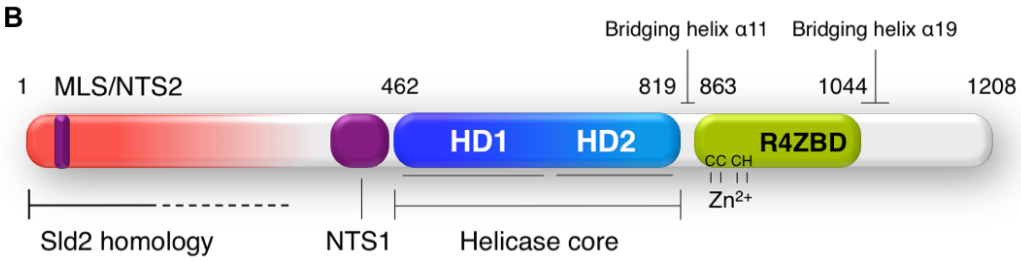
4.2.2.1. Overall structure of RecQ4$^{427\text{-}1116}$

The initial diffraction data that were collected on the first RecQ4$^{427\text{-}1116}$ protein crystals extended to a maximum resolution of about 4 Å. Although it was not possible to solve the structure by molecular replacement, utilizing the conserved helicase domains from other RecQ helicase as a template, the diffraction data indicated the presence of an anomalous scatterer. A subsequent fluorescence scan confirmed the presence of zinc in the RecQ4$^{427\text{-}1116}$ crystal. Manual optimization of the crystallization condition eventually resulted in improved protein crystals, diffracting to a maximum resolution of 2.75 Å. In addition to a high-resolution dataset at a wavelength of 0.9763 Å, diffraction data at the zinc peak wavelength of 1.2828 Å was recorded on individual protein crystals (Table 4-1). Using the CRANK2[236] pipeline, the structure was subsequently solved in space group $P3_121$ via the single wavelength anomalous dispersion (SAD) method with the diffraction data collected at the zinc peak energy. After the location of a single zinc ion with AFRO/CRUNCH[237,238], electron density modification and automated model building by PARROT[240] and BUCCANEER[241], respectively, the resulting fragmented RecQ4$^{427\text{-}1116}$ model was manually completed with COOT[242]. After alternating cycles of manual model building and automated refinement of structure coordinates, B-factors and TLS parameters using PHENIX[243], the final RecQ4$^{427\text{-}1116}$ model was refined to an R_{work} of 18.3% and an R_{free} of 23.1%. The model contains the RecQ4 aa 449-1111. However, two short disordered regions within the helicase core (aa 664-667 and 727-734) as well as two segments within the C-terminus (aa 858-882 and 975-983) are not resolved. The RecQ4$^{427\text{-}1116}$ structure (Fig. 4-8 A)

comprises the evolutionary conserved helicase core, encompassing HD1 (aa 466-663) and HD2 (aa 668-819), each adopting the typical RecA-like fold. The characteristic structural hallmark of a RecA-fold consists of a central parallel β-sheet that is encircled by several α-helices. Thus, the central β-sheet of the RecQ4 HD1 comprises six parallel β-strands, surrounded by six α-helices, while the smaller HD2 harbors six parallel β-strands that are surrounded by four α-helices. Further downstream, the RecQ4 model features a novel domain (aa 836-1044), which has so far not been observed in any other RecQ helicase (Fig. 4-8 A and B). One feature of this new domain is the coordination of a zinc-ion, for which it is subsequently designated as RecQ4-zinc-binding-domain (R4ZBD). The helicase core and the R4ZBD are structurally connected via two bridging helices, the first of which is sequentially located between HD2 and R4ZBD (helix α11, aa 823-835), while the second larger bridging helix is located downstream of the R4ZBD (helix α19, aa 1046-1079), leading back towards HD2. The last helix of the RecQ4$^{427\text{-}1116}$ model, helix α20 (aa 1095-1109), is folded back onto the distal half of the large bridging helix α19 and both secondary structure elements are tightly packed against HD2. The R4ZBD, combined with the bridging helices, resembles a lever-like structure with the HD2-bridging-helix interface representing the joint of the lever.

Figure 4-8: Overall structure of RecQ4$^{427\text{-}1116}$. (A) Structure of the RecQ4$^{427\text{-}1116}$ protein variant in cartoon representation shown from different perspectives. Front view (upper left), side view (upper right), top view (lower right) and back view (lower left). The model illustrates the structural arrangement of the conserved helicase core, consisting of HD1 (deep blue) and HD2 (light blue), and the newly identified RecQ4-zinc-binding domain (R4ZBD, depicted in olive), which are connected by the two bridging helices α11 and α19 (depicted in gray). The R4ZBD coordinates a zinc ion (cyan sphere). The N- and C-terminus of the RecQ4$^{427\text{-}1116}$ variant are indicated by N and C, respectively. **(B)** Updated illustration of the RecQ4 domain architecture, depicting the location and domain boundaries of the R4ZBD, including the three cysteines (C) and the histidine residue (H), which coordinate the zinc ion, as well as the location of the two bridging helices α11 and α19. This Figure has been partially adapted from Kaiser *et al.* (2017).


A
HD1
HD2
HD1
90°
N
C
Bridging helix α11
Zinc ion
Bridging helix α19
Zinc ion
R4ZBD
R4ZBD
180°
90°
HD2
HD1
HD2
HD1
Helix α20
C
R4ZBD
N
Bridging
helix α19
R4ZBD


B
Bridging helix α11
Bridging helix α19
1
MLS/NTS2
462
819
863
1044
1208
HD1
HD2
R4ZBD
CC CH
Zn2+
Sld2 homology
NTS1
Helicase core

Topologically, the R4ZBD can be separated into two halves (Fig. 4-9 A). The lower half features a three-stranded antiparallel β-sheet on top of a three-helix bundle and coordinates the zinc-ion via three cysteine residues and a histidine, which are all located within an extended loop region between helix α12 and β-strand β13. The upper half of the R4ZBD also comprises a three-stranded antiparallel β-sheet, adjacent to a four-helix bundle. Thus, the lower and upper halves of the R4ZBD both seem to adopt individual folds, while both are connected by two short linker segments between two β-strands of the upper and lower β-sheets.

The identity of the coordinated zinc-ion within the R4ZBD lower half was confirmed by recording diffraction data on a RecQ4$^{427\text{-}1116}$ protein crystal at wavelengths representing the zinc absorption edge peak-energy and an energy level that is slightly below the zinc absorption edge. The resulting anomalous signal from the zinc ion is maximal at the peak-energy and, consequently, reduced at the energy level that is below the absorption edge, thus confirming that the coordinated ion is indeed a zinc ion (Fig. 4-9 B).

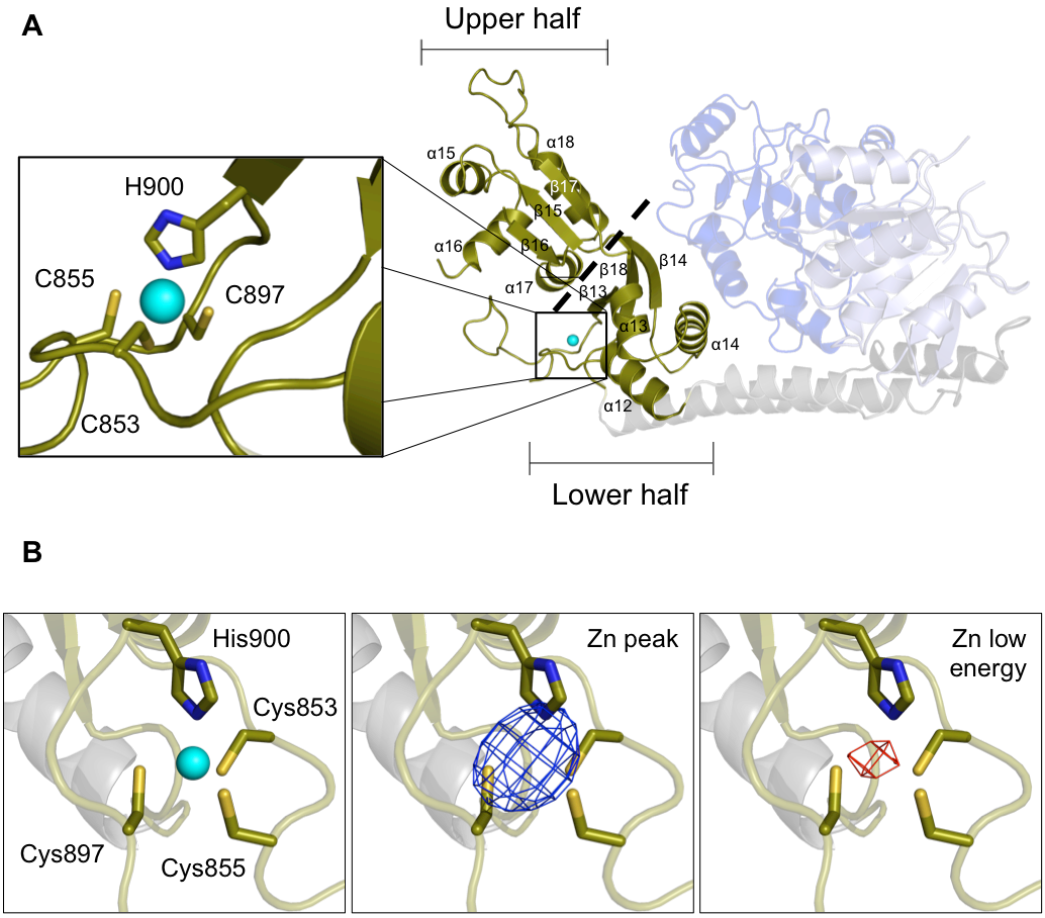


Figure 4-9: Structure of the R4ZBD. (A) The R4ZBD can be separated into an upper half and a lower half as indicated by the dashed line. Both halves comprise a three-stranded anti-parallel β-sheet and three (lower half) or four (upper half) α-helices. The lower half features a bound zinc ion (cyan sphere), which is coordinated via three cysteine residues (C853, C855, C897) and a histidine (H900). The insert provides an enlarged view on the zinc binding site, depicting the respective coordinating residues in stick representation. **(B)** Confirmation of the bound zinc ion by recording diffraction data at different wavelengths. **Left:** Position of the zinc ion (cyan sphere) within the R4ZBD. **Middle:** Anomalous density (blue mesh) at the zinc peak absorption energy (1.2823 Å). **Right:** Anomalous density (red mesh) at low energy (1.2831 Å). Both maps are shown at 10 σ. The anomalous signal at the zinc absorption peak energy is severely reduced for an energy level that is below the zinc absorption peak, thus confirming the coordination of the zinc ion within the R4ZBD domain. This Figure has been adapted from Kaiser *et al.* (2017).

Table 4-1: Data collection, phasing and refinement statistics

	High Resolution[1]	SAD Phasing[1]	Zinc Detection[2]	
Data collection				
Space group	$P3_121$	$P3_121$	$P3_121$	
Cell dimensions				
a, *b*, *c* (Å)	131.00, 131.00, 96.34	130.06, 130.06, 95.84	128.96, 128.96, 94.68	129.19, 129.19, 94.93
α, β, γ (°)	90, 90, 120			
			Peak	*Low Energy*
Wavelength	0.9763	1.2828	1.2823	1.2831
Resolution (Å)	44.34 – 2.75 (2.90 – 2.75)	48.56 -2.91 (3.09 – 2.91)	48.10 – 3.00 (3.18 – 3.00)	48.20 – 3.00 (3.18 – 3.00)
R_{pim}	1.9 (51.6)	2.3 (27.2)	2.5 (27.3)	2.4 (34.0)
CC1/2	99.9 (76.0)	100 (87.0)	99.9 (94.9)	100 (100)
$I / \sigma I$	21.7 (1.4)	25.0 (3.1)	26.0 (4.1)	27.9 (3.5)
Completeness (%)	99.8 (100)	99.5 (97.2)	100 (100)	100 (100)
Redundancy	14.3 (14.3)	19.6 (20.1)	19.8 (20.9)	19.9 (20.9)
Refinement				
Resolution (Å)	44.34 – 2.75			
No. reflections (free)	47984 (2459)			
R_{work} / R_{free}	18.3 / 23.1			
No. atoms				
Protein	4692			
Zn^{2+}	1			
B-factors				
Protein	108.8			
Zn^{2+}	83.1			
R.m.s deviations				
Bond lengths (Å)	0.004			
Bond angles (°)	0.681			

[1] High resolution and SAD datasets were recorded on individual crystals
[2] Peak- and low-energy datasets were recorded on the same crystal

4.2.2.2. Structural analysis of the R4ZBD

An analysis of the order of secondary structure elements revealed that the R4ZBD upper half is an insertion within the lower half, placed between β-strands β14 and β18, thus suggesting that the respective halves could exhibit individual functions (Fig. 4-10 A). In order to identify such potential functions, the R4ZBD structure – including the proximal half of bridging helix α19 (aa 836-1060) – was subjected to a DALI search against the PDB-90 database, which recognized homologous three-dimensional folds in 246 individual protein structures. The vast majority of these proteins (221/246) exhibited homology exclusively to secondary structure elements of the R4ZBD lower half. The ten most similar structures comprise nine bacterial transcription factors and one DNA-replication initiation factor from *E. coli* (Table 4-2). All of these proteins interact with dsDNA via a winged helix (WH)-like DNA-binding motif and the homology of the R4ZBD lower half is consistent with those WH-like DNA-interaction modules (Fig. 4-10 B). The key element of a prototypical WH-like DNA-interaction domain is the helix-turn-helix (HTH) motif, which mediates contacts with the DNA via an insertion of the second helix into the major or the minor DNA groove[244,245]. Based on the structural homology to WH-like domains, a putative HTH motif could also be identified within the lower half of the R4ZBD (Fig. 4-10 B). In addition, the structure of the isolated R4ZBD upper half (aa 942-1032) was subjected to a DALI PDB-90 search, however, compared to the entire R4ZBD, the outcome was less consistent regarding the function of the identified homologues. For the isolated R4ZBD upper half, the ten most homologous structures comprise two hydrolases, two transcription factors, an RNA-binding protein, a ligase, a transferase and proteins of unknown function, while the human methionine aminopeptidase 2 (PDB 5d6e) constitutes the closest homologue (Table 4-3). However, one of the top ten homologous structures for the isolated R4ZBD upper half is the bacterial transcription factor CtSR from *Geobacillus stearothermophilus*, a candidate that was also among the homologues of the R4ZBD lower half. Likewise, the structural homology between the R4ZBD

upper half and CtSR converges with the WH-like DNA-binding module of CtSR, suggesting that the R4ZBD upper half also represents a WH-like fold (Fig. 4-10 B). Similar to the R4ZBD lower half, the R4ZBD upper half also features a putative HTH motif (Fig. 4-10 A). Moreover, taking the particular order of secondary structure elements of the lower and upper R4ZBD WH domains into account, both folds can be assigned to two individual classes of WH domains. While the lower R4ZBD half represents the typical fold of a classical three-stranded WH domain, the upper half WH domain constitutes the order of secondary structure elements that is characteristic for the so-called methionine aminopeptidase-like WH domain[244] (Fig. 4-10 A).

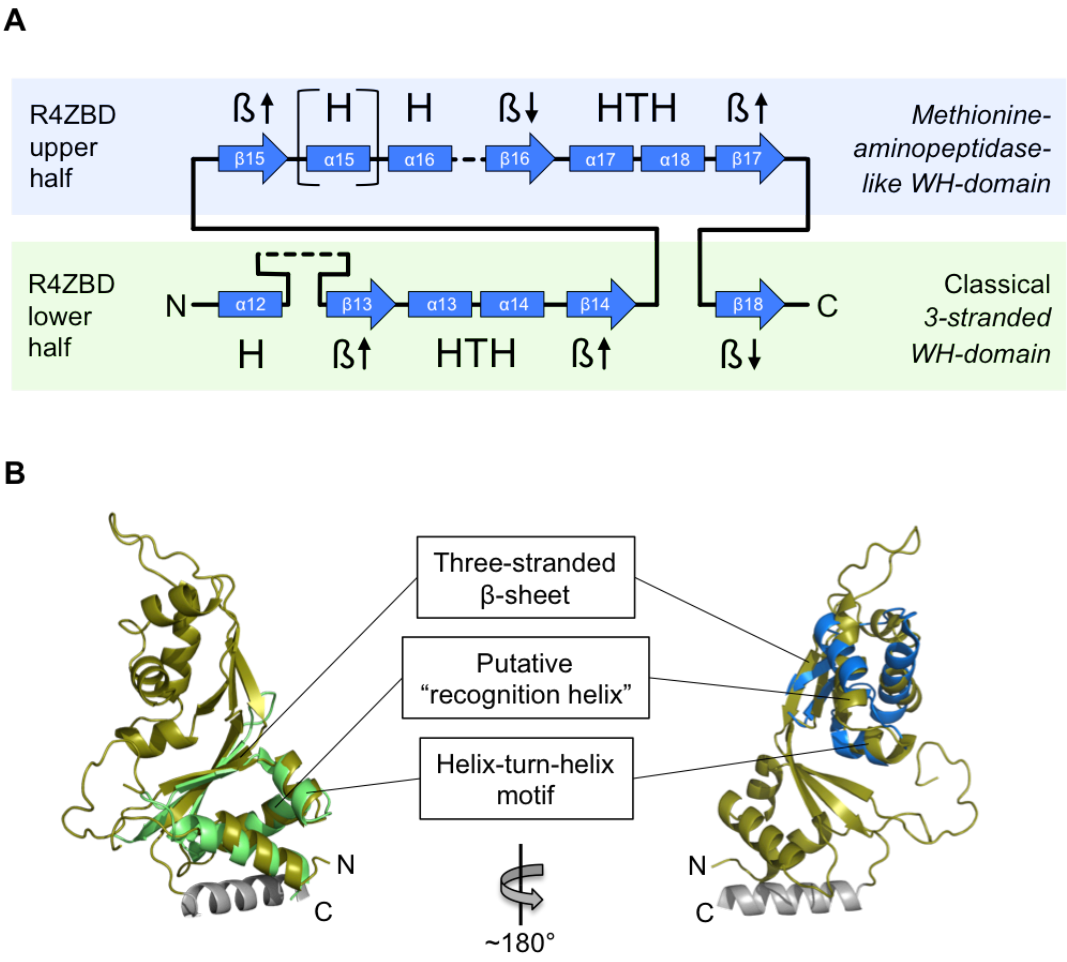


Figure 4-10: Structural analysis of the R4ZBD. (A) Arrangement of R4ZBD secondary structure elements for the upper and lower R4ZBD halves with α-helices and β-strands abbreviated as H and β, respectively. Annotation of secondary structure elements is as depicted in Fig. 4-9 A. The upper half represents an insertion within the lower half of the R4ZBD. The specific order of secondary structure elements of the lower half (green box) is characteristic for a classical three-stranded WH domain, while the arrangement of secondary structure elements of the upper half (blue box) is typical for a Methionine-aminopeptidase-like WH domain. **(B)** Superposition of WH domains identified by the DALI homology search with the lower and upper R4ZBD folds. **Left:** The WH domain of the bacterial transcription factor CtsR (green, PDB 3h0d). Right: WH domain of the human methionine aminopeptidase 2 (blue, PDB 5d6e). Orientation of the R4ZBD on the left is depicted as in Fig. 4-9 A, while the right side depicts the R4ZBD rotated by approx. 180° with regard to the orientation on the left. The proximal half of bridging helix α19 (depicted in gray) has been included into the homology search, yet was not present in any of the identified homologues. This Figure has been adapted from Kaiser *et al.* (2017).

Table 4-2: Top ten structural homologues of the DALI PDB search for the entire R4ZBD (aa 863-1060). Matches were obtained from the PDP90 database and are ranked according to their Z-value, representing a measure of homology. Abbreviations: Root mean square deviation (RMSD). Number of aligned amino acids (align). % Identity (% id.).

No	PDB / Chain	Z-value	RMSD	align	% id.	Protein class*
1	2fmy A	6.7	2.8 Å	71	15	DNA binding protein
2	3h0d A	6.6	2.6 Å	67	4	Transcription / DNA
3	1o57 A	6.3	2.5 Å	67	9	DNA binding protein
4	3iwz D	6.3	2.8 Å	65	14	Transcription
5	4r6i A	6.1	2.9 Å	79	11	Transcription
6	2oz6 A	6.1	2.2 Å	62	13	DNA binding protein
7	1ylf C	5.9	2.8 Å	79	9	Transcription regulation
8	5e44 A	5.6	3.1 Å	70	14	Transcription
9	2xhk B	5.6	2.7 Å	67	15	Transcription
10	2z9o A	5.6	3.2 Å	72	6	Replication / DNA

* according to the PDB classification

Table 4-3: Top ten structural homologues of the DALI PDB search for the isolated upper R4ZBD WH domain (aa 942-1032). Matches were obtained from the PDP90 database and are ranked according to their Z-value, representing a measure of homology. Abbreviations: Root mean square deviation (RMSD). Number of aligned amino acids (align). % Identity (% id.).

No	PDB / Chain	Z-value	RMSD	align	% id.	Protein class*
1	5d6e A	4.3	3.0 Å	59	19	Hydrolase
2	3ife A	3.8	3.2 Å	61	8	Hydrolase
3	3h0d A	3.8	2.5 Å	49	10	Transcription / DNA
4	4ney B	3.8	3.6 Å	59	8	De novo protein
5	3rir A	3.7	2.4 Å	47	9	Ligase
6	2mzj A	3.7	3.1 Å	61	10	RNA binding protein
7	1u8s B	3.7	2.9 Å	54	6	Transcription
8	3mgj A	3.6	3.4 Å	55	4	Unknown function
9	3mah A	3.5	2.4 Å	49	8	Transferase
10	2nzc D	3.5	2.8 Å	52	13	Unknown function

* according to the PDB classification

4.2.2.3. Structural comparison to other RecQ helicases

To analyze the structural similarities and differences between *Hs*RecQ4 and other RecQ helicases, the RecQ4$^{427\text{-}1116}$ model was superimposed with the structure of the bacterial *Ec*RecQ helicase (PDB 1oyw) and the structures of human *Hs*RecQ1 (PDB 2wwy) and *Hs*BLM (PDB 4o3m) (Fig. 4-11 A-D). HD2 represents the domain with the highest structural homology as well as the highest rigidity among all four RecQ helicases and therefore served as the structural basis for the superposition. Although structurally well conserved, HD1 adopts a range of different orientations towards HD2, illustrating the inherit degree of flexibility between the two RecA-like domains relative to each other. The structures of *Ec*RecQ, *Hs*RecQ1 and *Hs*BLM exhibit a conserved position of the RQC domain, which is located adjacent to HD2 (Fig. 4-11 B-D). In contrast, the R4ZBD of RecQ4 is located directly adjacent to HD1, thus assuming a unique position among the superimposed RecQ structures. The superposition further demonstrates that the R4ZBD and RQC domain each assume a unique structural fold. As described above, the R4ZBD exhibits an interconnected arrangement of two WH-like folds, each comprising a three-stranded anti-parallel β-sheet and a three-, respectively four-helix bundle that includes a putative HTH motif. RQC domains, on the other hand, consist of an individual zinc-binding domain (Zn) followed by a WH-fold (WH). The Zn domain is composed of four α-helices, and the last two in-line oriented α-helices coordinate a zinc ion via four cysteine residues (Fig. 4-12). The central element of the subsequent RQC-WH domain is the two-stranded β-hairpin, which is preceded by a helical bundle comprising three to four α-helices. Two additional β-strands, one N-terminal and one C-terminal to the WH-fold, form a second antiparallel β-sheet, which structurally separates the WH domain from the Zn domain. Regarding the specific classification of WH domains, the RQC-WH fold closely resembles the so-called “wHTH (winged HTH) with C-terminal helix”-class of WH domains[244] (Fig. 4-12), while the aforementioned N- and C-terminal β-strands are not

determinative for this WH class and can thus be considered as RecQ-characteristic structural extension elements of this specific WH group.

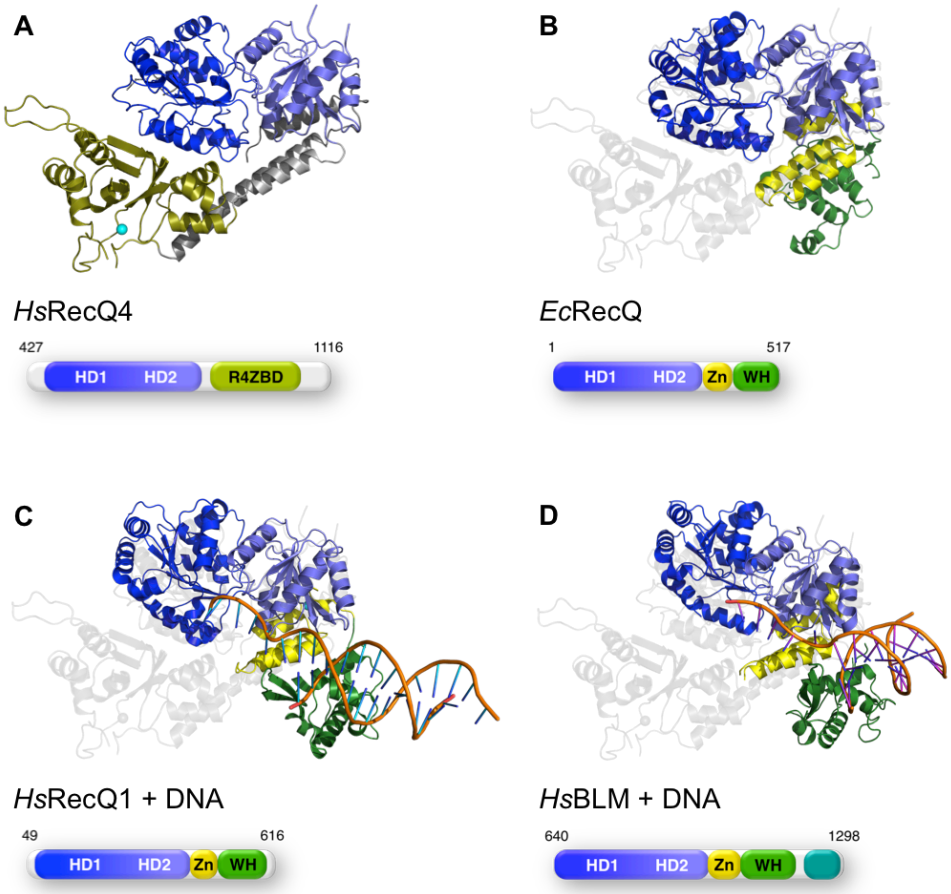


Figure 4-11: Structural comparison of RecQ4$^{427\text{-}1116}$ with other RecQ helicases. The model of human RecQ4$^{427\text{-}1116}$ **(A)** is superimposed with the structure of the bacterial RecQ helicase from *E. coli* **(B)** and the human RecQ structures of RecQ1 **(C)** and BLM **(D)**. All structures were superimposed with the isolated HD2 of RecQ4. The structure of *Hs*Recq4 is depicted in gray within each alignment for convenient comparison. The protein constructs that were crystallized are indicated below the structures. The alignments demonstrate that the R4ZBD of RecQ4 assumes a unique structural fold that does not resemble the typical structure of the RQC domain. Moreover, the R4ZBD acquires a unique position within the RecQ4 structure, adjacent to HD1, while the RQC domains of all other RecQ helicases are located in vicinity to HD2. The two DNA-bound complex structures **(C, D)** depict a conserved mode of DNA substrate binding by RecQ helicases, with the ssDNA entity bound across the helicase core and the dsDNA segment bound via the RQC WH domain. Structural alignments were performed in pymol. Color code and abbreviations as in Figure 1-2. PDB codes: *Hs*RecQ4 – 5LST. *Ec*RecQ – 1OYW. *Hs*RecQ1 – 2WWY. *Hs*BLM – 4O3M. Note: The *Hs*BLM structure (4O3M) also comprises the HRDC domain, which was omitted from the image for the purpose of a uniform comparison. This Figure has been adapted from Kaiser *et al.* (2017).

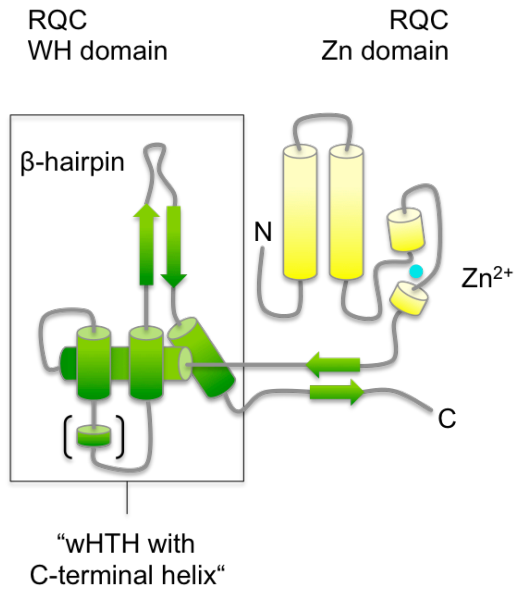


Figure 4-12: Secondary structure arrangement of the RQC domain. The RQC Zn domain (yellow) encompasses four α-helices, while the two short C-terminal helices coordinate the zinc ion (cyan sphere) between them. The conserved RQC WH domain (green) constitutes the typical fold of a "winged HTH motif (wHTH) with a C-terminal helix". The "wing" of this motif represents the antiparallel β-hairpin that is required for dsDNA strand separation. A short α-helix in the WH domain (depicted in brackets) is only occasionally present in the RQC domain WH structures.

4.2.2.4. RecQ4 mimics the conserved RQC-Zn domain

The superposition of the RecQ4 model with the structures of *Ec*RecQ, *Hs*RecQ1 and *Hs*BLM revealed an additional interesting structural feature of the RecQ4 protein. The conserved RQC Zn domain, which is present in all known RecQ helicase structures other than RecQ4, is mimicked by the two RecQ4 bridging helices, α11 and α19, as well as helix α20 (Fig. 4-13 A-C). The RecQ4 bridging helix α11 and the distal half of bridging helix α19 occupy the positions of the non-metal coordinating helices of the RQC Zn domain, while helix α20 adopts the location of the two in-line oriented zinc-binding helices. However, in contrast to all other RecQ helicases, RecQ4 does not coordinate a zinc ion at this position and does not feature any cysteine residues within secondary structure elements that mimic the helices of the RQC Zn domain, but two cysteine residues are present within the loop between bridging helix α19 and helix α20. Furthermore, an amino acid sequence analysis did not indicate any homology between the Zn domains of *Ec*RecQ, *Hs*RecQ1 or *Hs*BLM and the respective amino acid sequence of RecQ4, suggesting independent evolutionary pathways for the RQC Zn domain and their RecQ4 counterparts.

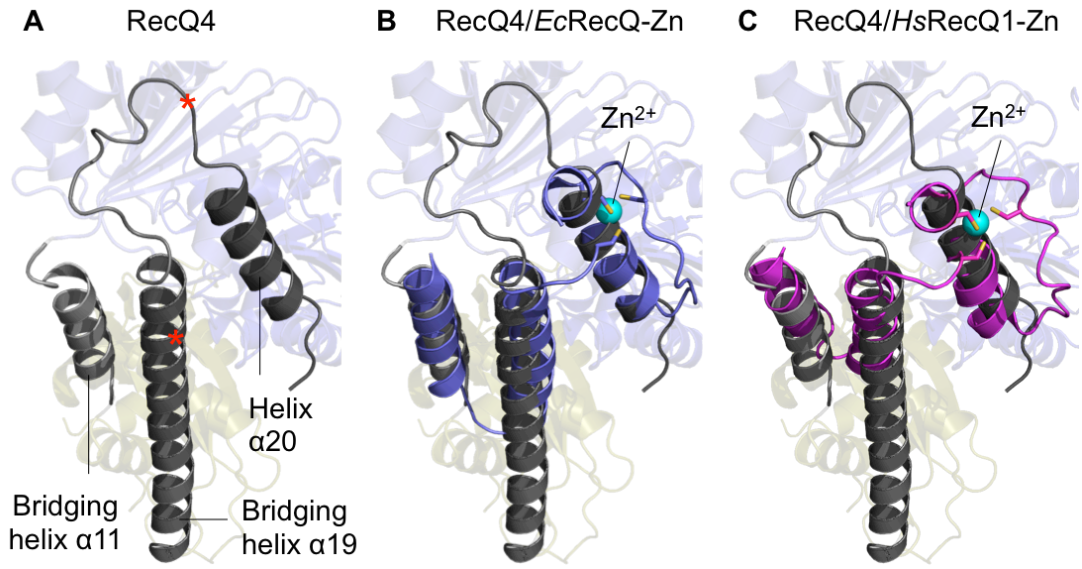


Figure 4-13: Three α-helices of RecQ4 mimic a RQC Zn domain-like arrangement. **(A)** The bottom perspective of the RecQ4$^{427\text{-}1116}$ structure depicts the location of the two bridging helices α11 and α19 as well as helix α20 (shown in gray). Red asterisks highlight the location of two early termination mutations within RecQ4 (R1072X and Q1091X), which are associated with the RAPADILINO syndrome (see Discussion). **(B)** Superposition of the three RecQ4 helices shown in A with the RQC Zn domain of *Ec*RecQ (depicted in blue). **(C)** Superposition of the three RecQ4 helices shown in A with the RQC Zn domain of *Hs*RecQ1 (depicted in magenta). **(B,C)** The superposition illustrates that the RecQ4 bridging helix α11, together with the distal half of bridging helix α19 and helix α20, occupy the exact same positions as the RQC Zn domain helices in *Ec*RecQ and *Hs*RecQ1. Alignments were extracted from the structural alignments shown in Figure 4-11. This Figure has been adapted from Kaiser *et al.* (2017).

4.3. Functional characterization of RecQ4 variants

4.3.1. RecQ4$^{427-1116}$ and RecQ4$^{427-1208}$

4.3.1.1. DNA binding studies via EMSA analyses

In order to investigate the DNA binding behavior of RecQ4$^{427-1116}$ and RecQ4$^{427-1208}$, increasing protein concentrations of each RecQ4 variant were incubated with a variety of different DNA substrates and subsequently analyzed on non-reducing SDS-gels via the electro mobility shift assay (EMSA) (Fig. 4-14 A-D). The tested DNA substrates included a Y-forked DNA with a 15-nucleotide (nt) dsDNA segment (generic sequence) and 15nt unpaired polyT flanks on the 3' end of the top DNA strand and the 5' end of the bottom DNA strand (Y-15-15), a 15nt 3'-overhang (OH) substrate (15nt3'OH), a 15nt 5' OH substrate (15nt5'OH) and a 15nt blunt end dsDNA.
The analysis reveals that both RecQ4 variants require a single-stranded DNA segment for efficient DNA binding as none of the 15nt blunt end DNA substrate was efficiently bound by either RecQ4 variant (Fig. 4-14 D). Both, RecQ4$^{427-1116}$ and RecQ4$^{427-1208}$, exhibit a preference for the Y-fork DNA compared to the 15ntOH DNA substrates. In comparison, RecQ4$^{427-1208}$ binds to Y-15-15 even more tightly than RecQ4$^{427-1116}$ as the binding event can be detected already at a 1:2 ratio (Fig. 4-14 A). The gradual appearance of three, respectively four species of protein-DNA complexes with increasing protein concentration suggests that Y-15-15 is gradually bound by multiple RecQ4 proteins with a maximum of three molecules for RecQ4$^{427-1116}$ or four molecules of RecQ4$^{427-1208}$, respectively. In comparison to Y-15-15, the 15nt3'OH substrate is significantly better bound by the C-terminally extended RecQ4$^{427-1208}$ than by the truncated RecQ4$^{427-1116}$ variant. While the former RecQ4 variant exhibits a single strong DNA-bound signal, that can be detected at a 1:4 ratio, RecQ4$^{427-1116}$ hardly binds to the 15nt3'OH DNA substrate as indicated by the very faint signals at the two highest protein concentrations (Fig. 4-14 B). A similar binding behavior for both RecQ4 variants was detected for the 15nt5'OH DNA substrate (Fig. 4-14 C).

However, in contrast to the 3'OH DNA, RecQ4$^{427-1208}$ exhibits two distinctive DNA-bound signals for the 15nt5'OH DNA, suggesting that the former DNA substrate is bound via one RecQ4 molecule while the latter is bound by up to two RecQ4$^{427-1208}$ molecules. Moreover, compared to the 15nt3'OH substrate, the 5'OH DNA seems to be bound less efficiently as the signals for the 15nt5'OH DNA-bound species start to appear at slightly higher protein concentrations for both, RecQ4$^{427-1116}$ and RecQ4$^{427-1208}$.

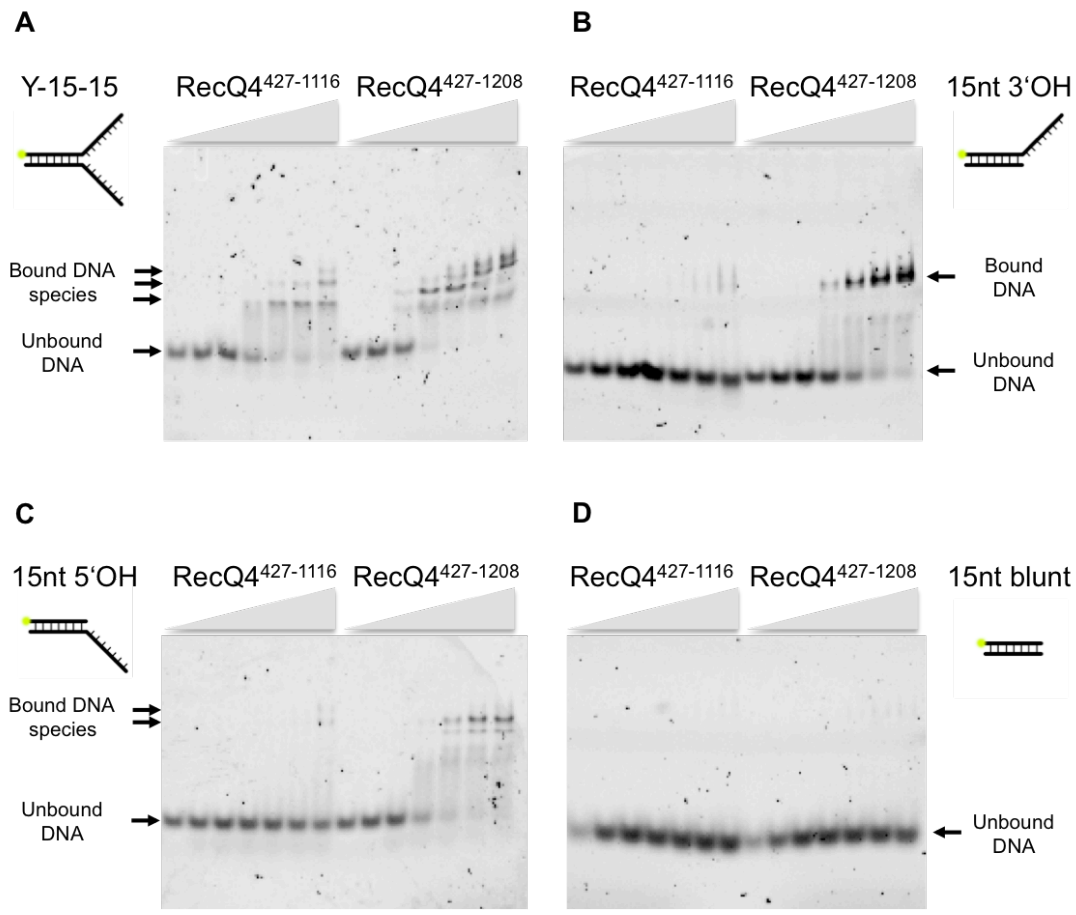


Figure 4-14: DNA binding analysis via EMSA studies of various DNA substrates. 100 nM of the DNA substrate were incubated with increasing molar ratios of either the RecQ4$^{427-1116}$ or the RecQ4$^{427-1208}$ variant (1:0, 1:1, 1:2, 1:4, 1:6, 1:8 and 1:10) and analyzed via non-reducing SDS-PAGE. **(A)** Binding of a Y-fork DNA substrate (Y-15-15) comprising 15nt polyT ssDNA flanks and a 15nt dsDNA segment. **(B)** Binding of a 15nt polyT 3'OH substrate (15nt3'OH). **(C)** Binding of a 15nt polyT 5'OH substrate (15nt5'OH). **(D)** Binding of a 15nt blunt end dsDNA substrate. A yellow dot on the DNA substrates indicates the modification of the top oligonucleotide with a cyanine-3 (Cy3) fluorescent label. Bands representing bound and unbound DNA species are indicated.

In order to analyze the potential influence of different additives on the DNA binding behavior of RecQ4$^{427-1208}$, a series of EMSAs were performed in the presence of 0.1 mg/mL BSA, 1 mM ADP, 1 mM of the non-hydrolysable ATP analogue AMP-PNP, 150 mM KCl and 30% (v/v) glycerol (Fig. 4-15).

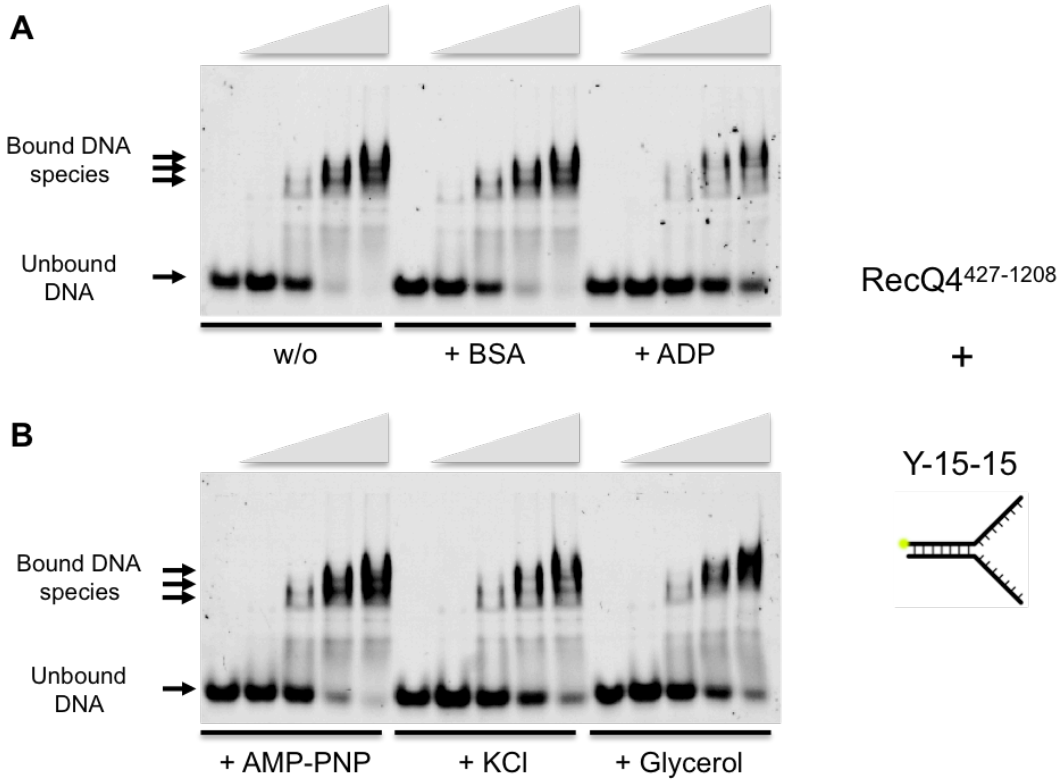


Figure 4-15: EMSA analysis in the presence of different additives. The influence of several additives on the binding of the Y-15-15 DNA substrate by RecQ4$^{427-1208}$ was analyzed. Increasing molar ratios of the RecQ4 variant (1:0, 1:2, 1:4, 1:6, 1:8) were incubated with 100 nM DNA substrate and analyzed via non-reducing SDS-PAGE. **(A)** Additives: none (w/o), 0.1 mg/mL BSA, 1mM ADP. **(B)** Additives: 1 mM AMP-PNP, 150 mM KCl, 30% (v/v) glycerol. Bands representing bound and unbound DNA species are indicated.

The EMSA analysis demonstrates that 0.1 mg/mL BSA and 1 mM AMP-PNP had no detectable influence on the DNA binding behavior of RecQ4$^{427-1208}$, while the presence of 150 mM KCl or 30% (v/v) glycerol led to a moderate decrease in DNA binding efficiency. Furthermore, 1 mM ADP supplementation resulted in a significantly reduced DNA binding behavior, suggesting that an ADP-bound conformation of the two RecA-like domains is

less prone to bind the Y-15-15 DNA substrate than a conformational state of the helicase core where either ATP/AMP-PNP or no nucleotide is bound.

4.3.1.2. Measuring DNA binding via fluorescence polarization

The EMSA studies were useful to differentiate between different substrates regarding their binding preference to the RecQ4 variants. Measuring fluorescence polarization (FP) data, on the other hand, represents a means to directly determine the amounts of free as well as bound DNA ligand at unperturbed equilibrium conditions in solution and thus permits the quantification of apparent binding parameters, e.g. the dissociation constant K_D. Accordingly, FP data were recorded for the two RecQ4 variants in order to further characterize their binding behavior to a range of different DNA substrates.

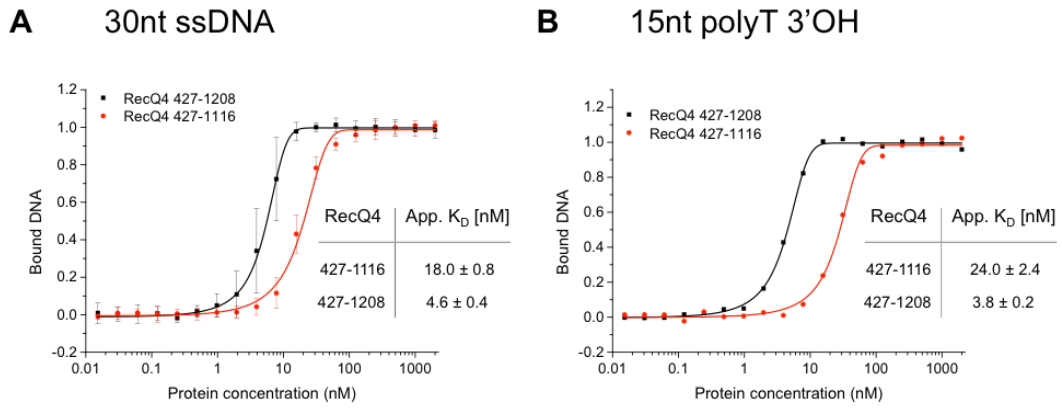


Figure 4-16: Measuring DNA binding via fluorescence polarization (FP). (A) Binding of a 30nt ssDNA substrate to RecQ4$^{427\text{-}1208}$ (black) and RecQ4$^{427\text{-}1116}$ (red). Both RecQ4 variants bind the ssDNA substrate with low nanomolar affinity. The apparent dissociation constant K_D is represented by the specific protein concentration at half-maximal binding. FP experiments have been performed in triplicates using protein from two different batches. Error bars represent 1 SD from the mean. This figure has been adapted from Kaiser *et al.* (2017). **(B)** DNA binding data for the 15nt3'OH substrate (as used in the helicase assays). DNA binding affinities for both RecQ4 variants are comparable to the 30nt ssDNA substrate.

The FP data for a 30nt ssDNA substrate revealed a low-nanomolar binding affinity for both RecQ4 variants (Fig. 4-16 A). With an apparent K_D of 18.0 ± 0.8 nM, the truncated RecQ4$^{427\text{-}1116}$ exhibited a weaker binding than the full C-terminal RecQ4$^{427\text{-}1208}$, for which an apparent K_D of 4.6 ± 0.4 nM was determined. Similar dissociation constants were observed for a 15nt polyT 3'OH DNA substrate with a 15nt dsDNA segment (15nt3'OH). Here, RecQ4$^{427\text{-}1116}$ exhibited an apparent K_D of 24 ± 2.4 nM, while RecQ4$^{427\text{-}1208}$ bound the 15nt3'OH substrate more efficiently with an apparent K_D of 3.8 ± 0.2 nM (Fig. 4-16 B).

Regarding the potential influence of additives on the binding characteristics of RecQ4 to a DNA substrate, FP experiments were also performed in the presence of 1 mM APM-PNP or 1 mM ADP (Fig. 4-17).

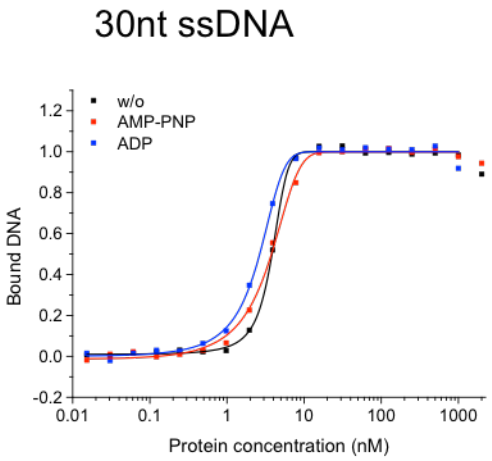


Figure 4-17: The influence of additives on DNA binding. Fluorescence polarization (FP) data were recorded for the binding of a 30nt ssDNA substrate by RecQ4$^{427\text{-}1208}$. The binding experiments were performed without any additive (black) or supplemented with 1 mM AMP-PNP (red) and 1 mM ADP (blue).

In comparison to the EMSA analysis, the FP data confirms that the addition of 1 mM AMP-PNP has no influence on the overall binding of RecQ4$^{427\text{-}1208}$ to the 30nt ssDNA substrate, as the data (red) overlaps well with the FP data for DNA substrate binding without (w/o) any additive (black). Similarly, the data for the addition of 1 mM ADP (blue) depicts the same binding characteristics as without additive in the FP experiments. In contrast, the EMSA studies indicated a significant reduction in the presence of 1 mM ADP. The reason for this discrepancy could be the use of different DNA substrates

that were utilized in both studies. While ssDNA is a rather flexible DNA substrate regarding the binding by the RecQ4 variants, the Y-forked DNA is more rigid and might thus be more easily influenced by additives that possibly alter the orientation of the two RecA-like domains towards each other, thus likely affecting the location of DNA-binding residues within. In this regard, testing the influence of additives on the DNA binding behavior of RecQ4 via FP experiments must be repeated using more physiological DNA substrates like Y-forked DNA or a 15nt3'OH substrate in order to obtain an unambiguous result.

Testing the binding of the two RecQ4 variants to a 15nt blunt end dsDNA substrate via FP revealed an unexpected behavior. In contrast to the EMSA studies, which illustrated a lack of binding for either RecQ4 variant to this particular DNA substrate, DNA binding analysis via FP indicated an interaction of the DNA substrate with RecQ4 (Fig. 4-18).

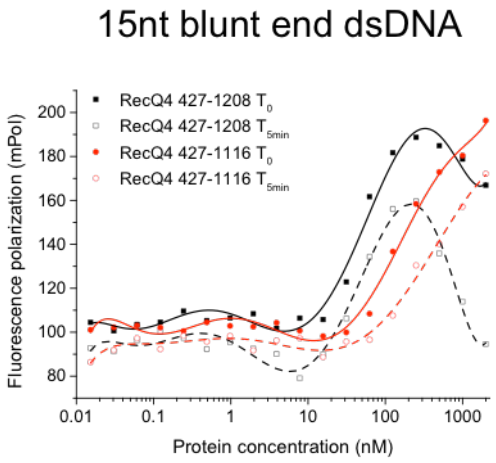


Figure 4-18: Fluorescence polarization (FP) DNA binding data for a 15nt blunt end DNA substrate. The graph illustrates the binding of a 15nt blunt end DNA substrate by RecQ4$^{427\text{-}1208}$ (black squares) and RecQ4$^{427\text{-}1116}$ (red circles), while full squares/circles represent the raw FP values at time point T_0 and empty squares/circles depict the raw FP data at a time point five minutes later (T_{5min}). FP values below a protein concentration of 10 nM fluctuate around 100 mPol, indicating unbound DNA, and start to rise for protein concentrations above 10 nM, suggesting DNA binding. The data were fitted by a 7th order polynomial fit. The curves for the different time points are non-identical for each RecQ4 variant, suggesting that the data were not recorded at equilibrium conditions.

As expected, the polarization values for the lowest RecQ4 protein concentrations were around 100 mPol (millipolarization unit, relative value), indicative for unbound DNA substrate (filled squares/circles and solid lines). However, FP values started to increase, beginning at a protein concentration of 10 nM or 50 nM and above for RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$, respectively, suggesting the physical interaction of the 15nt blunt end DNA substrate with RecQ4. In contrast to previous FP measurements, however, the binding data for both RecQ4 variants did not reach a discernable plateau phase and even declined for RecQ4$^{427\text{-}1208}$ protein concentrations above 250 nM. A typical FP experiment exhibited raw polarization values for 0.5 nM of a DNA substrate in the range of 100 mPol, representative for unbound substrate, to approximately 270 mPol for a completely bound DNA substrate (compare Fig. 4-19 A). However, the maximum values for the 15nt blunt end DNA substrate were entirely below 200 mPol for both RecQ4 variants, suggesting that the blunt end DNA substrate fraction is never entirely bound by either RecQ4 variant. Due to the lack of a plateau, conventional Boltzmann fitting was not applicable. Thus, in order to illustrate an appropriate approximation of the binding data, a 7th order polynomial fitting function was used instead. Furthermore, collecting polarization data on the exact same experiments five minutes after they have been initially recorded (empty squares/circles and dashed lines) revealed a deviation from the original data, suggesting that the initial experiment has not been recorded under equilibrium conditions. Three general trends can be observed regarding the change of FP values between the two time-separated experiments. (1) Apparent binding of the DNA substrate occurred at higher protein concentrations for both RecQ4 variants in the later data collection. (2) FP values are generally higher for the first experiment (T_0) than for the second data collection (T_{5min}) for both RecQ4 variants. (3) The decrease in the polarization values for the highest protein concentrations of the RecQ4$^{427\text{-}1208}$ variant is more pronounced for the second data collection (T_{5min}) than for the first run (T_0). Collectively, the data suggest that a fraction of the fluorescently labeled dsDNA substrate is bound by RecQ4, which was

followed by a subsequent processing of the DNA substrate that results in its physiological time-dependent alteration. Such a substrate processing could for example be achieved by an unexpected nuclease activity or an ATP independent DNA strand separation.

In order to gain further insights into these processes, two additional sets of FP experiments were performed. In a first experiment, FP data were recorded for the binding of the RecQ4$^{427\text{-}1208}$ variant to a 30nt ssDNA substrate at three different time points after the experiment was prepared (T_0, T_{5min} and T_{20min}) in order to detect a possible change of FP values over the course of 20 minutes (Fig. 4-19 A). The data illustrates that, irrespective of the time when the experiment was recorded, RecQ4$^{427\text{-}1208}$ bound to the 30nt ssDNA substrate in the same manner, generally exhibiting low-nanomolar affinities (7.6 nM, 7.6 nM and 8.4 nM for T_0, T_{5min} and T_{20min}, respectively). Especially over the first five minutes, the FP data were practically identical. The slight decrease in the maximum FP values for highest protein concentrations after 20 minutes can be ascribed to a potential nonspecific binding of the fluorescent dye to the reaction container, which would slightly decrease the available overall DNA substrate concentration. In general, however, the experiment rules out that RecQ4 degrades the ssDNA substrate over time, e.g. via a nuclease activity. In an additional set of FP experiments, the binding of RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ to either a 30nt ssDNA or a 15nt ssDNA substrate was recorded (Fig. 4-19 B). The data for the 30nt ssDNA (filled squares/circles and solid lines) exhibited the typical binding characteristics for both RecQ4 variants, as they have been described earlier. However, the binding data for the 15nt ssDNA substrate (empty squares/circles and dashed lines) illustrate some of the unexpected features that have been observed for the 15nt blunt end DNA substrate, e.g. a reduced binding affinity, the lack of a maximum binding plateau and the decline of FP values at the highest protein concentrations. Thus, regarding the blunt end DNA binding FP experiment (Fig. 4-18), it may be assumed that RecQ4 exerts an ATP-independent DNA strand separation mechanism

on the 15nt sdDNA substrate under the given experimental conditions, resulting in a fraction of fluorescently labeled 15nt ssDNA that is subsequently bound by RecQ4, leading to the DNA binding characteristics as observed for the 15nt ssDNA substrate (Fig. 4-19 B). However, why the binding of a 15nt ssDNA substrate is subsequently decreased at the highest protein concentrations cannot be explained at this point and requires further analysis.

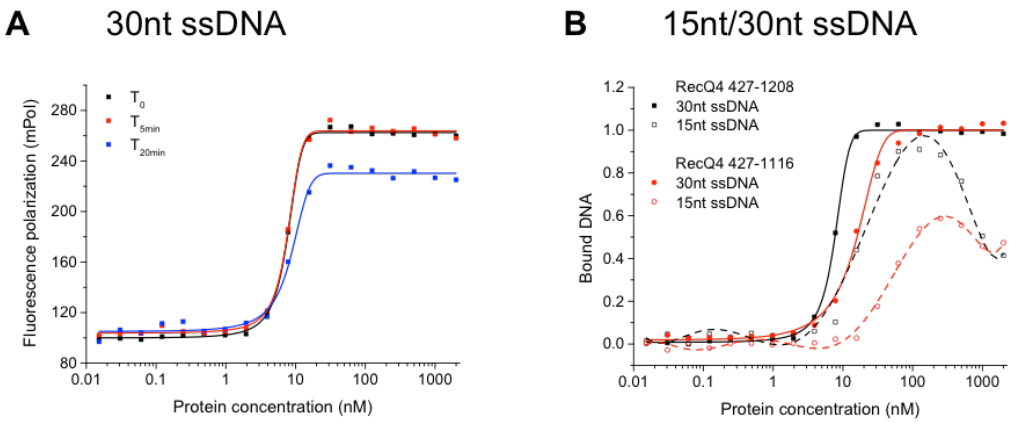


Figure 4-19: DNA binding for different ssDNA lengths at various time points. (A) Fluorescence polarization (FP) raw data for the binding of a 30nt ssDNA substrate by RecQ4$^{427\text{-}1208}$ at time point T_0 (black) as well as five (red) and twenty (blue) minutes after T_0 (T_{5min} and T_{20min}, respectively). **(B)** Normalized DNA binding data for two ssDNA substrates by RecQ4$^{427\text{-}1208}$ (black squares) and RecQ4$^{427\text{-}1116}$ (red circles). Full squares/circles and strait fits represent the data for a 30nt ssDNA substrate, while empty squares/circles and the dashed fits represent the binding data for a 15nt ssDNA substrate. The FP data for the 30nt ssDNA was fitted by the Boltzmann equation. The data for the 15nt ssDNA substrate was fitted by a 7th order polynomial fitting function.

4.3.1.3. *In vitro* helicase activity

The analysis of RecQ4 helicase activity was assessed by fluorescence-based *in vitro* helicase assays. During each experiment, the dsDNA strand separation activity of the RecQ4 variants was recorded by detecting the fluorescence intensity of a labeled DNA substrate (e.g. 5' top DNA strand labeled with cyanine-3). In a state where the fluorescently labeled top DNA

strand is bound by its complementary bottom oligonucleotide, forming the dsDNA segment, the fluorescence of the labeled top DNA strand is maximally suppressed by a respective quencher-molecule that is attached to the bottom strand (e.g. bottom DNA strand 3'-dabcyl). ATP-dependent DNA strand separation activity was initiated by the injection of ATP (time point T_0) and the change of fluorescence intensity was recorded over time as the quencher is separated from the fluorescent dye. Using this setup, the helicase activity of both variants, RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$ was investigated for a variety of different DNA substrates as well as varying assay conditions, e.g. different ATP- or protein concentrations.

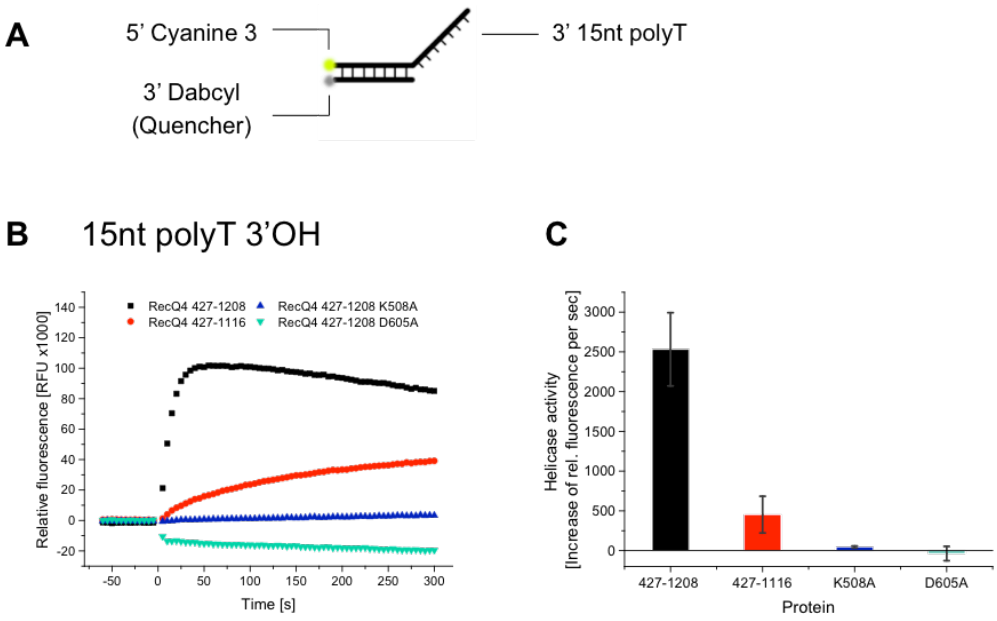


Figure 4-20: Helicase activities of RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$. (A) Illustration of the DNA substrate used for fluorescence-based helicase assays (15nt3'OH), featuring a 5' fluorescent cyanine-3 and a 3' 15nt polyT overhang (OH) on the top DNA strand and a 3' dabcyl quencher molecule at the complementary bottom DNA strand. **(B)** Typical fluorescence-based helicase assay performed on RecQ4$^{427\text{-}1208}$ (black), RecQ4$^{427\text{-}1116}$ (red) and the two ATPase impaired Walker mutants RecQ4$^{427\text{-}1208}$ K508A (blue) and RecQ4$^{427\text{-}1208}$ D605A (turquoise). Helicase activity was initiated by adding ATP at time point T_0. Increasing fluorescence indicates the separation of the Cy3-labeled top DNA strand from the quencher (dabcyl) labeled bottom DNA strand via ATP-dependent helicase activity of the RecQ4 variants. **(C)** Quantification of three individual helicase assays as depicted in A. Bars represent the linear increase in relative fluorescence 10 s after ATP addition. Error bars are 1 SD from the mean. This figure has been adapted from Kaiser *et al.* (2017).

The standard helicase DNA substrate constituted a 15nt polyT ssDNA 3'OH loading site and a 15nt dsDNA segment featuring a generic DNA sequence (15nt3'OH). After ATP addition, the RecQ4$^{427\text{-}1208}$ variant rapidly separated the dsDNA strands as indicated by the sharp increase in fluorescence intensity, reaching a maximum within the first 50 seconds of the experiment (Fig. 4-20). In contrast, the RecQ4$^{427\text{-}1116}$ variant separates the two DNA strands significantly slower and the fluorescence intensity consequently increased more steadily over the entire duration of the experiment. As a negative control, the ATPase deficient Walker-mutants RecQ4$^{427\text{-}1208}$ K508A (Walker A mutant) and RecQ4$^{427\text{-}1208}$ D605A (Walker B mutant) were tested and exhibited no detectable helicase activity, thus demonstrating that both RecQ4 variants are active helicases. The quantification of three individual helicase assays revealed that the helicase activity of RecQ4$^{427\text{-}1208}$ is approximately five-fold increased compared to RecQ4$^{427\text{-}1116}$, while the ATPase deficient Walker-mutants were consistently unable to separate the DNA double strands (Fig. 4-20 B).

In order to assess the processivity of the two RecQ4 variants, helicase activity was recorded for a set of Y-forked DNA substrates that exhibit different lengths of the dsDNA section (15nt, 18nt, 21nt or 24nt) while the unpaired 15nt polyT flanking sites were kept at a constant length of 15nt (Fig. 4-21). The helicase data for the different dsDNA lengths demonstrate that RecQ4$^{427\text{-}1208}$ is capable of efficiently separating a dsDNA segment of up to 18 base pairs, while the RecQ4$^{427\text{-}1116}$ variant was only able to displace the dsDNA of the 15nt Y-fork substrate. For a dsDNA segment length of 21nt, RecQ4$^{427\text{-}1208}$ exhibited a weak but detectable ATP-dependent helicase activity, while almost no DNA strand separation was detected for the 24nt dsDNA Y-fork substrate (Fig. 4-21 A). RecQ4$^{427\text{-}1116}$, on the other hand, features no helicase activity for either the 24nt or the 21nt dsDNA Y-fork. The helicase data for the 18nt dsDNA Y-fork substrate, however, exhibits a sudden but small increase of fluorescence intensity directly after ATP addition, which subsequently transitions into a steady plateau of about

20,000 RFU (Fig. 4-21 B). Since the initial burst of increasing fluorescence does not follow a smooth approximation to a maximum fluorescence intensity, as observed for the Y-15-15 DNA substrate, it can be assumed that this activity may originate from a small fraction of misaligned DNA substrate rather than from the actual helicase activity of the RecQ4$^{427\text{-}1116}$ variant on the correctly annealed Y-18-15 DNA substrate. The direct comparison between RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$ demonstrates that, as for the 15nt3'OH DNA, the C-terminally extended RecQ4 variant exhibits an increased helicase efficiency, indicated by the shorter amount of time that is required to reach a maximum fluorescence level (approx. five minutes) compared to RecQ4$^{427\text{-}1116}$ (approx. 15 minutes) as well as by their respective levels of maximum fluorescence intensity.

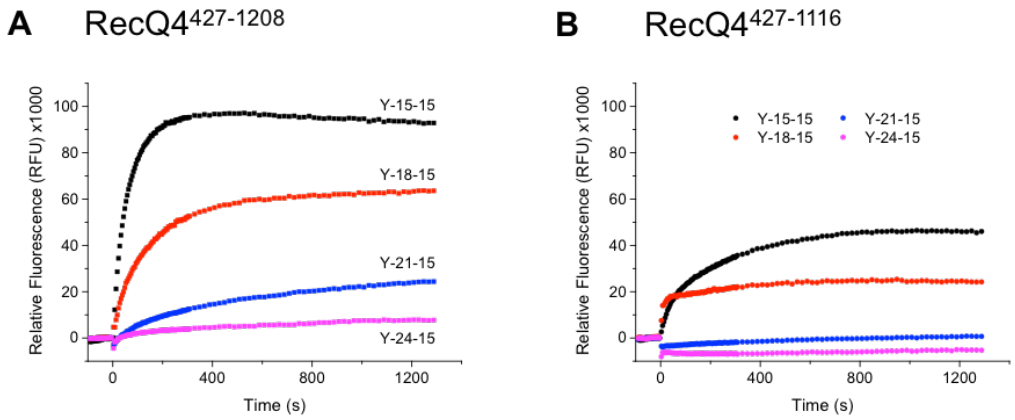


Figure 4-21: Helicase activities of RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$ on substrates with increasing dsDNA lengths. (A) RecQ4$^{427\text{-}1208}$ helicase assays for Y-forked DNA substrates with 15nt (black), 18nt (red), 21nt (blue) and 24nt (magenta) dsDNA lengths. **(B)** RecQ4$^{427\text{-}1116}$ helicase assays for the same substrates as in A. PolyT ssDNA flanking sites for the Y-forked substrates were kept at a constant length of 15nt.

Analyzing the dependency of the RecQ4$^{427-1208}$ helicase activity with respect to the protein concentration further provided functional insights regarding the helicase reaction. Foremost, the data for all tested protein concentrations illustrate a similar curve development, featuring a steep initial increase in fluorescence intensity after ATP addition, which gradually approximates a plateau phase and further steadily declines to moderately reduced fluorescence values (Fig. 4-22 A). The maximum fluorescence plateau, however, seems to be directly proportional to the amount of protein that is present in each reaction. Assuming that each RecQ4$^{427-1208}$ molecule would separate a bound DNA substrate by its ATP driven helicase activity and then immediately binds the next dsDNA substrate, thus constantly repeating the strand separation process, one would expect that the fluorescence intensity maxima of all protein concentrations should approximate each other over time, rather than reaching their individual plateau level. However, the data suggest that each steady-state phase after reaching the fluorescence plateau represents a mix of different reactions that take place at the same time. These reactions include the ATP-driven DNA strand separation of the original DNA substrate, the binding and ATP-driven translocation along the 15nt and 30nt ssDNA strands that were generated after DNA strand separation as well as a re-annealing of the top- and bottom DNA strands, the latter leading to a re-quenching of fluorescence. Furthermore, the raw data displays a difference in the slope of the increasing fluorescence intensity directly after ATP addition, suggesting a concentration dependent difference in the characteristic (and ideally constant) value for the helicase activity of RecQ4$^{427-1208}$ (Fig. 4-22 B). However, normalization of the raw data either based on maximum fluorescence or on protein concentration, reveals that the initial slope after ATP addition is comparable for all three different experimental setups, with only minor reductions, which are potentially based on the diminished DNA binding capabilities of RecQ4$^{427-1208}$ at 500 nM and 250 nM protein concentration.

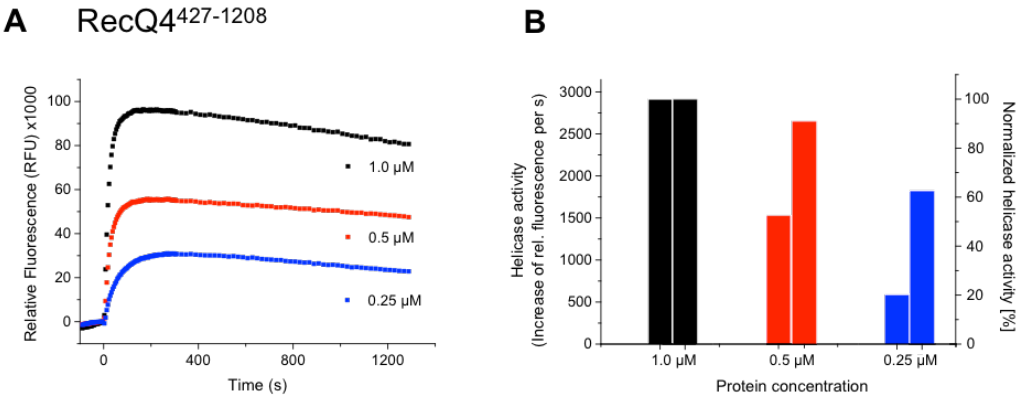


Figure 4-22: Helicase activities for different protein concentrations. (A) Helicase assays for 1.0 µM (black), 0.5 µM (red) and 0.25 µM of the RecQ4$^{427\text{-}1208}$ variant. **(B)** Quantification of helicase activity as depicted in A. Left bars of each protein concentration represent the linear increase of relative fluorescence 10 s after ATP addition (left axis). Right bars of each protein concentration represent the increase of relative fluorescence 10 s after ATP addition, normalized on the maximum fluorescence values (right axis).

The following set of helicase assays analyzed the influence of an extended ssDNA loading site and the performance of RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$ on a 30nt blunt end DNA substrate (Fig. 4-23). The fluorescence data for the RecQ4$^{427\text{-}1208}$ variant illustrates a similar helicase activity for both, the 15nt polyT 3'OH DNA and the 30nt polyT 3'OH substrate, including a characteristic increase of fluorescence upon ATP addition, a fluorescence plateau phase and a steady and moderate decrease to lower fluorescence values (Fig. 4-23 A). The fluorescence plateau for the 15nt3'OH substrate is not as high as for the 30nt3'OH DNA, however, the difference might rather originate from a decreased protein concentration based on a pipetting error, as comparable experiments generally illustrate a maximum fluorescence of approximately 100,000 RFU for the 15nt3'OH DNA instead of the 70,000 RFU that are depicted in this experiment (compare Fig. 4-20 A and Fig. 4-22 A). A difference between the two DNA substrates can be observed for the decrease of fluorescence during the steady state phase in the late stage of the reaction, which is more pronounced for the 30nt3'OH DNA substrate than for the 15nt-OH variant. A possible explanation for this effect might be that the 45nt ssDNA strand, which is generated after ATP-dependent strand separation, provides a larger substrate that enables the simultaneous binding of more RecQ4 molecules compared to the 30nt ssDNA strand that is

generated from the 15nt3'OH DNA. This increased fraction of bound RecQ4 for the 45nt ssDNA would consequently be prevented from binding and processing of the original dsDNA substrate, thus shifting the ratio of reactions during the steady state phase away from dsDNA strand separation and potentially increases the overall re-annealing probability. Similarly, the RecQ4$^{427\text{-}1116}$ variant displays a comparable helicase performance for the 15nt3'OH and the 30nt3'OH DNA substrates (Fig. 4-23 B).

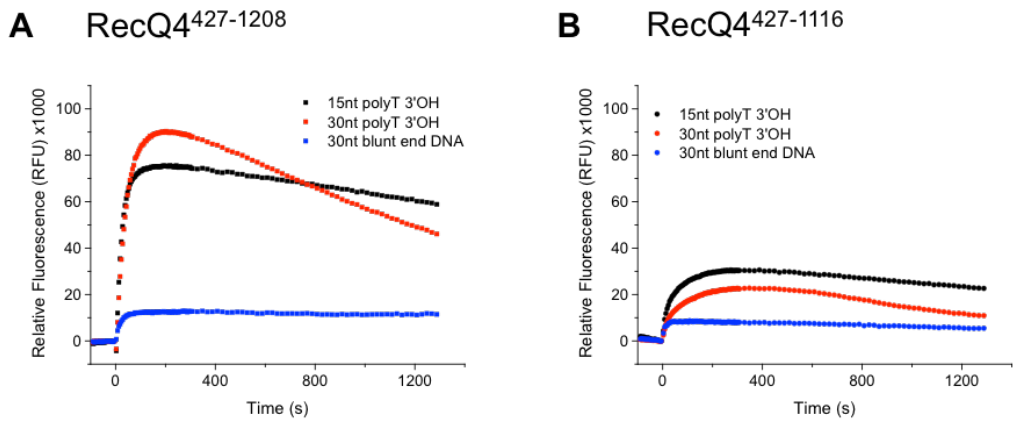


Figure 4-23: Influence of the length of the DNA substrate loading site on helicase activity and helicase performance on a blunt end DNA. **(A)** RecQ4$^{427\text{-}1208}$ helicase assays for a 15nt polyT 3'OH substrate (black), a 30nt polyT 3'OH substrate (red) and a 30nt dsDNA blunt end substrate (blue). **(B)** Helicase assays for RecQ4$^{427\text{-}1116}$ on the same set of DNA substrates as in A. The dsDNA segment of the two 3'OH substrates was kept constant at 15nt.

Although the maximum fluorescence plateau is slightly reduced for the latter DNA substrate, the tendency to an increased decline in fluorescence intensity can be observed as well, thus confirming the functional trends that are observed for the RecQ4$^{427\text{-}1208}$ variant. For both Recq4 variants, the fluorescence data demonstrate that neither protein is actively separating the 30nt blunt end DNA substrate as no substantial increase in fluorescence intensity can be observed after ATP addition. The minor but observable burst of fluorescence intensity for the blunt end substrate could originate from a small fraction of the DNA substrate that might have been annealed imperfectly during substrate preparation. In conclusion, increasing the length of the RecQ4 loading site had no significant effect on the performance of helicase activity by either RecQ4 variant.

The last set of helicase assays addressed the strand separation performance of RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$, initiated by varying concentrations of ATP (Fig. 4-24). As a general trend, the fluorescence data for both RecQ4 variants exhibited an ATP optimum, while a deviation towards lower or higher ATP concentrations led to an overall reduction of helicase activity. The recorded data for RecQ4$^{427\text{-}1208}$ illustrate two ATP-dependent features. Firstly, ATP concentrations around the optimum and below (e.g. 0.625 mM – 1.875 mM) exhibit the characteristic curve development as observed for previous experiments (Fig. 4-24 A). Hallmark elements of this characteristic behavior are the sudden increase in fluorescence intensity upon ATP addition and the moderate and steady decrease to lower fluorescence values during the steady state phase after the maximum fluorescence plateau has been reached. In contrast, ATP concentrations of 2.5 mM and 5 mM depict a slight and atypical increase of fluorescence intensity during the later stages of the experiments, while an ATP concentration of 7.5 mM exhibited no discernable helicase activity at all. The second noteworthy feature relates to the observed ATP independent strand separation activity of the RecQ4$^{427\text{-}1208}$ variant, which is illustrated by the H_2O negative control measurement (empty circles), in which H_2O instead of ATP was added. The steadily increasing fluorescence of the H_2O supplemented sample suggests that RecQ4$^{427\text{-}1208}$ is capable of separating the 15nt3'OH substrate in an ATP independent manner, a phenomenon that has been similarly observed for the fluorescence polarization experiments, during which the binding to a 15nt blunt end DNA substrate was investigated. Furthermore, the effect can be observed in the baseline recording section, 90 seconds before ATP injection is administered (Fig. 4-24 A). Supplementation of 7.5 mM ATP clearly counteracts this phenomenon, as no further increase in fluorescence intensity was recorded for this ATP concentration after ATP addition. In conclusion, the ATP dependent helicase data for RecQ4$^{427\text{-}1208}$ suggest that high ATP concentrations (2.5 mM ATP and more) reciprocally influence the interaction between the RecQ4 variant and the 15nt3'OH DNA substrate. The more ATP is applied, the less well the DNA substrate is bound, explaining

both, the decreased maximum fluorescence values for 2.5 mM and 5 mM ATP as well as the lack of ATP independent strand separation activity at 7.5 mM ATP. In contrast to RecQ4$^{427\text{-}1208}$, the C-terminally truncated RecQ4$^{427\text{-}1116}$ variant exhibits no such ATP independent dsDNA strand separation activity as illustrated by the steady fluorescence values of the H_2O negative control (empty squares) and during the 90 second pre-ATP injection phase (Fig. 4-24 B). However, as for RecQ4$^{427\text{-}1208}$, ATP concentrations of 2.5 mM and higher significantly decreased the maximum fluorescence intensity of the helicase reaction, thus suggesting that increased ATP concentrations equally suppress DNA substrate binding for the RecQ4$^{427\text{-}1116}$ variant.

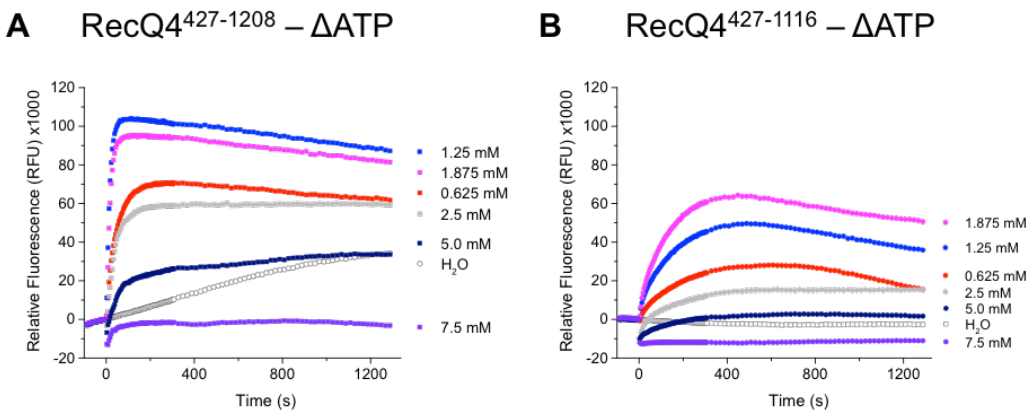


Figure 4-24: Helicase performance of RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$ at different ATP concentrations. (A) RecQ4$^{427\text{-}1208}$ helicase assays for a 15nt3'OH substrate at ATP concentrations of 0 mM (H_2O, empty circles), 0.625 mM (red), 1.25 mM (blue), 1.875 mM (magenta), 2.5 mM (gray), 5.0 mM (dark blue) and 7.5 mM (purple). **(B)** RecQ4$^{427\text{-}1116}$ helicase assays for a 15nt3'OH DNA substrate at the same ATP concentrations as in A. Fluorescence intensity for the zero ATP (H_2O) infection is depicted as empty squares. Color code for the respective ATP concentrations as in A.

4.3.1.4. ATPase activity

ATP hydrolysis rates for the RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ variants were determined via an *in vitro* NADH-coupled ATPase assay, in which the consumption of ATP is directly related to a decrease in the NADH concentration. As all RecQ helicases require a DNA cofactor for ATP hydrolysis, the ATPase rates for both RecQ4 variants were quantified for a variety of different ATP concentrations utilizing a saturating DNA concentration of 1 µM, with a 30nt polyT ssDNA substrate. The data for both RecQ variants follow typical Michaelis–Menten kinetics for a single-substrate biochemical reaction (Fig. 4-25 A). ATPase rates are practically zero without ATP, yet increase rapidly for the lowest tested ATP concentration and asymptotically approach a maximum rate (V_{max}) between 0.5 mM and 2.5 mM ATP. The data illustrate that there is essentially no difference between the two RecQ4 variants regarding their ATPase rates as both data fits overlap well and consequently exhibit similar kinetic parameters for V_{max} and K_M (Table 4-4). In addition to an ATP concentration range between 0 and 2.5 mM, hydrolysis rates were also determined for higher ATP concentrations, including up to 10 mM ATP (Fig. 4-25 B). The ATPase data for such increased ATP concentrations generally confirm the trend that was observed in the helicase assays, which were performed under similarly high ATP supplementation and show that the hydrolysis rates for both RecQ4 variants steadily decline with increasing ATP concentrations between 3.125 mM and 10 mM in a linear fashion. Similar to the helicase data, an explanation for this behavior could be that increased ATP concentrations gradually displace the ssDNA substrate from the RecQ4 protein, thus compromising the ability of RecQ4 to hydrolyze ATP.

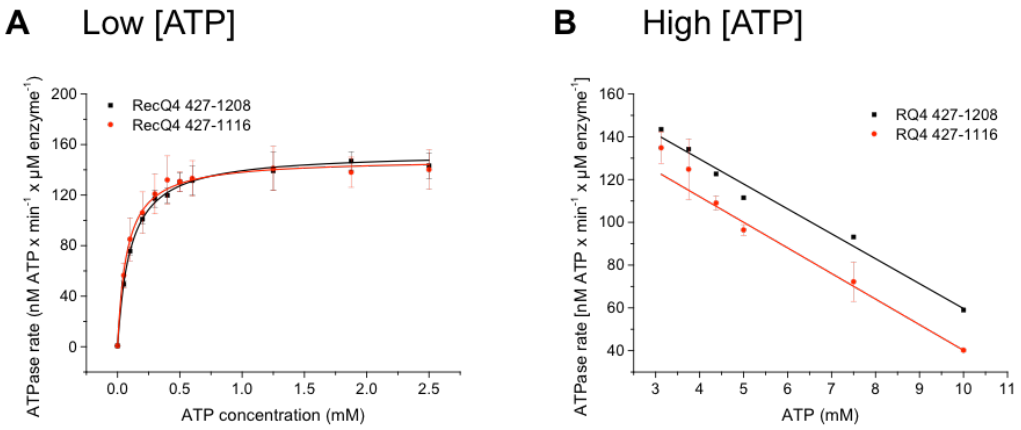


Figure 4-25: ATPase rates at low and high ATP concentrations. (A) ATPase rates for RecQ4$^{427\text{-}1208}$ (black) and RecQ4$^{427\text{-}1116}$ (red) at low ATP concentrations (0 – 2.5 mM). Data were collected for three independent experiments, using protein from two different purification batches and fitted by the Michaelis-Menten equation for single substrate enzyme kinetics. **(B)** ATP hydrolysis rates determined for RecQ4$^{427\text{-}1208}$ (black) and RecQ4$^{427\text{-}1116}$ (red) at high ATP concentrations (3.0 – 10.0 mM). Data were collected in two independent experiments and fitted by a linear fitting function.

Table 4-4: Kinetic fitting parameters for RecQ4$^{427\text{-}1208}$ and RecQ4$^{427\text{-}1116}$.

RecQ4 variant	427-1208	427-1116
V_{Max} [nM ATP x min^{-}1 x µM enzyme^{-1}]	153.7 ± 7.1	148.6 ± 24.1
K_M [nM]	103.9 ± 13.0	75.9 ± 65.1

4.3.2. *In vitro* characterization of RecQ4$^{427\text{-}1208}$ R4ZBD variants

The structural analysis of the novel RecQ4 R4ZBD revealed the presence of winged helix folds and a coordinated zinc ion, two features that are shared with the RecQ-conserved RQC domain. However, the analysis simultaneously highlights the substantial structural differences between the two domains. Nonetheless, these similarities argued for an analysis whether the R4ZBD could emulate the functions of the conserved RQC domain, e.g. facilitating dsDNA binding and DNA double strand separation. To explore this

possibility, several mutational variants of the RecQ4$^{427-1208}$ R4ZBD were generated. These variants included a deletion mutant, in which the entire upper WH domain was replaced by a double-glycine linker sequence. Additional R4ZBD variants encompassed single-, double- and triple amino acid substitutions within the lower R4ZBD WH domain. These mutations addressed positively charged and surface exposed residues that are potentially involved in DNA binding, and were substituted by alanine residues. (K1048A, R894A/R895A and K843A/R844A/R848A). With respect to the triple mutation in helix α12, an additional charge reversal mutation was created for the central arginine, e.g. K843A/R844E/R848A. Fig. 4-26 A provides an overview of the residues and protein folds that were manipulated and highlights their location within the R4ZBD. Subsequently, these RecQ4$^{427-1208}$ R4ZBD variants were tested for their DNA binding potential via fluorescence polarization (FP) and their ability to separate the 15nt3'OH DNA substrate by the fluorescence-based helicase assay (Fig. 4-26 B-D).

The analysis of the DNA binding behavior via FP determination (Fig. 4-26 B) illustrates the characteristic sigmoidal binding curve as observed for the previously tested RecQ4$^{427-1116}$ and RecQ4$^{427-1208}$ variants. The data demonstrate that neither the single amino acid substitutions nor the upper WH fold deletion had an influence on ssDNA binding as all fits for the individual FP experiments are overlapping well. The apparent dissociation constants for all R4ZBD variants are in a range of 1-3 nM, thus reflecting the ssDNA affinity of the wild-type RecQ4$^{427-1208}$ protein. In conclusion, the FP data suggest that the R4ZBD is not involved in the overall binding of a ssDNA substrate. The analysis of the helicase data for the R4ZBD variants further support this conclusion, as the fluorescence intensity curves of all R4ZBD variants follow the general characteristics of the RecQ4$^{427-1208}$ wild type (WT) protein (Fig. 4-26 C). Especially the lower WH domain single amino acid substitutions display the characteristic increase of fluorescence intensity directly after ATP addition, which is essentially indistinguishable from the WT RecQ4$^{427-1208}$ helicase (Fig. 4-26 D). Solely the upper WH domain deletion variant exhibits a reduced helicase activity after ATP

addition compared to the WT. However, as the final plateau of fluorescence intensity that is reached for the deletion variant is comparable to the fluorescence values of WT RecQ4$^{427-1208}$, it can be assumed that the upper WH domain (along with the lower WH domain) is unlikely to be directly involved in dsDNA strand separation. It is possible, however, that the upper WH domain aids in the guiding of the displaced DNA strand that is separated from its corresponding counter oligonucleotide. The lack of this domain therefore could either result in a hampered helicase velocity or favor a faster re-annealing of the two DNA strands, thus leading to the observed reduction of fluorescence intensity.

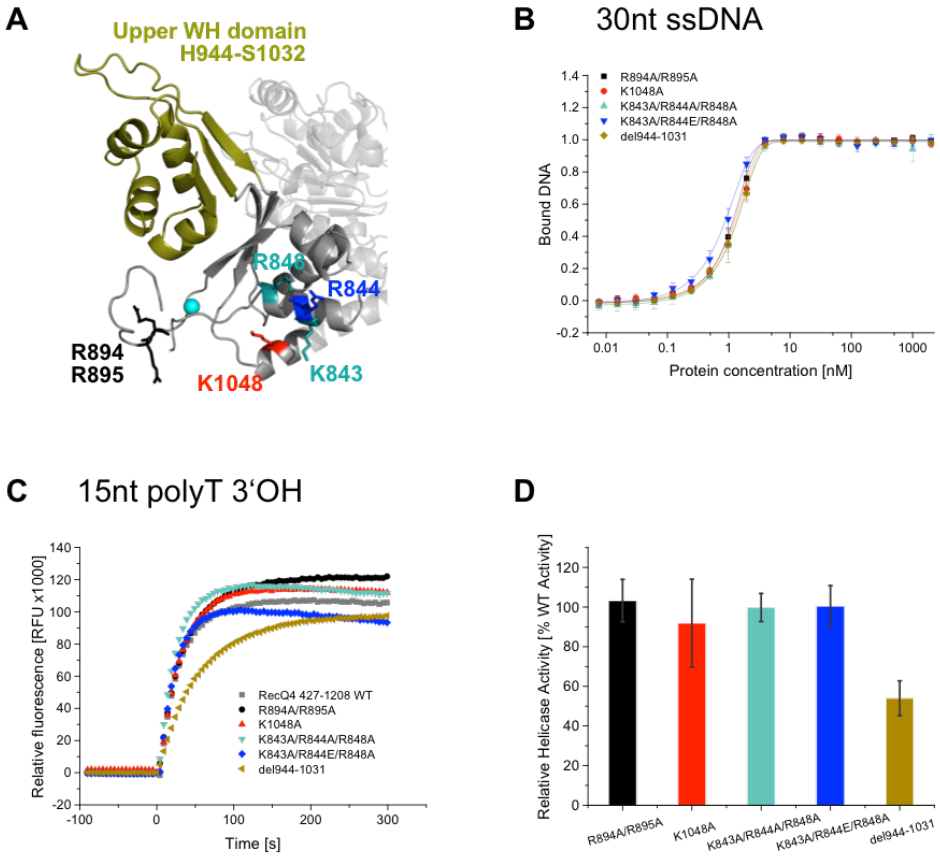


Figure 4-26: Analysis of RecQ4$^{427-1208}$ R4ZBD variants. (A) Overview of deletion- and mutational variants within the R4ZBD of RecQ4$^{427-1208}$. **(B)** Fluorescence polarization DNA binding data for a 30nt ssDNA substrate. **(C)** Fluorescence-based helicase assays for the RecQ4$^{427-1208}$ R4ZBD variants using a 15nt3'OH DNA substrate. **(D)** Quantification of the helicase activity for each R4ZBD variant based on three independent helicase assays. Bars represent the wild-type-normalized increase of relative fluorescence 10 s after ATP addition. All experiments were performed in triplicates, using protein from two independent purification batches. Error bars represent 1 SD from the mean. This Figure has been adapted from Kaiser *et al.* (2017).

4.3.3. *In vitro* characterization of RecQ4$^{427-1208}$ ARL variants

All structurally characterized RecQ helicases to date, that were co-crystallized with a DNA substrate, depict the binding of this substrate in a conserved fashion. In these models, the single stranded loading site of the DNA substrate is bound across the two RecA-like domains of the helicase core, while the dsDNA segment is coordinated by the RQC WH domain in a manner that positions the ß-hairpin of the WH domain in the vicinity of the ss/dsDNA junction. This arrangement requires the DNA substrate to enter the RecQ helicase from the HD2-site, leading towards HD1, thus implying a directed translocation along the ssDNA with the HD2 in the lead. As it was so far not possible to structurally verify if this DNA substrate trajectory is conserved in RecQ4, a functional analysis of the RecQ-conserved aromatic-rich loop (ARL) sequence was performed. The ARL in RecQ helicases constitutes a structural element that couples the binding of ssDNA to ATP hydrolysis. For the *E. coli* RecQ helicase, the ARL has been well studied, revealing characteristic functional features when specific single amino acids of the ARL were mutated. These properties provide a possibility to compare the effects of respective ARL amino acid substitutions within RecQ4, thus yielding functional insights regarding the DNA substrate binding properties of the RecQ4 helicase.

The ARL in RecQ4 exhibits four conserved amino acids: W613, H615, F617 and R618. All of these residues were individually substituted by a leucine and the functional characteristics of DNA binding, dsDNA strand separation and ATPase activity were subsequently determined via fluorescence polarization, the fluorescence-based helicase assay and the ATPase assay, respectively (Fig. 4-27). The DNA binding data of the FP experiments depict the characteristic sigmoidal curve for all ARL variants, exhibiting apparent dissociation constants that were indistinguishable from the RecQ4$^{427-1208}$ wild type (WT) protein (Fig. 4-27 A). Thus, each of the single amino acid ARL substitutions did not affect the overall DNA binding affinity. In contrast, determination of the ATPase rates for the individual ARL variants revealed

substantial functional differences (Fig. 4-27 B). In general, the data for each ARL substitution variant follow the typical Michaelis-Menten kinetics as described for the WT RecQ4$^{427\text{-}1208}$ and RecQ44$^{27\text{-}1116}$ proteins. However, the maximum ATP hydrolysis rates of each ARL substitution are significantly reduced compared to the RecQ4$^{427\text{-}1208}$ WT helicase. The variants W613L and H615L were least affected, both exhibiting approximately 60% of the WT ATPase rate, while ATP hydrolysis for F617L and R618L was more severely reduced to 35% and 21%, respectively, compared to the WT ATPase rate. As the concentration of the DNA cofactor influences the ability to hydrolyze ATP, it was further tested whether the ATPase rate of the ARL variants could be enhanced by increasing the amounts of the 30nt polyT ssDNA substrate. To this end, ATP hydrolysis data were collected at saturating ATP concentrations, while the amounts of the T30 DNA cofactor were varied (Fig. 4-27 C). The analysis of these results demonstrates that the original supplementation of 1 µM T30 ssDNA represents a saturating concentration, indicating that an additional increase of the ssDNA concentration would not lead to an elevation of the ATPase rates. Unexpectedly, the results demonstrate that the ARL variant F617L exhibits a small but steady ATPase activity even at the lowest ssDNA concentrations, e.g. at concentrations well below a 1:1 molar ratio of DNA and protein. In contrast, all other ARL variants including the WT helicase, displayed no detectable ATP hydrolysis at comparable ssDNA concentrations. The performance of the ARL variants in the helicase assays illustrates that all ARL substitutions except R618L were able to separate the 15nt3'OH substrate via their ATP-dependent helicase activity (Fig. 4-27 D). However, compared to the RecQ4$^{427\text{-}1208}$ WT protein, the ARL variants W613L, H615L and F617L display a reduction in helicase activity directly after ATP addition (Fig. 4-27 E). The ARL variants W613L and H615L were reduced to 58% and 51% of the WT helicase activity, while F617L still exhibited 85% of the RecQ4 WT dsDNA strand separation activity. This rather mild reduction for F617L was somewhat unexpected, considering the strong suppression of ATPase activity for this ARL substitution (35% of WT ATPase activity). In contrast to all other ARL

variants, R618L demonstrated only 2% of the WT helicase activity and can thus be considered as a helicase dead mutant.

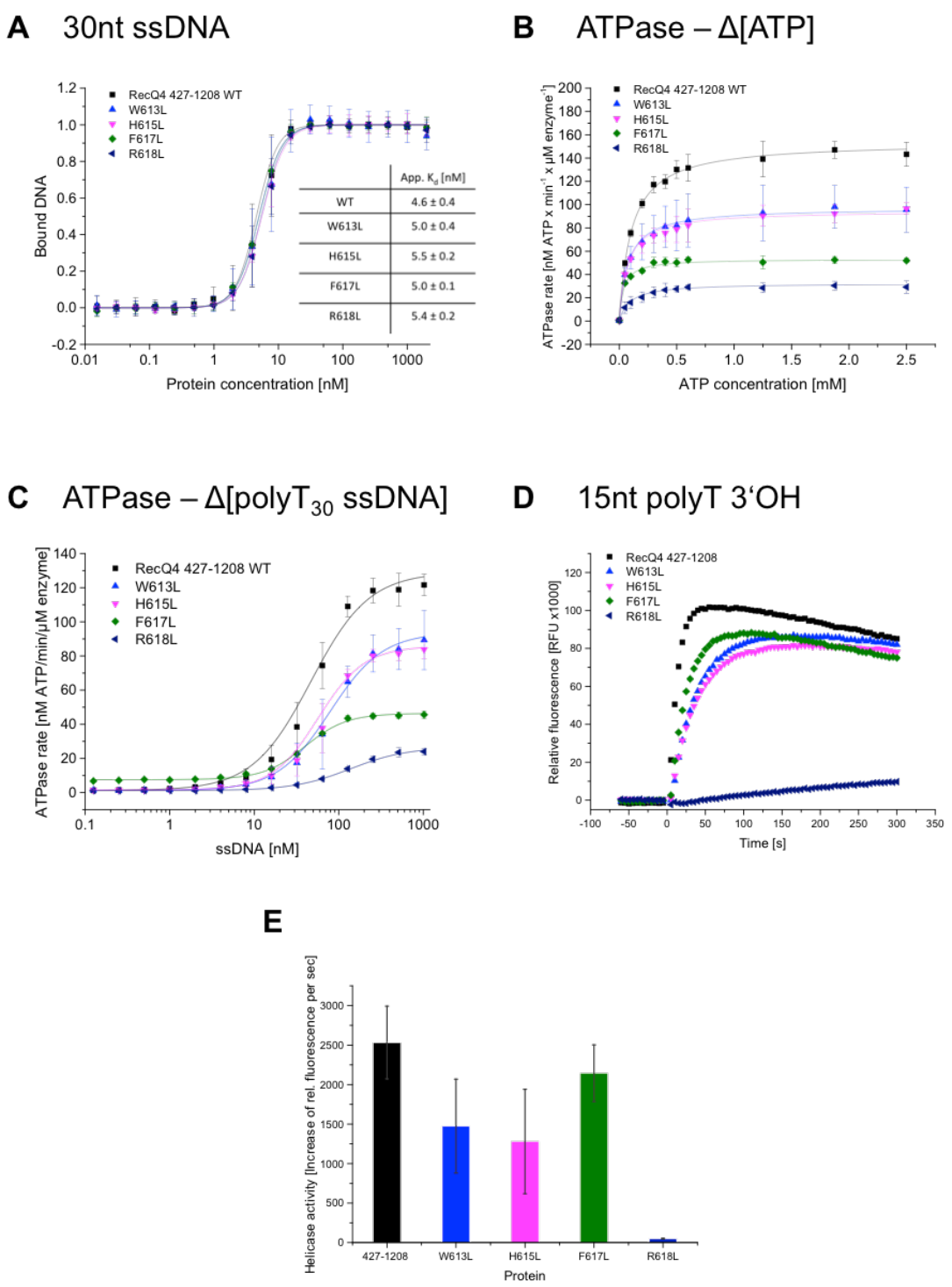


Figure 4-27: Analysis of RecQ4$^{427\text{-}1208}$ ARL variants. (A) Fluorescence polarization (FP) DNA binding data for the RecQ4$^{427\text{-}1208}$ ARL variants in comparison to the wild-type (WT) protein. Apparent dissociation constants (K_D) are listed in the table insert. **(B)** ATPase rates of the RecQ4$^{427\text{-}1208}$ ARL variants for an ATP concentration range of 0 - 2.5 mM. **(C)** ssDNA-dependent ATPase rates of the RecQ4$^{427\text{-}1208}$ ARL variants at saturating ATP concentrations (1.25 mM). **(D)** Helicase assays of the RecQ4$^{427\text{-}1208}$ ARL variants on a 15nt3'OH DNA substrate. **(E)** Quantification of the helicase activity for each ARL variant based on three independent helicase assays. Bars represent the increase of relative fluorescence 10 s after ATP addition. **(A-E)** All experiments were performed in triplicates, using protein from two independent purification batches. Error bars represent 1 SD from the mean. Color code: RecQ4$^{427\text{-}1208}$ WT – black. W613L – blue. H615L – pink. F617L – green. R618L – dark blue. This figure has been adapted from Kaiser *et al.* (2017).

This page has been left blank intentionally

5. Discussion

5.1. Revision of RecQ4 expression constructs

The RecQ4$^{427-1116}$ structure now provides the possibility to critically revise the initial RecQ4 variants and reevaluate their performance in expression and purification. In hindsight, it is evident that the early difficulties during protein purification were generally based either on unfortunate RecQ4 expression construct design, resulting in genuine protein destabilization and aggregation, or the association of the RecQ4 variant with DNA in case of a properly folded protein. Both events led to the common phenomenon of void volume elution during SEC analysis and exhibited an unfavorable 280/256 nm absorption ratio, which was commonly interpreted as protein aggregation. Regarding protein stability, the initial RecQ4$^{460-1208}$ variant only marginally deviates from the far more stable RecQ4$^{427-1208}$ construct. While the former tended to exhibit precipitation during protein concentration, the latter did not. Similarly, other RecQ4 constructs that started at aa 460 featured a comparably unstable behavior despite the fact that their protein boundaries were generally in accordance with the RecQ4 domain architecture and should thus be properly folded. Such constructs include for example the RecQ4$^{460-895}$ variant, which encompasses the helicase core, the first bridging helix and the first helix of the R4ZBD. The N-terminal Sumo/smt3-solubility-tag addition for RecQ4$^{460-1208}$ led to an enhanced protein expression, which further supports the notion that amino acids upstream of the helicase core generally stabilized the RecQ4 variants. In agreement with this observation, the RecQ4$^{427-1116}$ structure contains residues right in front of the helicase core, e.g. aa 449 – 465, which do not form any secondary structure elements such as helices or b-strands, yet adopt a discrete loop orientation in close proximity to HD1 that may stabilize its fold. An additional trend towards protein stability can be deduced from the parallel purification of the RecQ4$^{460-1047}$ and RecQ4$^{460-1115}$ variants (Fig. 4-3), which were originally created to analyze the influence of the C-terminus on the observed

aggregation behavior. While the slightly shorter RecQ4$^{460-1047}$ construct encompasses the helicase core as well as the entire R4ZBD, it lacks the second bridging helix (α19) and helix α20, which are both present in RecQ4$^{460-1115}$. The parallel purification of both RecQ4 variants illustrated that for RecQ4$^{460-1115}$ much higher absorption signals in the SEC chromatograms were observed compared the shorter RecQ4 variant in all three imidazole elution steps, however, especially the 100 mM imidazole elution analysis depicted far less protein in the SDS-gel than RecQ4$^{460-1047}$. This suggests that RecQ4$^{460-1115}$ bound much more DNA, whereas RecQ4$^{460-1047}$, being less stable due to the missing helices, cannot bind to DNA that efficiently and thus might primarily represent aggregated protein. Moreover, the RecQ4 structure illustrates that both, bridging helix α19 and helix α20, interact tightly with HD2 of the helicase core. Thus, they are likely required for the proper orientation of the R4ZBD within the RecQ4 structure and possibly support the stability of HD2. The only clear-cut examples for unfavorable construct design were the RecQ4 versions that ended with aa 997, e.g. RecQ4$^{450-997}$ and RecQ4$^{427-997}$. Although those variants contained the stabilizing N-terminal extensions, both were lacking the very C-terminal two β-strands of the R4ZBD, which are components of the upper and lower R4ZBD WH domain β-sheets. Thus, both RecQ4 variants likely exhibited an unfolded R4ZBD, resulting in the observed aggregation behavior during protein purification. Collectively, the sum of those observations emphasizes the two structural necessities for a stable RecQ4 construct design: A stabilizing N-terminal extension prior to the helicase core and the presence of the bridging helix α19 and helix α20, which seem to support the stability of HD2 and/or warrant the overall structural integrity.

5.2. RecQ4 structure and function

5.2.1. Structure predictions by Phyre2 and I-TASSER

A revision of the predicted RecQ4 secondary structure elements by the programs Phyre2 and I-TASSER reveals that both algorithms projected all secondary structure elements of the RecQ4$^{427-1116}$ model equally well and with very high accuracy. Boundaries for α-helices and β-strands mainly deviate by only one or two amino acids and were occasionally even perfectly predicted. One of the few differences between the Phyre2 and I-TASSER predictions falls into a region that is flexible and is thus not resolved in the model (aa 858-882). Therefore, it cannot be concluded which program predicted the correct secondary structure elements, as such areas might only assume a stable fold in the presence of a binding partner or a specific substrate. Compared to the twenty α-helices and eighteen β-strands that were correctly predicted and depicted in the RecQ4$^{427-1116}$ structure, only one case can be stated where both structure prediction programs incorrectly anticipated an α-helix that could not be confirmed (HD1, aa 587-594/595). In general, it is reasonable to presume that secondary structure predictions are accurate for protein domains that are highly conserved, such as the RecQ4 helicase core, as structural information of one or more homologue is likely available and the prediction algorithms can implement such information for the generation of a prognosis. However, the amino acid sequence of the R4ZBD exhibits no discernable homology to proteins other than RecQ4 paralogues from different species, none of which has been structurally described at the atomic level. The fact that the secondary structure elements of such unique RecQ4 segments were nonetheless predicted with very high accuracy, highlights the strong potential of these prediction algorithms regarding preliminary structure analysis and protein construct design.

5.2.2. The RecQ4 structure

The RecQ4 model presented here has been described in the publication by Kaiser *et al.* (2017) and represents the first high-resolution structural information of the RecQ4 helicase core as well as a large fraction of its unique C-terminus. The two conserved helicase domains adopt the expected RecA-like folds as observed in the structures of previously characterized RecQ helicases and facilitate ssDNA binding, ATP hydrolysis and the coupling of both of these functions to enable the directed 3'-5' translocation along a ssDNA substrate. Downstream of the helicase core, the RecQ4 structure exhibits the unique R4ZBD, which is composed of two interconnected WH-like folds, while a coordinated loop region within the lower WH domain features a complexed zinc ion. The new structural information is thus in contrast to previous reports that anticipated a RQC-like domain in place of the R4ZBD within the C-terminal region of RecQ4. Although the R4ZBD and the RQC domain share the characteristics of a WH-like fold and the coordination of a zinc ion, both domains adopt individual structures and assume distinct positions with respect to the conserved helicase core. The RQC domain is typically located adjacent to HD2, and thereby positions the RQC-WH domain in a prime orientation to perform its ascribed functions of dsDNA binding and dsDNA strand separation via the WH domain ß-hairpin element. In contrast, the R4ZBD of RecQ4 assumes an opposing location in close vicinity to HD1. Moreover, the R4ZBD does not feature a discernable dsDNA strand separation element, analogous to the ß-hairpin. Based on these structural features (or the lack thereof), it is unlikely that the R4ZBD assumes the functional role of the RQC domain in RecQ4. The structural analysis of the R4ZBD WH domains revealed that the upper and lower WH folds each represent the characteristics of a different WH domain type. The upper WH-domain (Fig. 4-10) was assigned to the methionine aminopeptidase-like WH class that features structural homology to a range of functionally different proteins[244]. Coincidentally, the closest structural homologue of the upper WH domain was indeed the human

methionine aminopeptidase 2. The functional diversity regarding the structural homologues suggests that this particular WH domain and its putative HTH motif are not necessarily involved in the binding of DNA but could instead function as a protein-protein interaction module. However, a possible role of the R4ZBD upper WH domain in the binding of a non-canonical RecQ4-specific DNA substrate should not be excluded as the list of structural homologues included proteins that are involved in nucleic acid binding. Moreover, the putative HTH motif of the upper R4ZBD WH domain features a set of basic residues that could facilitate electrostatic interactions to the negatively charged DNA phosphodiester backbone as well as some aromatic residues, which could be involved in interactions with DNA bases. The lower R4ZBD WH fold exhibits the structural characteristics of a classical three-stranded WH domain[244] and features a pronounced homology to bacterial proteins that are involved in transcription regulation. This is, in principle, a strong indication that the lower R4ZBD WH domain could be equally involved in the binding of a DNA substrate. However, especially the second helix of its putative HTH motif, which characteristically facilitates the binding to the minor or major groove of the DNA helix, features a set of acidic residues that partially mediate electrostatic contacts to the adjacent HD1, thus suggesting that this particular element of the WH domain is not involved in DNA binding. Nonetheless, it has to be considered that the typical WH DNA-binding mode, in which the second helix of the HTH motif is embedded within the minor or major groove of the DNA, is characteristic especially for DNA-sequence specific proteins such as transcription factors or repressors. RecQ helicases, however, act predominantly independent of DNA sequence. In this respect, analyses of DNA bound RecQ structures that feature the RQC domain confirmed that the second α-helix of the HTH-motif in the RQC-WH domain is located outside of the DNA grooves and that DNA sequence independent contacts with the negatively charged DNA backbone are mediated by basic residues that are located in loop regions prior and after that α-helix[245]. In summary, similar to the upper R4ZBD WH domain, structural information by itself only provides a first approximation towards the

true function of the lower WH fold, which is yet to be further analyzed by functional studies. Feasible possibilities range from a simple scaffold function to the role of a binding module either for other proteins or non-canonical DNA structures, which moreover could act either independently or in combination with the upper WH domain. The analysis of the zinc-binding site of the lower WH domain shows that the metal ion is coordinated by three cysteine residues and a histidine that are mainly located within an extended loop of approximately fifty amino acids, connecting α-helix α12 and β-strand β13 (Fig. 4-9). A large portion of this loop is flexible and is thus not resolved in the RecQ4 structure, however, it is possible that this part of the R4ZBD would adopt a defined fold in the presence of a DNA substrate or a protein binding partner. In line with this assumption, the secondary structure prediction by Phyre2 and I-TASSER suggested defined structure elements for this region. As the binding site of the metal ion is located mostly outside of any secondary structure element, it can be specified as structurally detached from the lower R4ZBD WH domain. It is reasonable to assume that the prime function of the bound zinc ion is the stabilization of the lower R4ZBD WH structure segment during protein translation. The arrangement of the R4ZBD structure leaves the major part of the lower WH domain incomplete during translation, requiring the upper WH domain to be folded and the subsequent translation of the β-strand β18, which complements the lower WH domain β-sheet. During this time, the presence of the bound zinc ion could stabilize the transition state and thereby prevent protein aggregation and/or degradation.

The R4ZBD is structurally connected to the helicase core via a set of two bridging helices. Bridging helix α11 represents a linker between HD2 and the N-terminus of the R4ZBD. In contrast, the second bridging helix α19 interacts tightly with structural elements of both, the lower R4ZBD WH domain and the HD2 and thus presumably represents an important structural segment that facilitates the correct ternary arrangement between the helicase core and the R4ZBD. Moreover, the two bridging helices and helix α20 mirror a structural composition as observed for the RQC Zn domain. The fact that a RecQ4 variant that is missing helix α20 ($RecQ4^{427-1090}$) could not be expressed in a

soluble manner suggests that this structural arrangement is highly important for protein stability. Similarly, a reduction in soluble protein expression was reported for other RecQ helicases that featured a compromised RQC Zn domain[246-248], supporting the notion that this domain and its counterpart in RecQ4 primarily facilitate structural integrity as a scaffold. The last 97 amino acids of the RecQ4 C-terminus (termed the CTD) are not part of the here presented RecQ4 model and remain to be structurally characterized. Both secondary structure prediction programs consistently anticipate an arrangement of five α-helices that are connected to helix α20 by a linker segment of about 25 amino acids. Assuming that the CTD forms an independent domain, it is tempting to speculate that this domain could adopt a similar orientation as the RQC WH fold observed in the other RecQ structures, given that this structure follows the RecQ4-counterpart of the RQC Zn domain. However, the typical RQC WH fold features at least two β-strands (Fig. 4-12), secondary structure elements that were not predicted by Phyre2 or I-TASSER for the RecQ4 CTD. Nonetheless, taking the functional importance of those C-terminal amino acids into account, it is reasonable to assume that the RecQ4 CTD would perform a RQC-WH like function, although it does certainly exhibit a different structural fold.

5.2.3. RecQ4 functions and the helicase mechanism

The functional analysis presented here focused on the activity of RecQ4 on standardized canonical DNA substrates, e.g. forked DNA and dsDNA overhang substrates, which can be separated through the ATP-dependent helicase/translocation activity of the RecQ4 protein. This is an important distinction from the more complex, e.g. non-canonical, DNA substrates such as D-loops or Holiday-Junctions, which could be bound and processed by RecQ4 as well, but were not analyzed in this study. These latter substrates may require structural or functional support from other RecQ4 domains and could thus rely on a more extended and elaborate mechanism to be unwound. The functional studies encompass the basic activities of the

RecQ4 helicase, e.g. binding of different DNA substrates and the activities of ATP hydrolysis and DNA double-strand separation. These functions were initially analyzed for the two wild-type (WT) RecQ4 variants, RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$, and were later utilized to investigate the functional importance of specific structural RecQ4 elements such as the R4ZBD and the ARL.

5.2.3.1. DNA binding

Initial DNA binding studies were pursued via EMSA experiments, testing a range of basic DNA substrates, including a Y-forked DNA, a dsDNA with a 3' or a 5' overhang (OH) as well as a blunt end DNA substrate. The fully C-terminal RecQ4$^{427\text{-}1208}$ exhibited strong substrate affinity for the Y-fork as well as both OH substrates, whereas only the Y-fork was comparably well bound by the C-terminally truncated RecQ4$^{427\text{-}1116}$. These results suggest two general DNA binding features for the RecQ4 helicase (Fig. 5-1). First, both single-stranded segments of the Y-fork DNA substrate seem to contribute to the overall binding affinity of the RecQ4$^{427\text{-}1116}$ protein (ssDNA binding via A and B) as both OH substrates, which are each missing one of the OH tails compared to the Y-fork DNA, were considerably less well bound by RecQ4$^{427\text{-}1116}$. While the 3'OH segment is likely to be bound by the conserved helicase core (A), as observed for other DNA-bound RecQ structures, the exact location of the second 5'OH binding site is yet to be determined and verified (B). Secondly, the C-terminal 92 amino acids (CTD) represent an additional DNA binding module (C). In case of the 3'- and 5'-OH substrates, the increased binding affinity by the CTD of RecQ4$^{427\text{-}1208}$ compensates for the reduced affinity that originates from the lack of one OH segment as compared to the Y-fork DNA. Since RecQ4$^{427\text{-}1116}$ does not feature the CTD, this variant cannot compensate for the missing binding affinity as observed for the RecQ4$^{427\text{-}1208}$ variant, resulting in the dramatically reduced affinity for the 3'- and 5'OH substrates.

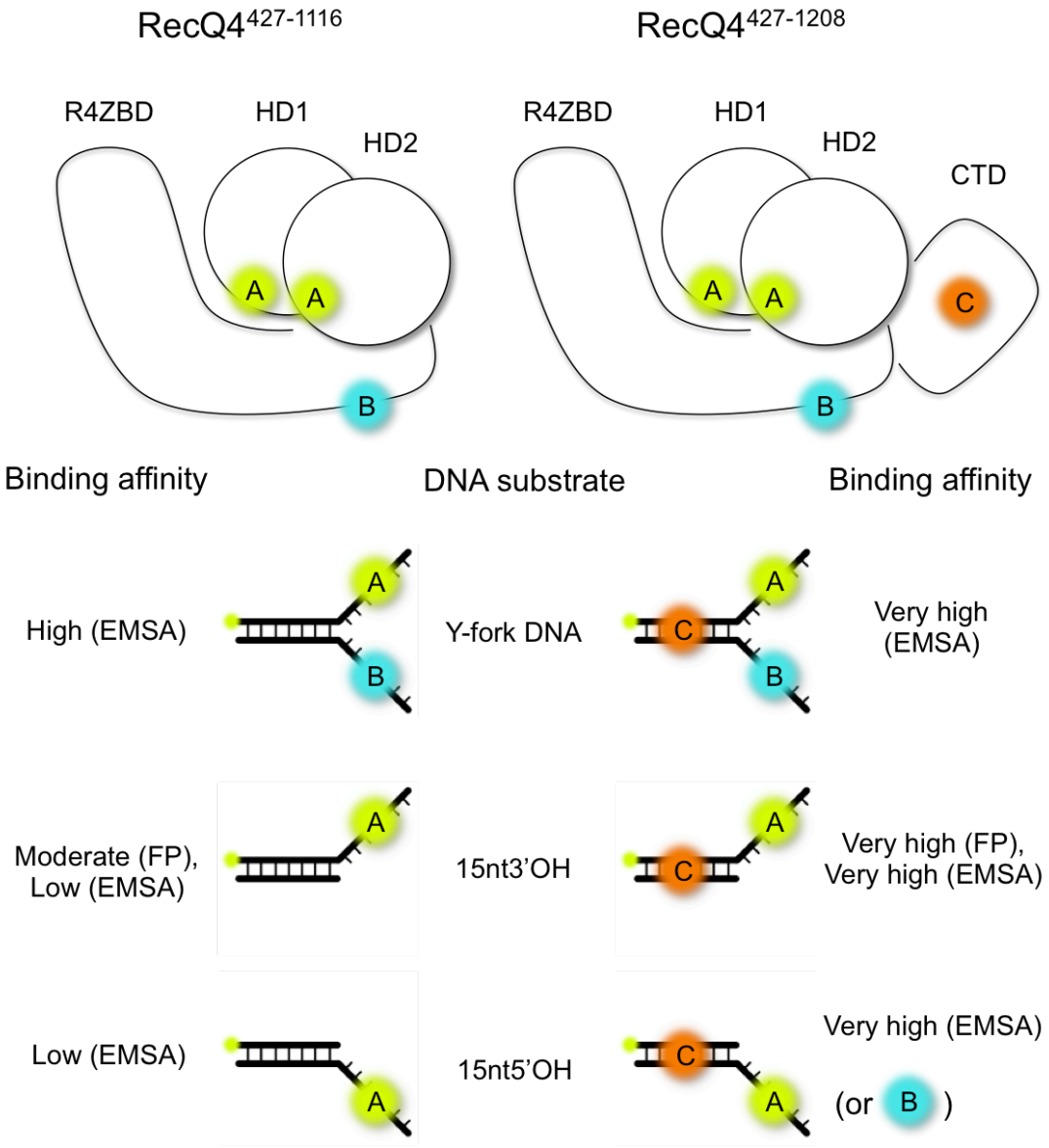


Figure 5-1: Possible DNA interaction sites within RecQ4. Based on the analysis of DNA binding data by EMSA and FP experiments for the RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ variants, three different DNA binding regions in RecQ4 helicase core and the C-terminus can be postulated. **Top:** Graphical illustrations of the two RecQ4 variants, highlighting the locations where DNA might be bound. The two helicase domains, HD1 and HD2, represent an established an evolutionary conserved ssDNA binding location (A, green). A second DNA binding region could be envisioned close to the RecQ4 bridging helices (B, cyan). Based on the DNA binding data, this region likely binds to ssDNA. The CTD constitutes the third DNA binding element (C, orange), presumably binding to dsDNA or strenghening the binding of a ss/dsDNA junction-containing substrate. **Bottom:** Illustration of DNA substrates used for EMSa and/or FP experiments, highlighting the DNA segments that are presumably bound by the two, respectively three, DNA binding regions of the two RecQ4 variants.

Moreover, the direct comparison of Y-fork DNA binding by the two RecQ4 variants illustrates an increased affinity of the C-terminally extended RecQ4$^{427\text{-}1208}$ variant, thus supporting the hypothesis that the CTD represents an additional DNA binding module. The fact that either RecQ4 variant does not bind a blunt end DNA substrate suggests that the major determinants for DNA binding are the single stranded DNA segments. Moreover, the data imply that the CTD DNA binding module rather strengthens the affinity of one or both ssDNA binding sites instead of increasing the affinity by directly binding to the dsDNA segment itself. The differences in the binding affinities for the RecQ4$^{427\text{-}1208}$ variant to either the 3'OH- or the 5'OH DNA suggests that such DNA substrates are generally bound in a preferred orientation and that this trajectory has an influence on the binding affinity. It can be assumed that the physiological substrate, which is the 3'OH DNA, is bound in a fixed position and the presence of the dsDNA segment exerts an increase of the binding affinity as it is locked and ready to be separated upon ATP addition. In contrast, the 5'OH DNA substrate can only be bound via the 5' ssDNA loading side and the dsDNA segment of this substrate cannot influence the DNA binding affinity as it is not oriented in the physiologically relevant position, e.g. in front of the protein rather than behind it.

In order to directly measure DNA affinities in solution, fluorescence polarization (FP) measurements were performed under equilibrium binding conditions. The FP data confirmed the general trend of an increased binding affinity for the RecQ4$^{427\text{-}1208}$ variant, which features the additional CTD DNA binding module. The results yielded apparent dissociation constants for both RecQ4 variants in the low nM range for a 30nt ssDNA substrate. Interestingly, RecQ4$^{427\text{-}1208}$ bound a 15nt3'OH substrate with comparable affinity, while binding of RecQ4$^{427\text{-}1116}$ to the 15nt3'OH DNA was slightly reduced compared to the ssDNA substrate. This observation supports the conclusions drawn from the EMSA studies, e.g. that RecQ4$^{427\text{-}1116}$ features two independent binding sites for the two Y-fork OH tails. A single 30nt ssDNA substrate essentially represents two 15nt OH segments that are connected to each other, thus enabling the binding of each half of the 30nt

ssDNA to its own binding site. In comparison, the 15nt3'OH substrate features only one 15nt ssDNA OH, leading to a reduced binding for the RecQ4$^{427-1116}$ variant, while the CTD of RecQ4$^{427-1208}$ can compensate for the lack of a second OH tail by an increased affinity through the CTD. However, the observed difference is rather small and the FP binding experiments for the 15nt3'OH substrate have only been performed once, thus a more thorough analysis of this effect is required to obtain an unambiguous conclusion. Compared to the FP data, the EMSA results imply a rather dramatic difference between the binding affinities for the ssDNA substrate and the 15nt3'OH DNA by the RecQ4$^{427-1116}$ variant, a trend that is not reflected by the FP data. It is possible that the sample treatment during the EMSA studies, e.g. application of an electric field that induces sample heating or the influence of the gel matrix, have had an effect on the obtained results, whereas FP experiments yielded quantitative data that reflect a more physiological environment. An analysis of the binding behavior of both RecQ4 variants to a 15nt blunt end DNA substrate yielded rather unexpected results. In analogy to the EMSA data, binding of the 15nt blunt end DNA substrate by either RecQ4 variant was not expected. However, the FP data depict a change in the polarization values with increasing protein concentrations, indicating that the fluorescently labeled DNA substrate was indeed bound by both RecQ4 variants. Moreover, a second determination of the polarization values, five minutes after the original data was recorded, depicted an additional change compared to the first measurements, indicating that the systems were not at equilibrium conditions when the first data were collected.

Subsequent FP experiments revealed that a comparable binding behavior of both RecQ4 variants can be observed for a 15nt ssDNA substrate. These results suggest, that both RecQ4 variants are able to interact with a blunt end dsDNA substrate in solution, resulting in a slow and ATP-independent separation of the dsDNA strands, which eventually led to the observation of the binding behavior that effectively reflects that of a 15nt ssDNA strand rather than the binding of the entire 15nt dsDNA. The comparison between

the two RecQ4 variants illustrates that this phenomenon is more pronounced for the C-terminally extended RecQ4$^{427\text{-}1208}$ than for RecQ4$^{427\text{-}1116}$ variant, which indicates that the increased binding affinity of the CTD generally supports the mechanism of this ATP-independent dsDNA strand separation. A possible explanation for this phenomenon may reside in the shortness of the dsDNA substrate. The melting temperature of the 15nt generic sequence dsDNA substrate was calculated to be approximately 40°C (OligoAnalyzer Tool, Integrated DNA Technologies). Based on the geometry of the standard B-form DNA in solution, two complementary 15nt DNA strands result in the stable formation of a DNA substrate with 1.5 helical turns. However, it is possible that the terminal bases of the DNA substrate, especially the two 3' T-A pairs, experience so-called DNA breathing due to thermal fluctuations, which led to a short temporary unpairing of the terminal base-pairs. It is likely that the presence of the RecQ4 variants further aggravate the re-pairing of the bases as they compete with the counter-oligonucleotides for the ssDNA fragments. This would also explain the enhanced ATP-independent DNA strand separation of RecQ4$^{427\text{-}1208}$, as this RecQ4 variant exhibits a higher affinity for ssDNA compared to RecQ4$^{427\text{-}1116}$. It is more challenging, however, to explain the reduction of FP values for the 15nt ssDNA substrate at the highest protein concentrations of both RecQ4 variants. Generally, decreased FP values indicate less bound DNA substrate, however, it is implausible that RecQ4 should exhibit reduced DNA binding affinity particularly at high protein concentrations. It is more likely that at high RecQ4 concentrations multiple molecules are bound to the same ssDNA substrate, which in case of the rather short 15nt ssDNA substrate may result in a shielding effect of the fluorescent dye, thus decreasing the intensities for parallel and perpendicular fluorescence. Based on the binding of additional RecQ4 molecules, the molecular mobility of the fluorescently labeled DNA substrate further decreases, technically resulting in increased FP values. Due to the reduction in fluorescence intensity, however, FP values effectively decrease, leading to the observed binding curves. Although this explanation seems to be rather far fetched, a similar behavior can be envisioned for the

FP experiments that depict the binding of a 30nt ssDNA substrate, as some results indicate a trend to declining FP values at the highest protein concentrations (compare Fig. 4-17). Based on these results, FP experiments to determine the binding of RecQ4 to blunt end dsDNA should ideally be repeated using longer dsDNA substrates to exclude a thermal-driven destabilization and investigate whether the ATP-independent dsDNA strand separation by RecQ4 is an experimental artifact or has functional relevance.
DNA binding was also assessed in the presence of several additives. The results of the EMSA studies suggested that the overall binding affinity for the Y-fork DNA substrate to RecQ4$^{427\text{-}1208}$ was only marginally reduced in the presence of 150 mM KCl, 30% (v/v) glycerol or 1mM ADP. The influence of 1 mM AMP-PNP and 1 mM ADP on DNA binding was further analyzed using the FP setup, which didn't support the EMSA results but suggested that none of these additives had any significant effect on the overall DNA binding.

5.2.3.2. Helicase activity and the helicase mechanism

Measuring dsDNA strand separation activities utilizing the fluorescence based helicase assay permitted a closer investigation of the RecQ4 helicase mechanism. A comparative analysis of RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ illustrates that both RecQ4 variants are actively separating the 15nt3'OH dsDNA substrate in an ATP-dependent manner. The data presented in Figure 4-20 depicts the true ATP-dependent helicase activity of both RecQ4 variants, as fluorescence values representing the thermal destabilization of the DNA substrate (e.g. control without protein) as well as ATP-independent helicase activity values (e.g. control with H_2O-supplementation instead of ATP) have been subtracted from the respective experimental values for RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$. The data reveal that the crystallized RecQ4$^{427\text{-}1116}$ variant is an active helicase, which should therefore exhibit all minimal structural elements that are required for the helicase mechanism. However, the CTD-containing RecQ4$^{427\text{-}1208}$ variant is about 5x more active than RecQ4$^{427\text{-}1116}$, suggesting that the increased DNA binding affinity by the

CTD has an influence on the helicase activity. As expected, no helicase activity was observed for the ATPase deficient RecQ4$^{427\text{-}1208}$ K508A and D605A Walker-mutants, confirming that both, RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$, are active ATP-dependent helicases. Two scenarios seem plausible to explain a possible helicase mechanism for RecQ4. In the first scenario, dsDNA strand separation is achieved by a wedge-like structural element that is present within the CTD of RecQ4 and thus assumes an analogous function as the ß-hairpin observed in the other human RecQ helicases. Such a wedge-like element would bind to the ss/dsDNA junction of a DNA substrate, thus explaining the increased DNA binding affinity for the RecQ4$^{427\text{-}1208}$ variant. Secondary structure predictions for the RecQ4 CTD suggest a structural composition of 4-5 α-helices, thus making it unlikely that the CTD would feature a true β-hairpin element. However, sequence analysis of the C-terminal domains from various RecQ4 paralogues revealed the presence of two highly conserved patches that contain several basic and aromatic amino acids (Fig. 5-2), which could potentially facilitate interactions to the negatively charged DNA backbone or the DNA bases, respectively. In line with this hypothesis, a recent structural and functional analysis of the human RecQ5 helicase demonstrated the functional importance of a single α-helix, which directly follows the RecQ5 Zn domain[62]. Experimental data suggest that this α-helix might adopt a wedge-like function comparable to the ß-hairpin element of the RQC domain in RecQ1, BLM and WRN. However, the aforementioned conserved patches of the RecQ4 CTD do not include the first CTD α-helix and are thus not congruent with the functional model suggested for the RecQ5 helicase. Moreover, the major drawback of this mechanism is that it does not explain the observed helicase activity for the RecQ4$^{427\text{-}1116}$ variant, which therefore would need to be ascribed to represent an experimental artifact that could potentially be based on the insufficient length of the dsDNA substrate.

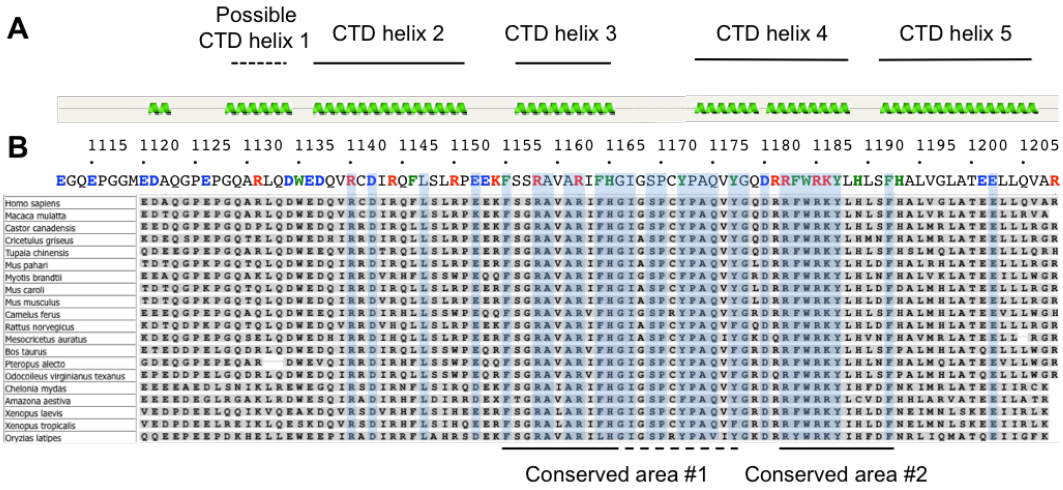


Figure 5-2: Sequence analysis of RecQ4 helicases from various species. (A) Secondary structure prediction by the Phyre2 algorithm for the RecQ4 CTD (aa 1111-1208). α-helices are depicted as green helices. **(B)** Sequence alignment of the RecQ4 CTD from twenty different species. Acidic (blue), basic (red) and aromatic residues (green) are highlighted in the human amino acid sequence. Sequence identity or homology of 100% is illustrated by blue boxes within the alignment. Conserved areas that feature positively charged and aromatic residues are indicated below the alignment.

The second scenario for a potential helicase mechanism of RecQ4 does not rely on the presence of a dsDNA strand separation wedge. Similar to bacterial RecQ helicases, for which a bend-angle driven dsDNA strand separation mechanism has been proposed, the CTD of RecQ4$^{427\text{-}1208}$ could arrange the DNA substrate in a specific orientation that would exert a destabilizing force on the paired bases of the ss/dsDNA junction (e.g. a critical bend-angle) and thereby ease dsDNA strand separation through ATP-driven translocation. This hypothesis solely relies on the increased DNA affinity by the presence of the CTD, which has been experimentally verified. Further support for this hypothesis requires experimental evidence that the CTD interacts with the dsDNA entity of the bound substrate, as such an interaction would be needed to enforce a critical bend angle at the ss/dsDNA junction (Fig. 5-3). However, the experimental data so far do not unambiguously support a direct interaction between the RecQ4 CTD and the dsDNA entity of a DNA substrate. Nonetheless, this alternative hypothesis provides an explanation for the detectable, albeit strongly reduced helicase activity of RecQ4$^{427\text{-}1116}$. As the CTD is not present in this particular RecQ4

variant, the dsDNA entity can be envisioned to be rather flexible and not constantly held in a fixed position. Thermal fluctuations or movements of the protein would allow the infrequent adaptation of near-critical bend-angle conformations for the DNA substrate, thus permitting a strongly reduced helicase activity of the RecQ4$^{427\text{-}1116}$ variant without a dsDNA separation wedge or the substrate-stabilizing CTD.

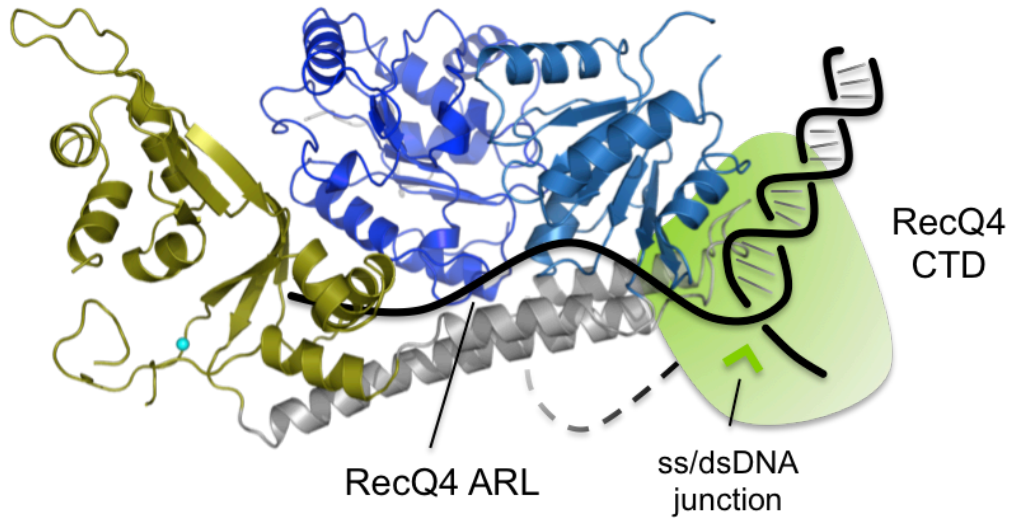


Figure 5-3: Potential model of DNA binding by RecQ4$^{427\text{-}1208}$. The RecQ4 CTD (green shape) presumably adopts the fold of a globular domain in close proximity to HD2 and the bridging helices (connected to helix α20 by the dashed linker). Helicase activity is either achieved by a wedge-like strand separation element within the RecQ4 CTD (green triangle) or by the specific arrangement of the DNA substrate (depicted in black), inducing a critical bend-angle at the ss/dsDNA junction. For the latter hypothesis, the CTD would orient the dsDNA entity of the substrate in a fixed position while the ssDNA segment is bound in the conserved fashion across the helicase core domains, in close proximity to the RecQ4 ARL.

Additional helicase experiments were mainly dedicated to explore the functional limitations of the RecQ4 helicase activity. The data illustrate for example that RecQ4$^{427\text{-}1208}$ is unable to separate any Y-fork DNA substrate that exhibits a dsDNA length of more than 18nt. Moreover, the C-terminally shortened RecQ4$^{427\text{-}1116}$ variant was only capable of separating a Y-forked dsDNA segment of up to 15nt. Overall, this indicates that the RecQ4 protein

is a rather weak and non-processive helicase compared to other RecQ family members. However, one has to consider that RecQ4 segments upstream of the helicase core or additional binding partners might increase the processivity of the RecQ4 enzyme. As expected, both RecQ4 variants were incapable of separating a 30nt blunt end DNA. Furthermore, an extension of the ssDNA loading site did not significantly increase helicase activity, thus arguing against the possibility of a collaborative effort in dsDNA strand separation by multiple in-line acting RecQ4 molecules. An interesting effect has been observed for the analysis of various ATP concentrations that were added last to the reaction mixture and initiated the helicase activity (Fig. 4-24). ATP concentrations of 2.5 mM and higher generally exhibited a strongly suppressing effect. Inhibition of the helicase mechanism by ATP itself is rather unlikely and a possible product-inhibition by ADP+P_i can be similarly excluded, as such high product concentrations are not present at the beginning of the reaction. Taking the results of the ATPase rate determinations into account, which are discussed in the next section, it is likely that the excessive presence of ATP obstructs the binding to the DNA, which consequently results in a recued fluorescence signal. The ATP molecule features a sugar-linked adenine base and a negatively charged triphosphate group, two characteristics that are shared with DNA. Thus, it is possible that the excessive presence of ATP shields basic residues that usually interact with the DNA backbone. Moreover, the adenine moiety of ATP might pair with the thymidine bases of the 15nt polyT ssDNA tail, which represents the loading site for RecQ4 on the DNA substrate in the helicase assay, thereby preventing the binding by the RecQ4 helicase. Regarding the previously discussed ATP-independent dsDNA strand separation activity, the H_2O control experiments that were recorded for the two RecQ4 variants in this experimental setup (Fig. 4-24), represent a prime illustration of this phenomenon. The data illustrate that RecQ4$^{427\text{-}1208}$ continuously exerts its ATP-independent dsDNA strand separation activity on the dsDNA helicase substrate, whereas this was not observed for RecQ4$^{427\text{-}1116}$. Considering the two discussed scenarios for the RecQ4 helicase mechanism, this behavior

strongly supports the hypothesis that the CTD may induce a destabilizing conformation within the DNA substrate, thus lowering the mechanical energy required to separate the two DNA strands. Due to the shortness of the dsDNA segment, thermal fluctuations might be sufficient to overcome the energetic barrier during the time when no ATP is present in case of the CTD containing RecQ4$^{427\text{-}1208}$ variant. Exploring the structural and functional mechanism of this ATP-independent destabilization of the dsDNA substrate might thus hold the key to determine whether the RecQ4$^{427\text{-}1208}$ CTD in fact binds to the dsDNA entity of the DNA and thereby enables the increased helicase activity of CTD-containing RecQ4 variants.

5.2.3.3. RecQ4 ATPase activity

The kinetic parameters for the ATP hydrolysis rates of RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ illustrate that both RecQ4 variants operate at a comparable velocity at any given ATP concentration. Similarly, the equivalent K_M values suggest an equal affinity for the ATP substrate. Thus, the data demonstrate that the C-terminal 92 amino acids of the RecQ4$^{427\text{-}1208}$ variant do not have any influence on the overall ATPase rate. If the characteristic step-size along the ssDNA per hydrolyzed ATP molecule can considered to be the same for both RecQ4 variants, it can be concluded that both enzymes operate at the same translocation velocity and that the differences in helicase activities between RecQ4$^{427\text{-}1116}$ and RecQ4$^{427\text{-}1208}$ are thus solely the result of the presence of the CTD, which in conclusion increases helicase effectively either actively by a wedge-like element or passively by ensuring a specific DNA substrate conformation that promotes strand separation. In analogy to the reduced helicase activity at high ATP concentrations, ATPase rates are gradually declining for ATP concentrations between 3 mM and 10 mM. These results support the above described hypothesis that the excessive presence of ATP might prevent the binding of the DNA to the RecQ4 protein, whose binding is, however, required to trigger ATP hydrolysis. In order to confirm this hypothesis, additional FP experiments should be performed that

investigate the binding affinities for a DNA substrate at for example high concentrations of either ADP or AMP-PNP.

5.2.4. Functional analysis of the R4ZBD mutants

The structural analysis of the R4ZBD suggests that it is unlikely to be involved in the processing of a canonical DNA substrate and may thus not be required for genuine helicase activity. In order to confirm this hypothesis several R4ZBD variants were generated and analyzed towards their ability to bind DNA and exert helicase activity. The data confirm that the R4ZBD WH domains are no major determinants for DNA binding. Similarly, all R4ZBD variants exhibited a WT-like helicase activity, thus providing functional evidence that the R4ZBD of RecQ4 does not assume any functions of the RQC domain that is present in other RecQ helicases. Solely the upper WH domain deletion variant exhibited a slight variation compared to the helicase performance of the WT protein, which might be based on secondary mechanistic effects as suggested in the experimental results section.

5.2.5. Functional analysis of the ARL variants

The here presented structural and functional analysis of RecQ4 implies that the R4ZBD is not required for dsDNA strand separation and that instead the CTD has a major influence with respect to DNA substrate binding and helicase activity. However, due to the lack of a DNA-bound crystal structure or structural information about the CTD itself, it is difficult to hypothesize whether RecQ4 separates the dsDNA via a conserved helicase mechanism or applies a RecQ unique strand separation technique. Therefore, a functional analysis of the RecQ4 ARL was performed to acquire mechanistic insights that could be compared to the results of the ARL studies on the *Ec*RecQ helicase and might thus provide additional information towards the RecQ4 helicase mechanism. The ARL in *Ec*RecQ recognizes the binding of

a DNA substrate (ssDNA) and subsequently enables ATP hydrolysis, presumably by delicate structural rearrangements within the helicase core[51,63]. In order to investigate whether the RecQ4 ARL serves the same purpose, the conserved ARL residues W613, H615, F617 and R618 were substituted for leucine and all variants were analyzed via FP DNA binding experiments, ATPase rate determination and the fluorescence-based helicase assay. In general, the RecQ4 ARL data yielded comparable results to the *Ec*RecQ ARL studies. None of the single amino acid substitutions affected overall DNA binding of a ssDNA substrate. However, ATPase rates as well as helicase activity were significantly impaired. While the ARL variants W613L and H615L were similarly reduced in their ATPase rates and helicase activity, F617L and R618L displayed a disproportional reduction of those two processes. F617L achieved only 35% of the maximum WT ATPase rate, yet its helicase activity was only reduced by 15% compared to the WT protein. In contrast, although R618L still exhibited 21% of the WT ATPase rate, the RecQ4 variant was entirely unable to separate the dsDNA substrate (Table 5-1). The data permits to draw conclusions about the functional relevance of each tested ARL residue. For example, the fact that W613L and H615L are equivalently reduced in ATPase rate and helicase activity suggests that both residues are involved in secondary structure rearrangements, which transfer the signal of DNA binding to enable ATP hydrolysis. The impaired signal transmission results in an equally impaired ATP hydrolysis rate and, consequently, a reduced translocation-based dsDNA strand separation. The F617L and R618L variants were even more impaired in their ability to hydrolyze ATP, suggesting that both residues are involved in ssDNA recognition, too. However, their disproportional reduction in helicase activity suggests an additional function in the helicase mechanism for those two residues. This function could be the direct involvement in ssDNA translocation. The F617L substitution exhibits a near WT-like helicase activity, while being severely reduced in ATP turnover rates. This suggests that F617 exhibits a decelerating function in ssDNA translocation, which is abolished by the leucine substitution. In other words, the RecQ4 F617L

variant can achieve much more translocation per hydrolyzed ATP molecule, e.g. displays an increased efficiency, as it features an enhanced helicase activity when the phenylalanine is not present. Moreover, the data for the F617L variant depict a faint but significant ATPase activity even in the absence of DNA, a phenomenon that has been similarly observed for the *Ec*RecQ phenylalanine substitution[63]. It is likely that the F617L replacement thus represents a physiological state of the RecQ4 (and *Ec*RecQ) protein that resembles ssDNA binding even in the absence of DNA.

Table 5-1: Summary of the RecQ4 ARL analysis presented in Fig. 4-27. Efficiency is defined as the amount of helicase activity (% WT Helicase) per ATPase activity (% WT ATPase) for each ARL variant. Efficiency values of 1 thus indicate a direct correlation between the two activities, whereas higher values (green) represent an increase in helicase activity per hydrolyzed ATP and lower values (red) symbolize a reduced helicase activity per hydrolyzed ATP.

	WT	**W613L**	**H615L**	**F617L**	**R618L**
% WT ATPase	100.0	63.2	61.7	34.7	21.1
% WT Helicase	100.0	58.2	50.5	84.7	2.0
Efficiency (Helicase/ATPase)	1.0	0.92	0.82	2.44	0.1

In contrast, the R618L mutation is completely deficient in helicase activity although still exhibiting 21% of WT ATPase activity. This correlation implies a propulsive function for R618 in the ssDNA translocation mechanism, which is uncoupled from ATP hydrolysis by the substitution with the leucine residue. Thus, R618L can still transmit the signal of ssDNA binding to enable ATP hydrolysis to a limited extend, however, it is unable to utilize the energy from ATP hydrolysis for protein translocation and, consequently, helicase activity. Combined, the mutational analysis of the RecQ4 ARL demonstrates that the binding mode for ssDNA across the surface of the helicase core, in close proximity to the ARL, is conserved between RecQ4 and the *Ec*RecQ helicase. Furthermore, helicase activity is likely to be achieved via the conserved translocation mechanism along the bound ssDNA upon ATP

hydrolysis. Thus, one can assume that the general helicase mechanism, which is dsDNA strand separation by ATP-dependent translocation, is conserved and that DNA binding by RecQ4 likely resembles a comparable DNA trajectory as depicted for other RecQ helicases that have been crystallized in the presence of a DNA substrate (compare Fig. 4-11 C and D). Importantly, this conserved orientation of the DNA substrate should position its dsDNA entity in close proximity to the RecQ4 CTD, which evidently has an influence on DNA binding and the helicase mechanism (Fig. 5-3).

5.3. A glimpse into RecQ4-associated disease mutations

The Human Gene Mutation Database (HGMD, http://www.biobase-international.com) is currently listing 70 individual RecQ4 mutations (feeless version) that are associated either with the Rothmund-Thomson Syndrome (RTS), RAPADILINO Syndrome (RAPA) or Baller-Gerold Syndrome (BGS). Mutation types include missense or nonsense mutations (30), splice-site alterations (11), deletions (25) and insertions (4). With the structure of the RecQ4 helicase in hand it is now possible to evaluate such mutations and analyze their impact on the structural integrity or the function of the protein (Fig. 5-4). However, such an assessment is not trivial as RecQ4-associated diseases are of the recessive type and thus pathogenesis is generally the result of biallelic mutations, which are often heterogenic. The structural analysis of RecQ4 mutations can thus only serve as a first approximation and individual mutations have to be thoroughly tested by functional assays or the analysis of protein stability, for example via CD spectroscopy. The analysis of nonsense mutations, which represents the alteration of a sense-codon to a stop-codon, is probably the most straightforward. Premature termination of the polypeptide chain typically results in the loss of functional elements or compromises the overall fold of the protein when the termination occurs within a domain itself. Similarly, nucleotide deletions or insertions usually cause a frame shift in the amino acid sequence, which sooner or later leads to a stop codon (e.g. Cys525AlafsX33, stop codon after 33 out-of-frame

amino acids subsequent to the frame-shift). In contrast, splice-site mutations either affect regulatory splice-sequence elements or take place within the intron regions themselves, resulting in critically shortened introns that are too small to be correctly processed. Such a pre-mRNA splice-site mutation results in the RecQ4 Ala420_Ala463del variant, which is missing a large part of the major RecQ4 NTS and represents a common RAPA founder mutation in the Finnish population[185]. Missense mutations, on the other hand, represent a very attractive variation of the wild-type protein as such single amino acid substitutions might provide the most detailed information with respect to the function of a particular amino acid or region of the protein. A selection of missense mutations, early termination mutations, duplications and deletions are illustrated in Fig. 5-4.

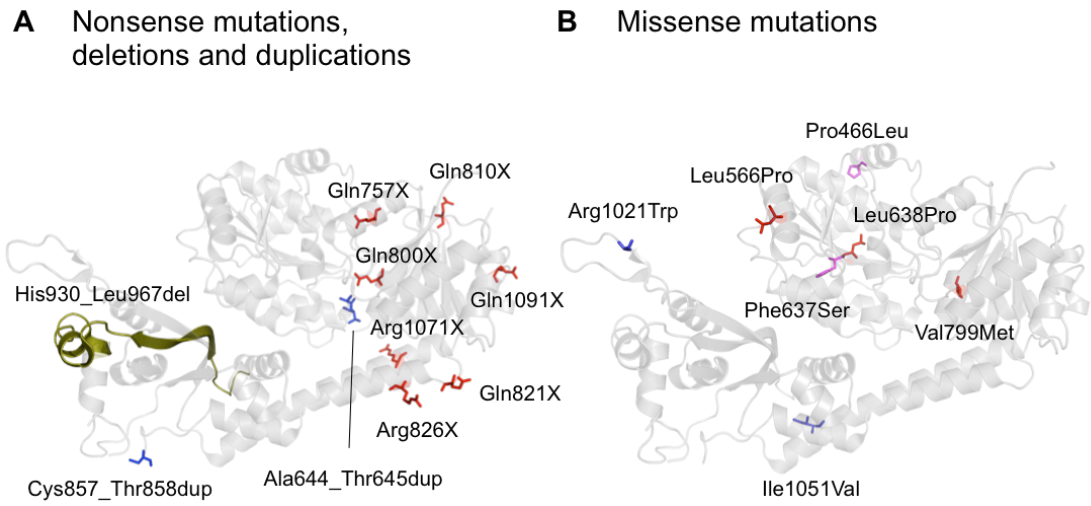


Figure 5-4: RecQ4 patient mutations within the RecQ4$^{427\text{-}1116}$ structure. (A) The location of nonsense mutations (red), deletions (olive) and duplications (blue). **(B)** Location of single amino acid substitutions that are associated with RTS (red), RAPA (magenta) or BGS (blue).

Considering the location of premature termination mutations within the RecQ4 protein (Fig. 5-4 A), it appears that they cluster either directly within HD2 or in secondary structure elements that are closely associated with HD2, e.g. the α-helices that mimic the RQC Zn domain. Early protein

terminations that occur within the HD2 are commonly associated with RTS, whereas premature terminations in the very C-terminus, e.g. Arg1071X and Gln1091X, are common in RAPA patients[183]. However, there are exceptions to this tendency. The two aforementioned early termination mutations are located within the large bridging helix (α19) and in the loop right before helix α20. Both are heterozygous with the RecQ4 Ala420_Ala463del variant, which, if correctly folded, is probably not imported into the nucleus based on its compromised NTS. An approach to express and purify a RecQ4$^{427\text{-}1090}$ variant in *E. coli*, representing the Gln1091X mutation, resulted in insoluble protein, which is a strong indication that the presence of helix α20 is essential for protein stability. This preliminary mutational analysis thus suggests that protein alterations, which compromise the function of the helicase core, yet preserve its (although sometimes only partial) structural presence, are more likely to cause RTS. In contrast, mutations that result in the depletion of the RecQ4 protein, either by sequestration from the nucleus or protein aggregation, are more commonly observed in RAPA patients. Interestingly, the two missense mutations that are associated with BGS are both located within the R4ZBD (Fig. 5-4 B) and are heterozygote with either a Glu166X mutation or the Cys525Alafs33X variant. An expression approach of the RecQ4$^{427\text{-}1208}$ R1021W variant (that is associated with BGS) resulted in soluble protein and subsequent functional assays showed no significant difference to WT RecQ4$^{427\text{-}1208}$. This result suggests that the development of BGS might thus be the result of an impaired RecQ4 interactome, presuming the R4ZBD represents indeed a module for protein-protein interactions. In general, however, all of these functional indications need to be verified and individually analyzed. For the early termination mutations, such a validation is best performed *in vivo*, investing the physical presence of the RecQ4 variants and their subcellular location via protein staining methods. The single amino acid substitutions that represent patient missense mutations, should be analyzed via *in vitro* function analysis or protein-interaction studies which may provide valuable information towards the functional basis of the corresponding RecQ4-associated diseases.

6. Future perspectives

Based on the comprehensive characterization of the RecQ4 helicase presented in this study, we obtained novel and valuable insights into the structure and function of this essential genome caretaker protein. Especially the structure of RecQ4 constitutes the last piece of the puzzle towards a unified structural knowledge of all five human RecQ helicases and provided first insights into RecQ4's unique C-terminus. However, some important questions remain unanswered and should thus be the focus of future investigations. One of the most intriguing issues concerns the structure of the RecQ4 CTD, e.g. the very C-terminal 92 amino acids, and their functional implication in dsDNA strand separation. Two reasonable mechanisms for RecQ4 helicase activity have been discussed here, which should be further addressed by structural and functional studies. For this purpose, a variety of new RecQ4 variants should be expressed and analyzed for their crystallization potential. One candidate for structural investigations is clearly the upper WH domain deletion variant, e.g. RecQ4$^{427\text{-}1208}$ del944-1032, which has already been utilized for functional investigations of the R4ZBD. Furthermore, a RecQ4 variant in which the entire R4ZBD is eliminated (e.g. RecQ4$^{427\text{-}1208}$ del839-1060) should be established and analyzed as well. In those deletion variants, the entire helicase core as well as the bridging helices and the CTD of RecQ4 would be present, thus permitting the structural characterization of the CTD in the context of its neighboring domains. The structure of RecQ4$^{427\text{-}1116}$ indicates that the CTD and the R4ZBD occupy opposing sides with respect to the helicase core and are therefore unlikely to interact with each other. Thus, it should be feasible to exclude the R4ZBD from the RecQ4 protein without affecting the structure or orientation of the CTD. Moreover, the R4ZBD represents a partially flexible and complex domain and, consequently, its removal (or parts of it) should ultimately increase the likelihood to produce diffraction-quality crystals for the subsequent structural characterization. In addition, the R4ZBD-deletion

variants can be utilized for DNA co-crystallization attempts, as the data suggests that the R4ZBD is not involved in canonical DNA substrate binding.
Lastly, a RecQ4 variant that solely comprises the residues of the CTD (e.g. RecQ4$^{1090-1208}$) would be valuable for structural as well as functional studies. The isolated RecQ4 CTD variant could provide the missing information regarding a possible interaction between the CTD and dsDNA, thus either supporting or refuting the proposed bend-angle driven RecQ4 helicase mechanism. If the RecQ4 deletion variants mentioned above fail to provide structural information via the crystallographic approach, the relatively short RecQ4 CTD variant would have suitable properties to be structurally characterized via NMR spectroscopy.
In addition to the structural approach, a functional analysis of the RecQ4 CTD will provide valuable complementary insights towards RecQ4s helicase mechanism. To probe the existence of a potential helical wedge-like element within the CTD, which could facilitate dsDNA strand separation, successive extension of the RecQ4$^{427-1116}$ variant, one CTD α-helix at a time, could identify such an element. One would expect a sudden increase of helicase activity from the low value of the RecQ4$^{427-1116}$ variant to an activity, which reflects the one of the full C-terminal RecQ4$^{427-1208}$ protein once the wedge-like element (if it exists) is included in the variant. Furthermore, a mutational screen addressing the highly conserved basic and aromatic amino acids of the CTD would further complement such functional studies. An analysis of these RecQ4 CTD mutation variants (analogous to the R4ZBD mutation variants) for DNA binding and helicase activity could identify whether and how the CTD would facilitate dsDNA strand separation via DNA binding, favoring the bend-angle hypothesis, or an aromatic separation pin, analogous to the ß-hairpin element of other human RecQ helicases.
A second hotspot for future studies constitutes the function of the novel RecQ4 R4ZBD. Our structural analysis suggests that the two R4ZBD WH domains could represent a module either for protein-protein interactions or to facilitate the binding of non-canonical DNA substrates, e.g. Holliday-Junctions, D-Loops or G-quaduplex DNA. Utilizing the aforementioned

R4ZBD deletion variants and/or the already existing R4ZBD mutational variants, one could study the binding to a variety of synthetic DNA substrates that represent such non-canonical DNA structures via, for example, fluorescence polarization. A promising candidate for a potential interaction with the RecQ4 protein is the poly (ADP-ribose) polymerase-1 (PARP-1). PARP-1 interacts with RecQ4 via its BRCT domain, an approximately 100 aa domain, through which it mediates multiple protein-protein interactions with numerous proteins involved in DNA repair pathways. The RecQ4 interaction site with respect to PARP-1 was broadly mapped to the C-terminus, ranging from aa 833-1208. Thus, the established RecQ4 protein variants, $RecQ4^{427\text{-}1116}$ and $RecQ4^{427\text{-}1208}$, partially or fully include the PARP-1 interaction site. The binding of both proteins should first be verified *in vitro* via analytical SEC using recombinantly expressed proteins, to determine and confirm the exact domain boundaries for the interaction of both proteins. Follow-up studies should then address the structural characterization via co-crystallization of the minimal complex and the analysis of functional binding properties, for example via isothermal titration calorimetry (ITC). The *in vitro* data should then be further complemented by *in vivo* studies, probing the functional RecQ4-PARP-1 interaction under DNA damage stress conditions in cell-based assays. Moreover, the structural and functional characterization of the RecQ4-PARP-1 interaction could be the framework for the screening of potential inhibitors, which modulate this interaction and might thus set the stage for the development of new therapeutic compounds that could be used in cancer therapy.

Finally, a considerable effort should be made to produce and purify the full-length RecQ4 helicase. So far, the large intrinsically unstructured N-terminus of RecQ4 has prevented our attempts to recombinantly express and purify full-length RecQ4 from *E. coli*. This issue could be overcome by RecQ4 co-expression with potential binding partners, such as p53 or XPA, which could stabilize the N-terminal part and thus possibly circumvent protein aggregation and/or degradation. Alternatively, recombinant expression in eukaryotic cell culture should be considered, utilizing for example human embryonic kidney

cells (HEK 293 cells) or insect cell culture. The RecQ4 N-terminus is so far both functionally and structurally poorly understood. It was suggested that this region is responsible for the strong DNA annealing activity of RecQ4 and potentially facilitates the interaction with a variety of non-canonical DNA substrates, yet lacks a defined structural fold. An *in vitro* analysis of the human full-length RecQ4, using the methods applied in this study, would thus greatly complement the functional conclusions that have been presented here.

7. Literature

1. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**, 585-98 (2014).
2. Lindahl, T. & Barnes, D.E. Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-33 (2000).
3. De Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-85 (2004).
4. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-15 (1993).
5. Nakamura, J. & Swenberg, J.A. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res* **59**, 2522-6 (1999).
6. Urban, V., Dobrovolna, J. & Janscak, P. Distinct functions of human RecQ helicases during DNA replication. *Biophys Chem* **225**, 20-26 (2017).
7. Swenberg, J.A. et al. Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci* **120 Suppl 1**, S130-45 (2011).
8. Busuttil, R., Bahar, R. & Vijg, J. Genome dynamics and transcriptional deregulation in aging. *Neuroscience* **145**, 1341-7 (2007).
9. Waters, L.S. et al. Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev* **73**, 134-54 (2009).
10. Oey, H. & Whitelaw, E. On the meaning of the word 'epimutation'. *Trends Genet* **30**, 519-20 (2014).
11. Sergiev, P.V., Dontsova, O.A. & Berezkin, G.V. Theories of aging: an ever-evolving field. *Acta Naturae* **7**, 9-18 (2015).
12. Lopez-Otin, C., Blasco, M.A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-217 (2013).
13. Vermeij, W.P., Hoeijmakers, J.H. & Pothof, J. Genome Integrity in Aging: Human Syndromes, Mouse Models, and Therapeutic Options. *Annu Rev Pharmacol Toxicol* **56**, 427-45 (2016).
14. Vijg, J. & Suh, Y. Genome instability and aging. *Annu Rev Physiol* **75**, 645-68 (2013).
15. Goto, M. Hierarchical deterioration of body systems in Werner's syndrome: implications for normal ageing. *Mech Ageing Dev* **98**, 239-54 (1997).
16. Navarro, C.L., Cau, P. & Levy, N. Molecular bases of progeroid syndromes. *Hum Mol Genet* **15 Spec No 2**, R151-61 (2006).
17. Pollex, R.L. & Hegele, R.A. Hutchinson-Gilford progeria syndrome. *Clin Genet* **66**, 375-81 (2004).
18. Redwood, A.B. et al. A dual role for A-type lamins in DNA double-strand break repair. *Cell Cycle* **10**, 2549-60 (2011).
19. Lehmann, A.R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev* **15**, 15-23 (2001).
20. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
21. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
22. Nitta, M. et al. Spindle checkpoint function is required for mitotic catastrophe induced by DNA-damaging agents. *Oncogene* **23**, 6548-58 (2004).
23. Allen, T.M. Ligand-targeted therapeutics in anticancer therapy. *Nat Rev Cancer* **2**, 750-63 (2002).
24. Wu, Y. Unwinding and rewinding: double faces of helicase? *J Nucleic Acids* **2012**, 140601 (2012).
25. Tuteja, N. & Tuteja, R. Unraveling DNA helicases. Motif, structure, mechanism and function. *Eur J Biochem* **271**, 1849-63 (2004).
26. Walker, J.E., Saraste, M., Runswick, M.J. & Gay, N.J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-

requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-51 (1982).
27. Subramanya, H.S., Bird, L.E., Brannigan, J.A. & Wigley, D.B. Crystal structure of a DExx box DNA helicase. *Nature* **384**, 379-83 (1996).
28. Fairman-Williams, M.E., Guenther, U.P. & Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20**, 313-24 (2010).
29. Singleton, M.R., Dillingham, M.S. & Wigley, D.B. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**, 23-50 (2007).
30. Eoff, R.L. & Raney, K.D. Helicase-catalysed translocation and strand separation. *Biochem Soc Trans* **33**, 1474-8 (2005).
31. Hickson, I.D. RecQ helicases: caretakers of the genome. *Nat Rev Cancer* **3**, 169-78 (2003).
32. Chu, W.K. & Hickson, I.D. RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer* **9**, 644-54 (2009).
33. Croteau, D.L., Popuri, V., Opresko, P.L. & Bohr, V.A. Human RecQ helicases in DNA repair, recombination, and replication. *Annu Rev Biochem* **83**, 519-52 (2014).
34. Nakayama, H. et al. Isolation and genetic characterization of a thymineless death-resistant mutant of Escherichia coli K12: identification of a new mutation (recQ1) that blocks the RecF recombination pathway. *Mol Gen Genet* **195**, 474-80 (1984).
35. Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev* **58**, 401-65 (1994).
36. Courcelle, J. & Hanawalt, P.C. RecQ and RecJ process blocked replication forks prior to the resumption of replication in UV-irradiated Escherichia coli. *Mol Gen Genet* **262**, 543-51 (1999).
37. Hanada, K. et al. RecQ DNA helicase is a suppressor of illegitimate recombination in Escherichia coli. *Proc Natl Acad Sci U S A* **94**, 3860-5 (1997).
38. Gangloff, S., McDonald, J.P., Bendixen, C., Arthur, L. & Rothstein, R. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol* **14**, 8391-8 (1994).
39. Stewart, E., Chapman, C.R., Al-Khodairy, F., Carr, A.M. & Enoch, T. rqh1+, a fission yeast gene related to the Bloom's and Werner's syndrome genes, is required for reversible S phase arrest. *EMBO J* **16**, 2682-92 (1997).
40. Groocock, L.M., Prudden, J., Perry, J.J. & Boddy, M.N. The RecQ4 orthologue Hrq1 is critical for DNA interstrand cross-link repair and genome stability in fission yeast. *Mol Cell Biol* **32**, 276-87 (2012).
41. Bernstein, K.A., Gangloff, S. & Rothstein, R. The RecQ DNA helicases in DNA repair. *Annu Rev Genet* **44**, 393-417 (2010).
42. Khadka, P., Croteau, D.L. & Bohr, V.A. RECQL5 has unique strand annealing properties relative to the other human RecQ helicase proteins. *DNA Repair (Amst)* **37**, 53-66 (2016).
43. Kawabe, T. et al. Differential regulation of human RecQ family helicases in cell transformation and cell cycle. *Oncogene* **19**, 4764-72 (2000).
44. Fang, H. et al. RecQL4 helicase amplification is involved in human breast tumorigenesis. *PLoS One* **8**, e69600 (2013).
45. Futami, K. & Furuichi, Y. RECQL1 and WRN DNA repair helicases: potential therapeutic targets and proliferative markers against cancers. *Front Genet* **5**, 441 (2014).
46. Futami, K., Ogasawara, S., Goto, H., Yano, H. & Furuichi, Y. RecQL1 DNA repair helicase: A potential tumor marker and therapeutic target against hepatocellular carcinoma. *Int J Mol Med* **25**, 537-45 (2010).
47. Sanada, S. et al. RECQL1 DNA repair helicase: a potential therapeutic target and a proliferative marker against ovarian cancer. *PLoS One* **8**, e72820 (2013).
48. Swan, M.K. et al. Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA. *Acta Crystallogr D Biol Crystallogr* **70**, 1465-75 (2014).
49. Pike, A.C. et al. Structure of the human RECQ1 helicase reveals a putative strand-separation pin. *Proc Natl Acad Sci U S A* **106**, 1039-44 (2009).

50. Kitano, K., Kim, S.Y. & Hakoshima, T. Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN. *Structure* **18**, 177-87 (2010).
51. Manthei, K.A., Hill, M.C., Burke, J.E., Butcher, S.E. & Keck, J.L. Structural mechanisms of DNA binding and unwinding in bacterial RecQ helicases. *Proc Natl Acad Sci U S A* **112**, 4292-7 (2015).
52. Liu, Z. et al. The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins. *Structure* **7**, 1557-66 (1999).
53. Harami, G.M., Nagy, N.T., Martina, M., Neuman, K.C. & Kovacs, M. The HRDC domain of E. coli RecQ helicase controls single-stranded DNA translocation and double-stranded DNA unwinding rates without affecting mechanoenzymatic coupling. *Sci Rep* **5**, 11091 (2015).
54. Kitano, K., Yoshihara, N. & Hakoshima, T. Crystal structure of the HRDC domain of human Werner syndrome protein, WRN. *J Biol Chem* **282**, 2717-28 (2007).
55. Sato, A. et al. Solution structure of the HRDC domain of human Bloom syndrome protein BLM. *J Biochem* **148**, 517-25 (2010).
56. Wu, L. et al. The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *EMBO J* **24**, 2679-87 (2005).
57. Newman, J.A. et al. Crystal structure of the Bloom's syndrome helicase indicates a role for the HRDC domain in conformational changes. *Nucleic Acids Res* **43**, 5221-35 (2015).
58. Vindigni, A., Marino, F. & Gileadi, O. Probing the structural basis of RecQ helicase function. *Biophys Chem* **149**, 67-77 (2010).
59. Lucic, B. et al. A prominent beta-hairpin structure in the winged-helix domain of RECQ1 is required for DNA unwinding and oligomer formation. *Nucleic Acids Res* **39**, 1703-17 (2011).
60. Kim, S.Y., Hakoshima, T. & Kitano, K. Structure of the RecQ C-terminal domain of human Bloom syndrome protein. *Sci Rep* **3**, 3294 (2013).
61. Bernstein, D.A., Zittel, M.C. & Keck, J.L. High-resolution structure of the E.coli RecQ helicase catalytic core. *EMBO J* **22**, 4910-21 (2003).
62. Newman, J.A., Aitkenhead, H., Savitsky, P. & Gileadi, O. Insights into the RecQ helicase mechanism revealed by the structure of the helicase domain of human RECQL5. *Nucleic Acids Res* **45**, 4231-4243 (2017).
63. Zittel, M.C. & Keck, J.L. Coupling DNA-binding and ATP hydrolysis in Escherichia coli RecQ: role of a highly conserved aromatic-rich sequence. *Nucleic Acids Res* **33**, 6982-91 (2005).
64. Dillingham, M.S., Soultanas, P. & Wigley, D.B. Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic Acids Res* **27**, 3310-7 (1999).
65. Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**, 635-47 (1997).
66. Seki, M. et al. Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12. *Nucleic Acids Res* **22**, 4566-73 (1994).
67. Wang, M. et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics* **11**, 492-500 (2012).
68. Berti, M. et al. Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition. *Nat Struct Mol Biol* **20**, 347-54 (2013).
69. Martin-Hernandez, K., Rodriguez-Vargas, J.M., Schreiber, V. & Dantzer, F. Expanding functions of ADP-ribosylation in the maintenance of genome integrity. *Semin Cell Dev Biol* **63**, 92-101 (2017).
70. Popuri, V. et al. The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J Biol Chem* **283**, 17766-76 (2008).
71. Cui, S. et al. Characterization of the DNA-unwinding activity of human RECQ1, a helicase specifically stimulated by human replication protein A. *J Biol Chem* **278**, 1424-32 (2003).

72. Bugreev, D.V., Mazina, O.M. & Mazin, A.V. Bloom syndrome helicase stimulates RAD51 DNA strand exchange activity through a novel mechanism. *J Biol Chem* **284**, 26349-59 (2009).
73. Sung, P. & Klein, H. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol* **7**, 739-50 (2006).
74. Branzei, D. & Foiani, M. RecQ helicases queuing with Srs2 to disrupt Rad51 filaments and suppress recombination. *Genes Dev* **21**, 3019-26 (2007).
75. Mazina, O.M., Rossi, M.J., Deakyne, J.S., Huang, F. & Mazin, A.V. Polarity and bypass of DNA heterology during branch migration of Holliday junctions by human RAD54, BLM, and RECQ1 proteins. *J Biol Chem* **287**, 11820-32 (2012).
76. Muzzolini, L. et al. Different quaternary structures of human RECQ1 are associated with its dual enzymatic activity. *PLoS Biol* **5**, e20 (2007).
77. Sharma, S., Phatak, P., Stortchevoi, A., Jasin, M. & Larocque, J.R. RECQ1 plays a distinct role in cellular response to oxidative DNA damage. *DNA Repair (Amst)* **11**, 537-49 (2012).
78. Sharma, S. & Brosh, R.M., Jr. Human RECQ1 is a DNA damage responsive protein required for genotoxic stress resistance and suppression of sister chromatid exchanges. *PLoS One* **2**, e1297 (2007).
79. Doherty, K.M. et al. RECQ1 helicase interacts with human mismatch repair factors that regulate genetic recombination. *J Biol Chem* **280**, 28085-94 (2005).
80. Parvathaneni, S., Stortchevoi, A., Sommers, J.A., Brosh, R.M., Jr. & Sharma, S. Human RECQ1 interacts with Ku70/80 and modulates DNA end-joining of double-strand breaks. *PLoS One* **8**, e62481 (2013).
81. Sharma, S. et al. RECQL, a member of the RecQ family of DNA helicases, suppresses chromosomal instability. *Mol Cell Biol* **27**, 1784-94 (2007).
82. Li, D. et al. Single nucleotide polymorphisms of RecQ1, RAD54L, and ATM genes are associated with reduced survival of pancreatic cancer. *J Clin Oncol* **24**, 1720-8 (2006).
83. Cybulski, C. et al. Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* **47**, 643-6 (2015).
84. Li, X.L. et al. Identification of RECQ1-regulated transcriptome uncovers a role of RECQ1 in regulation of cancer cell migration and invasion. *Cell Cycle* **13**, 2431-45 (2014).
85. Lu, X., Parvathaneni, S., Li, X.L., Lal, A. & Sharma, S. Transcriptome guided identification of novel functions of RECQ1 helicase. *Methods* **108**, 111-7 (2016).
86. Johnson, J.E., Cao, K., Ryvkin, P., Wang, L.S. & Johnson, F.B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res* **38**, 1114-22 (2010).
87. Nguyen, G.H. et al. Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc Natl Acad Sci U S A* **111**, 9905-10 (2014).
88. Zhang, P. et al. Identification of genes associated with cisplatin resistance in human oral squamous cell carcinoma cell line. *BMC Cancer* **6**, 224 (2006).
89. Arora, A. et al. Clinicopathological and Functional Significance of RECQL1 Helicase in Sporadic Breast Cancers. *Mol Cancer Ther* **16**, 239-250 (2017).
90. Viziteu, E. et al. RECQ1 helicase is involved in replication stress survival and drug resistance in multiple myeloma. *Leukemia* (2017).
91. Arai, A. et al. RECQL1 and WRN proteins are potential therapeutic targets in head and neck squamous cell carcinoma. *Cancer Res* **71**, 4598-607 (2011).
92. Mendoza-Maldonado, R. et al. The human RECQ1 helicase is highly expressed in glioblastoma and plays an important role in tumor cell proliferation. *Mol Cancer* **10**, 83 (2011).
93. Futami, K. et al. Induction of mitotic cell death in cancer cells by small interference RNA suppressing the expression of RecQL1 helicase. *Cancer Sci* **99**, 71-80 (2008).
94. Jewell, R. et al. Patterns of expression of DNA repair genes and relapse from melanoma. *Clin Cancer Res* **16**, 5211-21 (2010).
95. Sharma, S. An appraisal of RECQ1 expression in cancer progression. *Front Genet* **5**, 426 (2014).

96. Karow, J.K., Chakraverty, R.K. & Hickson, I.D. The Bloom's syndrome gene product is a 3'-5' DNA helicase. *J Biol Chem* **272**, 30611-4 (1997).
97. Ellis, N.A. et al. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655-66 (1995).
98. Shi, J. et al. A helical bundle in the N-terminal domain of the BLM helicase mediates dimer and potentially hexamer formation. *J Biol Chem* **292**, 5909-5920 (2017).
99. Zhong, S. et al. A role for PML and the nuclear body in genomic stability. *Oncogene* **18**, 7941-7 (1999).
100. Bischof, O. et al. Regulation and localization of the Bloom syndrome protein in response to DNA damage. *J Cell Biol* **153**, 367-80 (2001).
101. Bartos, J.D., Wang, W., Pike, J.E. & Bambara, R.A. Mechanisms by which Bloom protein can disrupt recombination intermediates of Okazaki fragment maturation. *J Biol Chem* **281**, 32227-39 (2006).
102. Neelsen, K.J. & Lopes, M. Replication fork reversal in eukaryotes: from dead end to dynamic response. *Nat Rev Mol Cell Biol* **16**, 207-20 (2015).
103. Kowalczykowski, S.C. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb Perspect Biol* **7**(2015).
104. Gravel, S., Chapman, J.R., Magill, C. & Jackson, S.P. DNA helicases Sgs1 and BLM promote DNA double-strand break resection. *Genes Dev* **22**, 2767-72 (2008).
105. Nimonkar, A.V. et al. BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev* **25**, 350-62 (2011).
106. Bugreev, D.V., Yu, X., Egelman, E.H. & Mazin, A.V. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev* **21**, 3085-94 (2007).
107. Wu, L. & Hickson, I.D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870-4 (2003).
108. Singh, T.R. et al. BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome. *Genes Dev* **22**, 2856-68 (2008).
109. Xu, D. et al. RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability. *Genes Dev* **22**, 2843-55 (2008).
110. German, J. Bloom's syndrome. *Dermatol Clin* **13**, 7-18 (1995).
111. Lillard-Wetherell, K. et al. Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. *Hum Mol Genet* **13**, 1919-32 (2004).
112. Opresko, P.L. et al. POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J Biol Chem* **280**, 32069-80 (2005).
113. Hagelstrom, R.T., Blagoev, K.B., Niedernhofer, L.J., Goodwin, E.H. & Bailey, S.M. Hyper telomere recombination accelerates replicative senescence and may promote premature aging. *Proc Natl Acad Sci U S A* **107**, 15768-73 (2010).
114. Rezazadeh, S. On BLM helicase in recombination-mediated telomere maintenance. *Mol Biol Rep* **40**, 3049-64 (2013).
115. Pickett, H.A. & Reddel, R.R. Molecular mechanisms of activity and derepression of alternative lengthening of telomeres. *Nat Struct Mol Biol* **22**, 875-80 (2015).
116. Sun, H., Karow, J.K., Hickson, I.D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem* **273**, 27587-92 (1998).
117. Huppert, J.L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res* **33**, 2908-16 (2005).
118. Wu, W.Q., Hou, X.M., Li, M., Dou, S.X. & Xi, X.G. BLM unfolds G-quadruplexes in different structural environments through different mechanisms. *Nucleic Acids Res* **43**, 4614-26 (2015).
119. Tangeman, L., McIlhatton, M.A., Grierson, P., Groden, J. & Acharya, S. Regulation of BLM Nucleolar Localization. *Genes (Basel)* **7**(2016).
120. Babbe, H. et al. Genomic instability resulting from Blm deficiency compromises development, maintenance, and function of the B cell lineage. *J Immunol* **182**, 347-60 (2009).
121. Bloom, D. Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs; probably a syndrome entity. *AMA Am J Dis Child* **88**, 754-8 (1954).

122. German, J., Archibald, R. & Bloom, D. Chromosomal Breakage in a Rare and Probably Genetically Determined Syndrome of Man. *Science* **148**, 506-7 (1965).
123. German, J., Crippa, L.P. & Bloom, D. Bloom's syndrome. III. Analysis of the chromosome aberration characteristic of this disorder. *Chromosoma* **48**, 361-6 (1974).
124. Hand, R. & German, J. A retarded rate of DNA chain growth in Bloom's syndrome. *Proc Natl Acad Sci U S A* **72**, 758-62 (1975).
125. Guo, R.B. et al. Structural and functional analyses of disease-causing missense mutations in Bloom syndrome protein. *Nucleic Acids Res* **35**, 6297-310 (2007).
126. Keller, C., Keller, K.R., Shew, S.B. & Plon, S.E. Growth deficiency and malnutrition in Bloom syndrome. *J Pediatr* **134**, 472-9 (1999).
127. Diaz, A., Vogiatzi, M.G., Sanz, M.M. & German, J. Evaluation of short stature, carbohydrate metabolism and other endocrinopathies in Bloom's syndrome. *Horm Res* **66**, 111-7 (2006).
128. Arora, H. et al. Bloom syndrome. *Int J Dermatol* **53**, 798-802 (2014).
129. Ozawa, T., Kondo, N., Motoyoshi, F., Kasahara, K. & Orii, T. DNA mutation induced in the sequence upstream of the secreted MYU C-terminal coding sequence by ultraviolet irradiation in the cell line of Bloom's syndrome. *Eur J Immunogenet* **22**, 163-9 (1995).
130. Evans, H.J., Adams, A.C., Clarkson, J.M. & German, J. Chromosome aberrations and unscheduled DNA synthesis in X- and UV-irradiated lymphocytes from a boy with Bloom's syndrome and a man with xeroderma pigmentosum. *Cytogenet Cell Genet* **20**, 124-40 (1978).
131. Cunniff, C., Bassetti, J.A. & Ellis, N.A. Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition. *Mol Syndromol* **8**, 4-23 (2017).
132. de Renty, C. & Ellis, N.A. Bloom's syndrome: Why not premature aging?: A comparison of the BLM and WRN helicases. *Ageing Res Rev* **33**, 36-51 (2017).
133. Mao, F.J., Sidorova, J.M., Lauper, J.M., Emond, M.J. & Monnat, R.J. The human WRN and BLM RecQ helicases differentially regulate cell proliferation and survival after chemotherapeutic DNA damage. *Cancer Res* **70**, 6548-55 (2010).
134. Nguyen, G.H. et al. A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem Biol* **20**, 55-62 (2013).
135. Yu, C.E. et al. Positional cloning of the Werner's syndrome gene. *Science* **272**, 258-62 (1996).
136. Huang, S. et al. The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. *Nat Genet* **20**, 114-6 (1998).
137. Gray, M.D. et al. The Werner syndrome protein is a DNA helicase. *Nat Genet* **17**, 100-3 (1997).
138. Sturzenegger, A. et al. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J Biol Chem* **289**, 27314-26 (2014).
139. Thangavel, S. et al. DNA2 drives processing and restart of reversed replication forks in human cells. *J Cell Biol* **208**, 545-62 (2015).
140. Constantinou, A. et al. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep* **1**, 80-4 (2000).
141. Kusumoto-Matsuo, R. et al. 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 (Albany NY)* **6**, 70-81 (2014).
142. Kusumoto, R. et al. Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. *Biochemistry* **47**, 7548-56 (2008).
143. Cooper, M.P. et al. Ku complex interacts with and stimulates the Werner protein. *Genes Dev* **14**, 907-12 (2000).
144. Ahn, B., Harrigan, J.A., Indig, F.E., Wilson, D.M., 3rd & Bohr, V.A. Regulation of WRN helicase activity in human base excision repair. *J Biol Chem* **279**, 53465-74 (2004).

145. Trego, K.S. et al. The DNA repair endonuclease XPG interacts directly and functionally with the WRN helicase defective in Werner syndrome. *Cell Cycle* **10**, 1998-2007 (2011).
146. Kamath-Loeb, A.S., Lan, L., Nakajima, S., Yasui, A. & Loeb, L.A. Werner syndrome protein interacts functionally with translesion DNA polymerases. *Proc Natl Acad Sci U S A* **104**, 10394-9 (2007).
147. Opresko, P.L. Telomere ResQue and preservation--roles for the Werner syndrome protein and other RecQ helicases. *Mech Ageing Dev* **129**, 79-90 (2008).
148. Brosh, R.M., Jr., Opresko, P.L. & Bohr, V.A. Enzymatic mechanism of the WRN helicase/nuclease. *Methods Enzymol* **409**, 52-85 (2006).
149. Crabbe, L., Verdun, R.E., Haggblom, C.I. & Karlseder, J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science* **306**, 1951-3 (2004).
150. Yankiwski, V., Marciniak, R.A., Guarente, L. & Neff, N.F. Nuclear structure in normal and Bloom syndrome cells. *Proc Natl Acad Sci U S A* **97**, 5214-9 (2000).
151. On cataract in conjunction with scleroderma. Otto Werner, doctoral dissertation, 1904, Royal Ophthalmology Clinic, Royal Christian Albrecht University of Kiel. *Adv Exp Med Biol* **190**, 1-14 (1985).
152. Friedrich, K. et al. WRN mutations in Werner syndrome patients: genomic rearrangements, unusual intronic mutations and ethnic-specific alterations. *Hum Genet* **128**, 103-11 (2010).
153. Oshima, J., Campisi, J., Tannock, T.C. & Martin, G.M. Regulation of c-fos expression in senescing Werner syndrome fibroblasts differs from that observed in senescing fibroblasts from normal donors. *J Cell Physiol* **162**, 277-83 (1995).
154. Salk, D., Bryant, E., Hoehn, H., Johnston, P. & Martin, G.M. Growth characteristics of Werner syndrome cells in vitro. *Adv Exp Med Biol* **190**, 305-11 (1985).
155. Melcher, R. et al. Spectral karyotyping of Werner syndrome fibroblast cultures. *Cytogenet Cell Genet* **91**, 180-5 (2000).
156. Saintigny, Y., Makienko, K., Swanson, C., Emond, M.J. & Monnat, R.J., Jr. Homologous recombination resolution defect in werner syndrome. *Mol Cell Biol* **22**, 6971-8 (2002).
157. Goto, M., Miller, R.W., Ishikawa, Y. & Sugano, H. Excess of rare cancers in Werner syndrome (adult progeria). *Cancer Epidemiol Biomarkers Prev* **5**, 239-46 (1996).
158. Lauper, J.M., Krause, A., Vaughan, T.L. & Monnat, R.J., Jr. Spectrum and risk of neoplasia in Werner syndrome: a systematic review. *PLoS One* **8**, e59709 (2013).
159. Agrelo, R. et al. Epigenetic inactivation of the premature aging Werner syndrome gene in human cancer. *Proc Natl Acad Sci U S A* **103**, 8822-7 (2006).
160. Slupianek, A. et al. BCR/ABL stimulates WRN to promote survival and genomic instability. *Cancer Res* **71**, 842-51 (2011).
161. Sakamoto, S. et al. Werner helicase relocates into nuclear foci in response to DNA damaging agents and co-localizes with RPA and Rad51. *Genes Cells* **6**, 421-30 (2001).
162. Futami, K., Takagi, M., Shimamoto, A., Sugimoto, M. & Furuichi, Y. Increased chemotherapeutic activity of camptothecin in cancer cells by siRNA-induced silencing of WRN helicase. *Biol Pharm Bull* **30**, 1958-61 (2007).
163. Aggarwal, M., Sommers, J.A., Shoemaker, R.H. & Brosh, R.M., Jr. Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. *Proc Natl Acad Sci U S A* **108**, 1525-30 (2011).
164. Moser, R. et al. MYC-driven tumorigenesis is inhibited by WRN syndrome gene deficiency. *Mol Cancer Res* **10**, 535-45 (2012).
165. Huang, S. et al. The spectrum of WRN mutations in Werner syndrome patients. *Hum Mutat* **27**, 558-67 (2006).
166. Chang, S. et al. Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nat Genet* **36**, 877-82 (2004).
167. Cheung, H.H. et al. Telomerase protects werner syndrome lineage-specific stem cells from premature aging. *Stem Cell Reports* **2**, 534-46 (2014).
168. Sangrithi, M.N. et al. Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell* **121**, 887-98 (2005).

169. Marino, F., Vindigni, A. & Onesti, S. Bioinformatic analysis of RecQ4 helicases reveals the presence of a RQC domain and a Zn knuckle. *Biophys Chem* **177-178**, 34-9 (2013).
170. Matsuno, K., Kumano, M., Kubota, Y., Hashimoto, Y. & Takisawa, H. The N-terminal noncatalytic region of Xenopus RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. *Mol Cell Biol* **26**, 4843-52 (2006).
171. Xu, X., Rochette, P.J., Feyissa, E.A., Su, T.V. & Liu, Y. MCM10 mediates RECQ4 association with MCM2-7 helicase complex during DNA replication. *EMBO J* **28**, 3005-14 (2009).
172. Thangavel, S. et al. Human RECQ1 and RECQ4 helicases play distinct roles in DNA replication initiation. *Mol Cell Biol* **30**, 1382-96 (2010).
173. Kitao, S. et al. Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes. *Genomics* **54**, 443-52 (1998).
174. De, S. et al. RECQL4 is essential for the transport of p53 to mitochondria in normal human cells in the absence of exogenous stress. *J Cell Sci* **125**, 2509-22 (2012).
175. Gupta, S. et al. RECQL4 and p53 potentiate the activity of polymerase gamma and maintain the integrity of the human mitochondrial genome. *Carcinogenesis* **35**, 34-45 (2014).
176. Burks, L.M., Yin, J. & Plon, S.E. Nuclear import and retention domains in the amino terminus of RECQL4. *Gene* **391**, 26-38 (2007).
177. Dietschy, T. et al. p300-mediated acetylation of the Rothmund-Thomson-syndrome gene product RECQL4 regulates its subcellular localization. *J Cell Sci* **122**, 1258-67 (2009).
178. Woo, L.L., Futami, K., Shimamoto, A., Furuichi, Y. & Frank, K.M. The Rothmund-Thomson gene product RECQL4 localizes to the nucleolus in response to oxidative stress. *Exp Cell Res* **312**, 3443-57 (2006).
179. Keller, H. et al. The intrinsically disordered amino-terminal region of human RecQL4: multiple DNA-binding domains confer annealing, strand exchange and G4 DNA binding. *Nucleic Acids Res* **42**, 12614-27 (2014).
180. Marino, F. et al. Structural and biochemical characterization of an RNA/DNA binding motif in the N-terminal domain of RecQ4 helicases. *Sci Rep* **6**, 21501 (2016).
181. Sedlackova, H., Cechova, B., Mlcouskova, J. & Krejci, L. RECQ4 selectively recognizes Holliday junctions. *DNA Repair (Amst)* **30**, 80-9 (2015).
182. Abe, T. et al. The N-terminal region of RECQL4 lacking the helicase domain is both essential and sufficient for the viability of vertebrate cells. Role of the N-terminal region of RECQL4 in cells. *Biochim Biophys Acta* **1813**, 473-9 (2011).
183. Siitonen, H.A. et al. The mutation spectrum in RECQL4 diseases. *Eur J Hum Genet* **17**, 151-8 (2009).
184. Kitao, S. et al. Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nat Genet* **22**, 82-4 (1999).
185. Siitonen, H.A. et al. Molecular defect of RAPADILINO syndrome expands the phenotype spectrum of RECQL diseases. *Hum Mol Genet* **12**, 2837-44 (2003).
186. Van Maldergem, L. et al. Revisiting the craniosynostosis-radial ray hypoplasia association: Baller-Gerold syndrome caused by mutations in the RECQL4 gene. *J Med Genet* **43**, 148-52 (2006).
187. Wang, L.L. et al. Clinical manifestations in a cohort of 41 Rothmund-Thomson syndrome patients. *Am J Med Genet* **102**, 11-7 (2001).
188. Mirabello, L., Troisi, R.J. & Savage, S.A. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. *Int J Cancer* **125**, 229-34 (2009).
189. Wang, L.L. et al. Association between osteosarcoma and deleterious mutations in the RECQL4 gene in Rothmund-Thomson syndrome. *J Natl Cancer Inst* **95**, 669-74 (2003).
190. Gripp, K.W. et al. Phenotype of the fibroblast growth factor receptor 2 Ser351Cys mutation: Pfeiffer syndrome type III. *Am J Med Genet* **78**, 356-60 (1998).

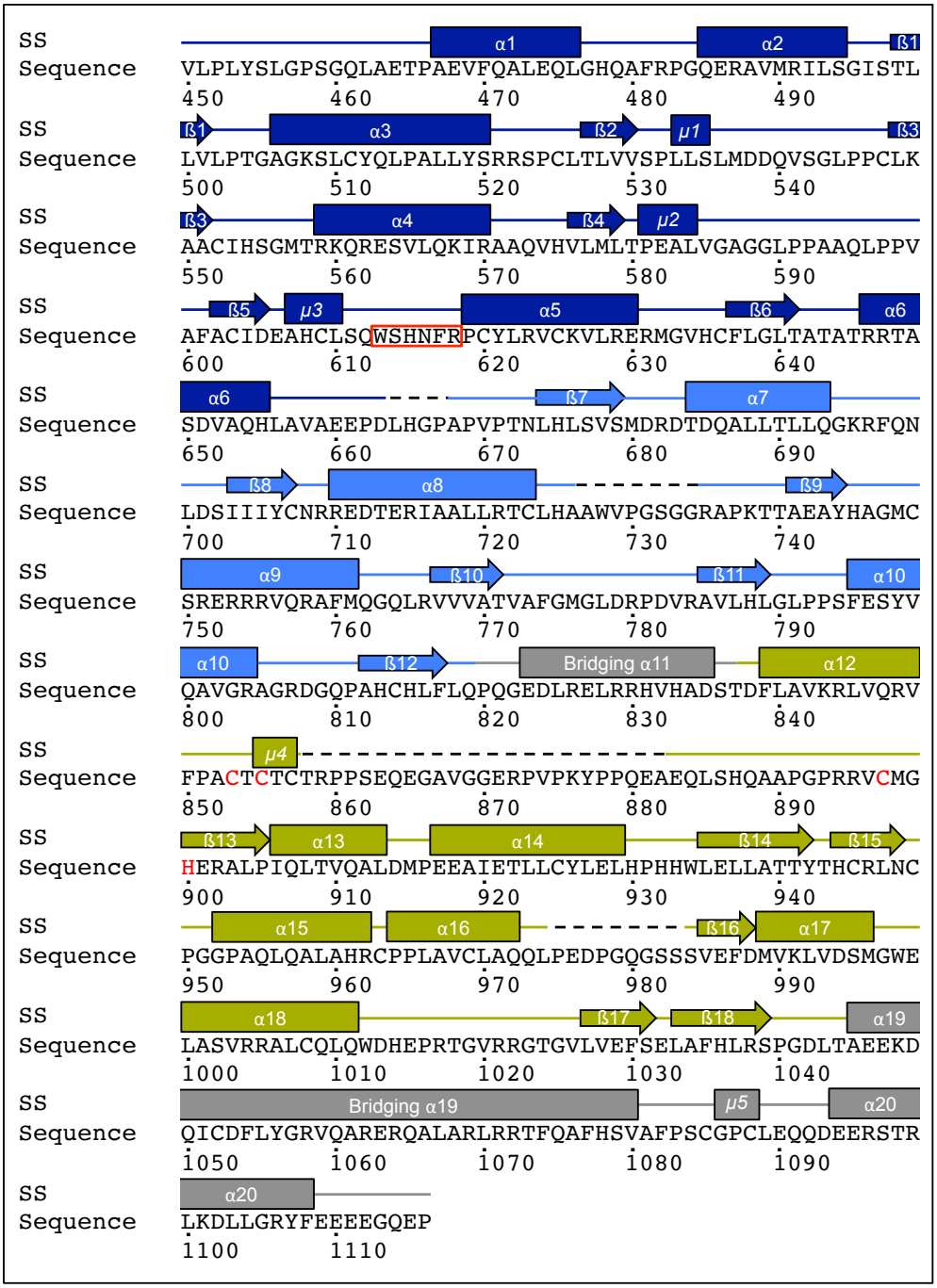
191. Gripp, K.W. et al. TWIST gene mutation in a patient with radial aplasia and craniosynostosis: further evidence for heterogeneity of Baller-Gerold syndrome. *Am J Med Genet* **82**, 170-6 (1999).
192. Ghosh, A.K. et al. RECQL4, the protein mutated in Rothmund-Thomson syndrome, functions in telomere maintenance. *J Biol Chem* **287**, 196-209 (2012).
193. Petkovic, M., Dietschy, T., Freire, R., Jiao, R. & Stagljar, I. The human Rothmund-Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. *J Cell Sci* **118**, 4261-9 (2005).
194. Singh, D.K. et al. The involvement of human RECQL4 in DNA double-strand break repair. *Aging Cell* **9**, 358-71 (2010).
195. Schurman, S.H. et al. Direct and indirect roles of RECQL4 in modulating base excision repair capacity. *Hum Mol Genet* **18**, 3470-83 (2009).
196. Jin, W. et al. Sensitivity of RECQL4-deficient fibroblasts from Rothmund-Thomson syndrome patients to genotoxic agents. *Hum Genet* **123**, 643-53 (2008).
197. Rossi, M.L., Ghosh, A.K., Kulikowicz, T., Croteau, D.L. & Bohr, V.A. Conserved helicase domain of human RecQ4 is required for strand annealing-independent DNA unwinding. *DNA Repair (Amst)* **9**, 796-804 (2010).
198. Ying, K.L., Oizumi, J. & Curry, C.J. Rothmund-Thomson syndrome associated with trisomy 8 mosaicism. *J Med Genet* **27**, 258-60 (1990).
199. Der Kaloustian, V.M., McGill, J.J., Vekemans, M. & Kopelman, H.R. Clonal lines of aneuploid cells in Rothmund-Thomson syndrome. *Am J Med Genet* **37**, 336-9 (1990).
200. Fan, W. & Luo, J. RecQ4 facilitates UV light-induced DNA damage repair through interaction with nucleotide excision repair factor xeroderma pigmentosum group A (XPA). *J Biol Chem* **283**, 29037-44 (2008).
201. Lu, H. et al. RECQL4 Promotes DNA End Resection in Repair of DNA Double-Strand Breaks. *Cell Rep* **16**, 161-73 (2016).
202. Shamanna, R.A. et al. RECQ helicase RECQL4 participates in non-homologous end joining and interacts with the Ku complex. *Carcinogenesis* **35**, 2415-24 (2014).
203. Kohzaki, M. et al. The helicase domain and C-terminus of human RecQL4 facilitate replication elongation on DNA templates damaged by ionizing radiation. *Carcinogenesis* **33**, 1203-10 (2012).
204. Su, Y. et al. Human RecQL4 helicase plays critical roles in prostate carcinogenesis. *Cancer Res* **70**, 9207-17 (2010).
205. Bakhanashvili, M. et al. p53 in mitochondria enhances the accuracy of DNA synthesis. *Cell Death Differ* **15**, 1865-74 (2008).
206. Lu, L. et al. RECQL4 Regulates p53 Function In Vivo During Skeletogenesis. *J Bone Miner Res* **30**, 1077-89 (2015).
207. Choi, Y.W. et al. Gene expression profiles in squamous cell cervical carcinoma using array-based comparative genomic hybridization analysis. *Int J Gynecol Cancer* **17**, 687-96 (2007).
208. Yokota, T. et al. Frequent multiplication of chromosomal region 8q24.1 associated with aggressive histologic types of breast cancers. *Cancer Lett* **139**, 7-13 (1999).
209. Ioannidis, P. et al. 8q24 Copy number gains and expression of the c-myc mRNA stabilizing protein CRD-BP in primary breast carcinomas. *Int J Cancer* **104**, 54-9 (2003).
210. Shimamoto, A., Nishikawa, K., Kitao, S. & Furuichi, Y. Human RecQ5beta, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3alpha and 3beta. *Nucleic Acids Res* **28**, 1647-55 (2000).
211. Ren, H. et al. 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. *Biochem J* **412**, 425-33 (2008).
212. Popuri, V., Tadokoro, T., Croteau, D.L. & Bohr, V.A. Human RECQL5: guarding the crossroads of DNA replication and transcription and providing backup capability. *Crit Rev Biochem Mol Biol* **48**, 289-99 (2013).

213. Aygun, O., Svejstrup, J. & Liu, Y. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proc Natl Acad Sci U S A* **105**, 8580-4 (2008).
214. Islam, M.N., Fox, D., 3rd, Guo, R., Enomoto, T. & Wang, W. RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. *Mol Cell Biol* **30**, 2460-72 (2010).
215. Kassube, S.A., Jinek, M., Fang, J., Tsutakawa, S. & Nogales, E. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nat Struct Mol Biol* **20**, 892-9 (2013).
216. Izumikawa, K. et al. Association of human DNA helicase RecQ5beta with RNA polymerase II and its possible role in transcription. *Biochem J* **413**, 505-16 (2008).
217. Li, M., Xu, X. & Liu, Y. The SET2-RPB1 interaction domain of human RECQ5 is important for transcription-associated genome stability. *Mol Cell Biol* **31**, 2090-9 (2011).
218. Rudolph, C.J., Dhillon, P., Moore, T. & Lloyd, R.G. Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair (Amst)* **6**, 981-93 (2007).
219. Garcia, P.L., Liu, Y., Jiricny, J., West, S.C. & Janscak, P. Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *EMBO J* **23**, 2882-91 (2004).
220. Ghosh, A., Rossi, M.L., Aulds, J., Croteau, D. & Bohr, V.A. Telomeric D-loops containing 8-oxo-2'-deoxyguanosine are preferred substrates for Werner and Bloom syndrome helicases and are bound by POT1. *J Biol Chem* **284**, 31074-84 (2009).
221. Speina, E. et al. Human RECQL5beta stimulates flap endonuclease 1. *Nucleic Acids Res* **38**, 2904-16 (2010).
222. Kanagaraj, R., Saydam, N., Garcia, P.L., Zheng, L. & Janscak, P. Human RECQ5beta helicase promotes strand exchange on synthetic DNA structures resembling a stalled replication fork. *Nucleic Acids Res* **34**, 5217-31 (2006).
223. Ramamoorthy, M. et al. RECQL5 cooperates with Topoisomerase II alpha in DNA decatenation and cell cycle progression. *Nucleic Acids Res* **40**, 1621-35 (2012).
224. Hu, Y. et al. Recql5 and Blm RecQ DNA helicases have nonredundant roles in suppressing crossovers. *Mol Cell Biol* **25**, 3431-42 (2005).
225. Hu, Y. et al. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev* **21**, 3073-84 (2007).
226. Popuri, V. et al. Recruitment and retention dynamics of RECQL5 at DNA double strand break sites. *DNA Repair (Amst)* **11**, 624-35 (2012).
227. Zheng, L. et al. MRE11 complex links RECQ5 helicase to sites of DNA damage. *Nucleic Acids Res* **37**, 2645-57 (2009).
228. Kanagaraj, R. et al. RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. *Nucleic Acids Res* **38**, 8131-40 (2010).
229. Tadokoro, T. et al. Human RECQL5 participates in the removal of endogenous DNA damage. *Mol Biol Cell* **23**, 4273-85 (2012).
230. Popuri, V. et al. RECQL5 plays co-operative and complementary roles with WRN syndrome helicase. *Nucleic Acids Res* **41**, 881-99 (2013).
231. Rodriguez-Lopez, A.M. et al. Characterisation of the interaction between WRN, the helicase/exonuclease defective in progeroid Werner's syndrome, and an essential replication factor, PCNA. *Mech Ageing Dev* **124**, 167-74 (2003).
232. Pike, A.C. et al. Human RECQ1 helicase-driven DNA unwinding, annealing, and branch migration: insights from DNA complex structures. *Proc Natl Acad Sci U S A* **112**, 4286-91 (2015).
233. Sharma, S., Doherty, K.M. & Brosh, R.M., Jr. Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. *Biochem J* **398**, 319-37 (2006).
234. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
235. Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).

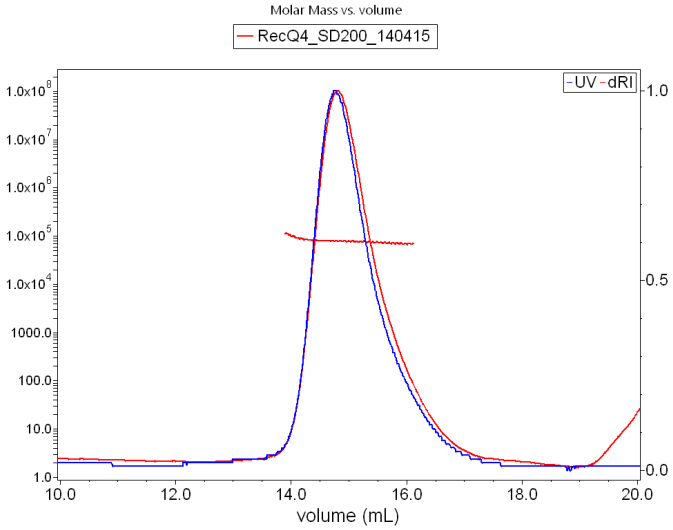
236. Skubak, P. & Pannu, N.S. Automatic protein structure solution from weak X-ray data. *Nat Commun* **4**, 2777 (2013).
237. Pannu, N.S. et al. Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr D Biol Crystallogr* **67**, 331-7 (2011).
238. de Graaff, R.A., Hilge, M., van der Plas, J.L. & Abrahams, J.P. Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data. *Acta Crystallogr D Biol Crystallogr* **57**, 1857-62 (2001).
239. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-67 (2011).
240. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr D Biol Crystallogr* **66**, 470-8 (2010).
241. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* **62**, 1002-11 (2006).
242. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
243. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
244. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. & Iyer, L.M. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* **29**, 231-62 (2005).
245. Harami, G.M., Gyimesi, M. & Kovacs, M. From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem Sci* **38**, 364-71 (2013).
246. Sami, F., Gary, R.K., Fang, Y. & Sharma, S. Site-directed mutants of human RECQ1 reveal functional importance of the zinc binding domain. *Mutat Res* **790**, 8-18 (2016).
247. Guo, R.B., Rigolet, P., Zargarian, L., Fermandjian, S. & Xi, X.G. Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase. *Nucleic Acids Res* **33**, 3109-24 (2005).
248. Liu, J.L., Rigolet, P., Dou, S.X., Wang, P.Y. & Xi, X.G. The zinc finger motif of Escherichia coli RecQ is implicated in both DNA binding and protein folding. *J Biol Chem* **279**, 42794-802 (2004).

Appendix

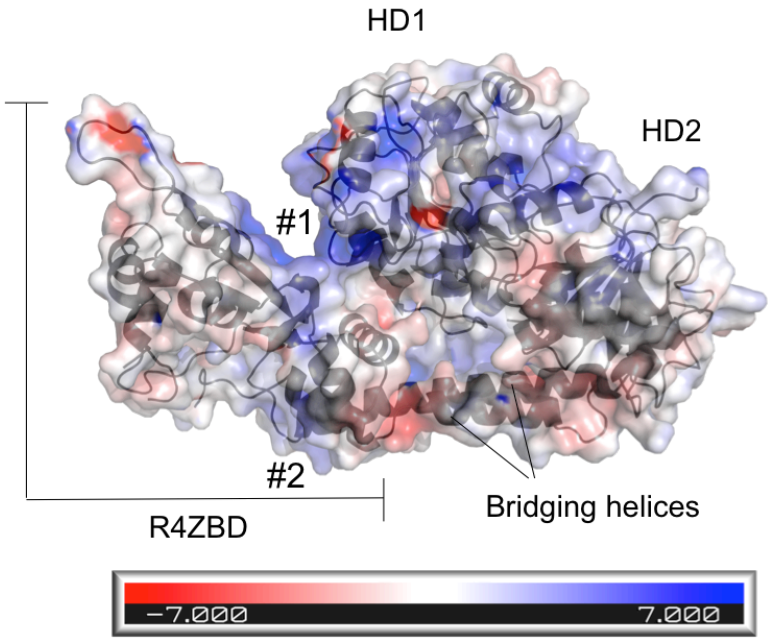
Supplementary Data



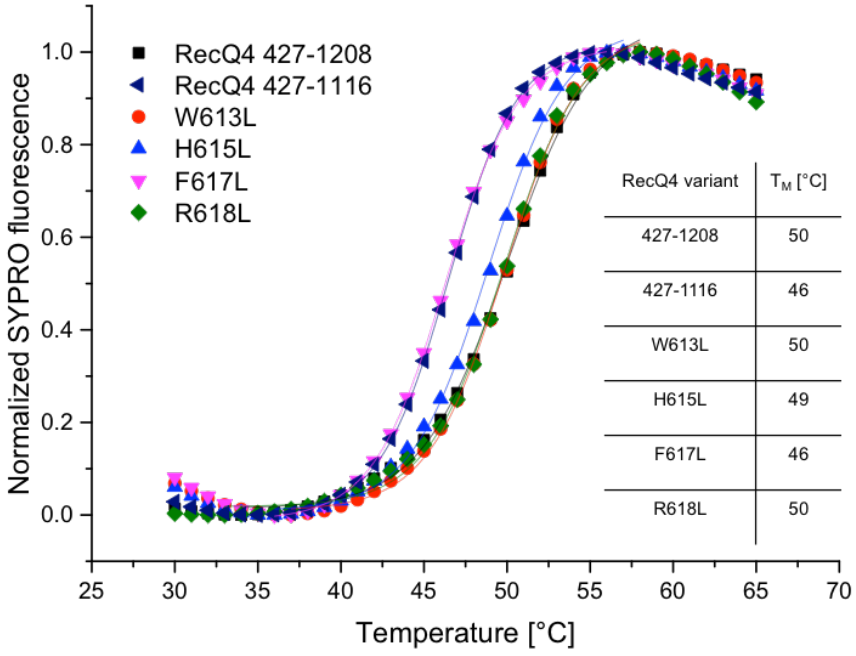
Annotation of secondary structure elements.
Annotation of secondary structure elements within the RecQ4$^{427\text{-}1116}$ structure based on the DSSP algorithm. Domain color code as in Fig. 4-8. Briefly: dark blue – HD1, light blue – HD2, olive – R4ZBD, gray – bridging helices. Helices are depicted as boxes. β-strands are represented by arrows. 3-10-helices are indicated by the symbol μ in italic. The ARL-sequence is highlighted by a red box and the Zn^{2+}-coordinating residues are depicted in red. Regions which are not resolved in the structure are indicated by dashed lines. This Figure has been adapted from the supplementary information to Kaiser *et al.*, 2017.



SEC-MALS analysis of the RecQ4$^{427\text{-}1116}$ variant.
Analytical SEC was performed in storage buffer, using a Superdex200 10/300 GL column at room temperature. The RI signal indicates homogenious protein with a determined molecular weight (M_W) of 77.54 kDa. With a calculated M_W of 76.3 kDa, the SEC-MALS analysis confirms that RecQ4$^{427\text{-}1116}$ is a monomer in solution.



Electrostatic surface potential map of the RecQ4$^{427\text{-}1116}$ model.
The electrostatic surface potential was calculated using the APBS plugin in PyMOL. Main RecQ4$^{427\text{-}1116}$ domains are indicated. Electronegative areas are shown in red. Blue color illustrates an electropositive surface potential. The model features a large electropositive surface area directly between the R4ZBD and the HD1 (#1). A second electropositive patch is located at the bottom of the lower R4ZBD half (#2). This Figure has been adapted from the supplementary information to Kaiser *et al.*, 2017.



Thermal stability analysis (Thermofluor assay) of the RecQ4 ARL variants.
Melting curves for all tested RecQ4 ARL variants display the gradual unfolding with increasing temperature. The melting temperature at a 50% unfolded state is represented by the T_M value. The melting curves indicate that all RecQ4 variants are properly folded as all variants display regular denaturing curves. The ARL variant F617L and RecQ4$^{427\text{-}1116}$ are less stable compared to other RecQ4 variants, indicated by their decreased T_M value. This Figure has been adapted from the supplementary information to Kaiser *et al.*, 2017.

Publications

The structural and functional characterization of human RecQ4 reveals insights into its helicase mechanism.
Kaiser S, Sauer F, Kisker C
Nat. Commun. 8, 15907 (2017)
doi: 10.1038/ncomms15907

Identification of Multifaceted Binding Modes for Pyrin and ASC Pyrin Domains Gives Insights into Pyrin Inflammasonme Assembly.
Vajjhala PR, Kaiser S, Smith SJ, Ong QR, Soh SL, Stacey KJ, Hill JM
J. Biol. Chem. 289, 23504-23519 (2014)
doi: 10.1074/jbc.M114.553305

Presentations and participations

(1) Poster presentation at the 7th International GSLS Student Symposium "EPOS" – 16/17.10.2012, Würzburg, Germany

(2) EMBO Workshop: High-Throughput Protein Production and Crystallization – 16-25.05.2013, Harwell, Oxford, UK

(3) Poster presentation at the 8th International GSLS Student Symposium "Scientific Crosstalk" – 09/10.10.2013, Würzburg, Germany

(4) Poster presentation at the 9th International GSLS Student Symposium "EUREKA! 2014" – 14/15.10.2014, Würzburg, Germany

(5) Participation at the 10th International GSLS Student Symposium "EUREKA! 2015" – 15/15.10.2015, Würzburg, Germany

(6) Poster presentation at FASEB Meeting "Helicases and Nucleic acid based machines: From mechanisms to insight into disease"
– 26-31.07.2015, Steamboat Springs, CO, USA

(7) Speaker at the 25th Annual Meeting of the German Crystallographic Society – 27-30.03.2017, Karlsruhe, Germany
A RecQ helicase in disguise – How the unconventional structure of the human RecQ4 helicase defines its special position in the RecQ family

Contributions

The here presented work was performed entirely by myself with the following exceptions:

(1) Processing of X-ray diffraction data, model building and structure refinement were performed by Dr. Florian Sauer (Research group of Prof. Caroline Kisker, Structural Biology, University of Würzburg)
(2) Several diffraction data sets that were used in this study were collected for me by my colleague Wolfgang Kölmel

Acknowledgements

At this point, I would like to express my gratitude to the numerous people that had an influence on this work as much as on myself. I thank Caroline Kisker for accepting me in her research group, her patience and her professional support over the last years, including many stimulating discussions about the wonderful world of DNA repair and RecQ helicases. I am very happy and grateful for my time with our PostDoc Dr. Florian Sauer, without whom this project would not have become what it is and who was, and still is, an inexhaustible source for strategies and solutions to every scientific problem you might have. I want to thank the wonderful people of our structural biology department, especially my colleague and flat mate Wolfgang Kölmel, as well as all the other people who provided such welcome and supportive atmosphere, at work and beyond. I want to thank our IT department, Dr. Bernhard Fröhlich and colleagues, for providing an incredible infrastructure and last-resort assistance with the occasional data rescue procedure. Finally, I want to thank the members of my thesis committee for their professional support, and I would like to acknowledge the great work of the administrative staff of the Graduate School of Life Sciences and say thanks for all the help you could want at the beginning, during and at the end of your PhD thesis time.

Last but not least, I am very grateful for the unconditional and everlasting support of my parents, Kerstin and Jürgen Kaiser.

Curriculum Vitae

Affidavit

I hereby confirm that my thesis entitled “A RecQ helicase in disguise: Characterization of the unconventional Structure and Function of the human Genome Caretaker RecQ4” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Wuerzburg, 29.09.2017

(Signature)

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass die Dissertation „Die unkonventionelle RecQ Helikase RecQ4: Charakterisierung der ungewöhnliche Struktur und Funktion eines essentiellen Beschützers des menschlichen Genoms“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eine kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 29.09.2017

(Unterschrift)